REFERENCES

[1] B. Arazi, "The optimal burst-error-correcting capability of the codes generated by $f(X) = (X^p + 1)(X^q + 1)/(X + 1)$," *Inform. Contr.*, vol. 39, pp. 303-314, 1978.

[2] A. Dür, "The decoding of extended Reed-Solomon codes," *Discrete Math.*, vol. 90, pp. 21-40, 1991.

[3] J. H. Grace and A. Young, *The Algebra of Invariants*. Cambridge: Cambridge Univ. Press, 1903.

[4] R. D. Jenks, R. S. Sutor, and S. M. Watt, "Scratchpad II: An abstract datatype system for mathematical computation," *Springer LNCS*, vol. 296, pp. 12-37, 1988.

[5] T. Kasami, "Comments on 'Determining the burst-correcting limit of cyclic codes,' " *IEEE Trans. Inform. Theory*, vol. IT-27, p. 812, Nov. 1981.

[6] B. W. Kernighan and D. M. Ritchie, *The C Programming Language*. New York: Prentice Hall, 1978.

[7] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice Hall, 1983.

[8] J. H. van Lint and R. M. Wilson, "On the minimum distance of cyclic codes," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 23-40, Jan. 1986.

[9] H. J. Matt and J. L. Massey, "Determining the burst-correcting limit of cyclic codes," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 289-297, May 1980.

[10] J. L. Massey and T. Schaub, "Linear complexity in coding theory," *Springer LNCS*, vol. 311, pp. 19-32, 1988.

[11] J. L. Massey, "Shift-register synthesis and BCH decoding," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 122-127, Jan. 1969.

[12] W. W. Peterson and E. J. Weldon, Jr., *Error-Correcting Codes*, 2nd ed. Cambridge, MA: M.I.T., 1972.

Construction of Asymptotically Good Low-Rate Error-Correcting Codes through Pseudo-Random Graphs

Noga Alon, Jehoshua Bruck, *Member, IEEE,* Joseph Naor, Moni Naor, Ron M. Roth, *Member, IEEE*

***Abstract*—A new technique, based on the pseudo-random properties of certain graphs, known as expanders, is used to obtain new simple explicit constructions of asymptotically good codes. In one of the constructions, the expanders are used to enhance Justesen codes by replicating, shuffling and then regrouping the code coordinates. For any fixed (small) rate, and for sufficiently large alphabet, the codes thus obtained lie above the Zyablov bound. Using these codes as outer codes in a concatenated scheme, a second asymptotic good construction is obtained which applies to small alphabets (say, GF(2)) as well. Although these concatenated codes lie below Zyablov bound, they are still superior to previously-known explicit constructions in the zero-rate neighborhood.**

***Index Terms*—Expanders, Justesen codes, Zyablov bound, independent sets.**

I. INTRODUCTION

An infinite sequence of codes $S = \{C_i\}_{i=1}^{\infty}$ over an alphabet Σ of q elements is called asymptotically good if the lengths n_i, sizes M_i and minimum distances d_i of the C_i's satisfy the following: 1) $\lim_{i\to\infty} n_i = \infty$ and 2) both the rate of the sequence $R \triangleq \liminf_{i\to\infty} \log_q M_i/n_i$, and its relative minimum distance $\delta \triangleq \liminf_{i\to\infty} d_i/n_i$, are strictly greater than zero.

By the Gilbert-Varshamov bound, for any $\delta \in [0, 1 - 1/q)$ there exists a good sequence of codes over Σ of relative minimum distance δ and of rate $R \geq R_{\mathrm{GV}}(\delta)$, where

$$R_{\mathrm{GV}}(\delta) \triangleq 1 - H_q(\delta), \tag{1}$$

and $H_q(x) \triangleq -x \cdot \log_q x - (1-x) \cdot \log_q (1-x) + x \cdot \log_q (q-1), 0 \leq x \leq 1 - 1/q$. Furthermore, the seminal works of Tsfasman *et al.* [10], [13], [25] show the existence of good code sequences beyond the Gilbert-Varshamov bound for $q \geq 46$.

A code sequence $S = \{C_i\}_{i=1}^{\infty}$ over an alphabet Σ is called *constructive* if there exists an algorithm that computes any codeword of C_i in time complexity which is polynomial in the length of C_i. In particular, if the codes C_i are linear, then S is constructive, if and only if the generator matrices of the C_i can be computed in polynomial-time.

A *parametric family of sequences* over an alphabet Σ, $|\Sigma| = q$, is a set of code sequences $\mathcal{S} = \{S(\delta)\}_{0 \leq \delta \leq 1-1/q}$ where each $S(\delta)$ is a code sequence of relative minimum distance $\geq \delta$ over Σ. For each family of code sequences we associate a function $R(\delta)$ which stands for the rate of $S(\delta)$.

A parametric family $\mathcal{S}$ is called *uniformly constructive* if 1) there exists a constant c, independent of δ, such that the encoding of a codeword of any code in $S(\delta)$ of length n can be carried out in n^c steps; and 2) $R(\delta) > 0$ whenever $\delta < 1 - 1/q$. Note that $R_{\mathrm{GV}}(\delta) > 0$ for $\delta < 1 - 1/q$, whereas by the Plotkin bound we must have $R(\delta) = 0$ for $\delta \geq 1 - 1/q$. Such a uniformity definition is aimed to characterize good low-rate code sequences that can be efficiently constructed, no matter how close the rate is to zero.

By using a concatenated code construction, with a Reed-Solomon code as the outer code and a code which attains the Gilbert-Varshamov bound as the inner code, one can obtain a family of constructive sequences whose rate function $R(\delta)$ satisfies the Zyablov bound $R(\delta) \geq R_{\mathrm{Zyablov}}(\delta)$ [29], where

$$R_{\mathrm{Zyablov}}(\delta) \triangleq \max_{\delta \leq \mu \leq 1 - \frac{1}{q}} \left(1 - H_q(\mu)\right)\left(1 - \frac{\delta}{\mu}\right). \tag{2}$$

However, searching for the inner code by any known algorithm requires time complexity that is exponential in the inner codelength. Hence, constructing the generator matrix of such a concatenated code of length n and relative minimum distance δ will require the order of $n^{c(\delta)}$ operations, where $\lim_{\delta \to 1-1/q} c(\delta) = \infty$. Hence, such a code sequence family is nonuniformly constructive.

The exponential search is avoided in Justesen codes [9] and in constructions derived thereof [22]-[24], [27], where the inner codes exhaust all members of Wozencraft's ensemble of randomly shifted codes. Justesen's construction is also "explicit" in the sense that once the rates of the inner and outer codes have been computed, the entries of the generator matrices of the codes can be written as closed formulas, and no searching is required. However, the rate function $R_{\mathrm{Jus}}(\delta)$, associated with Justesen's construction, vanishes for all $\delta > H_q^{-1}(\frac{1}{2})$, and $H_q^{-1}(\frac{1}{2})$ can be readily verified to be strictly smaller than $1 - 1/q$. Therefore, Justesen codes do not comply with requirement 2) of uniform constructiveness. The same

Manuscript received October 10, 1990. This work was done while N. Alon was on sabbatical from Tel-Aviv University. This work was presented in part at the IEEE International Symposium on Information Theory, Budapest, Hungary, June 24-28, 1991.

N. Alon is with the Department of Mathematics, Tel-Aviv University, Tel-Aviv 69978, Israel.

J. Bruck and M. Naor are with the IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120.

J. Naor and R. M. Roth are with the Department of Computer Science, Technion—Israel Institute of Technology, Haifa 32000, Israel.

IEEE Log Number 9104381.

holds also for some other known improvements on Justesen codes [23], [28].

Uniformly constructive families of codes over GF(q) were obtained by Weldon [27] and Sugiyama *et al.* [22], [24], where the outer Reed–Solomon codes were replaced by much longer codes over GF(q^m), at the expense of not attaining the Singleton bound. The rate $R_{\mathrm{SKHN}}(\delta)$ of the construction obtained in [24] satisfies

$$R_{\mathrm{SKHN}}(\delta) \geq \max_{\delta \leq \mu \leq 1-\frac{1}{q}} \left(1 - H_q(\mu)\right)\left(1 - \frac{\delta}{\mu}\left(1 + \ln\frac{\mu}{\delta}\right)\right). \tag{3}$$

Katsman, Tsfasman, and Vlădut [10] found a construction of algebraic–geometric codes which, when concatenated with specific inner codes, yield a uniformly constructive family that lies above the Zyablov bound. However, since the time complexity of finding the generator matrices of these codes is proportional to n^{32} [5], they can hardly be called constructive from any practical perspective. Apart from this construction, (3) yields the best uniformly constructive family for sufficiently low rates (i.e., when δ is close to $1 - 1/q$), to the best knowledge of the authors.

In this correspondence we introduce new simple uniformly constructive (in fact, explicit) families of asymptotically good codes, by applying a novel technique based on the pseudo-random characteristics of graphs known as *expanders*. More specifically, we make use of explicit constructions of families of Δ-regular undirected graphs $G = (V, E)$ with the following property: Fix some real number $\delta_0 \in (0, 1]$; then for any subset of vertices $B \subseteq V$ of size $\geq \delta_0 |V|$, the fraction of vertices in V that have at least one neighbor in B approaches unity "fast" as $\Delta \to \infty$. A precise definition of the expanders used, and their properties, are presented in Section II.

Given such a graph with $n = |V|$ vertices and a finite field Φ, we then show how to define a so-called *expander mapping* (or *expander code*) $C_{\exp}: \Phi^n \to (\Phi^\Delta)^n$, such that every input n-tuple over Φ of Hamming weight $\geq \delta_0 n$ is mapped into an output n-tuple over Φ^Δ whose Hamming weight (measured over Φ^Δ) is "close" to n. The notion of code amplification through expanders has been inspired by recent applications of expanders to deterministic simulation of randomized algorithms [1], [3], [6], [8], [11], [17]. In a way, the application of expanders presented in this correspondence can be viewed as an improvement on the method introduced in [17], in the sense that the codes that may be obtained are better.

These expander codes will serve as building blocks in our new asymptotically good constructions. The first construction, referred to as Construction $\mathscr{C}_1$, is obtained by taking the codewords of any good code sequence over a finite field Φ (say, Justesen codes), and then applying the expander code $C_{\exp}$, resulting in a code over the alphabet Φ^Δ whose rate is proportional to $1/\Delta$. The choice of Δ and the field Φ will depend on the prescribed size q of the underlying alphabet and the relative minimum distance δ. As we show in Section III, the rate $R_{\mathscr{C}_1}(\delta, q)$ of Construction $\mathscr{C}_1$ satisfies

$$R_{\mathscr{C}_1}(\delta, q) \geq \gamma_0(1 - \delta) - \frac{\gamma_1}{\log_2 q}, \tag{4}$$

for some positive constants γ_0 and γ_1. Note that, for sufficiently large q, (4) resembles the Singleton bound (or the rate attainable by the so-called modular code construction described in [10]), except for the multiplier γ_0 (which is approximately 0.021).

Construction $\mathscr{C}_1$ satisfies criterion 1) of the uniformity definition. As for criterion 2), the δ-interval for which $R_{\mathscr{C}_1}(\delta, q) = 0$ shrinks to zero length when $q \to \infty$; hence, $\mathscr{C}_1$ is "nearly-uniformly" constructive, and this fact will be exploited in our second construction. However, the significance of Construction $\mathscr{C}_1$ is manifest in the fact that, as a fairly simple construction, it exceeds the Zyablov bound for the zero-rate neighborhood and for sufficiently large alphabet sizes q.

When the size of the underlying alphabet is fixed (say, $q = 2$), Construction $\mathscr{C}_1$ fails to improve on previously-known constructions. However, we can use Construction $\mathscr{C}_1$ to introduce good code sequences over specific fields $F = \mathrm{GF}(q)$ by means of concatenation. The new codes will be referred to as Construction $\mathscr{C}_2$ and will be discussed in Section IV. Construction $\mathscr{C}_2$ is obtained by using Construction $\mathscr{C}_1$ over $\Sigma = (\mathrm{GF}(q^m))^\Delta$ as the outer code, with each output symbol (over Σ) undergoing a second level of encoding by codes of dimension $m\Delta$ over F. Such a scheme yields a uniformly constructive family of linear codes over F that satisfies the inequality

$$R_{\mathscr{C}_2}(\delta) \geq \max_{\delta \leq \mu \leq 1-\frac{1}{q}} \gamma_0\left(1 - H_q(\mu)\right)\left(1 - \frac{\delta}{\mu}\right). \tag{5}$$

The bound (5) resembles the Zyablov bound (2), except for the multiplier γ_0, due to which (5) lies beneath the curve (2). However, when the relative minimum distance δ is close enough to $1 - 1/q$, the right-hand side of (5) becomes larger than the right-hand side of (3). For instance, in the binary case ($q = 2$), the lower bound (5) exceeds the bound (3) for $0.45 \leq \delta \leq 0.5$, which corresponds to the low-rate range $R \leq 2.5 \times 10^{-6}$.

The significance of Construction $\mathscr{C}_2$ can be better illustrated if we express the rate R in terms of $\epsilon \stackrel{\Delta}{=} 1 - 1/q - \delta$. We take the binary case as a typical (and the most important) example. In this case, $\epsilon = \frac{1}{2} - \delta$, and, when ϵ is small, (5) becomes

$$R_{\mathscr{C}_2}\left(\frac{1}{2} - \epsilon\right) \geq \frac{16}{27 \ln 2}\gamma_0\epsilon^3 - O(\epsilon^4).$$

The same bound is obtained by (2) if we replace γ_0 by 1. Hence, the attainable rates in both the Zyablov bound and Construction $\mathscr{C}_2$ are of the same order i.e., proportional to ϵ^3. Repeating the calculation for (3), however, yields a lower bound which is proportional to ϵ^4. For comparison, it is worthwhile noting that, in terms of ϵ, the Gilbert–Varshamov bound for $q = 2$ takes the form

$$R_{\mathrm{GV}}\left(\frac{1}{2} - \epsilon\right) = \frac{2}{\ln^2}\epsilon^2 - O(\epsilon^4),$$

whereas the McEliece–Rodemich–Rumsey–Welch upper bound [15, p. 559] yields

$$R_{\mathrm{MRRW}}\left(\frac{1}{2} - \epsilon\right) = 2\epsilon^2 \log_2(1/\epsilon) + O(\epsilon^2).$$

Like in previous constructions [9], [22], [27], the inner code in Construction $\mathscr{C}_2$ is taken as Wozencraft's ensemble. It thus turns out that for any fixed q, the frequency of occurrence of each element of GF(q) in *any* nonzero codeword of $\mathscr{C}_2$ approaches $1/q$ as $\delta \to 1 - 1/q$ (and the codelength tends to infinity). In Section V, we present an application of this property to the so-called *t-independent set problem*, that is, finding a small set of vectors in $\{0, 1\}^m$ such that the subvectors obtained by extracting any t coordinates exhaust all 2^t binary t-tuples. Using a technique intro-

duced in [17], we construct such a set of size $ct2^{3t}$-log m for any fixed t and for sufficiently large m, where c is an absolute constant (independent of t). For related work see [4], [12], [21].

II. Pseudo-Random Graphs

Expanders are graphs which behave in many ways like sparse random graphs. Expanders, which are the subject of extensive literature, are, roughly, graphs in which every set of at most half of the vertices has many neighbors outside the set. As shown in [2], the expanding properties of a graph are closely related to the eigenvalues of its adjacency matrix. Since the property we need here is proved by using the eigenvalues, we do not mention the common definition of an expander, and only define the graphs we need in terms of their eigenvalues.

Let $G = (V,E)$ be a Δ-regular graph with n vertices and let $A = A_G = [a_{uv}]_{u,v\in V}$ be its adjacency matrix given by $a_{uv} = 1$ if $uv \in E$ and $a_{uv} = 0$, otherwise. Since G is Δ-regular the largest eigenvalue of A is Δ, corresponding to the all-one eigenvector. Let $\lambda_1, \cdots, \lambda_n$ be all the eigenvalues of G, (with multiplicities), where $\Delta = |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$. Define $\lambda(G) = |\lambda_2|$. As we show below, if $\lambda(G)$ is much smaller than Δ, then G has a strong pseudo-random property.

Theorem 1: Let $G = (V, E)$ be a Δ-regular graph with $n = |V|$ and $\lambda = \lambda(G)$. For a vertex $v \in V$ and a subset B of V denote by $N(v)$ the set of all neighbors of v in G, and let $N_B(v) = N(v) \cap B$ denote the set of all neighbors of v in B. Then, for every subset B of cardinality bn of V,

$$\sum_{v\in V} \left(|N_B(v)| - b\Delta\right)^2 \leq \lambda^2 b(1-b)n.$$

Observe that in a random Δ-regular graph each vertex v would tend to have about $b\Delta$ neighbors in each set of size bn. This theorem shows that if λ is much smaller than Δ then for most vertices v, $N_B(v)$ is not too far from $b\Delta$.

Proof: Let A be the adjacency matrix of G and define a vector $f: V \mapsto R$ by $f(v) = 1 - b$ for $v \in B$ and $f(v) = -b$ for $v \notin B$. Clearly $\Sigma_{v\in V} f(v) = 0$ i.e., f is orthogonal to the eigenvector of the largest eigenvalue of A. Therefore,

$$\langle Af, Af\rangle \leq \lambda^2 \langle f, f\rangle$$

($\langle \cdot, \cdot \rangle$ standing for scalar product of vectors). The right-hand side of the last inequality is $\lambda^2(bn(1-b)^2 + (1-b)nb^2) = \lambda^2 b(1-b)n$. The left-hand side is

$$\sum_{v\in V} \left((1-b)|N_B(v)| - b\left(\Delta - |N_B(v)|\right)\right)^2 = \sum_{v\in V} \left(|N_B(v)| - b\Delta\right)^2.$$

The desired result follows. □

Corollary 1: Let $G = (V, E)$ be a Δ-regular graph with $n = |V|$ and $\lambda = \lambda(G)$, and let B be a subset of cardinality bn of V. Let $t = |\{v \in V : N_B(v) = \emptyset\}|$ be the number of vertices of G that have no neighbors in B. Then

$$t \leq \frac{\lambda^2(1-b)n}{b\Delta^2}.$$

In particular, if $\lambda \leq 2\sqrt{\Delta - 1}$ then

$$t \leq \frac{4(\Delta-1)(1-b)n}{b\Delta^2} \leq \frac{4n}{b\Delta}.$$

Proof: Define $T = \{v \in V : N_B(v) = \emptyset\}$. For each vertex $v \in T$, $|N_B(v)| = 0$. Therefore, by Theorem 1:

$$tb^2\Delta^2 = \sum_{v\in T} \left(|N_B(v)| - b\Delta\right)^2 \leq \sum_{v\in V} \left(|N_B(v)| - b\Delta\right)^2 \leq \lambda^2 b(1-b)n.$$

This completes the proof. □

In view of the last two results, it is natural to ask how far from Δ the value of $\lambda(G)$ can be. It is known [2], [18] that the second largest eigenvalue of any Δ-regular graph with diameter k is at least $2\sqrt{\Delta - 1}\,(1 - O(1/k))$. Therefore, in any infinite family of Δ-regular graphs $\{G_i = (V_i, E_i)\}_{|V_i|\to\infty}$,

$$\limsup_{i\to\infty} \lambda(G_i) \geq 2\sqrt{\Delta - 1}. \tag{6}$$

Lubotzky, Phillips, and Sarnak [14], and independently, Margulis [16], gave, for every $\Delta = p + 1$ where p is a prime congruent to 1 modulo 4, explicit constructions of infinite families of Δ-regular graphs G_i with second largest eigenvalues $\lambda(G_i) \leq 2\sqrt{\Delta - 1}$. For the sake of completeness, we next describe these graphs.

For an integer m, denote by Z_m the ring of integers modulo m. Let p and π be unequal primes, both congruent to 1 modulo 4, such that p is a quadratic residue modulo π. Let $P = PSL(2, Z_\pi)$ denote the factor group of the group of all 2×2 matrices over Z_π with determinant 1 modulo its normal subgroup consisting of the identity I and its (additive) inverse $-I$. The elements of P are thus simply 2×2 matrices over Z_π of determinant 1, where both matrices A and $-A$ are regarded as the same element $\pm A$.

The graphs we describe are Cayley graphs of P i.e., their vertices are all $\pi(\pi^2 - 1)/2$ elements of P and two such elements A and B are adjacent, if and only if AB^{-1} belongs to a prescribed set Q of elements of P that we define next.

A well-known theorem of Jacobi asserts that the number of ways of representing a positive integer n as a sum of four squares is precisely eight times the sum of the divisors of n that are not divisible by 4. This easily implies that there are precisely $p + 1$ vectors $\boldsymbol{a} = [a_0, a_1, a_2, a_3]$, where a_0 is an odd positive integer, a_1, a_2, a_3 are even integers, and $a_0^2 + a_1^2 + a_2^2 + a_3^2 = p$. Associate each such vector $\boldsymbol{a}$ with the member

$$M_{\mathbf{a}} \triangleq \pm\frac{1}{\sqrt{p}}\begin{bmatrix} a_0 + ia_1 & a_2 + ia_3 \\ -a_2 + ia_3 & a_0 - ia_1 \end{bmatrix} \tag{7}$$

of P, where i is an integer satisfying $i^2 \equiv -1 (\mathrm{mod}\ \pi)$ (note that the determinant of M_a is 1 and that the square root of p modulo π does exist). Let Q be the set of the $p + 1$ matrices previously defined, and denote by $G(p, \pi)$ the Cayley graph of P with respect to this set Q. Thus $G(p, \pi)$ is a $(p + 1)$-regular graph with $\pi(\pi^2 - 1)/2$ vertices. It is shown in [14] that $\lambda(G(p, \pi)) \leq 2\sqrt{p}$ for every π. This upper bound is obtained by applying results of Eichler and Igusa concerning the Ramanujan conjecture. Eichler's proof relies on Weil's famous theorem known as the Riemann hypothesis for curves over finite fields [26]. Therefore, for every

fixed π, the family $\{G(p,\pi)\}_p$ is an optimal set of pseudo-random graphs as it attains the bound (6).

Although the construction given in [14] and [16] is proved only for primes π, a similar argument [20] shows that the analogous graphs defined for powers of π have the same properties. If π is a prime congruent to 1 modulo 4, p is a quadratic residue modulo π, and l is an integer, denote by P_l the factor group of the group of all 2×2 matrices with determinant 1 over Z_{π^l}, modulo its normal subgroup consisting of the identity I and its (additive) inverse $-I$. It is not too difficult to check that P_l has $\frac{1}{2}(\pi^{3l} - \pi^{3l-2})$ elements.

The graph $G(p, \pi, l)$ is defined as the Cayley graph of P_l with respect to the $p + 1$ generators M_a given by (7), except that now the square root i of -1, and that of p, are taken modulo π^l. Note that since p is a quadratic residue in Z_π, it is also a quadratic residue in Z_{π^l} for every $l \geq 1$. Moreover, an easy (though somewhat tedious) computation shows that if $\alpha^2 = b\pi + p$ for some integers α and b (i.e., α is a square root of p modulo π), then a square root β of p modulo π^l is obtained by

$$\beta = \alpha - \sum_{j=1}^{l-1} d_j \pi^j, \tag{8}$$

where

$$d_j \equiv \frac{c_{j-1} b^j}{(2\alpha)^{2j-1}} (\mathrm{mod}\, \pi^l), \tag{9}$$

and c_j is the jth Catalan number given by

$$c_j = \frac{1}{j+1} \binom{2j}{j}. \tag{10}$$

(These numbers appear frequently in Combinatorics, and their generating function $c(x) = \sum_{j=0}^{\infty} c_j x^j$ satisfies the relation $c(x) = 1 + xc^2(x)$; see [19, p. 82].)

Equations (8)–(10) enable us to compute the required square roots of p and -1 modulo π^l (needed for the computation of M_a) from the easy calculations of these roots in Z_π. This implies that the graphs $G(p, \pi, l)$ can be generated very efficiently. As is the case for $l = 1$, it can be shown that $\lambda(G(p, \pi, l)) \leq 2\sqrt{p}$ for all admissible π and l, making these graphs suitable for constructing the codes C_{exp}.

III. Good Codes over Large Alphabets

We start by describing the details of Construction $\mathscr{C}_1$ of designed relative minimum distance $\delta < 1$ over an alphabet Σ, $|\Sigma| = q$. Let ρ be a power of a prime (say, $\rho = 2$) and δ_0 be a positive real number smaller than $\frac{1}{2}$. The values of ρ and δ_0 are assumed to be fixed i.e., independent of δ and q.

Let Δ be the smallest integer that satisfies the inequality

$$\Delta \geq \frac{4(1/\delta_0 - 1)}{1 - \delta} \tag{11}$$

and such that $\Delta - 1$ is a prime congruent to 1 modulo 4. The code $\mathscr{C}_1$ involves two encoding levels. The first one is an $[n, r_0 n, \delta_0 n]$ Justesen code $\mathscr{C}_{\mathrm{Jus}}$ over the field $\Phi = \mathrm{GF}(\rho^m)$, where the values of m and r_0 are given by

$$m = \left\lfloor \frac{\log_\rho q}{\Delta} \right\rfloor \tag{12}$$

and

$$r_0 = \frac{1}{2}\left(1 - \frac{\delta_0}{H_{\rho^m}^{-1}(\frac{1}{2})}\right). \tag{13}$$

Since $\lim_{z \to \infty} H_z(\delta_0) = \delta_0 < \frac{1}{2}$, for sufficiently large m we have $H_{\rho^m}(\delta_0) < \frac{1}{2}$, in which case $r_0 > 0$ in (13) (in fact, when $\delta_0 < H_2^{-1}(\frac{1}{2}) \approx 0.11$, $H_{\rho^m}(\delta_0) \leq H_2(\delta_0) < \frac{1}{2}$ for every $m \geq 1$). Hence, for sufficiently large m, the code C_{Jus} of the previous parameters is, indeed, realizable, with Wozencraft's ensemble as inner codes of rate $\frac{1}{2}$ and the outer Reed–Solomon code having rate $2r_0$ [9]. The constant γ_1 in (4) will be adjusted so that the right-hand side of (4) be nonpositive whenever m in (12) is too small to let C_{Jus} be realized. We also assume that the length of C_{Jus} takes the values $n = \frac{1}{2}(\pi^{3l} - \pi^{3l-2})$ for some fixed prime π and for arbitrarily large l. Note that such lengths can always be attained for sufficiently large l by properly choosing the length of the outer Reed–Solomon code (possibly with appending a small number of zero coordinates to C_{Jus}).

The codewords of C_{Jus} then undergo a second coding level by the expander code C_{exp}, which maps n-tuples over Φ into n-tuples over $\hat{\Sigma} \triangleq \Phi^\Delta \cong \mathrm{GF}(\rho^{m\Delta})$. Since the overall code $\mathscr{C}_1$ will not be linear over Σ (though it will be over Φ), we may as well assume that $\hat{\Sigma} \subseteq \Sigma$. Let $G_{\mathrm{exp}} = G(\Delta - 1, \pi, l)$ be a pseudo-random graph with $n = \frac{1}{2}(\pi^{3l} - \pi^{3l-2})$ vertices and degree Δ, as defined in Section II. For each vertex i, $1 \leq i \leq n$, in G_{exp}, let $l_1(i), l_2(i), \cdots, l_\Delta(i)$ denote the set of vertices in G_{exp} that are adjacent to i, indexed according to some prespecified ordering. The encoding rule of C_{exp} is defined as follows: every input vector $\boldsymbol{u} = [u_1, u_2 \cdots u_n] \in \Phi^n$ is mapped into a codeword $\boldsymbol{c} = [c_1, c_2 \cdots c_n] \in \hat{\Sigma}^n$ with $c_i \triangleq [u_{l_1(i)} u_{l_2(i)} \cdots u_{l_\Delta(i)}]$, $1 \leq i \leq n$.

The code C_{exp} can be summarized explicitly in the following manner: Using the notations of Section II, let $P_l = \{\pm A_1, \pm A_2, \cdots, \pm A_n\}$ denote the set (of size $n = \frac{1}{2}(\pi^{3l} - \pi^{3l-2})$) of all 2×2 matrix inverse pairs $\pm A_i$ with determinant 1 over Z_{π^l}, and associate the ith coordinate of $\boldsymbol{u}$ with the matrix inverse pair $\pm A_i$. Let $\boldsymbol{a}_j = [a_{j,0}, a_{j,1}, a_{j,2}, a_{j,3}]$, $1 \leq j \leq \Delta$, exhaust all possible vectors such that $a_{j,0}$ is an odd positive integer, $a_{j,1}, a_{j,2}, a_{j,3}$ are even integers, and $a_{j,0}^2 + a_{j,1}^2 + a_{j,2}^2 + a_{j,3}^2 = \Delta - 1$. The code C_{exp} then maps each coordinate u_i of $\boldsymbol{u}$ into a Δ-tuple $c_i = [u_{l_1(i)} u_{l_2(i)} \cdots u_{l_\Delta(i)}]$, where the indices $l_j(i)$ are defined by $\pm A_{l_j(i)} = \pm A_i M_{a_j}$, and the M_{a_j} are given by (7). Recall that both square roots, $i = \sqrt{-1}$ and $\sqrt{p}$, are taken modulo π^l and can be computed efficiently by (8)–(10). Furthermore, the only searches required to construct C_{exp} are those of finding the smallest Δ that satisfies (11), and then computing all admissible vectors $\boldsymbol{a}_j$; these searches, in turn, require time complexity that is polynomial in Δ. Note also that C_{exp} is an additive group over $\hat{\Sigma}$ and, therefore, the Hamming distance between any two codewords $\boldsymbol{c}_1, \boldsymbol{c}_2 \in C_{\mathrm{exp}}$ equals the Hamming weight of $\boldsymbol{c}_1 - \boldsymbol{c}_2$, measured over $\hat{\Sigma}$.

The resulting overall code $\mathscr{C}_1$ is, therefore, of length n and rate r_0/Δ over $\hat{\Sigma}$, which translates into rate $(r_0/\Delta) \cdot \log_q |\hat{\Sigma}|$ over Σ. Observe that C_{exp}, as a code over $\hat{\Sigma}$, or Φ, is quite a bad one, since it just replicates and shuffles the input coordinates. However, the input to C_{exp} is not arbitrary, but rather codewords of C_{Jus}, the

minimum distance of which is at least $\delta_0 n$. This accounts for the bound (4), which is restated in the next lemma.

Lemma 1: There exist constants $\gamma_0 > 0, \gamma_1$ and $\delta_{\min} < 1$ such that for every $\delta \geq \delta_{\min}$

$$R_{C_1}(\delta, q) \geq \gamma_0(1-\delta) - \frac{\gamma_1}{\log_2 q}. \tag{14}$$

Proof: Let $\boldsymbol{c}$ be a codeword of $C_{\exp}$ over $\hat{\Sigma}$, corresponding to a nonzero input vector $\boldsymbol{u} \in C_{\text{Jus}}$, and let B be the set of vertices of $G_{\exp}$ associated with the nonzero coordinates in $\boldsymbol{u}$. The number of vertices in $G_{\exp}$ that have at least one neighbor in B is exactly the Hamming weight of $\boldsymbol{c}$, measured over $\hat{\Sigma}$. Therefore, by Corollary 1, the minimum distance d of $\mathscr{C}_1$, which is also the minimum Hamming weight of any nonzero codeword of $\mathscr{C}_1$, readily satisfies

$$n - d \leq \frac{4(\Delta-1)(1-\delta_0)n}{\delta_0 \Delta^2} \leq \frac{4}{\Delta}\left(\frac{1}{\delta_0} - 1\right)n \overset{(11)}{\leq} (1-\delta)n.$$

Hence, the relative minimum distance of $\mathscr{C}_1$ is at least δ.

We now express the rate of $\mathscr{C}_1$ in terms of δ and q. Let m_0 be the smallest positive integer greater than 4 for which $H_{\rho^{m_0}}(\delta_0) < \frac{1}{2}$, and assume that $m \geq m_0$; in this case we have $r_0 > 0$ in (13). The rate of $\mathscr{C}_1$ is given by

$$R_{\mathscr{C}_1}(\delta, q) = \frac{r_0}{\Delta} \cdot \log_q |\hat{\Sigma}| = \frac{r_0}{\Delta} \cdot \frac{m\Delta}{\log_\rho q} \tag{15}$$

$$\overset{(12)}{>} \frac{r_0}{\Delta} \cdot \left(1 - \frac{\Delta}{\log_\rho q}\right)$$

$$\overset{(13)}{=} \left(\frac{1}{2} - \frac{\delta_0}{2H^{-1}_{\rho^m}\left(\frac{1}{2}\right)}\right)\left(\frac{1}{\Delta} - \frac{1}{\log_\rho q}\right). \tag{16}$$

Now, it is easy to verify that $H_{\rho^m}(x) \leq x + 1/m\log_2\rho$ and, hence, $2H^{-1}_{\rho^m}(\frac{1}{2}) \geq 1 - 2/m\log_2\rho$. Also, since $m \geq m_0 > 4$, we have

$$\frac{2}{m\log_2\rho} \leq \frac{4}{(m+1)\log_2\rho} \overset{(12)}{<} \frac{4\Delta}{\log_2 q} \overset{(12)}{\leq} \frac{4}{m\log_2\rho} \leq \frac{4}{5} < 1.$$

Therefore,

$$\frac{1}{2H^{-1}_{\rho^m}\left(\frac{1}{2}\right)} \leq \frac{1}{1 - (4\Delta/\log_2 q)} \leq 1 + O\left(\frac{\Delta}{\log_2 q}\right). \tag{17}$$

Substituting (17) into (16) we obtain

$$R_{\mathscr{C}_1}(\delta, q) \geq \left(\frac{1}{2} - \delta_0 - O\left(\frac{\Delta}{\log_2 q}\right)\right)\left(\frac{1}{\Delta} - \frac{1}{\log_\rho q}\right)$$

$$= \frac{\frac{1}{2} - \delta_0}{\Delta} - O\left(\frac{1}{\log_2 q}\right) + O\left(\frac{\Delta}{\log_2^2 q}\right),$$

where we have absorbed the constant multipliers which depend on δ_0 and ρ in the $O(\cdot)$ expressions. Therefore, in terms of Δ and q, $R_{\mathscr{C}_1}(\delta, q)$ satisfies

$$R_{\mathscr{C}_1}(\delta, q) \geq \frac{\frac{1}{2} - \delta_0}{\Delta} - O\left(\frac{1}{\log_2 q}\right). \tag{18}$$

Now, by the prime number theorem for arithmetic progressions [7, ch. 7], the smallest Δ for which (11) holds also satisfies

$$\frac{1}{\Delta} \geq \frac{1-\delta}{4(1/\delta_0 - 1)}\left(1 - \theta(\delta)\right), \tag{19}$$

where $\lim_{\delta \to 1}\theta(\delta) = 0$. Plugging (19) into (18) we obtain

$$R_{\mathscr{C}_1}(\delta, q) \geq \underbrace{\left(\frac{\frac{1}{2} - \delta_0}{4(1/\delta_0 - 1)}\right)}_{\text{constant}}\left(1 - \theta(\delta)\right) \cdot (1-\delta) - O\left(\frac{1}{\log_2 q}\right). \tag{20}$$

Define

$$\alpha_0 \triangleq \frac{\frac{1}{2} - \delta_0}{4(1/\delta_0 - 1)}. \tag{21}$$

Assuming that $m \geq m_0$, we conclude that for every constant $\gamma_0 > \alpha_0$ there exists a real number $\delta_{\min} < 1$ (which depends on γ_0 and δ_0) such that

$$R_{\mathscr{C}_1}(\delta, q) \geq \gamma_0(1-\delta) - O\left(\frac{1}{\log_2 q}\right)$$

whenever $\delta \geq \delta_{\min}$.

Finally, we consider the case $m < m_0$, which corresponds to $\Delta > (\log_\rho q)/m_0$. By (19) we have

$$\frac{1-\delta}{4(1/\delta_0 - 1)}\left(1 - \theta(\delta)\right) \leq \frac{1}{\Delta} < \frac{m_0\log_2\rho}{\log_2 q} = O\left(\frac{1}{\log_2 q}\right).$$

Therefore, we may choose γ_1 to be large enough so that the right-hand side of (14) be nonpositive whenever $m < m_0$.

Remark 1: Referring to the notations of the last proof, the maximum value of α_0 in (21) is attained at

$$\delta_0 = \delta_{\max} \triangleq 1 - \frac{1}{\sqrt{2}} \approx 0.29, \tag{22}$$

in which case

$$\alpha_0 = \alpha_{\max} \triangleq \frac{1}{24 + 16\sqrt{2}} \approx 0.021. \tag{23}$$

Remark 2: The term $\theta(\delta)$ in (20) is identically zero if δ is taken from the infinite sequence

$$\delta_p = 1 - \frac{4}{p+1}\left(\frac{1}{\delta_0} - 1\right),$$

where p ranges over all primes congruent to 1 modulo 4. In such cases we can, therefore, take $\gamma_0 = \alpha_0$. If, in addition, Σ is taken as $\mathrm{GF}(\rho^{m(p+1)})$, then (16) becomes

$$R_{\mathscr{C}_1}\left(\delta_p, \rho^{m(p+1)}\right) = \frac{1 - \delta_0 / H_{\rho^m}^{-1}\left(\frac{1}{2}\right)}{2(p+1)} = \left(\frac{1 - \delta_0 / H_{\rho^m}^{-1}\left(\frac{1}{2}\right)}{8(1/\delta_0 - 1)}\right) \cdot (1 - \delta_p) \triangleq \alpha_0(\rho, \delta_0, m)(1 - \delta_p).$$

Clearly, for $\delta_0 = \delta_{\max}$ we have

$$\lim_{m \to \infty} \alpha_0(\rho, \delta_{\max}, m) = \alpha_{\max}.$$

Furthermore, for every *finite* $m \geq 5$ we also have $\alpha_0(\rho, \delta_{\max}, m) > 0$.

Comparing (14) with (2), we first note that, due to the Singleton bound, $1 - H_q(x)$ is bounded from above by $1 - x$ and, therefore,

$$R_{\mathrm{Zyablov}}(\delta) \leq \max_{\mu \geq 0} (1 - \mu)(1 - \delta/\mu) \leq \left(1 - \sqrt{\delta}\right)^2.$$

This implies that for relative minimum distances in the range $((1 - \gamma_0)/(1 + \gamma_0))^2 < \delta < 1$, the function $\delta \mapsto \gamma_0(1 - \delta)$ lies strictly above the curve $\delta \mapsto R_{\mathrm{Zyablov}}(\delta)$. Hence, for values of δ close to 1, and for sufficiently large q, Construction $\mathscr{C}_1$ lies above the Zyablov bound.

Finally, as for the explicitness of Construction $\mathscr{C}_1$, we have already pointed out that the only required searches are those of finding the minimum Δ that satisfies (11), and then finding all expressions for $\Delta - 1$ of the form of sums of four integer squares. However, since Δ is proportional to $1/R_{\mathscr{C}_1}(\delta, q)$, all the above searches can be carried out in time complexity that is polynomial in the inverse of the code rate (rather than polynomial in the code length). Once having the additive factorization of $\Delta - 1$, we can write explicit expressions for the entries of the generator matrix of $\mathscr{C}_{\mathrm{exp}}$ over Φ.

We remark that finding a polynomial-time decoding algorithm for $\mathscr{C}_1$ for correcting up to $(\delta n - 1)/2$ errors remains still an open problem.

IV. Good Codes over Specific Alphabets

We now use Construction $\mathscr{C}_1$ as an outer code in a concatenation scheme, obtaining a new code family over any finite field $F = \mathrm{GF}(q)$. Referring to the notations of Section III, we fix δ_0 to some real positive number $< \frac{1}{2}$ (say, to $\delta_{\max}$ as in (22)). For any $\eta \in [0, 1)$ let $\Delta(\eta)$ denote the smallest integer satisfying

$$\Delta(\eta) \geq \frac{4(1/\delta_0 - 1)}{1 - \eta}$$

and such that $\Delta(\eta) - 1$ is a prime congruent to 1 modulo 4 (see (11)).

Construction $\mathscr{C}_2$ over $F = \mathrm{GF}(q)$ is obtained as follows. As an outer code, we take Construction $\mathscr{C}_1$ of length n and relative minimum distance η over the alphabet $\Sigma = (\mathrm{GF}(q^m))^{\Delta(\eta)} \cong F^{m\Delta(\eta)}$. The inner code will be taken as a linear code over $\mathrm{GF}(q)$ of rate r, dimension $m\Delta(\eta)$ and relative minimum distance μ. The overall code is, therefore, a linear code over $\mathrm{GF}(q)$ of rate $R = r \cdot R_{\mathscr{C}_1}(\eta, q^{m\Delta(\eta)})$, relative minimum distance $\delta = \mu \cdot \eta$, and length $N = (nm/r)\Delta(\eta)$.

Since n is arbitrarily large, we may take Wozencraft's ensemble as the inner code, in which case we have $r \geq 1 - H_q(\mu)$ and, therefore,

$$R_{\mathscr{C}_2}(\mu \cdot \eta) \geq \left(1 - H_q(\mu)\right) \cdot R_{\mathscr{C}_1}\left(\eta, q^{m\Delta(\eta)}\right). \tag{24}$$

Note that (24) holds also for *fixed* values of m, in which case the parameters of the inner codes do not tend to infinity as $n \to \infty$. Theoretically, this would enable us to choose *specific* inner codes instead of Wozencraft's ensemble; however, for the low rates we are interested in there are not any known specific constructions that are above the Gilbert–Varshamov bound. In that case, we might as well let m go to infinity, and (24) then becomes

$$R_{\mathscr{C}_2}(\delta) \geq \left(1 - H_q(\mu)\right) \cdot R_{\mathscr{C}_1}(\delta/\mu, \infty) \overset{(4)}{\geq} \gamma_0\left(1 - H_q(\mu)\right)\left(1 - \frac{\delta}{\mu}\right). \tag{25}$$

The bound (5) is obtained by maximizing the right-hand side of (25) with respect to μ in the range $\delta \leq \mu \leq 1 - 1/q$.

As for the value of the constant γ_0 in (25), we note that when δ is close enough to $1 - 1/q$, δ/μ must be close to 1. Hence, in the zero-rate neighborhood, γ_0 can be any constant greater than $\alpha_{\max}$ (as in (23)).

The multiplier γ_0 in (25) can be slightly improved if we replace the C_{Jus} component in Construction $\mathscr{C}_1$ by a linear code C_{RS} over $\Phi = \mathrm{GF}(q^m)$ that consists of a concatenation of two Reed–Solomon codes. The code C_{RS} was used as the outer code by Sugiyama *et al.* in [22], where it was also shown that for a prescribed relative minimum distance δ_0, the rate $R_{\mathrm{RS}}(\delta_0)$ and length $N_{\mathrm{RS}}(\delta_0)$ of C_{RS} satisfy

$$R_{\mathrm{RS}}(\delta_0) \geq \left(1 - \sqrt{\delta_0}\right)^2$$

and

$$N_{\mathrm{RS}}(\delta_0) \geq q^{mq^{m\sqrt{R_{\mathrm{RS}}(\delta_0)}}}.$$

Although C_{RS} is not asymptotically good over the (fixed) field Φ (in the sense that $N_{\mathrm{Rs}}(\delta_0)$ cannot take arbitrarily large values), $N_{\mathrm{RS}}(\cdot)$ is large enough to let the whole Wozencraft's ensemble be concatenated to our modified Construction $\mathscr{C}_1$ (the proof of this assertion follows along the lines of that in [22]). We can now substitute $r_0 = (1 - \sqrt{\delta_0})^2$ in (15) and repeat the derivations of Lemma 1, ending by replacing the expression for α_0 in (21) by

$$\alpha_0 = \frac{\left(1 - \sqrt{\delta_0}\right)^2}{4(1/\delta_0 - 1)}. \tag{26}$$

The maximum of (26) is attained at $\delta_{\max} = (\sqrt{5} - 1)/2 \approx 0.62$, and the corresponding value of α_0 is given by

$$\alpha_{\max} = \frac{1}{10\sqrt{5} + 22} \approx 0.023.$$

V. Application to t-Independent Sets

In this section, we show how Construction $\mathscr{C}_2$ can be applied to obtain small t-independent sets. To this end, we first show that the

frequency of occurrence of each element of GF(q) in every nonzero codeword in these codes approaches $1/q$ as the length of the code tends to infinity.

Let C be an $[n, k]$ instance of Construction $\mathscr{C}_2$ over GF(q) for a prescribed relative minimum distance $\delta \triangleq 1 - 1/q - \epsilon$, where $\epsilon > 0$. Denote by n' the length of the inner (Wozencraft's ensemble) code in C, and let $\boldsymbol{c} = [\boldsymbol{c}_1\, \boldsymbol{c}_2 \cdots \boldsymbol{c}_{n/n'}]$ be a nonzero codeword of C, where each $\boldsymbol{c}_i$ stands for an inner codeword. Let μ be the value that maximizes the right-hand side of (5); note that, since $\epsilon > 0$, μ is strictly greater than δ. As the typical minimum distance of the inner Wozencraft's ensemble codes is $n'\mu$, for sufficiently large n (and n'), all but a negligible fraction of the nonzero $\boldsymbol{c}_i$ have Hamming weight $\geq n'\delta$; that is, virtually all of the nonzero $\boldsymbol{c}_i$ contain at most $n'(1 - \delta)$ zeros. A similar argument implies that in all but a negligible fraction of the nonzero $\boldsymbol{c}_i$, *any* element of GF(q) appears at most $n'(1 - \delta)$ times. Therefore, when $n \to \infty$, the frequency of occurrence of each nonzero element of GF(q) in any nonzero codeword of C becomes at most $1 - \delta = 1/q + \epsilon$. Furthermore, since the relative minimum distance of C is δ, the same upper bound holds for the frequency of occurrence of the zero element as well. This, in turn, implies that the frequency of each element of GF(q) in any nonzero codeword of C must as well be at least $1/q - (q - 1)\epsilon$. In particular, when $q = 2$, for every $\epsilon' > \epsilon$, the nonzero weights in such a code C are confined to the range $n(\frac{1}{2} \pm \epsilon')$ for sufficiently large n.

From now on we concentrate on the binary case. For fixed t, let G be a $k \times n$ generator matrix of the above code C where we set $\epsilon = 2^{-t-1}$. Also, let H be a $k \times m$ parity-check matrix of an $[m, m - k, t + 1]$ linear code over $F = \mathrm{GF}(2)$. Since every t columns in H are linearly independent, for any nonzero vector $\boldsymbol{y} \in F^n$ of weight $\leq t$ we have $H\boldsymbol{y} \neq \boldsymbol{0}$. Now, define the $m \times n$ matrix $A \triangleq H^T G$, where $(\cdot)^T$ stands for transposition. Consider a $t \times n$ matrix B consisting of t arbitrary distinct rows of A. It is easy to see that for every $\boldsymbol{u} \in F^t - \{\boldsymbol{0}\}$ we have $\boldsymbol{u}B \in C - \{\boldsymbol{0}\}$. We now claim that every binary t-tuple appears as a column in B.

For every $\boldsymbol{x} \in F^t$, denote by $n_{\boldsymbol{x}}$ the number of occurrences of $\boldsymbol{x}$ as a column in B. It can be readily verified that for any $\boldsymbol{u} \in F^t$,

$$n - 2\mathrm{wt}(\boldsymbol{u}B) = \sum_{\boldsymbol{x} \in F^t} (-1)^{\langle \boldsymbol{u}, \boldsymbol{x} \rangle} n_{\boldsymbol{x}} \tag{27}$$

where $\mathrm{wt}(\boldsymbol{u}B)$ stands for the Hamming weight of $\boldsymbol{u}B$ and $\langle \cdot, \cdot \rangle$ denotes scalar product of vectors (over F). Let $\boldsymbol{s}$ be the vector whose entries are given by $s_{\boldsymbol{x}} = n_{\boldsymbol{x}}/n$, $\boldsymbol{x} \in F^t$; similarly, let $\boldsymbol{w}$ be the vector whose entries are $w_{\boldsymbol{u}} = \mathrm{wt}(\boldsymbol{u}B)/n$, $\boldsymbol{u} \in F^t$. By (27) we have

$$1 - 2\boldsymbol{w} = \mathscr{H}_t \boldsymbol{s},$$

where $\mathscr{H}_t$ stands for the Sylvester-type $2^t \times 2^t$ Hadamard matrix [15, ch. 2, Section 3], and $\boldsymbol{1}$ is the all-one vector. Noting that $\mathscr{H}_t^{-1} = 2^{-t}\mathscr{H}_t$, we obtain

$$2^t \boldsymbol{s} = \mathscr{H}_t(1 - 2\boldsymbol{w}).$$

Now, let ϵ' be in the range $2^{-t-1}(= \epsilon) < \epsilon' < 1/(2^{t+1} - 2)$. Since $\boldsymbol{u}B$ is a codeword of C, for sufficiently large n we have $|1 - 2w_{\boldsymbol{u}}| \leq 2\epsilon'$ for any $\boldsymbol{u} \neq \boldsymbol{0}$. Therefore, for every $\boldsymbol{x} \in F^t$,

$$\begin{aligned} 2^t \cdot s_{\boldsymbol{x}} &= \sum_{\boldsymbol{u} \in F^t} (1 - 2w_{\boldsymbol{u}})(-1)^{\langle \boldsymbol{u}, \boldsymbol{x} \rangle} \\ &= 1 + \sum_{\boldsymbol{u} \neq \boldsymbol{0}} (1 - 2w_{\boldsymbol{u}})(-1)^{\langle \boldsymbol{u}, \boldsymbol{x} \rangle} \\ &\geq 1 - \sum_{\boldsymbol{u} \neq \boldsymbol{0}} |1 - 2w_{\boldsymbol{u}}| \\ &\geq 1 - (2^t - 1) \cdot 2\epsilon' \\ &\geq 0. \end{aligned}$$

Hence, given t, for sufficiently large n we have $s_{\boldsymbol{x}} > 0$ for all $\boldsymbol{x} \in F^t$, thus proving that every vector $\boldsymbol{x} \in F^t$ appears as a column in B. Now, since B is an arbitrary $t \times n$ submatrix of A, the columns of the latter form a t-independent set over $\{0, 1\}^m$ of size n.

Now, set H as a parity-check matrix of a (possibly punctured) binary BCH code of length m and designed minimum distance $t + 1$. In this case we have $k \leq \lceil t/2 \rceil \cdot \lceil \log_2(m + 1) \rceil$. Also, since C is an instance of Construction $\mathscr{C}_2$,

$$n \leq c_1 \cdot k/\epsilon^3 = c_2 \cdot 2^{3t} k,$$

for some absolute constants c_1 and c_2, independent of t. Hence, there exists an absolute constant c such that for every fixed t and for sufficiently large m, the previous t-independent construction is of size $\leq ct2^{3t}\log m$, thus improving on previously-known constructions. For comparison we note that the best-known lower bound on the size of t-independent sets is $\Omega(2^t \log m)$, whereas counting arguments provide the non-constructive upper bound $O(t2^t \log m)$ [12], [21].

The previous construction method for t-independent sets is based upon the technique introduced in [17] for obtaining so-called ϵ-bias probability spaces. For the sake of completeness, however, we reformulated the derivation for the special case of t-independent sets. The improvement over [17] in the size of the resulting t-independent set are possible since code C used here is better than the one used in [17].

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi, "Deterministic simulation in LOGSPACE," *Proc. 19th ACM Symp. Theory of Comput.*, 1987, pp. 132–140.

[2] N. Alon, "Eigenvalues and expanders," *Combinatorica*, vol. 6, pp. 83–96, 1986.

[3] ——, "Eigenvalues, geometric expanders, sorting in rounds, and Ramsey theory," *Combinatorica*, vol. 6, pp. 207–219, 1986.

[4] ——, "Explicit construction of exponential sized families of k-independent sets," *Discrete Math.*, vol. 58, pp. 191–193, 1986.

[5] D. Le Brigand, "On computational complexity of some algebraic curves over finite fields," *Lecture Notes in Comput. Sci.*, vol. 229, pp. 223–227, 1986.

[6] A. Cohen and A. Wigderson, "Dispersers, deterministic amplification and weak random sources," in *Proc. 30th Symp. Found. Comput. Sci.*, 1989, pp. 14–19.

[7] H. Davenport, *Multiplicative Number Theory*, second ed., revised by H. L. Montgomery. Berlin: Springer Verlag, 1980.

[8] R. Impagliazzo and D. Zuckerman, "Recycling random bits," in *Proc. 30th Symp. Found. Comput. Sci.*, 1989, pp. 248–253.

[9] J. Justesen, "A class of constructive asymptotically good algebraic codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 652–656, Sept. 1972.

[10] G. L. Katsman, M. A. Tsfasman, and S. G. Vlădut, "Modular curves and codes with a polynomial construction," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 353–355, Mar. 1984.

[11] R. M. Karp and N. Pippenger, "A time randomness tradeoff," presented at *AMS Conf. Probabilistic Computat. in Complexity*. Durham, NC, 1983.

[12] D. J. Kleitman and J. Spencer, "Families of k-independent sets," *Discrete Math.*, vol. 6, pp. 255–262, 1973.

[13] S. N. Litsyn and M. A. Tsfasman, "A note on lower bounds," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 705–706, Sept. 1986.

[14] A. Lubotzky, R. Phillips, and P. Sarnak, "Explicit expanders and the Ramanujan conjectures," in *Proc. 18th ACM Symp. Theory of Comput.*, 1986, 240–246; See also: "Ramanujan graphs," *Combinatorica,* vol. 8, pp. 261–277, 1988.
[15] F. J. MacWilliams and N.J.A. Sloane, *The Theory of Error-Correcting Codes.* Amsterdam: North Holland, 1977.
[16] G. A. Margulis, "Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and superconcentrators," *Prob. Inform. Transm.*, vol. 24, pp. 39–46, 1988.
[17] J. Naor and M. Naor, "Small-bias probability spaces: efficient constructions and applications," in *Proc. 22nd ACM Symp. Theory of Comput.*, 1990, pp. 213–223.
[18] A. Nilli, "On the second eigenvalue of a graph," *Discrete Math.*, vol. 91, pp. 207–210, 1991.
[19] J. Riordan, *Combinatorial Identities.* New York: John Wiley, 1968.
[20] P. Sarnak, private communication.
[21] G. Seroussi and N. H. Bshouty, "Vector sets for exhaustive testing of logic circuits," *IEEE Trans. Inform. Theory,* vol. 34, pp. 513–522, May 1988.
[22] Y. Sugiyama, M. Kasahara, S. Hirasawa, and T. Namekawa, "A modification of the constructive asymptotically good codes of Justesen for low rates," *Inform. Contr.*, vol. 25, pp. 341–350, 1974.
[23] ——, "A new class of asymptotically good codes beyond the Zyablov bound," *IEEE Trans. Inform. Theory,* vol. IT-24, pp. 198–204, Mar. 1978.
[24] ——, "Superimposed concatenated codes," *IEEE Trans. Inform. Theory,* vol. IT-26, pp. 735–736, Nov. 1980.
[25] M. A. Tsfasman, S. G. Vlădut, and Th. Zink, "Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound," *Math. Nachr.*, vol. 109, pp. 21–28, 1982.
[26] A. Weil, "Sur les courbes algébriques et les variétés qui sèn déduisent," *Actualités Sci. Ind.*, p. 1041, 1948.
[27] E. J. Weldon, Jr., "Justesen's construction—The low-rate case," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 711–713, Sept. 1973.
[28] ——, "Some results on the problem of constructing asymptotically good error-correcting codes," *IEEE Trans. Inform. Theory,* vol. IT-21, pp. 412–417, July 1975.
[29] V. V. Zyablov, "An estimate of the complexity of constructing binary linear cascade codes," *Probl. Inform. Transm.*, vol. 7, pp. 3–10, 1971.

On Generator Matrices of Codes

Juriaan Simonis

Abstract—**The if class of the q-ary linear codes of given length, dimension and minimum weight is nonempty, it is shown to contain a code whose generator matrix consists of words of minimum weight.**

Index Terms—**Linear codes, minimum weight, generator matrix.**

Recently, Dodunekov [1] showed that any q-ary linear code of length n, dimension k, and minimum weight d possesses a generator matrix consisting of codewords of weight $\leq d + t$, where

$$t := n - \sum_{j=0}^{k-1} \left\lceil \frac{d}{q^j} \right\rceil,$$

a nonnegative integer in virtue of the Griesmer bound. A perhaps more accessible reference is Theorem 2.6 in Dodunekov and Manev's paper [2] which, however, is restricted to the binary case.

Manuscript received March 7, 1991.
The author is with the Faculty of Technical Mathematics and Informatics, Delft University of Technology, Mekelweg 4, P.O. Box 5031, 2600 GA Delft, The Netherlands.
IEEE Log Number 9104809.

The following theorem, in a way a strengthening of Dodunekov's result, may prove useful in nonexistence proofs for linear codes.

Theorem: Any linear code $\mathscr{C} \subset \mathbb{F}_q^n$ of dimension k and minimum weight d can be transformed into a code $\mathscr{D} \subset \mathbb{F}_q^n$ with the same parameters such that $\mathscr{D}$ possesses a basis of weight d vectors.

Proof: In the sequel, the linear subspace of $\mathbb{F}_q^n$ spanned by a set of vectors $\boldsymbol{x}, \boldsymbol{y}, \cdots, \boldsymbol{z} \in \mathbb{F}_q^n$ will be denoted by $\langle \boldsymbol{x}, \boldsymbol{y}, \cdots, \boldsymbol{z} \rangle$. Let $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t\} \subset \mathscr{C}$ be a maximal set of independent codewords of weight d. Suppose that $t < k$. All codewords in the complement of the span $\langle \boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t \rangle$ of the $\boldsymbol{a}_i$ have weight $> d$. Pick a codeword $\boldsymbol{b}_1 \notin \langle \boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t \rangle$ of lowest weight, say $\tilde{d}$, and extend $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t, \boldsymbol{b}_1\}$ to a basis $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t, \boldsymbol{b}_1, \cdots, \boldsymbol{b}_{k-t}\}$ of the code $\mathscr{C}$. Now change $\boldsymbol{b}_1$ into a vector $\boldsymbol{b}_1'$ of weight d by changing $\tilde{d} - d$ of the nonzero coordinates into zero ones. Then, linear subspace

$$\mathscr{C}' := \langle \boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t, \boldsymbol{b}_1', \cdots, \boldsymbol{b}_{k-t} \rangle \subset \mathbb{F}_q^n$$

is a code of minimum weight d, because $\langle \boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t \rangle$ is unaltered and the words of $\mathscr{C} \setminus \langle \boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t \rangle$ have changed in at most $\tilde{d} - d$ coordinates. We claim that the dimension of $\mathscr{C}'$ is equal to k. For if $\dim \mathscr{C}' < k$, then $\boldsymbol{b}_1'$ would be a linear combination of the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_{k-t}$, and, thus, would be an element of the original code $\mathscr{C}$. Since the weight of both $\boldsymbol{b}_1'$ and $\boldsymbol{b}_1 - \boldsymbol{b}_1'$ is smaller than $\tilde{d}$, these vectors would in fact be contained in the linear subspace $\langle \boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t \rangle$ which contradicts the fact that $\boldsymbol{b}_1 \in \mathscr{C} \setminus \langle \boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_t \rangle$. So $\mathscr{C}'$ has the same parameters as $\mathscr{C}$ has, but the maximum number of independent weight d codewords in $\mathscr{C}'$ exceeds that of $\mathscr{C}$. The induction process is obvious. □

Remark: As one of the referees observed, it may happen that the resulting code $\mathscr{D}$ is contained in a coordinate hyperplane of $\mathbb{F}_q^n$ and thus has effective length $< n$. An example is the code $\mathscr{C} \subset \mathbb{F}_q^3$ with generator matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.

References

[1] S. D. Dodunekov, "Zamechanie o vesovoy strukture porozhdayushchikh matrits lineĭnykh kodov," *Probl. peredach. inform.*, vol. 26, pp. 101–104, 1990.
[2] S. D. Dodunekov and N. L. Manev, "An improvement of the Griesmer bound for some small minimum distances," *Discrete Appl. Math.*, vol. 12, pp. 103–114, 1985.

On Binary Cyclic Codes of Odd Lengths from 101 to 127

Dieter Schomaker and Michael Wirtz

Abstract—**All binary cyclic codes of odd lengths are checked from 101 to 127 to find codes which are better than those in a table by Verhoeff. There are five such cases, namely, [117, 36, 32], [117, 37, 29], [117, 42, 26], [117, 49, 24], and [127, 36, 35] cyclic codes. According to Verhoeff's table the previously known ranges of the highest minimum-**

Manuscript received January 31, 1991; revised August 8, 1991.
The authors are with the Mathematisches Institut der Universität Münster, Einsteinstraße 62, D-4400 Münster, Germany.
IEEE Log Number 9105484.