Small-bias sets from extended norm-trace codes

Gretchen L. Matthews [*,†] Justin D. Peachey [‡]

March 15, 2012

Abstract

As demonstrated by Naor and Naor [11] among others [1, 2], the construction of small-bias probability spaces, or small-bias sets, is connected to that of error-correcting codes. Small-bias sets are probability spaces that in some sense approximate larger ones. Error-correcting codes have provided explicit constructions of such spaces. For instance, the concatenation of a Reed-Solomon code with a Hadamard code provides a now standard construction. Recently, Ben-Aroya and Ta-Shma used Hermitian codes to construct small-bias sets [4]. In this paper, we consider small-bias sets constructed from the extended norm-trace function field $\mathbb{F}_{q^r}(x,y)/\mathbb{F}_{q^r}$ defined by $Tr_{\mathbb{F}_{q^r}/\mathbb{F}_q}(y) = x^u$ where q is a power of a prime, $r \geq 2$, and $u | \frac{q^r-1}{q-1}$; here, $Tr_{\mathbb{F}_{q^r}/\mathbb{F}_q}$ denotes the trace with respect to the extension $\mathbb{F}_{q^r}/\mathbb{F}_q$. The Hermitian function field $y^q + y = x^{q+1}$, its quotient $y^q + y = x^u$ where $u|q+1$, and the norm-trace function field given by $Tr_{\mathbb{F}_{q^r}/\mathbb{F}_q}(y) = N_{\mathbb{F}_{q^r}/\mathbb{F}_q}(x)$ are special cases of the extended norm-trace function field. We detail the resulting small-bias sets.

1 Introduction and preliminaries

Consider a binary random variable $X := x_1, \ldots, x_k$. Let Ω denote the associated sample space. As shown by Varizani in 1986 [13], the bits $x_1, \ldots, x_k$ of X are independent and uniformly distributed if and only if for all nonempty $T \subseteq \{1, \ldots, k\}$,

$$Prob\left(\sum_{i \in T} x_i = 0\right) = Prob\left(\sum_{i \in T} x_i = 1\right)$$

where the sums are taken in $\mathbb{F}_2$, the finite field with two elements. Of course, if these equivalent conditions are satisfied, then $\Omega = \mathbb{F}_2^k$, the set of binary vectors of length k, with the uniform distribution.

[*] Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975 email: gmatthe@clemson.edu

[†] G. L. Matthews' work is supported in part by NSF DMS-090169.

[‡] Department of Mathematical Sciences, Clemson University, Clemson, SC 29634-0975 email: jpeache@clemson.edu

For a fixed k, it is useful in a number of applications to have a sample space that is smaller than $\mathbb{F}_2^k$ yet retains some of its randomness properties. These applications include derandomization of algorithms, testing of combinatorial circuits, and automated theorem proving [11]. This need for probability spaces that, in some sense, approximate larger ones prompted the notion of a small-bias set.

Definition 1.1. *A subset $X \subseteq \mathbb{F}_2^k$ is ϵ-biased if and only if for all nonempty $T \subseteq \{1, \ldots, k\}$,*

$$\frac{1}{|X|} \left| \sum_{x \in X} (-1)^{\sum_{i \in T} x_i} \right| \leq \epsilon.$$

Example 1.2.

1. *Fix a positive integer k. Then the set $\mathbb{F}_2^k$ is 0-biased whereas the set $\{v\}$, for any $v \in \mathbb{F}_2^k$, is 1-biased and is not ϵ-biased for any $\epsilon < 1$.*

2. *Let* $X = \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\} \subseteq \mathbb{F}_2^3$. *Then X is $\frac{1}{2}$-biased. To see this, consider a nonempty subset $T \subseteq \{1, 2, 3\}$, and let*

$$S_T = \frac{1}{|X|} \left| \sum_{x \in X} (-1)^{\sum_{i \in T} x_i} \right|.$$

Note that if $T \subseteq \{2, 3\}$, then $S_T = 0$. In addition,

$$S_{\{1\}} = \frac{1}{4} \left| (-1)^1 + (-1)^1 + (-1)^0 + (-1)^1 \right| = \frac{1}{2}.$$

More generally, it is easy to check that $1 \in T$ implies $S_T = \frac{1}{2}$. Thus, S is $\frac{1}{2}$-biased.

Given an ϵ-biased set X, ϵ provides a measure of how far from uniform the distribution associated with X is. To make this precise, let U_k denote the uniform distribution on a variable with k bits, and let

$$\Delta(X, Y) := \frac{1}{2} \sum_{\alpha \in \{0,1\}^k} |Prob\,[X = \alpha] - Prob\,[Y = \alpha]|$$

be the statistical difference between two k-bit random variables X and Y (equivalently, the statistical difference between their distributions).

Remark 1.3. *[8] Suppose $X \subseteq \mathbb{F}_2^k$ is an ϵ-biased set. Then*

$$\epsilon \leq 2\Delta(X, U_k) \leq 2^{\frac{k}{2}} \epsilon.$$

Certainly, a set is 0-biased if and only if the associated random variable is uniformly distributed.

While a random set of size $\mathcal{O}\left(\frac{k}{\epsilon^2}\right)$ is ϵ-biased [5], there is a need for explicit constructions of small-bias sets. The goal of this paper is to construct ϵ-biased sets $X \subseteq \mathbb{F}_2^k$ for fixed k and ϵ with $|X|$ small.

Our primary tool in the construction of small-bias sets is error-correcting codes. Thus, this section concludes with terminology and notation from coding theory. Section 2 contains a tutorial on the construction of small-bias sets from linear codes, focusing on algebraic geometric codes in particular. This is followed by Section 3 detailing the application of algebraic geometric codes from the extended norm-trace function field.

Notation. The set of positive integers is denoted $\mathbb{Z}^+$. Given a prime power q and a positive integer k, $\mathbb{F}_q$ denotes the field with q elements and $\mathbb{F}_q^k$ denotes the set of vectors of length k with coordinates in $\mathbb{F}_q$. As usual, given $v \in \mathbb{F}^k$, the i^{th} coordinate of v is denoted by v_i. The weight of a vector $v \in \mathbb{F}^k$ is $wt(v) = \Big| \{i : v_i \neq 0\} \Big|$. Given a matrix A, $Row_i A$ denotes the i^{th} row of A and $Col_j A$ denotes the j^{th} column of A.

A linear code over $\mathbb{F}_q$ of length n and dimension k is called an $[n,k]_q$ code. The Hamming distance between words $w, w' \in \mathbb{F}^n$ is $d(w,w') := |\{i : w_i \neq w_i'\}|$. A linear code over $\mathbb{F}_q$ of length n, dimension k, and minimum distance d (resp. at least d) is called an $[n,k,d]_q$ (resp. $[n,k,\geq d]_q$) code.

Let $F/\mathbb{F}_q$ be an algebraic function field of genus g. Given a divisor A on F defined over $\mathbb{F}_q$, let $\mathcal{L}(A)$ denote the set of rational functions f on X defined over $\mathbb{F}_q$ such that $(f)+A$ is an effective divisor together with the zero function. Let $\ell(A)$ denote the dimension of $\mathcal{L}(A)$ as an $\mathbb{F}_q$-vector space. An algebraic geometric (or AG) code $C_{\mathcal{L}}(D,G)$ can be constructed using divisors $D = \sum_{i=1}^n P_i$ and G on F where $P_1, \ldots, P_n$ are pairwise distinct places of F of degree one none of which are in the support of G. In particular,

$$C_{\mathcal{L}}(D,G) := \{(f(P_1), \ldots, f(P_n)) : f \in \mathcal{L}(G)\}.$$

If $\deg G < n$, then $C_{\mathcal{L}}(D,G)$ is an $[n, \ell(G), \geq n - \deg G]_q$ code. If $\{f_1, \ldots, f_k\}$ is a basis for $\mathcal{L}(G)$, then

$$\begin{bmatrix} f_1(P_1) & f_1(P_2) & \ldots & f_1(P_n) \\ f_2(P_1) & f_2(P_2) & \ldots & f_2(P_n) \\ \vdots & \vdots & & \vdots \\ f_k(P_1) & f_k(P_2) & \ldots & f_k(P_n) \end{bmatrix}$$

is a generator matrix for $C_{\mathcal{L}}(D,G)$. General references for AG codes include [9, 12].

2 Balanced codes and small-bias sets

In this section, we review the explicit construction of small-bias sets from balanced codes.

Definition 2.1. *An ϵ-balanced code is a binary code C of length n such that for all nonzero $c \in C$*

$$\frac{1-\epsilon}{2} \leq \frac{wt(c)}{n} \leq \frac{1+\epsilon}{2}.$$

The relationship between ϵ-balanced codes and ϵ-biased sets may be seen in the following lemma.

Lemma 2.2. *Suppose C is an $[n,k]_2$ code which is ϵ-balanced and M is a generator matrix for C. Then*

$$X = \{Col_1 M, Col_2 M, \dots, Col_n M\} \subseteq \mathbb{F}_2^k$$

is an ϵ-biased set with cardinality $|X| \leq n$.

Proof. Suppose C is an $[n,k]_2$ code which is ϵ-balanced, and let

$$X = \{Col_1 M, Col_2 M, \dots, Col_n M\}$$

be the set of columns of a generator matrix M of C. Given nonempty $T \subseteq \{1, \dots, k\}$, define $v \in \mathbb{F}_2^k$ by $v_i = 1$ if and only if $i \in T$. Then

$$\begin{aligned}
\tfrac{1}{|X|} \left| \textstyle\sum_{x \in X} (-1)^{\sum_{i \in T} x_i} \right| &= \tfrac{1}{n} \left| \textstyle\sum_{j=1}^{n} (-1)^{vCol_j M} \right| \\
&= \tfrac{1}{n} \left| n - 2\ wt(vM) \right| \\
&\leq \tfrac{1}{n} n\epsilon = \epsilon.
\end{aligned}$$

Therefore, X is an ϵ-biased set. $\square$

To obtain ϵ-balanced codes, we utilize a Walsh-Hadamard code. Given a positive integer s, the Walsh-Hadamard code C_s is a $[2^s, s]_2$ code with generator matrix

$$M' = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_{2^s} \\ | & | & & | \end{bmatrix}$$

where $\mathbb{F}_2^s = \{v_1, \dots, v_{2^s}\}$. It is well-known that C_s is a constant-weight code, and

$$wt(c) = 2^{s-1}$$

for all codewords $c \in C \setminus \{0\}$ [3]. The concatenation of an $[n, k, \geq d]_{2^s}$ code C' with C_s is an $\frac{n-d}{n}$-balanced code C of length $2^s n$. To see this, let $\varphi : \mathbb{F}_{2^s} \to \mathbb{F}_2^s$ be an isomorphism and $\phi_s : \mathbb{F}_2^s \to C_s$ be an encoding map for C_s. Suppose $c \in C \setminus \{0\}$. Then

$$c = \left(\phi_s\left(\varphi\left(c_1'\right)\right), \phi_s\left(\varphi\left(c_2'\right)\right), \dots, \phi_s\left(\varphi\left(c_n'\right)\right)\right)$$

for some nonzero codeword $c' \in C'$. Notice that

$$2^{s-1} d \leq 2^{s-1} wt(c') \leq 2^{s-1} n$$

since $wt\left(\phi_s\left(\varphi\left(c_i'\right)\right)\right) = 2^{s-1}$ for each nonzero coordinate c_i' of the codeword c' and $d \leq wt(c') \leq n$. Hence, the criteria in Definition 2.1 are satisfied, and C is an $[n2^s, sk, \geq 2^{s-1}d]_2$ code which is $\frac{n-d}{n}$-balanced. This observation paired with Lemma 2.2 yields the following result.

Proposition 2.3. *Given an $[n,k,d]_{2^s}$ code C, the set of columns of a generator matrix for the concatenation of C with the Walsh-Hadamard code C_s is an $\frac{n-d}{n}$-biased set $X \subseteq \mathbb{F}_2^{sk}$ with $|X| \leq n2^s$.*

Example 2.4. *Consider the $[2^s, k, 2^s - k + 1]_{2^s}$ Reed-Solomon code. According to Proposition 2.3, this results in a $\frac{k}{2^s}$-bias set $X \subseteq \mathbb{F}_2^k$ of cardinality $\Big|X\Big| \leq 2^{2s}$. This now standard construction first appeared in [2].*

Of course, one may apply Proposition 2.3 to AG codes over finite fields of characteristic 2. The motivation for doing so is that Hermitian codes have produced explicit small-bias sets which improve over previously known constructions in the range $k^{-1.5} \leq \epsilon \leq k^{-0.5}$. Moreover, the small-bias set given by an AG code $C_{\mathcal{L}}(D,G)$ may be described explicitly from the divisors G and $D = Q_1 + \cdots + Q_n$. For easy reference, we record here the corollary one obtains from Proposition 2.3 when C is an AG code over $\mathbb{F}_{2^s}$.

Corollary 2.5. *An AG code $C_{\mathcal{L}}\left(D,G\right)$ of length n over $\mathbb{F}_{2^s}$, with $\deg G < n$, gives rise to a $\frac{\deg G}{n}$-biased set $X \subseteq \mathbb{F}_2^{s\ell(G)}$ with $|X| \leq n2^s$.*

Proof. Fix an algebraic function field $F/\mathbb{F}_{2^s}$. Consider the AG code $C_{\mathcal{L}}\left(D,G\right)$ where G and $D := P_1 + \cdots + P_n$ are divisors on F with $P_i \notin \text{supp } G$ for all i and *deg* $G < n$. Then $C_{\mathcal{L}}\left(D,G\right)$ is an $[n, \ell(G), \geq n - \deg G]_{2^s}$ code, and the result follows from Proposition 2.3. □

To describe explicity the elements of the set X given in Corollary 2.5, let $\{f_1, \ldots, f_k\}$ be a basis for $\mathcal{L}(G)$, and let C be the concatenation of $C_{\mathcal{L}}(D,G)$ and C_s as described above. Fix a generator γ of $\mathbb{F}_{2^s}^* := \mathbb{F}_{2^s} \setminus \{0\}$. Let $\varphi : \mathbb{F}_{2^s} \to \mathbb{F}_2^s$ be the isomorphism given by $\varphi\left(\gamma^i\right) = Row_{i+1}M'$ for $0 \leq i \leq s-1$. Then a generator matrix M for the concatenated code C is

$$M = \begin{bmatrix} \phi_s\left(\varphi(f_1(Q_1))\right) & \phi_s\left(\varphi(f_1(Q_2))\right) & \cdots & \phi_s\left(\varphi(f_1(Q_n))\right) \\ \phi_s\left(\varphi(\gamma f_1(Q_1))\right) & \phi_s\left(\varphi(\gamma f_1(Q_2))\right) & \cdots & \phi_s\left(\varphi(\gamma f_1(Q_n))\right) \\ \vdots & \vdots & & \vdots \\ \phi_s\left(\varphi(\gamma^{s-1} f_1(Q_1))\right) & \phi_s\left(\varphi(\gamma^{s-1} f_1(Q_2))\right) & \cdots & \phi_s\left(\varphi(\gamma^{s-1} f_1(Q_n))\right) \\ \vdots & \vdots & & \vdots \\ \phi_s\left(\varphi(f_k(Q_1))\right) & \phi_s\left(\varphi(f_k(Q_2))\right) & \cdots & \phi_s\left(\varphi(f_k(Q_n))\right) \\ \phi_s\left(\varphi(\gamma f_k(Q_1))\right) & \phi_s\left(\varphi(\gamma f_k(Q_2))\right) & \cdots & \phi_s\left(\varphi(\gamma f_k(Q_n))\right) \\ \vdots & \vdots & & \vdots \\ \phi_s\left(\varphi(\gamma^{s-1} f_k(Q_1))\right) & \phi_s\left(\varphi(\gamma^{s-1} f_k(Q_2))\right) & \cdots & \phi_s\left(\varphi(\gamma^{s-1} f_k(Q_n))\right) \end{bmatrix}.$$

The elements of the small-bias set X given in Corollary 2.5 are the columns of the matrix M. Therefore,

$$X = \left\{ \begin{bmatrix} \phi_s\left(\varphi\left(f_1\left(Q_{\lceil \frac{j}{2^s} \rceil}\right)\right)\right)_{j-(\lceil \frac{j}{2^s} \rceil -1)2^s} \\ \phi_s\left(\varphi\left(\gamma f_1\left(Q_{\lceil \frac{j}{2^s} \rceil}\right)\right)\right)_{j-(\lceil \frac{j}{2^s} \rceil -1)2^s} \\ \vdots \\ \phi_s\left(\varphi\left(\gamma^{s-1} f_1\left(Q_{\lceil \frac{j}{2^s} \rceil}\right)\right)\right)_{j-(\lceil \frac{j}{2^s} \rceil -1)2^s} \\ \vdots \\ \phi_s\left(\varphi\left(f_k\left(Q_{\lceil \frac{j}{2^s} \rceil}\right)\right)\right)_{j-(\lceil \frac{j}{2^s} \rceil -1)2^s} \\ \phi_s\left(\varphi\left(\gamma f_k\left(Q_{\lceil \frac{j}{2^s} \rceil}\right)\right)\right)_{j-(\lceil \frac{j}{2^s} \rceil -1)2^s} \\ \vdots \\ \phi_s\left(\varphi\left(\gamma^{s-1} f_k\left(Q_{\lceil \frac{j}{2^s} \rceil}\right)\right)\right)_{j-(\lceil \frac{j}{2^s} \rceil -1)2^s} \end{bmatrix} : 1 \leq j \leq n2^s \right\} \subseteq \mathbb{F}_2^{sk}$$

is a $\frac{\deg G}{n}$-bias set with cardinality $|X| \leq n2^s$.

In the next section, we apply the construction in Corollary 2.5 to extended norm-trace codes. This is prompted by the fact that Hermitian codes, which are known to produce improved small-bias sets, are among the extended norm-trace codes. Because the family of extended norm-trace codes is larger, there is the opportunity to obtain new small-bias sets with known parameters.

3 Extended norm-trace codes and associated ϵ-biased sets

In this section, we consider a generalization of the Hermitian function field, associated AG codes, and resulting small-bias sets. The extended norm-trace function field is studied in [6, 7, 10]. While the Hermitian function field is defined over $\mathbb{F}_{q^2}$, the extended norm-trace function field may be defined over $\mathbb{F}_{q^r}$ for any $r \geq 2$. Hence, this broader family of function fields provides codes over a wider range of alphabets than the Hermitian function field as well as a larger class of small-bias sets.

Definition 3.1. *Let q be a power of a prime, $r \geq 2$, and x be transcendental over $\mathbb{F}_{q^r}$. The extended norm-trace function field over $\mathbb{F}_{q^r}$ is $\mathbb{F}_{q^r}(x, y)$ where*

$$y^{q^{r-1}} + y^{q^{r-2}} + \cdots + y = x^u$$

and $u > 1$ is a divisor of $\frac{q^r-1}{q-1}$.

Example 3.2. *1. If $u = \frac{q^r-1}{q-1}$, then $\mathbb{F}_{q^r}(x, y)/\mathbb{F}_{q^r}$ is the norm-trace function field defined by*

$$Tr_{\mathbb{F}_{q^r}/\mathbb{F}_q}(y) = N_{\mathbb{F}_{q^r}/\mathbb{F}_q}(x),$$

where $Tr_{\mathbb{F}_{q^r}/\mathbb{F}_q}(y)$ (resp., $N_{\mathbb{F}_{q^r}/\mathbb{F}_q}(x)$) denotes the trace of y (resp., norm of x) with respect to a degree-r extension of $\mathbb{F}_q$.

2. *If $r = 2$ and $u = \frac{q^2-1}{q-1} = q+1$, then $\mathbb{F}_{q^2}(x, y)$ is the well-studied Hermitian function field with defining equation $y^q + y = x^{q+1}$.*

3. *Taking $r = 2$ and $u \mid \frac{q^2-1}{q-1}$ yields the quotient of the Hermitian function field defined by $y^q + y = x^u$ over $\mathbb{F}_{q^2}$.*

The extended norm-trace function field $F/\mathbb{F}_{q^r}$ has genus $g = \frac{(u-1)\left(q^{r-1}-1\right)}{2}$ and exactly

$$q^{r-1}(uq - u + 1) + 1$$

places of degree one. Moreover, it was shown in [10] that the dimension of the divisor αP_∞, where $\alpha \in \mathbb{Z}^+$ and P_∞ denotes the infinite place of F is

$$\ell(\alpha P_\infty) = \sum_{i=0}^{u-1} \max\left\{\left\lfloor \frac{\alpha - iq^{r-1}}{u} \right\rfloor + 1, 0\right\}. \tag{1}$$

Consider the AG code $C_{\mathcal{L}}(D, \alpha P_\infty)$ over the extended norm-trace function field, where $D = Q_1 + \cdots + Q_{q^{r-1}(uq-u+1)}$ is the sum of all places of degree one other than P_∞ and $\alpha < q^{r-1}(uq - u + 1)$. Then $C_{\mathcal{L}}(D, \alpha P_\infty)$ is a $\left[q^{r-1}(uq-u+1), \sum_{i=0}^{u-1} \max\left\{\left\lfloor \frac{\alpha - iq^{r-1}}{u} \right\rfloor + 1, 0\right\}, \geq q^{r-1}(uq-u+1) - \alpha\right]_{q^r}$ code.

Taking q to be a power of 2 and applying Corollary 2.5 to the code above yields a small-bias set as detailed in the next result.

Theorem 3.3. *Let $q = 2^s$, $r \geq 2$, and $u \mid \frac{q^r-1}{q-1}$. For every positive integer $\alpha < q^{r-1}(uq - u + 1)$, there exists an ϵ-biased set $X \subseteq \mathbb{F}_2^{rs\sum_{i=0}^{u-1}\max\left\{\left\lfloor \frac{\alpha - iq^{r-1}}{u}\right\rfloor + 1, 0\right\}}$ with $|X| \leq q^{2r-1}(uq - u + 1)$ and $\epsilon = \frac{\alpha}{q^{r-1}(uq-u+1)}$.*

Example 3.4. *[4] Take F to be the Hermitian function field defined by $y^q + y = x^{q+1}$ over $\mathbb{F}_{q^2}$, where $q^2 = 2^s$. The Hermitian code $C_{\mathcal{L}}(D, \alpha P_\infty)$ gives rise to a $\frac{\alpha}{q^3}$-biased set $X \subseteq \mathbb{F}_2^{s\ell(\alpha P_\infty)}$ with $|X| \leq q^5$.*

A key difference in the small-bias sets given in Theorem 3.3 and those in Example 3.4 is the range of values of k allowed in each construction. Theorem 3.3 yields small-bias sets $X \subseteq \mathbb{F}_2^k$ where $k = r \log q \; \ell(G)$, given that $C_{\mathcal{L}}(D, \alpha P_\infty)$ is a code over $\mathbb{F}_{q^r}$ and q is a power of two. Recall that as one considers divisors $G = \alpha P_\infty$, $\alpha \in \mathbb{Z}^+$, $\ell(G)$ takes on all positive integer values in the interval $[1, n - g]$. Hence, the small-bias sets X constructed from Hermitian codes (meaning those in Example 3.4) have the property that $X \subseteq \mathbb{F}_2^k$ where k is an even multiple of $\log q$ whereas those given by the more general family of extended norm-trace codes in Theorem 3.3 allow for $k \in r \log q \, \mathbb{Z}$ where $r \geq 2$. The following example illustrates this more general situation.

Example 3.5. *Consider the function field $F := \mathbb{F}_8(x,y)/\mathbb{F}_8$ where*

$$y^4 + y^2 + y = x^7.$$

Let $G = 15P_\infty$, and let D be the sum of all places of F of degree one other than those in the support of G. Thus, $C_{\mathcal{L}}(D,G)$ has length 32.

By [10], a basis for $\mathcal{L}(G)$ is $\{1, x, x^2, x^3, y, y^2, xy, x^2y\}$. Thus, a generator matrix for $C_{\mathcal{L}}(D,G)$ is

$$M := \begin{bmatrix} 1(P_1) & 1(P_2) & \cdots & 1(P_{32}) \\ x(P_1) & x(P_2) & \cdots & x(P_{32}) \\ x^2(P_1) & x^2(P_2) & \cdots & x^2(P_{32}) \\ x^3(P_1) & x^3(P_2) & \cdots & x^3(P_{32}) \\ y(P_1) & y(P_2) & \cdots & y(P_{32}) \\ y^2(P_1) & y^2(P_2) & \cdots & y^2(P_{32}) \\ xy(P_1) & xy(P_2) & \cdots & xy(P_{32}) \\ x^2y(P_1) & x^2y(P_2) & \cdots & x^2y(P_{32}) \end{bmatrix}.$$

Using the above information, we can construct a small-bias set by concatenating $C_{\mathcal{L}}(D,G)$ with the appropriate Walsh-Hadamard code. Let $\mathbb{F}_8 = \mathbb{F}_2(\gamma)$ where γ is a root of $x^3 + x + 1$. Let M' be a generator matrix for the Walsh-Hadamard code C_3. Define the map α as follows:

$$\alpha : 1 \to Row_1 M', \gamma \to Row_2 M', \gamma^2 \to Row_3 M'.$$

The rows of the generator matrix for the concatenated code are the images under α applied to the entries of the following rows:

$$Row_1 M, \gamma Row_1 M, \gamma^2 Row_1 M, \ldots, Row_7 M, \gamma Row_7 M, \gamma^2 Row_7 M.$$

The columns of this generator matrix for the concatenated code are the elements of the associated small-bias set.

Every ϵ-biased set $X \subseteq \mathbb{F}_2^k$ satisfies $\left|X\right| \geq \Omega\left(\min\left\{\frac{k}{\epsilon^2 \log \frac{1}{\epsilon}}, 2^k\right\}\right)$ [2]. With this in mind, we fix k and ϵ and consider $|X|$ for the construction given in Theorem 3.3. As we are interested in finding sets whose size is as small as possible and are given a lower bound on the size of these sets, we now give an upper bound on $|X|$. Moreover, for the sake of comparison with the lower bounds, we make use of big-O notation. Also, note that utilizing these bounds, we may compare our result to previous results given in [4]. Notice, for example, that the small-bias set X given in Example 2.4 has $|X| = \mathcal{O}\left(\frac{k^2}{\epsilon^2}\right)$.

Theorem 3.6. *For all k and ϵ such that $\frac{\epsilon}{\left(\log \frac{1}{\epsilon}\right)^{\frac{1}{\sqrt{l}}}} \leq k^{\frac{-1}{\sqrt{l}}}$ for some integer $l \geq 4$, there exists an ϵ- biased set $X \subseteq \mathbb{F}_2^{\Omega(k)}$ with cardinality $|X| = \mathcal{O}\left(\left(\frac{k}{\epsilon^{l-\sqrt{l}} \log \frac{1}{\epsilon}}\right)^{\frac{l+1}{l}}\right)$.*

Proof. Fix k and ϵ so that $\frac{\epsilon}{\left(\log \frac{1}{\epsilon}\right)^{\frac{1}{\sqrt{l}}}} \leq k^{\frac{-1}{\sqrt{l}}}$ for some positive integer $l \geq 4$. Choose

$$q \in \left[\left(\frac{k}{\epsilon^{l-\sqrt{l}} \log \frac{1}{\epsilon}} \right)^{\frac{1}{l}}, 2 \left(\frac{k}{\epsilon^{l-\sqrt{l}} \log \frac{1}{\epsilon}} \right)^{\frac{1}{l}} \right]$$

to be a power of 2, say $q = 2^s$. Then

$$\begin{aligned} \tfrac{1}{q} &\geq \tfrac{1}{2} \left(\tfrac{\epsilon^{l-\sqrt{l}} \log \frac{1}{\epsilon}}{k} \right)^{\frac{1}{l}} \\ &= \tfrac{1}{2} \epsilon^{\frac{l-\sqrt{l}}{l}} \left(\tfrac{\log \frac{1}{\epsilon}}{k} \right)^{\frac{1}{l}} \\ &\geq \tfrac{1}{2} \epsilon^{\frac{l-\sqrt{l}}{l}} \epsilon^{\frac{1}{\sqrt{l}}} = \tfrac{1}{2} \epsilon. \end{aligned}$$

We also have that

$$\frac{1}{q} \leq \frac{\epsilon^{\frac{l-\sqrt{l}}{l}} \left(\log \frac{1}{\epsilon}\right)^{\frac{1}{l}}}{k^{\frac{1}{l}}} \leq \epsilon^{\frac{l-\sqrt{l}}{l}}$$

since $\left(\frac{\log \frac{1}{\epsilon}}{k} \right)^{\frac{1}{\sqrt{l}}} \leq 1$. Hence,

$$\left(\frac{1}{q} \right)^{\frac{l}{l-\sqrt{l}}} \leq \epsilon \leq \frac{2}{q},$$

and

$$\log \frac{q}{2} \leq \log \frac{1}{\epsilon} \leq \left(\frac{l}{l-\sqrt{l}} \right) \log q.$$

It follows that $\log \frac{1}{\epsilon} = \Theta\left(\log q\right)$.

Set $r = \lfloor \frac{l+2}{3} \rfloor$, and let $\alpha = \frac{\epsilon q^{2r-1}}{2}$. Consider the norm-trace function field over $F/\mathbb{F}_{q^r}$. We claim that the set X of columns of a generator matrix for $C_{\mathcal{L}}(D, \alpha P_\infty)$ is an ϵ-biased set with $X \subseteq \mathbb{F}_2^{\Omega(k)}$ and $|X| = \mathcal{O}\left(\left(\frac{k}{\epsilon^{l-\sqrt{l}} \log \frac{1}{\epsilon}} \right)^{\frac{l+1}{l}} \right)$.

First, we prove that $X \subseteq \mathbb{F}_2^{\Omega(k)}$. Let $u = \frac{q^r-1}{q-1}$, and set $m = \lfloor \frac{\alpha}{q^{r-1}} \rfloor$. As stated in Equation (1),

$$\ell\left(\alpha P_\infty\right) = \sum_{i=0}^{u-1} \max \left\{ \lfloor \frac{\alpha - i q^{r-1}}{u} \rfloor + 1, 0 \right\}$$

which gives

$$\ell\left(\alpha P_\infty\right) \geq \sum_{i=0}^{m} \frac{\alpha - i q^{r-1}}{u}$$

since $m \leq u - 1$. Simplifying, we see that

$$\ell\left(\alpha P_\infty\right) \geq \frac{\alpha}{u}(m+1) - \frac{q^{r-1}}{u} \frac{m(m+1)}{2} \geq \frac{1}{2} \frac{\alpha}{u}(m+1)$$

as $m \leq \frac{\alpha}{q^r - 1}$. It then follows that

$$\ell\left(\alpha P_\infty\right) \geq \frac{1}{2}\left(\frac{\alpha}{u}\right)^2 \geq \frac{1}{32} q^l \epsilon^2.$$

As a result,

$$\ell\left(\alpha P_\infty\right) \geq \frac{k}{32 \log \frac{1}{\epsilon}} \geq \frac{l-\sqrt{l}}{32l} \frac{k}{\log q},$$

and $\ell\left(\alpha P_\infty\right) \in \Omega\left(\frac{k}{\log q}\right)$. This implies $X \subseteq \mathbb{F}_2^{\Omega(k)}$.

Next, we note that X is $\frac{\epsilon}{2}$-biased as $\frac{\alpha}{n} = \frac{\epsilon}{2}$. Because $\epsilon > \frac{\epsilon}{2}$, X is certainly ϵ-biased by definition.

Finally, it follows from Theorem 3.3 that $\mid X \mid \leq q^{3r-1}$. Therefore, $|X| = \mathcal{O}\left(\left(\frac{k}{\epsilon^{l-\sqrt{l}} \log \frac{1}{\epsilon}}\right)^{\frac{l+1}{l}}\right)$. □

By taking $l = 4$ in the previous theorem, one may recover the following result due to Ben-Aroya and Ta-Shma. The result when $l \geq 5$ is not covered in [4] and thus is a contribution on this work.

Corollary 3.7. *[4] For all k and ϵ such that $\frac{\epsilon}{\sqrt{\log \frac{1}{\epsilon}}} \leq \frac{1}{\sqrt{k}}$, there exists an ϵ-biased set $X \subseteq \mathbb{F}_2^{\Omega(k)}$ with cardinality $|X| = \mathcal{O}\left(\left(\frac{k}{\epsilon^2 \log \frac{1}{\epsilon}}\right)^{\frac{5}{4}}\right)$.*

Theorem 3.6 is an extension of Corollary 3.7 in that it applies to a wider range of values of k and ϵ. To see this, let $l > 4$. If

$$\frac{\epsilon}{\log\left(\frac{1}{\epsilon}\right)^{\frac{1}{\sqrt{l}}}} \leq \frac{1}{k^{\frac{1}{\sqrt{l}}}}, \tag{2}$$

but

$$\frac{1}{k^{\frac{1}{2}}} < \frac{\epsilon}{\log\left(\frac{1}{\epsilon}\right)^{\frac{1}{2}}}, \tag{3}$$

then we may apply Theorem 3.6, but Corollary 3.7 does not apply. Hence, Theorem 3.6 allows for the construction of larger families of small-bias sets than previously identified. Specifically, if we fix ϵ and choose k so that

$$\epsilon^{\sqrt{l}} \leq \frac{\log\left(\frac{1}{\epsilon}\right)}{k} < \epsilon^2,$$

then (2) and (3) hold. The following example provides an instance of this.

Example 3.8. *Let $\epsilon = \frac{1}{4}$, $l = 9$, and $k = 33$. Then (2) and (3) hold as $\frac{1}{64} \leq \frac{2}{k} < \frac{1}{16}$. Applying Theorem 3.6 results in a small-bias set X with $|X| = O\left(67584^{\frac{10}{9}}\right)$.*

References

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

[2] N. Alon, O. Goldreich, J. Hastad, and R. Peralta, Simple constructions of almost k–wise independent random variables, Random Structures Algorithms **3** (1992), no. 3, 289-303.

[3] S. Arora and B. Barak, Computational Complexity: A Modern Approach, Cambridge University Press, 2009.

[4] A. Ben-Aroya and A. Ta-Shma, Constructing small-bias sets from algebraic-geometric codes, FOCS'2009, 191–197.

[5] E. Ben-Sasson, M. Sudan, S. Vadhan, and A. Wigerson, Randomness-efficient low degree tests and short PCPs via epsilon-biased sets, Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, 612–621, ACM, New York, 2003,

[6] M. Bras-Amorós and M. E. O'Sullivan, Duality for some families of correction capability optimized evaluation codes, Adv. Math. Commun. **2** (2008), no. 1, 15–33.

[7] O. Geil, On codes from norm-trace curves, Finite Fields Appl. **9** (2003), no. 3, 351–371.

[8] O. Goldreich, Computational Complexity: A Conceptual Perspective, Cambridge University Press, 2008.

[9] T. Høholdt, J. H. van Lint, and R. Pellikaan, Algebraic geometry codes, in Handbook of Coding Theory, V. Pless, W. C. Huffman, and R. A. Brualdi, Eds., **1**, Elsevier, Amsterdam (1998), 871–961.

[10] G. L. Matthews and J. D. Peachey, Explicit bases for Riemann-Roch spaces of the extended norm-trace function field, with applications to AG codes and Weierstrass semigroups, preprint.

[11] J. Naor and M. Naor, Small-bias probability spaces: efficient construction and applications, SIAM J. Comput. **22** (1993), 838–856.

[12] H. Stichtenoth, Algebraic Function Fields and Codes, Springer Verlag, 2009.

[13] U. V. Vazirani, Randomness, adversaries, and computation, Ph.D. thesis, EECS, UC Berkeley, 1986.