Mathematika 63 (2017) 944–1040
doi:10.1112/S0025579317000316

NEW BOUNDS FOR SZEMERÉDI'S THEOREM, III: A POLYLOGARITHMIC BOUND FOR $r_4(N)$

BEN GREEN AND TERENCE TAO

Dedicated to the legacy of Klaus Roth

Abstract. Define $r_4(N)$ to be the largest cardinality of a set $A \subset \{1, \dots, N\}$ that does not contain four elements in arithmetic progression. In 1998, Gowers proved that

$$r_4(N) \ll N(\log\log N)^{-c}$$

for some absolute constant $c > 0$. In 2005, the authors improved this to

$$r_4(N) \ll N\mathrm{e}^{-c\sqrt{\log\log N}}.$$

In this paper we further improve this to

$$r_4(N) \ll N(\log N)^{-c},$$

which appears to be the limit of our methods.

Contents

§1. *Introduction.* Let $N \geqslant 100$ be a natural number (so that $\log\log N$ is positive). If $k \geqslant 3$ is a natural number we define $r_k(N)$ to be the largest cardinality of a set $A \subset [N] := \{1, \dots, N\}$ that does not contain an arithmetic progression of k distinct elements.

Klaus Roth proved in 1953 [24] that $r_3(N) \ll N(\log\log N)^{-1}$, and so in particular[1] $r_3(N) = o(N)$ as $N \to \infty$. Since Szemerédi's 1969 proof [29] that $r_4(N) = o(N)$, and his later proof [30] that $r_k(N) = o_k(N)$ for $k \geqslant 5$ (answering

Received 4 May 2017.
MSC (2010): 11B30 (primary).
[1] See §2 for the asymptotic notation used in this paper.

a question from [10]), it has been natural to ask for similarly effective bounds for these quantities. It is worth noting that the famous conjecture of Erdős [9] asserting that every set of natural numbers whose sum of reciprocals is divergent is equivalent to the claim that $\sum_{n=1}^{\infty} r_k(2^n)/2^n < \infty$ for all $k \geqslant 3$ (see [33, Exercise 10.0.6]).

A first attempt towards quantitative bounds for higher k was made by Roth in [25], who provided a new proof that $r_4(N) = o(N)$. A major breakthrough was made in 1998 by Gowers [11, 12], who obtained the bound $r_k(N) \ll_k N(\log\log N)^{-\epsilon_k}$ for each $k \geqslant 4$, where $\epsilon_k := 1/2^{2^{k+9}}$. In the other direction, a classical result of Behrend [2] shows that $r_3(N) \gg N\exp(-c\sqrt{\log N})$ for some absolute constant $c > 0$ (see [8, 20] for a slight refinement of this bound), and in [23] (see also [22]) the argument was generalized to give the bound $r_{1+2^k}(N) \gg_k N\exp(-c\log^{1/(k+1)} N)$ for any $k \geqslant 1$.

In the meantime, there has been progress on $r_3(N)$. Szemerédi (unpublished) obtained the bound $r_3(N) \ll Ne^{-c\sqrt{\log\log N}}$, and shortly thereafter Heath-Brown [21] and Szemerédi [32] independently obtained the bound $r_3(N) \ll N(\log N)^{-c}$ for some absolute constant $c > 0$. The best known value of c has been improved in a series of papers [4, 6, 7, 27, 28]. Sanders [28] was the first to show that any $c < 1$ is admissible, and Bloom [4] improved the factor of $\log\log N$ in Sanders's bound.

The only other direct progress on upper bounds for $r_k(N)$ is our previous paper [19], obtaining the bound $r_4(N) \ll Ne^{-c\sqrt{\log\log N}}$. The main objective of this paper is to obtain a bound for $r_4(N)$ of the same quality as the Heath-Brown and Szemerédi bound for $r_3(N)$.

THEOREM 1.1. *We have $r_4(N) \ll N(\log N)^{-c}$ for some absolute constant $c > 0$.*

An analogous result in finite fields was claimed (and published [15]) by us around 12 years ago, although an error in this paper came to light some years later. This was corrected around 5 years ago in [16]. These papers (like almost all of the previously cited quantitative results on $r_k(N)$) are based on the density increment argument of Roth [24]. However we will use a slightly different "energy decrement" and "regularity" approach here, inspired by the Khinchin-type recurrence theorems for length-four progressions established by Bergelson *et al* [3] in the ergodic setting, and by the authors [13] in the combinatorial setting.

§2. *Notation.* We use the asymptotic notation $X \ll Y$ or $X = O(Y)$ to denote $|X| \leqslant CY$ for some constant C. Given an asymptotic parameter N going to infinity, we use $X = o(Y)$ to denote the bound $|X| \leqslant c(N)Y$ for some function $c(N)$ of N that goes to zero as N goes to infinity. We also write $X \asymp Y$ for $X \ll Y \ll X$. If we need the implied constant C or decay function $c()$ to depend on an additional parameter, we indicate this by subscripts, e.g. $X = o_k(Y)$ denotes the bound $|X| \leqslant c_k(N)Y$ for a function $c_k(N)$ that goes to zero as $N \to \infty$ for any fixed choice of k.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We will frequently use probabilistic notation, and adopt the convention that boldface variables such as $\mathbf{a}$ or $\mathbf{r}$ represent random variables, whereas non-boldface variables such as a and r represent deterministic variables (or constants). We write $\mathbb{P}(E)$ for the probability of a random event E, and $\mathbb{E}\mathbf{X}$ and $\operatorname{Var}\mathbf{X}$ for the expectation and variance of a real or complex random variable $\mathbf{X}$; we also use $\mathbb{E}(\mathbf{X}|E) = \mathbb{E}\mathbf{X}1_E/\mathbb{P}(E)$ for the conditional expectation of $\mathbf{X}$ relative to an event E of non-zero probability, where of course 1_E denotes the indicator variable of E. In this paper, the random variables $\mathbf{X}$ of which we will compute expectations of will be discrete, in the sense that they take only finitely many values, so there will be no issues of measurability. The *essential range* of a discrete random variable $\mathbf{X}$ is the set of all values X for which $\mathbb{P}(\mathbf{X} = X)$ is non-zero.

By a slight abuse of notation, we also retain the traditional (in additive combinatorics) use for $\mathbb{E}$ as an average, thus $\mathbb{E}_{a\in A} f(a) := (1/|A|)\sum_{a\in A} f(a)$ for any finite non-empty set A and function $f : A \to \mathbb{C}$, where we use $|A|$ to denote the cardinality of A. Thus for instance $\mathbb{E}_{a\in A} f(a) = \mathbb{E} f(\mathbf{a})$ if $\mathbf{a}$ is drawn uniformly at random from A.

A function $f : A \to \mathbb{C}$ is said to be 1-*bounded* if one has $|f(a)| \leqslant 1$ for all $a \in A$. We will frequently rely on the following probabilistic form of the Cauchy–Schwarz inequality, the proof of which is an exercise.

LEMMA 2.1 (Cauchy–Schwarz). *Let A, B be sets, let $f : A \to \mathbb{C}$ be a 1-bounded function, and let $g : A \times B \to \mathbb{C}$ be another function. Let $\mathbf{a}, \mathbf{b}, \mathbf{b}'$ be discrete random variables in A, B, B' respectively, such that $\mathbf{b}'$ is a conditionally independent copy of $\mathbf{b}$ relative to $\mathbf{a}$, that is to say that*

$$\mathbb{P}(\mathbf{b} = b, \mathbf{b}' = b'|\mathbf{a} = a) = \mathbb{P}(\mathbf{b} = b|\mathbf{a} = a)\mathbb{P}(\mathbf{b} = b'|\mathbf{a} = a)$$

for all a in the essential range of $\mathbf{a}$ and all $b, b' \in B$. Then we have

$$|\mathbb{E} f(\mathbf{a})g(\mathbf{a}, \mathbf{b})|^2 \leqslant \mathbb{E} g(\mathbf{a}, \mathbf{b})\overline{g(\mathbf{a}, \mathbf{b}')}. \tag{2.1}$$

We will think of this lemma as allowing one to eliminate a factor $f(\mathbf{a})$ from a lower bound of the form $|\mathbb{E} f(\mathbf{a})g(\mathbf{a}, \mathbf{b})| \geqslant \eta$, at the cost of duplicating the factor g, and worsening the lower bound from η to η^2.

We also have the following variant of Lemma 2.1.

LEMMA 2.2 (Popularity principle). *Let $\mathbf{a}$ be a random variable taking values in a set A, and let $f : A \to [-C, C]$ be a function for some $C > 0$. If we have $\mathbb{E} f(\mathbf{a}) \geqslant \eta$ for some $\eta > 0$ then, with probability at least $\eta/2C$, the random variable $\mathbf{a}$ attains a value $a \in A$ for which $f(a) \geqslant \eta/2$.*

Proof. If we set $\Omega := \{a \in A : f(a) \geqslant \eta/2\}$, then

$$f(\mathbf{a}) \leqslant \frac{\eta}{2} + C1_{\mathbf{a}\in\Omega}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

and hence on taking expectations

$$\mathbb{E}f(\mathbf{a}) \leqslant \frac{\eta}{2} + C\mathbb{P}(\mathbf{a} \in \Omega).$$

This implies that

$$\mathbb{P}(\mathbf{a} \in \Omega) \geqslant \eta/2C$$

giving the claim. □

If $\theta \in \mathbb{R}$, we write $\|\theta\|_{\mathbb{R}/\mathbb{Z}}$ for the distance from θ to the nearest integer, and $e(\theta) = \mathrm{e}^{2\pi \mathrm{i}\theta}$. Observe from elementary trigonometry that

$$|e(\theta) - 1| = 2|\sin(\pi\theta)| \asymp \|\theta\|_{\mathbb{R}/\mathbb{Z}} \tag{2.2}$$

and hence also

$$1 - \cos(2\pi\theta) = 2|\sin(\pi\theta)|^2 \asymp \|\theta\|^2_{\mathbb{R}/\mathbb{Z}}. \tag{2.3}$$

We will also use the triangle inequalities

$$\|\theta_1 + \theta_2\|_{\mathbb{R}/\mathbb{Z}} \leqslant \|\theta_1\|_{\mathbb{R}/\mathbb{Z}} + \|\theta_2\|_{\mathbb{R}/\mathbb{Z}}; \qquad \|k\theta\|_{\mathbb{R}/\mathbb{Z}} \leqslant |k|\|\theta\|_{\mathbb{R}/\mathbb{Z}} \tag{2.4}$$

for $\theta_1, \theta_2 \in \mathbb{R}/\mathbb{Z}$ and $k \in \mathbb{Z}$ frequently in the sequel, often without further comment.

For any prime p, we (by slight abuse of notation) let $a \mapsto a/p$ be the obvious homomorphism from $\mathbb{Z}/p\mathbb{Z}$ to $\mathbb{R}/\mathbb{Z}$ that maps $a \pmod p$ to $a/p \pmod 1$ for any integer a. We then define $e_p : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ to be the character

$$e_p(a) := e\left(\frac{a}{p}\right) = \mathrm{e}^{2\pi \mathrm{i}a/p}$$

of $\mathbb{Z}/p\mathbb{Z}$.

§3. *High-level overview of argument.* We will establish Theorem 1.1 by establishing the following result, related to the Khinchin-type recurrence theorems mentioned earlier. It will be convenient to introduce the notation

$$\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f}) := \mathbb{E}\mathbf{f}(\mathbf{a})\mathbf{f}(\mathbf{a}+\mathbf{r})\mathbf{f}(\mathbf{a}+2\mathbf{r})\mathbf{f}(\mathbf{a}+3\mathbf{r})$$

whenever $\mathbf{a}, \mathbf{r}$ are random variables on $\mathbb{Z}/p\mathbb{Z}$ and $\mathbf{f} : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ is a random function; of course, the notation can also be applied to deterministic functions $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$. Later on we will also need the conditional variant

$$\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f}|E) := \mathbb{E}(\mathbf{f}(\mathbf{a})\mathbf{f}(\mathbf{a}+\mathbf{r})\mathbf{f}(\mathbf{a}+2\mathbf{r})\mathbf{f}(\mathbf{a}+3\mathbf{r})|E) \tag{3.1}$$

for some events E of non-zero probability. Informally, this quantity counts the density of arithmetic progressions $\mathbf{a}, \mathbf{a}+\mathbf{r}, \mathbf{a}+2\mathbf{r}, \mathbf{a}+3\mathbf{r}$ on the event E weighted by $\mathbf{f}$, where $\mathbf{a}, \mathbf{r}$ need not be drawn uniformly or independently (and $\mathbf{f}$ may also be coupled to $\mathbf{a}, \mathbf{r}$).

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

THEOREM 3.1. *Let p be a prime, let η be a real number with $0 < \eta \leqslant \frac{1}{10}$, and let $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ be a function. Then there exist random variables $\mathbf{a}, \mathbf{r} \in \mathbb{Z}/p\mathbb{Z}$, not necessarily independent, obeying the near-uniform distribution bound*

$$\mathbb{E} f(\mathbf{a}) = \mathbb{E}_{x \in \mathbb{Z}/p\mathbb{Z}} f(x) + O(\eta), \tag{3.2}$$

the recurrence property

$$\Lambda_{\mathbf{a},\mathbf{r}}(f) \geqslant (\mathbb{E} f(\mathbf{a}))^4 - O(\eta), \tag{3.3}$$

and the "thickness" bound

$$\mathbb{P}(\mathbf{r} = 0) \ll \exp(-\eta^{-O(1)})/p. \tag{3.4}$$

We note that a variant of Theorem 3.1 was established by us in [**13**] (answering a question in [**3**]), in which the random variable $\mathbf{a}$ was uniformly distributed in $\mathbb{Z}/p\mathbb{Z}$, the random variable $\mathbf{r}$ was uniformly distributed in a subset of $\mathbb{Z}/p\mathbb{Z}$ of size $\gg_\eta p$ and was independent of $\mathbf{a}$, and the condition (3.4) (which is crucial to the quantitative bound in Theorem 1.1) was not present. Compared to that result, Theorem 3.1 obtains the much more quantitative bound (3.4), but at the expense of no longer enforcing independence between $\mathbf{a}$ and $\mathbf{r}$. The use of non-independent random variables $\mathbf{a}, \mathbf{r}$ is an innovation of this current paper; it is similar to the technique in previous papers of using "factors" (finite partitions) to break up the domain $\mathbb{Z}/p\mathbb{Z}$ into smaller "atoms" such as Bohr sets and analyzing each atom separately. However there will be technical advantages from the more general framework of pairs of independent random variables $\mathbf{a}, \mathbf{r}$. In particular we will be able to avoid some of the boundary issues arising from irregularity of Bohr sets, by using the smoother device of "regular probability distributions" associated to such sets. Although f is allowed to attain negative values in Theorem 3.1, in our applications we shall only be concerned with the case when f is non-negative.

Let us now see how Theorem 1.1 follows from Theorem 3.1. Clearly we may assume that $N \geqslant 100$. Suppose that A is a subset of $\{1, \dots, N\}$ without any non-trivial four-term arithmetic progressions. By Bertrand's postulate, we may find a prime p between (for example) $2N$ and $4N$. If we define $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ to be the indicator function 1_A of A (viewed as a subset of $\mathbb{Z}/p\mathbb{Z}$), then we have

$$\mathbb{E}_{x \in \mathbb{Z}/p\mathbb{Z}} f(x) = \frac{|A|}{p} \tag{3.5}$$

and also

$$f(a) f(a + r) f(a + 2r) f(a + 3r) = 0 \tag{3.6}$$

whenever $a, r \in \mathbb{Z}/p\mathbb{Z}$ with r non-zero. Now let $\mathbf{a}, \mathbf{r}$ be as in Theorem 3.1, with η to be chosen later. From (3.2), (3.3), (3.5) we have

$$\Lambda_{\mathbf{a},\mathbf{r}}(f) \geqslant \left(\frac{|A|}{p}\right)^4 - O(\eta).$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

But by (3.6), (3.4), the left-hand side is $O(\exp(-\eta^{-O(1)})/p)$. Setting $\eta := c\log^{-c} p$ for a sufficiently small absolute constant $c > 0$, we conclude that

$$\left(\frac{|A|}{p}\right)^4 \ll \log^{-c} p$$

and hence $A \ll N\log^{-c/4} N$, giving Theorem 1.1.

Remark. As mentioned previously, the arguments in [**13**] established a bound of the form (3.3) with **a** and **r** independent, and also one could ensure that **a** was uniformly distributed over $\mathbb{Z}/p\mathbb{Z}$. As a consequence, one could establish a variant of Theorem 1.1, namely that for any $N \geqslant 1$, $\eta > 0$, and $A \subset [N]$, one had

$$\frac{|A \cap (A-r) \cap (A-2r) \cap (A-3r)|}{N} \geqslant \left(\frac{|A|}{N}\right)^4 - \eta$$

for $\gg_\eta N$ choices of $0 \leqslant r \leqslant N$. Unfortunately our methods do not seem to provide a good bound of this form due to our coupling together of **a** and **r**.

It remains to establish Theorem 3.1. As in [**3**, **13**], the lower bound (3.3) will ultimately come from the following consequence of the Cauchy–Schwarz inequality that counts solutions to the equation $x - 3y + 3z - w = 0$ for x, y, z, w in some subset of a compact abelian group; this inequality is a specific feature of the theory of length-four progressions that is not available for longer progressions[2].

LEMMA 3.2 (Application of Cauchy–Schwarz). *Let $G = (G, +)$ be a compact abelian group, let μ be the probability Haar measure on G, and let $F : G \to \mathbb{R}$ be a bounded measurable function. Then*

$$\int_G \int_G \int_G F(x)F(y)F(z)F(x-3y+3z)\,d\mu(x)\,d\mu(y)\,d\mu(z) \geqslant \left(\int_G F\,d\mu\right)^4.$$

Proof. Making the change of variables $w = x - 3y$ and using Fubini's theorem, the left-hand side may be rewritten as

$$\int_G \left(\int_G F(w+3y)F(y)\,d\mu(y)\right)^2 d\mu(w),$$

which by the Cauchy–Schwarz inequality is at least

$$\left(\int_G \int_G F(w+3y)F(y)\,d\mu(y)\,d\mu(w)\right)^2.$$

But by a further application of Fubini's theorem, the expression inside the square is $(\int_G F(x)\,d\mu(x))^2$. The claim follows. □

[2] For longer progressions, the relevant constraints coming from nilpotent algebra are significantly more complicated than a single linear equation; see [**35**]. In any event, the counterexamples in [**3**] indicate that no comparable positivity property with polynomial lower bounds will hold for higher length progressions.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

To see the relevance of this lemma to Theorem 3.1, and to motivate the strategy of proof of that theorem, let us first test that theorem on some key examples. To simplify the exposition, our discussion will be somewhat non-rigorous in nature; for instance, we will make liberal use of the non-rigorous symbol $\approx$ without quantifying the nature of the approximation.

Example 1 (A well-distributed pure quadratic factor). Let G be the d-torus $G = (\mathbb{R}/\mathbb{Z})^d$ for some bounded $d = O(1)$, and let $F : G \to [-1, 1]$ be a smooth function (independent of p); for instance, F could be a finite linear combination of characters $\chi : G \to S^1$ of G. Let $\alpha_1, \dots, \alpha_d \in \mathbb{Z}/p\mathbb{Z}$ be "generic" frequencies, in the sense that there are no non-trivial linear relations of the form

$$k_1\alpha_1 + \cdots + k_d\alpha_d = 0 \tag{3.7}$$

with $k_1, \dots, k_d = O(1)$ not all equal to zero. We also introduce some additional frequencies $\beta_1, \dots, \beta_d \in \mathbb{Z}/p\mathbb{Z}$, for which we impose no genericity restrictions. Let $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ be the function

$$f(a) := F(Q(a)),$$

where $Q : \mathbb{Z}/p\mathbb{Z} \to G$ is the quadratic polynomial

$$Q(a) := \left(\frac{\alpha_1 a^2 + \beta_1 a}{p}, \dots, \frac{\alpha_d a^2 + \beta_d a}{p}\right),$$

and where we use the obvious division by zero map $a \mapsto a/p$ from $\mathbb{Z}/p\mathbb{Z}$ to $\mathbb{R}/\mathbb{Z}$. For any tuples $k = (k_1, \dots, k_d) \in \mathbb{Z}^d \equiv \hat{G}$ and $\xi = (\xi_1, \dots, \xi_d) \in G$, we define the dot product

$$k \cdot \xi := k_1\xi_1 + \cdots + k_d\xi_d.$$

Because of our genericity hypothesis on the α_i, we see from Gauss sum estimates that

$$\mathbb{E}_{a \in \mathbb{Z}/p\mathbb{Z}} e(k \cdot Q(a)) \approx 0$$

for any bounded tuple $k \in \mathbb{Z}^d$ when p is large. By the Weyl equidistribution criterion, we thus see that when p is large, the quantity $(\alpha a^2 + \beta a)/p$ becomes equidistributed in G as a ranges over $\mathbb{Z}/p\mathbb{Z}$. In particular, as F was assumed to be smooth, we expect to have

$$\mathbb{E} f(\mathbf{a}) = \mathbb{E}_{a \in \mathbb{Z}/p\mathbb{Z}} f(a) \approx \int_G F(x)\, d\mu(x)$$

if $\mathbf{a}$ is drawn uniformly in $\mathbb{Z}/p\mathbb{Z}$. Now suppose that $\mathbf{r}$ is also drawn uniformly in $\mathbb{Z}/p\mathbb{Z}$, independently of $\mathbf{a}$. The tuple

$$(Q(\mathbf{a}), Q(\mathbf{a} + \mathbf{r}), Q(\mathbf{a} + 2\mathbf{r}), Q(\mathbf{a} + 3\mathbf{r})) \tag{3.8}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

will *not* become equidistributed in G^4, because of the elementary algebraic identity

$$Q(\mathbf{a}) - 3Q(\mathbf{a}+\mathbf{r}) + 3Q(\mathbf{a}+2\mathbf{r}) - Q(\mathbf{a}+3\mathbf{r}) = 0, \tag{3.9}$$

which is a discrete version of the fact that the third derivative of any quadratic polynomial vanishes. However, this turns out to be the *only* constraint on this tuple in the limit $p \to \infty$. Indeed, from the genericity hypothesis on the α_i, one can verify that the quadratic form

$$(a, r) \mapsto k_0 \cdot Q_0(a) + k_1 \cdot Q_0(a+r) + k_2 \cdot Q_0(a+2r) + k_3 \cdot Q_0(a+3r)$$

on $(\mathbb{Z}/p\mathbb{Z})^2$ for bounded tuples $k_0, k_1, k_2, k_3 \in \mathbb{Z}^d$ vanishes if and only if (k_0, k_1, k_2, k_3) is of the form $(k, -3k, 3k, -k)$ for some tuple k, where

$$Q_0(a) := \left(\frac{\alpha_1 a^2}{p}, \ldots, \frac{\alpha_d a^2}{p}\right)$$

denotes the purely quadratic component of $Q(a)$. Using this and a variant of the Weyl equidistribution criterion, one can eventually compute that

$$\Lambda_{\mathbf{a},\mathbf{r}}(f) \approx \int_G \int_G \int_G F(x)F(y)F(z)F(x-3y+3z)\, d\mu(x)\, d\mu(y)\, d\mu(z).$$

Applying Lemma 3.2, we conclude (a heuristic version of) Theorem 3.1 in this case, taking $\mathbf{a}$, $\mathbf{r}$ to be independent uniformly distributed variables on $\mathbb{Z}/p\mathbb{Z}$.

Example 2 (A well-distributed impure quadratic factor). Now we give a "local" version of the first example, in which the function f exhibits "locally quadratic" behaviour rather than "globally quadratic" behaviour. Let $\eta > 0$ be a small parameter, and suppose that p is very large compared to η. We suppose that the cyclic group $\mathbb{Z}/p\mathbb{Z}$ is somehow partitioned into a number $P_1, \ldots, P_m$ of arithmetic progressions; the number m of such progressions should be thought of as being moderately large (e.g. $m \sim \exp(1/\eta^{O(1)})$ for some parameter $\eta > 0$). Consider one such progression, for example $P_c = \{b_c + ns_c : 1 \leqslant n \leqslant N_c\}$ for some $b_c, s_c \in \mathbb{Z}/p\mathbb{Z}$ and some $N_c > 0$; one should think of N_c as being reasonably large, e.g. $N_c \gg \exp(-1/\eta^{O(1)})p$. To each such progression P_c, we associate a torus $G_c = (\mathbb{R}/\mathbb{Z})^{d_c}$ for some bounded d_c with probability Haar measure μ_c, a smooth function $F_c : G_c \to [-1, 1]$, and a collection $\xi_{c,1}, \ldots, \xi_{c,d_c} \in \mathbb{R}/\mathbb{Z}$ of frequencies that are generic in the sense that there does not exist any non-trivial relations of the form

$$k_1\xi_{c,1} + \cdots + k_{d_c}\xi_{c,d_c} = O\left(\frac{1}{N_c}\right) \pmod 1 \tag{3.10}$$

for bounded $k_1, \ldots, k_{d_c} \in \mathbb{Z}$. We then define the function $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ by setting

$$f(b_c + ns_c) := F_c(\xi_{c,d_1}n^2, \ldots, \xi_{c,d_c}n^2)$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

for $1 \leqslant c \leqslant m$ and $1 \leqslant n \leqslant N_c$. One could also add a lower order linear term to the phases $\xi_{c,i}n^2$, as in the preceding example, if desired, but we will not do so here to simplify the exposition slightly.

Within each progression P_c, a Weyl equidistribution analysis (using the genericity hypothesis) reveals that the tuple $(\xi_{c,d_1}n^2, \dots, \xi_{c,d_c}n^2)$ becomes equidistributed in G_c as p becomes large, so that

$$\mathbb{E}_{a\in P_c} f(a) \approx \int_{G_c} F_c(x)\, d\mu_c(x). \tag{3.11}$$

Now we define the random variables $\mathbf{a}, \mathbf{r} \in \mathbb{Z}/p\mathbb{Z}$ as follows. We first select a random element $\mathbf{c}$ from $\{1, \dots, m\}$ with $\mathbb{P}(\mathbf{c} = c) = |P_j|/p$ for $c = 1, \dots, m$. Conditioning on the event that $\mathbf{c}$ is equal to c, we then select $\mathbf{a}$ uniformly at random from P_c, and also select $\mathbf{r}$ uniformly at random from an arithmetic progression of the form

$$\{ns_c : |n| \leqslant \exp(-1/\eta^{-C})N_c\}, \tag{3.12}$$

with $\mathbf{a}$ and $\mathbf{r}$ independent after conditioning on $\mathbf{c} = c$. Note that $\mathbf{a}$ and $\mathbf{r}$ are only *conditionally* independent, relative to the auxiliary variable $\mathbf{c}$; if one does not perform this conditioning, then $\mathbf{a}$ and $\mathbf{r}$ become coupled to each other through their mutual dependence on $\mathbf{c}$.

Without conditioning on $\mathbf{c}$, the random variable $\mathbf{a}$ becomes uniformly distributed on $\mathbb{Z}/p\mathbb{Z}$, thus

$$\mathbb{E} f(\mathbf{a}) = \mathbb{E}_{a\in\mathbb{Z}/p\mathbb{Z}} f(a).$$

Also, from (3.11) we have the conditional expectation

$$\mathbb{E}(f(\mathbf{a})|\mathbf{c} = c) \approx \int_{G_c} F_c(x)\, d\mu_c(x).$$

A modification of the equidistribution analysis from the first example also gives

$$\Lambda_{\mathbf{a},\mathbf{r}}(f|\mathbf{c} = c) \gtrapprox \int_{G_c}\int_{G_c}\int_{G_c} F_c(x)F_c(y)F_c(z)F(x - 3y + 3z)\, d\mu_c(x)\, d\mu_c(y)\, d\mu_c(z),$$

where the conditional quartic form $\Lambda_{\mathbf{a},\mathbf{r}}(f|\mathbf{c} = c)$ was defined in (3.1), and hence by Lemma 3.2 we have

$$\Lambda_{\mathbf{a},\mathbf{r}}(f|\mathbf{c} = c) \gtrapprox (\mathbb{E}(f(\mathbf{a})|\mathbf{c} = c))^4.$$

Averaging in c (weighted by $\mathbb{P}(\mathbf{c} = c)$) to remove the conditional expectation on the left-hand side, and then applying Hölder's inequality, we obtain a heuristic version of Theorem 3.1 in this case.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Example 3 (A poorly distributed pure quadratic factor). We now return to the situation of the first example, except that we no longer impose the genericity hypothesis, that is to say we allow for a non-trivial relation of the form (3.7). Without loss of generality we can take the coefficient k_d of this relation to be non-zero. Because of this relation, the quantity $Q(\mathbf{a})$ studied in the first example and the tuple (3.8) may not necessarily be as equidistributed as before. However, we can use this irregularity of distribution to modify the representation of f (up to a small error) in such a manner as to reduce the number d of quadratic phases involved. Namely, we can write

$$f(a) := \tilde{F}\left(\tilde{Q}(a), \frac{\gamma a}{p}\right)$$

where

$$\tilde{Q}(a) := \left(\frac{k_d^{-1}\alpha_1 a^2 + k_d^{-1}\beta_1 a}{p}, \ldots, \frac{k_d^{-1}\alpha_{d-1} a^2 + k_d^{-1}\beta_{d-1} a}{p}\right),$$
$$\gamma := \beta_d + k_1 k_d^{-1}\beta_1 + \cdots + k_{d-1}k_d^{-1}\beta_{d-1},$$
$$\tilde{F}(x_1, \ldots, x_{d-1}, y) := F(k_d x_1, \ldots, k_d x_{d-1}, -k_1 x_1 - \cdots - k_{d-1}x_{d-1} + y)$$

and where we take advantage of the field structure of $\mathbb{Z}/p\mathbb{Z}$ to locate an inverse k_d^{-1} of k_d in this field. For our quantitative analysis we will run into a technical difficulty with this representation, in that the Lipschitz constant of $\tilde{F}$ will increase by an undesirable amount compared to that of F when one performs this change of variable, at least if one uses the standard metric on the torus. To fix this, we will eventually have to work with more general tori $\prod_{i=1}^{d} \mathbb{R}/\lambda_i\mathbb{Z}$ than the standard torus $(\mathbb{R}/\mathbb{Z})^d$, but we ignore this issue for now to continue with the heuristic discussion.

To remove the dependence on the linear phase $\gamma a/p$, we partition $\mathbb{Z}/p\mathbb{Z}$ into "(shifted) Bohr sets" $B_1, \ldots, B_m$ for some moderately large m (e.g. $m \sim \exp(1/\eta^{-C})$ for some constant $C > 0$), defined by

$$B_c := \left\{a \in \mathbb{Z}/p\mathbb{Z} : \frac{\gamma a}{p} \in \left[\frac{c-1}{m}, \frac{c}{m}\right) \pmod 1\right\}$$

for $c = 1, \ldots, m$. On each Bohr set B_c, we have the approximation

$$f(a) := \tilde{F}_c(\tilde{Q}(a))$$

where $\tilde{F}_c(x, y) := \tilde{F}(x, c/m)$. Using the heuristic that Bohr sets behave like arithmetic progressions, the situation is now similar to that in the second example, with the number of quadratic phases involved reduced from d to $d-1$, except that there may still be some non-trivial relations among the surviving quadratic phases (and one also now has some lower order linear terms in the quadratic phases). To deal with this difficulty, we turn now to the consideration of yet another example.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Example 4 (A poorly distributed impure quadratic factor). We now consider an example that is in some sense a combination of the second and third examples. Namely, we suppose we are in the same situation as in the second example, except that we allow some of the indices c to have "poor quadratic distribution" in the sense that they admit non-trivial relations of the form (3.10). Again we may assume without loss of generality that k_{d_c} is non-zero in such relations. Because of such relations, we no longer expect to have the equidistribution properties that were used in the second example. However, by modifying the calculations in the third example, we can obtain a new representation of f (again allowing for a small error) on each of the progressions P_c with poor quadratic distribution to reduce the number d_c of quadratic polynomials used in that progression by one. Iterating this process a finite number of times, one eventually returns to the situation in the second example in which no non-trivial relations occur, at which point one can (heuristically, at least) verify Theorem 3.1 in this case.

The situation becomes slightly more complicated if one adds a lower order linear term $\zeta_{c,i}n$ to the purely quadratic phases $\xi_{c,i}n^2$ appearing in the second example; this basically is the type of situation one encounters for instance at the conclusion of the third example. In this case, every time one converts a non-trivial relation of the form (3.10) on one of the cells P_c of the partition into a new representation of f on that cell, one must subdivide that cell P_j into smaller pieces, by intersecting P_j with various Bohr sets. However, the resulting sets still behave somewhat like arithmetic progressions, and it turns out that we can still iterate the construction a bounded number of times until no further non-trivial relations between surviving quadratic phases remain on any of the cells of the partition, at which point one can (heuristically, at least) verify Theorem 3.1 in this case (as well as in the case considered in the third example).

Example 5 (A pseudorandom perturbation of a pure quadratic factor). In all the preceding examples, the function $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ under consideration was "locally quadratically structured", in the sense that on local regions such as P_c, the function f could be accurately represented in terms of quadratic phase functions $a \mapsto Q(a)$. This is however not the typical behaviour expected for a general function $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$. A more representative example would be a function of the form

$$f(a) := f_1(a) + f_2(a),$$

where $f_1 : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ is a function of the type considered in the first example, thus

$$f_1(a) = F(Q(a))$$

for some quadratic function $Q : \mathbb{Z}/p\mathbb{Z} \to G$ into a torus $G = (\mathbb{R}/\mathbb{Z})^d$ and some smooth $F : G \to [-1, 1]$, and $f_2 : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ is a function that is *globally Gowers uniform* in the sense that

$$\mathbb{E} \prod_{(\omega_1,\omega_2,\omega_3)\in\{0,1\}^3} f_2(\mathbf{a} + \omega_1\mathbf{h}_1 + \omega_2\mathbf{h}_2 + \omega_3\mathbf{h}_3) \approx 0, \tag{3.13}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

where $\mathbf{a}, \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ are drawn independently and uniformly at random from $\mathbb{Z}/p\mathbb{Z}$. A typical example to keep in mind is when F (and hence f_1) takes values in $[0, 1]$, and $f = \mathbf{f}$ is a random function with $f(a)$ equal to 1 with probability $f_1(a)$ and 0 with probability $1 - f_1(a)$, independently as $a \in \mathbb{Z}/p\mathbb{Z}$ varies; then the $f_2(a)$ for $a \in \mathbb{Z}/p\mathbb{Z}$ become independent random variables of mean zero, and the global Gowers uniformity can be established with high probability using tools such as the Chernoff inequality.

From the standard theory of the Gowers norms (see e.g. [**33**, Ch. 11]), one can use the global Gowers uniformity of f_2, combined with a number of applications of the Cauchy–Schwarz inequality, to establish a "generalized von Neumann theorem" that, in our current context, implies that f and f_1 globally count approximately the same number of length-four progressions in the sense that

$$\Lambda_{\mathbf{a},\mathbf{r}}(f) \approx \Lambda_{\mathbf{a},\mathbf{r}}(f_1); \tag{3.14}$$

similarly one also has

$$\mathbb{E} f(\mathbf{a}) \approx \mathbb{E} f_1(\mathbf{a}). \tag{3.15}$$

As a consequence, Theorem 3.1 for such functions follows (heuristically, at least) from the analysis of the first example, at least if one assumes the genericity of the frequencies $\xi_1, \ldots, \xi_d$.

Example 6 (A pseudorandom perturbation of an impure quadratic factor). We now consider a situation that is to the second example as the fifth example was to the first. Namely, we consider a function of the form

$$f(a) := f_1(a) + f_2(a),$$

where $f_1 : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ is a function of the type considered in the second example, thus

$$f_1(b_c + ns_c) := F_c(\xi_{c,d_1} n^2, \ldots, \xi_{c,d_c} n^2)$$

for $c = 1, \ldots, m$ and $n = 1, \ldots, N_c$. As for the function $f_2 : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$, global Gowers uniformity of f_2 will be too weak of a hypothesis for our purposes, because the random variable $\mathbf{r}$ appearing in the second example is now localized to a significantly smaller region than $\mathbb{Z}/p\mathbb{Z}$. Instead, we will require the *local Gowers uniformity* hypothesis

$$\mathbb{E} \prod_{(\omega_1,\omega_2,\omega_3)\in\{0,1\}^3} f_2(\mathbf{a} + \omega_1\mathbf{h}_1 + \omega_2\mathbf{h}_2 + \omega_3\mathbf{h}_3) \approx 0, \tag{3.16}$$

where $\mathbf{a}$ is now the random variable from the second example (in particular, $\mathbf{a}$ depends on the auxiliary random variable $\mathbf{c}$), and once one conditions on an event $\mathbf{c} = c$ for $c = 1, \ldots, m$, one draws $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ independently of each other and from $\mathbf{a}$, and each $\mathbf{h}_i$ drawn uniformly from an arithmetic progression of the form

$$\{ns_c : |n| \leqslant \exp(-1/\eta^{-C_i})N_c\}, \tag{3.17}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

for some constant $C_i > 0$ (for technical reasons, it is convenient to allow these constants C_1, C_2, C_3 to be different from each other, and also to be larger than the constant C appearing in (3.12), so that $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ range over a narrower scale than $\mathbf{r}$). As with $\mathbf{a}$ and $\mathbf{r}$, the random variables $\mathbf{a}, \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$ are now only *conditionally* independent relative to the auxiliary variable $\mathbf{c}$, but are not independent of each other without this conditioning, as they are coupled to each other through $\mathbf{c}$.

As it turns out, once one assumes this local Gowers uniformity of f_2, one can modify the Cauchy–Schwarz arguments used to establish the global generalized von Neumann theorem to obtain the approximations (3.14), (3.15) for the random variables $\mathbf{a}, \mathbf{r}$ considered in the second example, at which point Theorem 3.1 for this choice of f follows (heuristically, at least) from the analysis of that example, at least if one assumes that there are no non-trivial relations of the form (3.10).

Example 7 (Non-pseudorandom perturbation of a pure quadratic factor). We now modify the fifth example by replacing the hypothesis (3.13) by its negation

$$\mathbb{E} \prod_{(\omega_1,\omega_2,\omega_3)\in\{0,1\}^3} f_2(\mathbf{a} + \omega_1\mathbf{h}_1 + \omega_2\mathbf{h}_2 + \omega_3\mathbf{h}_3) \gg 1 \tag{3.18}$$

(it is not difficult to show that the left-hand side is non-negative). In this case, the generalized von Neumann theorem used in that example does not give a good estimate. However, in this situation one can apply the inverse theorem for the Gowers norm established by us in [**14**]. To obtain good quantitative bounds, we will use the version of that theorem that involves local correlation with quadratic objects (as opposed to a somewhat weak global correlation with a single "locally quadratic" object). Namely, if (3.18) holds, then one can partition $\mathbb{Z}/p\mathbb{Z}$ into a moderately large (e.g. $O(\exp(1/\eta^{-O(1)}))$) number of pieces $P_1, \dots, P_m$, such that on each piece P_c, the function f_2 correlates with a "quadratically structured" object. The precise statement is somewhat technical to state, but one simple special case of this conclusion is that the pieces $P_1, \dots, P_m$ are arithmetic progressions as in the second example, and for a "significant number" of the progressions

$$P_c = \{b_c + ns_c : 1 \leqslant n \leqslant N_c\}$$

there exists a frequency $\xi_c \in \mathbb{R}/\mathbb{Z}$ such that

$$|\mathbb{E}_{1\leqslant n\leqslant N_c} f_2(b_c + ns_c)e(-\xi_c n^2)| \gg 1.$$

(In general, one would take P_c to be Bohr sets of moderately high rank, rather than arithmetic progressions, and the phase $a \mapsto \xi_c a^2/p$ would have to be replaced by a more general locally quadratic phase function on such a Bohr set, but we ignore these technicalities for the current informal discussion.) From this and the cosine rule, it is possible to find a function $g : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ that is equal to (the real part of) a scalar multiple of the quadratic phases

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

$b_c + ns_c \mapsto e(\xi_c n^2)$ on each progression P_c, such that $f_2 + g$ has an *energy decrement* compared to f_2 in the sense that

$$\mathbb{E}_{a\in\mathbb{Z}/p\mathbb{Z}}(f_2(a)+g(a))^2 \leqslant \mathbb{E}_{a\in\mathbb{Z}/p\mathbb{Z}} f_2(a)^2 - \eta^C \tag{3.19}$$

for some constant $C > 0$. In this situation, we can modify the decomposition $f = f_1 + f_2$ by adding g to f_2 and subtracting it from f_1. (Strictly speaking, this may make f_1 and f_2 range slightly outside of $[-1, 1]$, but because f itself ranges in $[-1, 1]$, it turns out to be relatively easy to modify f_1, f_2 further to rectify this problem.) The new function f_1 has a similar "quadratic structure" to the previous function f_1, except that the quadratic structure is now localized to the cells $P_1, \dots, P_m$ of the partition of $\mathbb{Z}/p\mathbb{Z}$, and the number of quadratic functions has been increased by one. If the new function f_2 is now locally Gowers uniform in the sense of (3.16), then we are now essentially in the situation of the sixth example (at least if there are no non-trivial relations of the form (3.10)), and we can (heuristically at least) conclude Theorem 3.1 in this case by the previous analysis. If f_2 is locally Gowers uniform but there are additionally some relations of the form (3.10), then one can hope to adapt the analysis of the fourth example to reduce the quadratic complexity of f_1 on all the poorly distributed cells, at which point one restarts the analysis. If however f_2 remains non-uniform, then we need to argue using the analysis of the next and final example.

Example 8 (Non-pseudorandom perturbation of an impure quadratic factor). Our final and most difficult example will be as to the sixth example as the seventh example was to the fifth. Namely, we modify the sixth example by assuming that the negation of (3.16) holds. Equivalently, one has the lower bound

$$\mathbb{E}\Big(\prod_{(\omega_1,\omega_2,\omega_3)\in\{0,1\}^3} f_2(\mathbf{a}+\omega_1\mathbf{h}_1+\omega_2\mathbf{h}_2+\omega_3\mathbf{h}_3)|\mathbf{c}=c\Big) \gg 1 \tag{3.20}$$

on the local Gowers norm for a "significant fraction" of the $c = 1, \dots, m$.

At the qualitative level, the inverse theorem in [**14**] for the global Gowers norm allows one to also deduce a similar conclusion starting from the hypothesis (3.20). However, the quantitative bounds obtained by this approach turn out to be too poor for the purposes of establishing Theorems 3.1 or 1.1. Instead, one must obtain a quantitative local inverse theorem for the Gowers norm that has reasonably good bounds (of polynomial type) on the amount of correlation that is (locally) attained. Establishing such a theorem is by far the most complicated and lengthy component of this paper, although broadly speaking it follows the same strategy as previous theorems of this type in [**11**, **14**]. If one takes this local inverse theorem for granted, then roughly speaking what we can then conclude from the hypothesis (3.20) is that for a significant number of $c = 1, \dots, m$, one can partition the cell P_c into subcells $P_{c,1}, \dots, P_{c,m_c}$, and locate a "locally quadratic phase function" $\phi_{c,i} : P_{c,i} \to \mathbb{R}/\mathbb{Z}$ on each such subcell (generalizing the functions $b_c + ns_c \mapsto e(\xi_c n^2)$ from the previous example), such that

$$|\mathbb{E}_{a\in P_{c,i}} f_2(b_{c,i})e(-\phi_{c,i}(a))| \gg 1$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

for a significant fraction of the c, i. Using this, one can again obtain an energy decrement of the form (3.19), where now g is (the real part of) a scalar multiple of the functions $a \mapsto e(\phi_{c,i}(a))$ on each $P_{c,i}$. By arguing as in the sixth example, one can then modify f_1 and f_2 in such a way that the "energy" $\mathbb{E} f_2(\mathbf{a})^2$ decreases significantly, while f_1 is now locally quadratically structured on a somewhat finer partition of $\mathbb{Z}/p\mathbb{Z}$ than the original partition $P_1, \dots, P_m$, with the number of quadratic phases needed to describe f_1 on each partition having increased by one. If the function f_2 is now locally Gowers uniform (with respect to a new set of random variables $\mathbf{a}, \mathbf{r}$ adapted to this finer partition), and there are no non-trivial relations of the form we can now (heuristically) conclude Theorem 3.1 from the analysis of the sixth example, assuming the addition of the new quadratic phase has not introduced relations of the form (3.10). If such relations occur, though, one can hope to adapt the analysis of the fourth example to reduce the quadratic complexity of the poorly distributed cells, perhaps at the cost of further subdivision of the cells. Finally, if the new version of f_2 remains non-uniform with respect to the finer partition, then one iterates the analysis of this example to reduce the energy of f_2 further. This process cannot continue indefinitely due to the non-negativity of the energy (and also because none of the other steps in the iteration will cause a significant increase in energy). Because of this, one can hope to cover all cases of Theorem 3.1 by some complicated iteration of the eight arguments described above.

Having informally discussed the eight key examples for Theorem 3.1, we return now to the task of proving this theorem rigorously.

It will be convenient to work throughout the rest of the paper with a fixed choice

$$1 < C_1 < C_2 < \cdots < C_5$$

of absolute constants, with each C_i assumed to be sufficiently large depending on the previous $C_1, \dots, C_{i-1}$. For instance, for sake of concreteness one could choose $C_i := 2^{2^{100i}}$; of course, other choices are possible. The implied constants in the $O(\,)$ notation will not depend on the C_i unless otherwise specified. These constants will serve as exponents for various scales η^{-C_i} that will appear in our analysis, with the point being that any scale of the form η^{-C_i} for $i = 2, \dots, 5$ is extremely tiny with respect to any polynomial combination of the previous scales $\eta^{-C_1}, \dots, \eta^{-C_{i-1}}$.

In all of the eight examples considered above, the function f was approximated by some "quadratically structured" function, usually denoted f_1, with the approximation being accurate in various senses with respect to some pair $(\mathbf{a}, \mathbf{r})$ of random variables. The rigorous argument will similarly approximate f by a quadratically structured object; it will be convenient to make this object a *random* function $\mathbf{f}$ rather than a deterministic one (though as it turns out, this function will become deterministic again once an auxiliary random variable $\mathbf{c}$ is fixed). The precise definition of "quadratically structured" will be rather technical, and will eventually be given in Definition 6.1. For now, we shall abstract the properties of "quadratic structure" that we will need,

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

in the following proposition involving an abstract directed graph $G = (V, E)$ (encoding the "structured local approximants"), which we will construct more explicitly later. We will shortly iterate this proposition to establish Theorem 3.1 and hence Theorem 1.1.

PROPOSITION 3.3 (Main proposition, abstract form). *Let η be a real number with $0 < \eta \leqslant \frac{1}{10}$, and let p be a prime with*

$$p \geqslant \exp(\eta^{-3C_5}). \tag{3.21}$$

Let $f : \mathbb{Z}/p\mathbb{Z} \to [0, 1]$ be a function. Then there exist the following:

(a) *a (possibly infinite) directed graph $G = (V, E)$, with elements $v \in V$ referred to as structured local approximants, and the notation $v \to v'$ used to denote the existence of a directed edge from one structured local approximant v to another v';*

(b) *a triple $(\mathbf{a}_v, \mathbf{r}_v, \mathbf{f}_v)$ associated to f and to each structured local approximant $v \in V$, where $\mathbf{a}_v, \mathbf{r}_v$ are random variables in $\mathbb{Z}/p\mathbb{Z}$, and $\mathbf{f}_v : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ is a random function (with $\mathbf{a}_v, \mathbf{r}_v, \mathbf{f}_v$ not assumed to be independent);*

(c) *a quadratic dimension $d_2(v) \in \mathbb{N}$ assigned to each vertex $v \in V$;*

(d) *a poorly distributed quadratic dimension $d_2^{\mathrm{poor}}(v) \in \mathbb{N}$ assigned to each vertex $v \in V$, with $0 \leqslant d_2^{\mathrm{poor}}(v) \leqslant d_2(v)$; and*

(e) *an initial approximant $v_0 \in V$, with $d_2(v_0) = 0$ (and hence $d_2^{\mathrm{poor}}(v_0) = 0$).*

Furthermore, whenever a structured local approximant $v_k \in V$ can be reached from v_0 by a path $v_0 \to v_1 \to \cdots \to v_k$ with $0 \leqslant k \leqslant 8\eta^{-2C_2}$, then the following properties are obeyed:

(i) *one has the "thickness" condition*

$$\mathbb{P}(\mathbf{r}_{v_k} = 0) \ll \exp(3\eta^{-C_5})/p; \tag{3.22}$$

(ii) *we have the almost uniformity condition*

$$|\mathbb{E} f(\mathbf{a}_{v_k}) - \mathbb{E}_{a \in \mathbb{Z}/p\mathbb{Z}} f(a)| \leqslant \eta; \tag{3.23}$$

(iii) *bad approximation implies energy decrement: if*

$$|\mathbb{E}\mathbf{f}_{v_k}(\mathbf{a}_{v_k}) - f(\mathbf{a}_{v_k})| > \eta \tag{3.24}$$

or

$$|\Lambda_{\mathbf{a}_{v_k}, \mathbf{r}_{v_k}}(\mathbf{f}_{v_k}) - \Lambda_{\mathbf{a}_{v_k}, \mathbf{r}_{v_k}}(f)| > \eta \tag{3.25}$$

then there exists a structured local approximant $v_{k+1} \in V$ with $v_k \to v_{k+1}$ such that

$$\mathbb{E}|f(\mathbf{a}_{v_{k+1}}) - \mathbf{f}_{v_{k+1}}(\mathbf{a}_{v_{k+1}})|^2 \leqslant \mathbb{E}|f(\mathbf{a}_{v_k}) - \mathbf{f}_{v_k} k(\mathbf{a}_{v_k})|^2 - \eta^{C_2}$$

and

$$d_2(v_{k+1}) \leqslant d_2(v_k) + 1.$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

(iv) *failure of "Khinchin-type recurrence" implies dimension decrement: if*

$$\Lambda_{\mathbf{a}_{v_k},\mathbf{r}_{v_k}}(\mathbf{f}_{v_k}) \leqslant (\mathbb{E}\mathbf{f}_{v_k}(\mathbf{a}_{v_k}))^4 - \eta, \tag{3.26}$$

then there exists a structured local approximant $v_{k+1} \in V$ with $v_k \to v_{k+1}$ obeying the bounds

$$\begin{gathered}\mathbb{E}|f(\mathbf{a}_{v_{k+1}}) - \mathbf{f}_{v_{k+1}}(\mathbf{a}_{v_{k+1}})|^2 \leqslant \mathbb{E}|f(\mathbf{a}_{v_k}) - \mathbf{f}_{v_k}(\mathbf{a}_{v_k})|^2 + \eta^{3C_2},\\ d_2(v_{k+1}) \leqslant d_2(v_k),\\ d_2^{\mathrm{poor}}(v_{k+1}) \leqslant d_2^{\mathrm{poor}}(v_k) - 1.\end{gathered}$$

The proof of this proposition will occupy the remainder of the paper. For now, let us see how this proposition implies Theorem 3.1. Let p, η, f be as in that theorem, and let $C_1, \dots, C_5$ be as above. If the largeness criterion (3.21) fails, then we may set $\mathbf{r} := 0$, $\mathbf{f} := f$, and draw $\mathbf{a}$ uniformly at random from $\mathbb{Z}/p\mathbb{Z}$, and it is easy to see that the conclusions of Theorem 3.1 are obeyed (with (3.3) following from Hölder's inequality). Thus we may assume without loss of generality that (3.21) holds.

Let $G = (V, E)$, v_0, $d_2(\,)$, $d_2^{\mathrm{poor}}(\,)$, and $(\mathbf{a}_v, \mathbf{r}_v, \mathbf{f}_v)$ be as in Proposition 3.3. Suppose first that there exists a structured local approximant $v_k \in V$ that can be reached from v_0 by a path of length at most $8\eta^{-2C_2}$, and for which none of the inequalities (3.24)–(3.26) hold, that is to say one has the bounds

$$|\mathbb{E}\mathbf{f}_{v_k}(\mathbf{a}_{v_k}) - f_{v_k}(\mathbf{a}_{v_k})| \leqslant \eta, \tag{3.27}$$

$$|\Lambda_{\mathbf{a}_{v_k},\mathbf{r}_{v_k}}(\mathbf{f}_{v_k}) - \Lambda_{\mathbf{a}_{v_k},\mathbf{r}_{v_k}}(f_{v_k})| \leqslant \eta \tag{3.28}$$

$$\Lambda_{\mathbf{a}_{v_k},\mathbf{r}_{v_k}}(\mathbf{f}_{v_k}) > (\mathbb{E}\mathbf{f}_{v_k}(\mathbf{a}_{v_k}))^4 - \eta. \tag{3.29}$$

From (3.29), (3.28), (3.27) and the triangle inequality (and the boundedness of $\mathbf{f}_{v_k}$, f) we conclude that

$$\Lambda_{\mathbf{a}_{v_k},\mathbf{r}_{v_k}}(f_{v_k}) > (\mathbb{E}f(\mathbf{a}_{v_k}))^4 - O(\eta);$$

combining this with (3.22) and (3.23) we see that the random variables $\mathbf{a}_{v_k}, \mathbf{r}_{v_k}$ obey the properties required of Theorem 3.1. Thus we may assume for sake of contradiction that this situation never occurs, which by Proposition 3.3 implies that whenever $v_k \in V$ is a structured local approximant that can be reached from v_0 by a path of length at most $8\eta^{-2C_2}$, then the conclusions of at least one of (iii) and (iv) hold. Iterating this we may therefore construct a path

$$v_0 \to v_1 \to \cdots \to v_{k_0+1}$$

with

$$k_0 := \lfloor 8\eta^{-2C_2} \rfloor, \tag{3.30}$$

such that for every $0 \leqslant k \leqslant k_0$, one either has the energy decrement bounds

$$\begin{gathered}\mathbb{E}|f(\mathbf{a}_{v_{k+1}}) - \mathbf{f}_{k+1}(\mathbf{a}_{v_{k+1}})|^2 \leqslant \mathbb{E}|f(\mathbf{a}_{v_k}) - \mathbf{f}_k(\mathbf{a}_{v_k})|^2 - \eta^{C_2},\\ d_2(v_{k+1}) \leqslant d_2(v_k) + 1\end{gathered}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

or the dimension decrement bounds

$$\mathbb{E}|f(\mathbf{a}_{v_{k+1}}) - \mathbf{f}_{k+1}(\mathbf{a}_{v_{k+1}})|^2 \leqslant \mathbb{E}|f(\mathbf{a}_{v_k}) - \mathbf{f}_k(\mathbf{a}_{v_k})|^2 + \eta^{3C_2},$$
$$d_2(v_{k+1}) \leqslant d_2(v_k),$$
$$d_2^{\text{poor}}(v_{k+1}) \leqslant d_2^{\text{poor}}(v_k) - 1.$$

Since v_0 already has the minimum quadratic dimension $d_2^{\text{poor}}(v_0) = 0$, we see that we must experience an energy decrement at the $k = 0$ stage. Also, if k is the jth index to experience an energy decrement, we see that $d_2^{\text{poor}}(v_{k+1}) \leqslant d_2(v_{k+1}) \leqslant j$, and so one can have at most j consecutive dimension decrements after the kth stage; in other words, we must experience another energy decrement within $j+1$ steps. By definition of k_0, we have $\sum_{0\leqslant j\leqslant 2\eta^{-C_2}}(j+1) < k_0$ if C_2 is large enough. We conclude that at least $2\eta^{-C_2}$ energy decrements occur within the path $v_0 \to \cdots \to v_{k_0+1}$. This implies that

$$\mathbb{E}|f(\mathbf{a}_{v_{k_0+1}}) - \mathbf{f}_{k_0+1}(\mathbf{a}_{v_{k_0+1}})|^2 \leqslant \mathbb{E}|f(\mathbf{a}_{v_0}) - \mathbf{f}_{k+1}(\mathbf{a}_{v_0})|^2 - (2\eta^{-C_2})\eta^{C_2} + k_0\eta^{3C_2}.$$

But if C_2 is sufficiently large, this implies from (3.30) that

$$\mathbb{E}|f(\mathbf{a}_{v_{k_0+1}}) - \mathbf{f}_{k_0+1}(\mathbf{a}_{v_{k_0+1}})|^2 < \mathbb{E}|f(\mathbf{a}_{v_0}) - \mathbf{f}_0(\mathbf{a}_{v_0})|^2 - 4$$

(for example), which leads to a contradiction because the left-hand side is clearly non-negative, and the right-hand side non-positive. This gives the desired contradiction that establishes Theorem 3.1 and hence Theorem 1.1.

It remains to establish Proposition 3.3. This will occupy the remaining portions of the paper.

§4. *Bohr sets.* To define and manipulate the "structured local approximants" that appear in Proposition 3.3, we will need to develop the theory of two mathematical objects. The first is that of a *Bohr set*, which will be covered in this section; the second is that of a *dilated torus*, which we will discuss in the next section.

Definition 4.1 (Bohr set). A subset S of $\mathbb{Z}/p\mathbb{Z}$ is said to be *non-degenerate* if it contains at least one non-zero element. In this case we define the dual S-norm

$$\|a\|_{S^\perp} := \sup_{\xi\in S} \left\| \frac{a\xi}{p} \right\|_{\mathbb{R}/\mathbb{Z}}$$

for any $a \in \mathbb{Z}/p\mathbb{Z}$, and then define the *Bohr set* $B(S, \rho) \subset \mathbb{Z}/p\mathbb{Z}$ for any $\rho > 0$ by the formula

$$B(S, \rho) := \{a \in \mathbb{Z}/p\mathbb{Z} : \|a\|_{S^\perp} < \rho\}$$

where $\|\theta\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance from θ to the nearest integer. We refer to S as the *set of frequencies* of the Bohr set, ρ as the *radius*, and $|S|$ as the *rank* of the Bohr set. We also define the shifted Bohr sets

$$n + B(S, \rho) := \{a + n : a \in B(S, \rho)\}$$

for any $n \in \mathbb{Z}/p\mathbb{Z}$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

From (2.4) we have the triangle inequalities

$$\|a+b\|_{S^\perp} \leqslant \|a\|_{S^\perp} + \|b\|_{S^\perp}; \qquad \|ka\|_{S^\perp} \leqslant |k|\|a\|_{S^\perp} \tag{4.1}$$

for $a, b \in \mathbb{Z}/p\mathbb{Z}$ and $k \in \mathbb{Z}$; also we trivially have

$$\|a\|_{S^\perp} \leqslant \|a\|_{(S')^\perp}$$

if $S \subset S'$ and $a \in \mathbb{Z}/p\mathbb{Z}$, or equivalently that $B(S', \rho) \subset B(S, \rho)$ for $\rho > 0$. We will frequently use these inequalities in the sequel, usually without further comment. In Lemma 4.6 below, we will show that $\|\|_{S^\perp}$ is "dual" to a certain word norm $\|\|_S$ on $\mathbb{Z}/p\mathbb{Z}$. One could also define Bohr sets in the case when S is degenerate, but this creates some minor complications in our arguments, so we remove this case from our definition of a Bohr set.

We have the following standard size bounds for Bohr sets, whose proof may be found in [**33**, Lemma 4.20].

LEMMA 4.2. *If $B(S, \rho)$ is a Bohr set, then $|B(S, \rho)| \geqslant \rho^{|S|}p$ and $|B(S, 2\rho)| \leqslant 4^{|S|}|B(S, \rho)|$.*

In previous work on Roth-type theorems, one sometimes restricts attention to *regular Bohr sets*, as first introduced in [**6**]; see [**33**, §4.4] for some discussion of this concept. Due to our use of the probabilistic method, we will be able to work with a technically simpler and "smoothed out" version of a regular Bohr set, which we call the *regular probability distribution* on a Bohr set.

Definition 4.3. Let $B(S, \rho)$ be a Bohr set. The *regular probability distribution* $\mathfrak{p}_{B(S,\rho)} : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ associated to $B(S, \rho)$ is the function defined by the formula

$$\mathfrak{p}_{B(S,\rho)}(a) := 2\int_{1/2}^{1} \frac{1_{B(S,t\rho)}(a)}{|B(S, t\rho)|}\, dt; \tag{4.2}$$

it is easy to see (from Fubini's theorem) that this is indeed a probability distribution on $\mathbb{Z}/p\mathbb{Z}$. A random variable $\mathbf{a} \in \mathbb{Z}/p\mathbb{Z}$ is said to be *drawn regularly* from $B(S, \rho)$ if it has probability density function $\mathfrak{p}_{B(S,\rho)}$, thus $\mathbb{P}(\mathbf{a} = a) = \mathfrak{p}_{B(S,\rho)}(a)$ for all $a \in \mathbb{Z}/p\mathbb{Z}$.

More generally, for any shifted Bohr set $n + B(S, \rho)$, we define the regular probability distribution $\mathfrak{p}_{n+B(S,\rho)} : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ by the formula

$$\mathfrak{p}_{n+B(S,\rho)}(a) := \mathfrak{p}_{B(S,\rho)}(a-n),$$

and say that $\mathbf{a}$ is *drawn regularly* from $n + B(S, \rho)$ if it has probability distribution $\mathfrak{p}_{n+B(S,\rho)}$.

Informally, to draw a random variable $\mathbf{a}$ regularly from $n + B(S, \rho)$, one should draw it uniformly from $n + B(S, \mathbf{t}\rho)$, where $\mathbf{t}$ is itself selected uniformly at random from the interval $[1/2, 1]$. Note that if $\mathbf{a}$ is drawn regularly from

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

$n + B(S, \rho)$, then $m + \mathbf{a}$ will be drawn regularly from $m + n + B(S, \rho)$ for any $m \in \mathbb{Z}/p\mathbb{Z}$, and similarly $k\mathbf{a}$ will be drawn from $kn + B(k^{-1} \cdot S, \rho)$ for any non-zero $k \in \mathbb{Z}/p\mathbb{Z}$, where $k^{-1} \cdot S := \{k^{-1}\xi : \xi \in S\}$ is the dilate of the frequency set S by k^{-1}.

From Lemma 4.2 we see that if $\mathbf{a}$ is drawn regularly from a shifted Bohr set $n + B(S, \rho)$, then

$$\mathbb{P}(\mathbf{a} = a) \leqslant \frac{1}{(\rho/2)^{|S|}p} \tag{4.3}$$

for all $a \in \mathbb{Z}/p\mathbb{Z}$. In practice, this will mean that the influence of any given value of $\mathbf{a}$ will be negligible.

The presence of the averaging parameter t in (4.2) allows for the following very convenient approximate translation-invariance property. Given two random variables $\mathbf{a}, \mathbf{a}'$ taking values in a finite set A, we define the *total variation distance* between the two to be the quantity

$$d_{\mathrm{TV}}(\mathbf{a}, \mathbf{a}') := \sum_{a \in A} |\mathbb{P}(\mathbf{a} = a) - \mathbb{P}(\mathbf{a}' = a)|,$$

or equivalently

$$d_{\mathrm{TV}}(\mathbf{a}, \mathbf{a}') = \sup_f |\mathbb{E}f(\mathbf{a}) - \mathbb{E}f(\mathbf{a}')|$$

where $f : A \to \mathbb{C}$ ranges over 1-bounded functions.

The next lemma gives some approximate translation-invariance properties of Bohr sets. Its proof is a thinly disguised version of the arguments of Bourgain [6].

LEMMA 4.4. *Let $n + B(S, \rho)$ be a shifted Bohr set, and let $\mathbf{a}$ be drawn regularly from $B(S, \rho)$. Let $B(S', \rho')$ be another Bohr set with $S' \supset S$.*

(i) *If $h \in B(S', \rho')$, then $\mathbf{a}$ and $\mathbf{a} + h$ differ in total variation by at most $O(|S|\rho'/\rho)$.*

(ii) *More generally, if $\mathbf{h}$ is a random variable independent of $\mathbf{a}$ that takes values in $B(S', \rho')$, then $\mathbf{a}$ and $\mathbf{a} + \mathbf{h}$ differ in total variation by at most $O(|S|\rho'/\rho)$.*

Proof. To prove (i), it suffices to show that

$$\mathbb{E}f(\mathbf{a} + h) = \mathbb{E}f(\mathbf{a}) + O\left(|S|\frac{\rho'}{\rho}\right)$$

for any 1-bounded function $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$; the claim (ii) then also follows by conditioning $\mathbf{h}$ to a fixed value $h \in B(S', \rho')$, then multiplying by $\mathbb{P}(\mathbf{h} = h)$ and summing over h.

By translating f by n, we may assume that $n = 0$. We may assume that $\rho' \leqslant \rho/10|S|$, as the claim is trivial otherwise.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

From (4.2) we have

$$\mathbb{E} f(\mathbf{a}) = 2\int_{1/2}^{1} \sum_{a \in \mathbb{Z}/p\mathbb{Z}} f(a) \frac{1_{B(S,t\rho)}(a)}{|B(S,t\rho)|}\, dt$$

and

$$\mathbb{E} f(\mathbf{a}+h) = 2\int_{1/2}^{1} \sum_{a \in \mathbb{Z}/p\mathbb{Z}} f(a) \frac{1_{B(S,t\rho)-h}(a)}{|B(S,t\rho)|}\, dt$$

so by the triangle inequality it suffices to show that

$$\int_{1/2}^{1} \frac{\sum_{a \in \mathbb{Z}/p\mathbb{Z}} |1_{B(S,t\rho)}(a) - 1_{B(S,t\rho)-h}(a)|}{|B(S,t\rho)|}\, dt \ll |S| \frac{\rho'}{\rho}. \tag{4.4}$$

By the triangle inequality, the integrand here is bounded above by 2. Also, from (4.1), we see that any a for which $1_{B(S,t\rho)-h}(a) \neq 1_{B(S,t\rho)}(a)$ lies in the "annulus" $B(S, t\rho + \rho')\backslash B(S, t\rho - \rho')$. We conclude that the left-hand side of (4.4) is bounded by

$$\int_{1/2}^{1} O\left(\min\left(\frac{|B(S,t\rho+\rho')| - |B(S,t\rho-\rho')|}{|B(S,t\rho-\rho')|}, 1\right)\right) dt$$

which, using the elementary bound $\min(x-1, 1) \ll \log x$ for $x \geqslant 1$, can be bounded in turn by

$$O\left(\int_{1/2}^{1} \log \frac{|B(S,t\rho+\rho')|}{|B(S,t\rho-\rho')|}\, dt\right).$$

The integral telescopes to

$$O\left(\int_{1}^{1+\rho'/\rho} \log |B(S,t\rho)|\, dt - \int_{1/2-\rho'/\rho}^{1/2} \log |B(S,t\rho)|\, dt\right)$$

which can be bounded in turn by

$$O\left(\frac{\rho'}{\rho} \log \frac{|B(S,2\rho)|}{|B(S,\rho/4)|}\right).$$

The claim now follows from Lemma 4.2. □

We will be interested in the Fourier coefficients $\mathbb{E} e_p(\lambda \mathbf{n}) = \mathbb{E} e(\lambda \mathbf{n}/p)$ of random variables $\mathbf{n}$ drawn regularly from Bohr sets $B(S, \rho)$. As was noted by Bourgain [6], these coefficients are controlled by a "word norm" $\|\|_S$, defined as follows.

Definition 4.5 (Word norm). If $S \subset \mathbb{Z}/p\mathbb{Z}$ is non-degenerate, and a is an element of $\mathbb{Z}/p\mathbb{Z}$, we define the *word norm* $\|a\|_S$ of a to be the minimum value of $\sum_{s \in S} |n_s|$, where $(n_s)_{s \in S} \in \mathbb{Z}^S$ ranges over tuples of integers such that one has a representation $a = \sum_{s \in S} n_s s$; note that such a representation always exists because S is non-degenerate.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Similarly to (4.1), we observe the triangle inequalities

$$\|a+b\|_S \leqslant \|a\|_S + \|b\|_S; \qquad \|ka\|_S \leqslant |k|\|a\|_S \tag{4.5}$$

for $a, b \in \mathbb{Z}/p\mathbb{Z}$ and $k \in \mathbb{Z}$, which we will use frequently in the sequel, often without further comment.

We now give a duality relationship between the word norm $\|\|_S$ and the dual S-norm $\|\|_{S^\perp}$.

LEMMA 4.6 (Duality). *Let S be a non-degenerate subset of $\mathbb{Z}/p\mathbb{Z}$, and let $\lambda \in \mathbb{Z}/p\mathbb{Z}$:*

(i) *for every $n \in \mathbb{Z}/p\mathbb{Z}$, one has $\|n\lambda/p\|_{\mathbb{R}/\mathbb{Z}} \leqslant \|n\|_{S^\perp}\|\lambda\|_S$;*
(ii) *conversely, if one has the estimate $\|n\lambda/p\|_{\mathbb{R}/\mathbb{Z}} \leqslant A\|n\|_{S^\perp}$ for some $A \geqslant 1$ and all $n \in \mathbb{Z}/p\mathbb{Z}$, then $\|\lambda\|_S \ll |S|^{3/2}A$.*

Proof. To prove (i), we simply observe (using (2.4)) that for any $n \in \mathbb{Z}/p\mathbb{Z}$, one has

$$\|n\lambda/p\|_{\mathbb{R}/\mathbb{Z}} = \left\|\sum_{\xi\in S} a_\xi \frac{n\xi}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \sum_{\xi\in S}|a_\xi|\left\|\frac{n\xi}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \sum_{\xi\in S}|a_\xi|\|n\|_{S^\perp} \leqslant \|\lambda\|_S\|n\|_{S^\perp}$$

as desired, where $\lambda = \sum_{\xi\in S} a_\xi \xi$ is a representation of λ that minimizes $\sum_{\xi\in S}|\xi|$.

Estimates such as (ii) go back to the work of Bourgain [**6**]. We will prove this claim by a Fourier-analytic argument. We may assume that $\|\lambda\|_S \geqslant |S|^{3/2}$, as the claim is trivial otherwise. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a non-negative smooth even function (not depending on p or λ) supported on $[-1, 1]$ and non-zero on $[-1/2, 1/2]$, whose Fourier transform $\hat{\psi}(\xi) := \int_{\mathbb{R}} \psi(x)e(-\xi x)\,dx$ is also non-negative. Set $N := |S|^{-1}\|\lambda\|_S$, so in particular $N \geqslant 1$. We consider the kernel $K_N : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ defined by

$$K_N(n) := \sum_{k\in\mathbb{Z}} e_p(kn)\psi\left(\frac{k}{N}\right);$$

by the Poisson summation formula we have

$$K_N(n \pmod p) = N\sum_{m\in\mathbb{Z}} \hat{\psi}\left(\frac{Nn}{p} - Nm\right)$$

for any integer n, so in particular K_N is non-negative.

By definition of N, the frequency λ has no representations of the form $\lambda = \sum_{\xi\in S} a_\xi\xi$ with $\sup_{\xi\in S}|a_\xi| < N$. Hence the Riesz-type product $\prod_{\xi\in S} K_N(\xi n)$,

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

when expanded, contains no terms of the form $e_p(\lambda n)$ or $e_p(-\lambda n)$, and is therefore orthogonal to $\cos(2\pi\lambda n/p)$. In particular we have the identity

$$\mathbb{E}_{n\in\mathbb{Z}/p\mathbb{Z}}\prod_{\xi\in S}K_N(\xi n)=\mathbb{E}_{n\in\mathbb{Z}/p\mathbb{Z}}\left(1-\cos\left(\frac{2\pi\lambda n}{p}\right)\right)\prod_{\xi\in S}K_N(\xi n).$$

On the other hand, from two applications of (2.3) we have

$$\begin{aligned}1-\cos\left(\frac{2\pi\lambda n}{p}\right)&\ll\left\|\frac{\lambda n}{p}\right\|^2_{\mathbb{R}/\mathbb{Z}}\leqslant A^2\|n\|^2_{S^\perp}\\&\leqslant A^2\sum_{\xi_0\in S}\left\|\frac{\xi_0 n}{p}\right\|^2_{\mathbb{R}/\mathbb{Z}}\leqslant A^2\sum_{\xi_0\in S}\left(1-\cos\left(\frac{2\pi\xi_0 n}{p}\right)\right).\end{aligned}$$

As K_N is non-negative, we conclude that

$$\begin{aligned}&\mathbb{E}_{n\in\mathbb{Z}/p\mathbb{Z}}\prod_{\xi\in S}K_N(\xi n)\\&\ll A^2\sum_{\xi_0\in S}\mathbb{E}_{n\in\mathbb{Z}/p\mathbb{Z}}\left(\left(\prod_{\xi\in S\setminus\xi_0}K_N(\xi n)\right)K_N(\xi_0 n)\left(1-\cos\left(\frac{2\pi\xi_0 n}{p}\right)\right)\right).\end{aligned}\tag{4.6}$$

We can expand $K_N(\xi_0 n)(1-\cos(2\pi\xi_0 n/p))$ as a Fourier series

$$\sum_{k\in\mathbb{Z}}e_p(kn)\left(\psi\left(\frac{k}{N}\right)-\frac{\psi((k-1)/N)+\psi((k+1)/N)}{2}\right).$$

The expression inside parentheses is only non-vanishing for $|k|\leqslant N+1$, and has magnitude $O(1/N^2)$. As ψ is non-negative everywhere and non-zero on $[-1/2,1/2]$, we thus have a pointwise estimate of the form

$$\psi\left(\frac{k}{N}\right)-\frac{\psi((k-1)/N)+\psi((k+1)/N)}{2}\ll\frac{1}{N^2}\sum_{j=-8}^{8}\psi\left(\frac{k}{N}-\frac{j}{4}\right)$$

(for example). By using the non-negativity of the Fourier coefficients of K_N, this gives the estimate

$$\begin{aligned}&\mathbb{E}_{n\in\mathbb{Z}/p\mathbb{Z}}\left(\prod_{\xi\in S\setminus\xi_0}K_N(\xi n)\right)K_N(\xi_0 n)\left(1-\cos\left(\frac{2\pi\xi_0 n}{p}\right)\right)\\&\ll\frac{1}{N^2}\mathbb{E}_{n\in\mathbb{Z}/p\mathbb{Z}}\prod_{\xi\in S}K_N(\xi n).\end{aligned}$$

Comparing this with (4.6), we conclude that $1\ll A^2|S|/N^2$, and the claim follows from the definition of N. □

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Next, we estimate the Fourier coefficients of a regular distribution on a Bohr set in terms of the word norm.

LEMMA 4.7. *Let S be a non-degenerate subset of $\mathbb{Z}/p\mathbb{Z}$. Suppose that $\mathbf{n}$ is drawn regularly from $B(S, \rho)$. Then we have*

$$\mathbb{E}e_p(\lambda \mathbf{n}) \ll \frac{|S|^{5/2}}{\rho \|\lambda\|_S}$$

for all $\lambda \in \mathbb{Z}/p\mathbb{Z}$, where we adopt the convention that the above estimate is vacuously true if $\|\lambda\|_S = 0$.

Proof. For any $h \in \mathbb{Z}/p\mathbb{Z}$, one has from Lemma 4.4 that

$$\mathbb{E}e_p(\lambda \mathbf{n}) = \mathbb{E}e_p(\lambda(\mathbf{n}+h)) + O\left(\frac{|S|\|h\|_{S^\perp}}{\rho}\right)$$

which we may rearrange as

$$(1 - e_p(\lambda h))\mathbb{E}e_p(\lambda \mathbf{n}) \ll \frac{|S|\|h\|_{S^\perp}}{\rho}.$$

Since $|1 - e_p(\lambda h)| \gg \|\lambda h/p\|_{\mathbb{R}/\mathbb{Z}}$, we conclude that

$$\left\|\frac{\lambda h}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \mathbb{E}e_p(\lambda \mathbf{n}) \ll \frac{|S|\|h\|_{S^\perp}}{\rho}.$$

Taking h so as to minimize the ratio $\|h\|_{S^\perp}/\|\lambda h/p\|_{\mathbb{R}/\mathbb{Z}}$, the claim follows from Lemma 4.6. □

We will take advantage of the fact that Bohr sets can be approximately described as generalized arithmetic progressions. A key lemma in this regard is the following.

LEMMA 4.8. *Let Γ be a lattice in $\mathbb{R}^d$. Then there exist linearly independent generators $v_1, \dots, v_d$ of Γ and real numbers $N_1, \dots, N_d > 0$ such that*

$$B_{\mathbb{R}^d}(0, O(d)^{-3d/2}t) \cap \Gamma \subset \left\{\sum_{i=1}^{d} n_i v_i : |n_i| < tN_i\right\} \subset B_{\mathbb{R}^d}(0, t) \cap \Gamma \quad (4.7)$$

for all $t > 0$, where $B_{\mathbb{R}^d}(0, r)$ is the open Euclidean ball of radius r in $\mathbb{R}^d$, and the n_i are understood to be integers. Furthermore, the determinant/covolume $\det(\Gamma)$ *obeys the bounds*

$$\det(\Gamma) = (2d)^{O(d)} \prod_{i=1}^{d} N_i^{-1}. \quad (4.8)$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Proof. Applying [34, Theorem 1.6], we can find elements $v_1, \dots, v_r$ of Γ for some $r \leqslant d$, linearly independent over the rationals, and real numbers $N_1, \dots, N_d > 0$ such that

$$B_{\mathbb{R}^d}(0, O(d)^{-3d/2}t) \cap \Gamma \subset \left\{ \sum_{i=1}^{r} n_i v_i : |n_i| < tN_i \right\} \subset B_{\mathbb{R}^d}(0,t) \cap \Gamma \quad (4.9)$$

for all $t > 0$, and such that

$$O(d)^{-7d/2}|B_{\mathbb{R}^d}(0,t) \cap \Gamma| \leqslant \left| \left\{ \sum_{i=1}^{r} n_i v_i : |n_i| < tN_i \right\} \right| \leqslant |B_{\mathbb{R}^d}(0,t) \cap \Gamma|.$$

(Strictly speaking, the statement of [34, Theorem 1.6] only claims the latter bound for $t = 1$, but the same argument gives the bound for all $t > 0$.) Sending t to infinity, we conclude that the $v_1, \dots, v_r$ generate Γ; since, by virtue of being a lattice, Γ is cocompact, this forces $d = r$. Also, volume packing arguments show that as $t \to \infty$, the cardinality $|B_{\mathbb{R}^d}(0,t) \cap \Gamma|$ is asymptotic to the measure of $B_{\mathbb{R}^d}(0,t)$ divided by $\det(\Gamma)$, while the cardinality of $|\{n_1v_1 + \cdots + n_dv_d : |n_i| \leqslant tN_i\}|$ is asymptotic to $\prod_{i=1}^{d}(2tN_i)$. We conclude (4.8) as desired. □

The following corollary describes how we may pick a "basis" for a Bohr set.

COROLLARY 4.9. *Let S be a non-degenerate subset of $\mathbb{Z}/p\mathbb{Z}$, and set $d := |S|$. Then there exist elements $a_1, \dots, a_d$ of $\mathbb{Z}/p\mathbb{Z}$ and real numbers $N_1, \dots, N_d > 0$ such that*

$$\prod_{i=1}^{d} N_i^{-1} = (2d)^{O(d)}p \quad (4.10)$$

and

$$\|a_i\|_{S^\perp} \leqslant N_i^{-1} \quad (4.11)$$

for all $i = 1, \dots, d$. Furthermore, for any $a \in \mathbb{Z}/p\mathbb{Z}$, there exists a representation

$$a = n_1a_1 + \cdots + n_da_d \quad (4.12)$$

with $n_1, \dots, n_d$ integers of size

$$n_i = (2d)^{O(d)}N_i\|a\|_{S^\perp} \quad (4.13)$$

for $i = 1, \dots, d$. Finally, if one imposes the additional condition $|n_i| < N_i/2$ for all $i = 1, \dots, d$, then there is at most one such representation of this form (4.12) for a given a.

Proof. For each $s \in S$, the fraction s/p can be viewed as an element of $\mathbb{R}/\mathbb{Z}$ of order at most p; as S is non-degenerate, we see that the tuple $(s/p)_{s\in S}$ is an element of the torus $(\mathbb{R}/\mathbb{Z})^S$ of order p. Let Γ be the preimage in $\mathbb{R}^S$ of the

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

group generated by this element, thus Γ is a lattice of $\mathbb{R}^S$ that contains $\mathbb{Z}^S$ as a sublattice of index p; in particular, Γ has determinant p. Applying Lemma 4.8, one can find generators $v_1, \ldots, v_d$ of Γ and real numbers $N_1, \ldots, N_d$ obeying (4.10) such that

$$B_{\mathbb{R}^S}(0, O(d)^{-3d/2}t) \cap \Gamma \subset \left\{ \sum_{i=1}^{d} n_i v_i : |n_i| < tN_i \right\} \subset B_{\mathbb{R}^S}(0, t) \cap \Gamma \quad (4.14)$$

for all $t > 0$.

By construction of Γ, we can find elements $a_1, \ldots, a_d$ of $\mathbb{Z}/p\mathbb{Z}$ such that

$$v_i = \left(\frac{a_i s}{p} \right)_{s \in S} \pmod{\mathbb{Z}^S} \quad (4.15)$$

for $i = 1, \ldots, d$. Applying (4.14) with t slightly larger than N_i^{-1} for some $i = 1, \ldots, d$, we see that $v_i \in B_{\mathbb{R}^d}(N_i^{-1})$, and hence by (4.15) we have (4.11).

Finally, if $a \in \mathbb{Z}/p\mathbb{Z}$, then by definition of Γ we can find an element x of Γ in the preimage of $(as/p)_{s \in S}$ such that each component of x has magnitude less than $\|a\|_{S^\perp}$; in particular, $x \in B_{\mathbb{R}^S}(0, \sqrt{d}\|a\|_{S^\perp})$. Applying (4.14), we conclude that $x = \sum_{i=1}^{d} n_i v_i$ for some integers $n_1, \ldots, n_d$ obeying (4.13), giving the desired representation (4.12).

Finally, we show uniqueness. If there were two representations of the form (4.12) with $|n_i| < N_i/2$ for all $i = 1, \ldots, d$, then there exists a tuple $(n_1', \ldots, n_d') \in \mathbb{Z}^d$, not identically zero, with $|n_i'| < N_i$ for all $i = 1, \ldots, d$ and $\sum_{i=1}^{d} n_i a_i = 0$, which implies that the vector $\sum_{i=1}^{d} n_i v_i$ lies in $\mathbb{Z}^S$. As the $v_1, \ldots, v_d$ are linearly independent, this vector must have magnitude at least 1; but this contradicts (4.7) (with $t = 1$). □

Linear and quadratic functions on Bohr sets. We will frequently need to deal with locally linear or quadratic functions on Bohr sets. We review the definitions of these now.

Definition 4.10. Let B be a subset of $\mathbb{Z}/p\mathbb{Z}$, and let $G = (G, +)$ be an abelian group. A function $\phi : B \to G$ is said to be *locally linear* on B if one has

$$\phi(n + h_1 + h_2) - \phi(n + h_1) - \phi(n + h_2) + \phi(n) = 0$$

whenever $n, h_1, h_2 \in \mathbb{Z}/p\mathbb{Z}$ are such that $n, n + h_1, n + h_2, n + h_1 + h_2 \in B$. Similarly, ϕ is said to be *locally quadratic* on B if one has

$$\sum_{(\omega_1, \omega_2, \omega_3) \in \{0,1\}^3} (-1)^{\omega_1 + \omega_2 + \omega_3} \phi(n + \omega_1 h_1 + \omega_2 h_2 + \omega_3 h_3) = 0 \quad (4.16)$$

whenever $n, h_1, h_2, h_3 \in \mathbb{Z}/p\mathbb{Z}$ are such that $n + \omega_1 h_1 + \omega_2 h_2 + \omega_3 h_3 \in B$ for all $(\omega_1, \omega_2, \omega_3) \in \{0, 1\}^3$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

A function $\psi : B \times B \to G$ is said to be *locally bilinear* on B if one has

$$\psi(h_1 + h_1', h_2) = \psi(h_1, h_2) + \psi(h_1', h_2)$$

whenever $h_1, h_1', h_2 \in B$ are such that $h_1 + h_1' \in B$, and similarly one has

$$\psi(h_1, h_2 + h_2') = \psi(h_1, h_2) + \psi(h_1, h_2')$$

whenever $h_1, h_2, h_2' \in B$ are such that $h_2 + h_2' \in B$.

Specializing (4.16) to the case $h_1 = h_2 = h_3 = h$, we conclude that

$$\phi(n) - 3\phi(n+h) + 3\phi(n+2h) - \phi(n+3h) = 0 \tag{4.17}$$

whenever $\phi : B \to G$ is locally quadratic on B and $n, n+h, n+2h, n+3h \in B$.

It is well known (from the Weyl exponential sum estimates) that quadratic exponential sums such as $\mathbb{E}_{1 \leqslant n \leqslant N} e(\alpha n^2 + \beta n)$ can only be large when the quadratic phase αn^2 is of "major arc" type in the sense that $k\alpha n^2$ is close to constant on the range $\{1, \dots, N\}$ of the summation variable n, for some bounded positive integer k. The following proposition is an analogue of this phenomenon on Bohr sets.

PROPOSITION 4.11 (Large local quadratic exponential sums). *Let $B(S, \rho)$ be a Bohr set, let $0 < \delta \leqslant 1/2$, let $\lambda, \mu : B(S, 10\rho) \to \mathbb{R}/\mathbb{Z}$ be locally linear maps, and let $\phi : B(S, 10\rho) \times B(S, 10\rho) \to \mathbb{R}/\mathbb{Z}$ be a locally bilinear phase such that*

$$|\mathbb{E} e(\phi(\mathbf{n}, \mathbf{m}) + \lambda(\mathbf{n}) + \mu(\mathbf{m}))| \geqslant \delta \tag{4.18}$$

if $\mathbf{n}, \mathbf{m}$ are drawn independently and regularly from $B(S, \rho)$. Then there exists a natural number

$$1 \leqslant k \leqslant \delta^{-O(C_1|S|^2)}$$

such that

$$\|k\phi(n, m)\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(C_1|S|^2)} \frac{\|n\|_S \|m\|_S}{\rho^2} \tag{4.19}$$

whenever $n, m \in B(S, \delta^{C_1}\rho/(C_1|S|)^{3|S|})$.

Proof. Let $d := |S|$, thus $d \geqslant 1$. By Corollary 4.9, we can find elements $a_1, \dots, a_d$ of $\mathbb{Z}/p\mathbb{Z}$ and real numbers $N_1, \dots, N_d$ obeying the conclusions of that corollary.

Suppose that $1 \leqslant i, j \leqslant d$ are such that $N_i, N_j \geqslant d/\delta^{C_1/2}\rho$ (we allow i and j to be equal). Then by (4.11) we have

$$\|a_i\|_{S^\perp}, \|a_j\|_{S^\perp} \leqslant d^{-1}\delta^{C_1/2}\rho.$$

We can control the coefficient $\phi(a_i, a_j)$ by the following argument. If we draw $\mathbf{b}_i$ and $\mathbf{b}_j$ uniformly from $\{b_i \in \mathbb{Z} : 1 \leqslant b_i \leqslant \delta^{C_1/4}N_i\rho/d\}$ and $\{b_j \in \mathbb{Z} : 1 \leqslant b_j \leqslant \delta^{C_1/4}N_j\rho/d\}$ respectively and independently of each other and of $\mathbf{n}, \mathbf{m}$,

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

then from two applications of Lemma 4.4 (comparing $\mathbf{n}$ with $\mathbf{n}+\mathbf{b}_i a_i$, and $\mathbf{m}$ with $\mathbf{m}+\mathbf{b}_j a_j$) we have

$$\begin{aligned}&\mathbb{E}e(\phi(\mathbf{n}+\mathbf{b}_i a_i, \mathbf{m}+\mathbf{b}_j a_j)+\lambda(\mathbf{n}+\mathbf{b}_i a_i)+\mu(\mathbf{m}+\mathbf{b}_j a_j))\\ &\quad = \mathbb{E}e(\phi(\mathbf{n},\mathbf{m})+\lambda(\mathbf{n})+\mu(\mathbf{m}))+O(\delta^{C_1/4})\end{aligned}$$

and hence from (4.18) (assuming C_1 large enough) we have

$$|\mathbb{E}e(\phi(\mathbf{n}+\mathbf{b}_i a_i, \mathbf{m}+\mathbf{b}_j a_j)+\lambda(\mathbf{n}+\mathbf{b}_i a_i)+\mu(\mathbf{m}+\mathbf{b}_j a_j))| \gg \delta.$$

By the pigeonhole principle, we can therefore find $n, m \in B(S,\rho)$ such that

$$|\mathbb{E}e(\phi(n+\mathbf{b}_i a_i, m+\mathbf{b}_j a_j)+\lambda(n+\mathbf{b}_i a_i)+\mu(m+\mathbf{b}_j a_j))| \gg \delta.$$

Using the local bilinearity of ϕ, the left-hand side may be written as

$$|\mathbb{E}e(\mathbf{b}_i\mathbf{b}_j\phi(a_i,a_j)+\alpha\mathbf{b}_i+\beta\mathbf{b}_j+\gamma)|$$

for some $\alpha,\beta,\gamma \in \mathbb{R}/\mathbb{Z}$ depending on i, j, n, m whose exact values are not of importance to us. Evaluating the expectations and using the triangle inequality, we conclude that

$$\mathbb{E}_{1\leqslant b_i\leqslant \delta^{C_1/4}N_i\rho/d}|\mathbb{E}_{1\leqslant b_j\leqslant \delta^{C_1/4}N_j\rho/d}e(b_j(b_i\phi(a_i,a_j)+\beta))| \gg \delta$$

and hence (by Lemma 2.2)

$$|\mathbb{E}_{1\leqslant b_j\leqslant \delta^{C_1/4}N_j\rho/d}e(b_j(b_i\phi(a_i,a_j)+\beta))| \gg \delta$$

for $\gg \delta^{C_1/4+1}N_i\rho/d$ values of b_i in the range $1 \leqslant b_i \leqslant \delta^{C_1/4}N_i\rho/d$. This average is a geometric series that can be explicitly computed, leading to the bound

$$\|b_i\phi(a_i,a_j)+\beta\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{d}{\delta^{C_1/4+1}N_j\rho}$$

for $\gg \delta^{C_1/4+1}N_i\rho/d$ values of b_i in the range $1 \leqslant b_i \leqslant \delta^{C_1/4}N_i\rho/d$. Applying [17, Lemma A.4] (which is really an observation of Vinogradov, used often in the theory of Weyl sums), we conclude that

$$\|k_{i,j}\phi(a_i,a_j)\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{d^2}{\delta^{O(C_1)}N_iN_j\rho^2}$$

for some natural number $k_{i,j}$ with $1 \leqslant k_{i,j} \ll \delta^{-O(C_1)}$. If we then "clear denominators" by defining

$$k := \prod_{1\leqslant i,j\leqslant d:N_i,N_j\geqslant d/\delta^{C_1/2}\rho} k_{i,j},$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

then $1 \leqslant k \ll \delta^{-O(C_1 d^2)}$ and

$$\|k\phi(a_i, a_j)\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{1}{\delta^{O(C_1 d^2)} N_i N_j \rho^2} \tag{4.20}$$

for all $1 \leqslant i, j \leqslant d$ with $N_i, N_j \geqslant d/\delta^{C_1/2}\rho$.

For any $n \in \mathbb{Z}/p\mathbb{Z}$, we see from Corollary 4.9 that we can find integers $n_1, \ldots, n_d$ with

$$n_i \ll (2d)^{O(d)} N_i \|n\|_{S^\perp}$$

such that

$$n = n_1 a_1 + \cdots + n_d a_d.$$

In particular, if $n \in B(S, \delta^{C_1}\rho/(C_1 d)^{3d})$, then n_i is only non-zero when $N_i \geqslant d/\delta^{C_1/2}\rho$. From these bounds, (4.20), and the local bilinearity of ϕ, we conclude (4.19) as desired. □

Local U^2-inverse theorem. The global inverse U^2 theorem, which is a simple and well-known exercise in discrete Fourier analysis, asserts that if a 1-bounded function $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ obeys the bound

$$|\mathbb{E} f(\mathbf{h}_0 + \mathbf{h}_1)\overline{f}(\mathbf{h}_0 + \mathbf{h}_1')\overline{f}(\mathbf{h}_0' + \mathbf{h}_1) f(\mathbf{h}_0' + \mathbf{h}_1')| \geqslant \eta \tag{4.21}$$

where $\mathbf{h}_0, \mathbf{h}_1, \mathbf{h}_0', \mathbf{h}_1'$ are drawn uniformly at random from $\mathbb{Z}/p\mathbb{Z}$, then there exists $\xi \in \mathbb{Z}/p\mathbb{Z}$ such that

$$|\mathbb{E} f(\mathbf{h}) e_p(-\xi\mathbf{h})| \geqslant \eta^{1/2} \tag{4.22}$$

where $\mathbf{h}$ is also drawn uniformly at random from $\mathbb{Z}/p\mathbb{Z}$.

In this section we give a local version of the above claim, in which the random variables $\mathbf{h}, \mathbf{h}_0, \mathbf{h}_1, \mathbf{h}_0', \mathbf{h}_1'$ are localized to a small Bohr set. If the rank of the Bohr set is bounded, one can modify the above arguments to obtain a reasonable inverse theorem of this nature, but in our application the rank of the Bohr set will be rather large, and it will be important that this rank does not affect the lower bound in correlations of the form (4.22). Fortunately, such a result is available, and will be crucial in the proofs of the two remaining claims (Corollary 4.13 and Theorem 8.1) needed to prove Theorem 1.1.

Here is a precise version of the claim.

THEOREM 4.12. *Let $S \subset \mathbb{Z}/p\mathbb{Z}$ be non-degenerate for some prime p, and let $0 < \eta < 1/2$. Let ρ_0, ρ_1 be real parameters with $0 < \rho_1 < \rho_0 < 1/2$ and such that*

$$\rho_0 > \frac{C|S|}{\eta^2}\rho_1 \tag{4.23}$$

for a sufficiently large absolute constant C. Let $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ be a 1-bounded function such that

$$|\mathbb{E} f(\mathbf{h}_0 + \mathbf{h}_1)\overline{f}(\mathbf{h}_0 + \mathbf{h}_1')\overline{f}(\mathbf{h}_0' + \mathbf{h}_1) f(\mathbf{h}_0' + \mathbf{h}_1')| \geqslant \eta, \tag{4.24}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

where $\mathbf{h}_0, \mathbf{h}_0', \mathbf{h}_1, \mathbf{h}_1'$ *are drawn independently and regularly from* $B(S, \rho_0)$, $B(S, \rho_0)$, $B(S, \rho_1)$, $B(S, \rho_1)$ *respectively. Then there exists* $\xi \in \mathbb{Z}/p\mathbb{Z}$ *such that*

$$\sum_{n_0 \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n}_0 = n_0)|\mathbb{E} f(n_0 + \mathbf{n}_1) e_p(-\xi \mathbf{n}_1)|^2 \geqslant \eta/2$$

where $\mathbf{n}_0, \mathbf{n}_1$ *are drawn independently and regularly from* $B(S, \rho_0)$, $B(S, \rho_1)$ *respectively.*

Proof. We thank Fernando Shao for supplying a proof of this result, which was considerably simpler than our original argument.

For this proof, which is Fourier-analytic in nature, it will be convenient to work explicitly with probability densities rather than probabilistic notation. (However, in the lengthier proof of the local inverse U^3 theorem given in the next section, the probabilistic notation will be significantly cleaner to use.) In this argument, all sums will be over $\mathbb{Z}/p\mathbb{Z}$. We abbreviate

$$\mathfrak{p}_i(h) := \mathfrak{p}_{B(S,\rho_i)}(h) = \mathbb{P}(\mathbf{h}_i = h)$$

for $i = 0, 1$ and $h \in \mathbb{Z}/p\mathbb{Z}$; clearly we have $\mathfrak{p}_i(h) \geqslant 0$ and

$$\sum_h \mathfrak{p}_i(h) = 1. \tag{4.25}$$

The hypothesis (4.24) may be written as

$$\Big| \sum_{h_0, h_0', h_1, h_1'} \mathfrak{p}_0(h_0)\mathfrak{p}_0(h_0')\mathfrak{p}_1(h_1)\mathfrak{p}_1(h_1') f(h_0 + h_1)\overline{f}(h_0 + h_1') \times \overline{f}(h_0' + h_1) f(h_0' + h_1') \Big| \geqslant \eta \tag{4.26}$$

and our goal is to locate $\xi \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_0} \mathfrak{p}_0(n_0) \Big| \sum_{n_1} \mathfrak{p}_1(n_1) f(n_0 + n_1) e_p(-\xi n_1) \Big|^2 \geqslant \eta/2.$$

The first step is to replace the factor $\mathfrak{p}_0(h_0)$ by the slightly different factor $\mathfrak{p}_0^{1/2}(h_0 + h_1)\mathfrak{p}_0^{1/2}(h_0 + h_1')$. If we use the elementary inequality $|x^{1/2} - y^{1/2}| \leqslant |x - y|^{1/2}$ for $x, y \geqslant 0$ and then apply Cauchy–Schwarz, Lemma 4.4, and (4.23), we see that

$$\begin{aligned} &\sum_{h_0} |\mathfrak{p}_0^{1/2}(h_0 + h_1) - \mathfrak{p}_0^{1/2}(h_0)|\mathfrak{p}_0^{1/2}(h_0) \\ &\quad \leqslant \sum_{h_0} |\mathfrak{p}_0(h_0 + h_1) - \mathfrak{p}_0(h_0)|^{1/2}\mathfrak{p}_0^{1/2}(h_0) \end{aligned}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

$$\leqslant \Big(\sum_{h_0 \in \mathbb{Z}/p\mathbb{Z}} |\mathfrak{p}_0(h_0+h_1) - \mathfrak{p}_0(h_0)| \Big)^{1/2}$$
$$= \Big(\sum_{h_0 \in \mathbb{Z}/p\mathbb{Z}} b_{h_1}(h_0)\mathfrak{p}_0(h_0+h_1) - b_{h_1}(h_0)\mathfrak{p}_0(h_0) \Big)^{1/2}$$
$$\ll \left(\frac{|S|\rho_1}{\rho_0} \right)^{1/2} \ll \frac{\eta}{C^{1/2}}$$

for any h_1 in the support of $\mathfrak{p}_1$, where the 1-bounded function b_{h_1} is given by $b_{h_1}(h_0) := \mathrm{sgn}(\mathfrak{p}_0(h_0+h_1) - \mathfrak{p}_0(h_0))$. Similarly we have

$$\sum_{h_0} |\mathfrak{p}_0^{1/2}(h_0+h_1') - \mathfrak{p}_0^{1/2}(h_0)|\mathfrak{p}_0^{1/2}(h_0+h_1) \ll \frac{\eta}{C^{1/2}}$$

whenever h_1' is also in the support of $\mathfrak{p}_1$; by the triangle inequality, we conclude that

$$\sum_{h_0} |\mathfrak{p}_0^{1/2}(h_0+h_1)\mathfrak{p}_0(h_0+h_1')^{1/2} - \mathfrak{p}_0(h_0)| \ll \frac{\eta}{C^{1/2}}$$

for all h_1, h_1' in the support of $\mathfrak{p}_1$. From the 1-boundedness of f and (4.25), we conclude that

$$\Big| \sum_{h_0,h_0',h_1,h_1'} |\mathfrak{p}_0^{1/2}(h_0+h_1)\mathfrak{p}_0^{1/2}(h_0+h_1') - \mathfrak{p}_0(h_0)|$$
$$\times \mathfrak{p}_0(h_0')\mathfrak{p}_1(h_1)\mathfrak{p}_1(h_1') f(h_0+h_1)\overline{f}(h_0+h_1')\overline{f}(h_0'+h_1) f(h_0'+h_1') \Big|$$
$$\ll \frac{\eta}{C^{1/2}}.$$

If C is large enough, the left-hand side is thus bounded by 0.1η (for example), so by (4.26) and the triangle inequality we conclude that

$$\Big| \sum_{h_0,h_0',h_1,h_1'} \mathfrak{p}_0^{1/2}(h_0+h_1)\mathfrak{p}_0^{1/2}(h_0+h_1')\mathfrak{p}_0(h_0')\mathfrak{p}_1(h_1)\mathfrak{p}_1(h_1')$$
$$\times f(h_0+h_1)\overline{f}(h_0+h_1')\overline{f}(h_0'+h_1) f(h_0'+h_1') \Big| \geqslant 0.9\eta$$

If we write

$$f_0(n) := f(n)\mathfrak{p}_0^{1/2}(n), \tag{4.27}$$

we may rewrite the above estimate as

$$\Big| \sum_{h_0,h_0',h_1,h_1'} \mathfrak{p}_0(h_0')\mathfrak{p}_1(h_1)\mathfrak{p}_1(h_1')$$
$$\times f_0(h_0+h_1)\overline{f_0}(h_0+h_1')\overline{f}(h_0'+h_1) f(h_0'+h_1') \Big| \geqslant 0.9\eta.$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

A similar argument then lets us replace $\mathfrak{p}_0(h_0')$ with $\mathfrak{p}_0^{1/2}(h_0'+h_1)\mathfrak{p}_0^{1/2}(h_0'+h_1')$, leaving us with

$$\Big|\sum_{h_0,h_0',h_1,h_1'} \mathfrak{p}_0(h_0'+h_1)^{1/2}\mathfrak{p}_0(h_0'+h_1')^{1/2}\mathfrak{p}_1(h_1)\mathfrak{p}_1(h_1') \times f_0(h_0+h_1)\overline{f_0}(h_0+h_1')\overline{f}(h_0'+h_1)f(h_0'+h_1')\Big| \geqslant 0.8\eta.$$

which we can simplify using (4.27) to

$$\Big|\sum_{h_0,h_0',h_1,h_1'} \mathfrak{p}_1(h_1)\mathfrak{p}_1(h_1')f_0(h_0+h_1)\overline{f_0}(h_0+h_1')\overline{f_0}(h_0'+h_1)f_0(h_0'+h_1')\Big| \geqslant 0.8\eta.$$

Making the change of variables $n := h_1 - h_1'$, we may rewrite the left-hand side as

$$\sum_n (\mathfrak{p}_1 * \tilde{\mathfrak{p}}_1)(n)|(f_0 * \tilde{f}_0)(n)|^2$$

where $\tilde{f}_0(n) := \overline{f_0}(-n)$, and similarly for p_1, and $f * g$ denotes the discrete convolution

$$f * g(n) := \sum_m f(m)g(n-m).$$

Using the Fourier transform, we may then rewrite the previous bound as

$$p^4 \sum_{\xi,\xi'} |\hat{\mathfrak{p}}_1(\xi')|^2|\hat{f}_0(\xi)|^2|\hat{f}_0(\xi+\xi')|^2 \geqslant 0.8\eta \tag{4.28}$$

where

$$\hat{f}(\xi) := \frac{1}{p}\sum_n f(n)e_p(-\xi n).$$

From (4.25), the 1-boundedness of f, and the Plancherel identity we have

$$\sum_\xi |\hat{f}_0(\xi)|^2 = \frac{1}{p}\sum_n |f_0(n)|^2 \leqslant \frac{1}{p}.$$

By this, (4.28), and the pigeonhole principle, we may therefore find $\xi \in \mathbb{Z}/p\mathbb{Z}$ such that

$$p^3 \sum_{\xi' \in \mathbb{Z}/p\mathbb{Z}} |\hat{p}_1(\xi')|^2|\hat{f}_0(\xi+\xi')|^2 \geqslant 0.8\eta.$$

By the Plancherel identity again, the left-hand side may be rewritten as

$$\sum_{n_0}\Big|\sum_{n_1} f_0(n_0-n_1)\mathfrak{p}_1(n_1)e_p(\xi n_1)\Big|^2$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

and hence (by replacing n_1 with $-n_1$ and using (4.27))

$$\sum_{n_0}\left|\sum_{n_1} f(n_0+n_1)\mathfrak{p}_0^{1/2}(n_0+n_1)\mathfrak{p}_1(n_1)e_p(-\xi n_1)\right|^2 \geqslant 0.8\eta.$$

By argument similar to those at the beginning of the proof, we may replace $\mathfrak{p}_0^{1/2}(n_0+n_1)$ by $\mathfrak{p}_0^{1/2}(n_0)$ and conclude that

$$\sum_{n_0}\left|\sum_{n_1} f(n_0+n_1)\mathfrak{p}_0^{1/2}(n_0)\mathfrak{p}_1(n_1)e(-\xi n_1)\right|^2 \geqslant 0.7\eta,$$

and the claim follows. □

As a corollary of this inverse theorem, we can establish that locally almost linear phases on Bohr sets can be approximated by globally linear phases; this will be needed in §7 to deal with poorly distributed quadratic factors.

Here is a precise statement.

COROLLARY 4.13. *Let $\phi : n_0 + B(S,\rho) \to \mathbb{R}/\mathbb{Z}$ be a function on a shifted Bohr set $n_0 + B(S,\rho)$ that is "locally almost linear" in the sense that one has the bound*

$$\|\phi(n_0+h+k)-\phi(n_0+h)-\phi(n_0+k)+\phi(n_0)\|_{\mathbb{R}/\mathbb{Z}} \leqslant A\frac{\|h\|_{S^\perp}\|k\|_{S^\perp}}{\rho^2} \tag{4.29}$$

for all $h,k \in B(S,\rho/2)$ and some $A \geqslant 1$. Then there exists $\xi \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\left\|\phi(n_0+h)-\phi(n_0)-\frac{\xi h}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \ll A^{1/2}|S|^4\frac{\|h\|_{S^\perp}}{\rho} \tag{4.30}$$

for all $h \in B(S,\rho)$.

Proof. By translating in space, we may normalize so that $n_0 = 0$; by shifting ϕ by a phase, we may also suppose that $\phi(0) = 0$. By replacing ρ with the smaller quantity $\rho/A^{1/2}$ if necessary, we may normalize A to be 1 (note that (4.30) is trivial for $\|h\|_{S^\perp} \geqslant \rho/A^{1/2}$). Thus, we now have a function $\phi : B(S,\rho) \to \mathbb{R}/\mathbb{Z}$ with $\phi(0) = 0$ such that the quantity

$$\partial^2\phi(h,k) := \phi(h+k)-\phi(h)-\phi(k) \tag{4.31}$$

obeys the bound

$$\|\partial^2\phi(h,k)\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{\|h\|_{S^\perp}\|k\|_{S^\perp}}{\rho^2} \tag{4.32}$$

for all $h,k \in B(S,\rho/2)$, and our task is to locate $\xi \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\left\|\phi(h)-\frac{\xi h}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \ll |S|^4\frac{\|h\|_{S^\perp}}{\rho} \tag{4.33}$$

for all $h \in B(S,\rho)$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Let $\rho_0 := \rho/100$, and set $\rho_1 := \rho/C|S|^3$ for some sufficiently large absolute constant C. If we let $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ be the 1-bounded function

$$f(x) := 1_{B(S,\rho)} e(\phi(x)) \tag{4.34}$$

and draw $\mathbf{h}_0, \mathbf{h}_0', \mathbf{h}_1, \mathbf{h}_1'$ independently and regularly from $B(S, \rho_0)$, $B(S, \rho_0)$, $B(S, \rho_1)$, $B(S, \rho_1)$ respectively, then from (4.31) we have

$$\begin{aligned} &f(\mathbf{h}_0 + \mathbf{h}_1)\overline{f}(\mathbf{h}_0 + \mathbf{h}_1')\overline{f}(\mathbf{h}_0' + \mathbf{h}_1) f(\mathbf{h}_0' + \mathbf{h}_1') \\ &\quad = e(\partial^2\phi(\mathbf{h}_0, \mathbf{h}_1) - \partial^2\phi(\mathbf{h}_0', \mathbf{h}_1) - \partial^2\phi(\mathbf{h}_0, \mathbf{h}_1') + \partial^2\phi(\mathbf{h}_0', \mathbf{h}_1')). \end{aligned}$$

Applying (4.32) and taking expectations, we conclude that

$$|\mathbb{E} f(\mathbf{h}_0 + \mathbf{h}_1)\overline{f}(\mathbf{h}_0 + \mathbf{h}_1')\overline{f}(\mathbf{h}_0' + \mathbf{h}_1) f(\mathbf{h}_0' + \mathbf{h}_1')| \geqslant 1/2$$

(for example). Applying Theorem 4.12 (which is applicable for C large enough), we may thus find $\xi \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_0 \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n}_0 = n_0)|\mathbb{E} f(n_0 + \mathbf{n}_1) e_p(-\xi \mathbf{n}_1)|^2 \geqslant 1/4$$

if $\mathbf{n}_0, \mathbf{n}_1$ are drawn independently and regularly from $B(S, \rho_0)$, $B(S, \rho_1)$ respectively. In particular, there exists $n \in B(S, \rho_0)$ such that

$$|\mathbb{E} f(n + \mathbf{n}_1) e_p(-\xi \mathbf{n}_1)| \geqslant 1/4.$$

By (4.34), (4.31) we have

$$f(n + \mathbf{n}_1) = e(\phi(\mathbf{n}_1) + \phi(n) + \partial^2\phi(n, \mathbf{n}_1))$$

so by (4.32) we conclude that

$$\left|\mathbb{E} e\left(\phi(\mathbf{n}_1) - \frac{\xi \mathbf{n}_1}{p}\right)\right| \gg 1. \tag{4.35}$$

For any $h \in B(S, \rho_1)$, we have from Lemma 4.4 that

$$\left|\mathbb{E} e\left(\phi(\mathbf{n}_1 + h) - \frac{\xi(\mathbf{n}_1 + h)}{p}\right) - \mathbb{E} e\left(\phi(\mathbf{n}_1) - \frac{\xi \mathbf{n}_1}{p}\right)\right| \ll |S| \frac{\|h\|_{S^\perp}}{\rho_1};$$

on the other hand, from (4.31) we have the identity

$$\begin{aligned} &\mathbb{E} e\left(\phi(\mathbf{n}_1 + h) - \frac{\xi(\mathbf{n}_1 + h)}{p}\right) \\ &\quad = e\left(\phi(h) - \frac{\xi h}{p}\right)\mathbb{E} e\left(\phi(\mathbf{n}_1) - \frac{\xi \mathbf{n}_1}{p} + \partial^2\phi(\mathbf{n}_1, h)\right). \end{aligned}$$

Combining this with (4.32), (4.35), and (2.2), we conclude that

$$\left\|\phi(h) - \frac{\xi h}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \asymp \left|e\left(\phi(h) - \frac{\xi h}{p}\right) - 1\right| \ll |S| \frac{\|h\|_{S^\perp}}{\rho_1}$$

for all $h \in B(S, \rho_1)$. As the claim (4.33) is trivial for $h \in B(S, \rho)\backslash B(S, \rho_1)$, the claim follows. □

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

§5. *Dilated tori.* As mentioned in Example 3 of §3, to maintain good quantitative control (and specifically, Lipschitz norm control) on the functions $F : G \to [-1, 1]$ used to build quadratic approximants, one needs to generalize the underlying domain G to more general tori than the standard tori $(\mathbb{R}/\mathbb{Z})^d$ with the usual norm structure. It turns out that it will suffice to work with *dilated tori* of the form

$$G = \prod_{i=1}^{d} (\mathbb{R}/\lambda_i \mathbb{Z}),$$

where $\lambda_1, \ldots, \lambda_d \geqslant 1$ are real numbers. One can view this dilated torus as the quotient of $\mathbb{R}^d$ by a dilated lattice $\Gamma := \prod_{i=1}^{d} \lambda_i \mathbb{Z}$. We can place a "norm" on G by declaring $\|x\|_G$ for $x \in G$ to be the Euclidean distance in $\mathbb{R}^d$ from x to Γ; this generalizes the norm $\|\|_{\mathbb{R}/\mathbb{Z}}$ from §2. This in turn defines a metric d_G on G by the formula

$$d_G(x, y) := \|x - y\|_G.$$

The *volume* $\mathrm{vol}(G)$ of a dilated torus is defined to be the product

$$\mathrm{vol}(G) := \prod_{i=1}^{d} \lambda_i = \det(\Gamma).$$

It will be important to keep this quantity under control during the iteration process. In particular, when transforming from one dilated torus to another, the volume of the new torus should behave like a linear function of the existing torus; anything worse than this (e.g. quadratic behaviour) will lead to undesirable bounds upon iteration.

We define the *Pontryagin dual* $\hat{G}$ of a dilated torus G to be the lattice

$$\hat{G} := \prod_{i=1}^{d} \frac{1}{\lambda_i} \mathbb{Z}.$$

Elements k of this dual will be called *dual frequencies* of the torus. If $k = (k_1, \ldots, k_d)$ is a dual frequency and $x = (x_1, \ldots, x_d)$ is an element of G, we define the dot product $k \cdot x \in \mathbb{R}/\mathbb{Z}$ in the usual fashion as

$$k \cdot x = k_1 x_1 + \cdots + k_d x_d$$

noting that this gives a well-defined element of $\mathbb{R}/\mathbb{Z}$.

A dual frequency k is said to be *irreducible* if it is non-zero, and not of the form $k = nk'$ for some other dual frequency k' and some natural number $n > 1$. If a dual frequency k is irreducible, then its orthogonal complement

$$k^{\perp} := \{x \in G : k \cdot x = 0\}$$

is a $(d-1)$-dimensional subtorus of G; it inherits a metric $d_{k^{\perp}}$ from the torus G it lies in. We will need to pass to such a complement when dealing with poorly distributed quadratic factors (as in the third or fourth examples in §3), however

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

we encounter the technical issue that these complements $k^\perp$ will not quite be of the form of a dilated torus. However, we will be able to transform $k^\perp$ into a dilated torus using a bilipschitz transformation, as the following result shows.

THEOREM 5.1. *Let $G = \prod_{i=1}^d (\mathbb{R}/\lambda_i\mathbb{Z})$ be a dilated torus, and let $k \in \hat{G}$ be an irreducible dual frequency of G. Then there exists a dilated torus $G' = \prod_{i=1}^{d-1}(\mathbb{R}/\lambda_i'\mathbb{Z})$ and a Lie group isomorphism $\psi : k^\perp \to G'$ obeying the bilipschitz bounds*

$$\|\psi\|_{\mathrm{Lip}}, \|\psi^{-1}\|_{\mathrm{Lip}} \ll d^{O(d)} \tag{5.1}$$

and such that one has the volume bound

$$\mathrm{vol}(G') = d^{O(d)}|k|\,\mathrm{vol}(G), \tag{5.2}$$

where $|k|$ denotes the Euclidean magnitude of k in $\mathbb{R}^d$.

Proof. The case $d = 0$ is vacuous and the case $d = 1$ is trivial, so we may assume $d > 1$. One can identify $k^\perp$ with the quotient V/Γ, where $V := \{x \in \mathbb{R}^d : k\cdot x = 0\}$ is the hyperplane in $\mathbb{R}^d$ orthogonal to k (now viewed as an element of $\mathbb{R}^d$), and $\Gamma := V \cap \prod_{i=1}^d (\lambda_i\mathbb{Z})$ is the restriction of the lattice $\prod_{i=1}^d(\lambda_i\mathbb{Z})$ to V.

As k is irreducible, there exists a vector e in the lattice $\prod_{i=1}^d(\lambda_i\mathbb{Z})$ with $k\cdot e = 1$; thus e has distance $1/|k|$ to V. One can form a fundamental domain of $\mathbb{R}^d/\prod_{i=1}^d(\lambda_i\mathbb{Z})$ by taking any fundamental domain for V/Γ and performing the Minkowski sum of that domain with the interval $\{te : 0 \leqslant t \leqslant 1\}$. By Fubini's theorem, the d-dimensional Lebesgue measure of such a sum will equal the $(d-1)$-dimensional Lebesgue measure of the fundamental domain of V/Γ and $1/|k|$; thus the covolume of $\prod_{i=1}^d(\lambda_i\mathbb{Z})$ in $\mathbb{R}^d$ equals $1/|k|$ times the covolume of Γ in V. As the former covolume (determinant) is $\prod_{i=1}^d \lambda_i = \mathrm{vol}(G)$, we conclude that Γ has covolume $|k|\,\mathrm{vol}(G)$ in V.

Applying Lemma 4.8, we can find linearly independent elements $v_1, \ldots, v_{d-1}$ generating Γ such that

$$B_V(0, O(d)^{-3d/2}t) \cap \Gamma \subset \left\{\sum_{i=1}^r n_i v_i : |n_i| \leqslant tN_i\right\} \subset B_V(0,t) \cap \Gamma \tag{5.3}$$

for all $t > 0$, where $B_V(0,r)$ is the Euclidean ball of radius r in V, and the n_i are understood to be integers, with the bound

$$\prod_{i=1}^{d-1} N_i^{-1} = (2d)^{O(d)}|k|\,\mathrm{vol}(G). \tag{5.4}$$

From (5.3) we conclude in particular that

$$O(d)^{-3d/2}N_i^{-1} \leqslant |v_i| \leqslant N_i^{-1} \tag{5.5}$$

for all $1 \leqslant i \leqslant d$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We now define the $(d-1)$-dimensional dilated torus

$$G' := \prod_{i=1}^{d-1} (\mathbb{R}/N_i^{-1}\mathbb{Z})$$

and the isomorphism $\phi : V/\Gamma \to G'$ by the formula

$$\phi\left(\sum_{i=1}^{d-1} t_i v_i \pmod{\Gamma}\right) := (t_1 N_1^{-1}, \dots, t_{d-1} N_{d-1}^{-1}) \quad \left(\mathrm{mod} \prod_{i=1}^{d-1} N_i^{-1}\mathbb{Z}\right)$$

for real numbers $t_1, \dots, t_{d-1}$. It is easy to see that this is a Lie group isomorphism, and the bound (5.2) follows from (5.4). It remains to establish the bilipschitz bounds (5.1). It suffices to show that the linear isomorphism

$$\sum_{i=1}^{d-1} t_i v_i \mapsto (t_1 N_1^{-1}, \dots, t_{d-1} N_{d-1}^{-1})$$

from V to $\mathbb{R}^{d-1}$, together with its inverse, have an operator norm of $O(d^{O(d)})$. For the inverse map, this is clear from (5.5). For the forward map, it suffices from Cramer's rule to show that

$$\frac{|v_1 \wedge \cdots \wedge v_{i-1} \wedge x \wedge v_{i+1} \wedge \cdots \wedge v_{d-1}|}{|v_1 \wedge \cdots \wedge v_{d-1}|} \ll \frac{d^{O(d)}}{\lambda_i'}$$

for all $i = 1, \dots, d-1$ and all unit vectors x in V. But from (5.5) the numerator is at most $\prod_{1 \leqslant i' \leqslant d-1 : i' \neq i} N_{i'}^{-1}$, while the denominator is the volume of a fundamental domain in V and is thus equal to $d^{O(d)} N_1^{-1} \dots N_{d-1}^{-1}$ thanks to (5.4). The claim follows. □

§6. *Constructing the approximants.* In this section we construct the abstract directed graph $G = (V, E)$ that appears in Proposition 3.3. For the rest of the paper, the prime p, the function $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$, and the parameter η with $0 < \eta \leqslant \frac{1}{10}$ are fixed, and we assume that (3.21) holds.

We begin with a description of the structured approximants $v \in V$.

Definition 6.1 (Structured local approximant). A *structured local approximant* is a tuple

$$v = (C, \mathbf{c}, (n_c + B(S_c, \rho_c))_{c \in C}, (G_c)_{c \in C}, (F_c)_{c \in C}, (\Xi_c)_{c \in C})$$

consisting of the following objects:

- a finite non-empty set C;
- a random variable $\mathbf{c}$, which we call the *label variable*, taking values in C;
- a shifted Bohr set $n_c + B(S_c, \rho_c)$ associated to each label $c \in C$;
- a dilated torus G_c associated to each label $c \in C$;

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

- a 1-Lipschitz function $F_c : G_c \to [-1, 1]$ associated to each label $c \in \mathbb{C}$; and
- a locally quadratic function $\Xi_c : n_c + B(S_c, \rho_c) \to G_c$ associated to each label $c \in C$.

We denote the collection of all structured local approximants (up to isomorphism[3]) as V. Given any structured local approximant $v \in V$, we define the random variables $(\mathbf{a}_v, \mathbf{r}_v, \mathbf{f}_v)$ associated to v by the following construction.

(1) First, let $\mathbf{c}$ be the random label variable appearing above.
(2) For each $c \in C$ in the essential range of $\mathbf{c}$, if we condition on the event $\mathbf{c} = c$, we draw $\mathbf{a}_v, \mathbf{r}_v$ independently and regularly from $n_c + B(S_c, \rho_c/2)$ and $B(S_c, \exp(-\eta^{-C_4})\rho_c)$ respectively, and then we let $\mathbf{f}_v$ be the function

$$\mathbf{f}_v(a) := F_c(\Xi_c(a)).$$

Thus $\mathbf{f}_v$ is deterministic when $\mathbf{c}$ is conditioned to be fixed, but random when $\mathbf{c}$ is allowed to vary.

We also define the following additional statistics of the structured local approximant v:

- the *waste* $\mathrm{waste}(v)$ is the quantity $|\mathbb{E} f(\mathbf{a}) - \mathbb{E}_{a \in \mathbb{Z}/p\mathbb{Z}} f(a)|$;
- the 1-*error* $\mathrm{Err}_1(v)$ is $|\mathbb{E}\mathbf{f}(\mathbf{a}) - \mathbb{E} f(\mathbf{a})|$;
- the 4-*error* $\mathrm{Err}_4(v)$ is $|\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f}) - \Lambda_{\mathbf{a},\mathbf{r}}(f)|$;
- the *energy* $\mathrm{Energy}(v)$ is $\mathbb{E}|f(\mathbf{a}) - \mathbf{f}(\mathbf{a})|^2$;
- the *linear rank* $d_1(v)$ is $\max_{c \in C} |S_c|$;
- the *quadratic dimension* $d_2(v)$ is $\max_{c \in C} \dim(G_c)$;
- the *linear scale* $\rho(v)$ is $\min_{c \in C} \rho_c$;
- the *quadratic volume* $\mathrm{vol}(v)$ is the quantity $\max_{c \in C} \mathrm{vol}(G_c)$;
- the *poorly distributed quadratic dimension* $d_2^{\mathrm{poor}}(v)$ is the maximum value of $\dim(G_c)$ over all poorly distributed c in the essential range of $\mathbf{c}$, or zero if no such c exists. Here, an element c in the essential range of $\mathbf{c}$ is said to be *poorly distributed* if one has

$$\Lambda_{\mathbf{a},\mathbf{r}}(f|\mathbf{c} = c) < \mathbb{E}(\mathbf{f}(\mathbf{a})|\mathbf{c} = c)^4 - \frac{\eta}{2}. \tag{6.1}$$

This gives the set V of structured local approximants for Proposition 3.3; we clearly have $0 \leqslant d_2^{\mathrm{poor}}(v) \leqslant d_2(v)$ for all $v \in V$.

We now also define the initial approximant.

Definition 6.2. The initial approximant $v_0 \in V$ is defined to be the tuple

$$v_0 = (C, \mathbf{c}, (n_c + B(S_c, \rho_c))_{c \in C}, (G_c)_{c \in C}, (F_c)_{c \in C}, (\Xi_c)_{c \in C})$$

defined as follows:

- $C := \mathbb{Z}/p\mathbb{Z}$, and $\mathbf{c}$ is drawn uniformly from C;

[3] This caveat is needed for the technical reason that V should be a set and not a proper class.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

- for each $c \in C$, we have $n_c := 0$, $S_c := \{1\}$, and $\rho_c := 1$;
- for each $c \in C$, the group G_c is the standard 0-torus $(\mathbb{R}/\mathbb{Z})^0$ (that is to say, a point);
- for each $c \in C$, the function $F_c : G_c \to [-1, 1]$ is the zero function $F_c(x) := 0$;
- for each $c \in C$, the function $\Xi_c : \mathbb{Z}/p\mathbb{Z} \to G_c$ is the unique (constant) map from $\mathbb{Z}/p\mathbb{Z}$ to the point G_c.

By chasing the definitions, we see that $\mathbf{a}_{v_0}$ is uniformly distributed in $\mathbb{Z}/p\mathbb{Z}$, and we can compute several of the statistics of the initial approximant v_0:

$$\text{waste}(v_0) = d_2^{\text{poor}}(v_0) = d_2(v_0) = 0; d_1(v_0) = \rho(v) = \text{vol}(v) = 1. \quad (6.2)$$

Now we define the edges of the graph $G(V, E)$.

Definition 6.3. We let E be the set of all directed edges $v \to v'$, where v, $v' \in V$ are structured local approximants such that

$$\begin{aligned} d_1(v') &\leqslant d_1(v) + \eta^{-C_2}, \\ d_2(v') &\leqslant d_2(v) + 1, \\ \rho(v') &\geqslant \exp(-\eta^{-C_5})\rho(v), \\ \text{vol}(v') &\leqslant \exp(\eta^{-C_3})\,\text{vol}(v), \\ |\text{waste}(v) - \text{waste}(v')| &\leqslant \eta^{C_3}. \end{aligned}$$

From this definition and (6.2) we have the following bounds on the various statistics of vertices of V that are not too far from the initial vertex v_0, assuming that each constant C_i is chosen sufficiently large depending on the preceding constants $C_1, \dots, C_{i-1}$.

LEMMA 6.4. *Suppose a vertex $v = v_k \in V$ can be reached from v_0 by a path $v_0 \to v_1 \to \cdots \to v_k$ with $0 \leqslant k \leqslant 8\eta^{-2C_2}$. Then we have*

$$d_1(v) \leqslant 8\eta^{-3C_2}, \quad (6.3)$$
$$d_2(v) \leqslant 8\eta^{-2C_2}, \quad (6.4)$$
$$\rho(v) \geqslant \exp(-\eta^{-2C_5}), \quad (6.5)$$
$$\text{vol}(v) \leqslant \exp(\eta^{-2C_3}), \quad (6.6)$$
$$\text{waste}(v) \leqslant \eta^{C_3/2}. \quad (6.7)$$

From (6.7) we see in particular that the almost uniformity axiom in Proposition 3.3(ii) is obeyed. The thickness axiom in Proposition 3.3(i) is also easy, as the following corollary shows.

COROLLARY 6.5. *Suppose a quadratic approximant $v = v_k \in V$ can be reached from v_0 by a path $v_0 \to v_1 \to \cdots \to v_k$ of length k at most $8\eta^{-2C_2}$. Then we have $\mathbb{P}(\mathbf{r}_v = 0) \ll \exp(\eta^{-C_5^2})/p$.*

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Proof. Write

$$v = (C, \mathbf{c}, (n_c + B(S_c, \rho_c))_{c\in C}, (G_c)_{c\in C}, (F_c)_{c\in C}, (\Xi_c)_{c\in C}).$$

It suffices to show that

$$\mathbb{P}(\mathbf{r}_v = 0 | \mathbf{c} = c) \ll \exp(\eta^{-C_5^2})/p$$

for each c in the essential range of $\mathbf{c}$. But once $\mathbf{c}$ is fixed to equal $\mathbf{c}$, then $\mathbf{r}_v$ is drawn regularly from $n_c + B(S_c, \exp(-\eta^{-C_4})\rho_c)$. By Lemma 6.4, S_c has cardinality at most $8\eta^{-3C_2}$ and ρ_c is at least $\exp(-\eta^{-2C_5})$. The claim now follows from Lemma 4.2. □

It remains to verify the last two axioms (iii), (iv) of Proposition 3.3. We isolate these statements formally, using Lemma 6.4 and Definition 6.3.

The first of these results, Theorem 6.6, states that "a bad approximation implies an energy decrement". The second, Theorem 6.7, states that "a bad lower bound implies a dimension increment".

THEOREM 6.6. *Let the notation and hypotheses be as above. Suppose that $v \in V$ is a structured local approximant obeying (6.3)–(6.6). If we have*

$$\mathrm{Err}_1(v) > \eta \tag{6.8}$$

or

$$\mathrm{Err}_4(v) > \eta \tag{6.9}$$

then there exists a structured local approximant v' obeying the bounds

$$d(v') \leqslant d(v) + \eta^{-C_2}, \tag{6.10}$$
$$d_2(v') \leqslant d_2(v) + 1, \tag{6.11}$$
$$\rho(v') \geqslant \exp(-\eta^{-C_5})\rho(v), \tag{6.12}$$
$$\mathrm{vol}(v') \leqslant \exp(\eta^{-C_3})\,\mathrm{vol}(v), \tag{6.13}$$
$$|\mathrm{waste}(v') - \mathrm{waste}(v)| \leqslant \eta^{C_3}, \tag{6.14}$$
$$\mathrm{Energy}(v') \leqslant \mathrm{Energy}(v) - \eta^{C_2}. \tag{6.15}$$

THEOREM 6.7. *Let the notation and hypotheses be as above. Suppose that $v \in V$ is a structured local approximant obeying (6.3)–(6.6), and let $\mathbf{a}_v, \mathbf{r}_v, \mathbf{f}_v$ be the random variables associated to v. If we have*

$$\Lambda_{\mathbf{a}_v,\mathbf{r}_v}(\mathbf{f}_v) \leqslant (\mathbb{E}\mathbf{f}_v(\mathbf{a}_v))^4 - \eta, \tag{6.16}$$

then there exists a quadratic approximant $v' \in V$ with

$$d(v') \leqslant d(v) + \eta^{-C_2}, \tag{6.17}$$
$$d_2(v') \leqslant d_2(v), \tag{6.18}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

$$d_2^{\mathrm{poor}}(v') \leqslant d_2^{\mathrm{poor}}(v) - 1, \tag{6.19}$$

$$\rho(v') \geqslant \exp(-\eta^{-C_5})\rho(v), \tag{6.20}$$

$$\mathrm{vol}(v') \leqslant \exp(\eta^{-C_3})\,\mathrm{vol}(v), \tag{6.21}$$

$$|\mathrm{waste}(v') - \mathrm{waste}(v)| \leqslant \eta^{C_3}, \tag{6.22}$$

$$\mathrm{Energy}(v') \leqslant \mathrm{Energy}(v) + \eta^{3C_2}. \tag{6.23}$$

It remains to prove Theorems 6.6 and 6.7. Theorem 6.6 will be proven in §8 using a difficult local inverse Gowers theorem, Theorem 8.1, that will be proven in later sections. Theorem 6.7, on the other hand, will not rely on the local inverse Gowers theorem; it is proven in §7.

§7. *Bad lower bound implies dimension decrement.* In this section we prove Theorem 6.7. Let the notation and hypotheses be as in Theorem 6.7. We abbreviate $\mathbf{a}_v, \mathbf{r}_v, \mathbf{f}_v$ as $\mathbf{a}, \mathbf{r}, \mathbf{f}$ respectively. We can write the left-hand side of (6.16) as $\mathbb{E}A(\mathbf{c})$, where for any $c \in C$, the quantity $A(c)$ is defined as the conditional expectation

$$A(c) := \Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f}|\mathbf{c} = c).$$

Similarly, we can write $\mathbb{E}\mathbf{f}(\mathbf{a}) = \mathbb{E}B(\mathbf{c})$, where $B(\mathbf{c}) := \mathbb{E}(\mathbf{f}(\mathbf{a})|\mathbf{c} = c)$. By (6.16) and Hölder's inequality, we thus have

$$\mathbb{E}B(\mathbf{c})^4 - A(\mathbf{c}) \geqslant \eta.$$

Applying Lemma 2.2, we must therefore have

$$\mathbb{P}(B(\mathbf{c})^4 - A(\mathbf{c}) > \eta/2) \gg \eta.$$

By (6.1), we conclude that $\mathbf{c}$ is poorly distributed with probability $\gg\eta$. In particular, there is at least one poorly distributed value of c.

Most of this section will be devoted to the proof of the following proposition, which roughly speaking asserts that when $\mathbf{c}$ is poorly distributed, there is a linear constraint between the quadratic frequencies that will ultimately allow us to decrease the poorly distributed quadratic dimension d_2^{poor}.

PROPOSITION 7.1. *Let c be a poorly distributed element of the essential range of $\mathbf{c}$. Then there exists a natural number m_c, a frequency $\xi_c \in \mathbb{Z}/p\mathbb{Z}$ and an irreducible dual frequency $k_c' \in \hat{G}_c$ with*

$$1 \leqslant m_c \ll \exp(\eta^{-4C_3}) \tag{7.1}$$

and

$$\exp(-\eta^{-4C_3}) \ll |k_c'| \ll \exp(\eta^{-3C_2}) \tag{7.2}$$

such that

$$\|k_c' \cdot \Xi_c(a + 2m_c h) - k_c' \cdot \Xi_c(a)\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-3C_4}) \tag{7.3}$$

for all $a \in B(S_c, \rho_c/2)$ and $h \in B(S_c \cup \{\xi_c\}, \exp(-\eta^{-5C_4})\rho)$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

A key technical point here is that the upper bound on $|k_c'|$ involves only C_2 and not C_3 or C_4; this is necessary to keep the bounds under control during the iteration process. However, we will be able to tolerate the presence of the C_3 and C_4 constants in the other components of Proposition 7.1.

Proof. We condition on the event $\mathbf{c} = c$. By Definition 6.1, the random variables $\mathbf{a}, \mathbf{r}$ are now independent and regularly drawn from $n_c + B(S_c, \rho_c/2)$ and $B(S_c, \exp(-\eta^{-C_4})\rho_c)$ respectively, while $\mathbf{f}(n) = F_c(\Xi_c(a))$. We conclude that

$$\begin{aligned}&\mathbb{E}(F_c(\Xi_c(\mathbf{a}))F_c(\Xi_c(\mathbf{a}+\mathbf{r}))F_c(\Xi_c(\mathbf{a}+2\mathbf{r}))F_c(\Xi_c(\mathbf{a}+3\mathbf{r}))|\mathbf{c}=c)\\ &\quad < \mathbb{E}(F_c(\Xi_c(\mathbf{a}))|\mathbf{c}=c)^4 - \eta/2.\end{aligned}$$

Since $\Xi_c : \mathbb{Z}/p\mathbb{Z} \to G_c$ is locally quadratic on $n_c + B(S_c, \rho_c)$, which contains the progression $\mathbf{a}, \mathbf{a}+\mathbf{r}, \mathbf{a}+2\mathbf{r}, \mathbf{a}+3\mathbf{r}$, we see from (4.17) that

$$\Xi_c(\mathbf{a}) - 3\Xi_c(\mathbf{a}+\mathbf{r}) + 3\Xi_c(\mathbf{a}+2\mathbf{r}) - \Xi_c(\mathbf{a}+3\mathbf{r}) = 0$$

and so the left-hand side can be written as

$$\mathbb{E}(F_c^{(3)}(\Xi_c(\mathbf{a}), \Xi_c(\mathbf{a}+\mathbf{r}), \Xi_c(\mathbf{a}+2\mathbf{r}))|\mathbf{c}=c),$$

where $F_c^{(3)} : G_c^3 \to [-1, 1]$ is the function

$$F_c^{(3)}(x_0, x_1, x_2) := F_c(x_0)F_c(x_1)F_c(x_2)F_c(x_0 - 3x_1 + 3x_2).$$

Applying Lemma 3.2, we have

$$\int_{G_c^3} F_c^{(3)}(x_0, x_1, x_2)\, d\mu_c(x_0)\, d\mu_c(x_1)\, d\mu_c(x_2) \geqslant \left(\int_{G_c} F_c(x)\, d\mu_c(x)\right)^4,$$

where μ_c is the probability Haar measure on G_c. By the triangle inequality, we conclude that at least one of the assertions

$$\begin{aligned}&\left|\mathbb{E}(F_c^{(3)}(\Xi_c(\mathbf{a}), \Xi_c(\mathbf{a}+\mathbf{r}), \Xi_c(\mathbf{a}+2\mathbf{r}))|\mathbf{c}=c)\right.\\ &\quad\left. - \int_{G_c^3} F_c^{(3)}(x_0, x_1, x_2)\, d\mu_c(x_0)\, d\mu_c(x_1)\, d\mu_c(x_2)\right| \gg \eta\end{aligned}$$

or

$$\left|\mathbb{E}(F_c(\Xi_c(\mathbf{a}))|\mathbf{c}=c) - \int_{G_c} F_c(x)\, d\mu_c(x)\right| \gg \eta$$

holds. Defining $\tilde{F} : G_c^3 \to [-1, 1]$ by

$$\begin{aligned}&\tilde{F}(x_0, x_1, x_2)\\ &\quad = \frac{1}{10}\left(F_c^{(3)}(x_0, x_1, x_2) - \int_{G_c^3} F_c^{(3)}(x_0, x_1, x_2)\, d\mu_c(x_0)\, d\mu_c(x_1)\, d\mu_c(x_2)\right)\end{aligned}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

in the former case and

$$\tilde{F}(x_0, x_1, x_2) := \frac{1}{10}\left(F_c(x_0) - \int_{G_c} F_c(x_0)\, d\mu_c(x_0)\right)$$

in the latter case, we see that $\tilde{F}$ is 1-Lipschitz and of mean zero, and

$$|\mathbb{E}(\tilde{F}(\mathbf{x}_c)|\mathbf{c} = c)| \gg \eta, \tag{7.4}$$

where $\mathbf{x}_c \in G_c^3$ is the random variable

$$\mathbf{x}_c := (\Xi_c(\mathbf{a}), \Xi_c(\mathbf{a}+\mathbf{r}), \Xi_c(\mathbf{a}+2\mathbf{r})).$$

The Weyl equidistribution criterion, applied in the contrapositive, then suggests that there should be a non-zero dual frequency $k = (k_1, k_2, k_3) \in \hat{G}_c^3$ to G_c^3 such that $\mathbb{E}(e(k \cdot \mathbf{x}_c)|\mathbf{c} = c)$ is large. The next lemma makes this intuition precise.

LEMMA 7.2 (Weyl equidistribution). *With the notation and hypotheses as above, there exists a non-zero dual frequency $k = (k_1, k_2, k_3) \in \hat{G}_c^3$ to G_c^3 with $|k| \ll \exp(O(\eta^{-3C_2}))$ such that*

$$|\mathbb{E}(k \cdot \mathbf{x}_c|\mathbf{c} = c)| \gg \exp(-O(\eta^{-3C_2}))/\mathrm{vol}(G_c).$$

A key point here is that the bound on $|k|$ does not depend on the volume of the dilated torus G_c, which will typically be much larger than η^{-2C_2-10}.

Proof. Write $G_c = \prod_{i=1}^d (\mathbb{R}/\lambda_i\mathbb{Z})$, thus $\lambda_1, \ldots, \lambda_d \geqslant 1$, and by (6.4) one has

$$d \leqslant 8\eta^{-2C_2}. \tag{7.5}$$

The bound (7.4) is not possible when $d = 0$, so we may assume $d \geqslant 1$. We can write $G_c^3 = \prod_{i=1}^{3d} (\mathbb{R}/\lambda_i\mathbb{Z})$, where we extend λ_i periodically with period d.

Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a fixed smooth even function supported on $[-1, 1]$ that equals 1 at the origin and whose Fourier transform $\hat{\varphi}(\xi) := \int_{\mathbb{R}} \phi(x)e(-x\xi)\, dx$ is non-negative; such a function may be easily constructed by convolving an L^2-normalized smooth function on $[0, 1]$ with its reflection. Let $A \geqslant 1$ be a parameter to be chosen later, and introduce the kernel $K : G_c^3 \to \mathbb{R}^+$ by the formula

$$K(t_1, \ldots, t_{3d}) := \prod_{i=1}^{3d} K_i(t_i)$$

for $t_i \in \mathbb{R}/\lambda_i\mathbb{Z}$, where

$$K_i(t_i) := \sum_{k_i \in (1/\lambda_i)\mathbb{Z}} \varphi\left(\frac{k_i}{A}\right) e(k_i t_i).$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

By Poisson summation, the K_i and hence K are non-negative. A Fourier-analytic calculation using the smoothness of φ gives

$$\int_{\mathbb{R}/\lambda_i\mathbb{Z}} K_i(t_i)\,\frac{dt_i}{\lambda_i} = 1$$

and

$$\int_{\mathbb{R}/\lambda_i\mathbb{Z}} K_i(t_i)\sin^2(\pi t_i/\lambda_i)\,\frac{dt_i}{\lambda_i} \ll \frac{1}{A^2\lambda_i^2}$$

(where the implied constant is allowed to depend on φ) and hence by (2.2) and Cauchy–Schwarz we have

$$\int_{\mathbb{R}/\lambda_i\mathbb{Z}} K_i(t_i)\|t_i\|_{\mathbb{R}/\mathbb{Z}}\,\frac{dt_i}{\lambda_i} \ll \frac{1}{A},$$

which on taking tensor products gives

$$\int_{G_c^3} K(x)\,d\mu_c^3(x) = 1$$

and

$$\int_{G_c^3} K(x)\|x\|_{G_c^3}\,d\mu_c^3(x) \ll \frac{d}{A},$$

where μ_c^3 is the Haar probability measure on G_c^3. If we then take the convolution

$$\tilde{F} * K(x) := \int_{G_c} \tilde{F}(x-y)K(y)\,d\mu_c^3(y)$$

then by the 1-Lipschitz nature of $\tilde{F}$ we see that

$$\tilde{F} * K(x) = \tilde{F}(x) + O\left(\frac{d}{A}\right).$$

Thus, if we choose

$$A := \frac{Cd}{\eta}$$

for a sufficiently large absolute constant C, we conclude from (7.4) that

$$|\mathbb{E}(\tilde{F} * K(\mathbf{x}_c)|\mathbf{c} = c)| \gg \eta.$$

However, by Fourier expansion and the fact that $\tilde{F}$ has mean zero,

$$\tilde{F} * K(\mathbf{x}_c) = \sum_{k\in\hat{G}_c^3\setminus\{0\}} \left(\prod_{i=1}^{3d}\varphi\left(\frac{k_i}{A}\right)\right)\widehat{\tilde{F}(k)}\mathbb{E}e(k\cdot\mathbf{x}),$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

where $k = (k_1, \dots, k_{3d})$ with $k_i \in (1/\lambda_i)\mathbb{Z}$ for $i = 1, \dots, 3d$, and

$$\widehat{\widetilde{F}}(k) := \int_{G_c^3} \tilde{F}(x) e(-k \cdot x)\, d\mu_c^3(x).$$

Using the triangle inequality and crudely bounding $|\widehat{\widetilde{F}}(k)|$ by 1, we conclude that

$$\sum_{k \in \hat{G}_c^3 \setminus \{0\}} \left(\prod_{i=1}^{d} \left| \varphi\left(\frac{k_i}{A}\right) \right| \right) |\mathbb{E}(e(k \cdot \mathbf{x}_c) | \mathbf{c} = c)| \gg \eta.$$

The summand is only non-vanishing when $\sup_i |k_i| \leqslant A$, so that

$$|k| \leqslant dA \ll \exp(O(\eta^{-3C_2}))$$

(thanks to (7.5) and the choice of A), and the number of such k is

$$O\left(\prod_{i=1}^{3d} (A\lambda_i) \right) \ll \exp(O(\eta^{-3C_2}))\, \mathrm{vol}(T).$$

Since φ is bounded, the claim now follows from the pigeonhole principle. □

We return to the proof of Proposition 7.1. Applying Lemma 7.2 and (6.5), we see that there exists a non-zero triplet $(k_c^0, k_c^1, k_c^2) \in \hat{G}_c^3$ with

$$|k_c^0|, |k_c^1|, |k_c^2| \ll \exp(\eta^{-3C_2}) \tag{7.6}$$

and

$$\mathbb{E}(e(k_c^0 \cdot \Xi_c(\mathbf{a}) + k_c^1 \cdot \Xi_c(\mathbf{a}+\mathbf{r}) + k_c^2 \cdot \Xi_c(\mathbf{a}+2\mathbf{r})) | \mathbf{c} = c) \gg \exp(-\eta^{-3C_3}). \tag{7.7}$$

Among other things, the non-zero nature of this triplet forces G_c to be non-trivial, and thus

$$d_2^{\mathrm{poor}}(v) \geqslant 1.$$

We also emphasize that the bound (7.6) involves C_2 rather than C_3; this will become important when establishing the important upper bound of (7.2) later in this proof.

We can use the exponential sum bound (7.7) to control the "second derivative" of Ξ_c. Indeed, for any $h_1, h_2 \in B(S_c, \rho_c/10)$, define the quantity $\partial^2 \Xi_c(h_1, h_2) \in \mathbb{R}/\mathbb{Z}$ by

$$\partial^2 \Xi_c(h_1, h_2) := \Xi_c(a + h_1 + h_2) - \Xi_c(a + h_1) - \Xi_c(a + h_2) + \Xi_c(a)$$

for any $a \in n_c + B(S_c, \rho/2)$. Since $\Gamma_{\mathbf{c}}$ is locally quadratic on $n_c + B(S_c, \rho)$, this quantity is well-defined, symmetric in h_1, h_2, and is also locally bilinear in h_1 and h_2.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

LEMMA 7.3. *Let the notation and hypotheses be as above. Then for any $i = 0, 1, 2$, we have*

$$|\mathbb{E}(e(2k_c^i \cdot \partial^2 \Xi_c(\mathbf{r}-\mathbf{r}', \mathbf{h}-\mathbf{h}'))|\mathbf{c}=c)| \gg \exp(-4\eta^{-3C_3}),$$

where, conditioning on the event $\mathbf{c} = c$, the random variables $\mathbf{r}, \mathbf{r}', \mathbf{h}, \mathbf{h}'$ are drawn independently and regularly from the Bohr sets $B(S_c, \exp(-\eta^{-C_4})\rho)$, $B(S_c, \exp(-\eta^{-C_4})\rho)$, $B(S_c, \exp(-\eta^{-2C_4})\rho)$, $B(S_c, \exp(-\eta^{-2C_4})\rho)$ respectively, independently of $\mathbf{a}$.

Proof. To simplify the notation we only consider the $i = 2$ case, as the $i = 0, 1$ cases are similar. This will be "Weyl differencing" argument that relies primarily on the Cauchy–Schwarz inequality.

Recall that after conditioning to the event $\mathbf{c} = c$, the random variable $\mathbf{a}$ is drawn regularly from $B(S_c, \rho/2)$. Using Lemma 4.4, we see that $\mathbf{a}$ and $\mathbf{a} - \mathbf{h}$ differ in total variation by $O(\exp(-\eta^{-C_4/2}))$, hence from (7.7) we have

$$|\mathbb{E}(e(k_c^0 \cdot \Xi_c(\mathbf{a}-\mathbf{h}) + k_c^1 \cdot \Xi_c(\mathbf{a}-\mathbf{h}+\mathbf{r}) + k_c^2 \cdot \Xi_c(\mathbf{a}-\mathbf{h}+2\mathbf{r}))|\mathbf{c}=c)| \gg \exp(-\eta^{-3C_3}).$$

Similarly we may use Lemma 4.4 to compare $\mathbf{r}$ and $\mathbf{r} + \mathbf{h}$, and conclude that

$$|\mathbb{E}(e(k_c^0 \cdot \Xi_c(\mathbf{a}-\mathbf{h}) + k_c^1 \cdot \Xi_c(\mathbf{a}+\mathbf{r}) + k_c^2 \cdot \Xi_c(\mathbf{a}+\mathbf{h}+2\mathbf{r}))|\mathbf{c}=c)| \gg \exp(-\eta^{-3C_3}).$$

By the pigeonhole principle (and independence of $\mathbf{a}, \mathbf{h}, \mathbf{r}$ relative to the event $\mathbf{c} = c$), we may thus find $a_c \in n_c + B(S_c, \rho/2)$ such that

$$|\mathbb{E}(e(k_c^0 \cdot \Xi_c(a_c-\mathbf{h}) + k_c^1 \cdot \Xi_c(a_c+\mathbf{r}) + k_c^2 \cdot \Xi_c(a_c+\mathbf{h}+2\mathbf{r}))|\mathbf{c}=c)| \gg \exp(-\eta^{-3C_3}).$$

Using the identity

$$\Xi_c(a_c+\mathbf{h}+2\mathbf{r}) = \Xi_c(a_c+\mathbf{h}) + \Xi_c(a_c+2\mathbf{r}) - \Xi_c(a_c) + \partial^2\Xi_c(2\mathbf{r}, \mathbf{h})$$

we can rewrite the left-hand side as

$$|\mathbb{E}(b_1(\mathbf{r})b_2(\mathbf{h})e(k_c^2 \cdot \partial^2\Xi_c(2\mathbf{r}, \mathbf{h}))|\mathbf{c}=c)| \gg \exp(-\eta^{-3C_3})$$

where $b_1, b_2 : B(S_c, \rho) \to \mathbb{C}$ are the 1-bounded functions

$$b_1(r) := e(k_c^1 \cdot \Xi_{\mathbf{c}}(a_c + r) + k_c^2 \cdot \Xi_c(a_c + 2r) - k_c^2 \cdot \Xi_c(a_c))$$

and

$$b_2(h) := e(k_c^0 \cdot \Xi_c(a_c - h) + k_c^2 \cdot \Xi_c(a_c + h)).$$

Applying Lemma 2.1 to eliminate the $\mathbf{b}_1(\mathbf{r})$ factor, we conclude that

$$|\mathbb{E}(b_2(\mathbf{h})\overline{b_2(\mathbf{h}')}e(k_c^2 \cdot \partial^2\Xi_c(2\mathbf{r}, \mathbf{h}-\mathbf{h}'))|\mathbf{c}=c)| \gg \exp(-2\eta^{-3C_3}).$$

Applying Lemma 2.1 again to eliminate the $b_2(\mathbf{h})\overline{b_2(\mathbf{h}')}$ factor, we obtain the claim. □

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We return to the proof of Proposition 7.1. Let $i = i_c \in \{0,1,2\}$ be such that k_c^i is non-zero. Let $\mathbf{r}, \mathbf{r}', \mathbf{h}, \mathbf{h}'$ be as in the above lemma, and let $\mathbf{h}''$ be a further independent copy of $\mathbf{h}$ or $\mathbf{h}'$, thus $\mathbf{h}''$ is also drawn regularly from $B(S_c, \exp(-\eta^{-2C_4})\rho)$ and independently of $\mathbf{r}, \mathbf{r}', \mathbf{h}, \mathbf{h}'$ (after conditioning on $\mathbf{c} = c$). Applying Lemma 4.4 to compare $\mathbf{r}$ with $\mathbf{r} + \mathbf{h}''$, we have

$$|\mathbb{E}(e(2k_c^i \cdot \partial^2 \Xi_c(\mathbf{r} - \mathbf{r}' + \mathbf{h}'', \mathbf{h} - \mathbf{h}'))|\mathbf{c} = c)| \gg \exp(-4\eta^{-3C_3}),$$

so by the pigeonhole principle we can find $r, r', h' \in B(S_c, \exp(-\eta^{-C_4})\rho_c)$ (depending on c, of course) such that

$$|\mathbb{E}(e(2k_c^i \cdot \partial^2 \Xi_c(r - r' + \mathbf{h}'', \mathbf{h} - h'))|\mathbf{c} = c)| \gg \exp(-4\eta^{-3C_3}).$$

By the local bilinearity of $\partial^2 \Xi_c$, we may thus have

$$|\mathbb{E}(e(2k_c^i \cdot \partial^2 \Xi_c(\mathbf{h}'', \mathbf{h}) + \psi(\mathbf{h}) + \psi''(\mathbf{h}''))|\mathbf{c} = c)| \gg \exp(-4\eta^{-3C_3})$$

for some locally linear functions $\psi, \psi'' : B(S_c, \rho/100) \to \mathbb{R}/\mathbb{Z}$ (which can depend on c).

Applying Proposition 4.11 (recalling from (6.3) that $|S_c| \leqslant 8\exp(-3C_2)$), we conclude that there exists a non-zero multiple $k_c \in \hat{G}_c$ of k_c^i with

$$k_c \ll \exp(\eta^{-4C_3}) \tag{7.8}$$

such that

$$\|k_c \cdot \partial^2 \Xi_c(n, m)\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(\eta^{-3C_4}) \frac{\|n\|_{S_c} \|m\|_{S_c}}{\rho_c^2} \tag{7.9}$$

for $n, m \in B(S_c, \exp(-\eta^{-3C_4})\rho_c)$.

Applying Corollary 4.13, we may thus find $\xi_c \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\left\| k_c \cdot \Xi_c(n_c + h) - k_c \cdot \Xi_c(n_c) - \frac{\xi_c h}{p} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(\eta^{-4C_4}) \frac{\|h\|_{S_c}}{\rho_c} \tag{7.10}$$

for all $n \in \mathbb{Z}/p\mathbb{Z}$ (of course, the bound is only non-trivial when h lies in the Bohr set $B(S_c, \exp(-\eta^{-4C_4})\rho)$).

The dual frequency $k_c \in \widehat{G_c}$ is non-zero, but not necessarily irreducible. However, we may write $k_c = m_c k_c'$ where m_c is a positive natural number and $k_c' \in \widehat{G_c}$ is irreducible, thus by (7.8) we have the bound (7.1). The same argument gives the bound $k_c' \ll \exp(\eta^{-4C_3})$, but this is not sufficient to establish the upper bound in (7.2). However, observe that k_c^i must also be a multiple of the irreducible vector k_c', and now the upper bound in (7.2) follows from (7.6).

We can also obtain a lower bound on k_c' by observing that the slab

$$\{x \in G_c : \|k_c' \cdot x\|_{\mathbb{R}/\mathbb{Z}} \leqslant \tfrac{1}{2}|k_c'|\}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

has measure at most $|k_c'|\,\mathrm{vol}(G_c)$, and contains the Euclidean ball of radius $1/2$ centred at the origin. This gives the lower bound

$$|k_c'| \gg \frac{1}{\dim(G_c)^{O(\dim(G_c))}\,\mathrm{vol}(G_c)}$$

which by (6.4), (6.6) gives the lower bound in (7.2).

Now let $a \in B(S_c, \rho_c/2)$ and $h \in B(S_c \cup \{\xi_c\}, \exp(-\eta^{-5C_4})\rho_c)$. Then we have

$$jh \in B(S_c, 2m_c \exp(-\eta^{-5C_4})\rho_c)$$

and

$$\left\| \frac{j\xi_c h}{p} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-5C_4})\rho_c$$

for all $j, 0 \leqslant j \leqslant 2m_c$. From (7.10) and (7.1), we conclude that

$$\|k_c \cdot \Xi_c(n_c + jh) - k_c \cdot \Xi_c(n_c)\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-4C_4})$$

(for example). On the other hand, from (7.9) we have

$$\|k_c \cdot (\Xi_c(a + jh) - \Xi_c(a) - \Xi_c(n_c + j\xi_c h) + \Xi_c(n_c))\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-4C_4})$$

and hence by the triangle inequality we have

$$\|k_c \cdot \Xi_c(a + jh) - k_c \cdot \Xi_c(a)\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-4C_4}) \tag{7.11}$$

for all $j, 0 \leqslant j \leqslant 2m_c$.

This is close to (7.3), but we will need to replace the dual frequency k_c here with the irreducible dual frequency k_c'. To do this, we first observe that as Ξ_c is locally quadratic on $n_c + B(S_c, \rho_c)$, we may write

$$\Xi_c(a + jh) = \alpha + \beta j + \gamma j^2 \tag{7.12}$$

for all $j, 0 \leqslant j \leqslant 2m_c$, and some $\alpha, \beta, \gamma \in G_c$ depending on c, a, h. Inserting this formula into the preceding estimate, we conclude that

$$\|j(k_c \cdot \beta) + j^2(k_c \cdot \gamma)\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-4C_4})$$

for $j, 0 \leqslant j \leqslant 2m_c$. Applying this for $j = 1, 2$ and using the triangle inequality, we have

$$\|k_c \cdot \beta\|, \|2(k_c \cdot \gamma)\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-4C_4}).$$

Since $2m_c k_c' = 2k_c$ and $(2m_c)^2 k_c' = (2m_c)2k_c$, we conclude in particular (using (7.1)) that

$$\|2m_c(k_c' \cdot \beta)\|_{\mathbb{R}/\mathbb{Z}}, \|(2m_c)^2(k_c' \cdot \gamma)\|_{\mathbb{R}/\mathbb{Z}} \ll \exp(-\eta^{-3C_4})$$

and thus by (7.12) we obtain (7.3) as desired. This finally completes the proof of Proposition 7.1. □

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We now return to the proof of Theorem 6.7. We are given a structured local approximant

$$v = (C, \mathbf{c}, (n_c + B(S_c, \rho_c))_{c\in C}, (G_c)_{c\in C}, (F_c)_{c\in C}, (\Xi_c)_{c\in C})$$

and need to construct a modification

$$v' = (C', \mathbf{c}', (n'_{c'} + B(S'_{c'}, \rho'_{c'}))_{c'\in C'}, (G'_{c'})_{c'\in C'}, (F'_{c'})_{c'\in C'}, (\Xi'_{c'})_{c'\in C'})$$

that somehow incorporates the linear constraint identified in Proposition 7.1 to decrement the poorly distributed quadratic dimension of v', in the spirit of the third and fourth examples in §3. To avoid confusion, we shall restore the subscripts $(\mathbf{a}_v, \mathbf{r}_v, \mathbf{f}_v)$ on the random variables associated to v as per Definition 6.1, to distinguish them from the corresponding random variables $(\mathbf{a}_{v'}, \mathbf{r}_{v'}, \mathbf{f}_{v'})$ that will be associated to v'.

We shall set $C' := (\mathbb{Z}/p\mathbb{Z}) \times C$, and let $\mathbf{c}'$ be the random variable

$$\mathbf{c}' := (\mathbf{a}_v, \mathbf{c}).$$

Clearly $\mathbf{c}'$ takes values in the non-empty finite set C'. Now we need to define $n'_{c'}$, $S'_{c'}$, $\rho'_{c'}$, $G'_{c'}$, $F'_{c'}$, $\Xi'_{c'}$ for any given $c' = (a, c)$ in C'. In the case where c is not poorly distributed, we simply carry over the corresponding data from v without further modification. That is to say, we define

$$(n'_{c'}, S'_{c'}, \rho'_{c'}, G'_{c'}, F'_{c'}, \Xi'_{c'}) := (n_c, S_c, \rho_c, G_c, F_c, \Xi_c)$$

whenever $c' = (a, c)$ with c not poorly distributed. If instead $c' = (a, c)$ with c poorly distributed, then we introduce the natural number m_c, the dual frequency $k'_c \in \hat{G}_c$, and the frequency $\xi_c \in \mathbb{Z}/p\mathbb{Z}$ from Proposition 7.1; of course we can arrange matters so that m_c, k'_c, ξ_c depend only on c and not on a. Because of (7.1) and the hypothesis (3.21), the quantity $2m_c$ is invertible in the field $\mathbb{Z}/p\mathbb{Z}$, and so we may define the dilate $(2m_c)^{-1} \cdot S_c$ of S_c inside $\mathbb{Z}/p\mathbb{Z}$, and can similarly define the dilate $(2m_c)^{-1}\xi_c$ of ξ_c. We will need to do this division here to cancel some denominators appearing later in the argument.

In this poorly distributed case, we define the "linear" data $n'_{c'}$, $S'_{c'}$, $\rho'_{c'}$ by

$$\begin{aligned} n'_{c'} &:= a, \\ S'_{c'} &:= (2m_c)^{-1} \cdot S_c \cup \{(2m_c)^{-1}\xi_c\}, \\ \rho'_{c'} &:= \exp(-\eta^{-6C_4})\rho_c, \end{aligned}$$

thus the shifted Bohr set $n'_{c'} + B(S'_{c'}, \rho'_{c'})$ will be a small subset of $n_c + B(S_c, \rho_c)$ in which the radius ρ_c has been reduced and an additional frequency $\xi_c/2m_c$ has been added. As we shall see, this particular choice of this linear data will allow us to utilize the approximate constraint (7.3).

The constraint (7.3) has the effect of approximately restricting Ξ_c (on a suitable Bohr set) to a coset of the orthogonal complement $(k'_c)^\perp = \{x \in G_c :$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

$k'_c \cdot x = 0\}$ of k'_c in G_c. Applying Theorem 5.1, (6.4), and the crucial bound (7.2), we may find a dilated torus $\tilde{G}_c = \prod_{i=1}^{\dim(G_c)-1} (\mathbb{R}/\tilde{\lambda}_{c,i}\mathbb{Z})$ with volume

$$\mathrm{vol}(\tilde{G}_c) \ll \exp(\eta^{-4C_2}) \,\mathrm{vol}(G_c) \tag{7.13}$$

as well as a Lie group isomorphism $\psi_c : (k'_c)^\perp \to \tilde{G}_c$ obeying the bilipschitz bounds

$$\|\psi\|_{\mathrm{Lip}}, \|\psi^{-1}\|_{\mathrm{Lip}} \leqslant \exp(\eta^{-4C_2}).$$

In particular, if we define the even more dilated torus

$$G'_c := \prod_{i=1}^{\dim(G_c)-1} (\mathbb{R}/\exp(\eta^{-4C_2})\tilde{\lambda}_{c,i}\mathbb{Z})$$

and let $\delta_c : G'_c \to \tilde{G}_c$ be the rescaling map

$$\delta_c : (x_i)_{i=1}^{\dim(G_c)-1} \mapsto (\exp(-\eta^{-4C_2})x_i)_{i=1}^{\dim(G_c)-1}$$

then we see that $\psi^{-1} \circ \delta_c : G'_c \to (k'_c)^\perp$ is a 1-Lipschitz Lie group isomorphism.

An element of $n'_{c'} + B(S'_c, \rho'_c)$ can be uniquely represented in the form $n'_{c'} + 2m_c h$ for $h \in B(S_c \cup \{\xi_c\}, \exp(-\eta^{-6C_4})\rho_c)$. From (7.3), we know that the point $\Xi_c(n'_{c'} + 2m_c h) - \Xi_c(n'_{c'})$ lies within a $O(\exp(-\eta^{-3C_4}))$-neighbourhood of the subtorus $(k'_c)^\perp$. Using the lower bound in (7.2), we can find a locally linear projection π_c from this neighbourhood to the subtorus itself (e.g. by viewing the subtorus locally as a graph in $\dim(G_c) - 1$ of the $\dim(G_c)$ coordinates and then projecting in the direction of the remaining coordinate), which moves each point in the neighbourhood by at most $O(\exp(-\eta^{-2C_4}))$. From the 1-Lipschitz nature of F_c, we thus have

$$\begin{aligned} &F_c(\Xi_c(n'_{c'} + 2m_c h)) \\ &\quad = F_c(\pi_c(\Xi_c(n'_{c'} + 2m_c h) - \Xi_c(n'_{c'})) + \Xi_c(n'_{c'})) + O(\exp(-\eta^{-2C_4})). \end{aligned}$$

We can rewrite this as

$$F_c(\Xi_c(n'_{c'} + 2m_c h)) = F'_{c'}(\Xi'_{c'}(n'_{c'} + 2m_c h)) + O(\exp(-\eta^{-2C_4})), \tag{7.14}$$

where $F'_{c'} : G'_c \to [-1, 1]$ is the 1-Lipschitz function

$$F'_{c'}(x) := F_c(\psi_c^{-1}(\delta_c(x))) + \Xi_c(n'_{c'})$$

and $\Xi'_{c'} : n'_{c'} + B(S'_c, \rho'_c) \to G'_c$ takes the form

$$\Xi'_{c'}(n'_{c'} + 2mch) := \delta_c^{-1}(\psi_c(\pi_c(\Xi_c(n'_{c'} + 2m_c h) - \Xi_c(n'_{c'})) + \Xi_c(n'_{c'}))).$$

The map $\Xi'_{c'}$ is the composition of a locally quadratic map with three locally linear maps, and is hence also locally quadratic. This concludes the construction

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

of all the required quadratic data $G'_{c'}$, $F'_{c'}$, $\Xi'_{c'}$ when c' arises from a poorly distributed c.

It remains to verify the claims (6.17)–(6.23) of Theorem 6.7. The claim (6.17) is clear; in fact, the frequency sets $S'_{c'}$ are either equal to their original counterparts S_c or have the addition of just one further frequency ξ_c, so we even obtain the improved bound $d(v') \leqslant d(v) + 1$ in our construction here. Since the dilated torus $G'_{c'}$ is either equal to G_c when c is not poorly distributed, or has one lower dimension than G_c if c is poorly distributed, we obtain the bounds (6.18), (6.19). Since $\rho'_{c'}$ is either equal to ρ_c when c is not poorly distributed, or $\exp(-\eta^{-6C_4})\rho_c$ when c is poorly distributed, we obtain (6.20) (with a little room to spare). As for the volume bound, $G'_{c'}$ clearly has the same volume as G_c when c is not poorly distributed, and when c is poorly distributed we have

$$\operatorname{vol}(G'_{c'}) = \exp(-\eta^{-4C_2}\dim(\tilde{G}_{c'}))\operatorname{vol}(\tilde{G}_{c'})$$

which by (7.13), (6.3) is bounded in turn by $\exp(-\eta^{-5C_2})\operatorname{vol}(G_c)$, which yields (6.21), again with a little bit of room to spare (because the bounds here only increased the volume by factors that involved C_2 rather than C_3).

Now we establish (6.22). From the triangle inequality we have

$$\begin{aligned}|\operatorname{waste}(v') - \operatorname{waste}(v)| &\leqslant |\mathbb{E}f(\mathbf{a}_{v'}) - \mathbb{E}f(\mathbf{a}_v)| \\ &\leqslant \sum_{c\in C}\mathbb{P}(\mathbf{c}=c)|\mathbb{E}(f(\mathbf{a}_{v'})|\mathbf{c}=c) - \mathbb{E}(f(\mathbf{a}_v)|\mathbf{c}=c)|\end{aligned}$$

so it will suffice to show that

$$|\mathbb{E}(f(\mathbf{a}_{v'})|\mathbf{c}=c) - \mathbb{E}(f(\mathbf{a}_v)|\mathbf{c}=c)| \leqslant \eta^{C_3} \tag{7.15}$$

for each c in the essential range of $\mathbf{c}$.

The claim is trivial when c is not poorly distributed, since in this case $\mathbf{a}_v$ and $\mathbf{a}_{v'}$ have identical distribution after conditioning to $\mathbf{c} = c$. If c is poorly distributed, then (after conditioning to $\mathbf{c} = c$) $\mathbf{a}_v$ is drawn regularly from $n_c + B(S_c, \rho_c/2)$, while $\mathbf{a}_{v'}$ has the distribution of $\mathbf{a}_v + 2m_c\mathbf{h}_c$ where $\mathbf{h}_c$ is drawn regularly from $B(S_c \cup \{\xi_c\}, \exp(-\eta^{-6C_4})\rho_c)$ independently of $\mathbf{a}_v$ (after conditioning to $\mathbf{c} = c$). The required bound (6.22) now follows from Lemma 4.4 (and (6.3)).

Finally, we prove (6.23). Our task is to show that

$$\mathbb{E}|f(\mathbf{a}_{v'}) - \mathbf{f}_{v'}(\mathbf{a}_{v'})|^2 \leqslant \mathbb{E}|f(\mathbf{a}_v) - \mathbf{f}_v(\mathbf{a}_v)|^2 + \eta^{3C_2}.$$

By the triangle inequality as before, it suffices to show that

$$\mathbb{E}(|f(\mathbf{a}_{v'}) - \mathbf{f}_{v'}(\mathbf{a}_{v'})|^2|\mathbf{c}=c) \leqslant \mathbb{E}(|f(\mathbf{a}_v) - \mathbf{f}_v(\mathbf{a}_v)|^2|\mathbf{c}=c) + \eta^{3C_2}$$

for all c in the essential range of $\mathbf{c}$. This is trivial for c not poorly distributed, so assume c is poorly distributed. From (7.14) we then have

$$\mathbf{f}_{v'}(\mathbf{a}_{v'}) = \mathbf{f}_v(\mathbf{a}_{v'}) + O(\exp(-\eta^{-2C_4}))$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

and also

$$\mathbf{f}_v(a) = F_c(\Xi_c(a))$$

for $a \in B(S_c, \rho_c)$, so by the triangle inequality it suffices to show that

$$\mathbb{E}(|f(\mathbf{a}_{v'}) - F_c(\Xi_c(\mathbf{a}_{v'}))|^2|\mathbf{c} = c) \leqslant \mathbb{E}(|f(\mathbf{a}_v) - F_c(\Xi_c(\mathbf{a}_v))|^2|\mathbf{c} = c) + \eta^{4C_2}$$

(for example). But this follows by repeating the proof of (7.15), with the function f replaced by $|f - F_c \circ \Xi_c|^2$. This completes the proof of Theorem 6.7.

§8. *Bad approximation implies energy decrement.* The remaining task in the paper is to prove Theorem 6.6. In this section we will establish this result contingent on a local inverse Gowers norm theorem (Theorem 8.1) that will be proven in later sections. We begin by stating the (rather technical) precise form of that theorem that we will need.

THEOREM 8.1 (Local inverse U^3 theorem). *Let p be a prime, and let S be a subset of $\mathbb{Z}/p\mathbb{Z}$ containing at least one non-zero element. Let η be a real parameter with $0 < \eta < \frac{1}{2}$. Let K be the quantity*

$$K := \frac{1}{\eta} + |S|, \tag{8.1}$$

and let $\rho_0, \rho_1, \rho_2, \ldots, \rho_{10}$ be real numbers satisfying

$$0 < \rho_{10} < \cdots < \rho_0 < 1/2$$

as well as the separation condition

$$\rho_{i+1} \geqslant \exp(K^{C_2})\rho_i \tag{8.2}$$

for all $i = 0, \ldots, 9$. Assume that the prime p is huge relative to the reciprocal of these parameters, in the sense that

$$p \geqslant \rho_{10}^{-K^{C_2^3}}. \tag{8.3}$$

Let $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ be a 1-bounded function such that

$$\begin{aligned}|\mathbb{E}f(\mathbf{h}_0 + \mathbf{h}_1 + \mathbf{h}_2)\overline{f}(\mathbf{h}_0 + \mathbf{h}'_1 + \mathbf{h}_2)\overline{f}(\mathbf{h}'_0 + \mathbf{h}_1 + \mathbf{h}_2)f(\mathbf{h}'_0 + \mathbf{h}'_1 + \mathbf{h}_2)\\ \times \overline{f}(\mathbf{h}_0 + \mathbf{h}_1 + \mathbf{h}'_2)f(\mathbf{h}_0 + \mathbf{h}'_1 + \mathbf{h}'_2)f(\mathbf{h}'_0 + \mathbf{h}_1 + \mathbf{h}'_2)\overline{f}(\mathbf{h}'_0 + \mathbf{h}'_1 + \mathbf{h}'_2)|\\ \geqslant \eta\end{aligned} \tag{8.4}$$

whenever $\mathbf{h}_0, \mathbf{h}'_0, \mathbf{h}_1, \mathbf{h}'_1, \mathbf{h}_2, \mathbf{h}'_2$ are drawn independently and regularly from $B(S, \rho_0)$, $B(S, \rho_0)$, $B(S, \rho_1)$, $B(S, \rho_1)$, $B(S, \rho_2)$, and $B(S, \rho_2)$ respectively. Then there exists a positive integer $k < \exp(K^{O(C_1)})$, a set $S' \subset \mathbb{Z}/p\mathbb{Z}$, $S' \supset S$, with

$$|S'| \leqslant |S| + O(\eta^{-O(C_1)}), \tag{8.5}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

a locally quadratic phase $\phi : B(S', \rho_9) \to \mathbb{R}/\mathbb{Z}$, *and a function* $\beta : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ *such that*

$$\sum_{n \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n} = n) \left| \mathbb{E} f(n + k\mathbf{m}) e\left(-\phi(\mathbf{m}) - \frac{\beta(n)\mathbf{m}}{p} \right) \right| \gg \eta^{O(C_1)} \tag{8.6}$$

if $\mathbf{n}, \mathbf{m}$ *are drawn independently and regularly from* $B_S(0, \rho_0)$ *and* $B_{S'}(0, \rho_{10})$ *respectively.*

Remarks. The parameters $\rho_3, \dots, \rho_8$ do not have any role in the statement of this result, but they appear in the proof. We have retained them to avoid a potentially confusing relabelling.

Informally, this theorem asserts that if f has a large U^3 norm on $B(S, \rho_0)$, then f will correlate with a locally quadratic phase $n + km \mapsto \phi(m) + \beta(n)m/p$ on translates $n + k \cdot B_{S'}(0, \rho_{10})$ of $k \cdot B_{S'}(0, \rho_{10})$, with polynomial bounds on the correlation. Although we will not make crucial use of this fact in our arguments, it may be noted that the homogeneous component ϕ of this locally quadratic phase does not depend on the translation parameter n. In the bounded rank case $|S| = O(1)$, a theorem very roughly of this form was established in [**14**]; the key point in Theorem 8.1 is that the inverse theory of [**14**] can be localized to a Bohr set without having the lower bound $\eta^{O(C_1)}$ on the correlation appearing in (8.6) depend on the rank $|S|$ or radius ρ_0 of the Bohr set (although these parameters certainly influence the range of the variables $\mathbf{n}, \mathbf{m}$ appearing in (8.6)).

The proof of Theorem 8.1 will occupy most of the remainder of this paper. To a large extent, it may be understood separately of our main arguments, requiring little of the notation of §3, for example. In this section, we will assume Theorem 8.1 and use it to establish Theorem 6.6.

For the remainder of this section, the notation and hypotheses will be as in Theorem 6.6. Namely, we fix a prime p, a function $f : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$, and a parameter $0 < \eta \leqslant 1/10$, and assume (3.21). We also suppose that

$$v = (C, \mathbf{c}, (n_c + B(S_c, \rho_c))_{c \in C}, (G_c)_{c \in C}, (F_c)_{c \in C}, (\Xi_c)_{c \in C})$$

is a structured local approximant obeying (6.3)–(6.6), and one of (6.8) or (6.9) holds. Our objective is to construct a structured local approximant

$$v' = (C', \mathbf{c}', (n'_{c'} + B(S'_{c'}, \rho'_{c'}))_{c' \in C'}, (G'_{c'})_{c' \in C'}, (F'_{c'})_{c' \in C'}, (\Xi'_{c'})_{c' \in C'})$$

obeying the bounds (6.10)–(6.15). The situation here is a formalization of Example 8 from §3.

Let $\mathbf{a} = \mathbf{a}_v, \mathbf{r} = \mathbf{r}_v, \mathbf{f} = \mathbf{f}_v$ be the random variables associated to v in Definition 6.1. We can unify the hypotheses (6.8), (6.9) by introducing the quadrilinear form

$$\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3) := \mathbb{E}\mathbf{f}_0(\mathbf{a})\mathbf{f}_1(\mathbf{a} + \mathbf{r})\mathbf{f}_2(\mathbf{a} + 2\mathbf{r})\mathbf{f}_3(\mathbf{a} + 3\mathbf{r}),$$

defined for arbitrary random (or deterministic) bounded functions $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3 : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$. From the definitions of Err_1 and Err_4 (just prior to (6.1)),

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

the hypothesis (6.8) may be written as

$$|\Lambda_{\mathbf{a},\mathbf{r}}(f,1,1,1)-\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f},1,1,1)|>\eta,$$

while (6.9) can be similarly written as

$$|\Lambda_{\mathbf{a},\mathbf{r}}(f,f,f,f)-\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f},\mathbf{f},\mathbf{f},\mathbf{f})|>\eta.$$

Applying the triangle inequality and the quadrilinearity of $\Lambda_{\mathbf{a},\mathbf{r}}$, we conclude that

$$|\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f}_0,\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3)|\gg\eta$$

for some random functions $\mathbf{f}_0,\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3$, each of which is either equal to 1, f, or $f-\mathbf{f}$, and with at least one of the functions $\mathbf{f}_0,\mathbf{f}_1,\mathbf{f}_2,\mathbf{f}_3$ equal to $f-\mathbf{f}$. For sake of concreteness we will assume that it is $\mathbf{f}_3$ that is equal to $f-\mathbf{f}$, thus

$$|\Lambda_{\mathbf{a},\mathbf{r}}(\mathbf{f}_0,\mathbf{f}_1,\mathbf{f}_2,f-\mathbf{f})|\gg\eta; \tag{8.7}$$

the other cases are treated similarly (with some changes to the numerical constants below) and are left to the interested reader.

We can write the left-hand side of (8.7) as

$$\left|\sum_{c\in C}\mathbb{P}(\mathbf{c}=c)\mathbb{E}(\mathbf{f}_0(\mathbf{a})\mathbf{f}_1(\mathbf{a}+\mathbf{r})\mathbf{f}_2(\mathbf{a}+2\mathbf{r})(f-\mathbf{f})(\mathbf{a}+3\mathbf{r})|\mathbf{c}=c)\right|.$$

Applying Lemma 2.2, we conclude that with probability $\gg\eta$, the variable $\mathbf{c}$ attains a value c for which we have the lower bound

$$|\mathbb{E}(\mathbf{f}_0(\mathbf{a})\mathbf{f}_1(\mathbf{a}+\mathbf{r})\mathbf{f}_2(\mathbf{a}+2\mathbf{r})(f-\mathbf{f})(\mathbf{a}+3\mathbf{r})|\mathbf{c}=c)|\gg\eta. \tag{8.8}$$

We now use a local version of the standard "generalized von Neumann theorem" argument (based on several applications of the Cauchy–Schwarz inequality) to obtain some local correlation of $f-f_c$ with a quadratic phase.

PROPOSITION 8.2. *Let the notation and hypotheses be as above. For each (a,c) in the essential range of $(\mathbf{a},\mathbf{c})$, there exists a natural number $k_{a,c}$ with*

$$1\leqslant k_{a,c}<\eta^{-C_3}, \tag{8.9}$$

a set $\tilde{S}_{a,c}\subset\mathbb{Z}/p\mathbb{Z}$ with $\tilde{S}_{a,c}\supset S_c$ and

$$|\tilde{S}_{a,c}|\leqslant|S_c|+\eta^{-C_2}, \tag{8.10}$$

and a locally quadratic function $\gamma_{n,a,c}:B(\tilde{S}_{a,c},\exp(-\eta^{-11C_4})\rho_c)\to\mathbb{R}/\mathbb{Z}$ for each $n\in\mathbb{Z}/p\mathbb{Z}$, such that

$$\begin{aligned}\operatorname{Re}\sum_{a,c\in\mathbb{Z}/p\mathbb{Z}}&\mathbb{P}(\mathbf{a}=a,\mathbf{c}=c)\\&\times\mathbb{E}((f-f_c)(a+6\mathbf{n}+6k_{a,c}\mathbf{m})e(-\gamma_{\mathbf{n},a,c}(\mathbf{m}))|\mathbf{a}=a,\mathbf{c}=c)\geqslant\eta^{C_2/10},\end{aligned} \tag{8.11}$$

where, after conditioning to the event $\mathbf{a}=a,\mathbf{c}=c$, the random variables $\mathbf{n}$ and $\mathbf{m}$ are drawn regularly and independently from the Bohr sets $B(S_c,\exp(-\eta^{-2C_4})\rho)$ and $B(\tilde{S}_{a,c},\exp(-\eta^{-12C_4})\rho_c)$ respectively.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Proof. Suppose for now that c obeys (8.8). From Definition 6.1, once we condition to the event $\mathbf{c} = c$, the random variables $\mathbf{a}, \mathbf{r}$ are independent and regularly drawn from $B(S_c, \rho_c/2)$ and $B(S_c, \exp(-\eta^{-C_4})\rho_c)$ respectively; from (6.4) we have the bounds

$$|S_c| \leqslant 8\eta^{-3C_2} \quad \text{and} \quad \rho_c \geqslant \exp(-\eta^{-2C_5}). \tag{8.12}$$

Also, the function $\mathbf{f}$ is now the deterministic function

$$f_c(a) := F_c(\Xi_c(a))$$

on the Bohr set $B(S_c, \rho_c)$, and $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2$ become deterministic functions $f_{0,c}$, $f_{1,c}$ and $f_{2,c}$ taking values in $[-2, 2]$. Thus we have

$$|\mathbb{E}(f_{0,c}(\mathbf{a}) f_{1,c}(\mathbf{a}+\mathbf{r}) f_{2,c}(\mathbf{a}+2\mathbf{r}) f_{3,c}(\mathbf{a}+3\mathbf{r})|\mathbf{c} = c)| \gg \eta$$

where $f_{3,c} := f - f_c$.

We now do a linear change of variable with conveniently chosen numerical coefficients that will facilitate a certain use of the Cauchy–Schwarz inequality to eliminate the bounded functions $f_{0,c}$, $f_{1,c}$, $f_{2,c}$, leaving only the function $f_{3,c}$. Continuing to condition on the event that $\mathbf{c} = c$, let $\mathbf{n}_1, \mathbf{n}_2$ and $\mathbf{n}_3$ be drawn regularly and independently from the Bohr sets $B(S_c, \exp(-\eta^{-2C_4})\rho_c)$, $B(S_c, \exp(-\eta^{-3C_4})\rho_c)$, and $B(S_c, \exp(-\eta^{-4C_4})\rho_c)$ respectively, independently of the previous random variables. We can use Lemma 4.4 (and (8.12)) to compare $\mathbf{a}$ with $\mathbf{a} - 3\mathbf{n}_2 - 12\mathbf{n}_3$, and conclude that

$$\begin{aligned}|\mathbb{E}(f_{0,c}(\mathbf{a} - 3\mathbf{n}_2 - 12\mathbf{n}_3) f_{1,c}(\mathbf{a}+\mathbf{r} - 3\mathbf{n}_2 - 12\mathbf{n}_3) f_{2,c}(\mathbf{a}+2\mathbf{r} - 3\mathbf{n}_2 - 12\mathbf{n}_3)\\ \times f_{3,c}(\mathbf{a}+3\mathbf{r} - 3\mathbf{n}_2 - 12\mathbf{n}_3)|\mathbf{c} = c)| \gg \eta.\end{aligned}$$

By another application of Lemma 4.4, we may compare $\mathbf{r}$ with $\mathbf{r} + 2\mathbf{n}_1 + 3\mathbf{n}_2 + 6\mathbf{n}_3$, and conclude that

$$\begin{aligned}|\mathbb{E}(f_{0,c}(\mathbf{a} - 3\mathbf{n}_2 - 12\mathbf{n}_3) f_{1,c}(\mathbf{a}+\mathbf{r} + 2\mathbf{n}_1 - 6\mathbf{n}_3) f_{2,c}(\mathbf{a}+2\mathbf{r} + 4\mathbf{n}_1 + 3\mathbf{n}_2)\\ \times f_{3,c}(\mathbf{a}+3\mathbf{r} + 6(\mathbf{n}_1+\mathbf{n}_2+\mathbf{n}_3))|\mathbf{c} = c)| \gg \eta.\end{aligned}$$

Finally, we use Lemma 4.4 to replace $\mathbf{a}$ by $\mathbf{a} - 3\mathbf{r}$, so that

$$\begin{aligned}|\mathbb{E}(f_{0,c}(\mathbf{a} - 3\mathbf{r} - 3\mathbf{n}_2 - 12\mathbf{n}_3) f_{1,c}(\mathbf{a} - 2\mathbf{r} + 2\mathbf{n}_1 - 6\mathbf{n}_3)\\ \times f_{2,c}(\mathbf{a} - \mathbf{r} + 4\mathbf{n}_1 + 3\mathbf{n}_2) f_{3,c}(\mathbf{a} + 6(\mathbf{n}_1+\mathbf{n}_2+\mathbf{n}_3))|\mathbf{c} = c)| \gg \eta.\end{aligned}$$

The purpose of this odd-seeming change of variables is that each of the functions $f_{0,c}$, $f_{1,c}$, $f_{2,c}$ now has an argument that involves only two of the three random variables $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$, while the argument of the key function $f_{3,c}$ depends on $\mathbf{n}_1$, $\mathbf{n}_2, \mathbf{n}_3$ only through their sum $\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3$.

One can achieve a similar effect for the other three choices $\mathbf{f}_0, \mathbf{f}_1, \mathbf{f}_2$ for key function by suitable adjustment to the constants above; we leave the details to the interested reader.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

By Lemma 2.2, we see that with probability $\gg \eta$ (conditioning on $\mathbf{c} = c$), the random variable $\mathbf{a}$ attains a value a such that

$$|\mathbb{E}(f_{0,c}(a - 3\mathbf{r} - 3\mathbf{n}_2 - 12\mathbf{n}_3) f_{1,c}(a - 2\mathbf{r} + 2\mathbf{n}_1 - 6\mathbf{n}_3) \times f_{2,c}(a - \mathbf{r} + 4\mathbf{n}_1 + 3\mathbf{n}_2) f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3))|\mathbf{a} = a, \mathbf{c} = c)| \gg \eta. \tag{8.13}$$

Let a be such that (8.13) holds. We can then find an $r \in \mathbb{Z}/p\mathbb{Z}$ (depending on a, c) such that

$$|\mathbb{E}(f_{0,c}(a - 3r - 3\mathbf{n}_2 - 12\mathbf{n}_3) f_{1,c}(a - 2r + 2\mathbf{n}_1 - 6\mathbf{n}_3) \times f_{2,c}(a - r + 4\mathbf{n}_1 + 3\mathbf{n}_2) f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3))|\mathbf{a} = a, \mathbf{c} = c)| \gg \eta.$$

We now suppress the additive structure on the first three arguments by rewriting the above bound as

$$|\mathbb{E}(f_{0,c,a}(\mathbf{n}_2, \mathbf{n}_3) f_{1,c,a}(\mathbf{n}_1, \mathbf{n}_3) \times f_{2,c,a}(\mathbf{n}_1, \mathbf{n}_2) f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3))|\mathbf{c} = c)| \gg \eta,$$

where $f_{0,c,a}, f_{1,c,a}, f_{2,c,a} : \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to [-2, 2]$ are bounded functions whose exact form

$$\begin{aligned} f_{0,c,a}(n_2, n_3) &:= f_{0,c}(a - 3r - 3n_2 - 12n_3), \\ f_{1,c,a}(n_1, n_3) &:= f_{1,c}(a - 2r + 2n_1 - 6n_3), \\ f_{2,c,a}(n_1, n_2) &:= f_{2,c}(a - r + 4n_1 + 3n_2) \end{aligned}$$

will not be relevant in the arguments that follow.

We can eliminate the factor $f_{0,c,a}$ using Lemma 2.1 to conclude that

$$|\mathbb{E}(f_{1,c,a}(\mathbf{n}_1, \mathbf{n}_3) f_{1,c,a}(\mathbf{n}_1', \mathbf{n}_3) f_{2,c,a}(\mathbf{n}_1, \mathbf{n}_2) f_{2,c,a}(\mathbf{n}_1', \mathbf{n}_2) \times f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3)) f_{3,c}(a + 6(\mathbf{n}_1' + \mathbf{n}_2 + \mathbf{n}_3))|\mathbf{a} = a, \mathbf{c} = c)| \gg \eta^2$$

where $\mathbf{n}_1'$ is an independent copy of $\mathbf{n}_1$ (and also independent of $\mathbf{n}_2, \mathbf{n}_3$) on the event $\mathbf{a} = a, \mathbf{c} = c$. We can similarly apply Lemma 2.1 to eliminate the $f_{1,c,a}(\mathbf{n}_1, \mathbf{n}_3) f_{1,c,a}(\mathbf{n}_1', \mathbf{n}_3)$ variables to conclude that

$$\begin{aligned} |\mathbb{E}(&f_{2,c,a}(\mathbf{n}_1, \mathbf{n}_2) f_{2,c,a}(\mathbf{n}_1', \mathbf{n}_2) f_{2,c,a}(\mathbf{n}_1, \mathbf{n}_2') f_{2,c,a}(\mathbf{n}_1', \mathbf{n}_2') \\ &\times f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3)) f_{3,c}(a + 6(\mathbf{n}_1' + \mathbf{n}_2 + \mathbf{n}_3)) \\ &\times f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2' + \mathbf{n}_3)) f_{3,c}(a + 6(\mathbf{n}_1' + \mathbf{n}_2' + \mathbf{n}_3))|\mathbf{a} = a, \mathbf{c} = c)| \gg \eta^4 \end{aligned}$$

and finally apply Lemma 2.1 to eliminate the $f_{2,c,a}$ terms and arrive at

$$\begin{aligned} |\mathbb{E}(&f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3)) f_{3,c}(a + 6(\mathbf{n}_1' + \mathbf{n}_2 + \mathbf{n}_3)) \\ &\times f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2' + \mathbf{n}_3)) f_{3,c}(a + 6(\mathbf{n}_1' + \mathbf{n}_2' + \mathbf{n}_3)) \\ &\times f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3')) f_{3,c}(a + 6(\mathbf{n}_1' + \mathbf{n}_2 + \mathbf{n}_3')) \\ &\times f_{3,c}(a + 6(\mathbf{n}_1 + \mathbf{n}_2' + \mathbf{n}_3')) f_{3,c}(a + 6(\mathbf{n}_1' + \mathbf{n}_2' + \mathbf{n}_3'))|\mathbf{a} = a, \mathbf{c} = c)| \gg \eta^8, \end{aligned}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

where $\mathbf{n}_2', \mathbf{n}_3'$ are independent copies of $\mathbf{n}_2, \mathbf{n}_3$ respectively on $\mathbf{a} = a, \mathbf{c} = c$, with $\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, \mathbf{n}_1', \mathbf{n}_2', \mathbf{n}_3'$ all independent relative to $\mathbf{a} = a, \mathbf{c} = c$.

We now apply Theorem 8.1, replacing η by a small multiple of η^8, and choosing $\rho_i := \exp(-\eta^{-(i+2)C_4})\rho$ for $i = 0, \dots, 10$, and using the bounds (8.12), (3.21) to justify the hypothesis (8.3). We conclude that for c obeying (8.8) and a obeying (8.13), we can find a natural number $k_{a,c}$ obeying (8.9), a set $\tilde{S}_{a,c}$ with $S_c \subset \tilde{S}_{a,c} \subset \mathbb{Z}/p\mathbb{Z}$ obeying (8.10), a locally quadratic function $\phi_{a,c} : B(\tilde{S}_{a,c}, \exp(-\eta^{-11C_4})\rho) \to \mathbb{R}/\mathbb{Z}$, and a function $\beta_{a,c} : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n} = n | \mathbf{a} = a, \mathbf{c} = c) \times |\mathbb{E}(f_3(a + 6n + 6k\mathbf{m})e(-\phi_{a,c}(\mathbf{m}) - \beta_{a,c}(n)\mathbf{m}) | \mathbf{a} = a, \mathbf{c} = c)| \gg \eta^{C_2/20}$$

if $\mathbf{n}, \mathbf{m}$ are drawn independently and regularly from $B(S_c, \exp(-\eta^{-2C_4})\rho_c)$ and $B(S_{a,c}, \exp(\eta^{-12C_4})\rho_c)$ respectively on the event $\mathbf{a} = a, \mathbf{c} = c$. Taking expectations in $\mathbf{a}$ (and choosing $S_{a,c} = S_c$, $\phi_{a,c} = 0$ and $\beta_{a,c} = 0$ if (8.8) or (8.13) is not satisfied), we conclude that

$$\sum_{n,a,c \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n} = n, \mathbf{a} = a, \mathbf{c} = c) \times |\mathbb{E}(f_3(a + 6n + 6k\mathbf{m})e(-\phi_{a,c}(\mathbf{m}) - \beta_{a,c}(n)\mathbf{m}) | \mathbf{a} = a, \mathbf{c} = c)| \geqslant \eta^{C_2/10}.$$

In particular, if we set $\gamma_{n,a,c}(m) := \phi_{a,c}(m) + \beta_{a,c}(n)m + \theta_{n,a,c}$ for a suitable phase $\theta_{n,a,c} \in \mathbb{R}/\mathbb{Z}$, then $\gamma_{n,a,c}$ is locally quadratic on $B(\tilde{S}_{a,c}, \exp(-\eta^{-11C_4})\rho)$ and

$$\mathrm{Re} \sum_{n,a,c \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n} = n, \mathbf{a} = a, \mathbf{c} = c) \times \mathbb{E}(f_3(a + 6n + 6k\mathbf{m})e(-\gamma_{n,a,c}(\mathbf{m})) | \mathbf{a} = a, \mathbf{c} = c)| \geqslant \eta^{C_2/10},$$

giving the claim. □

Let $\mathbf{n}, \mathbf{m}, k_{a,c}, \tilde{S}_{a,c}, \gamma_{n,a,c}$ be as in the above proposition. The conclusion (8.11) of Proposition 8.2 may be rewritten more compactly as

$$\mathrm{Re}\, \mathbb{E}((f - \mathbf{f})(\mathbf{a} + 6\mathbf{n} + 6k_{\mathbf{a},\mathbf{c}}\mathbf{m})e(-\gamma_{\mathbf{n},\mathbf{a},\mathbf{c}}(\mathbf{m}))) \geqslant \eta^{C_2/10}. \tag{8.14}$$

We now introduce the modified random function $\mathbf{f}' : \mathbb{Z}/p\mathbb{Z} \to [-2, 2]$ by the formula

$$\mathbf{f}'(l) := \mathbf{f}(l) + \eta^{C_2/2} \cos\left(2\pi \gamma_{\mathbf{n},\mathbf{a},\mathbf{c}}\left(\frac{l - \mathbf{a} - 6\mathbf{n}}{6k_{\mathbf{a},\mathbf{c}}}\right)\right), \tag{8.15}$$

where we extend $\gamma_{n,a,c}$ arbitrarily outside of $B(S_c', \exp(-\eta^{-11C_4})\rho_c)$. Note from (8.9) and (3.21) that we can divide by $6k_{\mathbf{a},\mathbf{c}}$ in $\mathbb{Z}/p\mathbb{Z}$ without difficulty.

We claim that the function $\mathbf{f}'$ is a little closer to f than $\mathbf{f}$ is.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

LEMMA 8.3. *We have*

$$\mathbb{E}|(f-\mathbf{f}')(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},\mathbf{c}}\mathbf{m})|^2 \leqslant \mathrm{Energy}(v)-\eta^{C_2}.$$

Proof. From (8.15) we have

$$\mathbf{f}'(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},\mathbf{c}}\mathbf{m}) = \mathbf{f}(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},\mathbf{c}}\mathbf{m}) + \eta^{C_2/2}\cos(2\pi\gamma_{\mathbf{n},\mathbf{a},\mathbf{c}}(\mathbf{m})),$$

and so

$$\begin{aligned}&|(f-\mathbf{f}')(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},\mathbf{c}}\mathbf{m})|^2\\ &\quad= |(f-\mathbf{f})(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},\mathbf{c}}\mathbf{m})|^2\\ &\qquad- 2\eta^{C_2/2}\mathbb{E}(f-\mathbf{f})(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},\mathbf{c}}\mathbf{m})\cos(2\pi\gamma_{\mathbf{n},\mathbf{a},\mathbf{c}}(\mathbf{m}))\\ &\qquad+ O(\eta^{C_2}).\end{aligned} \tag{8.16}$$

On the other hand, for any (a,c) in the essential range of $(\mathbf{a},\mathbf{c})$, we may use Lemma 4.4 to compare $\mathbf{n}$ with $\mathbf{n}+k_{\mathbf{a},\mathbf{c}}\mathbf{m}$, and conclude that

$$\begin{aligned}&\mathbb{E}(|(f-\mathbf{f})(a+6\mathbf{n}+6k_{a,c}\mathbf{m})|^2|\mathbf{a}=a,\mathbf{c}=c)\\ &\quad= \mathbb{E}(|(f-\mathbf{f})(a+6\mathbf{n})|^2|\mathbf{a}=a,\mathbf{c}=c)+O(\eta^{2C_3})\end{aligned}$$

(for example), and hence on taking expectations in $\mathbf{a}$

$$\begin{aligned}&\mathbb{E}(|(f-\mathbf{f})(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},c}\mathbf{m})|^2|\mathbf{c}=c)\\ &\quad= \mathbb{E}(|(f-\mathbf{f})(\mathbf{a}+6\mathbf{n})|^2|\mathbf{c}=c)+O(\eta^{2C_3}).\end{aligned}$$

Applying Lemma 4.4 again to compare $\mathbf{a}$ with $\mathbf{a}+6\mathbf{n}$, we conclude that

$$\mathbb{E}(|(f-\mathbf{f})(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},c}\mathbf{m})|^2|\mathbf{c}=c) = \mathbb{E}(|(f-\mathbf{f})(\mathbf{a})|^2|\mathbf{c}=c)+O(\eta^{2C_3}).$$

and hence on taking averages in $\mathbf{c}$

$$\mathbb{E}(|(f-\mathbf{f})(\mathbf{a}+6\mathbf{n}+6k_{\mathbf{a},c}\mathbf{m})|^2|\mathbf{c}=c) = \mathrm{Energy}(v)+O(\eta^{2C_3}). \tag{8.17}$$

Taking expectations in (8.16) and using (8.15), (8.17), we obtain the claim. □

There is a very minor technical issue that $\mathbf{f}'$ does not quite take values in $[-1,1]$, which is what is needed in the definition of an approximant. However, this is easily fixed by truncation, or more precisely by introducing the random function $\mathbf{f}'':\mathbb{Z}/p\mathbb{Z}\to[-1,1]$ defined by

$$\mathbf{f}''(l) := \min(\max(\mathbf{f}'(l),-1),1). \tag{8.18}$$

Since $f(l)$ already lies in $[-1,1]$, we see that $\mathbf{f}''(l)$ is at least as close to $f(l)$ as $\mathbf{f}'(l)$ is, thus we have the pointwise bound

$$|(f-\mathbf{f}'')(l)| \leqslant |(f-\mathbf{f}')(l)|$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

for any $l \in \mathbb{Z}/p\mathbb{Z}$. From the above lemma, we thus have

$$\mathbb{E}|(f - \mathbf{f}'')(\mathbf{a} + 6\mathbf{n} + 6k_{\mathbf{a},\mathbf{c}}\mathbf{m})|^2 \leqslant \mathrm{Energy}(v) - \eta^{C_2}. \tag{8.19}$$

We can now construct the new structured approximant

$$v' = (C', \mathbf{c}', (n'_{c'} + B(S'_{c'}, \rho'_{c'}))_{c' \in C'}, (G'_{c'})_{c' \in C'}, (F'_{c'})_{c' \in C'}, (\Xi'_{c'})_{c' \in C'})$$

as follows. We write the dilated torus G_c as $G_c = \prod_{i=1}^{\dim(G_c)} \mathbb{R}/\lambda_{i,c}\mathbb{Z}$.

(i) We set $C' := (\mathbb{Z}/p\mathbb{Z}) \times (\mathbb{Z}/p\mathbb{Z}) \times C$ and $\mathbf{c}' := (\mathbf{n}, \mathbf{a}, \mathbf{c})$.

(ii) If $c' = (n, a, c)$ is in C', we set

$$\begin{aligned} n'_{c'} &:= a + 6n, \\ S'_{c'} &:= (6k_{a,c})^{-1} \cdot \tilde{S}_{a,c}, \\ \rho'_{c'} &:= \exp(-\eta^{-12C_4})\rho_c, \\ G'_{c'} &:= \prod_{i=1}^{\dim(G_c)} (\mathbb{R}/100\lambda_{i,c}\mathbb{Z}) \times (\mathbb{R}/\mathbb{Z}). \end{aligned}$$

(iii) If $c' = (n, a, c)$ is in C', we define $F'_{c'} : G'_{c'} \to [-1, 1]$ to be the function

$$F'_{c'}(x, y) := \min\big(\max\big(F_c\big(\tfrac{1}{100} \cdot x\big) + \eta^{C_2/2}\cos(2\pi y), -1\big), 1\big)$$

for $x \in \prod_{i=1}^{\dim(G_c)}(\mathbb{R}/100\lambda_{i,c}\mathbb{Z})$ and $y \in \mathbb{R}/\mathbb{Z}$, where $x \mapsto \frac{1}{100} \cdot x$ is the obvious contraction map from $\prod_{i=1}^{\dim(G_c)}(\mathbb{R}/100\lambda_{i,c}\mathbb{Z})$ to $\prod_{i=1}^{\dim(G_c)}(\mathbb{R}/\lambda_{i,c}\mathbb{Z})$.

(iv) If $c' = (n, a, c)$ is in C', we define $\Xi'_{c'} : n'_{c'} + B(S'_{c'}, \rho'_{c'}) \to G'_{c'}$ by the formula

$$\Xi'_{c'}(l) := \left(100 \cdot \Xi_c(l), \gamma_{n,a,c}\left(\frac{l - a - 6n}{6k_{a,c}}\right)\right)$$

for $l \in n'_{c'} + B(S'_{c'}, \rho'_{c'})$ (which implies in particular that $(l - a - 6n)/6k_{a,c} \in B(\tilde{S}_{a,c}, \exp(-\eta^{-12C_4})\rho_c)$), where $x \mapsto 100 \cdot x$ is the obvious dilation map from $\prod_{i=1}^{\dim(G_c)}(\mathbb{R}/\lambda_{i,c}\mathbb{Z})$ to $\prod_{i=1}^{\dim(G_c)}(\mathbb{R}/100\lambda_{i,c}\mathbb{Z})$ (the inverse of the map $x \mapsto \frac{1}{100} \cdot x$ from part (iii)).

Since F_c is 1-Lipschitz, it is easy to see (thanks to the contraction by $\frac{1}{100}$) that $F'_{c'}$ is also 1-Lipschitz; similarly, as Ξ_c and $\gamma_{n,a,c}$ are locally quadratic on $n_c + B(S_c, \rho_c)$ and $B(\tilde{S}_{a,c}, \exp(\eta^{-11C_4})\rho_c)$ respectively, we see that $\Xi'_{c'}$ is also locally quadratic on $n'_{c'} + B(S'_{c'}, \rho'_{c'})$. From (8.15), (8.18), Definition 6.1, and the above constructions we see that

$$\mathbf{f}'' = \mathbf{f}_{v'}$$

and hence by (8.19)

$$\mathbb{E}|(f - \mathbf{f}_{v'})(\mathbf{a} + 6\mathbf{n} + 6k_{\mathbf{a},\mathbf{c}}\mathbf{m})|^2 \leqslant \mathrm{Energy}(v) - \eta^{C_2}.$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

From Definition 6.1 and the above constructions, we also see that $\mathbf{a}_{v'}$ has the same distribution as $\mathbf{a} + 6\mathbf{n} + 6k_{\mathbf{a},\mathbf{c}}\mathbf{m}$ (after conditioning to any positive probability event of the form $(\mathbf{n}, \mathbf{a}, \mathbf{c}) = (n, a, c)$), which gives the required energy decrement (6.15).

The bound (6.10) follows from (8.10), while from construction we clearly have $\dim(G'_{c'}) = \dim(G_c) + 1$, which gives (6.11). Since we have $\rho'_{c'} := \exp(-\eta^{-12C_4})\rho_c$, the bound (6.12) is clear; also, from (6.4) we have

$$\mathrm{vol}(G'_{c'}) = 100^{\dim(G'_{c'})}\,\mathrm{vol}(G_c) \leqslant \exp(O(\eta^{-2C_2}))\,\mathrm{vol}(G_c)$$

which gives (6.13). It remains to establish (6.14). By the definition of Err_1 (just before (6.1)) and the triangle inequality, it suffices to show that

$$|\mathbb{E} f(\mathbf{a}_{v'}) - \mathbb{E} f(\mathbf{a})| \leqslant \eta^{C_3}.$$

But as mentioned previously, $\mathbf{a}_{v'}$ has the same distribution as $\mathbf{a} + 6\mathbf{n} + 6k_{\mathbf{a},\mathbf{c}}\mathbf{m}$, and by using Lemma 4.4 as in the proof of Lemma 8.3 we have

$$\mathbb{E} f(\mathbf{a} + 6\mathbf{n} + 6k_{\mathbf{a},\mathbf{c}}\mathbf{m}) = \mathbb{E} f(\mathbf{a}) + O(\eta^{2C_3})$$

giving the claim. This completes the proof of Theorem 6.6, assuming the local inverse Gowers norm theorem (Theorem 8.1).

§9. *Local inverse U^3 theorem.* We now turn to the proof of Theorem 8.1, which is the last component needed in the proof of Theorem 1.1. Let us begin by recalling the setup of this theorem. We let S be a subset of $\mathbb{Z}/p\mathbb{Z}$, take a parameter η satisfying $0 < \eta < \frac{1}{2}$, and define the quantity K by (8.1), thus

$$\frac{1}{\eta}, \qquad |S| \leqslant K. \tag{9.1}$$

We suppose that

$$0 < \rho_{10} < \cdots < \rho_0 < \tfrac{1}{2}$$

are scales obeying the separation condition (8.2) and the largeness condition (8.3), and suppose that $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is a 1-bounded function obeying (8.4). Our task is to locate a natural number k with $k < \exp(K^{O(C_1)})$, a set S' with $S \subset S' \subset \mathbb{Z}/p\mathbb{Z}$ obeying (8.5), a locally quadratic phase $\phi : B(S', \rho_9) \to \mathbb{R}/\mathbb{Z}$, and a function $\beta : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ obeying (8.6). We will initially work at the scale ρ_0, but retreat to smaller scales as the argument progresses (mainly to ensure that the error terms in Lemma 4.4 are negligible), until we are working at the final scales ρ_9 and ρ_{10}. Let us comment once more that the intermediate scales $\rho_3, \ldots, \rho_8$ play no role in the actual statement of Theorem 8.1.

In this section, all sums will be over $\mathbb{Z}/p\mathbb{Z}$ unless otherwise stated.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

9.1. *First step: associate a frequency $\xi(n_2)$ to each derivative of f.* We now begin the (lengthy) proof of this theorem, which broadly follows the same inverse U^3 strategy in previous literature [**11**, **14**], but localized to a Bohr set, the key aim being to reduce the dependence of constants on the rank or radius of this Bohr set as much as possible.

The first step is to use the local inverse U^2 theorem (Theorem 4.12) to associate a frequency $\xi(n_2) \in \mathbb{Z}/p\mathbb{Z}$ to many "derivatives" $x \mapsto f(x+n_2)\overline{f(x)}$ of f.

THEOREM 9.2. *Let the notation and hypotheses be as in Theorem 8.1. Then there exists a set $\Omega \subset B(S, 2\rho_2)$ obeying the largeness condition*

$$\mathbb{P}(\mathbf{h}_2 - \mathbf{h}_2' \in \Omega) \geqslant \eta/4 \tag{9.2}$$

when $\mathbf{h}_2, \mathbf{h}_2'$ are drawn independently and regularly from $B(S, \rho_2)$, and a function $\xi : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_0 \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n}_0 = n_0)|\mathbb{E} f(n_0 + \mathbf{n}_1 + n_2)\overline{f}(n_0 + \mathbf{n}_1)e_p(-\xi(n_2)\mathbf{n}_1)|^2 \geqslant \frac{\eta}{8} 1_\Omega(n_2) \tag{9.3}$$

for all $n_2 \in \mathbb{Z}/p\mathbb{Z}$, and $\mathbf{n}_0, \mathbf{n}_1$ are drawn independently and regularly from $B(S, \rho_0)$, $B(S, \rho_1)$ respectively.

Proof. For each $n_2 \in \mathbb{Z}/p\mathbb{Z}$, let $f_{n_2} : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ denote the 1-bounded function

$$f_{n_2}(n) := f(n + n_2)\overline{f}(n).$$

Then we may rewrite the left-hand side of (8.4) as

$$\begin{aligned}|\mathbb{E} f_{\mathbf{h}_2 - \mathbf{h}_2'}(\mathbf{h}_0 + \mathbf{h}_2 + \mathbf{h}_1)\overline{f_{\mathbf{h}_2 - \mathbf{h}_2'}}(\mathbf{h}_0 + \mathbf{h}_2 + \mathbf{h}_1') \\ \times \overline{f_{\mathbf{h}_2 - \mathbf{h}_2'}}(\mathbf{h}_0' + \mathbf{h}_2 + \mathbf{h}_1) f_{\mathbf{h}_2 - \mathbf{h}_2'}(\mathbf{h}_0' + \mathbf{h}_2 + \mathbf{h}_1')|.\end{aligned}$$

By Lemma 4.4 and (8.2), the random variables $\mathbf{h}_0, \mathbf{h}_0'$ differ in total variation from $\mathbf{h}_0 + \mathbf{h}_2, \mathbf{h}_0' + \mathbf{h}_2$ respectively by at most $\eta/4$ (for example). We conclude that

$$|\mathbb{E} f_{\mathbf{h}_2 - \mathbf{h}_2'}(\mathbf{h}_0 + \mathbf{h}_1)\overline{f_{\mathbf{h}_2 - \mathbf{h}_2', 0}}(\mathbf{h}_0 + \mathbf{h}_1')\overline{f_{\mathbf{h}_2 - \mathbf{h}_2'}}(\mathbf{h}_0' + \mathbf{h}_1) f_{\mathbf{h}_2 - \mathbf{h}_2'}(\mathbf{h}_0' + \mathbf{h}_1')| \geqslant \eta/2.$$

By the triangle inequality, the left-hand side is at most

$$\sum_h \mathbb{P}(\mathbf{h}_2 - \mathbf{h}_2' = h)|\mathbb{E} f_h(\mathbf{h}_0 + \mathbf{h}_1)\overline{f_h}(\mathbf{h}_0 + \mathbf{h}_1')\overline{f_h}(\mathbf{h}_0' + \mathbf{h}_1) f_h(\mathbf{h}_0' + \mathbf{h}_1')|.$$

The inner expectation is bounded by 1. Applying Lemma 2.2 (with $\mathbf{a} = \mathbf{h}_2 - \mathbf{h}_2'$), we conclude that there is a set $\Omega \subset \mathbb{Z}/p\mathbb{Z}$ obeying (9.2) such that

$$|\mathbb{E} f_{n_2}(\mathbf{h}_0 + \mathbf{h}_1)\overline{f_{n_2}}(\mathbf{h}_0 + \mathbf{h}_1')\overline{f_{n_2}}(\mathbf{h}_0' + \mathbf{h}_1) f_{n_2}(\mathbf{h}_0' + \mathbf{h}_1')| \geqslant \eta/4$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

for all $n_2 \in \Omega$. Applying Theorem 4.12, we see that for each $n_2 \in \Omega$, there exists $\xi(n_2) \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_0 \in \mathbb{Z}/p\mathbb{Z}} \mathbb{P}(\mathbf{n} = n_0)|\mathbb{E} f_{n_2}(n_0 + \mathbf{n}_1) e_p(-\xi(n_2)\mathbf{n}_1)|^2 \geqslant \eta/8.$$

For $n_2 \notin \Omega$, we set $\xi(n_2)$ arbitrarily (e.g. to zero). The claim follows. □

9.3. *Second step: ξ is approximately linear 1% of the time.* The next step, following Gowers [11], is to obtain some approximate linearity control on the function $\xi : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. Define an *additive quadruple* to be a quadruplet $\vec{a} = (a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)}) \in (\mathbb{Z}/p\mathbb{Z})^4$ such that

$$a_{(1)} + a_{(2)} = a_{(3)} + a_{(4)}, \tag{9.4}$$

and let $\mathrm{Q} \subset (\mathbb{Z}/p\mathbb{Z})^4$ denote the space of all additive quadruples. We call an additive quadruple $(a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)}) \in \mathrm{Q}$ *bad* if

$$\|\xi(a_{(1)}) + \xi(a_{(2)}) - \xi(a_{(3)}) - \xi(a_{(4)})\|_S > \frac{K^{C_1}}{\rho_1}, \tag{9.5}$$

where the word norm $\|\|_S$ was defined in Definition 4.5. Let $\mathrm{BQ} \subset \mathrm{Q}$ denote the space of all bad additive quadruples.

THEOREM 9.4. *Let the notation and hypotheses be as in Theorem 8.1, and let Ω and $\xi : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ be as in Theorem 9.2. If $\mathbf{h}_2, \mathbf{h}_2', \mathbf{k}_2, \mathbf{k}_2'$ are drawn independently and regularly from $B(S, \rho_2)$, then with probability $\gg \eta^{O(1)}$, one has*

$$(\mathbf{h}_2 - \mathbf{h}_2', \mathbf{k}_2 - \mathbf{k}_2', \mathbf{k}_2 - \mathbf{h}_2', \mathbf{h}_2 - \mathbf{k}_2') \in \Omega^4 \cap (\mathrm{Q} \setminus \mathrm{BQ}). \tag{9.6}$$

Proof. Let $\mathbf{n}_0, \mathbf{n}_1$ be drawn independently and regularly from the Bohr sets $B(S, \rho_0)$, $B(S, \rho_1)$ respectively. From (9.3) we have

$$\sum_{n_0} \mathbb{P}(\mathbf{n}_0 = n_0)|\mathbb{E} f(n_0 + \mathbf{n}_1 + n_2)\overline{f}(n_0 + \mathbf{n}_1) e_p(-\xi(n_2)\mathbf{n}_1)| \gg \eta$$

for any $n_2 \in \Omega$. Using (9.2), we conclude that

$$\sum_{n_0} \sum_{n_2 \in \Omega} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{h}_2 - \mathbf{h}_2' = n_2)|\mathbb{E} f(n_0 + \mathbf{n}_1 + n_2)\overline{f}(n_0 + \mathbf{n}_1) \times e_p(-\xi(n_2)\mathbf{n}_1)| \gg \eta^2,$$

where $\mathbf{h}_2, \mathbf{h}_2'$ are drawn independently and regularly from $B(S, \rho_2)$, and are independent of $\mathbf{n}_0, \mathbf{n}_1$. By the pigeonhole principle, one can thus find $n_0 \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_2 \in \Omega} \mathbb{P}(\mathbf{h}_2 - \mathbf{h}_2' = n_2)|\mathbb{E} f(n_0 + \mathbf{n}_1 + n_2)\overline{f}(n_0 + \mathbf{n}_1) e_p(-\xi(n_2)\mathbf{n}_1)| \gg \eta^2.$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We can rewrite the left-hand side as

$$\mathbb{E}F_{n_0}(\mathbf{h}_2-\mathbf{h}_2')f(n_0+\mathbf{n}_1+\mathbf{h}_2-\mathbf{h}_2')\overline{f}(n_0+\mathbf{n}_1)e_p(-\xi(\mathbf{h}_2-\mathbf{h}_2')\mathbf{n}_1)$$

for some 1-bounded function $F_{n_0}:\mathbb{Z}/p\mathbb{Z}\to\mathbb{C}$ depending on n_0. Using Lemma 4.4 to compare $\mathbf{n}_1$ with $\mathbf{n}_1+\mathbf{h}_2'$, we conclude that

$$|\mathbb{E}F_{n_0}(\mathbf{h}_2-\mathbf{h}_2')f(n_0+\mathbf{n}_1+\mathbf{h}_2)\overline{f}(n_0+\mathbf{n}_1+\mathbf{h}_2')e_p(-\xi(\mathbf{h}_2-\mathbf{h}_2')(\mathbf{n}_1+\mathbf{h}_2'))|\gg\eta^2.$$

We rearrange the left-hand side as

$$\sum_{n_1}\mathbb{P}(\mathbf{n}_1=n_1)\mathbb{E}f(n_0+n_1+\mathbf{h}_2)\overline{f}(n_0+n_1+\mathbf{h}_2')G_{n_0,n_1}(\mathbf{h}_2,\mathbf{h}_2')$$

where $G_{n_0,n_1}:\mathbb{Z}/p\mathbb{Z}\times\mathbb{Z}/p\mathbb{Z}\to\mathbb{C}$ is the 1-bounded function

$$G_{n_0,n_1}(h_2,h_2'):=F_{n_0}(h_2-h_2')e_p(-\xi(h_2-h_2')(n_1+h_2')).\tag{9.7}$$

By Hölder's inequality, we conclude that

$$\sum_{n_1}\mathbb{P}(\mathbf{n}_1=n_1)|\mathbb{E}f(n_0+n_1+\mathbf{h}_2)\overline{f}(n_0+n_1+\mathbf{h}_2')G_{n_0,n_1}(\mathbf{h}_2,\mathbf{h}_2')|^4\gg\eta^{O(1)}.$$

From this point onward we cease to keep careful track of powers of η. On the other hand, by using two applications of Lemma 2.1 to eliminate the 1-bounded functions f, we have

$$\begin{aligned}&|\mathbb{E}f(n_0+n_1+\mathbf{h}_2)\overline{f}(n_0+n_1+\mathbf{h}_2')G_{n_0,n_1}(\mathbf{h}_2,\mathbf{h}_2')|^4\\&\quad\leqslant\mathbb{E}G_{n_0,n_1}(\mathbf{h}_2,\mathbf{h}_2')\overline{G_{n_0,n_1}}(\mathbf{h}_2,\mathbf{k}_2')\overline{G_{n_0,n_1}}(\mathbf{k}_2,\mathbf{h}_2')G_{n_0,n_1}(\mathbf{k}_2,\mathbf{k}_2')\end{aligned}$$

where $(\mathbf{k}_2,\mathbf{k}_2')$ is an independent copy of $(\mathbf{h}_2,\mathbf{h}_2')$. We thus have

$$\mathbb{E}G_{n_0,\mathbf{n}_1}(\mathbf{h}_2,\mathbf{h}_2')\overline{G_{n_0,\mathbf{n}_1}}(\mathbf{h}_2,\mathbf{k}_2')\overline{G_{n_0,\mathbf{n}_1}}(\mathbf{k}_2,\mathbf{h}_2')G_{n_0,\mathbf{n}_1}(\mathbf{k}_2,\mathbf{k}_2')\gg\eta^{O(1)}$$

which by the triangle inequality and (9.7) gives

$$\begin{aligned}&\sum_{h_2,k_2,h_2',k_2'}1_{h_2-h_2',k_2-k_2',k_2-h_2',h_2-k_2'\in\Omega}\mathbb{P}(\mathbf{h}_2=h_2;\mathbf{k}_2=k_2;\mathbf{h}_2'=h_2';\mathbf{k}_2'=k_2')\\&\quad\times|\mathbb{E}e_p(-(\xi(h_2-h_2')+\xi(k_2-k_2')-\xi(k_2-h_2')-\xi(h_2-k_2'))\mathbf{n}_1)|\\&\quad\gg\eta^{O(1)}.\end{aligned}$$

By Lemma 2.2, we conclude that with probability $\gg\eta^{O(1)}$, the tuple $(\mathbf{h}_2,\mathbf{k}_2,\mathbf{h}_2',\mathbf{k}_2')$ attains a value (h_2,k_2,h_2',k_2') for which

$$h_2-h_2',k_2-k_2',h_2-k_2',k_2-h_2'\in\Omega$$

and

$$|\mathbb{E}e_p(-(\xi(h_2-h_2')+\xi(k_2-k_2')-\xi(k_2-h_2')-\xi(h_2-k_2'))\mathbf{n}_1)|\gg\eta^{O(1)}\gg K^{-O(1)}\tag{9.8}$$

thanks to (9.1). Since $(h_2-h_2',k_2-k_2',h_2-k_2',k_2-h_2')$ is an additive quadruple, the claim now follows from Lemma 4.7, (8.2), and (9.1). □

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We localize this claim slightly, though for notational reasons we will not move from ρ_2 immediately to ρ_3 and beyond, but instead first work in some intermediate scales between ρ_2 and ρ_3. For any natural number j, define

$$\rho_{2,j} := \exp(-C_1 jK)\rho_2,$$

thus

$$\rho_2 = \rho_{2,0} > \rho_{2,1} > \cdots > \rho_{2,j} \geqslant \rho_3$$

if (for example) $j \leqslant K^{C_1^2}$.

It will be necessary to break the symmetry between the four components of an additive quadruple, by restricting the second component to a tiny Bohr set, the third component to a larger Bohr set, and the first and fourth components to an even larger Bohr set. More precisely, given an additive quadruple $\vec{a}_0 = (a_{(1),0}, a_{(2),0}, a_{(3),0}, a_{(4),0}) \in \mathrm{Q}$, a subset $S' \subset \mathbb{Z}/p\mathbb{Z}$, and radii $0 < r_2 \leqslant r_3 \leqslant r_4 \leqslant 1/2$, we say that a random additive quadruple $\vec{\mathbf{a}} = (\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a}_{(4)}) \in \mathrm{Q}$ is *centred at* $\vec{a}_0$ *with frequencies* S' *and scales* r_2, r_3, r_4 if $\mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a}_{(4)}$ are drawn independently and regularly from $a_{(2),0} + B(S', r_2)$, $a_{(2),0} + B(S', r_2)$, and $a_{(2),0} + B(S', r_2)$ respectively. Note that this property also describes the distribution of $\mathbf{a}_{(1)}$, since we have the constraint

$$\mathbf{a}_{(1)} = \mathbf{a}_{(3)} + \mathbf{a}_{(4)} - \mathbf{a}_{(2)}.$$

In practice, r_4 will be much larger than r_2, r_3, so (by Lemma 4.4) $\mathbf{a}_{(1)}$ will be approximately regularly drawn from $a_{(1),0} + B(S', r_4)$, but will be highly coupled to the other three components of the quadruple (in particular, it will stay close to $\mathbf{a}_{(4)}$). We thus see that for $i = 1, 2, 3, 4$, each $\mathbf{a}_{(i)}$ is either exactly or approximately drawn regularly from $a_{(i),0} + B(S', r_{l_i})$, where $l_i \in \{0, 1, 2\}$ is the quantity defined by the formulae

$$l_1 := 0; \qquad l_2 := 2; \qquad l_3 := 1; \qquad l_4 := 0. \tag{9.9}$$

COROLLARY 9.5. *Let the notation and hypotheses be as in Theorem 8.1, and let* Ω *and* ξ *be as in Theorem 9.2. Then there exists a random additive quadruple* $\vec{\mathbf{a}} \in \mathrm{Q}$ *centred at some quadruple* $\vec{a}_0 \in \mathrm{Q}$ *with frequencies* S *and scales* $\rho_{2,2}$, $\rho_{2,1}$, $\rho_{2,0}$, *such that* $\vec{\mathbf{a}} \in \Omega^4 \cap (\mathrm{Q} \setminus \mathrm{BQ})$ *with probability* $\gg \eta^{O(1)}$.

Proof. Let $\mathbf{h}_2, \mathbf{k}_2, \mathbf{h}_2', \mathbf{k}_2', \mathbf{n}_{2,1}, \mathbf{n}_{2,2}$ be drawn independently and regularly from $B(S, \rho_{2,0})$, $B(S, \rho_{2,0})$, $B(S, \rho_{2,0})$, $B(S, \rho_{2,0})$, $B(S, \rho_{2,1})$ and $B(S, \rho_{2,2})$ respectively. From Theorem 9.4, we have

$$(\mathbf{h}_2 - \mathbf{h}_2', \mathbf{k}_2 - \mathbf{k}_2', \mathbf{h}_2 - \mathbf{k}_2', \mathbf{k}_2 - \mathbf{h}_2') \in \Omega^4 \cap (\mathrm{Q} \setminus \mathrm{BQ})$$

with probability $\gg \eta^{O(1)}$. Using Lemma 4.4, we may replace $\mathbf{k}_2'$ by $\mathbf{k}_2' - \mathbf{n}_{2,2}$, and similarly replace $\mathbf{h}_2$ by $\mathbf{h}_2 + \mathbf{n}_{2,1} - \mathbf{n}_{2,2}$, to conclude that

$$(\mathbf{h}_2 - \mathbf{h}_2' + \mathbf{n}_{2,1}, \mathbf{k}_2 - \mathbf{k}_2' + \mathbf{n}_{2,2}, \mathbf{h}_2 - \mathbf{k}_2' + \mathbf{n}_{2,1}, \mathbf{k}_2 - \mathbf{h}_2') \in \Omega^4 \cap (\mathrm{Q} \setminus \mathrm{BQ})$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

with probability $\gg \eta^{O(1)}$. By the pigeonhole principle, we may thus find k_2, k_2', $h_2 \in \mathbb{Z}/p\mathbb{Z}$ such that

$$(h_2 - \mathbf{h}_2' + \mathbf{n}_{2,1}, k_2 - k_2' + \mathbf{n}_{2,2}, h_2 - k_2' + \mathbf{n}_{2,1}, k_2 - \mathbf{h}_2') \in \Omega^4 \cap (\mathrm{Q} \setminus \mathrm{BQ})$$

with probability $\gg \eta^{O(1)}$. The left-hand side is an additive quadruple centred at $(h_2, k_2 - k_2', h_2 - k_2', k_2)$ with frequencies S and scales $\rho_{2,2}$, $\rho_{2,1}$, $\rho_{2,0}$, and the claim follows. □

9.6. *Third step: ξ is approximately linear 99% of the time on a rough set.* The next general step in the standard inverse U^3 argument is to upgrade this weak additive structure, which is of a "1 percent" nature, to a more robust "99 percent" additive structure. There are two basic ways to proceed here. The first way is to invoke the Balog–Szemerédi–Gowers theorem [**1**, **11**], followed by standard sum set estimates including Freiman's theorem (see e.g. [**33**, Ch. 2]). It is likely that this approach will eventually work here, but these results need to be localized efficiently to Bohr sets, and also to allow for the fact that $\xi(a_{(1)})+\xi(a_{(2)})-\xi(a_{(3)})-\xi(a_{(4)})$ no longer vanishes, but instead has controlled word norm. This would require reworking of large portions of the standard additive combinatorics literature. We have thus elected instead to follow the second approach, also due to Gowers [**12**], in which a certain probabilistic argument is used to "purify" a 1 percent additive map to a 99 percent additive map, albeit on a set that has no particular structure itself. To deal with this set we will use a more recent innovation, namely a variant[4] of the arithmetic regularity lemma [**13**], [**18**] to make the subsets of $\mathbb{Z}/p\mathbb{Z}$ on which one has good control of ξ suitably "pseudorandom" in the sense of Gowers.

We turn to the details. We first locate a reasonably large quadruple of sets $A_{(1)}, A_{(2)}, A_{(3)}, A_{(4)}$ on which ξ is "almost a Freiman homomorphism" in the sense that most quadruples falling inside $A_{(1)} \times A_{(2)} \times A_{(3)} \times A_{(4)}$ are somewhat good. We call an additive quadruple $(a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)}) \in \mathrm{Q}$ *very bad* if

$$\|\xi(a_{(1)}) + \xi(a_{(2)}) - \xi(a_{(3)}) - \xi(a_{(4)})\|_S > \frac{1}{\rho_3}, \tag{9.10}$$

and let $\mathrm{VBQ} \subset \mathrm{BQ}$ denote the space of all very bad additive quadruples.

THEOREM 9.7. *Let the notation and hypotheses be as in Theorem 8.1, and let Ω and ξ be as in Theorem 9.2. Let $\vec{a}$ be the random additive quadruple from Corollary 9.5. Then there exist sets $A_{(1)}, A_{(2)}, A_{(3)}, A_{(4)} \subset \Omega$ such that*

$$\mathbb{E}W(\vec{a}) \gg \eta^{C_1+O(1)}, \tag{9.11}$$

where $W : \mathrm{Q} \to \mathbb{R}$ is the weight function

$$W(\vec{a}) := 1_{A_{(1)}\times A_{(2)}\times A_{(3)}\times A_{(4)}}(\vec{\mathbf{a}})(1 - \eta^{-C_1/100}1_{\mathrm{VBQ}}(\vec{\mathbf{a}})). \tag{9.12}$$

[4] The actual arithmetic regularity lemma, which creates arithmetic regularity on almost all regions of space, has quantitative bounds of tower-exponential type, which are far too poor for our application; however we will only need to create a single neighbourhood in which arithmetic regularity exists, and this can be done with much more efficient quantitative bounds.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

The idea here is that W is a weight function that strongly penalizes very bad quadruples, and so Theorem 9.7 is asserting that "most" of the quadruples in $A_{(1)} \times A_{(2)} \times A_{(3)} \times A_{(4)}$ are not very bad.

Proof. We will construct the sets $A_{(i)}$ by the probabilistic method, adapting an argument from [12] in which the $A_{(i)}$ are created by applying a number of random linear "filters" to the graph of ξ to eliminate most of the additive quadruples that are not (almost) preserved by ξ.

We turn to the details. Let m be the integer

$$m := \left\lfloor \frac{\log \eta^{C_1}}{3 \log 100} \right\rfloor. \tag{9.13}$$

We then select jointly independent random variables $\mathbf{h}_j \in \mathbb{Z}/p\mathbb{Z}$ and $\boldsymbol{\lambda}_j \in \mathbb{Z}/p\mathbb{Z}$ for each for $j = 1, \dots, m$, by selecting each h_j regularly from $B(S, \rho_2)$, and selecting $\boldsymbol{\lambda}_j$ uniformly at random from $\mathbb{Z}/p\mathbb{Z}$; we also choose these random variables to be independent of $\vec{\mathbf{a}}$. For $j = 1, \dots, m$, we then let $\boldsymbol{\Xi}_j : \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be the random map

$$\boldsymbol{\Xi}_j(n) := \xi(n)\mathbf{h}_j + \frac{\boldsymbol{\lambda}_j n}{p} \tag{9.14}$$

and then define the random sets

$$\mathbf{A}_{(i)} := \bigcap_{j=1}^{m} \mathbf{A}_{(i),j}$$

for $i = 1, 2, 3, 4$, where

$$\mathbf{A}_{(1),j} = \mathbf{A}_{(2),j} = \mathbf{A}_{(3),j} := \left\{ n \in \Omega : \|\boldsymbol{\Xi}_j(n)\|_{\mathbb{R}/\mathbb{Z}} \leqslant \tfrac{1}{200} \right\}$$

and

$$\mathbf{A}_{(4),j} := \left\{ n \in \Omega : \|\boldsymbol{\Xi}_j(n)\|_{\mathbb{R}/\mathbb{Z}} \leqslant \tfrac{1}{10} \right\}.$$

We will show that

$$\mathbb{E} 1_{A_{(1)} \times A_{(2)} \times A_{(3)} \times A_{(4)}}(\vec{\mathbf{a}}) \gg \eta^{O(1)} 100^{-3m} \tag{9.15}$$

and

$$\mathbb{E} 1_{A_{(1)} \times A_{(2)} \times A_{(3)} \times A_{(4)}}(\vec{\mathbf{a}}) 1_{\mathrm{BQ}}(\vec{a}) \ll 2^{-m} \times 100^{-3m} \tag{9.16}$$

which will give the claim thanks to (9.13) and (9.12), if C_1 is large enough.

We first show (9.15). By Corollary 9.5 and linearity of expectation, it suffices to show that

$$\mathbb{P}(a_{(i)} \in \mathbf{A}_{(i)} \text{ for } i = 1, 2, 3, 4) \gg 100^{-3m} \tag{9.17}$$

whenever $(a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)})$ lies in $\Omega^4 \cap (\mathrm{Q} \setminus \mathrm{BQ})$. Actually, we will only show the weaker assertion that (9.17) holds for all but at most $O(m^{O(1)} p^2)$ of the available additive quadruples $(a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)})$; this still suffices, since

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

by (4.3), (9.1) each exceptional additive quadruple is attained with probability $O(1/\rho_3^{O(K)}p^3)$, and the additional factor of p will dominate all the losses in m, K, ρ_3 thanks to (8.3), (9.13).

Fix an additive quadruple $\vec{a} = (a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)})$ in $\Omega^4 \cap (\mathrm{Q} \setminus \mathrm{BQ})$. The left-hand side of (9.17) factors as

$$\prod_{j=1}^{m} \mathbb{P}(a_{(i)} \in \mathbf{A}_{(i)} \text{ for } i = 1, 2, 3, 4) \tag{9.18}$$

so it will suffice to show that for each $j = 1, \dots, m$, one has

$$\mathbb{P}(a_{(i)} \in \mathbf{A}_{(i),j} \text{ for } i = 1, 2, 3, 4) \geqslant 100^{-3} - O\left(\frac{1}{m}\right)$$

for all but $O(m^{O(1)}p^2)$ quadruples $(a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)}) \in \mathrm{Q} \setminus \mathrm{BQ}$. Note however that from (9.14) we have

$$\begin{aligned}&\boldsymbol{\Xi}_j(a_{(1)}) + \boldsymbol{\Xi}_j(a_{(2)}) - \boldsymbol{\Xi}_j(a_{(3)}) - \boldsymbol{\Xi}_j(a_{(4)})\\ &\quad = (\xi(a_{(1)}) + \xi(a_{(2)}) - \xi(a_{(3)}) - \xi(a_{(4)}))\mathbf{h}_j\end{aligned}$$

and hence by the hypothesis $(a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)}) \in \mathrm{Q} \setminus \mathrm{BQ}$ and the range of $\mathbf{h}_j$ we have

$$\left\| \frac{\boldsymbol{\Xi}_j(a_{(1)}) + \boldsymbol{\Xi}_j(a_{(2)}) - \boldsymbol{\Xi}_j(a_{(3)}) - \boldsymbol{\Xi}_j(a_{(4)})}{p} \right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{1}{100}$$

(for example). In particular, we see from the triangle inequality that the claim $a_{(4)} \in \mathbf{A}_{(4),j}$ is implied by the claims $a_{(i)} \in \mathbf{A}_{(i),j}$ for $i = 1, 2, 3$. Thus it suffices to show that

$$\mathbb{P}(a_{(i)} \in \mathbf{A}_{(i),j} \text{ for } i = 1, 2, 3) \geqslant 100^{-3} - O\left(\frac{1}{m}\right)$$

for all but $O(m^{O(1)}p^2)$ triples $(a_{(1)}, a_{(2)}, a_{(3)}) \in (\mathbb{Z}/p\mathbb{Z})^3$, noting that $a_{(4)}$ is determined by $a_{(1)}, a_{(2)}, a_{(3)}$. We can write the left-hand side as

$$\mathbb{P}\left(\frac{(\xi(a_{(1)}), \xi(a_{(2)}), \xi(a_{(3)}))\mathbf{h}_j + (a_{(1)}, a_{(2)}, a_{(3)})\boldsymbol{\lambda}_j}{p} \in [-1/200, 1/200]^3 \right),$$

where we view the interval $[-1/200, 1/200]$ as a subset of $\mathbb{R}/\mathbb{Z}$. Thus it will suffice to show the equidistribution property

$$\inf_{x \in (\mathbb{R}/\mathbb{Z})^3} \mathbb{P}\left(\frac{(a_{(1)}, a_{(2)}, a_{(3)})\boldsymbol{\lambda}_j}{p} \in x + [-1/200, 1/200]^3 \right) \geqslant 100^{-3} - O\left(\frac{1}{m}\right).$$

Let $\psi : (\mathbb{R}/\mathbb{Z})^3 \to [0, 1]$ be a Lipschitz cutoff supported on $[-1/20, 1/20]^3$ that equals one on $[-1/200 + 1/m, 1/200 - 1/m]^3$ and has Lipschitz constant $O(m)$. Then we may lower bound the left-hand side by

$$\inf_{x \in (\mathbb{R}/\mathbb{Z})^3} \mathbb{E}_{\lambda \in \mathbb{Z}/p\mathbb{Z}} \psi\left(\frac{(a_{(1)}, a_{(2)}, a_{(3)})\lambda}{p} - x \right). \tag{9.19}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

By standard Fourier expansion (see e.g. [17, Lemma A.9]), we may write

$$\psi(y) = \sum_{k\in\mathbb{Z}^3:k=O(m^{O(1)})} c_k e(k\cdot y) + O\left(\frac{1}{m}\right)$$

for all $y \in (\mathbb{R}/\mathbb{Z})^3$ and some bounded Fourier coefficients $c_k = O(1)$; integrating in x, we see in particular that $c_0 = 10^{-3} + O(1/m)$. We may thus write (9.19) as

$$10^{-3} + O\left(\frac{1}{m}\right) + O\left(\sum_{k\in\mathbb{Z}^3\setminus\{0\}:k=O(m^{O(1)})} |\mathbb{E}_{\lambda\in\mathbb{Z}/p\mathbb{Z}} e_p(k\cdot(a_{(1)},a_{(2)},a_{(3)})\lambda)|\right)$$

which gives the desired claim as long as there are no relations of the form

$$k\cdot(a_{(1)},a_{(2)},a_{(3)}) = 0$$

for some non-zero $k \in \mathbb{Z}^3$ with $k = O(m^{O(1)})$. But it is easy to see that the number of $(a_{(1)},a_{(2)},a_{(3)})$ with such a relation is $O(m^{O(1)}p^2)$, thus concluding the proof of (9.15).

Now we show (9.16). By linearity of expectation as before, it suffices to show that

$$\mathbb{P}(a_{(i)} \in \mathbf{A}_{(i)} \text{ for } i = 1,2,3,4) \ll 2^{-m} \times 100^{-3m}$$

for all but $O(m^{O(1)}p^2)$ of the quadruples $(a_{(1)},a_{(2)},a_{(3)},a_{(4)})$ in VBQ. Using the factorization (9.18), it suffices to show that for each $j = 1,\dots,m$, one has

$$\mathbb{P}(a_{(i)} \in \mathbf{A}_{(i),j} \text{ for } i = 1,2,3,4) \leqslant 2^{-1} \times 100^{-3} + O\left(\frac{1}{m}\right)$$

for all but $O(m^{O(1)}p^2)$ of the quadruples $(a_{(1)},a_{(2)},a_{(3)},a_{(4)})$ in VBQ.

The left-hand side may be written as

$$\mathbb{P}\left(\frac{(\xi(a_{(1)}),\dots,\xi(a_{(4)}))\mathbf{h}_j}{p} + \vec{a}\boldsymbol{\lambda}_j \in [-1/200,1/200]^3 \times [-1/10,1/10]\right),$$

which we bound above by

$$\mathbb{P}\left((\xi(a_{(1)}),\xi(a_{(2)}),\xi(a_{(3)}))\mathbf{h}_j + (a_{(1)},a_{(2)},a_{(3)})\boldsymbol{\lambda}_j \in [-1/200,1/200]^3,\ \left\|\frac{\sigma\mathbf{h}_j}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{1}{8}\right),$$

where $\sigma := \xi(a_{(1)}) + \xi(a_{(2)}) - \xi(a_{(3)}) - \xi(a_{(4)})$. By arguing as in the proof of (9.15), we see that after deleting $O(m^{O(1)}p^2)$ exceptional tuples, one has

$$\sup_{x\in(\mathbb{R}/\mathbb{Z})^3} \mathbb{P}((a_{(1)},a_{(2)},a_{(3)})\boldsymbol{\lambda}_j \in x + [-1/200,1/200]^3) \leqslant 100^{-3} + O\left(\frac{1}{m}\right),$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

so by Fubini's theorem and the independence of $\mathbf{h}_j$ and $\boldsymbol{\lambda}_j$ it will suffice to show that

$$\mathbb{P}\left(\left\|\frac{\sigma \mathbf{h}_j}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{1}{8}\right) \leqslant 2^{-1} + O\left(\frac{1}{m}\right).$$

However, by Lemma 4.6 and the hypothesis $(a_{(1)}, a_{(2)}, a_{(3)}, a_{(4)}) \in \mathrm{VBQ}$ we may find $h \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\left\|\frac{\sigma h}{p}\right\|_{\mathbb{R}/\mathbb{Z}} > K^{-O(1)}\|h\|_{S^\perp}\rho_3.$$

In particular, h is non-zero. By repeatedly doubling h until $\|\eta h/p\|_{\mathbb{R}/\mathbb{Z}}$ exceeds $\frac{1}{4}$, we may also assume that

$$\frac{1}{2} \geqslant \left\|\frac{\eta h}{p}\right\|_{\mathbb{R}/\mathbb{Z}} > \frac{1}{4}$$

and thus

$$\|h\|_{S^\perp} \ll K^{O(1)}\rho_3.$$

From Lemma 4.4 we conclude that

$$\mathbb{P}\left(\left\|\frac{\eta(\mathbf{h}_j + h)}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{1}{8}\right) = \mathbb{P}\left(\left\|\frac{\eta \mathbf{h}_j}{p}\right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{1}{8}\right) + O\left(\frac{1}{m}\right).$$

But from the triangle inequality we see that the events $\|\eta(\mathbf{h}_j + h)/p\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{1}{8}$, $\|\eta \mathbf{h}_j/p\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{1}{8}$ are disjoint. The claim follows. □

9.8. *Fourth step: the rough set is pseudorandom in a Bohr set.* The sets $A_{(i)}$ provided by Theorem 9.7 are currently rather arbitrary. In particular we have no control on the pseudorandomness of these sets (as measured by local Gowers U^2 norms) in the Bohr sets we are working with. However, it is possible to use an "energy decrement argument" to pass to smaller[5] Bohr sets in which the sets $A_{(i)}$ do enjoy good pseudorandomness properties, basically by converting any large Fourier coefficient of any of the $A_{(i)}$ in a Bohr set into a refinement of the Bohr sets (which add the frequency of the large Fourier coefficient to the frequency set S) on which the indicator function $1_{A_{(i)}}$ has smaller variance. Furthermore, it is possible to shrink the Bohr sets in this fashion without destroying the conclusion (9.11) of Theorem 9.7.

Here is a precise statement.

[5] This is somewhat analogous to the variants of the Szemerédi regularity lemma [**31**] in which one locates a single regular pair inside an arbitrary large random graph. In contrast to the full regularity lemma that strives to ensure that *almost all* pairs are regular, the "one regular pair" versions of the lemma enjoy significantly better quantitative bounds. In our current application, such good quantitative bounds are essential, so we cannot appeal to analogues of the regularity lemma such as the arithmetic regularity lemma of the first author [**13**].

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

THEOREM 9.9. *Let the notation and hypotheses be as in Theorem 8.1, and let Ω and ξ be as in Theorem 9.2. Let $A_{(1)}, A_{(2)}, A_{(3)}, A_{(4)}, W$ be as in Theorem 9.7. Then there exists a natural number j, $j \leqslant \eta^{-10^3 C_1}$, an additive quadruple $\vec{a}_1 = (a_{(1),1}, a_{(2),1}, a_{(3),1}, a_{(4),1}) \in \mathrm{Q}$, and a set S_1, $S \subset S_1 \subset \mathbb{Z}/p\mathbb{Z}$ with $|S_1| \leqslant |S| + j$, with the following properties.*

(i) (Few very bad quadruples) *We have*

$$\mathbb{E}W(\vec{\mathbf{a}}) \gg \eta^{C_1+O(1)}, \tag{9.20}$$

where $\vec{\mathbf{a}}$ is a random additive quadruple centred at $\vec{a}_1$ with frequencies S_1 and scales $\rho_{2,j+2}$, $\rho_{2,j+1}$, and $\rho_{2,j}$.

(ii) (Local Fourier pseudorandomness) *For each $i = 1, 2, 3, 4$, we have*

$$\begin{aligned}|\mathbb{E} f_i(\mathbf{a}_{(i)} + \mathbf{h}_0 + \mathbf{h}_1) f_i(\mathbf{a}_{(i)} + \mathbf{h}_0 + \mathbf{h}_1') f_i(\mathbf{a}_{(i)} + \mathbf{h}_0' + \mathbf{h}_1) \\ \times f_i(\mathbf{a}_{(i)} + \mathbf{h}_0' + \mathbf{h}_1')| \leqslant \eta^{100C_1},\end{aligned}$$

where $f_i : \mathbb{Z}/p\mathbb{Z} \to [-1, 1]$ denotes the balanced function

$$f_i(a_{(i)}) := 1_{A_{(i)}}(a_{(i)}) - \alpha_i, \tag{9.21}$$

α_i denotes the mean

$$\alpha_i := \mathbb{E}1_{A_{(i)}}(\mathbf{a}_{(i)}), \tag{9.22}$$

and where $\mathbf{a}_{(i)}$ and $\mathbf{h}_0, \mathbf{h}_0', \mathbf{h}_1, \mathbf{h}_1'$ are drawn independently and regularly from the Bohr sets $a_{(i),1} + B(S_1, \rho_{2,j+l_i})$ and $B(S_1, \rho_{2,j+10})$, $B(S_1, \rho_{2,j+10})$, $B(S_1, \rho_{2,j+11})$, $B(S_1, \rho_{2,j+11})$ respectively, with the quantity l_i given by (9.9).

Proof. We will formulate the "energy decrement" argument here as a "score maximization" argument. Define a 4-*neighbourhood* to be a tuple

$$N = (\vec{a}_1, j, S_1),$$

where $\vec{a}_1 \in \mathrm{Q}$ is an additive quadruple, j is a natural number between 0 and $\eta^{-10^3 C_1}$, and S_1 is a subset of $\mathbb{Z}/p\mathbb{Z}$ containing S with $|S_1| \leqslant |S| + j$; we refer to j as the *depth* of the 4-neighbourhood N. Given such a neighbourhood, we define the *score* $\mathrm{Score}(N)$ of the 4-neighbourhood to be the quantity

$$\mathrm{Score}(N) := \mathbb{E}W(\vec{\mathbf{a}}) - \eta^{2C_1} \sum_{i=1}^{4} \mathrm{E}_i(N) - \eta^{10^3 C_1} j, \tag{9.23}$$

where $\vec{\mathbf{a}} = (\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a}_{(4)})$ is a random additive quadruple centred at $\vec{a}_1$ with frequencies S_1 and scales $\rho_{2,j+2}$, $\rho_{2,j+1}$, $\rho_{2,j}$, and E_i is the energy-type quantity

$$\mathrm{E}_i(N) := \mathrm{Var}\, 1_{A_{(i)}}(\mathbf{a}_{(i)}). \tag{9.24}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

If we define N_0 to be the 4-neighbourhood

$$N_0 := (\vec{a}_0, 0, S),$$

then Theorem 9.7 tells us that

$$\mathrm{Score}(N_0) \gg \eta^{C_1+O(1)}. \tag{9.25}$$

We choose

$$N := (\vec{a}_1, j, S_1)$$

to be a 4-neighbourhood that comes within $\eta^{10^3C_1}$ (for example) of maximizing the adjusted score. Then we must have

$$\mathrm{Score}(N) \geqslant \mathrm{Score}(N_0) - \eta^{10^3C_1} \gg \eta^{C_1+O(1)}$$

which from (9.23) implies the bound (9.20), as well as the bound

$$j \leqslant \eta^{-10^3C_1} - 10^3$$

(for example). It will then suffice to show that property (ii) of the theorem holds.

It remains to show (ii). Let $i = 1, 2, 3, 4$, and write

$$\vec{a}_1 = (a_{(1),1}, a_{(2),1}, a_{(3),1}, a_{(4),1}).$$

Suppose for contradiction that

$$|\mathbb{E} f_i(\mathbf{a}_{(i)}+\mathbf{h}_0+\mathbf{h}_1) f_i(\mathbf{a}_{(i)}+\mathbf{h}_0+\mathbf{h}'_1) f_i(\mathbf{a}_{(i)}+\mathbf{h}'_0+\mathbf{h}_1) f_i(\mathbf{a}_{(i)}+\mathbf{h}'_0+\mathbf{h}'_1)| > \eta^{100C_1}, \tag{9.26}$$

where f_i is given by (9.21), and $\mathbf{a}_{(i)}, \mathbf{h}_0, \mathbf{h}'_0, \mathbf{h}_1, \mathbf{h}'_1$ are drawn independently and regularly from the Bohr sets $a_{(i),1} + B(S_1, \rho_{2,j+l_i})$, $B(S_1, \rho_{2,j+10})$, $B(S_1, \rho_{2,j+10})$, $B(S_1, \rho_{2,j+11})$, $B(S_1, \rho_{2,j+11})$, with l_i given by (9.9).

We will use (9.26) to construct a random 4-neighbourhood $\mathbf{N}$ of depth $j + 20$ obeying the estimates

$$\mathbb{E}W(\mathbf{N}) = W(N) + O(\eta^{10^3C_1}) \tag{9.27}$$

and

$$\mathbb{E}\,\mathrm{E}_{i'}(\mathbf{N}) \leqslant \mathrm{E}_{i'}(N) - \eta^{500C_1} 1_{i=i'} + O(\eta^{10^3C_1}) \tag{9.28}$$

for $i' = 1, 2, 3, 4$. If we have the estimates (9.27), (9.28), we conclude from (9.23) and linearity of expectation that

$$\mathbb{E}\,\mathrm{Score}(\mathbf{N}) > \mathrm{Score}(N) + \eta^{600C_1},$$

contradicting the near-maximality of $\mathrm{Score}(N)$.

It remains to construct $\mathbf{N}$ obeying (9.27), (9.28). We begin by noting that for each $a_{(i)} \in \mathbb{Z}/p\mathbb{Z}$, the Gowers uniformity-type quantity

$$\mathbb{E} f_i(a_{(i)} + \mathbf{h}_0 + \mathbf{h}_1) f_i(a_{(i)} + \mathbf{h}_0 + \mathbf{h}'_1) f_i(a_{(i)} + \mathbf{h}'_0 + \mathbf{h}_1) f_i(a_{(i)} + \mathbf{h}'_0 + \mathbf{h}'_1)$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

can be factored as

$$\sum_{h_0,h_0'} \mathbb{P}(\mathbf{h}_0 = h_0, \mathbf{h}_0' = h_0')|\mathbb{E} f_i(a_{(i)} + h_0 + \mathbf{h}_1) f_i(a_{(i)} + h_0' + \mathbf{h}_1)|^2$$

and thus takes values between 0 and 1. By (9.26) and Lemma 2.2, we may thus find a set $E \subset \mathbb{Z}/p\mathbb{Z}$ with

$$\mathbb{P}(\mathbf{a}_{(i)} \in E) \gg \eta^{100C_1}$$

such that

$$\mathbb{E} f_i(a_{(i)} + \mathbf{h}_0 + \mathbf{h}_1) f_i(a_{(i)} + \mathbf{h}_0 + \mathbf{h}_1') f_i(a_{(i)} + \mathbf{h}_0' + \mathbf{h}_1) f_i(a_{(i)} + \mathbf{h}_0' + \mathbf{h}_1') \gg \eta^{100C_1}$$

for all $a_{(i)} \in E$. Applying Theorem 4.12, we may thus find, for each $a_{(i)} \in E$, a frequency $\xi(a_{(i)}) \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_0} \mathbb{P}(\mathbf{n}_0 = n_0)\mathbb{E}|\mathbb{E} f_i(a_{(i)} + n_0 + \mathbf{n}_1) e_p(-\xi(a_{(i)})\mathbf{n}_1)|^2 \gg \eta^{100C_1},$$

where $\mathbf{n}_0, \mathbf{n}_1$ are drawn independently and regularly from $B(S_1, \rho_{2,j_*+10})$ and $B(S_1, \rho_{2,j_*+11})$ respectively, independently of the $\mathbf{a}_{(i)}$.

If we define $\xi(a_{(i)})$ arbitrarily for $a_{(i)} \notin E$ (e.g. setting $\xi(a_{(i)}) = 0$), we thus have

$$\sum_{n_0,a_{(i)}} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{a}_{(i)} = a_{(i)})\mathbb{E}|\mathbb{E}(f_i(a_{(i)} + n_0 + \mathbf{n}_1) e_p(-\xi(a_{(i)})\mathbf{n}_1))|^2 \gg \eta^{200C_1}.$$

In particular, there exists a 1-bounded function $g : \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ such that

$$|\mathbb{E} g(\mathbf{n}_0, \mathbf{a}_{(i)}) f_i(\mathbf{a}_{(i)} + \mathbf{n}_0 + \mathbf{n}_1) e_p(-\xi(\mathbf{a}_{(i)})\mathbf{n}_1)| \gg \eta^{200C_1}. \tag{9.29}$$

We now construct the random 4-neighbourhood $\mathbf{N}$ as follows. We first construct a random additive quadruple $\vec{\mathbf{k}} = (\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$ centred at the origin $(0, 0, 0, 0)$ with frequency set S_1 and scales $\rho_{2,j+10+l_2-l_i}$, $\rho_{2,j+10+l_3-l_i}$, $\rho_{2,j+10+l_4-l_i}$, and independent of all previous random variables. We then set

$$\mathbf{N} := (\vec{\mathbf{a}} + \vec{\mathbf{k}}, j + 20, S_1 \cup \{\xi(\mathbf{a}_{(i)})\}).$$

It is easy to verify that $\mathbf{N}$ is a (random) 4-neighbourhood.

We now verify (9.27). The left-hand side of (9.27) can be expanded as

$$\mathbb{E} W(\vec{\mathbf{a}} + \vec{\mathbf{k}} + \vec{\mathbf{h}}),$$

where, once $\vec{\mathbf{a}}$ and $\vec{\mathbf{k}}$ are chosen, the random additive quadruple $\vec{\mathbf{h}} = (\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_4)$ is selected to be centred at $(0, 0, 0, 0)$ with frequencies $S_1 \cup \{\xi(\mathbf{a}_{(i)})\}$ and scales $\rho_{2,j+22}$, $\rho_{2,j+21}$, $\rho_{2,j+20}$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

From two applications of Lemma 4.4 (and the fact that $W = O(\eta^{-C_1/100})$), we have

$$\mathbb{E}W(\vec{\mathbf{a}} + \vec{\mathbf{k}} + \vec{\mathbf{h}}) = \mathbb{E}W(\vec{\mathbf{a}} + \vec{\mathbf{k}}) + O(\eta^{10^3 C_1}) = \mathbb{E}W(\vec{\mathbf{a}}) + O(\eta^{10^3 C_1})$$

(for example). The claim (9.27) now follows from (9.23).

Now we verify (9.28). By (9.24), we have

$$\mathrm{E}_{i'}(\mathbf{N}) = \sum_{\vec{a},\vec{k}} \mathbb{P}(\vec{\mathbf{a}} = \vec{a}, \vec{\mathbf{k}} = \vec{k}) \mathbb{E}|1_{A_{(i')}}(a_{(i')} + k_{i'} + \mathbf{h}_{i'}) - \alpha_{i',\vec{a},\vec{k}}|^2,$$

where $\vec{a} = (a_{(1)}, \dots, a_{(4)})$, $\vec{k} = (k_1, \dots, k_4)$, and $\alpha_{i',\vec{a},\vec{k}}$ is the quantity

$$\alpha_{i',\vec{a},\vec{k}} := \mathbb{E}1_{A_{(i')}}(a_{(i')} + k_{i'} + \mathbf{h}_{i'}). \tag{9.30}$$

By Pythagoras' theorem, we thus have

$$\mathrm{E}_{i'}(\mathbf{N}) = \sum_{\vec{a},\vec{k}} \mathbb{P}(\vec{\mathbf{a}} = \vec{a}, \vec{\mathbf{k}} = \vec{k}) \mathbb{E}|1_{A_{(i')}}(a_{(i')} + k_{i'} + \mathbf{h}_{i'}) - \alpha_{i'}|^2 - |\alpha_{i',\vec{a},\vec{k}} - \alpha_{i'}|^2,$$

where $\alpha_{i'}$ is defined in (9.22). We shall shortly establish the bound

$$|\alpha_{i',\vec{a},\vec{k}} - \alpha_{i'}|^2 \gg \eta^{400C_1} 1_{i'=i}. \tag{9.31}$$

Assuming this bound, we conclude that

$$\begin{aligned}\mathbb{E}\,\mathrm{E}_{i'}(\mathbf{N}) &\leqslant \sum_{\vec{a},\vec{k}} \mathbb{P}(\vec{\mathbf{a}} = \vec{a}, \vec{\mathbf{k}} = \vec{k}) \mathbb{E}|1_{A_{(i')}}(a_{(i')} + k_{i'} + \mathbf{h}_{i'}) - \alpha_{i'}|^2 \\ &= \mathbb{E}|1_{A_{(i')}}(\mathbf{a}_{(i')} + \mathbf{k}_{i'} + \mathbf{h}_{i'}) - \alpha_{i'}|^2 - \eta^{500C_1} 1_{i'=i}.\end{aligned}$$

By applying Lemma 4.4 twice as in the proof of (9.27) to replace $\mathbf{a}_{(i')} + \mathbf{k}_{i'} + \mathbf{h}_{i'}$ by $\mathbf{a}_{(i')}$ for $i' = 2, 3, 4$ (and by using Lemma 4.4 six times for $i' = 1$, after writing $\mathbf{a}_{(1)}$ in terms of $\mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a}_{(4)}$, and similarly for $\mathbf{k}_{(1)}$ and $\mathbf{h}_{(1)}$) we thus have

$$\mathbb{E}\,\mathrm{E}_{i'}(\mathbf{N}) \leqslant \mathbb{E}|1_{A_{(i')}}(\mathbf{a}_{(i')}) - \alpha_{i'}|^2 - \eta^{500C_1} 1_{i'=i} + O(\eta^{10^3 C_1}).$$

This will give (9.28) as soon as we establish (9.31). This is trivial for $i' \neq i$, so suppose that $i = i$. By (9.30) and (9.21), it suffices to show that

$$\sum_{\vec{a},\vec{k}} \mathbb{P}(\vec{\mathbf{a}} = \vec{a}, \vec{\mathbf{k}} = \vec{k}) |\mathbb{E} f_i(a_{(i)} + k_i + \mathbf{h}_i)|^2 \gg \eta^{400C_1}. \tag{9.32}$$

To prove this, we introduce random variables $\mathbf{n}_0, \mathbf{n}_1$ drawn independently and regularly from $B(S_1, \rho_{2,j+10})$ and $B(S_1, \rho_{2,j+11})$ independently of all previous variables. From (9.29) we have

$$|\mathbb{E} f_i(\mathbf{a}_{(i)} + \mathbf{n}_0 + \mathbf{n}_1) g(\mathbf{n}_0, \mathbf{a}_{(i)}) e_p(-\xi(\mathbf{a}_{(i)})\mathbf{n}_1)| \gg \eta^{200C_1}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

for some 1-bounded function g. After using Lemma 4.4 to compare $\mathbf{n}_1$ and $\mathbf{n}_1 + \mathbf{h}_i$ for each fixed choice of $\mathbf{n}_0$ and $\mathbf{a}_{(i)}$, we conclude that

$$|\mathbb{E} f_i(\mathbf{a}_{(i)} + \mathbf{n}_0 + \mathbf{n}_1 + \mathbf{h}_i) g(\mathbf{n}_0, \mathbf{a}_{(i)}) e_p(-\xi(\mathbf{a}_{(i)})(\mathbf{n}_1 + \mathbf{h}_i))| \gg \eta^{200C_1}.$$

But we have

$$\left\| \frac{\xi(\mathbf{a}_{(i)})\mathbf{h}_i}{p} \right\|_{\mathbb{R}/\mathbb{Z}} \leqslant \|\mathbf{h}_i\|_{S_1 \cup \{\xi(\mathbf{a}_{(i)})\}} \ll \rho_{j+l_i+20}$$

and hence by (2.2)

$$e_p(-\xi(\mathbf{a}_{(i)})(\mathbf{n}_1 + \mathbf{h}_i)) = e_p(-\xi(\mathbf{a}_{(i)})\mathbf{n}_1) + O(\eta^{10^3 C_1}).$$

We conclude that

$$|\mathbb{E}(f_i(\mathbf{a}_{(i)} + \mathbf{n}_0 + \mathbf{n}_1 + \mathbf{h}_i) g(\mathbf{n}_0, \mathbf{a}_{(i)}) e_p(-\xi(\mathbf{a}_{(i)})\mathbf{n}_1))| \gg \eta^{200C_1}.$$

For fixed choices of $\mathbf{a}_{(i)}, \mathbf{h}_{(i)}, \mathbf{n}_1$, we see from Lemma 4.4 that $\mathbf{k}_i$ and $\mathbf{n}_0 + \mathbf{n}_1$ differ in total variation by $O(\eta^{10^3 C_1})$. Thus we have

$$|\mathbb{E}(f_i(\mathbf{a}_{(i)} + \mathbf{k}_i + \mathbf{h}_i) g(\mathbf{k}_i - \mathbf{n}_1, a_{(i)}) e_p(-\xi(\mathbf{a}_{(i)})\mathbf{n}_1))| \gg \eta^{200C_1},$$

and the claim now follows after using Lemma 2.1 to eliminate the $g(\mathbf{k}_i - \mathbf{n}_1, a_{(i)}) e_p(-\xi(\mathbf{a}_{(i)})\mathbf{n}_1)$ factor. □

A useful consequence of the bounds in Theorem 9.9(ii) is the following weak mixing bound, which roughly speaking asserts that the convolution of $1_{A_{(i)}}$ with a bounded function is essentially constant.

LEMMA 9.10. *Let the notation and hypotheses be as above, and let Ω and ξ be as in Theorem 9.2. Let $A_{(1)}, \dots, A_{(4)}$ be as in Theorem 9.7, and let $j, a_{(1),*}, \dots, a_{(4),*}, S_1, f_1, \dots, f_4$ be as in Theorem 9.9. Then for any $i = 1, 2, 3, 4$, any $l_i < m \leqslant 10$, and any 1-bounded function $g : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$, one has*

$$\sum_n \mathbb{P}(\mathbf{n} = n)|\mathbb{E} f_i(n - \mathbf{k}) g(\mathbf{k})|^2 \ll \eta^{50C_1}, \tag{9.33}$$

where $\mathbf{n}, \mathbf{k}$ are drawn independently and regularly from $a_{(i),*} + B(S_1, \rho_{2,j})$ and $B(S_1, \rho_{2,j+m})$ respectively. Dually, for any 1-bounded function $G : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$, one has

$$\sum_k \mathbb{P}(\mathbf{k} = k)|\mathbb{E} f_i(\mathbf{n} - k) G(\mathbf{n})| \ll \eta^{25C_1}. \tag{9.34}$$

Proof. In preparation for invoking Theorem 9.9(ii), we introduce random variables $\mathbf{h}_0, \mathbf{h}_1, \mathbf{h}_1'$ drawn independently and regularly from $B(S_1, \rho_{2,j_*+10})$, $B(S_1, \rho_{2,j_*+11})$, and $B(S_1, \rho_{2,j_*+11})$ respectively, independently of $\mathbf{n}$ and $\mathbf{k}$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Using Lemma 4.4 to compare $\mathbf{n}, \mathbf{k}$ with $\mathbf{n} + \mathbf{h}_0$, $\mathbf{k} - \mathbf{h}_1$ respectively, we may transform (9.33) to the estimate

$$\sum_{n,h_0} \mathbb{P}(\mathbf{n} = n, \mathbf{h}_0 = h_0)|\mathbb{E}(f_i(n + h_0 - \mathbf{k} - \mathbf{h}_1)g(\mathbf{k} - \mathbf{h}_1))|^2 \ll \eta^{50C_1}.$$

By the triangle inequality in L^2, it thus suffices to show that

$$\sum_{n,h_0} \mathbb{P}(\mathbf{n} = n, \mathbf{h}_0 = h_0)|\mathbb{E}(f_i(n + h_0 - k - \mathbf{h}_1)g(k - \mathbf{h}_1))|^2 \ll \eta^{50C_1} \quad (9.35)$$

for all $k \in B(S_1, \rho_{2,j_*+m})$.

Fix k. We may expand out the left-hand side of (9.35) as

$$\mathbb{E} f_i(\mathbf{n} + \mathbf{h}_0 - \mathbf{h}_1 - k)g(k - \mathbf{h}_1) f_i(\mathbf{n} + \mathbf{h}_0 - \mathbf{h}_1' - k)g(k - \mathbf{h}_1').$$

Using Lemma 4.4 to compare $\mathbf{n}$ with $\mathbf{n} + \mathbf{h}_0 - \mathbf{h}_1 - \mathbf{h}_1' - k$, we can thus rewrite (9.35) as

$$|\mathbb{E} f_i(\mathbf{n} + \mathbf{h}_0 + \mathbf{h}_1')g(k - \mathbf{h}_1) f_i(\mathbf{n} + \mathbf{h}_0 + \mathbf{h}_1)g(k - \mathbf{h}_1')| \ll \eta^{50C_1},$$

which by the triangle inequality and the 1-boundedness of g would follow from

$$\sum_{n,h_1,h_1'} \mathbb{P}(\mathbf{n} = n, \mathbf{h}_1 = h_1, \mathbf{h}_1' = h_1)|\mathbb{E} f_i(n + \mathbf{h}_0 + h_1') f_i(n + \mathbf{h}_0 + h_1)| \ll \eta^{50C_1},$$

which by Cauchy–Schwarz will follow in turn from

$$\sum_{n,h_1,h_1'} \mathbb{P}(\mathbf{n} = n, \mathbf{h}_1 = h_1, \mathbf{h}_1' = h_1)|\mathbb{E} f_i(n + \mathbf{h}_0 + h_1') f_i(n + \mathbf{h}_0 + h_1)|^2 \ll \eta^{100C_1}.$$

But this follows from Theorem 9.9(ii) (relabelling $\mathbf{n}$ as $\mathbf{a}_{(i)}$).

Finally, we show (9.34). By subtracting $\mathbb{E}G(\mathbf{n})$ from G (and dividing by 2 to recover 1-boundedness), we may assume that $\mathbb{E}G(\mathbf{n}) = 0$. It then suffices to show that

$$\sum_k \mathbb{P}(\mathbf{k} = k)g(k)\mathbb{E}1_{A_{(i)}}(\mathbf{n} - k)G(\mathbf{n}) \ll \eta^{25C_1}$$

for any 1-bounded function g. But the left-hand side may be rearranged as

$$\sum_n \mathbb{P}(\mathbf{n} = n)G(n)(\mathbb{E}1_{A_{(i)}}(n - \mathbf{k})g(\mathbf{k}) - \alpha_i \mathbb{E}g(\mathbf{k})) \ll \eta^{25C_1},$$

and the claim follows from (9.33) and the Cauchy–Schwarz inequality. □

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

9.11. *Fifth step: a frequency function ξ' that is approximately linear 99% of the time on a Bohr neighbourhood.* The next step is to obtain additive structure on almost all of a Bohr neighbourhood, rather than just the subsets $A_{(i)}$.

THEOREM 9.12. *Let the notation and hypotheses be as in Theorem 8.1, and let ξ be as in Theorem 9.2. Let $A_{(1)}, \dots, A_{(4)}$ be as in Theorem 9.7, and let j, $a_{(1),1}, a_{(2),1}, a_{(3),1}, a_{(4),1}, S_1, \alpha_1, \dots, \alpha_4$ be as in Theorem 9.9. Let $a_1 \in \mathbb{Z}/p\mathbb{Z}$ be the quantity*

$$a_1 := a_{(1),1} + a_{(2),1} = a_{(3),1} + a_{(4),1},$$

and let $\mathbf{a}$ and $\mathbf{a}_{(2)}$ be drawn regularly and independently from $a_1 + B(S_1, \rho_{2,j})$ and $a_{(2),1} + B(S_1, \rho_{2,j+2})$ respectively. Then there is a function $\xi' : \mathbb{Z}/p\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$, such that with probability at least $1 - O(\eta^{C_1/200})$, the random variable $\mathbf{a}$ attains a value a for which we have the estimates

$$\mathbb{E}1_{A_{(2)}}(\mathbf{a}_{(2)})1_{A_{(1)}}(a - \mathbf{a}_{(2)}) = \alpha_1\alpha_2 + O(\eta^{20C_1}), \tag{9.36}$$

and

$$\mathbb{P}\left(a - \mathbf{a}_{(2)} \in A_{(1)}; a_{(2)} \in A_{(2)}; \|\xi'(a) - \xi(a - \mathbf{a}_{(2)}) - \xi(\mathbf{a}_{(2)})\|_S > \frac{1}{\rho_3}\right) \ll \eta^{C_1/200}\alpha_1\alpha_2. \tag{9.37}$$

Proof. Let $\mathbf{a}$ be drawn regularly from $a_1 + B(S_1, \rho_{2,j})$, and let $(\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a}_{(4)})$ be a random additive quadruple centred at $(a_{(1),1}, a_{(2),1}, a_{(3),1}, a_{(4),1})$ with frequencies S_1 and scales $\rho_{2,j+2}, \rho_{2,j+1}, \rho_{2,j}$, independently of $\mathbf{a}$. From the definition of an additive quadruple, we have $\mathbf{a}_{(1)} = \mathbf{a}_{(3)} + \mathbf{a}_{(4)} - \mathbf{a}_{(2)}$. From Theorem 9.9(i) we thus have

$$\mathbb{E}W(\mathbf{a}_{(3)} + \mathbf{a}_{(4)} - \mathbf{a}_{(2)}, \mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a}_{(4)}) \gg \eta^{C_1+O(1)}. \tag{9.38}$$

From Lemma 4.4 we see that once we condition $\mathbf{a}_{(2)}$ and $\mathbf{a}_{(3)}$ to be fixed, $\mathbf{a}_{(4)}$ and $\mathbf{a} - \mathbf{a}_{(3)}$ differ in total variation by $O(\eta^{100C_1})$. Thus we may replace $\mathbf{a}_{(4)}$ by $\mathbf{a} - \mathbf{a}_{(3)}$ in (9.38) to conclude that

$$\mathbb{E}W(\mathbf{a} - \mathbf{a}_{(2)}, \mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a} - \mathbf{a}_{(3)}) \gg \eta^{C_1+O(1)}.$$

If we then define

$$\sigma := \mathbb{E}1_{A_{(1)}}(\mathbf{a} - \mathbf{a}_{(2)})1_{A_{(2)}}(\mathbf{a}_{(2)})1_{A_{(3)}}(\mathbf{a}_{(3)})1_{A_{(4)}}(\mathbf{a} - \mathbf{a}_{(3)})$$

then from (9.12) we see that

$$\sigma \gg \eta^{C_1+O(1)} \tag{9.39}$$

and

$$\mathbb{E}1_{A_{(1)}}(\mathbf{a} - \mathbf{a}_{(2)})1_{A_{(2)}}(\mathbf{a}_{(2)})1_{A_{(3)}}(\mathbf{a}_{(3)})1_{A_{(4)}}(\mathbf{a} - \mathbf{a}_{(3)}) \times 1_{\mathrm{VBQ}}(\mathbf{a} - \mathbf{a}_{(2)}, \mathbf{a}_{(2)}, \mathbf{a}_{(3)}, \mathbf{a} - \mathbf{a}_{(3)}) \ll \eta^{-C_1/100}\sigma. \tag{9.40}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We can express σ in the form

$$\sigma = \mathbb{E} g_{12}(\mathbf{a}) g_{34}(\mathbf{a}), \tag{9.41}$$

where $g_{12}, g_{34} : \mathbb{Z}/p/\mathbb{Z} \to \mathbb{R}$ are the functions

$$g_{12}(a) := \mathbb{E} 1_{A_{(1)}}(a - \mathbf{a}_{(2)}) 1_{A_{(2)}}(\mathbf{a}_{(2)}) \tag{9.42}$$

and

$$g_{34}(a) := \mathbb{E} 1_{A_{(3)}}(\mathbf{a}_{(3)}) 1_{A_{(4)}}(a - \mathbf{a}_{(3)}).$$

From Lemma 9.10, we have

$$\sum_n \mathbb{P}(\mathbf{n} = n) |\mathbb{E} f_1(n - \mathbf{k}) 1_{A_{(2)}}(a_{(2),1} + \mathbf{k})|^2 \ll \eta^{50C_1}$$

if $\mathbf{n}, \mathbf{k}$ are drawn independently and regularly from $a_{(i),1} + B(S_1, \rho_{2,j})$ and $B(S_1, \rho_{2,j+m})$ respectively. Note that the pair $(\mathbf{n}, \mathbf{k})$ has the same distribution as $(\mathbf{a} - a_{(2),1}, \mathbf{a}_{(2)} - a_{(2),1})$, thus

$$\sum_a \mathbb{P}(\mathbf{a} = a) |\mathbb{E} f_1(a - \mathbf{a}_{(2)}) 1_{A_{(2)}}(\mathbf{a}_{(2)})|^2 \ll \eta^{50C_1}.$$

From (9.21), (9.22), (9.42) we have

$$\mathbb{E} f_1(a - \mathbf{a}_{(2)}) 1_{A_{(2)}}(\mathbf{a}_{(2)}) = g_{12}(a) - \alpha_1\alpha_2$$

and thus

$$\sum_a \mathbb{P}(\mathbf{a} = a) |g_{12}(a) - \alpha_1\alpha_2|^2 \ll \eta^{50C_1}. \tag{9.43}$$

Similarly we have

$$\sum_a \mathbb{P}(\mathbf{a} = a) |g_{34}(a) - \alpha_3\alpha_4|^2 \ll \eta^{50C_1}. \tag{9.44}$$

From Cauchy–Schwarz and the triangle inequality we conclude that

$$\sum_a \mathbb{P}(\mathbf{a} = a) |g_{12}(a) g_{34}(a) - \alpha_1\alpha_2\alpha_3\alpha_4| \ll \eta^{25C_1},$$

and hence by (9.41) and the triangle inequality

$$\sigma = \alpha_1\alpha_2\alpha_3\alpha_4 + O(\eta^{25C_1}). \tag{9.45}$$

In particular, from (9.39) one has

$$\alpha_1\alpha_2\alpha_3\alpha_4 \gg \eta^{C_1 + O(1)}. \tag{9.46}$$

From (9.45), (9.46) and (9.40) we have

$$\mathbb{E} h(\mathbf{a}) \ll \eta^{C_1/100} \alpha_1\alpha_2\alpha_3\alpha_4,$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

where

$$h(a) := \mathbb{E}W(a - \mathbf{a}_{(2)}, \mathbf{a}_{(2)}, \mathbf{a}_{(3)}, a - \mathbf{a}_{(3)}). \tag{9.47}$$

By Markov's inequality, we conclude that we have

$$h(\mathbf{a}) \ll \eta^{C_1/200}\alpha_1\alpha_2\alpha_3\alpha_4 \tag{9.48}$$

with probability $1 - O(\eta^{C_1/200})$. Similarly, from (9.43), (9.44) and Chebyshev's inequality we also have

$$g_{12}(\mathbf{a}) = \alpha_1\alpha_2 + O(\eta^{20C_1}) \tag{9.49}$$

and

$$g_{34}(\mathbf{a}) = \alpha_3\alpha_4 + O(\eta^{20C_1}) \tag{9.50}$$

with probability $1 - O(\eta^{C_1/200})$.

Now let a be a value of $\mathbf{a}$ be such that (9.48)–(9.50) hold. From (9.50) we have in particular that

$$\mathbb{E}1_{A_{(3)}}(\mathbf{a}_{(3)})1_{A_{(4)}}(a - \mathbf{a}_{(3)}) \gg \alpha_3\alpha_4;$$

comparing this with (9.48) and (9.47), we see that we may find $a_{(3)}(a) \in A_{(3)}$ (depending only on a) with $a - a_{(3)}(a) \in A_{(4)}$ such that

$$\begin{aligned}\mathbb{E}1_{A_{(1)}}(a - \mathbf{a}_{(2)})1_{A_{(2)}}(\mathbf{a}_{(2)})1_{\mathrm{VBQ}}(a - \mathbf{a}_{(2)}, \mathbf{a}_{(2)}, a_{(3)}(a), a - a_{(3)}(a))\\ \ll \eta^{C_1/200}\alpha_1\alpha_2.\end{aligned}$$

If we then set $\xi'(a) := \xi(a_{(3)}(a)) + \xi(a - a_{(3)}(a))$ (and define $\xi'(\mathbf{a})$ arbitrarily when (9.48), (9.49), or (9.50) fail), then the claims (9.36), (9.37) follow from (9.49) and the definition (9.10) of VBQ. □

The function ξ' has better additive structure than ξ, in that it respects almost all additive quadruples in a Bohr set, rather than almost all additive quadruples in a rough set. More precisely, we have the following.

PROPOSITION 9.13. *Let the notation and hypotheses be as in Theorem 9.12. Suppose that* $\mathbf{a}, \mathbf{a}', \mathbf{h}$ *are selected independently and regularly from* $a_1 + B(S_1, \rho_{2,j})$, $a_1 + B(S_1, \rho_{2,j})$, *and* $B(S_1, \rho_{2,j+3})$ *respectively. Then with probability* $1 - O(\eta^{C_1/200})$ *we have*

$$\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} + \mathbf{h}) - \xi'(\mathbf{a}') + \xi'(\mathbf{a}' + \mathbf{h})\|_S \leqslant \frac{4}{\rho_3}. \tag{9.51}$$

Proof. Let $\mathbf{a}_{(2)}$ be drawn regularly from $a_{(2),1} + B(S_1, \rho_{2,j+2})$, independently of $\mathbf{a}, \mathbf{a}', \mathbf{h}$. For each $a, a', h \in \mathbb{Z}/p\mathbb{Z}$, let $\mathbf{I}_{a,a',h}$ denote the random indicator variable

$$\mathbf{I}_{a,a',h} := 1_{A_{(2)}}(\mathbf{a}_{(2)})1_{A_{(2)}}(\mathbf{a}_{(2)} + h)1_{A_{(1)}}(a - \mathbf{a}_{(2)})1_{A_{(1)}}(a' - \mathbf{a}_{(2)}).$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Suppose that we can show that with probability $1 - O(\eta^{C_1/200})$, the triple $(\mathbf{a}, \mathbf{a}', \mathbf{h})$ attains a value (a, a', h) for which one has the estimates

$$\mathbb{E}\mathbf{I}_{a,a',h} \geqslant 0.9\alpha_1^2\alpha_2^2, \tag{9.52}$$

$$\mathbb{E}\mathbf{I}_{a,a',h}1_{\|\xi'(a)-\xi(a-\mathbf{a}_{(2)})-\xi(\mathbf{a}_{(2)})\|_S>1/\rho_3} \leqslant 0.1\alpha_1^2\alpha_2^2, \tag{9.53}$$

$$\mathbb{E}\mathbf{I}_{a,a',h}1_{\|\xi'(a')-\xi(a'-\mathbf{a}_{(2)})-\xi(\mathbf{a}_{(2)})\|_S>1/\rho_3} \leqslant 0.1\alpha_1^2\alpha_2^2, \tag{9.54}$$

$$\mathbb{E}\mathbf{I}_{a,a',h}1_{\|\xi'(a+h)-\xi(a-\mathbf{a}_{(2)})-\xi(\mathbf{a}_{(2)}+h)\|_S>1/\rho_3} \leqslant 0.1\alpha_1^2\alpha_2^2, \tag{9.55}$$

$$\mathbb{E}\mathbf{I}_{a,a',h}1_{\|\xi'(a'+h)-\xi(a'-\mathbf{a}_{(2)})-\xi(\mathbf{a}_{(2)}+h)\|_S>1/\rho_3} \leqslant 0.1\alpha_1^2\alpha_2^2. \tag{9.56}$$

Assuming these estimates, we conclude from the union bound that with probability $1 - O(\eta^{C_1/200})$, the random variable $(\mathbf{a}, \mathbf{a}', \mathbf{h})$ attains a value (a, a', h) for which there exists at least one element $a_{(2)}$ of $\mathbb{Z}/p\mathbb{Z}$ obeying the constraints

$$\begin{aligned}
a_{(2)}, a_{(2)} + h &\in A_{(2)},\\
a - a_{(2)}, a' - a_{(2)} &\in A_{(1)},\\
\|\xi'(a) - \xi(a - a_{(2)}) - \xi(a_{(2)})\|_S &\leqslant \frac{1}{\rho_3},\\
\|\xi'(a') - \xi(a' - a_{(2)}) - \xi(a_{(2)})\|_S &\leqslant \frac{1}{\rho_3},\\
\|\xi'(a + h) - \xi(a - a_{(2)}) - \xi(a_{(2)} + h)\|_S &\leqslant \frac{1}{\rho_3},\\
\|\xi'(a' + h) - \xi(a' - a_{(2)}) - \xi(a_{(2)} + h)\|_S &\leqslant \frac{1}{\rho_3}
\end{aligned}$$

and (9.51) then follows from the triangle inequality.

It remains to establish (9.52)–(9.56). We first prove (9.53). By Markov's inequality, it suffices to show that

$$\mathbb{E}\mathbf{I}_{\mathbf{a},\mathbf{a}',\mathbf{h}}1_{\|\xi'(\mathbf{a})-\xi(\mathbf{a}-\mathbf{a}_{(2)})-\xi(\mathbf{a}_{(2)})\|_S>1/\rho_3} \ll \eta^{C_1/200}\alpha_1^2\alpha_2^2.$$

We rewrite the left-hand side as

$$\mathbb{E}g_1(\mathbf{a}_{(2)})g_2(\mathbf{a}_{(2)})1_{A_{(2)}}(\mathbf{a}_{(2)})1_{A_{(1)}}(\mathbf{a} - \mathbf{a}_{(2)})1_{\|\xi'(\mathbf{a})-\xi(\mathbf{a}-\mathbf{a}_{(2)})-\xi(\mathbf{a}_{(2)})\|_S>1/\rho_3}$$

where

$$g_1(a_{(2)}) := \mathbb{E}1_{A_{(1)}}(\mathbf{a}' - a_{(2)})$$

and

$$g_2(a_{(2)}) := \mathbb{E}1_{A_{(2)}}(a_{(2)} + \mathbf{h}).$$

But from (9.37) we have

$$\mathbb{E}1_{A_{(2)}}(\mathbf{a}_{(2)})1_{A_{(1)}}(\mathbf{a} - \mathbf{a}_{(2)})1_{\|\xi'(\mathbf{a})-\xi(\mathbf{a}-\mathbf{a}_{(2)})-\xi(\mathbf{a}_{(2)})\|_S>1/\rho_3} \ll \eta^{C_1/200}\alpha_1\alpha_2,$$

from Lemma 4.4 one has

$$g_1(\mathbf{a}_{(2)}) = \alpha_1 + O(\eta^{10C_1})$$

and from (9.33) one has

$$g_2(\mathbf{a}_{(2)}) = \alpha_2 + O(\eta^{10C_1})$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

with probability $1 - O(\eta^{10C_1})$ (for example), with the trivial bound $g(\mathbf{a}_{(2)}) = O(1)$ otherwise, and the claim (9.53) then follows from (9.46).

The proofs of (9.54)–(9.56) are similar to (9.53) and are omitted. It thus remains to prove (9.52). From (9.34) and Markov's inequality, we see that with probability $1 - O(\eta^{C_1/200})$, the random variable $\mathbf{h}$ attains a value h for which

$$\mathbb{E}1_{A_{(2)}}(\mathbf{a}_{(2)})1_{A_{(2)}}(\mathbf{a}_{(2)} + h) \geqslant 0.99\alpha_2^2.$$

For any h obeying this inequality, define $E(h) \subset \mathbb{Z}/p\mathbb{Z}$ to be the set

$$E(h) := A_{(2)} \cap (A_{(2)} - h),$$

so that

$$\mathbb{P}(\mathbf{a}_{(2)} \in E(h)) \geqslant 0.99\alpha_2^2.$$

By (9.33) and the Chebyshev inequality, we conclude that with probability $1 - O(\eta^{C_1/200})$, the random variable $(\mathbf{a}, \mathbf{h})$ attains a value (a, h) for which one has

$$\mathbb{P}(\mathbf{a}_{(2)} \in E(h); a - \mathbf{a}_{(2)} \in A_{(1)}) \geqslant 0.98\alpha_1\alpha_2^2.$$

For any (a, h) of the above form, define $E'(a, h) \subset \mathbb{Z}/p\mathbb{Z}$ to be the set

$$E'(a, h) := \mathbb{E}(h) \cap (a - A_{(1)}),$$

then

$$\mathbb{P}(\mathbf{a}_{(2)} \in E'(a, h)) \geqslant 0.98\alpha_1\alpha_2^2.$$

By one last application of (9.33) and the Chebyshev inequality, we see that with probability $1 - O(\eta^{C_1/200})$, the random variable $(\mathbf{a}', \mathbf{a}, \mathbf{h})$ attains a value (a', a, h) for which one has

$$\mathbb{P}(\mathbf{a}_{(2)} \in E'(a, h); a' - \mathbf{a}_{(2)} \in A_{(1)}) \geqslant 0.97\alpha_1^2\alpha_2^2$$

which gives (9.52) as required. □

9.14. *Sixth step: a frequency function ξ'' that is approximately linear 100% of the time on a Bohr set.* We now use a standard "majority vote" argument to upgrade the "99% linear" structure of ξ' to a "100% linear" structure of a closely related function ξ'' (cf. [5]). More precisely, one has the following.

THEOREM 9.15. *Let the notation and hypotheses be as in Theorem 8.1. Let j, S_1 be as in Theorem 9.9, and let a_1, ξ' be as in Theorem 9.12. Then there is a function $\xi'' : B(S_1, \rho_3) \to \mathbb{Z}/p\mathbb{Z}$ such that*

$$\|\xi''(n + m) - \xi''(n) - \xi''(m)\|_S \leqslant \frac{24}{\rho_3} \tag{9.57}$$

for all $n, m \in B(S_1, \rho_3/2)$, and such that for any $n \in B(S_1, \rho_3)$, if $\mathbf{a}$ is drawn regularly from $a_1 + B(S_1, \rho_{2,j})$, one has

$$\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - n) - \xi''(n)\|_S \leqslant \frac{8}{\rho_3} \tag{9.58}$$

with probability $1 - O(\eta^{C_1/200})$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Proof. Let $\mathbf{a}, \mathbf{h}$ be drawn independently and regularly from $a_* + B(S_1, \rho_{2,j})$ and $B(S_1, \rho_{2,j+3})$ respectively. From Proposition 9.13 and the pigeonhole principle, we may find $a_0' \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\mathbb{P}\left(\|\xi'(\mathbf{a}) - \xi'(\mathbf{a}+\mathbf{h}) - \xi'(a_0') + \xi'(a_0'+\mathbf{h})\|_S \leqslant \frac{4}{\rho_3} \right) \geqslant 1 - O(\eta^{C_1/200}). \quad (9.59)$$

Fix this a_0'. Now let n by an arbitrary element of $B(S_1, \rho_3)$. Then using Lemma 4.4 to compare $\mathbf{a}$ with $\mathbf{a} - n$ and $\mathbf{h}$ with $\mathbf{h} + n$, we obtain

$$\mathbb{P}\left(\|\xi'(\mathbf{a} - n) - \xi'(\mathbf{a} + \mathbf{h}) - \xi'(a_0') + \xi'(a_0' + \mathbf{h} + n)\|_S \leqslant \frac{4}{\rho_3} \right) \geqslant 1 - O(\eta^{C_1/200}).$$

Combining this with (9.59) and the triangle inequality, we see that

$$\mathbb{P}\left(\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - n) + \xi'(a_0' + \mathbf{h}) - \xi'(a_0' + \mathbf{h} + n)\|_S \leqslant \frac{8}{\rho_3} \right) \geqslant 1 - O(\eta^{C_1/200}).$$

Thus, by the pigeonhole principle, we may find $h_n \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\mathbb{P}\left(\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - n) + \xi'(a_0' + h_n) - \xi'(a_0' + h_n + n)\|_S \leqslant \frac{8}{\rho_3} \right) \geqslant 1 - O(\eta^{C_1/200}).$$

If we thus define

$$\xi''(n) := \xi'(a_0' + h_n + n) - \xi'(a_0' + n)$$

then we have obtained (9.58).

Now suppose that $n, m \in B(S_1, \rho_3/2)$. From (9.58), we see that with probability at least $1 - O(\eta^{C_1/200})$ we have

$$\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - n) - \xi''(n)\|_S \leqslant \frac{8}{\rho_3},$$
$$\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - m) - \xi''(m)\|_S \leqslant \frac{8}{\rho_3},$$

and

$$\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - n - m) - \xi''(n + m)\|_S \leqslant \frac{8}{\rho_3}.$$

Using Lemma 4.4 to compare $\mathbf{a}$ with $\mathbf{a} - n$ in the second inequality, we also conclude

$$\|\xi'(\mathbf{a} - n) - \xi'(\mathbf{a} - n - m) - \xi''(m)\|_S \leqslant \frac{8}{\rho_3},$$

with probability $1 - O(\eta^{C_1/200})$. Thus there is a positive probability that the first, third, and fourth estimates hold simultaneously, and the claim (9.57) follows from the triangle inequality. □

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

The function ξ'' is still closely related to ξ, and in particular a variant of the correlation estimate (9.3) is obeyed by ξ''.

PROPOSITION 9.16. *Let the notation and hypotheses be as in the preceding theorem. Then there exist $a_0 \in B(S, 3\rho_2)$ and $\xi_0 \in \mathbb{Z}/p\mathbb{Z}$ such that*

$$\sum_{n_0,n} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n} = n)|\mathbb{E}f(n_0 + \mathbf{h} + a_0 - n)\overline{f}(n_0 + \mathbf{h})e_p((\xi''(n) - \xi_0)\mathbf{h})|^2 \gg \eta^{C_1+O(1)},$$

where $\mathbf{n}, \mathbf{n}_0, \mathbf{h}$ are drawn independently and regularly from the Bohr sets $B(S_1, \rho_3/4)$, $B(S, \rho_0)$, $B(S_1, \rho_4)$ respectively.

With this proposition and the previous theorem, we may now safely forget about the original function ξ, and work now with ξ''; the parameters a_1, j will also no longer be relevant.

Proof. Let $\mathbf{n}, \mathbf{a}, \mathbf{a}_{(2)}$ be drawn independently and regularly from $B(S_1, \rho_3/4)$, $a_1 + B(S_1, \rho_{2,j})$, and $B(S_1, \rho_{2,j+2})$ respectively. From (9.58) we have

$$\|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - \mathbf{n}) - \xi''(\mathbf{n})\|_S \ll \frac{1}{\rho_3}$$

with probability $1 - O(\eta^{C_1/200})$. Similarly, from (9.36), (9.37), (9.46) we see that with probability $1 - O(\eta^{C_1/200})$, the random variable $\mathbf{a}$ attains a value a for which

$$\mathbb{P}\left(a - \mathbf{a}_{(2)} \in A_{(1)}; \mathbf{a}_{(2)} \in A_{(2)}; \|\xi'(a) - \xi(a - \mathbf{a}_{(2)}) - \xi(\mathbf{a}_{(2)})\|_S \leqslant \frac{1}{\rho_3}\right) \gg \alpha_1\alpha_2.$$

Using Lemma 4.4 to compare $\mathbf{a}$ and $\mathbf{a} - \mathbf{n}$, we also see that with probability $1 - O(\eta^{C_1/200})$, the random variable $(\mathbf{a}, \mathbf{n})$ attains a value (a, n) for which

$$\mathbb{P}\Big(a - n - \mathbf{a}_{(2)} \in A_{(1)}; \mathbf{a}_{(2)} \in A_{(2)}; \|\xi'(a - n) - \xi(a - n - \mathbf{a}_{(2)}) - \xi(\mathbf{a}_{(2)})\|_S \leqslant \frac{1}{\rho_3}\Big) \gg \alpha_1\alpha_2.$$

From the union bound and Fubini's theorem, we conclude that with probability $\gg\alpha_1\alpha_2$, we simultaneously have the statements

$$\begin{aligned} \mathbf{a} - \mathbf{n} - \mathbf{a}_{(2)} &\in A_{(1)}, \\ \mathbf{a}_{(2)} &\in A_{(2)}, \\ \|\xi'(\mathbf{a}) - \xi'(\mathbf{a} - \mathbf{n}) - \xi''(\mathbf{n})\|_S &\ll \frac{1}{\rho_3}, \\ \|\xi'(\mathbf{a} - \mathbf{n}) - \xi(\mathbf{a} - \mathbf{n} - \mathbf{a}_{(2)}) - \xi(\mathbf{a}_{(2)})\|_S &\leqslant \frac{1}{\rho_3} \end{aligned}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

and hence by the triangle inequality

$$\|\xi'(\mathbf{a}) - \xi(\mathbf{a} - \mathbf{n} - \mathbf{a}_{(2)}) - \xi(\mathbf{a}_{(2)}) - \xi''(\mathbf{n})\|_S \ll \frac{1}{\rho_3}.$$

By the pigeonhole principle, we may thus find $a, a_{(2)} \in \mathbb{Z}/p\mathbb{Z}$ such that the statements

$$\begin{aligned} a - \mathbf{n} - a_{(2)} &\in A_{(1)}, \\ a_{(2)} &\in A_{(2)}, \\ \|\xi'(a) - \xi(a - \mathbf{n} - a_{(2)}) - \xi(a_{(2)}) - \xi''(\mathbf{n})\|_S &\ll \frac{1}{\rho_3} \end{aligned}$$

simultaneously hold with probability $\gg \alpha_1\alpha_2$, and thus with probability $\gg \eta^{C_1+O(1)}$ thanks to (9.46). Writing $a_0 := a - a_{(2)}$ and $\xi_0 := \xi(a_{(2)}) - \xi'(a)$, and recalling from Theorem 9.7 that $A_{(1)} \in S$, we thus have

$$\mathbb{P}(a_0 - \mathbf{n} \in S; \|\xi''(\mathbf{n}) + \xi(a_0 - \mathbf{n}) - \xi_0\|_S \ll 1/\rho_3) \gg \eta^{C_1+O(1)}.$$

In particular, since $\mathbf{n} \in B(S_1, \rho_3/4)$ and $S \subset B(S, 2\rho_2)$, we have $a_0 \in B(S, 3\rho_2)$.

Let $\mathbf{n}_0, \mathbf{n}_1$ be drawn independently and regularly from $B(S, \rho_0)$, $B(S, \rho_1)$ respectively, independently of all previous random variables. From the above estimate and (9.3), we see that with probability $\gg \eta^{C_1+O(1)}$, the random variable $\mathbf{n}$ attains a value n for which the statements

$$a_0 - n \in S \tag{9.60}$$

$$\|\xi''(n) + \xi(a_0 - n) - \xi_0\|_{S_1} \ll 1/\rho_3 \tag{9.61}$$

$$\begin{aligned} &\sum_{n_0} \mathbb{P}(\mathbf{n}_0 = n_0) \\ &\quad \times |\mathbb{E} f(n_0 + \mathbf{n}_1 + a_0 - n)\overline{f}(n_0 + \mathbf{n}_1) e_p(-\xi(a_0 - n)\mathbf{n}_1)|^2 \geqslant \eta/8 \end{aligned} \tag{9.62}$$

simultaneously hold.

Let n obey the above estimates (9.60)–(9.62). If we now draw $\mathbf{h}$ regularly from $B(S_1, \rho_4)$, then by using Lemma 4.4 to compare $\mathbf{n}_1$ with $\mathbf{n}_1 + \mathbf{h}$ in (9.62), we obtain

$$\begin{aligned} &\sum_{n_0} \mathbb{P}(\mathbf{n}_0 = n_0) |\mathbb{E} f(n_0 + \mathbf{n}_1 + \mathbf{h} + a_0 - n)\overline{f}(n_0 + \mathbf{n}_1 + \mathbf{h}) \\ &\quad \times e_p(-\xi(a_0 - n)(\mathbf{n}_1 + \mathbf{h}))|^2 \gg \eta \end{aligned}$$

and thus by the triangle inequality in L^2

$$\begin{aligned} &\sum_{n_0, n_1} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_1 = n_1) |\mathbb{E} f(n_0 + n_1 + \mathbf{h} + a_0 - n)\overline{f}(n_0 + n_1 + \mathbf{h}) \\ &\quad \times e_p(-\xi(a_0 - n)(n_1 + \mathbf{h}))|^2 \gg \eta. \end{aligned}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

We may delete the deterministic phase $e_p(-\xi(a_0-n)n_1)$ to obtain

$$\sum_{n_0,n_1}\mathbb{P}(\mathbf{n}_0=n_0,\mathbf{n}_1=n_1)|\mathbb{E}f(n_0+n_1+\mathbf{h}+a_0-n)\overline{f}(n_0+n_1+\mathbf{h})\\ \times e_p(-\xi(a_0-n)\mathbf{h})|^2\gg\eta.$$

Since $\mathbf{h}$ takes values in $B(S_1,\rho_4)$, we see from (9.61) that

$$e_p(-\xi(a_0-\mathbf{n})\mathbf{h})=e_p((\xi''(\mathbf{n})-\xi_0)\mathbf{h})+O(\eta^{100})$$

(for example), and so

$$\sum_{n_0,n_1}\mathbb{P}(\mathbf{n}_0=n_0,\mathbf{n}_1=n_1)|\mathbb{E}f(n_0+n_1+\mathbf{h}+a_0-n)\overline{f}(n_0+n_1+\mathbf{h})\\ \times e_p((\xi''(n)-\xi_0)\mathbf{h})|^2\gg\eta.$$

Using Lemma 4.4 to compare $\mathbf{n}_0$ with $\mathbf{n}_0+\mathbf{n}_1$, we conclude that

$$\sum_{n_0,n_1}\mathbb{P}(\mathbf{n}_0=n_0,\mathbf{n}_1=n_1)|\mathbb{E}f(n_0+\mathbf{h}+a_0-n)\overline{f}(n_0+\mathbf{h})e_p((\xi''(n)-\xi_0)\mathbf{h})|^2\\ \gg\eta.$$

Multiplying by $\mathbb{P}(\mathbf{n}=n)$ and summing in n, we obtain the claim. □

9.17. *Seventh step: derivatives of f correlate with a locally bilinear form.* We now pass to the "cohomological" phase of the argument, in which we remove the error $\xi''(n+m)-\xi''(n)-\xi''(m)$ in the linearity of ξ'' that appears in (9.57). This improved linearity of the form $(n,h)\mapsto\xi(n)h$ in the n aspect will come at the expense of the h aspect, which will now merely be locally linear instead of globally linear. However, this is a worthwhile tradeoff for our purposes (and in any event local linearity is more natural in this context than global linearity).

More precisely, the purpose of this subsection is to establish the following result towards the proof of Theorem 8.1.

THEOREM 9.18. *Let the notation and hypotheses be as in Theorem 8.1. Then there exists a set S_1 with $S\subset S_1\subset\mathbb{Z}/p\mathbb{Z}$ and $|S_1|\leqslant|S|+O(\eta^{-O(C_1)})$, a locally bilinear map*

$$\Xi:B(S_1,\rho_4)\times B(S_1,\rho_4)\to\mathbb{R}/\mathbb{Z},$$

a shift $a_1\in B(S,4\rho_2)$, and a frequency $\xi_1\in\mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_0,n_1}\mathbb{P}(\mathbf{n}_0=n_0,\mathbf{n}_1=n_1)\\ \times\left|\mathbb{E}f(n_0+\mathbf{m}_1+a_1-n_1)\overline{f}(n_0+\mathbf{m}_1)e\left(\Xi(n_1,\mathbf{m}_1)-\frac{\xi_1\mathbf{m}_1}{p}\right)\right|^2\\ \gg\eta^{C_1+O(1)} \tag{9.63}$$

if $\mathbf{n}_0,\mathbf{m}_1,\mathbf{n}_1$ are drawn independently and regularly from $B(S,\rho_0)$, $B(S_1,\rho_5)$, and $B(S_1,\rho_6)$ respectively.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Once the proof of this theorem is completed, the auxiliary data $\xi, \xi', \xi'', j, \Omega$, VBQ used in the previous parts of the section are no longer needed and may be discarded.

We now prove Theorem 9.18. Let j_*, S_1 be as in Theorem 9.9, let a_*, ξ' be as in Theorem 9.12, let $\xi'' : B(S_1, \rho_3) \to \mathbb{Z}/p\mathbb{Z}$ be as in Theorem 9.15, and let a_0, ξ_0 be as in Proposition 9.16. We will use a "cohomological" argument to construct the required bilinear map Ξ. Namely, we define the *cocycle* $\mu : B(S_1, \rho_3/2) \times B(S_1, \rho_3/2) \to \mathbb{Z}/p\mathbb{Z}$ to be the quantity

$$\mu(n, m) := \xi''(n + m) - \xi''(n) - \xi''(m). \tag{9.64}$$

Clearly (9.57) is symmetric, and we have the *cocycle equation*

$$\mu(n_1, n_2 + n_3) + \mu(n_2, n_3) = \mu(n_1, n_2) + \mu(n_1 + n_2, n_3) \tag{9.65}$$

as well as the auxiliary equations

$$\mu(n_1, n_2) = \mu(n_2, n_1); \qquad \mu(n_1, 0) = 0$$

whenever $n_1, n_2, n_3 \in B(S_1, \rho_3/4)$. From (9.57) we also have the estimate

$$\|\mu(n, m)\|_S \leqslant \frac{24}{\rho_3} \tag{9.66}$$

for all $n, m \in B(S_1, \rho_3/4)$.

To construct the bilinear map Ξ, we will show that a certain projection of μ is a "coboundary" is a certain sense. Let $\phi : \mathbb{Z}^S \to \mathbb{Z}/p\mathbb{Z}$ be the homomorphism

$$\phi((n_s)_{s\in S}) := \sum_{s\in S} n_s s.$$

From (9.66), we see that for each $n, m \in B(S_1, \rho_3/4)$ we have a representation of the form

$$\mu(n, m) = \phi(\tilde{\mu}(n, m)) \tag{9.67}$$

for some lift $\tilde{\mu}(n, m) \in \mathbb{Z}^S$ of size

$$|\tilde{\mu}(n, m)| \leqslant 24/\rho_3. \tag{9.68}$$

This lift $\tilde{\mu}(n, m)$ is only defined up to an element of the kernel $\ker(\phi) := \{p \in \mathbb{Z}^S : \phi(p) = 0\}$ of ϕ; to eliminate this ambiguity we will apply a projection. Since S contains a non-zero element, $\phi : \mathbb{Z}^S \to \mathbb{Z}/p\mathbb{Z}$ is a surjective homomorphism, and in particular, $\ker(\phi)$ is a sublattice of $\mathbb{Z}^S$ of index p. Applying Lemma 4.8, we may find generators $v_1, \dots, v_{|S|}$ of $\ker(\phi)$ and real numbers $N_1, \dots, N_{|S|} > 0$ with

$$\prod_{i=1}^{|S|} N_i = O(K)^{O(K)} p \tag{9.69}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

such that

$$B_{\mathbb{R}^S}(0, O(K)^{-3K/2}t) \cap \ker(\phi) \subset \{n_1 v_1 + \cdots + n_{|S|} v_{|S|} : |n_i| \leqslant t N_i\} \subset B_{\mathbb{R}^S}(0, t) \cap \ker(\phi) \tag{9.70}$$

for all $t > 0$.

By relabelling, we may take the N_i to be non-increasing. Let d, $0 \leqslant d \leqslant |S|$ be such that

$$N_1 \geqslant \cdots \geqslant N_d > \frac{\rho_3}{\exp(K^{C_1})} \geqslant N_{d+1} \geqslant \cdots \geqslant N_{|S|}. \tag{9.71}$$

From (9.69), (8.3) we see that d cannot equal $|S|$. Let V be the d-dimensional subspace of $\mathbb{R}^S$ spanned by $v_1, \ldots, v_d$, let $V^\perp$ be the orthogonal complement of V in $\mathbb{R}^S$, and let $\pi : \mathbb{R}^S \to V^\perp$ be the orthogonal projection.

We claim that $\pi(\tilde{\mu}(n, m))$ is now uniquely determined by $\mu(n, m)$ for n, $m \in B(S_1, \rho_3/4)$. Indeed, if $\tilde{\mu}(n, m)$ and $\tilde{\mu}'(n, m)$ both obeyed (9.67), (9.68), then their difference (call it w) would be of magnitude $O(1/\rho_3)$ and lies in the kernel of ϕ. By (9.70) with $t = \exp(-K^{C_1})\rho_3$, we conclude that w lies in V, and hence $\pi(\tilde{\mu}(n, m))$ and $\pi(\tilde{\mu}'(n, m))$ agree.

A variant of the above argument shows that $\pi \circ \tilde{\mu}$ also continues to obey the cocycle equation.

LEMMA 9.19 (Projected lift is a cocycle). *One has*

$$\pi(\tilde{\mu}(n_1, n_2 + n_3)) + \pi(\tilde{\mu}(n_2, n_3)) = \pi(\tilde{\mu}(n_1, n_2)) + \pi(\tilde{\mu}(n_1 + n_2, n_3))$$

and additionally

$$\pi(\tilde{\mu}(n_1, n_2)) = \pi(\tilde{\mu}(n_2, n_1)); \qquad \pi(\tilde{\mu}(n_1, 0)) = 0$$

for all $n_1, n_2, n_3 \in B(S_1, \rho_3/4)$.

Proof. By (9.68), the quantity $w := \tilde{\mu}(n_1, n_2 + n_3) + \tilde{\mu}(n_2, n_3) - \tilde{\mu}(n_1, n_2) - \tilde{\mu}(n_1 + n_2, n_3)$ has magnitude $O(1/\rho_3)$; by (9.67), (9.65), w lies in the kernel of ϕ. Repeating the previous arguments, we conclude that $w \in V$. Applying the homomorphism π, we obtain the first claim. The second claim is proven similarly. □

We can in fact make $\pi \circ \tilde{\mu}$ a coboundary, after shrinking the domain somewhat.

PROPOSITION 9.20 (Projected lift is a coboundary). *There exists a map* $F : B(S_1, 2\exp(-K^{C_1^2})\rho_3) \to V^\perp$ *with*

$$F(n) \ll \frac{K^{O(C_1)}}{\rho_3} \tag{9.72}$$

for all $n \in B(S_1, 2\exp(-K^{C_1^2})\rho_3)$, *such that*

$$\pi(\tilde{\mu}(n_1, n_2)) = F(n_1 + n_2) - F(n_1) - F(n_2)$$

for all $n_1, n_2 \in B(S_1, \exp(-K^{C_1^2})\rho_3)$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Proof. As a first attempt at constructing F, we introduce the average

$$F_1(n) := \mathbb{E}\pi(\tilde{\mu}(n, \mathbf{n}_3))$$

for $n \in B(S_1, \rho_3/4)$, where $\mathbf{n}_3$ is drawn regularly from $B(S_1, \rho_3/4)$. From (9.68) we have

$$|F_1(n)| \leqslant \frac{24}{\rho_3}$$

for all $n \in B(S_1, \rho_3/4)$. Also, since $|S_1| \ll K^{O(C_1)}$, if we replace n_3 by $\mathbf{n}_3$ in Lemma 9.19 and take expectations using Lemma 4.4, we conclude that

$$F_1(n_1) + F_1(n_2) = \pi(\tilde{\mu}(n_1, n_2)) + F_1(n_1 + n_2) + O\left(\frac{K^{O(C_1)}\|n_2\|_{S_1^\perp}}{\rho_3^2}\right)$$

for all $n_1, n_2 \in B(S_1, \rho_3/8)$.

If we now introduce the modified cocycle

$$\sigma_1(n_1, n_2) := \pi(\tilde{\mu}(n_1, n_2)) + F_1(n_1 + n_2) - F_1(n_1) - F_1(n_2)$$

for $n_1, n_2 \in B(S_1, \rho_3/8)$, then we have the cocycle equation

$$\sigma_1(n_1, n_2 + n_3) + \sigma_1(n_2, n_3) = \sigma_1(n_1, n_2) + \sigma_1(n_1 + n_2, n_3), \tag{9.73}$$

the auxiliary equations

$$\sigma_1(n_1, n_2) = \sigma_1(n_2, n_1); \qquad \sigma_1(n_1, 0) = 0$$

and the bound

$$\sigma_1(n_1, n_2) \ll \frac{K^{O(C_1)}\|n_2\|_{S_1^\perp}}{\rho_3^2} \tag{9.74}$$

for $n_1, n_2 \in B(S_1, \rho_3/16)$.

We now make σ_1 a coboundary by using a basis for $B(S_1, \rho_3/16)$. Set $d := |S_1| \leqslant K^{O(C_1)}$. By Corollary 4.9, we can find $a_1, \ldots, a_d$ of $\mathbb{Z}/p\mathbb{Z}$ and real numbers $N_1, \ldots, N_d > 0$ such that

$$\|a_i\|_{S_1^\perp} \leqslant N_i^{-1} \tag{9.75}$$

for all $i = 1, \ldots, d$, and such that for any $a \in \mathbb{Z}/p\mathbb{Z}$, there exists a representation

$$a = m_1 a_1 + \cdots + m_d a_d \tag{9.76}$$

with $m_1, \ldots, m_d$ integers of size

$$m_i \ll \exp(O(K^{O(C_1)}))N_i\|a\|_{S_1^\perp} \tag{9.77}$$

for $i = 1, \ldots, d$, with at most one such representation obeying the bounds $|m_i| < N_i/2$ for $i = 1, \ldots, d$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

By relabelling we may assume that $N_i \geqslant 32d'/\rho_3$ for $i = 1, \ldots, d'$ and $N_i < 32d'/\rho_3$ for $i = d'+1, \ldots, d$ for some $0 \leqslant d' \leqslant d$. By (9.75) we have $a_i \in B(S_1, \rho_3/32d')$ for all $i = 1, \ldots, d'$. In particular, from (9.73) we see that for any $n \in B(S_1, \rho_3/32)$ and $1 \leqslant i, j \leqslant d'$, we have

$$\sigma_1(n_1, a_i + a_j) + \sigma_1(a_i, a_j) = \sigma_1(n_1, a_i) + \sigma_1(n_1 + a_i, a_j)$$

and hence by swapping i and j and subtracting

$$\sigma_1(n_1 + a_j, a_i) - \sigma_1(n_1, a_i) = \sigma_1(n_1 + a_i, a_j) - \sigma_1(n_1, a_j).$$

Let $P \subset \mathbb{Z}^{d'}$ denote the collection of tuples $(m_1, \ldots, m_{d'}) \in \mathbb{Z}^{d'}$ with $|m_i| \leqslant \rho_3/2N_i$ for $i = 1, \ldots, d'$, and for each $m \in P$ and $i = 1, \ldots, d$, define the quantity

$$f_i(m) := \sigma_1(\phi(m), a_i)$$

where $\phi : \mathbb{Z}^{d'} \to \mathbb{Z}/p\mathbb{Z}$ is the homomorphism

$$\phi(m_1, \ldots, m_{d'}) := \sum_{k=1}^{d'} m_k a_k.$$

Then from (9.75) we have $\phi(P) \subset B(S_1, \rho_3/32)$. The above identity then says that the "1-form" $(f_1, \ldots, f_{d'})$ is "closed" or "curl-free" in the sense that

$$f_i(m + e_j) - f_i(m) = f_j(m + e_i) - f_j(m) \tag{9.78}$$

whenever $i, j = 1, \ldots, d'$ and $m, m + e_i, m + e_j \in P$, where $e_1, \ldots, e_{d'}$ is the standard basis for P. This implies that there exists a function $H : P \to V^{\perp}$ such that $F(0) = 0$ and $f_i(m) = H(m + e_i) - H(m)$ whenever $i = 1, \ldots, d$ and $m, m + e_i \in P$. Indeed, one can define H to be an "antiderivative" of the $(f_1, \ldots, f_{d'})$ by setting

$$H(m) := \sum_{l=0}^{L-1} f_{i_l}(m_l)$$

whenever $0 = m_0, \ldots, m_L = m$ is a path in P with $m_{l+1} = m_l + e_{i_l}$ for $l = 0, \ldots, L - 1$; a "homotopy" argument using (9.78) shows that the right-hand side does not depend on the choice of path. From (9.74), (9.75) we have

$$f_i(m) \ll \frac{K^{O(C_1)}}{N_i \rho_3^2}$$

for $m \in P$ and $i = 1, \ldots, d'$, which on "integrating" (and recalling that $d' \leqslant d \ll K^{O(C_1)}$) implies that

$$H(m) \ll \frac{K^{O(C_1)}}{\rho_3}$$

for all $m \in P$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Since $\sigma_1(0, e_i) = 0$, we have $f_i(0) = 0$ and hence $H(e_i) = 0$ for all $i = 1, \dots, d'$. Thus we have

$$\sigma_1(\phi(m), \phi(e_i)) = H(m + e_i) - H(m) - H(e_i)$$

whenever $m, m + e_i \in P$. An induction (on the magnitude of a vector m') using (9.73) then shows that

$$\sigma_1(\phi(m), \phi(m')) = H(m + m') - H(m) - H(m')$$

whenever $m, m', m+m' \in P$. Now, if $n \in B(S_1, 2\exp(-K^{C_1^2})\rho)$, then by (9.76), (9.77) we see that $n = \phi(m)$ for some $m \in P$. If we then define $F_2 : B(S_1, 2\exp(-K^{C_1^2})\rho) \to V^\perp$ by setting $F_2(n) := H(m)$, we conclude that

$$F_2(n) \ll \frac{K^{O(C_1)}}{\rho_3}$$

and

$$\sigma_1(n, n') = F_2(n + n') - F_2(n) - F_2(n')$$

for all $n, n' \in B(S_1, \exp(-K^{C_1^2})\rho)$. Setting $F := F_2 - F_1$, we obtain the claim. □

Let F be as in Proposition 9.20. We use F to construct the locally bilinear form $\Xi : B(S_1, \rho_4) \times B(S_1, \rho_4) \to \mathbb{R}/\mathbb{Z}$ as follows. We first define the locally linear map $\iota : B(S_1, \rho_4) \to \mathbb{R}^S$ by the formula

$$\iota(m) := \left(\left\{ \frac{ms}{p} \right\} \right)_{s \in S},$$

where $x \mapsto \{x\}$ is the signed fractional map from $\mathbb{R}/\mathbb{Z}$ to $(-1/2, 1/2]$; note that ι takes values in the box $[-\rho_4, \rho_4]^S$. We then define

$$\Xi(n, m) := \frac{\xi''(n)m}{p} - F(n) \cdot \iota(m) \tag{9.79}$$

for $n, m \in B(S_1, \rho_4)$, where $\cdot$ denotes the dot product on $\mathbb{R}^S$. It is clear that Ξ is locally linear in m; we also claim that it is locally linear in n, thus

$$\Xi(n_1 + n_2, m) - \Xi(n_1, m) - \Xi(n_2, m) = 0 \tag{9.80}$$

whenever $n_1, n_2, n_1 + n_2 \in B(S_1, \rho_4)$. By (9.64) and Proposition 9.20, the left-hand side of (9.80) may be written as

$$\frac{\mu(n_1, n_2)m}{p} - \pi(\tilde{\mu}(n_1, n_2)) \cdot \iota(m) \bmod 1.$$

From (9.67) we have

$$\frac{\mu(n_1, n_2)m}{p} = \tilde{\mu}(n_1, n_2) \cdot \iota(m) \bmod 1$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

so to prove (9.80), it suffices to show that $\iota(m)$ lies in $V^\perp$. This is equivalent to showing that $\iota(m) \cdot v_i = 0$ for $i = 1, \dots, d$. Since $v_i \in \ker(\phi)$, we have

$$\iota(m) \cdot v_i = 0 \bmod 1.$$

On the other hand, we have $\iota(m) = O(K^{1/2}\rho_4)$, and from (9.70) with $t = N_i^{-1}$ followed by (9.71), we have

$$|v_i| \leqslant N_i^{-1} < \frac{\exp(K^{C_1})}{\rho_3}$$

and hence $|\iota(m) \cdot v_i| < 1$. The claim follows.

Now we verify (9.63). Let a_0, ξ_0 be as in Proposition 9.16. Let $\mathbf{n}, \mathbf{n}_0, \mathbf{h}$, $\mathbf{n}_1$, $\mathbf{m}_1$ be drawn independently and regularly from the Bohr sets $B(S_1, \rho_3/4)$, $B(S, \rho_0)$, $B(S_1, \rho_4)$, $B(S_1, \rho_6)$, $B(S_1, \rho_5)$ respectively. From Proposition 9.16 we have

$$\sum_{n_0,n} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n} = n)|\mathbb{E} f(n_0 + \mathbf{h} + a_0 - n)\overline{f}(n_0 + \mathbf{h}) e_p((\xi''(n) - \xi_0)\mathbf{h})|^2 \gg \eta^{C_1+O(1)}.$$

Using Lemma 4.4 to replace $\mathbf{n}$ by $\mathbf{n} + \mathbf{n}_1$, and to replace $\mathbf{h}$ by $\mathbf{h} + \mathbf{m}_1$, we have

$$\sum_{n_0,n,n_1} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n} = n, \mathbf{n}_1 = n_1)|\mathbb{E} f(n_0 + \mathbf{h} + \mathbf{m}_1 + a_0 - n - n_1) \times \overline{f}(n_0 + \mathbf{h} + \mathbf{m}_1) e_p((\xi''(n + n_1) - \xi_0)(\mathbf{h} + \mathbf{m}_1))|^2 \gg \eta^{C_1+O(1)}$$

and thus by the triangle inequality we have

$$\sum_{n_0,n,n_1,h} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n} = n, \mathbf{n}_1 = n_1, \mathbf{h} = h)|\mathbb{E} f(n_0 + h + \mathbf{m}_1 + a_0 - n - n_1) \times \overline{f}(n_0 + h + \mathbf{m}_1) e_p((\xi''(n + n_1) - \xi_0)(h + \mathbf{m}_1))|^2 \gg \eta^{C_1+O(1)}.$$

The phase $e((\xi''(n + n_1) - \xi_0)h)$ is deterministic and may thus be omitted:

$$\sum_{n_0,n,n_1,h} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n} = n, \mathbf{n}_1 = n_1, \mathbf{h} = h)|\mathbb{E} f(n_0 + h + \mathbf{m}_1 + a_0 - n - n_1) \times \overline{f}(n_0 + h + \mathbf{m}_1) e_p((\xi''(n + n_1) - \xi_0)\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)}.$$

As the expectation only depends on the sum $n_0 + h$ rather than the individual variables n_0, h, we thus have

$$\sum_{n_0,n,n_1} \mathbb{P}(\mathbf{n}_0 + \mathbf{h} = n_0, \mathbf{n} = n, \mathbf{n}_1 = n_1)|\mathbb{E} f(n_0 + \mathbf{m}_1 + a_0 - n - n_1) \times \overline{f}(n_0 + \mathbf{m}_1) e_p((\xi''(n + n_1) - \xi_0)\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)}.$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

By Lemma 4.4 we may replace $\mathbf{n}_0 + \mathbf{h}$ here by $\mathbf{n}_0$. From (9.57) we have

$$\|(\xi''(n+n_1) - \xi''(n) - \xi''(n_1))\mathbf{m}_1\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{100C_1}$$

and so

$$\sum_{n_0,n,n_1} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n} = n, \mathbf{n}_1 = n_1)|\mathbb{E}f(n_0 + a_0 + \mathbf{m}_1 - n - n_1) \times \overline{f}(n_0 + \mathbf{m}_1)e_p((\xi''(n) + \xi''(n_1) - \xi_0)\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)}.$$

By the pigeonhole principle, there thus exists $n \in B(S_*, \rho_3/4)$ such that

$$\sum_{n_0,n_1} \mathbb{P}(\mathbf{n}_0 = n_0\mathbf{n}_1 = n_1)|\mathbb{E}f(n_0 + a_0 + \mathbf{m}_1 - n - n_1)\overline{f}(n_0 + \mathbf{m}_1) \times e_p((\xi''(n) + \xi''(n_1) - \xi_0)\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)},$$

which, if we write $a_1 := a_0 - n$ and $\xi_1 := \xi_0 - \xi''(n)$, simplifies to

$$\sum_{n_0,n_1} \mathbb{P}(\mathbf{n}_0 = n_0\mathbf{n}_1 = n_1)|\mathbb{E}f(n_0 + \mathbf{m}_1 + a_1 - n_1)\overline{f}(n_0 + \mathbf{m}_1) \times e_p((\xi''(n_1) - \xi_1)\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)}.$$

Since $a_0 \in B(S, 3\rho_2)$ and $n \in B(S_*, \rho_3/4)$, we have $a_1 \in B(S, 4\rho_2)$.

Now, from (9.79) one has

$$e_p(\xi''(n_1)\mathbf{m}_1) = e(\Xi(n_1, \mathbf{m}_1))e(-F(n_1) \cdot \iota(\mathbf{m}_1));$$

but since $\mathbf{m}_1 \in B(S_*, \rho_5)$, we have $\iota(\mathbf{m}_1) = O(K\rho_5)$, and hence by (9.72) we have

$$\|F(n_1) \cdot \iota(\mathbf{m}_1)\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{100C_1},$$

and so

$$\sum_{n_0,n_1} \mathbb{P}(\mathbf{n}_0 = n_0; \mathbf{n}_1 = n_1)|\mathbb{E}f(n_0 + \mathbf{m}_1 + a_1 - n_1)\overline{f}(n_0 + \mathbf{m}_1) \times e(\Xi(n_1, \mathbf{m}_1) - \xi_1\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)},$$

which gives (9.63). The proof of Theorem 9.18 is now complete.

9.21. *Eighth step: making the frequency function symmetric.* The next step is the "symmetry step" from [**14**, **26**], which uses the Cauchy–Schwarz inequality to ensure that Ξ is essentially symmetric.

THEOREM 9.22. *Let the notation and hypotheses be as in Theorem 9.18. For $n, m \in B(S_1, \rho_4)$, define*

$$\{n, m\} := \Xi(n, m) - \Xi(m, n).$$

Then there exists a natural number k with $1 \leqslant k \ll \exp(K^{O(C_1)})$ such that

$$\|k\{n, m\}\|_{\mathbb{R}/\mathbb{Z}} \leqslant \frac{\|n\|_{S_1^\perp}}{\rho_8}\frac{\|m\|_{S_1}}{\rho_8}$$

for all $n, m \in B(S_1, \rho_9)$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Proof. Let $\mathbf{n}_0, \mathbf{m}_1, \mathbf{n}_1$ be as in Theorem 9.18. From (9.63) and the pigeonhole principle, we may find $n_0 \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{n_1} \mathbb{P}(\mathbf{n}_1 = n_1)|\mathbb{E}f(n_0 + \mathbf{m}_1 + a_1 - n_1)\overline{f}(n_0 + \mathbf{m}_1) \times e(\Xi(n_1, \mathbf{m}_1) - \xi_1\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)}$$

which by the boundedness of the expectation implies

$$\sum_{n_1} \mathbb{P}(\mathbf{n}_1 = n_1)|\mathbb{E}f(n_0 + \mathbf{m}_1 + a_1 - n_1)\overline{f}(n_0 + \mathbf{m}_1) \times e(\Xi(n_1, \mathbf{m}_1) - \xi_1\mathbf{m}_1)| \gg \eta^{C_1+O(1)}$$

and thus we may find a 1-bounded function $b_1 : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ such that

$$|\mathbb{E}b_1(\mathbf{n}_1)f(n_0 + \mathbf{m}_1 + a_1 - \mathbf{n}_1)\overline{f}(n_0 + \mathbf{m}_1)e(\Xi(\mathbf{n}_1, \mathbf{m}_1) - \xi_1\mathbf{m}_1)| \gg \eta^{C_1+O(1)}.$$

Writing $b_2(n) := f(n_0 + a_1 + n)$ and $b_3(n) := \overline{f}(n_0 + \mathbf{m}_1)e(-\xi_1\mathbf{m}_1)$, we may simplify this as

$$|\mathbb{E}b_1(\mathbf{n}_1)b_2(\mathbf{m}_1 - \mathbf{n}_1)b_3(\mathbf{m}_1)e(\Xi(\mathbf{n}_1, \mathbf{m}_1))| \gg \eta^{C_1+O(1)}.$$

Using the Cauchy–Schwarz inequality (Lemma 2.1) to eliminate the $b_3(\mathbf{m}_1)$ factor, we conclude that

$$|\mathbb{E}b_1(\mathbf{n}_1)\overline{b_1}(\mathbf{n}_1')b_2(\mathbf{m}_1-\mathbf{n}_1)\overline{b_2}(\mathbf{m}_1-\mathbf{n}_1')e(\Xi(\mathbf{n}_1, \mathbf{m}_1)-\Xi(\mathbf{n}_1', \mathbf{m}_1))| \gg \eta^{2C_1+O(1)}$$

where $\mathbf{n}_1'$ is an independent copy of $\mathbf{n}_1$. Writing $\mathbf{k} := \mathbf{n}_1 + \mathbf{n}_1' - \mathbf{m}_1$, and noting from the local bilinearity of Ξ that

$$\begin{aligned}\Xi(\mathbf{n}_1, \mathbf{m}_1) - \Xi(\mathbf{n}_1', \mathbf{m}_1) &= \Xi(\mathbf{n}_1 - \mathbf{n}_1', \mathbf{m}_1)\\ &= \Xi(\mathbf{n}_1 - \mathbf{n}_1', \mathbf{n}_1 + \mathbf{n}_1' - \mathbf{k})\\ &= \Xi(\mathbf{n}_1, \mathbf{n}_1) - \Xi(\mathbf{n}_1', \mathbf{n}_1') + \{\mathbf{n}_1, \mathbf{n}_1'\}\\ &\quad - \Xi(\mathbf{n}_1, \mathbf{k}) + \Xi(\mathbf{n}_1', \mathbf{k})\end{aligned}$$

we conclude that

$$|\mathbb{E}b_3(\mathbf{n}_1, \mathbf{k})b_4(\mathbf{n}_1', \mathbf{k})e(\{\mathbf{n}_1, \mathbf{n}_1'\})| \gg \eta^{2C_1+O(1)},$$

where $b_3, b_4 : \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ are the 1-bounded functions

$$b_3(n_1, k) := b_1(n_1)\overline{b_2}(k - n_1)e(\Xi(n_1, n_1) - \Xi(n_1, k))$$

and

$$b_4(n_1', k) := \overline{b_1}(n_1')b_2(k - n_1')e(-\Xi(n_1', n_1') + \Xi(n_1', k)).$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

For fixed $\mathbf{n}_1, \mathbf{n}_1'$, we see from Lemma 4.4 that $\mathbf{k}$ differs from $\mathbf{m}_1$ in total variation by $O(\eta^{100C_1})$, and hence

$$|\mathbb{E}b_3(\mathbf{n}_1, \mathbf{m}_1)b_4(\mathbf{n}_1', \mathbf{m}_1)e(\{\mathbf{n}_1, \mathbf{n}_1'\})| \gg \eta^{2C_1+O(1)}.$$

By the pigeonhole principle, we may thus find $m_1 \in \mathbb{Z}/p\mathbb{Z}$ such that

$$|\mathbb{E}b_3(\mathbf{n}_1, m_1)b_4(\mathbf{n}_1', m_1)e(\{\mathbf{n}_1, \mathbf{n}_1'\})| \gg \eta^{2C_1+O(1)}.$$

Using Cauchy–Schwarz (Lemma 2.1) to eliminate $b_4(\mathbf{n}_1', m_1)$, and using the local bilinearity of $\{,\}$, we conclude that

$$|\mathbb{E}b_3(\mathbf{n}_1, m_1)\overline{b_3}(\mathbf{l}_1, m_1)e(\{\mathbf{n}_1 - \mathbf{l}_1, \mathbf{n}_1'\})| \gg \eta^{4C_1+O(1)},$$

where $\mathbf{l}_1$ is an independent copy of $\mathbf{n}_1$; using a further application of Cauchy–Schwarz (Lemma 2.1) to eliminate $b_3(\mathbf{n}_1, m_1)\overline{b_3}(\mathbf{l}_1, m_1)$, we conclude that

$$|\mathbb{E}e(\{\mathbf{n}_1 - \mathbf{l}_1, \mathbf{n}_1' - \mathbf{l}_1'\})| \gg \eta^{8C_1+O(1)},$$

where $\mathbf{l}_1'$ is an independent copy of $\mathbf{n}_1'$ (thus $\mathbf{n}_1, \mathbf{n}_1', \mathbf{l}_1, \mathbf{l}_1'$ are jointly independent and drawn regularly from $B(S_1, \rho_6)$). In particular, by the pigeonhole principle one can find $l_1, l_1' \in B(S_1, \rho_6)$ such that

$$|\mathbb{E}e(\{\mathbf{n}_1 - l_1, \mathbf{n}_1' - l_1'\})| \gg \eta^{8C_1+O(1)}.$$

By local bilinearity, one can rewrite $\{\mathbf{n}_1 - l_1, \mathbf{n}_1' - l_1'\}$ as $\{\mathbf{n}_1, \mathbf{n}_1'\}$ plus locally linear functions of $\mathbf{n}_1$ and $\mathbf{n}_1'$. The claim now follows from Proposition 4.11. □

9.23. *Ninth step: integrating the frequency function.* We may now finally prove Theorem 8.1. Let the notation and hypotheses be as in that theorem, let S_1 and Ξ be as in Theorem 9.18, and let k be as in Theorem 9.22. Thus if we let $\mathbf{n}_0, \mathbf{n}_1, \mathbf{m}_1$ be drawn independently and regularly from $B(S, \rho_0)$, $B(S_1, \rho_6)$, $B(S_1, \rho_5)$ respectively, we have

$$\sum_{n_0, n_1} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_1 = n_1)|\mathbb{E}f(n_0 + m_1 + a_1 - n_1)\overline{f}(n_0 + \mathbf{m}_1) \times e(\Xi(n_1, \mathbf{m}_1) - \xi_1\mathbf{m}_1)|^2 \gg \eta^{C_1+O(1)}. \tag{9.81}$$

Now let $\mathbf{n}_2, \mathbf{m}_2$ be drawn independently and regularly from the Bohr sets $B(S_1, \rho_9)$, $B(S_1, \rho_{10})$ respectively, independently of all previous random variables. By Lemma 4.4, we may replace $\mathbf{n}_1, \mathbf{m}_1$ by $\mathbf{n}_1 + 2k\mathbf{n}_2$ and $\mathbf{m}_1 + 2k\mathbf{m}_2$ in (9.81), leading to

$$\sum_{n_0, n_1, n_2} \mathbb{P}(\mathbf{n}_0 = n_0, \dots, \mathbf{n}_2 = n_2)|\mathbb{E}f(n_0 + \mathbf{m}_1 + 2k\mathbf{m}_2 + a_1 - n_1 - 2kn_2) \times \overline{f}(n_0 + \mathbf{m}_1 + 2k\mathbf{m}_2)e(\Xi(n_1 + 2kn_2, \mathbf{m}_1 + 2k\mathbf{m}_2) - \xi_1(\mathbf{m}_1 + 2k\mathbf{m}_2))|^2 \gg \eta^{C_1+O(1)}.$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Thus we may find $n_1 \in B(S_1, \rho_6)$, $m_1 \in B(S_1, \rho_5)$ such that

$$\sum_{n_0,n_2} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_2 = n_2)|\mathbb{E} f(n_0 + m_1 + 2k\mathbf{m}_2 + a_1 - n_1 - 2kn_2) \\ \times \overline{f}(n_0 + m_1 + 2k\mathbf{m}_2)e(\Xi(n_1 + 2kn_2, m_1 + 2k\mathbf{m}_2) - \xi_1(m_1 + 2k\mathbf{m}_2))|^2 \\ \gg \eta^{C_1+O(1)},$$

which we can simplify slightly as

$$\sum_{n_0,n_2} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_2 = n_2)|\mathbb{E} f(n_0 + 2k\mathbf{m}_2 + a_2 - 2kn_2) \\ \times \overline{f}(n_0 + m_1 + 2k\mathbf{m}_2)e(\Xi(n_1 + 2kn_2, m_1 + 2k\mathbf{m}_2) - 2k\xi_1\mathbf{m}_2)|^2 \\ \gg \eta^{C_1+O(1)},$$

where $a_2 := a_1 + m_1 - n_1$; since $a_1 \in B(S, 4\rho_2), m_1 \in B(S_1, \rho_5), n_1 \in B(S_1, \rho_6)$, we have $a_2 \in B(S, 5\rho_2)$. By the local bilinearity of Ξ, we have

$$\begin{aligned}\Xi(n_1 + 2kn_2, m_1 + 2k\mathbf{m}_2) &= \Xi(n_1, m_1) + 2k\Xi(n_2, m_1) + 2k\Xi(n_1, \mathbf{m}_2) + 4k^2\Xi(n_2, \mathbf{m}_2) \\ &= \Xi(n_1, m_1) + 2k\Xi(n_2, m_1) + 2k\Xi(n_1, \mathbf{m}_2) + 2k^2\Xi(n_2 + \mathbf{m}_2, n_2 + \mathbf{m}_2) \\ &\quad - 2k^2\Xi(n_2, n_2) - 2k^2\Xi(\mathbf{m}_2, \mathbf{m}_2) + 2k^2\{n_2, \mathbf{m}_2\}\end{aligned}$$

and so we have

$$\sum_{n_0,n_2} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_2 = n_2)|\mathbb{E} F(n_0, n_2 - \mathbf{m}_2)G(n_0, \mathbf{m}_2)e(2k^2\{n_2, \mathbf{m}_2\})|^2 \\ \gg \eta^{C_1+O(1)},$$

where

$$F(n, m) := f(n + a_2 - 2km)e(-k^2\Xi(m, m)) \tag{9.82}$$

and

$$G(n, m) := \overline{f}(n + m_1 + 2km)e(2k\Xi(n_1, m) - 2k^2\Xi(m, m) - 2k\xi_1 m).$$

By Theorem 9.22, one has $\|k\{\mathbf{n}_2, \mathbf{m}_2\}\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{100C_1}$, and thus

$$\sum_{n_0,n_2} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_2 = n_2)|\mathbb{E} F(n_0, n_2 - \mathbf{m}_2)G(n_0, \mathbf{m}_2)|^2 \gg \eta^{C_1+O(1)}.$$

By boundedness of the expectation, this implies that

$$\sum_{n_0,n_2} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_2 = n_2)|\mathbb{E} F(n_0, n_2 - \mathbf{m}_2)G(n_0, \mathbf{m}_2)| \gg \eta^{C_1+O(1)}$$

and thus

$$|\mathbb{E} F(\mathbf{n}_0, \mathbf{n}_2 - \mathbf{m}_2)G(\mathbf{n}_0, \mathbf{m}_2)H(\mathbf{n}_0, \mathbf{n}_2)| \gg \eta^{C_1+O(1)}$$

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

for some 1-bounded function $H : \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$. By Cauchy–Schwarz (Lemma 2.1), we thus have

$$|\mathbb{E}F(\mathbf{n}_0, \mathbf{n}_2 - \mathbf{m}_2)G(\mathbf{n}_0, \mathbf{m}_2)\overline{F}(\mathbf{n}_0, \mathbf{n}_2 - \mathbf{m}_2')\overline{G}(\mathbf{n}_0, \mathbf{m}_2)| \gg \eta^{2C_1+O(1)},$$

where $\mathbf{m}_2'$ is an independent copy of $\mathbf{m}_2$; by a second application of Cauchy–Schwarz (Lemma 2.1), we then have

$$|\mathbb{E}F(\mathbf{n}_0, \mathbf{n}_2-\mathbf{m}_2)\overline{F}(\mathbf{n}_0, \mathbf{n}_2-\mathbf{m}_2')\overline{F}(\mathbf{n}_0, \mathbf{n}_2'-\mathbf{m}_2)F(\mathbf{n}_0, \mathbf{n}_2'-\mathbf{m}_2')| \gg \eta^{4C_1+O(1)},$$

where $\mathbf{n}_2'$ is an independent copy of $\mathbf{n}_2$. Since the distributions of $\mathbf{m}_2, \mathbf{m}_2'$ are symmetric, we thus have

$$|\mathbb{E}F(\mathbf{n}_0, \mathbf{n}_2+\mathbf{m}_2)\overline{F}(\mathbf{n}_0, \mathbf{n}_2+\mathbf{m}_2')\overline{F}(\mathbf{n}_0, \mathbf{n}_2'+\mathbf{m}_2)F(\mathbf{n}_0, \mathbf{n}_2'+\mathbf{m}_2')| \gg \eta^{4C_1+O(1)}.$$

In particular, with probability $\gg\eta^{4C_1+O(1)}$, the random variable $\mathbf{n}_0$ attains a value n_0 for which

$$|\mathbb{E}F(n_0, \mathbf{n}_2+\mathbf{m}_2)\overline{F}(n_0, \mathbf{n}_2+\mathbf{m}_2')\overline{F}(n_0, \mathbf{n}_2'+\mathbf{m}_2)F(n_0, \mathbf{n}_2'+\mathbf{m}_2')| \gg \eta^{4C_1+O(1)}. \tag{9.83}$$

If n_0 is such that (9.83) holds, then we may apply Theorem 4.12 and conclude that there exists a frequency $\beta(n_0) \in \mathbb{Z}/p\mathbb{Z}$ such that

$$\left|\sum_{n_2} \mathbb{P}(\mathbf{n}_2 = n_2)\mathbb{E}F(n_0, n_2 + \mathbf{m}_2)e(-\beta(n_0)\mathbf{m}_2)\right| \gg \eta^{2C_1+O(1)}$$

and thus (defining $\beta(n_0)$ arbitrarily if (9.83) does not hold),

$$\sum_{n_0,n_2} \mathbb{P}(\mathbf{n}_0 = n_0, \mathbf{n}_2 = n_2)|\mathbb{E}F(n_0, n_2 + \mathbf{m}_2)e(-\beta(n_0)\mathbf{m}_2)| \gg \eta^{6C_1+O(1)}$$

and hence there exists $n_2 \in B(S_1, \rho_9)$ with

$$\sum_{n_0} \mathbb{P}(\mathbf{n}_0 = n_0)|\mathbb{E}F(n_0, n_2 + \mathbf{m}_2)e(-\beta(n_0)\mathbf{m}_2)| \gg \eta^{6C_1+O(1)}.$$

Applying (9.82), we conclude that

$$\sum_{n_0} \mathbb{P}(\mathbf{n}_0 = n_0)|\mathbb{E}f(n_0 + a_3 - 2k\mathbf{m}_2)e(-k^2\Xi(\mathbf{m}_2, \mathbf{m}_2) - \beta(n_0)\mathbf{m}_2)| \gg \eta^{6C_1+O(1)},$$

where $a_3 := a_2 - 2kn_2$; since $a_2 \in B(S, 5\rho_2)$, $n_2 \in B(S_1, \rho_9)$, and $k = O(\exp(K^{O(C_1)}))$, we have $a_3 \in B(S, 6\rho_2)$. In particular, by Lemma 4.4, $\mathbf{n}_0$ and $\mathbf{n}_0 + a_3$ differ in total variation by $O(\eta^{100C_1+O(1)})$, and thus

$$\sum_{n_0} \mathbb{P}(\mathbf{n}_0 = n_0)|\mathbb{E}f(n_0 - 2k\mathbf{m}_2)e(-k^2\Xi(\mathbf{m}_2, \mathbf{m}_2) - \beta(n_0)\mathbf{m}_2)| \gg \eta^{6C_1+O(1)}.$$

Theorem 8.1 then follows after a change of variables, noting that the map $\mathbf{m}_2 \mapsto \Xi(\mathbf{m}_2, \mathbf{m}_2)$ is locally quadratic on $B(S_1, \rho_9)$.

https://doi.org/10.1112/S0025579317000316 Published online by Cambridge University Press

Acknowledgements. The first author is supported by a Simons Investigator grant. The second author is supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1266164. Part of this paper was written while the authors were in residence at MSRI in spring 2017, which is supported by NSF grant DMS-1440140.

We are indebted to the anonymous referee for helpful corrections and suggestions. Finally, we thank any readers interested in the result of this paper for their patience. Most of the argument was worked out by us in 2005, and the result was claimed in [**19**], dedicated to Roth's 80th birthday. While a complete, though not very readable, version has been available on request since around 2012, it has taken us until now to create a potentially publishable manuscript.

References

1. A. Balog and E. Szemerédi, A statistical theorem of set addition. *Combinatorica* **14**(3) (1994), 263–268.
2. F. A. Behrend, On sets of integers which contain no three terms in arithmetic progression. *Proc. Nat. Acad. Sci.* **32** (1946), 331–332.
3. V. Bergelson, B. Host and B. Kra, Multiple recurrence and nilsequences. With an appendix by Imre Ruzsa. *Invent. Math.* **160**(2) (2005), 261–303.
4. T. F. Bloom, A quantitative improvement for Roth's theorem on arithmetic progressions. *J. Lond. Math. Soc.* (2) **93**(3) (2016), 643–663.
5. M. Blum, M. Luby and R. Rubinfeld, Self-testing/correcting with applications to numerical problems. *J. Comput. System Sci.* **47**(3) (1993), 549–595, Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (Baltimore, MD, 1990).
6. J. Bourgain, On triples in arithmetic progression. *Geom. Funct. Anal.* **9**(5) (1999), 968–984.
7. J. Bourgain, Roth's theorem on progressions revisited. *J. Anal. Math.* **104** (2008), 155–192.
8. M. Elkin, An improved construction of progression-free sets. *Israel J. Math.* **184** (2011), 93–128.
9. P. Erdős, Problems in number theory and combinatorics. In *Proceedings of the Sixth Manitoba Conference on Numerical Mathematics (University of Manitoba, Winnipeg, MB, 1976)* (*Congress. Numer. XVIII*), Utilitas Math. (Winnipeg, MB, 1977), 35–58.
10. P. Erdős and P. Turán, On some sequences of integers. *J. Lond. Math. Soc.* **11** (1936), 261–264.
11. W. T. Gowers, A new proof of Szemerédi's theorem for progressions of length four. *Geom. Funct. Anal.* **8**(3) (1998), 529–551.
12. W. T. Gowers, A new proof of Szemerédi's theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
13. B. J. Green, A Szemerédi-type regularity lemma in abelian groups, with applications. *Geom. Funct. Anal.* **15**(2) (2005), 340–376.
14. B. J. Green and T. C. Tao, An inverse theorem for the Gowers $U^3(G)$-norm. *Proc. Edinb. Math. Soc.* (2) **51**(1) (2008), 73–153.
15. B. J. Green and T. C. Tao, New bounds for Szemerédi's theorem, I: progressions of length 4 in finite field geometries. *Proc. Lond. Math. Soc.* (3) **98**(2) (2009), 365–392.
16. B. J. Green and T. C. Tao, New bounds for Szemerédi's theorem, Ia: progressions of length 4 in finite field geometries revisited. *Preprint*, 2012, arXiv:1205.1330.
17. B. J. Green and T. C. Tao, Quadratic uniformity of the Möbius function. *Ann. Inst. Fourier (Grenoble)* **58**(6) (2008), 1863–1935.
18. B. J. Green and T. C. Tao, An arithmetic regularity lemma, an associated counting lemma, and applications. In *An Irregular Mind* (*Bolyai Soc. Math. Stud.* **21**), János Bolyai Mathematical Society (Budapest, 2010), 261–334.
19. B. J. Green and T. C. Tao, New bounds for Szemerédi's theorem, II: a new bound for $r_4(N)$. In *Analytic Number Theory: Essays in Honour of Klaus Roth*, (eds W. W. L. Chen, W. T. Gowers, H. Halberstam, W. M. Schmidt and R. C. Vaughan), Cambridge University Press (Cambridge, 2009), 180–204.
20. B. J. Green and J. Wolf, A note on Elkin's improvement of Behrend's construction. In *Additive Number Theory*, Springer (New York, 2010), 141–144.
21. D. R. Heath-Brown, Integer sets containing no arithmetic progressions. *J. Lond. Math. Soc.* **35** (1987), 385–394.

22. I. Łaba and M. Lacey, On sets of integers not containing long arithmetic progressions. *Preprint*, 2001, arXiv:0108155.
23. R. A. Rankin, Sets of integers containing not more than a given number of terms in arithmetic progression. *Proc. Roy. Soc. Edinburgh Sect. A* **65** (1960/1961), 332–344.
24. K. F. Roth, On certain sets of integers. *J. Lond. Math. Soc.* **28** (1953), 245–252.
25. K. F. Roth, Irregularities of sequences relative to arithmetic progressions, IV. *Period. Math. Hungar.* **2** (1972), 301–326.
26. A. Samorodnitsky, Low-degree tests at large distances. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, ACM (New York, 2007), 506–515.
27. T. Sanders, On certain other sets of integers. *J. Anal. Math.* **116** (2012), 53–82.
28. T. Sanders, On Roth's theorem on progressions. *Ann. of Math.* (2) **174**(1) (2011), 619–636.
29. E. Szemerédi, On sets of integers containing no four elements in arithmetic progression. *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89–104.
30. E. Szemerédi, On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 299–345.
31. E. Szemerédi, Regular partitions of graphs. In *Problémes combinatoires et théorie des graphes (Colloq. Internat. CNRS, University of Orsay, Orsay, 1976)* (*Colloq. Internat. CNRS* **260**), CNRS (Paris, 1978), 399–401.
32. E. Szemerédi, Integer sets containing no arithmetic progressions. *Acta Math. Hungar.* **56**(1–2) (1990), 155–158.
33. T. C. Tao and V. H. Vu, *Additive Combinatorics* (*Cambridge Studies in Advanced Mathematics* **105**), Cambridge University Press (Cambridge, 2006).
34. T. C. Tao and V. H. Vu, John-type theorems for generalized arithmetic progressions and iterated sumsets. *Adv. Math.* **219** (2008), 428–449.
35. T. Ziegler, A non-conventional ergodic theorem for a nilsystem. *Ergod. Th. & Dynam. Sys.* **25**(4) (2005), 1357–1370.

Ben Green,
Mathematical Institute,
Andrew Wiles Building,
Radcliffe Observatory Quarter,
Woodstock Rd,
Oxford OX2 6GG,
U.K.
E-mail: ben.green@maths.ox.ac.uk

Terence Tao,
Department of Mathematics,
UCLA,
Los Angeles CA 90095-1555,
U.S.A.
E-mail: tao@math.ucla.edu