# A PRIME SYSTEM WITH MANY SELF-JOININGS

JON CHAIKA AND BRYNA KRA

Abstract. We construct a rigid, rank 1, prime transformation that is not quasi-simple and whose self-joinings form a Poulsen simplex. This seems to be the first example of a prime system whose self-joinings form a Poulsen simplex.

## 1. Introduction

A natural question is to find indecomposable structures, and we study this question in the setting of measurable dynamics. More precisely, we consider a measure preserving dynamical system $(Z, \mathcal{M}, \mu, T)$, where $Z$ is a set endowed with a $\sigma$-algebra $\mathcal{M}$, $\mu$ is a probability measure on the measure space $(Z, \mathcal{M})$, and $T\colon Z \to Z$ is a measurable transformation that preserves the measure $\mu$. Throughout this article, we assume that $(Z, \mathcal{M}, \mu)$ is a (non-atomic) Lebesgue space. A *factor* of a measure preserving system $(Z, \mathcal{M}, \mu, T)$ is a measure preserving system $(Z', \mathcal{M}', \mu', T')$ and a measurable map $\pi\colon Z \to Z'$ such that $\mu \circ \pi^{-1} = \mu'$ and $T' \circ \pi(x) = \pi \circ T(x)$ for $\mu$-almost all $x \in Z$. In this setting, the indecomposable structures are the *prime* transformations, which are transformations with no nontrivial (measurable) factors. That is, any factor map on $(Z, \mathcal{M}, \mu, T)$ is either an isomorphism or a map to a one point system. Historically, showing systems are prime has largely been accomplished by understanding the self-joinings of the system, that is, the $T \times T$ invariant measures on $Z \times Z$ with marginals $\mu$ on each of the coordinates. Our main result is that there exists a prime transformation with many self-joinings (the self-joinings form a Poulsen simplex) and the self-joinings can be large (there is a self-joining that does not arise as a distal extension of the system):

**Theorem 1.1.** *There exists a prime system $(Y, \mathcal{B}, \nu, T)$ that is rank 1, rigid, and has an ergodic self-joining $\eta$, which is not the product measure, such that $(Y \times Y, \mathcal{B} \times \mathcal{B}, \eta, T \times T)$ is not a distal extension of $(Y, \mathcal{B}, \nu, T)$. Moreover, the set of self-joinings of $Y$ is a Poulsen simplex.*

To highlight the novelty of our construction, we note that being not quasi-simple (or being not quasi-distal) is a residual property in the space of measure preserving transformations (endowed with the weak topology). This answers a question posed by Danilenko [5, Section 7, Question (iii)] who asked if the set of quasi-simple transformations and the set of *distal-simple* transformations are both meager. It is also a strengthening of a result of Ageev [2] who showed that being simple is meager.

The first author was partially supported by NSF grants DMS-135500 and DMS-452762, the Sloan foundation, Poincaré chair, and Warnock chair and the second author by NSF grant DMS-1800544. The authors thank M. Lemanczyk for helpful comments and questions.

1.1. **Context of the results.** The first systematic family of prime systems was introduced by Rudolph [28], based on Ornstein's counterexample machinery, and has been studied extensively since; for example, see [7, 9, 22, 15, 18, 29]. A system $(Y, \nu, T)$ has 2-*fold minimal self-joinings* if all of the ergodic self-joinings are either $\nu \times \nu$ or are concentrated on the graph $\{(x, T^j x)\}$ for some integer $j \geq 0$. Defining the natural generalization for $k$-fold minimal self-joinings for all $k \geq 2$, Rudolph showed that any system having minimal self-joinings is prime.

However, having minimal self-joinings is quite special, and so there was interest in more general criteria for obtaining prime systems. In this direction, Veech showed that a *2-simple* system is prime if it has no compact subgroups in its centralizer. Recall that a system is 2-simple (which Veech called *property* $S$) if the only ergodic self-joinings arise from the product measure and measures carried on graphs of transformations in the centralizer of the system. Simple systems have since been studied in a variety of contexts (see for example [31, 8, 14, 2, 5, 7, 6, 13, 30]). Veech's criterion gave rise to the first example of a rigid prime system, with the construction by del Junco and Rudolph [8] of a specific rigid, simple system that had no non-trivial compact subgroups in its centralizer. Glasner and Weiss [14] constructed an example of a prime system that is not simple, by taking a simple system and considering the factor corresponding to a non-normal maximal compact subgroup, again using Veech's criteria to show that the factor is prime since it arises from a maximal compact subgroup. In this example, as the subgroup is not normal, the factor itself is not simple, but the self-joinings of the factor of a simple system are always isometric extensions of the factor.

There are a few other known examples of prime systems. For example, King [21, Section 2] showed that the (proper) factors of rank 1 systems are rigid and so it follows that mildly mixing rank 1 systems are prime. Continuing in this vein, Thouvenot asked if mildly mixing rank 1 transformations have minimal self-joinings, and this difficult question remains open. Parreau and Roy [27] gave a construction of prime systems for some Poisson suspensions of (infinite measure preserving) prime systems, and it follows from results in [24] that the constructed systems are quasi-distal. In the same article, Parreau and Roy write "it is yet unknown whether prime rank one maps are always factors of simple systems." Our construction resolves this by producing a prime rank 1 system that is not the factor of a simple system (it is not quasi-simple).

This short list of examples basically includes all known prime systems, and one motivation for this work is to give a new construction of prime systems not relying on a paucity of joinings (as in the minimal self-joinings, simple, or factor of simple systems) or soft restrictions on the prime factors (as in the mildly mixing rank 1 or Poisson suspension of prime infinite measure preserving systems with additional properties).

Turning to the second conclusion of Theorem 1.1, we note that it is well-known that a residual set of measure preserving systems is rank 1 and rigid. King [23] showed that for a typical measure preserving transformation, its self-joinings form a Poulsen simplex (recall that a *Poulsen simplex* is a simplex such that the extreme points are dense). Putting this in context, Lindenstrauss, Olsen, and Sternfeld [25] proved that a Poulsen simplex is unique up to affine homeomorphism. Ageev showed that the typical transformation is not prime [1] and is not simple [2].

1.2. **A brief outline of the paper and a conjecture.** In Section 2, we introduce general concepts from ergodic theory. In Section 3, we define our system $(Y, \nu, T)$, as the first return map of an odometer to a compact set, and then we set up the basic notation used throughout and prove first results on the mixing properties of the system. In Section 4, we show that our system is not quasi-distal and that its self joinings form a Poulsen simplex. In Section 6, we show that our system is prime, building heavily on ingredients developed in Section 5. As our arguments are technical and require some additional development, we defer conceptual descriptions of the proofs to Sections 4 and 5, after the preliminary tools have already been defined.

Our methods for building self-joinings and building self-joinings that can not be distal extensions of the base system are fairly soft and general (if technical and involved).

Our proof that the transformation is prime is more combinatorial, making heavy use of the specific construction. This should not be surprising, because being prime is a meager property [1] in the space of measure preserving transformations (with the weak topology). Nevertheless, an ideology of this work is that it may still be a fairly common property. In particular, we conjecture that in some families of measure preserving transformations almost every system is prime. To be specific:

**Conjecture 1.** Almost every 3-IET is prime.

Although this may hold more generally for a $k$-IET, such a conjecture is out of reach at this point, but for a 3-IET, some of these tools are already developed with the methods of [4].

A second conjecture, closer to the work of this paper, is stated in Section 3, after we have developed some further background.

## 2. Definitions and notation

2.1. **Systems and joinings.** By a *measure preserving system* $(X, \mathcal{B}, \mu, T)$, we mean that $\mathcal{B}$ is the Borel $\sigma$-algebra for some compact metric topology on $X$, $(X, \mathcal{B}, \mu)$ is a probability space, and $T\colon X \to X$ is a measurable, measure preserving map. Throughout the paper, we generally omit the associated $\sigma$-algebra from the notation, assuming that any measure preserving system is endowed with the Borel $\sigma$-algebra. We say that the measure preserving system $(Y, \nu, S)$ is a *factor* of $(X, \mu, T)$ if there exists a measurable map $\pi\colon X \to Y$ such that $\pi \circ T = S \circ \pi$ and $\mu \circ \pi^{-1} = \nu$.

A *joining* of the ergodic measure preserving systems $(X_i, \mu_i, T_i)$ for $i = 1, 2$ is a $(T_1 \times T_2)$-invariant measure $\alpha$ on $X_1 \times X_2$ such that $\alpha$ projects to $\mu_1$ on the first coordinate and to $\mu_2$ on the second coordinate. A *self-joining* of a system is a joining of two copies of the same system. If $(X, \mu, T)$ is a measure preserving system, $J(n)$ denotes the *off diagonal joining* on $\{(x, T^n x)\}$, meaning that $J(n)$ is the measure on $X \times X$ such that for all $f \in C(X \times X)$

$$\int f(x, y)\, dJ(n) = \int f(x, T^n x)\, d\mu.$$

If $(X, \mu, T)$ is an ergodic measure preserving system, we say that the bounded linear operator $P\colon L^2(\mu) \to L^2(\mu)$ is a *Markov operator* if it satisfies:

(i) For all $f \in L^2(\mu)$ with $f \geq 0$, we have $Pf \geq 0$ and $P^* f \geq 0$.

(ii) $P\mathbf{1}_X = \mathbf{1}_X$ and $P^*\mathbf{1}_X = \mathbf{1}_X$, where $\mathbf{1}_A$ denotes the indicator function of the set $A$.

(iii) $PU_T = U_T P$, where $U_T\colon L^2(\mu) \to L^2(\mu)$ by $U_T f = f \circ T$.

Markov operators can be defined more generally for an operator mapping one measure preserving system to another, but our interest is when the operator arises as an integral of fibers of a factor and so we can take the map from a system to itself; see, for example, Glasner [12] for more on such operators. More precisely, if $(X, \mu, T)$ has a factor $(Y, \nu, S)$ with factor map $\pi$, then by integrating over the fibers of the factor map, we obtain a bounded linear operator $P\colon L^2(\mu) \to L^2(\mu)$, satisfying Properties (i)–(iii) and we call this the *Markov operator defined by* $\pi$. That is, by disintegration of measures there exist measures $\mu_y$ on $X$ such that $\mu = \int_Y \mu_y d\nu$ and $P(f)(x) = \int f d\mu_{\pi(x)}$. Note that joinings also give rise to Markov operators. However, these do not formally enter the arguments and so we do not discuss these Markov operators.

2.2. **Rigid rank one by cylinders.** As above, we assume that each system is endowed with its Borel $\sigma$-algebra, but we omit it from the notation.

**Definition 2.1.** *An invertible ergodic system $(Z, \lambda, R)$, where $Z \subset [0,1]$ and $\lambda$ denotes normalized (probability) Lebesgue measure restricted to $Z$, is* rigid rank one by cylinders *if there exist a sequence of intervals $(I_i)_{i\in\mathbb{N}}$, which we call* cylinders, *and a sequence of positive integers $(n_i)_{i\in\mathbb{N}}$ such that:*

*(i) For $0 \le i < n_k$, the iterate $R^i I_k$ is a cylinder having the same measure as $I_k$.*

*(ii) The cylinders $R^i I_k$ and $R^j I_k$ are pairwise disjoint for all $k$ and for $0 \le i < j < n_k$.*

*(iii) The measure $\lambda(\bigcup_{i=0}^{n_k-1} R^i I_k)$ tends to $1$ as $k \to \infty$.*

*(iv) The ratio $\frac{\lambda(R^{n_k} I_k \Delta I_k)}{\lambda(I_k)}$ tends to $0$ as $k \to \infty$.*

Note that rank 1 systems can be rigid without being rigid rank 1, but the rigidity of a rigid rank one system is not directly tied to the towers of the system. Also, note that cylinders in this setting are intervals in $[0,1]$, but we refer to them as cylinders in analogy with the symbolic setting. By a *symbolic system* $(X, T)$, we mean an infinite sequence space $X \subset \prod_{i=1}^{\infty} \mathcal{A}_i$, where each $\mathcal{A}_i$ is a finite alphabet, and $T\colon X \to X$ is a measurable map. We denote elements of the space as $\mathbf{x} = (x_i)_{i\in\mathbb{N}} \in X$, with the convention that a bold face letter $\mathbf{x}$ has its entries denoted as $x_i$. In a symbolic system $X$, a *cylinder set* $[w]$ *determined by a word* $w = w_1 \dots w_n$ is defined to be

$$[w] = \{\mathbf{x} \in X \colon x_i = w_i \text{ for all } 1 \le i \le n\}.$$

We also consider cylinders defined only by some entries $a_{i_1} \in \mathcal{A}_{i_1}, \dots, a_{i_k} \in \mathcal{A}_{i_k}$ defining the cylinder

$$\{\mathbf{x} \in X \colon x_{i_j} = a_{i_j} \text{ for all } 1 \le j \le k\}$$

and we refer to the $i_j$ as *defining indices* of the cylinder. The collection of cylinder sets forms a basis for the topology of $X$. When working with a symbolic system $(X, T)$, fixing initial entries corresponds to an interval in $[0,1]$, meaning that a cylinder set corresponds to an interval.

The first three conditions in the definition of rigid rank one by cylinders imply that $G$ is rank one, but in the general setting of a rank one transformation there is no requirement that the subsets $I_i$ are intervals. The fourth condition gives a sequence of times under which the transformation $R$ is rigid, meaning that along these times the iterates of $R$ approach the identity. Indeed, Condition (iv) implies that $\frac{\lambda(R^{n_k}R^iI_k \cap R^iI_k)}{\lambda(I_k)}$ is close to 1 for all large $k$ and $0 \le i < n_k$, and so using this with Conditions (i) and (iii), we have a rigidity sequence.

2.3. **Distal extensions.** We review the definitions of (measurable) isometric and distal extensions, as introduced by Parry [26]. These extensions were key in Furstenberg's proof [10] of Szemerédi's Theorem (see [11] for further background), and the definition we use comes from Zimmer [33, 34], who showed that a measurably distal system is equivalent to a (possibly transfinite) inverse limit of a tower of isometric extensions.

If $G$ is a compact group, $H \subset G$ is a closed subgroup, and $(X, \mu, T)$ is a Borel probability system, then a measurable map $\phi\colon X \to G$ is called a *cocycle* and the *extension of $G$ by $G/H$ given by the cocycle $\phi$* is defined to be the system $(X \times G/H, \mu \times m_{G/H}, T_\phi)$, where $T_\phi(x, \tilde{g}) = (Tx, \phi(x)\cdot\tilde{g})$ for $x \in X$ and $\tilde{g} \in G/H$ and $m_{G/H}$ is the Haar measure on $G/H$ (we use the convention that cosets in $G/H$ are denoted by $\tilde{\cdot}$). Defining the topology of the group $G$ by a distance $d_G$ that is invariant under right translation, and of course continuous with respect to translation on either side, we have an induced distance $d_{G/H}$ on $G/H$ and we have that the restriction of $T_\phi$ to each fiber of the natural projection map $X \times G/H \to X$ is continuous. The system $(X \times G/H, \mu \times m_{G/H}, T_\phi)$ is an *isometric extension of the system* $(X, \mu, T)$.

If $(X, \mu, T)$ and $(Y, \nu, S)$ are ergodic systems, then $(X, \mu, T)$ is a *distal extension* of $(Y, \nu, S)$ if it has a sequence of factors $X_\eta$ indexed by ordinals $\eta \le \eta_0$ for some countable ordinal $\eta_0$ such that $X_0 = Y$, $X_{\eta_0} = X$, $X_{\eta+1}$ is an isometric extension of $X_\eta$ for each $\eta$, and for each limit ordinal $\zeta \le \eta_0$ the system $X_\zeta$ is an inverse limit of the systems $X_\eta$ with $\eta \le \zeta$.

**Notation.** We use $d$ to denote the metric in various settings, with a subscript indicating the space as needed. Thus $d_G$ denotes the right invariant metric on the group $G$, $d_{G/H}$ denotes the induced distance on $G/H$.

## 3. Construction of the system

3.1. **Definition of the transformation $T$.** We begin by constructing an odometer. Set

$$X = \prod_i \{0, \dots, a_i - 1\}, \tag{1}$$

where

$$a_i = \begin{cases} 8 & \text{if } i \notin \{10^k \colon k \ge 2\} \\ k & \text{if } i = 10^k \text{ for some } k \ge 2. \end{cases}$$

We write elements $\mathbf{x} \in X$ as $\mathbf{x} = (x_i)_{i\in\mathbb{N}}$. Let $S$ denote the odometer on $X$, meaning that $S$ is addition by $(1, 0, 0, \dots)$ with carrying to the right. Thus

$$S(\mathbf{x}) = S(x_1, x_2, \dots, x_k, x_{k+1}, \dots) = (0, 0, \dots, 0, x_k + 1, x_{k+1}, \dots), \tag{2}$$

where $k$ is the least entry such that $x_k < a_k - 1$ and if there is no such $k$, then the odometer turns over and outputs the point $\mathbf{0} = (0, 0, \dots)$.

Set

$$Z_k = \{\mathbf{x} \in X \colon x_k = 7 \text{ and } x_i = a_i - 2 \text{ for all } i < k\} \tag{3}$$

and

$$W_k = \{\mathbf{x} \in X \colon x_i = a_i - 2 \text{ for all } i < 10^{2k} \text{ and } x_{10^{2k}} < a_{10^{2k}}/2\}. \tag{4}$$

Define

$$Y = X \setminus \Big( \bigcup_{\ell \notin \{10^k \colon k \geq 2\}} Z_\ell \cup \bigcup_{k=1}^{\infty} W_k \Big) \tag{5}$$

and define $T \colon Y \to Y$ to be the first return map of $S$ to $Y$. Throughout this paper, $T$ refers to this map and $d_Y$ is any metric on $Y$ giving rise to the product topology, viewing it as a subspace of $X$. When there is no confusion as to which metric is meant, we omit the subscript and just write $d$ for the metric on $Y$. As usual, we denote elements $\mathbf{y} \in Y$ as $\mathbf{y} = (y_i)_{i \in \mathbb{N}}$.

Define $D_k$ to be the cylinder sets with largest defining index $k$ in $X \setminus Y$. More explicitly, this means that:

$$D_k = \begin{cases} Z_k & \text{if } k \notin \{10^k \colon k \geq 2\} \\ W_\ell & \text{if } k = 10^{2\ell} \text{ for some } \ell \geq 1 \\ \emptyset & \text{if } k = 10^{2\ell+1} \text{ for some } \ell \geq 1. \end{cases} \tag{6}$$

The following result is standard:

**Lemma 3.1.** *The odometer $S$ is uniquely ergodic with respect to a probability measure $\mu$, and thus the first return map $T$ is uniquely ergodic with respect to the measure $\nu = \mu(Y)^{-1} \cdot \mu|_Y$.*

It follows immediately from the construction of the set $Y$ that its measure is strictly between 0 and 1, and so the maps $T$ and $S$ are not obviously isomorphic. In fact, they are not isomorphic, as $T$ is weakly mixing (see Proposition 3.4), while $S$ has purely discrete spectrum.

**Notation** (for the systems we study throughout this article)**.** Throughout this article, $X$ is the space defined by (1), $S$ is the odometer defined on $X$ as in (2), $\mu$ is the unique ergodic measure on this system, and $(X, \mu, S)$ is the odometer system thus defined. The space $Y$ is defined by (5) and $(Y, \nu, T)$ is the associated uniquely ergodic system defined by the first return map.

Both $(X, S)$ and $(Y, T)$ are measurable maps of compact metric spaces. The remainder of this paper is devoted to studying the properties of the system $(Y, \nu, T)$.

3.2. **An overview of the behavior in the system** $(Y, \nu, T)$**.** To give an idea of what types of behaviors built into the system $(Y, \nu, T)$ give rise to it being both prime and having many self-joinings, we summarize the types of irregularities that are built into the system in the construction of the towers (see [20] for the terminology) defining the system. Namely, there are four distinct types of irregularities:

(i) The alphabet size for the odometer is typically 8, but at stage $n = 10^k$, the alphabet has size $k$. This changing in the size of the indices is necessary to allow enough room for the constructions.

(ii) For any $n \neq 10^k$, before stacking the $n-1$ columns to obtain an $n$-tower, we delete one eighth of the right most tower. This allows separation of the indices in the set $E$.
(iii) For $n = 10^{2k}$, we have $2k$ Rokhlin towers and we remove one level from each of the first $k$ of them and none from the other $k$. This is used in our construction of joininngs (in the language of [20], the joinings are built using that the system has good linked approximation of type $(m, m+1)$).
(iv) For $n = 10^{2k+1}$, before stacking $n-1$ columns to obtain an $n$-tower, we make no change (meaning no deletion). This allows us to use results of Chaika and Eskin (Theorem 4.9) and King (Theorem 4.15) to ensure that we obtain a system that is rigid rank 1.

3.3. **A further conjecture.** Maintaining the notation of this section, we state a conjecture closely related to this subject:

**Conjecture 2.** Let $a_1, a_2, \ldots \in \mathbb{N}$ with $a_i \geq 2$ for all $i \in \mathbb{N}$. Let $T$ be the corresponding odometer viewed as a measure preserving map of $[0,1]$, meaning that if $x = \sum_{j=1}^{\infty} b_j \frac{1}{a_1 \cdot \ldots \cdot a_j}$ with $b_i \in \{0, \ldots, a_i - 1\}$, then $Tx = \sum_{j=1}^{\infty} c_j \frac{1}{a_1 \cdot \ldots \cdot a_j}$ where $c_k = b_k + 1$ if $k = \min\{j \colon b_j < a_j - 1\}$, $c_i = 0$ for all $i < k$ and $c_i = b_i$ for all $i > k$. For almost every $x \in [0,1]$, the first return map of $T$ to $[0,x]$ is prime.

3.4. **Weak mixing of the transformation $T$.** Our first goal is to show that the transformation $T$ is weakly mixing, and we start with a sufficient (but not necessary) condition for a transformation to be weakly mixing.

**Lemma 3.2.** *Assume that $(Z_1, \lambda, T_1)$ is an ergodic measure preserving system with respect to the Lebesgue measure $\lambda$. If there exist a constant $c > 0$, a sequence of integers $(n_i)_{i\in\mathbb{N}}$, and sequences of measurable sets $(A_i)_{i\in\mathbb{N}}$ and $(B_i)_{i\in\mathbb{N}}$ such that*

*(i) the measures $\lambda(A_i), \lambda(B_i) > c$ for all $i \in \mathbb{N}$,*
*(ii) the limit $\lim\limits_{i\to\infty} \int_{A_i} |T_1^{n_i}x - x|\, d\lambda(x) = 0$, and*
*(iii) the limit $\lim\limits_{i\to\infty} \int_{B_i} |T_1^{n_i}x - T_1x|\, d\lambda(x) = 0$,*

*then $T_1$ is weakly mixing.*

*Proof.* Assume that $f$ is an eigenfunction of $T_1$ with eigenvalue $\gamma \neq 1$. By Lusin's Theorem, for every $\varepsilon > 0$ there exists $\delta > 0$ and a measurable set $U$ with $\lambda(U) > 1 - \varepsilon$ such that if $|x - y| < \delta$ and $x, y \in U$, then $|f(x) - f(y)| < \varepsilon$. Choose $\varepsilon < \min\{\frac{1-|\gamma|}{9}, \frac{c}{9}\}$, where $c$ is the constant given in the statement. For all sufficiently large $i \in \mathbb{N}$, by hypothesis there exists a measurable set $A_i'$ with measure at least $\frac{c}{2}$ and integer $n_i$ such that if $x \in A_i'$, then $|T_1^{n_i}x - x| < \varepsilon$. It follows that there exists $x \in A_i' \cap U$ and $T_1^{n_i}x \in U$ and so

$$|f(x) - f(T_1^{n_i}x)| = |(1 - \gamma^{n_i})| \cdot |f(x)| = |1 - \gamma^{n_i}| < \varepsilon.$$

Similarly there exists $y \in B_i \cap U$ such that $|f(y) - T_1^{n_i+1}y| = |1 - \gamma^{n_i+1}| < \varepsilon$. If these two inequalities hold simultaneously, this contradicts the choice of $\varepsilon$, and so $\gamma = 1$. Since $T_1$ is ergodic, it follows that $f$ is constant almost everywhere and so $T_1$ is weakly mixing. $\square$

Set

$$q_i = \prod_{j=1}^{i-1} a_j. \tag{7}$$

Then $S^{q_i}(\mathbf{x})$ fixes the first $i-1$ positions of $\mathbf{x}$ and increments the entry in $x_i$ position by 1. All other entries remain the same unless the $i^{th}$ position was exactly $a_i - 1$, in which case the carrying continues until this process terminates.

Given $n \in \mathbb{N}$, we choose $c_i(n)$ such that

$$n = \sum c_i(n) q_i \quad \text{with } c_i(n) \in \left\{-\frac{a_{i+1}}{2}, \dots, \frac{a_{i+1}}{2}\right\}. \tag{8}$$

Note that there is no unique choice of these coefficients, but we can make a canonical choice by using the greedy algorithm to define the coefficients $c_i$. That is, we choose $i$ and $c_i$ such that $|n - c_i q_i|$ is minimal out of all possible $i \in \mathbb{N}$ and $c_i \in \left\{-\frac{a_{i+1}}{2}, \dots, \frac{a_{i+1}}{2}\right\}$, and then iteratively choose the next coefficient to be the maximal choice satisfying these conditions. If there is a tie, that is if $|n - c_j q_j| = |n - c_{j'} q_{j'}|$ is minimal, we choose $i = \min\{j, j'\}$. Once such a representation is fixed, our construction depends on this choice.

We define two functions from $\mathbb{N}$ to itself that allow us to move between studying properties of the odometer $S$ and those of the first return map $T$:

**Notation.** We introduce two functions to relate powers of T and S. These are useful in arguments throughout the paper, most immediately in the proof that T is weakly mixing (Proposition 3.4) below. For $\mathbf{y} \in Y$, define $\zeta_{\mathbf{y}} \colon \mathbb{N} \to \mathbb{N}$ to be the map taking the integer $n$ to the integer $m$ such that $S^m \mathbf{y} = T^n \mathbf{y}$.

For $\mathbf{y} \in Y$, define $\xi_{\mathbf{y}} \colon \mathbb{N} \to \mathbb{N}$ to be the map taking the integer $n$ to the least integer $m$ such that there exists $\ell \geq n$ satisfying $T^m \mathbf{y} = S^\ell \mathbf{y}$.

Let $\mathbf{0} \in Y$ denote the point consisting of all 0's. To keep track of the iterates of $S$ that fix the first $i$ positions, as determined by the $q_i$ defined in (7) and the expansion of any integer in the base determined by the sequence $q_i$, as defined in (8), we define

$$r_i = \xi_{\mathbf{0}}(q_i) \tag{9}$$

and define

$$d_i(n) = c_i(\zeta_{\mathbf{0}}(n)). \tag{10}$$

Thus the map $\zeta_{\mathbf{y}}$ maps an iterate of $T$ to an iterate of $S$ and the coefficients $c_i$ are changed into $d_i$, while the map $\xi_{\mathbf{y}}$ reverses this, taking an iterate of $S$ to an iterate of $T$. However they are not precisely inverses, as one can not regain all of the odometer $S$ from the first return $T$: if $S^i(\mathbf{x}) \notin Y$, then there is no corresponding $T$ time.

An easy analysis of the return times for the odometer $S$ leads to (we omit the proof):

**Lemma 3.3.** *If $\mathcal{C}$ is a cylinder defined by positions $< i$, then the sum $\sum_{j=0}^{q_i-1} \mathbf{1}_{\mathcal{C}}(S^j \mathbf{x})$ does not depend on $\mathbf{x} \in X$. Thus the sum $\sum_{j=0}^{q_i-1} \mathbf{1}_{Z_\ell}(S^j \mathbf{x})$ does not depend on $\mathbf{x} \in X$ for any $\ell$ such that $\ell < i$ and similarly, $\sum_{j=0}^{q_i-1} \mathbf{1}_{W_\ell}(S^j \mathbf{x})$ does not depend on $\mathbf{x} \in X$ for any $\ell$ such that $10^{2\ell} < i$.*

We use this to show:

**Proposition 3.4.** *The system $(Y, \nu, T)$ is weakly mixing.*

*Proof.* Assume $i = 10^k - k$ and set

$$U_i = \Big(\bigcup_{j=0}^{q_i} S^{-j}\big(\bigcup_{m=i+1}^{\infty} Z_m \cup \bigcup_{10^j > i}^{\infty} W_j\big)\Big)^c.$$

We claim that if $i \geq 10^8$, then $\mu(U_i) \geq 1 - \frac{1}{8} - \sum_{m \geq i+2} \frac{1}{8^{m-i+1}}$. Indeed, under this assumption $a_m \geq 8$ and so $\mu(W_m) = \frac{m}{q_{10^{2m}+1}}$ for $m \geq i$ and $\mu(Z_\ell) = \frac{1}{q_{\ell+1}}$ for all $\ell \geq i$. Thus, $q_i \mu(Z_\ell) \leq 8^{i-\ell-1}$ for all $\ell \geq i$. By the assumption on $i$, it follows that $\sum_{\{j\colon 10^{2j} > m\}} q_i \frac{j}{q_j} < \frac{1}{8}$.

Set

$$A_i = \{\mathbf{x} \in U_i \colon x_i \leq 4\}$$

and so $\nu(A_i) > \frac{1}{2} - \frac{1}{8}$. For $x \in A_i$, we have $S^j \mathbf{x} \notin D_\ell$ for any $0 \leq j \leq q_i$ and $\ell \geq i$ (recall that the sets $D_k$ are defined in (6) and $(S^j \mathbf{x})_i \neq a_i - 2$). Thus by Lemma 3.3 and the definition of $T$, and $r_i \in \mathbb{N}$ such that $T^{r_i} \mathbf{x} = S^{q_i} \mathbf{x}$, which by choice of $q_i$ is close to $\mathbf{x}$ (note that $r_i$ is defined in (9)). Set

$$B_i = \{\mathbf{x} \in U_i \colon x_i = 7, x_1 < 5, x_{i-1} < 6\} \supset \{\mathbf{x} \in X \colon x_{i+1} < 5, x_i = 7, x_{i-1} < 6, x_1 < 5\}.$$

Then $\mu(B_i) > \frac{1}{64}$ and $\nu(B_i \cap Y) > \frac{1}{128}$. For $\mathbf{x} \in B_i$, we have $S^j \mathbf{x} \in D_i = Z_{10^k - k}$ for some $0 < j < q_i$ and by definition $S^j \mathbf{x} \notin D_\ell$ for all $\ell > i$ (because $(S^j \mathbf{x})_i \neq a_i - 2$). Thus Lemma 3.3 implies that $T^{r_i} \mathbf{x} = S^{q_i + 1} \mathbf{x}$ (by our assumption that $x_1 < 5$, we have $S^{q_i+1} \mathbf{x} \in Y$). Thus the assumptions of Lemma 3.2 are verified for the measurable sets $A_i, B_i$, and sequence of integers $n_i = r_i$ with $i \in \{10^k - k \colon k \geq 8\}$. □

3.5. $T$ **is rigid rank one by cylinders.** We now show that the constructed system is rigid rank one by cylinders, using information on the odometer system $(X, S)$ to study the system $(Y, T)$. Recall that since the system $(X, S)$ is an odometer, fixing initial entries corresponds to an interval in $[0, 1)$.

**Lemma 3.5.** *The system $(Y, \nu, T)$ is rigid rank one by cylinders.*

*Proof.* Let $I_k$ be the cylinder set determined by the word of all 0's up to $10^{2k+1}$ and with any value between 0 and $2k + 1 - 5 = 2k - 4$ in the entry at $10^{2k+1}$. Let $n_k = r_{10^{2k+1}}$, as defined in (9). If $\mathbf{y} \in \bigcup_{i=0}^{2n_k - 1} T^i(I_k)$, then $y_{10^{2k+1}} < 2k + 1 - 3$ and so $(S^i \mathbf{y})_{10^{2k+1}} < 2k + 1 - 2$ for $0 \leq i \leq n_k$. Thus, $S^i \mathbf{y} \notin \bigcup_{\ell \geq 10^{2k+1}} D_\ell$ for $0 \leq i \leq n_k$ and so

$$\bigcup_{i=0}^{2n_k - 1} T^i(I_k) \cap \Big( \bigcup_{\ell = 10^{2k+1}+1}^{\infty} Z_\ell \cup \bigcup_{j=k+1}^{\infty} W_j \Big) = \emptyset.$$

Additionally, by Lemma 3.3 we have that $\sum_{j=0}^{q_{10^{2k+1}} - 1} \mathbf{1}_{\cup_{\ell < 10^{2k+1}} D_\ell} S^j \mathbf{x}$ is constant on $X$. Therefore $\xi_{\mathbf{y}}(q_{10^{2k+1}})$ is constant (and equal to $n_k = r_{10^{2k+1}}$) on this set.

For any $\mathbf{x} \in I_k$, we have that $(T^{n_k}(\mathbf{x}))_i = x_i$ for all $i \neq 10^{2k+1}$ and $(T^{n_k}(\mathbf{x}))_{10^{2k+1}} = x_{10^{2k+1}} + 1$. Thus

$$\mu(T^{n_k} I_k \cap I_k) = (1 - \frac{1}{2k - 4}) \mu(I_k),$$

establishing condition (iv) (after passing from $\mu$ to $\nu$) of the definition of rigid rank one by cylinders. For any $\mathbf{x} \in I_k$ and $0 < i < n_k$, we have $T^i(\mathbf{x})_j \neq 0$ for some $j < 10^{2k+1}$, and so condition (ii) follows. Since each $T^i I_k$ is either contained in or is disjoint from $Z_\ell$ and $W_s$ for $\ell < 10^{2k+1}$ and $s < 2k + 1$, and furthermore is disjoint from all other $Z_\ell$ and $W_s$, we have that $T^i I_k$ is a cylinder set for all $0 \leq i < q_k$, establishing condition (i). Finally condition (iii) follows since $\bigcup_{i=0}^{n_k - 1} T^i I_k$ contains all of $Y$ other than the cylinder sets defined by having entries at least $2k-4$ in the position $10^{2k+1}$. □

## 4. Joinings

In this section we prove that our system is not quasi-distal and that the self-joinings of the system form a Poulsen simplex. We start by proving Theorem 4.1, a general criterion for a system to not be quasi-simple. As simple extensions arise via quite a general construction, it is natural that this argument becomes technical. In Sections 4.2 and 4.3, we show that our system $(Y,\nu,T)$ verifies the assumptions of Theorem 4.1. The key results used for doing this are Proposition 4.4 and Lemma 4.13, and we include a paragraph after Proposition 4.4 for a description of its role. The motivating idea behind the proof of Proposition 4.4 comes from a modification of a construction of the first named author and Eskin [4, Section 3], and in Section 4.2.1, we verify that our system $(Y,\nu,T)$ satisfies the assumptions of the construction. Lemma 4.13 is general. The fact that our joinings form a Poulsen simplex is analogous to the previously mentioned construction in [4] and is established in Section 4.4 using only the results from Section 4.2.1 (and in particular does not require Proposition 4.4). Section 4.5 establishes that these properties are residual.

4.1. **Isometric and distal extensions.** Given systems $(Z_1,\zeta_1,T_1)$ and $(Z_2,\zeta_2,T_2)$, if $\eta$ is a measure on $Z_1\times Z_2$, we make a mild abuse of notation and let $\eta_x$ denote the measure on $Z_2$ that is defined for almost all $x\in Z_1$ by disintegrating the measure $\eta$ on the fiber $\{x\}\times Z_2$. We want to have a condition to rule out that $(Z_1\times Z_2,\eta,T_1\times T_2)$ is measurably isomorphic to $T_\phi\colon X\times G/H\to X\times G/H$ by $T_\phi(x,g)=(Tx,[\phi(x)]g)$. Note that the change in the second fiber of such a map is independent of $g$ (but may depend on $x$). Theorem 4.1 is the tool to do this, and we give a rough idea how the various the conditions in the hypotheses play different roles. Condition (ii) identifies what the change in the second fiber must be (note that it is allowed to depend on $x$) and condition (iii) says that this can not be the change. Since our isomorphism is only a measurable map, conditions (i), (iv) and (v) are to allow us to be able to apply Lusin's Theorem.

**Theorem 4.1.** *Assume $(Z_1,\zeta_1,T_1)$ and $(Z_2,\zeta_2,T_2)$ are ergodic, Borel probability systems such that $Z_1$ and $Z_2$ are compact metric spaces. Let $\eta$ be an ergodic joining of the systems $(Z_1,\zeta_1,T_1)$ and $(Z_2,\zeta_2,T_2)$, and let $c>0$. Assume that there exists $\hat\delta>0$, a sequence of integers $(n_i)_{i\in\mathbb{N}}$ tending to infinity, a sequence of integers $(L_i)_{i\in\mathbb{N}}$ such that $L_i>\hat\delta n_i$, and measurable sets $A_i\subset Z_1$ satisfying*

*(i) $\zeta_1(A_i)>c$ for all $i\in\mathbb{N}$.*

*Further assume that for each $x\in A_i$, there exist sets $C_i(x),E_i(x)\subset Z_2$, and $j_x\in[-n_i,n_i]$ (all depending on $x$) satisfying the following conditions:*

*(ii) $\sup\limits_{x\in A_i}\ \sup\limits_{y\in C_i(x)} \frac{1}{L_i}\sum_{\ell=0}^{L_i-1} d_{Z_2}(T_2^\ell T_2^{n_i}y, T_2^\ell T_2^{j_x}y)\to 0$.*

*(iii) For all $y\in E_i(x)$, we have*

$$\frac{1}{L_i}|\{0\le \ell\le L_i-1\colon d_{Z_2}(T_2^\ell T_2^{n_i}y, T_2^\ell T_2^{j_x}y)>c\}|>c.$$

*(iv) For all $x\in A_i$, $\eta_x(C_i(x)),\eta_x(E_i(x))>c$ .*

(v) *For any $c' > 0$, there exists $i_0$ such that for all $i \geq i_0$ and any $x \in A_i$ if we have balls $B(p_\ell, c') \subset Z_2$ satisfying $\eta_x(E_i(x) \cap \cup B(p_\ell, c') > c - c'$ then $\eta_x(C_i(x) \cap \cup B(p_\ell, 2c')) > c - 2c'$.*

*Then $\eta$ is not a distal extension of $(Z_1, \zeta_1, T_1)$.*

Note that this is a general result, holding for arbitrary measure preserving systems whose underlying spaces are compact metric spaces, and this result does not depend on the particular constructions we have for the systems $(Z_1, \zeta_1, T_1)$ and $(Z_2, \zeta_2, T_2)$. We further note that in (ii), we can not take $j_x = n_i$, as this would preclude Condition (iii). Note that since Condition (v) holds for arbitrarily small choice of $c'$, this rules out the possibility that the joining is carried on a finite union of graphs.

The proof of Theorem 4.1 proceeds by contradiction. We assume $(Z_1 \times Z_2, \eta, T_1 \times T_2)$ is an isometric extension of $(Z_1, \zeta_1, T_1)$, meaning that there exists a (measurable) isomorphism $\Psi\colon (Z_1 \times Z_2, \eta, T_1 \times T_2) \to (Z_1 \times G/H, \zeta_1 \times m_{G/H}, T_\phi)$ that is the identity on the first coordinate, and use this to derive a contradiction. Since a distal system can be decomposed as a tower of isometric extensions, we conclude that it can not be a distal extension.

Before turning to this proof, we start with some preliminaries and a lemma used to derive the contradiction.

Let $\mathcal{K}$ be a compact continuity set for $\Psi$ with $\eta(\mathcal{K}) > 1 - \frac{\hat{\delta}}{100}c^4$. Thus $\mathcal{K}$ is also a continuity set for $\pi_2 \circ \Psi$, where $\pi_2\colon Z_1 \times G/H \to G/H$ is the projection on the second coordinate. Choose $\delta > 0$ such that $d_{G/H}(g\tilde{h}, g\tilde{h}') < \frac{c}{8}$ whenever $d_{G/H}(\tilde{h}, \tilde{h}') < \delta$ and $g \in G$. Choose $\frac{c}{8} > \delta' > 0$ such that $d_{G/H}(\pi_2 \circ \Psi(x,y), \pi_2 \circ \Psi(x'y')) < \delta$ whenever $(x,y), (x',y') \in \mathcal{K}$ and $d_{Z_1 \times Z_2}\Big((x,y),(x',y')\Big) < \delta'$.

**Lemma 4.2.** *Under the assumptions of Theorem 4.1, there exist a pair of points $(x,y), (x,y') \in Z_1 \times Z_2$ and $b \in \mathbb{Z}$ such that*

(i) $(x,y), (x,y'), (T_1^b x, T_2^b y)$, *and* $(T_1^b x, T_2^b y') \in \mathcal{K}$;
(ii) $d_{Z_2}(y, y') < \delta'$;
(iii) $d_{Z_2}(T_2^b y, T_2^b y') > \frac{c}{2} - \frac{c}{8} > \frac{c}{3}$.

*Proof.* For all $L \geq 1$, we have that $\eta(\{(x,y)\colon \sum_{i=0}^{L-1} \mathbf{1}_{\mathcal{K}}\Big((T_1^i x, T_2^i y)\Big) < L - \frac{L\hat{\delta}}{10}c\}) < \frac{1}{10}c^3$. Choosing $c' = \frac{\delta'}{8}$ as in Condition (v) of Theorem 4.1, for all sufficiently large $i$, we can pick $x \in A_i$, $y \in C_i(x)$, and $y' \in E_i(x)$ satisfying $d_{Z_2}(y, y') < \delta'$ and the conditions

$$\sum_{i=0}^{L_i+n_i-1} \mathbf{1}_{\mathcal{K}}\Big((T_1^i x, T_2^i y)\Big) > L_i + n_i - \frac{L_i}{10}c,$$
$$\sum_{i=0}^{L_i+n_i-1} \mathbf{1}_{\mathcal{K}}\Big((T_1^i x, T_2^i y')\Big) > (L_i + n_i) - \frac{L_i}{10}c.$$

By Conditions (ii) and (iii) of Theorem 4.1, there exists $\ell$ such that the points $(T_1^{\ell+n_i}x, T_2^{\ell+n_i}y)$, $(T_1^{\ell+j_x}x, T_2^{\ell+j_x}y)$, $(T_1^{\ell+n_i}x, T_2^{\ell+n_i}y')$, and $(T_1^{\ell+j_x}x, T_2^{\ell+j_x}y')$ all lie in the set $\mathcal{K}$, while at the same time $d_{Z_2}(T_2^{\ell+n_i}y, T_2^{\ell+j_x}y) < \delta'$ and $d_{Z_2}(T_2^{\ell+n_i}y', T_2^{\ell+j_x}y') >$

$c$. Thus we can take $b$ to be one of $\ell + j_x$ or $\ell + n_i$. Indeed,

$$\max\Big\{d_{Z_2}(T^{\ell+n_i}y, T^{\ell+n_i}y'),\, d_{Z_2}(T^{\ell+j_x}y, T^{\ell+j_x}y')\Big\} \geq$$
$$d_{Z_2}(T^{\ell+n_i}y', T^{\ell+j_x}y') - d_{Z_2}(T^{\ell+n_i}y, T^{\ell+j_x}y). \qquad \square$$

*Proof of Theorem 4.1.* We first show that $\eta$ is not an isometric extension. Observe that if $g = \phi(T_1^{b-1}x) \cdot \ldots \cdot \phi(x)$, then $\pi_2 \circ \Psi(T_1^b x, T_2^b y) = g\pi_2 \circ \Psi(x, y)$ and $\pi_2 \circ \Psi(T_1^b x, T_2^b y') = g\pi_2 \circ \Psi(x, y')$. Because all four of the points $(T_1^{\ell+n_i}x, T_2^{\ell+n_i}y)$, $(T_1^{\ell+j_x}x, T_2^{\ell+j_x}y)$, $(T_1^{\ell+n_i}x, T_2^{\ell+n_i}y')$, and $(T_1^{\ell+j_x}x, T_2^{\ell+j_x}y')$ lie in the set $\mathcal{K}$, Conclusion (ii) of Lemma 4.2 implies that $d_{Z_2}(T_2^b y, T_2^b y') < \frac{c}{4}$, a contradiction of Conclusion (iii) of Lemma 4.2.

Now assume that $\eta$ is a distal extension of $(Z_1, \zeta_1, T_1)$. By the structure theorem for distal flows of Furstenberg [10] and Zimmer [34], the system $(Z_1 \times Z_2, \eta, T_1 \times T_2)$. is an inverse limit of systems, each of which is an isometric extension of the preceding one. Thus there is a factor of our distal extension, which is an isometric extension of $(Z_1, \zeta_1, T_1)$ , and which satisfies the assumptions of Theorem 4.1 (with different $c$). Indeed, by the definition of inverse limits, we can embed our distal extension into the product defining the inverse limit. This contradicts the previous paragraph. $\square$

4.2. **A self-joining that is not quasi-simple.** We apply Theorem 4.1 to establish part of Theorem 1.1:

**Theorem 4.3.** *There exists an non-trivial ergodic self-joining of $(Y, \nu, T)$ that is not a distal extension of $(Y, \nu, T)$.*

By non-trivial, we mean that joining is not trivial in the sense that it is not a product of the system with itself.

Before turning to the proof, we start with some preliminaries. If $(Y, T)$ is a compact metric space, let $\mathcal{M}(Y \times Y)$ denote the set of Borel probability measures on $Y$ and let $d_{\mathcal{M}(Y\times Y)}(\cdot, \cdot)$ denote the *Kantorovich-Rubenstein metric*, defined for Borel probability measures $\mu, \nu \in \mathcal{M}(Y \times Y)$ as

$$d_{\mathcal{M}(Y\times Y)}(\mu, \nu) := \sup\Big\{\Big|\int f d\mu - \int f d\nu\Big| \colon f \text{ is 1-Lipschitz function on } Y \times Y\Big\}.$$

This metric endows the set of Borel probability measures $\mathcal{M}(Y \times Y)$ on $Y \times Y$ with the weak*-topology. Similarly, define $d_{\mathcal{M}(Y)}$ to be the Kantorovich-Rubenstein metric on the $\mathcal{M}(Y)$.

Recall that $J(n)$ denotes the off diagonal joining on $\{(\mathbf{x}, T^n\mathbf{x})\}$, meaning that $J(n)$ is the measure on $X \times X$ defined by

$$\int f(\mathbf{x}, \mathbf{y})\, dJ(n) = \int f(\mathbf{x}, T^n\mathbf{x})\, d\mu.$$

Recall that if $\sigma$ is a self-joining of $(Y, \nu, T)$, we let $\sigma_{\mathbf{x}}$ denote the disintegration of $\sigma$ given by projection to the first coordinate, thought of as a measure on $Y$. Note that this is only defined $\nu$-almost everywhere and is slightly different than the usual disintegration of measures: it defines a measure on $Y$, rather than a measure on $Y \times Y$ that gives full measure to $\{\mathbf{x}\} \times Y$.

The main tool in establishing Theorem 4.3 is the following proposition:

**Proposition 4.4.** *For any $\varepsilon > 0$ and $k_1, \ldots, k_r \in \mathbb{Z}$, there exist $\ell_1, \ldots, \ell_{2r}$, $N, M, L \in \mathbb{Z}$, a set $A \subset Y$ with $\nu(A) > \frac{1}{99}$, and for each $\mathbf{x} \in A$ there exists $j_{\mathbf{x}} \in [-M, M]$ such that* *(a)* $\nu\Big(\big\{\mathbf{x}\colon d_{\mathcal{M}(Y)}\big((\frac{1}{2r}\sum_{n=1}^{2r}J(\ell_n))_{\mathbf{x}},(\frac{1}{r}\sum_{n=1}^{r}J(k_n))_{\mathbf{x}}\big) > \varepsilon\big\}\Big) < \varepsilon$. *(Recall our convention that the disintegration of measure on $Y\times Y$ by projection onto the first coordinate is a measure on $Y$.)*

*(b)*

$$\nu\big(\{\mathbf{x}\in A\colon \text{there exist reorderings } i_1,\dots,i_r \text{ of } 1,\dots,r \text{ and } i_{1+r},\dots,i_{2r} \text{ of } r+1,\dots,2r \text{ such that for all } 1\le s\le r,\ d_Y\big(T^{k_s}\mathbf{x},T^{\ell_{i_s}}\mathbf{x}\big)<\varepsilon \text{ and } d_Y\big(T^{k_s}\mathbf{x},T^{\ell_{i_{r+s}}}\mathbf{x}\big)<\varepsilon\}\big) > \nu(A)-\varepsilon.$$

*(c)* $d_{\mathcal{M}(Y\times Y)}\big(\frac{1}{N}\sum_{i=1}^{N}\delta_{(T^i\mathbf{x},T^iT^{\ell_n}\mathbf{x})},\frac{1}{r}\sum_{i=1}^{r}J(k_i)\big)\}\big) < \varepsilon$ *for all* $n\le 2r$ *and* $\mathbf{x}\in A$.

*(d)* $\frac{1}{L}\sum_{i=0}^{L-1}d_Y(T^{M+i}T^{\ell_n}\mathbf{x},T^{i+j_{\mathbf{x}}}T^{\ell_n}\mathbf{x})<\varepsilon$ *for all* $\mathbf{x}\in A$ *and* $n\le r$.

*Moreover, if we assume that there exist* $a,b\in\mathbb{N}$ *and* $c>0$ *such that* $d_Y(T^a\mathbf{x},T^b\mathbf{x}) > 4c+\varepsilon$ *for a set* $W$ *of* $\mathbf{x}$ *with* $\nu(W)=\frac{1}{2}$ *and*

$$|\{1\le n\le r\colon d_Y(T^{k_n}\mathbf{x},T^a\mathbf{x})<c\}| = |\{1\le n\le r\colon d_Y(T^{k_n}\mathbf{x},T^b\mathbf{x})<c\}| = \frac{r}{2} \tag{11}$$

*for all* $\mathbf{x}\in W$, *then*

*(e)* $\frac{1}{L}|\{0\le i\le L-1\colon d_Y(T^{M+i}T^{\ell_d}\mathbf{x},T^iT^{\ell_d}T^{j_{\mathbf{x}}}\mathbf{x}) > c$ *for all* $\mathbf{x}\in A$ *and* $r<d\le r+2\lceil\frac{1}{16}r\rceil\}| > \frac{1}{9}$.

The proof of this proposition occupies the rest of this section, starting with finding the first half of the $\ell_i$ and then the second half. Before we turn to this, we comment on the role that this proposition plays. To prove Theorem 4.3, we iteratively apply this proposition, and at the $k-1^{\text{st}}$ application obtain a joining that is the barycenter of $2^k$ off diagonal joinings. We then take the weak*-limit of this sequence of (non-ergodic) joinings and obtain $\sigma$, an ergodic joining that satisfies the criterion of Theorem 4.1. Using the proposition, we obtain a joining with the desired properties before passing to a limit. Before turning to the proof, we give some indication of the role of the various conditions in the statement. Conditions (a) and (b) are used to prove that $\sigma$ is ergodic, and Conditions (d) and (e) are used to show that $\sigma$ satisfies the assumptions of Theorem 4.1. More precisely, the sets $C_i(x)$ for $\sigma$ are approximated by $\{T^{\ell_n}\mathbf{x}\}_{n=1}^r$ in the sense that $\eta_x$ restricted to $C_i(x)$ is close to $\frac{1}{2r}\sum_{n=1}^r\delta_{T^{\ell_n}\mathbf{x}}$. Similarly, the sets $\{T^{\ell_n}\mathbf{x}\}_{n=r+1}^{r+2\lceil\frac{r}{16}\rceil}$ correspond to $E_i(x)$, $M$ corresponds to $n_i$, and $L$ to $L_i$. Conclusion (e) is the analog of (iii) in Theorem 4.1 and Conclusion (d) of (ii). Condition (v) in Theorem 4.1 corresponds to observing that (b) implies that for most $\mathbf{x}$, for any $r<d\le r+2\lceil\frac{r}{16}\rceil$ there exists $1\le i_d\le i_r$ such that $T^{\ell_d}\mathbf{x}$ is close to $T^{\ell_{i_d}}\mathbf{x}$. The relation between the pre-limit versions of the properties and the desired properties for the limiting measure $\sigma$ is addressed in Lemma 4.13.

4.2.1. *Finding* $\ell_1,\dots,\ell_r$. We now construct $\ell_1,\dots,\ell_r$ satisfying the conclusions of Proposition 4.4.

**Lemma 4.5.** *For all* $\varepsilon>0$, *there exists* $k_0\in\mathbb{N}$ *such that for all* $k>k_0$ *and* $1\le\ell\le k$*:*

*(i) If* $(T^i\mathbf{x})_{10^{2k}}<k$ *for all* $0\le i\le \ell r_{10^{2k}}$, *then* $d(T^{\ell r_{10^{2k}}}\mathbf{x},\mathbf{x})<\varepsilon$.
*Similarly, if* $(T^i\mathbf{x})_{10^{2k}}<k$ *for all* $0\ge i\ge -\ell r_{10^{2k}}$, *then* $d(T^{\ell r_{10^{2k}}}\mathbf{x},\mathbf{x})<\varepsilon$.

*(ii) If* $k \le (T^i\mathbf{x})_{10^{2k}} < 2k-2$ *for all* $0 \le i \le \ell(r_{10^{2k}}+1)$*, then* $d(T^{\ell(r_{10^{2k}}+1)}\mathbf{x},\mathbf{x}) < \varepsilon$.
*Similarly, if* $k \le (T^i\mathbf{x})_{10^{2k}} < 2k-2$ *for all* $0 \ge i \ge -\ell(r_{10^{2k}}+1)$*, then* $d(T^{\ell(r_{10^{2k}}+1)}\mathbf{x},\mathbf{x}) < \varepsilon$.

Recall from Section 3 that $d = d_Y$ is a metric giving the subspace topology for $Y$ coming from product topology on $X$.

*Proof.* We only include the proof of the first part of (ii), as the proofs of all four statements are similar. Thus we need to show that under the assumptions, $(T^{\ell(r_{10^{2k}}+1)}\mathbf{x})_i = x_i$ for all $i < 10^{2k}$. This statement immediately follows once we show that

$$T^{\ell(r_{10^{2k}}+1)}\mathbf{x} = S^{\ell q_{10^{2k}}}\mathbf{x}. \tag{12}$$

To prove (12), note that by assumption, $(T^i\mathbf{x})_{10^{2k}} < 2k-2$ for all $0 \le i \le \ell(r_{10^{2k}}+1)$, and so $S^i\mathbf{x} \notin \bigcup_{j>10^{2k}} D_j$ for all $0 \le i \le \ell q_{10^{2k}}$. (Recall that $D_j$ are defined in (6).) Similarly, by the assumption that $k \le (T^i\mathbf{x})_{10^{2k}}$ for all $0 \le i \le \ell(r_{10^{2k}}+1)$, we have that $S^i\mathbf{x} \notin D_{10^{2k}}$ for all $0 \le i \le \ell q_{10^{2k}}$. Thus for any such $\mathbf{x}$, $\zeta_{\mathbf{x}}(q_{10^{2k}}) = \zeta_{\mathbf{0}}(q_{10^{2\ell}})+1 = r_{10^{2k}}+1$. Indeed, there exists $0 < i < r_{10^{2k}}$ such that $S^i(\mathbf{0}) \in D_{10^{2k}}$. Iterating this process for $\mathbf{x}$, we obtain that $S^{q_{10^{2k}}}\mathbf{x} = T^{r_{10^{2k}}+1}\mathbf{x}, \dots, S^{\ell q_{10^{2k}}}\mathbf{x} = T^{\ell(r_{10^{2k}}+1)}\mathbf{x}$, thus proving the claim. (Note that by assumption, $S^i(T^{j(r_{10^{2k}}+1)}\mathbf{x}) \notin D_{10^{2k}}$ for any $0 \le j \le \ell$ and $0 \le i \le q_{10^{2k}}$.) If $k_0$ is large enough (depending on the metric, $d$, and $\varepsilon$), the lemma follows. $\square$

**Lemma 4.6.** *Let* $u, v \in \mathbb{Z}$ *and* $n = u + vr_{10^{2k}}$. *If*

$$x_{10^{2k}} \in [|u|+|v|+1, k-|u|-|v|-1], \tag{13}$$

*then* $(T^n\mathbf{x})_j = (T^u\mathbf{x})_j$ *for all* $j \neq 10^{2k}$. *Similarly, if*

$$x_{10^{2k}} \in [|u|+|v|+k+1, 2k-|u|-|v|-3], \tag{14}$$

*then* $(T^n\mathbf{x})_j = (T^{u-v}\mathbf{x})_j$ *for all* $j \neq 10^{2k}$.

*Proof.* These results follow from Lemma 4.5, and again we only prove the first part as the others are analogous. If $x_{10^{2k}} \in [|u|+|v|+1, k-|u|-|v|-1]$, then we apply the first part of Lemma 4.5 with $\ell = v$ to $T^u\mathbf{x}$. $\square$

**Corollary 4.7.** For all $\varepsilon > 0$ and $b, b' \in \mathbb{Z}$, there exists $k_0$ such that for all $\ell > k_0$ there exists $p_\ell \in \mathbb{Z}$, disjoint sets $A_\ell$, $B_\ell$, and a cylinder $J_\ell$ satisfying

(i) $T^{p_\ell}(\mathbf{x})_j = (T^b\mathbf{x})_j$ for all $j \neq 10^{2\ell}$ and $\mathbf{x} \in A_\ell$.
(ii) $T^{p_\ell}(\mathbf{x})_j = (T^{b'}\mathbf{x})_j$ for all $j \neq 10^{2\ell}$ and $\mathbf{x} \in B_\ell$.
(iii) $\nu(A_\ell)$, $\nu(B_\ell) > \frac{1}{2} - \varepsilon$.
(iv) $A_\ell = \bigcup_{i=0}^{\left(\ell-2(|b|-|b'|-1)\right)r_{10^{2\ell}}} T^iJ_\ell$
(v) $T^iJ_\ell \cap J_\ell = \emptyset$ for all $0 \le i \le 2(\ell-|b|-|b'|-1)r_{10^{2k}}$.

*Proof.* We apply Lemma 4.6 to $n = b + (b-b')r_{10^{2\ell}}$, and as in the lemma, we write $n = u + vr_{10^{2k}}$. (That is, $u = b$ and $v = b - b'$.) Since $b - (b-b') = b'$, by choosing

$$J_\ell = \{\mathbf{x}\colon x_j = 0 \text{ for all } j < 10^{2\ell} \text{ and } x_{10^{2\ell}} = |u|+|v|+1\},$$

the corollary follows with $p_\ell = b + (b-b')r_{10^{2\ell}}$. Indeed, $A_\ell$ satisfies (13) and

$$B_\ell = \bigcup_{i=(|u|+|v|+\ell)r_{10^{2\ell}}}^{(2\ell-|u|-|v|)r_{10^{2\ell}}} T^iJ_\ell$$

satisfies (14). Clearly the measures of each of these sets converge to $\frac{1}{2}$ as $\ell$ goes to infinity (for fixed $u, v$). $\square$

**Lemma 4.8.** *For any $\varepsilon > 0$ and $b \in \mathbb{Z}$, there exists $N \in \mathbb{N}$ such that for all $n \geq N$,*

$$\nu(\{\mathbf{x} \in Y : d_{\mathcal{M}(Y\times Y)}\big(\sum_{i=1}^{n} \delta_{(T^i\mathbf{x}, T^iT^b\mathbf{x})}, J(b)\big) > \varepsilon\}) < \varepsilon.$$

*Proof.* This follows immediately from the Ergodic Theorem and the compactness of the space of 1-Lipschitz functions with bounded integral. $\square$

We now combine these results with the strategy developed in [4] to build off diagonal joinings that are weak-*close to the barycenter of other off diagonal joinings. We begin by summarizing the results of [4], where the input is a sequence of numbers and sets with certain properties.

We assume that $c > 0$, $J_j$ is a sequence of cylinders, $m_j$ is a sequence of natural numbers, $b_j^{(1)}, \dots, b_j^{(d)}$ are sequences of integers, and $\hat{A}_j$, $\hat{B}_j$ and $U_j$ are sequences of sets, and $\varepsilon_j > 0$ satisfy the following properties:

(i) For all $j$, $\hat{A}_j = \bigcup_{\ell=1}^{m_j} T^\ell J_j \setminus U_j$.
(ii) For all $j$, $\hat{B}_j = Y \setminus (A_j \cup U_j)$.
(iii) For all $j$, $\nu(\hat{A}_j), \nu(\hat{B}_j) > c$.
(iv) The minimal return time to $J_j$ is at least $\frac{3}{2} m_j$.
(v) For all $j$, $\nu(U_j) < \varepsilon_j$.
(vi) For all $j$, $m_j \sum_{\ell=j+1}^{\infty} \nu(J_\ell) < \varepsilon_\ell$.
(vii) For all $j$, $\varepsilon_{j+1} \leq \varepsilon_j$ and $\sum_j \varepsilon_j < \infty$.
(viii) For any $\mathbf{x} \in \hat{A}_j$, we have $d(T^{b_j^{(p)}}\mathbf{x}, T^{b_{j-1}^{(p-1)}}\mathbf{x}) < \varepsilon_j$ and for any $\mathbf{x} \in \hat{B}_j$, we have $d(T^{b_j^{(p)}}\mathbf{x}, T^{b_{j-1}^{(p)}}\mathbf{x}) < \varepsilon_j$. (Note that $b_{j-1}^{(p-1)}$ is interpreted to be $b_{i-1}^{(d)}$ if $p = 1$.)
(ix) $d_{\mathcal{M}(Y\times Y)}\big(\frac{1}{L}\sum_{i=1}^{L}(T\times T)^i(J(b_i^{(p)})_{\mathbf{x}}), J(b_i^{(p)})\big) < \varepsilon_i$ for all $\mathbf{x} \in Y$, all $L \geq \frac{m_{k_{i+1}}}{9}$, and any $p \in \{1, \dots, d\}$.

**Theorem 4.9** (Chaika-Eskin [4, Proposition 3.1 and (the proof of) Corollary 3.3])**.** *Assuming sequences of numbers and sets satisfying* (i)- (ix)*, there exist $\rho < 1$, $C' > 0$ (depending only on $c$ and $d$) such that*

$$d_{\mathcal{M}(Y\times Y)}\big(J(b_k^{(p)}), \frac{1}{r}\sum_{p=1}^{r} J(b_i{}^{(p)})\big) \leq C' \sum_{q=i}^{k} \varepsilon_q + C'\rho^{k-i},$$

*whenever $k \geq i$ and $p \in \{1, \dots, r\}$. Moreover, if $\mathbf{x} \notin \bigcup_{q=i}^{k} U_q$, there is a reordering (which is allowed to depend on $\mathbf{x}$) $p_1, \dots, p_d$ with $d(T^{b_i^{(j)}}\mathbf{x}, T^{\hat{b}_k^{(p_j)}}\mathbf{x}) < \sum_{q=1}^{k} \varepsilon_q$ for all $1 \leq j \leq d$.*

**Remark 1.** The last statement of this theorem is not in the statement of Corollary 3.3, but follows by iterating (viii). The condition in (viii) is a slightly simpler condition than that in [4], where the conditional measure of an off diagonal joining on a fiber is used instead of the distance between points, but the condition in [4] follows immediately by using the definition of the Kantorovich-Rubinstein metric.

**Remark 2.** We iteratively apply the result of Theorem 4.9 for different (decreasing) choices of $\varepsilon_i$ and (increasing) $d$, with each choice satisfying all of the properties (i)–(ix). Corollary 4.7 is designed to ensure that conditions (i)-(ix) in the hypotheses of Theorem 4.9 hold. Indeed conclusions (i) and (ii) of Corollary 4.7 provide condition (viii). Conclusion (v) of Corollary 4.7 provides condition (iv), while conclusion (iv) provides condition (i). Lemma 4.8 provides condition (ix).

**Corollary 4.10.** For any $\varepsilon > 0$ and integers $b_1, \dots, b_d$, there exist integers $\hat{b}_1, \dots, \hat{b}_d$ such that

$$d_{\mathcal{M}(Y\times Y)}\big(J(\hat{b}_\ell), \frac{1}{d}\sum_{j=1}^{d} J(b_j)\big) < \varepsilon \tag{15}$$

for all $\ell \in \{1, \dots, d\}$. Moreover, we may assume that there is a set $\hat{W}$ of measure $1-\varepsilon$ such that for every $\mathbf{x} \in \hat{W}$, there is a reordering $p_1, \dots, p_d$ with $d(T^{b_j}\mathbf{x}, T^{\hat{b}_{p_j}}\mathbf{x}) < \varepsilon$ for all $1 \le j \le d$.

We note that the reordering in the second part of this statement depends on the particular $\mathbf{x}$.

*Proof.* By Corollary 4.7 and Lemma 4.8 , the result holds for $d = 2$. Indeed, given $b_1, b_2$, and $\varepsilon'$, Lemma 4.8 provides $L_0$ such that

$$\nu(\{x\colon d_{\mathcal{M}(Y\times Y)}\left(\frac{1}{L}\sum_{i=1}^{L}\delta_{(T\times T)^i(\mathbf{x}, T^{b_j}\mathbf{x})}, J(b_j)\right) > \varepsilon'\}) < \varepsilon'$$

for $j \in \{1, 2\}$ and $L \ge L_0$. Given this $L_0$, we apply Corollary 4.7 twice (for sufficiently large $\ell$ depending on $L_0$) to obtain sets $A$, $B$ and $p, p' \in \mathbb{Z}$ for $(b, b') = (b_1, b_2)$ and $(b, b') = (b_2, b_1)$ respectively as in the statement of the Corollary and such that $\nu(\cap_{i=0}^{L_0} A)$ and $\nu(\cap_{i=0}^{L_0} B)$ are at least $\frac{1}{2} - \varepsilon'$. Iterating this provides conditions (i)-(ix). In particular, in the next application we have $p, p'$ instead of $b_1, b_2$.

Moreover, we claim that we can simultaneously apply these results to $d$ different pairs $b_1, b_1', \dots, b_d, b_d'$ (the resulting common sets become $\hat{A}$ and $\hat{B}$). To see this, choose

$$\hat{A} = \{\mathbf{x}\colon x_{10^{2\ell}} \in [\max\{|b_i|\} + \max\{|b_i - b_i'|\} + 1, \ell - (\max\{|b_i|\} + \max\{|b_i - b_i'|\}) - 1]\}$$

and

$$\hat{B} = \{\mathbf{x}\colon x_{10^{2\ell}} \in [\ell + \max\{|b_i|\} + \max\{|b_i - b_i'|\} + 1, 2\ell - (\max\{|b_i|\} + \max\{|b_i - b_i'|\}) - 3]\}.$$

We apply this argument for the $d$ pairs $b_1^{(1)}, b_1^{(2)}; \dots; b_1^{(d)}, b_1^{(1)}$ to produce measures $b_2^{(1)}, \dots, b_2^{(d)}$. Note that on $\hat{A}$, there is a reordering of $1, \dots, d$, call it $p_1, \dots, p_d$, such that $d(T^{b_j}\mathbf{x}, T^{\hat{b}_{p_j}}\mathbf{x}) < \varepsilon$ for all $1 \le j \le d$ and $\mathbf{x} \in \hat{A}$. (In fact, this is the reordering $p_j = j$.) There is a similar reordering on $\hat{B}$ (this is the reordering $p_j = j - 1$ for $j \ne 1$ and $p_1 = d$). Inductively, given $b_j^{(1)}, \dots, b_j^{(d)}$ we apply this to the corresponding pairs $b_j^{(1)}, b_j^{(2)}; \dots; b_j^{(d)}, b_j^{(1)}$. Let $\hat{A}_1, \dots, \hat{A}_j$ and $\hat{B}_1, \dots, \hat{B}_j$ denote the corresponding sets, as above. By Theorem 4.9, there exists $j \in \mathbb{N}$ such that $b_j^{(i)}$ satisfy (15) for all $i = 1, \dots, d$. That is,

$$d_{\mathcal{M}(Y\times Y)}\big(b_j^{(i)}), \frac{1}{d}\sum_{\ell=1}^{d} J(b_\ell)\big) < \varepsilon.$$

Define

$$\hat{W} = \bigcap_{i=1}^{j} (\hat{A}_i \cup \hat{B}_i) \tag{16}$$

to be the intersection of the sets obtained at each step, and this satisfies the desired conclusion. □

We now combine these results to find $\ell_1, \ldots, \ell_r$. Given $\varepsilon > 0$, applying Corollary 4.10 to $k_1, \ldots, k_r$, we obtain $\ell_1, \ldots, \ell_r$ such that

$$\nu\big(\{\mathbf{x}\colon d_{\mathcal{M}(Y)}(\frac{1}{r}\sum_{i=1}^{r} J(\ell_i)_{\mathbf{x}}, \frac{1}{r}\sum_{i=1}^{r} J(k_i)_{\mathbf{x}}\big) > \varepsilon/2\}) < \varepsilon/2$$

(Condition (a) for the first $r$) and also satisfy the reordering condition on $\ell_1, \ldots, \ell_r$ in (b), where for each $\mathbf{x} \in \hat{W}$, the reordering is given by $p'_k = k - |\{1 \leq i \leq j\colon \mathbf{x} \in \hat{B}_i\}|$ and this difference is taken modulo $d$. (Note that as we have not yet introduced $\ell_{r+1}, \ldots, \ell_{2r}$, we have not yet fully established (a) or (b).) Towards obtaining Conclusion (c), for each $\ell_i$, choose $N_i$ such that for all $L \geq N_i$ we have

$$d_{\mathcal{M}(Y\times Y)}\big(\frac{1}{L}\sum_{j=0}^{L-1} \delta_{(T^j\times T^j)(\mathbf{x}, T^{\ell_i}\mathbf{x})}, J(\ell_i)\big) < \varepsilon.$$

4.2.2. *Finding $\ell_{r+1}, \ldots, \ell_{2r}$.* We start first by finding $\ell_{r+1}, \ldots, \ell_{2\lceil\frac{1}{16}r\rceil}$.

**Lemma 4.11.** *Under the assumptions of Proposition 4.4 including the additional assumption, there exists $J \subset \{1, \ldots, r\}$ with $|J| = 2\lceil\frac{1}{2\cdot 8}r\rceil$ and an order 2 bijection $\phi\colon J \to J$ such that*

$$\nu(\{\mathbf{x}\colon d(T^{\ell_i}\mathbf{x}, T^{\ell_{\phi(i)}}\mathbf{x}) > c\}) > \frac{1}{8} \tag{17}$$

*for all $i \in J$.*

*Proof.* First we claim that for each $i \leq r$, we have that

$$\Big|\Big\{j \leq r\colon \nu\big(\{\mathbf{x}\colon d(T^{\ell_i}\mathbf{x}, T^{\ell_j}\mathbf{x}) > c\}\big) > \frac{1}{8}\Big\}\Big| \geq \frac{r}{8}. \tag{18}$$

To justify (18), we limit our consideration to $W \cap \hat{W}$, where $\hat{W}$ is defined as in (16) as given in the proof of Corollary 4.10 and note that $\nu(W \cap \hat{W}) \geq \frac{1}{2} - \varepsilon > \frac{49}{100}$. If (18) does not hold,

$$\int_{W\cap\hat{W}} \frac{1}{r}|\{j\colon d(T^{\ell_j}\mathbf{x}, T^{\ell_i}\mathbf{x}) \leq c\}|\, d\nu \geq (\frac{49}{100} - \frac{1}{8})\frac{7}{8} > \frac{1}{2}\nu(W \cap \hat{W}).$$

It follows that there exists $\mathbf{x} \in (W \cap \hat{W})$ such that

$$|\{j : d(T^{\ell_j}\mathbf{x}, T^{\ell_i}\mathbf{x}) < c\}| > \frac{r}{2}.$$

Since $\mathbf{x} \in \hat{W}$, it follows that

$$|\{j\colon d(T^{k_j}\mathbf{x}, T^{\ell_i}\mathbf{x}) < c + \varepsilon\}| > \frac{r}{2}.$$

But since $\mathbf{x} \in W \cap \hat{W}$, we have that $d(T^{\ell_i}\mathbf{x}, T^a\mathbf{x})$ or $d(T^{\ell_i}\mathbf{x}, T^b\mathbf{x})$ is less than $c+\varepsilon$, all of these $T^{k_j}\mathbf{x}$ are at least $2c$ away from whichever of $T^a\mathbf{x}$ or $T^b\mathbf{x}$ that $T^{\ell_i}\mathbf{x}$ is not close to. This contradicts the fact that $\mathbf{x} \in W$.

Given (18), we can obtain our set of $J$, because until $|J| = \lceil \frac{1}{8}r \rceil$, we can always inductively pick any $i \notin J$ and find $j \notin J$ satisfying (17) and add them both into $J$, letting $\phi(i) = j$ and $\phi(j) = i$. Thus we can obtain a set $J$ whose cardinality is the smallest even number that is at least $\frac{1}{8}r$. □

**Lemma 4.12.** *Assume there exist $a, b \in \mathbb{Z}$ and $c > 0$ such that*

$$\nu(\{\mathbf{x}\colon d(T^a\mathbf{x}, T^b\mathbf{x}) > c\}) > \frac{1}{8}.$$

*Let*

$$G_k = \{\mathbf{x} \in Y \colon x_{10^{2k}} \in [\frac{1}{3}k + |a-b|, \frac{1}{2}k - 2 - |a-b|] \text{ and } x_{10^{2k}-1} = 3\}$$

*and set $d_k = a + (a-b)r_{10^{2k}}$. Then for every $\varepsilon > 0$, there exists $k_0$ such for all $k \geq k_0$ and $\mathbf{x} \in G_k$, there exists $j_{\mathbf{x}}$ satisfying*

*(i)* $d(T^{\ell+a+j_{\mathbf{x}}}\mathbf{x}, T^{\ell+a+\frac{k}{2}r_{10^{2k}}}\mathbf{x}) < \varepsilon$

*(ii)* $d(T^{d_k+\ell+j_{\mathbf{x}}+\frac{k}{2}r_{10^{2k}}}\mathbf{x}, T^{b+\ell}\mathbf{x}) < \varepsilon$

*for all $\ell \in [-r_{10^{2k}-1}, r_{10^{2k}-1}]$. Moreover, for all but a set of such $\mathbf{x}$ with measure at most $\varepsilon$, we have*

$$\Big|\{\ell \in [-r_{10^{2k}-1}, r_{10^{2k}-1}] \colon d(T^{a+\ell}\mathbf{x}, T^{b+\ell}\mathbf{x}) > c\}\Big| > \frac{1}{9} 2r_{10^{2k}-1}. \tag{19}$$

*Proof.* We apply the proof of Corollary 4.7 with $n = d_k$ to obtain the first 2 conditions. More precisely, by construction $G_k$ is a subset of $A_k \cup B_k$ where $A_k$ and $B_k$ given by the proof of Corollary 4.7. Choose $j_{\mathbf{x}} = \sum_{i=0}^{\zeta_{\mathbf{x}}(\frac{k}{2}r_{10^{2k}})-1} \mathbf{1}_{D_{10^{2k}}}(S^i\mathbf{x}) - \frac{k}{2}$. (Recall that $D_j$ is defined in (6).) Then since $x_{10^{2k}-1} = 3$, it follows that $j_{\mathbf{x}} = j_{T^{\ell'}\mathbf{x}}$ for all such $\mathbf{x}$ and $\ell' \in [-\frac{3}{2}r_{10^{2k}-1}, \frac{3}{2}r_{10^{2k}-1}]$. Indeed, because $\mathbf{y} \in D_{10^{2k}}$ implies $y_{10^{2k}-1} = 6$, for all $\mathbf{x}$ with $x_{10^{2k}-1} = 3$ and $\ell' \in [-\frac{3}{2}r_{10^{2k}-1}, \frac{3}{2}r_{10^{2k}-1}]$ we have

$$\sum_{i=0}^{\zeta_{\mathbf{x}}(\frac{k}{2}r_{10^{2k}})-1} \mathbf{1}_{D_{10^{2k}}}(S^i\mathbf{x}) - \sum_{i=0}^{\zeta_{T^{\ell'}\mathbf{x}}(\frac{k}{2}r_{10^{2k}})-1} \mathbf{1}_{D_{10^{2k}}}(S^i\mathbf{x}) = 0.$$

Choosing $k_0$ such that $|a|, |b| < r_{10^{2k_0}-1}$, the first 2 conditions hold. For the final condition, let $V = \{\mathbf{x}\colon d(T^a\mathbf{x}, T^b\mathbf{x}) > c\}$.

By the (mean) ergodic theorem, there exists $N \in \mathbb{N}$ such that

$$\nu\big(\{\mathbf{x}\colon \frac{1}{M}\sum_{i=0}^{M-1} \mathbf{1}_V(T^i\mathbf{x}) > (1-\varepsilon)\nu(V)\}\big) > 1 - \varepsilon$$

for all $M \geq N$. Choosing $r_{10^{2k}-1} > N$ we have (19). □

We now use this to define $\ell_j$ for $j \in \{r+1, \ldots, r + 2\lceil \frac{r}{16} \rceil\}$. Choose $J$ as in Lemma 4.11 and enumerate the elements of the set $J$ as $a_1, \ldots, a_{2\lceil \frac{r}{16} \rceil}$. Let $\ell_{r+j}$ be given by Lemma 4.12 applied with $a = a_j$ and $b = \ell_{\phi(a_j)}$ and where $k$ is chosen larger than the $k_0$ needed for the $2\lceil \frac{r}{16} \rceil$ applications of Lemma 4.12, as well as sufficiently large such that $\ell_i < \frac{1}{2}r_{10^{2k}-1}$ for all $1 \leq i \leq r$. For each such $j$, we define $\ell_{r+j}$ to be the corresponding $d_k$.

We now define $\ell_{r+2\lceil \frac{r}{16} \rceil+1}, \ldots, \ell_{2r}$. Define $\hat{V}_a$ to be the set of $\mathbf{x}$ such that $d(T^{\ell_j}\mathbf{x}, T^{\ell_j + r_{10^{2a+1}}}\mathbf{x}) < \varepsilon$ for all $j \in J^c$. Observe that for all large enough $a$, $\nu(\hat{V}_a) > 1 - \epsilon$. Let $i_1, ..., i_{r-2\lceil \frac{r}{16} \rceil}$ be an enumeration of $J$ and define $\ell_{j+r+2\lceil \frac{r}{16} \rceil} = \ell_{i_j} + r_{10^{2a+1}}$. So $\{\ell_j\}_{j=r+2\lceil \frac{r}{16} \rceil+1}^{2r} = \{\ell_{j'} + r_{10^{2a+1}}\}_{j' \in J^c}$.

We use this to prove the proposition:

*Proof of Proposition 4.4.* Choose $M = \frac{k}{2} r_{10^{2k}}$ and set

$$B = \left[\frac{1}{3}k + 2 \max_{j=1,\dots,2\lceil\frac{r}{16}\rceil}\{|\ell_{a_j}|\}, \frac{1}{2}k - 2 - 2\max_{j=1,\dots,2\lceil\frac{r}{16}\rceil}\{|\ell_{a_j}|\}\right]$$

and

$$A = \{\mathbf{x}\colon x_{10^{2k}-1} = 3 \text{ and } x_{10^{2k}} \in B\} \cap \hat{W} \cap \hat{V}.$$

By Conclusion (i) of Lemma 4.12 and our choice that $|\ell_j| < r_{10^{2k}-1}$, we have Conclusion (d) with $j_{\mathbf{x}}$ as in Lemma 4.12. By Corollary 4.10 (see also the last paragraph of Section 4.2.1) we have Conclusion (a) and Conclusion (b) for $\{\ell_i\}_{i=1}^r$. Since each $\ell_{a_1}, \dots, \ell_{a_{2\lceil\frac{r}{16}\rceil}}$ appears as both $j$ and $\phi(j)$ in our construction of $\{\ell_i\}_{i=r+1}^{r+2\lceil\frac{r}{16}\rceil}$, it follows outside a set of $\mathbf{x}$ with small measure, for each such $\mathbf{x}$ there exist, $p_1, \dots, p_{2\lceil\frac{r}{16}\rceil}$, a reordering of $\ell_i$ for $r < i \leq r + 2\lceil\frac{r}{16}\rceil$ such that $d_Y(T^{\ell_{a_i}}\mathbf{x}, T^{\ell_{p_i}}\mathbf{x}) < \varepsilon$. For the remaining $r < j \leq 2r$, the off diagonal joining $\ell_j$ is built to be $\varepsilon$ close to the corresponding $\ell_i$. Thus Conclusion (b) follows for $\{\ell_i\}_{i=r}^{2r}$. We have (c) for $N = \max\{N_i\}_{i=1}^r$ (see the end of Section 4.2.1). Finally, by (19) and (ii) of Lemma 4.12, we have (e). Indeed, by our choice of $A$ we can apply (ii) and by (19) this gives the desired distance bound of $T^{M+d_k+\ell}\mathbf{x}$ from $T^{d_k+\ell+j_{\mathbf{x}}}\mathbf{x}$. □

4.3. **Proof of Theorem 4.3.** Let $\mathcal{J}_\nu$ denote the self-joinings of $(Y, T, \nu)$. Recall that $\sigma_x$ denotes a measure on $Y$, and not on $\{x\} \times Y$.

**Lemma 4.13.** *Let $\varepsilon > 0$, $k_1, \dots, k_r \in \mathbb{Z}$, $\ell_1, \dots, \ell_{2r}, L, N, M \in \mathbb{Z}$, the set $A$, and $j_{\mathbf{x}} \in [-M, M]$ be as in Proposition 4.4.*

*There exists $\frac{1}{20} > \delta > 0$ such that if for some $\sigma \in \mathcal{J}_\nu$ we have*

$$\nu\Big(\big\{\mathbf{x}\colon d_{\mathcal{M}(Y)}\big(\sigma_{\mathbf{x}}, (\frac{1}{2r}\sum_{n=1}^{2r} J(\ell_n))_{\mathbf{x}}\big) > \delta\big\}\Big) < \delta, \tag{20}$$

*then there exists $\tilde{A} \subset A \subset Y$ and $\nu(\tilde{A}) > \frac{1}{999}$ such that*

*(i)* $\sigma(\{(\mathbf{x}, \mathbf{y}) \in \tilde{A} \times Y\colon d_{\mathcal{M}(Y\times Y)}(\frac{1}{N}\sum_{i=1}^N \delta_{(T^i\mathbf{x}, T^i\mathbf{y})}, \frac{1}{r}\sum_{i=1}^r J(k_i)) < 2\varepsilon\}) > \frac{9}{10}\nu(\tilde{A})$.

*(ii) for all $\mathbf{x} \in \tilde{A}$, there exists $C_{\mathbf{x}}$ with $\sigma_{\mathbf{x}}(C_{\mathbf{x}}) > \frac{1}{99999}$ and*

$$\frac{1}{L}\sum_{i=0}^{L-1} d(T^{M+i}\mathbf{y}, T^{j_{\mathbf{x}}+i}\mathbf{y}) < 2\varepsilon$$

*for all $\mathbf{y} \in C_{\mathbf{x}}$.*

*Moreover, under the additional assumption that there exist $a, b \in \mathbb{N}$ and $c > 0$ such that $d(T^a\mathbf{x}, T^b\mathbf{x}) > 3c$ for a set of $\mathbf{x}$ with measure $\frac{1}{2}$, then there exists $E_{\mathbf{x}}$ with $\sigma_{\mathbf{x}}(E_{\mathbf{x}}) > \frac{1}{99999}$ satisfying*

*(iii)* $\frac{1}{L}|\{0 \leq i < L\colon d(T^{M+i}\mathbf{y}, T^{j_{\mathbf{x}}+i}\mathbf{y}) > \frac{c}{2}\}| > \frac{c}{2}$ *for all $\mathbf{x} \in \tilde{A}$ and $\mathbf{y} \in E_{\mathbf{x}}$.*

*Proof.* Choose a compact set $G$ with $\nu(G) > 1 - \frac{\varepsilon}{10000r}$ such that $T^i|_G$ is (uniformly) continuous for all $|i| \leq \max\{N, L, M\}$. Let $\hat{G} = G \cap \bigcap_{n=1}^{2r} T^{-\ell_n} G$. There exists $\delta > 0$ such that if $\mathbf{x} \in G$ and $d(\mathbf{x}, \mathbf{y}) < \delta$, then

$$d(T^i\mathbf{x}, T^i\mathbf{y}) < \min\{\varepsilon, 10^{-7}\} \quad \text{for all } |i| \leq \max\{N, L, M\}. \tag{21}$$

Thus we can choose $A_1 = A \cap \hat{G}$. If $\mathbf{x} \in A_1$, $\mathbf{y}, \mathbf{y}' \in \hat{G}$, $d(\mathbf{y}, \mathbf{y}') < \delta$, and

$$d_{\mathcal{M}(Y\times Y)}\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{(T^i\mathbf{x},T^i\mathbf{y})}, \frac{1}{r}\sum_{i=1}^{r}J(k_i)\Big) < \varepsilon,$$

then by (21) the definition of $d_{\mathcal{M}(Y\times Y)}$ we have

$$d_{\mathcal{M}(Y\times Y)}\Big(\frac{1}{N}\sum_{i=1}^{N}\delta_{(T^i\mathbf{x},T^i\mathbf{y}')}, \frac{1}{r}\sum_{i=1}^{r}J(k_i)\Big) < 2\varepsilon.$$

Thus Condition (i) follows from Condition (c) of Proposition 4.4 (as well as (20) and the measure bound on $G$). Setting

$$C_{\mathbf{x}} = \hat{G} \cap \bigcup_{n=1}^{r} B(T^{\ell_n}\mathbf{x}, \delta),$$

then (ii) (without the measure bound) follows from (d). Setting

$$E_{\mathbf{x}} = \hat{G} \cap \bigcup_{n=r+1}^{r+2\lceil \frac{r}{16}\rceil} B(T^{\ell_n}\mathbf{x}, \delta),$$

then (iii) (without the measure bound) follows from (e). Let $\tilde{A}$ be the subset of $A_1$ such that (20) holds, $\sigma_{\mathbf{x}}(A_1) > \frac{19}{20} > \frac{9}{10} + \delta$ and so that $\sigma_{\mathbf{x}}(C_{\mathbf{x}})$, $\sigma_{\mathbf{x}}(E_{\mathbf{x}}) > \frac{1}{99999}$. $\square$

Before completing the proof of Theorem 4.3 we note the following. If $(X, T, \mu)$ is an ergodic Borel probability system and $X$ is a compact metric space, then $\mu$ is ergodic if and only if there exists a sequence $N_i \to \infty$ such that for every $f \in C(X)$,

$$\lim_{i\to\infty}\frac{1}{N_i}\sum_{j=0}^{N_i-1} f(T^j x) - \int f\,d\mu = 0 \tag{22}$$

for $\mu$-almost every $x \in X$.

*Proof of Theorem 4.3.* We now produce a joining as close (with respect to $d_{\mathcal{M}(Y\times Y)}$) as desired to $\frac{1}{2}(J(0) + J(1))$, which can thus be assumed to be different from the product joining. Let $k_1^{(1)} = 0$ and $k_2^{(1)} = 1$. We apply Proposition 4.4 with $\varepsilon < \frac{1}{10^{-3}} < \frac{1}{10}\min_{\mathbf{x}\in Y} d(T\mathbf{x}, \mathbf{x})$ to obtain $\ell_1, \dots, \ell_4$, which we denote as $k_1^{(2)}, \dots, k_4^{(2)}$. We also obtain $N_1, L_1, M_1$ and we then apply Lemma 4.13 to obtain $\delta_1$. Applying Proposition 4.4 with $\varepsilon_2 = \min\{\frac{1}{2^2}, \frac{\delta_1}{2}\}$ and $k_1^{(2)}, \dots, k_4^{(2)}$ and we also obtain $N_2, L_2, M_2$. We repeat the application of Lemma 4.13 to obtain $\delta_2$, which without loss of generality we can assume is less than $\frac{\delta_1}{2}$. We repeat this procedure, inductively obtaining $k_1^{(r)}, \dots, k_{2^r}^{(r)}$, which by Part (a) of Proposition 4.4 satisfies

$$\nu\Big(\Big\{\mathbf{x}\colon d_{\mathcal{M}(Y)}\Big(\frac{1}{2^r}\sum_{i=1}^{2^r}J(k_i^{(r)})_{\mathbf{x}}, \frac{1}{2^n}\sum_{i=1}^{2^n}J(k_i^{(n)})_{\mathbf{x}}\Big) > \delta_n\big(1 - \frac{1}{2^{r-n}}\big)\Big\}\Big) < \delta_n\big(1 - \frac{1}{2^{r-n}}\big)$$

for all $n < r$. Applying Lemma 4.13 to obtain $\delta_r$, which again we take to be bounded by $\frac{\delta_n}{2^{r-n}}$ for all $n < r$, we can repeat the application of Proposition 4.4, but with $\varepsilon_{r+1} = \min\{\frac{1}{2^{r+1}}, \frac{1}{2}\delta_r\}$.

We pass to the weak*-limit of $\frac{1}{2^r}\sum_{i=1}^{2^r} J(k_r^{(i)})$, which we denote $\sigma$. By construction, we have that

$$\nu\Big(\big\{\mathbf{x}\colon d_{\mathcal{M}(Y)}\big(\sigma_{\mathbf{x}}, \frac{1}{2^r}\sum_{i=1}^{2^r} J(k_i^{(r)})_x\big) > \delta_r)\big\}\Big) \le \delta_r \tag{23}$$

for all $r \in \mathbb{N}$. From this, it follows via Lemma 4.13 (i) and Proposition 4.4 (c),

$$\sigma\Big\{(\mathbf{x},\mathbf{y}) \in Y \times Y : d_{\mathcal{M}(Y\times Y)}\big(\frac{1}{N_r}\sum_{i=1}^{N_r}\delta_{(T^i\mathbf{x},T^i\mathbf{y})}, \frac{1}{2^{r-1}}\sum_{i=1}^{2^{r-1}} J(k_i^{(r-1)})\big) > \varepsilon_r\Big\}$$
$$< 2\varepsilon_r \le \frac{1}{2^{r-1}}.$$

Thus $\lim\limits_{r\to\infty}\frac{1}{N_r}\sum_{i=1}^{N_r}\delta_{(T^i\mathbf{x},T^i\mathbf{y})}$ is the weak*-limit of $\frac{1}{2^r}\sum_{i=1}^{2^r} J(k_i^{(r)})$, which is $\sigma$, and this holds for $\sigma$-almost every $(\mathbf{x},\mathbf{y}) \in Y \times Y$. By the criterion given in (22), it follows that $\sigma$ is ergodic.

Thus to complete the proof of the theorem, it suffices to show that the assumptions for Theorem 4.1 are satisfied. By (23) we can apply Lemma 4.13 with $\sigma$ and any $\frac{1}{2^r}\sum_{i=1}^{2^r} J(k_i^{(r)})$ that we have produced. This gives us a sequence of sets $\tilde{A}_m$ so that for every $\mathbf{x} \in \tilde{A}_m$ we have sets $C_{\mathbf{x}}^{(m)}, E_{\mathbf{x}}^{(m)}$ such that $\sigma_{\mathbf{x}}(C_{\mathbf{x}}^{(m)}), \sigma_{\mathbf{x}}(E_{\mathbf{x}}^{(m)}) > \frac{1}{99999}$ and

(i) $\frac{1}{L}\sum_{i=0}^{L-1} d(T^{M+i}\mathbf{y}, T^{j_{\mathbf{x}}+i}\mathbf{y}) < 2\varepsilon$ for all $\mathbf{x} \in \tilde{A}_m$ and $\mathbf{y} \in C_{\mathbf{x}}$ giving Theorem 4.1 (ii) as $m \to \infty$ and we can choose $\varepsilon \to 0$. (This uses Proposition 4.4, part (d) and Lemma 4.13, part (ii).)
(ii) $\frac{1}{L}|\{0 \le i < L\colon d(T^{M+i}\mathbf{y}, T^{j_{\mathbf{x}}+i}\mathbf{y}) > \frac{c}{2}\}| > \frac{c}{2}$ for all $\mathbf{x} \in \tilde{A}_m$ and $\mathbf{y} \in E_{\mathbf{x}}$ giving Theorem 4.1 (iii). (This uses Proposition 4.4, part (e) and Lemma 4.13, part (iii).)
(iii) The assumption that $\sigma_{\mathbf{x}}(C_{\mathbf{x}}), \sigma_{\mathbf{x}}(E_{\mathbf{x}}) > \frac{1}{99999}$ giving Theorem 4.1 (iv).
(iv) Proposition 4.4, part (b) applied to $r < d \le r + 2\lceil\frac{r}{16}\rceil$ combined with (20) imply Theorem 4.1 (v).

Note that strictly speaking, $M$, $L$, $C_{\mathbf{x}}$, $E_{\mathbf{x}}$, and $j_{\mathbf{x}}$ depend on $m$, but we omit the dependency for the sake of readability. The choices of $C_{\mathbf{x}}$ and $E_{\mathbf{x}}$ are given by Lemma 4.13 and we have that $M$ and $L$ are $M_m$ and $L_m$ introduced earlier in the proof (which are required input for using Lemma 4.13). Thus we have proven the assumptions needed to apply Theorem 4.1. □

4.4. **Poulsen simplex.** We have assembled the tools to prove the last part of Theorem 1.1, showing that the set of self-joinings of the constructed system form a *Poulsen simplex*, meanig that they form a simplex such that the extreme points are dense.

**Proposition 4.14.** *The self-joinings of the system $(Y, \nu, T)$ form a Poulsen simplex.*

In the proof, we make use of a result of King:

**Theorem 4.15** (King [23, EJCL Theorem])**.** *If $\eta$ is a self-joining of $(Y, \nu, T)$ and $T$ is rigid rank 1, then there exist real numbers $\alpha_i^{(k)} > 0$ such that $\sum_i \alpha_i^{(k)} J(i)$ converges in the weak*-topology to $\eta$.*

Note that in our setting, we can utlize this result, as rigid rank 1 transformations have *flat stacks*. In fact, King establishes that the ergodic self-joinings of transformations with flat stack lie in the weak closures of off diagonal joinings. A different proof of this result is given in [4, Corollary 2.3] (see also [3, Corollary 0.3]).

*Proof of Proposition 4.14.* By King's Theorem, it suffices to show that for any integers $n_1, \dots, n_k$ and positive rationals $\beta_1, \dots, \beta_k$ such that $\sum \beta_i = 1$, there exists $m$ such that $d_{\mathcal{M}(Y\times Y)}(J(m), \sum \beta_i J(n_i)) < \varepsilon$. Without loss of generality, we may assume that all of the rationals have a common denominator, writing $\beta_i = \frac{m_i}{r}$ where all $m_i$ are positive integers. By Corollary 4.10, applied to $n_1, \dots, n_k$ where each $n_i$ appears $m_i$ times, there exists $m$ such that $d_{\mathcal{M}(Y\times Y)}(J(m), \frac{1}{r}\sum_i \sum_{\ell=1}^{m_i} J(a_\ell^{(i)})) < \frac{\varepsilon}{2}$. Thus $d_{\mathcal{M}(Y\times Y)}(J(m), \sum \beta_i J(n_i)) < \varepsilon$. $\square$

### 4.5. **These properties are residual.**

**Theorem 4.16.** *A residual set of measure preserving transformations are not quasi-simple.*

If $(h_j)_{j\in\mathbb{N}}$ is a sequence of positive integers, we say a system $(X, T, \mu)$ *admits special linked approximation of type* $(h_j, h_j + 1)$ if for each $j \in \mathbb{N}$, there exist sets $A_j, C_j \subset X$ satisfying the following five conditions:

(i) $\lim_{j\to\infty} \mu(\bigcup_{i=0}^{h_j-1} T^i A_j) = \frac{1}{2} = \lim_{j\to\infty} \mu(\bigcup_{i=0}^{h_j} T^i C_j)$;

(ii) The sets $A_j, \dots, T^{h_j-1}A_j, C_j, \dots, T^{h_j}C_j$ are pairwise disjoint;

(iii) $\lim_{j\to\infty} \frac{\mu(T^{h_j}A_j\cap A_j)}{\mu(A_j)} = 1 = \lim_{j\to\infty} \frac{\mu(T^{h_j+1}C_j\cap C_j)}{\mu(C_j)}$;

(iv) Defining
$$\mathcal{R}_A^{(j)} = \bigsqcup_{i=0}^{h_j-1} T^i A_j \text{ and } \mathcal{R}_C^{(j)} = \bigsqcup_{i=0}^{h_j} T^i C_j,$$
there exist measurable sets $J_j \subset A_j$ and $a, b \in \mathbb{N}$ such that $J_j, \dots, T^{a+b-1}J_j$ are all pairwise disjoint, $T^i J_j \subset \mathcal{R}_A^{(j)}$ for all $0 \le i \le a-1$, and $T^i J_j \subset \mathcal{R}_C^{(j)}$ for all $a \le i \le a+b-1$ and $\lim_{j\to\infty} \mu(\bigcup_{i=0}^{a+b-1} T^i J_j) = 1$;

(v) For all $\varepsilon > 0$, there exist measurable sets $B_0^{(j)}, \dots, B_{h_j-1}^{(j)}$ and $\hat{B}_0^{(j)}, \dots, \hat{B}_{h_j}^{(j)} \in X$ of diameter at most $\varepsilon$ such that
$$\lim_{j\to\infty} \sum_{i=0}^{h_j-1} \mu(T^i A_j \setminus B_i^{(j)}) = 0 = \lim_{j\to\infty} \sum_{i=0}^{h_j} \mu(T^i C_j \setminus \hat{B}_i^{(j)}).$$

Condition (iv) distinguishes this from usual linked approximation, and is needed to carry out the arguments of Section 4.2. This property is a residual property in the space of measure preserving transformations. Indeed, it is conjugacy invariant, and nonempty. Halmos [16, Theorem 1] showed that the conjugacy class of any aperiodic measure preserving transformation is dense. Our conditions (i)-(v) are the intersection of a countable number of open conditions and so the property holds on a $G_\delta$ set. Thus it is a dense $G_\delta$, that is residual, property.

We say a system $(X, T, \mu)$ is *rigid rank 1* if there exist numbers $n_j$ and sets $I_j$ such that

(i) $\lim_{j\to\infty} \mu(\bigcup_{i=0}^{n_j-1} T^i I_j) = 1$;

(ii) The sets $I_j, \dots, T^{n_j-1}I_j$ are pairwise disjoint;

(iii) $\lim_{j\to\infty} \frac{\mu(T^{n_j} I_j \cap I_j)}{\mu(I_j)} = 1$;

(iv) For all $\varepsilon > 0$, there exist measurable sets $B_0^{(j)}, \ldots, B_{n_j-1}^{(j)} \in X$ of diameter at most $\varepsilon$ such that

$$\lim_{j\to\infty} \sum_{i=0}^{n_j-1} \mu(T^i I_j \setminus B_i^{(j)}) = 0.$$

Note that this property is stronger than being both rigid and rank 1. Similarly to the property of admitting a special linked approximation, rigid rank 1 is also a residual property in the space of measure preserving transformations.

Any transformation that both admits a special linked approximation of type $(h_j, h_{j+1})$ and is rigid rank 1 has a self-joining that is not a distal extension of $(X, T, \mu)$. Indeed, these transformations have the following property: for any pair of integers $a, b \in \mathbb{N}$ and $\varepsilon > 0$, there exists $m \in \mathbb{N}$ and a pair of sets $C, D$ with measure at least $\frac{1}{2} - \varepsilon$ so that

$$\mu(\{x \in C\colon d(T^a x, T^m x) > \varepsilon)\} < \varepsilon \text{ and } \mu(\{x \in D\colon d(T^b x, T^m x) > \varepsilon)\} < \varepsilon.$$

Using this property, rank 1 rigidity, and the ergodicity of $\mu$, our construction of the joining that is not a distal extension of $(X, T, \mu)$ proceeds similarly to Sections 4.2 and 4.3. More precisely, for sufficiently large $j$, we can choose $C = \bigcup_{i=0}^{h_j-1} T^i A_j$, $D = \bigcup_{i=0}^{h_j} T^i B_j$, and $m = a+(a-b)h_j$. The inductive construction of $\ell_1, \ldots, \ell_r$ proceeds verbatim. Similarly for $\ell_{r+1}, \ldots, \ell_{r+2\lceil \frac{r}{16} \rceil}$ is almost verbatim (the described set in Lemma 4.12 is less explicit) and the construction of $\ell_{r+2\lceil \frac{r}{16} \rceil + 1}, \ldots, \ell_{2r}$ is verbatim (making use of the property that our transformation is rigid).

**Remark 3.** Analogously, Proposition 4.14 can be generalized for any rigid rank 1 transformation that admits special linked approximation of type $(h_j, h_j+1)$. Using this, it follows that there is a residual set of measure preserving transformations such that their self-joinings form a Poulsen simplex.

## 5. Coding and results

Sections 5 and 6 are interrelated and technical, and these contain the arguments that rule out an arbitrary, non-trivial factor. We do this by studying the Markov operators. As our system $(Y, \nu, T)$ has many self-joinings, it also has many Markov operators on $L^2(\nu)$. The crux of our argument is that none of these Markov operators can be projections, other than those corresponding to trivial factors. Since each of the possible trivial factors give rise to a Markov operator that is itself a projection gives an indication that our arguments can not be too soft. These arguments are carried out in Section 6, which unfortunately is difficult to summarize at this point, as it rules out non-trivial Markov operators that are not projections by treating three possible cases. The rough idea is that if $T$ has a non-trivial factor $P$ with Markov operator $F$, then there exists a measurable set $A$ such that $\langle F\mathbf{1}_A, F\mathbf{1}_{T^{-1}A}\rangle = 0$ (see Lemma 5.6). The idea of the proof is that if $F'$ is the Markov operator corresponding to a non-trivial self-joining of $T$, then for any measurable set $A$ of positive measure, there exists some iterate $M$ of the operator such that

$$\langle F'^M \mathbf{1}_A, F'^M \mathbf{1}_{T^{-1}A}\rangle > 0. \tag{24}$$

To do this, we relate $F'$ to $\sum \alpha_i U_{T^i}$ (Theorem 5.1 and Corollary 5.2) and define a notion called $i$-friends adapted to this property in Section 5.2, showing that there is some small iterate $i$ of the transformation $T$ such that $\langle F'^M \mathbf{1}_A, F'^M \mathbf{1}_{T^{-i}A}\rangle > 0$. To study this quantity, we relate $\sum \alpha_i U_{T^i}$ to $\sum_{i=-r_N}^{r_N} \beta_i U_{T^i}$ (this is the idea of Section 5.3) for fixed $N$ depending on $A$. We show that there exists some $M$, which depends on the choice of $N$, such that

$$\langle \sum_{i=-r_N}^{r_N} \beta_i U_{T^i}^M \mathbf{1}_A, (\sum_{i=-r_N}^{r_N} \beta_i U_{T^i})^M \mathbf{1}_{T^{-1}A}\rangle > 0.$$

This argument covers the first case of Section 6. Now, though, we can (and do!) choose $\alpha_i$ such that $(\sum \alpha_i U_{T^i})^M$ is close to $F'^M$, but we can not conclude that $\sum_{i=-r_N}^{r_N} \beta_i U_{T^i}^M$ is close to $\sum \alpha_i U_{T^i}$; in particular, the closeness of $\sum_{i=-r_N}^{r_N} \beta_i U_{T^i}$ to $\sum \alpha_i U_{T^i}$ depends on $N$, but $M$ also depends on $N$. However, we can show that if these two quantitites are not close then (24) still holds (this corresponds to cases 2 and 3 of Section 6). Section 5 sets up the machinery for Section 6 and is perhaps even more opaque, though it is be motivated by explanations within that section. As we are only concerned with factors, our results are all stated for Markov operators corresponding to factors. (There are two simplifications in the above description: In reality, we can not just consider $\langle F'^M \mathbf{1}_A, F'^M \mathbf{1}_{T^{-1}A}\rangle$, and instead we must consider $\langle F'^M \mathbf{1}_A, F'^M \mathbf{1}_{T^{-i}A}\rangle$ for $i \in \{1, 2, 3\}$. Additionally we approximate $\sum \alpha_i U_{T^i}$ by $\sum \beta_i \mathbf{1}_{B_i} U_{T^i}$ where $B_i \subset Y$ are cylinders.)

Before starting, we also give a short overview of Section 5. In Section 5.2 we introduce a key definition *i-friends* and Lemma 5.6 whose contrapositive is used to show our system is prime. The application of this lemma (in Section 5.3) makes use of an elaborate inductive definition (see Definition 5.7), characterizing the relation between the transformation $T$ (or some of its small powers) and large powers of $T$. Informally, we call this reducing or the reduction of the power, as it gives us a procedure by which to replace higher powers of $T$ by lower ones. We study this procedure in Lemmas 5.9 and 5.10 and throughout Section 5.4, showing how it is connected to the notion of $n$-friends. This leads to a criterion for our process to be prime, developed in Section 5.5. Namely, using Propositions 5.15 and 5.20, we show that if $T$ has a non-trivial factor, then the inductive procedure given in Definition 5.7 only can produce small errors. We provide additional motivation throughout this section.

5.1. **The mechanism for showing $(Y, \nu, T)$ is prime.** Throughout this section, we continue to assume that $(X, \mu, S)$ and $(Y, \nu, T)$ are the systems defined in Section 3, maintaining all of the notation introduced in that section. The proof that $(Y, \nu, T)$ is prime is based on showing that a factor map is either an isomorphism or a map to the one point system. The first step is relating factor maps to linear combinations of powers of $T$ which holds for any rigid rank 1 transformation:

**Theorem 5.1.** *([4, Theorem 2.2]) If $P$ is a factor map of $T$ and $F$ is the corresponding Markov operator, then $F$ is the limit in the strong operator topology of linear combinations of powers of $U_T$ with non-negative coefficients.*

This theorem is stated in [4] for any Markov operator corresponding to any self-joining of any rigid rank 1 transformation.

**Corollary 5.2.** If $P$ is a factor map of $T$ and $F$ is the corresponding Markov operator, then there exists a sequence of convex combinations $\sum_{i\in\mathbb{Z}} \alpha_i^{(k)}$ satisfying $\sum_{i\in\mathbb{Z}} \alpha_i^{(k)} = 1$ and such that $\sum_{i\in\mathbb{Z}} \alpha_i^{(k)} U_{T^i} \to F$ in the strong operator topology as $k \to \infty$.

*Proof.* The existence of the sequence of $\alpha_i^{(k)}$ without the extra hypothesis that $\sum \alpha_i^{(k)} = 1$ for each $k$ follows from Theorem 5.1. For this last assumption, observe that $F\mathbf{1}_Y = \mathbf{1}_Y$, and so $\lim\limits_{k\to\infty} \sum_i \alpha_i^{(k)} U_{T^i}\mathbf{1}_Y = \mathbf{1}_Y$ and we may assume the (non-negative) coefficients add up to 1. Indeed, because $\alpha_i^{(k)}$ are all positive, $\| \sum_i \alpha_i^{(k)} U_{T^i}\mathbf{1}_Y\| = \sum_i \alpha_i^{(k)}$ and we see that we may assume the $\sum_i \alpha_i^{(k)}$ is a convex combination. $\square$

5.2. **A condition for a factor to be the one point system.** Recall that $Z_k$ and $W_k$ are defined in (3) and (4). Given $n \in \mathbb{N}$, we say that $\mathbf{x}, \mathbf{y} \in Y$ are *n-friends* if

$$\sum_{j=0}^{n} \mathbf{1}_{Z_k}(S^j\mathbf{x}) = \sum_{j=0}^{n} \mathbf{1}_{Z_k}(S^j\mathbf{y})$$

for all but one $k \in \mathbb{N}$,

$$\Big|\sum_{j=0}^{n} \mathbf{1}_{Z_k}(S^j\mathbf{x}) - \sum_{j=0}^{n} \mathbf{1}_{Z_k}(S^j\mathbf{y})\Big| = 1$$

for exactly one $k \in \mathbb{N}$, and

$$\sum_{j=0}^{n} \mathbf{1}_{W_\ell}(S^j\mathbf{x}) = \sum_{j=0}^{n} \mathbf{1}_{W_\ell}(S^j\mathbf{y})$$

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

**Lemma 5.3.** *If* $\mathbf{x}$ *and* $\mathbf{y}$ *are* $\zeta_{\mathbf{x}}(n)$*-friends, then* $0 < |\zeta_{\mathbf{y}}(n) - \zeta_{\mathbf{x}}(n)| \le 3$.

*Proof.* Since $Y = X \setminus \big(\bigcup_{\ell\notin\{10^k\colon k\ge 2\}} Z_\ell \cup \bigcup_{k=1}^{\infty} W_k\big)$ and since $\mathbf{x}$ and $\mathbf{y}$ are $n$-friends, it follows that

$$\Big|\sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_Y(S^j\mathbf{x}) - \sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_Y(S^j\mathbf{y})\Big| = 1.$$

Assume $\sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_Y(S^j\mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_Y(S^j\mathbf{x}) - 1$, and so $\zeta_{\mathbf{y}}(n) = \zeta_{\mathbf{x}}(n) + m$ where $m$ is the least integer such that

$$\sum_{j=1}^{m} \mathbf{1}_Y(S^j S^{\zeta_{\mathbf{x}}(n)}\mathbf{y}) = 1.$$

To prove the statement, we are left with showing that $m \le 3$. If $\mathbf{z} \in X$, $\ell \in \mathbb{Z}$, and $S^\ell\mathbf{z}, S^{\ell+1}\mathbf{z} \notin Y$, then one of the two iterates lies in $Z_1$ (the only $D_\ell$ with 1st index not 6) and the other lies in $\bigcup_{\ell=2}^{\infty} D_\ell$, and so $(S^{\ell+2}\mathbf{z})_1 \notin \{6, 7\}$ which means it lies in $Y$. It follows that $m \le 3$, completing the proof. $\square$

We record part of the proof for future reference:

**Corollary 5.4.** For every $\mathbf{x} \in Y$, $n \in \mathbb{Z}$ we have $|\zeta_{\mathbf{x}}(n)| \le 3|n|$.

**Notation 5.5.** We introduce notation (namely $\mathcal{H}_{n,\varepsilon}$) that is crucial for establishing that $T$ is prime, and is used extensively starting in Section 5.4. To motivate its meaning, sets with $k$-friends play a key role in the proof, being used in Lemma 5.6 to establish a criterion that rules factors not being to the one point system. To do this, we invoke Lusin's Theorem and use that many pairs of friends share their initial entries. We keep this in mind and make the definition precise.

Given $j \in \mathbb{Z}$, we say that 2 disjoint measurable sets $\mathcal{A}_j, \mathcal{B}_j \subset Y$ of equal measure and a measure preserving map $G_j \colon \mathcal{A}_j \to \mathcal{B}_j$ are an $(n,\varepsilon)$*-triple for* $j$ if $\nu(\mathcal{A}_j) = \nu(\mathcal{B}_j) > \varepsilon$, $\mathbf{x}$ and $G_j(\mathbf{x})$ are $\zeta_{\mathbf{x}}(j)$-friends, and $\mathbf{x}_k = G(\mathbf{x})_k$ for all $k \leq n$. (Note this terminology is local and is only used in this definition.) We define:

$$\mathcal{H}_{n,\varepsilon} = \{j\colon \text{ there exists an } (n,\varepsilon)\text{-triple for } j.\} \tag{25}$$

The next lemma is not used until Section 5.5, but as we aim to prove numbers are in $\mathcal{H}_{n,\varepsilon}$ in Section 5.4, and we set up useful definitions to do this in Section 5.3, it is placed here for motivation. In the next lemma we approximate a non-explicit measurable set by cylinders.

**Lemma 5.6.** *Assume that* $(Y,\nu,T)$ *has a non-trivial factor* $(Z,\rho,R)$ *with associated factor map* $P\colon Y \to Z$*. Let* $F\colon L^2(\nu) \to L^2(\nu)$ *be a Markov operator defined by* $P$ *and further assume that* $F$ *is the limit (as* $k \to \infty$*) of* $\sum \alpha_i^{(k)} U_{T^i}$*, in the strong operator topology where* $\alpha_i^{(k)} \geq 0$ *for all* $i,k$ *and* $\sum_i \alpha_i^{(k)} = 1$ *for all* $k$*. Then for all small enough* $\varepsilon > 0$ *there exists* $N_0 = N_0(F,\varepsilon)$ *such that for all* $N \geq N_0$ *and sufficiently large* $m$*,*

$$\sum_{j \in \mathcal{H}_{N,\varepsilon}} \alpha_j^{(m)} < \varepsilon.$$

Note that the proof only uses that the factor is not to a 1 point system, but is phrased this way for consistency with the results in Section 5.5.

*Proof.* Since $T$ is weakly mixing, $R$ is aperiodic and by Rokhlin's Lemma, for any $\delta > 0$, there exists $V \subset Z$ such that $\rho(V) > \frac{1}{4} - \delta$ and such that $V, RV, R^2V, R^3V$ are pairwise disjoint. Set

$$g = \mathbf{1}_V + \sqrt{-1}\mathbf{1}_{RV} - \mathbf{1}_{R^2V} - \sqrt{-1}\mathbf{1}_{R^3V}$$

and let $f = g \circ P$ be the pullback of $g$ to $Y$. Choose $\tilde{f}$, taking values in $\{\sqrt{-1}^j\}_{j=0}^3$, that is a finite linear combination of characteristic functions of cylinder sets such that $\nu(\{\mathbf{x}\colon \tilde{f}(\mathbf{x}) \neq f(\mathbf{x})\}) < \delta$, and let $k$ be the largest defining index out of all of these cylinder sets.

We claim that if $N > k+1$ and $n \in \mathcal{H}_{N,\hat{\varepsilon}}$, then

$$\nu(\{\mathbf{x}\colon |U_{T^n}\tilde{f}(\mathbf{x}) - \tilde{f}(\mathbf{x})| > \frac{1}{\sqrt{2}}\}) > \hat{\varepsilon} - 11\delta.$$

To prove the claim, assume that $G_n \colon \mathcal{A}_n \to \mathcal{B}_n$ is the measure preserving bijection given in the definition of $\mathcal{H}_{n,\varepsilon}$ and define

$$G(\mathbf{x}) = \begin{cases} G_n(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{A}_n \\ G_n^{-1}(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{B}_n \\ \mathbf{x} & \text{otherwise.} \end{cases}$$

We restrict our attention to the set of points $\mathbf{y}$ of measure at least $\varepsilon - 10\delta$ that satisfy the following properties:

(i) the points lie in $\mathcal{A}_n$
(ii) the points satisfy $\tilde{f}(\mathbf{y}) = f(\mathbf{y})$ and $\tilde{f}(G(\mathbf{y})) = f(G(\mathbf{y}))$.
(iii) the points are such that $P(T^n\mathbf{y})$ and $P(G(T^n\mathbf{y}))$ lie in $V \cup RV \cup R^2V \cup R^3V$.

Then for any such point $\mathbf{y}$, we have that $\tilde{f}(\mathbf{y}) = \tilde{f}(G(\mathbf{y}))$ (because $y_i = G(y)_i$ for every $i$ in the defining indices of the cylinders defining $\tilde{f}$) and furthermore for some $1 \le m \le 3$ (which may depend on $\mathbf{y}$) we have

$$\tilde{f}(T^n\mathbf{y}) = f(T^n\mathbf{y}) = \sqrt{-1}^m f(T^nG(\mathbf{y})) = \sqrt{-1}^m \tilde{f}(T^nG(\mathbf{y}))$$

(the second equality follows from Lemma 5.3). Thus either $\tilde{f}(\mathbf{y}) \neq \tilde{f}(T^n\mathbf{y})$ or $\tilde{f}(G(\mathbf{y})) \neq \tilde{f}(T^nG(\mathbf{y}))$. Since $\tilde{f}$ takes values in $\{\sqrt{-1}^j\}_{j=0}^3$, if $\tilde{f}(\mathbf{x}) \neq \tilde{f}(\mathbf{z})$ then $|\tilde{f}(\mathbf{x}) - \tilde{f}(\mathbf{z})| \ge \sqrt{2}$ and the claim follows.

By construction $\|F(\tilde{f}) - \tilde{f}\|_2 < 4\delta$. However, if

$$\|U_{T^{n_j}}\tilde{f} - \tilde{f}\|_2 > c, \ \gamma_j \ge 0, \text{ and } \sum \gamma_j \le 1,$$

then by taking a convex combination it follows that

$$\|\sum \gamma_j U_{T^{n_j}}\tilde{f} - \tilde{f}\|_2 > Cc^2.$$

Similarly if

$$\sum_{\{j:\ \|U_{T^{n_j}}\tilde{f} - \tilde{f}\|_2 > c\}} \gamma_j > \varepsilon,$$

then $\|\sum \gamma_j U_{T^{n_j}}\tilde{f} - \tilde{f}\|_2 > C\varepsilon^2c^2$. Since $\delta$ is arbitrary, the lemma follows. □

5.3. **Recoding of time scales.** This section is devoted to relating a large power to $T$ to a smaller power, or perhaps several smaller powers with accompanying subsets of $Y$. This procedure, which we call reducing or the reductions, is carried out via Definition 5.7, which also contains a parameter for testing how good this relation is. The defect of it is related to the notion of $n$-friends in the next section. To carry out the reduction, the next definition is a mechanism for computing the "order of magnitude" of the relevant power of $T$.

**Notation.** Let

$$\sigma_n = \max\{i\colon d_i(n) \neq 0\}, \tag{26}$$

where $d_i$ is defined as in (10).

Set

$$E = \{10^k\colon k \ge 2\}. \tag{27}$$

Now if $\sigma_m \notin E$ it is relatively easy to see that $m \in \mathcal{H}_{\sigma_m - 1,\varepsilon}$ (Lemma 5.11). If $\sigma_m \in E$ we seek to relate $m$ to $m'$ where $\sigma_{m'} < \sigma_m$. We now consider two motivating examples: $T^{r_{10^{2k+1}}+2} \approx T^2$, because off of a small measure set $(T^{r_{10^{2k+1}}+2}\mathbf{x})_j = (T^2\mathbf{x})_j$ for all $j < 10^{2k+1}$. There is a more complicated situation, $T^{r_{10^{2k}}+2}$ is roughly $T^2$ on $\{\mathbf{y} \in Y\colon \mathbf{y}_{10^{2k}} < k\}$ (off of a set of small measure) and $T^{r_{10^{2k}}+2}$ is roughly $T$ on $\{\mathbf{y} \in Y\colon \mathbf{y}_{10^{2k}} \ge k\}$ (off of a set of small measure). Note that $\sigma(r_{10^{2k+1}} + 2) = 10^{2k+1}$ and similarly for the other powers. We make this recoding precise below by triples which keep track of the new powers in the first coordinate, the set where this approximation is relevant in the second coordinate and the measure of the set where this approximation fails in the third coordinate. Note, the third coordinate can also be related to friends (Lemma 5.12). The next definition defines an inductive procedure, and the relevant initial conditions are deferred until Definition 5.8.

**Definition 5.7.** *Fix $N \in \mathbb{N}$ and $\varepsilon > 0$. For $r \geq 1$ and a set of triples $\mathfrak{H}_r(N,\varepsilon) \subset \mathbb{Z} \times \mathcal{B}(Y) \times [0,1]$, we inductively define the set of triples $\mathfrak{H}_{r+1}(N,\varepsilon)$ as follows: if $(j, A, \rho) \in \mathfrak{H}_r(N,\varepsilon)$ and at least one of the following conditions holds*

(i) $\sigma_j \notin E$
(ii) $j = 0$
(iii) $\sigma_j \leq N$
(iv) $\rho > \varepsilon$,

*then $(j, A, \rho) \in \mathfrak{H}_{r+1}(N,\varepsilon)$. Otherwise we modify the triple, depending on the value of $\sigma_j$:*

(i) *If $\sigma_j \in \{10^{2\ell+1} \colon \ell \geq 1\}$, then*

$$(j - d_{\sigma_j}(j) r_j, A, \rho + \frac{|d_{\sigma_j}(j)|}{a_{\sigma_j}}) \in \mathfrak{H}_{r+1}(N,\varepsilon).$$

(ii) *If $\sigma_j \in \{10^{2\ell} \colon \ell \geq 1\}$, then both*

$$(j - d_{\sigma_j}(j) r_j, A \cap \bigcup_{\ell < \frac{a_{\sigma_j}}{2}} \mathcal{C}_{\sigma_j}(\ell), \rho + \frac{|d_{\sigma_j}(j)|}{a_{\sigma_j}}) \in \mathfrak{H}_{r+1}(N,\varepsilon)$$

*and*

$$(j - d_{\sigma_j}(j) r_j + d_{\sigma_j}(j), A \cap \bigcup_{\ell \geq \frac{a_{\sigma_j}}{2}} \mathcal{C}_{\sigma_j}(\ell), \rho + \frac{|d_{\sigma_j}(j)|}{a_{\sigma_j}}) \in \mathfrak{H}_{r+1}(N,\varepsilon).$$

**Definition 5.8.** *Fix $N \in \mathbb{N}$ and $\varepsilon > 0$.*

*Define $\mathfrak{F}(N,\varepsilon)$ to be the collection of triples $\mathfrak{H}_r(N,\varepsilon)$ that stabilize with respect to $r$, meaning that*

$$\mathfrak{F}(N,\varepsilon) = \mathfrak{H}_r(N,\varepsilon) \text{ when } \mathfrak{H}_r(N,\varepsilon) = \mathfrak{H}_{r+1}(N,\varepsilon).$$

*Define $\mathcal{F}(N,\varepsilon)$ to keep track of the measure of the sets in $\mathfrak{F}(N,\varepsilon)$, meaning that*

$$\mathcal{F}(N,\varepsilon) = \{(n, \nu(A), \rho) \colon (n, A, \rho) \in \mathfrak{F}(N,\varepsilon)\}.$$

*If $\mathfrak{H}_0(N,\varepsilon) = (i, [0,1], 0)$ for some $i \in \mathbb{N}$, we define $\mathcal{F}_i(N,\varepsilon)$ to be the set $\mathcal{F}(N,\varepsilon)$.*

*Similarly, define $\mathfrak{H}_{r,i}(N,\varepsilon)$ to be $\mathfrak{H}_r(N,\varepsilon)$ if $\mathfrak{H}_0(N,\varepsilon) = \{(i, [0,1], 0)\}$. We similarly define $\mathfrak{F}_i(N,\varepsilon)$ to be $\mathfrak{F}(N,\varepsilon)$ when $\mathfrak{H}_0(N,\varepsilon) = \{(i, [0,1], 0)\}$.*

Note that $\sigma_n$ is defined in (26), $a_i$ are defined in (1), and $r_i$ are defined in (9). We state a lemma that motivates the sets given in Definition (5.8):

**Lemma 5.9.** *Given $n \in \mathbb{N}$, let $\mathcal{C}$ be a cylinder defined by positions in $E$ that are greater than $\sigma_n$.*

(i) *Assume $\sigma_n$ is an odd power of 10. Setting $\tilde{n} = n - d_{\sigma_n}(n) r_{\sigma_n}$, we have*

$$\nu(\{x \in \mathcal{C} \colon (T^{\tilde{n}}\mathbf{x})_i \neq T^n(\mathbf{x})_i \text{ for some } i < \sigma_n\}) < 4\nu(\mathcal{C}) \frac{|d_{\sigma_n}(n)|}{a_{\sigma_n}}.$$

(ii) *Assume $\sigma_n$ is an even power of 10. Setting $n' = n - d_{\sigma_n}(n)(r_{\sigma_n} - 1)$ and defining $A_1 = \{\mathbf{x} \colon x_{\sigma_n} \geq \frac{\sigma_n}{2}\}$, we have*

$$\nu(\{\mathbf{x} \in A_1 \colon (T^{n'}\mathbf{x})_i \neq T^n(\mathbf{x})_i \text{ for some } i < \sigma_{n'}\}) < 4\nu(\mathcal{C} \cap A_1) \frac{|d_{\sigma_n}(n)|}{a_{\sigma_n}}.$$

*Furthermore, setting $n'' = n - d_{\sigma_n}(n) r_{\sigma_n}$, and defining $A_2 = \{\mathbf{x} \colon x_{\sigma_n} < \frac{\sigma_n}{2}\}$, we have*

$$\nu(\{\mathbf{x} \in A_2 \colon (T^{n''}\mathbf{x})_i \neq T^n(\mathbf{x})_i \textit{ for some } i < \sigma_{n''}\}) < 4\nu(\mathcal{C} \cap A_2) \frac{|d_{\sigma_n}(n)|}{a_{\sigma_n}}.$$

*Proof.* For convenience, in this proof we assume $d_{\sigma_n} > 0$ (the case $d_{\sigma_n} < 0$ is similar). Recall that $D_j$ is defined in (6). Observe that if $T^n(\mathbf{x}) = S^{d_{\sigma_n}(n) q_{\sigma_n}}(T^{\tilde{n}}\mathbf{x})$, then $T^n(\mathbf{x})_j = T^{\tilde{n}}(\mathbf{x})_j$ for all $j < \sigma_n$, and (by Lemma 3.3) this holds if

$$\sum_{i=0}^{d_{\sigma_n}(n) q_{\sigma_n} - 1} \mathbf{1}_{\bigcup_{j=\sigma_n}^{\infty} D_j}(S^i \mathbf{x}) = 0. \tag{28}$$

First we consider the case of $\sigma_n = 10^j$ for $j$ odd.

Since $\sigma_n \in \{10^{2k+1} \colon k \geq 1\}$, we have that $D_{\sigma_n} = \emptyset$ and so (28) fails for a set of $\mathbf{x}$ of $\mu$ measure at most

$$\frac{d_{\sigma_n}(n)}{a_{\sigma_n}} q_{\sigma_n} \mu(\bigcup_{j=\sigma_n+1}^{\infty} D_j) \leq \frac{d_{\sigma_n}(n)}{\log_{10}(\sigma_n)}. \tag{29}$$

Furthermore,

$$\nu(\{\mathbf{x} \in Y \colon \sum_{i=0}^{d_{\sigma_n}(n) q_{\sigma_n} - 1} \mathbf{1}_{\bigcup_{j=\sigma_n}^{\infty} D_j}(S^i \mathbf{x}) \neq 0\}) \leq 3 d_{\sigma_n}(n) q_n \mu(Z_{\sigma_{n+1}}) = 3 \cdot \frac{1}{8} \frac{d_{\sigma_n}(n)}{a_{\sigma_n}}.$$

(Restricting to $\mathbf{x} \in Y$ and converting from $\mu$ to $\nu$ changes this by a factor of less than 3.)

The next two cases are similar, but a bit more complicated as $D_{\sigma_n}$ is not empty, but is equal to $W_\ell$ for some $\ell$. If $\mathbf{x} \in A_1$, then the conclusion holds if

$$\sum_{i=0}^{d_{\sigma_n}(n) q_{\sigma_n} - 1} \mathbf{1}_{\bigcup_{j=\sigma_n}^{\infty} D_j}(S^i T^{n'} \mathbf{x}) = 0.$$

Indeed, if $\sum_{i=0}^{q_{\sigma_n}-1} \mathbf{1}_{\bigcup_{j=\sigma_n}^{\infty} D_j}(S^i \mathbf{y}) = 0$, then this follows from Lemma 3.3 and the fact that $\sum_{i=0}^{q_{\sigma_n}-1} \mathbf{1}_{\bigcup_{j=\sigma_n}^{\infty} D_j}(S^i \mathbf{0}) = 1$. So,

$$T^{r_{\sigma_n}}(\mathbf{y})_j = \begin{cases} (T^{-1}\mathbf{y})_j & \text{for } j \neq \sigma_n \\ (T^{-1}\mathbf{y})_j + 1 & \text{for } j < \sigma_n. \end{cases}$$

Thus this case follows analogously to (29) above after estimating

$$\mu(\{\mathbf{y} \in A_1 \colon S^i \mathbf{y} \in A_1 \setminus (\cup_{j=\sigma_n}^{\infty} D_j) = A_1 \setminus (\cup_{j=\sigma_n+1}^{\infty} D_j) \text{ for all } i \leq d_{\sigma_n}(n) q_{\sigma_n}\}).$$

This is at most $\frac{d_{\sigma_n}(n)}{a_{\sigma_n}}$.

The third case is similar: if $\mathbf{x} \in A_2$, then the conclusion holds if

$$\sum_{i=0}^{d_{\sigma_n}(n) q_{\sigma_n} - 1} \mathbf{1}_{\bigcup_{j=\sigma_n+1}^{\infty} D_j}(S^i \mathbf{x}) = 0$$

and

$$\sum_{i=0}^{d_{\sigma_n}(n) q_{\sigma_n} - 1} \mathbf{1}_{D_{\sigma_n}}(S^i \mathbf{x}) = d_{\sigma_n}(n),$$

where $10^{2\ell} = \sigma_n$. The remainder of the proof is analogous to the first case. $\square$

Motivated by the sets in Lemma 5.9, we make a few more definitions. If $(n, A, \rho) \in \mathfrak{H}_{r,i}(N, \varepsilon)$, let

$$P_r(n, A, \rho) = \{\mathbf{x} \in A \colon (T^n\mathbf{x})_j \neq (T^i\mathbf{x})_j \text{ for some } j \leq \sigma_n\}$$

and

$$Q_r(n, A, \rho) = \{\mathbf{x} \in A \colon (T^n\mathbf{x})_j = (T^i\mathbf{x})_j \text{ for all } j \leq \sigma_n\}.$$

Define

$$\mathcal{P}_r = \bigcup_{(n,A,\rho)\in\mathfrak{H}_r(N,\varepsilon)} P_r(n, A, \rho) \tag{30}$$

and

$$\mathcal{Q}_r = \bigcup_{(n,A,\rho)\in\mathfrak{H}_r(N,\varepsilon)} Q_r(n, A, \rho). \tag{31}$$

**Lemma 5.10.** *Assume $\sigma_i \in E$ and let $A$ be a cylinder with all defining indices at least $\sigma_i$. Let $\mathfrak{H}_0(N, \varepsilon) = \{(i, [0, 1], 0)\}$. There exist cylinders $C_1, \dots, C_\ell$ defined in positions greater than or equal to $\sigma_i$ such that the following hold:*

*(i)* $A \cap \mathcal{P}_1 \subset \bigcup_{j=1}^{\ell} C_j$.
*(ii)* $99\nu(A \cap \mathcal{P}_1) > \nu(\bigcup_{j=1}^{\ell} C_j)$.

*Proof.* We treat $i$ with $\sigma_i \in \{10^{2k+1}\}$. Consider the set of $\mathbf{y} \in Y$ such that (28) fails. We cover this set by cylinders and show that $\nu(\mathcal{P})$ is proportional to the union of these cylinders. The set $D_\ell$ requires that $x_j = a_{j-2}$ for all $j < \ell$, and so $S^{-d_{\sigma_i}(i)q_{\sigma_i}}(\bigcup_{\ell=\sigma_i+1}^{\infty} D_\ell)$ is contained in at most $d_{\sigma_i}(\sigma_i)+1$ cylinders defined by the position $\sigma_i$. Furthermore,

$$\mathcal{P} \supset \{\mathbf{y} \in Y \colon \sum_{j=0}^{d_{\sigma_i}(i)q_{\sigma_i}-1} \mathbf{1}_{D_{\sigma_i+1}}(S^j\mathbf{y}) = 1 \text{ and } \sum_{j=0}^{n_{\mathbf{y}}-1} \mathbf{1}_{D_{\sigma_i+1}}(S^j\mathbf{y}) = 0\}$$

where $n_{\mathbf{y}}$ is the first coordinate of $(n_{\mathbf{y}}, B, \rho) \in \mathfrak{H}_1(N, \varepsilon)$ and $\mathbf{y} \in B$. This set has measure at least $(\frac{1}{8} - \frac{1}{10^3})\frac{d_{\sigma_i}(i)}{a_{\sigma_i}}$.

The argument for $i$ with $\sigma_i \in \{10^{2k}\}$ is similar, but slightly complicated analogously to the proof of Lemma 5.9, because $D_{\sigma_i} = W_{\frac{1}{2}\log_{10}(\sigma_i)}$. □

5.4. **Obtaining friends.** The motivation for this section is that imperfections in the reduction process are related to the presence of $n$-friends. Indeed, we show the parameter $\rho$ in Definition 5.7 is proportional to the measure of a set of points that have friends. The strategy is similar to the proof of Proposition 3.4. However, we use a more "pointwise approach." Rather than having two sets, $A_i$, $B_i$ such that $\xi_{\mathbf{z}}(i)$ is constant on each set and $\xi_{\mathbf{y}}(i) - \xi_{\mathbf{x}}(i) = 1$ for all $(\mathbf{y}, \mathbf{x}) \in B_i \times A_i$, we define a set $A$ and a map $G$ so that $\mathbf{x}$ and $G(\mathbf{x})$ are $\zeta_{\mathbf{x}}(m)-$friends for all $\mathbf{x} \in A$. In particular, we do not claim $\zeta_{\mathbf{z}}(m)$ is well behaved as $\mathbf{z}$ varies in $A$. (Recall that $\xi$ and $\zeta$ are morally "inverses" of each other and while $\xi$ was more convenient for the proof of Proposition 3.4, $\zeta$ is more convenient here and in the remainder of the proof.)

**Lemma 5.11.** *If $\sigma_m \notin E$, then $m \in \mathcal{H}_{\sigma_m-1,\frac{1}{99}}$. Furthermore, if $G \colon \mathcal{A}_m \to \mathcal{B}_m$ is the measure preserving bijection associated to $\mathcal{H}_{\sigma_m-1,\frac{1}{99}}$ as defined in (25), then $\mathcal{A}_m$ and $G(\mathcal{A}_m)$ can be chosen to be a union of cylinders whose defining indices are a subset of $\sigma_m - 1$, $\sigma_m$, and $\sigma_m + 1$.*

*Proof.* Assume $\sigma_m \notin E$ and set $k = \sigma_m$. Recall that $d_k = d_k(m)$ is defined in (10). Assume that $d_k \in \{1, 2, 3, 4\}$ (the case that $d_k \in \{-1, -2, -3, -4\}$ is analogous). Set $x_k = 0$, $x_{k-1} = 5$, and $x_{k+1} = 4$ for whichever of $k-1$ and $k+1$ do not lie in $E$. Whenever $k-1$ or $k+1$ lies in $E$, we stipulate that $x_{k-1}$ or $x_{k+1} \in (\frac{a_{k-1}}{2}, a_{k-1}-3)$. Set $y_k = 7$ and $x_j = y_j$ for all other $j$. We claim that if $\mathbf{x}, \mathbf{y} \in \mathbf{Y}$ are as above, then they are $\zeta_{\mathbf{x}} = \zeta_{\mathbf{x}}(m)$-friends. We first check that

$$\sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_\ell}(S^j \mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_\ell}(S^j \mathbf{x})$$

for all $\ell < k$. To see this, note that the inclusion $S^j \mathbf{z} \in Z_\ell$ depends only on $z_1, \dots, z_\ell$ and we have that $x_j = y_j$ for all $j < k$. Likewise if $10^{2\ell} < k$, then $\sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{W_\ell}(S^j \mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{W_\ell}(S^j \mathbf{x})$. Also note that because $S^j(\mathbf{x})_{k+1}, S^j(\mathbf{y})_{k+1} \neq a_{k+1} - 2$ for all $j \leq \zeta_{\mathbf{x}}$, we have

$$\sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_\ell}(S^j \mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_\ell}(S^j \mathbf{x}) = 0$$

for all $\ell > k + 1$ and

$$\sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{W_\ell}(S^j \mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{W_\ell}(S^j \mathbf{x}) = 0$$

for all $10^{2\ell} > k + 1$. Now, if $k + 1 \notin E$, then since $(S^j \mathbf{x})_{k+1}, (S^j \mathbf{y})_{k+1} \neq a_{k+1} - 1$ we have

$$\sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_{k+1}}(S^j \mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_{k+1}}(S^j \mathbf{x}) = 0.$$

If $k + 1 \in E$, then since $y_{k+1} = x_{k+1} > \frac{a_{k+1}}{2}$, we have

$$\sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{W_\ell}(S^j \mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{W_\ell}(S^j \mathbf{x}) = 0,$$

where $10^\ell = k + 1$.

Lastly, since $\zeta_{\mathbf{x}} > \frac{5}{8} a_k q_{k-1}$, we have that by the condition on the digits $k$ and $k - 1$ of $\mathbf{y}$,

$$\sum_{j=0}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_k}(S^j \mathbf{y}) = 1.$$

But since $\zeta_{\mathbf{x}} < 5q_k$, using that $x_k = 1$ we have that $(S^j \mathbf{x})_k < 7$ for all $0 \leq j \leq \zeta_{\mathbf{x}}$ and so $\sum_{j=1}^{\zeta_{\mathbf{x}}} \mathbf{1}_{Z_k}(S^j \mathbf{x}) = 0$. This proves the claim that $\mathbf{x}$ and $\mathbf{y}$ are $\zeta_{\mathbf{x}} = \zeta_{\mathbf{x}}(m)$-friends and $G$ is the bijection taking $\mathbf{x}$ to $\mathbf{y}$. (That is, changing the $k^{\text{th}}$ entry from 0 to 7.) □

**Lemma 5.12.** *If $\sigma_m \in E$, then there exist cylinder sets $K_1, \dots, K_r$ defined on the entries $\sigma_m + 1$, $\sigma_m$, and $\sigma_m - 1$ such that*

$$\nu\left(\bigcup_{j=1}^{r} K_j\right) > \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{|d_{\sigma_m}(m)|}{a_{\sigma_m}} \cdot \frac{1}{64} \tag{32}$$

*and there exists a measure preserving map* $G\colon \bigcup_{j=1}^r K_j \to Y \setminus \bigcup_{j=1}^r K_j$ *defined by changing the* $\sigma_m + 1$ *entry such that if* $\mathbf{x} \in \bigcup_{j=1}^r K_j$*, then* $\mathbf{x}$ *and* $G(\mathbf{x})$ *are* $\zeta_{\mathbf{x}}(m)$*-friends. Moreover,* $K_1, \dots K_r, G(K_1), \dots G(K_r)$ *are disjoint cylinders.*

*Proof.* Let $10^k = \sigma_m$ and assume that $d_{10^k}(m) > 0$ (the case that $d_{10^k}(m) < 0$ is similar). Let $x_{10^k+1} = 0$ and $y_{10^k+1} = 7$. Let

$$y_{10^k} \in \{k-2, k-3, \dots, k-1-\min\{d_{10^k}(m), \frac{1}{3}a_k\}\}$$

and set $y_{10^k-1} = 5$. Furthermore, set $x_\ell = y_\ell$ for all $\ell \neq 10^k+1$. It is straightforward that $\sum_{j=0}^{\zeta_{\mathbf{x}}(m)} \mathbf{1}_{Z_{10^k+1}}(S^j\mathbf{x}) = 0$ and $\sum_{j=0}^{\zeta_{\mathbf{x}}(m)} \mathbf{1}_{Z_{10^k+1}}(S^j\mathbf{y}) = 1$.

We claim that $\mathbf{1}_V(S^j\mathbf{x}) = \mathbf{1}_V(S^j\mathbf{y})$ for all $|j| \leq \zeta_{\mathbf{x}}(m)$, where $V$ is either $Z_\ell$ for $\ell \neq 10^k+1$ or $V$ is any $W_\ell$. To see this, for $Z_\ell$ with $\ell < 10^k$ and $W_\ell$ with $\ell \leq k$, this holds since $\mathbf{y}$ and $\mathbf{x}$ agree in the relevant entries. Furthermore, $(S^j\mathbf{x})_{10^k+1}$ and $(S^j\mathbf{y})_{10^k+1}$ are not 6 in this range and so we never land in $Z_\ell$ for $\ell > 10^k+1$ or in $W_\ell$ for $\ell > k$, proving the claim. Assuming $K_j$ is a cylinder set as in the statement, define $G(\mathbf{x})$ to change the $10^k+1$ entry from 0 to 7 (leaving all the other entries unchanged). Thus $G$ satisfies all of the announced properties.

Now

$$\mu(\{\mathbf{x}\colon x_{10^k} \in \{k-2, k-3, \dots, k-1-\min\{d_{10^k}(m), \frac{1}{3}a_k\}\},$$
$$x_{10^k+1} = 0, \text{ and } x_{10^k-1} = 5\}) \geq \frac{1}{3} \cdot \frac{|d_{\sigma_m}(m)|}{k} \cdot \frac{1}{64}.$$

Considering the set of such $\mathbf{x} \in Y$ so that $\mathbf{y} \in Y$ as well with $x_i = y_i$ for all $i \neq \sigma_m + 1 = 10^k+1$ and $y_{10^k+1} = 7$ and (trivially) converting to $\nu$ establishes (32). □

**Lemma 5.13.** *Assume* $r \geq 1$, $\varepsilon < \frac{1}{8\cdot 99}$, $\mathfrak{H}_0(N,\varepsilon) = \{(i,[0,1],0)\}$, $(n,A,\rho) \in \mathfrak{H}_r(N,\varepsilon)$ *and* $\sigma_n \in E$. *We can choose* $B_1, \dots, B_\ell \subset A$ *to be cylinders whose defining indices are at least* $\sigma_{n-1}$ *such that* $\nu(\bigcup_{j=1}^\ell B_j) > \frac{1}{6\cdot 64} \cdot \frac{|d_{\sigma_n}(n)|}{a_{\sigma_n}}$ *and such that there exists* $\tilde{B} \subset \bigcup_{j=1}^\ell B_j$ *with*

$$\nu(\tilde{B}) \geq (1 - 4\cdot 99\rho)\nu(\bigcup_{j=1}^{\ell} B_j), \tag{33}$$

$G\colon \tilde{B} \to Y$ *a measure preserving injection, defined by changing the* $\sigma_n + 1$ *position, and thus* $G(\tilde{B}) \subset A$*, such that* $\mathbf{x}$ *and* $G(\mathbf{x})$ *are* $\zeta_{\mathbf{x}}(i)$*-friends and* $x_j = G(\mathbf{x})_j$ *for all* $j < \sigma_n$.

*Proof.* Let $\hat{B}_1, \dots, \hat{B}_\ell$ be the cylinders and $\hat{G}$ be the function given by Lemma 5.12 applied with $m = n$. Set $B_i = \hat{B}_i \cap A$ and let $\tilde{B}$ be the set of points in $\bigcup_{j=1}^\ell B_j \cap \mathcal{Q}_r \cap G^{-1}(\mathcal{Q}_r)$. Let $(n', A, \rho') \in \mathfrak{H}_{r-1}(N,\varepsilon)$ be the predecessor of $(n,A,\rho)$. Since $\sigma_{n'} > \sigma_n + 1$, we have that if $\mathbf{x} \in \tilde{B}$ and $\mathbf{x}, \mathbf{y} \in \mathcal{Q}_r$ then for all $k$, we claim that

$$\sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_{Z_k}(S^j\mathbf{x}) - \sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_{Z_k}(S^j\mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}(i)} \mathbf{1}_{Z_k}(S^j\mathbf{x}) - \sum_{j=0}^{\zeta_{\mathbf{x}}(i)} \mathbf{1}_{Z_k}(S^j\mathbf{y}). \tag{34}$$

We first consider the case of $k < \sigma_{n'}$. The sums on the left hand side of (34) are either $\lfloor \frac{n}{q_k} \rfloor$ or $\lceil \frac{n}{q_k} \rceil$, while on the right hand side they are either $\lfloor \frac{i}{q_k} \rfloor$ or $\lceil \frac{i}{q_k} \rceil$; the choice of $\lfloor\cdot\rfloor$ or $\lfloor\cdot\rfloor + 1 = \lceil\cdot\rceil$ depends on comparing $x_j$ and $(S^n\mathbf{x})_j$ for the left sums on each side, and similarly $y_j$ and $(S^n\mathbf{y})_j$ for the right sums on each side, for $j \le k$. By our assumption that $\mathbf{x}, \mathbf{y} \in \mathcal{Q}_r$, we have that $(S^n\mathbf{x})_j$ is the same as $(S^i\mathbf{x})_j$ for all $j \le k < n'$ and so whether the first sum on the left hand side is the floor function or one more is the same for the first sum on the right hand side. The case of $\mathbf{y}$ is identical.

Next consider the case of $k > \sigma_{n+1}$ (since $\sigma_n + 1 < \sigma_{n'}$ this covers $k \ge \sigma_{n'}$). We have that $\sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_{Z_k}(S^j\mathbf{x}) = \lfloor\frac{\zeta_{\mathbf{x}}(n)}{q_k}\rfloor$ and $\sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_{Z_k}(S^j\mathbf{y}) = \lfloor\frac{\zeta_{\mathbf{y}}(n)}{q_k}\rfloor$ by the argument in Lemma 5.12. (This argument only depends on the cylinders with defining entries in positions $\sigma_n - 1$, $\sigma_n$, and $\sigma_n + 1$ that define the cylinders in the proof of Lemma 5.12 and these entries are the same for $\hat{B}_i$.) For the right hand side, as above $(S^n\mathbf{x})_j = (S^i\mathbf{y})_j$ for $j \in \{n, n+1\}$, so whether we take the floor or ceiling in the summands on the right hand side depends on $j > n+1$. These are the same for $\mathbf{x}$ and $\mathbf{y}$ by construction, giving (34). So in the left hand side both summands take the floor and for the right hand side they either both take the floor or both take the ceiling, establishing (34).

A similar computation yields

$$\sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_{W_\ell}(S^j\mathbf{x}) - \sum_{j=0}^{\zeta_{\mathbf{x}}(n)} \mathbf{1}_{W_\ell}(S^j\mathbf{y}) = \sum_{j=0}^{\zeta_{\mathbf{x}}(i)} \mathbf{1}_{W_\ell}(S^j\mathbf{x}) - \sum_{j=0}^{\zeta_{\mathbf{x}}(i)} \mathbf{1}_{W_\ell}(S^j\mathbf{y}). \tag{35}$$

To complete the lemma we are left with establishing (33). To check this, we claim that it suffices to show that $\mathcal{P}_r$ can be chosen to be unions of cylinders defined by entries with positions at least $10^{\log_{10}(\sigma_n)+1}$. This follows from the following:

**Claim.** For all $\delta > 0$, there exists $k \in \mathbb{N}$ such that if $C_1, C_2$ are cylinders with the smallest entry defining $C_2$ at least $k$ larger than the largest entry defining $C_1$, then

$$\frac{\nu(C_1 \cap C_2)}{\nu(C_1)\nu(C_2)} \in [1-\delta, 1+\delta]. \tag{36}$$

*Proof of Claim:* To check that the claim holds, let $L$ be the smallest entry defining $C_2$. Let $U_1, \ldots, U_m$ be the cylinders given by proscribing the first $L-1$ terms that intersect $Y$. All but one of these cylinders are also a cylinders in $X$, and so they all have the same $\nu$ measure. If $U_i$ is the one cylinder set in $Y$ that is not also a cylinder set in $X$, then $U_i$ has smaller $\nu$ measure than the other $m-1$ cylinders. Assume $U_1, \ldots, U_{m'}$ are those cylinders that are contained in $C_1$. If $i \in \{1, \ldots, m'\}$, then $\nu(C_1 \cap C_2) < \nu(C_1)\nu(C_2)$, but it is at least $\frac{m'-1}{m'}\nu(C_1)\nu(C_2)$. Similarly, if $i \notin \{1, \ldots, m'\}$, then $\nu(C_1 \cap C_2) < \nu(C_1)\nu(C_2)$, but it is at most $\frac{m}{m-1}\nu(C_1)\nu(C_2)$. Since $\tilde{B}$ is a union of the sets $B_i$ that pairwise satisfy this condition, the claim follows. □

The sufficient condition, that $\mathcal{P}_r$ can be chosen to be unions of cylinders defined by entries with positions at least $10^{\log_{10}(\sigma_n)+1}$, then follows by iterating Lemma 5.10 and our assumption that $\varepsilon < \frac{1}{8 \cdot 99}$ and so $\nu(\mathcal{P}_r) > \frac{1}{2}$; indeed, by Lemmas 5.9 and 5.10 at least $\nu(A \cap \mathcal{P}_r) > \nu(A)(1 - 99 \cdot 4 \cdot \frac{1}{8 \cdot 99})$. □

**Lemma 5.14.** *Assume $r \ge 1$, $\varepsilon < \frac{1}{8\cdot 99}$, $\mathfrak{H}_0(N,\varepsilon) = \{(i, [0,1], 0)\}$, $(n, A, \rho) \in \mathfrak{H}_r(N, \varepsilon)$, and $\sigma_n \notin E$. Let $(n', A', \rho') \in \mathfrak{H}_{r-1}(N,\varepsilon)$ so that $A \subset A'$. Then there exist cylinders $B_1, \ldots, B_\ell \subset A'$ defined by positions whose entries are at least $\sigma_n - 2$, such that there exists $\tilde{B} \subset \bigcup_{j=1}^{\ell} B_j$ with $\nu(\tilde{B}) > \frac{1}{999}\nu(A) \ge \frac{1}{2} \cdot \frac{1}{999}\nu(A')$ and a map*

*$G\colon \tilde{B} \to Y$ such that $\mathbf{x}$ and $G(\mathbf{x})$ are $\zeta_{\mathbf{x}}(i)$-friends for all $\mathbf{x} \in \bigcup_{j=1}^{\ell} B_j$. Moreover, $x_j = G(\mathbf{x})_j$ for all $j < \sigma_n$. Thus $G(\tilde{B}) \subset A'$ as well.*

*Proof.* We first prove the statement under the assumption that $\sigma_n \neq \sigma_{n'} - 1$. In this case, let $\hat{B}_1, \dots, \hat{B}_\ell$ be the cylinders and let $\hat{G}\colon \hat{A} \to \hat{B}$ be the map given by Lemma 5.11 for $m = n$. Let $B_j = \hat{B}_j \cap A$ and $\tilde{B} = \bigcup_{j=1}^{\ell} B_j \cap Q_r \cap \hat{G}^{-1} Q_r$.

Repeating the proof used to derive (36), we obtain cylinders defined in entries at most $\sigma_{n-2}$, with the entry before the last place defining the cylinders in $B_j$ (and also smaller than the cylinders defining $Q_r$ and $A$). One of these cylinders differs from the cylinder with the same defining entries in $X$. On all the other cylinders, $B_j$ intersects $A \cap Q_r$ as expected and the lemma follows using an argument analogous to the proof of (36).

Now we treat the remaining case, $\sigma_n = \sigma_{n'} - 1$ and the largest entry of the cylinders defining $B_j$ overlaps with the smallest entry of the cylinders defining $A$. In this case, we consider $A'$ whose defining entries are all larger than $\sigma_n$ (they are at least $10^{k+1}$ where the smallest entry defining $A$ is $10^k$). Then $(n', A', \rho')$ has two descendants in $\mathcal{H}_r(N, \varepsilon)$, $(n, A, \rho)$ and $(\tilde{n}, \tilde{A}, \tilde{\rho})$. One of these is $A \cap \{\mathbf{x}\colon x_{10^k} \geq \frac{k}{2}\}$ and so by the definition of $B_j$ in Lemma 5.11 has nonempty intersection with the cylinders $\hat{B}_1, \dots, \hat{B}_\ell$. The proof then follows as above, via the same arguments used to conclude the proof of Lemma 5.13. $\square$

5.5. **Restricting factors.** In this section, we develop our main criteria to rule out factors, Propositions 5.15 and 5.20 and Corollary 5.16. Morally, Proposition 5.15 and Corollary 5.16 rely on the assumption that the factor map is not to the 1 point system, while Proposition 5.20 relies on the assumption that the factor map is not an isomorphism, though for technical reasons it is helpful not to disambiguate these situations (in particular the proof of Proposition 5.20 uses Proposition 5.15).

For approximating, we make use of a metric giving rise to the strong operator topology on $L^2(\nu)$. While any such metric suffices for our purposes, it is convenient to choose one that simplifies the computations:

**Notation.** Let $\mathcal{B} = \mathcal{B}(L^2(\nu))$ denote the set of continuous linear operators on $L^2(\nu)$ and let $\{f_i\}_{i=1}^{\infty}$ be an orthonormal basis for $L^2(\nu)$ such that $\|f_i\|_\infty = 1$ for all $i$. Set $D\colon \mathcal{B} \times \mathcal{B} \to [0, \infty)$ to be the metric defined by

$$D(U, V) = \sum_{i=1}^{\infty} 2^{-i} \|U f_i - V f_i\|_2. \tag{37}$$

Note that restricting $D$ to the set (the choice of 10 is any arbitrary positive real)

$$\{(U, V) \in \mathcal{B} \times \mathcal{B}\colon \|U\|_{\mathrm{OP}} + \|V\|_{\mathrm{OP}} \leq 10\}$$

endows this set with the strong operator topology.

Recall that the notation $\mathcal{F}_i(N, \varepsilon)$ is introduced in Definition 5.8.

**Proposition 5.15.** *Assume that $(Y, \nu, T)$ has a non-trivial factor with associated factor map $P\colon Y \to Z$ and let $F\colon L^2(\nu) \to L^2(\nu)$ be the Markov operator defined by $P$ and $\lim_{k\to\infty} D(\sum_i \alpha_i^{(k)} U_{T^i}, F) = 0$ where $\alpha_i^{(k)} \geq 0$ for all $i, k$ and $\sum_i \alpha_i^{(k)} = 1$ for all $k$. Then for all $\varepsilon > 0$, there exists $N_0$ such that for all $N > N_0$ and all large enough $k$,*

$$\sum_i \alpha_i^{(k)} \sum_{\{(n,\beta,\rho) \in \mathcal{F}_i(N,\varepsilon)\colon\, \sigma_n > N\}} \beta < \varepsilon. \tag{38}$$

We record an immediate corollary for later use:

**Corollary 5.16.** Assume that $(Y,\nu,T)$ has a non-trivial factor with associated factor map $P\colon Y\to Z$ and let $F\colon L^2(\nu)\to L^2(\nu)$ be a Markov operator defined by $P$, then for all small enough $\varepsilon>0$ there exists $N_0$ and $\delta>0$ so that if $a_i\geq 0$, $\sum a_i=1$ and

$$D(\sum_i a_i U_{T^i},F)<\delta$$

then for all $N\geq N_0$

$$\sum_i a_i \sum_{\{(n,\beta,\rho)\in\mathcal{F}_i(N,\varepsilon)\colon\ \sigma_n>N\}}\beta<\varepsilon.$$

*Proof of Proposition 5.15.* Let $\varepsilon>0$ be small enough such that Lemma 5.6 can be satisfied. We proceed by contradiction, and show that if

$$\sum_i \alpha_i^{(k)} \sum_{\{(n,\beta,\rho)\in\mathcal{F}_i(N,\delta)\colon\ \sigma_n>N\}}\beta>10c,$$

then

$$\sum_{i\in\mathcal{H}_{N,\frac{c\delta}{9999}}}\alpha_i^{(k)}>c. \tag{39}$$

Then by taking $N$ sufficiently large, we obtain a contradiction via Lemma 5.6.

Let $J\subset\mathbb{N}$ be the set of indices $i$ such that

$$\sum_{\{(n,\beta,\rho)\in\mathcal{F}_i(N,\delta)\colon\ \sigma_n>N\}}\beta>c.$$

Since $\sum_{i\notin J}\alpha_i^{(k)}\sum_{\{(n,\beta,\rho)\in\mathcal{F}_i(N,\delta)\colon\ \sigma_n>N\}}\beta\leq c$, it follows that $\sum_{i\in J}\alpha_i^{(k)}>c$.

**Claim 5.17.** For any $i\in J$, we have $i\in\mathcal{H}_{N,\frac{c\delta}{9999}}$.

To check this, the triples defined in Definition 5.7 give the two possible reasons for $(n,A,\rho)\in\mathfrak{F}(N,\delta)$ with $\sigma_n>N$. The first is that $\sigma_n\notin E$, in which case Lemma 5.14 gives a set of points, contained in $A$, which have $i$-friends of measure at least $\frac{1}{999}\nu(A)$ and a map $G$ defined on these symbols, identifying the point with its friend so that $G(\mathbf{x})_j=x_j$ for all $j<\sigma_n$. The second is that $\rho>\delta$, in which case Lemma 5.13 similarly gives cylinders with measure at least $\frac{1}{6\cdot 64}\delta\cdot(1-4\cdot 99\delta)\nu(A)$. So if $i$ is such that $\sum_{\{(n,\beta,\rho)\in\mathcal{F}_i(N,\delta)\colon\ \sigma_n>N\}}\beta>c$, then since the $\beta$ are the measure of the sets $A$ mentioned in the previous 2 sentences, we have $i\in\mathcal{H}_{N,\frac{c\delta}{9999}}$. $\square$

The next lemma is another step in the chain of conditions on the approximations given by Theorem 5.1 to have a non-trivial factor, which started with Lemma 5.6.

**Lemma 5.18.** *Assume that $(Y,\nu,T)$ has a non-trivial factor with associated factor map $P\colon Y\to Z$ and assume that $F\colon L^2(\nu)\to L^2(\nu)$ is the Markov operator defined by $P$. Then for all $c,\varepsilon>0$, there exists $N_0$ such that for all $N>N_0$ we have that*

$$\sum_{i\in B_{c,\varepsilon}}\alpha_i^{(k)}<\sqrt{\varepsilon}, \tag{40}$$

*where*

$$B_{c,\varepsilon}=\Big\{i\colon D\big(U_{T^i},\sum_{(n,A,\beta)\in\mathfrak{F}_i(N,\varepsilon)}\mathbf{1}_A\cdot U_{T^n}\big)>10\sqrt{\varepsilon}+c\Big\}.$$

This notation $B_{c,\varepsilon}$ is local and only used until the end of this section.

*Proof.* We claim that for all $c > 0$ and small enough $\varepsilon > 0$, there exists $N$ such that if $S_1$, $S_2$ are measure preserving transformations such that $\nu(\{x\colon (S_1\mathbf{x})_j = (S_2\mathbf{x})_j \text{ for all } j < N\}) > 1-\varepsilon$ then $D(U_{S_1}, U_{S_2}) < c + 3\sqrt{\varepsilon}$.

To prove the claim, given $N \in \mathbb{N}$, set

$$A(N) = \{\mathbf{x} \in X\colon (S_1\mathbf{x})_j = (S_2\mathbf{x})_j \text{ for } j < N\}.$$

By Lusin's Theorem and uniform integrability, for any $f \in L^2(\nu)$, there exists $N \in \mathbb{N}$ such that if $A = A(N)$, then $\|(f\circ S_1)\mathbf{1}_A - (f \circ S_2)\mathbf{1}_A\|_2 < c$. As in the definition of $D$, let $\{f_i\}_{i=1}^{\infty}$ be an orthonormal basis of $L^2(\nu)$ with $\|f_i\|_\infty = 1$ for all $i \in \mathbb{N}$. Given $\varepsilon > 0$, choose $k$ such that $2^{-k} < \varepsilon$ and pick $N$ sufficiently large such that the associated set $A = A(N)$ ensures that $\|(f_i \circ S_1)\mathbf{1}_A - (f_i \circ S_2)\mathbf{1}_A\|_2 < c$ for $i = 1, \dots, k$. Then $\nu(A) > 1-\varepsilon$ for some $\varepsilon > 0$ and so (the definition of the metric $D$ is given in (37))

$$\begin{aligned} D(U_{S_1}, U_{S_2}) \le c + \sum_{i=1}^{\infty} 2^{-i}\Big(\int_{\mathbf{1}_{A^c}} |f_i\circ S_1 - f_i \circ f_2|^2 d\nu\Big)^{1/2} + \sum_{i=k+1}^{\infty} \|f_i \circ S_1 - f_i \circ S_2\|_2 \\ \le c + 2\sqrt{\varepsilon} + 2\varepsilon \le c + 3\sqrt{\varepsilon}, \end{aligned}$$

for all small enough $\varepsilon > 0$, proving the claim.

We now complete the proof by contraposition. Take $N_1 = N$ where $N$ is sufficiently large such that this claim holds. Then for any $N > N_1$, if

$$D(U_{T^i}, \sum_{\{(n,A,\rho)\in\mathfrak{F}_i(N,\varepsilon)\}} \mathbf{1}_A \cdot U_T^n) > c + 10\sqrt{\varepsilon},$$

then the claim implies that

$$\nu(\{\mathbf{x}\colon (T^i\mathbf{x})_j \ne (T^n\mathbf{x})_j \text{ for } j > N \text{ and } \mathbf{x} \in A \text{ where } (n, A, \rho) \in \mathfrak{F}_i(N,\varepsilon)\}) > 3\varepsilon.$$

By Lemma 5.9, for each $i$ such that this claim holds, we have

$$\sum_{\{(n,\beta,\rho)\in\mathcal{F}_i(N,\varepsilon)\}} \rho\beta > \varepsilon.$$

Further observe that if such a triple $(n, \beta, \rho)$ has $\rho > \varepsilon$, then $\sigma_n > N$. Assuming the negation of (40), we check that this gives the opposite of the bound in (38). Taking $N_0$ as in Proposition 5.15 for $F$ and $\varepsilon$, then if $N > N_0$ we have that $F$ is given by the factor map to the one point system. Taking $N_0$ in the statement of Lemma 5.18 to be the maximum of $N_1$ and the $N_0$ from Proposition 5.15, the statement follows. □

**Lemma 5.19.** *Let $([0, 1], R, \tau)$ be an ergodic measure preserving system and $(Z, R', \tau')$ be a factor with factor map $P$ that is not an isomorphism and let $F$ be the Markov operator defined by $P$. Then there exists a set $B \subset [0, 1)$ such that $\tau(B) \ge \frac{1}{3}$ and $(F\mathbf{1}_B)(x) < \frac{1}{2}$ for all $x \in B$.*

*Proof.* By disintegration of measures, there exist probability measures $\tau_{P(x)}$ carried on $P^{-1}(P(x))$ such that $\tau = \int_Z \tau_z d\tau'(z)$. Let $m_x = \text{ess inf}\{y\colon \tau_{P(x)}([0, y]) > \frac{1}{2}\}$. We claim that

$$\frac{1}{3} \le \tau_{P(x)}([0, m_x)) \le \frac{1}{2}. \tag{41}$$

Indeed, if $\tau_{P(x)}$ is non-atomic, then $\tau_{P(x)}([0, m_x)) = \frac{1}{2}$. If $\tau_{P(x)}$ is atomic, then by assumption there are at least two atoms and by ergodicity the atoms are of equal size. Because the largest atom of $\tau_{P(x)}$ has measure at most $\frac{1}{2}$, and there are a sum of atoms with measure between $\frac{1}{3}$ and $\frac{1}{2}$ inclusive, we have (41).

We claim that $x \mapsto m_x$ is a Borel measurable function. First recall that by disintegration of measures, the map $x \mapsto \tau_{P(x)}$ is Borel. Next, for every interval $[0, a]$, we have that the map from Borel measures to real numbers, $\mathfrak{p} \mapsto \mathfrak{p}([0, a])$ is a Borel map. It follows that the function $h(x, y) = \tau_{P(x)}[0, y]$ is Borel and so $h^{-1}\big((\frac{1}{2}, 1]\big)$ is Borel and for each $x$, $m_x = \text{ess inf}\{y : (x, y) \in h^{-1}\big((\frac{1}{2}, 1]\big)\}$. Note that since $\tau_{P(x)}$ is a probability measure this is the same as

$$\min\{\text{ess inf}\{y : (x, y) \in h^{-1}\big((\frac{1}{2}, 1]\big)\}, 1\}.$$

Now if $A \subset [0, 1]^2$ is Borel, then $x \mapsto \min\{\text{ess inf}\{y \colon (x, y) \in A\}, 1\} = m_x$ is Borel measurable. Indeed, the set of such $A$ is a monotone class containing the algebra generated by rectangles. To see this, for countable nested unions,

$$\max\{\text{ess inf}\{y \colon (x, y) \in \cup_{i=1}^{\infty} A_i\}, 1\} = \lim_{n\to\infty} \min\{\text{ess inf } \{y \colon (x, y) \in \cup_{i=1}^{n} A_i\}, 1\}.$$

For countable nested intersections, we have that $\min\{\text{ess inf}\{y \colon (x, y) \in \cap_{i=1}^{\infty} A_i\}, 1\}$ can be defined piecewise as

$$\begin{cases} 1 & \text{if } \{y : (x, y) \in \cap_{i=1}^{\infty} A_i\} \text{ is a zero set} \\ \lim\limits_{n\to\infty} \min\{\text{ess inf}\{y \colon (x, y) \in \cap_{i=1}^{n} A_i\}, 1\} & \text{otherwise.} \end{cases}$$

By the Monotone Class Theorem it is defined on the smallest $\sigma$-algebra containing rectangles, which is the Borel $\sigma$-algebra. We set $B = \bigcup_{x\in[0,1]} [0, m_x)$. Then for almost every $x$, $\tau_{P(x)}([0, m_x)) \leq \frac{1}{2}$ and so $F(\mathbf{1}_B)(x) \leq \frac{1}{2}\mathbf{1}_B(x)$ for all $x \in B$. $\square$

Note that this lemma holds for any Lebesgue space and in particular to $(Y, \nu, T)$.

**Proposition 5.20.** *Assume that $(Y, \nu, T)$ has a non-trivial factor with associated factor map $P \colon Y \to Z$ and let $F$ be the Markov operator defined a factor map $P$. Let $\lim\limits_{k\to\infty} D(\sum_i \alpha_i^{(k)} U_{T^i}, F) = 0$ where $\alpha_i^{(k)} \geq 0$ for all $i, k$ and $\sum_i \alpha_i^{(k)} = 1$ for all $k$. Then for all $\frac{1}{99} > \varepsilon > 0$, there exists $N_0$ such that for any $N > N_0$ and all sufficiently large enough $k$, we have*

$$\sum_i \alpha_i^{(k)} \sum_{\{(n,\beta,\rho)\in\mathcal{F}_i(N,\varepsilon) \colon n\neq 0\}} \beta > \frac{1}{3} \cdot (\frac{1}{2} - 3\varepsilon). \tag{42}$$

*Proof.* Because $F \neq \text{Id}$, for almost every $\mathbf{x} \in Y$ we have $P^{-1}(P\mathbf{x})$ is at least two points, and so there exists a set $B$ as in Lemma 5.19. By assumption and Lemma 5.18, there exist $k_1, N_1$ such that for all $k \geq k_1$ and $N > N_1$ we have

$$\nu(\{\mathbf{x} : |(F\mathbf{1}_B)(\mathbf{x}) - (\sum_i \alpha_i^{(k)} \sum_{(n,A,\rho)\in\mathfrak{F}_i(N,\varepsilon)} \mathbf{1}_A(\mathbf{x})(U_{T^n}\mathbf{1}_B)(\mathbf{x})| > \frac{\varepsilon}{9}\}) < \frac{\varepsilon}{9}. \tag{43}$$

By the non-negativity of the $\alpha_i^{(k)}$, for almost every $\mathbf{x} \in Y$ we have

$$(\sum_i \alpha_i^{(k)} \sum_{(n,A,\rho)\in\mathfrak{F}_i(N,\varepsilon)} \mathbf{1}_A U_{T^n}\mathbf{1}_B)(\mathbf{x}) \geq \sum_i \alpha_i^{(k)} \sum_{\{(n,A,\rho)\in\mathfrak{F}_i(N,\varepsilon) \colon n=0 \text{ and } \mathbf{x}\in A\}} \mathbf{1}_B(\mathbf{x}).$$

If (42) does not hold, then there is a set of $\mathbf{x}$ of measure at least $\frac{2}{3}+\varepsilon$ such that

$$\sum_i \alpha_i^{(k)}\Big(\sum_{\{(n,A,\rho)\in\mathfrak{F}_i(N,\varepsilon)\colon\, n=0 \text{ and } \mathbf{x}\in A\}} \beta\Big) > \frac{1}{2}+\varepsilon.$$

But then there is a set of $\mathbf{x}\in B$ of measure at least $\varepsilon$ such that for each $\mathbf{x}$ in this set,

$$\Big(\sum_i \alpha_i^{(k)} \sum_{(n,A,\rho)\in\mathfrak{F}_i(N,\varepsilon)} \mathbf{1}_A(x)U_{T^n}\mathbf{1}_B\Big)(\mathbf{x}) > \frac{1}{2}+\varepsilon \geq (F\mathbf{1}_B)(\mathbf{x})+\frac{\varepsilon}{2},$$

a contradiction of (43). □

## 6. The behavior of a projection

### 6.1. Overview of the proof that $(Y,\nu,T)$ is prime.

In this section, we show that our constructed system is prime:

**Theorem 6.1.** *The system $(Y,\nu,T)$ is prime.*

We start with an overview of the proof and then proceed to study different cases. We assume that $(Y,\nu,T)$ has a non-trivial factor $Z$ with factor map $P\colon Y\to Z$ and assume that $F\colon L^2(\nu)\to L^2(\nu)$ is a Markov operator defined by $P$. We further assume that $F$ is the limit, as $k\to\infty$, of $\sum\alpha_i^{(k)}U_{T^i}$ in the strong operator topology (Corollary 5.2). Given $\varepsilon>0$, by Proposition 5.15 we can assume that there exists $N_0\in\mathbb{N}$ such that for all $N>N_0$ and sufficiently large $k$, we have that

$$\sum_i \alpha_i^{(k)} \sum_{(n,A,\rho)\in\mathfrak{F}_i} \mathbf{1}_A U_{T^n}$$

gives a good approximation to $\sum\alpha_i^{(k)}U_{T^i}$ (which in turn leads to a good approximation for $F$). The general idea in the proof of Theorem 6.1 is that we rule out the possibility that $\sum\alpha_i^{(k)}U_{T^i}$ is close to a non-trivial projection. The key facts used are that the composition of projections is still a projection and by properties of the strong operator topology, we may assume that for any fixed $M$, for all large enough $k$

$$\underbrace{\Big(\sum\alpha_i^{(k)}U_{T^i}\Big)\circ\Big(\sum\alpha_i^{(k)}U_{T^i}\Big)\ldots\circ\Big(\sum\alpha_i^{(k)}U_{T^i}\Big)}_{M\text{ times}} \tag{44}$$

is close to $F^M=F$, and this is also close to $\sum\alpha_i^{(k)}U_{T^i}$. We then use the fact that (44) is

$$\sum_{(i_1,\ldots,i_M)}\Big(\prod_{m=1}^{M}\alpha_{i_m}^{(k)}\Big)U_{T^{\sum i_m}}, \tag{45}$$

and apply Definition 5.7 to (45). That is, we study

$$\sum_{(i_1,\ldots,i_M)}\Big(\prod_{m=1}^{M}\alpha_{i_m}^{(k)}\Big)\sum_{(n,A,\rho)\in\mathfrak{F}_{\sum_{j=1}^{M} i_j}(N,\frac{\varepsilon}{2})}\mathbf{1}_A U_{T^n}.$$

Treating 3 different cases, this allows us to produce friends and obtain a contradiction via Lemma 5.6. We now make this precise.

6.2. **Set up for the proof of Theorem 6.1.** We begin a proof by contradiction, assuming that there is a Markov operator $F$ coming from a non-trivial factor map. By Theorem 5.1, there exists $\alpha_i^{(k)} \geq 0$ with $\sum_i \alpha_i^{(k)} = 1$ for all $i$ such that $\sum_i \alpha_i^{(k)} U_{T^i}$ converges in the strong operator topology to $F$.

We assume that $\varepsilon > 0$ is sufficiently small such that all of the Lemmas and Propositions in Section 5 hold. That is, $\varepsilon < \min\{\frac{1}{8\cdot 99}, \frac{1}{9999}\}$ and small enough such that Lemma 5.6 holds. We also assume that

$$\varepsilon < \frac{1}{10^5 \cdot 9999}, \tag{46}$$

which is to be used in Lemma 6.6. Furthermore, we choose $N_1 > 6$ (this choice is made to simplify the analysis in the third case we consider) to be sufficiently large such that Lemma 5.6 holds for $\frac{N_1}{2}$ and $\varepsilon^8$ and such that

$$2^{-\frac{N_1}{2}} < \varepsilon^4. \tag{47}$$

Setting

$$G_N = \{n\colon \sum_{\{(i,c,\gamma)\in \mathcal{F}_n(\frac{N}{2},\varepsilon^2)\colon \sigma_i > \frac{N}{2}\}} c < \varepsilon^4\},$$

our choices imply that for all sufficiently large $k$,

$$\sum_{n\in G_N} \alpha_n^{(k)} > \frac{3}{4}. \tag{48}$$

Indeed, by Claim 5.17 if $n \notin G_N$ then $n \in \mathcal{H}_{N,\frac{\varepsilon^2\varepsilon^4}{9999}}$. Since $\frac{\varepsilon^6}{9999} > \varepsilon^8$, by Lemma 5.6 this contradicts that our factor map is not to the 1 point system. We set $s = \min\{10^j\colon 10^{j-1} \geq N\}$, set $s' = \min\{10^\ell\colon 10^\ell \geq N\}$, and recall that $r_i$ is defined in (9). Define

$$M = M_N = \frac{r_{s-2}}{r_N}. \tag{49}$$

Although $M = M_N$ depends on $N$, as $N$ is fixed at this point, we omit it from the notation, except at one step in the proof of Proposition 6.3. The motivation behind this definition of $M$ is given by the following lemma (this plays a role in the proof of Proposition 6.3):

**Lemma 6.2.** *For all sufficiently large $N$, we have that $M^{\frac{1}{8}} > r_{s'+1}$, where $M$ is defined as in* (49)*. Moreover, for any $\varepsilon > 0$, for all sufficiently large $N$ we have $8^{(1-\varepsilon)10^k} < r_{10^k-1} < 8^{(1+\varepsilon)10^k}$.*

*Proof.* We first claim that for all $\varepsilon > 0$, there exists $k$ such that

$$8^{(1-\varepsilon)10^k} < r_{10^k-1} < 8^{(1+\varepsilon)10^k}. \tag{50}$$

For all $\ell \notin E$, we have that $r_{\ell+1} = a_{\ell+1} r_\ell - 1$. Thus there exists $\ell_0$ with $r_{\ell+1} \geq 8^{1-\frac{\varepsilon}{2}} r_\ell$ for all $\ell \geq \ell_0$ and the lower bound follows. For the upper bound, $r_{\ell+1} < 8r_\ell$ for all $\ell \notin E$ and so there exists $\ell_0$ such that $\ell < (8^{\frac{\varepsilon}{2}})^{9\cdot 10^{\ell-1}}$ for all $\ell \geq \ell_0$. For any such $\ell$, we have $r_{10^{\ell-1}+j} < (8^{1+\frac{\varepsilon}{2}})^j r_{10^{\ell-1}}$ and the upper bound follows, completing the proof of (50).

For all large enough $N$, by (50) we have

$$M = M_N = \frac{r_{s-1}}{r_N} > \frac{\frac{1}{8} r_s}{r_{s'}} > \frac{8^{(1-\varepsilon)10^k}}{8^{(1+\varepsilon)10^{k-1}}} > 8 \cdot 8^{8.5(1+\varepsilon)10^{k-1}} > r_{s'+1}^8,$$

where the second to last inequality holds for all sufficiently large $N$. $\square$

6.3. **The three cases:** This leads us to consider three possibilities for the behavior of the projection $F$ on $L^2(\mu)$ (recall that $D$ is the metric defined in (37)):
**Case 1:**
(51)
$$D\Big(\sum_{(i_1,\dots,i_M)}\prod_{m=1}^{M}\alpha_{i_m}^{(k)}U_{T^{\sum i_m}},(\sum_i \alpha_i^{(k)}\sum_{\{j\colon (j,A,\rho)\in\mathfrak{F}_i(N,\varepsilon):\sigma_j<N\}}U_{T^j}\mathbf{1}_A\big)^M\Big)<\varepsilon^{\frac14}.$$

**Case 2:** Case 1 does not hold and

$$\sum_n \alpha_n^{(k)}\sum_{\{(j,\beta,\rho)\in\mathcal{F}_n(N,\varepsilon^2)\colon\,\sigma_j>N\}}\beta<\frac{\varepsilon^2}{M}. \tag{52}$$

**Case 3:** Case 1 does not hold and

$$\sum_n \alpha_n^{(k)}\sum_{\{(j,\beta,\rho)\in\mathcal{F}_n(N,\varepsilon^2)\colon\,\sigma_j>N\}}\beta\geq\frac{\varepsilon^2}{M}. \tag{53}$$

We analyze each of these cases separately.

6.4. **Case 1:** Fix $\varepsilon>0$ and assume that (51) holds. Roughly speaking, the assumption means that when we iterate the approximation of our transformation given by Definition 5.7 up to $M$ times, we remain close to the original map.

**Proposition 6.3.** *There exists $N_3$ such that if $N>N_3$ and $\gamma$ is a probability measure supported on $\{-2r_N,-2r_N+1,\dots,-1,0,1,\dots,2r_N\}$ with $\gamma(\{0\})<\frac17$, and $M$ corresponds to $N$ as in (49), then*

$$\gamma^M\big(\{(i_1,\dots,i_M)\colon r_{s-2}>|\sum_{j=1}^{M}i_j|>r_{s'+1}\}\big)>\frac19.$$

*Proof.* Let $(\Omega,\mathbb{P})$ be a probability space and let $F_1,\dots,F_M\colon(\Omega,\mathbb{P})\to\mathbb{Z}$ be a sequence of independent $\gamma$ distributed random variables, let $Z=\sum_{i=1}^M F_i-M\mathbb{E}_{\mathbb{P}}(F_1)$, and let $\sigma$ be the variance of $F_1$. Then the variance of $Z$ is $M\sigma$.

By our choice of $M=M_N$, we have that $|\sum_{j=1}^M i_j|<r_{s-1}$ for all $i_j$ in the support of $\gamma$ and the remainder of the proof is devoted to showing the lower bound.

By Lemma 6.2, for sufficiently large $N$, we have that $M^{\frac18}$ is bounded from below. The proof splits into two cases. In the first, $|\mathbb{E}_{\mathbb{P}}(F_1)|$ is not too small and we make use of Chebyshev's inequality (as in the proof of the weak law of large numbers). In the second case, $|\mathbb{E}_{\mathbb{P}}(F_1)|$ is small, and using the central limit theorem we show that for many $\omega$, $|Z(\omega)|\sim\sqrt{M\sigma}$.

First assume that $\frac{1}{99}<|\mathbb{E}_{\mathbb{P}}(F_1)|$. We compute $\mathbb{E}_{\mathbb{P}}(Z^2)$ and apply Chebyshev's inequality. By independence of the $F_i$, we have that

$$\mathbb{E}_{\mathbb{P}}(Z^2)=M\mathbb{E}_{\mathbb{P}}\big((F_1^2-2F_1\mathbb{E}_{\mathbb{P}}(F_1)+\mathbb{E}_{\mathbb{P}}(F_1)^2)\big).$$

Applying Hölder's Inequality, we bound $\|F_i^2\|$ by $\|F_i\|_\infty\cdot\|F_i\|$. By Lemma 6.2, we have that $\|F_i\|\leq 2r_N<M^{\frac18}$ and similarly, $\|F_i\|_\infty<M^{\frac18}$. Then $\mathbb{E}_{\mathbb{P}}(Z^2)\leq MM^{\frac14}$ and so it follows from Chebyshev's Inequality that $\mathbb{P}(\{x\colon|f(x)|>\varepsilon\int f^2d\mathbb{P}\})|<\frac{1}{\varepsilon^2\int f^2d\mathbb{P}}$. Therefore,

$$\mathbb{E}_{\mathbb{P}}(\{\omega\colon|Z(\omega)|<4M^{\frac34}\})>\frac12.$$

Since $\sum F_i = M\mathbb{E}_{\mathbb{P}}(F_1) + Z$ and $|M\mathbb{E}_{\mathbb{P}}(F_1)| > \frac{M}{99}$, we have that if $|Z(\omega)| < 4M^{\frac{3}{4}}$, then

$$|\sum_{i=1}^{M} F_i(\omega)| > \frac{M}{99} - 4M^{\frac{3}{4}} > M^{\frac{1}{2}}.$$

Now we consider the case that $\frac{1}{99} \geq |\mathbb{E}_{\mathbb{P}}(F_1)|$. Under this assumption, because $\gamma(\{0\}) < \frac{1}{7}$ and is supported on integers, we have that

$$|\mathbb{E}_{\mathbb{P}}(F_1 - \mathbb{E}_{\mathbb{P}}(F_1)^2)| \leq \frac{6}{7}(1 - \frac{1}{99})^2,$$

which implies that the variance $\sigma$ is at least $\frac{1}{8}$.

Let $\varepsilon > 0$. There exists $N_0$ such that for all $N \geq N_0$, and any probability measure $\mathfrak{p}$ on $[-r_N, r_N]$, we have

$$\frac{1}{(M_N)^{\frac{k}{2}}} \int |t|^k d\mathfrak{p}(t) < \frac{\varepsilon}{2^k}$$

for all $k \geq 3$. It follows that $\phi_t = \mathbb{E}e^{\frac{itZ}{\sqrt{M\sigma}}}$ is bounded by $-\frac{t^2}{2} + c(t)$, where $|c(t)| < \varepsilon$.

We use Lévy's Continuity Theorem to complete the proof. Namely, we repeat this process for a sequence of $N_j$ tending to infinity and obtain $Z_j, M_j$, and $\sigma_j$ such that $\phi_t^{(j)} = \mathbb{E}_{\mathbb{P}}e^{\frac{itZ_j}{\sqrt{M_j\sigma_j}}} \to e^{-\frac{t^2}{2}}$ pointwise. Since $e^{-t^2/2}$ is continuous at 0, it follows that that $\frac{Z_j}{\sqrt{M_j\sigma_j}}$ converges in distribution. Thus for sufficiently large $j$, it follows that the probability that $|Z_j| > \frac{\sqrt{M_j}}{16}$ is at least $1/9$. If the expectation of $F_1$ is nonnegative, this implies that

$$\gamma^M\Big(\{(i_1, \ldots, i_M)\colon \sum_{j=1}^{M} i_j > r_{s'+1}\}\Big) > \frac{1}{9}.$$

Otherwise,

$$\gamma^M\Big(\{(i_1, \ldots, i_M)\colon \sum_{j=1}^{M} i_j < -r_{s'+1}\}\Big) > \frac{1}{9},$$

and the result follows. □

**Notation:** Let $\tilde{\varepsilon}$ be as in Section 6.2, that is less than $\frac{1}{10^5 \cdot 9999}$ and small enough such that Lemma 5.6 holds. Let $N_2$ be chosen according to Corollary 5.16 for $\tilde{\varepsilon}^2$. Let $\tilde{N} \geq \max\{N_1, N_2, N_3\}$ and $M = M_{\tilde{N}}$. Let $\tilde{k}$ be chosen large enough such that

(A) $D((\sum_i \alpha_i^{(\tilde{k})} U_{T^i})^M, F) < \delta$ where $\delta$ is as in Corollary 5.16 for $\varepsilon = \tilde{\varepsilon}$ and $N_0 = \tilde{N}$

(B) (47) with $G_{\tilde{N}}$ and (48) hold.

*Concluding the proof of Theorem 6.1 in case 1:* Let

$$A = \{(i_1, \ldots, i_M) : \sigma_{\sum i_j} > \tilde{N} \text{ and } \sigma_{\sum i_j} \notin E\}.$$

Let $\gamma$ be the probability measure on $\mathbb{Z}$ given by $\gamma(i) = \alpha_i^{(\tilde{k})}$. By Proposition 5.20, we have $\gamma(0) < \frac{1}{7}$. By Proposition 6.3, applied to $\gamma^M$ we have that $\sum_A \prod_{j=1}^M \alpha_{i_j}^{(\tilde{k})} > \frac{1}{2}$. Indeed if $\sum_{j=1}^M i_j \in (r_{s'}, r_{s-1})$, then $\sigma_{\sum i_j} \notin E$. By (45) and our choice of $\tilde{k}$, we obtain a contradiction of Corollary 5.16. Thus this case can not occur.

6.5. **Case 2:** In the absence of the first case, we are left with showing that for at least $\tilde{\varepsilon}^2$ of the sums $\sum_{j=1}^M i_j$, for at least $\tilde{\varepsilon}^2$ points we have $\sum_{j=1}^M i_j$-friends. Roughly speaking, the idea is that $\prod_{j=1}^M U_T^{i_j}$ under iteration does not stay close to $\prod_{\mathfrak{F}_{i_j}(\tilde{N},\tilde{\varepsilon})} \sum_{(\ell,A,\rho)\in\mathfrak{F}_{i_j}(\tilde{N},\tilde{\varepsilon})} \mathbf{1}_A U_T^\ell$, and so the sum

$$\sum_{(i_1,\dots,i_M)\in\mathbb{Z}^M} (\prod_{j=1}^M \alpha_{i_j}^{(\tilde{k})}) \sum_{\{(j,\beta,\rho)\in\mathcal{F}_{\sum_{\ell=1}^M i_\ell}(\tilde{N},\tilde{\varepsilon})\colon \sigma_j>\tilde{N}\}} \beta$$

becomes significant. To make this precise, we deal with two cases separately, depending on the sizes of the sums in (52) and (53).

Thus as we are in Case 2, we assume that

$$\sum_n \alpha_n^{(\tilde{k})} \sum_{\{(j,\beta,\rho)\in\mathcal{F}_n(\tilde{N},\tilde{\varepsilon}^2)\colon \sigma_j>\tilde{N}\}} \beta < \frac{\tilde{\varepsilon}^2}{M}.$$

Given $n \in \mathbb{N}$, set $\mathfrak{H}_0 = \{(n,[0,1],\rho)\}$ and define the *reduction* $\mathrm{red}_k(n)[\mathbf{x}] = (m,A,\rho)$, where

- $(m,A,\rho)\in\mathfrak{H}_r(\tilde{N},\tilde{\varepsilon})$ for some (smallest) $r$,
- $\mathbf{x}\in A$, and
- $\sigma_m \le 10^k$ or $(m,A,\rho)\in\mathfrak{F}(N,\tilde{\varepsilon})$ (that is, $m\notin E$ or $\rho>\tilde{\varepsilon}$).

Let $\widetilde{\mathrm{red}}_k(n)[\mathbf{x}]$ denote the first coordinate of $\mathrm{red}_k(n)[\mathbf{x}]$. We say that the sum $\sum_j i_j$ is *treatable* if

$$\sum_{\{(k,\beta,\rho)\in\mathcal{F}_{i_j}(N,\varepsilon^2)\colon \sigma_k>N\}} \beta < \tilde{\varepsilon}$$

for all choices of $i_j$ and the sum $\sum_j i_j$ is $\mathbf{x}$-*treatable* if for all $i_j$, the elements $(n,A,\rho)\in\mathfrak{F}_{i_j}(N,\tilde{\varepsilon})$ satisfy $\mathbf{x}\in A$ has $\sigma_n\le N$. (Recall, $\mathfrak{F}_{i_j}(N,\tilde{\varepsilon})$ is not necessarily a singleton, but each $\mathbf{x}\in\mathbf{Y}$ is in the second coordinate of exactly one triple in $\mathfrak{F}_{i_j}(N,\tilde{\varepsilon})$.)

**Lemma 6.4.** *Assume* $\sum i_j$ *is* $\mathbf{x}$-treatable, $A\subset Y$ *so that* $(n,A,\rho)=\mathrm{red}_k(\sum_{j=1}^M i_j)[\mathbf{x}]$,

$$\widetilde{\mathrm{red}}_k(\sum_{j=1}^M i_j)[\mathbf{x}] \neq \sum_{j=1}^M \widetilde{\mathrm{red}}_k(i_j)[\mathbf{x}] \tag{54}$$

*and $k$ is maximal with this property. Then* $\mathrm{red}_{k+1}(\sum_{j=1}^M i_j)[\mathbf{x}]$ *or* $\mathrm{red}_k(\sum_{j=1}^M i_j)[\mathbf{x}]$ *lies in* $\mathfrak{F}_{\sum_{j=1}^M i_j}(\tilde{N},\tilde{\varepsilon})$.

*Proof.* Set

$$m=\sum_{j=1}^M \widetilde{\mathrm{red}}_{k+1}(i_j)[\mathbf{x}] = \widetilde{\mathrm{red}}_{k+1}(\sum_{j=1}^M i_j)[\mathbf{x}]. \tag{55}$$

First, because $i_j$ is $\mathbf{x}$-treatable, we have

$$\begin{aligned}\widetilde{\mathrm{red}}_{k+1}(i_j)[\mathbf{x}] - \widetilde{\mathrm{red}}_k(i_j)[\mathbf{x}] \in \{&d_{10^{k+1}}(\widetilde{\mathrm{red}}_{k+1}(i_j)[\mathbf{x}])r_{10^{k+1}},\\ &d_{10^{k+1}}(\widetilde{\mathrm{red}}_{k+1}(i_j)[\mathbf{x}])r_{10^{k+1}} + d_{10^{k+1}}(\widetilde{\mathrm{red}}_{k+1}(i_j)[\mathbf{x}])\}\end{aligned}$$

for all $j$. Let $n=\sum_j\left(\widetilde{\mathrm{red}}_{k+1}(i_j)[\mathbf{x}] - \widetilde{\mathrm{red}}_k(i_j)[\mathbf{x}]\right)$ and by our assumption that $k$ is maximal satisfying (54), $n\neq \widetilde{\mathrm{red}}_{k+1}(m)[\mathbf{x}] - \widetilde{\mathrm{red}}_k(m)[\mathbf{x}]$. We now treat a series of cases, each of which is straightforward. If $\sigma_{\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}]} > 10^{k+1}$, then because $|\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}]| \leq M r_{10^{k+1}+1} < r_{10^{k+2}-1}$, $\sigma_{|\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}]} \notin E$. Thus the reduction algorithm given in Definition 5.7 halts and the lemma follows. If $\sigma_{\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}]} = 10^{k+1}$, then by choice of $M$ and $N$ we have

$$|\sum_j \widetilde{\text{red}}_k(i_j)([\mathbf{x}])| < r_{10^{k+1}-2}. \tag{56}$$

Also, because $\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}] - \widetilde{\text{red}}_k(m)[\mathbf{x}], \widetilde{\text{red}}_{k+1}(i_j)[\mathbf{x}] - \widetilde{\text{red}}_k(i_j)[\mathbf{x}]$ are all multiples of either $r_{10^{k+1}}$ or $r_{10^{k+1}}+1$ (depending on $x_{10^{k+1}}$), we have that $\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}] - \widetilde{\text{red}}_k(m)[\mathbf{x}] = n + p$ where $|p| \geq r_{10^{k+1}}$.

By the algorithm for representing numbers in terms of $d_i$, we have that if $|p| \geq r_{\sigma_\ell+2}$ then $|p| > 5|\ell|$ and so

$$\sigma_{\ell+p} \geq \sigma_p - 1. \tag{57}$$

Thus by Equation (56),

$$\sigma_{\widetilde{\text{red}}_k(m)[\mathbf{x}]} = \sigma_{\sum_j \widetilde{\text{red}}_k(i_j)[\mathbf{x}]-p} \geq 10^{k+1} - 1 > 10^k.$$

Therefore, by the definition of $\text{red}_k(\cdot)[\mathbf{x}]$ we must have that $\text{red}_k(m)[\mathbf{x}] \in (N, \tilde{\varepsilon})$, that is, the algorithm halts in this case as well. In the final case, $\sigma_{\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}]} < 10^{k+1}$ we have that $n = 0$ and either $\sigma_{\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}]} \leq 10^k$ in which case $\text{red}_k(m)[\mathbf{x}] = \text{red}_{k+1}(m)[\mathbf{x}]$ and this is $\sum \widetilde{\text{red}}_k(i_j)[\mathbf{x}]$ a contradiction, or $10^k < \sigma_{\widetilde{\text{red}}_{k+1}(m)[\mathbf{x}]} < 10^{k+1}$ and so the algorithm stops. □

*Concluding the proof of Theorem 6.1 in case 2.* Let

$$I(\mathbf{x}, M) = \{(i_1, \dots, i_M)\colon \sum i_j \text{ is } \mathbf{x}\text{-treatable and } \sum_{j=1}^M \widetilde{\text{red}}(i_j)[\mathbf{x}] \neq \widetilde{\text{red}}(\sum_{j=1}^M i_j)[\mathbf{x}]\}.$$

We assume that we are not in the first case and moreover (52) holds. Thus for a set of $\mathbf{x}$ of measure at least $\tilde{\varepsilon}$, we have

$$\sum_{\{(i_1,\dots,i_M)\colon (i_1,\dots,i_M) \in I(\mathbf{x},M)\}} \prod_{j=1}^M \alpha_{i_j}^{(\tilde{k})} \geq \tilde{\varepsilon}.$$

For each such $\mathbf{x}$, $(i_1, \dots, i_M)$ there exists $m$, $A \subset [0,1]$, $\rho \geq 0$ such that $\mathbf{x} \in A$ and $(m, A, \rho) \in \mathfrak{F}_{\sum_{j=1}^M i_j}(\tilde{N}, \tilde{\varepsilon})$ and $\sigma_m = k > \tilde{N}$. Thus we have

$$\sum_{(i_1,\dots,i_M) \in \mathbb{Z}^M} (\prod_{j=1}^M \alpha_{i_j}^{(\tilde{k})}) \sum_{\{(j,\beta,\rho) \in \mathcal{F}_{\sum_{\ell=1}^M i_\ell}(\tilde{N},\tilde{\varepsilon})\colon \sigma_j > \tilde{N}\}} \beta \geq \tilde{\varepsilon}^2$$

By Corollary 5.16, and our choices of $\tilde{N}$, $\tilde{k}$, (see (A)) this establishes case 2. □

6.6. **Case 3 (we assume neither of the conditions in Case 1 or in Case 2 holds).** We say that $n$ is *good for reduction* if

$$\sum_{\{(i,c,\gamma) \in \mathcal{F}_n(\frac{\tilde{N}}{2}, \tilde{\varepsilon}^4)\colon \sigma_i > \frac{\tilde{N}}{2}\}} c < \tilde{\varepsilon}^4$$

and we say $n$ is *bad for reduction* if

$$\sum_{\{(i,c,\gamma)\in\mathcal{F}_n(\tilde{N},\tilde{\varepsilon}^2):\ \sigma_i>\tilde{N}\}} c > \tilde{\varepsilon}.$$

By our assumption on $\tilde{k}$ and the estimate for sufficiently large $\tilde{k}$ given in (48), we have:

**Lemma 6.5.** *Let* $G=\{\vec{i}\in\mathbb{Z}^M : i_j$ *is good for reduction for at least* $\frac{M}{2}$ *choices of* $j\}$, *then* $\sum_{\vec{i}\in G}\prod_{j=1}^M \alpha_{i_j}^{(\tilde{k})} \geq \frac{1}{2}$.

**Lemma 6.6.** *If* $j$ *is bad for reduction and* $k$ *and* $m$ *are good for reduction, then*

$$j-k+m\in\mathcal{H}_{\tilde{N},\tilde{\varepsilon}^4}.$$

*Similarly, if* $j$ *and* $m$ *are good for reduction and* $k$ *is bad for reduction, then*

$$j-k+m\in\mathcal{H}_{\tilde{N},\tilde{\varepsilon}^4}.$$

Note that we separate the roles of the terms $k$ and $m$ to make it easier to apply the lemma (see Corollary 6.7).

*Proof.* We establish the first claim, as the second is similar. First, if $j$ is bad for reduction, then by Claim 5.17 we have that $j\in\mathcal{H}_{\tilde{N},\frac{\tilde{\varepsilon}^3}{9999}}$. Taking $\mathcal{A}_j$ and $G_j$ as in Notation 5.5, we have that when $\mathbf{x}\in\mathcal{A}_j$ there exists $0<|\ell|\leq 3$ such that $(T^\ell T^j\mathbf{x})_i=(T^j(G_j\mathbf{x}))_i$ for all $i\leq\tilde{N}$. Recall that the sets $D_a$ are defined in (6). Now if $(T^{n+\ell+j}\mathbf{x})_i\neq(T^{n+j}(G_j\mathbf{x}))_i$ for some $i\leq\tilde{N}$, then there exists $a>\tilde{N}$ such that either $S^bT^{\ell+j}\mathbf{x}\in D_a$ for $|b|\leq 3|n|\leq|\zeta_{T^{\ell+j}\mathbf{x}}(b)|$ or $S^bT^jG_j\mathbf{x}\in D_a$ for $|b|\leq 3|n|\leq|\zeta_{T^jG_j\mathbf{x}}(b)|$. This uses Corollary 5.4. If $\sigma_n\leq\frac{\tilde{N}}{2}$, then the measure of such points is at most

$$4\cdot 3n\sum_{a>\tilde{N}}\mu(D_a)<2^{-\frac{\tilde{N}}{2}}.$$

Let $d\in\mathbb{Z}$, $(n,A,\rho)\in\mathfrak{F}_d(\frac{\tilde{N}}{2},\tilde{\varepsilon}^4)$, and $\mathbf{x}\in\mathcal{A}_j\cap A$ (which implies that $G_j(\mathbf{x})\in A$). If $(T^{\ell+j}\mathbf{x})_i=(T^j(G_j\mathbf{x}))_i$ for all $i\leq\tilde{N}$ but $(T^dT^\ell T^j\mathbf{x})_i=(T^dT^j(G_j\mathbf{x}))_i$ for some $i\leq\tilde{N}$, then $\mathbf{x}\in\mathcal{P}_r$ for some $r$. Since $k$ and $m$ are good for reduction, by iterating Lemma 5.10 when $d=k$ or $m$, we have that the measure of the set of such points is at most $40\tilde{\varepsilon}^4$. Combining these two estimates and considering $(n,A,\rho)\in\mathfrak{F}_d(\frac{\tilde{N}}{2},\tilde{\varepsilon}^4)$ with $\sigma_n>\frac{\tilde{N}}{2}$, we obtain that

$$j-k+m\in\mathcal{H}_{\tilde{N},\frac{\tilde{\varepsilon}^3}{9999}-2\cdot(\tilde{\varepsilon}^4+40\tilde{\varepsilon}^4+2^{-\frac{\tilde{N}}{2}})}.$$

By the assumptions (46) and (47) on $\tilde{N}$ and $\tilde{\varepsilon}$, the lemma follows. □

**Corollary 6.7.** Assume $\sum_{\ell=1}^M i_\ell$ is good for reduction and $j_\ell$ is such that $j_\ell=i_\ell$ except at one place where $i_\ell$ is good for reduction and $j_\ell$ is bad for reduction. Then $\sum_{\ell=1}^M j_\ell\in\mathcal{H}_{N,\varepsilon^4}$. Similarly if $j_\ell=i_\ell$ except at one place where $i_\ell$ is bad for reduction and $j_\ell$ is good for reduction, then $\sum_{\ell=1}^M j_\ell\in\mathcal{H}_{\tilde{N},\tilde{\varepsilon}^4}$.

*Proof.* We prove the first case and the second is similar. For concreteness we assume that $j_1\neq i_1$. So $\sum_{\ell=1}^M j_\ell=j_1-i_1+\sum_{\ell=1}^M i_\ell$ satisfies the assumptions of Lemma 6.6, completing the proof. □

Set $A = \{i\colon i \text{ is good for reduction}\}$, set $B = \{i\colon i \text{ is bad for reduction}\}$ and set $C = \mathbb{Z} \setminus (A \cup B)$. We define two closely related partitions $\mathcal{P}$ and $\mathcal{P}'$ of $\mathbb{Z}^M$. We index the partition elements of both $\mathcal{P}$ and $\mathcal{P}'$ by elements of $A^a \times B^b \times C^c$, where $a, b, c \geq 0$ and $a + b + c = M - 1$. Given some triple $(\vec{x}, \vec{y}, \vec{z}) \in A^a \times B^b \times C^c$, let $P_{(\vec{x},\vec{y},\vec{z})}$ be the set of all $M$-tuples $(i_1, \dots, i_M)$ such that there exists $e_1 < \dots < e_a$, $f_1 < \dots < f_b$, $g_1 < \dots < g_c$ satisfying that for each $\ell$ in the allowed ranges, $i_{e_\ell} = x_\ell$, $i_{f_\ell} = y_\ell$ and $i_{g_\ell} = z_\ell$. Moreover if $j$ is the unique element of $\{1, \dots, M\} \setminus \{e_1, \dots, e_a, f_1, \dots, f_b, g_1, \dots, g_c\}$ then $j > e_a$ and $i_j \in A$. Set

$$\mathcal{P} = \{P_{(\vec{x},\vec{y},\vec{z})} : (\vec{x}, \vec{y}, \vec{z}) \in A^a \times B^b \times C^c \text{ where } 0 \leq a, b, c \text{ and } a + b + c = M - 1\}.$$

We give another way to describe this. The partition elements are subsets of $\mathbb{Z}^M$ so that each $M$-tuple has $a + 1$ terms in $A$, $b$ terms in $B$ and $c$ terms in $C$. We fix all the terms that are in $B$ and the order they come in relative to the other terms that are in $B$ (but not relative to the terms that are in $A$ and $C$), and similarly for $C$. For $A$, we fix all but the last term that is in $A$ that appears and their order relative to the other terms of $A$ (but not relative to the terms of $B$ and $C$). The last term that is in $A$ which appears is allowed to be any element of $A$.

We now define $\mathcal{P}'$ by switching the roles of $A$ and $B$. That is, we define $P'_{(\vec{x},\vec{y},\vec{z})}$, to be the set of all $(i_1, \dots, i_M)$ such that there exists $e_1 < \dots < e_a$, $f_1 < \dots < f_b$, $g_1 < \dots < g_c$ such that for each $\ell$ in the allowed ranges, $i_{e_\ell} = x_\ell$, $i_{f_\ell} = y_\ell$, and $i_{g_\ell} = z_\ell$. Moreover if $j$ is the unique element of $\{1, \dots, M\} \setminus \{e_1, \dots, e_a, f_1, \dots, f_b, g_1, \dots, g_c\}$, then $j > f_b$ and $i_j \in B$. Set

$$\mathcal{P}' = \{P'_{(\vec{x},\vec{y},\vec{z})} : (\vec{x}, \vec{y}, \vec{z}) \in A^a \times B^b \times C^c \text{ where } 0 \leq a, b, c \text{ and } a + b + c = M - 1\}.$$

**Lemma 6.8.** *For any $a+b+c = M-1$, $(\vec{x}, \vec{y}, \vec{z}) \in A^a \times B^b \times C^c$, if any element of $P_{(\vec{x},\vec{y},\vec{z})}$ is good for reduction, then no element of $P'_{(\vec{x},\vec{y},\vec{z})}$ is good for reduction, and the analogous statement holds when the roles of $P_{(\vec{x},\vec{y},\vec{z})}$ and $P'_{(\vec{x},\vec{y},\vec{z})}$ are exchanged.*

*Proof.* Let $(i_1, \dots, i_M), (i'_1, \dots, i'_M) \in P_{(\vec{x},\vec{y},\vec{z})} \cup P'_{(\vec{x},\vec{y},\vec{z})}$. There exists a permutation $\pi$ such that $i_j = i'_{\pi(j)}$ except for at most one $j$. Moreover, if $k \in P_{(\vec{x},\vec{y},\vec{z})}$ and $\ell \in P'_{(\vec{x},\vec{y},\vec{z})}$, then this is a change from a good for reduction element to a bad for reduction element. By Corollary 6.7 if $\sum_{j=1}^M i_j$ is good for reduction, then $\sum_{j=1}^M i'_j$ is not good for reduction. Thus if there exists one element in $P_{(\vec{x},\vec{y},\vec{z})}$ that was good for reduction, then this argument shows every $\ell \in P'_{(\vec{x},\vec{y},\vec{z})}$ is not good for reduction and similarly vice versa. $\square$

We define one more partition $\mathcal{N}$ of $\mathbb{Z}^M$. If $u, v, w \geq 0$ and $u + v + w = M$, let

$$\mathcal{N}_{u,v,w} = \big\{(i_1, \dots, i_M)\colon |\{j\colon i_j \in A\}| = u,\ |\{j\colon i_j \in B\}| = v \text{ and } |\{j\colon i_j \in C\}| = w\big\}.$$

**Proposition 6.9.** *If the conditional probability that $\sum_{j=1}^M i_j$ is good for reduction given that $(i_1, \dots, i_M) \in \mathcal{N}_{(u,v,w)}$ is greater than $\frac{1}{2}$, then the conditional probability that $\sum_{j=1}^M i_j$ is not good for reduction given that $(i_1, \dots, i_M) \in \mathcal{N}_{(u+1,v-1,w)}$ is at least $\frac{1}{2}$*

Observe that both $\mathcal{P}$ and $\mathcal{P}'$ are finer partitions than $\mathcal{N}$.

**Lemma 6.10.** *For $(\vec{x}, \vec{y}, \vec{z}) \in A^a \times B^b \times C^c$, the conditional probability of an element in $\mathcal{N}_{(a+1,b,c)}$ is in $\mathcal{P}_{(\vec{x},\vec{y},\vec{z})}$ is the same as the conditional probability of an element in $\mathcal{N}_{(a,b+1,c)}$ being in $\mathcal{P}'_{(\vec{x},\vec{y},\vec{z})}$.*

*Proof.* Let $u_1 = \sum_{i\in A} \alpha_i^{(\tilde{k})}$, $u_2 = \sum_{i\in B} \alpha_i^{(\tilde{k})}$, and $u_3 = 1-(a+b) = \sum_{i\in C} \alpha_i^{(\tilde{k})}$. The conditional probability of being in a particular $P_{(\vec{x},\vec{y},\vec{z})}$ with $(\vec{x},\vec{y},\vec{z}) \in A^a \times B^b \times C^c$ given that one lies in $\mathcal{N}_{(a+1,b,c)}$ is

$$\prod_{i=1}^{a} \frac{v_i}{u_1} \prod_{i=1}^{b} \frac{w_i}{u_2} \prod_{i=1}^{c} \frac{x_i}{u_3}. \tag{58}$$

This is also the conditional probability of being in $P'_{(\vec{x},\vec{y},\vec{z})}$ given that one is in $\mathcal{N}_{(a,b+1,c)}$. □

Combining Lemmas 6.10 and 6.8, Proposition 6.9 follows.

If $(\Omega,\mathbb{P})$ is a probability space and $H\colon \Omega \to \{0,1,\ldots\}$ is $\mathbb{P}$ measurable, we say $i$ is $(H,\delta)$-*spread* if

$$\max\{\mathbb{P}(H^{-1}(i+1)), \mathbb{P}(H^{-1}(i-1))\} > \delta\mathbb{P}(H^{-1}(i)).$$

We say $H$ is $\delta$-*spread* if $\mathbb{P}(\bigcup_{i,\,(H,\delta)\text{-spread}} H^{-1}(i)) > \delta$.

**Lemma 6.11.** *There exists $C$ such that if $F_i\colon (\Omega,\mathbb{P}) \to \{0,1\}$ are independent, identically $\mathbb{P}$ distributed random variables satisfying $\frac{\delta}{K} \le \mathbb{P}(F_i^{-1}(0)) \le 1-\frac{\delta}{K}$, then $H(\omega) = \sum_{i=1}^K F_i(\omega)$ is $\min\{\frac{\delta^2}{C}, \frac{1}{C}\}$-spread.*

*Proof.* If $\delta < \frac{9}{K}$, this is straightforward and so we assume $\delta > \frac{9}{K}$. Let $p = \mathbb{P}(F_i^{-1}(0))$. Due to the symmetry, we can assume that $p \le \frac{1}{2}$, and by the assumption on $\delta$, we can assume that $p > \frac{9}{K}$. Thus

$$\begin{aligned}\frac{\mathbb{P}(\{\omega\colon \sum_{i=1}^K F_i(\omega) = n+1\})}{\mathbb{P}(\{\omega\colon \sum_{i=1}^K F_i(\omega) = n\})} &= \frac{\binom{K}{n+1}(p)^{n+1}(1-p)^{K-n-1}}{\binom{K}{n}(p)^n(1-p)^{K-n}} \binom{K}{n}(p/K)^n(1-(p/K))^{K-n} \\ &= \frac{K-n-1}{n-1}p(1-p)^{-1}.\end{aligned}$$

If $n \in [\frac{1}{3}Kp, \frac{5}{3}Kp]$, then this is greater than $\min\{\frac{1}{99}, \frac{1}{99}Kp\}$. Since $\frac{1}{99}Kp \ge \frac{1}{99}\delta$ the result follows if at least half of the $\omega$ lie in this range. To check this, note that we have

$$\int_\Omega (\sum_{i=1}^K F_i(\omega) - p)^2\, d\mathbb{P} = K\big((1-\frac{p}{K})^2\frac{K}{p} + (\frac{p}{K})^2(1-\frac{p}{K})\big) < \frac{3}{2}pK.$$

Thus by Chebyshev's inequality, $\mathbb{P}(\{\omega\colon |\sum_{i=1}^K F_i(\omega) - Kp| > 2\sqrt{pK}\}) < \frac{1}{2}$. Since $pK \ge 9$ we have that $2\sqrt{pK} \le \frac{2}{3}Kp$ establishing the necessary condition. □

*Concluding the proof of Theorem 6.1 in Case 3.* In this proof only, we introduce some terminology for clarity: we say $i$ is *decisive* if it is either good or bad for reduction. Let $\gamma^{(\tilde{k})}$ be the probability measure on $\{0,1\}$ defined by

$$\gamma^{(\tilde{k})}(\{0\}) = \frac{\sum_{i \text{ is bad for reduction}} \alpha_i^{(\tilde{k})}}{\sum_{i \text{ is decisive}} \alpha_i^{(\tilde{k})}}$$

and

$$\gamma^{(\tilde{k})}(\{1\}) = \frac{\sum_{i \text{ is good for reduction}} \alpha_i^{(\tilde{k})}}{\sum_{i \text{ is decisive}} \alpha_i^{(\tilde{k})}}.$$

Note that $\gamma^{(\tilde{k})}(\{0\})$ is the conditional probability that $i$ is *bad* for reduction given that it is decisive and $\gamma^{(\tilde{k})}(\{1\})$ is the conditional probability that $i$ is *good* for reduction given that it is decisive. As we are not in Case 2, it follows that $\gamma^{(\tilde{k})}(\{0\}) > \frac{\tilde{\varepsilon}^2}{M}$. Thus by Lemma 6.11, $\sum_{j=0}^{n} \gamma^{(\tilde{k})}$ is at least $\frac{\tilde{\varepsilon}^4}{64}$ spread. We partition $\mathbb{Z}^M$ into sets $\mathcal{N}_{(a,b,c)} \cup \mathcal{N}_{(a+1,b-1,c)}$ where $a$ is even. This gives rise to partitions of these elements into $P_{(\vec{x},\vec{y},\vec{z})}$ and $P'_{(\vec{x},\vec{y},\vec{z})}$. By Lemma 6.8, for each $(\vec{x},\vec{y},\vec{z})$ one of $P_{(\vec{x},\vec{y},\vec{z})}$ or $P'_{(\vec{x},\vec{y},\vec{z})}$ contain no good for reduction elements. By Lemma 6.10 and the fact that $\sum_{i=0}^{M-c-1} \gamma^{(\tilde{k})}$ is $\tilde{\varepsilon}^4$ spread (so long as $c < \frac{M}{8}$) it follows from Lemma 6.5 that at least $\frac{1}{2}\tilde{\varepsilon}^8$ of the points in $\mathcal{N}_{(a,b,c)} \cup \mathcal{N}_{(a+1,b-1,c)}$ are not good for reduction. Once again this contradicts Corollary 5.16 and our choices. $\square$

## References


[1] O. N. Ageev. The generic automorphism of a Lebesgue space conjugate to a G-extension for any finite abelian group G. *Dokl. Akad. Nauk.* **374** (2000), no. 4, 439–442.

[2] O. N. Ageev. A typical dynamical system is not simple or semisimple. *Ergodic Theory Dynam. Systems* **23** (2003), no. 6, 1625–1636

[3] J. Chaika. Self joinings of rigid rank one transformations arise as strong operator topology limits of convex combinations of powers. `arXiv:1901.08695`

[4] J. Chaika and A. Eskin. Self-joinings for 3-IETs. `arXiv:1805.11167`

[5] A. I. Danilenko On simplicity concepts for ergodic actions. *J. Anal. Math.* **102** (2007), 77–117.

[6] A. I. Danilenko and A. del Junco. Cut-and-stack simple weakly mixing map with countably many prime factors. *Proc. Amer. Math. Soc.* **136** (2008), no. 7, 2463–2472.

[7] A. del Junco. A simple map with no prime factors. *Israel J. Math.* **104** (1998), 301–320.

[8] A. del Junco and D. Rudolph. A rank-one, rigid, simple, prime map. *Ergodic Theory Dynam. Systems* **7** (1987), no. 2, 229–247.

[9] A. del Junco, M. Rahe, and L. Swanson. Chacon's automorphism has minimal self-joinings *J. Analyse Math.* **37** (1980), 276–284

[10] H. Furstenberg. The structure of distal flows. *Amer. J. Math.* **85** (1963), 477–515.

[11] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory.* Princeton University Press, Princeton 1981.

[12] E. Glasner. *Ergodic Theory via Joinings.* Mathematical Surveys and Monographs, **101**. American Mathematical Society, Providence, RI, 2003

[13] E. Glasner, B Host, and D. Rudolph. Simple systems and their higher order self-joinings. *Israel J. Math.* **78** (1992), no. 1, 131–142.

[14] E. Glasner and B. Weiss. A simple weakly mixing transformation with nonunique prime factors. *Amer. J. Math.* **116** (1994), no. 2, 361–375.

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

[16] P. Halmos. In general a measure preserving transformation is mixing. *Ann. of Math.* (2) **45** (1944), 786–792.

[17] É. Janvresse, E. Roy, and T. de la Rue. Poisson suspensions and Sushis. *Ann. Sci. Éc. Norm. Supér.* (4) **50** (2017), no. 6, 1301–1334.

[18] É. Janvresse, A. A. Prikhod'ko, T. de la Rue, and V. V. Ryzhikov. Weak limits of powers of Chacon's automorphism. *Ergodic Theory Dynam. Systems* **35**, no. 1, (2015), 128–141.

[19] É. Janvresse, T. de la Rue, and V. Ryzhikov. Around King's rank-one theorems: flows and n-actions. *Dynamical systems and group actions.* Contemp. Math. **567** Amer. Math. Soc., Providence, RI, (2012), 143–161.

[20] A. Katok and J. P. Thouvenot. Spectral properties and combinatorial constructions in ergodic theory. Handbook in Dynamical Systems, vol. 1B, Elsevier, (2005), 649–743.

[21] J. L. King. The commutant is the weak closure of the powers, for rank-1 transformations. *Ergod. Th. Dynam. Systems* **6** (1986), no. 3, 363–384.

[22] J. L. King. Joining-rank and the structure of finite rank mixing transformations. *J. Analyse Math.* **51** (1988), 182–227.
[23] J. King. Flat stacks, joining closure and genericity. Preprint.
[24] M. Lemanczyk, F. Parreau, and E. Roy. Joining primeness and disjointness from infinitely divisible systems. *Proc. Amerc. Math. Soc.* **139** (2011), no.1, 185–199.
[25] J. Lindenstrauss, G. Olsen, and Y. Sternfeld. The Poulsen simplex. *Ann. Inst. Fourier* (Grenoble) **28** (1978), no. 1, vi, 91–114.
[26] W. Parry. Zero entropy of distal and related transformations. Topological Dynamics (Symp. at Colorado State University, (1967)). Benjamin, New York, 1968.
[27] F. Parreau and E. Roy. Prime poisson suspensions. *Ergodic Theory Dynam. Systems* **35** (2015), no. 7, 2216–2230.
[28] D. Rudolph. An example of a measure preserving map with minimal self-joinings, and applications. *J. Analyse Math.* **35** (1979), 97–122.
[29] V. V. Ryzhikov. Bounded ergodic constructions, disjointness, and weak limits of powers. *Trans. Moscow Math. Soc.* 2013, 165-171.
[30] J.-P. Thouvenot. Les systèmes simples sont disjoints de ceux qui sont infiniment divisibles et plongeables dans un flot. *Colloq. Math.* **84/85** (2000), part 2, 481–483
[31] W. Veech. A criterion for a process to be prime. *Monatsh. Math.* **94** (1982), no. 4, 335–341.
[32] P. Walters. An introduction to ergodic theory. Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982.
[33] R. J. Zimmer. Extensions of ergodic actions and generalized discrete spectrum. *Bull. Amer. Math. Soc.* **81** (1975), 633–636.
[34] R. J. Zimmer. Ergodic actions with generalized discrete spectrum. *Illinois J. Math.* **20** (1976), no. 4, 555–588.

University of Utah, Salt Lake City, UT 84112
*E-mail address*: `chaika@math.utah.edu`

Northwestern University, Evanston, IL 60208
*E-mail address*: `kra@math.northwestern.edu`