Host-Kra-Ziegler factors

by **A. Leibman**

The *nonconventional*, or *multiple* ergodic averages

$$\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot \ldots \cdot T^{kn} f_k, \tag{1}$$

where T is a measure preserving transformation of a probability measure space X and $f_1, \ldots, f_k$ are (bounded) measurable functions on X, were introduced by H. Furstenberg in his ergodic theoretical proof of Szemerédi's theorem ([F]). For the needs of Szemerédi's theorem it was sufficent to show that, in the case $f_1 = \ldots = f_k > 0$, the liminf of the averages (1) is nonzero, and Furstenberg had confined himself to proving this fact. The question whether the limit of the multiple ergodic averages exists in L^1-sense was an open problem for more than twenty years, until it was answered positively by Host and Kra ([HK1]) and, independently, by Ziegler ([Z]). The way of solving this problem was prompted by Furstenberg: one has to determine a factor Z of X which is *characteristic* for the averages (1), which means that the limiting behavior of (1) only depends on the expectation of f_i with respect to Z: $\bigl\|\frac{1}{N}\sum_{n=1}^{N}\bigl(T^n f_1 \cdot \ldots \cdot T^{kn} f_k - T^n E(f_1|Z) \cdot \ldots \cdot T^{kn} E(f_k|Z)\bigr)\bigr\|_{L^1(X)} \longrightarrow 0$ for any $f_1, \ldots, f_k \in L^\infty(X)$. Once a characteristic factor Z has been found, the problem is restricted to the system (Z,T); one therefore succeeds if he/she manages to show that every system (X,T) possesses a characteristic factor with a relatively simple structure, so that the convergence of averages (1) can be easily established for it. For example, under the assumption that T is ergodic (which always may be done due to the ergodic decomposition theorem), one can show that the Kronecker factor K of X is characteristic for the two-term averages $\frac{1}{N}\sum_{n=1}^{N} T^n f_1 \cdot T^{2n} f_2$ (see [F]). Since K has structure of a compact abelian group on which T acts as a translation, it is not hard to see that the averages above converge for $f_1, f_2 \in L^\infty(K)$.

A k-step nilsystem is a pair (N,T) where N is a compact homogeneous space of a k-step nilpotent group G and T is a translation of N defined by an element of G. In the case G is a nilpotent Lie group, N is called *a nilmanifold*, and N is called *a pro-nilmanifold* if it is representable as an inverse limit of nilmanifolds. After Conze and Lesigne had shown ([CL1], [CL2], [CL3]) that the characteristic factor for the three-term multiple ergodic averages is a two-step nilsystem, it was natural to conjecture that the characteristic factor for the averages (1) with arbitrary k is a $(k-1)$-step nilsystem. Host-Kra and Ziegler have confirmed this conjecture by constructing such factors.

Ziegler's factors $Y_{k-1}(X,T)$, $k = 2,3,\ldots$, are characteristic for the averages of the form

$$\frac{1}{N}\sum_{n=1}^{N} T^{a_1 n} f_1 \cdot \ldots \cdot T^{a_k n} f_k \tag{2}$$

for any $a_1,\dots,a_k\in\mathbb{Z}$. Ziegler's construction is a (very complicated) extension of Conze-Lesigne's one: she obtains the factor $Y_k(X,T)$ as a product of $Y_{k-1}(X,T)$ and a compact abelian group H so that T acts as a skew-shift on $Y_k(X,T)=Y_{k-1}(X,T)\times H$, $T(y,h)=(Ty,h+\rho(y))$, with ρ satisfying certain conditions that allow one to impose on $Y_k(X,T)$ the structure of a k-step pro-nilmanifold with T being a translation on it. She also shows that $Y_{k-1}(X,T)$ is the minimal factor of X which is characteristic for all averages of the form (2), and the maximal factor of X having the structure of a $(k-1)$-step pro-nilmanifold.

Host and Kra used another, very elegant construction. They first describe the characteristic factor for the (numerical) averages of the form

$$\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{n_k=1}^{N_k}\cdots\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\int_X\prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}}T^{\epsilon_1 n_1+\ldots+\epsilon_k n_k}f_{\epsilon_1,\dots,\epsilon_k}.$$

Though this expression looks frightening, it is quite natural (for the case $k=2$ it is simply $\lim_{N_2\to\infty}\frac{1}{N_2}\sum_{n_k=1}^{N_k}\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\int_X f_{0,0}T^{n_1}f_{1,0}T^{n_2}f_{0,1}T^{n_1+n_2}f_{1,1}$) and the corresponding characteristic factor, which will be denoted by $Z_{k-1}(X,T)$, can be easily constructed. Then Host and Kra prove that, for each k, the factor $Z_k(X,T)$ possess structure of a k-step pro-nilmanifold, and that it is characteristic for ergodic averages of other sorts. In particular, it is shown in [HK1] that $Z_{k-1}(X,T)$ is characteristic for the averages (1), and in [HK1] that $Z_k(X,T)$ is characteristic for the averages of the form (2). We will undertake an alittle bit more detailed analysis to show that, actually, for $k\geq 2$ already $Z_{k-1}(X,T)$ is characteristic for the averages (2). This will imply that the Host-Kra factors $Z_{k-1}(X,T)$ coincide with the corresponding Ziegler factors $Y_{k-1}(X,T)$. Indeed, being a $(k-1)$-step pro-nilmanifold, $Z_{k-1}(X,T)$ is a factor of $Y_{k-1}(X,T)$; on the other hand, since $Y_{k-1}(X,T)$ is the minimal characteristic factor for the averages (2), it is a factor of $Z_{k-1}(X,T)$.

Let us settle terminology and notation. We will assume the measure spaces we deal with to be *regular*, that is, compact metric endowed with probability Borel measures. Though some specific measure on each measure space is meant, to simplify notation we will not usually specify it.

Given a measurable mapping $p\colon X\longrightarrow Y$ from a measure space $(X,\mathcal{B})$ onto a measure space $(Y,\mathcal{D})$, we will call Y *a factor* of X. $p^{-1}(\mathcal{D})$ is a sub-σ-algebra of $\mathcal{B}$, which we will identify with $\mathcal{D}$. Conversely, with any sub-σ-algebra of $\mathcal{B}$ a factor of X is associated. Let Y be a factor of $(X,\mathcal{B},\mu)$ and let $p\colon X\longrightarrow Y$ be the factorization mapping. We then have *the decomposition* $X=\bigcup_{y\in Y}X_y$ of X with respect to Y, where we put $X_y=p^{-1}(y)$, $y\in Y$, and equip each X_y with a Borel measure μ_y so that $\int_Y\mu_y=\mu$.

Let Y be a factor of $(X,\mathcal{B},\mu)$ and of $(X',\mathcal{B}',\mu')$ and let $p\colon X\longrightarrow Y$ and $p'\colon X'\longrightarrow Y$ be the factorization mappings. *The relative product* $X\times_Y X'$ is the space $\big\{(x,x')\in X\times X':p(x)=p'(x')\big\}$. Y is a factor of $X\times_Y X'$, and if $X=\bigcup_{y\in Y}X_y$, $X'=\bigcup_{y\in Y}X'_y$ are the decompositins of X and of X' with respect to Y, then $X\times_Y X'=\bigcup_{y\in Y}(X_y\times_Y X'_y)$ is the decomposition of $X\times_Y X'$. The measure on $X\times_Y X'$ is defined as $\int_Y\mu_y\times\mu'_y$. $X\times_Y X'$ is *a joining* of X and X', which means that both X and X' are factors of $X\times_Y X'$ and $\mathcal{B}\otimes\mathcal{B}'$ coincides with the Borel σ-algebra of this space.

Let T be a measure preserving transformation of a measure space $(X,\mathcal{B})$. If Y is a factor of X associated with a T-invariant sub-σ-algebra of $\mathcal{B}$, then the action of T reduces to a measure preserving action on Y, which we will also denote by T. In this situation, the restriction of $T\times T$ on $X\times_Y X$ is a measure preserving transformation of this space.

We will denote by $\mathcal{I}(X,T)$ the σ-algebra of T-invariant measurable subsets of X and by $I(X,T)$ the factor of X associated with $\mathcal{I}(X,T)$. To simplify notation, we will write $X\times_T X$ for $X\times_{I(X,T)} X$.

The Host-Kra factors of X with respect to T are constructed in the following way. One puts $X_T^{[0]}=X$, $T^{[0]}=T$, and when $X_T^{[k]}$ and $T^{[k]}$ have been defined for certain k, let $X_T^{[k+1]}=X_T^{[k]}\times_{T^{[k]}} X_T^{[k]}$ and let $T^{[k+1]}$ be the restriction of $T^{[k]}\times T^{[k]}$ on $X_T^{[k+1]}$. Then, for any $k=0,1,\ldots$, $X_T^{[k]}$ is a joining of 2^k copies of X. For $k=0,1,\ldots$, let $\mathcal{Z}_k(X,T)$ be the minimal σ-algebra on X such that $\mathcal{I}(X_T^{[k]},T^{[k]})\subseteq\mathcal{Z}_k(X,T)^{\otimes 2^k}$. *The k-th Host-Kra factor $Z_k(X,T)$ of X with respect to T* is the factor of X associated with $\mathcal{Z}_k(X,T)$. (See [HK1].)

Let $X=\bigcup_{\alpha\in J}X_\alpha$ be a partition of X into T-ivariant subsets. Since in distinct sets X_α "life goes independently", we have:

Lemma 1. *For any k, the spaces $X_T^{[k]}$, $I(X_T^{[k]},T^{[k]})$ and $Z_k(X,T)$ partition, respectively, to $\bigcup_{\alpha\in J}(X_\alpha)_T^{[k]}$, $\bigcup_{\alpha\in J}I((X_\alpha)_T^{[k]},T^{[k]})$ and $\bigcup_{\alpha\in J}Z_k(X_\alpha,T)$.*

In particular, when J is finite, this implies $\mathcal{I}(X_T^{[k]},T^{[k]})=\prod_{\alpha\in J}\mathcal{I}((X_\alpha)_T^{[k]},T^{[k]})$ and $\mathcal{Z}_k(X,T)=\prod_{\alpha\in J}\mathcal{Z}_k(X_\alpha,T)$.

Our first goal is to show that the Host-Kra factors associated with any nontrivial power of a measure preserving transformation are the same as for the transformation itself. In [HK2] this fact was established for totally ergodic transformations; we extend it to the general case.

Theorem 2. *For any $l\neq 0$ and $k\geq 1$ the k-th Host-Kra factor $Z_k(X,T^l)$ of X with respect to T^l coincides with the k-th Host-Kra factor $Z_k(X,T)$ of X with respect to T.*

We fix a nonzero integer l. It follows from Lemma 1 that it suffices to prove Theorem 2 for an ergodic T only. We first assume that T^l is also ergodic. Given a measure preserving transformation S of a measure space Y, let us denote by $\mathcal{E}_\lambda(Y,S)$ the eigenspace of S in $L^1(Y)$ corresponding to the eigenvalue λ, $\mathcal{E}_\lambda(Y,S)=\bigl\{f\in L^1(Y):S(f)=\lambda f\bigr\}$. In particular, $\mathcal{E}_1(Y,S)$ is the space of S-invariant integrable functions on Y, which we will denote by $\mathcal{L}(Y,S)$.

Lemma 3. ([HK2]) *Let S be a measure preserving transformation of a measure space Y. If S^l is ergodic, then $I(Y\times Y,S^l\times S^l)=I(Y\times Y,S\times S)$.*

Proof. S^l is ergodic means that $\mathcal{E}_\lambda(Y,S)=\{0\}$ for all $\lambda\neq 1$ with $\lambda^l=1$. We have $\mathcal{L}(Y\times Y,(S\times S)^l)\subseteq\mathrm{Span}\bigl\{\mathcal{E}_\lambda(Y\times Y,S\times S):\lambda^l=1\bigr\}$. For any $\lambda\in\mathbb{C}$, $|\lambda|=1$, the space $\mathcal{E}_\lambda(Y\times Y,S\times S)$ is spanned by the the functions of the form $f\otimes g$ where $f\in\mathcal{E}_{\lambda_1}(Y,S)$ and $g\in\mathcal{E}_{\lambda_2}(Y,S)$ with $\lambda_1+\lambda_2=\lambda$. For such a function, $fg\in\mathcal{E}_\lambda(Y,S)$. Thus, for any $\lambda\neq 1$ with $\lambda^l=1$ we have $\mathcal{E}_\lambda(Y\times Y,S\times S)=\{0\}$. Hence, $\mathcal{L}(Y\times Y,S^l\times S^l)\subseteq\mathcal{E}_1(Y\times Y,S\times S)=$

$\mathcal{L}(Y\times Y,S\times S)$. With the evident opposite inclusion $\mathcal{L}(Y\times Y,S\times S)\subseteq\mathcal{L}(Y\times Y,S^l\times S^l)$ this implies $\mathcal{I}(Y\times Y,S^l\times S^l)=\mathcal{I}(Y\times Y,S\times S)$. ■

Lemma 4. ([HK2]) *Let T be a measure preserving transformation of a measure space X. If T^l is ergodic then $X^{[k]}_{T^l}=X^{[k]}_T$ and $I\big(X^{[k]}_{T^l},(T^l)^{[k]}\big)=I\big(X^{[k]}_T,T^{[k]}\big)$ for all $k\geq 0$.*

Proof. For $k=0$ the statement is trivial. Assume by induction that, for some $k\geq 0$, $Y=X^{[k]}_{T^l}=X^{[k]}_T$ and $I=I(Y,(T^l)^{[k]})=I(Y,T^{[k]})$. Then $X^{[k+1]}_{T^l}=X^{[k+1]}_T=Y\times_I Y$. Let $Y=\bigcup_{\alpha\in I}Y_\alpha$ be the decomposition of Y with respect to I and for each $\alpha\in I$ let $S_\alpha=T^{[k]}|_{Y_\alpha}$. By the induction assumption S^l_α is ergodic on Y_α for every $\alpha\in I$, thus by Lemma 1 and Lemma 3 applied to the systems (Y_α,S_α),

$$I(Y\times_I Y,(T^l)^{[k]}\times(T^l)^{[k]})=\bigcup_{\alpha\in I}I\big(Y_\alpha\times Y_\alpha,S^l_\alpha\times S^l_\alpha\big)=\bigcup_{\alpha\in I}I\big(Y_\alpha\times Y_\alpha,S_\alpha\times S_\alpha\big)$$
$$=I(Y\times_I Y,T^{[k]}\times T^{[k]}).$$

■

It follows that $Z_k(X,T^l)=Z_k(X,T)$ for all $k\geq 0$, which proves Theorem 2 in the case T^l is ergodic.

Now assume that T is ergodic whereas T^l is not. We may assume that l is a prime integer. In this case X is partitioned, up to a subset of measure 0, to measurable subsets $X_0,\ldots,X_{l-1}$ such that $T(X_i)=X_{i+1}$ for all $i\in\mathbb{Z}_l$. (We identify $\{0,\ldots,l-1\}$ with $\mathbb{Z}_l=\mathbb{Z}/(l\mathbb{Z})$ in order to have $(l-1)+1=0$.)

Lemma 5. *Let X be a disjoint union of measure spaces $X_0,\ldots,X_{l-1}$ and let T be an invertible measure preserving transformation of X such that $T(X_i)=X_{i+1}$, $i\in\mathbb{Z}_l$. Then $X_0,\ldots,X_{l-1}\in\mathcal{Z}_1(X,T)$.*

Proof. We may assume that T is ergodic; otherwise we pass to the ergodic components of X with respect to T. Then $X^{[1]}_T=X^2$ and $T^{[1]}=T\times T$. The "diagonal" $W=X_0^2\cup\ldots\cup X_{l-1}^2\subseteq X^{[1]}_T$ is $T^{[1]}$-invariant and therefore W is $\mathcal{Z}_1(X,T)\otimes\mathcal{Z}_1(X,T)$-measurable. By the Fubini theorem the "fibers" $X_0,\ldots,X_{l-1}$ of W are $\mathcal{Z}_1(X,T)$-measurable. ■

Lemma 6. *Let Y be a disjoint union of measure spaces $Y_0,\ldots,Y_{l-1}$ and let S be an invertible measure preserving transformation of Y such that $S(Y_i)=Y_{i+1}$, $i\in\mathbb{Z}_l$. Then $Y\times_S Y$ is partitioned to $\bigcup_{i,j\in\mathbb{Z}_l}Y_{i,j}$ where $Y_{i,i}=Y_i\times_{S^l}Y_i$ for all $i\in\mathbb{Z}^l$, and for all $i,j,s,t\in\mathbb{Z}_l$, $(S^s\times S^t)|_{Y_{j,j}}$ is an isomorphism between $Y_{i,j}$ and $Y_{i+s,j+t}$. In particular, $(S\times S)(Y_{i,j})=Y_{i+1,j+1}$ for all i,j, thus the subsets $V_i=\bigcup_{j\in\mathbb{Z}_l}Y_{j,j+i}$, $i\in\mathbb{Z}_l$ are $S\times S$-invariant and partition $Y\times_S Y$, and $\mathrm{Id}_{Y_0}\times S^i$ is an isomorphism between V_0 and V_i.*

Proof. We first determine $I(Y,S)$. Let A be a measurable S-invariant subset of Y. Let $A_i=A\cap Y_i$, $i\in\mathbb{Z}_l$. Then A_0 is S^l-invariant, and $A_i=S^i(A_0)$ for $i\in\mathbb{Z}_l$. So, the mapping $A\mapsto A\cap Y_0$ is an isomorphism between $\mathcal{I}(Y,S)$ and $\mathcal{I}(Y_0,S^l)$, which induces an isomorphism (up to measure scaling) between $I(Y,S)$ and $I(Y_0,S^l)$.

Let $Y_0=\bigcup_{\alpha\in I}Y_{0,\alpha}$ be the decomposition of Y_0 with respect to $I=I(Y_0,S^l)$. For every $\alpha\in I$ and $i\in\mathbb{Z}_l\setminus\{0\}$ define $Y_{i,\alpha}=S^i(Y_{0,\alpha})$ and $Y_\alpha=\bigcup_{i\in\mathbb{Z}_l}Y_{i,\alpha}$. Then $Y=\bigcup_{\alpha\in I}Y_\alpha$

is the decomposition of Y with respect to I. We have

$$Y_S^{[1]}=\bigcup_{\alpha\in I}Y_\alpha\times_S Y_\alpha=\bigcup_{\alpha\in I}\bigcup_{i,j\in\mathbb{Z}_l}Y_{i,\alpha}\times Y_{j,\alpha}=\bigcup_{i,j\in\mathbb{Z}_l}\bigcup_{\alpha\in I}Y_{i,\alpha}\times Y_{j,\alpha}=\bigcup_{i,j\in\mathbb{Z}_l}Y_{i,j},$$

where $Y_{i,j}=\bigcup_{\alpha\in I}Y_{i,\alpha}\times Y_{j,\alpha}$. In particular, $Y_{i,i}=\bigcup_{\alpha\in I}Y_{i,\alpha}\times Y_{i,\alpha}=Y_i\times_{S^l}Y_i$ for all $i\in\mathbb{Z}_l$. ■

Lemma 7. *Let X be a disjoint union of measure spaces $X_0,\dots,X_{l-1}$ and let T be an invertible measure preserving transformation of X such that $T(X_i)=X_{i+1}$, $i\in\mathbb{Z}_l$. Then for any $k\geq 0$, $X_T^{[k]}$ can be partitioned, $X_T^{[k]}=\bigcup_{j=1}^{l^k}W_j$, into $T^{[k]}$-invariant measurable subsets $W_1,\dots,W_{l^k}$, such that $W_1=\bigcup_{i\in\mathbb{Z}_l}(X_i)_{T^l}^{[k]}$ with $T^{[k]}\big((X_i)_{T^l}^{[k]}\big)=(X_{i+1})_{T^l}^{[k]}$ for each i, and for each $j=2,\dots,l^k$ there exists an isomorphism $\tau_j\colon W_1\longrightarrow W_j$, which in each coordinate is given by a power of T (that is, if $\pi_n\colon X^{[k]}\longrightarrow X$, $n=1,\dots,2^k$, are the projection mappings, for each n there exists $m\in\mathbb{Z}$ such that $\pi_n\circ\tau_j=T^m\circ\pi_n|_{W_1}$).*

Proof. We use induction on k; for $k=0$ the statement is trivial. Assume that it holds for some $k\geq 0$. Then by Lemma 1, $X_T^{[k+1]}=\bigcup_{j=1}^{l^k}W_j\times_{T^{[k]}}W_j$. The isomorophisms τ_j between W_1 and W_j, commuting with $T^{[k]}$, induce isomorphisms $\tau_j\times\tau_j$ between $W_1\times_{T^{[k]}}W_1$ and $W_j\times_{T^{[k]}}W_j$, $j=1,\dots,l^k$, and $\tau_j\times\tau_j$ act on coordinates as powers of T if τ_j do. Thus, we may focus on $W_1\times_{T^{[k]}}W_1$ only.

By Lemma 6 applied to $W_1=\bigcup_{i\in\mathbb{Z}_l}(X_i)_{T^l}^{[k]}$ and $T^{[k]}|_{W_1}$, $W_1\times_{T^{[k]}}W_1$ is partitioned into $T^{[k]}\times T^{[k]}=T^{[k+1]}$-invariant subsets $V_0,\dots,V_{l-1}$ such that

$$V_0=\bigcup_{i\in\mathbb{Z}_l}(X_i)_{T^l}^{[k]}\times_{(T^{[k]})^l}(X_i)_{T^l}^{[k]}=\bigcup_{i\in\mathbb{Z}_l}(X_i)_{T^l}^{[k+1]}$$

and $V_1,\dots,V_{l-1}$ are isomorphic to V_0 by isomorphisms whose projections on the factors $(X_i)_{T^l}^{[k]}$ coincide with some powers of $T^{[k]}$. ■

End of the proof of Theorem 2. Assume that T is ergodic on X, l is a prime integer and T^l is not ergodic on X. Let $k\geq 1$. Ignoring a subset of measure 0 in X, partition X to measurable subsets $X_0,\dots,X_{l-1}$ such that, for each i, $T(X_i)=X_{i+1}$. Let $k\geq 1$ and let $W_1,\dots,W_{l^k}$ be as in Lemma 7. Since $X_0,\dots,X_{l-1}$ are T^l-invariant, by Lemma 1 we have $\mathcal{I}(X^{[k]},(T^l)^{[k]})=\prod_{i\in\mathbb{Z}_l}\mathcal{I}(X_i^{[k]},(T^l)^{[k]})$ and $\mathcal{Z}_k(X,T^l)=\prod_{i\in\mathbb{Z}_l}\mathcal{Z}_k(X_i,T^l)$. Any $T^{[k]}$-invariant measurbale subset A of $W_1=\bigcup_{i\in\mathbb{Z}_l}(X_i)_{T^l}^{[k]}$ has form $A=\bigcup_{i\in\mathbb{Z}_l}A_i$ where $A_i\in\mathcal{I}(X_i,(T^l)^{[k]})$ and $T^{[k]}(A_i)=A_{i+1}$, $i\in\mathbb{Z}_l$. Thus, $\mathcal{I}(W_1,T^{[k]})\subseteq\mathcal{I}(X^{[k]},(T^l)^{[k]})\subseteq\mathcal{Z}_k(X,T^l)^{\otimes 2^k}$. Since $\mathcal{Z}_k(X,T^l)$ is T-invariant and $W_n=\tau_n(W_1)$ where τ_n is an isomorphism acting on each coordinate as a power of T, $\mathcal{I}(W_n,T^{[k]})\subseteq\mathcal{Z}_k(X,T^l)^{\otimes 2^k}$ for any n. Hence, $\mathcal{Z}_k(X,T)\subseteq\mathcal{Z}_k(X,T^l)$.

We will now show that for any $i\in\mathbb{Z}_l$ and any $B\in\mathcal{I}(X_i^{[k]},(T^l)^{[k]})$ one has $B\in\mathcal{Z}_k(X,T)^{\otimes 2^k}$; this will imply that $\mathcal{Z}_k(X,T^l)\subseteq\mathcal{Z}_k(X,T)$. Put $A_j=(T^{[k]})^{j-i}(B)$, $j\in\mathbb{Z}_l$, and $A=\bigcup_{j\in\mathbb{Z}_l}A_j$. Then $A\in\mathcal{I}(W_1,T^{[k]})\subseteq\mathcal{Z}_k(X,T)^{\otimes 2^k}$. By Lemma 5, $X_i\in\mathcal{Z}_1(X,T)\subseteq\mathcal{Z}_k(X,T)$, thus $(X_i)_{T^l}^{[k]}\in\mathcal{Z}_k(X,T)^{\otimes 2^k}$, and therefore $B=A_i=A\cap(X_i)_{T^l}^{[k]}\in\mathcal{Z}_k(X,T)^{\otimes 2^k}$. ■

We now pass to our second result:

Theorem 8. *For any $k \geq 2$, any $d \in \mathbb{N}$, any linear functions $\varphi_1, \ldots, \varphi_k \colon \mathbb{Z}^d \longrightarrow \mathbb{Z}$ and any Følner sequence $\{\Phi_N\}_{N=1}^{\infty}$ in $\mathbb{Z}^d$, $Z_{k-1}(X,T)$ is a characteristic factor for the averages $\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N} T^{\varphi_1(u)}f_1 \cdot \ldots \cdot T^{\varphi_k(u)}f_k$ in $L^1(X)$, that is,*

$$\lim_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N} T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k - \frac{1}{|\Phi_N|}\sum_{u\in\Phi_N} T^{\varphi_1(u)}E\big(f_1|Z_{k-1}(X,T)\big)\cdot\ldots\cdot T^{\varphi_k(u)}E\big(f_k|Z_{k-1}(X,T)\big)\Big\|_{L^1(X)} = 0 \tag{3}$$

for any $f_1, \ldots, f_k \in L^\infty(X)$.

In order to prove Theorem 8 we will first show that $Z_{k-1}(X,T)$ is a characteristic factor for averages of a very special form. Let us bring more facts from [HK1]. Starting from this moment, we will only be considering real-valued functions on X. Given $f_0, f_1 \in L^\infty(X)$, by the ergodic theorem we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\int_X f_0\cdot T^n f_1 = \int_{I(X,T)} E\big(f_0, I(X,T)\big)\cdot E\big(f_1, I(X,T)\big) = \int_{X_T^{[1]}} f_0\otimes f_1.$$

Applying this twice we get, for $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1} \in L^\infty(X)$,

$$\begin{aligned}
&\lim_{N_2\to\infty}\frac{1}{N_2}\sum_{n_2=1}^{N_2}\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_2}\int_X f_{0,0}\cdot T^{n_1}f_{1,0}\cdot T^{n_2}f_3\cdot T^{n_1+n_2}f_{1,1}\\
&\quad=\lim_{N_2\to\infty}\frac{1}{N_2}\sum_{n_2=1}^{N_2}\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_2}\int_X (f_{0,0}\cdot T^{n_2}f_{0,1})\cdot T^{n_1}(f_{1,0}\cdot T^{n_2}f_{1,1})\\
&\quad=\lim_{N_2\to\infty}\frac{1}{N_2}\sum_{n_2=1}^{N_2}\int_{X^{[1]}}(f_{0,0}\otimes f_{1,0})\cdot T^{n_2}(f_{0,1}\otimes f_{1,1}) = \int_{X^{[2]}}(f_{0,0}\otimes f_{1,0})\otimes(f_{0,1}\otimes f_{1,1}).
\end{aligned}$$

By induction, for any k and any collection $f_{\epsilon_1,\ldots,\epsilon_k}\in L^\infty(X)$, $\epsilon_1,\ldots,\epsilon_k\in\{0,1\}$,

$$\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{n_k=1}^{N_k}\cdots\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\int_X\prod_{\epsilon_1,\ldots,\epsilon_k\in\{0,1\}}T^{\epsilon_1 n_1+\ldots+\epsilon_k n_k}f_{\epsilon_1,\ldots,\epsilon_k} = \int_{X^{[k]}}\bigotimes_{\epsilon_1,\ldots,\epsilon_k\in\{0,1\}}f_{\epsilon_1,\ldots,\epsilon_k}$$

(where the tensor product is taken in a certain order, which we do not specify here).

For $k\in\mathbb{N}$ and $f\in L^\infty(X)$ the seminorm $|||f|||_{T,k}$ associated with T is defined by $|||f|||_{T,k} = \big(\int_{X_T^{[k]}} f^{\otimes 2^k}\big)^{1/2^k}$. Equivalently,

$$|||f|||_{T,k}^{2^k} = \lim_{N_k\to\infty}\frac{1}{N_k}\sum_{n_k=1}^{N_k}\cdots\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\int_X\prod_{\epsilon_1,\ldots,\epsilon_k\in\{0,1\}}T^{\epsilon_1 n_1+\ldots+\epsilon_k n_k}f.$$

It is proved in [HK1] that for any $f_1,\dots,f_{2^k}\in L^\infty(X)$ one has

$$\Big|\int_{X_T^{[k]}}\bigotimes_{j=1}^{2^k} f_j\Big| \le \prod_{j=1}^{2^k} |||f_j|||_{T,k}.$$

For any $k\in\mathbb{N}$ and $f\in L^\infty(X)$ we have

$$|||f|||_{T,k}^{2^k}=\int_{X_T^{[k]}} f^{\otimes 2^k}=\int_{I(X_T^{[k-1]},T^{[k-1]})} E\big(f^{\otimes 2^{k-1}}|I(X_T^{[k-1]},T^{[k-1]})\big)^2.$$

Since $\mathcal{I}(X_T^{[k-1]},T^{[k-1]})\subseteq \mathcal{Z}_{k-1}(X,T)^{\otimes 2^{k-1}}$, $|||f|||_{T,k}=0$ whenever $E\big(f|Z_{k-1}(X,T)\big)=0$.

Proposition 9. *For any $k\ge 2$, nonzero integers $l_1,\dots,l_k$ and a collection $f_{\epsilon_1,\dots,\epsilon_k}\in L^\infty(X)$, $\epsilon_1,\dots,\epsilon_k\in\{0,1\}$, if $E\big(f_{\epsilon_1,\dots,\epsilon_k}|Z_{k-1}(X,T)\big)=0$ for some $\epsilon_1,\dots,\epsilon_k$ then*

$$\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{n_k=1}^{N_k}\cdots\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\int_X\prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l_1 n_1+\dots+\epsilon_k l_k n_k} f_{\epsilon_1,\dots,\epsilon_k}=0.$$

Proof. Let l be a common multiple of $l_1,\dots,l_k$. Since, by Theorem 2, $Z_{k-1}(X,T^l)=Z_{k-1}(X,T)$, $E\big(f_{\epsilon_1,\dots,\epsilon_k}|Z_{k-1}(X,T)\big)=0$ implies $|||f_{\epsilon_1,\dots,\epsilon_k}|||_{T^l,k}=0$.

Let $r_i=l/l_i$, $i=1,\dots,k$. We have

$$\begin{aligned}
&\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{n_k=1}^{N_k}\cdots\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\int_X\prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l_1 n_1+\dots+\epsilon_k l_k n_k} f_{\epsilon_1,\dots,\epsilon_k}\\
&=\frac{1}{r_1\dots r_k}\sum_{m_k=0}^{r_k-1}\cdots\sum_{m_1=0}^{r_1-1}\lim_{N_k\to\infty}\frac{1}{N_k}\sum_{n_k=1}^{N_k}\cdots\lim_{N_1\to\infty}\frac{1}{N_1}\sum_{n_1=1}^{N_1}\\
&\qquad\int_X\prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l n_1+\dots+\epsilon_k l n_k}\big(T^{\epsilon_1 l_1 m_1+\dots+\epsilon_k l_k m_k} f_{\epsilon_1,\dots,\epsilon_k}\big)\\
&=\frac{1}{r_1\dots r_k}\sum_{m_k=0}^{r_k-1}\cdots\sum_{m_1=0}^{r_1-1}\int_{X_{T^l}^{[k]}}\bigotimes_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{\epsilon_1 l_1 m_1+\dots+\epsilon_k l_k m_k} f_{\epsilon_1,\dots,\epsilon_k}.
\end{aligned}$$

And for any $m_{\epsilon_1,\dots,\epsilon_k}\in\mathbb{Z}$, $\epsilon_1,\dots,\epsilon_k\in\{0,1\}$,

$$\begin{aligned}
\Big|\int_{X_{T^l}^{[k]}}\bigotimes_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} T^{m_{\epsilon_1,\dots,\epsilon_k}} f_{\epsilon_1,\dots,\epsilon_k}\Big| &\le \prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} |||T^{m_{\epsilon_1,\dots,\epsilon_k}} f_{\epsilon_1,\dots,\epsilon_k}|||_{T^l,k}\\
&=\prod_{\epsilon_1,\dots,\epsilon_k\in\{0,1\}} |||f_{\epsilon_1,\dots,\epsilon_k}|||_{T^l,k}=0.
\end{aligned}$$

■

Let $\varphi\colon \mathbb{Z}^d \longrightarrow \mathbb{Z}$ be a nonzero linear function, that is, a function of the form $\varphi(n_1,\ldots,n_d) = a_1n_1 + \ldots + a_dn_d$ with $a_1,\ldots,a_d \in \mathbb{Z}$ not all zero. Then for any measure preserving system (Y,S), any $f \in L^1(Y)$ and any Følner sequence $\{\Phi_N\}_{N=1}^\infty$ in $\mathbb{Z}^d$ one has $\lim_{N\to\infty} \frac{1}{|\Phi_N|}\sum_{u\in\Phi_N} S^{\varphi(u)}f = E(f|I(Y,S^l)) = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^N S^{ln}f$, where $l = \gcd(a_1,\ldots,a_d)$. Applying this fact k times, we come to the following generalization of Proposition 9:

Proposition 10. *For any $k \geq 2$, positive integers $d_i \in \mathbb{N}$, nonzero linear functions $\varphi_i\colon \mathbb{Z}^{d_i} \longrightarrow \mathbb{Z}$, Følner sequences $\{\Phi_{i,N}\}_{N=1}^\infty$ in $\mathbb{Z}^{d_i}$, $i = 1,\ldots,k$, and a collection $f_{\epsilon_1,\ldots,\epsilon_k} \in L^\infty(X)$, $\epsilon_1,\ldots,\epsilon_k \in \{0,1\}$, if $E\big(f_{\epsilon_1,\ldots,\epsilon_k}|Z_{k-1}(X,T)\big) = 0$ for some $\epsilon_1,\ldots,\epsilon_k$ then*

$$\lim_{N_k\to\infty}\frac{1}{|\Phi_{k,N_k}|}\sum_{u_k\in\Phi_{k,N_k}}\cdots\lim_{N_1\to\infty}\frac{1}{|\Phi_{1,N_1}|}\sum_{u_1\in\Phi_{1,N_1}} \int_X \prod_{\epsilon_1,\ldots,\epsilon_k\in\{0,1\}} T^{\epsilon_1\varphi_1(u_1)+\ldots+\epsilon_k\varphi_k(u_k)} f_{\epsilon_1,\ldots,\epsilon_k} = 0.$$

The proof of Theorem 8 will be based on the following lemma:

Lemma 11. *For any linear funcions $\varphi_1,\ldots,\varphi_k\colon \mathbb{Z}^d \longrightarrow \mathbb{Z}$ and any $f_1,\ldots,f_k \in L^\infty(X)$,*

$$\begin{aligned}\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N} T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k\Big\|_{L^2(X)} \\ \leq \Big(\lim_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{(v_1,w_1)\in\Phi_{N_1}^2}\lim_{N_k\to\infty}\frac{1}{|\Phi_{N_k}|^2}\sum_{(v_k,w_k)\in\Phi_{N_k}^2}\cdots\lim_{N_2\to\infty}\frac{1}{|\Phi_{N_2}|^2}\sum_{(v_2,w_2)\in\Phi_{N_2}^2} \\ \int_X\prod_{\epsilon_1,\epsilon_2,\ldots,\epsilon_k\in\{0,1\}} T^{\epsilon_1\varphi_1(v_1-w_1)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)}f_1\Big)^{1/2^k} \\ \cdot\prod_{i=2}^k\|f_i\|_{L^\infty(X)}.\end{aligned}$$

Proof. Let $\{\Phi_N\}_{N=1}^\infty$ be a Følner sequence in $\mathbb{Z}^d$. We will use the van der Corput lemma in the following form: if $\{f_u\}_{u\in\mathbb{Z}^d}$ is a bounded family of elements of a Hilbert space, then

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N} f_u\Big\|^2 \leq \limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\langle f_u, f_{u+v-w}\rangle.$$

We may assume that $|f_2|,\dots,|f_k|\le 1$. By the van der Corput lemma we have:

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k\Big\|^2_{L^2(X)}$$
$$\le\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k$$
$$\cdot T^{\varphi_1(u+v-w)}f_1\cdot\ldots\cdot T^{\varphi_k(u+v-w)}f_k$$
$$=\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X T^{\varphi_1(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{\varphi_k(u)}(f_k\cdot T^{\varphi_k(v-w)}f_k)$$
$$=\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}\int_X T^{\varphi_1(u)-\varphi_k(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{\varphi_{k-1}(u)-\varphi_k(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\cdot(f_k\cdot T^{\varphi_k(v-w)}f_k)$$
$$=\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{v,w\in\Phi_{N_1}}\limsup_{N\to\infty}\int_X\Big(\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{(\varphi_1-\varphi_k)(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{(\varphi_{k-1}-\varphi_k)(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\Big)\cdot(f_k\cdot T^{\varphi_k(v-w)}f_k)$$
$$\le\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{(v,w)\in\Phi^2_{N_1}}\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{(\varphi_1-\varphi_k)(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{(\varphi_{k-1}-\varphi_k)(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\Big\|_{L^2(X)}.$$

By the induction hypothesis, applied to the linear functions $\varphi_i-\varphi_k\colon\mathbb{Z}^d\longrightarrow\mathbb{Z}$ and to the functions $f_i\cdot T^{\varphi_i(v-w)}f_i\in L^\infty(X)$, $i=1,\dots,k-1$, for any $(v,w)\in\mathbb{Z}^{2d}$ we have

$$\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{(\varphi_1-\varphi_k)(u)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\cdot\ldots$$
$$\cdot T^{(\varphi_{k-1}-\varphi_k)(u)}(f_{k-1}\cdot T^{\varphi_{k-1}(v-w)}f_{k-1})\Big\|_{L^2(X)}$$
$$\le\Big(\lim_{N_k\to\infty}\frac{1}{|\Phi_{N_k}|^2}\sum_{(v_k,w_k)\in\Phi^2_{N_k}}\cdots\lim_{N_2\to\infty}\frac{1}{|\Phi_{N_2}|^2}\sum_{(v_2,w_2)\in\Phi^2_{N_2}}$$
$$\int_X\prod_{\epsilon_2,\dots,\epsilon_k\in\{0,1\}}T^{\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)}(f_1\cdot T^{\varphi_1(v-w)}f_1)\Big)^{1/2^{k-1}}$$
$$=\Big(\lim_{N_k\to\infty}\frac{1}{|\Phi_{N_k}|^2}\sum_{(v_k,w_k)\in\Phi^2_{N_k}}\cdots\lim_{N_2\to\infty}\frac{1}{|\Phi_{N_2}|^2}\sum_{(v_2,w_2)\in\Phi^2_{N_2}}$$
$$\int_X\prod_{\epsilon_1,\epsilon_2,\dots,\epsilon_k\in\{0,1\}}T^{\epsilon_1\varphi_1(v-w)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)}f_1\Big)^{1/2^{k-1}}.$$

Thus,

$$
\begin{aligned}
&\limsup_{N\to\infty}\Big\|\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k\Big\|_{L^2(X)}\\
&\le\Big(\limsup_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{(v,w)\in\Phi_{N_1}^2}\Big(\lim_{N_k\to\infty}\frac{1}{|\Phi_{N_k}|^2}\sum_{(v_k,w_k)\in\Phi_{N_k}^2}\cdots\lim_{N_2\to\infty}\frac{1}{|\Phi_{N_2}|^2}\sum_{(v_2,w_2)\in\Phi_{N_2}^2}\\
&\qquad\int_X\prod_{\epsilon_1,\epsilon_2,\ldots,\epsilon_k\in\{0,1\}}T^{\epsilon_1\varphi_1(v-w)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)}f_1\Big)^{1/2^{k-1}}\Big)^{1/2}\\
&\le\Big(\lim_{N_1\to\infty}\frac{1}{|\Phi_{N_1}|^2}\sum_{(v,w)\in\Phi_{N_1}^2}\lim_{N_k\to\infty}\frac{1}{|\Phi_{N_k}|^2}\sum_{(v_k,w_k)\in\Phi_{N_k}^2}\cdots\lim_{N_2\to\infty}\frac{1}{|\Phi_{N_2}|^2}\sum_{(v_2,w_2)\in\Phi_{N_2}^2}\\
&\qquad\int_X\prod_{\epsilon_1,\epsilon_2,\ldots,\epsilon_k\in\{0,1\}}T^{\epsilon_1\varphi_1(v-w)+\epsilon_2(\varphi_1-\varphi_2)(v_2-w_2)+\ldots+\epsilon_k(\varphi_1-\varphi_k)(v_k-w_k)}f_1\Big)^{1/2^{k}}.
\end{aligned}
$$

■

Proof of Theorem 8. Because of the multilinearity of (3), it suffices to show that $\lim_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k=0$ in $L^1(X)$ whenever $E\big(f_1|Z_{k-1}(X,T)\big)=0$. We may assume that the functions $\varphi_1,\ldots,\varphi_k$ are all nonzero and distinct. Then, combining Lemma 11 and Proposition 10, applied to the nonzero linear functions $\varphi_1(v-w)$, $(\varphi_1-\varphi_2)(v-w)$, …, $(\varphi_1-\varphi_k)(v-w)$ on Z^{2d} and the Følner sequence $\{\Phi_N^2\}_{N=1}^\infty$ in $\mathbb{Z}^{2d}$, we get $\lim_{N\to\infty}\frac{1}{|\Phi_N|}\sum_{u\in\Phi_N}T^{\varphi_1(u)}f_1\cdot\ldots\cdot T^{\varphi_k(u)}f_k=0$ in $L^2(X)$ and so, in $L^1(X)$. ■

Bibliography

[CL1] J.P. Conze and E. Lesigne, Théorèmes ergodiques pour des mesures diagonales, *Bull. Soc. Math. France* **112** (1984), no. 2, 143–175.

[CL2] J.P. Conze and E. Lesigne, Sun un théorème ergodique pour des mesures diagonales, *Probabilités*, 1–31, Publ. Inst. Rech. Math. Rennes, 1987–1, Univ. Rennes I, Rennes, 1988.

[CL3] J.P. Conze and E. Lesigne, Sun un théorème ergodique pour des mesures diagonales, *C. R. Acad. Sci. Paris Sér. I Math.* **306** (1988), no. 12, 491–493.

[F] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. d'Analyse Math.* **31** (1977), 204–256.

[HK1] B. Host and B. Kra, Non-conventional ergodic averages and nilmanifolds, to appear in *Annals of Math.*

[HK2] B. Host and B. Kra, Convergence of polynomial ergodic averages, to appear in *Israel J. of Math.*

[Z] T. Ziegler, Universal characteristic factors and Furstenberg averages, preprint.