BSC THRESHOLDS FOR CODE ENSEMBLES BASED ON "TYPICAL PAIRS" DECODING*

SRINIVAS AJI[†], HUI JIN[†], AAMOD KHANDEKAR[†], DAVID J.C. MACKAY[‡], AND ROBERT J. MCELIECE[†]

Abstract. In this paper, we develop a method for closely estimating noise threshold values for ensembles of binary linear codes on the binary symmetric channel. Our method, based on the "typical pairs" decoding algorithm pioneered by Shannon, completely decouples the channel from the code ensemble. In this, it resembles the classical union bound, but unlike the union bound, our method is powerful enough to prove Shannon's theorem for the ensemble of random linear codes. We apply our method to find numerical thresholds for the ensembles of low-density parity-check codes, and "repeat-accumulate" codes.

1. Introduction. In this paper, we consider the performance of ensembles of codes on the binary symmetric channel. Our particular focus is on the question as to whether or not a given ensemble is "good," in the sense of MacKay [7]. In short, an ensemble of codes is said to be good, if there is a $p > 0$ such that the ensemble word error probability (with maximum-likelihood decoding) on a BSC with crossover probability p approaches zero as the block length approaches infinity. The largest such p for a given ensemble is called the (noise) *threshold* for the ensemble. Our main result (Theorem 4.1) is a technique for finding a lower bound on the ensemble threshold, which is based on the ensemble's weight enumerator.

Of course the classical union bound provides one way of using weight enumerators to estimate ensemble thresholds, but the estimates are poor. Gallager [4, Chapter 3] gave a variational method for upper bounding the probability of maximum-likelihood decoding error for an arbitrary binary code, or ensemble of codes (given an expression for the average weight-enumerator function) on a general class of binary-input channels. Gallager's technique, however, is quite complex, and even in the special case of the BSC it is difficult to apply to the problem of finding ensemble thresholds.[1]

In this paper we abandon the full maximum-likelihood decoder, and instead focus on a slightly weaker decoding algorithm, which is much easier to analyze, the *typical pairs* decoder. This technique was pioneered by

[*] The work of Aji, Jin, Khandekar, and McEliece on this paper was supported by NSF grant no. CCR-9804793, and grants from Sony, Qualcomm, and Caltech's Lee Center for Advanced Networking. David MacKay's work is supported by the Gatsby Charitable Foundation.

[†] Department of Electrical Engineering (136-93), California Institute of Technology, Pasadena, CA 91125.

[‡] Department of Physics, University of Cambridge, Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, United Kingdom.

[1] We have been able to show that the thresholds obtained by our method are the same as the best obtainable by the Gallager methodology.

Shannon [11, Theorem 11], but as far as we can tell was not used to analyze ensembles other than the ensemble of all codes (which we call the Shannon ensemble in Section 5, below) until the 1999 paper of MacKay [7], in which it was used to analyze certain ensembles of low-density-parity check codes. In brief, when presented with a received word $\mathbf{y}$, the typical pairs decoder seeks a codeword $\mathbf{x}$ such that the pair $(\mathbf{x}, \mathbf{y})$ belongs to the set T of "typical pairs." (We give a precise definition of T in section 2, which follows.) In Section 3, we develop an upper bound on the typical-pairs decoder's error probability (Theorem 3.1) which, like the classical union bound, decouples the code's weight enumerator from the channel, but unlike the union bound, when combined with the law of large numbers, gives good estimates for code thresholds (Theorem 4.1).

We then apply Theorem 4.1 to three families of binary code ensembles: (1) The Shannon ensemble, consisting of all linear codes of rate R; (2) the Gallager ensemble, consisting of (j, k) low-density parity-check codes; and (3) the ensemble of Repeat-Accumulate codes introduced by Divsalar, Jin and McEliece [2]. In the case of the Shannon ensembles, we show that our method yields thresholds identical to those implied by Shannon's theorem. Thus the typical sequence method, despite its suboptimality, loses nothing (in terms of coding thresholds) for the Shannon ensemble.

Finally, we compare our thresholds to the *iterative* thresholds for the Gallager and RA ensembles recently obtained by Richardson and Urbanke [10], in order to estimate the price paid in coding threshold for the benefits of iterative decoding. In most cases, this loss is quite small, and in the case of $j = 2$ LDPC codes, there appears to be no penalty at all.

The method described in this paper can be readily extended to many other channel models, including channels with memory (cf. [7, Section II]). This extension will be developed in a forthcoming paper, where the emphasis will be on the binary erasure channel and the additive Gaussian noise channel.

2. Typical pairs. Let T be a set of binary vectors of length n which is closed under coordinate permutations, and let $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ be the BSC noise vector, i.e., the Z_i's are i.i.d. random variables with common density

$$\Pr\{Z = 0\} = 1 - p, \qquad \Pr\{Z = 1\} = p.$$

If we define the set T to be a set of "typical" noise vectors, then T represents the typical channel outputs if the zero-word is transmitted, and the $T + \mathbf{x}$ represents the set of typical channel outputs if the codeword $\mathbf{x}$ is transmitted. In the typical-pairs decoder (to be defined shortly), decoder errors can result if the channel output is in the typical set of more than one codeword. We are therefore interested in the quantity $\Pr\{\mathbf{Z} \in T \cap (T + \mathbf{x})\}$.

If T is invariant under coordinate permutations, the probability $\Pr\{\mathbf{Z} \in T \cap (T + \mathbf{x})\}$ depends only on the weight of $\mathbf{x}$. Thus we define, for $h = 0, 1, \ldots, n$,

$$P_h(T) = \Pr\{\mathbf{Z} \in T \cap (T + \mathbf{x})\}, \tag{2.1}$$

where $\mathbf{x}$ is any vector of weight h. The quantity $P_h(T)$ is then the probability of error in a typical-set decoder in the case of a code having only two codewords separated by a Hamming distance h.

For example,

$$P_0(T) = \Pr\{\mathbf{Z} \in T\}. \tag{2.2}$$

Since any set T which is invariant under coordinate permutations must consist of all vectors of weight $k \in K$, where K is a subset of $\{0, 1, \ldots, n\}$, the probabilities $P_h(T)$ depend only on the set K. A short combinatorial calculation gives

$$P_h(T) = \sum_{k_1 \in K} p^{k_1}(1-p)^{n-k_1} \sum_{k_2 \in K} \binom{h}{(h+k_1-k_2)/2} \binom{n-h}{(k_1-h+k_2)/2}. \tag{2.3}$$

This is because a vector of weight k_1 has probability $p^{k_1}(1-p)^{n-k_1}$, and there are exactly $\binom{h}{(h+k_1-k_2)/2}\binom{n-h}{(k_1-h+k_2)/2}$ vectors of weight k_1, which have the property that when the first h components are complemented, i.e., the vector $\mathbf{x} = (\overbrace{11\cdots1}^{h}\overbrace{00\cdots0}^{n-h})$ is added, the resulting vector has weight k_2. Applying (2.3) to the case $h = 0$, we obtain

$$P_0(T) = \sum_{k \in K} p^k (1-p)^{n-k} \binom{n}{k},$$

in agreement with (2.2).

In our main application (Theorem 4.1) the set T will be the "typical sequences" of length n and so will be denoted by T_n. The definition of T_n is

$$T_n = \left\{ \mathbf{z} : \left| \frac{\mathrm{wt}(\mathbf{z})}{n} - p \right| \le \epsilon_n \right\}, \tag{2.4}$$

where ϵ_n is a sequence of real numbers approaching zero more slowly than $n^{-1/2}$, i.e., $\epsilon_n \sqrt{n} \to \infty$. Then by a straightforward extension of the weak law of large numbers,

$$\lim_{n \to \infty} \Pr\{\mathbf{Z} \in T_n\} = 1. \tag{2.5}$$

Furthermore, by defining $K_n = \{k : n(p - \epsilon_n) \le k \le n(p + \epsilon_n)\}$, and using the formula (2.3), it is relatively easy to prove that for any δ in the range $0 \le \delta \le 2p$, we have

$$\lim_{n \to \infty} -\frac{1}{n} \log P_{\delta n}(T_n) = K(\delta, p), \tag{2.6}$$

where $K(\delta, p)$ is given by the equivalent formulas

$$K(\delta, p) = H(p) - \delta \log 2 - (1-\delta) H\left(\frac{p - \delta/2}{1-\delta}\right) \tag{2.7}$$

$$= H(\delta) - pH\left(\frac{\delta}{2p}\right) - (1-p)H\left(\frac{\delta}{2(1-p)}\right), \tag{2.8}$$

where $H(x)$ is the entropy function, i.e., $H(x) = -x \log x - (1-x)\log(1-x)$. (These formulas are true only for $\delta < 2p$; for $\delta \geq 2p$, $K(\delta, p)$ is infinite, since $P_h(T_n) = 0$ for $h > 2n(p + \epsilon_n)$.) In Figure 1, we have plotted the function $K(\delta, p)$ for several values of p.[2]

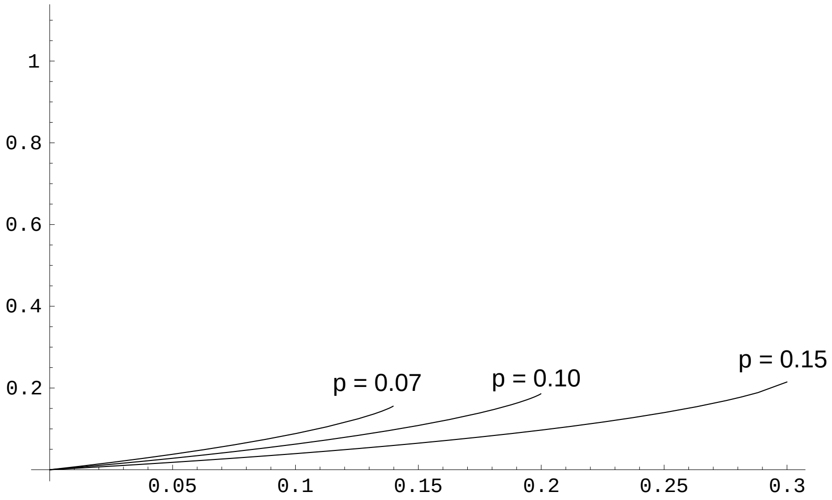

FIG. 1. *The Function $K(\delta, p)$ for $p = 0.07, 0.10., 0.15$.*

In fact, a closer examination of the limit in (2.6) shows that for a fixed value of p, the limit is *uniform*. That is, for a fixed p, there exists a sequence of positive numbers $\beta_n \to 0$, such that

$$\left| -\frac{1}{n} \log P_{\delta n}(T_n) - K(\delta, p) \right| < \beta_n \quad \text{for all } 0 < \delta < 2p. \tag{2.9}$$

Alternatively, we can write (2.9) as

$$P_h(T_n) = e^{-n(K(\delta,p)+o(1))}, \tag{2.10}$$

where $\delta = h/n$.

[2] In Figure 1, and all the other figures in the paper, computations using logarithms use natural logarithms.

3. The typical pairs decoding method. Suppose C is an (n, k) binary linear code, with weight enumerator $(A_0, A_1, \ldots, A_n)$, i.e., C contains exactly A_h words of Hamming weight h, for $h = 1, \ldots, n$. We suppose that at the transmitter, a codeword $\mathbf{x} \in \mathcal{C}$ is selected at random, transmitted over a BSC with crossover probability p, and received at the destination as $\mathbf{y}$. The T-decoder tries to infer $\mathbf{x}$ based on knowledge of the code $\mathcal{C}$, the noisy codeword $\mathbf{y}$, and the channel noise parameter p. The T-decoder works as follows.

For every codeword $\mathbf{x}_i$, the ith "pseudonoise" $\mathbf{z}_i = \mathbf{y} - \mathbf{x}_i$ is computed. If there are no indices i for which $\mathbf{z}_i \in T$, the decoder fails. Otherwise, among those indices such that $\mathbf{z}_i \in T$, the decoder choose one for which the Hamming weight $w(\mathbf{z}_i)$ is smallest. In short, the decoder chooses the *most likely codeword for which* $\mathbf{z}_i$ *is typical.* In what follows we do not distinguish between decoder error and failure, and denote the probability of decoder error (or failure) by P_E.

THEOREM 3.1. *If* P_E *denotes the probability that the* T*-decoder does not correctly identify the transmitted codeword, then*

$$P_E \leq (1 - P_0(T)) + \sum_{h=1}^{n} A_h \min(\beta^h, P_h(T)), \tag{3.1}$$

where $\beta = 2\sqrt{p(1-p)}$ *is the channel Bhattacharyya parameter.*[3]

Proof. Let $(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{M-1})$ be an ordering of the code with $\mathbf{x}_0$ being the all-zeros word, and suppose $\mathbf{x}_0$ is transmitted. For $i = 0, 1, \ldots, M-1$, define the following events:

$$\begin{aligned} T_i &= \{\mathbf{z}_i \in T\} && (\mathbf{z}_i \text{ is typical}) \\ V_i &= \{w(\mathbf{z}_i) \leq w(\mathbf{z}_0)\} && (\mathbf{z}_i \text{ is more likely than } \mathbf{z}_0) \\ S_i &= T_i \cap V_i && (\mathbf{z}_i \text{ is typical and is more likely than } \mathbf{z}_0) \end{aligned}$$

Then the T-decoder will fail only if at least one of the events $T_0', S_1, \ldots, S_{M-1}$ occurs. Thus if $\mathcal{E}$ denotes the event "T-decoder fails, given that $\mathbf{x}_0$ was transmitted," we have (here T_0' denotes the complement of T_0)

$$\begin{aligned} \mathcal{E} &= T_0' \cup \left(\textstyle\bigcup_{i=1}^{M-1} S_i\right) \\ &= T_0' \cup \left(T_0 \cap \textstyle\bigcup_{i=1}^{M-1} S_i\right) \\ &= T_0' \cup \left(\textstyle\bigcup_{i=1}^{M-1} (T_0 \cap S_i)\right). \end{aligned} \tag{3.2}$$

Therefore the probability of T-decoder error, given that $\mathbf{x}_0$ was transmitted, can be upper bounded as follows:

$$\Pr\{\mathcal{E} | \mathbf{x}_0\} \leq \Pr\{T_0' | \mathbf{x}_0\} + \sum_{i=1}^{M-1} \Pr\{T_0 \cap S_i | \mathbf{x}_0\}. \tag{3.3}$$

[3] The term β^h in (3.1) is present for technical reasons, e.g., the proof of Theorem 4.1. Normally, it will be smaller than the term $P_h(T)$ only for very small values of h.

But $\Pr\{T_0'|\mathbf{x}_0\} = 1 - \Pr\{T_0|\mathbf{x}_0\}$, and $\Pr\{T_0|\mathbf{x}_0\} = \Pr\{\mathbf{Z} \in T\} = P_0(T)$, from (2.2). Thus

$$\Pr\{T_0'|\mathbf{x}_0\} = 1 - P_0(T). \tag{3.4}$$

Also, since $S_i = T_i \cap V_i$, it follows that

$$\Pr\{T_0 \cap S_i|\mathbf{x}_0\} \leq \min(\Pr\{V_i|\mathbf{x}_0\}, \Pr\{T_0 \cap T_i|\mathbf{x}_0\}).$$

By the familiar union bound argument [8, Theorem 7.5], we have

$$\Pr\{V_i|\mathbf{x}_0\} \leq \beta^{h_i},$$

where h_i is the Hamming weight of $\mathbf{x}_i$.

Also note that by definition $\mathbf{z}_i = \mathbf{y} - \mathbf{x}_i$, and so we have, for $i = 1, \ldots, M-1$, $\mathbf{z}_i = \mathbf{z}_0 + (\mathbf{x}_0 - \mathbf{x}_i) = \mathbf{z}_0 - \mathbf{x}_i$, since $\mathbf{x}_0$ is the all-zeros word. Thus $T_i = \{\mathbf{z}_0 \in T + \mathbf{x}_i\}$, and so

$$\begin{aligned}\Pr\{T_0 \cap T_i|\mathbf{x}_0\} &= \Pr\{\mathbf{Z} \in T \cap (T + \mathbf{x}_i)\}\\ &= P_{h_i}(T)\end{aligned}$$

where h_i is the Hamming weight of $\mathbf{x}_i$. Hence

$$\begin{aligned}\sum_{i=1}^{M-1} \Pr\{T_0 \cap S_i|\mathbf{x}_0\} &\leq \sum_{i=1}^{M-1} \min(\beta^{h_i}, P_{h_i}(T)),\\ &= \sum_{h=1}^{n} A_h \min(\beta^h, P_h(T)),\end{aligned} \tag{3.5}$$

since there are exactly A_h words of Hamming weight h in $\mathcal{C}$. Combining (3.3) with (3.4) and (3.5), gives (3.1). □

4. Code ensembles. By an *ensemble* of linear codes we mean a sequence $\mathcal{C}_{n_1}, \mathcal{C}_{n_2}, \ldots$ of sets of linear codes of a common rate R, where $\mathcal{C}_{n_i}$ is a set of (n_i, k_i) codes with $k_i/n_i = R$. We assume that the sequence $n_1, n_2, \ldots$ approaches infinity. If C is an (n, k) code in the ensemble, we denote the weight enumerator of C by the list $A_0(C), A_1(C), \ldots, A_n(C)$. The *average weight enumerator* for the set $\mathcal{C}_n$ is defined as the list

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

where

$$\overline{A}_h^{(n)} \triangleq \frac{1}{|\mathcal{C}_n|} \sum_{C \in \mathcal{C}_n} A_h(C) \qquad \text{for } h = 0, 1, \ldots, n. \tag{4.1}$$

We define, for each n in the sequence $n_1, n_2, \ldots$, the function

$$r_n(\delta) \triangleq \frac{1}{n} \log \overline{A}_{\lfloor \delta n \rfloor}^{(n)} \qquad \text{for } 0 < \delta < 1, \tag{4.2}$$

Also, we define the *ensemble spectral shape* :

$$r(\delta) \triangleq \lim_{n\to\infty} r_n(\delta) \qquad \text{for } 0 < \delta < 1, \tag{4.3}$$

assuming that the limit exists. In this case, we may write

$$\overline{A}_h^{(n)} = e^{n(r(\delta)+o(1))}, \tag{4.4}$$

where $\delta = h/n$.

Now we apply Theorem 3.1, using the set T_n, defined in (2.4), to a code $C \in \mathcal{C}_n$:

$$P_E \le \eta_n + \sum_{h=1}^{n} A_h(C) P_h(T_n), \tag{4.5}$$

where $\eta_n = \Pr\{T_n'\} \to 0$ by (2.5). If we average (4.5) over all codes in the ensemble $\mathcal{C}_n$, we obtain the following upper bound on $\overline{P}_E^{(n)}$, the ensemble decoder error probability:

$$\overline{P}_E^{(n)} \le \eta_n + \sum_{h=1}^{n} \overline{A}_h^{(n)} P_h(T_n). \tag{4.6}$$

Replacing $\overline{A}_h^{(n)}$ with the right side of (4.4), and $P_h(T_n)$ with the right side of (2.10), (4.6) becomes

$$\overline{P}_E^{(n)} \le \eta_n + \sum_{h=1}^{n} e^{-n(K(\delta,p)-r(\delta)+o(1))}. \tag{4.7}$$

It now appears that if p is chosen so that the function $K(\delta,p) - r(\delta)$ is positive for all $0 < \delta < 1$, so that the exponent in the sum in (4.7) is always negative, the ensemble word error probability $\overline{P}_E^{(n)}$ will approach zero, as $n \to \infty$. This is in fact true, provided we make the following two technical assumptions about the behavior of $\overline{A}_h^{(n)}$, for $h = o(n)$.

- *Assumption 1.* There exist a sequence of integers d_n such that $d_n \to \infty$ and

$$\lim_{n\to\infty} \sum_{h=1}^{d_n} \overline{A}_h^{(n)} = 0. \tag{4.8}$$

(This assumption says, roughly, that the minimum distance of the ensemble is at least d_n.)

- *Assumption 2.* There exist a sequence of real numbers $\theta_n \ge 0$ such that

$$r_n(\delta) \le r(\delta) + \theta_n, \quad \text{where} \quad \lim_{n\to\infty} \frac{n\theta_n}{d_n} = 0. \tag{4.9}$$

We now state our main result:

THEOREM 4.1. *Suppose the code ensemble has spectral shape* $r(\delta)$, *and also that it satisfies Assumptions 1 and 2. Then if the crossover probability* $p < 1/2$ *of the channel satisfies*

$$K(\delta, p) > r(\delta) \qquad \text{for} \quad 0 < \delta < 2p,$$

then $\overline{P}_E^{(n)} \to 0$ *as* $n \to \infty$.

There is a slightly weaker version of Assumption 1 that guarantees that the ensemble *bit* error probability approaches zero:

- *Assumption* 1′. There exist a sequence of integers d_n such that $d_n \to \infty$ and

$$\lim_{n\to\infty} \sum_{h=1}^{d_n} \frac{h}{n} \overline{A}_h^{(n)} = 0. \tag{4.10}$$

The corresponding modification of Theorem 4.1 follows.

THEOREM 4.2. *Suppose the code ensemble has spectral shape* $r(\delta)$, *and also that it satisfies Assumptions* 1′ *and 2. Then if the crossover probability* $p < 1/2$ *of the channel satisfies*

$$K(\delta, p) > r(\delta) \qquad \text{for } 0 < \delta < 2p,$$

then $\overline{P}_b^{(n)} \to 0$ *as* $n \to \infty$, *where* P_b *denotes the T-decoder's bit error probability.*

(A proof of Theorem 4.1 will be found in the Appendix. The proof of Theorem 4.2 is similar and is omitted.)

In the following three sections, we will apply Theorem 4.1 to three different ensembles of binary linear codes: (1) The Shannon ensemble, consisting of all linear codes of rate R; (2) the Gallager ensemble, consisting of (j, k) low-density parity-check codes; and (3) the ensemble of Repeat-Accumulate codes introduced by Divsalar, Jin and McEliece [2].

5. The Shannon ensemble. For the set of random linear codes of rate R, we have

$$\overline{A}_h^{(n)} = \binom{n}{h} 2^{-n(1-R)}, \tag{5.1}$$

from which it follows via a routine calculation that

$$r(\delta) = H(\delta) - (1 - R)\log 2. \tag{5.2}$$

This function is shown for $R = 1/3$ in Figure 2.

To apply Theorem 4.1 to the Shannon ensemble,[4] for a given rate R we must find the largest p such that $K(\delta, p) > H(\delta) - (1 - R)\log 2$ for all $0 < \delta < 2p$.

[4] Assumptions 1 and 2 are satisfied with $d_n = Kn$ for a suitable positive constant $K = K(R)$, and $\theta_n = 0$.

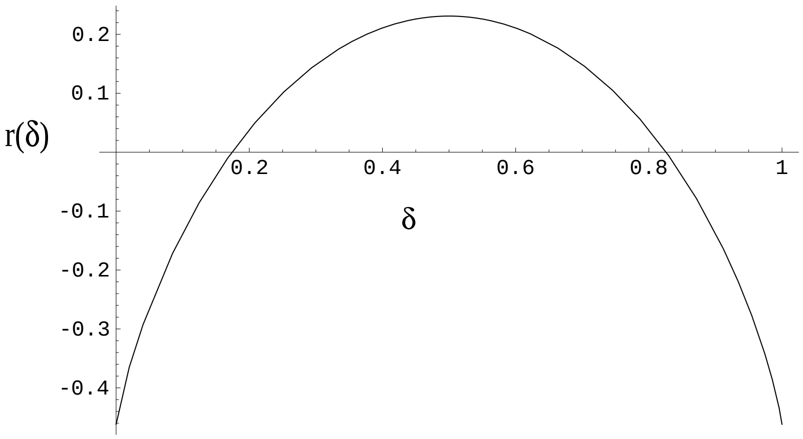

FIG. 2. *The function $r(\delta)$ for the ensemble of $R = 1/3$ linear codes.*

Using (5.2) and (2.8), this inequality becomes

$$pH\left(\frac{\delta}{2p}\right) + (1-p)H\left(\frac{\delta}{2(1-p)}\right) < (1-R)\log 2. \tag{5.3}$$

The maximum of the left side of (5.3) in the range $0 < \delta < 2p$ occurs at $\delta = 2p(1-p)$, and is $H(p)$. Thus the inequality $K(\delta, p) > H(\delta) - (1-R)\log 2$ required by Theorem 4.1 becomes simply $H(p) < (1-R)\log 2$, or $H_2(p) < 1 - R$, where $H_2(p)$ is the binary entropy function. Thus we have proved

THEOREM 5.1. *The ensemble of random linear codes of rate R is good on a BSC with crossover probability p if $H_2(p) < 1 - R$.*

The idea of the proof is illustrated in Figure 3, where we see the function $K(\delta, 0.174)$ just touching the $r(\delta)$ curve of Figure 2. This shows that the threshold for the ensemble of $R = 1/3$ linear codes is $p = 0.174$, which reflects the fact that $H_2(0.174) = 1 - 2/3$.

Of course, Theorem 5.1 is just Shannon's theorem for linear codes on the BSC. We have included it only to demonstrate that Theorem 4.1 is powerful enough to reproduce Shannon's theorem. In the next two sections we will apply it to more interesting ensembles.

6. The Gallager ensemble. In this section, we discuss the application of Theorem 4.1 to the ensemble of (j, k) low-density parity-check codes defined by Gallager [4].[5] In brief, every code in Gallager's (j, k) ensemble is defined by a parity-check matrix which has j ones in each column

[5] There are numerous ways to define this ensemble. The definition we follow was given by Gallager [4, Section 2.2], and differs, e.g. from the ensemble analyzed by MacKay in [7, Section II].

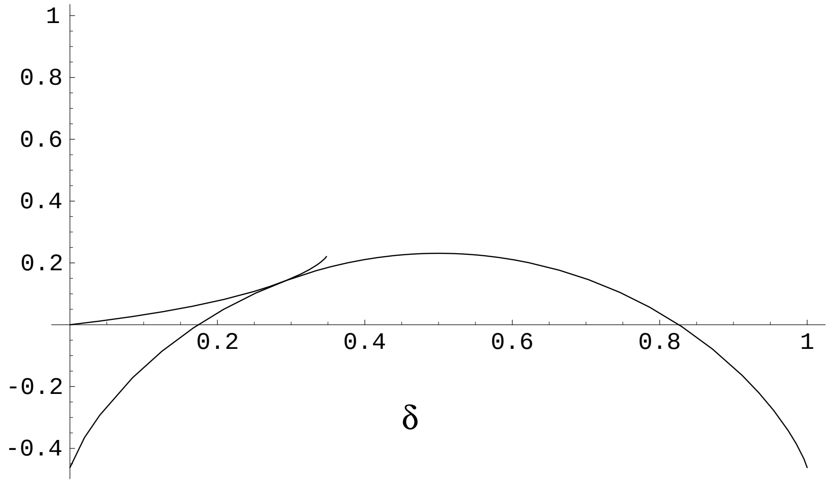

FIG. 3. *The function $r(\delta)$ for the ensemble of $R = 1/3$ linear codes, together with the function $K(\delta, p)$ for $p = 0.174$.*

and k ones in each row. The rate of each code in the ensemble is at least $R_{j,k} = 1 - (j/k)$.

The spectral shape $r_{j,k}(\delta)$ for the (j,k) ensemble was determined by Gallager. It can be expressed in parametric form, as follows:

$$\begin{aligned} \delta_{j,k}(s) &= \frac{1}{k}\frac{\partial\mu}{\partial s}(s,k) \\ r_{j,k}(s) &= \frac{j}{k}\left(\mu(s,k) - s\frac{\partial\mu}{\partial s}(s,k) + (k-1)\log 2\right) - (j-1)H\left(\frac{1}{k}\frac{\partial\mu}{\partial s}(s,k)\right) \end{aligned}$$

where the parameter s ranges from $-\infty$ to $+\infty$, and the function $\mu(s,k)$ is defined by

$$\mu(s,k) \triangleq \log\frac{(1+e^s)^k + (1-e^s)^k}{2^k}.$$

Figure 4 shows the function $r_{j,k}$ for $(j,k) = (3,6)$.

Given the spectral shape, it is an easy task to apply Theorem 4.1 to find the corresponding BSC ensemble thresholds.[6] A short table of these thresholds, together with the corresponding Shannon limit, is given below.

[6] To satisfy Assumptions 1 and 2 for $j \geq 3$, we can take $d_n = Kn$ for a suitable constant $K = K(j)$, and $\theta_n = 0$. For $j = 2$, we can prove the existence of a sequence of d_n's which satisfy Assumptions 1′ and 2 with $\theta_n = 0$, though we do not have an explicit expression for them.

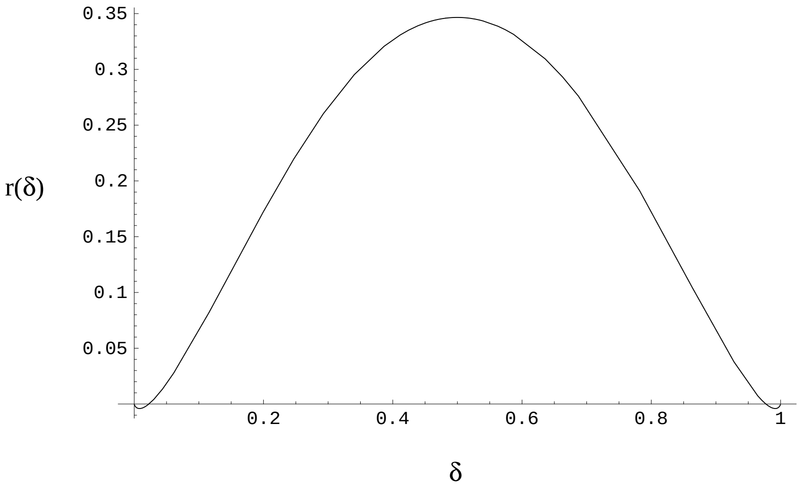

FIG. 4. *The function $r(\delta)$ for the ensemble of* (3, 6) *LDPC codes.*

(j,k)	$R_{j,k}$	$p_{j,k}$	RU limit	Shannon limit
(3,6)	1/2	0.0915	0.084	0.109
(3,5)	2/5	0.129	0.113	0.145
(4,6)	1/3	0.170	0.116	0.174
(3,4)	1/4	0.205	0.167	0.214
(2,3)	1/3	0.0670	0.0670	0.174
(2,4)	1/2	0.0286	0.0286	0.109

For example, consider the "$(3,5)$" line in the table. The corresponding Gallager ensemble consists of codes which have parity-check matrices with 3 ones per column and 5 ones per row. The rate of all codes in this ensemble at least $R_{3,5} = 1-(3/5) = 2/5$. Using Theorem 4.1, it is calculated that for any BSC with crossover probability $p < 0.129$, the $(3,5)$ ensemble is good, i.e., the average word error probability of the T-decoder approaches 0, as $n \to \infty$. This should be compared to the Shannon limit for the ensemble of all linear codes of rate $2/5$ (cf. Theorem 5.1), which is $p = 0.145$, which indicates the price which is paid for having the $(3,5)$ structure. Finally, we note that the Richardson-Urbanke limit [10] for the $(3,5)$ ensemble is $p = 0.113$, i.e., with belief propagation–style iterative decoding, the ensemble decoder error probability approaches 0 if and only if $p < 0.113$.

(The values $p_{j,k}$ for $(j,k) = (3,6)$, $(3,5)$, $(4,6)$, and $(3,4)$ given in the above table appear to agree with the values given by Gallager [4] in his Figure 3.5, although he gave no numerical values. However, as we mentioned above, we have been able to show that the thresholds obtained from our Theorem 4.1 are the same as the best obtainable using Gallager's methodology, so our threshold values are at least as good as Gallager's.)

We conclude this section with some remarks on the ensemble of $(2,k)$ LDPC codes. Originally dismissed by Gallager because their minimum distance is $O(\log n)$ [4, Theorem 2.5], they are nevertheless quite interesting, and are variously called "graph-theoretic," "circuit," or "cycle" codes [9, Section 5.8], [6] because of their close connection to finite undirected graphs. Using Theorem 4.2, we can show that for $p < p^*(k)$, the *bit error probability* for T-decoding of the $(2,k)$ ensemble approaches zero, where $p^*(k)$ is given by the exact formula

$$p^*(k) = \frac{1}{2}\left(1 - \sqrt{1 - \frac{1}{(k-1)^2}}\right). \tag{6.1}$$

(The ensemble word error probability does not approach zero for any $p > 0$.)

Furthermore, Wiberg [12, Example 5.1] showed that with *iterative* decoding, the ensemble of $(2,k)$ cycle codes has ensemble bit error probability approaching zero for $p < p^*(k)$. Numerically, the Richardson-Urbanke method appears to give the same value, so it seems safe to say that (6.1) gives the exact iterative threshold for the Gallager $(2,k)$ ensemble.[7]

Finally, it was shown by Decreusefond and Zémor [3] that for an "expurgated" ensemble of $(2,k)$ cycle codes, the *exact* maximum-likelihood BSC coding threshold is equal to $p^*(k)$. Since as we have seen, the threshold for the unexpurgated ensemble is at least this good, it seems very likely that $p^*(k)$ is the exact ML threshold for the unexpurgated ensemble as well. These results strongly suggest that that for $(2,k)$ cycle codes, the iterative and maximum-likelihood thresholds are the same, and are given by the formula (6.1).

7. The ensemble of repeat-accumulate codes. In brief, for an integer $q \geq 2$, the ensemble of q-repeat accumulate codes consists of those codes which can be encoded by the serial concatenation of a q-ary repetition encoder, followed by a pseudorandom permutation, followed by a rate 1 code with (square) generator matrix of generic shape

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The basic combinatorial fact about the ensemble of (qk,k) RA codes is the following formula for the average number of input words of weight w

[7] For a survey of iterative decoding of cycle codes, see [5].

which are encoded into output words of weight h [2, Eq. (5.4)]:

$$\overline{A}^{(qk)}_{w,h} = \frac{\binom{k}{w}\binom{qk-h}{\lfloor qw/2 \rfloor}\binom{h-1}{\lceil qw/2 \rceil - 1}}{\binom{qk}{qw}}. \tag{7.1}$$

It follows then that if $\overline{A}^{(qk)}_{h}$ denotes the average number of words of weight h in the ensemble,

$$\overline{A}^{(qk)}_{h} = \sum_{w=1}^{N} \overline{A}^{(qk)}_{w,h}. \tag{7.2}$$

From (7.1) and (7.2), it can be shown that the spectral shape $r(\delta)$ for the ensemble of q-RA codes is as follows :

$$r(\delta) = \max_{0 \le x \le 1/q} \left\{ -\frac{q-1}{q} H(qx) + (1-\delta) H(\frac{qx}{2(1-\delta)}) + \delta H(\frac{qx}{2\delta}) \right\}. \tag{7.3}$$

Figure 5 shows the $r(\delta)$ curve for the ensemble of $q = 3$ RA codes.[8]

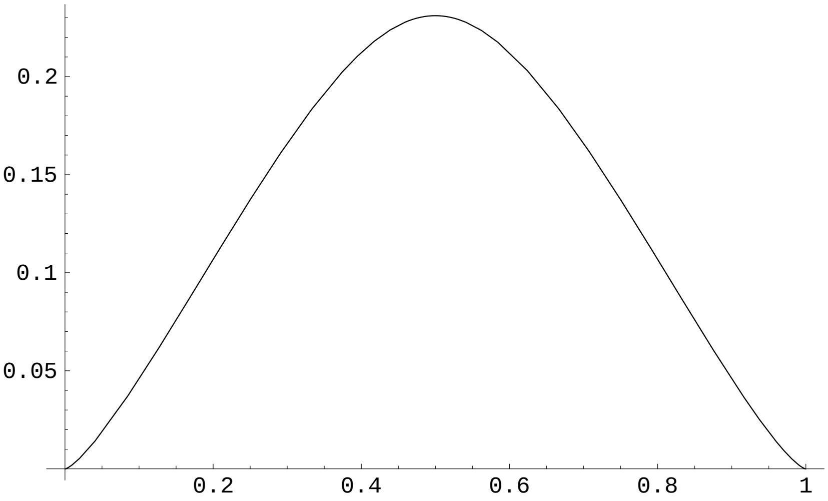

FIG. 5. *The function $r(\delta)$ for the ensemble of $R = 1/3$ RA codes.*

Combining (7.3) with Theorem 4.1, it is a straightforward computation to obtain the thresholds in the following table.

[8] To satisfy Assumptions 1 and 2 for $q \ge 3$, we can take $d_n = \log^2 n$ and $\theta_n = (K \log n)/n$ for suitable constants $K = K(q)$. For $q = 2$, we can only show the existence of a sequence d_n satisfying Assumptions 1′ and 2 by taking $d_n = 2$ and $\theta_n = (K \log n)/n$.

q	R_q	p_q	RU limit	Shannon limit
2	1/2	0.029*	–	0.109
3	1/3	0.132	0.142	0.174
4	1/4	0.191	0.188	0.215
5	1/5	0.228	0.216	0.243
6	1/6	0.254	0.235	0.264
7	1/7	0.274	0.250	0.281

For example, consider the $q = 3$ line of the table. It indicates that the common rate for all $q = 3$ RA codes is $R = 1/3$, and that this ensemble is good on any BSC with crossover probability $p < 0.132$. By way of comparison, the Shannon threshold for the ensemble of all rate 1/3 linear codes is seen to be $p < 0.174$. Finally, the Richardson-Urbanke iterative decoding threshold [Richardson and Urbanke, private commmunication] is $p < 0.142$. Since we can show that the T-decoding algorithm always gives the same ensemble threshold as does maximum-likelihood decoding, which must be at least as good as the iterative threshold, this apparently shows either that the thresholds given in Theorem 4.1 are not always the best possible for T-decoding, or that the R-U theorem is not correct for this ensemble. A resolution of this paradox would be very welcome.

Finally we note that for the ensemble of $q = 2$ RA codes, the word error probability for T-decoding does not approach zero for any $p > 0$, but, again by using Theorem 4.2, we can show that the ensemble *bit* error probability approaches zero for $p < 0.029$.

APPENDIX

A. Proof of Theorem 4.1. We first define the *ensemble threshold* as follows:

$$p_0 = \sup\{p : K(\delta, p) > r(\delta), 0 < \delta < 2p\}. \tag{A.1}$$

LEMMA A.1. *If $p < p_o$, then there exist real numbers $\alpha_0 > 0$ and $\gamma_0 > 0$, and a positive integer N_0, such that for $n \geq N_0$,*

$$\sum_{h=d_n}^{\alpha_0 n} \overline{A}_h^{(n)} \beta^h = O(e^{-d_n \gamma_0}),$$

where $\beta = 2\sqrt{p(1-p)}$.

Proof. Using the definition (2.7), It is straightforward to show that

$$\lim_{\delta \to 0} \frac{K(\delta, p_0)}{\delta} = \frac{\partial K(0, p_0)}{\partial \delta} = -\log \beta_0,$$

where $\beta_0 = 2\sqrt{p_0(1-p_0)}$. Hence for $p < p_0$, we have

$$\begin{aligned} \limsup_{\delta\to 0} \tfrac{r(\delta)}{\delta} &\leq \lim_{\delta\to 0} \frac{K(\delta,p_0)}{\delta} \\ &= -\log\beta_0 = 2\sqrt{p_0(1-p_0)} \\ &< -\log\beta = 2\sqrt{p(1-p)}. \end{aligned}$$

This, together with Assumption 2, implies that there exists $\alpha_0 > 0$, $\gamma_0 > 0$, and a positive integer N_0 such that for $n \geq N_0$, we have

$$\sup_{d_n/n \leq \delta < \alpha_0} \frac{r_n(\delta)}{\delta} < \frac{n\theta_n}{d_n} + \sup_{0\leq\delta<\alpha_0} \frac{r(\delta)}{\delta} < -\log\beta - \gamma_0.$$

Hence we have, for $n \geq N_0$,

$$\begin{aligned} \sum_{h=d_n}^{\alpha_0 n} \overline{A}_n^{(n)} \beta^h &= \sum_{h=d_n}^{\alpha_0 n} e^{-h(\log\beta - r_n(\delta)/\delta)} < \sum_{h=d_n}^{\alpha_0 n} e^{-h\gamma_0} \\ &< \sum_{h=d_n}^{\infty} e^{-h\gamma_0} = O(e^{-d_n\gamma_0}), \end{aligned}$$

which completes the proof. □

Now we can give the proof of Theorem 4.1. With the notation being as established above, we have, by Theorem 3.1, for $p < p_0$,

$$P_E \leq \sum_{h=1}^{d_n} \overline{A}_h^{(n)} + \sum_{h=d_n}^{\alpha_0 n} \overline{A}_h^{(n)} \beta^h + \sum_{h=\alpha_0 n}^{n} \overline{A}_h^{(n)} P_h(T) + o(n). \tag{A.2}$$

The first sum in (A.2) approaches zero by Assumption 1, the second sum approaches zero by Lemma A.1 together with the fact that $d_n \to \infty$. The third sum is

$$\begin{aligned} \sum_{h=\alpha_0 n}^{n} \overline{A}_h^{(n)} P_h(T) &= \sum_{h=\alpha_0 n}^{n} e^{-n(K(\delta,p)-r(\delta)+o(1))} \\ &\leq \sum_{h=\alpha_0 n}^{n} e^{-n(K(\delta,p)-K(\delta,p_0)+o(1))}, \end{aligned} \tag{A.3}$$

where the first line follows from (2.10) and Assumption 2, and the second line follows from the definition (A.1) of p_0.

Finally, let ϵ be such that

$$K(\delta,p) - K(\delta,p_0) \geq \epsilon \qquad \text{for } \alpha_o \leq \epsilon \leq 2p.$$

Then for n sufficiently large, the exponent in (A.3) will be $\geq \epsilon/2$, and so the sum will be upper bounded by $n \cdot e^{-n\epsilon/2}$, which goes to zero as $n \to \infty$. □

Acknowledgement. This paper is an outgrowth of conversations that took place at the Institute for Mathematical Analysis, Minneapolis, Minnesota, during the workshop on Graphical Models and Iterative Decoding that took place in August, 1999. The authors wish to thank IMA for its hospitality and conducive work environment.

REFERENCES

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

[2] D. Divsalar, H. Jin, and R. McEliece, "Coding Theorems for 'Turbo-Like' Codes." Proc. 1998 Allerton Conf., pp. 201–210.

[3] L. Decreusefond and G. Zémor, "On the error-correcting capabilities of cycle codes of graphs," *Combinatorics, Probability, and Computing*, vol. **6** (1997), pp. 27–38.

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

[5] G.B. Horn, "The iterative decoding of cycle codes," submitted to *IEEE Trans. Inform. Theory.*

[6] D. Jungnickel and S.A. Vanstone, "Graphical codes revisited," *IEEE Trans. Inform. Theory*, vol. IT-43 (Jan. 1997), pp. 136–146.

[7] 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.

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

[9] W.W. Peterson and E.J. Weldon, Jr., *Error-Correcting Codes, 2nd. ed.* Cambridge, Mass.: The MIT Press, 1972.

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

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

[12] N. Wiberg, *Codes and Decoding on General Graphs.* Linköping Studies in Science and Technology. Dissertation, no. **440**. Linköping University, Linköping, Sweden, 1996.