A set of 2-recurrence whose perfect squares do not form a set of measurable recurrence

JOHN T. GRIESMER

Department of Applied Mathematics and Statistics, Colorado School of Mines,
Golden, Colorado, USA
(e-mail: jtgriesmer@gmail.com)

Abstract. We say that $S \subseteq \mathbb{Z}$ is a set of k-recurrence if for every measure-preserving transformation T of a probability measure space (X, μ) and every $A \subseteq X$ with $\mu(A) > 0$, there is an $n \in S$ such that $\mu(A \cap T^{-n}A \cap T^{-2n} \cap \cdots \cap T^{-kn}A) > 0$. A set of 1-recurrence is called a set of measurable recurrence. Answering a question of Frantzikinakis, Lesigne, and Wierdl [Sets of k-recurrence but not $(k+1)$-recurrence. *Ann. Inst. Fourier (Grenoble)* **56**(4) (2006), 839–849], we construct a set of 2-recurrence S with the property that $\{n^2 : n \in S\}$ is not a set of measurable recurrence.

1. *Background and motivation*

A *probability measure-preserving system* (or *MPS*) is a quadruple $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a probability measure space and $T : X \to X$ is an invertible transformation preserving μ, meaning $\mu(T^{-1}A) = \mu(A)$ for every measurable set $A \subseteq X$.

We say that $S \subseteq \mathbb{Z}$ is a *set of measurable recurrence* if for every MPS $(X, \mathcal{B}, \mu, T)$ and every $A \subseteq X$ having $\mu(A) > 0$, there is an $n \in S$ such that $\mu(A \cap T^{-n}A) > 0$.

For a fixed $k \in \mathbb{N}$, we say S is a *set of k-recurrence* if under these hypotheses, there is an $n \in S$ such that $\mu(\bigcap_{j=0}^{k} T^{-jn}A) > 0$; in this terminology, a set of measurable recurrence is a set of 1-recurrence.

Finally, $S \subseteq \mathbb{Z}$ is a *set of Bohr recurrence* if for all $d \in \mathbb{N}$, every $\boldsymbol{\alpha} \in \mathbb{T}^d$, and all $\varepsilon > 0$, there is an $n \in S$ such that $\|n\boldsymbol{\alpha}\| < \varepsilon$ (see §3 for definitions and notation).

Frantzikinakis, Lesigne, and Wierdl [10] proved that if $k \in \mathbb{N}$ and $S \subseteq \mathbb{Z}$ is a set of k-recurrence, then $S^{\wedge k} := \{n^k : n \in S\}$ is a set of Bohr recurrence. They ask (the remarks following [10, Proposition 2.2]) whether this conclusion can be strengthened to '$S^{\wedge k}$ is a set of measurable recurrence,' and the subsequent articles [7, 8] reiterate ([8, Problem 5] of the current version at arXiv:1103.3808) this question. Our main result, Theorem 1.1, provides a negative answer for the case $k = 2$. For $k \geq 3$, the question remains open. A related question in [7] asks whether a set S which is a set of k-recurrence for every k

must have the property that $S^{\wedge 2}$ is a set of measurable recurrence. We discuss how our construction relates to these questions in §16.

THEOREM 1.1. *There is a set $S \subseteq \mathbb{Z}$ which is a set of 2-recurrence such that $S^{\wedge 2}$ is not a set of measurable recurrence.*

Reflecting on the known examples of sets of Bohr recurrence which are not sets of measurable recurrence, Frantzikinakis [8] predicts that an example of a set of 2-recurrence S where $S^{\wedge 2}$ is not a set of measurable recurrence will be rather complicated. Our example is indeed complicated: while built from well-known constituents using standard methods, the proof that it is a set of 2-recurrence uses several reductions—from general measure-preserving systems to totally ergodic systems to nilsystems to affine systems to Kronecker systems. The final reduction combines explicit computations of multiple ergodic averages in 2-step affine systems with classical estimates for three term arithmetic progressions in terms of Fourier coefficients.

1.1. *Outline of the article.* Our approach is similar to Kriz's construction [18] proving that there is a set of topological recurrence which is not a set of measurable recurrence. Very roughly, our example S in Theorem 1.1 is $\{n : n^2 \in R\}$, where R is Kriz's example. While this description is not quite correct, it may help those familiar with [18], [16] or [15] understand our construction.

The overall proof of Theorem 1.1 is presented at the end of §2. We outline its components here. Section 2 begins by collecting standard facts about the following finite approximations to recurrence properties.

Definition 1.2. Let $S \subseteq \mathbb{Z}$ and $k \in \mathbb{N}$. We say that S is (δ, k)-*recurrent* if for every MPS $(X, \mathcal{B}, \mu, T)$ and every $A \subseteq X$ with $\mu(A) > \delta$, we have $A \cap T^{-n}A \cap \cdots \cap T^{-kn}A \neq \varnothing$ for some $n \in S$.

We say that S is (δ, k)-*non-recurrent* if there is an MPS $(X, \mathcal{B}, \mu, T)$ and $A \subseteq X$ with $\mu(A) > \delta$ such that $A \cap T^{-n}A \cap \cdots \cap T^{-kn}A = \varnothing$.

We say S is δ-*non-recurrent* if it is $(\delta, 1)$-non-recurrent, meaning there is an MPS $(X, \mathcal{B}, \mu, T)$ and $A \subseteq X$ with $\mu(A) > \delta$ such that $A \cap T^{-n}A = \varnothing$ for all $n \in S$.

Remark 1.3. The condition $A \cap T^{-n}A \cap \cdots \cap T^{-kn}A \neq \varnothing$ in the definition of (δ, k)-recurrent may be replaced with $\mu(A \cap T^{-n}A \cap \cdots \cap T^{-kn}A) > 0$; cf. Lemma 15.1.

Lemma 2.1 says that if $S_1, S_2 \subseteq \mathbb{Z}$ are finite, δ_1-non-recurrent, and δ_2-non-recurrent, then for all sufficiently large m, $S_1 \cup mS_2$ is $2\delta_1\delta_2$-non-recurrent. Thus, if $S_1^{\wedge 2}$ and $S_2^{\wedge 2}$ are δ_1-non-recurrent and δ_2-non-recurrent, respectively, then $(S_1 \cup mS_2)^{\wedge 2}$ is $2\delta_1\delta_2$-non-recurrent for all sufficiently large m, as $(S_1 \cup mS_2)^{\wedge 2} = S_1^{\wedge 2} \cup m^2 S_2^{\wedge 2}$.

Lemma 2.3 says that $S \subseteq \mathbb{Z}$ is δ-non-recurrent if and only if for all $\delta' < \delta$ and all *finite* subsets $S' \subseteq S$, S' is δ'-non-recurrent. Likewise, if $S \subseteq \mathbb{Z}$ is $(\eta, 2)$-recurrent, then for all $\eta' < \eta$, there is a finite subset $S' \subseteq S$ which is $(\eta', 2)$-recurrent.

The proof of Theorem 1.1 is given at the end of §2; it explains in detail how finite approximations are assembled to form a 2-recurrent set whose perfect squares do not

form a set of measurable recurrence. This reduces the problem to proving Lemma 2.4, which states that the required finite approximations exist. These approximations are based on *Bohr–Hamming balls*, which we introduce in §3. Bohr–Hamming balls were used in [15, 18] to construct sets with prescribed recurrence properties. Fixing $\delta < \frac{1}{2}$ and $\eta > 0$, Lemmas 3.4 and 3.5 show that there is a Bohr–Hamming ball BH which is δ-non-recurrent, while $\sqrt{BH} := \{n \in \mathbb{N} : n^2 \in BH\}$ is $(\eta, 2)$-recurrent.

The proof of Lemma 3.5 occupies §§4–15. It is proved by estimating multiple ergodic averages of the form

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(n^2 \boldsymbol{\beta}) \int f \cdot f \circ T^n \cdot f \circ T^{2n} \, d\mu, \qquad (1.1)$$

where $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system, $f : X \to [0, 1]$ has $\int f \, d\mu > \delta$ for some prescribed $\delta > 0$, $\boldsymbol{\beta} \in \mathbb{T}^r$ for some $r \in \mathbb{N}$, and $g : \mathbb{T}^r \to [0, 1]$ is Riemann integrable. Under certain hypotheses on g, we will prove the limit in equation (1.1) is positive; this is the inequality in equation (4.7) in the proof of Lemma 3.5. In §4, we show how the general case may be reduced to that where T is totally ergodic. The remainder of the article, outlined in §5.2, is dedicated to analyzing the limit in equation (1.1) when T is totally ergodic. Section 8 shows that the totally ergodic case can be further reduced to the study of *standard* 2-*step Weyl systems*, and §§9–13 are dedicated to simplifying and estimating equation (1.1) for these systems.

Readers familiar with the theory of characteristic factors (especially [6]) may find it most profitable to read §§2, 3, 5, and 8 in detail, and skim §4.

2. *Constructing the example from finite approximations*
We first require some standard facts about the properties mentioned in Definition 1.2. The following is [16, Lemma 3.6]; it is essentially [18, Lemma 3.2]. Similar lemmas appear, often unnamed, in the variations on Kriz's example [5, 20, 21, 25].

LEMMA 2.1. *Let S_1, $S_2 \subseteq \mathbb{N}$ be finite. If S_1 and S_2 are δ-non-recurrent and η-non-recurrent, respectively, then for all sufficiently large $m \in \mathbb{N}$, $S_1 \cup mS_2$ is $2\delta\eta$-non-recurrent.*

LEMMA 2.2. *Let $m \in \mathbb{Z}$ and $\delta \geq 0$. If $S \subseteq \mathbb{Z}$ is $(\delta, 2)$-recurrent, then mS is also $(\delta, 2)$-recurrent.*

Proof. Fix $m \in \mathbb{Z}$ and let $S \subseteq \mathbb{Z}$ be a $(\delta, 2)$-recurrent set. Let $(X, \mathcal{B}, \mu, T)$ be an MPS, with $A \subseteq X$ having $\mu(A) > \delta$. Consider the MPS $(X, \mathcal{B}, \mu, T^m)$. Since $\mu(A) > \delta$, there exists $n \in S$ such that $\mu(A \cap (T^m)^{-n} A \cap (T^m)^{-2n} A) > 0$, meaning $\mu(A \cap T^{-mn} A \cap T^{-2(mn)} A) > 0$. Since $mn \in mS$, this proves mS is $(\delta, 2)$-recurrent. $\square$

Our proof of Lemma 2.4 uses the following compactness properties for recurrence.

LEMMA 2.3. *Let $k \in \mathbb{N}$ and $\delta \geq 0$. If $\delta' > \delta$ and every finite subset of S is (δ', k)-non-recurrent, then S is (δ, k)-non-recurrent.*

Consequently, if S is (δ, k)-recurrent, then for all $\delta' > \delta$, there is a finite $S' \subseteq S$ which is (δ', k)-recurrent.

We prove Lemma 2.3 in §15. A special case, which is easily adapted to prove the general case, appears in [**5**, Ch. 2].

Theorem 1.1 is proved by combining the following lemma with the others in this section.

LEMMA 2.4. *For all $\delta > 0$ and $\eta < 1/2$, there exists $S \subseteq \mathbb{Z}$ which is $(\delta, 2)$-recurrent such that $S^{\wedge 2}$ is η-non-recurrent.*

By Lemma 2.3, we can take S to be finite in Lemma 2.4.

Lemmas 3.4 and 3.5 will prove Lemma 2.4; the proof of Lemma 3.5 forms the majority of this article.

Proof of Theorem 1.1. Let $\delta < \delta' < \frac{1}{2}$. We will construct an increasing sequence of finite sets $S_1 \subseteq S_2 \subseteq \cdots$ so that S_n is $(1/n, 2)$-recurrent, and $S_n^{\wedge 2}$ is δ'-non-recurrent. Setting $S := \bigcup_{n=1}^{\infty} S_n$, we get that S is a set of 2-recurrence, while every finite subset of $S^{\wedge 2}$ is δ'-non-recurrent. Lemma 2.3 then implies S is δ-non-recurrent.

To define S_1, we apply Lemma 2.4 to find an $S_1 \subseteq \mathbb{Z}$ which is $(1, 2)$-recurrent, while $S_1^{\wedge 2}$ is δ_1-non-recurrent for some $\delta_1 > \delta'$. We define the remaining S_n inductively: suppose $n \in \mathbb{N}$ and that S_n has been chosen to be $(1/n, 2)$-recurrent, while $S_n^{\wedge 2}$ is δ_n-non-recurrent for some $\delta_n > \delta'$. Let $\eta < \frac{1}{2}$ so that $2\eta\delta_n > \delta'$. We will find $S_{n+1} \supset S_n$ so that S_{n+1} is $(1/(n + 1), 2)$-recurrent and $S_{n+1}^{\wedge 2}$ is $2\eta\delta_n$-non-recurrent. To do so, apply Lemma 2.4 to find a finite $R \subseteq \mathbb{Z}$ which is $(1/(n + 1), 2)$-recurrent such that $R^{\wedge 2}$ is η-non-recurrent. By Lemma 2.1, choose $m \in \mathbb{N}$ so that $(S_n^{\wedge 2}) \cup m^2(R^{\wedge 2})$ is $2\eta\delta_n$-non-recurrent. Now $S_{n+1} := S_n \cup mR$ is the desired set: mR is $(1/(n + 1), 2)$-recurrent, by Lemma 2.2, while $S_{n+1}^{\wedge 2} = (S_n^{\wedge 2}) \cup m^2(R^{\wedge 2})$. Since $2\eta\delta_n > \delta'$, this completes the inductive step of the construction. $\qquad\square$

3. *Approximate Hamming balls in $\mathbb{T}^r$ and Bohr–Hamming balls in $\mathbb{Z}$*

Let $\mathbb{T}$ denote the group $\mathbb{R}/\mathbb{Z}$ with the usual topology. For $x \in \mathbb{T}$, let $\tilde{x}$ denote the unique element of $[0, 1)$ such that $x = \tilde{x} + \mathbb{Z}$ and define $\|x\| := \min\{|\tilde{x} - n| : n \in \mathbb{Z}\}$. For $r \in \mathbb{N}$ and $\mathbf{x} = (x_1, \ldots, x_r) \in \mathbb{T}^r$, let $\|\mathbf{x}\| := \max_{j \leq r} \|x_j\|$.

For $\varepsilon > 0$, $r \in \mathbb{N}$, and $\mathbf{x} = (x_1, \ldots, x_r) \in \mathbb{T}^r$, let

$$w_\varepsilon(\mathbf{x}) := |\{j : \|x_j\| \geq \varepsilon\}|.$$

So $w_\varepsilon(\mathbf{x})$ is the number of coordinates of $\mathbf{x}$ differing from 0 by at least ε.

Definition 3.1. For $k < r \in \mathbb{N}$, $\mathbf{y} \in \mathbb{T}^r$, and $\varepsilon > 0$, we define the *approximate Hamming ball* of radius (k, ε) around $\mathbf{y}$ as

$$\text{Hamm}(\mathbf{y}; k, \varepsilon) := \{\mathbf{x} \in \mathbb{T}^r : w_\varepsilon(\mathbf{y} - \mathbf{x}) \leq k\}.$$

So $\text{Hamm}(\mathbf{y}; k, \varepsilon)$ is the set of $\mathbf{x} = (x_1, \ldots, x_r) \in \mathbb{T}^r$, where at most k coordinates x_i differ from y_i by at least ε.

If Z is a topological abelian group, we say $\alpha \in Z$ *generates* Z if the cyclic subgroup $\{n\alpha : n \in \mathbb{Z}\}$ is dense in Z. In other words, α generates Z if Z is the smallest closed subgroup containing α.

The *group rotation system* $(Z, \mathcal{B}, m_Z, R_\alpha)$, where $\mathcal{B}$ is the Borel σ-algebra on Z and m_Z is Haar measure on Z, is given by $R_\alpha z = z + \alpha$.

Definition 3.2. If $U = \text{Hamm}(\mathbf{y}; k, \varepsilon) \subseteq \mathbb{T}^r$ is an approximate Hamming ball and $\boldsymbol{\beta} \in \mathbb{T}^r$, the corresponding *Bohr–Hamming ball* of *radius* (k, ε) is

$$BH(\boldsymbol{\beta}, \mathbf{y}; k, \varepsilon) := \{n \in \mathbb{Z} : n\boldsymbol{\beta} \in U\}.$$

If $\boldsymbol{\beta}$ generates $\mathbb{T}^r$, we say that the corresponding Bohr–Hamming ball is *proper*.

We write m for Haar probability measure on $\mathbb{T}^r$. Lemmas 3.3 and 3.4 here are implicit in [18] and proved explicitly in [15].

LEMMA 3.3. *Let $k \in \mathbb{N}$ and $\eta < \frac{1}{2}$. For all sufficiently large $r \in \mathbb{N}$, there is an $\varepsilon > 0$ and $E \subseteq \mathbb{T}^r$ with $m(E) > \eta$ such that $E \cap (E + U) = \varnothing$, where $U = \text{Hamm}(\mathbf{y}; k, \varepsilon)$, with $\mathbf{y} = (\frac{1}{2}, \ldots, \frac{1}{2}) \in \mathbb{T}^r$.*

Lemma 3.3 is a consequence of [15, Lemma 7.1]. To derive the former from the latter, note that [15, Lemma 7.1] (in the case $p = 2$ there) provides sets $E, E' \subseteq \mathbb{T}^r$ with $\mu(E) > \eta$, an approximate Hamming ball U around $0_{\mathbb{T}^r}$ with radius (k, ε) for some $\varepsilon > 0$, such that $E + U \subseteq E'$ and $E' + (\frac{1}{2}, \ldots, \frac{1}{2})$ is disjoint from E'.

LEMMA 3.4. *Let $k \in \mathbb{N}$ and $\eta < \frac{1}{2}$. For all sufficiently large $r \in \mathbb{N}$, there is an $\varepsilon > 0$ such that for all $\boldsymbol{\beta} \in \mathbb{T}^r$, the Bohr–Hamming ball $BH(\boldsymbol{\beta}, \mathbf{y}; k, \varepsilon)$ is η-non-recurrent, where $\mathbf{y} = (\frac{1}{2}, \ldots, \frac{1}{2}) \in \mathbb{T}^r$.*

Proof. Let $\eta < \frac{1}{2}$ and choose r large enough to find the E and U provided by Lemma 3.3, with $m(E) > \eta$. Let $(X, \mathcal{B}, \mu, T) = (\mathbb{T}^r, \mathcal{B}, m, R_{\boldsymbol{\beta}})$ be the group rotation on $\mathbb{T}^r$ determined by $\boldsymbol{\beta}$. For $n \in BH(\boldsymbol{\beta}, \mathbf{y}; k, \eta)$, we have $R_{\boldsymbol{\beta}}^n E \subseteq E + U$, so $E \cap R_{\boldsymbol{\beta}}^n E = \varnothing$. Since $R_{\boldsymbol{\beta}}$ is invertible, this means $E \cap R_{\boldsymbol{\beta}}^{-n} E = \varnothing$, as well. $\square$

For $S \subseteq \mathbb{Z}$, let $\sqrt{S} := \{n \in \mathbb{Z} : n^2 \in S\}$.

LEMMA 3.5. *For all $\delta > 0$, there exists $k_0 \in \mathbb{N}$ such that for every $r \in \mathbb{N}$, every proper Bohr–Hamming ball $BH := BH(\boldsymbol{\beta}, \mathbf{y}; k, \varepsilon)$ with $k \geq k_0$, $\varepsilon > 0$, and $\mathbf{y} \in \mathbb{T}^r$, $\sqrt{BH}$ is $(\delta, 2)$-recurrent.*

Lemma 3.5 is proved using multiple ergodic averages and characteristic factors. The main argument is given in §4, using several reductions developed in §§4–14.

Proof of Lemma 2.4. Let $\delta > 0$ and $\eta < \frac{1}{2}$. Choose k large enough to satisfy the conclusion of Lemma 3.5. With this k, choose $r > k$ and ε small enough to satisfy the conclusion of Lemma 3.4. Let $\boldsymbol{\beta} \in \mathbb{T}^r$ be generating and let $BH = BH(\boldsymbol{\beta}, \mathbf{y}; k, \varepsilon)$, where $\mathbf{y} = (\frac{1}{2}, \ldots, \frac{1}{2}) \in \mathbb{T}^r$, so that BH is η-non-recurrent. Finally, let $S = \sqrt{BH}$, so that S is $(\delta, 2)$-recurrent, by Lemma 3.5. Since $S^{\wedge 2} \subseteq BH$, we get that $S^{\wedge 2}$ is η-non-recurrent, as desired. $\square$

3.1. *Cylinders and Fourier coefficients.* Here we define constituents of approximate Hamming balls.

Definition 3.6. Given $r \in \mathbb{N}$, $I \subseteq \{1, \ldots, r\}$, $\eta > 0$, and $\mathbf{y} \in \mathbb{T}^r$, define the η-*cylinder determined by I around* $\mathbf{y}$ to be

$$V_{I,\mathbf{y},\eta} := \{\mathbf{x} \in \mathbb{T}^r : \|x_i - y_i\| < \eta \text{ for all } i \in I\},$$

so that

$$U := \mathrm{Hamm}(\mathbf{y}; k, \eta) = \bigcup_{\substack{I \subseteq \{1,\ldots,r\} \\ |I| = r-k}} V_{I,\mathbf{y},\eta}. \tag{3.1}$$

We say that $g : \mathbb{T} \to \mathbb{R}$ is a *cylinder function subordinate to U* if $g = m(V)^{-1} 1_V$, where V is one of the cylinders $V_{I,\mathbf{y},\eta}$ in equation (3.1). Note that each cylinder function subordinate to U is supported on U.

Let $\mathcal{S}^1$ denote the circle group $\{z \in \mathbb{C} : |z| = 1\}$ with the usual topology and the group operation of complex multiplication. If Z is a compact abelian group with Haar probability measure m, $\widehat{Z}$ denotes its Pontryagin dual, meaning $\widehat{Z}$ is the group of continuous homomorphisms $\chi : Z \to \mathcal{S}^1$; such homomorphisms are called *characters* of Z. Given $f : Z \to \mathbb{C}$, its *Fourier transform* is $\hat{f} : \widehat{Z} \to \mathbb{C}$ given by $\hat{f}(\chi) = \int f \overline{\chi} \, dm$.

For $s \in Z$, let f_s be the translate of f defined by $f_s(x) := f(x + s)$. Then $\widehat{f_s}(\chi) = \chi(s)\hat{f}(\chi)$ for each $\chi \in \widehat{Z}$.

As usual, for $f, g : Z \to \mathbb{C}$, $f * g$ denotes their *convolution*, defined as $f * g(x) := \int f(t)g(x - t) \, dm(t)$. We will use the standard identity $\widehat{f * g} = \hat{f}\hat{g}$ (the Fourier transform turns convolution into pointwise multiplication).

Letting $\|f\| := (\int |f|^2 \, dm)^{1/2}$ denote the $L^2(m)$ norm of f, we have the standard Plancherel identity in equation (3.2), which leads to the subsequent lemma,

$$\sum_{\chi \in \widehat{Z}} |\hat{f}(\chi)|^2 = \|f\|^2. \tag{3.2}$$

LEMMA 3.7. *Let Z be a compact abelian group with Haar probability measure m and $f \in L^2(m)$. If $\|f\| \leq 1$ and $|\hat{f}(\chi_1)|, \ldots, |\hat{f}(\chi_k)|$ are the k largest values of $|\hat{f}|$, then $|\hat{f}(\chi)| < k^{-1/2}$ for all $\chi \in \widehat{Z} \setminus \{\chi_1, \ldots, \chi_k\}$.*

Proof. Let $S_1 = \{\chi_1, \ldots, \chi_k\}$ be the set of characters attaining the k largest values of $|\hat{f}|$, let $S_2 = \widehat{Z} \setminus S_1$, and let $c = \max\{|\hat{f}(\chi)| : \chi \in S_2\}$. By definition, we have $|\hat{f}(\chi)| \geq c$ for all $\chi \in S_1$.

We split the left-hand side of equation (3.2) into sums over $\chi \in S_1$ and $\chi \in S_2$, then subtract the sum over S_1 to get

$$\sum_{\chi \in S_2} |\hat{f}(\chi)|^2 = \|f\|^2 - \sum_{\chi \in S_1} |\hat{f}(\chi)|^2.$$

Since $|\hat{f}(\chi)| \geq c$ for all $\chi \in S_1$, the right-hand side is bounded above by $\|f\|^2 - kc^2$. Since $c \leq |\hat{f}(\chi)|$ for at least one $\chi \in S_2$, the left-hand side above is bounded below by c^2. So

$$c^2 \le \sum_{\chi \in S_2} |\hat{f}(\chi)|^2 = \|f\|^2 - \sum_{\chi \in S_1} |\hat{f}(\chi)|^2 \le 1 - kc^2,$$

which implies $c^2 \le 1 - kc^2$. Solving, we get $c \le (1+k)^{-1/2}$. This means $|\hat{f}(\chi)| < k^{-1/2}$ for all $\chi \in S_2$. $\qquad\square$

Remark 3.8. The exact form of the inequality in Lemma 3.7 is not important; we only need $\sup_{\chi \in \widehat{Z} \setminus \{\chi_1, \ldots, \chi_k\}} |\hat{f}(\chi)| \le c(k)$, where $c(k) \to 0$ as $k \to \infty$, uniformly for $\|f\| \le 1$.

Much of the proof of Lemma 3.5 is contained in Lemma 3.9. The actual application requires a technical generalization (Lemma 12.2).

LEMMA 3.9. *Fix $k < r \in \mathbb{N}$, and let $U \subseteq \mathbb{T}^r$ be an approximate Hamming ball of radius (k, η) with $\eta > 0$.*

(i) *Let $\chi_1, \ldots, \chi_k \in \widehat{\mathbb{T}}^r$ be non-trivial. Then there is a cylinder function g subordinate to U such that for all $s \in \mathbb{T}^r$, we have*

$$\widehat{g_s}(\chi_j) = 0 \quad \text{for each } j \le k.$$

(ii) *If $f \in L^2(m_{\mathbb{T}^r})$ with $\|f\| \le 1$, there is a cylinder function g subordinate to U so that*

$$|\widehat{f * g}(\chi)| < k^{-1/2} \quad \text{for all } \chi \in \widehat{\mathbb{T}}^d.$$

Proof. (i) Assuming k, r, χ_j, and U are as in the statement, we may write χ_j as

$$\chi_j(x_1, \ldots, x_r) = e\bigg(\sum_{l=1}^{r} n_l^{(j)} x_l \bigg), \tag{3.3}$$

where $e(t) := \exp(2\pi i t)$ and $n_l^{(j)} \in \mathbb{Z}$. Non-triviality of χ_j means that for each j, at least one of the $n_l^{(j)}$ is non-zero. So choose one such index l_j for each $j \le k$ and let $I = \{1, \ldots, r\} \setminus \{l_1, \ldots, l_k\}$. In case some of the l_j repeat, remove additional elements from I so that $|I| = r - k$.

Writing U as Hamm$(\mathbf{y}; k, \eta)$, let $V = V_{I, \mathbf{y}, \eta} = \{\mathbf{x} \in \mathbb{T}^d : \|y_l - x_l\| < \eta \text{ for all } l \in I\}$, so that $V \subseteq U$. Let $g := m(V)^{-1} 1_V$, so that g is a cylinder function subordinate to U, and let $j \le k$. To prove that $\hat{g}(\chi_j) = 0$, note that g does not depend on any of the coordinates x_{l_j}, so we can simplify the right-hand side of equation (3.3) as $e(\sum_{\substack{l=1 \\ l \ne l_j}}^{r} n_l^{(j)} x_l) e(n_{l_j}^{(j)} x_{l_j})$ and write $\hat{g}(\chi_j) = \int g \overline{\chi}_j \, dm$ as

$$\int_{\mathbb{T}^{r-1}} g(x_1, \ldots, x_r) e\bigg(-\sum_{\substack{l=1 \\ l \ne l_j}}^{r} n_l^{(j)} x_l \bigg) dx_1 \ldots dx_{l_j-1} \, dx_{l_j+1} \ldots dx_r \int_{\mathbb{T}} e(-n_{l_j}^{(j)} x_{l_j}) \, dx_{l_j}.$$

Since $\int e(-n_{l_j}^{(j)} x_{l_j}) \, dx_{l_j} = 0$, we conclude that $\hat{g}(\chi_j) = 0$ for each j. To complete the proof of part (i), we observe that $\widehat{g_s}(\chi) = \chi(s) \hat{g}(\chi)$ for each χ.

To prove part (ii), assume $f : \mathbb{T}^r \to \mathbb{C}$ has $\|f\| \le 1$, and let $|\hat{f}(\chi_1)|, \ldots, |\hat{f}(\chi_k)|$ be the k largest values of $|\hat{f}|$. By part (i), choose a cylinder function g subordinate to U so that $\hat{g}(\chi_j) = 0$ for these χ_j. Then $|\hat{f}(\chi)| < k^{-1/2}$ for all other χ, by

Lemma 3.7. Note that $|\hat{g}(\chi)| \leq 1$ for all $\chi \in \widehat{\mathbb{T}^d}$, since $\int |g|\, dm = 1$. We therefore have $\widehat{f * g}(\chi_j) = \hat{f}(\chi_j)\hat{g}(\chi_j) = 0$ for $j = 1, \ldots, k$, while $|\widehat{f * g}(\chi)| \leq |\hat{f}(\chi)| < k^{-1/2}$ for all other χ. $\square$

4. *Multiple ergodic averages*

Some of our reductions use facts from the general theory of nilsystems, mainly contained in [**6**, **9**]. Readers who want a general introduction to the theory can consult [**17**].

If $(X, \mathcal{B}, \mu, T)$ is an MPS and f is a bounded function on X, let

$$L_3(f, T) := \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f \cdot T^n f \cdot T^{2n} f \, d\mu.$$

The existence of this limit was established in [**11**, §**3**].

In this section, we prove Lemma 3.5 using Lemma 4.4, which estimates variants of $L_3(f, T)$. In §5.1, we state a more convenient form of Lemma 4.4 and outline its proof.

We will use the following known result, which follows by combining a special case of [**3**, Theorem 2.1] with the multidimensional Szemerédi theorem [**13**].

THEOREM 4.1. *For all $\delta > 0$, there exists $c(\delta) > 0$ such that for every MPS $(X, \mathcal{B}, \mu, T)$ and every $f : X \to [0, 1]$ with $\int f \, d\mu > \delta$, we have*

$$L_3(f, T) > c(\delta). \tag{4.1}$$

Definition 4.2. We say that $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ is *ergodic* if $\mu(A \triangle T^{-1}A) = 0$ implies $\mu(A) = 0$ or $\mu(A) = 1$ for every $A \in \mathcal{B}$. We say that $\mathbf{X}$ is *totally ergodic* if for every $m \in \mathbb{N}$, the system $(X, \mathcal{B}, \mu, T^m)$ is ergodic.

Remark 4.3. When determining whether a set is a set of k-recurrence, we may restrict our attention to *ergodic* MPSs where μ is a regular Borel measure on a compact metric space X; cf. [**4**, §§7.2.2 and 7.2.3].

When we say a sequence $(b_n)_{n \in \mathbb{N}}$ of natural numbers has *linear growth*, we mean that it is strictly increasing and $\limsup_{n \to \infty} b_n/n < \infty$. Note that a strictly increasing sequence has linear growth if and only if the set of terms $B = \{b_n : n \in \mathbb{N}\}$ satisfies $\underline{d}(B) := \liminf_{n \to \infty}(|B \cap \{1, \ldots, n\}|/n) > 0$. Enumerating the positive elements of $\sqrt{BH}$ in increasing order, where BH is a proper Bohr–Hamming ball always results in a sequence of linear growth. To see this, write $\sqrt{BH}$ as $\{n \in \mathbb{Z} : n^2 \boldsymbol{\beta} \in U\}$ for some approximate Hamming ball $U \subseteq \mathbb{T}^r$ and generator $\boldsymbol{\beta} \in \mathbb{T}^r$. Then,

$$\lim_{N \to \infty} \frac{|\sqrt{BH} \cap [1, \ldots N]|}{N} = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} 1_U(n^2 \boldsymbol{\beta}) = m(U),$$

by Weyl's theorem on uniform distribution of polynomials (see Lemma 10.3). Since $n = |\sqrt{BH} \cap [1, \ldots, b_n]|$, this implies b_n/n is bounded. Likewise, if g is a cylinder function subordinate to U (Definition 3.6), then enumerating $\{n \in \mathbb{N} : g(n^2 \boldsymbol{\beta}) > 0\}$ in increasing order results in a sequence of linear growth.

The next lemma says that $L_3(f, T)$ can be approximated by averaging over elements of $\sqrt{BH}$, provided $\mathbf{X}$ is *totally ergodic* and BH is a proper Bohr–Hamming ball of radius (k, η) with k sufficiently large. In passing to the general case, we need to consider $\sqrt{BH}/\ell := \{n \in \mathbb{Z} : \ell n \in \sqrt{BH}\}$.

LEMMA 4.4. *For all $\varepsilon > 0$, there is a $k \in \mathbb{N}$ such that for every totally ergodic MPS $(X, \mathcal{B}, \mu, T)$, every $f : X \to [0, 1]$, every proper Bohr–Hamming ball BH of radius (k, η) ($\eta > 0$), and all $\ell \in \mathbb{N}$, there is a sequence $b_n \in \sqrt{BH}/\ell$ having linear growth such that*

$$\lim_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} \int f \cdot T^{b_n} f \cdot T^{2b_n} f \, d\mu - L_3(f, T) \right| < \varepsilon \|f\|^2. \tag{4.2}$$

Consequently, if $\int f \, d\mu > \delta$ and k is sufficiently large (depending only on δ), we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f \cdot T^{b_n} f \cdot T^{2b_n} f \, d\mu > c(\delta)/2, \tag{4.3}$$

where $c(\delta)$ is defined in Theorem 4.1.

Lemma 5.1 is a convenient reformulation of Lemma 4.4. In §5.1, we outline its proof, which occupies the majority of this article.

Remark 4.5. We do not know whether the condition 'totally ergodic' can be replaced with 'ergodic' in Lemma 4.4. The main obstruction to this replacement is our lack of a convenient representation of ergodic, but not totally ergodic, 2-step affine nilsystems.

4.1. *Factors and extensions.* If $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ and $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ are MPSs, we say that $\mathbf{Y}$ is a *factor* of $\mathbf{X}$ if there is a measurable $\pi : X \to Y$ intertwining S and T, meaning

$$\pi(Tx) = S\pi(x) \quad \text{for } \mu\text{-almost every (a.e.) } x \in X,$$

and $\mu(\pi^{-1}D) = \nu(D)$ for all $D \in \mathcal{D}$. Strictly speaking, the factor is the *pair* $(\pi, \mathbf{Y})$, and we refer to 'the factor $\pi : \mathbf{X} \to \mathbf{Y}$'.

If $\pi : \mathbf{X} \to \mathbf{Y}$ is a factor and $f \in L^2(\mu)$ is equal μ-almost everywhere to a function of the form $g \circ \pi$, with $g \in L^2(\nu)$, we say that f is $\mathbf{Y}$-*measurable*. This is equivalent to saying that f is $\pi^{-1}(\mathcal{D})$-measurable (modulo μ). We denote by $P_{\mathbf{Y}}$ the orthogonal projection from $L^2(\mu)$ to the space of $\pi^{-1}(\mathcal{D})$-measurable functions. Given $f \in L^2(\mu)$, we identify $P_{\mathbf{Y}} f$ with $\tilde{f} \in L^2(\nu)$ satisfying $P_{\mathbf{Y}} f = \tilde{f} \circ \pi$.

We repeatedly use, without comment, the fact that $P_{\mathbf{Y}}$ is a positive operator preserving integration with respect to μ. In other words, if $f(x) \geq 0$ for μ-a.e. x, then $P_{\mathbf{Y}} f(x) \geq 0$ for μ-a.e. x, and $\int f \, d\mu = \int P_{\mathbf{Y}} f \, d\mu$. Consequently, $\sup f \geq \tilde{f}(y) \geq \inf f$ for ν-a.e. y and $\int \tilde{f} \, d\nu = \int f \, d\mu$.

Remark 4.6. When $\pi : \mathbf{X} \to \mathbf{Y}$ is a factor, we say that $\mathbf{X}$ is an *extension* of $\mathbf{Y}$. If we wish to prove an inequality on ergodic averages for a system $\mathbf{Y}$, it suffices to prove that inequality for an extension $\pi : \mathbf{X} \to \mathbf{Y}$, since the integrals $\int f_0 \cdot S^a f_1 \cdot S^b f_2 \, d\nu$ can be written as $\int h_0 \cdot T^a h_1 \cdot T^b h_2 \, d\mu$, where $h_i = f_i \circ \pi$. This observation will be used in §14.

4.2. *Reducing to total ergodicity.* The next lemma is used to deduce Lemma 3.5 from Lemma 4.4 and Theorem 4.1. Part (i) is a special case of [2, Corollary 4.6], and part (ii) is an immediate consequence of part (i). Here '**Y** is an inverse limit of ergodic nilsystems' means that for all $f \in L^{\infty}(\nu)$ and $\varepsilon > 0$, there is a factor $\pi : \mathbf{Y} \to \mathbf{Z}$, where $\mathbf{Z} = (Z, \mathcal{Z}, \eta, R)$ is an ergodic nilsystem and $\| f - P_{\mathbf{Z}} f \|_{L^1(\nu)} < \varepsilon$.

LEMMA 4.7. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be an ergodic measure-preserving system. There is a factor* $\pi : \mathbf{X} \to \mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ *which is an inverse limit of ergodic nilsystems such that:*
  (i) *for all* $f_i \in L^{\infty}(\mu)$, *letting* $\tilde{f}_i \circ \pi = P_{\mathbf{Y}} f_i$, *we have*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \left| \int f_0 \cdot T^n f_1 \cdot T^{2n} f_2 \, d\mu - \int \tilde{f}_0 \cdot S^n \tilde{f}_1 \cdot S^{2n} \tilde{f}_2 \, d\nu \right| = 0;$$

  (ii) *if* $(b_n)_{n \in \mathbb{N}}$ *is a sequence of linear growth, then*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \left| \int f_0 \cdot T^{b_n} f_1 \cdot T^{2b_n} f_2 \, d\mu - \int \tilde{f}_0 \cdot S^{b_n} \tilde{f}_1 \cdot S^{2b_n} \tilde{f}_2 \, d\nu \right| = 0.$$

To derive part (ii) from part (i), note that

$$\frac{1}{N} \sum_{n=1}^{N} \left| \int f \cdot T^{b_n} f \cdot T^{2b_n} f \, d\mu - \int \tilde{f} \cdot S^{b_n} \tilde{f} \cdot S^{2b_n} \tilde{f} \, d\nu \right|$$

$$\leq \frac{b_N}{N} \cdot \frac{1}{b_N} \sum_{n=1}^{b_N} \left| \int f_0 \cdot T^n f_1 \cdot T^{2n} f_2 \, d\mu - \int \tilde{f}_0 \cdot S^n \tilde{f}_1 \cdot S^{2n} \tilde{f}_2 \, d\nu \right|$$

$$\xrightarrow[N \to \infty]{} 0,$$

since b_N/N is bounded.

We get the next result by combining the definition of 'inverse limit' with the fact that for every ergodic nilsystem $(Y, \mathcal{D}, \nu, S)$, there is an $\ell \in \mathbb{N}$ such that the ergodic components of $(Y, \mathcal{D}, \nu, S^{\ell})$ are totally ergodic; see [6, Proposition 2.1] for justification.

LEMMA 4.8. *If* $(X, \mathcal{B}, \mu, T)$ *is an inverse limit of ergodic nilsystems,* $f : X \to [0, 1]$, *and* $\varepsilon > 0$, *there is a factor* $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ *and* $\ell \in \mathbb{N}$ *such that:*
  (i) $\| f - P_{\mathbf{Y}} f \|_{L^1(\mu)} < \varepsilon$;
  (ii) *the ergodic components of* $(Y, \mathcal{D}, \nu, S^{\ell})$ *are totally ergodic.*

Notation 4.9. When Y is the phase space of an ergodic nilsystem where $(Y, \mathcal{D}, \nu, S^{\ell})$ is totally ergodic, we will enumerate its connected components as $Y_1, \ldots, Y_M$, and write $\nu_i := (1/M)\nu|_{Y_i}$. Each $\mathbf{Y}_i := (Y_i, \mathcal{D}_i, \nu_i, S^{\ell})$ is an ergodic component of $(Y, \mathcal{D}, \nu_Y, S^{\ell})$. If $\mathbf{X}$ is an extension of $\mathbf{Y}$ with factor map $\pi : \mathbf{X} \to \mathbf{Y}$, we let $X_i = \pi^{-1}(Y_i)$, $\mu_i := (1/M)\mu|_{X_i}$, $\mathcal{B}_i := \{B \cap X_i : B \in \mathcal{B}\}$, and $\mathbf{X}_i = (X_i, \mathcal{B}_i, \mu_i, T^{\ell})$. It is easy to verify that $\mathbf{Y}_i$ is a factor of $\mathbf{X}_i$ with factor map $\pi|_{X_i}$.

Remark 4.10. Here we identify a technical difficulty common in multiple recurrence arguments. Readers familiar with the use of Markov's inequality to overcome this difficulty may skip to the proof of Lemma 3.5.

Our proof of Lemma 3.5 starts with an ergodic, but not totally ergodic, MPS $\mathbf{X} = (X, \mathcal{B}, \mu, T)$. By Lemma 4.7, it suffices to prove the lemma in the special case where $\mathbf{X}$ is an inverse limit of ergodic nilsystems, so we assume $\mathbf{X}$ is such an inverse limit. We then consider $f : X \to [0, 1]$ with $\int f \, d\mu > \delta$. The goal is to find an $\ell \in \mathbb{N}$ and a sequence (b_n) of elements of $\sqrt{BH}/\ell$ satisfying equation (4.7). The main difficulty arises when trying to exploit the structure of nilsystems: Lemma 4.4 requires total ergodicity, so we fix $\varepsilon > 0$ and choose a factor $\pi : \mathbf{X} \to \mathbf{Y}$ where $\mathbf{Y}$ is an ergodic nilsystem satisfying parts (i) and (ii) in Lemma 4.8. We choose ℓ so that the ergodic components of $(Y, \mathcal{D}, \nu, S^\ell)$ are totally ergodic, and we enumerate these components as $\mathbf{Y}_i = (Y_i, \mathcal{D}_i, \nu_i, S^\ell)$, $i = 1, \ldots, M$. With Notation 4.9 defined above, let $\tilde{f} \circ \pi = P_{\mathbf{Y}} f$ and $\tilde{f}_i = \tilde{f}|_{Y_i}$. Lemma 4.4 allows us to choose, *for each ergodic component* $\mathbf{Y}_i$ *where* $\int \tilde{f}_i \, d\nu_i > \delta/2$, a sequence $b_n^{(i)} \in \sqrt{BH}/\ell$ having linear growth, such that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int \tilde{f}_i \cdot S^{\ell b_n^{(i)}} \tilde{f}_i \cdot S^{2\ell b_n^{(i)}} \tilde{f}_i \, d\nu_i > c(\delta/2)/2. \tag{4.4}$$

The choice of $b_n^{(i)}$ depends on $\mathbf{Y}_i$, so equation (4.4) implies only that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int \tilde{f} \cdot S^{\ell b_n^{(i)}} \tilde{f} \cdot S^{2\ell b_n^{(i)}} \tilde{f} \, d\nu > \frac{1}{M} \frac{c(\delta/2)}{2}. \tag{4.5}$$

If M is large, then $\|f - P_{\mathbf{Y}} f\|_{L^1(\mu)}$ may be large compared with $(1/M)c(\delta/2)/2$, and equation (4.5) will not immediately imply equation (4.7). To overcome this obstacle, we want to find an i where equation (4.4) holds and $(1/M)\|f_i - P_{\mathbf{Y}} f_i\|_{L^1(\mu)}$ is sufficiently small to make $\int \tilde{f}_i \cdot S^{\ell a} \tilde{f}_i \cdot S^{\ell b} \tilde{f}_i \, d\nu_i$ close to $\int f_i \cdot T^{\ell a} f_i \cdot T^{\ell b} f_i \, d\mu_i$ for all a, b. Such an i is provided by two straightforward applications of Markov's inequality outlined in §15.3.

Before proving Lemma 3.5, we recall its statement: for all $\delta > 0$, there is a $k_0 \in \mathbb{N}$ such that for every proper Bohr–Hamming ball $BH := BH(\boldsymbol{\beta}, \mathbf{y}; k, \eta)$ with $k \geq k_0$, $\eta > 0$, and $\mathbf{y} \in \mathbb{T}^r$, $\sqrt{BH}$ is $(\delta, 2)$-recurrent.

Proof of Lemma 3.5, assuming Lemma 4.4. Let $\delta > 0$ and choose $k_0 \in \mathbb{N}$ so that for all $k \geq k_0$, the inequality in equation (4.3) holds in Lemma 4.4 with $c(\delta/2)$ in place of $c(\delta)$. Let BH be a proper Bohr–Hamming ball with radius (k, η) for some $\eta > 0$. It suffices to prove that for every MPS $(X, \mathcal{B}, \mu, T)$ with $A \subseteq X$ having $\mu(A) > \delta$,

$$\mu(A \cap T^{-n} A \cap T^{-2n} A) > 0 \quad \text{for some } n \in \sqrt{BH}. \tag{4.6}$$

By Remark 4.3, we need only consider *ergodic* MPSs. We will prove that if $\mathbf{X}$ is ergodic and $f : X \to [0, 1]$ has $\int f \, d\mu > \delta$, then there is a sequence of elements $b_n \in \sqrt{BH}$ with linear growth such that

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f \cdot T^{b_n} f \cdot T^{2b_n} f \, d\mu > 0. \tag{4.7}$$

The special case of equation (4.7) where $f = 1_A$ implies equation (4.6), as the integral then simplifies to $\mu(A \cap T^{-b_n} A \cap T^{-2b_n} A)$.

By part (ii) of Lemma 4.7, it suffices to prove equation (4.7) when $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ is an inverse limit of ergodic nilsystems. We now fix such an $\mathbf{X}$, and $f : X \to [0, 1]$ with $\int f \, d\mu > \delta$.

Let $\varepsilon = (\delta/24)c(\delta/2)$, and let $\pi : \mathbf{X} \to \mathbf{Y}$ be the factor provided by Lemma 4.8 for this ε, with $\ell \in \mathbb{N}$ chosen so that the ergodic components $\mathbf{Y}_i$ of $(Y, \mathcal{D}, \nu, S^\ell)$ are totally ergodic. Let M be the number of ergodic components (we can take $\ell = M$, but we do not need this fact) so that $\mu(Y_i) = 1/M$ for each i.

Let $X_i = \pi^{-1}(Y_i)$ and let $f_i = 1_{X_i} f$, so that the X_i partition X into sets of measure $1/M$, and $\sum_i \int f_i \, d\mu = \int f \, d\mu > \delta$. Observe that $P_{\mathbf{Y}} f_i$ is supported on X_i and $\int P_{\mathbf{Y}} f_i \, d\mu = \int f_i \, d\mu$ for each i.

Setting $\mathbf{Y}_i := (Y_i, \mathcal{D}_i, \nu_i, S^\ell)$, where $\nu_i := M\nu|_{Y_i}$, we get that $\mathbf{Y}_i$ is a totally ergodic MPS. Likewise, $\mathbf{X}_i := (X_i, \mathcal{B}_i, \mu_i, T^\ell)$, with $\mu_i := M\mu|_{X_i}$ is an MPS (possibly not ergodic), with $\pi|_{X_i} : X_i \to Y_i$ a factor map. To prove equation (4.7), we will find a sequence b_n of elements of $\sqrt{BH}/\ell$ having linear growth and $i \leq M$ with

$$\liminf_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f_i \cdot T^{\ell b_n} f_i \cdot T^{2\ell b_n} f_i \, d\mu > 0. \tag{4.8}$$

We claim that there is an i such that

$$\int f_i \, d\mu > \frac{\delta}{2M} \quad \text{and} \tag{4.9}$$

$$\| f_i - P_{\mathbf{Y}} f_i \|_{L^1(\mu)} < \frac{c(\delta/2)}{12M}. \tag{4.10}$$

This i is provided by Lemmas 15.5 and 15.6: setting

$$I := \left\{ i : \int f_i \, d\mu > \frac{\delta}{2M} \right\}, \quad J := \left\{ i : \int |f_i - P_{\mathbf{Y}} f_i| \, d\mu < \frac{c(\delta/2)}{12M} \right\},$$

we get $|I| > M\delta/2$ and $|J| > M(1 - 12\varepsilon/c(\delta/2)) = M(1 - \delta/2)$. Thus $|I| + |J| > M$, implying $I \cap J$ is non-empty.

Fix i satisfying inequalities (4.9) and (4.10). Note that inequality (4.9) and the definition of ν_i, μ_i, and $\tilde{f}_i$ imply

$$\int \tilde{f}_i \, d\nu_i > \delta/2. \tag{4.11}$$

Since $(Y_i, \mathcal{B}_i, \nu_i, S^\ell)$ is totally ergodic, we may apply Lemma 4.4 to choose a sequence of elements $b_n \in \sqrt{BH}/\ell$ having linear growth and satisfying

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int \tilde{f}_i \cdot S^{\ell b_n} \tilde{f}_i \cdot S^{2\ell b_n} \tilde{f}_i \, d\nu_i > c(\delta/2)/2. \tag{4.12}$$

Inequality (4.10), the bounds $\| f_i \|_\infty \leq 1$, $\| P_{\mathbf{Y}} f_i \|_\infty \leq 1$, and Lemma 15.7 imply

$$\left| \int f_i \cdot T^{\ell a} f_i \cdot T^{\ell b} f_i \, d\mu_i - \int P_{\mathbf{Y}_i} f_i \cdot T^{\ell a} P_{\mathbf{Y}_i} f_i \cdot T^{\ell b} P_{\mathbf{Y}_i} f_i \, d\mu_i \right| < \frac{1}{4} c(\delta/2) \tag{4.13}$$

for each $a, b \in \mathbb{N}$. Recalling the definition of μ_i and ν_i, we see that for all sufficiently large N,

$$\frac{1}{N} \sum_{n=1}^{N} \int f_i \cdot T^{\ell b_n} f_i \cdot T^{2\ell b_n} f_i \, d\mu$$

$$> \frac{1}{N} \sum_{n=1}^{N} \int \tilde{f_i} \cdot S^{\ell b_n} f_i \cdot S^{2\ell b_n} f_i \, d\nu - \frac{c(\delta/2)}{4M} \quad \text{by inequality (4.13)}$$

$$> \frac{c(\delta/2)}{2M} - \frac{c(\delta/2)}{4M} \quad\quad\quad\quad\quad\quad \text{by inequality (4.12)}$$

$$= \frac{c(\delta/2)}{4M}.$$

The above inequalities imply equation (4.8). Since $f \geq f_i$ pointwise and we chose $b_n \in \sqrt{BH}/\ell$, this implies equation (4.7) and completes the proof of Lemma 3.5. $\qquad\square$

5. *Reformulation of Lemma 4.4*

5.1. *Reformulation.* Lemma 4.4 is an immediate consequence of the following reformulation. This version allows us to apply the theory of characteristic factors.

LEMMA 5.1. *Let $k < r \in \mathbb{N}$, $\ell \in \mathbb{N}$, let $\boldsymbol{\beta} \in \mathbb{T}^r$ be generating, and let $U \subseteq \mathbb{T}^r$ be an approximate Hamming ball of radius (k, η) for some $\eta > 0$. For every totally ergodic MPS $(X, \mathcal{B}, \mu, T)$, and every measurable $f : X \to [0, 1]$, there is a cylinder function $g = m(V)^{-1} 1_V$ subordinate to U such that*

$$\lim_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} g(n^2 \ell^2 \boldsymbol{\beta}) \int f \cdot T^n f \cdot T^{2n} f \, d\mu - L_3(f, T) \right| < 2k^{-1/2} \|f\|^2. \quad (5.1)$$

While U does not depend on f in Lemma 5.1, the choice of g to satisfy equation (5.1) does depend on f.

We prove Lemma 5.1 in §14. The derivation of Lemma 4.4 from Lemma 5.1 is an instance of the following general principle: if a_n is a bounded sequence, $B \subseteq \mathbb{N}$ is enumerated as $\{b_1 < b_2 < \ldots\}$, and $d(B) := \lim_{N \to \infty} (|B \cap \{1, \ldots, N\}|/N) > 0$, then

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} a_{b_n} = \lim_{N \to \infty} \frac{1}{N d(B)} \sum_{n=1}^{N} 1_B(n) a_n$$

provided the limit on the right exists. Note that $(b_n)_{n \in \mathbb{N}}$ has linear growth if $d(B) > 0$.

We will apply this principle with $a_n = \int f \cdot T^n f \cdot T^{2n} f \, d\mu$ and $B = \{n : n^2 \ell^2 \boldsymbol{\beta} \in V\}$, where V is a cylinder contained in U. Then, $g = m(V)^{-1} 1_V$ is a cylinder function subordinate to U, and $g(n^2 \ell^2 \boldsymbol{\beta}) = d(B)^{-1} 1_B(n)$. The equation $d(B) = m(V)$ follows from Weyl's theorem on uniform distribution (cf. §10). Note that this B is contained in $\sqrt{BH}/\ell$, where BH is the Bohr–Hamming ball corresponding to U, with frequency $\boldsymbol{\beta}$.

Remark 5.2. The exact form of the bound in equation (5.1) is not important in the following. The only relevant property is that the coefficient of $\|f\|^2$ tends to 0 as $k \to \infty$.

5.2. *Outline of a special case of Lemma 15.1.* This outline highlights the key steps in our proof while avoiding some complications.

We begin with an arbitrary *totally ergodic* measure-preserving system $\mathbf{X} = (X, \mathcal{B}, \mu, T)$, $f : X \to [0, 1]$, and $k \in \mathbb{N}$. We let $r > k$, $\eta > 0$, and fix an approximate Hamming ball $U = \mathrm{Hamm}(\mathbf{y}; k, \eta) \subseteq \mathbb{T}^r$ and a generator $\boldsymbol{\beta} \in \mathbb{T}^r$. We want to find a cylinder function g subordinate to U so that

$$A_N(f, g) := \frac{1}{N} \sum_{n=1}^{N} g(n^2 \boldsymbol{\beta}) \int f \cdot T^n f \cdot T^{2n} f \, d\mu \tag{5.2}$$

satisfies $\lim_{N \to \infty} |A_N(f, g) - L_3(f, T)| < 2k^{-1/2} \|f\|^2$.

In §§6–8, we will reduce to the case where $\mathbf{X}$ is a *standard 2-step Weyl system*. This means that $(X, \mathcal{B}, \mu, T)$ can be realized with $X = \mathbb{T}^d \times \mathbb{T}^d$, $d \in \mathbb{N}$, $\mu = $ Haar probability measure on $\mathbb{T}^d \times \mathbb{T}^d$, and T is given by $T(x, y) = (x + \boldsymbol{\alpha}, y + x)$, for some generator $\boldsymbol{\alpha} \in \mathbb{T}^d$. The orbits of T can be computed explicitly: $T^n(x, y) = (x + n\boldsymbol{\alpha}, y + nx + \binom{n}{2}\boldsymbol{\alpha})$. This reduction relies on the theory of characteristic factors, especially [6, Theorem B].

To simplify this outline, we assume $r = d$ and $\boldsymbol{\beta} = \boldsymbol{\alpha}$. We write functions on $\mathbb{T}^d \times \mathbb{T}^d$ with variables displayed as $f(x, y)$, where $x, y \in \mathbb{T}^d$. Writing $m \times m$ for Haar probability measure on $\mathbb{T}^d \times \mathbb{T}^d$, we write $\int f \, dm \times m$ as $\int f(x, y) \, dx \, dy$, or $\int f\binom{x}{y} \, dx \, dy$ to save space. With these assumptions, the averages in equation (5.2) become

$$B_N := \frac{1}{N} \sum_{n=1}^{N} g(n^2\boldsymbol{\alpha}) \int_{\mathbb{T}^d \times \mathbb{T}^d} f\binom{x}{y} f\binom{x + n\boldsymbol{\alpha}}{y + nx + \binom{n}{2}\boldsymbol{\alpha}} f\binom{x + 2n\boldsymbol{\alpha}}{y + 2nx + \binom{2n}{2}\boldsymbol{\alpha}} \, dx \, dy.$$

Proposition 13.1 provides an explicit formula for $\lim_{N \to \infty} B_N$. Under the present assumptions, it says

$$\lim_{N \to \infty} B_N$$
$$= \int_{(\mathbb{T}^d)^4} f(x, y) f(x + s, y + t) \left(\int_{\mathbb{T}^d} f(x + 2s, y + 2t + 2w) g(w) \, dw \right) ds \, dt \, dx \, dy. \tag{5.3}$$

Write I for the right-hand side above, and define

$$f *_2 g(x, y) := \int f(x, y + 2w) g(w) \, dw.$$

Using Lemma 12.2 (a generalization of Lemma 3.9), we choose a cylinder function g subordinate to U such that $|\widehat{f *_2 g}(\chi, \psi)| < k^{-1/2}$ for all $(\chi, \psi) \in \widehat{\mathbb{T}}^d \times \widehat{\mathbb{T}}^d$ with ψ non-trivial. We set $f'(x) := \int f(x, y) \, dy$ and $J' := \int \int f'(x) f'(x + s) f'(x + 2s) \, dx \, ds$. By Lemma 11.1, the bound on $\widehat{f *_2 g}$ will imply

$$|I - J'| < k^{-1/2} \|f\|^2. \tag{5.4}$$

We can also prove (directly, or using Theorem 7.1) that $L_3(f, T) = J'$. Combining equation (5.4) with equation (5.3), we then have equation (5.1), completing the outline of this special case. The factor 2 on the right-hand side of equation (5.1) accounts for the reduction to Weyl systems.

In the general case, we must compute $\lim_{N\to\infty} A_N(f, g)$ for $d \neq r$ and $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$. The integral $\int f(x + 2s, y + 2t + 2w)g(w)\,dw$ in equation (5.3) will then be replaced by an integral over an *affine joining* of $\mathbb{T}^d$ with $\mathbb{T}^r$ (Definition 9.4), but the computation in this case is not substantially different from the outline above.

5.3. *Iterated integral notation.* When all variables are displayed and there is no chance of confusion, we may omit all but one of the integral signs and the subscripts indicating the domain of integration. So the integral in equation (5.3) may be written as

$$\int f(x, y)f(x + s, y + t)f(x + 2s, y + 2t + 2w)g(w)\,dw\,ds\,dt\,dx\,dy.$$

6. *Eigenvalues and ergodicity of products*

An *eigenfunction* of an MPS $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ with *eigenvalue* $\lambda \in \mathbb{C}$ is an $f \in L^2(\mu)$ satisfying $\|f\| \neq 0$ and $f \circ T = \lambda f$. Since $\int |f \circ T|\,d\mu = \int |f|\,d\mu$, we have $|\lambda| = 1$. We then have that $|f \circ T|$ is T-invariant, so if $\mathbf{X}$ is ergodic, we get that $|f|$ is equal μ-almost everywhere to a constant. We say an eigenvalue λ of $\mathbf{X}$ is *non-trivial* if $\lambda \neq 1$. Note that the eigenfunctions of $\mathbf{X}$ are the eigenvectors of the unitary operator $U_T : L^2(\mu) \to L^2(\mu)$ defined by $U_T f = f \circ T$.

Given two MPSs $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ and $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$, we form the product system $\mathbf{X} \times \mathbf{Y} = (X \times Y, \mathcal{B} \otimes \mathcal{D}, \mu \times \nu, T \times S)$. For $f \in L^2(\mu)$ and $g \in L^2(\nu)$, we write $f \otimes g$ for the function defined by $f \otimes g(x, y) = f(x)g(y)$.

We need some standard consequences of the following, which is the specialization of [**12**, Lemma 4.17, p. 91] to the case where $\mathcal{H} = L^2(\mu)$, $\mathcal{H}' = L^2(\nu)$ for MPSs $\mathbf{X}$ and $\mathbf{Y}$ as above, with unitary operators $Uf := f \circ T$ and $U'g := g \circ S$.

LEMMA 6.1. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be measure-preserving systems as above, and let* $\mathbf{X} \times \mathbf{Y}$ *be the product system. Let* $h \in L^2(\mu \times \nu)$ *be an eigenfunction of* $\mathbf{X} \times \mathbf{Y}$ *with eigenvalue* λ, *meaning* $h \circ (T \times S) = \lambda h$. *Then* $h = \sum c_n f_n \otimes g_n$, *where* $f_n \circ T = \lambda_n f_n$, $g_n \circ S = \lambda'_n g_n$, $\lambda_n \lambda'_n = \lambda$, *and the sequences* $\{f_n\}$, $\{g_n\}$ *are orthonormal in* $L^2(\mu)$ *and* $L^2(\nu)$, *respectively.*

To deduce Lemma 6.1 from of [**12**, Lemma 4.17], note that if μ and ν are measure spaces, $L^2(\mu \times \nu)$ is isomorphic to the tensor product $L^2(\mu) \otimes L^2(\nu)$, and the obvious isomorphism identifies $U_{T\times S}$ with $U_T \otimes U_S$.

The next lemma is a well-known consequence of Lemma 6.1; we omit its proof.

LEMMA 6.2. *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are ergodic MPSs, the product system* $\mathbf{X} \times \mathbf{Y}$ *is ergodic if and only if* $\mathbf{X}$ *and* $\mathbf{Y}$ *have no non-trivial eigenvalues in common.*

Another immediate consequence of Lemma 6.1 is the following lemma.

LEMMA 6.3. *If* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *and* $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ *are MPSs such that* $\mathbf{X} \times \mathbf{Y}$ *is ergodic and* $g \in L^2(\nu)$ *is orthogonal to every eigenfunction of* $\mathbf{Y}$, *then for every* $f \in L^2(\mu)$, $f \otimes g$ *is orthogonal to every eigenfunction of* $\mathbf{X} \times \mathbf{Y}$.

7. Eigenfunctions and the Kronecker factor

Every ergodic MPS $\mathbf{X}$ has a factor $\pi : \mathbf{X} \to \mathbf{Z}$ where $\mathbf{Z} = (Z, \mathcal{Z}, m, R)$ is a compact abelian group rotation such that every eigenfunction of $\mathbf{X}$ is $\pi^{-1}(\mathcal{Z})$-measurable. This factor is called the *Kronecker factor* of $\mathbf{X}$, and we write $\int_Z f(s) \, ds$ (or sometimes just $\int f(s) \, ds$) to abbreviate $\int f(s) \, dm(s)$.

The following result is proved in [11, §3]; we use the notation L_3 introduced in §4.

THEOREM 7.1. *If $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ is an ergodic MPS with Kronecker factor $\pi : \mathbf{X} \to \mathbf{Z}$, $\mathbf{Z} = (Z, \mathcal{Z}, m, R)$, and $f_i : X \to [0, 1]$, then*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f_1(T^n x) f_2(T^{2n} x) = \int_Z \tilde{f}_1(\pi(x) + s) \tilde{f}_2(\pi(x) + 2s) \, ds, \quad (in \ L^2(\mu)),$$

where $\tilde{f}_i \in L^\infty(m)$ satisfies $\tilde{f}_i \circ \pi = P_{\mathbf{Z}} f_i$. Furthermore,

$$L_3(f, T) = \int_Z \int_Z \tilde{f}(z) \tilde{f}(z + s) \tilde{f}(z + 2s) \, dz \, ds \quad for \ all \ f \in L^\infty(\mu). \tag{7.1}$$

7.1. Kronecker factor of a standard 2-step Weyl system.

A *standard 2-step Weyl system* is an MPS of the form $\mathbf{Y} = (Y, \mathcal{B}, m, S)$, where $Y = \mathbb{T}^d \times \mathbb{T}^d$, $d \in \mathbb{N}$, and $S : Y \to Y$ is defined as $S(x, y) = (x + \alpha, y + x)$, for some fixed $\alpha = (\alpha_1, \ldots, \alpha_d)$ generating $\mathbb{T}^d$. There is an explicit formula for the orbits of S:

$$S^n(x, y) = (x + n\alpha, y + nx + \binom{n}{2}\alpha), \tag{7.2}$$

which may be verified by induction. Ergodicity of $\mathbf{Y}$ is equivalent to α generating $\mathbb{T}^d$. For $d = 1$, this follows from [12, Proposition 3.11, p. 67], and the general case follows from a nearly identical proof. Also explained in [12] is the Kronecker factor of $\mathbf{Y}$: the eigenfunctions of $\mathbf{Y}$ are exactly the functions χ on Y defined by

$$\chi((x_1, \ldots, x_d), (y_1, \ldots, y_d)) := \exp(2\pi i (n_1 x_1 + \cdots + n_d x_d))$$

for some $n_j \in \mathbb{Z}$, so the group of eigenvalues of $\mathbf{Y}$ is $\{\exp(2\pi i (n_1 \alpha_1 + \cdots + n_d \alpha_d)) : n_j \in \mathbb{Z}\}$. Thus, the Kronecker factor of $\mathbf{Y}$ is obtained by setting $Z = \mathbb{T}^d$ and letting $\pi : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{T}^d$ be a projection onto the first coordinate. Since the span of the eigenfunctions of $\mathbf{Y}$ consists solely of those functions depending on the first coordinate, the orthogonal projection $P_{\mathbf{Z}} f(x, y)$ can be written as $(P_{\mathbf{Z}} f)(x, y) := \int f(x, y) \, dy$. Combining this with Theorem 7.1, we have the following observation.

Observation 7.2. The Kronecker factor $(Z, \mathcal{Z}, m, R)$ of a standard 2-step Weyl system $(\mathbb{T}^d \times \mathbb{T}^d, \mathcal{D}, \mu, S)$ is spanned by functions of the form $f(x, y) = g(x)$ (i.e. functions depending on only the first coordinate), and for all bounded $f : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{C}$, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \int f \cdot S^n f \cdot S^{2n} f \, d\mu = \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} f'(x) f'(x + s) f'(x + 2s) \, dx \, ds,$$

where $f' : \mathbb{T}^d \to \mathbb{C}$ is defined as $f'(x) := \int f(x, y) \, dy$.

8. *Reduction to Weyl systems*

The next lemma is one key step in the proof of Lemma 5.1. Its proof is similar to the proof of [**1**, Lemma 8.1].

LEMMA 8.1. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be a totally ergodic MPS and* $f : X \to [0, 1]$. *For all* $\varepsilon > 0$, *there is a factor* $\pi : \mathbf{X} \to \mathbf{Y}$ *such that:*

(i) $\mathbf{Y}$ *is a factor of a standard 2-step Weyl system;*

(ii) *setting* $\tilde{f} \circ \pi = P_{\mathbf{Y}} f$, *we have*

$$\lim_{N \to \infty} \left| \frac{1}{N} \sum_{n=1}^{N} g(n^2 \boldsymbol{\beta}) \int f \cdot T^n f \cdot T^{2n} f \, d\mu - g(n^2 \boldsymbol{\beta}) \int \tilde{f} \cdot S^n \tilde{f} \cdot S^{2n} \tilde{f} \, d\nu \right| < \varepsilon$$

for every continuous $g : \mathbb{T}^r \to [0, 1]$ *and every* $\boldsymbol{\beta} \in \mathbb{T}^r$, *for all* $r \in \mathbb{N}$.

If we assume $\boldsymbol{\beta}$ *generates* $\mathbb{T}^r$, *then item (ii) holds for every Riemann integrable* $g : \mathbb{T}^r \to [0, 1]$.

We prove Lemma 8.1 at the end of this section. Most of the proof is contained in the next lemma, an application of [**6**, Theorem B]. It concerns the *maximal 2-step affine factor* $\mathbf{A}_2$ of an ergodic MPS $\mathbf{X}$; see [**6**] for discussion and exposition. Additionally, we use the standard fact that the Kronecker factor of $\mathbf{X}$ is a factor of $\mathbf{A}_2$.

If $\mathbf{X}$ is an MPS, we write $\mathcal{E}(\mathbf{X})$ for the group of eigenvalues of $\mathbf{X}$ (see §6). We continue to write $e(t)$ for $\exp(2\pi i t)$, and we use the notation $P_{\mathbf{Y}}$ introduced in §4.1.

LEMMA 8.2. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be an ergodic measure-preserving system with maximal 2-step affine factor* $\mathbf{A}_2$ *and let* $\beta \in [0, 1)$. *Then* $\mathbf{A}_2$ *is characteristic for the averages*

$$B_N(f_1, f_2) := \frac{1}{N} \sum_{n=1}^{N} e(n^2 \beta) \cdot T^n f_1 \cdot T^{2n} f_2, \tag{8.1}$$

meaning

$$\lim_{N \to \infty} B_N(f_1, f_2) = \lim_{N \to \infty} B_N(P_{\mathbf{A}_2} f_1, P_{\mathbf{A}_2} f_2) \tag{8.2}$$

in $L^2(\mu)$, *for all bounded* f_1, f_2. *Furthermore, if* β *is irrational and*

$$\mathcal{E}(\mathbf{X}) \cap \{e(n\beta)\}_{n \in \mathbb{Z}} = \{1\}, \tag{8.3}$$

then $\lim_{N \to \infty} B_N(f_1, f_2) = 0$ *in* $L^2(\mu)$ *for all bounded measurable* f_i.

Remark 8.3. The existence of $\lim_{N \to \infty} B_N(f_1, f_2)$ is not immediately obvious, but the proof of Lemma 8.2 will show that it is a special case of the existence of limits of polynomial multiple ergodic averages found in [**6**].

Proof. We first dispense with the case where β is rational. In this case, the sequence $e(n^2 \beta)$ is periodic, so we fix a period $p \in \mathbb{N}$ such that $e((pn + q)^2 \beta) = e(q^2 \beta)$ for every n and $q \in \mathbb{N}$. For $0 \le r \le p$ and $N \in \mathbb{N}$, we can then write $B_{pN+r}(f_1, f_2)$ as

$$\frac{1}{pN + r} \left(\sum_{n=pN+1}^{pN+r} e(n^2 \beta) \cdot T^n f_1 \cdot T^{2n} f_2 + \sum_{q=1}^{p} e(q^2 \beta) \sum_{n=0}^{N-1} T^{pn+q} f_1 \cdot T^{2(pn+q)} f_2 \right).$$

For large N, the sum $\sum_{n=pN+1}^{pN+r} e(n^2\beta) \cdot T^n f_1 \cdot T^{2n} f_2$ can be ignored, and we have

$$\lim_{N\to\infty} B_N(f_1, f_2) = \lim_{N\to\infty} \frac{1}{p} \sum_{q=1}^{p} e(q^2\beta) \frac{1}{N} \sum_{n=0}^{N-1} T^{pn+q} f_1 \cdot T^{2(pn+q)} f_2.$$

Now, [6, Theorem A] implies that the Kronecker factor of $\mathbf{X}$ (which is itself a factor of $\mathbf{A}_2$) is characteristic for the averages above. This proves the first assertion of the lemma when β is rational.

We now assume $\beta \in (0, 1)$ is irrational and consider two cases, based on whether equation (8.3) holds. When equation (8.3) fails we write the coefficient $e(n^2\beta)$ in terms of $\bar{g} T^{p(n)} g$, where $g \in L^\infty(\mu)$ and p is a polynomial. When equation (8.3) holds, we write $e(n^2\beta)$ as $g_0 S^n g_1 S^{2n} g_2$, where $\mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ is an ergodic MPS such that $\mathbf{X} \times \mathbf{Y}$ is ergodic and $g_i \in L^\infty(\nu)$. In each case, we write $B_N(f_1, f_2)$ as a familiar multiple ergodic average and apply known results.

For the first case, we assume equation (8.3) fails. We fix $k \in \mathbb{N}$ such that $1 \neq e(k\beta) \in \mathcal{E}(\mathbf{X})$, meaning $e(k\beta)$ is a non-trivial eigenvalue of $\mathbf{X}$. Let $g \in L^2(\mu)$ be a corresponding eigenfunction, so that $g \circ T = e(k\beta)g$ and $|g| = 1$ μ-almost everywhere. Then $e(k\beta)^m = \bar{g} \cdot T^m g \mu$-almost everywhere. In particular,

$$e(k\beta)^{kn^2+2nj} = \bar{g} \cdot T^{kn^2+2jn} g \quad \text{for all } j, n \in \mathbb{Z} \tag{8.4}$$

in $L^2(\mu)$. Then,

$$\lim_{N\to\infty} B_N(f_1, f_2)$$

$$= \lim_{N\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} \frac{1}{N} \sum_{n=0}^{N-1} e((nk+j)^2\beta) T^{nk+j} f_1 \cdot T^{2nk+2j} f_2$$

$$= \lim_{N\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} \frac{1}{N} \sum_{n=0}^{N-1} e(j^2\beta) e(k\beta)^{kn^2+2jn} T^{nk+j} f_1 \cdot T^{2nk+2j} f_2$$

$$= \lim_{N\to\infty} \frac{1}{k} \sum_{j=0}^{k-1} e(j^2\beta) \frac{1}{N} \sum_{n=0}^{N-1} \bar{g} \cdot T^{kn^2+2jn} g \cdot T^{nk+j} f_1 \cdot T^{2nk+2j} f_2 \quad \text{by equation (8.4)}.$$

The polynomial exponents $p_1(n) = kn^2 + 2jn$, $p_2(n) = nk + j$, $p_3(n) = 2nk + 2j$ are, in the terminology of [6], *essentially distinct* and not *type* (e_1). Therefore, [6, Theorem B] asserts that f_1 and f_2 in equation (8.1) can be replaced with $P_{\mathbf{A}_2} f_1$ and $P_{\mathbf{A}_2} f_2$, respectively, without changing the value of the limit. This proves the first assertion of the lemma in the case where $\mathcal{E}(\mathbf{X}) \cap \{e(n\beta)\}_{n\in\mathbb{Z}} \neq \{1\}$.

Now we assume that equation (8.3) holds. We will prove that $\lim_{N\to\infty} B_N(f_1, f_2) = 0$ for all $f_1, f_2 \in L^\infty(\mu)$. This implies equation (8.2), since $P_{\mathbf{A}_2} f_i \in L^\infty(\mu)$.

Consider the system $\mathbf{Y} = (\mathbb{T}^2, \mathcal{D}, m, S)$, where $S(x, y) = (x + \beta, y + x)$; this $\mathbf{Y}$ is ergodic since β is irrational. As discussed in §7.1, the eigenvalues of $\mathbf{Y}$ are $\{e(n\beta)\}_{n\in\mathbb{Z}}$. Thus, $\mathbf{Y}$ has no non-trivial eigenvalues in common with $\mathbf{X}$, by equation (8.3). The product system $(\mathbb{T}^2 \times X, m \times \mu, S \times T)$ is therefore ergodic, by Lemma 6.2. We will write $B_N(f_1, f_2)$ as an element of $L^2(m \times \mu)$. First observe that for all $(x, y) \in \mathbb{T}^2$, we

have

$$e(n^2\beta) = e(y)e(-2y - 2nx - n(n+1)\beta)e(y + 2nx + n(2n+1)\beta)$$
$$= g_0 \cdot g_1(S^n(x, y)) \cdot g_2(S^{2n}(x, y)),$$

where $g_0(x, y) := e(y)$, $g_1(x, y) := e(-2y)$, $g_2(x, y) := e(y)$. So

$$e(n^2\beta) \cdot T^n f_1 \cdot T^{2n} f_2 = g_0 \otimes 1_X \cdot (S \times T)^n g_1 \otimes f_1 \cdot (S \times T)^{2n} g_2 \otimes f_2 \in L^2(m \times \mu). \tag{8.5}$$

When computing the limit of the averages of the right-hand side in equation (8.5), Theorem 7.1 allows us to replace each $g_i \otimes f_i$ with its projection $\phi_i := P_{\mathbf{Z}}(g_i \otimes f_i)$, where $\mathbf{Z}$ is the Kronecker factor of $\mathbf{Y} \times \mathbf{X}$. By Lemma 6.3 and Observation 7.2, $g_i \otimes f_i$ is orthogonal to every eigenfunction of $\mathbf{Y} \times \mathbf{X}$, so $\phi_i = 0$. Thus, the limit of the averages is 0 in $L^2(m \times \mu)$. Since $B_N(f_1, f_2)$ belongs to the natural embedding of $L^2(\mu)$ in $L^2(m \times \mu)$, this proves $\lim_{N\to\infty} B_N(f_1, f_2) = 0$ in $L^2(\mu)$. □

COROLLARY 8.4. *Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure-preserving system and let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r) \in \mathbb{T}^r$. If $g : \mathbb{T}^r \to \mathbb{C}$ is continuous and $f_i \in L^\infty(\mu)$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N g(n^2\boldsymbol{\beta}) \cdot T^n f_1 \cdot T^{2n} f_2 = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N g(n^2\boldsymbol{\beta}) \cdot T^n \bar{f}_1 \cdot T^{2n} \bar{f}_2$$

in $L^2(\mu)$, where $\mathbf{A}_2$ is the maximal 2-step affine factor of $\mathbf{X}$, and $\bar{f}_i = P_{\mathbf{A}_2} f_i$.

Proof. Uniformly approximating g by trigonometric polynomials, it suffices to prove the lemma in the case where g is a character of $\mathbb{T}^r$. In this case, we can write $g(n^2\boldsymbol{\beta})$ as $e(n^2\alpha)$ for some $\alpha \in [0, 1)$ and apply Lemma 8.2. □

8.1. *Nilsystems and their affine factors.* The following is a restatement of part (i) of Lemma 4.7.

LEMMA 8.5. *Let $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ be an ergodic MPS, $f_i \in L^\infty(\mu)$, and $\varepsilon > 0$. There is a factor $\pi : \mathbf{X} \to \mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ which is a 2-step nilsystem such that for every bounded sequence $(c_n)_{n\in\mathbb{N}}$, we have*

$$\limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^N c_n \int f_0 \cdot T^n f_1 \cdot T^{2n} f_2 \, d\mu - c_n \int \tilde{f}_0 \cdot S^n \tilde{f}_1 \cdot S^{2n} \tilde{f}_2 \, d\nu \right| < \varepsilon \sup_n |c_n|, \tag{8.6}$$

where $\tilde{f}_i \circ \pi := P_{\mathbf{Y}} f_i$.

When computing ergodic averages for ergodic 2-step affine nilsystems, the following lemma allows us to specialize to standard Weyl systems.

LEMMA 8.6. [9, Lemma 4.1] *Let $T : \mathbb{T}^d \to \mathbb{T}^d$ be defined by $T(x) = Ax + b$, where A is a $d \times d$ unipotent integer matrix and $b \in \mathbb{T}^d$. Assume furthermore that T is ergodic. Then T is a factor of an ergodic affine transformation $S : \mathbb{T}^d \to \mathbb{T}^d$, where $S = S_1 \times S_2 \times \cdots \times$*

S_s and for $r = 1, 2, \ldots, s$, $S_r : \mathbb{T}^{d_r} \to \mathbb{T}^{d_r}$ has the form

$$S_r(x_1, \ldots, x_{d_r}) = (x_1 + b_r, x_2 + x_1, \ldots, x_{d_r} + x_{d_r - 1})$$

for some $b_r \in \mathbb{T}$.

Although not explicitly stated in [9], the proof there allows us to conclude that we have $d_r \leq D$, where D is the degree of unipotency of A. Furthermore, if $(A - I)^2 = 0$, as is the case when T is 2-step affine, then we can take $d_r \leq 2$ for each r. For convenience, we may also assume that $d_r = 2$ for each r, and therefore $s = 1$. With these specializations, the system given by S above is a standard 2-step Weyl system.

Proof of Lemma 8.1. Fix a totally ergodic MPS $\mathbf{X} = (X, \mathcal{B}, \mu, T)$, bounded measurable functions f_i on X, $r \in \mathbb{N}$, and $\boldsymbol{\beta} \in \mathbb{T}^r$. Let $g : \mathbb{T}^r \to [0, 1]$ be continuous, and let $\varepsilon > 0$. Consider the averages

$$A_N := \frac{1}{N} \sum_{n=1}^{N} g(n^2 \boldsymbol{\beta}) \int f_0 \cdot T^n f_1 \cdot T^{2n} f_2 \, d\mu.$$

First apply Lemma 8.5 to find a 2-step nilsystem $\mathbf{Y}_0 = (Y_0, \mathcal{D}_0, \nu_0, S)$ satisfying equation (8.6) with $c_n = g(n^2 \boldsymbol{\beta})$, and write B_N for the averages

$$\frac{1}{N} \sum_{n=1}^{N} g(n^2 \boldsymbol{\beta}) \int \tilde{f}_0 \cdot S^n \tilde{f}_1 \cdot S^{2n} \tilde{f}_2 \, d\nu.$$

Our application of Lemma 8.5 means that $\limsup_{N \to \infty} |A_N - B_N| < \varepsilon$.

By Corollary 8.4, the factor $\mathbf{Y} := \mathbf{A}_2(\mathbf{Y}_0)$ is characteristic for the averages B_N: we may replace each $\tilde{f}_i$ with $P_{\mathbf{Y}} \tilde{f}_i$ without affecting $\lim_{N \to \infty} B_N$. The total ergodicity of $\mathbf{X}$ implies every factor of $\mathbf{X}$ is also totally ergodic; in particular, $\mathbf{A}_2(\mathbf{Y}_0)$ is totally ergodic. By Lemma 15.2, we conclude that $\mathbf{A}_2(\mathbf{Y}_0)$ is isomorphic to a unipotent 2-step affine transformation on a finite-dimensional torus, and Lemma 8.6 allows us to conclude that $\mathbf{A}_2(\mathbf{Y})$ is a factor of a standard 2-step Weyl system.

To obtain the remark after part (ii) in the statement of the lemma, apply Lemma 15.8 with $y_n = n^2 \boldsymbol{\beta}$ and $v_n = \int f \cdot T^n f \cdot T^{2n} f \, d\mu - \int \tilde{f} \cdot S^n \tilde{f} \cdot S^{2n} \tilde{f} \, d\nu$. We may apply Lemma 15.8 since the Weyl criterion implies $n^2 \boldsymbol{\beta}$ is uniformly distributed in $\mathbb{T}^r$ whenever $\boldsymbol{\beta}$ is generating. □

Remark 8.7. Our proof of Lemma 8.1 needs the hypothesis of total ergodicity to conclude that $\mathbf{A}_2(\mathbf{Y})$ is isomorphic to a 2-step affine transformation on a finite-dimensional torus. Without this hypothesis, $\mathbf{A}_2(\mathbf{X})$ may be more complicated: the underlying space may be disconnected, and may even have uncountably many connected components. In particular, the Kronecker factor of $\mathbf{X}$ could be isomorphic to a rotation on a compact uncountable totally disconnected abelian group (such as the profinite compactification of $\mathbb{Z}$). This would cause two problems in the following: first, in §13, we exploit the fact that the connected component of a closed subgroup Λ of $\mathbb{T}^d$ has finite index in Λ (although this may not be crucial); second, we simply lack a convenient algebraic description of affine systems

defined on disconnected groups, and such a description is required for our computation in Proposition 13.1.

For similar reasons, we cannot prove Lemma 8.1 starting with an arbitrary totally ergodic **X** and passing immediately to $\mathbf{A}_2(\mathbf{X})$. While disconnectedness will not be a problem, it is possible that the Kronecker factor of **X** is a group rotation on an infinite-dimensional torus, or a solenoid, and then $\mathbf{A}_2(\mathbf{X})$ could be an affine transformation on such a group, which does not fit the hypothesis of Lemma 8.6.

9. *Joinings of groups*

Given two compact abelian groups Z and W with cartesian product $Z \times W$, write π_1 and π_2 for the projection maps onto Z and W, respectively. We say a subgroup $\Gamma \subseteq Z \times W$ is a *joining of Z with W* if Γ is closed, and $\pi_1 : \Gamma \to Z$ and $\pi_2 : \Gamma \to W$ are both surjective.

Observation 9.1. If $\alpha \in Z$ and $\beta \in W$ are generating elements, then the closed subgroup Γ of $Z \times W$ generated by (α, β) is a joining of Z with W: $\pi_1(\Gamma)$ is generated by α and $\pi_2(\Gamma)$ is generated by β.

Joinings arise naturally in the computation of multiple ergodic averages. For example, let $\Gamma := \{(t, t)\} : t \in Z\}$ (= the diagonal of $Z \times Z$), so that Γ is a joining of Z with itself. Then we can write the integral on the right-hand side of equation (7.1) as

$$\int_\Gamma \int_Z f(x) f(x + \pi_1(t)) f(x + 2\pi_2(t)) \, dx \, dm_\Gamma(t). \tag{9.1}$$

The notation $\pi_i(t)$ will be cumbersome in our formulas, so we adopt the following abbreviation.

Notation 9.2. If Γ is a joining of Z with W and $t \in \Gamma$, we write t_1 for $\pi_1(t)$ and t_2 for $\pi_2(t)$.

So the integral in equation (9.1) can be written as $\int_\Gamma \int_Z f(x) f(x + t_1) f(x + 2t_2) \, dx \, dm_\Gamma(t)$.

The joinings we consider will be closed subgroups of $\mathbb{T}^d \times \mathbb{T}^r$; this allows us to exploit the well-known structure of such groups (detailed in [**24**], for example).

Observation 9.3. If Γ is a joining of $\mathbb{T}^d$ with $\mathbb{T}^r$, then its identity component is also a joining of these groups. To see this, note that since Γ is a closed subgroup of a finite-dimensional torus, its identity component Γ_0 has finite index in Γ. The images of π_1 and π_2 therefore have finite index in their respective codomains $\mathbb{T}^d$ and $\mathbb{T}^r$. Since these codomains are connected, they have no proper closed finite index subgroups, so the images $\pi_1(\Gamma_0)$, $\pi_2(\Gamma_0)$ must equal their respective codomains.

If G is a compact abelian group and H is a closed subgroup, m_H denotes Haar probability measure on H. If H' is a coset $H + t$ of H, $m_{H'}$ denotes Haar measure on H', i.e. the measure given by $\int f \, dm_{H'} := \int f(x + t) \, dm_H(x)$.

Definition 9.4. If Γ_0 is a joining of Z with W, Γ_j, $j \le k$ is a collection of cosets of Γ_0, and $c_j \in [0, 1]$ satisfy $\sum_j c_j = 1$, we say that the Γ_j and c_j form an *affine joining* Γ of Z

with W, and define integration over Γ by

$$\int f\, dm_\Gamma := \sum_j c_j \int f\, dm_{\Gamma_j}.$$

For example, $\Gamma_0 = \{(x, 2x) : x \in \mathbb{T}\}$, $\Gamma_1 = \{(x + \frac{1}{4}, 2x) : x \in \mathbb{T}\} \subseteq \mathbb{T} \times \mathbb{T}$, $c_0 = \frac{1}{3}$, and $c_1 = \frac{2}{3}$ determine an affine joining Γ of $\mathbb{T}$ with $\mathbb{T}$, and

$$\int f\, dm_\Gamma = \frac{1}{3} \int f(x, 2x)\, dx + \frac{2}{3} \int f\left(x + \frac{1}{4}, 2x\right) dx.$$

10. *Application of Kronecker's and Weyl's theorems*

The limits of ergodic averages we consider will be computed as integrals over affine joinings. To compute them explicitly, we need the following well-known results of Kronecker and Weyl.

Given a compact abelian group Z and $\alpha_1, \ldots, \alpha_d \in Z$, we write $\langle \alpha_1, \ldots, \alpha_d \rangle$ for the subgroup of Z generated by these elements. We write $\overline{\langle \alpha_1, \ldots, \alpha_d \rangle}$ for its closure.

LEMMA 10.1. (Kronecker's criterion) *Let $\alpha_1, \ldots, \alpha_d$ be elements of a compact abelian group Z. Then $\overline{\langle \alpha_1, \ldots, \alpha_d \rangle} = Z$ if and only if for every non-trivial character $\chi \in \widehat{Z}$, $\chi(\alpha_j) \neq 1$ for at least one of the α_j.*

Weyl's theorem on uniform distribution of polynomials ([**26**], or [**19**, Theorem 3.2]) says that if $p(x) = c_m x^m + c_{m-1} x^{m-1} + \cdots + c_0$ is a polynomial with real coefficients and at least one of the c_j with $j > 0$ is irrational, then

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e(p(n)) = 0.$$

As usual, $e(t)$ denotes $\exp(2\pi i t)$.

LEMMA 10.2. *Let Z be a compact abelian group, let $\alpha, \beta \in Z$, and let $\chi \in \widehat{Z}$ be such that $\chi(\alpha)$, $\chi(\beta)$ are not both roots of unity. Then $\lim_{N\to\infty}(1/N) \sum_{n=1}^{N} \chi(n\alpha + n^2\beta) = 0$.*

Proof. Write $\chi(n\alpha + n^2\beta)$ as $\chi(\alpha)^n \chi(\beta)^{n^2} = e(n\gamma_1 + n^2\gamma_2)$, where at least one of $\gamma_1, \gamma_2 \in [0, 1)$ is irrational. Weyl's theorem then implies the limit of the averages is 0. $\square$

LEMMA 10.3. *Let Z be a compact abelian group with Haar probability measure m and let α, β generate Z.*

(i) *If Z is connected, then for all continuous $f : Z \to \mathbb{C}$, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(n\alpha + n^2\beta) = \int f\, dm. \tag{10.1}$$

(ii) *If Z has finitely many connected components Z_j, then the limit above can be written as $\sum c_j \int f\, dm_{Z_j}$ for some non-negative c_j with $\sum c_j = 1$.*

(iii)   *For fixed β, if $\alpha, \alpha' \in Z$ are such that $\overline{\langle \alpha \rangle} = \overline{\langle \alpha' \rangle}$ and $\overline{\langle \alpha \rangle}$ is connected, we have*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(n\alpha + n^2 \beta) = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(n\alpha' + n^2 \beta). \qquad (10.2)$$

Proof. (i) Approximating f by trigonometric polynomials, it suffices to prove the special case where f is a non-trivial character $\chi \in \widehat{Z}$. Under this assumption, we will show that the limit of the averages in equation (10.1) is 0. In this case, $f(n\alpha + n^2 \beta) = \chi(\alpha)^n \chi(\beta)^{n^2}$. Connectedness of Z implies $\chi^n \neq 1$ for all $n \in \mathbb{N}$. Since α and β generate Z, Lemma 10.1 implies $\chi(\alpha)^n \neq 1$ or $\chi(\beta)^n \neq 1$ for all $n \in \mathbb{N}$. Lemma 10.2 then implies $\lim_{N \to \infty} (1/N) \sum_{n=1}^{N} \chi(n\alpha + n^2 \beta) = 0$.

(ii) Assuming Z has finitely many connected components Z_j and identity component Z_0, let $A_j := \{n \in \mathbb{Z} : n\alpha + n^2 \beta \in Z_j\}$, and let p be the index of Z_0 in Z. We claim that each A_j is a union of infinite arithmetic progressions of the form $p\mathbb{Z} + q$. To prove this, it suffices to prove $A_j + p = A_j$. To do so, observe that $p\alpha, p\beta \in Z_0$. We will show that if $n \in A_j$, then $n - p \in A_j$; in other words, if $n\alpha + n^2 \beta \in Z_j$, then $(n + p)\alpha + (n + p)^2 \beta \in Z_j$. Now fix n, j with $n\alpha + n^2 \beta \in Z_j$. Then

$$(n + p)\alpha + (n + p)^2 \beta = n\alpha + n^2 \beta + p\alpha + (2n + p)p\beta \in Z_j + Z_0 = Z_j,$$

as desired. Similarly, we can show that if $n \in A_j$, then $n + p \in A_j$, so that $A_j + p = A_j$.

Fix $q \in \mathbb{Z}$. We claim $\alpha_0 := (1 + 2q)p\alpha$ and $\beta_0 := p^2 \beta$ generate Z_0. To see this, note that the closed subgroup they generate is contained in Z_0, and has finite index in the subgroup generated by α and β, while Z_0 has no proper finite index closed subgroup.

We decompose the limit in equation (10.1) as

$$\frac{1}{p} \sum_{q=0}^{p-1} \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f((pn + q)\alpha + (pn + q)^2 \beta). \qquad (10.3)$$

Thinking of q as fixed, so that $(pn + q)\alpha + (pn + q)^2 \beta \in Z_i$ for some i, it suffices to prove that the limit is 0 when f is a character χ of Z which is not constant on Z_i (and therefore not constant on Z_0). We fix such a χ and write

$$\chi((pn + q)\alpha + (pn + q)^2 \beta) = \chi(q\alpha + q^2 \beta)\chi(n(1 + 2q)p\alpha + n^2 p^2 \beta)$$
$$= \chi(q\alpha + q^2 \beta)\chi(n\alpha_0 + n^2 \beta_0).$$

Since α_0 and β_0 generate Z_0, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \chi((pn + q)\alpha + (pn + q)^2 \beta) = \chi(q\alpha + q^2 \beta) \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \chi(n\alpha_0 + n^2 \beta_0) = 0,$$

by part (i). This shows that the averages in equation (10.3) converge to 0 when f is a character which is not constant on Z_i, completing the proof of part (ii).

To prove part (iii), fix $\alpha, \alpha', \beta \in Z$ and assume $H := \overline{\langle \alpha \rangle} = \overline{\langle \alpha' \rangle}$ is a connected subgroup of Z. It suffices to prove that equation (10.2) holds when f is a character χ of Z. If $\chi(H) = \{1\}$, then $\chi(n\alpha + n^2 \beta) = \chi(n\alpha' + n^2 \beta) = \chi(n^2 \beta)$, so the averages in equation (10.2) are equal. Now assume $\chi(H) \neq \{1\}$. We will prove that both sides of equation (10.2) are 0. First note that $\chi(H) = \{z \in \mathbb{C} : |z| = 1\}$, since H is compact and connected, and its

image under χ is a non-trivial compact connected subgroup of $\mathcal{S}^1$. Since α and α' generate dense subgroups of H, $\chi(\alpha)$ and $\chi(\alpha')$ generate dense subgroups of $\chi(H)$, and hence they are both not roots of unity. Lemma 10.2 then implies both limits in part (iii) are 0. $\square$

Remark 10.4. Part (iii) of Lemma 10.3 says that when β is fixed and $\overline{\langle \alpha \rangle} = \overline{\langle \alpha' \rangle}$ is connected, the c_j provided by part (ii) do not change when α' replaces α.

11. *The Roth integral and Fourier coefficients*

Let Z be a compact abelian group with Haar probability measure m and $f : Z \to [0, 1]$. We examine the multilinear form which 'counts 3-term arithmetic progressions' in the support of f:

$$I_3(f) := \int f(z) f(z + t) f(z + 2t) \, dm(z) \, dm(t).$$

Roth [22, 23] (cf. [14]) and Furstenberg [11] observed that if $|\hat{f}(\chi)|$ is small for all non-trivial $\chi \in \widehat{Z}$, then $I_3(f) \approx (\int f \, dm)^3$. Lemma 11.1 is a minor generalization of this fact; to state it, we first introduce some notation.

Let $W = Z/K$ be a quotient of Z by a closed subgroup K. For $f \in L^2(m)$, let

$$f'(z) := \int_K f(z + y) \, dm_K(y). \tag{11.1}$$

Let $\pi : Z \to W$ be the quotient map, and identify $\widehat{W}$ with $\{\chi \circ \pi : \chi \in \widehat{W}\} \subseteq \widehat{Z}$. We have

$$\widehat{f'}(\chi) = \begin{cases} \hat{f}(\chi) & \text{if } \chi \in \widehat{W}, \\ 0 & \text{if } \chi \notin \widehat{W}. \end{cases} \tag{11.2}$$

To see this, note that for $\chi \in \widehat{W}$, we have $\chi(z + y) = \chi(z)$ for all $y \in K$, so

$$\hat{f}(\chi) = \int f(z)\overline{\chi(z)} \, dm(z) = \int \int f(z + y)\overline{\chi(z + y)} \, dm(z) \, dm_K(y)$$

$$= \int \int f(z + y) \, dm_K(y)\overline{\chi(z)} \, dm(z)$$

$$= \int f'\overline{\chi} \, dm$$

$$= \widehat{f'}(\chi).$$

Now for $\chi \notin \widehat{W}$, there exists $t \in K$ such that $\chi(t) \neq 1$. Since $f'(z + s) = f'(z)$ for all $s \in K$, we have

$$\widehat{f'}(\chi) = \int f'(z)\overline{\chi(z)} \, dm(z) = \int f'(z + t)\overline{\chi(z + t)} \, dm(z)$$

$$= \int f'(z)\overline{\chi(z + t)} \, dm(z)$$

$$= \overline{\chi(t)} \int f'(z)\overline{\chi(z)} \, dm(z)$$

$$= \overline{\chi(t)}\widehat{f'}(\chi).$$

So $\widehat{f'}(\chi) = \overline{\chi(t)}\widehat{f'}(\chi)$, which is possible only if $\widehat{f'}(\chi) = 0$.

Below we will use dz and dt to indicate integration over all of Z with respect to the displayed variable. Integration over K will be indicated by dm_K.

LEMMA 11.1. *With Z, K, and W as defined above, let f_0, f_1, $f_2 \in L^\infty(m)$, and write*

$$I := \int \int f_0(z) f_1(z+t) f_2(z+2t) \, dz \, dt,$$

$$I_W := \int \int f_0'(z) f_1'(z+t) f_2'(z+2t) \, dz \, dt.$$

Suppose $|\hat{f}_2(\chi)| \leq \kappa$ for all $\chi \in \widehat{Z} \setminus \widehat{W}$. Assuming the map $\chi \mapsto \chi^2$ is injective on $\widehat{Z}$, we have

$$|I - I_W| \leq \kappa \|f_0\|_{L^2(m)} \|f_1\|_{L^2(m)}. \tag{11.3}$$

Proof. Let $I_2 = \int \int f_0(z) f_1(z+t) f_2'(z+2t) \, dz \, dt$. We will prove that

$$|I - I_2| \leq \kappa \|f_0\|_{L^2(m)} \|f_1\|_{L^2(m)} \tag{11.4}$$

and that $I_2 = I_W$. We first prove the special case where each f_i is a trigonometric polynomial. Expanding each f_i as $\sum_{\chi \in \widehat{Z}} \hat{f}_i(\chi) \chi$ and simplifying, we get

$$I = \sum_{\chi, \psi, \tau \in \widehat{Z}} \int \hat{f}_0(\chi) \chi(z) \hat{f}_1(\psi) \psi(z+t) \hat{f}_2(\tau) \tau(z+2t) \, dz \, dt$$

$$= \sum_{\chi, \psi, \tau \in \widehat{Z}} \int \hat{f}_0(\chi) \hat{f}_1(\psi) \hat{f}_2(\tau) \chi \psi \tau(z) \psi \tau^2(t) \, dz \, dt$$

$$= \sum_{\chi, \psi, \tau \in \widehat{Z}} \hat{f}_0(\chi) \hat{f}_1(\psi) \hat{f}_2(\tau) \int \chi \psi \tau(z) \, dz \int \psi \tau^2(t) \, dt.$$

At least one of $\int \psi \tau^2(t) \, dt$ or $\int \chi \psi \tau(z) \, dz$ is zero unless $\psi \tau^2$ and $\chi \psi \tau$ are both trivial; this triviality occurs exactly when $\psi = \tau^{-2}$ and $\chi = \tau$. The sum in the last line above may therefore be restricted to values of χ, ψ, and τ satisfying these identities, and we get

$$I = \sum_{\tau \in \widehat{Z}} \hat{f}_0(\tau) \hat{f}_1(\tau^{-2}) \hat{f}_2(\tau).$$

As noted in equation (11.2), $\hat{f}_2'(\tau) = \hat{f}_2(\tau)$ for $\tau \in \widehat{W}$ and $\widehat{f'}_2(\tau) = 0$ for $\tau \notin \widehat{Z} \setminus \widehat{W}$, so

$$I_2 = \sum_{\tau \in \widehat{W}} \hat{f}_0(\tau) \hat{f}_1(\tau^{-2}) \hat{f}_2(\tau).$$

Then,

$$|I - I_2| = \left| \sum_{\tau \notin \widehat{W}} \hat{f}_0(\tau) \hat{f}_1(\tau^{-2}) \hat{f}_2(\tau) \right|$$

$$\leq \sum_{\tau \in \widehat{Z}} \kappa |\hat{f}_0(\tau) \hat{f}_1(\tau^{-2})| \quad \text{since } |\hat{f}_2(\tau)| < \kappa \text{ for } \tau \notin \widehat{W}$$

$$\leq \kappa \|\hat{f}_0\|_{l^2} \|\hat{f}_1\|_{l^2} \qquad \text{Cauchy–Schwarz, assuming } \tau \mapsto \tau^2 \text{ is injective}$$

$$= \kappa \|f_0\|_{L^2(m)} \|f_1\|_{L^2(m)}, \quad \text{Plancherel,}$$

where $\| \cdot \|_{l^2}$ denotes the l^2 norm for functions on $\widehat{Z}$.

To prove $I_2 = I_W$, replace t with $t + s$ in the dt integral in I_2, then integrate s over K, using the fact that $f_2'(z+s) = f_2(z)$ for all $z \in Z$, $s \in K$:

$$
\begin{aligned}
I_2 &= \int \int f_0(z) f_1(z+t) f_2'(z+2t) \, dz \, dt \\
&= \int_K \int \int f_0(z) f_1(z+t+s) f_2'(z+2t+2s) \, dz \, dt \, dm_K(s) \\
&= \int \int f_0(z) \int_K f_1(z+t+s) \, dm_K(s) \, f_2'(z+2t) \, dz \, dt \\
&= \int \int f_0(z) f_1'(z+t) f_2'(z+2t) \, dz \, dt.
\end{aligned}
$$

A similar manipulation, replacing z with $z+s$, lets us replace f_0 with f_0', completing the proof that $I_2 = I_W$, and hence $|I - I_W| \le \kappa \|f_0\|_{L^2(m)} \|f_1\|_{L^2(m)}$. This proves eqaution (11.4). □

12. Annihilating characters

Let $d, r \in \mathbb{N}$, let $f : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{C}$, $g : \mathbb{T}^r \to \mathbb{C}$, and let Γ be an affine joining of $\mathbb{T}^d$ with $\mathbb{T}^r$ (Definition 9.4). The limits we compute in the proof of Lemma 5.1 will contain functions of the form $f *_\Gamma g : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{C}$, defined by

$$
f *_\Gamma g(x, y) := \int f(x, y + 2\pi_1(w)) g(\pi_2(w)) \, dm_\Gamma(w). \tag{12.1}
$$

The next two lemmas let us bound the Fourier coefficients of $f *_\Gamma g$. We use the abbreviation w_i for $\pi_i(w)$ introduced in Notation 9.2.

LEMMA 12.1. *Let $k < r \in \mathbb{N}$ and $U \subseteq \mathbb{T}^r$ be an approximate Hamming ball of radius (k, η), $\eta > 0$. Then we have the following.*

(i) *If Z is a compact abelian group, $\pi : Z \to \mathbb{T}^r$ is a continuous homomorphism, and $\chi_1, \ldots, \chi_k \in \widehat{Z}$ are non-trivial, then there is a cylinder function g subordinate to U such that $\widehat{g_s \circ \pi}(\chi_j) = 0$ for each j and each translate g_s of g.*

(ii) *If Γ is an affine joining of $\mathbb{T}^d$ with $\mathbb{T}^r$ and $\chi_1, \ldots, \chi_k \in \widehat{\mathbb{T}}^d$ are non-trivial, then there is a cylinder function g subordinate to U such that $\int \chi_j(w_1) g(w_2) \, dm_\Gamma(w) = 0$ for each $j \le k$.*

Proof. (i) Let $\chi_j \in \widehat{Z}$ for $j \le k$, and let K be the kernel of π. We first consider those χ_j where $\chi_j|_K$ is constant. In this case, χ_j can be written as $\chi_j' \circ \pi$, where $\chi_j' \in \widehat{\mathbb{T}}^r$, and $\int (g_s \circ \pi) \overline{\chi}_j \, dm$ can be written as

$$
\int g_s \overline{\chi}_j' \, dm_{\mathbb{T}^r}.
$$

So choose g by Lemma 3.9 to make these integrals vanish for such χ_j. For those j where $\chi_j|_K$ is not constant, write $\int_Z f(z) \, dz$ as $\int_Z \int_K f(z+t) \, dm_K(t) \, dm(z)$. Then

$$
\begin{aligned}
\widehat{g_s \circ \pi}(\chi_j) = \int g_s(\pi(z)) \overline{\chi_j(z)} \, dz &= \int \int g_s(\pi(z+t)) \overline{\chi_j(z+t)} \, dm_K(t) \, dz \\
&= \int g_s(\pi(z)) \overline{\chi_j(z)} \, dz \int_K \overline{\chi_j(t)} \, dm_K(t) \\
&= 0,
\end{aligned}
$$

where the last line follows from the fact that $\chi_j|_K$ is a non-trivial character of K.

(ii) Since Γ is an affine joining of $\mathbb{T}^d$ with $\mathbb{T}^r$, by definition, there is a joining Γ_0 so that the integral over Γ is a convex combination of integrals over translates of Γ_0. To prove part (ii), it therefore suffices to find a g subordinate to U so that $\int \chi(w_1)g(w_2)\,dm_{\Gamma+t}(w) = 0$ for every $t \in \mathbb{T}^d \times \mathbb{T}^r$. We will use the identity

$$\int \chi(w_1)g(w_2)\,dm_{\Gamma+t}(w) = \chi(\pi_1(t))\int_{\Gamma_0}\chi(w_1)g_{\pi_2(t)}(w_2)\,dm_{\Gamma_0}(w), \qquad (12.2)$$

which follows from the manipulations

$$\int_{\Gamma_0+t}\chi(w_1)g(w_2)\,dm_{\Gamma_0+t}(w) = \int_{\Gamma_0}\chi(\pi_1(w+t))g(\pi_2(w+t))\,dm_{\Gamma_0}(w)$$

$$= \int_{\Gamma_0}\chi(\pi_1(t))\chi(\pi_1(w))g(\pi_2(w)+\pi_2(t))\,dm_{\Gamma_0}(w)$$

$$= \chi(\pi_1(t))\int_{\Gamma_0}\chi(w_1)g_{\pi_2(t)}(w_2)\,dm_{\Gamma_0}(w).$$

We can consider the functions $z \mapsto \chi_j(\pi_1(z))$ as characters $\tilde{\chi}_j$ on Γ_0. These characters are non-trivial since $\pi_1 : \Gamma_0 \to \mathbb{T}^d$ is surjective, so we can apply part (i) of the present lemma (with Γ_0 in place of Z and π_2 in place of π) to find g subordinate to U so that $\int_{\Gamma_0}\chi_j(\pi_1(w))g_s(\pi_2(w))\,dm_{\Gamma_0}(w) =: \widehat{g_s \circ \pi_2}(\tilde{\chi}_j) = 0$ for every translate g_s of g and every $j \le k$. In light of equation (12.2), this proves part (ii). $\qquad\square$

The expression $f *_\Gamma g$ in the next lemma is defined in equation (12.1); for $h : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{C}$, $h' : \mathbb{T}^d \to \mathbb{C}$ is defined as $h'(x) := \int_{\mathbb{T}^d} h(x, y)\,dy$.

LEMMA 12.2. *Let $k, d, r \in \mathbb{N}$, and let Γ be an affine joining of $\mathbb{T}^d$ with $\mathbb{T}^r$. Let $U \subseteq \mathbb{T}^r$ be an approximate Hamming ball of radius (k, η) for some $\eta > 0$.*

Let $f : \mathbb{T}^d \times \mathbb{T}^d \to [0, 1]$. Then there is a cylinder function g subordinate to U such that

$$|\widehat{f *_\Gamma g}(\chi, \psi)| < k^{-1/2} \quad \text{whenever } \psi \text{ is non-trivial}, \qquad (12.3)$$

$$(f *_\Gamma g)' = f'. \qquad (12.4)$$

Proof. Let $(\chi, \psi)_j$, $j \le k$, denote the characters of $\mathbb{T}^d \times \mathbb{T}^d$ having the k largest values of $|\hat{f}(\chi, \psi)|$, among those where ψ is non-trivial. Lemma 3.7 implies

$$|\hat{f}(\chi, \psi)| < k^{-1/2} \text{ for all } (\chi, \psi) \notin \{(\chi, \psi)_1, \ldots, (\chi, \psi)_k\} \text{ with } \psi \text{ non-trivial}. \quad (12.5)$$

Applying part (ii) of Lemma 12.1 with ψ^2 in place of the χ_j, we may choose a cylinder function g subordinate to U such that

$$\int_\Gamma \psi(2w_1)g(w_2)\,dm_\Gamma(w) = 0 \text{ for every } \psi \text{ appearing in the } (\chi, \psi)_j \text{ selected above}. \qquad (12.6)$$

Expand f as a Fourier series $\sum_{(\chi,\psi)\in\widehat{\mathbb{T}}^d\times\widehat{\mathbb{T}}^d}\hat{f}(\chi,\psi)\chi\psi$, so that

$$f *_\Gamma g = \int f(x, y + 2w_1)g(w_2)\, dm_\Gamma(w)$$

$$= \sum_{(\chi,\psi)} \hat{f}(\chi,\psi)\chi(x)\int \psi(y + 2w_1)g(w_2)\, dm_\Gamma(w)$$

$$= \sum_{(\chi,\psi)} \hat{f}(\chi,\psi)\chi(x)\psi(y)\int \psi(2w_1)g(w_2)\, dm_\Gamma(w),$$

and

$$\widehat{f *_\Gamma g}(\chi,\psi) = \hat{f}(\chi,\psi)\int \psi(2w_1)g(w_2)\, dm_\Gamma(w). \tag{12.7}$$

The integral in equation (12.7) is 0 when (χ,ψ) are among the $(\chi,\psi)_j$, so the inequality in equation (12.3) is satisfied for these characters. For the remaining characters with ψ non-trivial, note that $\int |g|\, dm = 1$, so equation (12.7) implies $|\widehat{f *_\Gamma g}| \le |\hat{f}|$ everywhere. Now equations (12.5) and (12.6) imply equation (12.3). To prove equation (12.4), write

$$(f *_\Gamma g)'(x) = \int\int f(x, y + w_1)g(w_2)\, dm_\Gamma(w)\, dy$$

$$= \int\int f(x, y + w_1)\, dy\; g(w_2)\, dm_\Gamma(w)$$

$$= f'(x)\int g(w_2)\, dm_\Gamma(w)$$

$$= f'(x). \qquad \square$$

LEMMA 12.3. *With the hypotheses of Lemma 12.2, let $f : \mathbb{T}^d \times \mathbb{T}^d \to [0, 1]$, define $f' : \mathbb{T}^d \to [0, 1]$ by $f'(x) := \int_{\mathbb{T}^d} f(x, y)\, dy$, and let $J' := \int\int f'(x)f'(x+s)f'(x+2s)\, ds\, dx$. Then there is a cylinder function g subordinate to U such that*

$$J := \int f(x, y)f(x + s, y + t)f *_\Gamma g(x + 2s, y + 2t)\, ds\, dt\, dx\, dy$$

satisfies $|J - J'| < k^{-1/2}$.

Proof. Choose, by Lemma 12.2, a cylinder function g subordinate to U so that

$$|\widehat{f *_\Gamma g}(\chi,\psi)| < k^{-1/2} \quad \text{for all } \chi, \psi \text{ with } \psi \in \widehat{\mathbb{T}}^d \setminus \{0\}, \tag{12.8}$$

$$(f *_\Gamma g)' = f'. \tag{12.9}$$

Now we apply Lemma 11.1 with $Z = \mathbb{T}^d \times \mathbb{T}^d$, $W = \mathbb{T}^d$, $K = \{0\} \times \mathbb{T}^d$, $I = J$, and $I_W = \int f'(x)f'(x+s)(f *_\Gamma g)'(x+2s)\, dx\, ds$, using $\kappa = k^{-1/2}$ as supplied by equation (12.8). We conclude that $|J - I_W| < k^{-1/2}$. Since $(f *_\Gamma g)' = f'$, we have $I_W = J'$, so this is the desired conclusion. $\qquad \square$

13. *Averages for standard 2-step Weyl systems*

In the next section, we will prove Lemma 5.1 by reducing the general statement to the special case where the totally ergodic system under consideration is a standard 2-step Weyl system. Proposition 13.1 will then allow us to compute the limit of the multiple ergodic averages appearing in equation (5.1).

For the remainder of this section, we fix $d, r \in \mathbb{N}$, $\alpha \in \mathbb{T}^d$, $\beta \in \mathbb{T}^r$ and let $S : (\mathbb{T}^d)^2 \to (\mathbb{T}^d)^2$ be given by $S(x, y) = (x + 2\alpha, y + x)$. We assume α and β generate $\mathbb{T}^d$ and $\mathbb{T}^r$, respectively, and we write m for Haar probability measure on $\mathbb{T}^d$. We maintain the notational conventions introduced in §5 and the intervening sections.

PROPOSITION 13.1. *With d, r, α, β, and S defined above, there is an affine joining Γ of $\mathbb{T}^d$ with $\mathbb{T}^r$ such that for all Riemann integrable $g : \mathbb{T}^r \to \mathbb{R}$ and all bounded measurable $f : \mathbb{T}^d \times \mathbb{T}^d \to \mathbb{R}$, we have*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(n^2 \beta) \int f \cdot f \circ S^n \cdot f \circ S^{2n} \, d(m \times m)$$

$$= \int f(x, y) f(x + s, y + t) f *_\Gamma g(x + 2s, y + 2t) \, ds \, dt \, dx \, dy, \qquad (13.1)$$

where $f *_\Gamma g$ defined in equation (12.1).

Remark 13.2. We use '2α' in place of 'α' in our definition of S to simplify computations. Since every generating $\alpha \in \mathbb{T}^d$ can be written as $2\alpha'$, where α' is generating, there is no loss of generality.

We first prove Lemma 13.4, which provides explicit limits of polynomial averages on $(\mathbb{T}^d)^4 \times \mathbb{T}^r$. Lemma 13.5 then provides an explicit pointwise-almost everywhere limit for the relevant averages in equation (13.1) when f and g are continuous. Corollary 13.6 uses these to establish L^2 convergence with the same limit formula, for bounded measurable f and Riemann integrable g. Proposition 13.1 is then proved in the last paragraph of this section.

The following lemma is needed for the proof of Lemma 13.4; it is nothing but Fubini's theorem together with the translation invariance of Haar measure.

LEMMA 13.3. *Let ν be a Borel probability measure on $\mathbb{T}^d \times \mathbb{T}^r$, let m be Haar measure on $\mathbb{T}^d$, and let $h : \mathbb{T}^d \times (\mathbb{T}^d \times \mathbb{T}^r) \to \mathbb{C}$ be continuous. Then,*

$$\int h(t, w) \, d\nu(w) \, dm(t) = \int h(t - \pi_1(w), w) \, d\nu(w) \, dm(t),$$

where $\pi_1 : \mathbb{T}^d \times \mathbb{T}^r \to \mathbb{T}^d$ is the projection map.

Let $G = (\mathbb{T}^d)^4 \times \mathbb{T}^r$, with elements of G written $(z_1, z_2, z_3, z_4, z_5)$, $z_i \in \mathbb{T}^d$ for $i \leq 4$, $z_5 \in \mathbb{T}^r$. Let G_{3AP} be the closed connected subgroup $\{(s, t, 2s, 2t, 0) : s, t \in \mathbb{T}^d\} \subseteq G$.

LEMMA 13.4. *With α, β, d, r, and G as above, let* $\mathbf{u} = (0, \alpha, 0, 4\alpha, \beta) \in G$. *Then there is an affine joining* Γ *of* $\mathbb{T}^d$ *with* $\mathbb{T}^r$ *such that*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} F(n\mathbf{c} + n^2 \mathbf{u}) = \int F(s, t, 2s, 2t + 2w_1, w_2) \, dm_\Gamma(w) \, ds \, dt \qquad (13.2)$$

for every continuous $F : G \to \mathbb{C}$ and all $\mathbf{c} \in G$ such that $\overline{\langle \mathbf{c} \rangle} = G_{3AP}$.

Proof. Assume $\mathbf{c} \in G$ is such that $\overline{\langle \mathbf{c} \rangle} = G_{3AP}$. Let $\Lambda = \overline{\langle \mathbf{u} \rangle}$ and let $\Phi = \overline{\langle \mathbf{c}, \mathbf{u} \rangle}$. Note that $\Phi = G_{3AP} + \Lambda$. Also, Φ does not depend on $\mathbf{c}$ (assuming $\mathbf{c}$ generates G_{3AP}).

Since Φ is a closed subgroup of a finite-dimensional torus, its identity component Φ_0 has finite index in Φ. Part (ii) of Lemma 10.3 then provides cosets Φ_j of Φ_0 in Φ and non-negative c_j with $\sum c_j = 1$ such that for every continuous $F : G \to \mathbb{C}$, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} F(n\mathbf{c} + n^2 \mathbf{u}) = \sum c_j \int F \, dm_{\Phi_j}. \qquad (13.3)$$

Part (iii) of Lemma 10.3 implies that the c_j do not depend on $\mathbf{c}$, assuming $\overline{\langle \mathbf{c} \rangle} = G_{3AP}$ (which is connected). We will prove that for each coset Φ_j of Φ_0 in Φ, we can write

$$\int F \, dm_{\Phi_j} = \int F(s, t, 2s, 2t + 2w_1, w_2) \, dm_{\Lambda_j}(w) \, ds \, dt, \qquad (13.4)$$

where Λ_j is a coset of Λ_0 (= the identity component of Λ), and that Λ_0 can be viewed as a joining of $\mathbb{T}^d$ with $\mathbb{T}^r$. Combining equation (13.4) with equation (13.3), we get equation (13.2), where Γ is the affine joining of $\mathbb{T}^d$ with $\mathbb{T}^r$ determined by c_j and Λ_j.

Claim

(i) Λ is a joining of the closed subgroups $H_1 := \{(0, z, 0, 4z, 0) : z \in \mathbb{T}^d\}$ and $H_2 := \{(0, 0, 0, 0, v) : v \in \mathbb{T}^r\}$. Its identity component Λ_0 is also a joining of H_1 and H_2.

(ii) $\Phi_0 = G_{3AP} + \Lambda_0$ is the identity component of $G_{3AP} + \Lambda$.

(iii) Every coset of Φ_0 in Φ has the form $G_{3AP} + \Lambda_j$ where Λ_j is a coset of Λ_0 in Λ.

Part (i) of the claim follows from Observation 9.1, the fact that α and β are generating, and Observation 9.3.

To prove part (ii), note that G_{3AP} is closed and connected, so $G_{3AP} + \Lambda_0$ is a closed connected subgroup of $G_{3AP} + \Lambda$. Since Λ_0 is the identity component of Λ, which is a closed subgroup of a finite-dimensional torus, we see that Λ_0 has finite index in Λ. Thus, $G_{3AP} + \Lambda_0$ is a closed, connected, finite index subgroup of $G_{3AP} + \Lambda$, and therefore is its identity component. Part (iii) is an immediate consequence of part (ii) and the fact that G_{3AP} is connected.

The claim allows us to write integrals with respect to Haar measure over Φ_j explicitly. We write integration over a coset Λ_j of Λ_0 in Λ as

$$\int F \, dm_{\Lambda_j} = \int F(0, w_1, 0, 4w_1, w_2) \, dm_{\Lambda_j}(w), \qquad (13.5)$$

where the m_{Λ_j} on the right is viewed as Haar probability measure on a coset of a joining of $\mathbb{T}^d$ with $\mathbb{T}^r$; this identification is possible as H_1 and H_2 are isomorphic to $\mathbb{T}^d$ and $\mathbb{T}^r$,

respectively. We then write integration over Φ_j $(=G_{3AP} + \Lambda_j)$ as

$$\int F \, dm_{\Phi_j} = \int F(s, t + w_1, 2s, 2t + 4w_1, w_2) \, dm_{\Lambda_j}(w) \, ds \, dt. \tag{13.6}$$

This is justified by the fact that the above integral is invariant under translation by elements of G_{3AP} and by elements of Λ_0, so the above integral is indeed integration with respect to Haar probability measure on $G_{3AP} + \Lambda_j$.

We may replace t with $t - w_1$ in equation (13.6). To see this, first observe that the order of the outer integrals can be changed to $dt \, ds$. For a fixed $s \in \mathbb{T}^d$, define h_s on $\mathbb{T}^d \times (\mathbb{T}^d \times \mathbb{T}^r)$ by $h_s(t, w) := F(s, t + w_1, 2s, 2t + 4w_1, w_2)$. The right-hand side of equation (13.6) can then be written as $\int \int h_s(t, w) \, dm_{\Lambda_j}(w) \, dt \, ds$. We apply Lemma 13.3 with m_{Λ_0} in place of ν, and again change the order of integration. The integral in equation (13.6) therefore simplifies to yield equation (13.4), completing the proof. $\qquad\square$

An immediate consequence of Lemma 13.4 is that for every continuous $F : G \to \mathbb{C}$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} F(\mathbf{z} + n\mathbf{c} + n^2\mathbf{u})$$
$$= \int F(z_1 + s, z_2 + t, z_3 + 2s, z_4 + 2t + 2w_1, z_5 + w_2) \, dm_\Gamma(w) \, ds \, dt \tag{13.7}$$

for all $\mathbf{z} \in G$, all $\mathbf{c} \in G$ such that $\overline{\langle \mathbf{c} \rangle} = G_{3AP}$. This can be seen by applying Lemma 13.4 with the translate $F_\mathbf{z}$ in place of F.

LEMMA 13.5. *With the above $d, r, \alpha, \beta, S, G$, and the affine joining Γ provided by Lemma 13.4, there is a set $W \subseteq \mathbb{T}^d$ with $m(W) = 1$ such that*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(n^2\beta) \cdot f \circ S^n(x, y) \cdot f \circ S^{2n}(x, y)$$
$$= \int f(x + s, y + t) f(x + 2s, y + 2t + 2w_1) g(w_2) \, dm_\Gamma(w) \, ds \, dt. \tag{13.8}$$

for all $x \in W$, all $y \in \mathbb{T}^d$, and all continuous $f : (\mathbb{T}^d)^2 \to \mathbb{R}$, $g : \mathbb{T}^r \to \mathbb{R}$.

Proof. Write the terms on the left-hand side of equation (13.8) as

$$g(n^2\beta) f(S^n(x, y)) f(S^{2n}(x, y))$$
$$= F(x + 2n\alpha, y + nx + \tbinom{n}{2}2\alpha, x + 4n\alpha, y + 2nx + \tbinom{2n}{2}2\alpha, n^2\beta)$$
$$= F(\mathbf{z}_{x,y} + n\mathbf{c}_x + n^2\mathbf{u}),$$

where $F : G \to \mathbb{R}$ is given by $F(x, y, x', y', z) := f(x, y) f(x', y') g(z)$ and

$$\mathbf{z}_{x,y} = (x, y, x, y, 0) \quad \mathbf{c}_x = (2\alpha, x + \alpha, 4\alpha, 2(x + \alpha), 0) \quad \mathbf{u} = (0, \alpha, 0, 4\alpha, \beta).$$

If f and g are continuous, then F is continuous, and we may apply Lemma 13.4 (and equation (13.7) in particular) to conclude that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} F(\mathbf{z}_{x,y} + n\mathbf{c}_x + n^2 \mathbf{u})$$
$$= \int F(x + s, y + t, x + 2s, y + 2t + 2w_1, w_2) \, dm_\Gamma(w) \, ds \, dt$$

for all $x \in \mathbb{T}^d$ such that $\overline{\langle \mathbf{c}_x \rangle} = G_{3AP}$ and all $y \in \mathbb{T}^d$. This is equivalent to equation (13.8).

Let $W := \{x \in \mathbb{T}^d : \overline{\langle \mathbf{c}_x \rangle} = G_{3AP}\}$. To complete the proof, we will show that $m(W) = 1$. Note that $x \in W$ if and only if $\chi(\mathbf{c}_x) \neq 1$ for every non-trivial character χ of G_{3AP}, and there are only countably many characters, so it suffices to prove that for every such χ, $m(E_\chi) = 0$, where $E_\chi := \{x \in \mathbb{T}^d : \chi(\mathbf{c}_x) = 1\}$. Every character of G_{3AP} can be written as $\chi((s, t, 2s, 2t, 0)) = \exp(2\pi i(\mathbf{j} \cdot s + \mathbf{k} \cdot t))$ for some $\mathbf{j}, \mathbf{k} \in \mathbb{Z}^d$, and if χ is non-trivial, then $\mathbf{j}$ and $\mathbf{k}$ are not both 0. Thus, $\chi(\mathbf{c}_x) = 1$ if and only if $\mathbf{k} \cdot x = -(2\mathbf{j} + \mathbf{k}) \cdot \alpha$. When $\mathbf{k} = \mathbf{0}$ and $\mathbf{j} \neq \mathbf{0}$, we then have $E_\chi = \varnothing$. When $\mathbf{k} \neq \mathbf{0}$, we see that E_χ is contained in a coset of the closed proper subgroup $\{x \in \mathbb{T}^d : \mathbf{k} \cdot x = 0\}$, so that $m(E_\chi) = 0$. □

COROLLARY 13.6. *With d, r, α, β and S defined above, let $f \in L^\infty(m \times m)$ and let $g : \mathbb{T}^r \to \mathbb{R}$ be Riemann integrable. Define $A_N \in L^\infty(m \times m)$ by*

$$A_N := \frac{1}{N} \sum_{n=1}^{N} g(n^2 \beta) \cdot f \circ S^n \cdot f \circ S^{2n}$$

and let $A(x, y) := \int f(x + s, y + t) f(x + 2s, y + 2t + 2w_1) g(w_2) \, dm_\Gamma(w) \, ds \, dt$, where Γ is the affine joining given by Lemma 13.4. Then, $\lim_{N \to \infty} A_N = A$ in $L^2(m \times m)$.

Proof. Let W be the set provided by Lemma 13.5. To deduce Corollary 13.6, we first prove that equation (13.8) holds for all $x \in W$, $y \in \mathbb{T}^d$, assuming f is continuous and g is Riemann integrable. To prove this, assume we have such x, y, f, and g. Let $h_0^{(k)}, h_1^{(k)}$ be continuous functions on $\mathbb{T}^r$ satisfying $\inf g \leq h_0^{(k)} \leq g \leq h_1^{(k)} \leq \sup g$ pointwise, such that $\lim_{k \to \infty} \int h_1^{(k)} - h_0^{(k)} \, dm_{\mathbb{T}^r} = 0$. For each k, Lemma 13.4 says that equation (13.8) holds with $h_i^{(k)}$ in place of g. Applying Lemma 15.8 with $f \circ S^n(x, y) \cdot f \circ S^{2n}(x, y)$ in place of v_n, we see that

$$\lim_{N \to \infty} A_N(x, y) = \lim_{k \to \infty} \int f(x + s, y + t) f(x + 2s, y + 2t + 2w_1) h_0^{(k)}(w_2) \, dm_\Gamma(w) \, ds \, dt.$$
(13.9)

The pointwise inequalities $h_0^{(k)} \leq g \leq h_1^{(k)}$ and the assumption $\lim_{k \to \infty} \int h_1^{(k)} - h_0^{(k)} \, dm_{\mathbb{T}^d} = 0$ now imply that $\lim_{k \to \infty} h_0^{(k)} = g$ in $L^2(m_{\mathbb{T}^r})$. The limit on the right of equation (13.9) is therefore equal to $\int f(x + s, y + t) f(x + 2s, y + 2t + 2w_1) g(w_2) \, dm_\Gamma(w) \, ds \, dt$. This proves that A_N converges to A for $m \times m$ almost every (x, y), and the dominated convergence theorem then implies A_N converges to A in $L^2(m \times m)$, in the special case where f is continuous and g is Riemann integrable.

To prove the general case of Corollary 13.6, let $f \in L^\infty(m \times m)$ and let $\varepsilon > 0$. Assume, without loss of generality, that $\sup(|g|) \leq 1$. We write $\|\cdot\|$ for the L^2 norm given by $m \times m$. Let $f_0 : (\mathbb{T}^d)^2 \to \mathbb{R}$ be continuous with $\|f - f_0\| < \varepsilon$.

Let $A'_N := (1/N) \sum_{n=1}^N g(n^2\beta) \cdot f_0 \circ S^n \cdot f_0 \circ S^{2n}$, and let

$$A'(x, y) := \int f_0(x + s, y + t) f_0(x + 2s, y + 2t + 2w_1) g(w_2)\, dm_\Gamma(w)\, ds\, dt.$$

Note that

$$\|f \circ S^n \cdot f \circ S^{2n} - f_0 \circ S^n \cdot f_0 \circ S^{2n}\| \leq 2\|f - f_0\|,$$

hence $\|A_N - A'_N\| \leq 2(\sup |g|)\|f - f_0\|$ for every N, and similarly $\|A - A'\| \leq 2(\sup |g|)\|f - f_0\|$. Then

$$\begin{aligned}
\|A_N - A\| &= \|A_N - A'_N + A'_N - A' + A' - A\| \\
&\leq \|A_N - A'_N\| + \|A'_N - A'\| + \|A' - A\| \\
&< 2\varepsilon + \|A'_N - A'\| + 2\varepsilon \quad \text{assuming } \sup(|g|) \leq 1.
\end{aligned}$$

(Apply the identity $ab - cd = a(b - d) + (a - c)(d)$ with $a = f_0 \circ S^n$, $b = f_0 \circ S^{2n}$, $c = f \circ S^n$, $d = f \circ S^{2n}$, note that $\|f \circ S^n - f_0 \circ S^n\| = \|f - f_0\|$ and likewise for S^{2n}.)

From the first paragraph of this proof, we have $\|A'_N - A'\| \to 0$ as $N \to \infty$. Combining this with the above inequalities, we get $\limsup_{N\to\infty} \|A_N - A\| \leq 4\varepsilon$. Since ε was arbitrary and g is fixed, we have $\|A_N - A\| \to 0$ as $N \to \infty$. $\square$

Proposition 13.1 now follows from Corollary 13.6, observing that the left-hand side of equation (13.1) is $\lim_{N\to\infty} \int f(x, y) A_N(x, y)\, dx\, dy$, with A_N as in Corollary 13.6, and the right-hand side of equation (13.1) can be written as $\int f(x, y) A(x, y)\, dx\, dy$ (recalling the definition of $f *_\Gamma g$ from equation (12.1)).

14. *Proof of Lemma 5.1*

Recall the statement of Lemma 5.1: let $k < r \in \mathbb{N}$, $\ell \in \mathbb{N}$, let $\beta \in \mathbb{T}^r$ be generating, and let $U \subseteq \mathbb{T}^r$ be an approximate Hamming ball of radius (k, η) for some $\eta > 0$. For every totally ergodic MPS $(X, \mathcal{B}, \mu, T)$ and every measurable $f : X \to [0, 1]$, there is a cylinder function $g = m(V)^{-1} 1_V$ subordinate to U such that

$$\lim_{N\to\infty} \left| \frac{1}{N} \sum_{n=1}^N g(n^2\ell^2\beta) \int f \cdot T^n f \cdot T^{2n} f\, d\mu - L_3(f, T) \right| < 2k^{-1/2}\|f\|^2. \quad (14.1)$$

Proof. Let $(X, \mathcal{B}, \mu, T)$ be a totally ergodic MPS, $f : X \to [0, 1]$, and $k \in \mathbb{N}$. Let $U \subset \mathbb{T}^r$ be an approximate Hamming ball of radius (k, η), let $\beta \in \mathbb{T}^r$ be generating, and let $\ell \in \mathbb{N}$. Note that $\ell^2\beta$ is also generating. For a Riemann integrable $g : \mathbb{T}^r \to \mathbb{R}$, write

$$A(f, g) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N g(n^2\ell^2\beta) \int f \cdot T^n f \cdot T^{2n} f\, d\mu.$$

We will prove that there is a cylinder function g subordinate to U such that

$$|A(f, g) - L_3(f, T)| < 2k^{-1/2}\|f\|^2. \tag{14.2}$$

Let $M = 1/m(V)$, where V is one of the cylinders $V_{I,\mathbf{y},\eta}$ in equation (3.1). In other words, $M = \|g\|_\infty$ for each cylinder function g subordinate to U. Choose, by Lemma 8.1, a factor $\pi : \mathbf{X} \to \mathbf{Y} = (Y, \mathcal{D}, \nu, S)$ so that $\mathbf{Y}$ is a factor of a standard 2-step Weyl system, and such that for all Riemann integrable $g : \mathbb{T}^r \to [0, M]$, we have

$$|A(f, g) - B(\tilde{f}, g)| < \tfrac{1}{2}k^{-1/2}\|f\|^2, \tag{14.3}$$

where $\tilde{f} \circ \pi = P_{\mathbf{Y}} f$ and $B(\tilde{f}, g) := \lim_{N\to\infty}(1/N)\sum_{n=1}^{N} g(n^2\ell^2\boldsymbol{\beta}) \int \tilde{f} \cdot S^n \tilde{f} \cdot S^{2n} \tilde{f}\, d\nu$. Let

$$C(\tilde{f}) := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int \tilde{f} \cdot S^n \tilde{f} \cdot S^{2n} \tilde{f}\, d\nu,$$

so that $C(\tilde{f}) = B(\tilde{f}, \mathbf{1})$. Note that $A(f, \mathbf{1}) = L_3(f, T)$, so the special case of equation (14.3) with $g = \mathbf{1}$ yields

$$|C(\tilde{f}) - L_3(f, T)| < \tfrac{1}{2}k^{-1/2}\|f\|^2. \tag{14.4}$$

Let $\tilde{\mathbf{Y}} = (\tilde{Y}, \tilde{\mathcal{D}}, \tilde{\nu}, \tilde{S})$ be an extension of $\mathbf{Y}$ which is a standard 2-step Weyl system $(\mathbb{T}^d \times \mathbb{T}^d, \mathcal{B}_{\mathbb{T}^d \times \mathbb{T}^d}, m, \tilde{S})$, and view $\tilde{f}$ as a function on $\tilde{Y} = \mathbb{T}^d \times \mathbb{T}^d$ (cf. Remark 4.6). By Proposition 13.1, there is an affine joining Γ of $\mathbb{T}^d$ with $\mathbb{T}^r$ such that for each Riemann integrable $g : \mathbb{T}^r \to \mathbb{R}$, we have

$$B(\tilde{f}, g) = \int \tilde{f}(x, y)\tilde{f}(x + s, t + y)\tilde{f} *_\Gamma g(x + 2x, y + 2t)\, ds\, dt\, dx\, dy.$$

Let J denote the integral above, define $\tilde{f}' : \mathbb{T}^d \to [0, 1]$ by $\tilde{f}'(x) := \int \tilde{f}(x, y)\, dy$, and let

$$J' := \int \tilde{f}'(x)\tilde{f}'(x + s)\tilde{f}'(x + 2s)\, dx\, ds.$$

Choose, by Lemma 12.3, a cylinder function g subordinate to U so that

$$|J - J'| < k^{-1/2}\|\tilde{f}\|^2. \tag{14.5}$$

Observation 7.2 means $J' = C(\tilde{f})$, so equation (14.5) can be written as

$$|B(\tilde{f}, g) - C(\tilde{f})| < k^{-1/2}\|\tilde{f}\|^2. \tag{14.6}$$

Combining equation (14.6) with equations (14.4), (14.3), and the triangle inequality, we get equation (14.2), completing the proof. □

15. *Auxiliary lemmas*

In §15.1, we prove Lemma 2.3, essentially by repeating a routine proof of Furstenberg's correspondence principle. Section 15.2 explains a fact needed in the proof of Lemma 8.2, and §15.3 states two immediate consequences of Markov's inequality needed in the proof of Lemma 3.5.

15.1. *Compactness.* Here we write $[N]$ for the interval $\{0, 1, \ldots, N-1\}$ in $\mathbb{Z}$.

LEMMA 15.1. *Let $S \subseteq \mathbb{Z}$, $k \in \mathbb{N}$, and $\delta \geq 0$. The following conditions are equivalent.*

(i) *There is a measure preserving system $(X, \mathcal{B}, \mu, T)$ and $A \subseteq X$ with $\mu(A) > \delta$ such that $\mu(\bigcap_{j=0}^{k} T^{-js} A) = 0$ for all $s \in S$.*

(ii) *S is (δ, k)-non-recurrent, meaning condition (i) holds with $\bigcap_{j=0}^{k} T^{-js} A = \varnothing$ in place of $\mu(\bigcap_{j=0}^{k} T^{-js} A) = 0$.*

(iii) *There is a $\delta' > \delta$ such that for all $N \in \mathbb{N}$, there is a set $B_N \subseteq [N]$ with $|B_N| \geq \delta' N$ such that $\bigcap_{j=0}^{k} (B_N - js) = \varnothing$ for all $s \in S$.*

Proof. To prove condition (i) implies condition (ii), let A satisfy condition (i), and let $A' := A \setminus \bigcup_{s \in S} \bigcap_{j=0}^{k} T^{-js} A$. Then, $\mu(A') = \mu(A) > \delta$, while $A' \subseteq \bigcap_{j=0}^{k} T^{-js} A' \subseteq \bigcap_{j=0}^{k} T^{-js} A$ for every $s \in S$. Since A' is both a subset of and disjoint from $\bigcap_{j=0}^{k} T^{-js} A$, we have $\bigcap_{j=0}^{k} T^{-js} A' = \varnothing$ for every $s \in S$.

To prove condition (ii) implies condition (iii), suppose A satisfies condition (ii). Let δ' be such that $\mu(A) > \delta' > \delta$. Fixing $x \in X$ and setting $A_x := A \cap \{T^n x : n \in \mathbb{N}\}$, we have $\bigcap_{j=0}^{k} T^{-jn} A_x = \varnothing$. Setting $B_x := \{n \in \mathbb{Z} : T^n x \in A\}$, we have $\bigcap_{j=0}^{k} (B_x - jn) = \{n \in \mathbb{Z} : T^n x \in \bigcap_{j=0}^{k} T^{-jn} A\}$. Thus, $\bigcap_{j=0}^{k} (B_x - jn) = \varnothing$ whenever $\bigcap_{j=0}^{k} T^{-jn} A = \varnothing$.

Set $F_N := (1/N) \sum_{n=0}^{N-1} 1_A(T^n x)$. Then, $\int F_N(x) \, d\mu(x) = \mu(A)$. It follows that there is an $x \in X$ such that $F_N(x) \geq \delta'$. Our definition of F_N then implies $|B_x \cap [N]| \geq \mu(A) N$.

To prove condition (iii) implies condition (i), suppose condition (iii) holds. Let $X = \{0, 1\}^{\mathbb{Z}}$ with the product topology, and let $\mathcal{B}$ be the corresponding Borel σ-algebra. Let $T : X \to X$ be the left shift, meaning $(Tx)(n) = x(n+1)$. We will construct a Borel probability measure μ on $(X, \mathcal{B})$ and find a clopen set $A \subseteq X$ satisfying condition (i).

Let $A := \{x \in X : x(0) = 1\}$ (so A is the cylinder set where 1 appears at index 0). For each $N \in \mathbb{N}$, let $y_N := 1_{B_N} \in X$. Note that $1_A(T^n y_N) = 1$ if and only if $n \in B_N$, and similarly

$$1_{A \cap T^{-s} A \cap \cdots \cap T^{-ks} A}(T^n y_N) = 1 \quad \text{if and only if } n \in \bigcap_{j=0}^{k} (B_N - js). \qquad (15.1)$$

Form a measure μ_N on X defined by

$$\int f \, d\mu_N := \frac{1}{N} \sum_{n=0}^{N-1} f(T^n y_N).$$

Let μ be a weak* limit of the μ_N (that is, choose a convergent subsequence of μ_N and let μ be the limit). To see that μ is T-invariant, note that

$$\left| \int f \circ T \, d\mu_N - \int f \, d\mu_N \right| = \frac{1}{N} |f(T^N y_N) - f(y_N)| \leq \frac{2}{N} \sup |f|$$

for every N, so $\int f \circ T \, d\mu = \int f \, d\mu$ for every bounded continuous f. In particular, $\mu(T^{-1} C) = \int 1_C \circ T \, d\mu = \int 1_C \, d\mu = \mu(C)$ for every clopen set $C \subseteq X$. Since the clopen subsets of X generate the Borel σ-algebra of X, this proves that T preserves μ.

To see that $\mu(A) \geq \delta'$, note that

$$\mu(A) \geq \liminf_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} 1_A(T^n y_N) \geq \liminf_{N \to \infty} \frac{1}{N} |B_N| \geq \delta'.$$

To prove that $\mu(\bigcap_{j=0}^{k} T^{-js} A) = 0$ for all $s \in S$, fix $s \in S$ and note that equation (15.1) implies

$$\mu_N\left(\bigcap_{j=0}^{k} T^{-js} A \right) = \frac{1}{N} \sum_{n=0}^{N-1} 1_{A \cap T^{-s} A \cap \cdots \cap T^{-ks} A}(T^n y_N) \leq \frac{1}{N} \left| \bigcap_{j=0}^{k} (B_N - js) \right| = 0$$

for all $N \in \mathbb{N}$. Since $C := \bigcap_{j=0}^{k} T^{-js} A$ is clopen and μ is a weak* limit of the μ_N, we have $\mu(C) = \lim_{N \to \infty} \mu_N(C) = 0$. □

Recall the statement of Lemma 2.3: if $k \in \mathbb{N}$, $0 \leq \delta < \delta'$, and $S \subseteq \mathbb{Z}$ is such that every finite subset of S is (δ', k)-non-recurrent, then S is (δ, k)-non-recurrent.

Proof of Lemma 2.3. Suppose $S \subseteq \mathbb{Z}$, $k \in \mathbb{N}$, $0 \leq \delta < \delta'$, and that every finite subset of S is (δ', k)-non-recurrent. Applying Lemma 15.1 to the finite set $S_N := S \cap [-N, N]$, we may choose, for each N, a set $B_N \subseteq [N]$ such that $|B_N| > \delta' N$ and $\bigcap_{j=0}^{N} (B_N - js) = \varnothing$ for all $s \in S \cap [-N, N]$. Note that this implies $\bigcap_{j=0}^{N} (B_N - js) = \varnothing$ for all $s \in S$, since $B_N - js$ is disjoint from $[N]$ for every $s \in S \setminus [-N, N]$. This means S satisfies condition (iii) of Lemma 15.1, and we conclude that S is δ-non-recurrent. □

15.2. *The 2-step affine factor of a totally ergodic nilsystem.* A *nilsystem* is an MPS $(Y, \mathcal{D}, \nu, S)$ where $Y = G/\Gamma$, with G a nilpotent Lie group and Γ a cocompact discrete subgroup, $\mathcal{D}$ is the Borel σ-algebra of Y, ν is the unique probability measure on $(Y, \mathcal{D})$ invariant under left multiplication, and $Sy = ay$ for some fixed $a \in G$.

When G is a topological group, we write G_0 for the connected component of the identity. For Lie groups, G_0 is a closed subgroup of G. We will use the fact that an ergodic nilsystem $(G/\Gamma, \mathcal{B}, \mu, T)$ is totally ergodic if and only if G/Γ is connected.

Lemma 15.2 identifies the maximal 2-step affine factor of a totally ergodic nilsystem; the purpose of this subsection is to explain how it follows from the results of [6], where it is essentially proved but not explicitly stated.

LEMMA 15.2. *Let* $\mathbf{X} = (X, \mathcal{B}, \mu, T)$ *be a totally ergodic nilsystem. The maximal 2-step affine factor* $\mathbf{A}_2(\mathbf{X})$ *of* $\mathbf{X}$ *is isomorphic to* $(\mathbb{T}^d, \mathcal{B}, m, A)$, *where* $d \in \mathbb{N}$ *and* $A : \mathbb{T}^d \to \mathbb{T}^d$ *is a 2-step unipotent affine transformation.*

We will use the following standard fact about factors: let $\pi_i : \mathbf{X} \to \mathbf{X}_i = (X_i, \mathcal{B}_i, \nu_i, T_i)$, $i = 1, 2$, be two factors of a system where $(X_i, \mathcal{B}_i, \nu_i)$ are separable as measure spaces. Then, $\mathbf{X}_1$ and $\mathbf{X}_2$ are isomorphic (as measure-preserving systems) if the algebra of bounded $\mathbf{X}_1$-measurable functions is equal, up to μ-measure 0, to the algebra of bounded $\mathbf{X}_2$-measurable functions. We also need the following lemma from [9].

LEMMA 15.3. [**9**, Proposition 3.1] *Let $X = G/\Gamma$ be a connected nilmanifold such that G_0 is abelian. Then any nilrotation $T_a(x) = ax$ defined on X with Haar measure μ is isomorphic to a unipotent affine transformation U on some finite-dimensional torus.*

Remark 15.4. The computation in [**9**] showing that the transformation U is unipotent also shows that when G is k-step nilpotent, U is k-step unipotent.

We now explain how Lemma 15.2 follows from [**6**]. Let $\mathbf{X}$ be a totally ergodic nilsystem, $\mathbf{X} = (X, \mathcal{B}, \mu, T)$, where $X = G/\Gamma$, G being a nilpotent Lie group, Γ a cocompact lattice in G, and μ the unique left-translation invariant Borel probability measure on G/Γ, $Tx\Gamma := ax\Gamma$ for some fixed $a \in G$. It is shown by [**6**, Proposition 2.4] that the algebra of functions measurable with respect to $\mathbf{A}_2(\mathbf{X})$ coincides with the functions measurable with respect to the factor $\pi_2 : \mathbf{X} \to \mathbf{Y}$, where $\mathbf{Y} = (X', \mathcal{B}', \mu', T')$, $X' := G/(G_3[G_0, G_0]\Gamma)$, the factor map is given by $\pi_2(x\Gamma) := xG_3[G_0, G_0]\Gamma$, and $T'y = \pi_2(a)y$. Furthermore, it is easy to verify (given the background suggested in [**6**, §2.2]) that X' can be written as G'/Γ', where Γ' is a cocompact lattice in $G' := G/(G_3[G_0, G_0])$, and G' is a 2-step nilpotent Lie group with abelian identity component. It is stated by [**9**, Proposition 3.1] (cf. Remark 15.4 above) that $\mathbf{Y}$ is isomorphic to a 2-step unipotent affine transformation A on a finite-dimensional torus. Since the $\mathbf{A}_2(\mathbf{X})$-measurable functions coincide with the $\mathbf{Y}$-measurable functions, we get that $\mathbf{A}_2(\mathbf{X})$ is itself isomorphic to $\mathbf{Y}$.

15.3. *Consequences of Markov's inequality.* Let (X, μ) be a probability space partitioned into subsets X_i, $0 \le i \le M - 1$, with $\mu(X_i) = 1/M$ for each i, and let $f : X \to [0, 1]$ have $\int f \, d\mu > \delta$. Let $f_i := f 1_{X_i}$.

LEMMA 15.5. *With X, f, and X_i specified above, let $I := \{i : \int f_i \, d\mu > \delta/2M\}$. Then, $|I| > M\delta/2$.*

Proof. Let $I' := \{0, \dots, M - 1\} \setminus I$. Note that

$$\delta < \int f \, d\mu = \sum_{i \in I'} \int f_i \, d\mu + \sum_{i \in I} \int f_i \, d\mu \le \sum_{i \in I'} \frac{\delta}{2M} + \sum_{i \in I} \frac{1}{M} = \frac{\delta}{2M}(M - |I|) + \frac{1}{M}|I|,$$

so $\delta < \delta/2 + |I|/M(1 - \delta/2)$. This can be rearranged to $M\delta/2 < |I|(1 - \delta/2)$, which implies $M\delta/2 < |I|$. $\qquad \square$

LEMMA 15.6. *With X and X_i as defined above, let $c, \varepsilon > 0$ and assume $f, g : X \to \mathbb{R}$ satisfy $\|f - g\|_{L^1(\mu)} < \varepsilon$. Define*

$$J := \left\{ i : \int_{X_i} |f - g| \, d\mu < \frac{c}{M} \right\}.$$

Then, $|J| > M(1 - \varepsilon/c)$.

Proof. We estimate J', where $J' := \{0, \dots, M - 1\} \setminus J$. Let $\varepsilon_i := \int_{X_i} |f - g| \, d\mu$.

Note that $\sum_{i=0}^{M-1} \varepsilon_i = \|f - g\|_{L^1(\mu)} < \varepsilon$, so $J' = \{i : \varepsilon_i \ge c/M\}$ satisfies $|J'| \cdot c/M < \varepsilon$, meaning $|J'| < M\varepsilon/c$. Thus, $|J| = M - |J'| > M(1 - \varepsilon/c)$. $\qquad \square$

The next lemma is an immediate consequence of the triangle inequality and the identity

$$f_1 f_2 \cdots f_k - h_1 h_2 \cdots h_k = \sum_{i=1}^{k} f_1 \cdots f_{i-1}(f_i - h_i)h_{i+1} \cdots h_k.$$

LEMMA 15.7. *If (X, μ) is a probability space, $f_i, h_i : X \to [0, 1]$, $i = 1, \ldots, k$, and $\|f_i - h_i\|_{L^1(\mu)} < \varepsilon$ for each i, then $|\int f_1 f_2 \cdots f_k \, d\mu - \int h_1 h_2 \cdots h_k \, d\mu| < k\varepsilon$.*

15.4. *Convergence with Riemann integrable coefficients.* Let $r \in \mathbb{N}$. We say that a sequence $(y_n)_{n \in \mathbb{N}}$ of elements of $\mathbb{T}^r$ is *uniformly distributed* if

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(y_n) = \int g \, dm$$

for every continuous $g : \mathbb{T}^r \to \mathbb{C}$, where m is Haar probability measure on $\mathbb{T}^d$.

LEMMA 15.8. *Let $r \in \mathbb{N}$, let $(y_n)_{n \in \mathbb{N}}$ be a uniformly distributed sequence of elements of $\mathbb{T}^r$. If $(v_n)_{n \in \mathbb{N}}$ is a bounded sequence of real numbers such that $L(g) := \lim_{N \to \infty}(1/N) \sum_{n=1}^{N} g(y_n)v_n$ exists for every continuous $g : \mathbb{T}^r \to \mathbb{R}$, then $L(g)$ exists for all Riemann integrable g.*

Furthermore, if $h_0^{(k)}, h_1^{(k)}$ are continuous functions on $\mathbb{T}^r$ with $h_0^{(k)} \le g \le h_1^{(k)}$ pointwise and $\lim_{k \to \infty} \int h_1^{(k)} - h_0^{(k)} \, dm = 0$, then $L(g) = \lim_{k \to \infty} L(h_0^{(k)}) = \lim_{k \to \infty} L(h_1^{(k)})$.

Finally, if $C > 0$ and $|L(g)| \le C$ for every continuous $g : \mathbb{T}^r \to [0, 1]$, then $|L(g)| \le C$ for every Riemann integrable $g : \mathbb{T}^r \to [0, 1]$.

Proof. Note that it suffices to prove the statement under the additional assumption that $v_n \in [0, 1]$ for each n. The general case follows by linearity.

Let $g : \mathbb{T}^r \to \mathbb{R}$ be Riemann integrable. Let $\varepsilon > 0$, and choose continuous $g_0, g_1 : \mathbb{T}^r \to \mathbb{R}$ so that $g_0 \le g \le g_1$, $\int g_1 - g_0 \, dm < \varepsilon$. Let $A_N(g) := (1/N) \sum_{n=1}^{N} g(y_n)v_n$. We have

$$L(g_0) = \lim_{N \to \infty} A_N(g_0) \le \liminf_{N \to \infty} A_N(g) \le \limsup_{N \to \infty} A_N(g) \le \lim_{N \to \infty} A_N(g_1) = L(g_1)$$

(15.2)

and $L(g_1) - L(g_0) = L(g_1 - g_0) \le \lim_{N \to \infty}(1/N) \sum_{n=1}^{N} g_1(y_n) - g_0(y_n) = \int g_1 - g_0 \, dm < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, this proves that $A_N(g)$ converges, meaning $L(g)$ exists.

A nearly identical argument will prove the second assertion of the lemma. The third assertion follows from the second, by assuming $h_i^{(k)} : \mathbb{T}^r \to [0, 1]$. □

16. Remarks

16.1. *More general 2-recurrence.* We say that $S \subseteq \mathbb{Z}$ is *good for k-recurrence of powers* if for every MPS $(X, \mathcal{B}, \mu, T)$, every $A \subseteq X$ with $\mu(A) > 0$, and all $c_1, \ldots, c_k \in \mathbb{N}$, there is an $n \in S$ such that $A \cap T^{-c_1 n} A \cap \cdots \cap T^{-c_k n} A \neq \varnothing$.

It is asked in [8, Problem 5] whether $S \subseteq \mathbb{Z}$ being good for k-recurrence of powers implies $S^{\wedge k}$ is a set of measurable recurrence. Our proof of Theorem 1.1 does not

immediately resolve this question for $k = 2$, since we considered intersections of the form $A \cap T^{-n} A \cap T^{-2n} A$ (that is, $c_1 = 1$, $c_2 = 2$ only). We believe that our proof can be modified slightly to construct a set S which is good for 2-recurrence of powers such that $S^{\wedge 2}$ is not a set of measurable recurrence.

16.2. *Higher-order recurrence.* For $k \geq 3$, one possible approach to [8, Problem 5] would be to prove that the set S we construct in the proof of Theorem 1.1 is actually a set of k-recurrence, or to prove that our construction necessarily results in a set which is not a set of k-recurrence. While our construction does not appear to restrict $\mu(A \cap T^{-n} A \cap T^{-2n} A \cap T^{-3n} A)$ for $n \in S$, computations and estimates of

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} g(n^2 \beta) \int f \cdot f \circ T^n \cdot f \circ T^{2n} \cdot f \circ T^{3n} \, d\mu \qquad (16.1)$$

analogous to those in §§13–14 seem to require more intricate reasoning. It may not be possible to specialize the limit in equation (16.1) to affine systems. Perhaps one must consider arbitrary 2-step totally ergodic nilsystems, or even more general systems.

For $k \geq 3$, our approach to Theorem 1.1 leads to the following natural conjecture, an analogue of Lemma 3.5. Here, $BH^{1/k}$ denotes $\{n \in \mathbb{N} : n^k \in BH\}$.

Conjecture 16.1. Let $k \in \mathbb{N}$. For all $\delta > 0$, there exists $m_0 \in \mathbb{N}$ such that for every $r \in \mathbb{N}$, every proper Bohr–Hamming Ball $BH := BH(\boldsymbol{\beta}, \mathbf{y}, m, \varepsilon)$ with $m \geq m_0$, $\varepsilon > 0$ and $\mathbf{y} \in \mathbb{T}^r$, $BH^{1/k}$ is (δ, k)-recurrent.

Conjecture 16.1 could be proved with appropriate higher-order analogues of Lemma 8.1, Proposition 13.1, and Lemma 11.1. For $k \geq 3$, it seems very unlikely that a reduction to 2-step affine systems will be possible, and for $k \geq 4$, it is nearly certain that explicit computations must be carried out for essentially arbitrary $(k-1)$-step totally ergodic nilsystems. These computations seem forbidding, so we hope a more qualitative approach can be developed.

REFERENCES

[1] E. Ackelsberg, V. Bergelson and A. Best. Multiple recurrence and large intersections for abelian group actions. *Discrete Anal.* **18** (2021), 91 pp.

[2] V. Bergelson, B. Host and B. Kra. Multiple recurrence and nilsequences. *Invent. Math.* **160**(2) (2005), 261–303, with an appendix by I. Ruzsa.

[3] V. Bergelson, B. Host, R. McCutcheon and F. Parreau. Aspects of uniformity in recurrence. *Colloq. Math.* **84/85** (2000), 549–576, dedicated to the memory of A. Iwanik.

[4] M. Einsiedler and T. Ward. *Ergodic Theory: With a View Towards Number Theory (Graduate Texts in Mathematics, 259)*. Springer, London, 2011.

[5] A. H. Forrest. Recurrence in dynamical systems: a combinatorial approach. *PhD Thesis*, The Ohio State University, ProQuest LLC, Ann Arbor, MI, 1990.

[6] N. Frantzikinakis. Multiple ergodic averages for three polynomials and applications. *Trans. Amer. Math. Soc.* **360**(10) (2008), 5435–5475.

[7] N. Frantzikinakis. Powers of sequences and recurrence. *Proc. Lond. Math. Soc. (3)* **98**(2) (2009), 504–530.

[8] N. Frantzikinakis. Some open problems on multiple ergodic averages. *Bull. Hellenic Math. Soc.* **60** (2016), 41–90.

[9] N. Frantzikinakis and B. Kra. Polynomial averages converge to the product of integrals. *Israel J. Math.* **148** (2005), 267–276.

[10] N. Frantzikinakis, E. Lesigne and M. Wierdl. Sets of k-recurrence but not $(k+1)$-recurrence. *Ann. Inst. Fourier (Grenoble)* **56**(4) (2006), 839–849.

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

[12] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory* (Princeton University Press, Princeton, NJ, 1981), M. B. Porter. Lectures (82j:28010).

[13] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Anal. Math.* **34** (1978), 275–291.

[14] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11**(3) (2001), 465–588.

[15] J. T. Griesmer. Separating Bohr denseness from measurable recurrence. *Discrete Anal.* **9** (2021), 20 pp.

[16] J. T. Griesmer. Separating topological recurrence from measurable recurrence: exposition and extension of Kriz's example. *Preprint*, 2022, arXiv:2108.01642.

[17] B. Host and B. Kra. *Nilpotent Structures in Ergodic Theory (Mathematical Surveys and Monographs, 236)*. American Mathematical Society, Providence, RI, 2018.

[18] I. Kříž. Large independent sets in shift-invariant graphs: solution of Bergelson's problem. *Graphs Combin.* **3**(2) (1987), 145–158.

[19] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences (Pure and Applied Mathematics)*. Wiley-Interscience, New York, 1974.

[20] R. McCutcheon. Three results in recurrence. *Ergodic Theory and Its Connections with Harmonic Analysis (Alexandria, 1993) (London Mathematical Society Lecture Note Series, 205)*. Eds. K. E. Petersen and I. Salama. Cambridge University Press, Cambridge, 1995, pp. 349–358.

[21] R. McCutcheon. *Elemental Methods in Ergodic Ramsey Theory (Lecture Notes in Mathematics, 1722)*. Springer-Verlag, Berlin, 1999.

[22] K. F. Roth. Sur quelques ensembles d'entiers. *C. R. Math. Acad. Sci. Paris* **234** (1952), 388–390.

[23] K. F. Roth. On certain sets of integers. *J. Lond. Math. Soc. (2)* **28** (1953), 104–109.

[24] W. Rudin. *Fourier Analysis on Groups (Interscience Tracts in Pure and Applied Mathematics, 12)*. Wiley-Interscience Publishers, New York, 1962.

[25] B. Weiss. *Single Orbit Dynamics (CBMS Regional Conference Series in Mathematics, 95)*. American Mathematical Society, Providence, RI, 2000.

[26] H. Weyl. Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.* **77**(3) (1916), 313–352.