ANNALES
**HENRI LEBESGUE**

JON CHAIKA

DONALD ROBERTSON

# A RANK ONE MILD MIXING SYSTEM WITHOUT MINIMAL SELF JOININGS

## UN SYSTÈME DE RANG UN MODÉRÉMENT MÉLANGEANT SANS AUTOCOUPLAGES MINIMAUX

ABSTRACT. — We show that there is a rank 1 transformation that is mildly mixing but does not have minimal self-joinings, answering a question of Thouvenot.

RÉSUMÉ. — Nous montrons l'existence d'une transformation de rang 1 qui est modérément mélangeante mais admet des autocouplages non minimaux, ce qui répond à une question de Thouvenot.

## 1. Introduction

A *joining* of two measure-preserving system $(X, \mu, T)$ and $(Y, \nu, S)$ is any $T \times S$ invariant measure on $X \times Y$ that is mapped to $\mu$ by the first coordinate projection

and to $\nu$ by the second. The product measure $\mu \otimes \nu$ is always a joining. The existence of other joinings of $(X \mu, T)$ and $(Y, \nu, S)$ detects whether the systems have properties in common. In particular, if $(X, \mu, T)$ and $(Y, \nu, S)$ have no other joinings they tend to have orthogonal dynamical properties. When $\mu \otimes \nu$ is the only joining the systems $(X, \mu, T)$ and $(Y, \nu, S)$ are called *disjoint.* Joinings and disjointness were first studied by Furstenberg [Fur67]. The survey of de la Rue [Rue23] is a good introduction.

A non-trivial system $(X, \mu, T)$ is never disjoint from itself: there are always *off-diagonal* joinings obtained by embedding $X$ in $X \times X$ using the map $x \mapsto (x, T^n x)$ for any $n \in \mathbb{Z}$. When these and the product measure are the only ergodic joinings of $(X, \mu, T)$ with itself the system $(X, \mu, T)$ is said to have *minimal self-joinings.* Rudolph [Rud79] gave the first example of a system with minimal self-joinings.

King [Kin88] showed that all mixing rank one systems have minimal self-joinings. The survey of Ferenczi [Fer97] collects many equivalent definitions of rank one and we record here the definition most useful for our purposes.

DEFINITION 1.1 ([Fer97, Definition 6]). — *A system* $(X, \mu, T)$ *is* rank one *if for every measurable set* $A \subset X$ *and every* $\epsilon > 0$ *there is a measurable set* $F \subset X$ *and* $h \in \mathbb{N}$ *and a measurable set* $A' \subset X$ *such that:*
  - *the sets* $F, TF, \ldots, T^{h-1}F$ *are pairwise disjoint,*
  - $\mu(A' \triangle A) < \epsilon$,
  - $\mu(F \cup TF \cup \cdots \cup T^{h-1}F) > 1 - \epsilon$,
  - $A'$ *belongs to the* $\sigma$-*algebra generated by* $\{F, TF, \ldots, T^{h-1}F\}$,

*all hold.*

Thouvenot asked the following question.

QUESTION 1.2 ([Ryz20, § 1.2 in Section 12]). — *Does a non-rigid rank one transformation without factors have the minimal self-joinings property?*

To make sense of this question we recall the following definitions.

DEFINITION 1.3. — *Fix a system* $(X, \mu, T)$. *Say that* $f \in \mathsf{L}^2(X, \mu)$ *is* rigid *for* $T$ *if there is a sequence* $\mathbf{r}$ *in* $\mathbb{Z}$ *with* $|r(n)| \to \infty$ *and* $T^{r(n)}f \to f$ *in* $\mathsf{L}^2(X, \mu)$. *In this situation we say that* $f$ *is* rigid *along* $\mathbf{r}$. *Any such sequence is called an* $f$- rigidity sequence. *One says that* $(X, \mu, T)$ *is* mild mixing *if the only rigid functions are the constant functions.*

King [Kin86, p. 377] proved that all non-trivial factors of rank one systems are rigid. As mild mixing systems cannot have non-trivial rigid factors, Question 1.2 is equivalent to asking whether mildly mixing rank-one systems have minimal self-joinings. We construct a mild-mixing rank-one system that does not have minimal self-joinings, thereby answering negatively Question 1.2.

THEOREM 1.4. — *There is a mild mixing rank one system* $(X, \mu, T)$ *without minimal self-joinings.*

Ryzhikov has results connected to this theorem. On the one hand he showed that such an example can not be "bounded" rank 1. That is, bounded rank 1 mildly mixing systems have minimal self-joinings [Ryz13]. On the other hand, he showed that for

any $\epsilon > 0$ there is a mildly mixing transformation without minimal self-joinings that has a "local rank" of $1 - \epsilon$ [Ryz12].

## 1.1. Outline of proof

We define in Section 2 a measure-preserving system $(X, \mu, T)$. The following properties of the system are established in Sections 3, 4 and 5 respectively.

THEOREM 1.5. — *The system constructed in Section 2 is rank one.*

THEOREM 1.6. — *The system constructed in Section 2 does not have the minimal self-joinings property.*

THEOREM 1.7. — *The system constructed in Section 2 is mild mixing.*

Our system will be the induced transformation of an odometer on an explicit subset obtained by removing a countable number of cylinders. First consider the odometer

$$\Omega = \prod_{i \in \mathbb{N}} \{0, \ldots, 9\}$$

on which $S : \Omega \to \Omega$ is the usual "addition with carry". The product of the uniform measures on $\{0, \ldots, 9\}$, denoted $\nu$, is $S$-invariant and the system $(\Omega, \nu, S)$ is ergodic. Put

$$W_i = \{x \in \Omega : x(1) = 7, \ldots, x(i-1) = 7, x(i) = 8\}$$

for all $i \geqslant 2$ and $W_1 = \{x \in \Omega : x(1) = 8\}$. The closed subset

$$X = \Omega \backslash \bigcup_{i \in \mathbb{N}} W_i$$

has positive measure and we can therefore consider the induced system $(X, \mu, T)$ where $\mu$ is the normalization of the restriction of $\nu$ to $X$ and $T$ is the first return map defined by

$$T(x) = S^{\min\{n \in \mathbb{N} : S^n(x) \in X\}}(x)$$

for (in this case) all $x \in X$.

The removal of the cylinder sets $W_i$ dramatically alters the dynamical properties of the system. While $(\Omega, \nu, S)$ is rigid, the system $(X, \mu, T)$ is mild mixing. This is due to the abundance for any large $n \in \mathbb{N}$ of pairs of nearby points $x, y$ with the property that $T^n(x)$ and $T^{n+1}(y)$ are also nearby. For example, if $n = 10^i$ and $x(i) = 8$ and $y(i) = 9$ and $x(j) = y(j)$ for all $j < i$ and $(x(1), \ldots, x(i-1))$ precedes $(7, \ldots, 7)$ lexicographically in the collection of cylinders that intersect $X$ then:

- $x$ and $y$ are close,
- there will be $1 \leqslant k \leqslant n$ with $S^k(x) \in W_i$,
- for no $1 \leqslant k \leqslant n$ will $S^k(y)$ belong to any $W_j$,

so $T^n(x)$ will be close to $T^{n+1}(y)$. One can use such points to show that every rigid function is approximately $T$-invariant and therefore constant by ergodicity. For more general $n$ the decimal expansion of $n$ dictates the scale $i$ for which the removed cylinder $W_i$ can be used to "pick up invariance" in the above sense.

It is likely, using an argument analogous to that of [FHZ05] for showing that linear recurrent 3-IETs have minimal self-joinings, that pairs of points with the above behaviour can be used to prove the system $(X, \mu, T)$ has the much stronger property of minimal self-joinings.

To produce a system that does not have minimal self-joinings we build on a method from [CE21] for producing ergodic joinings that are neither one of the off-diagonal joinings nor the product measure. To do so we modify $\Omega$ to be of the form

$$\Omega = \prod_{i \in \mathbb{N}} \{0, \ldots, c(i)\}$$

where $c(i) = 9$ unless $i = 2^{j+5}$ in which case $c(i) = 2^{j+6} - 1$; and by modifying $W_i$ when $i = 2^{j+5}$ to consist of many more cylinders. The formal definition is in Section 2.

In modifying the odometer to allow the removal of many cylinders at certain scales we are able to find for each $k \in \mathbb{N}$ sets $A_k$ and $B_k$ of measure about 0.5 and times $\mathsf{p}(k)$, $\mathsf{q}(k)$ with

$$(1.1) \qquad\qquad T^{\mathsf{p}(k)} \approx I \text{ on } A_k \quad \text{and} \quad T^{\mathsf{p}(k)} \approx T^{\mathsf{q}(k)} \text{ on } B_k$$

and the relation $q_{k+1} = 2p_k - q_k$. We use these properties to build a non-trivial joining in Section 4. See in particular the paragraph after Proposition 4.3, for how these properties facilitate building the joining.

Having modified the sets $W_i$ it is no longer clear that the resulting system $(X, \mu, T)$ is mild mixing; the bulk of our effort is in verifying that property. We wish to pursue the same strategy as described above, but the decimal expansion of a time $n$ may no longer be relevant. Instead one must expand

$$(1.2) \qquad n = e(\mathsf{J}(n))\, \mathsf{p}(\mathsf{J}(n)) + e(\mathsf{J}(n) - 1)\, \mathsf{p}(\mathsf{J}(n) - 1) + \cdots + e(1)\, \mathsf{p}(1)$$

where $k \mapsto \mathsf{p}(k)$ is a specific sequence of quickly growing return times for our system $(X, \mu, T)$.

When the scale $\mathsf{J}(n)$ of the expansion (1.2) is not of the form $2^{j+5}$ we can again "pick up invariance" using the single removed cylinder $W_{\mathsf{J}(n)}$. However, if $\mathsf{J}(n)$ is of the form $2^{j+5}$ we cannot take this approach: the set $W_{\mathsf{J}(n)}$ that we remove consists in this case of too many cylinders to run that argument. To overcome this difficulty we appeal to the relations (1.1) which will enable us to "reduce" our attention to $n - e(\mathsf{J}(n))\, \mathsf{p}(\mathsf{J}(n))$ and $n - (\mathsf{p}(\mathsf{J}(n)) - \mathsf{q}(\mathsf{J}(n)))e(\mathsf{J}(n))$ on $A_{\mathsf{J}(n)}$ and $B_{\mathsf{J}(n)}$ respectively. A single application of this reduction procedure may not suffice, as these reduced times may themselves not have expansions at scales where one can pick up invariance. Moreover, even after several iterations of this reduction procedure we might not be in a position to pick up invariance. To overcome this difficulty we consider also the reductions of a multiple $sn$ of $n$. Our main technical result (Proposition 5.20) keeps track of what happens as we apply the reduction procedure to $n$ and $sn$ simultaneously. The upshot is that the reduction of $sn$ is roughly equal to $s$ multiplied by the reduction of $n$. Fixing in advance a relatively small scale compared to that of $n$ we choose $s$ large enough that $s$ multiplied by the reduction of $n$ must "overflow" and have an expansion at which one can again pick up invariance.

This approach is quite technical, and it is natural to wonder whether the careful bookkeeping is necessary. Two facts suggest this care is important. The first fact is the real possibility of "secret" rigidity times. Such "secret" times are present in exchanges of three intervals (see for instance Case 2 in the proof of [FHZ05, Theorem 5.1]) and transformations such as Katok's map (see [Fer97, Section 1.4.3]). The second fact is that there are some natural candidates for rigidity sequences for our transformation: we build our joining by showing that there is a sequence $n_1, n_2, \ldots \in \mathbb{Z}$ so that the corresponding sequence of off-diagonal joinings is Cauchy in the weak* topology; it is then natural to wonder if say, $n_{i+1} - n_i$ is a rigidity sequence. It is an open question of King [Kin86, p. 382] whether rank one transformations always have that $\{(id \times T^n)_* \mu\}_{n \in \mathbb{Z}}$ is weakly dense in the set of ergodic self-joinings.

We hope that our strategy will help to further differentiate mildly mixing rank one transformations from mixing rank one transformations. We comment further on this in Remark 3.1.

## 1.2. Outline of paper

In Section 2 we define our system $(X, \mu, T)$ and introduce the numeration system relevant to it. Additionally, we prove some basic properties of the system including that it is partially rigid and weakly mixing. We prove in Sections 3 and 4 that our system is rank one and has a non-trivial ergodic joining, respectively. The most involved argument (that $(X, \mu, T)$ is mild mixing) is in Section 5. We include two appendices: one proves a minor modification of work in [CE21] needed in Section 4 and the other lists notation used throughout the paper.

## 2. Building the system

In this section we define our measure-preserving system $(X, \mu, T)$ from an odometer $(\Omega, \nu, S)$ by taking $X$ to be a subset of $\Omega$ and $T$ to be the induced transformation and $\mu$ to be the normalized restriction of $\nu$ to $X$. In fact $X$ will be $\Omega$ less a countable union of odometer cylinders and thus a closed subset of $\Omega$.

### 2.1. An odometer

Define
$$d(j) = 2^{j+5}$$
$$a(j) = 2^{j+5}$$
for all $n \in \mathbb{N}$. Define
$$c(n) = \begin{cases} 9 & n \notin \{a(j) : j \in \mathbb{N}\} \\ 2d(j) + 1 & n = a(j) \end{cases}$$

for all $n \in \mathbb{N}$ and put

$$\Omega = \prod_{n \in \mathbb{N}} \{0, \dots, c(n)\}$$

with the product topology.

Given $x, y \in \Omega$ and $k \in \mathbb{N}$ write

$$x \overset{k}{=} y$$

when $x(i) = y(i)$ for all $1 \leqslant i \leqslant k$.

For concreteness, define a metric on $\Omega$ by

$$\mathsf{d}(x, y) = \begin{cases} 0 & \text{if } x = y \\ 2^{-j} & \text{where } j = \min\{i : x(i) \neq y(i)\} \end{cases}$$

for all $x, y \in \Omega$. Define $S : \Omega \to \Omega$ as follows: put $S(c) = 0$ and, for every $x \neq c$ define

$$S(x(1), x(2), \dots) = (0, \dots, 0, x(j) + 1, x(j + 1), \dots)$$

where $j = \min\{k \in \mathbb{N} : x(k) < c(k)\}$. The map $S$ is a homeomorphism of $\Omega$ and preserves the product Borel measure

$$\nu = \bigotimes_{k \in \mathbb{N}} \frac{1}{c(k) + 1} \big(\delta_0 + \dots + \delta_{c(k)}\big)$$

on $\Omega$.

By a *cylinder* we mean any subset of $\Omega$ of the form

$$\{x \in \Omega : x(i(1)) = b(1), \dots, x(i(k)) = b(k)\}$$

for some $k \in \mathbb{N}$ and values $0 \leqslant b(j) \leqslant c(i(j))$ for each $1 \leqslant j \leqslant k$. We call $i(1), \dots, i(k)$ the *defining indices* of the cylinder. The *length* of a cylinder is the numbers of coordinates it fixes. We sometimes write $|C|$ for the length of the cylinder $C$. We use the notation

(2.1)  $[b(1) \ b(2) \dots b(k)]$
$$= \{x \in \Omega : x(1) = b(1) \bmod c(1), \dots, x(k) = b(k) \bmod c(k)\}$$

given prescribed values $b(1), \dots, b(k) \in \mathbb{N} \cup \{0\}$ for cylinders defined by the first $k$ coordinates. We for example write $[0^k]$ for the cylinder $[00 \dots 0]$ of length $k$ and $[7^{k-1}8]$ for the cylinder $[77 \dots 78]$ of length $k$.

We sometimes refer to the index $n \in \mathbb{N}$ defining an odometer digit as a *scale*. When $n$ belongs to $\{a(j) : j \in \mathbb{N}\}$ we call $n$ an *unbounded* scale; otherwise $n$ is a *bounded* scale. For any index $n \geqslant 2$ we picture the odometer as a collection of $t(n)$ cylinders arranged as in Figure 2.1 mapped to one another by the transformation $S$.

For each $s \in \mathbb{N}$ define $\lambda(s)$ to be the largest unbounded scale strictly smaller than $s$. Thus $\lambda(a(j) + 1) = a(j)$ and $\lambda(a(j)) = a(j - 1)$. For example $\lambda(2^{j+5}) = 2^{j+4}$. We will sometimes iterate $\lambda$ so that $\lambda^2(a(j) + 1) = a(j - 1)$. For example $\lambda^3(2^{j+5} + 6) = 2^{j+3}$.

Put

$$t(n) = \prod_{j=1}^{n} 1 + c(j)$$

for each $n \in \mathbb{N}$. Note that $t(n)$ is the minimal $m \in \mathbb{N}$ with $S^m[0^{n+1}] = [0^n 1]$.
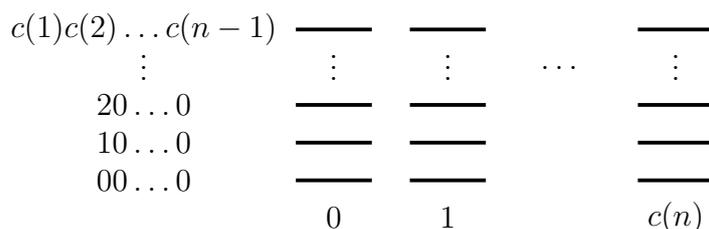
*Figure 2.1. A partition of an $\Omega$ into cylinders (drawn schematically as intervals) of length $n$. Rows are labeled by the first $n-1$ coordinates and columns are labeled by the $n^{\text{th}}$ coordinate. The transformation $S$ maps a cylinder to the one above, with the cylinder at the top of each column mapped to the bottom cylinder of the tower to the right, and the top right cylinder mapped to the bottom left cylinder.*

## 2.2. Inductive removals to define the system

For each $k \in \mathbb{N}$ we define a union $W_k$ of cylinders of length $k$ that will be removed from $\Omega$ to form our space $X$. When $k \notin \{a(n) : n \in \mathbb{N}\}$ the set $W_k$ will consist of the single cylinder $[7^{k-1}8]$ of length $k$. When $k$ is equal to some $a(n)$ the set $W_k$ will consist of a large number of cylinders of length $k$ defined by an inductive procedure. The formal construction follows.

Put $Y_0 = \Omega$. Put $W_1 = [8]$ and for each $1 < k < a(1)$ define

$$W_k = \{x \in \Omega : x(i) = 7, \quad \forall \, 1 \leqslant i < k \quad \text{and} \quad x(k) = 8\} = \left[7^{k-1}8\right]$$

so that $\nu(W_k) = 1/t(k)$ for all $1 \leqslant k < a(1)$. Write

$$Y_k = \Omega \setminus (W_1 \cup \cdots \cup W_k)$$

and let $S|Y_k$ be the induced transformation on $Y_k$ for all $1 \leqslant k < a(1)$. See Figure 2.2 for a schematic. For each $1 \leqslant k < a(1)$ define $\mathsf{q}(k) = 0$ and $\mathsf{p}(k)$ to be the minimal $m \in \mathbb{N}$ with $(S|Y_k)^m[0^{k-1}0] = [0^{k-1}1]$. For example $\mathsf{p}(1) = 1$ and $\mathsf{p}(2) = 9$ and $\mathsf{p}(3) = 89$.

Put $\mathsf{Q}(1) = \mathsf{q}(a(1)) = 1$ and define

$$W_{a(1)} = \bigcup_{j=1}^{\mathsf{Q}(1)} \left\{x \in Y_{a(1)-1} : x(a(1)) \geqslant d(1) + 1 \quad \text{and} \quad \left(S|Y_{a(1)-1}\right)^j(x) \in \left[0^{a(1)-1}\right]\right\}$$

$$= \bigcup_{i=d(1)+1}^{2d(1)+1} \left[9^{a(1)-1}i\right]$$

cf. Figure 2.3. We have $\nu(W_{a(1)}) = (d(1) + 1)\,\mathsf{Q}(1)/t(a(1))$. Put

$$Y_{a(1)} = \Omega \setminus \left(W_1 \cup \cdots \cup W_{a(1)}\right)$$

and let $S|Y_{a(1)}$ be the transformation induced on $Y_{a(1)}$ by $S$. Analogous to what we did before, define $\mathsf{p}(a(n))$ to be the minimal $m \in \mathbb{N}$ with $(S|Y_{a(1)})^m[0^{a(1)-1}0] = [0^{a(1)-1}1]$. Write $\mathsf{P}(1) = \mathsf{p}(a(n))$.
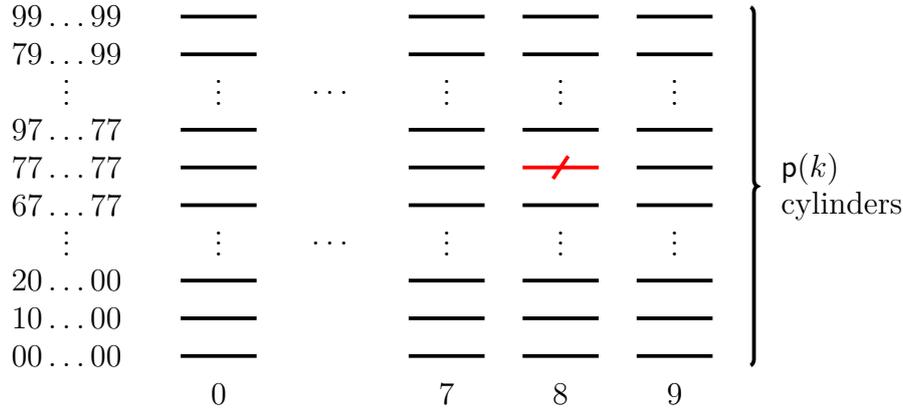
*Figure 2.2. The picture of $Y_{k-1}$ at scale $k$ for some $1 < k < a(1)$. The marked cylinder is $W_k$. Note that the rows labeled $89 \ldots 99$ and $87 \ldots 77$ are not shown as they were removed via $W_1$.*
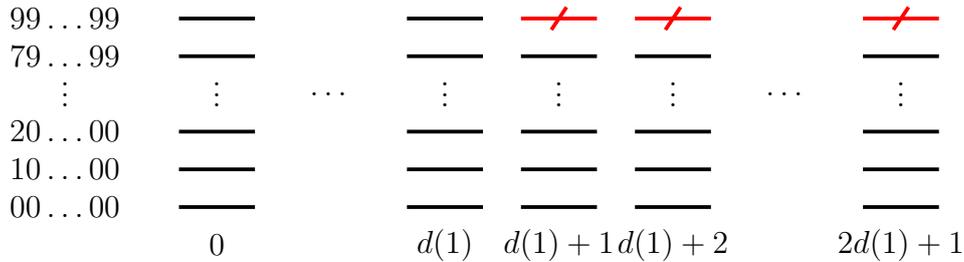


*Figure 2.3. The picture of $Y_{a(1)-1}$ at scale $a(1)$. The marked cylinders constitute $W_{a(1)}$. Note that the row labeled $89 \ldots 99$ is not shown as it was removed via $W_1$.*

We now proceed inductively, supposing that subsets $W_1, \ldots, W_{a(n)}$ of $\Omega$ and $\mathsf{p}(1), \ldots, \mathsf{p}(a(n))$ in $\mathbb{N}$ and $\mathsf{q}(1), \ldots, \mathsf{q}(a(n))$ in $\mathbb{N}$ have been defined, with $Y_i = \Omega \setminus (W_1 \cup \cdots \cup W_i)$ for all $1 < i \leqslant a(n)$ and $\mathsf{P}(i) = \mathsf{p}(a(i))$ and $\mathsf{Q}(i) = \mathsf{q}(a(i))$ for all $1 \leqslant i \leqslant n$. Define

$$W_k = \{x \in \Omega : x(k) = 7 \quad \forall \ 1 \leqslant i < k \quad \text{and} \quad x(k) = 8\} = \left[ 7^{k-1} 8 \right]$$

for each $a(n) < k < a(n+1)$. Note that $W_k$ was not removed at any earlier stage. We have

$$(2.2) \qquad\qquad\qquad \nu(W_k) = \frac{1}{t(k)}$$

for all $a(n) < k < a(n+1)$. Let

$$Y_k = \Omega \setminus (W_1 \cup \cdots \cup W_k) = Y_{a(n)} \setminus \left( W_{a(n)+1} \cup \cdots \cup W_k \right)$$

for all $a(n) < k < a(n+1)$. Define $\mathsf{p}(k)$ to be the minimal $m \in \mathbb{N}$ with

$$(S|Y_k)^m \left[ 0^{k-1} 0 \right] = \left[ 0^{k-1} 1 \right]$$

and define $\mathsf{q}(k) = 0$ for all $a(n) < k < a(n+1)$.

Next, let $\mathsf{Q}(n+1) = \mathsf{q}(a(n+1)) = 2\mathsf{P}(n) - \mathsf{Q}(n)$ and put

$$W_{a(n+1)} = \bigcup_{j=1}^{\mathsf{Q}(n+1)} \left\{ \begin{array}{l} x \in Y_{a(n+1)-1} : x(a(n+1)) \geqslant d(n+1) + 1 \\ \qquad\qquad \text{and} \quad \left(S|Y_{a(n+1)-1}\right)^j(x) \in \left[0^{a(n+1)-1}\right] \end{array} \right\}$$

which is shown in Figure 2.4. We have

(2.3)
$$\nu\left(W_{a(n+1)}\right) = \frac{(d(n+1)+1)\,\mathsf{Q}(n+1)}{t(a(n+1))}$$

Let

$$Y_{a(n+1)} = \Omega \setminus \left(W_1 \cup \cdots \cup W_{a(n+1)}\right)$$

and define $\mathsf{p}(a(n)+1) \in \mathbb{N}$ to be minimal satisfying

$$\left(S|Y_{a(n+1)}\right)^{\mathsf{p}(a(n+1))}\left[0^{a(n+1)-1}0\right] = \left[0^{a(n+1)-1}1\right]$$

and put $\mathsf{P}(n+1) = \mathsf{p}(a(n+1))$ to complete the inductive construction. It is immediate from Figures 2.2 and 2.4 that

(2.4)
$$(c(n)+1)\,\mathsf{p}(n) \leqslant t(n)$$

for all $n \in \mathbb{N}$ and in particular, $t(n)$ is the number of levels of $Y$ at stage $n$ and $p(n)$ is morally the number of levels of $X$ at $p(n-1)$.

LEMMA 2.1. — $\sum_{k \in \mathbb{N}} \nu(W_k) < \frac{1}{8}$.

*Proof.* — First estimate

(2.5)
$$\begin{aligned} \frac{(d(n)+1)\,\mathsf{Q}(n)}{t(a(n))} &= \frac{d(n)+1}{c(a(n))+1} \cdot \frac{\mathsf{Q}(n)}{t(a(n))-1} \leqslant \frac{2\mathsf{P}(n-1)}{t(a(n))-1} \leqslant \frac{t(a(n-1))}{t(a(n))-1} \\ &= \frac{1}{10^{a(n)-a(n-1)-1}} \end{aligned}$$

using (2.4). Then use (2.2) and (2.3) to get

$$\sum_{k \in \mathbb{N}} \nu(W_k) \leqslant \sum_{k \in \mathbb{N}} \frac{1}{t(k)} + \sum_{n \in \mathbb{N}} \frac{(d(n)+1)\,\mathsf{Q}(n)}{t(a(n))} \leqslant \sum_{n \in \mathbb{N}} \frac{1}{10^n} + \frac{1}{10^{a(n+1)-a(n)-1}}$$

which is less than $1/8$. □

By Lemma 2.1 the set

$$X = \Omega \setminus \bigcup_{k \in \mathbb{N}} W_k$$

has positive measure. It is a closed subset of $\Omega$ as each $W_k$ is open. The set

$$G = \left\{ \begin{array}{l} x \in X : S^n(x) \in X_0 \quad \text{for infinitely many } n \in \mathbb{N} \\ \qquad \text{and } S^{-n}(x) \in X_0 \quad \text{for infinitely many } n \in \mathbb{N} \end{array} \right\}$$

has full measure in $X$ by Poincaré recurrence. We define $T : X \to X$ by taking $T$ to be the first return map on $G$ and the identity map on $X \setminus G$. Define $\mu(B) = \nu(B \cap X)/\nu(X)$ for every set $B \subset X$ of the form $C \cap X$ with $C$ a Borel subset of $\Omega$. Our system is $(X, \mu, T)$. As $(\Omega, \nu, S)$ is ergodic $(X, \mu, T)$ is as well.
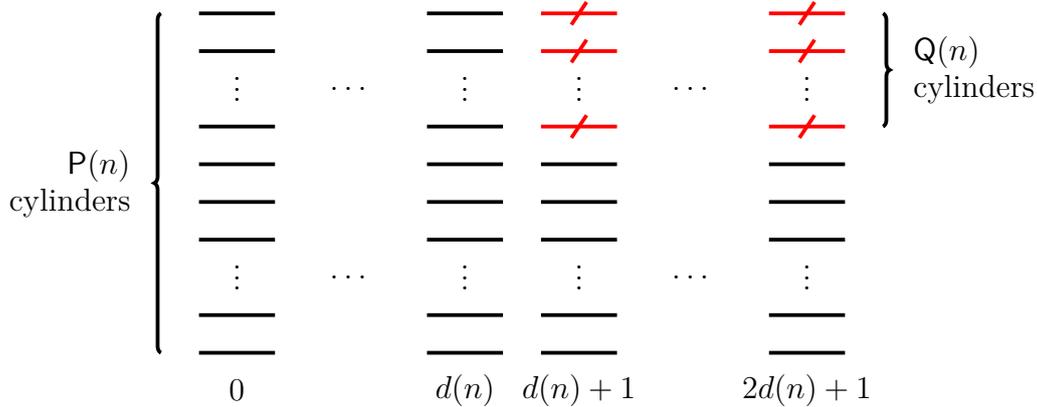
*Figure 2.4. The picture of $Y_{a(n)-1}$ at scale $a(n)$. The marked cylinders consti-
tute $W_{a(n)}$.*

## 2.3. Measures of cylinders

We will note that our choices for $d(n)$ and $a(n)$ give the following properties.

(P1)
$$\mathsf{P}(n) \leqslant t(a(n) - 1)$$

(P2)
$$\sum_{n=1}^{\infty} \frac{\mathsf{P}(n)}{\mathsf{P}(n+1)} < \frac{1}{10^5}$$

We will want often to study partial $T$ orbits as orbits of some $S|Y_n$. For convenience we define for any $r \in \mathbb{Z}$ and any $b \in \mathbb{N}$

$$\mathsf{V}(b, r) = \bigcup_{a(m)>b} \left\{ x \in X : (S|Y_b)^i(x) \in W_{a(m)} \quad \text{for some} \quad -|r| \leqslant i \leqslant |r| \right\}$$

to consist of those points whose $S|Y_b$ orbits visit some finer $W_{a(m)}$ between times $-|r|$ and $|r|$. When $x$ does not belong to $\mathsf{V}(b, r)$ we have

$$T^i(x) = (S|Y_b)^i(x) \quad \forall -|r| \leqslant i \leqslant |r|$$

unless some $(S|Y_b)^i(x)$ belongs to some $[7^i 8]$ with $i \geqslant b$.

If a cylinder set $C$ of length $k$ does not belong to $W_1 \cup \cdots \cup W_k$ then $\mu$ assigns it positive measure. In other words, any cylinder set $C$ not removed by scale $|C|$ has positive measure in $X$. Slightly more is true: the following lemma will be used to prove $\mu(C)t(|C|)$ is bounded away from zero over such cylinder sets.

LEMMA 2.2. — *If $C \neq [7^{|C|}]$ is a cylinder not contained in $W_1 \cup \cdots \cup W_{|C|}$ then*

$$\nu(C \setminus X) \leqslant \frac{4}{t(|C|)} \sum_{a(n)>|C|} \frac{\mathsf{P}(n-1)}{\mathsf{P}(n)}$$

*holds.*

*Proof.* — Fix a cylinder set $C \neq [7^{|C|}]$ disjoint from $W_1 \cup \cdots \cup W_{|C|}$. We can only possibly have $C \cap W_j \neq \varnothing$ when $j = a(n) > |C|$ for some $n \in \mathbb{N}$. For any such $n$, note that $C$ consists of $t(a(n))/t(|C|)$ cylinders of length $a(n)$ and that at most

$$\frac{2(d(n)+1)\,\mathsf{Q}(n)}{(2d(n)+1)\mathsf{P}(n)} \leqslant \frac{2\,\mathsf{Q}(n)}{\mathsf{P}(n)} \leqslant \frac{4\mathsf{P}(n-1)}{\mathsf{P}(n)}$$

of them belong to $W_{a(n)}$. Therefore

$$\nu\big(C \cap W_{a(n)}\big) \leqslant \frac{4}{t(|C|)} \frac{\mathsf{P}(n-1)}{\mathsf{P}(n)}$$

from which the conclusion follows. $\qquad\square$

LEMMA 2.3. — *There is $\xi > 0$ such that*

$$\mu(C) \geqslant \frac{\xi}{t(|C|)}$$

*whenever $C$ is a cylinder set not contained in $W_1 \cup \cdots \cup W_{|C|}$.*

*Proof.* — If $C = [7^{|C|}]$ then

$$\nu(C \cap W_n) \leqslant \frac{1}{t(n)}$$

whenever $n > |C|$ is a bounded scale. Thus for every $C$ not contained in $W_1 \cup \cdots \cup W_{|C|}$ we have

$$\nu(C \setminus X) \leqslant \sum_{n > |C|} \frac{1}{t(n)} + \frac{4}{t(|C|)} \sum_{a(n) > |C|} \frac{\mathsf{P}(n-1)}{\mathsf{P}(n)} \leqslant \frac{1}{t(|C|)}\Big(\frac{1}{9} + \frac{4}{10^5}\Big)$$

after Lemma 2.2 and P2. Thus

$$\xi = \frac{55}{63} \frac{1}{\nu(X)}$$

suffices. $\qquad\square$

LEMMA 2.4. — *Let $D \subset \Omega$ be a cylinder whose largest defining index is $r$ and let $C \subset \Omega$ be a cylinder whose smallest defining index is $r + k$ for some $k \geqslant 1$. Further assume $D \cap X \neq \varnothing$ and $D \cap C \cap X \neq \varnothing$. Then*

$$\frac{\mu(D \cap C \cap X)}{\mu(D)\mu(C)} \geqslant 1 - \frac{1}{2^k}$$

*holds.*

*Proof.* — We first consider the case when $D$ has every index defined up to $r$ and the indices defining $C$ are from $r+1$ to $b$. (This is a special version of the case $k=1$.) Thus $C \cap D$ is a single cylinder defined at entries 1 to $b$ that is non-empty. Thus $C \cap D \subset Y_b$ and so $\mu(C \cap D) = \nu(X)^{-1}(\frac{1}{t(b)} - \nu(C \cap D \cap X^c))$. Thus we compute

$$\nu(Y_b \setminus X) \leqslant \sum_{j > b} \frac{1}{t(j)} + \sum_{a(j) > b} \frac{4\mathsf{P}(j-1)}{\mathsf{P}(j)} < \frac{1}{8}\nu(C \cap D) = \frac{1}{8}\frac{1}{t(b)}$$

so

$$\mu(C \cap D) \geqslant \left(\frac{7}{8}\right)^2 \frac{1}{t(r)} \frac{1}{\nu(X)} \frac{t(r)}{t(b)} \frac{1}{\nu(X)} \geqslant \left(\frac{7}{8}\right)^2 \mu(C)\mu(D)$$

establishing the first case.

We now treat $D$ defined at all indices at most $r$ and $C$ is defined at all indices from $r + k$ to $b$. We consider $D \cap Y_{r+k-1}$, a union of $N \geqslant 8^k$ cylinders. Now

$$(2.6) \qquad \left| \mu(D) - \frac{N}{\mathsf{p}(r+k+1)} \right| \leqslant \max\left\{ \nu(Y_{r+k-1} \setminus X)\nu(X)^{-1}, \ \frac{\nu(X)^{-1}}{\nu(Y_{r+k-1})^{-1}} \right\}.$$

Indeed we estimate the maximum amount that could be removed by going from $Y_{r+k-1}$ to $X$ and assume it is all removed from $D$ and on the other hand, if nothing is removed from $D$, how much we have to scale it to obtain its proportion of $X$ from its proportion of $Y_{r+k-1}$. Similarly

$$(2.7) \qquad \left| \mu(C) - \frac{\mathsf{p}(r+k+1)}{\mathsf{p}(b+1)} \right| \leqslant \max\left\{ \nu(Y_b \setminus X)\nu(X)^{-1}, \ \frac{\nu(X)^{-1}}{\nu(Y_b)^{-1}} \right\}.$$

At most one of the $N$ cylinders defined at entries $1$ to $r+k-1$ intersects $\bigcup_{j \notin \{a(k)\}} W(j)$. For each of $N-1$ other cylinders, we intersect them with $C$ and obtain cylinders defined at entries from $1$ to $b$ which do not intersect $W_j$ for $j \notin \{a(i)\}_{i\in\mathbb{N}}$. Name these cylinders $E_1, \ldots, E_{N-1}$ and note that

$$\nu\left( \bigcup_{i=1}^{N-1} E_i \cap C \cap Y_b^c \right) < \sum_{a(j)>b} \frac{4}{t(b)} \frac{\mathsf{P}(j-1)}{\mathsf{P}(j)} < \frac{1}{2^{b+5}} \frac{1}{\mathsf{p}(b+1)}$$

holds. We obtain that

$$\mu(C \cap D) \geqslant \frac{N-1}{\mathsf{p}(b+1)} - \frac{1}{2^{b+5}} \frac{1}{\mathsf{p}(b+1)}.$$

Since $N \geqslant 8^k$, combining this with (2.6) and (2.7) we have the claim.

The general case is similar: treat $D$ as a union of $A$ cylinders defined on every index from $1$ to $r + k - 1$ and $C$ as a union of $B$ cylinders defined on every index from $r + k$ to $b$. □

## 2.4. Expansion and reduction

Given $r \in \mathbb{Z}$ we will analyze $T^r$ by expanding $r$ using a numeration system closely tied to the dynamics of $T$. The numeration system is based on the heights $\mathsf{p}(j)$ of the left-most towers in the scale $j$ picture. At unbounded scales we might instead have chosen to use $\mathsf{p}(j) - \mathsf{q}(j)$ in our numeration system as each height is witnessed by about half of the points in the system. Although we can only use one in our numeration system, we will in effect keep track of both in the technical parts of our arguments using the reductions in Definition 2.11. To define this expansion we need the following lemma.

LEMMA 2.5. — *We have*

$$\left(\frac{c(n)+1}{2}+1\right)\mathsf{p}(n) \leqslant \mathsf{p}(n+1)$$

*for all $n \in \mathbb{N}$.*

*Proof.* — We have

$$\mathsf{p}(n+1) = \begin{cases} (c(n)+1)\,\mathsf{p}(n) - 1 & n \text{ a bounded scale} \\ (c(n)+1)\,\mathsf{p}(n) - \mathsf{q}(n)(c(n)+1)/2 & n \text{ an unbounded scale} \end{cases}$$

from which the statement follows. $\qquad\square$

DEFINITION 2.6. — *Fix $r$ in $\mathbb{Z}$. By the* tower expansion *of $r$ we mean the expansion*

(2.8)
$$r = \sum_{j=1}^{\mathsf{J}(r)} e(j)\,\mathsf{p}(j)$$

*with $e(\mathsf{J}(r)) \neq 0$ and $2|e(j)| \leqslant c(j)+1$ chosen greedily. Formally, this means the following: take*

$$\mathsf{J}(r) = \max\{j \in \mathbb{N} : e \longmapsto |r - e\,\mathsf{p}(j)| \quad \text{is not minimized by } e = 0\}$$

*with the convention that $\mathsf{J}(0) = 0$. Define $\mathsf{e}(r) = e(\mathsf{J}(r))$ to be the associated minimizing value, choosing $|\mathsf{e}(r)|$ as large as possible whenever there is a choice. Since $\mathsf{p}(1) = 1$ both $1$ and $-1$ have tower expansions, and since $|r - \mathsf{e}(r)\,\mathsf{p}(\mathsf{J}(r))| < |r|$ one can define the tower expansion of any integer by induction on $|r|$. Indeed letting $r' = r - e(r)J(r)$ we can inductively (using $|r'| < |r|$) assume $r' = \sum_{i=0}^{J(r')} e(i)p(i)$ and represent $r = e(r)p(\mathsf{J}(r)) + \sum_{i=0}^{J(r')} e(i)p(i)$. Since $J(r') < J(r)$ this gives an expansion for $r$. We will refer to $\mathsf{J}(r)$ as the* scale *of $r$. We will say that $r$ is* bounded *if $\mathsf{J}(r)$ is a bounded scale, and that $r$ is* unbounded *if $\mathsf{J}(r)$ is an unbounded scale.*

*Example 2.7.* — We have $\mathsf{p}(1) = 1$, $\mathsf{p}(2) = 9$, $\mathsf{p}(3) = 89$, $p(4) = 889$ and $c(i)+1 = 10$ for $i < a(1)$ so

$$74 = 3\,\mathsf{p}(1) - 2\,\mathsf{p}(2) + \mathsf{p}(3)$$
$$150 = -\,\mathsf{p}(1) - 3\,\mathsf{p}(2) + 2\,\mathsf{p}(3)$$

are tower expansions.

We conclude this section with some basic results about our expansion.

LEMMA 2.8. — *If $2|e(j)| \leqslant c(j)+1$ for all $1 \leqslant j \leqslant J$ then*

$$\left|\sum_{j=1}^{J} e(j)\,\mathsf{p}(j)\right| \leqslant \left(\frac{c(J)+1}{2}+1\right)\mathsf{p}(J)$$

*holds.*

*Proof.* — The proof is by induction. One calculates

$$\left| \sum_{j=1}^{K+1} e(j)\, \mathsf{p}(j) \right| \leqslant \left( \frac{c(K)+1}{2} + 1 \right) \mathsf{p}(K) + \frac{c(K+1)+1}{2}\, \mathsf{p}(K+1)$$

and applies Lemma 2.5. $\qquad\square$

COROLLARY 2.9. — *For all $r \in \mathbb{Z}$ we have $|r| \leqslant \frac{3}{4}\, \mathsf{p}(\mathsf{J}(r)+1)$.*

*Proof.* — This follows from Lemma 2.5. If $J = \mathsf{J}(r)$ is a bounded scale then this follows because $\mathsf{p}(J+1) = 10\,\mathsf{p}(J) - 1$ and thus $(\frac{c(J)+1}{2}+1)\,\mathsf{p}(J) = 6\,\mathsf{p}(J) < \frac{3}{4}\,\mathsf{p}(J+1)$. Similarly if $J = a(k)$ is an unbounded scale, $\mathsf{p}(J+1) = 2(d(k)+1)\,\mathsf{p}(J) - d(k)\,\mathsf{q}(J)$ and so $(\frac{2d(k)+2}{2}+1)\,\mathsf{p}(J) < \frac{3}{4}\,\mathsf{p}(J+1)$. $\qquad\square$

LEMMA 2.10. — *For every $m \in \mathbb{Z}$ with $\mathsf{J}(m)$ unbounded we have*

$$2|\,\mathsf{e}(m)\,\mathsf{q}(\mathsf{J}(m))| \leqslant \mathsf{p}(\mathsf{J}(m) - 1).$$

*Proof.* — Since $|\,\mathsf{e}(m)| \leqslant \frac{c(\mathsf{J}(m))+1}{2}$ we can estimate

$$|\,\mathsf{e}(\mathsf{J}(m))\,\mathsf{q}(\mathsf{J}(m))| \leqslant \frac{c(\mathsf{J}(m))+1}{2} \cdot 2\,\mathsf{p}(\lambda(\mathsf{J}(m))) \leqslant 10^{-2^4}\,\mathsf{p}(\mathsf{J}(m) - 1)$$

as required. $\qquad\square$

When $r$ is at an unbounded scale a definite proportion of points $x$ with $x(\mathsf{J}(r)) < c(\mathsf{J}(r))/2$ satisfy $T^{\mathsf{p}(\mathsf{J}(r))}(x) \approx x$ (see Corollaries 5.14(1) and 5.16(1)) and a definite proportion of points $x$ with $x(\mathsf{J}(r)) > c(\mathsf{J}(r))/2$ satisfy $T^{\mathsf{p}(\mathsf{J}(r))-\mathsf{q}(\mathsf{J}(r))}(x) \approx x$ (see Corollaries 5.14(2) and 5.16(2)). Motivated by this, the next definition will be used to relate $T^r(x)$ for $r$ at unbounded scales to $T^{r'}(x)$ for $r'$ at smaller scales.

DEFINITION 2.11. — *Define the* left *and* right reductions *by*

$$\mathsf{L}(r) = r - \mathsf{e}(r)\,\mathsf{p}(\mathsf{J}(r))$$
$$\mathsf{R}(r) = r - \mathsf{e}(r)\,\mathsf{p}(\mathsf{J}(r)) + \mathsf{e}(r)\,\mathsf{q}(\mathsf{J}(r))$$

*respectively whenever $\mathsf{J}(r)$ is an unbounded scale.*

A definite proportion of points $x$ satisfy $T^r(x) \approx T^{\mathsf{L}(r)}(x)$ and a definite proportion of points $x$ satisfy $T^r(x) \approx T^{\mathsf{R}(r)}(x)$ whenever $\mathsf{J}(r)$ is unbounded.

LEMMA 2.12. — *Fix $r \in \mathbb{Z}$ with $\mathsf{J}(r)$ unbounded:*
(1) $|\mathsf{L}(r)| \leqslant \mathsf{p}(\mathsf{J}(r))$,
(2) $|\mathsf{R}(r)| \leqslant \mathsf{p}(\mathsf{J}(r)) - \mathsf{q}(\mathsf{J}(r))$.

*Proof.* — For (1), apply Lemma 2.8 with $J = \mathsf{J}(r) - 1$. For (2), apply Lemma 2.8 with $J = \mathsf{J}(r) - 1$ and Lemma 2.10. $\qquad\square$

The following lemmas describe how reductions and scales interact.

LEMMA 2.13. — *Fix $m \in \mathbb{Z}$ with $\mathsf{J}(m)$ an unbounded scale. We have*

$$\max\{\mathsf{J}(\mathsf{L}(m)), \mathsf{J}(\mathsf{R}(m))\} \geqslant \lambda(\mathsf{J}(m))$$

*and if $\mathsf{L}(m)$ and $\mathsf{R}(m)$ are both at an unbounded scale then neither is at scale $\mathsf{J}(m)$ and at least one of $\mathsf{J}(\mathsf{L}(m)) = \lambda(\mathsf{J}(m))$ and $\mathsf{J}(\mathsf{R}(m)) = \lambda(\mathsf{J}(m))$ holds.*

*Proof.* — Since

(2.9) $\qquad \mathsf{R}(m) - \mathsf{L}(m) = \mathsf{e}(m)\,\mathsf{q}(\mathsf{J}(m)) = 2\,\mathsf{e}(m)\,\mathsf{p}(\lambda(\mathsf{J}(m))) - \mathsf{e}(m)\,\mathsf{q}(\lambda(\mathsf{J}(m)))$

we have in general

(2.10) $\qquad\qquad\qquad \max\{|\mathsf{R}(m)|,\ |\mathsf{L}(m)|\} \geqslant \dfrac{3}{4}\,\mathsf{p}(\lambda(m))$

and so the larger of $\mathsf{R}(m)$, $\mathsf{L}(m)$ in absolute value cannot be at a scale smaller than $\lambda(\mathsf{J}(m))$, establishing the first claim.

We now assume that both $\mathsf{L}(m), \mathsf{R}(m)$ are at unbounded scales. It is immediate that (in general) $\mathsf{J}(\mathsf{L}(m)) < \mathsf{J}(m)$. As $\mathsf{L}(m)$ is assumed to be at an unbounded scale we must have $\mathsf{J}(\mathsf{L}(m)) \leqslant \lambda(\mathsf{J}(m))$. That implies

$$|\mathsf{L}(m)| \leqslant \frac{\mathsf{p}(\lambda(\mathsf{J}(m)) + 1)}{2}$$

and, as we have

$$|\,\mathsf{e}(m)\,\mathsf{q}(\mathsf{J}(m))| \leqslant \mathsf{p}(\mathsf{J}(m) - 1)$$

from Lemma 2.10, the triangle inequality gives

$$2|\mathsf{R}(m)| \leqslant \mathsf{p}(\lambda(\mathsf{J}(m)) + 1) + 2\,\mathsf{p}(\mathsf{J}(m) - 1) < \mathsf{p}(\mathsf{J}(m))$$

which means $\mathsf{J}(\mathsf{R}(m)) < \mathsf{J}(m)$. As $\mathsf{R}(m)$ is also assumed to be at an unbounded scale we have $\mathsf{J}(\mathsf{R}(m)) \leqslant \lambda(\mathsf{J}(m))$ as well. Thus when they are both at an unbounded scale, it must be that one of $\mathsf{R}(m)$ and $\mathsf{L}(m)$ is at scale $\lambda(\mathsf{J}(m))$. $\qquad\square$

LEMMA 2.14. — *Let $m \in \mathbb{Z}$ be at an unbounded scale $a(b)$. At least one of $\mathsf{J}(\mathsf{R}(m)) < \mathsf{J}(m)$ and $|\mathsf{R}(m)| < \frac{3}{4}\,\mathsf{p}(\mathsf{J}(m))$ is true. Also $\mathsf{J}(\mathsf{L}(m)) < \mathsf{J}(m)$.*

*Proof.* — This follows by the greedy algorithm to define $\mathsf{p}$ and the fact that $(2d(b) + 2)\,\mathsf{q}(\mathsf{J}(m))$ is much smaller than $\frac{1}{8}\,\mathsf{p}(\mathsf{J}(m))$. $\qquad\square$

## 2.5. Some dynamical properties

We mention here briefly some dynamical properties of $(X, \mu, T)$ that follow quickly from the construction.

### 2.5.1. Partial rigidity

The task of proving $(X, \mu, T)$ is mild mixing is made difficult by partial rigidity in the system, which we remark upon here. Recall that $(X, \mu, T)$ is *partially rigid* if there is a constant $\delta > 0$ such that, for every measurable $A \subset X$ one has $\mu(A \cap T^n A) \geqslant \delta\mu(A)$ infinitely often. Using Lusin's theorem, one can deduce partial rigidity from the following proposition.

PROPOSITION 2.15. — *For all $\epsilon > 0$*

$$\lim_{n \to \infty} \mu\Big(\Big\{x \in X : d\big(T^{\mathsf{p}(n)}x, x\big) < \epsilon\Big\}\Big) \geqslant \frac{1}{9}$$

*holds.*

Let $k_x : \mathbb{Z} \to \mathbb{Z}$ by $k_x(n) = m$ where $S^m(x) = T^n(x)$. It is not hard to check that

$$\nu(\{x \in X \subset \Omega : k_x(\mathsf{p}(n)) = t(n-1)\}) \geqslant \frac{1}{9}$$

for all large enough $n$. Indeed, this follows from estimating $1 - t(n-1)\nu(\bigcup_{k \geqslant n} W_k)$. Note that there are two cases, one where $n$ is at an unbounded scale and one where $n$ is at a bounded scale.

### 2.5.2. Weak mixing

We prove here that $(X, \mu, T)$ is weak mixing. We follow roughly a well-known argument but give the proof from scratch both for completeness and because the argument proceeds along lines similar to, but much simpler than, those in the proof of our main result. More precisely, in both arguments we study $T$ by relating it to $S|Y_m$ for appropriate $m$. It is particularly important to examine the effect of $W_m$.

PROPOSITION 2.16. — $(X, \mu, T)$ is weakly mixing.

*Proof.* — Fix an eigenfunction $f$ of $T$ with eigenvalue $\lambda$. Recall that ergodicity of $T$ gives $|f(x)| = 1$ almost surely. We must prove that $\lambda = 1$. Fix $\epsilon > 0$ and $s > 0$. There is a compact set $G \subset X$ with $\nu(G) \geqslant 1 - s$ on which $f$ is uniformly continuous. Fix $\delta > 0$ so that $\mathsf{d}(x, y) < \delta$ guarantees $|f(x) - f(y)| < \epsilon$ whenever $\{x, y\} \subset G$. Fix a bounded scale $n \in \mathbb{N}$ with $2^{-(n-1)} < \delta$ and $n - 1$ an unbounded scale. We wish to choose points $x, y$ in $X$ with all the following properties.

(1) $$\left\{x, y, T^{2\mathsf{p}(n)+1}(x),\ T^{2\mathsf{p}(n)}(y)\right\} \subset G$$

(2) $$x(i) = y(i), \quad \forall\, 1 \leqslant i \leqslant n-1$$

(3) $$x(n) = 4 \quad \text{and} \quad x(n) = 7$$

(4) $$T^i(x) = (S|Y_n)^i(x) \quad \text{and} \quad T^i(y) = (S|Y_n)^i(y) \quad \forall\, 1 \leqslant i \leqslant 2\mathsf{p}(n)+1$$

(5) $$f\left(T^{2\mathsf{p}(n)+1}(x)\right) = \lambda^{2\mathsf{p}(n)+1} f(x) \quad \text{and} \quad f\left(T^{2\mathsf{p}(n)}(y)\right) = \lambda^{2\mathsf{p}(n)} f(y)$$

Indeed, if we are able to do this then we have

- $|f(x) - f(y)| < \epsilon$ by properties (1) and (2).
- $T^{2\mathsf{p}(n)+1}(x)$ and $T^{2\mathsf{p}(n)}(y)$ agree in their first $n-1$ coordinates by (2), (3), (4) and property (5). Indeed our third condition and the definition of $W_n$ gives that $(S|Y_n)^{2\mathsf{p}(n)+1}x \overset{n}{=} (S|Y_n)x$ and $(S|Y_n)^{2\mathsf{p}(n)}y \overset{n}{=} (S|Y_n)y$.
- $|f(T^{2\mathsf{p}(n)+1}(x)) - f(T^{2\mathsf{p}(n)}(y))| < \epsilon$ by property (1) and the preceding item from which we conclude that $|1 - \lambda| < 2\epsilon$.

To obtain the five properties we note that $G \cap T^{-2\mathsf{p}(n)}G$ and $G \cap T^{-2\mathsf{p}(n)-1}G$ are both very large if $s$ is small enough, and that $\mathsf{V}(n, 2\mathsf{p}(n)+1)$ will have small measure if $n$ is large, so that we can guarantee the first four properties by shrinking $s$. (5) holds almost surely. $\square$

## 3. The rank one property

We prove here that our system is rank one.

*Proof of Theorem 1.5.* — Fix $A \subset X$ measurable and $\epsilon > 0$. For each $n \in \mathbb{N}$ let $h(n) \in \mathbb{N}$ be minimal with the property that $(S|Y_{a(n)})^{h(n)}[87^{a(n)-1}]$ is the lexicographically largest cylinder of length $a(n)$ in $Y_{a(n)}$. Certainly $h(n) < t(a(n))$. All subsequent removals $W_k$ with $k \notin \{a(m) : m > n\}$ are contained within $[7^{a(n)}]$. Let

$$F_n = \left[87^{a(n)-1}\right] \setminus \mathsf{V}\big(a(n), h(n)\big) \subset X$$

which is shown in Figure 3.1. We now bound the measure removed by removing $\mathsf{V}(a(n), h(n))$. For each $m > n$ we remove from $[87^{a(n)-1}]$ at most $d(m) + 1$ cylinders of length $a(m)$. Thus

$$\mu\Big(X \setminus F_n \cup T^1 F_n \cup \cdots \cup T^{h(n)-1} F_n\Big)$$

$$\leqslant \frac{1}{\nu(X)}\left(8\mathsf{P}(n)\nu\big(\left[0^{a(n)}\right]\big) + \sum_{m>n} h(n)\frac{d(m)+1}{t(a(m))}\right)$$

$$\leqslant \frac{8}{\nu(X)}\frac{t(a(n)-1)}{t(a(n))} + \frac{2}{\nu(X)}\sum_{m>n}\frac{t(a(n))d(m)}{t(a(m))}$$

$$\leqslant \frac{8}{\nu(X)}\frac{1}{2d(n)+2} + \frac{2}{\nu(X)}\sum_{m>n}\frac{1}{10^{a(m)-a(n)-1}}$$

where the term $8\mathsf{P}(n)\nu([0^{a(n)}])$ accounts for the cylinders of length $a(n)$ that precede $[87^{a(n)-1}]$ lexicographically and we have used $\mathsf{P}(n) \leqslant t(a(n) - 1)$. We thus have

$$\mu\Big(F_n \cup TF_n \cup \cdots \cup T^{h(n)} F_n\Big) \longrightarrow 1$$
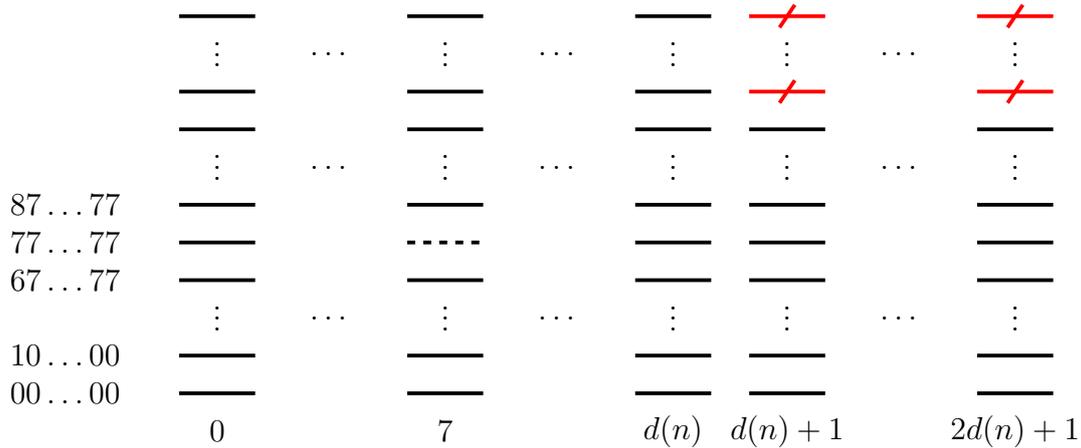
as $n \to \infty$.



Figure 3.1. The picture of $Y_{a(n)-1}$ at scale $a(n)$. The marked cylinders make up $W_{a(n)}$. The dashed cylinder contains all $W_j$ with $j > a(n)$ and $j$ a bounded scale. A subset of $[87^{a(n)-1}]$ will be used as the set $F$ exhibiting $T$ as rank one.

Choose $n$ so large that there exists cylinders $C_1, \ldots, C_t \subset Y$ with defining coordinates $1, \ldots, a(n)$ and

- $\mu(A \triangle (X \cap C_1 \cup \cdots \cup X \cap C_t)) < \epsilon/100\nu(X)$,
- $8\mathsf{P}(n)\nu([0^{a(n)}]) < \epsilon/10$,
- $\mu(X \setminus F_n \cup T^1 F_n \cup \cdots \cup T^{h(n)-1}F_n) < \epsilon/2$,

all hold. Then the $\mu$ measure of the intersection of $A$ with the union of all cylinders that precede $[87^{a(n)-1}]$ lexicographically is at most $\epsilon/10\nu(X)$. For $0 \leqslant i < h(n)$ put $A_i = T^i F_n$ if

$$(C_1 \cup \cdots \cup C_t) \cap T^i F_n \neq \varnothing$$

and $A_i = \varnothing$ otherwise. The set $A' = A_0 \cup \cdots \cup A_{h(n)-1}$ is measurable with respect to the $\sigma$-algebra generated by $\{F_n, TF_n, \ldots, T^{h(n)-1}F_n\}$ and $\mu(A' \triangle A) < \epsilon$. $\qquad\square$

*Remark 3.1.* — It is natural to try to determine whether our procedure can be modified to construct a mildly mixing rank 1 transformation $T$ with the property that $T^n$ has rank two for all $n \geqslant 2$. (Note that $T$ being mild mixing precludes $T^n$ from being rank one: indeed $T$ is in its centralizer and were $T^n$ rank one it could be concluded that $T$ is rigid by the weak closure theorem. We thank Mariusz Lemanczyk for pointing this out.) Indeed, one could try to modify our construction (perhaps by removing some extra levels at certain bounded scales and changing the definition of $\mathsf{Q}(n)$) so that for every $n$ there are infinitely many $k$ so that at the unbounded scales $a(k)$ we have $\mathsf{P}(k)$ and $\mathsf{P}(k) - \mathsf{Q}(k)$ both relatively prime to $n$. One could use the following definitions.

$$F_k = \bigcup_{i=0}^{n-1} (S|Y_{a(k)})^i \left[87^{a(k)-1}\right] \setminus \mathsf{V}(a(k), \mathsf{p}(k) + 2)$$

$$F_k = \bigcup_{i=0}^{n-1} (S|Y_{a(k)})^i \left[87^{a(k)-2} d(k) + 2\right] \setminus \mathsf{V}(a(k), \mathsf{p}(k) + 2).$$

Letting

$$h = \min\left\{ i > 0 : (S|Y_{a(k)})^{in} F \cap \left[0^{a(k)-1} d(k)\right] \neq \varnothing \right\} - 1$$

and

$$h' = \min\left\{ i > 0 : (S|Y_{a(k)})^{in} F' \cap \left[0^{a(k)}\right] \neq \varnothing \right\} - 1$$

we see our transformation is rank 2.

Having done the modification to the construction, one still needs to check that the proof of mild mixing goes through. We do not pursue this here.

# 4. An ergodic joining

The purpose of this section is to construct an ergodic self-joining of $(X, \mu, T)$ that is neither the product measure nor an off-diagonal joining. The construction is an application of [CE21, Proposition 3.1]; to prepare for its application in Section 4.2 we give in Section 4.1 a technical description of towers for $(X, \mu, T)$.

We briefly describe the proof. We find a pair of sequences of integers $\alpha(i)$ and $\gamma(i)$ and measurable sets $A(i)$, $B(i)$ so that

- $T^{\alpha(i)}|_{A(i)} \approx T^{\gamma(i-1)}$ and $T^{\alpha(i)}|_{B(i)} \approx T^{\alpha(i-1)}$
- $T^{\gamma(i)}|_{A(i)} \approx T^{\alpha(i-1)}$ and $T^{\gamma(i)}|_{B(i)} \approx T^{\gamma(i-1)}$.

We use this to show

- $\nu_{\alpha(i)}$ and $\mu_{\gamma(i)}$ are Cauchy with the same limit.
- This limit is close to $c(i)\nu_{\alpha(k)} + (1 - c(i))\nu_{\gamma(k)}$ for all $k$.
- This common limit is neither one-to-one on almost every fiber nor the product measure.
- The limit is ergodic.

This strategy is from [CE21] where it was used to build joinings of interval exchange transformations. Proposition 4.3 is most of the technical argument from that paper adapted to our setting. Proposition 4.2 proves the required facts about our system to apply Proposition 4.3. We describe the choices of the $\alpha(i)$ and $\gamma(i)$, between the statement of Proposition 4.3 and the proof of Theorem 1.6. The sets $A(i)$, $B(i)$ are given in the statements of Proposition 4.2.

## 4.1. A tower

In the picture at scale $a(k)$ we have a tower

$$\mathcal{T} = \bigcup_{n=0}^{(d(k)+1)\mathsf{P}(k)-1} (S|Y_{a(k)})^n \left[0^{a(k)}\right] \subset Y_{a(k)}$$

of height $(d(k) + 1)\mathsf{P}(k)$ for the transformation $S|Y_{a(k)}$. We will wish often to think of $\mathcal{T} \cap X$ as a tower for $T$ on which $T^n$ and $S|Y_{a(j)}^n$ agree for all $-2\mathsf{P}(k) \leqslant n \leqslant 2\mathsf{P}(k)$ for most points $x$ in $\mathcal{T} \cap X$. In this subsection we produce subsets $J(k)$ of $[0^{a(k)}]$ that are the bases of towers for $T$ with such properties. We will do so by removing "columns" from the tower.

DEFINITION 4.1. — *By a* column *at scale* $i$ *we mean a subset of* $\Omega$ *defined by fixing values for all coordinates* $i \leqslant n \leqslant I$. *We call* $I \geqslant i$ *the* rank *of the column. The* base *of a column is its intersection with* $[0^{i-1}]$.

For example $\{x \in \Omega : x(4) = 2, x(5) = 3\}$ is a column at scale 4 and rank 5 while

$$\{x \in \Omega : 5 \leqslant x(a(j)) \leqslant 10\}$$

is a union of six columns at scale $a(j)$ and rank $a(j)$.

PROPOSITION 4.2. — *For every* $k \in \mathbb{N}$ *there are sets* $J(k), U(k) \subset X$ *and a constant* $c > 0$ *with the following properties.*
(PJ1) $\min\{n \in \mathbb{N} : T^n(x) \in J(k)\} \geqslant (d(k)+1)(2\mathsf{P}(k)-\mathsf{Q}(k))-3$ *for every* $x \in J(k)$,
(PJ2) $(d(k) + 1)\mathsf{P}(k) \sum_{j>k} \mu(J(j)) \to 0$.
*Writing*

$$A(k) = \left(\bigcup_{n=0}^{(d(k)+1)\mathsf{P}(k)-1} T^n J(k)\right) \setminus U(k) \quad \text{and} \quad B(k) = \left(X \setminus A(k)\right) \setminus U(k)$$

*for every* $k \in \mathbb{N}$ *we have all the following properties.*

(PJ3) $T^n(x) = (S|Y_{a(k)})^n(x)$ for every $x \in A(k)$ and every $-2\mathsf{P}(k) \leqslant n \leqslant 2\mathsf{P}(k)$.

(PJ4) For every $x \in A(k)$ we have $(S|Y_{a(k)})^{\mathsf{P}(k)}(x) \overset{a(k)-1}{=} x$.

(PJ5) $T^n(x) = (S|Y_{a(k)})^n(x)$ for every $x \in B(k)$ and every

$$-2\mathsf{P}(k) + 2\,\mathsf{Q}(k) \leqslant n \leqslant 2\mathsf{P}(k) - 2\,\mathsf{Q}(k).$$

(PJ6) For every $x \in B(k)$ we have $(S|Y_{a(k)})^{\mathsf{P}(k)-\mathsf{Q}(k)}(x) \overset{a(k)-1}{=} x$.

(PJ7) $\mu(U(k)) \leqslant \frac{15}{2^k}$,

(PJ8) $\mu(A(k)) \geqslant c$ and $\mu(B(k)) \geqslant c$.

*Proof.* — For each $k \in \mathbb{N}$ define

$$V(k) = \mathsf{V}\Big(a(k), (d(k) + 1)(2\mathsf{P}(k) - \mathsf{Q}(k)) - 1\Big)$$

and put $J(k) = [0^{a(k)}] \setminus V(k)$. Note that $J(k) \subset X$ for all $k$. For every $x \in J(k)$ and every $0 \leqslant n \leqslant (d(k) + 1)(2\mathsf{P}(k) - \mathsf{Q}(k)) - 3$ we have

$$(4.1) \qquad T^n(x) = (S|Y_{a(k)})^n(x) \quad \text{or} \quad T^n(x) = (S|Y_{a(k)})^{n+1}(x)$$

and

$$(4.2) \qquad (S|Y_{a(k)})^j(x) \notin \bigcup_{\substack{j > a(k) \\ j \notin \{a(i)\}}} W_j \implies T^n(x) = (S|Y_{a(k)})^n(x).$$

Indeed, if $T^n(x) \neq (S|Y_{a(k)})^n(x)$ for some $n$ in the above range then

$$(S|Y_{a(k)})^j(x) \in \bigcup_{i > a(k)} W_i$$

for some $0 < j \leqslant n$. But as $x \notin V(k)$ this restriction becomes

$$(S|Y_{a(k)})^j(x) \in \bigcup_{\substack{i > a(k) \\ i \text{ bounded}}} W_i$$

and

$$\min\left\{ \ell > 0 \,\middle|\, \begin{matrix} (S|Y_{a(k)})^\ell x \in W_j \quad \text{for some } x \in W_i \setminus V(k) \\ \text{with } i \neq j, \, i, j \in \{a(k)+1, a(k)+2, \dots\} \setminus \{a(r)\} \end{matrix} \right\}$$
$$\geqslant \mathsf{p}(a(k)+1) - 1$$

so for each $x \in J(k)$

$$\left| \left\{ 0 \leqslant j \leqslant n : (S|Y_{(a(k))})x \in \bigcup_{\ell > a(k)} W_\ell \right\} \right| \leqslant 1$$

and

$$\left| \left\{ 0 \geqslant j \geqslant -n : (S|Y_{(a(k))})x \in \bigcup_{\ell > a(k)} W_\ell \right\} \right| \leqslant 1$$

which establishes (PJ1) because $(d(k) + 1)(2\mathsf{P}(k) - \mathsf{Q}(k))$ is the first return time to $[0^{a(k)}]$ for $S|Y_{a(k)}$.

For (PJ4) and (PJ6) note that the definition of $A(k)$ gives that $x \in A(k)$ implies $0 \leqslant x(\ell) \leqslant d(k) + 1$ and $y \in A(k)^c \supset B(k)$ implies $0 \leqslant y(\ell) \leqslant 2d(k) + 1$, consulting Figure 2.4 gives these claims.

For (PJ2) note that $\mu(J(k)) \leqslant 1/\nu(X)t(a(k) + 1)$ giving

$$(4.3) \quad (d(k) + 1)\mathsf{P}(k) \sum_{j>k} \mu(J(j)) \leqslant \frac{1}{\nu(X)} \sum_{j>k} \frac{(d(k) + 1)\mathsf{P}(k)}{t(a(j) + 1)} \leqslant \frac{1}{\nu(X)} \sum_{j>k} \frac{1}{10^{a(j)-a(k)}}$$

because $(d(k) + 1)\mathsf{P}(k) \leqslant t(a(k))$ for all $j > k$.

To get (PJ3) and (PJ5) we define the sets $U(k)$ in terms of columns. First define

$$(4.4) \quad U_0(k)$$

$$= \Big\{ x \in \Omega : x(a(k) \notin \{0, 1, 5, 6, 7, 8, 9, d(k) - 1, \dots, d(k) + 2, 2d(k), 2d(k) + 1\} \Big\}$$

which is a union of columns at scale $a(k)$. We have

$$\nu(U_0(k)) = \frac{13\mathsf{P}(k)}{t(a(k))} \leqslant \frac{13}{d(k)}$$

and therefore

$$(4.5) \qquad\qquad \mu(U_0(k) \cap X) \leqslant \frac{13}{d(k)\nu(X)} \leqslant \frac{15}{2^k}$$

which goes to zero as $k \to \infty$.

Now, if $x \in X \setminus \big( U_0(k) \cup V(k) \big)$ then for all $0 \leqslant |n| \leqslant \mathsf{P}(k)$ because $x \notin V(k)$ and $(T^n x)(a(k)) \neq 7$, as in (4.2), $(S|Y_{a(k)})^n x = T^n x$.

Setting $U(k) = U_0(k) \cup V(k)$ we have (PJ3) and (PJ5). Using (4.5) we have

$$\mu(U(k)) \leqslant \frac{14}{2^k} + \frac{1}{\nu(X)}(2((d(k) + 1)(2P(k) - Q(k)) - 1) + 1)\nu\left( \bigcup_{j>k} W_{a(j)} \right) \leqslant \frac{15}{2^k}$$

which is (PJ7).

Lastly, (PJ8) is a consequence of $\{x \in Y_{a(k)} : x(k) \leqslant d(k)\}$ and $\{x \in Y_{a(k)} : x(k) > d(k)\}$ both having measure at least $\frac{1}{4}$ and (PJ7). $\qquad\square$

## 4.2. Building the joining

In this section we prove Theorem 1.6. We begin by recalling that the Kantorovich–Rubinstein metric on the space $\mathcal{M}_1(V)$ of Borel probability measures on a compact metric space $V$ is defined by

$$\mathsf{d_{KR}}(\eta, \xi) = \sup\{|\eta(f) - \xi(f)| : f \quad \text{is 1-Lipschitz on } X \text{ with respect to } \mathsf{d}|_{X \times X}\}$$

for all $\eta, \xi$ in $\mathcal{M}_1(V)$.

The joining $\lambda$ will be built using a minor modification of [CE21, Proposition 3.1] (in the case $d = 2$). We highlight the differences in bold and in the appendix we prove the version of [CE21, Corollary 3.3], which is the only part of the proof that requires significant modification. Note that (J7) is a strengthening of [CE21, Proposition 3.1 (B)], the corresponding condition in [CE21], and this is used for (A.5) in the appendix.

PROPOSITION 4.3. — *Fix sequences*

- $k \mapsto J(k)$ *of subsets of* $X$,
- $k \mapsto U(k)$ *of measurable subsets of* $X$,
- $k \mapsto r(k)$ *of natural numbers,*
- $k \mapsto \alpha(k)$ *and* $k \mapsto \gamma(k)$ *of integers,*
- $k \mapsto \epsilon(k)$ *of positive real numbers,*

*and set*

$$(4.6) \qquad A(k) = \bigcup_{i=0}^{r(k)-1} T^i(J(k)) \backslash U(k) \qquad B(k) = (X \setminus A(k)) \setminus U(k)$$

*and let* $\nu_i$ *on* $X \times X$ *be the push-forward of* $\mu$ *under the map* $x \mapsto (x, T^i(x))$. *Assume all of the following.*

(J1) *There is* $c > 0$ *such that for all* $k \in \mathbb{N}$ *one has* $\mu(A(k)) \geqslant c$ *and* $\mu(B(k)) \geqslant c$.
(J2) *For all* $k \in \mathbb{N}$ *the minimal return time of* $T$ *on* $J(k)$ *is at least* $\frac{3}{2}r(k)$.
(J3) *For all* $k \in \mathbb{N}$ *we have* $\mu(U(k)) < \epsilon(k)$.
(J4) $\lim_{k \to \infty} r(k) \sum j > k \mu(J(j)) = 0$.
(J5) *The sequence* $k \mapsto \epsilon(k)$ *is non-increasing and summable.*
(J6) *For every* $k \in \mathbb{N}$ *and every* $x \in A(k)$ *one has*

$$(4.7) \qquad \mathsf{d}\big(T^{\alpha(k)}x, T^{\gamma(k-1)}x\big) < \epsilon(k) \quad \text{and} \quad \mathsf{d}\big(T^{\gamma(k)}x, T^{\alpha(k-1)}x\big) < \epsilon(k)$$

*and for every* $k \in \mathbb{N}$ *and every* $x \in B(k)$ *one has*

$$(4.8) \qquad \mathsf{d}\big(T^{\alpha(k)}x, T^{\alpha(k-1)}x\big) < \epsilon(k) \quad \text{and} \quad \mathsf{d}\big(T^{\gamma(k)}x, T^{\gamma(k-1)}x\big) < \epsilon(k).$$

(J7) *For every* $k \in \mathbb{N}$ *the estimates*

$$(4.9) \quad \mu\left(\bigcup_{9L \geqslant r(k+1)} \left\{x \in X : \mathsf{d}_{\mathsf{KR}}\left(\nu_{\alpha(k)}, \frac{1}{L}\sum_{i=1}^{L}(T \times T)^i \delta\big(x, T^{\alpha(k)}x\big)\right) > \epsilon(k)\right\}\right)$$
$$< \epsilon(k)$$

$$(4.10) \quad \mu\left(\bigcup_{9L \geqslant r(k+1)} \left\{x \in X : \mathsf{d}_{\mathsf{KR}}\left(\nu_{\gamma(k)}, \frac{1}{L}\sum_{i=1}^{L}(T \times T)^i \delta\big(x, T^{\gamma(k)}x\big)\right) > \epsilon(k)\right\}\right)$$
$$< \epsilon(k)$$

*both hold.*

*Then*

$$\lim_{j \to \infty} \nu_{\alpha(j)} = \lim_{j \to \infty} \frac{\nu_{\alpha(j)} + \nu_{\gamma(j)}}{2} = \lim_{j \to \infty} \nu_{\gamma(j)}$$

*and in particular the above limits exist. Moreover the common limiting value* $\lambda$ *is ergodic for* $T \times T$ *and there is* $C > 0$ *such that*

$$\mathsf{d}_{\mathsf{KR}}\left(\lambda, \frac{\nu_{\alpha(k)} + \nu_{\gamma(k)}}{2}\right) \leqslant C \sum_{i > k} \epsilon(i)$$

*for all* $k \in \mathbb{N}$.

In order to apply Proposition 4.3 to establish Theorem 1.6 the central assumption to establish is (J6). The times $\alpha(k)$ and $\gamma(k)$ will be chosen based on the picture at the unbounded scale $a(k)$. The set $A(k)$ will consist of points in the left tower at scale $a(k)$ while the set $B(k)$ will consist of points in the right tower at scale $a(k)$. Indeed, if $x \in A(x)$, then by (4.4) we have $x(a(k)) \in [2, d(k) - 2]$ and so $T^n(x)$ is in the left tower at scale $a(k)$ for all $0 \leqslant |n| \leqslant \mathsf{P}(k)$. Analogously for $x \in B(k)$ and the right tower. Recall that $\mathsf{P}(k)$ is the height of the left tower and $\mathsf{P}(k) - \mathsf{Q}(k)$ is the height of the right tower at scale $a(k)$. Taking

$$\alpha(0) = 1, \qquad\qquad \alpha(1) = \mathsf{P}(1),$$
$$\alpha(2) = \mathsf{P}(2) - \mathsf{P}(1) + 1, \qquad\qquad \alpha(3) = \mathsf{P}(3) - \mathsf{P}(2) + \mathsf{P}(1),$$

and

$$\gamma(0) = 0, \qquad\qquad \gamma(1) = -\mathsf{P}(1) + 1,$$
$$\gamma(2) = -\mathsf{P}(2) + \mathsf{P}(1), \qquad\qquad \gamma(3) = -\mathsf{P}(3) + \mathsf{P}(2) - \mathsf{P}(1) + 1$$

and using the inductive definition $\mathsf{Q}(n + 1) = 2\mathsf{P}(n) - \mathsf{Q}(n)$ we have for example that $T^{\alpha(3)} \approx T^{\gamma(2)}$ and $T^{\gamma(3)} \approx T^{\alpha(2)}$ for many points in $A(3)$ and $T^{\alpha(3)} \approx T^{\alpha(2)}$ and $T^{\gamma(3)} \approx T^{\gamma(2)}$ for many points in $B(3)$.

*Proof of Theorem 1.6.* — We will apply Proposition 4.3 to produce an ergodic joining that is neither the product measure nor an off diagonal joining. Let the constant $c > 0$ and the sequences $k \mapsto J(k)$ and $k \mapsto U(k)$ be those from Proposition 4.2. Define $r(k) = (d(k) + 1)\mathsf{P}(k) - 1$ and put $\epsilon(k) = 15 \cdot 2^{-k}$ With these choices the sets $A(k)$ and $B(k)$ in the hypothesis of Proposition 4.3 are the same as the sets $A(k)$ and $B(k)$ in the statement of Proposition 4.2. In particular we get (J1) from (PJ8) and (J2) from

$$3(d(k) + 1)\mathsf{P}(k) \leqslant 4(d(k) + 1)(\mathsf{P}(k) - \mathsf{Q}(k))$$

and (PJ1). Our choice of $\epsilon(k)$ gives (J3) immediately and (J4) is just (PJ2). We get (J5) from (PJ7).

To establish (J6) and (J7) we need to define the sequences $\alpha(k)$ and $\gamma(k)$. Put $\alpha(0) = 1$ and $\gamma(0) = 0$ and inductively define

$$\alpha(k + 1) = \mathsf{P}(k + 1) + \gamma(k) \quad \text{and} \quad \gamma(k + 1) = -\mathsf{P}(k + 1) + \alpha(k)$$

for all $k \in \mathbb{N}$. One can check by induction that

$$(4.11) \quad \alpha(k) + \mathsf{Q}(k) = \alpha(k - 1) + \mathsf{P}(k) \quad \text{and} \quad \gamma(k) + \mathsf{P}(k) = \mathsf{Q}(k) + \gamma(k - 1)$$

for all $k \in \mathbb{N}$.

We verify that (J6) holds. We calculate

$$T^{\alpha(k)}(x) = T^{\mathsf{P}(k) + \gamma(k-1)}(x) \stackrel{a(k)-1}{=} T^{\gamma(k-1)}(x) \implies \mathsf{d}\Big(T^{\alpha(k)}(x), T^{\gamma(k-1)}(x)\Big) \leqslant \epsilon(k)$$

and

$$T^{\gamma(k)}(x) = T^{-\mathsf{P}(k) + \alpha(k-1)}(x) \stackrel{a(k)-1}{=} T^{\alpha(k-1)}(x) \implies \mathsf{d}\Big(T^{\gamma(k)}(x), T^{\alpha(k-1)}(x)\Big) \leqslant \epsilon(k)$$

using (PJ3) and (PJ4), which gives (4.7). Similarly, for all $x \in B(k)$ we have

$$T^{\alpha(k)}(x) \stackrel{a(k)-1}{=} T^{\alpha(k) - \mathsf{P}(k) + \mathsf{Q}(k)}(x) = T^{\alpha(k-1)}(x) \implies \mathsf{d}\Big(T^{\alpha(k)}(x), T^{\alpha(k-1)}(x)\Big) < \epsilon(k)$$

and

$$T^{\gamma(k)}(x) \overset{a(k)-1}{=} T^{\gamma(k)+\mathsf{P}(k)-\mathsf{Q}(k)}(x) = T^{\gamma(k-1)}(x) \Longrightarrow \mathsf{d}\Big(T^{\gamma(k)}(x), T^{\gamma(k-1)}(x)\Big) < \epsilon(k)$$

using (PJ5) and (PJ6). This establishes (4.8) and thus (J6).

Finally, we establish (4.9) to give half of (J7), the other half, (4.10), is analogous and omitted. Observe that $\mathsf{p}(n+1)$ is the number of cylinders of length $n$ that non-trivially intersect $X$. Also, observe that if $x \notin \mathsf{V}(n, \mathsf{p}(n+1))$ then either

- $\{T^j x\}_{0 \leqslant j \leqslant \mathsf{p}(n+1)}$ intersects every cylinder of length $n$ in $Y_n$ exactly once
- or it does not intersect $[7^n]$ and it intersects one such cylinder twice.

Combining these two facts, we have that if $x, y \notin \mathsf{V}(\mathsf{p}(n+1)+a(k), n)$ then

$$(4.12) \quad \left| \frac{1}{\mathsf{p}(n+1)} \sum_{i=0}^{\mathsf{p}(n+1)-1} f\Big(T^i x, T^{a(k)} T^i x\Big) - f\Big(T^i y, T^{a(k)} T^i y\Big) \right|$$

$$\leqslant 2^{-n} \mathsf{Lip}(f) + \frac{a(k)+2}{\mathsf{p}(n+1)} \cdot \mathrm{diam}(X) \mathsf{Lip}(f).$$

Indeed, if $T^i x, T^j y$ are in the same cylinder of length $n$ and $T^{a(k)+i} x, T^{a(k)+j} y$ are in the same cylinder of length $n$ then

$$\left| f\Big(T^i x, T^{a(k)+i} x\Big) - f\Big(T^j y, T^{a(k)+j} y\Big) \right| \leqslant 2^{-n-1} \mathsf{Lip}(f).$$

This allows us to treat at least $\mathsf{p}(n+1) - (a(k)+2)$ terms in each summand and we estimate trivially on the rest.

Thus, if $x \notin V(n, \mathsf{p}(n+1)+a(k))$, $f$ is 1-Lipschitz and $\|f\|_{\sup} = 1$ we have

$$(4.13) \quad \left| \int_X f d\mu - \frac{1}{\mathsf{p}(n+1)} \sum_{i=0}^{\mathsf{p}(n+1)} f(T^i x) \right|$$

$$\leqslant 2\mu\Big( \mathsf{V}\Big(n, \mathsf{p}(n+1)+a(k)\Big)\Big) + 2^{-n} \frac{a(k)}{\mathsf{p}(n+1)}.$$

We now let $n = \frac{3}{2} \cdot a(k)$ and we have

$$(4.14) \quad \mu\Big(\mathsf{V}(n, \mathsf{p}(n+1)+a(k))\Big) \leqslant 2 \cdot \Big(\mathsf{p}(n+1)+a(k)\Big)\nu(X)^{-1}\nu\left(\bigcup_{a(j)>n} W_{a(j)}\right)$$

$$\leqslant 8 \cdot \frac{10^{1.6a(k)}}{10^{a(k+1)-a(k)-1}}.$$

This uses (2.5), the fact that $|\mathsf{p}(n+1)| \leqslant 10^n \cdot 2^{k+6}$.

To complete the proof we need straightforward estimate: Assume

$$\mu\left(\left\{x : \left|\frac{1}{r}\sum_{i=0}^{r-1} f\Big(T^i x, T^{i+m} x\Big)\right| > \epsilon\right\}\right) < \delta$$

then for $Lr \leqslant N < (L+1)N$

$$(4.15) \quad \mu\left(\left\{x : \left|\frac{1}{N}\sum_{i=0}^{N-1} f\Big(T^i x, T^{i+m} x\Big)\right| > \epsilon + \sqrt{\delta}\|f\|_{\sup} + \frac{2}{L}\right\}\right) < \sqrt{\delta}.$$

Indeed, we say $j$ is bad for $x$ if

$$\left| \frac{1}{r} \sum_{i=jr}^{(j+1)r-1} f\left(T^i x, T^{i+m} x\right) \right| > \epsilon.$$

Now

$$\mu\left(\{x : |\{0 \leqslant j \leqslant L-1 : j \quad \text{is bad for } x\}| \geqslant \left\lceil \sqrt{\delta} L \right\rceil\}\right) < \sqrt{\delta}.$$

On this set the estimate in (4.15) holds.

Observing that $\frac{1}{9} r(k+1) \geqslant 9^{n-a(k+1)-2} \mathsf{p}(n)$, plugging in (4.13) and (4.14) into (4.15) completes the proof that the assumptions of Proposition 4.3 hold.

We now have to prove that our limiting joining, $\sigma$, is neither one-to-one nor the product measure. To do this we require some additional notation. Let $\tau$ be a coupling of $\mu$ on $X \times X$. Applying disintegration of measures to projection onto the first coordinate, one defines a $\mu$-a.e. uniquely defined family of measure $\tau_x$ so that

- $\tau_x$ is supported on $\{x\} \times X$,
- $\int \tau_x d\mu(x) = \tau$.

Observe that

$$(4.16) \quad \left\{x : d_{KR}\left(\sigma_x, \frac{1}{2}\left[\delta\left(x, T^{\gamma(k)} x\right) + \delta\left(x, T^{\alpha(x)} x\right)\right]\right) \leqslant 2^{-a(k)}\right\}$$

$$\subset \bigcup_{j>k} A(k) \cup B(k) = \left(\bigcup_{j>k} U(k)\right)^c.$$

The containment uses (J6) and the equality uses (4.6). Following an analysis as in the proof of (J7), applying this with $k = 1$ shows that $\sigma$ is not the product measure. To prove that the joining is not one-to-one on a.e. fiber we use the following:

LEMMA 4.4 ([CE21, Lemma 4.7]). — *If $\sigma$ is a measure that is one-to-one on almost every fiber then $\sigma$ cannot be the weak-\* limit of a sequence of measures $\tau_i$ that are two-to-one on almost every fiber and such that there exists a $\delta$ satisfying*

$$\mu(\{x : \operatorname{diam}(\operatorname{supp}(\tau_i|_x) > \delta\}) > 3/4$$

*for infinitely many $i$ where $\tau_i|_x$ is the a.e. defined measure given by applying disintegration of measures to the projection of $\tau$ onto the first coordinate.*

Using (J3), (J5) and (4.16) provides the assumption of the above lemma. $\qquad \square$

*Remark 4.5.* — One can use (4.16) a little more and show that $\sigma$ is two-to-one on a.e. fiber. Indeed, this equation gives the atoms in $\sigma|_x$. This fact is not needed for any statements of this paper and we omit the details.

# 5. Mild mixing

In this section we will prove Theorem 1.7. Given a non-constant function $f$ in $\mathsf{L}^2(X, \mu)$ we wish to prove that no sequence $\boldsymbol{r}$ in $\mathbb{Z}$ with $|r(n)| \to \infty$ satisfies $\|f - T^{r(n)} f\| \to 0$.

In Section 5.1 we define the essential notion of "buddies" and give in Theorem 5.5 a criterion for a non-constant $f \in \mathsf{L}^2(X, \mu)$ not to have $\boldsymbol{r}$ as a rigidity sequence. The thrust of the criterion is that the existence of a uniformly positive proportion of $(r(\tau), \tau)$ buddies as $\tau \to \infty$ precludes $f$ from being rigid along $\boldsymbol{r}$.

Thus, given $\boldsymbol{r}$ we need to be able to produce buddies. There are certain situations where producing buddies is relatively straightforward. Those are the situations where we do not need to carry out the reduction procedure mentioned in the introduction. In those situations are able to instead produce "friendly pairs" which are defined in Section 5.2.

In fact, we would prefer to work with friendly pairs throughout and avoid altogether the notion of buddies. However, when using the reduction procedure (see Section 2.4) to prove Proposition 5.20 we are unable to work directly with friendly pairs.

We give in Section 5.3 a technical summary of the reduction procedure. In Section 5.4 we given some preparatory results describing what happens when we apply the reduction procedure once. The main technical result is proved in Section 5.5. Finally, the proof of Theorem 1.7 is given in Section 5.6.

## 5.1. Buddies

Essential to our proof is the following definition.

DEFINITION 5.1 (Buddies). — *Fix $r, j \in \mathbb{N}$. We will say that $(y, z)$ in $X \times X$ are $(r, j)$ buddies if the following properties all hold.*

(1) $y \overset{j-1}{=} z$.

(2) *There exists $0 < |i| < 9$ so that $(S|Y_{j-1}^i \circ T^r)(y) \overset{j-1}{=} (T^r)(z)$*

*Remark 5.2.* — In our argument, we will have $|i| = 1$. However, in related arguments, $|i|$ will often be in a larger range (see e.g. [CK21]). As such we phrase things more generally to have Theorem 5.5 in a more general situation.

The points $y$ and $z$ are $(r, j)$ buddies if they are initially close (in that they agree in the first $j - 1$ coordinates) and through $r$ iterations of $T$ differ by at least one and at most eight iterations of $S|Y_{j-1}$.

*Example 5.3.* — Fix $k \in \mathbb{N}$ a bounded scale. When $y \in X \cap [67^{k-2}7]$ with $S(y) = T(y)$ and $z \in X \cap [67^{k-2}8]$ with $S^2(y) = T(y)$ the points $y$ and $z$ are $(1, k)$ buddies.

We record here a lemma about buddies that will be needed later.

LEMMA 5.4. — *Assume $r, u \in \mathbb{N}$, $j < J \in \mathbb{N}$ and $y, z \in X$ so that*

(1) $T^r(y) \overset{J-1}{=} T^u(y)$,

(2) $T^r(z) \overset{J-1}{=} T^u(z)$,

(3) $(y, z)$ are $(r, j)$-buddies,

*then $(y, z)$ are $(u, j)$-buddies.*

*Proof.* — The condition (1) of being $(a, j)$ buddies is independent of $a$ and so we just need to check condition (2). Condition (2) of being $(a, j)$-buddies depends only on the first $j - 1$ coordinates of $T^a$. Indeed, $S|Y_{j-1}$ depends only on these coordinates. So by conditions (1)–(2) above, $(y, z)$ being $(r, j)$-buddies and being $(u, j)$-buddies is equivalent. □

The following theorem is our main criterion for non-rigidity of $f \in \mathsf{L}^2(X, \mu)$ along a sequence **r** in $\mathbb{Z}$.

THEOREM 5.5. — *Let $f : X \to \mathbb{C}$ be a non-constant $\mathsf{L}^2(X, \mu)$ function and let* **r** *be a sequence in $\mathbb{Z}$ with $|r(n)| \to \infty$. Assume there exists $\rho > 0$ so that for all $\tau \in \mathbb{N}$ there exists $i_0 \in \mathbb{N}$ so that for all $i \geqslant i_0$ there exists a Borel set $\Gamma_i \subset X$ and a measure preserving function $\phi_i : X \to X$ with*

- *$\mu(\Gamma_i) > \rho$,*
- *for all $z \in \Gamma_i$ we have that $(\phi_i(z), z)$ are $(r(i), \tau)$-buddies*

*all hold. Then* **r** *is not a rigidity sequence for $f$.*

LEMMA 5.6. — *For all $\epsilon > 0$ and $B \in \mathbb{N}$ there exists $\tau$ so that if $r \in \mathbb{Z}$, $\Gamma$ is a Borel set and $\phi : \Gamma \to X$ is a measure preserving function satisfying $(\phi(z), z)$ are $(r, \tau)$-buddies for then*

$$\mu\left(\left\{ y \in \Gamma \,\middle|\, \begin{array}{l} a \,\exists\, \ell \text{ with } 0 < |\ell| < 10 \text{ so that for all } -B \leqslant a \leqslant B \\ \qquad \text{we have } \mathsf{d}\left(\left(T^{r+a} \circ \phi_i\right)(y),\; T^{r+a+\ell}(y)\right) > \epsilon \end{array} \right\}\right) < \epsilon.$$

*Proof.* — The definition of $(y, z)$-buddies gives that for any $\delta > 0$ there exists $j_0$ so that for all $j' \geqslant j_0$ we have that for every $(y, z)$ which are $(r, j')$-buddies we have that there exists $0 < |k| < 10$ so that

$$(5.1) \qquad \min_{0 < |k| < 10} \mathsf{d}\left((T^r \circ \phi_i)(y),\quad \left((S|Y_{j'-1})^k \circ T^r\right)(y)\right) < \delta.$$

Because $T^k : X \to X$ is measurable for all $k \in \mathbb{Z}$, by Lusin's theorem and the fact that continuous functions on compact sets are uniformly continuous, for any $B \in \mathbb{N}$ there exists $\delta > 0$ and $K \subset X$, compact with $\mu(K) > 1 - \epsilon$, so that for all $x, y \in K$ with $d(x, y) < \delta$ and $-B \leqslant a \leqslant B$ we have

$$\mathsf{d}(T^a x, T^a y) < \epsilon.$$

Combining this with (5.1) gives the lemma. □

*Proof of Theorem 5.5.* — Let $\rho > 0$ be given and $f$ be a non-constant $\mathsf{L}^2(\mu)$ function. Because $T$ is weakly mixing and thus every power of $T$ is ergodic, for all $0 < |\ell| < 9$ there exists $c_\ell > 0$ so that

$$\int \left| f - f \circ T^\ell \right| \mathsf{d}\mu > c_\ell$$

holds. Let $\widetilde{c} = \min\{1, c_\ell : 0 < |\ell| < 9\}$ and $0 < \varepsilon < \frac{\widetilde{c}}{9} \frac{\rho}{400}$. By the ergodic theorem (either Birkhoff or Von Neumann), for each $0 < |\ell| < 9$ there exists $B_\ell \in \mathbb{N}$ so that for all $B \geqslant B_\ell$ we have

$$(5.2) \qquad \mu\left(\left\{ y : \left| \frac{1}{B+1} \sum_{k=0}^{B} \left| f(T^k y) - f(T^{\ell+k} y) \right| - \int \left| f - f \circ T^\ell \right| \mathsf{d}\mu \right| > \varepsilon \right\}\right) < \varepsilon.$$

Denote these sets $A_{\ell,B}$. Let $\widetilde{B} = \max\{B_\ell : 0 < |\ell| < 9\}$. By Lusin's theorem, there exists $K := K_{\varepsilon,\tilde{B}} \subset X$ compact so that $\mu(K) > 1 - \varepsilon$ and $f \circ T^i$ restricted to $K$ is continuous for all $-9 < i < \widetilde{B} + 9$. Because continuous functions on compact sets are uniformly continuous, there exists $\tau_\varepsilon$ so that when $x, y \in K_{\varepsilon,B}$ and $x \stackrel{\tau_\varepsilon}{=} y$ then

$$\tag{5.3} \left| f(T^j x) - f(T^j y) \right| < \varepsilon.$$

We now assume that the assumption of Theorem 5.5 holds and obtain $\tau$ as in Lemma 5.6 for $\epsilon = \min\{2^{-\tau_\varepsilon}, \varepsilon\}$ and let $\tau' = \max\{\tau_\varepsilon, \tau\}$. We choose $i_0$ as in the statement of Theorem 5.5 with the $\rho$ we have been given and $\tau = \tau'$.

By Lemma 5.6 there exists $0 < |\ell| < 9$ (depending on $i$) and a set $G_i$ so that $\mu(G_i) \geqslant \frac{1}{18}\mu(\Gamma_i) - \varepsilon > \frac{1}{20}\rho$ and

$$\tag{5.4} \left(T^{n_i+a} \circ \phi_i\right)(y) \stackrel{\tau}{=} \left(T^{n_i+a+\ell}\right)(z) \quad \text{for all } -\widetilde{B} \leqslant a \leqslant \widetilde{B} \quad \text{and all } z \in G_i$$

both hold.

Now let $D$ denote the set of $z$ so that:

- $z \in K$,
- $\phi_i(z) \in K$,
- $T^{n_i}(z) \in K$,
- $(T^{n_i} \circ \phi_i)(z) \in K$,
- $z \in G_i$,
- $T^{n_i}(z) \in A_{\ell,\tilde{B}}$.

Observe that $\mu(D) = \mu(G_i) - \mu(A_{\ell,\tilde{B}}^c) - 4\mu(K^c) > \frac{1}{40}\rho$. For all $z \in D$ we have by (5.2), (5.3) and (5.4) that

$$\left| \frac{1}{\widetilde{B}+1}\sum_{j=0}^{\tilde{B}} \left| f\left(T^{n_i+j}(z)\right) - f\left(\left(T^{n_i+j} \circ \phi_i\right)(z)\right) \right| - \int \left| f - f \circ T^\ell \right| \, \mathsf{d}\mu \right| < 2\varepsilon$$

and so

$$\frac{1}{\widetilde{B}+1}\sum_{j=0}^{\tilde{B}} \left| f\left(T^{n_i+j}(z)\right) - f\left((T^{n_i+j} \circ \phi_i)(z)\right) \right| > \widetilde{c} - 2\varepsilon$$

as well. Because the triangle inequality gives

$$\left| f(T^{n_i+j}(z)) - f\left((T^{n_i+j} \circ \phi_i)(z)\right) \right| - \left| f(T^j(z)) - f\left((T^j \circ \phi_i)(z)\right) \right|$$
$$\leqslant \left| f(T^{n_i+j}(z)) - f(T^j(z)) \right| + \left| f\left((T^{n_i+j} \circ \phi_i)(z)\right) - f\left((T^j \circ \phi_i)(z)\right) \right|$$

we can estimate

$$\frac{1}{\widetilde{B}+1}\sum_{j=0}^{\tilde{B}} \int_D \left| f(T^{n_i+j}(z)) - f(T^j(z)) \right| + \left| f\left((T^{n_i+j} \circ \phi_i)(z)\right) - f\left((T^j \circ \phi_i)(z)\right) \right| \, \mathsf{d}\mu$$

$$\geqslant \mu(D)\widetilde{c} - 2\varepsilon - \frac{1}{\widetilde{B}+1}\sum_{j=0}^{\tilde{B}} \int_D \left| f(T^j(z)) - f\left((T^j \circ \phi_i)(z)\right) \right| \, \mathsf{d}\mu \geqslant \frac{1}{80}\rho\widetilde{c}$$

where the last inequality uses (5.3). Thus

$$(5.5) \qquad \int_X |f(T^{n_i}x) - f(x)| \geqslant \frac{1}{160}\rho\widetilde{c}$$

proving the Theorem 5.5. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In fact the proof gives the following.

COROLLARY 5.7. — *For any non-constant $f \in \mathsf{L}^2(X,\mu)$ and any $\rho > 0$ there exists $\tau' := \tau'(\rho, f) \in \mathbb{N}$ and $\delta := \delta(\rho, f) > 0$ so that if*

- *$m \in \mathbb{Z}$,*
- *$\Gamma$ is a Borel set,*
- *$\phi : \Gamma \to X$ is $\mu$-measure preserving*

*satisfy*

- *$\mu(\Gamma) \geqslant \rho$ and*
- *for all $z \in \Gamma$ we have that $(z, \phi(z))$ are $(m, \tau)$-buddies*

*then*

$$\int_X |f - f \circ T^m| \, \mathsf{d}\mu > \delta$$

*holds.*

## 5.2. Buddies from friendly pairs

We define friendly pairs and prove that friendly pairs are always buddies. We then show how to produce friendly pairs at bounded scales, establishing that if $f \in \mathsf{L}^2$, non-constant and $\boldsymbol{r}$ is a sequence such that $\mathsf{J}(r(n))$ is at a bounded scale for infinitely many $n$ then $\boldsymbol{r}$ is not an $f$-rigidity sequence (see Corollary 5.11).

DEFINITION 5.8 (Friendly pairs). — *Fix $r \in \mathbb{Z}$ and $j \in \mathbb{N}$ with $j$ at a bounded scale. We will say that $(y, z)$ in $X \times X$ is an $(r, j)$ friendly pair if the following properties all hold.*

(F1) $y(n) \overset{j-1}{=} z(n)$ for all $1 \leqslant n \leqslant j - 1$,
(F2) $T^n(y) = (S|Y_j)^n(y)$ for all $1 \leqslant |n| \leqslant |r| + 1$ with $sgn(n) = sgn(r)$,
(F3) $T^n(z) = (S|Y_j)^n(z)$ for all $1 \leqslant |n| \leqslant |r| + 1$ with $sgn(h) = sgn(r)$,
(F4) $(S|Y_{j-1})^n(y) \notin W_j$ for all $1 \leqslant |n| \leqslant |r| + 1$ with $sgn(h) = sgn(r)$,
(F5) $(S|Y_{j-1})^h(z) \in W_j$ for a unique $1 \leqslant |h| \leqslant |r|$ with $sgn(h) = sgn(r)$.

The following lemma says that friendly pairs are always buddies.

LEMMA 5.9. — *If $(y, z)$ is an $(r, j)$ friendly pair then $(y, z)$ are $(r, j)$-buddies.*

*Proof.* — By conditions (F3) and (F5) we have that $T^r(z) = (S|Y_{j-1})^{r+1}(z)$ and by conditions (F2) and (F4) we have that $T^r(y) = (S|Y_{j-1})^r(y)$ and $T^{r+1}(y) = (S|Y_{j-1})^{r+1}(y)$. By condition (F1) we have that the first $j - 1$ entries of $(S|Y_{j-1})^k(y)$ are the same as the first $j - 1$ entries of $(S|Y_{j-1})^k(z)$ for all $k$ and in particular for $k \in \{r, r + 1\}$ proving the second condition in the definition of buddies. This proves Lemma 5.9. $\qquad\square$
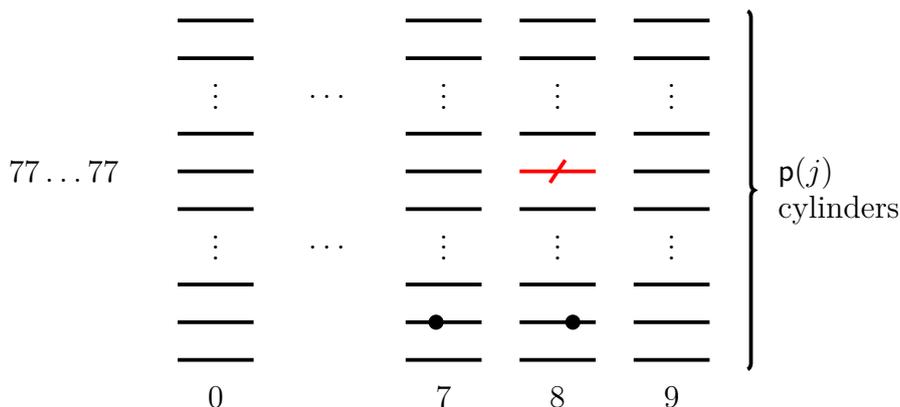
Figure 5.1. The picture of $Y_{j-1}$ at the bounded scale $j$. The marked interval is $W_j$. The marked points form a $(\mathsf{p}(j), j)$ friendly pair if both points are outside $\mathsf{V}(j, \mathsf{p}(j))$.

There are some situations where it is relatively straightforward to produce friendly pairs. The first, formalized by the next result, is when $r \in \mathbb{Z}$ is at a bounded scale. The second, given by Proposition 5.12, is when $r$ is small for its scale and $\mathsf{J}(r) - 1$ is a bounded scale. The second applies in particular when $r$ is at an unbounded scale, but small at that scale. The salient point is that neither of these situations require us to reduce to smaller scales.

THEOREM 5.10. — *For all $r \in \mathbb{Z}$ with $\mathsf{J}(r)$ bounded there is a set $B \subset X$ with the following properties.*

- $\mu(B) \geqslant 10^{-5}$.
- *There is $\Phi : B \setminus \mathsf{V}(\mathsf{J}(r), 8r) \to X$ a measure-preserving map that is a bijection onto its image such that $(\Phi(z), z)$ is are $(r, \mathsf{J}(r))$-friends for all $z \in B \setminus \mathsf{V}(\mathsf{J}(r), 8r)$.*

*Proof.* — Fix $r \in \mathbb{Z}$ with $\mathsf{J}(r)$ bounded. We prove the case $r < 0$. The case $r > 0$ is similar and therefore omitted. Note that

$$(5.6) \qquad \qquad \mathsf{p}(\mathsf{J}(r) + 1)/2 \geqslant |r| \geqslant \mathsf{p}(\mathsf{J}(r))/2$$

holds from the definition of the expansion. Define

$$(5.7) \quad B = \Bigg\{ x \in X : x(\mathsf{J}(r)) = 8,$$
$$x(\mathsf{J}(r) + 1) \in \left[ \frac{(c(\mathsf{J}(r) + 1) + 1)}{5}, \ \frac{(c(\mathsf{J}(r) + 1) + 1)}{2} \right]$$
$$\text{and } \left( S|Y_{\mathsf{J}(r)-1} \right)^i (x) \in W_{\mathsf{J}(r)} \text{ for some } -|r| < i \leqslant -1 \Bigg\}$$

and note that $\nu(B) \geqslant 10^{-3}$ by independence so $\mu(B) \geqslant 10^{-3}$ by definition of $\mu$.

Given $y \in B \setminus \mathsf{V}(\mathsf{J}(r), 8r)$ define $\Phi(y)$ by $(\Phi(y))(\mathsf{J}(r)) = 6$ and $(\Phi(y))(n) = y(n)$ otherwise. Fix $z \in B \setminus \mathsf{V}(\mathsf{J}(r), 8r)$. The pair $(\Phi(z), z)$ satisfies (F1). Observe that

$(S|Y_{\mathsf{J}(r)-1})^h(z) \in W_{\mathsf{J}(r)}$ for exactly one $r < h \leqslant -1$ because of (5.6) giving (F5). We also have (F3) because

$$(S|Y_{\mathsf{J}(r)})^h(z) \notin \bigcup_{i > \mathsf{J}(r)} W_i$$

for all $-r \leqslant h \leqslant -1$ follows upon combining $z \notin \mathsf{V}(\mathsf{J}(r), 8r)$ with the condition on $z(\mathsf{J}(r) + 1)$. As $(\Phi(z))(\mathsf{J}(r)) = 6$ we have $(S|Y_{\mathsf{J}(r)-1})^h(\Phi(z)) \notin W_{\mathsf{J}(r)}$ for all $-r-1 \leqslant h \leqslant -1$. That gives (F4). Next note that $\Phi(z) = T^n(z)$ for some $4r < n < 0$ which implies $\Phi(z) \in X$. As $z \notin \mathsf{V}(\mathsf{J}(r), 8r)$, we have

$$(S|Y_{\mathsf{J}(r)-1})^h(\Phi(z)) \notin \bigcup_{a(k) > \mathsf{J}(r)} W_{a(k)}$$

which, together with the fact that $(\Phi(z))(\mathsf{J}(r)) = 6$ implies

$$(S|Y_{\mathsf{J}(r)-1})^h(\Phi(z)) \notin \bigcup_{j > \mathsf{J}(r)} W_j$$

for all $r \leqslant h \leqslant 0$, gives (F2). Thus $(\Phi(z), z)$ is an $(r, \mathsf{J}(r))$ friendly pair. $\qquad\square$

COROLLARY 5.11. — *If* $\mathbf{r}$ *is a sequence in* $\mathbb{Z}$ *with* $|r(i)| \to \infty$ *and* $r(i)$ *at a bounded scale infinitely often then* $\mathbf{r}$ *is not a rigidity sequence for any non-constant* $\mathsf{L}^2(X, \mu)$ *function.*

*Proof.* — Pass to a subsequence with the property that $r(i)$ is at a bounded scale for all $i \in \mathbb{N}$. Let $B_i$ and $\Phi_i$ result from applying Theorem 5.10 to each $r(i)$. By Lemma 5.9 and Theorem 5.5 it suffices to show

$$\mu\Big(B_i \cap \mathsf{V}\big(\mathsf{J}(r(i)), \, 8r(i)\big)\Big) < \frac{8}{9} \cdot \frac{1}{10^5}$$

for all large enough $i$. Fix $i$ and write $r$ for $r(i)$ and $B$ for $B_i$. For convenience of exposition we assume $r < 0$. Because of the fast decay of $j \mapsto \mu(W_{a(j)})$, it suffices to show that if $a(k)$ is the smallest unbounded scale strictly greater than $\mathsf{J}(r)$ we have

$$\mu\Big(\big\{x \in B : (S|Y_{\mathsf{J}(r)})^i(x) \in W_{a(k)} \quad \text{for some} -8|r| < i < 8|r|\big\}\Big) < \frac{3}{4}\mu(B).$$

Observe that

$$\big\{x \in X : (S|Y_{\mathsf{J}(r)})^i(x) \in W_{a(k)} \quad \text{for some} -8|r| < i < 8|r|\big\}$$

is contained in $\{x \in X : x(a(k)) \in [0, 9] \cup [d(k) - 8, 2d(k) + 1]\}$. From the definition of $B$, (5.7), if $\mathsf{J}(r) = a(k) - 1$ then

$$\big\{x \in B : (S|Y_{\mathsf{J}(r)})^i(x) \notin W_{\mathsf{J}(r)+1} = W_{a(k)} \quad \forall -8|r| < i < 8|r|\big\}$$

and otherwise

$$\nu\Big(B \setminus \big\{x \in X : (S|Y_{\mathsf{J}(r)})^i(x) \in W_{a(k)} \quad \text{for some} -8|r| < i < 8|r|\big\}\Big)$$

$$\geqslant \frac{d(k)}{2d(k) + 2}\nu(B)$$

and analogously to the proof of Lemma 2.4 we have our estimate. $\qquad\square$

We conclude this section with a second situation in which friends can be produced directly: when the time $r$ is relatively small for its scale and $\mathsf{J}(r) - 1$ is a bounded scale.

PROPOSITION 5.12. — *Fix $r \in \mathbb{Z}$. If $\mathsf{J}(r) - 1$ is bounded and $|r| \leqslant \frac{3}{4}\,\mathsf{p}(\mathsf{J}(r))$ then there exists a Borel set $B$ with $\mu(B) > 10^{-5}$ and $\Phi : B \setminus \mathsf{V}(\mathsf{J}(r), 2r) \to X$ measure preserving bijection onto its image, so that for all $x \in B$, $(x, \Phi(x))$ are $(r, \mathsf{J}(r) - 1)$ friends.*

*Proof.* — By Theorem 5.10 may assume $r$ is at an unbounded scale. Thus $\mathsf{J}(r) \in \{a(\ell)\}$. For readability, in this proof we write $b$ for $\mathsf{J}(r)$. It suffices to construct a set of $(x, x')$ that form an $(r, b-1)$-friendly pair. We present this in the case that $r > 0$. The case $r < 0$ is similar. The pairs are formed by $x$ so that $x(b-1) = 9$,

$$(5.8) \qquad x(b) < d(b) - 2, \quad \text{and} \quad x(b+1) \notin \{6, 7, 8\}.$$

and $x'$ so that

$$(5.9) \qquad x'(b-1) = 6 \quad \text{and} \quad x'(\ell) = x(\ell) \quad \forall \ \ell \neq b - 1.$$

Observe that $x \overset{b-2}{=} y$ gives (F1). Also, from (5.8) and (5.9), $(S|_{Y_{b-1}})^k(x') \in W_{b-1}$ for some $0 < k < i$ and this is not true for $x$. This gives (F5) and (F4). Further observe that $(S|_{Y_{b-1}})^i\{x, x'\} \cap W_\ell = \varnothing$ for all $0 < k < i+1$ and $\ell \geqslant b$ with $\ell \notin \{a(j)\}$. From these pairs we remove any pair $x, x'$ so that at least one is in $\mathsf{V}(b, i)$ gives (F2) and (F3). Because $x' = T^n(x)$ for $-i < n < 0$, we have that $x' \in \mathsf{V}(b, i)$ implies $x \in \mathsf{V}(b, 2i)$. Thus $(x, x')$ are $(i, b-1)$-friends whenever $x \notin \mathsf{V}(b, 2i)$. Defining $\Phi$ to change the $b-1$ digit from 9 to 6 we obtain the proposition. $\qquad \square$

## 5.3. The reduction argument

We have established via Corollary 5.11 that any non-constant $f \in \mathsf{L}^2(X, \mu)$ cannot be rigid along $\boldsymbol{r}$ if, infinitely often, the time $r(n)$ is at a bounded scale. It remains (after passing to a subsequence if necessary) to consider the complementary case: that $r(n)$ is at an unbounded scale for all $n \in \mathbb{N}$. Suppose therefore that every term of the sequence $\boldsymbol{r}$ has tower expansion

$$r(n) = \sum_{j=1}^{\mathsf{J}(r(n))} e_n(j)\,\mathsf{p}(j)$$

with $\mathsf{J}(r(n))$ unbounded for all $n \in \mathbb{N}$. This is the most technical part of the argument.

We are unable to produce $(r(n), \mathsf{J}(r(n)) - 1)$ friends or buddies directly when $r(n)$ is at an unbounded scale as each $W_{a(j)}$ consists of a large number of cylinders. Our strategy in this case is to fix in advance an unbounded scale $\tau$ that is much smaller than $\mathsf{J}(r(n))$ and to understand $T^{r(n)}$ at scale $\tau$ instead of the more natural scale of $\mathsf{J}(r(n))$. To begin to understand $T^{r(n)}$ at lower scales we use the following observation: in the picture at scale $\mathsf{J}(r(n))$ there is a large set of points $A_1^{(1)}$ where $T^{\mathsf{p}(\mathsf{J}(r(n)))} \approx \mathsf{Id}$ and a large proportion of points $A_2^{(1)}$ where $T^{\mathsf{p}(\mathsf{J}(r(n))) - \mathsf{q}(\mathsf{J}(r(n)))} \approx \mathsf{Id}$.

We then have $T^{r(n)} \approx T^{\mathsf{L}(r(n))}$ and $T^{r(n)} \approx T^{\mathsf{R}(r(n))}$ on sets $A_1^{(1)}$ and $A_2^{(1)}$ respectively where

$$\mathsf{L}(m) = m - \mathsf{e}(m)\,\mathsf{p}(\mathsf{J}(m))$$
$$\mathsf{R}(m) = m - \mathsf{e}(m)\,\mathsf{p}(\mathsf{J}(m)) + \mathsf{e}(m)\,\mathsf{q}(\mathsf{J}(m))$$

are the reduction operations that we will apply iteratively.

A consequence of the above is that we need to keep track not only of the times we obtain by applying $\mathsf{L}$ and $\mathsf{R}$ but also paired sets of points where the approximations hold. As we repeatedly reduce (i.e. compose with $\mathsf{L}$ and $\mathsf{R}$ again and again) there is a proliferation of such pairs of times and sets. We need to analyze how they are related, whether their associated times are at bounded or unbounded scales, and various other features. We prepare to undertake this analysis in Section 5.4 by carefully characterizing (cf. Corollary 5.14 and Corollary 5.16) which points are well-behaved under a single application of $\mathsf{L}$, which are well-behaved under a single application of $\mathsf{R}$, and which are neither.

The analysis of repeated reduction is the content of Proposition 5.20. It records in (E1) through (E7) what happens as we reduce a time $m$ at an unbounded scale $\mathsf{J}(m)$ to a smaller unbounded scale $\lambda^I(\mathsf{J}(m))$ above our target unbounded scale $\tau$. (E1) and (E2) records the data we keep track of: pairs of points and times $(A^{(i)}, n_j^{(i)})$ that result from composing $i$ times in various ways the maps $\mathsf{L}$ and $\mathsf{R}$ and a repository $A_0^{(i)}$ of "left over" points. (E3) records how times for different values of $i$ are related.

We see from (E4) that a positive proportion of the leftover points $A_0^{(i)}$ belong to friendly pairs. The sets $A_0^{(1)}, A_0^{(2)}, \ldots, A_0^{(I)}$ are nested, and it may be that $\mu(A_0^{(i)})$ is large enough for E4 to be compatible with Corollary 5.7.

This leaves the case that $A_0^{(I)}$ is not particularly large. When this happens we need to worry about the possibility that some scales $\mathsf{J}(n_j^{(i)})$ are smaller than $\tau$. It is certainly possible (consider $m = \mathsf{p}(\mathsf{J}(m)) + 1$ for example) that scales drop precipitously when $\mathsf{L}$ or $\mathsf{R}$ is applied. However (cf. Lemma 2.13) it turns out that $\mathsf{L}(m)$ and $\mathsf{R}(m)$ cannot both behave badly in this sense. To make use of this along the reduction procedure we pair up our times $n_1^{(i)}, \ldots, n_{k(i)}^{(i)}$ for each $1 \leqslant i \leqslant I$ in such a way that at most one of them is at a scale coarser than $\tau$. The properties we need of this pairing are given by (E5) and (E7).

All of this puts us in the following position: we have for each large enough $n$ a description of what happens as each $r(n)$ is reduced $I(n)$ times to the target unbounded scale $\tau$. If for infinitely many $n$ the set $A_0^{I(n)}$ is large enough to apply Corollary 5.7 then $\boldsymbol{r}$ is not a rigidity sequence for any non-constant $f \in \mathsf{L}^2(X, \mu)$. If that does not happen, we are left with times $n_1^{(I(n))}, \ldots, n_{k(I(n))}^{(I(n))}$ at least half of which are at scale $\tau$ and behave like $r(n)$ with respect to $T$. As $\tau$ is relatively small we can scale each of these times by a relatively small amount $d$ so that half of them have expansions at the bounded scale $\tau + 1$. We will be able to conclude the whole argument if it is the case that the times

$$dn_1^{(I(n))}, dn_2^{(I(n))}, \ldots, dn_{k(I(n))}^{(I(n))}$$

are what result if the whole procedure above is instead applied to the scaled sequence $d\boldsymbol{r}$. Thus, when reducing we want also to keep track of how scaling affects everything. What we need for that is recorded in (E6).

## 5.4. Reducing at a single scale

LEMMA 5.13. — *Fix* $r \in \mathbb{N}$ *with* $\mathsf{J}(r)$ *an unbounded scale* $a(b)$ *and fix* $x \in X$.
  (1) *If* $1 \leqslant x(\mathsf{J}(r)) \leqslant d(b) - \mathsf{e}(r)$ *then* $(S|_{Y_{\mathsf{J}(r)}})^r(x)$ *and* $(S|_{Y_{\mathsf{J}(r)}})^{\mathsf{L}(r)}(x)$ *agree in their first* $\mathsf{J}(r) - 1$ *coordinates.*
  (2) *If* $d(b) + 2 \leqslant x(\mathsf{J}(r)) \leqslant c(b) - \mathsf{e}(r)$ *then* $(S|_{Y_{\mathsf{J}(r)}})^r(x)$ *and* $(S|_{Y_{\mathsf{J}(r)}})^{\mathsf{R}(r)}(x)$ *agree in their first* $\mathsf{J}(r) - 1$ *coordinates.*
  (3) *If* $\mathsf{e}(r) \geqslant 11$ *and* $c(b) - \mathsf{e}(r) + 10 \leqslant x(\mathsf{J}(r)) \leqslant c(b)$ *then* $((S|_{Y_{\mathsf{J}(r)}})^i(x))(\mathsf{J}(r)) = 0$ *for some* $0 \leqslant i < r$. *If* $x(\mathsf{J}(r) + 1) = 7$ *as well there is* $0 \leqslant i \leqslant r$ *with* $(S|_{Y_{\mathsf{J}(r)}})^i(x) \in W_{\mathsf{J}(r)+1}$.

*Proof.*

(1). — The effect of $(S|_{Y_{\mathsf{J}(r)}})^{r-\mathsf{L}(r)} = (S|_{Y_{\mathsf{J}(r)}})^{\mathsf{e}(r)\,\mathsf{p}(r)}$ on any such $x$ is to increase the $\mathsf{J}(r)$ coordinate by the amount $\mathsf{e}(r)$ to at most $d(b)$. When we apply $(S|_{Y_{\mathsf{J}(r)}})^{\mathsf{L}(r)}$ to both $x$ and $((S|_{Y_{\mathsf{J}(r)}})^{r-\mathsf{L}(r)})(x)$ the resulting points continue to agree in their first $\mathsf{J}(r) - 1$ coordinates by Lemma 2.12. Indeed, when $\mathsf{L}(r)$ is negative the condition $1 \leqslant x(\mathsf{J}(r))$ together with $|\mathsf{L}(r)| \leqslant \mathsf{p}(\mathsf{J}(r))$ avoids $((S|_{Y_{\mathsf{J}(r)}})^{\mathsf{L}(r)})(x)$ having a negative $\mathsf{J}(r)$ coordinate, and when $\mathsf{L}(r)$ is positive $|\mathsf{L}(r)| \leqslant \mathsf{p}(\mathsf{J}(r)) - \mathsf{q}(\mathsf{J}(r))$ prevents $(S|_{Y_{\mathsf{J}(r)}})^r(x)$ from having a $\mathsf{J}(r)$ coordinate strictly larger than $d(b) + 1$.

(2). — The effect of $(S|_{Y_{\mathsf{J}(r)}})^{r-\mathsf{R}(r)}$ on any such $x$ is to increase the $\mathsf{J}(r)$ coordinate by the amount $\mathsf{e}(r)$ to at most $c(b)$. When we apply $(S|_{Y_{\mathsf{J}(r)}})^{\mathsf{R}(r)}$ to both $x$ and $((S|_{Y_{\mathsf{J}(r)}})^{r-\mathsf{R}(r)})(x)$ the resulting points continue to agree in their first $\mathsf{J}(r) - 1$ coordinates by Lemma 2.12. Indeed, when $\mathsf{R}(r)$ is negative the condition $d(b) + 2 \leqslant x(\mathsf{J}(r))$ together with $|\mathsf{R}(r)| \leqslant \mathsf{p}(\mathsf{J}(r)) - \mathsf{q}(\mathsf{J}(r))$ avoids $((S|_{Y_{\mathsf{J}(r)}})^{\mathsf{R}(r)})(x)$ having a $\mathsf{J}(r)$ coordinate less than or equal to $d(b)$, and when $\mathsf{R}(r)$ is positive $|\mathsf{R}(r)| \leqslant \mathsf{p}(\mathsf{J}(r)) - \mathsf{q}(\mathsf{J}(r))$ prevents $(S|_{Y_{\mathsf{J}(r)}})^r(x)$ from going too far up the first tower i.e. having $\mathsf{J}(r)$ coordinate greater than 0 or have $\mathsf{J}(r)$ coordinate equal to 0 and being in a level that would be in $W_{\mathsf{J}(r)}$ on the right side towers.

(3). — Note that $|r - \mathsf{e}(r)\,\mathsf{p}(\mathsf{J}(r))| \leqslant \mathsf{p}(\mathsf{J}(r))$ by Corollary 2.9 because
$$\mathsf{J}(r - \mathsf{e}(r)\,\mathsf{p}(r)) \leqslant \mathsf{J}(r) - 1$$
so that for any $x$ that satisfies $c(b) - \mathsf{e}(r) + 10 \leqslant x(\mathsf{J}(r)) \leqslant c(b)$ there is $0 \leqslant j \leqslant r$ such that the $\mathsf{J}(r)$ coordinate of $(S|_{Y_{\mathsf{J}(r)}})^j(x)$ is 7. In particular, there must be some $0 \leqslant i \leqslant r$ for which $(S|_{Y_{\mathsf{J}(r)}})^i(x)$ belongs to $[7^{\mathsf{J}(r)}]$ and if additionally $x(\mathsf{J}(r) + 1) = 7$ then we will have $(S|_{Y_{\mathsf{J}(r)}})^j(x)$ in $W_{\mathsf{J}(r)+1}$. Indeed, $(S|_{Y_{\mathsf{J}(r)}})^\ell(x)(\mathsf{J}(r))$ going from $c(b)$ to 0 makes $(S|_{Y_{\mathsf{J}(r)}})^\ell(x)(\mathsf{J}(r) + 1)$ go from 7 to 8. $\square$

COROLLARY 5.14. — *Fix* $r \in \mathbb{N}$ *with* $\mathsf{J}(r)$ *an unbounded scale* $a(b)$ *and fix* $x \in X$.
  (1) *If* $9 \leqslant x(\mathsf{J}(r)) \leqslant d(b) - \mathsf{e}(r)$ *and* $x \notin \mathsf{V}(\mathsf{J}(r), r)$ *then*
$$(S|_{Y_{\mathsf{J}(r)}})^r(x) = T^r(x) \qquad\qquad (S|_{Y_{\mathsf{J}(r)}})^{\mathsf{L}(r)}(x) = T^{\mathsf{L}(r)}(x)$$

and these two points agree on the first $\mathsf{J}(r) - 1$ coordinates.
(2) If $d(b) + 2 \leqslant x(\mathsf{J}(r)) \leqslant c(b) - \mathsf{e}(r)$ and $x \notin \mathsf{V}(\mathsf{J}(r), r)$ then

$$(S|Y_{\mathsf{J}(r)})^r(x) = T^r(x) \qquad\qquad (S|Y_{\mathsf{J}(r)})^{\mathsf{R}(r)}(x) = T^{\mathsf{R}(r)}(x)$$

and these two points agree in their first $\mathsf{J}(r) - 1$ coordinates.
(3) If $\mathsf{e}(r) \geqslant 11$ and $c(b) - \mathsf{e}(r) + 10 \leqslant x(\mathsf{J}(r)) \leqslant c(b)$ and $x(\mathsf{J}(r) + 1) = 7$ and $x \notin \mathsf{V}(\mathsf{J}(r), r + 1)$ then $(y, x)$ are $(r, \mathsf{J}(r) + 1)$-friends whenever $y$ satisfies $x(k) = y(k)$ for all $k \leqslant \mathsf{J}(r)$ and $y(\mathsf{J}(r) + 1) = 3$ and $y \notin \mathsf{V}(\mathsf{J}(r), r)$.

*Proof.*

(1). — The condition $x(\mathsf{J}(r)) \geqslant 9$ guarantees the $\mathsf{J}(r)$ coordinate of $(S|Y_{\mathsf{J}(r)})^i(x)$ is not 7 for all $-|\mathsf{L}(r)| \leqslant i \leqslant r$ by Lemma 2.12. Together with the assumption that $x$ does not belong to $\mathsf{V}(\mathsf{J}(r), r)$ it follows that $(S|Y_{\mathsf{J}(r)})^i(x) \notin W_k$ for all $-|\mathsf{L}(r)| \leqslant i \leqslant r$ and all $k > \mathsf{J}(r)$. Thus $(S|Y_{\mathsf{J}(r)})^i(x) = T^i(x)$ for all $-|\mathsf{L}(r)| \leqslant i \leqslant r$. As $x$ satisfies in particular the hypothesis of (1) we get the desired agreement in the first $\mathsf{J}(r) - 1$ coordinates.

(2). — Note there is no possibility of $(S|Y_{\mathsf{J}(r)})^i(x)$ belonging to some $W_k$ with $k > \mathsf{J}(r)$ for some $-|\mathsf{R}(r)| \leqslant i \leqslant r$ and we have the desired conclusion by appealing to (2).

(3). — We need to verify the properties in Definition 5.8. Condition (F1) is one of our assumptions. To prove (F2) and (F4) combine $y \notin \mathsf{V}(\mathsf{J}(r), r)$ with $y(\mathsf{J}(r) + 1) = 3$ to get $(S|Y_{\mathsf{J}(r)})^i(y) \notin W_k$ for all $k \geqslant \mathsf{J}(r) + 1$ and all $1 \leqslant i \leqslant r$. For (F3) and (F5) combine $x \notin \mathsf{V}(\mathsf{J}(r), r + 1)$ with the facts that there is a unique $1 \leqslant i \leqslant r$ for which $(S|Y_{\mathsf{J}(r)})^i(x)$ belongs to $[7^{\mathsf{J}(r)}]$ and that $x(\mathsf{J}(r) + 1) = 7$ allows (3) to guarantee there is $1 \leqslant i < r$ with $(S|Y_{\mathsf{J}(r)})^i(x)$ in $W_{\mathsf{J}(r)+1}$. $\qquad\square$

We have additionally the following reduction statements when the time $r$ is negative. The proofs are entirely analogous.

LEMMA 5.15. — *Fix $r \in -\mathbb{N}$ with $\mathsf{J}(r)$ an unbounded scale $a(b)$ and fix $x \in X$.*
(1) If $-\mathsf{e}(r) + 1 \leqslant x(\mathsf{J}(r)) \leqslant d(b) - 1$ then $(S|Y_{\mathsf{J}(r)})^r(x)$ and $(S|Y_{\mathsf{J}(r)})^{\mathsf{L}(r)}(x)$ agree in their first $\mathsf{J}(r) - 1$ coordinates.
(2) If $d(b) + 2 - \mathsf{e}(r) \leqslant x(\mathsf{J}(r)) \leqslant c(b) - 2$ then $(S|Y_{\mathsf{J}(r)})^r(x)$ and $(S|Y_{\mathsf{J}(r)})^{\mathsf{R}(r)}(x)$ agree in their first $\mathsf{J}(r) - 1$ coordinates.
(3) If $\mathsf{e}(r) \leqslant -11$ and $10 \leqslant x(\mathsf{J}(r)) \leqslant -\mathsf{e}(r) + 1$ then $((S|Y_{\mathsf{J}(r)})^i(x))(\mathsf{J}(r)) = 0$ for some $0 \geqslant i \geqslant r$. If $x(\mathsf{J}(r) + 1) = 8$ as well there is $0 \geqslant i \geqslant r$ with $(S|Y_{\mathsf{J}(r)})^i(x) \in W_{\mathsf{J}(r)+1}$.

COROLLARY 5.16. — *Fix $r \in -\mathbb{N}$ with $\mathsf{J}(r)$ an unbounded scale $a(b)$ and fix $x \in X$.*
(1) If $-\mathsf{e}(r) + 10 \leqslant x(\mathsf{J}(r)) \leqslant d(b) - 1$ and $x \notin \mathsf{V}(\mathsf{J}(r), r)$ then

$$(S|Y_{\mathsf{J}(r)})^r(x) = T^r(x) \qquad\qquad (S|Y_{\mathsf{J}(r)})^{\mathsf{L}(r)}(x) = T^{\mathsf{L}(r)}(x)$$

and these two points agree in the first $\mathsf{J}(r) - 1$ coordinates.
(2) If $d(b) + 2 - \mathsf{e}(r) \leqslant x(\mathsf{J}(r)) \leqslant c(b) - 2$ and $x \notin \mathsf{V}(\mathsf{J}(r), r)$ then

$$(S|Y_{\mathsf{J}(r)})^r(x) = T^r(x) \qquad\qquad (S|Y_{\mathsf{J}(r)})^{\mathsf{R}(r)}(x) = T^{\mathsf{R}(r)}(x)$$

and these two points agree in their first $\mathsf{J}(r) - 1$ coordinates.

(3) *If* $\mathsf{e}(r) \leqslant -11$ *and* $10 \leqslant x(\mathsf{J}(r)) \leqslant -\mathsf{e}(r) + 1$ *and* $x(\mathsf{J}(r) + 1) = 8$ *and*
$x \notin \mathsf{V}(\mathsf{J}(r), r + 1)$ *then* $(y, x)$ *are* $(r, \mathsf{J}(r) + 1)$-*friends whenever* $y$ *satisfies*
$x(k) = y(k)$ *for all* $k \leqslant \mathsf{J}(r)$ *and* $y(\mathsf{J}(r) + 1) = 3$ *and* $y \notin \mathsf{V}(\mathsf{J}(r), r)$.

COROLLARY 5.17. — *Fix* $r \in \mathbb{Z}$ *with* $\mathsf{J}(r)$ *an unbounded scale* $a(b)$ *and* $|\mathsf{e}(r)| < 10$.
*There exists* $B \in X$ *measurable and* $\Phi : B \to X$ *bijective and measurable on its
image so that* $(x, \phi(x))$ *are* $(r, \mathsf{J}(r) + 1)$-*friends and*

$$\mu(B) \geqslant \frac{1}{2} \frac{1}{20} \frac{1}{20} \frac{1}{d(b)}.$$

*Proof.* — We prove the case $r > 0$. The case $r < 0$ is analogous. Let $J = \mathsf{J}(r)$ and
let

$$B = \left\{ \begin{array}{l} y \in \Omega : (S|Y_{J(r)})^i(x) \in W_{\mathsf{J}(r)+1} \\ \text{for some } 0 \leqslant i < r \quad \text{and} \quad x(\mathsf{J}(r) + 2) = 5 \end{array} \right\} \cap X$$

which is the intersection with $X$ of a union of cylinders of length $\mathsf{J}(r)$. Let $\phi : B \to X$
by $\phi(x)_j = x_j$ for all $j \neq J + 1$ and $\phi(x)_{J+1} = 4$. As in the proof of Corollaries 5.14
and 5.16 (3) we have that $x, \phi(x)$ are $(r, J + 1)$-friends so long as $x \notin V(b, r)$. For
our measure estimate, we have that

$$\mu\left(\left\{x : T^i x \in S^{-1} W_{J+1} \quad \text{for some } 0 \leqslant i < r\right\}\right) > \frac{r}{t(J+1)} > \frac{1}{20 d(J)}.$$

By our independence result, Lemma 2.4 applied with $k = 1$, and the fact that there
are 10 symbols we have that adding the condition that $x(J + 2) = 5$ reduces the
measure by less than a factor of $2 \cdot 10$. Lastly, the condition not being in $V(b, r)$
removes less than

$$2 \cdot r \sum_{a(j) > J(r)} \frac{Q(j)}{P(j)} < 2 \cdot 10 p(\mathsf{J}(r)) \sum_{a(j) > J(r)} \frac{Q(j)}{P(j)} < \frac{1}{2} \cdot \frac{1}{20} \frac{1}{20} \frac{1}{d(J)}$$

giving the claim.                                                                 $\square$

By similar estimates to what we have done earlier Corollary 5.14 (3), Corollary 5.16 (3) and Corollary 5.17 give the following as well.

LEMMA 5.18. — *Fix* $r \in \mathbb{N}$ *with* $\mathsf{J}(r)$ *unbounded. With*

$A_0 = \{x \in X : x \text{ does not satisfy first assumption of Corollaries 5.14 (1) or (2)}\}$

*there is* $B \subset A_0$, *and* $\Phi : B \to X$, *a measure preserving injection so that*
$\mu(B) > \mu(A_0) 10^{-6}$ *and* $(x, \Phi(x))$ *are* $(r, \mathsf{J}(r) + 1)$-*friends.*
*The same is true with*

$A_0 = \{x \in X : x \text{ does not satisfy first assumption of Corollaries 5.16 (1) or (2)}\}$

*instead.*

## 5.5. Key result for unbounded scales

In this section we state and prove the key result needed for the proof of Theorem 1.7.
Fix an unbounded scale $\tau \in \mathbb{N}$. Given $r \in \mathbb{Z}$ with $\mathsf{J}(r) > \tau$ the proposition keeps
track – using Corollary 5.14 and Corollary 5.16 – of the behavior of points under

first $T^{\mathsf{L}(r)}$ and $T^{\mathsf{R}(r)}$ and then under $T^{\mathsf{L}(\mathsf{L}(r))}, T^{\mathsf{R}(\mathsf{L}(r))}, T^{\mathsf{L}(\mathsf{R}(r))}$ and $T^{\mathsf{R}(\mathsf{R}(r))}$ and so on until we have applied compositions of $\mathsf{L}$ and $\mathsf{R}$ enough times to arrive at the scale $\tau$. There are various intricacies to keep track of. For example, it may be the case that $\mathsf{L}(\mathsf{R}(\mathsf{L}(r)))$ is at a bounded scale, it may be that $\mathsf{e}(\mathsf{L}(\mathsf{L}(r)))$ is either very small or very large, and it may be that some compositions of $\mathsf{L}$ and $\mathsf{R}$ reduce $r$ to a scale smaller than $\tau$. We begin with a preparatory lemma.

LEMMA 5.19. — *If $0 < |k| \leqslant 2d(r) + 3$ then either $\mathsf{J}(k\,\mathsf{Q}(r)) = a(r-1)$ or $k\,\mathsf{Q}(r)$ is at a bounded scale. Moreover, either*

$$\mathsf{J}\big(\mathsf{L}(k\,\mathsf{Q}(r))\big) = \mathsf{J}\big(\mathsf{R}(k\,\mathsf{Q}(r))\big) = a(r-2)$$

*or at least one of $\mathsf{J}(\mathsf{L}(k\,\mathsf{Q}(r)))$, $\mathsf{J}(\mathsf{R}(k\,\mathsf{Q}(r)))$ is at a bounded scale. Moreover, in the first case, both $\mathsf{J}(\mathsf{L}(k\,\mathsf{Q}(r)))$ and $\mathsf{J}(\mathsf{R}(k\,\mathsf{Q}(r)))$ have the form $k'\,\mathsf{Q}(r-1)$ for $|k'| \leqslant 2d(r-2) + 3$.*

*If $\mathsf{J}(b) = a(r-\ell)$ for $\ell \geqslant 2$ and $\mathsf{J}(k\,\mathsf{Q}(r) + b) < a(r)$ then $\mathsf{J}(k\,\mathsf{Q}(r) + b) = a(r-1)$ or $k\,\mathsf{Q}(r) + b$ is at a bounded scale. Moreover*

$$\mathsf{J}\big(\mathsf{L}(k\,\mathsf{Q}(r) + b)\big) = \mathsf{J}\big(\mathsf{R}(k\,\mathsf{Q}(r) + b)\big) = a(r-2)$$

*or at least one of $\mathsf{J}(\mathsf{L}(k\,\mathsf{Q}(r) + b))$, $\mathsf{J}(\mathsf{R}(k\,\mathsf{Q}(r) + b))$ is at a bounded scale greater than $a(r-2)$. Moreover, in the first case, both $\mathsf{J}(\mathsf{L}(k\,\mathsf{Q}(r) + b))$ and $\mathsf{J}(\mathsf{R}(k\,\mathsf{Q}(r) + b))$ have the form $k\,\mathsf{Q}(r-1) + b$ for $|k| < 2d(r-1) + 2 = c(a(r-1)) + 1$.*

*Proof.* — The first claim is because $\mathsf{P}(r-1) < \mathsf{Q}(r) < \mathsf{p}(a(r-1) + 1)$ and $(2d(r) + 2)\,\mathsf{p}(a(r-1) + 1)$ is much smaller than $\mathsf{p}(a(r) - 1)$. In particular, $k\,\mathsf{Q}(r)$ is at a scale between $a(r-1)$ and $a(r) - 1$ and any such scale other than $a(r-1)$ is a bounded scale.

Now we consider the reductions $\mathsf{L}$ and $\mathsf{R}$. Recall, $k\,\mathsf{Q}(r) = k(2\mathsf{P}(r-1) - \mathsf{Q}(r-1))$, so if it is not at a bounded scale, $\mathsf{J}(k\,\mathsf{Q}(r)) = a(r-1)$ and by the greedy algorithm, $\mathsf{e}(k\,\mathsf{q}(r)) = 2k$ because $(2d(r) + 3)\,\mathsf{Q}(r-1)$ is much smaller than $\mathsf{P}(r-1)$. If $\mathsf{J}(\mathsf{L}(k\,\mathsf{Q}(r)))$ is at an unbounded scale, similarly to above $\mathsf{J}(\mathsf{L}(k\,\mathsf{Q}(r)) = \mathsf{J}(k\,\mathsf{Q}(r-1)) = a(r-2)$. Because $\mathsf{R}(k\,\mathsf{Q}(r)) = 2k\,\mathsf{Q}(r-1) + \mathsf{L}(k\,\mathsf{Q}(r)) = 3k\,\mathsf{Q}(r-1)$ we have that it is either at a bounded scale or at $a(r-2)$. Now for the bound on $k'$ if it was not true $\mathsf{L}(\mathsf{J}(\mathsf{Q}(r)))$ would be at a bounded scale.

The second paragraph follows similarly to the above. In particular, for the "moreover" parts of the second paragraph, $b$ is small enough that adding $b$ does not affect the greedy algorithm. □

PROPOSITION 5.20. — *Fix $\tau \in \mathbb{N}$ a target scale greater than or equal to $a(10)$. Fix $m \in \mathbb{N}$ with $\mathsf{J}(m)$ unbounded and $\mathsf{J}(m) > \tau$. Fix $I \in \mathbb{N}$ with $\lambda^I(\mathsf{J}(m)) \geqslant \tau$. For every $1 \leqslant i \leqslant I$ there is a tuple*

$$(5.10) \qquad \left( A_0^{(i)}, A_1^{(i)}, n_1^{(i)}, \ldots, A_{k(i)}^{(i)}, n_{k(i)}^{(i)} \right)$$

*with all the following properties.*

(E1) $A_j^{(i)} \subset X$ *for all $0 \leqslant j \leqslant k(i)$ and $n_j^{(i)} \in \mathbb{Z}$ for all $1 \leqslant j \leqslant k(i)$ and $k(i) \in 2\mathbb{N} \cup \{0\}$.*

(E2) *The sets $A_0^{(i)}, \ldots, A_{k(i)}^{(i)}$ form a measurable partition of $X$ in which each $A_j^{(i)}$ is the intersection with $X$ of a finite union of cylinders, each of which is defined by fixing values at unbounded scales between $\mathsf{J}(m)$ and $\lambda^{i-1}(\mathsf{J}(m))$.*

(E3) *Every $1 \leqslant j \leqslant k(i)$ has an ancestor $1 \leqslant \alpha(j) \leqslant k(i-1)$. If $j \neq 0$ and $x \in A_j^{(i)}$ and*

$$(5.11) \qquad x \notin \bigcup_{s=0}^{i} \mathsf{V}\Big(\mathsf{J}\big(n_{\alpha^s(j)}^{(i-s)}\big), n_{\alpha^s(j)}^{(i-s)}\Big)$$

*then*

$$(5.12) \qquad (T^m(x))(\ell) = \Big(T^{n_j^{(i)}}(x)\Big)(\ell), \quad \forall\, \ell \leqslant \lambda^{i-1}(\mathsf{J}(m)) - 1.$$

(E4) *There exists $F \subset A_0^{(i)}$ with $\mu(F) > 10^{-9}\mu(A_0^{(i)})$ and $\psi : F \to X$ measure preserving and $j \geqslant \tau$ so that $(x, \psi(x))$ are $(m,j)$ buddies and in particular $(m,\tau)$ buddies.*

(E5) *The numbers $1, \ldots, k(i)$ are paired so that each $1 \leqslant j \leqslant k(i)$ is in exactly one pair. Moreover, whenever $j$ and $h$ are paired then at least one of $n_j^{(i)}$ and $n_h^{(i)}$ is at scale $\lambda^i(\mathsf{J}(m))$.*

(E6) *If $j$ and $h$ are paired with $\lambda^i(\mathsf{J}(m)) = \mathsf{J}(n_j^{(i)}) \neq \mathsf{J}(n_h^{(i)})$ then*

$$n_j^{(i)} = k\,\mathsf{q}\Big(\lambda^{i-1}(\mathsf{J}(m))\Big) + n_h^{(i)} \quad \text{with} \quad 2|k| \leqslant c\Big(\lambda^{i-1}\mathsf{J}(m)\Big) + 1.$$

(E7) *If $j$ and $h$ are paired the cylinders defining $A_j^{(i)}$ and $A_h^{(i)}$ differ in at most one unbounded scale. In that index one of them satisfies the first hypothesis of either Lemma 5.13(1) or Lemma 5.15(1) and the other satisfies the first hypothesis of either Lemma 5.13(2) or Lemma 5.15(2) respectively with $\max\{|n_j|, |n_h|\}$ in place of $r$ therein.*

*Proof.* — Let $\tau$, $m$ and $I$ be as in the statement. The proof is recursive.

*Initial step.* — First suppose both $\mathsf{J}(\mathsf{L}(m))$ and $\mathsf{J}(\mathsf{R}(m))$ are unbounded scales. By Lemma 2.13 neither can be at scale $\mathsf{J}(m)$ and at least one is at scale $\lambda(\mathsf{J}(m))$. If $m$ is positive let $A_1^{(1)}$ and $A_2^{(1)}$ be the sets of points in $X$ satisfying the hypothesis $9 \leqslant x(\mathsf{J}(m)) \leqslant d(b) - \mathsf{e}(r)$ of Corollary 5.14(1) and the hypothesis $d(b) + 2 \leqslant x(\mathsf{J}(m)) \leqslant c(b) - \mathsf{e}(r)$ of Corollary 5.14(2) respectively. When $m$ is negative use corresponding hypotheses of Corollaries 5.16(1) and (2) respectively instead. In either case define $A_0^{(1)}$ to be the intersection with $X$ of the cylinders defined by the $\mathsf{J}(m)$ coordinate that contribute neither to $A_1^{(1)}$ nor to $A_2^{(1)}$. Thus $A_0^{(1)}, A_1^{(1)}, A_2^{(1)}$ is a measurable partition of $X$ and (E2) holds. Set $k(1) = 2$ and $n_1^{(1)} = \mathsf{L}(m)$ and $n_2^{(1)} = \mathsf{R}(m)$. Thus (E1) holds. We get (E3) from the conclusions of Corollary 5.14(1) and (2) respectively. (When $m$ is negative use Corollary 5.16 instead.) As $k(1) = 2$ we are forced to pair $n_1^{(1)}$ and $n_2^{(1)}$. By Lemma 2.13 at least one of $n_1^{(1)}$ and $n_2^{(2)}$ is at scale $\lambda(\mathsf{J}(m))$ giving (E5). As $|\mathsf{L}(m) - \mathsf{R}(m)| = |\mathsf{e}(m)|\,\mathsf{q}(\mathsf{J}(m))$ and $2|\mathsf{e}(m)| \leqslant c(\mathsf{J}(m)) + 1$ we have (E6). Property (E7) holds as the cylinders defining $A_1^{(1)}$ and $A_2^{(1)}$ are defined in the $\mathsf{J}(m)$ coordinate alone to satisfy the appropriate hypotheses.

Secondly, suppose instead that at least one of $\mathsf{L}(m)$ and $\mathsf{R}(m)$ is at a bounded scale greater than or equal to $\tau$. In this case set $k(1) = 0$ and $A_0^{(1)} = X$. All properties except (E4) are immediate or vacuous. We now treat (E4), assuming without loss of generality that $\mathsf{J}(\mathsf{L}(m)) = k > \tau$, a bounded scale. By Theorem 5.10 we get sufficiently many $(\mathsf{L}(m), \mathsf{J}(\mathsf{L}(m)))$-friends. By Corollary 5.14 (1) if $m > 0$ or Corollary 5.16 (1) if $m < 0$, Lemmas 5.9 and 5.4 we have (E4).

Finally, suppose one of $\mathsf{L}(m)$, $\mathsf{R}(m)$ is at an unbounded scale and the other is at a bounded scale less than $\tau$. (We will not find both $\mathsf{L}(m)$ and $\mathsf{R}(m)$ at bounded scales less than $\tau$ because $\mathsf{L}(m) - \mathsf{R}(m) = \mathsf{e}(m)\,\mathsf{q}(\mathsf{J}(m))$ forces at least one of $\mathsf{L}(m)$ and $\mathsf{R}(m)$ to be at a scale at least as large as $\lambda(\mathsf{J}(m)) - 1$.) Then we proceed as in the first case, except that for (E5) we instead note here that if $\mathsf{L}(m)$ is not at scale $\lambda(\mathsf{J}(m))$ then $\mathsf{J}(\mathsf{L}(m)) \leqslant \lambda^2(\mathsf{J}(m))$ and so $\mathsf{J}(\mathsf{R}(m)) = \lambda(\mathsf{J}(m))$ because $\mathsf{R}(m) = \mathsf{L}(m) + \mathsf{e}(m)\,\mathsf{q}(\mathsf{J}(m))$ gives $\mathsf{J}(\mathsf{R}(m)) \geqslant \lambda(\mathsf{J}(m))$ and it cannot be larger than $\lambda(\mathsf{J}(m))$.

*The recursion.* — For the inductive step, suppose for some $1 \leqslant i < I$ that a tuple

$$\left( A_0^{(i)}, A_1^{(i)}, n_1^{(i)}, \ldots, A_{k(i)}^{(i)}, n_{k(i)}^{(i)} \right)$$

satisfying all the properties has been produced. If $k(i) = 0$ we set $k(i+1) = 0$ and $A_0^{(i+1)} = A_0^{(i)}$. All properties continue to hold.

Assuming, then, that $k(i) > 0$, we proceed by reducing further each $n_j^{(i)}$ unless one of the following holds. Say that the pair $h \neq j$ in $\{1, \ldots, k(i)\}$ is *bad* if at least one of the following is true.

- $\mathsf{J}(n_j^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $n_j^{(i)} < \frac{3}{4}\,\mathsf{p}(\lambda^i(\mathsf{J}(m)))$,
- $\mathsf{J}(n_j^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $\mathsf{L}(n_j^{(i)})$ is at a bounded scale not smaller than $\tau$,
- $\mathsf{J}(n_j^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $\mathsf{R}(n_j^{(i)})$ is at a bounded scale not smaller than $\tau$,
- $\mathsf{J}(n_h^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $n_h^{(i)} < \frac{3}{4}\,\mathsf{p}(\lambda^i(\mathsf{J}(m)))$,
- $\mathsf{J}(n_h^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $\mathsf{L}(n_h^{(i)})$ is at a bounded scale not smaller than $\tau$,
- $\mathsf{J}(n_h^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $\mathsf{R}(n_h^{(i)})$ is at a bounded scale not smaller than $\tau$.

When the pair to which $1 \leqslant j \leqslant k(i)$ belongs is not bad we call $j$ *reducible* if $\mathsf{J}(n_j^{(i)}) = \lambda^i(\mathsf{J}(m))$. Note that (E5) says that if neither element of a pair is bad then at least one element of each pair is reducible. As a reducible $j$ does not belong to a bad pair it must be the case that $\mathsf{L}(n_j^{(i)})$ and $\mathsf{R}(n_j^{(i)})$ are at unbounded scales. For each $1 \leqslant j \leqslant k(i)$ that is reducible set

$$F(j) = \begin{cases} \left\{ x \in A_j^{(i)} : x\left( \mathsf{J}\left( n_j^{(i)} \right) \right) \in [0, 8] \cup [d - e + 1, d + 1] \cup [c - e + 1, c] \right\} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } n_j^{(i)} > 0 \\ \left\{ x \in A_j^{(i)} : x\left( \mathsf{J}\left( n_j^{(i)} \right) \right) \in [0, 9 - e] \cup [d, d + 1 - e] \cup [c - 1, c] \right\} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } n_j^{(i)} < 0 \end{cases}$$

where $c = c(\lambda^i(\mathsf{J}(m)))$ and $d = d(\lambda^i(\mathsf{J}(m)))$ and $e = \mathsf{e}(n_j^{(i)})$. For each $1 \leqslant h \leqslant k(i)$ that is neither in a bad pair nor reducible set

$$F(h) = \begin{cases} \left\{ x \in A_h^{(i)} : x\left(\mathsf{J}\left(n_j^{(i)}\right)\right) \in [0, 8] \cup [d - e + 1, d + 1] \cup [c - e + 1, c] \right\} \\ \hspace{8cm} \text{if } n_k^{(i)} > 0 \\ \left\{ x \in A_h^{(i)} : x\left(\mathsf{J}\left(n_j^{(i)}\right)\right) \in [0, 9 - e] \cup [d, d + 1 - e] \cup [c - 1, c] \right\} \\ \hspace{8cm} \text{if } n_k^{(i)} < 0 \end{cases}$$

where $c = c(\lambda^i(\mathsf{J}(m)))$ and $d = d(\lambda^i(\mathsf{J}(m)))$ and $e = \mathsf{e}(n_k^{(i)})$ and $j$ is the (necessarily reducible) pair of $h$ as in (E7).

Note all $1 \leqslant h \leqslant k(i)$ that are neither bad nor reducible are paired with an index that is reducible and not bad. We define

$$(5.13) \qquad A_0^{(i+1)} = A_0^{(i)} \cup \bigcup_{j \text{ in a bad pair}} A_j^{(i)} \cup \bigcup_{j \text{ reducible}} F(j) \cup \bigcup_{\substack{h \text{ neither bad} \\ \text{nor reducible}}} F(h)$$

and note that the sets making up (5.13) are pairwise disjoint as $A_0^{(i)}, \ldots, A_{k(i)}^{(i)}$ is a partition of $X$. We now verify (E4) for $A_0^{(i+1)}$ by checking (E4) holds for each of its constituents.

- It follow by the inductive assumption for $A_0^{(i)}$.
- Fix a bad pair $h \neq j$. We first check that

$$(5.14) \qquad \frac{1}{2} < \frac{\mu\left(A_j^{(i)}\right)}{\mu\left(A_k^{(i)}\right)} < 2$$

holds. By (E2) and (E7) there exist 3 unions of cylinders, $B, C, D$ all defined only by fixing indices at unbounded scales greater than $\tau$ so that $A_j^{(i)} = B \cap C$ and $A_k^{(i)} = B \cap D$. What is more, by (E7) and our construction, up to flipping the role of $C$ and $D$, $C$ is defined by all of the left towers at one unbounded index, say $a(n)$ and $D$ is defined by all of the right towers $a(n)$. Moreover, this index is smaller than the indices defining $B$. Now by a similar computation to those we have done before

$$1 < \frac{(d(n) + 1)\mathsf{P}(n) - \nu(X)^{-1}\nu\left(\bigcup_{\ell > n} W_{a(\ell)}\right)}{(d(n) + 1)(\mathsf{P}(n) - \mathsf{Q}(n))} \leqslant \frac{\mu(C)}{\mu(D)} < 1 + 2\frac{\mathsf{Q}(n)}{\mathsf{P}(n)}$$

and (5.14) follows from Lemma 2.4 and the fact that the index defining $C$ and $D$ is at least $2^{2^5}$ smaller than the smallest index defining $B$.

With (5.14) established it suffices to find a positive proportion of $(m, \tau)$ buddies in either $A_j^{(i)}$ or $A_k^{(i)}$. How this is done depends on why the pair $h, j$ is bad.

— If $\mathsf{J}(n_j^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $s \in \{\mathsf{L}(n_j^{(i)}), \mathsf{R}(n_j^{(i)})\}$ at a bounded scale not smaller than $\tau$ apply Theorem 5.10 with $s$ in place of $r$ and note that the set $B$ therein is defined by coordinates no larger than $\mathsf{J}(s) + 1$ while $A_j^{(i)}$ is defined by coordinates no smaller than $\lambda^{i-1}(\mathsf{J}(m))$. We may therefore

apply Lemma 2.4 to get

$$\mu\Big(B \cap A_j^{(i)}\Big) \geqslant \Big(1 - \frac{1}{2^{2^5}}\Big)\mu(B)\mu\Big(A_j^{(i)}\Big)$$

whence a positive proportion of $A_j^{(i)}$ intersects $B$. Removing $\mathsf{V}(\mathsf{J}(s), 8s)$ and (5.11) leaves us with a positive proportion of $(m, \tau)$ buddies, as in the proof of Corollary 5.11.

- If $\mathsf{J}(n_j^{(i)}) = \lambda^i(\mathsf{J}(m))$ and $n_j^{(i)} < \frac{3}{4}\mathsf{p}(\lambda^i(\mathsf{J}(m)))$ apply Proposition 5.12.
- If instead $\mathsf{J}(n_h^{(i)}) = \lambda^i(\mathsf{J}(m))$ the arguments are exactly the same upon replacing $j$ with $h$.

- When $j$ is reducible or neither bad nor reducible, we must show that $F(j)$ has a positive proportion of $(m, j)$-buddies for $j \geqslant \tau$.
  - We claim that it suffices to show this when $j$ is reducible, because the measure of the paired cylinders differs by less than a factor of 2 and when $h$ is neither bad nor reducible it is paired with $j$ which is reducible. Indeed, by (E7) they differ by at most one unbounded symbol. We consider the other (necessarily unbounded) entries defining these sets (E2) and apply Lemma 2.4 to these cylinders and the cylinders defined by the (necessarily unbounded) position where they are defined to differ.
  - By Lemma 5.18, if $j$ is reducible the set of $(n_j^{(i)}, \lambda^i(\mathsf{J}(m)))$-friends is at least $\frac{1}{10^6}\mu(F(j))$.
  - By Lemmas 5.9 and 5.4 pairs that are $(n_j^{(i)}, \lambda^i(\mathsf{J}(m)))$-friends and for which (5.12) hold for both elements of the pair are automatically $(m, \tau)$-buddies.

Now we define $k(i+1)$ and the sets $A_1^{(i+1)}, \ldots, A_{k(i+1)}^{(i+1)}$ and the times $n_1^{(i+1)}, \ldots,$ $n_{k(i+1)}^{(i+1)}$. Write

$$G = \{1 \leqslant j \leqslant k(i) : j \text{ reducible}\} \cup \{1 \leqslant h \leqslant k(i) : h \text{ neither bad nor reducible}\}$$

which consists of those $1 \leqslant j \leqslant k(i)$ belonging to a pair that is not bad. Temporarily write $c = c(\lambda^i(\mathsf{J}(m)))$ and $d = d(\lambda^i(\mathsf{J}(m)))$. There are two cases to consider.

- When $j \in G$ is reducible and its pair is too, write $e = \mathsf{e}(n_j^{(i)})$ and define the sets

$$L_j^{(i)} = \Big\{x \in A_j^{(i)} : x\Big(\mathsf{J}\Big(n_j^{(i)}\Big)\Big) \in [9, d - e]\Big\}$$
$$R_j^{(i)} = \Big\{x \in A_j^{(i)} : x\Big(\mathsf{J}\Big(n_j^{(i)}\Big)\Big) \in [d + 2, c - e]\Big\}$$

if $n_j^{(i)} > 0$ and

$$L_j^{(i)} = \Big\{x \in A_j^{(i)} : x\Big(\mathsf{J}\Big(n_j^{(i)}\Big)\Big) \in [10 - e, d - 1]\Big\}$$
$$R_j^{(i)} = \Big\{x \in A_j^{(i)} : x\Big(\mathsf{J}\Big(n_j^{(i)}\Big)\Big) \in [d + 2 - e, c - 2]\Big\}$$

if $n_j^{(i)} < 0$. In this case form the pairs $(L_j^{(i)}, \mathsf{L}(n_j^{(i)}))$ and $(R_j^{(i)}, \mathsf{R}(n_j^{(i)}))$.

- For neither bad nor reducible $h \in G$ its paired index $j \in G$ is necessarily reducible by (E5). Again write $e = \mathsf{e}(n_j^{(i)})$ and in this case define

$$L_h^{(i)} = \left\{ x \in A_h^{(i)} \; : \; x\left(\mathsf{J}\left(n_j^{(i)}\right)\right) \in [9, d - e] \right\}$$
$$R_h^{(i)} = \left\{ x \in A_h^{(i)} \; : \; x\left(\mathsf{J}\left(n_j^{(i)}\right)\right) \in [d + 2, c - e] \right\}$$

if $n_j^{(i)} > 0$ and

$$L_h^{(i)} = \left\{ x \in A_h^{(i)} \; : \; x\left(\mathsf{J}\left(n_j^{(i)}\right)\right) \in [10 - e, d - 1] \right\}$$
$$R_h^{(i)} = \left\{ x \in A_h^{(i)} \; : \; x\left(\mathsf{J}\left(n_j^{(i)}\right)\right) \in [d + 2 - e, c - 2] \right\}$$

if $n_j^{(i)} < 0$. In this case form the pairs $(L_h^{(i)}, n_h^{(i)})$ and $(R_h^{(i)}, n_h^{(i)})$.

Take $k(i+1) = 2|G|$. Let $(A_1^{(i+1)}, n_1^{(i+1)}), \ldots, (A_{k(i+1)}^{(i+1)}, n_{k(i+1)}^{(i+1)})$ be an enumeration of the $k(i+1)$ pairs that result from the above two cases. For each $1 \leqslant j \leqslant k(i+1)$ let $1 \leqslant \alpha(j) \leqslant k(i)$ be the index for which $n_{\alpha(j)}^{(i)}$ determines $n_j^{(i+1)}$ as the above pairs are formed. For each $g \in G$ the sets $F(g)$, $L_g^{(i)}$, $R_g^{(i)}$ partition $A_g^{(i)}$. It is immediate from the recursion and the definitions that (E1) and (E2) are satisfied. We have already verified (E4).

We next verify (E3). Fix $1 \leqslant j \leqslant k(i+1)$. If its ancestor $1 \leqslant \alpha(j) \leqslant k(i)$ is such that $n_j^{(i+1)} = n_j^{(i)}$ then there is nothing to check. Suppose instead that its ancestor is a value $1 \leqslant \alpha(j) \leqslant k(i)$ for which $n_j^{(i+1)} = \mathsf{L}(n_{\alpha(j)}^{(i)})$ or $n_j^{(i+1)} = \mathsf{R}(n_{\alpha(j)}^{(i)})$. Fix $x \in A_j^{(i+1)}$ with $j \neq 0$. Write $n = n_{\alpha(j)}^{(i)}$ and let $r \in \{\mathsf{L}(n), \mathsf{R}(n)\}$ be the reduction we are considering. By either Corollary 5.14 or Corollary 5.16 we have that $(T^n x)(\ell) = (T^r x)(\ell)$ for $\ell \leqslant \mathsf{J}(n) - 1$, so long as $x \notin \mathsf{V}(\mathsf{J}(n), n)$. Invoking (E3) gives $(T^n x)(\ell) = (T^m x)(\ell)$ for $\ell \leqslant \lambda^{i-1} \mathsf{J}(m) - 1$ so long as (5.11) holds. Combining the two verifies (E3).

It remains to define how $1, \ldots, k(i+1)$ are paired and verify (E5), (E6), (E7). There are two cases to consider:

- When $j, h \in G$ are paired and both are reducible, pair the descendants of $j$ and pair the descendants of $h$.
- When $j, h \in G$ are paired and neither is bad and only $j$ is reducible then necessarily $h$ is neither bad nor reducible. Thus $\mathsf{J}(n_h^{(i)}) < \lambda^i(\mathsf{J}(m))$ and since $h$ is not bad, $J(n_h^{(i)}) \leqslant \lambda^{i+1}(\mathsf{J}(m))$. There are two sub-cases to consider:
  (1) If $\mathsf{J}(n_h^{(i)}) = \lambda^{i+1}(\mathsf{J}(m))$ then pair the descendants of $j$ and pair the descendants of $h$.
  (2) If $\mathsf{J}(n_h^{(i)}) < \lambda^{i+1}(\mathsf{J}(m))$ we form the pairing so that $L_j^{(i)}$ is with $L_h^{(i)}$ and $R_j^{(i)}$ is with $R_h^{(i)}$.

First we verify (E7). When the descendants $h'$, $h''$ of some $1 \leqslant h \leqslant k(i)$ are paired we have

$$\left\{ A_{h'}^{(i+1)}, A_{h''}^{(i+1)} \right\} = \left\{ L_h^{(i)}, R_h^{(i)} \right\}$$

and therefore the sets $A_{h'}^{(i+1)}$, $A_{h''}^{(i+1)}$ disagree only in the $\mathsf{J}(n_{j'}^{(i)})$ coordinate, where $j' = h$ when $h$ is reducible and $j'$ is the pair of $h$ when $h$ is not reducible. In either

case $\mathsf{J}(n_{j'}^{(i)})$ is unbounded. It remains to verify that the pairing satisfies (E7) when we are in the second subcase. In that case the additional restriction defining $L_j^{(i)}$ and $L_h^{(i)}$ are the same because both are determined by $n_j^{(i)}$ so $L_j^{(i)}$ and $L_h^{(i)}$ will only disagree in the coordinate where $A_j^{(i)}$ and $A_h^{(i)}$ disagreed. The same is true for $R_j^{(i)}$ and $R_h^{(i)}$.

To conclude we check (E5) and (E6) simultaneously. Fix again $1 \leqslant j, h \leqslant k(i)$ paired. The construction is such that each descendent is paired with exactly one other descendant. Now we check at least one descendant in each pair has its associated time at scale $\lambda^{i+1}(m)$. If both $n_j^{(i)}$ and $n_h^{(i)}$ are reducible we can apply Lemma 2.13 to both separately. Note that in this case (2.9) gives (E6). This is also the justification of Subcase (1). We are left with Subcase (2) where $n_j^{(i)}$ is reducible and $n_h^{(i)}$ is neither bad nor reducible. We are done if $\mathsf{J}(n_h^{(i)}) = \lambda^{i+1}(\mathsf{J}(m))$, so we assume $\mathsf{J}(n_h^{(i)}) \leqslant \lambda^{i+2}(\mathsf{J}(m))$. By (E6)

$$ n_j^{(i)} = k\, \mathsf{q}\Big(\lambda^{i-1}\mathsf{J}(m)\Big) + n_h^{(i)} \quad \text{with} \quad 0 < |k| \leqslant 2d(\lambda^i) + 3 $$

and (applying Lemma 5.19 with $b = n_h^{(i)}$ and using that $n_j^{(i)}$ is not bad) we have that both $\mathsf{L}(n_j^{(i)})$ and $\mathsf{R}(n_j^{(i)})$ have the form

$$ k_L'\, \mathsf{q}\Big(\lambda^i \mathsf{J}(m)\Big) + n_h^{(i)} \quad \text{and} \quad k_R'\, \mathsf{q}\Big(\lambda^i \mathsf{J}(m)\Big) + n_h^{(i)} $$

with $0 < |k_L'|, |k_R'| \leqslant c(\lambda^i(\mathsf{J}(m))) + 1$. So both pairs created from $n_h^{(i)}$ and $n_j^{(i)}$ satisfy (E5), with the terms $\mathsf{L}(n_j^{(i)})$ and $\mathsf{R}(n_j^{(i)})$ satisfying

$$ \mathsf{J}\Big(\mathsf{L}\big(n_j^{(i)}\big)\Big) = \lambda^{i+1}(\mathsf{J}(m)) = \mathsf{J}\Big(\mathsf{R}\big(n_j^{(i)}\big)\Big) $$

as desired. Similarly, we have both pairs satisfy (E6). □

PROPOSITION 5.21. — *Fix $\tau = a(\beta) \geqslant a(10)$ a target unbounded scale and $m \in \mathbb{N}$ with $\mathsf{J}(m) = a(b)$ unbounded and $b > \beta$. Let $I \in \mathbb{N}$ and the tuples $(A_0^{(i)}, A_1^{(i)}, n_1^{(i)}, \ldots, A_{k(i)}^{(i)}, n_{k(i)}^{(i)})$ for $1 \leqslant i \leqslant I = b - \beta$ result from an application of Proposition 5.20 to $m$. In particular, $\mathsf{J}(n_j^{(i)}) \leqslant \tau$ and for at least half of the $j$ we have $\mathsf{J}(n_j^{(I)}) = \tau$. If $\mu(A_0^{(I)}) < \frac{1}{10^{50}}$ and tuples $(\overline{A}_0^{(i)}, \overline{A}_1^{(i)}, \overline{n}_1^{(i)}, \ldots, \overline{A}_{\overline{k}(i)}^{(i)}, \overline{n}_{\overline{k}(i)}^{(i)})$ for $1 \leqslant i \leqslant K = b - \beta$ result from applying Proposition 5.20 to $6d(\beta)m$ then $\mu(\overline{A}_0^{(K)}) > \frac{1}{10^{50}}$.*

This follows from repeated applications of the following result.

LEMMA 5.22. — *Fix $m \in \mathbb{Z}$ with $\mathsf{J}(m) = a(b)$ an unbounded scale.*

*If $\mathsf{R}(m)$ is at an unbounded scale and $k \in \mathbb{Z}$ satisfies $|k| \leqslant 6d(b)$ and $km$ is at an unbounded scale then one of the following is true:*

- *$\mathsf{R}(km) = k\mathsf{R}(m)$.*
- *$\mathsf{R}(km)$ is at a bounded scale.*
- *At least one of $\mathsf{R}(m)$, $\mathsf{R}(km)$ is both at scale $\mathsf{J}(m)$ and less than $\frac{3}{4}\mathsf{p}(\mathsf{J}(m))$.*

*Moreover, if $\mathsf{R}(km)$ is at scale $\lambda(\mathsf{J}(m))$ then $|k| < 6d(a^{-1}(\mathsf{J}(\mathsf{R}(m))))$.*

If $\mathsf{L}(m)$ is at an unbounded scale and $k \in \mathbb{Z}$ satisfies $|k| \leqslant 6d(b)$ and $km$ is at an unbounded scale then one of the following is true:

- $\mathsf{L}(km) = k\mathsf{L}(m)$,
- $\mathsf{L}(km)$ is at a bounded scale.

Moreover, if $\mathsf{L}(km)$ is at scale $\lambda(\mathsf{J}(m))$ then $|k| < 6d(a^{-1}(\mathsf{J}(\mathsf{L}(m))))$.

*Proof of Proposition 5.21 assuming Lemma 5.22.* — Consider any sequence $m, n^{(1)}$, $n^{(2)}, \ldots, n^{(I)}$ of reductions of $m$. The corresponding sequence for $km$ is $km, kn^{(1)}$, $kn^{(2)}, \ldots$ until one of $kn^{(i)}$ is in a bad pair. This follows by Lemma 5.22 and the definition of bad pairs in the proof of Proposition 5.20. Now if $kn^{(i)}$ is bad the corresponding points are put in $A_0$ and we have proportional friends. On the other hand $6d(\beta)n^{(I)} > \mathsf{p}(\tau)$ so $A^{(I-1)}$ is put in $A_0^{(I)}$ if it is not put in earlier. $\qquad \square$

*Proof of Lemma 5.22.* — We present the proof in the case of $\mathsf{R}$. The case of $\mathsf{L}$ is similar. First, if $km$ is at an unbounded scale, then $\mathsf{J}(km) = \mathsf{J}(m)$ because the bound on $k$ gives $\mathsf{J}(km) \leqslant \mathsf{J}(m) + 1$. If $\mathsf{R}(km) \neq k\mathsf{R}(m)$ then

$$
\left| \mathsf{R}(km) - k\mathsf{R}(m) \right|
$$
$$
= \left| km - \mathsf{e}(km)\Big( \mathsf{p}(\mathsf{J}(m)) - \mathsf{q}(\mathsf{J}(m)) \Big) - k\Big( m - \mathsf{e}(m)\big( \mathsf{p}(\mathsf{J}(m)) - \mathsf{q}(\mathsf{J}(m)) \big) \Big) \right|
$$
$$
\geqslant \mathsf{p}(\mathsf{J}(m)) - \mathsf{q}(\mathsf{J}(m))
$$

and by Lemma 2.14 they are both at most $\frac{3}{4}\mathsf{p}(\mathsf{J}(m))$. Lastly, if $\mathsf{J}(m) = a(\gamma))$ and $6d(\gamma + 1) \geqslant |k| \geqslant 6d(\gamma)$ then $\mathsf{J}(km)$ is strictly between $a(\gamma)$ and $a(\gamma + 1)$. Thus it is at a bounded scale. $\qquad \square$

## 5.6. Proof of Theorem 1.7

We are ready to conclude that $(X, \mu, T)$ is mild mixing.

*Proof of Theorem 1.7.* — Fix a non-constant function $f$ in $\mathsf{L}^2(X, \mu)$ and a sequence $\boldsymbol{r}$ in $\mathbb{Z}$ with $|r(n)| \to \infty$. Set $\rho = 10^{-50} \cdot 10^{-9}$. Let $\delta$ and $\tau'$ be as in Corollary 5.7 for our choices of $f$ and $\rho$. Fix $\tau = a(\beta) > \tau'$ an unbounded scale. It suffices to show that at least one of $\boldsymbol{r}$ or $6d(\beta)\boldsymbol{r}$ is not a rigidity sequence for $f$. Indeed if $\boldsymbol{r}$ is a rigidity sequence for $f$ then so is $k\boldsymbol{r}$ for any fixed $k \in \mathbb{Z} \setminus \{0\}$. By Theorem 5.10 and by passing to a subsequence if necessary we may first assume for all $n \in \mathbb{N}$ that $\mathsf{J}(r(n))$ and $\mathsf{J}(6d(\beta)r(n))$ are unbounded scales and that $\mathsf{J}(r(n)) \geqslant \tau$.

Fix $I(n)$ and $K(n)$ maximal with $\lambda^{I(n)}(r(n)) \geqslant \tau$ and $\lambda^{K(n)}(6d(\beta)r(n)) \geqslant \tau$ respectively. Apply Proposition 5.20 to $r(n)$ and $6d(\beta)r(n)$. Let $\Gamma_n \subset A_0^{(I(n))}$ and $\overline{\Gamma}_n \subset \overline{A}_0^{(K(n))}$ be the respective sets furnished by (E4). There are measure preserving maps $\phi_n : \Gamma_n \to X$ and $\overline{\phi}_n : \overline{\Gamma}_n \to X$ with the following properties.

- For each $z \in \Gamma_n$ we have that $(z, \phi_n(z))$ are $(r(n), \tau)$ buddies.
- For each $z \in \overline{\Gamma}_n$ we have that $(z, \overline{\phi}_n(z))$ are $(6d(\beta)r(n), \tau)$ buddies.
- $\mu(\Gamma_n) \geqslant 10^{-9}\mu(A_0^{(I(n))})$
- $\mu(\overline{\Gamma}_n) \geqslant 10^{-9}\mu(\overline{A}_0^{(K(n))})$

By Proposition 5.21

$$\max\left\{\mu\left(A_0^{(I(n))}\right), \mu\left(\overline{A}_0^{(K(n))}\right)\right\} \geqslant 10^{-50}$$

and thus Corollary 5.7 gives

$$\max\left\{\int\left|f - f\circ T^{r(n)}\right|\,\mathsf{d}\mu, \int\left|f - f\circ T^{6d(\beta)r(n)}\right|\,\mathsf{d}\mu\right\} > \delta$$

proving Theorem 1.7. $\qquad\square$

## Appendix A. Changes for the proof of Proposition 4.3

In this section we prove the version of [CE21, Corollary 3.3] that is needed to prove Proposition 4.3.

The next lemma and its proof are unchanged from [CE21, Lemma 3.2].

LEMMA A.1. — *Given $c > 0$ and $d \in \mathbb{N}$ there exists $\rho < 1$, $C$ so that if $0 < \delta_i < 1/2$ and $a_i, b_i$ are such that $a_i, b_i > c$ and $1 \geqslant a_i + b_i > 1 - \delta_i$ and also $0 \leqslant \zeta_i^{(\ell)} \leqslant 1$ are sequences of real numbers for each $\ell \in \{1, \ldots, d\}$ satisfying*

(A.1) $$\left|\zeta_i^{(\ell)} - \left(a_i\zeta_{i-1}^{(\ell-1)} + b_i\zeta_{i-1}^{(\ell)}\right)\right| < \delta_{i-1}$$

*then*

$$\left|\zeta_i^{(s)} - \frac{1}{d}\sum_{\ell=1}^{d}\zeta_k^{(\ell)}\right| \leqslant C\sum_{j=k}^{i-1}\left(\delta_i + \frac{\delta_i}{1-\delta_i}\right) + C\rho^{i-k}$$

*for all $k \geqslant 0$, $i > k$ and $s \in \{1, \ldots, d\}$.*

COROLLARY A.2. — *Under the assumptions of Proposition 4.3 there exist $\rho < 1$, $C' > 0$ so that $d_{KR}(\nu_k^{(\ell)}, \frac{1}{d}\sum_{\ell=1}^{d}\nu_b^{(\ell)}) \leqslant C'\sum_{j=b}^{k}\epsilon_j + C'\rho^{k-b}$ whenever $k \geqslant b$ and $\ell \in \{1, \ldots, d\}$.*

*Proof of Corollary A.2.* — First notice that by (J6) we have that

(A.2) $$d_{KR}\left(\nu_{\alpha(j)}|_{A(j)}, \ \nu_{\gamma(j-1)}|_{A(j)}\right) < \epsilon(j)$$
$$\text{and} \quad d_{KR}\left(\nu_{(\alpha(j)}|_{B(j)}, \ \nu_{\alpha(j-1)}|_{B(j)}\right) < \epsilon(j).$$

We now claim that there exists a constant $D$ so that for all $j$,

(A.3) $$d_{KR}\left(\frac{1}{\mu(A(j))}\nu_{\gamma(j-1)}\bigg|_{A(j)}, \ \nu_{\gamma(j-1)}\right) < D\left(\epsilon(j-1) + \epsilon(j)\right)$$

By definition of $A(j)$ we have that

$$\int f\,d\nu(\gamma(j-1))|_{A(j)} = \int_{\bigcup_{i=0}^{r(j)-1} T^i J(j)\setminus U(j)} f\left(x, T^{\gamma(j-1)}x\right).$$

Now,

$$
\text{(A.4)} \quad \left| \int_{\bigcup_{i=0}^{r(j)-1} T^i J(j)} f \circ (T \times T)^i \left( x, T^{\gamma(j-1)} x \right) \right.
$$
$$
\left. - \sum_{j=0}^{(1-\epsilon(j-1))r(j)-1} \frac{1}{\epsilon(j-1)r(j)} \int_{\bigcup_{i=0}^{\epsilon(j-1)r(j)} T^{i+j} J(j)} f \circ (T \times T)^i \left( x, T^{\gamma(j-1)} x \right) \right|
$$
$$
\leqslant 2(\epsilon(j-1)r(j)) \mu(J(j)) \| f \|_{\sup}.
$$

To see (A.4), the integral has the form

$$
\int_{\bigcup_{i=0}^{r(j)} T^i J(j)} f \left( x, T^{\gamma(j-1)} x \right) - c(x) f \left( x, T^{\gamma(j-1)} x \right)
$$

where $0 \leqslant c(x) \leqslant 1$ for all $x \in \bigcup_{i=0}^{r(j)-1} T^i J(j)$ and $c(x) = 1$ for all

$$
x \in \bigcup_{i=e(j-1)r(j)}^{(1-e(j-1))r(j)-1} T^i J(j),
$$

giving (A.4).

By (J7),

$$
\text{(A.5)} \quad \left| \frac{1}{(1-\epsilon(j-1))r(j)} \sum_{j=0}^{(1-\epsilon(j-1))r(j)} \frac{1}{\mu(J(j))\epsilon(j-1)r(j)} \right.
$$
$$
\left. \int_{\bigcup_{i=0}^{\epsilon(j-1)r(j)-1} T^i T^k J(j)} f \circ (T \times T)^i \left( x, T^{\gamma(j-1)} x \right) d\mu - \int_{X \times X} f \, d\nu_{\gamma(j-1)} \right|
$$
$$
\leqslant 3 \| f \|_{\text{Lip}} \epsilon(j-1) + 4 \| f \|_{\sup} \epsilon(j).
$$

To see this, we split the left hand side into two parts: the $x \in \bigcup_{k=0}^{(1-\epsilon(j-1)r(j)} T^k J_j$ for which (4.9) holds and the complement that we estimate trivially.

Because $A(j) = \bigcup_{j=0}^{r(j)-1} T^j J(j) \setminus U(j)$ we now use (J3) (i.e. the size estimate on $U_j$),

$$
\text{(A.6)} \quad \left| \frac{1}{(1-\epsilon(j-1))r(j)} \sum_{j=0}^{(1-\epsilon(j-1))r(j)} \frac{1}{\lambda(J(j))\epsilon(j-1)r(j)} \right.
$$
$$
\left. \int_{\bigcup_{i=0}^{\epsilon(j-1)r(j)-1} T^i T^k J(j)} f \circ (T \times T)^i \left( x, T^{\gamma(j-1)} x \right) \chi_{U(j)}(x) d\mu \right|
$$
$$
\leqslant \| f \|_{\sup} \mu(U(j)) \leqslant \| f \|_{\sup} \epsilon(j).
$$

Combining these estimates we have (A.3).

Similarly, by partitioning $B_j$ into $D_{\frac{r_j}{2}}, \dots$ where

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

we get that there exists $E > 0$ so that

$$
\text{(A.7)} \quad d_{KR} \left( \frac{1}{\mu(B(j))} \nu_{(\alpha(j-1))} \bigg|_{B(j)}, \ \nu_{(\alpha(j-1))} \right) < E(\epsilon(j-1) + \epsilon(j)).
$$

We have analogous versions of these estimates swapping the roles of $\alpha(j)$ for $\gamma(j)$, $\alpha(j-1)$ for $\gamma(j-1)$ and $\gamma(j-1)$ for $\alpha(j-1)$.

So for any 1-Lipschitz function, $f$, with $\|f\|_{\sup} \leqslant 1$, we claim that we may apply Lemma A.1 to $\zeta_i^{(1)} = \int f d\nu_{\alpha(i)}$, $\zeta_i^{(2)} = \int f d\nu_{\gamma(i)}$ with $c = c$, $\delta_{j-1} = (E + D)(\epsilon(j-1) + \epsilon(j))$, $a_j = \mu(A(j))$ and $b_j = \mu(B(j))$. To verify (A.1), note that

$$\left| \int f d\nu_{\alpha(j)} - \int_{A(j)} f d\nu_{\alpha(j)} - \int_{B(j)} f d\nu_{\alpha(j)} \right| \leqslant \|f\|_{\sup} \lambda(U(j)) < \epsilon(j)$$

and similarly for $\gamma(j)$. Thus, by (A.2)

$$\left| \int f d\nu_{\alpha(j)} - \int_{A(j)} f \, d\nu_{\gamma(j-1)} - \int_{B(j)} f d\nu_{\alpha(j-1)} \right| \leqslant 2\epsilon(j).$$

Then, by (A.3) and (A.7),

$$\left| \int f d\nu_{\alpha(j)} - \left( \mu(A(j)) \int f d\nu_{\gamma(j-1)} + \mu(B(j)) \int f d\nu_{\alpha(j)} \right) \right|$$
$$\leqslant (E + D)(\epsilon(j-1) + \epsilon(j)).$$

This and the analogous result for $\nu_{\gamma(j)}$, completes the verification of (A.1), and, in view of Lemma A.1, the proof of Corollary A.2. $\qquad\square$

# Appendix B. Notation

The following symbols - used throughout the paper - are recalled here for reference.

### Odometer notation

- $\Omega$ is the odometer space,
- $n \mapsto a(n)$ enumerates those sparse positions where the odometer has many digits,
- $n \mapsto 2d(n) + 2$ is the number of digits in the odometer at position $a(n)$,
- $n \mapsto c(n) + 1$ is the number of digits in the $n^{\text{th}}$ position of the odometer,
- $n \mapsto t(n)$ is the number of cylinders of length $n$ in $\Omega$,
- $\nu$ is the normalized Haar measure on $\Omega$.

### System notation

- $n \mapsto W_n$ describes the cylinders of length $n$ removed from the odometer.
- $Y_n = \Omega \setminus (W_1 \cup \cdots \cup W_n)$,
- $X$ is the space $\Omega \setminus \bigcup_{n \in \mathbb{N}} W_n$,
- $T$ is the transformation we consider on $X$,
- $\mu$ is the normalized restriction of $\nu$ to $X$,
- $\mathsf{V}(b, r) = \bigcup_{a(m) > b} \{ x \in X : (S|Y_b)^i(x) \in W_{a(m)} \text{ for some } -|r| \leqslant i \leqslant |r| \}$.

## Specific times

- $\mathsf{p}(n)$ is the minimal natural number for which $(S|Y_n)^{\mathsf{p}(n)}[0^{n-1}0] = [0^{n-1}1]$,
- $\mathsf{P}(n) = \mathsf{p}(a(n))$ is the height of the left tower at scale $a(n)$,
- $\mathsf{Q}(n) = 2\mathsf{P}(n-1) - \mathsf{Q}(n-1)$ is the number of levels removed from $\Omega$ at scale $a(n)$,
- $\mathsf{q}(n) = \mathsf{Q}(a^{-1}(n))$ when $n$ unbounded and zero otherwise.

## Expansion notation

- $\mathsf{e}(r)$ is the leading coefficient in the tower expansion of $r \in \mathbb{Z}$,
- $\mathsf{J}(r)$ is the leading scale in the tower expansion of $r \in \mathbb{Z}$.

## Reduction notation

- $\mathsf{L}(r) = r - \mathsf{e}(r)\,\mathsf{p}(\mathsf{J}(r))$ whenever $\mathsf{J}(r)$ is unbounded,
- $\mathsf{R}(r) = r - \mathsf{e}(r)\,\mathsf{p}(\mathsf{J}(r)) + \mathsf{e}(r)\,\mathsf{q}(\mathsf{J}(r))$ whenever $\mathsf{J}(r)$ is unbounded,
- $\lambda(n)$ is the largest unbounded scale strictly smaller than $n$.

## BIBLIOGRAPHY

[CE21] Jon Chaika and Alex Eskin, *Self-joinings for 3-IETs*, J. Eur. Math. Soc. **23** (2021), no. 8, 2707–2731. ↑100, 101, 114, 115, 117, 121, 141

[CK21] Jon Chaika and Bryna Kra, *A prime system with many self-joinings*, J. Mod. Dyn. **17** (2021), 213–265. ↑122

[Fer97] Sébastien Ferenczi, *Systems of finite rank*, Colloq. Math. **73** (1997), no. 1, 35–65. ↑98, 101

[FHZ05] Sébastien Ferenczi, Charles Holton, and Luca Q. Zamboni, *Joinings of three-interval exchange transformations*, Ergodic Theory Dyn. Syst. **25** (2005), no. 2, 483–502. ↑100, 101

[Fur67] Harry Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Syst. Theory **1** (1967), 1–49. ↑98

[Kin86] Jonathan L. King, *The commutant is the weak closure of the powers, for rank-1 transformations*, Ergodic Theory Dyn. Syst. **6** (1986), no. 3, 363–384. ↑98, 101

[Kin88] ———, *Joining-rank and the structure of finite rank mixing transformations*, J. Anal. Math. **51** (1988), 182–227. ↑98

[Rud79] Daniel J. Rudolph, *An example of a measure preserving map with minimal self-joinings, and applications*, J. Anal. Math. **35** (1979), 97–122. ↑98

[Rue23] Thierry de la Rue, *Joinings in ergodic theory*, Ergodic theory, Encyclopedia of complexity and systems science, Springer, 2023, pp. 149–168. ↑98

[Ryz12] Valeriĭ V. Ryzhikov, *On local rank, joinings and asymptotic properties of measure-preserving actions*, 2012, https://arxiv.org/abs/1205.7081. ↑99

[Ryz13] ———, *Bounded ergodic constructions, disjointness, and weak limits of powers*, Trans. Mosc. Math. Soc. **2013** (2013), 165–171. ↑98

[Ryz20] ———, *Measure-preserving rank one transformations*, Trans. Mosc. Math. Soc. **81** (2020), 229–259. ↑98

Jon CHAIKA
Office 217 Fuld Hall
Einstein Dr, Princeton
NJ 08540 (USA)

jonchaika@gmail.com

Donald ROBERTSON
Department of Mathematics
University of Manchester
Oxford Road, Manchester
M13 9PL (United Kingdom)

donald.robertson@manchester.ac.uk