J. Eur. Math. Soc. 23, 2707–2731 (2021) DOI 10.4171/JEMS/1069

Jon Chaika · Alex Eskin

Self-joinings for 3-IETs

Received August 21, 2018 and in revised form July 30, 2019

Abstract. We show that typical interval exchange transformations on three intervals are not 2-simple answering a question of Veech. Moreover, the set of self-joinings of almost every 3-IET is a Poulsen simplex.

Keywords. Simple, interval exchange, joining

1. Introduction

Definition 1.1. Let $(X, \mathcal{B}, \mu, T)$ be a probability measure preserving system. A *self-joining* is a $T \times T$ invariant measure on $X \times X$ with marginals μ.

Definition 1.2. $(X, \mathcal{B}, \mu, T)$ is called 2-*simple* if every ergodic self-joining, other than $\mu \times \mu$, is one-to-one on almost every fiber.

Definition 1.3. A *Poulsen simplex* is a metrizable simplex whose extreme points are dense.

Lindenstrauss, Olsen and Sternfeld proved that a Poulsen simplex is unique up to affine homeomorphism [11].

Definition 1.4. A 3-*interval exchange transformation* is defined by three non-negative numbers ℓ_1, ℓ_2, ℓ_3. It is $T : [0, \ell_1 + \ell_2 + \ell_3) \to [0, \ell_1 + \ell_2 + \ell_3)$ given by

$$T(x) = \begin{cases} x + \ell_2 + \ell_3 & \text{if } x < \ell_1, \\ x + \ell_3 - \ell_1 & \text{if } \ell_1 \leq x < \ell_1 + \ell_2, \\ x - (\ell_1 + \ell_2) & \text{otherwise.} \end{cases}$$

Theorem 1.5. *Almost every* 3*-IET is not* 2*-simple. Also, its self-joinings form a Poulsen simplex.*

Note that $T \times T$ has topological entropy 0.

J. Chaika: Department of Mathematics, University of Utah,
Salt Lake City, UT 84112, USA; e-mail: chaika@math.utah.edu

A. Eskin: Department of Mathematics, University of Chicago,
Chicago, IL 60637, USA; e-mail: eskin@math.uchicago.edu

Mathematics Subject Classification (2020): Primary 37E05

The first part of Theorem 1.5 answers a question of Veech in the negative [16, Question 4.9]. (In [16] "2-simple" is called "Property S.")

Recall that a measure preserving system is called *prime* if it has no non-trivial factors. In the paper [16] mentioned above, Veech classified the factors of 2-simple systems, and so a natural question remains:

Question 1.6. Is almost every 3-IET prime?

It is also natural to wonder what happens for IETs with other permutations and flows on translation surfaces. It is likely that our techniques can show that residual sets of interval exchange transformations on more intervals, and flows on translation surfaces of genus greater than 1 are not simple, but we do not see how they can be applied to almost every flow on translation surface or IET with different permutation.

To prove Theorem 1.5 we define in Section 2 a distiguished class of self-joinings called "shifted power joinings." In Section 2 we also show that a special type of transformations called "rigid rank 1 by intervals" (which includes IET's by [17, Part 1, Theorem 1.4]) have the property that linear combinations of shifted power joinings are dense in their self-joinings. M. Lemańczyk brought to our attention that this result was proved in an unpublished paper of J. King [10]. We then prove that almost every 3-IET has the property that its ergodic self-joinings are dense in linear combinations of the shifted power joinings. We do this by producing an abstract criterion (Section 3) and showing 3-IETs satisfy this criterion (Section 4).

Context of our results. Before Veech's work, D. Rudolph introduced the notion of *minimal self-joinings*, using it as a fruitful class of examples, including examples of prime systems [12]. The property of 2-simple generalizes minimal self-joinings and in particular, no rigid system has minimal self-joinings. The typical IET is rigid [17, Part 1, Theorem 1.4], so the typical IET does not have minimal self-joinings, but there are rigid 2-simple systems. Ageev proved that the set of measure preserving transformations which are not 2-simple contains a dense G_δ, i.e. it is a residual set (with the topology being the so called *weak topology*) [1]. Our construction can be modified to give a new proof of this fact.

Our result that the self-joinings form a Poulsen simplex is also perhaps a little unexpected. Many examples of systems whose set of invariant measures forms a Poulsen simplex are well known, but typically these systems are of high complexity, satisfying some form of *specification*. In contrast, our examples have very low complexity, as $T \times T$ has *quadratic block growth*. Since systems of *linear block growth* have only finitely many ergodic measures [2], such a system cannot have the set of its invariant measures forming a Poulsen simplex (though as our examples show its Cartesian product could). We remark that in the previously mentioned unpublished work, J. King proved that a residual set of measure preserving transformations (which therefore must include rank 1 transformations) have the property that their self-joinings form a Poulsen simplex [10], giving many (non-explicit) entropy zero examples. Our result is perhaps still surprising, because we treat a previously considered family of examples and we show typicality in a metric, rather than topological setting.

Two key steps are showing that the typical 3-IET admit good linked $(n_j, n_j + 1)$ approximation [7] (see Remark 4.8 and the proof of Proposition 4.5) and that this implies the existence of all sorts of ergodic joinings (see Proposition 3.1). Some consequences of transformations with $(n_j, n_j + 1)$ approximation were studied by Ryzhikov [14] and as a result we get some spectral consequences for T^n and $T \times \cdots \times T$ (n times) (see Remark 4.6). Related to this, the methods of the paper show that Katok maps are not simple, answering [4, Question 7.1].

2. Joinings of rigid rank 1 transformations come from limits of linear combinations of powers

Let $([0, 1], \mathcal{M}, \lambda, T)$ be an ergodic invertible transformation.

Definition 2.1. We say T is *rigid rank* 1 *by intervals* if there exists a sequence of intervals $I_1, I_2, \ldots$ and natural numbers $n_1, n_2, \ldots$ such that:

- $T^i I_k$ is an interval with $\mathrm{diam}(T^i I_k) = \mathrm{diam}(I_k)$ for all $0 \le i < n_k$.
- $T^i I_k \cap T^j I_k = \emptyset$ for all k and $0 \le i < j < n_k$.
- $\lim_{k\to\infty} \lambda(\bigcup_{i=0}^{n_k-1} T^i I_k) = 1$.
- $\lim_{k\to\infty} \lambda(T^{n_k} I_k \triangle I_k)/\lambda(I_k) = 0$.

This is a condition meaning that our transformation is well approximated by periodic transformations. A similar condition, admitting *cyclic approximation by periodic transformations*, was considered in [8].

Let

$$\mathcal{R}_k = \bigcup_{i=0}^{n_k-1} T^i I_k, \quad \hat{\mathcal{R}}_k = \bigcup_{i=0}^{n_k-1} T^i (I_k \cap T^{-n_k} I_k \cap T^{n_k} I_k),$$
$$\tilde{\mathcal{R}}_k = \bigcup_{i=0}^{n_k-1} T^i (I_k \cap T^{-n_k} I_k \cap T^{-2n_k} I_k \cap T^{n_k} I_k \cap T^{2n_k} I_k), \tag{2.1}$$

Then $\mathcal{R}_k$ is the Rokhlin tower over I_k, $\hat{\mathcal{R}}_k$ is the Rokhlin tower over $I_k \cap T^{-n_k} I_k \cap T^{n_k} I_k$, and $\tilde{\mathcal{R}}_k$ is the Rokhlin tower over $\bigcap_{i=-2}^{2} T^{in_k} I_k$. We have

$$\hat{\mathcal{R}}_k = \{x : T^i x \in \mathcal{R}_k \text{ for all } -n_k < i < n_k\}, \tag{2.2}$$
$$\tilde{\mathcal{R}}_k = \{x : T^i x \in \mathcal{R}_k \text{ for all } -2n_k < i < 2n_k\}. \tag{2.3}$$

Heuristically one can think of $\mathcal{R}_k$ as the set of points we can control. The sets $\hat{\mathcal{R}}_k$ and $\tilde{\mathcal{R}}_k$ let us control the points for long orbit segments, which is necessary for some of our arguments.

Lemma 2.2. $\lim_{k\to\infty} \lambda(\tilde{\mathcal{R}}_k) = 1 = \lim_{k\to\infty} \lambda(\mathcal{R}_k) = \lim_{k\to\infty} \lambda(\hat{\mathcal{R}}_k)$.

Proof. By the third condition in the definition of rigid rank 1 by intervals we have $\lim_{k\to\infty} \lambda(\mathcal{R}_k) = 1$. By (2.1),

$$\lambda(\hat{\mathcal{R}}_k) \geq \lambda(\mathcal{R}_k) - n_k\lambda(I_k \setminus (T^{n_k} I_k \cup T^{-n_k} I_k)) \geq \lambda(\mathcal{R}_k) - 2n_k\lambda(I_k \setminus T^{n_k} I_k),$$

and thus by the fourth condition of that definition, $\lim_{k\to\infty} \lambda(\hat{\mathcal{R}}_k) \to 1$. Similarly, $\lim_{k\to\infty} \lambda(\tilde{R}_k) = 1$. □

Definition 2.3 (Shifted power joining). Let (X, T, μ) be a measure preserving dynamical system. A self-joining of (X, T, μ) that gives full measure to $\{(x, T^a x)\}$ for some $a \in \mathbb{Z}$ with $a \neq 0$ is called a *shifted power joining*. These have also been called *off-diagonal joinings*.

Let $\iota : [0, 1] \to [0, 1]$ by $x \mapsto (x, x)$. Let $\mu = \iota_*\lambda$. Shifted power joinings have the form $(\mathrm{id} \times T^a)_*\mu$ for some $a \in \mathbb{Z} \setminus \{0\}$.

The operator A_σ and convergence in the strong operator topology. Let σ be a self-joining of (T, λ). Let σ_x be the corresponding measure on $[0, 1]$ coming from disintegrating along σ on the fiber $\{x\} \times [0, 1]$. Define $A_\sigma : L^2(\lambda) \to L^2(\lambda)$ by $A_\sigma(f)[x] = \int f\, d\sigma_x$.

Recall that one calls the topology of pointwise convergence on $L^2(\lambda)$ the *strong operator topology*. That is, $A_1, A_2, \ldots$ converges to A_∞ in the strong operator topology if and only if $\lim_{i\to\infty} \|A_i f - A_\infty f\|_2 = 0$ for all $f \in L^2(\lambda)$.

Theorem 2.4. *Assume* $([0, 1], T, \lambda)$ *is rigid rank* 1 *by intervals and* σ *is a self-joining of* $([0, 1], T, \lambda)$. *Then* A_σ *is the strong operator topology* (*SOT*) *limit of linear combinations, with non-negative coefficients, of powers of* U_T*, where* $U_T : L^2([0, 1], \lambda) \to L^2([0, 1], \lambda)$ *denotes the Koopman operator* $U_T(f) = f \circ T$.

Corollary 2.5 (J. King). *Any self-joining of a transformation rigid rank* 1 *by intervals is a weak-* limit of linear combinations of shifted power joinings.*

These results (or very closely ones) were established earlier by J. King [10] using a different proof. In fact he shows that if the joining in Corollary 2.5 is ergodic then there is no need to take a linear combination. See also [6, Theorem 7.1]. There is an open question of whether this result is true for general rank 1 systems [9, p. 382]. Ryzhikov has a series of results in this direction: see for example [13] and [15].

2.1. *Proof of Theorem 2.4*

Lemma 2.6. *For each* $0 \leq j < n_k$ *we have*

$$n_k \int_{T^j I_k} \sigma_x(\mathcal{R}_k^c)\, d\lambda(x) \leq \lambda(\tilde{\mathcal{R}}_k^c). \tag{2.4}$$

Remark. Note that n_k is roughly $\lambda(T^j I_k)^{-1}$.

Proof of Lemma 2.6. Suppose $0 \le j < n_k$, and suppose $x \in T^j I_k$. From (2.3) we have $T^i \mathcal{R}_k^c \subset \tilde{\mathcal{R}}_k^c$ for all $-n_k < i < n_k$. We claim that

$$\sigma_x(\mathcal{R}_k^c) \le \sigma_{T^\ell x}(\tilde{\mathcal{R}}_k^c) \quad \text{for all } -n_k < \ell < n_k. \tag{2.5}$$

Indeed, $\sigma_x(\mathcal{R}_k^c) = \sigma_{T^\ell x}(T^\ell \mathcal{R}_k^c) \le \sigma_{T^\ell x}(\tilde{\mathcal{R}}_k^c)$. Integrating (2.5) we get

$$\int_{T^j I_k} \sigma_y(\mathcal{R}_k^c)\, d\lambda(y) \le \int_{T^{j+\ell} I_k} \sigma_z(\tilde{\mathcal{R}}_k^c)\, d\lambda(z) \quad \text{for all } -n_k < \ell < n_k. \tag{2.6}$$

Since we can choose ℓ in (2.6) so that $j + \ell$ takes any value in $[0, n_k - 1] \cap \mathbb{Z}$, we get

$$\int_{T^j I_k} \sigma_y(\mathcal{R}_k^c)\, d\lambda(y) \le \min_{0 \le i < n_k} \int_{T^i I_k} \sigma_z(\tilde{\mathcal{R}}_k^c)\, d\lambda(z). \tag{2.7}$$

Now

$$\sum_{i=0}^{n_k-1} \int_{T^i I_k} \sigma_y(\tilde{\mathcal{R}}_k^c)\, d\lambda(y) \le \int_{[0,1]} \sigma_y(\tilde{\mathcal{R}}_k^c)\, d\lambda(y) \le \lambda(\tilde{\mathcal{R}}_k^c),$$

where the last estimate uses the fact that σ has projections λ. So we obtain

$$\min_{0 \le i < n_k} \int_{T^i I_k} \sigma_x(\tilde{\mathcal{R}}_k^c)\, d\lambda(x) \le \frac{1}{n_k} \lambda(\tilde{\mathcal{R}}_k^c). \tag{2.8}$$

Now the estimate (2.4) follows from (2.7) and (2.8). □

We want to guess coefficients c_j such that σ is close to $\sum_{j=0}^{n_k-1} c_j (\mathrm{id} \times T^i)_* \mu$. The next lemma comes up with a candidate pointwise version. Theorem 2.4 and Corollary 2.5 follow because by Egorov's theorem this choice is almost constant on most of the $T^\ell I_k$ and the subsequent lemma (Lemma 2.8) shows that they are almost T-invariant.

Lemma 2.7. *Let $x \in \hat{\mathcal{R}}_k \cap T^j I_k$ where $0 \le j < n_k$. Define $c_i(x) = \sigma_x(T^a I_k \cap \hat{\mathcal{R}}_k)$ where $0 \le a < n_k$ and $i + j \equiv a \pmod{n_k}$. For all 1-Lipschitz f we have*

$$\left| A_\sigma f(x) - \sum_{i=0}^{n_k-1} c_i(x) f(T^i x) \right| \le \mathrm{diam}(I_k) + 2\|f\|_{\sup} \sigma_x(\hat{\mathcal{R}}_k^c).$$

Morally $c_j(x)$ is the σ_x measure of the level in R_k that is j levels above the level x is on. Because $j + \ell$ can be bigger than n_k the definition is slightly more complicated. Note that the $c_j(x)$ are non-negative.

Proof of Lemma 2.7. Suppose $x \in \hat{\mathcal{R}}_k \cap T^j I_k$. First notice that if $y, z \in T^i I_k$ for some $0 \le i < n_k$ then $d(y, z) \le \mathrm{diam}(I_k)$. So if $j + \ell < n_k$ we have

$$\left| \int_{\hat{R}_k \cap T^{j+\ell} I_k} f\, d\sigma_x - c_j(x) f(T^i x) \right| \le \|f\|_{\mathrm{Lip}}\, \mathrm{diam}(I_k). \tag{2.9}$$

If $j+\ell \geq n_k$ then $|c_j(x) - \sigma_x(\hat{R}_k \cap T^{j+\ell} I_k)| \leq \sigma_x(\tilde{R}_k^c \cap T^{j+\ell} I_k)$ because if $y \in \tilde{R}_k$ then $T^{\pm n_k} y \in \hat{R}_k$. So for any j we have

$$\left| \int_{\hat{R}_k \cap T^{j+\ell} I_k} f\, d\sigma_x - c_j(x) f(T^j x) \right| \leq \|f\|_{\mathrm{Lip}} \operatorname{diam}(I_k) + \|f\|_{\sup} \sigma_x(\tilde{R}_k^c \cap T^{j+\ell} I_k). \tag{2.10}$$

By (2.2), for all $0 \leq \ell < n_k$, $T^{-\ell} \hat{R}_k \subset R_k$. Therefore, $\hat{R}_k \subset \bigcup_{i=\ell}^{\ell+n_k-1} T^i I_k$ for all $0 \leq \ell < n_k$. By summing over j in (2.10) we obtain

$$\left| \int_{\hat{R}_k} f\, d\sigma_x - \sum_{j=0}^{n_k-1} c_i(x) f(T^i x) \right| \leq \|f\|_{\mathrm{Lip}} \operatorname{diam}(I_k) + \|f\|_{\sup} \sigma_x(\tilde{R}_k^c). \tag{2.11}$$

In view of the fact that

$$\left| \int_{\hat{R}_k^c} f\, d\sigma_x \right| \leq \|f\|_{\sup} \sigma_x(\tilde{R}_k^c), \tag{2.12}$$

we obtain the lemma. □

Lemma 2.8. *Suppose* $0 \leq \ell < n_k$. *If* $x \in T^\ell I_k$ *and* $-\ell \leq i < n_k - \ell$ *then*

$$\sum_{j=0}^{n_k-1} |c_j(x) - c_j(T^i x)| \leq 2\sigma_x(\tilde{R}_k^c).$$

Proof. Suppose $0 \leq \ell < n_k$, $0 \leq j < n_k$, and $-\ell \leq i < n_k - \ell$. First note that if $0 \leq m < n_k$ and $z \in T^m I_k \cap \hat{R}_k$ then by (2.1), we have $T^s z \in T^{m+s} I_k \cap \hat{R}_k$ for all $-m \leq s < n_k - m$. Thus, if $j + \ell < n_k$ and $i + j + \ell < n_k$ then

$$\sigma_{T^i x}(T^{i+j+\ell} I_k \cap \hat{R}_k) = \sigma_x(T^{j+\ell} I_k \cap T^{-i} \hat{R}_k) = \sigma_x(T^{j+\ell} I_k \cap \hat{R}_k).$$

This gives $c_j(x) = c_j(T^i x)$ if $j + \ell < n_k$ and $i + j + \ell < n_k$. By a similar reasoning we have $c_j(x) = c_j(T^i x)$ if $j + \ell \geq n_k$ and $i + j + \ell \geq n_k$.

Now assume that $j + \ell < n_k$ and $i + j + \ell \geq n_k$. Then

$$c_j(T^i x) = \sigma_{T^i x}(T^{i+j+\ell-n_k} I_k \cap \hat{R}_k) = \sigma_x(T^{j+\ell-n_k} I_k \cap T^{-i} \hat{R}_k). \tag{2.13}$$

Also,

$$c_j(x) = \sigma_x(T^{j+\ell} I_k \cap \hat{R}_k). \tag{2.14}$$

Now because $\tilde{R}_k \subset \bigcap_{i=-n_k}^{n_k} T^i \hat{R}_k$, if $z \in T^{i+j+\ell-n_k} I_k \cap \tilde{R}_k$, then $z \in T^{j+\ell-n_k} I_k \cap T^{-i} \hat{R}_k$ and $z \in T^{j+\ell} I_k \cap \hat{R}_k$. Therefore, the symmetric difference between $T^{j+\ell-n_k} I_k \cap T^{-i} \hat{R}_k$ and $T^{j+\ell} I_k \cap \hat{R}_k$ is contained in the union of $T^{i+j+\ell-n_k} I_k \cap \tilde{R}_k^c$ and $T^{j+\ell} I_k \cap \tilde{R}_k^c$. Thus, in view of (2.13) and (2.14),

$$|c_j(x) - c_j(T^i x)| \leq \sigma_x(T^{j+\ell+i-n_k} \tilde{R}_k^c) + \sigma_x(T^{j+\ell} \tilde{R}_k^c).$$

The last case, where $j + \ell \geq n_k$ and $0 \leq i + j + \ell < n_k$, gives analogous bounds. So we bound $\sum_{i=0}^{n_k-1} |c_j(x) - c_j(T^i x)|$ by $2 \sum_{i=0}^{n_k-1} \lambda(T^i I_k \cap \tilde{R}_k^c) \leq 2\lambda(\tilde{R}_k^c)$ and obtain the lemma. □

Let d_{KR} denote the Kantorovich–Rubinstein metric on measures,

$$d_{\mathrm{KR}}(\mu, \nu) = \sup\left\{\left|\int f\,d\mu - \int f\,d\nu\right| : f \text{ is 1-Lipschitz}\right\}.$$

The next lemma is an immediate consequence of this definition.

Lemma 2.9. *If f is 1-Lipschitz and $d_{\mathrm{KR}}(\sigma_x, \sigma_y) < \epsilon$ then $|A_\sigma f(x) - A_\sigma f(y)| < \epsilon$.*

We say $0 \le j < n_k$ is *k-good* if there exists y_j in $T^j I_k$ such that at least $1-\epsilon$ proportion of the points in $T^j I_k$ have their disintegration ϵ close to y_j, that is

$$\lambda(\{x \in T^j I_k : d_{\mathrm{KR}}(\sigma_x, \sigma_{y_j}) < \epsilon\}) \ge (1-\epsilon)\lambda(I_k).$$

Lemma 2.10. *For all $\epsilon > 0$ there exists k_0 such that for all $k > k_0$ we have*

$$|\{0 \le j < n_k : j \textit{ is k-good }\}| > (1-\epsilon)n_k.$$

Proof. By Luzin's Theorem there exists a compact set K of measure at least $1-\epsilon^2/4$ such that the map $y \mapsto \sigma_y$ is continuous with respect to the usual metric on $[0,1]$ and the metric d_{KR} on measures. Because K is compact, this map is uniformly continuous and so there exists $\delta > 0$ such that if $x, y \in K$ and $|x-y| < \delta$ then $d_{\mathrm{KR}}(\sigma_x, \sigma_y) < \epsilon$. We choose k so that $\mathrm{diam}(I_k) < \delta$ and $\lambda([0,1] \setminus \mathcal{R}_k) < \epsilon^2/4$. Let

$$\eta = \frac{1}{n_k}|\{0 \le j < n_k : \lambda(T^j I_k \cap K^c) > \epsilon\lambda(I_k)\}|.$$

Then, because the $T^j I_k$ are disjoint and of equal size and $\bigcup_{j=0}^{n_k-1} T^j I_k = \mathcal{R}_k$, it is clear that

$$\eta\epsilon \le \frac{\lambda(K^c \cap \mathcal{R}_k)}{\lambda(\mathcal{R}_k)} \le \frac{\epsilon^2/4}{1-\epsilon^2/4} < \frac{\epsilon^2}{2},$$

and thus $\eta < \epsilon/2$. This completes the proof of the lemma. □

Notation. If j is k-good let

$$G_j = \left\{x \in T^j I_k : \lambda(\{y \in T^j I_k : d_{\mathrm{KR}}(\sigma_x, \sigma_y) < 2\epsilon\}) > (1-\epsilon)\lambda(I_k)\right\},$$

i.e. G_j is the set of points that are almost continuity points of the map $x \mapsto \sigma_x$ (restricted to $T^j I_k$). We set $G_j = \emptyset$ if j is not k-good.

Lemma 2.11. *For all $\epsilon > 0$ there exists k_1 such that for all $k > k_1$ there exist $0 \le \ell < n_k$ and $y_k \in T^\ell I_k \cap \hat{\mathcal{R}}_k$ such that $\sigma_{y_k}(\tilde{\mathcal{R}}_k^c) < \epsilon$ and*

$$|\{-\ell \le j < n_k - \ell : T^j y_k \in G_{\ell+j} \textit{ and } j \textit{ is k-good}\}| > (1-12\epsilon)n_k. \tag{2.15}$$

Proof. If j is k-good then

$$\lambda(G_j) > (1-\epsilon)\lambda(I_k).$$

Let $\mathcal{R}_k^* = \bigcup_{j=0}^{n_k-1} G_j$. Notice that $\lim_{k\to\infty} \lambda(\bigcup_{i=0}^{n_k-1} T^i I_k) = \lim_{k\to\infty} \lambda(\mathcal{R}_k) = 1$ and so for all large enough k (such that $\lambda(\mathcal{R}_k)$ is close to 1 and Lemma 2.10 holds) we have

$$\lambda(\mathcal{R}_k^*) \geq (1-\epsilon)^2 \lambda(\mathcal{R}_k) > 1 - 3\epsilon.$$

By a straightforward L^1 estimate, we have

$$\sum_{\ell=0}^{n_k-1} \lambda\big(\{y \in T^\ell I_k : |\{-\ell \leq j < n_k - \ell : G_j = \emptyset \text{ or } T^j y \notin G_{j+\ell}\}| \geq 12\epsilon n_k\}\big) < \frac{3\epsilon}{12} = \frac{\epsilon}{4}.$$

Therefore, the measure of the set of y_k satisfying (2.15) (for some ℓ) is at least $1/2$.

Recall that by Lemma 2.2 we have $\lim_{k\to\infty} \lambda(\tilde{\mathcal{R}}_k^c) = 0$ and so for k large enough,

$$\lambda(\{y : \sigma_y(\tilde{\mathcal{R}}_k) > \epsilon\}) < 1/3.$$

Thus, we can pick y_k satisfying the conditions of the lemma. □

Proof of Theorem 2.4. For each k large enough that Lemmas 2.10 and 2.11 hold and $\mathrm{diam}(I_k) < \epsilon$ and $\lambda(\mathcal{R}_k^c) < \epsilon$, let y_k be as in the statement of Lemma 2.11 and assume it is in $T^\ell I_k$ for some $0 \leq \ell < n_k$.

Step 1: We show that for all 1-Lipschitz functions f with $\|f\|_{\sup} \leq 1$ we have

$$\lim_{k\to\infty} \Big\| A_\sigma f - \sum_{i=0}^{n_k-1} c_i(y_k) U_T^i f \Big\|_2 = 0.$$

First, observe that by Lemma 2.7 and the fact that $\|f\|_{\sup} \leq 1$,

$$\Big| A_\sigma f(T^j y_k) - \sum_{i=0}^{n_k-1} c_i(T^j y_k) f(T^{i+j} y_k) \Big| < \mathrm{diam}(I_k) + 2\sigma_{T^j y_k}(\hat{\mathcal{R}}_k^c) \leq \mathrm{diam}(I_k) + 2\sigma_{y_k}(\tilde{\mathcal{R}}_k^c).$$

By our assumptions that $\mathrm{diam}(I_k) < \epsilon$ and $\sigma_{y_k}(\tilde{\mathcal{R}}_k^c) < \epsilon$ we have

$$\Big| A_\sigma f(T^j y_k) - \sum_{i=0}^{n_k-1} c_i(T^j y_k) f(T^{i+j} y_k) \Big| < 3\epsilon.$$

From Lemma 2.9 we know that if x satisfies

$$d_{\mathrm{KR}}(\sigma_x, \sigma_{T^j y_k}) < \epsilon \tag{2.16}$$

then

$$\Big| A_\sigma f(x) - \sum_{i=0}^{n_k-1} c_i(T^j y_k) f(T^{i+j} y_k) \Big| < 4\epsilon.$$

Let V denote the set of x satisfying (2.16) and such that $x \in T^{\ell+j} I_k \cap \hat{\mathcal{R}}_k$ for $-\ell \le j < n_k - \ell$. Then, for $x \in V$, we have $T^i x, T^{i+j} y_k \in T^{i+\ell+j \,(\mathrm{mod}\, n_k)} I_k$ for all $0 \le i < n_k$ since $-n_k < i,\ i+j < n_k$ (by (2.2)). Thus for any $x \in V$,

$$\left| A_\sigma f(x) - \sum_{i=0}^{n_k-1} c_i(T^j y_k) f(T^i x) \right| < 4\epsilon + \mathrm{diam}(I_k).$$

Recalling that by assumption $\mathrm{diam}(I_k) < \epsilon$ and invoking Lemma 2.8 we have

$$\int_V \left| A_\sigma f(x) - \sum_{j=0}^{n_k-1} c_j(y_k) f(T^j x) \right|^2 d\lambda(x) \le (5\epsilon + \sigma_y(\tilde{\mathcal{R}}_k))^2 < (6\epsilon)^2.$$

Since y_k satisfies the assumptions of Lemma 2.11 and $\lambda(\tilde{\mathcal{R}}_k^c) < \epsilon$, we have

$$\lambda(V^c) < 2\epsilon n_k \lambda(I_k) + \epsilon. \tag{2.17}$$

Estimating trivially on V^c we have

$$\begin{aligned} \left\| A_\sigma f - \sum_{j=0}^{n_k-1} c_j(y_k) f \circ T^j \right\|_2^2 &= \int_0^1 \left| A_\sigma f(x) - \sum_{j=0}^{n_k-1} c_j(y_k) f(T^j x) \right|^2 d\lambda(x) \\ &\le (6\epsilon)^2 + \|f\|_{\sup}^2 (2\epsilon n_k \lambda(I_k) + \epsilon). \end{aligned}$$

Since $\|f\|_{\sup} \le 1$ and ϵ is arbitrary, this establishes Step 1.

Step 2: Completing the proof. The idea of the proof is that by Step 1 and linearity we have the limit on a dense set in L^2. Since the functions on L^2 we consider have operator norm uniformly bounded (by 1), they are an equicontinuous family and so convergence on a dense set implies convergence.

To complete the formal proof of the theorem, observe that for any z we have $\sum c_i(z) = \sum |c_i(z)| \le \sigma_z([0,1])$ and we may assume that $\sigma_z([0,1]) = 1$.[1] So

$$\left\| \sum_{i=0}^{n_k-1} c_i(y_k) U_T^i \right\|_{op} \le 1 \quad \text{for all } k.$$

Therefore since we have shown $\lim_{k\to\infty} \|A_\sigma f - \sum_{i=0}^{n_k-1} c_i(y_k) U_T^i f\|_2 = 0$ for a set of f with dense span in L^2 (namely, the 1-Lipschitz functions with $\|f\|_{\sup} \le 1$), we know that for all $f \in L^2$ we have $\lim_{k\to\infty} \|A_\sigma f - \sum_{i=0}^{n_k-1} c_i(y_k) U_T^i f\|_2 = 0$. This is the definition of strong operator convergence. □

Proof of Corollary 2.5. Let $\hat{\delta}_p$ denote the point mass at p. By the proof of the theorem there exists y_k such that

$$d_{\mathrm{KR}}\left(\sigma_x, \sum_{j=0}^{n_k-1} c_j(y_k) \hat{\delta}_{(x, T^i x)}\right) < 5\epsilon$$

for all $x \in V$. By (2.17) we may assume $\lambda(V^c)$ is as small as we want. The corollary follows. □

[1] It is 1 for all but a measure zero set of z and we may change the disintegration on this zero set.

3. An abstract criterion

Let (S, Y, λ) be a uniquely ergodic topological dynamical system. Let $\hat{\delta}_p$ denote a point mass at p. Note we will consider the metric d_{KR} on the Borel probability measures on $Y \times Y$ (which is a weak-* closed set since Y is compact) and the measures $\hat{\delta}_p$ for $p \in Y \times Y$. If μ is a measure on $Y \times Y$, let $(\mu)_x$ be the disintegration of μ along $\{x\} \times Y$.

Motivated by Corollary 2.5, we wish to build *ergodic* joinings that are close to finite linear combinations of shifted power joinings. For example we wish to have ergodic measures with d_{KR} distance ϵ from the joining that gives measure $1/2$ to $\{(x, x)\}$ and measure $1/2$ to $\{(x, Sx)\}$. Naively, one wants to find a sequence of shifted power joinings that spend half their time close to $\{(x, x)\}$ and half their time shadowing $\{(x, Sx)\}$. Taking a weak-* limit of these we wish to have a measure close to the joining that gives measure $1/2$ to $\{(x, x)\}$ and measure $1/2$ to $\{(x, Sx)\}$.

Our approach will be to do this inductively, to have sequences of measures ν_i and μ_i such that ν_0 is the shifted power joining supported on $\{(x, x)\}$ and μ_0 is the joining supported on $\{(x, Sx)\}$. Inductively, μ_{i+1} spends a definite proportion of its time near μ_i and a definite proportion near ν_i, and similarly for ν_{i+1}. That is, we want to have sets A_{i+1} and B_{i+1} such that if $x \in A_{i+1}$ then $(\nu_{i+1})_x$ is close to $(\mu_i)_x$ and $(\mu_{i+1})_x$ is close to $(\nu_i)_x$, and if $x \in B_{i+1}$ then $(\nu_{i+1})_x$ is close to $(\nu_i)_x$ and $(\mu_{i+1})_x$ is close to $(\mu_i)_x$. Clearly we want the union of A_i and B_i to have almost full measure and it is helpful that they each have measure at least $c > 0$. This is not quite good enough, in particular if A_i and B_i were constant sequences. We now make the next technical proposition to overcome these issues and additionally guarantee that the limiting joining is ergodic.

Of course we want to consider the case of a linear combination of d off-diagonal joinings. That is, if we are given a finite number of shifted power joinings $\nu_0^{(1)}, \dots, \nu_0^{(d)}$ we wish to approximate $d^{-1} \sum_{i=1}^{d} \nu_0^{(i)}$. We do this analogously to the previous case. Indeed, we have A_1, B_1 and $\{\nu_1^{(i)}\}_{i=1}^{d}$ such that $(\nu_1^{(i)})_x$ is close to $(\nu_0^{(i-1)})_x$ for $x \in A_1$ (where $i - 1$ is interpreted as d if $i = 1$) and to $(\nu_0^{(i)})_x$ for $x \in B_1$. We repeat this and obtain $\{\nu_2^{(i)}\}_{i=1}^{d}$, A_2 and B_2. Now $(\nu_2^{(i)})_x$ is close to $(\nu_0^{(i-2)})_x$ for $x \in A_1 \cap A_2$. We continue repeating to approximate $d^{-1} \sum_{i=1}^{d} \nu_0^{(i)}$.

Proposition 3.1 makes this precise. Conditions (a)–(e) are the basic setup, condition (A) gives the inductive switching as above and condition (B) lets us rule out a previously mentioned issue to show that the weak-* limit of the ν_i and μ_i is close to $\frac{1}{2}(\mu_0 + \nu_0)$ and moreover that it is ergodic.

Proposition 3.1. *Let J_k be a sequence of intervals, U_k be a sequence of measurable sets, r_k be a sequence of natural numbers, $n_k^{(\ell)}$ be sequences of natural numbers for $\ell \in \{1, \dots, d\}$, and $\epsilon_j > 0$ be a sequence of real numbers. Let $A_k = \bigcup_{i=1}^{r_k} S^i(J_k) \setminus U_k$ and $B_k = A_k^c \setminus U_k$. Let $\nu_k^{(\ell)}$ be the unique $S \times S$-invariant probability measure supported on $\{(x, S^{n_k^{(\ell)}} x)\}$. Assume that:*

(a) *There exists $c > 0$ such that for all k we have $\lambda(A_k) > c$ and $\lambda(B_k) > c$.*
(b) *The minimal return time of S to J_k is at least $\frac{3}{2} r_k$.*

(c) $\lambda(U_k) < \epsilon_k$.
(d) $\lim_{k\to\infty} r_k \sum_{i>k} \lambda(J_i) = 0$.
(e) *The* ϵ_i *are non-increasing and* $\sum \epsilon_j < \infty$.

Assume moreover that:

(A) *For any* $x \in A_k$ *we have* $d_{\mathrm{KR}}((\nu_k^{(\ell)})_x, (\nu_{k-1}^{(\ell-1)})_x) < \epsilon_k$ *and for any* $x \in B_k$ *we have* $d_{\mathrm{KR}}((\nu_k^{(\ell)})_x, (\nu_{k-1}^{(\ell)})_x) < \epsilon_k$. *Note that* $\nu_{k-1}^{(\ell-1)}$ *is interpreted to be* $\nu_{k-1}^{(d)}$ *if* $\ell = 1$.
(B) $d_{\mathrm{KR}}(L^{-1}\sum_{i=1}^{L}(S\times S)^i(\nu_k^{(\ell)})_x, \nu_k^{(\ell)}) < \epsilon_k$ *for all* $x \in X$, *all* $L \geq r_{k+1}/9$ *and any* $\ell \in \{1, \ldots, d\}$.[2]

Then the weak- limit of any* $\nu_k^{(\ell)}$ *(as* $k \to \infty$*) is the same as the weak-* limit of* $d^{-1}\sum_{\ell=1}^{d}\nu_k^{(\ell)}$ *as* $k \to \infty$. *In particular these limits exist. Call this measure* μ. *It is ergodic and there exists* C *such that* $d_{\mathrm{KR}}(\mu, d^{-1}\sum_{\ell=1}^{d}\nu_k^{(\ell)}) \leq C\sum_{j=k}^{\infty}\epsilon_j$.

Note that the system $(Y\times Y, S\times S, \nu_k^{(\ell)})$ is isomorphic to (S, Y, λ), and that $(\nu_j^{(\ell)})_x$ is a point mass at $(x, S^{n_j^{(\ell)}}x)$.

Remark 3.2. To connect this to the remarks above, consider the case that the $\nu_0^{(\ell)}$ are given shifted power joinings and we want an ergodic measure close to $d^{-1}\sum\nu_0^{(\ell)}$. Of course this only treats particular types of linear combinations, but if our system is rigid (as in the case of transformations rigid rank 1 by intervals), for any shifted power joining we have different shifted power joinings close to it. For example, if we want to approximate $\tilde{\nu} = \frac{2}{3}(T^n\times \mathrm{id})_*\lambda + \frac{1}{3}(T^m\times\mathrm{id})_*\lambda$ we choose k so that $T^k \approx \mathrm{id}$. This means

$$\tilde{\nu} \approx \tfrac{1}{3}(T^{n+k}\times\mathrm{id})_*\lambda + \tfrac{1}{3}(T^n\times\mathrm{id})_*\lambda + \tfrac{1}{3}(T^m\times\mathrm{id})_*\lambda,$$

and this is the measure we approximate as above. This lets us treat general linear combinations of shifted power joinings.

Remark 3.3. One can drop the assumption that (S, Y, λ) is uniquely ergodic. In this case one replaces (B) by

$$\lambda\left(\left\{x : d_{\mathrm{KR}}\left(\frac{1}{L}\sum_{i=1}^{L}(S\times S)^i(\nu_k^{(\ell)})_x, \nu_k^{(\ell)}\right) > \epsilon_k \text{ for some } L \geq r_{k+1}/9\right\}\right) < \epsilon_k.$$

This requires some straightforward changes to the estimates in the proof of Corollary 3.5 and to the definition of the set G_k in the proof of Proposition 3.1.

[2] Note that since $S\times S$ on $\{(x, S^{n_j^{(\ell)}}x)\}$ is uniquely ergodic, such an r_{k+1} always exists [3, Proposition 4.7.1].

3.1. Proof of Proposition 3.1

Lemma 3.4. *Given $c > 0$ and $d \in \mathbb{N}$ there exist $\rho < 1$ and C such that if $0 < \delta_i < 1/2$ and a_i, b_i are such that $a_i, b_i > c$ and $1 \geq a_i + b_i > 1 - \delta_i$ and also $0 \leq \gamma_i^{(\ell)} \leq 1$ are sequences of real numbers for each $\ell \in \{1, \dots, d\}$ satisfying*

$$|\gamma_i^{(\ell)} - (a_i \gamma_{i-1}^{(\ell-1)} + b_i \gamma_{i-1}^{(\ell)})| < \delta_{i-1}, \tag{3.1}$$

then

$$\left|\gamma_i^{(s)} - \frac{1}{d}\sum_{\ell=1}^{d} \gamma_k^{(\ell)}\right| \leq C \sum_{j=k}^{i-1}\left(\delta_j + \frac{\delta_j}{1-\delta_j}\right) + C\rho^{i-k}$$

for all $k \geq 0$, $i > k$ and $s \in \{1, \dots, d\}$.

Proof. Let $\hat{\gamma}_k^{(\ell)} = \gamma_k^{(\ell)}$ and inductively let

$$\hat{\gamma}_i^{(\ell)} = \frac{a_i}{a_i + b_i}\hat{\gamma}_{i-1}^{(\ell-1)} + \frac{b_i}{a_i + b_i}\hat{\gamma}_{i-1}^{(\ell)}.$$

Observe that

$$|\hat{\gamma}_i^{(\ell)} - \gamma_i^{(\ell)}| \leq \left|\frac{a_i}{a_i + b_i}(\hat{\gamma}_{i-1}^{(\ell-1)} - \gamma_{i-1}^{(\ell-1)}) + \frac{b_i}{a_i + b_i}(\hat{\gamma}_{i-1}^{(\ell)} - \gamma_{i-1}^{(\ell)})\right| + \left|\frac{a_i}{a_i + b_i}\gamma_{i-1}^{(\ell-1)} + \frac{b_i}{a_i + b_i}\gamma_{i-1}^{(\ell)} - \gamma_i^{(\ell)}\right|.$$

The second term is at most $\frac{\delta_{i-1}}{1-\delta_{i-1}} + \delta_{i-1}$ and we inductively see that $|\hat{\gamma}_i^{\ell} - \gamma_i^{(\ell)}| \leq \sum_{j=k}^{i-1}\big(\delta_j + \frac{\delta_j}{1-\delta_j}\big)$.

Thus it suffices to show that there exist C, ρ such that

$$\left|\hat{\gamma}_i^{(s)} - \frac{1}{d}\sum_{\ell=1}^{d} \gamma_k^{(\ell)}\right| < C\rho^{i-k}.$$

To see this note that $\hat{\gamma}_{i+d}^{(s)} = \sum c_{\ell,s}\hat{\gamma}_i^{(\ell)}$ where $1 \geq c_{\ell,s} > \zeta > 0$ for some fixed ζ depending only on c and d. Consider the matrix A_i which has (ℓ, s) entry $c_{\ell,s}$. This matrix is a definite contraction in the Hilbert projective metric. Indeed, for every ζ there exists $\theta > 0$ such that if M is a positive matrix where the ratio of every pair of entries is at most ζ and v, w are any vectors in the positive cone then $D_{\mathrm{HP}}(Mv, Mw) < \theta D_{\mathrm{HP}}(v, w)$ where D_{HP} denotes the Hilbert projective metric. Now $\hat{\gamma}_{k+rd}^{(\ell)}$ is the ℓ^{th} entry of $A_k A_{k+d} \dots A_{k+(r-1)d}\tilde{\gamma}$ where $\tilde{\gamma}$ is the vector whose i^{th} entry is $\hat{\gamma}_i^{(k)}$. Since each A_{i+jd} is a definite contraction in the Hilbert projective metric, we see that $|\hat{\gamma}_{i+rd}^{(\ell)} - \hat{\gamma}_{i+rd}^{(\ell')}|$ decays exponentially in r. It is straightforward to check that

$$\frac{1}{d}\sum_{\ell=1}^{d} \hat{\gamma}_i^{(\ell)} = \frac{1}{d}\sum_{\ell=1}^{d} \hat{\gamma}_k^{(\ell)} = \frac{1}{d}\sum_{\ell=1}^{d} \gamma_k^{(\ell)}$$

and so $|\hat{\gamma}^{(\ell)}_{k+rd} - d^{-1}\sum_{\ell=1}^{d}\gamma^{(\ell)}_k|$ decays exponentially in r. After choosing $C > \rho^{-d}$ we get

$$\left|\hat{\gamma}^{(\ell)}_{k+j} - \frac{1}{d}\sum_{\ell=1}^{d}\gamma^{(\ell)}_k\right| < C\rho^j. \qquad \square$$

Corollary 3.5. *Under the assumptions of Proposition 3.1 there exist $\rho < 1$ and $C' > 0$ such that*

$$d_{\mathrm{KR}}\left(\nu^{(\ell)}_k, \frac{1}{d}\sum_{\ell=1}^{d}\nu^{(\ell)}_b\right) \le C'\sum_{j=b}^{k}\epsilon_j + C'\rho^{k-b}$$

whenever $k \ge b$ and $\ell \in \{1, \dots, d\}$.

Remark. Corollary 3.5 establishes all the conclusions of Proposition 3.1 except the ergodicity of μ.

Proof of Corollary 3.5. First notice that by (A) we have

$$d_{\mathrm{KR}}(\nu^{(\ell)}_j|_{A_j}, \nu^{(\ell-1)}_{j-1}|_{A_j}) < \epsilon_j \quad \text{and} \quad d_{\mathrm{KR}}(\nu^{(\ell)}_j|_{B_j}, \nu^{(\ell)}_{j-1}|_{B_j}) < \epsilon_j. \tag{3.2}$$

We now claim that for all ℓ,

$$d_{\mathrm{KR}}\left(\frac{1}{\lambda(A_j)}\nu^{(\ell)}_{j-1}|_{A_j}, \nu^{(\ell)}_{j-1}\right) < \epsilon_{j-1} + 2\epsilon_j + \frac{\epsilon_j}{c^2}. \tag{3.3}$$

Indeed, for f 1-Lipschitz with $\|f\|_{\sup} \le 1$ we have

$$\begin{aligned}
&\frac{1}{\lambda(J_j)r_j}\int_{A_j} f\, d\nu^{(\ell)}_{j-1} \\
&\quad = \frac{1}{\lambda(J_j)r_j}\int_{\bigcup_{i=1}^{r_j} S^i J_j \setminus U} f\, d\nu^{(\ell)}_{j-1} = \frac{1}{\lambda(J_j)r_j}\sum_{i=1}^{r_j}\int_{J_j} f\circ S^i(x)\chi_{U^c}(S^i x)\, d\nu^{(\ell)}_{j-1} \\
&\quad = \frac{1}{\lambda(J_j)r_j}\sum_{i=1}^{r_j}\int_{J_j} f\circ S^i(x)\, d\nu^{(\ell)}_{j-1} - \frac{1}{\lambda(J_j)r_j}\sum_{i=1}^{r_j}\int_{J_j} f\circ S^i(x)\chi_U(S^i x)\, d\nu^{(\ell)}_{j-1}.
\end{aligned}$$

By (B),

$$\left|\frac{1}{\lambda(J_j)r_j}\sum_{i=1}^{r_j}\int_{J_j} f\circ S^i(x)\, d\nu^{(\ell)}_{j-1} - \int f\, d\nu^{(\ell)}_{j-1}\right| \le \epsilon_{j-1},$$

and by (c) (the size estimate on U_j),

$$\left|\frac{1}{\lambda(J_j)r_j}\sum_{i=1}^{r_j}\int_{J_j} f\circ S^i(x)\chi_U(S^i x)\, d\nu^{(\ell)}_{j-1}\right| \le \|f\|_{\sup}\lambda\Big(U_j \cap \bigcup_{i=1}^{r_j} S^i J_j\Big) \le 2\epsilon_j.$$

Then (3.3) follows because

$$\left|\frac{1}{r_j\lambda(J_j)} - \frac{1}{\lambda(A_j)}\right| \le \left|\frac{1}{r_j\lambda(J_j)} - \frac{1}{r_j\lambda(J_j) - \lambda(U_j)}\right| \le \frac{\epsilon_j}{c^2}.$$

Similarly, by partitioning B_j into $D_{r_j/2}, \ldots$ where

$$D_\ell = \left\{x \in S^{r_j} J_j : \min\{i > 0 : S^i x \in J_j\} = \ell\right\},$$

we get

$$d_{\mathrm{KR}}\left(\frac{1}{\lambda(B_j)} \nu_{j-1}^{(\ell)}|_{B_j}, \nu_{j-1}^{(\ell)}\right) < \epsilon_{j-1} + 2\epsilon_j + \frac{\epsilon_j}{c^2}. \tag{3.4}$$

So for any 1-Lipschitz function f with $\|f\|_{\sup} \le 1$, we may apply Lemma 3.4 to $\gamma_i^{(\ell)} = \int f\, d\nu_i^{(\ell)}$ with $c = c$, $\delta_{j-1} = \epsilon_{j-1} + 4\epsilon_j + \epsilon_j/c^2$, $a_j = \lambda(A_j)$ and $b_j = \lambda(B_j)$. To verify (3.1), note that

$$\left| \int f\, d\nu_i^{(\ell)} - \int_{A_i} f\, d\nu_i^{(\ell)} - \int_{B_i} f\, d\nu_i^{(\ell)} \right| \le \|f\|_{\sup} \lambda(U_i) < \epsilon_i$$

and so by (3.2),

$$\left| \int f d\nu_i^{(\ell)} - \int_A f\, d\nu_{i-1}^{(\ell-1)} - \int_B f\, d\nu_{i-1}^{(\ell)} \right| \le 2\epsilon_i.$$

Then, by (3.3) and (3.4),

$$\left| \int f d\nu_i^{(\ell)} - \left(\lambda(A_i) \int f\, d\nu_{i-1}^{(\ell-1)} + \lambda(B_i) \int f\, d\nu_{i-1}^{(\ell)}\right) \right| \le \epsilon_{j-1} + 4\epsilon_j + \frac{\epsilon_j}{c^2}.$$

This completes the verification of (3.1), and, in view of Lemma 3.4, the proof of Corollary 3.5. □

To complete the proof of Proposition 3.1, we need to prove that μ is ergodic. We start with the following:

Lemma 3.6. *It suffices to show that for any $\epsilon > 0$ and $M \in \mathbb{N}$ there exist $c > 0$ and $G \subset Y \times Y$ with $\mu(G) > c$ and such that for $(x, y) \in G$ there exists $L > M$ with $d_{\mathrm{KR}}(L^{-1} \sum_{i=1}^{L} \hat{\delta}_{(S\times S)^i(x,y)}, \mu) < \epsilon$.*

To prove Lemma 3.6 we use the following consequence of the ergodic decomposition.

Lemma 3.7. *Let $\tilde{T} : \tilde{Y} \to \tilde{Y}$ be a measurable map of a σ-compact metric space and $\tilde{\mu}$ be an invariant measure. For $\tilde{\mu}$-almost every $z \in \tilde{Y}$ the sequence $N^{-1} \sum_{i=0}^{N-1} \delta_{\tilde{T}z}$ converges to an ergodic measure in the weak-* topology. (The measure is allowed to depend on the point.)*

Proof. $\tilde{\mu}$ has an ergodic decomposition $\tilde{\mu} = \int_{\tilde{Y}} \tilde{\mu}_y\, d\tilde{\mu}$ where $\tilde{\mu}_y$ is an ergodic probability measure with $\tilde{\mu}_y(\{z : \tilde{\mu}_z = \tilde{\mu}_y\}) = 1$ for $\tilde{\mu}$-almost every y. For each y, let

$$Z_y = \left\{z : \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^i x) = \int f\, d\tilde{\mu}_y \text{ for every } f \in C_c(Y)\right\}.$$

Because there is a countable $\|\cdot\|_{\sup}$-dense subset of $C_c(Y)$, by the Birkhoff Ergodic Theorem we have $\tilde{\mu}_y(Z_y) = 1$ for all y. Then $\bigcup Z_y$ has full $\tilde{\mu}$-measure and satisfies the conclusion of the lemma. □

Proof of Lemma 3.6. By our assumptions, a positive μ-measure set of (x, y) has the property that μ is a weak-* limit point of $L^{-1}\sum_{i=1}^{L}\hat{\delta}_{(S\times S)^i(x,y)}$. Indeed, choose a sequence of $\epsilon_i > 0$ converging to 0 and observe that any point in the set $\limsup G_{\epsilon_i}$ has this property. Throwing out a set of μ-measure zero where the limit may not exist, Lemma 3.5 implies this is the unique weak-* limit point and it is ergodic. □

We now identify a set of full measure for μ. As a preliminary, by assumptions (e) and (A) of Proposition 3.1 we find that if $x \notin \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} U_k$ (this is a full measure condition) then there exists $p_1(x), \dots, p_d(x)$ such that

$$\lim_{i\to\infty}\{(\nu_i^{(\ell)})_x\}_{\ell=1}^{d} = \{\hat{\delta}_{p_1(x)}, \dots, \hat{\delta}_{p_d(x)}\}.$$

Lemma 3.8. $\mu\big(\{(x, p_1(x)), \dots, (x, p_d(x))\}_{x\notin\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}U_k}\big) = 1.$

Proof. It is straightforward to see that for any $f \in C(Y \times Y)$ we have

$$\lim_{i\to\infty}\int_{Y\times Y} f\, d\Big(\sum_{\ell=1}^{d}\frac{1}{d}\nu_i^{(\ell)}\Big) = \int_Y \frac{1}{d}\sum_{\ell=1}^{d} f(x, p_\ell(x))\, d\lambda.$$

By Corollary 3.5, the left hand side is $\int f\, d\mu = \lim_{i\to\infty}\int_{Y\times Y} f d\nu_i^{(\ell)}$ for every ℓ, establishing the lemma. □

Proof of Proposition 3.1. Let G_k be the set of all $x \in Y$ such that:

(1) $S^i x \notin \bigcup_{j=k+1}^{\infty} J_j \cup S^{r_j} J_j$ for all $0 \le i \le r_{k+1}/9$.
(2) $|\{0 \le i \le r_{k+1}/9 : S^i x \in \bigcup_{j=k}^{\infty} U_j\}| < 4\sum_{j=k}^{\infty}\frac{\epsilon_j}{9} r_{k+1}$.
(3) $x \notin \bigcup_{j=k}^{\infty} U_j$.

Claim 3.9. *For all large enough k we have $\lambda(G_k) \ge 1/2$.*

This is a straightforward measure estimate using assumptions (c), (d) and (e).

Suppose $x \in G_k$ and $y \in \mathrm{supp}(\mu_x)$. The next claim shows that there exists ℓ_j such that $\lim(\nu_j^{(\ell_j)})_x$ is the point mass at y.

Claim 3.10. *There exists ℓ' such that $d_{\mathrm{KR}}(\hat{\delta}_y, (\nu_k^{(\ell')})_x) < 3\sum_{j=k+1}^{\infty}\epsilon_j$. Also*

$$d_{\mathrm{KR}}\Big(\frac{9}{r_{k+1}}\sum_{i=1}^{r_{k+1}/9}\hat{\delta}_{(S\times S)^i(x,y)}, \nu_k^{(\ell')}\Big) < C''\sum_{j=k}^{\infty}\epsilon_j.$$

Proof of Claim 3.10. We first state the following straightforward consequence of condition (A) of Proposition 3.1 (by considering if $x \in A_k$ or $x \in B_k$):

Lemma 3.11. *Let $a \in \{1, \ldots, d\}$. If $S^j x \notin J_k \cup S^{r_k} J_k$ for $0 \le j \le L$ then there exists ℓ (it is either a or $a+1$) such that $d_{\mathrm{KR}}((\nu_k^{(\ell)})_{S^j x}, (\nu_{k-1}^{(a)})_{S^j x}) < \epsilon_k$ for any $0 \le j \le L$ with $S^j x \notin U_k$.*

By iterating we obtain:

Corollary 3.12. *For all $j > k$, $\ell \in \{1, \ldots, d\}$ and $x \in G_k$ there exists ℓ' such that*

$$d((\nu_k^{(\ell)})_{S^i x}, (\nu_j^{(\ell')})_{S^i x}) < 2 \sum_{s=k+1}^{j} \epsilon_s$$

for any $0 \le i \le r_{k+1}/9$ with $S^i x \notin \bigcup_{s=k+1}^{j} U_s$.

Note that if $L \ge r_{k+1}/9$ then by condition (B) of the proposition we obtain

$$d_{\mathrm{KR}}\left(\frac{1}{L}\sum_{j=1}^{L}(\nu_k^{(a)})_{S^j x}, \nu_k^{(a)}\right) < \epsilon_k. \tag{3.5}$$

By Corollary 3.12 there exists ℓ such that if $S^i x \notin \bigcup_{\ell=k+1}^{\infty} U_\ell$ then for some ℓ we have $d_{\mathrm{KR}}(\delta_{(S^i \times S^i)(x,y)}, (\nu_k^{(\ell)})_{S^i x}) \le \sum_{j=k+1}^{\infty} \epsilon_j$ (for $0 \le i \le r_{k+1}/9$). With (3.5) this gives

$$d_{\mathrm{KR}}\left(\sum_{i=1}^{r_{k+1}/9} \delta_{(S\times S)^i(x,y)}, \nu_k^{(a)}\right) < \epsilon_k + 2\sum_{j=k+1}^{\infty}\epsilon_j + 4\sum_{j=k+1}^{\infty}\epsilon_j. \qquad \square$$

This completes the proof of Proposition 3.1 by verifying Lemma 3.6 since for all $\epsilon > 0$ there exists k_0 such that for all $k \ge k_0$ and $\ell \in \{1, \ldots, d\}$ we have $d_{\mathrm{KR}}(\mu, \nu_k^{(\ell)}) < \epsilon$ (by Corollary 3.5). $\square$

4. Proof of Theorem 1.5

In this section, we will verify the conditions of Proposition 3.1.

Before beginning the proof we set up a geometric context connected to our situation. A 3-IET with lengths ℓ_1, ℓ_2 and ℓ_3 is a rescaling of the Poincaré first return map of rotation by $\frac{\ell_2+\ell_3}{\ell_1+2\ell_2+\ell_3}$ to the interval $[0, \frac{\ell_1+\ell_2+\ell_3}{\ell_1+2\ell_2+\ell_3}) \subset [0,1)$ [8, Section 8]. If ω_{sq} denotes the area 1 square torus oriented horizontally and vertically, observe that rotation by α corresponds to the first return map of the vertical flow on $\begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix}\omega_{sq}$ to a horizontal side, which is also the time 1 map of that flow.

To set up the geometric context, let $\mathcal{M}_{1,2}$ denote the moduli space of area 1 tori with two marked points where we allow the marked points to coincide. Note that $\mathcal{M}_{1,2}$ is isomorphic to $(\mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2)/(\mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2)$. For $\omega \in \mathcal{M}_{1,2}$ let F_ω^t denote the vertical flow on ω, which corresponds to left multiplication by the element $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ltimes \begin{pmatrix} 0 \\ t \end{pmatrix}$. Let $\hat{\omega} \in \mathcal{M}_{1,2}$ be the square torus with two marked points a distance $1/2$ apart on the same horizontal line segment. Let $\mathcal{S} \subset \mathcal{M}_{1,2}$ be the set of surfaces ω such that $F_\omega^1 p$ is on the same horizontal as p and its distance along this horizontal is at most $1/2$. That is, if p is one marked point the other marked point is at $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ltimes \begin{pmatrix} s \\ 0 \end{pmatrix}$ where $s \leq 1/2$.

Let T be a 3-IET. It arises as the first return map of a rotation R_α to an interval K. Let

$$\psi_M(x) = \sum_{\ell=0}^{M-1} \chi_K(R_\alpha^\ell x).$$

Then, for any $x \in K$ such that $R_\alpha^M x \in K$,

$$T^{\psi_M(x)} x = R_\alpha^M x. \tag{4.1}$$

Let $\omega_T \in \mathcal{M}_{1,2}$ be the torus defined by taking the torus $\begin{pmatrix} 1 & -\alpha \\ 0 & 1 \end{pmatrix} \omega_{sq}$ and marking two points on the bottom horizontal line that are $|K|$ apart. Whenever convenient, in what follows we will consider K as being embedded in ω_T and $g_t K$ as being embedded in $g_t \omega_T$ where $g_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$. Here we are identifying g_t with the matrix $\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \ltimes \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and think of g_t as acting on $\mathcal{M}_{1,2} \cong (\mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2)/(\mathrm{SL}(2,\mathbb{Z}) \ltimes \mathbb{Z}^2)$ by left multiplication. Thus, for any $M \in \mathbb{N}$ we can identify $\psi_M(x)$ as the intersection number between K and a vertical line of length M on ω_T starting at an x (see Figure 1). Using this as a definition, we can make sense of ψ_M for all $M \in \mathbb{R}^+$.

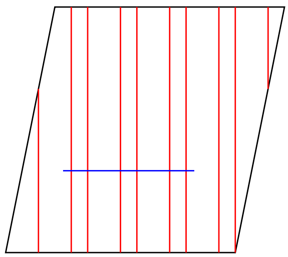

Fig. 1. The torus ω_T. A vertical segment of length M (colored red) intersects a horizontal slit (colored blue) of length $|K|$.

If we embed K in ω_T, then for $x \in K$ and $M \in \mathbb{N}$, we have

$$T^{\phi_M(x)} = F^M_{\omega_T}(x) \quad \text{if } F^M_{\omega_T}(x) \in K \tag{4.2}$$

where $\phi_M(x)$ is the number of intersections between a vertical line of length M starting at x and K.

Lemma 4.1. *For almost every T the torus $\hat{\omega}$ is a limit point of $\{g_t \omega_T\}_{t \geq 0}$.*

Proof. Let U^+ denote the subgroup $\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \ltimes \begin{pmatrix} * \\ 0 \end{pmatrix}$ of $\mathrm{SL}(2,\mathbb{R}) \ltimes \mathbb{R}^2$. Then U^+ is the expanding horospherical subgroup with respect to the action of g_t, or in other words, the orbits of U^+ are the unstable manifolds for the flow g_t.

By construction, the map $T \mapsto \omega_T$ projects to a positive measure subset $\mathcal{D}$ of a single U^+ orbit on $\mathcal{M}_{1,2}$. Moreover, the pushforward of the Lebesgue measure on the space of 3-IET's to $\mathcal{D}$ is absolutely continuous with respect to the pushforward of the Haar measure on U^+ to $\mathcal{D}$. The lemma then follows from the ergodicity of g_t. □

Corollary 4.2. *For every $\delta > 0$ and almost all T, there exists arbitrarily large $t > 0$ with $g_t\omega_T \in B(\hat{\omega}, \delta) \cap \mathcal{S}$.*

Proof. Since $\hat{\omega}$ is square, for $p \in \hat{\omega}$ we have $F^1_{\hat{\omega}} p = p$. Therefore, for $\omega' \in B(\hat{\omega}, \delta')$ and $p \in \omega'$, $F^1_{\omega'} p$ is within $c_1(\delta')$ of p, where $c_1(\delta') \to 0$ as $\delta' \to 0$. Write $F^1_{\omega'} p - p = (v_1, v_2)$, and note that for $\delta' > 0$ sufficiently small and for small $s \in \mathbb{R}$, for $p \in g_s\omega'$, we have

$$F^1_{g_s\omega'} p - p = (e^{-s} v_1, 1 - e^s + e^s v_2).$$

Therefore, given $\omega' \in B(\hat{\omega}, \delta)$, we can choose $s \in \mathbb{R}$, with $|s| < c_2(\delta')$ where $c_2(\delta') \to 0$ as $\delta \to 0$, such that $1 - e^s + e^s v_2 = 0$, i.e. $g_s\omega' \in \mathcal{S}$. We have $g_s\omega' \in B(\hat{\omega}, c_3(\delta'))$ with $c_3(\delta') \to 0$ as $\delta' \to 0$.

Suppose T is such that $\hat{\omega}$ is a limit point of $\{g_t\omega_T\}_{t \geq 0}$. Choose $\delta' > 0$ such that $c_3(\delta') < \delta$ and choose t' such that $g_{t'}\omega_T \in B(\hat{\omega}, \delta')$ and then let $t = t' + s$ where s is as in the previous paragraph. Then $g_t\omega_T \in B(\hat{\omega}, \delta) \cap \mathcal{S}$ as required. □

We now apply g_t to Figure 1 with $t = \log M$. Note that $\psi_M(x)$ is also the intersection number between a vertical segment γ_1 of length 1 and a horizonal slit γ_2 of length $M|K|$ (see Figure 2). From now on, we assume that $g_t\omega_T \in B(\hat{\omega}, \delta) \cap \mathcal{S}$ for some $\delta \ll 1$.

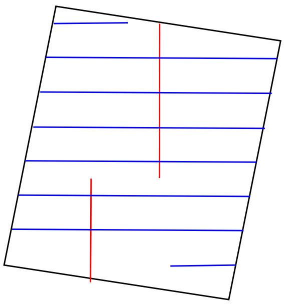

Fig. 2. The torus $g_{\log M}\omega_T$: A vertical segment γ_1 of length 1 (red) intersects a horizontal slit γ_2 of length $M|K|$ (blue). If we also assume that $g_{\log M}\omega_T \in B(\hat{\omega}, \delta) \cap \mathcal{S}$ then the torus $g_{\log M}\omega_T$ is nearly square, and the two endpoints of γ_1 are on the same horizontal line segment, of length at most $1/2 + O(\delta)$.

The following lemma refers to Figure 3.

Lemma 4.3. *There exists m such that if the green segment does not cross the purple segment then the number of times a trajectory of length 1 crosses $g_t K$ (the blue lines) is either m or $m + 1$. Moreover, it is $m + 1$ if the trajectory does not cross the (horizontal) purple segment, and m if it does.*

In other words, for the set of points x whose green segment does not cross the purple segment, $\phi_M(x)$ is m if its red segment crosses the purple segment (where ϕ_M is as in (4.2)) and $\phi_M(x) = m + 1$ if it does not.

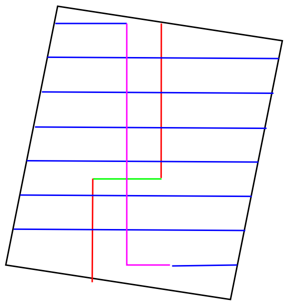

Fig. 3. Closing the curves. We complete the vertical segment γ_1 to a closed curve $\hat{\gamma}_1$ by adding a horizontal segment ζ_1 (green). Note that since $g_t\omega_T \in B(\hat{\omega}, \delta) \cap \mathcal{S}$, the length of ζ_1 is at most $1/2 + O(\delta)$. Similarly, we close up the horizontal slit γ_2 to obtain a closed curve $\hat{\gamma}_2$ by adding in a horizontal segment ζ_2 and a vertical segment ζ_2' (purple).

Note that because Figure 3 represents $g_{\log M}\omega_T$, vertical trajectories of length 1 in Figure 3 correspond to vertical trajectories of length M on ω_T.

Proof of Lemma 4.3. Indeed, the family of curves we define are all homotopic and so their intersection with $\hat{\gamma}_2$ is all the same. So for such curves, if the green and purple segments have intersection number zero then the intersection of the red segment and the blue segment depends only on the intersection of the purple segment and the red segment, which by construction is either 0 or 1. □

For the remainder, let λ denote 1-dimensional Lebesgue measure restricted to K.

Lemma 4.4. *For all $\epsilon > 0$ there exists $\delta > 0$ so that if $\omega \in B(\hat{\omega}, \delta) \cap \mathcal{S}$ and the flow F^s_ω is minimal then there exist $\rho < \epsilon$ and $L \in \mathbb{N}$ such that for any interval J with $|J| = \rho$ we have:*

- $\lambda(\bigcup_{s\in[0,L)} F^s_\omega J) > 1 - \epsilon$.
- *For all* $0 \le s < \ell < L$ *we have* $F^s_\omega J \cap F^\ell_\omega J = \emptyset$.
- $F^1_\omega J$ *is horizontally adjacent to* J.

Proof. Suppose p is a point in ω, and $\omega \in \mathcal{S}$. Then $F^1_\omega p$ is horizontally adjacent to p. For all $\epsilon > 0$ there exists $\delta > 0$ such that if ω is in $\mathcal{S} \cap B(\hat{\omega}, \delta)$ then $F^1_\omega p$ is translated by less than $\epsilon/9$. Since the vertical flow on ω is minimal, $F^1_\omega p \neq p$. Therefore, $F^1_\omega p$ is translated horizontally by some $\rho > 0$. Let J be a horizontal interval of length ρ. We choose $L = \min\{s > 0 : F^s_\omega J \cap J \neq \emptyset\}$. We have $\lambda(\bigcup_{s\in[0,L)} F^s_\omega J) > 1 - \rho$. Indeed, $\bigcup_{s\in[0,L+1)} F^s_\omega J = \omega$. □

Proposition 4.5. *For any $a_1, \ldots, a_d, b_1, \ldots, b_d \in \mathbb{Z}$ and $c/5 > \epsilon > 0$ there exist $\delta, t_0 > 0$ such that if $g_t\omega_T \in B(\hat{\omega}, \delta) \cap \mathcal{S}$, $\lambda(K) > c$ and $t > t_0$ then there exist*

- $n_1, \ldots, n_d \in \mathbb{Z}$ *and* $r, L \in \mathbb{N}$,
- *an interval* $J \subset K$ *and a measurable set* $B \subset K$

such that the minimal return time (under T) to J is at least $\frac{3}{2}r$, and for $A = \bigcup_{i=0}^{r} T^i J$ we have $\lambda(A) > \frac{1}{2}\lambda(K) - \epsilon$, $\lambda(B) > \frac{1}{2}\lambda(K) - \epsilon$ and the sets A and B satisfy

$$d(T^{n_\ell}x, T^{a_\ell}x) < \epsilon \quad \text{for all } x \in A \text{ and } \ell \in \{1, \dots, d\}, \tag{4.3}$$

$$d(T^{n_\ell}x, T^{b_\ell}x) < \epsilon \quad \text{for all } x \in B \text{ and } \ell \in \{1, \dots, d\}. \tag{4.4}$$

Moreover, $T^i J \cap T^j J = \emptyset$ for all $0 \le i < j \le \frac{3}{2}r$. Lastly, if $\nu^{(a_\ell)}$ is the joining supported on $\{(x, T^{a_\ell}x)\}$ then for all $x \in A$ and $\ell \in \{1, \dots, d\}$ we have

$$d_{\mathrm{KR}}\left(\frac{1}{L}\sum_{i=0}^{L-1} \delta_{(T^i x, T^{i+n_\ell}x)}, \nu^{(a_\ell)}\right) < 2\epsilon \tag{4.5}$$

and if $\nu^{(b_\ell)}$ is the joining supported on $\{(x, T^{b_\ell}x)\}$ then for all $x \in B$ and $\ell \in \{1, \dots, d\}$ we have

$$d_{\mathrm{KR}}\left(\frac{1}{L}\sum_{i=0}^{L-1} \delta_{(T^i x, T^{i+n_\ell}x)}, \nu^{(b_\ell)}\right) < 2\epsilon. \tag{4.6}$$

Remark 4.6. Specializing to the case where $d = 1$, $a = 0$ and $b = k$, we see that $\frac{1}{2}(\mathrm{id} + T^k)$ is in the weak closure of the powers of T. Veech showed that almost every 3-IET has simple spectrum [17, Theorem 1.3]. Combining these two facts with Ryzhikov's [14, Theorem 6.1(3, 4)] we find that the spectra of T^n and $T \times \dots \times T$ (n times) are simple for all $n > 0$.

Proof of Proposition 4.5. In view of Lemma 4.4, we can choose δ so small that for any $\omega \in B(\hat{\omega}, \delta)$,

(i) the horizontal purple line has length between $1/2 - \epsilon/4$ and $1/2 + \frac{\epsilon}{4}$ (which we can do because the two marked points on $\hat{\omega}$ are $\frac{1}{2}$ apart),
(ii) $F^1_\omega x = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right) \ltimes \left(\begin{smallmatrix}\rho\\ 0\end{smallmatrix}\right) x$ where $0 < \rho \le \frac{\epsilon}{10 \max\{|a_\ell| + |b_\ell|\}}$.

Because $T \times T$ is uniquely ergodic on the supports of $\nu^{(a_\ell)}$ and $\nu^{(b_\ell)}$, there exists L_ℓ such that if $d(p_i, T^{a_\ell + i}y) < \epsilon$ for all $0 \le i \le L$ then

$$d_{\mathrm{KR}}\left(\frac{1}{L}\sum_{i=0}^{L-1} \delta_{(T^i y, p_i)}, \nu^{(a_\ell)}\right) < 2\epsilon, \quad d_{\mathrm{KR}}\left(\frac{1}{L}\sum_{i=0}^{L-1} \delta_{(T^i y, p_i)}, \nu^{(b_\ell)}\right) < 2\epsilon \tag{4.7}$$

for all $L \ge L_\ell$ and $y \in K$. Indeed, $T \times T$ is uniquely ergodic on $\mathrm{supp}(\nu^{(a)})$ and $\mathrm{supp}(\nu^{(b)})$, and uniquely ergodic systems have uniform convergence of Birkhoff averages of continuous functions (see for example [3, Proposition 4.7.1]). Let $L_0 = \max L_\ell$. We choose t_0 so large that any vertical trajectory of length e^{t_0} on ω_T crosses K at least L_0 times. We further assume $L_0 > \max\{|a_\ell|, |b_\ell|\}$.

We now set about defining J and A. Let V be the horizontal purple line segment. Let ρ be as in the previous lemma for $g_t\omega_T$. For any horizontal interval I on $g_t\omega_T$ of length ρ we have one of the following mutually exclusive possibilities:

(a) $\bigcup_{s\in[0,1)} F^s(I) \cap V = \emptyset$.
(b) There exists $s \in [0, 1)$ such that $F^s(I) \subset V$.
(c) $\bigcup_{s\in[0,1)} F^s(I) \cap (\partial V) \neq \emptyset$.

Note that by Lemma 4.3 there exists m such that if $I \subset K$ and (b) holds then $\phi_M(x) = m$ and similarly if $I \subset K$ and (a) holds then $\phi_M(x) = m + 1$.

Let $\hat{A}$ be the set of points in $g_t\omega_T$ which belong to some horizontal interval of length ρ satisfying (b). Let

$$\tilde{A} = \bigcap_{s\in[-2-\max\{|a_\ell-b_\ell|\},2+\max\{|a_\ell-b_\ell|\}]} F^s_{g_t\omega_T}\hat{A}. \tag{4.8}$$

Let $\rho > 0$ be given by Lemma 4.4 and I be an interval of length ρ in $\tilde{A} \cap g_t K$ such that $F^{-1}_{g_t\omega_T} I \not\subset \tilde{A}$. Now $F^1 I$ is horizontally adjacent to I, and so $F^j_{g_t\omega_T} I$ is horizontally $j\rho$ away from I. So by our assumption on the length of I, we have

$$F^j_{g_t\omega_T} I \subset \tilde{A} \quad \text{for all } 0 \le j \le |V|/\rho - 2(2 + \max\{|a_\ell + b_\ell|\}) - 3 \equiv \hat{p}. \tag{4.9}$$

(Note that by (ii) and the fact that $|V| > 1/2 - \epsilon$ we have $\hat{p} \ge 1$.)

We now use what we have done for the flow on $g_t\omega_T$ to establish some of our claims about the IET T. Let r be the cardinality of the set of intervals of length ρ in $\bigcup_{s\in[0,\hat{p})} F^s_{g_t\omega_T} I \cap K$. Note that because in our set $\hat{A}$ a vertical trajectory of length 1 crosses $g_t K \subset g_t\omega_T$ exactly m times, we have $r = m\hat{p}$.

Let $A' = g_{-t}\hat{A} \cap K \subset \omega_T \cap K$, which we may as well consider as a subset of the domain of T (because it is contained in K). Note that

$$\phi_{e^t}(x) = m \quad \text{for } x \in A'. \tag{4.10}$$

We also have, for all $x \in A'$,

$$d(F^{e^t}_{\omega_T} x, x) = e^{-t}\rho, \tag{4.11}$$

because when we apply g_{-t} to pull our dynamics from $g_t\omega_T$ back to ω_T, we contract horizontal distances by e^{-t}. It follows from (4.2), (4.10) and (4.11) that

$$d(T^m x, x) = e^{-t}\rho \quad \text{for all } x \in A'. \tag{4.12}$$

Let J denote the interval corresponding to I in the domain of our IET T. That is, we consider $J = g_{-t}I \subset K \subset \omega_T$, which since it is in K we consider as an interval in the domain of T. Let $A = \bigcup_{i=0}^{r-1} T^i J$, which we can consider as a subset of $K \subset \omega_T$. We now claim that

$$A \subset g_{-t}\tilde{A} \cap K. \tag{4.13}$$

Indeed, by (4.9), we have

$$F^{se^t}_{\omega_T} J \subset g_{-t}\tilde{A} \quad \text{for all } 0 \le s \le \hat{p}. \tag{4.14}$$

It follows, in view of (4.10), that for $x \in J$,

$$\phi_{\hat{p}e^t}(x) = \sum_{k=0}^{\hat{p}-1} \phi_{e^t}(F^{ke^t}_{\omega_T} x) = m\hat{p} = r. \tag{4.15}$$

By (4.2) we have, for $x \in J$ and $i \in \mathbb{N}$,

$$T^i x = F^s_{\omega_T} x \quad \text{where } s \text{ is such that } \phi_s(x) = i.$$

Since for a fixed $x \in J$, the map $s \mapsto \phi_s(x)$ is increasing, for $0 < i < r$ we have in view of (4.15),

$$T^i x = F^s_{\omega_T} x \quad \text{where } s < \hat{p}.$$

This together with (4.14) implies (4.13). The same argument shows that

$$T^u x \in A' \quad \text{for } x \in A \text{ and } |u| \leq m(\max\{|a_\ell - b_\ell|\} + 1). \tag{4.16}$$

Let $n_\ell = a_\ell + (a_\ell - b_\ell)m$. We claim that for all $x \in A$,

$$d(T^{n_\ell}x, T^{a_\ell}x) \leq d(T^{a_\ell}x, T^{a_\ell}x) + \sum_{i=1}^{|a-b|} d(T^{im+a_\ell}x, T^{(i-1)m+a_\ell}x) \leq \epsilon e^{-t} \leq \epsilon.$$

Indeed, by (4.16) and (4.12) we have $d(T^{jm+a_\ell}x, T^{(j-1)m+a_\ell}x) = \rho e^{-t}$ for all $|j| \leq |a_\ell - b_\ell|$, because $|a_\ell| < m$. We obtain the second inequality by (ii).

We now show that for all $x \in A$,

$$d_{\mathrm{KR}}\left(\frac{1}{m}\sum_{i=0}^{m-1} \delta_{(T^i x, T^{i+n_\ell}x)}, \nu^{(a_\ell)}\right) < 3\epsilon.$$

By construction, if $x \in A$ then $T^i x \in A'$ for all $-m \leq i \leq m$. Consequently, $d(T^{i+n_\ell}x, T^{i+a_\ell}x) < \epsilon$ for all $|i| \leq |m|$. So by (4.7) and the fact that $m \geq L_0 \geq L_\ell$ we have our condition on d_{KR}.

We now show that $\lambda(A) > \frac{1}{2}\lambda(K) - \epsilon$. This follows from the fact that by (ii) the measure of the set of $x \in g_t\omega_T$ such that $F^\ell_{\omega_T}x$ crosses the horizontal purple strip for $0 \leq \ell \leq 1$ and $-1 \leq \ell \leq 0$, and $F^s_{g_t\omega_T}x$ does not have this property for some $-1 \leq s \leq 1$, has measure at most $2\frac{\epsilon}{10\max\{|a|+|b|\}}$. By our condition on the length of the purple horizontal strip, the measure condition on A is proved.

The fact that the return time of T to J is at most $\frac{3}{2}r$ follows from the fact that the measure of A^c is at most $\frac{1}{2}\lambda(K) + \epsilon$ and so the orbit of J after leaving A and before returning to J has measure at least $\frac{1}{2}\lambda(K) - \epsilon - \epsilon > \frac{1}{2}\lambda(A)$. So J has at least $\frac{1}{2}r$ images outside of A before part of it returns.

We now similarly define $B \subset A^c$ with the desired properties. First let

$$\hat{B} = \left\{x \in g_t\omega_T : \bigcup_{s\in[-3-\max\{|a-b|\}, 3+\max\{|a-b|\}]} F^s_{g_t\omega_T}(x) \cap V = \emptyset\right\}.$$

Similarly to before let $\tilde{B} = \bigcap_{s\in[-1,1]} F^s_{g_t\omega_T}\hat{B}$ and $B = g_{-t}\tilde{B} \cap K \subset \omega_T$, considered as a subset of the domain of T. Now as above, by Lemma 4.3 if $x \in \tilde{B}$ then we see that a vertical trajectory of length 1 or -1 emanating from x crosses $g_t K$ exactly $m+1$ times. Moreover, $F^s_{g_t\omega_T}x$ has this property for all $-|a_\ell - b_\ell| \le s \le |a_\ell - b_\ell|$. Since $n_\ell = b_\ell + (m+1)(a_\ell - b_\ell)$, for any $x \in B$ and $|i| \le m$ we have

$$d(T^{n_\ell}T^i x, T^{b_\ell}T^i x) \le \sum_{j=1}^{|a_\ell - b_\ell|} d(T^{j(m+1)}x, T^{(j-1)(m+1)}x) \le \epsilon.$$

Thus, as above we have $d_{\mathrm{KR}}(m^{-1}\sum_{i=0}^{m-1}\delta_{(T^i x, T^{i+n_\ell}x)}, \nu^{(b_\ell)}) < 3\epsilon$ for all $x \in B$. The fact that $\lambda(B) > \lambda(K) - \epsilon$ is proved similarly to the case of $\lambda(A)$ above. □

Proof that Proposition 4.5 implies the assumptions of Proposition 3.1. We prove the case $d = 1$; the case of larger d is analogous. Choose ϵ_i satisfying assumption (e). Define $a_1, b_2 = a$ and $b_1, a_2 = b$. We apply Proposition 4.5 to the pairs (a_1, b_1), (a_2, b_2) and $\epsilon = \epsilon_1/2$ to obtain $n_1^{(1)}, n_1^{(2)}, A_1, B_1$ and r_1. Note that $U_1 = (A \cup B)^c$ and its measure is less than $1 - 2(1/2 - \epsilon_1/2) = \epsilon_1$, (4.3) and (4.4) imply (A), and (4.5) and (4.6) imply (B). We repeat this procedure with n_1 and n_2 in place of a and b, and ϵ_2 in place of ϵ_1, and obtain $n_2^{(1)}, n_2^{(2)}, A_2, B_2, r_2$ and J_2. We further require that the interval J_2 satisfies $r\lambda(J_2) < \epsilon_2$. Iterating this we have the conditions of Proposition 3.1. □

Proof of Theorem 1.5. Let μ be an invariant measure for $T \times T$. By Corollary 2.5 there exist $n_1, \dots, n_d$ such that if ν_i is the joining supported on $\{(x, T^{n_i}x)\}$ then $d_{\mathrm{KR}}(\mu, d^{-1}\sum_{i=1}^d \nu_i) < \epsilon$. By the above, Proposition 4.5 implies we can satisfy the assumptions of Proposition 3.1 with $\sum \epsilon_i < \epsilon$. By Proposition 3.1 we obtain an ergodic measure within $C\epsilon$ of $d^{-1}\sum_{i=1}^d \nu_i$. This establishes that the joinings form a Poulsen simplex.

It remains to prove that there is an ergodic self-joining that is neither $\lambda \times \lambda$ nor one-to-one on almost every fiber. Let $\nu_0^{(1)}$ be the self-joining carried by $\{(x, x)\}$ and $\nu_0^{(2)}$ be the self-joining carried by $\{(x, Tx)\}$. Let $\epsilon_i > 0$ satisfy

$$d(x, Tx) > 40C\sum_{i=1}^{\infty}\epsilon_i, \tag{4.17}$$

$$d_{\mathrm{KR}}\big(\lambda \times \lambda, \tfrac{1}{2}(\nu_0^{(1)} + \nu_0^{(2)})\big) > 4C\sum_{i=1}^{\infty}\epsilon_i, \tag{4.18}$$

where C is as in the conclusion of Proposition 3.1. We apply Proposition 3.1 for these ϵ_i as above to obtain $\nu_i^{(1)}, \nu_i^{(2)}$ and their weak-* limit ν_∞, an ergodic measure which by (4.18) is not $\lambda \times \lambda$. The following lemma show ν_∞ cannot be one-to-one on almost every fiber.

Lemma 4.7. *If μ is a measure that is one-to-one on almost every fiber then μ cannot be the weak-* limit of a sequence of measures $\tilde{\nu}_i$ that are two-to-one on almost every fiber and such that*

$$\lambda(\{x : \mathrm{diam}(\mathrm{supp}\,(\tilde{\nu}_i)_x) > \delta\}) > 3/4$$

for infinitely many i.

Proof. There exists $f : [0,1) \to [0,1)$ measurable such that μ is carried by $\{(x, f(x))\}$. By Luzin's Theorem there exists $\mathcal{K}$ compact with $\lambda(\mathcal{K}) > \frac{99}{100}$ such that $f|_{\mathcal{K}}$ is uniformly continuous. Let $s > 0$ be such that $d(f(x), f(y)) < \delta/8$ for all $x, y \in \mathcal{K}$ with $d(x, y) < s$. Choose an interval I with $|I| \leq s$, $\lambda(I \cap \mathcal{K}) > \frac{99}{100}\lambda(I)$ and

$$\lambda(\{x \in I : \operatorname{diam}(\operatorname{supp}(\tilde{\nu}_i)_x) > \delta\}) > \tfrac{1}{2}|I| \tag{4.19}$$

for infinitely many i. Let $p = f(x)$ for some $x \in I \cap \mathcal{K}$ and let $g : [0,1) \times [0,1) \to \mathbb{R}$ be a 1-Lipschitz function such that

- $g|_{I^c \times [0,1)} \equiv 0$,
- $g|_{I \times B(p,\delta/4)} \equiv 0$,
- $g(x, y) = \min\{d(x, \partial I), d(y, \partial B(p, \delta/4)), \delta/4\}$ for all $(x, y) \in I \times (B(p, \delta/4))^c$.

Now $\int g\, d\mu \leq .01|I| \cdot \|g\|_{\sup} \leq .01|I| \cdot \min\{\delta/4, |I|/2\}$. On the other hand if $\tilde{\nu}_i$ satisfies (4.19) then on a set of $x \in I$ of measure at least $|I|/3$ we find that one of the two points in $\operatorname{supp}(\tilde{\nu}_i)_x$ is at least $\delta/2$ away from p. A subset of these x of measure at least $|I|/6$ satisfies $d(x, \partial I) \geq \frac{1}{12}|I|$. So $\int g\, d\tilde{\nu}_i \geq (|I|/6)\min\{\delta/4, |I|/12\}$. Since g is 1-Lipschitz it follows that

$$d_{\mathrm{KR}}(\mu, \tilde{\nu}_i) > |I| \min\left\{|I|\left(\tfrac{1}{72} - \tfrac{1}{200}\right), \tfrac{\delta}{24} - \tfrac{\delta}{400}\right\},$$

proving the lemma. □

Let $\tilde{\nu}_i = \frac{1}{2}(\nu_i^{(1)} + \nu_i^{(2)})$. Since by (4.17) they satisfy the condition in the lemma, we see that T is not 2-simple. □

Remark 4.8. Morally, for our construction of joinings we use the fact that our transformation T preserves λ and has the following properties: There exists $r > 0$, an infinite sequence of numbers $n_1, n_2, \ldots$ and intervals $I_1, I_2, \ldots$ and $J_1, J_2, \ldots$ such that

- $I_k, TI_k, \ldots, T^{n_k - 1}I_k, J_k, TJ_k, \ldots, T^{n_k}J_k$ are disjoint intervals,
- $\lim_{k\to\infty} \lambda(\bigcup_{i=0}^{n_k-1} T^i I_k \cup \bigcup_{j=0}^{n_k} T^j J_k) = 1$,
- $\lambda(\bigcup_{i=0}^{n_k-1} T^i I_k)$ and $\lambda(\bigcup_{i=0}^{n_k} T^i J_k)$ are both at least r for all k,
- $\lim_{k\to\infty} \frac{\lambda(I_k \cap T^{n_k} I_k)}{\lambda(I_k)} = 1 = \lim_{k\to\infty} \frac{\lambda(J_k \cap T^{n_k+1} J_k)}{\lambda(J_k)}$.

Acknowledgments. We thank El Abdaloui, M. Lemańczyk and V. Ryzhikov for numerous illuminating discussions on connections of other results with this work. We thank the referee for valuable corrections and suggestions to improve the paper.

The research of J. Chaika was supported in part by NSF grants DMS-135500 and DMS-1452762, the Sloan foundation, Poincaré chair, Warnock chair.

The research of A. Eskin is partially supported by the Simons Foundation and NSF grant DMS 1201422.

References

[1] Ageev, O. N.: A typical dynamical system is not simple or semisimple. Ergodic Theory Dynam. Systems **23**, 1625–1636 (2003) Zbl 1049.37002 MR 2032480

[2] Boshernitzan, M.: A unique ergodicity of minimal symbolic flows with linear block growth. J. Anal. Math. **44**, 77–96 (1984/85) Zbl 0602.28008 MR 801288

[3] Brin, M., Stuck, G.: Introduction to Dynamical Systems. Cambridge Univ. Press, Cambridge (2002) Zbl 1319.37001 MR 1963683

[4] Ferenczi, S., Holton, C., Zamboni, L. Q.: Joinings of three-interval exchange transformations. Ergodic Theory Dynam. Systems **25**, 483–502 (2005) Zbl 1076.37002 MR 2129107

[5] Fogg, N. P.: Substitutions in Dynamics, Arithmetics and Combinatorics. Lecture Notes in Math. 1794, Springer, Berlin (2002) Zbl 1014.11015 MR 1970385

[6] Janvresse, E., de la Rue, T., Ryzhikov, V.: Around King's rank-one theorems: flows and $\mathbb{Z}^n$-actions. In: Dynamical Systems and Group Actions, Contemp. Math. 567, Amer. Math. Soc., Providence, RI, 143–161 (2012) Zbl 1303.37001 MR 2931916

[7] Katok, A.: Combinatorial Constructions in Ergodic Theory and Dynamics. Univ. Lecture Ser. 30, Amer. Math. Soc., Providence, RI (2003) Zbl 1030.37001 MR 2008435

[8] Katok, A. B., Stepin, A. M.: Approximations in ergodic theory. Uspekhi Mat. Nauk **22**, no. 5, 81–106 (1967) (in Russian) Zbl 0172.07202 MR 0219697

[9] King, J.: The commutant is the weak closure of the powers, for rank-1 transformations. Ergodic Theory Dynam. Systems **6**, 363–384 (1986) Zbl 0595.47005 MR 863200

[10] King, J.: Flat stacks, joining closure and genericity. Preprint

[11] Lindenstrauss, J., Olsen, G., Sternfeld, Y.: The Poulsen simplex. Ann. Inst. Fourier (Grenoble) **28**, no. 1, vi, 91–114 (1978) Zbl 0363.46006 MR 500918

[12] Rudolph, D. J.: An example of a measure preserving map with minimal self-joinings, and applications. J. Anal. Math. **35**, 97–122 (1979) Zbl 446.28018 MR 555301

[13] Ryzhikov, V. V.: Mixing, rank and minimal self-joining of actions with invariant measure. Mat. Sb. **183**, 133–160 (1992) (in Russian) Zbl 0768.28008 MR 1180921

[14] Ryzhikov, V. V.: Weak limits of powers, the simple spectrum of symmetric products, and mixing constructions of rank 1. Mat. Sb. **198**, 137–159 (2007) (in Russian) Zbl 1161.37011 MR 2354530

[15] Ryzhikov, V. V.: Self-joinings of rank-one actions and applications. In: École de Théorie Ergodique, Sémin. Congr. 20, Soc. Math. France, Paris, 193–206 (2010) Zbl 1234.28018 MR 2856515

[16] Veech, W. A.: A criterion for a process to be prime. Monatsh. Math. **94**, 335–341 (1982) Zbl 0499.28016 MR 685378

[17] Veech, W. A.: The metric theory of interval exchange transformations. I. Generic spectral properties. Amer. J. Math. **106**, 1331–1359 (1984) Zbl 0631.28006 MR 765582