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

NOGA ALON*† JEHOSHUA BRUCK* JOSEPH NAOR‡ MONI NAOR*
RON M. ROTH*§

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.

---

*IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, CA 95120.

†On sabbatical from the Department of Mathematics, Tel-Aviv University, Tel-Aviv 69978, Israel.

‡Computer Science Department, Stanford University, Stanford, CA 94305.

§On leave from the Computer Science Department, Technion – Israel Institute of Technology, Haifa 32000, Israel.

1. 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: (i) $\lim_{i\to\infty} n_i = \infty$; and (ii) both the rate of the sequence $R \triangleq \liminf_{i\to\infty} \frac{\log_q M_i}{n_i}$, and its relative minimum distance $\delta \triangleq \liminf_{i\to\infty} \frac{d_i}{n_i}$, are strictly greater than zero.

By the Gilbert-Varshamov bound, for any $\delta \in \left[0, 1 - \frac{1}{q}\right)$ 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 $H_q(x) \triangleq -x \cdot \log_q x - (1-x) \cdot \log_q(1-x) + x \cdot \log_q(q-1)$, $0 \leq x \leq 1 - \frac{1}{q}$. Furthermore, the seminal works of Tsfasman et al. [9, 10, 21] 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 $\mathcal{S} = \{S(\delta)\}_{0 \leq \delta \leq 1 - \frac{1}{q}}$ 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 (i) 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 (ii) $R(\delta) > 0$ whenever $\delta < 1 - \frac{1}{q}$. Note that $R_{\mathrm{GV}}(\delta) > 0$ for $\delta < 1 - \frac{1}{q}$, whereas by the Plotkin bound we must have $R(\delta) = 0$ for $\delta \geq 1 - \frac{1}{q}$. Such a uniformity definition is aimed to characterize good low-rate code sequences which 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

$R(\delta) \geq R_{\text{Zyablov}}(\delta)$ [25], where

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

However, searching for the inner code by any known algorithm requires time complexity which is exponential in the inner code length. 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 $\lim_{\delta \to 1-\frac{1}{q}} c(\delta) = \infty$. Hence, such a code sequence family is non-uniformly constructive.

The exponential search is avoided in Justesen codes [8] and in constructions derived thereof [18, 19, 20, 23], 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 $R_{\text{Jus}}(\delta)$, 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 $1 - \frac{1}{q}$. Therefore, Justesen codes do not comply with requirement (ii) of uniform constructiveness. The same holds also for some other known improvements on Justesen codes [19, 24].

Uniformly constructive families of codes over $GF(q)$ were obtained by Weldon [23] and Sugiyama et al. [18, 20], 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 $R_{\text{SKHN}}(\delta)$ of the construction obtained in [20] satisfies

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

Katsman, Tsfasman and Vlădut [9] 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} [4], 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 $1 - \frac{1}{q}$), to the best knowledge of the authors.

In this paper 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 which 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 2.

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*) $C_{\exp} : \Phi^n \to \left(\Phi^\Delta\right)^n$, 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, 5, 7, 14].

Specifiaclly, [14] consider the notion of ϵ-bias probability spaces which are...

These expander codes will serve as building blocks in our new asymptotically good constructions. The first construction, referred to as Construction $\mathcal{C}_1$, 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 3, the rate $R_{\mathcal{C}_1}(\delta, q)$ of Construction $\mathcal{C}_1$ satisfies

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

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 [9]), except for the multiplier γ_0 (which is approximately 0.021).

Construction $\mathcal{C}_1$ satisfies criterion (i) of the uniformity definition. As for criterion (ii), the δ-interval for which $R_{\mathcal{C}_1}(\delta, q) = 0$ shrinks to zero length when $q \to \infty$; hence, $\mathcal{C}_1$ is 'nearly-uniformly' constructive, and this fact will be exploited in our second construction. However, the significance of Construction $\mathcal{C}_1$ 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 $\mathcal{C}_1$ fails to improve on previously-known constructions. However, we can use Construction $\mathcal{C}_1$ to introduce good code sequences over specific fields $F = GF(q)$ by means of concatenation. The new codes will be referred to as Construction $\mathcal{C}_2$ and will be discussed in Section 4. Construction $\mathcal{C}_2$ is obtained by using Construction $\mathcal{C}_1$ over $\Sigma = \left(GF(q^m)\right)^\Delta$ 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 which satisfies the inequality

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

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 $1 - \frac{1}{q}$, 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 $\mathcal{C}_2$ can be better illustrated if we express the rate R in terms of $\epsilon \triangleq 1 - \frac{1}{q} - \delta$. 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

$$R_{\mathcal{C}_2}(\tfrac{1}{2} - \epsilon) \;\geq\; \tfrac{16}{27 \ln 2}\, \gamma_0\, \epsilon^3 - O(\epsilon^4) .$$

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 $\mathcal{C}_2$ 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

$$R_{\mathrm{GV}}(\tfrac{1}{2} - \epsilon) \;=\; \tfrac{2}{\ln 2}\, \epsilon^2 - O(\epsilon^4) ,$$

whereas the McEliece-Rodemich-Rumsey-Welch upper bound [12, p. 559] yields

$$R_{\mathrm{MRRW}}(\tfrac{1}{2} - \epsilon) \;=\; 2\, \epsilon^2 \log_2(1/\epsilon) + O(\epsilon^2) .$$

2. 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 $\lambda_1, \ldots, \lambda_n$ be all the eigenvalues of G, (with multiplicities), where $\Delta = |\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_n|$. 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,*

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

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. The above 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 $\sum_{v \in V} f(v) = 0$ i.e., f is orthogonal to the eigenvector of the largest eigenvalue of A. Therefore

$$(Af, Af) \leq \lambda^2 (f, f)$$

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

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

The desired result follows. $\square$

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*

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

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

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

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

$$\begin{aligned} tb^2\Delta^2 &= \sum_{v\in T}\Big(|N_B(v)| - b\Delta\Big)^2 \\ &\le \sum_{v\in V}\Big(|N_B(v)| - b\Delta\Big)^2 \;\le\; \lambda^2 b(1-b)n \, . \end{aligned}$$

This completes the proof. $\square$

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, 15] that the second largest eigenvalue of any Δ-regular graph with diameter k is at least $2\sqrt{\Delta-1}\Big(1 - O(1/k)\Big)$. Therefore, in any infinite family of Δ-regular graphs $\{G_i = (V_i, E_i)\}_{|V_i|\to\infty}$,

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

Lubotzky, Phillips and Sarnak [11], and independently, Margulis [13], 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 $\lambda(G_i) \le 2\sqrt{\Delta-1}$. 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 which 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 which are not divisible by 4. This easily implies that there are precisely $p+1$ vectors $\mathbf{a} = [a_0, a_1, a_2, a_3]$, 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 $\mathbf{a}$ with the member

$$M_{\mathbf{a}} \triangleq \pm \frac{1}{\sqrt{p}} \begin{bmatrix} a_0 + \imath a_1 & a_2 + \imath a_3 \\ -a_2 + \imath a_3 & a_0 - \imath a_1 \end{bmatrix} \tag{7}$$

of P, where $\imath$ is an integer satisfying $i^2 \equiv -1 \pmod{\pi}$ (note that the determinant of $M_{\mathbf{a}}$ is 1 and that the square root of p modulo π does exist). Let Q be the set of the $p+1$ matrices defined above, 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 [11] 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 [22]. 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 [11] and [13] is proved only for primes π, a similar argument [17] 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 $M_{\mathbf{a}}$ given by (7), except that now the square root $\imath$ 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

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

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 [16, 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 $M_{\mathbf{a}}$) 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}.

3. Good codes over large alphabets

We start by describing the details of Construction $\mathcal{C}_1$ 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 which 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 $\mathcal{C}_1$ involves two encoding levels. The first one is an $[n, r_0 n, \delta_0 n]$ Justesen code C_{Jus} over the field $\Phi = GF(\rho^m)$, 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 $\delta_0 < H_2^{-1}(\frac{1}{2}) \approx 0.11$, $H_{\rho^m}(\delta_0) \le H_2(\delta_0) < \frac{1}{2}$ for every $m \ge 1$). Hence, for sufficiently large m, the code C_{Jus} of the above parameters is, indeed, realizable, with Wozencraft's ensemble as inner codes of rate $\frac{1}{2}$ and the outer Reed-Solomon code having rate $2r_0$ [8]. The constant γ_1 in (4) will be adjusted so that the right-hand side of (4) be non-positive 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 $\hat{\Sigma} \triangleq \Phi^\Delta \cong GF(\rho^{m\Delta})$. Since the overall code $\mathcal{C}_1$ will not be linear over Σ (though it will be over Φ), we may as well assume that $\hat{\Sigma} \subseteq \Sigma$. Let $G_{\text{exp}} = G(\Delta - 1, \pi, l)$ be a pseudo-random graph with $n = \frac{1}{2}(\pi^{3l} - \pi^{3l-2})$ vertices and degree Δ, as defined in Section 2. For each vertex i, $1 \le i \le n$, in G_{exp}, let $\ell_1(i), \ell_2(i), \ldots, \ell_\Delta(i)$ denote the set of vertices in G_{exp} which are adjacent to i, indexed according to some pre-specified ordering. The encoding rule of C_{exp} is defined as follows: every input vector $\mathbf{u} = [u_1\, u_2\, \ldots\, u_n] \in \Phi^n$ is mapped into a codeword $\mathbf{c} = [c_1\, c_2\, \ldots\, c_n] \in \hat{\Sigma}^n$, with $c_i \triangleq [u_{\ell_1(i)}\, u_{\ell_2(i)}\, \ldots\, u_{\ell_\Delta(i)}]$, $1 \le i \le n$. Note that C_{exp} is an additive group over $\hat{\Sigma}$ and, therefore, the Hamming distance between any two codewords $\mathbf{c}_1, \mathbf{c}_2 \in C_{\text{exp}}$ equals the Hamming weight of $\mathbf{c}_1 - \mathbf{c}_2$, measured over $\hat{\Sigma}$.

The resulting overall code $\mathcal{C}_1$ 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 re-stated in the next lemma.

Lemma 1. *There exist constants* $\gamma_0 > 0$, γ_1 *and* $\delta_{\min} < 1$ *such that for every* $\delta \geq \delta_{\min}$

$$R_{\mathcal{C}_1}(\delta, q) \;\geq\; \gamma_0(1-\delta) \;-\; \frac{\gamma_1}{\log_2 q}\;. \tag{14}$$

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

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

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

We now express the rate of $\mathcal{C}_1$ 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 $\mathcal{C}_1$ is given by

$$\begin{aligned} R_{\mathcal{C}_1}(\delta, q) \;&=\; \frac{r_0}{\Delta}\cdot \log_q |\hat{\Sigma}| \;=\; \frac{r_0}{\Delta}\cdot\frac{m\Delta}{\log_\rho q} && (15)\\ &\stackrel{(12)}{>}\; \frac{r_0}{\Delta}\cdot\left(1 - \frac{\Delta}{\log_\rho q}\right) \\ &\stackrel{(13)}{=}\; \left(\frac{1}{2} - \frac{\delta_0}{2H^{-1}_{\rho^m}(\frac{1}{2})}\right)\left(\frac{1}{\Delta} - \frac{1}{\log_\rho q}\right)\;. && (16)\end{aligned}$$

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

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

Therefore,

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

Substituting (17) into (16) we obtain

$$\begin{aligned} R_{\mathcal{C}_1}(\delta, q) \;&\geq\; \left(\frac{1}{2} - \delta_0 - O\left(\frac{\Delta}{\log_2 q}\right)\right)\left(\frac{1}{\Delta} - \frac{1}{\log_\rho q}\right) \\ &=\; \frac{\frac{1}{2}-\delta_0}{\Delta} \;-\; O\left(\frac{1}{\log_2 q}\right) + O\left(\frac{\Delta}{\log_2^2 q}\right)\;,\end{aligned}$$

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, $R_{\mathcal{C}_1}(\delta, q)$ satisfies

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

Now, by the Prime Number Theorem for arithmetic progressions [6, Ch. 7], the smallest Δ for which (11) holds also satisfies

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

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

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

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

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

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

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

Therefore, we may choose γ_1 to be large enough so that the right-hand side of (14) be non-positive 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 $GF(\rho^{m(p+1)})$, then (16) becomes

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

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,

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

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

Finally, as for the explicitness of Construction $\mathcal{C}_1$, we note that the only required searches are those of finding the minimum Δ which satisfies (19), and then finding all expressions for $\Delta - 1$ of the form of sums of four integer squares. However, since Δ is proportional to $1/R_{\mathcal{C}_1}(\delta, q)$, all the above searches can be carried out in time complexity which 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 C_{exp} over Φ.

4. Good codes over specific alphabets

We now use Construction $\mathcal{C}_1$ as an outer code in a concatenation scheme, obtaining a new code family over any finite field $F = GF(q)$. Referring to the notations of Section 3, 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 $\Delta - 1$ is a prime congruent to 1 modulo 4 (see (11)).

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

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

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 aren't any known specific constructions which are above the Gilbert-Varshamov bound. In that case, we might as well let m go to infinity, and (24) then becomes

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

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

As for the value of the constant γ_0 in (25), we note that when δ is close enough to $1 - \frac{1}{q}$, 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 $\mathcal{C}_1$ by a linear code C_{RS} over $\Phi = GF(q^m)$ which consists of a concatenation of two Reed-Solomon codes. The code C_{RS} was used as the outer code by Sugiyama et al.

in [18], 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 $\mathcal{C}_1$ (the proof of this assertion follows along the lines of that in [18]). We can now substitute $r_0 = \left(1 - \sqrt{\delta_0}\right)^2$ 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 $\delta_{\max} = \frac{\sqrt{5}-1}{2} \approx 0.62$, and the corresponding value of α_0 is given by

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

References

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

[2] N. Alon, *Eigenvalues and expanders, Combinatorica,* 6 (1986), 83–96.

[3] N. Alon, *Eigenvalues, geometric expanders, sorting in rounds, and Ramsey Theory, Combinatorica* 6 (1986), 207–219.

[4] D. Le Brigand, *On computational complexity of some algebraic curves over finite fields, Lecture Notes in Computer Science,* 229 (1986), 223–227.

[5] A. Cohen, A. Wigderson, *Dispersers, deterministic amplification and weak random sources, Proc. 30-th Symp. Found. of Comp. Science* (1989), 14–19.

[6] H. Davenport, *Multiplicative Number Theory,* Second Edition, revised by H.L. Montogomery, Springer Verlag, Berlin, 1980.

[7] R. Impagliazzo, D. Zuckerman, *Recycling random bits, Proc. 30-th Symp. Found. of Comp. Science* (1989), 248–253.

[8] J. Justesen, *A class of constructive asymptotically good algebraic codes, IEEE Trans. Inform. Theory,* IT-18 (1972), 652–656.

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

[10] S.N. Litsyn, M.A. Tsfasman, *A note on lower bounds, IEEE Trans. Inform. Theory,* IT-32 (1986), 705–706.

[11] A. Lubotzky, R. Phillips and P. Sarnak, *Explicit expanders and the Ramanujan conjectures, Proc. 18-th ACM Symp. Theory of Comput.* (1986), 240–246.
See also: A. Lubotzky, R. Phillips and P. Sarnak, *Ramanujan graphs, Combinatorica,* 8 (1988), 261–277.

[12] F.J. MacWilliams, N.J.A. Sloane, *The Theory of Error-Correcting Codes,* North Holland, Amsterdam, 1977.

[13] G.A. Margulis, *Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and superconcentrators, Prob. Inform. Trans.,* 24 (1988), 39–46.

[14] J. Naor, M. Naor, *Small-bias probability spaces: efficient constructions and applications, Proc. 22-nd ACM Symp. Theory of Comput.* (1990), 213–223.

[15] A. Nilli, *On the second eigenvalue of a graph, Discrete Math.,* to appear.

[16] J. Riordan, *Combinatorial Identities,* John Wiley, New York, 1968.

[17] P. Sarnak, Private communication.

[18] Y. Sugiyama, M. Kasahara, S. Hirasawa, T. Namekawa, *A modification of the constructive asymptotically good codes of Justesen for low rates, Inform. Control,* 25 (1974), 341–350.

[19] Y. Sugiyama, M. Kasahara, S. Hirasawa, T. Namekawa, *A new class of asymptotically good codes beyond the Zyablov bound, IEEE Trans. Inform. Theory,* IT-24 (1978), 198–204.

[20] Y. Sugiyama, M. Kasahara, S. Hirasawa, T. Namekawa, *Superimposed concatenated codes, IEEE Trans. Inform. Theory,* IT-26 (1980), 735–736.

[21] M.A. Tsfasman, S.G. Vlădut, Th. Zink, *Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound, Math. Nachr.,* 109 (1982), 21–28.

[22] A. Weil, *Sur les courbes algébriques et les variéstés qui sèn déduisent*, *Actualités Sci. Ind.,* 1041 (1948).

[23] E.J. Weldon, Jr., *Justesen's construction — the low-rate case, IEEE Trans. Inform. Theory,* IT-19 (1973), 711–713.

[24] E.J. Weldon, Jr., *Some results on the problem of constructing asymptotically good error-correcting codes, IEEE Trans. Inform. Theory,* IT-21 (1975), 412–417.

[25] V.V. Zyablov, *An estimate of the complexity of constructing binary linear cascade codes, Probl. Inform. Trans.,* 7 (1971), 3–10.