# HIGHER UNIFORMITY OF BOUNDED MULTIPLICATIVE FUNCTIONS IN SHORT INTERVALS ON AVERAGE

KAISA MATOMÄKI, MAKSYM RADZIWIŁŁ, TERENCE TAO, JONI TERÄVÄINEN, AND TAMAR ZIEGLER

Abstract. Let $\lambda$ denote the Liouville function. We show that, as $X \to \infty$,

$$\int_X^{2X} \sup_{\substack{P(Y)\in\mathbb{R}[Y]\\ \deg P\le k}} \left| \sum_{x\le n\le x+H} \lambda(n)e(-P(n)) \right| dx = o(XH)$$

for all fixed $k$ and $X^\theta \le H \le X$ with $0 < \theta < 1$ fixed but arbitrarily small. Previously this was only established for $k \le 1$. We obtain this result as a special case of the corresponding statement for (non-pretentious) 1-bounded multiplicative functions that we prove.

In fact, we are able to replace the polynomial phases $e(-P(n))$ by degree $k$ nilsequences $\overline{F}(g(n)\Gamma)$. By the inverse theory for the Gowers norms this implies the higher order asymptotic uniformity result

$$\int_X^{2X} \|\lambda\|_{U^{k+1}([x,x+H])}\, dx = o(X)$$

in the same range of $H$.

We present applications of this result to patterns of various types in the Liouville sequence. Firstly, we show that the number of sign patterns of the Liouville function is superpolynomial, making progress on a conjecture of Sarnak about the Liouville sequence having positive entropy. Secondly, we obtain cancellation in averages of $\lambda$ over short polynomial progressions $(n + P_1(m), \dots, n + P_k(m))$, which in the case of linear polynomials yields a new averaged version of Chowla's conjecture.

We are in fact able to prove our results on polynomial phases in the wider range $H \ge \exp((\log X)^{5/8+\varepsilon})$, thus strengthening also previous work on the Fourier uniformity of the Liouville function.

## Contents

## 1. Introduction

Let $\lambda\colon \mathbb{N} \to \{-1,+1\}$ denote the Liouville function, that is to say the completely multiplicative function with $\lambda(p) = -1$ for all primes $p$; we extend $\lambda$ by zero to the integers. In [23] it was shown that this function exhibited cancellation on almost all short intervals $[x, x+H]$ in the sense that[1]

$$\int_X^{2X} \left| \sum_{x \leq n \leq x+H} \lambda(n) \right| dx = o(HX) \tag{1}$$

as $X \to \infty$, whenever $H = H(X)$ was a function of $X$ that went to infinity as $X \to \infty$; see also [22] for a simpler proof of (1) in the case of "polynomially large intervals", in which $H = X^\theta$ for a fixed $0 < \theta < 1$. In [23], [22] the qualitative gain $o(HX)$ over the trivial bound $O(HX)$ was improved to a more quantitative bound, but in this paper we will focus only on qualitative estimates. The bounds for $\lambda$ also extend to the closely related Möbius function $\mu$, but for the sake of discussion we shall restrict attention initially to the Liouville function $\lambda$.

In [25] the estimate (1) was generalized to

$$\sup_{\alpha \in \mathbb{R}} \int_X^{2X} \left| \sum_{x \leq n \leq x+H} \lambda(n) e(-\alpha n) \right| dx = o(HX) \tag{2}$$

as $X \to \infty$, for any $H = H(X)$ that went to infinity as $X \to \infty$, where we adopt the usual notation $e(\alpha) := e^{2\pi i \alpha}$. Informally, this asserts that $\lambda$ does not asymptotically exhibit any correlation with a *fixed* linear phase $n \mapsto e(\alpha n)$ in short intervals on average. The question was then raised in [34, Section 4] as to whether the stronger estimate

$$\int_X^{2X} \sup_{\alpha \in \mathbb{R}} \left| \sum_{x \leq n \leq x+H} \lambda(n) e(-\alpha n) \right| dx = o(HX) \tag{3}$$

could be established. This is not known unconditionally, although as observed in [35] it can be deduced from the Chowla conjecture [4]. However, in a recent paper [26] the bound (3) was established in the regime where $H = X^\theta$ for a fixed $0 < \theta < 1$; the case $\theta > 5/8$ without needing the $x$-average was previously established by Zhan in [41] (and Zhan's result was recently improved to $\theta > 3/5$ in [28]).

[1] See Section 2 for our asymptotic notation conventions.

For any non-negative integer $k \geq 0$, any interval $[x, x+H]$, and any function $f\colon \mathbb{Z} \to \mathbb{C}$, define the *weak Gowers uniformity norm*

$$\|f\|_{u^{k+1}([x,x+H])} := \sup_{P \in \mathrm{Poly}_{\leq k}(\mathbb{R}\to\mathbb{R})} \frac{1}{H} \left| \sum_{x \leq n \leq x+H} \lambda(n) e(-P(n)) \right| \tag{4}$$

where $\mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is the $k+1$-dimensional vector space of polynomial maps[2] $P\colon \mathbb{R} \to \mathbb{R}$ of degree at most $k$. This norm is indeed much weaker than the usual Gowers norm, in the sense that it is well-known (see [11, §4]) that it does not control linear patterns of complexity $\geq 2$. Nevertheless, we will need the weak Gowers uniformity result in Theorem 1.3 below in order to establish the strong Gowers uniformity result in Theorem 1.5.

The result in [26] is then equivalent to the bound

$$\int_X^{2X} \|\lambda\|_{u^2([x,x+H])}\, dx = o(X)$$

as $X \to \infty$, with $H = X^\theta$ for a fixed $0 < \theta < 1$; the corresponding (and weaker) bound for the $u^1$ norm follows from the earlier result in [23] or [22]. Our first main result extends these bounds to higher orders of uniformity:

**Corollary 1.1** (Liouville does not correlate with polynomial phases on short intervals on average)**.** *Let $k \geq 0$ be a non-negative integer, and let $0 < \theta < 1$ be fixed. Then we have*

$$\int_X^{2X} \|\lambda\|_{u^{k+1}([x,x+H])}\, dx = o(X) \tag{5}$$

*as $X \to \infty$, where $H := X^\theta$.*

**Remark 1.2.** *In Theorem 1.8 below we show that Corollary 1.1 holds for $H$ as small as $\exp((\log X)^{5/8+\varepsilon})$ for any fixed $\varepsilon > 0$.*

We remark that previously this was known in the $k \geq 1$ cases for $\theta > 2/3$ by [27, Theorem 1.4]. In fact, in this regime a uniform bound $\sup_{x\in[X,2X]} \|\lambda\|_{u^{k+1}([x,x+H])} = o(1)$ is established. It is natural to conjecture that such uniform bounds extend to all $\theta > 0$, but this seems well beyond the reach of the methods in this paper.

In fact (as in [26]), we can generalize Corollary 1.1 to the case where the Liouville function $\lambda$ is replaced by a more general "non-pretentious" 1-bounded multiplicative function. Recall that a multiplicative function $f\colon \mathbb{N} \to \mathbb{C}$ is said to be 1-bounded if $|f(n)| \leq 1$ for all $n \in \mathbb{N}$. To motivate the "non-pretentiousness" hypothesis, we consider (as was done in [26] in the $k = 1$ case) the character

$$f(n) := n^{it}\chi(n), \tag{6}$$

formed by multiplying an "Archimedean character" $n \mapsto n^{it}$ for some real number $t$ with $|t| \leq \varepsilon X^{k+1}/H^{k+1}$ for some small $\varepsilon > 0$, and a Dirichlet character $\chi$ of some bounded conductor $q$. Observe that $f$ is completely multiplicative and 1-bounded, and a Taylor

[2] In the sum in (4), only the values of $P$ on the integers $\mathbb{Z}$ are relevant, but in our later analysis it will be convenient to evaluate such polynomials at non-integer values as well.

expansion with remainder of the phase $n \mapsto \frac{t}{2\pi}\log n$ of the Archimedean character $n^{it} = e(\frac{t}{2\pi}\log n)$ around a given point $x \in [X, 2X]$ yields a decomposition of the form

$$n^{it} = e(P_x(n)) + O_k(\varepsilon) \tag{7}$$

for all $n \in [x, x+H]$ and some polynomial $P_x \in \mathrm{Poly}_{\le k}(\mathbb{R} \to \mathbb{R})$ depending on $x$ (and $t$). This together with the $q$-periodicity of $\chi$ can be used to imply that

$$\int_X^{2X} \|f\|_{u^{k+1}([x,x+H])}\, dx \gg_{k,q} X$$

if $1 \le H \le X$ are sufficiently large.

Our next result asserts that this is essentially the only obstruction to extending Corollary 1.1 to more general 1-bounded multiplicative functions. Following Granville and Soundararajan [12], we define the distance function

$$\mathbb{D}(f, g; X) := \Big(\sum_{p \le X} \frac{1 - \mathrm{Re}(f(p)\overline{g(p)})}{p}\Big)^{1/2},$$

and further define the quantity

$$M(f; X, Q) := \inf_{|t| \le X} \inf_{\substack{\chi \bmod q \\ q \le Q}} \mathbb{D}(f, n \mapsto \chi(n)n^{it}; X).$$

Informally, $M(f; X, Q)$ is small whenever $f$ is close to a function of the form (6) with $|t| \le X$ and $\chi$ of conductor at most $Q$. We then have

**Theorem 1.3** (Non-pretentious multiplicative functions do not correlate with polynomial phases on short intervals on average)**.** *Let $k \ge 0$ be a non-negative integer, and let $0 < \theta < 1/2$. Suppose that $f\colon \mathbb{N} \to \mathbb{C}$ is a multiplicative 1-bounded function, and suppose that $X \ge 1$, $X^\theta \le H \le X^{1-\theta}$, and $\eta > 0$ are such that*

$$\int_X^{2X} \|f\|_{u^{k+1}([x,x+H])}\, dx \ge \eta X.$$

*Then one has*

$$M(f; CX^{k+1}/H^{k+1}, Q) \ll_{k,\eta,\theta} 1$$

*for some $Q, C \ll_{k,\eta,\theta} 1$.*

The upper bound $H \le X^{1-\theta}$ here is for minor technical reasons and it is likely that one can replace it with $H \le X$; however our main interest is in the opposite regime when $H$ is as small as possible. Standard calculations regarding the "non-pretentious" nature of the Liouville function (using the Vinogradov–Korobov zero-free region for $L$-functions) allow one to deduce Corollary 1.1 from Theorem 1.3; see for instance [25, (1.12)]. The $k = 0$ case of this theorem follows from the results in [23], and the $k = 1$ case is established[3] in [26,

[3]In that paper the constant $C$ appearing in the above theorem was worsened to $H^\rho$ for some arbitrarily small constant $\rho > 0$, but we have found a way to modify the arguments to eliminate that power loss in this result. In fact, it will be important in the induction arguments used to establish Theorem 1.5 below that such losses are avoided.

Theorem 1.4]. Our focus here shall accordingly be on the higher order case $k \geq 2$, which we will establish by generalizing the techniques in [26] to the polynomial phase setting (and in fact further to nilsequences, which are needed in proving our Theorem 1.5 on genuine Gowers norms of multiplicative functions).

As a corollary of Theorem 1.3 and the decomposition (7) we can also control the correlation of non-pretentious multiplicative functions with Archimedean characters on short intervals on average:

**Corollary 1.4** (Non-pretentious multiplicative functions do not correlate with Archimedean characters on short intervals on average)**.** *Let $k \geq 0$ be a non-negative integer, and let $0 < \theta < 1/2$. Suppose that $f\colon \mathbb{N} \to \mathbb{C}$ is a multiplicative 1-bounded function, and suppose that $X \geq 1$, $\varepsilon > 0$, $X^\theta \leq H \leq X^{1-\theta}$, and $\eta > 0$ are such that*

$$\int_X^{2X} \sup_{|t| \leq \varepsilon X^{k+1}/H^{k+1}} \left| \sum_{x \leq n \leq x+H} f(n) n^{it} \right| dx \geq \eta H X.$$

*Then one has*

$$M(f; CX^{k+1}/H^{k+1}, Q) \ll_{k,\eta,\varepsilon,\theta} 1$$

*for some $Q, C \ll_{k,\eta,\varepsilon,\theta} 1$.*

We also note that He and Wang [19] recently proved that

$$\sup_{P \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})} \int_X^{2X} \left| \sum_{x \leq n \leq x+H} \lambda(n) e(-P(n)) \right| dx = o(HX)$$

for any $H$ tending to infinity with $X$, and they also proved an analogous estimate for nilsequences. This statement with the supremum *outside* the integral unfortunately does not lead to control on Gowers norms (or weak Gowers norms) of $\lambda$ over short intervals and is accordingly closer in spirit to [25] than to the current paper. It is the case with the supremum *inside* the integral (as in Theorems 1.3 and 4.3) that we need for the applications in this paper, and such estimates would lead to a proof of the logarithmically averaged Chowla and Sarnak conjectures (via [35, Theorem 1.8]) if one was able to take the interval length $H$ to grow sufficiently slowly in them; see Proposition 1.7.

As indicated above, we can strengthen Theorem 1.3 further. For any non-negative integer $k \geq 0$, and any function $f\colon \mathbb{Z} \to \mathbb{C}$ with finite support, define the (unnormalized) Gowers uniformity norm

$$\|f\|_{U^{k+1}(\mathbb{Z})} := \left( \sum_{y, h_1, \dots, h_{k+1} \in \mathbb{Z}} \prod_{\omega \in \{0,1\}^{k+1}} \mathcal{C}^{|\omega|} f(y + \omega_1 h_1 + \cdots + \omega_{k+1} h_{k+1}) \right)^{1/2^{k+1}}$$

where $\omega = (\omega_1, \dots, \omega_{k+1})$, $|\omega| := \omega_1 + \cdots + \omega_{k+1}$, and $\mathcal{C}\colon z \mapsto \overline{z}$ is the complex conjugation map. Then for any interval $[x, x+H]$ with $H \geq 1$ and any $f\colon \mathbb{Z} \to \mathbb{C}$ (not necessarily of

finite support), define the *Gowers uniformity norm over* $[x, x+H]$ by

$$\|f\|_{U^{k+1}([x,x+H])} := \|f1_{[x,x+H]}\|_{U^{k+1}(\mathbb{Z})}/\|1_{[x,x+H]}\|_{U^{k+1}(\mathbb{Z})} \tag{8}$$

where $1_{[x,x+H]}\colon \mathbb{Z} \to \mathbb{C}$ is the indicator function of $[x, x+H]$. We then have

**Theorem 1.5** (Non-pretentious multiplicative functions are Gowers uniform on short intervals on average)**.** *Let $k \geq 0$ be a non-negative integer, and let $0 < \theta < 1/2$. Suppose that $f\colon \mathbb{N} \to \mathbb{C}$ is a multiplicative 1-bounded function (extended by zero to the remaining integers), and suppose that $X \geq 1$, $X^\theta \leq H \leq X^{1-\theta}$, and $\eta > 0$ are such that*

$$\int_X^{2X} \|f\|_{U^{k+1}([x,x+H])}\, dx \geq \eta X.$$

*Then one has*

$$M(f; CX^{k+1}/H^{k+1}, Q) \ll_{k,\eta,\theta} 1 \tag{9}$$

*for some $Q, C \ll_{k,\eta,\theta} 1$.*

The corresponding statement on correlations of $f$ with nilsequences $n \mapsto F(g(n)\Gamma)$ on intervals $[x, x+H]$, which we will use to derive Theorem 1.5 (and which in fact is equivalent to it), is given as Theorem 4.3.

In particular, using the non-pretentious nature of the Liouville function, this theorem yields the following corollary.

**Corollary 1.6** (Gowers uniformity of Liouville on short intervals on average)**.** *Let an integer $k \geq 0$ and $0 < \theta \leq 1$ be fixed. Then for $H \geq X^\theta$ we have*

$$\int_X^{2X} \|\lambda\|_{U^{k+1}([x,x+H])}\, dx = o(X). \tag{10}$$

Note that in the corollary above the case of larger values of $H \geq X^{1-o(1)}$ follows from the case $H = X^\theta$ by a simple averaging argument (by first using the inverse theorem for the Gowers norms to express (10) in terms of the correlation of $\lambda$ with nilsequences on $[x, x+H]$, and then partitioning this interval into subintervals of length $\asymp X^{1-\varepsilon}$). This partially verifies [35, Conjecture 1.6], which asserted that this estimate (or more precisely, a slightly weaker logarithmically averaged version of this estimate) held whenever $H = H(X)$ went to infinity as $X \to \infty$. Fully resolving this conjecture would have many implications, including the (logarithmically averaged) Chowla and Sarnak conjectures; see [34], [37] and [10] for the best currently known results in this direction). Correspondingly, the partial result (10) allows us to make progress on some problems concerning the Liouville function, including its word complexity and an averaged version of Chowla's conjecture, which we discuss in Subsection 1.2.

Regarding previous results on Gowers norms of non-pretentious multiplicative functions, a result of Frantzikinakis and Host [9] (generalizing work of Green and Tao [15]) establishes the "long sum" endpoint case of Theorem 1.5 (corresponding to the case $H = X$, which is strictly speaking not covered by the above theorem), showing that $\|f\|_{U^{k+1}[1,X]} = o(1)$

under the assumption that $\mathbb{D}(f, n \mapsto \chi(n)n^{it}; X) \to \infty$ as $X \to \infty$ for any fixed real number $t$ and Dirichlet character $\chi$.

It is not difficult to establish a general estimate of the form

$$\|f\|_{u^{k+1}([x,x+H])} \ll_k \|f\|_{U^{k+1}([x,x+H])}$$

for any $f\colon \mathbb{Z} \to \mathbb{C}$; this can be established for instance by a minor modification of the arguments in [38, §11.2]. Thus Theorem 1.5 implies Theorem 1.3. The converse implication is also routine for $k = 0, 1$, but as is now well known (see e.g., [38, Proposition 11.8]), for higher $k$ the polynomial phases $n \mapsto e(P(n))$ appearing in the definition of the weak Gowers norms (4) are insufficient to control the full Gowers norms (8). To bridge the gap, one needs to replace these polynomial phases by more general *nilsequences* $n \mapsto F(g(n)\Gamma)$. The polynomial phases correspond to nilsequences on filtered nilmanifolds $G/\Gamma$ with $G$ abelian. We will thus first prove Theorem 1.3 in Section 3 to treat the case of abelian $G$, and then use a different and more delicate argument (presented in Section 4 and outlined in Subsection 1.3) to handle the non-abelian case.

1.1. **Connection with the Chowla and Sarnak conjectures.** As already mentioned, estimates such as (10) with slowly growing $H$ are closely tied to the Chowla and Sarnak conjectures. The logarithmically averaged Chowla conjecture states that

$$\sum_{n \leq x} \frac{\lambda(a_1 n + b_1) \cdots \lambda(a_k n + b_k)}{n} = o(\log x)$$

whenever $a_i, b_i$ are natural numbers[4] with $a_i b_j \neq a_j b_i$ for $i \neq j$. The logarithmically averaged Sarnak conjecture in turn is the statement that

$$\sum_{n \leq x} \frac{\lambda(n) a(n)}{n} = o(\log x)$$

for every bounded, deterministic sequence $a\colon \mathbb{N} \to \mathbb{C}$ (in the sense that $a$ has zero topological entropy). See [6] for a survey of previous work on these two conjectures.

In [35], it was shown that the logarithmically averaged Chowla conjecture and the logarithmically averaged Sarnak conjecture are equivalent, and that both would also follow from (10) being true for every $H = H(X)$ tending to infinity with $X$. In fact these two conjectures are equivalent to the logarithmic version of (10) in this regime, which states that

$$\int_1^X \frac{\|\lambda\|_{U^{k+1}[x,x+H]}}{x}\, dx = o(\log X) \tag{11}$$

whenever $H = H(X)$ goes to infinity with $X$. Thus, a potential strategy towards proving the logarithmic Chowla and Sarnak conjectures emerges from the possibility of lowering the value of $H = H(X)$ in Theorem 1.5. We observe in Section 5.3 that we in fact do not need (11) for arbitrarily slowly growing $H$ to deduce the logarithmic Chowla conjecture; it instead suffices to prove it for $H \geq (\log X)^\eta$ for every $\eta > 0$.

[4]In this paper the natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$ begin with 1.

**Proposition 1.7.** *Suppose that for every natural number $k$ and every $\eta > 0$ for $H = H(X) = (\log X)^\eta$ we have*

$$\int_1^X \frac{\|\lambda\|_{U^{k+1}[x,x+H]}}{x}\,dx = o(\log X).$$

*Then the logarithmic Chowla conjecture holds.*

This proposition will be proved in Subsection 5.3.

Thus, in order to prove the logarithmic Chowla conjecture, it would suffice to bridge the gap between $H \geq X^\eta$ (which is the range where Corollary 1.6 is valid) and $H \leq (\log X)^\eta$ in Proposition 1.7. In Section 6, we already show that, at least in the case of our result on the weak Gowers norms (Theorem 1.3), we may lower the admissible $H$ to $H \geq \exp((\log X)^c)$ for some $c > 0$.

**Theorem 1.8** (Shortening the intervals)**.** *Let $k$ be a natural number, and let $\theta > 5/8$ and $\rho > 0$ be fixed. Suppose that $f\colon \mathbb{N} \to \mathbb{C}$ is a multiplicative 1-bounded function (extended by zero to the remaining integers), and suppose that $X \geq 1$, $X^\theta \geq H \geq \exp((\log X)^\theta)$, and $\eta > 0$ are such that*

$$\int_X^{2X} \|f\|_{u^{k+1}([x,x+H])}\,dx \geq \eta X. \tag{12}$$

*Then one has*

$$M(f; X^{k+1}/H^{k+1-\rho}, Q) \ll_{k,\eta,\theta} 1 \tag{13}$$

*for some $Q \ll_{k,\eta,\theta,\rho} 1$.*

It is conceivable that a careful reworking of the nilsequence part of our arguments in Section 4 would yield a similar regime $H \geq \exp((\log X)^{1-\delta})$ for Theorem 1.5; we do not pursue this here, however (see however Remark 4.27).

The exponent $5/8$ appearing in Theorem 1.8 is significant as it shows that it is possible to control $\|f\|_{u^{k+1}[x,x+H]}$ on average over $x$ without establishing cancellations in short sums over primes of the form $\sum_{H\leq p\leq 2H} p^{it}$ (with $t$ of size $X^k$). Instead, we show using general Dirichlet polynomial techniques that the set of points $t$ at which the above Dirichlet polynomial exhibits no cancellation is sparse. We note that the smallest $H$ for which $\sum_{H\leq p\leq 2H} p^{it}$ is known to exhibit cancellations for $t$ of size $X^k$ is $H = \exp((\log X)^{2/3+\varepsilon})$. We also note that the proof of Theorem 1.8 crucially relies on cancellation in short sums of multiplicative functions outside a power-saving exceptional set, proved in [24] as an improvement to [23]. See Remark 6.7 on how in the case of $f = \lambda$ the weaker range $H \geq \exp((\log X)^{2/3+\varepsilon})$ and can be obtained using only the method of [22].

It seems nonetheless that the lower bound for $H$ in Theorem 1.8 is close to the breaking point of several arguments in our proof. Firstly, for $H$ much smaller than $\exp((\log X)^c)$ with $c > 0$ it appears difficult to show (using general Dirichlet polynomial techniques) that for a large proportion of values $|t| \leq X^{O(1)}$ the sum $\sum_{H\leq p\leq 2H} p^{it}$ exhibits cancellations. Secondly, in the graph-theoretic part of our arguments factors of the type $\ell!$ with $\ell \asymp \frac{\log X}{\log H}$ arise, and while these are harmless for $H \geq X^\eta$, they become problematic in the regime

$H \leq \exp((\log X)^\theta)$, in particular if $\theta < 1/2$. Despite these limitations, at least if one works with certain model cases of the problem (such as a "99% version" of Theorem 1.5, where $\eta$ is very close to 1) and assumes GRH, then one should be able to push $H$ further down.

Handling the regime $H \in [(\log X)^\eta, (\log X)^{\eta^{-1}}]$, at the very least, would likely necessitate an entirely new idea for several reasons. Firstly, even under GRH cancellation in the Dirichlet polynomials $\sum_{H^\varepsilon \leq p \leq 2H^\varepsilon} \chi(p)p^{it}$ is known essentially only for $H \gg_{\varepsilon,\kappa} (\log X)^{(2+\kappa)\varepsilon^{-1}}$. Secondly, the arguments for solving "approximate functional equations" involving phase functions that are used in this paper do not seem to work (even in model cases) for such $H$, as such arguments rely on the "modulus" $\prod_{H^\varepsilon \leq p \leq 2H^\varepsilon} p$ being much larger than $X$ (see footnote 14). Thirdly, the entropy decrement argument (which is applied to prove Proposition 1.7 that the $H \geq (\log X)^\eta$ range of (11) implies the logarithmic Chowla conjecture) is restricted to the regime $H \leq (\log X)^\eta$, as it is based on equidistribution of the integers in $[1, X]$ modulo $\prod_{H^A \leq p \leq 2H^A} p$ for $A \geq 1$ large enough (see however the recent work [1] for a quantitatively stronger alternative replacement to the entropy decrement method in the case of two-point correlations).

### 1.2. **Applications.**

1.2.1. *Sign patterns of the Liouville function.* Let

$$s(k) := |\{v \in \{-1,+1\}^k : \ v = (\lambda(n+1), \dots, \lambda(n+k)) \text{ for some } n \in \mathbb{N}\}| \tag{14}$$

be the number of sign patterns of length $k$ in the Liouville sequence. A direct consequence of Chowla's conjecture (or its logarithmic version) is that $s(k) = 2^k$ for all $k$ and that each pattern of length $k$ occurs with positive lower density; yet, this remains unknown (apart from the $k \leq 4$ cases handled in [37]). In fact, known lower bounds on $s(k)$ are far from exponential; Frantzikinakis and Host [10] proved that $s(k)/k \to \infty$ as $k \to \infty$, and recently this was improved by McNamara [29] to $s(k) \gg k^2$. In fact, both in [29] and [10] a stronger result was proved, namely that $\lambda$ is orthogonal (with logarithmic averages) to any sequence having $o(k^2)$ (respectively $O(k)$) sign patterns of length $k$. Let us also remark that the validity of the $2j$-point Chowla conjecture for any fixed $j$ implies by a simple moment computation that there are $\gg k^j$ sign patterns of length $k$ that occur with positive density (so, in particular, $s(k) \gg k^j$). As an application of Theorem 1.5, we prove a superpolynomial lower bound on $s(k)$.

**Theorem 1.9** (The Liouville function has superpolynomial number of patterns)**.** *We have $s(k) \gg_A k^A$ for every $A \geq 1$.*

In fact, we prove a more general result (Theorem 5.1), which shows that any improvement in the range of validity of (10) leads to an improvement in the lower bound on $s(k)$. In particular, if (10) holds for $H \geq \exp((\log X)^{1-\delta})$, then $s(k) \gg_\varepsilon k^{(\log k)^{\delta/(1-\delta)-\varepsilon}}$. See also Theorem 5.4 for a generalization to multiplicative functions other than the Liouville function.

Theorem 1.9 can be viewed as progress towards a conjecture of Sarnak in [32] that the Furstenberg systems of the Liouville function have positive entropy (so that in particular

$s(k) \gg c^k$ for some $c > 1$). Sarnak highlighted this as a key special case of his Möbius randomness conjecture. It is worth noting that, as was observed in [32], one easily sees that the Möbius system has positive entropy, but this amounts solely to the fact that the distribution of squarefree numbers is very well understood and therefore this does not imply anything about the Liouville system (indeed, Sarnak says in [32] that it appears "pretty hard to show that $\lambda$ is not deterministic"). In this connection, it would be very interesting to say more about the frequency of the superpolynomially many patterns produced by Theorem 1.9.

The proof of Theorem 1.9 involves a different approach than the previous sign pattern arguments, utilizing a type of "structure and randomness" dichotomy (meaning that if there are few sign patterns, then the Liouville function is easier to understand, and we can leverage this to eventually get a contradiction); see Section 5 for the proof and Subsection 1.3.2 for its outline.

1.2.2. *Polynomial averages of the Liouville function.* As another application of Theorem 1.5, we use it to establish cancellation in averages

$$\mathbb{E}_{n\le X}\mathbb{E}_{m\le X^{1/d}}\lambda(n+P_1(m))\cdots\lambda(n+P_k(m))$$

of the Liouville function along polynomial progressions $(n+P_1(m),\dots,n+P_k(m))$ (with $d$ being the maximum degree of the polynomials $P_i$). Averages along polynomial progressions are natural objects in additive combinatorics and ergodic theory, and a particularly important result concerning them is the polynomial Szemerédi theorem of Bergelson and Leibman [2] that guarantees for any non-constant polynomials $P_i(x) \in \mathbb{Z}[x]$ with $P_i(0) = 0$ the existence of a polynomial progression $n + P_1(m),\dots,n + P_k(m)$ inside any positive density subset of the integers. This was generalized to polynomial progressions inside the primes in [39]. However, when one is considering polynomial progressions weighted by an oscillating function (such as $\lambda$), these results do not apply (as they are lower bound results).

It was later shown in [40, Theorem 1.4] that if the assumption $P_i(0) = 0$ for all $i$ is replaced with the polynomials $P_i - P_j$ having degree $d$ for all $i \neq j$ (where $d$ is the maximum of the degrees of $P_l$) one has an asymptotic for polynomial patterns $(n + P_1(m),\dots,n + P_k(m))$ weighted by the von Mangoldt function (and the same argument works for the Liouville function). Here we remove this assumption on the degree $d$ coefficients of the $P_i$ being distinct in the case of the Liouville weight, thus obtaining a result that works for any polynomial patterns (that are not of "infinite complexity", such as the pattern $(n+1, n+2,\dots,n+k)$). Moreover, we can take the $m$ average in our results to be of subpolynomial size, which is important for Corollary 1.11 below.

**Theorem 1.10** (Polynomial averages of the Liouville function)**.** *Let $k, r \ge 1$ be integers, and let $P_1,\dots,P_k$ be polynomials in $\mathbb{Z}[x_1,\dots,x_r]$ with degrees $\le d$. Suppose that $P_i - P_j$ is nonconstant for all $i \neq j$. Then for any fixed $0 < \varepsilon < 1/d$ we have*

$$\mathbb{E}_{\mathbf{m}\in[X^\varepsilon]^r}|\mathbb{E}_{n\le X}\lambda(n+P_1(\mathbf{m}))\cdots\lambda(n+P_k(\mathbf{m}))| = o(1).$$

*Here, $[N]^r$ stands for the $r$-dimensional discrete box $\{1,\dots,N\}^r$.*

Specializing to linear polynomials, the following result on Chowla's conjecture with a short one-variable average is an immediate corollary (in fact, this corollary could also be obtained more directly from our Gowers uniformity result, Corollary 1.6; see footnote 21).

**Corollary 1.11** (Chowla's conjecture with a short average)**.** *Let $k \geq 1$ be an integer, and let $a_1, \dots, a_k \geq 0$ be distinct. Let $\varepsilon > 0$ be arbitrary. Then we have*

$$\mathbb{E}_{h \leq X^\varepsilon} |\mathbb{E}_{n \leq X} \lambda(n + a_1 h) \cdots \lambda(n + a_k h)| = o(1).$$

We remark that the Theorem 1.10 (and hence Corollary 1.11) continues to hold, with essentially the same proof, if $k - 1$ of the $k$ occurrences of $\lambda$ in the correlation average are replaced with arbitrary fixed 1-bounded sequences.

Taking $h$ bounded in Corollary 1.11 would amount to settling Chowla's conjecture. Previously, the result of Corollary 1.11 was only known for $k \leq 2$ (using the main result of [25]), and for $k = 3$ without the absolute values around the $n$ average (using [26, Corollary 1.5]). Note that for $k \geq 3$ the averaged Chowla conjecture of [25] is not applicable in the setting above, since that result requires averaging over $k - 1$ independent short variables.

We can also prove an asymptotic similar to the one in Theorem 1.10 for averages of the von Mangoldt function if one of the terms in the progression is assigned the Liouville weight (but perhaps surprisingly the proof does not apply if the weight $\lambda$ is replaced with the constant weight 1).

**Theorem 1.12** (Polynomial averages of the von Mangoldt function with Liouville twist)**.** *Let $k, r \geq 1$ be integers, and let $P_1, \dots, P_k$ be non-constant polynomials in $\mathbb{Z}[x_1, \dots, x_r]$ with degrees $\leq d$. Suppose that $P_i - P_j$ is nonconstant for all $i \neq j$. Let $\Lambda$ be the von Mangoldt function. Then for any fixed $0 < \varepsilon < 1/d$ we have*

$$\mathbb{E}_{\mathbf{m} \in [X^\varepsilon]^r} |\mathbb{E}_{n \leq X} \lambda(n + P_1(\mathbf{m})) \Lambda(n + P_2(\mathbf{m})) \cdots \Lambda(n + P_k(\mathbf{m}))| = o(1).$$

We remark that the theorem continues to hold, with essentially the same proof, when the occurrences of $\Lambda$ in the correlation average are replaced with arbitrary fixed sequences that are bounded by $\Lambda$ in modulus.

These results will be established in Section 7.

## 1.3. **Overview of proofs.**

1.3.1. *The higher order uniformity theorem.* Let us outline the proof of Corollary 1.6; the proof of the more general Theorem 1.5 follows along similar lines. By the inverse theorem for the Gowers norms, Corollary 1.6 is equivalent to a discorrelation estimate between the Liouville function and nilsequences; more precisely

$$\int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{Z} \to G)} \left| \sum_{n \in [x, x+H]} \lambda(n) \overline{F}(g(n)\Gamma) \right| dx = o(HX), \tag{15}$$

where $G/\Gamma$ is any fixed[5] degree $k$ filtered nilmanifold, $F : G/\Gamma \to \mathbb{C}$ is any fixed Lipschitz function, and the supremum is over all polynomial sequences $g(n)$ taking values in the Lie group $G$ (for all the relevant definitions and for the precise statement, see Section 4).

By using an induction on the dimension of $G$ we may assume that the function $F$ is "irreducible" in a certain technical sense, which roughly means that the nilsequences $n \mapsto F(g(n)\Gamma)$ are "orthogonal" to all lower dimensional nilsequences. We split the proof of this estimate (15) into two cases that are analyzed separately, the case of abelian $G$ and the case of non-abelian $G$.

For abelian $G$, the nilsequences that arise on the filtered nilmanifold $G/\Gamma$ are (Lipschitz functions of) polynomial phases $n \mapsto e(P(n))$ with $\deg(P) \le k$, so this case reduces to the polynomial phase case. This case is handled in Section 3 and is already sufficient for proving Corollary 1.1. Here the task is to establish structure in phase functions $P_x \in \mathrm{Poly}_{\le k}(\mathbb{Z} \to \mathbb{R})$ satisfying

$$\left| \sum_{n \in [x, x+H]} \lambda(n) e(P_x(n)) \right| dx \gg H \tag{16}$$

for $\gg X$ choices of $x \in [X, 2X] \cap \mathbb{Z}$, and eventually to exploit that structure to show that such functions do not exist. In order to talk about polynomials being equal up to negligible contributions, we introduce an equivalence relation on them; in this sketch, we say that $P_x \sim Q_x$ if[6] $P_x(n) = \varepsilon(n) Q_x(n)$ holds on the underlying interval $[x, x+H]$ for some polynomial $\varepsilon(n)$ which is "smooth" in the sense that $|\varepsilon^{(\ell)}(n)| \ll H^{-\ell}$ for all $\ell \le k$. Note that if we can show that

$$e(P_x(n)) \approx e\left(\frac{T}{2\pi} \log n + \gamma(n)\right), \quad n \in [x, x+H], \tag{17}$$

with $\gamma \in \mathrm{Poly}_{\le k}(\mathbb{R} \to \mathbb{R})$ being a $O(1)$-integral polynomial (that is, it maps from $q\mathbb{Z}$ to $\mathbb{Z}$ for some $q = O(1)$) and $T$ independent of $x$ and of polynomial size in $X$, then $e(P_x(n))$ is essentially a twist of the Archimedean character $n \mapsto n^{iT}$, so we can use the results from [23], [25] to obtain the desired contradiction.

As in the linear phase case handled in [26], we begin by establishing an "approximate functional equation"[7] for the polynomial function $P_x(n)$ in (16). Note that if $p \le H^{\varepsilon}$ is a prime, then if $\lambda$ correlates with $e(P_x(n))$ on $[x, x+H]$, then $\lambda$ correlates with $e(P_x(pn))$ on $[x/p, (x+H)/p]$, for "most" choices of $p$ (this is a standard Turán–Kubilius argument; see Proposition 3.4). Similarly, for "most" $y \in [X, 2X]$ and primes $q \le H^{\varepsilon}$, we must have that $\lambda$ correlates with $e(P_y(qn))$ on $[y/q, (y+H)/q]$. Now, if $|x/p - y/q| \le H/(2\max\{p, q\})$,

[5]We note that the notion of "complexity" of nilmanifolds plays no role in this paper, unlike in e.g. [15], since the inverse theorem supplies us with a single nilmanifold $G_\eta/\Gamma_\eta$ such that $\|f\|_{U^k[N]} \ge \eta$ with $f$ 1-bounded implies that $f$ correlates with a nilsequence on $G_\eta/\Gamma_\eta$.

[6]The actual equivalence relation used in Section 3 is slightly more elaborate; it also allows for a factor $\gamma(n)$ which is a rational polynomial. To show ideas, let us work with this slightly simpler equivalence in which we allow the "Archimedean" error $\varepsilon$ but not the "non-Archimedean" error $\gamma$.

[7]Our use of the term approximate functional equation of course differs from its meaning in the context of $L$-functions.

then the intervals $[x/p,(x+H)/p],[y/q,(y+H)/q]$ have large intersection, and since by the large sieve for polynomial phases (Proposition 3.3) there can only be boundedly many polynomial phases that $\lambda$ correlates with on an interval, we can say that

$$e(P_x(pn)) \approx e(P_y(qn)), \quad n \in [x/p,(x+H)/p]$$

for "most" $p,q \in [P,2P]$ and $x,y \in [X,2X]$ with $x/p = y/q + O(H/P)$ for some $P \leq H^{\varepsilon}$. This corresponds to an approximate equality of polynomials modulo 1, but using a suitable version of the Chinese remainder theorem (Proposition 3.5), and shifting $P_x, P_y$ by integer amounts, which is always allowable, we can eventually upgrade this to an equality modulo the product $\prod_{p' \in \mathcal{P}} p'$, where $\mathcal{P}$ is a "large" set of primes in $[P,2P]$, and thus with our choice of $H$ the modulus is enormous compared to $X$, so we can essentially treat this as a genuine equality in $\mathbb{R}$. In this way, we can essentially pass to the approximate functional equation

$$P_x(pn) \sim P_y(qn) \tag{18}$$

for "most" $p,q \in [P,2P]$ and $x,y \in [X,2X]$ with $x/p = y/q + O(H/P)$.

If we now form a graph $\mathcal{G}$ on $[X,2X] \cap \mathbb{Z}$ by connecting $x,y$ whenever $x/p = y/q + O(H/P)$ and $x,y,p,q$ are satisfying the above conditions, we obtain a graph whose structure governs the solutions to (18). In particular, (a known case of) Sidorenko's conjecture tells us that $\mathcal{G}$ contains many configurations $\mathcal{C}$ consisting of two $\ell$-cycles and an edge between them, for $\ell > \log X / \log P$. When we unwrap what this means in terms of approximate functional equations, we obtain (in Proposition 3.7) the approximate dilation invariance

$$P_x(a_x n) \sim P_x(b_x n) \tag{19}$$

for many pairs $(a_x, b_x)$ that are of polynomial size in $X$ (more precisely, products of $\ell$ primes from $[P,2P]$), and relatively close to each other (with $\frac{a_x - b_x}{a_x} \asymp \frac{H}{X}$).

We then "solve" the approximate equation (19) using properties of the underlying polynomial algebra, with the conclusion that $P_x$ must locally "pretend" to be a character:

$$e(P_x(n)) \approx e\left(\frac{T_x}{2\pi} \log n + \gamma_x(n)\right),$$

where $\gamma_x$ is $O(X^{O(1)})$-rational (in a sense specified in Section 4) and $T_x = O(X^{k+1}/H^{k+1})$; see Proposition 3.8 for a precise statement. Moreover, the quantities $T_x$ can now be shown to satisfy the approximate functional equation

$$T_x = T_y + O(X/H)$$

when $x/p = y/q + O(\frac{H}{PX})$, for "most" $x,y,p,q$. As in [26], using mixing properties of the graph $\mathcal{G}$ arising from cancellation in $\sum_{P \leq p \leq 2P} p^{it}$ for $|t| \ll X^{O(1)}$, we may deduce from this that $T_x = T_0 + O(X/H)$ for some $T_0$ of polynomial size and for a "most" values of $x$. Further, we also have (modulo integer-valued polynomials) the relation

$$\gamma_x(pn) = \gamma_y(qn)$$

for the same tuples $(x, y, p, q)$, and solving this eventually leads to $\gamma_x(n)$ being $O(1)$-rational (with a bit more work than in [25], where $\gamma_x(n)$ was just of the form $\frac{a}{q'}n$). Putting everything together, we reach the relation (17), which was enough for finishing the proof.

For $G$ non-abelian, we can use some of the above arguments, but certain additional difficulties (indicated below) arise that necessitate a more involved analysis involving quantitative nilalgebra and some refinements on the graph theory side. Note that by the factorization theorem for nilsequences [16], we have a similar splitting of polynomials $g : \mathbb{Z} \to G$ to a smooth part, an equidistributed part and a rational part, so we may define a similar equivalence relation for these sequences as for polynomial phases. Moreover, we can make sense of the sequence $g(n)$ evaluated at real $n$ and we can define the size of an element of $G$; see Section 4 for details.

Up until the approximate functional equation (18) (now with $g_x(n)$ in place of $P_x(n)$), the arguments in the polynomial phase case are sufficiently general to work equally well for nilsequences. We can also obtain the analogue of (19) similarly but, perhaps surprisingly, in the nilsequence setting the solutions to (19) for a given pair $(a_x, b_x)$ are *not* all approximate characters (see (95) for a counterexample). We thus must proceed more carefully and extract more information from the fact that (19) holds for an extremely large family of pairs $(a_x, b_x)$. It turns out that the pathological solutions to (19) for a given $(a_x, b_x)$ generally do not obey (19) for other pairs $(a'_x, b'_x)$, but demonstrating that requires some work.

The way we obtain the required extra information is by generalizing the graph theory argument from [26] a bit (to configurations of two cycles of unequal length connected by an edge), and this extra flexibility allows us to obtain

$$g_x((1+\theta)t) \sim g_x(t)\gamma_{x,\theta}(t), \quad t \in [x, x+H] \tag{20}$$

for $t \in \mathbb{R}$ and for a "very dense" set of real numbers $\theta = O(H/X)$ (as opposed to just a few such numbers), where $\gamma_{x,\theta}$ is $Q$-rational with $Q \gg \prod_{p\in[P,2P]} p^{\varepsilon}$ (this notion makes sense in Lie algebras; see Section 4). This is the outcome of Proposition 4.19.

**Remark 1.13.** *As indicated above, while in the case of polynomial phases it suffices to have equation* (20) *hold for a single $\theta$, in the more general nilsequence case this condition is insufficient due to the existence of exotic "approximately multiplicative" nilsequences. Consider for example $\phi(n) = F(g(n)\Gamma)$ where*

$$g(n) = e_1^{T_1 \log n} e_2^{T_2 \log n} e_{12}^{-\frac{T_1 T_2}{2}(\log n)^2}$$

*where here $e_1, e_2, e_{12}$ are the generators of the free 2-step 3 dimensional nilpotent Lie group, $\Gamma$ the standard lattice. By Taylor approximation of the logarithm function, $g(n)$ differs from a polynomial sequence by a negligible amount. Moreover, $g((1+\theta)t) = g(1+\theta)g(t)$ so that so one would get $\phi((1+\theta)n) \sim \phi(n)$ if $g(1+\theta)$ is very close to $\Gamma$, independent of $n$.*

It is a fact (following from the Baker–Campbell–Hausdorff formula) that if $n \mapsto \gamma_{x,\theta}(n)$ is simultaneously very rational and of polynomial size, then it is a constant; thus, $\gamma_{x,\theta}(n) =: \gamma_{x,\theta}$. Make in (20) the change of variables $1 + \theta_x = e^{\alpha/N}$ with $\alpha \sim 1$ restricted to a very

dense set of numbers and with $N = X/H$. Then

$$g_x(e^{\alpha/N}t) \sim g_x(t)\gamma_{x,\alpha}, \quad t = x + O(H),$$

so by iterating

$$g_x(e^{n\alpha/N}t) \sim g_x(t)\gamma_{x,\alpha}^n$$

for all integers $n = O(1)$. In fact, by an interpolation lemma (Lemma 2.3), we will be able to boost this to real $n$ as well. Now we essentially have a two-variable functional equation for $g_x$, which after some manipulation gives us

$$g_x(y) \sim T^{N\log(y/x)}, \quad y = x + O(H), \tag{21}$$

and for some $T = T_x \in G$ of polynomial size. Here, $T$ is given by the relation

$$T^{\alpha s} \sim \gamma_{x,\alpha}^s$$

for $s = O(1)$ and for a dense set of $\alpha \sim 1$ (cf. Proposition 4.20). This is still not enough for us, since when $G$ is non-abelian, $y \mapsto F(T^{N\log(y/x)}\Gamma)$ need not resemble a character at all. With some extra work, which involves quantitative equidistribution theory of nilsequences and the mixing lemma to carefully analyze the compatibility between (20) and (21), we eventually show that $T = O(1)T_0$, where $T_0$ is of polynomial size and lies either in the center of $G$ or in a proper rational subgroup of $G$. In the case that $G$ is non-abelian, the former case is contained in the latter. This is then finally enough, since the $O(1)$ error turns out to be negligible by Taylor expansion, and if $T$ lies in a proper rational subgroup, we ascend to a group of lower dimension, so we can apply induction to conclude. Thus $n \mapsto T^{N\log(n/x)}$ must essentially be a polynomial function on an *abelian* nilmanifold, meaning that it is a classical polynomial. This reduces us back to the polynomial phase case, whose proof we outlined above.

1.3.2. *The sign patterns result.* We then sketch the proof of Theorem 1.9. Suppose for the sake of contradiction that $s(k) \ll k^A$ for some $A$ and for $k$ belonging to an infinite set $\mathcal{K}$. Then, expanding the (logarithmic) density of each sign pattern of length $k$ as a correlation, we must have

$$C := \frac{1}{\log x}\sum_{n\le x}\frac{\lambda(n+h_1)\cdots\lambda(n+h_j)}{n} \gg_k 1$$

for $k \in \mathcal{K}$ and for some distinct $h_1, \dots, h_j \in [1,k]$. The entropy decrement argument developed in [34] (see also [37]), allows one to write $C$ as a double average:

$$C = (-1)^k\frac{\log P}{P}\sum_{P\le p\le 2P}\frac{1}{\log x}\sum_{n\le x}\frac{\lambda(n+ph_1)\cdots\lambda(n+ph_j)}{n} + o(1), \tag{22}$$

where $P = P(x)$ is suitable. However, $P$ has to be very small here (namely $P \ll (\log x)^{o(1)}$), which is by far too small in order to apply Corollary 1.6. Instead, we leverage the assumption that $\lambda$ is assumed to have few sign patterns to show that the entropy decrement argument can be replaced with a quantitatively much stronger method of moments computation, and this eventually allows us to obtain (22) for $P \gg X^\varepsilon$ (along a suitable sequence of values

of $X$ depending on $\mathcal{K}$). Then we are in a position to apply Corollary 1.6, and we conclude from the generalized von Neumann theorem that actually $C = o(1)$, which is the desired contradiction.

1.4. **Acknowledgments.** This work was initiated at the American Institute of Mathematics workshop on Sarnak's conjecture in December 2018. KM was supported by Academy of Finland grant no. 285894. MR acknowledges the support of NSF grant DMS-1902063 and a Sloan Fellowship. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1764034. JT was supported by a Titchmarsh Fellowship. TZ was supported by ERC grant ErgComNum 682150.

We are grateful to the anonymous referees for their extremely careful reading of the paper and for numerous helpful comments and remarks that improved the presentation of this paper. We thank Amita Malik, Redmond McNamara and Peter Sarnak for helpful discussions.

## 2. Notation and preliminaries

We use the asymptotic notation $X \ll Y$, $X = O(Y)$ or $Y \gg X$ to denote the estimate $|X| \leq CY$ for some absolute constant $C$ (in case of $Y \gg X$ we also require that $X \geq 0$). If we allow the constant $C$ to depend on parameters, we will indicate this by subscripts unless otherwise specified, thus for instance $X = O_k(Y)$ denotes the estimate $|X| \leq C_k Y$ for some $C_k$ depending on $k$. We also write $X \asymp Y$ for $X \ll Y \ll X$.

Several of the concepts defined in this paper (e.g., "large family", "smooth polynomial", "comparable interval", etc.) will rely on the above notation, and thus involve some unspecified implicit constants. If a proposition involves such notation in both its hypotheses and conclusion, then the implied constants in the conclusions are always permitted to depend on the implied constants in the hypotheses.

All intervals in this paper will be closed. If $I$ is an interval, we use $|I|$ to denote its Lebesgue measure and $x_I$ to denote its midpoint, thus $I = [x_I - \frac{|I|}{2}, x_I + \frac{|I|}{2}]$. For any $x \in \mathbb{R}$, we define the normalized distance

$$\langle x \rangle_I := \frac{\operatorname{diam}(I \cup \{x\})}{|I|} \tag{23}$$

and similarly for an interval $J$

$$\langle J \rangle_I := \frac{\operatorname{diam}(I \cup J)}{|I|}. \tag{24}$$

We say that two intervals $I, J$ are *comparable*[8], and write $I \sim J$, if we have $\langle I \rangle_J, \langle J \rangle_I \ll 1$, or equivalently if $|I| \asymp |J| \asymp \operatorname{diam}(I \cup J)$. Note that this is an equivalence relation up to modification of the implied constants; for instance if $I \sim J$ and $J \sim K$ then $I \sim K$, where the implied constants in the latter relation can differ from those in the former.

[8]Here and throughout the paper, definitions such as this one that depend on an implicit asymptotic parameter are only called in the presence of such parameters (which will be the parameters in Theorem 1.5).

If $F$ is a finite set, we use $\#F$ to denote its cardinality. If $E$ is a set, we use $1_E$ to denote its indicator function, thus $1_E(n) = 1$ when $n \in E$ and $1_E(n) = 0$ otherwise. Similarly, for any statement $S$, we define the indicator $1_S$ to equal 1 when $S$ is true and 0 otherwise.

For any subset $E$ of the real line, we use $a + E := \{a + x : x \in E\}$ to denote the translation of $E$ by a shift $a \in \mathbb{R}$, and $\lambda E := \{\lambda x : x \in E\}$ to denote the dilation of $E$ by a factor $\lambda > 0$. For instance if $I, J$ are intervals, then $I \sim J$ if and only if $\lambda I \sim \lambda J$. If $f\colon \mathbb{R} \to S$ is any function taking values in some set $S$, we use $f(\lambda\cdot)\colon \mathbb{R} \to S$ to denote the dilated function $t \mapsto f(\lambda t)$. For an interval $I$ and function $g$, we also use the pushforward notation $\lambda_*(I, g) := \big(\lambda I, g\left(\frac{1}{\lambda}\cdot\right)\big)$.

If $a, b$ are elements of an additive group $(G, +)$, and $H$ is a subgroup of $G$, we write $a = b \bmod H$ to denote the claim that $a - b \in H$; by abuse of notation we also use $a \bmod H$ to denote the element $a + H$ of the quotient group $G/H$. Similarly, if $G = (G, \cdot)$ is a multiplicative group and $H$ is a normal subgroup, we write $a = b \bmod H$ to denote the claim that $ab^{-1} \in H$.

Summations and products over the symbol $p$ (or $p'$, etc.) are always understood to be over primes unless otherwise specified, and similarly sums over $n$ are understood to be over integers unless otherwise specified.

In Section 5, we will need some averaging notation. For a function $f : A \to \mathbb{C}$ defined on a set $A$ with $A \subset \mathbb{N}$ nonempty, define its unweighted and logarithmic average over $A$ by

$$\mathbb{E}_{n\in A} f(n) := \frac{1}{|A|}\sum_{n\in A} f(n) \quad \text{and} \quad \mathbb{E}^{\log}_{n\in A} f(n) := \frac{1}{\sum_{n\in A}\frac{1}{n}}\sum_{n\in A}\frac{f(n)}{n},$$

respectively. Thus in particular for a bounded function $f : \mathbb{N} \to \mathbb{C}$ we have

$$\mathbb{E}^{\log}_{n\le x} f(n) = \frac{1}{\log x}\sum_{n\le x}\frac{f(n)}{n} + o(1), \quad \text{and} \quad \mathbb{E}_{x\le p\le 2x} f(p) = \frac{1}{x/\log x}\sum_{x\le p\le 2x} f(p) + o(1).$$

If $\mathcal{P}$ is a collection of prime numbers, we use $\prod \mathcal{P}$ to denote the product of its elements:

$$\prod \mathcal{P} := \prod_{p\in\mathcal{P}} p.$$

For any $P \ge 2$, we let $\pi_0(P)$ denote the quantity

$$\pi_0(P) := \frac{P}{\log P}.$$

Note that from the prime number theorem, we see that for sufficiently large $P$, the number of primes in $[P, 2P]$ or $[P/2, P]$ is comparable to $\pi_0(P)$. Accordingly, we say that a set of primes in $[P, 2P]$ or $[P/2, P]$ is *large* if its cardinality is $\gg \pi_0(P)$. Observe that if $\mathcal{P}$ is a large set of primes in $[P, 2P]$ or $[P/2, P]$, then we have an exponential lower bound

$$\prod \mathcal{P} \gg \exp(cP) \tag{25}$$

for some $c \gg 1$. In practice, this lower bound means that $\prod \mathcal{P}$ is so large compared with the many "polynomial size" quantities we will encounter in the course of our arguments that this modulus is effectively infinite.

For a smooth function $f\colon \mathbb{R} \to \mathbb{C}$, we use $f^{(j)}$ to denote the $j^{\text{th}}$ derivative for $j \geq 0$. We recall the *Bernstein inequality* (see e.g., [31, p. 146])

$$\sup_{t\in I} |f^{(1)}(t)| \ll_k |I|^{-1} \sup_{t\in I} |f(t)| \tag{26}$$

for all polynomials $f \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$, and hence on iteration

$$\sup_{t\in I} |f^{(j)}(t)| \ll_k |I|^{-j} \sup_{t\in I} |f(t)| \tag{27}$$

for any $j \geq 0$ (note that $f^{(j)}$ vanishes for $j > k$). From Taylor expansion we then also have

$$|f^{(j)}(t')| \ll_k |I|^{-j} \langle t' \rangle_I^{k-j} \sup_{t\in I} |f(t)| \tag{28}$$

for any $t' \in \mathbb{R}$ and $j \geq 0$, using the notation (23).

If $\delta > 0$, we use $\mathrm{Poly}_{\leq k}(\delta\mathbb{Z} \to \mathbb{Z})$ to denote the subgroup of the additive group $\mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ consisting of polynomials $\gamma$ such that $\gamma(\delta\mathbb{Z}) \subset \mathbb{Z}$; we refer to these polynomials as $\frac{1}{\delta}$-*integral* polynomials. We have the following explicit description of these groups:

**Lemma 2.1** (Discrete Taylor expansion)**.** *For any $\delta > 0$ and $k \geq 0$, the space $\mathrm{Poly}_{\leq k}(\delta\mathbb{Z} \to \mathbb{Z})$ consists precisely of those functions $\gamma : \mathbb{R} \to \mathbb{R}$ of the form*

$$\gamma(t) := \sum_{j=0}^{k} c_j \binom{t/\delta}{j}$$

*for some integers $c_0, \dots, c_k$, where $\binom{x}{j} := \frac{x(x-1)\dots(x-j+1)}{j!}$.*

In some parts of the paper we will also use a non-abelian version of Lemma 2.1 (see Lemma B.2).

*Proof.* By rescaling we may take $\delta = 1$. The claim is trivial for $k = 0$, so suppose inductively that $k \geq 1$ and that the claim has already been proven for $k-1$. The polynomials $\binom{\cdot}{j}$ for $j = 0, \dots, k$ all lie in $\mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$, and hence so do all integer linear combinations $\sum_{j=0}^{k} c_j \binom{\cdot}{j}$. Conversely, suppose that $\gamma \in \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$. On taking $k^{\text{th}}$ divided differences, we see that the $k^{\text{th}}$ derivative $\gamma^{(k)}$ (which is a constant) is equal to an integer $c_k$. Thus the polynomial $\gamma - c_k\binom{\cdot}{k}$ has vanishing $k^{\text{th}}$ derivative and thus lies in $\mathrm{Poly}_{\leq k-1}(\mathbb{Z} \to \mathbb{Z})$. The claim now follows from the induction hypothesis. □

We will need the following application of Bezout's identity:

**Lemma 2.2** (Bezout identity)**.** *Let $a, b$ be coprime natural numbers, and let $k \geq 0$. Then for any $\lambda > 0$ we have*

$$\mathrm{Poly}_{\leq k}\left(\frac{\lambda}{a}\mathbb{Z} \to \mathbb{Z}\right) + \mathrm{Poly}_{\leq k}\left(\frac{\lambda}{b}\mathbb{Z} \to \mathbb{Z}\right) = \mathrm{Poly}_{\leq k}(\lambda\mathbb{Z} \to \mathbb{Z})$$

*and*

$$\mathrm{Poly}_{\leq k}\left(\frac{\lambda}{a}\mathbb{Z} \to \mathbb{Z}\right) \cap \mathrm{Poly}_{\leq k}\left(\frac{\lambda}{b}\mathbb{Z} \to \mathbb{Z}\right) = \mathrm{Poly}_{\leq k}\left(\frac{\lambda}{ab}\mathbb{Z} \to \mathbb{Z}\right).$$

Thus for instance every 1-integral polynomial can be decomposed as the sum of an $a$-integral and a $b$-integral polynomial, and a polynomial is $ab$-integral if and only if it is both $a$-integral and $b$-integral.

*Proof.* See Appendix C. □

We will need a variant of the Bernstein inequality for exponential polynomials, that is to say real linear combinations of exponential monomials $t \mapsto t^j \exp(\alpha t)$ for some non-negative integers $j$ and real numbers $\alpha$:

**Lemma 2.3** (Bernstein inequality for exponential polynomials)**.** *Let $d_1, \dots, d_k$ be non-negative integers, and let $N_0$ be a sufficiently large natural number depending on $k, d_1, \dots, d_k$. Let $\alpha_1, \dots, \alpha_k$ be real numbers whose absolute values are sufficiently small depending on $k, d_1, \dots, d_k, N_0$. Let $P : \mathbb{R} \to \mathbb{R}$ be a real linear combination of the exponential monomials $t \mapsto t^j \exp(\alpha_i t)$ for $i = 1, \dots, k$ and $0 \leq j \leq d_i$. Then for any interval $I$ and any non-negative integer $m$ one has, for all $t \in I$,*

$$|P^{(m)}(t)| \ll_{k,d_1,\dots,d_k,m,N_0,I} \sup_{n=1,\dots,N_0} |P(n)|. \tag{29}$$

*Proof.* See Appendix A. □

## 3. Local correlations with polynomial phases

In this section, we establish Theorem 1.3, which implies Corollary 1.1 as a special case. Our arguments shall follow those in [26] (although they will be reformulated in a more general and algebraic setting that applies to relevant collections of phase functions, such as polynomial phases and later to nilsequences in Section 4). Some familiarity with the arguments in [26] will be presumed in this section.

Let $k, \theta, f, X, \eta, H$ be as in Theorem 1.3. To simplify the notation we now allow all implied constants in the asymptotic notation to depend on $k, \theta, \eta$, thus for instance

$$\int_X^{2X} \|f\|_{u^{k+1}([x,x+H])}\, dx \gg X. \tag{30}$$

We can assume that $X$ is sufficiently large depending on $k, \theta, \eta$, since the claim is trivial otherwise. We can also assume[9] $k \geq 1$, since the $k = 0$ case follows similarly to [25, Theorem A.1][10].

It will be convenient to abstract the properties of the polynomial phases one is testing against, as this will allow us to easily generalize many of the arguments in this section to the case of nilsequence correlations in Section 4. Define a *local polynomial phase* to be a pair $\phi = (I, P)$, where $I$ is an interval in $\mathbb{R}$ and $P \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is a polynomial. We let $\Phi$ denote the set of all local polynomial phases $(I, P)$, and $\Phi_I$ the set of local polynomial phases $(I, P)$ with a given $I$. Intuitively, $(I, P)$ should be viewed as an abstraction of the phase function $t \mapsto e(P(t))$ on the interval $I$. If $\phi = (I, P)$ is a local polynomial phase and $f : \mathbb{Z} \to \mathbb{C}$ is a function, we define the correlation

$$\langle f, \phi \rangle := \frac{1}{|I|} \sum_{n \in I} f(n) e(-P(n)). \tag{31}$$

Thus we have

$$\|f\|_{u^{k+1}([x,x+H])} = \sup_{\phi \in \Phi_{[x,x+H]}} |\langle f, \phi \rangle|$$

and thus from (30)

$$\int_X^{2X} \sup_{\phi \in \Phi_{[x,x+H]}} |\langle f, \phi \rangle| \, dx \gg X. \tag{32}$$

Recall from Section 2 that given any local polynomial phase $\phi = (I, P) \in \Phi$ and a scaling factor $\lambda > 0$, we define the rescaling (or pushforward) $\lambda_* \phi \in \Phi$ by the formula

$$\lambda_* \phi := \left( \lambda I, P\left( \frac{1}{\lambda} \cdot \right) \right).$$

Note that this gives a multiplicative action on $\Phi$, in the sense that

$$(\lambda_1)_*((\lambda_2)_* \phi) = (\lambda_1 \lambda_2)_* \phi$$

whenever $\phi \in \Phi$ and $\lambda_1, \lambda_2 > 0$.

Following [26, §2], we perform a convenient discretization. Define an $(X, H)$-*family of intervals* to be a finite collection $\mathcal{I}$ of intervals of length $H$ contained in $[X/10, 10X]$ such that any pair of intervals in $\mathcal{I}$ are separated by a distance at least $500H$. We say that such a family $\mathcal{I}$ is *large* if $\#\mathcal{I} \gg X/H$. By repeating the proof of [26, Lemma 2.1] (which is a pigeonholing argument) using (32) as a starting point, one obtains a large $(X, H)$-family of intervals $\mathcal{I}$, such that for each $I \in \mathcal{I}$ one can find $\phi_I \in \Phi_I$ such that

$$|\langle f, \phi_I \rangle| \gg 1. \tag{33}$$

[9]Indeed, from the results in [26] we can almost assume $k \geq 2$, except for the problem that those results contain an additional loss of $H^\rho$ in the conclusion that is not conceded here. In any case, the arguments here will also recover the $k = 1$ case without difficulty.

[10]The only difference is that one needs to, in the formula below [25, Theorem A.2], treat the integral over $|t| \geq CX/(2H)$ by the mean value theorem to be able to work with $M(f; CX/H, Q)$ instead of $M(f; X, Q)$.

We remark that this step does not require any properties of the polynomial space $\mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$, as it only uses the fact that $e(P(n))$ is 1-bounded for every $P$ in this space.

The next step is to use the multiplicativity of $f$ to relate the various $\phi_I$ to each other. We need a key definition, given as Definition 3.1 below. Given an interval $I$ in $\mathbb{R}$, we say that a map $\varepsilon \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is *smooth* on $I$ if one has the bound

$$|\varepsilon(t)| \ll 1$$

for all $t \in I$, which by (28) also implies that

$$\left|\frac{d^j}{dt^j}\varepsilon(t)\right| \ll |I|^{-j}\langle t\rangle_I^{k-j}$$

for all $j \geq 0$ and $t \in \mathbb{R}$. In particular, if $\varepsilon$ is smooth on $I$, then it is also smooth on $I'$ for any $I' \sim I$.

**Definition 3.1** (Comparability of polynomial phases)**.** *Given two local polynomial phases $\phi_1 = (I_1, P_1), \phi_2 = (I_2, P_2)$ of $\Phi$ and a scaling factor $\delta > 0$, we define the relation*

$$\phi_1 \sim_\delta \phi_2$$

*to hold if $I_1 \sim I_2$, and we have a splitting*

$$P_1 = \varepsilon + P_2 + \gamma,$$

*where $\varepsilon, \gamma \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ are polynomials obeying the following axioms:*

(i) *($\varepsilon$ smooth) $\varepsilon$ is smooth on $I_1$.*
(ii) *($\gamma$ is $\frac{1}{\delta}$-integral) $\gamma \in \mathrm{Poly}_{\leq k}(\delta\mathbb{Z} \to \mathbb{Z})$.*

Informally, the relation $\phi_1 \sim_\delta \phi_2$ asserts that $\phi_1$ "pretends to be" $\phi_2$ on the discrete set $I_1 \cap \delta\mathbb{Z}$. Technically, this is not a single relation, but a family of relations, depending on the choices of implied constants appearing in (i), but we shall abuse notation by referring to $\sim_\delta$ as a single relation. It obeys the following basic properties:

**Proposition 3.2** (Basic properties of $\sim_\delta$)**.** *Let $\delta > 0$, and let $\phi, \phi', \phi'' \in \Phi$.*

(i) *(Equivalence relation) We have $\phi \sim_\delta \phi$, and if $\phi \sim_\delta \phi'$ then $\phi' \sim_\delta \phi$. Finally, if $\phi \sim_\delta \phi'$ and $\phi' \sim_\delta \phi''$ then $\phi \sim_\delta \phi''$, where we allow the implied constants in the latter relations to depend on the implied constants in the former relations.*
(ii) *(Dilation invariance) If $\phi \sim_\delta \phi'$ and $\lambda > 0$, then $\lambda_*\phi \sim_{\lambda\delta} \lambda_*\phi'$.*
(iii) *(Sparsification) If $\phi \sim_\delta \phi'$, then $\phi \sim_{\ell\delta} \phi'$ for any natural number $\ell$.*

*Proof.* These are immediate from Definition 3.1, together with the previously made observation that a polynomial smooth on an interval $I$ is automatically smooth on all comparable intervals $I' \sim I$. □

The relevance of this relation to the correlations (33) comes from the following lemma.

**Proposition 3.3** (Large sieve)**.** *Let $I$ be an interval of some length $|I| \geq 1$, and let $f : \mathbb{Z} \to \mathbb{C}$ be a function bounded in magnitude by 1. Suppose that for each $i = 1, \dots, K$ there is an interval $I_i \sim I$ and a local polynomial phase $\phi_i \in \Phi_{I_i}$ such that*

$$|\langle f, \phi_i \rangle| \gg 1.$$

*Then either*

$$K \ll 1$$

*or there exists $1 \leq i < j \leq K$ such that*

$$\phi_i \sim_1 \phi_j.$$

*Proof.* Write $\phi_i = (I_i, P_i)$ and $H = |I|$. By (31), for each $1 \leq i \leq K$, we can find a real number $\theta_i$ such that

$$\mathrm{Re}\left( e(\theta_i) \sum_{n \in I_i} f(n) e(-P_i(n)) \right) \gg H$$

and hence on summing in $i$ and rearranging

$$\mathrm{Re}\left( \sum_{n \in I} f(n) \sum_{i=1}^{K} 1_{I_i}(n) e(\theta_i) e(-P_i(n)) \right) \gg HK.$$

By Cauchy-Schwarz we conclude that

$$\sum_{n \in I} \left| \sum_{i=1}^{K} 1_{I_i}(n) e(\theta_i) e(-P_i(n)) \right|^2 \gg HK^2.$$

The left-hand side can be rearranged as

$$\sum_{i=1}^{K} \sum_{j=1}^{K} e(\theta_j - \theta_i) \sum_{n \in I_i \cap I_j} e(P_i(n) - P_j(n)).$$

Thus, by the pigeonhole principle and triangle inequality, there exists $1 \leq i \leq K$ such that

$$\sum_{j=1}^{K} \left| \sum_{n \in I_i \cap I_j} e(P_i(n) - P_j(n)) \right| \gg HK,$$

and hence

$$\left| \sum_{n \in I_i \cap I_j} e(P_i(n) - P_j(n)) \right| \gg H \tag{34}$$

for $\gg K$ choices of $j = 1, \dots, K$. Fix this choice of $i$.

Let $n_I$ denote an integer point in $I$. For each $j$ such that (34) holds, we write

$$P_i(t) - P_j(t) = \sum_{l=0}^{k} \alpha_{j,l} (t - n_I)^l$$

for some real coefficients $\alpha_{j,l}$. Then we have

$$\left| \sum_{n \in (I_i - n_I) \cap (I_j - n_I)} e\left( \sum_{l=0}^{k} \alpha_{j,l} n^l \right) \right| \gg H$$

Applying Weyl sum estimates such as [33, Lemma 1.1.16], we conclude that there exists a natural number $1 \leq q_j \ll 1$ such that

$$\|q_j \alpha_{j,l}\|_{\mathbb{R}/\mathbb{Z}} \ll H^{-l}$$

for $l = 0, \dots, k$, where $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance of $x$ to the nearest integer. In particular there exist natural numbers $1 \leq a_{j,l} \leq q_j$ such that

$$\left\| \alpha_{j,l} - \frac{a_{j,l}}{q_j} \right\|_{\mathbb{R}/\mathbb{Z}} \ll H^{-l}.$$

The total number of tuples $(q_j, a_{j,1}, \dots, a_{j,k})$ is $O(1)$. Thus by the pigeonhole principle, either $K \ll 1$, or else there exist $1 \leq j < j' \leq K$ such that $q_j = q_{j'}$ and $a_{j,l} = a_{j',l}$ for all $l = 0, \dots, K$. In particular, by the triangle inequality we have

$$\|\alpha_{j,l} - \alpha_{j',l}\|_{\mathbb{R}/\mathbb{Z}} \ll H^{-l}$$

for $l = 0, \dots, K$, so we can write $\alpha_{j',l} = \varepsilon_{j,j',l} + \alpha_{j,l} + \gamma_{j,j',l}$ for some integer $\gamma_{j,j',l}$ and some real number $\varepsilon_{j,j',l} = O(H^{-l})$. This gives the decomposition

$$P_j(t) = \sum_{l=0}^{k} \varepsilon_{j,j',l} (t - n_I)^l + P_{j'}(t) + \sum_{l=0}^{k} \gamma_{j,j',l} (t - n_I)^l.$$

Comparing this with Definition 3.1, we see that

$$\phi_j \sim_1 \phi_{j'},$$

and the proposition follows. □

Using this proposition, we can obtain

**Proposition 3.4** (Scaling down)**.** *Let $2 \leq P \leq Q \leq H \leq X$ and let $f : \mathbb{N} \to \mathbb{C}$ be a 1-bounded multiplicative function. Suppose there exists a large $(X, H)$-family $\mathcal{I}$ and a local polynomial phase $\phi_I \in \Phi_I$ associated to each interval $I \in \mathcal{I}$ such that*

$$|\langle f, \phi_I \rangle| \gg 1$$

*for all $I \in \mathcal{I}$. Assuming that $P, \frac{\log Q}{\log P}$ are sufficiently large (depending on the implied constants in the above hypotheses), there exist $P' \in [P, Q/2]$, a large $(\frac{X}{P'}, \frac{H}{P'})$-family $\mathcal{I}'$, and a function $\phi'_{I'} \in \Phi_{I'}$ associated to each $I' \in \mathcal{I}'$, such that*

$$|\langle f, \phi'_{I'} \rangle| \gg 1$$

*for all $I' \in \mathcal{I}'$. Furthermore, for each $I' \in \mathcal{I}'$, one can find $\gg \pi_0(P')$ pairs $(I, p')$, where $I \in \mathcal{I}$ and $p'$ is a prime in $[P', 2P']$, such that the rescaled interval $\frac{1}{p'}I$ lies within $3\frac{H}{P'}$ of $I'$, and such that*

$$(\frac{1}{p'})_* \phi_I \sim_1 \phi'_{I'}. \tag{35}$$

*Proof.* From Proposition 3.3 and the greedy algorithm, we can associate to each interval $I$ of length $H \geq 1$ and any $\eta' > 0$ a family $\phi_1, \dots, \phi_K \in \Phi_I$ of local polynomial phases with $K = O_{\eta'}(1)$ such that whenever one has

$$|\langle f, \phi \rangle| \geq \eta'$$

for some $\phi \in \Phi_J$ with $J \subset I$ and $|J| \geq \eta'|I|$, then one has

$$\phi \sim_1 \phi_i$$

for some $i = 1, \dots, K$ (if we permit implied constants in the $\sim_1$ notation to depend on $\eta'$). The claim now follows by repeating the proof of [26, Proposition 3.1] (which is a Turán–Kubilius argument), using the above claim as a substitute for [26, Lemma 2.2]. For the convenience of the reader we sketch the main ideas of this argument as follows. First, by using [26, Proposition 2.5] and the multiplicative nature of $f$, one can deduce that

$$|\langle f, (\frac{1}{p'})_* \phi_I \rangle| \gg 1$$

for many $I \in \mathcal{I}$ and many primes $p' \in [P, Q]$, and thence (by the pigeonhole principle) for many $I \in \mathcal{I}$ and $p' \in [P', 2P']$ for a suitable $P'$. By further pigeonholing, we may arrange matters so that the intervals $\frac{1}{p'}I$ lie close to intervals $I'$ in a suitable large $(\frac{X}{P'}, \frac{H}{P'})$-family $\mathcal{I}'$. Using the previously mentioned claim, one can then show that many of the $(\frac{1}{p'})_* \phi_I$ associated to a given interval $I'$ are related via the $\sim_1$ relation to a suitable phase $\phi'_{I'}$, which will give the claim. □

We also need the following version of the Chinese remainder theorem[11]. This proposition turns out to be very useful in what follows, since it allows us to upgrade equivalences

[11]The reason we call this a Chinese remainder theorem is that it allows us to combine $\bmod p$ conditions for different primes $p$.

between different exponential phases up to the point where the modulus is so large that we must have a genuine equality in $\mathbb{R}$.

**Proposition 3.5** (Chinese remainder theorem)**.** *Let $I$ be an interval of some length $|I| \geq 1$, and let $\mathcal{P}$ be a finite collection of primes.*

(i) *Suppose that $\phi \in \Phi_I$, and that for each $p \in \mathcal{P}$ there exists $\phi_p \in \Phi$ such that*

$$\phi_p \sim_1 \phi.$$

*Then there exists $\tilde{\phi} \in \Phi_I$ such that*

$$\phi_p \sim_{\frac{1}{p}} \tilde{\phi}$$

*for all $p \in \mathcal{P}$, and furthermore $\langle f, \phi\rangle = \langle f, \tilde{\phi}\rangle$ for all $f : \mathbb{Z} \to \mathbb{C}$.*

(ii) *Suppose that $\phi \in \Phi_I$ and $\phi' \in \Phi$ are such that*

$$\phi \sim_{\frac{1}{p}} \phi'$$

*for all $p \in \mathcal{P}$, and suppose $|I|$ is sufficiently large (depending on the implied constants in the $\sim_{\frac{1}{p}}$ notation). Then*

$$\phi \sim_{\frac{1}{\prod \mathcal{P}}} \phi'.$$

*Proof.* See Appendix C. □

One can now conclude

**Proposition 3.6** (Building a family of related local polynomial phases)**.** *Let the hypotheses be as in Theorem 1.3. Let $\varepsilon > 0$ be sufficiently small depending on $k, \theta, \eta$, and suppose that $X$ is sufficiently large depending on $\theta, \eta, \varepsilon, k$. Then there exist $P', P'' \in [X^{\varepsilon^2/2}, X^{\varepsilon}]$, a large $(\frac{X}{P'P''}, \frac{H}{P'P''})$-family $\mathcal{I}''$, and local polynomial phases $\phi''_{I''} \in \Phi_{I''}$ for each $I'' \in \mathcal{I}''$ such that*

$$|\langle f, \phi''_{I''}\rangle| \gg 1 \tag{36}$$

*for all $I'' \in \mathcal{I}''$. Furthermore, there exists a collection $\mathcal{Q}$ of $\gg \pi_0(P')^2 \frac{X}{H}$ quadruples $(I''_1, I''_2, p'_1, p'_2)$ with $I''_1, I''_2$ distinct intervals in $\mathcal{I}''$ and $p'_1, p'_2$ distinct primes in $[P', 2P']$, such that $I''_1$ lies within $50\frac{H}{P'P''}$ of $\frac{p'_2}{p'_1} I''_2$ (so in particular $\frac{1}{p'_2} I''_1 \sim \frac{1}{p'_1} I''_2$), and such that*

$$(\frac{1}{p'_2})_* \phi''_{I''_1} \sim_{\frac{1}{p''}} (\frac{1}{p'_1})_* \phi''_{I''_2} \tag{37}$$

*for a large set of primes $p''$ in $[P''/2, P'']$. (The implied constants in the conclusions may depend on the implied constants in the hypotheses.)*

*Proof.* One basically repeats [26, Proof of Proposition 3.2] more or less verbatim, but replacing [26, Proposition 3.1] by Proposition 3.4. For the convenience of the reader we now outline some more details of the argument. By two applications of Proposition 3.4 (arguing exactly as in the proof of [26, Proposition 3.2] down to [26, (41)]), we can find $P' \in [X^{\varepsilon^2}, X^{\varepsilon}]$ and $P'' \in [(X/P')^{\varepsilon^2}, (X/P')^{\varepsilon}] \subset [X^{\varepsilon^2/2}, X^{\varepsilon}]$, an $(X/P', H/P')$-family $\mathcal{I}'$ of

intervals, an $(X/P'P'', H/P'P'')$-family $\mathcal{I}''$ of intervals, and functions $\phi'_{I'}, \phi''_{I''} \in \Phi$ associated to each $I' \in \mathcal{I}', I'' \in \mathcal{I}''$ with the following properties:

- One has (36) for all $I'' \in \mathcal{I}''$.
- For each $I' \in \mathcal{I}'$, there are $\gg \pi_0(P')$ pairs $(I, p')$ with $I \in \mathcal{I}$ and $p'$ a prime in $[P', 2P']$ such that $I/p'$ lies within $3H/P'$ of $I'$ and
$$(\frac{1}{p'})_* \phi_I \sim_1 \phi'_{I'}. \tag{38}$$
- For each $I'' \in \mathcal{I}''$, there are $\gg \pi_0(P'')$ pairs $(I', p'')$ with $I' \in \mathcal{I}'$ and $p''$ a prime in $[P''/2, P'']$ such that $I'/p''$ lies within $3\frac{H}{P'P''}$ of $I''$, and
$$(\frac{1}{p''})_* \phi_{I'} \sim_1 \phi''_{I''}. \tag{39}$$

Note that the property (36) only depends on the values of $\phi''_{I''}$ on the integers. Thus, by Proposition 3.5(i), we may without loss of generality upgrade (39) to

$$(\frac{1}{p''})_* \phi_{I'} \sim_{\frac{1}{p''}} \phi''_{I''} \tag{40}$$

without impacting (36) or any of the other properties listed above. Henceforth we shall assume that (40) holds. Applying Cauchy-Schwarz (as in the continuation of the proof of [26, Proposition 3.2] down to [26, (43)]), we can now find $\gg \pi_0(P')^2 \pi_0(P'') \frac{X}{H}$ octuplets[12] $(I, I'_1, I'_2, I''_1, I''_2, p'_1, p'_2, p'')$ where

- $I \in \mathcal{I}$, $I'_1, I'_2 \in \mathcal{I}'$, $I''_1, I''_2 \in \mathcal{I}''$;
- $p'_1, p'_2$ are primes in $[P', 2P']$, and $p''$ is a prime in $[P''/2, P'']$, with $p'_1 \neq p'_2$;
- For $i = 1, 2$, $\frac{1}{p'_i} I$ lies within $3\frac{H}{P'}$ of $I'_i$, and $\frac{1}{p''} I'_i$ lies within $3\frac{H}{P'P''}$ of $I''_i$.
- For each $i = 1, 2$, we have
$$(\frac{1}{p'_i})_* \phi_I \sim_1 \phi'_{I'_i} \tag{41}$$
and
$$(\frac{1}{p''})_* \phi'_{I'_i} \sim_{\frac{1}{p''}} \phi''_{I''_i}. \tag{42}$$

From (41) and Proposition 3.2(ii) we have for $i = 1, 2$ that

$$(\frac{1}{p'_i p''})_* \phi_I \sim_{\frac{1}{p''}} (\frac{1}{p''})_* \phi'_{I'_i}$$

and hence by (42) and Proposition 3.2(i)

$$(\frac{1}{p'_i p''})_* \phi_I \sim_{\frac{1}{p''}} \phi''_{I''_i}$$

[12]For a visualization of the dependencies between the intervals $I$, $I'_1$, $I''_1$, $I'_2$ and $I''_2$, we refer to [26, Figure 8].

and thus by Proposition 3.2(ii), (iii)

$$(\frac{1}{p'_1p'_2p''})_*\phi_I \sim_{\frac{1}{p''}} (\frac{1}{p'_{3-i}})_*\phi''_{I''_i}$$

and thus by Proposition 3.2(i) we obtain (37). The proposition now follows by repeating the remainder of the proof of [26, Proposition 3.2] (where one estimates how many quadruples arise from these octuplets). □

One should think of the set $\mathcal{Q}$ of quadruples $e = (I''_1, I''_2, p'_1, p'_2)$ produced by the above proposition as a family of "edges" of a certain graph with vertex set $\mathcal{I}''$. Now, we adapt the graph-theoretic arguments in [26, §4] to locate lots of quadruples $e = (I''_1, I''_2, p'_1, p'_2)$ in $\mathcal{Q}$ for which one has a lot of structural control on the local polynomial phases $\phi''_{I''_1}, \phi''_{I''_2}$, and their relationship to each other. For the rest of this section we introduce the quantities

$$N := \#\mathcal{I}'' \asymp \frac{X}{H} \quad \text{and} \quad d := \pi_0(P')^2. \tag{43}$$

We say that a quantity $a$ is of *polynomial size* if one has $a = O(X^{O(1)})$. For instance, $P', P'', H, X, N, d$ are all of polynomial size.

**Proposition 3.7** (Local structure of $\phi''$)**.** *Let the hypotheses be as in Theorem 1.3, and let $\varepsilon, X, P', P'', \mathcal{I}'', \phi''_{I''}, \mathcal{Q}$ be as in Proposition 3.6. Let $\ell_1$, $\ell_2$ be even integers such that*

$$d^{\ell_1}, d^{\ell_2} \geq N^2 d^{10}. \tag{44}$$

*(Note from the lower bound on $P'$ that we can choose $\ell_1, \ell_2 = O_\varepsilon(1)$). We allow implied constants to depend on $\varepsilon, \ell_1, \ell_2$. Then, for a subset $\mathcal{Q}'$ of the quadruples $e = (I''_1, I''_2, p'_1, p'_2)$ in $\mathcal{Q}$ of cardinality $\gg dN$, one can find a collection $\mathcal{A}_e$ of quadruples $\vec{a} = (a_1, a_2, b_1, b_2)$ of natural numbers of cardinality $\asymp d^{\ell_1+\ell_2}/N^2$, and a large collection $\mathcal{P}_{e,\vec{a}}$ of primes in $[P''/2, P'']$ associated to each $\vec{a} \in \mathcal{A}_e$, with the following properties:*

(i) *One has*

$$(\frac{1}{p'_2})_*\phi''_{I''_1} \sim_{\frac{1}{\prod \mathcal{P}_{e,\vec{a}}}} (\frac{1}{p'_1})_*\phi''_{I''_2}. \tag{45}$$

*Here the implied constants in the equivalence relation do not depend on $\ell_1$ or $\ell_2$.*

(ii) *For $i = 1, 2$, $a_i, b_i$ are products of $\ell_i$ primes in $[P', 2P']$; in particular*

$$a_i, b_i \asymp (P')^{\ell_i}, \tag{46}$$

*so $a_i, b_i$ are of polynomial size. Furthermore, we have*

$$a_i - b_i \asymp \frac{1}{N} a_i. \tag{47}$$

(iii) *For $i = 1, 2$, we have the approximate dilation invariance*

$$(\frac{1}{a_i})_*\phi''_{I''_i} \sim_{\frac{1}{\prod \mathcal{P}_{e,\vec{a}}}} (\frac{1}{b_i})_*\phi''_{I''_i}. \tag{48}$$

*Here the implied constants in the equivalence relation may depend on $\ell_i$, but not on the complementary parameter $\ell_{3-i}$.*

For the arguments in this section, one could take the parameters $\ell_1, \ell_2$ to be equal to each other, but in the next section it will be convenient to allow $\ell_1, \ell_2$ to be distinct (in fact in that section we will take $\ell_1$ to be very large compared to $\ell_2$). The specified dependence of parameters in (45), (48) on $\ell_1, \ell_2$ will be of no relevance in the current arguments, but will be crucially exploited in the next section.

*Proof.* Running the proof of [26, Proposition 4.1] all the way down to [26, (53)] (with the role of $k$ replaced by $\ell_1$ and $\ell_2$, noting that the argument works perfectly well when the two cycles in the graph have different length), with Proposition 3.6 playing the role of [26, Proposition 3.2], we conclude that we can find $\gg d^{\ell_1+\ell_2+1}/N$ $(\ell_1+\ell_2)$-tuples

$$\vec{I}'' := (I''_{j,i})_{i=1,2;j\in\{0,1,\dots,\ell_i-1\}} \in (\mathcal{I}'')^{\ell_1+\ell_2}$$

which are "non-degenerate and very good" in the sense that they obey the following axioms:

(i) If $i = 1, 2$ and $j = 0, \dots, \ell_i - 1$ then there exist (uniquely determined) distinct primes $p'_{1,j,i}, p'_{2,j,i} \in [P', 2P']$ such that $I''_{j+1,i}$ lies within $100\frac{H}{P'P''}$ of $\frac{p'_{1,j,i}}{p'_{2,j,i}} I''_{j,i}$ (with the cyclic convention $I''_{\ell_i,i} = I''_{0,i}$). In particular $\frac{1}{p'_{1,j,i}} I''_{j+1,i} \sim \frac{1}{p'_{2,j,i}} I''_{j,i}$.

(ii) There also exist distinct primes $p'_1, p'_2 \in [P', 2P']$ such that $(I''_{0,1}, I''_{0,2}, p'_1, p'_2) \in \mathcal{Q}$. In particular, $I''_{0,2}$ lies within $100\frac{H}{P'P''}$ of $\frac{p'_1}{p'_2} I''_{0,1}$ and hence $\frac{1}{p'_1} I''_{0,2} \sim \frac{1}{p'_2} I''_{0,1}$.

(iii) For $i = 1, 2$, the primes $p'_{1,j,i}$, $j = 0, \dots, \ell_i - 1$ are distinct from the primes $p'_{2,j,i}$, $j = 0, \dots, \ell_i - 1$. In particular we have the non-degeneracy condition

$$\prod_{j=0}^{\ell_i-1} p'_{2,j,i} - \prod_{j=0}^{\ell_i-1} p'_{1,j,i} \neq 0 \tag{49}$$

for $i = 1, 2$.

(iv) There exists a large collection $\mathcal{P}(\vec{I}'')$ of primes in $[P''/2, P'']$ such that

$$(\frac{1}{p'_{2,j,i}})_* \phi''_{I''_{j,i}} \sim_{\frac{1}{Q}} (\frac{1}{p'_{1,j,i}})_* \phi''_{I''_{j+1,i}} \tag{50}$$

for all $j = 0, \dots, \ell_i - 1$ and $i = 1, 2$, and similarly

$$(\frac{1}{p'_2})_* \phi''_{I''_{0,1}} \sim_{\frac{1}{Q}} (\frac{1}{p'_1})_* \phi''_{I''_{0,2}}, \tag{51}$$

where $Q$ is the modulus

$$Q := \prod \mathcal{P}(\vec{I}''). \tag{52}$$

The relationships between the intervals $I''_{j,i}$ can be schematically described by an $\ell_1$-cycle and an $\ell_2$-cycle linked by an edge; see [26, Figure 10] for an example of this diagram in the case $\ell_1 = \ell_2 = 4$.

We note that in [26] the distinctness of the primes $p'_{1,j,i}$ and the primes $p'_{2,j,i}$ in (iii) was not established. However one can obtain this reduction as follows. For the sake of notation we eliminate the contribution of the case when one has a collision $p'_{1,0,1} = p'_{2,0,1}$;

the other cases are treated similarly. Firstly observe that from iterating axiom (i) using the equivalence relation and dilation invariance properties of $\sim$, we have

$$\frac{\prod_{j=0}^{\ell_i-1} p'_{1,j,i}}{\prod_{j=0}^{\ell_i-1} p'_{2,j,i}} I''_{0,i} \sim I''_{0,i}$$

and hence

$$\left|\prod_{j=0}^{\ell_i-1} p'_{2,j,i} - \prod_{j=0}^{\ell_i-1} p'_{1,j,i}\right| \lesssim \frac{1}{N}(P')^{\ell_i} \tag{53}$$

for $i = 1, 2$. If $p'_{1,0,1} = p'_{2,0,1}$, we can cancel one factor in the $i = 1$ case and conclude that

$$\left|\prod_{j=1}^{\ell_1-1} p'_{2,j,1} - \prod_{j=1}^{\ell_1-1} p'_{1,j,1}\right| \lesssim \frac{1}{N}(P')^{\ell_1-1}.$$

Using [26, Lemma 2.6], the number of primes $p'_{1,j,i}, p'_{2,j,i}$ that can obey all these constraints is bounded by

$$\ll \pi_0(P')\frac{d^{\ell_1-1}}{N}\frac{d^{\ell_2}}{N} \ll \frac{d^{\ell_1+\ell_2-1/2}}{N^2}.$$

Since the tuple $\vec{I}''$ is determined by the quadruple $(I''_{0,1}, I''_{0,2}, p'_1, p'_2) \in \mathcal{Q}$ and the above primes, and since $I''_{0,1}, I''_{0,2}$ uniquely determine $p'_1, p'_2$, we conclude that the number of tuples of this type is bounded by $O(d^{\ell_1+\ell_2+1/2}/N)$, and so these tuples can be removed without significantly affecting the total number of tuples. Similarly for other collisions.

In a similar spirit, we may improve the non-degeneracy bound property (49) to

$$\left|\prod_{j=0}^{\ell_i-1} p'_{2,j,i} - \prod_{j=0}^{\ell_i-1} p'_{1,j,i}\right| \gg \frac{1}{N}(P')^{\ell_i} \tag{54}$$

by the following argument. Suppose that we had

$$\left|\prod_{j=0}^{\ell_i-1} p'_{2,j,i} - \prod_{j=0}^{\ell_i-1} p'_{1,j,i}\right| \leq c\frac{1}{N}(P')^{\ell_i}$$

for some $i = 1, 2$, and some $c > 0$ to be chosen later. From (53) with $i$ replaced by $3 - i$ we also have

$$\left|\prod_{j=0}^{\ell_{3-i}-1} p'_{2,j,3-i} - \prod_{j=0}^{\ell_{3-i}-1} p'_{1,j,3-i}\right| \lesssim \frac{1}{N}(P')^{\ell_{3-i}} \tag{55}$$

By two applications of [26, Lemma 2.6], the number of tuples $(p'_{l,j,i})_{l,i=1,2;j=0,\dots,\ell_i-1}$ of primes in $[P', 2P']$ with these properties is $O(cd^{\ell_1+\ell_2}/N^2)$. Since the tuple $\vec{I}''$ is determined by $(I''_{0,1}, I''_{0,2}, p'_1, p'_2)$ and the above primes, we conclude that the number of tuples $\vec{I}''$ arising in this fashion is at most $O(cd^{\ell_1+\ell_2+1}/N)$. For $c$ small enough, this is less than (say) half of

the tuples of $\vec{I}''$ currently under consideration, so on removing those tuples we obtain the bound (54).

If we apply Proposition 3.2(ii) to (50) with the dilation factor

$$\left(\prod_{0\leq j'<j} p'_{1,j',i}\right)\left(\prod_{j<j'<\ell_i} p'_{2,j',i}\right)$$

we conclude that

$$(\frac{1}{a_{j,i}})_*\phi''_{I''_{j,i}} \sim_{\frac{1}{Q}} (\frac{1}{a_{j+1,i}})_*\phi''_{I''_{j+1,i}}$$

for $j = 0,\dots,\ell-1$, where

$$a_{j,i} \coloneqq \left(\prod_{0\leq j'<j} p'_{1,j',i}\right)\left(\prod_{j\leq j'<\ell_i} p'_{2,j',i}\right)$$

for $j = 0,\dots,\ell_i$. Observe that the intervals $\frac{1}{a_{j,i}}I''_{j,i}$ all have length $\asymp (P')^{-\ell_i}\frac{H}{P'P''}$ and are comparable to each other in the sense of the relation $\sim$. Applying Proposition 3.2(i), (iii) $O(\ell_i)$ times, we conclude that

$$(\frac{1}{a_{0,i}})_*\phi''_{I''_{0,i}} \sim_{\frac{1}{Q}} (\frac{1}{a_{\ell_i,i}})_*\phi''_{I''_{0,i}}.$$

Since $a_{0,i}, a_{\ell_i,i}$ are the product of $\ell_i$ distinct primes in $[P', 2P']$, we have

$$a_{0,i}, a_{\ell_i,i} \asymp (P')^{\ell_i}. \tag{56}$$

Also, from the fundamental theorem of arithmetic, once one fixes $I''_{0,1}, I''_{0,2}$, each quadruplet $(a_{0,1}, a_{\ell_1,1}, a_{0,2}, a_{\ell_2,2})$ is associated to at most $O(1)$ tuples $\vec{I}''$ (note that from the above axiom (i) that $I''_{j+1,i}$ is uniquely determined by $I''_{j,i}$, $p'_{1,j,i}, p'_{2,j,i}$).

On the other hand, since $\frac{1}{a_{0,i}}I''_{0,i} \sim \frac{1}{a_{\ell_i,i}}I''_{0,i}$, we have

$$\left(\frac{1}{a_{\ell_i,i}} - \frac{1}{a_{0,i}}\right)(P')^{-\ell_i}\frac{X}{P'P''} \ll (P')^{-\ell_i}\frac{H}{P'P''}$$

which simplifies using (56), (43) to

$$a_{\ell_i,i} - a_{0,i} \ll \frac{(P')^{\ell_i}}{N}.$$

From (54) we get the corresponding lower bound. If we set $a_i$ to be the larger of $a_{\ell_i,i}, a_{0,i}$ and $b_i$ to be the smaller, then we have the properties claimed in (ii), (iii) of the proposition, while (i) follows from (51).

The counting argument at the end of the proof of [26, Proposition 4.1] (which is based on the estimate in [26, Lemma 2.6]) shows that each quadruple $e$ in $\mathcal{Q}$ is associated to at most $O(d^{\ell_1+\ell_2}/N^2)$ tuples $\vec{I}''$ of the above form, and $\mathcal{Q}$ has cardinality $O(dN)$, hence there is a subset $\mathcal{Q}'$ of $\mathcal{Q}$ of cardinality $\gg dN$ such that each $e \in \mathcal{Q}'$ is associated to $\asymp d^{\ell_1+\ell_2}/N^2$

tuples $\vec{I}''$, which by the previous discussion generates $\asymp d^{\ell_1+\ell_2}/N^2$ quadruples $(a_1, b_1, a_2, b_2)$ obeying the required properties (i), (ii), (iii). The claim follows. □

In this section the precise values of $\ell_1, \ell_2$ are not important; we can select them to be any bounded even integers obeying (44). In [26], $\ell_1, \ell_2$ were essentially chosen to be the minimal even integer obeying (44), so that one could make $a_i - b_i$ as small as possible; however this will convey no significant advantage in our current arguments.

While the above proposition produces a large family $\mathcal{A}_e$ of quadruples $\vec{a}$ associated to each $e \in \mathcal{Q}'$, in the argument below it will suffice to just use a single such quadruple $\vec{a}$; this was also the case in the previous paper [26]. However, when we work with nilsequences in the next section, it will become necessary to use multiple quadruples $\vec{a}$ for each $e \in \mathcal{Q}'$.

Thus far we have not exploited the polynomial phase structure of functions in $\mathcal{P}$ beyond the properties in Proposition 3.2 and Proposition 3.3. Now we make heavier use of this structure in order to "solve" the approximate dilation invariance relation (48) produced by Proposition 3.7, using just a single quadruple from $\mathcal{A}_e$. The following proposition asserts, roughly speaking, that this equation is only solvable when the local polynomial phases $\phi''_{I''_i}(t)$ "pretend" to be like the character $t^{iT}$ on $I''_i$ for some real number $T = T_{I''_1, I''_2}$. Let us say that a polynomial $\gamma \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is *$Q$-rational* for some $Q$ if it lies in $\mathrm{Poly}_{\leq k}(\frac{q}{Q}\mathbb{Z} \to \mathbb{Z})$ for some natural number $q$ of polynomial size.

**Proposition 3.8** (Solving the approximate dilation invariance)**.** *Let the notation and hypotheses be as in Proposition 3.7, and write $\phi''_{I''} = (I'', P_{I''})$ for each $I''$. Then for any of the quadruples $e = (I''_1, I''_2, p'_1, p'_2) \in \mathcal{Q}'$, and any $\vec{a} = (a_1, b_1, a_2, b_2)$ in $\mathcal{A}_e$, there exists a real number*

$$T = T_{I''_1, I''_2} \ll N^{k+1}$$

*and decompositions*

$$P_{I''_i}(t) = \varepsilon_i(t) + \frac{T}{2\pi} \log t + \gamma_i(t)$$

*for $i = 1, 2$ and $t > 0$, where $\varepsilon_i \colon \mathbb{R}^+ \to \mathbb{R}$ is a smooth function obeying the derivative bounds*

$$\varepsilon_i^{(j)}(t) \ll_j |I''_i|^{-j}$$

*for all $j \geq 0$ and $t \in I''_i$, and $\gamma_i$ is a $Q$-rational polynomial with*

$$Q := \prod \mathcal{P}_{e,\vec{a}}.$$

*Here $T, \varepsilon_i$ and $\gamma_i$ may depend on $e$ and $\vec{a}$.*
*Also, we have*

$$\gamma_1(p'_2 \cdot) = \gamma_2(p'_1 \cdot) \bmod \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z}). \tag{57}$$

*Proof.* We abbreviate $P_{I''_i}$ as $P_i$. From (48) we have an identity of the form

$$P_i(a_i t) = \varepsilon''_i(t) + P_i(b_i t) + \gamma'_i(t)$$

where $\varepsilon_i'' \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is smooth on $\frac{1}{a_i} I_1''$ and $\gamma_i' \in \mathrm{Poly}_{\leq k}(\frac{1}{Q}\mathbb{Z} \to \mathbb{Z})$ is $Q$-integral; by a change of variables, we can write this as

$$P_i(a_i t) = \varepsilon_i'(a_i t) + P_i(b_i t) + \gamma_i'(t) \tag{58}$$

where $\varepsilon_i' \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is now smooth on $I_1''$. Taking $k^{\text{th}}$ derivatives to make all functions independent of $t$, we conclude in particular that

$$a_i^k P_i^{(k)} = a_i^k (\varepsilon_i')^{(k)} + b_i^k P_i^{(k)} + (\gamma_i')^{(k)}$$

or equivalently

$$q_i P_i^{(k)} = a_i^k (\varepsilon_i')^{(k)} + (\gamma_i')^{(k)} \tag{59}$$

where[13] $q_i \coloneqq a_i^k - b_i^k$. As $\gamma_i$ is $Q$-integral, we see on taking $k^{\text{th}}$ divided differences (or using Lemma 2.1) that $\gamma_i^{(k)}$ is an integer multiple $c_i Q^k$ of $Q^k$. Thus

$$q_i P_i^{(k)} = O(a_i^k |I_i''|^{-k}) + c_i Q^k$$

From (46) we also know that $q_i$ is a natural number of polynomial size; and from the mean value theorem and (47) we have

$$q_i \asymp \frac{a_i - b_i}{a_i} a_i^k \asymp \frac{1}{N} a_i^k.$$

We thus have

$$P_i^{(k)} = \frac{c_i}{q_i} Q^k + O(N |I_i''|^{-k}).$$

Recalling that $x_{I_i''}$ is the midpoint of $I_i''$, we can write the above estimate as

$$P_i^{(k)} = \frac{c_i}{q_i} Q^k + \frac{(-1)^{k-1}(k-1)!}{2\pi} \frac{T_i}{x_{I_i''}^k} \tag{60}$$

for some real number $T_i$ with the bounds

$$T_i \ll N \left( \frac{X}{P'P''} \right)^k |I_i''|^{-k} \ll N^{k+1}.$$

Motivated by the Taylor expansion around $x_{I_i''}$, we write

$$P_i(t) = \tilde{\varepsilon}_i(t) + \frac{T_i}{2\pi} \log t + \tilde{P}_i(t) + \tilde{\gamma}_i(t) \tag{61}$$

for $t \in \mathbb{R}^+$, where $\tilde{\varepsilon}_i : \mathbb{R}^+ \to \mathbb{R}$ is the Taylor remainder

$$\tilde{\varepsilon}_i(t) = -\frac{T_i}{2\pi} \log t + \frac{T_i}{2\pi} \log x_{I_i''} + \sum_{j=1}^{k} \frac{(-1)^{j-1} T_i}{2\pi j} \frac{(t - x_{I_i''})^j}{x_{I_i''}^j}$$

[13]Note that this choice of $q_i$ explains why our bound on $q_i$ in this lemma is a lot weaker that in [26, Proposition 4.1], even if we try to take $\ell_1, \ell_2$ to be as small as possible. Indeed, if $q_i = a_i^k - b_i^k$ with $a_i - b_i$ small, $a_i^k - b_i^k$ may still be relatively large.

which is a smooth function obeying the bounds

$$\tilde{\varepsilon}_i^{(j)}(t) \ll_j (H/P'P'')^{-j}$$

for $j \geq 0$ and $t \in I_i''$, and $\tilde{\gamma}_i \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is the function

$$\tilde{\gamma}_i(t) \coloneqq \frac{c_i}{q_i}\binom{Qt}{k},$$

and

$$\tilde{P}_i(t) \coloneqq P_i(t) - \tilde{\gamma}_i(t) - \sum_{j=1}^{k} \frac{(-1)^{j-1}T_i}{2\pi j}\frac{(t - x_{I_i''})^j}{x_{I_i''}^j} - \frac{T_i}{2\pi}\log x_{I_i''}$$

is an element of $\mathrm{Poly}_{\leq k-1}(\mathbb{R} \to \mathbb{R})$. We can then write by (58) and (61)

$$\tilde{P}_i(a_i t) = \varepsilon_i^*(a_i t) + \tilde{P}_i(b_i t) + \gamma_i^*(t) \tag{62}$$

for $t \in \mathbb{R}^+$, where

$$\gamma_i^*(t) \coloneqq \gamma_i'(t) + \tilde{\gamma}_i(b_i t) - \tilde{\gamma}_i(a_i t) + \left\lfloor \frac{T_i}{2\pi}\log\frac{b_i}{a_i} \right\rfloor$$

and

$$\varepsilon_i^*(t) \coloneqq \varepsilon_i'(t) + \tilde{\varepsilon}_i\left(\frac{b_i}{a_i}t\right) - \tilde{\varepsilon}_i(t) + \left\{ \frac{T_i}{2\pi}\log\frac{b_i}{a_i} \right\}.$$

By construction, $\gamma_i^*$ is an element of $\mathrm{Poly}_{\leq k}(\frac{q_i}{Q}\mathbb{Z} \to \mathbb{Z})$ that has vanishing $k^{\text{th}}$ derivative, so $\gamma_i^*$ in fact lies in $\mathrm{Poly}_{\leq k-1}(\mathbb{R} \to \mathbb{R})$. From (62) we conclude that $\varepsilon_i^*(a_i t)$ also lies in $\mathrm{Poly}_{\leq k-1}(\mathbb{R} \to \mathbb{R})$, and from the triangle inequality we have

$$(\varepsilon_i^*)^{(j)}(t) \ll (H/P'P'')^{-j}$$

for all $j \geq 0$ and $t \in I_i''$. In conclusion, $\tilde{P}_i$ obeys similar properties to $P_i$ except that all polynomials involved have degree at most $k-1$ instead of at most $k$, and the polynomial $\gamma_i^*$ lies in $\mathrm{Poly}_{\leq k-1}(\frac{q_i}{Q}\mathbb{Z} \to \mathbb{Z})$ rather than $\mathrm{Poly}_{\leq k-1}(\frac{1}{Q}\mathbb{Z} \to \mathbb{Z})$. One can iterate this procedure $k$ times and after collecting terms in the telescoping series, one ends up with a decomposition of the form

$$P_i(t) = \varepsilon_i^{**}(t) + \frac{T_i^{**}}{2\pi}\log t + P_i^{**} + \gamma_i^{**}(t)$$

for $t \in \mathbb{R}^+$, where $T_i^{**}$ is a real number with

$$T_i^{**} \ll N^{k+1},$$

$\varepsilon_i^{**} : \mathbb{R}^+ \to \mathbb{R}$ is a smooth function obeying the derivative estimates

$$(\varepsilon_i^{**})^{(j)}(t) \ll_j (H/P'P'')^{-j}$$

for all $j \geq 0$ and $t \in I_i''$, $P_i^{**} \in \mathbb{R}$ is a constant, and $\gamma_i^{**}$ is $Q$-rational. By splitting $P_i^{**}$ into integer and fractional parts and redistributing these parts to $\gamma_i^{**}$ and $\varepsilon_i^{**}$ respectively, we may assume that $P_i^{**} = 0$, thus

$$P_i(t) = \varepsilon_i^{**}(t) + \frac{T_i^{**}}{2\pi}\log t + \gamma_i^{**}(t) \tag{63}$$

for $t \in \mathbb{R}^+$.

This is almost what we need for the claims of the proposition (excluding (57)), except that the two real numbers $T_1^{**}, T_2^{**}$ are allowed to be unequal. From (51) and Definition 3.1 we have

$$P_1(p_2' t) = \varepsilon^\dagger(p_2' t) + P_2(p_1' t) + \gamma^\dagger(t)$$

for $t \in \mathbb{R}^+$, where $\varepsilon^\dagger \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ is smooth on $I''_{0,1}$, and $\gamma^\dagger$ is $Q$-integral. Inserting (63), we conclude that

$$\frac{T_1^{**}}{2\pi} \log(p_2' t) = \varepsilon^{\dagger\dagger}(p_2' t) + \frac{T_2^{**}}{2\pi} \log(p_1' t) + \gamma^{\dagger\dagger}(t)$$

where $\varepsilon^{\dagger\dagger} \colon \mathbb{R}^+ \to \mathbb{R}$ is given by the formula

$$\varepsilon^{\dagger\dagger}(p_2' t) \coloneqq \varepsilon^\dagger(p_2' t) + \varepsilon_2^{**}(p_1' t) - \varepsilon_1^{**}(p_2' t)$$

and obeys the derivative estimates

$$(\varepsilon^{\dagger\dagger})^{(j)}(t) \ll_j (H/P'P'')^{-j}$$

for all $j \geq 0$ and $t \in I_i''$, and $\gamma^{\dagger\dagger}$ is given by the formula

$$\gamma^{\dagger\dagger}(t) \coloneqq \gamma^\dagger(t) + \gamma_2^{**}(p_1' t) - \gamma_1^{**}(p_2' t). \tag{64}$$

and in particular is $Q$-rational. Let $n_{I''_{0,1}}$ be an integer point of $I''_{0,1}$. From Lemma 2.1 we see that the first derivative $(\gamma^{\dagger\dagger})'(n_{I''_{0,1}})$ takes values in $\frac{Q}{q^k k!}\mathbb{Z}$ for some $q$ of polynomial size. We conclude that

$$\frac{T_1^{**}}{2\pi n_{I''_{0,1}}} = O\left(\frac{P'P''}{H}\right) + \frac{T_2^{**}}{2\pi n_{I''_{0,1}}} \bmod \frac{Q}{q^k k!}\mathbb{Z}. \tag{65}$$

Since $\mathcal{P}_{I_1'', I_2''}$ is a large set of primes in $[P''/2, P'']$, we see from (25) that $Q \gg \exp(cP'')$ for some $c \gg 1$, so in particular[14] $Q$ exceeds $X^C$ for any fixed $C$ if $X$ is large enough. But both sides of (65) are of polynomial size, and thus have magnitude less than $\frac{Q}{2q^k k!}$ for $X$ large enough. Hence we may remove the modulus restriction and conclude that

$$\frac{T_1^{**}}{2\pi n_{I''_{0,1}}} = O\left(\frac{P'P''}{H}\right) + \frac{T_2^{**}}{2\pi n_{I''_{0,1}}}$$

which we can rearrange using (43) as

$$T_1^{**} = T_2^{**} + O(N).$$

If we set $T \coloneqq T_1^{**}$,

$$\gamma_i(t) \coloneqq \gamma_i^{**}(t) + \left\lfloor \frac{T_i^{**} - T}{2\pi} \log n_{I''_{0,1}} \right\rfloor, \text{ and } \varepsilon_i(t) \coloneqq \varepsilon_i^{**}(t) + \frac{T_i^{**} - T}{2\pi} \log t - \left\lfloor \frac{T_i^{**} - T}{2\pi} \log n_{I''_{0,1}} \right\rfloor,$$

[14] We remark that it is this need for $Q$ to be bigger than $X$ that puts a limit on the range of $H$ where one could possibly prove Theorem 1.3 using the strategy of this paper. Since $P'' \leq H^\varepsilon$, we must have $H \geq (\log x)^A$ for any fixed $A$. It turns out that there are further restrictions on the size of $H$ in our proof, coming from the graph theory part of the proof, where factors of $\ell!$ appear, and also from the Vinogradov–Korobov zero-free region. For these reasons, $H$ actually needs to be at least $\exp((\log X)^c)$ for some $c \geq 1/2$.

we obtain all the required claims except for (57). But observe that the previous argument in fact showed that the first derivative of $\gamma^{\dagger\dagger}$ vanished at all integer points of $I''_{0,1}$, and thus vanished identically thanks to Lagrange interpolation; hence $\gamma^{\dagger\dagger}$ is in fact an integer constant. The claim (57) now follows from (64) since $\gamma^\dagger$ is already 1-integral. □

We now follow the arguments in [26, §5] (starting after the proof of [26, Corollary 5.2]. Let $\delta > 0$ be a sufficiently small quantity (depending on $k, \varepsilon, \eta, \theta$) to be chosen later. We assume $X$ (and hence $H$) to be sufficiently large depending on $\delta$, and allow implied constants to depend on $\delta$. Define a *good quadruple* to be a tuple $(I'', T, q, \gamma)$ with $I'' \in \mathcal{I}''$, $T$ a real number with

$$|T| \leq \frac{1}{\delta} N^{k+1}, \tag{66}$$

and $q$ a natural number with

$$1 \leq q \leq X^{1/\delta} \tag{67}$$

and $\gamma$ an element of $\mathrm{Poly}_{\leq k}(\frac{q}{\prod \mathcal{P}}\mathbb{Z} \to \mathbb{Z})$ for some collection $\mathcal{P}$ of primes in $[P''/2, P'']$ of cardinality $\geq \delta \pi_0(P'')$ that do not divide $q$, such that we have a decomposition

$$P_{I''}(t) = \varepsilon(t) + \frac{T}{2\pi} \log t + \gamma(t) \tag{68}$$

for all $t > 0$, where $\phi''_{I''} = (I'', P_{I''})$, and $\varepsilon \colon \mathbb{R}^+ \to \mathbb{R}$ is a smooth function obeying the estimates

$$|\varepsilon^{(j)}(t)| \leq \frac{1}{\delta} (H/P'P'')^{-j} \tag{69}$$

for $t \in I''$ and $0 \leq j \leq k$. We also require that $q$ is the least natural number for which $\gamma \in \mathrm{Poly}_{\leq k}(q\mathbb{Z} \to \mathbb{Z})$.

We will shortly show that Proposition 3.7 yields a lot of pairs of "compatible" good quadruples.

Each interval $I''$ is only associated with a small number of essentially distinct good quadruples. Indeed, we have

**Proposition 3.9.** *Let $I'' \in \mathcal{I}''$, let $K$ be a sufficiently large natural number depending on $\delta$, and let $(I'', T_j, q_j, \gamma_j), j = 1, \dots, K$ be a collection of good quadruples associated to the interval $I''$. Then there exist $1 \leq j < j' \leq K$ with the following properties:*

(i) $q_j = q_{j'}$.
(ii) $\gamma_j = \gamma_{j'} \bmod \mathbb{Z}$. *(Here we view $\mathbb{Z} \subset \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ as the group of constant integer functions).*
(iii) $T_j = T_{j'} + O(N)$.

*(Recall that we allow implied constants to depend on $\delta$.)*

*Proof.* We modify the proof of [26, Proposition 5.3]. For $j = 1, \dots, K$, let $\mathcal{P}_j$ denote the set of primes in $[P''/2, P'']$ associated to the good quadruple $(I'', T_j, q_j, \gamma_j)$. Then

$$\sum_{p'' \in [P''/2, P'']} \sum_{j=1}^{K} 1_{p'' \in \mathcal{P}_j} \gg K \delta \pi_0(P'')$$

and hence by the prime number theorem we have that

$$\sum_{j=1}^{K} 1_{p'' \in \mathcal{P}_j} \gg K\delta$$

for $\gg \delta\pi_0(P'')$ primes $p'' \in [P''/2, P'']$. For $K$ large in terms of $\delta$, we can then find $j, j' \in \{1, \dots, K\}$ such that $\mathcal{P} := \mathcal{P}_j \cap \mathcal{P}_{j'}$ contains $\gg_\delta \pi_0(P'')$ primes $p'' \in [P''/2, P'']$.

From (68), we have for all $j = 1, \dots, K$ that

$$P_{I''}(t) = \varepsilon_j(t) + \frac{T_j}{2\pi} \log t + \gamma_j(t)$$

for all $t > 0$, where $\phi''_{I''} = (I'', P'')$ and $\varepsilon_j \colon \mathbb{R}^+ \to \mathbb{R}$ is smooth with $\varepsilon_j^{(l)}(t) \ll (H/P'P'')^{-l}$ for all $t \in I''$ and $0 \leq l \leq k$. Taking first derivatives, we see that the function

$$\varepsilon_j'(t) + \frac{T_j}{2\pi t} + \gamma_j'(t) \tag{70}$$

is independent of $j$. We now specialize $t$ to an integer point $n_{I''}$ of $I''$. From Lemma 2.1, we have $\gamma_j'(n_I) \in \frac{\prod \mathcal{P}}{q_j^k k!}\mathbb{Z}$. Thus we have

$$\frac{T_j}{2\pi n_{I''}} = \frac{T_{j'}}{2\pi n_{I''}} + O\left(\frac{P'P''}{H}\right) \bmod \frac{\prod \mathcal{P}}{q_j^k q_{j'}^k k!}\mathbb{Z}$$

for all $j, j' \in \{1, \dots, K\}$. Both sides of this equation are of polynomial size, while the modulus $\frac{\prod \mathcal{P}}{q_j^k q_{j'}^k k!}$ is far larger than this thanks to (25). We may thus remove the modulus and conclude that

$$\frac{T_j}{2\pi n_{I''}} = \frac{T_{j'}}{2\pi n_{I''}} + O\left(\frac{P'P''}{H}\right)$$

and hence by (43)

$$T_{j'} = T_j + O(N),$$

giving the conclusion (iii). If we now return to the independence of (70) in $j$, we conclude that

$$\gamma_j'(t) - \gamma_{j'}'(t) = O\left(\frac{P'P''}{H}\right)$$

for all $t \in I''$. By the Bernstein inequality (27), we can thus obtain the bound

$$\gamma_j^{(l)}(n_I) - \gamma_{j'}^{(l)}(n_I) = O\left(\left(\frac{P'P''}{H}\right)^l\right)$$

for all $1 \leq l \leq k$. On the other hand, from Lemma 2.1 the left-hand side lies in $\frac{\prod \mathcal{P}}{q_j^k q_{j'}^k k!}\mathbb{Z}$. Using (25) as before, we conclude that

$$\gamma_j^{(l)}(n_I) - \gamma_{j'}^{(l)}(n_I) = 0$$

for $1 \leq l \leq k$, hence by Taylor expansion $\gamma_j$ and $\gamma_{j'}$ differ by a constant, which must lie in $\mathbb{Z}$ since $\gamma_j, \gamma_{j'} \in \mathrm{Poly}_{\leq k}(q_j q_{j'}\mathbb{Z} \to \mathbb{Z})$. This gives the conclusion (ii). Finally, since $q_j$ is the

minimal natural number for which $\gamma_j \in \mathrm{Poly}_{\leq k}(q_j\mathbb{Z} \to \mathbb{Z})$, and $\gamma_j, \gamma_{j'}$ differ by an integer shift, we conclude (i). □

From this and the greedy algorithm, we conclude the following analogue of [26, Corollary 5.4]:

**Corollary 3.10.** *For each $I'' \in \mathcal{I}''$ there exists a set $\mathcal{F}(I'')$ of triples $(T', q, \gamma')$ of cardinality*

$$\#\mathcal{F}(I'') \ll 1$$

*such that for any good quadruple $(I'', T, q, \gamma)$ there exists a real number $T'$ and a $\gamma' = \gamma \bmod \mathbb{Z}$ such that $(T', q, \gamma') \in \mathcal{F}(I'')$ and*

$$T = T' + O(N).$$

Henceforth we fix the finite sets $\mathcal{F}(I'')$. Now we can obtain many pairs of compatible good quadruples:

**Proposition 3.11.** *For $\gg N\pi_0(P')^2$ pairs $(I_1'', I_2'') \in (\mathcal{I}'')^2$, there exist $T_1, T_2, q, \gamma_1, \gamma_2$ with $(T_i, q, \gamma_i) \in \mathcal{F}(I_i'')$ for $i = 1, 2$ and*

$$T_2 = T_1 + O(N) \tag{71}$$

*Furthermore, for each such pair, there exist primes $p_1', p_2' \in [P', 2P']$ coprime to $q$ such that $I_1''$ lies within $100\frac{H}{P'P''}$ of $\frac{p_2'}{p_1'}I_2''$ with*

$$\gamma_1(p_2'\cdot) = \gamma_2(p_1'\cdot) \bmod \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z}). \tag{72}$$

*Proof.* This will be a modification of the arguments used to establish [26, Proposition 5.5]. From Propositions 3.7 and 3.8, we can find a collection $\mathcal{Q}'$ of quadruples $e = (I_1'', I_2'', p_1', p_2')$ in $\mathcal{Q}$ of cardinality $\gg N\pi_0(P')^2$, such that to each such quadruple $e$ there exists $T, \varepsilon_1, \varepsilon_2, \gamma_1, \gamma_2, Q$ obeying the conclusions of Proposition 3.8 (for some quadruple $\vec{a}$, which will play no further role in the arguments). In particular, each $e \in \mathcal{Q}'$ generates a pair of good quadruples $(I_1'', T_1, q_1, \gamma_1)$, $(I_2'', T_2, q_2, \gamma_2)$ for some $\gamma_1 \in \mathrm{Poly}_{\leq k}(q_1\mathbb{Z} \to \mathbb{Z})$, $\gamma_2 \in \mathrm{Poly}_{\leq k}(q_2\mathbb{Z} \to \mathbb{Z})$ obeying (71), (72). By Corollary 3.10 we may adjust these good quadruples so that $(T_i, q_i, \gamma_i) \in \mathcal{F}(I_i'')$ for $i = 1, 2$.

At present it is possible that $p_i'$ divides $q_j$ for some $i, j = 1, 2$. But, as noted in [26, Proposition 5.5], for each $q_j$ there are only at most $O(1)$ such $p_i'$ that can do this, and by the bounded cardinality of the $\mathcal{F}(I_i'')$, the total number of quadruples $e = (I_1'', I_2'', p_1', p_2')$ that generate such a situation is $O(N\pi_0(P'))$, which is negligible compared to the cardinality of $\mathcal{Q}'$. Thus by refining $\mathcal{Q}'$ we may assume that $p_1', p_2'$ do not divide $q_1$ or $q_2$.

We now claim that $q_1$ and $q_2$ are equal. By the definition of a good quadruple, $\gamma_1$ lies in $\mathrm{Poly}_{\leq k}(q_1\mathbb{Z} \to \mathbb{Z})$; by (72) this implies that $\gamma_2$ lies in $\mathrm{Poly}_{\leq k}(p_2'q_1\mathbb{Z} \to \mathbb{Z})$. On the other hand, $q_2$ is the minimal natural number for which $\gamma_2$ lies in $\mathrm{Poly}_{\leq k}(q_2\mathbb{Z} \to \mathbb{Z})$; by Lemma 2.2, this implies that $q_2$ divides $p_2'q_1$, and similarly $q_1$ divides $p_1'q_2$. Since $p_1', p_2'$ do not divide $q_1, q_2$, we obtain $q_1 = q_2$, and the claim follows. □

As in [26, §5], on the space $Z$ of triples $(T, q, \gamma)$ with $T \in \mathbb{R}$, $q \geq 1$, $\gamma \in \mathrm{Poly}(q\mathbb{Z} \to \mathbb{Z})$ we define the metric

$$d((T_1, q_1, \gamma_1), (T_2, q_2, \gamma_2)) \coloneqq c(\delta)\frac{1}{N}|T_1 - T_2| + 1_{q_1 \neq q_2} + \frac{1}{100}1_{\gamma_1 \neq \gamma_2}$$

with some sufficiently small constant $c(\delta) > 0$. Proposition 3.11 provides one with a collection $\mathcal{S}$ of sextuples $(I_1'', I_2'', (T_1, q_1, \gamma_1), (T_2, q_2, \gamma_2), p_1', p_2')$ of cardinality $\gg N\pi_0(P')^2$ such that

$$d((T_1, q_1, \gamma_1), (T_2, q_2, \gamma_2)) \leq \frac{1}{10}.$$

Applying the mixing lemma in [26, Corollary 5.2], we conclude that there exists a triple $(T_0, q_0, \gamma_0) \in Z$ and a collection $\mathcal{T}$ of quadruples $(I'', T, q, \gamma)$ with $I'' \in \mathcal{I}''$, $(T, q, \gamma) \in \mathcal{F}(I'')$, and $d((T, q, \gamma), (T_0, q_0, \gamma_0)) \leq \frac{1}{5}$ such that

$$\#\mathcal{T} \gg N$$

and such that there are $\gg Nd$ sextuples $(I_1'', I_2'', (T_1, q', \gamma_1), (T_2, q', \gamma_2), p_1', p_2')$ such that $(I_i'', T_i, q', \gamma_i) \in \mathcal{T}$ and $p_1', p_2'$ distinct primes in $[P', 2P']$ with $I_1''$ lying within $100\frac{H}{P'P''}$ of $\frac{p_2'}{p_1'}I_2''$ (so in particular $I_1'' \sim \frac{p_2'}{p_1'}I_2''$), with $p_1', p_2'$ coprime to $q'$, and obeying the properties (71), (72).

In particular, if $(I'', T, q, \gamma) \in \mathcal{T}$, then $q = q_0$ and

$$T = T_0 + O(N). \tag{73}$$

From this and (66), we conclude in particular that

$$T_0 \ll N^{k+1}. \tag{74}$$

At present our upper bound (67) on $q = q_0$ is quite large (and significantly worse than in [26]). Nevertheless, we can improve the bound on $q_0$ after first establishing the following variant of [26, Lemma 2.6]:

**Lemma 3.12.** *Let $m, \ell \in \mathbb{N}$ and $P', N \geq 3$ be such that $(P')^{\ell-1} \gg N$. Let $q \geq 1$. Then the number of $2\ell$-tuples $(p_{1,1}', \dots, p_{1,\ell}', p_{2,1}', \dots, p_{2,\ell}')$ of primes in $[P', 2P']$ not dividing $q$ obeying the condition*

$$\left|\prod_{j=1}^{\ell} p_{2,j}' - \prod_{j=1}^{\ell} p_{1,j}'\right| \leq C\frac{(P')^\ell}{N}$$

*and*

$$\prod_{j=1}^{\ell}(p_{2,j}')^m = \prod_{j=1}^{\ell}(p_{1,j}')^m \bmod q$$

*for some $C \geq 1$ is bounded by*

$$\ll_{\ell,C,m} \frac{d^\ell}{N}\left(\frac{m^{\omega(q)}}{\phi(q)} + \frac{1}{\log N}\right),$$

*where $\omega(q)$ denotes the number of prime factors of $q$.*

*Proof.* This follows the same Dirichlet character argument used to prove [26, Lemma 2.6], with the one main difference being that the indicator $1_{\chi=\chi_0}$ is replaced by $1_{\chi^m=\chi_0}$. This latter condition is attained for at most $m^{\omega(q)}$ characters $\chi$ with period $q$, explaining the additional factor of $m^{\omega(q)}$ here compared with [26, Lemma 2.6]. □

We now have

**Proposition 3.13.** $q_0 \ll 1$.

*Proof.* This will be a modification of the proof of [26, Proposition 5.6], using Lemma 3.12 in place of [26, Lemma 2.6]. Let $\ell$ be the first even natural number such that $d^\ell \geq N^{2+\varepsilon}$. Arguing as in the proof of [26, Proposition 5.6], we can find $\gg d^\ell$ tuples

$$(Q_0, \dots, Q_{\ell-1}) \in \mathcal{T}^\ell$$

such that if we write $Q_j = (I_j'', T_j, q_0, \gamma_j)$ for $j = 0, \dots, \ell$ (with the convention $Q_\ell = Q_0$) then for each $j = 0, \dots, \ell-1$, there exist primes $p'_{j,1}, p'_{j,2} \in [P', 2P']$ such that

$$\gamma_j(p'_{j,2}\cdot) = \gamma_{j+1}(p'_{j,1}\cdot) \bmod \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$$

and such that $I_j'' \sim \frac{p'_{j,2}}{p'_{j,1}} I_{j+1}''$. From the first claim we have

$$\gamma_j\left((\prod_{i=0}^{j-1} p'_{i,1})(\prod_{i=j}^{\ell-1} p'_{i,2})\cdot\right) = \gamma_{j+1}\left((\prod_{i=0}^{j} p'_{i,1})(\prod_{i=j+1}^{\ell-1} p'_{i,2})\cdot\right) \bmod \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$$

for $j = 0, \dots, \ell-1$, which by transitivity implies that

$$\gamma_0\left((\prod_{i=0}^{\ell-1} p'_{i,2})\cdot\right) = \gamma_0\left((\prod_{i=0}^{\ell-1} p'_{i,1})\cdot\right) \bmod \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z}). \tag{75}$$

Similarly, we have that $I_0'' \sim \frac{\prod_{i=0}^{\ell-1} p'_{i,2}}{\prod_{i=0}^{\ell-1} p'_{i,1}} I_0''$, which implies that

$$\prod_{i=0}^{\ell-1} p'_{i,2} - \prod_{i=0}^{\ell-1} p'_{i,1} \ll \frac{(P')^\ell}{N}.$$

Now we analyze the condition (75). We write the polynomial $\gamma_0$ as

$$\gamma_0(t) = \sum_{m=0}^{k} \frac{a_m}{b_m} t^m$$

where $b_m$ are natural numbers and each $a_m$ is an integer coprime to $b_m$. Clearly $\gamma_0 \in \mathrm{Poly}_{\leq k}(b_1 \dots b_k\mathbb{Z} \to \mathbb{Z})$, and hence $q_0 \leq b_1 \dots b_k$. In particular, there exists $1 \leq m \leq k$ such that $b_m \geq q_0^{1/k}$. From (75) and Lemma 2.1, and extracting the $t^m$ coefficient, we see that

$$\left(\prod_{i=0}^{\ell-1} p'_{i,2}\right)^m \frac{a_m}{b_m} = \left(\prod_{i=0}^{\ell-1} p'_{i,1}\right)^m \frac{a_m}{b_m} \bmod \frac{1}{k!}\mathbb{Z}$$

and hence

$$\left(\prod_{i=0}^{\ell-1} p'_{i,2}\right)^m = \left(\prod_{i=0}^{\ell-1} p'_{i,1}\right)^m \mod \frac{b_m}{(b_m, k!)}.$$

By Lemma 3.12 (and bounding $\frac{m^{\omega(q)}}{\phi(q)} \ll q^{-1/2}$, say), we conclude that the total number of tuples of primes $(p'_{i,1}, p'_{i,2})_{0\le i<\ell}$ is at most

$$\ll \frac{d^\ell}{N}\left(q_0^{-1/2k} + \frac{1}{\log X}\right).$$

Since there are $\ll N$ choices for the interval $I''_1$, and $I''_1$ and $(p'_{i,1}, p'_{i,2})_{0\le i<\ell}$ determine the other $I''_j$, and we have $\#F(I''_j) \ll 1$, we deduce that the number of tuples $(Q_0, \dots, Q_{\ell-1}) \in \mathcal{T}^\ell$ is in fact $\ll d^\ell(q_0^{-1/2k} + (\log X)^{-1})$. Comparing with the lower bound we had for the number of these tuples, we must have

$$q_0^{-1/2k} + \frac{1}{\log X} \gg 1,$$

giving the claim. □

Let $(I'', T, q_0, \gamma) \in \mathcal{T}$, then from (36) one has

$$\left|\sum_{n\in I''} f(n)e(-P_{I''}(n))\right| \gg |I''|.$$

Let $H^* := c\frac{H}{P'P''}$ for a sufficiently small $c > 0$. Then one has

$$\sum_{n\in I''} f(n)e(-P_{I''}(n)) = \frac{1}{H^*}\int_{I''} \sum_{n\in[x,x+H^*]} f(n)e(-P_{I''}(n))\, dx + O(H^*)$$

and thus by the triangle inequality we have (for $c$ small enough)

$$\int_{I''} \left|\sum_{n\in[x,x+H^*]} f(n)e(-P_{I''}(n))\right| dx \gg |I''|H^*. \tag{76}$$

For $n \in [x, x+H^*] \cap \mathbb{Z}$, we have from (68) that

$$P_{I''}(n) = \varepsilon(n) + \frac{T}{2\pi}\log n + \gamma(n);$$

from (69) we have

$$\varepsilon(n) = \varepsilon(x) + O(c)$$

while from (73) one has

$$\frac{T}{2\pi}\log n = \frac{T_0}{2\pi}\log n + \frac{T-T_0}{2\pi}\log x + O(c).$$

The effect of the $O(c)$ error to (76) is negligible if $c$ is small enough, and the constant terms $\varepsilon(x), \frac{T-T_0}{2\pi}\log x$ disappear once the absolute value signs in (76) are applied. We conclude that

$$\int_{I''}\left|\sum_{n\in[x,x+H^*]} f(n)n^{-iT_0}e(-\gamma(n))\right| dx \gg |I''|H^*.$$

The function $e(-\gamma(n))$ is periodic modulo $q_0$. Since $q_0 = O(1)$, we can expand $e(\gamma(n))$ as a linear combination of $O(1)$ functions of the form $1_{q_1|n}\chi(n/q_1)$, where $q_1$ divides $q_0$ and $\chi$ is a Dirichlet character of period $q_0/q_1$. We conclude that there exists $q_1, \chi$ of this form such that

$$\int_{I''}\left|\sum_{n\in[x,x+H^*]} f(n)n^{-iT_0}1_{q_1|n}\overline{\chi}(n/q_1)\right| dx \gg |I''|H^*.$$

Since each $I''$ is associated to $O(1)$ quadruples in $\mathcal{T}$, there are $\gg X/H$ intervals $I'' \in \mathcal{I}''$ for which we have an estimate of this form. At present $q_1, \chi$ can depend on $I''$, but there are only $O(1)$ choices for these quantities, so by the pigeonhole principle we may make $q_1, \chi$ independent of $I''$, while still retaining $\gg X/H$ intervals. Summing in these intervals, we conclude that

$$\int_{X/4P'P''}^{4X/P'P''}\left|\sum_{n\in[x,x+H^*]} f(n)n^{-iT_0}1_{q_1|n}\overline{\chi}(n/q_1)\right| dx \gg \frac{X}{P'P''}H^*.$$

Arguing exactly as in the final part of [26, §5] (namely, applying the complex-valued version [25] of the main result from [23]), we conclude that

$$M(f;T,Q) \ll 1$$

for some $T \ll \frac{X^{k+1}}{H^{k+1}}$ and $Q \ll 1$, and Theorem 1.3 follows.

## 4. Local correlation with nilsequences

4.1. **The set-up.** In this section we prove Theorem 1.5. Our argument shall closely follow in large parts the proof of Theorem 1.3, except that the space $\Phi$ of local polynomial phases will be replaced by a different family $\Psi$ of local nilsequences, and significantly more effort needs to be expended to "solve" the approximate dilation invariance "equations".

Recall that a *degree $k$ filtered nilmanifold* $G/\Gamma$ is a quotient space $G/\Gamma$, where

- $G$ is a connected, simply connected Lie group equipped with a filtration $G_\bullet = (G_i)_{i\geq 0}$ of closed connected subgroups $G_i$, with $G_0 = G_1 = G$, $G_i \supset G_{i+1}$ for all $i$, $G_i = \{1\}$ for $i > k$, and $[G_i, G_j] \subset G_{i+j}$ for $i, j \geq 0$ (note in particular that this implies that $G$ is nilpotent);
- $\Gamma$ is a discrete subgroup of $G$ such that the subgroups $\Gamma_i := G_i \cap \Gamma$ are cocompact subgroups of $G_i$ for each $i$, so that the quotient spaces $G_i/\Gamma_i$ are all compact.

Let $G$ be a connected, simply connected nilpotent Lie group. Then $G$ is isomorphic to a matrix Lie group (a Lie group consisting of invertible $n \times n$ complex matrices for some $n$); see e.g., [20, Proposition 16.2.6], and so for the following discussion we may assume

without loss of generality that $G$ is a matrix Lie group. The Lie algebra of $G$, defined as the tangent space of $G$ at the identity, will be denoted $\log G$. The matrix exponential map $\exp\colon \log G \to G$ is then a diffeomorphism (see e.g., [20, Corollary 11.2.7]), and hence we have a well-defined logarithm map $\log\colon G \to \log G$ inverting this map; similarly we have the diffeomorphism $\log\colon G_i \to \log G_i$ where $\log G_i$ is the Lie algebra of $G_i$. We define exponentiation $g^t$ for any $g \in G$ and $t \in \mathbb{R}$ by the familiar formula

$$g^t \coloneqq \exp(t \log g), \tag{77}$$

so in particular $\log(g^t) = t \log g$. We place an arbitrary Euclidean metric on the vector space $\log G$, and allow implied constants to depend on $G$ and this metric. If $g \in G$ and $X > 0$, we then write $g = O(X)$ as shorthand for $|\log g| = O(X)$. We also place an arbitrary smooth metric $d$ on $G/\Gamma$ (for instance, one could take the Carnot–Carathéodory metric associated to the metric on $\log G$, although it is not essential here that we do so), and define the Lipschitz norm of a function $F\colon G/\Gamma \to \mathbb{C}$ to be

$$\|F\|_{\mathrm{Lip}} \coloneqq \sup_{x \in G/\Gamma} |F(x)| + \sup_{x,y \in G/\Gamma : x \neq y} \frac{|F(x) - F(y)|}{d(x,y)}$$

and call a function $F$ *Lipschitz continuous* if its Lipschitz norm is finite.

The presence of the logarithm here may seem strange to those accustomed to more "abelian" analysis, but for nilpotent groups (written multiplicatively) one should view log, exp, and $(g,t) \mapsto g^t$ as polynomial maps, as the following example illustrates:

**Example 4.1** (Heisenberg group)**.** *Take $G$ to be the Heisenberg group $G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}$, with filtration $G_0 = G_1 = G$, $G_2 = \begin{pmatrix} 1 & 0 & \mathbb{R} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, and $G_i = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}$ for all $i > 2$. Then $\log G = \begin{pmatrix} 0 & \mathbb{R} & \mathbb{R} \\ 0 & 0 & \mathbb{R} \\ 0 & 0 & 0 \end{pmatrix}$ and*

$$\exp \begin{pmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & x & z + \frac{xy}{2} \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix},$$

*and hence*

$$\log \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & x & z - \frac{xy}{2} \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix}$$

*for any $x, y, z \in \mathbb{R}$. In particular we have*

$$\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}^t = \begin{pmatrix} 1 & tx & tz + \frac{t(t-1)}{2} xy \\ 0 & 1 & ty \\ 0 & 0 & 1 \end{pmatrix}$$

*for any* $x, y, z, t \in \mathbb{R}$, *and* $\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} = O(X)$ *if and only if* $x, y, z - \frac{xy}{2} = O(X)$.

From the identity $\log g^{-1} = -\log g$ we see that if $g = O(X)$ then $g^{-1} = O(X)$. Similarly, from the Baker–Campbell–Hausdorff formula (175), (176) we see that that $\log(gh)$ is a polynomial function of $\log g, \log h$ (with degree and coefficients $O(1)$), and hence if $g, h = O(X)$ then $gh = O(X^{O(1)})$.

We define $\mathrm{Poly}(\mathbb{R} \to G)$ to be the space of all maps $g \colon \mathbb{R} \to G$ of the form

$$g(t) \coloneqq \exp(\sum_{i=0}^{k} X_i t^i)$$

where $X_i \in \log G_i$ for $i = 0, \dots, k$. From the Baker–Campbell–Hausdorff formula (175), (176), (179) we see that $\mathrm{Poly}(\mathbb{R} \to G)$ is a group with respect to multiplication. For any $\delta > 0$, we define $\mathrm{Poly}(\delta\mathbb{Z} \to G)$ to be the set of all maps $g \colon \delta\mathbb{Z} \to G$ such that

$$\partial_{h_1} \dots \partial_{h_i} g(t) \in G_i$$

for all $i \geq 0$ and $h_1, \dots, h_i, t \in \delta\mathbb{Z}$, where $\partial_h g(t) \coloneqq g(t+h)g(t)^{-1}$. We similarly define $\mathrm{Poly}(\delta\mathbb{Z} \to \Gamma)$ by replacing $G_i$ with $\Gamma_i$ in the above definition; equivalently, $\mathrm{Poly}(\delta\mathbb{Z} \to \Gamma)$ consists of those elements of $\mathrm{Poly}(\delta\mathbb{Z} \to G)$ that take values in $\Gamma$. We refer to elements of $\mathrm{Poly}(\mathbb{R} \to G)$ and $\mathrm{Poly}(\delta\mathbb{Z} \to G)$ as *polynomial maps*. We have the following basic fact:

**Lemma 4.2.** *Let* $\delta > 0$. *Then every element* $\tilde{g}$ *of* $\mathrm{Poly}(\mathbb{R} \to G)$ *restricts to an element* $g$ *of* $\mathrm{Poly}(\delta\mathbb{Z} \to G)$*; conversely, every element* $g$ *of* $\mathrm{Poly}(\delta\mathbb{Z} \to G)$ *has a unique extension to an element* $\tilde{g}$ *of* $\mathrm{Poly}(\mathbb{R} \to G)$. *Finally,* $\mathrm{Poly}(\delta\mathbb{Z} \to \Gamma)$ *forms a group.*

*Proof.* See Appendix B. □

In view of this lemma we shall abuse notation by identifying $\mathrm{Poly}(\delta\mathbb{Z} \to G)$ with $\mathrm{Poly}(\mathbb{R} \to G)$, and viewing each of the $\mathrm{Poly}(\delta\mathbb{Z} \to \Gamma)$ as subgroups of $\mathrm{Poly}(\mathbb{R} \to G)$. We will refer to polynomial maps in $\mathrm{Poly}(\delta\mathbb{Z} \to \Gamma)$ as being $\frac{1}{\delta}$*-integral*.

Applying the inverse conjecture for the Gowers norms as in [35, §4], [17, §C] we see that Theorem 1.5 follows from (and is in fact equivalent to) the following claim:

**Theorem 4.3** (Non-pretentious multiplicative functions do not correlate with nilsequences on short intervals on average)**.** *Let $k \geq 0$ be a non-negative integer, and let $0 < \theta < 1$. Let $G/\Gamma$ be a degree $k$ filtered nilmanifold, and let $F\colon G/\Gamma \to \mathbb{C}$ be a Lipschitz function. Suppose that $f\colon \mathbb{N} \to \mathbb{C}$ is a multiplicative 1-bounded function, and suppose that $X \geq 1$, $X^\theta \leq H \leq X^{1-\theta}$, and $\eta > 0$ are such that*

$$\int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n \in [x, x+H]} f(n) \overline{F}(g(n)\Gamma) \right| dx \geq \eta H X.$$

*Then one has*

$$M(f; CX^{k+1}/H^{k+1}, Q) \ll_{k,\eta,\theta,F,G/\Gamma} 1 \tag{78}$$

*for some $C, Q \ll_{k,\eta,\theta,F,G/\Gamma} 1$.*

We note that in order to prove Theorem 1.5 it suffices to prove Theorem 4.3 with $F$ fixed since by Arzelà–Ascoli the family of Lipschitz functions $F$ on $G/\Gamma$ of bounded norm is precompact in the uniform topology, and moreover we can modify $F$ in the uniform norm by anything less than $\eta/10$, say, without significantly affecting the assumption of Theorem 4.3 (i.e changing $\geq \eta HX$ to $\geq \eta HX/2$, say). As a result we can restrict to a finite set of $F$'s and thus to a fixed $F$ by pigeonholing.

As in the previous section, at present it is only the values of $g$ on $\mathbb{Z}$ that are relevant, but once one begins exploiting the dilation structure of $\mathbb{R}$ it becomes convenient to view $g$ as a polynomial map on all of $\mathbb{R}$ and not just on $\mathbb{Z}$. As remarked in the introduction, in [19] a variant of this estimate was established in which the supremum in $g$ was placed outside the integral, and in which $H$ was allowed to grow in $X$ arbitrarily slowly rather than at a polynomial rate; see also [7] for an earlier partial result in this direction.

We prove Theorem 4.3 by induction on the dimension $\dim(G/\Gamma) = \dim(G)$ of the nilmanifold $G/\Gamma$ (keeping $k$ fixed). When $\dim(G/\Gamma) = 0$, the function $F(g(n)\Gamma)$ is constant, and the claim corresponds to $k = 0$ case of Theorem 1.3 which in turn essentially followed from the result in [25]. Hence we assume inductively that $\dim(G/\Gamma) \geq 1$, and that the claim has already been proven for all $G$ of smaller dimension. We now fix $k, \eta, \theta, F, G/\Gamma$, and allow implied constants to depend on these quantities. Thus we have

$$\int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n \in [x, x+H]} f(n) \overline{F}(g(n)\Gamma) \right| dx \gg HX, \tag{79}$$

and our objective is to show that

$$M(f; CX^{k+1}/H^{k+1}, Q) \ll 1$$

for some $C, Q = O(1)$. We may normalize $F$ to be bounded in magnitude by 1, so that the sequences $n \mapsto \overline{F}(g(n)\Gamma)$ are 1-bounded. As in the previous section, we also introduce a small parameter $\varepsilon > 0$ that can depend on $k, \eta, \theta, F, G/\Gamma$, and allow implied constants to also depend on $\varepsilon$ unless otherwise specified.

4.2. **Initial reductions.** We first make a minor but convenient reduction, namely that we restrict to the case when $f$ is completely multiplicative rather than merely multiplicative (cf. [34, Proposition 10]). If we let $f_1$ be the completely multiplicative function that equals $f$ at each prime $p$, then we can write $f$ as a Dirichlet convolution $f(n) = \sum_{d=1}^\infty 1_{d|n} f_1(\frac{n}{d}) h(d)$ for some multiplicative function $h$ with $h(p) = 0$ and $|h(p^j)| \leq 2$ for all $j \geq 2$ (in fact $h(p^j) = f(p^j) - f(p)f(p^{j-1})$). From (79) and the triangle inequality, we thus have

$$\sum_{d=1}^\infty |h(d)| \int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n \in [x,x+H]} 1_{d|n} f_1(\frac{n}{d}) \overline{F}(g(n)\Gamma) \right| dx \gg HX.$$

From Euler products we see that $\sum_{d=1}^\infty \frac{|h(d)|}{d^{2/3}} \ll 1$ (say), so by the pigeonhole principle there exists $d \geq 1$ such that

$$\int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n \in [x,x+H]} 1_{d|n} f_1(\frac{n}{d}) \overline{F}(g(n)\Gamma) \right| dx \gg d^{-2/3} HX.$$

The left-hand side can be trivially bounded by $O(d^{-1}HX)$, hence $d = O(1)$. Making the change of variables $n = dn'$ and $x = dx'$, we then have

$$\int_{X/d}^{2X/d} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n' \in [x',x'+H/d]} f_1(n') \overline{F}(g(dn')\Gamma) \right| dx \gg (H/d)(X/d).$$

Note that if $g$ lies in $\mathrm{Poly}(\mathbb{R} \to G)$ then the dilation $g(d\cdot)$ does also. Applying Theorem 4.3 for the completely multiplicative function $f_1$ (adjusting $\theta$ slightly to retain the hypothesis $X^\theta \leq H \leq X^{1-\theta}$), we conclude that

$$M(f_1; C(X/d)^{k+1}/(H/d)^{k+1}, Q) \ll 1$$

and the claim follows.

It remains to establish the claim for completely multiplicative $f$. Assume for contradiction that this claim is false. Then we can find a sequence $X = X_{\mathrm{n}} \geq 1$ of real numbers and a sequence $f = f_{\mathrm{n}}$ of 1-bounded completely multiplicative functions, such that (79) holds uniformly in n, but such that

$$M(f; CX^{k+1}/H^{k+1}, Q) \to \infty \tag{80}$$

as $\mathrm{n} \to \infty$ for any fixed $Q, C$, where $H = H_{\mathrm{n}}$ lies in the interval $[X_{\mathrm{n}}^\theta, X_{\mathrm{n}}^{1-\theta}]$. Among other things, this implies that $X \to \infty$ as $\mathrm{n} \to \infty$. We now restrict attention to n sufficiently large, so that $X$ can be made larger than any fixed constant. Henceforth we suppress the dependence of $X, H, f$ on n. We refer to a quantity as *fixed* if it is independent of n, and use the asymptotic notation $Y = o(Z)$ to denote the claim $|Y| \leq c(\mathrm{n})Z$ for some quantity $c(\mathrm{n})$ that may depend on fixed quantities, but goes to zero as $\mathrm{n} \to \infty$. From the induction

hypothesis, we conclude that

$$\int_X^{2X} \sup_{\tilde{g}\in \mathrm{Poly}(\mathbb{R}\to\tilde{G})} \left| \sum_{n\in[x,x+H]} f(n)\overline{\tilde{F}}(\tilde{g}(n)\tilde{\Gamma}) \right| dx = o(HX)$$

whenever $\tilde{G}/\tilde{\Gamma}$ is a fixed degree $k$ filtered nilmanifold of dimension strictly less than that of $G/\Gamma$, and $\tilde{F}: \tilde{G}/\tilde{\Gamma} \to \mathbb{C}$ is a fixed Lipschitz function. More generally, for any fixed Dirichlet character $\chi$, we see from (80) and enlarging $Q$ that

$$M(f\chi; CX^{k+1}/H^{k+1}, Q) \to \infty$$

for any fixed $C$, and hence

$$\int_X^{2X} \sup_{\tilde{g}\in \mathrm{Poly}(\mathbb{R}\to\tilde{G})} \left| \sum_{n\in[x,x+H]} f(n)\chi(n)\overline{\tilde{F}}(\tilde{g}(n)\tilde{\Gamma}) \right| dx = o(HX).$$

By multiplicative Fourier expansion we thus have

$$\int_X^{2X} \sup_{\tilde{g}\in \mathrm{Poly}(\mathbb{R}\to\tilde{G})} \left| \sum_{n\in[x,x+H]} f(n)1_{n=a \bmod q}\overline{\tilde{F}}(\tilde{g}(n)\tilde{\Gamma}) \right| dx = o(HX) \tag{81}$$

for any fixed natural number $q$ and any fixed $a$ coprime to $q$. Because $f$ is completely multiplicative, we also see that the same claim is true when $a$ shares a common factor $d$ with $q$, after rescaling $X, H, x, n$ by $d$ as before (and expressing sum over the shrunken interval $[x/d, x/d + H/d]$ as an average of sums over intervals of length $(X/H)^{\theta/2}$, plus negligible error).

Among other things, this allows us to eliminate "major arc" cases of (79). Define a *rational* subgroup of $G$ to be a closed subgroup $\tilde{G}$ of $G$ for which $\tilde{G} \cap \Gamma$ is cocompact in $\tilde{G}$.

**Proposition 4.4** (Major arc case)**.** *Assume that $f$ satisfies* (80)*. Let $\tilde{G}$ be a fixed connected rational subgroup of $G$, and suppose that $\tilde{G}$ is a proper subgroup in the sense that* $\dim(\tilde{G}) < \dim(G)$ *(or equivalently[a], $\tilde{G} \neq G$). We endow $\tilde{G}$ with the filtration $\tilde{G}_i := G_i \cap \tilde{G}$ induced from $G$. Let $q$ be a fixed natural number, and let $E$ be a fixed compact subset of $\tilde{G}$. Then*

$$\int_X^{2X} \sup_{\substack{\varepsilon\in E \\ \tilde{g}\in \mathrm{Poly}(\mathbb{R}\to\tilde{G}) \\ \gamma\in \mathrm{Poly}(q\mathbb{Z}\to\Gamma)}} \left| \sum_{n\in[x,x+H]} f(n)\overline{F}(\varepsilon\tilde{g}(n)\gamma(n)\Gamma) \right| dx = o(HX).$$

[a]This is because $\tilde{G} \neq G$ is equivalent to $\log \tilde{G}$ being a proper subspace of $\log G$.

*Proof.* Since $F$ is a Lipschitz function, and $E$ is compact it suffices to verify the Theorem for a single choice of $\varepsilon$. Next, we claim that the quotient space $\mathrm{Poly}(q\mathbb{Z} \to \Gamma)/\mathrm{Poly}(\mathbb{Z} \to \Gamma)$ is finite. Indeed, from Taylor expansion we see that if $\gamma \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma)$, then $\gamma(\mathbb{Z})$ takes values in the group $\Gamma'$ generated by the roots $\{\gamma^{1/q^k} : \gamma \in \Gamma\}$ of $\Gamma$. As noted at the end of Appendix B, $\Gamma$ has finite index in $\Gamma'$, so there are only finitely many possibilities for

the tuple $(\gamma(0),\dots,\gamma(k))$ modulo right multiplication by elements of $\Gamma^{k+1}$. As this tuple uniquely determines the polynomial map $\gamma$, we conclude that there are only finitely many possibilities for $\gamma$ modulo right multiplication by elements of $\mathrm{Poly}(\mathbb{Z}\to\Gamma)$, giving the claim.

Since the quantity $F(\tilde{g}(n)\gamma(n)\Gamma)$ is unaffected if one multiplies $\gamma$ on the right by an element of $\mathrm{Poly}(\mathbb{Z}\to\Gamma)$, we see that we may restrict $\gamma$ without loss of generality to a set of coset representatives of the finite quotient space $\mathrm{Poly}(q\mathbb{Z}\to\Gamma)/\mathrm{Poly}(\mathbb{Z}\to\Gamma)$. Thus, by the triangle inequality, it suffices to prove the claim for a single fixed choice of $\gamma$.

Fix $\gamma$. As $\Gamma$ has finite index in $\Gamma'$, there is a finite index subgroup $\Gamma_*$ of $\Gamma$ which is normal in $\Gamma'$ (for instance, one can take $\Gamma_*$ to be the kernel of the left-action of $\Gamma$ on the finite space $\Gamma/\Gamma'$).

The sequence $n\mapsto\gamma(n)\Gamma_*$ is then a polynomial map from $\mathbb{Z}$ to the finite group $\Gamma'/\Gamma_*$ (it is the composition of $\gamma\in\mathrm{Poly}(\mathbb{Z}\to\Gamma')$ with the quotient homomorphism $\pi$ from $\Gamma'$ to $\Gamma'/\Gamma_*$, where we equip $\Gamma'/\Gamma_*$ with the filtration $\pi(\Gamma'_i)$) and is hence periodic of some fixed period $Q$; this implies that $n\mapsto\gamma(n)\Gamma$ depends only on the residue class $n$ mod $Q$. By the triangle inequality, it now suffices to show that

$$\int_X^{2X}\sup_{\tilde{g}\in\mathrm{Poly}(\mathbb{R}\to\tilde{G})}\left|\sum_{n\in[x,x+H]}f(n)1_{n=a\bmod Q}\overline{F}(\tilde{g}(n)\gamma_0\Gamma)\right|\,dx=o(HX) \tag{82}$$

for any fixed $a$ and any fixed $\gamma_0\in\Gamma'$.

Since $\tilde{G}\cap\Gamma$ is cocompact in $\tilde{G}$, so is $\tilde{G}\cap\Gamma_*$. As $\Gamma_*$ is normalized by $\gamma_0$, this implies that $\gamma_0^{-1}\tilde{G}\gamma_0\cap\Gamma_*$ is cocompact in $\gamma_0^{-1}\tilde{G}\gamma_0$, so in particular the group $\gamma_0^{-1}\tilde{G}\gamma_0$ is rational. If we let $\tilde{F}:\gamma_0^{-1}\tilde{G}\gamma_0/(\gamma_0^{-1}\tilde{G}\gamma_0\cap\Gamma_*)\to\mathbb{C}$ be the function

$$\tilde{F}(\gamma_0^{-1}\tilde{g}\gamma_0\Gamma_*):=F(\tilde{g}\gamma_0\Gamma)$$

then $\tilde{F}$ is Lipschitz, and the left-hand side of (82) can be rewritten (after conjugating $\tilde{g}$ by $\gamma_0$) as

$$\int_X^{2X}\sup_{\tilde{g}\in\mathrm{Poly}(\mathbb{R}\to\gamma_0^{-1}\tilde{G}\gamma_0)}\left|\sum_{n\in[x,x+H]}f(n)1_{n=a\bmod Q}\overline{\tilde{F}}(\tilde{g}(n)\Gamma^*)\right|\,dx.$$

Here of course we give $\gamma_0^{-1}\tilde{G}\gamma_0$ the filtration $(\gamma_0^{-1}\tilde{G}\gamma_0)_i=\gamma_0^{-1}\tilde{G}_i\gamma_0$, and note that composition with the Lie group isomorphism $g\mapsto\gamma_0^{-1}g\gamma_0$ gives an isomorphism between $\mathrm{Poly}(\mathbb{R}\to\tilde{G})$ and $\mathrm{Poly}(\mathbb{R}\to\gamma_0^{-1}\tilde{G}\gamma_0)$. Since the dimension of the nilmanifold $\gamma_0^{-1}\tilde{G}\gamma_0/(\gamma_0^{-1}\tilde{G}\gamma_0\cap\Gamma_*)$ is strictly less than that of $G/\Gamma$, the claim now follows from (81). □

We now eliminate some components of $F$ that arise from lower dimensional nilmanifolds[15]. Suppose that there is a non-trivial normal rational connected closed subgroup $N$ of $G$. Then inside the Hilbert space $L^2(G/\Gamma)$ of square-integrable functions on $G/\Gamma$ (with respect to the Haar probability measure $\mu_{G/\Gamma}$) there is the closed subspace $L^2(G/\Gamma)^N$ of functions that are invariant with respect to the left-action of $N$; from normality this space is also preserved by the left-action of $G$.

[15]The need for dealing with these arises from the large sieve for nilsequences that we present as Lemma 4.11.

**Proposition 4.5** (Invariant case)**.** *Assume that $f$ satisfies* (80)*. If $N$ is a fixed non-trivial normal connected rational subgroup of $G$, and $F_N \in L^2(G/\Gamma)^N$ is a fixed Lipschitz continuous function, then*

$$\int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n \in [x, x+H]} f(n) \overline{F_N}(g(n)\Gamma) \right| \, dx = o(HX).$$

*Proof.* Let $\pi \colon G \to G/N$ be the quotient map from $G$ to $G/N$. As $N$ is normal, closed, and connected, $G/N$ is also a nilpotent connected, simply connected[16] Lie group, with a degree $k$ filtration $(G/N)_j := \pi(G_j)$. Because $\Gamma$ is discrete and cocompact in $G$ and $N \cap \Gamma$ is discrete and cocompact in $N$, we see that $\pi(\Gamma) \equiv \Gamma/(N \cap \Gamma)$ is discrete and cocompact in $\pi(G) = G/N$. Thus $\pi(G)/\pi(\Gamma)$ is a degree $k$ filtered nilmanifold, whose dimension $\dim(G) - \dim(N)$ is strictly less than that of $G/\Gamma$. Then we can write $F_N = \tilde{F} \circ \tilde{\pi}$ for some $\tilde{F} \colon \pi(G)/\pi(\Gamma) \to \mathbb{C}$ with $\tilde{\pi} \colon G/\Gamma \to \pi(G)/\pi(\Gamma)$ is the obvious projection; this function $F_N$ can be seen to also be Lipschitz continuous by working in local coordinates. Since $\pi \circ g \in \mathrm{Poly}(\mathbb{R} \to \pi(G))$ whenever $g \in \mathrm{Poly}(\mathbb{R} \to G)$, the claim now follows from (81). □

We let $F \mapsto \mathbf{E}(F|N)$ denote the orthogonal projection from $L^2(G/\Gamma)$ to $L^2(G/\Gamma)^N$; it can be described explicitly as

$$\mathbf{E}(F|N)(g\Gamma) = \int_{N/(N\cap\Gamma)} F(gx) \, d\mu_{N/(N\cap\Gamma)}(x)$$

for almost every $g \in G$, where we view $N/(N \cap \Gamma)$ as a subset of $G/\Gamma$ in the natural fashion. One can check (using the normality of $N$ and the uniqueness of the Haar probability measure $\mu_{N/(N\cap\Gamma)}$) that this gives a well-defined self-adjoint projection from $L^2(G/\Gamma)$ to $L^2(G/\Gamma)^N$, and so must indeed agree with the orthogonal projection to the latter space. It is also clear from this definition that if $F$ is Lipschitz continuous then so is $\mathbf{E}(F|N)$. In particular, from Proposition 4.5 one can remove the component $\mathbf{E}(F|N)$ from $F$ while making a negligible impact to (79). In our arguments we would like to perform this maneuver not for a single $N$, but for a large (but fixed) finite collection of such $N$. To do this we need the following observation:

**Lemma 4.6** (Composition of projections)**.** *Let $N_1, N_2$ be two normal connected rational subgroups of $G$. Then $N_1 N_2$ is also a normal connected rational subgroup, and*

$$\mathbf{E}(\mathbf{E}(F|N_1)|N_2) = \mathbf{E}(F|N_1 N_2)$$

*for all $F \in L^2(G/\Gamma)$. In particular (since $N_1 N_2 = N_2 N_1$), the projections $F \mapsto \mathbf{E}(F|N_1)$ and $F \mapsto \mathbf{E}(F|N_2)$ commute with each other.*

*Proof.* It is clear that $N_1 N_2$ is a normal connected subgroup of $G$. Because $N_1 \cap \Gamma$ is cocompact in $N_1$ and $N_2 \cap \Gamma$ is cocompact in $N_2$, and $N_1$ is normal, $(N_1 \cap \Gamma)(N_2 \cap \Gamma)$ is

[16] Indeed, from the Baker–Campbell–Hausdorff formula the space $G/N$ is homeomorphic to the vector space $\log G / \log N$.

cocompact[17] in $N_1N_2$, so $N_1N_2$ is rational. The function

$$F - \mathbf{E}(\mathbf{E}(F|N_1)|N_2) = (F - \mathbf{E}(F|N_1)) + (\mathbf{E}(F|N_1) - \mathbf{E}(\mathbf{E}(F|N_1)|N_2))$$

is orthogonal to $L^2(G/\Gamma)^{N_1} \cap L^2(G/\Gamma)^{N_2} = L^2(G/\Gamma)^{N_1N_2}$. The function

$$\mathbf{E}(\mathbf{E}(F|N_1)|N_2)$$

is clearly $N_2$-invariant, and can also be seen to be $N_1$-invariant using the normality of $N_2$. Thus $\mathbf{E}(\mathbf{E}(F|N_1)|N_2)$ lies in $L^2(G/\Gamma)^{N_1N_2}$, and is thus the orthogonal projection of $F$ to this space. The claim follows. □

Given any fixed finite collection $N_1, \dots, N_\ell$ of non-trivial normal connected rational subgroups $N_1, \dots, N_\ell$ of $G$, let $\Pi_{N_j} \colon L^2(G/\Gamma) \to (L^2(G/\Gamma)^{N_j})^\perp$ denote the complementary orthogonal projection to $L^2(G/\Gamma)^{N_j}$, thus

$$\Pi_{N_j} F \coloneqq F - \mathbf{E}(F|N_j).$$

From the above lemma, the $\Pi_{N_j}$ all commute with each other. Let $\Pi_{N_1,\dots,N_\ell} \coloneqq \Pi_{N_1} \dots \Pi_{N_\ell}$ denote the composition of these projections. Then one can express $F - \Pi_{N_1,\dots,N_\ell} F$ as a finite sum of Lipschitz functions, each of which lies in one of the $L^2(G/\Gamma)^{N_j}$. From Proposition 4.5 and the triangle inequality, we thus have

$$\int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n \in [x, x+H]} f(n) \overline{(F - \Pi_{N_1,\dots,N_\ell} F)}(g(n)\Gamma) \right| dx = o(HX) \tag{83}$$

as n $\to \infty$.

We can also use Theorem 1.3, proven in the previous section, to obtain

**Proposition 4.7.** *Let the hypotheses be as in Theorem 4.3, but assume that $f$ satisfies* (80)*. Then $G$ is not abelian.*

*Proof.* Suppose for contradiction that $G$ was abelian, then $G/\Gamma$ is a connected abelian Lie group and is therefore a torus (this follows for instance from Pontryagin duality). One can approximate $F$ uniformly by finite linear combinations of characters $e(\xi)$, where $\xi \colon G/\Gamma \to \mathbb{R}/\mathbb{Z}$ are continuous homomorphisms. By the triangle inequality (and passing to a subsequence of $X$ if necessary), we may thus find $\xi$ such that

$$\int_X^{2X} \sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \left| \sum_{n \in [x, x+H]} f(n) e(-\xi(g(n)\Gamma)) \right| dx \gg HX.$$

But from Taylor expansion we see that $t \mapsto \xi(g(t)\Gamma)$ is of the form $t \mapsto P(t) \bmod \mathbb{Z}$ for some $P \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$, and Theorem 1.3 supplies the required contradiction. □

[17] Indeed, we have $N_1 = K_1(N_1 \cap \Gamma)$ and $N_2 = K_2(N_2 \cap \Gamma)$ for some compact $K_1, K_2$, hence $N_1N_2 = N_1K_2(N_2 \cap \Gamma) = K_2N_1(N_2 \cap \Gamma) = K_2K_1(N_1 \cap \Gamma)(N_2 \cap \Gamma)$, giving the cocompactness.

4.3. **Studying the structure of local nilsequences.** Now we start following the arguments of the previous section. Define a *local nilsequence* to be a pair $\phi = (I, g)$, where $I$ is an interval and $g \in \mathrm{Poly}(\mathbb{R} \to G)$. We let $\Psi$ be the collection of all local nilsequences $\phi = (I, g)$, and $\Psi_I$ to be the collection of local nilsequences $(I, g)$ with a fixed choice of $I$. One should view $(I, g)$ as an abstraction of the function $t \mapsto F(g(t)\Gamma)$ on $I$. For any $\phi = (I, g) \in \Psi$ and $f \colon \mathbb{R} \to \mathbb{C}$, we define the correlation

$$\langle f, \phi \rangle := \frac{1}{|I|} \sum_{n \in I} f(n) \overline{F}(g(n)\Gamma),$$

where $F : G/\Gamma \to \mathbb{C}$ is understood to be a fixed Lipschitz function, with $G/\Gamma$ a fixed filtered nilmanifold. As before we have the dilation action

$$\lambda_*(I, g) := \left(\lambda I, g\left(\frac{1}{\lambda}\cdot\right)\right)$$

for any $(I, g) \in \Psi$ and $\lambda > 0$. The family $\Psi$ will play the role of the family $\Phi$ from the preceding section (which can be viewed as the special case when $G/\Gamma = \mathbb{R}/\mathbb{Z}$ with the filtration $G_j = \mathbb{R}$ for $j \le k$ and $G_j = \{0\}$ for $j > k$, and $F(x) := e(x)$). From (79) we have

$$\int_X^{2X} \sup_{\phi \in \Psi_{[x,x+H]}} |\langle f, \phi \rangle| \, dx \gg X$$

and hence by repeating the proof of [26, Lemma 2.1] as in the previous section, we can find a large $(X, H)$-family of intervals $\mathcal{I}$, such that for each $I \in \mathcal{I}$ one can find $\phi_I \in \Psi_I$ such that $|\langle f, \phi_I \rangle| \gg 1$.

For subsequent analysis we will need to somehow import the decay estimates in Proposition 4.4 and (83) into this context. This is achieved via the following application of Markov's inequality. Call a $(X, H)$-family of intervals *small* if it has cardinality $o(X/H)$.

**Proposition 4.8** (Local decay outside of exceptional set)**.** *Assume that $f$ satisfies* (80)*. Let $1 \le P \le X^{2\varepsilon}$, and let $\mathcal{I}'$ be a $(X/P, H/P)$-family of intervals. Then there exists a small exceptional subset $\mathcal{E}$ of $\mathcal{I}'$ such that the following properties hold uniformly for all $I \in \mathcal{I}' \backslash \mathcal{E}$:*

(i) *(Major arc estimate) If $\tilde{G}$ is a fixed connected closed proper rational subgroup of $G$, $E$ is a fixed compact subset of $\tilde{G}$, and $q$ is a fixed natural number, then*

$$\sup_{\substack{\varepsilon \in E \\ \tilde{g} \in \mathrm{Poly}(\mathbb{R} \to \tilde{G}) \\ \gamma \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma)}} \sup_{I' \subset 500I} \left| \sum_{n \in I'} f(n) \overline{F}(\varepsilon \tilde{g}(n) \gamma(n) \Gamma) \right| \, dx = o(H/P)$$

*where $I'$ ranges over all intervals contained in $500I$.*

(ii) *(Invariant estimate) For any fixed finite collection $N_1, \dots, N_\ell$ of non-trivial normal connected rational subgroups $N_1, \dots, N_\ell$ of $G$, one has*

$$\sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \sup_{I' \subset 500I} \left| \sum_{n \in [x, x+H]} f(n) \overline{(F - \Pi_{N_1, \dots, N_\ell} F)}(g(n)\Gamma) \right| = o(H/P).$$

*Proof.* We begin with (i). We will shortly establish that

$$\sum_{I\in\mathcal{I}'} \sup_{\substack{\varepsilon\in B_{\tilde{G}}(1,r)\\ \tilde{g}\in \mathrm{Poly}(\mathbb{R}\to\tilde{G})\\ \gamma\in\mathrm{Poly}(q\mathbb{Z}\to\Gamma)}} \sup_{I'\subset 500I} \left|\sum_{n\in I'} f(n)\overline{F}(\varepsilon\tilde{g}(n)\gamma(n)\Gamma)\right| = o(X/P) \tag{84}$$

for each fixed $\tilde{G}, q, r$, where $B_{\tilde{G}}(1,r)$ denotes the ball of radius $r$ centred at the identity in $\tilde{G}$, and the decay rate in the $o(X)$ right-hand side may depend on $\tilde{G}, q$. Assuming this bound for the moment, we can perform the following "diagonalization" argument. There are only countably many rational subgroups $\tilde{G}$ of $G$ (because $\log \tilde{G}$ can be described as a subspace of $\log G$ cut out by equations with rational coefficients). Enumerate the countable set of triples $(\tilde{G}, q, r)$ with $r$ a natural number as $(\tilde{G}_i, q_i, r_i)$. For each $i$, we see from (84), the triangle inequality, and Markov's inequality that we can find an exceptional set $\mathcal{E}_i \subset \mathcal{I}$ of cardinality at most $\frac{1}{i}\frac{X}{H}$, and a threshold $\mathrm{x}_i$, such that

$$\sum_{j\le i} \sup_{\substack{\varepsilon\in B_{\tilde{G}}(1,r_j)\\ \tilde{g}\in \mathrm{Poly}(\mathbb{R}\to\tilde{G}_j)\\ \gamma\in\mathrm{Poly}(q_j\mathbb{Z}\to\Gamma)}} \sup_{I'\subset 500I} \left|\sum_{n\in I'} f(n)\overline{F}(\varepsilon\tilde{g}(n)\gamma(n)\Gamma)\right| \le \frac{1}{i}H/P$$

whenever $x \ge \mathrm{x}_i$ and $I \in \mathcal{I}\backslash\mathcal{E}_i$. By increasing the $\mathrm{x}_i$ as necessary we may assume that $\mathrm{x}_{i+1} > \mathrm{x}_i$ for all $i$. If we now set $\mathcal{E} := \mathcal{E}_{i_*}$, where $i_*$ is the largest natural number for which $x \ge \mathrm{x}_{i_*}$, then $\mathcal{E}$ is well-defined for sufficiently large $x$, and the claim (i) follows (since any compact set $E$ is a subset of some ball $B_{\tilde{G}}(1,r)$).

It remains to verify (84). Set $H^* := (X/P)^{\theta/2}$. Then we can use the triangle inequality to write

$$\sum_{n\in I'} f(n)\overline{F}(\varepsilon\tilde{g}(n)\gamma(n)\Gamma) \ll \frac{1}{H^*}\int_{I'} \left|\sum_{n\in[x,x+H^*]} f(n)\overline{F}(\varepsilon\tilde{g}(n)\gamma(n)\Gamma)\right| dx + O(H^*)$$

and thus (since the intervals $500I$ in $\mathcal{I}'$ have bounded overlap in $[X/2P, 4X/P]$) we can bound the left-hand side of (84) by

$$\frac{1}{H^*}\int_{X/2P}^{4X/P} \sup_{\substack{\varepsilon\in E\\ \tilde{g}\in \mathrm{Poly}(\mathbb{R}\to\tilde{G})\\ \gamma\in\mathrm{Poly}(q\mathbb{Z}\to\Gamma)}} \left|\sum_{n\in[x,x+H^*]} f(n)\overline{F}(\varepsilon\tilde{g}(n)\gamma(n)\Gamma)\right| dx + o(X/P)$$

and the claim now follows from Proposition 4.4 (which is also valid if one replaces $X$ by a quantity comparable to $X/P$).

The claim (ii) is proven similarly (using (83) in place of Proposition 4.4), noting that there are only countably many rational closed connected subgroups $N$ of $G$ (since such groups are determined by their intersection $N \cap \Gamma$ with $\Gamma$, which is a finitely generated subgroup of the countable group $\Gamma$), and hence only countably many finite tuples $(N_1, \dots, N_\ell)$. □

Thus, for instance, using this proposition (with $P = 1$ and $\mathcal{I}' = \mathcal{I}$), we could now delete a small set of intervals from $\mathcal{I}$ and assume without loss of generality that the conclusions of this proposition hold for all $I \in \mathcal{I}$. As it turns out, however, it will be more useful to apply this proposition to a different family $\mathcal{I}'$ of intervals than $\mathcal{I}$, as we shall shortly see.

As in the preceding section, the next step is to relate the various $\phi_I$ to each other. We need a variant of Definition 3.1. If $I$ is an interval, we say that a polynomial map $\varepsilon \in \mathrm{Poly}(\mathbb{R} \to G)$ is *smooth* on $I$ if $\varepsilon(t) = O(1)$ for all $t \in I$. Taking logarithms and applying (28) to the polynomial map $\log \varepsilon \colon \mathbb{R} \to \log G$, this implies in particular that $|\frac{d^j}{dt^j} \log \varepsilon(t)| \ll |I|^{-j} \langle t \rangle_I^{O(1)}$ for all $j \geq 0$ and $t \in \mathbb{R}$. In particular $\varepsilon$ is also smooth on any interval $I' \sim I$ that is comparable to $I$. Also observe from the Baker–Campbell–Hausdorff formula (176) that if $\varepsilon_1, \varepsilon_2$ are both smooth on $I$, then so are $\varepsilon_1^{-1}$ and $\varepsilon_1 \varepsilon_2$ (with slightly different implied constants).

**Definition 4.9** (Comparability of nilsequences)**.** *Given two local nilsequences $\phi = (I, g), \phi' = (I', g') \in \Psi$ and a scaling factor $\delta > 0$, we define the relation*

$$\phi \sim_\delta \phi'$$

*to hold if $I \sim I'$, and we have the relation*

$$g(t) = \varepsilon(t) g'(t) \gamma(t)$$

*for all $t \in \mathbb{R}$, where $\varepsilon, \gamma \in \mathrm{Poly}(\mathbb{R} \to G)$ are polynomials obeying the following axioms:*

(i) *($\varepsilon$ smooth) $\varepsilon$ is smooth on $I$.*
(ii) *($\gamma$ is $\frac{1}{\delta}$-integral) $\gamma \in \mathrm{Poly}(\delta\mathbb{Z} \to \Gamma)$.*

We have the following analogue of Proposition 3.2:

**Proposition 4.10** (Basic properties of $\sim_\delta$)**.** *Let $\delta > 0$, and let $\phi, \phi', \phi'' \in \Psi$.*

(i) *(Equivalence relation) We have $\phi \sim_\delta \phi$, and if $\phi \sim_\delta \phi'$ then $\phi' \sim_\delta \phi$. Finally, if $\phi \sim_\delta \phi'$ and $\phi' \sim_\delta \phi''$ then $\phi \sim_\delta \phi''$, where we allow the implied constants in the latter relations to depend on the implied constants in the former relations.*
(ii) *(Dilation invariance) If $\phi \sim_\delta \phi'$ and $\lambda > 0$, then $\lambda_* \phi \sim_{\lambda\delta} \lambda_* \phi'$.*
(iv) *(Sparsification) If $\phi \sim_\delta \phi$, then $\phi \sim_{l\delta} \phi$ for any natural number $l$.*

*Proof.* These are immediate from Definition 4.9, together with the previous observation that a polynomial map that is smooth on $I$ is also smooth on $I'$ for any $I' \sim I$, and the observation that the product of two polynomial maps smooth on $I$ is also smooth on $I$. □

Now we have the analogue of Proposition 3.3:

**Proposition 4.11** (Large sieve)**.** *Let $I$ be an interval of some length $|I| \geq 1$, and let $f \colon \mathbb{Z} \to \mathbb{C}$ be a function bounded in magnitude by 1. Suppose that for each $i = 1, \dots, K$ there is an interval $I_i \sim I$ and a local nilsequence $\phi_i \in \Psi_{I_i}$ such that*

$$|\langle f, \phi_i \rangle| \gg 1. \tag{85}$$

*Then at least one of the following claims hold:*

(i) $K \ll 1$.
(ii) *There exist* $1 \leq i < j \leq K$ *such that* $\phi_i \sim_1 \phi_j$.
(iii) *(Correlation with major arc nilsequence) There is a connected closed proper rational subgroup* $\tilde{G}$ *of* $G$ *(drawn from a fixed finite collection of such subgroups) and a natural number* $q$ *(drawn from a fixed finite collection of such numbers) and a compact subset* $E$ *of* $\tilde{G}$ *(again drawn from a fixed finite collection) such that*

$$\sup_{\substack{\varepsilon \in E \\ \tilde{g} \in \mathrm{Poly}(\mathbb{R} \to \tilde{G}) \\ \gamma \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma)}} \sup_{I' \subset 500I} \left| \sum_{n \in I'} f(n) \overline{F}(\varepsilon \tilde{g}(n) \gamma(n) \Gamma) \right| \, dx \gg |I|.$$

(iv) *(Correlation with invariant nilsequence) There is a tuple* $(N_1, \dots, N_\ell)$ *of non-trivial normal connected rational subgroups* $N_1, \dots, N_\ell$ *of* $G$ *(drawn from a fixed finite collection of such subgroups) such that*

$$\sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \sup_{I' \subset 500I} \left| \sum_{n \in I'} f(n) \overline{(F - \Pi_{N_1, \dots, N_\ell} F)}(g(n)\Gamma) \right| \gg |I|.$$

As one might expect, we will be able to use Proposition 4.8 to eliminate the options (iii), (iv) from this proposition, after removing a small set of exceptional intervals.

*Proof.* We let $K_0$ be a sufficiently large fixed natural number (depending on $F, G/\Gamma$), to be chosen later, and write $\phi_j = (I_j, g_j)$. We can assume that $K \geq K_0$, since otherwise we are in case (i). We will initially just analyze the first $K_0$ local nilsequences $\phi_j$, and return to the remaining $\phi_j$ later.

Let $S \colon \mathbb{R}^+ \to \mathbb{R}^+$ be a sufficiently rapidly growing but fixed function depending on $F, G/\Gamma, K_0$ to be chosen later. The tuple $\vec{g} := (g_1, \dots, g_{K_0})$ can be viewed as a polynomial map in the product group $G^{K_0}$ (endowed with the obvious filtration $(G^{K_0})_j := G_j^{K_0}$). The subgroup $\Gamma^{K_0}$ is a discrete cocompact lattice in $G^{K_0}$. We may thus apply the quantitative factorization theorem in [16, Theorem 1.19], using the function $S$ in place of the function $M \mapsto M^A$, to obtain a factorization

$$\vec{g} = \vec{\varepsilon}\, \vec{g}'\, \vec{\gamma} \tag{86}$$

where $\vec{\varepsilon}, \vec{g}', \vec{\gamma} \in \mathrm{Poly}(\mathbb{R} \to G^{K_0})$ obey the following properties for some quantity $1 \leq M \ll_{K_0, S} 1$:

(i) (Smoothness) One has $|\log \vec{\varepsilon}(t)| \leq M$ for all $t \in I$ (and hence by (27), $|\frac{d^j}{dt^j} \log \vec{\varepsilon}(t)| \ll M|I|^{-j}$ for all $t \in I$ and $j \geq 0$).
(ii) (Equidistribution) $\vec{g}'$ takes values in some rational connected closed subgroup $\vec{G}'$ of $G^{K_0}$ which is $M$-rational (in the sense of [16, Definition 2.5], using some arbitrarily chosen Mal'cev basis on $G^{K_0}$), and is totally $1/S(M)$-equidistributed in the sense that

$$\left| \frac{1}{\#P} \sum_{n \in P} \vec{F}(\vec{g}'(n)\vec{\Gamma}') - \int_{\vec{G}'/\vec{\Gamma}'} \vec{F} \, d\mu_{\vec{G}'/\vec{\Gamma}'} \right| \leq \frac{1}{S(M)} \|\vec{F}\|_{\mathrm{Lip}}$$

for any Lipschitz function $\vec{F}\colon \vec{G}'/\vec{\Gamma}' \to \mathbb{C}$, and any arithmetic progression $P$ in $I \cap \mathbb{Z}$ of length at least $\frac{1}{S(M)}|I|$, where $\vec{\Gamma}' := \vec{G}' \cap \Gamma^{K_0}$ (and we endow $\vec{G}'/\vec{\Gamma}'$ with the metric induced from $(G/\Gamma)^{K_0}$).

(iii) (Rationality) $\vec{\gamma}(\mathbb{Z})\Gamma^{K_0}$ takes values in the set $\{\vec{\gamma}\Gamma^{K_0} : \vec{\gamma}^q \in \Gamma^{K_0}\}$ for some $1 \leq q \leq M$, and the sequence $n \mapsto \vec{\gamma}(n)\Gamma^{K_0}$ is periodic on $\mathbb{Z}$ with period at most $M$.

Arguing as in the proof of [16, Corollary 1.20], this gives a summation formula of the form

$$\sum_{n \in I'} \vec{F}(\vec{g}(n)\vec{\Gamma}) = \sum_{i=1}^{s} A_i \int_{x_i \vec{G}' y_i \Gamma^{K_0}/\Gamma^{K_0}} \vec{F}\, d\mu_{x_i \vec{G}' y_i \Gamma^{K_0}/\Gamma^{K_0}} + O_{M,K_0}\left(\frac{\|\vec{F}\|_{\mathrm{Lip}}}{S(M)^{1/2}}|I|\right) \quad (87)$$

for any interval $I' \subset I$, where the $A_i$ are positive quantities summing to $O(|I|)$, the $x_i$ are elements of $\vec{G}^{K_0}$ with $\log x_i = O_M(1)$, of magnitude $O_M(1)$, and the $y_i$ are elements of $\vec{G}^{K_0}$ with $y_i^q \in \Gamma$ (the argument proceeds by splitting $I'$ into $O_{M,K_0}(S(M)^{1/2})$ arithmetic progressions of diameter $O_{M,K_0}(\frac{|I|}{S(M)^{1/2}})$ and spacing equal to the period of $\vec{\gamma}\Gamma^{K_0}$). One could be more precise about the values of $A_i, x_i, y_i$ here, as well as provide upper bounds on the quantity $s$ but it will not be necessary for our argument to do so.

We write $\vec{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_{K_0})$, $\vec{g}' = (g_1', \dots, g_{K_0}')$, and $\vec{\gamma} = (\gamma_1, \dots, \gamma_{K_0})$. We now divide into several cases, depending on the nature of $\vec{G}'$. For each $1 \leq j \leq K_0$, let $\pi_j\colon G^{K_0} \to G$ be the projection to the $j^{\text{th}}$ factor of $G$. Then $\pi_j(\vec{G}')$ is a closed connected rational subgroup of $G$. Suppose that there exists $j$ for which $\pi_j$ is not surjective, so that $\pi_j(\vec{G}')$ is a proper subgroup of $G$. Because $\vec{G}'$ is $M$-rational, it belongs to a fixed finite family of subgroups of $G^{K_0}$, and hence $\pi_j(\vec{G}')$ also belongs to a fixed finite family of subgroups. From (86), (85) we have

$$\left|\sum_{n \in I_j} f(n)\overline{F}(\varepsilon_j(n) g_j'(n) \gamma_j(n)\Gamma)\right| \gg |I|,$$

so in particular $|I_j| \gg |I|$. Let $\sigma > 0$ be a small quantity to be chosen later. Then by covering $I_j$ by intervals $I_j'$ of length $\sigma|I|$ and using the pigeonhole principle, we can find one such interval $I_j'$ for which

$$\left|\sum_{n \in I_j'} f(n)\overline{F}(\varepsilon_j(n) g_j'(n) \gamma_j(n)\Gamma)\right| \gg \sigma|I|.$$

From property (i) and (26) we see that

$$F(\varepsilon_j(n) g_j'(n) \gamma_j(n)\Gamma) = F(\varepsilon_j(x_{I_j'}) g_j'(n) \gamma_j(n)\Gamma) + O_M(\sigma)$$

for $n \in I_j'$. For $\sigma$ sufficiently small depending on $M$ (but with $\sigma \asymp_M 1$) we can then neglect the error term and conclude that

$$\left|\sum_{n \in I_j'} f(n)\overline{F}(\varepsilon_j(x_{I_j'}) g_j'(n) \gamma_j(n)\Gamma)\right| \gg \sigma|I|,$$

and now we have conclusion (iii) of the proposition.

Henceforth we now assume that $\pi_j$ is surjective for all $1 \leq j \leq K_0$. For distinct $i, j \in \{1, \dots, K_0\}$, consider the group

$$N_{i,j} \coloneqq \{\pi_i(\vec{h}) : \vec{h} \in \vec{G}'; \pi_j(\vec{h}) = 1\}.$$

This is a normal connected closed rational subgroup of $G$; indeed one can check that

$$\log N_{i,j} = \{\tilde{\pi}_i(\vec{h}) : \vec{h} \in \log \vec{G}'; \tilde{\pi}_j(\vec{h}) = 0\}$$

where $\tilde{\pi}_i \colon \log G^{K_0} \to \log G$ are the coordinate projections, and then the claims are easily verified. Let $\Pi$ be the projection on $L^2(G/\Gamma)$ formed by composing together the $\Pi_{N_{i,j}}$ for all distinct $i, j \in \{1, \dots, K_0\}$ for which $N_{i,j}$ is not trivial. Note that because $\vec{G}'$ belongs to a fixed finite family of subgroups of $G^{K_0}$, $N_{i,j}$ belongs to a fixed finite family of subgroups of $G$ (depending on $M, K_0$). Thus, if

$$\left| \sum_{n \in I_i} f(n) \overline{(F - \Pi F)(g_i(n)\Gamma)} \right| \gg |I|$$

for some $i = 1, \dots, K_0$, then we have conclusion (iv) of the proposition. Otherwise, by (85) and the triangle inequality, we may assume that

$$\left| \sum_{n \in I_i} f(n) \overline{\Pi F(g_i(n)\Gamma)} \right| \gg |I|$$

for all $i = 1, \dots, K_0$. We may now apply Cauchy–Schwarz as in the proof of Proposition 3.3 and conclude that

$$\sum_{i=1}^{K_0} \sum_{j=1}^{K_0} \left| \sum_{n \in I_i \cap I_j} \Pi F(g_i(n)\Gamma) \overline{\Pi F(g_j(n)\Gamma)} \right| \gg K_0^2 |I|. \tag{88}$$

We now dispose of the diagonal terms by claiming that

$$\sum_{n \in I_i} |\Pi F(g_i(n)\Gamma)|^2 \ll |I| \tag{89}$$

for each $i$. A key point here is that the implied constant does not depend on $K_0, M$. Here we have a technical difficulty because $\Pi F$ is not well controlled in $L^\infty(G/\Gamma)$ norm (one has a $L^\infty$ bound of $O_{K_0,M}(1)$ rather than $O(1)$); however it is still bounded in $L^2(G/\Gamma)$ by 1 since $\Pi$ is an orthogonal projection, and it also has a Lipschitz norm of $O_{K_0,M}(1)$. Nevertheless, by applying the formula (87), one can write the left-hand side of (89) as

$$\sum_{j=1}^{s} A_j \int_{x_j \vec{G}' y_j \Gamma^{K_0}/\Gamma^{K_0}} |\Pi F \circ \pi_i|^2 \, d\mu_{x_j \vec{G}' y_j \Gamma^{K_0}/\Gamma^{K_0}} + O_{M,K_0}\left( \frac{1}{S(M)^{1/2}} |I| \right)$$

for some $A_j, x_j, y_j$ (which can depend on $i$) with the properties listed after (87). As $\pi_i$ is surjective, it pushes forward Haar measure to Haar measure by the uniqueness properties

of Haar measure, so the above estimate simplifies to

$$\sum_{j=1}^{s} A_j \int_{G/\Gamma} |\Pi F|^2 \, d\mu_{G/\Gamma} + O_{M,K_0}\left(\frac{1}{S(M)^{1/2}}|I|\right).$$

Since the $L^2$ norm of $\Pi F$ is bounded by 1, and $\sum_{j=1}^{s} A_j = O(|I|)$, we obtain the claim (89) if $S$ is chosen to be sufficiently rapidly growing.

Using (89) to remove the diagonal terms from (88), we conclude (for $K_0$ large enough) that there exist distinct $i, j \in \{1, \dots, K_0\}$ such that

$$\left| \sum_{n \in I_i \cap I_j} \Pi F(g_i(n)\Gamma)\overline{\Pi F(g_j(n)\Gamma)} \right| \gg |I|.$$

Applying (87), we can bound the left-hand side by

$$\sum_{l=1}^{s} A_l \int_{x_l \vec{G}' y_l \Gamma^{K_0}/\Gamma^{K_0}} (\Pi F \circ \pi_i)\overline{(\Pi F \circ \pi_j)} \, d\mu_{x_l \vec{G}' y_l \Gamma^{K_0}/\Gamma^{K_0}} + O_{M,K_0}\left(\frac{|I|}{S(M)}\right)$$

for some $A_l, x_l, y_l$ obeying the properties after (87); in particular, for $S$ sufficiently rapidly growing, there exists $l$ such that

$$\int_{x_l \vec{G}' y_l \Gamma^{K_0}/\Gamma^{K_0}} (\Pi F \circ \pi_i)\overline{(\Pi F \circ \pi_j)} \, d\mu_{x_l \vec{G}' y_l \Gamma^{K_0}/\Gamma^{K_0}} \neq 0.$$

We can project the nilmanifold $x_l \vec{G}' y_l \Gamma^{K_0}/\Gamma^{K_0}$ down to $(G/\Gamma)^2$ using the projection map $(\pi_i, \pi_j)$ to the $i, j$ coordinates. The image of this nilmanifold is then invariant under the left action of the normal group $N_{i,j} \times \{1\}$. If $N_{i,j}$ is non-trivial, then $\Pi F$ has mean zero along all orbits of $N_{i,j}$ by construction, and the above integral will vanish. Thus $N_{i,j}$ must be trivial. A similar argument shows that $N_{j,i}$ is trivial.

Now consider the subgroup

$$G_{i,j} \coloneqq \{(\pi_i(\vec{g}), \pi_j(\vec{g})) : \vec{g} \in \vec{G}'\}$$

of $G^2$; this is a closed connected rational subgroup of $G^2$. By the preceding discussion, the projections $\pi_1 \colon G_{i,j} \to G$, $\pi_2 \colon G_{i,j} \to G$ are both surjective and injective. By the Goursat lemma, $G_{i,j}$ then takes the form

$$G_{i,j} = \{(g, \phi_{i,j}(g)) : g \in G\} \tag{90}$$

for some group isomorphism $\phi_{i,j} \colon G \to G$. As there are $O_{K_0,M}(1) = O_{K_0,S}(1)$ possible choices for $\vec{G}'$, there are $O_{K_0,S}(1)$ choices of $\phi_{i,j}$. As $G_{i,j}$ is rational, the map $\phi_{i,j}$ (when expressed in the standard basis for $\log G$) is a polynomial map with rational coefficients, hence (by Baker–Campbell–Hausdorff) $\phi_{i,j}(\Gamma)$ is covered by finitely many translates of $\Gamma$, and conversely; thus $\phi_{i,j}(\Gamma)$ must be commensurate with $\Gamma$, in the sense that $\phi_{i,j}(\Gamma) \cap \Gamma$ has finite index in $\phi_{i,j}(\Gamma)$ or $\Gamma$. Since $\vec{g}'$ takes values in $\vec{G}'$, we see from (90) that

$$g'_j = \phi_{i,j}(g'_i)$$

and thus by (86)

$$g_j = \varepsilon_{i,j}\phi_{i,j}(g_i)\gamma_{i,j} \tag{91}$$

where

$$\varepsilon_{i,j} \coloneqq \varepsilon_j \phi_{ij}(\varepsilon_i)^{-1}$$

and

$$\gamma_{i,j} \coloneqq \phi_{i,j}(\gamma_i)^{-1}\gamma_j.$$

From the smoothness properties of $\varepsilon_i, \varepsilon_j$ we see that

$$\log \varepsilon_{i,j}(t) \ll_{K_0,M} 1$$

for $t \in I$, and hence (since $M = O_{K_0,S}(1)$)

$$\log \varepsilon_{i,j}(t) \ll_{K_0,S} 1. \tag{92}$$

By rationality, the functions $\phi_{i,j}(\gamma_i)\Gamma$, $\gamma_j\Gamma$ each map $\mathbb{Z}$ to $\{\gamma\Gamma : \gamma^q \in \Gamma\}$ for some $q = O_{K_0,M}(1) = O_{K_0,S}(1)$ and are also periodic with period $O_{K_0,S}(1)$, which (as discussed at the end of Appendix B) implies that

$$\gamma_{i,j}(\mathbb{Z})\Gamma \subset \{\gamma\Gamma : \gamma^q \in \Gamma\} \tag{93}$$

for some $q = O_{K_0,S}(1)$, and $\gamma_{i,j}$ is periodic with period $O_{K_0,S}(1)$.

Call a pair $(i,j)$ of distinct elements of $\{1,\dots,K\}$ *good* if there is an identity of the form (91), where $\phi_{i,j}$ ranges over one of $O_{K_0,S}(1)$ isomorphisms of $G$, $\varepsilon_{i,j}$ obeys (92) on $I$, and $\gamma_{i,j}$ obeys (93) for some $q = O_{K_0,S}(1)$ and is periodic with period $O_{K_0,S}(1)$. By relabeling, we have shown that every $K_0$-element subset of $\{1,\dots,K\}$ contains a good pair $(i,j)$. Averaging over all such subsets, we conclude that there are $\gg_{K_0} K^2$ good pairs. In particular, by the pigeonhole principle, there exists $i \in \{1,\dots,K\}$ such that $(i,j)$ is good for $\gg_{K_0} K$ values of $j$. Note that there are only $O_{K_0,S}(1)$ possible values of $\phi_{i,j}$ and of the coset $\gamma_{i,j}\mathrm{Poly}(\mathbb{Z} \to \Gamma)$. Thus, if $K$ is large enough, we see from the pigeonhole principle that there exist distinct $j, j'$ such that $\phi_{i,j} = \phi_{i,j'}$ and $\gamma_{i,j}\mathrm{Poly}(\mathbb{Z} \to \Gamma) = \gamma_{i,j'}\mathrm{Poly}(\mathbb{Z} \to \Gamma)$. From (91) we conclude that

$$g_j = \varepsilon_{i,j}\varepsilon_{i,j'}^{-1} g_{j'} \gamma_{i,j'}^{-1}\gamma_{i,j}$$

and hence by Definition 4.9

$$\phi_j \sim_1 \phi_{j'}$$

and the claim follows. □

Using this proposition as in the previous section (but now also using Proposition 4.8 to eliminate the unwanted options (iii), (iv) from Proposition 4.11), we obtain the following variant of Proposition 3.4.

**Proposition 4.12** (Scaling down). *Let $2 \leq P \leq Q \leq H \leq X$ and let $f\colon \mathbb{N} \to \mathbb{C}$ be a 1-bounded completely multiplicative function. Assume that $P, \frac{\log Q}{\log P}$ are sufficiently large (depending on the parameters $k, \theta, \eta$). Suppose there exists a large $(X,H)$-family $\mathcal{I}$ and a local nilsequence $\phi_I \in \Psi_I$ associated to each interval $I \in \mathcal{I}$ such that*

$$|\langle f, \phi_I \rangle| \gg 1$$

*holds for all $I \in \mathcal{I}$. Then there exist $P' \in [P, Q/2]$, a large $(\frac{X}{P'}, \frac{H}{P'})$-family $\mathcal{I}'$, and a local nilsequence $\phi'_{I'} \in \Psi_{I'}$ associated to each $I' \in \mathcal{I}'$, such that*

$$|\langle f, \phi'_{I'} \rangle| \gg 1$$

*for all $I' \in \mathcal{I}'$. Furthermore, for each $I' \in \mathcal{I}'$, one can find $\gg \pi_0(P')$ pairs $(I, p')$, where $I \in \mathcal{I}$ and $p'$ is a prime in $[P', 2P']$, such that the rescaled interval $\frac{1}{p'}I$ lies within $3\frac{H}{P'}$ of $I'$, and such that*

$$(\frac{1}{p'})_* \phi_I \sim_1 \phi'_{I'}. \tag{94}$$

*Proof.* Repeat the proof of [26, Proposition 3.1] down to the paragraph after (36). Then one can find $P' \in [P, Q/2]$, and a collection $\mathcal{I}_2$ of intervals in $[0, 10X/P']$ that are separated by distance at least $2H/P'$, with the property that for $\gg \frac{X}{H}\pi_0(P')$ pairs $(I, p')$ with $I \in \mathcal{I}$ and $p'$ a prime in $[P', 2P']$, $\frac{1}{p'}I$ lies within $3\frac{H}{P'}$ of some interval $I' \in \mathcal{I}_2$, and furthermore

$$|\langle f, (\frac{1}{p'})_* \phi_I \rangle| \gg 1.$$

Note that each $I'$ is associated to at most $O(\pi_0(P'))$ such pairs. In particular we have the freedom to remove a small set of intervals from $\mathcal{I}'$ without significantly diminishing the set of pairs $(I, p')$ in the above claims.

From Proposition 4.11 and the greedy algorithm, we see that for each $I' \in \mathcal{I}_2$, at least one of the following claims hold:

(i) There is a family $\phi_{I',1}, \dots, \phi_{I',K_{I'}} \in \Psi$ of functions with $K_{I'} = O(1)$ such that whenever $(I, p')$ is one of the above pairs with $\frac{1}{p'}I$ within $3\frac{H}{P'}$ of $I'$, one has

$$(\frac{1}{p'})_* \phi_I \sim_1 \phi_{I',K_i}$$

for some $i = 1, \dots, K_{I'}$.

(ii) There is a connected closed proper rational subgroup $\tilde{G}$ of $G$ (drawn from a fixed finite collection of such subgroups) and a natural number $q$ (drawn from a fixed finite collection of such numbers) and a compact subset $E$ of $\tilde{G}$ (again drawn from a fixed finite collection) such that

$$\sup_{\substack{\varepsilon \in E \\ \tilde{g} \in \mathrm{Poly}(\mathbb{R} \to \tilde{G}) \\ \gamma \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma)}} \sup_{J \subset 500I'} \left| \sum_{n \in J} f(n) \overline{F}(\varepsilon \tilde{g}(n) \gamma(n) \Gamma) \right| \, dx \gg \frac{H}{P'}.$$

(iii) There is a tuple $(N_1, \dots, N_\ell)$ of non-trivial normal connected rational subgroups $N_1, \dots, N_\ell$ of $G$ (drawn from a fixed finite collection of such subgroups) such that

$$\sup_{g \in \mathrm{Poly}(\mathbb{R} \to G)} \sup_{J \subset 500I'} \left| \sum_{n \in J} f(n) \overline{(F - \Pi_{N_1,\dots,N_\ell} F)}(g(n)\Gamma) \right| \gg \frac{H}{P'}.$$

By Proposition 4.8, we can eliminate the options (ii), (iii) by removing a small set of intervals from $\mathcal{I}_2$, leaving only option (i). One can now continue the proof of [26, Proposition 3.1] (making only the obvious changes) to conclude the proposition. □

We continue to follow the line of argument from the previous section. We will need an analogue of Lemma 2.2 for nilsequences:

**Lemma 4.13** (Bezout identity)**.** *Let $a, b$ be coprime natural numbers, and let $\lambda > 0$. Then*

$$\mathrm{Poly}(\frac{\lambda}{a}\mathbb{Z} \to \Gamma) \cdot \mathrm{Poly}(\frac{\lambda}{b}\mathbb{Z} \to \Gamma) = \mathrm{Poly}(\lambda\mathbb{Z} \to \Gamma)$$

*and*

$$\mathrm{Poly}(\frac{\lambda}{a}\mathbb{Z} \to \Gamma) \cap \mathrm{Poly}(\frac{\lambda}{b}\mathbb{Z} \to \Gamma) = \mathrm{Poly}(\frac{\lambda}{ab}\mathbb{Z} \to \Gamma).$$

*Proof.* See Appendix C. □

As a consequence, we can now establish the analogue of Proposition 3.5 for nilsequences (though with a slightly weaker version of part (ii)):

**Proposition 4.14** (Chinese remainder theorem)**.** *Let $I$ be an interval of some length $|I| \geq 1$, and let $\mathcal{P}$ be a finite collection of primes.*

(i) *Suppose that $\phi \in \Psi_I$, and for each $p \in \mathcal{P}$ there exists $\phi_p \in \Psi$ such that*

$$\phi_p \sim_1 \phi.$$

*Then there exists $\phi' \in \Psi_I$ such that*

$$\phi_p \sim_{\frac{1}{p}} \phi'$$

*for all $p \in \mathcal{P}$, and furthermore $\langle f, \phi \rangle = \langle f, \phi' \rangle$ for all $f \colon \mathbb{Z} \to \mathbb{C}$.*

(ii) *Suppose that $\phi \in \Psi_I$ and $\phi' \in \Psi$ are such that*

$$\phi \sim_{\frac{1}{p}} \phi'$$

*for all $p \in \mathcal{P}$, and suppose $|I|$ is sufficiently large (depending on the implied constants in the $\sim_{\frac{1}{p}}$ notation). Then there is a subset $\mathcal{P}'$ of $\mathcal{P}$ with $\#\mathcal{P}' \gg \#\mathcal{P}$ such that*

$$\phi \sim_{\frac{1}{\prod \mathcal{P}'}} \phi'.$$

*Proof.* See Appendix C □

One can now conclude an analog of Proposition 3.6:

**Proposition 4.15** (Building a family of related local nilsequences)**.** *Let the hypotheses be as in Theorem 4.3. Let $\varepsilon > 0$ be sufficiently small depending on $k, \theta, \eta$, and suppose that $X$ is sufficiently large depending on $\theta, \eta, \varepsilon, k$. Then there exist $P', P'' \in [X^{\varepsilon^2/2}, X^{\varepsilon}]$, a large $(\frac{X}{P'P''}, \frac{H}{P'P''})$-family $\mathcal{I}''$, and a local nilsequence $\phi''_{I''} \in \Psi_{I''}$ for each $I'' \in \mathcal{I}''$ such that* (36) *holds for all $I'' \in \mathcal{I}''$; also, each $I'' \in \mathcal{I}''$ obeys the conclusions (i), (ii) of Proposition 4.8 (with $P = P'P''$). Furthermore, there exist a collection $\mathcal{Q}$ of $\gg \pi_0(P')^2 \frac{X}{H}$ quadruples $(I''_1, I''_2, p'_1, p'_2)$ with $I''_1, I''_2$ distinct intervals in $\mathcal{I}''$ and $p'_1, p'_2$ distinct primes in $[P', 2P']$, such that $I''_1$ lies within $50\frac{H}{P'P''}$ of $\frac{p'_2}{p'_1} I''_2$ (so in particular $I''_1 \sim \frac{p'_2}{p'_1} I''_2$), and such that* (37) *holds for a large set of primes $p''$ in $[P''/2, P'']$.*

*Proof.* One repeats the proof of Proposition 3.6 verbatim, using Propositions 4.10, 4.12, 4.14 in place of Propositions 3.2, 3.4, 3.5. To ensure the conclusions (i), (ii) of Proposition 4.8, one simply removes the exceptional set produced by that proposition, which has only a negligible impact on the cardinality of $\mathcal{Q}$. □

For the rest of this section we introduce the quantities

$$N := \#\mathcal{I}'' \asymp \frac{X}{H}$$

and

$$d := \pi_0(P')^2$$

as in the previous section. We now establish an analog of Proposition 3.7:

**Proposition 4.16** (Local structure of $\phi''$)**.** *Let the hypotheses be as in Theorem 4.3, and let $\varepsilon, X, P', P'', \mathcal{I}'', \phi''_{I''}$ be as in Proposition 4.15. Let $\ell_1, \ell_2$ be bounded even integers obeying* (44)*. We allow implied constants to depend on $\varepsilon, \ell_1, \ell_2$. Then, for a subset $\mathcal{Q}'$ of the quadruples $e = (I''_1, I''_2, p'_1, p'_2)$ in $\mathcal{Q}$ of cardinality $\gg dN$, one can find a collection $\mathcal{A}_e$ of quadruples $\vec{a} = (a_1, a_2, b_1, b_2)$ of natural numbers of cardinality $\asymp d^{\ell_1+\ell_2}/N^2$, and a large collection $\mathcal{P}_{e,\vec{a}}$ of primes in $[P''/2, P'']$ associated to each $\vec{a} \in \mathcal{A}_e$, obeying the properties (i), (ii), (iii) of Proposition 3.7. In particular, the implied constants in* (45) *do not depend on $\ell_1, \ell_2$, and the implied constants in* (48) *may depend on $\ell_i$ but do not depend on $\ell_{3-i}$.*

*Proof.* One repeats the proof of Proposition 3.7, using Propositions 4.10, 4.15, 4.14 in place of Propositions 3.2, 3.6, 3.5. Note that Proposition 4.14(ii) will force us to refine the set of primes $\mathcal{P}_{I''_1, I''_2}$ somewhat, but it will still remain large. □

In the previous section, the values of $\ell_1, \ell_2$ were not of particular significance. In this section it will be convenient to choose $\ell_1$ to be significantly larger than $\ell_2$, because we will need to work with many quadruples simultaneously.

4.4. **Solving the approximate dilation invariance.** The next step is to solve the approximate dilation invariance equation (48) for a given quadruple $e$. In the previous section, we were able to obtain a satisfactory description of the solutions just by using a single choice of $\vec{a} = (a_1, b_1, a_2, b_2) \in \mathcal{A}_e$; see Proposition 3.8. Here, however, the situation will be more complicated, because for each $\vec{a}$ there can be some unwanted "exotic" solutions $\phi''_{I''_i}$ to (48)

that do not pretend to behave like a character $t^{iT}$, and which therefore cannot be treated using the results from [23], [25]. For instance, consider the situation in which

$$\phi''_{I''_1} = (I''_1, t \mapsto \gamma^{P(t)}) \tag{95}$$

for some $\gamma = \gamma_{a_1,b_1} \in \Gamma$ of polynomial size $\gamma = X^{O(1)}$ and some polynomial $P(t)$ which is a partial Taylor expansion of the analytic function $t \mapsto \frac{\log t}{\log(a_1/b_1)}$ around the midpoint $x_{I''_1}$ of $I''_1$. If the filtration $G_i$ is defined suitably, $t \mapsto \gamma^{P(t)}$ will be a polynomial map. On the other hand, since

$$\gamma^{\frac{\log(a_1 t)}{\log(a_1/b_1)}} = \gamma^{\frac{\log(b_1 t)}{\log(a_1/b_1)}}\gamma$$

one can verify that the approximate dilation invariance (48) will be obeyed for $i = 1$ if $P(t)$ is a sufficiently long partial Taylor expansion of $t \mapsto \frac{\log t}{\log(a_1/b_1)}$. If $\gamma$ is a central element of $G$, the local nilsequence (95) will then "pretend" to be like $t^{iT}$ for some $T$ depending on $\gamma$ and $\log(a_1/b_1)$, but if $\gamma$ is not central then one would not expect this to be the case in general. As a consequence, merely having (48) for a single tuple $\vec{a}$ will be insufficient for our arguments. However, as we shall see, if have the approximate dilation invariance (48) holds for a very "dense" collection of ratios $a_1/b_1$, then one cannot have a representation such as (95) for all of these $a_1/b_1$ simultaneously unless the bases $\gamma$ involved are essentially central, or if $\phi''_{I''_1}$ can be modeled by a lower dimensional nilsequence. Actually the first case is contained in the second thanks to Proposition 4.7, so we will be able to proceed via Proposition 4.4.

We now begin the formal arguments. The first step is to decouple the "continuous" (or "Archimedean") aspects of the equation (48) (associated to the smooth polynomial maps $\varepsilon$ in Definition 4.9 and the dilation structure in (48)) from the "rational" (or "non-Archimedean") aspects (associated to the rational maps $\gamma$ in Definition 4.9). It will be possible to do this thanks to the exponentially large size of the modulus $\prod \mathcal{P}_{e,\vec{a}}$ occurring in (48), which enable a sort of "Lefschetz principle" to pass to the continuous setting. To describe this more precisely we need some more notation. As in the previous section, a quantity $a$ (which could be a number or an element of $G$ or $\log G$) is said to be of *polynomial size* if $a = O(X^{O(1)})$. We similarly say that a map $g \in \mathrm{Poly}(\mathbb{R} \to G)$ is of *polynomial size* if the coefficients $g_0, \dots, g_k$ of the Taylor expansion

$$g(t) = g_0 g_1^{\binom{t}{1}} \dots g_k^{\binom{t}{k}}$$

of $g$ around the origin are all of polynomial size. Observe from many applications of the Baker–Campbell–Hausdorff formula (Appendix B) that a polynomial map $g \in \mathrm{Poly}(\mathbb{R} \to G)$ is of polynomial size if and only if the polynomial map $\log g \colon \mathbb{R} \to \log G$ has all coefficients of its Taylor expansion around the origin of polynomial size. In particular (from a further application of Baker–Campbell–Hausdorff) if $g, h \in \mathrm{Poly}(\mathbb{R} \to G)$ are of polynomial size then so are $g^{-1}$ and $gh$ (though with different implied constants in the $O()$ notation); also one has $g(t)$ of polynomial size whenever $g, t$ are. Next, for any modulus $Q > 0$, we say that a map $\gamma \in \mathrm{Poly}(\mathbb{R} \to G)$ is *$Q$-rational* if $\gamma \in \mathrm{Poly}(\frac{q}{Q}\mathbb{Z} \to \Gamma)$ for some natural number $q$ of polynomial size. From Lemma 4.13 (and a rescaling by $q$) we see that if $\gamma, \gamma' \in \mathrm{Poly}(\mathbb{R} \to G)$

are $Q$-rational, then so are $\gamma^{-1}$ and $\gamma\gamma'$, again with different implied constants in the $O()$ notation. The key fact that allows us to decouple is the following "transversality" between the collection of maps of polynomial size and the collection of maps that are extremely rational.

**Lemma 4.17** (Transversality)**.** *Let $\mathcal{P}$ be a large set of primes in $[P''/2, P'']$. Suppose that $g \in \mathrm{Poly}(\mathbb{R} \to G)$ is both of polynomial size and $Q$-rational, where $Q = \prod \mathcal{P}$. Then $g$ is equal to a constant $g(t) = \gamma$ for some $\gamma \in \Gamma$ of polynomial size.*

*Proof.* The group element $g(0)$ lies in $\Gamma$ and is of polynomial size. By dividing this out we may assume $g(0) = 1$. We first prove the claim for abelian groups $G$. Since $g$ in a map in $\mathrm{Poly}(\frac{q}{Q}\mathbb{Z} \to \Gamma)$ we have

$$g(\frac{q}{Q}t) = \sum_{i=0}^{k} a_i \binom{t}{i}$$

with $a_i \in \mathbb{Z}$. So that

$$g(t) = \sum_{i=0}^{k} a_i \binom{\frac{Q}{q}t}{i}.$$

Since $g$ is polynomial size we conclude that $[\frac{Q}{q}]^k \frac{1}{k!} a_k$ must be of polynomial size; as $q$ is also of polynomial size, we therefore have

$$a_k = O(X^{O(1)} Q^{-k}).$$

On the other hand, as $\mathcal{P}$ is a large set of primes in $[P''/2, P'']$, we have from (25) that

$$Q \gg \exp(X^{\varepsilon^2/3})$$

(say). Since $a_k \in \mathbb{Z}$ we conclude that $a_k = 0$. Proceeding by induction we obtain that $a_i = 0$ for all $i > 0$.

Now that if $g \in \mathrm{Poly}(\mathbb{R} \to G)$ is of polynomial size then $\bar{g} = g[G, G] \in \mathrm{Poly}(\mathbb{R} \to G/[G, G])$ is also of polynomial size, since if $g(t) = g_0 g_1^{\binom{t}{1}} \dots g_k^{\binom{t}{k}}$ then $\bar{g}(t) = \bar{g}_0 \bar{g}_1^{\binom{t}{1}} \dots \bar{g}_k^{\binom{t}{k}}$, where $\bar{g}_i = g_i[G, G]$. Consider $\bar{g}$ now as a polynomial map in $\mathrm{Poly}(\frac{q}{Q}\mathbb{Z} \to \Gamma/[\Gamma, \Gamma])$, then by Lemma B.2 we have the Taylor expansion

$$\bar{g}\left(\frac{q}{Q}t\right) = \bar{\gamma}_0 \bar{\gamma}_1^{\binom{t}{1}} \dots \bar{\gamma}_k^{\binom{t}{k}}$$

where $\bar{\gamma}_i = \gamma_i[\Gamma, \Gamma]$. By the claim for abelian groups we have for each $i \geq 1$ that $\bar{\gamma}_i = 1$, so that $\gamma_i \in [\Gamma, \Gamma]$. The claim now follows by induction on the derived sequence. □

Using this lemma, we obtain the following.

**Proposition 4.18** (Splitting). *Let the notation and hypotheses be as in Proposition 4.16, and assume $\ell_1 \geq \ell_2$. Let $e = (I_1'', I_2'', p_1', p_2') \in \mathcal{Q}'$ and $\vec{a} = (a_1, b_1, a_2, b_2)$ in $\mathcal{A}_e$, and write $\phi_{I_i''} = (I_i'', g_{I_i''})$ for $i = 1, 2$ and some $g_{I_i''} \in \mathrm{Poly}(\mathbb{R} \to G)$. Then we may factor*

$$g_{I_i''} = \tilde{g}_{e,\vec{a},i}\gamma_{e,\vec{a},i} \tag{96}$$

*for $i = 1, 2$, where $\tilde{g}_{e,\vec{a},i} \in \mathrm{Poly}(\mathbb{R} \to G)$ is of polynomial size (with exponents that can depend on $\ell_2$, but are independent of $\ell_1$) and $\gamma_{e,\vec{a},i}$ is $\prod \mathcal{P}_{e,\vec{a}}$-rational. Furthermore, we have the approximate dilation invariance*

$$\tilde{g}_{e,\vec{a},i}\left(\frac{a_i}{b_i}\cdot\right) = \varepsilon_i \tilde{g}_{e,\vec{a},i}\gamma_i \tag{97}$$

*for some $\gamma_i = \gamma_{i,e,\vec{a}} \in \Gamma$ of polynomial size, and some $\varepsilon_i = \varepsilon_{i,e,\vec{a}} \in \mathrm{Poly}(\mathbb{R} \to G)$ that is smooth on $I_i''$. In a similar vein we have*

$$\tilde{g}_{e,\vec{a},1}(p_2'\cdot) = \varepsilon^\dagger \tilde{g}_{e,\vec{a},2}(p_1'\cdot) \tag{98}$$

*for some $\varepsilon^\dagger = \varepsilon^\dagger_{e,\vec{a}} \in \mathrm{Poly}(\mathbb{R} \to G)$ that is smooth on $\frac{1}{p_2'}I_1''$, and*

$$\gamma_{e,\vec{a},1}(p_2'\cdot) = \gamma_{e,\vec{a},2}(p_1'\cdot)\gamma^\dagger \tag{99}$$

*for some $\gamma^\dagger = \gamma^\dagger_{e,\vec{a}} \in \mathrm{Poly}(\frac{1}{\prod \mathcal{P}_{e,\vec{a}}}\mathbb{Z} \to \Gamma)$.*

The fact that the polynomial size bounds for $\tilde{g}_{e,\vec{a},1}$ depend only on the smaller exponent $\ell_2$ rather than the larger one $\ell_1$ will be crucial in our subsequent analysis.

*Proof.* Let $i = 1, 2$, and set $Q := \prod \mathcal{P}_{e,\vec{a}}$. From (48) and Definition 4.9 one has

$$g_{I_i''}(a_i\cdot) = \varepsilon_i^* g_{I_i''}(b_i\cdot)\gamma_i^* \tag{100}$$

where $\gamma_i^*$ is $Q$-rational and $\varepsilon_i^*$ is smooth on $\frac{1}{a_i}I_i''$. Applying (28) to the polynomial $\log \varepsilon_i^*$ we conclude that $\varepsilon_i^*$ is of polynomial size (with exponents that do not depend on $\ell_1, \ell_2$). We now claim inductively for every $j = 1, \ldots, k+1$ that we can factor

$$g_{I_i''} = \tilde{g}_{e,\vec{a},i,j} g_{e,\vec{a},i,j}\gamma_{e,\vec{a},i,j} \tag{101}$$

where $\tilde{g}_{e,\vec{a},i,j} \in \mathrm{Poly}(\mathbb{R} \to G)$ is of polynomial size (with exponents that may depend on $\ell_i$ but not on $\ell_{3-i}$), $\gamma_{e,\vec{a},i,j} \in \mathrm{Poly}(\mathbb{R} \to G)$ is $Q$-rational, and $g_{e,\vec{a},i,j} \in \mathrm{Poly}(\mathbb{R} \to G_j)$ takes values in $G_j$; setting $j = k+1$ then gives the desired claim (96) for $i = 2$ at least; for $i = 1$ we will have the issue that the exponents depend on $\ell_1$ rather than $\ell_2$, but we will return to fix this issue later.

The inductive claim is trivial for $j = 1$ (set $g_{e,\vec{a},i,1} = g_{I_i''}$ with $\tilde{g}_{e,\vec{a},i,1}, \gamma_{e,\vec{a},i,1}$ trivial); now suppose that the claim has been established for some $1 \leq j \leq k$. In this argument all exponents are allowed to depend on $\ell_i$ but not on $\ell_{3-i}$. Then from (100) we see that

$$g_{e,\vec{a},i,j}(a_i\cdot) = \varepsilon_j g_{e,\vec{a},i,j}(b_i\cdot)\gamma_j$$

for some $\varepsilon_j$ of polynomial size and $Q$-rational $\gamma_j$ (we suppress the dependence of these maps on $e, \vec{a}, i$ for brevity). Quotienting by $G_j$ we see that $\varepsilon_j^{-1}$ and $\gamma_j$ agree modulo $G_j$, and

hence by Lemma 4.17 applied to $G/G_j$ are both equal modulo $G_j$ to a constant $\gamma \in \Gamma$ of polynomial size. Thus we have

$$g_{e,\vec{a},i,j}(a_i\cdot) = \tilde{\varepsilon}_j \gamma^{-1} g_{e,\vec{a},i,j}(b_i\cdot)\gamma\tilde{\gamma}_j$$

for some $\tilde{\varepsilon}_j$ of polynomial size taking values in $G_j$, and $Q$-rational $\tilde{\gamma}_j$ taking values in $G_j$. In the abelian group $G_j/G_{j+1}$, we thus have the identity

$$g_{e,\vec{a},i,j}(a_i\cdot) = \tilde{\varepsilon}_j g_{e,\vec{a},i,j}(b_i\cdot)\tilde{\gamma}_j \bmod G_{j+1}$$

and thus on taking logarithms and working in the abelian Lie algebra $\log G_j/\log G_{j+1}$ (noting from Appendix B that the logarithm map is a homomorphism from $G_j/G_{j+1}$ to $\log G_j/\log G_{j+1}$), we have from (180) that

$$\log g_{e,\vec{a},i,j}(a_i\cdot) = \log\tilde{\varepsilon}_j + \log g_{e,\vec{a},i,j}(b_i\cdot) + \log\tilde{\gamma}_j \bmod \log G_{j+1}.$$

For $d = 0,\dots,j$, we may differentiate $d$ times at 0 and rearrange to conclude that

$$(a_i^d - b_i^d)(\log g_{e,\vec{a},i,j})^{(d)}(0) = (\log\tilde{\varepsilon}_j)^{(d)}(0) + (\log\tilde{\gamma}_j)^{(d)}(0) \bmod \log G_{j+1}.$$

As $\tilde{\varepsilon}_j$ is of polynomial size, we have

$$(\log\tilde{\varepsilon}_j)^{(d)}(0) = O(X^{O(1)}).$$

Similarly, as $\tilde{\gamma}_j$ is $Q$-rational, $(\log\tilde{\gamma}_j)^{(d)}(0) \bmod \log G_{j+1}$ takes values in $\frac{Q}{q}\log\Gamma_j \bmod \log G_{j+1}$ for some positive integer $q$ of polynomial size. Since $a_i^d - b_i^d$ is also a positive integer of polynomial size, we conclude that

$$(\log g_{e,\vec{a},i,j})^{(d)}(0) = O(X^{O(1)}) + \frac{Q}{q_d}\gamma_d \bmod \log G_{j+1}$$

for some $\gamma_d \in \Gamma_j$ and positive integer $q_d$ of polynomial size. By Taylor expansion (and clearing denominators with the $q_d$), we may then write

$$\log g_{e,\vec{a},i,j} = \log g_j^* + \log\gamma_j^* \bmod \log G_{j+1}$$

where $g_j^* \in \mathrm{Poly}(\mathbb{R}\to G_j)$ is of polynomial size and $\gamma_j^* \in \mathrm{Poly}(\mathbb{R}\to G_j)$ is $Q$-rational. Exponentiating (noting that $G_j/G_{j+1}$ is abelian), we conclude that

$$g_{e,\vec{a},i,j} = g_j^* g_{e,\vec{a},i,j+1}\gamma_j^*$$

for some $g_{e,\vec{a},i,j+1} \in \mathrm{Poly}(\mathbb{R}\to G_{j+1})$. Inserting this into (101) we close the induction and establish (96) (with the above caveat regarding the exponents depending on $\ell_i$ rather than $\ell_2$).

From (45) we have

$$g_{I_1''}(p_2'\cdot) = \varepsilon^\dagger g_{I_2''}(p_1\cdot)\gamma^\dagger$$

for some $\varepsilon^\dagger \in \mathrm{Poly}(\mathbb{R} \to G)$ smooth on $\frac{1}{p_2'} I_1''$, and some $\gamma^\dagger \in \mathrm{Poly}(\frac{1}{Q}\mathbb{Z} \to \Gamma)$; in particular, $\gamma^\dagger$ is $Q$-rational with exponents that do not depend on $\ell_1$ or $\ell_2$. Combining this with (96) and rearranging, we see that

$$\tilde{g}_{e,\vec{a},2}(p_1'\cdot)^{-1}(\varepsilon^\dagger)^{-1}\tilde{g}_{e,\vec{a},1}(p_2'\cdot) = \gamma_{e,\vec{a},2}(p_1'\cdot)\gamma^\dagger\gamma_{e,\vec{a},1}(p_2'\cdot)^{-1}.$$

The left-hand side is of polynomial size and the right-hand side is $Q$-rational. Here the exponents depend on both $\ell_1, \ell_2$; since $\ell_1 \geq \ell_2$, we can view these exponents as depending on $\ell_1$ only. Applying Lemma 4.17, both sides are equal to a constant $\gamma \in \Gamma$ of polynomial size (with exponents depending on $\ell_1, \ell_2$). By multiplying $\tilde{g}_{e,\vec{a},1}$ on the right by $\gamma^{-1}$ (and $\gamma_{e,\vec{a},1}$ on the left by $\gamma$), we can assume that $\gamma = 1$, without significantly worsening any of the claimed properties of these objects, thus we may assume without loss of generality that $\gamma = 1$. Once one makes this normalization, one obtains the factorizations (98), (99). Furthermore, since the right-hand side of (98) is of polynomial size with exponents depending only on $\ell_2$, the left-hand side is also. Hence we have now resolved the previously mentioned caveat in (96) in that the exponents for the polynomial size nature of $\tilde{g}_{e,\vec{a},1}$ were depending on $\ell_1$ rather than $\ell_2$.

Inserting (96) back into (100) and rearranging, we conclude that

$$\tilde{g}_{e,\vec{a},i}(b_i\cdot)^{-1}(\varepsilon_i^*(t))^{-1}\tilde{g}_{e,\vec{a},i}(a_i t) = \gamma_{e,\vec{a},i}(b_i t)\gamma_i^*(t)\gamma_{e,\vec{a},i}^{-1}(a_i t).$$

As the left-hand side is of polynomial size and the right-hand side is $Q$-rational, we conclude from Lemma 4.17 that both sides are equal to a constant $\gamma_i \in \Gamma$ of polynomial size. This rearranges to give

$$\tilde{g}_{e,\vec{a},i}(a_i t) = \varepsilon_i^*(t)\tilde{g}_{e,\vec{a},i}(b_i t)\gamma_i$$

and therefore the claim (97) follows from reparameterizing $t$ and defining $\varepsilon_i(t) \coloneqq \varepsilon_i^*(a_i t)$. □

At this point we encounter a minor technical complication due to the fact that the factors $\tilde{g}_{e,\vec{a},i}, \gamma_{e,\vec{a},i}$ generated by the above proposition depend on $\vec{a}$, so in particular as one varies $a_i, b_i$ the polynomial map $\tilde{g}_{e,\vec{a},i}$ appearing in relations such as (97) also varies. Fortunately, using some arguments of a graph theoretic nature, and taking advantage of the ability to make the two parameters $\ell_1, \ell_2$ differ significantly from each other, we can eliminate this dependence:

**Proposition 4.19** (Approximate dilation invariance for a dense set of dilations)**.** *Let $e = (I_1'', I_2'', p_1', p_2') \in \mathcal{Q}'$, and let $g_{I_1''}, g_{I_2''} \in \mathrm{Poly}(\mathbb{R} \to G)$ be the maps associated to $\phi_{I_1''}, \phi_{I_2''}$. Assume that $\ell_1$ is sufficiently large depending on $\ell_2$. Then there is a large set $\mathcal{P}_e$ of primes in $[P''/2, P'']$ and a factorization*

$$g_{I_i''} = \tilde{g}_{e,i}\gamma_{e,i} \tag{102}$$

*for each $i = 1, 2$, where $\tilde{g}_{e,i} \in \mathrm{Poly}(\mathbb{R} \to G)$ is of polynomial size and $\gamma_{e,i}$ is $\prod \mathcal{P}_e$-rational, one has the relation*

$$\tilde{g}_{e,1}(p_2'\cdot) = \varepsilon^\dagger \tilde{g}_{e,2}(p_1\cdot) \tag{103}$$

*for some $\varepsilon^\dagger \in \mathrm{Poly}(\mathbb{R} \to G)$ that is smooth on $\frac{1}{p_2'}I_1''$, and one has the relation*

$$\gamma_{e,1}(p_2'\cdot) = \gamma_{e,2}(p_1'\cdot)\gamma^\dagger \tag{104}$$

*for some $\gamma^\dagger \in \mathrm{Poly}(\frac{1}{\prod \mathcal{P}_e}\mathbb{Z} \to \Gamma)$. (In all these cases we permit the exponents to depend on both $\ell_1$ and $\ell_2$.) Furthermore:*

(i) *There exists a measurable subset $\Omega_e$ of the interval $[1+\frac{1}{CN}, 1+\frac{C}{N}]$ for some fixed constant $C > 0$ of measure $\gg 1/N$, such that for each $\alpha \in \Omega_e$ one has the approximate dilation invariance*

$$\tilde{g}_{e,1}(\alpha\cdot) = \varepsilon_\alpha \tilde{g}_{e,1}\gamma_\alpha \tag{105}$$

*for some $\gamma_\alpha \in \Gamma$ of polynomial size, and some $\varepsilon_\alpha \in \mathrm{Poly}(\mathbb{R} \to G)$ that is smooth on $I_1''$.*

(ii) *We have $\tilde{g}_{e,1}(x_{I_1''}) = O(1)$.*

*Proof.* We first observe that we may drop the conclusion (ii) as follows. Suppose we have already obtained all the conclusions of the proposition other than (ii). Then $\tilde{g}_{e,1}(x_{I_1''})$ is already of polynomial size. Since $G/\Gamma$ is compact, we may write

$$\tilde{g}_{e,1}(x_{I_1''}) = O(1)\gamma$$

for some $\gamma \in \Gamma$ of polynomial size. If we then multiply $\tilde{g}_{e,1}$ on the right by $\gamma^{-1}$, multiply $\gamma_{e,1}$ and $\gamma^\dagger$ on the left by $\gamma$, and replace the lattice element $\gamma_\alpha$ appearing in (105) by $\gamma\gamma_\alpha\gamma^{-1}$, we thus see that we may recover the claimed property (ii), without significantly impacting any of the other claims.

Henceforth we focus on establishing the remaining conclusions of the proposition. For $i = 1, 2$, let $V_i$ denote the set of ratios $\frac{a_i}{b_i}$ of coprime positive integers $a_i, b_i$ that are products of $\ell_i$ primes in $[P', 2P']$ with

$$\frac{a_i}{b_i} - 1 \asymp \frac{1}{N} \asymp \frac{H}{X}.$$

By [26, Lemma 2.6], $V_i$ has cardinality $O(d^{\ell_i}/N)$. From Proposition 4.16, we see that for any $e \in \mathcal{Q}'$, the set

$$E_e \coloneqq \left\{ (\frac{a_1}{b_1}, \frac{a_2}{b_2}) : (a_1, b_1, a_2, b_2) \in \mathcal{A}_e \right\}$$

is a subset of $V_1 \times V_2$ of cardinality $\gg d^{\ell_1+\ell_2}/N^2$, thus $\#V_i \asymp d^{\ell_i}/N$ and $\#E_e \asymp (\#V_1)(\#V_2)$. We view $E_e$ as a dense bipartite graph on $V_1, V_2$. Each edge $\vec{a} = (\frac{a_1}{b_1}, \frac{a_2}{b_2})$ in $E_e$ is associated

to a large set of primes $\mathcal{P}_{e,\vec{a}} \coloneqq \mathcal{P}_{e,(a_1,b_1,a_2,b_2)}$ in $[P''/2, P'']$. In particular

$$\sum_{\vec{a}\in E_e} \#\mathcal{P}_{e,\vec{a}} \gg (\#V_1)(\#V_2)\pi_0(P'')$$

which we rearrange as

$$\sum_{p''\in[P''/2,P'']} \sum_{v_2\in V_2} \#\{v_1 \in V_1 : (v_1,v_2)\in E_e; p'' \in \mathcal{P}_{e,(v_1,v_2)}\} \gg \pi_0(P'')(\#V_1)(\#V_2).$$

By Cauchy–Schwarz, this implies that

$$\sum_{p''\in[P''/2,P'']} \sum_{v_2\in V_2} \#\{v_1 \in V_1 : (v_1,v_2)\in E_e; p'' \in \mathcal{P}_{e,(v_1,v_2)}\}^2 \gg \pi_0(P'')(\#V_1)^2(\#V_2),$$

which we rearrange as

$$\sum_{(v_1,v_2)\in E_e} \sum_{v_1'\in V_1:(v_1',v_2)\in E_e} \#(\mathcal{P}_{e,(v_1,v_2)} \cap \mathcal{P}_{e,(v_1',v_2)}) \gg \pi_0(P'')(\#V_1)^2(\#V_2).$$

Hence by the pigeonhole principle there exists $(v_1,v_2)\in E_e$ for which

$$\sum_{v_1'\in V_1:(v_1',v_2)\in E_e} \#(\mathcal{P}_{e,(v_1,v_2)} \cap \mathcal{P}_{e,(v_1',v_2)}) \gg \pi_0(P'')\#V_1$$

which implies that

$$\#(\mathcal{P}_{e,(v_1,v_2)} \cap \mathcal{P}_{e,(v_1',v_2)}) \gg \pi_0(P'')$$

and $(v_1',v_2)\in E_e$ for all $v_1'$ in a subset $V_e$ of $V_1$ of cardinality $\gg \#V_1 \gg d^{\ell_1}/N$.

Set $\mathcal{P}_e \coloneqq \mathcal{P}_{e,(v_1,v_2)}$. From Proposition 4.18 applied to the quadruple $(v_1,v_2)$, we obtain factorizations

$$g_{I_i''} = \tilde{g}_{e,i}\gamma_{e,i} \tag{106}$$

for $i=1,2$, where $\tilde{g}_{e,i} = \tilde{g}_{e,(v_1,v_2),i} \in \mathrm{Poly}(\mathbb{R}\to G)$ is of polynomial size and $\gamma_{e,i} = \gamma_{e,(v_1,v_2),i}$ is $\prod \mathcal{P}_e$-rational, obeying

$$\tilde{g}_{e,1}(p_2'\cdot) = \varepsilon^\dagger \tilde{g}_{e,2}(p_1\cdot) \tag{107}$$

for some $\varepsilon^\dagger = \varepsilon^\dagger_{e,(v_1,v_2)} \in \mathrm{Poly}(\mathbb{R}\to G)$ that is smooth on $\frac{1}{p_2'}I_1''$. For any $v_1'\in V_e$, we also have a factorization

$$g_{I_1''} = \tilde{g}_{e,(v_1',v_2),1}\gamma_{e,(v_1',v_2),1}, \tag{108}$$

where $\tilde{g}_{e,(v_1',v_2),1}$ is of polynomial size and $\gamma_{e,(v_1',v_2),1}$ is $\prod \mathcal{P}_{e,(v_1',v_2)}$-rational, and

$$\tilde{g}_{e,(v_1',v_2),1}(v_1'\cdot) = \varepsilon_{v_1'}\tilde{g}_{e,(v_1',v_2),1}\gamma_{v_1'} \tag{109}$$

for some $\varepsilon_{v_1'}$ smooth on $I_1''$ and $\gamma_{v_1'}\in\Gamma$ of polynomial size. From (106), (108) we have

$$\tilde{g}_{e,1}^{-1}\tilde{g}_{e,(v_1',v_2),1} = \gamma_{e,1}\gamma_{e,(v_1',v_2),1}^{-1}.$$

The left-hand side is of polynomial size and the right-hand side is $\prod(\mathcal{P}_{e,(v_1,v_2)} \cap \mathcal{P}_{e,(v_1',v_2)})$-rational. By Lemma 4.17, both sides are then equal to a constant $\gamma^*_{v_1'} \in \Gamma$ of polynomial size, thus

$$\tilde{g}_{e,(v_1',v_2),1} = \tilde{g}_{e,1}\gamma^*_{v_1'}.$$

We conclude from (109) that

$$\tilde{g}_{e,1}(v_1'\cdot) = \varepsilon_{v_1'}\tilde{g}_{e,1}\tilde{\gamma}_{v_1'}$$

for all $t \in \mathbb{R}$, where $\tilde{\gamma}_{v_1'} \coloneqq \gamma^*_{v_1'}\gamma_{v_1'}(\gamma^*_{v_1'})^{-1}$ is an element of $\Gamma$ of polynomial size. This gives the bound (105) for all $\alpha$ in the discrete set $V_e$. This is not yet what we need because $V_e$ has measure zero. However we can use the hypothesis that $\ell_1$ is large compared to $\ell_2$ to remove the discretization as follows. Recall from Proposition 4.18 that $\tilde{g}_{e,1}$ is of polynomial size, with exponents depending only on the smaller parameter $\ell_2$ and not on the larger parameter $\ell_1$. As a consequence, if (105) holds for some real number $\alpha = 1 + O(\frac{1}{N})$, then one can perturb $\alpha$ by at most $d^{-\ell_1/10}$ (say) and still retain (105) with only a negligible change in all the implied constants. Hence we have (105) for all $\alpha \in \Omega_e$, where $\Omega_e$ is the $d^{-\ell_1/10}$-neighborhood of $V_e$. We have

$$\int_{\Omega_e} \sum_{\alpha \in V_e} 1_{[\alpha - d^{-\ell_1/10}, \alpha + d^{-\ell_1/10}]}(\beta)\, d\beta = 2d^{-\ell_1/10}\#V_e \gg d^{9\ell_1/10}/N.$$

To obtain the desired lower bound of $\gg 1/N$ on the measure of $\Omega_e$, it suffices to establish the pointwise bound

$$\sum_{\alpha \in V_e} 1_{[\alpha - d^{-\ell_1/10}, \alpha + d^{-\ell_1/10}]}(\beta) \ll d^{9\ell_1/10}$$

for any $\beta = 1 + O(1/N)$. The left-hand side can be written as

$$\#\{\alpha \in V_e : |\alpha - \beta| \le d^{-\ell_1/10}\}.$$

This in turn can be bounded by the number of pairs $(a, b) \in S^2$ with $\frac{a}{b} = \beta + O(d^{-\ell_1/10})$, where $S$ is the collection of products of $\ell_1$ primes in $[P', 2P']$. This can then be bounded by

$$d^{\ell_1/10} \int_0^\infty f(t) f(\beta t) \frac{dt}{t}$$

where

$$f(t) \coloneqq \#(S \cap [(1 - C_1 d^{-\ell_1/10})t, (1 + C_1 d^{-\ell_1/10})t])$$

for some absolute constant $C_1 > 0$. By Cauchy–Schwarz, the previous expression may be bounded by

$$d^{\ell_1/10} \int_0^\infty f(t)^2 \frac{dt}{t}$$

which is in turn bounded by the number of pairs $(a, b) \in S^2$ with $\frac{a}{b} = 1 + O(d^{-\ell_1/10})$. Applying [26, Lemma 2.6], this quantity is $O(d^{9\ell_1/10})$, and the claim follows. □

Now that we have established an approximate dilation invariance (105) for a large set of dilation parameters $\alpha$, we can begin solving this equation effectively. The first step is as follows.

**Proposition 4.20.** *Let $e, I_1'', \tilde{g}_{e,1}, \Omega_e, \gamma_\alpha$ be as in Proposition 4.19. Then for any $\alpha \in \Omega_e$, we have the estimate*

$$\tilde{g}_{e,1}(t) = O(1)\gamma_\alpha^{\frac{\log(t/x_{I_1''})}{\log \alpha}} \tag{110}$$

*for real $t$ with $\langle t \rangle_{I_1''} \ll 1$. As a consequence, for any $\alpha, \alpha' \in \Omega_e$, we have*

$$\gamma_{\alpha'}^s = O(1)\gamma_\alpha^{s\frac{\log \alpha'}{\log \alpha}}. \tag{111}$$

*for all $s = O(1)$.*

*Proof.* From iterating (105) we see that for any fixed natural number $n$ and any $\alpha \in \Omega_e$ we have

$$\tilde{g}_{e,1}(\alpha^n x_{I_1''}) = O(1)\tilde{g}_{e,1}(x_{I_1''})\gamma_\alpha^n$$

which we rearrange as

$$\tilde{g}_{e,1}(\exp(n \log \alpha) x_{I_1''})\gamma_\alpha^{-n} = O(1). \tag{112}$$

The left-hand side is a (matrix-valued) exponential polynomial in $n$, with the exponents in the exponentials being bounded multiples of $\log \alpha$ and thus of size $O(1/N)$. Applying Lemma 2.3 to each component of this matrix-valued function, we conclude that (112) holds for all *real* $n = O(1)$. Rearranging using the fact that $\log \alpha \asymp \frac{1}{N}$, we conclude the estimate (110). Applying this estimate twice we conclude that

$$\tilde{g}_{e,1}(e^{s \log \alpha'} x_{I_1''}) = O(1)\gamma_{\alpha'}^s$$

and

$$\tilde{g}_{e,1}(e^{s \log \alpha'} x_{I_1''}) = O(1)\gamma_\alpha^{s\frac{\log \alpha'}{\log \alpha}}$$

for $\alpha, \alpha' \in \Omega_e$ and $s = O(1)$, giving (111). □

Now we give some satisfactory control on $\tilde{g}_{e,1}$, which roughly speaking asserts that $\tilde{g}_{e,1}$ "pretends to be like" $t \mapsto T^{\log(t/x_{I_1''})}$ for some $T$ which is either nearly central, or nearly contained in a proper subgroup of $G$. Following [16], we define a *horizontal character* to be a continuous additive homomorphism $\eta : G \to \mathbb{R}/\mathbb{Z}$ that annihilates $\Gamma$; its derivative $d\eta : \log G \to \mathbb{R}$ at the identity is then a linear functional on $\log G$, and is related to $\eta$ by the formula

$$\eta(g) = d\eta(\log g) \bmod \mathbb{Z}, \tag{113}$$

as can be seen by starting with the formula $\eta(g) = n\eta(\exp(\frac{1}{n}\log g))$ and taking limits as $n \to \infty$. In particular, $\eta$ is the descent of the homomorphism $d\eta \circ \log : G \to \mathbb{R}$ to $\mathbb{R}/\mathbb{Z}$.

**Example 4.21.** *Let $G$ be the Heisenberg group from Example 4.1, and let $\Gamma$ be the lattice*

$$\Gamma := \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.$$

*Then every horizontal character $\eta\colon G \to \mathbb{R}/\mathbb{Z}$ takes the form*

$$\eta\left(\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}\right) = ax + by \bmod 1$$

*for some integers $a, b$, and the corresponding map $d\eta\colon \log G \to \mathbb{R}$ is given by*

$$d\eta\left(\begin{pmatrix} 0 & x & z \\ 0 & 1 & y \\ 0 & 0 & 0 \end{pmatrix}\right) = ax + by.$$

**Proposition 4.22** (Description of $\tilde{g}_{e,1}$)**.** *Let $e, \tilde{g}_{e,1}, I_1''$ be as in Proposition 4.19.*

*(1) Then there exists $T = T_e \in G$ of polynomial size such that the map*

$$t \mapsto \log\left(\tilde{g}_{e,1}(t) T^{-\log(t/x_{I_1''})}\right) \tag{114}$$

*is bounded by $O(1)$ and has a Lipschitz norm of $O(|I_1''|^{-1})$ whenever $\langle t \rangle_{I_1''} \ll 1$.*

*(2) There is a non-trivial horizontal character $\eta = \eta_e : G \to \mathbb{R}/\mathbb{Z}$ such that $d\eta : \log G \to \mathbb{R}$ has operator norm $O(1)$, and such that*

$$d\eta(\log T) = O(N). \tag{115}$$

*Proof.* Let $\Omega_e$ and $\gamma_\alpha$ be as in Proposition 4.19. Let $\alpha_0$ be an arbitrary element of $\Omega_e$, and let $T \in G$ be the quantity

$$T := \gamma_{\alpha_0}^{\frac{1}{\log \alpha_0}}.$$

Since $\gamma_\alpha$ is of polynomial size and $\alpha_0 - 1 \asymp \frac{1}{N}$, we see that $T$ is also of polynomial size. From (110) one has

$$\tilde{g}_{e,1}(t) = O(1) T^{\log(t/x_{I_1''})} \tag{116}$$

whenever $\langle t \rangle_{I_1''} \ll 1$. In particular, after making the substitution $u := N \log(t/x_{I_1''})$, the function

$$u \mapsto \log\left(\tilde{g}_{e,1}(e^{u/N} x_{I_1''}) T^{-u/N}\right)$$

is bounded for $u = O(1)$. By the Baker–Campbell–Hausdorff formula (see Appendix B), this map is an exponential polynomial involving $O(1)$ terms with exponents of order $O(1/N)$. (Note that the quantity $T^{-u/N} = \exp(-u \log T/N)$ is actually a polynomial in $u$, rather than an exponential polynomial, due to the nilpotent nature of $G$.) Applying Lemma 2.3, we conclude that this map has a Lipschitz constant of $O(1)$. Undoing the substitution, we obtain the claims regarding (114).

Applying (110) again and combining with (116), we see that

$$\gamma_\alpha^s = O(1) T^{s \log \alpha}$$

for all $\alpha \in \Omega_e$ and $s = O(1)$. If we then write

$$g_\alpha := T^{\log \alpha} \gamma_\alpha^{-1}$$

then $g_\alpha = T^{\log \alpha} \bmod \Gamma$ and $g_\alpha = O(1)$. Furthermore, for any $s = O(1)$ we have

$$\begin{aligned} T^{s \log \alpha} g_\alpha T^{-s \log \alpha} &= T^{(s+1) \log \alpha} \gamma_\alpha^{-1} T^{-s \log \alpha} \\ &= O(1) \gamma_\alpha^{s+1} \gamma_\alpha^{-1} (O(1) \gamma_\alpha^s)^{-1} \\ &= O(1), \end{aligned}$$

and thus

$$T^t g_\alpha T^{-t} = O(1)$$

for all $\alpha \in \Omega_e$ and $t = O(\frac{1}{N})$. Taking logarithms and applying the Lie algebra identity (178), we may rewrite this as

$$e^{t \mathrm{ad}_{\log T}} \log g_\alpha = O(1) \tag{117}$$

for all $t = O(\frac{1}{N})$ and $\alpha \in \Omega_e$.

Let $C_0 > 0$ be a sufficiently large fixed quantity to be chosen later. Suppose first that $\frac{1}{N} \mathrm{ad}_{\log T}$ has operator norm less than $C_0$. The map $\mathrm{ad} : X \mapsto \mathrm{ad}_X$ is a fixed linear map from $\log G$ to the space $\mathrm{End}(\log G)$ of linear endomorphisms of $\log G$, and its kernel is $\log Z(G)$ where $Z(G)$ is the center of $G$. The image of $\frac{1}{N} \log T$ under this map has size $O(C_0)$, hence $\frac{1}{N} \log T$ lies at a distance $O(C_0)$ from $\log Z(G)$. On the other hand, from Proposition 4.7, $\log Z(G)$ is a proper normal subalgebra of the Lie algebra $\log G$; using Mal'cev bases (for definition, see Appendix B) it can also be seen to be rational. By lifting a non-trivial horizontal character of $G/Z(G)$ (which can be in turn obtained by lifting a non-trivial character from the horizontal torus formed by quotienting out $G/Z(G)$ by both $\Gamma Z(G)/Z(G)$ and the commutator group $[G/Z(G), G/Z(G)]$), we may thus find a fixed non-trivial horizontal character $\eta$ that annihilates $\log Z(G)$ and such that $d\eta$ has operator norm $O(1)$, so that (115) holds, in which case we are done.

Henceforth we may assume that $\frac{1}{N} \mathrm{ad}_{\log T}$ has operator norm at least $C_0$. As $\frac{1}{N} \mathrm{ad}_{\log T}$ is nilpotent, we conclude (on finite Taylor expansion of the logarithm map) that the linear map $e^{\frac{1}{N} \mathrm{ad}_{\log T}}$ has operator norm $\gg C_0^c$ for some constant $c > 0$. From this and the singular value decomposition, we conclude that the set

$$\Omega = \{x \in \log G : e^{\frac{1}{N} \mathrm{ad}_{\log T}} x = O(1)\}$$

lies in the $O(C_0^{-c})$-neighbourhood of a hyperplane $\Pi$ in $\log G$. From (117) we conclude that for $\alpha \in \Omega_e$, $\log g_\alpha$ lies within $O(C_0^{-c})$-neighbourhood of $\Pi$. Since we have $g_\alpha = O(1)$ and $g_\alpha = T^{\log \alpha} \bmod \Gamma$, we thus have

$$T^{\log \alpha} \Gamma = g_\alpha \Gamma \in \{\kappa \exp(h) \Gamma : \kappa \in G; h \in \Pi; \kappa = O(C_0^{-c}); h = O(1)\}. \tag{118}$$

Thus, for $t = O(1/N)$ in a set of measure $\asymp 1/N$, $T^t \Gamma$ is contained in the $O(C_0^{-c})$-neighbourhood of the set

$$\Sigma := \{\exp(h) \Gamma : h \in \Pi; h = O(1)\}.$$

Discretising this using the polynomial size of $T$, we conclude (for $A > 0$ a large enough constant) that $T^t\Gamma$ lies in the $O(C_0^{-c})$ neighbourhood of $\Sigma$ for $\gg X^A/N$ values of $t = O(1/N)$ with $t \in X^{-A}\mathbb{Z}$. If $C_0$ is large enough, this implies that the sequence $n \mapsto T^{X^{-A}n}\Gamma$ fails to be $C_0^{-C}$-equidistributed on the interval $[-CX^A/N, CX^A/N] \cap \mathbb{Z}$ for some fixed $C > 0$, in the sense of [16, Definition 1.2] (by testing this equidistribution hypothesis against a suitable cutoff function adapted to the $O(C_0^{-c})$-neighbourhood of $\Sigma$). Applying [16, Theorem 1.16], this implies that there is a non-trivial horizontal character $\eta\colon G \to \mathbb{R}/\mathbb{Z}$ with $d\eta$ having operator norm[18] $O(C_0^{O(1)})$, such that

$$\|\eta(T^{X^{-A}n}) - \eta(T^{X^{-A}(n-1)})\|_{\mathbb{R}/\mathbb{Z}} \ll X^{-A}N$$

for $n \in [-CX^A/N, CX^A/N] \cap \mathbb{Z}$, which by (113) implies that

$$X^{-A}d\eta(\log T) = O(X^{-A}N) \bmod \mathbb{Z}.$$

For $A$ large enough, both sides here are less than $1/2$ in magnitude, so we may remove the mod $\mathbb{Z}$ constraint. The claim follows. □

**Remark 4.23.** *Proposition 4.22(2) is the first place where the non-abelian nature of $G$ plays a role. Part (1) of Proposition 4.22 is valid for abelian groups as well. However in part (2), if the group $G$ is abelian, then we can not find a character $\eta$ with the desired properties since the action of $\mathrm{ad}_{\log T}$ is trivial.*

Having established satisfactory control on the "continuous" (or "Archimedean") component $\tilde{g}_{e,1}$ on the factorization from Proposition 4.19, we now need to control the "rational" (or "non-Archimedean") component $\gamma_{e,i}$, with the ultimate aim being to establish an analogue of Proposition 3.13. We begin with a variant of Corollary 3.10. We view $\Gamma$ as a subgroup of $\mathrm{Poly}(\mathbb{R} \to G)$, by identifying each element $\gamma$ of $\Gamma$ with the constant polynomial map $t \mapsto \gamma$. In particular we may form the quotient space $\Gamma\backslash\mathrm{Poly}(\mathbb{R} \to G)$.

**Proposition 4.24.** *For each $I'' \in \mathcal{I}''$ there exists a set $\mathcal{F}(I'')$ of elements of the quotient space $\Gamma\backslash\mathrm{Poly}(\mathbb{R} \to G)$ of cardinality $O(1)$ such that for any quadruple $e = (I_1'', I_2'', p_1', p_2') \in \mathcal{Q}'$ and functions $\gamma_{e,i}$ as in Proposition 4.19, one has*

$$\Gamma\gamma_{e,i} \in \mathcal{F}(I'') \tag{119}$$

*if $i = 1, 2$ and $I_i'' = I''$. Furthermore, each element in $\mathcal{F}(I'')$ is 1-rational, that is to say it lies in $\Gamma\backslash\mathrm{Poly}(q\mathbb{Z} \to \Gamma)$ for some positive integer $q$ of polynomial size.*

*Proof.* We just prove the claim for $i = 1$, as the $i = 2$ case is similar, and then we can obtain the joint case $i = 1, 2$ by taking the union of the two sets $\mathcal{F}(I'')$ thus produced. We let $\mathcal{F}(I'')$ be the collection of all cosets $\Gamma\gamma_{e,1}$ whenever $e = (I_1'', I_2'', p_1', p_2') \in \mathcal{Q}'$ with $I_1'' = I''$. Since $\gamma_{e,i}$ is $Q$-rational for some natural number $Q$, it is also 1-rational.

[18] More precisely, [16, Theorem 1.16] shows that $\eta$, when expressed in Mal'cev coordinates, is given by a linear functional with coefficients $O(C_0^{O(1)})$, from which it is easy to verify that $d\eta$ is also a linear functional with coefficients $O(C_0^{O(1)})$.

Clearly we have the property (119) by definition for $i = 1$. To complete the proof of the proposition, we need to show that $\mathcal{F}(I'')$ has cardinality $O(1)$. Suppose for contradiction that $\mathcal{F}(I'')$ has cardinality at least $K$ for some large fixed $K$ to be chosen later. By construction, we can then find $K$ quadruples $e_j = (I'', I''_{2,j}, p'_{1,j}, p'_{2,j}) \in \mathcal{Q}'$ for $j = 1, \dots, K$ and associated factorizations

$$g_{I''} = \tilde{g}_{e_j,1}\gamma_{e_j,1} \tag{120}$$

for $j = 1, \dots, K$, with $\tilde{g}_{e_j,1} \in \mathrm{Poly}(\mathbb{R} \to G)$ of polynomial size and $\gamma_{e_j,1} \prod \mathcal{P}_{e_j}$-rational for some large set $\mathcal{P}_{e_j}$ of primes in $[P''/2, P'']$, and such that the cosets $\Gamma\gamma_{e_j,1}$ are all distinct. As each $\mathcal{P}_j$ is large, we have

$$\sum_{j=1}^{K} \#\mathcal{P}_{e_j} \gg K\pi_0(P'').$$

The left-hand side can be written as $\sum_{p\in[P''/2,P'']} \#\{1 \le j \le K : p \in \mathcal{P}_j\}$. By Cauchy–Schwarz we then have

$$\sum_{p\in[P''/2,P'']} \#\{1 \le j \le K : p \in \mathcal{P}_{e_j}\}^2 \gg K^2\pi_0(P'').$$

The left-hand side may be written as

$$\sum_{1\le j,j'\le K} \#(\mathcal{P}_{e_j} \cap \mathcal{P}_{e_{j'}}).$$

For $K$ large enough, we may delete the diagonal contribution $j = j'$ and then use the pigeonhole principle to conclude that there exists $1 \le j < j' \le K$ for which $\mathcal{P}_{e_j} \cap \mathcal{P}_{e_{j'}}$ is large. For this $j, j'$, we use (120) to conclude that

$$\tilde{g}^{-1}_{e_{j'},1}\tilde{g}_{e_j,1} = \gamma_{e_{j'},1}\gamma^{-1}_{e_j,1}.$$

The left-hand side is of polynomial size, and the right-hand side is $\prod(\mathcal{P}_{e_j} \cap \mathcal{P}_{e_{j'}})$-rational. By Lemma 4.17, we conclude that $\gamma_{e_{j'},1}\gamma^{-1}_{e_j,1} \in \Gamma$, thus $\Gamma\gamma_{e_j,1} = \Gamma\gamma_{e_{j'},1}$, contradicting the construction of the $e_j$. The claim follows. □

We can now establish an analogue of Proposition 3.13 that dramatically improves the bound on $q$.

**Proposition 4.25.** *There exists a subset $\mathcal{Q}''$ of $\mathcal{Q}'$ of cardinality $\gg dN$, such that for each $e \in \mathcal{Q}''$ and functions $\gamma_{e,i}$ as in Proposition 4.19, one has $\gamma_{e,1} \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma)$ for some $q = O(1)$.*

*Proof.* Let $q_0$ be a sufficiently large fixed quantity to be chosen later. Suppose for contradiction that the proposition fails, then we can find a subset $\mathcal{Q}''$ of $\mathcal{Q}'$ of cardinality at least $\frac{1}{2}\#\mathcal{Q}' \gg dN$ such that $\gamma_{e,1} \notin \mathrm{Poly}(q\mathbb{Z} \to \Gamma)$ for any $1 \le q \le q_0$ and $e = (I''_1, I''_2, p'_1, p'_2) \in \mathcal{Q}''$. By Proposition 4.24, $\Gamma\gamma_{e,1} \in \mathcal{F}(I''_1)$. By randomly selecting one element $F_{I''_1}$ from each

$\mathcal{F}(I_1'')$ and using the probabilistic method, we conclude that for at least one such choice of elements $F_{I_1''}$, there is a subset $\mathcal{Q}'''$ of $\mathcal{Q}''$ of cardinality $\gg dN$ such that

$$\Gamma\gamma_{e,i} = F_{I_i''}$$

for all $e = (I_1'', I_2'', p_1', p_2') \in \mathcal{Q}'''$ and $i = 1, 2$. In particular, we have

$$F_{I_1''} \notin \Gamma\backslash\mathrm{Poly}(q\mathbb{Z} \to \Gamma)$$

whenever $e = (I_1'', I_2'', p_1', p_2') \in \mathcal{Q}'''$ and $1 \leq q \leq q_0$.

Let $\ell$ be a bounded integer, large enough so that $d^\ell \geq Nd^{10}$. Viewing $\mathcal{Q}'''$ as a (directed) graph with vertex set $\mathcal{I}''$ and applying the Blakley–Roy inequality [3] (see also [30]) and Cauchy–Schwarz to count cycles of length $2\ell$ in this graph, we conclude that there exist $\gg d^{2\ell}$ $2\ell$-tuples

$$(I_{j,i}'')_{0\leq j\leq \ell-1; i=1,2} \in (\mathcal{I}'')^{2\ell} \tag{121}$$

with the property that for each $0 \leq j \leq \ell - 1$, there exists primes $p_{j,1}', p_{j,2}', p_{j,3}', p_{j,4}' \in [P', 2P']$ such that

$$(I_{j,1}'', I_{j,2}'', p_{j,1}', p_{j,2}'), (I_{j+1,1}'', I_{j,2}'', p_{j,3}', p_{j,4}') \in \mathcal{Q}'''$$

for $j = 0, \dots, \ell - 1$, with the periodic convention $I_{\ell,1}'' = I_{0,1}''$. In particular, $I_{j,1}''$ lies within $O(\frac{H}{P'P''})$ of $\frac{p_{j,2}'}{p_{j,1}'} I_{j,2}''$, and similarly $I_{j+1,1}''$ lies within $O(\frac{H}{P'P''})$ of $\frac{p_{j,4}'}{p_{j,3}'} I_{j,2}''$. Iterating this we conclude that $I_{0,1}''$ lies within $O(\frac{H}{P'P''})$ of $\prod_{j=0}^{\ell-1} \frac{p_{j,4}' p_{j,1}'}{p_{j,3}' p_{j,2}'} I_{0,1}''$, which implies that

$$|a - b| \ll \frac{1}{N}(P')^{2\ell}. \tag{122}$$

where

$$a \coloneqq \prod_{j=0}^{\ell-1} p_{j,4}' p_{j,1}'$$

and

$$b \coloneqq \prod_{j=0}^{\ell-1} p_{j,3}' p_{j,2}'.$$

By the pigeonhole principle, we may find an $I_{0,1}'' \in \mathcal{I}''$ which is associated to a family $\mathcal{T}$ of tuples (121) of cardinality

$$\#\mathcal{T} \gg d^{2\ell}/N. \tag{123}$$

We now fix this interval $I_{0,1}''$ and the family $\mathcal{T}$.

Let $q$ be the least positive integer for which

$$F_{I_{0,1}''} \in \Gamma\backslash\mathrm{Poly}(q\mathbb{Z} \to \Gamma).$$

By construction we have

$$q_0 < q \ll X^{O(1)}.$$

Intuitively, the lower bound $q > q_0$ means that polynomials in the coset $F_{I_{0,1}''}$ have at least one coefficient with some "large denominator" $n_m$. The strategy is to locate this coefficient

and this denominator, and then to study the equation (104) to obtain some non-trivial congruence conditions relating $a$ and $b$ modulo $n_m$ which will restrict the size of $\mathcal{T}$ enough to obtain a contradiction.

We turn to the details. We arbitrarily select a coset representative $\gamma_{0,1} \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma)$ of $F_{I''_{0,1}}$. For any $l = 1, \dots, k+1$, we let $\gamma_{0,1} \bmod G_l$ be the projection to $\mathrm{Poly}(q\mathbb{Z} \to \Gamma G_l/G_l) \subset \mathrm{Poly}(\mathbb{R} \to G/G_l)$, and let $q_l$ be the least positive integer for which $\gamma_{0,1} \bmod G_l \in \mathrm{Poly}(q_l\mathbb{Z} \to \Gamma G_l/G_l)$. Then $q_1 = 1$, $q_{k+1} = q \geq q_0$, and from Lemma 4.13 we have $q_i | q_{i+1}$ for $i = 1, \dots, k$. In particular, by the pigeonhole principle we can find $l \in \{1, \dots, k\}$ such that

$$q_l \leq q_0^{\frac{i-1}{k}}$$

and

$$q_{l+1} > q_0^{\frac{i}{k}} \geq q_0^{\frac{1}{k}} q_l. \tag{124}$$

We now fix this $l$. By lifting the Taylor coefficients of $\gamma_{0,1} \bmod G_l$ from $G/G_l$ back to $G$, we can factor

$$\gamma_{0,1} = \gamma'_{0,1}\gamma''_{0,1} \tag{125}$$

where $\gamma''_{0,1} \in \mathrm{Poly}(q_l\mathbb{Z} \to \Gamma)$ and $\gamma'_{0,1} \in \mathrm{Poly}(\mathbb{R} \to G_l)$, hence also $\gamma'_{0,1} \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma_l)$. We then see that $q_{l+1}$ is the least multiple of $q_l$ for which $\gamma'_{0,1} \bmod G_{l+1} \in \mathrm{Poly}(q_{l+1}\mathbb{Z} \to \Gamma_l G_{l+1}/G_{l+1})$. If we perform the Taylor expansion

$$\gamma'_{0,1}(t) = g_0 g_1^t \dots g_k^{t^k/k!} \tag{126}$$

for $g_0, \dots, g_k \in G_l$, then on setting $t = 0$ we conclude that $g_0 \in \Gamma_l$; also, by taking repeated differences with spacing $q_{l+1}$, we see that for each $m = 1, \dots, k$ we have $g_m^{a_m} \in \Gamma_l G_{l+1}$ for some positive integer $a_m$ of polynomial size. Note that $g_1, \dots, g_m$ do not depend on the choice of coset representative $\gamma_{0,1}$. If we let $n_m$ be the least positive integer such that $g_m^{n_m q_l^m/m!} \in \Gamma G_{l+1}$, we see that each $n_m$ is of polynomial size and

$$\gamma'_{0,1} \bmod G_{l+1} \in \mathrm{Poly}(k! n_1 \dots n_k q_l \mathbb{Z} \to \Gamma_l G_{l+1}/G_{l+1})$$

and thus

$$n_1 \dots n_k q_l \gg q_{l+1}$$

so by (124) and the pigeonhole principle we can find $m = 1, \dots, k$ such that

$$n_m \gg q_0^{1/k^2}. \tag{127}$$

Henceforth we fix this $m$. We will shortly use this large integer $n_m$ as a modulus to which one can apply Lemma 3.12. A key technical point is that this modulus does not depend on the tuples in $\mathcal{T}$.

Next, we claim that after removing a negligible fraction of tuples from the family $\mathcal{T}$, we may assume that none of the $p'_{j,i}$ divide $n_m$. For sake of notation let us just remove the contribution where $p'_{0,1}$ divides $n_m$. There are $O(N)$ choices for $I''_{0,1}$. As $n_m$ is of polynomial size and $p'_{0,1} \in [P', 2P']$, we see that there are only $O(1)$ choices for $p'_{0,1}$. After fixing this choice, there are at most $O(\pi_0(P')^{2\ell-1}) = O(d^{\ell-1/2})$ choices for the remaining choices of $p'_{j,4}, p'_{j,1}$, $j = 0, \dots, \ell-1$. Then we see from (122) and the fundamental theorem

of arithmetic that there are $O(\frac{1}{N}(P')^{2\ell}) = O(d^{\ell+o(1)}/N)$ choices for the $p'_{j,3}, p'_{j,2}$. After making all these choices, the tuple (121) is fixed, so the total number of tuples generated in this fashion is $O(d^{2\ell-1/2+o(1)}/N)$, which is negligible. Similarly for the cases when some other prime $p'_{j,i}$ divides $n_m$.

For each $0 \leq j \leq \ell$ and $i = 1, 2$, let $\gamma_{j,i} \in \mathrm{Poly}(\mathbb{R} \to G)$ be a representative of the coset $f_{I''_{j,i}} \in \Gamma\backslash\mathrm{Poly}(\mathbb{R} \to G)$, thus $f_{I''_{j,i}} = \Gamma\gamma_{j,i}$; for $(j,i) = (0,1)$ we use the same choice $\gamma_{0,1}$ of coset representative that was made earlier. From (104) we have for all $0 \leq j \leq \ell - 1$ that

$$\gamma_{j,1}(p'_{j,2}\cdot) = \gamma_j \gamma_{j,2}(p'_{j,1}\cdot)\gamma_j^\dagger$$

for some $\gamma_j \in \Gamma$, and some $\gamma_j^\dagger \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$, and similarly

$$\gamma_{j+1,1}(p'_{j,4}\cdot) = \tilde{\gamma}_j \gamma_{j,2}(p'_{j,3}\cdot)\tilde{\gamma}_j^\dagger$$

for all $t \in \mathbb{R}$ and some $\tilde{\gamma}_j \in \Gamma$, and some $\tilde{\gamma}_j^\dagger \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$. Concatenating these estimates, we conclude that

$$\gamma_{0,1}(a\cdot) = \gamma\gamma_{0,1}(b\cdot)\gamma^\dagger$$

for some $\gamma \in \Gamma$ and $\gamma^\dagger \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$. By (125), this implies that

$$\gamma'_{0,1}(a\cdot) = \gamma\gamma'_{0,1}(b\cdot)\gamma^{-1}\tilde{\gamma}^\dagger$$

where $\tilde{\gamma}^\dagger \in \mathrm{Poly}(q_l\mathbb{Z} \to \Gamma)$. Since $\gamma'_{0,1}(a\cdot)$ and $\gamma\gamma'_{0,1}(b\cdot)\gamma^{-1}$ both take values in $G_l$, $\tilde{\gamma}^\dagger$ does also, thus $\tilde{\gamma}^\dagger \in \mathrm{Poly}(q_l\mathbb{Z} \to \Gamma_l)$. If we now project to the torus $G_l/(\Gamma_l G_{l+1})$, we see that

$$\gamma'_{0,1}(aq_l n) = \gamma'_{0,1}(bq_l n) \bmod \Gamma_l G_{l+1}$$

for all integers $n$. Using the Taylor expansion (126), we conclude on taking $m$ divided differences with spacing $q_l$ at the origin that

$$g_m^{(aq_l)^m} = g_m^{(bq_l)^m} \bmod \Gamma_l G_{l+1}$$

and hence by definition of $n_m$

$$a^m = b^m \bmod n_m.$$

Applying Lemma 3.12, we can then bound the number $\#\mathcal{T}$ of tuples as

$$\#\mathcal{T} \ll \frac{d^{2\ell}}{N}\left(\frac{k^{\omega(n_m)}}{\phi(n_m)} + \frac{1}{\log N}\right)$$

which by the divisor bound and (127) gives

$$\#\mathcal{T} \ll q_0^{-\frac{1}{2k^2}}\frac{d^{2\ell}}{N}$$

which contradicts the lower bound (123) if $q_0$ is large enough. □

Note that each $I''_1$ appears in at most $O(d)$ quadruples $e = (I''_1, I''_2, p'_1, p'_2) \in \mathcal{Q}''$. Combining this observation with Propositions 4.25, 4.22, we conclude

**Corollary 4.26.** *For all $I''$ in a large subcollection $\mathcal{I}'''$ of $\mathcal{I}''$, we can find a representation*

$$F(g_{I''}\Gamma) = F(\tilde{g}_{I''}\gamma_{I''}\Gamma) \tag{128}$$

*for some $\tilde{g}_{I''}, \gamma_{I''} \in \mathrm{Poly}(\mathbb{R} \to G)$, and $T_{I''} \in G$ of polynomial size such that*

(i) *The map*

$$t \mapsto \log(\tilde{g}_{I''}(t) T_{I''}^{-\log(t/x_{I''})}) \tag{129}$$

*is bounded by $O(1)$ and has a Lipschitz norm of $O(|I''|^{-1})$ whenever $\langle t \rangle_{I''} \ll 1$.*

(ii) *There is a non-trivial horizontal character $\eta_{I''} \colon G \to \mathbb{R}/\mathbb{Z}$ such that the derivative $d\eta_{I''} \colon \log G \to \mathbb{R}$ has operator norm $O(1)$, and such that*

$$d\eta_{I''}(\log T_{I''}) = O(N). \tag{130}$$

(iii) *$\gamma_{I''} \in \mathrm{Poly}(q_{I''}\mathbb{Z} \to \Gamma)$ for some $q_{I''} = O(1)$.*

Let $I'', \tilde{g}_{I''}, \gamma_{I''}, T_{I''}, \eta_{I''}, q_{I''}$ be as in the above corollary. Observe that as the number of possible $q_{I''}$ is bounded, we may refine the family $\mathcal{I}'''$ of intervals in the above corollary by a bounded factor to assume that

$$q_{I''} = q$$

is independent of $I''$ (one could also simply clear denominators here). In a similar spirit, as $\eta_{I''}$ takes values in the lattice of horizontal characters (which one can identify with the Pontryagin dual of the torus $G/\Gamma[G,G]$) and is a bounded distance away from the origin, there are only finitely many choices for $\eta_{I''}$, so we may assume that

$$\eta_{I''} = \eta$$

is independent of $I''$.

Now we will be able to descend from $G$ to the lower-dimensional nilpotent group $\ker(\eta)$ as follows. Since $\eta \colon G \to \mathbb{R}/\mathbb{Z}$ is a homomorphism to the abelian group $\mathbb{R}/\mathbb{Z}$, it annihilates the commutator group $[G,G]$, and hence (by (177)) the derivative map $d\eta \colon \log G \to \mathbb{R}$ annihilates the commutator algebra $[\log G, \log G]$. In particular, from the Baker–Campbell–Hausdorff formula, we have

$$d\eta\left(\log\left(\tilde{g}_{I''}(t) T_{I''}^{-\log(t/x_{I''})}\right)\right) = d\eta(\log \tilde{g}_{I''}(t)) - \log(t/x_{I''}) d\eta(\log T_{I''}).$$

If we then apply $d\eta$ to (129), we conclude that the map

$$t \mapsto d\eta(\log \tilde{g}_{I''}(t)) - \log(t/x_{I''}) d\eta(\log T_{I''})$$

has a Lipschitz norm of $O(|I''|^{-1})$ whenever $\langle t \rangle_{I''} \ll 1$. Combining this with (130), we see that the map

$$t \mapsto d\eta(\log \tilde{g}_{I''}(t))$$

also has a Lipschitz norm of $O(|I''|^{-1})$ in this region. From the definition of $\mathrm{Poly}(\mathbb{R} \to G)$, this map is also a polynomial of degree $k$, with the $t^j$ coefficient lying in $d\eta(\log G_j)$ for each $j \geq 0$. By the Bernstein inequality (27), we may thus write

$$d\eta(\log \tilde{g}_{I''}(t)) = \sum_{j=0}^{k} \theta_j (t - x_{I''})^j$$

where the $\theta_j$ are real numbers with $\theta_j \in d\eta(\log G_j)$ and $\theta_j = O(|I''|^{-j})$. Lifting this polynomial back to $G$, we may thus write

$$\log \tilde{g}_{I''}(t) = \sum_{j=0}^{k} X_j (t - x_{I''})^j \bmod \ker(d\eta)$$

for some $X_j \in \log G_j$ with $X_j = O(|I''|^{-j})$. If we set

$$\varepsilon_{I''}(t) \coloneqq \exp(\sum_{j=0}^{k} X_j (t - x_{I''})^j)$$

then $\varepsilon_{I''} \in \mathrm{Poly}(\mathbb{R} \to G)$ is smooth on $I''$, and if we then define $g^*_{I''} \colon \mathbb{R} \to G$ to be the map for which

$$\tilde{g}_{I''}(t) = \varepsilon_{I''}(t) g^*_{I''}(t)$$

then from the Baker–Campbell–Hausdorff formula (176) we see that $g^*_{I''} \in \mathrm{Poly}(\mathbb{R} \to \ker(\eta))$ takes values in the kernel $\ker(\eta) = \exp(\ker(d\eta))$ of $G$, which is a proper rational normal subgroup of $G$. By (128), (36) we then have

$$\left| \sum_{n \in I''} f(n) \overline{F}(\varepsilon_{I''}(n) g^*_{I''}(n) \gamma_{I''}(n) \Gamma) \right| \gg |I''|.$$

Let $H^* \coloneqq c \frac{H}{P'P''}$ for a sufficiently small absolute constant $c > 0$. Then we have

$$\int_{I''} \left| \sum_{n \in [x, x+H^*]} f(n) \overline{F}(\varepsilon_{I''}(n) g^*_{I''}(n) \gamma_{I''}(n) \Gamma) \right| \, dx \gg |I''| H^*.$$

As $\varepsilon_{I''}$ is smooth on $I''$, $\varepsilon_{I''}(n)$ is $O(1)$ and varies by at most $O(c)$ on $[x, x + H^*]$, hence by the Lipschitz nature of $F$

$$\int_{I''} \left| \sum_{n \in [x, x+H^*]} f(n) \overline{F}(\varepsilon_{I''}(x) g^*_{I''}(n) \gamma_{I''}(n) \Gamma) \right| \, dx \gg |I''| H^*.$$

Summing over $I'' \in \mathcal{I}'''$, we conclude that

$$\int_X^{2X} \sup_{\varepsilon \in E; \tilde{g} \in \mathrm{Poly}(\mathbb{Z} \to \ker(\eta)); \gamma \in \mathrm{Poly}(q\mathbb{Z} \to \Gamma)} \left| \sum_{n \in [x, x+H]} f(n) \overline{F}(\varepsilon \tilde{g}(n) \gamma(n) \Gamma) \right| \, dx \gg HX$$

for some compact set $E \subset G$. But this contradicts Proposition 4.4. This contradiction (finally!) concludes the proof of Theorem 1.5.

**Remark 4.27.** *It seems plausible that the proof of Theorem 1.5, combined with the quantitative work in Section 6 for lowering the value of $H$, would allow lowering the length of the intervals to $H \geq \exp((\log X)^{1-\delta})$ for some $\delta > 0$. We do not pursue this further here, however, as that would further lengthen this paper. Let us note, however, that at least the convenient notion of polynomially large elements in Lie groups used in this section would*

*have to be replaced with a more cumbersome notation in the case where $H$ is no longer polynomially large in terms of $X$.*

## 5. Sign patterns

5.1. **The Liouville case.** Our main goal in this section is to use Theorem 1.5 to prove Theorem 1.9, which asserts a superpolynomial lower bound on the number $s(k)$ of sign patterns of the Liouville function, defined in (14). We will also prove a generalization of Theorem 1.9 to sign patterns of other multiplicative functions (Theorem 5.4), and prove Proposition 1.7.

Regarding Theorem 1.9, we will in fact prove a more general implication, which gives a lower bound on $s(k)$ whenever one has local Gowers uniformity of the Liouville function on short intervals:

**Theorem 5.1** (From local Gowers uniformity to lower bounds on sign patterns)**.** *Let $0 < \kappa < 1/2$. Let $\Psi : \mathbb{R}_{\geq 1} \to \mathbb{R}$ be a strictly increasing function with $X \leq \Psi(X) \leq \exp(X^{1/2-\kappa})$ for all large enough $X$. Suppose that* (10) *holds for $H(X) = \Psi^{-1}(X^\eta)$ for every fixed $\eta > 0$. Then $s(k) \geq \Psi(k)$ for all large enough $k$.*

Taking $\Psi(X) = X^A$ and applying Theorem 1.5, we see that Theorem 1.9 follows directly from the above theorem. Furthermore, we have the following conditional corollary.

**Corollary 5.2.** *Let $\varepsilon > 0$. Assuming that* (10) *holds with $H(X) = \exp((\log X)^{1-\delta})$ for some $\delta \in (0,1)$, we have $s(k) \gg_\varepsilon k^{(\log k)^{\delta/(1-\delta)-\varepsilon}}$. Further, assuming* (10) *with $H(X) = (\log X)^C$ for some $C > 2$, we have $s(k) \gg_\varepsilon \exp(k^{1/C-\varepsilon})$.*

**Remark 5.3.** *In the proof of Theorem 5.1 below, one may on first reading want to assume that $\Psi(X) = X^A$, which corresponds to $H(X) = X^{o(1)}$, in which case we wish to show that $s(k) \gg_A k^A$ for all $A$. This simplifies various expressions involved; in particular expressions involving $\Psi$ are just large powers of the argument and expressions involving $\Psi^{-1}$ are small powers of the argument.*

We now begin the proof of Theorem 5.1. Fix $\kappa > 0$; we allow all implied constants to depend on $\kappa$. Suppose for the sake of contradiction that $s(m) < \Psi(m)$ for infinitely many $m$. We will use this to show that $s(k) = 2^k$ for all $k$. Since $\Psi(k) < 2^k$ for all sufficiently large $k$, this will give the required contradiction.

Fix $k$; we now allow all implied constants to depend on $k$. We now select additional parameters $\varepsilon, w, m, R, x$, arranged so that

$$k \ll \frac{1}{\varepsilon} \ll w \ll m \ll R \ll x,$$

by the following scheme.

- First, we choose $\varepsilon > 0$ to be a sufficiently small quantity depending on $k, \kappa$.
- Then we choose a quantity $w > 1$ to be sufficiently large depending on $\varepsilon, k, \kappa$.
- Next, we choose $m$ to be a natural number with $s(m) < \Psi(m)$ that is sufficiently large (depending on $w, \varepsilon, k, \kappa$). Such an $m$ always exists by hypothesis.

- One then sets $R := \Psi(m)^{\varepsilon^{-2}}$ and $x := \Psi(m)^{\varepsilon^{-3}}$.

By construction and the hypothesis $X \leq \Psi(X) \leq \exp(X^{1/2-\kappa})$, we have $R = x^{\varepsilon}$,

$$(\log x)^{2+\kappa} \leq m \leq x^{\varepsilon^3}, \tag{131}$$

and

$$s(m) < x^{\varepsilon^3}. \tag{132}$$

Now suppose for contradiction that $s(k) < 2^k$. Then by (14) there exists a sign pattern $(\varepsilon_1, \dots, \varepsilon_k) \in \{-1, +1\}^k$ which never occurs in the Liouville sequence, so in particular

$$\mathbb{E}^{\log}_{n \leq x} 1_{\lambda(n+1)=\varepsilon_1} \cdots 1_{\lambda(n+k)=\varepsilon_k} = 0. \tag{133}$$

Writing $1_{\lambda(n+j)=\varepsilon_j} = \frac{1+\varepsilon_j \lambda(n+j)}{2}$, we may expand the left-hand side of (133) as the sum of the $2^k$ quantities of the form

$$(\prod_{l=1}^{i} \varepsilon_{\ell_l}) 2^{-k} \mathbb{E}^{\log}_{n \leq x} \lambda(n+\ell_1) \cdots \lambda(n+\ell_i), \quad \text{where} \quad \{\ell_1, \dots, \ell_i\} \subset \{1, 2, \dots, k\}.$$

The $i = 0$ term is equal to $2^{-k}$. Thus by the pigeonhole principle, there must exist $1 \leq i \leq k$ and $1 \leq \ell_1 < \ldots < \ell_i \leq k$ for which the correlation

$$C := \mathbb{E}^{\log}_{n \leq x} \lambda(n+\ell_1) \cdots \lambda(n+\ell_i) \tag{134}$$

is such that

$$|C| \gg 1. \tag{135}$$

The precise choice of $i, \ell_1, \dots, \ell_i$ may depend on $x$, but this will not concern us. Henceforth let $i, \ell_1, \dots, \ell_i$ be chosen so that (135) holds.

Set $P := \frac{m}{3k}$. By using the multiplicativity relation $\lambda(pn) = -\lambda(n)$ and the fact that the correlation $C$ in (134) involves a logarithmic average, for all primes $p \leq 2P$ we deduce

$$\begin{aligned} C &= (-1)^i \mathbb{E}^{\log}_{n \leq x} \lambda(pn + p\ell_1) \cdots \lambda(pn + p\ell_i) \\ &= (-1)^i \mathbb{E}^{\log}_{n' \leq px} \lambda(n' + p\ell_1) \cdots \lambda(n' + p\ell_i) p1_{p|n'} + O(\varepsilon^3) \\ &= (-1)^i \mathbb{E}^{\log}_{n' \leq x} \lambda(n' + p\ell_1) \cdots \lambda(n' + p\ell_i) p1_{p|n'} + O(\varepsilon^3), \end{aligned}$$

where the final estimate follows from (131). Hence, by averaging over $p$,

$$C = (-1)^i \mathbb{E}_{P \leq p < 2P} \mathbb{E}^{\log}_{n \leq x} \lambda(n + p\ell_1) \cdots \lambda(n + p\ell_i) p1_{p|n} + O(\varepsilon^3).$$

The contribution of $n \leq R = x^{\varepsilon}$ to the average is trivially $\ll \varepsilon$, so

$$C = (-1)^i \mathbb{E}_{P \leq p < 2P} \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + p\ell_1) \cdots \lambda(n + p\ell_i) p1_{p|n} + O(\varepsilon), \tag{136}$$

We will shortly exploit the sign pattern bound (132) to obtain the bound

$$\mathbb{E}_{P \leq p < 2P} \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + p\ell_1) \cdots \lambda(n + p\ell_i)(p1_{p|n} - 1) \ll \varepsilon. \tag{137}$$

Assuming this bound for the moment, we may then simplify (136) to

$$C = (-1)^i \mathbb{E}_{P \leq p < 2P} \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + p\ell_1) \cdots \lambda(n + p\ell_i) + O(\varepsilon).$$

For $d \in [P, 2P]$, the von Mangoldt function $\Lambda(d)$ is equal to $(1 + O(\varepsilon)) \log P$ when $d$ is prime and is only nonzero (and of size $O(\log P)$) for $O(P^{1/2+\varepsilon})$ other values of $d$. Since $P$ is large compared to $\varepsilon$, we easily conclude that

$$C = (-1)^i \mathbb{E}_{P \leq d < 2P} \Lambda(d) \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + d\ell_1) \cdots \lambda(n + d\ell_i) + O(\varepsilon)$$

We now apply the "$W$-trick". If we set $W := \prod_{p \leq w} p$ and split $d$ into residue classes $b$ mod $W$, then the contribution of the non-primitive classes $(b, W) > 1$ is negligible, and we have

$$C = (-1)^i \mathbb{E}_{\substack{1 \leq b \leq W \\ (b,W)=1}} \mathbb{E}_{P/W \leq d < 2P/W} \Lambda_{W,b}(d) \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + (Wd+b)\ell_1) \cdots \lambda(n + (Wd+b)\ell_i) + O(\varepsilon) \tag{138}$$

where $\Lambda_{W,b}(d) := \frac{\phi(W)}{W} \Lambda(Wd + b)$, and $\phi$ is the Euler totient function. By splitting the average over $n$ into intervals of length $P/W$ and applying the Gowers uniformity of $\Lambda_{W,b}(d) - 1$ (established in [14], [15], [17]) as in [35, Proposition 3.3], we find

$$\mathbb{E}_{P/W \leq d < 2P/W} (\Lambda_{W,b}(d) - 1) \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + (Wd + b)\ell_1) \cdots \lambda(n + (Wd + b)\ell_i) \ll \varepsilon$$

for any $b \in (\mathbb{Z}/W\mathbb{Z})^\times$ (here we use the fact that $P$ is large compared to $W, \varepsilon$). We conclude that

$$C = (-1)^i \mathbb{E}_{\substack{1 \leq b \leq W \\ (b,W)=1}} \mathbb{E}_{P/W \leq d < 2P/W} \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + (Wd + b)\ell_1) \cdots \lambda(n + (Wd + b)\ell_i) + O(\varepsilon),$$

or equivalently

$$C = (-1)^i \frac{W}{\phi(W)} \mathbb{E}_{P \leq d < 2P} 1_{(d,W)=1} \mathbb{E}^{\log}_{R \leq n \leq x} \lambda(n + d\ell_1) \cdots \lambda(n + d\ell_i) + O(\varepsilon).$$

Splitting the $n$ sum into intervals of length $m = 3kP$ and using the triangle inequality, we obtain

$$C \ll \frac{W}{\phi(W)} \mathbb{E}_{P \leq d < 2P} \mathbb{E}^{\log}_{n \leq x} |\mathbb{E}_{n \leq n' \leq n+m} \lambda(n' + d\ell_1) \cdots \lambda(n' + d\ell_i)| + \varepsilon.$$

Embedding $[n, n + m]$ into a cyclic group of prime order, and applying the generalized von Neumann theorem in the form of [14, Proposition 7.1], we have

$$\frac{W}{\phi(W)} \mathbb{E}_{P \leq d < 2P} |\mathbb{E}_{n \leq n' \leq n+m} \lambda(n' + d\ell_1) \cdots \lambda(n' + d\ell_i)| \ll O_W(\kappa(\|\lambda\|_{U^k[n,n+m]})) + \varepsilon$$

for some bounded function $\kappa(x)$ tending to 0 as $x \to 0$, and so we conclude that

$$C \ll O_W(\mathbb{E}^{\log}_{n \leq x} \kappa(\|\lambda\|_{U^k[n,n+m]})) + \varepsilon. \tag{139}$$

Since $m = \Psi^{-1}(x^{\varepsilon^3})$, we conclude from the assumption of the theorem (and the fact that $x$ is sufficiently large depending on $w, k, \varepsilon$) that

$$C \ll \varepsilon,$$

but this contradicts (135) for $\varepsilon$ small enough.

To conclude the proof of Theorem 5.1, it remains to establish the bound (137). This is reminiscent of the bounds one can establish by entropy decrement arguments as seen for

instance in [35]; however the size of $P$ compared to $x$ is too large here for such methods to apply (and furthermore these methods need to exclude an exceptional set of bad scales $P$). The key observation is that one can instead exploit the small number (132) of sign patterns of length $m = 3kP$ to obtain a strong estimate via the moment method. Firstly, by approximate translation invariance we can write

$$\mathbb{E}_{P\leq p<2P}\mathbb{E}^{\log}_{R\leq n\leq x}\lambda(n+p\ell_1)\cdots\lambda(n+p\ell_i)(p1_{p|n}-1)$$

as

$$\mathbb{E}_{P\leq p<2P}\mathbb{E}^{\log}_{R\leq n\leq x}\lambda(n+j+p\ell_1)\cdots\lambda(n+j+p\ell_i)(p1_{p|n+j}-1)+O(\varepsilon)$$

for any $1\leq j\leq P$, thus on averaging we may also write it as

$$\mathbb{E}^{\log}_{R\leq n\leq x}\mathbb{E}_{P\leq p<2P}\mathbb{E}_{j\leq P}\lambda(n+j+p\ell_1)\cdots\lambda(n+j+p\ell_i)(p1_{p|n+j}-1)+O(\varepsilon).$$

Thus by the triangle inequality, it suffices to show that

$$\mathbb{E}^{\log}_{R\leq n\leq x}\left|\mathbb{E}_{P\leq p<2P}\mathbb{E}_{j\leq P}\lambda(n+j+p\ell_1)\cdots\lambda(n+j+p\ell_i)(p1_{p|n+j}-1)\right|\ll\varepsilon.$$

By the triangle inequality, the quantity inside the absolute values is bounded by $O(1)$. Thus it will suffice to establish the probability bound

$$\mathbb{P}^{\log}_{R\leq n\leq x}\left(\left|\mathbb{E}_{P\leq p<2P}\mathbb{E}_{j\leq P}\lambda(n+j+p\ell_1)\cdots\lambda(n+j+p\ell_i)(p1_{p|n+j}-1)\right|\geq\varepsilon\right)\ll\varepsilon$$

where $\mathbb{P}^{\log}_{R\leq n\leq x}(A)\coloneqq\mathbb{E}^{\log}_{R\leq n\leq x}1_A(n)$ is the probability measure associated to the averaging operator $\mathbb{E}^{\log}_{R\leq n\leq x}$.

Observe that the numbers $n+j+p\ell_l$ that appear in this expression all lie in the interval $\{n+1,\dots,n+m\}$. By (132), there are at most $x^{\varepsilon^3}$ possible choices for the sign pattern $(\lambda(n+1),\dots,\lambda(n+m))$. Thus, by the union bound, it will suffice to show that

$$\mathbb{P}^{\log}_{R\leq n\leq x}\left(\left|\mathbb{E}_{P\leq p<2P}\mathbb{E}_{j\leq P}a_{j+p\ell_1}\cdots a_{j+p\ell_i}(p1_{p|n+j}-1)\right|\geq\varepsilon\right)\ll\varepsilon x^{-\varepsilon^3}\tag{140}$$

for each choice of sign pattern $(a_1,\dots,a_m)\in\{-1,+1\}^m$.

Fix $a_1,\dots,a_m$. Let $2r$ be the largest even integer such that $P^{2r}\leq x^{\varepsilon^2}$. From (131) and the definition $P=m/(3k)$ we observe the estimates

$$\frac{1}{\varepsilon}\ll r\asymp\varepsilon^2\frac{\log x}{\log P}\ll\varepsilon^2\frac{\log x}{\log\log x}.\tag{141}$$

From Markov's inequality we may bound the left-hand side of (140) by

$$\varepsilon^{-2r}\mathbb{E}^{\log}_{R\leq n\leq x}\left|\mathbb{E}_{P\leq p<2P}\mathbb{E}_{j\leq P}a_{j+p\ell_1}\cdots a_{j+p\ell_i}(p1_{p|n+j}-1)\right|^{2r}$$

which by expanding out the $2r^{\text{th}}$ power and applying the triangle inequality is bounded by

$$\varepsilon^{-2r}\mathbb{E}_{P\leq p_1,\dots,p_{2r}<2P}\mathbb{E}_{j_1,\dots,j_{2r}\leq P}|\mathbb{E}^{\log}_{R\leq n\leq x}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r})|$$

where $\xi_p(n)\coloneqq p1_{p|n}-1$. From (141) we have $\varepsilon^{2r+1}\gg x^{-\varepsilon^3}$, so it will thus suffice to establish the estimate

$$\mathbb{E}_{P\leq p_1,\dots,p_{2r}<2P}\mathbb{E}_{j_1,\dots,j_{2r}\leq P}|\mathbb{E}^{\log}_{R\leq n\leq x}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r})|\ll x^{-2\varepsilon^3}.\tag{142}$$

For any given $p_1, \dots, p_{2r}, j_1, \dots, j_{2r}$, the function $n \mapsto \xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r})$ is periodic of period $Q := p_1 \dots p_{2r}$ and has magnitude at most $Q$. We have

$$\mathbb{E}^{\log}_{R\leq n\leq x}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r}) = \mathbb{E}^{\log}_{R\leq n\leq x}\xi_{p_1}(n+h+j_1)\dots\xi_{p_{2r}}(n+h+j_{2r}) + O\left(\frac{Q^2}{R\log x}\right)$$

for any $1 \leq h \leq Q$. Averaging in $h$ and using the periodicity, we conclude that

$$\mathbb{E}^{\log}_{R\leq n\leq x}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r}) = \mathbb{E}_{n\in\mathbb{Z}/Q\mathbb{Z}}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r}) + O\left(\frac{Q^2}{R\log x}\right)$$

where we view $\xi_{p_1}, \dots, \xi_{p_{2r}}$ as functions on $\mathbb{Z}/Q\mathbb{Z}$ in the obvious fashion. Since

$$Q^2 \leq (2P)^{4r} \leq 2^{4r}x^{2\varepsilon^2} \ll x^{3\varepsilon^2}$$

(by (141)) and $R = x^\varepsilon$, we see that the $Q^2/(R\log x)$ error is negligible. Thus it will suffice to show that

$$\mathbb{E}_{P\leq p_1,\dots,p_{2r}<2P}\mathbb{E}_{j_1,\dots,j_{2r}\leq P}|\mathbb{E}_{n\in\mathbb{Z}/Q\mathbb{Z}}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r})| \ll x^{-2\varepsilon^3}. \tag{143}$$

If one of the primes $p_i$ is distinct from all the others, then the inner average $\mathbb{E}_{n\in\mathbb{Z}/Q\mathbb{Z}}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r})$ vanishes from the Chinese remainder theorem, since $\xi_{p_i}(n+j_i)$ is periodic with mean zero with period $p_i$, and all other factors have period coprime to $p_i$. Thus we may restrict attention to those tuples $(p_1, \dots, p_{2r})$ in which each prime $p_i$ appears at least twice, hence there are at most $r$ distinct primes in this tuple. The number of such tuples can then be bounded crudely by $O(r^2\pi_0(P))^r$, by first selecting $r$ primes in $[P, 2P]$ (for which there are $O(\pi_0(P))^r$ choices), and then assigning each $p_1, \dots, p_{2r}$ to one of these primes (for which there are $r^{2r}$ choices). Thus the proportion of such tuples amongst all primes $P \leq p_1, \dots, p_{2r} < 2P$ is $O(r^2\pi_0(P)^{-1})^r$. If $(p_1, \dots, p_{2r})$ is such a tuple, then from the triangle inequality one has

$$\begin{aligned}
&\mathbb{E}_{j_1,\dots,j_{2r}\leq P}|\mathbb{E}_{n\in\mathbb{Z}/Q\mathbb{Z}}\xi_{p_1}(n+j_1)\dots\xi_{p_{2r}}(n+j_{2r})| \\
&\leq \mathbb{E}_{n\in\mathbb{Z}/Q\mathbb{Z}}\mathbb{E}_{j_1,\dots,j_{2r}\leq P}|\xi_{p_1}(n+j_1)|\dots|\xi_{p_{2r}}(n+j_{2r})| \\
&= \mathbb{E}_{n\in\mathbb{Z}/Q\mathbb{Z}}\prod_{i=1}^{2r}\mathbb{E}_{j\leq P}|\xi_{p_i}(n+j)| \\
&\leq O(1)^r
\end{aligned}$$

since $\mathbb{E}_{j\leq P}|\xi_{p_i}(n+j)| \ll 1$ for any $i$. Thus we can bound the left-hand side of (143) by $O(r^2\pi_0(P)^{-1})^r$. But from (141), (131) we have $r^2\pi_0(P)^{-1} \ll P^{-c}$ for some $c > 0$ depending only on $\kappa$, hence by (141) the left-hand side of (143) is $O(x^{-c'\varepsilon^2})$ for some $c' > 0$ depending on $\kappa$, and the claim follows. This concludes the proof of Theorem 5.1.

5.2. **Generalization to other multiplicative functions.** The above proof can be generalized to produce a result about patterns in more general multiplicative functions.

**Theorem 5.4.** *Let $g : \mathbb{N} \to \mu_\ell$ be a multiplicative function taking values in the roots of unity of order $\ell \geq 2$, and suppose that $\mathbb{D}(g^j, \chi; X) \xrightarrow{X\to\infty} \infty$ for all Dirichlet characters $\chi$ and for all $1 \leq j \leq \ell - 1$. Then the number*

$$s_g(k) := \{v \in \mu_\ell^k : \ v = (g(n+1), \dots, g(n+k)) \text{ for some } n \in \mathbb{N}\}$$

*of value patterns of $g$ of length $k$ satisfies $s_g(k) \gg_A k^A$.*

We remark that a similar result holds (with essentially the same proof) for any 1-bounded multiplicative function $g : \mathbb{N} \to \mathbb{C}$ such that $\inf_{|t|\leq X^{k+1}} \mathbb{D}(g^j, \chi(n)n^{it}; X) \xrightarrow{X\to\infty} \infty$ for all $j \geq 1$. In this case, the "sign patterns" would be defined as occurrences of a pattern $(g(n+1), \dots, g(n+k)) \in I_1 \times \cdots \times I_k$, where $I_i$ are arcs of the unit circle of the form $[e(m_i/\ell), e((m_i+1)/\ell)]$ with $0 \leq m_i \leq \ell - 1$. We leave the details of this generalization to the interested reader.

*Proof.* (Sketch) The proof follows along similar lines as that of Theorem 5.1. We assume for the sake of contradiction that $s_g(m) \leq m^A$ for infinitely many $m$ and aim to deduce that

$$C := \mathbb{E}^{\log}_{n\leq x} g^{a_1}(n+\ell_1) \cdots g^{a_j}(n+\ell_j) = o(1) \tag{144}$$

for any nonempty set $\{\ell_1, \dots, \ell_j\} \subset \{1, 2, \dots, k\}$ with the $\ell_i$ distinct, and for any integers $a_1, \dots, a_j \in [1, \ell - 1]$. Once we have proved (144), we use the expansion

$$1_{g(n)=e(a/\ell)} = \frac{1}{\ell} \sum_{j=0}^{\ell-1} g(n)^j e(-\frac{aj}{\ell})$$

for the indicator functions of the level sets to obtain $s_g(k) = \ell^k$ for any $k$, which gives the desired contradiction.

The main difficulty[19] is that the factor $(-1)^i$ that appeared in the proof of Theorem 5.1 must now be replaced by $g(p)^{-a_1-\cdots-a_j}$. One can still repeat the proof of Theorem 5.1 with obvious modifications down to (138), where the right-hand side is now up to $O(\varepsilon)$ equal to

$$\mathbb{E}_{\substack{1\leq b\leq W\\(b,W)=1}} \mathbb{E}_{P/W\leq d<2P/W} g(d)^{-a_1-\cdots-a_j} \Lambda_{W,b}(d) \mathbb{E}^{\log}_{R\leq n\leq x} g^{a_1}(n+(Wd+b)\ell_1) \cdots g^{a_j}(n+(Wd+b)\ell_j).$$

The weight $g(d)^{-a_1-\cdots-a_j}$ now prevents one from applying the Gowers uniformity theory for the von Mangoldt function [14], [15], [17]. However, the function $g(d)^{-a_1-\cdots-a_j}\Lambda_{W,b}(d)$ is still dominated pointwise by $\Lambda_{W,b}(d)$, which is a pseudorandom majorant in the sense of [14]. One can then apply the generalized von Neumann theorem (essentially in the form of [14, Proposition 7.1]), and reduce matters to showing that

$$\mathbb{E}^{\log}_{n\leq x} \|g^j\|_{U^{k-1}[n,n+m]} = o(1)$$

[19] A much more minor difficulty is that $g$ is now only assumed to be multiplicative rather than completely multiplicative, so that the identity $g(n) = g(p)^{-1}g(pn)$ only holds when $n$ is not divisible by $p$. However, as we will be working with moderately large primes $p$, the contribution of those $n$ which are divisible by $p$ can easily be seen to be negligible.

whenever $1 \leq j \leq \ell - 1$ and $m \gg x^\theta$ for some $\theta > 0$. This Gowers norm bound then follows from Theorem 1.5, once we show that $M(f; x^{k+1}, Q) \to \infty$ as $x \to \infty$ for any given $k$ and $Q$. By [21, Lemma 3.1] (which is a pretentious triangle inequality argument), and the fact that $\mathbb{D}(f, g; x) = \mathbb{D}(f, g; x^{k+1}) + O_k(1)$, we have

$$M(g; x^{k+1}, Q) \geq \inf_{\substack{\chi \bmod q \\ q \leq Q \\ |t| \leq x^{k+1}}} \mathbb{D}(g\overline{\chi}, n \mapsto n^{it}; x) \geq \frac{1}{2kQ} \min\{(\log\log x)^{1/2}, \mathbb{D}(g\overline{\chi}, 1; x)\} - O_{k,Q}(1),$$

and the right-hand side is tending to infinity with $x$ by assumption. This completes the proof. □

5.3. **Uniformity at very small scales.** We now give a proof of Proposition 1.7 that states that the estimate (11) at scale $H = (\log x)^\eta$ is enough to deduce the logarithmic Chowla conjecture (and hence in fact (11) for *any* $H = H(X)$ tending to infinity, thanks to the results in [35]).

*Proof of Proposition 1.7.* Let $k$ be a natural number, and let be $h_1, \dots, h_k$ given shifts. Let $x$ be large enough, and denote the correlation along these shifts by

$$C := \mathbb{E}^{\log}_{n \leq x} \lambda(n + h_1) \cdots \lambda(n + h_k).$$

For any fixed $\varepsilon > 0$, we wish to show that $|C| \ll \varepsilon$. We begin by applying the entropy decrement argument in the slightly refined form given in [36, Theorem 3.1] (the original argument from [34] is able to locate a good scale on any interval $I$ with $\sum_{m \in I} \frac{1}{m \log m} \gg \varepsilon^{-10}$, whereas the refined one is able to locate a good scale on any interval with $\sum_{m \in I} \frac{1}{m} \gg \varepsilon^{-10}$).

By [36, Theorem 3.1], we deduce that

$$C = (-1)^k \mathbb{E}_{2^m \leq p \leq 2^{m+1}} \mathbb{E}^{\log}_{n \leq x} \lambda(n + ph_1) \cdots \lambda(n + ph_k) + O(\varepsilon) \tag{145}$$

for all $m \leq \log\log X$ outside of an exceptional set $\mathcal{M} \subset [1, \log\log x] \cap \mathbb{N}$ with

$$\sum_{m \in \mathcal{M}} \frac{1}{m} \ll \varepsilon^{-3}.$$

In particular, we can locate some $m$ with the property (145) belonging to the range $m \in [\varepsilon' \log\log x, \frac{1}{10} \log\log x]$ with $\varepsilon' := \exp(-\varepsilon^{-10})$. Let $P = 2^m \geq (\log x)^{\varepsilon'/2}$, where $m$ has this value. Then, by introducing the von Mangoldt weight, we have

$$C = (-1)^k \mathbb{E}_{P \leq d \leq 2P} \Lambda(d) \mathbb{E}^{\log}_{n \leq x} \lambda(n + dh_1) \cdots \lambda(n + dh_k) + O(\varepsilon)$$

As in the proof of Theorem 5.1, we may split $d$ into residue classes $\pmod W$ with $W = \prod_{p \leq w} p$ and $w = w(x)$ tending to infinity slowly enough, and then apply the Gowers uniformity of the $W$-tricked von Mangoldt function and the generalized von Neumann theorem (as in [36, Section 5]) to conclude that

$$C = (-1)^k \frac{W}{\phi(W)} \mathbb{E}_{P \leq d \leq 2P} 1_{(d,W)=1} \mathbb{E}^{\log}_{n \leq x} \lambda(n + dh_1) \cdots \lambda(n + dh_k) + O(\varepsilon).$$

Arguing as in the proof of (139), we have

$$C \ll O_W(\mathbb{E}^{\log}_{n\le x}\kappa(\|\lambda\|_{U^k[n,n+3kP]})) + \varepsilon.$$

Since $P \ge (\log x)^\eta$ where $\eta = \varepsilon'/2$, the hypothesis of the theorem will then give $C = O(\varepsilon)$ if we assume $x$ sufficiently large depending on $w$. □

## 6. Reducing the length of the intervals

In this section we indicate the changes needed to the proof of Theorem 1.1 to obtain Theorem 1.8. Up to Proposition 3.7 (corresponding to the work up to [26, Section 4]), everything works for smaller $H$ as well, except in the statement of Proposition 3.6 the range for $P', P''$ is now $[H^{\varepsilon^2/2}, H^\varepsilon]$.

To proceed, we will need the following variant of Lemma 3.12 in which the implied constants do not depend on the number of primes in the product. Crude bounds suffice here and stronger bounds would not be useful as we in any case lose factors like $\ell!$ in our arguments.

**Lemma 6.1** (Counting nearby products of primes)**.** *Let $m, \ell, q \in \mathbb{N}$ and $P', N \ge 3$. Then the number of $2\ell$-tuples $(p'_{1,1}, \dots, p'_{1,\ell}, p'_{2,1}, \dots, p'_{2,\ell})$ of primes in $[P', 2P']$ obeying the conditions*

$$\left|\prod_{j=1}^{\ell} p'_{2,j} - \prod_{j=1}^{\ell} p'_{1,j}\right| \le C \cdot \frac{(2P')^\ell}{N}$$

*and*

$$\prod_{j=1}^{\ell} (p'_{2,j})^m = \prod_{j=1}^{\ell} (p'_{1,j})^m \bmod q$$

*for some $C \ge 1$ is bounded by*

$$\ll C\ell!^2 (2P')^\ell m^{\omega(q)} \left(\frac{(2P')^\ell}{Nq} + 1\right).$$

*Proof.* Since every integer has at most $\ell!$ representations as a product of $\ell$ primes, the number of prime tuples we need to count is at most $\ell!^2$ times the number of integers $n_1, n_2 \le (2P')^\ell$ for which

$$|n_1 - n_2| \le C \cdot \frac{(2P')^\ell}{N} \qquad \text{and} \qquad n_1^m = n_2^m \bmod q.$$

The claim follows by noticing that there are $(2P')^\ell$ choices for $n_1$, and after fixing it, there are at most $m^{\omega(q)}$ choices for $n_2 \bmod q$. □

Let us now get back to Proposition 3.7 corresponding to [26, Proposition 4.1]. In our setting we obtain the following variant, where we for simplicity restrict to the case $\ell_1 = \ell_2 = \ell$ and a single quadruple $\vec{a}$ corresponding to each $e \in \mathcal{Q}$ as this is sufficient for the polynomial phase case.

**Proposition 6.2** (Local structure of $\phi''$)**.** *Let the hypotheses be as in Theorem 1.8, and let $\varepsilon, X, P', P'', \mathcal{I}'', \phi''_{I''}, \mathcal{Q}$ be as in Proposition 3.6 (except now $P', P'' \in [H^{\varepsilon^2/2}, H^{\varepsilon}]$). Let $\ell$ be an even integer such that*

$$N^2 d^{10} \leq d^\ell = O(N^{O(1)}). \tag{146}$$

*We allow implied constants to depend on $\varepsilon, \eta$ and $\theta$. There exists a constant $c = c(\varepsilon, \eta, \theta)$ such that, for a subset $\mathcal{Q}'$ of the quadruples $e = (I''_1, I''_2, p'_1, p'_2)$ in $\mathcal{Q}$ of cardinality $\gg c^\ell dN$, one can find a quadruple $\vec{a} = (a_1, a_2, b_1, b_2)$ of natural numbers, and a collection $\mathcal{P}_e$ of primes in $[P''/2, P'']$ with $|\mathcal{P}_e| \gg (\log X)^{-10\ell}\pi_0(P')$, with the following properties:*

(i) *One has*

$$\frac{1}{p'_2} \circ \phi''_{I''_1} \sim_{\frac{1}{\prod \mathcal{P}_e}} \frac{1}{p'_1} \circ \phi''_{I''_2}. \tag{147}$$

(ii) *For $i = 1, 2$, $a_i, b_i$ are products of $\ell_i$ primes in $[P', 2P']$; in particular*

$$(P')^\ell \leq a_i, b_i \leq (2P')^\ell. \tag{148}$$

*Furthermore we have*

$$0 \neq a_i - b_i \ll \frac{C^\ell}{N} a_i, \tag{149}$$

*where $C$ is an absolute constant.*

(iii) *For $i = 1, 2$, we have the approximate dilation invariance*

$$\frac{1}{a_i} \circ \phi''_{I''_i} \sim_{\frac{1}{\prod \mathcal{P}_{e,\vec{a}}}} \frac{1}{b_i} \circ \phi''_{I''_i}. \tag{150}$$

*Here (abusing the notation) the implied constants depend linearly on $\ell$.*

*Sketch of proof.* The proof is very similar to the proofs of Propositions 3.7 and [26, Proposition 4.1]: One makes two cycles of length $\ell$ joined by a "middle edge". The main difference is that now $\ell \asymp (\log x)^{1-\theta}$, so $\ell$ is no longer a constant.

Since the number of edges in the graph is $\gg \frac{X}{H} P'^2/(\log P')^2$, the number of such constellations gets reduced by a factor $c'^\ell$ (with certain constant $c' \in (0,1)$). Hence the Cauchy-Schwarz argument at the end naturally only gives us $\gg c'^\ell X/H \cdot \pi_0(P')^2$ middle edges.[20] Since $P'$ is larger than $(c\ell \log P')^{O(\ell)}$, Lemma 6.1 is sufficient to show that degenerate cases involving repeating primes or products are negligible as before.

Since the constellation involves $2\ell + 1$ edges, the intersection

$$\mathcal{P}(\overrightarrow{I''}) := \mathcal{P}(\{I''_{0,1}, I''_{0,2}\}) \cap \bigcap_{j=1}^{k} \bigcap_{i=1,2} \mathcal{P}(\{I''_{j,i}, I''_{j+1,i}\})$$

that appears in [26, (52)] is now expected to be only of size $c^\ell \pi_0(P')$ for some constant $c > 0$, so $\delta$ in [26, (52))] cannot anymore be taken to be a constant but can be at most $c^\ell$. In fact to compensate for losses in Lemma 6.1 we choose $\delta$ in [26, (52))] to be $(\log X)^{-10\ell}$. Then in the argument below [26, (52))] the number of candidate tuples is at most $\ell!^4 P'^{4\ell+1}/N$ and

[20] This might be fixable through arguing more carefully removing some edges before running the argument but this would be of no importance.

so the expected number of good tuples obeying [26, (52))] is $\ll (\log X)^{-10\ell}\ell!^4 P'^{4\ell+1}/N \ll (\log X)^{-\ell} d^{2\ell+1}/N$ whereas with probability $\gg 1$, there are $\gg c^\ell d^{2\ell+1}/N$ non-degenerate good tuples. Hence one can indeed find a deterministic choice of $\mathbf{p}$ such that there are $\gg c^\ell d^{2\ell+1}/N$ very good tuples, i.e. tuples for which

$$\#\mathcal{P}(\overrightarrow{I''}) > (\log X)^{-10\ell}\pi_0(P')$$

as desired. □

Lowering $H$ does not affect solving the approximate dilation invariance in Proposition 3.8, except that the bounds for $T$ and the smoothness of $\varepsilon_i^{(j)}(t)$ get worsened by $C^\ell$ for a constant $C$. Since $\mathcal{P}(\overrightarrow{I''})$ now of size $\gg (\log X)^{-10\ell}\pi_0(P')$, we now need to take $K \gg (\log X)^{10\ell}$ in Proposition 3.9, so in Corollary 3.10 we now have $\#\mathcal{F}(I'') \ll (\log X)^{10\ell}$. Proposition 3.11 works without changes but now it provides only $\gg c^\ell X/H\pi_0(P')^2$ pairs $(I_1'', I_2'')$.

To proceed, we need an adequate version of the mixing lemma:

**Lemma 6.3** (Mixing lemma)**.** *Let $X, V \geq 3, 2 \leq P \leq H$. Let $\mathcal{A}_1, \mathcal{A}_2$ be two $(X,H)$-families of intervals. Write*

$$\mathcal{V} = \left\{ \xi \in [-X/H, X/H] \colon \Big| \sum_{P \leq p \leq 2P} p^{2\pi i \xi} \Big| \geq PV^{-1} \right\}.$$

*Then the number of quadruplets $(J_1, J_2, p_1, p_2)$ with $J_1 \in \mathcal{A}_1, J_2 \in \mathcal{A}_2$, $p_1, p_2$ primes in $[P, 2P]$, and $I_1$ lying within $100H$ of $\frac{p_2}{p_1} I_2$ is*

$$\ll |\mathcal{V}|(\#\mathcal{A}_1)(\#\mathcal{A}_2)\frac{H}{X}\left(\frac{P}{\log P}\right)^2 + (\#\mathcal{A}_1)^{1/2}(\#\mathcal{A}_2)^{1/2}P^2V^{-2}. \tag{151}$$

*Proof.* As in [26, Proof of Lemma 5.1], the number of quadruplets in question is bounded by

$$\ll \frac{H}{X}\int_{|\xi| \leq \frac{X}{H}} |S_1(\xi)||S_2(\xi)||T(\xi)|^2 \, d\xi \tag{152}$$

where

$$S_i(\xi) := \sum_{I \in \mathcal{A}_i} e(\xi \log x_I)$$

for $i = 1, 2$ and

$$T(\xi) := \sum_{P \leq p \leq 2P} p^{2\pi i \xi}. \tag{153}$$

Splitting the integral in (152) according to whether $\xi \in \mathcal{V}$, we obtain that (152) is at most

$$\frac{H}{X}|\mathcal{V}| \sup_{|\xi|\in\mathcal{V}} |S_1(\xi)S_2(\xi)T(\xi)^2| + \frac{H}{X}P^2V^{-2}\int_{|\xi|\le\frac{X}{H}} |S_1(\xi)||S_2(\xi)|\, d\xi$$
$$\ll \frac{H}{X}|\mathcal{V}|(\#\mathcal{A}_1)(\#\mathcal{A}_2)\left(\frac{P}{\log P}\right)^2 + \frac{H}{X}P^2V^{-2}\left(\int_{|\xi|\le\frac{X}{H}} |S_1(\xi)|^2\, d\xi \int_{|\xi|\le\frac{X}{H}} |S_2(\xi)|^2\, d\xi\right)^{1/2}.$$

From the large sieve inequality (see e.g. [26, Lemma 2.3]) we have

$$\int_{|\xi|\le\frac{X}{H}} |S_i(\xi)|^2 \ll \#\mathcal{A}_i\frac{X}{H}, \tag{154}$$

and the claim follows. □

Note that the size of $\mathcal{V}$ above is at most twice the size of the maximal one-spaced subset of $\mathcal{V}$ (meaning a set where any two points are at least one apart). The needed bound for $|\mathcal{V}|$ in our situation is provided by the following lemma. The requirement $\theta > 5/8$ comes from it as for smaller $\theta$ we do not know how to obtain $|\mathcal{V}| = P^{o(1)}$.

**Lemma 6.4.** *Let $\theta \in (5/8, 1)$ be fixed, $H = \exp((\log X)^\theta)$ and $P = \exp(\varepsilon(\log X)^\theta)$ for some $\varepsilon > 0$, and let $V = (\log X)^{100\ell}$, where $\ell \asymp (\log X)^{1-\theta}$. Let $\mathcal{U}$ be a set of one-spaced points $\xi \in [-X/H, X/H]$ for which*

$$\Big|\sum_{p\sim P} p^{2\pi i\xi}\Big| \ge PV^{-1}.$$

*Then, for some $\varepsilon' > 0$, we have*

$$\#\mathcal{U} \ll \exp((\log X)^{\theta-\varepsilon'}) = P^{o(1)}.$$

**Remark 6.5.** *From the proof of Lemma 6.4, it will be clear that the larger $\theta > 5/8$ is, the better the bound we can obtain on $\#\mathcal{U}$. In fact, for $\theta = 2/3 + \varepsilon$ the Vinogradov–Korobov bound (see [22, Lemma 2]) directly gives $\mathcal{U} \subset [-V^2, V^2]$, so that $\#\mathcal{U} \ll V^2 \ll \exp((\log X)^{1-\theta+\varepsilon^2})$, say. Nevertheless, here the main interest is in the smallest value of $\theta$ for which $\#\mathcal{U} \ll \exp((\log X)^{\theta-\varepsilon'})$ holds, so this aspect is not optimized.*

*Proof.* Let $T(\chi)$ be as in (153). We apply [24, Lemma 4.4], which is a variant of the Halász–Montgomery estimate that uses Vinogradov's bound on $\sum_{P\le n\le 2P} n^{it}$ as an input (see also Lemma 6.6 below with $q = 1$). This gives that uniformly for $\eta \in (0, 1)$ and integers $k \ge 0$ we have

$$\#\mathcal{U}\cdot\Big(\frac{P}{V}\Big)^{2k} \ll \sum_{t\in\mathcal{U}} |T(\xi)|^{2k} \ll \Big((2P)^k + \#\mathcal{U}\cdot X^{5\eta^3/2}(\log X)^{2/3}\cdot(2P)^{k(1-\eta/4)}\Big)k!\cdot(2P)^k. \tag{155}$$

This means that we have the bound

$$\#\mathcal{U} \ll (4kV)^{2k}$$

whenever $X^{5\eta^{3/2}}(\log X)^{3/2}k^{2k}P^{k(2-\eta/4)} = o((P/V)^{2k})$. The latter holds whenever

$$X^{5\eta^{3/2}}k^{2k}\cdot\exp(k(\log X)^{1-\theta}(\log\log X)^2) = o(\exp(\tfrac{\varepsilon}{4}\eta k(\log X)^\theta)),$$

which in turn follows from

$$5\eta^{3/2}\log X + 2k\log k + k(\log X)^{1-\theta}(\log\log X)^2 < \tfrac{\varepsilon}{5}\eta k(\log X)^\theta.$$

This holds (assuming already $k = (\log X)^{O(1)}$ and letting $\delta$ be a small positive constant) if

$$\begin{cases} k \geq \eta^{1/2}(\log X)^{1-\theta+\delta} \\ \eta \geq (\log X)^{-\theta+\delta} \\ \eta \geq (\log X)^{1-2\theta+\delta}. \end{cases}$$

For $\theta < 1$, the third condition is more demanding than the second and thus we can set $\eta = (\log X)^{1-2\theta+\delta}$ and $k = (\log X)^{3/2-2\theta+2\delta}$. With these choices the first term dominates in (155) and we obtain the upper bound

$$\#\mathcal{U} \ll (4kV)^{2k} \ll (\log X)^{300\ell k} \ll \exp((\log X)^{5/2-3\theta+3\delta})$$

The claim follows as $5/2 - 3\theta < \theta$ since $\theta > 5/8$. □

Now this leads to approximate ergodicity [26, Corollary 5.2] except that now we have either

$$\frac{MK^3}{\delta} \gg (\log X)^{100\ell}$$

or a collection $\mathcal{T}$ as in [26, Corollary 5.2] but with

$$\#\mathcal{T} \gg \exp(-(\log X)^{\theta-\varepsilon})\frac{\delta}{MK^3}\frac{X}{H}. \tag{156}$$

We can apply this with $\delta = c^\ell$, $K \asymp (\log X)^{10\ell}$, $M = 100$ and $r = 1/10$ to get conclusions between Proposition 3.11 and Lemma 3.12, except that now have the weaker lower bound $\#\mathcal{T} \gg \exp(-(\log X)^{\theta-\varepsilon})X/H$.

As for the analogue of Proposition 3.13, we can use the same argument as in its proof to obtain upper and lower bounds for the number of certain tuples $(Q_0,\dots,Q_{\ell-1}) \in \mathcal{T}^\ell$: The lower bound we get is $\gg c^\ell d^\ell$ (with $d := (P'/\log P')^2$) and the upper bound (from Lemma 6.1) is

$$\ll \ell!^2(2P')^\ell k^{\omega(q)}\left(\frac{(2P')^\ell}{q_0^{1/k}N}+1\right)$$

Combining the lower and upper bounds, we obtain $q_0 \ll (\log X)^{O(\ell)}$.

Now, repeating the arguments after Proposition 3.13, we see that there are at least $\gg \exp(-(\log X)^{\theta-\varepsilon})X/(P'P'')$ integers $X/(2P'P'') \leq x \leq X/(P'P'')$ for which

$$\Big|\sum_{n\in[x,x+H^*]} f(n)n^{-iT}e(-\gamma(n))\Big| \gg H^* \tag{157}$$

with $H^* := C^{-\ell}H/(P'P'')$ and $\gamma(t) = \sum_{j=0}^k c_j\binom{t/q_0}{j}$, where $c_j$ are integers.

Now we will obtain a contradiction as in Section 3, except due to worse bounds for $\mathcal{T}^*$ and $q_0$ we need to use results from [24] where one obtains a polynomial saving in the exceptional set for averages of multiplicative functions in short intervals (in the special case $f = \lambda$ and $\theta = 2/3 + \varepsilon$ arguments of [23] actually suffice — see Remark 6.7 below). Also since $q_0$ is not bounded, we need to treat the $q$-aspect non-trivially.

As in [23, 24] we first restrict $n$ to a set of numbers with factors of convenient sizes. For this, let $\delta$ be small in terms of the implied constant above and define $\mathcal{S}$ as in [24, Proof of Theorem 1.7 in Section 11], i.e. choose in [24, Section 9] the parameters $\eta = 1/150, \nu_1 = \delta^2/4000, \nu_2 = 1/10, Q_1 = H^*$ and $P_1 = Q_1^{\delta/4}$, so that $J = 1$, $P_2 = X^{\nu_1}, Q_2 = P_3 = X^{\sqrt{\nu_1\nu_2}}$ and $Q_3 = X^{\nu_2}$ and $\mathcal{S}$ consists of numbers with a prime factor on each interval $(P_j, Q_j]$ with $j = 1, 2, 3$.

Using the linear sieve (cf. [24, Proof of Theorem 1.7 in Section 11]), we see that $n \notin \mathcal{S}$ make a negligible contribution of

$$H^* \sum_{1\le i\le 3} \frac{\log P_i}{\log Q_i} \ll \delta H^*,$$

to (157) and so we have $\gg \exp(-(\log X)^{\theta-\varepsilon})X/(P'P'')$ integers $X/(2P'P'') \le x \le X/(P'P'')$ for which

$$\Big| \sum_{\substack{n\in[x,x+H^*]\\ n\in\mathcal{S}}} f(n)n^{-iT}e(-\gamma(n))\Big| \gg H^*.$$

Splitting into residue classes $a \bmod q_0$ and then splitting according to $q_2 = \gcd(a, q_0)$, we see that

$$\sum_{q_2:\ q_0=q_1q_2} \Big| \sum_{\substack{b \bmod q_1\\ (b,q_1)=1}} e(-\gamma(bq_2)) \sum_{\substack{n\in\mathcal{S}\\ n\in[x/q_2,(x+H^*)/q_2]\\ n=b \bmod q_1}} f(n)n^{-iT}\Big| \gg H^*$$

for $\gg \exp(-(\log X)^{\theta-\varepsilon})X/(P'P'')$ integers $X/(2P'P'') \le x \le X/(P'P'')$. This implies that for some choice of $q_0 = q_1q_2$, we have

$$\Big| \sum_{\substack{b \bmod q_1\\ (b,q_1)=1}} e(-\gamma(bq_2)) \sum_{\substack{n\in\mathcal{S}\\ n\in[x,x+H^*/q_2]\\ n=b \bmod q_1}} f(n)n^{-iT}\Big| \gg \frac{\phi(q_1)}{q_1q_2}H^*$$

for $\gg \exp(-(\log X)^{\theta-\varepsilon})X/(q_2P'P'')$ integers $X/(2q_2P'P'') \le x \le X/(q_2P'P'')$. Moving into characters, the left-hand side is at most

$$\frac{1}{\phi(q_1)} \sum_{\chi \bmod q_1} \Big| \sum_{\substack{b \bmod q_1\\ (b,q_1)=1}} e(-\gamma(bq_2))\overline{\chi(b)}\Big| \cdot \Big| \sum_{\substack{n\in\mathcal{S}\\ n\in[x,x+H^*/q_2]}} f(n)\chi(n)n^{-iT}\Big|.$$

Recall that $\gamma$ is a polynomial phase of degree $k$. By [5, Corollary 1.1] and the Chinese reminder theorem we have, for every $\chi$,

$$\Big| \sum_{\substack{b \bmod q_1 \\ (a,q_1)=1}} e(-\gamma(bq_2))\overline{\chi(b)}\Big| = O(q_1^{1-1/(k+1)}), \tag{158}$$

so that

$$\sum_{\chi \bmod q_1} \Big| \sum_{\substack{n\in\mathcal{S} \\ n\in[x,x+H^*/q_2]}} f(n)\chi(n)n^{-iT}\Big| \gg q_1^{1/(k+2)}H^*/q_2 \tag{159}$$

for $\gg \exp(-(\log X)^{\theta-\varepsilon})X/(q_2P'P'')$ integers $X/(2q_2P'P'') \le x \le X/(q_2P'P'')$.

Now, if $q_1 \le Q$ for a constant $Q \ll_{k,\eta,\theta,\rho} 1$ to be determined later, we have, for some $\chi \pmod{q_1}$,

$$\sum_{\chi \bmod q_1} \Big| \sum_{\substack{n\in\mathcal{S} \\ n\in[x,x+H^*/q_2]}} f(n)\chi(n)n^{-iT}\Big| \gg_{k,\eta,\theta,\rho} H^*/q_2$$

for $\gg \exp(-(\log X)^{\theta-\varepsilon})X/(q_2P'P'')$ integers $X/(2q_2P'P'') \le x \le X/(q_2P'P'')$. By [24, Theorem 9.2(i)] this implies that

$$\Big| \sum_{\substack{n\in\mathcal{S} \\ X<n\le 2X}} f(n)\chi(n)n^{-iT+it_0}\Big| \gg_{k,\eta,\theta,\rho} X,$$

for some $|t_0| \le X$, which in turn by inclusion-exclusion and Halász's theorem implies (13) since $|T| \le C^\ell(X/H)^{k+1} \le X^{k+1}/H^{k+1-\rho/2}$.

Let us now turn to the case $q_1 \ge Q$. The proof of [24, Proposition 8.3] (taking $\mathcal{V}_1 = \emptyset$ in the proof of [24, Proposition 8.3] and bounding $R_C(1+it)$ trivially) gives

$$\begin{aligned}\frac{1}{H^*/q_2} \sum_{\substack{n\in\mathcal{S} \\ n\in[x,x+H^*/q_2]}} f(n)\chi(n)n^{-iT} &= A(x,H^*/q_2,\mathcal{U}) + O\Big(\frac{1}{H^*/q_2}\Big) \\ &+ O\Big(\Big(\sum_{\substack{A=2^j \\ P_3/2\le A\le Q_3}} \sum_{t\in\mathcal{W}^*(\chi)} |Q_{3,A}(\chi,1+it)|^2 \sum_{\substack{B=2^j \\ P_2/2\le B\le Q_2}} \sum_{t\in\mathcal{W}^*(\chi)} |Q_{2,B}(\chi,1+it)|^2\Big)^{1/2}\Big),\end{aligned} \tag{160}$$

where

$$\mathcal{W}^*(\chi) \subset \{|t| \le X\colon \max_B |Q_{2,B}(\chi,1+it)| \ge X^{-\nu_1^3/320}\}$$

is one-spaced,

$$Q_{j,D}(\chi,s) := \sum_{\substack{D<p\le 2D \\ P_j<p\le Q_j}} \frac{\chi(p)}{p^s},$$

and $A(x,H^*/q_2,\mathcal{U})$ satisfies [24, (46)].

As in [24, Proof of Theorem 9.2(ii)] with same choices of $\mathcal{U}$ and $d_n$, we have $|A(x,H^*/q_2,\mathcal{U})| \ll H^{*-\delta/5000}$ except for $\ll XH^{*-\delta/5000}$ values $X/(2q_1P'P'') \le x \le X/(q_1P'P'')$. Summing

over $\chi \bmod q_1$ and taking the union bound, the contribution from $A(x, H^*/q_2, \mathcal{U})$ is acceptable.

Given all this, (159) implies that

$$\sum_{\chi \bmod q_1} \sum_{A} \sum_{t\in\mathcal{W}^*(\chi)} |Q_{3,A}(\chi, 1+it)|^2 \sum_{\chi \bmod q_1} \sum_{B} \sum_{t\in\mathcal{W}^*(\chi)} |Q_{2,B}(\chi, 1+it)|^2 \gg Q^{2/(k+2)} \tag{161}$$

In [24] this sort of term with $q_1 = 1$ is dealt with using [24, Lemma 4.4] which is a large values result of Halász –Montgomery type that uses Ford's bound (see [8, Theorem 1])

$$|\zeta(\sigma+it)| \ll 1 + |t|^{\frac{9}{2}(1-\sigma)^{3/2}} (\log(|t|+2))^{2/3} \quad \text{for } 1/2 \le \sigma \le 1.$$

for $\zeta(s)$. As pointed out by Ford, $L(s,\chi) = q^{-s}\sum_{m=1}^{q} \chi(m)\zeta(s, m/q)$, where $\zeta(s,u) = \sum_{n=0}^{\infty}(n+u)^{-s}$ is the Hurwitz zeta function, so that [8, Theorem 1] also gives

$$|L(\sigma+it,\chi)| \ll q^{1-\sigma}|t|^{\frac{9}{2}(1-\sigma)^{3/2}} (\log(|t|+2))^{2/3} + \frac{q^{1-\sigma}}{1-\sigma} \quad \text{for } 1/2 \le \sigma < 1.$$

Using this in the proof of [24, Lemma 4.4], we get the following variant.

**Lemma 6.6.** *Let $T \ge 3$ $q \ge 1$ and let $\mathcal{T}$ be a set of pairs $(\chi, t)$, where $\chi$ is a Dirichlet character $\bmod q$ and $t \in [-T, T]$ such that if $(\chi, t_1), (\chi, t_2) \in \mathcal{T}$, then $|t_1 - t_2| \ge 1$. Let $P(s,\chi) = \sum_{N<p\le 2N} a(p)\chi(p)p^{it}$ be a Dirichlet polynomial of length $N \le T^2$ whose coefficients are supported on primes. Then, for any $\varepsilon', \eta \in (0, 1/2)$,*

$$\sum_{(\chi,t)\in\mathcal{T}} |P(\chi, it)|^2 \ll_{\varepsilon'} \Big(\frac{N}{\log N} + |\mathcal{T}| \cdot (q^{\eta} T^{\frac{9}{2}\eta^{3/2}} (\log T)^{2/3} + q^{\eta}/\eta) \cdot N^{1-\eta(1-\varepsilon')}\Big) \sum_{N<p\le 2N} |a(p)|^2.$$

Using this and arguing as in [24, Proof of Proposition 8.3], we obtain

$$\sum_{\chi \bmod q_1} \sum_{A} \sum_{t\in\mathcal{W}^*(\chi)} |Q_{3,A}(\chi, 1+it)|^2 \sum_{\chi \bmod q_1} \sum_{B} \sum_{t\in\mathcal{W}^*(\chi)} |Q_{2,B}(\chi, 1+it)|^2 \ll 1$$

which contradicts (161) once the constant $Q$ is large enough. Hence Theorem 1.8 follows.

**Remark 6.7.** *We remark that the special case $f = \lambda$ of Theorem 1.8 with the weaker value $\theta = 2/3 + \varepsilon$ can be proved more simply by relying only on* [22] *as follows. Firstly note that, by Remark 6.5, we can replace $\exp(-(\log X)^{\theta-\varepsilon})$ with $\exp(-(\log X)^{1-\theta-\varepsilon^2})$ in* (156) *and on later occurrences. Note also that in this case $q_1, q_2 \ll \exp((\log X)^{1/3-\varepsilon/2})$.*

*We must then show that* (159) *with $f = \lambda$ cannot hold for $\gg \exp(-(\log X)^{1-\theta-\varepsilon^2})X/(q_2 P' P'')$ integers $X/(2q_2P'P'') \le x \le X/(q_2P'P'')$. We have the Vinogradov–Korobov zero-free region for $L$-functions of the form*

$$L(s,\chi) \ne 0, \quad \sigma \ge 1 - \frac{c_0}{\log q_1 + (\log(|t|+3))^{2/3}(\log\log(|t|+3))^{1/3}} \tag{162}$$

*for all $\chi \pmod{q_1}$, apart from possibly one real zero corresponding to one real character. In case an exceptional character exists, $q_1 \gg_A (\log X)^A$. The contribution of an exceptional character to* (159) *is trivially negligible, so we may assume that in* (159) *we only sum over*

*characters* $\chi \pmod{q_1}$ *satisfying* (162). *Moreover, we may assume that the set* $\mathcal{S}$ *in* (159) *is instead simply defined as the set of* $n$ *having a prime factor from* $[P, Q]$, *with* $Q = H^*$, $P = Q^{\delta/4}$. *We again claim that* (159) *fails, which will then provide the desired contradiction.*

*To show this claim, we apply the proof method of* [22] *to the multiplicative function* $\lambda(n)\chi(n)n^{-iT}$, *summed over* $n \in \mathcal{S}$. *Reducing matters from short sums to Dirichlet polynomials by Parseval-type arguments, as in* [22, Section 4], *we can reduce the claim to*

$$\int_{[-T_1,T_1]\setminus[T-T_0,T+T_0]} |P(1+it)|^2|Q(1+it)|^2\,dt \ll \exp(-(\log X)^{1/3-\varepsilon/10}), \tag{163}$$

*where* $T_0 = \exp((\log X)^{1/3-\varepsilon/10})$ *and* $T_1 = X\exp((\log X)^{1/3-\varepsilon/10})/H^*$, *and we have* $P(s) = \sum_{p\in I}\chi(p)p^{s-iT}$ *for some interval* $I \subset [P, Q]$ *and* $Q(s) = \sum_{X/Q\le n\le X} a_n n^s$ *for some* $|a_n| \le 1$. *As in* [22], *applying the pointwise Vinogradov–Korobov bound to* $P(s)$ *and the mean value theorem to* $Q(s)$, (163) *follows.*

## 7. Polynomial averages of the Liouville function

In this section, we prove Theorems 1.10 and 1.12. Note that Corollary 1.11 is a special case[21] of Theorem 1.10 where we take $P_i(m) = a_i m$.

*Proof of Theorems 1.10 and 1.12.* We borrow notation from [40]. Note that the claim of Theorem 1.10 follows from

$$\mathbb{E}_{\mathbf{m}\in[X^\varepsilon]^r}\mathbb{E}_{n\le X}c_X(\mathbf{m})\lambda(n+P_1(\mathbf{m}))\lambda(n+P_2(\mathbf{m}))\cdots\lambda(n+P_k(\mathbf{m})) = o(1) \tag{164}$$

for an arbitrary unimodular sequence $c_X(\mathbf{m})$. Denoting $W = \prod_{p\le w} p$, where $w$ tends to infinity very slowly in terms of $X$, and splitting $n$ and $m$ into residue classes $\pmod W$ in the statement of Theorem 1.12, that theorem in turn reduces to

$$\mathbb{E}_{\mathbf{m}\in[L]^r}\mathbb{E}_{n\le X/W}c_X(\mathbf{m})\lambda_{b_1,W}(n+P_1'(\mathbf{m}))\Lambda_{b_2,W}(n+P_2'(\mathbf{m}))\cdots\Lambda_{b_k,W}(n+P_k'(\mathbf{m})) = o(1) \tag{165}$$

uniformly for unimodular sequences $c_X(\mathbf{m})$, for $X^\varepsilon/W \ll L \ll X^\varepsilon$, for $1 \le b_1, \dots, b_k \le W$ coprime to $W$, and for $P_1', \dots, P_k'$ polynomials in $\mathbb{Z}[x_1, \dots, x_r]$ with $P_i' - P_j'$ non-constant for $i \ne j$, and $\deg P_i' \le d$, and the coefficients of $P_i'$ bounded by $W^{1/\kappa}$ in absolute value for some constant $\kappa > 0$ (cf. [40, Section 5] for this reduction). Here we have denoted $\lambda_{b,W}(n) := \lambda(Wn+b)$, and recall that $\Lambda_{b,W}(n) := \phi(W)/W \cdot \Lambda(Wn+b)$. We now see that in fact both Theorem 1.10 and 1.12 will follow once we prove (165) in a form where some copies of $\Lambda$ are allowed to be replaced with $\lambda$.

Let $A = W^{1/\kappa}$, so that the absolute values of the coefficients of $P_i'$ are bounded by $A$. Recall $d = \max_i \deg P_i'$. We set $N = \lfloor X/W \rfloor$, so that $L = o(N^{1/d})$. Consider functions $f_1, \dots, f_k : [N] \to \mathbb{C}$ with $|f_i| \ll \Lambda_{b_i,W} + 1$ and $|f_1| \le 1$. Extend the $f_i$ to functions $\tilde{f}_i : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ by making them $N$-periodic. Observe that

$$\mathbb{E}_{\mathbf{m}\in[L]^r}\mathbb{E}_{n\le N}c_X(\mathbf{m})f_1(n+P_1'(\mathbf{m}))\cdots f_k(n+P_k'(\mathbf{m})) \tag{166}$$

[21] This special case could in fact be proved more directly without considering polynomial progressions, instead combining the generalized von Neumann theorem with Corollary 1.6.

is up to $o(1)$ error equal to

$$\mathbb{E}_{\mathbf{m}\in[L]^r} c_X(\mathbf{m})\mathbb{E}_{n\in\mathbb{Z}/N\mathbb{Z}}\tilde{f}_1(n+P_1'(\mathbf{m}))\cdots\tilde{f}_k(n+P_k'(\mathbf{m})), \tag{167}$$

since the components of the $\mathbf{m}$ variable in (166) are bounded by $\eta X^{1/d}$ for some $\eta > 0$ small enough in terms of $A$, so that wraparound issues are negligible.

This latter expression is in turn bounded using van der Corput's inequality (see e.g. [13, Formula (4.1)]) by

$$\ll (\mathbb{E}_{h\in\mathbb{Z}/N\mathbb{Z}}|\mathbb{E}_{\mathbf{m}\in[L]^r}\mathbb{E}_{n\in\mathbb{Z}/N\mathbb{Z}}\Delta_h\tilde{f}_1(n+P_1'(\mathbf{m}))\cdots\Delta_h\tilde{f}_k(n+P_k'(\mathbf{m}))|)^{1/2},$$

where $\Delta_h f(x) := f(x+h)\overline{f(x)}$.

By [40, Theorem 13], for any polynomials $P_i'$ as in Theorems 1.10, 1.12, we have

$$|\mathbb{E}_{n\in\mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\mathbf{m}\in[L]^r}\Delta_h\tilde{f}_1(n+P_1'(\mathbf{m}))\cdots\Delta_h\tilde{f}_k(n+P_k'(\mathbf{m}))| = o(1),$$

provided that

$$\mathbb{E}_{\mathbf{t}\in[A^{-1}L]^r}\|\Delta_h\tilde{f}_1\|_{\square^{D'}_{Q_1(\mathbf{t})[-A^{-1}L,A^{-1}L],\dots,Q_{D'}(\mathbf{t})[-A^{-1}L,A^{-1}L]}} = o(1) \tag{168}$$

for any fixed $D' \geq 1$ and any polynomials $Q_1,\dots,Q_{D'} \in \mathbb{Z}[t_1,\dots,t_r]$ not identically zero and with coefficients of size $O(A^{O(1)})$, where

$$\|f\|_{\square^d_{C_1,\dots,C_d}} := \left(\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{h_1\in C_1-C_1}\cdots\mathbb{E}_{h_d\in C_d-C_d}\prod_{\omega\in\{0,1\}^d}\mathcal{C}^{|\omega|}f(x+\omega\cdot\mathbf{h})\right)^{1/2^d}$$

is a Gowers box norm of order $d$ and $\mathcal{C}$ is the complex conjugation operator, and we used the notation $q[-N,N] := [-qN,qN]\cap q\mathbb{Z}$. Thus we may control polynomial averages with averaged Gowers box norms. Further, by a concatenation theorem, namely [40, Theorem 9] (with $d_0 = 1$ there), we have (168) provided that

$$\|\Delta_h\tilde{f}_1\|_{U^{D''}_{q[1,A^{-2D''}L]}} = o(1) \tag{169}$$

holds for all fixed $D'' \geq 1$ and all $1 \leq q \leq A^{D''}$, where $\|f\|_{U^d_C} := \|f\|_{\square^d_{C,\dots,C}}$.

Averaging this over $h$, we now conclude that the desired bound for (167) follows from

$$\mathbb{E}_{h\in\mathbb{Z}/N\mathbb{Z}}\|\Delta_h\tilde{f}_1\|^{2^{D''}}_{U^{D''}_{q[1,A^{-2D''}L]}} = o(1).$$

Expanding out the Gowers norm above, we see that this claim in turn reduces to

$$\|\tilde{f}_1\|_{U^{D''+1}_{\mathbb{Z}/N\mathbb{Z},q[1,A^{-2D''}L],\dots,q[1,A^{-2D''}L]}} = o(1). \tag{170}$$

Since wraparound issues are again negligible, we can split the average over $\mathbb{Z}/N\mathbb{Z}$ implicit in (170) into intervals of length $\asymp L$ and apply the generalized von Neumann theorem, thus reducing the proof of (170) to

$$\sup_{A^{-c}L\leq M\leq A^c L}\mathbb{E}_{n\leq N-M}\|f_1\|_{U^{D''+1}[n,n+M]} = o(1) \tag{171}$$

for any constant $c \geq 1$.

Now specialize to the case where $f_1$ is the ($W$-tricked) Liouville function $\lambda_{b,W}(n)1_{[N]}(n)$ (and $N = \lfloor X/W \rfloor$ as before). By making a change of variables, and extending the range of the supremum in $W1_{m\equiv b \pmod W}$, we reduce (171) to

$$\sup_{N^{\varepsilon/2}\leq M\leq N^{2\varepsilon}} \mathbb{E}_{n\leq W(N-M)} \|\lambda \cdot W1_{\cdot\equiv b \pmod W} 1_{[WN]}\|_{U^{D''+1}[n,n+M]} = o(1). \tag{172}$$

The factor $1_{[WN]}$ can be removed, since the contribution to the $n$ average from the range $WN - O(M) \leq n \leq W(N-M)$ is negligible. By Fourier expanding $1_{\cdot\equiv b \pmod W}$ in terms of additive characters, and applying the triangle inequality (and recalling that $w$ tends to infinity arbitrarily slowly) we reduce[22] to proving (171) also without the factor $W1_{\cdot\equiv b \pmod W}$.

By our main theorem, Theorem 1.5, we have (172) without the term $W1_{m\equiv b \pmod W}1_{[WN]}$, and therefore taking above $f_i \in \{\lambda_{b_i,W}, \Lambda_{b_i,W}\}$ for $1 \leq i \leq k$, both Theorem 1.10 and Theorem 1.12 follow. □

## Appendix A. Bernstein inequality for exponential polynomials

In this appendix we establish the Bernstein inequality for exponential polynomials, Lemma 2.3. We begin with a bound for the number of zeroes of such polynomials:

**Lemma A.1.** *Let $\alpha_1, \dots, \alpha_k$ be real numbers, let $d_1, \dots, d_k$ be non-negative integers, and let $P : \mathbb{R} \to \mathbb{R}$ be a real linear combination of the exponential monomials $t \mapsto t^j \exp(\alpha_i t)$ for $i = 1, \dots, k$ and $0 \leq j \leq d_i$. Then if $P$ is not identically zero, it has at most $k + \sum_{i=1}^k d_i$ zeroes.*

*Proof.* The claim is trivial for $k = 0$, so suppose that $k \geq 1$ and that the claim has already been proven for $k-1$. We now fix $k$ and induct on $\sum_{i=1}^k d_i$. By reordering we may assume that $d_1 \leq d_2 \leq \cdots \leq d_k$. By multiplying $P$ by $t \mapsto \exp(-\alpha_1 t)$ we may assume that $\alpha_1 = 0$. If $d_1$ vanishes, then the derivative $P'$ is a linear combination of the exponential monomials $t \mapsto t^j \exp(\alpha_i t)$ with $2 \leq i \leq k$ and $0 \leq j \leq d_i$, so the claim follows from the outer induction hypothesis on $k$ and Rolle's theorem. If instead $d_1$ does not vanish, then $P'$ is of the same form as $P$ but with $d_1$ replaced by $d_1 - 1$, thus by the induction hypothesis it either vanishes identically or has at most $k + (\sum_{i=1}^k d_i) - 1$ zeros. The claim now follows from Rolle's theorem. □

*Proof of Lemma 2.3.* We allow all implied constants to depend on $k, d_1, \dots, d_k, m, I$. Let $N_0$ be large enough in terms of $k, d_1, \dots, d_k$. We may normalize $\sup_{n=1,\dots,N_0} |P(n)| = 1$. The claim is trivial if $P$ is constant, so we may assume that $P$ is non-constant. By Lemma A.1 the exponential polynomial $P(t)$ then attains the values $\pm 1$ at most $O(1)$ times, so the set $\{t \in \mathbb{R} : |P(t)| \leq 1\}$ is the union of $O(1)$ intervals (possibly of infinite or zero length). As this set contains $\{1, \dots, N_0\}$, we conclude from the pigeonhole principle (for $N_0$ large

[22] Note that even though the Fourier expansion of $1_{\cdot\equiv b \pmod W}$ followed by the triangle inequality loses a multiplicative factor of $W$, this loss is harmless, since $w$, and hence $W$, can be assumed to tend to infinity much slower than the decay rate of (172) without the $W1_{\cdot\equiv b \pmod W}1_{[WN]}$ factor.

enough in terms of $d_1, \dots, d_k$) that this set also contains an interval $[n, n+1]$ for some $n = 1, \dots, N_0 - 1$.

Now observe that $P$ solves the ordinary differential equation

$$\prod_{i=1}^{k} \left( \frac{d}{dt} - \alpha_i \right)^{d_i+1} P(t) = 0.$$

Writing $D := \sum_{i=1}^{k} (d_i + 1) = O(1)$ and $\varepsilon := \sup_{1 \leq i \leq k} |\alpha_i|$ (where, by assumption, $\varepsilon$ is small enough in terms of $k, d_1, \dots, d_k, N_0$), we can write this equation as

$$P^{(D)}(t) + c_{D-1} P^{(D-1)}(t) + \cdots + c_0 P(t) = 0 \tag{173}$$

where the coefficients $c_0, \dots, c_{D-1}$ are of size $O(\varepsilon)$. In terms of the $D$-dimensional vector

$$v(t) := \begin{pmatrix} P(t) \\ \vdots \\ P^{(D-1)}(t) \end{pmatrix}$$

one can write this differential equation as a first-order system

$$\frac{d}{dt} v(t) = (U + E) v(t)$$

where $U$ is the shift matrix

$$U := \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

and $E$ is a $t$-independent matrix of dimension $D$ with all entries being of size $O(\varepsilon)$. The solution of this equation is

$$v(t) = \exp((t-n)(U+E)) v(n).$$

By the continuity of the matrix exponential we then have

$$v(t) = \exp((t-n)U) v(n) + O(\varepsilon \|v(n)\|) \tag{174}$$

whenever $|t - n| = O(1)$ (here $\|\cdot\|$ denotes the Euclidean norm of a vector). In particular, we have the approximate Taylor expansion

$$P(t) = \sum_{j=0}^{D-1} \frac{(t-n)^j}{j!} P^{(j)}(n) + O(\varepsilon \|v(n)\|).$$

Since $|P(t)| \leq 1$ for $t \in [n, n+1]$, we conclude that

$$\sum_{j=0}^{D-1} \frac{(t-n)^j}{j!} P^{(j)}(n) \ll 1 + \varepsilon \|v(n)\|$$

for $t \in [n, n+1]$. From (27) applied to the polynomial in $t$ on the left-hand side we have that

$$|P^{(j)}(n)| \ll 1 + \varepsilon \|v(n)\|.$$

We conclude that

$$\|v(n)\| \ll 1 + \varepsilon \|v(n)\|$$

and hence for $\varepsilon$ small enough we see that all the components of $v(n)$ are $O(1)$. Inserting this back into (174) we conclude that (29) holds for all $m \leq D-1$; the remaining cases then follow by differentiating the equation (173) $m - D$ times and using induction on $m$. □

## Appendix B. The Baker–Campbell–Hausdorff formula and its consequences

In this section, we review some standard facts about connected, simply connected nilpotent Lie groups $G$ and their Lie algebras $\log G$. As mentioned in Section 4, all connected, simply connected nilpotent Lie groups are isomorphic to matrix algebras, so we shall abuse notation in this appendix by viewing elements of $G$ and $\log G$ as matrices (in particular we identify the Lie group exponential with the matrix exponential).

If $G$ is a simply connected nilpotent Lie group with some filtration $(G_i)_{i \geq 0}$ with $G_i = 0$ for $i > k$, we can define the operation $* : \log G \times \log G \to \log G$ by the formula

$$X * Y \coloneqq \log(\exp(X)\exp(Y)) \tag{175}$$

for all $X, Y \in \log G$, or equivalently

$$\log(gh) = \log g * \log h$$

for all $g, h \in G$. For instance, in the Heisenberg group example from Example 4.1, we have

$$\begin{pmatrix} 0 & x_1 & z_1 \\ 0 & 0 & y_1 \\ 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 0 & x_2 & z_2 \\ 0 & 0 & y_2 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & x_1 + x_2 & z_1 + z_2 + \frac{x_1 y_2 - x_2 y_1}{2} \\ 0 & 0 & y_1 + y_2 \\ 0 & 0 & 0 \end{pmatrix}.$$

The operation $*$ is clearly a group operation on $\log G$ (with identity 0 and inverse map $X \mapsto -X$). The *Baker–Campbell–Hausdorff formula* gives an explicit description of this operation. As is well known, $\log G$ is a nilpotent Lie algebra, using the usual matrix commutator $[X, Y] \coloneqq XY - YX$ as the Lie bracket; see [20, Corollary 11.2.7]. For any $X \in \log G$, we can then define the adjoint representation $\mathrm{ad}_X : \log G \to \log G$ to be linear map

$$\mathrm{ad}_X(Y) \coloneqq [X, Y].$$

As $\log G$ is a nilpotent Lie algebra, $\mathrm{ad}_X$ is a nilpotent linear transformation, thus $\mathrm{ad}_X^m = 0$ for some natural number $m$; more generally, for any $X, Y \in \log G$, any word in $\mathrm{ad}_X, \mathrm{ad}_Y$ of length greater than or equal to some threshold $m$ will vanish (in fact, by the inclusion (179) established below, one can take $m$ to equal the degree $k$ of the filtration). The Baker–Campbell–Hausdorff formula then states

$$X * Y = X + \int_0^1 \psi(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}) Y \, dt,$$

where $e^{\mathrm{ad}_X} = \sum_{n=0}^\infty \frac{1}{n!}\mathrm{ad}_X^n$ is the matrix exponential of $\mathrm{ad}_X$, and $\psi$ is the function

$$\psi(x) \coloneqq \frac{x\log x}{x-1} = 1 + \frac{x-1}{2} - \frac{(x-1)^2}{6} + \dots;$$

see for instance [18, Theorem 3.3] or [20, Proposition 3.4.4]. Note that from the nilpotent nature of $\log G$ that we can truncate the Taylor series for the matrix exponential and the function $\psi$ to some finite threshold $m$, so that

$$X * Y = X + Y + P(\mathrm{ad}_X, \mathrm{ad}_Y)Y \tag{176}$$

for some (non-commutative) polynomial $P$ of two variables of bounded degree and coefficients that are rational numbers of bounded height, where the constant term of $P$ vanishes and the linear term is equal to $\frac{1}{2}\mathrm{ad}_X$ (the contribution of $\mathrm{ad}_Y$ can be deleted from the linear term since $\mathrm{ad}_Y Y = 0$). The first few terms of this formula are

$$\begin{aligned} X * Y &= X + Y + \frac{1}{2}\mathrm{ad}_X Y + \frac{1}{12}(\mathrm{ad}_X^2 - \mathrm{ad}_Y\mathrm{ad}_X)Y + \dots \\ &= X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}([X,[X,Y]] - [Y,[X,Y]]) + \dots, \end{aligned}$$

although we will not need the explicit form of these terms beyond the quadratic case. From (176) we conclude in particular that $X * Y$ is a polynomial combination of $X, Y$, with bounded degree and coefficients. As one particular consequence of this formula, we see that

$$(tX) * (tY) * (-tX) * (-tY) = t^2[X,Y] + O(t^3)$$

as $t \to 0$ for any $X, Y \in \log G$, so the Lie bracket can be recovered from $*$ by the limiting formula

$$[X,Y] = \lim_{t\to 0} \frac{(tX) * (tY) * (-tX) * (-tY)}{t^2}, \tag{177}$$

which can also be established directly from (175) and Taylor expansion of the matrix exponential (this is also [20, (3.14)]).

Another closely related identity to the Baker–Campbell–Hausdorff formula is

$$e^{\mathrm{ad}_X} Y = \exp(X) Y \exp(-X)$$

for any $X, Y \in \log G$; see [18, Proposition 2.25]. As $\exp(C^{-1}YC) = C^{-1}\exp(Y)C$ for any invertible $C$, we have

$$\exp(\exp(X)Y\exp(-X)) = \exp(X)\exp(Y)\exp(-X)$$

and thus

$$\exp(e^{\mathrm{ad}_X}Y) = \exp(X)\exp(Y)\exp(-X)$$

for all $X, Y \in \log G$, or equivalently

$$\log(hgh^{-1}) = e^{\mathrm{ad}_{\log h}} \log g \tag{178}$$

for all $g, h \in G$.

By definition, the groups $G_i$ in the filtration $(G_i)$ are closed subgroups of $G$, and thus are themselves Lie groups with a Lie algebra $\log G_i$ that are subalgebras of $\log G$; see [20, Proposition 9.3.9], [18, Proposition 3.14]. In particular, the exponential map $\exp : \log G \to$

$G$ descends to a diffeomorphism $\exp : \log G_i \to G_i$, so $G_i$ is simply connected. The group $G_i$ is nilpotent simply connected, and $G_{i+1}$ is a closed simply connected nilpotent subgroup, thus $G_i/G_{i+1}$ is simply connected. If $X \in \log G_i$ and $Y \in \log G_j$, then from the filtration property $[G_i, G_j] \subset G_{i+j}$ and (175) we see that $(tX) * (tY) * (-tX) * (-tY) \in \log G_{i+j}$ for any $t > 0$; inserting this into (177) we conclude that $[X, Y] \in \log G_{i+j}$, thus we have the Lie algebra filtration property

$$[\log G_i, \log G_j] \subset \log G_{i+j}. \tag{179}$$

In particular, each of the $\log G_i$ are normal Lie subalgebras of $\log G$. From the Baker–Campbell–Hausdorff formula (176) and (179) we then also have

$$X * Y = X + Y \bmod \log G_{i+1}$$

whenever $i \geq 1$ and $X, Y \in \log G_i$, or equivalently

$$\log(gh) = \log(g) + \log(h) \bmod \log G_{i+1} \tag{180}$$

whenever $i \geq 1$ and $g, h \in G_i$. Thus $\log G_i / \log G_{i+1}$ is an abelian Lie algebra for any $i \geq 1$, and the logarithm map descends to a homomorphism from the multiplicative group $G_i/G_{i+1}$ to the additive group $\log G_i / \log G_{i+1}$.

**Lemma B.1** (Taylor expansion)**.** *Let $d \geq 1$ be a natural number, and let $g \in \mathrm{Poly}(\mathbb{Z} \to G)$. Then there exist unique* Taylor coefficients *$g_j \in G_j$ such that*

$$g(n) = \prod_j g_j^{\binom{n}{j}}.$$

*Proof.* This is a special case of [17, Lemma B.9]. □

Now we can prove Lemma 4.2.

*Proof of Lemma 4.2.* We may rescale $\delta = 1$. The fact that $\mathrm{Poly}(\mathbb{Z} \to G)$ forms a group is the Leibman–Lazard theorem; see e.g., [17, Corollary B.4]. Now suppose that $\tilde{g} \in \mathrm{Poly}(\mathbb{R} \to G)$, thus we have a Taylor expansion

$$\log \tilde{g}(t) = \sum_{i=0}^{k} X_i t^i$$

for some $X_i \in \log G_i$. For any $j \geq 0$, let $V_j$ denote the vector space of polynomial maps $p \colon \mathbb{R} \to \log G$ of the form

$$p(t) = \sum_{0 \leq i \leq k-j} Y_i t^i$$

where $Y_i \in \log G_{i+j}$ for all $i$, thus $\log \tilde{g} \in V_0$. One can check that the $V_j$ are decreasing with

$$[V_i, V_j] \subset V_{i+j} \tag{181}$$

and $V_i = 0$ for $i > k$; in particular, the $V_i$ are each Lie algebras. We now claim by induction that

$$\log \partial_{h_1} \dots \partial_{h_j} \tilde{g} \in V_j$$

for all $j \geq 0$ and $h_1, \dots, h_j \in \mathbb{R}$. This claim is already established for $j = 0$. If it holds for some $j$, and $h_{j+1} \in \mathbb{R}$, then by using the fact that $(t+h_{j+1})^i$ differs from $t^i$ by a polynomial of degree at most $i-1$ in $t$, we see that

$$\log \partial_{h_1} \dots \partial_{h_j} \tilde{g}(\cdot + h_{j+1}) = \log \partial_{h_1} \dots \partial_{h_j} \tilde{g} \bmod V_{j+1}$$

and hence by the Baker–Campbell–Hausdorff formula (176)

$$\log \partial_{h_1} \dots \partial_{h_j} \tilde{g}(\cdot + h_{j+1}) * (-\log \partial_{h_1} \dots \partial_{h_j} \tilde{g}) \in V_{j+1}.$$

But by (175) the left-hand side is equal to $\log \partial_{h_1} \dots \partial_{h_{j+1}} \tilde{g}$, closing the induction. Applying this with $j = k$ and $h_1, \dots, h_j, t \in \mathbb{Z}$, we conclude that the restriction of $\tilde{g}$ to $\mathbb{Z}$ lies in $\mathrm{Poly}(\mathbb{Z} \to G)$.

Now suppose that $g \in \mathrm{Poly}(\mathbb{Z} \to G)$. Any such element can be expressed uniquely as a Taylor expansion

$$g(n) = g_0 g_1^{\binom{n}{1}} \dots g_k^{\binom{n}{k}}$$

for all $n \in \mathbb{Z}$ and some $g_j \in G_j$; see [17, Lemma B.9]. Using the real exponentiation (77), we can extend $g$ to the map

$$\tilde{g}(t) = g_0 g_1^{\binom{t}{1}} \dots g_k^{\binom{t}{k}} \tag{182}$$

and from many applications of the Baker–Campbell–Hausdorff formula (175), (176), (179) one sees that $\tilde{g}$ is now an element of $\mathrm{Poly}(\mathbb{R} \to G)$. This establishes existence. To show uniqueness, it suffices by the group property to check the case $g = 1$. Then any extension $\tilde{g}$ is such that $\log \tilde{g}(n) = 0$ for every integer $n$; since $\log \tilde{g}$ is also a polynomial, $\log \tilde{g}$ vanishes identically, hence $\tilde{g}$ must be 1, giving uniqueness. □

As a corollary we obtain:

**Lemma B.2** (Non-abelian Discrete Taylor expansion)**.** *For any $\delta > 0$, the space* $\mathrm{Poly}(\delta\mathbb{Z} \to G)$ *consists precisely of those functions $\gamma : \mathbb{R} \to G$ of the form*

$$\gamma(t) \coloneqq \prod_{j=0}^{k} g_j^{\binom{t/\delta}{j}}$$

*for some $g_j \in G_j$, where $\binom{x}{j} \coloneqq \frac{x(x-1)\dots(x-j+1)}{j!}$.*

If $\Gamma$ is a cocompact discrete subgroup of $G$ with each $\Gamma_i \coloneqq \Gamma \cap G_i$ cocompact in $G_i$, then there exists a *Mal'cev basis* for $\Gamma$, by which we mean a linear basis $X_1, \dots, X_{\dim G}$ for $\log G$ with the property that $X_{\dim G - \dim G_i + 1}, \dots, X_{\dim G}$ form a basis for $\log G_i$ for each $i$ (so in particular $[X_i, X_j]$ lies in the span of $X_{\max(i,j)+1}, \dots, X_{\dim G}$ for any $1 \leq i, j \leq \dim G$), and

$$\Gamma = \{\exp(n_1 X_1) \cdots \exp(n_{\dim G} X_{\dim G}) : n_1, \dots, n_{\dim G} \in \mathbb{Z}\}.$$

See [16, §2] for details. From this and many applications of the Baker–Campbell–Hausdorff formula, we see that for any $1 \leq i, j \leq \dim G$, the coefficients of $[X_i, X_j]$ in the basis

$X_{\max(i,j)+1}, \dots, X_{\dim G}$ are rational numbers with denominator $O(1)$, and thus every element of $\Gamma$ can be written in the form

$$\exp(\frac{1}{Q_1}(n_1 X_1 + \cdots + n_{\dim G} X_{\dim G})) \tag{183}$$

for some integers $n_1, \dots, n_{\dim G}$ and some natural number $Q_1 = O(1)$ depending only on $G$ and the Mal'cev basis; conversely, there exists a natural number $Q_2 = O(1)$ such that every expression of the form

$$\exp(Q_2(n_1 X_1 + \cdots + n_{\dim G} X_{\dim G}))$$

with $n_1, \dots, n_{\dim G} \in \mathbb{Z}$ lies in $\Gamma$. One consequence of this and the Baker–Campbell–Hausdorff formula is that, for any fixed natural number $q = O(1)$, the set $\{\gamma \in G : \gamma^q \in \Gamma\}$ generates a group, all of whose elements are of the form

$$\exp(\frac{1}{Q}(n_1 X_1 + \cdots + n_{\dim G} X_{\dim G}))$$

for some $Q$ depending on $G$, $q$, and the Mal'cev basis; in particular, this group contains only finitely many cosets of $\Gamma$, so that $\Gamma$ is a finite index subgroup of it. As one particular corollary of this, we see that if $\gamma_1, \gamma_2 \in G$ are such that $\gamma_1^{q_1}, \gamma_2^{q_2} \in \Gamma$ for some natural numbers $q_1, q_2 = O(1)$, then one has $(\gamma_1 \gamma_2)^q \in \Gamma$ for some $q = O(1)$.

## Appendix C. Bezout's identity and the Chinese remainder theorem for polynomial spaces

In this section, we prove various versions of Bezout's identity and the Chinese remainder theorem for polynomial maps, either into the circle $\mathbb{R}/\mathbb{Z}$ or into more general filtered nilpotent Lie groups.

### C.1. **Bezout-type identities.**

*Proof of Lemma 2.2.* We may normalize $\lambda = 1$. We begin with the first claim. It suffices to establish the inclusion

$$\mathrm{Poly}_{\leq k}\left(\frac{1}{a}\mathbb{Z} \to \mathbb{Z}\right) + \mathrm{Poly}_{\leq k}\left(\frac{1}{b}\mathbb{Z} \to \mathbb{Z}\right) \supset \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$$

as the opposite inclusion is trivial. That is, it suffices to show that every $\gamma \in \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$ may be split as $\gamma = \gamma_a + \gamma_b$ where $\gamma_a \in \mathrm{Poly}_{\leq k}\left(\frac{1}{a}\mathbb{Z} \to \mathbb{Z}\right)$ and $\gamma_b \in \mathrm{Poly}_{\leq k}\left(\frac{1}{b}\mathbb{Z} \to \mathbb{Z}\right)$.

We prove this by induction on $k$. The claim is trivial for $k = 0$, so suppose that $k \geq 1$ and that the claim has already been proven for $k - 1$. From Lemma 2.1 we can write $\gamma(t) = c\binom{t}{k} + \gamma^*(t)$ for some integer $c$ and $\gamma^* \in \mathrm{Poly}_{\leq k-1}(\mathbb{R} \to \mathbb{R})$. By Bezout's identity we may write $c = qa^k + rb^k$ for some integers $q, r$, thus

$$\gamma(t) = q\binom{at}{k} + r\binom{bt}{k} + \gamma^{**}(t)$$

for some $\gamma^{**} \in \mathrm{Poly}_{\leq k-1}(\mathbb{R} \to \mathbb{R})$. As $\gamma(\mathbb{Z}) \subset \mathbb{Z}$, also $\gamma^{**}(\mathbb{Z}) \subset \mathbb{Z}$; so by the induction hypothesis we may write $\gamma^{**}(t) = \gamma_a^{**}(t) + \gamma_b^{**}(t)$ where $\gamma_a^{**} \in \mathrm{Poly}_{\leq k-1}\left(\frac{1}{a}\mathbb{Z} \to \mathbb{Z}\right)$ and $\gamma_b^{**} \in \mathrm{Poly}_{\leq k-1}\left(\frac{1}{b}\mathbb{Z} \to \mathbb{Z}\right)$. Setting $\gamma_a(t) \coloneqq q\binom{at}{k} + \gamma_a^{**}(t)$ and $\gamma_b(t) \coloneqq r\binom{bt}{k} + \gamma_b^{**}(t)$ closes the induction.

Now we prove the second claim. Again it suffices to prove the inclusion

$$\mathrm{Poly}_{\leq k}\left(\frac{1}{a}\mathbb{Z} \to \mathbb{Z}\right) \cap \mathrm{Poly}_{\leq k}\left(\frac{1}{b}\mathbb{Z} \to \mathbb{Z}\right) \subset \mathrm{Poly}_{\leq k}\left(\frac{1}{ab}\mathbb{Z} \to \mathbb{Z}\right)$$

as the opposite inclusion is trivial, and we may again inductively assume that $k \geq 1$ and that the claim has already been proven for $k-1$.

If $\gamma \in \mathrm{Poly}_{\leq k}\left(\frac{1}{a}\mathbb{Z} \to \mathbb{Z}\right) \cap \mathrm{Poly}_{\leq k}\left(\frac{1}{b}\mathbb{Z} \to \mathbb{Z}\right)$, then from Lemma 2.1 we see that the derivative $\gamma^{(k)}$ (which is a constant) is an integer multiple of both $a^k$ and $b^k$, hence can be written as $c(ab)^k$ for some integer $c$. Thus we may write $\gamma(t) = c\binom{abt}{k} + \gamma^*(t)$ for some integer $c$ and $\gamma^* \in \mathrm{Poly}_{\leq k-1}(\mathbb{R} \to \mathbb{R})$. One then easily checks that

$$\gamma^* \in \mathrm{Poly}_{\leq k-1}\left(\frac{1}{a}\mathbb{Z} \to \mathbb{Z}\right) \cap \mathrm{Poly}_{\leq k-1}\left(\frac{1}{b}\mathbb{Z} \to \mathbb{Z}\right)$$

and the claim now follows from the induction hypothesis and Lemma 2.1. □

*Proof of Lemma 4.13.* We again normalize $\lambda = 1$. We begin with the first claim. As $\mathrm{Poly}(\mathbb{Z} \to \Gamma)$ is a group that contains[23] $\mathrm{Poly}(\frac{1}{a}\mathbb{Z} \to \Gamma), \mathrm{Poly}(\frac{1}{b}\mathbb{Z} \to \Gamma)$, we clearly have the inclusion

$$\mathrm{Poly}(\frac{1}{a}\mathbb{Z} \to \Gamma) \cdot \mathrm{Poly}(\frac{1}{b}\mathbb{Z} \to \Gamma) \subset \mathrm{Poly}(\mathbb{Z} \to \Gamma)$$

and it now suffices to show that any $\gamma \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$ can be factored as $\gamma = \gamma_a\gamma_b$, where $\gamma_a \in \mathrm{Poly}(\frac{1}{a}\mathbb{Z} \to \Gamma)$ and $\gamma_b \in \mathrm{Poly}(\frac{1}{b}\mathbb{Z} \to \Gamma)$.

Set $\Gamma_i \coloneqq G_i \cap \Gamma$ for all $i$. If $\gamma$ lies in $\mathrm{Poly}(\mathbb{Z} \to \Gamma_{k+1})$ then the claim is trivial since $\Gamma_{k+1} = \{1\}$, so now suppose by downward induction that $\gamma$ lies in $\mathrm{Poly}(\mathbb{Z} \to \Gamma_i)$ for some $1 \leq i \leq k$, and that the claim has already been proven for $\gamma$ in $\mathrm{Poly}(\mathbb{Z} \to \Gamma_{i+1})$. By Lemma B.2 we have a Taylor expansion of the form

$$\gamma(t) = \prod_j \gamma_j^{\binom{t}{j}}.$$

Since for $t \in \mathbb{Z}$ we have $\gamma(t) \in \Gamma_i$ we get by induction on $n$ that $\gamma_j \in \Gamma_i$. If we let $\pi_i \colon \Gamma_i \to \Gamma_i/\Gamma_{i+1}$ be the quotient map, then since $\Gamma_i/\Gamma_{i+1}$ is abelian we get for $t \in \mathbb{Z}$

$$\pi_i(\gamma(t)) = \prod_{j=0}^{i} \pi_i(\gamma_j)^{\binom{t}{j}}.$$

[23]We remind here that, by Lemma 4.2, the group $\mathrm{Poly}(\delta\mathbb{Z} \to \Gamma)$ can be (by an abuse of notation) interpreted as a subgroup of $\mathrm{Poly}(\mathbb{R} \to \Gamma)$.

By Lemma 2.2, we can split each $\binom{t}{j}$ as $P_{a,j}(t) + P_{b,j}(t)$ for $t \in \mathbb{R}$ and some $P_{a,j} \in \mathrm{Poly}_{\leq j}(\frac{1}{a}\mathbb{Z} \to \mathbb{Z})$ and $P_{b,j} \in \mathrm{Poly}_{\leq j}(\frac{1}{b}\mathbb{Z} \to \mathbb{Z})$. Setting

$$\gamma'_a(t) \coloneqq \prod_{j=0}^{i} \gamma_j^{P_{a,j}(t)}; \quad \gamma'_b(t) \coloneqq \prod_{j=0}^{i} \gamma_j^{P_{b,j}(t)}$$

for all $t \in \mathbb{R}$, we see that $\gamma'_a \in \mathrm{Poly}(\frac{1}{a}\mathbb{Z} \to \Gamma)$, $\gamma'_b \in \mathrm{Poly}(\frac{1}{b}\mathbb{Z} \to \Gamma)$, and

$$\gamma = \gamma'_a \sigma \gamma'_b$$

for some $\sigma \in \mathrm{Poly}(\mathbb{Z} \to \Gamma_{i+1})$. The claim now follows from the induction hypothesis.

Now we prove the second claim. We show by downwards induction on $k$ that for each $1 \leq i \leq k+1$ and $\gamma \in \mathrm{Poly}(\frac{1}{a}\mathbb{Z} \to \Gamma_i) \cap \mathrm{Poly}(\frac{1}{b}\mathbb{Z} \to \Gamma_i)$ one has $\gamma \in \mathrm{Poly}(\frac{1}{ab}\mathbb{Z} \to \Gamma_i)$. The claim is trivially true for $i = k+1$, so suppose that $1 \leq i \leq k$ and that the claim has already been proven for $i+1$. From two applications of Lemma B.2 and with $\pi_i$ as above, we have

$$\pi_i(\gamma(t)) = \prod_{j=0}^{i} \pi_i(\gamma_{j,a})^{\binom{at}{j}} \tag{184}$$

for all $t \in \frac{1}{a}\mathbb{Z}$ and some $\gamma_{j,a} \in \Gamma_i$, and

$$\pi_i(\gamma(t)) = \prod_{j=0}^{i} \pi_i(\gamma_{j,b})^{\binom{bt}{j}} \tag{185}$$

for all $t \in \frac{1}{b}\mathbb{Z}$ and some $\gamma_{j,b} \in \Gamma_i$. Specializing to $t \in \mathbb{Z}$ and comparing the top order coefficients of these polynomials (using the uniqueness of the Taylor expansion) in the abelian group $\Gamma_i/\Gamma_{i+1}$, we conclude that

$$\pi_i(\gamma_{i,a})^{a^i} = \pi_i(\gamma_{i,b})^{b^i}.$$

As $a^i, b^i$ are coprime, the Bezout identity allows one to express 1 as an integer combination of $a^i, b^i$. We conclude that there exists $\gamma_i \in \Gamma_i$ such that $\pi_i(\gamma_{i,a}) = \pi(\gamma_i)^{b^i}$ and $\pi_i(\gamma_{i,b}) = \pi(\gamma_i)^{a^i}$. If one then divides out the polynomial $t \mapsto \gamma_i^{\binom{abt}{i}}$ (which lies in $\mathrm{Poly}(\frac{1}{ab}\mathbb{Z} \to \Gamma_i)$) from $\gamma$ (either on the right or left), one ends up with a polynomial in $\gamma \in \mathrm{Poly}(\frac{1}{a}\mathbb{Z} \to \Gamma_i) \cap \mathrm{Poly}(\frac{1}{b}\mathbb{Z} \to \Gamma_i)$ which has an expansion similar to that of (184), (185) but with the $j = i$ term absent. Repeating this argument we may eliminate all the other factors in (184), (185) by dividing out appropriate sequences in $\mathrm{Poly}(\frac{1}{ab}\mathbb{Z} \to \Gamma_i)$, until $\pi_i(\gamma(n))$ is identically equal to 1 on both $\frac{1}{a}\mathbb{Z}$ and $\frac{1}{b}\mathbb{Z}$, so that $\gamma$ now lies in $\mathrm{Poly}(\frac{1}{a}\mathbb{Z} \to \Gamma_{i+1}) \cap \mathrm{Poly}(\frac{1}{b}\mathbb{Z} \to \Gamma_{i+1})$, and the claim now follows from the induction hypothesis. □

## C.2. **Chinese remainder theorems.**

*Proof of Proposition 3.5.* We begin by proving an auxiliary claim, namely that if $a_1, \dots, a_m$ are coprime natural numbers, and $\gamma_1, \dots, \gamma_m \in \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$, then there exists $\gamma \in \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$ such that $\gamma_i - \gamma \in \mathrm{Poly}_{\leq k}(\frac{1}{a_i}\mathbb{Z} \to \mathbb{Z})$ for $i = 1, \dots, m$. It suffices to verify this when $m = 2$, as this also implies the $m = 1$ case, and the higher $m$ cases also follow

from induction. From the first claim of Lemma 2.2 we can write $\gamma_1 - \gamma_2 = \gamma_1^* - \gamma_2^*$ where $\gamma_1^* \in \mathrm{Poly}_{\leq k}(\frac{1}{a_1}\mathbb{Z} \to \mathbb{Z})$ and $\gamma_2^* \in \mathrm{Poly}_{\leq k}(\frac{1}{a_2}\mathbb{Z} \to \mathbb{Z})$. The claim now follows by setting $\gamma := \gamma_1 - \gamma_1^* = \gamma_2 - \gamma_2^*$.

Now we prove (i). Write $\phi = (I, P)$ and $\phi_p = (I_p, P_p)$. From Definition 3.1, we have

$$P_p = \varepsilon_p + P + \gamma_p$$

where $\varepsilon_p \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ obeys the smoothness bounds in Definition 3.1(i), and $\gamma_p \in \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$. From the previous claim, there exists $\gamma \in \mathrm{Poly}_{\leq k}(\mathbb{Z} \to \mathbb{Z})$ such that $\gamma_p - \gamma \in \mathrm{Poly}_{\leq k}(\frac{1}{p}\mathbb{Z} \to \mathbb{Z})$ for each $p$. If one then sets $\tilde{\phi} := (I, P + \gamma)$, one obtains the claim (i).

Now we prove (ii). Write $\phi = (I, P)$ and $\phi' = (I', P')$. From hypothesis we may write

$$P(t) = \epsilon_p(t) + P'(t) + \gamma_p(t)$$

for all $p \in \mathcal{P}$ and some $\epsilon_p, \gamma_p \in \mathrm{Poly}_{\leq k}(\mathbb{R} \to \mathbb{R})$ obeying the properties in Definition 3.1. In particular, we see that $\epsilon_p(t) + \gamma_p(t)$ is independent of $p$. Setting $n_I$ to be an integer point in $I$, we then have that $\epsilon_p(n_I)$ mod 1 is independent of $p$. Since also $\epsilon_p(n_I) = O(1)$, we may subtract a bounded integer from each $\epsilon_p$ and add it to $\gamma_p$ to assume without loss of generality that $\epsilon_p(n_I)$ is independent of $p$. Since $\epsilon_p(n+1) = \epsilon_p(n) + O(1/|I|)$ for all $n \in I \cap \mathbb{Z}$, and $\epsilon_p(n)$ mod 1 is independent of $p$, we conclude from induction (for $|I|$ large enough) that $\epsilon_p(n)$ is independent of $p$ for all $n \in I \cap \mathbb{Z}$, which by Lagrange interpolation (or Lemma 2.3) implies that $\epsilon_p = \epsilon$ is independent of $p$. This implies that $\gamma_p = \gamma$ is also independent of $p$. Since $\gamma \in \mathrm{Poly}_{\leq k}(\frac{1}{p}\mathbb{Z} \to \mathbb{Z})$ for all $p \in \mathcal{P}$, we see from iterating the second claim of Lemma 2.2 that $\gamma \in \mathrm{Poly}_{\leq k}(\frac{1}{\prod \mathcal{P}}\mathbb{Z} \to \mathbb{Z})$, and the claim follows. □

*Proof of Proposition 4.14.* As with the proof of Proposition 3.5, we begin by proving an auxiliary claim, namely that if $a_1, \ldots, a_m$ are coprime natural numbers, and $\gamma_1, \ldots, \gamma_m \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$, then there exists $\gamma \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$ such that $\gamma^{-1}\gamma_i \in \mathrm{Poly}(\frac{1}{a_i}\mathbb{Z} \to \Gamma)$ for $i = 1, \ldots, m$. As before it suffices from induction to verify the $m = 2$ case. From the first claim of Lemma 4.13 we can write $\gamma_1^{-1}\gamma_2 = (\gamma_1^*)^{-1}\gamma_2^*$ where $\gamma_1^* \in \mathrm{Poly}(\frac{1}{a_1}\mathbb{Z} \to \Gamma)$ and $\gamma_2^* \in \mathrm{Poly}(\frac{1}{a_2}\mathbb{Z} \to \Gamma)$. The claim now follows by setting $\gamma := \gamma_1(\gamma_1^*)^{-1} = \gamma_2(\gamma_2^*)^{-1}$.

Now we prove (i). From Definition 4.9, if we write $\phi = (I, g)$ and $\phi_p = (I_p, g_p)$, we have

$$g_p = \varepsilon_p g \gamma_p$$

where $\varepsilon_p \in \mathrm{Poly}(\mathbb{R} \to G)$ obeys the smoothness bounds in Definition 3.1(i), and $\gamma_p \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$. From the previous claim, there exists $\gamma \in \mathrm{Poly}(\mathbb{Z} \to \Gamma)$ such that $\gamma^{-1}\gamma_p \in \mathrm{Poly}(\frac{1}{p}\mathbb{Z} \to \mathbb{Z})$ for each $p$. If one then sets $\phi' := (I, g\gamma)$, one obtains the claim (i).

Now we prove (ii). Write $\phi = (I, g)$ and $\phi' = (I', g')$. From hypothesis we may write

$$g = \epsilon_p g' \gamma_p \tag{186}$$

for all $p \in \mathcal{P}$ and some $\epsilon_p, \gamma_p \in \mathrm{Poly}(\mathbb{R} \to G)$ obeying the properties in Definition 4.9. Let $n_I$ be an integer point in $I$. The points $\log \epsilon_p(n_I)$ take values in a ball of size $O(1)$ around the origin in $\log G$. Let $\delta > 0$ be a small, fixed constant (depending on $k, \varepsilon, \theta, G/\Gamma, F$). By the pigeonhole principle, one can find a subcollection $\mathcal{P}'$ of $\mathcal{P}$ with $\#\mathcal{P}' \gg_\delta \#\mathcal{P}$ such that

$\log \epsilon_p(n_I) = \epsilon_0 + O(\delta)$ for some $\epsilon_0 = O(1)$. From Bernstein's inequality (26) (applied to the function that expresses the distance between $\log \varepsilon_p(t)$ and $\varepsilon_0$) we also have $\log \epsilon_p(t) = \epsilon_0 + O(\delta)$ whenever $t = n_I + O(\delta|I|)$. From (186) one has

$$(g')^{-1}\epsilon_p^{-1}\epsilon_{p'}g' = \gamma_p\gamma_{p'}^{-1}. \tag{187}$$

Now suppose that $t$ is an integer with $t = n_I + O(\delta|I|)$. By the Baker–Campbell–Hausdorff formula (176), the quantity

$$\epsilon_p(t)^{-1}\epsilon_{p'}(t) = \exp((-\log \epsilon_p(t)) * \log \epsilon_{p'}(t)) = \exp((-\epsilon_0 + O(\delta)) * (\epsilon_0 + O(\delta)))$$

lies within $O(\delta)$ of the identity, hence the conjugate $g'(t)^{-1}\epsilon_p(t)^{-1}\epsilon_{p'}(t)g'(t)$ lies within $O(\delta)$ of the identity when projected to the abelian group $G/G_2$. On the other hand by (183), the projection of $\gamma_p(t)\gamma_{p'}(t)^{-1}$ to $G/G_2$ is rational in the sense that it lies in the image of $\Gamma$ when raised to some power $q = O(1)$. For $\delta$ small enough, these facts are only compatible if the projection of both sides of (187) to $G/G_2$ is trivial, that is to say both sides of (187) lie in $G_2$, so $\epsilon_p(t)^{-1}\epsilon_{p'}(t)$ also lies in $G_2$. Now one can project to the abelian group $G_2/G_3$ and repeat the above arguments to show that both sides of (187) lie in $G_3$ (for $\delta$ small enough). Continuing this argument we conclude that both sides of (187) are in fact trivial for all integers $t = n_I + O(\delta|I|)$, and hence by Lagrange interpolation (for $|I|$ large enough) for all real $t$ also. In particular, $\gamma_p = \gamma$ is independent of $p$. From the second part of Lemma 4.13 we conclude that $\gamma \in \mathrm{Poly}(\frac{1}{\prod \mathcal{P}'}\mathbb{Z} \to \mathbb{Z})$, and the claim follows. □

## References


[1] H. A. Helfgott and M. Radziwiłł. Expansion, divisibility and parity. *arXiv e-prints*, page arXiv:2103.06853, March 2021.

[2] V. Bergelson and A. Leibman. Polynomial extensions of van der Waerden's and Szemerédi's theorems. *J. Amer. Math. Soc.*, 9(3):725–753, 1996.

[3] G. R. Blakley and P. Roy. A Hölder type inequality for symmetric matrices with nonnegative entries. *Proc. Amer. Math. Soc.*, 16:1244–1245, 1965.

[4] S. Chowla. *The Riemann hypothesis and Hilbert's tenth problem.* Mathematics and Its Applications, Vol. 4. Gordon and Breach Science Publishers, New York-London-Paris, 1965.

[5] T. Cochrane and Z. Zheng. Pure and mixed exponential sums. *Acta Arith.*, 91(3):249–278, 1999.

[6] S. Ferenczi, J. Kułaga-Przymus, and M. Lemańczyk. Sarnak's conjecture: what's new. In *Ergodic theory and dynamical systems in their interactions with arithmetics and combinatorics*, volume 2213 of *Lecture Notes in Math.*, pages 163–235. Springer, Cham, 2018.

[7] L. Flaminio, K. Frączek, J. Kułaga-Przymus, and M. Lemańczyk. Approximate orthogonality of powers for ergodic affine unipotent diffeomorphisms on nilmanifolds. *Studia Math.*, 244(1):43–97, 2019.

[8] K. Ford. Vinogradov's integral and bounds for the Riemann zeta function. *Proc. London Math. Soc. (3)*, 85(3):565–633, 2002.

[9] N. Frantzikinakis and B. Host. Higher order Fourier analysis of multiplicative functions and applications. *J. Amer. Math. Soc.*, 30(1):67–157, 2017.

[10] N. Frantzikinakis and B. Host. The logarithmic Sarnak conjecture for ergodic weights. *Ann. of Math. (2)*, 187(3):869–931, 2018.

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

[12] A. Granville and K. Soundararajan. Large character sums: pretentious characters and the Pólya-Vinogradov theorem. *J. Amer. Math. Soc.*, 20(2):357–384, 2007.

[13] B. Green and T. Tao. An inverse theorem for the Gowers $U^3(G)$ norm. *Proc. Edinb. Math. Soc. (2)*, 51(1):73–153, 2008.
[14] B. Green and T. Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010.
[15] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)*, 175(2):541–566, 2012.
[16] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012.
[17] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$-norm. *Ann. of Math. (2)*, 176(2):1231–1372, 2012.
[18] B. Hall. *Lie groups, Lie algebras, and representations*, volume 222 of *Graduate Texts in Mathematics*. Springer, Cham, second edition, 2015. An elementary introduction.
[19] X. He and Z. Wang. Möbius disjointness for nilsequences along short intervals. *Trans. Amer. Math. Soc.*, 374(6):3881–3917, 2021.
[20] J. Hilgert and K-H. Neeb. *Structure and geometry of Lie groups*. Springer Monographs in Mathematics. Springer, New York, 2012.
[21] O. Klurman and A. P. Mangerel. On the orbits of multiplicative pairs. *Algebra Number Theory*, 14(1):155–189, 2020.
[22] K. Matomäki and M. Radziwiłł. A note on the Liouville function in short intervals. *arXiv e-prints*, page arXiv:1502.02374, Feb 2015.
[23] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals. *Ann. of Math. (2)*, 183(3):1015–1056, 2016.
[24] K. Matomäki and M. Radziwiłł. Multiplicative functions in short intervals II. *arXiv e-prints*, page arXiv:2007.04290, July 2020.
[25] K. Matomäki, M. Radziwiłł, and T. Tao. An averaged form of Chowla's conjecture. *Algebra Number Theory*, 9(9):2167–2196, 2015.
[26] K. Matomäki, M. Radziwiłł, and T. Tao. Fourier uniformity of bounded multiplicative functions in short intervals on average. *Invent. Math.*, 220(1):1–58, 2020.
[27] K. Matomäki and X. Shao. Discorrelation between primes in short intervals and polynomial phases. *Int. Math. Res. Not. IMRN*, (16):12330–12355, 2021.
[28] K. Matomäki and J. Teräväinen. On the Möbius function in all short intervals. To appear in J. Eur. Math. Soc., page arXiv:1911.09076, Nov 2019.
[29] R. McNamara. Sarnak's conjecture for sequences of almost quadratic word growth. *Ergodic Theory Dynam. Systems*, 41(10):3060–3115, 2021.
[30] H. P. Mulholland and C. A. B. Smith. An inequality arising in genetical theory. *Amer. Math. Monthly*, 66:673–683, 1959.
[31] V. V. Prasolov. *Polynomials*, volume 11 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2004. Translated from the 2001 Russian second edition by Dimitry Leites.
[32] P. Sarnak. Mobius randomness and dynamics. *Not. S. Afr. Math. Soc.*, 43(2):89–97, 2012.
[33] T. Tao. *Higher order Fourier analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.
[34] T. Tao. The logarithmically averaged Chowla and Elliott conjectures for two-point correlations. *Forum Math. Pi*, 4:e8, 36, 2016.
[35] T. Tao. Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. In *Number theory—Diophantine problems, uniform distribution and applications*, pages 391–421. Springer, Cham, 2017.
[36] T. Tao and J. Teräväinen. Odd order cases of the logarithmically averaged Chowla conjecture. *J. Théor. Nombres Bordeaux*, 30(3):997–1015, 2018.
[37] T. Tao and J. Teräväinen. The structure of logarithmically averaged correlations of multiplicative functions, with applications to the Chowla and Elliott conjectures. *Duke Math. J.*, 168(11):1977–2027, 2019.

[38] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
[39] T. Tao and T. Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Math.*, 201(2):213–305, 2008.
[40] T. Tao and T. Ziegler. Polynomial patterns in the primes. *Forum Math. Pi*, 6:e1, 60, 2018.
[41] T. Zhan. On the representation of large odd integer as a sum of three almost equal primes. *Acta Math. Sinica (N.S.)*, 7(3):259–272, 1991. A Chinese summary appears in Acta Math. Sinica **3**5 (1992), no. 4, 575.

Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland
*Email address*: ksmato@utu.fi

Department of Mathematics, Caltech, 1200 E California Blvd, Pasadena, CA, 91125, USA
*Email address*: maksym.radziwill@gmail.com

Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles, CA, 90095, USA
*Email address*: tao@math.ucla.edu

Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, United Kingdom

Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland
*Email address*: joni.p.teravainen@gmail.com

Einstein Institute of Mathematics, Givaat Ram The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem 91904, Israel
*Email address*: tamarz@math.huji.ac.il