To be submitted to *IEEE Trans. Inform. Theory*
Draft of 11 Jan 2001 12:17 p.m.

Coding Theorems for Turbo Code Ensembles*

Hui Jin and Robert J. McEliece
Deprtment of Electrical Engineering
California Institute of Technology

Abstract.

This paper is devoted to a Shannon-theoretic study of turbo codes. We prove that ensembles of parallel and serial turbo codes are "good" in the following sense. For a turbo code ensemble defined by a fixed set of component codes (subject only to mild necessary restrictions), there exists a positive number γ_0 such that for any binary-input memoryless channel whose Bhattacharyya noise parameter is less than γ_0, the average maximum-likelihood decoder error probability approaches zero, at least as fast as $n^{-\beta}$, where β is the "interleaver gain" exponent defined by Benedetto et al. in 1996.

1. Introduction.

The invention of turbo codes in 1993 [5], and the explosion of research that followed, has revolutionized every aspect of channel coding. Turbo codes appear to offer nothing less than a solution to the challenge issued by Shannon in 1948 [25]: to devise practical methods of communicating reliably at rates near channel capacity. And while there has been a good deal of excellent theoretical work on turbo codes, it seems fair to say that practice still leads theory by a considerable margin. In particular, there has been little previous Shannon-theoretic work on turbo codes. By "Shannon-theoretic" we mean a study of the average performance of the codes in the turbo-code ensemble under maximum-likelihood decoding. Of course, there is little possibility that maximum-likelihood decoding of turbo codes can be implemented practically, but since the turbo decoding algorithm seems to be, in most cases, a close approximation to MLD, it is important to know the MLD potential for this class of codes. In any case, this paper is devoted to a Shannon-theoretic study of turbo codes. In particular, it may be viewed as an elaboration of the following remark, which was made in [21]:

"The presence [in turbo-codes] of the pseudorandom interleavers between the component codes ensures that the resulting overall code behaves very much like a long random code, and by Shannon's theorems, a long random code is likely to be 'good'. . . ."

In this paper, building on ideas pioneered by Benedetto et al. [2, 4], we will prove that turbo codes are indeed good, in the following sense. For any turbo code ensemble, parallel or serial, defined by a fixed set of component codes (subject only to to mild necessary restrictions), there exists a positive number γ_0, such that on any binary-input memoryless

* This research was supported by NSF grant no. CCR-9804793, and grants from Sony, Qualcomm, and Caltech's Lee Center for Advanced Networking.

channel whose Bhattacharyya noise parameter is less than γ_0, the average maximum-likelihood decoder error probability approaches zero, at least as fast as $n^{-\beta}$, where β is the (ensemble-dependent) "interleaver gain" exponent defined by Benedetto et al. in 1996 [2]. (For an exact statement of these results, see Section 8, Theorems 8.1 and 8.4, below.)

Here is an outline of the paper:

• Section 2: A definition of the parallel and serial turbo code ensembles.

• Section 3. A discussion of general code ensembles, and their weight enumerators.

• Section 4. The Bhattacharyya noise parameter and the union bound, for binary input discrete memoryless channels.

• Section 5. A coding theorem for general code ensembles, combining the ensemble weight enumerator with the union bound.

• Section 6. Estimates (upper bounds) of the weight enumerators of the parallel turbo code ensembles defined in Section 2.

• Section 7. Estimates (upper bounds) of the weight enumerators of the serial turbo code ensembles defined in Section 2.

• Section 8. Statement and proof of the main results.

• Section 9. Examples: The Berrou-Galvieux-Thitimajshima ensemble, and the ensemble of RA codes.

• Section 10. Discussion and conclusions.

• Appendix A. Combinatorial facts about convolutional codes.

• Appendix B. Some useful inequalities.

• Appendix C. Extension of main theorems to bit error probability.

2. The Turbo Code Ensembles.

The general structure of a parallel turbo code is shown in Figure 1. There are J interleavers (pseudorandom scramblers) $P_1, P_2, \ldots, P_J$ and J recursive convolutional encoders $E_1, E_2, \ldots, E_J$. An information block of length k is scrambled by interleaver P_i and then encoded (and truncated) by E_i, producing a codeword of length n_i, for $i = 1, 2, \ldots, J$. These J codewords are then sent to the channel. The overall code is therefore a (n, k) linear block code, with $n = \sum_{i=1}^{J} n_i$. If $R_i = k/n_i$ is the rate of the ith component code C_i, then overall code rate is easily seen to be $R = (\sum_{i=1}^{J} R_i^{-1})^{-1}$. Because there are $k!$ choices for each interleaver,[1] there are a large number of codes with the structure shon in Figure 1. We call this set of codes the $[E_1 \| E_2 \| \cdots \| E_J]$ ensemble. (We will define a code ensemble more precisely in Section 3, below.)

---

[1] Without loss of generality, we may assume that P_1 is the identity permutation, so that there are really only $J - 1$ interleavers.

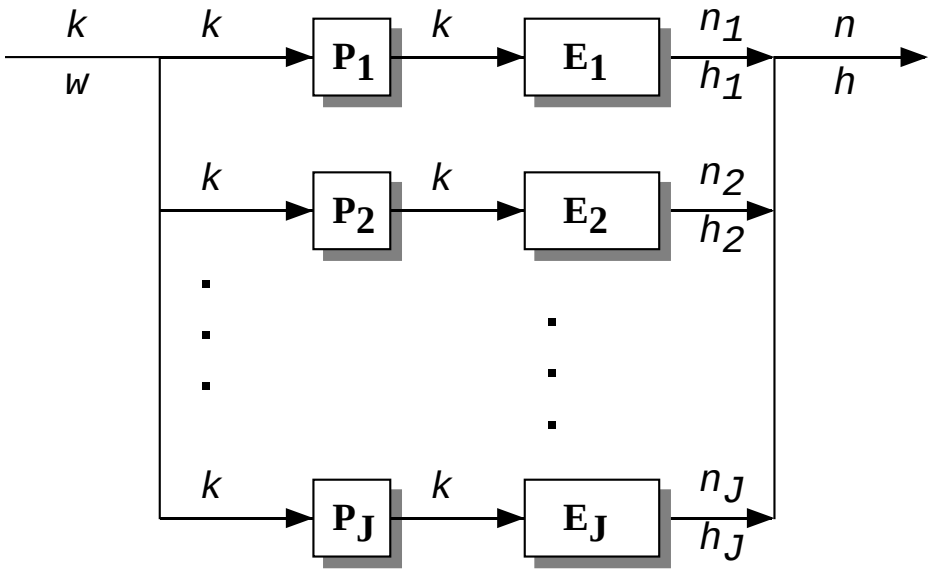

Figure 1. Encoder for a parallel turbo code. The numbers above the input-output lines indicate the length of the corresponding block, and those below the lines indicate (when present) the Hamming weight of the block.

Our first main result (Theorem 8.1) is that if $J \geq 2$, the $[E_1 \| E_2 \| \cdots \| E_J]$ ensemble is "good," in the sense defined in Section 1.

A serial turbo code has the general structure shown in Figure 2. An information block of length k is encoded by an outer encoder E_1 into a codeword of length N, which is scrambled by an interleaver P, and then encoded by an inner encoder E_2 into a codeword of length n. The outer code C_1 is a truncated convolutional code,[2] and the inner code C_2 is a truncated recursive convolutional code. The overall code is therefore an (n, k) linear block code, with rate $R = R_1 R_2$, where R_1 is the rate of the outer code and R_2 is the rate of the inner code. Because of the choices for the interleaver, there are $N!$ codes with the structure shown in Figure 2. We call this set of codes the $[E_1 \Rightarrow E_2]$ ensemble.

Our second main result (Theorem 8.4) is that if the minimum distance of the outer code C_1 is at least three, the $[E_1 \Rightarrow E_2]$ ensemble is also "good."

3. Code Ensembles, in General.

Parallel and serial turbo codes are important examples of code ensembles, but our results can be applied to other ensembles, as well. In this section we will give a general definition of a code ensemble.

[2] We note that a block code can be viewed as a convolutional code without memory, so that E_1 may be a block encoder.

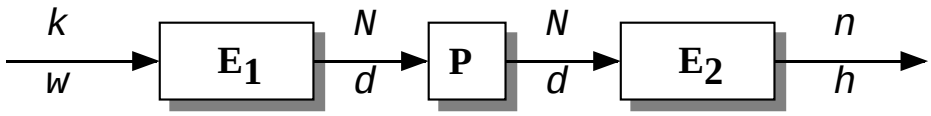

Figure 2. Encoder for a serial turbo code. As in Figure 1, the numbers above the input-output lines indicate the length of the corresponding block, and those below the lines indicate the Hamming weight of the block.

By an *ensemble* of linear codes, then, we mean a sequence $\mathcal{C}_{n_1}, \mathcal{C}_{n_2}, \ldots$ of sets of linear codes, where $\mathcal{C}_{n_i}$ is a set of (n_i, k_i) codes with common rate $R_i = k_i/n_i$. We assume that the sequence $n_1, n_2, \ldots$ approaches infinity, and that $\lim_{i\to\infty} R_i = R$, where R is called the rate of the ensemble.

We shall be concerned with the weight structure of the ensemble, and with this in mind we introduce some notation. If C is an (n, k) linear code, we denote its weight enumerator by the list $A_0(C), A_1(C), \ldots, A_n(C)$. In other words, $A_h(C)$ is the number of words of weight h in C, for $h = 0, 1, \ldots, n$. When no ambiguity is likely to occur, we denote the weight enumerator simply by $A_0, A_1, \ldots, A_n$. We will also need the *cumulative weight enumerator*

$$A_{\le h} = \sum_{d=1}^{h} A_d \qquad \text{for } h = 1, \ldots, n. \tag{3.1}$$

In words, $A_{\le h}$ is the number of *nonzero* codewords of weight $\le h$.

When the code C is viewed as the set of possible outputs of a particular encoder E, we denote by $A_{w,h}^{(E)}$ the number of $(\boldsymbol{x}, \boldsymbol{y})$ pairs where the encoder input $\boldsymbol{x}$ has weight w and the corresponding encoder output $\boldsymbol{y}$ (codeword) has weight h. Usually the encoder will be understood, and the simpler notation $A_{w,h}$ will do. The set of numbers $A_{w,h}$ is called the input-output weight enumerator (IOWE) for the code. In analogy with (3.1) we define the cumulative input-output weight enumerator (CIOWE):

$$A_{w,\le h} = \sum_{d=1}^{h} A_{w,d}. \tag{3.2}$$

Returning now to the ensemble, we define the *average weight enumerator* for the set $\mathcal{C}_n$ as the list

$$\overline{A}_0^{(n)}, \overline{A}_1^{(n)}, \ldots, \overline{A}_n^{(n)},$$

where

$$\overline{A}_h^{(n)} \triangleq \frac{1}{|\mathcal{C}_n|} \sum_{C \in \mathcal{C}_n} A_h(C) \qquad \text{for } h = 0, 1, \ldots, n. \tag{3.3}$$

Similarly, we define the average cumulative weight enumerator $\overline{A}^{(n)}_{\leq h}$, the average IOWE $\overline{A}^{(n)}_{w,h}$, and the average CIOWE $\overline{A}^{(n)}_{w,\leq h}$.

For each n in the sequence $n_1, n_2, \ldots$, the nth *spectral shape* function is defined as

$$r_n(\delta) \stackrel{\Delta}{=} \frac{1}{n} \log \overline{A}^{(n)}_{\lfloor \delta n \rfloor} \qquad \text{for } 0 < \delta < 1. \tag{3.4}$$

Thus $\overline{A}^{(n)}_h = e^{n r_n(\delta)}$, where $\delta = h/n$.

Finally, we define the *asymptotic spectral shape* :

$$r(\delta) \stackrel{\Delta}{=} \lim_{n \to \infty} r_n(\delta) \qquad \text{for } 0 < \delta < 1, \tag{3.5}$$

provided the limit exists. In this case, we can say, roughly, that for large n, if the ratio $\delta = h/n$ is fixed, then

$$\overline{A}^{(n)}_h \sim e^{n r(\delta)}.$$

It is worth noting here that the main difficulty in proving our main results (Theorems 8.1 and 8.4) is that we are unable to compute $r(\delta)$ for the $[E_1 \| E_2 \| \cdots \| E_J]$ and $[E_1 \Rightarrow E_2]$ ensembles. Instead, we have had to resort to upper bounds on $r(\delta)$ (see (6.8) and (7.8)), based on the work of Kahale and Urbanke [18], which render our results existence theorems only.

4. Memoryless Binary-Input Channels and the Union Bound.

Since turbo codes, as we have defined them, are binary codes, we consider using them on memoryless binary input channels. Such a channel has binary input alphabet $\{0, 1\}$ and arbitrary output alphabet Ω. If the channel input is a binary random variable X, then the channel output is a random variable Y. If Ω is finite, then Y is characterized by transition probabilities $p(y|0)$, $p(y|1)$, i.e., for $y \in \Omega$,

$$\begin{aligned} p(y|0) &= \Pr\{Y = y | X = 0\} \\ p(y|1) &= \Pr\{Y = y | X = 1\}. \end{aligned}$$

If Ω is a subset of R^r, where R is the real line, then Y is characterized by transition probability densities $p(y|0)$, $p(y|1)$, i.e., if S is a measurable subset of Ω,

$$\begin{aligned} \int_S p(y|0) dy &= \Pr\{Y \in S | X = 0\} \\ \int_S p(y|1) dy &= \Pr\{Y \in S | X = 1\}. \end{aligned}$$

The "noisiness" of the channel can be summarized by the *Bhattacharyya noise parameter* γ, which is defined by

$$\gamma = \sum_{y \in \Omega} \sqrt{p(y|0) p(y|1)}, \tag{4.1}$$

if Ω is finite and

$$\gamma = \int_{\Omega} \sqrt{p(y|0)p(y|1)}dy, \tag{4.2}$$

if $\Omega = R^r$. It is easy to see (by the Cauchy-Schwarz inequality) that $\gamma \leq 1$, with equality if and only if $p(y|0) = p(y|1)$ for all y, in which case the channel has capacity zero.[3]

For example, for a binary erasure channel with erasure probabiity p, we have

$$\gamma_{\mathrm{BEC}} = p.$$

Similarly, for a binary symmetric channel with crossover probability p we have

$$\gamma_{\mathrm{BSC}} = 2\sqrt{p(1-p)}.$$

Also, for the asymmetric "Z" channel, we have

$$\gamma_{\mathrm{Z}} = \sqrt{p}.$$

For an additive gaussian channel with $\Omega = R$ and

$$p(y|0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-1)^2/2\sigma^2}$$
$$p(y|1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y+1)^2/2\sigma^2},$$

a short calculation using (4.2) gives

$$\gamma_{AGC} = e^{-1/2\sigma^2}.$$

As a final example, for the binary input coherent Rayleigh fading channel with perfect channel state information available to the receiver, we have $\Omega = R \times R^+$, and for $(y, a) \in \Omega$

$$p(y, a|0) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-a)^2/2\sigma^2} 2ae^{-a^2}$$
$$p(y, a|1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y+a)^2/2\sigma^2} 2ae^{-a^2}.$$

In this case (4.2) yields

$$\gamma_{RF,CSI} = 1 + \frac{1}{2\sigma^2}.$$

[3] The so-called cutoff rate for the channel is $R_0 = 1 - \log_2(1+\gamma)$, which is positive if and only if the capacity is positive, i.e., $\gamma < 1$.

The importance of γ is that γ^h is an upper bound on the maximum-likelihood decoder error probability for a binary code with two codewords separated by a Hamming distance of h (see [20, Theorem 7.5]). It follows that for an (n, k) binary linear code with A_h codewords of weight h, we have the following upper bound, usually called the union bound, on the ML decoder word error probability:

$$\begin{aligned} P_W &\leq \sum_{h=1}^{n} A_h \gamma^h \\ &= \sum_{h=1}^{n} A_h e^{-\alpha h}, \end{aligned} \tag{4.3}$$

where $\alpha = -\log \gamma \geq 0$ is what we shall call the *noise exponent* for the channel. Since as noted above $\gamma \leq 1$, we have $a \geq 0$, with equality if and only if the channel has zero capacity.. Similarly, we can use the union bound to estimate the ML decoder *bit* error probability:

$$P_b \leq \sum_{h=1}^{n} \sum_{w=1}^{k} \frac{w}{k} A_{w,h} \gamma^h,$$

where $A_{w,k}$ is the input-outut weight enumerator of the code.

Since the union bound is linear on weight enumerators, it also applies to ensembles of codes, with A_h replaced by $\overline{A}_h^{(n)}$, the average number of codewords of weight h in $\mathcal{C}_n$:

$$\overline{P}_W^{(n)} \leq \sum_{h=1}^{n} \overline{A}_h^{(n)} e^{-\alpha h} \tag{4.4}$$

$$= \sum_{h=1}^{n} e^{-n(\alpha\delta - r_n(\delta))}, \tag{4.5}$$

where in (4.5) $\delta = h/n$. For the ensemble bit error probability we have correspondingly

$$\overline{P}_b^{(n)} \leq \sum_{h=1}^{n} \sum_{w=1}^{k} \frac{w}{k} \overline{A}_{w,h}^{(n)} e^{-\alpha h}.$$

5. A Coding Theorem.

In this section, by combining the spectral shape function with the union bound, we obtain an upper bound on the ML decoder word error probability for an ensemble of binary linear codes (Theorem 5.1). It shows that under certain conditions, there exists a threshold c_0 such that if the channel noise exponent α exceeds c_0, the ensemble word error probability approaches 0. We shall see that the low-weight codewords in the ensemble determine whether or not the threshold c_0 is finite.

To begin, we introduce some notation. First, let D_n be a fixed sequence of integers satisfying

$$\frac{D_n}{n^\epsilon} \to 0, \qquad \text{for all } \epsilon > 0 \tag{5.1}$$

$$\frac{\log n}{D_n} \to 0. \tag{5.2}$$

For example, $D_n = \log^2 n$ will do. Second, we define the *noise thresholds* for the ensemble:

$$c_0^{(n)} \triangleq \sup_{D_n/n < \delta \le 1} r_n(\delta)/\delta \tag{5.3}$$

$$c_0 \triangleq \limsup_{n\to\infty} c_0^{(n)}. \tag{5.4}$$

Finally, the nth *innominate sum* is defined as follows:

$$Z^{(n)}(D) \triangleq \sum_{h=1}^{D} \overline{A}_h^{(n)},$$

where D is an integer with $1 \le D \le n$. In words, $Z^{(n)}(D)$ is the average number of words of weight $\le D$ for a code in the set C_n. (Incidentally, it is also an upper bound on the probability that the minimum distance of a code in C_n is $\le D$.)

5.1 Theorem. *Suppose the ensemble threshold c_0 defined in (5.4) is finite, and the channel error exponent α satisfies $\alpha > c_0$. Then if $\overline{P}_W^{(n)}$ denotes the ensemble maximum-likelihood decoder error probability, there exists an integer n_0 and positive constants K and ϵ such that for $n > n_0$,*

$$\overline{P}_W^{(n)} \le Z^{(n)}(D_n) + Ke^{-\epsilon D_n}. \tag{5.5}$$

Proof: Since the channel error exponent α is nonnegative, we have

$$A_h e^{-\alpha h} \le A_h.$$

Therefore by (4.4) and (4.5),

$$\begin{aligned} \overline{P}_W^{(n)} &\le \sum_{h=1}^{D_n} \overline{A}_h^{(n)} + \sum_{h>D_n} \overline{A}_h^{(n)} e^{-\alpha h} \\ &= Z^{(n)}(D_n) + \sum_{h>D_n} e^{-h(\alpha - r_n(\delta)/\delta)}. \end{aligned} \tag{5.6}$$

If $\alpha > c_0$, then there exists an integer n_0, and an $\epsilon > 0$ such that for $n > n_0$, $\alpha - c_0^{(n)} > \epsilon$. Hence for $n > n_0$ and $h > D_n$, we have

$$\alpha - \frac{r_n(\delta)}{\delta} \geq \alpha - c_0^{(n)} > \epsilon,$$

so that

$$e^{-h(\alpha - r_n(\delta)/\delta)} \leq e^{-h\epsilon}. \tag{5.7}$$

Thus

$$\sum_{h > D_n} e^{-h(\alpha - r_n(\delta)/\delta)} \leq \sum_{h > D_n} e^{-h\epsilon} = K e^{-D_n \epsilon}, \tag{5.8}$$

where $K = e^{-\epsilon}/(1 - e^{-\epsilon})$. Substituting (5.8) into (5.6), we have (5.5). ▪

5.2 Corollary. *If, in addition, $Z^{(n)}(D_n) = O(n^{-\beta})$ where $\beta > 0$, then for $\alpha > c_0$,*

$$P_W^{(n)} = O(n^{-\beta}). \tag{5.9}$$

Proof: Note that $n^{-\beta} = e^{-\beta \log n}$. The result now follows from (5.5) and (5.2).

The question as to whether c_0 is finite is partially answered by the following two technical results.

5.3 Theorem. *For a code ensemble $\mathcal{C}$, the code threshold c_0 is finite if and only if for all sequences ϵ_n such that $\epsilon_n > D_n/n$ and $\epsilon_n \to 0$,*

$$c_0' = \lim_{n \to \infty} \sup_{D_n/n < \delta < \epsilon_n} r_n(\delta)/\delta \tag{5.10}$$

is finite.

Proof: Clearly

$$\sup_{D_n/n < \delta < \epsilon_n} \frac{r_n(\delta)}{\delta} \leq \sup_{D_n/n < \delta \leq 1} \frac{r_n(\delta)}{\delta},$$

so that if c_0 as defined in (5.3) is finite, so is c_0', for any choice of ϵ_n.

To complete the proof, we will show that if c_0' is finite, so is c_0, or rather the contrapositive, i.e., $c_0 = \infty$ implies $c_0' = \infty$. If c_0 is infinite, then there is a convergent subsequence $\delta_n \to \delta_0$ such that $D_n/n < \delta_n \leq 1$ with

$$\lim_{n \to \infty} \frac{r_n(\delta_n)}{\delta_n} = \infty. \tag{5.11}$$

If $\delta_0 > 0$, note that $\overline{A}_h^{(n)} \le \binom{n}{h} \le e^{nH(\delta)}$,[4] hence $r_n(\delta) = \log \overline{A}_h^{(n)}/n \le H(\delta)$. Thus

$$\lim_{n\to\infty} \frac{r_n(\delta_n)}{\delta_n} \le \frac{H(\delta_0)}{\delta_0},$$

which contradicts (5.11). Thus $\delta_0 = 0$. Hence if we define $\epsilon_n = \min(2\delta_n, 1)$, we have

$$\sup_{D_n/n<\delta<\epsilon_n} \frac{r_n(\delta)}{\delta} \ge \frac{r_n(\delta_n)}{\delta_n} \to \infty.$$

Thus (5.11) diverges, which shows that c_0' is infinite. ▪

5.4 Corollary. *If there exists a function $s(\delta)$ and constants $\gamma_n = O(D_n/n)$ such that $r_n(\delta) \le \gamma_n + s(\delta)$ for all sufficiently small δ and all sufficiently large n, then the ensemble noise threshold c_0 is finite provided*

$$\limsup_{\delta\to 0} \frac{s(\delta)}{\delta} < \infty. \tag{5.12}$$

Proof: We use Theorem 5.3. Thus let ϵ_n be a sequence such that $\epsilon_n > D_n/n$ and $\epsilon_n \to 0$. Then

$$\begin{aligned}
\lim_{n\to\infty} \sup_{D_n/n<\delta<\epsilon_n} r_n(\delta)/\delta &\le \lim_{n\to\infty} \sup_{D_n/n<\delta<\epsilon_n} (\gamma_n + s(\delta))/\delta \\
&\le \limsup_{n\to\infty} (n\gamma_n/D_n) + \lim_{n\to\infty} \left(\sup_{0<\delta<\epsilon_n} s(\delta)/\delta \right) \\
&\le K + \limsup_{\delta\to 0} s(\delta)/\delta \\
&< \infty.
\end{aligned}$$

Thus by Theorem 5.3, the code threshold c_0 is finite. ▪

6. Weight Enumerator Estimates for Parallel Turbo Code Ensembles.

For the $[E_1 \| E_2 \| \cdots \| E_J]$ ensemble, the average IOWE can be obtained from the IOWEs of the component codes using the "uniform interleaver" technique [2]:

$$\overline{A}_{w,h}^{(n)} = \frac{1}{\binom{k}{w}^{J-1}} \sum_{\sum_{i=1}^J h_i = h} \prod_{i=1}^{J} A_{w,h_i}^{[i]}, \tag{6.1}$$

[4] We have collected several useful inequalities on binomial coefficients in Appendix B.

where $A^{[i]}_{w,h_i}$ is the IOWE for the ith component code C_i (see Figure 1 for notation). Therefore

$$\begin{aligned}
\overline{A}^{(n)}_{w,\le h} &= \frac{1}{\binom{k}{w}^{J-1}} \sum_{\sum_{i=1}^{J} h_i \le h} \prod_{i=1}^{J} A^{[i]}_{w,h_i} \\
&\le \frac{1}{\binom{k}{w}^{J-1}} \sum_{h_1=1}^{h} \cdots \sum_{h_J=1}^{h} \prod_{i=1}^{j} A^{[i]}_{w,h_i} \\
&= \frac{1}{\binom{k}{w}^{J-1}} \prod_{i=1}^{J} (\sum_{h_i=1}^{h} A^{[i]}_{w,h_i}) \\
&= \frac{\prod_{i=1}^{J} A^{[i]}_{w,\le h}}{\binom{k}{w}^{J-1}}. \qquad (6.2)
\end{aligned}$$

Next, we apply the bound of Theorem A.3 from Appendix A to each $A^{[i]}_{w,\le h}$ in (6.2). The truncation length for each C_i is less than its code length n_i, which in turn is strictly less than $n = \sum_i n_i$. Defining $\eta = \max_i \eta_i$, and noting that the binomial coefficient $\binom{n}{j}$ is an increasing function of n, we obtain

$$A^{[i]}_{w,\le h} \le \theta_i^w \sum_{j=0}^{\lfloor w/2 \rfloor} \binom{n}{j} \binom{\eta h}{w-j}. \qquad (6.3)$$

If $\eta h < n$, then by Proposition B.1, $\binom{n}{j}\binom{\eta h}{w-j}$ attains its maximum for $0 \le j \le \lfloor w/2 \rfloor$ at $j = \lfloor w/2 \rfloor$. Thus (provided $h < n/\eta$), each $A^{[i]}_{w,\le h}$ can be bounded as follows:

$$\begin{aligned}
A^{[i]}_{w,\le h} &\le \theta_i^w \left(\lfloor \frac{w}{2} \rfloor + 1 \right) \binom{n}{\lfloor w/2 \rfloor} \binom{\eta h}{\lceil w/2 \rceil} \\
&\le (2\theta_i)^w \binom{n}{\lfloor w/2 \rfloor} \binom{\eta h}{\lceil w/2 \rceil} \qquad (\text{since } \lfloor w/2 \rfloor + 1 \le 2^w). \qquad (6.4)
\end{aligned}$$

Combining (6.2) and (6.4), we obtain

$$\overline{A}^{(n)}_{w,\le h} \le \theta^w \frac{\binom{n}{\lfloor w/2 \rfloor}^J \binom{\eta h}{\lceil w/2 \rceil}^J}{\binom{k}{w}^{J-1}},$$

for some constant $\theta > 1$. Consequently, for small δ, $\overline{A}^{(n)}_{\le h}$ can be upper bounded as follows:

$$\begin{aligned}
\overline{A}^{(n)}_{\le h} &\le \sum_{w=1}^{\mu h} A^{(n)}_{w,\le h} \\
&\le \sum_{w=1}^{\mu h} \theta^w \frac{\binom{n}{\lfloor w/2 \rfloor}^J \binom{\eta h}{\lceil w/2 \rceil}^J}{\binom{k}{w}^{J-1}}. \qquad (6.5)
\end{aligned}$$

(The sum in (6.5) stops at μh rather than k because of Theorem A.1). Equation (6.5) will be used to bound the innominate sum $Z^{(n)}(D_n)$ that appears in Theorem 5.1.

To bound $r_n(\delta)$ for small δ, we simplify (6.5), by replacing the summation with the maximum term times the number of terms. Since $\binom{\eta h}{l} < 2^{\eta h}$ for any integer l, and $\mu h \leq n$, we have

$$\overline{A}^{(n)}_{\leq h} \leq n 2^{J\eta h} \theta^{\mu h} \max_{1 \leq w \leq \mu h} \frac{\binom{n}{\lfloor w/2 \rfloor}^J}{\binom{k}{w}^{J-1}}. \tag{6.6}$$

Using the inequalities in (B.3), we have

$$\begin{aligned} \frac{\binom{n}{\lfloor w/2 \rfloor}^J}{\binom{k}{w}^{J-1}} &\leq \frac{e^{nJH(x/2)}}{e^{nR(J-1)H(x/R)}} (k+1)^{J-1} \\ &\leq \frac{e^{nJH(x/2)}}{e^{nR(J-1)H(x/R)}} n^{J-1}, \end{aligned} \tag{6.7}$$

where $R = k/n$ (the rate of the overall code), and $x = w/n$. Combining (3.4) with (6.6) and (6.7), we have

$$\begin{aligned} r_n(\delta) &= \frac{1}{n} \log \overline{A}^{(n)}_h \\ &\leq \frac{1}{n} \log \overline{A}^{(n)}_{\leq h} \\ &\leq J \frac{\log n}{n} + T\delta + \sup_{0 < x \leq \mu\delta} \{ JH(\frac{x}{2}) - R(J-1)H(\frac{x}{R}) \}, \end{aligned} \tag{6.8}$$

where T is a constant. Equation (6.8) will used with Theorem 5.4 to prove that c_0 is finite for the $[E_1 \| \cdots \| E_J]$ ensemble.

7. Weight Enumerator Estimates for Serial Turbo Code Ensembles.

For the $[E_1 \Rightarrow E_2]$ ensemble, the average IOWE can be obtained from the weight enumerator of the outer code C_1 and the IOWE of the inner code C_2 [4] (see Figure 2 for notation):

$$\overline{A}^{(n)}_h = \sum_{d=1}^{N} \frac{A^{[1]}_d A^{[2]}_{d,h}}{\binom{N}{d}}. \tag{7.1}$$

Hence

$$\overline{A}^{(n)}_{\leq h} = \sum_{d=1}^{N} \frac{A^{[1]}_d A^{[2]}_{d,\leq h}}{\binom{N}{d}}. \tag{7.2}$$

Since if $A^{[2]}_{d,h} \neq 0$, d is less than μh by Theorem A.1 (where $\mu = \mu(E_2)$), applying Theorem A.2 to the outer code C_1 and Theorem A.3 to the inner code C_2 with L_1 as the trellis length for C_1 and L_2 as the trellis length for C_2, we obtain

$$\begin{aligned} \overline{A}^{(n)}_{\leq h} &= \sum_{d=1}^{\mu h} \frac{A^{[1]}_d A^{[2]}_{d,\leq h}}{\binom{N}{d}} \\ &\leq \sum_{d=1}^{\mu h} \theta^d \frac{\binom{L_1}{\lfloor d/d_1 \rfloor}}{\binom{N}{d}} \sum_{j=0}^{\lfloor d/2 \rfloor} \binom{L_2}{j}\binom{\eta h}{d-j}, \end{aligned} \tag{7.3}$$

where d_1 is the free distance of C_1. If C_1 is an (n_1, k_1, m_1) code of rate $R_1 = k_1/n_1$, and C_2 is a (n_2, k_2, m_2) code with rate $R_2 = k_2/n_2$, then we have $L_1 = N/n_1$, $L_2 = n/n_2$, and $N = R_2 n$, so that

$$\begin{aligned} L_1 &= \frac{k_2}{n_1 n_2} n = \alpha n \\ N &= \frac{k_2}{n_2} n = \beta n \\ L_2 &= \frac{1}{n_2} n = \gamma n, \end{aligned}$$

where $\alpha = k_2/n_1 n_2$, $\beta = k_2/n_2$, and $\gamma = 1/n_2$. Thus (7.3) becomes

$$\overline{A}^{(n)}_{\leq h} \leq \sum_{d=1}^{\mu h} \theta^d \frac{\binom{\alpha n}{\lfloor d/d_1 \rfloor}}{\binom{\beta n}{d}} \sum_{j=0}^{\lfloor d/2 \rfloor} \binom{\gamma n}{j}\binom{\eta h}{d-j}. \tag{7.4}$$

For $\delta = h/n$ small enough, we have $\eta h = \eta \delta n < n$, hence

$$\binom{\gamma n}{j}\binom{\eta h}{d-j} \leq \binom{\gamma n}{\lfloor d/2 \rfloor}\binom{\eta h}{\lceil d/2 \rceil}, \tag{7.5}$$

for any $0 \leq j \leq \lfloor d/2 \rfloor$ by Proposition B.1. Therefore, replacing the inner sum in (7.4) with $\lfloor d/2 \rfloor + 1$ times the right side of (7.5), we have

$$\begin{aligned} \overline{A}^{(n)}_{\leq h} &\leq \sum_{d=1}^{\mu h} \theta^d \frac{\binom{\alpha n}{\lfloor d/d_1 \rfloor}}{\binom{\beta n}{n}} (\lfloor d/2 \rfloor + 1) \binom{\gamma n}{\lfloor d/2 \rfloor}\binom{\eta h}{\lceil d/2 \rceil} \\ &\leq \sum_{d=1}^{\mu h} (2\theta_1)^d \frac{\binom{\alpha n}{\lfloor h/d_1 \rfloor}}{\binom{\beta n}{d}} \binom{\gamma n}{\lfloor d/2 \rfloor}\binom{\eta h}{\lceil d/2 \rceil}, \end{aligned} \tag{7.6}$$

(The last inequality because $\lfloor d/2 \rfloor + 1 \leq 2^d$.) The inequality (7.6) will be used to bound the innominate sum $Z^{(n)}(D_n)$.

To bound $r_n(\delta)$, we further simplify (7.6). Using the inequality $\binom{\eta h}{l} < 2^{\eta h}$, and bounding the summation in (7.6) by the number of terms times the maximum term, we have

$$\overline{A}^{(n)}_{\leq h} \leq n\theta^{\mu h}2^{\eta h} \max_{1\leq d\leq \mu h} \binom{\alpha n}{\lfloor h/d_1 \rfloor}\binom{\gamma n}{\lfloor d/2 \rfloor}/\binom{\beta n}{d}. \tag{7.7}$$

Using techniques like those that led from (6.6) to (6.8), the spectral shape can thus be upper bounded by the following expression, where $x = d/n$:

$$r_n(\delta) \leq 2\frac{\log n}{n} + T\delta + \sup_{0<x\leq \mu\delta}\{\alpha H(\frac{x}{d_1\alpha}) + \gamma H(\frac{x}{2\gamma}) - \beta H(\frac{x}{\beta})\}, \tag{7.8}$$

where T is a constant. Equation (7.8) will used with Corollary 5.4 to prove that c_0 is finite for the $[E_1 \Rightarrow E_2]$ ensemble.

8. Proof of Main Results.

In this section, we give the proofs of our main results, viz. Theorems 8.1 and 8.4. These theorems first appeared as conjectures, implicitly in [2] and [4] and explicitly in [11]. Theorem 8.1 can be summarized, using the language of [2] and [4], by saying that the $[E_1 \| \cdots \| E_J]$ ensemble has word error probability interleaving gain exponent $-J+2$, and bit error probability interleaving gain exponent $-J+1$. Theorem 8.4 can be summarized by saying that the $[E_1 \Rightarrow E_2]$ ensemble has word error probability interleaving gain exponent $-\lfloor \frac{d_1-1}{2} \rfloor$, and bit error probability interleaving gain exponent $-\lfloor \frac{d_1+1}{2} \rfloor$, where d_1 is the minimum distance of the outer code C_1.

8.1 Theorem. *For the $[E_1 \| \cdots \| E_J]$ ensemble, if $J \geq 2$, there exists a positive number c_0, such that for any binary-input memoryless channel whose noise exponent α satisfies $\alpha > c_0$, we have*

$$\overline{P}^{(n)}_W = O(n^{-J+2+\epsilon})$$
$$\overline{P}^{(n)}_b = O(n^{-J+1+\epsilon}),$$

for any $\epsilon > 0$.

Proof: (We restrict our attention to the statement about $\overline{P}^{(n)}_W$. The extension to $\overline{P}^{(n)}_b$ is explained in Appendix C.) Given Theorem 5.1 and Corollary 5.2, it will be sufficient to prove the following two lemmas.

8.2 Lemma. *For the $[E_1 \| \cdots \| E_J]$ ensemble, if $J \geq 2$, c_0 is finite.*

Proof: We use Corollary 5.4, with the upper bound (6.8) on the code spectral shape:

$$\gamma_n = \frac{J\log n}{n}$$
$$s(\delta) = T\delta + \sup_{0<x\leq \mu\delta}(JH(\frac{x}{2}) - R(J-1)H(\frac{x}{R})).$$

To show that $\limsup s(\delta)/\delta < \infty$, we need to show that the following limit is finite:

$$\lim_{\delta\to 0}\frac{1}{\delta}\sup_{0<x<\mu\delta}(JH(\frac{x}{2})-R(J-1)H(\frac{x}{R})).$$

But by Proposition B.3, this is true, since $J/2-R(J-1)/R=-J/2+1\le 0$, for $J\ge 2$. ▪

8.3 Lemma. *For the $[E_1\|\cdots\|E_J]$ ensemble, if $J\ge 2$,*

$$Z^{(n)}(D_n)=O(n^{-J+2+\epsilon}),$$

for any positive ϵ.

Proof: Using the upper bound (6.5) on $\overline{A}^{(n)}_{\le h}$, we have

$$\begin{aligned}
Z^{(n)}(D_n)&=\sum_{h=1}^{D_n}\overline{A}^{(n)}_n=\overline{A}^{(n)}_{\le D_n}\\
&\le\sum_{w=1}^{\mu D_n}\theta^w\frac{\binom{n}{\lfloor w/2\rfloor}^J\binom{\eta D_n}{\lceil w/2\rceil}^J}{\binom{Rn}{w}^{J-1}}\\
&\overset{(a)}{\le}\sum_{w=1}^{\mu D_n}\Theta^w n^{J\lfloor w/2\rfloor-(J-1)w}D_n^{(2J-1)w}\qquad(8.1)\\
&\overset{(b)}{=}O(n^{-J+2+\epsilon}).
\end{aligned}$$

In (a), we have used the following inequalities (see (B.2)): $\binom{n}{\lfloor w/2\rfloor}\le n^{\lfloor w/2\rfloor}$; $\binom{\eta D_n}{\lceil w/2\rceil}\le(\eta D_n)^{\lceil w/2\rceil}<(\eta D_n)^w$;$\binom{Rn}{w}\ge(Rn)^w/w^w\ge(Rn)^w/(\mu D_n)^w$. Here Θ represents a new constant. For (b), the sum in (8.1) can be upper bounded by μD_n times the largest term, which by Proposition B.2 is the $w=2$ term for large enough n. Notice $D_n=o(n^\epsilon)$ for any positive ϵ by (5.1). ▪

Next, we prove the corresponding theorem for serial turbo codes.

8.4 Theorem. *For the $[E_1\Rightarrow E_2]$ ensemble, with E_2 recursive, if the free distance of the outer code satisfies $d_1\ge 3$, there exists a positive number c_0, such that for any binary-input memoryless channel whose noise exponent α satisfies $\alpha>c_0$,*

$$\begin{aligned}
\overline{P}^{(n)}_W&=O(n^{-\lfloor\frac{d_1-1}{2}\rfloor+\epsilon})\\
\overline{P}^{(n)}_b&=O(n^{-\lfloor\frac{d_1+1}{2}\rfloor+\epsilon})
\end{aligned}$$

for arbitrary $\epsilon>0$.

(We again restrict our attention to $\overline{P}^{(n)}_W$, leaving $\overline{P}^{(n)}_W$ to Appendix C.) Again, because of Theorem 5.1 and Corollary 5.2, it's sufficient to prove the following two lemmas.

8.5 Lemma. *For the $[E_1 \Rightarrow E_2]$ ensemble, if the free distance of the outer code satisfies $d_1 \geq 2$, c_0 is finite.*

Proof: Corollary 5.4, together with eq. (7.8), makes it sufficient to show:

$$\lim_{\delta \to 0} \frac{1}{\delta} \sup_{0<x<\mu\delta} (aH(\frac{x}{d_1 \alpha}) + \gamma H(\frac{x}{2\gamma}) - \beta H(\frac{x}{\beta})) < \infty.$$

But by Proposition B.3, this is true, since $1/d_1 + 1/2 - 1 \leq 0$, for $d_1 \geq 2$. ▪

8.6 Lemma. *For the $[E_1 \Rightarrow E_2]$ ensemble, if $d_1 \geq 3$,*

$$Z^{(n)}(D_n) = O(n^{-\lfloor \frac{d_1-1}{2} \rfloor + \epsilon})$$

for arbitrary $\epsilon > 0$.

Proof: With the bound (7.6), we have

$$\begin{aligned} Z^{(n)}(D_n) &= \overline{A}^{(n)}_{\leq D_n} \\ &\leq \sum_{d=d_1}^{\mu D_n} \theta^d \frac{\binom{\alpha n}{\lfloor d/d_1 \rfloor}}{\binom{\beta n}{d}} \binom{n}{\lfloor d/2 \rfloor} \binom{\eta D_n}{\lceil d/2 \rceil} \\ &\overset{(a)}{\leq} \sum_{d=d_1}^{\mu D_n} \Theta^d n^{\lfloor d/d_1 \rfloor - \lceil d/2 \rceil} D_n^{d+\lceil d/2 \rceil} \\ &\overset{(b)}{=} O(n^{-\lfloor \frac{d_1-1}{2} \rfloor + \epsilon}). \end{aligned}$$

In step (a), we have used the following inequalities (see (B.2)): $\binom{\alpha n}{\lfloor d/d_1 \rfloor} \leq (\alpha n)^{\lfloor d/d_1 \rfloor}$; $\binom{\beta n}{d} \geq (\beta n)^d / d^d \geq (\beta n)^d / (\mu D_n)^d$; $\binom{n}{\lfloor d/2 \rfloor} \leq n^{\lfloor d/2 \rfloor}$; and $\binom{\eta D_n}{\lceil d/2 \rceil} \leq (\eta D_n)^{\lceil d/2 \rceil}$. For step (b), the sum is upper bounded by μD_n times the biggest term, which by Prop. B.2 is the $d = d_1$ term, as n becomes large. The conclusion follows, since $D_n = o(n^\epsilon)$ for any positive ϵ. ▪

9. Examples.

It is interesting to consider the CCSDS "standard" $R = 1/3$ turbo code [6] in the light of our results. This turbo code is a parallel concatenation with $J = 2$ recursive convolutional component codes, $R_1 = 1/2$, $R_2 = 1$, and overall rate $R = 1/3$. The two encoders are described by the transfer functions

$$G_1(D) = (1, \frac{1 + D + D^3 + D^4}{1 + D^3 + D^4}), \qquad G_2(D) = \frac{1 + D + D^3 + D^4}{1 + D^3 + D^4}.$$

Experimental evidence, together with density evolution analysis [10], with this ensemble on the AWGN channel suggests that for any value of E_b/N_0 greater than around -0.05

dB,[5] the bit error probability can be made arbitrarily small, in approximately inverse proportion to the block size, but the word error probability does not go to zero. If we apply Theorem 8.1 to this same ensemble, we get no quantitative information about the noise threshold, but we find that above the threshold, we have (ignoring the "$+\epsilon$" in the exponent) $\overline{P}_b^{(n)} = O(1/n)$, and $\overline{P}_W^{(n)} = O(1)$, in gratifying agreement with experiment. It is important to bear in mind, however, that: (1) the experiments are with suboptimum iterative decoding, whereas Theorem 8.1 deals with maximum-likelihod decoding; (2) Theorem 8.1 only provides an upper bound on code performance, and does not preclude the possibility that a more rapid decrease in decoder error probability is possible; and (3) experiments always deal with particular interleavers, whereas Theorem 8.1 treats the average over all interleavers.

The repeat-accumulative (RA) codes we introduced in [12] are serial turbo code ensembles with a $R_1 = 1/q$ q-fold repetition code as the outer code, and a $R_2 = 1$ recursive convolutional code, with transfer function $1/(1+D)$, as the inner code. The outer code has minimum distance $d_1 = q$. Hence, by Theorem 8.4, on all memoryless binary input channels, RA codes have word error probability approaching zero for $q \geq 3$ and bit error probability approaching zero for $q \geq 2$. For this ensemble, we can say something quantitative about the noise thresholds, since we can compute the exact spectral shape [12]:

$$r(\delta) = \max_{0 \leq x \leq 1/q} \{ -\frac{q-1}{q} H(qx) + (1-\delta) H(\frac{qx}{2(1-\delta)} + \delta H(\frac{qx}{2\delta}) \}.$$

Two short tables of these thresholds, on the binary symmetric channel and the Gaussian channel respectively, are given below.

q	R	γ_q	Shannon Limit
3	1/3	0.091	0.133
4	1/4	0.132	0.191
5	1/5	0.163	0.228
6	1/6	0.187	0.254
7	1/7	0.207	0.274

Table 1. RA ensemble thresholds on the BSC, obtained from the union bound.

In Table 1, the noise threshold γ_q is given as the largest value of the channel crossover probability for which the union bound guarantees good code performance for the corresponding RA ensemble. In Table 2, the threshold is given as the largest value of E_b/N_0 for which the union bound guarantees good performance. If the union bound is replaced with a more powerful tool, these thresholds can be considerably improved. For example,

[5] The Shannon limit for $R = 1/3$ codes on the AWGN channel is -0.495 dB.

q	R	γ_q (dB)	Shannon Limit (dB)
3	1/3	2.20	−0.495
4	1/4	1.93	−0.794
5	1/5	1.80	−0.963
6	1/6	1.72	−1.071
7	1/7	1.67	−1.150

Table 2. RA ensemble thresholds on the AWGN, obtained using the union bound.

q	R	UB:γ_q	TP:γ_q	Shannon Limit
3	1/3	0.091	0.132	0.174
4	1/4	0.132	0.191	0.215
5	1/5	0.163	0.228	0.243
6	1/6	0.187	0.254	0.265
7	1/7	0.207	0.274	0.281

Table 3. Comparison of RA ensemble thresholds using the union bound to those obtainable using the "typical pairs," technique on the BSC.

q	R	UB: γ_q	TP: γ_q	Shannon Limit (dB)
3	1/3	2.20	0.739	−0.495
4	1/4	1.93	−0.078	−0.794
5	1/5	1.80	−0.494	−0.963
6	1/6	1.72	−0.742	−1.071
7	1/7	1.67	−0.905	−1.150

Table 4. Comparison of RA ensemble thresholds using the union bound to those obtainable using the "typical pairs," technique on the AWGN channel.

using the "typical pairs" method, we can obtain the thresholds in Table 3 for RA codes on the BSC [1], and those in Table 4 for the AWGN channel [17].

10. Discussion and Conclusions.

The results in this paper are in a sense the culmination of a series of earlier papers [1, 9, 11, 12, 15]. In those papers we were interested in computing channel noise thresholds for specific code ensembles on specific channels; in this paper we have considered general

ensembles on general channels. However, we have paid a price for this generality: whereas in the earlier papers our estimates for the noise thresholds were computed numerically, in this paper we only prove the existence of the thresholds. To get good numerical thresholds using our methodology would require at least two improvements. First, we would have to replace the union bound with a more powerful technique; and second, we would need much more accurate estimates for the asymptotic weight spectrum $r(\delta)$ of the ensembles in question.

We have already addressed the first of these two problems. In refs. [1, 9, 11, 12, 15] we have developed a tool, the "typical pairs" method, which is capable of reproducing Shannon's theorem for the ensemble of random linear codes. However, the typical pairs method, despite its power, is useless unless one has an exact or near-exact expression for the asymptotic weight spectrum $r(\delta)$ of the ensemble in question. This is the second, and more difficult, of the needed improvements. To date, we can give good estimates for $r(\delta)$ in only three cases: the ensemble of all linear codes of rate R (here $r(\delta) = H(\delta) - (1-R)$), the ensemble of Gallager (j,k) low-density parity-check codes [14], and the ensmble of RA codes [1]. A method for computing $r(\delta)$ for other ensembles, in particular the turbo-code ensembles, would be very welcome.

Our main results provide only upper bounds on $\overline{P}_W^{(n)}$ and $\overline{P}_b^{(n)}$, but based on experimental evidence we conjecture that these bounds are close to best possible, viz., for any channel with $\gamma < \gamma_0$, $\overline{P}_W^{(n)} = \Omega(n^{-\beta})$. More generally, for any binary-input discrete memoryless channel, we conjecture that either

$$\lim_{n\to\infty} \overline{P}_W^{(n)} = 1$$

or

$$\overline{P}_W^{(n)} = \Theta(n^{-\beta}).$$

If these conjectures are true, it follows that the interleaving gain exponent β is an important measure of the ensemble's performance, and not just an artifact of our method of proof.

Finally, we mention the important alternative approach to this problem recently announced by Richardson and Urbanke [24]. This work extends their earlier, landmark work on low-density parity-check codes [23], and deals directly with the performance of iterative decoding. They show, for any $J = 2$, rate $1/3$ parallel turbo ensemble, on a extensive class of symmetric binary-input channels, the existence of a noise threshold σ^*, such that if the noise is below σ^*, the ensemble bit error probability can be made arbitrarly small, whereas if the noise exceeds σ^*, the ensemble bit error probability is bounded away from zero. Furthermore, they describe a numerical algorithm that can be used to find the exact value of σ^* in many cases. In many ways this work surpasses ours for the (ensemble, channel) pairs to which it applies. The only pieces of our main results apparently not present in R-U is quantitative information about the rate at which $\overline{P}_b^{(n)}$ approaches zero, and information about the word error probability. We predict that the RU analysis can be extended to the general $[E_1 \| E_2 \| \cdots \| E_J]$ and $[E_1 \Rightarrow E_2]$ ensembles, and to all memoryless binary-input channels, in which case there will be little of remaining interest in the present paper.

Appendix A. Combinatorial Facts About Truncated Convolutional Codes.

In this Appendix we shall state for reference three useful combinatorial facts about the weight structure of convolutional codes, due essentially to Kahale and Urbanke [18]. (Although Theorems A.2 and A.3 were stated in [18] only for systematic rate 1/2 codes, the proofs given apply in the generality we state.)

A.1 Theorem. *(The η-μ theorem.) For a non-catastrophic convolutional encoder E, there exists a constant μ, $\mu = \mu(E)$, such that if the output weight is h, then the input weight is at most μh. Also, there is a constant $\eta = \eta(E)$ such that if a codeword in the truncated code consists of several detours, of total length L_0, then the codeword weight d satisfies $d \geq L\eta$.*

In what follows, $A_h^{(L)}$ denotes the number of codewords of weight h in the Lth truncation of the code and $A_{w,h}^{(L)}$ denotes the corresponding number of codewords with input weight w and output weight h. Thus $A_h^{(L)} = \sum_{w=1}^{k} A_{w,h}^{(L)} =$ (by Theorem A.1) $= \sum_{w=1}^{\mu h} A_{w,h}^{(L)}$. Similarly, $A_{w,\leq h}^{(L)}$ denotes the the number of codewords with input weight w and output weight less than or equal to h, i.e., $A_{w,\leq h}^{(L)} = \sum_{d=1}^{h} A_{w,d}^{(L)}$.

A.2 Theorem. *(Cf. [18, Lemma 3]) Let C be an (n, k, m) convolutional code, as represented by a noncatastrophic encoder E. Then for the $(nL, kL - m)$ block code obtained by truncating C at depth L,*

$$A_h^{(L)} \leq \theta^h \binom{L}{\lfloor h/d_1 \rfloor}, \tag{A.1}$$

where d_1 is the free distance of the code, and θ is a constant independent of h and n.

We define a *recursive* convolutional code to be one for which any input of weight 1 produces an output of infinite weight.

A.3 Theorem. *(Cf. [18, Lemma 1]) Let C be an (n, k, m) recursive convolutional code, with corresponding noncatastrophic encoder E. Then for the $(nL, kL - m)$ block code obtained by truncating the E-trellis representation of C at depth L,*

$$A_{w,\leq h}^{(L)} \leq \theta^w \sum_{j=0}^{\lfloor w/2 \rfloor} \binom{L}{j} \binom{\eta h}{w - j}, \tag{A.2}$$

where θ and η are constants independent of w, h, and n. (For the significance of η, see Theorem A.1.)

Appendix B. Some Useful Inequalities.

Suppose n, k are positive integers, $1 \leq k \leq n$. Then

$$\binom{n}{k} \leq 2^n \tag{B.1}$$

$$\frac{n^k}{k^k} \le \binom{n}{k} \le n^k \tag{B.2}$$

$$e^{nH(k/n)}/(n+1) \le \binom{n}{k} \le e^{nH(k/n)}. \tag{B.3}$$

(For (B.3), see [7, Example 12.1.3, p. 284].)

B.1 Proposition. *If $n \ge m$, $0 \le j \le \lfloor w/2 \rfloor$, then*

$$\binom{n}{j}\binom{m}{w-j} \le \binom{n}{\lfloor w/2 \rfloor}\binom{m}{w-\lfloor w/2 \rfloor} = \binom{n}{\lfloor w/2 \rfloor}\binom{m}{\lceil w/2 \rceil}.$$

Proof: It suffices to show that $f(j) = \binom{n}{j}\binom{m}{w-j}$ is an increasing function of j, for $0 \le j \le \lfloor w/2 \rfloor$. Consider the ratio

$$\frac{f(j)}{f(j-1)} = \frac{n-j+1}{m-w+j}\frac{w-j+1}{j}, \qquad \text{for } j \ge 1.$$

since $w-j+1 \ge j$ and $n-j+1 \ge m-w+j$, we have $f(j)/f(j-1) \ge 1$. Hence the conclusion follows. ▪

B.2 Proposition.

(1) Given $F_n(w) = \Theta^w n^{J\lfloor w/2 \rfloor - (J-1)w} D_n^{(2J-1)w}$, $1 \le w \le \mu D_n$, $F_n(2)$ will be the largest term as n becomes large.

(2) Given $G_n(d) = \Theta^d n^{\lfloor d/d_1 \rfloor + \lfloor d/2 \rfloor - d} D_n^{2d}$, $d_1 \le d \le \mu D_n$, $G_n(d_1)$ will be the largest term as n becomes large.

Proof: (1): It is easy to show that $F_n(w)$ satisfies $F_n(1) \ge F_n(3) \ge F_n(5) \ge \ldots$ and $F_n(2) \ge F_n(4) \ge F_n(6) \ge \ldots$ as n gets large by taking the ratio of two consecutive terms. Verifying that $F_n(2) \ge F_n(1)$ for large n, we have the claim.

(2): Similarly, we can show $G_n(d_1) \ge G_n(d_1+1) \ge \ldots \ge G_n(\mu D_n)$ by taking the ratio of two consecutive terms. ▪

B.3 Proposition. *Given real numbers α_i, β_i for $i = 1, \ldots, n$, with $\beta_i \ge 0$, define*

$$\Delta = \sum_{i=1}^{n} \alpha_i \beta_i,$$

and for $\mu > 0$, let

$$L = \lim_{\delta \to 0} \frac{1}{\delta} \left(\sup_{0 < x < \mu\delta} \sum_{i=1}^{n} \alpha_i H(\beta_i x) \right). \tag{B.4}$$

Then $L < +\infty$, if $\Delta \le 0$.

Proof: (Sketch). It is easy to see that for small x,

$$H(x) = x \log \frac{1}{x} + x + O(x^2),$$

and so

$$\sum_i \alpha_i H(\beta_i x) = \Delta x \log \frac{1}{x} + \left(\sum_i \alpha_i \beta_i (1 + \log \frac{1}{\beta_i}) \right) x + O(x^2).$$

If $\Delta < 0$, the first term in the above expansion dominates, and the result follows immediately (indeed, the limit is 0). If $\Delta = 0$ we have

$$\sum_i \alpha_i H(\beta_i x) = \left(\sum_i \alpha_i \beta_i \log(\frac{1}{\beta_i}) \right) x + O(x^2),$$

in which casee the "sup" in (B.4) is attained at $x = \mu\delta$ as $\delta \to 0$, and the limit is finite. ▪

Appendix C. Bit Error Probability vs. Word Error Probability.

The union bound on the bit error probability for maximum likelihood decoding of an (n, k) binary linear code C with IOWE $(A_{w,k})$ over a memoryless binary input channel has the following form:

$$P_b \le \sum_{h=1}^{n} \sum_{w=1}^{k} \frac{w}{k} A_{w,h} e^{-\alpha h}. \tag{C.1}$$

In this appendix, we will state, and sketch a proof of, a theorem on the ensemble bit error probability $\overline{P}_b^{(n)}$, analogous to Theorem 5.1 (which deals with word error probability). To that end, we define another innominate sum:

$$Y^{(n)}(D) \stackrel{\Delta}{=} \sum_{h=1}^{D} \sum_{w=1}^{k} \frac{w}{k} \overline{A}_{w,h}^{(n)}. \tag{C.2}$$

C.1 Theorem. *If the threshold c_0 defined in (5.4) is finite, then if $\alpha > c_0$, there exists an integer n_0 and positive constants K and ϵ such that for $n \ge n_0$,*

$$\overline{P}_b^{(n)} \le Y^{(n)}(D_n) + K e^{-\epsilon D_n}. \tag{C.3}$$

Proof: (Sketch.) Beginning with eq. (C.1), we have

$$\overline{P}_b^{(n)} \le \sum_{h=1}^{n} \sum_{w=1}^{k} \frac{w}{k} \overline{A}_{w,h}^{(n)} e^{-\alpha h}$$

$$\leq \sum_{h=1}^{D} \sum_{w=1}^{k} \frac{w}{k} \overline{A}_{w,h}^{(n)} + \sum_{h>D} \sum_{w=1}^{k} \frac{w}{k} \overline{A}_{w,h}^{(n)} e^{-\alpha h}$$

$$= Y^{(n)}(D) + \sum_{h>D} \sum_{w=1}^{k} \frac{w}{k} \overline{A}_{w,h}^{(n)} e^{-\alpha h}$$

$$\leq Y^{(n)}(D) + \sum_{h>D} \sum_{w=1}^{k} \overline{A}_{w,h}^{(n)} e^{-\alpha h}$$

$$= Y^{(n)}(D) + \sum_{h>D} \overline{A}_{h}^{(n)} e^{-\alpha h}. \qquad (C.4)$$

Theorem C.1 now follows almost immediately from (C.4) and the proof of Theorem 5.3. ▪

C.2 Corollary. *If in addition, $Y^{(n)}(D_n) = O(n^{-\beta})$, where $\beta > 0$, then for $\alpha > c_0$,*

$$\overline{P}_b^{(n)} = O(n^{-\beta}) \qquad (C.5)$$

The following lemma shows how the results on word error probability can be easily extended to bit error probability. In essence, Lemma C.3 shows that $Z^{(n)}(D_n) = O(n^{-\beta})$ if and only if $Y^{(n)}(D_n) = O(n^{-\beta+1})$.

C.3 Lemma. *There exists a positive constant μ, such that*

$$Z^{(n)}(D_n)/k \leq Y^{(n)}(D_n) \leq \mu D_n Z^{(n)}(D_n)/k.$$

Proof: Applying $w/k \geq 1/k$ to (C.2), we obtain the left inequality. From Prop. A.1 we know that if $\overline{A}_{w,h}^{(n)} \neq 0$, then $w \leq \mu h$. Thus if $h \leq D_n$, and $\overline{A}_{w,h}^{(n)} \neq 0$, then $w \leq \mu h \leq \mu D_n$. The right hand inequality then follows if we upper bound w/k by $\mu D_n/k$ in (C.2). Finally, since $k = Rn$, where R is the rate of the ensemble, it follows that $Z^{(n)}(D_n) = O(n^{-\beta})$ iff $Y^{(n)}(D_n) = O(n^{-\beta+1})$. ▪

Acknowledgements.
We thank Dariush Divsalar, who did the density evolution analysis for the $R = 1/3$ CCDSD code, and Rüdiger Urbanke for assuring us that the bounds in Appendix A indeed follow from the results in [18].

References.

1. S. M. Aji, H. Jin, D. MacKay, and R. J. McEliece, "BSC thresholds for code ensembles based on 'Typical Pairs' decoding." To appear in *Proc. IMA Workshop on Codes and Graphs*, Minneapolis, MN, August 1999.

2. S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes", *IEEE Trans. Inform. Theory*, vol. 42, no. 2 (March 1996), pp. 409–428.

3. S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes," *IEEE Trans. Comm.*, vol. 44, no. 5, (May 1996) pp. 591–600.

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

5. C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: turbo codes," *Proc. 1993 IEEE International Conference on Communications*, Geneva, Switzerland (May 1993), pp. 1064–1070.

6. Consultative Committee for Space Data Systems (CCSDC), "Telemetry Channel Coding," vol. 101.0-B-4, Blue Book, issue 4, May 1999. (Available at `http://www.ccsds.org.documents/pdf/CCSDS-101.0-B-4.pdf`).

7. T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley and Sons, 1991.

8. D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," *TDA Progress Report*, vol. 42-139 (July-Sept. 1999). Available at `http://tmo.jpl.nasa.gov/tmo/progress_report/42-139/139L.pdf`.

9. D. Divsalar, S. Dolinar, H. Jin and R. McEliece, "AWGN coding theorems from ensemble weight enumerators," Proc. 2000 *International Symposium on Information Theory*, p. 458.

10. D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *in preparation*.

11. D. Divsalar and R. J. McEliece, "On the design of concatenated coding systems with interleavers," JPL TMO Progress Report vol. 2-134 (August 15, 1998), pp. 1–22. (`http://tmo.jpl.nasa.gov/tmo/progress_report/42-134/134D.pdf` .)

12. D. Divsalar, H. Jin, and R. McEliece. "Coding theorems for 'Turbo-Like' codes." Proc. 1998 Allerton Conf., pp. 201–210.

13. D. Divsalar and F. Pollara, "On the design of turbo codes," *TDA Progress Report* vol. 42-123 (November 15, 1995), pp. 99–121.

14. R. Gallager, *Low-Density Parity-Check Codes*. Cambridge, Mass.: The M.I.T. Press, 1963.

15. H. Jin and R. J. McEliece, "AWGN coding theorems for serial turbo codes," Proc. 37th Allerton Conf. on Communication, Computation and Control, Allerton, IL. (Sept. 1999), pp. 893–894.

16. H. Jin and R. J. McEliece, "RA codes achieve AWGN channel capacity," Proc. 13th Int. Symp. AAECC-13 (1999) (Springer Lecture Notes in Computer Science no. 1719), pp. 10–18.

17. H. Jin and R. J. McEliece, "Typical pairs decoding on the AWGN channel," Proc.

2000 International Symp. Inform. Theory and its Applications, pp. 180–183.

18. N. Kahale and Rüdiger Urbanke, "On the minimum distance of parallel and serially concatenated codes," submitted to *IEEE Trans. Inform. Theory*.

19. D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory* vol. IT-45 (March 1999), pp. 399–431.

20. R. J. McEliece, *The Theory of Information and Coding*. Reading, Mass.: Addison-Wesley, 1977.

21. R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng, "Turbo decoding as as instance of Pearl's 'Belief Propagation' algorithm," *IEEE J. Selected Areas in Communication*, vol. 16, no. 2 (Feb. 1998), pp. 140–152.

22. T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of provably good low-density parity-check codes," submitted to *IEEE Trans. Inform. Theory*.

23. T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," submitted to *IEEE Trans. Inform. Theory*.

24. T. Richardson and R. Urbanke, "Thresholds for turbo codes," Proc. 2000 *International Symposium on Information Theory*, P. 317.

25. C. E. Shannon, *The Mathematical Theory of Information*. Urbana, IL: University of Illinois Press, 1949 (reprinted 1998).