IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 38, NO. 2, MARCH 1992

REFERENCES


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 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 lies above the Zyablov bound. Using these codes as outer codes in a concatenated scheme, a second asymptotic good construction is obtained, which lies above the Zyablov bound.

I. INTRODUCTION

An infinite sequence of codes \( S = \{C_i\}_i \) over an alphabet \( \Sigma \) 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 \( R \) and the relative minimum distance \( \delta \) are strictly greater than zero.
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(qm), at the expense of not attaining the Singleton bound. The rate $R_{SKHN}(\delta)$ of the construction obtained in [24] satisfies

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

(3)

Katsman, Tsfasman, and Vladut [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 $n^3$ [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 $\delta$ 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 $\Delta$-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 a graph with $n = |V|$ vertices and a finite field $\mathbb{F}$, we then show how to define a so-called expander mapping (or expander code) $C_{exp}: \mathbb{F}^n \to (\mathbb{F}^\Delta)^n$, such that every input $n$-tuple over $\mathbb{F}$ of Hamming weight $\geq \delta_0$ is mapped into an output $n$-tuple over $\mathbb{F}^\Delta$ whose Hamming weight (measured over $\mathbb{F}^\Delta$) 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 [11], [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 $C_1$, is obtained by taking the codewords of any good code sequence over a finite field $\mathbb{F}$ (say, Justesen codes), and then applying the expander code $C_{exp}$, resulting in a code over the alphabet $\mathbb{F}^\Delta$ whose rate is proportional to $1/\Delta$. The choice of $\Delta$ and the field $\mathbb{F}$ will depend on the prescribed size $q$ of the underlying alphabet and the relative minimum distance $\delta$. As we show in Section III, the rate $R_{C_1}(\delta, q)$ of Construction $C_1$ satisfies

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

(4)

for some positive constants $\gamma_0$ and $\gamma_1$. 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 $\gamma_0$ (which is approximately 0.021).

Construction $C_1$ satisfies criterion 1) of the uniformity definition. As for criterion 2), the $\delta$-interval for which $R_{C_1}(\delta, q) = 0$ shrinks to zero length when $q \to \infty$; hence, $C_1$ is "nearly-uniformly" constructive, and this fact will be exploited in our second construction. However, the significance of Construction $C_1$ 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 $C_1$ fails to improve on previously-known constructions. However, we can use Construction $C_1$ to introduce good code sequences over specific fields $F = GF(q)$ by means of concatenation. The new codes will be referred to as Construction $C_2$ and will be discussed in Section IV. Construction $C_2$ is obtained by using Construction $C_1$ over $\Sigma = (GF(q^m))^n$ as the outer code, with each output symbol (over $\Sigma$) 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

$$R_{C_2}(\delta) \geq \max_{\delta \leq \mu \leq 1 - \frac{1}{q}} \gamma_0 (1 - H_q(\mu)) \left(1 - \frac{\delta}{\mu}\right).$$

(5)

The bound (5) resembles the Zyablov bound (2), except for the multiplier $\gamma_0$, due to which (5) lies beneath the curve (2). However, when the relative minimum distance $\delta$ 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 $R \lesssim 2.5 \times 10^{-6}$.

The significance of Construction $C_2$ can be better illustrated if we express the rate $R$ in terms of $\epsilon = 1 - 1/q - \delta$. We take the binary case as a typical (and the most important) example. In this case, $\epsilon = 1 - \delta$, and, when $\epsilon$ is small, (5) becomes

$$R_{C_2}(\frac{1}{2} - \epsilon) \geq \frac{16}{27} \ln 2 \gamma_0 \epsilon^3 - O(\epsilon^5).$$

The same bound is obtained by (2) if we replace $\gamma_0$ by 1. Hence, the attainable rates in both the Zyablov bound and Construction $C_2$ are of the same order i.e., proportional to $\epsilon^3$. Repeating the calculation for (3), however, yields a lower bound which is proportional to $\epsilon^5$. For comparison, it is worthwhile noting that, in terms of $\epsilon$, the Gilbert-Varshamov bound for $q = 2$ takes the form

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

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

$$R_{MRW}(\frac{1}{2} - \epsilon) = 2\epsilon^2 \log_2 (1/\epsilon) + O(\epsilon^2).$$

Like in previous constructions [9], [22], [27], the inner code in Construction $C_2$ 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 $C_2$ approaches $1/q$ as $\delta \to 1 - 1/q$ (and the code length 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 $\{0, 1\}^n$ 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 \( c2^{3\lambda}\log m \) 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 \( \Delta \)-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 \( \Delta \)-regular the largest eigenvalue of \( A \) is \( \Delta \), 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 \( \lambda(G) = |\lambda_1| \). As we show below, if \( \lambda(G) \) is much smaller than \( \Delta \), then \( G \) has a strong pseudo-random property.

**Theorem I:** Let \( G = (V, E) \) be a \( \Delta \)-regular graph with \( n = |V| \) and \( \lambda = \lambda(G) \). For a vertex \( v \in V \) and a subset \( B \) of \( V \) such that \( v \in B \), let \( N_B(v) \) denote 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 \( \Delta \)-regular graph each vertex \( v \) would tend to have about \( b\Delta \) neighbors in each set of size \( bn \). This theorem shows that if \( \lambda \) is much smaller than \( \Delta \) 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 \( f: V \rightarrow \mathbb{R} \) 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,

\[
(\langle A f, A f \rangle) \leq \lambda^2(f, f)
\]

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

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

The desired result follows.

**Corollary I:** Let \( G = (V, E) \) be a \( \Delta \)-regular graph with \( n = |V| \) and \( \lambda = \lambda(G) \), and let \( B \) be a subset of cardinality \( bn \) of \( V \). Let \( t = \left| \{v \in V : N_B(v) = \emptyset \} \right| \) 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:

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

This completes the proof.

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

\[
\limsup \chi(G_i) \geq 2\sqrt{\Delta - 1}.
\]

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 \( \Delta \)-regular graphs \( G \), 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 \( \pi \) be unequal primes, both congruent to 1 modulo 4, such that \( p \) is a quadratic residue modulo \( \pi \). Let \( P = \text{PSL}(2, Z_\pi) \) denote the factor group of the group of all \( 2 \times 2 \) matrices over \( Z_\pi \) with determinant 1 modulo its normal subgroup consisting of the identity \( I \) and its (additive) inverse \(-I\). The elements of \( P \) are thus simply \( 2 \times 2 \) matrices over \( Z_\pi \) 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 \( a = [a_0, a_1, a_2, a_3] \) where \( a_0 \) is an odd positive integer, \( a_1, a_2, a_3 \) are even integers, and \( a_0^2 + a_1^2 + a_2^2 + a_3^2 = p \). Associated each such vector \( a \) with the member

\[
M_a \triangleq \pm \frac{1}{\sqrt{p}} \begin{bmatrix}
  a_0 + ia_1 & a_2 + ia_3 \\
  -a_2 + ia_3 & a_0 - ia_1
\end{bmatrix}
\]

of \( P \), where \( i \) is an integer satisfying \( i^2 = -1 \mod \pi \) (note that the determinant of \( M_a \) is 1 and that the square root of \( p \) modulo \( \pi \) 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 \( \chi(G(p, \pi)) \leq 2\sqrt{p} \) for every \( \pi \). 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

\[
\end{document}
fixed $\pi$, 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 $\pi$, a similar argument [20] shows that the analogous graphs defined for powers of $\pi$ have the same properties. If $\pi$ is a prime congruent to 1 modulo 4, $p$ is a quadratic residue modulo $\pi$, and $l$ is an integer, denote by $P_l$ the factor group of the group of all $2 \times 2$ matrices with determinant 1 over $Z_q$, 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 $\frac{1}{2}(\pi^{1/2} - \pi^{-1/2})$ elements.

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

$$\beta = a - \sum_{j=1}^{l-1} d_j \pi^j,$$

where

$$d_j = \frac{c_j \cdot b^j}{(2\alpha)^{\frac{j}{2}-1} \pmod{\pi}},$$

and $c_j$ is the $j$th Catalan number given by

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

(These numbers appear frequently in Combinatorics, and their generating function $c(x) = \sum_{n=0}^{\infty} c_n x^n$ satisfies the relation $c(x) = 1 + xe^{c(x)}$; see [19, p. 821].)

Equations (8)-(10) enable us to compute the required square roots of $p$ and $-1$ modulo $\pi$ (needed for the computation of $M_q$) from the easy calculations of these roots in $Z_q$. This implies that the graphs $G(p, \pi, l)$ can be generated very efficiently. As is the case from the easy calculations of these roots in $Z_q$, the codes $C_\exp$ 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 $l$ and $q$, making these graphs suitable for constructing the codes $C_\exp$.

### III. GOOD CODES OVER LARGE ALPHABETS

We start by describing the details of Construction $\mathcal{E}_t$ of designed relative minimum distance $\delta < 1$ over an alphabet $\Sigma$, $|\Sigma| = q$. Let $\rho$ be a power of a prime (say, $\rho = 2$) and $\delta_0$ be a positive real number smaller than $\frac{1}{2}$. The values of $\rho$ and $\delta_0$ are assumed to be fixed, i.e., independent of $\delta$ and $q$.

Let $\Delta$ be the smallest integer that satisfies the inequality

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

and such that $\Delta - 1$ is a prime congruent to 1 modulo 4. The code $\mathcal{E}_t$ involves two encoding levels. The first one is an $(\delta, r_0, \delta_0, n)$ Justesen code $\mathcal{E}_{\text{just}}$ over the field $\Phi = GF(\rho^n)$, where the values of $m$ and $r_0$ are given by

$$m = \left\lceil \frac{\log_\rho q}{\Delta} \right\rceil$$

and

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

Since $\lim_{\rho \to \infty} H_\rho(\delta_0) = \delta_0 < \frac{1}{2}$, for sufficiently large $m$ we have $H_\rho(\delta_0) < \frac{1}{2}$, in which case $r_0 > 0$ in (13) (in fact, when $\delta_0 < H_\rho(\frac{1}{2}) = 0.11$, $H_\rho(\delta_0) \leq H_\rho(\delta_0) < \frac{1}{2}$ for every $m \geq 1$). Hence, for sufficiently large $m$, the code $C_\exp$ 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 $\gamma_1$ in (4) will be adjusted so that the right-hand side of (4) be nonpositive whenever $m$ in (12) is too small to let $C_\text{just}$ be realizable. We also assume that the length of $C_\text{just}$ takes the values $n = \frac{1}{2}(\pi^{1/2} - \pi^{-1/2})$ for some fixed prime $\pi$ and for arbitrarily large $n$. Note that such lengths can always be attained for sufficiently large $m$ by properly choosing the length of the outer Reed-Solomon code (possibly with appending a small number of zero coordinates to $C_\text{just}$).

The codewords of $C_\text{just}$ then undergo a second coding level by the expander code $C_\exp$, which maps $n$-tuples over $\Phi$ into $n$-tuples over $\Sigma \equiv \Phi^k \equiv GF(\rho^n)$. Since the overall code $\mathcal{E}_t$ will not be linear over $\Sigma$ (though it will be over $\Phi$), we may as well assume that $\Sigma \subseteq \Phi$. Let $G_\exp = G(\Delta - 1, \pi, \delta)$ be a pseudo-random graph with $\pi = \frac{1}{2}(\pi^{1/2} - \pi^{-1/2})$ vertices and degree $\Delta$, as defined in Section II. For each vertex $i$, $1 \leq i \leq n$, in $G_\exp$, let $I(i), I(i)'$, $\cdots$, $I(i)''$ denote the set of vertices in $G_\exp$ that are adjacent to $i$, indexed according to some preselected ordering. The encoding rule of $C_\exp$ is defined as follows: every input vector $u = [u_1, u_2, \cdots, u_n]$ in $\Phi^n$ is mapped into a codeword $c = [c_1, c_2, \cdots, c_n]$ in $\Sigma^n$ with $c_i \equiv [u_{i1}, u_{i1}(I(i))', u_{i1}(I(i))'', \cdots, u_{i1}(I(i))''']$, $1 \leq i \leq n$.

The code $C_\exp$ can be summarized explicitly in the following manner: Using the notations of Section II, let $P_l = \{ \pm A_1, \pm A_2, \cdots, \pm A_\delta \}$ denote the set (of size $n = \frac{1}{2}(\pi^{1/2} - \pi^{-1/2})$) of all $2 \times 2$ matrix inverse pairs $\pm A_\delta$ with determinant 1 over $Z_q$, and associate the $i$th coordinate of $u$ with the matrix inverse pair $\pm A_j$. Let $a_{l,j} = [a_{l,j}, a_{j,l}, a_{j,l}']$, $1 \leq j \leq \Delta$, exhaust all possible vectors such that $a_{l,j}$ is an odd positive integer, $a_{j,l}, a_{j,l}'$, $a_{j,l}''$ are even integers, and $a_{j,l} + a_{j,l}' + a_{j,l}'' = \Delta - 1$. The code $C_\exp$ then maps each coordinate $u_i$ of $u$ into a $\Delta$-tuple $c_i = [u_{i1}(I(i))', u_{i1}(I(i))'', \cdots, u_{i1}(I(i))''']$, where the indices $I(i)$ are defined by $\pi \Delta A_{l,j}$, $\pi A_{l,j}$, $\pi A_{l,j}'$, and the $M_{l,j}$ are given by (7). Recall that both square roots, $l = \sqrt{-1}$ and $\sqrt{p}$, are taken modulo $\pi'$ and can be computed efficiently by (8)-(10). Furthermore, the only searches required to construct $C_\exp$ are those of finding the smallest $\Delta$ that satisfies (11), and then computing all admissible vectors $a_j$; these searches, in turn, require time complexity that is polynomial in $\Delta$.

Note also that $C_\exp$ is an additive group over $\Sigma$ and, therefore, the Hamming distance between any two codewords $c_1, c_2$ in $C_\exp$ equals the Hamming weight of $c_1 - c_2$, measured over $\Sigma$. The resulting overall code $\mathcal{E}_t$ is, therefore, of length $n$ and rate $r_0/\Delta$ over $\Sigma$, which translates into rate $(r_0/\Delta) \cdot \log_\rho q/\log_\rho q$ over $\Sigma$. Observe that $C_\exp$, as a code over $\Sigma$, or $\Phi$, 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 $C_\text{just}$, 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 \( \delta_{\text{min}} < 1 \) such that for every \( \delta \geq \delta_{\text{min}} \)

\[
R_{c}(\delta, q) \geq \gamma_0 (1 - \delta) - \frac{\gamma_1}{\log_2 q}.
\]

(14)

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

\[
n - d \leq 4(\Delta - 1)(1 - \delta_0)n \leq \frac{4}{\Delta} \left( \frac{1}{\delta_0} - 1 \right) n \leq (1 - \delta)n.
\]

(11)

Hence, the relative minimum distance of \( \mathcal{V}_1 \) is at least \( \delta \).

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

\[
R_{\mathcal{V}_1}(\delta, q) = \frac{r_0}{\Delta} \log_q |\Sigma| = \frac{r_0}{\Delta} \cdot \frac{m \Delta}{\log_2 q}
\]

(15)

Now, it is easy to verify that \( H_{r,0}(x) \leq x + 1/m \log_2 p \) and, hence, \( 2H_{r,0}(\frac{1}{2}) \geq 1 - 2/m \log_2 p \). Also, since \( m \geq m_0 > 4 \), we have

\[
\frac{2}{m \log_2 p} \leq \frac{4}{(m + 1) \log_2 q} \leq \frac{4}{\log_2 q} \leq \frac{4}{m \log_2 p} \leq 5 < 1.
\]

Therefore,

\[
\frac{1}{2H_{r,0}(\frac{1}{2})} \leq 1 - \frac{1}{(4\Delta / \log_2 q)} \leq 1 + O\left( \frac{\Delta}{\log_2 q} \right).
\]

(17)

Substituting (17) into (16) we obtain

\[
R_{\mathcal{V}_1}(\delta, q) \geq \left( \frac{1}{2} - \delta_0 - O\left( \frac{\Delta}{\log_2 q} \right) \right) \left( \frac{1}{\Delta} - \frac{1}{\log_2 q} \right)
\]

\[
= \frac{1}{\Delta} - \frac{\delta_0}{\log_2 q} + O\left( \frac{\Delta}{\log_2^2 q} \right).
\]

where we have absorbed the constant multipliers which depend on \( \delta_0 \) and \( \rho \) in the \( O(\cdot) \) expressions. Therefore, in terms of \( \Delta \) and \( q, R_{\mathcal{V}_1}(\delta, q) \) satisfies

\[
R_{\mathcal{V}_1}(\delta, q) \geq \frac{1}{\Delta} - \frac{\delta_0}{\log_2 q} - O\left( \frac{1}{\log_2 q} \right).
\]

(18)

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

\[
\frac{1}{\Delta} \geq \frac{1}{4} \left( \frac{1}{\delta_0 - 1} \right) (1 - \theta(\delta)).
\]

(19)

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

\[
R_{\mathcal{V}_1}(\delta, q) \geq \left( \frac{1 - \delta_0}{4(1/\delta_0 - 1)} \right) (1 - \theta(\delta)) \cdot (1 - \delta) - O\left( \frac{1}{\log_2 q} \right).
\]

(20)

Define

\[
\alpha_0 \triangleq \frac{1 - \delta_0}{4(1/\delta_0 - 1)}.
\]

(21)

Assuming that \( m \geq m_0 \), we conclude that for every constant \( \gamma_0 > 0 \), there exists a real number \( \delta_{\text{min}} \) (which depends on \( \gamma_0 \) and \( \delta_0 \)) such that

\[
R_{\mathcal{V}_1}(\delta, q) \geq \gamma_0 (1 - \delta) - O\left( \frac{1}{\log_2 q} \right)
\]

whenever \( \delta \geq \delta_{\text{min}} \).

Finally, we consider the case \( m < m_0 \), which corresponds to \( \Delta > (\log_2 q)/m_0 \). By (19) we have

\[
\frac{1 - \delta}{4(1/\delta_0 - 1)} (1 - \theta(\delta)) \leq \frac{1}{\Delta} \leq \frac{m_0 \log_2 p}{\log_2 q} = O\left( \frac{1}{\log_2 q} \right).
\]

Therefore, we may choose \( \gamma_1 \) 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 \( \alpha_0 \) in (21) is attained at

\[
\delta_0 = \delta_{\text{max}} \triangleq 1 - \frac{1}{\sqrt{2}} \approx 0.29.
\]

(22)

in which case

\[
\alpha_0 = \alpha_{\text{max}} \triangleq \frac{1}{24} + 16 \sqrt{2} \approx 0.021.
\]

(23)

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

\[
\delta_\rho = 1 - \frac{4}{p + 1} \left( \frac{1}{\delta_0} - 1 \right).
\]
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, \( \Sigma \) is taken as \( \text{GF}(q^{m(p+1)}) \), then (16) becomes

\[
R_{\gamma}(\delta, q^{m(p+1)}) = \frac{1 - \delta_0/H_{\Sigma}(\delta)}{2(p + 1)}
\]

\[
= \left( 1 - \frac{\delta_0}{H_{\Sigma}(\delta)} \right) \cdot (1 - \delta_p)
\]

\[
\approx \alpha_0(\rho, \delta_0, m)(1 - \delta_p).
\]

Clearly, for \( \delta_0 = \delta_{\text{max}} \) we have

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

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

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

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

This implies that for relative minimum distances in the range \((1 - \gamma_0)(1 + \gamma_0)^{-1} < \delta < 1\), the function \( \delta \to \gamma_0(1 - \delta) \) lies strictly above the curve \( \delta \to R_{\text{Zyablov}}(\delta) \). Hence, for values of \( \delta \) close to 1, and for sufficiently large \( q \), Construction \( \gamma_1 \) lies above the Zyablov bound.

Finally, as for the explicitness of Construction \( \gamma_1 \), we have already pointed out that the only required searches are those of finding the minimum \( \Delta \) that satisfies (11), and then finding all expressions for \( \Delta - 1 \) of the form of sums of four integer squares. However, since \( \Delta \) is proportional to \( 1/R_{\gamma}(\delta, q) \), 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 \( \gamma_{\text{op}} \) over \( \Phi \).

We remark that finding a polynomial-time decoding algorithm for \( \gamma_1 \), for correcting up to \((\delta - 1)/2\) errors remains still an open problem.

IV. Good Codes over Specific Alphabets

We now use Construction \( \gamma_1 \), as an outer code in a concatenation scheme, obtaining a new code family over any finite field \( \Phi = \text{GF}(q) \). Referring to the notations of Section III, we fix \( \delta_0 \) to some real positive number \( < \frac{1}{2} \) (say), to \( \delta_{\text{max}} \) as in (22)). For any \( \eta \in [0, 1) \) let \( \Delta(\eta) \) denote the smallest integer satisfying

\[
\Delta(\eta) = \alpha_0(\rho, \delta_{\text{max}}, m) = \alpha_{\text{max}}.
\]

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

Construction \( \gamma_2 \) over \( \Phi = \text{GF}(q) \) is obtained as follows. As an outer code, we take Construction \( \gamma_1 \) of length \( n \) and relative minimum distance \( \eta \) over the alphabet \( \Sigma = \text{GF}(q^{m(\rho + 1)}) \). The inner code will be taken as a linear code over \( \text{GF}(q) \) of rate \( r \), dimension \( m\Delta(\eta) \) and relative minimum distance \( \mu \). The overall code is, therefore, a linear code over \( \text{GF}(q) \) of rate \( R = r \cdot R_{\gamma}(\eta, q^{m(\rho + 1)}) \), 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 \( r \geq 1 - H_\Sigma(\mu) \) and, therefore,

\[
R_{\gamma}(\eta, q^{m(\rho + 1)}) \geq (1 - H_\Sigma(\mu)) \cdot R_{\gamma}(\eta, q^{m\Delta(\eta)}).
\]

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

\[
R_{\gamma}(\delta) \geq (1 - H_\Sigma(\mu)) \cdot R_{\gamma}(\delta/\mu, \infty)
\]

\[
\geq \gamma_0(1 - H_\Sigma(\mu)) \left( 1 - \frac{\delta}{\mu} \right).
\]

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

As for the value of the constant \( \gamma_0 \) in (25), we note that when \( \delta \) is close enough to \( 1 - 1/q \), \( \delta/\mu \) must be close to 1. Hence, in the zero-rate neighborhood, \( \gamma_0 \) can be any constant greater than \( \alpha_{\text{max}} \) (as in (23)).

The multiplier \( \gamma_0 \) in (25) can be slightly improved if we replace the \( C_{\text{in}} \) component in Construction \( \gamma_1 \) by a linear code \( C_{\text{in}} \) over \( \Phi = \text{GF}(q^m) \) that consists of a concatenation of two Reed–Solomon codes. The code \( C_{\text{in}} \) was used as the outer code by Sugiyama et al. in [22], where it was also shown that for a prescribed relative minimum distance \( \delta_0 \), the rate \( R_{\text{in}}(\delta_0) \) and length \( N_{\text{in}}(\delta_0) \) of \( C_{\text{in}} \) satisfy

\[
R_{\text{in}}(\delta_0) \geq (1 - \sqrt{\delta_0})^2
\]

and

\[
N_{\text{in}}(\delta_0) \geq q^{m\Delta(\delta_0)} \sqrt{\text{dim}(\gamma_{\text{in}}(\delta_0))}.
\]

Although \( C_{\text{in}} \) is not asymptotically good over the (fixed) field \( \Phi \) (in the sense that \( N_{\text{in}}(\delta_0) \) cannot take arbitrarily large values), \( N_{\text{in}}(\cdot) \) is large enough to let the whole Wozencraft's ensemble be concatenated to our modified Construction \( \gamma_1 \) (the proof of this assertion follows along the lines of that in [22]). We can now substitute \( \gamma_0 = (1 - \sqrt{\delta_0})^2 \) in (15) and repeat the derivations of Lemma 1, ending by replacing the expression for \( \alpha_0 \) in (21) by

\[
\alpha_0 = \frac{(1 - \sqrt{\delta_0})^2}{4(1/\delta_0 - 1)}.
\]

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

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

V. Application to \( t \)-Independent Sets

In this section, we show how Construction \( \gamma_2 \) 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 \(\varepsilon_2\) over GF(q) for a prescribed relative minimum distance \(d = 1 - 1/q - \epsilon\), where \(\epsilon > 0\). Denote by \(n'\) the length of the inner (Wozencraft’s ensemble) code in \(C\), and let \(\mu = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{n'/\sqrt{n'}}\}\) be a nonzero codeword of \(C\), where each \(\varepsilon_i\) stands for an inner codeword. Let \(\mu\) be the value that maximizes the right-hand side of (5); note that, since \(\epsilon > 0\), it is strictly greater than \(\delta\). 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 \(c_i\) have Hamming weight \(\geq n'\delta\); that is, virtually all of the nonzero \(c_i\) contain at most \(n'(1 - \delta)\) zeros. A similar argument implies that in all but a negligible fraction of the nonzero \(c_i\), 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 \(\delta\), 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 \(n'(\pm \epsilon')\) 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 \(\epsilon' = 2^{t+1}/t\). Also, let \(H\) be a \(k \times m\) parity-check matrix of an \([m, m - k, t + 1]\) linear code over GF(2). Since every \(t\) columns of \(H\) are linearly independent, for any nonzero vector \(x \in F^m\) of weight \(\leq t\) we have \(Hx^T \neq 0\). Now, define the \(m \times n\) matrix \(A = H^T G\), 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 \(u \in F^t\) that does not appear in \(B\) the Hamming weight of \(\sum_{u \in F^t} (x, u)\) for any \(x \in F^t\) is \(\leq t - 2\). Let \(s_j\) be the vector whose entries are given by \(s_j = n_j/n, x \in F^t\); similarly, let \(w\) be the vector whose entries are \(w_u = wt(uB)/n, u \in F^t\). By (27) we have

\[
1 - 2w = s_j, 
\]

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

\[
2s_j = s_j(1 - 2w). 
\]

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

\[
2\epsilon \cdot s_j = \sum_{u \in F^t} (1 - 2w_u)(-1)^{(u, x)} 
= 1 + \sum_{u \in F^t} (1 - 2w_u)(-1)^{(u, x)} 
\geq 1 - 2(1 - 2\epsilon) \cdot 2^{t+1/2} 
\geq 0. 
\]

Hence, given \(t\), for sufficiently large \(n\) we have \(s_j > 0\) for all \(x \in F^t\), thus proving that every vector \(x \in F^t\) 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 \([0, 1]^m\).

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 \(k \leq \lfloor (1/2) \cdot \log_2(m + 1) \rfloor\). Also, since \(C\) is an instance of Construction \(\varepsilon_2\),

\[
n \leq c_1 \cdot k / \epsilon^3 = c_1 \cdot 2^{3/2} k, 
\]

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 sufficiently large \(m\), the previous \(t\)-independent construction is of size \(\leq 2^{2\log m}\), 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 \(O(2^t \log m)\) [12, 21].

The previous construction method for \(t\)-independent sets is based upon the technique introduced in [17] for obtaining so-called \(t\)-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 is possible since code \(C\) used here is better than the one used in [17].

REFERENCES


On Generator Matrices of Codes

Jurianna Simonsi

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

\[
t := n - \sum_{j=0}^{k-1} \left\lfloor \frac{d}{q^j} \right\rfloor
\]

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.

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 \( C \in \F_q^m \) of dimension \( k \) and minimum weight \( d \) can be transformed into a code \( D \in \F_q^n \) with the same parameters such that \( D \) possesses a basis of weight \( d + t \) vectors.

**Proof:** In the sequel, the linear subspace of \( \F_q^m \) spanned by a set of vectors \( x, y, \ldots, z \in \F_q^m \) will be denoted by \( \langle x, y, \ldots, z \rangle \). Let \( \{a_1, a_2, \ldots, a_r\} \subseteq \F_q^m \) be a maximal set of independent codewords of weight \( d \). Suppose that \( r < k \). All codewords in the complement of the span \( \langle a_1, a_2, \ldots, a_r \rangle \) of the \( a_i \) have weight \( \geq d \). Pick a codeword \( x \not\in \langle a_1, a_2, \ldots, a_r \rangle \) of lowest weight, say \( d \), and extend \( \{a_1, a_2, \ldots, a_r, x\} \) to a basis \( \{a_1, a_2, \ldots, a_r, b_1, \ldots, b_{k-r}\} \) of the code \( C \). Now change \( b_i \) into a vector \( b'_i \) of weight \( d \) by changing \( d - d \) of the nonzero coordinates into zero ones. Then, linear subspace

\[
\langle a_1, a_2, \ldots, a_r, b'_1, \ldots, b'_{k-r} \rangle \subseteq \F_q^m
\]

is a code of minimum weight \( d \), because \( \{a_1, a_2, \ldots, a_r\} \) is unaltered and the words of \( \langle a_1, a_2, \ldots, a_r, x \rangle \) have changed in at most \( d - d \) coordinates. We claim that the dimension of \( \langle a_1, a_2, \ldots, a_r, x \rangle \) is equal to \( k \). For if \( \dim \langle a_1, a_2, \ldots, a_r, x \rangle < k \), then \( b'_i \) would be a linear combination of the vectors \( a_1, a_2, \ldots, a_r, b_1, \ldots, b_{k-r} \), and, thus, would be an element of the original code \( C \). Since the weight of both \( b'_i \) and \( b_i - b'_i \) is smaller than \( d \), these vectors would in fact be contained in the linear subspace \( \langle a_1, a_2, \ldots, a_r, x \rangle \) which contradicts the fact that \( b_i \not\in \langle a_1, a_2, \ldots, a_r, x \rangle \). So \( \langle a_1, a_2, \ldots, a_r, x \rangle \) has the same parameters as \( C \) has, but the maximum number of independent weight \( d \) codewords in \( \langle a_1, a_2, \ldots, a_r, x \rangle \) exceeds that of \( C \). The induction process is obvious.

**Remark:** As one of the referees observed, it may happen that the resulting code \( D \) is contained in a coordinate hyperplane of \( \F_q^m \) and thus effective length \( < n \). An example is the code \( C \subseteq \F_q^n \) with generator matrix

\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 1
\end{bmatrix}
\]

REFERENCES


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-