

The Bombieri–Vinogradov Theorem for Nilsequences

Xuancheng Shao*

Joni Teräväinen†

Received 12 June 2020; Revised 8 April 2021; Published 11 October 2021

Abstract: We establish results of Bombieri–Vinogradov type for the von Mangoldt function $\Lambda(n)$ twisted by a nilsequence. In particular, we obtain Bombieri–Vinogradov type results for the von Mangoldt function twisted by any polynomial phase $e(P(n))$; the results obtained are as strong as the ones previously known in the case of linear exponential twists. We derive a number of applications of these results. Firstly, we show that the primes p obeying a “nil-Bohr set” condition, such as $\|\alpha p^k\| < \varepsilon$, exhibit bounded gaps. Secondly, we show that the Chen primes are well-distributed in nil-Bohr sets, generalizing a result of Matomäki. Thirdly, we generalize the Green–Tao result on linear equations in the primes to primes belonging to an arithmetic progression to large modulus $q \leq x^\theta$, for almost all q .

Key words and phrases: Bombieri–Vinogradov theorem, Nilsequences, Gowers norms

1 INTRODUCTION

The celebrated Bombieri–Vinogradov theorem states that

$$\sum_{d \leq x^{1/2-\varepsilon}} \max_{(c,d)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) - \frac{x}{\varphi(d)} \right| \ll_{A,\varepsilon} \frac{x}{(\log x)^A},$$

thus proving equidistribution of the von Mangoldt function in all residue classes to almost all moduli $d \leq x^{1/2-\varepsilon}$. The $x^{1/2-\varepsilon}$ threshold can further be improved to $x^{1/2}/(\log x)^B$ for suitable $B = B(A)$, but either with or without this improvement the conclusion can be stated as saying that the primes have *level of distribution* $1/2$.

*Supported by the NSF grant DMS-1802224.

†Supported by a Titchmarsh Research Fellowship.

It is natural to study whether other sequences related to the primes or various arithmetic functions also satisfy bounds of Bombieri–Vinogradov type. Bombieri–Vinogradov type estimates for general 1-bounded multiplicative functions were investigated in [15, 14], and there are a number of works on the level of distribution of the smooth numbers [9, 21] and of the k -fold divisor functions $d_k(n)$ (e.g. [12, 13, 22] for all moduli and [7, 8] for almost all moduli), and the literature is swarming with many other interesting examples besides.

Our object in this paper is to obtain level of distribution results for another natural class of functions, namely twists $\Lambda(n)e(P(n))$ of the von Mangoldt function by polynomial phases and, more generally, nilsequences. There has been previous work on the case of linear polynomials P ; we recall these results later in this introduction. Let us first present the necessary definitions for stating our main theorems; in Section 2 we present their applications.

1.1 Results for nilsequence twists

In order to state our results for nilsequence twists, we need a few definitions. The results for polynomial phases will be deduced as special cases in Subsection 1.2, and they do not require knowledge of nilsequences. For an in-depth discussion of nilsequences and their importance in additive combinatorics, see [34].

Definition 1.1 (Nilsequences). Let G be a connected, simply-connected nilpotent Lie group, and let $\Gamma \leq G$ be a lattice. By a *filtration* $G_\bullet = (G_i)_{i=0}^\infty$ on G , we mean an infinite sequence of subgroups of G (which are also connected, simply-connected nilpotent Lie groups) such that

$$G = G_0 = G_1 \supset G_2 \supset \cdots$$

and such that the commutators satisfy $[G_i, G_j] \subset G_{i+j}$, and with the additional properties that $\Gamma_i := \Gamma \cap G_i$ is a lattice in G_i for $i \geq 0$ and $G_{s+1} = \{\text{id}\}$ for some s .

The least such s is called the *degree* of $G_\bullet$.

A *polynomial sequence* on G (adapted to the filtration $G_\bullet$) is any sequence $g : \mathbb{Z} \rightarrow G$ satisfying the derivative condition

$$\partial_{h_1} \cdots \partial_{h_k} g(n) \in G_k$$

for all $k \geq 0$, $n \in \mathbb{Z}$ and all $h_1, \dots, h_k \in \mathbb{Z}$, where $\partial_h g(n) := g(n+h)g(n)^{-1}$ is the discrete derivative with shift h .

Finally, if $\varphi : G/\Gamma \rightarrow \mathbb{C}$ is Lipschitz with respect to a natural metric on G/Γ (induced by a Mal'cev basis; see definition below), we call a sequence of the form $n \mapsto \varphi(g(n)\Gamma)$ a *nilsequence*.

Since nilsequences are a vast class of functions, it is natural to restrict to those that have “bounded complexity”. This is made precise in the following definition.

Definition 1.2 (Bounded complexity nilsequences). For a positive integer s and real numbers $\Delta, K \geq 2$, we define $\Psi_s(\Delta, K)$ to be the class of all nilsequences $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ of the form $\psi(n) = \varphi(g(n)\Gamma)$, where

1. G/Γ is a nilmanifold of dimension at most Δ , equipped with a filtration $G_\bullet$ of degree at most s and a K -rational Mal'cev basis $\mathcal{X}$ (defined in [19, Definition 2.1, Definition 2.4]);
2. $g : \mathbb{Z} \rightarrow G$ is a polynomial sequence adapted to $G_\bullet$;

3. $\varphi: G/\Gamma \rightarrow \mathbb{C}$ is a Lipschitz function with $\|\varphi\|_{\text{Lip}} \leq 1$, where the Lipschitz norm is defined as

$$\|\varphi\|_{\text{Lip}} = \|\varphi\|_{\infty} + \sup_{x,y \in G/\Gamma, x \neq y} \frac{|F(x) - F(y)|}{d_{\mathcal{X}}(x,y)}.$$

The metric $d_{\mathcal{X}}$ appearing above is defined in [19, Definition 2.2]) using the Mal'cev basis $\mathcal{X}$.

The important examples to keep in mind are the polynomial phase functions $\psi(n) = e(g(n))$, where g is a polynomial of degree at most s . Here the relevant nilmanifold is $G/\Gamma = \mathbb{R}/\mathbb{Z}$ with the simple filtration $G_k = \mathbb{R}/\mathbb{Z}$ for $k \leq s$ and $G_k = \{0\}$ for $k > s$. We also note that any *bracket polynomial*, such as $e(P_1(n)[P_2(n)])$, with $P_1(x), P_2(x) \in \mathbb{R}[x]$, is essentially a nilsequence, in the sense that (by smoothing the fractional part function a bit, as in [34, p. 102]) this function for $n \leq x$ be written as a linear combination with bounded coefficients of $\ll (\log x)^A$ nilsequences $\psi \in \Psi_s(\Delta, K)$ for some $s, \Delta, K \ll 1$, plus an error term that is $O((\log x)^{-A/10})$ in ℓ^1 norm.

We make a technical remark on the function φ appearing in Definition 1.2(3). When applying the machinery of nilsequences to problems in additive combinatorics, it is arguably more convenient to work with smooth functions φ (with controlled smoothness norms) instead of Lipschitz functions, although the use of Lipschitz functions has by now become standard. All the results proved in this paper would remain true with this alternative definition of nilsequences.

In order to state our main theorems, we need the W -trick. For $w \geq 2$, we write $\mathcal{P}(w) := \prod_{p \leq w} p$.

Theorem 1.3. *Let an integer $s \geq 1$, large real numbers $A, \Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/4)$ be given. Then for any $x \geq 2$, we have*

$$\sum_{d \leq x^{1/4-\varepsilon}} \max_{(c,d)=1} \sup_{\psi \in \Psi_s(\Delta, \log x)} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) \psi(n) - \frac{dW}{\varphi(dW)} \sum_{\substack{n \leq x \\ (n,W)=1 \\ n \equiv c \pmod{d}}} \psi(n) \right| \ll_{s,A,\Delta,\varepsilon} \frac{x}{(\log x)^A},$$

where $W = \mathcal{P}((\log x)^C)$ for some constant $C = C(A, s, \Delta, \varepsilon)$.

We can increase the level of distribution to $1/3$ if the nilsequence ψ is *fixed* (does not depend on d).

Theorem 1.4. *Let an integer $s \geq 1$, large real numbers $A, \Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/3)$ be given. Then for any nilsequence $\psi \in \Psi_s(\Delta, \log x)$ and $x \geq 2$, we have*

$$\sum_{d \leq x^{1/3-\varepsilon}} \max_{(c,d)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) \psi(n) - \frac{dW}{\varphi(dW)} \sum_{\substack{n \leq x \\ (n,W)=1 \\ n \equiv c \pmod{d}}} \psi(n) \right| \ll_{s,A,\Delta,\varepsilon} \frac{x}{(\log x)^A},$$

where $W = \mathcal{P}((\log x)^C)$ for some constant $C = C(A, s, \Delta, \varepsilon)$.

We can further increase the level of distribution to $1/2$ if c is fixed (does not depend on d) and the absolute value inside the d sum is replaced by a well-factorable weight, defined in the following definition.

Definition 1.5 (Well-factorable sequences). We say that a sequence (λ_d) of real numbers is *well-factorable* of level D if (λ_d) is supported on $d \in [1, D]$ and for any $1 \leq R, S \leq D$ with $RS = D$ one can write $\lambda_d = \beta * \gamma(d) := \sum_{d=d_1 d_2} \beta(d_1) \gamma(d_2)$ for some sequences $\beta(d), \gamma(d)$ of modulus at most 1 and such that β is supported on $[1, R]$ and γ is supported on $[1, S]$.

Well-factorable weights arise in many sieve problems due to the fact that the linear sieve weights (introduced by Iwaniec) are a bounded linear combination of well-factorable weights; see [11, Lemma 12.16]. Bombieri–Friedlander–Iwaniec [4] famously broke the $1/2$ barrier for the level of distribution in the Bombieri–Vinogradov inequality, provided that the absolute value signs are replaced by well-factorable weights and the residue class is fixed for all q . In this setting, we can also do better than Theorems 1.3 and 1.4, despite not being able to break the $1/2$ barrier.

Theorem 1.6. *Let integers $s \geq 1$, $c \neq 0$, large real numbers $A, \Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/2)$ be given. Then for any nilsequence $\psi \in \Psi_s(\Delta, \log x)$ and any well-factorable sequence (λ_d) of level $x^{1/2-\varepsilon}$ with $x \geq 2$, we have*

$$\left| \sum_{\substack{d \leq x^{1/2-\varepsilon} \\ (d,c)=1}} \lambda_d \left(\sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) \psi(n) - \frac{dW}{\phi(dW)} \sum_{\substack{n \leq x \\ (n,W)=1 \\ n \equiv c \pmod{d}}} \psi(n) \right) \right| \ll_{s,A,\Delta,\varepsilon,c} \frac{x}{(\log x)^A},$$

where $W = \mathcal{P}((\log x)^C)$ for some constant $C = C(A, s, \Delta, \varepsilon)$.

Remark 1.7. In Theorems 1.3, 1.4, and 1.6, in order to get an arbitrary power of log saving in the error term, it is necessary to perform a “W-trick” to overcome the fact that functions such as $\Lambda(n)e(\frac{an^s}{q})$ are not equidistributed in all residue classes when $q \leq (\log x)^C$. Since this leads to the choice $W = \mathcal{P}((\log x)^C)$ which is rather large, we are unable to perform the same W-trick as in [17] which compares the function $n \mapsto \frac{\phi(W)}{W} \Lambda(Wn+1)$ to 1. We have adapted an alternative approach which, roughly speaking, compares Λ to (a normalized version of) the function $n \mapsto 1_{(n,W)=1}$.

Remark 1.8. One can also obtain similar results for the Möbius function, without the need for the W-trick and with no main term. Specifically, with the notations and assumptions of Theorem 1.3, we have

$$\sum_{d \leq x^{1/4-\varepsilon}} \max_{(c,d)=1} \sup_{\psi \in \Psi_s(\Delta, \log x)} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \mu(n) \psi(n) \right| \ll_{s,A,\Delta,\varepsilon} \frac{x}{(\log x)^A}.$$

With the notations and assumptions of Theorem 1.4, we have

$$\sum_{d \leq x^{1/3-\varepsilon}} \max_{(c,d)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \mu(n) \psi(n) \right| \ll_{s,A,\Delta,\varepsilon} \frac{x}{(\log x)^A}.$$

With the notations and assumptions of Theorem 1.6, we have

$$\left| \sum_{\substack{d \leq x^{1/2-\varepsilon} \\ (d,c)=1}} \lambda_d \left(\sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \mu(n) \psi(n) \right) \right| \ll_{s,A,\Delta,\varepsilon,c} \frac{x}{(\log x)^A}.$$

These statements can be proved with arguments almost identical to Theorems 1.3, 1.4 and 1.6, using the fact that the Möbius function obeys an identity analogous to Vaughan’s identity for Λ .

In fact, our results above also apply to a number of other multiplicative functions, such as the divisor functions $d_k(n)$ (for all values of k , including non-integer and complex values) and the indicator function $1_S(n)$ of sums of two squares. For these functions, however, one would have to modify the main term sum involving $(n, W) = 1$ in Theorems 1.3, 1.4 and 1.6. For simplicity, and to reduce repetition in the arguments, we only state Bombieri–Vinogradov type theorems for these functions in the equidistributed (“minor arc”) case where there is no main term; see Theorem 4.6.

1.2 Results for polynomial phase twists

Since polynomial phases of the form $e(P(n))$ are examples of nilsequences of bounded complexity, Theorems 1.3, 1.4 and 1.6 immediately imply as special cases Bombieri–Vinogradov type estimates for $\Lambda(n)e(P(n))$. We state these below, since they are of independent interest and since they will be utilized in deriving some of the applications of our results.

Corollary 1.9. *Let an integer $s \geq 1$, a large real number $A \geq 2$, and a small real number $\varepsilon \in (0, 1/4)$ be given. Then for any $x \geq 2$, we have*

$$\sum_{d \leq x^{1/4-\varepsilon}} \max_{(c,d)=1} \sup_{\deg(P) \leq s} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n)e(P(n)) - \frac{dW}{\varphi(dW)} \sum_{\substack{n \leq x \\ (n,W)=1 \\ n \equiv c \pmod{d}}} e(P(n)) \right| \ll_{s,A,\varepsilon} \frac{x}{(\log x)^A},$$

where $W = \mathcal{P}((\log x)^C)$ for some constant $C = C(A, s, \varepsilon)$.

Corollary 1.10. *Let an integer $s \geq 1$, a large real number $A \geq 2$, and a small real number $\varepsilon \in (0, 1/3)$ be given. Then for any polynomial $P(x) \in \mathbb{R}[x]$ of degree s and $x \geq 2$, we have*

$$\sum_{d \leq x^{1/3-\varepsilon}} \max_{(c,d)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n)e(P(n)) - \frac{dW}{\varphi(dW)} \sum_{\substack{n \leq x \\ (n,W)=1 \\ n \equiv c \pmod{d}}} e(P(n)) \right| \ll_{s,A,\varepsilon} \frac{x}{(\log x)^A},$$

where $W = \mathcal{P}((\log x)^C)$ for some constant $C = C(A, s, \varepsilon)$.

Corollary 1.11. *Let integers $s \geq 1$, $c \neq 0$, a large real number $A \geq 2$, and a small real number $\varepsilon \in (0, 1/2)$ be given. Then for any polynomial $P(x) \in \mathbb{R}[x]$ of degree s and any well-factorable sequence (λ_d) of level $x^{1/2-\varepsilon}$ with $x \geq 2$, we have*

$$\left| \sum_{\substack{d \leq x^{1/2-\varepsilon} \\ (d,c)=1}} \lambda_d \left(\sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n)e(P(n)) - \frac{dW}{\varphi(dW)} \sum_{\substack{n \leq x \\ (n,W)=1 \\ n \equiv c \pmod{d}}} e(P(n)) \right) \right| \ll_{s,A,\varepsilon,c} \frac{x}{(\log x)^A},$$

where $W = \mathcal{P}((\log x)^C)$ for some constant $C = C(A, s, \varepsilon)$.

Remark 1.12. We could also obtain analogous results for bracket polynomial phases, so for example $e(P_1(n)[P_2(n)])$, where P_1, P_2 are polynomials. One simply needs the fact that these functions are well-approximable by nilsequences of bounded complexity, a property that was noted above. We leave the details to the interested reader.

Previous results related to our main theorems are as follows:

- For the Möbius function, it was established in [32] for $Q < x^{1/2}$ that

$$\max_{(c,d)=1} \sup_{\psi \in \Psi_s(\Delta, K)} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \mu(n)\psi(n) \right| \ll_{s,A,\Delta,K} \frac{x \log \log x}{Q(\log(x/Q^2))} \quad (1.1)$$

for almost all $d \in [Q, 2Q]$, in the sense that the number of exceptional d is $\ll_A Q(\log x)^{-A}$. Although this result is applicable for $d \leq x^{1/2-\varepsilon}$, it saves a factor of $(\log \log x)/(\log x)$ at best. The proof relies crucially on almost all numbers having prime factors in various suitable ranges, and hence it does not work for the case of primes, i.e. the von Mangoldt function.

- For linear exponentials (that is, $s = 1$), Theorem 1.4 was proved by Todorova and Tolev [36], and for quadratic phase functions (which is a special case of the $s = 2$ case) by Tolev [37].
- It was shown by Matomäki [26], improving on work of Mikawa [31], that the well-factorable level of distribution estimate given by Theorem 1.6 holds for $s = 1$.

1.3 Notation

We use the usual asymptotic notation $\ll$, $\gg$, $O(\cdot)$, $o(\cdot)$, $\asymp$. Dependence of these symbols on parameters is indicated whenever such a dependence occurs (so, for example, $o_{w,x \rightarrow \infty}(1)$ is a quantity depending on w and x and tending to 0 as $x \rightarrow \infty$).

We use Λ to denote the von Mangoldt function, μ to denote the Möbius function, φ to denote the Euler phi function, d_k to denote the k -fold divisor function (with $d(n) := d_2(n)$) and (a, b) to denote the greatest common divisor of a and b . We also let $\mathcal{P}(w) := \prod_{p \leq w} p$.

Let $\|\cdot\|_{U^k(\mathbb{Z}/N\mathbb{Z})}$ stand for the usual Gowers norm over the cyclic group $\mathbb{Z}/N\mathbb{Z}$. Given a function $f : \mathbb{Z} \rightarrow \mathbb{C}$ supported on $[N] := \{1, 2, \dots, N\}$, we define its Gowers norm $\|f\|_{U^k[N]}$ over the interval $[N]$ as

$$\|f\|_{U^k[N]} := \frac{\|f \cdot 1_{[N]}\|_{U^k(\mathbb{Z}/N'\mathbb{Z})}}{\|1_{[N]}\|_{U^k(\mathbb{Z}/N'\mathbb{Z})}},$$

where $N' > 2N$ (say $N' = 2N + 1$ for concreteness) and $f \cdot 1_{[N]}$ and $1_{[N]}$ are extended to $\mathbb{Z}/N'\mathbb{Z}$ in the natural way.

If A is any nonempty, finite set and $f : A \rightarrow \mathbb{C}$ is a function, we use the averaging notation

$$\mathbb{E}_{a \in A} f(a) := \frac{1}{|A|} \sum_{a \in A} f(a).$$

For $x \in \mathbb{R}$, we use $\|x\|$ to denote the distance from x to its nearest integer.

2 APPLICATIONS

We now present several applications of our Bombieri–Vinogradov type theorems to problems related to Diophantine properties of the primes, as well as to additive combinatorics.

2.1 Bounded gaps between primes in Bohr sets

Our first application generalizes the celebrated result of Zhang [39], Maynard [29] and Tao (unpublished) on bounded gaps between primes. Subsequent to these works, a number of interesting subsets of the primes have also been shown to exhibit bounded gaps. See [1], [30] for primes in short intervals, [35], [30] for work on Chebotarev sets, and [2], [6] for work on primes in Beatty sequences $(\lfloor \alpha n + \beta \rfloor)_{n \geq 1}$.

As a consequence of our main theorems, we are able to add to this list that the primes lying in a *nil-Bohr set* exhibit bounded gaps. This generalizes the result of Baker and Zhao [2] mentioned above, which corresponds to classical Bohr sets of a special form.

Nil-Bohr sets were introduced by Host and Kra in [24] and are a natural generalization of classical Bohr sets to the setting of higher order Fourier analysis. For the convenience of the reader, we first define classical Bohr sets and then nil-Bohr sets.

Definition 2.1 (Bohr sets). Let $r \geq 0$ and $\rho > 0$. Then for any real numbers $\alpha_1, \dots, \alpha_r \in \mathbb{R}$ the set

$$B = \{n \in \mathbb{Z} : \|\alpha_i n\| < \rho \ \forall 1 \leq i \leq r\}$$

is called a (classical) *Bohr set* (of rank r), where $\|x\|$ denotes the distance from x to its nearest integer.

Definition 2.2 (nil-Bohr sets). Let U be an open subset of $\mathbb{C}$, and let $\psi : G/\Gamma \rightarrow \mathbb{C}$ be a nilsequence defined on some nilmanifold. Then the set

$$B = \{n \in \mathbb{Z} : \psi(n) \in U\}$$

is called a *nil-Bohr set*.

Note that any Bohr set is also a nil-Bohr set. In fact, a Bohr set of any rank r can be represented as a nil-Bohr set, by taking $G/\Gamma = (\mathbb{R}/\mathbb{Z})^r$ and ψ, U appropriately. Moreover, any polynomial Bohr set of the form

$$\{n \in \mathbb{Z} : \|Q(n)\| < \rho\}$$

is a nil-Bohr set for any polynomial $Q(x)$ and any $\rho \in (0, 1/2)$; take $\psi(n) = e(Q(n))$ as the nilsequence in the definition, and take U to be $\{z \in \mathbb{C} \setminus \{0\} : \arg(z) \in (-2\pi\rho, 2\pi\rho)\}$. Any nil-Bohr set that is “irrational” in a suitable sense (see Remark 2.4) contains infinitely many primes. This is however not a trivial fact; if $Q(x)$ is a polynomial with irrational leading coefficient, then proving the asymptotic

$$|\{p \leq x : \|Q(p)\| < \rho\}| = (2\rho + o(1)) \frac{x}{\log x} \quad (2.1)$$

is related to Weyl’s equidistribution criterion, exponential sum estimates in the Waring–Goldbach problem, and, in the case of general nil-Bohr sets, to the Green–Tao result [18] on Möbius orthogonality with nilsequences. Also note that nil-Bohr sets (as well as classical Bohr sets) can be rather irregular: in the asymptotic formula (2.1) it is not possible to specify the $o(1)$ term without imposing a restriction on how fast the denominators of the continued fraction convergents of the leading coefficient of $Q(x)$ grow.

In what follows, we say that a set $S \subset \mathbb{N}$ has *bounded gaps* if there exists $C > 0$ such that the inequality $0 < s_1 - s_2 \leq C$ has infinitely many solutions with $s_1, s_2 \in S$.

Theorem 2.3 (Bounded gaps between primes in nil-Bohr sets). *Let $Q(x) \in \mathbb{R}[x]$ be a polynomial with at least one irrational coefficient which is not the constant term. Let $\rho \in (0, 1/2)$, and form the nil-Bohr set*

$$B = \{n \in \mathbb{Z} : \|Q(n)\| < \rho\}.$$

Then the subset of the primes $\mathbb{P} \cap B$ has bounded gaps.

Remark 2.4. It is only for the sake of simplicity that we restrict to nil-Bohr sets of this form; likely similar arguments could be made to work for any nil-Bohr set B that satisfies the following two natural properties: (i) B is *irrational* in the sense that, in the notation of Definition 4.1, for some constant $c > 0$ and infinitely many $x \in \mathbb{N}$ we have $\psi \in \Psi_s(\Delta, K; x^{-c}, x)$ with s, Δ, K fixed. (ii) B is *dense* in the sense that $|B \cap [1, x]| \gg x$. We leave the details of this generalization to the interested reader.

2.2 Chen primes in Bohr sets

Our next application involves Chen primes, which are primes p such that $p + 2 \in P_2$, where P_2 is the set of positive integers with at most two prime factors. We write

$$\mathcal{P}_{\text{Chen}} = \{p \in \mathbb{P} : p + 2 \in P_2\}.$$

A celebrated result of Chen [5] shows that $\mathcal{P}_{\text{Chen}}$ is infinite. It is moreover a sparse subset of the primes, since it can be shown to satisfy $|\mathcal{P}_{\text{Chen}} \cap [1, x]| \asymp x/(\log x)^2$.

It was shown by Matomäki [26] that the Chen primes are well-distributed in classical Bohr sets, meaning that there exists $\theta > 0$ such that

$$\|\alpha p\| < p^{-\theta}$$

has infinitely many solutions in Chen primes p for any fixed irrational α .

We generalize this by proving that the Chen primes are well-distributed in more general nil-Bohr sets.

Theorem 2.5 (Chen primes in nil-Bohr sets). *Let $Q(x) \in \mathbb{R}[x]$ be a polynomial of degree $s \geq 1$ such that Q has at least one irrational coefficient which is not the constant term. Then for some constant $\theta_s > 0$ (independent of Q) there are infinitely many solutions to*

$$\|Q(p)\| < p^{-\theta_s}, \quad p \in \mathcal{P}_{\text{Chen}}.$$

Remark 2.6. As in the case of Theorem 2.3, it should be possible to generalize our proof method to showing that for any irrational and dense nil-Bohr set B we have infinitely many solutions to $p \in \mathcal{P}_{\text{Chen}} \cap B$. We leave the details to the interested reader.

2.3 Linear equations in primes in arithmetic progressions

The next application is a generalization of the groundbreaking result proved by Green–Tao [17, 18] and Green–Tao–Ziegler [20] that one can asymptotically count the number of solutions to any linear system of equations in the primes (of finite complexity, which excludes for instance counting twin primes or solutions to the binary Goldbach problem). It is natural to consider this problem for subsets of the primes as well, in particular for primes belonging to an arithmetic progression. We show that in the rather large range $q \leq x^\theta$, for suitable $\theta > 0$, we still obtain asymptotics for linear equations in primes restricted to a congruence class $a \pmod{q}$, for almost all choices of q .

In what follows, we set

$$\Lambda_{a,q}(n) := \frac{\varphi(q)}{q} \Lambda(qn + a), \tag{2.2}$$

which is a normalized version of the von Mangoldt function restricted to the arithmetic progression $a \pmod{q}$. Extend $\Lambda_{a,q}(n)$ to integers $n \leq 0$ by setting it to be zero at them. Further, for a system $\Psi(n) = (L_1(\mathbf{n}), \dots, L_t(\mathbf{n}))$ of affine linear forms in d variables we define its *size* at scale N to be

$$\|\Psi\|_N := \sum_{i=1}^t \sum_{j=1}^d |L_i(e_j) - L_i(0)| + \sum_{i=1}^t \left| \frac{L_i(0)}{N} \right|,$$

where $e_1, \dots, e_d$ are the standard basis vectors in $\mathbb{Z}^d$.

Theorem 2.7 (Linear equations in primes in arithmetic progressions to large moduli). *Let $\varepsilon > 0$ and $A, t, d, M \geq 1$ be given. Let $x \geq 10$ and $Q \leq x^{1/3-\varepsilon}$. Then for all but $\ll_{\varepsilon, A, t, d, M} Q/(\log x)^A$ choices of $1 \leq q \leq Q$ the following holds. For every $\mathbf{a} \in (\mathbb{Z}/q\mathbb{Z})^d$ and every finite complexity tuple $\Psi = (L_1(\mathbf{n}), \dots, L_t(\mathbf{n}))$ of non-constant affine-linear forms in d variables of size $\|\Psi\|_x \leq M$ we have*

$$\sum_{\mathbf{n} \in [1, x]^d} \Lambda(L_1(q\mathbf{n} + \mathbf{a})) \cdots \Lambda(L_t(q\mathbf{n} + \mathbf{a})) = \beta_\infty \prod_p \beta_{p, a, q} + o_{t, d, M}(x^d), \quad (2.3)$$

where the product on the right-hand side is convergent, $\beta_{p, a, q} \geq 0$, $0 \leq \beta_\infty \ll x^d$, and the local factors $\beta_{p, a, q}$ are given by

$$\beta_{p, a, q} := \mathbb{E}_{\mathbf{n} \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^t \Lambda_{\mathbb{Z}/p\mathbb{Z}}(L_i(q\mathbf{n} + \mathbf{a})),$$

and $\Lambda_{\mathbb{Z}/p\mathbb{Z}}(b) = \frac{p}{p-1} 1_{(n, p)=1}$, and $\beta_\infty = \text{vol}_d([1, x]^d \cap \Psi^{-1}(\mathbb{R}_+^t))$.

We also obtain a similar theorem (without the main term) with the Möbius function in place of the von Mangoldt function; see Proposition 11.1. Note that the condition $Q \leq x^{1/3-\varepsilon}$ actually corresponds to a $1/4$ level of distribution, since the primes being counted are of size $\approx Qx$.

Remark 2.8. Theorem 2.7 can be viewed as generalizing a result of Bienvenu [3] on linear equations in primes in the Siegel–Walfisz regime, corresponding to $q \ll_A (\log x)^A$ in (2.3). As in the Siegel–Walfisz theorem, there are no exceptional moduli q in [3, Theorem 1.3]. The same applies to our Theorem 2.7 as well, since if we are considering moduli of size $\leq Q = (\log x)^A$, the theorem gives $\ll_A (\log x)^{-2A}$ exceptional moduli, and this quantity is clearly less than 1 for x large.

In the course of proving Theorem 2.7, we obtain Gowers uniformity of primes in almost all arithmetic progressions.

Theorem 2.9 (Gowers uniformity of primes in arithmetic progressions to large moduli). *Let $\varepsilon > 0$ and $A, k \geq 1$ be given. Let $x \geq 10$, $Q \leq x^{1/3-\varepsilon}$, and $w \geq 1$. Then for all but $\ll_{\varepsilon, A, k} Q/(\log x)^A$ choices of $1 \leq q \leq Q$ with $\mathcal{P}(w) \mid q$ the following holds. For every $(a, q) = 1$, the function $\Lambda_{a, q}$ defined by (2.2) satisfies the Gowers norm bound*

$$\|\Lambda_{a, q} - 1\|_{U^k[x]} = o_{w \rightarrow \infty}(1) + o_{w, x \rightarrow \infty}(1).$$

We conclude with an immediate corollary of Theorem 2.7 to long arithmetic progressions in the primes.

Corollary 2.10 (Green–Tao theorem for primes in arithmetic progressions to large moduli). *Let $x \geq 10$, and let $k \geq 1$, $A \geq 1$ and $\varepsilon > 0$ be given. Also let $Q \leq x^{1/4-\varepsilon}$. Then for all but $\ll_{\varepsilon, A, k} Q/(\log x)^A$ moduli $q \leq Q$, each of the sets $\mathbb{P} \cap (q\mathbb{Z} + a)$ with $(a, q) = 1$ contains a nontrivial k -term arithmetic progression all of whose elements are $\leq x$.*

2.4 Future work

In ongoing work with P.-Y. Bienvenu, we develop a transference principle for arbitrary systems of linear equations (not necessarily translation-invariant) in the spirit of [16, 17], generalizing the

transference principle of [27]. This new transference principle requires as a key input that the sparse set we want to transfer is “well-distributed” in nil-Bohr sets, which using our Bombieri–Vinogradov theorems (Theorems 1.3, 1.4, 1.6) we can verify for many subsets of the primes of interest. In particular, we will be able to generalize the Green–Tao linear equations in primes result to Chen primes.

3 QUANTITATIVE EQUIDISTRIBUTION OF POLYNOMIAL SEQUENCES

To understand the main idea behind the proof of our main results (Theorems 1.3, 1.4 and 1.6), it is not essential to know the precise definitions around nilsequences, as long as one is willing to accept certain results about nilsequences as black boxes, many of which can be found in [19]. The readers are thus encouraged to consider the special case when the nilsequence $\psi(n)$ is a polynomial phase function (Corollaries 1.9, 1.10 and 1.11). The purpose of this section is to collect definitions and results about equidistribution of polynomial sequences that are used in our proofs. These can be thought of as generalizations of major and minor arc analysis to treat exponential sums, and of Weyl’s inequality.

Definition 3.1 (Equidistribution). Let G/Γ be a nilmanifold equipped with a Mal’cev basis $\mathcal{X}$. Let $N \geq 1$ and $\delta > 0$.

1. A finite sequence $(g(n)\Gamma)_{1 \leq n \leq N}$ is said to be δ -*equidistributed* if we have

$$\left| \frac{1}{N} \sum_{n=1}^N F(g(n)\Gamma) - \int_{G/\Gamma} F \right| \leq \delta \|F\|_{\text{Lip}}$$

for all Lipschitz functions $F : G/\Gamma \rightarrow \mathbb{C}$. Here the integral is with respect to the unique Haar measure on G/Γ .

2. A finite sequence $(g(n)\Gamma)_{1 \leq n \leq N}$ is said to be *totally* δ -*equidistributed* if we have

$$\left| \frac{1}{|P|} \sum_{n \in P} F(g(n)\Gamma) - \int_{G/\Gamma} F \right| \leq \delta \|F\|_{\text{Lip}}$$

for all Lipschitz functions $F : G/\Gamma \rightarrow \mathbb{C}$ and all arithmetic progressions $P \subset [N]$ of length at least δN .

Definition 3.2 (Horizontal characters). Let G/Γ be an m -dimensional nilmanifold equipped with a Mal’cev basis $\mathcal{X}$. A *horizontal character* is a homomorphism $\eta : G \rightarrow \mathbb{R}/\mathbb{Z}$ which annihilates Γ . It is called *nontrivial* if it is not identically zero. Any horizontal character can be written as $\eta(x) = k \cdot \psi(x)$ for some $k \in \mathbb{Z}^m$, where $\psi : G \rightarrow \mathbb{R}^m$ is the Mal’cev coordinate map of G (see [19, Definition 2.1]). The *modulus* of η is defined by $|\eta| := |k|$.

If G/Γ is equipped with a filtration of degree s , η is a horizontal character on G/Γ , and g is a polynomial sequence on G , then $\eta \circ g : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a genuine polynomial of degree at most s . This follows from Mal’cev coordinates of polynomial sequences; see [19, Lemma 6.7], for example.

Definition 3.3 (Smoothness norms). Let $g: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ be a (genuine) polynomial of degree s , written uniquely as

$$g(n) = \alpha_0 + \alpha_1 \binom{n}{1} + \cdots + \alpha_s \binom{n}{s}.$$

For any $N \geq 1$ we define the *smoothness norm*

$$\|g\|_{C^\infty(N)} := \sup_{1 \leq j \leq s} N^j \|\alpha_j\|_{\mathbb{R}/\mathbb{Z}}.$$

One can easily relate the smoothness norm with the traditional coefficients of a polynomial (see [18, Lemma 3.2]).

Lemma 3.4. *Let $g: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ be a polynomial $g(n) = \beta_0 + \beta_1 n + \cdots + \beta_s n^s$, and let $N \geq 1$. There exists a positive integer $q = O_s(1)$ such that*

$$\|q\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll_s N^{-j} \|g\|_{C^\infty(N)}$$

for each $1 \leq j \leq s$. In the other direction, we have

$$\|g\|_{C^\infty(N)} \ll_s \sup_{1 \leq j \leq s} N^j \|\beta_j\|_{\mathbb{R}/\mathbb{Z}}.$$

We now relate equidistribution of a polynomial sequence with smoothness norms.

Theorem 3.5 (Quantitative Leibman theorem). *Let $s, \Delta \geq 1$, $0 < \delta < 1/2$ and $N \geq 1$. Let G/Γ be a nilmanifold of dimension Δ , equipped with a filtration $G_\bullet$ of degree s and a $1/\delta$ -rational Mal'cev basis. Let $g: \mathbb{Z} \rightarrow G$ be a polynomial sequence adapted to $G_\bullet$. If $(g(n)\Gamma)_{1 \leq n \leq N}$ is not totally δ -equidistributed, then there is a nontrivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ with $|\eta| \ll \delta^{-O_{s,\Delta}(1)}$ such that*

$$\|\eta \circ g\|_{C^\infty(N)} \ll \delta^{-O_{s,\Delta}(1)}.$$

Proof. This is proved in [19, Theorem 2.9] under the marginally stronger hypothesis that $(g(n))_{1 \leq n \leq N}$ is not δ -equidistributed. Our version can be deduced from [19, Theorem 2.9] using the same arguments as those from [10, Theorem 5.2], as follows. Let $F: G/\Gamma \rightarrow \mathbb{C}$ be a Lipschitz function and let $P \subset [N]$ be an arithmetic progression of length $|P| \geq \delta N$, such that

$$\left| \frac{1}{|P|} \sum_{n \in P} F(g(n)\Gamma) - \int_{G/\Gamma} F \right| > \delta \|F\|_{\text{Lip}}.$$

Let q be the common difference of P (so that $q \leq \delta^{-1}$), and write $P = \{qm + a: 1 \leq m \leq |P|\}$ for some $|a| \leq q$. Consider the polynomial sequence $g': \mathbb{Z} \rightarrow G$ defined by

$$g'(m) = g(qm + a).$$

It follows that $(g'(m))_{1 \leq m \leq |P|}$ is not δ -equidistributed in G/Γ , and hence, by [19, Theorem 2.9], there exists a nontrivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ with $|\eta| \ll \delta^{-O_{s,\Delta}(1)}$ such that

$$\|\eta \circ g'\|_{C^\infty(|P|)} \ll \delta^{-O_{s,\Delta}(1)}.$$

If we write

$$\eta \circ g(n) = \sum_{i=0}^s \alpha_i n^i,$$

then

$$\eta \circ g'(m) = \sum_{i=0}^s \alpha_i (qm + a)^i = \sum_{j=0}^s \left(\sum_{i \geq j} \alpha_i \binom{i}{j} a^{i-j} \right) q^j m^j$$

In view of the bound for the smoothness norm of $\eta \circ g'$ and Lemma 3.4, it follows that

$$\left\| q^j \sum_{i \geq j} \alpha_i \binom{i}{j} a^{i-j} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_{s,\Delta}(1)} |P|^{-j} \ll \delta^{-O_{s,\Delta}(1)} N^{-j}$$

for each $1 \leq j \leq s$. A simple induction argument then shows that

$$\|(aq)^s s! \alpha_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_{s,\Delta}(1)} N^{-j}$$

for each $1 \leq j \leq s$. It follows that (again using Lemma 3.4)

$$\|(aq)^s s! \eta \circ g\|_{C^\infty(N)} \ll \delta^{-O_{s,\Delta}(1)}.$$

This completes the proof since $\|(aq)^s s! \eta\| = (aq)^s s! \|\eta\| \ll \delta^{-O_{s,\Delta}(1)}$. $\square$

We have the following converse to the quantitative Leibman theorem (see [10, Lemma 5.3]).

Lemma 3.6. *Let $s, \Delta \geq 1$. Let G/Γ be a nilmanifold of dimension Δ , equipped with a filtration $G_\bullet$ of degree s . Let $g : \mathbb{Z} \rightarrow G$ be a polynomial sequence adapted to $G_\bullet$. Suppose that $N \geq CD$ for some sufficiently large constant $C = C(s, \Delta)$. If $\|\eta \circ g\|_{C^\infty(N)} \leq D$ for some nontrivial horizontal character $\eta : G \rightarrow \mathbb{R}/\mathbb{Z}$ with $\|\eta\| \leq D$, then $(g(n))_{1 \leq n \leq N}$ is not totally cD^{-2} -equidistributed in G/Γ , where $c = c(s, \Delta) > 0$ is a positive constant.*

Finally, we will need the following factorization theorem for polynomial sequences [19, Theorem 1.19], which decomposes an arbitrary polynomial sequence into a smooth part, an equidistributed part, and a rational part.

Theorem 3.7 (Factorization theorem). *Let s be a positive integer and let $\Delta, M_0 \geq 2$. Let G/Γ be a nilmanifold of dimension at most Δ , equipped with a filtration $G_\bullet$ of degree at most s and an M_0 -rational Mal'cev basis $\mathcal{X}$. Let $g : \mathbb{Z} \rightarrow G$ be a polynomial sequence adapted to $G_\bullet$.*

For any $A > 0$, we may find $M \in [M_0, M_0^{O_{\Delta,s,\Delta}(1)}]$, a sub-nilmanifold G'/Γ' of G/Γ , a Mal'cev basis $\mathcal{X}'$ for G'/Γ' in which each element is an M -rational combination of the elements of $\mathcal{X}$, and a decomposition $g = \varepsilon g' \gamma$ into polynomial sequences ε, g', γ with the following properties:

1. $\varepsilon : \mathbb{Z} \rightarrow G$ is (M, N) -smooth in the sense that $d_{\mathcal{X}}(\varepsilon(n), \text{id}) \leq M$ and $d_{\mathcal{X}}(\varepsilon(n), \varepsilon(n-1)) \leq M/N$ for each $1 \leq n \leq N$;
2. $g' : \mathbb{Z} \rightarrow G'$ takes values in G' , and $\{g'(n)\Gamma'\}_{1 \leq n \leq N}$ is totally M^{-A} -equidistributed in G'/Γ' ;
3. $\gamma : \mathbb{Z} \rightarrow G$ is M -rational in the sense that for each $n \in \mathbb{Z}$, $\gamma(n)^r \in \Gamma$ for some $1 \leq r \leq M$. Moreover, γ is periodic with period at most M .

4 BOMBIERI–VINOGRADOV THEOREMS FOR EQUIDISTRIBUTED NILSEQUENCES

In this section we reduce our problem to studying those nilsequences ψ that are “equidistributed”. In the case when $\psi(n) = e(\alpha n^s)$, this corresponds to α lying in “minor arcs”.

Definition 4.1 (Equidistributed nilsequences). Recall the definition of $\Psi_s(\Delta, K)$ in Definition 1.2. For $\eta \in (0, 1)$ and $x \geq 2$, we define $\Psi_s(\Delta, K; \eta, x)$ to be the class of those nilsequences $\psi \in \Psi_s(\Delta, K)$ of the form $\psi(n) = \varphi(g(n)\Gamma)$ that obey the following additional conditions:

1. the finite sequence $(g(n)\Gamma)_{1 \leq n \leq 10x}$ is totally η -equidistributed in G/Γ (defined in [19, Definition 1.2]);
2. the Lipschitz function φ satisfies $\int_{G/\Gamma} \varphi = 0$ (Here the integral is with respect to the unique Haar measure on G/Γ).

We will loosely call those nilsequences in $\Psi_s(\Delta, K; \eta, x)$ η -equidistributed.

4.1 Statements of results for equidistributed nilsequences

The following theorems show that our main theorems hold for η -equidistributed nilsequences when $0 < \eta \leq (\log x)^{-O_A(1)}$. Moreover, the error term for such nilsequences is even power-saving if $\eta = x^{-\delta}$ for some constant $\delta > 0$.

Theorem 4.2. *Let an integer $s \geq 1$, a large real number $\Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/4)$ be given. There exists a constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$, such that for $x \geq 2$, $\eta > 0$ we have*

$$\sum_{d \leq x^{1/4-\varepsilon}} \max_{(c,d)=1} \sup_{\psi \in \Psi_s(\Delta, \eta^{-\kappa}; \eta, x/d)} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) \psi((n-c)/d) \right| \ll \eta^\kappa x (\log x)^2.$$

Theorem 4.3. *Let an integer $s \geq 1$, a large real number $\Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/3)$ be given. There exists a constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$, such that for any $x \geq 2$, $\eta > 0$ and any nilsequence $\psi \in \Psi_s(\Delta, \eta^{-\kappa}; \eta, x)$ we have*

$$\sum_{d \leq x^{1/3-\varepsilon}} \max_{(c,d)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) \psi(n) \right| \ll \eta^\kappa x (\log x)^2.$$

Theorem 4.4. *Let integers $s \geq 1$, $c \neq 0$, a large real number $\Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/2)$ be given. There exists a constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$, such that for any well-factorable sequence (λ_d) of level $x^{1/2-\varepsilon}$ with $x \geq 2$ and any nilsequence $\psi \in \Psi_s(\Delta, \eta^{-\kappa}; \eta, x)$ with $\eta > 0$, we have*

$$\left| \sum_{\substack{d \leq x^{1/2-\varepsilon} \\ (d,c)=1}} \lambda_d \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) \psi(n) \right| \ll \eta^\kappa x (\log x)^2.$$

Remark 4.5. Theorems 4.2, 4.3, 4.4 hold equally well with Λ replaced by μ , since the proofs rely on type I and II estimates and μ satisfies a similar Vaughan identity as Λ does.

As mentioned earlier, our proof methods also apply to other arithmetic functions than Λ and μ , in particular to the functions d_k and 1_S (and other sequences that satisfy an identity of Heath-Brown’s type).

Theorem 4.6. *Theorems 4.2, 4.3 and 4.4 continue to hold with Λ replaced by the k -fold divisor function d_k with $k \in \mathbb{C}$ fixed, or Λ replaced by 1_S , where S is the set of natural numbers expressible as the sum of two squares.*

In our deductions of Theorems 1.3, 1.4, and 1.6 from Propositions 4.2, 4.3, and 4.4, we also need the analogues of Propositions 4.2, 4.3, and 4.4 with Λ replaced by the function $n \mapsto 1_{(n,W)=1}$. Proposition 4.7 below will be used in the deduction of Theorem 1.3, and Proposition 4.8 below will be used in the deductions of Theorems 1.4 and 1.6.

Proposition 4.7. *Let an integer $s \geq 1$, large real numbers $A, \Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/2)$ be given. There exists a constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$, such that for $x \geq 2$, $\eta > 0$ we have*

$$\sum_{d \leq x^{1/2-\varepsilon}} \max_{(c,d)=1} \sup_{\psi \in \Psi_s(\Delta, \eta^{-\kappa}; \eta, x/d)} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ (n,W)=1}} \psi((n-c)/d) \right| \ll \eta^\kappa x \log x + O_A(x(\log x)^{-A}),$$

where $W = \mathcal{P}(w)$ for some $w \leq \exp(\sqrt{\log x})$.

Proposition 4.8. *Let an integer $s \geq 1$, large real numbers $A, \Delta \geq 2$, and a small real number $\varepsilon \in (0, 1/2)$ be given. There exists a constant $\kappa = \kappa(s, \Delta, \varepsilon) > 0$, such that for any $x \geq 2$, $\eta > 0$ and any nilsequence $\psi \in \Psi_s(\Delta, \eta^{-\kappa}; \eta, x)$ we have*

$$\sum_{d \leq x^{1/2-\varepsilon}} \max_{(c,d)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ (n,W)=1}} \psi(n) \right| \ll \eta^\kappa x \log x + O_A(x(\log x)^{-A}),$$

where $W = \mathcal{P}(w)$ for some $w \leq \exp(\sqrt{\log x})$.

The proofs of Theorems 4.2–4.6 and of Propositions 4.7 and 4.8 will be given in Section 7, using the type I and type II estimates established in Sections 5 and 6.

4.2 Proof that Theorem 4.2 implies Theorem 1.3

We now show that these equidistributed cases for Λ imply the general cases of our main theorems. The proof of the analogous results for the Möbius function as mentioned in Remark 1.8 is completely similar, using Remark 4.5 (and in fact easier due to not needing the W -trick).

We start with the deduction of Theorem 1.3 from Theorem 4.2. We may assume that x is sufficiently large in terms of $A, s, \Delta, \varepsilon$. Let $1 \leq Q \leq x^{1/4-\varepsilon}$. For $q \in [Q, 2Q]$, let

$$f_q(n) := \Lambda(n) - \frac{qW}{\varphi(qW)} 1_{(n,W)=1}.$$

Let $\mathcal{B}$ be the set of $q \in [Q, 2Q]$ such that

$$\max_{(a,q)=1} \sup_{\psi \in \Psi_s(\Delta, \log x)} \left| \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} f_q(n) \psi(n) \right| \geq \frac{x}{q(\log x)^{2A}}.$$

We may assume that $|\mathcal{B}| \geq Q(\log x)^{-2A}$ (for some choice of Q), since otherwise the claim of Theorem 1.3 follows from the triangle inequality. For each $q \in \mathcal{B}$, we may pick some $(a, q) = 1$ and some $\psi \in \Psi_s(\Delta, \log x)$ such that

$$\left| \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} f_q(n) \psi((n-a)/q) \right| \geq \frac{x}{q(\log x)^{2A}}, \quad (4.1)$$

where we used the simple observation that if $\psi \in \Psi_s(\Delta, \log x)$ then the dilation $\psi'(n) = \psi(qn+a)$ is also in $\Psi_s(\Delta, \log x)$.

Lemma 4.9. *Let $\mathcal{B}$ and f_q be defined as above. For $q \in \mathcal{B}$ and any $B \geq 2$, we may find*

- $M \in [(\log x)^{3A}, (\log x)^{O_{A,B,s,\Delta}(1)}]$;
- $y \in [x(\log x)^{-O_{A,B,s,\Delta}(1)}, x]$;
- $q' = tq$ for some positive integer $t \leq M$;
- a residue class $a' \pmod{q'}$ with $0 \leq a' < q'$ and $a' \equiv a \pmod{q}$,

such that at least one of the following two conclusions hold:

1. For some $\psi' \in \Psi_s(\Delta, M^{O_{s,\Delta}(1)}; M^{-B}, y/q')$ we have

$$\left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \psi'((n-a')/q') \right| \gg \frac{x}{q'} (\log x)^{-O_{A,B,s,\Delta}(1)}.$$

2. We have

$$\left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,B,s,\Delta}(1)}.$$

Proof. This will be deduced from the factorization theorem for polynomial sequences (Theorem 3.7), in a similar way as in [32, Lemma 2.4]. Roughly speaking, the nilsequence $n \mapsto \psi((n-a)/q)$ becomes equidistributed after passing to a sub-progression of the form $\{n \in I : n \equiv a' \pmod{q'}\}$. On this sub-progression, we can write $\psi = \psi' + c_0$ for some equidistributed ψ' as in conclusion (1), where c_0 is the average of ψ on the sub-progression.

We now turn to the details. Write $\psi(n) = \phi(g(n)\Gamma)$ for some Lipschitz function ϕ on G/Γ and some polynomial sequence g . Let D be a constant sufficiently large in terms of B, s, Δ . By Theorem 3.7 applied with $M_0 = (\log x)^{3A}$, we may find $M \in [(\log x)^{3A}, (\log x)^{O_{A,D,s,\Delta}(1)}]$, a sub-nilmanifold G'/Γ' of G/Γ , a Mal'cev basis $\mathcal{X}'$ for G'/Γ' in which each element is an M -rational combination of the elements of $\mathcal{X}$, and a decomposition $g = \varepsilon g' \gamma$ into polynomial sequences ε, g', γ with the following properties:

1. ε is $(M, x/q)$ -smooth;
2. g' takes values in G' , and $\{g'(n)\}_{1 \leq n \leq x/q}$ is totally M^{-D} -equidistributed in G'/Γ' ;
3. γ is M -rational and is periodic with period $t \leq M$.

After a change of variables $n = qm + a$, we may write (4.1) as

$$\left| \sum_{m \leq x/q} f_q(qm + a) \varphi(g(m)\Gamma) \right| \geq \frac{x}{q} (\log x)^{-2A}.$$

Dividing $[0, x/q]$ into $O(M^2)$ intervals of equal length and then further dividing them into residue classes modulo t , we may find an interval $I \subset [0, x]$ with $|I| \asymp x/M^2$ and some residue class $b \pmod{t}$, such that

$$\left| \sum_{\substack{m \equiv b \pmod{t} \\ qm + a \in I}} f_q(qm + a) \varphi(\varepsilon(m)g'(m)\gamma(m)\Gamma) \right| \gg \frac{|I|}{qt} (\log x)^{-2A}. \quad (4.2)$$

Pick any m_0 counted in the sum (i.e. $m_0 \equiv b \pmod{t}$ and $qm_0 + a \in I$). Note that

$$\begin{aligned} & \left| \sum_{\substack{m \equiv b \pmod{t} \\ qm + a \in I}} f_q(qm + a) (\varphi(\varepsilon(m)g'(m)\gamma(m)\Gamma) - \varphi(\varepsilon(m_0)g'(m)\gamma(m)\Gamma)) \right| \\ & \ll (\log x) \sum_{\substack{m \equiv b \pmod{t} \\ qm + a \in I}} |\varphi(\varepsilon(m)g'(m)\gamma(m)\Gamma) - \varphi(\varepsilon(m_0)g'(m)\gamma(m)\Gamma)| \\ & \ll (\log x) \sum_{\substack{m \equiv b \pmod{t} \\ qm + a \in I}} d_{\mathcal{X}}(\varepsilon(m)g'(m)\gamma(m), \varepsilon(m_0)g'(m)\gamma(m)) \\ & = (\log x) \sum_{\substack{m \equiv b \pmod{t} \\ qm + a \in I}} d_{\mathcal{X}}(\varepsilon(m), \varepsilon(m_0)), \end{aligned}$$

where we used the right invariance of $d_{\mathcal{X}}$. By the smoothness property of ε , each term $d_{\mathcal{X}}(\varepsilon(m), \varepsilon(m_0))$ is $\ll M^{-1} \ll (\log x)^{-3A}$. Hence the expression above is $\ll \frac{|I|}{qt} (\log x)^{-3A+1}$, which is negligible compared with (4.2), so we can replace $\varepsilon(m)$ in (4.2) by $\varepsilon(m_0)$. Moreover, we may clearly replace $\gamma(m)$ in (4.2) by $\gamma(m_0)$, since γ has period t . Hence, after a change of variables $m = t\ell + b$, (4.2) becomes

$$\left| \sum_{\ell: qt\ell + qb + a \in I} f_q(qt\ell + qb + a) \varphi(\varepsilon(m_0)g'(t\ell + b)\gamma(m_0)\Gamma) \right| \gg \frac{|I|}{qt} (\log x)^{-2A}.$$

By writing the interval I as a difference of two intervals $[1, y_2] \setminus [1, y_1]$, we have for either $y = y_1$ or $y = y_2$,

$$\left| \sum_{\ell: qt\ell + qb + a \leq y} f_q(qt\ell + qb + a) \varphi(\varepsilon(m_0)g'(t\ell + b)\gamma(m_0)\Gamma) \right| \gg \frac{x}{qt} (\log x)^{-O_{A,B,S,\Delta}(1)}.$$

Note that for this to hold, we necessarily have $y \geq x(\log x)^{-O_{A,B,S,\Delta}(1)}$. Now let $\tilde{g}$ be the polynomial sequence defined by

$$\tilde{g}(\ell) = \gamma(m_0)^{-1} g'(t\ell + b) \gamma(m_0),$$

taking values in $\tilde{G} = \gamma(m_0)^{-1} G' \gamma(m_0)$, and let φ' be the Lipschitz function on $\tilde{G}/\tilde{\Gamma}$ (where $\tilde{\Gamma} = \tilde{G} \cap \Gamma$) defined by

$$\varphi'(x) = \varphi(\varepsilon(m_0)\gamma(m_0)x).$$

Then we have

$$\left| \sum_{\ell: q\ell + qb + a \leq y} f_q(qt\ell + qb + a) \phi'(\tilde{g}(\ell)\tilde{\Gamma}) \right| \gg \frac{x}{qt} (\log x)^{-O_{A,B,s,\Delta}(1)}.$$

Setting $q' = qt$ and $a' = qb + a$, the inequality above can be rewritten as

$$\left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \phi'(\tilde{g}((n - a')/q')\tilde{\Gamma}) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,B,s,\Delta}(1)}.$$

Let $c_0 = \int \phi'$ and $\tilde{\phi} = \phi' - c_0$, so that $\int \tilde{\phi} = 0$. Then either conclusion (2) holds, in which case we are done, or else we have

$$\left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \tilde{\phi}(\tilde{g}((n - a')/q')\tilde{\Gamma}) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,B,s,\Delta}(1)}.$$

Using some standard “quantitative nil-linear algebra” arguments (see the claim at the end of [18, Section 2]), one can verify the following properties:

- $\tilde{G}/\tilde{\Gamma}$ is a nilmanifold of dimension at most Δ , equipped with a filtration $\tilde{\mathcal{X}}$ of degree at most s and a $M^{O_{s,\Delta}(1)}$ -rational Mal'cev basis;
- $\|\tilde{\phi}\|_{\text{Lip}(\tilde{\mathcal{X}})} \leq M^{O_{s,\Delta}(1)} \leq (\log x)^{O_{A,D,s,\Delta}(1)}$;
- $\tilde{g}$ is a polynomial sequence on $\tilde{G}$ adapted to $\tilde{\mathcal{X}}$, and $\{\tilde{g}(\ell)\}_{1 \leq \ell \leq y/q'}$ is totally $M^{-cD+O(1)}$ -equidistributed in $\tilde{G}/\tilde{\Gamma}$ for some small constant $c = c(s, \Delta) > 0$.

By taking D large enough in terms of B, s, Δ , we may ensure that $M^{-cD+O(1)} \leq M^{-B}$. Hence conclusion (1) follows by taking $\psi'(\ell) = \tilde{\phi}(\tilde{g}(\ell)\tilde{\Gamma}) / \|\tilde{\phi}\|_{\text{Lip}(\tilde{\mathcal{X}})}$. $\square$

We now apply Lemma 4.9, with B a constant sufficiently large in terms of $A, s, \Delta, \varepsilon$. Note that the quantities M, y, t produced in Lemma 4.9 all depend on q . However, each of them can be chosen from at most $(\log x)^{O_{A,s,\Delta,\varepsilon}}$ possibilities: For t this is clear since $t \leq M$; For y , we may require in the proof that it is of the form $y = \lfloor kx/M^2 \rfloor$ for some positive integer $k \leq M^2$; For M , we may require that it is of the form $M = (\log x)^k$ for a positive integer $k = O_{A,s,\Delta,\varepsilon}(1)$. Hence, after restricting to a subset $\mathcal{B}'$ of $\mathcal{B}$ of size $|\mathcal{B}'| \gg Q(\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}$, we may find

- $M \in [(\log x)^{3A}, (\log x)^{O_{A,s,\Delta,\varepsilon}(1)}]$,
- $y \in [x(\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}, x]$;
- $q' = \ell q$ for some positive integer $\ell \leq M$,

such that at least one of the following two conclusions hold:

- For each $q \in \mathcal{B}'$ we have

$$\max_{\substack{a' \pmod{q'} \\ (a', q) = 1}} \sup_{\psi' \in \Psi_s(\Delta, M^{O_{s,\Delta}(1)}, M^{-B}, y/q')} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \psi'((n - a')/q') \right| \gg \frac{x}{q'} (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}. \quad (4.3)$$

- For each $q \in \mathcal{B}'$ we have

$$\max_{\substack{a' \pmod{q'} \\ (a', q) = 1}} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}. \quad (4.4)$$

First assume that (4.4) holds. We will show that this is a contradiction to the classical Bombieri–Vinogradov inequality. By the definition of $f_q(n)$, we have

$$\sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) = \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} \Lambda(n) - \frac{qW}{\varphi(qW)} \sum_{\substack{n \leq y \\ (n,W)=1 \\ n \equiv a' \pmod{q'}}} 1. \quad (4.5)$$

Case $(a', q') > 1$:

First consider the case when $(a', q') > 1$. Since $a' \equiv a \pmod{q}$ and $(a, q) = 1$, we must have $(a', q') = (a', \ell) \leq \ell \leq (\log x)^{O_{A,s,\Delta,\varepsilon}(1)}$. Hence we can ensure that $(a', q') \mid W^\infty$ by choosing C in the definition of W large enough in terms of $A, s, \Delta, \varepsilon$, and thus the second sum on the right hand side of (4.5) is empty, so (4.5) is $O(\log x)$ in this case, contradicting (4.4).

Case $(a', q') = 1$:

Now consider the case when $(a', q') = 1$. Let D be a constant sufficiently large in terms of $A, s, \Delta, \varepsilon$. By the classical Bombieri–Vinogradov theorem, we have

$$\sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} \Lambda(n) = \frac{y}{\varphi(q')} (1 + O((\log x)^{-D}))$$

for almost all $q \in \mathcal{B}'$. Let $\mathcal{B}''$ be the set of $q \in \mathcal{B}'$ satisfying the estimate above, so that $|\mathcal{B}''| \gg Q(\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}$. By the fundamental lemma of sieve theory [25, Lemma 6.3], the second term on the right hand side of (4.5) is

$$\frac{qW}{\varphi(qW)} \sum_{\substack{n \leq y \\ (n,W)=1 \\ n \equiv a' \pmod{q'}}} 1 = (1 + O((\log x)^{-D})) \frac{qW}{\varphi(qW)} \cdot \frac{y}{q'} \prod_{p|W, p \nmid q'} (1 - p^{-1}).$$

Since $qW/\varphi(qW) = q'W/\varphi(q'W)$ by the fact that $\ell \mid W^\infty$, the main term above can be computed to be

$$\frac{qW}{\varphi(qW)} \cdot \frac{y}{q'} \prod_{p|W, p \nmid q'} (1 - p^{-1}) = \frac{q'W}{\varphi(q'W)} \cdot \frac{y}{q'} \cdot \frac{\varphi(q'W)}{W\varphi(q')} = \frac{y}{\varphi(q')}.$$

Combining the above, we see that (4.5) is $\ll \frac{y}{\varphi(q')} (\log x)^{-D}$, contradicting (4.4) once D is large enough.

Now assume that (4.3) holds. We will show that this is a contradiction to either Theorem 4.2 or Proposition 4.7. By the definition of $f_q(n)$, either

$$\max_{\substack{a' \pmod{q'} \\ (a', q) = 1}} \sup_{\psi' \in \Psi_s(\Delta, M^{O_{s,\Delta}(1)}; M^{-B}, y/q')} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} \Lambda(n) \psi'((n - a')/q') \right| \gg \frac{x}{q'} (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)},$$

or

$$\max_{\substack{a' \pmod{q'} \\ (a', q')=1}} \sup_{\psi' \in \Psi_s(\Delta, M^{O_{s,\Delta}(1)}; M^{-B}, y/q')} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'} \\ (n, W)=1}} \psi'((n-a')/q') \right| \gg \frac{x}{q'} (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}$$

holds for $q \in \mathcal{B}'$. In both cases we must have $(a', q') = 1$, since otherwise the sum over y is $\ll \log x$. Since $y \geq x(\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}$, we have

$$q' = \ell q \leq x^{1/4-\varepsilon} (\log x)^{O_{A,s,\Delta,\varepsilon}(1)} \leq y^{1/4-\varepsilon/2}.$$

It follows that either

$$\sum_{q' \leq y^{1/4-\varepsilon/2}} \max_{(a', q')=1} \sup_{\psi' \in \Psi_s(\Delta, M^{O_{s,\Delta}(1)}; M^{-B}, y/q')} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} \Lambda(n) \psi' \left(\frac{n-a'}{q'} \right) \right| \gg x (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}, \quad (4.6)$$

or

$$\sum_{q' \leq y^{1/4-\varepsilon/2}} \max_{(a', q')=1} \sup_{\psi' \in \Psi_s(\Delta, M^{O_{s,\Delta}(1)}; M^{-B}, y/q')} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'} \\ (n, W)=1}} \psi' \left(\frac{n-a'}{q'} \right) \right| \gg x (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}. \quad (4.7)$$

On the other hand, Theorem 4.2 implies that the left-hand side of (4.6) is $\ll M^{-\kappa B} x (\log x)^2 \ll x (\log x)^{-\kappa B+2}$ for some $\kappa = \kappa(s, \Delta, \varepsilon) > 0$. This is a contradiction to (4.6) once B is large enough. Similarly, Proposition 4.7 implies that the left-hand side of (4.7) is $\ll M^{-\kappa B} x (\log x)^2 + O_D(x (\log x)^{-D})$ for any $D \geq 2$. This is a contradiction to (4.7) once B and D are large enough.

4.3 Proof that Theorems 4.3 and 4.4 imply Theorems 1.4 and 1.6

In the deductions of Theorems 1.4 and 1.6, we have a fixed nilsequence ψ and we will apply the factorization theorem to reduce to the case when it is equidistributed at scale x . In comparison, we had nilsequences ψ_q varying with q in Section 4.2 and we applied the factorization theorem to reduce to the case when each ψ_q is equidistributed at scale x/q . Apart from this difference, the rest of the arguments are very much the same, so we will be brief with the arguments in this section.

We focus on the deduction of Theorem 1.4 from Theorem 4.3, since the deduction of Theorem 1.6 from Theorem 4.4 is analogous. We may assume that x is sufficiently large in terms of $A, s, \Delta, \varepsilon$. Let $1 \leq Q \leq x^{1/3-\varepsilon}$. For $q \in [Q, 2Q]$, let f_q be defined as before. Let $\mathcal{B}$ be the set of $q \in [Q, 2Q]$ such that

$$\max_{(a,q)=1} \left| \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} f_q(n) \psi(n) \right| \geq \frac{x}{q (\log x)^{2A}}.$$

We may assume that $|\mathcal{B}| \geq Q (\log x)^{-2A}$ (for some choice of Q), since otherwise the claim of Theorem 1.4 follows from the triangle inequality.

Lemma 4.10. *Let $\mathcal{B}$ and f_q be defined as above. For $q \in \mathcal{B}$ and any $B \geq 2$, we may find*

- $M \in [(\log x)^{3A}, (\log x)^{O_{A,B,s,\Delta}(1)}];$

- $y \in [x(\log x)^{-O_{A,B,s,\Delta}(1)}, x]$;
- $q' = tq$ for some positive integer $t \leq M$;
- a residue class $a' \pmod{q'}$ with $0 \leq a' < q'$ and $a' \equiv a \pmod{q}$,

such that at least one of the following two conclusions hold:

1. For some $\psi' \in \Psi_s(\Delta, M^{O_{s,\Delta}(1)}; M^{-B}, y)$ independent of q we have

$$\left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \psi'(n) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,B,s,\Delta}(1)}.$$

2. We have

$$\left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,B,s,\Delta}(1)}.$$

We omit the proof of Lemma 4.10, since its deduction from the factorization theorem for polynomial sequences (Theorem 3.7) is completely analogous to that of Lemma 4.9.

We now apply Lemma 4.10, with B a constant sufficiently large in terms of $A, s, \Delta, \varepsilon$. After restricting to a subset $\mathcal{B}'$ of $\mathcal{B}$ of size $|\mathcal{B}'| \gg Q(\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}$, we may find

- $M \in [(\log x)^{3A}, (\log x)^{O_{A,s,\Delta,\varepsilon}(1)}]$,
- $y \in [x(\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}, x]$;
- $q' = \ell q$ for some positive integer $\ell \leq M$,

(with all of M, y, ℓ independent of q), such that at least one of the following two conclusions hold:

- For each $q \in \mathcal{B}'$ we have

$$\max_{\substack{a' \pmod{q'} \\ (a', q)=1}} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \psi'(n) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}. \quad (4.8)$$

- For each $q \in \mathcal{B}'$ we have

$$\max_{\substack{a' \pmod{q'} \\ (a', q)=1}} \left| \sum_{\substack{n \leq y \\ n \equiv a' \pmod{q'}}} f_q(n) \right| \gg \frac{x}{q'} (\log x)^{-O_{A,s,\Delta,\varepsilon}(1)}. \quad (4.9)$$

(4.9) is analyzed in the same way as (4.4); it leads to a contradiction to the classical Bombieri–Vinogradov inequality. (4.8) is analyzed in the same way as (4.3); it leads to a contradiction to either Theorem 4.3 or Proposition 4.8.

5 TYPE I ESTIMATES

Our remaining task for proving the main theorems is proving the results for equidistributed nilsequences from Section 4. By applying Vaughan's identity, our sum of interest

$$\sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \Lambda(n) \psi((n-c)/d)$$

can be decomposed in terms of type I sums of the form

$$\sum_{\substack{mn \leq x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m \psi((mn-c)/d) \quad \text{or} \quad \sum_{\substack{mn \leq x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m (\log n) \psi((mn-c)/d)$$

for some $M \leq x^{1/3}$, and type II sums of the form

$$\sum_{\substack{mn \leq x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m b_n \psi((mn-c)/d)$$

for some $x^{1/3} \leq M \leq x^{2/3}$. We begin by analyzing the type I sums. In our argument we will repeatedly use the following recurrence result, which is a standard consequence of Weyl's inequality (see [19, Lemma 4.5]).

Lemma 5.1. *Let $0 < \delta < 1/2$ and $0 < \varepsilon < \delta/4$. Let $g(n)$ be a polynomial of degree s . Let $N \geq \delta^{-C}$ for some sufficiently large constant $C = C(s)$, and suppose that for at least δN values of the integers $1 \leq n \leq N$ we have $\|g(n)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon$. Then there exists a positive integer $q \ll \delta^{-O_s(1)}$ such that*

$$\|q\alpha_i\|_{\mathbb{R}/\mathbb{Z}} \ll \varepsilon \delta^{-O_s(1)} N^{-i}$$

for each $0 \leq i \leq s$, where α_i is the coefficient of n^i in $g(n)$.

Note that the assumption $N \geq \delta^{-C}$ is necessary, even though this is not explicitly mentioned in [19, Lemma 4.5]. However, when $g(n)$ is a monomial $g(n) = \alpha_s n^s$, the conclusion of Lemma 5.1 holds for any $N \geq 1$. Note also that [19, Lemma 4.5] only gives the conclusion for $1 \leq i \leq s$, but the bounds for $\|q\alpha_i\|$ combined with our assumption that $\|g(n)\| \leq \varepsilon$ for some $1 \leq n \leq N$ gives the conclusion for $i = 0$.

Throughout the rest of the paper, we will use the notation

$$\|a\|_p := \left(\frac{1}{M} \sum_m |a_m|^p \right)^{1/p}, \quad (5.1)$$

where $M \geq 1$ is the smallest element of the support of $\{a_m\}$. The following type I estimate suffices for Theorems 4.2.

Proposition 5.2 (type I estimate with supremum). *Let $x \geq 2$ and $\varepsilon > 0$. Let $1 \leq M \leq x^{1/2}$ and $1 \leq D \leq x^{1/2-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, and $0 < \delta < 1/2$. Then there exists a large constant $C = C(s, \Delta, \varepsilon)$ such that*

$$\sum_{D \leq d \leq 2D} \max_{(c,d)=1} \sup_{\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x/D)} \left| \sum_{\substack{mn \leq x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m \psi((mn-c)/d) \right| \ll \delta x \|a\|_2$$

for any sequence $\{a_m\}$.

By applying the triangle inequality and Cauchy–Schwarz, Proposition 5.2 reduces to the following slightly more general bound (replacing δ by δ^2 there). Note that, in (5.2) below, the maximum is taken over all residue classes $c \pmod{d}$, not just those coprime with d .

Proposition 5.3. *Let $x \geq 2$ and $\varepsilon > 0$. Let $1 \leq M \leq x^{1/2}$ and $1 \leq D \leq x^{1/2-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, and $0 < \delta < 1/2$. Then there exists a large constant $C = C(s, \Delta, \varepsilon)$ such that*

$$\sum_{D \leq d \leq 2D} \max_c \sup_{\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x/D)} \sum_{\substack{M \leq m \leq 2M \\ (m, d)=1}} \left| \sum_{\substack{n \leq x/m \\ n \equiv cm^{-1} \pmod{d}}} \psi((mn - c)/d) \right| \ll \delta x. \quad (5.2)$$

Proof. For ease of notation we will suppress the dependence of implied constants on s, Δ, ε . We may assume that $\delta > x^{-1/C}$, since otherwise $\Psi_s(\Delta, \delta^{-1}; \delta^C, x/D)$ is empty.

Let T be the set of $(d, m) \in [D, 2D] \times [M, 2M]$ for which the inner sum over n in (5.2) is at least $\delta x/(DM)$. It suffices to show that

$$|T| \ll \delta DM.$$

For each $d \in [D, 2D]$, let $\psi_d \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x/D)$ be such that the supremum in (5.2) is attained by ψ_d . Also let $c_d \in [0, d)$ be such that the maximum over c in (5.2) is attained by this value. Write $\psi_d(n) = \phi_d(g_d(n)\Gamma_d)$ for some Lipschitz function ϕ_d on a nilmanifold G_d/Γ_d and some polynomial sequence g_d , with $\{g_d(n)\}_{1 \leq n \leq x/D}$ totally δ^C -equidistributed. For $(d, m) \in T$, consider the polynomial sequence $h_{d,m}$ defined by

$$h_{d,m}(n) := g_d(mn + b_{d,m}),$$

where $b_{d,m}$ is the unique number satisfying $1 \leq b_{d,m} \leq m$ and $db_{d,m} \equiv -c_d \pmod{m}$. Then by a change of variables we have

$$\sum_{\substack{n \leq x/m \\ n \equiv c_d m^{-1} \pmod{d}}} \psi_d((mn - c_d)/d) = \sum_{n \leq x/(md)} \phi_d(h_{d,m}(n)\Gamma_d) + O(1).$$

If $(d, m) \in T$, then the sums above have size at least $\delta x/(2DM)$, and thus the sequence $(h_{d,m}(n)\Gamma_d)_{1 \leq n \leq x/(DM)}$ fails to be totally δ^2 -equidistributed. Hence by Theorem 3.5, there is a nontrivial horizontal character $\chi_{d,m}: G_d \rightarrow \mathbb{R}/\mathbb{Z}$ with $\|\chi_{d,m}\| \ll \delta^{-O(1)}$ such that

$$\|\chi_{d,m} \circ h_{d,m}\|_{C^\infty(x/DM)} \ll \delta^{-O(1)}.$$

Since the number of such characters is $\delta^{-O(1)}$, there is a subset $T' \subset T$ with $|T'| \gg \delta^{O(1)}|T|$ and a nontrivial horizontal character χ (not depending on d, m) with $\|\chi\| \ll \delta^{-O(1)}$, such that

$$\|\chi \circ h_{d,m}\|_{C^\infty(x/DM)} \ll \delta^{-O(1)}$$

for all $(d, m) \in T'$. Suppose, for the purpose of contradiction, that $|T| \gg \delta^3 DM$, so that $|T'| \gg \delta^{O(1)} DM$. Then there are at least $\delta^{O(1)} D$ values of d such that $(d, m) \in T'$ for at least $\delta^{O(1)} M$ values of m . For each such d , we will show that $\{g_d(n)\}_{1 \leq n \leq x/D}$ is not totally δ^C -equidistributed, a contradiction.

Fix any such d for the rest of the proof. For $(d, m) \in T'$, we may explicitly write

$$\chi \circ h_{d,m}(n) = \sum_{i=0}^s \beta_{d,m,i} n^i$$

for some coefficients $\beta_{d,m,i} \in \mathbb{R}$. Then by Lemma 3.4, there is a positive integer $q = O(1)$ such that

$$\|q\beta_{d,m,i}\| \ll (x/DM)^{-i} \|\chi \circ h_{d,m}\|_{C^\infty(x/DM)} \ll (x/DM)^{-i} \delta^{-O(1)}. \quad (5.3)$$

Using the definition of $h_{d,m}$, we see that

$$\chi \circ h_{d,m}(n) = \chi \circ g_d(mn + b_{d,m}).$$

Thus if we write

$$\chi \circ g_d(n) = \sum_{i=0}^s \alpha_{d,i} n^i$$

for some coefficients $\alpha_{d,i} \in \mathbb{R}$, then by the binomial formula we have the relations

$$\alpha_{d,i} m^i = \beta_{d,m,i} - m^i \sum_{i < j \leq s} \binom{j}{i} \alpha_{d,j} b_{d,m}^{j-i}. \quad (5.4)$$

At this point we simply have a Diophantine problem to solve: we know that the $\beta_{d,m,i}$'s lie in "major arcs" (of appropriate width) for many m , and we would like to deduce that the $\alpha_{d,i}$'s also lie in major arcs (of appropriate width). We claim that there is a positive integer $k \ll \delta^{-O(1)}$, such that

$$\|kq\alpha_{d,i}\| \ll (x/D)^{-i} \delta^{-O(1)} \quad (5.5)$$

for each $1 \leq i \leq s$. To prove the claim we proceed by induction. For $i = s$ we have $\alpha_{d,s} m^s = \beta_{d,m,s}$. From (5.3) we see that

$$\|q\alpha_{d,s} m^s\| \ll (x/DM)^{-s} \delta^{-O(1)}$$

for at least $\delta^{O(1)}M$ values of m . By our assumption that $\delta > x^{-1/C}$ for some sufficiently large C , we have $(x/DM)^{-s} \leq x^{-\varepsilon} \leq \delta^{O(1)}$. Hence we may apply Lemma 5.1 to the polynomial $m \mapsto q\alpha_{d,s} m^s$ to conclude that

$$\|kq\alpha_{d,s}\| \ll (x/D)^{-s} \delta^{-O(1)}$$

for some positive integer $k \ll \delta^{-O(1)}$, completing the proof of the base case $i = s$. Now take $1 \leq i < s$, and assume that (5.5) with some $k \ll \delta^{-O(1)}$ has already been proven for larger values of i . By (5.3), (5.4) and the induction hypothesis, we see that

$$\|kq\alpha_{d,i} m^i\| \ll \|kq\beta_{d,m,i}\| + \sum_{i < j \leq s} M^j \|kq\alpha_{d,j}\| \ll (x/DM)^{-i} \delta^{-O(1)}$$

holds for at least $\delta^{O(1)}M$ values of m . By Lemma 5.1 applied to the polynomial $m \mapsto kq\alpha_{d,i} m^i$, there exists a positive integer $k' \ll \delta^{-O(1)}$ such that

$$\|k'kq\alpha_{d,i}\| \ll (x/D)^{-i} \delta^{-O(1)}.$$

Replacing k by $k'k$ completes the induction step, and finishes the proof of the claim (5.5).

Finally, by Lemma 3.4 we have

$$\|kq\chi \circ g_d\|_{C^\infty(x/D)} \ll \delta^{-O(1)}.$$

Since $kq\chi$ is nontrivial and $\|kq\chi\| \ll \delta^{-O(1)}$, Lemma 3.6 implies that $\{g_d(n)\}_{1 \leq n \leq x/D}$ is not totally $\delta^{O(1)}$ -equidistributed, a contradiction if C is chosen large enough. $\square$

The following type I estimate suffices for Theorems 4.3 and 4.4.

Proposition 5.4 (type I estimate). *Let $x \geq 2$ and $\varepsilon > 0$. Let $1 \leq M \leq x^{1/2}$ and $1 \leq D \leq x^{1/2-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, and $0 < \delta < 1/2$. Let $\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x)$ for some sufficiently large constant $C = C(s, \Delta, \varepsilon)$. Then*

$$\sum_{D \leq d \leq 2D} \max_{(c,d)=1} \left| \sum_{\substack{mn \leq x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m \psi(mn) \right| \ll \delta x \|a\|_2$$

for any sequence $\{a_m\}$.

As with Proposition 5.2, by applying the triangle inequality and Cauchy–Schwarz, we reduce Proposition 5.4 to the following slightly more general statement, with no coprimality condition on $c \pmod{d}$.

Proposition 5.5. *Let $x \geq 2$ and $\varepsilon > 0$. Let $1 \leq M \leq x^{1/2}$ and $1 \leq D \leq x^{1/2-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, and $0 < \delta < 1/2$. Let $\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x)$ for some sufficiently large constant $C = C(s, \Delta, \varepsilon)$. Then*

$$\sum_{D \leq d \leq 2D} \max_c \sum_{\substack{M \leq m \leq 2M \\ (m,d)=1}} \left| \sum_{\substack{mn \leq x \\ n \equiv cm^{-1} \pmod{d}}} \psi(mn) \right| \ll \delta x. \quad (5.6)$$

Proof. Again, for ease of notation we will suppress the dependence of implied constants on s, Δ, ε . As before, we may assume that $\delta > x^{-1/C}$. Let $c_d \in [0, d)$ be such that the inner sum in (5.6) is maximal with $c = c_d$. Write $\psi(n) = \varphi(g(n)\Gamma)$ for some Lipschitz function φ on a nilmanifold G/Γ and some polynomial sequence g , with $\{g(n)\}_{1 \leq n \leq x}$ totally δ^C -equidistributed. This is the same sum Σ that appears in the proof of Proposition 5.2 if $\psi_d(n) = \varphi(g_d(n))$ with $g_d(n) = g(dn + c_d)$.

Suppose that the desired bound (5.6) fails. Then, following the proof of Proposition 5.2 up to the claim (5.5), we find a nontrivial horizontal character χ with $\|\chi\| \ll \delta^{-O(1)}$, such that if we write

$$\chi \circ g_d(n) = \sum_{i=0}^s \alpha_{d,i} n^i,$$

then there is a positive integer $q \ll \delta^{-O(1)}$ with

$$\|q\alpha_{d,i}\| \ll (x/D)^{-i} \delta^{-O(1)} \quad (5.7)$$

for each $1 \leq i \leq s$ and at least $\delta^{O(1)}D$ values of $D \leq d \leq 2D$. Now, if we write

$$\chi \circ g(n) = \sum_{i=0}^s \beta_i n^i,$$

then from the relations $g_d(n) = g(dn + c_d)$ we see that

$$\alpha_{d,i} = d^i \sum_{j \leq s} \binom{j}{i} \beta_j c_d^{j-i}. \quad (5.8)$$

We are then left with the Diophantine problem of deducing from (5.7) that there exists a positive integer $k \ll \delta^{-O(1)}$, such that

$$\|kq\beta_i\| \ll x^{-i} \delta^{-O(1)} \quad (5.9)$$

for each $1 \leq i \leq s$. We prove (5.9) by induction. Since $\alpha_{d,s} = d^s \beta_s$, we see that

$$\|q\beta_s d^s\| \ll (x/D)^{-i} \delta^{-O(1)}$$

for at least $\delta^{O(1)}D$ values of d . It follows from Lemma 5.1 that there is a positive integer $k \ll \delta^{-O(1)}$ such that (5.9) holds for $i = s$.

Now take $1 \leq i < s$, and assume that (5.9) with some $k \ll \delta^{-O(1)}$ has already been proven for larger values of i . By (5.7), (5.8) and the induction hypothesis, we see that

$$\|kq d^i \beta_i\| \ll \|kq \alpha_{d,i}\| + \sum_{i < j \leq s} D^j \|kq \beta_j\| \ll (x/D)^{-i} \delta^{-O(1)}$$

holds for at least $\delta^{O(1)}D$ values of d . By Lemma 5.1 again, there exists a positive integer $k' \ll \delta^{-O(1)}$ such that

$$\|k' k q \beta_i\| \ll x^{-i} \delta^{-O(1)}.$$

Replacing k by $k'k$ completes the induction step.

Finally, by Lemmas 3.4 and 3.6, (5.9) implies that $\{g(n)\}_{1 \leq n \leq x}$ is not totally $\delta^{O(1)}$ -equidistributed, a contradiction if C is chosen large enough. $\square$

6 TYPE II ESTIMATES

We proceed to the type II estimates, which are more delicate than the type I estimates. In particular, each of Theorems 1.3, 1.4 and 1.6 requires a different estimate. Despite this, the general structure of the proofs is still similar:

1. We start with appropriate applications of Cauchy–Schwarz to eliminate the sequences $\{a_m\}, \{b_n\}$.
2. Assuming that the desired estimate fails, we then use the quantitative Leibman theorem (Theorem 3.5) to obtain recurrence properties of certain polynomials.
3. At this point we face a purely elementary Diophantine problem of deducing from the recurrence properties that the coefficients lie in appropriate major arcs.

The readers are again encouraged to consider the special case when the nilsequences are polynomial phase functions, in which case step (2) above is essentially no longer needed. Propositions 6.3, 6.7, and 6.5 below are the required type II estimates for Theorems 4.2, 4.3, and 4.4, respectively.

We first need a lemma that generalizes Lemma 5.1 to polynomials in more than one variables.

Lemma 6.1. *Let $0 < \delta < 1/2$ be a parameter, and let $k, d \geq 1$, $0 < \varepsilon < \delta/4^k$. Let $P(x_1, \dots, x_k)$ be a polynomial of total degree d with real coefficients, and let $N_1, \dots, N_k \geq \delta^{-C}$ for some sufficiently large constant $C = C(d, k)$. Suppose that for proportion $\geq \delta$ of the integer tuples $(x_1, \dots, x_k) \in [N_1] \times \dots \times [N_k]$ we have*

$$\|P(x_1, \dots, x_k)\| \leq \varepsilon.$$

Then there exists an integer $1 \leq q \ll \delta^{-O_{d,k}(1)}$ such that

$$\|q\alpha_{(i_1, \dots, i_k)}\| \ll \varepsilon \delta^{-O_{d,k}(1)} N_1^{-i_1} \dots N_k^{-i_k},$$

for each $i_1, \dots, i_k \geq 0$, where $\alpha_{(i_1, \dots, i_k)}$ the coefficient of $x_1^{i_1} \dots x_k^{i_k}$ in P .

Proof. We perform an induction on k . The $k = 1$ case follows immediately from Lemma 5.1. Suppose that the case k has been proved and that we are considering case $k + 1$. Consider those values of $1 \leq y \leq N_{k+1}$ with the property that $\|P(x_1, \dots, x_k, y)\| \leq \varepsilon$ for at least $\delta/2$ -proportion of $(x_1, \dots, x_k) \in [N_1] \times \dots \times [N_k]$. The proportion of such $y \in [N_{k+1}]$ is at least $\delta/2$. For each such y , consider the k -variable polynomial Q_y given by

$$Q_y(x_1, \dots, x_k) = P(x_1, \dots, x_k, y).$$

The coefficient of $x_1^{i_1} \dots x_k^{i_k}$ in Q_y is given by

$$\sum_{i_{k+1} \geq 0} \alpha_{(i_1, \dots, i_{k+1})} y^{i_{k+1}}.$$

Hence, by the induction hypothesis applied to Q_y , there exists a positive integer $q' = q'(y) \ll \delta^{-O_{d,k}(1)}$ such that

$$\left\| q' \sum_{i_{k+1} \geq 0} \alpha_{(i_1, \dots, i_{k+1})} y^{i_{k+1}} \right\| \ll \varepsilon \delta^{-O_{d,k}(1)} N_1^{-i_1} \dots N_k^{-i_k} \quad (6.1)$$

for each $i_1, \dots, i_k \geq 0$. By the pigeonhole principle, there exists a value of $q' \ll \delta^{-O_{d,k}(1)}$ such that the above inequality holds for $\delta' \geq \delta^{O_{d,k}(1)}$ proportion of $y \in [N_{k+1}]$. For fixed $i_1, \dots, i_k$, the left-hand side in (6.1) is a polynomial in y . Since $N_1, \dots, N_k \geq \delta^{-C(d,k)}$, we may ensure that the upper bound in (6.1) is less than $\delta'/4^k$ by choosing $C(d, k)$ large enough. Hence we may apply Lemma 5.1 to obtain a positive integer $q = q(i_1, \dots, i_k) \ll \delta^{-O_{d,k}(1)}$, such that

$$\|qq' \alpha_{(i_1, \dots, i_{k+1})}\| \ll \varepsilon \delta^{-O_{d,k}(1)} N_1^{-i_1} \dots N_{k+1}^{-i_{k+1}}$$

for each $i_{k+1} \geq 0$. Finally, we can make this q independent of $(i_1, \dots, i_k)$ by taking it to be the product of all the $q(i_1, \dots, i_k)$. $\square$

Recall that, in the single variable case (Lemma 5.1), the assumption $N \geq \delta^{-C}$ can be removed if g is a monomial. This phenomenon extends to the multi-variable case as follows.

Lemma 6.2. *Let $0 < \delta < 1/2$ be a parameter, and let $k, d \geq 1$, $0 < \varepsilon < \delta/4^k$. Let $N_1, \dots, N_k$ be positive integers. For a sufficiently large constant $C = C(d, k)$, let $J = \{1 \leq j \leq k : N_j \geq \delta^{-C}\}$. Let $P(x_1, \dots, x_k)$ be a polynomial of total degree d , with $\alpha_{(i_1, \dots, i_k)}$ the coefficient of $x_1^{i_1} \dots x_k^{i_k}$ in P . Let $a_1, \dots, a_k \geq 0$. Assume that $\alpha_{(i_1, \dots, i_k)} = 0$ whenever $i_j = a_j$ for each $j \in J$ and $(i_1, \dots, i_k) \neq (a_1, \dots, a_k)$.*

Suppose that for proportion $\geq \delta$ of the integer tuples $(x_1, \dots, x_k) \in [N_1] \times \dots \times [N_k]$ we have

$$\|P(x_1, \dots, x_k)\| \leq \varepsilon.$$

Then there exists an integer $1 \leq q \ll \delta^{-O_{d,k,C}(1)}$ such that

$$\|q \alpha_{(a_1, \dots, a_k)}\| \ll \varepsilon \delta^{-O_{d,k}(1)} N_1^{-a_1} \dots N_k^{-a_k}.$$

In short, we can still obtain Diophantine information on the coefficient $\alpha_{(a_1, \dots, a_k)}$, provided that $x_1^{a_1} \dots x_k^{a_k}$ is the only term in P whose x_j -degree is equal to a_j for each $j \in J$.

Proof. If $J = \emptyset$, then P is the monomial $P(x_1, \dots, x_k) = \alpha_{(a_1, \dots, a_k)} x_1^{a_1} \cdots x_k^{a_k}$, and the conclusion holds trivially.

Now suppose that J is nonempty. By the pigeonhole principle, there exists x'_j for each $j \notin J$, such that for proportion $\geq \delta$ of the integer tuples $(x'_j)_{j \in J} \in \prod_{j \in J} [N_j]$ we have $\|P(x'_1, \dots, x'_k)\| \leq \varepsilon$. By fixing these choices of x'_j for $j \notin J$, we may view P as a polynomial in the variables $(x_j)_{j \in J}$ and apply Lemma 6.1. Our assumptions imply that the coefficient of $\prod_{j \in J} x_j^{a_j}$ is $\alpha_{(a_1, \dots, a_k)} \prod_{j \notin J} (x'_j)^{a_j}$. Hence there exists a positive integer $q \ll \delta^{-O_{d,k}(1)}$ such that

$$\left\| q \alpha_{(a_1, \dots, a_k)} \prod_{j \notin J} (x'_j)^{a_j} \right\| \ll \varepsilon \delta^{-O_{d,k}(1)} \prod_{j \in J} N_j^{-a_j}.$$

The conclusion follows after replacing q by $q \prod_{j \notin J} x_j^{a_j}$, since $x_j \leq N_j \leq \delta^{-C}$ for $j \notin J$. $\square$

We are now ready to state and prove our type II estimates. Recall our notation $\|a\|_p$ from (5.1).

Proposition 6.3 (type II estimate with supremum). *Let $x \geq 2$ and $M \in [x^{1/4}, x^{3/4}]$. Let $1 \leq D \leq M^{1/2-\varepsilon}$ for some small $\varepsilon > 0$. Let $s \geq 1$, $\Delta \geq 2$, and $0 < \delta < (\log x)^{-1}$. Then for a large constant $C = C(s, \Delta, \varepsilon)$ we have*

$$\sum_{D \leq d \leq 2D} \max_{(c,d)=1} \sup_{\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C x/D)} \left| \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m b_n \psi((mn - c)/d) \right| \ll \delta x \|a\|_2 \|b\|_4,$$

for any sequences $\{a_m\}, \{b_n\}$.

Proof. As in the proof of the type I estimates, we may assume that $\delta > x^{-1/C}$, since otherwise $\Psi_s(\Delta, \delta^{-1}; \delta^C x/D)$ is empty. We need to bound the sum

$$\Sigma := \sum_{D \leq d \leq 2D} \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M \\ mn \equiv c_d \pmod{d}}} a_m b_n \psi_d((mn - c_d)/d)$$

for any choice of $1 \leq c_d \leq d$ with $(c_d, d) = 1$ and $\psi_d \in \Psi_s(\Delta, \delta^{-1}; \delta^C x/D)$. This can be rewritten as

$$\Sigma = \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M}} a_m b_n F(mn),$$

where the function F is defined by

$$F(k) := \sum_{\substack{D \leq d \leq 2D \\ k \equiv c_d \pmod{d}}} \psi_d((k - c_d)/d)$$

for $x \leq k \leq 2x$. By the Cauchy–Schwarz inequality, we deduce

$$\Sigma^2 \ll M \|a\|_2^2 \sum_{N/2 \leq n_1, n_2 \leq 2N} |b_{n_1}| |b_{n_2}| \left| \sum_{\substack{M \leq m \leq 2M \\ x/n_1 \leq m \leq 2x/n_1 \\ x/n_2 \leq m \leq 2x/n_2}} F(mn_1) \overline{F(mn_2)} \right|,$$

where $N = x/M$. By Cauchy–Schwarz, this is further

$$\ll MN \|a\|_2^2 \|b\|_4^2 \left(\sum_{N/2 \leq n_1, n_2 \leq 2N} \left| \sum_{\substack{M \leq m \leq 2M \\ x/n_1 \leq m \leq 2x/n_1 \\ x/n_2 \leq m \leq 2x/n_2}} F(mn_1) \overline{F(mn_2)} \right|^2 \right)^{1/2}.$$

Using the trivial bound $O(M)$ for the inner sum over m , we get

$$\Sigma^2 \ll M^{1/2} x \|a\|_2^2 \|b\|_4^2 \left(\sum_{N/2 \leq n_1, n_2 \leq 2N} \left| \sum_{\substack{M \leq m \leq 2M \\ x/n_1 \leq m \leq 2x/n_1 \\ x/n_2 \leq m \leq 2x/n_2}} F(mn_1) \overline{F(mn_2)} \right| \right)^{1/2}.$$

If the desired bound $\Sigma \ll \delta x \|a\|_2 \|b\|_4$ fails, then

$$\sum_{N/2 \leq n_1, n_2 \leq 2N} \left| \sum_{\substack{M \leq m \leq 2M \\ x/n_1 \leq m \leq 2x/n_1 \\ x/n_2 \leq m \leq 2x/n_2}} F(mn_1) \overline{F(mn_2)} \right| \geq \delta^4 N^2 M. \quad (6.2)$$

We are now in a position to apply [32, Lemma 3.3]. For it to be applicable, we need the following two conditions:

$$D^2 \leq \delta^C M, \quad N^{-c} < \delta^4 < (\log D)^{-1}$$

for some sufficiently large constant $C = C(s, \Delta)$ and sufficiently small constant $c = c(s, \Delta) > 0$. Both conditions are satisfied by our assumption on δ . Hence [32, Lemma 3.3] implies that $\psi_d \notin \Psi_s(\Delta, \delta^{-1}; \delta^{O_{s,\Delta}(1)}, x/D)$ for some d , a contradiction. $\square$

Remark 6.4. We briefly mention here how the sum (6.2) is treated in [32, Lemma 3.3], and where the key condition $D^2 \leq \delta^C M$ comes from. After expanding out using the definition of F , the sum over m can be written as

$$\sum_m F(mn_1) \overline{F(mn_2)} = \sum_{d_1, d_2} \sum_{m: mn_i \equiv c_{d_i} \pmod{d_i}} \psi_{d_1}((mn_1 - c_{d_1})/d_1) \overline{\psi_{d_2}((mn_2 - c_{d_2})/d_2)}.$$

The inner sum above is of length $\asymp M/D^2$ (assuming that $(d_1, d_2) = 1$), and can be expressed in the form (after a change of variables)

$$\sum_{m \in J(d_1, d_2, n_1, n_2)} \psi_{d_1, d_2, n_1, n_2}(m)$$

for an appropriate interval $J(d_1, d_2, n_1, n_2)$ and nilsequence $\psi_{d_1, d_2, n_1, n_2}$. One can apply the quantitative Leibman theorem to obtain recurrence property for the underlying polynomial sequences involved in $\psi_{d_1, d_2, n_1, n_2}$, for many (d_1, d_2, n_1, n_2) , and then proceed to solve the Diophantine problem that shows up. Since the length of this sum is M/D^2 , these recurrence properties are only meaningful when $M/D^2 \geq \delta^{-C}$ for some large constant C . We also remark here that the original statement in [32, Lemma 3.3] has an error: the assumption $10Q^2 \leq L$ there should be $Q^2 \leq \delta^C L$ for some sufficiently large constant C in terms of s, d . This is needed in the proof in order for Lemma 5.1 to be applicable.

We shall then state and prove a type II estimate with well-factorable weights.

Proposition 6.5 (Well-factorable type II estimate). *Let $x \geq 2$, let $c \neq 0$ be a fixed integer, and let $M \in [x^{1/2}, x^{3/4}]$. Let $\varepsilon > 0$ be a small constant, and let (λ_d) be a well-factorable sequence of level $x^{1/2-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, $0 < \delta < 1/2$. Then for a large constant $C = C(s, \Delta, \varepsilon)$ and $\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x)$, we have*

$$\left| \sum_{\substack{d \leq x^{1/2-\varepsilon} \\ (d, c) = 1}} \lambda_d \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m b_n \psi(mn) \right| \ll \delta x \|a\|_2 \|b\|_4 (\log x)^3, \quad (6.3)$$

for any sequences $\{a_m\}, \{b_n\}$.

By the well-factorability of $(\lambda_d)_{d \leq x^{1/2-\varepsilon}}$, there exist $(\gamma_d), (\theta_d)$ with $|\gamma_d|, |\theta_d| \leq 1$ such that $\lambda_d = \sum_{d=d_1 d_2} \gamma_{d_1} \theta_{d_2}$, and with (γ_d) and (θ_d) being supported on $[1, U]$ and $[1, V]$, respectively, where

$$U = x^{1-\varepsilon}/M, \quad V = Mx^{-\frac{1}{2}}. \quad (6.4)$$

Therefore, Proposition 6.5 is an immediate consequence of the following more general statement.

Proposition 6.6. *Let $x \geq 2$, let $c' \neq 0$ be a fixed integer, and let $M \in [x^{1/2}, x^{3/4}]$. Let $\varepsilon > 0$ be a small constant, and let $U, V \geq 1$ satisfy $U \leq x^{1-\varepsilon}/M$, $UV \leq x^{1/2-\varepsilon}$, and $UV^2 \leq Mx^{-\varepsilon}$. Let $s \geq 1$, $\Delta \geq 2$, $0 < \delta < 1/2$. Then for a large constant $C = C(s, \Delta, \varepsilon)$ and $\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x)$, we have*

$$\sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v \leq 2V \\ (v, c'u) = 1}} \max_{(c, u) = 1} \left| \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{u} \\ mn \equiv c' \pmod{v}}} a_m b_n \psi(mn) \right| \ll \delta x \|a\|_2 \|b\|_4 \log x,$$

for any sequences $\{a_m\}, \{b_n\}$.

Proof. As in the proof of the type I estimates, we suppress the dependence of the implied constants on s, Δ, ε in the proof, and we may assume that $\delta > x^{-1/C}$. We may also assume that $U \geq x^{\varepsilon/2}$, since otherwise $UV^2 \leq Mx^{-\varepsilon}$ implies $UV \leq M^{1/2}x^{-\varepsilon/4}$, in which case the claim follows from Proposition 6.3.

By replacing absolute values with an arbitrary unimodular sequence, the left-hand side of (6.3) can be written as

$$\sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v \leq 2V \\ (v, c'u) = 1}} \theta_{u,v} \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M \\ mn \equiv c_u \pmod{u} \\ mn \equiv c' \pmod{v}}} a_m b_n \psi(mn) := \Sigma \quad (6.5)$$

for some $|\theta_{u,v}| \leq 1$ and some c_u coprime to u . By assumption, we have

$$UV^2 \leq Mx^{-\varepsilon}, \quad x^{\varepsilon/2} \leq U \leq x^{1-\varepsilon}/M. \quad (6.6)$$

By the triangle inequality, we have

$$\Sigma \ll \sum_{U \leq u \leq 2U} \sum_{M \leq m \leq 2M} |a_m| \left| \sum_{\substack{V \leq v \leq 2V \\ (v, c'u) = 1}} \theta_{u,v} \sum_{\substack{x/m \leq n \leq 2x/m \\ mn \equiv c_u \pmod{u} \\ mn \equiv c' \pmod{v}}} b_n \psi(mn) \right|.$$

Let $N := x/M$. By applying the Cauchy–Schwarz inequality and expanding, we get

$$\Sigma^2 \ll MU \|a\|_2^2 \sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ (v_1, c'u) = (v_2, c'u) = 1}} \theta_{u, v_1} \overline{\theta_{u, v_2}} \sum_{N/2 \leq n_1, n_2 \leq 2N} b_{n_1} \overline{b_{n_2}} \sum_{\substack{x \leq mn_1, mn_2 \leq 2x \\ M \leq m \leq 2M \\ mn_i \equiv c_u \pmod{u} \\ mn_i \equiv c' \pmod{v_i}}} \psi(mn_1) \overline{\psi(mn_2)}.$$

Using the triangle inequality, the trivial inequality $|b_{n_1} b_{n_2}| \ll |b_{n_1}|^2 + |b_{n_2}|^2$, and symmetry, we have

$$\Sigma^2 \ll MU \|a\|_2^2 \sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ (v_1, c'u) = (v_2, c'u) = 1}} \sum_{\substack{N/2 \leq n_1, n_2 \leq 2N \\ n_1 \equiv n_2 \pmod{u(v_1, v_2)}}} |b_{n_1}|^2 \left| \sum_{\substack{x \leq mn_1, mn_2 \leq 2x \\ M \leq m \leq 2M \\ mn_i \equiv c_u \pmod{u} \\ mn_i \equiv c' \pmod{v_i}}} \psi(mn_1) \overline{\psi(mn_2)} \right|,$$

where we are free to add the constraint $n_1 \equiv n_2 \pmod{u(v_1, v_2)}$ since otherwise the inner sum over m is empty. Note that the inner sum over m is bounded trivially by $O(M/u[v_1, v_2]) = O(M(v_1, v_2)/UV^2)$, since $u[v_1, v_2] \ll UV^2 \leq Mx^{-\varepsilon}$ by (6.6). Separating out the terms with $n_1 = n_2$ and with $(v_1, v_2) > \delta^{-2}$, and applying the Cauchy–Schwarz inequality, we get

$$\Sigma^2 \ll MU \|a\|_2^2 \left(\Sigma_0 + \Sigma'_0 + \Sigma_1^{1/2} \Sigma_2^{1/2} \right),$$

where

$$\begin{aligned} \Sigma_0 &= \sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ (v_1, c'u) = (v_2, c'u) = 1}} \sum_{N/2 \leq n \leq 2N} |b_n|^2 \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M \\ mn \equiv c_u \pmod{u} \\ mn \equiv c' \pmod{v_i}}} |\psi(mn)|^2, \\ \Sigma'_0 &= \sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ (v_1, c'u) = (v_2, c'u) = 1 \\ (v_1, v_2) > \delta^{-2}}} \sum_{\substack{N/2 \leq n_1, n_2 \leq 2N \\ n_1 \equiv n_2 \pmod{u(v_1, v_2)} \\ n_1 \neq n_2}} |b_{n_1}|^2 \left| \sum_{\substack{x \leq mn_1, mn_2 \leq 2x \\ M \leq m \leq 2M \\ mn_i \equiv c_u \pmod{u} \\ mn_i \equiv c' \pmod{v_i}}} \psi(mn_1) \overline{\psi(mn_2)} \right| \\ \Sigma_1 &= \sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ (v_1, c'u) = (v_2, c'u) = 1 \\ (v_1, v_2) \leq \delta^{-2}}} \sum_{\substack{N/2 \leq n_1, n_2 \leq 2N \\ n_1 \equiv n_2 \pmod{u(v_1, v_2)} \\ n_1 \neq n_2}} |b_{n_1}|^4, \\ \Sigma_2 &= \sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ (v_1, c'u) = (v_2, c'u) = 1 \\ (v_1, v_2) \leq \delta^{-2}}} \sum_{\substack{N/2 \leq n_1, n_2 \leq 2N \\ n_1 \equiv n_2 \pmod{u(v_1, v_2)} \\ n_1 \neq n_2}} \left| \sum_{\substack{x \leq mn_1, mn_2 \leq 2x \\ M \leq m \leq 2M \\ mn_i \equiv c_u \pmod{u} \\ mn_i \equiv c' \pmod{v_i}}} \psi(mn_1) \overline{\psi(mn_2)} \right|^2. \end{aligned}$$

For Σ_0 , using the trivial bound for the inner sum over m we get

$$\Sigma_0 \ll \frac{M}{UV^2} \sum_{U \leq u \leq 2U} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ (v_1, c'u) = (v_2, c'u) = 1}} (v_1, v_2) \sum_{N/2 \leq n \leq 2N} |b_n|^2.$$

Since

$$\sum_{V \leq v_1, v_2 \leq 2V} (v_1, v_2) \leq \sum_{v_0 \leq 2V} \sum_{\substack{V \leq v_1, v_2 \leq 2V \\ v_0 | v_1, v_0 | v_2}} v_0 \ll \sum_{v_0 \leq 2V} \frac{V^2}{v_0} \ll V^2 \log x,$$

we have

$$\Sigma_0 \ll MN \|b\|_2^2 \log x = x \|b\|_2^2 \log x.$$

For Σ'_0 , using the trivial bound again for the inner sum over m , we see that the contribution to Σ'_0 from those terms with $(v_1, v_2) = v_0$ for a given $v_0 \leq 2V$ is

$$\ll U \left(\frac{V}{v_0} \right)^2 \left(\sum_{n_1} |b_{n_1}|^2 \right) \frac{N}{U v_0} \cdot \frac{M v_0}{U V^2} = \frac{N x}{U v_0^2} \|b\|_2^2.$$

Summing this over all $v_0 \geq \delta^{-2}$ gives

$$\Sigma'_0 \ll \delta^2 \frac{N x}{U} \|b\|_2^2.$$

For Σ_1 we clearly have

$$\Sigma_1 \ll U V^2 \left(\sum_n |b_n|^4 \right) \frac{N}{U} = V^2 N^2 \|b\|_4^4.$$

Thus it suffices to show that

$$\Sigma_2 \ll \frac{\delta^4 x^2}{U^2 V^2}. \quad (6.7)$$

Indeed, once (6.7) is established, we can combine it with our bounds for $\Sigma_0, \Sigma'_0, \Sigma_1$ to get

$$\begin{aligned} \Sigma^2 &\ll MU \|a\|_2^2 \left(x \|b\|_2^2 \log x + \delta^2 \frac{N x}{U} \|b\|_2^2 + V N \|b\|_4^2 \cdot \frac{\delta^2 x}{U V} \right) \\ &\ll (M U x + \delta^2 x^2) (\log x) \|a\|_2^2 \|b\|_4^2. \end{aligned}$$

Since $M U x \leq x^{2-\varepsilon} \leq \delta^2 x^2$ by (6.6) and our assumption on δ , this gives the desired bound for Σ .

Now we turn to bounding Σ_2 . Let Γ be the set of (n_1, n_2, u, v_1, v_2) with

$$\begin{aligned} U \leq u \leq 2U, \quad V \leq v_1, v_2 \leq 2V, \quad N/2 \leq n_1, n_2 \leq 2N, \quad n_1 \neq n_2, \\ (u, c) = (v_1, c') = (v_2, c') = 1, \quad (v_1, v_2) \leq \delta^{-2}, \quad n_1 \equiv n_2 \pmod{u(v_1, v_2)} \end{aligned}$$

such that

$$\left| \sum_{\substack{x \leq m n_1, m n_2 \leq 2x \\ M \leq m \leq 2M \\ m n_i \equiv c_u \pmod{u} \\ m n_i \equiv c' \pmod{v_i}}} \psi(m n_1) \overline{\psi(m n_2)} \right| \geq \frac{\delta^2 M}{U V^2}. \quad (6.8)$$

Suppose, for the purpose of contradiction, that (6.7) fails. Then we must have

$$|\Gamma| \geq \delta^{O(1)} V^2 N^2.$$

For $1 \leq \ell \leq \delta^{-2}$, let Γ_ℓ be the set of $(n_1, n_2, u, v_1, v_2) \in \Gamma$ such that $(v_1, v_2) = \ell$. By the pigeonhole principle we have

$$|\Gamma_\ell| \gg \delta^{O(1)} V^2 N^2 \quad (6.9)$$

for some ℓ . Fix such an ℓ .

Let $\psi = \varphi \circ g$. For $(n_1, n_2, u, v_1, v_2) \in \Gamma_\ell$, note that the congruence condition in the sum in (6.8) is equivalent to $m \equiv B \pmod{uv_1v_2/\ell}$ for some $B = B(n_1, n_2, u, v_1, v_2) \in [1, uv_1v_2/\ell]$. Let $h = h_{n_1, n_2, u, v_1, v_2} : \mathbb{Z} \rightarrow G \times G$ be the polynomial sequence defined by

$$h(n) := (g(uv_1v_2/\ell \cdot n_1n + n_1B), g(uv_1v_2/\ell \cdot n_2n + n_2B)),$$

and $\varphi \otimes \overline{\varphi} : G \times G \rightarrow \mathbb{C}$ be the function defined by

$$\varphi \otimes \overline{\varphi}(x_1, x_2) := \varphi(x_1) \overline{\varphi(x_2)}$$

for $x_1, x_2 \in G$. After a change of variables, (6.8) implies that

$$\left| \sum_{n \in I} (\varphi \otimes \overline{\varphi}) \circ h(n) \right| \gg \frac{\delta^2 M}{UV^2},$$

where $I = I_{u, v_1, v_2, n_1, n_2}$ is an interval contained in $[1, 2\ell M/UV^2]$ and $|I| \asymp \ell M/UV^2$. Since $\ell \leq \delta^{-2}$, we have

$$\frac{\delta^2 M}{UV^2} \gg \delta^{O(1)} |I|.$$

Hence the polynomial sequence $(h(n))_{n \leq 2\ell M/UV^2}$ fails to be totally $\delta^{O(1)}$ -equidistributed, and hence by Theorem 3.5, there is a nontrivial horizontal character $\chi = \chi(n_1, n_2, u, v_1, v_2)$ on $G \times G$ with $\|\chi\| \ll \delta^{-O(1)}$, such that

$$\|\chi \circ h\|_{C^\infty(2\ell M/UV^2)} \ll \delta^{-O(1)}. \quad (6.10)$$

After pigeonholing (which replaces Γ_ℓ by a subset of it, but the lower bound in (6.9) remains valid with potentially larger implied constants), we may assume that χ is independent of n_1, n_2, u, v_1, v_2 . We can write $\chi = (\chi_1, \chi_2)$, where χ_1, χ_2 are horizontal characters on G with $\|\chi_i\| \ll \delta^{-O(1)}$, at least one of which, say χ_1 , is nontrivial, so that $\chi(x_1, x_2) = \chi_1(x_1) + \chi_2(x_2)$ for $x_1, x_2 \in G$. If we write

$$\chi_1 \circ g(n) = \sum_{j=0}^s \alpha_j n^j, \quad \chi_2 \circ g(n) = \sum_{j=0}^s \beta_j n^j,$$

then we can write

$$\chi \circ h(n) = \sum_{j=0}^s \alpha_j (uv_1v_2/\ell \cdot n_1n + n_1B)^j + \sum_{j=0}^s \beta_j (uv_1v_2/\ell \cdot n_2n + n_2B)^j = \sum_{j=0}^s \gamma_j n^j$$

for some coefficients $\gamma_j = \gamma_j(n_1, n_2, u, v_1, v_2)$. In particular,

$$\gamma_s(n_1, n_2, u, v_1, v_2) = \alpha_s (uv_1v_2n_1/\ell)^s + \beta_s (uv_1v_2n_2/\ell)^s.$$

It follows from (6.10) that

$$\|\gamma_j(n_1, n_2, u, v_1, v_2)\| \ll \frac{\delta^{-O(1)}}{|I|^j} \quad (6.11)$$

for each $1 \leq j \leq s$ and $(n_1, n_2, u, v_1, v_2) \in \Gamma_\ell$. We are now left with the problem of deducing the Diophantine properties of the α_j 's and β_j 's from those of the γ_j 's, which will eventually contradict the equidistribution assumption on g .

Consider the 5-variable polynomial

$$Q(n_1, z, u, v'_1, v'_2) = \gamma_s(n_1, n_1 + u\ell z, u, \ell v'_1, \ell v'_2).$$

By (6.11), we have

$$\|Q(n_1, z, u, v'_1, v'_2)\| \ll \frac{\delta^{-O(1)}}{|I|^s} \ll \delta^{-O(1)} \left(\frac{UV^2}{M} \right)^s$$

for $\gg \delta^{O(1)} V^2 N^2$ values of (n_1, z, u, v'_1, v'_2) , with $n_1 \asymp N$, $z \asymp N/(\ell U)$, $u \asymp U$, $v'_1, v'_2 \asymp V/\ell$. Note that

- the sizes of n_1, z are both $\geq \delta^{O(1)} x^{\varepsilon/2} \geq \delta^{-O(1)}$ by (6.6);
- the term $(\alpha_s + \beta_s) \ell^s \cdot (uv'_1 v'_2 n_1)^s$ is the only one appearing in $Q(n_1, z, u, v'_1, v'_2)$ whose degrees in the n_1, z variables are $s, 0$, respectively;
- the term $\beta_s \ell^{2s} \cdot u^{2s} (v'_1 v'_2 z)^s$ is the only one appearing in $Q(n_1, z, u, v'_1, v'_2)$ whose degrees in the n_1, z variables are $0, s$, respectively.

Hence Lemma 6.2 is applicable to Q and implies that there exists $1 \leq q \ll \delta^{-O(1)}$ such that

$$\begin{aligned} \|q \ell^s (\alpha_s + \beta_s)\| &\ll \delta^{-O(1)} \left(\frac{UV^2}{M} \right)^s \frac{1}{(UV^2 N)^s} \ll \delta^{-O(1)} x^{-s}, \\ \|q \ell^{2s} \beta_s\| &\ll \delta^{-O(1)} \left(\frac{UV^2}{M} \right)^s \frac{1}{(UV^2 N)^s} \ll \delta^{-O(1)} x^{-s}. \end{aligned}$$

Hence $q_s = q \ell^{2s} \ll \delta^{-O(1)}$ has the property that

$$\|q_s \alpha_s\| \ll \delta^{-O(1)} x^{-j}, \quad \|q_s \beta_s\| \ll \delta^{-O(1)} x^{-j}.$$

We then show by induction that for $1 \leq j \leq s$ and some $1 \leq q_j \ll \delta^{-O(1)}$ with $q_{j+1} \mid q_j$, we have

$$\|q_j \alpha_j\| \ll \delta^{-O(1)} x^{-j}, \quad \|q_j \beta_j\| \ll \delta^{-O(1)} x^{-j}. \quad (6.12)$$

The case $j = s$ has been proved, so we may assume that all the cases $s \geq j' > j$ have been proved and that we are considering case j . The coefficient $\gamma_j(n_1, n_2, u, v_1, v_2)$ can be explicitly written as

$$\begin{aligned} \gamma_j &= \alpha_j(uv_1 v_2 n_1 / \ell)^j + \beta_j(uv_1 v_2 n_2 / \ell)^j + \\ &\quad \sum_{j < j' \leq s} \binom{j'}{j} \left(\alpha_{j'}(uv_1 v_2 n_1 / \ell)^j (n_1 B)^{j'-j} + \beta_{j'}(uv_1 v_2 n_2 / \ell)^j (n_2 B)^{j'-j} \right) \end{aligned}$$

Since (6.12) holds for all indices $j' > j$, one easily verifies that

$$\left\| q_{j'} \alpha_{j'} (uv_1 v_2 n_1 / \ell)^j (n_1 B)^{j'-j} \right\| \ll \frac{\delta^{-O(1)}}{|I|^{j'}}, \quad \left\| q_{j'} \beta_{j'} (uv_1 v_2 n_2 / \ell)^j (n_2 B)^{j'-j} \right\| \ll \frac{\delta^{-O(1)}}{|I|^{j'}}.$$

Defining $q_{j+1} = \prod_{j' < j \leq s} q_{j'}$, we see that the coefficient $\gamma_j(n_1, n_2, u, v_1, v_2)$ may be written in the form

$$\|q_{j+1}\gamma_j\| = \|q_{j+1}\alpha_j(uv_1v_2n_1/\ell)^j + q_{j+1}\beta_j(uv_1v_2n_2/\ell)^j\| + O(\delta^{-O(1)}|I|^{-j-1}).$$

Thus by (6.11) we have

$$\|q_{j+1}\alpha_j(uv_1v_2n_1/\ell)^j + q_{j+1}\beta_j(uv_1v_2n_2/\ell)^j\| \ll \frac{\delta^{-O(1)}}{|I|^j}$$

for $(n_1, n_2, u, v_1, v_2) \in \Gamma_\ell$. Precisely as in the case $j = s$ (with s in that argument now replaced by j), this implies (6.12) for j . Now, for $q = q_1 \ll \delta^{-O(1)}$ we have for every $1 \leq j \leq s$ that

$$\|q\alpha_j\| \ll \delta^{-O(1)}x^{-j}, \quad \|q\beta_j\| \ll \delta^{-O(1)}x^{-j}.$$

Finally, by Lemma 3.4 we have

$$\|q\chi_1 \circ g\|_{C^\infty(x)} \ll \delta^{-O(1)}.$$

Since $q\chi_1$ is nontrivial and $\|q\chi_1\| \ll \delta^{-O(1)}$, Lemma 3.6 implies that $\{g(n)\}_{1 \leq n \leq x}$ is not totally $\delta^{O(1)}$ -equidistributed, a contradiction if C is chosen large enough. $\square$

Proposition 6.7 (type II estimate with a fixed nilsequence). *Let $x \geq 2$, let $\varepsilon > 0$ be a small constant, let $M \in [x^{1/2}, x^{2/3+\varepsilon/2}]$. Let $s \geq 1$, $\Delta \geq 2$, $0 < \delta < 1/2$. Then for a large constant $C = C(s, \Delta, \varepsilon)$ and $\psi \in \Psi_s(\Delta, \delta^{-1}; \delta^C, x)$, we have*

$$\sum_{d \leq x^{1/3-\varepsilon}} \max_{(c,d)=1} \left| \sum_{\substack{x \leq mn \leq 2x \\ M \leq m \leq 2M \\ mn \equiv c \pmod{d}}} a_m b_n \psi(mn) \right| \ll \delta x \|a\|_2 \|b\|_4 (\log x)^3, \quad (6.13)$$

for any sequences $\{a_m\}, \{b_n\}$.

Proof. This follows directly from Proposition 6.6 (with ε replaced by $\varepsilon/3$ there), taking $U \leq x^{1/3-\varepsilon}$, $V = 1$ there. $\square$

7 COMBINING THE TYPE I AND TYPE II ESTIMATES

The type I and II estimates in the previous sections have been tailored so that we will be able to shortly conclude the proofs of Theorems 4.2, 4.3, 4.4 by appealing to Vaughan's identity, stated here for convenience.

Lemma 7.1 (Vaughan's identity). *Let $x \geq 2$ and $x^{2/3} < n \leq x$. Then we can write*

$$\Lambda(n) = \sum_{j \leq J} (\alpha_j * \delta_j(n) + \beta_j * \gamma_j(n)),$$

where

$$(i) \ J \ll (\log x)^{10};$$

- (ii) $|\alpha_j(n)|, |\beta_j(n)|, |\gamma_j(n)| \ll d_3(n)(\log x)$ for each $j \leq J$, and $\alpha_j, \beta_j, \gamma_j$ are each supported inside dyadic intervals;
- (iii) $\text{supp}(\alpha_j) \subset [1, 2x^{1/3}]$ and $\text{supp}(\beta_j), \text{supp}(\gamma_j) \subset [x^{1/3}, x^{2/3}]$ for each $j \leq J$.
- (iv) $\delta_j(n) = 1$ or $\delta_j(n) = \log n$.

Proof. This follows from the formulation of Vaughan's identity in [25, Proposition 13.4], taking $y = z = x^{1/3}$ there and splitting the variables into dyadic intervals. $\square$

Proof of Theorems 4.2, 4.3, 4.4. The bound $\ll \eta^\kappa x (\log x)^2$ in the statements of the theorems follows trivially from the Brun–Titchmarsh inequality if $\eta^\kappa > 1/\log x$, so we may assume that $\eta \leq (\log x)^{-1/\kappa}$ for small enough $\kappa > 0$. Then it suffices to prove a bound of the form $\ll \eta^\kappa x (\log x)^{O(1)}$ for the sums in Theorems 4.2, 4.3, 4.4.

Let $\varepsilon > 0$ be small. Note that by the triangle inequality and crude estimation, it suffices to prove each proposition with the n sum ranging over any dyadic interval $[x', 2x'] \subset [x^{1-\varepsilon/10}, x]$ instead of $[1, x]$. Now, by Vaughan's identity, for $n \in [x', 2x']$ we can write $\Lambda(n)$ as a sum of $\ll (\log x')^{10}$ type I convolutions $\alpha * \delta(n)$ with $\delta \in \{1, \log\}$ and type II convolutions $\beta * \gamma(n)$, where α, β, γ satisfy condition (2) of Lemma 7.1 and $\text{supp}(\alpha) \subset [1, (x')^{1/3+\varepsilon}]$, $\text{supp}(\beta), \text{supp}(\gamma) \subset [(x')^{1/3}, (x')^{2/3+\varepsilon}]$. By the triangle inequality, it suffices to prove each of Theorems 4.2, 4.3, 4.4 with $\Lambda(n)$ replaced by one of these type I or type II convolutions.

For the type I convolutions, we can readily apply Proposition 5.2 in the case of Theorem 4.2 and Proposition 5.4 in the case of Propositions 4.3, 4.4 (possibly applying partial summation first to get rid of the $\log n$ factor in $\alpha * \log(n)$). For the type II convolutions, in turn, we can appeal to Proposition 6.3 in the case of Theorem 4.2, Proposition 6.7 in the case of Theorem 4.3, and Proposition 6.5 in the case of Theorem 4.4 (taking the sequence a_m in these propositions to be the one of β_m, γ_m whose support is contained in $[(x')^{1/2}, x']$). This completes the proof. $\square$

In view of the reduction in Section 4 to the equidistributed case, our main Theorems 1.3, 1.4 and 1.6 follow from Theorems 4.2, 4.3, 4.4, together with Propositions 4.7, 4.8. Hence, to complete the proofs of our main theorems, it now suffices to prove these two propositions.

Proof of Propositions 4.7 and 4.8. We prove Proposition 4.7 by replacing the condition $(n, W) = 1$ by a standard sieve weight, and then using our type I estimate (Proposition 5.2). The proof of Proposition 4.8 is completely analogous, using Proposition 5.4 instead of Proposition 5.2.

Let $D = x^{0.01}$, and let $(\lambda_\ell^\pm)_{\ell \leq D}$ be upper and lower linear sieve weights, defined in [25, Section 6.2]. Note that if $\psi \in \Psi_s(\Delta, \eta^{-\kappa}; \eta, x/d)$ then its real and imaginary parts lie in $\Psi_s(\Delta, \eta^{-\kappa}; \eta, x/d)$ as well. Thus, by splitting ψ into its real and imaginary parts we may restrict attention to real-valued ψ .

Apply the upper bound sieve to the non-negative function $\psi + 1$ to obtain

$$\begin{aligned} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ (n, W) = 1}} (\psi((n-c)/d) + 1) &\leq \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \left(\sum_{\ell | (n, W)} \lambda_\ell^+ \right) (\psi((n-c)/d) + 1) \\ &= \sum_{\ell | W} \lambda_\ell^+ \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ \ell | n}} [\psi((n-c)/d) + 1]. \end{aligned}$$

Apply the lower bound sieve to the constant function 1 to obtain

$$\sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ (n, W) = 1}} 1 \geq \sum_{\ell | W} \lambda_{\ell}^{-} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ \ell | n}} 1.$$

Using the fundamental lemma of sieve theory [25, Lemma 6.3] (with $g(\ell) = \frac{1}{\ell} 1_{(\ell, d) = 1}$ and $s = (\log D)/(\log w)$ there), we get

$$\sum_{\ell | W} \lambda_{\ell}^{\pm} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ \ell | n}} 1 = \frac{x}{d} \prod_{\substack{p | W \\ p \nmid d}} \left(1 - \frac{1}{p}\right) + O_A \left(\frac{x}{d(\log x)^{2A}} \right).$$

It follows that

$$\sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ (n, W) = 1}} \psi((n - c)/d) \leq \sum_{\ell | W} \lambda_{\ell}^{+} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ \ell | n}} \psi((n - c)/d) + O_A \left(\frac{x}{d(\log x)^{2A}} \right).$$

By applying the inequality above with ψ replaced by $-\psi$, we deduce that

$$\left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ (n, W) = 1}} \psi((n - c)/d) \right| \leq \left| \sum_{\ell | W} \lambda_{\ell}^{+} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ \ell | n}} \psi((n - c)/d) \right| + O_A \left(\frac{x}{d(\log x)^{2A}} \right).$$

Thus it suffices to show that

$$\sum_{d \leq x^{1/2-\varepsilon}} \max_{(c, d) = 1} \sup_{\psi \in \Psi_s(\Delta, \eta^{-\kappa}; \eta, x/d)} \left| \sum_{\ell | W} \lambda_{\ell}^{+} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d} \\ \ell | n}} \psi((n - c)/d) \right| \ll \eta^{\kappa} x \log x.$$

By a change of variables $n = \ell m$, we can rewrite the expression inside the absolute value sign in the form

$$\sum_{\substack{\ell m \leq x \\ \ell m \equiv c \pmod{d}}} a_{\ell} \psi((\ell m - c)/d),$$

where $a_{\ell} = \lambda_{\ell}^{+} 1_{\ell | W}$. Since (λ_{ℓ}^{+}) (and thus (a_{ℓ})) is supported on $\ell \leq D = x^{0.01}$, the desired estimate follows from our type I estimate (Proposition 5.2) by taking $\kappa = C^{-1}$, where $C = C(s, \Delta, \varepsilon)$ is the constant from Proposition 5.2. $\square$

The proof of Theorem 4.6 proceeds similarly to the proofs of Theorems 4.2, 4.3, 4.4, except that we need an analogue of Heath-Brown's identity for the multiplicative functions $d_k(n)$ and $1_S(n)$. This follows from the following general lemma on decompositions of multiplicative functions.

Lemma 7.2 (A Heath-Brown type decomposition for multiplicative functions). *Let $\kappa, Q, n_0, A \geq 1$ and $\varepsilon > 0$ be fixed. Let $f : \mathbb{N} \rightarrow \mathbb{C}$ be a multiplicative function satisfying $|f(n)| \leq d_{\kappa}(n)$ and such that the sequence $(f(p))_{p \geq n_0}$ is periodic with period $Q \in \mathbb{N}$. Then for $x \geq 2$ and $1 \leq n \leq x$ we can write*

$$f(n) = \sum_{j \leq J} \alpha_{1,j} * \cdots * \alpha_{R,j}(n) + O(E(n)),$$

where

- (i) $J \ll (\log x)^{O(1)}$ and $R \ll 1$;
- (ii) All the $\alpha_{i,j}(n)$ are $\ll d(n)^{O(1)}$ in absolute value;
- (iii) All the $\alpha_{i,j}(n)$ are supported on intervals of the form $I_{i,j} = [y, y']$ with $y < y' \leq 2y$, and if $y > x^\varepsilon$, then we have $\alpha_{i,j}(n) = \chi(n)1_{[y,y']}(n)$ or $\alpha_i(n) = \chi(n)(\log n)1_{[y,y']}(n)$, where χ is a Dirichlet character (depending on i, j) whose modulus divides Q ;
- (iv) We have

$$E(n) \ll ((\log x)^{O(1)}(1_{\exists p > w: p^2 | n} + 1_{\exists m > x^\varepsilon: m | n, m \text{ is } w\text{-smooth}}) + (\log x)^{-A})d_\kappa(n),$$

where $w = x^{1/(\lceil 3/\varepsilon \rceil (\log \log x))}$.

Proof. This is a minor modification of arguments in [28, Section 5]. For the sake of completeness, we provide the details.

We may assume that x is large enough. In what follows, all constants in $O(\cdot)$ notation may depend on the fixed quantities $\kappa, Q, n_0, \varepsilon, A$. Let $C = \lceil 3/\varepsilon \rceil$, $K = C \log \log x$, and $w = x^{1/K}$. Then we may factorize

$$f(n) = \sum_{\substack{0 \leq k \leq K \\ 0 \leq \ell \leq C}} \frac{1}{k! \ell!} \sum_{\substack{n = mp_1 \cdots p_k q_1 \cdots q_\ell \\ w < p_i \leq x^{1/C} \\ q_i > x^{1/C} \\ p' | m \implies p' \leq w}} f(p_1) \cdots f(p_k) f(q_1) \cdots f(q_\ell) f(m) + O(d_\kappa(n) 1_{\exists p > w: p^2 | n}),$$

where the error term arises from those n that have are divisible by the square of some prime $> w$. Next, by the orthogonality of characters, for any prime $q > \max\{n_0, Q\}$ we can write

$$f(q) = \frac{1}{\varphi(Q)} \sum_{\chi \pmod{Q}} \sum_{1 \leq b \leq Q} a_b \overline{\chi(b)} \chi(p), \quad (7.1)$$

where $a_b \in \mathbb{C}$ is such that $f(p) = a_b$ for all primes $p \equiv b \pmod{Q}$, $p \geq n_0$. Applying (7.1) to each of $f(q_i)$, we can decompose $f(n)$ as a linear combination (with $O(1)$ coefficients) of $O(1)$ sums of the form

$$\sum_{\substack{0 \leq k \leq K \\ 0 \leq \ell \leq C}} \frac{1}{k! \ell!} \sum_{\substack{n = mp_1 \cdots p_k q_1 \cdots q_\ell \\ w < p_i \leq x^{1/C} \\ q_i > x^{1/C} \\ p' | m \implies p' \leq w}} f(p_1) \cdots f(p_k) \chi_1(q_1) \cdots \chi_\ell(q_\ell) f(m) + O(d_\kappa(n) 1_{\exists p > w: p^2 | n})$$

where χ_i are characters to modulus Q . We split each q_i into short intervals of the form $[x(1 - 1/(\log x)^{A+1})^{j+1}, x(1 - 1/(\log x)^A)^j]$, where $j \geq 0$. Then $f(n)$ becomes a linear combination of $O(1)$ sums of the form

$$\sum_{I_1, \dots, I_\ell \in \mathcal{J}} \sum_{\substack{0 \leq k \leq K \\ 0 \leq \ell \leq C}} \frac{1}{k! \ell!} \sum_{\substack{n = mp_1 \cdots p_k q_1 \cdots q_\ell \\ w < p_i \leq x^{1/C} \\ q_i \in I_i \\ p' | m \implies p' \leq w}} f(p_1) \cdots f(p_k) \chi_1(q_1) \cdots \chi_\ell(q_\ell) f(m) + O(d_\kappa(n) 1_{\exists p > w: p^2 | n}),$$

where $\mathcal{J}$ is the collection of intervals of the form $I = [x(1 - 1/(\log x)^{A+1})^{j+1}, x(1 - 1/(\log x)^{A+1})^j]$, with $j \geq 0$ and $I \subset [x^{1/C}, x]$.

We then add the von Mangoldt weight to each of the q_i variables by writing, for $n \in I = [y, y(1 + (\log x)^{-A-1})]$,

$$1_{\mathbb{P}}(n) = \frac{\Lambda(n)}{\log n} + O(1_{n \text{ not squarefree}}) = \frac{\Lambda(n)}{\log y} + O((\log x)^{-A}) + O(1_{n \text{ not squarefree}}).$$

Estimating the contribution of the two error terms trivially by the triangle inequality, the result is that $f(n)$ can be written as a linear combination (with $O(1)$ coefficients) of $O((\log x)^{(A+2)C})$ sums of the form

$$\sum_{\substack{0 \leq k \leq K \\ 0 \leq l \leq C}} \frac{1}{k!l!} \sum_{\substack{n = mp_1 \cdots p_k m_1 \cdots m_l \\ w < p_i \leq x^{1/C} \\ q_i \in I_i \\ p' | m \implies p' \leq w \\ m \leq x^\varepsilon}} f(p_1) \cdots f(p_k) f(m) \prod_{j=1}^l \Lambda(m_j) \chi_j(m_j),$$

plus an error term

$$E(n) \ll d_\kappa(n) ((\log x)^{O(1)} 1_{\exists p > w: p^2 | n} + (\log x)^{O(1)} 1_{\exists m > x^\varepsilon: m | n, m \text{ is } w\text{-smooth}}) + (\log x)^{-A},$$

where the second summand in the upper bound for $E(n)$ arose from the restriction to $m \leq x^\varepsilon$. For the summands with $k \ll 1$, we could now use Heath–Brown’s identity ([25, Proposition 13.3] with $K = \lceil 1/\varepsilon \rceil$ there) to each of $\Lambda(m_i)$ to obtain a decomposition of the desired form. To make the argument work for large k , it suffices to show that

$$g(n) := \sum_{0 \leq k \leq K} \frac{1}{k!} \sum_{\substack{n = p_1 \cdots p_k \\ w < p_i \leq x^{1/C}}} f(p_1) \cdots f(p_k)$$

can be written as a linear combination of $O((\log x)^{O(1)})$ convolutions of the form $\beta_{1,i} * \cdots * \beta_{M,i}$ with $M \ll 1$ and $|\beta_{i,j}| \ll (\log x)^{O(1)}$ and with each $\beta_{i,j}$ supported in $[x^{1/C}, x^{3/C}]$. But this follows from a simple grouping of $p_1 \cdots p_k$ into subproducts of size $\in [x^{1/C}, x^{3/C}]$, detailed in [28, Section 5]. $\square$

We are now in a position to also conclude the proof of Theorem 4.6 that involves the functions d_k and 1_S .

Proof of Theorem 4.6. Let $\varepsilon > 0$ be small. As previously, it suffices to prove a dyadic version of Theorem 4.6, where we sum over $n \in [x', 2x'] \subset [x^{1-\varepsilon/10}, x]$ instead of $n \leq x$. Let $f(n) = d_k(n)$ with $k \in \mathbb{C}$ fixed, or $f(n) = 1_S(n)$. By Lemma 7.2 (with $Q = 4$ and $\kappa = |k| + 1$), we can decompose f as a sum of $\ll (\log x)^{O(1)}$ convolutions each of which is of the form $\alpha_1 * \cdots * \alpha_R(n)$ with α_i as in Lemma 7.2, plus an error term $E(n)$ that is of the form given in the lemma.

We will show that each of the convolutions $\alpha_1 * \cdots * \alpha_R(n)$ can be written as either a type I convolution $\alpha * \delta(n)$ or a type II convolution $\beta * \gamma(n)$, where

- (i) $|\alpha(n)| \ll (\log x)^{O(1)} d(n)^{O(1)}$ and $\text{supp}(\alpha) \subset [1, C(x')^{1/3}]$ for some $C \ll 1$ and $\delta(n) = \chi(n)$ or $\delta(n) = \chi(n) \log n$ with χ of modulus dividing 4.

(ii) $|\beta(n)|, |\gamma(n)| \ll (\log x)^{O(1)} d(n)^{O(1)}$ and $\text{supp}(\beta) \subset [(x')^{1/3}, (x')^{2/3}]$;

Let α_i be supported inside $[y_i, 2y_i]$. Since n is supported on $[x', 2x']$, we must have $y_1 \cdots y_R \asymp x'$. If $y_i \geq (x')^{2/3}$ for some i , then we can form a type I convolution by taking $\delta = \alpha_i$, $\alpha = \alpha_1 * \cdots * \alpha_{i-1} * \alpha_{i+1} * \cdots * \alpha_R$ (or $\alpha(n) = 1_{n=1}$ if $R = 1$), and now we have written $\alpha_1 * \cdots * \alpha_R$ in the form (1) above. If instead $y_i \in [(x')^{1/3}, (x')^{2/3}]$ for some i , then we form a type II convolution of the form (2) above by defining $\beta = \alpha_i$, $\gamma = \alpha_1 * \cdots * \alpha_{i-1} * \alpha_{i+1} * \cdots * \alpha_R$. If all $y_i \leq (x')^{1/3}$, then choose j to be the smallest index with $y_1 \cdots y_j \geq (x')^{1/3}$. We also must have $y_1 \cdots y_j \leq (x')^{2/3}$ since $y_j \leq (x')^{1/3}$. Therefore, taking $\beta = \alpha_1 * \cdots * \alpha_j$ and $\gamma = \alpha_{j+1} * \cdots * \alpha_R$ we have formed the desired type II convolution.

Now that each $\alpha_1 * \cdots * \alpha_R$ has been written as a type I or II convolution satisfying (1) or (2) above, we can apply the same argument that was used to conclude the proofs of Theorems 4.2, 4.3, 4.4 (noting that the $\chi(n)$ weight is harmless in our type I Bombieri–Vinogradov estimates by splitting n into progressions $(\bmod 4)$) to see that the analogue of each of those theorems holds with $\alpha_1 * \cdots * \alpha_R$ in place of $\Lambda(n)$.

To conclude the proof of Theorem 4.6, it now suffices to show that

$$\sum_{d \leq x^{1/2-\varepsilon}} \max_{(c,d)=1} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} |E(n)| \ll_B x / (\log x)^B, \quad (7.2)$$

To prove (7.2), we first note that by Cauchy–Schwarz and Shiu’s bound [33], for any $d \leq x^{1/2-\varepsilon}$ and c coprime to d we have

$$\begin{aligned} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}, \\ \exists w < p \leq x^{1/2}: p^2 | n}} d_{|k|+1}(n) &\ll \left(\sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} d_{|k|+1}(n)^2 \right)^{1/2} \left(\sum_{\substack{n \leq x \\ n \equiv c \pmod{d}, \\ \exists w < p \leq x^{1/2}: p^2 | n}} 1 \right)^{1/2} \\ &\ll (\log x)^{(|k|+1)^2/2} \left(\frac{x}{dw^{1/2}} + \frac{x^{3/4}}{d^{1/2}} \right). \end{aligned}$$

Hence, summing over $d \leq x^{1/2-\varepsilon}$, the term $(\log x)^{O(1)} 1_{\exists p > w: p^2 | n} d_{|k|+1}(n)$ in the definition of $E(n)$ gives an admissible contribution on the left of (7.2). We also have by Shiu’s bound, for any $d \leq x^{1/2-\varepsilon}$ and c coprime to d , that

$$\sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} d_{|k|+1}(n) \ll \frac{x(\log x)^{|k|}}{d},$$

so the contribution of the term $(\log x)^{-A} d_{|k|+1}(n)$ in the definition of $E(n)$ is also acceptable, provided that A is chosen large in terms of B .

Lastly, we deal with the term $(\log x)^{O(1)} 1_{\exists m > x^\varepsilon: m|n, m \text{ is } w\text{-smooth}} \cdot d_{|k|+1}(n)$ in the definition of $E(n)$. Let $\rho_w(n)$ denote the indicator function of w -smooth numbers. It suffices to show that

$$\sum_{d \leq x^{1/2-\varepsilon}} \max_{(c,d)=1} \sum_{\substack{n_1 n_2 \leq x \\ n_1 n_2 \equiv c \pmod{d} \\ n_1 > x^\varepsilon}} d_{|k|+1}(n_1) \rho_w(n_1) d_{|k|+1}(n_2) \ll_B x / (\log x)^B. \quad (7.3)$$

By a bilinear Bombieri–Vinogradov estimate [25, Theorem 17.4] (with $\Delta = 1$ there), for any $x^\varepsilon \leq N \leq x$ we have

$$\begin{aligned} & \sum_{d \leq x^{1/2-\varepsilon}} \max_{(c,d)=1} \sum_{\substack{n \leq x \\ n \equiv c \pmod{d}}} \sum_{\substack{n=n_1 n_2 \\ N \leq n_1 \leq 2N}} \rho_w(n_1) d_{|k|+1}(n_1) d_{|k|+1}(n_2) \\ & \ll \left(\sum_{N \leq n_1 \leq 2N} \rho_w(n_1)^2 d_{|k|+1}(n_1)^2 \right)^{1/2} \left(\sum_{n_2 \leq x/N} d_{|k|+1}(n_2)^2 \right)^{1/2}. \end{aligned}$$

Now, applying Cauchy–Schwarz to separate $d_{|k|+1}(n_1)^2$ and $\rho_w(n_1)^2$ and recalling that $w = x^{1/(\lceil 3/\varepsilon \rceil (\varepsilon \log \log x))}$ and applying a standard upper bound for smooth numbers [23, Theorem 1.1], this is bounded by $\ll x(\log x)^{O(1)} \exp(-(\log \log x)(\log \log \log x)) \ll_B x/(\log x)^B$.

Summing dyadically over N , we conclude that (7.3) holds, so (7.2) follows. $\square$

8 SIEVE LEMMAS

The rest of the paper is devoted to the proofs of our applications (Theorems 2.3, 2.5 and 2.7). We first formulate weighted versions of the sieves of Maynard and Chen, which will be needed subsequently.

Definition 8.1. We say that a k -tuple $(h_1, \dots, h_k)$ of integers is *admissible* if for every prime p there exists $a \in \mathbb{Z}$ such that $p \nmid a + h_i$ for all $1 \leq i \leq k$.

Proposition 8.2 (Maynard’s sieve). *For any $\theta \in (0, 1)$, $C_0 \geq 1$ and $k \in \mathbb{N}$, there exist $C = C(\theta)$ and $\sigma = \sigma(\theta, k)$ such that the following holds.*

Let $(\omega_n)_{n \leq x}$ be any sequence of nonnegative real numbers with $x \geq x_0$, and let $(L_1, \dots, L_k)$ be an admissible k -tuple of linear forms with $L_i(n) = a_i n + b_i$ and $1 \leq a_i, |b_i| \leq (\log x)^{1/100}$. Suppose that (ω_n) satisfies the following hypotheses:

(i) (Prime number theorem) *For all $1 \leq i \leq k$ and some $\delta > 0$ we have*

$$\frac{1}{k} \sum_{i=1}^k \frac{\varphi(a_i)}{a_i} \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \omega_n \geq \frac{\delta}{\log x} \sum_{n \leq x} \omega_n;$$

(ii) (Well-distribution in arithmetic progressions) *We have*

$$\sum_{r \leq x^\theta} \max_{c \pmod{r}} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{r}}} \omega_n - \frac{1}{r} \sum_{n \leq x} \omega_n \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{101k^2}}.$$

(iii) (Bombieri–Vinogradov) *For all $1 \leq i \leq k$ we have*

$$\sum_{r \leq x^\theta} \max_{(L_i(c), r)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{r} \\ L_i(n) \in \mathbb{P}}} \omega_n - \frac{1}{\varphi_{L_i}(r)} \sum_{\substack{n \leq x \\ L_i(n) \in \mathbb{P}}} \omega_n \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{101k^2}},$$

where for $L = \ell n + b$ we define $\varphi_L(n) := \varphi(|\ell|n)/\varphi(|\ell|)$.

(iv) (Brun–Titchmarsh) We have

$$\max_{\substack{c \pmod{r} \\ n \equiv c \pmod{r}}} \sum_{\substack{n \leq x \\ n \equiv c \pmod{r}}} \omega_n \leq \frac{C_0}{r} \sum_{n \leq x} \omega_n,$$

for all $1 \leq r \leq x^\theta$.

Then, for $x \gg_{k, \theta, C_0} 1$ we have

$$\sum_{\substack{n \leq x \\ |[L_1(n), \dots, L_k(n)] \cap \mathbb{P}| \geq C^{-1} \delta \log k \\ p | \prod_{i=1}^k L_i(n) \implies p > x^\sigma}} \omega_n \gg_{k, \theta, \delta} \frac{\mathfrak{S}(L_1, \dots, L_k)}{(\log x)^k} \sum_{n \leq x} \omega_n,$$

where the singular series $\mathfrak{S}(L_1, \dots, L_k)$ is given by

$$\mathfrak{S}(L_1, \dots, L_k) := \prod_p \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{|\{n \in \mathbb{Z}/p\mathbb{Z} : L_1(n) \cdots L_k(n) \equiv 0 \pmod{p}\}|}{p}\right) > 0.$$

Proof. This is [27, Theorem 6.2], which adds weights to the corresponding statement in [30] (in [27, Theorem 6.2] the result is stated for dyadic sums, but this clearly makes no difference in the proof). $\square$

Proposition 8.3 (Chen’s sieve). *Let $(\omega_n)_{n \leq x}$ be any sequence of nonnegative real numbers, and let x be large enough. Let $\varepsilon > 0$ be a small enough absolute constant. Suppose that (ω_n) satisfies the following hypotheses:*

(i) (Bombieri–Vinogradov with well-factorable weights) We have

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r,2)=1}} \lambda_r \left(\sum_{\substack{n \leq x \\ n+2 \equiv 0 \pmod{r} \\ n \in \mathbb{P}}} \omega_n - \frac{1}{\phi(r)} \sum_{n \leq x} \frac{\omega_n}{\log(n+1)} \right) \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{10}}$$

for any λ_r that is either well-factorable of level $x^{1/2-\varepsilon}$ or a convolution of the form $\lambda = 1_{p \in [P, P']} * \lambda'$ with λ' well-factorable of level $x^{1/2-\varepsilon}/P$ and $2P \geq P' \geq P \in [x^{1/10}, x^{1/3-\varepsilon}]$.

(ii) (Bombieri–Vinogradov for almost primes with well-factorable weights) For $j \in \{1, 2\}$ we have

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r,2)=1}} \lambda_r \left(\sum_{\substack{n \leq x \\ n \equiv 0 \pmod{r} \\ n+2 \in B_j}} \omega_n - \frac{1}{\phi(r)} \sum_{\substack{n \leq x \\ n+2 \in B_j}} \omega_n \right) \right| \ll \frac{\sum_{n \leq x} \omega_n}{(\log x)^{10}},$$

where λ_r is any well-factorable sequence of level $x^{1/2-\varepsilon}$ and

$$B_1 = \{p_1 p_2 p_3 : x^{1/10} \leq p_1 \leq x^{1/3-\varepsilon}, x^{1/3-\varepsilon} \leq p_2 \leq (2x/p_1)^{1/2}, p_3 \geq x^{1/10}\},$$

$$B_2 = \{p_1 p_2 p_3 : x^{1/3-\varepsilon} \leq p_1 \leq p_2 \leq (2x/p_1)^{1/2}, p_3 \geq x^{1/10}\}.$$

(iii) $((\omega_n)$ is not concentrated on almost primes): For $j \in \{1, 2\}$ we have

$$\sum_{\substack{n \leq x \\ n \in B_j}} \omega_n \leq (1 + \varepsilon) |B_j \cap [1, x]| \cdot \frac{1}{x} \sum_{n \leq x} \omega_n.$$

Then we have

$$\sum_{\substack{n \leq x \\ n \in \mathbb{P} \\ n+2 \in P_2 \\ p|n+2 \Rightarrow p \geq x^{1/10}}} \omega_n \geq \frac{\delta_0}{(\log x)^2} \sum_{n \leq x} \omega_n - O(x^{0.9} \max_n \omega_n),$$

for some absolute constant $\delta_0 > 0$.

Proof. This is essentially [27, Theorem 6.4], which adds weights to Chen's sieve argument. Note that since the sieve weights appearing in the proof of the weighted Chen's sieve in [27, Appendix A] are supported on squarefree numbers, the quantity $\mu(r)^2 \lambda_r$ in [27, Hypothesis 6.3] can be replaced with just λ_r . Note also that the $1 + \varepsilon$ factor in Hypothesis (iii) was $1 + o(1)$ there, but inspecting the proof in [27, Appendix A] (in particular Subsection A.5), this relaxation makes no difference. $\square$

9 BOUNDED GAPS BETWEEN PRIMES IN NIL-BOHR SETS

In this section we shall prove Theorem 2.3. We begin with Weyl's exponential sum estimate and an analogous estimate over the primes, formulated in a way that is convenient for us.

Lemma 9.1 (Weyl sum bounds over the integers and primes). *Let $s \geq 1$ be fixed. There exists $C_s > 0$ such that the following holds. Let $Q(x) = \alpha_s x^s + \dots + \alpha_1 x + \alpha_0$ with $\alpha_0, \dots, \alpha_s \in \mathbb{R}$. Let $W \geq 1$ be fixed. Lastly, let $x \geq 2$, $\delta \in (0, 1)$ and $1 \leq h \leq W$ be such that*

$$\left| \sum_{n \leq x} e(Q(n)) \right| > \delta x \quad \text{or} \quad \left| \sum_{\substack{n \leq x \\ Wn+h \in \mathbb{P}}} e(Q(n)) \right| > \delta \frac{W}{\phi(W)} \frac{x}{\log x}.$$

Then there exists $1 \leq \ell \ll \delta^{-C_s}$ such that $\|\ell \alpha_j\| \ll \delta^{-C_s} (\log x)^{C_s} / x^j$ for all $1 \leq j \leq s$.

Proof. In the case of the Weyl sum over the integers, this follows from [19, Proposition 9.3]. In the case of the exponential sum over the primes, in turn, we may combine Theorem 4.2 (taking $\psi(n) = Q((n-h)/W)$ there and considering only the $d = W$ term in the sum) with the quantitative Leibman theorem (Theorem 3.5) and Lemma 3.4 to obtain the result. $\square$

We then state a lemma that provides a convenient minorant function for $1_{\|x\| < \rho}$.

Lemma 9.2 (Vinogradov). *For any $\rho \in (0, 1/4)$ and $\eta \in (0, \rho/2)$, there exists a function $g = g_{\rho, \eta}$ such that we have the minorant property $0 \leq g(x) \leq 1_{\|x\| < \rho}$ and such that:*

- (i) $g(x)$ is 1-periodic and $g(x) = 0$ for $x \in [-1/2, 1/2] \setminus (-\rho, \rho)$.

(ii) We have a Fourier expansion

$$g(x) = 2(1 - \eta)\rho + \sum_{|j| > 0} c_j e(jx),$$

where $|c_j| \leq 10\rho$.

(iii) We have

$$\sum_{|j| > K} |c_j| \leq \frac{10\eta^{-1}}{K}$$

for any $K \geq 1$.

Proof. This follows from [38, Lemma 12]. $\square$

Let $k \in \mathbb{N}$ be any fixed natural number. For the rest of this section, we fix the choice $\eta = 1/(10000k^2)$ in the construction of the minorant function in Lemma 9.2, that is, we work with the function

$$g = g_{\rho, 1/(10000k^2)}.$$

Before embarking on the main proof, we show that if $(h_1, \dots, h_k)$ is a suitably chosen tuple, then the sequence $g(Q(n+h_1)) \cdots g(Q(n+h_k))$ has nearly the expected mean value $(2\rho)^k$ in the following quantitative sense.

Lemma 9.3 (Correlations of $g(Q(n))$). *Let $k \geq 1$ and $\rho \in (0, 1/4)$ be fixed. Let $g = g_{\rho, 1/(10000k^2)}$ be the function constructed above. Also let $Q(y) \in \mathbb{R}[y]$ be a fixed non-constant polynomial with irrational leading coefficient, and let $\mathcal{F} : \mathbb{N} \rightarrow \mathbb{N}$ be any fixed function. Then there exists $H \geq \mathcal{F}(k)$ (depending on all the above quantities) and a k -tuple $(h_1, \dots, h_k) \in [1, H]^k$ of distinct integers such that*

$$\sum_{n \leq x} g(Q(n+h_1)) \cdots g(Q(n+h_k)) \geq 0.99(2\rho)^k x \quad (9.1)$$

for $x \geq x_0$ large enough.

Proof. We may assume that $\mathcal{F}$ grows fast enough. Write $Q(y) = \alpha_s y^s + \cdots + \alpha_1 y + \alpha_0$ with $\alpha_s \in \mathbb{R} \setminus \mathbb{Q}$. By Dirichlet's approximation theorem, we can find infinitely many pairs (a', q') of coprime integers such that $|\alpha_s - a'/q'| \leq 1/(q')^2$. Choose a pair for which $q' \geq \mathcal{F}(k + \lceil \rho^{-1} \rceil)$. We take $H = (q')^2$. Note that we have the hierarchy $x \gg H \gg k + \rho^{-1}$ with each parameter large enough in terms of the ones to the right of it.

From the union bound it follows that if we can show that each of (i) and (ii) is individually satisfied by proportion $> 1/2$ of tuples $(h_1, \dots, h_k) \in [1, H]^k$ with h_i distinct, then there exists a choice of $(h_1, \dots, h_k)$ that obeys each of the two properties.

To show that proportion $> 1/2$ of tuples $(h_1, \dots, h_k)$ satisfy (i), by the pigeonhole principle it certainly suffices to show that

$$\sum_{1 \leq h_1, \dots, h_k \leq H} \sum_{n \leq x} g(Q(n+h_1)) \cdots g(Q(n+h_k)) \geq 0.999(2\rho)^k x \cdot H^k. \quad (9.2)$$

By writing out the Fourier series for $g(x)$ from Lemma 9.2, and truncating it at height $K = (\log H)^{10}$, the left-hand side of (9.2) becomes

$$\begin{aligned} & (2\rho(1 - \frac{1}{10000k^2}))^k x \cdot H^k + \sum_{\substack{|j_1|, \dots, |j_k| \leq K \\ (j_1, \dots, j_k) \neq (0, \dots, 0)}} c_{j_1} \cdots c_{j_k} \sum_{n \leq x} \prod_{i=1}^k \sum_{h \leq H} e(j_i Q(h+n)) \\ & + O_k(H^k \frac{x}{(\log H)^{10}}). \end{aligned} \quad (9.3)$$

To bound the sum over $j_1, \dots, j_k$ in (9.3), we note that if $j_m \neq 0$ then by Lemma 9.1, and our choice $H = (q')^{1/2}$ where $|\alpha_s - a'/q'| \leq 1/(q')^2$, we have

$$\left| \sum_{h \leq H} e(j_m Q(h+n)) \right| \ll H^{1-\gamma_s} \quad (9.4)$$

for some constant $\gamma_s > 0$. Since j_m cannot all be zero in the first sum in (9.3), we see that the whole expression (9.3) is

$$\geq (2\rho(1 - \frac{1}{10000k^2}))^k x \cdot H^k + O(xH^{k-\gamma_s}(\log H)^{10k}) - O_k(\frac{xH^k}{(\log H)^{10}}) \geq 0.999(2\rho)^k xH^k,$$

recalling that H is large enough in terms of k and ρ . This gives the desired inequality. $\square$

We will need the following lemma on simultaneous equidistribution of minor arc polynomial phases.

Lemma 9.4 (Equidistribution of minor arc polynomial phases). *Let $s \geq 1$, $k \geq 1$ and $D \geq 1$ be fixed. Then there exists $c_s > 0$ such that the following holds.*

Let $Q(y) = \alpha_s y^s + \cdots + \alpha_1 y + \alpha_0$ with $\alpha_0, \alpha_1, \dots, \alpha_s \in \mathbb{R}$. Let $1 \leq a \leq q$ be such that $(a, q) = 1$ and $|\alpha_s - a/q| \leq 1/q^2$. Let $x = q^2$. Let $|j_i|, |h_i| \leq x^{1/(10s)}$ be any integers such that the h_i are distinct and $j_1 Q(y+h_1) + \cdots + j_k Q(y+h_k)$ is not constant. Then, for x large enough,

1. *the polynomial sequence $(j_1 Q(n+h_1) + \cdots + j_k Q(n+h_k) + \mathbb{Z})_{n \leq x}$ on the nilmanifold $\mathbb{R}/\mathbb{Z}$ is totally x^{-c_s} -equidistributed;*
2. *the nilsequence $(e(j_1 Q(n+h_1) + \cdots + j_k Q(n+h_k)))_{n \leq x}$ has mean value $\ll x^{-c_s}$.*

Proof. Let

$$g(y) := j_1 Q(y+h_1) + \cdots + j_k Q(y+h_k) = \sum_{0 \leq d \leq s} \beta_d y^d \quad (9.5)$$

for some $\beta_d \in \mathbb{R}$. Then $e(g(y))$ is a nilsequence on $\mathbb{R}/\mathbb{Z}$ of degree $\leq s$, dimension 1 and Lipschitz constant $O(1)$. Note that $\int_{\mathbb{R}/\mathbb{Z}} e(z) dz = 0$. Hence, by the quantitative Leibman theorem (Theorem 3.5), the sequence $(g(n) + \mathbb{Z})_{n \leq x}$ is totally x^{-c_s} -equidistributed with mean value $\ll x^{-c_s}$ for some $c_s > 0$, provided that there does not exist any integer $1 \leq \ell \leq x^{1/10}$ (say) satisfying $\|\ell g\|_{C^\infty(x)} \leq x^{1/10}$. Suppose for the sake of contradiction that such an integer ℓ exists. By Lemma 3.4, the existence of ℓ implies the existence of another integer $1 \leq \ell' \ll_s x^{1/10}$ such that

$$\|\ell' \beta_d\| \ll_s x^{1/10-d} \quad \text{for all } 1 \leq d \leq s. \quad (9.6)$$

By the binomial formula, we see that

$$\beta_d = \sum_{d \leq m \leq s} J_{d,m} \alpha_m, \quad \text{where} \quad J_{d,m} := \binom{m}{d} \sum_{1 \leq r \leq k} j_r h_r^{m-d}. \quad (9.7)$$

We have the crude upper bound

$$|J_{d,m}| \leq 2^s \cdot k \cdot x^{1/(10s)+1/10} < x^{1/4}/2 \quad (9.8)$$

if x is large enough.

Now, using $|\alpha_s - a/q| \leq 1/q^2$, (9.6) and (9.8), we have

$$x^{-9/10} \gg_s \|\ell' \beta_s\| = \|\ell' J_{s,s} \alpha_s\| \geq \left\| \ell' J_{s,s} \frac{a}{q} \right\| - \frac{\ell' |J_{s,s}|}{q^2} \geq \left\| \ell' J_{s,s} \frac{a}{q} \right\| - \frac{1}{2q}.$$

Since $1/(2q) = 1/(2x^{1/2})$, and $\ell' |J_{s,s}| < q$, this is a contradiction unless $J_{s,s} = 0$. We will show by backwards induction on d that $J_{d,m} = 0$ for all $d \leq m \leq s$. The case $d = s$ has been handled. Suppose that $J_{d+1,m} = 0$ for all $d+1 \leq m \leq s$. Note that then $J_{d,m} = \binom{m}{d} / \binom{m+1}{d+1} \cdot J_{d+1,m+1} = 0$ for $d \leq m \leq s-1$. Therefore, by (9.6) and (9.8), we have

$$x^{-9/10} \gg_s \|\ell' \beta_d\| = \left\| \ell' \sum_{d \leq m \leq s} J_{d,m} \alpha_m \right\| = \|\ell' J_{d,s} \alpha_s\| \geq \left\| \ell' J_{d,s} \frac{a}{q} \right\| - \frac{\ell' |J_{d,s}|}{q^2} \geq \left\| \ell' J_{d,s} \frac{a}{q} \right\| - \frac{1}{2q}.$$

Again since $1/(2q) = 1/(2x^{1/2})$ and $\ell' |J_{d,s}| < q$, this is a contradiction unless $J_{d,s} = 0$.

We have now shown that $J_{d,m} = 0$ for all $1 \leq d \leq m \leq s$. But then, by (9.5) and (9.7), the polynomial $g(y)$ is constant, contrary to assumption. $\square$

Proof of Theorem 2.3. Let $s \geq 1$, $k \in \mathbb{N}$ and $\rho > 0$ be fixed, with k large enough in terms of s and ρ small enough in terms of s, k . Let $k < w \ll_Q 1$ be such that, denoting $W = \prod_{p \leq w} p$, all the rational coefficients of the polynomial $Q(Wy + 1)$ (if there are any) are integers. Now, by restricting to the smaller nil-Bohr set $B' = B \cap (W\mathbb{Z} + 1)$, and denoting $Q_1(y) = Q(Wy + 1)$, it suffices to prove that the set

$$\{Wn + 1 \in \mathbb{P} : \|Q_1(n)\| < \rho\}$$

has bounded gaps, with Q_1 now having the form

$$Q_1(y) = \alpha_s y^s + \cdots + \alpha_1 y + \alpha_0, \quad \alpha_1, \dots, \alpha_s \in (\mathbb{R} \setminus \mathbb{Q}) \cup \{0\}, \quad \alpha_s \neq 0. \quad (9.9)$$

In particular, the leading coefficient α_s of Q_1 is irrational. We shall rename Q_1 as Q for simplicity.

Let H be an integer that satisfies the conclusion of Lemma 9.3 (for some fast-growing but fixed function $\mathcal{F}$). Note that H depends only on k, ρ, Q .

Since α_s is irrational, we may find an infinite set $\mathcal{Q}$ of integers $q \geq 1$ such that for some a coprime to q we have $|\alpha_s - a/q| \leq 1/q^2$. Set $x = q^2$ with q ranging over $\mathcal{Q}$, and note that x tends to infinity along an infinite subsequence of the integers. We are going to show that $\{Wn + 1 \in \mathbb{N} : \|Q(n)\| < \rho\}$ contains infinitely many pairs of primes differing by $\leq WH$.

To prove this we shall apply the weighted version of Maynard's sieve given by Proposition 8.2. Let $g = g_{\rho, 1/(10000k^2)}$ be the minorant for $1_{\|x\| < \rho}$ constructed before. We set our weight ω_n to be

$$\omega_n = g(Q(n + h_1)) \cdots g(Q(n + h_k)),$$

where $(h_1, \dots, h_k) \in [1, H]^k$ is any k -tuple of distinct elements satisfying the conditions of Lemma 9.3.

Note that ω_n is a nonnegative minorant for the indicator function of $\{n : \|Q(n+j)\| < \rho \ \forall j \leq k\}$. By Lemma 9.3, we have

$$\sum_{n \leq x} \omega_n \geq 0.99(2\rho)^k x. \quad (9.10)$$

We will show that the weight ω_n satisfies the hypotheses of Proposition 8.2, taking $\theta = c_s/10, \delta = 1/10, C_0 = 10$ there, where c_s is the constant in Lemma 9.4. After that from Proposition 8.2 we obtain the lower bound

$$\sum_{\substack{n \leq x \\ |\{W(n+h_1)+1, \dots, W(n+h_k)+1\} \cap \mathbb{P}| \geq (1/(10C_s)) \log k}} \omega_n \gg_{k,s,\rho} \frac{1}{(\log x)^k} \sum_{n \leq x} \omega_n \gg_{k,\rho} \frac{x}{(\log x)^k},$$

for some constant $C_s \geq 1$. By choosing $k = \lceil e^{20C_s} \rceil$, this then proves that gaps of length at most $W \max_{1 \leq i \leq j \leq k} |h_i - h_j| \leq WH$ occur infinitely often in our Bohr set¹.

We shall now inspect hypotheses (i)–(iv) of Proposition 8.2 for x large enough in terms of k, s, ρ, H . They take the forms

(i) Prime number theorem: For $1 \leq h \leq H$,

$$\sum_{\substack{n \leq x \\ W(n+h)+1 \in \mathbb{P}}} \omega_n \geq \frac{1}{10} \frac{W}{\varphi(W)} \sum_{n \leq x} \omega_n;$$

(ii) Level of distribution $c_s/10$ for ω_n :

$$\sum_{r \leq x^{c_s/10}} \max_{c \pmod{r}} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{r}}} \omega_n - \frac{1}{r} \sum_{n \leq x} \omega_n \right| \ll \frac{x}{(\log x)^{101k^2}};$$

(iii) Level of distribution $c_s/10$ for $\omega_n 1_{W(n+h_i)+1 \in \mathbb{P}}$ for $1 \leq i \leq k$:

$$\sum_{r \leq x^{c_s/10}} \max_{(L_i(c), r)=1} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{r} \\ W(n+h_i)+1 \in \mathbb{P}}} \omega_n - \frac{\varphi(W)}{\varphi(Wr)} \sum_{\substack{n \leq x \\ W(n+h_i)+1 \in \mathbb{P}}} \omega_n \right| \ll \frac{x}{(\log x)^{101k^2}};$$

(iv) A Brun–Titchmarsh type bound for ω_n :

$$\max_{c \pmod{r}} \sum_{\substack{n \leq x \\ n \equiv c \pmod{r}}} \omega_n \leq \frac{10}{r} \sum_{n \leq x} \omega_n,$$

uniformly for $1 \leq r \leq x^{c_s/10}$.

¹In fact, if one lets $k \rightarrow \infty$ one sees that for any m there are infinitely many intervals of bounded length containing m primes from our nil-Bohr set.

Note that any nonnegative constant sequence satisfies hypotheses (i)–(iv) by the prime number theorem, the Bombieri–Vinogradov theorem, and the Brun–Titchmarsh inequality. By Lemma 9.2, we may expand ω_n as a Fourier series:

$$\omega_n = (2\rho(1 - 1/(10000k^2)))^k + \sum_{\substack{j_1, \dots, j_k \\ (j_1, \dots, j_k) \neq (0, \dots, 0)}} c_{j_1} \cdots c_{j_k} e(j_1 Q(n+1) + \cdots + j_k Q(n+k)),$$

where the coefficients c_j have the properties stated in Lemma 9.2. Then truncate the Fourier expansion of ω_n to $|j_i| \leq M := (\log x)^{200k^2}$. Recalling properties (ii) and (iii) of the Fourier expansion of $g(x)$ as well as (9.10), it suffices to verify that

$$\omega'_n := e(j_1 Q(n+1) + \cdots + j_k Q(n+k))$$

for any integer tuple $(j_1, \dots, j_k) \in [-M, M]^k \setminus \{0\}$ for which $j_1 Q(y+h_1) + \cdots + j_k Q(y+h_k)$ is non-constant, satisfies the following hypotheses:

(i')

$$\left| \sum_{\substack{n \leq x \\ W(n+h)+1 \in \mathbb{P}}} \omega'_n \right| + \left| \sum_{n \leq x} \omega'_n \right| \ll \frac{x}{(\log x)^{200k^3}};$$

(ii')

$$\sum_{r \leq x^{c_s/10}} \max_{c \pmod{r}} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{r}}} \omega'_n - \frac{1}{r} \sum_{n \leq x} \omega'_n \right| \ll \frac{x}{(\log x)^{200k^3}};$$

(iii')

$$\sum_{r \leq x^{c_s/10}} \max_{\substack{W(c+h_i)+1, r=1}} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{r} \\ W(n+h_i)+1 \in \mathbb{P}}} \omega'_n - \frac{\varphi(W)}{\varphi(Wr)} \sum_{\substack{n \leq x \\ W(n+h_i)+1 \in \mathbb{P}}} \omega'_n \right| \ll \frac{x}{(\log x)^{200k^3}};$$

(iv')

$$\max_{c \pmod{r}} \left| \sum_{\substack{n \leq x \\ n \equiv c \pmod{r}}} \omega'_n \right| = o\left(\frac{x}{r(\log x)^{200k^3}}\right) \quad (9.11)$$

for $1 \leq r \leq x^{c_s/10}$.

Note that by Lemma 9.4 (and the assumption $|\alpha_s - a/q| \leq 1/q^2$ with $q = x^{1/2}$) the sequence ω'_n is totally x^{-c_s} -equidistributed as a nilsequence and has mean $\ll x^{-c_s}$ over $[1, x]$. Hypothesis (i') then follows immediately by combining Lemma 9.1 with the converse to the quantitative Leibman theorem (Theorem 3.6).

Then, to deal with (ii') and (iv'), note that as ω'_n is a degree $\leq s$ nilsequence that is totally x^{-c_s} -equidistributed and has mean value $\ll x^{-c_s}$, also for any $1 \leq c \leq r \leq x^{c_s/2}$ the sequence $(\omega'_{rm+c})_{m \leq x/r}$ is a degree $\leq s$ nilsequence that is $x^{-c_s/2}$ -equidistributed on $[x/r, 2x/r]$ with mean value $\ll x^{-c_s/2}$, so (9.11) automatically holds, in fact in a stronger form where the right-hand side is replaced with $o((x/r)^{1-c_s/9})$. This stronger form immediately implies hypothesis (ii') as well.

Lastly, we deal with (iii'). Note that for any c coprime to d we have

$$\begin{aligned} \sum_{\substack{n \leq x \\ n \equiv c \pmod{r} \\ W(n+h_i)+1 \in \mathbb{P}}} \omega'_n &= \sum_{\substack{m \leq Wx \\ m \equiv 1 \pmod{W} \\ m \equiv W(c+h_i)+1 \pmod{d}}} 1_{\mathbb{P}}(m) \omega'_{(m-1)/W-h_i} + O(1) \\ &\ll_W \max_{\substack{(c', dW)=1 \\ c' \equiv 1 \pmod{W}}} \left| \sum_{\substack{m \leq Wx \\ m \equiv c' \pmod{dW}}} 1_{\mathbb{P}}(m) \omega'_{(m-1)/W-h_i} \right| + 1. \end{aligned}$$

Since ω'_n is totally x^{-c_s} -equidistributed with mean value $\ll x^{-c_s}$, also $\omega'_{(n'-1)/W-h_i}$ is totally $x^{-c_s/2}$ -equidistributed with mean value $\ll x^{-c_s/2}$. Thus, hypothesis (iii') follows from our Bombieri–Vinogradov theorem for equidistributed nilsequences (Theorem 4.3) after applying partial summation to replace $1_{\mathbb{P}}(m)$ with $\Lambda(m)$.

Now hypotheses (i')–(iv') have been verified, so Theorem 2.3 follows. $\square$

10 CHEN PRIMES IN NIL-BOHR SETS

Our task in this section is to prove Theorem 2.5 on Chen primes in nil-Bohr sets.

Proof of Theorem 2.5. Let $Q(x) = \alpha_s x^s + \dots + \alpha_1 x + \alpha_0$. By assumption, we can find at least one $1 \leq j \leq s$ such that α_j is irrational. By Dirichlet's approximation theorem, we can then find infinitely many pairs (a, q) of coprime integers with $|\alpha_j - a/q| \leq 1/q^2$. Restrict to those x that can be written as $x = q^2$ for some such q ; this is a sequence of integers that tends to infinity.

Let $\theta_s > 0$ be a small enough constant. Let $g = g_{\rho, 1/2}$ be the minorant function arising from Lemma 9.2 with $\rho = x^{-\theta_s}$. Then we can estimate

$$\sum_{\substack{p \leq x \\ p \in \mathcal{P}_{\text{Chen}}}} 1_{\|Q(p)\| < p^{-\theta_s}} \geq \sum_{\substack{p \leq x \\ p \in \mathcal{P}_{\text{Chen}}}} g(Q(p)). \quad (10.1)$$

Define $\omega_n := g(Q(n))$. By applying Chen's sieve in the form of Proposition 8.3 to $(\omega_n)_{n \leq x}$, we see that the right-hand side of (10.1) is

$$\gg \frac{1}{(\log x)^2} \sum_{n \leq x} g(Q(n)) - O(x^{0.9}),$$

provided that hypotheses (i)–(iii) of Proposition 8.3 are satisfied.

Expanding out the Fourier series of $g(x)$ given by Lemma 9.2 and truncating it from height $M := x^{3\theta_s}$, we can estimate

$$\begin{aligned} \frac{1}{(\log x)^2} \sum_{n \leq x} g(Q(n)) &= \frac{1}{2} x^{-\theta_s} \frac{x}{(\log x)^2} + \frac{1}{(\log x)^2} \sum_{0 < |j| \leq M} c_j \sum_{n \leq x} e(jQ(n)) + O(x^{1-2\theta_s}) \\ &= \left(\frac{1}{2} + O(x^{-c_s}) \right) \frac{x^{1-\theta_s}}{(\log x)^2} \end{aligned} \quad (10.2)$$

for x large enough and some constant $c_s > 0$, where for the last inequality we used Lemma 9.1 and the assumption that $x = q^2$ where q satisfies $|\alpha_j - a/q| \leq 1/q^2$ (so $\|\ell \alpha_j\| \geq \frac{1}{2q}$ for $1 \leq \ell \leq q/2$).

Therefore, to conclude the proof of Theorem 2.5, it suffices to verify hypotheses (i)–(iii) of Proposition 8.3 for $\omega_n = g(Q(n))$. Since

$$\omega_n = \frac{1}{2}x^{-\theta_s} + \sum_{0 < |j| \leq M} c_j e(jQ(n)) + O(x^{1-2\theta_s})$$

with $|c_j| \ll x^{-\theta_s}$, and since any nonnegative constant sequence satisfies hypotheses (i)–(iii), it suffices to verify that (using the notation of Proposition 8.3) for $0 < |j| \leq M = x^{3\theta_s}$ the sequence $\omega'_n := e(jQ(n))$ satisfies

(i')

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r,2)=1}} \lambda'_r \sum_{\substack{n \leq x \\ n+2 \equiv 0 \pmod{r} \\ n \in \mathbb{P}}} \omega'_n \right| \ll x^{1-4\theta_s}, \quad (10.3)$$

(ii')

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r,2)=1}} \lambda_r \sum_{\substack{n \leq x \\ n \equiv 0 \pmod{r} \\ n+2 \in B_i}} \omega'_n \right| \ll x^{1-4\theta_s}, \quad (10.4)$$

(iii')

$$\left| \sum_{\substack{n \leq x \\ n \in B_i}} \omega'_n \right| + \left| \sum_{\substack{n \leq x \\ n+2 \in B_i}} \omega'_n \right| \ll x^{1-4\theta_s}, \quad (10.5)$$

where λ_r is well-factorable of level $x^{1/2-\varepsilon}$, λ'_r is either well-factorable of level $x^{1/2-\varepsilon}$ or $\lambda'_r = 1_{p \in [P, P']} * \lambda''_r$ with λ''_r well-factorable of level $x^{1/2-\varepsilon}/P$ and $2P \geq P' \geq P \in [x^{1/10}, x^{1/3-\varepsilon}]$.

Note that ω'_n is a degree s nilsequence and that, by the quantitative Leibman theorem and the fact that $\|\ell\alpha_j\| \geq \frac{1}{2q}$ for $1 \leq \ell \leq q/2$, it is totally x^{-c_s} -equidistributed for some small constant $c_s > 0$. Moreover, by Lemma 9.1, ω'_n has mean value $\ll x^{-c_s}$ over $[1, x]$.

We first verify hypothesis (i'). By applying partial summation, it suffices to prove (i') with the condition $n \in \mathbb{P}$ replaced by the weight $\Lambda(n)$. Then, by Vaughan's identity, it suffices to prove that

$$\left| \sum_{\substack{r \leq x^{1/2-\varepsilon} \\ (r,2)=1}} \lambda'_r \sum_{\substack{mn \leq x \\ mn \equiv -2 \pmod{r} \\ M \leq m \leq 2M}} a_m b_n \omega'_{mn} \right| \ll x^{1-5\theta_s},$$

where either $M \in [x^{1/2}, x^{2/3+o(1)}]$ and $|a_m|, |b_m| \ll d(n)^{O(1)}$ (type II case), or $M \leq x^{1/3+o(1)}$ and $b_n \in \{1, \log n\}$ (type I case). Note that ω'_n is a degree s nilsequence which is totally x^{-c_s} -equidistributed and has mean $\ll x^{-c_s}$ on $[1, x]$. Now, by applying Proposition 5.2 in the type I case and Proposition 6.6 in the type II case (and making use of the “partial well-factorability” of λ'_r to write it as a convolution of two factors supported on $[1, x^{1-\varepsilon/2}/M]$ and $[1, Mx^{-1/2}]$), we obtain for the left-hand side of (10.3) a bound of $\ll x^{1-c_{s,\varepsilon}}$, where $c_{s,\varepsilon}$ is small enough. If we require that $\theta_s < c_{s,\varepsilon}/5$, the bound obtained is good enough.

For verifying hypothesis (ii'), we note that

$$1_{B_1}(n) = \sum_{\substack{n=mp \\ p \geq x^{1/10}}} a_m, \quad 1_{B_2}(n) = \sum_{\substack{n=mp \\ p \geq x^{1/10}}} a'_m, \quad (10.6)$$

where a_m is the indicator of m being of the form $p_1 p_2$ with $x^{1/10} \leq p_1 \leq x^{1/3-\varepsilon}$ and $x^{1/3-\varepsilon} \leq p_2 \leq ((2x+2)/p_1)^{1/2}$ and a'_m is the indicator of m being of the form $p_1 p_2$ with $x^{1/3-\varepsilon} \leq p_1 \leq p_2 \leq ((2x+2)/p_1)^{1/2}$. Thus $1_{B_1}(n), 1_{B_2}(n)$ are both type II convolutions with the m variable supported in $[x^{1/3}, x^{2/3+o(1)}]$. Hence, from Proposition 6.5, we obtain for the left-hand side of (10.4) a bound of $\ll x^{1-c'_{s,\varepsilon}}$ for some constant $c'_{s,\varepsilon} > 0$. If we require that $\theta_s < c'_{s,\varepsilon}/4$, the bound is good enough.

Lastly, to handle hypothesis (iii'), we appeal to (10.6) again to reduce matters to type II sums, and then apply Proposition 6.7 (taking only the $d = 1$ term in the sum there). We obtain for the left-hand side of (10.5) a bound of $\ll x^{1-c''_{s,\varepsilon}}$ for some $c''_{s,\varepsilon} > 0$. If we require that $\theta_s < c''_{s,\varepsilon}/4$, the bound obtained is strong enough.

The proof is now complete. $\square$

11 LINEAR EQUATIONS IN PRIMES IN ARITHMETIC PROGRESSIONS

We now turn to the proof of Theorem 2.7, which also includes the proof of Theorem 2.9.

Proof of Theorem 2.7. We adapt some arguments from [17]. Let $w = w(x)$ be a positive integer that tends to $+\infty$ slowly enough. By dividing the variables $\mathbf{n}$ into residue classes modulo $\mathcal{P}(w)$, it suffices to show that the result holds when $\mathcal{P}(w) \mid q$, with $O_A(Q(\log x)^{-A})$ exceptions. The details of this reduction is similar to the argument in [17, Section 5].

Henceforth assume that $\mathcal{P}(w) \mid q$. By the generalized von Neumann theorem [17, Proposition 7.1] and the relative version of the inverse theorem for the Gowers norms [17, Proposition 10.1] (see also [20, Theorem 1.3]), we have (2.3) for q outside an exceptional set of size $\ll_A Q(\log x)^{-A}$, provided that the following two conditions hold.

(i) For all but $\ll_A Q(\log x)^{-A}$ choices of $q \leq Q$ the following holds. For each invertible residue class $a \pmod{q}$, the function $\Lambda_{a,q}(n) := \frac{\varphi(q)}{q} \Lambda(qn + a)$ is majorized by $C_0 v(n)$, where v is D -pseudorandom with D large enough in terms of t, d, M (see [17, Section 6] for the definition of the pseudorandomness conditions).

(ii) For all but $\ll_A Q(\log x)^{-A}$ choices of $q \leq Q$ we have the Gowers norm bound $\|\Lambda_{a,q} - 1\|_{U^k[x]} = o_{w \rightarrow \infty}(1) + o_{w^*x \rightarrow \infty}(1)$ for all $k \geq 1$. (This is precisely the content of Theorem 2.9.)

Verifying (i):

For showing condition (i), we follow [17, Appendix D] that establishes the analogous claim for $q = 1$, indicating the necessary modifications (see also [3, Proposition 6.1] that handles the case $q \ll_A (\log x)^A$). Let

$$\mathcal{Q} := \{q \leq Q : \Omega(q) \leq C \log \log x\},$$

where C is a large enough constant and $\Omega(q)$ denotes the number of prime factors of q with multiplicities. We have

$$|[1, Q] \setminus \mathcal{Q}| \ll 2^{-C \log \log x} \sum_{q \leq Q} 2^{\Omega(q)} \ll Q(\log x)^{-C/2},$$

so it suffices to prove condition (i) for $q \in \mathcal{Q}$.

As in [17, Appendix D], we define

$$\begin{aligned} v(n) &:= \frac{1}{2} + \frac{1}{2} \tilde{v}(n), \\ \text{with } \tilde{v}(n) &= \frac{\varphi(q)}{q} \Lambda_{\chi, R, 2}(qn + a) := \frac{\varphi(q)}{q} (\log R) \left(\sum_{\substack{d|qn+a \\ d \leq R}} \mu(d) \chi\left(\frac{\log d}{\log R}\right) \right)^2, \end{aligned}$$

where χ is a smooth function compactly supported in $[-1, 1]$ and $\chi(0) = 1$, with $0 \leq \chi(y) \leq 1$ everywhere. Note that since $\mathcal{P}(w) \mid q$, the W -trick is already incorporated in the definition of $\Lambda_{\chi, R, 2}$.

Let N be a prime of size $(C_D + o(1))x$ for large enough $C_D > 0$, and extend $v(n)$ to $[1, N]$ by defining it to be $= 1$ elsewhere. Embed $v(n)$ into the cyclic group $\mathbb{Z}/N\mathbb{Z}$ in the obvious way. Then the function v is our choice of a pseudorandom measure.

We will inspect that the Goldston–Yıldırım estimate of [17, Theorem D.3] (with $N = x$) holds for the family of linear forms $\mathbf{n} \mapsto (L_1(q\mathbf{n} + a), \dots, L_t(q\mathbf{n} + a))$. The case $q = 1$ is [17, Theorem D.3]. We will then deduce the linear forms and correlation conditions for $\tilde{v}(n)$ by following [17, pp. 75–77]. Both of these arguments ([17, Theorem D.3] and [17, pp. 75–77]) go through in our setting with the following minor modifications:

- The factor $e^{O(X)}(\log R)^{-1/2}$ in [17, Theorem D.3] is $o(1)$ and therefore harmless, since

$$X \ll_{t, d, L} 1 + \sum_{p|q} p^{-1/2} \ll \log \log \log N$$

by the assumption $q \in \mathcal{Q}$ and the fact that the “exceptional” primes p in [17, Theorem D.3] are either $O_{d, t, L}(1)$ or divide q .

- In the proof of [17, Theorem D.3] there are a few conditions (such as $\alpha(p, B) = 1/p$ and $\beta_p = 1 + O(1/p)$) that only hold for the primes $p \nmid q$ in our setting (as opposed to all large enough primes). This makes little difference in the argument, since we can separate the contribution of the primes $p \mid q$ from the rest and the contribution of the rest of the primes gives the correct local factors, whereas for $p \mid q$ quantities such as $E_{p, \xi}$ in [17, Lemma D.5] are easy to compute, which ultimately leads to the value $\beta_p = p/\varphi(p)$ for $p \mid q$, as desired.
- When verifying the linear forms and correlation conditions in [17, pp. 75–77], one should replace W with q and conditions such as $p \leq w$ with $p \mid q$ and $p > w$ with $p \nmid q$. The estimates in [17, p. 77] go through verbatim with these modifications.

Thus condition (i) has been verified.

Verifying (ii):

We utilize Theorem 1.3, which implies that for any $A \geq 2$ we have

$$\max_{(a,q)=1} \sup_{\psi \in \Psi_k(\Delta, \log x)} \frac{q}{\varphi(q)} \left| \sum_{n \leq x} \Lambda_{a,q}(n) \psi(n) - \frac{\varphi(q)V}{\varphi(qV)} \sum_{\substack{n \leq x \\ (qn+a,V)=1}} \psi(n) \right| \ll_{k,A,\Delta,\varepsilon} \frac{x}{(\log x)^A}, \quad (11.1)$$

for all but $O(Q(\log x)^{-A})$ values of $q \leq Q$, where $V = \mathcal{P}((\log x)^B)$ for some sufficiently large constant $B = B(A, k, \Delta) > 0$. Let

$$\tilde{\Lambda}_{a,q}(n) = \frac{\varphi(q)V}{\varphi(qV)} \cdot 1_{(qn+a,V)=1}$$

be the function appearing in the second sum in (11.1). We claim that this function (embedded to a cyclic group) is D -pseudorandom for any fixed D . To see this, let χ be a smooth function supported on $[-1, 1]$ with $\chi(y) = 1$ for $|y| \leq 1/2$ and $0 \leq \chi(y) \leq 1$ everywhere. Let

$$\Lambda_{\chi,R,2}^{(V)}(n) := (\log R) \left(\sum_{\substack{d|n \\ d|V \\ d \leq R}} \mu(d) \chi\left(\frac{\log d}{\log R}\right) \right)^2,$$

where $R = N^\gamma$ for small enough $\gamma > 0$. Then we have

$$\tilde{\Lambda}_{a,q}(n) = (1 + o(1)) \frac{\varphi(q)V}{\varphi(qV)(\log R)} \Lambda_{\chi,R,2}^{(V)}(qn+a) + O\left(\frac{\varphi(q)V}{\varphi(qV)} E(qn+a)\right),$$

where

$$E(n) = d(n)^2 1_{\exists m|n, m \geq R^{1/2}, m \text{ is } V\text{-smooth}}.$$

Similarly as in the proof of Theorem 4.6, the term $E(qn+a)$ (which is bounded by a divisor-type function) is negligible in the linear forms and correlation conditions for $(a, q) = 1$ and all but $\ll_A Q/(\log x)^A$ values of $q \leq Q$. The function $\frac{\varphi(q)V}{\varphi(qV)(\log R)} \Lambda_{\chi,R,2}^{(V)}(qn+a)$ (embedded into a cyclic group) in turn is a pseudorandom measure for $(a, q) = 1$ and $q \leq Q$, outside an exceptional set of moduli q of size $\ll_A Q/(\log x)^A$, by the same argument that was used to verify condition (i). Thus $\tilde{\Lambda}_{a,q}(n)$ itself is a pseudorandom measure.

Now by [20, Theorem 1.3], [17, Proposition 10.1] and condition (i), (11.1) implies

$$\|\Lambda_{a,q} - \tilde{\Lambda}_{a,q}\|_{U^k[x]} = o(1)$$

for almost all q . Thus it remains to establish that $\|\tilde{\Lambda}_{a,q} - 1\|_{U^k[x]} = o(1)$. But this follows from the pseudorandomness of $\tilde{\Lambda}_{a,q}$ together with [16, Lemma 5.2]. $\square$

Lastly, in view of Remark 1.8, we note that as a consequence of the proof method (in fact with some simplifications in the sense that we do not need pseudorandom majorants), we obtain the following result for the Möbius function.

Corollary 11.1. *Let $\varepsilon > 0$ and $A, t, d, M \geq 1$ be given. Let $x \geq 10$ and $Q \leq x^{1/3-\varepsilon}$. Then for all but $\ll_{\varepsilon, A, t, d, M} Q/(\log x)^A$ choices of $1 \leq q \leq Q$ the following holds. For every $\mathbf{a} \in (\mathbb{Z}/q\mathbb{Z})^d$ and every finite complexity tuple $\Psi = (L_1(\mathbf{n}), \dots, L_t(\mathbf{n}))$ of non-constant affine-linear forms in d variables of size $\|\Psi\| \leq M$ we have*

$$\sum_{\mathbf{n} \in [1, x]^d} \mu(L_1(q\mathbf{n} + \mathbf{a})) \cdots \mu(L_t(\mathbf{n} + \mathbf{a})) = o_{t, d, M}(x^d). \quad (11.2)$$

ACKNOWLEDGMENTS

We thank Ben Green, James Maynard and Terence Tao for helpful comments. We are grateful to the anonymous referee for a careful reading of the paper and for comments that improved the quality of this paper.

REFERENCES

- [1] R. Alweiss and S. Luo. Bounded gaps between primes in short intervals. *Res. Number Theory*, 4(2):Paper No. 15, 27, 2018. 6
- [2] R. C. Baker and L. Zhao. Gaps between primes in Beatty sequences. *Acta Arith.*, 172(3):207–242, 2016. 6
- [3] P.-Y. Bienvenu. A higher-dimensional Siegel-Walfisz theorem. *Acta Arith.*, 179(1):79–100, 2017. 9, 50
- [4] E. Bombieri, J. B. Friedlander, and H. Iwaniec. Primes in arithmetic progressions to large moduli. *Acta Math.*, 156(3-4):203–251, 1986. 4
- [5] J. R. Chen. On the representation of a larger even integer as the sum of a prime and the product of at most two primes. *Sci. Sinica*, 16:157–176, 1973. 8
- [6] L. Chua, S. Park, and G. D. Smith. Bounded gaps between primes in special sequences. *Proc. Amer. Math. Soc.*, 143(11):4597–4611, 2015. 6
- [7] É. Fouvry, E. Kowalski, and P. Michel. On the exponent of distribution of the ternary divisor function. *Mathematika*, 61(1):121–144, 2015. 2
- [8] É. Fouvry and M. Radziwiłł. Level of distribution of unbalanced convolutions. *arXiv e-prints*, page arXiv:1811.08672, November 2018. 2
- [9] E. Fouvry and G. Tenenbaum. Répartition statistique des entiers sans grand facteur premier dans les progressions arithmétiques. *Proc. London Math. Soc. (3)*, 72(3):481–514, 1996. 2
- [10] N. Frantzikinakis and B. Host. Higher order Fourier analysis of multiplicative functions and applications. *J. Amer. Math. Soc.*, 30(1):67–157, 2017. 11, 12
- [11] J. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010. 4

- [12] J. B. Friedlander and H. Iwaniec. The divisor problem for arithmetic progressions. *Acta Arith.*, 45(3):273–277, 1985. 2
- [13] J. B. Friedlander and H. Iwaniec. Incomplete Kloosterman sums and a divisor problem. *Ann. of Math. (2)*, 121(2):319–350, 1985. With an appendix by Bryan J. Birch and Enrico Bombieri. 2
- [14] A. Granville and X. Shao. When does the Bombieri-Vinogradov theorem hold for a given multiplicative function? *Forum Math. Sigma*, 6:e15, 23, 2018. 2
- [15] A. Granville and X. Shao. Bombieri-Vinogradov for multiplicative functions, and beyond the $x^{1/2}$ -barrier. *Adv. Math.*, 350:304–358, 2019. 2
- [16] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math. (2)*, 167(2):481–547, 2008. 9, 52
- [17] B. Green and T. Tao. Linear equations in primes. *Ann. of Math. (2)*, 171(3):1753–1850, 2010. 4, 8, 9, 50, 51, 52
- [18] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math. (2)*, 175(2):541–566, 2012. 7, 8, 11, 17
- [19] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math. (2)*, 175(2):465–540, 2012. 2, 3, 10, 11, 12, 13, 21, 42
- [20] 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. 8, 50, 52
- [21] A. J. Harper. Bombieri-Vinogradov and Barban-Davenport-Halberstam type theorems for smooth numbers. *arXiv e-prints*, page arXiv:1208.5992, August 2012. 2
- [22] D. R. Heath-Brown. The divisor function $d_3(n)$ in arithmetic progressions. *Acta Arith.*, 47(1):29–56, 1986. 2
- [23] A. Hildebrand and G. Tenenbaum. Integers without large prime factors. *J. Théor. Nombres Bordeaux*, 5(2):411–484, 1993. 40
- [24] B. Host and B. Kra. Nil-Bohr sets of integers. *Ergodic Theory Dynam. Systems*, 31(1):113–142, 2011. 6
- [25] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004. 18, 35, 36, 38, 40
- [26] K. Matomäki. A Bombieri-Vinogradov type exponential sum result with applications. *J. Number Theory*, 129(9):2214–2225, 2009. 6, 8
- [27] K. Matomäki and X. Shao. Vinogradov’s three primes theorem with almost twin primes. *Compos. Math.*, 153(6):1220–1256, 2017. 10, 41, 42

- [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, November 2019. [37](#), [38](#)
- [29] J. Maynard. Small gaps between primes. *Ann. of Math. (2)*, 181(1):383–413, 2015. [6](#)
- [30] J. Maynard. Dense clusters of primes in subsets. *Compos. Math.*, 152(7):1517–1554, 2016. [6](#), [41](#)
- [31] H. Mikawa. On exponential sums over primes in arithmetic progressions. *Tsukuba J. Math.*, 24(2):351–360, 2000. [6](#)
- [32] X. Shao. Gowers norms of multiplicative functions in progressions on average. *Algebra Number Theory*, 11(4):961–982, 2017. [5](#), [15](#), [28](#)
- [33] P. Shiu. A Brun-Titchmarsh theorem for multiplicative functions. *J. Reine Angew. Math.*, 313:161–170, 1980. [39](#)
- [34] T. Tao. *Higher order Fourier analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012. [2](#), [3](#)
- [35] J. Thorner. Bounded gaps between primes in Chebotarev sets. *Res. Math. Sci.*, 1:Art. 4, 16, 2014. [6](#)
- [36] T. L. Todorova and D. I. Tolev. On the distribution of αp modulo one for primes p of a special form. *Math. Slovaca*, 60(6):771–786, 2010. [6](#)
- [37] D. I. Tolev. Additive problems with prime numbers of special type. *Acta Arith.*, 96(1):53–88, 2000. [6](#)
- [38] I. M. Vinogradov. *The method of trigonometrical sums in the theory of numbers*. Dover Publications, Inc., Mineola, NY, 2004. Translated from the Russian, revised and annotated by K. F. Roth and Anne Davenport, Reprint of the 1954 translation. [43](#)
- [39] Y. Zhang. Bounded gaps between primes. *Ann. of Math. (2)*, 179(3):1121–1174, 2014. [6](#)

AUTHORS

Xuancheng Shao
 Department of Mathematics
 University of Kentucky
 Lexington, KY, 40506, USA
xuancheng.shao@uky.edu

Joni Teräväinen
 Mathematical Institute
 University of Oxford
 Radcliffe Observatory Quarter
 Woodstock Rd
 Oxford OX2 6GG, UK
joni.teravainen@maths.ox.ac.uk