Pseudorandom binary functions on Bratteli diagrams

by

Katalin Gyarmati
Eötvös Loránd University, Department of Algebra and Number Theory
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary
e-mail: gykati@cs.elte.hu

Pascal Hubert
Aix-Marseille Université, CNRS, Centrale Marseille,
I2M, UMR7373, 13453 Marseille, France
e-mail: pascal.hubert@univ-amu.fr

András Sárközy
Eötvös Loránd University, Department of Algebra and Number Theory
H-1117 Budapest, Pázmány Péter sétány 1/C, Hungary
e-mail: sarkozy@cs.elte.hu

Abstract

In two earlier papers the authors studied pseudorandomness of binary functions defined on uniform trees and rooted plane trees, i.e., of functions of the type $f : \mathcal{V}(T) \to \{-1, +1\}$ where $\mathcal{V}(T)$ is the vertex set of the tree. Here we extend the problem further by considering Bratteli diagrams.

1 Introduction

Recently in a series of papers a new constructive and quantitative approach has been developed to study pseudorandomness of binary sequences

$$E_N = \{e_1, \dots, e_N\} \in \{-1, +1\}^N.$$

2010 Mathematics Subject Classification: 11K99, 11Z99.
Key words and phrases: Bratteli diagram, binary function, pseudorandomness.
Research partially supported by the NKFIH grant K119528.

In particular, first in [16] the following measures of pseudorandomness were introduced: the *well-distribution measure* of E_N is defined by

$$W(E_N) = \max_{a,b,t} \left| \sum_{j=0}^{t-1} e_{a+jb} \right|$$

where the maximum is taken over all $a, b, t \in \mathbb{N}$ with $1 \le a \le a+(t-1)b \le N$, the correlation measure *of order* k of E_N is defined as

$$C_k(E_N) = \max_{M,\mathbf{D}} \left| \sum_{n=1}^{M} e_{n+d_1} \dots e_{n+d_k} \right|$$

where the maximum is taken over all $\mathbf{D} = (d_1, \dots, d_k)$ and M such that $0 \le d_1 < \dots < d_k \le N - M$, and the normality measure of order k of E_N is defined as

$$N_k(E_N) = \max_{X \in \{-1,+1\}^k} \max_{0<M<N+1-k} \left| \left| \{ n : 0 \le n < M, (e_{n+1}, \dots, e_{n+k}) = \mathbf{X} \} \right| - \frac{M}{2^k} \right|.$$

We note that the combination of the well-distribution measure and correlation measure of order k, called *combined pseudorandom measure of order* k, was also introduced in [16]:

$$Q_k(E_N) = \max_{a,b,t,\mathbf{D}} \left| \sum_{j=0}^{t} e_{a+jb+d_1} e_{a+jb+d_2} \dots e_{a+jb+d_k} \right|$$

where the maximum is taken over all $a, b, t, \mathbf{D} = (d_1, d_2, \dots, d_k)$ such that $0 < d_1 < d_2 < \dots < d_k$ and all the subscripts $a + jb + d_i$ belong to $\{1, 2, \dots, N\}$. The advantage of using this measure is that clearly we have $\max\{W(E_N), C_k(E_N)\} \le Q_k(E_N)$, thus it suffices to give an upper bound for Q_k. On the other hand, restricting ourselves to the study of Q_k has two disadvantages: first, the formulas are much more complicated than in case of W and C_k, and secondly, it can be a useful information to know if, say, W is small while C_k is large or vice versa, but studying only Q_k these facts cannot be detected. Thus typically both the well-distribution and correlation measures are studied instead of the combined measures; in this paper, we will also use this approach.

Then the sequence E_N is considered to be a "good" pseudorandom sequence if both $W(E_N)$ and $C_k(E_N)$ (at least for "small" k) are "small" in terms of N; in particular, both are $o(N)$ as $N \to \infty$. (It was shown in [16] that the normality measures can be estimated in terms of the correlation measures.) Indeed, later Cassaigne, Mauduit and Sárközy [5] proved

that this terminology is justified since for almost all $E_N \in \{-1,+1\}^N$ both $W(E_N)$ and $C_k(E_N)$ are less than $N^{1/2}(\log N)^C$ (see also [2], [15]). It was also shown in [16] that the Legendre symbol forms a "good" pseudorandom sequence.

[16] was followed by numerous papers written on pseudorandomness of binary sequences.

In particular, the problem of pseudorandomness of sequences of k symbols was studied first by Mauduit and Sárközy [17]. They introduced the following definitions and notations (which we will need later):

Let $k \in \mathbb{N}$, $k \geq 2$, and let $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$ be a finite set ("alphabet") of k symbols ("letters"), and consider a sequence $E_N = (e_1, e_2, \ldots, e_N) \in \mathcal{A}^N$ of these symbols. They introduced the following new measures of pseudorandomness (motivated by the definition of the normality measure above, we may formulate these definitions in terms of any fixed ℓ-tuple, also called as "word", $\left(a_{i_1}, a_{i_2}, \ldots, a_{i_\ell}\right) \in \mathcal{A}^\ell$ occurring with the expected *frequency* in certain positions in E_N). For $a \in \mathcal{A}$ and positive integers M, u, v write

$$\sigma(E_N, a, M, u, v) = \left|\{j : 0 \leq j \leq M-1,\, e_{u+jv} = a\}\right|$$

and for $w = \left(a_{i_1}, a_{i_2}, \ldots, a_{i_\ell}\right) \in \mathcal{A}^\ell$, positive integers M, u, v and $\mathbf{D} = (d_1, d_2, \ldots, d_\ell)$ with non-negative integers $d_1 < d_2 < \cdots < d_\ell$ write

$$\tau(E_N, w, M, \mathbf{D}) = \left|\left\{n : 1 \leq n \leq M, \left(e_{n+d_1}, e_{n+d_2}, \ldots, e_{n+d_\ell}\right) = w\right\}\right|.$$

Then the *f-well-distribution* ("*f*" for "frequency") *measure* of E_N is defined as

$$\delta(E_N) = \max_{a,M,u,v} \left|\sigma(E_N, a, M, U, V) - \frac{M}{k}\right|$$

where the maximum is taken over all $a \in \mathcal{A}$ and u, v, M with $u + (M-1)v \leq N$, while the *f-correlation measure of order* ℓ of E_N is defined as

$$\gamma_\ell(E_N) = \max_{w,M,D} \left|\tau(E_N, w, M, \mathbf{D}) - \frac{M}{k^\ell}\right|$$

where the maximum is taken over all $w \in \mathcal{A}'$, and $\mathbf{D} = (d_1, \ldots, d_\ell)$ and M such that $M + d_\ell \leq N$. It was shown later by Bérczi [3] that for almost all $E_N \in \mathcal{A}^N$ both these measures are small in terms of N.

Moreover, in [12] and [13] we studied pseudorandomness of binary functions defined on almost uniform trees and on rooted plane trees, respectively. In this paper our goal is to extend the problem further by studying binary functions on *Bratteli diagrams*.

The notion of Bratteli diagram was introduced in 1972 by O. Bratteli [4]. They are used in the theory of algebras, dimension groups, substitutions,

transcendence, geometry and theoretical physics; the connection between Bratteli diagrams and these fields is discussed in [8]. (See also [7] for further applications.) The notion of Bratteli diagram is defined in [8] in the following way:

"**Definition 1.** A *Bratteli diagram* is an infinite directed graph (V, E) such that the vertex set V and the edge set E can be partitioned into finite sets

$$V = V_0 \cup V_1 \cup V_2 \cup \dots \quad \text{and} \quad E = E_1 \cup E_2 \cup \dots$$

with the following properties:

i) $V_0 = \{v_0\}$ is a one-point set.

ii) $r(E_n) \subseteq V_n$, $s(E_n) \subseteq V_{n-1}$, $n = 1, 2, \dots$ where r is the associated range map and s is the associated source map. Also, $s^{-1}(v) \neq \emptyset$ for all $v \in V$ and $r^{-1}(v) \neq \emptyset$ for all $v \in V \setminus V_0$.

There is an obvious notion of *isomorphism* between Bratteli diagrams (V, E) and (V', E'): namely, there exist a pair of bijections between V and V' and between E and E', respectively, preserving the gradings and intertwining

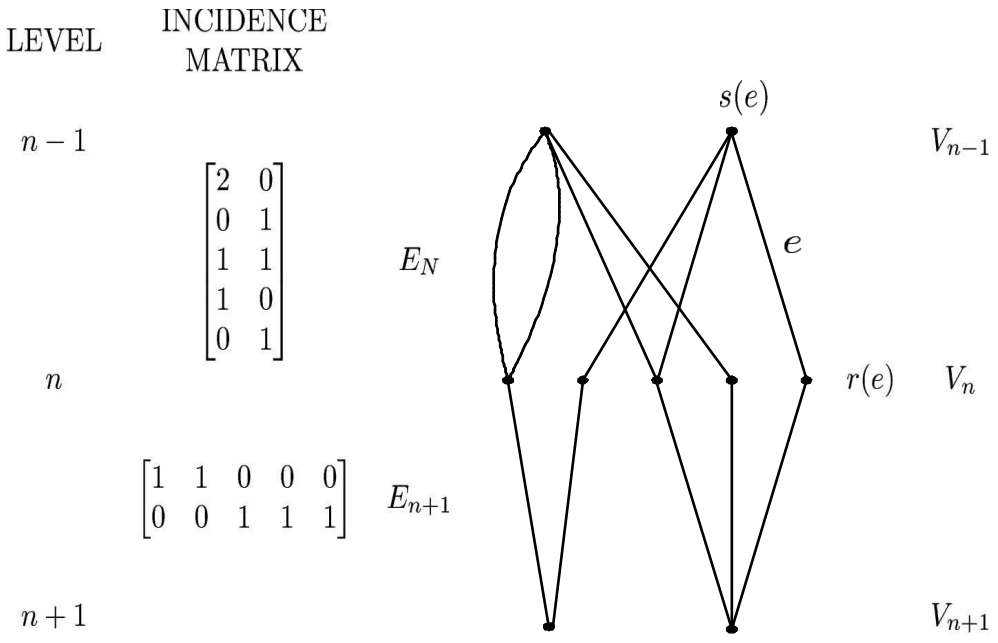

Figure 1"

Observe that clearly

Proposition 1. *Every rooted plane tree (of the type studied in* [13]*) is a Bratteli diagram.*

Thus, indeed, the study of Bratteli diagrams is a natural continuation and extension of the papers [12] and [13].

On the other hand, the reverse of this statement is not true; the Bratteli diagrams are much more general than the rooted plane trees. The most significant differences are that, unlike in rooted plane trees, it may occur in Bratteli diagrams that two vertices are joined by two different edges:

(1.1) there are $e, e' \in E$, $e \neq e'$ with $s(e) = s(e')$, $r(e) = r(e')$, and two different vertices have the same "child":

(1.2) there are $e, e' \in E$, $e \neq e'$ with $s(e) \neq s(e')$, $r(e) = r(e')$.

A survey of the papers written on Bratteli diagrams is presented in [7]. However, in order to adjust the definition of Bratteli diagrams to our goals we will modify it slightly. First, note that finite Bratteli diagrams also occur in the applications (the first papers of this type are [20] and [9]), besides the study of pseudorandomness of *infinite* Bratteli diagrams can be reduced to the study of *finite* ones by truncating them; thus we will consider here only *finite* Bratteli diagrams. Next, the situation described in (1.1) might make the study of pseudorandomness quite complicated, besides in many applications this case cannot occur, thus we will exclude this situation in the definition. So that throughout the rest of this paper we will use the following modified definition:

Definition 1'. A *Bratteli diagram* is a *finite* directed graph (V, E) such that the vertex set V and the edge set E can be partitioned into (finitely many) finite sets

$$V = V_0 \cup V_1 \cup \ldots \cup V_h \quad \text{and} \quad E = E_1 \cup E_2 \cup \ldots \cup E_h \tag{1.3}$$

so that both properties i) and ii) in Definition 1 hold, moreover,

iii) if $e, e' \in E$ and they join the same vertices, then we have $e = e'$.

The binary functions on Bratteli diagrams can be defined by extending the definition of binary functions on trees presented in [12] and [13]:

Definition 2. If $\mathcal{B} = (V, E)$ is a Bratteli diagram, then a function f of the type $f : V \to \{-1, +1\}$ is called a *binary function on* $\mathcal{B}$.

First in Section 2 we will present further notations and definitions related to Bratteli diagrams. Next in Section 3 we will introduce the measures of pseudorandomness of binary functions on Bratteli diagrams. Then in Section 4 we will present the construction of a binary function with strong pseudorandom properties on *"smooth"* Bratteli diagrams. In Section 5 we will study the connection between the different types of measures introduced in Section 3. In Section 6 we will study the measures of pseudorandomness of binary functions on *general* Bratteli diagrams. Section 7 will contain a few further remarks.

2 Notations and definitions related to Bratteli diagrams

We will use the words vertex (=node), root, successor (=child), leaf, path, distance, height in the same sense as we did for trees in [12] and [13] (see also [6], [19] for the same terminology). If the vertex P' is a successor (or child) of the vertex P, then P is said to be a *parent* of P'. The number of successors of the vertex P is called the *outdegree* of P and it is denoted by $d^+(P)$, while the number of parents of P is called the *indegree* of P and it is denoted by $d^-(P)$. If $\mathcal{B} = (V, E)$ is a Bratteli diagram where V is of form (1.3), then we say that the vertices $P \in V_n$ are at *level* n, and we also say that these vertices form the $n+1$-st *row* of $\mathcal{B}$. Now consider again this Bratteli diagram $\mathcal{B} = (V, E)$ with V, E of form (1.3), and for any integers m, n with $0 \leq m \leq n \leq h$, let $\mathcal{B}_{m,n}$ denote the diagram formed by the vertices belonging to $V_m, V_{m+1}, \ldots, V_n$, and the edges belonging to $E_{m+1}, E_{m+2}, \ldots, E_n$ (in other words, we keep the vertices in the $m+1$-st, $m+2$-nd, $\ldots$, $n+1$-st row and the edges running between these vertices). Such a diagram $\mathcal{B}_{m,n}$ will be called a *subdiagram* of $\mathcal{B}$. (Note that in general $\mathcal{B}_{m,n}$ is *not* a Bratteli diagram since for $m > 0$ it is not rooted.)

We will also use the following notations:

The number of the vertices of the Bratteli diagram $\mathcal{B} = (V, E)$ will be denoted by $N = N(\mathcal{B})$: $N = N(\mathcal{B}) = |V|$. The height of the Bratteli diagram $\mathcal{B}$ will be denoted by $h = h(\mathcal{B})$. We will denote the number of vertices in the i-th row V_{i-1} (i.e., at level $i-1$) by $y_i = y_i(\mathcal{B})$ ($= t_{i-1}$ according to the notation quoted in Definition 1), and we will denote these vertices (moving from left to right) by $P_{\mathcal{B}}(i, 1)$, $P_{\mathcal{B}}(i, 2), \ldots, P_{\mathcal{B}}(i, y_i)$; if $\mathcal{B}$ is fixed, then we will drop the subscript $\mathcal{B}$. Clearly we have

$$N = N(\mathcal{B}) = y_1 + y_2 + \cdots + y_{h+1}.$$

We will also use the following alternative notation for the vertices. The root is denoted by R_1: $R_1 = P(1,1)$, the vertices in the second row V_1 by $R_2, R_3, \ldots, R_{y_2+1}$: $R_2 = P(2,1)$, $R_3 = P(2,2), \ldots, R_{y_2+1} = P(2,y_2)$, the vertices in the third row V_2 by $R_{y_2+2}, R_{y_2+3}, \ldots, R_{y_2+y_3+1} : R_{y_2+2} = P(3,1)$, $R_{y_2+3} = P(3,2), \ldots, R_{y_2+y_3+1} = P(3,y_3)$ and so on, finally R_N denotes the last vertex in the last row $V_h : R_N = P(h+1, y_{h+1})$.

To the binary function $f : V \to \{-1,+1\}$ defined on the Bratteli diagram $\mathcal{B}$ we will assign the *binary sequence*

$$E_N = E_N(f,\mathcal{B}) = (e_1, e_2, \ldots, e_N) \in \{-1,+1\}^N \tag{2.1}$$

defined by

$$e_n = f(R_n) \quad \text{for} \quad n = 1,2,\ldots,N. \tag{2.2}$$

Consider a path with endpoints R_i, R_j with $i < j$ (so that Q_i is the endpoint closer to the root). This path will be denoted by $\mathcal{P}(R_i, R_j)$. A *path* $\mathcal{P}(Q_i, Q_j)$ consists of edges joining vertices $R_i = P(x,k_x)$, $P(x+1,k_{x+1})$, $P(x+2,k_{x+2}), \ldots, R_j = P(y,k_y)$ taken from consecutive rows, the *distance* between these vertices $R_i = P(x,k_x)$ and $R_j = P(y,k_y)$ is $y-x$, and the *height* of a subdiagram with first row at level x and last row at level y is $y-x$.

If we want to introduce measures of pseudorandomness for binary functions on Bratteli diagrams, then clearly we need some restrictions on the structure of the diagram since it seems to be hopeless to give a definition which can be used for any Bratteli diagram. Thus we will define certain special types of Bratteli diagrams which occur most frequently in the applications, and here we will focus on Bratteli diagrams of this type.

Definition 3. If $K \in \mathbb{N}$, $\mathcal{B} = (V,E)$ is a Bratteli diagram with V, E of form (1.3), and we have

$$|V_i| \le K \quad \text{for all} \quad i \in \{0,1,\ldots,h\},$$

then we say that $\mathcal{B}$ is *K-bounded*.

Definition 4. If $k \in \mathbb{N}$, $\mathcal{B} = (V,E)$ is a Bratteli diagram with V, E of form (1.3), and we have

$$|V_i| = K \quad \text{for all} \quad i \in \{1,2,\ldots,h\},$$

then we say that $\mathcal{B}$ is *K-uniform*.

The *stationary* Bratteli diagrams form an important special class of the K-uniform Bratteli diagrams. We adapt the definition presented in [8] to our notation:

Definition 5. "A Bratteli diagram $\mathcal{B} = (V, E)$ (with V, E of form (1.3)) is *stationary* if $K = |V_1| = |V_2| = \cdots = |V_h|$ and if (by an appropriate labelling of the vertices) the incidence matrix between levels n and $n+1$ is the same $K \times K$ matrix C for all $n = 1, 2, \ldots, h-1$. In other words, beyond level 1 the diagram repeats. (Clearly we may label the vertices $P(n+1, 1)$, $P(n+1, 2), \ldots, P(n+1, K)$ in V_n as $V(n, a_1), \ldots, V(n, a_K)$, where $A = \{a_1, \ldots, a_K\}$ is a set of K distinct symbols.)"

(Roughly, a K-uniform Bratteli diagram is stationary if by changing the order of the vertices in every row it can be achieved that beyond level 1 the diagram repeats.)

An even more special class of K-uniform Bratteli diagrams is formed by the *K-regular* Bratteli diagrams:

Definition 6. A K-uniform Bratteli diagram is said to be *K-regular* if every element of the incidence matrices between level n and $n+1$ is 1 for all $n = 1, 2, \ldots, h-1$. (In other words, every vertex in the $n+1$-st row is joined with every vertex in the $n+2$-nd row.)

Definition 7. The Bratteli diagram $\mathcal{B} = (V, E)$ (with V, E of form (1.3)) is said to be *(a, q)-periodic* if a, q are positive integers such that $a \equiv h (\mathrm{mod}\ q)$, for $a \leq n \leq h - q$ we have $|V_n|$ $(= y_{n+1})$ $= |V_{n+q}|$ $(= y_{n+q+1})$, and for $a \leq n < h-q$, $1 \leq u \leq y_{n+1}$, $1 \leq v \leq y_{n+2}$ the vertices $P(n+1, u), P(n+2, v)$ are joined if and only if the vertices $P(n+q+1, u)$, $P(n+q+2, v)$ are joined (so that the incidence matrix between levels n and $n+1$ is the same as the incidence matrix between levels $n+q$ and $n+q+1$, moreover, for $m, n \in \mathbb{N}$, $X \in \mathbb{N} \cup \{0\}$, $a \leq m < n \leq h - x$, $m \equiv n$ (mod q) the subdiagrams $B_{m,m+x}$ and $B_{n,n+x}$ are isomorphic). A $(1, q)$-periodic Bratteli diagram is said to be *purely periodic* with period q.

Note that a K-regular Bratteli diagram is purely periodic with period 1.

If a Bratteli diagram $\mathcal{B}$ is (x, y)-periodic for some pair x, y, then there is a unique pair (a, q) such that $\mathcal{B}$ is (a, q)-periodic but it is *not* (a', q')-periodic for any pair a', q' with $q' < q$, or $q = q'$ and $a' < a$; we will say that $\mathcal{B}$ is a *primitive (a, q)-periodic* Bratteli diagram. From now on we will restrict ourselves to periodic Bratteli diagrams which are *primitive (a, q)-periodic*; clearly this can be done without the loss of generality.

Throughout the rest of this paper we will consider K-bounded (a, q)-periodic Bratteli diagrams with K, a, q fixed and $h \to +\infty$ (apart from the last two sections).

3 The measures of pseudorandomness for binary functions on "smooth" Bratteli diagrams

If we want to introduce the measures of pseudorandomness for binary functions on Bratteli diagrams, then we may consider the following basic requirements:

R1) For a fixed Bratteli diagram we have to be able to give a good upper bound M for the *maximum* of the value of the measure to be introduced over all binary functions on the given Bratteli diagram.

R2) For a fixed Bratteli diagram we have to be able to show that for *almost all* binary functions defined on the given Bratteli diagram the value of the measure to be introduced is "much smaller" than M defined in R1); in particular, it must be $o(M)$. (If this requirement holds and the value of the measure is small on a certain binary function, then this fact can be considered as a good pseudorandom property of the given binary function.)

R3) We have to be able to show that the value of the measure to be introduced is small in terms of M (defined in R1)) for at least on certain *special* binary functions defined on any given Bratteli diagram. (We will usually apply the *Legendre symbol* $\left(\frac{n}{p}\right)$ to present constructions of this type. Note that we do not define the value of $\left(\frac{n}{p}\right)$ for $p \mid n$.)

R4) The measures to be introduced must be pairwise independent, i.e., for any pair of them either one of them can be large while the other one is small.

R5) Since by Proposition 1 the Bratteli diagrams include the rooted plane trees, thus the measures of pseudorandomness of binary functions defined on *Bratteli diagrams* must be extensions (or at least variants) of the measures defined on rooted plane trees in [13].

Starting out from requirement R5) here we will introduce two groups of pseudorandom measures in the same way as in [13]: horizontal measures and vertical measures. In order to avoid lengthy repetitions, here we will not recall the definitions given in [13] for trees, we will present only their adaptations to the case of Bratteli diagrams. However, these measures to be introduced will function well only under the same assumption that we also had in [13]: we have to assume that every vertex not in the last row has non-zero outdegree.

Definition 8. If every vertex not in the last row has non-zero outdegree (i.e., all the leafs are in the last row), then the Bratteli diagram is called *complete*.

Adding the assumption of completeness to our earlier assumptions on the Bratteli diagrams to consider we may say that we will restrict ourselves to studying Bratteli diagrams of the following type:

Definition 9. If a Bratteli diagram is K-bounded, primitive (a, q)-periodic and complete, then it is said to be a *smooth* (a, q) *Bratteli diagram.*

The family of the smooth (a, q) Bratteli diagrams with K, a, q, h given will be denoted by $\mathbb{B}(K, a, q, h)$. We will study $\mathbb{B}(K, a, q, h)$ for K, a, q fixed and $h \to +\infty$.

First we will define the *horizontal* measures. As in case of trees, we may define the horizontal well-distribution, correlation and normality measures of a binary function f defined on a Bratteli diagram $\mathcal{B}$ considering the corresponding measures of the binary sequence $E_N = E_N(f, \mathcal{B})$ assigned to f and $\mathcal{B}$ by (2.1) and (2.2):

Definition 10. The horizontal well-distribution measure, the correlation measure of order k and normality measure of order k of the binary function f defined on the Bratteli diagram $\mathcal{B}$ are defined as

$$W^H(f, \mathcal{B}) = W\big(E_N(f, \mathcal{B})\big),$$
$$C_k^H(f, \mathcal{B}) = C_k\big(E_N(f, \mathcal{B})\big)$$

and

$$N_k^H(f, \mathcal{B}) = N_k\big(E_N(f, \mathcal{B})\big),$$

respectively.

The case of the *vertical* measures is more difficult. In order to introduce the vertical well-distribution and correlation measures for trees, in [13] we considered *all* the path $\mathcal{P}$ starting from the root and ending in the last ($h+1$-st) row:

$$\mathcal{P}\big(P(1,1), P(2,i_2), \ldots, P(h+1, i_{H+1})\big), \tag{3.1}$$

and to each of these paths we assigned the binary sequence

$$\begin{aligned} G(\mathcal{P}) &= \big(g_1(\mathcal{P}), g_2(\mathcal{P}), \ldots, g_{h+1}(\mathcal{P})\big) \\ &= \big(f(P(1,1)), f(P(2,i_2)), \ldots, f(P(h+1, i_{h+1}))\big). \end{aligned} \tag{3.2}$$

Then we used the well-distribution measures and correlation measures of these sequences to define the corresponding vertical measures. However, here we ran into a serious difficulty: there are "too many" paths $\mathcal{P}$ of the form (3.1). Indeed, the number of these paths is equal to the number of vertices

in the last row which is usually of order of magnitude N (in case of trees). A further difficulty is that the length of the paths of this form is h which is very small, much smaller than N. In order to get around these difficulties, we had to introduce "weak" and "strong" measures, and even so these measures are not quite satisfactory, we have to combine them with other measures.

Here the situation is very much different. Namely, the length of the paths of form (3.1) is much longer: $h+1$ which is $\gg \frac{N}{K}$ for a K-bounded Bratteli diagram but, on the other hand, their number is even greater: e.g., for the especially important K-regular Bratteli diagrams the number of paths of form (3.1) is

$$K^h = K^{(N-1)/K}$$

which, since K is fixed, is exponentially large in terms of N, and thus one would need to control the measures of so many sequences which would be a hopeless task. Thus we have to restrict the number of sequences considered significantly. In case of smooth Bratteli diagrams the most natural way of doing this is to switch to the study of *k-symbol sequences* as described below.

Let $\mathcal{B}$ be a smooth Bratteli diagram with $\mathcal{B} = (V, E) \in \mathbb{B}(K, a, q, h)$ and V, E of form (1.3). Let a_0 denote the smallest integer with

$$a_0 > a \quad \text{and} \quad a_0 - 1 \equiv h \ (q); \tag{3.3}$$

then we have

$$a < a_0 \leq a + q. \tag{3.4}$$

Write

$$M = \frac{h - a_0 + 1}{q} \tag{3.5}$$

(note that $M \in \mathbb{N}$ by (3.3), (3.4) and $h \to \infty$) so that for fixed K, a, q and $h \to +\infty$ we have

$$M = \frac{h}{q} + O(1) > \frac{N}{Kq} + O(1). \tag{3.6}$$

Consider the subdiagrams

$$\mathcal{B}^{(i)} = \mathcal{B}_{a_0+(i-1)q, a_0+iq-1} \quad \text{for} \ \ i = 1, 2, \ldots, M$$

(by (3.4) and (3.5) here we have $0 < a_0+(i-1)q \leq a_0+iq-1 \leq a_0+Mq-1 \leq h$). Then

$$\mathcal{B}^{(i)} \cong \mathcal{B}^{(i')} \quad \text{for} \ \ 1 \leq i, i' \leq M.$$

It follows that the total number of vertices belonging to $\mathcal{B}^{(i)}$ is the same for every i; denote this number of vertices by Y. For every $i \in \{1, 2, \ldots, M\}$, let R_{X_i} be the first vertex in the first row of $\mathcal{B}^{(i)}$ so that the last vertex in the last row is R_{X_i+Y-1}. Assign the Y-tuple

$$\mathcal{E}_i\big(f(R_{X_i}), f(R_{X_i+1}), \ldots, f(R_{X_i+Y-1})\big) \in \{-1, +1\}^Y \tag{3.7}$$

to each $i \in \{1, 2, \ldots, M\}$, and define the sequence E'_M by

$$E'_M = (\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_M). \tag{3.8}$$

Write $\mathcal{A} = \{-1, +1\}^Y$ and $k = |\mathcal{A}| = 2^Y$. Then clearly we have

$$E'_M \in \mathcal{A}^M,$$

and the elements of the sequence E'_M belong to the "k-letter alphabet" $\mathcal{A}$. Thus we may reduce the study of the pseudorandomness of the function $f : \mathcal{B} \to \{-1, +1\}$ to the study of the "k-symbol sequence" E'_M:

Definition 11. Using the notations above, we define the *vertical frequency measures* of the binary function f defined on the Bratteli diagram $\mathcal{B} \subset \mathbb{B}(K, a, q, h)$ in the following way: the *vertical F-well-distribution* ("F" for "frequency", "V" for vertical) *measure* of f is defined as

$$W^{FV}(f, \mathcal{B}) = \delta(E'_M),$$

while the *vertical F-correlation measure of order ℓ* of f is defined as

$$C_\ell^{FV}(f, \mathcal{B}) = \gamma_\ell(E'_M).$$

(We remark that the values of f assumed at the vertices belonging to the first a rows are not included in these definitions. These values can be omitted since the number of these vertices is bounded.)

It remains to show that the horizontal and vertical measures introduced by us satisfy requirements R1)–R5) formulated above. This is trivial in case of requirement R1): clearly, the value of each of the measures introduced is at most N (=the number of vertices of the Bratteli diagram). It also follows easily from the analogous results on binary sequences [5], [2], [15] and k symbol sequences [3] that for any fixed smooth (a, q) Bratteli diagram and for a random binary function defined on it the measures above are small. It is more difficult to show that the requirements R3) and R4) also hold; this we will show in the next two sections. Finally, we have seen that the *horizontal* measures are the extensions of the analogous measures introduced for trees in [13] but this can not be achieved in case of the *vertical* measures. However, in the last Section 6 we will also introduce further vertical measures which are closer to the ones used for trees, besides they also can be used in the aperiodic case but, on the other hand, they have some significant disadvantages.

4 A "good" construction on smooth Bratteli diagrams

We will show that by using the Legendre symbol one can define a binary function with strong pseudorandom properties on any *smooth* Bratteli diagram. We will use the same notation as in Section 3.

Theorem 1. *If $\mathcal{B} = (V, E) \subset \mathbb{B}(K, a, q, h)$, p is the smallest prime with $p > N$ $(= |V|)$ and the binary function $f : \mathcal{B} \to \{-1, +1\}$ is defined by*

$$f(R_n) = \left(\frac{n}{p}\right) \quad \text{for } n = 1, 2, \ldots, N, \tag{4.1}$$

and $\ell \in \mathbb{N}$, $\ell < N$, then we have

$$W^H(f, \mathcal{B}) < 26N^{1/2} \log N, \tag{4.2}$$

$$C_\ell^H(f, \mathcal{B}) < 26\ell N^{1/2} \log N, \tag{4.3}$$

$$N_\ell^H(f, \mathcal{B}) < 26\ell N^{1/2} \log N, \tag{4.4}$$

$$W^{FV}(f, \mathcal{B}) < 26KqN^{1/2} \log N \tag{4.5}$$

and

$$C_\ell^{FV}(f, \mathcal{B}) < 26\ell KqN^{1/2} \log N. \tag{4.6}$$

Proof. We apply Definition 10 with the sequence

$$E_N(f, \mathcal{B}) = \left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \ldots, \left(\frac{N}{p}\right)\right),$$

and we use $N < p \leq 2N$, and the upper bound

$$Q_k(E_{p-1}) = Q_k\left(\left(\frac{1}{p}\right), \left(\frac{2}{p}\right), \ldots \left(\frac{p-1}{p}\right)\right) < 9p^{1/2} \log p$$

given in Theorem 1 in [16]. Then we get

$$W^H(f, \mathcal{B}) = W(E_N(f, \mathcal{B})) = Q_1(E_N) \leq Q_1(E_{p-1}) < 9p^{1/2} \log p < 26N^{1/2} \log N$$

and

$$\begin{aligned} C_\ell^H(f, \mathcal{B}) = C_\ell\big(E_N(f, \mathcal{B})\big) &\leq Q_\ell(E_N) \\ &\leq Q_\ell(E_{p-1}) < 9\ell p^{1/2} \log p < 26\ell N^{1/2} \log N \end{aligned} \tag{4.7}$$

which proves (4.2) and (4.3), while (4.4) follows from (4.7) and the inequality

$$N_\ell(E_N) \le \max_{1\le t\le \ell} C_\ell(E_N)$$

presented as Proposition 1 in [16].

Next we will prove (4.5). By the definitions of $x_i, Y, \mathcal{E}_i$ in (3.7) and f in (4.1) we have

$$\mathcal{E}_i = \left(\left(\frac{x_i}{p}\right), \left(\frac{x_i+1}{p}\right), \ldots, \left(\frac{x_i+Y-1}{p}\right)\right) \tag{4.8}$$

with

$$x_i = x_1 + (i-1)Y \quad \text{for } i = 1, 2, \ldots, M. \tag{4.9}$$

In order to prove the upper bound (4.5) for $W^{FV}(f, \mathcal{B}) = \delta(E'_M)$ (see Definition 11) we have to estimate $\sigma(E'_M, a, Z, u, v)$ for the sequence E'_M in (3.8) with the $\mathcal{E}'_i$s defined by (4.8), and for $a = (r_0, r_1, \ldots, r_{Y-1}) \in \{-1, +1\}^Y$ and positive integers Z, u, v such that

$$u + (Z-1)v \le M. \tag{4.10}$$

If

$$1 \le i \le M, \tag{4.11}$$

then clearly we have

$$\begin{aligned} &\frac{1}{2^Y}\prod_{t=0}^{Y-1}\left(r_t\left(\frac{x_i+t}{p}\right)+1\right) \\ &= \begin{cases} 1 & \text{if } \mathcal{E}_i = \left(\left(\frac{x_i}{p}\right), \ldots, \left(\frac{x_i+Y-1}{p}\right)\right) = (r_0, \ldots, r_{Y-1}) = a, \\ 0 & \text{if } \mathcal{E}_i \ne a; \end{cases} \end{aligned} \tag{4.12}$$

note that it follows from (4.11) and $0 \le t < Y$ that

$$0 < x_i + t \le N < p. \tag{4.13}$$

Thus for every $a = (r_0, r_1, \ldots, r_t)$ and u, v, Z satisfying (4.10) we have

$$\begin{aligned} \sigma(E'_M, a, Z, u, v) &= \left|\{j : 0 \le j \le Z-1, \mathcal{E}_{u+jv} = a\}\right| \\ &= \sum_{j=0}^{Z-1} \frac{1}{2^Y}\prod_{t=0}^{Y-1}\left(r_t\left(\frac{x_{u+jv}+t}{p}\right)+1\right) \end{aligned}$$

$$= \frac{1}{2^Y} \sum_{j=0}^{Z-1} \prod_{t=0}^{Y-1} \left(r_t \left(\frac{x_{u+jv}+t}{p} \right) + 1 \right)$$
$$= \frac{1}{2^Y} \sum_{j=0}^{Z-1} \left(1 + \sum_{\mathcal{T} \subset \{0,1,\dots,Y-1\}} \prod_{t \in \mathcal{T}} r_t \left(\frac{x_{u+jv}+t}{p} \right) \right)$$
$$= \frac{Z}{2^Y} + \frac{1}{2^Y} \sum_{\mathcal{T} \subset \{0,1,\dots,Y-1\}} \left(\prod_{t \in \mathcal{T}} r_t \right) \sum_{j=0}^{Z-1} \left(\frac{\prod_{t \in \mathcal{T}} (x_{u+jv}+t)}{p} \right)$$

whence

$$\left| \sigma(E'_M, a, Z, u, v) - \frac{Z}{2^Y} \right| \le \frac{1}{2^Y} \sum_{\mathcal{T} \subset \{0,1,\dots,Y-1\}} \left| \sum_{j=0}^{Z-1} \left(\frac{\prod_{t \in \mathcal{T}} (x_{u+jv}+t)}{p} \right) \right|. \tag{4.14}$$

By (4.9) the absolute value of the last sum can be rewritten in the following way:

(4.15)
$$\left| \sum_{j=0}^{Z-1} \left(\frac{\prod_{t \in \mathcal{T}} (x_{u+jv}+t)}{p} \right) \right| = \left| \sum_{j=0}^{Z-1} \left(\frac{\prod_{t \in \mathcal{T}} \big((x_1 + (u+jv-1)Y) + t \big)}{p} \right) \right|$$
$$= \left| \sum_{j=0}^{Z-1} \left(\frac{\prod_{t \in \mathcal{T}} \big((x_1 + (u-1)Y + t) + jvY \big)}{p} \right) \right|.$$

If

$$Z > 1, \tag{4.16}$$

then it follows from (4.10), $u > 0$ and $M \le N < p$ that $(0 <)\, v < M < p$, thus

$$(v, p) = 1. \tag{4.17}$$

Moreover, for any $a_0 \le n < h + 1 - q$ clearly we have

$$0 < Y = |V_n| + |V_{n+1}| + \cdots + |V_{n+q-1}| \le Kq = O(1) < h \le N < p \tag{4.18}$$

(for h large enough) so that we also have

$$(Y, p) = 1. \tag{4.19}$$

By (4.17) and (4.19) the modulo p multiplicative inverse of vY exists, thus it follows from (4.15) that

$$\begin{aligned}(4.20)\qquad &\left|\sum_{j=0}^{Z-1}\left(\frac{\prod_{t\in\mathcal{T}}(x_{u+jv}+t)}{p}\right)\right| \\ &=\left|\sum_{j=0}^{Z-1}\left(\frac{\prod_{t\in\mathcal{T}}\big(x_1(vY)^{-1}+(u-1)v^{-1}+t(vY)^{-1}+j\big)}{p}\right)\right| \\ &=\left|\sum_{j=0}^{Z-1}\left(\frac{g_{\mathcal{T}}(j)}{p}\right)\right|\end{aligned}$$

where $g_{\mathcal{T}}(x)\in\mathbb{F}_p[x]$ is a polynomial of degree

$$(4.21)\qquad 1\le \deg\, g_{\mathcal{T}}(x)=|\mathcal{T}|\le Y$$

such that it has no multiple zero (since the elements of $\mathcal{T}\subset\{0,1,\dots,Y-1\}$ are incongruent modulo p). Moreover, by (4.10) we have

$$(4.22)\qquad 0<x_1\le x_{u+jv}+t\le x_{u+(Z-1)v}+Y-1\le x_M+Y-1\le N<p,$$

thus

$$(4.23)\qquad g_{\mathcal{T}}(j)\ne 0\quad\text{for}\quad 0\le j\le Z-1.$$

Now we need the following lemma:

Lemma 1. *If p is a prime, $g(x)\in\mathbb{F}_p[x]$ is of degree k and such that it has no multiple zero, and α,β are real numbers with $0<\alpha<\beta<p$ and $g(n)\ne 0$ for $\alpha\le n\le\beta$, then we have*

$$\left|\sum_{\alpha\le n\le\beta}\left(\frac{g(n)}{p}\right)\right|<9kp^{1/2}\log p.$$

Proof. This is a special case of Corollary 1 in [16] (which was derived there from Weil's theorem [21]). □

By using this lemma, we get from (4.18), (4.20), (4.21) and (4.23) that for Z satisfying (4.16) we have

$$(4.24)\qquad \left|\sum_{j=0}^{Z-1}\left(\frac{\prod_{t\in\mathcal{T}}(x_{u+jv}+t)}{p}\right)\right|=\left|\sum_{j=0}^{Z-1}\left(\frac{g_{\mathcal{T}}(j)}{p}\right)\right|$$

$$< 9Yp^{1/2}\log p \le 9Kqp^{1/2}\log p < 26KqN^{1/2}\log N,$$

and this inequality also holds trivially when $Z = 1$ since then the sum to be estimated consists of a single term of absolute value 1.

By (4.14) and (4.24) we have

$$\left|\sigma(E'_M, a, Z, u, v) - \frac{Z}{2^Y}\right| < \frac{1}{2^Y}\sum_{\mathcal{T}\subset\{0,1,\dots,Y-1\}} 26KqN^{1/2}\log N$$
$$= \frac{26}{2^Y}KqN^{1/2}\log N \sum_{\mathcal{T}\subset\{0,1,\dots,Y-1\}} 1 < 26KqN^{1/2}\log N$$

for every $a \in \{-1,+1\}^Y$ and u, Z, v satisfying (4.10), which proves

$$\delta(E'_M) < 26KqN^{1/2}\log N.$$

By Definition 11, (4.5) follows from this.

Finally, in order to prove (4.6) we have to estimate $\tau(E'_M, W, Z, \mathbf{D})$ (see the definitions of τ and γ_ℓ in Section 1 and Definition 11) for the sequence E'_M in (3.8) with the $\mathcal{E}'_i$s defined by (4.8), and for $w = (a_{i_1}, a_{i_2}, \dots, a_{i_\ell}) \in \mathcal{A}^\ell$ with $a_{ij} = \left(r_0^{(j)}, r_1^{(j)}, \dots, r_{Y-1}^{(j)}\right) \in \{-1,+1\}^Y$ (for $j = 1, 2, \dots, \ell$), $Z \in \mathbb{N}$ and $\mathbf{D} = (d_1, d_2, \dots, d_\ell)$ with non-negative integers $d_1 < d_2 < \cdots < d_\ell$ such that

$$Z + d_\ell \le M. \tag{4.25}$$

Again we use (4.12) to compute $\tau(E'_M, w, Z, \mathbf{D})$:

$$\tau(E'_M, w, Z, \mathbf{D}) = \left|\left\{n :\ 1 \le n \le Z, (\mathcal{E}_{n+d_1}, \dots, \mathcal{E}_{n+d_\ell}) = (a_{i_1}, \dots, a_{i_\ell})\right\}\right|$$
$$= \sum_{n=1}^{Z}\prod_{j=1}^{\ell}\frac{1}{2^Y}\prod_{t=0}^{Y-1}\left(r_t^{(j)}\left(\frac{x_{n+d_j}+t}{p}\right)+1\right)$$
$$= \frac{1}{2^{\ell Y}}\sum_{n=1}^{Z}\prod_{j=1}^{\ell}\prod_{t=0}^{Y-1}\left(r_t^{(j)}\left(\frac{x_{n+d_j}+t}{p}\right)+1\right)$$

so that writing $\mathcal{S} = \left\{(j,t) : j \in \{1,\dots,\ell\},\ t \in \{0,\dots,Y-1\}\right\}$ we have

$$\tau(E'_M, w, Z, \mathbf{D}) = \frac{1}{2^{\ell Y}}\sum_{n=1}^{Z}\left(1 + \sum_{\mathcal{S}'\subset\mathcal{S}}\prod_{(j,t)\in\mathcal{S}'} r_t^{(j)}\left(\frac{x_{n+dj}+t}{p}\right)\right)$$
$$= \frac{Z}{2^{\ell Y}} + \frac{1}{2^{\ell Y}}\sum_{n=1}^{Z}\sum_{\mathcal{S}'\subset\mathcal{S}}\left(\prod_{(j,t)\in\mathcal{S}'} r_t^{(j)}\right)\left(\frac{\prod_{(j,t)\in\mathcal{S}'}(x_{n+d_j}+t)}{p}\right)$$

$$= \frac{Z}{2^{\ell Y}} + \frac{1}{2^{\ell Y}} \sum_{\mathcal{S}' \subset \mathcal{S}} \Big(\prod_{(j,t)\in\mathcal{S}'} r_t^{(j)} \Big) \sum_{n=1}^{Z} \left(\frac{\prod\limits_{(j,t)\in\mathcal{S}'} (x_{n+d_j} + t)}{p} \right)$$

whence, using (4.9)

(4.26)
$$\left| \tau(E'_M, w, Z, \mathbf{D}) - \frac{Z}{2^{\ell Y}} \right| \le \frac{1}{2^{\ell Y}} \sum_{\mathcal{S}' \subset \mathcal{S}} \left| \sum_{n=1}^{Z} \left(\frac{\prod\limits_{(j,t)\in\mathcal{S}'} (x_{n+d_j} + t)}{p} \right) \right|$$
$$= \frac{1}{2^{\ell Y}} \sum_{\mathcal{S}' \subset \mathcal{S}} \left| \sum_{n=1}^{Z} \left(\frac{\prod\limits_{(j,t)\in\mathcal{S}'} (x_1 + (n + d_j - 1)Y + t)}{p} \right) \right|.$$

Now we will prove that the factors in the last product are incongruent modulo p:

(4.27)
$$x_1 + (n + d_j - 1)Y + t \not\equiv x_1 + (n + d_{j'} - 1)Y + t' \pmod{p}$$
$$\text{for } (t, j), (t', j') \in \mathcal{S}',\ (t, j) \ne (t', j').$$

It suffices to show that

(4.28)
$$d_j Y + t \not\equiv d_{j'} Y + t' \quad \text{for } (t, j), (t', j') \in \mathcal{S}',\ (t, j) \ne (t', j').$$

If $j = j'$, then this is trivial since $0 \le t,\ t' < Y < p$ and $t \ne t'$. If, say, $j < j'$, then by (4.25) we have

$$0 < Y - t \le (d_{j'} - d_j)Y + t' - t = (d_{j'} Y + t') - (d_j Y + t)$$
$$\le d_{j'} Y + t' \le d_\ell Y + Y - 1 \le (M - 1)Y + Y - 1 < MY \le N < p$$

whence (4.28) follows, which proves (4.27). By (4.19) the modulo p multiplicative inverse of Y exists, thus in the last sum in (4.26) we have

(4.29)
$$\left| \sum_{n=1}^{Z} \left(\frac{\prod\limits_{(j,t)\in\mathcal{S}'} (x_1 + (n + d_j - 1)Y + t)}{p} \right) \right|$$
$$= \left| \sum_{n=1}^{Z} \left(\frac{\prod\limits_{(j,t)\in\mathcal{S}'} \big(((x_1 + t)Y^{-1} + d_j - 1) + n\big)}{p} \right) \right|,$$

and denoting the polynomial in the numerator by $G_{\mathcal{S}'}(n)$:

$$G_{\mathcal{S}'}(n) = \prod_{(j,t)\in\mathcal{S}'} \big(((x_1+t)Y^{-1}+d_j-1)+n\big), \tag{4.30}$$

the factors in this product are also incongruent modulo p, thus it has no multiple zero. Moreover, arguing as in (4.22), by using (4.25) one can show that the factors of the product in (4.26) are nonzero, so that

$$G_{\mathcal{S}'}(n) \neq 0 \quad \text{for} \quad 1 \le n \le Z. \tag{4.31}$$

Thus the polynomial $G_{\mathcal{S}'}(n)$ of degree

$$1 \le \deg G_{\mathcal{S}'}(n) = |\mathcal{S}'| \le |\mathcal{S}| \le \ell Y \tag{4.32}$$

satisfies the assumptions in Lemma 1 with $G_{\mathcal{S}'}$, 1 and Z in place of g, α and β respectively. By using Lemma 1 we get from (4.18), (4.26), (4.29), (4.30), (4.31) and (4.32) that

$$\begin{aligned}
&\left|\tau(E'_M, w, Z, \mathbf{D}) - \frac{Z}{2^{\ell V}}\right| \\
&\le \frac{1}{2^{\ell Y}} \sum_{\mathcal{S}'\subset\mathcal{S}} \left|\sum_{n=1}^{Z} \left(\frac{G_{\mathcal{S}'}(n)}{p}\right)\right| \le \frac{1}{2^{\ell Y}} \sum_{\mathcal{S}'\subset\mathcal{S}} 9\ell Y p^{1/2}\log p \\
&= \frac{9\ell Y}{2^{\ell Y}} p^{1/2}\log p \sum_{\mathcal{S}'\subset\mathcal{S}} 1 < \frac{9\ell Y}{2^{\ell Y}} p^{1/2}\log p\, 2^{|\mathcal{S}|} \\
&\le 9\ell Y p^{1/2}\log p < 26\ell K q N^{1/2}\log N
\end{aligned}$$

for every $w, Z, \mathbf{D}$ satisfying (4.25) which proves

$$\gamma_\ell(E'_M) < 26\ell K q N^{1/2}\log N.$$

By Definition 11, (4.6) follows from this which completes the proof of the theorem. □

We remark that the *vertical* E normality measure could be defined (and also handled for the construction presented in Theorem 1) analogously to the *horizontal* normality measure, we leave the details to the reader.

5 The connection between the horizontal measures and the vertical frequency measures

Ideally, one might like to show that requirement R4) (presented in Section 4) holds, in other words:

R4a) There exists a smooth Bratteli diagram and a binary function defined on it such that its *horizontal* measures are small but one of the corresponding *vertical frequency* measures is large;

R4b) There exists a smooth Bratteli diagram and a binary function defined on it such that its *vertical frequency* measures are small but one of the corresponding *horizontal* measures is large.

However, there are significant difficulties in trying to prove these facts. First, as already the proofs in Section 4 show, it is rather difficult to estimate the vertical frequency measures, and it is even more difficult to find constructions where *both* the horizontal measures and the vertical frequency measures can be handled. Secondly, it could be shown that the horizontal and vertical frequency measures are not quite independent: there is a certain weak connection between them. Due to these difficulties we have achieved only a partial success in studying the requirements above. We can prove a slightly weaker form of requirement R4a):

Theorem 2. *For any fixed $k \in \mathbb{N}$ and infinitely many $N \in \mathbb{N}$ there is a k-regular Bratteli diagram $\mathcal{B} = (V, E)$ with $|V| = N$ and a binary function $f : V \to \{-1, +1\}$ defined on it such that $W^H(f, \mathcal{B})$, $C_2^H(f, \mathcal{B})$, $C_3^H(f, \mathcal{B})$, $\dots$, $C_{2k}^H(f, \mathcal{B})$ are small:*

$$\max\bigl\{W^H(f,\mathcal{B}), C_2^H(f,\mathcal{B}), C_3^H(f,\mathcal{B}), \dots, C_{2k}^H(f,\mathcal{B})\bigr\} \ll N^{1/2}\log N, \tag{5.1}$$

but $W^{FV}(f, \mathcal{B})$ is large (it is "midway" between the optimal upper bound $N^{1/2}(\log N)^c$ (see [2]) and the trivial bound N):

$$W^{FV}(f,\mathcal{B}) \gg \Bigl(N \max_{2\le i\le 2k} C_i^H(f,\mathcal{B})\Bigr)^{1/2}. \tag{5.2}$$

(Here $\gg$ is Vinogradov's notation: we write $\varphi(n) \gg \psi(n)$ if there is a $c > 0$ such that $|\psi(n)| < c\varphi(n)$ for all n, and if c may depend on a parameter k, then we write $\varphi(n) \gg_k \psi(n)$.) Note that it was proved in [1] that

$$\min_{E_N\in\{-1,+1\}^N} C_2(E_N) > \frac{1}{\sqrt{6}} N^{1/2},$$

thus it follows from (5.2) that

$$W^{FV}(f,\mathcal{B}) \gg N^{3/4} \tag{5.3}$$

(while if k is fixed, then for a random k symbol sequence E_M of length $M \sim \frac{N}{k}$ with large probability we have

$$\delta(E_M) \ll_k M^{1/2}(\log M)^{1/2} \ll_k N^{1/2}(\log N)^{1/2}$$

as it was shown in [3].

We remark that we think that (5.2) in Theorem 2 is sharp in the sense that it follows from an assumption of type (5.1) that

$$W^{FV}(f, \mathcal{B}) \ll N^{3/4}(\log N)^c$$

(compare this with (5.3)); certain heuristic arguments support this statement. Moreover, we can show that requirement R4b) fails, and the horizontal measures must be greater than the corresponding vertical frequency measures multiplied by a constant small enough (in terms of ℓ in case of the correlation of order ℓ).

Although we prove only a weak form of requirement R4a), still we think that the use of the vertical frequency measure is justified. Namely, Theorem 2 shows that it may occur that a binary function defined on a smooth Bratteli diagram passes the tests based on the use of the horizontal measures, however, it has certain atypical vertical structure which can be detected by using the vertical frequency measures.

However, because of the difficulties described above, it would need a very lengthy and complicate argument to give a rigorous and detailed proof of Theorem 2 and to provide a detailed study of the facts stated in our two remarks above. Thus here we will restrict ourselves to present only a sketch of the proof of Theorem 2, and we omit the computations.

Sketch of the proof of Theorem 2. We will give a constructive proof based on a combination of ideas used in [14] and [18] (but the principle adapted from [18] could be replaced by some other principles appearing in papers surveyed in [11]).

Let k be a fixed positive integer, p a large prime, and let $\mathcal{B} = (V, E)$ be a k-regular Bratteli diagram with $|V| = N = k\left[\frac{p}{k}\right] + 1$ vertices. Now we will define a binary function $f : V \to \{-1, +1\}$. First we define a binary sequence

$$G_{[p/k]} = (g_1, g_2, \ldots, g_{[p/k]}) \in \{-1, +1\}^{[p/k]}$$

by

$$g_i = \begin{cases} +1 & \text{if } r_p(h(i)) \le (1-c)\frac{p}{2} \\ -1 & \text{if } r_p(h(i)) > (1-c)\frac{p}{2} \end{cases}$$

for $i = 1, 2, \ldots, \left[\frac{p}{k}\right]$ where $h(x) \in \mathbb{F}_p[x]$ is a polynomial of degree, say, $d = 2k+1$ with no multiple zero in $\overline{\mathbb{F}}_p$ (= the algebraic closure of $\mathbb{F}_p$), c is defined as $c = p^{-1/4}(\log p)^{1/2}$ and $r_p(s)$ denotes the integer of smallest absolute value congruent to s modulo p. By using the incomplete version of Weil's theorem

[21], it can be shown by a little computation that

$$W(G_{[p/k]}) = c\frac{p}{k} + O\big(\sqrt{p}\log p\big) = \frac{p^{3/4}}{k}(\log p)^{1/2} + O\big(\sqrt{p}\log p\big), \tag{5.4}$$

$$C_2(G_{[p/k]}) = c^2\frac{p}{k} + O\big(\sqrt{p}\log p\big) = O\big(\sqrt{p}\log p\big) \tag{5.5}$$

and for $3 \le i$,

$$C_i(G_{[p/k]}) = O\big(\sqrt{p}\log p\big). \tag{5.6}$$

Next we define a binary sequence $(e_1, e_2, \dots, e_N)$ by

$$e_i = \begin{cases} +1 & \text{if } r_p(h(i)) < \frac{p}{2} \text{ and } i \not\equiv 1 \pmod{k} \\ -1 & \text{if } r_p(h(i)) > \frac{p}{2} \text{ and } i \not\equiv 1 \pmod{k} \end{cases}$$

and

$$e_i = e_{i-k+1}e_{i-k+2}\dots e_{i-1}g_{(i-1)/k} \quad \text{if } i \equiv 1 \pmod{k}$$

(for $1 \le i \le N$). Then define the binary function $f : V \to \{-1, +1\}$ by

$$f(R_i) = e_i \quad (\text{for } \ i = 1, 2, \dots, N)$$

where $R_1, R_2, \dots, R_N$ are the vertices of $\mathcal{B}$ labelled as described in Section 3. It can be shown by some computation that

$$W^{FV}(f, \mathcal{B}) \gg_k W(G_{[p/k]}) \gg_k p^{3/4}(\log p)^{1/2} = (1+o(1))N^{3/4}(\log N)^{1/2}, \tag{5.7}$$

$$\max_{2\le i\le 2k} C_i^H(f, \mathcal{B}) \ll_k \max\Big\{\max_{2\le i\le 2k} C_i(G_{[p/k]}), \sqrt{p}\log p\Big\},$$

and by (5.5) and (5.6),

$$\max_{2\le i\le 2k} C_i^H(f, \mathcal{B}) \ll_k \sqrt{p}\log p = (1+o(1))N^{1/2}(\log N).$$

Similarly, it can be deduced easily that

$$W^H(f, \mathcal{B}) \ll_k \sqrt{p}\log p = (1+o(1))\sqrt{N}\log N,$$

and this completes the proof of (5.1).

(5.2) follows trivially from (5.1) and (5.7). □

On the other hand, we think that Theorem 2 is sharp and the following inequality always holds for k-smooth Bratteli diagrams:

$$W^{FV}(f, \mathcal{B}) \ll_k \Big(N \max_{2\le i\le 2k} C_i^H(f, \mathcal{B})\Big)^{1/2}.$$

6 The measures of pseudorandomness for binary functions on general Bratteli diagrams

So far we have studied the most important special case when the Bratteli diagram considered is "smooth", i.e., it is K-bounded, complete and primitive, (a, q)-periodic. The case when the Bratteli diagram is *aperiodic* occurs less frequently, besides it is more difficult to handle, thus we will discuss this case only briefly and we omit the details. *We will still assume that the Bratteli diagram is K-bounded and complete,* but we will drop the assumption on periodicity.

The horizontal measures introduced in Section 3 for smooth Bratteli diagrams can be defined and used in this general case in exactly the same way. On the other hand, the vertical frequency measures also introduced in Section 3 are defined by using the periodicity of the given diagram strongly, thus we have to replace these measures by other measures of "vertical nature". These new measures will be the vertical analogues of the horizontal measures introduced in Section 3. In order to define them first we have to introduce some new notations.

First we define a reordering $S_1, S_2, \dots, S_N$ of the vertices of the Bratteli diagram $\mathcal{B}$. Let us take the first vertex from each row starting with the first row (consisting of the root) and ending with the last row; these vertices will be $S_1, S_2, \dots, S_{h+1}$ so that (using the notations introduced in Section 2) we have $S_1 = R_1 = P(1,1)$, $S_2 = R_2 = P(2,1), \dots, S_{h+1} + P(h+1,1)$. Now consider every row containing at least 2 vertices; suppose the i_1-st, i_2-nd, ..., i_k-th row (with $i_1 < i_2 < \dots < i_k$) are these rows. Then from each of these rows we take the second vertex, and these vertices will be $S_{h+2}, \dots, S_{h+k+1}$: $S_{h+2} = P(i_1, 2)$, $S_{h+3} = P(i_2, 2)$, ..., $S_{h+k+1} = P(i_k, 2)$. Next we take every row, say, the j_1-st, j_2-nd, ..., j_ℓ-th row (with $j_1 < j_2 < \dots < j_\ell$) which contains at least 3 vertices, and taking the third vertex from each of these rows we get $S_{h+k+2}, S_{h+k+3}, \dots, S_{h+k+\ell+1}$: $S_{h+k+2} = P(j_1, 3), \dots, S_{h+k+\ell+1} = P(j_\ell, 3)$. We continue the labelling of the vertices in this way, finally, S_N will be the last vertex in the last of the rows containing the maximal number of vertices. We present an example for this relabelling of the vertices in Figure 2.

Let $\mathcal{B} = (V, E)$ be a Bratteli diagram and $f : V \to \{-1, +1\}$ a binary function on it. Write

$$E'_N = E'_N(f, \mathcal{B}) = (e'_1, e'_2, \dots, e'_N) = \big(f(S_1), f(S_2), \dots, f(S_N)\big). \tag{6.1}$$

Then the *linear vertical measures* of f can be defined as the corresponding measures of this binary sequence E'_N:

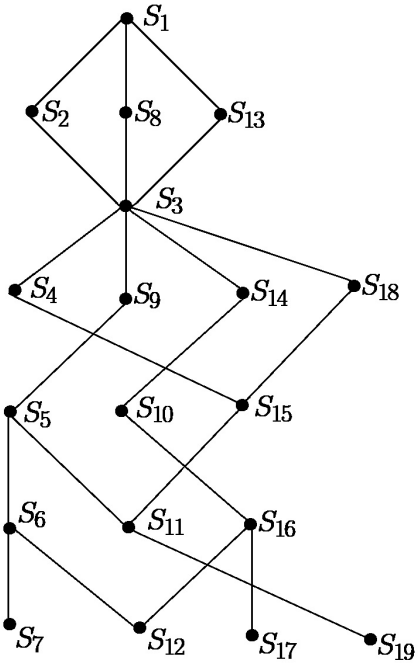

Figure 2

Definition 12. The linear vertical well-distribution measure, correlation measure of order k and normality measure of order k of the binary function f defined on the Bratteli diagram $\mathcal{B}$ are defined as

$$W^{LV}(f,\mathcal{B}) = W\big(E'_N(f,\mathcal{B})\big),$$
$$C_k^{LV}(f,\mathcal{B}) = C_k\big(E'_N(f,\mathcal{B})\big)$$

and

$$N_k^{LV}(f,\mathcal{B}) = N_k\big(E'_N(f,\mathcal{B})\big),$$

respectively.

Clearly, these measures can be used in both the periodic and aperiodic case. Unfortunately, comparing these new measures with the vertical frequency measures they have two great disadvantages: first, for a general aperiodic Bratteli diagram it seems to be very difficult to construct a binary function on it for which both the horizontal measures and these new linear vertical measures are small, and secondly, it could be shown with some work that in the most important periodic case the horizontal measures and the new measures are not quite independent. On the other hand, we can prove that in general the horizontal measures and the new vertical measures are independent in the sense that a binary function can be "good" for the

horizontal measures and "bad" for the horizontal measures and vice versa. Unfortunately, we have not been able to find relatively simple proofs for these results. Our constructions (using again the Legendre symbol) and proofs (using Weil's theorem) are quite complicated and lengthy, thus we do not include them here. It might be an interesting (but not easy) task to look for simpler proofs.

7 Remarks

In order to study pseudorandomness of binary functions defined on *Bratteli diagrams* we have introduced certain measures. In Section 4 we illustrated the applicability of our measures by presenting a construction which is "good" in terms of our measures. For this purpose we used the Legendre symbol construction described in (4.1). This construction is the adaptation of the construction given in [16] for binary functions defined on binary *sequences* with strong pseudorandom properties. In the applications it is usually not enough to have just one "good" function, one may need large *families* of them. In case of Bratteli diagrams the simplest way to construct such a family is to adapt the construction given in [10], and to replace $\left(\frac{n}{p}\right)$ in (4.1) by $\left(\frac{f(n)}{p}\right)$ where $f(x) \in \mathbb{F}_p[x]$ is a polynomial satisfying certain conditions. Since that many further constructions have been given for "good" binary functions defined on binary *sequences* and also for large families of them (see the survey paper [11]); most of these constructions can be adapted to binary functions defined on *Bratteli diagrams*.

Observe that the values of the pseudorandom measures introduced by us depend only on the configuration of the vertices and the values -1 or $+1$ assigned them, but they are independent of the position of the edges connecting them. Thus the values of these measures do not change if we delete or add edges (so that the prescribed properties of the edge set should still hold). One also might like to study the pseudorandomness of the distribution of the values assigned to the endpoints of the edges. However, this problem seems to be even more difficult than the one considered in this paper, thus another paper should be devoted to it.

References

[1] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, Measures of pseudorandom for finite sequences: minimal values, *Probab. Comput.* **15** (2006), no. 1-2, 1–29.

[2] N. Alon, Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, Measures of pseudorandomness for finite sequences: typical values, *Proc. London Math. Soc.* **95** (2007), 778–812.

[3] G. Bérczi, On finite pseudorandom sequences of k symbols, *Period. Math. Hungar.* **47** (2003), 29–44.

[4] O. Bratteli, Inductive limits of finite-dimensional C^*-algebras, *Trans. Amer. Math. Soc.* **171** (1972), 195–234.

[5] J. Cassaigne, C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences VII: The measures of pseudorandomness, *Acta Arith.* **103** (2002), 97–118.

[6] M. Drmota, *Random trees. An Interplay between Combinatorics and Probability*, Springer, Vienna, 2009.

[7] F. Durand, Combinatorics on Bratteli diagrams and dynamical systems, *Combinatorics, Automata and Number Theory, Series Encyclopedia of Mathematics and its Applications* **135**, Cambridge University Press, 2010, 338–386.

[8] F. Durand, B. Host and C. Skau, Substitution dynamical systems, Bratteli diagrams and dimension groups, *Ergodic Theory Dynam. Systems* **19** (1999), 953–993.

[9] F. M. Goodman, P. de la Harpe and V. F. R. Jones, *Coxeter graphs and towers of algebras*, Mathematical Sciences Research Institute Publications, 14, Springer-Verlag, New York, 1989. X+288 pp.

[10] L. Goubin, C. Mauduit and A. Sárközy, Construction of large families of pseudorandom binary sequences, *J. Number Theory* **106** (2004), 56–69.

[11] K. Gyarmati, Measures of pseudorandomness. Finite fields and their applications, 43–64, *Radon Ser. Comput. Appl. Math.* **11**, De Gruyter, Berlin, 2013.

[12] K. Gyarmati, P. Hubert and A. Sárközy, Pseudorandom binary functions on almost uniform trees, *J. Comb. Number Theory* **2** (2010), 1–24.

[13] K. Gyarmati, P. Hubert and A. Sárközy, Pseudorandom binary functions on rooted plane trees, *J. Comb. Number Theory* **4** (2012), 1–19.

[14] P. Hubert and A. Sárközy, On p-pseudorandom binary sequences, *Periodica Math. Hungar.* **49** (2004), 73–91.

[15] Y. Kohayakawa, C. Mauduit, C. G. Moreira and V. Rödl, Measures of pseudorandomness for finite sequences: minimum and typical values, *Proceedings of WORDS'03, 159–169, TUCS Gen. Publ.* 27, Turku Cent. Comput. Sci., Turku, 2003.

[16] C. Mauduit and A. Sárközy, On finite pseudorandom binary sequences, I. Measure of pseudorandomness, the Legendre symbol, *Acta Arith.* **82** (1997), 365–377.

[17] C. Mauduit and A. Sárközy, On finite pseudorandom sequences of k symbols, *Indag. Mathem.* **13**(1), 89–101.

[18] C. Mauduit, J. Rivat and A. Sárközy, Construction of pseudorandom binary sequences using additive characters, *Monatsh. Math.* **141** (2004), 197–208.

[19] J.-P. Serre, *Trees*, Springer Monographs in Math., 2nd ed., Springer, Berlin, 2003.

[20] A. M. Vershik and S. V. Kerov, Locally semisimple algebras. Combinatorial theory and the K_0 functor. (In Russian) *Current problems in mathematics. Newest results*, Vol. 26, 3–56, 260, Itogi Nauki i Tekhniki, Akad. Nauk SSSR, Vsesoyuz. Inst. Nauchn. i. Tekhn. Inform., Moscow, 1985.

[21] A. Weil, Sur les curbes algébriques et les variétés qui s'en déduisent, *Acta Sci. Ind.* **1041**, Hermann, Paris, 1948.