Large families of pseudorandom sequences of k symbols and their complexity, Part II

R. Ahlswede, C. Mauduit and A. Sárközy*

Dedicated to the memory of Levon Khachatrian

*Research partially supported by Hungarian National Foundation for Scientific Research, Grant No. T043623.

6 Introduction

We continue the investigation of Part I, keep its terminology, and also continue the numbering of sections, equations, theorems etc., and also use the Reference List at the end of Part I. Consequently we start here with section 6. As mentioned in Section 5 we present now criteria for a triple (r,t,p) to be k–admissible.

Then we consider the f–complexity (extended now to k–ary alphabets) $\Gamma_k(\mathcal{F})$ of a family $\mathcal{F}$. It serves again as a performance parameter of key spaces in cryptography. We give a lower bound for the f–complexity for a family of the type constructed in Part I. In the last sections we explain what can be said about the theoretically best families $\mathcal{F}$ with respect to their f–complexity $\Gamma_k(\mathcal{F})$.

We begin with straightforward extensions of the results of [4] for $k=2$ to general k by using the same Covering Lemma of [1].

But then we give an improvement (also of the earlier results) with respect to balancedness with the help of another old Covering Lemma from [1]. Finally this will again be improved by a more recent result on edge–coverings of hypergraphs from [2]. This has become a basic tool in Information Theory, for instance in the Theory of Identification. In the present context it gives families with a very strong balancedness property. A quantum theoretical analogue became a key tool for quantum channels [3]. It invites to investigate our cryptographical concepts in the quantum world.

7 Sufficient criteria for k–admissibility

We have shown in Part I that the assumption on the k–admissibility in Theorem 2 cannot be dropped. Thus in order to be able to use the construction in Theorem 2, we need criteria for a triple (r,t,p) being k–admissible. We will prove three sufficient criteria of this type:

Theorem 3.

(i) If $k,r,t\in\mathbb{N}$, $1\le t\le k$, p is a prime and $r<p$, then the triple (r,t,p) is k–admissible.

(ii) If $k,r,t\in\mathbb{N}$, p is a prime and

$$(4t)^r < p, \tag{7.1}$$

then (r,t,p) is k–admissible.

(iii) If $k\in\mathbb{N}$, $k\ge 2$, the prime factorization of k is $k=q_1^{\alpha_1}\dots q_s^{\alpha_s}$ (where $q_1,\dots,q_s$ are distinct primes and $\alpha_1,\dots,\alpha_s\in\mathbb{N}$), and p is a prime such that each of $q_1,\dots,q_s$ is a primitive root modulo p, then for every pair $r,t\in\mathbb{N}$ with $r,t<p$, the triple (r,t,p) is k–admissible.

Note that in the $k=2$ special case this theorem gives Theorem 2 in [6].

Proof.

(i) Assume that contrary to the assertion, there are $k, r, t \in \mathbb{N}$ and a prime p so that

$$1 \le t \le k, \tag{7.2}$$

$$r < p, \tag{7.3}$$

and the triple (r, t, p) is not k–admissible, i.e., there is an $\mathcal{A} \subset \mathbb{Z}_p$ and a k–set $\mathcal{B}$ whose elements belong to $\mathbb{Z}_p$ such that $|\mathcal{A}| = r$, $|\mathcal{B}| = t$ (multiple elements counted with their multiplicity) and the number of solutions of (3.1) is divisible by k for all $c \in \mathbb{Z}$.

Consider any $c \in \mathcal{A} + \mathcal{B}$ ($\mathcal{A}, \mathcal{B}$ are non–empty, thus $\mathcal{A} + \mathcal{B}$ is also non–empty). Since for this c (3.1) has at least one solution and the number of solutions is always divisible by k, thus (3.1) must have at least k solutions. On the other hand, clearly (3.1) may have at most $|\mathcal{B}| = t$ solutions so that we must have

$$|\mathcal{B}| = t \ge k. \tag{7.4}$$

It follows from (7.2) and (7.4) that

$$|\mathcal{B}| = t = k. \tag{7.5}$$

Since $\mathcal{B}$ is a k–set, thus the multiplicity of each element is $\le k - 1$. Thus it follows from (7.5) that $\mathcal{B}$ must have at least two distinct elements: say, $b_o, b_o + d \in \mathcal{B}$, $d \ne 0$. Every element of $\mathcal{A} + b_o$ must have (at least) k representations in form (3.1) whence, by (7.5), it follows easily that they also have a representation in the form $(a + b_o + d)$ with $a \in \mathcal{A}$ whence $\mathcal{A} + b_o = \mathcal{A} + b_o + rd$ for all $r \in \mathbb{N}$, thus $\mathcal{A} + b_o = \mathcal{A} + b_o + s$ for any $s \in \mathbb{Z}_p$, in particular for any $s \in \mathcal{A} + b_o$. Hence, $\mathcal{A} + b_o$ is an additive subgroup of $\mathbb{Z}_p$ thus $\mathcal{A} = \mathcal{A} + b_o = \mathbb{Z}_p$ which contradicts $|\mathcal{A}| = r$ and (7.3).

(ii) The proof is nearly the same as the proof of Theorem 2, (ii) in [6]. Thus we will omit most of the details here, we will present only those critical steps where a slight modification is needed.

Assume that r, t, p satisfy (7.1), $\mathcal{A} \subset \mathbb{Z}_p$, $\mathcal{B}$ is a k–set whose elements belong to $\mathbb{Z}_p$, $|\mathcal{A}| = r$ and $|\mathcal{B}| = t$ (multiple elements counted with their multiplicity). It suffices to show that then there is a $c \in \mathbb{Z}_p$ for which the number of solutions of (3.1) (the b's counted with multiplicity) is greater than 0 and less than k. To show this, it suffices to prove that there are $m \in \mathbb{N}$, $c' \in \mathbb{Z}_p$ such that $(m, p) = 1$, and the number of solutions of

$$ma + mb = c', \;\; a \in \mathcal{A}, \;\; b \in \mathcal{B} \tag{7.6}$$

is greater than 0 and less than k. Again, the proof of this is based on Lemma 3 in [6]. We start out from this lemma, and we proceed in the same way as in [6]. In particular, we define $m, b_i, b_j, r_1, r_k, a_n, a_v$ in the same way. Then again, the numbers

$$mb_i + r_k = mb_j + ma_v$$

and

$$mb_j + r_1 = mb_j + ma_u$$

do not have any further representations in form (7.6). Since $\mathcal{B}$ is a k–set, thus the multiplicity of both b_i and b_j is less than k. Thus these numbers have more than 0 and less than k representations in form (7.6) (counting the b's with multiplicity) which completes the proof.

(iii) From a practical point of view this seems to be the most important of the three criteria. Namely, this criterion enables us to control even correlations of very high order provided that there are "many" primes p such that each of $q_1, \dots, q_s$ is a primitive root modulo p. Partly because of the importance of this criterion, partly in order to help to understand the notion of k–admissibility and the related difficulties better, we will give a detailed discussion of this case in the next section. This discussion will lead not only to the proof of criterion (iii), but it will also provide negative examples. We will also show that, most probably, there are "many" primes p of the type described in (iii).

8 k–good primes. Negative examples

Definition 5. A number $m \in \mathbb{N}$ is said to be *k–good* if for any pair $r, t \in \mathbb{N}$ with $r < m$, $t < m$, the triple (r, t, m) is k–admissible. If for all $r < m$, $t < m$ the triple (r, t, m) is (k, k)–admissible, then m is said to be (k, k)–good.

Theorem 4. *If $k \in \mathbb{N}$, $k \geq 2$, the prime factorization of k is $k = q_1^{\alpha_1} \dots q_s^{\alpha_s}$ (where $q_1, \dots, q_s$ are distinct primes and $\alpha_1, \dots, \alpha_s \in \mathbb{N}$) and p is an odd prime such that each of $q_1, \dots, q_s$ is a primitive root modulo p, then p is k–good.*

Proof of Theorem 4. We will need the following lemma:

Lemma 4. *If p is an odd prime and q is a prime which is a primitive root modulo p, then the polynomial $x^{p-1} + x^{p-2} + \dots + x + 1$ is irreducible over F_q.*

Proof of Lemma 4. This is a trivial consequence of Theorem 2.47 in [11, p. 62].

We will prove the assertion of Theorem 4 by contradiction: assume that contrary to the statement of the theorem, there is a set $\mathcal{A} \subset \mathbb{Z}_p$ and a k–set $\mathcal{B}$ whose elements belong to $\mathbb{Z}_p$ so that

$$|\mathcal{A}| = r < p, \ |\mathcal{B}| = t < p \tag{8.1}$$

and the sum $\mathcal{A} + \mathcal{B}$ has property P_k.

If $\mathcal{C}$ is a multiset whose elements belong to $\mathbb{Z}_p$, then let $Q_{\mathcal{C}}(x)$ denote the polynomial $\sum\limits_{c \in \mathcal{C}} x^{s(c)}$ where $s(c)$ denotes the least non–negative element of the residue class c modulo p, and the elements c of $\mathcal{C}$ are to be taken with their multiplicity (so that if c occurs with multiplicity M in $\mathcal{C}$, then there is a term $Mx^{s(c)}$ appearing in $Q_{\mathcal{C}}(x)$). If $\mathcal{C}$ is a multiset of elements of $\mathbb{Z}_p$ and $u \in \mathbb{Z}_p$, then clearly we have $(x^p - 1) \mid x^u Q_{\mathcal{C}}(x) - Q_{\mathcal{C}+u}(x)$ (in $\mathbb{Z}[x]$). It follows that $(x^p - 1) \mid \big(Q_{\mathcal{A}}(x) Q_{\mathcal{B}}(x) - Q_{\mathcal{A}+\mathcal{B}}(x)\big)$:

$$Q_{\mathcal{A}}(x) Q_{\mathcal{B}}(x) = Q_{\mathcal{A}+\mathcal{B}}(x) + (x^p - 1) G(x) \text{ with } G(x) \in \mathbb{Z}[x]. \tag{8.2}$$

Write $Q_{\mathcal{B}}(x)=\sum_{j=0}^{p-1} v_j x^j$ so that the v_j's are the multiplicities of the elements $j \in \mathbb{Z}_p$ in $\mathcal{B}$. It follows that $0 \le v_j \le k-1$ for all $0 \le j \le p-1$, and since

$$|\mathcal{B}| = \sum_{j=0}^{p-1} v_j > 0,$$

thus we have

$$(v_0, v_1, \dots, v_{p-1}) \le k-1.$$

It follows that there is an i with $1 \le i \le s$, $q_i^{\alpha_i} \nmid (v_0, v_1, \dots, v_{p-1})$. Write

$$q_i^{\beta} \| (v_0, v_1, \dots, v_{p-1}) \tag{8.3}$$

so that

$$0 \le \beta < \alpha_i. \tag{8.4}$$

Then every coefficient of $Q_{\mathcal{B}}(x)$ is divisible by q_i^{β}. Since $\mathcal{A}+\mathcal{B}$ has property P_k, the coefficients of $Q_{\mathcal{A}+\mathcal{B}}(x)$ are divisible by k and thus also by q_i^{β}. Thus by (8.2), every coefficient of $(x^p-1)G(x)$ must be also divisible by q_i^{β}. Since the polynomial x^p-1 is primitive (a polynomial $\in \mathbb{Z}[x]$ is said to be primitive if the greatest common divisor of its coefficients is 1), and by Gauss' lemma the product of primitive polynomials is also primitive, thus it follows that the coefficients of $G(x)$ are also divisible by q_i^{β}. Thus we may simplify (8.2) so that we divide the coefficients of $Q_{\mathcal{B}}(x)$, $Q_{\mathcal{A}+\mathcal{B}}(x)$ and $G(x)$ by q_i^{β}:

$$Q_{\mathcal{A}}(x)\left(\frac{1}{q_i^{\beta}} Q_{\mathcal{B}}(x)\right) = \left(\frac{1}{q_i^{\beta}} Q_{\mathcal{A}+\mathcal{B}}(x)\right) + (x^p-1)\left(\frac{1}{q_i^{\beta}} G(x)\right). \tag{8.5}$$

Since this equation holds over $\mathbb{Z}$ thus it also holds over $\mathbb{Z}_{q_i}$, i.e., over F_{q_i}, in other words, we may consider (8.5) modulo q_i. The coefficients of $Q_{\mathcal{A}+\mathcal{B}}(x)$ are divisible by $q_i^{\alpha_i}$, thus by (8.4), the polynomial $\frac{1}{q_i^{\beta}} Q_{\mathcal{A}+\mathcal{B}}(x)$ is the zero polynomial. Since $(x^{p-1}+x^{p-2}+\dots+1) \mid (x^p-1)$, thus it follows from (8.5) that

$$(x^{p-1}+x^{p-2}+\dots+1) \mid Q_{\mathcal{A}}(x)\left(\frac{1}{q_i^{\beta}} Q_{\mathcal{B}}(x)\right).$$

By Lemma 4 the polynomial $x^{p-1}+x^{p-2}+\dots+1$ is irreducible over F_{q_i}. Thus it follows that either

$$(x^{p-1}+x^{p-2}+\dots+1) \mid Q_{\mathcal{A}}(x) \tag{8.6}$$

or

$$(x^{p-1}+x^{p-2}+\dots+1) \mid \left(\frac{1}{q_i^{\beta}} Q_{\mathcal{B}}(x)\right); \tag{8.7}$$

note that by (8.3), the polynomial $\frac{1}{q_i^{\beta}} Q_{\mathcal{B}}(x)$ is not the 0 polynomial. Since by the definitions of $Q_{\mathcal{A}}(x)$ and $Q_{\mathcal{B}}(x)$ these polynomials are of degree at most $p-1$, thus it would follow from (8.6) and (8.7) that $Q_{\mathcal{A}}(x)$, resp. $Q_{\mathcal{B}}(x)$ is a (non–zero) constant multiple of $x^{p-1}+x^{p-2}+\cdots+1$, whence $|\mathcal{A}| \geq p$, resp. $|\mathcal{B}| \geq p$. This contradicts (8.1) which completes the proof of Theorem 4.

In Section 4 we mentioned that there are negative examples with sums $\mathcal{A}+\mathcal{B}$ having property P_k, i.e., examples for primes p which are not k–good. Now we will present examples of this type.

First we recall that in the $k=2$ special case in [6] we proved that a prime p is 2–good if and only if 2 is a primitive root modulo p. There we presented several examples for sums $\mathcal{A}+\mathcal{B}$ possessing property P_2 (so that for the corresponding primes p, 2 is not a primitive root modulo p). Some of these examples follow:

Example 1. If $p=7$, $d=\{0,1,3\}$ and $\mathcal{B}=\{0,1,2,4\}$, then $\mathcal{A}+\mathcal{B}$ possesses property P_2 so that the triples $(3,4,7)$ and $(4,3,7)$ are not 2–admissible.

Example 2. If $p=17$, $\mathcal{A}=\{0,3,4,5,8\}$ and $\mathcal{B}=\{0,3,4,5,6,9\}$, then $\mathcal{A}+\mathcal{B}$ has property P_2 so that $(5,6,17)$ and $(6,5,17)$ are not 2–admissible.

Example 3. If $p=31$, $\mathcal{A}=\{0,2,5\}$ and $\mathcal{B}=\{0,2,4,5,6,8,9,13,14,15,16,17,20,21,23,26\}$, then $\mathcal{A}+\mathcal{B}$ has property P_2, thus $(3,16,31)$ and $(16,3,31)$ are not 2–admissible.

One might like to present similar negative examples for other k (and p) values as well. To find examples of this type, one has to consider the reverse of the argument in the proof of Theorem 4. We obtain that for fixed k and p, we have to look for non–trivial factorization of x^p-1 over $\mathbb{Z}_k$ of the form

$$x^p - 1 = Q_1(x)Q_2(x) \tag{8.8}$$

with

$$Q_1(x) = \sum_{a\in\mathcal{A}} x^a$$
$$Q_2(x) = \sum_{d\in\mathcal{D}} t_d x^d.$$

(Here "non–trivial" means that both $Q_1(x)$ and $Q_2(x)$ have at least 2 terms.)

If we find a factorization of this form, then defining $\mathcal{B}$ so that it contains the elements $d\in\mathcal{D}$ each with multiplicity t_d, the sum $\mathcal{A}+\mathcal{B}$ possesses property P_k so that the triple $(|\mathcal{A}|,|\mathcal{B}|,p)$ is not k–admissible. The difficulty is that not only we have to find a non–trivial factorization of form (8.8), but also there is the additional restriction that all the coefficients of $Q_1(x)$ must be 0 or 1. This is the reason for that if k is a prime, then for $k>2$ we can give only a sufficient condition for p being k–good. On the other hand, combining the proof of Theorem 4 and the argument above, we can prove that if k is a prime then a prime p is (k,k)–good if and only if k is a primitive root modulo p. (In [6] we proved this in the special case $k=2$.)

Example 4. If $p=13$, then we have

$$x^{13}-1 = (1+x+x^4+x^6)(2+x+2x^2+x^3+2x^5+x^7)$$

over $\mathbb{Z}_3$. It follows that, writing $\mathcal{A}=\{0,1,4,6\}$, $\mathcal{B}=\{0,0,1,2,2,3,5,5,7\}$, the sum $\mathcal{A}+\mathcal{B}$ possesses property P_3, so that $(4,9,13)$ is not 3–admissible, and thus $p=13$ is not 3–good.

If we have a negative example for a certain $k \in \mathbb{N}$ and prime p, and $k \mid k'$, then one can use this example to construct negative examples for k' and p. E.g., starting out from Example 3, we obtain the following negative example for $k = 6$ and $p = 31$:

Example 5. If $p = 31$, $\mathcal{A} = \{0, 2, 4, 5, 6, 8, 9, 13, 14, 15, 16, 17, 20, 21, 23, 26\}$ and $\mathcal{B} = \{0, 0, 0, 2, 2, 2, 5, 5, 5\}$, then $\mathcal{A} + \mathcal{B}$ has property P_6, thus $(16, 9, 31)$ is not 6–admissible.

Finally, we will study the following question: is it true that for any $k \in \mathbb{N}$, $k \geq 2$ there are infinitely many k–good primes? Based on Theorem 4 and considering the work related to Artin's conjecture [8], [9] one would expect that the answer is affirmative, however, this is certainly beyond reach at the moment. On the other hand, we can prove that the affirmative answer would follow from Schinzel's Hypothesis H [15], [16] (see also [7, p. 21]) which generalizes the twin prime conjecture:

Hypothesis H. If $k \in \mathbb{N}$, $F_1, \dots, F_k$ are distinct irreducible polynomials in $\mathbb{Z}[x]$ (with positive leading coefficients) and the product polynomial $F = F_1 \dots F_k$ has no fixed prime divisor, then there exist infinitely many integers n such that each $F_i(n)$ $(i = 1, \dots, k)$ is a prime.

Theorem 5. *If Hypothesis H is true, then for any primes $q_1 < \dots < q_s$ there are infinitely many primes p so that each of $q_1, \dots, q_s$ is a primitive root modulo p.*

Proof of Theorem 5. Let $r_1, \dots, r_t$ be the odd primes amongst $q_1, \dots, q_s$ (i.e., $\{r_1, \dots, r_t\} = \{q_1, \dots, q_s\} \setminus \{2\}$). For $i = 1, \dots, t$, let u_i denote an arbitrary quadratic non–residue modulo r_i. Consider the linear congruence system

$$\begin{aligned} 4x + 1 &\equiv u_1 \pmod{r_1} \\ &\vdots \\ 4x + 1 &\equiv u_t \pmod{r_t}. \end{aligned}$$

Clearly, each of these linear congruences can be solved, and the moduli are coprime, thus this system has a unique solution modulo $r_1 \dots r_t$. Let p_o be a positive element of this residue class so that

$$4p_o + 1 \equiv u_i \pmod{r_i} \text{ (for } i = 1, \dots, t). \tag{8.9}$$

Write

$$F_1(n) = p_o + nr_1 \dots r_t$$

and

$$F_2(n) = 4F_1(n) + 1 = (4p_o + 1) + 4nr_1 \dots r_t.$$

We will show that $F = F_1F_2$ has no fixed prime divisor. $F_2(n)$ is always odd and, $r_1 \dots r_t$ being odd, thus $F_1(n)$ is odd infinitely often, whence $F_1(n)F_2(n)$ is also odd infinitely often. For $i = 1, 2, \dots, t$, the number u_i is a quadratic non–residue modulo r_i, thus u_i cannot be congruent to 0 or 1 modulo r_i. By (8.9), it follows that

$$4F_1(n) \equiv 4p_o \equiv u_i - 1 \not\equiv 0 \pmod{r_i}$$

and

$$F_2(n) \equiv 4p_o + 1 \equiv u_i \not\equiv 0 \pmod{r_i}$$

so that $\big(r_i, F_1(n)F_2(n)\big) = 1$ for all i. Finally, if v is a prime different from each of $2, r_1, \ldots, r_t$, then

$$F_1(n)F_2(n) \equiv 0 \pmod{v} \tag{8.10}$$

is a quadratic congruence which has at most 2 solutions modulo v. Since $v > 2$, there is at least one residue class modulo v which does not satisfy (8.10), so far all n from this residue class $v \nmid F_1(n)F_2(n)$.

Thus, indeed, F_1F_2 has no fixed prime divisor, the polynomials $F_1, F_2 \in \mathbb{Z}[x]$ are linear and thus irreducible in $\mathbb{Z}[x]$, and their leading coefficients are postive, so that all the conditions in Hypothesis H hold. Since now this hypothesis is assumed to be true, thus there are infinitely many $n \in \mathbb{N}$ so that both

$$z = F_1(n) = p_o + nr_1 \ldots r_t \tag{8.11}$$

and

$$p = F_2(n) = 4z + 1 = (4p_o + 1) + 4nr_1 \ldots r_t \tag{8.12}$$

are primes. We will show that for such an n large enough, each of $2, r_1, \ldots, r_t$ is a primitive root modulo $p = p(n)$.

Since $p - 1 = 4z$ and z is a prime, thus all the positive divisors of $p - 1$ are $1, 2, 4, z, 2z$ and $4z$. Thus if $(g, p) = 1$ and g is not a primitive root modulo p, then we must have either

$$g^4 \equiv 1 \pmod{p} \tag{8.13}$$

or

$$g^{\frac{p-1}{2}} \equiv 1 \pmod{p}. \tag{8.14}$$

Since now p is assumed to be large, thus (8.13) does not hold for $g = 2, r_1, \ldots, r_t$. Thus if one of these numbers is not a primitive root modulo p, then it must satisfy (8.14) whence, by Euler's lemma,

$$\left(\frac{g}{p}\right) = +1$$

(where $\left(\frac{g}{p}\right)$ denotes the Legendre symbol). Thus it suffices to show that

$$\left(\frac{g}{p}\right) = -1 \text{ for } g = 2, r_1, \ldots, r_t. \tag{8.15}$$

By (8.12) we have $p = 4z + 1$ where z is an odd prime, and thus p is of form $8k + 5$, whence (8.15) follows if $g = 2$. If $g = r_i$, $1 \le i \le t$, then by the quadratic reciprocity law we have

$$\left(\frac{r_i}{p}\right) = (-1)^{\frac{r_i-1}{2}\cdot\frac{p-1}{2}} \left(\frac{p}{r_i}\right). \tag{8.16}$$

By (8.12), $\frac{p-1}{2} = 2z$ is even and thus

$$(-1)^{\frac{r_i-1}{2}\cdot\frac{p-1}{2}} = +1. \tag{8.17}$$

Moreover, by (8.9) and (8.12) we have

$$p \equiv 4p_0 + 1 \equiv u_i \pmod{r_i}$$

whence, by the definition of u_i,

$$\left(\frac{p}{r_i}\right) = \left(\frac{u_i}{r_i}\right) = -1. \tag{8.18}$$

(8.15) with r_i in place of g follows from (8.16), (8.17) and (8.18), and this completes the proof of Theorem 5.

9 Extension of the notion of f–complexity and a construction with high f–complexity

In [4] we introduced the notion of f–complexity ("f" for family) of families of binary sequences. This notion can be generalized easily to families of k symbols:

Definition 6. If $\mathcal{A}$ is a set of k symbols, $N, t \in \mathbb{N}$, $t < N$, $(\varepsilon_1, \dots, \varepsilon_t) \in \mathcal{A}^t$, $i_1, \dots, i_t$ are positive integers with $1 \le i_1 < \dots < i_t \le N$, and we consider sequences $E_N = (e_1, \dots, e_N) \in \mathcal{A}^N$ with

$$e_{i_1} = \varepsilon_1, \dots, e_{i_t} = \varepsilon_t, \tag{9.1}$$

then $(e_{i_1}, \dots, e_{i_t}; \varepsilon_1, \dots, \varepsilon_t)$ is said to be a *specification* of E_N of length t or a t*–specification* of E_N.

Definition 7. The f*–complexity* of a family $\mathcal{F}$ of sequences $E_N \in \mathcal{A}^N$ of k symbols is defined as the greatest integer t so that for any t–specification (9.1) there is at least one $E_N \in \mathcal{F}$ which satisfies it. The f–complexity of $\mathcal{F}$ is denoted by $\Gamma_k(\mathcal{F})$. (If there is no $t \in \mathbb{N}$ with the property above, we set $\Gamma_k(\mathcal{F}) = 0$.)

Note that the $k = 2$ special case of this definition is the notion of f–complexity of families of binary sequences introduced in [4].

One might like to show that the family constructed in Theorem 2, or at least a slightly modified version of it, is also of high f–complexity. Unfortunately, we have been able to prove only a partial result in this direction: we can handle only the case when k, the size of the alphabet is a prime number (this, of course, includes the binary case). We will explain the difficulties arising in the case of composite k later. We hope to return to this case in a subsequent paper, and there we will present other constructions where the f–complexity can be handled also for composite k.

Theorem 5. *Assume that k, p are prime numbers, χ is a (multiplicative) character modulo p of order k (so that $k \mid p-1$), $H \in \mathbb{N}$, $H < p$. Consider all the polynomials $f(x) \in F_p[x]$ with the properties that*

$$0 < \deg f(x) \le H \tag{9.2}$$

and

$$\text{in } \bar{F}_p \text{ the multiplicity of each zero of } f(x) \text{ is less than } k. \tag{9.3}$$

For each of these polynomials $f(x)$, consider the sequence $E_p = E_p(f) = (e_1, \dots, e_p)$ of k–th roots of unity defined as in Theorem 2:

$$e_n = \begin{cases} \chi\big(f(n)\big) & \text{for } \big(f(n), p\big) = 1 \\ +1 & \text{for } p \mid f(n). \end{cases}$$

Then we have

$$\delta(E_p) < 11Hp^{1/2}\log p. \tag{9.4}$$

Moreover, if $\ell \in \mathbb{N}$ and

(i) either

$$(4H)^\ell < p \tag{9.5}$$

(ii) or k is a primitive root modulo p and $\ell < p$,

then also

$$\gamma_\ell(E_p) < 10\ell Hkp^{1/2}\log p \tag{9.6}$$

holds. Finally, we have

$$\Gamma_k(\mathcal{F}) \geq H. \tag{9.7}$$

Proof of Theorem 5. The proof is a combination and extension of Theorem 1 in [4] and Theorem 2 above, thus we will leave some details to the reader.

In order to prove (9.4), we argue in the same way as in the proof of (3.2) in the proof of Theorem 2. Again we set $g(x) = f(u + xv)$ and $\chi_1 = \chi^t$ with

$$1 \leq t \leq k - 1. \tag{9.8}$$

Then by (9.3) the multiplicity of the zeros of $g(x)$ is less than k, and since the order of χ is k and k is now a prime number, thus it follows from (9.8) that the character χ_1 is also of order k. Thus by Lemma 2, again (3.16) holds with H in place of h, and then we may complete the proof of (9.4) in the same way as the proof of (3.2) was completed.

Similarly, in order to prove (9.6), we argue as in the proof of (3.3) in the proof of Theorem 2. We define $B, f_1(x)$ and $G(x)$ as there: $f(x) = Bf_1(x)$, $f_1(x)$ is unitary,

$$G(x) = f_1(x + d_1)^{t_1} \dots f_1(x + d_\ell)^{t_\ell} \tag{9.9}$$

with

$$0 \leq t_1, \dots, t_\ell \leq k - 1, \ (t_1, \dots, t_\ell) \neq (0, \dots, 0), \tag{9.10}$$

and again we get that (3.25) and (3.26) hold, and it suffices to show that the analogue of Lemma 3 holds:

Lemma 5. *If k, f, H, ℓ are defined as in Theorem 5, then $G(x)$ has at least one zero (in $\bar{F}_p$) whose multiplicity is not divisible by k.*

Indeed, assuming that Lemma 5 holds, the proof of (9.6) can be completed in the same way (with H in place of h) as the proof of (3.3) using Lemma 3. Thus it remains to prove Lemma 5.

Proof of Lemma 5. We argue as in the proof of Lemma 3, i.e., we consider the same equivalence relation as there, then we write $f_1(x)$ as the product of irreducible polynomials over F_p, and finally we group these factors so that in each group the equivalent irreducible factors are collected. However, there is a crucial difference compared with Lemma 3: while in Theorem 2 we assumed that $f(x)$ has no multiple zero, now this condition is relaxed to the weaker condition (9.3). It follows that now the irreducible factors may have an exponent not exceeding $k-1$. So now a typical group of the equivalent irreducible factors looks like $\varphi(x+a_1)^{s_1}, \dots, \varphi(x+a_r)^{s_r}$ where

$$1 \le s_1, \dots, s_r \le k-1. \tag{9.11}$$

Then writing $G(x)$ in (9.9) as the product of irreducible polynomials over F_p, all the polynomials $\varphi(x+a_i+d_j)$ with $1 \le i \le r$, $1 \le j \le \ell$ occur amongst the factors, and for fixed i, j such a factor occurs with exponent exactly $s_i t_j$. Since now k is a prime, thus it follows from (9.10) and (9.11) that

$$\text{if } s_i t_j > 0 \text{ then } k \nmid s_i t_j. \tag{9.12}$$

The conclusion of Lemma 5 fails, i.e., the multiplicity of each of the zeros of $G(x)$ is divisible by k if and only if each of the factors $\varphi(x+a_i+d_j)$ occurs with an exponent divisible by k. This is so if and only if the following holds: if $\mathcal{A}$ denotes the k–set whose elements are $a_1, \dots, a_r$, each a_i taken with multiplicity s_i, and $\mathcal{B}$ denotes the k–set whose elements are $d_1, \dots, d_\ell$, each d_j taken with multiplicity t_j, then $\mathcal{A}+\mathcal{B}$ possesses property P_k. Take any of the groups formed by the equivalent irreducible factors (by (9.2) there is at least one such group), and consider the corresponding sum $\mathcal{A}+\mathcal{B}$ with property P_k. Then $(|\mathcal{A}|, |\mathcal{B}|, p)$ is **not** a (k,k)–admissible triple, and here we have

$$|\mathcal{A}| = \sum_{i=1}^{r} s_i \le \sum_{i=1}^{r} (k-1) = r(k-1) \le (\deg f_1)(k-1) \le H(k-1)$$

and

$$|\mathcal{B}| = \sum_{j=1}^{\ell} t_j \le \ell(k-1).$$

It remains to show that assuming either (i) or (ii) (in Theorem 5), this is impossible.

(Observe that now we are studying (k,k)–admissibility instead of the k–admissibility occuring in the proof of Theorem 2; this is the price paid for relaxing the condition on the zeros of the polynomial $f(x)$ which is necessary for controlling the f–complexity. It is much more difficult to control (k,k)–admissibility than k–admissibility, since if we study (k,k)–admissibility then the set $\mathcal{A}$ in the sums $\mathcal{A}+\mathcal{B}$ considered also can be a multiset, thus we have more flexibility in constructing negative examples. Indeed, when k is composite, and both $\mathcal{A}$ and $\mathcal{B}$ can be k–sets, then it is easy to give negative examples of the type described in Example 5; this is why we cannot control the f–complexity for composite k.)

Assume first that (i) holds. Let $\bar{\mathcal{A}}$ and $\bar{\mathcal{B}}$ denote the set of the distinct elements of $\mathcal{A}$, resp. $\mathcal{B}$: $\bar{\mathcal{A}} = \{a_1, \dots, a_r\}$, $\bar{\mathcal{B}} = \{d_1, \dots, d_\ell\}$. Then by (9.2) and (9.5) we have

$$(4r)^\ell \leq (4 \deg f_1)^\ell = (4 \deg f)^\ell \leq (4H)^\ell < p$$

so that (9.1) in Theorem 3, (ii) holds with r and ℓ in place of t, tesp. r. Thus the argument in the proof of Theorem 3, (ii) can be used with $k = 2$, and then we obtain that there is a $c \in \mathbb{Z}_p$ which has a unique representation in the form

$$d_j + a_i = c,\ d_j \in \bar{\mathcal{B}}, a_i \in \bar{\mathcal{A}}.$$

It follows that, considering also multiplicities,

$$a_i + d_j = c,\ a_i \in \mathcal{A}, d_\ell \in \mathcal{B}$$

has exactly $s_i t_j (> 0)$ solutions. By (9.12), this contradicts the assumption that $\mathcal{A} + \mathcal{B}$ has property P_k which completes the proof in this case.

Assume now that (ii) holds. Then we use the notations of the proof of Theorem 4, so that, by (9.11), $Q_{\mathcal{A}}(x) = \sum\limits_{i=1}^{r} s_i x^{s(a_i)} \in F_k[x]$, by (9.10) $Q_{\mathcal{B}}(x) = \sum\limits_{j=1}^{\ell} t_j x^{s(d_j)} \in F_k[x]$, and, since $\mathcal{A} + \mathcal{B}$ possesses property P_k, thus $Q_{\mathcal{A}+\mathcal{B}}(x) = 0$ in $F_k[x]$. Again, (8.2) holds, whence it follows that $x^{p-1} + x^{p-2} + \dots + x + 1$ divides $Q_{\mathcal{A}}(x) Q_{\mathcal{B}}(x)$. Since it is now assumed that k is a primitive root modulo p, thus by Lemma 4 the polynomial $x^{p-1} + x^{p-2} + \dots + x + 1$ is irreducible over F_k. It follows that $x^{p-1} + x^{p-2} + \dots + x + 1$ divides either $Q_{\mathcal{A}}(x)$ or $Q_{\mathcal{B}}(x)$, so that either $Q_{\mathcal{A}}(x)$ or $Q_{\mathcal{B}}(x)$ is a constant multiple of this polynomial, but this is impossible by $r \leq \deg f \leq H < p$ and $\ell < p$, and this completes the proof of (9.6)

It remains to prove (9.7). As in [4], we use

Lemma 6. *If T is a field and $g(x) \in T[x]$ is a non–zero polynomial, then it can be written in the form*

$$g(x) = \big(h(x)\big)^k g^*(x) \tag{9.13}$$

where the multiplicity of each zero of $g^*(x)$ (in $\bar{F}_p$) is less than k.

Proof of Lemma 6. The $k = 2$ special case of this lemma was stated and proved in [4] as Lemma 1, and the general case presented here can be proved in the same way, thus we leave the details to the reader.

To prove (9.7), we have to show that for any specification of length H:

$$e_{i_1} = \varepsilon_1, \dots, e_{i_H} = \varepsilon_H \ (i_1 < \dots < i_H), \tag{9.14}$$

there is a polynomial $f(x) \in F_p[x]$ which satisfies (9.2) and (9.3) so that $E_p = E_p(f) \in \mathcal{F}$, and this sequence $E_p = E_p(f)$ satisfies the specification (9.14).

By $H < p$, there is an integer i_{H+1} with $0 < i_{H+1} \leq p$, $i_{H+1} \notin \{i_1, \dots, i_H\}$. Let ε_0 be a k–th root of unity with

$$\varepsilon_0 \neq 1, \tag{9.15}$$

and set

$$\varepsilon_{H+1} = \varepsilon_0 \varepsilon_1. \tag{9.16}$$

Denote the distinct k–th roots of unity by $\varphi_1, \dots, \varphi_k$, let $v_1, \dots, v_k$ be integers with

$$\chi(v_i) = \varphi_i \text{ (for } i = 1, \dots, k),$$

and define $y_1, \dots, y_{H+1}$ by

$$y_i = v_z \text{ where } z = z(i) \text{ is defined by } \varphi_z = \varepsilon_i. \tag{9.17}$$

By the well–known theorem on interpolation, there is a unique polynomial $g(x) \in F_p[x]$ with

$$\deg g(x) \leq H \tag{9.18}$$

and

$$g(i_j) = y_j \text{ for } j = 1, \dots, H+1. \tag{9.19}$$

(This polynomial can be determined by using either Lagrange interpolation or Newton interpolation.) By Lemma 6 (with $T = F_p$), this polynomial $g(x)$ can be written in form (9.13). Let

$$f(x) = g^*(x). \tag{9.20}$$

Then by Lemma 6, (9.3) holds. It follows from (9.13), (9.18) and (9.20) that

$$\deg f(x) = \deg g^*(x) \leq \deg g(x) \leq H. \tag{9.21}$$

By (9.17) and (9.19) we have

$$g(i_j) = y_j = v_{z(j)}$$

so that

$$\chi\big(g(i_j)\big) = \chi(v_{z(j)}) = \varphi_{z(j)} \ (\neq 0) \tag{9.22}$$

and thus

$$\big(g(i_j), p\big) = 1 \text{ for } j = 1, \dots, H+1. \tag{9.23}$$

By (9.13), (9.17), (9.20), (9.22) and (9.23) we have

$$\chi\big(g(i_j)\big) = \chi\big(\big(h(i_j)\big)^k\big)\chi\big(g^*(i_j)\big) = \chi\big(f(i_j)\big) = \varphi_{z(j)} = \varepsilon_j \text{ for } j = 1, \dots, H+1. \tag{9.24}$$

It follows from (9.15), (9.16) and (9.24) that

$$\chi\big(f(i_1)\big) \neq \chi\big(f(i_{H+1})\big)$$

and thus $f(x)$ is not constant, i.e.,

$$\deg f(x) > 0. \tag{9.25}$$

(9.2) follows from (9.21) and (9.25). Finally, it follows from (9.24) and the definition of $E_p(f)$ that $E_p(f)$ satisfies the specification (9.14) and this completes the proof of the theorem.

10 On the cardinality of a smallest family achieving a prescribed f–complexity and multiplicity

We introduce first k–ary extensions of two quantities studied in [4].

Definition 8. For positive integers $j \leq K \leq N, M$ and the alphabet $\mathcal{A} = \{a_1, \dots, a_k\}$ set

$$S(N,j,M,k) = \min\big\{|\mathcal{F}| : \mathcal{F} \subset \mathcal{A}^N, \text{ for every } (\varepsilon_1, \dots, \varepsilon_j) \in \mathcal{A}^j \text{ and } 1 \leq i_1 < \dots < i_j \leq N$$
$$\text{there are at least } M \text{ members } E_N = (e_1, \dots, e_N) \text{ of } \mathcal{F} \text{with } j\text{–specification } (e_{i_1}, \dots, e_{i_j}; \varepsilon_1, \dots, \varepsilon_j)\big\}. \tag{10.1}$$

We also say for the $\mathcal{F}$'s considered here that they *cover* every j–specification with multiplicity $\geq M$.

In particular for $M = 1$ and $j = K$ we get

$$S(N,K,k) \triangleq S(N,K,1,k) = \min\big\{|\mathcal{F}| : \mathcal{F} \subset \mathcal{A}^N, \Gamma_k(\mathcal{F}) = K\big\}, \tag{10.2}$$

which counts how many sequences $E_N \in \mathcal{A}^N$ are needed to cover all K specifications, that is, to have f–complexity $\Gamma_k(\mathcal{F}) = K$.

Finding the number can be formulated as a covering problem for the hypergraph

$$\mathcal{H}(N,K,k) = \big(\mathcal{V}(N,K,k), \mathcal{E}(N,k)\big),$$

where $\mathcal{E}(N,k) = \mathcal{A}^N$ is the edge set and the vertex set $\mathcal{V}(N,K,k)$ is defined as the set of K–specifications for $\mathcal{A}^N$ or, equivalently, as set of $(N-K)$–dimensional subcubes of $\mathcal{A}^N$ and thus

$$|\mathcal{V}(N,K,k)| = \binom{N}{K} k^K, |\mathcal{E}(N,k)| = k^N \tag{10.3}$$

$E_N \in \mathcal{E}(N,k)$ contains specification V iff E_N"$\in$"V. We derive now bounds on $S(N,K,k)$ and use (as in [4] for $k = 2$)

Lemma 7. (Covering Lemma 1 of [1]) *For any hypergraph $(\mathcal{V}, \mathcal{E})$ with*

$$\min_{v \in \mathcal{V}} \deg(v) \geq d \tag{10.4}$$

there exists a covering $\mathcal{C} \in \mathcal{E}$ with

$$|\mathcal{C}| \leq \left\lceil \frac{|\mathcal{E}|}{d} \log |\mathcal{V}| \right\rceil.$$

Theorem 6. *The cardinality $S(N,K,k)$ of a smallest family $\mathcal{F} \subset \mathcal{A}^N$ with f–complexity $\Gamma_k(\mathcal{F}) = K$ satisfies*

$$k^K \leq S(N,K,k) \leq k^K \log \binom{N}{K} k^K \leq k^K K \log N \quad (K \geq k^3).$$

Proof: Application of Lemma 7 to our hypergraph $\mathcal{H}(N,K,k)$ yields with $d = k^{N-K}$ a family $\mathcal{F}$ with $\Gamma_k(\mathcal{F}) \geq K$,

$$|\mathcal{F}| \leq \left\lceil \frac{k^N}{k^{N-K}} \log \binom{N}{K} k^K \right\rceil \leq k^K K \log N \quad (K \geq k^3)$$

and thus the upper bound for $S(N,K,k)$.

On the other hand one edge E_N covers exactly $\binom{N}{K}$ K–specifications and therefore by (10.3) necessarily as lower bound we have

$$S(N,K,k) \geq k^K.$$

We explaned already in [4] that in order to make it difficult for an eavesdropper to identify a key $E_N \in \mathcal{F}$, when he has observed j positions, we must leave him many options. This can be achieved by constructing a family $\mathcal{F}$ of high f–complexity $\Gamma_k(\mathcal{F})$. Indeed for $j < \Gamma_k(\mathcal{F})$ the multiplicity $M_j(\mathcal{F})$, that is, the least multiplicity of every j–specification satisfies

$$M_j(\mathcal{F}) \geq k^{\Gamma_k(\mathcal{F})-j}, \tag{10.5}$$

because a j–specification can be extended to as many $\Gamma_k(f)$–specifications with the same support. Therefore

$$\min_{\mathcal{F}:\Gamma_k(\mathcal{F})\geq K} M_j(\mathcal{F}) \geq k^{K-j} \tag{10.6}$$

and thus

$$S(N,j,k^{K-j},k) \leq S(N,K,k) \leq k^K K \log N \quad (K \geq k^3). \tag{10.7}$$

On the other hand, since $|\mathcal{V}(N,j,k)| = \binom{N}{j} k^j$ and an edge E_N covers exactly $\binom{N}{j}$ j–specifications, necessarily

$$S(N,j,k^{K-j},k) \geq k^{K-j} \binom{N}{j} k^j \binom{N}{j}^{-1} = k^K. \tag{10.8}$$

Quite surprisingly, for $K \log N$ small relative to k^K the two bounds are very close to each other. The fact that $S(N,K,k)$ and therefore f–complexity contains almost complete information about the quantity $S(N,j,k^{K-j},k)$ measuring multiplicity for the eavesdropper demonstrates the usefulness of our complexity measure. We summarize these findings.

Theorem 7. *The cardinality $S(N,j,k^{K-j},k)$ of a smallest family $\mathcal{F} \subset \mathcal{A}^N$ which covers every j–specification with multiplicity $\geq k^{K-j}$ satisfies for all $j \leq K \leq N$*

$$k^K \leq S(N,j,k^{K-j},k) \leq S(N,K,k) \leq k^K K \log N \quad (K \geq k^3).$$

11 Balanced families with prescribed f–complexity

Definition 9. A family $\mathcal{F} \subset \mathcal{A}^N$ with f–complexity $\Gamma_k(\mathcal{F}) = K$ is said to be c–balanced for some constant $c \in \mathbb{N}$, if no K–specification is covered by more than c sequences $E_N \in \mathcal{F}$.

We improve now Theorem 6 by adding c–balancedness.

Theorem 8. *For $c = \log|\mathcal{V}(N,K,k)| = \log\binom{N}{K}k^K \leq K\log N$ $(K \geq k^3)$ the smallest c–balanced family $\mathcal{F} \subset \mathcal{A}^N$ with f–complexity $\Gamma_k(\mathcal{F}) = K$ has a cardinality meeting the bounds on $S(N,K,k)$ in Theorem 6.*

Proof: We replace Lemma 7 by a lemma on balanced coverings.

Definition 10. A covering $\mathcal{C} \triangleq \{E_1,\dots,E_L\}$ of a hypergraph $\mathcal{H} = (\mathcal{V},\mathcal{E})$ is called *c–balanced* for some constant $c \in \mathbb{N}$, if no vertex occurs in more than c edges of $\mathcal{C}$.

Lemma 8. (Covering Lemma 3 of [1, Part II]) *A hypergraph $\mathcal{H} = (\mathcal{V},\mathcal{E})$ with maximum and minimum degrees $d_{\max} \triangleq \max\limits_{v\in\mathcal{V}} \deg(v)$ and $d_{\min} \triangleq \min_{v\in\mathcal{V}} \deg(v) > 0$ has a c–balanced covering $\mathcal{C} = \{E_1,\dots,E_L\}$ if*

(a) $L \geq \lceil |\mathcal{E}|d_{\min}^{-1} \cdot \log|\mathcal{V}|\rceil + 1$

(b) $c \leq L \leq c|\mathcal{E}|d_{\max}^{-1}$

(c) $\exp\left\{-D\left(\lambda\|\frac{d_{\max}}{|\mathcal{E}|}\right)L + \log|\mathcal{V}|\right\} < \frac{1}{2}$ *for* $\lambda \triangleq \frac{c}{L}$

(Here D denotes the Kullbach–Leibler divergence.)

Using Lemma 8 with $d_{\min} = d_{\max} = d = k^{N-K}$ and

$$c = \log|\mathcal{V}| = \log|\mathcal{B}(N,K,k)| = \log\binom{N}{K}k^K \leq K\log N \quad (K \geq k^3)$$

we get a c–balanced covering of said cardinality.

Remark: Using Theorem 8 also the bounds in Theorem 7 can be obtained in a c–balanced way with $c = K\log N$ by the previous reasoning.

Next we go for improvements of the balancedness property. It is known from probability theory that for large deviations the following inequality holds:

For an i.i.d. sequence $Z_1, Z_2, \dots, Z_L$ of RV with values in $[0,1]$ and expection $EZ_i = \mu$ for $0 < \varepsilon < 1$

$$Pr\left\{\frac{1}{L}\sum_{i=1}^{L} Z_i \notin \big[(1-\varepsilon)\mu, (1+\varepsilon)\mu\big]\right\} \leq 2\exp\left(-L\frac{\varepsilon^2\mu}{2\ell n2}\right).$$

This can be used to establish another balancedness property, which also gives a bound from below, but in exchange most, but not necessarily all vertices satisfy it. This suggests to apply a more recent auxiliary result.

Lemma 9. [2] *Let $\mathcal{H} = (\mathcal{V},\mathcal{E})$ be an e–uniform hypergraph (all edges cardinality $= e$) and P a PD on $\mathcal{E}$. Consider a PD Q on $\mathcal{V}$: $Q(v) \triangleq \sum\limits_{E\in\mathcal{E}} P(E)\frac{1}{e}1_E(v)$.*

Fix $\varepsilon, \tau > 0$, and define the set of vertices $\mathcal{V}_0 = \left\{ v \in \mathcal{V} : Q(v) < \frac{\tau}{|\mathcal{V}|} \right\} \subset \mathcal{V}$, then there exist edges $E^{(1)}, \dots, E^{(L)} \in \mathcal{E}$ such that for

$$\bar{Q}(v) \triangleq \frac{1}{L} \sum_{i=1}^{L} \frac{1}{e} 1_{E^{(i)}}(v)$$

(i) $Q(\mathcal{V}_0) \leq \tau$

(ii) $(1-\varepsilon)Q(v) \leq \bar{Q}(v) \leq (1+\varepsilon)Q(v)$ *for all* $v \in \mathcal{V} \smallsetminus \mathcal{V}_0$

(iii) $L \leq \left\lceil \frac{|\mathcal{V}|}{e} \frac{2\ell n 2 \log(2|\mathcal{V}|)}{\varepsilon^2 \tau} \right\rceil$.

We apply this lemma now to the e–uniform hypergraph $\mathcal{H}(N,K,k)$, whose edges have cardinality $e = \binom{N}{K}$. First notice that

$$L \leq \frac{\binom{N}{K} k^K}{\binom{N}{K}} \frac{3}{\varepsilon^2 \tau} \log \binom{N}{M} k^K = \frac{3}{\varepsilon^2 \tau} k^K \log \binom{N}{K} k^K \leq \frac{3}{\varepsilon^2 \tau} k^K K \log N (K \geq k^3).$$

Except for the constant $\frac{3}{\varepsilon^2 \tau}$ this is our previous bound.
Next choose as P the uniform PD on $\mathcal{E}(N,k)$. Then for all vertices $v \in \mathcal{V}(N,K,k)$

$$Q(v) = \sum_{E_N \in \mathcal{E}(N,k)} k^{-N} \binom{N}{K}^{-1} 1_{E_N}(v) = k^{-N} \binom{N}{K}^{-1} \deg(v) = k^{-N} \binom{N}{K}^{-1} k^{N-K} = \frac{1}{\binom{N}{K} k^K} \tag{11.1}$$

and for $v \in \mathcal{V} \smallsetminus \mathcal{V}_0$

$$(1-\varepsilon) Le\ Q(v) \leq \sum_{i=1}^{L} 1_{e^{(i)}}(v) \leq (1+\varepsilon) Le\ Q(v)$$

and for $\tau = 3/4$

$$(1-\varepsilon) \frac{4}{\varepsilon^2} K \log N \leq \sum_{i=1}^{L} 1_{E^{(i)}}(v) \leq (1+\varepsilon) \frac{4}{\varepsilon^2} K \log N. \tag{11.2}$$

This implies the uniformity property

$$\begin{aligned} \frac{1-\varepsilon}{1+\varepsilon} &\leq \min_{v,v' \in \mathcal{V} \smallsetminus \mathcal{V}_0} \left(\sum_{i=1}^{L} 1_{E^{(i)}}(v) \right) \left(\sum_{i=1}^{L} 1_{E^{(i)}}(v') \right)^{-1} \\ &\leq \max_{v,v' \in \mathcal{V} \smallsetminus \mathcal{V}_0} \left(\sum_{i=1}^{L} 1_{E^{(i)}}(v) \right) \left(\sum_{i=1}^{L} 1_{E^{(i)}}(v') \right)^{-1} \leq \frac{1+\varepsilon}{1-\varepsilon}. \end{aligned} \tag{11.3}$$

By choosing τ small most vertices are in $\mathcal{V} \smallsetminus \mathcal{V}_0$.

Now comes a **surprise**. Our hypergraph has strong symmetries and by (11.1) $Q(v)$ is independent of v. Therefore for $\tau = 3/4 < 1$ $\mathcal{V}_0 = \phi$ and (11.3) holds for all vertices. We have established

Theorem 10. *For every* $\varepsilon \in (0,1)$ *there is a family* $\mathcal{F} \subset \mathcal{A}^N$ *with* f*–complexity* $\Gamma_K(\mathcal{F}) = K$, $k^K \leq |\mathcal{F}| \leq \frac{4}{\varepsilon^2} k^K \log N (K \geq k^3)$ *such that for every* K*–specification the number of sequences* $E_N \in \mathcal{F}$ *which cover this specification lies between* $\frac{4(1-\varepsilon)}{\varepsilon^2} K \log N$ *and* $\frac{4(1+\varepsilon)}{\varepsilon^2} K \log N$.

12 Conclusion

We have constructed large families of sequences of k symbols with strong pseudorandom properties. We have also introduced and studied the notion of f–complexity of families of sequences of k symbols, and we have shown that the f–complexity of the family constructed by us is large if k, the size of the alphabet is a prime number but we have not been able to control the case when k is composite. We have also shown what are essentially minimal cardinalities of families with prescribed complexity and which additional multiplicity properties they may have.

One might like to construct families of large complexity for composite k as well; we will return to this problem in a subsequent paper.