

THE ERGODIC THEORETICAL PROOF OF SZEMERÉDI'S THEOREM

BY H. FURSTENBERG, Y. KATZNELSON AND D. ORNSTEIN¹

Introduction. In 1975, E. Szemerédi proved the following theorem conjectured some forty years earlier by Erdős and Turán:

THEOREM I. *Let $\Lambda \subset Z$ be a subset of the integers of positive upper density, then Λ contains arbitrarily long arithmetic progressions.*

Partial results were obtained previously by K. F. Roth (1952) who established the existence of arithmetic progressions of length three in subsets of Z of positive upper density, and by E. Szemerédi (1969) who proved the existence of progressions of length four.

In 1976 Furstenberg noticed that the statement of Theorem I is equivalent to a statement about "multiple recurrence" of measure-preserving transformations, namely

THEOREM II. *Let (X, \mathfrak{B}, μ) be a probability measure space, let T be an invertible, measure-preserving transformation on (X, \mathfrak{B}, μ) , and let $A \in \mathfrak{B}$ be a set of positive measure. Then for any positive integer k , there exists a subset $B \subset A$ with $\mu(B) > 0$ and an integer $n \geq 1$ with*

$$T^n B \subset A, T^{2n} B \subset A, \dots, T^{(k-1)n} B \subset A$$

or what amounts to the same,

$$\mu \left(\bigcap_{j=0}^{k-1} T^{-jn} A \right) > 0.$$

It turned out to be possible to give an ergodic theoretic proof of Theorem II, thereby providing a new proof of Szemerédi's theorem.

Various elements of Furstenberg's original proof were simplified by Katznelson and Ornstein, and making use of this it became possible to prove a generalization of Theorem II with T, T^2, \dots, T^k replaced by any commuting set of measure-preserving transformations (cf. [FK]). This result leads to an analogue of Szemerédi's theorem for Z' and, in fact, this proof proceeding by way of ergodic theory is the only one available so far for this analogue.

Received by the editors March 11, 1982.

1980 *Mathematics Subject Classification.* Primary 05, 10, 28, 60.

¹Partial support of the second and third authors was given by National Science Foundation grant MCS81-07092.

Our purpose here is to give an exposition, as widely accessible as possible, of the ergodic theoretic proof of Theorem I. For a detailed account of the interrelation between dynamics and combinatorial number theory the reader is referred to [F].

Theorem 2 is valid for all measure-preserving systems but not for the same reason. There are two distinct phenomena, mutually exclusive, which account for the existence of positive measure intersections of the form $\bigcap_{j=0}^{k-1} T^{-j}A$. One, compactness, is seen in the case of group rotations (T being a translation by a group element on the Haar measure space of a compact group) where for appropriate values of n , T^n is "close" to the identity so that $T^{-j}A$ differs from A by very little, for $0 < j < k$, and $\mu(\bigcap_{j=0}^{k-1} T^{-j}A)$ is very close to $\mu(A)$. The other phenomenon is that of weak mixing, defined by the condition that for every set A , $\mu(A \cap T^{-n}A) \sim \mu(A)^2$ for most n . It can then be proved (cf. §3) that $\mu(\bigcap_{j=0}^{k-1} T^{-j}A) \sim \mu(A)^k$ for most values of n .

It is not true that these two phenomena are complementary; there exist systems $(X, \mathfrak{B}, \mu, T)$ which are neither weakly mixing nor group rotations. However, if $(X, \mathfrak{B}, \mu, T)$ is not weakly mixing then there exists a nontrivial T -invariant sigma-algebra $\mathfrak{B}_1 \subset \mathfrak{B}$ such that T restricted to \mathfrak{B}_1 acts like a group rotation, so that in any case the assertion of Theorem II is valid for all the sets A in some T -invariant sigma algebra of \mathfrak{B} . The strategy of the proof of Theorem II is (a) to show that there exists a T -invariant subsigma-algebra $\mathfrak{B}_1 \subset \mathfrak{B}$ which is maximal, with respect to inclusion, in the class of T -invariant subalgebras of \mathfrak{B} for which the statement of Theorem II is valid. (b) Assuming $\mathfrak{B}_1 \neq \mathfrak{B}$, study the behavior of sets $A \subset \mathfrak{B}$ under T "relative to \mathfrak{B}_1 " and show that either we have "relative weak mixing" or else there exist $\mathfrak{B}_2 \supset \mathfrak{B}_1$ for which the action of T is "relatively compact". In either case we show that there exists a bigger subalgebra for which the statement of Theorem II is valid, contradicting the maximality of \mathfrak{B}_1 . This implies $\mathfrak{B}_1 = \mathfrak{B}$ and completes the proof.

In §1 we shall show that Theorems I and II are equivalent. In §2 we verify Theorem II in two special cases. The next three sections give a limited version of Theorem II, but the ideas used there form a basis for the general arguments used subsequently. The formal proof of Theorem II begins in §6 in which the notion of a factor system is described, and is carried out in a series of four steps which take up the last four sections.

1. We inherit the translation from Z as the "shift" homeomorphism $T\{\omega_n\} = \{\omega_{n+1}\}$ and with it the possibility to check whether an arithmetic progression, say $\{a + jb\}_{j=0}^{k-1}$, is contained in a set A . Formally, if we denote by A_0 the subset of 2^Z defined by the condition $\omega_0 = 1$, and by $\bar{\omega}$ the indicator function of Λ , then

$$(1.1) \quad \{a + jb\}_{j=0}^{k-1} \subset \Lambda \Leftrightarrow \bar{\omega} \in \bigcap_{j=0}^{k-1} T^{-(a+jb)}A_0.$$

It is not enough at this point to just check that the intersection in (1.1) is not empty; it certainly is not empty, containing all the elements of 2^Z which contain $\{a + jb\}_{j=0}^{k-1}$. Most of these points are irrelevant, the only relevant ones

being $\bar{\omega}$ itself and any ω which agrees with $\bar{\omega}$ on a set containing our given arithmetic progression. If we only want to know that Λ contains an arithmetic progression of length k and step b it is enough to know that $\bigcap_{j=0}^{k-1} T^{-jb}A_0$ contains some translate of $\bar{\omega}$. We remove the irrelevant points from 2^Z and consider the subspace $X =$ the closure in 2^Z of $\{T^{j\bar{\omega}}\}_{j=-\infty}^{\infty}$. X is clearly invariant under T and, writing $A = A_0 \cap X$, we see that Λ contains arithmetic progressions of length k if, and only if, for some $b \neq 0$, $\bigcap_{j=0}^{k-1} T^{-j}bA$ is nonempty (since, being open, it must contain translates of $\bar{\omega}$, the set of which is dense in X).

If we assume now that Λ has positive upper density we can construct a probability measure μ on X which is T -invariant and such that $\mu(A) > 0$. Once we have this, Theorem II applies and gives us nonempty intersections of the form $\bigcap_{j=0}^{k-1} T^{-j}bA$ and Theorem I follows.

Let $\{[a_n, b_n]\}$ be a sequence of intervals such that $b_n - a_n \rightarrow \infty$ and

$$\lim_{n \rightarrow \infty} |\Lambda \cap [a_n, b_n]| / (b_n - a_n) \rightarrow d > 0.$$

Put $\mu_n = (b_n - a_n)^{-1} \sum_{x=a_n}^{b_n-1} \delta_{T^x \bar{\omega}}$, where δ_x denotes the unit point mass at x . μ_n is a probability measure on X which, for $n \rightarrow \infty$, becomes more and more T -invariant. Specifically,

$$T\mu_n - \mu_n = (b_n - a_n)^{-1} (\delta_{T^{b_n} \bar{\omega}} - \delta_{T^{a_n} \bar{\omega}})$$

and its total mass is bounded by $2(b_n - a_n)^{-1}$. If μ is any ω^* -limit point of μ_n then μ is clearly T -invariant and, as $\mu_n(A) = |\Lambda \cap [a_n, b_n]| / (b_n - a_n)$, we obtain $\mu(A) = d > 0$. We have now proved that Theorem II implies Theorem I.

To prove that Theorem I implies Theorem II we first deduce a finite version of Theorem I.

THEOREM I_F. For every $\varepsilon > 0$ and positive integer k , there exists $N = N(\varepsilon, k)$ such that if Λ is a set of integers contained in some interval $[a, b]$ such that $b - a > N$ and $|\Lambda| \geq \varepsilon(b - a)$, then Λ contains an arithmetic progression of length k .

The implication $I_F \Rightarrow I$ is obvious and we claim that we also have $I \Rightarrow I_F$. In fact if I_F were false we would have an $\varepsilon > 0$ and a positive integer k such that for every N there exists a sequence Λ_N carried by an interval $[a_N, b_N]$, $b_N - a_N > N$, $|\Lambda_N| > \varepsilon(b_N - a_N)$ and Λ_N does not contain an arithmetic progression of length k . The properties listed for Λ_N are translation invariant so we may assume that the Λ_N 's are well separated, say $a_{N+1} > b_N + (b_{N+1} - a_{N+1})$. Writing $\Lambda = \bigcup \Lambda_N$ we see that Λ has positive upper density ($\geq \varepsilon$) and contains no arithmetic progression of length k since, because of the separation, such progression will have to be contained in one of the Λ_N 's. This would contradict Theorem I.

Theorem I_F makes it clear that the existence of arbitrarily long arithmetic progressions is insured also for a class of sequences of density zero; diminishing density can be compensated by the size of the intervals on which the density is checked. For precise statements in this direction we need an estimate

of $N(\varepsilon, k)$ of Theorem I_F. Szemerédi's proof consists in giving bounds for $N(\varepsilon, k)$. The "ergodic proof", at least so far, gives no estimates of $N(\varepsilon, k)$. We now deduce the following consequence of I_F.

THEOREM III. *Let $\varepsilon > 0$ and k be given and write $N_1 = N(\varepsilon/2, k)$ (of I_F). Let (X, \mathfrak{B}, μ) be a probability measure space and $B_l \in \mathfrak{B}$, $\mu(B_l) \geq \varepsilon$ for $l = 1, 2, \dots, N_1$. Then there exists an arithmetic progression of length k in $\{1, \dots, N_1\}$, say $\{a + mb\}_{m=0}^{k-1}$, such that*

$$(1.2) \quad \mu \left(\bigcap_{m=0}^{k-1} B_{a+mb} \right) > \frac{\varepsilon}{2} N_1^{-2}.$$

PROOF. For $x \in X$ write $\Lambda(x) = \{l; 1 \leq l \leq N_1, x \in B_l\}$. We have

$$(1.3) \quad \int |\Lambda(x)| d\mu = \sum_{l=1}^{N_1} \mu(B_l) \geq N_1 \varepsilon$$

and consequently

$$(1.4) \quad \mu(\{x; |\Lambda(x)| \geq \varepsilon N_1/2\}) \geq \varepsilon/2.$$

By I_F and the choice of N_1 , for each point x of the set appearing in (1.4), $\Lambda(x)$ contains an arithmetic progression of length k , say $\{a(x) + mb(x)\}_{m=0}^{k-1}$. There are fewer than N_1 choices for either $a(x)$ or $b(x)$ which implies, in view of (1.4), that for some pair (a, b) we have (1.2).

The proof of Theorem II, assuming we know I (and hence III), is done by writing $B_l = T^{-l}A$ and noticing that

$$\mu \left(\bigcap_{m=0}^{k-1} T^{-(a+bm)}A \right) = \mu \left(T^{-a} \bigcap_{m=0}^{k-1} T^{-bm}A \right).$$

With a and b chosen so that the measure in question is positive and with $B = \bigcap_{m=0}^{k-1} T^{-bm}A$, we obtain the conclusion of Theorem II.

2. Two special cases. In this section we discuss two examples of measure-preserving systems for which the assertion of Theorem II is readily established. By appropriately generalizing these two examples we shall obtain a strategy for proving Theorem II "step by step".

Our first example is that of a *Bernoulli system*. A Bernoulli system is the dynamical system that corresponds to a stochastic process of infinitely many independent, identically distributed, "Bernoulli" trials. To be precise, a Bernoulli system consists of a space Ω which is the space of all sequences $\{\omega_n\}_{n \in \mathbb{Z}}$ with values in a finite set, say $\Gamma = \{1, 2, \dots, r\}$. A σ -algebra of sets \mathfrak{B} is obtained in Ω by letting \mathfrak{B} be the smallest σ -algebra for which each $\omega \rightarrow \omega_n$ is measurable. The probability measure μ on \mathfrak{B} is the product measure

$$\mu\{\omega_i = j_1, \omega_{i_2} = j_2, \dots, \omega_{i_n} = j_n\} = p_{j_1} p_{j_2} \cdots p_{j_n}$$

where p_1, p_2, \dots, p_r is a probability distribution on Γ : $p_i \geq 0$, $\sum_{i=1}^r p_i = 1$. Finally the measure-preserving transformation T in this system is the shift $T\{\omega_n\} = \{\omega_{n+1}\}$.

In the case of a Bernoulli system, Theorem II follows from the following stronger assertion.

PROPOSITION 2.1. *If $(X, \mathfrak{B}, \mu, T)$ is a Bernoulli system and A_0, A_1, \dots, A_k are $k + 1$ sets in \mathfrak{B} , then as $n \rightarrow \infty$*

$$(2.1) \quad \mu(A_0 \cap T^{-n}A_1 \cap T^{-2n}A_2 \cap \cdots \cap T^{-kn}A_k) \rightarrow \mu(A_0)\mu(A_1) \cdots \mu(A_k).$$

Thus if $A \in \mathfrak{B}$, $\mu(A) > 0$, we will have for n sufficiently large,

$$\mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0$$

which, of course, implies Theorem II.

The proof of Proposition 2.1 is seen easily once it is noticed that the assertion of the proposition will follow for all $k + 1$ -tuples of sets if it is known for A_i in a "dense" subfamily of \mathfrak{B} . Elementary considerations show that if \mathfrak{B}_0 is an algebra of sets spanning \mathfrak{B} as a σ -algebra, then \mathfrak{B}_0 is dense in \mathfrak{B} in the sense that for $A \in \mathfrak{B}$ and $\varepsilon > 0$ there exists $A' \in \mathfrak{B}_0$ with $\mu(A \Delta A') < \varepsilon$, where $A \Delta A'$ denotes the symmetric difference $(A \setminus A') \cup (A' \setminus A)$.

In our case we may take \mathfrak{B}_0 to be the algebra of cylinder sets, i.e., sets in Ω defined by conditions on finitely many coordinates: $A = \{\omega; (\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_m}) \in \bar{A} \subset \Delta^m\}$. Now if A_0, A_1, \dots, A_k are cylinder sets we have, in fact,

$$\mu(A_0 \cap T^{-n}A_1 \cap \cdots \cap T^{-kn}A_k) = \mu(A_0)\mu(A_1) \cdots \mu(A_k)$$

as soon as n is so large that the sets of defining coordinates for $T^{-n}A_l$, $l = 0, \dots, k$, are disjoint. This establishes the proposition.

Our second example is of a rather different nature. Perhaps the most trivial case of Theorem II occurs for T periodic, i.e., $T^p = \text{identity}$ for some p . Not quite so trivial is the case of T "almost periodic", a phenomenon that is exemplified by $T = \text{irrational rotation of the circle}$. To be precise, let X be the circle which we represent as the reals modulo the integers, $X = R/Z$, \mathfrak{B} the σ -algebra of Borel subsets of X , μ Lebesgue measure and $T: X \rightarrow X$ defined by $Tx = x + \alpha$ for any fixed α . We now have

PROPOSITION 2.2. *With $(X, \mathfrak{B}, \mu, T)$ as above, $A \in \mathfrak{B}$ with $\mu(A) > 0$, we have for each $k = 1, 2, \dots$,*

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N \mu(A \cap T^{-l}A \cap \cdots \cap T^{-kn}A) > 0.$$

PROOF. From the fact that $\int_A 1_A(x+y) d\mu(x)$ is a continuous function of y we see that for any $\varepsilon > 0$ there exists $\delta > 0$ so that if $|y| < \delta$, $\mu(A \cap (A - y)) > \mu(A) - \varepsilon$. From this it follows that

$$\mu(A \cap (A - y) \cap (A - 2y) \cap \cdots \cap (A - ky)) > \mu(A) - (k + 1)\varepsilon.$$

Choose $\varepsilon < (k + 1)^{-1}\mu(A)$ and taking the corresponding δ , set $S_\delta = \{n; n\alpha \in (-\delta, \delta) \pmod{1}, n \geq 1\}$. We see that if $n \in S_\delta$,

$$\mu(A \cap T^{-n}A \cap T^{-2n}A \cap \cdots \cap T^{-kn}A) > \mu(A) - (k + 1)\varepsilon > 0.$$

The proposition now follows from the fact that for any α and $\delta > 0$, the set S has positive density. This latter fact is trivial if α is rational, and for α irrational it is also easily deduced since for large l , $\{0, \alpha, 2\alpha, \dots, l\alpha\}$ is 2δ -dense in R/Z and so the gap between successive numbers in S_δ cannot exceed $l + 1$.

Let us notice that in both of these cases we establish a sharper version of Theorem II. It is this version which we shall in fact obtain in general. Namely, we shall prove

THEOREM IV. For any measure-preserving system $(X, \mathfrak{B}, \mu, T)$ and $A \in \mathfrak{B}$ with $\mu(A) > 0$, and for any $k = 1, 2, \dots$

$$(2.2) \quad \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > 0.$$

REMARK 1. We do not know if the limit in (2.2) exists in general. In the two examples under discussion the \liminf may be replaced by \lim .

REMARK 2. In the Bernoulli case the intersection $A_0 \cap T^{-n}A_1 \cap \dots \cap T^{-kn}A_k$ is eventually nonempty for any sets A_0, A_1, \dots, A_k with positive measure. In the "almost periodic" case we may have $A_0 \cap T^{-n}A_1 \cap T^{-2n}A_2 = \emptyset$ for all n . The assertion (2.2) holds only for multiple intersections involving the same set.

In the next two sections we shall extend the phenomenon encountered in the examples of this section to as wide a class of systems as possible. The Bernoulli example will appear as a special case of *weak mixing systems* and the almost periodic example a special case of *compact systems*. These two notions play a central role in the sequel. The manner in which (2.2) will be established in these two classes of cases will be quite different. For weak mixing systems (2.2) will be a consequence of "mixing" which takes place for any $k + 1$ -tuple of sets; for any $A_0, A_1, \dots, A_k \in \mathfrak{B}$, the measure $\mu(A_0 \cap T^{-n}A_1 \cap \dots \cap T^{-kn}A_k)$ is close to $\mu(A_0)\mu(A_1) \dots \mu(A_k)$ for most n , $n \rightarrow \infty$. In the case of compact systems, the translate $T^{-n}A$ of a set $A \in \mathfrak{B}$ returns sufficiently closely to A , that the iterated translates $T^{-2n}A, \dots, T^{-kn}A$ all overlap, and this for a set of n of positive density.

3. Weak mixing systems. A system $(X, \mathfrak{B}, \mu, T)$ is *mixing* if for any two sets $A, B \in \mathfrak{B}$, $\lim_{n \rightarrow \infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$, so that asymptotically the portion of A that arrives in B after n steps is proportional to the measure of B . The system is *weak mixing* if, instead, one only has

$$(3.1) \quad \lim_{n \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N (\mu(A \cap T^{-n}B) - \mu(A)\mu(B))^2 = 0.$$

A weak mixing system is clearly ergodic, since for any two sets A, B of positive measure some intersection $A \cap T^{-n}B$ is nonempty, and so an invariant set must have either measure 0 or its complement will have measure 0. (3.1) is a special case of

$$(3.2) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \left(\int f T^n g d\mu - \int f d\mu \int g d\mu \right)^2 = 0$$

where $f, g \in L^2(X, \mathfrak{B}, \mu)$ and $T^n g$ is, by definition, the function $g(T^n x)$, (3.1) is obtained from (3.2) by restricting the latter to functions taking on the values 0, 1 (indicator functions). On the other hand, (3.2) follows for all functions in $L^2(X, \mathfrak{B}, \mu)$ once it is known for functions spanning linearly a dense subset; hence (3.1) \Rightarrow (3.2).

We also remark that if $(X, \mathfrak{B}, \mu, T)$ is weak mixing, so is the product system $(X \times X, \mathfrak{B} \times \mathfrak{B}, \mu \times \mu, T \times T)$. For, once again, it suffices to check (3.2) for a spanning system of functions in $L^2(X \times X)$. Such a spanning system consists of functions $f \otimes g$ where $f \in \mathfrak{C}(x_1, x_2) = f(x_1)g(x_2)$. Now (3.2) is equivalent to the assertion that for any $\varepsilon > 0$

$$\left| \int f T^n g d\mu - \int f d\mu \int g d\mu \right| < \varepsilon$$

but for a set of n of zero density. The same assertion now follows for the tensor products, since

$$\begin{aligned} \int f_1 \otimes f_2 (T \times T)^n g_1 \otimes g_2 d(\mu \times \mu) &= \int f_1 T^n g_1 d\mu \int f_2 T^n g_2 d\mu, \\ \int f_1 \otimes f_2 d(\mu \times \mu) &= \int f_1 d\mu \int f_2 d\mu, \\ \int g_1 \otimes g_2 d(\mu \times \mu) &= \int g_1 d\mu \int g_2 d\mu. \end{aligned}$$

The proof of Theorem IV in the case of weak mixing systems will follow from the next theorem which, in effect, states that a weak mixing system is "weak mixing of all orders".

THEOREM 3.1. If $(X, \mathfrak{B}, \mu, T)$ is a weak mixing system and A_0, A_1, \dots, A_k are sets in \mathfrak{B} , then

$$(3.3) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N (\mu(A_0 \cap T^{-n}A_1 \cap T^{-2n}A_2 \cap \dots \cap T^{-kn}A_k) - \mu(A_0)\mu(A_1) \dots \mu(A_k))^2 = 0.$$

The property described in this theorem is one of several possible analogues of the notion of (strong) mixing of all orders. If $(X, \mathfrak{B}, \mu, T)$ is mixing of all orders then for all $A_0, A_1, \dots, A_n \in \mathfrak{B}$,

$$\lim_{\substack{n_j \rightarrow \infty \\ |n_j - n_l| \rightarrow \infty}} \mu(A_0 \cap T^{-n_1}A_1 \cap T^{-n_2}A_2 \cap \dots \cap T^{-n_k}A_k) = \mu(A_0)\mu(A_1) \dots \mu(A_k).$$

It is still unknown if mixing implies mixing of all orders, or even if mixing implies

$$\lim_{n \rightarrow \infty} \mu(A_0 \cap T^{-n}A_1 \cap T^{-2n}A_2) = \mu(A_0)\mu(A_1)\mu(A_2).$$

Before giving the proof we will briefly sketch the strategy. Let $k = 2$. $T^n f_1 \cdot T^{2n} f_2$ will have the right product with f_0 for most n [(3.4)] if $(1/N) \sum_{n=1}^N T^n f_1 \cdot T^{2n} f_2$ is close to a constant in the L_2 norm [(3.5)]. To check

the above we need to show that $\int T^i f_1 T^{2i} f_2 T^j f_1 T^{2j} f_2$ is close on the average to $\int T^i f_1 \cdot \int T^{2i} f_2 \cdot \int T^j f_1 \cdot \int T^{2j} f_2$. Write those products where $j - i$ is fixed as

$$\int T^i (f_1 \cdot T^{(j-i)} f_1) \cdot T^{2i} (f_2 \cdot T^{2(j-i)} f_2) = \int (f_1 \cdot T^{j-i} f_1) \cdot T^i (f_2 T^{2(j-i)} f_2).$$

Weak mixing says that if i varies through enough values (depending on $j - i$) then the average value of the above integral will be $\int f_1 T^{(j-i)} f_1 \cdot \int f_2 T^{2(j-i)} f_2$ and for most $j - i$ this will equal $(\int f_1)^2 \cdot (\int f_2)^2$. In order to make the above argument work we need to have i vary through a large number of values compared to $j - i$. Therefore we check the average of the L_2 norms of $(1/H) \sum_{n=0}^{j+H} T^n f_1 \cdot T^{2n} f_2$ (3.9).

PROOF. To prove Theorem 3.1 we prove the following two variants of (3.3) by induction on k . Here $f_0, f_1, \dots, f_k \in L^\infty(X, \mathfrak{B}, \mu)$.

$$(3.4)_k \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \left[\int_{I=0}^k T^{In} f_1 d\mu - \prod_{I=0}^k \int f_1 d\mu \right]^2 = 0,$$

$$(3.5)_k \quad \lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N T^{In} f_1 - \prod_{I=1}^k \int f_1 d\mu \right\|_{L^2(X)} = 0.$$

Clearly (3.3) is a special case of (3.4) $_k$. (3.5) $_k$ refers to convergence in $L^2(X, \mathfrak{B}, \mu)$, and constitutes a generalized mean ergodic theorem for weak mixing systems.

Recall that if $(X, \mathfrak{B}, \mu, T)$ is weak mixing so is the product system (which we abbreviate $X \times X$). The strong convergence in (3.5) $_k$ implies weak convergence, so that (3.5) $_k$ implies

$$(3.6) \quad \frac{1}{N} \sum_{n=1}^N \int_{I=0}^k \int_{I=1}^k T^{In} f_1 d\mu \rightarrow \prod_{I=0}^k \int f_1 d\mu.$$

Since $X \times X$ is also weak mixing, we obtain the analogue of (3.6) replacing f_1 by $f_1 \otimes f_1$ and T by $T \times T$. The integrals on $X \times X$ become products of integrals on X and we obtain

$$(3.7) \quad \frac{1}{N} \sum_{n=1}^N \left[\int_{I=0}^k \int_{I=1}^k T^{In} f_1 d\mu \right]^2 \rightarrow \prod_{I=0}^k \left[\int f_1 d\mu \right]^2.$$

Now it is an elementary exercise to see that if $(1/N) \sum_{n=1}^N a_n \rightarrow \alpha$ and $(1/N) \sum_{n=1}^N a_n^2 \rightarrow \alpha^2$, then $(1/N) \sum_{n=1}^N (a_n - \alpha)^2 \rightarrow 0$. Thus (3.6) and (3.7) give us (3.4) $_k$, i.e., (3.5) $_k \Rightarrow$ (3.4) $_k$ (more precisely, the validity of (3.5) $_k$ for all weak mixing systems implies the same for (3.4) $_k$). Now (3.4) $_1$ is simply (3.2) and so our theorem will be proved if we show that (3.4) $_{k-1} \Rightarrow$ (3.5) $_k$.

We claim that to prove (3.5) $_k$ in general it suffices to consider the case where some $\int f_0 d\mu = 0$ so that (3.5) $_k$ can be replaced by

$$(3.5)'_k \quad \frac{1}{N} \sum_{n=1}^N \prod_{I=1}^k T^{In} f_1 \rightarrow 0$$

in $L^2(X, \mathfrak{B}, \mu)$. The reason for this is the identity

$$(3.8) \quad \prod_{I=1}^k a_I - \prod_{I=1}^k b_I = \sum_{j=1}^k \left(\prod_{I=1}^{j-1} a_I \right) (a_j - b_j) \left(\prod_{I=j+1}^k b_I \right);$$

alternatively: the reason for this is that the identity allows us to replace $\prod_{I=1}^k T^{In} f_1 - \prod_{I=1}^k \int f_1 d\mu$ by a sum of products satisfying the above condition. Now set

$$\psi_N = \frac{1}{N} \sum_{n=1}^N \prod_{I=1}^k T^{In} f_1$$

and let $N \rightarrow \infty$. We fix a large number H to be determined presently and rewrite

$$(3.9) \quad \psi_N = \frac{1}{N} \sum_{j=1}^{j+H-1} \left[\frac{1}{N} \sum_{n=j}^{j+H-1} \prod_{I=1}^k T^{In} f_1 \right] + \psi'_N = \psi''_N + \psi'_N.$$

Since the f_i are bounded, $\psi''_N \rightarrow 0$ uniformly as $O(H/N)$ as $N \rightarrow \infty$, and it will suffice to show that for appropriate H , $\limsup_{N \rightarrow \infty} \|\psi'_N\|_{L^2(X)} < \epsilon$, where ψ'_N denotes the first summand in (3.9). By convexity,

$$\psi''_N \leq \frac{1}{N} \sum_{j=1}^{j+H-1} \left[\frac{1}{H} \sum_{n=j}^{j+H-1} \prod_{I=1}^k T^{In} f_1 \right]^2,$$

and so

$$\begin{aligned} \|\psi'_N\|_{L^2(X)}^2 &\leq \frac{1}{N} \sum_{j=1}^{j+H-1} \frac{1}{H^2} \sum_{n,m=j}^{j+H-1} \int \prod_{I=1}^k T^{In} f_1 T^{Im} f_1 d\mu \\ &= \frac{1}{NH^2} \sum_{j=1}^{j+H-1} \sum_{n,m=j}^{j+H-1} \int \prod_{I=1}^k T^{In} (f_1 T^{(m-n)} f_1) d\mu. \end{aligned}$$

Since T , and therefore T^n , is measure preserving, we can replace each $T^{In}()$ by $T^{(j-1)n}()$ in the foregoing expression. Thus

$$(3.10) \quad \|\psi'_N\|_{L^2(X)}^2 \leq \frac{1}{NH^2} \sum_{j=1}^{j+H-1} \sum_{n,m=j}^{j+H-1} \int \prod_{I=0}^{I=k-1} T^{In} (f_{I+1} T^{(j+1)(m-n)} f_{I+1}) d\mu.$$

But now notice that the integrals occurring are those that occur in (3.4) $_{k-1}$ with the functions f_I replaced by $g_{I,m-n} = f_{I+1} T^{(j+1)(m-n)} f_{I+1}$. If we rewrite (3.10) as

$$(3.11) \quad \|\psi'_N\|_{L^2(X)}^2 \leq \frac{1}{H} \sum_{r=1-H}^{H-1} \left(1 - \frac{|r|}{H} \right) \left(\frac{1}{N} \sum_{n=1}^N \int \prod_{I=1}^{I=k-1} T^{In} g_{I,r} d\mu \right)$$

we can use (3.4) $_{k-1}$ to replace the right-hand side by

$$\frac{1}{H} \sum_{r=1-H}^{H-1} \left(1 - \frac{|r|}{H} \right) \left(\prod_{I=0}^{I=k-1} \int g_{I,r} d\mu \right) + \frac{\epsilon}{2}$$

for N large. On the other hand, some $\int f_0 d\mu = 0$ and by (3.2) most of the terms $\int g_{0-1,r} d\mu$ are small for H large, since

$$\int g_{0-1,r} d\mu = \int f_0 T^{10r} f_0 d\mu$$

which is, on the average, close to $(\int f_0 d\mu)^2 = 0$. Since everything in sight is bounded, we can choose H so large that

$$\left| \frac{1}{H} \sum_{r=1-H}^{H-1} \left(1 - \frac{|r|}{H} \right) \left(\prod_{l=0}^{k-1} \int g_{l,r} d\mu \right) \right| < \frac{\epsilon}{2}$$

and so

$$\limsup_{N \rightarrow \infty} \|\psi_N\|_{L^2(X)}^2 < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This completes the induction and proves Theorem 3.1.

4. Compact syst vs. We now abstract from the example of the irrational rotation of the circle the property that enables one to obtain a simple verification of

$$(2.2 \text{ bis}) \quad \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \mu(A \cap T^{-i}A \cap \dots \cap T^{-kn}A) > 0$$

for $\mu(A) > 0$.

DEFINITION 4.1. The system $(X, \mathfrak{B}, \mu, T)$ is compact if for every $f \in L^2(X, \mathfrak{B}, \mu)$ the closure in $L^2(X, \mathfrak{B}, \mu)$ of the orbit $\{f, Tf, T^2f, \dots, T^n f, \dots\}$ is compact.

The topology of $L^2(X, \mathfrak{B}, \mu)$ to which compactness and closure refer in this definition is the strong, or norm, topology. It is easily shown that if the unitary operator $f \rightarrow Tf$ has discrete spectrum then the system $(X, \mathfrak{B}, \mu, T)$ is compact. The converse is also true but somewhat more difficult to prove.

THEOREM 4.1. If $(X, \mathfrak{B}, \mu, T)$ is compact then for any $f \in L^\infty(X, \mathfrak{B}, \mu), f \geq 0$ but f not a.e. 0,

$$(4.1) \quad \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \int f T^i f T^{2i} f \dots T^{kn} f d\mu > 0.$$

Specializing to the case $f = 1_A$, the characteristic function of a set $A \in \mathfrak{B}$ with $\mu(A) > 0$, we obtain (2.2).

PROOF. Let $a = \int f^{k+1} d\mu$. We have $a > 0$. We can assume without loss of generality that $0 \leq f \leq 1$. Choose $\epsilon < a/(k+1)$. Let g_0, \dots, g_k be measurable functions with $0 \leq g_i \leq 1$ and with $\|f - g_i\| < \epsilon, i = 0, 1, \dots, k$. Then

$$\left| \int \prod_{i=0}^k g_i d\mu - \int f^{k+1} d\mu \right| \leq \sum_{j=0}^k \int \prod_{i=0}^{j-1} g_i |g_j - f| f^{k-j} d\mu \leq (k+1)\epsilon < a$$

using the identity (3.8). If we set $a' = a - (k+1)\epsilon$ then we shall have $\int \prod_{i=0}^k g_i d\mu \geq a'$. This observation will be used to prove (4.1) since we shall

show that for a set of n of positive lower density, $\|T^{ln} - f\| < \epsilon$ for $l = 0, 1, \dots, k$. In fact, if we show that $\|T^{nf} - f\| < \epsilon/k$ for a set of n of positive lower density, then since T is measure preserving it will follow that $\|T^{(l+1)nf} - T^{ln} f\| < \epsilon/k$ for these n , and by the triangle inequality, $\|T^{ln} f - f\| < \epsilon$ for $l = 0, 1, \dots, k$.

The property in question follows, however, directly from the compactness of the orbit closure $\overline{\{T^n f\}} \subset L^2(X, \mathfrak{B}, \mu)$. For $\{T^n f, n = 0, 1, 2, \dots\}$ being totally bounded we can find a finite subset $\{T^{n_1} f, T^{n_2} f, \dots, T^{n_r} f\}$ which is ϵ/k -separated: $\|T^{n_i} f - T^{n_j} f\| \geq \epsilon/k$, and which has the maximum cardinality r of such a subset. Now for any n , $\{T^{n+n_1} f, T^{n+n_2} f, \dots, T^{n+n_r} f\}$ is again ϵ/k -separated and has the same cardinality. Thus for some i , $\|T^{n+n_i} f - f\| < \epsilon/k$. This shows that $\|T^{nf} - f\| < \epsilon/k$ for a set of n of positive lower density (in fact, a set of n with bounded gaps), and this completes the proof of the theorem.

5. Weak mixing and compact factors. The two special cases of Theorem IV dealt with in the foregoing sections are clearly very different from one another, and, as is easily shown, are mutually exclusive.

In terms of the spectral decomposition of the unitary operator defined by T on $L^2(X, \mathfrak{B}, \mu)$, these cases correspond respectively to purely continuous spectrum (on the subspace of functions with 0 integral) and to pure point spectrum.

These two cases together are far from exhausting all possibilities; nevertheless, we shall show in this section that taking into account both weak mixing and compact systems, one can establish Theorem IV for some "factor" of an arbitrary system. This will be the first step of our proof of Theorem IV. The proof will then proceed by showing that in a like manner one can establish Theorem IV for larger and larger factors until one arrives at the given system. First, let us explain what is a factor.

Let $(X, \mathfrak{B}, \mu, T)$ be a measure-preserving system. \mathfrak{B} is a σ -algebra of subsets of X , and T is measurable with respect to \mathfrak{B} , so that $T^{-1}\mathfrak{B} \subset \mathfrak{B}$. For sets of \mathfrak{B} we have $\mu(T^{-1}A) = \mu(A)$. Now suppose $\mathfrak{B}_1 \subset \mathfrak{B}$ is another σ -algebra which is T -invariant: $T^{-1}\mathfrak{B}_1 \subset \mathfrak{B}_1$. We can form a new measure-preserving system $(X, \mathfrak{B}_1, \mu, T)$ and we refer to this as a factor of $(X, \mathfrak{B}, \mu, T)$. We shall presently explain why the new system is called a factor of the original system. A factor $(X, \mathfrak{B}_1, \mu, T)$ will be said to be nontrivial if \mathfrak{B}_1 contains sets of measure strictly between 0 and 1. The purpose of this section is to prove the following.

THEOREM 5.1. A system $(X, \mathfrak{B}, \mu, T)$ is weak mixing iff it has no nontrivial compact factors.

As a result, if a system is not weak mixing, some nontrivial factor is compact. Thus in any case either Theorem 3.1 or Theorem 4.1 may be used to establish Theorem IV for some nontrivial factor of an arbitrary system.

We have already noted in §3 that if a system $(X, \mathfrak{B}, \mu, T)$ is weak mixing it is ergodic. We now prove a converse.

PROPOSITION 5.2. *If the system $(X \times X, \mathfrak{B} \times \mathfrak{B}, \mu \times \mu, T \times T)$ is ergodic then $(X, \mathfrak{B}, \mu, T)$ is weak mixing.*

PROOF. Here we shall use the weakest version of the ergodic theorem for ergodic systems. Namely, if a system $(X, \mathfrak{B}, \mu, T)$ is ergodic, then for any $g \in L^2(X, \mathfrak{B}, \mu)$, the averages

$$\frac{1}{N+1} (g(x) + g(Tx) + \dots + g(T^Nx))$$

converge to $\int g d\mu$ in the weak topology of $L^2(X, \mathfrak{B}, \mu)$. (See [H, RN].) In other words, for $f, g \in L^2(X, \mathfrak{B}, \mu)$,

$$(5.1) \quad \frac{1}{N+1} \sum_{n=0}^N \int f T^n g d\mu \rightarrow \int f d\mu \int g d\mu.$$

If we assume $T \times T$ is ergodic then so is T and we have (5.1) as well as the corresponding assertion for $f \otimes f$ and $g \otimes g$. (Recall that $f \otimes f(x_1, x_2) = f(x_1)f(x_2)$.) This gives us

$$(5.2) \quad \frac{1}{N+1} \sum_{n=0}^N \left(\int f T^n g d\mu \right)^2 \rightarrow \left(\int f d\mu \right)^2 \left(\int g d\mu \right)$$

since, e.g., $\int f \otimes f d\mu \times \mu = (\int f d\mu)^2$. But now (5.1) and (5.2) together give us (by the remark following (3.7), i.e., a sequence is almost constant if the L_1 and L_2 norms are almost the same)

$$(5.3) \quad \frac{1}{N+1} \sum_{n=0}^N \left(\int f T^n g d\mu - \int f d\mu \int g d\mu \right)^2 \rightarrow 0$$

which proves the proposition.

Recall now the definition of compact systems. A system $(X, \mathfrak{B}, \mu, T)$ was said to be compact if for every $f \in L^2(X, \mathfrak{B}, \mu)$ the orbit closure $\overline{\{T^n f\}} \subset L^2(X, \mathfrak{B}, \mu)$ is compact. Now this may happen for some f and not for others. Let us say $f \in L^2(X, \mathfrak{B}, \mu)$ is AP (almost periodic) if its orbit closure is compact.

PROPOSITION 5.3. *If for a system $(X, \mathfrak{B}, \mu, T)$ the square $T \times T$ is not ergodic, then there is a nonconstant $f \in L^2(X, \mathfrak{B}, \mu)$ which is AP.*

PROOF. Let $H(x, x')$ be a nonconstant $T \times T$ -invariant function in $L^2(X \times X)$. We can suppose T itself to be ergodic; otherwise a T -invariant function on X would provide the desired AP function. The function $\int H(x, x') d\mu(x)$ is seen to be T -invariant, hence constant, and adding a constant to H we may suppose this vanishes. Since H is not identically 0, there must be some $\varphi \in L^2(X, \mathfrak{B}, \mu)$ with $\int H(x, x')\varphi(x') d\mu(x') \neq 0$ for a set, of x , of positive measure. It follows that

$$(5.4) \quad f(x) = \int H(x, x')\varphi(x') d\mu(x')$$

is also nonconstant, since $\int f(x) d\mu(x) = \int \varphi(x')\int H(x, x') d\mu(x) d\mu(x') = 0$. Now the function in (5.4) is AP. For

$$T^n f(x) = \int H(T^n x, x')\varphi(x') d\mu(x') = \int H(T^n x, T^n x')\varphi(T^n x') d\mu(x')$$

by the invariance of μ , or

$$T^n f(x) = \int H(x, x')T^n \varphi(x') d\mu(x').$$

If $\tilde{H}: L^2(X, \mathfrak{B}, \mu) \rightarrow L^2(X, \mathfrak{B}, \mu)$ denotes the integral operator

$$\tilde{H}\psi(x) = \int H(x, x')\psi(x') d\mu(x')$$

then $\overline{\{T^n f\}} = \overline{\{\tilde{H}(T^n \varphi)\}}$. However it is well known that the operator \tilde{H} is a compact operator [RN] and since the norms of $T^n \varphi$ are constant $\{\tilde{H}(T^n \varphi)\}$ is compact. This completes the proof.

REMARK. We will sketch a more complete picture that comes out of the above argument. (However, it is very helpful when we relativize later, and necessary for several commuting transformations, that we prove only what we actually need.) First, note that we get an invariant distance between points of X , $d(x_1, x_2) = \int |H(x_1, x') - H(x_2, x')| dx'$. If we identify points that are 0 distance apart and consider only sets that respect this identification we get an invariant σ -algebra or factor. (Factors will be discussed later and this remark may be clearer after that discussion.) T acting on this factor is an isometry of a metric space. We would then argue that this metric space is totally bounded. (For any ϵ there are at most a finite number, $N(\epsilon)$, of points ϵ apart. This is so because the set of points at distance $< \frac{1}{2}\epsilon$ from some x has measure $> \alpha > 0$ and by ergodicity this is true for a.e. x . If T is an isometry of a totally bounded metric space and f any function defined on this space then the orbit closure $\overline{\{T^n f\}}$ is compact.)

We turn now to the proof of Theorem 5.1.

PROOF OF THEOREM 5.1. Assume first that $(X, \mathfrak{B}, \mu, T)$ is weak mixing. We shall show that there can be no nonconstant functions $f \in L^2(X, \mathfrak{B}, \mu)$ with compact orbit closure $\overline{\{T^n f\}} \subset L^2(X, \mathfrak{B}, \mu)$. Indeed, the proof of Theorem 4.1 shows that if f has compact orbit closure, for any $\epsilon > 0$, there exists a subset $S \subset N$ of positive lower density so that for $n \in S$, $\|f - T^n f\|_{L^2(X)} < \epsilon$. On the other hand, (3.2) implies that for any $\delta > 0$, $f, g \in L^2(X, \mathfrak{B}, \mu)$

$$\left| \int f T^n g d\mu - \int f d\mu \int g d\mu \right| < \delta$$

but for a set n of density 0. In particular,

$$\left| \int f T^n f d\mu - \left(\int f d\mu \right)^2 \right| < \delta$$

but for a set of n of density 0. For some $n \in S$ we will have

$$\left| \int f^2 d\mu - \left(\int f d\mu \right)^2 \right| < \delta + \epsilon \|f\|_{L^2(X)},$$

and since ε, δ are arbitrary $\int f^2 d\mu = (\int f d\mu)^2$, which for real f implies $f = \text{constant a.e.}$

Now suppose that $(X, \mathfrak{B}, \mu, T)$ is not weak mixing. By Proposition 5.2, there is a nontrivial invariant function on $X \times X$. By Proposition 5.3, there exists a nonconstant $f \in L^2(X, \mathfrak{B}, \mu, T)$ which is AP. We now construct a nontrivial σ -algebra \mathfrak{B}_1 , invariant with respect to T , so that the factor $(X, \mathfrak{B}_1, \mu, T)$ is compact. First, recall that a subset of a complete metric space has compact closure iff for every $\varepsilon > 0$, the subset may be covered by finitely many balls of radius $\leq \varepsilon$. Using this it is easy to verify that the set of $\varphi \in L^2(X, \mathfrak{B}, \mu)$ which are AP is a closed linear subspace of $L^2(X, \mathfrak{B}, \mu)$. It can also be seen that it is closed under lattice operations $\varphi_1, \varphi_2 \rightarrow \max(\varphi_1, \varphi_2), \min(\varphi_1, \varphi_2)$. From this it follows that if f is AP and \mathfrak{B}_0 is the smallest σ -algebra of sets with respect to which f is measurable then each $A, A \in \mathfrak{B}_0$ is AP. Since φ is AP iff $T\varphi$ is AP, the same is true for $\mathfrak{B}_1 =$ the smallest σ -algebra of sets with respect to which f, Tf, T^2f, \dots are measurable. Finally, if each $A, A \in \mathfrak{B}_1$ is AP, so is each $\varphi \in L^2(X, \mathfrak{B}_1, \mu)$. It follows that this factor $(X, \mathfrak{B}_1, \mu, T)$ is compact. This completes the proof of Theorem 5.1.

We now summarize what has been done in the last three sections. Our goal is to prove Theorem IV which asserts that for any measure-preserving system $(X, \mathfrak{B}, \mu, T)$ and $A \in \mathfrak{B}$ with $\mu(A) > 0$, and for any k

$$(2.2 \text{ bis}) \quad \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > 0.$$

If the system $(X, \mathfrak{B}, \mu, T)$ is weak mixing, then we are done by Theorem 3.1. If the system $(X, \mathfrak{B}, \mu, T)$ is not weak mixing, it possesses a factor $(X, \mathfrak{B}_1, \mu, T)$ which is nontrivial and which is compact. Finally by Theorem 4.1, for this system the assertion of Theorem IV is valid. Thus, in any case, Theorem IV is valid for a nontrivial factor of the given system.

What we have done so far will appear as the first step in the proof of Theorem IV. Let us now make the strategy of proof more precise. Fix the measure-preserving system $(X, \mathfrak{B}, \mu, T)$. For any T -invariant σ -algebra $\mathfrak{B}_1 \subset \mathfrak{B}$ (i.e., for any factor $(X, \mathfrak{B}_1, \mu, T)$), we can ask if the conclusion of Theorem IV is valid for elements of \mathfrak{B}_1 ; that is, do we have

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mu(A \cap T^{-n}A \cap \dots \cap T^{-kn}A) > 0$$

for every $k \in \mathbb{Z}^+$ and $A \in \mathfrak{B}_1$ such that $\mu(A) > 0$? If the answer is "yes" we say that the action of T on \mathfrak{B}_1 is "Szemerédi" (shortened to SZ). We have shown that for some nontrivial factor, the action of T is SZ. Naturally, if $\mathfrak{B}_1 \supset \mathfrak{B}_2$ and the action of T on \mathfrak{B}_1 is SZ, so is the action of T on \mathfrak{B}_2 .

Consider the family of all factors (T -invariant sub- σ -algebras) of \mathfrak{B} ordered by inclusion. We shall show in the sequel that

(a) the set \mathfrak{F} of factors for which T is SZ contains a maximal element.

(b) No proper factor can be maximal in \mathfrak{F} .

Clearly (a) and (b) together imply that the action of T on the full σ -algebra \mathfrak{B} is SZ, and with that the proof of Theorem IV will be complete.

The proof of (a) is not hard and we shall dispose of it rather quickly. We shall then turn to the proof of (b). The proof of (b) will be carried out by relativizing the ideas in §§3-5. Namely we shall define the notion of a factor $(X, \mathfrak{B}_2, \mu, T)$ being *relatively weak mixing* with respect to another factor $(X, \mathfrak{B}_1, \mu, T)$, $\mathfrak{B}_2 \supset \mathfrak{B}_1$. We shall also speak of *relative compactness* of factors. In both of these cases we shall show that if the action of T is SZ for the smaller factor, it is so also for the larger factor. Finally we show that if for some factor $(X, \mathfrak{B}_1, \mu, T)$, the original system $(X, \mathfrak{B}, \mu, T)$ fails to be a relatively weak mixing extension, then there exists $\mathfrak{B}'_1 \not\supseteq \mathfrak{B}_1$ with $(X, \mathfrak{B}'_1, \mu, T)$, a relatively compact extension of $(X, \mathfrak{B}_1, \mu, T)$. This implies that if $\mathfrak{B}_1 \neq \mathfrak{B}$, $(X, \mathfrak{B}_1, \mu, T)$ cannot be a maximal member of \mathfrak{F} , proving (b).

6. Factors. In this section we shall elaborate on the notion of a factor of a measure-preserving system. The following is an important example of a factor of a system. Let $(Y, \mathfrak{D}, \nu, S)$ be a measure-preserving system, and suppose $(Z, \mathfrak{E}, \theta)$ is a measure space. Suppose we have a map $y \rightarrow \sigma(y)$ of Y into the measure-preserving maps of $(Z, \mathfrak{E}, \theta)$ such that $(y, z) \rightarrow \sigma(y)z$ is a measurable function from $Y \times Z$ to Z with respect to the σ -algebra $\mathfrak{D} \times \mathfrak{E}$ on $Y \times Z$. It is easy to see that if we set

$$T(y, z) = (\sigma y, \sigma(y)z)$$

then T is measure preserving on $(Y \times Z, \mathfrak{D} \times \mathfrak{E}, \nu \times \theta)$. Setting $X = Y \times Z$, $\mathfrak{B} = \mathfrak{D} \times \mathfrak{E}$, $\mu = \nu \times \theta$, we obtain a measure-preserving system $(X, \mathfrak{B}, \mu, T)$. We shall say that $(X, \mathfrak{B}, \mu, T)$ is a *skew product* of $(Y, \mathfrak{D}, \nu, S)$ with $(Z, \mathfrak{E}, \theta)$. Now let $\pi: X \rightarrow Y$ be the projection $\pi(y, z) = y$ and set $\mathfrak{B}_1 = \pi^{-1}(\mathfrak{D}) \subset \mathfrak{B}$. There is a one-one correspondence between sets of \mathfrak{D} and sets of \mathfrak{B}_1 , and for $D \in \mathfrak{D}$, $\pi^{-1}(SD) = T\pi^{-1}(D)$. Thus we can identify the "factor" $(X, \mathfrak{B}_1, \mu, T)$ with $(Y, \mathfrak{D}, \nu, S)$, the latter being the image of $(X, \mathfrak{B}, \mu, T)$ under the map π . It will often be helpful to think of this case as the typical example of a factor. Indeed, if $(X, \mathfrak{B}, \mu, T)$ is an ergodic system, a theorem of Rokhlin [R] asserts that for any T -invariant σ -algebra $\mathfrak{B}_1 \subset \mathfrak{B}$ we can find a system $(Y, \mathfrak{D}, \nu, S)$ so that $(X, \mathfrak{B}, \mu, T)$ can be identified as a skew product over $(Y, \mathfrak{D}, \nu, S)$ and \mathfrak{B}_1 arises in the manner described.

We shall not need the full strength of Rokhlin's theorem, but rather a consequence of it which is valid quite generally (even without ergodicity). In the case of a skew product, Fubini's theorem gives

$$\mu(A) = \int \theta(A_y) d\nu(y)$$

for $A \in \mathfrak{D} \times \mathfrak{E}$, where $A_y = \{z: (y, z) \in A\}$. We can also write

$$\theta(A_y) = (\delta_y \times \theta)(A)$$

where δ_y is the point mass concentrated at $y \in Y$. Thus we can "disintegrate" the measure μ into a family of measures $\mu_y = \delta_y \times \theta$ for which

$$(6.1) \quad \mu(A) = \int \mu_y(A) d\nu(y),$$

the measure μ_y being concentrated on the fiber $\pi^{-1}(y) \subset X$. Moreover we have

$$(6.2) \quad \mu_y(T^{-1}A) = \mu_{S_y}(A).$$

For the sequel we shall not need to retain the full details of the skew product construction. We shall however make use of the following description which can be applied without real loss of generality to any system and a factor of it. Namely, given a system $(X, \mathfrak{B}, \mu, T)$ and an invariant sub- σ -algebra $\mathfrak{B}_1 \subset \mathfrak{B}$, we can suppose that there exists a system $(Y, \mathfrak{D}, \nu, S)$ and a measurable, measure-preserving map $\pi: X \rightarrow Y$ such that $\mathfrak{B}_1 = \pi^{-1}(\mathfrak{D}) = \{\pi^{-1}(D): D \in \mathfrak{D}\}$. π maps the action of T into that of S : $\pi(Tx) = S\pi(x)$. Moreover we can find a measurable map $\gamma \rightarrow \mu_\gamma$ of Y into measures on (X, \mathfrak{B}) with μ_γ supported on the inverse image $\pi^{-1}(\gamma)$, and such that (6.1) and (6.2) are valid. We can identify the factor $(X, \mathfrak{B}_1, \mu, T)$ with the system $(Y, \mathfrak{D}, \nu, S)$, since in terms of the measures of sets and the respective actions of S and T on sets, $D \rightarrow \pi^{-1}(D)$ gives an isomorphism of the two systems.

Assume now $g \in L^2(X, \mathfrak{B}, \mu)$ and let $(Y, \mathfrak{D}, \nu, S)$ be a factor of $(X, \mathfrak{B}, \mu, T)$. We denote by $E(g|\mathfrak{D})$ the function "g conditioned on \mathfrak{D} " defined by

$$(6.3) \quad E(g|\mathfrak{D})(y) = \int g d\mu_y.$$

Another way of obtaining the same function—or rather the corresponding function in $L^2(X, \mathfrak{B}_1, \mu)$ is to regard $L^2(X, \mathfrak{B}_1, \mu)$ as a closed subspace of $L^2(X, \mathfrak{B}, \mu)$, and to let $g \rightarrow E(g|\mathfrak{D})$ represent the orthogonal projection of $L^2(X, \mathfrak{B}, \mu) \rightarrow L^2(X, \mathfrak{B}_1, \mu) \simeq L^2(Y, \mathfrak{D}, \nu)$. It will be convenient to identify the latter two spaces, and so we shall identify functions on Y with the corresponding functions on X , measurable with respect to \mathfrak{B}_1 . Having said this we can formulate two facts which follow from (6.2) and (6.3).

$$(6.4) \quad E(gf|\mathfrak{D}) = gE(f|\mathfrak{D}) \quad \text{if } g \text{ is measurable } \mathfrak{B}_1.$$

$$(6.5) \quad SE(f|\mathfrak{D}) = E(Tf|\mathfrak{D}).$$

Finally we shall define the *fiber-square* or the *relative-square* of a system $(X, \mathfrak{B}, \mu, T)$ with respect to a factor $\mathfrak{D} = (Y, \mathfrak{D}, \nu, S)$.

Here it will be convenient to use the picture of $(X, \mathfrak{B}, \mu, T)$ as a skew product of $(Y, \mathfrak{D}, \nu, S)$ with a space $(Z, \mathfrak{E}, \theta)$. We then set $\tilde{X} = Y \times Z \times Z$ and let $\tilde{\mathfrak{B}} = \mathfrak{D} \times \mathfrak{E} \times \mathfrak{E}$ with $\tilde{\mu} = \nu \times \theta \times \theta$. If $T(y, z) = (Sy, \sigma(y)z)$ then $\tilde{T}(y, z, z') = (Sy, \sigma(y)z, \sigma(y)z')$. In general if π denotes the measure-preserving map from X to Y we can set $\tilde{X} = \bigcup_{y \in Y} \pi^{-1}(y) \times \pi^{-1}(y) \subset X \times X$, with $\tilde{\mu} = \int \tilde{\mu}_y d\nu(y)$ where $\tilde{\mu}_y = \mu_y \times \mu_y$. $\tilde{\mathfrak{B}}$ is then the restriction of $\mathfrak{B} \times \mathfrak{B}$ to $\tilde{X} \subset X \times X$.

We denote the system $(\tilde{X}, \tilde{\mathfrak{B}}, \tilde{\mu}, \tilde{T})$ by $X \times_Y X$.

An identity that will be used repeatedly is

$$(6.6) \quad \begin{aligned} \int_{X \times_Y X} f(y, z) f(y, z') d\tilde{\mu}(y, z, z') \\ &= \iint f(y, z) f(y, z') d\mu_y(z) d\mu_y(z') d\nu(y) \\ &= \int E(f|\mathfrak{D})^2 d\nu. \end{aligned}$$

7. Maximal SZ factors. Let us now return to assertion (a) at the end of §5. Namely it is claimed that the family \mathfrak{F} of factors of $(X, \mathfrak{B}, \mu, T)$ which are SZ has a maximal element. Let $\{\mathfrak{B}_\alpha\}$ be a totally ordered (by inclusion) family of factors and recall that $\sup_\alpha \mathfrak{B}_\alpha$ is, by definition, the σ -algebra spanned by $\bigcup \mathfrak{B}_\alpha$. More concretely, a set $A \in \mathfrak{B}$ belongs to $\sup \mathfrak{B}_\alpha$ if for every $\varepsilon > 0$ there exist some $A_0 \in \bigcup_\alpha \mathfrak{B}_\alpha$ such that

$$\mu(A \setminus A_0) + \mu(A_0 \setminus A) < \varepsilon.$$

It is clear that if \mathfrak{B}_α is T -invariant for every α , so is $\sup \mathfrak{B}$. The key to the proof of (a) is the following

PROPOSITION 7.1. *Let $\{\mathfrak{B}_\alpha\}$ be a totally ordered family of factors and assume $\mathfrak{B}_\alpha \in \mathfrak{F}$ (that is, the action of T on \mathfrak{B}_α is SZ) for every α . Then $\sup \mathfrak{B}_\alpha \in \mathfrak{F}$.*

PROOF. Let $A \in \sup \mathfrak{B}_\alpha$, $\mu(A) > 0$, and k be fixed. Take $\eta = \frac{1}{2}(k+1)^{-1}$ and $A'_0 \in \mathfrak{B}_{\alpha_0}$ such that

$$(7.1) \quad \mu(A \setminus A'_0) + \mu(A'_0 \setminus A) < \frac{1}{4}\eta\mu(A).$$

We now apply to the factor $(X, \mathfrak{B}_{\alpha_0}, \mu, T)$ the description given above. Namely we suppose given a system $(Y, \mathfrak{D}_0, \nu, T_0)$ and a map $\pi: X \rightarrow Y$ as before so that $\mathfrak{B}_{\alpha_0} = \pi^{-1}(\mathfrak{D}_0)$. $A'_0 \in \mathfrak{B}_{\alpha_0}$ corresponds to a set $A''_0 \in \mathfrak{D}_0$, $A'_0 = \pi^{-1}(A''_0)$. By (7.1) $\mu(A'_0) \geq \mu(A) - \frac{1}{4}\eta\mu(A) > 0$. We claim that the set of $y \in A''_0$ such that $\mu_y(A) < 1 - \eta$ has measure less than $\frac{1}{4}\mu(A)$. For otherwise

$$\begin{aligned} \mu(A'_0 \setminus A) &= \int_{A''_0} \mu_y(A'_0 \setminus A) d\nu(y) \\ &= \int_{A''_0} (1 - \mu_y(A)) d\nu(y) \geq \frac{1}{4}\eta\mu(A) \end{aligned}$$

since for $y \in A''_0$, $\mu_y(A'_0) = 1$, and this inequality contradicts (7.1). Denote by A_0 the subset of points $y \in A''_0$ for which $\mu_y(A) > 1 - \eta$. $A_0 \in \mathfrak{D}_0$ and $\nu(A_0) > \nu(A'_0) - \frac{1}{4}\mu(A) = \mu(A'_0) - \frac{1}{4}\mu(A) > \frac{1}{2}\mu(A)$. Since the action of T on \mathfrak{B}_{α_0} is SZ, or equivalently the action of T_0 on \mathfrak{D}_0 is SZ, we have

$$(7.2) \quad \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \nu(A_0 \cap T_0^{-j}A_0 \cap \dots \cap T_0^{-kj}A_0) = a > 0.$$

We claim now that for every j

$$(7.3) \quad \frac{1}{2}\nu(A_0 \cap T_0^{-j}A_0 \cap \dots \cap T_0^{-kj}A_0) < \mu(A \cap T^{-j}A \cap \dots \cap T^{-kj}A).$$

In part, on account of (6.1), the latter will follow if we show that for $y \in A_0 \cap T_0^{-j}A_0 \cap \dots \cap T_0^{-kj}A_0$,

$$(7.4) \quad \mu_y(A \cap T^{-j}A \cap \dots \cap T^{-kj}A) > \frac{1}{2}.$$

But if $y \in T_0^{-l}A_0$, $l = 0, 1, \dots, k$, we obtain from the definition of A_0 and (6.2) that $\mu_y(T_0^{-l}A) > 1 - \eta$. The intersection of $k+1$ sets each having probability $> 1 - \eta$ has itself probability $> 1 - (k+1)\eta = 1/2$ and (7.4) follows. This

proves (7.3) which together with (7.2) implies

$$\liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N (A \cap T^{-j}A \cap \dots \cap T^{-k}A) \geq \frac{a}{2} > 0,$$

and this completes the proof of Proposition 7.1.

The proof of part (a) now follows immediately from Zorn's lemma and Proposition 7.1.

8. Relative weak mixing. Let $(X, \mathfrak{B}, \mu, T)$ be a measure-preserving system and let $\mathfrak{Q} = (Y, \mathfrak{D}, \nu, S) \simeq (X, \mathfrak{B}, \mu, T)$ be a factor. It will be convenient to assume that X is ergodic.

DEFINITION. X is *weakly mixing relative to* Y if $X \times_Y X$ is ergodic.

LEMMA 8.1. Let $(X, \mathfrak{B}, \mu, T)$ be a relatively weak mixing extension of $(Y, \mathfrak{D}, \nu, S)$ and let $f, g \in L^\infty(X, \mathfrak{B}, \mu)$. Then

$$(8.1) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int [E(fT^n g | \mathfrak{Q}) - E(f | \mathfrak{Q})S^n E(g | \mathfrak{Q})]^2 d\nu = 0.$$

PROOF. By (6.4) and (6.5) we may assume with no loss of generality that $E(f | \mathfrak{Q}) = 0$ (write $f = (f - E(f | \mathfrak{Q})) + E(f | \mathfrak{Q})$) and note that (8.1) is trivial if f is measurable \mathfrak{B}_1). Now we change the scene to $X \times_Y X$. Put $f \otimes f(y, z, z') = f(y, z)f(y, z')$ and $g \otimes g(y, z, z') = g(y, z)g(y, z')$. We have to prove

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \int [E(fT^n g | \mathfrak{Q})]^2 d\nu &= \lim_{N \rightarrow \infty} \frac{1}{N} \int \sum f \otimes f \tilde{T}^n (g \otimes g) d\tilde{\mu} \\ &= \lim_{N \rightarrow \infty} \int f \otimes f \left(\frac{1}{N} \sum_{n=1}^N \tilde{T}^n (g \otimes g) \right) d\tilde{\mu} = 0. \end{aligned}$$

But since \tilde{T} is ergodic (the relative weak mixing) we have by the ergodic theorem

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum \tilde{T}^n (g \otimes g) = \int g \otimes g d\tilde{\mu} = \text{constant}$$

and

$$\int f \otimes f d\tilde{\mu} = \int E(f | \mathfrak{Q})^2 d\nu = 0,$$

which proves (8.2) and hence (8.1).

If f and g are the characteristic functions of A and B , (8.1) can be rewritten as

$$(8.3) \quad \int \frac{1}{N} \sum_{n=1}^N |\mu_y(A \cap T^{-n}B) - \mu_y(A)\mu_y(T^{-n}B)|^2 d\nu(y) \rightarrow 0.$$

This shows that in case of relative weak mixing, for most n and most $y \in Y$, the sets A and $T^n B$ are almost independent with respect to μ_y .

LEMMA 8.2. Let $(X, \mathfrak{B}, \mu, T)$ be a relative weak mixing extension of $\mathfrak{Q} = (Y, \mathfrak{D}, \nu, S)$. Then $(X \times_Y X, \mathfrak{B}, \tilde{\mu}, \tilde{T})$ is also a relatively weak mixing extension of \mathfrak{Q} .

PROOF. We denote $X \times_Y X$ by \tilde{X} and $\tilde{X} \times_Y \tilde{X}$ by \hat{X} . We want to show that $(\hat{X}, \hat{\mathfrak{B}}, \hat{\mu}, \hat{T})$ is ergodic. For this it suffices to show that for a dense set of functions $F, G \in L^2(\hat{X}, \hat{\mathfrak{B}}, \hat{\mu})$

$$(8.4) \quad \frac{1}{N} \sum_{n=1}^N \int \hat{F} \hat{T}^n G d\hat{\mu} \rightarrow \int F d\hat{\mu} \int G d\hat{\mu}.$$

It therefore suffices to verify (8.4) for F and G of the form

$$\begin{aligned} F(y, z_1, z_2, z_3, z_4) &= f_1(y, z_1)f_2(y, z_2)f_3(y, z_3)f_4(y, z_4), \\ G(y, z_1, z_2, z_3, z_4) &= g_1(y, z_1)g_2(y, z_2)g_3(y, z_3)g_4(y, z_4). \end{aligned}$$

We then have

$$\begin{aligned} \int \hat{F} \hat{T}^n G d\hat{\mu} &= \int \left[\int f_1 T^n g_1 d\mu_y \int f_2 T^n g_2 d\mu_y \int f_3 T^n g_3 d\mu_y \int f_4 T^n g_4 d\mu_y \right] d\nu \\ &= \int E(f_1 T^n g_1 | \mathfrak{Q}) E(f_2 T^n g_2 | \mathfrak{Q}) E(f_3 T^n g_3 | \mathfrak{Q}) E(f_4 T^n g_4 | \mathfrak{Q}) d\nu. \end{aligned}$$

But now in light of Lemma 8.1, each expression $E(f_i T^n g_i | \mathfrak{Q})$ can be replaced by $E(f_i | \mathfrak{Q})S^n E(g_i | \mathfrak{Q})$ and so the left-hand side of (8.4) can be replaced by

$$\begin{aligned} \frac{1}{N} \sum_{n=1}^N \int E(f_1 | \mathfrak{Q}) E(f_2 | \mathfrak{Q}) E(f_3 | \mathfrak{Q}) E(f_4 | \mathfrak{Q}) \\ \cdot S^n [E(g_1 | \mathfrak{Q}) E(g_2 | \mathfrak{Q}) E(g_3 | \mathfrak{Q}) E(g_4 | \mathfrak{Q})] d\nu. \end{aligned}$$

Since $(Y, \mathfrak{D}, \nu, S)$ is ergodic

$$\frac{1}{N} \sum_{n=1}^N S^n [E(g_1 | \mathfrak{Q}) E(g_2 | \mathfrak{Q}) E(g_3 | \mathfrak{Q}) E(g_4 | \mathfrak{Q})]$$

$$\rightarrow \left[\int g_1 d\mu_y \int g_2 d\mu_y \int g_3 d\mu_y \int g_4 d\mu_y \right] d\nu = \int G d\hat{\mu},$$

and (8.4) follows.

We now show how relative weak mixing implies "relative weak mixing of all orders".

THEOREM 8.3. Let $(X, \mathfrak{B}, \mu, T)$ be a relative weak mixing extension of $(Y, \mathfrak{D}, \nu, S)$. Then if $f_l \in L^\infty(X, \mathfrak{B}, \mu)$, $l = 0, 1, \dots, k$, we have

$$(8.5) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \left[E \left(\prod_{l=0}^k T^{ln} f_l | \mathfrak{Q} \right) - \prod_{l=0}^k S^{ln} E(f_l | \mathfrak{Q}) \right]^2 d\nu = 0,$$

and

$$(8.6) \quad \lim_{N \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=1}^N \left(\prod_{l=1}^k T^{ln} f_l - \prod_{l=1}^k T^{ln} E(f_l | \mathfrak{Q}) \right) \right\|_{L^2} = 0.$$

Note that $E(f_1|\mathfrak{Q})$ is regarded both as a function in $L^\infty(X, \mathfrak{B}, \mu)$ and in $L^\infty(Y, \mathfrak{Q}, \nu)$.

PROOF. We proceed by induction on k . Lemma 8.1 gives us (8.5)₁ and (8.6)₁ follows from the (norm) ergodic theorem. We now assume that (8.5) _{$k-1$} is valid in full generality, that is, for all relatively weak mixing extensions of \mathfrak{Q} (in particular, by 3.2, for $X \times_\gamma X$) and prove

- (i) (8.5) _{$k-1$} \Rightarrow (8.6) _{k} ,
 - (ii) (8.6) _{k} (for $X \times_\gamma X$) \Rightarrow (8.5) _{k} (for X).
- We begin with (ii) which is easier and follows the line of proof of Lemma 8.1. If f_0 is measurable $\mathfrak{B}_1 = \pi^{-1}(\mathfrak{Q})$, the integrals in (8.5) _{k} have the form

$$(8.7) \quad \int f_0^2 \left[E \left(\prod_{l=1}^k T^{ln} f_l | \mathfrak{Q} \right) - \prod_{l=1}^k S^{ln} E(f_l | \mathfrak{Q}) \right]^2 d\nu \\ \leq \sup |f_0|^2 \int \left[E \left(\prod_{l=0}^{k-1} T^{ln} f_{l+1} | \mathfrak{Q} \right) - \prod_{l=0}^{k-1} S^{ln} E(f_{l+1} | \mathfrak{Q}) \right]^2 d\nu$$

(use (6.4), (6.5) and the fact that S^n is measure preserving) and (8.5) _{k} is reduced to the case (8.5) _{$k-1$} . This enables us to assume, as in the proof of Lemma 8.1, that $E(f_0|\mathfrak{Q}) = 0$. With this, (8.5) _{k} takes the form (see (6.6))

$$(8.8) \quad \lim_{N \rightarrow \infty} \int f_0 \otimes f_0 \left(\frac{1}{N} \sum_{n=1}^N \prod_{l=1}^k \tilde{T}^{ln} f_l \otimes f_l \right) d\tilde{\mu}.$$

By (8.6) _{k} , applied to $X \times_\gamma X$, the limit in (8.8) is the same as

$$\lim_{N \rightarrow \infty} \int f_0 \otimes f_0 \left(\frac{1}{N} \sum_{n=1}^N \prod_{l=1}^k \tilde{T}^{ln} E(f_l \otimes f_l | \mathfrak{Q}) \right) d\tilde{\mu}$$

which is 0 (for all N) since the sum appearing there is constant on fibers and $E(f_0|\mathfrak{Q}) = 0$.

Now to the proof of implication (i). First, write

$$(8.9) \quad \prod_{l=1}^k T^{ln} f_l - \prod_{l=1}^k T^{ln} E(f_l | \mathfrak{Q}) \\ = \sum_{j=1}^k \left(\prod_{l=1}^{j-1} T^{ln} f_l \right) T^{jn} \left(f_j - E(f_j | \mathfrak{Q}) \right) \prod_{l=j+1}^k T^{ln} E(f_l | \mathfrak{Q})$$

and notice that this enables us to prove (8.6) _{k} under the additional assumption that for some l_0 , $1 \leq l_0 \leq k$, $E(f_{l_0}|\mathfrak{Q}) = 0$. We now have to show that under this assumption, $\lim \|\psi_N\|_{L^2} = 0$ where $\psi_N = (1/N) \sum_{n=1}^N \prod_{l=1}^k T^{ln} f_l$.

Rewrite

$$\psi_N = \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{H} \sum_{n=j}^{j+H-1} \prod_{l=1}^k T^{ln} f_l \right) + O(H/N)$$

(where H will be chosen large but much smaller than N). By the convexity of the function $\varphi(x) = x^2$, we have (up to $O(H/N)$)

$$\psi_N^2 \leq \frac{1}{N} \sum_{j=1}^N \left(\frac{1}{H} \prod_{n=j}^{j+H-1} \prod_{l=1}^k T^{ln} f_l \right)^2.$$

By integration and the fact that T is measure preserving,

$$(8.10) \quad \|\psi_N\|_{L^2}^2 \leq \frac{1}{N} \sum_{j=1}^N \frac{1}{H^2} \sum_{n=m=j}^{j+H-1} \int \prod_{l=1}^k T^{ln} f_l \prod_{l=1}^k T^{lm} f_l d\mu \\ = \frac{1}{NH^2} \sum_{j=1}^N \sum_{n,m=j}^{j+H-1} \int \prod_{l=1}^k T^{(l-1)n} (f_l T^{(m-n)} f_l) d\mu.$$

Set $r = m - n$, notice that a pair (n, m) appears in (8.10) only if $|r| = |m - n| < H$, and then for $H - |r|$ values of j , and rewrite (8.10) as

$$(8.11) \quad \|\psi_N\|_{L^2}^2 \leq \frac{1}{H} \sum_{r=1-H}^{H-1} \left(1 - \frac{|r|}{H} \right) \left[\frac{1}{N} \sum_{n=1}^N \int \prod_{l=1}^k T^{(l-1)n} (f_l T^{lr} f_l) d\mu \right] \\ + O(H/N).$$

By (8.5) _{$k-1$} , for a fixed H and every r such that $|r| < H$ we can replace the integrated terms in (8.11), provided N is large enough, by $\prod_{l=1}^k T^{(l-1)n} E(f_l T^{lr} f_l | \mathfrak{Q})$ and obtain

$$\|\psi_N\|_{L^2}^2 \leq \frac{1}{H} \sum_{r=1-H}^{H-1} \left(1 - \frac{|r|}{N} \right) \left[\frac{1}{N} \sum_{n=1}^N \int \prod_{l=1}^k T^{(l-1)n} E(f_l T^{lr} f_l | \mathfrak{Q}) d\mu \right] \\ + O(H/N).$$

Now estimate the integrals appearing in (8.12) by

$$\|E(f_0 T^{0r} f_0 | \mathfrak{Q})\|_{L^2} \prod_{l \neq l_0} \|f_l\|_\infty$$

and recall that $E(f_{l_0}|\mathfrak{Q}) = 0$ so that by Lemma 8.1 most of the terms appearing in (8.12) are small provided H is large enough. Since all the terms are bounded by $\prod \|f_l\|_\infty^2$ and most of them are small, their average (8.12) is small, arbitrarily for large H and N , and the proof is complete.

THEOREM 8.4. Let $(X, \mathfrak{B}, \mu, T)$ be a relatively weak mixing extension of $(Y, \mathfrak{Q}, \nu, S)$. If the action of S on \mathfrak{Q} is SZ, then so is the action of T on \mathfrak{B} .

PROOF. Let $A \in \mathfrak{B}$, $\mu(A) > 0$ and write f for the indicator function of A . Let $a > 0$ be small enough to insure that, for $A_1 = \{y: E(f|\mathfrak{Q}) \geq a\}$, $\nu(A_1) > 0$. By (6.3) and Theorem 8.3 (and $E(f|\mathfrak{Q}) \geq a|_{A_1}$) we have for all k

$$\frac{1}{N} \sum_{l=1}^N \mu \left(\bigcap_{j=0}^k T^{-lj} A \right) > \frac{1}{2} a^{k+1} \cdot \frac{1}{N} \sum_{l=1}^N \mu \left(\bigcap_{j=0}^k S^{-lj} A_1 \right),$$

provided N is large enough.

9. Compact extensions. The notion of compact extension which we discuss in this section, is a natural generalization of group translations to the relative case. We shall prove the analogue of Theorem 8.4 for compact extensions and later show that whenever X is a nonrelatively weak mixing extension of \mathfrak{A} , there exists an intermediate extension which is compact. (Thus, in either case, a proper factor \mathfrak{B} cannot be maximal with respect to the property that the action of T is SZ, which completes the proof of Theorem IV.)

A function $f \in L^2(X, \mathfrak{B}, \mu)$ is said to be almost periodic (AP) relative to the factor \mathfrak{A} if for every $\delta > 0$ there exist functions $g_1, \dots, g_n \in L^2(X, \mathfrak{B}, \mu)$ such that for every $j \in Z$, $\inf_{1 \leq s \leq n} \|T^j f - g_s\|_{L^2(\mu_s)} < \delta$ for almost all $y \in \mathfrak{A}$. We denote the subset of AP functions in $L^2(X, \mathfrak{B}, \mu)$ by AP.

DEFINITION. $(X, \mathfrak{B}, \mu, T)$ is a compact extension of $(Y, \mathfrak{A}, \nu, S)$ if AP is dense in $L^2(X, \mathfrak{B}, \mu)$.

We now show that the property SZ lifts to compact extensions.

THEOREM 9.1. *If $(X, \mathfrak{B}, \mu, T)$ is a compact extension of $(Y, \mathfrak{A}, \nu, S)$ and if the action of S on (Y, \mathfrak{A}, ν) is SZ, then so is the action of T on (X, \mathfrak{B}, μ) .*

PROOF. Let $A \in \mathfrak{B}$, $\mu(A) > 0$, and k be given. We have to prove

$$(*) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{j=1}^N \mu \left(\bigcap_{l=1}^k T^{-j} A \right) > 0$$

which clearly follows from the same inequality holding for a subset of A . We remove from A its portions sitting on fibers for which $\mu_y(A) \leq \frac{1}{2}\mu(A)$. This removes less than half the measure of A and we may therefore assume without loss of generality that $\mu_y(A) \geq \alpha = \frac{1}{2}\mu(A)$ for $y \in A_1$, $\nu(A_1) > \frac{1}{2}\mu(A)$, and $\mu_y(A) = 0$ for $y \notin A_1$.

Denote $f = 1_{A_1}$, the indicator function of A . Our next step is to show that there is no loss of generality in assuming that f is AP. By our assumption there exist, for every $\epsilon > 0$, an AP function f' such that $\|f - f'\|_{L^2(\mu)} < \epsilon^2$. This implies $\|f - f'\|_{L^2(\mu_s)} < \epsilon$ for all y outside a set E_ϵ such that $\nu(E_\epsilon) < \epsilon^2$. If we denote by A_ϵ the set obtained from A by removing the part of A included in the fibers above points in E_ϵ and by f_ϵ the corresponding indicator function, we have that on every fiber and for every $j \in Z$, either

$$\|T^j f_\epsilon - T^j f\|_{L^2(\mu_s)} < \epsilon \quad \text{or} \quad \|T^j f_\epsilon\|_{L^2(\mu_s)} = 0.$$

For $\delta > 0$ let g_1, \dots, g_m be functions as assured by Definition 4.1 for f' , write $g_0 = 0$ and notice that

$$(9.1) \quad \inf_{0 \leq s \leq m} \|T^j f_\epsilon - g_s\|_{L^2(\mu_s)} < \delta + \epsilon, \quad \text{for a.e. } y, j \in Z.$$

Notice that (4.1) remains valid if we replace A_ϵ and correspondingly f_ϵ by its intersection with sets in $\pi^{-1}(\mathfrak{A})$ (i.e., replace f_ϵ by zero on some fibers). We repeat the procedure for a sequence $\{\epsilon_j\}$ going to zero fast enough to insure $\sum \epsilon_j^2 < \frac{1}{2}\mu(A)$, thus removing from A less than half of its measure and obtaining a set whose indicator function is AP.

The condition $f \in \text{AP}$ is equivalent to requiring that the sequence $\{T^j f\}_{j \in Z}$ be totally bounded, or relatively compact, in $L^2(\mu_y)$ for almost all y . Since T

maps μ_y onto $\mu_{S^k y}$, it is clear that $\{T^j f\} \subset L^2(\mu_y)$ is isometric with $\{T^j f\} \subset L^2(\mu_{S^k y})$. We could thus require the total boundedness only for a set of positive measure on Y and the ergodicity of S on Y would give it almost everywhere in a uniform manner.

We shall be considering vectors of the form $(f, T^n f, T^{2n} f, \dots, T^{kn} f)$ on a fiber above $y \in Y$. Denote by $\bigoplus_{l=0}^k L^2(\mu_y)$ the direct sum of $k + 1$ copies of $L^2(\mu_y)$ endowed with the norm $\|(f_0, f_1, \dots, f_k)\|_y = \max_l \|f_l\|_{L^2(\mu_y)}$. It is clear that if $f \in \text{AP}$, the set $\mathcal{L}(k, f) = \{(f, T^n f, \dots, T^{kn} f)\}_{n \in Z}$ is totally bounded in $\bigoplus_{l=0}^k L^2(\mu_y)$ for almost all $y \in Y$, in fact uniformly in $y \in Y$.

So we assume now that $f = 1_A$ is AP and keep the notation $A_1 = \{y; \mu_y(A) > \mu(A)/2\} = \{y; \mu_y(A) > 0\}$. Let $k > 0$ be given; we write

$$\mathcal{L}(k, f, y) = \{(f, T^n f, \dots, T^{kn} f)_y\}_{n \in Z} \subset \bigoplus_{l=0}^k L^2(\mu_y).$$

As we mentioned before, $\mathcal{L}(k, f, y)$ are totally bounded uniformly in y . We are interested only in $y \in A_1$ and for these points, in the elements of $\mathcal{L}(k, f, y)$ for which all the components are nonzero (hence have norm $\geq \frac{1}{2}\mu(A)$ in $L^2(\mu_y)$). We denote this subset of $\mathcal{L}(k, f, y)$ by $\mathcal{L}^*(k, f, y)$, and remark that these are still uniformly totally bounded. For $y \in A_1$ and $\epsilon > 0$ let $M(\epsilon, y)$ denote the maximum cardinality of an ϵ -separated subset of $\mathcal{L}^*(k, f, y)$, that is, a subset such that the $\|\cdot\|_y$ distance between any two of its members exceeds ϵ . The uniform total boundedness of $\mathcal{L}^*(k, f, y)$ clearly implies that $M(\epsilon, y)$ is bounded on A_1 . For every y , $M(\epsilon, y)$ is a monotone decreasing function of ϵ , integer valued, hence locally constant except at the countable set of "critical" ϵ 's where it has jump discontinuity. As a function of y , $M(\epsilon, y)$ is clearly measurable and we can find some $\epsilon_0 < \mu(A)/10k$, $\eta > 0$, and $A_2 \subset A_1$, $\nu(A_2) > 0$, so that $M(\epsilon, y)$ is constant, say M , for $\epsilon_0 - \eta \leq \epsilon \leq \epsilon_0$ and $y \in A_2$.

Take $y_0 \in A_2$ and find integers m_1, \dots, m_M so that $\{(f, T^{m_1} f, \dots, T^{k m_1} f)_{y_0}\}_{j=1}^M$ is a maximal ϵ_0 -separated set in $\mathcal{L}^*(k, f, y_0)$. Consider the function $\|T^{l m_1} f - T^{l m_2} f\|_{L^2(\mu_{y_0})}$, for $1 \leq i < j \leq M$ and $l = 0, 1, \dots, k$, as functions on Y ; these are measurable and we can suppose that y_0 has been chosen so that each neighborhood of the values of these functions at y_0 occurs with positive measure in the set A_2 . We now let A_3 be the subset of A_2 of points y for which, for each i, j, l , $1 \leq i < j \leq M, 0 \leq l \leq k$,

$$(9.2) \quad \|T^{l m_1} f - T^{l m_2} f\|_{L^2(\mu_y)} > \|T^{l m_1} f - T^{l m_2} f\|_{L^2(\mu_{y_0})} - \eta.$$

This will be a subset with $\nu(A_3) > 0$.

We use now the assumption that the action of S on Y is SZ, applying it to A_3 . Let $n \in Z$ such that $\nu(\bigcap_{l=0}^k S^{-n} A_3) > 0$ and let $y \in \bigcap_{l=0}^k S^{-n} A_3$. We have $S^{n l} y \in A_3$ for $l = 0, \dots, k$, and on the other hand, by the definition of $\mathcal{L}^*(k, f, y)$, $A_3 \subset \bigcap_{l=0}^k S^{-l m_j} A_1$ for $j = 1, \dots, M$. Thus $S^{(n+m_j)y} \in A_1$ for $l = 0, \dots, k$ and $j = 1, \dots, M$. We claim that the vectors $\{(f, T^{n+m_j} f, \dots, T^{k(n+m_j) f})_{y}\}_{j=1}^M$ are $\epsilon_0 - \eta$ separated in $\mathcal{L}^*(k, f, y)$, hence form a maximal such set which is therefore $\epsilon_0 - \eta$ dense in $\mathcal{L}^*(k, f, y)$. To prove the separation take $j \neq i$; by the definition of the norm $\|\cdot\|_y$ there exists

some $l, 0 < l \leq k$, such that

$$\|T^{lm_j}f - T^{lm_l}f\|_{L^2(\mu_{y_0})} \geq \epsilon_0;$$

therefore, $\|T^{lm_j}f - T^{lm_l}f\|_{L^2(\mu_{S^{ln_j}})} \geq \epsilon_0 - \eta$ by (9.2), since the points $S^{ln_j}y$ are all inside A_3 . We have $(f, f, \dots, f)_y \in \mathcal{E}^*(k, f, y)$, and hence, for an appropriate $j, (f, T^{n+m_j}f, \dots, T^{k(n+m_j)}f)_y$ is ϵ_0 -close to it. By the choice of ϵ_0 , this implies that

$$\mu_y \left(\bigcap_{l=0}^k T^{-l(n+m_j)}A \right) = \int \prod_{l=0}^k T^{-(n+m_j)l} f d\mu_y > \frac{9}{10} \mu_y(A) > \frac{1}{3} \mu(A).$$

The index j depends on y , but if we sum over j we will have for each $y \in \bigcap_{l=0}^k S^{-ln}A_3$

$$\sum_{j=1}^M \mu_y \left(\bigcap_{l=0}^k T^{-l(n+m_j)}A \right) > \frac{1}{3} \mu(A).$$

Integrating over $\bigcap_{l=0}^k S^{-ln}A_3$ we obtain

$$\sum_{j=1}^M \mu \left(\bigcap_{l=0}^k T^{-l(n+m_j)}A \right) \geq \frac{\mu(A)}{3} \nu \left(\bigcap_{l=0}^k S^{-ln}A_3 \right).$$

Finally averaging for $1 \leq n \leq N$ and passing to the limit $N \rightarrow \infty$, we obtain

$$M \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{p=1}^N \mu \left(\bigcap_{l=0}^k T^{-lp}A \right) \geq \frac{\mu(A)}{3} \liminf_{N \rightarrow \infty} \frac{1}{N} \sum_{p=1}^N \left(\bigcap_{l=0}^k S^{-lp}A_3 \right).$$

This completes the proof of Theorem 9.1.

10. Existence of compact extensions. Let us take account of where we stand. By Proposition 7.1 we know that for any $(X, \mathfrak{B}, \mu, T)$ there exists a maximal factor for which the action of T is SZ. By Theorem 8.4 we know that no proper extension of this factor could be relatively weak mixing, for otherwise this extension would enjoy the property SZ. By Theorem 9.1 no proper extension of this factor may be relatively compact. We complete the argument proving Theorem IV by showing in this section that if $(X, \mathfrak{B}, \mu, T)$ is a proper extension of $(Y, \mathfrak{D}, \nu, S)$ and is not relatively weak mixing, there necessarily exists a proper subextension of $(Y, \mathfrak{D}, \nu, S)$ which is relatively compact. It follows that $(X, \mathfrak{B}, \mu, T)$ already has the property SZ.

THEOREM 10.1. *If $\mathfrak{X} = (X, \mathfrak{B}, \mu, T)$ is an extension of $\mathfrak{Y} = (Y, \mathfrak{D}, \nu, S)$ which is not relatively weak mixing, then there exists an intermediate factor \mathfrak{X}^* between \mathfrak{Y} and \mathfrak{X} so that \mathfrak{X}^* is a compact extension of \mathfrak{Y} .*

PROOF. It will be convenient to use the Rokhlin picture and represent \mathfrak{X} as a skew product: $(X, \mathfrak{B}, \mu) = (Y, \mathfrak{D}, \nu) \times (Z, \mathfrak{C}, \theta)$ with $T(y, z) = (Sy, \sigma(y)z)$. We also suppose, again for convenience, that \mathfrak{X} is ergodic. Since \tilde{T} on $X \times_Y X$ is not ergodic, there exists a bounded function $H(x, x')$ on $X \times_Y X$ which is invariant under \tilde{T} , but is not a function of x alone or x' alone. We set

$X \times_Y X = Y \times Z \times Z$ and write $H(x, x') = H(y, z, z')$. There exists a function $\varphi \in L(X, \mathfrak{B}, \mu)$, so that the convolution $H * \varphi$ defined by

$$(10.1) \quad H * \varphi(y, z) = \int H(y, z, z') \varphi(y, z') d\theta(z')$$

is not a function of y alone. We have

$$(10.2)$$

$$\begin{aligned} T(H * \varphi)(y, z) &= H * \varphi(Sy, \sigma(y)z) = \int H(Sy, \sigma(y)z, z') \varphi(Sy, z') d\theta(z') \\ &= \int H(Sy, \sigma(y)z, \sigma(y)z') \varphi(Sy, \sigma(y)z') d\theta(z') \\ &= \int H(y, z, z') \varphi(Sy, \sigma(y)z') d\theta(z') = H * T\varphi \end{aligned}$$

using the fact that $\sigma(y)$ is measure preserving and that H is \tilde{T} -invariant.

For each y the integral operator (10.1) is compact and it follows that for every $\delta > 0$, there exists an integer $M = M(y, \delta)$ such that $\{T^j/(H * \varphi)\}_{j=-M}^M = \{H * T^j\varphi\}_{j=-M}^M$ is δ -dense in $\{T^j/(H * \varphi)\}_{j \in \mathbb{Z}}$ in the $L^2(\mu_y)$ norm. For every $\epsilon > 0$ we can now take $M_{\epsilon, \delta}$ big enough to insure $M(y, \delta) < M_{\epsilon, \delta}$ for all y outside some set $E(\delta, \epsilon)$ such that $\nu(E(\delta, \epsilon)) < \epsilon$. We repeat this argument for a sequence $\{\delta_j\}, \delta_j \rightarrow 0$ and $\{\epsilon_j\}, \sum_{j=1}^{\infty} \epsilon_j$ arbitrarily small and write

$$(10.3) \quad f(y, z) = \begin{cases} 0 & \text{if } y \in \bigcup E(\delta_j, \epsilon_j), \\ H * \varphi & \text{otherwise.} \end{cases}$$

Clearly $\|f - H * \varphi\|_{L^2} \leq \|H\|_{L^\infty} \|\varphi\|_{L^\infty} \sum_{j=1}^{\infty} \epsilon_j$ (which is arbitrarily small) and for every δ the family $\{0\} \cup \{T^j/(H * \varphi)\}_{j=-M}^M$ is δ -dense, for M large enough, in $\{T^j/f\}_{j \in \mathbb{Z}}$ in the $L^2(\mu_y)$ norm for every y . If we denote by \mathfrak{F} the algebra spanned by $\{H * \varphi; H \in L^\infty(X \times_Y X), TH = H, \varphi \in L^\infty(X)\}$ then by (10.2) \mathfrak{F} is T -invariant and the AP functions in \mathfrak{F} are dense in \mathfrak{F} . Let \mathfrak{B}^* be the smallest sub- σ -algebra of \mathfrak{B} with respect to which all the elements of \mathfrak{F} are measurable. \mathfrak{B}^* clearly contains (properly) $\mathfrak{B}_1 = \pi^{-1}(\mathfrak{D})$ and the T -invariance of \mathfrak{F} implies the same for \mathfrak{B}^* . \mathfrak{F} is dense in $L^2(X, \mathfrak{B}^*, \mu)$ and so the set of AP functions is dense in $L^2(X, \mathfrak{B}^*, \mu)$. If we denote the factor corresponding to $(X, \mathfrak{B}^*, \mu, T)$ by \mathfrak{X}^* then \mathfrak{X}^* is a compact extension of \mathfrak{Y} . This completes the proof of Theorem 10.1.

We remark that throughout our use of ergodicity has been merely for convenience and essentially the same arguments will work in general without the use of the Rokhlin theorem representing an extension as a skew product. In any case we have achieved a proof of Theorem IV in the case that $(X, \mathfrak{B}, \mu, T)$ is ergodic. Now in fact using the decomposition of an arbitrary measure into ergodic components it is easy to see that Theorem IV, once established for ergodic systems, is proved for arbitrary measure-preserving systems.

REFERENCES

- [F] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton Univ. Press, Princeton, N. J., 1981.
- [FK] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. d'Analyse Math. **34** (1978), 275–291.
- [H] P. R. Halmos, *Lectures on ergodic theory*, Chelsea, New York, 1960.
- [R] V. A. Rokhlin, *Selected topics from the metric theory of dynamical systems*, Amer. Math. Soc. Transl. (2) **49** (1966), 171–209.
- [RN] F. Riesz and B. Sz. Nagy, *Functional analysis*, Ungar, New York, 1955.

INSTITUTE OF MATHEMATICS, HEBREW UNIVERSITY OF JERUSALEM, JERUSALEM, ISRAEL

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305