Local Randomness in Candidate One–Way Functions

H. Niederreiter[1] and C.P. Schnorr[2]

[1] Österreichische Akademie der Wissenschaften, Institut für Informationsverarbeitung, Sonnenfelsgasse 19, A–1010 Wien, Austria
e-mail: nied@qiinfo.oeaw.ac.at
[2] Fachbereich Mathematik / Informatik, Universität Frankfurt, Postfach 111932, D–6000 Frankfurt/M., Germany
e-mail: schnorr@informatik.uni-frankfurt.de

Abstract. We call a distribution on n–bit strings (ε, e)–locally random, if for every choice of $e \leq n$ positions the induced distribution on e–bit strings is in the L_1–norm at most ε away from the uniform distribution on e–bit strings. We establish local randomness in polynomial random number generators (RNG) that are candidate one–way functions. Let N be a squarefree integer and let $f_1, \dots, f_\ell$ be polynomials with coefficients in $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$. We study the RNG that stretches a random $x \in \mathbb{Z}_N$ into the sequence of least significant bits of $f_1(x), \dots, f_\ell(x)$. We show that this RNG provides local randomness if for every prime divisor p of N the polynomials $f_1, \dots, f_\ell$ are linearly independent modulo the subspace of polynomials of degree ≤ 1 in $\mathbb{Z}_p[x]$. We also establish local randomness in polynomial random function generators. This yields candidates for cryptographic hash functions. The concept of local randomness in families of functions extends the concept of universal families of hash functions by Carter and Wegman (1979). The proofs of our results rely on upper bounds for exponential sums.

1 Introduction and Summary

A major open problem in cryptography is to establish one–way functions. While we cannot prove one–wayness it makes sense to analyse candidate one–way functions and to prove properties of these functions that are useful in cryptographic applications. We call a distribution on n–bit strings (ε, e)–locally random if for every choice of $e \leq n$ positions the induced distribution on e–bit strings is in the L_1–norm at most ε away from the uniform distribution on e–bit strings. We prove (ε, e)–local randomness for large classes of candidate one–way functions and candidate cryptographic hash functions.

We show that ℓ–tuples of polynomials $(f_1, \dots, f_\ell) \in \mathbb{Z}_N[x]^\ell$ with fixed coefficients in $\mathbb{Z}_N$ and for arbitrary odd squarefree N provide local randomness if for every prime divisor p of N the polynomials $f_1, \dots, f_\ell$ are linearly independent modulo the subspace of polynomials of degree ≤ 1 in $\mathbb{Z}_p[x]$. To give an example let N be prime $N > 2^n$, let $f_1, \dots, f_\ell \in \mathbb{Z}_N[x]$ be any polynomials that are linearly independent modulo the subspace of polynomials of degree ≤ 1 in $\mathbb{Z}_N[x]$. We prove in Corollary 2 that for random $x \in \mathbb{Z}_N$ the bit string

$$(f_1(x)[1, \ldots, f_\ell(x)[1)$$

consisting of the parity bits $f_i(x)[1$ of the residues $f_i(x) \bmod N$ in $[0, N-1]$ is (ε, e)–locally random provided that ε, n, ℓ and e satisfy the inequality

$$(1) \; 2^{-n/2}(2n \log 2)^{e+1} 2\ell \leq \varepsilon \;,$$

where $\log$ denotes the natural logarithm. E.g. we can choose $n \geq 64$, $\ell = \lfloor 2^{n/7} \rfloor$, $\varepsilon = 2^{-n/7}$, $e = \lfloor n/(7 \log n) \rfloor$. Our main result comprises the case that N is an arbitrary odd squarefree integer, that the output contains several bits from each of the residues $f_i(x) \bmod N$, $i = 1, \ldots, \ell$, and that x is chosen to be random in a subinterval $[0, M-1]$ of $[0, N-1]$.

Note that the above function

$$(2) \; [0, N-1] \ni x \;\mapsto\; (f_1(x)[1, \ldots, f_\ell(x)[1)$$

is a candidate one–way function. No inversion algorithm is known that is polynomial time in $\min(\ell, \log_2 N)$. So far the one–wayness of the function (2) has only be proved for random RSA–moduli N and RSA–polynomials $f_i = x^{e^i}$ (see below) provided that the RSA–scheme is secure. It is however possible that this one–way function is more secure than the RSA–scheme. We are not aware of any inversion algorithm which for RSA–moduli N runs in time $\min(2^\ell, N)^{o(1)}$. On the other hand the RSA–scheme can be broken by factoring N using only $\exp(\sqrt{\log N \log\log N})$ many steps. Is there any inversion algorithm that uses knowledge of the factorization of RSA–numbers N ? Is there any inversion algorithm that uses the structure of particular odd moduli N and of particular non–constant polynomials f_i ? Of course the function (2) can easily be inverted for $N = 2$ since $f_i(x)[1$ only depends on $x[1 = x \bmod 2$. Also the problem of inverting is trivial for constant functions as $f_i(x) = x^{N-1} (\bmod N)$ with N prime. Are there more exceptions? Almost nothing is known about the problem to invert (2). However if we cannot even find inverting algorithms for particular cases given the factorization of the modulus then this may be a sign that the function (2) is a truly one–way function.

It is important that the source of randomness in $(f_1(x)[1, \ldots, f_\ell(x)[1)$ is the random argument x while the coefficients of $f_1, \ldots, f_\ell$ are all fixed. Such functions are cryptographically interesting. A well known example is the random number generator (RNG) related to the RSA–scheme by ALEXI, CHOR, GOLDREICH and SCHNORR (1988) and MICALI, SCHNORR (1991). E.g. let N be the product of two large random primes and let the integer $e \geq 3$ be relatively prime to $\varphi(N)$. Then the mapping

$$[0, N-1] \ni x \;\mapsto\; (x^e[1, x^{e^2}[1, \ldots, x^\ell[1)$$

where x^{e^i} is taken modulo N, is a perfect (in the sense of YAO (1982) and BLUM, MICALI (1982)) RNG provided that the RSA–scheme is secure.

The functions $x \mapsto (f_1(x)[1, \ldots, f_\ell(x)[1)$ extend the class of polynomial random number generators (RNG) proposed by MICALI and SCHNORR (1991) which stretch a random seed $x \in [1, N2^{-k}]$ into a polynomial residue $P(x)(\bmod N)$. Micali and Schnorr prove that the m least significant bits of $P(x)(\bmod N)$ are in the L_1–norm at most $O(N^{-1/2}2^{k+m}(\log N)^2 \deg_N(P))$ away from the uniform distribution provided that N is prime and $\deg_N(P) \geq 2$ where $\deg_N(P)$ is the degree of P when P is considered modulo N.

So far local randomness has mainly been studied in functions that are easy to invert, see ALON, BABAI, ITAI (1986), LUBY (1986), SCHNORR (1988), MAURER, MASSEY (1989), NAOR, NAOR (1990), NISAN (1990) and ALON, GOLDREICH, HASTAD, PERALTA (1990). Most of these constructions are methodically simple and are not directed towards cryptographic applications. They aim at minimizing the number of random bits that are used in randomised algorithms. Merely the quadratic character construction by ALON et alii (1990) is similar to our generator, it relies on Weil's theorem. Our proof of local randomness relies on upper bounds for exponential sums and an inequality on quantitative Fourier inversion. We use upper bounds for the discrepancy of polynomial residues from NIEDERREITER (1977) and we extend these bounds from prime moduli to arbitrary squarefree moduli.

We also establish random function generators, associated with fixed polynomials, that provide local randomness. These generators are candidates for cryptographic hash functions. We associate with a polynomial $P \in \mathbb{Z}_N[x]$ of degree d a polynomial function family $P_z(y) = P(z+y)$ where z is the function name and y is the input. For fixed $k, m \leq \log_2 N$ we associate with a random $z \in \mathbb{Z}_N$ a random function

$$P_z^m : [0, 2^k - 1] \longrightarrow \{0,1\}^m \ , \quad y \longmapsto P(z+y)\ [m$$

where $P(z+y)[m$ denotes the bit string consisting of the m least significant bits of the residue $P(z+y) \bmod N$ in $[0, N-1]$.

We call a function family $\{P_z\}$ (ε, e)–*locally random* if for random z and for any e distinct points $y_1, \ldots, y_e$ the distribution of the em–bit string $P_z(y_1) \cdots P_z(y_e)$ is in the L_1–norm at most ε away from the uniform distribution on em–bit strings.

We prove in Theorem 6 that the above family of functions $\{P_z^m\}$ is (ε, e)–locally random, if N is prime, $d = \deg P$ satisfies $e + 1 \leq d < N$ and if

$$N^{-1/2}(\log N)^{e+1}2^{em+2}d \leq \varepsilon.$$

A family of functions is an e–universal family of hash functions as introduced by CARTER and WEGMAN (1979) if and only if it is $(0, e)$–locally random. Our hash functions require fewer random bits than those of Carter and Wegman since we only randomize the input of the polynomial whereas Carter and Wegman randomize all its coefficients. The main point however is that our hash–functions are – if $\deg P$ is sufficiently large – candidates for cryptographically secure hashing whereas the Carter–Wegman hash functions are easy to invert. Thus for the first time we establish local randomness in families of cryptographic hash functions.

2 Random Number Generators that Provide Statistical Local Randomness

We present in Theorem 1 our main result and we derive from it RNG's that are locally random. In order to prove Theorem 1 we establish in Theorem 3 an upper bound on the discrepancy for multidimensional polynomial number sequences. This upper bound relies on an upper bound for exponential sums given in Lemma 4 and on an inequality of Niederreiter (1977) on quantitative Fourier inversion.

Notation. Let $p_1, \dots, p_r$ be r distinct primes, $N = p_1 \cdots p_r$ (i.e. N is squarefree) and $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$. Let $\mathbf{F} = (f_1, \dots, f_\ell)$ be an ℓ–tuple of polynomials $f_j \in \mathbb{Z}[x]$, $j = 1, \dots, \ell$. We denote by $d_i(f_j)$ the degree of f_j when f_j is considered $\bmod p_i$ and we put $d_i(\mathbf{F}) = \max_{1 \le j \le \ell} d_i(f_j)$. We define $c_i(\mathbf{F}) = \min(d_i(\mathbf{F}) - 1, \sqrt{p_i})$ for $i = 1, \dots, r$ and $c(\mathbf{F}) = \prod_{i=1}^r (c_i(\mathbf{F}) + 1)$. We call $\mathbf{F}$ N–*admissible* if for every prime divisor p_i of N the polynomials $f_1, \dots, f_\ell$ are linearly independent modulo the subspace of polynomials of degree ≤ 1 in $\mathbb{Z}_{p_i}[x]$. In this case we also call the set of polynomials $f_1, \dots, f_\ell$ N–admissible. Thus $f_1, \dots, f_\ell$ are N–admissible if for $i = 1, \dots, r$ and for all $a_1, \dots, a_\ell \in \mathbb{Z}$ either the polynomial $\sum_{j=1}^{\ell} a_j f_j (\bmod p_i)$ is non–linear or $a_1 = \dots = a_\ell = 0 (\bmod p_i)$.

We let $\log N$ denote the natural logarithm of N. We identify $\mathbb{Z}_N$ with the integer interval $[0, N-1]$. We abbreviate the set $\{0,1\}^n$ as I_n and we identify the integer interval $[0, 2^n - 1]$ with I_n . If $y \in [0, N-1] = \mathbb{Z}_N$ and $n \le \log_2 N$ we let $y[n \in I_n$ denote the bit string consisting of the n least significant bits of y. Let $\mathbb{N}$ denote the set of positive integers.

A collection of m least significant output bits. We associate with $\mathbf{F} = (f_1, \dots, f_\ell) \in (\mathbb{Z}[x])^\ell$, $N \in \mathbb{N}$ and $\mathbf{m} = (m_1, \dots, m_\ell) \in \mathbb{N}^\ell$ the mapping

$$\mathbf{F}^{\mathbf{m}} : [0, N-1] \to I_m \ , \quad x \mapsto \prod_{j=1}^{\ell} (f_j(x)[m_j) \quad \text{with} \quad f_j(x) \in \mathbb{Z}_N ,$$

where $m = \sum_{j=1}^{\ell} m_j$ and $\prod$ is the concatenation of strings. The mapping $\mathbf{F}^{\mathbf{m}}$ outputs a collection of m least significant bits of $\mathbf{F}(x)$, where $\mathbf{F}(x)$ is taken modulo N.

Our main theorem provides explicit estimates for the max–norm difference between the distribution induced by $\mathbf{F}^{\mathbf{m}}(x)$ for random $x \in [0, M-1] \subset [0, N-1]$, N–admissible $\mathbf{F}$ and the uniform distribution on $\{0,1\}^m$.

Theorem 1. *Let N be odd and squarefree, let $\mathbf{F}, \mathbf{m}, m, \mathbf{F}^{\mathbf{m}}$ be as above and let $\mathbf{F}$ be N–admissible. Then for $N > 148$, $1 \le M \le N$ and random $x \in [0, M-1]$ we have that*

$$\max_{z \in \{0,1\}^m} \left|\mathrm{prob}[\mathbf{F}^{\mathbf{m}}(x) = z] - 2^{-m}\right| \le \frac{4}{M}\sqrt{N}(\log N)^{l+1} c(\mathbf{F}) \ .$$

The condition that $\mathbf{F}$ is N–admissible cannot be completely removed from Theorem 1. Theorem 1 does not hold for linear polynomials $f_1, \ldots, f_\ell$ with $\ell \ge 2$. This is because the least significant bits in two linear polynomials are highly correlated. On the other hand our proof shows that Theorem 1 holds for a single polynomial of degree 1 in the case that $N = M$.

For example let $N > 2^{512}$ be prime and let $d = 2^{32}$. Then the polynomials $x^2, \ldots, x^d$ are N–admissible. Consider for random $x \in [0, N-1]$ the bit string $(x^2[1, \ldots, x^d[1) \in I_{d-1}$. For any choice of 24 bit positions $2 \le i_1 < i_2 \cdots < i_{24} \le 2^{32}$ and every $z \in \{0,1\}^{24}$ we have that

$$\left|\mathrm{prob}(x^{i_1}[1 \cdots x^{i_{24}}[1 = z) - 2^{-24}\right| \ < \ 2^{-44} \ .$$

This follows from Theorem 1 with $\ell = 24$, $f_j = x^{i_j}$ for $j = 1, \ldots, 24$, $N = M$ and $c(\mathbf{F}) \le 2^{32}$.

Definition. A random variable y ranging over a finite set S is called *statistically random* within ε (in S) if $\sum_{s \in S} \ |\mathrm{prob}(y = s) - 1/\#S| \le \varepsilon$, i.e. the L_1–norm statistical difference of y from the uniform distribution on S is at most ε.

Definition. A probability distribution D on I_n is called (ε, e)*–locally random* if for any sequence of positions $1 \le j_1 < j_2 < \cdots < j_e \le n$ the substring $(y_{j_1}, \ldots, y_{j_e}) \in I_e$ of a D–random string $y = (y_1, \ldots, y_n)$ is statistically random within ε.

Using Theorem 1 we can stretch a short random seed into a long bit string that is "locally random".

Corollary 2. *Let $N = p_1 \cdots p_r$ be a product of r distinct odd primes, $1 \le M \le N$ and $N > 148$. Let $f_1, \ldots, f_\ell \in \mathbb{Z}[x]$ be polynomials of degree at most d that are N–admissible. Then for random $x \in [0, M-1]$ the bit string $(f_1(x)[1, \ldots, f_\ell(x)[1) \in I_\ell$ with $f_j(x) \in \mathbb{Z}_N$ is (ε, e)–locally random with $\varepsilon = 2\frac{\sqrt{N}}{M}(2\log N)^{e+1} d^r$ for $e = 1, \ldots, \ell$.*

Proof. Let $1 \le j_1 < j_2 < \cdots < j_e \le \ell$ be any sequence of e output bit positions. We apply Theorem 1 with $\mathbf{F} = (f_{j_1}, \ldots, f_{j_e})$, $\mathbf{m} = (1, \ldots, 1) \in \mathbb{N}^e$ and $m = \ell = e$. The L_1–norm difference between the distribution induced by $\mathbf{F}^{\mathbf{m}}(x) \in \{0,1\}^e$ and the uniform distribution is at most 2^e–times the max–norm difference. We have $c(\mathbf{F}) \le d^r$, and thus by Theorem 1 $\mathbf{F}(x)$ is statistically random within $2\frac{\sqrt{N}}{M}(2\log N)^{e+1} d^r$. □

The *discrepancy* $D_M^{(\ell)} = D_M^{(\ell)}(\mathbf{y}_1, \ldots, \mathbf{y}_M)$ of M points $\mathbf{y}_1, \ldots, \mathbf{y}_M \in [0,1)^\ell$ is defined to be

$$D_M^{(\ell)}(\mathbf{y}_1, \ldots, \mathbf{y}_M) = \sup_{\mathcal{I}} |F_M(\mathcal{I}) - V(\mathcal{I})|$$

where $\mathcal{I}$ ranges over all half–open subintervals $\mathcal{I}$ of $[0,1)^\ell$, i.e.

$$\mathcal{I} = \left\{(z_1, \ldots, z_\ell) \in [0,1)^\ell \,\middle|\, a_i \le z_i < b_i \text{ for } i = 1, \ldots, \ell\right\}$$

with $0 \le a_i < b_i \le 1$ for $i = 1, \ldots, \ell$. $V(\mathcal{I})$ is the volume of $\mathcal{I}$ and $F_M(\mathcal{I}) = M^{-1}\#\{k \,|\, \mathbf{y}_k \in \mathcal{I}\}$.

The proof of Theorem 1 relies on the following upper bound for the discrepancy of multidimensional polynomial sequences. For a real number a we let $\{a\}$ denote the residue of $a \bmod \mathbb{Z}$ in the real interval $[0,1)$.

Theorem 3. *Let N be squarefree and let $D_M^{(\ell)}$ be the discrepancy of the M points $\left(\left\{\frac{f_1(k)}{N}\right\}, \ldots, \left\{\frac{f_\ell(k)}{N}\right\}\right) \in [0,1)^\ell$ for $k = 1, \ldots, M$. If $\mathbf{F} = (f_1, \ldots, f_\ell)$ is N–admissible then $D_M^{(\ell)} \le \frac{4B}{M}\sqrt{N}(\log N)^{\ell+1} c(\mathbf{F})$ for $1 \le M \le N$ and $N > 148$. Here $B = \sqrt{2}$ if N is even and $B = 1$ if N is odd.*

The proof is based on a bound for exponential sums. For $f \in \mathbb{Z}[x]$ and $n \in \mathbb{N}$ define

$$S(f, n) = \sum_{x=1}^{n} e\left(\frac{f(x)}{n}\right) \text{ with } e(u) = e^{2\pi\sqrt{-1}u} \text{ for } u \in \mathbb{R}.$$

The proofs for Lemmata 4 and 5 are omitted due to lack of space. They are contained in the complete paper that is to appear in Siam J. Computing.

Lemma 4. *If $N = p_1 \ldots p_r$ is squarefree, B is as in Theorem 3, and $f \in \mathbb{Z}[x]$ is arbitrary, then $|S(f,N)| \le B\sqrt{N}\prod_{i=1}^{r} c_i'(f)$, where $c_i'(f) = \sqrt{p_i}$ if $d_i(f) < 1$ and $c_i'(f) = \min(d_i(f) - 1, \sqrt{p_i})$ if $d_i(f) \ge 1$.*

Lemma 5. *Let $D_M^{(\ell)}$ be the discrepancy of the M points $\mathbf{y}_k \in [0,1)^\ell$ for $k = 1, \ldots, M$ and let $D_N^{(\ell+1)}$ be the discrepancy of the N points $\left(\mathbf{y}_k, \frac{k-1}{N}\right)$ for $k = 1, \ldots, N$. Then $D_M^{(\ell)} \le \frac{N}{M} D_N^{(\ell+1)}$ for $1 \le M \le N$.*

Proof of Theorem 3. Let N have r distinct prime factors. Put $C_\ell(N) = (-N/2, N/2]^\ell \cap \mathbb{Z}^\ell$, $C_\ell^*(N) = C_\ell(N)\backslash\{\mathbf{0}\}$ (here we use, as in NIEDERREITER (1977), the interval $(-N/2, N/2]$ rather than $[0, N))$. For $\mathbf{h} = (h_1, \ldots, h_\ell) \in C_\ell(N)$ we put

$$r(\mathbf{h}, N) = \prod_{j=1}^{\ell} r(h_j, N) \quad \text{with} \quad r(h_j, N) = \begin{cases} 1 & \text{if } h_j = 0 \\ N \sin \frac{\pi |h_j|}{N} & \text{if } h_j \neq 0. \end{cases}$$

By Lemma 2.2 of Niederreiter (1977), we get

$$(3) \quad D_N^{(\ell+1)} \leq \frac{\ell+1}{N} + \frac{1}{N} \sum_{\mathbf{h} \in C^*_{\ell+1}(N)} \frac{1}{r(\mathbf{h}, N)} \left| S(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x, N) \right|,$$

where $\mathbf{h} = (h_1, \ldots, h_{\ell+1})$ and $f_1, \ldots, f_\ell \in \mathbb{Z}[x]$. By Lemma 4

$$(4) \quad |S(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x, N)| \leq B N^{1/2} \prod_{i=1}^{r} c_i'(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x).$$

If $\mathbf{h} \in C^*_{\ell+1}(N)$ with $(h_1, \ldots, h_\ell) = \mathbf{0}$, then $c_i'(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x) = c_i'(h_{\ell+1} x) = 0$ for some i, namely when $h_{\ell+1} \neq 0 \bmod p_i$, and so

$$\prod_{i=1}^{r} c_i'(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x) = 0.$$

Thus we only have to consider those $\mathbf{h} \in C^*_{\ell+1}(N)$ with $(h_1, \ldots, h_\ell) \neq \mathbf{0}$. We split up the set of $(h_1, \ldots, h_\ell) \in C_\ell^*(N)$ according to the set of i's for which $d_i(h_1 f_1 + \ldots + h_\ell f_\ell) \leq 1$. For $I \subseteq A_r := \{1, 2, \ldots, r\}$ we put $H(I) = \{(h_1, \ldots, h_\ell) \in C_\ell^*(N) : d_i(h_1 f_1 + \ldots + h_\ell f_\ell) \leq 1$ if and only if $i \in I\}$. If $(h_1, \ldots, h_\ell) \in H(I)$ and $i \in A_r \backslash I$, then for any $h_{\ell+1} \in C_1(N)$ we have

$$d_i(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x) = d_i(h_1 f_1 + \ldots + h_\ell f_\ell) \geq 2.$$

Since $d_i(h_1 f_1 + \ldots + h_\ell f_\ell) \leq d_i(\mathbf{F})$, it follows that

$$c_i'(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x) \leq c_i(\mathbf{F}).$$

Using the trivial bound $c_i'(f) \leq p_i^{1/2}$, we obtain

$$\prod_{i=1}^{r} c_i'(h_1 f_1 + \ldots + h_\ell f_\ell + h_{\ell+1} x) \leq \prod_{i \in I} p_i^{1/2} \cdot \prod_{i \in A_r \backslash I} c_i(\mathbf{F})$$

for any $(h_1, \ldots, h_\ell) \in H(I)$ and $h_{\ell+1} \in C_1(N)$. Together with (3) and (4) this yields

$$D_N^{(\ell+1)} \leq \frac{\ell+1}{N} + B N^{-1/2} \sum_{I \subseteq A_r} \prod_{i \in I} p_i^{1/2} \cdot \prod_{i \in A_r \backslash I} c_i(\mathbf{F}) \sum_{\mathbf{h} \in H(I)} \frac{1}{r(\mathbf{h}, N)} \sum_{h_{\ell+1} \in C_1(N)} \frac{1}{r(h_{\ell+1}, N)}.$$

Using the inequality

$$(5) \quad \sum_{h \in C_1^*(m)} \frac{1}{r(h,m)} < \frac{2}{\pi} \log m + \frac{2}{5} \quad \text{for} \quad m \geq 2$$

from Niederreiter [14, (2.7)] this yields

$$D_N^{(\ell+1)} < \frac{\ell+1}{N} + BN^{-1/2}\left(\frac{2}{\pi}\log N + \frac{7}{5}\right) \sum_{I \subseteq A_r} \prod_{i \in I} p_i^{1/2} \cdot \prod_{i \in A_r \setminus I} c_i(\mathbf{F}) \sum_{\mathbf{h} \in H(I)} \frac{1}{r(\mathbf{h}, N)}.$$

By the assumption of the theorem $(f_1, \ldots, f_\ell)$ is N–admissible. Therefore if $(h_1, \ldots, h_\ell) \in H(I)$ we get $h_k = 0 \bmod p_i$ for $i \in I$ and $1 \leq k \leq \ell$, thus $h_k = 0 \bmod \prod_{i \in I} p_i$ for $1 \leq k \leq \ell$. Therefore with $L = \prod_{i \in I} p_i$ we obtain

$$\begin{aligned}
\sum_{\mathbf{h} \in H(I)} \frac{1}{r(\mathbf{h}, N)} \leq \sum_{\substack{\mathbf{h} \in C_\ell^*(N) \\ \mathbf{h} = \mathbf{0} \bmod L}} \frac{1}{r(\mathbf{h}, N)} &= \sum_{\mathbf{h} \in C_\ell(N/L)} \frac{1}{r(L\mathbf{h}, N)} - 1 \\
&= \left(\sum_{h \in C_1(N/L)} \frac{1}{r(Lh, N)} \right)^\ell - 1 \\
&= \left(1 + \sum_{h \in C_1^*(N/L)} \frac{1}{r(Lh, N)} \right)^\ell - 1 \\
&= \left(1 + \frac{1}{L} \sum_{h \in C_1^*(N/L)} \frac{1}{r(h, N/L)} \right)^\ell - 1 \\
&< \left(1 + \frac{1}{L} \left(\frac{2}{\pi} \log \frac{N}{L} + \frac{2}{5} \right) \right)^\ell - 1 \qquad \text{(by the inequality (5))} \\
&< \frac{\ell}{L} \left(\frac{2}{\pi} \log N + \frac{7}{5} \right)^\ell,
\end{aligned}$$

where we applied the mean–value theorem in the last step. It follows that

$$\begin{aligned}
D_N^{(\ell+1)} &< \frac{\ell+1}{N} + B\ell N^{-1/2} \left(\frac{2}{\pi} \log N + \frac{7}{5} \right)^{\ell+1} \sum_{I \subseteq A_r} \prod_{i \in I} p_i^{-1/2} \cdot \prod_{i \in A_r \setminus I} c_i(\mathbf{F}) \\
&= \frac{\ell+1}{N} + B\ell N^{-1/2} \left(\frac{2}{\pi} \log N + \frac{7}{5} \right)^{\ell+1} \prod_{i=1}^{r} (c_i(\mathbf{F}) + p_i^{-1/2}) \\
&< BN^{-1/2} (\log N)^{\ell+1} c(\mathbf{F}) \left(\frac{\ell+1}{N^{1/2} (\log N)^{\ell+1}} + \ell \left(\frac{2}{\pi} + \frac{7}{5 \log N} \right)^{\ell+1} \right) \\
&< BN^{-1/2} (\log N)^{\ell+1} c(\mathbf{F}) \left(\frac{1}{60} (\ell+1) 5^{-\ell} + \ell (\frac{2}{\pi} + \frac{7}{25})^{\ell+1} \right)
\end{aligned}$$

$$< 4BN^{-1/2}(\log N)^{\ell+1}c(\mathbf{F})$$

provided that $\log N \geq 5$, i.e. that $N > 148$. Together with Lemma 5 we get the result of Theorem 3. $\square$

Proof of Theorem 1 Let N be an odd squarefree integer and $\overline{f}_j \in \mathbb{Z}[x]$ be polynomials such that $\overline{f}_j(x) = 2^{-m_j} f_j(x) (\mathrm{mod} N)$ for $j = 1, \ldots, \ell$. Application of Theorem 3 to $\overline{\mathbf{F}} = (\overline{f}_1, \ldots, \overline{f}_\ell)$ shows that the discrepancy $\overline{D}_M^{(\ell)}$ of $(\{\bar{f}_1(k)/N\}, \ldots, \{\bar{f}_\ell(k)/N\})$ for $k = 1, \ldots, M$ satisfies

$$\overline{D}_M^{(\ell)} \leq \frac{4}{M}\sqrt{N}(\log N)^{\ell+1}c(\mathbf{F}),$$

where we use that $c(\mathbf{F}) = c(\overline{\mathbf{F}})$. We apply to this inequality the equivalence

$$\{\overline{f}_j(x)/N\} \in [k_j 2^{-m_j}, (k_j+1)2^{-m_j}) \qquad \Longleftrightarrow$$

$$[f_j(x)]_N = -k_j N (\mathrm{mod} 2^{m_j}) \qquad \text{for} \quad j = 1, \ldots, \ell \ ,$$

where $[f_j(x)]_N$ is the residue of $f_j(x) \bmod N$ in $[0, N-1]$, and $0 \leq k_j < 2^{m_j}$. To see the equivalence we note that $\{\overline{f}_j(x)/N\} \in [k_j 2^{-m_j}, (k_j+1)2^{-m_j})$ implies that there is an integer y satisfying

$$k_j N \leq y < (k_j+1)N, \ y = f_j(x) \bmod N, \ y = 0 \bmod 2^{m_j} \ ,$$

and thus $[f_j(x)]_N = -k_j N (\mathrm{mod} 2^{m_j})$. This proves one direction of the equivalence and the converse direction is an immediate consequence.

We see from the above inequality and the equivalence that for every $y \in \{0,1\}^m$

$$\Big|\frac{1}{M}\#\{x \in [1, M] : \mathbf{F}^{\mathbf{m}}(x) = y\} - \frac{1}{2^m}\Big| \leq \frac{4}{M}\sqrt{N}(\log N)^{\ell+1}c(\mathbf{F}).$$

$\square$

The above proof of Theorem 1 extends to the following larger class of functions $\mathbf{F}^{\mathbf{u}}$. Let the polynomials $f_1, \ldots, f_\ell \in \mathbb{Z}_N[x]$ be N–admissible and let $u_1, \ldots, u_\ell$ be integers that are relatively prime to N, $\mathbf{F} = (f_1, \ldots, f_\ell)$ and $\mathbf{u} = (u_1, \ldots, u_\ell)$. Define $\mathbf{F}^{\mathbf{u}}$ as

$$\mathbf{F}^{\mathbf{u}} : [0, N-1] \ni x \ \mapsto \ \big(\ (f_i(x) \bmod N) \bmod u_i \ \big| \ \text{for } i = 1, \ldots, \ell).$$

Corollary 6. *For $N > 148$, $1 \leq M \leq N$ and random $x \in [0, M-1]$ the max–norm difference between the distribution induced by $\mathbf{F}^{\mathbf{u}}(x)$ and the uniform distribution on $[0, u_1 - 1] \times \cdots \times [0, u_\ell - 1]$ is at most $\frac{4}{M}\sqrt{N}(\log N)^{\ell+1}c(\mathbf{F})$.*

Theorem 1 deals with the particular case that the integers u_i are powers of 2. It is necessary that $u_1, \ldots, u_\ell$ are relatively prime to N. The proof of the Corollary uses the polynomials $\bar{f}_j = u_j^{-1} f_j (\mathrm{mod} N)$ and thus requires a division by u_j modulo N.

3 Random Function Generators that Provide Statistical Local Randomness

Let $H_{k,\ell} = I_\ell^{I_k}$ = "the set of functions $f : I_k \to I_\ell$". A *random function generator* F is an efficient algorithm that generates from names $x \in I_n$ a function $f_x = F(x,*) \in H_{k,\ell}$.

We call a probability distribution D on $H_{k,\ell}$ (ε,e)–*locally random* if for random f, $f \in_D H_{k,\ell}$, for any set of e distinct inputs $y_1,\ldots,y_e \in I_k$ the concatenated output $f(y_1)f(y_2)\cdots f(y_e) \in I_{e\ell}$ is statistically random within ε.

The concept of (ε,e)–locally random distribution D on $H_{k,\ell}$ extends the concept of *universal hash functions* of Carter and Wegman (1979). If D is $(0,e)$–locally random then for any distinct inputs $y_1,\ldots,y_e \in I_k$ the bit string $f(y_1)f(y_2)\cdots f(y_e) \in I_e$ is truly random, i.e. D is the probability distribution of an e–universal family of hash functions in the sense of Carter and Wegman.

Carter and Wegman show how to generate an e–universal family of hash functions in $H_{k,k}$ from ke random bits. Let $K = GF(2^k)$ be the field with 2^k elements. If $(a_0,\ldots,a_{e-1}) \in K^e$ is random then the polynomial $P = \sum_{i=0}^{e-1} a_i x^i \in K[x]$ yields an e–universal family of hash functions in $H_{k,k}$.

Let N be a prime and $P \in \mathbb{Z}_N[x]$ be a polynomial with coefficients in the field $\mathbb{Z}_N$. We associate with P and $k,\ell \in \mathbb{N}$, $k,\ell \le \log_2 N$, the function

$$P^\ell : \mathbb{Z}_N \times [0, 2^k - 1] \to I_\ell\,,\ (z,y) \mapsto P(y+z)[\ell.$$

Here we let $P(y+z)[\ell$, for $\ell \le \log_2 N$, denote the bit string consisting of the ℓ least significant bits of the residue of $P(y+z) \bmod N$ that is in $\mathbb{Z}_N$. We let $P_z^\ell : I_k \to I_\ell$ denote the function $P^\ell(z,*)$.

Theorem 7. *Let N be prime, $N > 148$, $P \in \mathbb{Z}_N[x]$, $k,\ell \le \log_2 N$, let $P_z^\ell : I_k \to I_\ell$ be as above and let $e+1 \le \deg P < N$. Then for random $z \in \mathbb{Z}_N$ the family of functions $\{P_z^\ell\}$ is (ε,e)–locally random with $\varepsilon = N^{-1/2}(\log N)^{e+1}2^{e\ell+2}\deg P$.*

Proof. Let $d = \deg P$, let $y_1,\ldots,y_e \in \mathbb{Z}_N$ be pairwise distinct and let $f_i \in \mathbb{Z}_N[x]$ be the polynomial $f_i(x) = P(y_i + x)$ for $i = 1,\ldots,e$. We next show that the polynomials $f_1,\ldots,f_e$ are linearly independent modulo the subspace of polynomials of degree ≤ 1 in $\mathbb{Z}_N[x]$. For suppose that there are $b_1,\ldots,b_e \in \mathbb{Z}_N$ such that

$$\deg\left(\sum_{i=1}^{e} b_i P(y_i + x)\right) \le 1.$$

Then for $j = d - e + 1, \ldots, d$ the j–th derivative of this linear combination vanishes at $x = 0$, hence

$$\sum_{i=1}^{e} b_i P^{(j)}(y_i) = 0 \quad \text{for} \quad j = d - e + 1, \ldots, d.$$

It is sufficient to prove that the coefficient matrix $[P^{(j)}(y_i)]_{\substack{1 \le i \le e \\ d-e+1 \le j \le d}}$ is non-singular since this implies that $b_1 = \cdots = b_e = 0$. Suppose that there exist $h_{d-e+1}, \ldots, h_d \in \mathbb{Z}_N$ such that

$$\sum_{j=d-e+1}^{d} h_j P^{(j)}(y_i) = 0 \quad \text{for} \quad 1 \le i \le e.$$

Put $g(x) = \sum_{j=d-e+1}^{d} h_j P^{(j)}(x)$, then $g(y_i) = 0$ for $1 \le i \le e$. Since $y_1, \ldots, y_e$ are distinct and $\deg(g) \le d - (d - e + 1) = e - 1$ we have $g = 0$, so

$$\sum_{j=d-e+1}^{d} h_j P^{(j)}(x) = 0.$$

Comparing coefficients of x^{e-1} we get $h_{d-e+1} = 0$ (the coefficient of x^{e-1} in $P^{(d-e+1)}$ is nonzero since $d < N$). Continuing in this manner, we obtain $h_{d-e+1} = \cdots = h_d = 0$.
Since $f_1, \ldots, f_e$ are linearly independent modulo $\mathbb{Z}_N + x\mathbb{Z}_N$ we can apply Theorem 1 to $\mathbf{F} = (f_1, \ldots, f_e)$. Since $\prod_{j=1}^{e} f_j(z)[\ell \in I_{e\ell}$ the m in Theorem 1 is $e\ell$. The ℓ in Theorem 1 is e. Hence $\prod_{j=1}^{e} P(y_j + z)[\ell \in I_{e\ell}$ is statistically random within $\varepsilon = N^{-1/2}(\log N)^{e+1} d\, 2^{e\ell+2}$. Therefore $\{P_z^\ell\}$ is (ε, e)–locally random. $\square$

References

[1] Alexi, W., Chor, B., Goldreich, O. and Schnorr, C.P.: RSA and Rabin Functions: certain parts are as hard as the whole. SIAM J. Comput., 17, 2 (1988), pp. 194 – 208.

[2] Alon, N., Babai, L. and Itai, A.: A fast and simple randomised parallel algorithm for the maximal independent set problem. J. of Alg. 7 (1986), pp. 567 – 583.

[3] Alon, N., Goldreich, O., Hastad, J. and Peralta, R.: Simple constructions of almost k–wise independent random variables. Proceedings of the 31st IEEE Symposium on Foundations of Computer Science (1990) pp. 544 – 552.

[4] Blum, L, Blum, M., and Shub, M.: A simple unpredictable pseudo–random number generator. SIAM J. Comput. 15 (1986), pp. 364 – 383.

[5] Blum, M. and Micali, S.: How to generate cryptographically strong sequences of pseudo–random bits. Proceedings of the 23rd IEEE Symposium on Foundations of Computer Science, IEEE, New York (1982); also SIAM J. Comput. 13 (1984), pp. 850–864.

[6] Carlitz, L. and Uchiyama, S.: Bounds for exponential sums. Duke Math. J. 24, (1957), pp. 37 – 41.

[7] Carter, L. and Wegman, M.: Universal hash functions. J. Comp. and Syst. Sci. 18, (1979) pp. 143 – 154.

[8] Goldreich, O., Goldwasser, S. and Micali, S.: How to Construct Random Functions. Proceedings of the 25th IEEE Symposium on Foundations of Computer Science, New York, (1984); also Journal ACM 33, 4 (1986), pp. 792–807.

[9] Lidl, R., and Niederreiter, H.: Finite Fields. Reading: Addison–Wesley 1983.

[10] Luby, M.: A simple parallel algorithm for the maximal independent set problem. SIAM J. Comput., 15 (1986), pp. 1036 – 1053.

[11] Maurer, U. M., and Massey, J.L.: Perfect local randomness in pseudo–random sequences. Proceedings Crypto '89, Lecture Notes in Computer Science, Vol. 435, Springer–Verlag 1990, pp. 100–112.

[12] Micali, S. and Schnorr, C.P.: Efficient, perfect polynomial random number generators. J. of Cryptology 3, (1991), pp. 157 – 172.

[13] Naor, J. and Naor, M.: Small–bias Probability Spaces: Efficient Constructions and Applications. Proceedings of the 22nd ACM Symposium on Theory of Computing (1990), pp. 213–223.

[14] Niederreiter, H.: Pseudo–random numbers and optimal coefficients. Advances in Math. 26, (1977) pp. 99 – 181.

[15] Nisan, N.: Pseudorandom generators for space–bounded computation. Proceedings of the 22nd ACM Symposium on Theory of Computing (1990), pp. 204–208.

[16] Schnorr, C.P.: On the construction of random number generators and random function generators. Proc. EUROCRYPT '88, Lecture Notes in Computer Science, Vol. 330, Springer–Verlag 1988, pp. 225–232.

[17] Weil, A.: On some exponential sums. Proc. Nat. Acad. Sci. USA 34, (1948), pp. 204 – 207.

[18] Yao, A.C.: Theory and applications of trapdoor functions. Proceedings of the 23rd IEEE Symposium on Foundations of Computer Science, IEEE, New York (1982), pp. 80–91.