ISRAEL JOURNAL OF MATHEMATICS, Vol. 44, No. 2, 1983

A FAMILY OF COUNTEREXAMPLES IN ERGODIC THEORY

BY

ANDRÉS DEL JUNCO[†]

ABSTRACT

Let T_α be the translation $x \mapsto x+\alpha$ (mod 1) of $[0,1)$, α irrational. Let T be the Lebesgue measure-preserving automorphism of $X=[0,3/2)$ defined by $Tx = x+1$ for $x \in [0,1/2)$, $Tx = T_\alpha(x-1)$ for $x \in [1,3/2)$ and $Tx = T_\alpha x$ for $x \in [1/2,1)$, i.e. T is T_α with a tower of height one built over $[0,1/2)$. If α is poorly approximable by rationals (there does not exist $\{p_n/q_n\}$ with $|\alpha - p_n/q_n| = o(q_n^{-2})$) and λ is a measure on X^k all of whose one-dimensional marginals are Lebesgue and which is $\bigotimes_{i=1}^k T^l$ invariant and ergodic ($l>0$) then λ is a product of off-diagonal measures. This property suffices for many purposes of counterexample construction. A connection is established with the POD (proximal orbit dense) condition in topological dynamics.

§1. Introduction

Ornstein [11] constructed an example of an automorphism of a probability space which commutes only with its powers and has only the trivial factor algebras. Other examples have been found since then — perhaps the simplest is the weak-mixing but not mixing automorphism of Chacón (see [5], [7]). All of these examples are constructed with malice aforethought, rather than being automorphisms which occur "naturally". It is reasonable to ask whether this behaviour, which may be regarded as pathological, can occur within some natural family of automorphisms, for example interval exchanges on $[0,1)$. Our purpose here is to give a natural family of examples — in fact they are interval exchanges on three intervals.

To describe this family, which was first considered in a topological context by Furtstenberg, Keynes and Shapiro [4], let T_α be the translation $x \mapsto x+\alpha$ (mod 1) on $[0,1)$ and let T be T_α delayed by one unit of time on $[0,1/2)$, or, in the

[†] Research supported in part by NSF contract MCS-8003038.
Received July 22, 1981 and in revised form August 25, 1982

language of skyscrapers, T_α with a second story added above $[0,1/2)$. Identifying $T[0,1/2)$ with $[1,3/2)$, T can be defined explicitly on $[0,3/2)$ by

$$Tx = \begin{cases} x+1, & x \in [0,1/2) \\ T_\alpha(x-1), & x \in [1,3/2) \\ T_\alpha x, & x \in [1/2,1). \end{cases}$$

Say that α is well approximable by rationals if there exist rationals p_n/q_n such that $|\alpha - p_n/q_n| = o(q_n^{-2})$ and poorly approximable otherwise. We will say that an automorphism S of a probability space $(Y, \mathcal{F}, \nu)$ is simple if any measure λ on $(Y^2, \mathcal{F}^2)$, both of whose marginals are ν, which is $S \times S$ invariant and ergodic must be $\nu \times \nu$ or an off-diagonal measure, that is, a measure of the form $\nu_k(A) = \mu\{x : (x, S^k x) \in A\}$. Simplicity is sufficient to show that S is prime and has trivial centralizer (see [12]). We will show that if α is poorly approximable then T is simple. Thus, for example, one can take α a quadratic surd, say $\sqrt{2}$, and T will be prime with trivial centralizer. It is well known that there are uncountably many poorly approximable α (it also follows from our Lemma 2.3) although the set of such has measure 0.

Simplicity of an automorphism is just the lowest level of the property of minimal self-joinings (MSJ) introduced by Rudolph (see [12] for the definition). T is not MSJ because it is clearly isomorphic to its inverse. We can show however that if α is poorly approximable and $l > 0$, then any $\bigotimes_{i=1}^k T^l$ invariant and ergodic measure on X^k with marginals μ must be a product of off-diagonals. An off-diagonal measure ν on $X^{k'}$ is one of the form $\nu(A) = \bigotimes_{i=1}^{k'} T^{l'(i)} \mu_\Delta$ where μ_Δ is diagonal measure on $X^{k'}$. The main interest of an MSJ automorphism is its versatility as a source of counterexample constructions (see [12]) and we observe that this restricted version of MSJ is sufficient for most such constructions. In fact it follows from results in [6] that any measure on X^k with marginals μ which is $\bigotimes_{i=1}^k T^{l(i)}$ invariant and ergodic, where $l(i) > 0\ \forall i$, must be a product of off-diagonals.

In [4] a topological analogue of simplicity called proximal orbit density (POD) was defined and established for (a topological model of) T whenever α is irrational (and also for more general "holding" sets $[0,\beta)$, $\beta \neq n\alpha$, in place of $[0,1/2)$). The definition of POD predates any explicit mention of simplicity in ergodic theory. We establish a sufficient condition, satisfied by the topological model of T when α is poorly approximable, for simplicity to imply POD.

In section 2 we describe symbolic dynamics for T following [10] and [1] which reduce the study of T to a shift. We then show that T is weakly-mixing for any irrational α. Although this is essentially known, even in a more general context (it follows from results in [13], [14]) we include a proof here using the block structure of T-names, as a warm-up for the arguments of sections 3 and 4. In section 3 we show that T is simple for any poorly approximable α. Section 4 extends this to higher cartesian powers. In Section 5 we establish some connections with the POD condition and conclude with some open problems.

We shall assume throughout the remainder of the paper that $\alpha < 1/2$. There is no loss of generality since the T's corresponding to α and to $1-\alpha$ are isomorphic in an obvious way.

§2. Symbolic dynamics for T and weak-mixing

In this section we will analyze the structure of T-names. This analysis borrows heavily from [10] and [1]. As shown in [1] the structure of T_α-names depends on the modified continued fraction expansion (m.c.f.e.) of $\beta = 2\alpha$. To define this expansion, for $x \in (0,1]$ let

$$n(x) = \left[\frac{1}{x}\right] + 1, \qquad s(x) = n(x) - \frac{1}{x},$$

so $s(x) \in (0,1]$. Set $n_k(x) = n(s^k x)$, $k \geqq 0$. In [10] it is shown that the partial quotients

$$1/n_0 - 1/n_1 - 1/n_2 \cdots - 1/n_k$$

converge to x as $k \to \infty$. The m.c.f.e. of x is

$$1/n_0 - 1/n_1 - \cdots.$$

Note that $n_k(x) \geqq 2$. Now set $\beta = 2\alpha$ and $n_k = n_k(\beta)$.

We begin our analysis with T_β-names, for which we use the partition $\{P_0, P_1\}$,

$$P_0 = [0, 1-\beta), \qquad P_1 = [1-\beta, 1).$$

By the T_β-name of $x \in (0,1]$ we mean the sequence $\xi \in \{0,1\}^{\mathbf{Z}}$ defined by $\xi(i) = \delta$ if $T_\beta^i(x) \in P_\delta$. Define inductively the k-blocks $B_k(0)$ and $B_k(1)$ of 0's and 1's, $k = 1, 2, \cdots$, by

$$B_1(0) = 0^{n_0 - 1}1, \qquad B_1(1) = 0^{n_0 - 2}1,$$

$$B_{k+1}(0) = B_k(0)^{n_k - 1} B_k(1), \qquad B_{k+1}(1) = B_k(0)^{n_k - 2} B_k(1).$$

Note that if $n_{k+1} = 2$, $B_{k+1}(1) = B_k(1)$. We refer to $B_k(0)$ and $B_k(1)$ as T_β-k-blocks.

LEMMA 2.1. *Every T_β-name can be decomposed, in a unique way, as a concatenation of T_β-k-blocks. If $1B_k(\delta)$, $\delta = 0$ or 1, occurs in a T_β-name then the $B_k(\delta)$ is one of the blocks in this decomposition.*

PROOF. To avoid typographical difficulties in this proof we will write $\tau = T_\beta$ when dealing with induced automorphisms. We have $n_0 - 1 \leqq 1/\beta < n_0$ so

(i) $(n_0 - 1)\beta \leqq 1 < n_0\beta$.

Since $\beta n_0 - 1 = \beta s(\beta)$, (i) implies that the induced automorphism $\tau_{[0\beta)}$ is given by

$$\tau_{[0\beta)}(x) = x + \beta s(\beta) \pmod{\beta}.$$

Let $r_{[0,\beta)}$ denote the T_β-return time function to $[0,\beta)$. By the T_β-n-name of x we mean the T_β-name restricted to $[0, n-1]$. (i) implies that for $x \in [\beta - \beta s(\beta))$ the T_β-$r_{[0\beta)}(x)$-name of x is $B_1(1)$, while for $x \in [0, \beta - \beta\sigma(\beta))$ it is $B_1(0)$. This implies that any T_β-name is a concatenation of 1-blocks, and that if $B_1(0)$ or $B_1(1)$ occurs at i in the T_β-name of x and is preceded by a 1 then $T^i_\beta x \in [0, \beta - \beta\sigma(\beta))$ or $[\beta - \beta\sigma(\beta), \beta)$ respectively.

Now set $\beta_k = \beta s(\beta) \cdots s^k(\beta)$ and suppose that for some k we have shown

$I(k)$: $\tau_{[0,\beta_k)}(x) = x + \beta_{k+1} \bmod \beta_k$,

$J(k)$: for $x \in [0, \beta_k - \beta_{k+1})$ the T_β-$r_{[0,\beta_k)}(x)$-name of x is $B_k(0)$ and for $x \in [\beta_k - \beta_{k+1}, \beta_k)$ it is $B_k(1)$.

$K(k)$: if $B_k(0)$ or $B_k(1)$ occurs at i in the T_β-name of x and is preceded by a 1 then $T^i_\beta x \in [0, \beta_k - \beta_{k+1})$ or $[\beta_k - \beta_{k+1}, \beta_k)$.

(We have just shown $I(1)$, $J(1)$ and $K(1)$.) Note that $J(k)$ implies a decomposition of any T_β-name into k-blocks and $K(k)$ implies the desired uniqueness of that decomposition. By $I(k)$ and the analysis of the previous paragraph, applied to $s^{k+1}\beta$ rather than β and scaled down to $[0, \beta_k)$, we see that

$$\tau_{[0,\beta_{k+1})}(x) = x + \beta_k s^{k+1}(\beta) s(s^{k+1}\beta) = x + \beta_{k+2},$$

so we have $I(k+1)$. Moreover if we consider $\tau_{[0,\beta_k)}$-names with respect to $\{P^k_0 = [0, \beta_k - \beta_{k+1}),\ P^k_1 = [\beta_k - \beta_{k+1}, \beta_k)\}$ and denote by $r^k_{[0,\beta_{k+1})}$ the return time to $[0, \beta_{k+1})$ under $\tau_{[0,\beta_k)}$, again as in the last paragraph we see that for $x \in [0, \beta_{k+1} - \beta_{k+2})$ this $\tau_{[0,\beta_k)}$-$r^k_{[0,\beta_{k+1})}(x)$-name of x is $0^{n_k - 1}1$ while for $x \in [\beta_{k+1} - \beta_{k+2}, \beta_{k+1})$ it is $0^{n_k - 2}1$. By $J(k)$ we get the T_β-$r_{[0,\beta_{k+1})}(x)$-name of x by replacing 0's and 1's in the $\tau_{[0,\beta_k)}$-$r^k_{[0,\beta_{k+1})}(x)$-name by $B_k(0)$ and $B_k(1)$. In view of the definition of $B_{k+1}(0)$ and $B_{k+1}(1)$ this means that $J(k+1)$ holds. Finally if

$B_{k+1}(\delta)$, $\delta = 0$ or 1, occurs at 0 in the T_β-name of x and is preceded by a 1, then by $K(k)$ its component k blocks are blocks of the decomposition implied by $J(k)$, and $0^{n_{k+1}-1}1$ or $0^{n_{k+1}-2}1$ occurs at 0 in the $\tau_{[0,\beta_k)}$-name of x and is preceded by a 1. Thus we see that $x \in [0, \beta_{k+1} - \beta_{k+2})$ or $[\beta_{k+1} - \beta_{k+2}, \beta_k)$, which gives $K(k+1)$. □

We shall refer to k-blocks in the k-block decomposition of a T_β-name as genuine k-blocks. Note that for example $B_k(1)$ occurs "falsely" in $B_k(0)$.

We will need characterizations of irrationality and poor approximability of α (equivalently of $\beta = 2\alpha$) in terms of the m.c.f.e. of β. The first of these can be found in [10].

LEMMA 2.2 (Keane). *β is rational if and only if $n_k = 2$ for all sufficiently large k.*

LEMMA 2.3. *β is well approximable if and only if $\{n_k\}$ is unbounded or contains arbitrarily long runs of 2's.*

PROOF. For $x \in (0,1]$, x is small if and only if $n(x)$ is large. On the other hand if $1-x$ is small $n(x)$ is 2 and $1 - s(x)$ is again small, so $\{n_k(x)\}$ begins with a long run of 2's. Also, if $x < 1 - \delta$ and $x > \frac{1}{2}$ then $s(x) = 2 - 1/x < 1 - \delta/x < 1 - 2\delta$. It follows that if x begins with a long run of 2's then $1 - x$ is small. Summarizing, what we have to show is that β is well approximable if and only if $\underline{\lim} \| s^k(\beta) \| = 0$, where $\| x \| = \min(\{x\})$ for any $x \in \mathbf{R}$, $\{x\}$ the fractional part of x.

Now suppose β is well approximable, $\delta \in (0,1)$ and choose ε such that

$$\delta > 2\varepsilon \prod_{j=2}^{\infty} \frac{j^2}{j^2 - 1}.$$

(In particular $1 > \delta > \varepsilon$.) Now find p/q such that $|\beta - p/q| < \varepsilon/q^2$ or, equivalently, find a q such that $\| q\beta \| < \varepsilon/q$. We want to show $\| s^k(\beta) \| < \delta$ for some k and we may suppose $p < q$, otherwise $\| \beta \| < \varepsilon/q^2 < \delta$ and we're done. Then we have

$$\text{(i)} \qquad \left| \frac{1}{\beta} - \frac{q}{p} \right| < \frac{\varepsilon}{\beta pq} < \frac{\varepsilon}{\left(\frac{p}{q} - \frac{\varepsilon}{q^2} \right) pq} \leqq \frac{\varepsilon}{p^2 - \frac{p}{q}} < \frac{\varepsilon}{p^2 - 1},$$

since $\varepsilon < 1$. Setting

$$q_1 = p < q \text{ and } \varepsilon_1 = \begin{cases} \dfrac{\varepsilon p^2}{p^2 - 1} & \text{if } p > 1 \\[2ex] \dfrac{\varepsilon}{p^2 - \frac{p}{q}} \leqq 2\varepsilon & \text{if } p = 1 \end{cases},$$

(i) implies $\|q_1/\beta\| < \varepsilon_1/q_1$ so

$$\|q_1 s(\beta)\| < \varepsilon_1/q_1.$$

Applying the same argument to $s(\beta)$ and continuing in this way we obtain

$$q = q_0 > q_1 > q_2 > \cdots > q_k = 1$$

and $\varepsilon = \varepsilon_0, \varepsilon_1, \cdots, \varepsilon_k$ such that

$$\varepsilon_i = \frac{\varepsilon_{i-1} q_i^2}{q_i^2 - 1} \qquad \text{for } i < k, \quad \varepsilon_k < 2\varepsilon_{k-1},$$

and $\|q_i s^i(\beta)\| < \varepsilon_i/q_i$. Thus for $i < k$

$$\varepsilon_i = \varepsilon \prod_{j=1}^{i} \frac{q_j^2}{q_j^2 - 1} < \varepsilon \prod_{j=2}^{\infty} \frac{j^2}{j^2 - 1} < \frac{\delta}{2}.$$

In particular $\varepsilon_i < \delta/2 < 1$ which is necessary to continue the argument at each stage. Also $\varepsilon_k < 2\varepsilon_{k-1} < \delta$ so

$$\|s^k(\beta)\| = \|q_k s^k(\beta)\| < \delta/q_k = \delta,$$

as desired.

For the converse suppose that $\underline{\lim}\|s^k(\beta)\| = 0$, $\delta \in (0,1)$ and choose ε such that

$$\varepsilon \prod_{t=1}^{\infty} (1 - [t(t+1)]^{-1})^{-1} < \delta$$

and k such that $\|s^k(\beta)\| < \varepsilon$. Note that if $x \in (0,1]$ and $\|qs(x)\| < \varepsilon/q$ with $\varepsilon < 1$ then $\|q/x\| < \varepsilon/q$ so there is a p such that

$$\left| q\frac{1}{x} - p \right| < \frac{\varepsilon}{q},$$

and necessarily $p \geqq q$ since $\varepsilon < 1$. If $p > q$ we have

$$|q - px| < \frac{\varepsilon}{q\frac{1}{x}} < \frac{\varepsilon}{q\left(\frac{p}{q} - \frac{\varepsilon}{q^2}\right)} < \frac{\varepsilon}{p - \frac{1}{q}}.$$

Thus, setting $\varepsilon' = \varepsilon p/(p - 1/q)$ and $q' = p$ we have

$$\|q'x\| < \varepsilon'/q'.$$

If $p = q$

(ii) $|q - qx| < \varepsilon x/q < \varepsilon/q$

so $\|qx\| < \varepsilon/q$ so we may take $q' = q$, $\varepsilon' = \varepsilon$ to get (ii). Now set $q_0 = 1$, $\varepsilon_0 = \varepsilon$, so

$\|q_0 s^k(\beta)\| < \varepsilon / q_0$. By the above remarks we obtain $q_0 \leqq q_1 \cdots \leqq q_k$ and $\varepsilon_0, \varepsilon_1, \cdots, \varepsilon_k$ such that

$$\varepsilon_{i+1} = \frac{\varepsilon_i q_{i+1}}{q_{i+1} - \dfrac{1}{q_i}} \quad \text{if } q_{i+1} > q_i, \qquad \varepsilon_{i+1} = \varepsilon_i \quad \text{if } q_{i+1} = q_i,$$

$$\|q_i s^{k-i}(\beta)\| < \varepsilon_i / q_i.$$

Thus

$$\varepsilon_r = \varepsilon \prod_{\substack{i=0 \\ q_{i+1} > q_i}}^{r-1} \left(1 - \frac{1}{q_{i+1} q_i}\right)^{-1} < \varepsilon \prod_{t=1}^{\infty} [1 - (t(t+1))^{-1}]^{-1} < \delta.$$

In particular each $\varepsilon_r < 1$ so the argument can be continued. Taking $r = k$

$$\|q_k \beta\| < \delta / q_k.$$

Since δ is arbitrary, β is well approximable. □

If ξ is a finite string we denote its length by $|\xi|$.

LEMMA 2.4. *If β is poorly approximable there is a $c > 0$ such that $|B_k(1)| > (\frac{1}{2} + 2c)|B_k(0)|$ for all k such that $n_{k-1} > 2$.*

PROOF. First observe that if $n_{k-1} > 2$ then $|B_k(1)| > \frac{1}{2}|B_k(0)|$. If k^* is the least integer greater than k for which $n_{k^*-1} > 2$ then $k^* - k$ is bounded (Lemma 2.3) and $B_{k^*-1}(1) = B_k(1)$. Since $\{n_k\}$ is also bounded there is a $c' > 0$ such that $|B_{k^*-1}(1)| > \frac{1}{2}|B_k(0)| > c'|B_{k^*-1}(0)|$. This implies that there is a $c > 0$ such that $|B_{k^*}(1)| > (\frac{1}{2} + 2c)|B_{k^*}(0)|$. □

Next we look at T_α-names, for which we use the partition $\{Q_1, Q_{-1}\}$, $Q_1 = [0, \frac{1}{2})$, $Q_{-1} = [\frac{1}{2}, 1)$. Define k-blocks $A_k(0)$, $A_k(-0)$, $A_k(1)$ and $A_k(-1)$ $(0 \neq -0!)$ by

$$A_1(0) = 1^{n_0}, \quad A_1(-0) = -A_1(0),$$

$$A_1(1) = 1^{n_0 - 1}, \qquad A_1(-1) = -A_1(1),$$

$$A_{k+1}(0) = A_k(\varepsilon_1 0) A_k(\varepsilon_2 0) A_k(\varepsilon_3 0) \cdots A_k(\varepsilon_{n_k - 1} 0) A_k(\varepsilon_{n_k} 1),$$

$$A_{k+1}(-0) = -A_{k+1}(0),$$

$$A_{k+1}(1) = A_k(\varepsilon_1 0) A_k(\varepsilon_2 0) A_k(\varepsilon_3 0) \cdots A_k(\varepsilon_{n_k - 2} 0) A_k(\varepsilon_{n_k - 1} 1),$$

$$A_{k+1}(-1) = -A_{k+1}(1),$$

$$A_{k+1}(-1) = -A_{k+1}(1),$$

where the signs ε_i are given by $\varepsilon_i = (-1)^{p(i-1)}$, $p = 0$ or 1 according to whether $A_k(0)$ ends in $+1$ or -1. In other words, in $k+1$-blocks there is always a sign change across the end of a k-block. We adopt the convention that when we write a string of k-blocks

$$A_k(\pm\delta_1)A_k(\pm\delta_2)\cdots A_k(\pm\delta_r), \qquad \delta_i = 0 \text{ or } 1$$

the leftmost sign is arbitrary and the others are chosen so that there is a sign change across the end of each k-block. We now have two types of k-blocks, the A_k's and the B_k's, but no confusion should arise if we keep in mind that the B_k's occur in T_β-names and the A_k's, as we shall see, occur in T_α-names. When it is necessary to make a distinction we will speak of T_β-k-blocks and T_α-k-blocks.

LEMMA 2.5. *Every T_α-name is for each k uniquely a concatenation of k-blocks, in such a way that a sign change occurs across the end of each k-block. If $-1A_k(0)\pm 1$ or $-1A_k(1)\pm 1$ appears in a T_α-name the $A_k(0)$ or $A_k(1)$ is one of the k-blocks of this unique decomposition. The same holds for appearances of $1A_k(-0)\pm 1$ and $1A_k(-1)\pm 1$. Any $A_k(\pm 0)$ or $A_k(\pm 1)$ in the k-block decomposition of a name is preceded and followed by $A_k(\pm 1)$'s, which are not necessarily part of the decomposition, but whose sign is governed by the requirement that there be a sign change across the end of a k-block.*

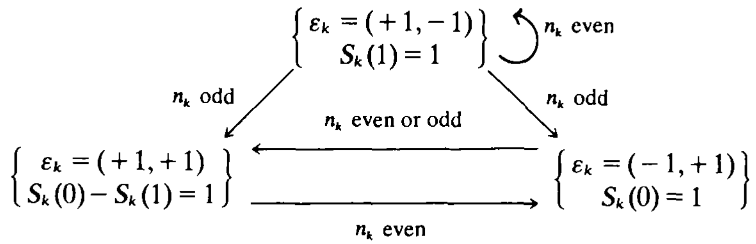
PROOF. Write $P(x) = \varepsilon$ if $x \in P_\varepsilon$. Let $\varphi : x \mapsto 2x \pmod 1$ so that $\varphi T_\alpha = T_\beta \varphi$. Note that $P(x) \neq P(T_\alpha x)$ precisely when $x \in [\frac{1}{2} - \alpha, \frac{1}{2}) \cup [1-\alpha, 1)$, that is $\varphi(x) \in [1-\beta, 1)$. (Recall that $\alpha < \frac{1}{2}$.) Thus $P(T_\alpha^i x) \neq P(T_\alpha^{i+1} x)$ precisely when $T_\beta^i \varphi(x) \in [1-\beta, 1)$, that is when the T_β-name of $\varphi(x)$ has a 1 in the ith co-ordinate.

Now let ξ be the T_α-name of x and η the T_β-name of $\varphi(x)$. Since $B_1(0)$ and $B_1(1)$ have 1's only in the rightmost position we see that where there is a genuine $B_1(0)$ or $B_1(1)$ in η there is a $A_1(\pm 0)$ or $A_1(\pm 1)$ in ξ. The converse is false — $A_1(1)$ appears inside $A_1(0)$ — but we do have that if $-1A_1(0)-1$ or $-1A_1(1)-1$ appears in ξ the $A_1(0)$ or $A_1(1)$ corresponds to a genuine 1-block in η. One then argues by induction that genuine k-blocks in η correspond to k-blocks in ξ (and there is a sign change across the end of each of these k-blocks in ξ because k-blocks in η end in 1) and that any appearance of a k-block in ξ with its flanking ± 1's must correspond to a genuine k-block in η. Thus we can also speak of genuine k-blocks in ξ.

It is immediate by induction that $A_k(0)$ begins with $A_k(1)$ and (without induction) that $A_k(0)$ ends with $A_k(\pm 1)$. It follows that a genuine $A_k(\pm 0)$ or $A_k(\pm 1)$ (in fact it is only $A_k(\pm 1)$'s which may be "false") is flanked by (possibly false) $A_k(\pm 1)$'s whose sign is evidently governed as claimed. □

We would like to know how the signs of the k-blocks in a T_α-name succeed each other. Since $A_k(0)$ and $A_k(1)$ begin with 1 what we need to know for each k is the vector $\varepsilon_k = (\varepsilon_k(0), \varepsilon_k(1))$, called the kth parity state, where $A_k(\delta)$ ends with $\varepsilon_k(\delta)$, $\delta = 0$ or 1. Denote by $S_k(\delta)$ the sum of the entries in $A_k(\delta)$. Our next lemma, taken from [1], is a diagram which shows how the n_k govern the parity state transitions (and in particular that $(-1,-1)$ is not a possible state) and also that each parity state entails a certain condition on $S_k(0)$ and $S_k(1)$. Note that $\varepsilon_1 = (+1,+1)$ and $S_1(0) - S_1(1) = 1$.

LEMMA 2.6.



For example, if $\varepsilon_k = (+1,+1)$ then $S_k(0) - S_k(1) = 1$ and if also n_k is odd then $\varepsilon_k = (+1,-1)$ (and $S_k(1) = 1$). Verifying the diagram is a direct check.

Finally we come to T-names for which we use the partition

$$R = \{R_1 = [0,\tfrac{1}{2}) \cup T[0,\tfrac{1}{2}), R_{-1} = [\tfrac{1}{2},1)\}.$$

R is a generator for T since Q is a generator for T_α. Let the k-block (for T-names) $C_k(\delta)$, $\delta = \pm 0$ or ± 1, be obtained by inserting $a + 1$ after each $+1$ in $A_k(\delta)$. Note that, for example, $C_k(-0) \neq -C_k(0)$. Evidently the $C_k(\delta)$'s obey the same inductive formation rule as the $A_k(\delta)$'s. Expressions like $C_k(\pm\delta_1)\cdots C_k(\pm\delta_r)$ obey the same convention that we adopted for the A_k's. Also $[C_k(\pm\delta)]^l$ is shorthand for $C_k(\pm\delta)\cdots C_k(\pm\delta)$ (l terms). Since T-names are obtained by doubling the 1's in T_α-names the following lemma is an immediate consequence of Lemma 2.5.

LEMMA 2.7. *Every T-name is, for each k, uniquely a concatenation of $C_k(\delta)$'s, $\delta = \pm 0, +1$. If $-1C_k(\delta)\pm 1$ appears in a T-name the $C_k(\pm\delta)$ is one of the k-blocks in the k-block decomposition (called a genuine k-block) and is preceded and followed by $C_k(\pm 1)$'s. The kth parity state governs the succession of signs of k-blocks in a T-name in the same way as in a T_α-name.*

We adopt the shorthand $C_k(\delta) = \delta_k$. Also if $C = C_k(\delta)$ we write $C^- = C_k(-\delta)$ $(C^- \neq -C)$. If $\xi = C_1 \cdots C_r$ is a string of r k-blocks we write $\xi_1^- = C_1^- \cdots C_r$. If ξ

is itself a k'-block, $k' > k$, $\xi = C_k(\delta)$ then $C_k(-\delta) = C_1^- \cdots C_r^-$, so there is no ambiguity in this notation. Observe that $S_k(\delta) = |\delta_k| - |-\delta_k|$ so that Lemma 2.6 translates into certain statements about the $|\delta_k|$, for example, $\varepsilon_k = (+1, +1)$ implies $|0_k - 1_k| - |-0_k\, 1_k| = 1$, which will be the key to our later arguments.

If $\{a_k\}$ and $\{b_k\}$ are sequences $a_k \sim b_k$ means $a_k / b_k \to 1$ and $a_k \lesssim b_k$ means $\lim a_k / b_k \leqq 1$.

LEMMA 2.8. *$|C_k(\delta)| \sim |C_k(-\delta)|$ for $\delta = 0$ or 1. If α is poorly approximable there are constants $0 < c$, $C < \infty$ such that $|1_k| > (\frac{1}{2} + 2c)|0_k|$ for each k such that $n_{k-1} > 2$, and the distance between successive $\pm 0_k$'s in $x \in X$ is bounded by $C|0_k|$ and also the distance between successive $\pm 1_k$'s is bounded by $C|0_k|$.*

PROOF. By unique ergodicity of T_α

$$|C_k(\delta)| \sim \tfrac{3}{2}|A_k(\delta)| = \tfrac{3}{2}|A_k(-\delta)| \sim |C_k(-\delta)|.$$

Thus the existence of c follows from Lemma 2.4. The existence of C is an easy consequence of the boundedness of $\{n_k\}$ and of runs of 2's in $\{n_k\}$. □

LEMMA 2.9. *If $C_1 \cdots C_r$ is a string of k-blocks the frequency of occurrence of $C_1 \cdots C_r$ among strings of r k-blocks in $x \in X$ is the same as that of $C_1^- \cdots C_r^-$.*

PROOF. By unique ergodicity of T_α any $x \in [0,1)$ is generic — that is, its T_α-name ξ has the right frequencies of occurrence of finite strings. But the T_α-name of $x + \frac{1}{2}$, which is also generic, is $-\xi$. This means that Lemma 2.9 is true for T_α-k-blocks in T_α-names and Lemma 2.9 itself follows immediately. □

If $x \in X$ the genuine k-block in x containing the 0th co-ordinate of x is called the time 0 k-block in x. If this block has length l the k-block level of x is the integer i, $0 \leqq i < l$ which gives the position of time 0 in the time 0 k-block, that is the time 0 k-block in $T^{-i}x$ begins at time 0. Note that the partitions P_k of x according to the time 0 k-block and k-block level of x increase and generate $\mathcal{B}$ up to null sets.

We conclude this section by establishing weak-mixing, using some variations on a standard argument (e.g. [2,3]).

THEOREM 2.10. *T is weak-mixing for all irrational α.*

PROOF. Since β is irrational $n_k > 2$ infinitely often. We continue to use the notation k^* for the least integer greater than k such that $n_{k^*-1} > 2$. By Lemma 2.6 either $\varepsilon_k = (+1, +1)$ infinitely often or $\varepsilon_k = (+1, -1)$ eventually. In the second case n_k is even eventually so we can find infinitely many k such that

$n_k > 3$. In the first case we can find infinitely many k such that $\varepsilon_k = (+1, +1)$ and either $n_{k-1} > 2$ or $n_{k-2} > 2$. To see this choose $\varepsilon_{k'} = (+1, +1)$ and, if $n_{k'-1} = 2$, suppose $n_{k'-l-1} > 2$, $n_{k'-i} = 2$ for $1 \leqq i \leqq l$. Then by Lemma 2.6 $\varepsilon_{k'-i}$ alternates between $(+1, +1)$ and $(-1, +1)$ as i runs from 1 to l so we may take $k = k' - l$ or $k' - l + 1$.

Now if $f \circ T = \lambda f$ then $|f| = 1$ since T is ergodic, so given $\varepsilon > 0$, by the above remarks we can find k as large as we wish and $\bar{f}$ such that $\int |\bar{f} - f| \, d\mu < \varepsilon$, $|\bar{f}| = 1$, $\bar{f}(x)$ depends only on the time 0 k-block and k-block level of x and one of the following holds:

(i) $\varepsilon_k = (+1, +1)$, $n_k > 2$,

(ii) $\varepsilon_k = (+1, +1)$, $n_k = 2$ and either $n_{k-1} > 2$ or $n_{k-2} > 2$,

(iii) $\varepsilon_k = (+1, -1)$, $n_k > 3$.

In each case we will argue that λ is close to 1.

If (i) holds choose $k+1$-blocks C and D such that the frequency of occurrence of CD among $k+1$-block pairs in $x \in X$ is greater than 1/16. (This is possible since there are only four types of $k+1$-blocks.) Note that C and D both begin with $\pm 0_k$ since $n_k > 2$ and that the signs of k-blocks in x alternate since $\varepsilon_k = (+1, +1)$, so we see that C and D contain a common substring γ consisting of at least $n_k - 2$ k-blocks $\pm 0_k$. Suppose that γ occurs at the ith and jth co-ordinates of C and D respectively and set $l^+ = j - i + |C|$.

Since $\bar{f}$ depends only on the time 0 k-block and k-block level, $\bar{f} \circ T^{l^+}(T^i x) = \bar{f}(T^i x)$ for indices i within the γ in the C of any CD in x. Thus for such indices i

$$|\bar{f} \circ T^{l^+}(T^i x) - \lambda^{l^+} \bar{f}(T^i x)| = |1 - \lambda^{l^+}|,$$

since $|f| = 1$. These indices occur in x with frequency greater than

$$\frac{|\gamma|}{|CD|} \frac{1}{16} > \frac{(n_k - 2)|-0_k|}{2n_k |0_k|} \frac{1}{16}$$

$$> \left(\frac{1}{2} - \frac{1}{3}\right) \frac{1}{16} \frac{|-0_k|}{|0_k|} > \frac{1}{200}$$

for large k ($|0_k| \sim |-0_k|$). Moreover since $f \circ T^{l^+} = \lambda^{l^+} f$

$$2\varepsilon > \|\bar{f} \circ T^{l^+} - f \circ T^{l^+}\|_1 + \|\lambda^{l^+} f - \lambda^{l^+} \bar{f}\|_1$$

$$> \int |\bar{f} \circ T^{l^+} - \lambda^{l^+} \bar{f}| \, d\mu$$

$$> |1 - \lambda^+| \frac{1}{200},$$

by calculating the integral as a time average over an orbit, so

$$|1-\lambda^{l^+}| < 400\varepsilon .$$

Now by Lemma 2.9, $C\ D^-$ also occurs with frequency more than 1/16. If γ^- occurs in C^- and D at the i'th and j'th co-ordinates respectively and $l^- = j' - i' + |C\ |$ then we can argue just as above that

$$|1-\lambda^{l}\ | < 400\varepsilon ,$$

so

(iv) $|1-\lambda^{l^+\ l}| < 800\varepsilon .$

Since C ends with a single $\pm 1_k$ and since the signs of k-blocks in x alternate, l^+ is the length of a string ξ of r k-blocks, r even, consisting of $r-1$ $\pm 0_k$'s and one $\pm 1_k$. Since l^- is the length of ξ^- we see that

$$|l^+ - l\ \ | = |0_k - 1_k| - |-0_k\ 1_k| = 1$$

(Lemma 2.6) so (iv) becomes $|1-\lambda| < 800\varepsilon$.

In case (ii) let $t \geqq 1$ be the least positive integer such that $n_{k+t} > 2$, so we have

$$0_{k+t} = 0_k - 1_k\, 1_k \cdots \pm 1_k \quad (t\ \pm 1_k\text{'s}),$$

$$1_{k+t} = 1_k .$$

Since $n_{k+t} > 2$ both 0_{k+t+1} and 1_{k+t+1} begin with 0_{k+t} so $x \in X$ is a concatenation of blocks A, A^-, B, B^- where $A = 0_{k+t}$ and $B = 0_{k+t} \pm 1_k$. Choose C and D blocks of this type such that CD occurs in x with frequency greater than 1/16. Since C and D both contain at least one $\pm 1_k$, C and D have a common substring γ consisting of at least t $\pm 1_k$'s. Now if l^+ is the distance between the initial indices of γ in C and in D we can argue as in (i) that

$$\begin{aligned} 2\varepsilon &> |1-\lambda^{l^+}| \frac{|\gamma|}{|CD|} \frac{1}{16} \\ &\gtrsim |1-\lambda^{l^+}| \frac{t|1_k|}{2(|0_k| + (t+1)|1_k|)} \frac{1}{16} \\ &\gtrsim |1-\lambda^{l^+}| \frac{t|1_k|}{2(t+4)|1_k|} > |1-\lambda^{l^+}| \frac{1}{11}\frac{1}{16}, \end{aligned}$$

where we have used $n_{k-1} > 2$ or $n_{k-2} > 2$ to see that $|0_k| \leqslant 3|1_k|$. Thus, for large k, $|1-\lambda^{l^+}| < 400\varepsilon$. Considering C^- and D^- and the corresponding l^- as before we see that $|1-\lambda^{l^-}| < 400\varepsilon$ and since $|l^+ - l^-| = 1$ as before we find that $|1-\lambda| < 800\varepsilon$.

In case (iii) choose a $k+1$-block pair CD with frequency at least 1/16. Both C and D contain at least $n_k - 2 \pm 0_k$'s so they have a common γ consisting of at least $n_k - 3 \pm 0_k$'s. (If n_k were 3 one might have $CD = 0_k - 1_k - 0_k 1_k$, for example.) Defining l^+ as before

$$2\varepsilon > |1-\lambda^{l^+}| \frac{|\gamma|}{|CD|}\frac{1}{16}$$

$$\geq |1-\lambda^{l^+}| \frac{(n_k-3)|0_k|}{2n_k|0_k|}\frac{1}{16} > |1-\lambda^{l^+}|\frac{1}{10}\frac{1}{16},$$

so for large k, $|1-\lambda^{l^+}| < 400\varepsilon$. Now since the final $\pm 1_k$ in C is followed by a $\pm 0_k$ in D of the same sign, l^+ is the length of a string of r k-blocks, r odd, including just one $\pm 1_k$. Thus if l^- is defined as before

$$|l^+ - l^-| = |1_k| - |-1_k| = 1,$$

so $|1-\lambda| < 800\varepsilon$. Since ε is arbitrary $\lambda = 1$. □

§3. Simplicity of T

Our goal in this section is to prove

THEOREM 3.1. *If α is poorly approximable then T is simple.*

COROLLARY 3.2. *If α is poorly approximable then T is prime and has trivial centralizer.*

To prove 3.1 suppose λ is $T \times T$ invariant and ergodic with marginals μ and choose a sequence $z = (x, y) \in X \times X$ which is generic for λ. Identifying $X \times X$ with $(\{1,-1\}\times\{1,-1\})^{\mathbb{Z}}$ we will write z as a sequence of pairs

$$z = \cdots \begin{matrix} x(-1)x(0)x(1) \\ y(-1)y(0)y(1) \end{matrix} \cdots$$

Say a finite string $\gamma \in (\{1,-1\}\times\{1,-1\})^l$ is ε-r-generic ($l > r$) if the frequency of occurrence in it of any string ξ in $(\{1,-1\}\times\{1,-1\})^r$ is within ε of $\lambda(\xi)$. Given ε and r if k is sufficiently large then any substring γ of z of length greater than $c\,|1_k|$ beginning no more than $10C|0_{k^*}|$ away from time 0 in z is ε-r-generic (see Lemma 2.8 for the definitions of c and C). Fix such a k such that also $n_{k-1} > 2$ and call such a γ substantial. We will say that a configuration of k-blocks in z is forcing if it forces the existence of substantial strings γ_1 and γ_2 in z such that

$$\gamma_1 = \begin{matrix} \xi_1 \\ \eta_1 \end{matrix} \cdots \begin{matrix} \xi_l \\ \eta_l \end{matrix} \qquad \gamma_2 = \begin{matrix} \xi_1 \\ \eta_0 \end{matrix} \cdots \begin{matrix} \xi_l \\ \eta_{l-1} \end{matrix}.$$

Note that this implies 2ε-$(r-1)$-$I\times T$-invariance of λ in the sense that for any cylinder E of length $r-1$ in $X\times X$

$$|\lambda(E)-\lambda((I\times T)E)|<2\varepsilon.$$

We introduce the following convenient notation for forcing configurations. If $A_1\cdots A_r$ and $B_1\cdots B_r$ are strings of k-blocks in x and y we write

$$\begin{array}{c|c} A_1\cdots A_{i-1} & A_i\cdots A_r \\ B_1\cdots B_{i-1} & B_i\cdots B_r \end{array}$$

to mean that the initial $\pm 1_k$'s in A_i and B_i (which may be all of A_i and B_i) overlap in at least $c\,|1_k|$ indices and the same is (necessarily) true for the final $\pm 1_k$'s in A_{i-1} and B_{i-1} (but not necessarily for other A_j and B_j). We say the initial (final) $\pm 1_k$'s in A_i and B_i (A_{i-1} and B_{i-1}) are left (right) aligned. To see that a configuration is forcing we will usually have to look at (possibly) false $\pm 1_k$'s as described in Lemma 2.7 and we indicate these by $\pm\tilde{1}_k$. The substantial strings γ_1 and γ_2 of a forcing configuration will always be the overlap of the aligned initial or final $\pm 1_k$'s in some pair of k-blocks in x and y and we indicate which k-blocks by underlining. Here are some examples of the use of this notation.

If $\varepsilon_k=(+1,+1)$ the configuration

F(i)

$$\begin{array}{c|} -0_k \\ 0_k \end{array}$$

is forcing because it implies

$$\begin{array}{c|c|c} +\tilde{1}_k & -0_k & \tilde{1}_k \\ \underline{-\tilde{1}_k} & 0_k & \underline{-\tilde{1}_k} \end{array}$$

and the separation between the initial indices of the $\pm 1_k$'s upstairs is $|1_k-0_k|$ while downstairs it is $|-1_k0_k|$ and $|1_k-0_k|-|-1_k0_k|=1$. (Here we have ignored the difference in length between 0_k and -0_k in order to left align them. This is unimportant because of Lemma 2.8 — the initial $\pm 1_k$'s will overlap in almost $c\,|1_k|$ — and we will do this consistently.) Other examples of forcing configurations are (with the explanation in each case following the colon):

$\underline{\varepsilon_k=(+1,+1):}$

F(ii)

$$\begin{array}{c|} -0_k\,1_k \\ -1_k\,0_k \end{array} \quad : \quad \begin{array}{|c|c} -0_k\,1_k & -\tilde{1}_k \\ \underline{-1_k}\,0_k & \underline{-\tilde{1}_k} \end{array}$$

$\underline{\varepsilon_k = (+1, -1):}$

F(iii)

$$\begin{array}{r|} -0_k \\ 0_k \end{array} \quad : \quad \underline{\begin{array}{r|r|} \bar{1}_k & \bar{1}_k \\ -\bar{1}_k & -\bar{1}_k \end{array}}$$

F(iv)

$$\begin{array}{r|} -1_k \\ 1_k \end{array} \quad : \quad \underline{\begin{array}{r|r} -1_k & -\bar{1}_k \\ 1_k & \bar{1}_k \end{array}}$$

F(v)

$$\begin{array}{|r} -1_k \\ 0_k \end{array} \quad : \quad \underline{\begin{array}{r|r} -\bar{1}_k & -1_k \\ \bar{1}_k & \bar{1}_k \end{array}}$$

F(vi)

$$\begin{array}{r|} 1_k \\ 0_k \end{array} \quad : \quad \underline{\begin{array}{r|r} 1_k & \bar{1}_k \\ -\bar{1}_k & \bar{1}_k \end{array}}$$

$\underline{\varepsilon_k = (-1, +1):}$

F(vii)

$$\begin{array}{r|} -0_k \\ 0_k \end{array} \quad : \quad \underline{\begin{array}{r|} \bar{1}_k \\ -\bar{1}_k \end{array}} \; \underline{\begin{array}{r|} -0_k \\ 0_k \end{array}}$$

F(viii)

$$\begin{array}{rr|} 0_k & 1_k \\ 1_k & -0_k \end{array} \quad : \quad \begin{array}{|rr|} 0_k & 1_k \\ \underline{1_k} & -\underline{0_k} \end{array}$$

In addition the negative of any forcing configuration is forcing.

We will say that a finite string occurring at i in x (or y) attracts itself (with shift $j-i$) if the same string occurs at j in y (or x) and $|j-i| < c|1_k|$. The key to Theorem 3.1 is the following lemma.

LEMMA 3.3. *Either there is a forcing configuration of k-blocks or $x[-C|0_{k^*}|, C|0_{k^*}|]$ attracts itself.*

PROOF. We assume that there is no forcing and show that the attraction holds. Let us say strings in x and y overlap substantially if their overlap is greater than $c|1_k|$ in length. To prove the lemma we distinguish six cases according to the values of ε_k and k^*. Note that all strings involved in our argument are well within the good range $[-10C|0_{k^*}|, 10C|0_{k^*}|]$.

Case I: $\varepsilon_k = (+1, +1)$, $k^* = k+1$

Let A be the block 0_{k+1} if $n_k = n_{k^*-1}$ is odd or 1_{k+1} if $n_k - 1$ is odd. Thus A ends with 1_k and any occurrence of A in x is preceded by (a genuine) -1_k.

Suppose now that A occurs in $x[-2C|0_{k^*}|, 2C|0_{k^*}|]$ so we see

(i) $-1_k\, 0_k - 0_k \cdots - 0_k\, 1_k \quad (r \ \pm 0_k\text{'s})$.

Because there are at most $r+1$ $\pm 0_k$'s between $\pm 1_k$'s in y and $|1_k| > (\frac{1}{2}+2c)|0_k|$ ($n_{k-1} > 2$ and Lemma 2.8) one sees that there must be a $\pm 1_k$ in y overlapping (i) substantially. We suppose first that it is a -1_k and show that (i) attracts itself.

We claim that the -1_k in y is aligned with the -1_k of (i). If this were not the case we would be able either to align it with the 1_k of (i) or right align it with the leftmost 0_k of (i) or left align it with some other $\pm 0_k$ of (i). (Here we are again using $|1_k| > (\frac{1}{2}+2c)|0_k|$.) In the first case we have

$$\begin{array}{c|c} -0_k & 1_k \\ 0_k & -1_k \end{array} \Rightarrow \mathrm{F(i)},$$

in the second case

$$\begin{array}{c|c} 0_k & -0_k \\ -1_k & 0_k \end{array} \Rightarrow \mathrm{F(i)},$$

and in the third case either

$$\begin{array}{c|c} -0_k & 0_k \\ 0_k & -1_k \end{array} \Rightarrow \mathrm{F(i)}$$

or

$$\begin{array}{|l|l} -0_k \quad\ \cdots - 0_k\, 1_k & -\bar{1}_k \\ \underline{-1_k}\, 0_k \cdots - 0_k\, 0_k & \underline{-\bar{1}_k} \end{array}.$$

(Notice that in the last forcing configuration above we do see a 0_k below the 1_k because the number of $\pm 0_k$'s between $\pm 1_k$'s in y is at least $r-1$.) This establishes our claim.

Now look at the next $\pm 1_k$ to the right of the -1_k in y which we have just aligned. If it is not a $+1_k$ we see either

$$\begin{array}{c|c|c} -1_k & 0_k \ldots & -0_k \quad 1_k \\ -1_k & 0_k & -1_k \quad 0_k \end{array} \Rightarrow \mathrm{F(ii)}$$

or

$$\begin{array}{c|c|c} -1_k & \ldots & 1_k \quad -0_k \\ -1_k & & 0_k \quad -1_k \end{array} \Rightarrow \mathrm{F(ii)}.$$

Thus it is a 1_k, so we have shown that (i), and hence A, attracts itself. To do this we started by assuming a -1_k in y overlapped (i) substantially but if it were a $+1_k$ we could argue similarly by working towards the -1_k in (i). Also, the same argument works for A^- and, of course, A's and A^-'s in $y\,[-2C|0_{k^*}|, 2C|0_{k^*}|]$ also attract themselves.

Thus any A or A^- in x or y which overlaps the index set $[-C|0_{k^*}|, C|0_{k^*}|]$ attracts itself. By Lemma 2.8 there must be such an A or A^- in x and in y. Moreover $(1-c)|1_k|$, the "range of attraction", is less than $\frac{1}{2}|1_{k^*}|$ $(n_{k-1} > 2!)$, that is, one half the minimal length of a k^*-block. From this it follows that all the A's and A^-'s overlapping $[-C|0_{k^*}|, C|0_{k^*}|]$ attract themselves with the same shift and hence $x[-C|0_{k^*}|, C|0_{k^*}|]$ attracts itself.

Case II: $\varepsilon_k = (+1, +1)$, $k^* = k + l + 1$, $l \geqq 1$

In this case

$$0_{k+l} = 0_k(\pm 1_k)^l,$$

$$1_{k+l} = 1_k.$$

Since $n_{k^*-1} > 2$ all k^*-blocks begin with a $\pm 0_{k+l}$ so one never sees two $\pm 1_{k+l}$'s in a row in a T-name. Thus x and y are concatenations of the blocks $D_1^{\pm}$ and $D_2^{\pm}$ where $D_1 = 0_k(\pm 1_k)^l$ and $D_2 = 0_k(\pm 1_k)^{l+1}$. Let A be D_1 if l is even or D_2 if $l+1$ is even so A ends with a 1_k. If A occurs in $x[-2C|0_{k^*}|, 2C|0_{k^*}|]$ it is followed by -0_k so we see

(ii) $0_k - 1_k \cdots 1_k - 0_k$ $(r\ \pm 1_k\text{'s}, r \geqq 2)$.

Because there are at most $r+1$ $\pm 1_k$'s between $\pm 0_k$'s in y we see that there must be a $\pm 0_k$ in y left aligned with one of the k-blocks of (ii). Let us assume that it is a 0_k in y and show that A attracts itself (the argument is similar for -0_k). We claim that it must be aligned with the 0_k of A. To see this note that if it is not aligned as claimed, since it cannot be aligned with the -0_k because of F(i), then we see either

$$\left\lfloor\begin{matrix} 1_k \\ 0_k \end{matrix}\right. \quad \begin{matrix} -1_k \\ -1_k \end{matrix} \cdots \begin{matrix} 1_k \\ 1_k \end{matrix} \quad \begin{matrix} -0_k \\ -1_k \end{matrix} \left\lfloor\begin{matrix} 1_k \\ \bar{1}_k \end{matrix}\right.$$

or

$$\left\lfloor\begin{matrix} -1_k \\ 0_k \end{matrix}\right. \cdots \left\lfloor\begin{matrix} -0_k \\ \bar{1}_k \end{matrix}\right. .$$

Note that in both cases the rightmost 1_k in y may be false but it is preceded by a genuine 1_k because the number of $\pm 1_k$'s between $\pm 0_k$'s in y is at most $r+1$. This establishes the claim.

Now the next $\pm 0_k$ in y to the right of the 0_k we have just aligned cannot be a 0_k because then we would see either

$$\begin{array}{c|c|cc} 0_k & & 1_k & -0_k \\ 0_k & \cdots & 0_k & -1_k \end{array} \Rightarrow \text{F(ii)}$$

or

$$\begin{array}{c|c|cc} 0_k & & -0_k & 1_k \\ 0_k & \cdots & -1_k & 0_k \end{array} \Rightarrow \text{F(ii)}.$$

Thus we have shown that A attracts itself, and the same is true of A . Now by Lemma 2.8 there must be an occurrence of $\pm 0_{k^*}$ in $x[-C|0_{k^*}|, C|0_{k^*}|]$. Since $0_{k^*} = 0_{k+l+1}$ begins with 0_{k+l} and ends with $\pm 0_{k+l}$ $\pm 1_{k+l}$ we see that either A or A^- occurs in $x[-C|0_{k^*}|, C|0_{k^*}|]$. Since $(1-c)|1_k| < \frac{1}{2}\min(|D_1|,|D_2|)$ we can finish the argument as in case I.

Case III: $\varepsilon_k = (+1, -1)$, $k^* = k+1$

Let A be 0_{k+1} if n_k is even or 1_{k+1} if $n_k - 1$ is even, so that A ends with -1_k. Suppose that A occurs in x within the good range, so we see

(iii) $1_k\, 0_k - 0_k \cdots 0_k - 1_k$ (r $\pm 0_k$'s).

As in case I there is a $\pm 1_k$ in y overlapping (iii) substantially, which we assume is a 1_k, leaving the other case to the reader. The 1_k in y must be aligned with the 1_k or (iii), for otherwise we could align it with the -1_k, which is F(iv), or (as in I) right align it with the leftmost 0_k of (i) or left align it with some other $\pm 0_k$ of (i). This would give either F(vi), F(v) or

$$\begin{array}{|ccccc|c} -0_k & 0_k & -0_k & \cdots & -1_k & -\tilde{1}_k \\ \underline{-0_k} & 1_k & 0_k & & 0_k & \underline{-\tilde{1}_k} \end{array}.$$

Similar arguments show that the next $\pm 1_k$ in y to the right of the one we have just aligned is a -1_k so (iii) attracts itself. The argument is concluded as in I.

Case IV: $\varepsilon_k = (+1, -1)$, $k^* = k + l + 1$, $l \geqq 1$

In this case $0_{k+l} = 0_k(-1_k)^l$ and if 0_{k+l} occurs in x we see

(iv) $0_k - 1_k \cdots - 1_k - 0_k$ (l -1_k's)

As in II there is a $\pm 0_k$, say a 0_k, in y left-aligned with some k-block of (iv). In fact it must be aligned with some k-block of (iv). In fact it must be aligned with the 0_k of (iv), for otherwise we get F(v) or F(iii). (If initially we found a -0_k in y we should work with right, rather than left, alignment.) If the next $\pm 0_k$ to the right in y, which is in fact a -0_k, is separated by $l+1$ -1_k's, rather than l -1_k's, one sees

$$\begin{array}{c|cccc|} \tilde{1}_k & 0_k & \cdots & -0_k & 1_k \\ \underline{\tilde{1}_k} & 0_k & & -1_k & \underline{-0_k} \end{array}.$$

Thus (iv) attracts itself and the argument is concluded as in II.

Case V: $\varepsilon_k = (-1, +1)$, $k^* = k+1$

If a 0_{k+1} occurs in the good range of x we see

(v) $-1_k\, 0_k \cdots 0_k\, 1_k$ ($n_k - 1$ 0_k's, $n_k > 2$).

We can find a $\pm 1_k$ in y overlapping (v) substantially and we assume it is a 1_k. It must be aligned with the 1_k of (v), for otherwise it is aligned with the -1_k

$$\begin{array}{r|r} -1_k & 0_k \\ 1_k & -0_k \end{array} \Rightarrow \mathrm{F(vii)}$$

or left aligned with the rightmost 0_k of (v)

$$\begin{array}{|ll} 0_k & 1_k \\ 1_k & -0_k \end{array} \Rightarrow \mathrm{F(viii)}$$

or right aligned with some other 0_k

$$\begin{array}{r|r} 0_k & 0_k \\ 1_k & -0_k \end{array} \Rightarrow \mathrm{F(vii)}.$$

If the next -1_k to the left in y is separated by $n_k - 2$ 0_k's, instead of $n_k - 1$, we see

$$\begin{array}{rr|c|l} -1_k & 0_k & \cdots & 1_k \\ -0_k & -1_k & & 1_k \end{array} \Rightarrow \mathrm{F(viii)}.$$

Thus (v) attracts itself and we are done as in I.

Case VI: $\varepsilon_k = (-1, +1)$, $k^* = k + l + 1$, $l \geqq 1$

Let A be $0_k(\pm 1_k)^l$ if l is odd or $0_k(\pm 1_k)^{l+1}$ if $l+1$ is odd. If A occurs in x we see

(vi) $0_k\, 1_k - 1_k \cdots 1_k - 0_k$ (r $\pm 1_k$'s)

and as in II we find a $\pm 0_k$, say a 0_k, in y left aligned with some k-block of (vi). It must be aligned with the 0_k of (vi) for otherwise it is aligned with the -0_k (F(vii)) or we see

$$\begin{array}{|rrcr|r} -1_k & 1_k & \cdots & -0_k & -1_k \\ \underline{0_k} & 1_k & & -1_k & \underline{\bar{1}_k} \end{array}$$

or

$$\begin{array}{|lrr|} 1_k & -1_k & -0_k \\ \underline{0_k} & 1_k & \underline{\bar{1}_k} \end{array}.$$

If the next $\pm 0_k$ to the right in y is not a -0_k we see

$$\begin{matrix} 1_k & -0_k \\ 0_k & 1_k \end{matrix}\Bigg| \quad \text{or} \quad \begin{matrix} -0_k & -1_k \\ -1_k & 0_k \end{matrix}\Bigg|,$$

both F(viii). Thus A attracts itself which is enough as in II.

This concludes the proof of Lemma 3.3. □

We can now finish proving Theorem 3.1. If there is a forcing configuration for infinitely many k such that $n_{k-1} > 2$ then by the remarks at the beginning of this section λ is ε-r-$I \times T$ invariant for every ε and r ($\varepsilon \to 0$ and $r \to \infty$ as $k \to \infty$). Thus λ is $I \times T$ invariant and this implies that λ is $\mu \times \mu$ [7, prop. 2].

Otherwise by Lemma 3.3, for sufficiently large k such that $n_{k-1} > 2$, $x[-C|0_{k^*}|, C|0_{k^*}|]$ attracts itself with shift i_{k^*}, $|i_{k^*}| < (1-c)|1_k|$. C can, of course, be as large as we wish, in particular large enough so that $x[-C|0_{k^*}|, C|0_{k^*}|]$ contains k^{**}-blocks, which, like $x[-C|0_{k^*}|, C|0_{k^*}|]$, attract themselves with shift i_{k^*}. However, as pieces of $x[-C|0_{k^{**}}|, C|0_{k^{**}}|]$ they also attract themselves with shift $i_{k^{**}}$. Now

$$|i_{k^*}| + |i_{k^{**}}| < (1-c)(|1_k| + |1_{k^*}|) < (1-c)|1_{k^{**}}|,$$

so by the uniqueness of k^{**}-block appearances (Lemma 2.7) we see that $i_{k^*} = i_{k^{**}}$. Similarly $i_{k^{**}} = i_{k^{***}}$ etc., so y is a shift of x and λ is an off-diagonal measure. This concludes the proof of Theorem 3.1. □

REMARK. If there is a sequence of fractions p_n/q_n with p_n and q_n co-prime and q_n even such that $|\alpha - p_n/q_n| = o(q_n^{-2})$ then it is easy to see that $T^{3q_n/2}$ converges to the identity in the weak topology. It follows that the closure of $\{T^k : k \in \mathbf{Z}\}$ is a perfect subset of a complete metric space and so must be uncountable. (This observation can be found in [9, chapter 5, paragraph 2.1.4].) This means the centralizer of T is uncountable so T is not simple.

§4. Higher cartesian products

THEOREM 4.1. *If α is poorly approximable, $t > 0$ and ν is a $\bigotimes_{i=1}^{t} T^i$-invariant and ergodic measure on X^t with marginals μ then ν is a product of off-diagonal measures.*

PROOF. We begin by observing that it suffices to handle the case $t = 1$ for if $S = \bigotimes_{i=1}^{t} T$ and ν is S^t invariant and ergodic with marginals μ then we may form $\nu' = t^{-1}(\nu + S\nu + \cdots + S^{t-1}\nu)$ which is S invariant with marginals μ. Moreover ν' is ergodic for S: if $SE = E$ then $S^tE = E$ so $\nu(E) = \nu(SE) = \cdots =$

$\nu(S^{i-1}E)$ is 0 or 1. Thus, if we already have the case $t=1$, ν' is a product of off-diagonals, which allows us to conclude that ν' is also S^t ergodic, since T is weak-mixing. Since each $S^i\nu$ is S^t invariant and ν' is an average of these we can conclude that each $S^i\nu$, in particular ν, coincides with ν'. Thus we shall assume hereafter that $t=1$.

We write $T^{(l)}$ for $\bigotimes_{i=1}^{l} T$. Let $\delta_1 = -0, \delta_2 = \cdots = \delta_{l-1} = +0$, $s \in \mathbf{Z}^+$ and let E_k^i be the set of $x \in X$ such that the 0th co-ordinate of x sits inside the first k-block of the $k+s$-block $C_{k+s}(\delta_i)$, where it is understood that these are the time 0 k- and $k+s$-blocks in x. For $x=(x_1,\cdots,x_{l-1}) \in X^{l-1}$ set $\chi_k(x)=\prod_{i=1}^{l-1} 1_{E_{k}^{i}}(x_i)$. Denote product measure on X^{l-1} by μ^{l-1}.

Lemma 4.2. $\mu^{l-1}\{x \in X^{l-1} : \chi_k(x)=1 \text{ i.o.}\}=1$.

Proof. For any $x \in X$, $C_{k+s}(\varepsilon_i)$ occupies a proportion of x which is bounded below as $k \to \infty$. Also by the boundedness of $\{n_k\}$ the first k-block in $C_{k+s}(\varepsilon_i)$ occupies a bounded below proportion of $C_{k+s}(\varepsilon_i)$. Thus $\mu(E_k^i)$ is bounded below and hence so is $\mu^{l-1}\{x : \chi_k(x)=1\}$. It follows that $\mu^{l-1}\{x : \chi_k(x)=1 \text{ i.o.}\}>0$.

Now for μ-a.a. x it is true that for sufficiently large k the 0th co-ordinate in x is not the first or last co-ordinate in the time 0 k-block in x, because k-block lengths grow exponentially. Thus we see that for μ^{l-1}-a.a. $x \in X^{l-1}$, $\chi_k(x)=\chi_k(T^{(l-1)}x)$ for sufficiently large k and so the set $\{x \in X^{l-1} : \chi_k(x)=1 \text{ i.o.}\}$ is $T^{(l-1)}$ invariant (μ^{l-1}-a.e.). Since μ^{l-1} is ergodic for $T^{(l-1)}$ it follows that this set has measure 0 or 1, but we have already seen it cannot be 0. □

Continuing with the proof of 4.1 it is sufficient, now that an induction has been started by Theorem 3.1, to assume that every $l-1$ dimensional marginal is μ^{l-1} and show that ν is μ^l. (See [7] for details of this reasoning.) Thus when we choose $x=(x_1,\cdots,x_l) \in X^l$ generic for ν we may assume by Lemma 4.2 that $\chi_k(x_1,\cdots,x_{l-1})=1$ i.o.

We adopt a notation analogous to that of section 3: if $A_1,\cdots,A_l$ are k-blocks in $x_1,\cdots,x_l$, $l>2$, we write

$$\begin{vmatrix} A_1 \\ \vdots \\ A_l \end{vmatrix.} \quad \text{or} \quad \begin{vmatrix.} A_1 \\ \vdots \\ A_l \end{vmatrix}$$

if the initial or final 1_k's have a common overlap greater than $\frac{1}{2}c\,|1_k|$. For $l=2$ alignment will continue to mean what it did in Section 3 — overlap greater than $c\,|1_k|$. We will also say that $A_1,\cdots,A_l$ are strongly aligned if the initial indices of all the A_i are no more than $\frac{1}{2}c\,|1_k|$ distant from the initial index of A_1. By choosing s in the definition of χ_k sufficiently large we see that there are infinitely many k such that the time 0 $k+s$-block in x_1 is -0_{k+s}, the time 0 $k+s$-blocks in $x_2,\cdots,x_{l-1}$ are all 0_{k+s} and these $k+s$-blocks are strongly aligned, since their initial k-blocks overlap and these initial k-blocks are a small fraction of the length of $k+s$-blocks. Moreover for $0 \leqq i < s$ the time 0 $k+s-i$ blocks in x_1 and $x_2,\cdots,x_l$ will also be -0_{k+s-i} and 0_{k+s-i} respectively and they will be strongly aligned if i is small compared to s. Choosing $n_{k+s-i} > 2$ and reindexing we have infinitely many k such that: (i) $n_k > 2$; (ii) the time 0 k-blocks in x_1 and $x_2,\cdots,x_l$ are -0_k and 0_k respectively; (iii) these k-blocks are strongly aligned; (iv) each is the first k-block in its time 0 k^*-block.

Now for any $\varepsilon > 0$ and $r \in \mathbf{Z}^+$ there is a k satisfying (i), (ii), (iii) and (iv) such that any finite string in x of length greater than $\frac{1}{2}c\,|1_k|$ occurring no more than $10|0_k|$ away from time 0 is ε-r-generic for ν. We are now going to argue that

(v) ν is 2ε-$(r-1)$-invariant under $\bigotimes_{i=1}^{l} T^{u_i}$ where $u_1,\cdots,u_l$ depend on k, but are bounded, with at least one $u_i = 0$ and one $u_i \neq 0$.

Once we have (v) we can finish the argument by noticing that some configuration of u_i's occurs for infinitely many k so we actually get that ν is $\bigotimes_{i=1}^{l} T^{u_i}$ invariant for some $u_1,\cdots,u_l$ as in (v). Since $\bigotimes_{u_i\neq 0} T^{u_i}$ is ergodic, ν is the product of two lower dimensional marginals ([7, proposition 2]) and these are product measures by assumption.

To prove (v) we again consider six cases as in Theorem 3.1, but they will go more quickly this time.

Case I: $k^* = k+1$, $\varepsilon_k = (+1,+1)$

We can find either a $\pm 0_k$ or a $\pm 1_k$ in x_l which is left aligned with the time 0 -0_k in x_1, so this k-block in x_l and the initial $\pm 1_k$'s in the time 0 k-blocks in $x_1,\cdots,x_{l-1}$ have a common overlap of at least $\frac{1}{2}c\,|1_k|$, that is, they are aligned. In the case of a 0_k in x_l we get the following configuration whose "forcing" character is indicated in a manner completely analogous to section 3:

$$\left|\begin{array}{c} \tilde{1}_k \\ -\tilde{1}_k \\ \\ \underline{-1_k} \end{array}\right. \quad \begin{array}{c} -0_k \\ 0_k \\ \vdots \\ 0_k \end{array} \quad \left|\begin{array}{c} \tilde{1}_k \\ -\tilde{1}_k \\ \\ \underline{-\tilde{1}_k} \end{array}\right. .$$

In this configuration the position of the leftmost $\tilde{1}_k$ of x_1 relative to the $-\tilde{1}_k$ of, say, x_2 is one space further to the left than that of the rightmost $\tilde{1}_k$ in x_1 relative to the $-\tilde{1}_k$ in x_2, while the $-\tilde{1}_k$'s in $x_3,\cdots,x_l$ experience no such shift. Thus we obtain (v) with $u_1 = 1$, $u_2 = u_3 = \cdots = u_l = 0$. Similarly in case of a -0_k in x_l we get forcing with $u_1 = u_l = 1$, other $u_i = 0$. If it is a $\pm 1_k$ in $x_1,\cdots,x_l$, since the time

0 k-blocks in $x_1, \cdots, x_l$ are the initial k-blocks in their $k+1$-blocks, they are preceded by $\pm 1_k$'s, so we see

$$\begin{array}{|c c|c} -0_k & 1_k & -0_k \\ 0_k & -1_k & 0_k \\ & \vdots & \\ \underline{\pm 1_k} & \pm 0_k & \underline{\pm 1_k} \end{array}$$

and we have $u_1 = -1$, $u_2 = \cdots = u_{l-1} = 0$, $u_l = -1$ or 0.

Case II: $\varepsilon_k = (+1, +1)$, $k^* = k + m + 1$, $m \geqq 1$

Consider the time 0 $k+m$ block in x_1 which is $-0_{k+m} = -0_k(\pm 1_k)^m$ and is followed by a $\pm 0_k$ because it is the first $k+m$-block in a 0_{k^*}. The situation is similar in $x_2, \cdots, x_{l-1}$ so in $x_1, \cdots, x_{l-1}$ we see

(vi)

$$\begin{array}{c c c} -0_k & 1_k & -1_k \cdots \pm 0_k \\ 0_k & -1_k & 1_k \cdots \mp 0_k \\ & & \vdots \\ 0_k & -1_k & 1_k \cdots \mp 0_k \end{array}$$

with the $\pm 0_k$'s strongly aligned. If a $\pm 0_k$ in x_l is aligned with any $\pm 0_k$ of (vi) we get forcing immediately as in I. Otherwise we can find a $\pm 0_k$ in x_l right aligned with some $\pm 1_k$ of (vi), so we see

$$\begin{array}{|c c c c|} -0_k & 1_k & \cdots & -1_k \\ 0_k & -1_k & \cdots & 1_k \\ & \vdots & & \\ \underline{\pm 1_k} & \mp 1_k & \cdots & \underline{\pm 0_k} \end{array} \quad \text{or} \quad \begin{array}{|c c c c|c} -0_k & 1_k & \cdots & 1_k & -\bar{1}_k \\ 0_k & -1_k & \cdots & -1_k & \bar{1}_k \\ & & \vdots & & \\ \underline{\mp 1_k} & \pm 1_k & \cdots & \pm 0_k & \underline{\mp \bar{1}_k} \end{array}$$

(note that the $\pm 0_k$'s in x_l are separated by at least m $\pm 1_k$'s.)

Cases III and IV: $\varepsilon_k = (+1, -1)$

Looking at the time 0 k-blocks in $x_1, \cdots, x_{l-1}$ and left aligning something in x_l with them we see

$$\begin{array}{c|c} -1_k & -0_k \\ 1_k & 0_k \\ & \vdots \\ 1_k & 0_k \\ \underline{\pm 1_k} & \underline{\pm 0_k} \end{array} \quad \text{or} \quad \begin{array}{c|c} -1_k & -0_k \\ 1_k & 0_k \\ & \vdots \\ 1_k & 0_k \\ \underline{\pm \bar{1}_k} & \underline{\pm 1_k} \end{array}$$

Case V: $\varepsilon_k = (-1, +1)$, $k^* = k + 1$

The time 0 k^*-block in x_1 begins with at least two -0_k's and there must be a $\pm 0_k$ in x_l aligned with one of them. Thus we see

$$
\begin{array}{|c}
-0_k \\
0_k \\
\vdots \\
0_l \\
\underline{\pm 0_k}
\end{array}
\qquad
\begin{array}{|c}
-\tilde{1}_k \\
\tilde{1}_k \\
\\
\tilde{1}_k \\
\underline{\pm \tilde{1}_k}
\end{array}
$$

Case VI: $\varepsilon_k = (-1, +1)$, $k^* = k + m + 1$, $m \geqq 1$

This is similar to II. We can rule out the possibility that a $\pm 0_k$ in x_l aligns with a $\pm 0_k$ in x_1, so as in II we must have

$$
\left|\begin{array}{ccccc}
-0_k & -1_k & +1_k & \cdots & -1_k \\
0_k & 1_k & -1_k & \cdots & 1_k \\
 & & \vdots & & \\
0_k & 1_k & -1_k & \cdots & 1_k \\
\underline{\mp 1_k} & \pm 1_k & \cdots & \mp 1_k & \underline{\pm 0_k}
\end{array}\right|
\quad \text{or} \quad
\begin{array}{|cccc}
-0_k & -1_k & \cdots & 1_k \\
0_k & 1_k & \cdots & -1_k \\
 & \vdots & & \\
0_k & 1_k & & -1_k \\
\underline{\pm 1_k} & \mp 1_k & & \pm 0_k
\end{array}
\quad
\begin{array}{|c}
-\tilde{1}_k \\
\tilde{1}_k \\
\\
\tilde{1}_k \\
\underline{\pm \tilde{1}_k}
\end{array}
$$

This concludes the proof of 4.1. □

§5. Connection with topological analogues and other remarks

Suppose Ψ is a homeomorphism of a compact metric space (Y, d). According to Furstenberg, Keynes and Shapiro [4] (Y, Ψ) is proximal orbit dense (POD) if it is minimal, no factor of Ψ is a rotation on a finite number of points, and whenever $x, y \in Y$ with $x \neq y$ there exists $\{n_i\} \subset \mathbf{Z}$ and $k \neq 0$ such that $d(\Psi^{n_i}\Psi^k x, \Psi^{n_i} y) \to 0$. They showed, among other things, that a class of flows, which includes (topological models of) the T's of this paper for all irrational α, consists of POD flows and that POD flows are topologically prime and have trivial topological centralizer. There is a close analogy between simplicity and the POD property. In particular if (Y, Ψ) is POD then the only $\Psi \times \Psi$-minimal subsets of $Y \times Y$ are the "off-diagonals" $\Delta_k = \{(y, \Psi^k y) : y \in Y\}$. This weaker property is easily connected with simplicity.

PROPOSITION 5.1. *If (Y, Ψ) is strictly ergodic with unique invariant probability μ and the measure-preserving system (Ψ, μ) is simple then the only minimal subsets of $Y \times Y$ are the off-diagonals.*

PROOF. By minimality $\operatorname{supp}\mu = Y$. If $M \subset Y \times Y$ is minimal, then letting ν be an invariant ergodic probability with $\operatorname{supp}\nu = M$, ν has marginals μ by unique ergodicity. Since (Ψ, μ) is simple ν is an off-diagonal or product measure. In the first case $M = \operatorname{supp}\nu = \Delta_k$ and the second case cannot occur because $\operatorname{supp}(\mu \times \mu) = Y \times Y$ which is not minimal. □

The next proposition shows that simplicity implies POD under certain conditions which are satisfied by T for poorly approximable α. Say that the flow (Y,Ψ) is weakly distal if for any $x,y\in Y$, $(2n+1)^{-1}\Sigma_{i=-n}^{n}d(\Psi^i x,\Psi^i y)\to 0$ implies $x=y$. Obviously (Y,Ψ) distal implies (Y,Ψ) weakly distal, but the converse is not true as we shall see. Note that $(2n+1)^{-1}\Sigma_{i=-n}^{n}d(\Psi^i x,\Psi^i y)\to 0$ is equivalent to the statement that for each $\varepsilon>0$, $\{i\in\mathbf{Z}: d(\Psi^i x,\Psi^i y)>\varepsilon\}$ has density 0, so the definition of weak distality is independent of the choice of metric. Also if Ψ is the shift on $I^{\mathbf{Z}}$, I finite, then $(2n+1)^{-1}\Sigma_{i=-n}^{n}d(\Psi^i x,\Psi^i y)\to 0$ (d any metric inducing the product topology) is equivalent to the statement

$$\bar{d}(x,y)=\lim_{n\to\infty}\frac{1}{2n+1}\#\{-n\leqq i\leqq n: x(i)\neq y(i)\}=0,$$

by the same density characterization. We will also use the notation

$$\bar{d}_{\pm}=\lim_{n\to\pm\infty}\frac{1}{n}\#\{0\leqq i\leqq n-1: x(i)\neq y(i)\}.$$

PROPOSITION 5.2. *If Y is infinite and (Y,Ψ) is strictly ergodic and weakly distal with invariant probability μ and (Ψ,μ) is simple then (Y,Ψ) is* POD.

PROOF. First, since Y is infinite and minimal, (Ψ,μ) is not a finite rotation so its simplicity implies that it is weak-mixing (see the remark at the end of section 1 in [12], where the non-atomic nature of μ is assumed though not explicitly mentioned). It follows that a continuous eigenfunction of Ψ must be constant μ-a.e., hence constant, since $\operatorname{supp}\mu = Y$. Thus Ψ has no finite rotation factors.

Now if x, $y\in Y$ with $x\neq y$ choose a measure ν on $Y\times Y$ such that (x,y) is quasi-generic for ν, that is, there is $\{n_i\}\subset\mathbf{Z}^+$ such that

$$\text{(i)}\qquad \text{w}^*\text{-lim}\frac{1}{2n_i+1}\sum_{n=-n_i}^{n_i}\delta_{(\Psi^n x,\Psi^n y)}=\nu,$$

δ_z the point mass at z. Since ν is $\Psi\times\Psi$-invariant (but not necessarily ergodic) and has marginals μ by unique ergodicity, we can write

$$\text{(ii)}\qquad \nu=c(\mu\times\mu)+\sum_{k=-\infty}^{\infty}c_k\mu_k,$$

where μ_k denotes off-diagonal measure on Δ_k. (Here one uses the fact that the extreme points of the convex set of $\Psi\times\Psi$-invariant measures with marginals μ are all ergodic.) If $c_k>0$ for some $k\neq 0$ let $h(t)$ be a continuous function on $\mathbf{R}^+$

such that $h(0)=1$ and $h(t)=0$ for $t \geqq \varepsilon$ and set $f(x,y)=h(d(\Psi^k x, y))$. Integrating (i) against f we find that

$$\frac{1}{2n_i+1}\sum_{n=-n_i}^{n_i} f(\Psi^n x, \Psi^n y) \to \int f d\nu \geqq c_k > 0,$$

so there must be values of n such that $f(\Psi^n x, \Psi^n y) > 0$, that is $d(\Psi^k \Psi^n x, \Psi^n y) < \varepsilon$. Since this is true for all ε we obtain $\{n_i\}$ such that $d(\Psi^{n_i}\Psi^k x, \Psi^{n_i} y) \to 0$. Similarly, if $c>0$, define f in the same way (for any choice of $k \neq 0$) and again $\int f d\nu \geqq c \int f d\mu \times \mu > 0$, since $f>0$ on a neighborhood of Δ_k and $\operatorname{supp} \mu \times \mu = Y \times Y$.

Thus we may assume c_0 is the only non-zero coefficient in (ii), so $\nu = \mu_0$, for all choices of $\{n_i\}$ giving w*-convergence. Thus we actually have

$$\text{w*-lim} \frac{1}{2n+1}\sum_{i=-n}^{n} \delta_{(\Psi^i x, \Psi^i y)} = \mu_0.$$

In particular

$$\frac{1}{2n+1}\sum_{i=-n}^{n} d(\Psi^n x, \Psi^n y) \to \int d(x,y) d\mu_0 = 0,$$

so $x=y$ since Ψ is weakly distal. □

We now describe the appropriate topological model of T, as defined in [4]. Let

$$p:[0,1) \to \{1,-1\}^Z$$

be the map which associates to $x \in [0,1)$ its T_α-name as described in Section 2. Observe that for x, $y \in [0,1)$, if we allow y to converge to x from the right then $p(y) \to p(x)$. Choosing $\{n_i\}$ such that $T_\alpha^{n_i}(0) \to x$ from the right we see that $p(x) \in \bar{\mathcal{O}}_\sigma(p(0))$, the closure of the orbit of $p(0)$ under the shift σ. Thus if we set $Z = (p[0,1))^-$, Z coincides with $\bar{\mathcal{O}}_\sigma(p(0))$. Note that $p(0)$ is the same whether one uses $[0,\frac{1}{2})$ to define T_α-names or $[0,\frac{1}{2}]$ as in [4].

The flow (Y,Ψ) of [4] is obtained by adding a homeomorphic copy of $B=\{z \in Z : x(0)=1\}$ to Z and delaying σ by one time unit on B. Thus it is topologically conjugate in an obvious way to the shift acting on the closed shift invariant set of sequences in $\{1,-1\}^Z$ which are obtained by doubling the 1's in each sequence $z \in Z$ (sequences $z \in B$ give rise to two doublings, one the shift of the other). We take this as the definition of (Y,Ψ).

PROPOSITION 5.3. *If α is poorly approximable, (Y,Ψ) is strictly ergodic, weakly distal and* POD.

Proof. We claim that each $z \in Z$ is, for each k, uniquely a concatenation of T_α-k-blocks having the structure of a T_α-name as described in Lemmas 2.5 and 2.6. (This does not mean that every $z \in Z$ is a T_α-name although all but countably many are, by [4, proposition 1.1].) To see this note that for any n, $z[-n,n]$ is part of a T_α-name and hence is a concatenation of T_α-k-blocks. The same is true of $z[-n',n']$ for $n' > n$. Each full k-block which together with its flanking ± 1's appears in $z[-n,n]$ also appears in $z[-n',n']$, which is part of a T_α-name, so by Lemma 2.5 that full k-block must be one of the k-blocks in the k-block structure of $z[-n',n']$. Thus as n increases the k-block structures of $z[n,n]$ extend each other so we see that z is a concatenation of k-blocks. The remaining properties of 2.5 and 2.6 follow immediately as they are "local" properties. This establishes the claim and as a consequence we have that every $y \in Y$ is uniquely a concatenation of T-k-blocks with the properties of Lemma 2.7.

For $z \in \{1,-1\}^{\mathbf{Z}}$ denote by z_+ the restriction of z to $[0,\infty)$. Taking $\{n_i\}$ such that $n_i \to +\infty$ and $T^{n_i}0 \to 0$ from the right we see that any finite segment of $p(0)$ appears in $p(0)_+$, hence in $A_k(0)$ for some k, since $p(0)_+$ begins with $A_k(0)$. On the other hand for any $z \in Z$, $A_k(0)$ appears in z (in fact $A_k(0)$ appears in any k'-block for k' sufficiently large) so $p(0) \in \bar{\mathcal{O}}_\sigma(z)$. Thus (Z,σ) is minimal. Moreover, by unique ergodicity of T_α, given r, for sufficiently large k any T_α-k-block has approximately correct frequency of finite strings of length less than r. Since any $z \in Z$ is a concatenation of k-blocks we obtain the unique ergodicity of (Z,σ). Minimality and unique ergodicity of (Y,Ψ) follow immediately. (So far we have used only irrationality of α.) If ν is the unique invariant probability on Y then the measure-preserving system (Ψ,ν) is isomorphic to T.

In the proof of Theorem 3.1 we showed that if y_1 and y_2 are T-names and $y_1 \notin \mathcal{O}_\sigma(y_2)$ then (y_1,y_2) contains a forcing configuration of k blocks for infinitely many k. However we only used the k-block structure of y_1 and y_2 so the same is true for y_1, $y_2 \in Y$. Any sufficiently long segment of a k-block differs substantially in $\bar{d}$ from its shift, so we see that

$$\overline{\lim}\, \frac{1}{2n+1} \# \{-n \leqq i \leqq n : y_1(i) \neq y_2(i)\} > 0.$$

On the other hand if $y_1 = \Psi^k y_2$, $k \neq 0$, then $\bar{d}(y_1,y_2)$ exists and is not 0 because (Ψ^k,ν) is ergodic (even weak-mixing). Thus (Y,Ψ) is weakly distal and by Proposition 3.2 it is POD. □

Since POD flows can never be distal Proposition 3.3 furnishes examples of weakly distal but not distal flows.

In connection with Propositions 3.2 and 3.3 we mention another example of a POD flow. Chacón's example of a weakly-mixing but not mixing automorphism (see [3], also [5], [7]) can be described as follows. Define k-blocks x_k by

$$x_1 = 0010, \qquad x_{k+1} = x_k x_k 1 x_k$$

and let $\bar{\mathcal{O}} \subset \{0,1\}^Z$ consist of those sequences x such that each finite segment of x is a segment of x_k for some k. It is easy to see that equivalently $x \in \bar{\mathcal{O}}$ if and only if x is, for each k, uniquely a concatenation of k-blocks with a 1 interposed between some k-block pairs. Then $\bar{\mathcal{O}}$ is closed, shift invariant and strictly ergodic with invariant probability μ, say. It was shown in [7] that (σ, μ) is simple (even has minimal self-joinings). Moreover the argument of [7] essentially showed that if x, $y \in \bar{\mathcal{O}}$ and $x \notin \mathcal{O}_\sigma(y)$ then one has "forcing" k-block configurations infinitely often, so that $\bar{d}(x, y) \neq 0$. Since $\bar{d}(x, \sigma^k x) \neq 0$ for $k \neq 0$ $(\bar{\mathcal{O}}, \sigma)$ is weakly distal and hence POD.

We conclude by mentioning several questions raised by this work. We have seen that if $|\alpha - p_n/q_n| = o(q_n^{-2})$ with $(p_n, q_n) = 1$ and q_n even, then T is not simple. It is likely that this is true without any restriction on q_n so T is simple if and only if α is poorly approximable. However one can ask whether T may still be prime when α is well approximable. This would require a new approach as previous proofs of primality have always been something close to establishing simplicity. If α is evenly well approximable T admits good cyclic approximation in the sense of [9]. This is incompatible with triviality of the centralizer, but not, as far as is known, with primality. A proof of primality for some S admitting good cyclic approximation might give some approach to the question of the category, with respect to the weak topology, of the class of prime automorphisms. One could also ask about existence of roots of T for well approximable α. Examples of automorphisms admitting good cyclic approximation without roots are known [2] but it is not known what the category of the class of such automorphisms is.

Finally, the question of loose Bernoullicity of the cartesian square $T \times T$ is open for any irrational α and also for Chacón's automorphism.

References

1. J. Aaronson and M. Keane, *The visits to zero of some deterministic random walks*, preprint.

2. R. V. Chacón, *Transformations having continuous spectrum*, J. Math. Mech. **16** (1966), 399–415.

3. R. V. Chacón, *Weakly mixing transformations which are not strongly mixing*, Proc. Am. Math. Soc. **22** (1969), 559–562.

4. H. Furstenberg, H. Keynes and L. Shapiro, *Prime flows in topological dynamics*, Isr. J. Math. **14** (1973), 26–38.

5. A. del Junco, *A simple measure-preserving transformation with trivial centralizer*, Pac. J. Math. **79** (1978), 357–362.

6. A. del Junco and D. Rudolph, *Universally disjoint measure-preserving systems*, to appear.

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

8. S. Kakutani, *Examples of ergodic measure-preserving transformations which are weakly mixing but not strongly mixing*, Springer Lecture Notes in Math. **318** (1973), 143–149.

9. A. B. Katok, Ya. G. Sinai and A. M. Stepin, *Theory of dynamical systems and general transformation groups with invariant measure*, J. Sov. Math. **7** (1977), 974–1065.

10. M. Keane, *Irrational rotations and quasi-ergodic measures*, Publ. des Séminaires de Math. (Fasc. I Prob.), Rennes, 1970–71.

11. D. S. Ornstein, *On the root problem in ergodic theory*, Proc. Sixth Berkeley Symp. Math. Stat. Prob. Vol. II, University of California Press, 1967, pp. 347–356.

12. D. Rudolph, *An example of a measure-preserving map with minimal self-joinings, and applications*, J. Analyse Math. **35** (1979), 97–122.

13. W. Veech, *Strict ergodicity in zero-dimensional dynamical systems and the Kronecker–Weyl theorem* mod 2, Trans. Am. Math. Soc. **140** (1969), 1–33.

14. W. Veech, *Well distributed sequences of integers*, Trans. Am. Math. Soc. **161** (1971), 63–70.

DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
COLUMBUS, OH 43210 USA