band-limited functions. The decoders are still convolutional, and thus democratic; in his construction, the rate of the exponential decay of the error, shown here to be necessarily strictly inferior to 1, is smaller than 1 by several orders of magnitude.

Acknowledgment

The authors would like to thank Ron DeVore and C. Sinan Güntürk for helpful discussions concerning the topic of this paper. They are also grateful to the referees for helpful comments. I. Daubechies would also like to thank the Institute for Advanced Study in Princeton for its hospitality during the writing of this correspondence.

References

[1] J. C. Candy and G. C. Temes, Eds., *Oversampling Delta–Sigma Data Converters Theory, Design and Simulation*. New York: IEEE Press, 1992.

[2] I. Daubechies and R. DeVore, "Reconstructing a bandlimited function from very coarsely quantized data: A family of stable sigma–delta modulators of arbitrary order," *Ann. Math.*, to be published.

[3] I. Daubechies, R. DeVore, C. S. Güntürk, and V. Vaishampayan, "Exponential precision in A/D conversion with an imperfect quantizer," paper, submitted for publication.

[4] P. Frankl and Z. Füredi, "A short proof for a theorem of Harper about Hamming-spheres," *Discr. Math.*, vol. 34, pp. 311–313, 1981.

[5] R. M. Gray, "Spectral analysis of quantization noise in single-loop sigma–delta modulator with dc input," *IEEE Trans. Commun.*, vol. COM-35, pp. 481–489, 1987.

[6] R. M. Gray, W. Chou, and P. W. Wong, "Quantization noise in single-loop sigma–delta modulation with sinusoidal inputs," *IEEE Trans. Commun.*, vol. 37, pp. 956–968, Sept. 1989.

[7] S. Güntürk, "Improved error estimates for first order sigma–delta modulation," in *Sampling Theory and Applications, SampTA'99*, Leon, Norway, Aug. 1999.

[8] ——, "Reconstructing a bandlimited function from very coarsely quantized data: Improving the error estimate for first order sigma–delta modulators," in preparation.

[9] ——, "Harmonic analysis of two problems in signal compression," Ph.D. dissertation, Program in Applied and Computational Mathematics, Princeton Univ., Princeton, NJ, Sept. 2000.

[10] ——, "On democratic encoding and decoding," in preparation.

[11] C. S. Güntürk, J. C. Lagarias, and V. Vaishampayan, "Robustness of single loop sigma–delta modulation for constant inputs," *IEEE Trans. Inform. Theory*, submitted for publication.

[12] S. Hein, K. Ibrahim, and A. Zakhor, "New properties of sigma–delta modulators with dc inputs," *IEEE Trans. Commun.*, vol. 40, pp. 1375–1387, Aug. 1992.

[13] S. Hein and A. Zakhor, "Optimal decoding for data acquisition applications of sigma delta modulators," *IEEE Trans. Signal Processing*, vol. 41, pp. 602–616, Feb. 1993.

[14] ——, *Sigma Delta Modulators: Nonlinear Decoding Algorithms and Stability Analysis*. Dordrecht, The Netherlands: Kluwer, 1993.

[15] ——, "Reconstruction of oversampled band-limited signals from $\Sigma\Delta$ encoded binary sequences," *IEEE Trans. Signal Processing*, vol. 42, pp. 799–811, Apr. 1994.

[16] S. Konyagin, private communication, 1998.

[17] S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta–Sigma Data Converters Theory, Design and Simulation*. New York: IEEE Press, 1997.

[18] A. Zakhor, "Lower bounds on the MSE of the single loop sigma delta modulator," in *Proc. 23rd Asilomar Conf. Signals, Systems and Computers*, Oct. 1989, pp. 849–853.

Error Exponents of Expander Codes

Alexander Barg, *Senior Member, IEEE,* and Gilles Zémor

***Abstract*—We show that expander codes attain the capacity of the binary-symmetric channel under iterative decoding. The error probability has a positive exponent for all rates between zero and channel capacity. The decoding complexity grows linearly with code length.**

***Index Terms*—Expander code, iterative decoding, Ramanujan graph.**

I. Introduction

Constructing families of codes of growing length N with low error probability and large minimum distance is one of the main problems of coding theory. The first result that gave us codes together with a polynomial decoding algorithm that achieves an exponentially small error probability goes back to Forney [3], and relies on a concatenated construction. Specifically, it states that for any $\varepsilon > 0$ there exists an infinite family of easily constructible codes of rate R that are decodable with complexity $O(N^2)$ and error probability $P_e \le 2^{-Nf_1(R,\, p)}$, where p is the bit transition probability of the binary symmetric channel and where

$$f_1(R,\, p) = \max_{R \le R_0 < 1 - H(p)} E(R_0,\, p)(1 - R/R_0) - \varepsilon. \qquad (1)$$

Here $E(R_0,\, p)$ is the "random coding exponent" [4] and $H(\cdot)$ is the binary entropy function. Thus, $f_1(R,\, p) > 0$ for all rates R up to the channel capacity. A similar idea was used by Zyablov in [10] to construct codes with polynomial decoding complexity and relative distance arbitrarily close to

$$\delta(R) = (1 - R/R_0)H^{-1}(1 - R_0) \qquad (R \le R_0 \le 1). \qquad (2)$$

Thus, for these codes we have $\delta(R) > 0$ for any value R of the code rate $0 \le R < 1$. These results underwent a number of improvements (surveyed, for instance, in [2]), but until recently no codes were known with lower decoding complexity and nonvanishing relative distance and/or error exponent for nontrivial values of the code rate.

The first result of this kind [7] gives us a family of codes, based on a graph-theoretic approach of [8] and constructions of Ramanujan graphs [5], [6]. The result is as follows.

Theorem 1 [7]: For any $\varepsilon > 0$, there exists a polynomial-time constructible family of codes with distance $\delta - \varepsilon$ and rate $1 - 2H(\sqrt{\delta})$ for which any $\alpha < \delta/48$ fraction of errors can be corrected by a circuit of size $O(N \log N)$ and depth $O(\log N)$. The complexity of a sequential implementation of this decoding is $O(N)$.

This result was a remarkable novelty because not only did it decrease the complexity of decoding but it was also the first time that concatenation was not used to construct families of binary codes with a nontrivial rate and a nonzero relative minimal distance. It implies the existence of linear-time decodable codes with $R > 0$ for $0 \le \delta < 0.0121$.

Sipser–Spielman's decoding algorithm was modified in [9] where the factor 48 was improved to 4. In the present correspondence, we study the error probability for iterative decoding of expander codes and

Manuscript received September 29, 2000; revised February 12, 2001.

A. Barg is with Bell Labs, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: abarg@research.bell-labs.com).

G. Zémor is with École Nationale Supérieure des Télécommunications, 75 634 Paris 13, France (e-mail: zemor@infres.enst.fr).

Communicated by S. Shamai, Guest Editor.

Publisher Item Identifier S 0018-9448(02)04034-8.

modified expander codes that we shall introduce. We shall use the decoding algorithm of [9].

Our main result is as follows.

Theorem 2: For a given rate R, any $\varepsilon > 0$, and $\alpha < 1$ there exists a polynomial-time constructible family of codes of length N such that $P_e(C, p) \leq 2^{-\alpha N f_2(R, p)}$, where

$$f_2(R, p) = \max_{R \leq R_0 < 1 - H(p)} E(R_0, p)(R_0 - R)/2 - \varepsilon.$$

The decoding complexity of these codes is the same as in Theorem 1.

Thus, $f_2(R, p) > 0$ for all $0 \leq p < 1 - H(p)$. Comparing this result with (1), we see that $f_1(R, p) > f_2(R, p)$ for all R, p, so in terms of error probability expander codes do not surpass concatenated codes, but they do give an alternative to classical concatenation and they have a lower decoding complexity.

II. Hard Iterative Decoding of Expander Codes

A. Code Construction, Notation

Setup—The Graph G, Vectors, and Subvectors: Let G be a bipartite graph with vertex set $A \cup B$ where $|A| = |B| = n$ and where every edge of G has one endpoint in A and one in B. Any vertex $v \in A$ ($v \in B$) will be called left (right) vertex. Let E be the edge set of G. The *neighborhood* of a vertex v, denoted by E_v, is the set of edges incident to v. Suppose G is Δ-regular, i.e., the neighborhood E_v of any vertex v contains exactly Δ edges. The total number of edges is therefore $N = \Delta n$. Number the edges of G, i.e., let $E = \{1, 2, \ldots, N\}$, and for any vertex v define $v(1), v(2), \ldots, v(\Delta)$ to be some ordering of the edges of the neighborhood E_v of v. Let $x = (x_1, x_2, \ldots, x_N)$ be any vector of $\{0, 1\}^N$. The neighborhood E_v of every vertex v of G induces a subvector of x of length Δ, namely, $x_v = (x_{v(1)}, x_{v(2)}, \ldots, x_{v(\Delta)})$.

The Code C: For every $v \in A \cup B$, let C_v be some code of length Δ and dimension k_v. Let us define the code C to be the set of binary vectors $x = (x_1, x_2, \ldots, x_N)$ of $\{0, 1\}^N$ such that for every vertex v of G, x_v is a codeword of C_v. To keep the construction manageable, one usually chooses C_v among a limited set of small codes, e.g., one sets $C_v = C_0$ for a given fixed code C_0; asymptotic behavior can then be studied by letting the number of vertices n of the graph G go to infinity. In this correspondence, we shall use two constituent codes C_0 and C_1 of length Δ, and we shall set $C_v = C_0$ for every left vertex $v \in A$ and $C_v = C_1$ for every right vertex $v \in B$.

Parameters: Let $[\Delta, k_0 = R_0\Delta, d_0]$ be the parameters (length, dimension, minimum distance) of C_0 and $[\Delta, k_1 = R_1\Delta, d_1]$ be those of C_1. Let $K = RN$ be the dimension of C. The code C is linear and its redundancy is at most the sum of the redundancies of the constituent codes [8], so that we have $N - K \leq \sum_{v \in A \cup B} \Delta - k_v$. Whence

$$R \geq R_0 + R_1 - 1. \tag{3}$$

Remark: To obtain a reduction in decoding complexity from (1) and (2) it is essential to notice that Δ is a constant independent of n; thus, the construction falls in the class of low-density parity-check codes. If Δ is allowed to grow, the construction would be similar to the classical ones; for instance, taking $G = K_{n,n}$ (a complete bipartite graph) yields a standard direct product construction of C_0 and C_1.

B. Decoding

We shall use the decoding algorithm of [9]. Let us briefly recall its description. Let $x \in \{0, 1\}^N$ be the received vector. The first iteration, let us call it a left-decoding step, consists of applying in parallel, for every left vertex $v \in A$, complete decoding of the subvector x_v induced by the neighborhood of v. In other words, a left-decoding step is a function L that decodes x into a vector $y = L(x)$ of $\{0, 1\}^N$ where, for every $v \in A$, the vector $y_v = (y_{v(1)}, y_{v(2)}, \ldots, y_{v(\Delta)})$ is one of the codewords of C_0 closest to $x_v = (x_{v(1)}, x_{v(2)}, \ldots, x_{v(\Delta)})$. The next iteration, a right-decoding step, is a function R: $y \mapsto z$ defined similarly with A replaced by B, and C_0 replaced by C_1. The next iterations alternately repeat left-decoding and right-decoding steps, i.e., alternately apply parallel decoding to the subvectors induced by the vertices of A and to the subvectors induced by the vertices of B. This produces a sequence of vectors, let us call it the decoding sequence

$$x^{(0)} = x, \; x^{(1)} = L(x^{(0)}), \; x^{(2)} = R(x^{(1)}), \; x^{(3)} = L(x^{(2)}) \cdots. \tag{4}$$

The procedure stops if it encounters a fixed point or after having moved $O(\log N)$ steps, where the constant can be expressed explicitly via the parameters of G, C_0, C_1. The number of gates at each round is at most $O(N \log N)$.

Without loss of generality, because C is linear, we may suppose that the initial uncorrupted codeword is the zero vector so that $x = (x_1 \ldots, x_N)$ is actually the error vector. Let x be a vector $x \in \{0, 1\}^N$ and let $y = L(x)$. It will be useful to identify vectors of $\{0, 1\}^N$ with their supports, i.e., think of x and y as the edge sets of subgraphs of G. Let $v \in A$ be a left vertex. Let us say that v is a (left) *survivor* of x if $y_j = 1$ for some $j \in E_v$. In other words, if x is an error vector, the survivors of x are all the vertices incident to the edge set represented by $L(x)$, i.e., the error vector *after* the next decoding step. Similarly, let $y' = R(x)$, and define a right vertex $v \in B$ to be a (right) survivor of x if $y'_j = 1$ for some $j \in E_v$.

Note that if x has no left survivors (right survivors) then $L(x) = 0$ ($R(x) = 0$). In [9], it is proved that if the number s of left survivors of x is small enough, then the number s' of right survivors of $y = L(x)$ is strictly smaller, i.e., $s' \leq \beta s$ with $\beta < 1$. By left–right symmetry (when the two codes C_1 and C_0 are the same) one obtains that the decoding sequence (4) converges to the zero vector in a number of steps logarithmic in n.

Let $x \in \{0, 1\}^N$ be a vector and $v \in A \cup B$ a vertex. The number of ones of x in the neighborhood v, $|E_v \cap \operatorname{supp}(x)|$, will be called the x-*degree* of v. It should be clear that we have the following.

Proposition 1: The minimum x-degree $\phi(x)$ of any survivor of x (left or right) satisfies

$$\phi(x) \geq \min(d_0/2, d_1/2).$$

Define the *fan* of a code C as follows:

$$\phi = \phi(C) := \min_{x \in \{0,1\}^N} \phi(x).$$

We emphasize this parameter rather than write simply $d_1/2$ or $d_0/2$ as in [9] because in Section IV we shall give a modified construction of an expander code that improves upon Proposition 1.

The key result of [9] that we need to recall here can be reformulated in terms of ϕ as follows.

Proposition 2: Suppose $2\phi > 3\lambda$ where λ is the second largest eigenvalue of the adjacency matrix of G. Let $\alpha < 1$. Then there exists $\beta < 1$ such that: if the number s of left survivors (right survivors) of x satisfies

$$s \leq \alpha n(\phi - \lambda)/\Delta$$

then the number s' of right survivors (left survivors) of $L(x)$ ($R(x)$) satisfies

$$s' \leq \beta s.$$

Recall that when G is a Ramanujan graph then $\lambda/\Delta \leq 2\sqrt{\Delta - 1}/\Delta$ which vanishes when Δ grows.

The proof of Proposition 2 is that of [9, Lemma 5 and Theorem 6] with $d_0/2$ replaced by ϕ: we refer the reader to [9] for the details. We remark that to prove that L (resp., R) has the claimed "contraction" property it suffices to replace ϕ in the claim with $d_1/2$ (resp., $d_0/2$).

III. Error Probability

Let us submit the expander code C of the preceding section to the binary-symmetric channel with transition probability p. Let $P_e(C, p)$ be the probability that the iterative decoding algorithm of the preceding section fails. We wish to study the asymptotic behavior of $P_e(C, p)$, i.e., Δ will be large, but fixed, and the number n of vertices of A or B will be allowed to go to infinity.

The idea developed in this section is that whenever p is less than the decoding threshold of the first code C_0, then the first decoding iteration will leave a very small proportion of bits in error. Therefore, the second code C_1 need have only a small redundancy for the iterative algorithm to converge.

The goal of this section is to prove the following theorem.

Theorem 3: For a given rate R, any $\varepsilon > 0$, and $\alpha < 1$ there exists an expander code C of length N such $P_e(C, p) \leq 2^{-\alpha N f_3(R, p)}$, where

$$f_3(R, p) = \max_{R \leq R_0 < 1 - H(p)} E(R_0, p) H^{-1}(R_0 - R)/2 - \varepsilon. \qquad (5)$$

Proof: Given a small $\varepsilon_1 > 0$, let us chose the value of Δ so that the error probability of complete decoding of the code C_0 is bounded above by $\pi(R_0, p) = 2^{-\Delta(E(R_0, p) - \varepsilon_1)}$, where $E(R_0, p)$ is the random coding exponent. Furthermore, let us make sure that the chosen Δ is sufficiently large so that the assumption of Proposition 2 is satisfied. This is possible since for any rate R_0 or R_1 one can choose the codes C_0 (C_1) of sufficiently large length Δ so that their distance d_0 (or d_1) is much greater than $\sqrt{\Delta}$, so $2\phi \geq d > 3\lambda$ (recall the Ramanujan property of G). In particular, let us assume that

$$2\phi/\Delta = \min(H^{-1}(1 - R_0), H^{-1}(1 - R_1)) - \varepsilon_2. \qquad (6)$$

Our main point here is that whenever the proportion of left survivors of a random error vector is strictly less than $(\phi - \lambda)/\Delta$ then, by Proposition 2, the decoding algorithm converges correctly. We have, therefore, for any $\alpha < 1$ and $R_0 > R$

$$P_e(C, p) \leq \sum_{i \geq i^*} \binom{n}{i} \pi(R_0, p)^i (1 - \pi(R_0, p))^{n-i} \qquad (7)$$

where $i^* = \alpha n(\phi - \lambda)/\Delta$. Note that for any fixed α we have $\pi(R_0, p) < i^*/n$ for Δ sufficiently large. So the dominating term in the above sum will be the one with $i = i^*$; all the other ones contribute to a nonexponential factor only. The opposite of the logarithm of this term has the form

$$\begin{aligned} &\alpha n\Delta \left(\frac{\phi}{\Delta} - \frac{\lambda}{\Delta}\right) (E(R_0, p) - \varepsilon_1) - \log \binom{n}{i^*} \\ &\quad - (n - i^*) \log(1 - \pi(R_0, p)) \\ &\geq \alpha n\Delta \left[\frac{\phi}{\Delta} E(R_0, p) - \frac{\phi}{\Delta} \varepsilon_1 \right. \\ &\quad \left. - \frac{2}{\sqrt{\Delta}} E(R_0, p) + O(\Delta^{-1})\right] \\ &\geq \alpha N \tfrac{1}{2} E(R_0, p) \min(H^{-1}(1 - R_0), H^{-1}(R_0 - R)) \\ &\quad + O(\Delta^{-1/2}). \end{aligned}$$

the second inequality by (6) and (3).

The smallest error probability is therefore obtained by computing the maximum over R_0 of $f_3^*(R, R_0, p)$ where

$$f_3^*(R, R_0, p) = \frac{1}{2} E(R_0, p) \min(H^{-1}(1 - R_0), H^{-1}(R_0 - R)) - \varepsilon \qquad (R \leq R_0 < 1 - H(p)).$$

It is straightforward to check numerically that the error exponent reduces to

$$f_3(R, p) = \max_{R \leq R_0 < 1 - H(p)} E(R_0, p) H^{-1}(R_0 - R)/2 - \varepsilon. \qquad \square$$

This result already gives codes with positive error exponent for all code rates less than capacity. In the next section we modify the code construction and prove Theorem 2.

IV. Replicated Expander Codes

A. Construction

We now modify the construction of the codes of Section II-B as follows: we stay with the same bipartite graph G, we keep its edges numbered $1, \ldots N$, $N = n\Delta$, and for every $v \in A \cup B$ keep the ordering $v(1), \ldots, v(\Delta)$ of the neighborhood E_v of v. But this time, we choose the constituent codes C_v (in practice there will again be two of them, $C_v = C_0$ for $v \in A$ and $C_v = C_1$ for $v \in B$) to be of length $t\Delta$ for some fixed integer $t \geq 1$. The length of the new code will be tN. For a vector $x \in \{0, 1\}^{tN}$ we modify the definition of the subvector x_v to be

$$x_v = \left(x_{t(v(i)-1)+j}\right)_{i=1, \ldots, \Delta, j=1, \ldots t}$$

so that now $x_v \in \{0, 1\}^{t\Delta}$. The new code C is defined to be the set of vectors $x \in \{0, 1\}^{tN}$ such that for every $v \in A \cup B$, $x_v \in C_v$. It should be clear that if R_0 and R_1 are again the rates of C_0 and C_1 then the linear code C has again rate R satisfying (3). This construction can be thought of alternately as replicating t times every edge in G.

B. Minimum Distance

We can view every t-tuple as an element of the additive group of $\mathbb{F}_{2^t}$, so that, by separating vectors $x \in \{0, 1\}^{t\Delta}$ into subvectors of length t

$$(x_1, \ldots x_t), (x_{t+1}, \ldots, x_{2t}) \cdots$$

and so on, we have a natural additive mapping of $\{0, 1\}^{t\Delta}$ onto $(\mathbb{F}_{2^t})^{\Delta}$. The linear binary code C_0 (and also C_1) can, therefore, be also thought of as a 2^t-ary additive code. Let us call $D_0(C_0)$ the minimum 2^t-ary Hamming distance of C_0: in other words, D_0 is the smallest number of nonzero t-tuples

$$\left(x_{t(i-1)+1}, \ldots, x_{t(i-1)+t-1}, x_{ti}\right), \qquad 1 \leq i \leq \Delta$$

of any nonzero codeword $x = (x_1, \ldots, x_{t\Delta}) \in C_0$ (sometimes called the *phased bursts* weight).

Consider briefly the case when $C_1 = C_0$. As in [7], we can invoke the Alon–Chung lemma [1].

Lemma 4 (Alon–Chung): Let G be a Δ-regular graph on n' vertices with second largest eigenvalue λ. Let S be a subset of vertices. Then the average degree $\overline{d}_S$ of the subgraph induced by S satisfies

$$\overline{d}_S \leq \Delta \frac{|S|}{n'} + \lambda \left(1 - \frac{|S|}{n'}\right).$$

Let D be the minimum distance of C and let $D/(tN)$ be its relative minimum distance. Let $\delta_0 = d_0/(t\Delta)$ be the relative minimum distance of C_0. We obtain the followng theorem.

Theorem 5: The relative minimum distance of C satisfies

$$\frac{D}{tN} \geq \delta_0 \frac{D_0/\Delta - \lambda/\Delta}{1 - \lambda/\Delta}.$$

Proof: Any nonzero codeword of C represents the set of edges of a subgraph of G with minimum (and hence average) degree at least D_0. Note that here $n' = 2n$. Apply the Alon–Chung lemma to obtain that the number of vertices $|S|$ of the subgraph is at least $n'(D_0 - \lambda)/(\Delta - \lambda)$ and argue that the number of edges of the subgraph, i.e., the weight of the codeword, is at least $d_0|S|/2$. □

Note that when $t = 1$ (the usual case) we have $D_0/\Delta = \delta_0$. The point of this generalized construction is that when $t > 1$ we can have $D_0/\Delta > \delta_0 = d_0/(t\Delta)$ in which case Theorem 5 gives an improved lower bound.

In particular, for large Δ and large t we get the following result.

Proposition 3: Let R_0 be fixed, $0 < R_0 < 1$, and let $\varepsilon > 0$. For Δ and t big enough, there exists a code C of length $t\Delta$ and rate R_0 such that

1) $\delta_0 = d_0/(t\Delta) > H^{-1}(1 - R_0) - \varepsilon$;
2) $D_0 \geq (1 - R_0)\Delta - \varepsilon$.

To see this, simply recall that, asymptotically, most binary linear codes lie on the binary Varshamov–Gilbert bound $R_0 = 1 - H(\delta)$ and most 2^t-ary linear codes lie on the 2^t-ary Varshamov–Gilbert bound which gets, as t goes to infinity, arbitrarily close to the Singleton bound $R_0 + D_0/\Delta \leq 1$. This last fact is classically stated for $\mathbb{F}_{2^t}$-linear codes, but the reader will be readily convinced that this carries over without any difficulty to 2^t-ary additive codes. Indeed, this follows by the usual procedure of augmenting the code C_0 with cosets from the quotient group $(\mathbb{F}_{2^t})^{\Delta}/C_0$.

C. Iterative Decoding

We now switch to the iterative decoding of replicated expander codes. We now again use two constituent codes C_0 and C_1 of length $t\Delta$. The algorithm is really the same as in Section II-B, namely, define $L(x)$ $(R(x))$ to be a vector y such that for every $v \in A$ $(v \in B)$ y_v is one of the codewords of C_0 (C_1) closest to x_v. As before, let the decoding sequence of a vector $x \in \{0, 1\}^{tN}$ be as in (4). Let $y = L(x)$ and $y' = R(x)$ and define as before a left (right) survivor of x to be a vertex $v \in A$ $(v \in B)$ such that y_v (y'_v) is a nonzero vector of $\{0, 1\}^{t\Delta}$. Finally, let the x-degree of a vertex $v \in A \cup B$ be the number of $j \in E_v$ such that the subvector of $\{0, 1\}^t$

$$\left(x_{t(j-1)+1}, x_{t(j-1)+2}, \ldots, x_{t(j-1)+t-1}, x_{tj}\right)$$

is nonzero.

Proposition 1 can now be replaced by the following.

Proposition 4: The minimum x-degree $\phi(x)$ of any survivor of x (left or right) satisfies

$$\phi(x) \geq \min(D_0/2, D_1/2).$$

Proposition 2 (see [9, Lemma 5]) is really a pure graph-theoretic statement, and it continues to hold when ϕ is now the new minimum value of $\phi(x)$ when x ranges over all $x \in \{0, 1\}^{tN}$.

We can now replace ϕ in the upper bound (7) on the decoding error probability $P_e(C, p)$ by any quantity arbitrarily close to $(1 - R_1)\Delta/2$. Otherwise the calculation in the proof of Theorem 3 is unchanged, which finally establishes Theorem 2.

V. Comments

1) The decoding algorithm of [9] and Section II-B is especially adapted to parallel computations. As in [7], however, it can be adapted so as to yield a sequential algorithm with complexity $O(N)$. Here is a simple way of doing this. Let the first L iteration be executed sequentially, i.e., one decodes, one after the other, every subcode associated to every one of the vertices of A. Note that when this is done, every left subvector y_v, $v \in A$, of the resulting vector $y = L(x)$ has zero syndrome. Let $S(y) \subset B$ be the set of right vertices v for which y_v has *nonzero* syndrome. Let $z = R(y)$ be the vector after the next iteration, and notice that all the left vertices w that are *not* neighbors of some $v \in S(y)$ must be left untouched by the decoding procedure, i.e., $z_w = y_w$, so that they stay with zero syndrome. This means that at the *next* iteration, when computing $L(z)$, all these left vertices w need not be examined at all.

   The trick is, therefore, to keep track of every vertex that corresponds to a nonzero syndrome. Specifically, at iteration i, when computing $x^{(i+1)}$ from $x^{(i)}$, we decode only subvectors $x_v^{(i)}$ for vertices v belonging to a privileged subset S^i that was computed at the previous iteration, and we prepare the way for the next iteration by computing S^{i+1} which is the set of neighboring vertices of all those $v \in S^i$ for which $x_v^{(i)}$ had nonzero syndrome.

   Finally, let $y = L(x)$ and notice that any vertex v such that y_v has nonzero syndrome must be a neighbor of left survivor of x. Now whenever Proposition 2 applies, the number of survivors decreases geometrically with the number of iterations, so that the total number of survivors throughout the total number of iterations is linear in n. Therefore, the sum, over all i, of the number of vertices for which $x^{(i)}$ has nonzero syndrome is also linear in n, and so is the total number of vertices examined by the decoding procedure.

2) In the case $C_0 = C_1$, the distance bound obtained in Proposition 3 for replicated expander codes gives the asymptotic result.

Theorem 6: For any R_0, $0 < R_0 < 1$, and $\varepsilon > 0$ there exists an expander code C of rate $R \geq 2R_0 - 1$ and relative distance $\delta = (1 - R_0)H^{-1}(1 - R_0) - \varepsilon$. Iterative decoding applied to this code corrects any $\alpha < \delta/4$ fraction of errors.

This is an improvement over Theorem 1: in particular, we obtain codes C with positive rates for all $0 \leq \delta < 0.055$.

3) As mentioned earlier, replicated expander codes can be thought of as expander codes with the graph G replaced by a graph with multiple edges. Such a graph is a graph with many small cycles. We have found here that replicated expander codes have a better behavior under iterative decoding than the unreplicated ones: this gives substance to the empirical finding sometimes claimed in the turbo coding community that "iterative decoding sometimes works better when the underlying graph has small cycles."

References

[1] N. Alon and F. R. K. Chung, "Explicit construction of linear sized tolerant networks," *Discr. Math.*, vol. 72, pp. 15–19, 1988.

[2] I. Dumer, "Concatenated codes and their multilevel generalizations," in *Handbook of Coding Theory*, V. Pless and W. C. Huffman, Eds. Amsterdam, The Netherlands: Elsevier, 1998, vol. II, pp. 1911–1988.

[3] G. D. Forney, Jr., *Concatenated Codes*. Cambridge, MA: MIT Press, 1966.

[4] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[5] A. Lubotsky, R. Philips, and P. Sarnak, “Ramanujan graphs,” *Combinatorica*, vol. 8, no. 3, pp. 261–277, 1988.

[6] G. A. Margulis, “Explicit group theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators,” *Probl. Inform. Transm.*, vol. 24, no. 1, pp. 39–46, 1988.

[7] M. Sipser and D. A. Spielman, “Expander codes,” *IEEE Trans. Inform. Theory*, vol. 42, pp. 1710–1722, Nov. 1996.

[8] M. Tanner, “A recursive approach to low-complexity codes,” *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 533–547, Sept. 1981.

[9] G. Zémor, “On expander codes,” *IEEE Trans. Inform. Theory (Special Issue on Codes on Graphs and Iterative Algorithms)*, vol. 47, pp. 835–837, Feb. 2001.

[10] V. V. Zyablov, “An estimate of complexity of constructing binary linear cascade codes,” *Probl. Inform. Transm.*, vol. 7, no. 1, pp. 3–10, 1971.