INTEGERS: ELECTRONIC JOURNAL OF COMBINATORIAL NUMBER THEORY 8 (2008), #A31

# ON THE DISTRIBUTION OF THE ELLIPTIC SUBSET SUM GENERATOR OF PSEUDORANDOM NUMBERS

**Edwin D. El-Mahassni**
*Department of Computing, Macquarie University, North Ryde, NSW 2109, Australia*
edwinelm@ics.mq.edu.au

*Received: 7/26/07, Revised: 6/23/08, Accepted: 7/1/07, Published: 7/18/08*

**Abstract**

We show that for almost all choices of parameters, the elliptic subset sum pseudorandom number generator produces a sequence of uniformly distributed pseudorandom numbers. The result is useful for both cryptographic and Quasi Monte Carlo applications and relies on bounds of exponential sums.

## 1. Introduction

Let $p \geq 1$ be a prime. Let $\mathbf{E}$ be an elliptic curve over $\mathbb{F}_p$, given by an affine *Weierstrass equation* of the form

$$y^2 + (a_1x + a_3)y = x^3 + a_2x^2 + a_4x + a_6;$$

see [1, 10]. Assume now that $p \geq 5$, and so the Weierstrass equation simplifies to

$$Y^2 = X^3 + aX + b, \quad a, b \in \mathbb{F}_p$$

where we also request that $4a^3 + 27b^2 \neq 0$.

It is known (see [1, 10]) that the set $\mathbf{E}(\mathbb{F}_p)$ of $\mathbb{F}_p$-rational points of $\mathbf{E}$ forms an Abelian group under an appropriate composition rule (which we denote by $\oplus$) and with the point at infinity $\mathcal{O}$ as the neutral element. We also recall the Hasse bound

$$|\#\mathbf{E}(\mathbb{F}_p) - p - 1| \leq 2p^{1/2},$$

where $\#\mathbf{E}(\mathbb{F}_p)$ is the number of $\mathbb{F}_p$-rational points, including the point at infinity. However, in this paper, we will only concentrate on $\mathbf{E}(\mathbb{F}_p)$, the set of $\mathbb{F}_p$ rational points on $\mathbf{E}$. Further, for a point $P \in \mathbf{E}(\mathbb{F}_p)$, with $P \neq \mathcal{O}$, we write $x(P)$ and $y(P)$ for its affine coordinates.

Next, we proceed to describe the *elliptic subset sum generator*. Let $(u(n))$ be a linear recurrence sequence defined over $\mathbb{F}_2$. For a prime integer $p$, the elliptic subset sum

generator of pseudorandom numbers, is defined as follows. Given an $r$-dimensional vector $\mathbf{Z} = (Z_1, \ldots, Z_r) \in (\mathbf{E}(\mathbb{F}_p))^r$ of points one can consider the sequence

$$V_{\mathbf{Z}}(n) = \sum_{i=1}^{r} u(n+i-1)Z_i, \quad n = 1, \ldots, N \le \tau \le p$$

of elements of $\mathbf{E}(\mathbb{F}_p)$, where the sequence $u(n)$ has period $\tau$.

Alternatively, when this generator is defined over a residue ring modulo $m$, the subset sum generator is also known as the *knapsack generator*. It was originally introduced in [9], and studied in [7], see also [6, Section 6.3.2], and [8, Section 3.7.9]. More recently, a bound on the multidimensional distribution of the subset sum generator of pseudorandom numbers has been given in [2].

A natural analogue of [2] is to study the distribution of $x(V_{\mathbf{Z}}(n))$. Here, using some previous results of functions on elliptic curves, we show for the one-dimensional case, that for almost all choices of points $\mathbf{Z} = (Z_1, \ldots, Z_r) \in (\mathbf{E}(\mathbb{F}_p))^r$, the vector $(x(V_{\mathbf{Z}}(n)))$ is uniformly and independently distributed. In fact we use the classical number-theoretic notion of *discrepancy* to give a quantitative form of this property. It can also be reformulated in terms of the $\varepsilon$-bias of the most significant bits of the elements of the generating sequences, which is more common in cryptographic literature.

Our method is based on some simple bounds on exponential sums and the famous *Erdös-Turán* inequality (see Lemma 2 below) which relates the deviation from uniformity of distribution, that is, discrepancy, and the corresponding exponential sums.

## 2. Preliminaries

Here we present several necessary technical tools.

We say that the linear recurrence sequence $u(n)$ of elements of $\mathbb{F}_2$ is of order $r \ge 1$ with characteristic polynomial

$$g(T) = T^r + c_{r-1}T^{r-1} + \ldots + c_1T_0 + c_0 \in \mathbb{F}_2[T]$$

if

$$u(n+r) + c_{r-1}u(n+r-1) + \ldots + c_1u(n+1) + c_0u(n) = 0, \quad n = 1, 2, \ldots,$$

and it does not satisfy any shorter linear recurrence relation, see [5], Chapter 8.

It is easy to see that the set of all sequences with the same characteristic polynomial $g$ form a linear space $\mathcal{L}(g)$ over $\mathbb{F}_2$.

We also need the following property of sequences from $\mathcal{L}(g)$ with irreducible $g$ which is essentially [5], Theorem 8.28.

**Lemma 1.** *If $g \in \mathbb{F}_2[T]$ is irreducible over $\mathbb{F}_2$ then all nonzero sequences from $\mathcal{L}(g)$ are purely periodic with the same period.*

In this paper, we will assume that the period is of length $\tau$.

For a real $z$ and an integer $q$ we use the notation

$$\mathbf{e}(z) = \exp(2\pi i z) \quad \textit{and} \quad \mathbf{e}_q(z) = \exp(2\pi i z/q).$$

We need the identity (see Exercise 11.a in Chapter 3 of [11])

$$\sum_{\eta=0}^{M-1} \mathbf{e}_M(\eta\lambda) = \begin{cases} 0, & \text{if} \quad \lambda \not\equiv 0 \pmod M, \\ M, & \text{if} \quad \lambda \equiv 0 \pmod M. \end{cases} \tag{1}$$

We also make use of the inequality

$$\sum_{\eta=0}^{l-1} \left| \sum_{\lambda=1}^{M} \mathbf{e}_l(\eta\lambda) \right| = O(l \log l), \tag{2}$$

which holds for any integers $l$ and $M, 1 \le M \le l$, (see [11], Chapter 3, Exercise 11c).

For a sequence of $N$ points

$$\Gamma = (\gamma_x)_{x=1}^{N}$$

in the unit interval, we denote its *discrepancy* by $D_\Gamma$. That is,

$$D_\Gamma = \sup_{B \subseteq [0,1)} \left| \frac{\mathcal{T}_\Gamma(B)}{N} - |B| \right|,$$

where $\mathcal{T}_\Gamma(B)$ is the number of points of the sequence $\Gamma$ in the interval

$$B = [\alpha, \beta) \subseteq [0, 1)$$

and the supremum is taken over all such intervals.

As we have mentioned, one of our basic tools to study the uniformity of distribution is the Erdös-Turán inequality, which we present in a slightly weaker form than that given by Theorem 1.21 of [3].

**Lemma 2.** *For any integer $L > 1$ and any sequence $\Gamma$ of $N$ points, the bound*

$$D_\Gamma = O\left( \frac{1}{L} + \frac{1}{N} \sum_{0<a<L} \frac{1}{a} \left| \sum_{n=1}^{N} \mathbf{e}(a\gamma_n) \right| \right)$$

*on the discrepancy $D_\Gamma$ holds, where the sum is taken over all $a \in \mathbb{F}_p$ with $0 < a < L$.*

## 3. Main Result

We denote by $D_{\mathbf{Z}}(N)$ the discrepancy of the points

$$\left( \frac{x(V_{\mathbf{Z}}(n))}{p} \right), \qquad n = 1, \ldots, N.$$

**Theorem 3.** *Let $(u(n))$ be a linear recurrence sequence of order $r \le p^{1/2}$ which is purely periodic with period $\tau$ and suppose that its characteristic polynomial is irreducible over $\mathbb{F}_2$. Then, for any $\delta > 0$ and $1 \le N \le \tau$, and for all but at most $O(\delta p^r)$ vectors of $\mathbf{Z} \in (\mathbf{E}_(\mathbb{F}_p))^r$, the bound*

$$D_{\mathbf{Z}}(N) = O\left(\delta^{-1} N^{-1/2} \log p \log^2 \tau\right)$$

*holds if $p \ge N^2$, and*

$$D_{\mathbf{Z}}(N) = O\left(\delta^{-1} p^{-1/4} \log p \log^2 \tau\right)$$

*otherwise.*

*Proof.* Let $L = p$. Then by Lemma 2, we have

$$D_{\mathbf{Z}}(N) = O\left(\frac{1}{p} + \frac{1}{N} \sum_{0<a<p} \frac{1}{a} \left| \sum_{n=1}^{N} \mathbf{e}_p(ax(V_{\mathbf{Z}}(n))) \right| \right).$$

Let $N_\mu = \min(2^\mu, \tau)$, $\mu \ge 0$ is the set of positive integers. Define $k$ by the inequality $N_{k-1} < N \le N_k$, that is, $k = \lceil \log_2 N \rceil$. Then from (1) we derive

$$\sum_{n=1}^{N} \mathbf{e}_p(ax(V_{\mathbf{Z}}(n))) = \frac{1}{N_k} \sum_{n=1}^{N_k} \sum_{\lambda=1}^{N} \sum_{\eta=0}^{N_k} \mathbf{e}_p(ax(V_{\mathbf{Z}}(n))) \mathbf{e}_{N_k}\left(\eta(n-\lambda)\right).$$

Hence,

$$D_{\mathbf{Z}}(N) = O\left(\frac{1}{p} + \frac{1}{NN_k} \Delta_{\mathbf{Z}}(k)\right) \tag{3}$$

where

$$\Delta_{\mathbf{Z}}(k) = \sum_{0<a<p} \frac{1}{a} \sum_{\eta=0}^{N_k} \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k}(-\eta\lambda) \right| \left| \sum_{n=1}^{N_k} \mathbf{e}_p(ax(V_{\mathbf{Z}}(n))) \mathbf{e}_{N_k}\left(\eta n\right) \right|.$$

Applying the Cauchy inequality we derive

$$\left( \sum_{\mathbf{Z} \in (\mathbf{E}(\mathbb{F}_p))^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_p(ax(V_{\mathbf{Z}}(n))) \mathbf{e}_{N_k}\left(\eta n\right) \right| \right)^2$$

$$\le (p^{1/2}+1)^{2r} \sum_{\mathbf{Z} \in (\mathbf{E}(\mathbb{F}_p))^r} \left| \sum_{n=1}^{N_k} \mathbf{e}_p(ax(V_{\mathbf{Z}}(n))) \mathbf{e}_{N_k}\left(\eta n\right) \right|^2$$

$$= O\left( p^r \sum_{n,l=1}^{N_k} \mathbf{e}_{N_k}\left(\eta(n-l)\right) \sum_{\mathbf{Z} \in (\mathbf{E}(\mathbb{F}_p))^r} \mathbf{e}_p(a(x(V_{\mathbf{Z}}(n)) - x(V_{\mathbf{Z}}(l)))) \right).$$

Using our upper bound for $r$, we then note that

$$\sum_{\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r} \mathbf{e}_p(a(x(V_{\mathbf{Z}}(n)) - x(V_{\mathbf{Z}}(l)))) \leq (p^{1/2}+1)^{2r} = O(p^r) \tag{4}$$

when $n = l$.

For $n \neq l$, then $(u(n), \ldots, u(n+r-1)) \neq (u(l), \ldots, u(l+r-1))$. To see this, suppose the converse is true and equality holds. Then, since $u(\alpha)$ is of order $r$, we also have $u(n+r) = u(l+r)$, and thus by induction for any $k \geq 0$, we have $u(n+k) = u(l+k)$. But then, $|n-l|$ is a period, which is impossible by our assumption on $n$ and $l$.

This means there is some $h$ with $u(n+h-1) \neq u(l+h-1) \pmod 2$, where $0 \leq h \leq r-1$. Thus, there is point $Z_h \in \mathbf{E}(\mathbb{F}_p)$ which appears in $x(V_{\mathbf{Z}}(n))$ but not in $x(V_{\mathbf{Z}}(l))$ (that is $u(n+h-1) = 1$ and $u(l+h-1) = 0$ or viceversa. We choose the former (the other case can be similarly estimated).

Let $\mathbf{Z}^{(h)} = (Z_1, \ldots, Z_{h-1}, Z_{h+1}, \ldots, Z_r) \in (\mathbf{E}(\mathbb{F}_p))^{r-1}$.

Thus,

$$\begin{aligned}&\sum_{\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r} \mathbf{e}_p(a(x(V_{\mathbf{Z}}(n)) - x(V_{\mathbf{Z}}(l))))\\ &\quad= \sum_{\mathbf{Z}^{(h)}\in(\mathbf{E}(\mathbb{F}_p))^{r-1}} \sum_{Z_h\in\mathbf{E}(\mathbb{F}_p)} \mathbf{e}_p(a(x(Q_n(\mathbf{Z}^{(h)}) + Z_h) - x(Q_l(\mathbf{Z}^{(h)})))),\end{aligned}$$

where $Q_n(\mathbf{Z}^{(h)})$ and $Q_l(\mathbf{Z}^{(h)})$ are some expressions which depend on $\mathbf{Z}^{(h)}$, but not on $Z_h$. Now, by Corollary 1 of [4],

$$\begin{aligned}&\left|\sum_{Z_h\in\mathbf{E}(\mathbb{F}_p)} \mathbf{e}_p(a(x(Q_n(\mathbf{Z}^{(h)}) + Z_h) - x(Q_l(\mathbf{Z}^{(h)}))))\right|\\ &\quad= \left|\sum_{Z_h\in\mathbf{E}(\mathbb{F}_p)} \mathbf{e}_p(a(x(Q_n(\mathbf{Z}^{(h)}) + Z_h)))\right| = O(p^{1/2})\end{aligned} \tag{5}$$

Therefore

$$\sum_{\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r} \mathbf{e}_p(a(x(V_{\mathbf{Z}}(n)) - x(V_{\mathbf{Z}}(l)))) = O(p^{r-1/2}),$$

when $n \neq l$.

Now, since

$$\left|\sum_{n,l=1,n=l}^{N_k} \mathbf{e}_{N_k}(\eta(n-l))\right| = N_k,$$

and

$$\left|\sum_{n,l=1,n\neq l}^{N_k} \mathbf{e}_{N_k}(\eta(n-l))\right| \leq N_k^2,$$

we have by (4) and (5)

$$\sum_{\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r}\left|\sum_{n=1}^{N_k}\mathbf{e}_p(ax(V_{\mathbf{Z}}(n)))\mathbf{e}_{N_k}(\eta n)\right| = O\left(N_k p^{r-1/4}+N_k^{1/2}p^r\right).$$

Hence, using (2), we obtain,

$$\begin{aligned}\sum_{\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r}\Delta_{\mathbf{Z}}(k) &= \sum_{0<a<p}\frac{1}{a}\sum_{\eta=0}^{N_k}\left|\sum_{\lambda=1}^{N}\mathbf{e}_{N_k}(-\eta\lambda)\right| \\ &\qquad \sum_{\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r}\left|\sum_{n=1}^{N_k}\mathbf{e}_p\left(ax(V_{\mathbf{Z}}(n))\right)\mathbf{e}_{N_k}(\eta n)\right| \\ &= O\left(p^r\left(N_k^{1/2}+N_kp^{-1/4}\right)\sum_{0<a<p}\frac{1}{a}\sum_{\eta=0}^{N_k}\left|\sum_{\lambda=1}^{N}\mathbf{e}_{N_k}(-\eta\lambda)\right|\right) \\ &= O\left(p^r\left(N_k^{3/2}+N_k^2p^{-1/4}\right)k\sum_{0<a<p}\frac{1}{a}\right) \\ &= O\left(p^r\left(N_k^{3/2}+N_k^2p^{-1/4}\right)\log\tau\log p\right),\end{aligned}$$

because $k=O(\log\tau)$.

This implies that for any $k$ the number of vectors $\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r$ with

$$\Delta_{\mathbf{Z}}(k)\ge\delta^{-1}\left(N_k^{3/2}+N_k^2p^{-1/4}\right)\log p\log^2\tau \tag{6}$$

is at most $O(\delta p^r\log^{-1}\tau)$. Therefore, we have that the number of vectors $\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r$ satisfying (6) for at least one $k=1,\dots,\lceil\log\tau\rceil$ is at most $O(\delta p^r)$. For other $\mathbf{Z}\in(\mathbf{E}(\mathbb{F}_p))^r$, from (3), we obtain

$$\begin{aligned}&D_{\mathbf{Z}}(N)\\ &\quad=O\left(\frac{1}{p}+\frac{1}{NN_k}\Delta_{\mathbf{Z}}(k)\right)=O\left(\delta^{-1}N^{-1}\left(N_k^{1/2}+N_kp^{-1/4}\right)\log p\log^2\tau\right).\end{aligned}$$

Taking into account the inequality $N^{-1}N_k^{1/2}\le 2N^{-1/2}$, we obtain the desired result. □

## 4. Remarks

We remark that this technique can not unfortunately be employed to establish a similar result for the multidimensional case. In this instance, we can not make use of Corollary 1

of [4]. As such, it would be interesting if, indeed, a similar result could be established for dimensions greater than one. Also, we note that although in the proof of the main theorem we could have replaced $N_k$ by $N$, this would have resulted in a weaker result as the statement would apply to a fixed $N$. That is, for every $N$, the exception set of "unsuitable" $\mathbf{Z}$, which would satisfy (6), depends on $N$. In particular, our result means that after removing a small exceptional "bad" set of vectors, the bound holds uniformly for all $N$. We do not see how to obtain such a result without introducing the numbers $N_k$ and thus more complicated exponential sums. Finally, we conclude by saying that instead of employing powers of 2 we could have used any other real number greater than 1, such as $1+10^{-100}$ or $10^{100}$. However, if we want to fit everything, then $N_k$ should grow a little slower than a geometric progression, as it saves working with double logarithms.

## 5. Acknowledgements

The author is very grateful to Igor Shparlinski for suggesting this problem and for useful discussion.

## References

[1] I. Blake, G. Seroussi, N. Smart, Elliptic curves in cryptography, London Math. Soc., Lecture Note Series, **265**, Cambridge Univ. Press (1999).

[2] A. Conflitti, I. E. Shparlinski, On the multidimensional distribution of the subset sum generator of pseudorandom numbers', Mathematics of Computation, **73**, 1005–1011 (2004).

[3] M. Drmota, R. Tichy, Sequences, discrepancies and applications, Springer-Verlag, Berlin (1997).

[4] D. R. Kohel, I. E. Shparlinski, Exponential sums and group generators for elliptic curves over finite fields, Lect. Notes in Comp. Sci., Springer-Verlag, Berlin, **1838**, 395–404 (2000).

[5] R. Lidl, H. Niederreiter, Finite fields, Cambridge University Press, Cambridge (1997).

[6] A. J. Menezes, P. C. van Oorschot, S. A. Vanstone, Handbook of applied cryptography, CRC Press, Boca Raton, FL (1996).

[7] R. A. Rueppel, Analysis and design of stream ciphers, Springer-Verlag, Berlin (1986).

[8] R. A. Rueppel, Stream ciphers, Contemporary cryptology: The science of information integrity, 65-134, IEEE Press, NY (1992).

[9] R. A. Rueppel, J. L Massey,Knapsack as nonlinear function, IEEE Intern. Symp. of Inform. Theory, **46**, IEEE Press, NY (1985).

[10] J. H. Silverman, The arithmetic of elliptic curves, Springer-Verlag, Berlin (1995).

[11] I. M. Vinogradov, Elements of number theory, Dover Publ., New York (1954).