Construction of Pseudorandom Binary Sequences using Additive Characters over GF(2^k)

JÁNOS FOLLÁTH
Faculty of Computer Science, University of Debrecen
Alvégi út 15, H-5000 Szolnok, Hungary, E-mail: follathj@inf.unideb.hu

In a series of papers Mauduit and Sárközy (partly with coauthors) studied finite pseudorandom binary sequences and they constructed sequences with strong pseudorandom properties. In these constructions fields with prime order were used. In this paper a new construction is presented, which is based on finite fields of order 2^k.

Categories and Subject Descriptors: E.3 [**Data**]: Data Encryption; G.2.0 [**Mathematics of Computing**]: Discrete Mathematics—*General*; G.3 [**Mathematics of Computing**]: Probability and Statistics

General Terms: Security,Design

Additional Key Words and Phrases: Binary sequence, Character sums, Normality measure, Pseudorandom

1. INTRODUCTION

Mauduit and Sárközy studied pseudorandom binary sequences of

$$E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N$$

and introduced the following measures of pseudorandomness of binary sequences [Mauduit and Sárközy 1997]:

Definition 1. The well-distribution measure of E_N is defined as

$$W(E_N) = \max_{a,b,t} \left| \sum_{j=0}^{t-1} e_{a+jb} \right|,$$

where the maximum is taken over all a, b, t with $a, b, t \in \mathbb{N}, 1 \le a + b \le a + tb \le N$.

Definition 2. The correlation measure of order k of E_N is:

$$C_k(E_N) = \max_{M,D} \left| \sum_{n=1}^{M} e_{n+d_1} e_{n+d_2} \dots e_{n+d_k} \right|,$$

where the maximum is taken over all $D = (d_1, \dots, d_k)$ ($d_1 < \dots < d_k$ are non-negative integers) and $M \in \mathbb{N}$ with $M + d_k \le N$.

Definition 3. Combined (well-distribution-correlation) PR-measure of order k E_N is defined as:

$$Q_k(E_N) = \max_{a,b,t,D} |Z(a, b, t, D)|$$

Mathematics subject classification number: 68P25, 11T23.

where

$$Z(a,b,t,D)=\left|\sum_{j=0}^{t} e_{a+jb+d_1}e_{a+jb+d_2}\dots e_{a+jb+d_k}\right|. \tag{1.1}$$

Definition 4. Combined PR-measure of E_N:

$$Q(E_N)=\max_{k\le \log N} Q_k(E_N).$$

A pseudo-random sequence is considered good if both $W(E_N)$ and $C_k(E_N)$ (at least for small k) are small in terms of N: both of them are $o(N)$ as $N\to\infty$. Later Cassaigne, Mauduit and Sárközy [Cassaigne et al. 2002] showed that this terminology is justified since for almost all $E_N\in\{-1,+1\}^N$, both $W(E_N)$ and $C_k(E_N)$ are less than $N^{1/2}(\log N)^c$. Moreover, in [Mauduit and Sárközy 1997] was shown that $\left\{\left(\frac{n}{p}\right)\right\}$, $0<n<p$, where $\left(\frac{n}{p}\right)$ is the Legendre symbol, forms a good pseudorandom sequence. Later in [Goubin et al. 2004] and [Gyarmati et al. 2005] this construction was extended to a large family of good pseudorandom sequences. The later construction and its pseudorandom measures are described in the following theorem (proved in [Gyarmati et al. 2005]):

Theorem 1. *Let p be an odd prime, $\lambda\in\mathbb{F}_p^*$ be of multiplicative order T and $f(x)\in\mathbb{F}_p[x]$ be of degree k and not of the form $cx^\alpha(g(x))^2$ with $c\in\mathbb{F}_p,\alpha\in\mathbb{N},g(x)\in\mathbb{F}_p[x]$. Define the sequence $E_T=\{e_1,\dots,e_T\}$ by*

$$e_n=\begin{cases}\left(\frac{f(\lambda^n)}{p}\right) & if\ \ p\nmid f(\lambda^n),\\ 1 & if\ \ p\mid f(\lambda^n).\end{cases}$$

Then we have

$$W(E_T)<5kp^{1/2}\log p$$

Moreover, assume that also $l\in\mathbb{N}$, T is a prime, and either $\min\{(4k)^l,(4l)^k\}\le T$ or 2 is a primitive root modulo T. Then we also have

$$C_l(E_T)\le 5klp^{1/2}\log p.$$

There were further good sequences constructed in [Sárközy 2001]. These sequences are based on the notion of index and multiplicative characters play a crucial role in the proofs. Many other sequences were studied, but the upper bounds obtained for $W(E_N)$ and $C_l(E_N)$ were much weaker, than for the constructions mentioned. One of these weaker constructions is still of particular interest: in [Mauduit et al. 2004] in the proof the properties of additive characters were utilized.The construction and its properties are summarized in the following theorem:

Theorem 2. *Let p an odd prime, $f(x)\in\mathbb{F}_p[x]$ of degree d, and define $E_p=\{e_1,\dots,e_p\}$ by*

$$e_n=\begin{cases}+1 & if\ \ 0\le r_p(f(n))<p/2,\\ -1 & if\ \ p/2\le r_p(f(n))<p\end{cases}$$

where $r_p(n)$ *denotes the unique* $r \in \{0, \dots, p-1\}$ *such that* $n \equiv r \pmod p$. *Then we have*

$$W(E_p) \ll dp^{1/2}(\log p)^2.$$

For $2 \le l \le d-1$ *we also have*

$$C_l(E_p) \ll dp^{1/2}(\log p)^{l+1}.$$

All the above mentioned constructions are using prime fields of odd characteristic. It is natural idea to try to give further good sequences by using fields with characteristic 2. The application of fields of characteristic 2 enables one to use additive characters in a more direct way than in Theorem 2.

The generator presented in [Goubin et al. 2004] not only has good pseudorandom measures, but also possesses the strict avalanche property [Tóth 2007]. There are results about its family complexity [Ahlswede et al. 2003], the computational complexity of the best known attacks and an extremly fast implementation [Hoffstein and Lieman 2001]. Furthermore the bound on its higher order correlation measure enables one to estimate its linear complexity profile [Brandstätter and Winterhof 2006]. Although its security cannot be proven by reduction, the above enumerated arguments make the generator a good candidate for cryptographic use. The main result of this paper and the first step in a search for pseudorandom sequence families of similar good quality as the above mentioned construction is to consider the pseudorandom measures of binary sequences based on additive characters over $\mathbb{F}_q$.

2. THE PSEUDORANDOMNESS OF THE TRACE

In this paper a new pseudorandom sequence generator is proposed, whose pseudorandom measures are better than of the top quality constructions (like in Theorem 1).

THEOREM 3. *Let* $\mathbb{F}_q$ *be a finite field of characteristic two and its multiplicative group of prime order. Let* χ *be a non principal additive character, and* α *a primitive element of* $\mathbb{F}_q$ *and let* $f(x) \in \mathbb{F}_q[x]$ *of odd degree* d *and let* I *be the set of exponents in the terms with nonzero coefficient in* $f(x)$. *For the minimal polynomial of* α^i *over* $\mathbb{F}_2$ *write* $m_i(x)$. *Let*

$$E_{q-1} = \{\chi(f(\alpha^1)), \chi(f(\alpha^2)), \dots, \chi(f(\alpha^{q-1}))\} \in \{-1, +1\}^{q-1},$$

Let $D' \subseteq \{1, \dots, q-1\}$ *such that* $\prod_{i \in I} m_i(x)$ *does not divide the polynomial* $d(x) = \sum_{d_i \in D'} x^{d_i}$, *then*

$$\max_{a,b,t,D'} |Z(a,b,t,D')| \le 9dq^{1/2} \log q. \tag{2.1}$$

Let $D \subseteq \{1, \dots, q-1\}$ *such that* $\prod_{i \in I} m_i(x)$ *divides the polynomial* $d(x) = \sum_{d_i \in D} x^{d_i}$, *then*

$$\max_{a,b,t,D} |Z(a,b,t,D)| = \max_{D}(q - 1 - \max_{d_i \in D} d_i). \tag{2.2}$$

Remark 1. In the $D = \{1, \dots, k\}$ case, with $|I| > k$ the divisibility property of the first part and (2.1) hold.

THEOREM 4. *Let $\mathbb{F}_q$ be a finite field of characteristic two and its multiplicative group of prime order. Let χ be a non principal additive character, and α a primitive element of $\mathbb{F}_q$ and let $f(x) \in \mathbb{F}_q[x]$ of odd degree $d \geq \log q$ and let the coefficients of its terms be zero if and only if the term has an even exponent. If*

$$E_{q-1} = \{\chi(f(\alpha^1)), \chi(f(\alpha^2)), \ldots, \chi(f(\alpha^{q-1}))\} \in \{-1, +1\}^{q-1},$$

then :

$$Q(E_N) \leq 9dq^{1/2} \log q. \tag{2.3}$$

COROLLARY 1. *Obviously we have:*

$$W(E_N) \leq 9dq^{1/2} \log q$$

furthermore for $l \leq d+1$

$$C_l(E_N) \leq 9dq^{1/2} \log q.$$

3. CHARACTER SUMS

To give an upper bound for incomplete character sums we will use the following inequality:

LEMMA 1. *If $m \in \mathbb{N}$, the function $g(x) : \mathbb{Z} \to \mathbb{C}$ is periodic with period m, and X, Y are real numbers with $Y > 0$ then*

$$\left| \sum_{X < n \leq X+Y} g(n) \right| \leq \frac{Y+1}{m} \left| \sum_{n=1}^{m} g(n) \right| + \sum_{1 \leq |h| \leq m/2} |h|^{-1} \left| \sum_{n=1}^{m} g(n) e\left(\frac{hn}{m}\right) \right|.$$

PROOF. This is implicit in [Vinogradov 2003], and it is presented in this explicit form in [Tietäväinen 1988] and [Friedlander and Iwaniec 1993] and it is also related to the Erdős-Turán inequality. □

To complete the proof we will also have to give an upper bound for the following character sum:

LEMMA 2. *Let χ be a non-principal additive character and α a primitive element of $\mathbb{F}_q$. Let $h \in \mathbb{Z}$, $h \not\equiv 0 \bmod (q-1)$ and let $f(x) \in \mathbb{F}_q[x]$ with its degree d being odd. Then*

$$\left| \sum_{n=1}^{q-1} \chi(f(\alpha^n)) e\left(\frac{hn}{q-1}\right) \right| \leq dq^{1/2}.$$

PROOF. This is an immediate consequence of Theorem 2G in [Schmidt 1976]. □

LEMMA 3. *Suppose that χ is a non-principal additive character and α a primitive element in $\mathbb{F}_q$, furthermore let $f(x) \in \mathbb{F}_q[x]$ be of odd degree d. Let X, Y be real numbers with $0 \leq X < X+Y \leq q-1$. Then*

$$\left| \sum_{X < n \leq X+Y} \chi(f(\alpha^n)) \right| < 9dq^{1/2} \log q.$$

PROOF. Applying Lemma 1 with $m = q-1$ and $g(x) = \chi(f(\alpha^x))$:

$$\left|\sum_{X<n\le X+Y}\chi(f(\alpha^n))\right| \le \frac{Y+1}{q-1}\left|\sum_{n=1}^{q-1}\chi(f(\alpha^n))\right| + \sum_{1\le|h|\le\frac{q-1}{2}}|h|^{-1}\left|\sum_{n=1}^{q-1}\chi(f(\alpha^n))e\left(\frac{hn}{q-1}\right)\right|.$$

By applying Lemma 2 and Weil's theorem (Theorem 5.38 in [Lidl and Niederreiter 1997]) we obtain:

$$\begin{aligned}\left|\sum_{X<n\le X+Y}\chi(f(\alpha^n))\right| &< 2dq^{1/2} + 2\sum_{1\le h\le\frac{q-1}{2}}|h|^{-1}dq^{1/2}\\ &< 2dq^{1/2}(1+(1+\log\left(\frac{q-1}{2}\right))) < 2dq^{1/2}(2+\log q)\\ &\le 2dq^{1/2}\left(\frac{\log q}{\log 2}+\log q\right) < 9dq^{1/2}\log q\end{aligned}$$

and this completes the proof. □

4. PROOF OF THEOREM 3

Let $Z(a,b,t,D)$ defined by (1.1), for $k < q-1$ we have

$$|Z(a,b,t,D)| = \left|\sum_{n=0}^{t}\chi(f(\alpha^{a+nb+d_1}))\chi(f(\alpha^{a+nb+d_2}))\dots\chi(f(\alpha^{a+nb+d_k}))\right|$$

for all $a, b, t, D = (d_1, \dots, d_k)$ such that

$$a + nb + d_l \in \{1, \dots, q-1\} \text{ for } n = 0, 1, \dots, t \text{ and } l = 1, \dots, k. \tag{4.1}$$

Suppose that $f(x) = \sum_{i=0}^{d} a_i x^i$. Write $a_{ij} = a_i\alpha^{i(a+d_j)}$ and $f_j(x) = \sum_{i=0}^{d} a_{ij}x^i$. Then

$$|Z(a,b,t,D)| = \left|\sum_{n=0}^{t}\chi(f_1(\alpha^{nb}))\chi(f_2(\alpha^{nb}))\dots\chi(f_k(\alpha^{nb}))\right| = \left|\sum_{n=0}^{t}\chi(\tilde{f}(\alpha^{nb}))\right|$$

where $\tilde{f}(x) = \sum_{i=0}^{d} a'_i x^i$, $a'_i = \sum_{j=0}^{k} a_{ij} = \sum_{j=0}^{k} a_i(\alpha^i)^{a+d_j}$. Let $\tilde{d}(x) = x^a d(x) = \sum_{j=0}^{k} x^{a+d_j}$ now $a'_i = a_i\tilde{d}(\alpha^i)$. $\tilde{d}(x)$ can be treated as a polynomial over $\mathbb{F}_2$ and so follows, that $\tilde{d}(\alpha^i)$ is zero if and only if the minimal polynomial $m_i(x)$ of α^i divides $\tilde{d}(x)$. Since $m_i(x) \neq x$ and is irreducible it divides $\tilde{d}(x)$ if and only if it also divides $d(x) = \sum_{j=0}^{k} x^{d_j}$. Consequently $\tilde{f}(x)$ is zero polynomial if and only if $\prod_{i:a_i\neq 0} m_i(x)$ divides $d(x)$. In this case we obtain

$$|Z(a,b,t,D)| = \left|\sum_{n=0}^{t}\chi(\tilde{f}(\alpha^{nb}))\right| = t$$

and since (4.1) $t \le q-1-\max_{1\le l\le k} d_l$ this proves (2.2).

If $\prod_{i:a_i\neq 0} m_i(x)$ does not divide $d(x)$, then with the notation $\beta=\alpha^b$ we obtain:

$$|Z(a,b,t,D)|=\left|\sum_{n=0}^{t}\chi(\tilde{f}(\beta^n))\right|.$$

Since $\mathbb{F}_q^*$ is of prime order, β is a primitive element of $\mathbb{F}_q$ if and only if $b\neq q-1$. If $b=q-1$ then $b=1,\ a=d_1=0$, thus $|Z(a,b,t,D)|=|\chi(f(\alpha^{a+nb+d_1}))|=1$ and the theorem is proved. Otherwise we can apply Lemma 3, which implies

$$|Z(a,b,t,D)|=\left|\sum_{n=0}^{t}\chi(\tilde{f}(\beta^n))\right|\leq 9dq^{1/2}\log q$$

and this proves (2.1) and completes the proof of Theorem 3.

5. PROOF OF THEOREM 4

Put $g(x)=l.c.m.\{m_1(x),m_3(x),\ldots,m_d(x)\}$. Then $g(x)$ is a generator polynomial of a BCH code C over $\mathbb{F}_q$. Then by the BCH bound (Theorem 8 on page 201 in [MacWilliams and Sloane 1977]) the minimum distance of C is at least $d+1$. Let $M(x)=\prod_{i=0}^{\frac{d-1}{2}} m_{2i+1}(x)$. Since $g(x)|M(x)$, $M(x)$ divides $d(x)\in\mathbb{F}_2[x]$ only if $d(x)$ is a codeword. It follows that $M(x)$ can only divide polynomials whose weight is greater than d. Now applying Theorem 3 completes the proof.

Remark 2. The proof can be applied with a slightly modified constrains on $f(x)=\sum_{i=0}^{d}a_ix^i$. If there is an integer a such that for each $j=a,\ldots,a+\log q-1$ exists a α^{c_j} conjugate of α^j such that $a_{c_j}\neq 0$ then (2.3) continues to hold.

6. PARAMETER SELECTION

The selection of the field is very restricted by Theorem 3: $q-1$ must be a Mersenne-prime. This means only a few candidates to a practical application: we know presently 44 Mersenne primes ([Me]). Therefore by this pseudorandom generator construction one should take a fixed field, best fitting to the application, and define the sequence family by the other parameters. There are three basic approaches to accomplish this:

(1) Variable χ with fixed α and $f(x)$
(2) Variable α with fixed χ and $f(x)$
(3) Variable $f(x)$ with fixed α and χ

In the first two cases presently one should use field size corresponding to the 12th Mersenne prime ($k=127$) or greater ($k=521$ and $k=607$ are the other two reasonable candidates) to resist brute force attacks. The third case offers much more versatility and a possibility to study the properties of the pseudorandom sequence family with more sophisticated methods. This question will be considered in a subsequent paper.

7. CONCLUSIONS

These sequences can be computed fast: using normal basis representation and the algorithm described in [Reyhani-Masoleh and Hasan 2003] this construction

is asymptotically faster than the previous constructions mentioned in the Introduction. Although the pseudorandom measures of the generator are even better than of the previous top quality generators, it has the same drawback as the other additive character based generator ([Mauduit et al. 2004]). This means that the correlations of large order can be large (remind that even by the other constructions mentioned, the upper bound to the correlation measure becomes trivial above order $\frac{q^{1/2}}{c \log q}$). This generator is a reasonable option for situations where computation time is more important than the control over the large order correlations.

REFERENCES

Mersenne primes: History, theorems and lists. http://primes.utm.edu/mersenne/.

Ahlswede, R., Khachatrian, L., Mauduit, C., and Sárközy, A. 2003. A complexity measure for families of binary sequences. *Period. Math. Hungar. 46,* 2 (June), 107–118.

Brandstätter, N. and Winterhof, A. 2006. Linear complexity profile of binary sequences with small correlation measure. *Period. Math. Hungar. 52,* 2 (June), 1–8.

Cassaigne, J., Mauduit, C., and Sárközy, A. 2002. On finite pseudorandom binary sequences VII: The measures of pseudorandomness. *Acta Arith. 103,* 2, 97–118.

Friedlander, J. and Iwaniec, H. 1993. Estimates for character sums. *Proc. Amer. Math. Soc. 119*, 365–372.

Goubin, L., Mauduit, C., and Sárközy, A. 2004. Construction of large families of pseudorandom binary sequences. *Journal of Number Theory 106,* 1, 56–69.

Gyarmati, K., Sárközy, A., and Pethő, A. 2005. On linear recursion and pseudorandomness. *Acta Arith. 118,* 4, 359–374.

Hoffstein, J. and Lieman, D. 2001. *Cryptography and Computational Number Theory*. Progress in Computer Science and Applied Logic, vol. 20. Birkhäuser Verlag, Chapter The Distribution of the Quadratic Symbol in Function Fields and a Faster Mathematical Stream Cipher, 59–68.

Lidl, R. and Niederreiter, H. 1997. *Finite Fields*. Encyclopedia of Mathematics, vol. 20. Cambridge University Press.

MacWilliams, F. J. C. and Sloane, N. J. A. 1977. *The Theory of Error-Correcting Codes*. North-Holland Mathematical Library, vol. 16. North-Holland Publishing Company.

Mauduit, C., Rivat, J., and Sárközy, A. 2004. Construction of pseudorandom binary sequences using additive characters. *Monatsh. Math. 141,* 3, 197–208.

Mauduit, C. and Sárközy, A. 1997. On finite pseudorandom binary sequences I: Measure of pseudorandomness, the Legendre symbol. *Acta Arith. 82,* 4, 365–377.

Reyhani-Masoleh, A. and Hasan, A. 2003. Fast normal basis multiplication general purpose processors. *IEE transactions on computers 52,* 11, 1379–1390.

Sárközy, A. 2001. A finite pseudorandom binary sequence. *Studia Sci. Math. Hungar. 38*, 377–384.

Schmidt, W. 1976. *Equations over Finite Fields. An Elementary Approach*. Lecture Notes in Mathematics, vol. 536. Springer-Verlag.

Tietäväinen. 1988. *Algebra, some current trends*. Lecture Notes in Math., vol. 1352. Springer Verlag, New York, Chapter Incomplete sums and two applications of Deligne's result, 190–205.

Tóth, V. 2007. Collision and avalanche effect in families of pseudorandom binary sequences. *Period. Math. Hungar. 55,* 2 (November), 185–196.

Vinogradov, I. M. 2003. *Elements of Number Theory*. Dover Publications.