J. Math. Crypt. **3** (2009), 149–163

DOI 10.1515 / JMC.2009.007

Subset sum pseudorandom numbers: fast generation and distribution

Joachim von zur Gathen and Igor E. Shparlinski

Communicated by Tran van Trung

Abstract. We show how to accelerate the subset sum pseudorandom number generator with arbitrary weights. Some special choices of weights speed up the naive usage of this generator without losing the property of uniform distribution which has recently been established in the general case. Our results confirm that this generator can be useful for both cryptographic and Quasi Monte Carlo applications.

Keywords. Pseudorandom numbers, subset sum problem, knapsack problem, exponential sums.

AMS classification. 11K45, 68P25, 94A60.

1 Introduction

We consider the following type of pseudorandom generator. We have a sequence $u_0, u_1, \ldots$ of integers, a sequence $\mathbf{w} = (w_0, \ldots, w_{n-1})$ of n elements of an (additively written Abelian) group $\mathcal{R}$, and the sequence $v_{\mathbf{w}} = v_{\mathbf{w}}(0), v_{\mathbf{w}}(1), \ldots$ of elements of $\mathcal{R}$ generated according to the formula

$$v_{\mathbf{w}}(i) = \sum_{0 \le j < n} u_{i+j} w_j \tag{1.1}$$

for $i \ge 0$. For a multiplicatively written group, we would interpret (1.1) as $v_{\mathbf{w}}(i) = w_0^{u_i} \cdot w_1^{u_{i+1}} \cdots w_{n-1}^{u_{i+n-1}}$.

A particularly useful way of obtaining an integer sequence $u_0, u_1, \ldots$ is to take a linear recurrence sequence (see [9, Chapter 8]) $\widetilde{u}_0, \widetilde{u}_1, \ldots$ over some ring $\mathcal{Q}$, together with some way of mapping an element $\widetilde{u} \in \mathcal{Q}$ to an integer $u \in \mathbb{Z}$. In this paper only such sequences of integers are considered.

If we take $\mathcal{Q} = \mathbb{Z}_q$, the residue ring modulo $q \ge 2$, then we can lift in a natural way a ring element $\widetilde{u}$ to the unique integer $u \in \{0, \ldots, q-1\}$ with $\widetilde{u} = (u \mod q)$. This produces the required sequence $u_0, u_1, \ldots$ in $\mathbb{Z}$. The case where q is prime, that is $\mathcal{Q} = \mathbb{F}_q$ is a finite field of q elements, is of special interest.

In fact throughout the paper, slightly violating notations, we do not distinguish between the sequences $u_0, u_1, \ldots$ and $\widetilde{u}_0, \widetilde{u}_1, \ldots$. In particular, we say that the sequence of integers $u_0, u_1, \ldots$ is a linear recurrence sequence over $\mathcal{Q}$ if the sequence $\widetilde{u}_0, \widetilde{u}_1, \ldots$ is.

When we further specialize q to 2, then $u_0, u_1, \ldots$ is a sequence of bits which is a linear recurrence sequence modulo 2, and we call the sequence generated by (1.1) a *subset sum* generator over $\mathcal{R}$. Given $v_{\mathbf{w}}(i)$ and $\mathbf{w}$, the determination of $u_i, \ldots, u_{i+n-1}$

is an instance of the subset sum problem (over $\mathcal{R}$), whose general case (over $\mathbb{Z}$) is NP-complete.

For $\mathcal{R} = \mathbb{Z}_r$ this generator, which is also known as *knapsack generator*, has been introduced in [16] and studied in [14], see also [11, Section 6.3.2] and [15, Section 3.7.9]. In [4] results on the uniformity of distribution for these generators have been obtained.

For cryptographic applications, it is usually recommended to use $\mathcal{R} = \mathbb{Z}_{2^n}$ and a binary linear recurrence sequence of order n and of maximal period $\tau = 2^n - 1$, however here we consider more general settings. Although the results of [7] suggest that this generator should be used with care, no major attack against it is known.

We remark that even if $u_0, u_1, \ldots$ is generated via a linear recurrence sequence and although we use a simple "linear" rule (1.1), they belong to different rings and as a result produce a highly nonlinear sequence. For example, it is asserted in [16] that the linear complexity of such a sequence (with $\mathcal{Q} = \mathbb{Z}_2$ and $\mathcal{R} = \mathbb{Z}_r$) is sufficiently large. However, one can probably increase the nonlinearity and thus the security properties of this construction by choosing groups $\mathcal{R}$ of more complicated structure, for example groups of points of elliptic curve over finite fields, see [10] for some uniformity of distribution results for the elliptic curve analogue of the subset sum generator.

It has been shown in [4] that the multidimensional distribution of the subset sum generator is close to uniform. In the special case $\mathcal{R} = \mathbb{Z}_{2^n}$ and of weights $w_j = 2^{n-j-1}$, $j = 0, \ldots, n-1$, this generator is well known in the theory of Quasi Monte Carlo methods. An exhaustive survey of known results about the distribution of this and more general generators can be found in [12, Chapter 9]. The known results on the uniformity of distribution for this deterministic choice of weights are weaker than the results of [4] which, however, apply only to a randomized choice of weights. It is also clear that the former choice of weights corresponds to easy instances of the knapsack problem and thus is probably not suitable for cryptographic applications.

For $\mathcal{Q} = \mathbb{F}_2$ a straight-forward evaluation of (1.1) takes at most n additions in $\mathbb{F}_2$ and n additions in $\mathcal{R}$ (this is why this case is theoretically and practically quite attractive). Here we show that in fact there are more efficient algorithms which apply to arbitrary rings $\mathcal{Q}$ and $\mathcal{R}$. We also obtain new results about the uniformity of distribution of the corresponding sequences.

To be more precise, the results of this paper are of two types.

- Efficient algorithms when $\mathcal{R}$ is a ring, namely
  - with $O(\log n \mathrm{loglog} n)$ operations for one element as in (1.1), in an amortized sense,
  - with a constant number of operations for a special choice of $\mathbf{w}$.
- Results on the distribution, namely upper bounds on the discrepancy
  - in the general case of a ring $\mathcal{R}$, and where $q = \#\mathcal{Q}$ is small,
  - in the special case mentioned above, with $\mathcal{Q} = \mathbb{Z}_2$,

Throughout the paper, the implied constants in symbols "O" may occasionally, where obvious, depend on the integer parameter $\nu \geq 1$, and are absolute otherwise.

2 Fast algorithms

2.1 General seeds over rings

Here we assume that $\mathcal{R}$ is a ring and consider the task of generating n arbitrary consecutive values $v_{\mathbf{w}}(i), \ldots, v_{\mathbf{w}}(i+n-1)$ of our generator. Thus, we are given $2n-1$ terms $u_i, \ldots, u_{i+2n-2}$ and n weights $w_0, \ldots, w_{n-1} \in \mathcal{R}$ and have to compute the product of the $n \times n$ Hankel matrix

$$\begin{pmatrix} u_i & u_{i+1} & \cdots & u_{i+n-1} \\ \vdots & \vdots & & \vdots \\ u_{i+n-1} & u_{i+n} & \cdots & u_{i+2n-2} \end{pmatrix}$$

with the vector $w = (w_0, \ldots, w_{n-1})^t \in \mathcal{R}^n$.

It is well known that this can be done efficiently, by reducing it to polynomial multiplication. Let M denote a multiplication time, so that two polynomials in $\mathcal{R}[x]$ of degree at most n can be multiplied with at most $\mathsf{M}(n)$ operations in $\mathcal{R}$. Since we can choose $\mathsf{M}(n) = O(n \log n \operatorname{loglog} n)$, see [6, Section 8.3] and the product of an $n \times n$ Hankel matrix with a vector can be computed with $O(\mathsf{M}(n))$ ring operations see [13, Chapter 2], we obtain:

Theorem 2.1. *There is an algorithm which generates n consecutive values $v_{\mathbf{w}}(i+j)$ for $j = 0, \ldots, n-1$ with $O(\log n \operatorname{loglog} n)$ operations in $\mathcal{R}$ per element.*

It is natural to ask whether in the case of $q = 2$ one can get a further speed-up.

Question 2.2. In Theorem 2.1 we do not use the fact that all u_i are 0 or 1. Can one compute the product of a 0-1-polynomial with a general polynomial significantly faster than in the general case?

2.2 Special seeds over rings

We now introduce a special choice of weights which helps to speed up generation. Namely we consider a ring $\mathcal{R}$ and weights of the form $w_j = yz^{n-j}$ for $j = 0, \ldots, n-1$ and some $y, z \in \mathcal{R}$. We denote by $v_{y,z}$ the corresponding sequence given by (1.1).

Theorem 2.3. *After a preprocessing stage requiring $O(n)$ operations in $\mathcal{R}$, one can generate elements of the sequence $v_{y,z}(i)$ consecutively with at most one general multiplication, two multiplications by integers in the range $\{0, \ldots, q-1\}$ and two additions in $\mathcal{R}$ per element.*

Proof. One verifies the identity

$$v_{y,z}(i+1) = z v_{y,z}(i) - yz^{n+1} \cdot u_i + yz \cdot u_{i+n}$$

for $i = 0, 1, \ldots$. Thus after precomputing yz^j for $0 \le j \le n+1$ and $-yz^{n+1}$ and calculating $v_{y,z}(0)$, with a total of $O(n)$ operations in $\mathcal{R}$, each further element can be generated with one multiplication by z, two multiplications by control values u_j, and two additions in $\mathcal{R}$. □

In the case $q = 2$, if one ignores the multiplications by 0 or 1, then only one multiplication and two additions are used per element.

In Section 3 we show that if $\mathcal{R} = \mathbb{F}_p$ where p is prime, then for any $y \in \mathbb{F}_p^*$ and almost all $z \in \mathbb{F}_p$ the sequence $v_{y,z}(i)$ has attractive uniform distribution properties.

We remark that in residue rings $\mathcal{R} = \mathbb{Z}_r$ one can obtain an additional speed-up by choosing a reasonable small value of z, so that multiplication by z is faster than generic modular multiplication modulo r. In Section 3 we show that if y is chosen at random from $\mathbb{F}_p$, but z is chosen from a small interval $[0, H]$, then with overwhelming probability the sequence $v_{y,z}(i)$ remains uniformly distributed.

A further advantage of these "dependent" weights is that they only require two random elements of $\mathcal{R}$ for the seed rather than the n random elements as in the case of independent weights. It is not clear whether this specialization affects the security of the generator.

3 Uniformity of distribution

3.1 Preparations

Our method is based on some simple bounds on exponential sums and the famous *Koksma–Szüsz inequality* (see Lemma 3.2 below) which relates the deviation from uniformity of distribution, that is, the discrepancy, and the corresponding exponential sums.

Here we present several necessary technical tools.

We say that a linear recurrence sequence u_i of elements of $\mathbb{F}_q$ is of order n with characteristic polynomial

$$f(T) = T^n + c_{n-1}T^{n-1} + \ldots + c_1T + c_0 \in \mathbb{F}_q[T]$$

if

$$u_{i+n} + c_{n-1}u_{i+n-1} + \ldots + c_1u_{i+1} + c_0u_i = 0, \qquad i = 0, 1, \ldots,$$

and it does not satisfy any shorter linear relation, see [9, Chapter 8].

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

We also need the following property of sequences from $\mathcal{L}(f)$ with irreducible f which is essentially [9, Theorem 8.28].

Lemma 3.1. *If $f \in \mathbb{F}_q[T]$ is irreducible over $\mathbb{F}_q$ then all nonzero sequences from $\mathcal{L}(f)$ are purely periodic with the same period.*

For a real z and an integer M we use the notation

$$\mathbf{e}(z) = \exp(2\pi iz) \qquad \text{and} \qquad \mathbf{e}_M(z) = \exp(2\pi iz/M).$$

We need the identity (see Exercise 11.a in Chapter 3 of [18])

$$\sum_{\eta=0}^{M-1} \mathbf{e}_M(\eta\lambda) = \begin{cases} 0, & \text{if } \lambda \not\equiv 0 \pmod M, \\ M, & \text{if } \lambda \equiv 0 \pmod M. \end{cases} \tag{3.1}$$

We also make use of the inequality

$$\sum_{\eta=0}^{M-1} \left| \sum_{\lambda=1}^{N} \mathbf{e}_M\left(\eta\lambda\right) \right| = O\left(M \log M\right), \tag{3.2}$$

which holds for any integers $M \geq N \geq 1$, see [18, Chapter III, Exercise 11c].

For a sequence of N points

$$\Gamma = \left(\gamma_{0,x}, \ldots, \gamma_{\nu-1,x}\right)_{x=1}^{N} \tag{3.3}$$

in the ν-dimensional unit cube, we denote its *discrepancy* by D_Γ. That is,

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

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

$$B = [\alpha_0, \beta_0) \times \ldots \times [\alpha_{\nu-1}, \beta_{\nu-1}) \subseteq [0,1)^\nu$$

and the supremum is taken over all such boxes.

As we have mentioned, one of our basic tools to study the uniformity of distribution is the Koksma–Szüsz inequality, which we present in a slightly weaker form than that given by Theorem 1.21 of [5].

For an integer vector $\mathbf{a} = (a_1, \ldots, a_\nu) \in \mathbb{Z}^\nu$ we define

$$|\mathbf{a}| = \max_{j=1,\ldots,\nu} |a_j|, \qquad \rho(\mathbf{a}) = \prod_{j=1}^{\nu} \max\{|a_j|, 1\}. \tag{3.4}$$

Lemma 3.2. *For any integer $L > 1$ and any sequence Γ of N points* (3.3) *the bound*

$$D_\Gamma = O\left(\frac{1}{L} + \frac{1}{N} \sum_{0<|\mathbf{a}|<L} \frac{1}{\rho(\mathbf{a})} \left| \sum_{x=1}^{N} \mathbf{e}\left(\sum_{j=0}^{\nu-1} a_j \gamma_{j,x}\right) \right|\right)$$

on the discrepancy D_Γ holds, where $|\mathbf{a}|$, $\rho(\mathbf{a})$ are defined by (3.4) *and the sum is taken over all integer vectors*

$$\mathbf{a} = (a_0, \ldots, a_{\nu-1}) \in \mathbb{Z}^\nu$$

with $0 < |\mathbf{a}| < L$.

Finally, we recall the *Weil bound* in its classical form given in Example 12 of Appendix 5 of [19]; see also Theorem 5.41 and comments to Chapter 5 of [9].

Lemma 3.3. *For any prime p and polynomial $f(X) \in \mathbb{Z}[X]$ of degree d which is not constant modulo p we have*

$$\left| \sum_{x=1}^{p} \mathbf{e}_p\left(f(x)\right) \right| \leq dp^{1/2}.$$

3.2 General seeds over $\mathcal{R} = \mathbb{Z}_r$

We denote by $D^\nu_{\mathbf{w}}(N)$ the discrepancy of the points

$$\left(\frac{v_{\mathbf{w}}(i)}{r}, \ldots, \frac{v_{\mathbf{w}}(i+\nu-1)}{r}\right), \qquad i = 1, \ldots, N.$$

In the case $q = 2$, it has been shown in [4] that for almost all weights $\mathbf{w} \in \mathbb{Z}_r^m$, the discrepancy $D^\nu_{\mathbf{w}}(N)$ is $O\left(N^{-1/2}(\log r)^\nu (\log \tau)^2\right)$ for any $\nu \le n$ (we recall that the implied constants may depend on ν).

Here we extend the result of [4] and obtain a similar upper bound on $D^\nu_{\mathbf{w}}(N)$ in the case $\mathcal{Q} = \mathbb{F}_q$ where q is prime and $\mathcal{R} = \mathbb{Z}_r$ where $r \ge 2$ is an integer.

Theorem 3.4. *Assume that $\mathcal{Q} = \mathbb{F}_q$ where q is prime and $\mathcal{R} = \mathbb{Z}_r$ where $r \ge 2$ is integer. Let the linear recurrence sequence u_i be purely periodic with period τ and order n and let its characteristic polynomial be irreducible over $\mathbb{F}_q$. Then for any $\delta > 0$, and any $\nu \le n$ for all $\mathbf{w} \in \mathbb{Z}_r^n$ except at most $O(\delta r^n)$ of them, for all $1 \le N \le \tau$ the bound*

$$D^\nu_{\mathbf{w}}(N) = O\left(\frac{q}{r} + \delta^{-1} N^{-1/2} (\log r)^\nu (\log \tau)^2\right)$$

holds.

Proof. From Lemma 3.2, used with $L = \lfloor r/q\nu \rfloor$ (thus $1/L = O(q/r)$), we derive

$$D^\nu_{\mathbf{w}}(N) = O\left(\frac{q}{r} + \frac{1}{N} \sum_{0 < |\mathbf{a}| < r/q\nu} \frac{1}{\rho(\mathbf{a})} \left| \sum_{m=1}^{N} \mathbf{e}_r\left(\sum_{j=0}^{\nu-1} a_j v_{\mathbf{w}}(m+j)\right)\right|\right).$$

Let $N_\mu = 2^\mu$, $\mu = 0, 1, \ldots$. Define k by the inequality $N_{k-1} < N \le N_k$, that is, $k = \lceil \log_2 N \rceil$. Then from (3.1) we derive

$$\sum_{m=1}^{N} \mathbf{e}_r\left(\sum_{j=0}^{\nu-1} a_j v_{\mathbf{w}}(m+j)\right) = \frac{1}{N_k} \sum_{m=1}^{N_k} \sum_{\lambda=1}^{N} \sum_{\eta=0}^{N_k} \mathbf{e}_r\left(\sum_{j=0}^{\nu-1} a_j v_{\mathbf{w}}(m+j)\right) \mathbf{e}_{N_k}\left(\eta(m-\lambda)\right).$$

Hence,

$$D^\nu_{\mathbf{w}}(N) = O\left(\frac{q}{r} + \frac{1}{N N_k} \Delta^\nu_{\mathbf{w}}(k)\right), \tag{3.5}$$

where

$$\Delta^\nu_{\mathbf{w}}(k) = \sum_{0 < |\mathbf{a}| < r/q\nu} \frac{1}{\rho(\mathbf{a})} \sum_{\eta=0}^{N_k} \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k}(-\eta\lambda) \right| \left| \sum_{m=1}^{N_k} \mathbf{e}_r\left(\sum_{j=0}^{\nu-1} a_j v_{\mathbf{w}}(m+j)\right) \mathbf{e}_{N_k}(\eta m) \right|.$$

Applying the Cauchy inequality we derive

$$\left(\sum_{\mathbf{w}\in\mathbb{Z}_r^n}\left|\sum_{m=1}^{N_k}\mathbf{e}_r\left(\sum_{j=0}^{\nu-1}a_jv_{\mathbf{w}}(m+j)\right)\mathbf{e}_{N_k}(\eta m)\right|\right)^2$$
$$\le r^n\sum_{\mathbf{w}\in\mathbb{Z}_r^n}\left|\sum_{m=1}^{N_k}\mathbf{e}_r\left(\sum_{j=0}^{\nu-1}a_jv_{\mathbf{w}}(m+j)\right)\mathbf{e}_{N_k}(\eta m)\right|^2$$
$$= r^n\sum_{m,l=1}^{N_k}\mathbf{e}_{N_k}(\eta(m-l))\sum_{\mathbf{w}\in\mathbb{Z}_r^n}\mathbf{e}_r\left(\sum_{j=0}^{\nu-1}a_j\left(v_{\mathbf{w}}(m+j)-v_{\mathbf{w}}(l+j)\right)\right).$$

By the definition of $v_{\mathbf{w}}(i)$ we have

$$\sum_{\mathbf{w}\in\mathbb{Z}_r^n}\mathbf{e}_r\left(\sum_{j=0}^{\nu-1}a_j\left(v_{\mathbf{w}}(m+j)-v_{\mathbf{w}}(l+j)\right)\right)$$
$$=\prod_{h=1}^{n}\sum_{z_h\in\mathbb{Z}_r}\mathbf{e}_r\left(w_h\sum_{j=0}^{\nu-1}a_j\left(u_{m+j+h-2}-u_{l+j+h-2}\right)\right).$$

The product is equal to r^n if for every $h=1,\dots,n$

$$\sum_{j=0}^{\nu-1}a_j\left(u_{m+j+h-2}-u_{l+j+h-2}\right)\equiv 0 \pmod r \tag{3.6}$$

otherwise it vanishes.

Therefore

$$\sum_{\mathbf{w}\in\mathbb{Z}_r^n}\left|\sum_{m=1}^{N_k}\mathbf{e}_r\left(\sum_{j=0}^{\nu-1}a_jv_{\mathbf{w}}(m+j)\right)\mathbf{e}_{N_k}(\eta m)\right|\le r^nT_k^{1/2} \tag{3.7}$$

where T_k is the number of pairs (m,l) with $1\le m,l\le N_k$ for which (3.6) holds for every $h=1,\dots,n$.

Because $u_x\in\{0,\dots,q-1\}$ for all integers $x\ge 1$ and $0\le|a_j|<r/q\nu$, the congruence (3.6) becomes an equation

$$\sum_{j=0}^{\nu-1}a_j\left(u_{m+j+h-2}-u_{l+j+h-2}\right)=0,\qquad h=1,\dots,n.$$

Let us write $a_j=q^\alpha b_j$ where q^α is the largest power of q which divides every a_j for $0\le j<\nu$. In particular, at least one of the b_j is relatively prime to q. Then the previous

equation becomes

$$\sum_{j=0}^{\nu-1} b_j \left(u_{m+j+h-2} - u_{l+j+h-2}\right) = 0, \qquad h = 1, \ldots, n. \tag{3.8}$$

Considering the equation (3.8) in $\mathcal{Q} = \mathbb{F}_q$, we derive

$$w_{m+h} \equiv w_{l+h} \pmod{q}, \qquad h = 1, \ldots, n,$$

where

$$w_x = \sum_{j=0}^{\nu-1} b_j u_{x+j-2}$$

is a nonzero sequence over $\mathbb{F}_q$ because at least one b_j with $0 \le j < \nu$ is relatively prime to q and $\nu \le r$. Taking into account that w_x is a linear recurrence sequence of order r (with the same characteristic polynomial as u_x) we obtain

$$w_{m+x} \equiv w_{l+x} \pmod{q}, \qquad x = 1, 2, \ldots. \tag{3.9}$$

Because the characteristic polynomial of u is irreducible, by Lemma 3.1 the linear recurrence sequence w_x has the same period τ. Therefore (3.9) implies that $n \equiv l \pmod{\tau}$ which yields the inequality $T_k \le N_k(\lfloor N_k/\tau \rfloor + 1) \le 2N_k$, because $N_k = 2N_{k-1} < 2N \le 2\tau$.

Thus by (3.7) we have

$$\sum_{\mathbf{w} \in \mathbb{Z}_r^n} \left| \sum_{m=1}^{N_k} \mathbf{e}_r \left(\sum_{j=0}^{\nu-1} a_j v_{\mathbf{w}}(m+j) \right) \mathbf{e}_{N_k}(\eta m) \right| \le 2^{1/2} r^n N_k^{1/2}.$$

Hence recalling (3.2) we obtain

$$\begin{aligned}
\sum_{\mathbf{w} \in \mathbb{Z}_r^n} \Delta_{\mathbf{w}}^{\nu}(k) &= \sum_{0 < |\mathbf{a}| < r/q\nu} \frac{1}{\rho(\mathbf{a})} \sum_{\eta=0}^{N_k} \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k}(-\eta\lambda) \right| \\
&\qquad \sum_{\mathbf{w} \in \mathbb{Z}_r^n} \left| \sum_{m=1}^{N_k} \mathbf{e}_r \left(\sum_{j=0}^{\nu-1} a_j v_{\mathbf{w}}(m+j) \right) \mathbf{e}_{N_k}(\eta m) \right| \\
&= 2^{1/2} r^n N_k^{1/2} \sum_{0 < |\mathbf{a}| < r/q\nu} \frac{1}{\rho(\mathbf{a})} \sum_{\eta=0}^{N_k} \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k}(-\eta\lambda) \right| \\
&= O\left(r^n N_k^{3/2} k \sum_{0 < |\mathbf{a}| < r/q\nu} \frac{1}{\rho(\mathbf{a})} \right) \\
&= O\left(r^n N_k^{3/2} k (\log r)^{\nu} \right) \\
&= O\left(r^n N_k^{3/2} (\log r)^{\nu} \log \tau \right),
\end{aligned}$$

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

This implies that for any k the number of vectors $\mathbf{w} \in \mathbb{Z}_r^n$ with

$$\Delta_{\mathbf{w}}^{\nu}(k) \ge \delta^{-1} N_k^{3/2} (\log r)^{\nu} (\log \tau)^2$$

is at most $O(\delta m^r (\log \tau)^{-1})$. Therefore, we have that the number of vectors $\mathbf{w} \in \mathbb{Z}_r^n$ with

$$\Delta_{\mathbf{w}}^{\nu}(k) \ge \delta^{-1} N_k^{3/2} (\log r)^{\nu} (\log \tau)^2$$

for at least one $k = 1, \ldots, \lceil \log \tau \rceil$ is at most $O(\delta r^n)$. For other $\mathbf{w} \in \mathbb{Z}_r^n$, from (3.5), we obtain

$$D_{\mathbf{w}}^{\nu}(N) = O\left(\frac{q}{r} + \frac{1}{NN_k} \Delta_{\mathbf{w}}^{\nu}(k)\right) = O\left(\frac{q}{r} + \delta^{-1} N^{-1} N_k^{1/2} (\log r)^{\nu} (\log \tau)^2\right).$$

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

In particular, if $q \le r^{1/2}$ then the first term in the bound of Theorem 3.4 never dominates and the bound takes that same form as that of [4], obtained for $q = 2$.

3.3 Special seeds over $\mathcal{R} = \mathbb{F}_p$

Here we obtain an upper bound on the discrepancy of the sequence $v_{y,z}(n)$ in the case $\mathcal{Q} = \mathbb{F}_q$, $\mathcal{R} = \mathbb{F}_p$ where q and p are prime.

We denote by $D_{y,z}^{\nu}(N)$ the discrepancy of the points

$$\left(\frac{v_{y,z}(i)}{p}, \ldots, \frac{v_{y,z}(i+\nu-1)}{p}\right), \qquad i = 1, \ldots, N.$$

Theorem 3.5. *Assume that $\mathcal{Q} = \mathbb{F}_q$ and that $\mathcal{R} = \mathbb{F}_p$ where q and p are prime. Let the linear recurrence sequence u_i be purely periodic with period τ and order n and let its characteristic polynomial be irreducible over $\mathbb{F}_q$. Then for any $\delta > 0$, positive integer $\nu \le n$ and $y \in \mathbb{F}_p^*$, for all $z \in \mathbb{F}_p$ except at most $O(\delta p)$ of them, and for all $1 \le N \le \tau$ we have*

$$D_{y,z}^{\nu}(N) \le \delta^{-1} \left(\frac{1}{N^{1/2}} + \frac{n^{1/2}}{p^{1/4}}\right) (\log p)^{\nu} (\log \tau)^2.$$

Proof. From Lemma 3.2, used with $L = \lfloor p/\nu \rfloor$, we derive

$$D_{y,z}^{\nu}(N) = O\left(\frac{1}{p} + \frac{1}{N} \sum_{0 < |\mathbf{a}| < p/\nu} \frac{1}{\rho(\mathbf{a})} \left| \sum_{m=1}^{N} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right)\right|\right).$$

Let $N_\mu = 2^\mu$ for $\mu \ge 0$. Define k by the inequality $N_{k-1} < N \le N_k$, that is, $k = \lceil \log_2 N \rceil$. Then from (3.1) we derive

$$\sum_{m=1}^{N} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right) = \frac{1}{N_k} \sum_{m=1}^{N_k} \sum_{\lambda=1}^{N} \sum_{\eta=0}^{N_k} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right) \mathbf{e}_{N_k}\left(\eta(m-\lambda)\right).$$

Hence,

$$D^\nu_{y,z}(N) = O\left(\frac{1}{p} + \frac{1}{NN_k}\Delta^\nu_{y,z}(k)\right), \tag{3.10}$$

where

$$\Delta^\nu_{y,z}(k) = \sum_{0<|\mathbf{a}|<p/\nu} \frac{1}{\rho(\mathbf{a})} \sum_{\eta=0}^{N_k} \left|\sum_{\lambda=1}^{N} \mathbf{e}_{N_k}\left(-\eta\lambda\right)\right| \left|\sum_{m=1}^{N_k} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right) \mathbf{e}_{N_k}\left(\eta m\right)\right|.$$

For $0 \le \eta \le N_k$, we define

$$\sigma(\eta) = \sum_{z\in\mathbb{F}_p} \left|\sum_{m=1}^{N_k} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right) \mathbf{e}_{N_k}\left(\eta m\right)\right|.$$

Applying the Cauchy inequality, we derive

$$\sigma(\eta)^2 \le p \sum_{z\in\mathbb{F}_p} \left|\sum_{m=1}^{N_k} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right) \mathbf{e}_{N_k}\left(\eta m\right)\right|^2$$
$$= p \sum_{m,l=1}^{N_k} \mathbf{e}_{N_k}\left(\eta(m-l)\right) \sum_{z\in\mathbb{F}_p} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j \left(v_{y,z}(m+j) - v_{y,z}(l+j)\right)\right).$$

By the definition of $v_{y,z}(i)$ we have

$$\sum_{z\in\mathbb{F}_p} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j \left(v_{y,z}(m+j) - v_{y,z}(l+j)\right)\right) = \sum_{z\in\mathbb{F}_p} \mathbf{e}_p\left(y\sum_{h=1}^{n} z^h \sum_{j=0}^{\nu-1} a_j \left(u_{m+j+h-2} - u_{l+j+h-2}\right)\right).$$

The sum is equal to p if for every $h = 1, \ldots, n$

$$\sum_{j=0}^{\nu-1} a_j \left(u_{m+j+h-2} - u_{l+j+h-2}\right) \equiv 0 \pmod{p}, \tag{3.11}$$

otherwise, by Lemma 3.3 its absolute value is at most $np^{1/2}$.

Therefore

$$\sigma(\eta)^2 \le p^2 T_k + np^{3/2}(N_k^2 - T_k) \le p^2 T_k + np^{3/2} N_k^2,$$

and we derive

$$\sigma(\eta) = O\left(pT_k^{1/2} + n^{1/2} N_k p^{3/4}\right), \tag{3.12}$$

where T_k is the number of pairs (m, l), $1 \le m, l \le N_k$, for which (3.11) holds for every $h = 1, \ldots, n$. It has been shown in [4] that $T_k \le 2N_k$.

Thus by (3.12) we have

$$\sigma(\eta) = O\left(pN_k^{1/2} + n^{1/2} N_k p^{3/4}\right).$$

Hence recalling (3.2) we obtain

$$\begin{aligned}
\sum_{z \in \mathbb{F}_p} \Delta_{y,z}^{\nu}(k) &= \sum_{0 < |\mathbf{a}| < p/\nu} \frac{1}{\rho(\mathbf{a})} \sum_{\eta=0}^{N_k} \sigma(\eta) \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k}\left(-\eta\lambda\right) \right| \\
&= O\left(\left(pN_k^{1/2} + n^{1/2} N_k p^{3/4}\right) \sum_{0 < |\mathbf{a}| < p/\nu} \frac{1}{\rho(\mathbf{a})} \sum_{\eta=0}^{N_k} \left| \sum_{\lambda=1}^{N} \mathbf{e}_{N_k}\left(-\eta\lambda\right) \right| \right) \\
&= O\left(k \left(pN_k^{3/2} + n^{1/2} N_k^2 p^{3/4}\right) \sum_{0 < |\mathbf{a}| < p/\nu} \frac{1}{\rho(\mathbf{a})} \right) \\
&= O\left(k \left(pN_k^{3/2} + n^{1/2} N_k^2 p^{3/4}\right) (\log p)^{\nu} \right) \\
&= O\left(\left(pN_k^{3/2} + n^{1/2} N_k^2 p^{3/4}\right) (\log p)^{\nu} \log \tau \right),
\end{aligned}$$

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

This implies that for any k the number of $z \in \mathbb{F}_p$ with

$$\Delta_{y,z}^{\nu}(k) \ge \delta^{-1} \left(N_k^{3/2} + n^{1/2} N_k^2 p^{-1/4}\right) (\log p)^{\nu} \log \tau$$

is at most $O(\delta p (\log \tau)^{-1})$. Therefore, we have that the number of $z \in \mathbb{F}_p$ with

$$\Delta_{y,z}^{\nu}(k) \ge \delta^{-1} \left(N_k^{3/2} + n^{1/2} N_k^2 p^{-1/4}\right) (\log p)^{\nu} \log \tau$$

for at least one $k = 1, \ldots, \lceil \log \tau \rceil$ is at most $O(\delta p)$. For other $z \in \mathbb{F}_p$, from (3.10), we obtain

$$\begin{aligned} D^{\nu}_{y,z}(N) &= O\left(\frac{1}{p} + \frac{1}{NN_k}\Delta^{\nu}_{y,z}(k)\right) \\ &= O\left(\delta^{-1}\left(N^{-1}N_k^{1/2} + n^{1/2}N^{-1}N_k p^{-1/4}\right)(\log p)^{\nu}(\log \tau)^2\right). \end{aligned}$$

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

We now show that if y is also randomized, we can obtain a stronger discrepancy bound. Moreover, the result remains nontrivial even if z is chosen from a relatively small interval.

Theorem 3.6. *Assume that $\mathcal{Q} = \mathbb{F}_q$ and that $\mathcal{R} = \mathbb{F}_p$ where q and p are prime. Let the linear recurrence sequence u_i be purely periodic with period τ and order n and let its characteristic polynomial be irreducible over $\mathbb{F}_q$. Then for any $\delta > 0$ and positive integers $H \le p$, $\nu \le n$, for all $y \in \mathbb{F}_p^*$ and $z \in [0, H-1]$ except at most $O(\delta pH)$ of them, for all $1 \le N \le \tau$ we have*

$$D^{\nu}_{y,z}(N) \le \delta^{-1}\left(\frac{1}{N^{1/2}} + \frac{n^{1/2}}{H^{1/2}}\right)(\log p)^{\nu}(\log \tau)^2.$$

Proof. As in the proof of Theorem 3.5 we see that our results depend on the following sum:

$$\Sigma(\eta) = \sum_{y \in \mathbb{F}_p^*} \sum_{z=0}^{H-1} \left| \sum_{m=1}^{N_k} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right) \mathbf{e}_{N_k}(\eta m)\right|.$$

Adding the term corresponding to $y = 0$ and using the Cauchy inequality, we find

$$\begin{aligned} \Sigma(\eta)^2 &\le pH \sum_{y \in \mathbb{F}_p} \sum_{z=0}^{H-1} \left| \sum_{m=1}^{N_k} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j v_{y,z}(m+j)\right) \mathbf{e}_{N_k}(\eta m)\right|^2 \\ &= pH \sum_{m,l=1}^{N_k} \mathbf{e}_{N_k}(\eta(m-l)) \\ &\qquad \times \sum_{z=0}^{H-1} \sum_{y \in \mathbb{F}_p} \mathbf{e}_p\left(\sum_{j=0}^{\nu-1} a_j \left(v_{y,z}(m+j) - v_{y,z}(l+j)\right)\right). \end{aligned}$$

By the definition of $v_{y,z}(n)$ we have

$$\sum_{z=0}^{H-1} \sum_{y \in \mathbb{F}_p} \mathbf{e}_p \left(\sum_{j=0}^{\nu-1} a_j \left(v_{y,z}(m+j) - v_{y,z}(l+j) \right) \right)$$
$$= \sum_{z=0}^{H-1} \sum_{y \in \mathbb{F}_p} \mathbf{e}_p \left(y \sum_{h=1}^{n} z^h \sum_{j=0}^{\nu-1} a_j \left(u_{m+j+h-2} - u_{l+j+h-2} \right) \right).$$

The inner sum is equal to p if

$$\sum_{h=1}^{n} z^h \sum_{j=0}^{\nu-1} a_j \left(u_{m+j+h-2} - u_{l+j+h-2} \right) = 0 \tag{3.13}$$

and vanishes otherwise. If for every $h = 1, \ldots, n$ we have (3.11) then (3.13) holds for H values of $z \in [0, H-1]$, otherwise it holds for at most n values.

Therefore

$$\Sigma(\eta) = O\left(pHN_k^{1/2} + n^{1/2} H^{1/2} N_k p \right).$$

Hence, as in the proof of Theorem 3.5, we obtain

$$\sum_{y \in \mathbb{F}_p^*} \sum_{z=0}^{H-1} \Delta_{y,z}^{\nu}(k) = O\left(\left(pHN_k^{3/2} + n^{1/2} H^{1/2} N_k^2 p \right) (\log p)^{\nu} \log \tau \right),$$

and the result follows. □

Corollary 3.7. *Assume that $\mathcal{Q} = \mathbb{F}_q$ and that $\mathcal{R} = \mathbb{F}_p$ where q and p are prime. Let the linear recurrence sequence u_i be purely periodic with period τ and order n and let its characteristic polynomial be irreducible over $\mathbb{F}_q$. Then for any $\delta > 0$ and a positive integer $\nu \le n$, for all $y, z \in \mathbb{F}_p^*$ except at most $O(\delta p^2)$ of them, for all $1 \le N \le \tau$ we have*

$$D_{y,z}^{\nu}(N) \le \delta^{-1} \left(\frac{1}{N^{1/2}} + \frac{n^{1/2}}{p^{1/2}} \right) (\log p)^{\nu} (\log \tau)^2.$$

4 Remarks

We remark that the result of Theorems 3.5 and 3.6 can be extended to more general classes of characteristic polynomials. However, as we have mentioned, the case of the most practical interest is $\tau = q^n - 1$ which implies that the characteristic polynomial is primitive, and thus irreducible, over $\mathbb{F}_q$, see [9, Section 3.1].

We can also extend Theorems 3.5 and 3.6 to more general finite fields and residue rings modulo squarefree numbers. Arbitrary residue rings can be studied by our method as well but in this case one obtains much weaker results because the Weil bound needs to be replaced by the Hua Loo Keng bound, see [1, 2, 3, 17], in the proof of Theorem 3.5. Also, one needs to use [8] to estimate the number of solutions of polynomial congruences (which we estimated simply as n in the proof of Theorem 3.6).

Finally we remark that one can also consider similar generators defined by a linear recurrence sequence over $\mathbb{F}_3$, which we assume to be represented by the elements $\{-1, 0, 1\}$ to avoid ring multiplication in the computation of the corresponding sequence.

Acknowledgments. The authors are grateful to Richard Brent for attracting their attention to the problem of fast generation of sequences of pseudorandom numbers.

References

[1] T. Cochrane, *Exponential sums modulo prime powers*, Acta Arith. 101 (2002), pp. 131–149.

[2] T. Cochrane and Z. Y. Zheng, *On upper bounds of Chalk and Hua for exponential sums*, Proc. Amer. Math. Soc. 129 (2001), pp. 2505–2516.

[3] ———, *A survey on pure and mixed exponential sums modulo prime powers*. Proc. Illinois Millennial Conf. on Number Theory, 1, pp. 271–300. A. K. Peters, Natick, MA, 2002.

[4] A. Conflitti and I. E. Shparlinski, *On the multidimensional distribution of the subset sum generator of pseudorandom numbers*, Math. Comp. 73 (2004), pp. 1005–1011.

[5] M. Drmota and R. Tichy, *Sequences, discrepancies and applications*. Springer-Verlag, Berlin, 1997.

[6] J. von zur Gathen and J. Gerhard, *Modern computer algebra*. Cambridge University Press, 2003.

[7] J. von zur Gathen and I. E. Shparlinski, *Predicting subset sum pseudorandom number generators*. Proc. 11th Workshop on Selected Areas in Cryptography, Waterloo, 2004, Lect. Notes in Comp. Sci. 3357, pp. 241–251. Springer-Verlag, Berlin, 2005.

[8] S. V. Konyagin and T. Steger, *On the number of solutions of polynomial congruences*, Matem. Zametki 55 (1994), pp. 73–79. (In Russian).

[9] R. Lidl and H. Niederreiter, *Finite fields*. Cambridge University Press, Cambridge, 1997.

[10] E. El Mahassni, *On the distribution of the elliptic subset sum generator of pseudorandom numbers*, Integers (to appear).

[11] A. J. Menezes, P. C. van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC Press, Boca Raton, FL, 1996.

[12] H. Niederreiter, *Random number generation and Quasi–Monte Carlo methods*. SIAM Press, 1992.

[13] Y. Pan, *Structured matrices and polynomials. Unified superfast algorithms*. Birkhäuser, Boston, MA, 2001.

[14] R. A. Rueppel, *Analysis and design of stream ciphers*. Springer-Verlag, Berlin, 1986.

[15] ———, *Contemporary cryptology: The science of information integrity*, ch. Stream ciphers, pp. 65–134, IEEE Press, NY, 1992.

[16] R. A. Rueppel and J. L. Massey, *Knapsack as a nonlinear function*. IEEE Intern. Symp. of Inform. Theory, p. 46. IEEE Press, NY, 1985.

[17] S. B. Stečkin, *An estimate of a complete rational exponential sum*, Proc. Math. Inst. Acad. Sci. USSR, Moscow 143 (1977), pp. 188–207. (In Russian).

[18] I. M. Vinogradov, *Elements of number theory*. Dover Publ., New York, 1954.

[19] A. Weil, *Basic number theory*. Springer-Verlag, New York, 1974.

Received 1 August, 2008; revised 31 May, 2009

Author information

Joachim von zur Gathen, B-IT, Universität Bonn, 53113 Bonn, Germany.
Email: gathen@bit.uni-bonn.de

Igor E. Shparlinski, Department of Computing, Macquarie University, Sydney, NSW 2109, Australia.
Email: igor@comp.mq.edu.au