

# Twofold mixing implies threefold mixing for rank one transformations

STEVEN ARTHUR KALIKOW

Mathematics Department, Stanford University, Stanford CA 94305, USA

(Received 9 May 1983 and revised 23 August 1983)

*Abstract.* This paper addresses the following long-standing open question: If a stationary transformation on a probability space obeys the property

$$\lim_{M \rightarrow \infty} P(T^{-M}(A_1) \cap A_2) = P(A_1)P(A_2)$$

for all measurable sets  $A_1, A_2$ , does it follow that

$$\lim_{\substack{M \rightarrow \infty \\ N \rightarrow \infty}} P(T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap (A_3)) = P(A_1)P(A_2)P(A_3)$$

for all measurable sets  $A_1, A_2, A_3$ ? Here we answer the question affirmatively for a certain class of transformations.

## 1. Introduction

Let  $T$  be a stationary transformation on  $(\Omega, \mathcal{A}, P)$  where  $\Omega$  is a probability space endowed with  $\sigma$ -algebra  $\mathcal{A}$  and probability measure  $P$  on  $\mathcal{A}$ .  $T$  is called *mixing* (or *twofold mixing*) if

$$\lim_{M \rightarrow \infty} P(T^{-M}(A_1) \cap A_2) = P(A_1)P(A_2),$$

for all  $A_1, A_2 \in \mathcal{A}$ .  $T$  is called *threefold mixing* if

$$\lim_{\substack{M \rightarrow \infty \\ N \rightarrow \infty}} P(T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3) = P(A_1)P(A_2)P(A_3), \quad (1.1)$$

for all  $A_1, A_2, A_3 \in \mathcal{A}$ .

The problem, originally posed by Rohlin [2] and reproduced in the very last sentence of Paul Halmos' book [1], is whether or not twofold mixing implies threefold mixing. J. P. Thouvenot demonstrated that if it is true that twofold mixing implies threefold mixing for transformations of zero entropy, then it is true in general, (the reader need not know the meaning of zero entropy in order to read this paper).

This paper discusses a certain subclass of the zero entropy transformations known as the rank 1 transformations, to be defined in the next section, and proves that twofold mixing implies threefold mixing for rank 1 transformations. The way Halmos introduces the problem, I suspect he and Rohlin were expecting that twofold and threefold mixing were not equivalent and indeed, at first glance, it does not seem that there is any reason why they should be. If this was truly their expectation, then I think they would find our partial result unexpected.

2. Definition of rank one transformation

Let  $\{q_i\}_{i=1}^\infty$  be a sequence of positive integers and let  $\{a_{i,j}\}_{i \in \mathbb{N}, j \in \{1,2,\dots,q_i\}}$  be a doubly indexed sequence of non-negative integers. We define an increasing sequence of words, called *n-blocks*,  $n = 0, 1, 2, \dots$  by induction on  $n$ . A 0-block is just the word 'a'. For  $n > 0$ , an *n-block* is defined as follows. List an  $n - 1$ -block. Then list the letter 'b'  $a_{n,1}$  times. Then list another  $n - 1$ -block. Then list 'b'  $a_{n,2}$  times etc., concluding with a listing of 'b'  $a_{n,q_n}$  times. The resulting word is an *n-block*.

Let  $F$  be a finite word, i.e. a finite sequence of *a*'s and *b*'s. The frequency of occurrence of the word  $F$  in an *n-block* (e.g. the frequency of occurrence of 'aa' in *aaab* is  $\frac{2}{3}$  because 2 of the 3 two letter subwords of 'aaab' are 'aa') approaches a limit as  $n$  approaches  $\infty$ . Let  $P(F)$  be that limit.

Let  $\Omega = \{a, b\}^{\mathbb{Z}}$  be the space of doubly infinite sequences of *a*'s and *b*'s and let  $\mathcal{A}$  be the  $\sigma$ -algebra generated by finite cylinder sets (i.e. finite words.) It can easily be seen that we can induce a consistent stationary measure on  $\mathcal{A}$  by assigning to each finite word  $F$  the probability  $P(F)$  defined above. For each  $\omega \in \Omega$ ,  $i \in \mathbb{Z}$  let  $\omega_i$  be the  $i$ 'th letter of  $\omega$ . Now define the transformation  $T$  by  $(T(\omega))_i = \omega_{i+1}$ .  $(\Omega, \mathcal{A}, P, T)$  is called a *rank one transformation*. Ergodic theorists could simply define a rank one transformation to be any transformation in which every finite collection of sets can be arbitrarily well approximated by a union of rungs of a Rohlin tower, but the definition given above is equivalent and is the one used in this paper.

The values of  $\{q_i\}$  and  $\{a_{i,j}\}$  determine exactly which rank one transformation  $T$  we obtain. However, the terms  $\{q_i\}$  and  $\{a_{i,j}\}$  will never again occur in this paper. The remainder of § 2 is devoted to developing the language to be used in this paper for discussing rank 1 transformations.

From here on, presume the measure to be non-atomic (otherwise it cannot be mixing). The support of the measure is on words consisting of 0-blocks within 1-blocks within 2-blocks etc. The non-atomic presumption forces these words to be aperiodic. If the word is aperiodic, given the entire word it is possible to determine where a given letter is in an *n-block*. This is not obvious. Suppose the 2-block, to be denoted by  $B$ , is *aabaab*, (i.e.  $B$  is an abbreviation for the finite word *aabaab*). Suppose the 3-block is *BBbB*. Suppose you see  $\downarrow$ *aabaab*, where the arrow indicates the origin. You might suspect that the 'a' on the origin is the first 'a' of its 2-block but with more of the word, *aab\downarrow**aabaabaabb* you would realize that it is the fourth letter of its 2-block. It is an exercise to show that if the entire word is aperiodic, the entire word always resolves the question uniquely. This will be set forth and proved in the appendix.

The significance of this result is that it enables us to consider an *n-block* as a Rohlin tower as follows; (do not panic if you do not know what a Rohlin tower is, we are about to define it).

Every *n-block* is a finite word of *a*'s and *b*'s. Define  $h_n$  to be the length of that word.

Define  $f_n : \Omega \rightarrow \{\text{'error'}, 1, 2, \dots, h_n\}$  by  $f_n(\omega) = i$  if  $\omega_0$  (recall that  $\omega_0$  is the zero letter of  $\omega$ ) is the  $i$ 'th letter of its *n-block* and  $f_n(\omega) = \text{'error'}$  if  $\omega_0$  is one of the *b*'s which are not in any *n-block*.

Define  $R_{n,i} = \{\omega : f_n(\omega) = i\}$ . The finite sequence  $(R_{n,1}, R_{n,2}, \dots, R_{n,h_n})$  is called the  $n$ -tower.  $R_{n,i}$  is called the  $i$ 'th rung of the  $n$ -tower,  $\{\omega : f_n(\omega) = \text{error}\}$  is called the error set for the  $n$ -tower or merely the  $n$ -error set.

$\Omega$  is the disjoint union of the rungs and error set of the  $n$ -tower and  $T(R_{n,i}) = R_{n,i+1}$  for all  $i < h_n$ . This is precisely the definition of a Rohlin tower.

For each  $\omega \in \Omega$ , we will now define a quantity  $K_n(\omega)$ . Heuristically,  $K_n(\omega)$  can be defined to be the amount of time  $\omega$  will spend in the  $n$ -error set after leaving the  $n$ -tower, before re-entering the  $n$ -tower. We now define  $K_n(\omega)$  rigorously.

If  $f_n(\omega) \neq \text{error}$ , we define  $K_n(\omega)$  to be the number of  $b$ 's between the  $n$ -block containing  $\omega_0$  and the following  $n$ -block, i.e. if  $f_n(\omega) = i$ ,  $K_n(\omega)$  is the least  $j$  such that  $f_n(T^{j+1+h_n-i}(\omega)) = 1$ . If  $f_n(\omega) = \text{error}$  we leave  $K_n(\omega)$  undefined.

Fix  $\omega$  and a positive integer  $L$ . Let

$$\theta = \{i : 0 \leq i \leq L-1 \text{ and } f_n(T^i(\omega)) \neq \text{error}\}.$$

If  $f_n(\omega) = f_n(T^L\omega) = 1$ , then

$$\sum_{i \in \theta} \left( \frac{K_n(T^i\omega)}{h_n} \right) = L - \# \theta$$

where  $\#$  means 'number of elements in'. Thus if  $f_n(\omega) = f_n(T^L\omega) = 1$ , the average value of  $K_n(T^i\omega)$  over all  $i \in \theta$  is

$$\frac{1}{\# \theta} \left( \sum_{i \in \theta} K_n(T^i(\omega)) \right) = h_n \left( \frac{L - \# \theta}{\# \theta} \right).$$

Fix  $\omega$  with  $f_n(\omega) = 1$  and one can choose arbitrarily large  $L$  with  $f_n(T^L\omega) = 1$ . By the Birkhoff ergodic theorem (all rank one transformations are ergodic), as  $L$  approaches  $\infty$ ,  $(L - \# \theta) / \# \theta$  approaches  $P(n\text{-error set})$  and  $\# \theta / L$  approaches  $1 - P(n\text{-error set})$  so  $(L - \# \theta) / L$  approaches  $P(n\text{-error set}) / (1 - P(n\text{-error set}))$  and the average value of  $K_n(T^i(\omega))$  over all  $i \in \theta$  approaches  $E(K_n(\omega) | f_n(\omega) \neq \text{error})$ .

We have

$$E(K_n(\omega) | f_n(\omega) \neq \text{error}) = h_n P(n\text{-error set}) / (1 - P(n\text{-error set})) \tag{2.1}$$

Our non-atomic presumption implies that the word 'a' has positive probability. (Otherwise the all 'b' word would have probability 1). In order for 'a' to have positive probability, it can easily be seen that

$$\lim_{n \rightarrow \infty} P(n\text{-error set}) = 0 \tag{2.2}$$

(2.1) and (2.2) give

$$\lim_{n \rightarrow \infty} (1/h_n) E(K_n(\omega) | f_n(\omega) \neq \text{error}) = 0 \tag{2.3}$$

Since the  $n$ -towers generate the entire  $\sigma$ -algebra, and each rung of an  $n$ -tower is a union of rungs of an  $n+1$ -tower, it follows that every measurable set can be approximated arbitrarily well in measure by a union of rungs of some  $n$ -tower. By 'in measure' we mean that the distance between two sets is the measure of their symmetric difference. For any  $m$ , any union of rungs of the  $m$ -tower is a union of rungs of the  $n$ -tower for all  $n > m$ .

Throughout the remainder of this paper, presume  $A_1, A_2, A_3$  are three fixed unions of rungs of some fixed  $m_0$ -tower. By the previous paragraph it follows that if we can prove that rank one mixing implies (1.1) for  $A_1, A_2, A_3$  then we are done. During this paper, whenever we request that  $n$  be chosen sufficiently large so that some conditions are obeyed, or merely request that  $n$  be chosen sufficiently large, it should also be presumed that  $n$  is being chosen larger than  $m_0$ . For any such  $n$ , each  $R_{n,i}$  is either contained in  $A_1$  (respectively  $A_2, A_3$ ) or disjoint from  $A_1$  (respectively  $A_2, A_3$ ).

3. *Overview of proof*

Fix  $\omega \in \Omega$  and associate each integer  $i$  with the point  $T^i(\omega)$ . Fix a large  $M, N$  and for a set  $S$  of integers, we will say that  $S$  is *good* if the fraction of elements of  $S$  whose associated points are in  $T^{-(M+N)}A_1 \cap T^{-M}A_2 \cap A_3$  is approximately  $P(A_1)P(A_2)P(A_3)$ . Our goal is to prove that some pre-chosen interval of integers  $[0, L] \cap \mathbb{Z}$ , ('interval of integers' will always mean an interval intersected with  $\mathbb{Z}$ ), is usually good. This will be accomplished by expressing most of  $[0, L] \cap \mathbb{Z}$  as a disjoint union of good intervals of integers.

In § 5, we will develop certain intervals of integers which will be referred to as *n-generations*. For fixed  $n$ , the *n-generations* are disjoint and cover most of the integers. Each *n-generation* is a subset of an  $n + 1$ -generation. The main result of § 5 is that for fixed  $n$ , two successive *n-generations* are unlikely to both fail to be good.

In § 6, the result of § 5 is used to cover most of  $[0, L] \cap \mathbb{Z}$  with good generations. Let  $n_1 \ll n_2 \ll \dots \ll n_k$ . By the result of § 5 nearly half the  $n_k$ -generations are good. Of these  $n_k$ -generations which are bad (not good) nearly half of the  $n_{k-1}$ -generations within them are good. This nesting procedure is continued and this covers most of  $[0, L] \cap \mathbb{Z}$ .

4. *Implications of mixing*

In this section we assume only that  $T$  is a mixing transformation, we do not assume it to be rank 1.

LEMMA 4.1. *Let  $A$  be a measurable set. For all  $\epsilon > 0$ , there exists  $K$  such that for all  $k \geq K$ , for any sequence  $n_1, n_2, \dots, n_k$  with  $k$  terms,*

$$P(|(1/k) \# (A \cap \{T^{n_1}(\omega), T^{n_2}(\omega), \dots, T^{n_k}(\omega)\}) - P(A)| \geq \epsilon) < \epsilon.$$

*Proof.* For every  $m$ , let  $X_m(\omega)$  be defined by

$$X_m(\omega) = \begin{cases} 1 & \text{if } T^m(\omega) \in A, \\ 0 & \text{otherwise.} \end{cases}$$

Let  $\delta > 0$ . By mixing, select  $I$  so that if  $|m - n| > I$ ,

$$|E(X_m X_n) - (P(A))^2| < \delta.$$

Fix a  $K$  and select integers  $n_1 < n_2 < \dots < n_k$ . Since, for all  $i, n_{i+1} - n_i \geq 1$  it follows that for  $|i - j| > I, |n_i - n_j| > I$ . Let

$$\Gamma = (1/k) \# (A \cap \{T^{n_1}(\omega), T^{n_2}(\omega), \dots, T^{n_k}(\omega)\}) = \left(\frac{1}{k}\right) \sum_{i=1}^k X_{n_i}.$$

$E(\Gamma) = P(A)$  and

$$\begin{aligned} \text{(Variance of } \Gamma) &= E(\Gamma^2) - (E\Gamma)^2 = E\left(\left(\frac{1}{k^2}\right) \sum_{i=1}^k \sum_{j=1}^k X_{n_i} X_{n_j}\right) - (P(A))^2 \\ &= E\left(\left(\frac{1}{k^2}\right) \sum_{i=1}^k \sum_{j=1}^k (X_{n_i} X_{n_j} - (P(A))^2)\right) \\ &< \left(\frac{1}{k^2}\right) \sum_{i=1}^k \sum_{j=1}^k \delta + \left(\frac{1}{k^2}\right) \sum_{\{i,j\}, |i-j| \leq I} (1) \\ &< \delta + \frac{(2I+1)k}{k^2} = \delta + \frac{2I+1}{k}. \end{aligned}$$

It follows by Chebychevs inequality that  $(P((\Gamma - PA) > \epsilon)) < (\sqrt{\delta + (2I+1)/K})/\epsilon$  which can be made less than  $\epsilon$  by first choosing  $\delta$  small, finding  $I$  in terms of  $\delta$ , then choosing  $K$  large. □

LEMMA 4.2. For any measurable sets  $B_1, B_2$  and any  $\epsilon > 0$ , there exists  $N$  such that for any  $M > N$  and  $n > N$ ,

$$E(|(1/n) \# (T^{-M}(B_1) \cap B_2 \cap \{\omega, T(\omega), \dots, T^{n-1}(\omega)\}) - P(B_1)P(B_2))|) < \epsilon.$$

*Proof.* For mixing transformations  $T$ , it is known that  $T \times T$  is ergodic. By the Birkhoff ergodic theorem, this implies that if  $\omega_1, \omega_2$  are independently chosen from  $\Omega$ , (i.e.  $(\omega_1, \omega_2)$  is chosen from  $\Omega \times \Omega$  in accordance with measure  $P \times P$ ) then we can select  $N_1$  so that

$$E(|(1/N_1) \# \{i: 0 \leq i \leq N_1 - 1 \text{ and } T^i(\omega_1) \in B_1 \text{ and } T^i(\omega_2) \in B_2\} - P(B_1)P(B_2))|) < \epsilon/2.$$

Since  $T$  is mixing, for  $N_2$  sufficiently large, the joint distribution of the events  $\{T^i(\omega_1) \in B_1, T^i(\omega_2) \in B_2\}_{0 \leq i \leq N_1-1}$  is very close to the joint distribution of the events  $\{T^{i+N_2}(\omega) \in B_1, T^i(\omega) \in B_2\}_{0 \leq i \leq N_1-1}$ , so we can choose  $N_2$  such that for all  $M \geq N_2$ ,

$$E(|(1/N_1) \# (T^{-M}(B_1) \cap B_2 \cap \{\omega, T(\omega), \dots, T^{N_1-1}(\omega)\}) - P(B_1)P(B_2))|) < \epsilon/2.$$

Select an integer  $N_3 > 2/\epsilon$ . Let  $N = \max\{N_1, N_3, N_2\}$ . Suppose  $M > N$  and  $n > N$ . Then there exist integers  $K$  and  $R$  such that,  $0 \leq R < N_1, K \geq N_3 > 2/\epsilon$  and  $n = N_1 K + R$ .

$$\begin{aligned} &E\left(\left|\frac{1}{n} \# (T^{-M}(B_1) \cap B_2 \cap \{\omega, T(\omega), \dots, T^{n-1}(\omega)\}) - P(B_1)P(B_2)\right|\right) \\ &\leq \frac{1}{K} \sum_{i=1}^K \frac{1}{N_1} E(|\# (T^{-M}(B_1) \cap B_2 \cap \{T^{(i-1)N_1}\omega, T^{(i-1)N_1+1}(\omega), \dots, T^{iN_1-1}(\omega)\}) \\ &\quad - P(B_1)P(B_2))|) + R/n \\ &\leq \epsilon/2 + 1/K < \epsilon. \end{aligned}$$
□

5. Proof that at least one of almost every two successive  $n$ -generations are good

For the remainder of this paper,  $T$  is mixing and rank 1.  $A_1, A_2, A_3$  are unions of rungs of an  $m_0$ -tower and from here on in, whenever we refer to an  $n$ -tower it is presumed that  $n > m_0$ .

Define  $l_1(n, M, N, \omega)$ ,  $l_2(n, M, N, \omega)$ ,  $l_3(n, M, N, \omega)$  to be  $f_n(T^{M+N}\omega)$ ,  $f_n(T^M\omega)$ ,  $f_n(\omega)$  respectively. We will merely write  $l_1(\omega)$ ,  $l_2(\omega)$ ,  $l_3(\omega)$  when  $n, M, N$  are understood or just  $l_1, l_2, l_3$  when  $\omega$  is also understood. Note that

$$R_{n,l_1} \subset A_1, R_{n,l_2} \subset A_2, R_{n,l_3} \subset A_3 \quad \text{iff} \quad \omega \in T^{-(M+N)}A_1 \cap T^{-M}(A_2) \cap A_3. \quad (5.1)$$

We now define an  $n$ -generation for  $N, M, \omega$ . Some or all of  $n, N, M, \omega$  will be dropped when they are understood, (e.g. we can just refer to an  $n$ -generation, or merely a generation). Fix  $n, N, M$ .  $i$  is said to be *the beginning of a generation* if the set  $\{l_1(T^i(\omega)), l_2(T^i(\omega)), l_3(T^i(\omega))\}$  contains 1 but does not contain error.  $i$  is said to be *the end of a generation* if the set  $\{l_1(T^i(\omega)), l_2(T^i(\omega)), l_3(T^i(\omega))\}$  contains  $h_n$  but does not contain error. For  $c, d \in \mathbb{Z}$ ,  $[c, d]$  is said to be a *generation* if  $c$  is the beginning of a generation and  $d$  is the least integer such that  $d \geq c$  and  $d$  is the end of a generation. Here  $[c, d]$  is short for  $[c, d] \cap \mathbb{Z}$ .

*Remark (5.2).* Fixing  $n, M, N, \omega$ , the generations are all disjoint and their union is the collection of all integers  $i$  such that error  $\notin \{l_1(T^i(\omega)), l_2(T^i(\omega)), l_3(T^i(\omega))\}$ . For any such integer  $i$ ,  $i$  is in generation  $[c, d]$  where  $c = i + 1 - \min\{l_1(T^i(\omega)), l_2(T^i(\omega)), l_3(T^i(\omega))\}$  and  $d = i + h_n - \max\{l_1(T^i(\omega)), l_2(T^i(\omega)), l_3(T^i(\omega))\}$ . For any  $j$  in that generation  $l_k(T^j(\omega)) = l_k(T^i(\omega)) + j - i$  for  $k \in \{1, 2, 3\}$ . Thus, by (5.1),  $l_1(T^i(\omega)), l_2(T^i(\omega)), l_3(T^i(\omega))$  not only determine the generation containing  $i$ , but also determine which values of  $j$  in that generation have the property  $T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3$ .

When  $n, M, N, \omega$  are all understood we will refer to the generation containing 0 as *the current generation*. The generation following the current generation will be referred to as *the successor generation*. If we replace  $i$  with 0 in remark 5.2, we see that knowledge of the values of  $l_1, l_2, l_3$  provides sufficient information for studying completely the current generation. We will now develop additional notation for studying the successor generation.

Let  $K_n(\omega)$  be defined as in § 2. Let  $K_{n,1}(\omega), K_{n,2}(\omega), K_{n,3}(\omega)$  be  $K_n(T^{(M+N)}(\omega)), K_n T^M(\omega), K_n(\omega)$  respectively. Either or both of the terms  $\omega, n$  can be dropped when understood (e.g. we can just refer to  $K_1, K_2, K_3$  when  $\omega$  and  $n$  are both held fixed). Suppose  $l_1, l_2, l_3$  have a unique maximum and that  $\sigma$  is a permutation of  $\{1, 2, 3\}$  such that  $l_{\sigma(1)} \leq l_{\sigma(2)} < l_{\sigma(3)}$ . We define  $\hat{K}$  to be  $K_{\sigma(3)}$  (or we can write  $\hat{K}_n(\omega)$  to indicate  $n$  and  $\omega$ ). Note that  $\hat{K}$  is well defined if  $l_1, l_2, l_3$  have a unique maximum.  $(l_1, l_2, l_3, \hat{K})$  is said to be *acceptable* if  $l_1, l_2, l_3$  have a unique maximum and  $\hat{K} < l_{\sigma(3)} - l_{\sigma(2)}$  where  $\sigma$  is a permutation of  $\{1, 2, 3\}$  with  $l_{\sigma(1)} \leq l_{\sigma(2)} < l_{\sigma(3)}$ .

*Remark (5.3).* Suppose  $(l_1, l_2, l_3, \hat{K})$  is acceptable and that  $\sigma$  is a permutation of  $\{1, 2, 3\}$  with  $l_{\sigma(1)} \leq l_{\sigma(2)} < l_{\sigma(3)}$ . Then the successor generation is  $[c, d]$  where  $c = h_n + 1 + \hat{K} - l_{\sigma(3)}$  and  $d = h_n - l_{\sigma(2)}$ . For any integer  $j \in [c, d]$ ,  $l_i(T^j\omega) = l_i(\omega) + j$  for  $i \in \{\sigma(1), \sigma(2)\}$  and  $l_{\sigma(3)}(T^j(\omega)) = l_{\sigma(3)}(\omega) + j - h_n - \hat{K}$ . Thus, by (5.1), knowledge of  $(l_1, l_2, l_3, \hat{K})$  gives sufficient information not only to determine the successor generation, but also to determine which values of  $j$  in the successor generation satisfy

$$T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3.$$

A generation  $[c, d]$  is said to be  $\epsilon$ -good if

$$\left| \frac{1}{d-c+1} \#(T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3 \cap \{T^i(\omega) : c \leq i \leq d\}) - P(A_1)P(A_2)P(A_3) \right| < \epsilon.$$

By remarks (5.2) and (5.3),  $(l_1, l_2, l_3, \hat{K})$ , if acceptable, determines whether or not the current generation, and whether or not the successor generation is  $\epsilon$ -good.  $\epsilon$ -bad means not  $\epsilon$ -good.

*Remark.* (5.4). We now give an overview of the program for the remainder of § 5. First we define a small bad set off which  $(l_1, l_2, l_3, \hat{K})$  is acceptable and pairs of the three  $T^i, T^{M+i}, T^{N+M+i}$ , mix well as  $i$  runs over either the current or the successor generation. Then comes the heart of the proof, lemmas 5.18 and 5.26, where we use lemma 4.1 to exhibit a number  $Q$  dependent only on  $\epsilon$  and not on  $l_1, l_2, l_3$  or  $n$  such that off the bad set, conditioned on knowledge of  $l_{\sigma(1)}$  and  $l_{\sigma(3)}$  there are at most  $2Q$  choices of  $l_{\sigma(2)}$  which make the current generation  $\epsilon$ -bad, and conditioned on knowledge of  $l_1, l_2, l_3$  there are at most  $2Q$  small values of  $\hat{K}$  which make the successor generation  $\epsilon$ -bad. Next we show that for  $n$  large,  $\hat{K}$  does not take on any fixed value with high probability. We conclude by arguing that for given  $l_{\sigma(1)}, l_{\sigma(3)}$  there are only  $2Q$  values of  $l_{\sigma(2)}$  for each of which there are only  $2Q$  values of  $\hat{K}$  so that altogether there are only  $(2Q)(2Q) = 4Q^2$  values of  $\hat{K}$  for which any value of  $l_{\sigma(3)}$  makes both the current and successor generation bad. Since each of those  $4Q^2$  values has small probability it is unlikely for both generations to be bad.

The crux of the proof is as follows. To analyze  $l_1, l_2, l_3$  directly we would need threefold mixing, which we are not given. However we can study  $l_{\sigma(1)}, l_{\sigma(3)}, \hat{K}$  with just twofold mixing and the previous paragraph reduces the study of  $l_1, l_2, l_3$  to  $l_{\sigma(1)}, l_{\sigma(3)}, \hat{K}$ .

We will abbreviate the vector  $(l_1, l_2, l_3)$  to just  $\vec{l}$  and abbreviate  $(l_1, l_2, l_3, K_1, K_2, K_3)$  to  $(\vec{l}, \vec{K})$ . Let  $\delta > 0$ . We will call  $\vec{l}$   $\delta$ -reasonable if

$$\text{error} \notin \{l_1, l_2, l_3\}$$

and

$$(1 - \sqrt{\delta})h_n \geq \max \{|l_1 - l_2|, |l_1 - l_3|, |l_2 - l_3|\}$$

and

$$\sqrt{\delta}h_n < \min \{|l_1 - l_2|, |l_1 - l_3|, |l_2 - l_3|\}$$

and call  $(\vec{l}, \vec{K})$   $\delta$ -reasonable if  $\vec{l}$  is  $\delta$ -reasonable and  $\delta h_n > \max \{K_1, K_2, K_3\}$ . If  $(\vec{l}, \vec{K})$  (respectively  $\vec{l}$ ) is not  $\delta$ -reasonable we will call it  $\delta$ -unreasonable. Notice that if  $(\vec{l}, \vec{K})$  is  $\delta$ -reasonable, for any  $\delta \leq 1$ , then  $(l_1, l_2, l_3, \hat{K})$  is acceptable.

LEMMA (5.5). *Let  $0 < \delta < 1$ . Then if  $n$  is chosen so that (probability of  $n$ -error set)  $< \delta^2 / (1 + \delta^2)$ , there exists  $N_n$  so that if  $M, N > N_n$ , then*

$$P((\vec{l}, \vec{K}) \text{ is } \delta\text{-unreasonable}) < 3\delta^2 + 3\delta + 18\sqrt{\delta},$$

(recall that  $(\vec{l}, \vec{K})$  depends on  $n, M, N$ ).

*Proof.*

$$P(\text{error} \in \{l_1, l_2, l_3\}) \leq 3P(n\text{-error}) < 3\delta^2. \tag{5.6}$$

Conditioning on the event  $\{\text{error} \notin \{l_1, l_2, l_3\}\}$ , each  $l_i$  takes on values 1 through  $h_n$  with uniform distribution. If the  $l_i$ 's were pairwise independent  $P(|l_i - l_j| < \sqrt{\delta} h_n)$  and  $P(|l_i - l_j| > (1 - \sqrt{\delta})h_n)$  would both be less than  $2\sqrt{\delta}$ ,  $i, j \in \{1, 2, 3\}$ ,  $i \neq j$ . When  $M$  and  $N$  are sufficiently large the  $l_i$ 's become close to pairwise independent. Therefore, there exists  $N_n$  such that for  $M > N_n$ ,  $N > N_n$ ,  $P(|l_i - l_j| < \sqrt{\delta} h_n)$  and  $P(|l_i - l_j| > (1 - \sqrt{\delta})h_n)$  are both less than  $3\sqrt{\delta}$ . Thus, for  $M > N_n$ ,  $N > N_n$ ,

$$P((1 - \sqrt{\delta})h_n < \max\{|l_1 - l_2|, |l_1 - l_3|, |l_2 - l_3|\} \text{ or } \sqrt{\delta} h_n > \min\{|l_1 - l_2|, |l_1 - l_3|, |l_2 - l_3|\} | \text{error} \notin \{l_1, l_2, l_3\}) < 18\sqrt{\delta}. \tag{5.7}$$

In the statement of the lemma we are given  $P(n\text{-error}) < \delta^2/(1 + \delta^2)$ . By (2.1), for  $i \in \{1, 2, 3\}$ ,  $E(K_i | l_i \neq \text{error}) = h_n \delta^2$ . Thus

$$P(K_i \geq h_n \delta | l_i \neq \text{error}) < \delta \quad \text{for } i \in \{1, 2, 3\}. \tag{5.8}$$

The lemma follows from 5.6, 5.7 and 5.8. □

Our next lemma essentially says that if  $(\vec{l}, \vec{K})$  is  $\delta$ -reasonable for small  $\delta$ , then a certain pair of  $T^i, T^{M+i}, T^{M+N+i}$  mixes well as  $i$  runs over the current generation, and another pair mixes well as  $i$  runs over the successor generation. Let  $\sigma$  be a permutation of  $\{1, 2, 3\}$  such that  $l_{\sigma(1)} \leq l_{\sigma(2)} \leq l_{\sigma(3)}$ . Let

$$\begin{aligned} Y_1 &= \{i: 1 \leq i \leq h_n - l_{\sigma(3)} + l_{\sigma(1)}, R_{n,i} \subset A_{\sigma(1)} \text{ and } R_{n,i-l_{\sigma(1)}+l_{\sigma(3)}} \subset A_{\sigma(3)}\}, \\ Y_2 &= \{i: 1 \leq i \leq l_{\sigma(3)} - l_{\sigma(2)}, R_{n,i+l_{\sigma(1)}+h_n-l_{\sigma(3)}} \subset A_{\sigma(1)} \text{ and } R_{n,i+l_{\sigma(2)}+h_n-l_{\sigma(3)}} \subset A_{\sigma(2)}\}. \end{aligned} \tag{5.9}$$

LEMMA (5.10). *Let  $\delta > 0$ . If  $n$  is chosen sufficiently large and  $\vec{l}$  is  $\delta$ -reasonable, then if we define  $\sigma$  to be the permutation of  $\{1, 2, 3\}$  such that  $l_{\sigma(1)} < l_{\sigma(2)} < l_{\sigma(3)}$ , ( $\delta$ -reasonable implies the  $l_i$ 's unequal) and define  $Y_1, Y_2$  as in (5.9) then:*

$$|(\#(Y_1)/(h_n - l_{\sigma(3)} + l_{\sigma(1)})) - P(A_{\sigma(1)})P(A_{\sigma(3)})| < \delta \tag{5.11}$$

and

$$|(\#(Y_2)/(l_{\sigma(3)} - l_{\sigma(2)})) - P(A_{\sigma(1)})P(A_{\sigma(2)})| < \delta. \tag{5.12}$$

*Proof.* Select  $\hat{\delta} > 0$  so that  $\hat{\delta} < \delta^2/4$ . By lemma 4.2, select  $n_1$  so that for any  $n_2, n_3 \geq n_1$ ,

$$E(|(1/n_2)[\#(A_{\sigma(1)} \cap T^{-n_3}(A_{\sigma(3)}) \cap \{\omega, T(\omega), \dots, T^{n_2-1}(\omega)\})] - P(A_{\sigma(1)})P(A_{\sigma(3)})|) < \hat{\delta}. \tag{5.13}$$

Now choose  $n$  sufficiently large that

$$P(n\text{-error}) < \hat{\delta} \text{ and } h_n > n_1/\hat{\delta}. \tag{5.14}$$

For all  $i$  such that  $1 \leq i \leq h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1$ , let

$$\begin{aligned} Q_i &= \{j: i \leq j \leq i + n_1 - 1, R_{n,j} \subset A_{\sigma(1)}, R_{n,j-l_{\sigma(1)}+l_{\sigma(3)}} \subset A_{\sigma(3)}\} \\ &= \{j: i \leq j \leq i + n_1 - 1, j \in Y_1\}. \end{aligned}$$

A simple counting argument indicates that if we sum  $(1/n) \sum_{i=1}^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} \# Q_i$ , each element of  $Y_1$  gets counted exactly once except for those  $j \in Y_1$  where  $1 \leq j \leq n_1$  or  $h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 < j \leq h_n - l_{\sigma(3)} + l_{\sigma(1)}$  each of which is counted less than once.



It follows that

$$\left| \# Y_1 - \left( \frac{1}{n_1} \right)^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} \sum_{i=1}^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} \# Q_i \right| < 2n_1. \tag{5.15}$$

For each  $\hat{\omega} \in \Omega$ , (we use  $\hat{\omega}$  instead of  $\omega$  because  $\omega$  is used to define the vector  $(\vec{l}, \vec{K})$  and  $\hat{\omega}$  will be allowed to vary after  $(\vec{l}, \vec{K})$  is selected and held fixed) let

$$N(\hat{\omega}) = \# \{j: 0 \leq j < n_1 - 1 \text{ and } T^j(\hat{\omega}) \in A_{\sigma(1)} \cap T^{l_{\sigma(1)} - l_{\sigma(3)}} A_{\sigma(3)}\}.$$

Since  $\vec{l}$  is  $\delta$ -reasonable,  $l_{\sigma(3)} - l_{\sigma(1)} > \sqrt{\delta} h_n > \hat{\delta} h_n > n_1$ , the last inequality by (5.14). Applying (5.13), with  $n_2, n_3$  replaced by  $n_1, l_{\sigma(3)} - l_{\sigma(1)}$  respectively, we get

$$E(|(1/n_1)N(\hat{\omega}) - P(A_{\sigma(1)})P(A_{\sigma(3)})|) < \hat{\delta}. \tag{5.16}$$

Furthermore, for  $\hat{\omega} \in R_{n,i}$ ,  $1 \leq i \leq h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1$ , it is immediate from the definitions of  $N(\hat{\omega})$  and of  $Q_i$  that  $N(\hat{\omega}) = \# Q_i$ . By (5.14), each rung has probability more than  $(1 - \hat{\delta})/h_n$ . Applying (5.16),

$$\begin{aligned} \hat{\delta} &> ((1 - \hat{\delta})/h_n) \sum_{i=1}^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} E(|(1/n_1)N(\hat{\omega}) - P(A_{\sigma(1)})P(A_{\sigma(3)})| | \hat{\omega} \in R_{n,i}) \\ &= ((1 - \hat{\delta})/h_n) \sum_{i=1}^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} |(1/n_1) \# Q_i - P(A_{\sigma(1)})P(A_{\sigma(3)})| \\ &\geq ((1 - \hat{\delta})/h_n) (1/n) \left[ \sum_{i=1}^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} (\# Q_i) \right] \\ &\quad - (h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1) P(A_{\sigma(3)})P(A_{\sigma(1)}) \\ &\geq ((1 - \hat{\delta})/h_n) \left( |\# Y_1 - (h_n - l_{\sigma(3)} + l_{\sigma(1)})P(A_{\sigma(1)})P(A_{\sigma(3)})| \right. \\ &\quad \left. - n_1 P(A_{\sigma(1)})P(A_{\sigma(3)}) - \left| \left( \frac{1}{n} \right)^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} \sum_{i=1}^{h_n - l_{\sigma(3)} + l_{\sigma(1)} - n_1 + 1} \#(Q_i) - \#(Y_1) \right| \right) \\ &\geq ((1 - \hat{\delta})/h_n) (|\# Y_1 - (h_n - l_{\sigma(3)} + l_{\sigma(1)})P(A_{\sigma(1)})P(A_{\sigma(3)})| - 3n_1). \end{aligned}$$

The last inequality follows from (5.15) and the fact that  $P(A_{\sigma(1)})P(A_{\sigma(3)}) < 1$ . Solving for  $|\# Y_1 - (h_n - l_{\sigma(3)} - l_{\sigma(1)})P(A_{\sigma(3)})P(A_{\sigma(1)})|$  and then dividing through by  $h_n - l_{\sigma(3)} - l_{\sigma(1)}$  gives

$$\left| \frac{1}{h_n - l_{\sigma(3)} + l_{\sigma(1)}} \#(Y_1) - P(A_{\sigma(3)})P(A_{\sigma(1)}) \right| \leq \frac{h_n \hat{\delta}}{(1 - \hat{\delta})(h_n - l_{\sigma(3)} + l_{\sigma(1)})} + \frac{3n_1}{h_n - l_{\sigma(3)} + l_{\sigma(1)}}. \tag{5.17}$$

Because  $\vec{l}$  is  $\delta$ -reasonable  $h_n - l_{\sigma(3)} + l_{\sigma(1)} > \sqrt{\delta} h_n$ . Thus

$$\frac{h_n \hat{\delta}}{(1 - \hat{\delta})(h_n - l_{\sigma(3)} + l_{\sigma(1)})} < \frac{\hat{\delta}}{(1 - \hat{\delta})\sqrt{\delta}} < \frac{\hat{\delta}}{\sqrt{\delta}},$$

and applying (5.14),

$$3(n_1)/(h_n - l_{\sigma(3)} + l_{\sigma(1)}) < 3n_1/\sqrt{\delta} h_n < 3\hat{\delta}/\sqrt{\delta}.$$

Thus the right-hand side of (5.17) is less than  $4\hat{\delta}/\sqrt{\delta} < \delta$  and (5.11) is proved. (5.12) is proved similarly.  $\square$

*Remark.* We are about to prove the heart of the proof, lemmas (5.18) and (5.26). This remark is unnecessary in reading this paper, but it should be noted that these lemmas and their proofs are modifications of a more generalized, easier to state, and easier to prove result: for every two-fold mixing transformation  $T$ , and set  $A_1$ , and any  $\varepsilon > 0$ , there exists an integer  $Q$  such that for any set  $A_2$  with  $P(A_2) > \varepsilon$ ,

$$\#\{i: |P(T^{-i}(A_1) \cap A_2) - P(A_1)P(A_2)| > \varepsilon\} < Q.$$

This result follows easily from lemma 4.1 and is much simpler to prove than lemmas (5.18) and (5.26). Here is a corollary of that result:

Let  $T$  be a mixing transformation and  $A_1, A_2, A_3$  be any three sets. Then for any  $\varepsilon > 0$ ,  $\#\{i: |P(A_1 \cap T^{-i}(A_2) \cap T^{-n}(A_3)) - P(A_1)P(A_2)P(A_3)| > \varepsilon\}$  is bounded above with a bound that does not depend on  $n$  except that  $n$  be sufficiently large.

LEMMA (5.18). *Let  $\varepsilon > 0$  be given. Select  $\delta > 0$  with  $\delta + (\sqrt{\delta}/1 - \delta) < \varepsilon$ . Then there is a positive integer  $Q$  so that if  $n$  is chosen large enough so that  $P(n\text{-error set}) < \delta$  and so that lemma (5.10) holds, if  $\sigma$  is defined by  $l_{\sigma(1)} < l_{\sigma(2)} < l_{\sigma(3)}$ , if  $\sigma, l_{\sigma(1)}$  and  $l_{\sigma(3)}$  are known, and it is known that  $\bar{l}$  is  $\delta$ -reasonable, then there are less than  $2Q$  possible values for  $l_{\sigma(2)}$  (between  $l_{\sigma(1)}$  and  $l_{\sigma(3)}$ ) such that the current generation is forced to be  $\varepsilon$ -bad.*

*Proof.* For simplicity, presume  $\sigma$  is the identity map. By (4.1), select  $Q$  so that for any  $Q$  integers  $n_1, n_2, \dots, n_Q$ ,

$$E \left( \left| \frac{1}{Q} \#(A_2 \cap \{T^{n_1}(\omega), T^{n_2}(\omega), \dots, T^{n_Q}(\omega)\}) - P(A_2) \right| \right) < \delta. \tag{5.19}$$

Now suppose we can make  $2Q$  choices,  $n_1, n_2, \dots, n_{2Q}$ , of values for  $l_2$ , between  $l_1$  and  $l_3$ , such that  $(l_1, l_2, l_3)$  force the current generation to be  $\varepsilon$ -bad. For any particular choice of  $l_2$ , let

$$\theta_{l_1, l_2, l_3} = \{i: 1 \leq i \leq h_n - l_3 + l_1, R_{n,i} \subset A_1, R_{n, i-l_1+l_2} \subset A_2, R_{n, i-l_1+l_3} \subset A_3\}.$$

The current generation is  $\varepsilon$ -good iff

$$|(1/(h_n - l_3 + l_1)) \#(\theta_{l_1, l_2, l_3}) - P(A_1)P(A_2)P(A_3)| \leq \varepsilon.$$

Thus, for  $1 \leq i \leq 2Q$ ,

$$|(1/(h_n - l_3 + l_1)) \# \theta_{l_1, n_i, l_3} - P(A_1)P(A_2)P(A_3)| > \varepsilon.$$

Hence, either there are  $Q$  values for  $i$  such that

$$(1/h_n - l_3 + l_1) \# \theta_{l_1, n_i, l_3} - P(A_1)P(A_2)P(A_3) > \varepsilon \tag{5.20}$$

or there are  $Q$  values for  $i$  such that

$$(1/(h_n - l_3 + l_1)) \# \theta_{l_1, n_i, l_3} - P(A_1)P(A_2)P(A_3) < -\varepsilon.$$

In either case the remainder of our proof by contradiction is essentially the same, so assume without loss of generality that (5.20) holds for  $1 \leq i \leq Q$ .

Let  $Y_1$  be as in (5.9) (keep in mind that here we presume  $\sigma$  to be the identity transformation). Notice that for any choice of  $l_2$ ,  $\theta_{l_1, l_2, l_3} \subset Y_1$ . For  $j \in Y_1$ , let

$$c_j = \#\{i: 1 \leq i \leq Q \text{ and } j \in \theta_{l_1, n_i, l_3}\}.$$

Then by (5.20),

$$\sum_{j \in Y_1} c_j = \sum_{i=1}^Q \# \theta_{l_1, n_i, l_3} > Q(h_n - l_3 + l_1)(P(A_1)P(A_2)P(A_3) + \varepsilon). \tag{5.21}$$

For  $j \in Y_1, j \in \theta_{l_1, n, l_3}$  iff  $R_{n, j-l_1+n_i} \subset A_2$  iff for all  $\hat{\omega} \in R_{n, j}, T^{n_i-l_1}(\hat{\omega}) \in A_2$ . Thus for  $j \in Y_1$  and  $\hat{\omega} \in R_{n, j}$

$$\#\{i: 1 \leq i \leq Q \text{ and } T^{n_i-l_1}(\hat{\omega}) \in A_2\} = c_j. \tag{5.22}$$

Since  $P(n\text{-error}) < \delta$ , each rung has probability more than  $(1/h_n)(1-\delta)$ . Since  $T^{-l_1}(\hat{\omega})$  has the same distribution as  $\hat{\omega}$ , by (5.19) and (5.22)

$$\delta > E(|(1/Q)\#\{A_2 \cap \{T^{n_1}(T^{-l_1}\hat{\omega}), \dots, T^{n_Q}(T^{-l_1}\hat{\omega})\}\} - P(A_2)|)$$

$$> (1/h_n)(1-\delta)$$

$$\times \sum_{j \in Y_1} E(|(1/Q)\#\{A_2 \cap \{T^{n_1-l_1}(\hat{\omega}), \dots, T^{n_Q-l_1}(\hat{\omega})\}\} - P(A_2)| | \hat{\omega} \in R_{n, j})$$

$$> (1/h_n)(1-\delta) \sum_{j \in Y_1} |(1/Q)c_j - P(A_2)| \geq (1/h_n)(1-\delta) |(1/Q) \sum_{j \in Y_1} c_j - (\#Y_1)P(A_2)|$$

$$\geq (1/h_n)(1-\delta) \left( \left| \left[ (1/Q) \sum_{j \in Y_1} c_j \right] - P(A_1)P(A_2)P(A_3)(h_n - l_3 + l_1) \right| \right. \\ \left. - |P(A_1)P(A_2)P(A_3)(h_n - l_3 + l_1) - \#(Y_1)P(A_2)| \right)$$

By (5.21),

$$\left| (1/Q) \left( \sum_{j \in Y_1} c_j \right) - P(A_1)P(A_2)P(A_3)(h_n - l_3 + l_1) \right| \geq \varepsilon(h_n - l_3 + l_1). \tag{5.24}$$

By (5.11)

$$|P(A_1)P(A_2)P(A_3)(h_n - l_3 + l_1) - (\#Y_1)P(A_2)| < \delta P(A_2)(h_n - l_3 + l_1) \\ \leq \delta(h_n - l_3 + l_1). \tag{5.25}$$

Since  $\vec{I}$  is assumed to be  $\delta$ -reasonable,  $(h_n - l_3 + l_1)/h_n > \sqrt{\delta}$ . Plugging (5.24) and (5.25) into the right hand side of (5.23) gives

$$\delta > (1/h_n)(1-\delta)(\varepsilon - \delta)(h_n - l_3 + l_1) > (\sqrt{\delta})(1-\delta)(\varepsilon - \delta),$$

and hence  $\varepsilon < \sqrt{\delta}/(1-\delta) + \delta$  which contradicts the choice of  $\delta$  in the statement of lemma (5.18). □

LEMMA (5.26). Let  $\varepsilon > 0$  be given. Select  $\delta, 0 < \delta < 1$  such that

$$((\delta/(\sqrt{\delta} - \delta))(1/(1-\delta)) + 4) < \varepsilon.$$

Then there is a positive integer  $Q$  so that if  $n$  is chosen large enough so that  $P(n\text{-error set}) < \delta$  and so that lemma 5.10 holds, then if  $l_1, l_2, l_3$  are known and it is known that  $(\vec{I}, \vec{K})$  is  $\delta$ -reasonable, then there are less than  $2Q$  possible values for  $\hat{K}$  (since we presume  $(\vec{I}, \vec{K})$   $\delta$ -reasonable only values  $\hat{K}$  with  $\hat{K} < \delta h_n$  are considered) such that  $l_1, l_2, l_3, \hat{K}$  force the successor generation to be  $\varepsilon$ -bad.

*Proof.* Without loss of generality presume  $l_1 < l_2 < l_3$  so that  $\hat{K} = K_3$ . By lemma (4.1) select  $Q$  so that for any  $Q$  integers  $n_1, n_2, \dots, n_Q$ .

$$E(|(1/Q)\#\{i: 1 \leq i \leq Q \text{ and } T^{n_i}(\omega) \in A_3\} - P(A_3)|) < \delta. \tag{5.27}$$

Now suppose we can make  $2Q$  choices for  $K_3, n_1, n_2, \dots, n_{2Q}$ , all less than  $\delta h_n$ , and each of which force the successor generation to be  $\varepsilon$ -bad. For any particular choice of  $K_3$ , let

$$\theta_{l_1, l_2, l_3, K_3} = \{i: K_3 + 1 \leq i \leq l_3 - l_2, R_{n, i+h_n+l_1-l_3} \subset A_1, \\ R_{n, i+h_n+l_2-l_3} \subset A_2 \text{ and } R_{n, i-k_3} \subset A_3\}.$$

To say that  $K_3$  makes the successor generation is  $\epsilon$ -bad is precisely the statement

$$\left| \frac{1}{l_3 - l_2 - K_3} \# \theta_{l_1, l_2, l_3, K_3} - P(A_1)P(A_2)P(A_3) \right| > \epsilon,$$

so arguing as we did in the proof of lemma (5.18) we can assume without loss of generality that for all  $i, 1 \leq i \leq Q$

$$\frac{1}{l_3 - l_2 - n_i} \# \theta_{l_1, l_2, l_3, n_i} - P(A_1)P(A_2)P(A_3) > \epsilon. \tag{5.28}$$

Let  $Y_2$  be as in (5.9), where  $\sigma$  is the identity transformation. Then:

$$(5.29) \text{ For any } i, i \in \theta_{l_1, l_2, l_3, K_3} \text{ iff } i \in Y_2, i \geq K_3 + 1 \text{ and } R_{n, i - K_3} \subset A_3.$$

Let  $\hat{\theta}_{l_1, l_2, l_3, K_3} = \theta_{l_1, l_2, l_3, K_3} \setminus \{i: 1 \leq i \leq \delta h_n\}$  and  $\hat{Y}_2 = Y_2 \setminus \{i: 1 \leq i \leq \delta h_n\}$ . Then by (5.29), if  $K_3 < \delta h_n$ ,

$$(5.30) \text{ } i \in \hat{\theta}_{l_1, l_2, l_3, K_3} \text{ iff } i \in \hat{Y}_2 \text{ and } R_{n, i - K_3} \subset A_3.$$

For each  $i \in \hat{Y}_2$ , let  $c_i = \#\{j: 1 \leq j \leq Q \text{ and } i \in \hat{\theta}_{l_1, l_2, l_3, n_j}\}$ . Then by (5.28),

$$\begin{aligned} \sum_{i \in \hat{Y}_2} c_i &= \sum_{j=1}^Q \# \hat{\theta}_{l_1, l_2, l_3, n_j} \geq \sum_{j=1}^Q (\# \theta_{l_1, l_2, l_3, n_j} - \delta h_n) \\ &> ((l_3 - l_2 - \delta h_n)(P(A_1)P(A_2)P(A_3) + \epsilon) - \delta h_n)Q; \end{aligned} \tag{5.31}$$

the last inequality follows from (5.28), noting that  $n_j < \delta h_n$  for each  $j, 1 \leq j \leq Q$ . Suppose  $i \in \hat{Y}_2$ . For  $\hat{\omega} \in R_{n, i}, 1 \leq j \leq Q, T^{-n_j}(\hat{\omega}) \in R_{n, i - n_j}$  and thus for such  $\hat{\omega}$ , by (5.30),  $T^{-n_j}(\hat{\omega}) \in A_3$  iff  $i \in \hat{\theta}_{l_1, l_2, l_3, n_j}$ . Hence, for  $\hat{\omega} \in R_{n, i}, i \in \hat{Y}_2$ ,

$$\#\{j: 1 \leq j \leq Q \text{ and } T^{-n_j}(\hat{\omega}) \in A_3\} = c_i. \tag{5.32}$$

Each rung has probability more than  $(1/h_n)(1 - \delta)$  so by (5.27) and (5.32),

$$\begin{aligned} \delta &> E(|(1/Q) \# \{j: 1 \leq j \leq Q \text{ and } T^{-n_j}(\hat{\omega}) \in A_3\} - P(A_3)|) \\ &> (1/h_n)(1 - \delta) \sum_{i \in \hat{Y}_2} |(1/Q)c_i - P(A_3)| \\ &\geq (1/h_n)(1 - \delta) \left| (1/Q) \left( \sum_{i \in \hat{Y}_2} c_i \right) - P(A_3) \# \hat{Y}_2 \right| \\ &\geq (1/h_n)(1 - \delta) \left( \left| (1/Q) \left( \sum_{i \in \hat{Y}_2} c_i \right) - l_3 - l_2 - \delta h_n \right| P(A_1)P(A_2)P(A_3) \right. \\ &\quad \left. - \delta h_n(P(A_1)P(A_2)P(A_3)) \right. \\ &\quad \left. - |(l_3 - l_2)P(A_1)P(A_2)P(A_3) - P(A_3) \# (Y_2)| - |P(A_3)(\#(Y_2) - \# \hat{Y}_2)| \right). \end{aligned} \tag{5.33}$$

By (5.31),  $(1/Q) (\sum_{i \in \hat{Y}_2} c_i) - (l_3 - l_2 - \delta h_n)P(A_1)P(A_2)P(A_3) > \epsilon(l_3 - l_2 - \delta h_n) - \delta h_n$ . Thus

$$\left| (1/Q) \left( \sum_{i \in \hat{Y}_2} c_i \right) - (l_3 - l_2 - \delta h_n)P(A_1)P(A_2)P(A_3) \right| > \epsilon(l_3 - l_2 - \delta h_n) - \delta h_n. \tag{5.34}$$

Applying (5.12),  $\sigma$  is here the identity transformation, gives

$$|(l_3 - l_2)P(A_1)P(A_2)P(A_3) - P(A_3) \# (Y_2)| < \delta(l_3 - l_2)P(A_3). \tag{5.35}$$

The definition of  $\hat{Y}_2$  gives

$$|\# Y_2 - \# \hat{Y}_2| \leq \delta h_n. \tag{5.36}$$

Plugging (5.34), (5.35) and (5.36) into the right hand side of (5.33) and using  $h_n > l_3 - l_2 > \sqrt{\delta} h_n$  gives

$$\begin{aligned} \delta &> (1/h_n)(1 - \delta)(\epsilon(l_3 - l_2 - \delta h_n) - \delta h_n - \delta h_n P(A_1)P(A_2)P(A_3) \\ &\quad - \delta(l_3 - l_2)P(A_3) - P(A_3)\delta h_n) \\ &> (1 - \delta)(\epsilon(\sqrt{\delta} - \delta) - \delta - \delta P(A_1)P(A_2)P(A_3) - \delta P(A_3) - \delta P(A_3)) \\ &> (1 - \delta)(\epsilon(\sqrt{\delta} - \delta) - 4\delta); \end{aligned}$$

thus  $(\delta/(\sqrt{\delta} - \delta))(1/(1 - \delta) + 4) > \epsilon$ , contradicting the choice of  $\delta$  in the statement of lemma (5.26). □

*Remark.* In the last paragraph of remark (5.4), we indicated that we only need twofold mixing to study  $l_{\sigma(1)}, l_{\sigma(3)}, \hat{K}$ . The next lemma makes this statement precise.

LEMMA (5.37). *Let  $\alpha, \beta$  be two distinct elements of  $\{1, 2, 3\}$  then:*

(5.38) *conditioned on  $l_\alpha \neq \text{error}$ ,  $l_\alpha$  is independent of  $K_\alpha$ ; and*

(5.39) *for any  $\delta > 0$ , if  $M, N$  are chosen large enough, then*

$$|P(l_\beta = i | l_\alpha, K_\alpha) - P(l_\beta = i)| < \delta \quad \text{for all } i.$$

(Recall that  $l_\alpha, l_\beta, K_\beta$  are functions of  $M$  and  $N$  and  $n$ . Note that you can choose  $n$  before  $\delta$ , i.e. you can let  $\delta$  depend on  $n$ .)

*Proof:* Assume without loss of generality that  $\alpha = 1$  and  $\beta = 2$ .  $K_1$  is the size of the  $n$ -error set above  $T^{M+N}(\omega)$ . This is clearly independent of what rung of the tower  $T^{M+N}(\omega)$  is in and hence (5.38) holds. (5.39) follows immediately from twofold mixing once one observes that  $l_2$  is information about  $T^M(\omega)$ , and  $l_1, K_1$  are information about  $T^{M+N}(\omega)$ . □

LEMMA (5.40). *Let  $\delta > 0$ . Then for  $n$  sufficiently large,  $P(K_i = j) < \delta$  for all  $j$  and all  $i \in \{1, 2, 3\}$ .*

*Proof.* First select  $m$  so that  $1/h_m < \delta/2$  and let  $S$  be the first rung of the  $m$ -tower. Then  $P(S) < \delta/2$ . By mixing, select  $L$  so that  $P(T^L(\omega) \in S | \omega \in S) < \delta$ . Suppose  $n > m$  is chosen large enough so that  $h_n > L$ . We will show for any such  $n$  that  $P(K_i = j) < \delta$  for all  $j$  and all  $i \in \{1, 2, 3\}$ .

Assume without loss of generality that  $i = 1$ .  $K_1$  is independent of what rung of the  $n$ -tower  $\omega$  is on and  $S$ , being a rung of the  $m$ -tower, is a union of rungs of the  $n$ -tower. Thus

$$P(K_1 = j | K_1 \text{ is defined}) = P(K_1 = j | \omega \in S).$$

If  $K_1 = j$ , then  $T^{h_n+j}(\omega)$  is on the same rung as  $\omega$  and thus if  $\omega \in S$ ,  $T^{h_n+j}(\omega) \in S$ . Hence

$$P(K_1 = j) \leq P(K_1 = j | K_1 \text{ is defined}) = P(K_1 = j | \omega \in S) \leq P(T^{h_n+j}(\omega) \in S | \omega \in S) < \delta. \quad \square$$

Theorem 1a and theorem 1b are called theorems, rather than lemmas, because all the previous lemmas are used to prove theorem 1 and once theorem 1 is proved, none of those lemmas are used again.

THEOREM 1a. *Let  $\epsilon > 0$  be given. Then there exists  $\hat{n}$  so that if  $n > \hat{n}$ , there exists  $M_n$  so that if  $M, N > M_n$ , the probability that  $\omega$  is between  $n$ -generations, or that both this  $n$ -generation and the successor  $n$ -generation are  $\epsilon$ -bad is less than  $\epsilon$ .*

*Proof.* Basically, the proof will consist of indicating certain ‘bad sets’ such that if  $\omega$  avoids these bad sets either the generation containing  $\omega$  or the successor generation is  $\varepsilon$ -good, and then proving that once  $n, M, N$  are properly chosen, the probability of the union of these bad sets is less than  $\varepsilon$ . Select  $\delta > 0$  to be small enough to obey all the conditions of lemmas (5.18) and (5.26) and so that

$$3\delta^2 + 147\delta + 18\sqrt{\delta} < \varepsilon. \tag{5.41}$$

Let  $\sigma(1), \sigma(2), \sigma(3)$  be any permutation of the numbers  $\{1, 2, 3\}$ . Let  $(l_{\sigma(1)}, l_{\sigma(2)}, K_{\sigma(3)})$  be called *double bad* if  $l_{\sigma(1)} < l_{\sigma(2)}$  and there exists a value for  $l_{\sigma(3)}$  such that  $l_{\sigma(1)} < l_{\sigma(2)} < l_{\sigma(3)}$  and  $(l_{\sigma(1)}, l_{\sigma(2)}, l_{\sigma(3)}, K_{\sigma(3)})$  is reasonable, (this means that there exists  $K_{\sigma(1)}$  such that  $(\vec{l}, \vec{K})$  is reasonable), but both the current generation and the successor generation for that four-tuple are  $\varepsilon$ -bad.

The bad sets are as follows:

- $A_1 = \{(\vec{l}, \vec{K}) \text{ is unreasonable}\};$
- $A_2 = \{\omega : (l_1, l_2, K_2) \text{ is double bad}\};$
- $A_3 = \{\omega : (l_2, l_1, K_1) \text{ is double bad}\};$
- $A_4 = \{\omega : (l_1, l_3, K_3) \text{ is double bad}\};$
- $A_5 = \{\omega : (l_3, l_1, K_1) \text{ is double bad}\};$
- $A_6 = \{\omega : (l_2, l_3, K_3) \text{ is double bad}\};$
- $A_7 = \{\omega : (l_3, l_2, K_2) \text{ is double bad}\}.$

It is immediate from the definitions of these sets that if  $\omega$  is in none of them, then  $\omega$  is in a generation and either that generation or the successor is  $\varepsilon$ -good. All that remains is to choose  $\hat{n}$ , and then for  $n > \hat{n}$  choose  $M_n$  so that if  $M, N > M_n$ , then  $M, N$  and  $n$  force all the bad sets to be small.

Choose  $Q$  so that if  $n$  is chosen large enough, the conclusion of lemmas (5.18) and (5.26) hold for that  $Q$ . Choose  $\hat{n}$  so that if  $n > \hat{n}$ ,  $n$  is large enough for the previous sentence to hold,  $P(n\text{-error set}) < \delta^2/1 + \delta^2$ , and lemma (5.40) holds where  $\delta$  is replaced by  $\delta/Q^2$ . Next choose  $M_n$  so that if  $M, N > M_n$  lemma (5.5) holds and lemma (5.37) holds with  $\delta$  replaced by  $(1 - \delta)/2h_n$ , (recall that it is acceptable for  $\delta$  to depend on  $n$  in lemma 5.37).

Now fix  $n > \hat{n}$  and  $M, N > M_n$ . By lemma (5.5),

$$P(A_1) < 3\delta^2 + 3\delta + 18\sqrt{\delta}. \tag{5.42}$$

We now estimate  $P(A_2)$ . By lemma (5.18) there are less than  $2Q$  ways to choose  $l_3$  so that  $l_1 < l_3 < l_2$  and  $(l_1, l_2, l_3)$  are  $\delta$ -reasonable and  $\varepsilon$ -bad. Once such an  $l_3$  is chosen, it follows by lemma (5.26) that there are less than  $2Q$  ways to choose  $\hat{K}$  (which in this case is  $K_2$ ) with  $(l_1, l_2, l_3, K_2)$  reasonable, and for which the successor generation is  $\varepsilon$ -bad. Thus, if  $l_1 < l_2$  there are at most  $(2Q)(2Q) = 4Q^2$  values of  $K_2$  so that  $(l_1, l_2, K_2)$  is double bad. Suppose  $l_1, l_2$  are known, say  $l_1 = a, l_2 = b$ . Let  $d_1, d_2, \dots, d_{4Q^2}$  be  $4Q^2$  values including all values of  $K_2$  where  $(l_1, l_2, K_2)$  is double bad. Recall that we chose  $\hat{n}$  so that  $n > \hat{n}$  implies lemma (5.40) holds with  $\delta$  replaced by  $\delta/Q^2$ . For such  $n$ ,

$$P(K_2 = d_i) < \delta/Q^2 \quad \text{for all } i, 1 \leq i \leq 4Q^2; \tag{5.43}$$

$$P(K_2 = d_i | l_1 = a, l_2 = b) = P(K_2 = d_i, l_1 = a, l_2 = b) / P(l_1 = a, l_2 = b)$$

$$= P(l_1 = a | K_2 = d_i, l_2 = b) P(k_2 = d_i, l_2 = b) / P(l_1 = a, l_2 = b); \tag{5.44}$$

the events  $K_2 = d_i, l_2 = b$  are independent conditional on the event  $l_2 \neq \text{error}$ .

$$P(K_2 = d_i, l_2 = b | l_2 \neq \text{error}) = P(k_2 = d_i | l_2 \neq \text{error}) P(l_2 = b | l_2 \neq \text{error}). \tag{5.45}$$

Both  $(K_2 = d_i) \subset (l_2 \neq \text{error})$  and  $(l_2 = b) \subset (l_2 \neq \text{error})$  so (5.45) becomes

$$P(K_2 = d_i, l_2 = b) / P(l_2 \neq \text{error}) = (P(K_2 = d_i) / P(l_2 \neq \text{error})) \times (P(l_2 = b) / P(l_2 \neq \text{error})).$$

$P(l_2 \neq \text{error}) > \frac{1}{2}$  so we have

$$P(K_2 = d_i, l_2 = b) < 2P(K_2 = d_i)P(l_2 = b).$$

Plugging this into the right hand side of equation (5.44) we get

$$P(K_2 = d_i | l_1 = a, l_2 = b) < 2P(l_1 = a | K_2 = d_i, l_2 = b) P(K_2 = d_i) / P(l_1 = a | l_2 = b) \tag{5.46}$$

Recall that we chose  $M_n$  so that  $M, N > M_n$  implies lemma (5.37) holds with  $\delta$  replaced by  $(1 - \delta) / 2h_n$ . Plugging this and (5.43) into (5.46) gives

$$P(k_2 = d_i | l_1 = a, l_2 = b) < 2((P(l_1 = a) + \frac{1}{2}(1 - \delta) / h_n) / (P(l_1 = a) - \frac{1}{2}(1 - \delta) / h_n)) \delta / Q^2 < 2[\frac{3}{2}P(l_1 = a) / \frac{1}{2}P(l_1 = a)] \delta / Q^2 = 6\delta / Q^2.$$

The second inequality holds because  $(1 - \delta) / h_n < (1 - P(n\text{-error set})) / h_n = P(l_1 = a)$ . For  $(l_1, l_2, K_2)$  to be double bad, where  $l_1 = a, l_2 = b$ , it is necessary that  $K_2 = d_i$  for some  $i, 1 \leq i \leq 4Q^2$ . Thus by (5.46),

$$P((l_1, l_2, K_2) \text{ double bad} | l_1 = a, l_2 = b) < (6\delta / Q^2) 4Q^2 = 24\delta.$$

Since  $a$  and  $b$  are arbitrary,  $P(A_2) < 24\delta$ . By similar reasoning,  $P(A_i) < 24\delta$  for  $2 \leq i \leq 7$ . By (5.41) and (5.42),

$$P\left(\bigcup_{i=1}^7 A_i\right) = P(A_1) + \sum_{i=2}^7 P(A_i) = 3\delta^2 + 3\delta + 18\sqrt{\delta} + (6 \times 24)\delta = 3\delta^2 + 147\delta + 18\sqrt{\delta} < \epsilon. \quad \square$$

Just as the successor generation refers to the generation following the current generation so we refer to the generation preceding the current generation as the *preceding generation*.

**THEOREM 1b.** *Let  $\epsilon > 0$  be given. Then there exists  $\hat{n}$  so that if  $n > \hat{n}$ , there exists  $M_n$  so that if  $M, N > M_n$ , the probability that  $\omega$  is between  $n$ -generations, or that both this  $n$ -generation and the preceding  $n$ -generation are  $\epsilon$ -bad is less than  $\epsilon$ .*

*Proof.* The proof is exactly analogous to the proof of theorem 1a (e.g. instead of  $K_n(\omega)$  define  $\hat{K}_n(\omega)$  to be the number of  $b$ 's between the  $n$ -block containing  $\omega_0$  and the previous  $n$ -block and then define  $\hat{K}_{n,1}\omega, \hat{K}_{n,2}\omega, \hat{K}_{n,3}\omega$  to be  $\hat{K}_n T^{M+N}(\omega), \hat{K}_n T^M(\omega), \hat{K}_n(\omega)$  respectively).  $\square$

6. Conclusion: the nesting argument

**THEOREM 2.** *Rank 1 twofold mixing implies threefold mixing.*

*Proof.* Our goal is to select  $\hat{N}$  so that for  $M, N > \hat{N}$ ,

$$|P(T^{(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3) - P(A_1)P(A_2)P(A_3)|$$

is small. We begin this proof by selecting a sequence of terms, the last of which is  $\hat{N}$ . First, fix  $\varepsilon > 0$ . Choose  $g$  to be an even integer with

$$\left(\frac{1}{1 + \varepsilon}\right)^{g/2} < \varepsilon. \tag{6.1}$$

Let

$$\delta = \left(\frac{\varepsilon}{g}\right)^2. \tag{6.2}$$

We now define an increasing sequence of integers of length  $g + 1$ ;  $n_0, n_1, n_2, \dots, n_g$ . To begin with, choose  $n_0$  bigger than the  $\hat{n}$  of theorem 1a and theorem 1b with  $\varepsilon$  replaced by  $\delta$ . By ‘ $\varepsilon$  replaced by  $\delta$ ’, we mean the following. Theorem 1a and b say that for any  $\varepsilon > 0$  something is true. In particular it’s true for the  $\delta$  of (6.2). For that  $\delta$  there exists  $\hat{n}$  and we choose  $n_0$  bigger than that  $\hat{n}$ . For  $i < g$  choose  $n_{i+1}$  sufficiently larger than  $n_i$  so that

$$\delta^2 h_{n_{i+1}} > h_{n_i}. \tag{6.3}$$

Recall that  $h_n$  is the height of an  $n$ -tower. Note that all  $n_i$  are bigger than  $n_0$  and hence are bigger than the  $\hat{n}$  of theorems 1a and b with  $\varepsilon$  replaced by  $\delta$ . We now choose  $\hat{N}$ . For  $i \neq j, i, j \in \{1, 2, 3\}$ , and any fixed  $n, l_i(n, M, N), l_j(n, M, N)$  become close to independent as  $M$  and  $N$  grow large. In particular, if  $l_i$  and  $l_j$  were completely independently valued from 1 to  $h_n$ , then  $P(|l_i - l_j| < \varepsilon h_n) < 2\varepsilon$ , because for every value  $l_i$  might take, there are at most  $2\varepsilon h_n$  values  $l_j$  can have where  $|l_i - l_j| < \varepsilon h_n$  and the probability of taking on one of  $2\varepsilon h_n$  values is  $2\varepsilon$ . Therefore, for  $\hat{N}$  chosen large enough, for any  $M, N > \hat{N}$

$$P(|l_i(n_k, M, N) - l_j(n_k, M, N)| < \varepsilon h_{n_k}) < 2\varepsilon, \tag{6.4}$$

for all  $i, j \in \{1, 2, 3\}$  and all  $k \in \{0, 1, \dots, g\}$ .

For  $i \in \{0, 1, \dots, g\}$ , let  $G_i(\omega)$  be the event that  $\omega$  is between  $n_i, M, N$ -generations or that both this  $n_i, M, N$ -generation and next one is  $\delta$ -bad or this  $n_i$ -generation and the previous one is  $\delta$ -bad then by theorem 1a and b with  $\varepsilon$  replaced by  $\delta$ , if  $\hat{N}$  is chosen large enough, then for all  $M, N > \hat{N}$

$$P(G_i(\omega)) < 2\delta \quad \text{for } i \in \{0, 1, \dots, g\}. \tag{6.5}$$

We choose  $\hat{N}$  large enough so that (6.4) and (6.5) hold. From here on in  $M, N$  are two fixed values larger than  $\hat{N}$ .

*Definition.* A finite set  $S \subset \mathbb{Z}$  will be called  $\omega$ -good if

$$\left| \frac{1}{\#S} \# \{i \in S: T^{i+M+N}(\omega) \in A_1, T^{i+M}(\omega) \in A_2, T^i(\omega) \in A_3\} - P(A_1)P(A_2)P(A_3) \right| < \delta.$$

We will just say ‘good’ if  $\omega$  is understood. Note that the disjoint union of good sets is good. Note also that a  $\delta$ -good  $n$ -generation for  $\omega$  is an  $\omega$ -good interval of integers.

Let  $L = h_{n_g}$ . We will show that  $P(T^{-(M+N)}A_1 \cap T^{-M}(A_2) \cap A_3)$  is close to  $P(A_1)P(A_2)P(A_3)$  by exhibiting, for most  $\omega$ , a large subset of  $0, 1, \dots, L$  which is



good, (by 'large' we mean most of  $0, 1, \dots, L$ ). We now construct for each  $\omega$  a sequence of three set partitions of  $\{0, 1, \dots, L\}$ , each used to construct the next, such that for most  $\omega$  one set in the last partition is the large good set desired.

$$(I_{1,g}(\omega), I_{2,g}(\omega), I_{3,g}(\omega)), (I_{1,g-1}(\omega), I_{2,g-1}(\omega), I_{3,g-1}(\omega)), \dots, (I_{1,0}(\omega), I_{2,0}(\omega), I_{3,0}(\omega)))$$

is our sequence of three set partitions of  $0, 1, \dots, L$  to be constructed in that order by induction.  $\omega$  will be omitted from expressions  $I_{i,j}(\omega)$  when understood.  $I_{1,g} = \{0, 1, \dots, L\}$ ,  $I_{2,g} = \emptyset$ ,  $I_{3,g} = \emptyset$ . Now suppose  $I_{1,i+1}$ ,  $I_{2,i+1}$ ,  $I_{3,i+1}$  have been constructed. We define  $I_{3,i}(\omega)$  to be the union of  $I_{3,i+1}(\omega)$  with all  $\delta$ -good  $n_i$ -generations for  $\omega$  contained in  $I_{1,i+1}(\omega)$ . We define  $I_{2,i}(\omega)$  to be the union of  $I_{2,i+1}(\omega)$  and all integers in  $I_{1,i+1}(\omega)$  not in an  $n_i$ -generation for  $\omega$  and all  $n_i$ -generations for  $\omega$  which are subsets of  $I_{1,i+1}(\omega)$  which are not  $\delta$ -good and which are smaller than  $\delta h_{n_i}$  in size.  $I_{1,i} = \{0, \dots, L\} \setminus (I_{2,i} \cup I_{3,i})$ . In the special case where  $i = g - 1$  we will alter this definition slightly. We define  $I_{2,g-1}$  and  $I_{3,g-1}$  just as we did above except that we also include, in  $I_{2,g-1}$ , the intersection with  $\{0, 1, 2, \dots, L\}$  of the  $n_{g-1}$ -generations (of which there are at most two) which intersect but are not contained in  $\{0, 1, \dots, L\}$ . As above, we define  $I_{1,g-1}$  to be  $\{0, \dots, L\} \setminus (I_{2,g-1} \cup I_{3,g-1})$ .

We now discuss what  $I_{1,i}$  looks like for each  $i$ .  $I_{2,g-1} \cup I_{3,g-1}$  is a union of  $n_{g-1}$ -generations with the integers between  $n_{g-1}$ -generations and the incomplete  $n_{g-1}$ -generations on the edge of  $\{0, \dots, L\}$ . That leaves a union of complete  $n_{g-1}$ -generations for  $I_{1,g-1}$ . Each of these is a union of  $n_{g-2}$ -generations with integers between  $n_{g-2}$ -generations. Continuing by induction we see that each  $I_{1,i}$  is a union of  $n_i$ -generations.

We can see by induction that each  $I_{3,i}$  is a disjoint union of  $\delta$ -good generations and hence is a good set. In particular,  $I_{3,0}$  is good. Our goal is to establish that for most  $\omega$ ,  $I_{3,0}(\omega)$  is the desired large good set mentioned earlier. Thus we must only establish that for most  $\omega$ ,  $I_{3,0}(\omega)$  is large.

We now proceed to establish the set of  $\omega$  for which  $I_{3,0}(\omega)$  is large. For each  $\omega$  and each  $i$ , let  $C_i(\omega)$  be the event that the previous  $n_i$ -generation, this  $n_i$ -generation or next  $n_i$ -generation is less than  $\epsilon h_n$  in size. In order for  $C_i(\omega)$  to occur, it is necessary that  $|l_j - l_k| < \epsilon h_{n_i}$  for some  $j, k \in \{1, 2, 3\}$  or for  $l_j < \epsilon h_{n_i}$  for some  $j \in \{1, 2, 3\}$ . For every fixed  $j$ ,  $P(l_j < \epsilon h_{n_i})$  is approximately  $\epsilon$  so  $P(l_j < \epsilon h_{n_i} \text{ for some } j) \approx 3\epsilon$ . By (6.4),  $P(|l_j - l_k| < \epsilon h_{n_i} \text{ for some } j \text{ and } k) < 6\epsilon$ . Thus  $P(C_i(\omega)) < 9\epsilon$ . Since the probability law on  $\omega$  is the same as the probability law on  $T^j(\omega)$ , for any  $j$ ,  $P(C_i(T^j(\omega))) < 9\epsilon$ . Thus, letting

$$\hat{C}_i(\omega) = \#\{j: 1 \leq j \leq L \text{ and } C_i(T^j(\omega))\}$$

we have

$$E\left((1/g) \sum_{i=0}^{g-1} (1/L) \hat{C}_i(\omega)\right) < 9\epsilon.$$

Thus

$$P\left(\left[ (1/g) \sum_{i=0}^{g-1} (1/L) \hat{C}_i(\omega) \right] > \sqrt{\epsilon}\right) < 9\sqrt{\epsilon}. \tag{6.6}$$

For  $\omega$  to be in an  $n_i$ -generation that is smaller than  $\delta h_{n_i}$ , it is necessary that  $l_1, l_2$  or  $l_3$  be less than  $\delta h_{n_i}$ . The probability of this is not much more than  $3\delta$  and is definitely less than  $4\delta$ . Let  $D_i(\omega)$  be the event that  $\omega$  is in an  $n_i$ -generation that is smaller than  $\delta h_{n_i}$ . Let  $\hat{D}_i(\omega) = \#\{j: 1 \leq j \leq L \text{ and } D_i(T^j(\omega))\}$ . Then  $E((1/g) \sum_{i=0}^{g-1} (1/L) \hat{D}_i(\omega)) < 4\delta$  and thus

$$P\left(\left[(1/g) \sum_{i=0}^{g-1} (1/L) \hat{D}_i(\omega)\right] > \sqrt{\delta}\right) < 4\sqrt{\delta}. \tag{6.7}$$

Repeating the reasoning of the past two paragraphs on  $G_i(\omega)$ , by (6.5), if we let  $\hat{G}_i(\omega) = \#\{j: 1 \leq j \leq L \text{ and } G_i(T^j(\omega))\}$  then

$$P\left((1/g) \sum_{i=0}^{g-1} (1/L) \hat{G}_i(\omega) > \sqrt{\delta}\right) < 2\sqrt{\delta} \tag{6.8}$$

$\omega$  will be called *special* if

$$\begin{aligned} (1/g) \sum_{i=0}^{g-1} (1/L) \hat{C}_i(\omega) &\leq \sqrt{\varepsilon}; \\ (1/g) \sum_{i=0}^{g-1} (1/L) \hat{D}_i(\omega) &\leq \sqrt{\delta}; \text{ and} \\ (1/g) \sum_{i=0}^{g-1} (1/L) \hat{G}_i(\omega) &\leq \sqrt{\delta}. \end{aligned}$$

It is for special  $\omega$  that we will prove  $I_{3,0}(\omega)$  is large. By (6.6), (6.7) and (6.8) we have

$$P(\omega \text{ is not special}) < 9\sqrt{\varepsilon} + 6\sqrt{\delta} < 15\sqrt{\varepsilon}. \tag{6.9}$$

From here on in, until stated otherwise,  $\omega$  is presumed to be a fixed special point.

We will prove that  $I_{3,0}$  is large by proving that  $I_{2,0}$  is small and  $I_{1,0}$  is small. First we prove that  $I_{2,0}$  is small. We will say that an integer  $j \in \{0, 1, \dots, L\}$  is *added to*  $I_{2,i}$  if it is in  $I_{2,i} \setminus I_{2,i+1}$ . Thus  $I_{2,0}$  is precisely the integers added to  $I_{2,i}$  for some  $i$ .

The number of integers added to  $I_{2,g-1}$  (because they are among the at most two  $n_{g-1}$ -generations which intersect, but are not contained in  $\{0, \dots, L\}$ ) is less than  $2h_{n_{g-1}}$  which by (6.3) is less than  $2\delta^2 h_{n_g} = 2\delta^2 L$ .

For any  $j \in \{0, 1, \dots, L\}$ , if  $j$  is added to  $I_{2,i}$  because it is in an  $n_i$ -generation of size less than  $\delta h_{n_i}$ , then we have  $D_i(T^j(\omega))$ . Thus the number of integers added to  $I_{2,i}$  for any  $i$  for this reason is no more than  $\sum_{i=0}^{g-1} \hat{D}_i(\omega) \leq \sqrt{\delta} gL = \varepsilon L$ . The inequality follows because  $\omega$  is special; the equality follows from (6.2).

For each  $j \in \{0, 1, \dots, L\}$ , if  $T^j(\omega)$  is added to  $I_{2,i}$  for some  $i$  because it is in between  $n_i$ -generations, then we have  $G_i(T^j(\omega))$ . Thus the total number of such integers is less than or equal to  $\sum_{i=0}^{g-1} \hat{G}_i(\omega) \leq \sqrt{\delta} gL = \varepsilon L$ .

Combining the results of the previous three paragraphs we get that

$$\#I_{2,0} \leq (2\delta^2 + 2\varepsilon)L < 4\varepsilon L. \tag{6.10}$$

Now we show that  $I_{1,0}$  is small. As already mentioned  $I_{1,i+1}$  consists of  $n_{i+1}$ -generations. All of these  $n_{i+1}$ -generations are more than  $\delta h_{n_{i+1}}$  in size because the definitions of  $I_{2,i+1}$  and  $I_{3,i+1}$  force all generations smaller than  $\delta h_{n_{i+1}}$  to be in one of those two sets.

Now we focus on a fixed  $n_{i+1}$ -generation in  $I_{1,i+1}$ , call it  $r$ , and see how much of it is in  $I_{i,1}$ . Let  $H_{0,r}$  be the total number of integers in  $r$ . Then

$$H_{0,r} > \delta h_{n_{i+1}}. \tag{6.11}$$

A certain amount of  $r$ , which we will call  $H_{1,r}$ , lies outside of  $n_i$ -generations. The remaining  $H_{0,r} - H_{1,r}$  integers lie in  $n_i$ -generations which we will list in order in pairs,  $E_{1,1}, E_{1,2}, E_{2,1}, E_{2,2}, E_{3,1}, E_{3,2} \dots$ . If the number of  $n_i$ -generations in  $r$  is odd we will call the last one  $E$ . Let  $H_{2,r} \neq \#E, H_{2,r} = 0$  if the number of  $n_i$ -generations in  $r$  is even.

$$H_{2,r} \leq h_{n_i} \tag{6.12}$$

Let  $H_{3,r}$  be the number of integers in pairs  $E_{i,1}, E_{i,2}$  such that both  $E_{i,1}$  and  $E_{i,2}$  are  $\delta$ -bad. Let  $H_{4,r}$  be the number of integers in pairs  $E_{i,1}, E_{i,2}$  in which either  $E_{i,1}$  or  $E_{i,2}$  is less than  $\epsilon h_{n_i}$  in size.

If we start with  $r$ , remove the integers between  $n_i$ -generations, remove  $E$  if it exists, and remove the pairs of generations in which both are  $\delta$ -bad or at least one is smaller than  $\epsilon h_{n_i}$ , you get at least  $H_{0,r} - H_{1,r} - H_{2,r} - H_{3,r} - H_{4,r}$  integers remaining. We say ‘at least’ not ‘exactly’ because  $H_{3,r}$  and  $H_{4,r}$  may count some of the same pairs of generations. These remaining integers form pairs  $E_{i,1}, E_{i,2}$  of which at least one is  $\delta$ -good and that one is at least  $\epsilon h_{n_i}$  in size. Since the other one is at most  $h_{n_i}$  in size, the fraction of integers in the pair which go into  $I_{i,3}$  is at least  $\epsilon/(1 + \epsilon)$  (i.e. the  $\delta$ -good generation goes into  $I_{i,3}$ ). Since this is true for each of these remaining pairs, the amount of  $r$  going into  $I_{3,i}$  is at least  $(\epsilon/(1 + \epsilon))(H_{0,r} - H_{1,r} - H_{2,r} - H_{3,r} - H_{4,r})$ .

Now, instead of considering  $r$  to be a fixed  $n_{i+1}$ -generation in  $I_{1,i+1}$ , consider  $r$  to be a variable  $n_{i+1}$ -generation in  $I_{1,i+1}$ . Altogether,

$$(6.13) \text{ the amount of } I_{1,i+1} \text{ in } I_{3,i} \text{ is at least } (\epsilon/(1 + \epsilon)) \sum_r (H_{0,r} - H_{1,r} - H_{2,r} - H_{3,r} - H_{4,r}) \text{ where the sum runs over all } n_{i+1}\text{-generations in } I_{1,i+1};$$

$$(6.14) \sum_r H_{0,r} = \#I_{1,i+1}.$$

Integers counted by  $H_{1,r} + H_{3,r}$  are all integers in  $r$  either between  $n_i$ -generations or integers where this generation and next one are  $\delta$ -bad, or integers where this generation and the previous one are  $\delta$ -bad. All are integers  $j$  where we have  $G_i(T^j(\omega))$ . Hence

$$\sum_r H_{1,r} + H_{3,r} \leq \hat{G}_i(\omega). \tag{6.15}$$

By (6.3), (6.11) and (6.12),  $H_{2,r} < \delta H_{0,r}$ . Hence by (6.14)

$$\sum_r H_{2,r} < \delta \cdot \#I_{1,i+1} \leq \delta L. \tag{6.16}$$

Every integer counted by  $H_{4,r}$  is an integer  $j$  where we have  $C_i(T^j(\omega))$ . Thus

$$\sum_r H_{4,r} \leq \hat{C}_i(\omega). \tag{6.17}$$

By (6.13), (6.14), (6.15), (6.16) and (6.17):

$$(6.18) \text{ The amount of } I_{1,i+1} \text{ in } I_{3,i} \text{ is at least } (\epsilon/1 + \epsilon)(\#I_{1,i+1} - \hat{G}_i(\omega) - \delta L - \hat{C}_i(\omega)).$$

Since  $\omega$  is special

$$(1/g) \sum_{i=1}^g (1/L)(\hat{G}_i(\omega) + \delta L + \hat{C}_i(\omega)) \leq \sqrt{\epsilon} + \sqrt{\delta} + \delta < 3\sqrt{\epsilon}.$$

Thus there are at least  $g/2$  (recall that  $g$  is an even integer) values of  $i$ ;  $i_1, i_2, \dots, i_{g/2}$ , such that for each  $j$ ,  $1 \leq j \leq g/2$ ,  $0 \leq i_j \leq g-1$  and

$$(1/L)(\hat{G}_{i_j} + \delta L + \hat{C}_{i_j}) < 6\sqrt{\epsilon}. \tag{6.19}$$

Choose them so that  $i_1 > i_2 > \dots > i_{g/2}$ . By (6.18) and (6.19),

$$\begin{aligned} \#I_{1,i_1} &\leq \#I_{1,i_1+1} - \frac{\epsilon}{1+\epsilon} (\#I_{1,i_1+1} - 6\sqrt{\epsilon}L) \leq L - \frac{\epsilon}{1+\epsilon} (L - 6\sqrt{\epsilon}L) \\ &= L \left( \frac{1}{1+\epsilon} + \frac{\epsilon}{1+\epsilon} 6\sqrt{\epsilon} \right). \end{aligned} \tag{6.20}$$

Since  $\#I_{1,i_2+1} \leq \#I_{1,i_1}$ , we have, by (6.18), (6.19), and (6.20),

$$\begin{aligned} \#I_{1,i_2} &\leq \#I_{1,i_2+1} - \frac{\epsilon}{1+\epsilon} (\#I_{1,i_2+1} - 6\sqrt{\epsilon}L) \\ &\leq L \left( \frac{1}{1+\epsilon} + \frac{\epsilon}{1+\epsilon} 6\sqrt{\epsilon} \right) - \frac{\epsilon}{1+\epsilon} \left( L \left( \frac{1}{1+\epsilon} + \frac{\epsilon}{1+\epsilon} 6\sqrt{\epsilon} \right) - 6\sqrt{\epsilon}L \right) \\ &= L \left( \left( \frac{1}{1+\epsilon} \right)^2 + \frac{1}{1+\epsilon} \frac{\epsilon}{1+\epsilon} 6\sqrt{\epsilon} + \frac{\epsilon}{1+\epsilon} 6\sqrt{\epsilon} \right). \end{aligned}$$

Continuing this reasoning by induction we have that

$$\begin{aligned} \#I_{1,i_{g/2}} &\leq L \left( \frac{1}{1+\epsilon}^{g/2} + \frac{\epsilon}{1+\epsilon} \left( \frac{1}{1+\epsilon}^{(g/2)-1} + \frac{1}{1+\epsilon}^{(g/2)-2} + \dots + \frac{1}{1+\epsilon} + 1 \right) 6\sqrt{\epsilon} \right) \\ &\leq L \left( \frac{1}{1+\epsilon}^{g/2} + 6\sqrt{\epsilon} \frac{\epsilon}{1+\epsilon} \sum_{i=0}^{\infty} \left( \frac{1}{1+\epsilon} \right)^i \right) \\ &= L \left( \left( \frac{1}{1+\epsilon} \right)^{g/2} + 6\sqrt{\epsilon} \right) < L(\epsilon + 6\sqrt{\epsilon}) < 7\sqrt{\epsilon}L. \end{aligned}$$

The second to last inequality follows from (6.1). Thus  $\#I_{1,0} \leq \#I_{1,i_{g/2}} < 7\sqrt{\epsilon}L$ . Combining this with (6.10) we get that

$$\#I_{3,0} = L - \#I_{1,0} - \#I_{2,0} \geq L - 7\sqrt{\epsilon}L - 4\epsilon L > L(1 - 11\sqrt{\epsilon}), \tag{6.21}$$

and we have proved that  $I_{3,0}$  is a large good set. All that remains is to show that the existence of a large good subset of  $\{0, \dots, L\}$  for most  $\omega$  implies that

$$\left| P(T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3) - P(A_1)P(A_2)P(A_3) \right|$$

is small. Since  $I_{3,0}$  is good

$$\left| \frac{1}{\#I_{3,0}} \# \{j \in I_{3,0} : T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3\} - P(A_1)P(A_2)P(A_3) \right| < \delta; \tag{6.22}$$

$$\left| \frac{1}{L - \#I_{3,0}} \# \{j : 1 \leq j \leq L \text{ and } j \notin I_{3,0} \text{ and } T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3\} - P(A_1)P(A_2)P(A_3) \right| < 1 \tag{6.23}$$

because it is the absolute value of the difference between two values in  $[0, 1]$ . By (6.21), (6.22), and (6.23),

$$\begin{aligned}
 & |(1/L) \#\{j: 1 \leq j \leq L \text{ and } T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3\} \\
 & \quad - P(A_1)P(A_2)P(A_3)| \\
 & \leq \frac{\#I_{3,0}}{L} \left| \frac{1}{\#I_{3,0}} \#\{j \in I_{3,0}: T^j(\omega) \in (T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3)\} \right. \\
 & \quad \left. - P(A_1)P(A_2)P(A_3) \right| \\
 & \quad + \frac{L - \#I_{3,0}}{L} \left| \frac{1}{L - \#I_{3,0}} \#\{j: 1 \leq j \leq L \text{ and } j \notin I_{3,0} \text{ and} \right. \\
 & \quad \left. T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3\} - P(A_1)P(A_2)P(A_3) \right| \\
 & \leq \frac{\#I_{3,0}}{L} \delta + \frac{L - \#I_{3,0}}{L} < \delta + 11\sqrt{\varepsilon} < 12\sqrt{\varepsilon}. \tag{6.24}
 \end{aligned}$$

We now remove the assumption that  $\omega$  is special. We just showed that if  $\omega$  is special (6.24) holds. For any  $j$ ,

$$\begin{aligned}
 P(T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3) &= P(T^{-j}(T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3)) \\
 &= E(1_{T^j(\omega) \in T^{-(M+N)}A_1 \cap T^{-M}(A_2) \cap A_3}).
 \end{aligned}$$

Here  $1_{T^j(\omega) \in T^{-(M+N)}A_1 \cap T^{-M}A_2 \cap A_3}$  is the indicator function of the event  $T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}A_2 \cap A_3$ . Let

$$J(\omega) = \#\{j: 1 \leq j \leq L \text{ and } T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3\}.$$

By (6.25),

$$\begin{aligned}
 P(T^{-(M+N)}A_1 \cap T^{-M}A_2 \cap A_3) &= E\left((1/L) \sum_{j=1}^L 1_{T^j(\omega) \in T^{-(M+N)}(A_1) \cap T^{-M}(A_2) \cap A_3}\right) \\
 &= E((1/L)J(\omega)).
 \end{aligned}$$

Thus by (6.24), the fact that  $0 < (1/L)J(\omega) < 1$  and  $0 < P(A_1)P(A_2)P(A_3) < 1$ , and (6.9) we get

$$\begin{aligned}
 & |P(T^{-(M+N)}A_1 \cap T^{-M}(A_2) \cap A_3) - P(A_1)P(A_2)P(A_3)| \\
 & = |E((1/L)J(\omega) - P(A_1)P(A_2)P(A_3))| \\
 & \leq E(|(1/L)J(\omega) - P(A_1)P(A_2)P(A_3)|)1_{\omega \text{ is special}} \\
 & \quad + E(|(1/L)J(\omega) - P(A_1)P(A_2)P(A_3)|)1_{\omega \text{ is not special}} \\
 & \leq 12\sqrt{\varepsilon} + P(\omega \text{ is not special}) \leq 12\sqrt{\varepsilon} + 15\sqrt{\varepsilon} = 27\sqrt{\varepsilon}. \quad \square
 \end{aligned}$$

**Appendix**

The purpose of the appendix is to resolve the seeming ambiguity in the block structure used to define  $n$ -towers in a rank one transformation. We prove here that if  $\omega$  is any non periodic word of ‘ $a$ ’s and ‘ $b$ ’s chosen as in our construction of a rank one transformation, then for every  $n$  the word uniquely determines which rung of the  $n$ -tower the word is in.

First, note that we can assume without loss of generality that each  $n$ -block starts and ends with an ‘ $a$ ’. Otherwise, simply remove the starting and ending ‘ $b$ ’s in each  $n$ -block and make them intermediate ‘ $b$ ’s in the construction of the  $n + 1$ -block. The words will remain the same and if it is now possible uniquely to resolve the question of which rung of the  $n$ -tower the word is in, do so and then when the starting and ending ‘ $b$ ’s are returned to the  $n$ -blocks we removed them from, it is easy to see that the non-ambiguity of which rung of the  $n$ -tower the word is in is preserved.

We now prove the result by contradiction. We prove that if the block structure of a word is ambiguous, then the word is periodic. Suppose that by looking at the word  $\omega$  it is possible for the word to be in either the  $i$ ’th or the  $j$ ’th rung of its  $n$ -tower,  $i < j$  (i.e. we are supposing that the block structure of  $\omega$  is ambiguous). Let  $B$  be an  $n$ -block. Then there must be two doubly infinite expansions of ‘ $B$ ’s and ‘ $b$ ’s which we will call forms 1 and 2, such that when we expand each  $B$  into its component ‘ $a$ ’s and ‘ $b$ ’s both forms reduce to the same word  $\omega$ , whose zero’th component is the  $i$ ’th term of its ‘ $B$ ’ in form 1, and the  $j$ ’th term of its ‘ $B$ ’ in form 2,  $i < j$ . Let  $B_{1,0}$  be the ‘ $B$ ’ in form 1 containing the origin, and  $B_{1,1}, B_{1,2}, \dots$ , be the ‘ $B$ ’s following  $B_{1,0}$  in form 1, listed in order, and  $B_{1,-1}, B_{1,-2}, \dots$ , be the ‘ $B$ ’s preceding  $B_{1,0}$  in form 1 in reverse order. Similarly, list the ‘ $B$ ’s of form 2,  $\dots$ ,  $B_{2,-3}, B_{2,-2}, B_{2,-1}, B_{2,0}, B_{2,1}, \dots$ , where  $B_{2,0}$  contains the origin. Let  $h_n$  be the height of the  $n$ -tower, i.e. the length of ‘ $B$ ’. Let  $r$  be the number of ‘ $b$ ’s between  $B_{1,0}$  and  $B_{1,1}$  and let  $\hat{r}$  be the number of ‘ $b$ ’s between  $B_{2,0}$  and  $B_{2,1}$ .

Since  $\omega_0$  is the  $i$ ’th entry of  $B_{1,0}$ ,  $\omega_{1-i}$  is the first entry of  $B_{1,0}$  and is thus an ‘ $a$ ’.  $\omega_{1-i}$  is the  $j + 1 - i$ ’th entry of  $B_{2,0}$  so since  $\omega_{1-i}$  is ‘ $a$ ’, the  $j + 1 - i$ ’th entry of ‘ $B$ ’ is ‘ $a$ ’.  $\omega_{1-i} + h_n + \hat{r}$  is the  $j + 1 - i$ ’th entry of  $B_{2,1}$  and is thus an ‘ $a$ ’. However,  $\omega_{h_n-i}$  is the last entry of  $B_{1,0}$  so that  $\omega_{h_n-i+1}, \omega_{h_n-i+2}, \dots, \omega_{h_n-i+r}$  are all ‘ $b$ ’s so  $1 - i + h_n + \hat{r} > h_n - i + r; 1 + \hat{r} > r;$

$$\hat{r} \geq r. \tag{6.26}$$

$\omega_{2h_n+\hat{r}-j}$  is the last entry of  $B_{2,1}$  and is thus an ‘ $a$ ’. We will now show that  $\omega_{2h_n+\hat{r}-j}$  is in  $B_{1,1}$ . Since it is an ‘ $a$ ’, it must be in  $B_{1,k}$  for some  $k$ . It cannot be in  $B_{1,k}$  for  $k \leq 0$  because  $2h_n + \hat{r} - j > h_n - i$  and  $\omega_{h_n-i}$  is the last entry of  $B_{1,0}$ . Suppose it is in  $B_{1,k}$  for some  $k \geq 2$ . Because we are assuming that  $\omega_{2h_n+\hat{r}-j}$  is in  $B_{1,k}$  for some  $k \geq 2$  and because  $\omega_{h_n-i+r+1}$  is the first entry of  $B_{1,1}$  it follows that

$$h_n + \hat{r} - j \geq h_n - i + r + 1. \tag{6.27}$$

We also have

$$h_n - i + r + 1 \geq h_n - i + 1 > h_n - j + 1. \tag{6.28}$$

$\omega_{h_n-j}$  is the last entry of  $B_{2,0}$  so  $\omega_{h_n-j+1}, \omega_{h_n-j+2}, \dots, \omega_{h_n-j+\hat{r}}$  are all ‘ $b$ ’s’. In particular, by (6.27) and (6.28)  $\omega_{h_n-i+r+1}$  is a ‘ $b$ ’. This is impossible because  $\omega_{h_n-i+r+1}$  is the first entry of  $B_{1,1}$  and is thus an ‘ $a$ ’.

Now that we’ve proved that  $\omega_{2h_n+\hat{r}-j}$  is in  $B_{1,1}$  it can readily be verified that  $\omega_{2h_n+\hat{r}-j}$  is the  $h_n + \hat{r} - r + i - j$ ’th entry of  $B_{1,1}$ . Thus, since  $\omega_{2h_n+\hat{r}-j}$  is an ‘ $a$ ’ and  $\omega_{h_n+\hat{r}-r-j}$  is the  $h_n + \hat{r} - r + i - j$ ’th entry of  $B_{1,0}$ , it follows that  $\omega_{h_n+\hat{r}-r-j}$  is an ‘ $a$ ’.  $\omega_{h_n-j}$  is the

last entry of  $B_{2,0}$  so by definition of  $\hat{r}$ ,  $\omega_{h_n-j+1}, \omega_{h_n-j+2}, \dots, \omega_{h_n-j+\hat{r}}$  are all 'b's'. It follows from the last two sentences that  $h_n + \hat{r} - r - j < h_n - j + 1$ ;  $\hat{r} < r + 1$ ;

$$\hat{r} \leq r. \tag{6.29}$$

From (6.26) and (6.29) we have that  $r = \hat{r}$ . We have shown that the distance between  $B_{2,0}$  and  $B_{2,1}$  is the same as the distance from  $B_{1,0}$  and  $B_{1,1}$  namely  $r$ . By translating  $\omega$ , we could use the same argument to prove that the distance between  $B_{1,0}$  and  $B_{1,1}$  is the same as the distance between  $B_{2,1}$  and  $B_{2,2}$ , namely  $r$ . Continue the argument proving successively that  $r$  is the distance between  $B_{1,1}$  and  $B_{1,2}$ ,  $B_{2,2}$  and  $B_{2,3}$ ,  $B_{1,2}$  and  $B_{1,3}$ ,  $B_{2,3}$  and  $B_{2,4}$  etc. Going backwards, we can prove successively that  $r$  is the distance between  $B_{1,-1}$  and  $B_{1,0}$ ,  $B_{2,-1}$  and  $B_{2,0}$ ,  $B_{1,-2}$  and  $B_{1,-1}$  etc. Therefore  $\omega$  is periodic. □

REFERENCES

[1] P. R. Halmos. *Lectures on Ergodic Theory*. Chelsea Publishing Co: New York, (1956).  
 [2] V. A. Rohlin. On endomorphisms of compact commutative groups. *Izv. Acad. Nauk. SSSR*, 13 (1949), 329-340. (In Russian.)