RIGIDITY IN DYNAMICS AND MÖBIUS DISJOINTNESS

ADAM KANIGOWSKI, MARIUSZ LEMAŃCZYK, AND MAKSYM RADZIWIŁŁ

ABSTRACT. Let (X, T) be a topological dynamical system. We show that if each invariant measure of (X, T) gives rise to a measure-theoretic dynamical system that is either:

  a. rigid along a sequence of "bounded prime volume" or

  b. admits a polynomial rate of rigidity on a linearly dense subset in $C(X)$

then (X, T) satisfies Sarnak's conjecture on Möbius disjointness. We show that the same conclusion also holds if there are countably many invariant ergodic measures, and each of them satisfies a. or b. This recovers some earlier results and implies Sarnak's conjecture in the following new cases: for almost every interval exchange map of d intervals with $d \geq 2$, for $C^{2+\epsilon}$-smooth skew products over rotations and $C^{2+\epsilon}$-smooth flows (without fixed points) on the torus. In particular, these are improvements of earlier results of respectively Chaika-Eskin, Wang and Huang-Wang-Ye. We also discuss some purely arithmetic consequences for the Liouville function.

1. INTRODUCTION

Let (X, d) be a compact metric space and $T : X \to X$ a homeomorphism. We call (X, T) a *topological dynamical system*. Sarnak's conjecture [30] asserts that if (X, T) is of zero entropy, then

$$(1) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(T^n x)\mu(n) = 0$$

for all continuous $f : X \to \mathbb{R}$ and all $x \in X$; here μ stands for the Möbius function.

Let $M(X, T)$ denote the set of Borel T-invariant probability measures on X; this is a simplex whose set of extremal points coincides with $M^e(X, T)$, the set of ergodic measures. Any measure ν in $M(X, T)$ gives rise to a measure theoretic dynamical system $(X, \mathcal{B}, \nu, T)$ with $\mathcal{B} = \mathcal{B}(X)$ the Borel σ-algebra of subsets of X. Such a system is called *rigid* if there exists a strictly increasing sequence $\{q_n\}_{n \geq 1}$ such that for every $g \in L^2(X, \nu)$, $g \circ T^{q_n} \to g$ in $L^2(X, \nu)$ as $n \to \infty$. Rigid systems have (measure-theoretic) entropy zero, so by the variational principle, one naturally expects the validity of (1) in the topological systems whose all measures yield rigidity.

Considerable progress in understanding Sarnak's conjecture has been accomplished following the result of Matomäki and the third author [27] which roughly says that the behavior of μ on typical short intervals is the same as its behavior on intervals $[1, N]$. In trying to prove (1) for rigid systems we need a variant of [27] that holds for almost all short arithmetic progressions. Such an extension is obtained in Theorem 3.1 (see also [23] for a very recent related result). Somewhat surprisingly we encounter a significant technical complication

that require us to assume that $\sum_{p|q} 1/p$ is not too large compared to $\sum_{p \leq H} 1/p$. Roughly speaking this means that we need to avoid those q that have a lot of prime factors $\leq H$. We call $\sum_{p|q} 1/p$ the *prime volume* of q. The prime volume grows slowly with q: in Lemma 3.2 and Corollary 3.3, we observe that

$$(2) \qquad \sum_{p|q} \frac{1}{p} \leq \log\log\log q + O(1)$$

but "most" of the time, the prime volume of q stays bounded. Precisely, if we set

$$(3) \qquad \mathcal{D}_j := \Big\{ q \in \mathbb{N} \ : \ \sum_{p|q} \frac{1}{p} < j \Big\};$$

where obviously, $\mathcal{D}_j \subset \mathcal{D}_{j+1}$ for $j \in \mathbb{N}$, then

$$(4) \qquad d(\mathcal{D}_j) \to 1 \text{ when } j \to \infty.$$

Note that the density $d(\mathcal{D}_j)$ of $\mathcal{D}_j$ exists for each j and equals $\mathbb{P}(\sum_p X_p < j)$ with X_p a sequence of Bernoulli independent random variables with $\mathbb{P}(X_p = 1) = 1 - \mathbb{P}(X_p = 0) = 1/p$.

Before we state our main results, we need to define the class of dynamical systems for which our results will be applicable.

Definition 1.1. Let (X, T) be a topological dynamical system. We say that (X, T) is *good* if for every $\nu \in M(X, T)$ at least one of the following conditions holds:

(BPV rigidity): $(X, \mathcal{B}, \nu, T)$ is rigid along a sequence $\{q_n\}_{n \geq 1}$ with **bounded prime volume**, i.e. there exists j such that $\{q_n\}_{n \geq 1} \subset \mathcal{D}_j$;

(PR rigidity): $(X, \mathcal{B}, \nu, T)$ has **polynomial rate** of rigidity: there exists a linearly dense (in $C(X)$) set $\mathcal{F} \subset C(X)$ such that for each $f \in \mathcal{F}$ we can find $\delta > 0$ and a sequence $\{q_n\}_{n \geq 1}$ satisfying

$$(5) \qquad \sum_{j=-q_n^\delta}^{q_n^\delta} \|f \circ T^{jq_n} - f\|_{L^2(\nu)}^2 \to 0.$$

Here is our first result.

Theorem 1.1. *Let (X, T) be a **good** topological dynamical system. Then (X, T) is Möbius disjoint, that is, (1) holds for all continuous $f : X \to \mathbb{R}$ and all $x \in X$.*

To put our second result in context we recall also a recent result of Frantzikinakis-Host [11] according to which Sarnak's conjecture holds in the **logarithmic form** for all zero entropy topological dynamical systems with at most countably many ergodic invariant measures.

Theorem 1.2. *Let (X, T) be a topological dynamical system. Suppose that the set of invariant ergodic measures $M^e(X, T)$ is countable. If for every $\nu \in M^e(X, T)$ the measure theoretic dynamical system $(X, \mathcal{B}, \nu, T)$ satisfies either BPV rigidity or PR rigidity (see Definition 1.1), then (X, T) is Möbius disjoint; that is, (1) holds for all continuous $f : X \to \mathbb{R}$ and all $x \in X$.*

One of the main techniques for establishing Sarnak's conjecture is the Daboussi-Delange-Kátai-Bourgain-Sarnak-Ziegler criterion (colloquially known as DDKBSZ, see [4], [21]) and its measure-theoretic counterpart called AOP [1]. This criterion reduces establishing Möbius disjointness to understanding the *joinings* of T^p and T^q for various sufficiently large primes p and q. For rigid systems the set of joinings is typically very complicated, making it impossible to appeal to the DDKBSZ criterion in this setting. In fact, only two general criteria that go beyond the DDKBSZ criterion are known:

(1) Möbius disjointness holds for dynamical systems whose all invariant measures yield systems with discrete spectrum [15] (see also Cor. 3.20 in [9]).
(2) Möbius disjointness holds for systems with measures of subpolynomial complexity [16].

Note that the dynamical systems in (1) are rigid, since systems with discrete spectrum have a countable group of eigenvalues. In fact all the examples given in [16] for which the second criterion (2) applies are also rigid. Hence, our result can be seen as a new one on the short list of those which are complementary to the DDKBSZ criterion.

In Section 3.2, we give a short outline of the differences between BPV rigidity and PR rigidity, and discuss their common features. Both conditions introduce a kind of stability in our Möbius disjointness results: if in one uniquely ergodic model of a measure-preserving automorphism, BPV or PR rigidity holds then in all remaining uniquely ergodic models of the automorphism we have Möbius disjointness (see Corollary 4.2 and Theorem 4.1). This applies, in particular, to all examples which are considered in Corollaries 1.3, 1.4, 3.7 and 3.8.

Let us now highlight the old and new results that follow as special cases of Theorem 1.1 and Theorem 1.2. In Subsection 3.3, we prove the following corollary:

Corollary 1.3. *For every $d \geq 2$, almost every interval exchange transformation (IET) of d intervals is Möbius disjoint.*

Let us recall that recently Chaika and Eskin [6] proved Corollary 1.3 **for** $d = 3$ (see also [3], [10], [20] for Möbius orthogonality for some subclasses of interval exchange transformations with 3 intervals).

By $\mathbb{T}$ we denote the additive circle which we identify with $[0, 1)$ with addition mod 1, and we consider now *Anzai skew products* on $\mathbb{T}^2$, that is, systems of the form

$$T_\phi(x, y) := (x + \alpha, y + \phi(x)),$$

for irrational $\alpha \in \mathbb{T}$ and $\phi : \mathbb{T} \to \mathbb{T}$ a continuous function on the circle. If ϕ is of zero topological degree, that is, $\phi(t) = \widetilde{\phi}(t) \mod 1$, where $\widetilde{\phi} : \mathbb{R} \to \mathbb{R}$ is absolutely continuous and 1-periodic, then $T_\phi^{q_n} \to \mathrm{Id}$ **uniformly** along the sequence $\{q_n\}$ of best rational approximations to α [14]. In [34], Wang proved that all such analytic Anzai skew products are Möbius disjoint (for an earlier result see Liu-Sarnak [26]), see also [16] for an extension of this result to the C^∞ case. In Section 3.4 we will also prove the following corollary:

Corollary 1.4. *For every $\epsilon > 0$, each irrational α and ϕ of zero topological degree and of class $C^{2+\epsilon}$, the corresponding Anzai skew product T_ϕ is Möbius disjoint.*

Another important class of systems with the property that $T^{q_n} \to \mathrm{Id}$ (uniformly) along some subsequence $\{q_n\}$ is given by smooth (area preserving) flows on $\mathbb{T}^2$. In fact, whenever such a flow has no fixed points, it is rigid. In particular, the recent result of [17] that such C^∞ flows are Möbius disjoint is a particular case of our main result whenever $\{q_n\}$ is of bounded prime volume. If we drop the assumption of the boundedness of the prime volume of $\{q_n\}$ then still Möbius disjointness will hold whenever the roof function is smooth enough, see Corollary 3.7. Finally, in Corollary 3.8, we prove new Möbius disjointness results for so called Rokhlin extensions.

Furthermore, Theorem 2.1, which also holds for the Liouville function λ instead of μ, has some purely arithmetic consequences for the correlations of the Liouville function. Matomäki and the third author showed in [27] that

$$(6) \qquad \limsup_{N \to \infty} \left| \frac{1}{N} \sum_{n \le N} \lambda(n)\lambda(n+h) \right| < 1 - \eta(h)$$

for some $\eta(h) > 0$ and all $h \ge 1$. We can weaken the dependence of η on h, along a subsequence of any given subsequence of N.

Corollary 1.5. *For each sequence $\{N_n\}_{n \ge 1} \subset \mathbb{N}$ there exists a subsequence $\{N_{n_k}\}$ such that for each $j \ge 1$ and some $\eta > 0$ (depending on the subsequence and j), we have:*

$$\lim_{k \to \infty} \frac{1}{N_{n_k}} \sum_{n \le N_{n_k}} \lambda(n)\lambda(n+h) < 1 - \eta,$$

for all $h \in \mathcal{D}_j$.

The proof of the above corollary uses the notion of *Furstenberg systems* associated to λ and will be given in Section 2.1.

We remark that Corollary 1.5 is related to a recent result of Tao and Terävänen [31] according to which Chowla's conjecture holds for "most" subsequences. However given a particular subsequence $\{N_{n_k}\}$ their result cannot guarantee the existence of even one subsequence of $\{N_{n_k}\}$ along which Chowla would hold. This is not surprising, since such a result would settle Chowla's conjecture. This marks an important but subtle distinction between their result and our Corollary 1.5.

2. Completely rigid points in topological dynamics

To prove Theorem 1.1 and Theorem 1.2 we will need the notion of completely rigid points, or, taking into account Definition 1.1, nicely rigid points.

Definition 2.1. Given a topological dynamical system (X, T) a point $x \in X$ is *quasi-generic with respect to a measure ν* if there exists a subsequence $M_k \to \infty$ along which we have,

$$\frac{1}{M_k} \sum_{m \leq M_k} \delta_{T^m x} \to \nu \tag{7}$$

weakly in the space of probability measures on X.

Notice that by (7), necessarily, $\nu \in M(X, T)$.

We are now ready to define *completely rigid (nicely rigid) points* in analogy with the *completely deterministic points* introduced in [18], [35].

Definition 2.2. Given a topological dynamical system (X, T), a point $x \in X$ is *completely rigid* if each measure $\nu \in M(X, T)$ for which x is quasi-generic yields a rigid measure-theoretic dynamical system. If, additionally, each such ν satisfies either BPV or PR rigidity condition then x is called *nicely rigid*.

Alternatively, we can restate the definition of x being nicely rigid by requiring that whenever

$$\frac{1}{M_k} \sum_{m \leq M_k} \delta_{T^m x} \to \nu,$$

then $(X, \mathcal{B}(X), \nu, T)$ satisfies either BPV rigidity or PR rigidity condition. In Section 3.6, we will prove the following more general version of Theorem 1.1.

Theorem 2.1. *Assume that (X, T) is a topological dynamical system. Assume that $x \in X$ is nicely rigid. Then*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(T^n x) \mu(n) = 0 \tag{8}$$

for each $f \in C(X)$.

It follows that whenever in (X, T) all points are nicely rigid, the system (X, T) satisfies Sarnak's conjecture (when PR rigidity condition holds, we obtain (8) for $f \in \mathcal{F}$ (and all $x \in X$), but it is not hard to see that (8) is a closed condition in $C(X)$). Theorem 1.1 is now an immediate consequence of Theorem 2.1.

2.1. Consequences for multiplicative functions.

Remark 2.2. A closer look at the proof of Theorem 2.1 shows that it has a "local" character. Namely, to show (8) for **fixed** $x \in X$ and $f \in C(X)$ we need to know that for each measure ν for which x is quasi-generic we have, for some $\{q_n\}$, $f \circ T^{q_n} \to f$ in $L^2(X, \nu)$ and either $\{q_n\} \subset \mathcal{D}_j$ (for some $j \geq 1$) or (for some $\delta > 0$) (5) holds. We will say that in this situation (x, f) satisfies either BPV or PR rigidity, respectively.

Note also that if $\frac{1}{N_k} \sum_{n \leq N_k} \delta_{T^n x} \to \nu$, and we are interested in establishing (8) along $\{N_k\}$, then we need $\{q_n\}$ (depending now on (f, x) and $\{N_k\}$) to satisfy the above conditions. In such a situation we will say that (x, f, ν) satisfies either BPV or PR rigidity, depending on whether $\{q_n\} \subset \mathcal{D}_j$ or whether (5) holds.

By treating μ as a point in the space $\{-1, 0, 1\}^{\mathbb{Z}}$ ($\mu(-n) = \mu(n)$) on which we have the action of the left shift S, we obtain the Möbius subshift (X_μ, S), where $X_\mu := \overline{\{S^n\mu : n \in \mathbb{Z}\}}$. On X_μ we consider all measures κ for which μ is quasi-generic (i.e. $\kappa = \lim_{k\to\infty} \frac{1}{N_k} \sum_{n\leq N_k} \delta_{S^n\mu}$). Each measure-theoretic system $(X_\mu, \mathcal{B}(X_\mu), \kappa, S)$ is called a *Furstenberg system* of μ.

Consider now the Möbius subshift (X_μ, S) with the function $\theta : X_\mu \to \{0, \pm 1\}$, $\theta(y) = y(0)$ for $y \in X_\mu$. Since the set of square-free numbers has positive density, μ is not orthogonal to itself (so (8) is not satisfied for $\theta \in C(X_\mu)$ and $\mu \in X_\mu$ **along any subsequence of** $\mathbb{N}$), Theorem 2.1 implies:

Corollary 2.3. *For no $\kappa \in M(X_\mu, S)$ for which μ is quasi-generic, (μ, θ, κ) satisfy either BPV or PR rigidity. In particular, no Furstenberg system $(X_\mu, \mathcal{B}(X_\mu), \kappa, S)$ of μ is BPV or PR rigid. The same holds for the Liouville function λ.*

Let us continue with λ.

Assume that we are given any Furstenberg system $(X_\lambda, \mathcal{B}(X_\lambda), \kappa, S)$ of λ, that is, assume that $\frac{1}{N_k} \sum_{n\leq N_k} \delta_{S^n\lambda} \to \kappa$. Then

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{n\leq N_k} \lambda(n)\lambda(n+h) = \lim_{k\to\infty} \frac{1}{N_k} \sum_{n\leq N_k} (\theta \cdot \theta \circ S^h)(S^n\lambda) =$$

$$\int \theta \cdot \theta \circ S^h \, d\kappa = \widehat{\sigma}_\theta[h]$$

for each $h \geq 1$ ($\widehat{\sigma}_\theta(h)$ stands for the hth Fourier coefficient of the spectral measure σ_θ of $\theta \in L^2(X_\lambda, \kappa)$). Note that $\int_{X_\lambda} \theta^2 \, d\kappa = 1$, so if $\widehat{\sigma}_\theta[q_n] \to 1$ along a subsequence $\{q_n\}$, then θ would be rigid. Since Corollary 2.3 holds, we can now prove Corollary 1.5:

Proof of Corollary 1.5. By the above, we have that there exists $\eta_0 > 0$ such that

$$\limsup_{\mathcal{D}_j \ni h \to \infty} \lim_{k\to\infty} \frac{1}{N_k} \left| \sum_{n\leq N_k} \lambda(n)\lambda(n+h) \right| < 1 - \eta_0.$$

To complete the proof, we use (6) (for "small" h). $\qquad\square$

3. PROPERTIES OF MÖBIUS FUNCTION AND THE PROOF OF THEOREM 2.1

3.1. **Möbius function on short intervals along arithmetic progressions.** We will prove the following result.

Theorem 3.1. *Let $A > 100$ and $\varepsilon \in (0, \frac{1}{100})$ be given. For all $X > X_0(\varepsilon, A)$, $H > H_0(\varepsilon, A)$ and $q \leq (\log X)^A$ such that*

$$(9) \qquad\qquad \sum_{p|q} \frac{1}{p} \leq (1-\varepsilon) \sum_{p\leq H} \frac{1}{p}$$

we have,

$$\frac{1}{qXH} \sum_{a<q} \int_X^{2X} \Big| \sum_{\substack{x \le n \le x+qH \\ n \equiv a \pmod q}} \mu(n) \Big| dx \le \varepsilon.$$

One can roughly think of the condition on q as excluding those q that are divisible by primorials such as $\prod_{p \le H} p$. It is a generic condition as we show in the lemma below.

Lemma 3.2. *Let $\varepsilon \in (0, \frac{1}{100})$ be given. For any $A > 100$,*

$$\#\Big\{ q \le X : \sum_{p|q} \frac{1}{p} \ge (1-\varepsilon) \sum_{p \le H} \frac{1}{p} \Big\} \ll_A X(\log H)^{-A}.$$

Furthermore, for all q sufficiently large,

$$\sum_{p|q} \frac{1}{p} \le \log \log \log q + O(1).$$

Therefore, (9) holds for all $H > \exp((\log \log q)^{1+2\varepsilon})$ and q sufficiently large in terms of ε^{-1}.

Proof. By Chernoff's bound the number of exceptions is bounded by,

$$\ll (\log H)^{-2A(1-\varepsilon)} \sum_{q \le X} \exp\Big(2A \sum_{p|q} \frac{1}{p}\Big).$$

By [32, Theorem 5, Ch. III.3] the above is,

$$\ll (\log H)^{-2A(1-\varepsilon)} X \prod_{p \le X} \Big(1 + \frac{e^{2A/p}-1}{p}\Big) \ll_A X(\log H)^{-A}.$$

For the second claim we notice that q has at most $2\log q$ distinct prime factors, since $\prod_{p<2\log q} p > q$ for all q sufficiently large. Therefore,

$$\sum_{p|q} \frac{1}{p} \le \sum_{p \le 2\log q} \frac{1}{p} \le \log \log \log q + O(1)$$

as claimed. ☐

The upper bound $\log \log \log q$ is attained for example by primorials $q = \prod_{p \le K} p$ once K is sufficiently large. The first part of the above lemma implies the following:

Corollary 3.3. *For $j \in \mathbb{N}$ let $\mathcal{D}_j$ be as in (3). Then $\lim_{j \to +\infty} d(\mathcal{D}_j) = 1$.*

Proof. Let $A = 1$, $\varepsilon = 1/2$ and let $H = \exp(j)$. Then

$$D_j^c \subset \Big\{ q \in \mathbb{N} : \sum_{p|q} \frac{1}{p} \ge (1-\varepsilon) \sum_{p \le H} \frac{1}{p} \Big\}.$$

The statement then immediately follows by Lemma 3.2 ☐

Before we prove the above theorem, let us state the following immediate corollary.

Corollary 3.4. *For each $\varepsilon \in (0, \frac{1}{100})$ there exists L_0 such that for each $L \geq L_0$ and $q \geq 1$ with*

$$\sum_{p|q} \frac{1}{p} \leq (1 - \varepsilon) \sum_{p \leq L} \frac{1}{p}$$

we can find $M_0 = M_0(q, L)$ such that for all $M \geq M_0$, we have

$$(10) \qquad \sum_{j=0}^{M/Lq} \sum_{a=0}^{q-1} \left| \sum_{\substack{m \in [z+jLq, z+(j+1)Lq) \\ m \equiv a \bmod q}} \mu(m) \right| \leq \varepsilon M$$

for some $0 \leq z < Lq$.

Proof. Partitioning into dy-adic intervals we readily see that the conclusion of Theorem 3.1 might have been as-well stated with an integration over $0 \leq x \leq X$ instead of an integration over $X \leq x \leq 2X$. It follows from this that for any $\varepsilon > 0$ and $q \geq 1$ and all L and M sufficiently large in terms of ε and q,

$$\sum_{0 \leq z < qL} \left(\sum_{0 < j < M/(Lq)} \sum_{a < q} \left| \sum_{\substack{n \in [z+jqL, z+(j+1)qL] \\ n \equiv a \bmod q}} \mu(n) \right| \right) \leq \varepsilon M \cdot Lq.$$

The claim now follows by the pigeon-hole principle. $\square$

Proof of Theorem 3.1. Grouping terms according to $d = (a, q)$ we have the bound,

$$\frac{1}{qXH} \sum_{a < q} \int_X^{2X} \left| \sum_{\substack{x \leq n \leq x+qH \\ n \equiv a \pmod q}} \mu(n) \right| dx \leq \frac{1}{qXH} \sum_{d|q} \sum_{\substack{0 \leq a < q/d \\ (a, q/d)=1}} \int_X^{2X} \left| \sum_{\substack{x/d \leq n \leq x/d+(q/d)H \\ n \equiv a \pmod{q/d}}} \mu(dn) \right| dx.$$

Let $\varepsilon > 0$. We now introduce the set $\mathcal{S}_d$ which consists of integers n that have at least one prime divisor in each of the intervals $[P_i, Q_i]$ where $Q_1 = dH$, $P_1 = \log^{2000} Q_1$ and P_j, Q_j are as in [27, equation (4)], that is

$$P_j = \exp(j^{4j}(\log Q_1)^{j-1} \cdot \log P_1) , \ Q_j = \exp(j^{4j+2}(\log Q_1)^j) , \ j \leq J$$

with J the largest index such that $Q_J \leq \exp(\sqrt{\log X})$. We then find that the contribution of the integers not in $\mathcal{S}_d$ is

$$\leq \frac{1}{qXH} \sum_{d|q} \sum_{\substack{0 \leq a < q/d \\ (a, q/d)=1}} \int_X^{2X} \left(\sum_{\substack{x/d \leq n \leq x/d+(q/d)H \\ n \equiv a \pmod{q/d} \\ n \notin \mathcal{S}_d}} 1 \right) dx.$$

By Fubini's theorem, this is

$$(11) \qquad \leq \frac{1}{X} \sum_{d|q} \sum_{\substack{X/d \leq n \leq 2X/d \\ (n, q/d)=1 \\ n \notin \mathcal{S}_d}} 1.$$

Notice that, for any $j \geq 1$,

$$(12) \qquad \sum_{\substack{P_j \leq p \leq Q_j \\ p \nmid q}} \frac{1}{p} \geq \sum_{P_j \leq p \leq Q_j} \frac{1}{p} - \sum_{p | q} \frac{1}{p} \geq \frac{\varepsilon}{2} \log \log H + 2 \log j.$$

Therefore, by a standard sieve bound,

$$\sum_{\substack{X/d \leq n \leq 2X/d \\ (n,q/d)=1 \\ n \notin \mathcal{S}_d}} 1 \ll \frac{X}{d} \frac{\varphi(q/d)}{q/d} \sum_{j=1}^{J} \prod_{\substack{P_j \leq p \leq Q_j \\ p \nmid q}} \left(1 - \frac{1}{p}\right) \ll (\log H)^{-\varepsilon/2} \frac{\varphi(q/d)}{q} \cdot X \leq \varepsilon \cdot \frac{\varphi(q/d)}{q} \cdot X$$

for all sufficiently large H, and therefore, (11) is $\ll \varepsilon$. Thus, it suffices to bound

$$\frac{1}{qXH} \sum_{d|q} \sum_{\substack{0 \leq a < q/d \\ (a,q/d)=1}} \int_X^{2X} \Big| \sum_{\substack{x/d \leq n \leq x/d+(q/d)H \\ n \equiv a \pmod{q/d} \\ n \in \mathcal{S}_d}} \mu(dn) \Big| dx$$

$$\leq \left(\frac{1}{qXH} \sum_{d|q} \sum_{\substack{0 \leq a < q/d \\ (a,q/d)=1}} X\right)^{1/2} \cdot \left(\frac{1}{qXH} \sum_{d|q} \sum_{\substack{0 \leq a < q/d \\ (a,q/d)=1}} \int_X^{2X} \Big| \sum_{\substack{x/d \leq n \leq x/d+(q/d)H \\ n \equiv a \pmod{q/d} \\ n \in \mathcal{S}_d}} \mu(dn) \Big|^2 dx\right)^{1/2}.$$

Therefore, to conclude, it will suffice to show that

$$\frac{1}{qXH^2} \sum_{d|q} \sum_{\substack{0 \leq a < q/d \\ (a,q/d)=1}} \int_X^{2X} \Big| \sum_{\substack{x/d \leq n \leq x/d+(q/d)H \\ n \equiv a \pmod{q/d} \\ n \in \mathcal{S}_d}} \mu(dn) \Big|^2 dx \ll \varepsilon^2.$$

We express the condition $n \equiv a \pmod{q/d}$ using Dirichlet characters. This allows us to re-write the above equation as,

$$\frac{1}{qXH} \sum_{d|q} d \cdot \frac{1}{\varphi(q/d)} \sum_{\chi \pmod{q/d}} \int_{X/d}^{2X/d} \Big| \sum_{\substack{x \leq n \leq x+(q/d)H \\ n \in \mathcal{S}_d}} \mu(dn)\chi(n) \Big|^2 dx.$$

We now claim that a variant of [27, Lemma 14] gives,

$$\frac{1}{X/d} \frac{1}{((q/d)H)^2} \int_{X/d}^{2X/d} \Big| \sum_{\substack{x \leq n \leq x+(q/d)H \\ n \in \mathcal{S}_d}} \mu(dn)\chi(n) \Big|^2 dx$$

$$(13)$$

$$\ll \frac{1}{(\log X)^{100A}} + \int_{(\log X)^{100A}}^{X/(qH)} |D_d(1+it,\chi)|^2 dt + \max_{T > X/(qH)} \frac{X/(qH)}{T} \int_T^{2T} |D_d(1+it,\chi)|^2 dt,$$

where

$$D_d(1+it,\chi) := \sum_{\substack{X \leq n \leq 4X \\ n \in \mathcal{S}_d}} \frac{\mu(dn)\chi(n)}{n^{1+it}}.$$

To obtain this we repeat the proof in [27, Lemma 14] but choose $T_0 = (\log X)^{100A}$, $h_2 = X/(\log X)^{1000A}$ and notice that the prime number theorem in arithmetic progressions gives

$$\frac{1}{h_2}S_2(x) \ll (\log X)^{-10000A}.$$

We will focus only on bounding the first integral in (13) since the contribution of the second integral is handled by simply repeating the argument. Therefore, it remains to obtain a bound of the form $\ll \varepsilon^2$ for,

$$\frac{1}{qH^2} \sum_{d|q} \frac{1}{\varphi(q/d)} \sum_{\chi \pmod{q/d}} \left(\frac{qH}{d}\right)^2 \int_{(\log X)^{100A}}^{X/(qH)} |D_d(1+it,\chi)|^2 d.t$$

We re-write this as

$$(14) \qquad \sum_{d|q} \frac{q/d}{\varphi(q/d)} \cdot \frac{1}{d} \sum_{\chi \pmod{q/d}} \int_{(\log X)^{100A}}^{X/(qH)} |D_d(1+it,\chi)|^2 dt.$$

We claim that a slight modification of [27, Proposition 1] gives the following "hybrid variant" of [27, Proposition 1],

$$(15) \qquad \sum_{\chi \pmod{q/d}} \int_{(\log X)^{100A}}^{X/(qH)} |D_d(1+it,\chi)|^2 dt \ll_\eta \left(\frac{(\log Q_1)^{1/3}}{P_1^{1/6-\eta}} + \frac{1}{(\log X)^{50A}}\right) \cdot \frac{\varphi(q/d)}{q/d}$$

for any $\eta > 0$. Inserting this "hybrid variant" into (14) shows that (14) is

$$\ll \sum_{d|q} \frac{1}{d} \cdot \left((\log dH)^{-2} + \frac{1}{(\log X)^{10A}}\right) \ll_\varepsilon \frac{1}{\log H} + \frac{1}{\log^A X}$$

(since $\sum_{d|q} d^{-1} \ll \log H$ by the assumption $\sum_{p|q} p^{-1} \leq \log\log H$) which is less than ε for sufficiently large H and X.

Therefore, all that remains to be done is to explain how to obtain (15). We repeat the argument in [27, Proposition 1] with the following differences:

- Instead of $Q_{v,H_j}(s)$ we work with,

$$Q_{v,H_j}(s,\chi) := \sum_{\substack{P_j \leq q \leq Q_j \\ e^{v/H_j} \leq q \leq e^{(v+1)/H_j}}} \frac{\chi(q)}{q^s}$$

and, instead of $R_{v,H_j}(s)$, we work with

$$R_{v,H_j}(s,\chi) := \sum_{\substack{Xe^{-v/H_j} \leq m \leq 2Xe^{-v/H_j} \\ m \in \mathcal{S}_j}} \frac{\mu(dm)\chi(m)}{m^s} \cdot \frac{1}{\#\{P_j \leq p \leq Q_j : p|m\} + 1},$$

where $\mathcal{S}_j$ is the set of those integers which have at least one prime factor in every interval $[P_i, Q_i]$ with $i \neq j$ and $i \leq J$.

- We define $\mathcal{T}_j$ as the set of those (t, χ) with $(\log X)^A \leq t \leq X/(qH)$ and χ of modulus q/d for which $|Q_{v,H_j}(1 + it, \chi)| < e^{-\alpha_j v/H_j}$ for all v, and $\mathcal{U}$ as the set of those (t, χ) that do not belong to any of the $\mathcal{T}_1 \cup \ldots \cup \mathcal{T}_j$.

- In Section 8.1 instead of using the standard mean-value theorem we use the following "hybrid mean-value theorem" (see [29, Theorem 6.4] for a proof),

$$(16) \qquad \sum_{\chi \pmod{q/d}} \int_0^T \Big| \sum_{X \leq n \leq 4X} \frac{a(n)}{n^{1+it}} \Big|^2 \ll \Big(\frac{X}{qH} \cdot \varphi(q/d) + X \Big) \sum_{\substack{X \leq n \leq 4X \\ (n, q/d) = 1}} \frac{|a(n)|^2}{n^2}.$$

This leads to a bound for E_1 which is

$$E_1 \ll \frac{(\log Q_1)^{1/3}}{P_1^{1/6 - \eta}} \cdot \Big(\varphi\Big(\frac{q}{d}\Big) \frac{Q_1}{qH} + 1 \Big) \cdot \frac{\varphi(q/d)}{q/d} \ll \frac{(\log Q_1)^{1/3}}{P_1^{1/6 - \eta}} \cdot \frac{\varphi(q/d)}{q/d}$$

and explains the choice of Q_1.

- In Section 8.2 we again appeal to the "hybrid mean-value theorem" instead of the standard mean-value theorem. In particular, repeating the proof of Lemma 13 with the hybrid mean-value theorem we obtain,

$$\sum_{\chi \pmod{q/d}} \int_{(\log X)^A}^{X/(qH)} |Q_{r,H_{j-1}}(1 + it, \chi)^{\ell_{j,r}} \cdot R_{v,H_j}(1 + it, \chi)|^2 dt$$

$$\ll \Big(\frac{\varphi(q/d)}{qH} + Q_{j-1} \Big) \exp(2\ell_{j,r} \log \ell_{j,r}) \cdot \frac{\varphi(q/d)}{q/d}.$$

This leads to the bound

$$E_j \ll \frac{1}{j^2} \cdot \frac{1}{P_1} \cdot \frac{\varphi(q/d)}{q/d}.$$

- In Section 8.3 it suffices to bound

$$\sum_{(t,\chi) \in \mathcal{T}} |Q_{v,H}(1 + it, \chi) R_{v,H}(1 + it, \chi)|^2$$

with the sum taken over a 1-well spaced set of tuples (t, χ) (that is if $(t, \chi) \neq (t', \chi)$ then $|t - t'| \geq 1$) with $\mathcal{T} \subset \mathcal{U}$. By the prime number theorem, we have $Q_{v,H}(1 + it, \chi) \ll (\log X)^{-50A}$ for any $(t, \chi) \in \mathcal{T}$. Therefore, it remains to bound,

$$(\log X)^{-100A} \sum_{(t,\chi) \in \mathcal{T}} |R_{v,H}(1 + it, \chi)|^2.$$

By a hybrid version of Halasz's Lemma (see [29, Theorem 8.3]) this is

$$\ll (\log X)^{-100A} \Big(X e^{-v/H} + |\mathcal{T}| \cdot \Big(\frac{q}{d} \cdot \frac{X}{qH} \Big)^{1/2} \Big) \frac{\log X}{X e^{-v/H}}.$$

Moreover, repeating the proof of [27, Lemma 8] using the hybrid mean-value (16) we obtain that $|\mathcal{T}| \ll (X/(dH))^{1/2-\eta+o(1)}$. From this it follows that the above expression is $\ll (\log X)^{-100A}$ as needed.

Summing the error term that we obtain in the above modifications of Section 8.1, 8.2 and 8.3, we obtain the claimed bound (15). □

3.2. **A short discussion of BPV and PR rigidity.** In this short section we would like to point out differences and common features of BPV and PR rigidity, defined in Definition 1.1. BPV rigidity puts arithmetic restrictions on rigidity times but its advantage is that it depends only on the measure-theoretic properties of ν, hence it does not "see" topological properties of (X, T). On the other hand, PR rigidity does not impose any arithmetic restrictions on rigidity times, however it takes into account the topological properties of (X, T) expressed by a speed of rigidity for $\nu \in M(X, T)$ and certain continuous functions (namely $f \in \mathcal{F}$). If in a concrete situation we have a measure-theoretic system and BPV rigidity holds then Möbius disjointness holds in **all** uniquely ergodic models of the system (hence, in Corollary 1.3, we obtain that in **each** uniquely ergodic model of an IET from the set C^c Möbius disjointness holds). In contrast PR rigidity applies only in **selected** uniquely ergodic models. However whenever PR rigidity holds, we will show in Corollary 4.2 that not only Möbius disjointness holds but, in fact, we have the strong MOMO property in this model. This, by Theorem 4.1, implies that Möbius disjointness holds in **all** uniquely ergodic systems of the original measure-theoretic system (hence, also for uniquely ergodic systems considered in Corollary 1.4, we have Möbius disjointness in all other uniquely ergodic models).

To illustrate this, consider an ergodic automorphism with discrete spectrum. By Halmos-von Neumann theorem, it has a uniquely ergodic model which is given as a rotation $Tx = x + a$ on a (metric) compact, abelian group X considered with Haar measure. Now, $\mathcal{F}$ is given by the group of characters of X each of which is an eigenfunction. If $f \in \mathcal{F}$ corresponds to an eigenvalue $e^{2\pi i\alpha}$ with $\alpha \in [0, 1)$ irrational then (5) is satisfied as it is $\ll \|q_n\alpha\|^2 \sum_{j=-q_n^\delta}^{q_n^\delta} j^2 \to 0$ (for a small $\delta > 0$), where q_n, $n \geq 1$, stands for the sequence of denominators of α. If α is rational then consider just multiples of its denominators. It follows that **all** uniquely ergodic models of systems with discrete spectrum are Möbius disjoint (cf. [15] for the first proof of this result). Note that BPV rigidity is satisfied for each irrational rotation $Tx = x + \alpha$ on $\mathbb{T}$ (this follows from Vinogradov's theorem on equidistribution of $p\alpha$, $p \in \mathbb{P}$) but it is not clear how to apply it to any ergodic discrete spectrum automorphism.

Finally, note that a natural situation in which PR rigidity holds is the case of *pointwise rigid* homeomorphisms satisfying the *topological PR rigidity* condition: for some $\{q_n\}$, $\delta > 0$ and each $x \in X$, we have

$$(17) \qquad \sum_{j=-q_n^\delta}^{q_n^\delta} d(T^{jq_n}x, x) \to 0.$$

Indeed, assuming that the metric d is bounded by 1, for $\mathcal{F}$ we can take the family of Lipschitz continuous functions and for each such function and each $\nu \in M(X, T)$, condition (5) follows automatically from (17). For different concepts of rigidity in topological dynamics, see [13].

3.3. **Interval exchange transformations.** In this section we will prove Corollary 1.3. We first recall the following result of Chaika:

Theorem 3.5 (Corollary 3, [5]). *For every $d \in \mathbb{N}$ and $\epsilon > 0$ there exists $0 < a(\epsilon) < 1$ such that for every $E \subset \mathbb{N}$ satisfying $\underline{d}(E) \geq a(\epsilon)$, there exists a set of d IET's of measure at least $1 - \epsilon$ such that each IET in this set has a rigidity sequence in E.*

Using the above result we are able to prove Corollary 1.3:

Proof of Corollary 1.3. Let $\epsilon_n = \frac{1}{n}$ and let j_n be such that $d(D_{j_n}) \geq a(\epsilon_n)$ (such j_n exists by Corollary 3.3). Then, by Theorem 3.5, all but a set C_n of measure ϵ_n of d IET's have a rigidity sequence in $\mathcal{D}_{j_n}$ (we can WLOG assume that such IET's are also uniquely ergodic). Hence, BPV rigidity (see Definition 1.1) is satisfied for every d IET outside C_n. Hence, by Theorem 1.1, every such IET is Möbius disjoint. It is then enough to notice that the measure of $C = \bigcap_n C_n$ is 0 and every IET outside C is Möbius disjoint. $\square$

3.4. **Anzai skew products.** In this section we will prove Corollary 1.4. By an abuse of notation, we can treat ϕ as a $C^{2+\epsilon}$ real-valued function on $\mathbb{T}$: $\phi \in C^{2+\epsilon}(\mathbb{T})$, and therefore

$$\phi(x) = \sum_{j \in \mathbb{Z}} a_j e_j(x) \quad \text{where} \quad |a_j| \ll j^{-2-\epsilon} \text{ and } a_0 = 0.$$

Let $\{q_n\}$ denote the sequence of denominators of α. We have the following lemma:

Lemma 3.6. *There exists a subsequence $\{q_{n_m}\}$ such that*

$$\sup_{(x,y) \in \mathbb{T}^2} \sup_{|k| \leq q_{n_m}^{\epsilon^2/4}} d(T_{\alpha,\phi}^{kq_n}(x,y), (x,y)) \ll q_{n_m}^{-\epsilon^2/2}.$$

Before we prove the above lemma, let us show how it implies Corollary 1.4.

Proof of Corollary 1.4. Let $\{q_{n_m}\}$ denote the sequence from Lemma 3.6. We have

$$\sum_{|k| \leq q_{n_m}^{\epsilon^2/4}} d(T_{\alpha,\phi}^{kq_n}(x,y), (x,y)) \ll q_{n_m}^{-\epsilon^2/4},$$

so condition (17) is satisfied and therefore PR rigidity is satisfied for the sequence $\{q_{n_m}\}$, $\delta = \epsilon^2/4$ and the family $\mathcal{F}$ of Lipschitz functions. The statement then follows by Theorem 1.1. $\square$

So, it only remains to prove Lemma 3.6.

Proof of Lemma 3.6. Notice that $T_{\alpha,\phi}^{kq_n}(x,y) = (x+kq_n\alpha, y+S_{kq_n}(\phi)(x))$, where $S_r(\phi)(x) := \phi(x) + \phi(x+\alpha) + \ldots + \phi(x+(r-1)\alpha)$, for $r \geq 1$.

For every $n \in \mathbb{N}$, we have $S_{q_n}(\phi)(x) = \sum_{j \in \mathbb{Z}} a_j S_{q_n}(e_j)(x)$, where $e_j(x) := e^{2\pi i j x}$. Moreover,

$$S_{q_n}(e_j)(x) = e_j(x) \frac{1 - e_j(q_n \alpha)}{1 - e_j(\alpha)}.$$

For $|j| \geq q_n$, we have

$$\left| a_j \frac{1 - e_j(q_n \alpha)}{1 - e_j(\alpha)} \right| \leq |j|^{-2-\epsilon} q_n.$$

Therefore (for every $n \in \mathbb{N}$),

(18)
$$\left| \sum_{|j| \geq q_n} a_j S_{q_n}(e_j)(x) \right| \leq q_n \sum_{|j| \geq q_n} |j|^{-2-\epsilon} < q_n^{-\epsilon}.$$

Moreover, notice that for $0 < |j| < q_n$, using $\|j q_n \alpha\| < j/q_{n+1}$ and $|1 - e(j)| \sim \|j\alpha\|$, we obtain

$$\left| a_j \frac{1 - e_j(q_n \alpha)}{1 - e_j(\alpha)} \right| \ll q_{n+1}^{-1} |j|^{-1-\epsilon} \|j\alpha\|^{-1}.$$

Notice that

(19)
$$\left| \sum_{|j| < q_n} a_j S_{q_n}(e_j)(x) \right| \leq q_{n+1}^{-1} \sum_{|j| < q_n} |j|^{-1-\epsilon} \|j\alpha\|^{-1}.$$

Since $\|q_n \alpha\| < q_{n+1}^{-1} < q_n^{-1}$, for $|k| \leq q_n^{\epsilon^2/2}$, we have $\|k q_n \alpha\| \leq q_n^{-1+\epsilon/4}$. So, by the cocycle identity for $S_r(\phi)$, the lemma follows by showing that there exists a subsequence $\{q_{n_m}\}$ satisfying

(20)
$$\sup_{x \in \mathbb{T}} |S_{q_{n_m}}(\phi)(x)| \ll q_{n_m}^{-\epsilon/8}.$$

We consider two cases:
A. There exists a subsequence $\{q_{n_m}\}$ such that $q_{n_m+1} > q_{n_m}^{1+\epsilon/4}$. In this case we will show that (20) holds for $\{q_{n_m}\}$.
B. For all (sufficiently large) n, $q_{n+1} \leq q_n^{1+\epsilon/4}$.
Note moreover that $\sup_{|j| < q_n} \|j\alpha\|^{-1} < 2q_n$. Using this if we are in case **A.**, along the sequence $\{q_{n_m}\}$, (19) is bounded above by

$$q_{n+1}^{-1} \sum_{|j| < q_n} |j|^{-1-\epsilon} \|j\alpha\|^{-1} \ll q_{n_m}^{-\epsilon/4},$$

which, together with (18) finishes the proof if **A** holds.
If **B.** holds, we will show that (20) holds for $n_m = m$. Let $0 < k \leq n$. Notice that (19) is bounded above by

$$q_{n+1}^{-1} \sum_{|j| < q_k} |j|^{-1-\epsilon} \|j\alpha\|^{-1} + q_{n+1}^{-1} \sum_{q_k \leq |j| < q_n} |j|^{-1-\epsilon} \|j\alpha\|^{-1}.$$

For the first summand above we use that $\sup_{|j|<q_k} \|j\alpha\|^{-1} < 2q_k$ and therefore,

$$q_{n+1}^{-1} \sum_{|j|<q_k} |j|^{-1-\epsilon} \|j\alpha\|^{-1} \ll \frac{q_k}{q_{n+1}}.$$

Moreover, since $\sup_{|j|<q_n} \|j\alpha\|^{-1} < 2q_n < 2q_{n+1}$,

$$q_{n+1}^{-1} \sum_{q_k \leq |j| < q_n} |j|^{-1-\epsilon} \|j\alpha\|^{-1} \ll \sum_{q_k \leq |j| < q_n} |j|^{-1-\epsilon} \ll q_k^{-\epsilon}.$$

Putting the above bounds together we get that (19) is

$$\ll \frac{q_k}{q_{n+1}} + q_k^{-\epsilon}.$$

We now choose k so that $q_k \in [q_n^{1/4}, q_n^{1/2}]$. Notice that this is possible since we are in Case **B.** (and so for every sufficiently large a, there exists $k_0 \in \mathbb{N}$ such that $q_{k_0} \in [a, a^2]$). Then (19) is $\ll q_n^{-\epsilon/4}$. Using this and (18), we get that (20) holds for $n_m = m$. This finishes the proof. □

3.5. **Smooth flows on $\mathbb{T}^2$ and Rokhlin extensions.** Using the same arguments as in Section 3.4, we can provide some new instances of Möbius disjointness. Recall that smooth time changes of the linear flow on $\mathbb{T}^2$ are given by (here $\alpha \in \mathbb{T}$ is irrational) $\frac{dx}{dt} = \frac{\alpha}{F(x,y)}$, $\frac{dy}{dt} = \frac{1}{F(x,y)}$ for a smooth, positive function $F : \mathbb{T}^2 \to \mathbb{R}$. They have a special representation as a (special) flow $T^f = (T_t^f)_{t \in \mathbb{R}}$ over the irrational rotation $Tx = x + \alpha$ and under a smooth function $f : \mathbb{T} \to \mathbb{R}$ satisfying (for simplicity) $\int f\, dLeb = 1$ (in fact, each smooth area-preserving flow on $\mathbb{T}^2$ has such a representation whenever the flow has no fixed points). The space $\mathbb{T}^f$ has a natural metric D making it a compact metric space, and

(21) $$D(T_t(x,y),(x,y)) \leq |t| \text{ whenever } t \text{ is small enough.}$$

We have

$$D(T_{q_n}^f(x,y),(x,y)) = D(T_{q_n - S_{q_n}(f)(x)}^f(T_{S_{q_n}(f)(x)}^f(x,y)),(x,y)) \leq$$

$$D(T_{q_n - S_{q_n}(f)(x)}^f(T^{q_n}x,y),(T^{q_n}x,y)) + D((T^{q_n}x,y),(x,y)).$$

Using (21) and (20), we obtain PR rigidity condition (along the rigidity time $\{q_{n_m}\}$ for the special flow T^f). We have proved:

Corollary 3.7. *Time-1 maps of the special flows T^f over irrational rotation $Tx = x + \alpha$ and under $C^{2+\epsilon}$ roof functions f are Möbius disjoint.*

Another class of examples is given by Rokhlin extensions. Here, we fix an irrational rotation $Tx = x + \alpha$ and a smooth zero mean $f : \mathbb{T} \to \mathbb{R}$. Assume that $\mathcal{R} = (R_t)_{t \in \mathbb{R}}$ is a (measurable) flow on $(Y, \mathcal{C}, \kappa)$. Then on the product space $X \times Y$ we have the corresponding Rokhlin extension $T_{f,\mathcal{R}}$ given by

$$T_{f,\mathcal{R}}(x,y) = (Tx, R_{f(x)}(y)).$$

(Note that Anzai skew products are special cases of this construction in which $Y = \mathbb{T}$ and $\mathcal{R}$ stands for the linear flow on $\mathbb{T}$.) If $\mathcal{R}$ is a smooth flow on a compact manifold Y with metric ρ then (for the product metric D on $X \times Y$), we have

$$D(T_{f,\mathcal{R}}^{q_n}(x,y)(x,y)) = d(T^{q_n}x, x) + \rho(R_{S_{q_n}(f)(x)}(y), y),$$

if the flow $\mathcal{R}$ is Lipschitz continuous, then (20) will again be satisfied and we obtain the following:

Corollary 3.8. *Assume that $f \in C^{2+\epsilon}(\mathbb{T})$. Then for **all** Lipschitz continuous $\mathcal{R}$, Rokhlin extensions $T_{f,\mathcal{R}}$ are Möbius disjoint.*

Notice that one can take $\mathcal{R} = (h_t)_{t \in \mathbb{R}}$ to be the horocycle flow. In this case the fiber dynamics is very different from the case of Anzai skew-products (mixing with countable Lebesgue spectrum) and, by the above Corollary, Möbius disjointness still holds.

Remark 3.9. If the sequence $\{q_n\}$ of denominators of α has bounded prime volume (that BPV rigidity holds) then the above corollary is true for **all continuous** flows $\mathcal{R}$. In fact, the Möbius disjointness holds for uniquely ergodic models for **all measurable** flows $\mathcal{R}$, to be compared with a result of [24].

3.6. **Proof of Theorem 2.1.**

Proof. To simplify notation, we will assume that x is generic for ν which yields a rigid system with a rigidity time $\{q_n\}$ along which either BPV or PR rigidity holds (see Definition 1.1).

Fix a continuous $f : X \to \mathbb{R}$, where in case PR rigidity is satisfied, we assume additionally that $f \in \mathcal{F}$. Select $L_n \to \infty$ slowly enough to have

$$\sum_{j=-L_n+1}^{L_n-1} \|f \circ T^{jq_n} - f\|_{L^2(\nu)}^2 \to 0.$$

Note that such a sequence obviously exists when BPV rigidity (along $\{q_n\}$) is satisfied, while in the case of PR rigidity, we simply take $L_n = q_n^\delta$. Fix $\varepsilon > 0$ (sufficiently small). Then, for n large enough (which we fix)

$$(22) \qquad \int_X \sum_{j=-L_n+1}^{L_n-1} \left| f \circ T^{jq_n} - f \right|^2 d\nu < \varepsilon.$$

Since x is generic for ν, by (22), we obtain

$$\lim_{M \to \infty} \frac{1}{M} \sum_{m \leq M} \left(\sum_{j=-L_n+1}^{L_n-1} \left| f(T^{jq_n+m}x) - f(T^m x) \right|^2 \right) < \varepsilon.$$

Hence, for some $M_0 = M_0(\varepsilon)$ and every $M > M_0$, we have

$$(23) \qquad \frac{1}{M} \sum_{m \leq M} \left(\sum_{j=-L_n+1}^{L_n-1} \left| f(T^{jq_n+m}x) - f(T^m x) \right|^2 \right) < \varepsilon.$$

We say that $m \leq M$ is *good* if

$$(24) \qquad \sum_{j=-L_n+1}^{L_n-1} \left| f(T^{jq_n+m}x) - f(T^m x) \right|^2 < \varepsilon^{1/2}.$$

Then by Markov's inequality,

$$(25) \qquad |\{m \leq M : m \text{ is good}\}| > (1 - \varepsilon^{1/2})M.$$

We also assume that (10) holds for M (with ε replaced by $\varepsilon/\|f\|_\infty$). Moreover, as M is arbitrarily large compared to $L_n q_n$, no harm to assume that $z = 0$; indeed, otherwise replace intervals $[jL_n q_n, (j+1)L_n q_n)$ by $[z + jL_n q_n, z + (j+1)L_n q_n)$ in the reasoning below. So now, we write $[0, M] = \bigcup_{j=0}^{M/(L_n q_n)}[jL_n q_n, (j+1)L_n q_n)$ (we do not pay attention to the last interval as M is arbitrarily large with respect to $L_n q_n$). We say that an interval $[jL_n q_n, (j+1)L_n q_n)$ is *good* if the number of good m in it is at least $(1 - \varepsilon^{1/4})L_n q_n$. It follows that

$$(26) \qquad \left| \left\{ j \leq \frac{M}{L_n q_n} : [jL_n q_n, (j+1)L_n q_n) \text{ is good} \right\} \right| \geq (1 - \varepsilon^{1/4})\frac{M}{L_n q_n}.$$

Indeed, if K denotes the number of good intervals, then by (25),

$$(1 - \varepsilon^{1/2})M \leq \left(\frac{M}{L_n q_n} - K \right)(1 - \varepsilon^{1/4})L_n q_n + KL_n q_n,$$

which implies that $(\epsilon^{1/4} - \epsilon^{1/2})\frac{M}{L_n q_n} \leq \epsilon^{1/4}K$ and this proves (26).

Now, take $[jL_n q_n, (j+1)L_n q_n)$ and assume that it is good. We will consider numbers in this interval mod q_n. We say that $a \in \{0, 1, \ldots, q_n - 1\}$ is good if there exists $m = m_a \in [jL_n q_n, (j+1)L_n q_n)$ which is good and $m \equiv a \mod q_n$. Note that

$$(27) \qquad |\{a \in \{0, 1, \ldots, q_n - 1\} : a \text{ is good}\}| \geq (1 - \varepsilon^{1/4})q_n.$$

Indeed, if a is bad, then it produces L_n of bad m. Note also that whenever a is good (and $m_a \in [jL_n q_n, (j+1)L_n q_n)$ is good with $m_a \equiv a \mod q_n$) then for each $m_1, m_2 \in [jL_n q_n, (j+1)L_n q_n)$, with $m_i \equiv a \mod q_n$, $i = 1, 2$, by (24) for m_a, we have

$$(28) \quad |f(T^{m_1}x) - f(T^{m_2}x)| \leq |f(T^{m_1}x) - f(T^{m_a}x)| + |f(T^{m_a}x) - f(T^{m_2}x)| \leq 2\varepsilon^{1/4}.$$

We fix $[jL_n q_n, (j+1)L_n q_n)$ which is good and we evaluate the relevant part of the sum $\frac{1}{M}\sum_{m \leq M} f(T^m x)\mu(m)$ using (27) and then (28):

$$\left| \sum_{m \in [jL_n q_n, (j+1)L_n q_n)} f(T^m x)\mu(m) \right| \leq$$

$$\sum_{\substack{a=0 \\ a \text{ good}}}^{q_n-1} \left| \sum_{\substack{m \in [jL_n q_n, (j+1)L_n q_n) \\ m \equiv a \mod q_n}} f(T^m x)\mu(m) \right| + \|f\|_\infty \cdot \varepsilon^{1/4}L_n q_n \leq$$

$$\|f\|_\infty \sum_{\substack{a=0 \\ a \text{ good}}}^{q_n-1} \left| \sum_{\substack{m\in[jL_nq_n,(j+1)L_nq_n) \\ m\equiv a \bmod q_n}} \mu(m) \right| + 2\varepsilon^{1/4}L_nq_n + \|f\|_\infty \cdot \varepsilon^{1/4}L_nq_n.$$

We have (in view of (26) and the estimate above)

$$\left| \sum_{m\le M} f(T^m x)\mu(m) \right| = \left| \sum_{j=0}^{M/(L_nq_n)} \sum_{m\in[jL_nq_n,(j+1)L_nq_n)} f(T^m x)\mu(m) \right| \le$$

$$\sum_{\substack{j=0 \\ j \text{ good}}}^{M/(L_nq_n)} \left| \sum_{m\in[jL_nq_n,(j+1)L_nq_n)} f(T^m x)\mu(m) \right| + \varepsilon^{1/2}\frac{M}{L_nq_n}\cdot\|f\|_\infty L_nq_n \le$$

$$\sum_{\substack{j=0 \\ j \text{ good}}}^{M/(L_nq_n)} \left(\|f\|_\infty \sum_{\substack{a=0 \\ a \text{ good}}}^{q_n-1} \left| \sum_{\substack{m\in[jL_nq_n,(j+1)L_nq_n) \\ m\equiv a \bmod q_n}} \mu(m) \right| + 2\varepsilon^{1/4}L_nq_n + \|f\|_\infty\cdot\varepsilon^{1/4}L_nq_n \right) +$$

$$\varepsilon^{1/2}M\|f\|_\infty = \|f\|_\infty \sum_{\substack{j=0 \\ j \text{ good}}}^{M/(L_nq_n)} \sum_{\substack{a=0 \\ a \text{ good}}}^{q_n-1} \left| \sum_{\substack{m\in[jL_nq_n,(j+1)L_nq_n) \\ m\equiv a \bmod q_n}} \mu(m) \right| + \mathrm{O}(\varepsilon^{1/7}M).$$

The last expression is $\mathrm{O}(\varepsilon^{1/10}M)$ since we have assumed (10) in Corollary 3.4 to hold for $z=0$. Indeed, when BPV rigidity holds, we can apply Corollary 3.4, since $L_n \to \infty$, so $\log\log L_n \to \infty$ and the prime volume of $\{q_n\}$ is bounded, while if PR rigidity is satisfied, we first use the second statement in Lemma 3.2 to see that Corollary 3.4 is applicable (recall that $L_n = q_n^\delta$). The result follows. $\square$

Remark 3.10. Note that the argument used at the beginning of the proof gives the following. If T is uniquely ergodic and $\|\sum_{j=-L_n}^{L_n} f\circ T^{jq_m} - f\|_{L^2(\nu)} < \varepsilon$ then for all $M \ge M_0$ we have

$$\frac{1}{M}\sum_{m\le M}\sum_{j=-L_n}^{L_n}\|f\circ T^{jq_n+m} - f\circ T^m\|_{C(X)} < \varepsilon.$$

4. SYSTEMS WITH COUNTABLY MANY ERGODIC RIGID MEASURES. PROOF OF THEOREM 1.2

Following [2] and [12], a dynamical system (X,T) is said to satisfy the strong MOMO property if for each increasing sequence (b_k) of natural numbers, $b_{k+1} - b_k \to \infty$, and each $f \in C(X)$, we have

$$(29) \qquad \frac{1}{b_K}\sum_{k<K}\left\| \sum_{b_k\le n<b_{k+1}} \mu(n)f\circ T^n \right\|_{C(X)} \to 0 \text{ when } K\to\infty.$$

Even though the strong MOMO property looks stronger than the original Möbius disjoint-ness, as proved in [2], Sarnak's conjecture is equivalent to the fact that all zero entropy systems satisfy the strong MOMO property. Moreover, the strong MOMO property implies uniform (in $x \in X$) convergence in (1).

Given $x \in X$, by $V(x) \subset M(X,T)$ we denote the set of measures for which x is quasi-generic.

In what follows we need an extension of the main result from [2].

Let $((Z_i, \mathcal{D}_i, \kappa_i, R_i))_{i \geq 1}$, be a sequence of ergodic dynamical systems (this means that we admit a repetition of the same dynamical system infinitely many times). Consider the following three conditions:

(PF1) For each $i \geq 1$, there is a topological system (Y_i, S_i) satisfying the strong MOMO property and for some $\mu_i \in M^e(Y_i, S_i)$, the measure-theoretic systems $(Z_i, \mathcal{D}_i, \kappa_i, R_i)$ and $(Y_i, \mathcal{B}(Y_i), \mu_i, S_i)$ are (measure-theoretically) isomorphic.

(PF2) For each topological system (X, T) and $x \in X$ satisfying:

- $V(x) \subset \left\{ \sum_{j \geq 1} \alpha_j \mu_j : \mu_j \in M(X,T), \alpha_j \geq 0 \text{ for } j \geq 1, \sum_{j \geq 1} \alpha_j = 1 \right\}$,
- the (measure-theoretic) systems $(X, \mathcal{B}(X), \mu_j, T)$ and $(Z_{i_j}, \mathcal{D}_{i_j}, \kappa_{i_j}, R_{i_j})$ (for some $i_j \geq 1$) are measure-theoretically isomorphic for each $j \geq 1$,

we have that the point x satisfies the μ-Sarnak property:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n \leq N} f(T^n x) \mu(n) = 0.$$

(PF3) For each topological system (Y, S) for which $M^e(Y, S) = \{\nu_j : j \geq 1\}$ and, for each $j \geq 1$, the (measure-theoretic) systems $(Y, \mathcal{B}(Y), \nu_j, S)$ and $(Z_{i_j}, \mathcal{D}_{i_j}, \kappa_{i_j}, R_{i_j})$ (for some $i_j \geq 1$) are measure-theoretically isomorphic, we have that (Y, S) satisfies the strong MOMO property.

Theorem 4.1. *Conditions* **(PF1)**, **(PF2)** *and* **(PF3)** *are equivalent.*

Theorem 4.1, which is of independent interest, will be proved in the appendix in a more general setting. It is an extension of Main Theorem in [2] in which only finitely many measures are used, all of them giving rise to the same measure theoretic dynamical system.

If R is an ergodic automorphism on a standard probability space $(Z, \mathcal{D}, \rho)$ which is rigid then it remains rigid in every of its uniquely ergodic models. Therefore, such a model is Möbius disjoint whenever we have BPV rigidity or PR rigidity (by Theorem 2.1). In fact, more is true.

Corollary 4.2. *Assume that (Y, S) is uniquely ergodic, with the unique invariant measure ν which yields either BPV rigidity or PR rigidity. Then (Y, S) satisfies the strong MOMO property.*

Proof. I. Consider first BPV rigidity. We check **(PF2)**. So, assume that (X, T) is a topological system and $x \in X$ satisfies $V(x) \subset \left\{ \sum_{j \geq 1} \alpha_j \mu_j : \alpha_j \geq 0, \sum_{j \geq 1} \alpha_j = 1 \right\}$, where

μ_j yields a system measure-theoretically isomorphic to $(Y, \mathcal{B}(Y), \nu, S)$ (we use Theorem 4.1 with the constant sequence equal to $(Y, \mathcal{B}(Y), \nu, S)$). Since all measures μ_j yield the same (up to isomorphism) system, all the measures $\sum_{j \geq 1} \alpha_j \mu_j$ yield systems having a common rigidity sequence. Hence, x is completely BPV rigid. It follows from Theorem 2.1 that x satisfies the μ-Sarnak property and **(PF2)** holds. Hence **(PF3)** holds which completes the proof.

II. Assume that we have PR rigidity. Taking into account the definition of strong MOMO property, we need to estimate $\frac{1}{b_K} \sum_{k<K} \left\| \sum_{b_k \leq m < b_{k+1}} \mu(m) f \circ T^m \right\|_{C(X)}$. We essentially repeat the proof of Theorem 2.1. In view of Remark 3.10, for $M \geq M_0$,

$$\frac{1}{M} \sum_{m \leq M} \sum_{j=-L_n}^{L_n} \| f \circ T^{jq_n+m} - f \circ T^m \|_{C(X)} < \varepsilon.$$

Assume for a while that the numbers b_k are all multiples of $L_n q_n$. Then the proof of Theorem 2.1 readily repeats (with $M = b_K$). If not, remembering that $b_{k+1} - b_k \to \infty$, we can choose a sequence $\{b'_k\}$, so that: b'_k are multiples of $L_n q_n$ (for k sufficiently large) and $|b_k - b'_k| \leq L_n q_n$. Now,

$$\left| \sum_{k<K} \left\| \sum_{b_k \leq m < b_{k+1}} \mu(m) f \circ T^m \right\|_{C(X)} - \sum_{k<K} \left\| \sum_{b'_k \leq m < b'_{k+1}} \mu(m) f \circ T^m \right\|_{C(X)} \right| =$$

$$o(b_K) + 2\|f\|_{C(X)} K L_n q_n,$$

which is sufficient for our proof. $\quad\square$

In fact, still more is true (we stress that the automorphisms in the sequence $(Z_i, \mathcal{D}_i, \kappa_i, R_i)$ below need not have a common rigid sequence).

Corollary 4.3. *Assume that $((Z_i, \mathcal{D}_i, \kappa_i, R_i))_{i \geq 1}$ is a sequence of ergodic automorphisms each of which is BPV or PR rigid. Assume that we have a topological system (Y, S) for which $M^e(Y, S) = \{\nu_j : j \geq 1\}$ and, for each $j \geq 1$, the (measure-theoretic) systems $(Y, \mathcal{B}(Y), \nu_j, S)$ and $(Z_{i_j}, \mathcal{D}_{i_j}, \kappa_{i_j}, R_{i_j})$ (for some $i_j \geq 1$) are measure-theoretically isomorphic. Then (Y, S) satisfies the strong MOMO property.*

Proof. It follows from Corollary 4.2 that **(PF1)** is satisfied. By Theorem 4.1, **(PF3)** holds, so the result follows. $\quad\square$

Proof of Theorem 1.2. The result follows directly from Corollary 4.3.

5. Appendix: Proof of Theorem 4.1

The proof of Theorem 4.1 does not use any special property of μ except of boundedness. Therefore, in what follows we consider the notion of strong $\boldsymbol{u}$-OMO and Theorem 4.1 in which the Möbius function μ has been replaced by $\boldsymbol{u} : \mathbb{N} \to \mathbb{C}$ any bounded arithmetic function.

In order to prove the equivalence of conditions **(PF1)-(PF3)**, we need the following two auxiliary lemmas. The proofs follow the same lines as the proof of Lemma 17 in [2] and the arguments just before this lemma.

Lemma 5.1. *Let (X,T) be a topological system and $\nu \in M(X,T)$. Suppose that $C \subset X$ is a compact subset such that $\nu(C) > 1 - \varepsilon^2$ for some $0 < \varepsilon < 1$. Then, for every $L \geq 1$, the set*

$$B_L(C, \varepsilon) := \Big\{ x \in C : \frac{1}{L} \sum_{l < L} \chi_C(T^l x) > 1 - \varepsilon \Big\}$$

is compact and $\nu(B_L(C, \varepsilon)) > 1 - \varepsilon$.

Lemma 5.2. *Let (X,T) be a topological system. Let $x \in X$ and $C \subset X$ be a compact subset. Suppose that $\nu \in M(X,T)$ and $(N_i)_{i \geq 1}$ is an increasing sequence of natural numbers such that*

$$\frac{1}{N_i} \sum_{n < N_i} \delta_{T^n x} \to \nu$$

weakly. Then, for every $\eta > 0$, we have

$$\limsup_{i \to \infty} \frac{1}{N_i} \#\big\{ 0 \leq n < N : d(T^n x, C) \geq \eta \big\} \leq \nu(X \setminus C).$$

Proof. of Theorem 4.1 **(PF3)** $\Rightarrow$ **(PF1)** To obtain **(PF1)**, for each $i \geq 1$, using Jewett-Krieger theorem, choose (Y_i, S_i) being a uniquely ergodic model of $(Z_i, \mathcal{D}_i, \kappa_i, R_i)$. Then apply apply **(PF3)** for (Y_i, S_i).

(PF1) $\Rightarrow$ **(PF2)** (is a modification of the proof from [2]). Suppose contrary to our claim that there exist a topological system (X,T), $x \in X$, a continuous function $f : X \to \mathbb{C}$ with $\|f\|_\infty \leq 1$ and $0 < \varepsilon < 1/2$ such that:

(i) $V(x) \subset \Big\{ \sum_{j \geq 1} \alpha_j \mu_j : \mu_j \in M(X,T), \alpha_j \geq 0 \text{ for } j \geq 1, \sum_{j \geq 1} \alpha_j = 1 \Big\}$;

(ii) for each $j \geq 1$ there exists $i_j \geq 1$ such that the measure-theoretic systems $(X, \mathcal{B}(X), \nu_j, T)$ and $(Z_{i_j}, \mathcal{D}_{i_j}, \kappa_{i_j}, R_{i_j})$ are measure-theoretically isomorphic;

(iii) $\limsup_{N \to \infty} \Big| \frac{1}{N} \sum_{n \leq N} f(T^n x) \boldsymbol{u}(n) \Big| > 7\varepsilon > 0$.

Therefore, we can find an increasing sequence $(N_i)_{i \geq 1}$ such that

$$\frac{1}{N_i} \sum_{n < N_i} \delta_{T^n x} \to \nu \in V(x) \quad \text{weakly as} \quad i \to \infty;$$

$$(30) \qquad \Big| \frac{1}{N_i} \sum_{n \leq N_i} f(T^n x) \boldsymbol{u}(n) \Big| > 7\varepsilon \quad \text{for all} \quad i \geq 1.$$

By condition (i), $\nu = \sum_{j \geq 1} \alpha_j \nu_j$ for a sequence $(\alpha_j)_{j \geq 1}$ of positive numbers with $\sum_{j \geq 1} \alpha_j = 1$ so that $(\nu_j)_{j \geq 1}$ is a sequence of different T-invariant ergodic measures satisfying (ii). Choose a natural number t such that

$$(31) \qquad \sum_{j > t} \alpha_j < \varepsilon/2.$$

In view of (ii) and **(PF1)**, for every $1 \leq j \leq t$, there exists a topological system (Y_j, S_j) satisfying the strong $\boldsymbol{u}$-OMO property and an invariant ergodic measure $\mu_j \in M^e(Y_j, S_j)$ so that the measure-theoretic systems $(X, \mathcal{B}(X), \nu_j, T)$ and $(Y_j, \mathcal{B}(Y_j), \mu_j, S_j)$ are measure-theoretically isomorphic.

Since ν_j for $1 \leq j \leq t$ are different ergodic T-invariant measures, we can find T-invariant Borel subsets $X_j \subset X$ for $1 \leq j \leq t$ with $\nu_j(X_j) = 1$, and measure-theoretic isomorphisms $\phi_j : (X_j, \mathcal{B}(X_j), \nu_j, T) \to (Y_j, \mathcal{B}(Y_j), \mu_j, S_j)$. By Lusin's theorem, for every $1 \leq j \leq t$ there exists a compact subset $W_j \subset X_j$ such that $\nu_j(W_j) > 1 - \varepsilon^4$ and the restriction $\phi_j : W_j \to \phi_j(W_j)$ is a homeomorphism.

In view of Lemma 5.1, for every $L \geq 1$ and $1 \leq j \leq t$ the set $B_L(W_j, \varepsilon^2)$ is compact and

$$\nu_j(B_L(W_j, \varepsilon^2)) > 1 - \varepsilon^2 \geq 1 - \varepsilon/2.$$

Let $B_L := \bigcup_{1 \leq j \leq t} B_L(W_j, \varepsilon^2)$. Then B_L is a compact subset of X such that $\nu_j(B_L) > 1 - \varepsilon/2$ for every $1 \leq j \leq t$. Since $\nu = \sum_{j \geq 1} \alpha_j \nu_j$ and $\sum_{j > t} \alpha_k < \varepsilon/2$ we have

$$\nu(B_L) > 1 - \varepsilon.$$

Therefore, by Lemma 5.2, for every $\eta > 0$ and $L \geq 1$, we have

$$(32) \qquad \limsup_{i \to \infty} \frac{1}{N_i} \# \big\{ 0 \leq n < N_i : d(T^n x, B_L) \geq \eta \big\} < \varepsilon.$$

For every $L \geq 1$ let $\eta(L)$ be a positive number such that

$$d(w, w') < \eta(L) \Rightarrow \forall_{0 \leq l < L} \ \ d(T^l w, T^l w') < \varepsilon.$$

Fix an increasing sequence $(L_i)_{i \geq 1}$ of natural numbers. In view of (32), there exists $(M_i)_{i \geq 1}$ a subsequence of $(N_i)_{i \geq 1}$ such that $M_i - M_{i-1} \to +\infty$,

$$(33) \qquad L_i < \varepsilon M_i \quad \text{for every} \quad i \geq 1,$$

and, setting $M_0 = 0$, we have

$$(34) \qquad \frac{1}{M_i - M_{i-1}} \# \big\{ M_{i-1} \leq n < M_i : d(T^n x, B_{L_i}) \geq \eta(L_i) \big\} < \varepsilon \text{ for } i \geq 1.$$

A natural number b is said to be *good* if, for some $i \leq 1$, we have $M_{i-1} \leq b < M_i$ and $d(T^b x, B_{L_i}) < \eta(L_i)$.

Let us define the increasing sequence $(b_i)_{i \geq 1}$ of good numbers inductively in the following way: set $b_0 = 0$ and let $b_1 \geq 1$ be the smallest good number. Assume that $M_{i-1} \leq b_k < M_i$ is defined for some $k \geq 1$. Then we define b_{k+1} as the smallest good number $b \geq b_k + L_i$. As $b_{k+1} - b_k \geq L_i$ and $L_i \to +\infty$, we have $b_{k+1} - b_k \to +\infty$.

For every $1 \leq j \leq t$ the function $f \circ \phi_j^{-1}$ is continuous on the compact set $\phi_j(W_j) \subset Y_j$. Then, by Tietze extension theorem, there is a continuous function $g_j : Y_j \to \mathbb{C}$ such that $\|g_j\|_\infty \leq \|f\|_\infty \leq 1$ and $g_j = f \circ \phi_j^{-1}$ on W_j. Since ϕ_j establishes a measure-theoretical isomorphism between $(X, \mathcal{B}(X), \nu_j, T)$ and $(Y_j, \mathcal{B}(Y_j), \mu_j, S_j)$, we have

$$(35) \qquad w \in W_j \cap T^{-l} W_j \ \Rightarrow \ f(T^l w) = g_j(\phi_j(T^l w)) = g_j(S_j^l \phi_j(w)).$$

Since b_k is good, there exist $1 \leq j_k \leq t$ and $x_k \in B_{L_i}(W_{j_k}, \varepsilon^2)$ such that $d(T^{b_k}x, x_k) < \eta(L_i)$. It follows that

$$|f(T^{b_k+l}x) - f(T^l x_k)| < \varepsilon \quad \text{for} \quad 1 \leq l < L_i.$$

Since $x_k \in B_{L_i}(W_{j_k}, \varepsilon^2)$, the average frequency of the orbit $\{T^l x_k : 0 \leq l < L_i\}$ in W_{j_k} is at least $1 - \varepsilon^2$. Therefore, by setting $y_k := \phi(x_k) \in Y_{j_k}$, we obtain

$$\sum_{l < L_i} |f(T^{b_k+l}x) - g_{j_k}(S^l_{j_k} y_k)| \leq 2\varepsilon^2 L_i \|f\|_\infty + \varepsilon L_i.$$

In view of (34), it follows that

$$\sum_{b_k \leq s < b_{k+1}} |f(T^s x) - g_{j_k}(S^s_{j_k}(S^{-b_k}_{j_k} y_k))|$$

$$\leq \varepsilon(b_{k+1} - b_k)(\|f\|_\infty + 1) + 2\|f\|_\infty \#\{b_k \leq s < b_{k+1}; s \text{ is not good}\}$$

$$\leq 4\varepsilon(b_{k+1} - b_k).$$

Hence, for every $K \geq 1$, we have

$$\frac{1}{b_K} \sum_{s < b_K} |f(T^s x) - g_{j_k}(S^s_{j_k}(S^{-b_k}_{j_k} y_k))| \leq 4\varepsilon.$$

By the strong $\boldsymbol{u}$-OMO property of the topological systems (Y_j, S_j) for $1 \leq j \leq t$, we have

$$\lim_{K \to \infty} \frac{1}{b_K} \sum_{k < K} \left| \sum_{b_k \leq s < b_{k+1}} g_{j_k}(S^s_{j_k}(S^{-b_k}_{j_k} y_k))\boldsymbol{u}(s) \right| = 0.$$

It follows that

$$\limsup_{K \to \infty} \frac{1}{b_K} \left| \sum_{s < b_K} f(T^s x)\boldsymbol{u}(s) \right| \leq 4\varepsilon.$$

For every $i \geq 1$ denote by K_i the largest K such that $b_K \leq M_i$. Then

$$M_i - b_{K_i} \leq L_i + \#\{n < M_i : n \text{ is not good}\} < 2\varepsilon M_i$$

and

$$\frac{1}{M_i} \left| \sum_{s < M_i} f(T^s x)\boldsymbol{u}(s) \right| \leq \frac{1}{b_{K_i}} \left| \sum_{s < b_{K_i}} f(T^s x)\boldsymbol{u}(s) \right| + \frac{M_i - b_{K_i}}{M_i}$$

$$\leq \frac{1}{b_i K} \left| \sum_{s < b_{K_i}} f(T^s x)\boldsymbol{u}(s) \right| + 2\varepsilon.$$

Hence

$$\limsup_{i \to \infty} \frac{1}{M_i} \left| \sum_{s < M_i} f(T^s x)\boldsymbol{u}(s) \right| \leq 6\varepsilon,$$

contrary to (30) (as (M_i) is a subsequence of (N_i)).

(PF2) $\Rightarrow$ **(PF3)** We fix $f \in C(Y)$, (b_k) satisfying $b_{k+1} - b_k \to \infty$ and $(y_k) \subset Y$. Let Σ_3 be the group of roots of unity of degree 3. Choose $e_k \in \Sigma_3$ so that

$$e_k \left(\sum_{b_k \le n < b_{k+1}} f(S^n(S^{-b_k} y_k)) \boldsymbol{u}(n) \right)$$

is in the cone $\{0\} \cup \{z \in \mathbb{C} : \arg(z) \in [-\pi/3, \pi/3]\}$. Set

$$x = (x_n) \in (Y \times \Sigma_3)^{\mathbb{N}}, \text{ where } x_n = (S^{n-b_k} y_k, e_k) \text{ for } b_k \le n < b_{k+1}$$

and

$$X_x := \overline{\{T^n x : n \ge 0\}},$$

where T stands for the shift on $(Y \times \Sigma_3)^{\mathbb{N}}$. We need to determine all ergodic (T-invariant) measures on X_x. Recall that each such measure has a generic point. On the other hand, the analysis done in the proof of Corollary 9 [2], shows that if $z \in X_x$ is a quasi-generic point for a measure $\rho \in M(X_x, T)$, then the projection on the first coordinate in $X_x \subset (Y \times \Sigma_3)^{\mathbb{N}}$-coordinate ρ-a.e. intertwines T and $S \times Id$. It follows that if ρ is ergodic, then the projection $\rho^{(1)}$ of ρ is an $S \times Id$-invariant measure which is ergodic (for this map). By the classical disjointness result *ergodicity* $\perp$ *identity*, it follows that $\rho^{(1)} = \kappa \otimes \delta_a$, where $\kappa \in M^e(Y, S)$ and $a \in \Sigma_3$. Furthermore, ρ is the image of $\rho^{(1)}$ via the map

$$(Id \times Id) \times (S \times Id) \times (S^2 \times Id) \times \ldots,$$

so $M^e(X_x, T)$ is countable and each ergodic measure yields a system which is measure-theoretically isomorphic to a system $(Y, \mathcal{B}(Y), \nu, S)$ for some $\nu \in M^e(Y, S)$. We now apply **(PF2)** to x and the function

$$((v_0, a_0), (v_1, a_1), \ldots) \mapsto a_0 f(v_0)$$

in exactly the same manner as in [2] to conclude. $\square$

REFERENCES

[1] H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, T. de la Rue, *The Chowla and the Sarnak conjectures from ergodic theory point of view*, Discrete and Continuous Dynamical Systems **37** (2017), 2899-2944.

[2] H. El Abdalaoui, J. Kułaga-Przymus, M. Lemańczyk, T. de la Rue, *Möbius disjointness for models of an ergodic system and beyond*, Israel J. Math. **228** (2018), 707-751.

[3] J. Bourgain, *On the correlation of the Möbius function with rank-one systems*, J. Anal. Math. **120** (2013), 105-130.

[4] J. Bourgain, P. Sarnak, T. Ziegler, *Disjointness of Möbius from horocycle flows, From Fourier analysis and number theory to Radon transforms and geometry*, Dev. Math., vol. 28, Springer, New York, 2013, pp. 67-83.

[5] J. Chaika, *Every ergodic transformation is disjoint from almost every interval exchange transformation*, Ann. of Math, 175 (2012)(1), pp. 237-253.

[6] J. Chaika, A. Eskin, *Möbius disjointness for intervl exchange transformations on three intervals*, to appear in J. Modern Dynam. arXiv 1606.02357v2.

[7] S. Donoso, S. Song, *Uniformly rigid models for rigid actions*, Studia Math. **236** (2017), 13-31.

[8] S. Ferenczi, C. Holton, L. Zamboni, *Joinings of three-interval exchange transformations*, Ergodic Theory Dynam. Systems **25** (2005), 483-502.

[9] S. Ferenczi, J. Kułaga-Przymus, M. Lemańczyk, *Sarnak's Conjecture – what's new*, in: Ergodic Theory and Dynamical Systems in their Interactions with Arithmetics and Combinatorics, CIRM Jean-Morlet Chair, Fall 2016, Editors: S. Ferenczi, J. Kułaga-Przymus, M. Lemańczyk, Lecture Notes in Mathematics 2213, Springer International Publishing, pp. 418.

[10] S. Ferenczi, C. Mauduit, *On Sarnak's conjecture and Veech's question for interval exchanges*, J. Analyse Math. **134** (2018), 545-573.

[11] N. Frantzikinakis, B. Host, *The logarithmic Sarnak conjecture for ergodic weights*, Annals Math. **187** (2018), 869-931.

[12] A. Gomilko, M. Lemańczyk, T. de la Rue, *Möbius orthogonality in density for zero entropy dynamical systems*, arXiv:1905.06563.

[13] E. Glasner, D. Maon, *Rigidity in topological dynamics*, Ergodic Theory Dynam. Systems **9** (1989), 309-320.

[14] M. Herman, *Sur la conjugaison différentiabledes difféomorphismes du cercle à des rotations*, Publ. IHES **49** (1979), 5-234.

[15] W. Huang, Z. Wang, and G. Zhang, *Möbius disjointness for topological models of ergodic systems with discrete spectrum*, to appear in J. Modern Dynamics.

[16] W. Huang, Z. Wang, and X. Ye, *Measure complexity and Möbius disjointness*, to appear in Advances Math., arxiv 1707.06345.

[17] W. Huang, L. Xu, *Special Flow, Weak Mixing and Complexity*, Communications in Math. and Statistics **7** (2019), 85-122.

[18] T. Kamae, *Subsequences of normal sequences*, Israel J. Math. **16** (1973), 121-149.

[19] A. Kanigowski, M. Lemańczyk, C. Ulcigrai, *On disjointness of some parabolic flows*, arXiv:1810.11576.

[20] D. Karagulyan, *Hausdorff dimension of a class of three interval-exchange maps*, arXiv:1701.01984v2

[21] I. Kátai, *A remark on a theorem of H. Daboussi*, Acta Math. Hungar. **47**(1-2) (1986), 223-225.

[22] A. Katok, *Interval exchange transformations and some special flows are not mixing*, Israel J. Math. 35 (1980), 301-310.

[23] O. Klurman, A. Mangerel, J. Teräväinen, *Multiplicative functions in short arithmetic progressions*, pre-print

[24] J. Kułaga-Przymus, M. Lemańczyk, *Möbius disjointness along ergodic sequences for uniquely ergodic actions*, to appear in Ergodic Theory Dynam. Systems published online 2018, arXiv: 1703.02347.

[25] P. Kunde, *Uniform rigidity sequences for weakly mixing diffeomorphisms on $\mathbb{D}^m$, $\mathbb{T}^m$ and $\mathbb{S}^1 \times [0,1]^{m-1}$*, J. Math. Anal. Appl. **462** (2018), 1398-1424.

[26] J. Liu, P. Sarnak, *The Möbius function and distal flows*, Duke Math. J. **164** (2015), 1353 - 1399.

[27] K. Matomäki, M. Radziwiłł, *Multiplicative functions in short intervals*, Annals Math. (2) **183** (2016), 1015-1056.

[28] M. Matomäki, M. Radziwiłł, and T. Tao, *An averaged form of Chowla's conjecture*, Algebra & Number Theory **9** (2015), 2167-2196.

[29] H. Montgomery, *Topics in multiplicative number theory*, Lecture Notes in Mathematics **227**, 1971, Springer-Verlag

[30] P. Sarnak, *Three lectures on the Möbius function, randomness and dynamics*, http://publications.ias.edu/sarnak/.

[31] T. Tao, J. Teräväinen, *The structure of correlations of multiplicative functions at almost all scales, with applications to the Chowla and Elliott conjectures*, Algebra and Number Theory, to appear

[32] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge Studies in Advanced Mathematics **46**, 1995, Cambridge University Press

[33] W.A. Veech, *The metric theory of interval exchange transformations.I. Generic spectral properties*, Amer. J. Math. **106** (1984), 1331-1359.

[34] Z. Wang, *Möbius disjointness for analytic skew products*, Inventiones Math. **209** (2017), 175-196.

[35] B. Weiss, *Normal sequences as collectives*, Proc. Symp. on Topological Dynamics and ergodic theory, Univ. of Kentucky, 1971.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND AT COLLEGE PARK, COLLEGE PARK, MD 20740, USA, E-MAIL ADDRESS: ADKANIGOWSKI@GMAIL.COM

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE, NICOLAUS COPERNICUS UNIVERSITY, CHOPIN STREET 12/18, 87-100 TORUŃ, POLAND, E-MAIL ADDRESS: MLEM@MAT.UMK.PL

DEPARTMENT OF MATHEMATICS, CALTECH, 1200 E CALIFORNIA BLVD, PASADENA, CA 91125, USA, E-MAIL ADDRESS: MAKSYM.RADZIWILL@GMAIL.COM