Finite Fields and Their Applications **7,** 318–326 (2001)
doi.10.1006/ffta.2000.0291, available online at http://www.idealibrary.com on IDEAL®

On the Uniformity of Distribution of the Naor–Reingold Pseudo-Random Function

Igor E. Shparlinski[1]

Department of Computing, Macquarie University, New South Wales 2109, Australia
E-mail: igor@comp.mq.edu.au

Communicated by Harald Niederreiter

Received May 4, 1999; revised March 7, 2000; published online March 29, 2001

We show that the new pseudo-random number function, introduced recently by M. Naor and O. Reingold, possesses one more attractive and useful property. Namely, it is proved that for almost all values of parameters it produces a uniformly distributed sequence. The proof is based on some recent bounds of character sums with exponential functions. © 2001 Academic Press

Key Words: pseudo-random numbers; discrepancy; exponential functions; character sums.

1. INTRODUCTION

Let p be an n-bit prime, $2^{n-1} \le p \le 2^n - 1$, and let l be a prime divisor of $p-1$.

Denote by $\mathbb{F}_p$ the finite field of p elements and select an element $g \in \mathbb{F}_p^*$ of multiplicative order l. We recall that $\vartheta \in \mathbb{F}_p^*$ is of multiplicative order t if and only if

$$\vartheta^i \neq 1,\ 1 \le i \le t-1, \qquad \vartheta^t = 1.$$

Then for each n-dimensional vector $\mathbf{a} = (a_1, \ldots, a_n) \in (\mathbb{Z}/l)^n$ one can define the function

$$f_{\mathbf{a}}(x) = g^{a_1^{x_1} \cdots a_n^{x_n}} \in \mathbb{F}_p,$$

[1] Supported in part by ARC Grant A69700294.

1071-5797/01 $35.00

where $x = x_1 \dots x_n$ is the bit representation of an integer x, $0 \le x \le 2^n - 1$, with some extra leading zeros if necessary.

In [11] Naor and Reingold proposed the function $f_{\mathbf{a}}(x)$ as an efficient pseudo-random function (for a randomly chosen vector $\mathbf{a} \in (\mathbb{Z}/l)^n$). It is shown in [11] that this function can be computed in parallel by threshold circuits of bounded depth and polynomial size and also that it has some very desirable security property, provided certain standard cryptographic assumptions hold.

Here we show that this function has one more useful feature, which comes as an additional bonus to the aforementioned cryptographic properties of $f_{\mathbf{a}}(x)$. Namely, we prove that for almost all vectors $\mathbf{a} \in (\mathbb{Z}/l)^n$, the sequence $f_{\mathbf{a}}(x)$, $x = 0, 1, \dots, 2^n - 1$, is asymptotically uniformly distributed.

We note that although this property does not seem to have any immediate cryptographic implications, the inverse fact, that is, nonuniformity of distribution, if true, would have disastrous consequences for applications of this function. Besides this, studying the uniformity of distribution of interesting functions is a very attractive number theoretic question. Our main tool is the bound of character sums with exponential functions which is due to Konyagin and the author [7], which in turn is based on the estimate of Gaussian sums of Heath-Brown and Konyagin [6]. Previously known bounds of character sums with exponential functions, which are due to Korobov [8, 9] and Niederreiter [12, 13], can also be used; however, they imply weaker results.

We also remark that an exponential lower bound on the *linear complexity* of this generator has been obtained in [4, 16]. In [2] this bound has been extended to *nonlinear complexity*.

Finally, for the *elliptic curve* version of this generator similar results have been obtained in [15, 17].

2. PREPARATIONS

We identify $\mathbb{F}_p$ with the set $\{0, \dots, p-1\}$.

For a set $\mathcal{M} \subseteq \mathbb{F}_p$ we define the *discrepancy* $D(\mathcal{M})$ modulo p as

$$D(\mathcal{M}) = \sup_{\mathcal{I} \subseteq [0,1]} \left| \frac{N(\mathcal{I})}{\#\mathcal{M}} - |\mathcal{I}| \right|,$$

where $N(\mathcal{I})$ is the number of fractional parts $\{m/p\}$ with $m \in \mathcal{M}$ which hit the interval $\mathcal{I} = [\alpha, \beta] \subseteq [0, 1]$ of length $|\mathcal{I}| = \beta - \alpha$.

We denote by $D_{l,p,g}(\mathbf{a})$ the discrepancy of the set $\{f_{\mathbf{a}}(x) \mid x = 0, 1, \dots, 2^n - 1\}$. We show that $D_{l,p,g}(\mathbf{a}) = o(1)$ for all except possibly $o(l^n)$ vectors $\mathbf{a} \in (\mathbb{Z}/l)^n$, provided that $l \ge p^{1/3+\varepsilon}$ with any fixed $\varepsilon > 0$.

Throughout the paper the implied constants in symbols O and $\ll$ are absolute (we recall that $A \ll B$ is equivalent to $A = O(B)$).

We also denote by $\log u$ the binary logarithm of a real u and put

$$\mathbf{e}_p(a) = \exp(2\pi i a/p), \qquad a \in \mathbb{F}_p.$$

Thus $\mathbf{e}_p(a)$ is a nontrivial additive character of $\mathbb{F}_p$.

We need a form of the *Erdös–Turán inequality* which relates the discrepancy and character sums; see Corollary 1.1 to Chapter 1 of [10] or Corollary 3.11 of [13].

Lemma 1. *For any set $\mathcal{M} \subseteq \mathbb{F}_p$ the bound*

$$D(\mathcal{M}) \ll \frac{1}{p} + \frac{1}{\#\mathcal{M}} \sum_{h=1}^{p-1} \frac{1}{h} \left| \sum_{m \in \mathcal{M}} \mathbf{e}_p(hm) \right|$$

holds.

We also need the following upper bound on character sums with exponential functions which is essentially Theorem 3.4 of [7].

Lemma 2. *Let p be prime and let $\vartheta \in \mathbb{F}_p^*$ be of multiplicative order t modulo p. Then the bound*

$$\max_{\gcd(h,p)=1} \left| \sum_{r=0}^{t-1} \mathbf{e}_p(h\vartheta^r) \right| \ll B(t,p),$$

where

$$B(t,p) = \begin{cases} p^{1/2}, & \text{if } t \ge p^{2/3}, \\ p^{1/4}t^{3/8}, & \text{if } p^{2/3} > t \ge p^{1/2}, \\ p^{1/8}t^{5/8}, & \text{if } p^{1/2} > t \ge p^{1/3}, \end{cases}$$

holds.

This bound is nontrivial for $t \ge p^{1/3+\varepsilon}$ with any fixed $\varepsilon > 0$. Our next result shows that for almost all primes p even much shorter sums admit a nontrivial estimate. It readily follows from Theorem 5.5 of [7]; see also remarks after the proof of this result in [7].

Lemma 3. *Let $\varepsilon > 0$ be sufficiently small and let an integer t and a real $A > 2$ be such that there are at least $U \ge t^{A-1-\varepsilon}$ primes $p \equiv 1 \pmod t$ with $t^A \le p \le 2t^A$. Then for all except possibly $o(U)$ such primes p and any fixed* *integer $k \geq 2$ the bound*

$$\max_{\gcd(h,p)=1} \left| \sum_{r=0}^{t-1} \mathbf{e}_p(h\vartheta_p^r) \right| \ll t(t^{-(2k-A)/2k^2} + t^{-(A-2)/2k^2+\varepsilon})$$

holds for all elements ϑ_p of multiplicative order t modulo p.

3. MAIN RESULTS

Now we are prepared to prove our main results.

THEOREM 4. *The bound*

$$\frac{1}{l^n} \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} D_{l,p,g}(\mathbf{a})^2 \ll \Delta(l,p)$$

holds, where

$$\Delta(l,p) = \begin{cases} p^{1-\gamma} l^{-1} \log^2 p, & \text{if } l \geq p^{\gamma}, \\ p l^{-2} \log^2 p, & \text{if } p^{\gamma} > l \geq p^{2/3}, \\ p^{1/2} l^{-5/4} \log^2 p, & \text{if } p^{2/3} > l \geq p^{1/2}, \\ p^{1/4} l^{-3/4} \log^2 p, & \text{if } p^{1/2} > l \geq p^{1/3}, \end{cases}$$

and $\gamma = 2.5 - \log 3 = 0.9150\ldots$.

Proof. We may assume that p is large enough, in particular that $n \geq 3$. From Lemma 1 and the Cauchy inequality we conclude that

$$\begin{aligned} \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} D_{l,p,g}(\mathbf{a})^2 &\ll \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} \left(\frac{1}{p} + \frac{1}{2^n} \sum_{h=1}^{p-1} \frac{1}{h} \left| \sum_{x=0}^{2^n-1} \mathbf{e}_p(hf_{\mathbf{a}}(x)) \right| \right)^2 \\ &\ll \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} \left(\frac{1}{p^2} + \frac{1}{2^{2n}} \left(\sum_{h=1}^{p-1} \frac{1}{h} \left| \sum_{x=0}^{2^n-1} \mathbf{e}_p(hf_{\mathbf{a}}(x)) \right| \right)^2 \right) \\ &\ll \frac{l^n}{p^2} + \frac{1}{2^{2n}} \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} \sum_{j=1}^{p-1} \frac{1}{j} \sum_{h=1}^{p-1} \frac{1}{h} \left| \sum_{x=0}^{2^n-1} \mathbf{e}_p(hf_{\mathbf{a}}(x)) \right|^2. \end{aligned}$$

Therefore,

$$\sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} D_{l,p,g}(\mathbf{a})^2 \ll \frac{l^n}{p^2} + \frac{\log p}{2^{2n}} \sum_{h=1}^{p-1} \frac{1}{h} W_h, \tag{1}$$

where

$$W_h = \sum_{\mathbf{a}\in(\mathbb{Z}/l)^n} \left| \sum_{x=0}^{2^n-1} \mathbf{e}_p(hf_{\mathbf{a}}(x)) \right|^2 .$$

We recall that $|z|^2 = z\bar{z}$ for any complex z and that $\overline{\mathbf{e}_p(a)} = \mathbf{e}_p(-a)$ for any real a. Then, it is easy to see that replacing the square of the inner sum by a double sum and changing the order of summation we obtain

$$W_h = \sum_{x,y=0}^{2^n-1} \sum_{\mathbf{a}\in(\mathbb{Z}/l)^n} \mathbf{e}_p(h(f_{\mathbf{a}}(x) - f_{\mathbf{a}}(y))).$$

If $x = y$ the inner sum is equal to l^n.

Now we consider the case $x \neq y$. We say that $x \succ y$ if $x_i \geq y_i$, $i = 1, \dots, n$, where $x = x_1 \dots x_n$ and $y = y_1 \dots y_n$ are the bit representation of x and y.

We also say that integers x and y are *comparable* if either $x \succ y$ or $y \succ x$.

If $x \neq y$ and $x \succ y$ we fix i, $1 \leq i \leq n$, with $x_i = 1$, $y_i = 0$.

We see that the term $f_{\mathbf{a}}(y)$ does not depend on a_i.

Let the vector $(z_1, \dots, z_{n-1})$ be formed by all the bits of x except x_i, that is, $z_k = x_k$ if $1 \leq k < i$ and $z_k = x_{k+1}$ if $i \leq k \leq n-1$. Therefore,

$$\left| \sum_{\mathbf{a}\in(\mathbb{Z}/l)^n} \mathbf{e}_p(h(f_{\mathbf{a}}(x) - f_{\mathbf{a}}(y))) \right| \leq \sum_{\mathbf{b}\in(\mathbb{Z}/l)^{n-1}} \left| \sum_{r=0}^{l-1} \mathbf{e}_p(h\vartheta_{\mathbf{b},x}^r) \right|,$$

where $\mathbf{b} = (b_1, \dots, b_{n-1})$ and

$$\vartheta_{\mathbf{b},x} = g^{b_1^{z_1} \dots b_{n-1}^{z_{n-1}}}.$$

We see that if

$$b_1 \dots b_{n-1} \not\equiv 0 \qquad (\mathrm{mol}\ l)$$

then, because l is prime, $\vartheta_{\mathbf{b},x}$ is of multiplicative order l. Hence the bound of Lemma 2 applies to the inner sum. For other $O(nl^{n-2})$ vectors $\mathbf{b}$ we estimate the inner sum trivially by l.

It is easy to see that there are

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = 3^n$$

pairs of (x, y), $0 \le x, y \le 2^n - 1$, with $x \succ y$. Thus this part of the sum can be estimated as

$$\left| \sum_{\substack{x,y=0 \\ x \ne y,\ x \succ y}}^{2^n - 1} \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} \mathbf{e}_p(h(f_{\mathbf{a}}(x) - f_{\mathbf{a}}(y))) \right| \ll 3^n(nl^{n-1} + l^{n-1}B(l, p)).$$

The case $x \ne y$ and $y \succ x$ can be considered quite analogously.

Finally, let us consider pairs of x and y which are not comparable. In this case there are i and j, $1 \le i, j \le n$, with $x_i = y_j = 1$ and $x_j = y_i = 0$. We see that the term $f_{\mathbf{a}}(y)$ does not depend on a_i and the term $f_{\mathbf{a}}(x)$ does not depend on a_j.

Let the vector $(z_1, \dots, z_{n-2})$ be formed by all the bits of x except x_i and x_j; that is, $z_k = x_k$ if $1 \le k < I$, $z_k = x_{k+1}$ if $I \le k < J - 1$, and $z_k = x_{k+2}$ if $J - 1 \le k \le n - 2$, where $I = \min\{i, j\}$ and $J = \max\{i, j\}$. We also form the vector $(w_1, \dots, w_{n-2})$ in a similar way from all the bits of y except y_i and y_j.

Therefore,

$$\left| \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} \mathbf{e}_p(h(f_{\mathbf{a}}(x) - f_{\mathbf{a}}(y))) \right| \le \sum_{\mathbf{b} \in (\mathbb{Z}/l)^{n-2}} \left| \sum_{r=0}^{l-1} \mathbf{e}_p(h\lambda_{\mathbf{b},x}^r) \right| \left| \sum_{s=0}^{l-1} \mathbf{e}_p(h\mu_{\mathbf{b},y}^s) \right|,$$

where $\mathbf{b} = (b_1, \dots, b_{n-2})$,

$$\lambda_{\mathbf{b},x} = g^{b_1^{z1} \dots b_{n-2}^{zn-2}} \qquad \text{and} \qquad \mu_{\mathbf{b},y} = g^{b_1^{w1} \dots b_{n-2}^{wn-2}}.$$

We see that if

$$b_1 \dots b_{n-2} \not\equiv 0 \qquad (\text{mol } l)$$

then, because l is prime, $\lambda_{\mathbf{b},x}$ and $\mu_{\mathbf{b},y}$ are both of multiplicative order l. Hence the bound of Lemma 2 applies to both inner sums. For other $O(nl^{n-3})$ vectors $\mathbf{b}$ we estimate the inner sums trivially by l each.

Therefore, for each pair of x and y which are not comparable the bound

$$\left| \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} \mathbf{e}_p(h(f_{\mathbf{a}}(x) - f_{\mathbf{a}}(y))) \right| \ll nl^{n-1} + l^{n-2}B(l, p)^2$$

holds.

Putting everything together and taking into account that $2^n = O(p)$ and $3^n = O(p^\alpha)$, where $\alpha = \log 3$, we derive

$$W_h \ll 2^n l^n + 3^n(nl^{n-1} + l^{n-1}B(l, p)) + 2^{2n}(nl^{n-1} + l^{n-2}B(l, p)^2)$$

$$\ll pl^n + np^\alpha l^{n-1} + p^\alpha l^{n-1}B(l, p) + np^2 l^{n-1} + p^2 l^{n-2}B(l, p)^2.$$

It is easy to see that the terms including $B(l,p)$ dominate all other terms. Thus

$$W_h \ll p^{\alpha} l^{n-1} B(l,p) + p^2 l^{n-2} B(l,p)^2. \tag{2}$$

Combining (1) and (2), we derive

$$\frac{1}{l^n} \sum_{\mathbf{a} \in (\mathbb{Z}/l)^n} D_{l,p,g}(\mathbf{a})^2 \ll \frac{1}{p^2} + \frac{\log p}{2^{2n}} \sum_{h=1}^{p-1} \frac{p^{\alpha} l^{-1} B(l,p) + p^2 l^{-2} B(l,p)^2}{h}$$

$$\ll (p^{\alpha-2} l^{-1} B(l,p) + l^{-2} B(l,p)^2) \log^2 p.$$

Remarking that the first term in the numerator dominates if and only if $l \geq p^{\gamma}$, we obtain the desired result. ■

In particular, if the vector $\mathbf{a} \in (\mathbb{Z}/l)^n$ is chosen uniformly at random then for any $\delta > 0$ with probability at least $1 - \delta$

$$D_{l,p,g}(\mathbf{a}) \ll \delta^{-1} \Delta(l,p)^{1/2}.$$

Lemma 3 can be used in a similar way to produce the following result which essentially tells that even elements of very small period are likely to produce uniformly distributed sequences.

THEOREM 5. *Let $\varepsilon > 0$ be sufficiently small and let a prime l and a real $A > 2$ be such that there are at least $U \geq l^{A-1-\varepsilon}$ primes $p \equiv 1 \, (\mathrm{mol}\, l)$ with $l^A \leq p \leq 2l^A$. For each such prime p let us fix an element $g_p \in \mathbb{F}_p$ of order l. Then,for any fixed integer $k \geq 2$,for all, except possibly $o(U)$, such primes p, if the vector $\mathbf{a} \in (\mathbb{Z}/l)^n$ is chosen uniformly at random then for any $\delta > 0$ with probability at least $1 - \delta$*

$$D_{l,p,g_p}(\mathbf{a}) \ll \delta^{-1}(l^{-(2k-A)/2k^2+\varepsilon} + l^{-(A-2)/2k^2+\varepsilon}) \log p.$$

4. REMARKS

It is easy to see that the bound of Theorem 4 is nontrivial beginning with $l \geq p^{1/3+\varepsilon}$ with any fixed $\varepsilon > 0$. It is also useful to recall that there exist infinitely many primes p such that $p - 1$ has a prime divisor $l > p^{0.677}$; see [1]. For such p and l we see that $D_{l,p,g}(\mathbf{a}) \leq l^{-0.26}$ for almost all $\mathbf{a} \in (\mathbb{Z}/l)^n$. Moreover, it is expected that $l = (p-1)/2$ is prime for infinitely many primes p. Such pairs of p and l are of special value for cryptography. For them we deduce that $D_{l,p,g}(\mathbf{a}) \leq l^{-0.41}$ for almost all $\mathbf{a} \in (\mathbb{Z}/l)^n$.

It is well known that assuming the extended Riemann hypothesis one can select any $A > 2$ in Lemma 3 and Theorem 5; see Chapter 20 of [3] or Section 5 of Chapter 7 of [14]. In fact only such values of A are of our interest because otherwise Lemma 2 provides stronger results. The author is unaware of any unconditional results of such kind but they can probably be obtained as well (possibly with much larger value of A); see [5]. It is also useful to remark that if $A \geq 3$ is an integer then the optimal choice for k is $k = A - 1$ which produces the estimate

$$D_{l,p,g_p}(\mathbf{a}) \ll l^{-(A-2)/2(A-1)^2+\varepsilon}$$

for almost all p satisfying the conditions of Theorem 5 and almost all vectors $\mathbf{a} \in (\mathbb{Z}/l)^n$.

Analogues of Theorems 4 and 5 can also be obtained for other pseudo-random functions from [11]. The same method can also be used to study the distribution of $f_{\mathbf{a}}(x)$ for $x = 0, 1, \ldots, N-1$ with $N \leq 2^n$.

Finally, it would also be interesting to study the distribution of k-tuples $(f_{\mathbf{a}}(x), \ldots, f_{\mathbf{a}}(x+k-1))$.

REFERENCES

1. R. C. Baker and G. Harman, Shifted primes without large prime factors, *Acta Arith.* **83** (1998), 331–361.
2. W. Banks, F. Griffin, D. Lieman, and I. E. Shparlinski, Non-linear complexity of the Naor–Reingold pseudo-random function, *in* "Proc. the 2nd Intern. Conf. on Information Security and Cryptology, Seoul, 1999," *Lecture Notes in Computer Science*, Vol. 1787, pp. 53–59, Springer-Verlag, Berlin, 2000.
3. H. Davenport, "Multiplicative Number Theory," Markham, Chicago, 1967.
4. F. Griffin and I. E. Shparlinski, On the linear complexity of the Naor–Reingold pseudo-random function, *in* "Proc. 2nd Intern. Conf. on Information and Communication Security, Sydney, 1999," *Lecture Notes in Computer Science*, Vol. 1726, pp. 301–308. Springer-Verlag, Berlin, 1999.
5. D. R. Heath-Brown, Zero-free regions for Dirichlet L-functions and the least prime in an arithmetic progression, *Proc. London Math. Soc.* **64** (1991), 265–338.
6. D. R. Heath-Brown and S. Konyagin, New bounds for Gauss sums derived from kth powers, and for Heilbronn's exponential sum, *Quart. J. Math. Oxford*, to appear.
7. S. V. Konyagin and I. E. Shparlinski, "Character Sums with Exponential Functions and Their Applications," Cambridge Univ. Press, Cambridge, UK, 1999.
8. N. M. Korobov, On the distribution of digits in periodic fractions, *Mat. Sb.* **18** (1972), 659–676.
9. N. M. Korobov, "Exponential Sums and Their Applications," Kluwer Academic, Dordrecht, 1992.
10. H. L. Montgomery, "Ten Lectures on the Interface between Analytic Number Theory and Harmonic Analysis," CBMS Regional Conference Series in Math., Vol. 84, *Amer. Math. Soc., Providence*, RI, 1994.

11. M. Naor and O. Reingold, Number-theoretic constructions of efficient pseudo-random functions, *in* "Proc. 38th IEEE Symp. on Foundations of Comp. Sci., 1997," pp. 458–467.

12. H. Niederreiter, Quasi-Monte Carlo methods and pseudo-random numbers, *Bull. Amer. Math. Soc.* **84** (1978), 957–1041.

13. H. Niederreiter, "Random Number Generation and Quasi–Monte Carlo Methods," SIAM, Philadelphia, 1992.

14. K. Prachar, "Primzahlverteilung," Springer-Verlag, Berlin, 1957.

15. I. E. Shparlinski, "On the Naor–Reingold Pseudo-random Number Function from Elliptic Curves," *Appl. Algebra Engng. Comm. Comput.* **11** (2000), 27–34.

16. I. E. Shparlinski, "Linear complexity of the Naor-Reingold pseudo-random function," *Inform. Process Lett.* **95** (2000), 95–99.

17. I. E. Shparlinski and J. H. Silverman, "On the linear complexity of the Naor-Reingold pseudo-random function from elliptic curves," Des. Codes Cryptogr., to appear.