

Sarnak’s conjecture for sequences of almost quadratic word growth

REDMOND MCNAMARA

UCLA Los Angeles, CA, USA

(e-mail: rmcnamara@math.ucla.edu)

(Received 24 April 2019 and accepted in revised form 17 June 2020)

Abstract. We prove the logarithmic Sarnak conjecture for sequences of subquadratic word growth. In particular, we show that the Liouville function has at least quadratically many sign patterns. We deduce the main theorem from a variant which bounds the correlations between multiplicative functions and sequences with subquadratically many words which occur with positive logarithmic density. This allows us to actually prove that our multiplicative functions do not locally correlate with sequences of subquadratic word growth. We also prove a conditional result which shows that if the $(\kappa - 1)$ -Fourier uniformity conjecture holds then the Liouville function does not correlate with sequences with $O(n^{t-\varepsilon})$ many words of length n where $t = \kappa(\kappa + 1)/2$. We prove a variant of the 1-Fourier uniformity conjecture where the frequencies are restricted to any set of box dimension less than 1.

Key words: arithmetic and algebraic dynamics, number theory, classical ergodic theory
2020 Mathematics Subject Classification: 37A45, 11N37 (Primary); 11K65 (Secondary)

1. Introduction

The prime number theorem states that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N} \Lambda(n) = 1,$$

where $\Lambda(n)$ is the von Mangoldt function, equal to $\log p$ if n is a power of a prime p and 0 otherwise. (We refer the reader to §1.1 for an explanation of the \mathbb{E} notation.) This is equivalent to the estimate

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N} \lambda(n) = 0,$$

where $\lambda(n) = (-1)^{\# \text{ of prime factors of } n}$ is the Liouville function. Dirichlet's theorem on prime numbers in arithmetic progressions morally follows from the estimate

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N} \mathbb{1}_{n \equiv r \pmod{d}} \lambda(n) = 0,$$

for any d and r . Taking linear combinations, we find that for any periodic function f ,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N} f(n) \lambda(n) = 0.$$

Equivalently, for any function $F : S^1 \rightarrow \mathbb{C}$ and any rational angle α ,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N} F(\alpha^n) \lambda(n) = 0.$$

The analogous estimate when α is irrational and F is a continuous function was proved by Vinogradov and was a key ingredient in his proof that any sufficiently large odd number is the sum of three primes. Green and Tao proved that

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N} F(g^n \Gamma) \lambda(n) = 0$$

where G is a nilpotent Lie group, g is an element of G , Γ is a cocompact lattice and F is a continuous function $F : G/\Gamma \rightarrow \mathbb{C}$. A version of this statement was a key ingredient in their proof with Tamar Ziegler that counts the solutions to almost any system of linear equations over the primes. This motivates the following conjecture, due to Sarnak.

Conjecture 1.1. (Sarnak; see [Sar12]) For any topological dynamical system (X, T) with zero entropy, any continuous function $F : X \rightarrow \mathbb{C}$ and any point x in X ,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N} F(T^n x) \lambda(n) = 0.$$

Tao introduced the following variant,

Conjecture 1.2. (Logarithmically averaged Sarnak conjecture) For any topological dynamical system (X, T) with zero entropy, any continuous function $F : X \rightarrow \mathbb{C}$ and any point x in X ,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} F(T^n x) \lambda(n) = 0.$$

Many instances of Sarnak's conjecture have been proven. A few examples are [Bou13a, BSZ13, Bou13b, DK15, EAKL16, EAKaPLdIR17, FJ18, HLSY17, LS15, MMR14, Mü17, Pec18, Vee17, Wan17], but we stress that this is an incomplete list.

Definition 1.3. A word ϵ of length k is an element of \mathbb{C}^k . Let k be a natural number, let $\epsilon \in \mathbb{C}^k$ and let $b : \mathbb{N} \rightarrow \mathbb{C}$. We say that ϵ occurs as a word of b if there exists a natural number n such that $b(n+h) = \epsilon_h$ for all $h \leq k$. We say that ϵ occurs with (upper) logarithmic density $\delta \in \mathbb{R}$ if

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \mathbb{1}_{\epsilon_h = b(n+h) \text{ for all } h \leq k} = \delta.$$

In this paper, when we refer to log-density we mean upper logarithmic density. A word ϵ whose entries are all ± 1 is called a sign pattern. We say that b has subquadratic word

growth if b takes finitely many possible values and the number of words of length k that occur with positive upper logarithmic density is $o(k^2)$.

Then a particular case of Sarnak’s conjecture predicts that, for any bounded sequence $b: \mathbb{N} \rightarrow \mathbb{C}$ with subexponential word growth,

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} b(n)\lambda(n) = 0.$$

Because λ correlates with itself, this in particular implies that the number of sign patterns of λ of length k is exponential in k . Frantzikinakis and Host [FH18b] proved the special case where b has linear word growth. In this paper, we prove the following special case.

THEOREM 1.4. *Let b be a bounded sequence with subquadratic word growth. Then*

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} b(n)\lambda(n) = 0.$$

Previously Hildebrand [Hil86] showed that all eight sign patterns of length 3 occur infinitely often. Matomäki, Radziwiłł and Tao [MRT16] showed all eight sign patterns of length 3 occur with positive density. Tao and Teräväinen [TT17b] proved that all 16 sign patterns of length 4 occur with positive density, using an argument communicated to them by Matomäki and Sawin. Tao and Teräväinen [TT17b] also showed the number of sign patterns of length k is at least $2k + 8$ for $k \geq 4$. Frantzikinakis and Host [FH18b] showed that the number of sign patterns is superlinear. In particular, Theorem 1.4 implies that λ does not have subquadratically many sign patterns. We actually prove something slightly stronger.

THEOREM 1.5. *There is a constant $\delta > 0$ such that λ has at least δk^2 many sign patterns of length k .*

Tao [Tao17a] showed that the log Sarnak conjecture is equivalent to the following Fourier uniformity conjecture for every natural number t .

Conjecture 1.6. (t-Fourier uniformity) Let G be a nilpotent Lie group of step t , Γ a cocompact lattice and $F: G/\Gamma \rightarrow \mathbb{C}$ a continuous function. Then

$$\lim_{H \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{g \in G} |\mathbb{E}_{h \leq H} \lambda(n+h)F(g^h\Gamma)| = 0.$$

Tao [Tao17a] also showed that this is equivalent to the log-Chowla conjecture for every t .

Conjecture 1.7. (Logarithmic Chowla conjecture) For every natural number t and every distinct natural number h_1, \dots, h_t , we have

$$\lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \lambda(n+h_1) \cdots \lambda(n+h_t) = 0.$$

A function $a: \mathbb{N} \rightarrow \mathbb{C}$ is said to be unpretentious, non-pretentious or strongly aperiodic if there exists a function ϕ from \mathbb{N} to \mathbb{N} such that, for all natural numbers A , for all Dirichlet characters χ of period at most A , for all natural numbers N sufficiently large, and for all

real numbers $|t| \leq AN$, we have

$$\sum_{p \leq N} \frac{1 - \operatorname{Re}(a(p)\chi(p)p^{-it})}{p} \geq \phi(A),$$

and $\phi(A) \rightarrow \infty$ as $A \rightarrow \infty$. The main goal of this paper is to prove the following theorems.

THEOREM 1.8. *Let $a: \mathbb{N} \rightarrow S^1$ be an unpretentious completely multiplicative function taking values in the unit circle. Let $b: \mathbb{N} \rightarrow \mathbb{C}$ be a finite-valued 1-bounded function. Suppose further that for any $\delta > 0$ there are infinitely many k such that the number of words of b of length k that occur with positive upper logarithmic density is at most δk^2 . Then*

$$\lim_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} a(n)b(n)| = 0.$$

We also obtain a conditional version of this result.

THEOREM 1.9. *Let κ be a natural number. Set $t = \binom{\kappa+1}{2}$. Let $a: \mathbb{N} \rightarrow S^1$ be an unpretentious completely multiplicative function taking values in the unit circle so that the local $(\kappa - 1)$ -Fourier uniformity conjecture holds for a . Let $b: \mathbb{N} \rightarrow \mathbb{C}$ be a finite-valued 1-bounded function. Suppose further that for some $\epsilon > 0$ there are infinitely many k such that the number of words of b of length k that occur with positive upper logarithmic density is at most $k^{t-\epsilon}$. Then*

$$\lim_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} a(n)b(n)| = 0.$$

We note that this result matches the numerology in [Saw20] and may be almost the best possible result one can obtain with purely dynamical methods. We also note that even the 1-Fourier uniformity conjecture is still unknown and so this theorem currently has no unconditional content. We also obtain a version of the theorem where b need not take only finitely many values and we only have information about the number of ‘approximate’ words.

Definition 1.10. We say that a sequence b has at most h words of length k up to ϵ rounding if there exists a set Σ of words of length k such that for all $n \in \mathbb{N}$ there is an $\epsilon \in \Sigma$ such that $|b(n + j) - \epsilon_j| \leq \epsilon$ for all $j \leq k$ and the cardinality of Σ is at most h . We say that b has at most h words of length k that occur with positive logarithmic density up to ϵ rounding if we only require $|b(n + j) - \epsilon_j| \leq \epsilon$ for a set of n of lower logarithmic density 1.

THEOREM 1.11. *Let $c > 0$ and $\epsilon > 0$. Then if ϵ is sufficiently small depending on c the following statement holds. Let $a: \mathbb{N} \rightarrow S^1$ be an unpretentious completely multiplicative function taking values in the unit circle. Let $b: \mathbb{N} \rightarrow \mathbb{C}$ be a 1-bounded function with entropy zero. Suppose further that for every $\delta > 0$ there are infinitely many k such that the number of words of b of length k that occur with positive logarithmic density up to ϵ*

rounding is at most δk . Then

$$\limsup_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} a(n)b(n)| \leq c.$$

In fact, this works for any ε satisfying $c^2 > 2\varepsilon$.

We list a few new applications of this theorem.

Proof of Theorem 1.5. Apply Theorem 1.8 to $a = b = \lambda$. □

THEOREM 1.12. *If S is a finite set of sequences of subquadratic word growth and a is an unpretentious completely multiplicative function taking values in the unit circle then*

$$\lim_{H \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\phi \in S} |\mathbb{E}_{h \leq H} a(n+h)\phi(h)| = 0.$$

Remark 1.13. We remark that since the set S is finite, it is enough to show that any one function does not locally correlate with a . However, we also remark that it is generally harder to show that a does not locally correlate with b than it is to show that a does not correlate with b . For Theorem 1.12, we need to use that Theorem 1.8 allows us to handle the case where b may have many words which occur with 0 log-density but still only subquadratically many words which occur with positive log-density. Theorem 1.12 in the linear word growth case seems to follow implicitly from [GLdLR19].

Proof. For convenience, we will assume that 0 is in S . Let $\varepsilon > 0$. We aim to show that

$$\limsup_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\phi \in S} |\mathbb{E}_{h \leq H} a(n+h)\phi(h)| = O(\varepsilon).$$

Suppose not. We will now use an argument of [Tao17a] (see §5 of that paper) to show that a must be correlate with a ‘ticker tape’ function. We define S_ε to be the set of sequences of the form $\phi'(n) = e(\alpha)\phi(n)$ where ϕ is an element of S and α is a rational number with denominator $O(\varepsilon)$. By the pigeonhole principle, for any ϕ in S and any natural numbers H and n in \mathbb{N} there exists a rational number α with denominator $O(\varepsilon)$ such that

$$\text{Re}(|\mathbb{E}_{h \leq H}^{\log} \phi(h)a(n+h)| - \mathbb{E}_{h \leq H}^{\log} e(\alpha)\phi(h)a(n+h)) = O(\varepsilon).$$

Therefore, we may assume for the sake of contradiction that for some $\phi_{n,H}$ in S_ε ,

$$\limsup_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \text{Re}(\mathbb{E}_{n \leq N}^{\log} \mathbb{E}_{h \leq H} a(n+h)\phi_{n,H}(h)) \gg \varepsilon.$$

By a diagonalization argument, we may find a sequence H_i and N_i of natural numbers both tending to infinity and functions $\phi_{n,i} = \phi_{n,H_i}$ such that $N_{i+1} \gg N_i \gg H_i$ a

$$\lim_{i \rightarrow \infty} \text{Re}(\mathbb{E}_{n \leq N_i}^{\log} \mathbb{E}_{h \leq H_i} a(n+h)\phi_{n,i}(h)) \gg \varepsilon.$$

Since the functions ϕ in S_ε and a are bounded, for i sufficiently large there exists a set A_i of natural numbers of lower logarithmic density $\gg \varepsilon$ in the interval $[1, N_i]$ such that for n in A_i ,

$$\text{Re}(\mathbb{E}_{h \leq H_i} a(n+h)\phi_{n,i}(h)) \gg \varepsilon.$$

By a greedy algorithm, we can select a subset B_i of A_i of upper logarithmic density at least ε/H_i in $[1, N_i]$ that is at least H_i separated (meaning distinct points of B_i differ by at least H_i). Now define the 'ticker tape' function ψ as follows:

$$\psi(n + h) = \phi_{n,i}(h),$$

for all n in B_i between N_{i-1} and N_i and $h \leq H_i$. If m is not of the form $n + h$ for n in B_i between N_{i-1} and N_i and $h \leq H_i$ then we set $\psi(m) = 0$. Thus,

$$\limsup_{N \rightarrow \infty} \operatorname{Re}(\mathbb{E}_{n \leq N}^{\log} a(n)\psi(n)) \gg \varepsilon^2.$$

Now we aim to show that ψ has subquadratically many words of length k that occur with positive upper logarithmic density. Let k be a natural number and let ε be a word of length k which occurs in ψ with positive upper logarithmic density. Consider the set C of natural numbers m such that m is within k of an element n of B_i or $B_i + H_i$ for some i . Then since elements of B_i are at least H_i separated, the upper logarithmic density of C in $[N_{i-1}, N_i]$ is at most $2k/H_i$ which clearly tends to 0 as i tends to infinity. Since $N_i \gg N_{i-1}$, we may assume that the log-density of $[1, N_{i-1}]$ in the interval $[1, N_i]$ is also $o(1)$. Thus, C has log-density 0. Therefore, if ε occurs with positive log-density then $(\psi(n + h))_{h=1}^k = \varepsilon$ for a positive density set of n not in C . Since S_ε has only finitely many members, we get that there exists ϕ in S_ε such that for a positive upper logarithmic density set of n , $\psi(n + h) = \phi(n + h) = \varepsilon_h$ for all $h \leq k$. Thus, ψ has subquadratic word growth and a does not correlate with ψ by Theorem 1.8, which gives a contradiction. □

THEOREM 1.14. *Let C be a subset of $[0, 1]$ of upper box dimension less than 1. Then if a is an unpretentious completely multiplicative function taking values in the unit circle,*

$$\lim_{H \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in C} |\mathbb{E}_{h \leq H} a(n + h)e(h\alpha)| = 0.$$

Remark 1.15. In particular, this implies that if C is the middle thirds Cantor set then

$$\lim_{H \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in C} |\mathbb{E}_{h \leq H} a(n + h)e(h\alpha)| = 0.$$

Of course, the result also applies to a large family of other fractals. The author does not know of any results in the literature where this is established for any infinite set. He does not know of any proof for any set of positive box dimension which does not use Theorem 1.11.

Proof. Suppose the upper box dimension of $C \subset S^1$ is less than $d < 1$. Let $\varepsilon > 0$. As in the proof of Theorem 1.12, we assume that

$$\limsup_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in C} |\mathbb{E}_{h \leq H} a(n + h)e(h\alpha)| \gg \varepsilon,$$

and derive a contradiction. As before, there is a ticker tape function $\psi : \mathbb{N} \rightarrow \mathbb{C}$ such that

$$\limsup_{N \rightarrow \infty} \operatorname{Re}(\mathbb{E}_{n \leq N}^{\log} \mathbb{E}_{h \leq H} a(n + h)\psi(h)) \gg \varepsilon^2,$$

of the following form: there exist sequences of natural numbers N_i and H_i tending to infinity with $N_{i+1} \gg N_i \gg H_i$, a sequence of H_i -separated sets B_i , and $\psi(n + h) =$

$e(\beta_n)e(\alpha_n h)$ for some rational β_n of denominator at most $O(\varepsilon)$, some α_n in C and for all n in some set $B_i \cap [N_{i-1}, N_i]$ and $h \leq H_i$. We set $\psi(m) = 0$ for all natural numbers m not of this form. As before, for any natural number k , the natural numbers m that are within k of a number n in B_i or $B_i + H_i$ have log-density 0.

Let k be a natural number sufficiently large depending on C and ε . Let $\varepsilon' = \varepsilon^2$. Then because C has upper box dimension less than d there exists a collection of at most $(k/\varepsilon')^d$ intervals of length ε'/k covering C . If two points on the circle α and α' differ by at most ε'/k then by the triangle inequality, for all $h \leq k$, we have that $|e(h\alpha) - e(h\alpha')| \leq \varepsilon'$. Therefore, the number of sign patterns of ψ that occur with positive log-density up to ε' rounding is sublinear. In particular, for any $\delta > 0$, there are fewer than δk sign patterns that occur with positive log-density up to ε^2 rounding. By Theorem 1.11, we get a contradiction. \square

On the surface, this argument appears to be very close to the t -Fourier uniformity conjecture, which Tao introduced in [Tao17a] and proved was equivalent to the log-Chowla and log-Sarnak conjectures. (For recent significant progress on the Fourier uniformity conjecture, see [MRT18].) If you wanted to prove the Fourier uniformity conjecture in the case $d = 1$, namely that

$$\lim_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in \mathbb{R}} |\mathbb{E}_{h \leq H} a(n+h)e(h\alpha)| = 0,$$

the ticker tape functions that you would need λ to be orthogonal to have $\sim \varepsilon^{-1}k$ sign patterns of length k up to ε rounding. Thus, one might hope that a simple argument could adjust the constants in Theorem 1.11 and thereby prove the Fourier uniformity conjecture. However, there is a major theoretical obstacle to further progress. Frantzikinakis and Host [FH18b] introduced the dynamical system $(S^1 \times S^1, dx, T, \mathcal{B})$ where $T(\alpha, \beta) = (\alpha, \alpha\beta)$. Sawin [Saw20] showed that this dynamical system with some additional structure is a dynamical model for the Liouville function (a notion which we will precisely define later). This is an obstruction to solving the Fourier uniformity conjecture purely with dynamical methods and without any new input from number theory. Sawin [Saw20] further showed that there are dynamical models for the Liouville function which have only polynomially many sign patterns. Explicitly, consider the following function \tilde{a} which behaves almost like a multiplicative function: we partition the natural numbers into intervals with the length of the intervals slowly tending to infinity. For instance, we could split all the numbers between 10^{10^r} and $10^{10^{r+1}}$ into blocks of length $\sim n$. Then on each interval I we pick a random phase α_I in S^1 uniformly and independently. Then we set \tilde{a} to be the function obtained by rounding the function which sends $n \mapsto e(\alpha_I n)$ for n in I . In formulas, we set $\tilde{a}(n) = 2\mathbb{1}_{\text{Re } e(\alpha_I n) > 0} - 1$ for n in I . We remark that the dynamical model for this sequence is isomorphic to the product of the dynamical system introduced by [FH18b] with $\widehat{\mathbb{Z}}$ (again, we defer the precise definition until later). Clearly, \tilde{a} is not multiplicative. However, it is ‘statistically’ multiplicative in the sense that, with high probability, for any sign pattern ϵ of length k , for any m and for large N ,

$$\mathbb{E}_{n \leq N}^{\log} \mathbb{1}_{\tilde{a}_{n+h} = \epsilon_h \text{ for all } h \leq k} \approx \mathbb{E}_{n \leq N}^{\log} m \mathbb{1}_{m|n} \mathbb{1}_{\tilde{a}_{n+mh} = -\epsilon_h \text{ for all } h \leq k}.$$

This function clearly does not satisfy the 1-Fourier uniformity conjecture, and [Saw20] showed that it has quadratically many sign patterns that occur with positive upper logarithmic density even though it does satisfy a version of [MRT15]. If we had used a random $(\kappa - 1)$ -degree polynomial instead of a random linear polynomial, we would get a function which is again statistically multiplicative but which fails the $(\kappa - 1)$ -Fourier uniformity conjecture, and [Saw20] showed that it has $\lesssim k^{\binom{\kappa+2}{2}}$ sign patterns of length k . However, the author is unaware of any ‘dynamical’ techniques that distinguish these statistically multiplicative functions from the Liouville function. This is made precise with Definition 1.18.

We give one last application.

THEOREM 1.16. *Again, suppose that a is an unpretentious completely multiplicative function taking values in the unit circle. There is a set $C \subset [0, 1]$ of Hausdorff dimension 1 such that*

$$\lim_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in C} |\mathbb{E}_{h \leq H} a(n+h)e(h\alpha)| = 0.$$

Proof. The main idea is to combine Theorem 1.14 with a diagonalization argument. For a disjoint collection of intervals $\mathcal{J} = \{J\}$ and a natural number n we define $D_n(\mathcal{J})$ to be the set of intervals obtained by taking each J , removing a ball of diameter $|J|/n$ around the center of the interval J , taking the two remaining intervals, then taking the union over all J in \mathcal{J} .

We construct C inductively as follows. Start with any interval I and set $\mathcal{J}_2 = \{I\}$. Assume inductively that we have constructed \mathcal{J}_{n-1} . Then we apply D_n again and again. Let

$$C_n = \bigcap_{m \in \mathbb{N}} \bigcup_{J \in D^m \mathcal{J}_{n-1}} J.$$

Since C_n has box dimension $(\log n - 1)/(\log n)$, we know by Theorem 1.14 that there exists a natural number H_n such that if $H \geq H_n$ then

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in C} |\mathbb{E}_{h \leq H} a(n+h)e(h\alpha)| \leq \frac{1}{n}.$$

Then define

$$\mathcal{J}_n = D_n^{H_n} \mathcal{J}_{n-1}.$$

We set

$$C = \bigcap_{n \geq 2} \bigcup_{J \in \mathcal{J}_n} J.$$

Clearly, the Hausdorff dimension of C is at least $(-\log 2)/(\log((n - 1)/2n))$ for every n and therefore the Hausdorff dimension is precisely 1. Now we verify that C has the desired property. For each natural number m , by enlarging the set we are maximizing over,

we have that

$$\begin{aligned} & \limsup_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in C} |\mathbb{E}_{h \leq H} a(n+h)e(h\alpha)| \\ & \lesssim \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\alpha \in J \in \mathcal{J}_m} |\mathbb{E}_{h \leq H_m} a(n+h)e(h\alpha)|. \end{aligned}$$

Since every element $\alpha \in J \in \mathcal{J}_m$ is in $D_m^{mH_m}(\mathcal{J}_{m-1})$ there exists $\beta = \beta_\alpha$ depending on α such that β is in C_m and the distance from α to β is no more than $1/H_m m$. Therefore, for all $h \leq H_m$, αh is within $1/m$ of βh . Thus,

$$\lesssim \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\beta \in C_m} |\mathbb{E}_{h \leq H_m} a(n+h)e(h\beta)| + \frac{1}{m}.$$

However, by our choice of H_m , we have

$$\lesssim \frac{1}{m}.$$

Since m was arbitrary, we obtain the desired result. □

Remark 1.17. We have stated our main theorems in the case where a is completely multiplicative and takes values in the unit circle. We remark that these assumptions can be weakened to include all multiplicative functions taking values in the unit disk. The reduction from multiplicative functions taking values in the unit disk to multiplicative functions taking values in the unit circle is essentially due to Tao (see [Tao16, Proposition 2.1]). The reduction from multiplicative functions to completely multiplicative functions (say, both taking values in the unit circle) is carried out in Appendix B. The argument is rather short and was essentially communicated to me by Tao. However, it may be more broadly known and I make no claim of originality.

We now sketch an outline of an argument that is morally very similar to the main argument in this paper. However, for the moment we will work in a more concrete setting. To make this argument rigorous, it is much easier to pass to the dynamical context. Suppose that b is a sequence with quadratic word growth rate and that

$$\limsup_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} \lambda(n)b(n)| > c > 0.$$

Then we can fix a natural number k and average over translates,

$$\limsup_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} \mathbb{E}_{h \leq k} \lambda(n+h)b(n+h)| > c.$$

Fix a large natural number P with $N \gg P \gg k$. Because λ also has a multiplicative symmetry, we can average over dilates

$$\limsup_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} \mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{h \leq k} \lambda(pn+ph)b(n+h)| > c.$$

Moving the absolute values inside and crudely replacing b by the worst word of length k , we get

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\epsilon} \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \lambda(pn+ph)\epsilon_h| > c,$$

where the supremum is taken over all words ϵ of b . Tao’s entropy decrement argument, introduced in [Tao16], allows us to replace pn by n :

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\epsilon} \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \lambda(n + ph) \epsilon_h| > c.$$

Now if λ behaves randomly, then we already know that λ is orthogonal to b . Therefore, if λ correlates with b it must have some structure. Morally, [FH18b] says we can break up λ into a structured part and a random part, and that all the correlation comes from the structured part. Host [HK05] proves that the structured part must take the form of a nilsequence. For the purposes of this sketch, we will focus on the case where there exist α_n and β_n irrational such that

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\epsilon} \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} e(\alpha_n(ph)^2 + \beta_n ph) \epsilon_h| > c.$$

By Hölder’s inequality,

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\epsilon} \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} e(\alpha_n(ph)^2 + \beta_n ph) \epsilon_h|^4 > c^4.$$

By the pigeonhole principle, since there are only δk^2 sign patterns, there is a sign pattern ϵ such that

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} e(\alpha_n(ph)^2 + \beta_n ph) \epsilon_h|^4 > \delta^{-1} k^{-2} c^4.$$

Expanding everything out and using that $\delta \leq c^4/2$,

$$\begin{aligned} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} |\mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{j \in [k]^4} e(\alpha_n p^2 (j_1^2 + j_2^2 - j_3^2 - j_4^2) \\ + \beta_n p (j_1 + j_2 - j_3 - j_4))| > 2k^{-2}. \end{aligned}$$

If $j_1^2 + j_2^2 - j_3^2 - j_4^2 \neq 0$ or $j_1 + j_2 - j_3 - j_4 \neq 0$ then for P large, by the circle method,

$$\mathbb{E}_{P/2 < p \leq P} e(\alpha_n p^2 (j_1^2 + j_2^2 - j_3^2 - j_4^2) + \beta_n p (j_1 + j_2 - j_3 - j_4)) \approx 0.$$

The analogue of the circle method for more general nilpotent Lie groups was introduced in [GT12a], [GT10] and [GTZ12]. The analogue of the step where we conclude that the sums of powers is 0 for more general nilpotent Lie groups is an argument of [Fra17]. Thus the only contribution is from the terms where $j_1^2 + j_2^2 - j_3^2 - j_4^2 = 0$ and $j_1 + j_2 - j_3 - j_4 = 0$. But it is easily seen from Newton’s identities for symmetric polynomials that this only happens for the $2k^2$ ‘diagonal’ terms. Thus, we get

$$2k^{-2} > 2k^{-2},$$

which of course provides a contradiction. For the proof of Theorem 1.9, we need to use not only the theory of symmetric polynomials but also [BDG16].

1.1. *Background and notation.* Suppose $a(n)$ is a 1-bounded, unpretentious multiplicative function with $|a(n)| = 1$ for all n . Let $b(n)$ a sequence where only $o(k^2)$ or $O(k^{t-\epsilon})$ many sign patterns occur with positive log-density. The usual construction of a Furstenberg system (see [FKO82]) for (a, b) proceeds as follows. Consider the point (a, b) in the space

of pairs of sequences. Then apply a random shift to this deterministic variable, $(T^n a, T^n b)$. This gives a random variable in the space of pairs of sequences. The distribution of this random variable is then a shift invariant measure on the space of pairs of sequences. Furthermore, if f is the function on the space of pairs of sequences that evaluates the first sequence at 1 and f' is the function which evaluates the second sequence at 1 then

$$f(T^n a, T^n b) f'(T^n a, T^n b) = a(n + 1) b(n + 1),$$

which is the sequence whose average value we care about. Therefore, if the average of $a(n)b(n)$ is greater than c in absolute value then

$$\left| \int f \cdot f' \right| > c,$$

as well. Of course, it does not really make sense to take a random natural number. Instead, one must shift by a random natural number in a large but finite interval whose length tends to infinity, then find a subsequence of the random variables that converges in distribution. This corresponds to taking a weak-* limit of the corresponding measures.

However, we take a slightly modified approach. The reason is that the function a has some additional symmetry, namely $a(nm) = a(n)a(m)$. As such, the probability that some word occurs (i.e. that $a(n + h) = \epsilon_h$ for $h = 1, \dots, k$ and for n randomly chosen between 1 and N) is the same as the probability that $a(pn + ph) = a(p) \cdot \epsilon_h$ for $h = 1, \dots, k$ and for n chosen randomly between 1 and N . That is the same as p times the probability that for a randomly chosen n between 1 and pN one has $a(n + ph) = a(p) \cdot \epsilon_h$ for $h = 1, \dots, k$ and p divides n . Just flipping everything around, the probability that a random n between 1 and pN satisfies $a(n + ph) = a(p) \cdot \epsilon_h$ and is divisible by p is $1/p$ times the probability that a random n between 1 and N satisfies $a(n + h) = \epsilon_h$. We want our dynamical system to capture this symmetry. There are two difficulties which arise when we want to translate this symmetry to our dynamical system. The first is that the interval keeps changing: the distribution of $T^n a$ might be very different on the intervals from 1 to N and from 1 to pN so when we take a weak limit along a subsequence of intervals, the distribution $T^n a$ for shifts in one interval might approximate our invariant measure while shifts along the other interval might not. The fix for this problem is to use log-averaging. After we weight each natural number n by $1/n$, the probability that a random n will be between N and pN is $\sim (\log p)/(\log N)$ which tends to 0 as N tends to infinity. Therefore, the distribution of $T^n a$ for a random n between 1 and N is very close to the distribution of $T^n a$ for a random n between 1 and pN as long as we choose n randomly using logarithmic weights. The other problem is that our dynamical system does not have a good notion of ‘being divisible’ by a number. To remedy this, we make use of the profinite completion of the integers

$$\widehat{\mathbb{Z}} = \prod_p \mathbb{Z}_p,$$

where p is always restricted to be a prime and \mathbb{Z}_p is the p -adic integers $\mathbb{Z}_p = \lim_{\leftarrow} \mathbb{Z}/p^k\mathbb{Z}$, that is, the inverse limit of $\mathbb{Z}/p^k\mathbb{Z}$ for all k . For each natural number n , we get an element of $\widehat{\mathbb{Z}}$ by reducing $n \bmod p^k$ for every prime p and every natural number k . Then to build our dynamical system, we take the space of triples consisting of two sequences and a profinite integer and for a logarithmically randomly chosen integer n we

consider the random variable $(T^n a, n, T^n b)$ in this space. The distribution of this random variable is a shift invariant measure. Furthermore, we have the following symmetry. Let $Y = \overline{\{T^n b : n \in \mathbb{N}\}}$ and $X = (S^1)^\mathbb{N} \times \widehat{\mathbb{Z}}$. Define the function

$$M : X \rightarrow \widehat{\mathbb{Z}},$$

by projecting onto the $\widehat{\mathbb{Z}}$ coordinate in X ,

$$M : (\alpha, r) \mapsto r.$$

Define the function

$$I_p : M^{-1}(p\widehat{\mathbb{Z}}) \rightarrow X$$

by 'zooming in' by a factor of p and multiplying by $\overline{a(p)}$ on the first factor and dividing by p on the second,

$$I_p : (\alpha(n), r) \mapsto (\overline{a(p)}\alpha(pn), r/p),$$

where r/p is the unique element of $\widehat{\mathbb{Z}}$ such that $p \cdot (r/p) = r$. Then if ν is our invariant measure on $X \times Y$ and μ is its first marginal, I_p pushes forward μ restricted to $M^{-1}(p\widehat{\mathbb{Z}})$ to $(1/p)\mu$. Formally, we make the following definition.

Definition 1.18. Let (X, μ, T) be a dynamical system, let $f : X \rightarrow \mathbb{C}$ be a measurable function, let $M : X \rightarrow \widehat{\mathbb{Z}}$ be a measurable function, and for each m let $I_m : M^{-1}(m\widehat{\mathbb{Z}}) \rightarrow X$ be a measurable function. We say that (X, μ, T, f, M, I_m) is a dynamical model for a if the following statements hold.

- $M \circ T = M + 1$ almost everywhere.
- $I_m \circ T^m = T \circ I_m$ almost everywhere in $M^{-1}(m\widehat{\mathbb{Z}})$.
- I_m pushes forward the measure μ restricted to $M^{-1}(m\widehat{\mathbb{Z}})$ to $(1/m)\mu$. Symbolically, for any function ϕ in $L^1(\mu)$ we have

$$\int_X \phi(x)\mu(dx) = \int_X m \mathbb{1}_{x \in M^{-1}(m\widehat{\mathbb{Z}})} \phi(I_m(x))\mu(dx).$$

- $f \circ I_m = \overline{a(m)} \cdot f$ almost everywhere in $M^{-1}(m\widehat{\mathbb{Z}})$.
- For all m and n , $I_{nm} = I_n \circ I_m$ almost everywhere in $M^{-1}(mn\widehat{\mathbb{Z}})$.

We also require the following property that [Saw20] does not impose.

- For any natural number m and any measurable subset A of \mathbb{C}^m ,

$$\begin{aligned} \mu\{x \in X : (f(T^1x), \dots, f(T^mx)) \in A\} \\ \leq d^{\log}\{n \leq N : (a(n+1), \dots, a(n+m)) \in A\}, \end{aligned}$$

where d^{\log} denotes upper logarithmic density. We remark that we can also fix a Banach limit p -lim extending the usual limit functional and require that equality in the previous equation holds for any limit taken with respect to that Banach limit. For more details, see [Tao17b].

Let $(X \times Y, \nu, T)$ be a joining of two dynamical systems X and Y . Suppose that μ is the first marginal and (X, μ, T, f, M, I_m) is a dynamical model for a . Let f' be a measurable function on $X \times Y$ which is Y measurable. We say that $(X \times Y, \nu, T, f, f', M, I_m)$ is a joining of a dynamical model of a with b if we also have that, for any natural number m

and any measurable subset A of \mathbb{C}^m ,

$$\begin{aligned} & \nu\{(x, y) \in X \times Y : (f'(T^1 y), \dots, f'(T^m y)) \in A\} \\ & \leq d^{\log}\{n \leq N : (b(n+1), \dots, b(n+m)) \in A\} \end{aligned}$$

where d^{\log} denotes upper logarithmic density. We could also require that the joint statistics of (f, f') agree with the joint statistics of (a, b) but this is not necessary for our argument.

Remark 1.19. The preceding definition was used first in [Tao17b] and generalized in [Saw20].

We abuse notation and denote all transformations by the letter T . We also remark that, for the proof of Theorems 1.8 and 1.9, f' only takes finitely many values.

We now specify some notation used in the main argument.

- We fix an unpretentious 1-bounded multiplicative function a . (For the definition of unpretentious, see [MRT15]. We will only really use that a is unpretentious in Theorem 2.1. We remark that the Liouville function is unpretentious). We fix constants $t \in \mathbb{N}$, $c > 0$ and $\delta > 0$. We fix a 1-bounded function b with at most $o(k^2)$ or $k^{t-\varepsilon}$ many words of length k occurring with positive upper logarithmic density for all $k \in \mathcal{K}$ where \mathcal{K} is some fixed infinite set. We suppose that

$$\limsup_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} a(n)b(n)| > c.$$

We fix $\eta > 0$ such that

$$\limsup_{N \rightarrow \infty} |\mathbb{E}_{n \leq N}^{\log} a(n)b(n)| > c + \eta.$$

- We use the following theorem of [FH18a].

THEOREM 2.15. [FH18a, Theorem 1.5] *There exists a joining of a dynamical model for a with b , $(X \times Y, T, \nu, f, f', M, I_m)$, satisfying*

$$\left| \int_{X \times Y} f(x, y) f'(x, y) \nu(dx dy) \right| > c + \eta,$$

and if μ is the first marginal then the ergodic components (X, μ_ω, T) are isomorphic to products of Bernoulli systems with the Host–Kra factor of (X, μ_ω, T) .

Because the statement here is slightly different than that in [FH18a, Theorem 1.5], we will go through the details in Appendix A. We fix such a system. We will always denote by μ the first marginal of ν . We also fix ergodic decompositions $\nu = \int_{\Omega} \nu_\omega d\omega$ and $\mu = \int_{\Omega} \mu_\omega d\omega$. We define the words of length k of f' to be those words ϵ of length k such that the set of (x, y) such that $f'(T^h x, T^h y) = \epsilon_h$ for all $h \leq k$ has positive measure. We note that the set of words of f' is a subset of the set of words of b that occur with positive upper log density: after all, if $f'(T^h x, T^h y) = \epsilon_h$ then by definition of a joining of a dynamical model a with b ,

$$\begin{aligned} 0 & < \mu\{(x, y) \in X \times Y : f'(T^h y) = \epsilon_h \text{ for all } h \leq k\} \\ & \leq d^{\log}\{n \in \mathbb{N} : b(n+h) = \epsilon_h \text{ for all } h \leq k\}, \end{aligned}$$

where d^{\log} denotes upper logarithmic density.

- G will always refer to a nilpotent Lie group. G_s will always refer to the s th step in the lower central series. Γ will always refer to a cocompact lattice in G , meaning that G_s/Γ_s is compact for every s . g, σ , and τ will always refer to group elements. \mathcal{B} will always refer to the Borel sigma algebra. We will fix a particular G, Γ and g following Corollary 2.20. For more on this see [GT12b].
- For a non-empty, finite set A and $\phi: A \rightarrow \mathbb{C}$, we denote $\mathbb{E}_{n \in A} \phi(n) = (1/\#A) \cdot \sum_{n \in A} \phi(n)$. For $A \subset \mathbb{N}$, we denote

$$\mathbb{E}_{n \in A}^{\log} \phi(n) = \frac{1}{\sum_{n \in A} (1/n)} \sum_{n \in A} \frac{\phi(n)}{n}.$$

This notation is due to Frantzikinakis (see [Fra17]). We always restrict p to be prime. By definition a nilsystem is a dynamical system $(G/\Gamma, dx, T, \mathcal{B})$ where G is a nilpotent Lie group, Γ is a cocompact subgroup, dx is Haar measure, there exists g such that $T(x) = gx$ and \mathcal{B} is the Borel sigma algebra. A nilsequence is a sequence of the form $F(g^n \Gamma)$ where G is a nilpotent Lie group, Γ is a cocompact lattice in G , g is an element in G , and $F: G/\Gamma \rightarrow \mathbb{C}$ is a continuous function. Suppose G is an s -step nilpotent Lie group so that G_s is an abelian group and G_s/Γ_s is a compact abelian group. Then a nilcharacter Φ is a function $G/\Gamma \rightarrow \mathbb{C}$ such that there exists a character $\xi: G_s/\Gamma_s \rightarrow S^1$ called the frequency of Φ such that, for all x in G/Γ and u in G_s , we have $\Phi(ux) = \xi(u\Gamma_s)\Phi(x)$. We will abuse notation and identify ξ with the function on G_s that maps $u \mapsto \xi(u\Gamma_s)$. We say that ξ is non-trivial if there exists u in G_s such that $\xi(u) \neq 1$. We say that ξ is non-trivial on the identity component if we can find a u in the identity component of G_s such that $\xi(u) \neq 1$.

- For Theorem 2.14, we will adopt conventions from the theory of Shannon entropy. In particular, $H(x)$ will denote the Shannon entropy of x and $I(x, y)$ will denote the mutual information between x and y . For more details, see [Tao16].
- We will always denote by \mathcal{Z} the smallest sigma algebra on X generated by the union of the sigma algebras corresponding to each of the Host–Kra factors. We will denote

$$B = \{(x, y) \in X \times Y : f'(T^n y) \text{ is eventually periodic as a function of } n\}.$$

Since whether (x, y) is in B depends only y , we will abuse notation and also use $B = \{y \in Y : f'(T^n y) \text{ is eventually periodic as a function of } n\}$.

- For a complex number z , a set A and a real number w we say that $z = O_A(w)$ and $z \lesssim_A w$ if there exists a constant C depending on A but not z and w such that $|z| \leq Cw$. If there are more subscripts we mean that the constant may depend on more parameters. For instance, by $\lesssim_{A,u,K}$ we mean that the implied constant can depend on A, u , and K .

2. Main argument

In this section we prove Theorems 1.8 and 1.9. In §3 we explain how to adapt the proof to handle Theorem 1.11.

We remark that much of the notation, including X, Y, μ, ν, f, f', a , and b , was defined in §1.1.

We start off with a theorem by [MRT15], relying on work in [MR16]. This is a special case of our theorem, so it is no surprise that we need this result.

THEOREM 2.1. ([MRT15, Theorem 1.7]; see also [MR16]) *Let a be a bounded, non-pretentious multiplicative function. Let θ be a periodic sequence. Then*

$$\lim_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} |\mathbb{E}_{h \leq H} a(n+h)\theta(h)| = 0.$$

This theorem says that a does not locally correlate with periodic functions. Eventually, we plan to use a local argument. In particular, our argument will only work for those points where f' does not behave locally like a periodic function. Therefore, we need to exclude any contribution to the integral coming from points where f' behaves like a periodic function. That is the content of the following corollary.

COROLLARY 2.2. *Let $B = \{(x, y) \in X \times Y : f'(T^n y) \text{ is eventually periodic as a function of } n\}$. Then*

$$\int_B f(x)f'(y)v(dx dy) = 0.$$

Proof. In this proof, we introduce some notation which will not be used in the rest of the paper. Because T preserves v and because B is T -invariant, we can average over shifts:

$$\int_B f(x)f'(y)v(dx dy) = \lim_{H \rightarrow \infty} \int_B \mathbb{E}_{h \leq H} f(T^h x)f'(T^h y)v(dx dy).$$

We know f' takes only finitely many values. There are only countably many different periodic sequences taking values in a finite alphabet. Therefore, it suffices to prove that if B_θ is the set of points (x, y) on which $f'(T^h y)$ is eventually equal to the periodic function θ then

$$0 = \lim_{H \rightarrow \infty} \int_{B_\theta} \mathbb{E}_{h \leq H} f(T^h x)f'(T^h y)v(dx dy).$$

Let $\varepsilon > 0$. By Theorem 2.1, for H sufficiently large,

$$\varepsilon^3 \gg \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{j \in \mathbb{N}} |\mathbb{E}_{h \leq H} a(n+h)\theta(h+j)|. \tag{1}$$

We claim that, translating this to the dynamical world using the definition of a dynamical model for a ,

$$v\left\{ (x, y) : \limsup_{H \rightarrow \infty} |\mathbb{E}_{h \leq H} f(T^h x)\theta(h)| > \varepsilon \right\} \leq \varepsilon.$$

After all, by Chebyshev’s inequality, for any H such that (1) holds,

$$d^{\log} \left\{ n \in \mathbb{N} : \sup_{j \in \mathbb{N}} |\mathbb{E}_{h \leq H} a(n+h)\theta(h+j)| \geq \varepsilon \right\} \ll \varepsilon^2,$$

where d^{\log} denotes upper logarithmic density. Fix such an H for the moment and fix a natural number $H' > H$. In fact, more is true. Let S be the subset of the natural numbers such that n is in S if and only if there exists a natural number $H' \geq H_n \geq H$ such that

$$|\mathbb{E}_{h \leq H_n} a(n+h)\theta(h)| \geq \varepsilon.$$

Let S' denote the union of all the intervals $[n + 1, n + H_n]$ for n in S . We claim there is a subcollection \mathcal{I} of these intervals which covers S' and such that each natural number is contained in at most two intervals in \mathcal{I} . This is a somewhat standard covering lemma, but we include the details for the interested reader. For instance, consider the following construction. Let \mathcal{I}_0 denote the empty set. Then assuming we have constructed \mathcal{I}_ℓ for some natural number ℓ , let m denote the smallest natural number in S' not contained in the union of the intervals in \mathcal{I}_ℓ . (If no such m exists, then just set $\mathcal{I}_{\ell+1} = \mathcal{I}_\ell$.) Let n be a natural number maximizing $n + H_n$ subject to the constraints that n is in S and m is in $[n + 1, n + H_n]$. Such an n exists because m is in S' . Then let $\mathcal{I}_{\ell+1} = \mathcal{I}_\ell \cup \{[n + 1, n + H_n]\}$. Now we check that $\mathcal{I} = \bigcup \mathcal{I}_\ell$ has the desired property. First, for any m in S , m is clearly contained in the union of the intervals in \mathcal{I}_m . Thus, \mathcal{I} covers S' . Second, suppose that m in S' is contained in I_1, I_2 , and I_3 , with I_ℓ chosen before $I_{\ell+1}$ for $\ell = 1, 2$. Suppose that \mathcal{I}_{ℓ_i} is the first set of the form \mathcal{I}_ℓ where I_i is contained in \mathcal{I}_{ℓ_i} . Then the union of the intervals in \mathcal{I}_{ℓ_1} contains m . Thus, there exists m' in S' such that $I_2 = [n + 1, n + H_n]$ was chosen to maximize $n + H_n$ subject to the constraint that m' is in I_2 . Since we assumed m was contained in I_2 , we have that $n < m$. Now let $I_3 = [n' + 1, n' + H_{n'}]$. Since I_3 also contains m , n' is also less than m which is in turn less than m' . But I_2 maximized $n + H_n$ over all intervals containing m' and if $n' + H_{n'}$ were larger than $n + H_n$ which is larger than m' , then I_3 would contained m' as well. Thus $n' + H_{n'} \leq n + H_n$ and therefore any point contained in I_3 is already contained in the union of the intervals in \mathcal{I}_{ℓ_2} . Therefore, I_3 should not have been selected for \mathcal{I}_{ℓ_3} , which leads to a contradiction. Thus, every natural number is covered at most twice by the union of the intervals in \mathcal{I} . If

$$|\mathbb{E}_{h \leq H_n} a(n + h)\theta(h)| \gtrsim \varepsilon$$

then

$$\mathbb{E}_{h \leq H_n} |\mathbb{E}_{h' \leq H} a(n + h + h')\theta(h + h')| \gtrsim \varepsilon.$$

Therefore, for at least $\varepsilon \cdot H_n$ points $n + h'$ in the interval $[n + 1, n + H_n]$,

$$\sup_{j \in \mathbb{N}} |\mathbb{E}_{h \leq H} a(n + h + h')\theta(h + j)| \gtrsim \varepsilon.$$

However, we know that

$$d^{\log} \left\{ n \in \mathbb{N} : \sup_{j \in \mathbb{N}} |\mathbb{E}_{h \leq H} a(n + h)\theta(h + j)| \geq \varepsilon \right\} \ll \varepsilon^2,$$

and that each such natural number is contained in at most two intervals of the form $[n + 1, n + H_n]$ in \mathcal{I} . We conclude that, by Chebyshev's inequality, the logarithmic density of S' is at most ε . Therefore, the logarithmic density of S is at most ε . This means precisely that

$$d^{\log} \left\{ n \in \mathbb{N} : \sup_{L \in [H, H']} |\mathbb{E}_{h \leq L} a(n + h)\theta(h)| \geq \varepsilon \right\} \leq \varepsilon.$$

The condition $\sup_{L \in [H, H']} |\mathbb{E}_{h \leq L} a(n+h)\theta(h)| \geq \varepsilon$ depends measurably on $(a(n+1), \dots, a(n+H'))$ so by definition of a dynamical model for a ,

$$\nu \left\{ (x, y) : \sup_{L \in [H, H']} |\mathbb{E}_{h \leq L} f(T^h x)\theta(h)| \geq \varepsilon \right\} \leq \varepsilon.$$

Since this is true for all H' , we get that

$$\nu \left\{ (x, y) : \sup_{L \geq H} |\mathbb{E}_{h \leq L} f(T^h x)\theta(h)| \geq \varepsilon \right\} \leq \varepsilon.$$

Since this is true for all $\varepsilon > 0$, for all (x, y) outside a set of measure 0, we have

$$\lim_{H \rightarrow \infty} \mathbb{E}_{h \leq H} f(T^h x)\theta(h) = 0.$$

For $(x, y) \in B_\theta$, we know that $f'(T^h y) = \theta(h)$ for h sufficiently large, so for $(x, y) \in B_\theta$ outside a set of measure 0, we know $\mathbb{E}_{h \leq H} f(T^h x)f'(T^h y) \rightarrow 0$. By the dominated convergence theorem, we have

$$0 = \lim_{H \rightarrow \infty} \int_{B_\theta} \mathbb{E}_{h \leq H} f(T^h x)f'(T^h y)\nu(dx dy)$$

as desired. □

We will also need the following result later. It states that f does not correlate locally with periodic functions.

COROLLARY 2.3. *Let $\mu = \int_{\Omega} \mu_\omega d\omega$ be an ergodic decomposition of μ . For almost every ω , for all 1-bounded functions $\phi: X \rightarrow \mathbb{C}$ such that, for μ_ω -almost every x , $\phi(T^h x)$ is periodic in h , we have*

$$\int_X f(x)\phi(x)\mu_\omega(dx) = 0.$$

Proof. Let d be a natural number. Then we claim that

$$\lim_{H \rightarrow \infty} \lim_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} a(n+h)\theta(h)| = 0,$$

where S_d is the set of $d!$ periodic, 1-bounded functions. Since the supremum is over a finite set, this directly follows from Theorem 2.1. Let $\varepsilon > 0$. For H sufficiently large,

$$\lim_{N \rightarrow \infty} \sup_{n \leq N} \mathbb{E}_{n \leq N}^{\log} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} a(n+h)\theta(h)| \leq \varepsilon^3.$$

Therefore, as in the proof of Proposition 2.2,

$$\nu \left\{ (x, y) : \limsup_{H \rightarrow \infty} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} f(T^h x)\theta(h)| > \varepsilon \right\} \leq \varepsilon.$$

Since this is true for all ε , we get that

$$\int_X \limsup_{H \rightarrow \infty} \sup_{\theta \in S_d} |\mathbb{E}_{h \leq H} f(T^h x)\theta(h)| \mu(dx) = 0.$$

Therefore, there exists $\Omega_d \subset \Omega$ of full measure such that for ω in Ω_d , we have

$$\int_X \limsup_{H \rightarrow \infty} \sup_{\theta \in \mathcal{S}_d} |\mathbb{E}_{h \leq H} f(T^h x) \theta(h)| \mu_\omega(dx) = 0.$$

Now let ω be an element of Ω_d for all d and let ϕ be a 1-bounded function such that $\phi(T^h x)$ is periodic in h for μ_ω -almost every x . Suppose that there exists $\varepsilon > 0$ such that

$$\left| \int_X f(x) \phi(x) \mu_\omega(dx) \right| > \varepsilon.$$

Then by translation invariance, we know that

$$\left| \limsup_{H \rightarrow \infty} \int_X \mathbb{E}_{h \leq H} f(T^h x) \phi(T^h x) \mu_\omega(dx) \right| \geq \varepsilon.$$

Let X_d be the set of all points x such that $\phi(T^h x)$ is periodic with period at most d . Note by assumption that $\mu_\omega(\bigcup X_d) = 1$. Then by dominated convergence, there exists d such that

$$\int_{X_d} \limsup_{H \rightarrow \infty} |\mathbb{E}_{h \leq H} f(T^h x) \phi(T^h x)| \mu_\omega(dx) > \frac{\varepsilon}{2}.$$

Since $\phi(T^h x)$ is $d!$ periodic for every x in X_d , this integral is bounded by

$$\int_X \limsup_{H \rightarrow \infty} \sup_{\theta \in \mathcal{S}_d} |\mathbb{E}_{h \leq H} f(T^h x) \theta(h)| \mu_\omega(dx),$$

which gives a contradiction. □

For the proof of Theorem 1.9, we also need an upgraded version of Corollary 2.3 under the assumption that the $(\kappa - 1)$ -Fourier uniformity conjecture holds.

PROPOSITION 2.4. *Suppose that the $(\kappa - 1)$ -Fourier uniformity conjecture holds, that is, for every nilpotent Lie group G of step less than κ , every cocompact lattice Γ and every continuous function $F : G/\Gamma \rightarrow \mathbb{C}$,*

$$\lim_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{g \in G} |\mathbb{E}_{h \leq H} a(n+h) F(g^h \Gamma)| = 0.$$

Then for almost every ω , we have the following property: for every nilpotent Lie group G of step less than κ , every cocompact lattice Γ , every continuous function $F : G/\Gamma \rightarrow \mathbb{C}$ and every function ϕ on X such that for μ_ω -almost every x there exist x' in G/Γ and g in G we have $\phi(T^h x) = F(g^h x')$ for all h in \mathbb{N} , we have that

$$\int_X f(x) \phi(x) \mu_\omega(dx) = 0.$$

Proof. In the proof of this proposition, we will introduce some notation which will not be used in the rest of the paper. By, for instance, [HK18, Ch. 10, Theorem 28] there are only countably many pairs (G, Γ) up to isomorphism of G/Γ . Thus, we can fix a sequence (G_i, Γ_i) of nilpotent Lie groups of step less than κ and cocompact lattices such that, for any nilpotent Lie group G of step less than κ and for any cocompact lattice Γ there exist a natural number i and a Lie group isomorphism $\psi : G_i \rightarrow G$ such that $\psi(\Gamma_i) = \Gamma$.

By Stone–Weierstrass, there exists a countable, uniformly dense subset of the continuous functions on G_i/Γ_i . Fix such a subset and call it \mathcal{F}_i . We are assuming the $(\kappa - 1)$ -Fourier uniformity conjecture:

$$\lim_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{g \in G_i} |\mathbb{E}_{h \leq H} a(n+h)F(g^h \Gamma_i)| = 0,$$

for all i and all F a continuous function on G/Γ . By [Fra17, §4.5, Step 4] we also get that, for all i and F as before,

$$\lim_{H \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H} a(n+h)F(g^h x)| = 0.$$

Fix a natural number i for the moment and a function F in \mathcal{F}_i . For each $\varepsilon > 0$ there exists H_ε such that

$$\limsup_{N \rightarrow \infty} \mathbb{E}_{n \leq N}^{\log} \sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H_\varepsilon} a(n+h)F(g^h x)| \ll \varepsilon^3.$$

Therefore, by Chebyshev’s inequality,

$$d^{\log} \left\{ n \in \mathbb{N} : \sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H_\varepsilon} a(n+h)F(g^h x)| \geq \varepsilon \right\} \ll \varepsilon^2,$$

where d^{\log} denotes the upper logarithmic density. Note that

$$\sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H_\varepsilon} a(n+h)F(g^h x)|$$

depends measurably on $(a(n+1), \dots, a(n+H_\varepsilon))$. Thus, there exists some set A in $\mathbb{C}^{H_\varepsilon}$ such that

$$\sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H_\varepsilon} a(n+h)F(g^h x)| \geq \varepsilon$$

if and only if $(a(n+1), \dots, a(n+H_\varepsilon))$ are in A . Therefore, we know that

$$d^{\log} \{n \in \mathbb{N} : (a(n+1), \dots, a(n+H_\varepsilon)) \in A\} \ll \varepsilon^2.$$

By the definition of a dynamical model for a ,

$$\mu \{x' \in X : (f(T^1 x'), \dots, f(T^{H_\varepsilon} x')) \in A\} \ll \varepsilon^2.$$

Unpacking definitions, we get

$$\mu \left\{ x' \in X : \sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H_\varepsilon} f(T^h x')F(g^h x)| \geq \varepsilon \right\} \ll \varepsilon^2.$$

We call this set

$$\left\{ x' \in X : \sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H_\varepsilon} f(T^h x')F(g^h x)| \geq \varepsilon \right\} = S_\varepsilon.$$

Remember that S_ε implicitly depends on i and F . By the definition of the ergodic decomposition, we have that

$$\mu(S_\varepsilon) = \int_{\Omega} \mu_\omega(S_\varepsilon) d\omega.$$

Therefore, by another application of Chebyshev's inequality, we find that

$$|\{\omega \in \Omega : \mu_\omega(S_\varepsilon) \leq \varepsilon\}| \geq 1 - \varepsilon.$$

We call this set $K_\varepsilon = \{\omega \in \Omega : \mu_\omega(S_\varepsilon) \leq \varepsilon\}$. Of course K_ε depends on i and F . Define

$$\Omega_{i,F} = \bigcap_{m \in \mathbb{N}} \bigcup_{r \geq m} K_{1/r},$$

and define

$$\Omega' = \bigcap_{i \in \mathbb{N}} \bigcap_{F \in \mathcal{F}} \Omega_{i,F}.$$

Since $|K_\varepsilon| \geq 1 - \varepsilon$, we know that for any m , we have $|\bigcup_{r \geq m} K_{1/r}| = 1$ and therefore $|\Omega'| = 1$.

Now we check that Ω' has the desired properties. Thus, fix ω in Ω' , ϕ a measurable function on X , G a nilpotent Lie group of step less than κ , Γ a cocompact lattice and F' a function on G/Γ . Suppose that for μ_ω -almost every x in X , there exists x' in G/Γ such that $\phi(T^h x) = F'(g^h x')$ for some g in G . Fix $\varepsilon > 0$. We aim to show

$$\left| \int_X f(x)\phi(x)\mu_\omega(dx) \right| \lesssim \varepsilon \cdot (\|F'\|_{L^\infty} + 1).$$

Fix i in the natural numbers such that (G, Γ) is isomorphic to (G_i, Γ_i) . Fix $\psi : G_i \rightarrow G$ an isomorphism such that $\psi(\Gamma_i) = \Gamma$. Fix F in \mathcal{F}_i such that $\|F \circ \psi - F'\|_{L^\infty} \leq \varepsilon$. Then ω is in Ω' so ω is in $\Omega_{i,F}$ and therefore there exists $r > 1/\varepsilon$ such that ω is in $K_{1/r}$. Therefore, for some $H = H_{1/r}$,

$$\mu_\omega \left\{ x' \in X : \sup_{\substack{g \in G_i \\ x \in G_i/\Gamma_i}} |\mathbb{E}_{h \leq H} f(T^h x') F(g^h x)| \geq \varepsilon \right\} \leq \varepsilon.$$

By the triangle inequality,

$$\mu_\omega \left\{ x' \in X : \sup_{\substack{g \in G \\ x \in G/\Gamma}} |\mathbb{E}_{h \leq H} f(T^h x') F'(g^h x)| \geq 2\varepsilon \right\} \leq \varepsilon.$$

Next, we use that ϕ locally looks like F' :

$$\mu_\omega \{ x' \in X : |\mathbb{E}_{h \leq H} f(T^h x') \phi(T^h x')| \geq 2\varepsilon \} \leq \varepsilon.$$

Bounding the exceptional points by the L^∞ norm, we get that

$$\int_X |\mathbb{E}_{h \leq H} f(T^h x) \phi(T^h x)| \mu_\omega(dx) \lesssim \varepsilon \cdot (\|F'\|_{L^\infty} + 1).$$

By the triangle inequality,

$$\left| \int_X \mathbb{E}_{h \leq H} f(T^h x) \phi(T^h x) \mu_\omega(dx) \right| \lesssim \varepsilon \cdot (\|F'\|_{L^\infty} + 1).$$

By translation invariance,

$$\left| \int_X f(x) \phi(x) \mu_\omega(dx) \right| \lesssim \varepsilon \cdot (\|F'\|_{L^\infty} + 1).$$

This completes the proof. □

PROPOSITION 2.5. *Let (X, μ, T) be a (topologically) compact, invertible, not necessarily ergodic dynamical system. Let $\mu = \int_\Omega \mu_\omega d\omega$ be an ergodic decomposition. Recall that, for each ω , the Host–Kra factor \mathcal{Z}_ω is defined up to sets of μ_ω -measure 0. For each ω , fix such a Host–Kra factor. For instance, one could use any definition of the Host–Kra factor and then add all sets of μ_ω -measure 0 to obtain the complete Host–Kra factor. Then there exists a sigma algebra \mathcal{Z} on X such that, for any measurable set A , A is \mathcal{Z} measurable if and only if there exists a full measure subset $\Omega' \subset \Omega$ such that for all ω in Ω' , A is \mathcal{Z}_ω measurable. This implies that a function f in $L^\infty(\mu)$ is \mathcal{Z} measurable if and only if there exists a full measure subset $\Omega' \subset \Omega$ such that f is \mathcal{Z}_ω measurable for every ω in Ω' .*

Proof. Let \mathcal{Z} be the set of measurable subsets of X such that there exists a full measure set $\Omega_A \subset \Omega$ such that for all ω in Ω_A , A is \mathcal{Z}_ω measurable. For each such set, fix such an Ω_A . Let A_1, A_2, A_3, \dots be a countable list of sets in \mathcal{Z} . Consider

$$\Omega' = \bigcap_{i \in \mathbb{N}} \Omega_{A_i}.$$

Because Ω' is the intersection of countably many full measure sets, it has full measure. Let ω be an element of Ω' . Then for every natural number i , A_i is \mathcal{Z}_ω measurable. Because \mathcal{Z}_ω is a sigma algebra, that implies the countable intersection and countable union of the sets A_i are also \mathcal{Z}_ω measurable. Thus the intersection $\bigcap A_i$ and union $\bigcup A_i$ are both \mathcal{Z}_ω measurable for a full measure subset $\Omega' \subset \Omega$ and thus, by definition of \mathcal{Z} , \mathcal{Z} is closed under countable unions and intersections. If A is in \mathcal{Z} , then A is in \mathcal{Z}_ω for every ω in Ω_A . Since \mathcal{Z}_ω is a sigma algebra, the complement A^c is also in \mathcal{Z}_ω for every ω in Ω_A . By definition of \mathcal{Z} , we conclude that \mathcal{Z} is closed under complements. Obviously X and \emptyset are in \mathcal{Z} so \mathcal{Z} is a sigma algebra.

Lastly, we check that a function f in $L^\infty(\mu)$ is \mathcal{Z} measurable if and only if it is \mathcal{Z}_ω measurable for a full measure set of ω . First, suppose there exists a full measure subset $\Omega' \subset \Omega$ such that, for ω in Ω' , f is \mathcal{Z}_ω measurable. Let A be a measurable subset of \mathbb{C} . Then since f is \mathcal{Z}_ω measurable for any ω in Ω' , $f^{-1}(A)$ is in \mathcal{Z}_ω for any ω in Ω' . Therefore, by definition of \mathcal{Z} , $f^{-1}(A)$ is in \mathcal{Z} so f is \mathcal{Z} measurable.

Now suppose f is \mathcal{Z} measurable. We approximate f by simple functions f_i . For instance, we can take $f_i(x) = k \cdot 2^{-i}$ if $f(x)$ is between $k \cdot 2^{-i}$ and $(k + 1) \cdot 2^{-i}$ for any natural number k . Then $f_i \rightarrow f$ in $L^1(\mu)$ and also in $L^1(\mu_\omega)$ for any ω by the dominated convergence theorem. For each i , the function f_i has only finitely many distinct level sets. Because f is \mathcal{Z} measurable, the level sets of f_i are \mathcal{Z} measurable. Therefore, there exists

a full measure subset $\Omega_i \subset \Omega$ such that f_i is \mathcal{Z}_ω measurable for all ω in Ω_i . Let

$$\Omega' = \bigcap_{i \in \mathbb{N}} \Omega_i.$$

Then since Ω' is the intersection of sets of full measure, Ω' has full measure. For each ω in Ω' , f_i is \mathcal{Z}_ω measurable for all natural numbers i . But $f_i \rightarrow f$ in $L^1(\mu_\omega)$ so the limit f is also \mathcal{Z}_ω measurable for all ω in Ω' . □

Definition 2.6. By Proposition 2.5, there exists a sigma algebra \mathcal{Z} such that a $L^\infty(\mu)$ function f is \mathcal{Z} measurable if and only if it is \mathcal{Z}_ω measurable for almost every ω in Ω . We fix such a sigma algebra and call it the Host–Kra sigma algebra for (X, μ, T) .

PROPOSITION 2.7. *Let f be a function in $L^\infty(\mu)$. Then there exists a set Ω' of full measure in Ω such that for ω in Ω' ,*

$$\mathbb{E}^\mu[f|\mathcal{Z}] = \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega],$$

μ_ω -almost everywhere.

Proof. First, we need the following quick ergodic-theoretic fact. The space X can be essentially partitioned into pieces where each piece carries all the mass of an ergodic component. More precisely, there exists a map $\omega' : X \rightarrow \Omega$ such that

$$\int_\Omega \int_X \phi(x) \mu_\omega(dx) d\omega = \int_\Omega \int_X \phi(x) \mathbb{1}_{\omega=\omega'(x)} \mu_\omega(dx) d\omega$$

for any integrable ϕ . For instance, in the usual construction of an ergodic decomposition, one can take Ω to be the set of atoms of X with respect to the invariant sigma algebra \mathcal{I} . Then let $\int \psi(x') \mu_{[x]}(dx') = \mathbb{E}[\psi|\mathcal{I}](x)$ where x is any point in the atom $[x]$. In this case the map ω' just sends x to the atom containing x .

By Proposition 2.5, there is a set Ω_0 of full measure such that $\mathbb{E}^\mu[f|\mathcal{Z}]$ is \mathcal{Z}_ω measurable for every ω in Ω_0 . We also require that, for ω in Ω_0 ,

$$\|f\|_{L^\infty(\mu_\omega)} \leq \|f\|_{L^\infty(\mu)}$$

which holds for a full measure set of ω . Fix such an Ω_0 . Since X is compact, there exists a countable uniformly dense subset of the space of continuous functions. Fix such a subset and fix an order on that subset $f_1, f_2, f_3, f_4, \dots$. Again by Proposition 2.5, there exists a full measure subset Ω_i of Ω such that for ω in Ω_i , the function $\mathbb{E}^\mu[f_i|\mathcal{Z}]$ is \mathcal{Z}_ω measurable. Let

$$\Omega' = \bigcap_{i \geq 0} \Omega_i.$$

Since each Ω_i has full measure and there are only countably many choices of i , we conclude that Ω' has full measure. Now let ω be an element of Ω' and suppose for the sake of contradiction that

$$\mathbb{E}^\mu[f|\mathcal{Z}] \neq \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega],$$

meaning equality does not hold up to sets of μ_ω -measure 0. The conditional expectation is uniquely defined by two properties, namely that $\mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$ is \mathcal{Z}_ω measurable and that

$$\int_X \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega](x)\phi(x)\mu_\omega(dx) = \int_X f(x)\phi(x)\mu_\omega(dx),$$

for any \mathcal{Z}_ω measurable function ϕ in $L^\infty(\mu_\omega)$. If $\mathbb{E}^\mu[f|\mathcal{Z}]$ satisfies the same properties then $\mathbb{E}^\mu[f|\mathcal{Z}]$ equals $\mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$ μ_ω -almost everywhere. We know since ω is in Ω' which is contained in Ω_0 that $\mathbb{E}^\mu[f|\mathcal{Z}]$ is \mathcal{Z}_ω measurable. Therefore, there exists ϕ in $L^\infty(\mu_\omega)$ such that

$$\int_X \mathbb{E}^\mu[f|\mathcal{Z}](x)\phi(x)\mu_\omega(dx) \neq \int_X f(x)\phi(x)\mu_\omega(dx).$$

By subtracting off the appropriate multiple of $\mathbb{E}^\mu[f|\mathcal{Z}]$, we may assume that ϕ is μ_ω -orthogonal to $\mathbb{E}^\mu[f|\mathcal{Z}]$. Multiplying by a scalar, we may assume that $\langle f, \phi \rangle_{L^2(\mu_\omega)}$ is a positive real number greater than 1.

For each ω in Ω' such that $\mathbb{E}^\mu[f|\mathcal{Z}] \neq \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$, we showed there exists ϕ a \mathcal{Z}_ω measurable function such that $\langle \mathbb{E}^\mu[f|\mathcal{Z}], \phi \rangle_{L^2(\mu_\omega)} = 0$ and $\langle \phi, f \rangle_{L^2(\mu_\omega)} > 1$. Let ϕ be such a function. Suppose for the moment that $\|\phi\|_{L^2(\mu_\omega)} < C$. Since f_1, f_2, f_3, \dots are dense in $L^2(\mu_\omega)$, for any ε and for any power $p \geq 2$ we can find an i such that $\|\phi - f_i\|_{L^p(\mu_\omega)} \leq \varepsilon$. This implies, by Cauchy–Schwarz, that

$$\langle \mathbb{E}^\mu[f|\mathcal{Z}], f_i \rangle_{L^2(\mu_\omega)} \leq \varepsilon \|f\|_{L^\infty(\mu)} \tag{2}$$

and

$$\langle \phi, f_i \rangle_{L^2(\mu_\omega)} > 1 - \varepsilon. \tag{3}$$

We also need a quantitative way of saying that f_i is close to being \mathcal{Z}_ω measurable. One option is to use the Host–Kra norms defined for an ergodic system in [HK05, §3.5]. Let $\|\phi\|_{k,\omega}$ denote the k th Host–Kra norm. The key feature of the Host–Kra norms is that a function ϕ is \mathcal{Z}_ω measurable if and only if $\|\phi\|_{k,\omega} = 0$ for all k (see [HK05, Lemma 4.3]). We claim that $\|\phi\|_{k,\omega}$ is a measurable function of ω . After all, by definition $\|\phi\|_{k,\omega}^{2^k}$ is the integral of some fixed function on X^{2^k} , namely $(x_1, \dots, x_{2^k}) \mapsto \phi(x_1) \dots \phi(x_{2^k})$ with respect to some measure (namely $\mu_\omega^{[k]}$ defined in [HK05, §3.1]) which depends measurably on ω . Thus, we can find also f_i with

$$\|f_i\|_{k,\omega} \leq 2\varepsilon, \tag{4}$$

for all $k \leq 1/\varepsilon$. If we also know that

$$\|\phi\|_{L^p(\mu_\omega)} < C$$

for some constant C then also by the triangle inequality,

$$\|f_i\|_{L^p(\mu_\omega)} \leq C + \varepsilon. \tag{5}$$

Fix a constant C . Now we define a function $i : \Omega \times \mathbb{R}_{>0} \rightarrow \mathbb{N} \cup \{\infty\}$ as follows. Let $i(\omega, \varepsilon)$ be the first index such that all four inequalities (2)–(5) are satisfied with $p = 1/\varepsilon$ if such an i exists and $+\infty$ otherwise. Note that i implicitly depends on C . Let E denote the set of ω such that $i(\omega, \varepsilon)$ is finite for all ε . In particular, if $\mathbb{E}^\mu[f|\mathcal{Z}] \neq \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$ then $\omega \in E$.

is in this set for some choice of C . Thus, we may assume for the sake of contradiction that the measure of E is positive. Let

$$\psi_\varepsilon(x) = \begin{cases} f_{i(\omega,\varepsilon)}(x), & \omega \in E, \omega'(x) = \omega, \\ 0, & \text{otherwise.} \end{cases}$$

Since, for all $p < 1/\varepsilon$,

$$\|\psi_\varepsilon\|_{L^p(\mu)} = \int_\Omega \|\psi_\varepsilon\|_{L^p(\mu_\omega)} d\omega \leq C + \varepsilon,$$

we can take an $L^p(\mu)$ weak-* limit $\psi_\varepsilon \rightarrow \psi$ for some subsequence of epsilons tending to 0. By a diagonalization argument, we can ensure that this weak-* limit exists for all $p < \infty$. By (4), we conclude that ψ is \mathcal{Z}_ω measurable for each ω in E . If ω is not in E , then $\psi = 0$ on a set of μ_ω full measure, so ψ is measurable with respect to \mathcal{Z}_ω for a full measure set of ω in Ω , so by definition ψ is \mathcal{Z} measurable. Furthermore, by (3),

$$\langle \phi, \psi \rangle_{L^2(\mu_\omega)} \geq 1,$$

so we conclude that

$$\langle \phi, \psi \rangle_{L^2(\mu)} \geq |E|$$

by integrating in ω . On the other hand, by (2),

$$\langle \mathbb{E}^\mu[f|\mathcal{Z}], \psi \rangle_{L^2(\mu_\omega)} = 0.$$

This contradicts the definition of $\mathbb{E}^\mu[f|\mathcal{Z}]$. Thus,

$$\mathbb{E}^\mu[f|\mathcal{Z}] = \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega]$$

for almost every ω in Ω . □

A crucial input is the following theorem of [FH18a]. This theorem says that if a correlates with b then it does so for some algebraic reason. In particular, any correlation between f and f' is due solely to some locally algebraic structure in f .

THEOREM 2.8. ([FH18a, Theorem 1.5]; see also Appendix A) *Let μ be the first marginal of ν corresponding to the factor X . Then the ergodic components (X, μ_ω, T) of μ are isomorphic to the product of a Bernoulli system with the Host–Kra factor of (X, μ_ω, T) .*

To use this theorem, we need the following result, which essentially appears in [FH18b].

LEMMA 2.9. ([FH18b]; see the proof of Lemma 6.2) *Suppose that $(X, \mu_\omega, T) \cong (W, dw, T) \times (Z, dz, T)$ where W is a Bernoulli system, Z is a zero entropy system and μ_ω is the first marginal of ν_ω . Then for any function $\phi: X \rightarrow \mathbb{C}$ and any function $\psi: Y \rightarrow \mathbb{C}$ we have*

$$\int_{X \times Y} \phi(x)\psi(y)\nu_\omega(dx dy) = \int_{X \times Y} \mathbb{E}^{\nu_\omega}[\phi|Z](x)\psi(y)\nu_\omega(dx dy)$$

where $\mathbb{E}^{\nu_\omega}[\phi|Z]$ denotes the conditional expectation of ϕ with respect to the measure ν_ω and the sigma algebra of Z -measurable functions.

Proof. By density, it suffices to consider the case $\phi(w, z) = \phi_W(w)\phi_Z(z)$. Because any joining of the Bernoulli system W and the zero entropy system $Z \times Y$ is trivial (i.e. is equipped with the product measure), we can break up the integral

$$\begin{aligned} & \int_{W \times Z \times Y} \phi_W(w)\phi_Z(z)\psi(y)v_\omega(dw dz dy) \\ &= \int_W \phi_W(w)v_\omega(dw dz dy) \cdot \int_{Z \times Y} \phi_Z(z)\psi(y)v_\omega(dw dz dy) \\ &= \int_{X \times Y} \mathbb{E}^{v_\omega}[\phi|Z](z)\psi(y)v_\omega(dw dz dy). \end{aligned} \quad \square$$

We also need the following result, which says that conditional expectation is essentially local.

COROLLARY 2.10. *Let X, Y, v, f, f', c , and η be as in §1.1. Let B be as in Corollary 2.2. Then*

$$\left| \int_{B^c} \mathbb{E}_{h \leq k} T^h (\mathbb{E}^v[f|Z] \cdot f') dv \right| > c + \eta.$$

Proof. Recall that

$$\left| \int_{X \times Y} f(x) \cdot f'(y)v(dx dy) \right| > c + \eta.$$

By Corollary 2.2, we have that

$$\left| \int_{B^c} f(x) \cdot f'(y)v(dx dy) \right| > c + \eta.$$

Since B^c is T invariant and v is T invariant, we can average over shifts:

$$\left| \int_{B^c} \mathbb{E}_{h \leq k} f(T^h x) \cdot f'(T^h y)v(dx dy) \right| > c + \eta.$$

Next, we disintegrate the measure v ,

$$\left| \int_\Omega \int_{B^c} \mathbb{E}_{h \leq k} f(T^h x) \cdot f'(T^h y)v_\omega(dx dy) d\omega \right| > c + \eta.$$

Notice, for each h , THAT $f'(T^h y)\mathbb{1}_{y \notin B}$ is a function on Y . By Theorem 2.8, (X, μ_ω, T) is isomorphic to a product of a Bernoulli factor with the Host–Kra factor for almost every ω . Since the Host–Kra factor has entropy zero, the hypotheses of Lemma 2.9 are satisfied. Thus, by Lemma 2.9,

$$\left| \int_\Omega \int_{B^c} \mathbb{E}_{h \leq k} \mathbb{E}^{\mu_\omega}[f|Z_\omega](T^h x) \cdot f'(T^h y)v_\omega(dx dy) d\omega \right| > c + \eta.$$

By Proposition 2.7,

$$\left| \int_\Omega \int_{B^c} \mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|Z](T^h x) \cdot f'(T^h y)v_\omega(dx dy) d\omega \right| > c + \eta.$$

By definition of the ergodic decomposition,

$$\left| \int_{B^c} \mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^h x) \cdot f'(T^h y) \nu(dx dy) \right| > c + \eta.$$

This completes the proof. □

Now we forget everything about the joining of X and Y and reduce to the worst case scenario, where we choose the worst possible y in Y for each x in X .

COROLLARY 2.11. *Let $X, Y, \nu, \mu, f, f', c,$ and η be as in §1.1. Let \mathcal{Z} be as in Definition 2.6. Let B be as in Corollary 2.2. Since whether $(x, y) \in B$ only depends on y , we abuse notation and write $y \in B$ to mean $(x, y) \in B$ for some x . Then,*

$$\int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^h x) \cdot f'(T^h y)| \mu(dx) > c + \eta,$$

where the supremum is an essential supremum taken with respect to the second marginal of ν .

We will need the following lemma, which states that conditioning with respect to a conditional measure is essentially the same as conditioning with respect to the original measure.

LEMMA 2.12. *Let A be a positive measure set in \mathcal{Z} and denote $\mu_A(S) = \mu(S|A)$. Then for any measurable function f ,*

$$\mathbb{E}^{\mu_A}[f|\mathcal{Z}] = \mathbb{E}^\mu[f|\mathcal{Z}],$$

μ_A -almost everywhere (i.e. for μ -almost every point in A).

Proof. Let C be another set in \mathcal{Z} . Then

$$\int_C \mathbb{E}^\mu[f|\mathcal{Z}](x) \mu_A(dx) = \int_{C \cap A} \frac{1}{\mu(A)} \mathbb{E}^\mu[f|\mathcal{Z}](x) \mu(dx).$$

Since A is in \mathcal{Z} , we know that $A \cap C$ is in \mathcal{Z} . By definition of conditional expectation, this is

$$\begin{aligned} &= \frac{1}{\mu(A)} \int_{C \cap A} f \mu(dx) \\ &= \int_{C \cap A} f \mu_A(dx). \end{aligned}$$

This is the defining property of $\mathbb{E}^{\mu_A}[f|\mathcal{Z}]$. Since conditional expectation is well defined up to sets of measure 0, we obtain the result. □

The system X possesses an extra symmetry that most dynamical systems do not have, a dilation symmetry. In fact, it possesses a whole family of dilation symmetries. It is not obvious which dilation makes the problem easiest. Therefore, instead of choosing a particular dilation, we use a random dilation.

PROPOSITION 2.13. *Let P be any natural number. Then*

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})}(x) \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx) > c + \eta,$$

where p is always restricted to be prime.

Proof. By Corollary 2.11 we have

$$\int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^h x) \cdot f'(T^h y)| \mu(dx) > c + \eta.$$

Now we use that I_p pushes forward $p \mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})}\mu$ to μ for every p and average in p :

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^h I_p x) \cdot f'(T^h y)| \mu(dx) > c + \eta.$$

Because $I_p \circ T^{hp}(x) = T^h \circ I_p(x)$ for almost every x in $M^{-1}(p\widehat{\mathbb{Z}})$, we have that

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](I_p T^{ph} x) \cdot f'(T^h y)| \mu(dx) > c + \eta.$$

Next we use the standard fact that

$$\mathbb{E}^\mu[f|\mathcal{Z}] \circ I_p = \mathbb{E}^{I_{p*}\mu}[f \circ I_p | I_p^{-1}(\mathcal{Z})]$$

$I_{p*}\mu$ -almost everywhere, where $I_{p*}\mu$ is the pushforward of μ . Since $I_{p*}\mu = (1/p)\mu$, we can replace $I_{p*}\mu$ by μ . Note that I_p defines a factor map between $(M^{-1}(p\widehat{\mathbb{Z}}), p\mu, T^p)$ and (X, μ, T) . Since Host–Kra factors are functorial, the Host–Kra factor for $(M^{-1}(p\widehat{\mathbb{Z}}), p\mu, T^p)$ factors onto the Host–Kra factor for (X, μ, T) . Thus, $I_p^{-1}(\mathcal{Z})$ is contained in the Host–Kra factor of some dynamical system and thus corresponds to an inverse limit of nilsystems. This is all we actually need for our purposes. However, for the sake of avoiding notation, we also prove that

$$\mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \mathbb{E}^\mu[f \circ I_p | I_p^{-1}(\mathcal{Z})] = \mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \overline{a(p)} \mathbb{E}^\mu[f|\mathcal{Z}].$$

That $f \circ I_p = \overline{a(p)}f$ follows from the definition of I_p . If $\mathcal{Z}_i(T^p)$ denotes the i th Host–Kra factor for T^p and $\mathcal{Z}_i(T)$ denotes the i th Host–Kra factor for T , then any T^p invariant subset of the cube X^{2^i} is an element of the Kronecker factor, that is, the first Host–Kra factor for $(X^{2^i}, T, \mu^{[i]})$ (where $\mu^{[i]}$ is the measure on the cube defined in [HK05, §3]). Since the Host–Kra factor of an ergodic system is the smallest sigma algebra generating the invariant factor on the cube, we conclude that $\mathcal{Z}_i(T^p) \subset \mathcal{Z}_{i+1}(T)$ so $I_p^{-1}(\mathcal{Z}) \subset \mathcal{Z} \cap M^{-1}(p\widehat{\mathbb{Z}})$. In fact, as in Appendix A, the Host–Kra factor for X is a joining of the Host–Kra factor on the space of sequences $D^{\mathbb{Z}}$ and $\widehat{\mathbb{Z}}$. On the second factor, I_p acts by division by p . On the first factor, $I_p \circ T^p = T \circ I_p$ and so, on each ergodic component of the first factor, I_p acts by multiplication by p up to a possible translation. Multiplication by p is a local isomorphism of any nilmanifold that does not contain p torsion. However, by Corollary 2.3, f is already orthogonal to all p torsion. Thus,

$$\mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \mathbb{E}^\mu[f \circ I_p | I_p^{-1}(\mathcal{Z})] = \mathbb{1}_{M^{-1}(p\widehat{\mathbb{Z}})} \overline{a(p)} \mathbb{E}^\mu[f|\mathcal{Z}].$$

Combined with Lemma 2.12 and the fact that $M^{-1}(p\hat{\mathbb{Z}})$ is T^p invariant and therefore an element of \mathcal{Z} , we get

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu \overline{a(p)} [f | \mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx) > c + \eta.$$

Recall that $|a(p)| = 1$ for all p . Thus, $\overline{a(p)}$ merely gets absorbed into the absolute value:

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu [f | \mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx) > c + \eta. \quad \square$$

2.1. *The entropy decrement argument.* Next, we use the entropy decrement method to replace $p \mathbb{1}_{M^{-1}(\hat{\mathbb{Z}})}$ by its average, 1. This is essentially due to Tao, but because our statement is slightly different we reproduce the argument. For the definitions of entropy, conditional entropy, mutual information, and conditional mutual information, see [Tao16].

Let x' be a random variable distributed according to μ and fix a natural number P . From this, we get the following two random variables. Set $X_P = (x_1, \dots, x_{(k+1)P})$ where $x_i = \mathbb{E}^\mu [f | \mathcal{Z}](T^i x')$, and set $Y_P = (M(x') \bmod p)_{P/2 < p \leq P}$ in $\prod_{P/2 < p \leq P} \mathbb{Z}/p\mathbb{Z}$. Denote $Y_P \bmod p = y_p$ so that $Y_P = (y_p)_{P/2 < p \leq P}$. Note that Y_P is uniformly distributed in $\prod_{P/2 < p \leq P} \mathbb{Z}/p\mathbb{Z}$ and that the distribution of X_P is the same as the distribution of $T^i X_P$ for any i because μ is translation invariant. Technically, if $\mathbb{E}^\mu [f | \mathcal{Z}]$ takes infinitely many values then we will have to round $\mathbb{E}^\mu [f | \mathcal{Z}](T^i x')$ so that each x_i takes values in a finite set, but this slightly annoying detail may be delayed for the moment. We want to study the following integral:

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu [f | \mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx).$$

By translation invariance, this is equal to

$$\mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{i \leq P} \int_X p T^i \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu [f | \mathcal{Z}](T^{ph+i}x) \cdot f'(T^h y)| \mu(dx).$$

Notice that this is the expected value of some function of X_P and Y_P . In particular, we are interested in

$$\mathbb{E} (\mathbb{E}_{P/2 < p \leq P} f_{P,p}(X_P, y_p))$$

where $f_{P,p}: \mathbb{C}^{(k+1)P} \times \mathbb{Z}/p\mathbb{Z} \rightarrow \mathbb{C}$ is defined by the formula

$$f_{P,p}(X_P, y_p) = \mathbb{E}_{i \leq P} p \mathbb{1}_{y_p=i} \sup_{y \notin B} |\mathbb{E}_{h \leq k} x_{hp+i} f(Ty)|.$$

Define

$$f_P(X_P, Y_P) = \mathbb{E}_{P/2 < p \leq P} f_{P,p}(X_P y_P).$$

Thus, we are interested in

$$\mathbb{E} f_P(X_P, Y_P).$$

We would like to say that X_P and Y_P are very close to independent for some large choice of P . Let W_P be a random variable with the same distribution as Y_P but which is independent

of X_P . We would like to say that

$$\mathbb{E}[f_P(X_P, Y_P)] \approx \mathbb{E}[f_P(X_P, W_P)].$$

A property like this actually holds in a more general setting, which we take the liberty of stating now.

THEOREM 2.14. ([Tao16, §3]; see also [Mor18], [TT17b, Lemma 3.4 and Proposition 3.5], and [TT17a, §4]) *Let A be a finite set and let C be a natural number. For each P a power of 2, let $X_P = (x_1, \dots, x_{CP})$ be a sequence of random variables with x_i taking values in A and let Y_P be a random variable that is uniformly distributed in $\prod_{P/2 < p \leq P} \mathbb{Z}/p\mathbb{Z}$. We write $Y_P = (y_p)_{P/2 < p \leq P}$ where $y_p = Y_P \bmod p$. We further assume that for different values of P , the random variables Y_P are jointly independent, meaning $(y_p)_{p \leq P}$ is uniformly distributed in $\prod_{p \leq P} \mathbb{Z}/p\mathbb{Z}$ for all P that are powers of 2. Suppose that, for any natural numbers i and m such that $i + m \leq CP$, we have that the distribution of (x_1, \dots, x_m) is equal to the distribution of $(x_{i+1}, \dots, x_{i+m})$. Furthermore, suppose that for any P , any element b in $\prod_{p \leq P} \mathbb{Z}/p\mathbb{Z}$, and any S a measurable subset of \mathbb{C}^m ,*

$$\mathbb{P}((x_1, \dots, x_m) \in S \mid (y_p)_{p \leq P} = b) = \mathbb{P}((x_{i+1}, \dots, x_{i+m}) \in S \mid (y_p)_{p \leq P} = b + i).$$

For each p with $P/2 < p \leq P$, let $f_{P,p}$ be a 1-bounded function $A^{CP} \times \mathbb{Z}/p\mathbb{Z} \rightarrow \mathbb{C}$ and let $f_P(X_P, Y_P) = \mathbb{E}_{P/2 < p \leq P} f_{P,p}(X_P, y_p)$. Let W_P be a random variable with the same distribution as Y_P but which is independent of X_P . Then

$$\liminf_{P \rightarrow \infty} \mathbb{E}[|f_P(X_P, Y_P) - f_P(X_P, W_P)|] = 0.$$

Proof. Fix P a large power of 2 and $\varepsilon > 0$. By replacing $f_{P,p}(a, b)$ by $f_{P,p}(a, b) - f_{P,p}(a, W_P)$ we may assume that $f_{P,p}(a, W_P) = 0$ for all a . To prove the theorem, first we need a very good understanding of the case where X_P and Y_P are independent. In that case, even if we know the exact value of X_P , f_P is still a sum of independent random variables $f_{P,p}(a, W_P)$ and therefore exhibits concentration. This is formalized in Hoeffding’s inequality, which says that large collections of independent random variables exhibit concentration.

LEMMA 2.15. (Hoeffding’s inequality) *Suppose Z_1, \dots, Z_n are independent random variables taking values in $[-2, 2]$. Then*

$$\mathbb{P}(|Z_1 + \dots + Z_n - \mathbb{E}[Z_1 + \dots + Z_n]| > t) \leq \exp(-nt^2/16).$$

Let a be an element of A^{CP} . We apply Hoeffding’s inequality to the random variables $f_{P,p}(a, Y_P)$ (we remind the reader that there are roughly $P/(2 \log P)$ such terms, by the prime number theorem):

$$\mathbb{P}(|f_P(a, W_P)| > \varepsilon) \leq \exp\left(-\frac{\varepsilon^2 P}{40 \log P}\right). \tag{6}$$

Next, we aim to show that if Y_P is not necessarily independent of X_P but nearly independent of Y_P , we still can obtain a good bound. To do this, we use a Pinsker-type inequality.

LEMMA 2.16. [TT17a, Lemma 3.4] *Let Y be a random variable taking values in a finite set, let W be a uniformly distributed random variable on the same set, and let E be a set. Then*

$$\mathbb{P}(Y \in E) \leq -\frac{H(W) - H(Y) + \log 2}{\log \mathbb{P}(W \in E)}.$$

Let a be an element of A^{CP} . Let E be the set of b in $\prod_{P/2 < p \leq P} \mathbb{Z}/p\mathbb{Z}$ such that $|f_P(a, b)| > \varepsilon$. By (6), we know that

$$\mathbb{P}(W_P \in E) \leq \exp\left(-\frac{\varepsilon^2 P}{40 \log P}\right).$$

Applying Lemma 2.16 to $\mathbb{P}(\cdot | X_P = a)$, we find that

$$\mathbb{P}(|f(a, Y_P)| > \varepsilon | X_P = a) \leq -\frac{(H(W_P) - H(Y_P | X_P = a) + \log 2)40 \log P}{\varepsilon^2 P}.$$

Note that

$$\sum_a \mathbb{P}(X_P = a)(H(W_P) - H(Y_P | X_P = a)) = H(W_P) - H(Y_P | X_P) = I(X_P, Y_P),$$

where the last equality follows since $H(W_P) = H(Y_P)$ since the two random variables have the same distribution. Therefore, summing over a , we get

$$\mathbb{P}(|f(X_P, Y_P)| > \varepsilon) \leq -\frac{(I(X_P, Y_P) + \log 2)40 \log P}{\varepsilon^2 P}.$$

If

$$I(X_P, Y_P) \lesssim \frac{\varepsilon P}{\log P}$$

then

$$\mathbb{P}(|f(X_P, Y_P)| > \varepsilon) \lesssim \varepsilon.$$

This would complete the proof. Let $Y_{\leq P/2} = (y_p)_{p \leq P/2}$. Fix b' an element of $\prod_{p \leq P/2} \mathbb{Z}/p\mathbb{Z}$. Then we may repeat the previous argument with $\mathbb{P}(\cdot | Y_{\leq P/2} = b')$ to conclude that

$$\mathbb{P}(|f(X_P, Y_P)| > \varepsilon) \leq -\frac{(I(X_P, Y_P | Y_{\leq P/2}) + \log 2)40 \log P}{\varepsilon^2 P},$$

and therefore if

$$I(X_P, Y_P | Y_{\leq P/2}) \lesssim \frac{\varepsilon^3 P}{\log P}$$

then

$$\mathbb{P}(|f(X_P, Y_P)| > \varepsilon) \lesssim \varepsilon$$

and therefore

$$\mathbb{E}|f(X_P, Y_P)| \lesssim \varepsilon.$$

Let P_0 be a power of 2. We will try to show that there exists $P \geq P_0$ such that

$$\mathbb{E}|f(X_P, Y_P)| \lesssim \varepsilon.$$

This would complete the proof. Suppose not. Then

$$I(X_P, Y_P | Y_{\leq P/2}) \gg \frac{\varepsilon^3 P}{\log P},$$

for all $P \geq P_0$. By definition of mutual information,

$$H(X_P | Y_{\leq P}) = H(X_P | Y_{\leq P/2}) - I(X_P, Y_P | Y_{\leq P/2})$$

where $Y_{\leq P} = (y_p)_{p \leq P}$. By assumption, we have a lower bound for the mutual information

$$\leq H(X_P | Y_{\leq P/2}) - \frac{\varepsilon^3 P}{\log P}.$$

By subadditivity of entropy,

$$\leq H(X_{P/2} | Y_{\leq P/2}) + H(x_{CP/2+1}, \dots, x_{CP} | Y_{\leq P/2}) - \frac{\varepsilon^3 P}{\log P} \tag{7}$$

where $X_{P/2} = (x_1, \dots, x_{CP/2})$. Since $(x_1, \dots, x_{CP/2})$ has the same distribution as $(x_1, \dots, x_{CP/2})$, for any set S in $\mathbb{C}^{CP/2}$ and for any b' in $\prod_{p \leq P/2} \mathbb{Z}/p\mathbb{Z}$,

$$\mathbb{P}((x_1, \dots, x_{CP/2}) \in S | Y_{\leq P/2} = b') = \mathbb{P}((x_{CP/2+1}, \dots, x_{CP}) \in S | Y_{\leq P/2} = b' + CP/2).$$

Since the entropy of a random variable only depends on its distribution, we conclude that, for all b' ,

$$H(x_1, \dots, x_{CP/2} | Y_{\leq P/2} = b') = H(x_{CP/2+1}, \dots, x_{CP} | Y_{\leq P/2} = b' + CP/2).$$

Since $Y_{\leq P/2}$ is uniformly distributed, for all b' ,

$$\mathbb{P}(Y_{P/2} = b') = \mathbb{P}(Y_{P/2} = b' + CP/2).$$

Therefore, summing in b' ,

$$\begin{aligned} & H(x_1, \dots, x_{CP/2} | Y_{\leq P/2}) \\ &= \sum_{b'} \mathbb{P}(Y_{P/2} = b') H(x_1, \dots, x_{CP/2} | Y_{\leq P/2} = b') \\ &= \sum_{b'} \mathbb{P}(Y_{P/2} = b' + CP/2) H(x_{CP/2+1}, \dots, x_{CP} | Y_{\leq P/2} = b' + CP/2) \tag{8} \\ &= H(x_{CP/2+1}, \dots, x_{CP} | Y_{\leq P/2}). \end{aligned}$$

Applying (8) to (7),

$$H(X_P | Y_{\leq P}) \leq 2H(X_{P/2} | Y_{\leq P/2}) - \frac{\varepsilon^3 P}{\log P}.$$

We just obtained an upper bound for $H(X_P | Y_{\leq P})$. We can apply the same argument to obtain an upper bound for $H(X_{P/2} | Y_{\leq P/2})$:

$$\leq 4H(X_{P/4} | Y_{\leq P/4}) - 2 \frac{\varepsilon P/2}{\log P/2} - \frac{\varepsilon^3 P}{\log P}$$

where $X_{P/2} = (x_1, \dots, x_{CP/4})$ and where $Y_{\leq P/4} = (y_p)_{p \leq P/4}$. Applying this argument inductively, if $P = 2^m \cdot P_0$ then

$$\leq 2^m \left(H(X_{P_0} | Y_{\leq P_0}) - \varepsilon \sum_{j=1}^m \frac{P_0}{j} \right). \tag{9}$$

However, $\sum_{m \leq \log_2 P/P_0} (1/m) \sim \log \log P$ so, for large P ,

$$\varepsilon \sum_{j=1}^m \frac{P_0}{j} \gg CP_0 \log |A| \geq H(X_{P_0} | Y_{\leq P_0}).$$

Combining this with (7),

$$H(X_P, Y_{\leq P}) < 0$$

which is impossible. □

Applying Theorem 2.14 to our situation yields the following corollary.

COROLLARY 2.17. *Let $X, Y, \mu, f, f', M, I_p, c$, and η be as in §1.1. Let \mathcal{Z} be as in Definition 2.6. Let B be as in Corollary 2.2. We have*

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \sup_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu [f | \mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx) > c.$$

Proof. Recall that, for all natural numbers P ,

$$\mathbb{E}_{P/2 < p \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}})} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu [f | \mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx) > c + \eta.$$

By translation invariance, for all natural numbers P ,

$$\mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{i \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}}+i)} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu [f | \mathcal{Z}](T^{ph+i}x) \cdot f'(T^h y)| \mu(dx) > c + \eta.$$

Let x' be a random variable with distribution μ . Fix $\varepsilon > 0$ small. We will require that $\varepsilon < 10 \cdot \eta$. Let ϕ be a measurable function on X which uniformly approximates $\mathbb{E}^\mu [f | \mathcal{Z}]$, that is,

$$\|\phi - \mathbb{E}^\mu [f | \mathcal{Z}]\|_{L^\infty} < \varepsilon.$$

For instance, $\phi(x)$ could be obtained by rounding $\mathbb{E}^\mu [f | \mathcal{Z}](x)$ to the closest element of $(\varepsilon/10) \cdot \mathbb{Z}[i]$. By the triangle inequality,

$$\mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{i \leq P} \int_X p \mathbb{1}_{M^{-1}(p\hat{\mathbb{Z}}+i)} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu \phi(T^{ph+i}x) \cdot f'(T^h y)| \mu(dx) > c + \eta - \varepsilon.$$

For each natural number P , let $X_P = (x_1, \dots, x_{CP})$ where $x_i = \phi(T^i x')$ and where $C = k + 1$. Let $Y_P = (y_p)_{P/2 < p \leq P}$ where $y_p = M(x') \bmod p$. For each natural number P , let

$$f_{P,p}(X_P, y_p) = \mathbb{E}_{i \leq P} p \mathbb{1}_{y_p=i} \sup_{y \notin B} |\mathbb{E}_{h \leq k} x_{hp+i} f(Ty)|.$$

Define

$$f_P(X_P, Y_P) = \mathbb{E}_{P/2 < p \leq P} f_{P,p}(X_P Y_P).$$

Unpacking definitions, for every natural number P ,

$$\mathbb{E} f_P(X_P, Y_P) > c + \eta - \varepsilon.$$

Now we check the hypotheses of Theorem 2.14. Because ϕ takes only finitely many values, x_i takes values in a finite set. For all natural numbers P , since the distribution of $(y_p)_{p \leq P}$ is a $+1$ invariant measure on $\prod_{p \leq P} \mathbb{Z}/p\mathbb{Z}$, it must be the uniform distribution. Since μ is translation invariant, for any natural numbers i and m and any subset E of X^m ,

$$\mathbb{P}((Tx', \dots, T^m x') \in E) = \mathbb{P}((T^{i+1}x', \dots, T^{i+m}x') \in E).$$

Applying this to the preimage under (ϕ, \dots, ϕ) of an arbitrary subset S of \mathbb{C}^m reveals that the distribution of (x_1, \dots, x_m) is the same as the distribution of $(x_{i+1}, \dots, x_{i+m})$. Similarly, if b is an element in $\prod_{p \leq P} \mathbb{Z}/p\mathbb{Z}$ and if E is the preimage under (ϕ, \dots, ϕ) of an arbitrary set S intersected with the set of points z in X such that $M(z) = b \pmod{\prod_{p \leq P} p}$ then we conclude that

$$\mathbb{P}((x_1, \dots, x_m) \in S | (y_p)_{p \leq P} = b) = \mathbb{P}((x_{i+1}, \dots, x_{i+m}) \in S | (y_p)_{p \leq P} = b + i).$$

For each natural number P and each prime $P/2 < p \leq P$, for at most two values of $i \leq P$ is it true that $y_p = i \pmod p$. Therefore, at most two terms in the sum

$$\mathbb{E}_{i \leq P} p \mathbb{1}_{y_p=i} \sup_{y \notin B} |\mathbb{E}_{h \leq k} x_{hp+i} f(Ty)|$$

are non-zero. Therefore $f_{P,p}$ is bounded by 2. Let W_P be a random variable with the same distribution as Y_P . Then by Theorem 2.14,

$$\liminf_{P \rightarrow \infty} \mathbb{E}[|f_P(X_P, Y_P) - f_P(X_P, W_P)|] = 0.$$

Since, for any natural number P ,

$$\mathbb{E} f_P(X_P, Y_P) > c + \eta - \varepsilon,$$

we conclude that

$$\limsup_{P \rightarrow \infty} \mathbb{E} f_P(X_P, W_P) > c + \eta - \varepsilon.$$

Unpacking definitions, this proves

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{i \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \phi(T^{ph+i}x) \cdot f'(T^h y)| \mu(dx) > c + \eta - \varepsilon.$$

By the triangle inequality,

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{i \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|Z](T^{ph+i}x) \cdot f'(T^h y)| \mu(dx) > c + \eta - 2\varepsilon.$$

Since ε was arbitrary,

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \mathbb{E}_{i \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|Z](T^{ph+i}x) \cdot f'(T^h y)| \mu(dx) > c.$$

By translation invariance,

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx) > c.$$

This completes the proof. □

2.2. *Nilsystems and algebraic structure.* Now we want to use [HK05] to show that $\mathbb{E}^\mu[f|\mathcal{Z}]$ has some local algebraic structure. This algebraic structure makes $\mathbb{E}^\mu[f|\mathcal{Z}]$ much easier to understand than f .

PROPOSITION 2.18. *Let ω be an element of Ω such that*

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > c.$$

Then for almost all such choices for ω , there exists a collection of nilsystems $(G(j)/\Gamma(j), dx, g(j), \mathcal{B})$, 1-bounded functions F_j , and factor maps $\psi_j: X \rightarrow G(j)/\Gamma(j)$ so that F_j is a nilcharacter on $G(j)/\Gamma(j)$ with frequency non-trivial on the identity component and such that, after identifying F_j with a function on X , we have that $\sum F_j = F$ satisfies $\|F\|_\infty \leq 1$ and

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} F(T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > c.$$

Proof. We are given that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > c.$$

Recall that by Lemma 2.7, we know that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^{\mu_\omega}[f|\mathcal{Z}_\omega](T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > c.$$

By [HK05] Theorem 10.1, $(X, \mu_\omega, T, \mathcal{Z})$ is isomorphic to an inverse limit of nilsystems. Therefore, there exist a nilsystem $(G/\Gamma, dx, g, \mathcal{B})$ and a factor map $\psi: X \rightarrow G/\Gamma$ such that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^{\mu_\omega}[f|\psi^{-1}(\mathcal{B})](T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > c.$$

We denote $F = \mathbb{E}^{\mu_\omega}[f|\psi^{-1}(\mathcal{B})]$. By a Fourier decomposition, we may write F as a sum of nilcharacters, $F = \sum_\xi F_\xi$. For each ξ , either ξ is non-trivial on the identity component of G/Γ or ξ is trivial on the identity component. If ξ is trivial on the identity component and the step s of G is greater than 1, then ξ is actually trivial on G_s . That is because, for any σ in G , the multiplication by σ map $\sigma: G/\Gamma \rightarrow G/\Gamma$ is continuous so it takes components to components. Let σ_* : components of $G \rightarrow$ components of G be the induced map on components and let τ be any other element of G . Then if σ and τ are in the same component of G then for any σ' in G , multiplication by σ' on the right is also continuous, so $\sigma\sigma'$ is in the same component as $\tau\sigma'$ so $\sigma_* = \tau_*$. We return to the

general case where σ and τ are not necessarily in the same component. Also note that, for any element γ in Γ , $(\gamma\sigma)_* = [\gamma, \sigma]_* \sigma_* \gamma_* = [\gamma, \sigma]_* \sigma_*$. Pick n, m, γ and γ' such that $g^n \gamma$ is in the same component as σ and $g^m \gamma'$ is in the same component as τ . Thus $[\sigma, \tau]_* = [g^n \gamma, g^m \gamma']_* = \pi_* [g^n, g^m]_*$ where π is an element of higher order. Of course $[g^n, g^m] = e$ and by induction we get that $[\sigma, \tau]_*$ is the identity and therefore $[\sigma, \tau]\gamma$ is in the identity component for some γ . Therefore, if $s > 1$, the function F_ξ descends to a function on $(G/G_s)/(\Gamma/\Gamma_s)$. By induction, we can almost prove the theorem, namely we can find a collect of nilsystems $(G(j)/\Gamma(j), dx, g(j), \mathcal{B})$ and functions F_j and factor maps $\psi_j: X \rightarrow G(j)/\Gamma(j)$ so that F_j is a nilcharacter on $G(j)/\Gamma(j)$ with frequency non-trivial on the identity component or $G(j)$ is abelian and such that, after identifying F_j with a function on X , we have that $\sum F_j = F$ satisfies $\|F\|_\infty \leq 1$ and

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} F(T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > c.$$

It remains to observe that the case of a locally constant function on an abelian group cannot occur by Corollary 2.3 as follows: we can think of the F_j s as all functions on the group G with some additional equivariance properties; by construction the different F_j s have different frequencies so if F_r is a locally constant function on an abelian group and thus is locally periodic, meaning $F_r(T^h x)$ is a periodic function of h , then by Corollary 2.3,

$$0 = \int_X f \cdot \overline{F_r} \mu_\omega(dx).$$

Since F_r is $\psi^{-1}(\mathcal{B})$ measurable,

$$= \int_{G/\Gamma} F \cdot \overline{F_r} dx.$$

Since all the F_j s have different frequencies, they are all orthogonal to each other,

$$= \int_{G/\Gamma} F_r \cdot \overline{F_r} dx. \quad \square$$

Remark 2.19. If we also know the $(\kappa - 1)$ -Fourier uniformity conjecture then the step of all nilpotent Lie groups is at least κ by Proposition 2.4 (plugging $\overline{F_r} = \phi$ into the statement of that proposition).

COROLLARY 2.20. *There exist a natural number L independent of k , a nilpotent Lie group G of step s , a cocompact subgroup Γ , an ergodic element g in G , and a nilcharacter Φ with non-trivial frequency even when restricted to the identity component,*

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)| dx > \frac{c}{L},$$

where \mathcal{K} as defined in §1.1 is an infinite set such that for k in \mathcal{K} , the number of words of length k of f' is $o(k^2)$ if $t = 2$ or $O(k^{t-\varepsilon})$ if $t \neq 2$ for some ε .

Proof. By Corollary 2.17,

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu(dx) > c.$$

Thus, for a positive measure set of ω ,

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > c.$$

By Proposition 2.18, we know that for almost every ω ,

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \sum_j F_j(T^{ph}x) \cdot f'(T^h y) \right| \mu_\omega(dx) > c,$$

where F_j is as in Proposition 2.18. Fix such an ω .

Since the sum $\sum_j F_j$ converges in $L^2(\mu_\omega)$, there exists a natural number L independent of k such that $\|\sum_{j \leq L/2} F_j - \sum_j F_j\|_{L^2(\mu_\omega)} < c/2$. By the triangle inequality

$$\begin{aligned} & \limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \sum_{j \leq L/2} F_j(T^{ph}x) \cdot f'(T^h y) \right| \mu_\omega(dx) \\ & + \int_X \sup_{y \notin B} \mathbb{E}_{h \leq k} \left| \sum_{j > L/2} F_j(T^{ph}x) \cdot f'(T^h y) \right| \mu_\omega(dx) > c. \end{aligned}$$

The second term is bounded by

$$\int_X \left| \sum_{j > L/2} F_j(x) \right| \mu_\omega(dx),$$

using that μ_ω is shift invariant and f' is 1-bounded. By Cauchy–Schwarz, this term is bounded by $\|\sum_{j \leq L/2} F_j - \sum_j F_j\|_{L^2(\mu_\omega)} < c/2$. Thus, by the triangle inequality,

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} \left| \mathbb{E}_{h \leq k} \sum_{j \leq L/2} F_j(T^{ph}x) \cdot f'(T^h y) \right| \mu_\omega(dx) > \frac{c}{2}.$$

By the pigeonhole principle, there exists some F_j such that

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} F_j(T^{ph}x) \cdot f'(T^h y)| \mu_\omega(dx) > \frac{c}{L}.$$

Renaming everything gives the conclusion. We remark that the corollary just stated that such L, G, Γ, g , and Φ exist and therefore the statement of the corollary allows L to depend on G, Γ , and all the other data that comes from ω . The remainder of the argument essentially takes place inside a single ergodic component and so how the constants vary from component to component is not important for our purposes. □

For the remainder of §2, we fix G, Γ, g , and Φ . We let $c_0 = c/L$. For the next few pages, we fix an integer k in \mathcal{K} such that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)| dx > c_0.$$

We will later send k to infinity. The following lemma does two things. First, it uses Hölder’s inequality to raise the exponent of $|\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)|$. We want this term raised to an even power because we want to expand out the product and get rid of the absolute values which are less ‘algebraic’ and therefore harder to understand directly using the theory of nilpotent Lie groups. We also want this even power to be larger the more oscillatory our function Φ is. This is because the more Φ oscillates, the more cancelation we expect in larger and larger products. The larger the power we use, the smaller the fraction of terms which do not exhibit cancelation is. Second, we use the pigeonhole principle. This lemma and the following lemma are where we make essential use of our bound on the word growth rate of b .

LEMMA 2.21. *Recall that b had at most $k^{1-\varepsilon}$ words of length k occurring with positive upper logarithmic density for k in \mathcal{K} , or b has $o(k^2)$ many words of length k that occur with positive upper logarithmic density if $t = 2$. Fix δ a constant that is small even when compared to c_0 . Then for each k in \mathcal{K} there is a word $\epsilon = (\epsilon_1, \dots, \epsilon_k)$ of length k such that*

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > k^{-t+\varepsilon} c_0^{2t},$$

for $t \neq 2$, and when $t = 2$ we have

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > \delta^{-1} k^{-t} c_0^{2t}.$$

Proof. We know that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)| dx > c_0.$$

By Hölder’s inequality, we have

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)|^{2t} dx > c_0^{2t}.$$

Because each term is non-negative, we can replace the essential sup by a sum over words that occur with positive log-density:

$$\sum_{\epsilon} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > c_0^{2t}.$$

We assumed that the number of words occurring with positive logarithmic density and therefore the number of terms in the sum is at most δk^t when $t = 2$ or $k^{t-\varepsilon}$ when $t \neq 2$. By the pigeonhole principle, when $t = 2$ there is a word such that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > \delta^{-1} k^{-t} c_0^{2t},$$

and similarly for $t \neq 2$ we have

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > k^{-t+\varepsilon} c_0^{2t},$$

which completes the proof. □

We need a slightly different estimate for the abelian case. The key to the next lemma is the idea that if $e(\alpha h)$ correlates with ϵ_h for $h \leq k$ then $e(\alpha h)$ also must correlate with translates of ϵ of size $\sim k$. Thus, in the abelian case, the previous lemma is rather lossy. When we replace the sup by a sum, we should gain an extra power of k .

LEMMA 2.22. *For $t = 2$, for all k in \mathcal{K} , there is a word ϵ such that*

$$\begin{aligned} & \limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell} x) \cdot \epsilon_h|^{2t-2} dx \\ & > \frac{c_0^{2t-2}}{9} \cdot \left\lfloor \frac{c_0^{2t-2} k}{6} \right\rfloor \delta^{-1} k^{-t}. \end{aligned}$$

Proof. Again, we know that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph} x) \cdot f'(T^h y)| dx > c_0.$$

By Holder's inequality,

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph} x) \cdot f'(T^h y)|^{2t-2} dx > c_0^{2t-2}.$$

Again we want to replace $F'(T^h y)$ by a sum over words. Let P be a number satisfying

$$\mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph} x) \cdot f'(T^h y)|^{2t-2} dx > c_0^{2t-2}.$$

Let A be the set of x such that

$$\mathbb{E}_{P/2 < p \leq P} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph} x) \cdot f'(T^h y)|^{2t-2} > \frac{2c_0^{2t-2}}{3}.$$

Therefore, the measure of A is at least $c^{2t-2}/3$. We want to show that for μ -almost every x in A , there are at least $\lfloor c^{2t-2} k/6 \rfloor$ many distinct words of f' such that

$$\mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell} x) \cdot f'(T^h y)|^{2t-2} > \frac{c_0^{2t-2}}{3}.$$

Let y be an element of B^c such that the words of $f'(T^h y)$ are words of Y and such that

$$\mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph} x) \cdot f'(T^h y)|^{2t-2} > \frac{2c_0^{2t-2}}{3}.$$

Denote by $\epsilon(m)$ the word of length k whose h th entry is $\epsilon(m)_h = f'(T^{m+h} y)$. If the words $\epsilon(m)$ are distinct for $m = 1, \dots, \lfloor c_0^{2t} k/6 \rfloor$ then by the triangle inequality,

$$\begin{aligned} & \mathbb{E}_{P/2 < p \leq P} \sup_{y \notin B} |\mathbb{E}_{h \leq k} \Phi(g^{ph+pm} x) \cdot \epsilon(m)_h|^{2t-2} \\ & = \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph+pm} x) \cdot f'(T^{h+pm} y)|^{2t-2}. \end{aligned}$$

We note that all but $2mk$ terms in the average are the same if we replace $\Phi(g^{ph+pm}x) \cdot f'(T^{h+pm}y)$ with $\Phi(g^{ph}x) \cdot f'(T^h y)$. Thus

$$\begin{aligned} &\geq \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)|^{2t-2} - \frac{2m}{k} \\ &\geq \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)|^{2t-2} - \frac{c_0^{2t-2}}{3}. \end{aligned}$$

Suppose for a moment that instead the words $\epsilon(m)$ are not distinct for $m = 1, \dots, \lfloor c_0^{2t-2}k/6 \rfloor$. Then there exists a minimum j such that $\epsilon(1), \dots, \epsilon(j)$ are not distinct. Fix such a j for the remainder of the proof. Thus, there exists some $1 \leq d < j$ such that $\epsilon(j) = \epsilon(j - d)$. We claim that $\epsilon(j - d)$ is d -periodic: that is because $\epsilon(j - d)_h = \epsilon(j)_h = \epsilon(j - d)_{h+d}$. Furthermore, if $\epsilon(j - d - 1)_1 = \epsilon(j - 1)_1$ then since $\epsilon(j - d - 1)_h = \epsilon(j - d)_{h-1} = \epsilon(j)_{h-1} = \epsilon(j - 1)_h$ for all $h > 1$, we clearly have $\epsilon(j - d - 1) = \epsilon(j - 1)$ and j is not minimal. For the rest of the proof, let r be the minimum number such that $r \geq j - d$ and $\epsilon(r)$ is not d periodic. For y not in B , we can find such an r because $f'(T^h y)$ is not eventually periodic. Since $\epsilon(r)$ is not d periodic but $\epsilon(r - 1)$ is d periodic and is equal to $\epsilon(q)$ for some q between $j - d$ and $j - 1$, we have that $\epsilon(r)_k \neq \epsilon(r)_{k-d}$ but $\epsilon(r)_h = \epsilon(r)_{h-d}$ for all other $h \leq k$. We claim that the words $1, \dots, j - 1$ and $r, \dots, r + \lfloor c_0^{2t-2}k/6 \rfloor - j + 1$ are all distinct. The reason is that for all m between 1 and $j - d$, we have that $\epsilon(m)_h = \epsilon(j - d)_{h+m-j+d}$ for all $h > m - j + d$ and precisely no larger range of h , and for all m between r and $r + \lfloor c_0^{2t-2}k/6 \rfloor - j + 1$ we have that $\epsilon(m)_h = \epsilon(j - d)_{h+m-r}$ for all $1 \leq h < k - m + r$ and precisely no larger range of h . For m between $j - d$ and $j - 1$, $\epsilon(m)$ is d periodic but because j was the minimal natural number such that $\epsilon(1), \dots, \epsilon(j)$ are not distinct, we have that the $\epsilon(m)$ for m between $j - d$ and $j - 1$ are still distinct. For m between 1 and $j - d$ and m' between r and $r + \lfloor c_0^{2t-2}k/6 \rfloor - j + d$ we have that the intervals $h > m - j + d$ and $h < k - m + r$ meet so the previous argument shows that $\epsilon(m) \neq \epsilon(m')$. A similar triangle inequality computation shows that for m between r and $r + \lfloor c_0^{2t-2}k/6 \rfloor - j + 1$ we still have

$$\begin{aligned} &\mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot \epsilon(m)_h|^{2t-2} \\ &= \mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot f'(T^{h+m}y)|^{2t-2} \\ &\geq \mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot f'(T^{h+r-1}y)|^{2t-2} - \frac{2(m - r + 1)}{k}. \end{aligned}$$

Next, we use that $\epsilon(r - 1) = \epsilon(q)$ for some q between $j - d$ and $j - 1$:

$$\begin{aligned} &\geq \mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot f'(T^{h+q}y)|^{2t-2} - \frac{2(m - r + 1)}{k} \\ &\geq \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot f'(T^h y)|^{2t-2} - \frac{c_0^{2t-2}}{3}. \end{aligned}$$

This proves the claim that for x in A there are at least $\lfloor c_0^{2t-2}k/6 \rfloor$ distinct words ϵ of f' such that

$$\mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot \epsilon_h|^{2t-2} > \frac{c_0^{2t-2}}{3}.$$

Summing over words we get that for almost every x in A ,

$$\begin{aligned} & \sum_{\epsilon \text{ a word of } f'} \mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot \epsilon_h|^{2t-2} \\ & > \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor \cdot \frac{c_0^{2t-2}}{3}. \end{aligned}$$

Next, we use that $\mu(A) > c_0^{2t-2}/3$:

$$\begin{aligned} & \int_X \sum_{\epsilon \text{ a word of } f'} \sup_{\ell \in \mathbb{N}} \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot \epsilon_h|^{2t-2} dx \\ & > \frac{c_0^{2t-2}}{3} \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor \cdot \frac{c_0^{2t-2}}{3}. \end{aligned}$$

Sending P to infinity and using the pigeonhole principle, we deduce that for some word ϵ ,

$$\begin{aligned} & \limsup_{P \rightarrow \infty} \int_X \sup_{\ell \in \mathbb{N}} \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot \epsilon_h|^{2t-2} dx \\ & > \frac{c_0^{2t-2}}{9} \cdot \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor \delta^{-1} k^{-t}. \quad \square \end{aligned}$$

Remark 2.23. For G abelian and therefore Φ a character, we have

$$\begin{aligned} & \limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot \epsilon_h|^{2t-2} dx \\ & = \limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t-2} dx. \end{aligned}$$

Therefore, by choosing δ sufficiently small, in the abelian case, we get

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \epsilon_h \Phi(g^{ph}x)|^2 dx \gg k^{-1}.$$

The next theorem contradicts the previous two lemmas and proves Theorem 1.8. In its proof, we rely heavily on [BDG16, Fra17, GTZ12, GT12a, GT10].

THEOREM 2.24. *Recall that, after Corollary 2.20, we fixed a nilpotent Lie group G , a cocompact lattice Γ , a nilcharacter Φ with non-trivial frequency on the identity component ξ , and an element g which acts ergodically on G/Γ , such that $\|\Phi\|_{L^\infty_x} = 1$. Recall that the step s of G is at least κ , where $t = \binom{\kappa+1}{2}$. Let ϵ be a sequence of words implicitly depending on k . Let $\varepsilon > 0$. Then*

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \epsilon_h \Phi(g^{ph}x)|^{2t} dx \cdot k^{t-\varepsilon} = 0.$$

If $t = 2$ then we do not need the epsilon loss and instead get the estimate

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \epsilon_h \Phi(g^{ph}x)|^{2t} dx \cdot k^t \leq C_s.$$

This contradicts Lemmas 2.21 and 2.22 as follows. When G is abelian and thus $t = 2$, Lemma 2.22 states that there is a word ϵ of length k such that

$$\begin{aligned} & \limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} \sup_{\ell \in \mathbb{N}} |\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \cdot \epsilon_h|^{2t-2} dx \\ & > \frac{c_0^{2t-2}}{9} \cdot \left\lfloor \frac{c_0^{2t-2}k}{6} \right\rfloor \delta^{-1} k^{-t}, \end{aligned}$$

for any δ we choose so long as k is chosen from the set \mathcal{K} of natural numbers such that f' has fewer than δk^2 words of length k . Thus, picking δ small (in particular, smaller than, say, $C_1^{-1} 100c_0^4$), we find that

$$\limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t-2} dx > k^{-t} C_1,$$

contradicting Theorem 2.24. (Note that we have replaced $|\mathbb{E}_{h \leq k} \Phi(g^{ph+\ell}x) \epsilon_h|$ by the same expression without the shift in ℓ as in Remark 2.23.) Similarly, if the group is not abelian, Lemma 2.21 states that there exists a word ϵ of length k such that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > k^{-t+\epsilon} c_0^{2t},$$

for $t \neq 2$, and when $t = 2$ we have

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > \delta^{-1} k^{-t} c_0^{2t}.$$

When $t = 2$, again by picking δ small, this time smaller than $C_s^{-1} c_0^{2t}$, we have that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > C_s k^{-t},$$

again contradicting Theorem 2.24. Finally, when $t \neq 2$,

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x) \cdot \epsilon_h|^{2t} dx > k^{-t+\epsilon} c_0^{2t}$$

contradicts

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \int_X \mathbb{E}_{P/2 < p \leq P} |\mathbb{E}_{h \leq k} \epsilon_h \Phi(g^{ph}x)|^{2t} dx \cdot k^{t-\epsilon} = 0$$

from Theorem 2.24.

Thus, the rest of this section is devoted to showing that Theorem 2.24 is true. Suppose not, and for the moment fix k in \mathcal{K} such that

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \epsilon_h \Phi(g^{ph}x)|^{2t} dx \gg k^{-t+\epsilon}.$$

The first step is to replace averages over primes by uniform averages over natural numbers. To do this, we need the machinery of Green and Tao [GT12a, GT10] and Green, Tao, and Ziegler [GTZ12]. By the triangle inequality, we may replace averages over primes by averages weighted by the von Mangoldt function. We have

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < n \leq P} \int_X \Lambda(n) |\mathbb{E}_{h \leq k} \epsilon_h \Phi(g^{ph}x)|^{2t} dx \gg k^{-t+\epsilon}.$$

We denote $\psi_x(m) = \Phi(g^m x)$. We expand:

$$\limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < n \leq P} \int_X \Lambda(n) \mathbb{E}_{J \in [k]^{2t} \in J} \psi_x(nj_1) \cdots \psi_x(nj_t) \cdot \overline{\psi_x(nj_{t+1})} \cdots \overline{\psi_x(nj_{2t})} dx \gg k^{-t+\epsilon},$$

where ϵ_j is a phase given by the formula $\epsilon_J = \epsilon_{j_1} \cdots \epsilon_{j_t} \cdot \overline{\epsilon_{j_{t+1}}} \cdots \overline{\epsilon_{j_{2t}}}$.

We say that $J \in [k]^s$ is diagonal if $\#\{m \leq s : j_m = h\} = \#\{m > s : j_m = h\}$ for all $h \leq k$. We say that J solves the Vinogradov mean value problem if, for all m between 1 and s , we have

$$j_1^m + \cdots + j_t^m = j_{t+1}^m + \cdots + j_{2t}^m.$$

Every diagonal J also solves Vinogradov’s mean value problem. We rely on the following Theorem due to Bourgain, Demeter, and Guth which says that those account for ‘most’ solutions, up to a constant.

THEOREM 2.25. [BDG16, Theorem 1.1] *For all ϵ and s there exists a constant $C_{s,\epsilon}$ such that the number of solutions to the Vinogradov mean value problem is less than $C_{s,\epsilon} k^{t+\epsilon}$ where $t \leq \binom{s+1}{2}$.*

We will show that if J does not solve Vinogradov’s mean value problem then J does not contribute to the sum. Thus, fix J which does not solve Vinogradov’s mean value theorem and suppose that

$$\limsup_{P \rightarrow \infty} \left| \mathbb{E}_{P/2 < n \leq P} \int_X \Lambda(n) \psi_x(nj_1) \cdots \psi_x(nj_t) \cdot \overline{\psi_x(nj_{t+1})} \cdots \overline{\psi_x(nj_{2t})} dx \right| \gtrsim_{k,s,c} 1.$$

We denote

$$\Psi_x(n) = \psi_x(nj_1) \cdots \psi_x(nj_t) \cdot \overline{\psi_x(nj_{t+1})} \cdots \overline{\psi_x(nj_{2t})}.$$

Fix a subsequence such that

$$\lim_{P \in I} \left| \mathbb{E}_{P/2 < n \leq P} \int_X \Lambda(n) \Psi_x(n) dx \right| = \limsup_{P \rightarrow \infty} \left| \mathbb{E}_{P/2 < n \leq P} \int_X \Lambda(n) \Psi_x(n) dx \right| \gtrsim_{k,s,c} 1,$$

where I is some infinite subset of the natural numbers and where the implied constant may depend on Ψ_x . Fix a large number W , a product of many small primes. We will later choose exactly how large W must be. We pass to a subsequence where the following limit

exists for each $b \leq W$:

$$\lim_{P \in I'} \left| \mathbb{E}_{P/2 < Wn \leq P} \int_X \Lambda(Wn + b) \Psi_x(Wn + b) dx \right|,$$

where I' is an infinite subset of I . We may do this by a diagonalization argument. By the triangle inequality,

$$\mathbb{E}_{b < W} \lim_{P \in I'} \left| \mathbb{E}_{P/2 < Wn + b \leq P} \int_X \Lambda(Wn + b) \Psi_x(Wn + b) dx \right| \gtrsim_{k,s,c} 1,$$

where the implied constant does not depend on W . Note that because $b < W$, we miss at most one term by changing the bounds of the sum from $P/2 < Wn + b \leq P$ to $P/2 < Wn \leq P$. Since W is much smaller than P , this is an acceptable error. Note that if b is not coprime to W , then

$$\lim_{P \in I'} \mathbb{E}_{P/2 < Wn \leq P} \int_X \Lambda(Wn + b) \Psi_x(Wn + b) dx = 0,$$

because $Wn + b$ is never prime. By the pigeonhole principle, there exists $b < W$ such that

$$\lim_{P \in I'} \left| \mathbb{E}_{P/2 < Wn \leq P} \int_X \frac{W}{\varphi(W)} \Lambda(Wn + b) \Psi_x(Wn + b) dx \right| \gtrsim_{k,s,c} 1,$$

where again the implied constant does not depend on W and where $\varphi(W)$ is Euler's totient function, the function which counts the number of residue classes mod W that are coprime to W . Denote $(W/\varphi(W))\Lambda(Wn + b) = \Lambda_{b,W}$. Then we can write our expression as a sum of two terms

$$\begin{aligned} & \lim_{P \in I'} \left| \mathbb{E}_{P/2 < Wn \leq P} \int_X \Lambda_{b,W}(n) \Psi_x(Wn + b) dx \right| \\ &= \lim_{P \in I'} \left| \mathbb{E}_{P/2 < Wn \leq P} \int_X (\Lambda_{b,W}(n) - 1) \Psi_x(Wn + b) + \Psi_x(Wn + b) dx \right| \end{aligned}$$

To handle the first term, we need the following theorems of Green and Tao and of Green, Tao, and Ziegler.

THEOREM 2.26. [GT10, Proposition 11.2] *Let G/Γ be a degree s filtered nilmanifold, and let $M > 0$. Suppose that $F(g^n x)_{n=1}^\infty$ is a bounded nilsequence on G/Γ with Lipschitz constant at most M , where F is a function on G/Γ , g is an element of G , and x is a point in G/Γ . Let $\varepsilon \in (0, 1)$ and P a large natural number. Then we may decompose*

$$F(g^n x) = F_1(n) + F_2(n),$$

where $F_1: \mathbb{N} \rightarrow [-1, 1]$ is a sequence with Lipschitz constant $O_{M,\varepsilon,G/\Gamma}(1)$ and obeying the dual norm bound

$$\|F_1\|_{U^{s+1}[P/2 < Wn \leq P]^*} = O_{M,\varepsilon,G/\Gamma}(1),$$

while $F_2: \mathbb{N} \rightarrow \mathbb{R}$ obeys the uniform bound

$$\|F_2\|_\infty \leq \varepsilon.$$

Note that the bound $\|F_1\|_{U^{s+1}[N]^*} = O_{M,\varepsilon,G/\Gamma}(1)$ is uniform in the element g . We also need the following theorem of Green, Tao, and Ziegler. The proof of this theorem is spread out over [GTZ12, GT12a, GT10], making it somewhat hard to give a specific theorem number. Essentially, if the Gowers norm were big then the inverse conjecture for the Gowers norms would imply that the Möbius function correlates with a nilsequence which it does not by the Möbius nilsequence conjecture. In [GT10], Theorem 7.2 states that the theorem follows from the Möbius nilsequence conjecture and the inverse conjecture for the Gowers norms. The first of these conjectures is an immediate consequence of [GT12a, Theorem 1.1]. The second of these conjectures is [GTZ12, Theorem 1.3].

THEOREM 2.27. ([GTZ12]; see also [GT12a, GT10]) *With all the notation as before,*

$$\|\Lambda_{b,W} - 1\|_{U^{s+1}[P/2 < Wn \leq P]} = o_{W \rightarrow \infty}(1).$$

Thus, our nilsequence Ψ_x can be written as a sum $\Psi_x = F_1 + F_2$ where F_1 and F_2 implicitly depend on x and enjoy the following properties. F_2 is uniformly small, so

$$\limsup_{P \in I'} \left| \mathbb{E}_{P/2 < Wn \leq P} (\Lambda_{b,W}(n) - 1) F_1(Wn + b) \right|$$

can be estimated by simply moving the absolute values inside. The remaining term is bounded in dual norm, so

$$(\Lambda_{b,W} - 1)(n) \cdot F_1(n) \leq \|\Lambda_{b,W} - 1\|_{U^{s+1}[P/2 < Wn \leq P]} \cdot \|F_1\|_{U^{s+1}[P/2 < Wn \leq P]^*}$$

which tends to 0. For a similar argument, see the proof of [GT10, Proposition 10.2]. It may also be possible to circumvent the use of [GTZ12] by using [GT12a, Theorem 7.1]. Putting this together, we get that

$$\limsup_{P \in I'} \left| \mathbb{E}_{P/2 < Wn \leq P} \int_X (\Lambda_{b,W}(n) - 1) \Psi_x(Wn + b) dx \right| = o_{W \rightarrow \infty}(1).$$

As such, for W sufficiently large, by the triangle inequality,

$$\liminf_{P \in I'} \left| \mathbb{E}_{P/2 < Wn \leq P} \int_X \Psi_x(Wn + b) dx \right| \gtrsim_{k,s,c} 1.$$

So far we exploited cancellation in the $\Lambda_{b,W}(n) - 1$ term and simply boundedness in the $\Psi_x(n)$ term. Next, we will try to exploit cancellation in Ψ_x to obtain a contradiction. To exploit this cancellation we interpret the average as an integral over a complicated nilmanifold, then use the fact that the frequency of Φ is non-trivial on the identity component of G/Γ and therefore non-trivial on every component of G/Γ . Let $G^{2t} = G \times \dots \times G$ be the product of G with itself $2t$ many times and let $\mathbf{G} = (g^{Wj_1}, g^{Wj_2}, \dots, g^{Wj_{2t}})$ be the element of G^{2t} whose ℓ th coordinate is g^{Wj_ℓ} . For any σ in G let $\Delta\sigma = (\sigma, \dots, \sigma)$ be the element of G^{2t} whose entries are all σ and let G_Δ be the set of all the elements of the form $\Delta\sigma$. Define

$$\mathbf{G} = \overline{\langle \mathbf{G}, G_\Delta, \Gamma \times \dots \times \Gamma \rangle},$$

the closure of the group generated by \mathbf{G} , G_Δ , and Γ^{2t} inside G^{2t} . Our sequence Ψ_x is a nilsequence on \mathbf{G}/Γ^{2t} . Consider the sequence of ‘empirical’ measures on G^{2t}/Γ^{2t} ,

$$\rho_P = \mathbb{E}_{P/2 < W_n \leq P} (g^{(W_n+b)j_1}, \dots, g^{(W_n+b)j_{2t}})_* (\Delta_* dx),$$

where $\Delta_* dx$ is the Haar measure on $G_\Delta/(\Gamma^{2s} \cap G_\Delta)$ and where $*$ denotes the pushforward. By construction, if $\Xi: G^{2t}/\Gamma^{2t} \rightarrow \mathbb{C}$ is defined by

$$\Xi(x_1, \dots, x_{2t}) = \prod_{j=1}^t \Phi(x_j) \cdot \prod_{j=t+1}^{2t} \overline{\Phi}(x_j),$$

then

$$\mathbb{E}_{P/2 < W_n \leq P} \int_X \Psi_x(W_n + b) dx = \int_{G^{2t}/\Gamma^{2t}} \Xi \rho_P(dx).$$

By the Banach–Alaoglu theorem, there is a further subsequence along which the empirical measures converge weakly,

$$\lim_{P \in I''} \rho_P \xrightarrow{*} \rho,$$

where I'' is an infinite subset of I' . Note that, by summation by parts, ρ_P is almost invariant by \mathbf{G} in the following sense:

$$\mathbf{g}_* \rho_P = \rho_P + O\left(\frac{W}{P}\right).$$

Therefore ρ is actually \mathbf{G} invariant. Since ρ is an average of G_Δ invariant measures, ρ is also G_Δ invariant. Of course ρ is also Γ^{2t} invariant because Γ^{2t} acts trivially on G^{2t}/Γ^{2t} . Since stabilizers of measures are closed, ρ is invariant under \mathbf{G} . By the classification of invariant measures, we know that ρ is actually (a translate of) Haar measure on some nilmanifold \mathbf{X} . Next we need the following result, essentially due to Frantzikinakis [Fra17].

LEMMA 2.28. ([Fra17]; see §5.7 and especially the proof of Proposition 5.7) *With all the notation as before, for any $u \in G_s$ and $m \leq s$, we have $(u^{(W_{j_\ell})^m})_{\ell=1}^{2t} \in \mathbf{G}$.*

We include the proof for completeness and because our result differs very slightly from the way it was stated in [Fra17].

Proof. We split Lemma 2.28 into three claims.

CLAIM 2.29. *Let m and ℓ be natural numbers. If g_1 is in \mathbf{G}_m and g_2 and g_3 are in \mathbf{G}_r then there exists σ in \mathbf{G}_{m+r+1} such that*

$$[g_1, g_2] \cdot [g_1, g_3] = [g_1, g_2 \cdot g_3] \cdot \sigma.$$

Moreover, σ depends continuously on $g_1, g_2,$ and g_3 . In fact, this holds for any nilpotent Lie group, not just \mathbf{G} .

CLAIM 2.30. *For any m between 1 and s and any element τ in G_m and in the identity component (which is automatic for $m > 1$), there exists an element σ in $(G_{m+1})^{2t}$ such*

that

$$(\tau^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma \in \mathbf{G}.$$

CLAIM 2.31. For any natural number r between 1 and s and any natural number m between 1 and r and for any τ in G_r , there exists an element σ in $(G_{r+1})^{2s}$ such that

$$(\tau^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma \in \mathbf{G}.$$

We remark that taking $\tau = u$ in Claim 2.31 gives Lemma 2.28.

Proof of Claim 2.29. The proof is simply a computation. For any g_1, g_2 , and g_3 as above,

$$\begin{aligned} [g_1, g_2] \cdot [g_1, g_3] &= g_1 g_2 g_1^{-1} g_2^{-1} [g_1, g_3] \\ &= g_1 g_2 g_1^{-1} [g_1, g_3] g_2^{-1} \pmod{G_{m+r+1}} \\ &= g_1 g_2 g_3 g_1^{-1} g_3^{-1} g_2^{-1} \pmod{G_{m+r+1}} \\ &= [g_1, g_2 g_3] \pmod{G_{m+r+1}}. \end{aligned} \quad \square$$

Proof of Claim 2.30. We prove Claim 2.30 by induction on m . First, suppose $m = 1$. Consider the torus $Z = G/G_2\Gamma$. Let π be the projection map $\pi : G \rightarrow Z$. Then since g acts ergodically on G/Γ , we know $\pi(g)$ is an ergodic element in Z . Therefore, for any $\pi(\tau)$ in G/G_2 , note that $\pi(\tau)$ is in the orbit of $\pi(g)$. By the definition of \mathbf{G} , $(\pi(g^{W_{j\ell}}))_{\ell=1}^{2t}$ is an element of $\pi^{2t}(\mathbf{G})$. Thus, for any τ in G , $(\pi(\tau^{W_{j\ell}}))_{\ell=1}^{2t}$ is an element of $\pi^{2t}(\mathbf{G})$ so, by definition of the quotient,

$$(\tau^{W_{j\ell}})_{\ell=1}^{2t} \cdot \sigma \in \mathbf{G}$$

for some σ in G_2^{2t} .

Next, assume by induction that Claim 2.30 holds for m . We will try to prove the claim for $m + 1$. We begin with the case where τ is the commutator of two elements of the following form. Suppose that there exist g_1 in G and g_2 in G_m such that $[g_1, g_2] = \tau$. By assumption, there exist σ_1 in G_2^{2t} and σ_2 in G_{m+1}^{2t} such that

$$(g_1^{W_{j\ell}})_{\ell=1}^{2t} \cdot \sigma_1 \in \mathbf{G} \text{ and } (g_2^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma_2 \in \mathbf{G}.$$

Since \mathbf{G} is a group, we conclude that the commutator is in \mathbf{G} .

$$[(g_1^{W_{j\ell}})_{\ell=1}^{2t} \cdot \sigma_1, (g_2^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma_2] \in \mathbf{G}.$$

Using Claim 2.29 repeatedly, this is

$$([g_1, g_2]^{W_{j\ell}^{m+1}})_{\ell=1}^{2t} \sigma \in \mathbf{G},$$

for some σ in G_{m+2}^{2t} .

Finally, we note that commutators generate G_{m+1} so it suffices to show that if τ_1 and τ_2 are elements of G_{m+1} that satisfy Claim 2.30 then so does their product. After all, if

$$(\tau_1^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot \sigma_1 \in \mathbf{G} \quad \text{and} \quad (\tau_2^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot \sigma_2 \in \mathbf{G},$$

where σ_1 and σ_2 are in G_{m+2}^{2t} , then

$$\begin{aligned} (\tau_1^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot \sigma_1 \cdot (\tau_2^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot \sigma_2 &= (\tau_1^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot (\tau_2^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot \sigma \\ &= ((\tau_1 \tau_2)^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot \sigma' \\ &= (\tau^{(W_{j\ell})^{m+1}})_{\ell=1}^{2t} \cdot \sigma', \end{aligned}$$

where σ and σ' are in G_{m+2}^{2t} . This completes the proof of Claim 2.30. □

Proof of Claim 2.31. First, if $m = r$ then we are done by Claim 2.30. Thus, we will assume $m < r$.

Second, we check that if τ_1 and τ_2 are in G_r and satisfy Claim 2.31 then so does their product. By assumption, we may write

$$(\tau_i^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma_i \in \mathbf{G},$$

where σ_i is an element of $(G_{r+1})^{2s}$ and $i = 1, 2$. Then the product is given by

$$\begin{aligned} (\tau_1^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma_1 \cdot (\tau_2^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma_2 \\ = (\tau_1^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot (\tau_2^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma_1 \cdot [\sigma_1^{-1}, (\tau_2^{-(W_{j\ell})^m})_{\ell=1}^{2t}] \cdot \sigma_2. \end{aligned}$$

Then we use that $\tau_1^{(W_{j\ell})^m} \cdot \tau_2^{(W_{j\ell})^m} = (\tau_1 \tau_2)^{(W_{j\ell})^m}$ up to higher-order terms:

$$= ((\tau_1 \tau_2)^{(W_{j\ell})^m})_{\ell=1}^{2t} \pmod{(G_{r+1})^{2s}}.$$

Therefore, it suffices to prove Claim 2.31 in the case where $\tau = [g_1, g_2]$ where g_1 is in G_m and g_2 is in G_{r-m} because such commutators generate G_r as a group up to higher-order corrections.

By Claim 2.30, there exists σ in $(G_{m+1})^{2s}$ such that

$$(g_1^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma \in \mathbf{G}.$$

We also know, because \mathbf{G} contains diagonal elements, that $(g_2)_{\ell=1}^{2t}$ is an element of \mathbf{G} . We conclude that

$$[(g_1^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma, (g_2)_{\ell=1}^{2t}] \in \mathbf{G}.$$

By Claim 2.29, this is given by

$$(\tau^{(W_{j\ell})^m})_{\ell=1}^{2t} \cdot \sigma' \in \mathbf{G},$$

for some σ' in G_{r+1} . □

This completes the proof of Lemma 2.28 by plugging in $r = s$. □

Since the frequency ξ of Φ is non-trivial on the identity component, there exists an element u in the identity component of G_s such that $(1/2\pi i) \log \xi(u)$ is irrational. Fix such a u . Now since J does not solve Vinogradov’s mean value problem there exists $m \leq s$ such that $j_1^m + \dots + j_t^m - j_{t+1}^m - \dots - j_{2t}^m \neq 0$. Fix such an m . Then the map $G_s \rightarrow G_s$ given by $v \mapsto v^{(j_1^m + \dots + j_t^m - j_{t+1}^m - \dots - j_{2t}^m)W^m}$ has image both open and closed so u is in the

image. For more details, see [Fra17]. Fix a v such that $v \mapsto u$. Then by Lemma 2.28, $(v^{(Wj_\ell)^m})_{\ell=1}^{2t} \in \mathbf{G}$. As such

$$\begin{aligned} \int_{\mathbf{X}} \Xi(x) \rho(dx) &= \int_{\mathbf{X}} \Xi(vx) \rho(dx) \\ &= \int_{\mathbf{X}} \xi(u) \Xi(x) \rho(dx) \\ &= 0. \end{aligned}$$

This gives a contradiction. We conclude that the terms which do not solve Vinogradov's mean value problem do not contribute to our sum.

For every $2t$ -tuple j_1, \dots, j_{2t} in $[k]^{2t}$, we have that for all p ,

$$|\Phi(g^{pj_1}x) \dots \Phi(g^{pj_t}x) \cdot \overline{\Phi}(g^{pj_{t+1}}x) \dots \overline{\Phi}(g^{pj_{2t}}x)| \leq 1,$$

simply using a trivial L^∞ bound. For every $2t$ -tuple which does not solve Vinogradov's mean value problem we have

$$\begin{aligned} \limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \\ \int_X \Phi(g^{pj_1}x) \dots \Phi(g^{pj_t}x) \cdot \overline{\Phi}(g^{pj_{t+1}}x) \dots \overline{\Phi}(g^{pj_{2t}}x) = 0. \end{aligned}$$

Therefore, the average is bounded by the fraction of terms which solve Vinogradov's mean value problem. There are no more than $C_{s,\varepsilon} k^{t+.5\varepsilon}$ such solutions by Bourgain, Demeter, and Guth's result (Theorem 2.25). Thus

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x)|^{2t} dx \cdot k^{t-\varepsilon} = 0$$

in the case $t \neq 2$, and

$$\limsup_{k \in \mathcal{K}} \limsup_{P \rightarrow \infty} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \Phi(g^{ph}x)|^{2t} dx \cdot k^t \leq C$$

in the case $t = 2$. After all, since diagonal solutions are the only solutions to Vinogradov's mean value problem in the case of two variables and one equation (i.e. $j_1 = j_2$), there is no ε loss when $t = 2$. Thus, we obtain Theorem 2.24 and in turn Theorems 1.8 and 1.9.

3. Proof of Theorem 1.11

The proof of Theorem 1.11 is essentially the proof of Theorem 1.8 with a few minor simplifications. As before, suppose not. Then as before, we can find a joining such that

$$\left| \int_{X \times Y} f(x) f'(y) \nu(dx dy) \right| > c.$$

As before, we can apply [FH18a] such that

$$\left| \int_{X \times Y} \mathbb{E}^\mu[f|\mathcal{Z}](x) f'(y) \nu(dx dy) \right| > c.$$

Unlike before, we do not need to restrict the integral to B . As before, we can average over translates

$$\int_{X \times Y} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^h x) f'(T^h y)| \nu(dx dy) > c.$$

As before, we can take an essential supremum over y ,

$$\int_X \sup_{y \in Y} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^h x) f'(T^h y)| \mu(dx) > c.$$

As before, we can apply the entropy decrement argument: for some $P \gg k$, we have

$$\mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \in Y} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph} x) f'(T^h y)| \mu(dx) > c.$$

We can use the Cauchy–Schwarz inequality,

$$\mathbb{E}_{P/2 < p \leq P} \int_X \sup_{y \in Y} |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph} x) f'(T^h y)|^2 \mu(dx) > c^2.$$

This time, we would like to replace f' by a sum over words of length k up to ε rounding. In the no-rounding case, we knew that words of f' were words of b . We double-check that a similar result holds for words up to constant rounding. In particular, fix k such that there are at most δk words of length k that occur with positive log density up to ε rounding. Thus, we can fix a set Σ of words of length k such that $\#\Sigma \leq \delta k$ and for all n outside a set of 0 log density there exists an ϵ in Σ such that $|b(n+h) - \epsilon_h| \leq \varepsilon$. Translating this to the dynamical setting,

$$\nu\{(x, y) : \text{there exists } \epsilon \text{ in } \Sigma \text{ such that } |f'(T^h y) - \epsilon_h| \leq \varepsilon\} = 0.$$

Therefore, we can replace f' by a sum over words as before:

$$\sum_{\epsilon} \mathbb{E}_{P/2 < p \leq P} \int_X |\mathbb{E}_{h \leq k} \mathbb{E}^\mu[f|\mathcal{Z}](T^{ph} x) \epsilon_h|^2 \mu(dx) > c^2 - 2\varepsilon.$$

Notice that this time, when we replace $f'(T^h y)$ by a word, we incur an error of ε . Now the rest of the argument runs exactly the same as before. In fact, after pigeonholing, any dependence on b completely drops out of the argument.

Acknowledgements. Special thanks to Terence Tao for sharing many ideas on an earlier version of this paper and for his many helpful comments. Also, special thanks to Nikos Frantzikinakis for pointing out a number of ways to strengthen the main theorem of this paper. I would also like to thank Tim Austin, Björn Bringmann, Alex Dobner, Asgar Jamneshan, Bernard Host, Gyu Eun Lee, Zane Li, Adam Lott, Maksym Radziwiłł, Bar Roytman, Chris Shriver, Joni Teräväinen and Alex Wertheim for many helpful discussions. Thanks to Will Sawin for suggesting I use Vinogradov’s mean value theorem to improve an earlier version of Theorem 1.9. I would lastly like to thank the anonymous reviewer for their extremely helpful comments. Some of this work was completed while the author was at the American Institute for Mathematics workshop on Sarnak’s conjecture.

A. Appendix. Frantzikinakis–Host and dynamical models

Tao [Tao17b] shows that there is a joining $(X_0 \times Y, \nu_0, T, f, f', M, I_m)$ of a dynamical model for a with b where $X_0 = D^{\mathbb{Z}} \times \widehat{\mathbb{Z}}$ is the space of sequences in the unit disk with the product topology, T is the shift map on $D^{\mathbb{Z}}$ and $+1$ on $\widehat{\mathbb{Z}}$, f is the evaluation at 0 map, M is projection onto the second factor, and $I_m((x(n))_{n \in \mathbb{Z}}, r) = ((\overline{a(m)}x(mn))_{n \in \mathbb{Z}}, r/m)$ whenever r is in $m\widehat{\mathbb{Z}}$. Call μ_0 the pushforward of ν_0 onto X_0 . Of course, X_0 factors onto $D^{\mathbb{Z}}$ by projection onto the first factor. Call ρ the pushforward of ν onto $D^{\mathbb{Z}}$. Frantzikinakis and Host [FH18a, Proposition 4.2] showed that $(D^{\mathbb{Z}}, T, \rho)$ is a factor of a system $(\tilde{X}, \tilde{\rho}, T)$ where $\tilde{X} = (D^{\mathbb{Z}})^{\mathbb{Z}}$, T is the shift map and there exists a natural number d so that if \mathcal{P}_d is the set of primes which are 1 mod d then

$$\int_{\tilde{X}} \prod_{j=-K}^K F_j(T^j x) \tilde{\rho}(dx) = \lim_{N \rightarrow \infty} \mathbb{E}_{p \in \mathcal{P}_d \cap [N]} \int_{D^{\mathbb{Z}}} \prod_{j=-K}^K F_j(T^{pj} x) \rho(dx),$$

where K is any natural number, the functions F_j are any bounded measurable functions depending only on the zeroth coordinate and by [FH18a] the limit always exists. We fix such a d . By [FH18a, Theorem 4.5], each ergodic component of \tilde{X} is isomorphic to a product of a Bernoulli system with an inverse limit of nilsystems. Thus, we get a joining of \tilde{X} with X_0 over their common factor $D^{\mathbb{Z}}$. Call this joining (X, μ, T) . We also get a joining of $X_0 \times Y$ and X over their common factor X_0 , which we call $(X \times Y, \nu, T)$. Explicitly, this joining is defined as follows. A point in $(X \times Y, \nu, T)$ can be thought of as a triple of points (x_1, x_2, y) with x_1 in \tilde{X} and (x_2, y) in $X_0 \times Y$. Since $X_0 = D^{\mathbb{Z}} \times \widehat{\mathbb{Z}}$, we have that $x_2 = (x_3, r)$ for some x_3 in $D^{\mathbb{Z}}$ and r in $\widehat{\mathbb{Z}}$. The measure is supported on triples where $\pi(x_1) = x_3$, so we will often forget x_1 and simply write a point in $X \times Y$ as a triple (x, r, y) with x in \tilde{X} , r in $\widehat{\mathbb{Z}}$, and y in Y . The measure is given explicitly by the following formula: if K is a natural number, F_j are bounded measurable functions on $D^{\mathbb{Z}}$ depending only on the zeroth coordinate, ϕ is a bounded measurable function on $\widehat{\mathbb{Z}}$, and ψ is a bounded measurable function on Y , then

$$\begin{aligned} & \int_{X \times Y} \prod_{j=-K}^K F_j(T^j x) \cdot \phi(r) \cdot \psi(y) \nu(dx \, dr \, dy) \\ &= \lim_{N \rightarrow \infty} \mathbb{E}_{p \in \mathcal{P}_d \cap [N]} \int_{X_0 \times Y} \prod_{j=-K}^K F_j(T^{pj} x) \cdot \phi(r) \cdot \psi(y) \nu_0(dx \, dr \, dy). \end{aligned}$$

We will proceed to check that $X \times Y$ has all the desired properties. We define $M: X \rightarrow \widehat{\mathbb{Z}}$ by taking an element (x, r) with x in \tilde{X} and r in $\widehat{\mathbb{Z}}$ to r . Let x be an element of \tilde{X} . We will write $x = (x_n)_{n \in \mathbb{Z}}$ for a sequence of elements x_n in $D^{\mathbb{Z}}$ and write $x_n(k) \in D$ for the k th element of the sequence x_n . Let $\iota_m((x(k))_{k \in \mathbb{Z}}) = \overline{a(m)}(x(mk))_{k \in \mathbb{Z}}$. We define $I_m(x, r) = (\iota_m((x_{nm})_{n \in \mathbb{Z}}), r/m)$. Explicitly

$$I_m(x, r) = \left((\overline{a(m)}x_{nm}(km))_{k \in \mathbb{Z}}, \frac{r}{m} \right)$$

whenever r is in $m\widehat{\mathbb{Z}}$. We define $f: X \times Y \rightarrow \mathbb{C}$ by the formula $f(x, r, y) = x_0(0)$. This is just the pullback of $f: X_0 \times Y \rightarrow \mathbb{C}$ under the factor map $X \times Y \rightarrow X_0 \times Y$. We

define $f': X \times Y \rightarrow \mathbb{C}$ by pulling back $f': X_0 \times Y \rightarrow \mathbb{C}$ under the same factor map (i.e. $f'(x, r, y) = y$). Now we check the following points.

- $M(T(x, r)) = M(Tx, r + 1) = r + 1 = M(x, r) + 1$.
- We have

$$\begin{aligned} I_m \circ T^m(x, r) &= \left(\overline{(a(m)x_{nm+m}(km)_{k \in \mathbb{Z}})_{n \in \mathbb{Z}}}, \frac{r+m}{m} \right) \\ &= \left(\overline{(a(m)x_{(n+1)m}(km)_{k \in \mathbb{Z}})_{n \in \mathbb{Z}}}, \frac{r}{m} + 1 \right) \\ &= T \left(\overline{(a(m)x_{nm}(km)_{k \in \mathbb{Z}})_{n \in \mathbb{Z}}}, \frac{r}{m} \right) \\ &= T \circ I_m(x, r) \end{aligned}$$

for any m and whenever r is in $m\widehat{\mathbb{Z}}$.

- Let K be a natural number and $F_j: \tilde{X} \rightarrow \mathbb{C}$ be a sequence of bounded measurable functions depending only on 0. Let ϕ be a function which is measurable with respect to $\widehat{\mathbb{Z}}$. Then for any m ,

$$\begin{aligned} &\int_X \mathbb{1}_{M^{-1}(m\widehat{\mathbb{Z}})}(x) \phi(I_m x) \cdot \prod_{j=-K}^K F_j(T^j I_m x) \mu(dx) \\ &= \lim_{N \rightarrow \infty} \mathbb{E}_{p \in \mathcal{P}_d \cap [N]} \int_{X_0} \mathbb{1}_{r \in m\widehat{\mathbb{Z}}} \phi\left(\frac{r}{m}\right) \cdot \prod_{j=-K}^K F_j \circ \iota_m(T^{pmj} x) \mu_0(dx \, dr), \end{aligned}$$

by definition of μ . Next, we use that $\iota_m \circ T^{pmj} = T^{pj} \circ \iota_m$:

$$= \lim_{N \rightarrow \infty} \mathbb{E}_{p \in \mathcal{P}_d \cap [N]} \int_{X_0} \mathbb{1}_{r \in m\widehat{\mathbb{Z}}} \phi\left(\frac{r}{m}\right) \cdot \prod_{j=-K}^K F_j \circ T^{pj} \circ \iota_m(x) \mu_0(dx \, dr).$$

Because X_0 is a dynamical model for a , it possesses a dilation symmetry:

$$= \lim_{N \rightarrow \infty} \mathbb{E}_{p \in \mathcal{P}_d \cap [N]} \int_{X_0} \frac{1}{m} \phi(r) \cdot \prod_{j=-K}^K F_j(T^{pj} x) \mu_0(dx \, dr).$$

Finally, we apply the definition of μ one more time:

$$= \int_X \frac{1}{m} \phi(x) \cdot \prod_{j=-K}^K F_j(T^j x) \mu(dx).$$

- For any natural number m and any r in $m\widehat{\mathbb{Z}}$, we have

$$f(I_m(x, r)) = f\left(\left(\overline{(a(m)x_{nm}(km)_{k \in \mathbb{Z}})_{n \in \mathbb{Z}}}, \frac{r}{m}\right)\right) = \overline{(a(m)x_0(0))} = \overline{(a(m))} f(x, r).$$

- Clearly, for any natural numbers m and h ,

$$I_h I_m(x, r) = \left(\overline{(a(mh)x_{nmh}(kmh)_{k \in \mathbb{Z}})_{n \in \mathbb{Z}}}, \frac{r}{mh} \right) = I_m I_h(x, r),$$

for any r in $hm\widehat{\mathbb{Z}}$.

- Since f and f' are pulled back from $X_0 \times Y$, the ‘statistics’ of $f: X \times Y \rightarrow \mathbb{C}$ will be the same as the statistics of $f: X_0 \times Y \rightarrow \mathbb{C}$ and similarly for f' .

Therefore, $(X \times Y, \nu, T, f, f', I_m, M)$ is a joining of a dynamical model for a with b .

Let (X, μ_ω, T) be an ergodic component of (X, μ, T) which joins the corresponding ergodic component $(\tilde{X}, \tilde{\rho}_\omega, T)$ of $(\tilde{X}, \tilde{\rho}, T)$ with $\widehat{\mathbb{Z}}$. Note that $\widehat{\mathbb{Z}}$ is already an ergodic inverse limit of nilsystems: after all, it is an inverse limit of the ergodic systems of the form $\mathbb{Z}/m\mathbb{Z}$ and the inverse limit of ergodic systems is ergodic. By [FH18a], there exist a Bernoulli system (W, dw, T) and an inverse limit of nilsystems (Z_0, dz, T) such that $(\tilde{X}, \tilde{\rho}_\omega, T) \cong (W, dw, T) \times (Z_0, dz, T)$. Therefore (X, μ_ω, T) is isomorphic to $(W \times Z_0 \times \widehat{\mathbb{Z}}, \mu', T)$ where μ' is some mystery measure and where T is just the product transformation. We can think of this system as a joining of $(W \times Z_0, dw \times dz, T)$ with $(\widehat{\mathbb{Z}}, dz, T)$ or as a joining of (W, dw, T) with $(Z_0 \times \widehat{\mathbb{Z}}, \zeta, T)$ where ζ is some unknown measure given by pushing forward μ' onto $Z_0 \times \widehat{\mathbb{Z}}$. Next, we claim that any ergodic joining of two inverse limits of nilsystems is in fact isomorphic to an inverse limit of nilsystems. After all, if Z_1 and Z_2 are two nilsystems and ζ is an ergodic invariant measure on $Z_1 \times Z_2$, then ζ is a translate of Haar measure on some closed subgroup by measure classification for nilsystems. Thus $(Z_1 \times Z_2, \zeta, T) \cong (Z_3, dz, T)$ for some nilsystem Z_3 . Taking inverse limits, $(Z_0 \times \widehat{\mathbb{Z}}, \zeta, T)$ is isomorphic to an inverse limit of nilsystems (Z, dz, T) . Because (Z, dz, T) is an inverse limit of nilsystems, it has zero entropy so the only possible joining of (Z, dz, T) with the Bernoulli system (W, dw, T) is the trivial joining (i.e. μ' is the product measure $dw \times dz$). Lastly, we claim that Z is isomorphic to the Host–Kra factor of (X, μ_ω, T) . Since the Host–Kra factor $\mathcal{Z}(X)$ is isomorphic to an inverse limit of nilsystems, it has zero entropy, so any factor map from $W \times Z$ to $\mathcal{Z}(X)$ where W is Bernoulli necessarily factors through Z . Thus Z factors onto $\mathcal{Z}(X)$. Of course, since X factors onto Z , the Host–Kra factor for X factors onto the Host–Kra factor for Z . Implicitly in [HK05] and explicitly, for instance, in [HK18, Ch. 12], for any nilsystem (Z_1, dz, T) the Host–Kra factor of Z_1 is Z_1 . Thus, taking inverse limits gives that the Host–Kra factor of Z is Z , so $\mathcal{Z}(X) \cong Z$. This completes the proof.

B. Appendix. Reduction to the completely multiplicative case

We have stated our main theorems in the case where a is completely multiplicative. In this appendix, we show that these assumptions can be weakened to include all multiplicative functions. For example, we will show that Theorem 1.8 holds in this generality. The same argument works for Theorems 1.9 and 1.11 (although in this last case, the way that c depends on ε gets worse). The argument here will be entirely formal, using nothing of the proof of Theorem 1.8 and only the result. However, we remark that the interested reader could check that the proof we give can be adapted to the more general case of multiplicative functions. The main difference is that now the dynamical model for a does not satisfy the identity that the pushforward of μ restricted to $M^{-1}(m\widehat{\mathbb{Z}})$ is $(1/m)\mu$ but instead we incur a $1/m$ error, that is, for all ϕ in satisfying $\|\phi\|_{L^\infty(\mu)} \leq 1$ we have

$$\int_X \phi(x)\mu(dx) = \int_X m \mathbb{1}_{x \in M^{-1}(m\widehat{\mathbb{Z}})} \phi(I_m(x))\mu(dx) + O\left(\frac{1}{m}\right)$$

This introduces an error term of size $O(1/P)$ in Corollary 2.17 which tends to 0 as P tends to infinity.

However, here we proceed just using the statement of Theorem 1.8. Famously, we can write

$$\mu(n) = \sum_{d^2|n} \lambda\left(\frac{n}{d^2}\right) \mu(d),$$

where λ is the Liouville function and μ is the Möbius function which agrees with the Liouville function on squarefree numbers and vanishes on numbers which are not squarefree. Of course,

$$\lambda\left(\frac{n}{d^2}\right) = \lambda(n),$$

but we write it this way to suggest that the convolution identity

$$\mu = \lambda * \phi,$$

where $\phi(d^2) = \mu(d)$, may be generalized. In fact, for any multiplicative function a taking values on the unit circle, we may write

$$a = a_1 * a_2$$

where a_1 is some completely multiplicative function taking values on the unit circle and a_2 is a (possibly unbounded) multiplicative function supported on numbers of the form d^k for some natural numbers d and k with $k \geq 2$. To prove this is possible, it suffices to check it is possible on prime powers since both sides are multiplicative. For any prime p , we define

$$a_1(p) = a(p)$$

and so

$$a_1(p) \cdot a_2(1) + a_1(1) \cdot a_2(p) = a(p) \cdot 1 + 1 \cdot 0 = a(p).$$

We also want

$$a(p^2) = a_1(p^2) \cdot 1 + 1 \cdot a_2(p^2),$$

so we choose

$$\begin{aligned} a_2(p^2) &= a(p^2) - a_1(p^2) \\ &= a(p^2) - a(p)^2. \end{aligned}$$

Iteratively, we may define

$$\begin{aligned} a_2(p^k) &= a(p^k) - \sum_{0 \leq i < k} a_1(p^{k-i}) a_2(p^i) \\ &= a(p^k) - \sum_{0 \leq i < k} a(p)^{k-i} a_2(p^i). \end{aligned}$$

Since whether a is unpretentious or not depends only on the behavior of a at primes; clearly if a is unpretentious then so is a_1 .

Informally, the probability that a random number is divisible by d is roughly $1/d$. Thus, the expected number of times that any number of the form d^k for $k \geq 2$ divides a random natural number is at most

$$\sum_{d \geq 2} \sum_{k \geq 2} \frac{1}{d^k}$$

which is summable. Thus the tails

$$\sum_{d \geq C} \sum_{k \geq 2} \frac{1}{d^k}$$

and

$$\sum_{d \geq 2} \sum_{k \geq C} \frac{1}{d^k}$$

tend to zero as C tends to infinity. Let S be the set of natural numbers n for which d^k divides n for $d, k \geq 2$ implies $d, k \leq C$. The previous analysis says most numbers are in S . Fix a function b as in the statement of Theorem 1.8, that is, a bounded function such that for any $\delta > 0$ there are infinitely many k such that the number of words of b of length k that occur with positive upper logarithmic density is at most δk^2 . Our goal will be to show that for N large,

$$|\mathbb{E}_{n \leq N}^{\log} a(n)b(n)|$$

is small, say less than a constant times some small positive number ε . If C is sufficiently large depending on ε but still very small compared to N , we may modify a on the set of numbers outside S . In particular, a is given by the formula

$$a(n) = \sum_{\ell|n} a_1\left(\frac{n}{\ell}\right)a_2(\ell).$$

For most numbers, this is the same as

$$= \sum_{\substack{\ell|n \\ \ell \leq C^C}} a_1\left(\frac{n}{\ell}\right)a_2(\ell).$$

That formula works as long as n is not divisible by a number of the form d^k where either d or k is greater than C . In that exceptional case when n is divisible by a number of the form d^k with d or k greater than C , we can write n as $i \cdot j$ where i is not divisible by any number greater than C^C and is as large as possible given that constraint. We conclude that if $\ell \leq C^C$ and $\ell|n$ then $\ell|i$. Thus, expanding the definitions and using multiplicativity,

$$\begin{aligned} a_1(j) \cdot a(i) &= a_1(j) \sum_{\substack{\ell|i \\ \ell \leq C^C}} a_1\left(\frac{i}{\ell}\right)a_2(\ell) \\ &= \sum_{\substack{\ell|i \\ \ell \leq C^C}} a_1\left(\frac{ij}{\ell}\right)a_2(\ell) \end{aligned}$$

$$= \sum_{\substack{\ell|n \\ \ell \leq C^C}} a_1\left(\frac{n}{\ell}\right) a_2(\ell)$$

We conclude that the formula

$$\sum_{\ell|n} a_1\left(\frac{n}{\ell}\right) a_2(\ell)$$

is bounded and agrees with $a(n)$ all but at most

$$\sum_{d \geq C} \sum_{k \geq 2} \frac{1}{d^k} + \sum_{d \geq 2} \sum_{k \geq C} \frac{1}{d^k}$$

of the time. Thus, it suffices to show

$$\left| \mathbb{E}_{n \leq N}^{\log} \sum_{\substack{m\ell=n \\ \ell \leq C^C}} a_1(m) a_2(\ell) b(n) \right|$$

is small. By changing variables and applying Fubini,

$$= \left| \sum_{\ell \leq C^C} a_2(\ell) \mathbb{E}_{m \leq N/\ell}^{\log} a_1(m) b(m\ell) \right|$$

Fix a natural number ℓ . Notice that every word of length k of the function $m \mapsto b(m\ell)$ embeds in a word of b of length $k \cdot \ell$. Thus, it is easy to check that $m \mapsto b(m\ell)$ still satisfies the conditions of Theorem 1.8. Therefore, as N tends to infinity, the previous expression tends to 0.

REFERENCES

- [BDG16] J. Bourgain, C. Demeter and L. Guth. Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three. *Ann. of Math. (2)* **184**(2) (2016), 633–682.
- [Bou13a] J. Bourgain. Möbius-Walsh correlation bounds and an estimate of Mauduit and Rivat. *J. Anal. Math.* **119** (2013), 147–163.
- [Bou13b] J. Bourgain. On the correlation of the Moebius function with rank-one systems. *J. Anal. Math.* **120** (2013), 105–130.
- [BSZ13] J. Bourgain, P. Sarnak and T. Ziegler. Disjointness of Moebius from horocycle flows. *From Fourier Analysis and Number Theory to Radon Transforms and Geometry*. Springer, New York, NY, 2013, pp. 67–83.
- [DK15] T. Downarowicz and S. Kasjan. Odometers and Toeplitz Systems revisited in the context of Sarnak's conjecture. *Studia Math.* **229**(1) (2015), 45–72.
- [EAKaPLdIR17] E. H. El Abdalaoui, J. K. Przymus, M. Lemańczyk and T. de la Rue. The Chowla and the Sarnak conjectures from ergodic theory point of view. *Discrete Contin. Dyn. Syst.* **37**(6) (2017), 2899–2944.
- [EAKL16] E. H. El Abdalaoui, S. Kasjan and M. Lemańczyk. 0-1 sequences of the Thue-Morse type and Sarnak's conjecture. *Proc. Amer. Math. Soc.* **144**(1) (2016), 161–176.
- [FH18a] N. Frantzikinakis and B. Host. Furstenberg systems of bounded multiplicative functions and applications. *Int. Math. Res. Not.* doi:<https://doi.org/10.1093/imrn/rnz037>. Published online 21 February 2019.
- [FH18b] N. Frantzikinakis and B. Host. The logarithmic Sarnak conjecture for ergodic weights. *Ann. of Math. (2)* **187**(3) (2018), 869–931.
- [FJ18] A.-H. Fan and Y. Jiang. Oscillating sequences, MMA and MMLS flows and Sarnak's conjecture. *Ergod. Th. & Dynam. Sys.* **38**(5) (2018), 1709–1744.
- [FKO82] H. Furstenberg, Y. Katznelson and D. Ornstein. The ergodic theoretical proof of Szemerédi's theorem. *Bull. Amer. Math. Soc. (N.S.)* **7**(3) (1982), 527–552.

- [Fra17] N. Frantzikinakis. Ergodicity of the Liouville system implies the Chowla conjecture. *Discrete Anal.* **41**(19) (2017).
- [GLdLR19] A. Gomilko, M. Lemańczyk and T. de La Rue. Möbius orthogonality in density for zero entropy dynamical systems. *Preprint*, 2019, arXiv:<https://arxiv.org/abs/1905.06563>.
- [GT10] B. Green and T. Tao. Linear equations in primes. *Ann. of Math* (2) **171**(3) (2010), 1753–1850.
- [GT12a] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math.* (2) **175**(2) (2012), 541–566.
- [GT12b] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math.* **175** (2012), 465–540.
- [GTZ12] B. Green, T. Tao and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$ -norm. *Ann. of Math.* (2) **176**(2) (2012), 1231–1372.
- [Hil86] A. Hildebrand. On consecutive values of the Liouville function. *Enseign. Math.* (2) **32**(3–4) (1986), 219–226.
- [HK05] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math.* (2) **161**(1) (2005), 397–488.
- [HK18] B. Host and B. Kra. *Nilpotent Structures in Ergodic Theory (Mathematical Surveys and Monographs, 236)*, American Mathematical Society, Providence, RI, 2018.
- [HLSY17] W. Huang, Z. Lian, S. Shao and X. Ye. Sequences from zero entropy noncommutative toral automorphisms and Sarnak conjecture. *J. Differential Equations* **263**(1) (2017), 779–810.
- [LS15] J. Liu and P. Sarnak. The Möbius function and distal flows. *Duke Math. J.* **164**(7) (2015), 1353–1399.
- [MMR14] B. Martin, C. Mauduit and J. Rivat. Théorème des nombres premiers pour les fonctions digitales. *Acta Arith.* **165**(1) (2014), 11–45.
- [Mor18] J. Moreira. *Tao's proof of (logarithmically averaged) Chowla's conjecture for two point correlations* (2018), available at <https://joelmoreira.wordpress.com/2018/11/04/taos-proof-of-logarithmically-averaged-chowlas-conjecture-for-two-point-correlations/>.
- [MR16] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math.* (2) **183**(3) (2016), 1015–1056.
- [MRT15] K. Matomäki, M. Radziwiłł and T. Tao. An averaged form of Chowla's conjecture. *Algebra Number Theory* **9**(9) (2015), 2167–2196.
- [MRT16] K. Matomäki, M. Radziwiłł and T. Tao. Sign patterns of the Liouville and Möbius functions. *Forum Math. Sigma* **4** (2016), e14, 44.
- [MRT18] K. Matomäki, M. Radziwiłł and T. Tao. Fourier uniformity of bounded multiplicative functions in short intervals on average. *Invent. Math.* **220** (2020), 59.
- [Mül17] C. Müllner. Automatic sequences fulfill the Sarnak conjecture. *Duke Math. J.* **166**(17) (2017), 3219–3290.
- [Pec18] R. Peckner. Möbius disjointness for homogeneous dynamics. *Duke Math. J.* **167**(14) (2018), 2745–2792.
- [Sar12] P. Sarnak. Möbius randomness and dynamics. *Not. S. Afr. Math. Soc.* **43**(2) (2012), 89–97.
- [Saw20] W. Sawin. Dynamical models for Liouville and obstructions to further progress on sign patters. *J. Number Theory* **213** (2020), 1–15.
- [Tao16] T. Tao. The logarithmically averaged Chowla and Elliott conjectures for two-point correlations. *Forum Math. Pi* **4** (2016), e8, 36.
- [Tao17a] T. Tao. Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. *Number Theory—Diophantine Problems, Uniform Distribution and Applications*. Springer, Cham, 2017, pp. 391–421.
- [Tao17b] T. Tao. *Furstenberg limits of the Liouville function* (2017), available at <https://terrytao.wordpress.com/2017/03/05/furstenberg-limits-of-the-liouville-function/>.
- [TT17a] T. Tao and J. Teräväinen. Odd order cases of the logarithmically averaged Chowla conjecture. *J. Théor. Nombres Bordeaux* **30** (2018), 997–1015, available at https://jtnb.centre-mersenne.org/article/JTNB_2018__30_3_997_0.pdf.
- [TT17b] T. Tao and J. Teräväinen. The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures. *Duke Math. J.* **168** (2019), 1977–2027.
- [Vee17] W. A. Veech. Möbius orthogonality for generalized Morse-Kakutani flows. *Amer. J. Math.* **139**(5) (2017), 1157–1203.
- [Wan17] Z. Wang. Möbius disjointness for analytic skew products. *Invent. Math.* **209**(1) (2017), 175–196.