MATHEMATICS OF COMPUTATION
Volume 70, Number 236, Pages 1575–1589
S 0025-5718(00)01283-7
Article electronically published on October 17, 2000

ON THE DISTRIBUTION OF THE POWER GENERATOR

JOHN B. FRIEDLANDER AND IGOR E. SHPARLINSKI

Abstract. We present a new method to study the power generator of pseudorandom numbers modulo a Blum integer m. This includes as special cases the RSA generator and the Blum–Blum–Shub generator. We prove the uniform distribution of these, provided that the period $t \geq m^{3/4+\delta}$ with fixed $\delta > 0$ and, under the same condition, the uniform distribution of a positive proportion of the leftmost and rightmost bits. This sharpens and generalizes previous results which dealt with the RSA generator, provided the period $t \geq m^{23/24+\delta}$. We apply our results to deduce that the period of the binary sequence of the rightmost bit has exponential length.

1. Introduction

Let $e \geq 2$, $m \geq 1$ and ϑ be integers such that $\gcd(\vartheta, m) = 1$. Then one can define the sequence (u_n) by the recurrence relation

$$u_n \equiv u_{n-1}^e \pmod m, \quad 0 \leq u_n \leq m-1, \qquad n = 1, 2, \dots, \tag{1.1}$$

with the *initial value* $u_0 = \vartheta$.

This sequence is known as the *power generator* of pseudorandom numbers and has many applications to cryptography, see [1, 7, 18, 29]. In two special cases $\gcd(e, \varphi(m)) = 1$, where $\varphi(m)$ is the Euler function, and $e = 2$ this sequence is known as the *RSA generator* and as the *Blum–Blum–Shub generator*, respectively.

It is obvious that the sequence (1.1) eventually becomes periodic with some period t. In this paper we assume that the sequence (u_n) is *purely periodic*. It is easy to see that if $\gcd(e, \varphi(m)) = 1$, then this is always the case, otherwise we can consider a shift of the original sequence.

Although various properties have been studied in a number of works (see [1, 2, 6, 7, 8, 12, 15, 18, 20, 29]), very few unconditional results are known. One such result is due to [6]. It is shown in that paper that the rightmost bit of the Blum–Blum–Shub generator takes values 0 and 1 almost equally often, provided that the period is large enough. In [10], by using a completely different method, relying on some results of [3, 4], it has been proved that if the period of the RSA generator is large enough, namely $t \geq m^{23/24+\delta}$ for some fixed $\delta > 0$, then the elements of this sequence are uniformly distributed modulo m and a positive proportion of the

Received by the editor April 30, 1999 and, in revised form, November 10, 1999.

2000 *Mathematics Subject Classification.* Primary 11L07, 11T71, 94A60; Secondary 11T23, 11K45.

Key words and phrases. Pseudorandom numbers, RSA generator, Blum–Blum–Shub, exponential sums.

The first author was supported in part by NSERC grant A5123 and by an NEC grant to the Institute for Advanced Study.

The second author was supported in part by ARC grant A69700294.

rightmost and leftmost bits is uniformly distributed. Lower bounds on the linear complexity of this generator have been given in [12, 28].

Here we introduce a new method which allows us to improve the results of [10], which now are nontrivial beginning with periods $t \geq m^{3/4+\delta}$, and extend them to the general power generator. This approach is also used to study the distribution of several consecutive terms of the Blum–Blum–Shub generator. In some sense this approach is based on a combination of some ideas of [12, 28] with a general approach to studying non-linear pseudorandom number generators invented in the series of papers [13, 14, 23, 24, 25, 26].

It is useful to remark that it has been shown in [11] that provided the parameters of the power generator (1.1) are selected at random, then the period t will be very close to m.

In this paper we prove the result in the most important case for applications when $m = pl$ where p and l are distinct primes of approximately the same order. Such numbers are called *Blum integers* (sometimes given with certain additional conditions such as that $p \equiv l \equiv 3 \pmod 4$). However, the same results hold for m prime, as is easy to see from our proof, and similar, but somewhat weaker, results can be obtained for arbitrary composite moduli as well.

We remark that for small e some nontrivial results can be derived from the estimates of the paper [24] which, although rather weak, apply to general polynomial generators.

Throughout the paper the implied constants in symbols "O", "$\gg$" and "$\ll$" may occasionally, where obvious, depend on the small positive parameters ε and δ and an integer parameter $\nu \geq 1$, and are absolute otherwise (we recall that $A \ll B$ and $B \gg A$ are equivalent to $A = O(B)$). Moreover, any expression involving ε, for example a bound of the form $A \ll B^{\varepsilon}$, means that this holds for any $\varepsilon > 0$, not just for some fixed value. The same convention is not used for any other variable, for example δ.

Let $\omega(k)$ denote the number of distinct prime divisors of an integer $k \geq 1$. We use the well-known bounds

$$\omega(k) \ll \frac{\log k}{\log\log(k+2)} \qquad \text{and} \qquad \varphi(k) \gg \frac{k}{\log\log(k+2)}. \tag{1.2}$$

Also, $\log z$ denotes the binary logarithm.

2. Exponential sums

We define exponential sums

$$S_a = \sum_{n=1}^{t} \mathbf{e}_m\left(au_n\right),$$

where

$$\mathbf{e}_d(z) = \exp(2\pi i z/d).$$

We obtain a nontrivial upper bound for the sums S_a and derive (see Theorem 3.2) the uniformity of distribution modulo m of the elements u_n, $n = 1, \dots, t$, provided that $t > m^{3/4+\delta}$ with a fixed positive δ.

These exponential sums enter into our problem by means of the following well-known basic identity (see Problem 11.a to Chapter 3 of [30]).

Lemma 2.1. *For any integer u*

$$\sum_{\lambda=0}^{m-1} \mathbf{e}_m(\lambda u) = \begin{cases} 0, & \text{if } u \not\equiv 0 \pmod m; \\ m, & \text{if } u \equiv 0 \pmod m. \end{cases}$$

We also need the following estimate (see Problem 11.c to Chapter 3 of [30]).

Lemma 2.2. *For any integer $H \ge 0$,*

$$\sum_{a=1}^{m-1} \left| \sum_{u=0}^{H} \mathbf{e}_m(au) \right| = O(m \log m).$$

Our strongest tool is the Weil bound for exponential sums which we present in the following form (see Chapter 5 of [19]).

Lemma 2.3. *For any prime p and any polynomial $f(X) \in \mathbb{Z}[X]$ of degree $d \ge 1$ which is not identical to a constant modulo p, the bound*

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

holds.

Below we use the following well-known consequence of the sieve of Eratosthenes.

Lemma 2.4. *For any integers $q, J \ge 1$,*

$$\sum_{\substack{j=1 \\ \gcd(j,q)=1}}^{J} 1 = \frac{\varphi(q)}{q} J + O(2^{\omega(q)}).$$

Proof. Indeed, using the Möbius function $\mu(d)$ over the divisors of q to detect the co-primality condition and interchanging the order of summation, we obtain the Legendre formula

$$\sum_{\substack{j=1 \\ \gcd(j,q)=1}}^{J} 1 = \sum_{d|q} \mu(d) \left\lfloor \frac{J}{d} \right\rfloor = J \sum_{d|q} \frac{\mu(d)}{d} + O\left(\sum_{d|q} |\mu(d)| \right)$$

from which the result follows at once (see Section 4 of Chapter 2 of [30]). □

We reduce the problem of estimating sums S_a to certain sums with polynomials. In [24] the above bound has been used to obtain a nontrivial estimate for very general pseudorandom number generators. Nevertheless, for our purposes a direct application of the bound would give us rather weak results, the problem being that the dependence therein on the degree is not very good and the polynomials to which our argument naturally leads are in some cases of very high degree. Our next lemma, combinatorial in nature, shows that in the special case of the power generator there exists a subset of these polynomials, which contains sufficiently many that we may restrict our consideration to it and simultaneously consists of polynomials of sufficiently low degree that, when applied to them, the Weil bound will be a strong one.

Lemma 2.5. *Let τ be the multiplicative order of e modulo an integer $T \geq 1$. Then for any fixed $\delta > 0$ and any integer $h \geq T^{\delta}$ there exists an integer r with*

$$\gcd(r,T) = 1$$

and such that the congruence

$$re^k \equiv y \pmod{T}, \qquad 1 \leq k \leq \tau,\ 0 \leq y \leq h-1,$$

has

$$L_r(h) \gg \frac{\tau h}{T}$$

solutions.

Proof. For each $k = 1, \dots, \tau$ and every y, $0 \leq y \leq h-1$ with $\gcd(y,T) = 1$, the integer r, $1 \leq r \leq T$, such that the above congruence holds is uniquely determined and satisfies $\gcd(r,T) = 1$ since $\gcd(e,T) = 1$. Hence we have

$$\sum_{\substack{r=1\\ \gcd(r,T)=1}}^{T} L_r(h) \gg \tau \sum_{\substack{y=0\\ \gcd(y,T)=1}}^{h-1} 1.$$

Since $h \geq T^{\delta}$, by Lemma 2.4 and by (1.2) we have

$$\sum_{\substack{y=0\\ \gcd(y,T)=1}}^{h-1} 1 \gg \frac{\varphi(T)h}{T},$$

and substituting this we get

$$\sum_{\substack{r=1\\ \gcd(r,T)=1}}^{T} L_r(h) \gg \frac{\tau\varphi(T)h}{T}.$$

Hence for some choice of r we have the required lower bound for $L_r(h)$. □

Lemma 2.6. *If the sequence (u_n), given by (1.1), is purely periodic with period t, then for any integers $\lambda \geq 0$ and $\mu \geq 1$ the sequence $(u_{\lambda+\mu n})$ is purely periodic with period $t/\gcd(\mu,t)$.*

Proof. Let T be the multiplicative order of ϑ modulo m. Because (u_n) is purely periodic with period t, then

$$\vartheta \equiv u_0 \equiv u_t = \vartheta^{e^t} \pmod{m}.$$

Hence $e^t \equiv 1 \pmod{T}$. Therefore we conclude that $\gcd(e,T) = 1$ and that t is the multiplicative order of e modulo T. Put

$$\rho = \vartheta^{e^{\lambda}} \qquad \text{and} \qquad f = e^{\mu}.$$

We see that the multiplicative order of ρ modulo m is T as well and the multiplicative order of f modulo T is $t/\gcd(\mu,t)$. □

Now we are prepared to formulate our main estimates. However, for notational simplicity, we first define two functions which will appear repeatedly in all of our main estimates. For $\nu \geq 1$, an integer parameter, we define

$$\alpha(\nu) = \frac{2\nu+1}{2\nu(\nu+1)} \qquad \text{and} \qquad \beta(\nu) = \frac{3\nu+2}{4\nu(\nu+1)}. \tag{2.1}$$

Theorem 2.7. *Let $m = pl$ where p and l are two distinct primes with*

$$\gcd(p-1, l-1) = \Delta.$$

If the sequence (u_n), given by (1.1), is purely periodic with period t, then for any integer $\nu \geq 1$ the bound

$$\max_{\gcd(a,m)=1} |S_a| \ll \Delta^{\alpha(\nu)} t^{1-\alpha(\nu)} m^{\beta(\nu)}$$

holds with $\alpha(\nu)$, $\beta(\nu)$ given by (2.1).

Proof. Denote by t_p and t_l the periods of (u_n) modulo p and l, respectively.

First of all we consider the case $\gcd(t_p, t_l) = 1$; thus $t = t_p t_l$. Moreover in this case we can find integers q_p and q_l such that

$$q_p \equiv 1 \pmod{t_p}, \qquad q_p \equiv 0 \pmod{t_l}$$

and

$$q_l \equiv 0 \pmod{t_p}, \qquad q_l \equiv 1 \pmod{t_l}.$$

We can also find integers Q and R such that $Ql + Rp = 1$. Then we have

$$\begin{aligned} S_a &= \sum_{n=0}^{t_p-1} \sum_{k=0}^{t_l-1} \mathbf{e}_m \left(a\vartheta^{e^{nq_p+kq_l}} \right) = \sum_{n=0}^{t_p-1} \sum_{k=0}^{t_l-1} \mathbf{e}_m \left(a(Ql+Rp)\vartheta^{e^{nq_p+kq_l}} \right) \\ &= \sum_{n=0}^{t_p-1} \sum_{k=0}^{t_l-1} \mathbf{e}_p \left(aQ\vartheta^{e^{nq_p+kq_l}} \right) \mathbf{e}_l \left(aR\vartheta^{e^{nq_p+kq_l}} \right). \end{aligned}$$

Remarking that

$$\vartheta^{e^{nq_p+kq_l}} \equiv \vartheta^{e^{nq_p}} \pmod{p} \qquad \text{and} \qquad \vartheta^{e^{nq_p+kq_l}} \equiv \vartheta^{e^{kq_l}} \pmod{l},$$

we derive

$$S_a = \sum_{n=0}^{t_p-1} \mathbf{e}_p \left(aQ\vartheta^{e^{nq_p}} \right) \sum_{k=0}^{t_l-1} \mathbf{e}_l \left(aR\vartheta^{e^{kq_l}} \right) = \sum_{n=0}^{t_p-1} \mathbf{e}_p \left(A\vartheta^{e^n} \right) \sum_{k=0}^{t_l-1} \mathbf{e}_l \left(B\vartheta^{e^k} \right),$$

where $A = aQ$ and $B = aR$. We are going to apply Lemma 2.5 with modulus T being the multiplicative order of ϑ modulo p. As explained in the proof of Lemma 2.6 the order of e modulo T is just $\tau = t_p$. For some $\delta > 0$ and $h \geq T^\delta$ to be chosen later we select r as in Lemma 2.5. Let $\mathcal{L}$ denote the set of k which satisfy the corresponding congruence. Put $L = |\mathcal{L}|$. Then

$$\sum_{n=0}^{t_p-1} \mathbf{e}_p \left(A\vartheta^{e^n} \right) = \frac{1}{L} \sum_{k\in \mathcal{L}} \sum_{n=0}^{t_p-1} \mathbf{e}_p \left(A\vartheta^{e^{n+k}} \right).$$

By the Hölder inequality we have

$$\begin{aligned} \left| \sum_{n=0}^{t_p-1} \mathbf{e}_p \left(A\vartheta^{e^n} \right) \right|^{2\nu} &\leq L^{-2\nu} t_p^{2\nu-1} \sum_{n=0}^{t_p-1} \left| \sum_{k\in \mathcal{L}} \mathbf{e}_p \left(A\vartheta^{e^{n+k}} \right) \right|^{2\nu} \\ &= L^{-2\nu} t_p^{2\nu-1} \sum_{n=0}^{t_p-1} \left| \sum_{k\in \mathcal{L}} \mathbf{e}_p \left(A\vartheta^{e^n e^k} \right) \right|^{2\nu}. \end{aligned}$$

Let $d = (p-1)/T$. Obviously, the powers ϑ^{e^n}, $n = 0, \ldots, t_p-1$, are pairwise distinct and are each of the form x^d modulo p. Moreover, for each n there are precisely d

values of $x = 1, \dots, p-1$ which give rise to this value of ϑ^{e^n}. Thus by replacing ϑ^{e^n} by x^d we may write this last sum over n as d^{-1} times the corresponding sum over x, where x runs over a certain subset of the nonzero residue classes modulo p. Using positivity, we may majorize this last sum by the same sum over all classes modulo p. In this way we obtain

$$\begin{aligned}
&\left|\sum_{n=0}^{t_p-1} \mathbf{e}_p\left(A\vartheta^{e^n}\right)\right|^{2\nu} \\
&\quad\le L^{-2\nu} d^{-1} t_p^{2\nu-1} \sum_{x=0}^{p-1} \left|\sum_{k\in\mathcal{L}} \mathbf{e}_p\left(Ax^{de^k}\right)\right|^{2\nu} \\
&\quad= L^{-2\nu} d^{-1} t_p^{2\nu-1} \sum_{j_1,\dots,j_\nu\in\mathcal{L}} \ \sum_{k_1,\dots,k_\nu\in\mathcal{L}} \\
&\qquad\times \sum_{x=0}^{p-1} \mathbf{e}_p\left(A\left(x^{de^{j_1}} + \ldots + x^{de^{j_\nu}} - x^{de^{k_1}} - \ldots - x^{de^{k_\nu}}\right)\right) \\
&\quad= L^{-2\nu} d^{-1} t_p^{2\nu-1} \sum_{j_1,\dots,j_\nu\in\mathcal{L}} \ \sum_{k_1,\dots,k_\nu\in\mathcal{L}} \\
&\qquad\times \sum_{x=0}^{p-1} \mathbf{e}_p\left(A\left(x^{dre^{j_1}} + \ldots + x^{dre^{j_\nu}} - x^{dre^{k_1}} - \ldots - x^{dre^{k_\nu}}\right)\right)
\end{aligned}$$

because $\gcd(r,T) = 1$. For the case that $(k_1, ..., k_\nu)$ is a permutation of $(j_1, ..., j_\nu)$, we must use the trivial bound and this gives a contribution $L^\nu p$. In case this does not happen (there are at most $L^{2\nu}$ ways) the inner sum above is a character sum with a polynomial of degree at most dh. By Lemma 2.3 each of these terms contributes at most $dhp^{\frac{1}{2}}$. Thus

$$\left|\sum_{n=0}^{t_p-1} \mathbf{e}_p\left(A\vartheta^{e^n}\right)\right|^{2\nu} \ll L^{-2\nu} d^{-1} t_p^{2\nu-1} \left(L^\nu p + L^{2\nu} dhp^{\frac{1}{2}}\right)$$

and so

$$\begin{aligned}
\left|\sum_{n=0}^{t_p-1} \mathbf{e}_p\left(A\vartheta^{e^n}\right)\right| &\ll t_p^{1-1/2\nu}\left(L^{-1/2}p^{1/2\nu}d^{-1/2\nu} + h^{1/2\nu}p^{1/4\nu}\right) \\
&\ll t_p^{1-1/2\nu}\left(L^{-1/2}T^{1/2\nu} + h^{1/2\nu}p^{1/4\nu}\right).
\end{aligned}$$

By Lemma 2.5 we have $L \gg t_p h/T$. We substitute this in, use the trivial bound $T \le p$, and equalize by choosing

$$h = \left\lceil p^{(2\nu+1)/2(\nu+1)} t_p^{-\nu/(\nu+1)} \right\rceil \ge p^{1/2(\nu+1)},$$

which thus satisfies $h \ge T^\delta$ with $\delta = 1/2(\nu+1)$. After a simple computation we obtain

$$\left|\sum_{n=0}^{t_p-1} \mathbf{e}_p\left(A\vartheta^{e^n}\right)\right| \ll t_p^{1-\alpha(\nu)} p^{\beta(\nu)},$$

and similarly

$$\left|\sum_{k=0}^{t_l-1} \mathbf{e}_l \left(B\vartheta^{e^k}\right)\right| \ll t_l^{1-\alpha(\nu)} l^{\beta(\nu)}$$

from which the result follows, provided that $\gcd(t_p, t_l) = 1$.

In the general case we put $\mu = \gcd(t_p, t_l)$ and remark that

$$S_a = \sum_{n=0}^{t-1} \mathbf{e}_m (au_n) = \sum_{\lambda=1}^{\mu} \sum_{n=0}^{t/\mu-1} \mathbf{e}_m (au_{\lambda+n\mu}) .$$

Using Lemma 2.6 one easily verifies that $(u_{\lambda+n\mu})$ has relatively prime periods modulo p and l (because they are divisors of t_p/μ and t_l/μ, respectively). By the above bound we obtain

$$S_a \ll \mu(t/\mu)^{1-\alpha(\nu)} m^{\beta(\nu)}. \tag{2.2}$$

Now from $\mu \le \gcd(p-1, l-1) = \Delta$ the result follows. □

It is easy to check that for a given $\delta > 0$ one may, by selecting sufficiently large ν, obtain a bound which is nontrivial for all $t \ge \Delta m^{3/4+\delta/2}$. If also $\gcd(p-1, l-1) \ll m^{\delta/2}$, then it suffices to take $t \ge m^{3/4+\delta}$. Under these conditions an asymptotically optimal choice for small δ is $\nu = \lceil 1/2\delta \rceil$, which gives the bound $O(t^{1-c\delta^2})$ with an absolute constant $c > 0$. For sufficiently small δ, this holds with any value $c < 2/3$. On the other hand, for large values of t the choice $\nu = 1$ becomes optimal, producing the bound

$$\max_{\gcd(a,m)=1} |S_a| \ll \Delta^{3/4} t^{1/4} m^{5/8}.$$

In the most interesting case that $t \gg m^{1-\delta/2}$ and $\Delta \ll m^{\delta/2}$ with a small δ, selecting $\nu = 1$ we obtain

$$\max_{\gcd(a,m)=1} |S_a| \ll t^{7/8+\delta}. \tag{2.3}$$

Moreover, as the results of [11] show, for a random choice of the parameters of the generator this case occurs almost always.

We note that although the condition $\gcd(p-1, l-1) \ll m^{\varepsilon}$ is satisfied for almost all pairs of primes p and l, in fact our method works without any restriction on $\gcd(p-1, l-1)$. Obviously, the parameter μ in the proof of Theorem 2.7 can be estimated from the inequality $t \le m/\mu$ which, being substituted in (2.2) leads to the estimate

$$\max_{\gcd(a,m)=1} |S_a| \ll t^{1-(2\nu+1)/\nu(\nu+1)} m^{(7\nu+4)/4\nu(\nu+1)},$$

which holds for any p and l. One can also modify the scheme a little to prove results under the natural assumption that the primes p and l are about the same order. For example, for the Blum integers $m = pl$ with $p < l \le p^{1+\varepsilon}$ and $p \equiv l \equiv 3 \pmod 4$ we have

$$\max_{\gcd(a,m)=1} |S_a| \ll t^{1-(3\nu+1)/2\nu(2\nu+1)} m^{(5\nu+2)/4\nu(2\nu+1)+\varepsilon}.$$

Several more statements of this kind can be proved as well; however in the case $\gcd(p-1, l-1) \ll m^{\varepsilon}$ the bound of Theorem 2.7 supersedes them all.

To study the distribution of s-tuples $(u_n, \dots, u_{n+s-1})$ we need to estimate more general sums. For an integer vector $\mathbf{a} = (a_0, \dots, a_{s-1}) \in \mathbb{Z}^s$ we define the exponential sum

$$S_{\mathbf{a}} = \sum_{n=1}^{t} \mathbf{e}_m \left(\sum_{i=0}^{s-1} a_i u_{n+i} \right).$$

Although in the next result we estimate these sums only for the special case of the Blum–Blum–Shub generator, it can be extended to any power generator with a small value of e.

Theorem 2.8. *Let $m = pl$ where p and l are two distinct primes with*

$$\gcd(p-1, l-1) = \Delta.$$

If the sequence (u_n), given by (1.1) *with $e = 2$, is purely periodic with period t, then for any integer $\nu \ge 1$ and any dimension $s \ge 1$, the bound*

$$\max_{\gcd(a_0,\dots,a_{s-1},m)=1} |S_{\mathbf{a}}| \ll 2^{s\nu/(\nu+1)} \Delta^{\alpha(\nu)} t^{1-\alpha(\nu)} m^{\beta(\nu)}$$

holds with $\alpha(\nu)$, $\beta(\nu)$ given by (2.1).

Proof. We put

$$h = \left\lceil 2^{-2s\nu/(\nu+1)} p^{(2\nu+1)/2(\nu+1)} t_p^{-\nu/(\nu+1)} \right\rceil \gg p^{1/2(\nu+1)}$$

and proceed as in Theorem 2.7 getting the sum

$$\sum_{x=0}^{p-1} \mathbf{e}_p \left(A \sum_{i=0}^{s-1} \sum_{\eta=1}^{\nu} a_i \left(x^{dr2^{j_\eta+i}} - x^{dr2^{k_\eta+i}} \right) \right),$$

which is an exponential sum with a polynomial of degree at most $2^{s-1}h$. Continuing as in the proof of Theorem 2.7, after simple calculations we obtain the desired result. □

We remark that both Theorem 2.7 and Theorem 2.8 apply to prime moduli m as well. In fact in this case even stronger results can be obtained, see [9].

3. Leftmost and Rightmost Bits

Here we show that a positive proportion of the leftmost and rightmost bits of (u_n) are uniformly distributed. Let σ be a binary string of length k. Denote by $L(\sigma)$ the number of $n = 1, \dots, t$, such that σ is the string of the k least significant bits of u_n.

Theorem 3.1. *Let $m = pl$, where p and l are two distinct odd primes with*

$$\gcd(p-1, l-1) = \Delta.$$

If the sequence (u_n) given by (1.1) *is purely periodic with period t, then for any integer $\nu \ge 1$ the bound*

$$\sup_{\sigma} \left| L(\sigma) - t2^{-k} \right| \ll \Delta^{\alpha(\nu)} t^{1-\alpha(\nu)} m^{\beta(\nu)} \log m$$

holds with $\alpha(\nu)$, $\beta(\nu)$ given by (2.1) *and where the supremum is taken over all binary strings σ of length k.*

Proof. We denote by $\overline{\sigma}$ the integer whose binary representation coincides with σ and put $K = 2^k$, $H = \lfloor (m-1-\overline{\sigma})/K \rfloor$. We remark that $L(\sigma)$ is equal to the number $W(\sigma)$ of solutions of the congruence

$$u_n \equiv Kx + \overline{\sigma} \pmod m, \qquad 1 \le n \le t,\ 0 \le x \le H.$$

Thus, using Lemma 2.1 we write

$$\begin{aligned} W(\sigma) &= \frac{1}{m} \sum_{n=1}^{t} \sum_{x=0}^{H} \sum_{a=0}^{m-1} \mathbf{e}_m \left(a \left(u_n - Kx - \overline{\sigma} \right) \right) \\ &= \frac{1}{m} \sum_{a=0}^{m-1} \mathbf{e}_m \left(-a\overline{\sigma} \right) S_a \sum_{x=0}^{H} \mathbf{e}_m(-aKx). \end{aligned}$$

The term corresponding to $a = 0$ equals

$$t(H+1)m^{-1} = t2^{-k} + O(1),$$

which gives the main term of the desired formula, apart from an admissible error. To estimate the contribution R of the remaining terms, we apply Theorem 2.7 getting

$$\begin{aligned} R \quad &\ll \quad \Delta^{\alpha(\nu)} t^{1-\alpha(\nu)} m^{-1+\beta(\nu)} \sum_{\substack{a=1 \\ \gcd(a,m)=1}}^{m-1} \left| \sum_{x=0}^{H} \mathbf{e}_m(-aKx) \right| \\ &\quad + m^{-1} \sum_{a=1}^{l-1} |S_{ap}| \left| \sum_{x=0}^{H} \mathbf{e}_l(-aKx) \right| + m^{-1} \sum_{a=1}^{p-1} |S_{al}| \left| \sum_{x=0}^{H} \mathbf{e}_p(-aKx) \right|. \end{aligned}$$

We still need to estimate the sums $|S_{ap}|$ and $|S_{al}|$. Let t_p be the period of the sequence (u_n) modulo p so that $t \le t_p(l-1)$. It is easy to check that the bound of Theorem 2.7 applies to the case when m is prime as well, and moreover the factor related to the $\gcd(p-1, l-1)$ does not appear in this case. Therefore,

$$S_{al} = \frac{t}{t_p} \sum_{n=1}^{t_p} \mathbf{e}_p \left(a u_n \right) \ll l t_p^{1-\alpha(\nu)} p^{\beta(\nu)}$$

if $\gcd(a,p) = 1$. A similar estimate also holds for S_{ap}.

Remarking that $\gcd(K,m) = 1$, we see that $-aK$ can be replaced by just a. Therefore, applying Lemma 2.2, after some simple calculations we obtain the desired estimate. □

Virtually the same proof yields the same result for the most significant bits. One simply replaces $Kx + \overline{\sigma}$ by $x + M\overline{\sigma}$ for suitable M in the above congruence for u_n. Since the most significant bits of a number are the ones most responsible for locating it as a point on the line, this case may also be formulated somewhat differently. For an interval $\mathcal{I} = [r, r+h-1]$, where $0 \le r \le r+h-1 \le m-1$ of length $\#\mathcal{I} = h$, we denote by $N(\mathcal{I})$ the number of $n = 1, \ldots, t$ for which u_n belongs to the interval $\mathcal{I}$. In this form there is an alternative well-known way of deriving the result from the exponential sum bound (see for example [5]) which, however, does not apply to the least significant bits. The result is

Theorem 3.2. *Let $m = pl$, where p and l are two distinct odd primes with*

$$\gcd\left(p-1, l-1\right) = \Delta.$$

If the sequence (u_n) given by (1.1) *is purely periodic with period t, then for any integer $\nu \ge 1$ the bound*

$$\sup_{\mathcal{I}\subseteq[0,m-1]} \left| N(\mathcal{I}) - \frac{t}{m}\#\mathcal{I} \right| \ll \Delta^{\alpha(\nu)} t^{1-\alpha(\nu)} m^{\beta(\nu)} \log m$$

holds with $\alpha(\nu)$, $\beta(\nu)$ given by (2.1).

In the case $\gcd(p-1, l-1) \ll m^{\varepsilon}$, the error terms in Theorems 3.1 and 3.2 take the form $t^{1-(2\nu+1)/2\nu(\nu+1)} m^{(3\nu+2)/4\nu(\nu+1)+\varepsilon}$.

With very little change we can use Theorem 2.8 to also study the multidimensional distribution of the Blum–Blum–Shub generator. Let $\Sigma = (\sigma_0, \dots, \sigma_{s-1})$ be a collection of s binary strings of length k. Denote by $L(\Sigma)$ the number of $n = 1, \dots, t$, such that σ_i is the string of the k least significant bits of u_{n+i}, $i = 0, \dots, s-1$.

Combining Theorem 2.8 with the same arguments used in the proof of Theorem 3.1, we obtain

Theorem 3.3. *Let $m = pl$, where p and l are two distinct odd primes with*

$$\gcd(p-1, l-1) = \Delta.$$

There exists an absolute constant $C > 0$ such that if the sequence (u_n), given by (1.1) *with $e = 2$, is purely periodic with period t, then for any integer $\nu \ge 1$ and every integer s the bound*

$$\sup_{\Sigma} \left| L(\Sigma) - t2^{-ks} \right| \ll \Delta^{\alpha(\nu)} t^{1-\alpha(\nu)} m^{\beta(\nu)} (C \log m)^s$$

holds with $\alpha(\nu)$, $\beta(\nu)$ given by (2.1) *and where the supremum is taken over all collections Σ of s binary strings of length k.*

For an s-dimensional box

$$\mathcal{B} = [r_1, r_1 + h_1 - 1] \times \cdots \times [r_s, r_s + h_s - 1],$$

where $0 \le r_i \le r_i + h_i - 1 \le m - 1$, $i = 1, \dots, s$, of size $\#\mathcal{B} = h_1 \cdots h_s$, we denote by $N(\mathcal{B})$ the number of integers $n = 1, \dots, t$ for which the s-tuple $(u_n, \dots, u_{n+s-1})$ belongs to the box $\mathcal{B}$.

Accordingly, in this case Theorem 2.8 yields

Theorem 3.4. *Let $m = pl$, where p and l are two distinct primes with*

$$\gcd(p-1, l-1) = \Delta.$$

There exists an absolute constant $C > 0$ such that if the sequence (u_n), given by (1.1) *with $e = 2$, is purely periodic with period t, then for any integer $\nu \ge 1$ and every integer s the bound*

$$\sup_{\mathcal{B}\subseteq[0,m-1]^s} \left| N(\mathcal{B}) - \frac{t}{m}\#\mathcal{B} \right| \ll \Delta^{\alpha(\nu)} t^{1-\alpha(\nu)} m^{\beta(\nu)} (C \log m)^s$$

holds with $\alpha(\nu)$, $\beta(\nu)$ given by (2.1).

We see that if $\gcd(p-1, l-1) \ll m^{\delta/2}$ and $t \ge m^{3/4+\delta}$, then for sufficiently large ν depending on δ the bounds of Theorems 3.3 and 3.4 are nontrivial, that is of the form $o(t)$, for all $s \le c\delta^2 \log m / \log\log m$ with an absolute constant $c > 0$. In particular, under this condition Theorem 3.4 gives the statement of uniform distribution of s-tuples produced by the Blum–Blum–Shub generator.

4. Period of the Blum–Blum–Shub bit generator

Let us consider the binary sequence (ξ_n) where ξ_n is the rightmost bit of u_n. We note that many cryptographic applications make use of this sequence instead of the original sequence (u_n) (see [1, 6, 7, 8, 15, 18, 20, 29]). On the other hand, although the period t of the sequence (u_n) admits a reasonably simple number theoretic characterization via the Carmichael function, the period τ of the sequence (ξ_n) does not seem to be easy to evaluate. Aside from the trivial property $\tau | t$ and a certain lower bound for the Blum–Blum–Shub generator, valid for some very special moduli m (see Section 14.8 of [7]), nothing is known about the period τ.

Here, using Theorem 3.3 we easily derive that if, for some fixed $\delta > 0$, the primes p and l and the period t satisfy the conditions $\gcd(p-1, l-1) \ll m^{\delta/2}$ and $t \geq m^{3/4+\delta}$, then τ is exponentially large in the bit size of m. Indeed, taking $k = 1$ in Theorem 3.3, we have already noted that $L(\Sigma) > 0$ provided that $s \leq c(\delta) \log m / \log\log m$ for some constant $c(\delta) > 0$. For such s there are at least 2^s distinct s-tuples of the form $(\xi_n, \ldots, \xi_{n+s-1})$, so that, choosing s as large as permitted, we find

$$\tau \geq 2^s \gg 2^{c(\delta) \log m / \log\log m}.$$

Now we improve this bound by using a modified approach which allows us to show that all $L(\Sigma) > 0$ for a slightly larger value of s by counting, instead of $L(\Sigma)$ itself, a weighted sum over the integers n which contribute to it.

Theorem 4.1. *Let $m = pl$, where p and l are distinct odd primes with*

$$\gcd(p-1, l-1) = \Delta.$$

If the sequence (u_n), given by (1.1) with $e = 2$, is purely periodic with period $t \geq \Delta m^{3/4+\delta}$ for some fixed $\delta > 0$, then there exists a constant $\gamma(\delta) > 0$ such that

$$\tau \gg m^{\gamma(\delta)}$$

holds.

Proof. Using Theorem 2.8 with sufficiently large ν, we derive that there exists $\eta > 0$ such that

$$S_{\mathbf{a}} \ll 2^s t^{1-\eta} \tag{4.1}$$

for any vector $\mathbf{a} = (a_0, \ldots, a_{s-1})$ with $\gcd(a_0, \ldots, a_{s-1}, m) = 1$. Let t_p and t_l be the periods of the sequence (u_n) modulo p and l, respectively. It is also easy to see that the same considerations also give the bounds

$$S_{l\mathbf{a}} \ll 2^s l t_p^{1-\eta} \qquad \text{and} \qquad S_{p\mathbf{a}} \ll 2^s p t_l^{1-\eta} \tag{4.2}$$

for vectors with $\gcd(a_0, \ldots, a_{s-1}, p) = 1$ and $\gcd(a_0, \ldots, a_{s-1}, l) = 1$, respectively.

Let $\sigma = (\sigma_1, \ldots, \sigma_s)$ be a binary string of length $s \geq 1$. Denote by $Q(\sigma)$ the number of $n = 1, \ldots, t$, such that $(\xi_n, \ldots, \xi_{n+s-1}) = \sigma$.

Put $H = \lfloor m/4 \rfloor$. Denote by $W(\sigma)$ the number of solutions of the system of congruences

$$u_{n+i} \equiv 2(H + x_i - y_i) + \sigma_i \pmod{m}, \qquad 0 \leq i \leq s-1,$$

where

$$1 \leq n \leq t, \qquad 0 \leq x_0, \ldots, x_{s-1}, y_0, \ldots, y_{s-1} \leq H - 1.$$

It is obvious that if $W(\sigma) > 0$, then $Q(\sigma) > 0$.

Using Lemma 2.1 we obtain

$$\begin{aligned}
W(\sigma) &= \frac{1}{m^s}\sum_{n=1}^{t}\sum_{x_0,\dots,x_{s-1}=0}^{H-1}\sum_{y_0,\dots,y_{s-1}=0}^{H-1} \\
&\qquad \times \sum_{a_0,\dots,a_{s-1}=0}^{m-1} \mathbf{e}_m\left(\sum_{i=0}^{s-1} a_i\left(u_{n+i}-2x_i+2y_i-2H-\sigma_i\right)\right) \\
&= \frac{1}{m^s}\sum_{a_0,\dots,a_{s-1}=0}^{m-1} \mathbf{e}_m\left(-\sum_{i=0}^{s-1} a_i\left(2H+\sigma_i\right)\right) S_{\mathbf{a}} \\
&\qquad \times \sum_{x_0,\dots,x_{s-1}=0}^{H-1}\sum_{y_0,\dots,y_{s-1}=0}^{H-1} \mathbf{e}_m\left(-2\sum_{i=0}^{s-1} a_i\left(x_i-y_i\right)\right) \\
&= \frac{1}{m^s}\sum_{a_0,\dots,a_{s-1}=0}^{m-1} \mathbf{e}_m\left(-\sum_{i=0}^{s-1} a_i\left(2H+\sigma_i\right)\right) S_{\mathbf{a}} \prod_{i=0}^{s-1}\left|\sum_{x_i=0}^{H-1}\mathbf{e}_m\left(2a_ix_i\right)\right|^2 .
\end{aligned}$$

The term corresponding to $a_0 = \cdots = a_{s-1} = 0$ equals $tH^{2s}m^{-s}$. To estimate the contribution of the terms with $\gcd(a_0,\dots,a_{s-1},m) = 1$, we apply (4.1) together with the identity (m is odd)

$$\sum_{a_0,\dots,a_{s-1}=0}^{m-1}\prod_{i=0}^{s-1}\left|\sum_{x_i=0}^{H-1}\mathbf{e}_m\left(2a_ix_i\right)\right|^2 = \left(\sum_{a=0}^{m-1}\left|\sum_{x=0}^{H-1}\mathbf{e}_m\left(ax\right)\right|^2\right)^s = m^sH^s.$$

To estimate the contribution of the terms with $\gcd(a_0,\dots,a_{s-1},m) = p$ and $\gcd(a_0,\dots,a_{s-1},m) = l$, we use (4.2) together with the analogue of the above identity modulo l and p, respectively. This gives

$$W(\sigma) = tH^{2s}m^{-s} + O(2^st^{1-\eta}H^s) = tH^s\left(H^sm^{-s} + O(2^st^{-\eta})\right).$$

We can assume that $m \geq 15$, thus

$$H^sm^{-s} \geq \left(\frac{m-3}{4m}\right)^s \geq 5^{-s}.$$

Therefore there is a constant $\gamma(\delta) > 0$ such that $W(\sigma) > 0$, provided that we have $s \leq \gamma(\delta)\log m$, and from this the result follows. □

Although in this section we have been restricting to $k = 1$ we may remark that a result corresponding to the above holds for general k. It is easy to see that if $t \ll m \ll t^{1+\varepsilon}$, then the bound (2.3) implies that $\tau \geq m^{1/24-\varepsilon}$. As we have mentioned, the results of [11] imply that this inequality is most typical.

This also implies a lower bound for the linear complexity $\mathcal{L}$ of the sequence (ξ_n) over $\mathbb{F}_2$. We recall that the linear complexity of a periodic sequence is defined as the smallest possible order of a linear recurrence relation which this sequence satisfies, (see [7, 20, 27]). It is clear that all $\mathcal{L}$-tuples $(\xi_n,\dots,\xi_{n+\mathcal{L}-1})$, $n = 1,\dots,\tau$, are pairwise distinct. Hence $\tau \leq 2^{\mathcal{L}}$ and from Theorem 4.1 we derive the bound

$$\mathcal{L} \geq \gamma(\delta)\log m. \tag{4.3}$$

5. Remarks

As we have already mentioned, these results also hold for m prime (with a slightly simpler version of the same proof and without the factor depending on Δ) and similar results can be obtained for the case of a general square-free modulus m.

Theorems 2.8, 3.3 and 3.4 can be extended to any small value of e. On the other hand, it is not clear how to study the joint distribution of s-tuples $(u_n, \dots, u_{n+s-1})$ if e is large. Even the case of pairs (u_n, u_{n+1}) is of interest.

It would be important to replace the logarithmic lower bound (4.3) by a stronger result. The linear complexity of the original sequence (u_n) has been studied in [12, 28], with bounds obtained much stronger than (4.3), but the method of these papers cannot be applied to the sequence (ξ_n).

It would also be interesting to extend the results of this paper to the case of the *exponential generator*

$$v_n \equiv g^{v_{n-1}} \pmod p, \quad 0 \le v_n \le p-1, \qquad n = 1, 2, \dots ,$$

where g is a primitive root modulo a prime p, which also has numerous cryptographic applications [7, 18]. Although we have not been able to obtain nontrivial results about the distribution of this generator, we remark that exponential sums still help to extract some nontrivial information about it. Indeed, let us consider the sequence of binary strings $\left(\sigma_n^{(k)}\right)$ generated by the k rightmost bits of v_n, $n = 1, 2, \dots$. Let the sequence (v_n) be purely periodic with period t, and let τ be the period of the sequence $\left(\sigma_n^{(k)}\right)$. Because there are at most $O(p2^{-k})$ integers v with given k rightmost bits, then obviously $\tau \ge 2^k t/p$. We show that exponential sums implies a stronger bound. Indeed, using the well-known bound (see [16, 17, 21, 22])

$$\max_{\gcd(a,p)=1} \left| \sum_{x=1}^{p-1} \mathbf{e}_p\left(ag^x\right) \mathbf{e}_{p-1}(bx) \right| \ll p^{1/2} \log p,$$

one derives that the number of x, $0 \le x \le p-1$ such that both x and the remainder of g^x modulo p have given k rightmost bits, is $p2^{-2k} + O(p^{1/2} \log p)$. On the other hand, all $v_{k\tau}$ for $k = 0, \dots , t/\tau$, have the same k rightmost bits, and the same is true for $v_{k\tau+1} \equiv g^{v_{k\tau}} \pmod p$ as well. Therefore, if $k \le (0.25 - \varepsilon) \log p$ (that is, if we use less than a quarter of bits of v_n), then $\tau \gg 2^{2k} t/p$. This, rather weak, improvement of the trivial bound makes us mildly optimistic about the possibility of obtaining more interesting results about the exponential generator.

Acknowledgment

We wish to thank the referee for a patient and careful reading which unearthed a number of minor inaccuracies in the original version of the paper.

References

[1] L. Blum, M. Blum and M. Shub, 'A simple unpredictable pseudorandom number generator', *SIAM J. Comp.*, **15** (1986), 364–383. MR **87k:**65007

[2] J. J. Brennan and B. Geist, 'Analysis of iterated modular exponentiation: The orbit of x^α mod N', *Designs, Codes and Cryptography*, **13** (1998), 229–245. MR **99b:**11086

[3] R. Canetti, J. B. Friedlander, S. Konyagin, M. Larsen, D. Lieman and I. E. Shparlinski, 'On the statistical properties of Diffie–Hellman distributions', *Israel J. Math.* (to appear).

[4] R. Canetti, J. B. Friedlander and I. E. Shparlinski, 'On certain exponential sums and the distribution of Diffie–Hellman triples', *J. London Math. Soc.*, **59** (1999), 799–812. MR **2000g:**11079

[5] J. H. H. Chalk, 'The Vinogradov–Mordell–Tietäväinen inequalities', *Indag. Math.*, **42** (1980), 367–374. MR **82d:**10053

[6] T. W. Cusick, 'Properties of the x^2 mod N pseudorandom number generator', *IEEE Trans. Inform. Theory*, **41** (1995), 1155–1159. MR **96k:**65006

[7] T. W. Cusick, C. Ding and A. Renvall, *Stream Ciphers and Number Theory*, Elsevier, Amsterdam, 1998. MR **99h:**94045

[8] R. Fischlin and C. P. Schnorr, 'Stronger security proofs for RSA and Rabin bits', *Lect. Notes in Comp. Sci.*, Springer-Verlag, Berlin, **1233** (1997), 267–279. CMP 98:09

[9] J. B. Friedlander, J. Hansen and I. E. Shparlinski, 'Character sums with exponential functions', *Mathematika* (to appear).

[10] J. B. Friedlander, D. Lieman and I. E. Shparlinski, 'On the distribution of the RSA generator', *Proc. Intern. Conf. on Sequences and Their Applications (SETA'98), Singapore*, Springer-Verlag, London, 1999, 205-212.

[11] J. B. Friedlander, C. Pomerance and I. E. Shparlinski, 'Period of the power generator and small values of Carmichael's function', *Math. Comp.* (to appear).

[12] F. Griffin and I. E. Shparlinski, 'On the linear complexity profile of the power generator', *Preprint*, 1998, 1–11.

[13] F. Griffin, H. Niederreiter and I. E. Shparlinski, 'On the distribution of nonlinear recursive congruential pseudorandom numbers of higher orders, *Proc. the 13th Symp. on Appl. Algebra, Algebraic Algorithms, and Error-Correcting Codes*, Hawaii, 1999, Lect. Notes in Comp. Sci., Springer-Verlag, Berlin, 1999, v.1719, 87–93.

[14] J. Gutierrez, H. Niederreiter and I. E. Shparlinski, 'On the multidimensional distribution of inversive congruential pseudorandom numbers in parts of the period', *Monatsh. Math.* **129** (2000) 31–36. CMP 2000:08

[15] J. Håstad and M. Näslund, 'The security of individual RSA bits', *Proc 39th IEEE Symp. on Foundations of Comp. Sci.*, 1998, 510–519.

[16] N. M. Korobov, 'On the distribution of digits in periodic fractions', *Math. USSR – Sbornik*, **18** (1972), 659–676. MR **54:**12619

[17] N. M. Korobov, *Exponential sums and their applications*, Kluwer Acad. Publ., Dordrecht, 1992. MR **93a:**11068

[18] J. C. Lagarias, 'Pseudorandom number generators in cryptography and number theory', *Proc. Symp. in Appl. Math.*, Amer. Math. Soc., Providence, RI, **42** (1990), 115–143. MR **92f:**11109

[19] R. Lidl and H. Niederreiter, *Finite fields*, Cambridge University Press, Cambridge, 1997. MR **97i:**11115

[20] A. J. Menezes, P. C. van Oorschot and S. A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, FL, 1997. MR **99g:**94015

[21] H. Niederreiter, 'Quasi-Monte Carlo methods and pseudo-random numbers', *Bull. Amer. Math. Soc.*, **84** (1978), 957–1041. MR **80d:**65016

[22] H. Niederreiter, *Random number generation and Quasi–Monte Carlo methods*, SIAM Press, Philadelphia, 1992. MR **93h:**65008

[23] H. Niederreiter and I. E. Shparlinski, 'On the distribution of inversive congruential pseudorandom numbers in parts of the period', *Math. Comp.* (to appear).

[24] H. Niederreiter and I. E. Shparlinski, 'On the distribution and lattice structure of nonlinear congruential pseudorandom numbers', *Finite Fields and Their Appl.*, **5** (1999), 246–253. CMP 99:17

[25] H. Niederreiter and I. E. Shparlinski, 'On the distribution of inversive congruential pseudorandom numbers modulo a prime power', *Acta Arithm.* (to appear).

[26] H. Niederreiter and I. E. Shparlinski, 'On the distribution of pseudorandom numbers and vectors generated by inversive methods', *Appl. Algebra in Engin., Commun. and Computing*, **10** (2000) 189–202. CMP 2000:11

[27] R. A. Rueppel, 'Stream ciphers', *Contemporary cryptology: The science of information integrity*, IEEE Press, NY, 1992, 65–134. CMP 93:08

[28] I. E. Shparlinski, 'On the linear complexity of the power generator', *Designs, Codes and Cryptography* (to appear).

[29] D. R. Stinson, *Cryptography: Theory and Practice*, CRC Press, Boca Raton, FL, 1995. MR **96k:**94015

[30] I. M. Vinogradov, *Elements of Number Theory*, Dover Publ., NY, 1954. MR **15:**933e

Department of Mathematics, University of Toronto, Toronto, Ontario M5S 3G3, Canada

E-mail address: frdlndr@math.toronto.edu

Department of Computing, Macquarie University, Sydney, New South Wales 2109, Australia

E-mail address: igor@ics.mq.edu.au