Efficient constructions of $\epsilon-$biased arrays, $\epsilon-$dependent arrays and authentication codes

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 6, 1998

1 Introduction

The concepts of limited dependence and limited bias have manifold applications in cryptography and complexity theory. We mention universal hashing, authentication, resiliency against correlation attacks, pseudorandomness, block ciphers, derandomization, two-point based sampling, testing of combinatorial circuits (see [17, 13, 16, 2, 8, 20, 3]). A basic notion underlying these concepts are families of $\epsilon-$biased random variables. The heart of the influential papers by Naor and Naor [16] and by Alon, Goldreich, Håstad and Peralta [2] are constructions of families of binary $\epsilon-$biased random variables. In this paper we analyze, simplify and generalize these constructions. Our language is coding-theoretic and it will turn out that virtually all basic constructions are coding-theoretic.

Definition 1. *An $(n, k)_p-$array $\mathcal{A}$ is an array with n rows and k columns, where the entries are taken from a set with p elements.*

Typically in complexity-theoretic applications one interprets the rows of the array as the points of a probability space with uniform distribution, where each column describes a random variable. One seeks to construct small probability spaces such that the random variables are not far from being independent. When used for the purposes of authentication of messages the columns of the array are identified with the source, each row is a key and the entries are authenticators. In order to simplify description we chose to use neutral notation. Also, from the point of view of applications it is advisable not to restrict to the binary case $p = 2$. We consider the case that p is a prime.

Definition 2. *Let p be a prime, $v = (v_1, v_2, \ldots, v_n) \in \mathbb{F}_p^n$. For every $i \in \mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ let $\nu_i(v)$ be the frequency of i as an entry of v and let $\delta_i(v)$ be defined by $\nu_i(v) = \frac{n}{p} + \delta_i(v)$. Let ζ be a primitive complex $p-$th root of unity. The* **bias** *of v is defined as*

$$bias(v) = \frac{1}{n} \mid \sum_{i \in F_p} \delta_i(v)\zeta^i \mid = \frac{1}{n} \mid \sum_{i \in F_p} \nu_i(v)\zeta^i \mid$$

Observe that $\sum_{i \in F_p} \delta_i(v) = 0$. We see that $0 \leq bias(v) \leq 1$ for every vector v. It is also clear that $bias(v) = 1$ if and only if v is a constant vector and that $bias(v) = bias(v + i)$ for every $i \in \mathbb{F}_p$.

Definition 3. *Let $0 \leq \epsilon < 1$. An $(n, k)_p-$array is $\epsilon-$***biased** *if every non-trivial linear combination of its columns has bias $\leq \epsilon$.*

A (binary) array $(n, k)_2$ is $\epsilon-$biased if and only if the code generated by the columns has dimension k and all nonzero codewords have weight between $1 - \frac{\epsilon}{2}$ and $1 + \frac{\epsilon}{2}$. In order to obtain an efficient description we first introduce some useful concepts.

2 Asymptotically good families of codes

We say that (δ, R) is **asymptotically reachable** by linear $q-$ary codes if there is an infinite family $\mathcal{C}_i$ of codes $[n_i, k_i, d_i]_q$ 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. The classical **Gilbert-Varshamov bound** guarantees the existence of asymptotically good families. It has the drawback of not being constructive. The **Zyablov bound** [21] is weaker.

Theorem 1 (Zyablov bound). *Let q be a prime power. For every $0 \leq \mu \leq (q-1)/q$ the line L_μ in the $(\delta, R)-$plane connecting $(0, 1 - H_q(\mu)$ and $(\mu, 0)$ is asymptotically reachable. Here $H_q(x) = \frac{1}{log(q)}(H(x) + xlog(q-1))$ is the q-ary entropy function and $H(x) = H_2(x) = -xlog(x) - (1-x)log(1-x)$ is the binary entropy function. Logarithms are binary.*

The original proof of Theorem 1 (concatenation of a Reed-Solomon code as outer code and a code attaining the Gilbert-Varshamov bound as inner code) may be called semi-constructive. The first efficient construction of asymptotically nontrivial binary codes is due to Justesen [10]. It yields the lines L_μ where $\mu \leq H^{-1}(1/2) \sim 0.11$, hence a special case of the Zyablov bound. Shen [19] has shown that application of Justesen's method to a family of algebraic-geometric codes yields a constructive proof of the Zyablov bound in the binary case. A generalization to arbitrary q is in [4]. It should be noted that in order to reach values of δ close to $(q-1)/q$, these methods apply as ingredients codes derived from algebraic curves with many rational places and high genus. The efficiency of the code construction is therefore bounded by the complexity of the description of the underlying curves. The curves described in [19] and [6] are rather efficiently constructed. The result on algebraic curves, which is needed in the construction, may be formulated as follows:

Theorem 2. *Let p be a prime and $e \geq 1$ a rational number. Then 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 $\mathbb{F}_{p^m}$ with more than p^{em} rational points and genus g_m, such that $lim_{m \in \mathcal{M}} \frac{g_m}{p^{em}} = 0$. The codes derived via the standard mechanism of AG-codes have parameters*

$$[p^{em}, k, p^{em} - k - g_m]_{p^m}$$

3 A direct coding-theoretic construction

In [2] three direct constructions are presented, each of which results in $\epsilon-$biased $(n, k)_2-$arrays, where approximately $n = (k/\epsilon)^2$. An analysis of the so-called **Powering construction** reveals that it is based on Reed-Solomon codes and may be generalized as follows:

Theorem 3 (generalized powering construction). *Let $\mathcal{C}$ be a code $[n,k,d]_q$, where $q = p^m$. Let $v(1),\dots,v(k)$ be a basis of $\mathcal{C}$. Construct a $(qn,k)_p$-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 $tr : \mathbb{F}_q \longrightarrow \mathbb{F}_p$ is the trace. Then $A(\mathcal{C})$ has bias $\leq 1 - \frac{d}{n}$.

Proof. Let $\alpha = (\alpha_1,\dots,\alpha_k) \neq (0,\dots,0)$. 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

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

□

Apply Theorem 3 to the codes from Theorem 2:

Theorem 4. *For every prime p, every rational number $e \geq 1$ and $0 < \epsilon < 1$ there is a family of $(n_i,k_i)_p-$arrays, which are ϵ_i- biased such that $lim_i \epsilon_i = \epsilon$ and*

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

This produces families of $\epsilon-$biased $(n,k)_p-$arrays, where asymptotically $n = (\frac{k}{\epsilon})^a$, for every $a > 1$. Here $a = (e+1)/e$ in the terminology of Theorem 4. Moreover these constructions are explicit. The Reed-Solomon codes are obtained from the projective line ($g = 0$). They yield $e = 1$, hence $a = 2$. In characteristic 2 this is the original powering construction. The first layer in the construction of [6] are the Hermitian curves. These satisfy the assumptions of Theorem 2 for $e = 1.5$ and therefore yield $a = 5/3$. In order to demonstrate that constructions involving algebraic curves can be made efficient we give an explicit description of the Hermitian codes.

4 Hermitian codes for the user

We describe how to obtain generator matrices for the Hermitian codes. Consider the field extension $\mathbb{F}_{q^2} \mid \mathbb{F}_q$ and the corresponding trace tr and norm N, where $tr(x) = x + x^q, N(x) = x^{q+1}$. Our codes are defined over $\mathbb{F}_{q^2}$ and have length q^3.

The coordinates are parametrized by the pairs (α, β), where $N(\alpha) = tr(\beta)$. So we need to calculate traces and norms of all elements in the field and to list all these pairs in some order. There are precisely q^3 such pairs.

The general build-up: We construct a $(q^3, q^3)-$ matrix G with entries from $\mathbb{F}_{q^2}$. The first k rows of G generate the $k-$dimensional Hermitian code. It has parameters

$$[q^3, k, q^3 - k + 1 - g]_{q^2}, \text{ where } g = \binom{q}{2}.$$

The pole-order test: For $n = 0, 1, 2, \ldots$ we have to decide if n is a **pole-order** or not. If n is a pole-order we determine its **coordinate vector** (i, j). This is done as follows: Let r be the remainder of $n \bmod q$, where $0 \leq r \leq q - 1$ and $-s$ the (negative) remainder of $n \bmod q + 1$, where $0 \leq s \leq q$. Then n is a pole-order if and only if

$$x = \frac{n - r}{q} \geq \frac{n + s}{q + 1} = y.$$

If $n \geq 2g - 2$, then the pole-order test does not need to be performed. Every such number is a pole-order. If n is a pole-order, then $n = (q + 1)i + qj$, where $i = (x - y)q + r, j = s$. The coordinate vector of n is (i, j).

Constructing the rows of G **:** Let $u_1 = 0, u_2 = q, u_3 = q + 1 \dots$ be the first pole-orders. If u_k has coordinate-vector (i, j), then the entry of row k of G in coordinate (α, β) is $\beta^i \alpha^j$.

5 Block covering systems

Definition 4. *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*

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

This is a very natural combinatorial notion. Known sources for good BCS are graphs with extremal Eigenvalue properties (see [16] and [1]) and orthogonal arrays (the so-called **two-point based sampling**, see [7]). BCS may equivalently be described as bipartite graphs with certain expansion properties. In this language they are known as **bounded strong concentrators** (see [5]). The following is proved in [1].

Theorem 5. *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* $BCS(n, \Delta, n, \gamma, \frac{\gamma}{1-\gamma}(\frac{\lambda_1}{\Delta})^2)$.

6 A construction involving BCS

The construction of $\epsilon-$biased $(n, k)-$arrays as given in [16] proceeds in several steps. The first step yields a family of codes $[k^a, k, d]_2$, where d and $a \geq 1$ are constant. It is easy to see that $a > 2$ because of the use of a 2-dimensional Reed-Solomon code as one of the ingredients. In particular the family of codes is asymptotically trivial. An analysis of the remaining steps leads to the following natural generalization:

Theorem 6. *Assume the following exist:*

- *A linear code* $\mathcal{C}$ *with parameters* $[n, k, d]_p$.
- *A* $BCS(n, l, b, 1 - \frac{d}{n}, \epsilon)$ *for some* l.

- *An ϵ_0-biased $(n_0, l)_p-$array $\mathcal{A}$.*

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

Proof. At first we construct a p^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 $V(l,p)$, the vector space of dimension l over $\mathbb{F}_p$. Clearly $\tilde{\mathcal{C}}$ is linear over $\mathbb{F}_p$. 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 linear code obtained by concatenation with $\tilde{\mathcal{C}}$ as outer code and the code generated by the columns of $\mathcal{A}$ as inner code. Our end product is the transpose of a generator matrix of $\mathcal{C}'$. Clearly this is an $(n_0 b, k)_p-$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

$$bias(v') \leq \frac{\epsilon b n_0 + (1-\epsilon) b \epsilon_0 n_0}{b n_0} = \epsilon + \epsilon_0(1-\epsilon).$$

□

Apply Theorem 6 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 [14, 5]) on n vertices with valency Δ via Theorem 5. It has parameters $BCS(n, \Delta, n, 1-\delta, \frac{1-\delta}{\delta}\frac{4}{\Delta})$. The weakly biased arrays are derived from Theorem 4, an ϵ_0-biased $(n_0, \Delta)_p-$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 $(N, K)_p-$ array, where $N = O(K/\epsilon^{2(e+1)/e})$. We sum up:

Theorem 7. *Let p be a prime. For every $a > 1$ there are efficient constructions for infinite families of $\epsilon-$ biased $(n,k)_p-$arrays, where asymptotically $n \leq (k/\epsilon)^a$. For every $b > 2$ there are efficient constructions for infinite families of $\epsilon-$ biased $(n,k)_p-$arrays, where asymptotically $n = O(k/\epsilon^b)$.*

7 Weak bias and weak dependency

Definition 5. *Let $0 \leq \epsilon < 1$. An $(n,k)_p-$array is* t**-wise** $\epsilon-$**biased** *if every nontrivial linear combination of at most t of its columns has bias $\leq \epsilon$.*

Definition 6. *Let $0 \leq \epsilon < 1$. An $(n,k)_p-$array $\mathcal{A}$ is* t**-wise** ϵ **-dependent** *if for every set U of $s \leq t$ columns and every $a \in I\!F_p^s$ the frequency $\nu_U(a)$ of rows of $\mathcal{A}$, whose projection onto U equals a satisfies*

$$\mid \frac{\nu_U(a)}{n} - 1/p^s \mid \leq \epsilon.$$

The notion of a t-wise ϵ -dependent array generalizes the notion of an orthogonal array of strength t. An array is t-wise independent (=0-dependent) if and only if it is an orthogonal array of strength t.

The construction of t-wise $\epsilon-$biased arrays is essentially from [16].

Theorem 8. *Let the following be given:*

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

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

Proof. Let H be a check matrix of the code, hence an $(k,N)_p$-array each t columns of which are linearly independent. Our array is the product $\mathcal{A} = BH$. Denote by r_i the rows of B, by c_j the columns of $H, i = 1, \ldots, n; j = 1, \ldots, N$. The entries of BH are the dot products $r_i \cdot c_j$. Let $0 \neq \lambda = (\lambda_1, \ldots, \lambda_N)$ of weight $\leq t$. The corresponding linear combination of columns of $\mathcal{A}$ has entry $v_i = r_i \cdot (\sum_j \lambda_j c_j)$ in row i. The defining property of H shows that $\sum_j \lambda_j c_j \neq 0$. Vector $v = (v_i)$ is therefore a nontrivial linear combination of the columns of B, hence has bias $\leq \epsilon$. $\square$

8 Fourier transform

Let $G = I\!F_p^k$ be the elementary abelian group of order $g = p^k$. Consider the space $L^2(G)$ of complex-valued functions on G, with scalar product $<,>$, where

$$< f, g >= \frac{1}{g} \sum_{x \in G} f(x) \cdot \overline{g(x)}.$$

Here $x = (x_1, x_2, \ldots, x_k) \in G$. Let ζ be a primitive complex $p-$th root of unity. Every element $z = (z_1, z_2, \ldots, z_k)$ yields a **character** ϕ_z, where

$$\phi_z(x) = \zeta^{\sum x_i \cdot z_i} = \zeta^{x \cdot z}.$$

The characters of G form an orthonormal basis of $L^2(G)$.

Definition 7. *Let $f \in L^2(G)$. Express $f(x)$ in terms of the characters in the form*

$$f = \sum_{z \in G} \hat{f}(z) \phi_z.$$

The complex numbers $\hat{f}(z)$ are the **Fourier coefficients** *of f. The* **mass** *of f is defined as*

$$mass(f) = g < f, 1 >= g\hat{f}(0) = \sum_{x \in G} f(x).$$

Finally we define the **bias** *of $z \in G$ with respect to function f by*

$$bias_f(z) = \frac{g}{mass(\mid f \mid)} \mid \hat{f}(z) \mid$$

The orthonormality of the characters implies

$$\hat{f}(z) =< f, \phi_z >= \frac{1}{g} \sum_{x \in G} f(x) \zeta^{-x \cdot z}.$$

Definition 8. *Let $\mathcal{A}$ be an $(n, k)_p-$array with entries in $\mathbb{F}_p$. The characteristic function $f(x) = \chi_{\mathcal{A}}(x)$ counts with which multiplicity x occurs as a row of $\mathcal{A}$. We define the Fourier transform $\hat{\mathcal{A}}(z)$, the mass and the bias of $\mathcal{A}$ by identifying $\mathcal{A}$ with its characteristic function and applying Definition 7. Observe $mass(\mathcal{A}) = n$.*

Lemma 1. *Let $f = \chi_{\mathcal{A}}$ be the characteristic function of an $(n, k)_p-$array $\mathcal{A}$ and $z \in G = \mathbb{F}_p^k$. Then $bias_f(z)$ in the sense of Definition 7 is the same as $bias(v)$ in the sense of Definition 3, where v is the linear combination of the columns of $\mathcal{A}$ with z as vector of coefficients.*

This is easily checked. Together with Definition 7 it implies the following characterization of t-wise ϵ -biased arrays in terms of their Fourier transform.

Theorem 9. *An $(n, k)_p-$array $\mathcal{A}$ is t-wise ϵ -biased if and only if its Fourier coefficients $\hat{\mathcal{A}}(z)$ satisfy*

$$\mid \hat{\mathcal{A}}(z) \mid \leq \frac{n}{g} \epsilon \text{ for } 0 < wt(z) \leq t.$$

9 The relation between weak dependency and weak bias

Definition 9. *Let $\mathcal{A}$ be an $(n,k)_p-$array, $f \in L^2(G)$. The* **expectation** *$E_{\mathcal{A}}(f)$ with respect to $\mathcal{A}$ is defined as the statistical expected value of f when the rows of $\mathcal{A}$ are used as a probability space with uniform distribution:*

$$E_{\mathcal{A}}(f) = \sum_{x \in G} \frac{\chi_{\mathcal{A}}(x)}{n} f(x)$$

The following lemma was inspired by [9]:

Lemma 2. *Let $\mathcal{A}$ be an $(n,k)_p-$array. Then the following are equivalent:*

- *$\mathcal{A}$ is t-wise $\epsilon-$dependent.*
- *for every subset U of $l \leq t$ columns and every $f \in L^2((\mathbb{F}_p)^U)$ the following holds:*

$$\mid E_U(f) - mass(f)/p^l \mid \leq \epsilon \cdot mass(\mid f \mid).$$

  Here the expectation $E_U(f)$ is with respect to the restriction of $\mathcal{A}$ to the columns in U.

Proof. If f varies over the characteristic functions $\chi_a, a \in (\mathbb{F}_p)^U$, then the second condition expresses the definition of t-wise $\epsilon-$dependent arrays. Assume now that $\mathcal{A}$ is t-wise $\epsilon-$dependent, fix U and denote by $\nu_U(a)$ the multiplicity of a as a row in the restriction to U. Consider a general function $f = \sum_a c_a \chi_a$. Because of the linearity of the expectation and the mass, we get $\mid E_U(f) - \frac{mass(f)}{p^l} \mid = \mid \sum_a c_a (\frac{\nu_U(a)}{n} - \frac{1}{p^l}) \mid \leq \sum_a \mid c_a \mid \cdot \epsilon = \epsilon \cdot mass(\mid f \mid)$. $\square$

Lemma 3. *Let $\mathcal{A}$ be an $(n,k)_p-$array. Then the following are equivalent:*

- *$\mathcal{A}$ is t-wise $\epsilon-$biased.*
- *For every set U of $l \leq t$ columns and every $0 \neq z \in G$ with support in U the restriction g_z of the character ϕ_z to $\mathbb{F}_p^U$ satisfies*

$$\mid E_U(g_z) \mid \leq \epsilon$$

Proof. We have $\mid E(g_z) \mid = \mid \sum_{x\in F_p^U} \frac{\nu_U(x)}{n} \zeta^{-x\cdot z} \mid$. Comparison with Definition 2 shows that $\mid E(g_z) \mid$ is the bias of the linear combination of columns of $\mathcal{A}$ with $-z$ as set of coefficients. □

It can be proved, using Lemma 2 and Lemma 3, that every t-wise ϵ -dependent array is i-wise $p^i\epsilon$-biased for all $0 < i \leq t$. More important for our purposes is the other direction:

Theorem 10. *Let $\mathcal{A}$ be a $p-$ary array, which is i-wise δ_i-biased $(i = 1, 2, \dots, t)$. Then $\mathcal{A}$ is t-wise ϵ -dependent, where $\epsilon = (\sum_{i=1}^t \binom{t}{i}(p-1)^i\delta_i)/p^t$.*
In particular: t-wise $\epsilon-$biased implies t-wise $\epsilon'-$dependent for some $\epsilon' < \epsilon$.

Proof. Let $\mid U \mid = l \leq t, g_z$ as above, $h = p^l, a \in \mathbb{F}_p^U$. It is easy to check that $\chi_a = \frac{1}{h}\sum_z \zeta^{-a\cdot z} g_z$. By assumption $\mid E(g_z) \mid \leq \delta_{wt(z)}, z \neq 0$. Thus $\mid E(\chi_a) - \frac{1}{h} \mid = \frac{1}{h} \mid \sum_{z\neq 0} \zeta^{-a\cdot z} E(g_z) \mid \leq \frac{1}{h}\sum_{i=1}^l \binom{l}{i}(p-1)^i\delta_i$. The δ_i form by definition a non-decreasing series. It is easily seen that the last expression increases with l. This leads to the statement of the theorem. □

It follows from Theorem 10 that a t-wise $p^{-t}-$biased array has the property that the projection onto any set of t columns is onto. This property is being exploited for the testing of combinatorial circuits (see [16, 18]).

10 Authentication codes: an example

An $(N, m)_p-$array is $(\delta, t)-$**almost strongly universal** (short $(\delta, t)-$ASU) if for every set $U = U_0 \cup \{u\}$ of t columns and every $a' \in \mathbb{F}_p^{t-1}, x \in \mathbb{F}_p$ the frequencies $\nu_{U_0}(a')$ and $\nu_U(a', x)$ satisfy

$$\mid \nu_U(a', x)/\nu_{U_0}(a') \mid \leq \delta.$$

It was proved in [13] that a t-wise $\epsilon-$ dependent array is $(\delta, t)-$ASU, where $\delta = (p^{-t}+\epsilon)/(p^{-(t-1)}-\epsilon)$. Because of Theorem 10 we see that t-wise $\epsilon-$biased arrays yield authentication codes $(\delta, t)-$ASU. We start from $\epsilon-$biased arrays constructed out of Hermitian codes via Theorem 3. Consider the Hermitian code of dimension $k \sim \frac{1}{2}p^m$ defined over $\mathbb{F}_{p^m}$, where m is even (see Section 4). The parameters of the code are $[p^{3m/2}, \frac{1}{2}p^m, p^{3m/2} - p^m]$. Theorem 3 yields an $\epsilon-$ biased $(p^{5m/2}, \frac{1}{2}p^m)_p-$array, where $\epsilon = p^{-m/2}$. Apply Theorem 8 with the code $[p^{2i}, p^{2i}-3i, 4]_p$, obtained by the product construction (see [15]) from the affine part of the ovoid. Here $k \sim \frac{1}{2}p^m \sim 3i$. The result is a

$$\text{3-wise } p^{-m/2}- \text{ biased } (p^{5m/2}, p^{\frac{1}{3}p^m})_p - \text{array.}$$

In case $m=6$ we obtain a $(p^{15}, p^{\frac{1}{3}p^6})_p-$array, which is $(\delta, 3)-$ASU, where $\delta \sim 2/p$. This can be used to authenticate a source with $\frac{1}{3}p^6 log(p)$ bits with $15log(p)$ key bits, where the same authenticator is used for 3 consecutive messages. Choose for instance $p \sim 2^{41}$. The opponent's probability of success is limited by 2^{-40} in this case. We can authenticate a source of up to 2^{249} bits, using 605 key bits.

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,* Random Structures and Algorithms **3**(1992), 289-304, preliminary version: Symposium 31st FOCS 1990,544-553.

[3] J.Boyar, G.Brassard and R.Peralta: *Subquadratic zero-knowledge,* JACM **42**(1995),1169-1193.

[4] J.Bierbrauer and H.Schellwat: *Constructive asymptotic lower bounds on linear codes and weakly biased arrays,* submitted to IEEE Transaction on Information Theory.

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

[6] 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.

[7] 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.

[8] J.L.Carter,M.N.Wegman: *Universal Classes of Hash Functions,* J.Computer and System Sci. 18(1979), 143-154.

[9] B.Chor,O.Goldreich,J.Håstad,J.Friedman, S.Rudich and R.Smolensky: *The bit extraction problem or t-resilient functions,* Proc. 26th Symp. on Foundations of Computer Science (1985),396-407.

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

[11] N.Kahale : *Better expansion for Ramanujan graphs, Proc. 33rd FOCS 1992,* 398-404.

[12] N.Kahale : *Eigenvalues and expansion of regular graphs, Journal of the ACM* **42** (1995), 1091-1106.

[13] K.Kurosawa, T.Johansson and D.Stinson: *Almost k-wise independent sample spaces and their cryptologic applications, Lecture Notes in Computer Science* **1233**(1997),409-421 (Advances in Cryptology, Eurocrypt 97).

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

[15] A.C. Mukhopadhyay: *Lower bounds on $m_t(r,s)$, Journal of Combinatorial Theory A* **25**(1978),1-13.

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

[17] M.Naor and O.Reingold: *On the construction of pseudo-random permutations:Luby-Rackoff revisited,* Proceedings STOC 29(1997),189-199,
complete version under
http://theory.lcs.mit.edu/ tcryptol/homepage.html

[18] G.Seroussi and N.H.Bshouty: *Vector sets for exhaustive testing of logic circuits, IEEE Transaction on Information Theory* **34**(1988),513-522.

[19] 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.

[20] M.N.Wegman and J.L.Carter: *New Hash Functions and Their Use in Authentication and Set Equality,* J.Computer and System Sci. 22(1981),265-279.

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