Constructive asymptotic lower bounds on linear codes and weakly biased arrays

Jürgen Bierbrauer
Department of Mathematical Sciences
Michigan Technological University
Houghton, Michigan 49931 (USA)
and
Holger Schellwat
Department of Technology and Science
University of Örebro (Sweden)

December 2, 1998

1 Introduction

Let q be a prime-power,$0 \leq \delta, R \leq 1$. We say that (δ, R) is **asymptotically reachable** if there is an infinite family $\mathcal{C}_i$ of linear $q-$ary codes $[n_i, k_i, d_i]$ such that $n_i, k_i, d_i \longrightarrow \infty$ for $i \longrightarrow \infty$ and $lim_{i \longrightarrow \infty} \frac{k_i}{n_i} \geq R, lim_{i \longrightarrow \infty} \frac{d_i}{n_i} \geq \delta$. Let L be a line in the $(\delta, R)-$plane. We say that L is asymptotically reachable if all points on L in the first quadrant are asymptotically reachable by linear q-ary codes. A classical lower bound is the Gilbert-Varshamov bound:

Theorem 1 *Let $H(x) = H_2(x) = -xlog(x) - (1-x)log(1-x)$ be the binary entropy function ($0 \leq x \leq 1$) and*

$$H_q(x) = \frac{1}{log(q)}(H(x) + xlog(q-1))$$

the q-ary entropy function. Then all points on the curve

$$R = 1 - H_q(\delta)$$

are reachable by linear q-ary codes.

A drawback of the simple Theorem 1 is that it is essentially an existence result. Families of codes meeting the bound cannot be constructed efficiently. The **Zyablov bound** [15] states that the line $R=(1-H_q(\mu))(1-\frac{\delta}{\mu})$ is asymptotically reachable for all $0\leq\mu\leq(q-1)/q$ by codes, which are obtained by standard concatenation with a Reed-Solomon code as an outer code and a code reaching the Gilbert-Varshamov bound as inner code. This bound may be called semi-constructive. A first constructive lower bound was obtained by Justesen [8]. We use Justesen's idea and algebraic-geometric codes to show that the Zyablov bound is efficiently constructible.

2 AG-codes and the Justesen construction

We start with a generalization of Justesen's construction.

Definition 1 (generalized Justesen construction) *Let q be a prime-power, m a natural number and $e\geq 1$ a rational number such that em is integer. Let $\mathcal{C}_m$ be a linear q^m-ary code of length q^{em}. Choose $\mathbb{F}_q$-linear injective mappings as follows:*

$$\alpha:\mathbb{F}_{q^m}\longrightarrow\mathbb{F}_q^m,\beta:\mathbb{F}_{q^m}\longrightarrow\mathbb{F}_{q^{em}},\gamma:\mathbb{F}_{q^{em}}\longrightarrow\mathbb{F}_q^{em}.$$

Let $\mathbb{F}_{q^{em}}=\{u_i\}$. Also, let $s\leq em$ and let $\pi:\mathbb{F}_q^{em}\longrightarrow\mathbb{F}_q^{em-s}$ be a projection. We define an $\mathbb{F}_q$-linear code $\tilde{\mathcal{C}}_m$ by concatenation and projection, as follows: let $(v_i)\in C_m$. The corresponding codeword of $\tilde{\mathcal{C}}_m$ is obtained by mapping the entry v_i to

$$\tilde{v_i}=(\alpha(v_i),\pi(\gamma(u_i\beta(v_i))))\in\mathbb{F}_q^{m(e+1)-s}$$

The Justesen construction proper is obtained in the special case when $e=1,q=2,\mathcal{C}_m$ is a Reed-Solomon code, and $s=0$. In that case π is the identity and $\tilde{v_i}=(\alpha(v_i),\gamma(u_i\beta(v_i)))\in\mathbb{F}_2^{2m}$. Consider the case when $s=0$. The basis for Justesen's construction is the following observation, which obviously is true: whenever $i\neq j$ and $v_iv_j\neq 0$, then the $m(e+1)-$tuples $\tilde{v_i}$ and $\tilde{v_j}$ are different. In the general case of Definition 1 we are led to the following:

Lemma 1 (Justesen's observation) *In the situation of Definition 1 let $0\neq(v_i)\in\mathcal{C}_m$. Then for every $x\in\mathbb{F}_q^{m(e+1)-s}$ there are at most q^s indices i such that $\tilde{v_i}=x$.*

Theorem 2 *Let q be a prime-power and $e \geq 1$ a rational number. Assume there is an infinite set $\mathcal{M}$ of natural numbers such that em is a natural number for all $m \in \mathcal{M}$ and for every $m \in \mathcal{M}$ there is an algebraic curve $\mathcal{K}_m$ defined over $\mathbb{F}_{q^m}$ with more than q^{em} rational points and genus g_m, such that $lim_{m\in\mathcal{M}} \frac{g_m}{q^{em}} = 0$. Then the generalized Justesen construction from Definition 1 based on the standard AG-codes derived from $\mathcal{K}_m$ yields a family of linear $q-$ary codes, which show that the line $R = (1 - H_q(\mu))(1 - \frac{\delta}{\mu})$ is asymptotically reachable for all $0 \leq \mu \leq H_q^{-1}(\frac{e}{e+1})$.*

Proof: Fix R_0, where $0 < R_0 < 1$. It follows from the standard mechanism of AG-codes that there is a q^m-ary linear code $\mathcal{C}_m$ with parameters

$$[q^{em}, R_0 q^{em}, (1 - R_0)q^{em} - g_m].$$

Use the terminology of Definition 1. Code $\tilde{\mathcal{C}}_m$ has length $(m(e+1) - s)q^{em}$ and dimension $R_0 m q^{em}$. Its rate is therefore $R_0 \frac{m}{m(e+1)-s}$. Let $n = m(e+1) - s$. Fix a constant c such that $\frac{1}{e+1} \leq c \leq 1$ and choose s such that $lim_m \frac{m}{n} = c$. It follows from Lemma 1 that whenever $\sum_{j=0}^{r} \binom{n}{j}(q-1)^j \leq Dq^{-s}$, then $d \geq \sum_{j=0}^{r} j\binom{n}{j}(q-1)^j$. Here $D \geq (1 - R_0)q^{em} - g_m$ is the minimum distance of $\mathcal{C}_m$. We use the fact that the term for $j = r$ is asymptotically dominating and that asymptotically $\binom{n}{i} \sim 2^{nH(i/n)}$. Take the binary logarithm and divide by n. With $x = \frac{r}{n}$ the condition becomes after simplification

$$H(x) + xlog(q-1) \leq \frac{me - s}{n} log(q) + \frac{1}{n} log(1 - R_0 - \frac{g_m}{q^{me}}).$$

Because of our condition on the genus the last summand on the right vanishes in the limit. As $lim_m \frac{me-s}{n} = 1 - c$ the condition is asymptotically equivalent with

$$H_q(x) \leq 1 - c.$$

We can choose $x = H_q^{-1}(1-c)$. It follows $\delta = \lim_m \frac{d}{nq^{me}} \geq (1-R_0)H_q^{-1}(1-c)$. A formal proof of this fact follows from [14],Lemma (9.2.2) in the binary case. Expressing R_0 in terms of the rate R we obtain $\delta \geq (1 - (e+1)R)H_q^{-1}(\frac{e}{e+1})$. Solving for R yields the line $R = c - \frac{c}{H_q^{-1}(1-c)}\delta$. Let $\mu = H_q^{-1}(1 - c)$. This yields the claim of the theorem.∎

Theorem 2 shows that families of codes attaining the Zyablov bound can be obtained by applying the Justesen construction to suitable families of algebraic-geometric codes. The resulting families of codes are just as constructive as the underlying algebraic curves are. The construction given by Garcia and Stichtenoth in [5] is rather explicit. This family of algebraic curves satisfies the assumptions of Theorem 2 for all rational $e \geq 1$. It follows that the corresponding codes can be used to construct families of linear q-ary codes meeting the Zyablov bound for all q. The first layer in the construction of [5] are the Hermitian curves. These satisfy the assumptions of Theorem 2 for $e = 1.5$. In the binary case the Suzuki curves (see [7]) reach $e = 2$. They yield asymptotically nontrivial binary families of relative distance $\delta \leq H^{-1}(2/3) = 0.1739$.

3 Weakly biased arrays

Definition 2 *An $(n,m)-$array is a binary array with n rows and m columns.*

In the applications the rows of the array form a probability space Ω, with uniform distribution. Each column describes a random variable defined on Ω with values in $\mathbb{F}_2$. We have chosen to use a neutral terminology. The following is a definition in the style of [11]. Let $x_1, x_2, \ldots, x_k$ be $\{0,1\}$ random variables and $0 \leq \epsilon \leq 1$. The probability distribution defined by the x_i is $\epsilon-$biased if for every nonempty subset $\emptyset \neq S \subseteq \{1,2,\ldots,k\}$ we have

$$\mid Prob(\sum_{i \in S} x_i = 0) - Prob(\sum_{i \in S} x_i = 1) \mid \leq \epsilon$$

We use our terminology to simplify:

Definition 3 *Let $v = (v_1, v_2, \ldots, v_n) \in \mathbb{F}_2^n$. The* **bias** *of v is defined as $bias(v) = \mid 1 - \frac{2w}{n} \mid$, where $w = wt(v)$ is the weight of v.*

We see that $bias(v) = \mid Prob(x_i = 0) - Prob(x_i = 1) \mid$ if the indices i are chosen uniformly at random. It is also clear that $bias(v) = bias(v+1)$. The definition of an $\epsilon-$biased binary probability distribution as given in the beginning of this section is equivalent to the following:

Definition 4 *Let $0 \leq \epsilon < 1$. An $(n,k)-$array is $\epsilon-$***biased** *if every nontrivial linear combination of its columns has bias $\leq \epsilon$.*

Let $\mathcal{C}$ be the linear code generated by the columns of the $\epsilon-$**biased** $(n,k)-$array $\mathcal{A}$. If $dim(\mathcal{C}) < k$, then the 0-word is a nontrivial linear combination of the columns of $\mathcal{A}$. This shows that $\mathcal{A}$ cannot have a bias < 1. The same is true if the all-1 word $\mathbf{1}$ is in $\mathcal{C}$. We can therefore assume that $dim(\mathcal{C}) = k$ (equivalently $\mathcal{A}$ has maximal rank k) and that $\mathbf{1} \notin \mathcal{C}$. Let $0 \neq v \in \mathcal{C}, wt(v) = w$. The condition $bias(v) \leq \epsilon$ can be written equivalently as

$$\frac{1-\epsilon}{2} \leq \frac{w}{n} \leq \frac{1+\epsilon}{2}$$

Let d be the minimal distance and D the maximal distance of $\mathcal{C}$. The bias-condition is then equivalent to

$$\frac{d}{n} \geq \frac{1-\epsilon}{2} \text{ and } \frac{D}{n} \leq \frac{1+\epsilon}{2}$$

The latter condition can be written equivalently as $\frac{n-D}{n} \geq \frac{1-\epsilon}{2}$. We have seen the following:

Theorem 3 *Let $0 \leq \epsilon < 1$. The following are equivalent:*

- *An (n,k)-array, which is $\epsilon-$biased.*
- *A binary linear code $\mathcal{C}$ of length n and dimension k, whose minimal distance d and maximal distance D satisfy*

$$Min\{\frac{d}{n}, \frac{n-D}{n}\} \geq \frac{1-\epsilon}{2}$$

The $\epsilon-$biased array is simply the transpose of a generator matrix of code $\mathcal{C}$. Let $\mathcal{C}' =< \mathcal{C}, \mathbf{1} >$ be the $(k+1)-$dimensional code generated by $\mathcal{C}$ and $\mathbf{1}$. It is obvious that the minimum distance of $\mathcal{C}'$ is precisely $Min\{d, n-D\}$. This gives us the possibility to express the defining property of binary $\epsilon-$biased arrays in terms of the minimum distance of binary linear codes, as follows:

Theorem 4 *Let $0 \leq \epsilon < 1$. The following are equivalent:*

- *An (n,k)-array, which is $\epsilon-$biased.*

- *A binary linear code $\mathcal{C}'$ of length n and dimension $k+1$, which contains* $\mathbf{1}$ *and whose minimal distance d satisfies*

$$\frac{d}{n} \geq \frac{1-\epsilon}{2}$$

The **Gilbert-Varshamov bound** gives us an existence result for binary $\epsilon-$biased arrays. In a slightly enhanced version the Gilbert-Varshamov bound states the following (see [3]):

Theorem 5 *Let $V(n,i) = \sum_{j=0}^{i} \binom{n}{j}$. If $V(n-1, d-2) < 2^{n-k}$, then every linear code $[n,l,d]_2$, where $l \leq k$ is embeddable in a code $[n,k,d]_2$.*

If we apply this to the case where $[n,l,d]_2 = [n,1,n]_2$ (the code generated by the all-1 word), then Theorem 4 yields the following:

Theorem 6 *If $V(n-1,d-2) < 2^{n-k-1}$, then there is an $(n,k)-$array, which is $\{1-2\delta\}-$biased. Here $\delta = d/n$, the relative minimum distance.*

The asymptotic version of the Gilbert-Varshamov bound says in the binary case that if $\delta \leq 1/2$ and $R \leq 1 - H(\delta)$, then there is an infinite series of binary linear codes whose lengths n_i, dimensions k_i and distances d_i go to infinity such that $lim(k_i/n_i) \geq R, lim(d_i/n_i) \geq \delta$. Clearly this remains true when we choose $\mathbf{1}$ to be contained in each of these codes. Here H is the binary entropy function. The following is obtained:

Theorem 7 *Let $\delta \leq 1/2$ and $R = 1 - H(\delta) = 1 + \delta log_2(\delta) + (1-\delta)log_2(1-\delta)$. Then there is a family of $(1-2\delta)-$ biased $(n_i,k_i)_2-$arrays such that $lim(k_i/n_i) \geq R$.*

A serious drawback in the application of the Gilbert-Varshamov bound is that these codes are not really efficiently constructed.

It is natural to ask questions of the following type:

Problem 1

1. *What is the smallest exponent a such that $n = (\frac{k}{\epsilon})^a$ is asymptotically reachable by $\epsilon-$biased $(n,k)-$arrays?*
2. *What is the smallest exponent b such that $n = \frac{k}{\epsilon^b}$ is asymptotically reachable by $\epsilon-$biased $(n,k)-$arrays?*

For which exponents a,b can we find explicit efficient constructions?

We intend to show that for every $a > 1$ and $b > 2$ we can find efficient constructions.

4 The Justesen construction and weakly biased arrays

Consider the codes constructed from Theorem 2 in case $q = 2, s = 0$. Lemma 1 shows that we can derive a bound on the number of zeroes in an arbitrary codeword, which certainly is not inferior to our bound on the weight. It follows from Theorem 3 that the transpose of the generator matrix of each of our codes is a binary array, which is $(1-2\delta)-$biased. Choose some $\mu < 1/2$ and $\delta = 2\mu - 1/2$. Let e such that $\mu \leq H^{-1}(\frac{e}{e+1})$. We obtain binary arrays with 2^{em} rows and $R2^{em}$ columns, where $R = (1-H(\mu))(1-\frac{\delta}{\mu}) = (1-H(\mu))(1-2\mu)/(2\mu)$, which is $\epsilon-$biased, where $\epsilon = 2(1-2\mu)$. In particular for each such choice of μ we obtain an infinite family of binary $\epsilon-$ biased $(n,k)-$arrays such that $n = \frac{k}{\epsilon}\frac{4\mu}{1-H(\mu)}$.

5 The generalized powering construction

A very useful direct construction is in [2]. These authors show that $n = (\frac{k}{\epsilon})^2$ can always be reached. Three variants are considered in [2]. These are related to shift-register sequences, quadratic residues and finite fields. All constructions yield the same parameters. We start from the third of these constructions, termed the **Powering construction** by the authors. It is easily seen that this construction may alternatively be described as being based on Reed-Solomon codes over a field $\mathbb{F}_{2^m}$. A conceptual description involves the trace. The construction is easier to understand in a generalized form.

Theorem 8 (generalized powering construction) *Let $\mathcal{C}$ be a code $[n,k,d]_q$, where $q = 2^m$. Let $v(1),\dots,v(k)$ be a basis of $\mathcal{C}$. Construct a $(qn,k)_2$-array $A(\mathcal{C})$ as follows: the columns are indexed by the basis elements $v(j)$, the rows are indexed by pairs (i,u), where i is a coordinate of the ambient space $\mathbb{F}_q^n$ containing the code $\mathcal{C}$ and $u \in \mathbb{F}_q$. The entry in row (i,u) and column j is*

$$tr(uv(j)_i).$$

Here $v(j)_i$ is the entry of $v(j)$ in coordinate i and $tr : \mathbb{F}_q \longrightarrow \mathbb{F}_2$ is the trace. Then $A(\mathcal{C})$ has bias $\leq 1 - \frac{d}{n}$.

Proof: Let $\alpha = (\alpha_1, \ldots, \alpha_k) \neq (0, \ldots, 0)$. Let c be the corresponding linear combination of the columns of $A(\mathcal{C})$. The entry of c in row (i, u) is

$$\sum_{j=1}^{k} \alpha_j tr(uv(j)_i) = tr((\sum_j \alpha_j v(j)_i)u) = tr(v_i u).$$

Here $0 \neq v = \sum_j \alpha_j v(j) \in \mathcal{C}$. We imagine the rows of the array as being subdivided in n segments according to the value of i. If $v_i \neq 0$, then the corresponding segment of c is unbiased. If $v_i = 0$, then the corresponding segment is 0. We obtain

$$bias(c) \leq \frac{(n-d)q}{nq} = 1 - \frac{d}{n}.$$

■

We see that array $A(\mathcal{C})$ may be best described in terms of a generator matrix of $\mathcal{C}$. The elements $v(j)_i$ are the entries of a generator matrix. The basic idea of randomization is the same in Theorem 8 and in the Justesen construction (see Definition 1). Use of Reed-Solomon codes leads to arrays $(2^{2m}, k)_2$, where $\epsilon < k/2^m$. If we express n in terms of k and ϵ we obtain $n = (k/\epsilon)^2$, as promised. Reed-Solomon codes are algebraic-geometric codes corresponding to genus 0. It is natural to ask if AG-codes derived from curves of higher genus yield stronger results. This is indeed the case.

Theorem 9 *Assume there is an algebraic curve defined over $\mathbb{F}_q, q = 2^m$, with more than n rational points and genus g. Let $0 < R < 1$ such that Rn is integer. Then we can construct a $(qn, Rn)_2$-array which is $\epsilon-$biased, where $\epsilon < R + \frac{g}{n}$.*

This follows from the standard mechanism of AG-codes and Theorem 8. Observe that the weakly biased array will be just as constructive as the underlying curve. An asymptotic version is obtained by using families of curves. We have mentioned that efficient explicit constructions exist for families of AG-codes satisfying the assumptions of Theorem 2. If we use the codes derived from these curves in Theorem 9 the following asymptotic result is obtained:

Theorem 10 *For every $e \geq 1$ and every $0 < \epsilon < 1$ there is a family of $(n_i, k_i)_2-$arrays, which are ϵ_i- biased such that $lim_i \epsilon_i = \epsilon$ and*

$$lim_i(\frac{n_i}{(k_i/\epsilon_i)^{(e+1)/e}}) = 1.$$

We see that we find families of $\epsilon-$biased $(n, k)_2-$arrays, where asymptotically $n = (\frac{k}{\epsilon})^a$, for every $a > 1$. Moreover these constructions are explicit. Let us illustrate with the Hermitian curves in characteristic 2. These are defined over $\mathbb{F}_{q_0^2}$, where $q_0 = 2^m$. They have $q_0^3 + 1$ rational places and genus $\binom{q_0}{2}$. Application of Theorem 9 yields a $(q_0^5, Rq_0^3)-$array, which is $\epsilon-$ biased, where $\epsilon = R + \binom{q_0}{2}/q_0^3 < R + 2^{-(m+1)}$. Asymptotically for large m this yields $\epsilon-$biased arrays $((\frac{k}{\epsilon})^{5/3}, k)$ for every ϵ. Similarly the Suzuki codes (see [7]) asymptotically achieve the value $n = (\frac{k}{\epsilon})^{3/2}$.

6 Constructions using expander graphs

In order to attack the second instance of Problems 1 we use expander graphs. The following concept is suggested by an analysis of the constructions in [11] and [1].

Definition 5 *A* **block covering system** *$BCS(n, l, b, \gamma, \epsilon)$ is a family $\mathcal{B}$ of l-element subsets of a fixed $n-$set Ω, where $\mid \mathcal{B} \mid= b$ and the following holds: whenever $W \subset \Omega, \mid W \mid= w$ such that $1 - \frac{w}{n} \leq \gamma$, then*

$$\frac{\textit{number of blocks avoiding } W}{b} \leq \epsilon.$$

Here it is permitted that a point occurs in a block with multiplicity more than 1. It is also permitted that different blocks are made of the same set of points, even with the same multiplicities. Known sources for good BCS are graphs with extremal Eigenvalue properties (see the papers mentioned above) and orthogonal arrays (see [6]). BCS may equivalently be described as bipartite graphs with certain expansion properties. In this language they are known as **bounded strong concentrators** (see [4]). The following is proved in [1].

Theorem 11 *Let Γ be a graph on n vertices, which is connected and regular of valency Δ. Let λ_1 be the second-largest Eigenvalue of Γ. The family $\mathcal{B} = \{i^\perp \mid i \in V\}$ is a $BCS(n, \Delta, n, \gamma, \frac{\gamma}{1-\gamma}(\frac{\lambda_1}{\Delta})^2)$.*

The following is a variant of the main construction of [1]:

Theorem 12 *Assume the following exist:*

- *A binary linear code $\mathcal{C}$ with parameters $[n, k, d]_2$.*
- *A $BCS(n, l, b, 1 - \frac{d}{n}, \epsilon)$ for some l.*
- *An ϵ_0-biased $(n_0, l)_2-$array A.*

Then we can construct an $(n_0 b, k)-$array, which is $\{\epsilon + \epsilon_0(1-\epsilon)\}-$biased.

Proof: At first we construct a 2^l- ary additive code $\tilde{\mathcal{C}}$ of length b, with just as many codewords as $\mathcal{C}$ and minimum distance (equal to minimum weight) $\geq (1-\epsilon)b$, as follows: Let $v \in \mathcal{C}$. The corresponding codeword of $\tilde{\mathcal{C}}$ is $\tilde{v} = (\tilde{v}_B)$, where B varies through the blocks. Here $\tilde{v}_B = (v_i)_{i \in B}$. Observe that we interpret $\tilde{\mathcal{C}}$ as defined over $V(l, 2)$, the vector space of dimension l over $\mathbb{F}_2$. Clearly $\tilde{\mathcal{C}}$ is linear over $\mathbb{F}_2$. This shows that minimum weight and minimum distance agree. Let $v \neq 0$. Then $\tilde{v}_B = 0$ if and only if $B \cap supp(v) = \emptyset$. The definition of a BCS shows that the number of such blocks is at most $\epsilon \cdot b$.

Let $\mathcal{C}'$ be the binary linear code obtained by concatenation with $\tilde{\mathcal{C}}$ as outer code and the code generated by the columns of A as inner code. Our end product is the transpose of a generator matrix of $\mathcal{C}'$. Clearly this is an $(n_0 b, k)-$array. We need to control the bias of an arbitrary codeword $0 \neq v' \in \mathcal{C}'$. If $B \cap supp(v) = \emptyset$, then v'_B is a 0-tuple of length n_0. By definition of a BCS there are at least $(1-\epsilon)b$ blocks B such that $B \cap supp(v) \neq \emptyset$. Each corresponding section v'_B is a nonzero codeword of the inner code, hence of bias $\leq \epsilon_0$. It follows that v'_B can have at most $\epsilon_0 n_0$ more zeroes than ones. The total surplus of zeroes in v' is therefore limited by $\epsilon b n_0 + (1-\epsilon) b \epsilon_0 n_0$. Division by the length bn_0 yields the desired upper bound on the bias.■

Apply Theorem 12 with the following ingredients: Code $\mathcal{C}$ belongs to a family of asymptotically nontrivial codes with relative minimum distance δ. The BCS is derived from Ramanujan graphs (see [10]) on n vertices with valency Δ via Theorem 11. It has parameters $BCS(n, \Delta, n, 1-\delta, \frac{1-\delta}{\delta}\frac{4}{\Delta})$. The weakly biased array belongs to a family constructed in Section 5, an ϵ_0-biased $(n_0, \Delta)-$ array, where $n_0 = (\Delta/\epsilon_0)^{(e+1)/e}$. Choose $\epsilon_0 = \epsilon/2$ and Δ such that $\frac{1-\delta}{\delta}\frac{4}{\Delta} \leq \epsilon/2$. We obtain an $\epsilon-$biased $(N, K)-$ array, where $N = O(K/\epsilon^{2(e+1)/e})$.

7 Weakly biased arrays and covering arrays

Call an $(n, k)-$array t-**wise** $\epsilon-$**biased** if any (n, t)-subarray is $\epsilon-$biased. The following construction is essentially from [11].

Theorem 13 *Let the following be given:*

- *An $(n, k)-$array B, which is $\epsilon-$biased.*
- *A linear code $[N, N-k, t+1]_2$.*

Then we can construct an $(n, N)-$array, which is t-wise $\epsilon-$biased.

In the binary case this yields at the same time a construction for **covering arrays.** Here a **covering array** $(n, N, t)_q$ is an $(n, N)-$array with entries from an alphabet of size q such that every $(n, t)-$subarray contains every vector of length t as a row. The following is essentially from [1]:

Theorem 14 *Let the following be given:*

- *An $(n, k)-$array B, which is $\epsilon-$biased.*
- *A linear code $[N, N-k, t+1]_2$.*

Consider the t-wise $\epsilon-$biased $(n, N)_2-$array $\mathcal{A}$ constructed in Theorem 13. If $\epsilon \leq 2^{-t}$, then $\mathcal{A}$ is a covering array $(n, N, t)_2$.

Let us apply this to BCH-codes and our $\epsilon-$biased arrays. Let $N = 2^i$ and consider (extended) BCH-codes with designed distance $t+1$. We have $k \leq it$. Recall that asymptotically for large k we can construct ϵ -biased $(n, k)-$ arrays, where $n \sim (k/\epsilon)^a$, for any $a > 1$. This yields $(n, N, t)-$arrays, where $n = O((t2^t log(N))^a)$, for any $a > 1$. This comes close to the non-constructive upper bound of $n = O(t2^t log(N))$ mentioned in [1]. We remark that the theory can easily be generalized from the binary case to the case of $I\!\!F_p$ as ground field. It follows that similar bounds on $(n, N, t)_p-$covering arrays can be obtained.

8 Conclusion

We have generalized the Justesen construction in a natural way and shown that application to suitable families of algebraic-geometric codes yields efficient constructions of families of linear codes meeting the Zyablov bound. A similar idea was used by Shen [12] in the binary case. This construction relies on a family of codes, which generalize the Hermitian codes. In the original Justesen construction the inner code has length $2m$. A generalization to larger length has been considered in [13]. The use of the projection π in Definition 1 follows [14]. In [9] modular curves are used for the construction of codes that beat the Zyablov bound. [1] presents a construction of asymptotically nontrivial families of codes based on expander graphs. This paper also contains valuable remarks on the history of the problem of efficient constructions for asymptotically nontrivial codes. We have then used asymptotically nontrivial codes, algebraic curves and expander graphs to obtain improved constructions for weakly biased binary arrays and binary covering arrays.

References

[1] N.Alon,J.Bruck,J,Naor,M.Naor,R.Roth: *Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs, IEEE Transaction on Information Theory* **38**(1992),509-516.

[2] N.Alon,O.Goldreich,J.Håstad,R.Peralta: *Simple constructions of almost k-wise independent random variables,* Symposium 31st FOCS 1990,544-553.

[3] J.Bierbrauer and Y.Edel: *Lengthening and the Gilbert-Varshamov bound, IEEE Transaction on Information Theory* **43**(1997),991-992.

[4] O.Gabber and Z.Galil: *Explicit construction of linear sized superconcentrators, Journal on Computer and System Science* **22**(1981),407-420.

[5] A.Garcia and H.Stichtenoth: *A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound, Inventiones Mathematicae* **121**(1995),211-222.

[6] K.Gopalakrishnan and D.R.Stinson: *A simple analysis of the error probability of two-point based sampling,* Information Processing Letters **60**(1996),91-96.

[7] J.P.Hansen: *Group codes on Deligne-Lusztig varieties,* Coding Theory and Algebraic Geometry: Proceedings Luminy 1991, LNM, Springer **1518** (1992).

[8] J.Justesen: *A class of asymptotically good algebraic codes,* IEEE Transactions on Information Theory **18**(1972),652-656.

[9] G.L.Katsman,M.A.Tsfasman,S.G.Vladut: *Modular curves and codes with a polynomial construction,* IEEE Transaction on Information Theory **30**(1984),353-355.

[10] A.Lubotzky, R.Phillips and P.Sarnak: *Explicit expanders and the Ramanujan conjectures,* Proceedings 18th ACM Symposium Theory of Computing 1986, 240-246.

[11] J.Naor,M.Naor: *Small-bias probability spaces: efficient constructions and applications,* SIAM Journal of Computing **22**(1993),838-856; preliminary version in Proceedings STOC 1990, 213-223.

[12] B.Z.Shen: *A Justesen construction of binary concatenated codes that asymptotically meet the Zyablov bound for low rate,* IEEE Transaction on Information Theory **39**(1993),239-242.

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

[14] J.H.van Lint: *Introduction to Coding Theory,* Springer 1992.

[15] V.V.Zyablov: *An estimate of the complexity of constructing binary linear cascade codes,* Problems in Information transmission **7**(1971),3-10.