

This is a repository copy of *Roth–Waring–Goldbach*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/111119/>

Version: Accepted Version

Article:

Chow, Samuel Khai Ho (2017) Roth–Waring–Goldbach. *International Mathematics Research Notices*. 2341–2374. ISSN: 1687-0247

<https://doi.org/10.1093/imrn/rnw307>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

ROTH–WARING–GOLDBACH

SAM CHOW

ABSTRACT. We use Green’s transference principle to show that any subset of the d th powers of primes with positive relative density contains nontrivial solutions to a translation-invariant linear equation in $d^2 + 1$ or more variables, with explicit quantitative bounds.

1. INTRODUCTION

Waring’s problem [34] dates back to 1770, and asks how large s has to be in terms of d to ensure that if n is a large positive integer then

$$x_1^d + \cdots + x_s^d = n \tag{1.1}$$

has a solution $\mathbf{x} \in \mathbb{N}^s$. The Hardy–Littlewood circle method has been a particularly effective approach to such problems, with the best results due to Wooley — see [32, 36], as well as [8, §2] and [37]. The circle method has also been used to solve the ternary Goldbach problem, and other problems concerning the addition of primes [15, 33]. Since Hua [18], many authors have enjoyed working on the Waring–Goldbach problem, which considers prime solutions to (1.1) — see [20, 22, 23, 28, 29], for instance. The circle method has again been the weapon of choice, with the main technical issue being the study of exponential sums over primes [19].

Roth’s theorem [25] states that if $A \subset [N]$ contains no nontrivial three-term arithmetic progressions then $|A| \ll \frac{N}{\log \log N}$. This bound has since been improved, most recently by Bloom [5]. Such results are interesting because they identify patterns in the set A without assuming anything about its structure. Three-term arithmetic progressions pertain to the diophantine equation

$$x - 2y + z = 0,$$

and much of the arithmetic combinatorics literature surrounds linear equations. Smith [26], Keil [21] and Henriot [16, 17] have considered higher degree systems with the property that the solution set is invariant under translations and dilations. This property allows the use of a density increment strategy, which is the standard approach to Roth’s theorem.

In 2005, Green [14] famously solved a problem of Roth–Goldbach type for three primes. He devised a means of transferring Roth-type results from the integers to the primes. Using this mechanism, he showed that any subset of the

2010 *Mathematics Subject Classification*. 11B30, 11D72, 11P32, 11P55, 37A45.

Key words and phrases. Arithmetic combinatorics, diophantine equations, primes, circle method, restriction theory.

primes with positive *relative density* contains nontrivial three-term arithmetic progressions. Recently Browning and Prendiville [10] have shown Green's transference method to be versatile, establishing a theorem of Roth–Waring type for five squares. They were able to transfer results from the integers to the squares, thereby obtaining a Roth-type bound for a quadratic equation without the property of translation-dilation invariance. The present article combines aspects of Roth's theorem, Waring's problem and Goldbach problems.

Let c_1, \dots, c_s be nonzero integers such that

$$c_1 + \dots + c_s = 0. \quad (1.2)$$

Let K be a union of k proper subspaces of the rational hyperplane

$$c_1 x_1 + \dots + c_s x_s = 0, \quad (1.3)$$

each of which contains the diagonal

$$\{(x, \dots, x) : x \in \mathbb{Q}\}. \quad (1.4)$$

Let $d \geq 2$ be an integer, and let A be a set of primes in $[X] := \{1, 2, \dots, X\}$ such that the only solutions $\mathbf{x} \in A^s$ to

$$c_1 x_1^d + \dots + c_s x_s^d = 0 \quad (1.5)$$

have $(x_1^d, \dots, x_s^d) \in K$.

Theorem 1.1. *Assume $s \geq C(d)$, where $C(2) = 5$, $C(3) = 9$, $C(4) = 15$ and*

$$C(d) = d^2 + 1 \quad (d \geq 5).$$

Then

$$|A| \ll_{\mathbf{c}, k, \varepsilon} \frac{X}{\log X} (\log \log \log \log X)^{\frac{2-s}{d} + \varepsilon}. \quad (1.6)$$

Loosely, this says that any subset of relative density $(\log \log \log \log)^{(2-s)/d + \varepsilon}$ within the primes contains nontrivial solutions to (1.5). One could choose

$$K = \bigcup_{i \neq j} \{\mathbf{x} \in \mathbb{Q}^s : x_i = x_j, \mathbf{c} \cdot \mathbf{x} = 0\},$$

for instance, which was the choice of Keil [21] and Henriot [16] in their work on diagonal quadrics in dense variables. Thus, any positive density subset of the primes contains a solution to (1.5) with pairwise distinct coordinates. The reason for having a notion of trivial solutions is that, by (1.2), the diagonal (1.4) lies within the solution set of (1.5).

The assumption (1.2) is necessary for Theorem 1.1. Indeed, if $c_1 + \dots + c_s \neq 0$, then fix a prime ℓ that does not divide $c_1 + \dots + c_s$. The set

$$A = \{p \in [X] : p \text{ prime}, p \equiv 1 \pmod{\ell}\}$$

has positive relative density in the set of primes up to X , so if X is large then (1.6) does not hold. Moreover, if $x_1, \dots, x_s \in A$ then

$$c_1 x_1^d + \dots + c_s x_s^d \equiv c_1 + \dots + c_s \not\equiv 0 \pmod{\ell}.$$

In particular, there are no solutions $\mathbf{x} \in A^s$ to (1.5). This construction shows that Theorem 1.1 cannot hold without the condition (1.2).

The strategy is to use Green’s transference technology [14] to transfer Roth-type results from the integers to the set of d th powers of primes. The protagonist is a measure ν on some interval $[N]$, where N can be thought of as X^d . Morally $\nu(n)$ should be $dp^{d-1} \log p$, if $n = p^d$ for some prime $p \leq X$, and zero otherwise. The measure ν has become known as a majorant; a *majorant* on $[N]$ is a function $\nu : \mathbb{Z} \rightarrow [0, \infty)$ with support in $[N]$. Our majorant has the additional normalisation property that

$$\|\nu\|_1 \sim N. \quad (1.7)$$

We refer the curious reader to the expository article [24] for more on the history and terminology of the transference principle.

The point is that our set A can be lifted to $[N]$ and weighted by ν to behave like a dense subset — not of $\mathcal{P}_X := \{\text{prime } p \leq X\}$, but of $[N]$. Bloom’s theorem [4] then ensures that the ν -weighted solution count is large in terms of the density — see [10, §2] and [24, §1.2]. Since A has only K -trivial solutions to (1.5), in the sense that the only solutions $\mathbf{x} \in A^s$ to (1.5) have $(x_1^d, \dots, x_s^d) \in K$, we also obtain an upper bound for this count. Combining the two inequalities reaps a density bound of the shape (1.6).

Browning and Prendiville [10] have distilled the method into the following ingredients.

- (1) Density transfer. We shall lift our set $A \subset \mathcal{P}_X$ to a set $\mathcal{A} \subset [N]$ in the support of our majorant ν . With

$$\delta := |A| \frac{\log X}{X}, \quad (1.8)$$

we show that \mathcal{A} has a ν -weighted density of at least δ^d in $[N]$. In other words, we establish the bound

$$\sum_{n \in \mathcal{A}} \nu(n) \gg \delta^d N. \quad (1.9)$$

- (2) Fourier decay. The majorant ν has *Fourier decay of level θ* if

$$\|\hat{\nu} - \widehat{1_{[N]}}\|_\infty \leq \theta N.$$

We shall demonstrate a quantitatively $o(1)$ level of Fourier decay.

- (3) Restriction estimate. The majorant ν satisfies a *restriction estimate at exponent u* if

$$\sup_{|\phi| \leq \nu} \int_{\mathbb{T}} |\hat{\phi}(\alpha)|^u d\alpha \ll_u \|\nu\|_1^u N^{-1}.$$

- (4) K -trivial saving. For $\eta > 0$, the majorant ν *saves η on K -trivial solutions* if

$$\sum_{\mathbf{x} \in K} \prod_{i=1}^s \nu(x_i) \ll_{k,s,d,\eta} \|\nu\|_1^s N^{-1-\eta}.$$

Of these, the most technically demanding are Fourier decay and the restriction estimate, especially the latter. For both, it is necessary to have good pointwise estimates for certain exponential sums over primes. These exponential sums are given by the Fourier transform of our majorant. To ensure the necessary Fourier decay, we use the W -trick [14], which circumvents technical difficulties arising from the fact that the prime d th powers are not equidistributed in congruence classes to small moduli.

The number of variables required in Theorem 1.1 is determined by the restriction estimate. Since we have good control on the growth of the weights involved, we shall see that the restriction estimate can be derived from a moment estimate for a simpler exponential sum. This leads to the following strengthening of Theorem 1.1.

Theorem 1.2. *Let $t \geq d$ be an integer such that the number of solutions $\mathbf{z} \in [X]^{2t}$ to*

$$z_1^d + \cdots + z_t^d = z_{t+1}^d + \cdots + z_{2t}^d \quad (1.10)$$

is $O_{t,d,\varepsilon}(X^{2t-d+\varepsilon})$, and assume $s > 2t$. Then we have (1.6).

To deduce Theorem 1.1, we apply Theorem 1.2 with the particular choice

$$t = \begin{cases} \lfloor d^2/2 \rfloor, & \text{if } d \neq 4 \\ 7, & \text{if } d = 4. \end{cases} \quad (1.11)$$

For this to be valid, we need to check that if t is given by (1.11) then the number of solutions $\mathbf{z} \in [X]^{2t}$ to (1.10) is indeed $O_{t,d,\varepsilon}(X^{2t-d+\varepsilon})$. For $(d, t) = (2, 2)$, it is known that (1.10) has $O(X^2 \log X)$ solutions $\mathbf{z} \in [X]^4$; the paper [3] discusses this and much more. For $(d, t) = (3, 4)$ it is known that there are $O(X^5)$ solutions — this follows, for instance, from the methods of [30].

For $d \geq 5$, one can show that if t is given by (1.11) then (1.10) has $O(X^{2t-d})$ solutions $\mathbf{z} \in [X]^{2t}$. This is a consequence of the main conjecture in Vinogradov's mean value theorem, which was recently established by Bourgain, Demeter and Guth [9]. Indeed, the equation (1.10) is a direct analogue of the equation

$$z_1^d + \cdots + z_{2t}^d = n \quad (n \text{ large and fixed}) \quad (1.12)$$

addressed in [35, Theorem 4.1]. That theorem, which was previously conditional on [9], tells us that (1.12) has $O(X^{2t-d})$ solutions $\mathbf{z} \in [X]^{2t}$, since

$$2t \geq d^2 - 1 \geq d^2 + 1 - \left\lfloor \frac{\log d}{\log 2} \right\rfloor.$$

One can follow the proof of [35, Theorem 4.1] verbatim, to show that (1.10) has $O(X^{2t-d})$ solutions $\mathbf{z} \in [X]^{2t}$. The point is that when $s \geq \frac{1}{2}d(d+1)$, the displayed equation in that proof tells us that we save $X^{d+1-\varepsilon}$ on the $(2s)$ th moment on minor arcs, which is far more than the $X^{d+\varepsilon}$ that we need to save; the interpolation procedure in [35, §3] then secures a saving of $X^{d+\varepsilon}$ on the $(2t)$ th moment.

For $(d, t) = (4, 7)$, we again follow the proof of [35, Theorem 4.1], interpolating on minor arcs between an eighth and a twentieth moment. Hua’s lemma [31, Lemma 2.5] yields the eighth moment bound $O(X^{5+\varepsilon})$. On minor arcs, the equation displayed in the proof of [35, Theorem 4.1] gives us the twentieth moment bound $O(X^{15+\varepsilon})$. The treatment of the major arcs is standard [31, §4.4], and we conclude that the number \mathcal{N} of solutions $\mathbf{z} \in [X]^{14}$ to (1.10) satisfies

$$\mathcal{N} \ll (X^{5+\varepsilon})^{1/2} (X^{15+\varepsilon})^{1/2} = X^{10+\varepsilon}.$$

Theorem 1.2 also enables a septenary result for cubes, assuming the so-called Hooley Riemann hypothesis (HRH); see [11, §6]. The statement below follows easily from Theorem 1.2 and [11, Lemma 6.2].

Corollary 1.3. *Assume HRH, $d = 3$ and $s \geq 7$. Then we have (1.6).*

We now comment on the relevance of restriction theory [6, 7, 17, 27]. The number of variables required to implement the circle method is often governed by the exponent at which we know a sharp moment estimate for an exponential sum. When the variables are restricted to lie in a set A , the relevant exponential sums necessarily come with weights supported on A . The key ingredient for such problems, therefore, is a moment estimate for an exponential sum with fairly arbitrary weights. Restriction theory concerns inequalities between norms of Fourier transforms, which is the same as bounding moments of weighted exponential sums.

Finally, we feel it is appropriate to describe the difficulties involved in proving the restriction estimate, and to outline our strategy for doing so. Using our hypothesis on t , step one is to deduce an ‘almost-sharp’ restriction estimate at exponent $2t$: this fails to be sharp by a factor of X^ε (see Lemma 5.3). If one could obtain a power saving on traditional minor arcs, for Weyl sums over primes, then a standard epsilon-removal process would complete the proof (see [6, §4]); the problem is that current technology only allows us to save a logarithmic factor here (see Lemma 3.1). Our approach is to introduce a related majorant μ , not involving primes. Step two is to establish a restriction inequality for μ at some intermediate exponent, sharp up to a logarithmic factor (see Lemma 5.2). Step three is to bootstrap this to the full-strength restriction estimate for our prime power majorant ν .

Note that for the specific value (1.11) of t , we know that the equation (1.10) has $O(X^{2t-d} \log X)$ solutions, at least when $d \neq 4$: recall the discussion following (1.11). Thus, when $d \neq 4$, the restriction estimate relevant to Theorem 1.1 is substantially easier to prove — one deals with the logarithmic factor directly using the methods in [6, §4]. Our more general approach reduces the number of variables required when $d = 4$, enables Corollary 1.3, and also anticipates future improvements in our understanding of the diophantine equation (1.10).

We organise thus. In §2, we construct our majorant ν , confirm (1.7), define our lifted set \mathcal{A} , and establish the density transfer inequality (1.9). In §3, we use the circle method to study the Fourier transform $\hat{\nu}$. The analysis therein allows us to establish Fourier decay in §4. In §5, we use Bourgain’s methods

[6] to prove that ν satisfies the relevant restriction estimate. We check in §6 that ν saves $1/s$ on K -trivial solutions, before putting it all together to prove Theorem 1.2 in §7.

We adopt the convention that ε denotes an arbitrarily small positive real number, so its value may differ between instances. The symbol p is reserved for primes. For $x \in \mathbb{R}$ and $q \in \mathbb{N}$, put $e(x) = e^{2\pi i x}$ and $e_q(x) = e^{2\pi i x/q}$. Boldface is used for vectors, for instance we abbreviate (x_1, \dots, x_n) to \mathbf{x} , and define $|\mathbf{x}| = \max(|x_1|, \dots, |x_n|)$. For $x \in \mathbb{R}$, let $\|x\|$ be the distance from x to the nearest integer. Let \mathcal{P} denote the set of primes. For $Y \in \mathbb{N}$, let $[Y] = \{1, 2, \dots, Y\}$ and $\mathcal{P}_Y = \mathcal{P} \cap [Y]$. We make use of the offset logarithmic integral $\text{Li}(x) = \int_2^x \frac{dt}{\log t}$.

We write \mathbb{T} for the torus \mathbb{R}/\mathbb{Z} . We use Landau and Vinogradov notation: for functions f and positive-valued functions g , write $f \ll g$ or $f = O(g)$ if there exists a constant C such that $|f(x)| \leq Cg(x)$ for all x . If S is a set, we denote the cardinality of S by $|S|$ or $\#S$. The pronumeral X denotes a positive integer, sufficiently large in terms of constants such as $d, c_1, \dots, c_s, k, \sigma, \sigma_0$ and ε . Moreover, we put $L = \log X$ throughout. We write C_1, C_2, \dots for positive constants that appear in the course of our proofs.

For $r \geq 1$ and $f : \mathbb{Z} \rightarrow \mathbb{C}$, we define the L^r -norm by

$$\|f\|_r = \left(\sum_n |f(n)|^r \right)^{1/r}.$$

When $\|f\|_1 < \infty$, we also define the Fourier transform of f by

$$\begin{aligned} \hat{f} : \mathbb{T} &\rightarrow \mathbb{C}, \\ \hat{f}(\alpha) &= \sum_n f(n)e(\alpha n). \end{aligned}$$

The author would like to thank his PhD advisor Trevor Wooley very much for his guidance. Thanks also to Tim Browning and Sean Prendiville for fruitful conversations. We are grateful to the anonymous referees for detailed reports containing several helpful suggestions.

2. THE W -TRICK

We begin by defining our majorant ν . As discussed, we apply the W -trick [14] from the outset, so that we later obtain sufficient Fourier decay. Let

$$w = \frac{1}{2} \log \log X, \quad W = 4d^3 \prod_{p \leq w} p. \quad (2.1)$$

Note the factor of d^3 included in the definition (2.1) of W ; this special feature comes into play during the case analysis in §4. Since X is large, it follows from the prime number theorem that

$$W \leq e^{2w} = \log X = L. \quad (2.2)$$

For $b \in [W]$ with

$$-b \in (\mathbb{Z}/W\mathbb{Z})^{\times d} := \{z^d : z \in (\mathbb{Z}/W\mathbb{Z})^\times\}, \quad (2.3)$$

let

$$\sigma(b) = \#\{z \in [W] : z^d \equiv -b \pmod{W}\}. \quad (2.4)$$

We begin with the observation that $\sigma(b)$ does not, in fact, depend on b . To verify this it suffices, by the Chinese remainder theorem, to show that if $r \in \mathbb{N}$ and $p^r \parallel W$ then

$$\#\{z \pmod{p^r} : z^d \equiv -b \pmod{p^r}\}$$

is the same for each b satisfying (2.3). If $p \neq 2$ then this follows easily using a primitive root (see [1, Ch. 10]). For $p = 2$, we can instead use the fact that any odd residue class is representable uniquely as $(-1)^u 5^v$, with $u \pmod{2}$ and $v \pmod{2^{r-2}}$ (see [13, Ch. 4]). We conclude that $\sigma(b)$ is the same for each $b \in -(\mathbb{Z}/W\mathbb{Z})^{\times d}$.

To ensure density transfer, we choose b to maximise the ν_b -measure of \mathcal{A}_b , where

$$\nu_b(n) = \begin{cases} \frac{\varphi(W)}{W\sigma(b)} dp^{d-1} \log p, & \text{if } Wn - b = p^d \text{ with } p \in \mathcal{P}_X \\ 0, & \text{otherwise} \end{cases} \quad (2.5)$$

and

$$\mathcal{A}_b = \{n \in \mathbb{Z} : Wn - b = p^d \text{ for some } p \in A\}. \quad (2.6)$$

Let

$$N = \lfloor X^d/W \rfloor + 1. \quad (2.7)$$

Lemma 2.1 (Density transfer). *Assume that $\delta \geq CX^{-1}$, for some large positive constant C . Then there exists $b \in [W]$ such that $-b \in (\mathbb{Z}/W\mathbb{Z})^{\times d}$ and*

$$\sum_{n \in \mathcal{A}_b} \nu_b(n) \gg \delta^d N. \quad (2.8)$$

Proof. We implicitly embed $-(\mathbb{Z}/W\mathbb{Z})^{\times d}$ into $[W]$, in the obvious way. We use a standard averaging argument, noting first that

$$\sum_{b \in -(\mathbb{Z}/W\mathbb{Z})^{\times d}} \sum_{n \in \mathcal{A}_b} \nu_b(n) = \sum_{p \in A} \frac{\varphi(W)}{W\sigma(b)} dp^{d-1} \log p - \sum_{p \in A, p \leq w} \frac{\varphi(W)}{W\sigma(b)} dp^{d-1} \log p.$$

By over-counting, and by recalling that $\sigma(b)$ is the same for each $b \in -(\mathbb{Z}/W\mathbb{Z})^{\times d}$, we deduce that

$$|-(\mathbb{Z}/W\mathbb{Z})^{\times d}| = \frac{\varphi(W)}{\sigma(b)}.$$

Thus, if b is chosen to maximise $\sum_{n \in \mathcal{A}_b} \nu_b(n)$, then

$$\sum_{n \in \mathcal{A}_b} \nu_b(n) \geq -1 + \sum_{p \in A} W^{-1} dp^{d-1} \log p,$$

since X is large compared to d .

A crude lower bound for $\sum_{p \in A} p^{d-1} \log p$ is given by the sum of $p^{d-1} \log p$ over the first $|A|$ primes p . By (1.8) and the prime number theorem, we now have

$$\sum_{p \in A} p^{d-1} \log p \geq \sum_{p \leq (1-\varepsilon)\delta X} p^{d-1} \log p \gg \delta^d X^d;$$

we have used the assumption that $\delta \geq CX^{-1}$. Hence

$$\sum_{n \in \mathcal{A}_b} \nu_b(n) \gg \delta^d X^d / W \gg \delta^d N.$$

□

The assumption that $\delta \geq CX^{-1}$ is harmless in the context of Theorem 1.2, for if $\delta < CX^{-1}$ then we certainly have (1.6). We do not expect that the density transfer inequality (2.8) is best possible, but sharpening the lower bound to δN would not affect the eventual outcome, so we make do with our crude inequality.

We henceforth fix b as in Lemma 2.1, and write

$$\mathcal{A} = \mathcal{A}_b, \quad \nu = \nu_b, \tag{2.9}$$

so that we have (1.9). Note that our majorant ν is supported on $[N]$, and that

$$\|\nu\|_\infty \ll X^{d-1} L. \tag{2.10}$$

Next, we verify (1.7). The proof is standard, but we nonetheless present it, as it will prepare us well for the next section.

We compute:

$$\begin{aligned} \|\nu\|_1 &= \sum_{\substack{p \leq X: \\ p^d \equiv -b \pmod{W}}} \frac{\varphi(W)}{W\sigma(b)} dp^{d-1} \log p \\ &= \sigma(b)^{-1} \sum_{\substack{z \in [W]: \\ z^d \equiv -b \pmod{W}}} \frac{\varphi(W)}{W} \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} dp^{d-1} \log p. \end{aligned} \tag{2.11}$$

The inner sum is treated using Abel summation. For $n \in [X]$, put

$$A_n = \sum_{\substack{p \leq n: \\ p \equiv z \pmod{W}}} 1.$$

Note that since $(b, W) = 1$ we must also have $(z, W) = 1$. The bound (2.2) allows us to apply Siegel–Walfisz [19, Lemma 7.14], so

$$A_n = \frac{\text{Li}(n)}{\varphi(W)} + O(Xe^{-C_1\sqrt{L}}).$$

With $g(n) = dn^{d-1} \log n$, we have

$$\begin{aligned} \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} dp^{d-1} \log p &= \sum_{n=2}^X (A_n - A_{n-1})g(n) \\ &= A_X g(X+1) + \sum_{n=2}^X A_n (g(n) - g(n+1)). \end{aligned}$$

The mean value theorem tells us that $g(n) - g(n+1) \ll X^{d-2}L$. In light of (2.2) and (2.7), we now have

$$\begin{aligned} \varphi(W) \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} dp^{d-1} \log p &= \text{Li}(X)g(X+1) + \sum_{n=2}^X \text{Li}(n) \cdot (g(n) - g(n+1)) \\ &\quad + O(Ne^{-C_2\sqrt{L}}). \end{aligned}$$

As $\text{Li}(2) = 0$, we therefore have

$$\begin{aligned} \varphi(W) \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} dp^{d-1} \log p &= \sum_{n=3}^X g(n) \int_{n-1}^n \frac{dx}{\log x} + O(Ne^{-C_2\sqrt{L}}) \\ &= \sum_{n=3}^X dn^{d-1} \int_{n-1}^n \frac{\log n}{\log x} dx + O(Ne^{-C_2\sqrt{L}}). \end{aligned}$$

When $2 \leq n-1 < x < n$, the mean value theorem tells us that

$$\log n = \log x + O(1/n).$$

Hence

$$\varphi(W) \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} dp^{d-1} \log p = \sum_{n \leq X} dn^{d-1} + O(Ne^{-C_2\sqrt{L}}) = X^d + O(Ne^{-C_2\sqrt{L}}).$$

Substituting this into (2.11), and recalling (2.4) and (2.7), yields

$$\|\nu\|_1 = X^d/W + O(Ne^{-C_2\sqrt{L}}) = N + O(Ne^{-C_2\sqrt{L}}),$$

confirming (1.7).

3. EXPONENTIAL SUMS

We wish to investigate

$$\begin{aligned} \hat{\nu}(\alpha) &= \frac{\varphi(W)}{W\sigma(b)} \sum_{\substack{p \leq X: \\ p^d \equiv -b \pmod{W}}} (dp^{d-1} \log p) e(\alpha(p^d + b)/W) \\ &= \frac{\varphi(W)e(\alpha b/W)}{W\sigma(b)} \sum_{\substack{z \in [W]: \\ z^d \equiv -b \pmod{W}}} \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} (dp^{d-1} \log p) e(\alpha p^d/W), \quad (3.1) \end{aligned}$$

so we focus on the inner sum. We begin with a Hardy–Littlewood dissection, dissecting \mathbb{T} into major arcs \mathfrak{M} and minor arcs \mathfrak{m} . Let σ_0 be a large positive constant, and let σ be a much larger positive constant. For $q \in \mathbb{N}$ and $a \in \mathbb{Z}$, let $\mathfrak{M}(q, a)$ be the set of $\alpha \in \mathbb{T}$ such that $|\alpha - a/q| \leq L^\sigma X^{-d}$. Let $\mathfrak{M}(q)$ be the union of the sets $\mathfrak{M}(q, a)$ over integers a such that $(a, q) = 1$, and let \mathfrak{M} be the union of the sets $\mathfrak{M}(q)$ over $q \leq L^\sigma$. Put $\mathfrak{m} = \mathbb{T} \setminus \mathfrak{M}$. By identifying \mathbb{T} with a unit interval, we may write

$$\mathfrak{M}(q) = \bigcup_{\substack{a=0 \\ (a,q)=1}}^{q-1} \mathfrak{M}(q, a).$$

Lemma 3.1. *If $\alpha \in \mathfrak{m}$ then $\hat{\nu}(\alpha) \ll NL^{-\sigma_0}$.*

Proof. Let $\alpha \in \mathfrak{m}$. By Dirichlet’s approximation theorem [31, Lemma 2.1], we obtain relatively prime integers q and a such that $1 \leq q \leq X^d L^{-\sigma}$ and $|q\alpha - a| \leq L^\sigma X^{-d}$. Now $|\alpha - a/q| \leq L^\sigma X^{-d}$ so, as $\alpha \notin \mathfrak{M}$, we must have $q > L^\sigma$. Thus, with $\beta = \alpha - a/q$, we have

$$|\beta| \leq \frac{L^\sigma}{qX^d} \leq X^{-d}.$$

Let $z \in [W]$ with $z^d \equiv -b \pmod{W}$. By partial summation, we have

$$\begin{aligned} \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} (dp^{d-1} \log p) e(\alpha p^d/W) &= \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} e(\beta p^d/W) (dp^{d-1} \log p) e_{Wq}(ap^d) \\ &= A_\diamond(X) f(X) - \int_1^X A_\diamond(t) f'(t) dt, \end{aligned}$$

where $f(t) = e(\beta t^d/W) dt^{d-1} \log t$ and

$$A_\diamond(t) = \sum_{\substack{p \leq t: \\ p \equiv z \pmod{W}}} e_{Wq}(ap^d).$$

Note that $f'(t) \ll X^{d-2}L$, and that $|A_\diamond(t)| \leq t$. Thus,

$$\begin{aligned} \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} (dp^{d-1} \log p) e(\alpha p^d/W) &= A_\diamond(X) f(X) - \int_{XL^{-2\sigma_0}}^X A_\diamond(t) f'(t) dt \\ &\quad + O(NL^{-\sigma_0}). \end{aligned}$$

In view of (2.4) and (3.1), we now have

$$\hat{\nu}(\alpha) \ll NL^{-\sigma_0} + X^{d-1}L \sup_{XL^{-2\sigma_0} < t \leq X} |A_\diamond(t)|.$$

It remains to estimate $A_\diamond(t)$ when $XL^{-2\sigma_0} < t \leq X$. We use [19, Theorem 10] for this. In order to apply this result, we need to control size of the denominator

$$q^* := \frac{Wq}{(a, Wq)} = \frac{Wq}{(a, W)},$$

in terms of t . Recalling (2.2), we have

$$L^\sigma < q \leq q^* \leq Wq \leq X^d L^{1-\sigma} \leq t^d L^{-\sigma/2}.$$

As $L \ll \log t \ll L$, we may thus invoke [19, Theorem 10], which tells us that

$$A_\diamond(t) \ll XL^{-\sigma_0-1}W^{-1}.$$

Hence $\hat{\nu}(\alpha) \ll NL^{-\sigma_0}$. \square

On major arcs we can decompose our Fourier transform into archimedean and non-archimedean components. When $(z, W) = 1$, let

$$S_q^*(a, z) = \sum_{\substack{r \bmod q: \\ (z+Wr, Wq)=1}} e_q\left(a \frac{(z+Wr)^d + b}{W}\right) \quad (3.2)$$

and

$$I(\beta) = \int_0^N e(\beta t) dt. \quad (3.3)$$

Lemma 3.2. *Let $\alpha \in \mathfrak{M}(q, a)$ with $(a, q) = 1$ and $q \leq L^\sigma$, and put*

$$\beta = \alpha - a/q \in [-L^\sigma X^{-d}, L^\sigma X^{-d}]. \quad (3.4)$$

Then

$$\hat{\nu}(\alpha) = I(\beta)\sigma(b)^{-1} \sum_{\substack{z \in [W]: \\ z^d \equiv -b \bmod W}} \frac{\varphi(W)}{\varphi(Wq)} S_q^*(a, z) + O(Ne^{-C_4\sqrt{L}}). \quad (3.5)$$

Proof. With (3.1) in mind, we initially fix $z \in [W]$ with $z^d \equiv -b \bmod W$, and study

$$\sum_{\substack{p \leq X: \\ p \equiv z \bmod W}} (dp^{d-1} \log p) e(\alpha p^d/W).$$

For $n \in [X]$, let

$$S_n = \sum_{\substack{p \leq n: \\ p \equiv z \bmod W}} e_{Wq}(ap^d).$$

Then

$$S_n = O(Wq) + \sum_{\substack{r \bmod q: \\ (z+Wr, Wq)=1}} e_{Wq}(a(z+Wr)^d) \sum_{\substack{p \leq n: \\ p \equiv z+Wr \bmod Wq}} 1.$$

As $n \leq X$ and $Wq \leq L^{\sigma+1}$, the inner sum is amenable to Siegel–Walfisz [19, Lemma 7.14], and so

$$S_n = \frac{\text{Li}(n)}{\varphi(Wq)} V_q(a, z) + O(Xe^{-C_3\sqrt{L}}),$$

where

$$V_q(a, z) = \sum_{\substack{r \bmod q: \\ (z+Wr, Wq)=1}} e_{Wq}(a(z+Wr)^d). \quad (3.6)$$

With $f(t) = e(\beta t^d/W) dt^{d-1} \log t$, we have

$$\begin{aligned} \sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} (dp^{d-1} \log p) e(\alpha p^d/W) &= \sum_{n=2}^X (S_n - S_{n-1}) f(n) \\ &= S_X f(X+1) + \sum_{n=2}^X S_n (f(n) - f(n+1)). \end{aligned}$$

As $|\beta| \leq L^\sigma X^{-d}$, the mean value theorem implies that

$$f(n) - f(n+1) \ll X^{d-2} L^{\sigma+1}.$$

Hence

$$\begin{aligned} &\sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} (dp^{d-1} \log p) e(\alpha p^d/W) \\ &= \frac{V_q(a, z)}{\varphi(Wq)} \left[\text{Li}(X) f(X+1) + \sum_{n=2}^X \text{Li}(n) \cdot (f(n) - f(n+1)) \right] + O(Ne^{-C_4\sqrt{L}}). \end{aligned}$$

As $\text{Li}(2) = 0$, we thus have

$$\begin{aligned} &\sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} (dp^{d-1} \log p) e(\alpha p^d/W) \\ &= \frac{V_q(a, z)}{\varphi(Wq)} \sum_{n=3}^X \int_{n-1}^n \frac{f(n)}{\log x} dx + O(Ne^{-C_4\sqrt{L}}). \end{aligned} \tag{3.7}$$

When $n-1 < x < n$, the mean value theorem reveals that

$$f(n) = f(x) + O(X^{d-2} L^{\sigma+1}),$$

and so

$$\begin{aligned} \sum_{n=3}^X \int_{n-1}^n \frac{f(n)}{\log x} dx &= \int_2^X dx^{d-1} e(\beta x^d/W) dx + O(X^{d-1} L^{\sigma+1}) \\ &= WI(\beta) + O(X^{d-1} L^{\sigma+1}). \end{aligned}$$

Substituting this into (3.7), and noting that $|V_q(a, z)| \leq q \leq L^\sigma$, gives

$$\sum_{\substack{p \leq X: \\ p \equiv z \pmod{W}}} (dp^{d-1} \log p) e(\alpha p^d/W) = \frac{W}{\varphi(Wq)} V_q(a, z) I(\beta) + O(Ne^{-C_4\sqrt{L}}).$$

Substituting this into (3.1), and recalling (2.4), gives

$$\hat{\nu}(\alpha) = \frac{\varphi(W) e(\alpha b/W)}{\varphi(Wq) \sigma(b)} \sum_{\substack{z \in [W]: \\ z^d \equiv -b \pmod{W}}} V_q(a, z) I(\beta) + O(Ne^{-C_4\sqrt{L}}).$$

From (3.2) and (3.6), we see that

$$e_{Wq}(ab) V_q(a, z) = S_q^*(a, z).$$

Since

$$e(\beta b/W)I(\beta) = \int_0^N e(\beta(t + b/W)) dt = I(\beta) + O(1),$$

we finally have (3.5). \square

4. FOURIER DECAY

In this section, we establish that ν has Fourier decay of level $O(w^{\varepsilon-1/2})$. In other words, we prove that if $\alpha \in \mathbb{T}$ then

$$\hat{\nu}(\alpha) - \widehat{1_{[N]}}(\alpha) \ll_{\varepsilon} w^{\varepsilon-1/2} N. \quad (4.1)$$

By a geometric series, we have

$$\widehat{1_{[N]}}(\alpha) = \sum_{n \leq N} e(\alpha n) \ll \|\alpha\|^{-1}. \quad (4.2)$$

First suppose $\alpha \in \mathfrak{m}$. By Dirichlet's approximation theorem, we obtain relatively prime integers q and a such that $1 \leq q \leq L^{\sigma}$ and $|q\alpha - a| \leq L^{-\sigma}$. As $\alpha \notin \mathfrak{M}$, we must have $|\alpha - a/q| > L^{\sigma} X^{-d}$, so

$$\widehat{1_{[N]}}(\alpha) \ll \|\alpha\|^{-1} \ll \frac{q}{\|q\alpha\|} \ll X^d L^{-\sigma}.$$

Recalling (2.2) and (2.7), we now have $\widehat{1_{[N]}}(\alpha) \ll NL^{1-\sigma}$. Coupling this with Lemma 3.1, using the triangle inequality, yields

$$\hat{\nu}(\alpha) - \widehat{1_{[N]}}(\alpha) \ll NL^{1-\sigma} + NL^{-\sigma_0}.$$

Upon recalling the definition (2.1) of w , we conclude that (4.1) holds for $\alpha \in \mathfrak{m}$.

Next we consider the case in which $q = 1$ and $\alpha \in \mathfrak{M}(q)$. In other words,

$$\|\alpha\| \leq L^{\sigma} X^{-d}.$$

From (2.4) and (3.5), we see that

$$\hat{\nu}(\alpha) = I(\alpha) + O(Ne^{-C_4\sqrt{L}}).$$

By Euler–Maclaurin summation [31, Eq. (4.8)], we have

$$\widehat{1_{[N]}}(\alpha) - I(\alpha) \ll 1 + N\|\alpha\| \ll 1 + NL^{\sigma} X^{-d} \ll L^{\sigma}.$$

The triangle inequality now gives

$$\hat{\nu}(\alpha) - \widehat{1_{[N]}}(\alpha) \ll Ne^{-C_4\sqrt{L}}.$$

Recalling (2.1), we conclude that (4.1) holds whenever $\alpha \in \mathfrak{M}(1)$.

Finally, let $\alpha \in \mathfrak{M}(q, a)$ with $2 \leq q \leq L^{\sigma}$ and $(a, q) = 1$, and put (3.4). Since $q \geq 2$, we must have $|a| \geq 1$. Substituting

$$\|\alpha\| \geq q^{-1} - |\beta| \geq q^{-1} - L^{\sigma} X^{-d} \gg q^{-1}$$

into (4.2) gives

$$\widehat{1_{[N]}}(\alpha) \ll q \ll L^{\sigma}. \quad (4.3)$$

By (2.4), (3.5) and the trivial estimate $|I(\beta)| \leq N$, we have

$$\hat{\nu}(\alpha) \ll N e^{-C_4 \sqrt{L}} + \frac{N}{\varphi(q)} \sup_z |S_q^*(a, z)|, \quad (4.4)$$

where the supremum is over $z \in [W]$ such that $(z, W) = 1$.

We now study the sums $S_q^*(a, z)$. Let $z \in [W]$ with $(z, W) = 1$. By (3.2), we have

$$S_q^*(a, z) = e_{Wq}(a(z^d + b)) S_q^\diamond(a, z), \quad (4.5)$$

where

$$S_q^\diamond(a, z) = \sum_{\substack{r \bmod q: \\ (z+Wr, Wq)=1}} e_q \left(a \sum_{\ell=1}^d \binom{d}{\ell} W^{\ell-1} z^{d-\ell} r^\ell \right).$$

As $(z + Wr, W) = (z, W) = 1$, we have the slightly simpler expression

$$S_q^\diamond(a, z) = \sum_{\substack{r \bmod q: \\ (z+Wr, q)=1}} e_q \left(a \sum_{\ell=1}^d \binom{d}{\ell} W^{\ell-1} z^{d-\ell} r^\ell \right).$$

Let $q = uv$, where u is w -smooth and $(v, W) = 1$. Since $(u, v) = 1$, a standard calculation reveals that

$$S_q^\diamond(a, z) = S_u^\diamond(a_1, z) S_v^\diamond(a_2, z), \quad (4.6)$$

where $a_1 = av^{-1} \in (\mathbb{Z}/u\mathbb{Z})^\times$ and $a_2 = au^{-1} \in (\mathbb{Z}/v\mathbb{Z})^\times$ (see [31, Lemma 2.10]). First consider

$$S_u^\diamond(a_1, z) = \sum_{\substack{r \bmod u: \\ (z+Wr, u)=1}} e_u \left(a_1 \sum_{\ell=1}^d \binom{d}{\ell} W^{\ell-1} z^{d-\ell} r^\ell \right).$$

As u is w -smooth and $(z, W) = 1$, the condition $(z + Wr, u) = 1$ is always met, and so

$$S_u^\diamond(a_1, z) = \sum_{r \bmod u} e_u \left(a_1 \sum_{\ell=1}^d \binom{d}{\ell} W^{\ell-1} z^{d-\ell} r^\ell \right). \quad (4.7)$$

We now borrow a strategy employed in [10, §5]. Let $h = (u, W)$, and put $u = hu'$ and $W = hW'$, noting that $(u', W') = 1$. Writing $r = r_1 + u'r_2$, with $r_1 \bmod u'$ and $r_2 \bmod h$, yields

$$\begin{aligned} S_u^\diamond(a_1, z) &= \sum_{\substack{r_1 \bmod u' \\ r_2 \bmod h}} e_{hu'} \left(a_1 \sum_{\ell=1}^d \binom{d}{\ell} (hW')^{\ell-1} z^{d-\ell} (r_1 + u'r_2)^\ell \right) \\ &= \sum_{r_1=0}^{u'-1} e_{hu'} \left(a_1 \sum_{\ell=1}^d \binom{d}{\ell} (hW')^{\ell-1} z^{d-\ell} r_1^\ell \right) \\ &\quad \sum_{r_2=0}^{h-1} e_h \left(a_1 \sum_{\ell=1}^d \binom{d}{\ell} (hW')^{\ell-1} z^{d-\ell} (u')^{\ell-1} r_2^\ell \right). \end{aligned}$$

The inner sum is

$$\sum_{r_2 \bmod h} e_h(da_1 z^{d-1} r_2),$$

which vanishes unless $h \mid da_1 z^{d-1}$. As $(h, a_1) = (h, z) = 1$, we conclude that

$$S_u^\diamond(a_1, z) = 0, \quad \text{if } (u, W) \nmid d, \quad (4.8)$$

while if $h \mid d$ then

$$S_u^\diamond(a_1, z) = h \sum_{r_1=0}^{u'-1} e_u \left(a_1 \sum_{\ell=1}^d \binom{d}{\ell} W^{\ell-1} z^{d-\ell} r_1^\ell \right). \quad (4.9)$$

Next consider

$$e_{Wv}(a_2 z^d) S_v^\diamond(a_2, z) = \sum_{\substack{r \bmod v: \\ (z+Wr, v)=1}} e_v \left(a_2 \frac{(z+Wr)^d}{W} \right).$$

As $(v, W) = 1$, we can change variables by $t = zW^{-1} + r \in \mathbb{Z}/v\mathbb{Z}$, which gives

$$e_{Wv}(a_2 z^d) S_v^\diamond(a_2, z) = \sum_{\substack{t \bmod v: \\ (t, v)=1}} e_v(a_2 W^{d-1} t^d).$$

Since $(a_2 W^{d-1}, v) = 1$, we may apply [19, Lemma 8.5], which tells us that

$$S_v^\diamond(a_2, z) \ll v^{1/2+\varepsilon} \ll q^{1/2+\varepsilon}. \quad (4.10)$$

As $q \geq 2$, we must have (i) $u \nmid d$, (ii) $1 \neq q \mid d$, or (iii) $u \mid d$ and $q > w$. **Case:** $u \nmid d$. Suppose for a contradiction that $(u, W) \mid d$. Then for all primes p we have

$$\min(\text{ord}_p(u), \text{ord}_p(W)) \leq \text{ord}_p(d).$$

Since $\text{ord}_p(W) > \text{ord}_p(d)$ whenever $p \leq w$, and since u is w -smooth, this tells us that $u \mid d$, contradicting this case. Hence $(u, W) \nmid d$, so by (4.8) we have $S_u^\diamond(a_1, z) = 0$. Therefore $S_q^*(a, z)$ vanishes, by (4.5) and (4.6). Now (3.5) gives

$$\hat{v}(\alpha) \ll N e^{-C_4 \sqrt{L}}. \quad (4.11)$$

Case: $1 \neq q \mid d$. In this case $v = 1$ and $q = u$. Further,

$$h = (u, W) = (q, W) = q,$$

since $q \mid d \mid W$. So $u' = 1$, and from (4.9) we see that $S_q^\diamond(a, z) = q$. By (4.5), we therefore have

$$\sum_{\substack{z \in [W]: \\ z^d \equiv -b \bmod W}} S_q^*(a, z) = q \sum_{\substack{z \in [W]: \\ z^d \equiv -b \bmod W}} e_{Wq}(a(z^d + b)).$$

We will find that this sum vanishes, which is the point of the W -trick. Write

$$z \equiv x + \frac{W}{d} y \bmod W$$

with $x \bmod W/d$ and $y \bmod d$. Note that

$$z^d \equiv \left(x + \frac{W}{d} y \right)^d \equiv x^d \bmod W,$$

in view of the definition (2.1) of W . Hence

$$q^{-1} \sum_{\substack{z \in [W]: \\ z^d \equiv -b \pmod{W}}} S_q^*(a, z) = \sum_{\substack{x \in [W/d]: \\ x^d \equiv -b \pmod{W}}} e_{Wq}(a(x^d + b)) \sum_{y \pmod{d}} e_q(ax^{d-1}y).$$

As $(b, W) = 1$ and $q \mid W$, we have $(x, q) = 1$. Since $q \neq 1$ and $(a, q) = 1$, and since $q \mid d$, we deduce that the inner sum vanishes. Therefore

$$\sum_{\substack{z \in [W]: \\ z^d \equiv -b \pmod{W}}} S_q^*(a, z) = 0.$$

Substituting this into (3.5) yields (4.11).

Case: $u \mid d$ and $q > w$. By (4.7), we have

$$|S_u^\diamond(a_1, z)| \leq u \leq d \ll 1.$$

Now (4.5), (4.6) and (4.10) give

$$S_q^*(a, z) \ll q^{1/2+\varepsilon}. \quad (4.12)$$

Substituting this into (4.4) yields

$$\hat{\nu}(\alpha) \ll N e^{-C_4 \sqrt{L}} + q^{2\varepsilon-1/2} N \ll w^{2\varepsilon-1/2} N.$$

Recalling (2.1), we see that we have

$$\hat{\nu}(\alpha) \ll w^{\varepsilon-1/2} N$$

in all three cases. Coupling this with (4.3) yields (4.1). We conclude that the majorant ν has Fourier decay of level $O(w^{\varepsilon-1/2})$.

Note that the inequality (4.12) is valid in all three cases. We record the following estimate for later use.

Lemma 4.1. *Let $\alpha \in \mathfrak{M}(q, a)$ with $1 \leq q \leq L^\sigma$ and $(a, q) = 1$. Then*

$$\hat{\nu}(\alpha) \ll q^{\varepsilon-1/2} \min\{N, |\alpha - a/q|^{-1}\} + N e^{-C_4 \sqrt{L}}. \quad (4.13)$$

Proof. Put (3.4). By (2.4) and (3.5), we have

$$\hat{\nu}(\alpha) \ll \varphi(q)^{-1} \sup_z |I(\beta) S_q^*(a, z)| + N e^{-C_4 \sqrt{L}}.$$

The integral (3.3) admits the standard estimate

$$I(\beta) \ll \min\{N, \|\beta\|^{-1}\} = \min\{N, |\alpha - a/q|^{-1}\},$$

so by (4.12) we have

$$\begin{aligned} \hat{\nu}(\alpha) &\ll \varphi(q)^{-1} q^{1/2+\varepsilon} \min\{N, |\alpha - a/q|^{-1}\} + N e^{-C_4 \sqrt{L}} \\ &\ll q^{2\varepsilon-1/2} \min\{N, |\alpha - a/q|^{-1}\} + N e^{-C_4 \sqrt{L}}. \end{aligned}$$

□

5. THE RESTRICTION ESTIMATE

Let $t \geq d$ be an integer such that the number of solutions $\mathbf{z} \in [X]^{2t}$ to (1.10) is $O_{t,d,\varepsilon}(X^{2t-d+\varepsilon})$. In this section, we show that ν satisfies a restriction estimate at any exponent $u > 2t$. The following lemma suffices, by (1.7).

Lemma 5.1. *Let $\phi : \mathbb{Z} \rightarrow \mathbb{C}$ with $|\phi| \leq \nu$, and let $u > 2t$ be a real number. Then*

$$\int_{\mathbb{T}} |\hat{\phi}(\alpha)|^u d\alpha \ll_u N^{u-1}.$$

We proceed in stages. The factor of X^ε in the assumed bound on the number of solutions to (1.10) is a formidable hurdle, as Lemma 3.1 fails to provide a power saving on minor arcs. We shall reduce this to a logarithmic factor, at some intermediate exponent v ; here v is a real number with

$$2t < v < u. \quad (5.1)$$

In order to obtain a power saving on minor arcs, we introduce thicker major arcs, and use them to study the auxiliary majorant

$$\mu(n) = \sigma(b)^{-1} \sum_{\substack{x \in [X]: \\ Wn-b=x^d}} dx^{d-1},$$

wherein we recall (2.4). The reader should rest assured that the Möbius function does not appear in this manuscript, and so μ is always defined as above. Observe that

$$\nu(n) \leq L \cdot \mu(n) \quad (n \in \mathbb{Z}).$$

We shall prove the following restriction estimate for μ , which will serve as a platform from which to attack Lemma 5.1.

Lemma 5.2. *Let $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ with $|\psi| \leq \mu$, and let $v > 2t$ be a real number. Then*

$$\int_{\mathbb{T}} |\hat{\psi}(\alpha)|^v d\alpha \ll_v N^{v-1} L^v.$$

Let us explain how this implies Lemma 5.1, following Bourgain's strategy [6, §4]. We apply Lemma 5.2 with $\psi = L^{-1}\phi$, and with v in the range (5.1), obtaining

$$\int_{\mathbb{T}} |\hat{\phi}(\alpha)|^v d\alpha \ll N^{v-1} L^{2v}. \quad (5.2)$$

In this section only, we denote by δ an arbitrary parameter in the range

$$0 < \delta < 1,$$

for consistency with previous literature. This is not to be confused with the density δ defined in (1.8).

Consider the large spectra

$$\mathcal{R}_\delta = \{\alpha \in \mathbb{T} : |\hat{\phi}(\alpha)| > \delta N\},$$

which may be regarded as level sets. Note from (1.7) that

$$\|\hat{\phi}\|_\infty \leq \|\phi\|_1 \leq \|\nu\|_1 < (1 + \varepsilon)N.$$

By a standard argument involving dyadic intervals, it suffices to show that if $\varepsilon_0 > 0$ then

$$\text{meas}(\mathcal{R}_\delta) \ll_{\varepsilon_0} \frac{1}{\delta^{v+\varepsilon_0} N} \quad (5.3)$$

(see the discussion surrounding [10, Lemma 6.3]).

In our quest to establish (5.3), we begin by noting that if $\delta \leq L^{-2v/\varepsilon_0}$ then by (5.2) we have

$$(\delta N)^v \text{meas}(\mathcal{R}_\delta) \leq \int_{\mathbb{T}} |\hat{\phi}(\alpha)|^v d\alpha \ll N^{v-1} L^{2v},$$

whereupon

$$\text{meas}(\mathcal{R}_\delta) \ll \frac{L^{2v}}{\delta^v N} \ll \frac{1}{\delta^{v+\varepsilon_0} N}.$$

Thus, we may assume that

$$L^{-2v/\varepsilon_0} < \delta < 1. \quad (5.4)$$

Let $\theta_1, \dots, \theta_R$ be N^{-1} -spaced points in \mathcal{R}_δ . As $v > 2t \geq 2d$, it suffices to show that

$$R \ll_{\varepsilon_0} \delta^{-2d-\varepsilon_0}. \quad (5.5)$$

Put

$$\gamma = d + \varepsilon_0/3. \quad (5.6)$$

Routinely, as in [6, §4] and [10, §6], we have

$$\delta^{2\gamma} N^\gamma R^2 \ll \sum_{1 \leq r, r' \leq R} |\hat{\nu}(\theta_r - \theta_{r'})|^\gamma. \quad (5.7)$$

This calculation may be found in [10, §6]; see also (5.18) and its subsequent derivation.

Recall our Hardy–Littlewood dissection from §3. In this dissection, we now specify that σ_0 is large in terms of ε_0 and u . Consider

$$\theta = \theta_r - \theta_{r'}$$

in the summand on the right hand side of (5.7). By Lemma 3.1, the contribution from $\theta \in \mathfrak{m}$ to the right hand side of (5.7) is $O(R^2 N^\gamma L^{-\sigma_0 \gamma})$, with σ_0 large. By (5.4), this is $o(\delta^{2\gamma} N^\gamma R^2)$. Hence

$$\delta^{2\gamma} N^\gamma R^2 \ll \sum_{\substack{1 \leq r, r' \leq R: \\ \theta \in \mathfrak{M}}} |\hat{\nu}(\theta_r - \theta_{r'})|^\gamma. \quad (5.8)$$

Let $Q = C_5 + \delta^{-5}$, with C_5 a large positive constant. By (4.13), the contribution to the right hand side of (5.8) from denominators $q > Q$ is bounded, up to a constant, by

$$R^2 N^\gamma (Q^{\varepsilon-\gamma/2} + e^{-C_4 \gamma \sqrt{L}}).$$

This is negligible compared to the left hand side of (5.8), by (5.4) and the fact that C_5 is large. We thus conclude from (4.13), (5.4) and (5.8) that

$$\delta^{2\gamma} R^2 \ll \sum_{q \leq Q} \sum_{\substack{a \bmod q \\ (a,q)=1}} \sum_{1 \leq r, r' \leq R} \frac{q^{\varepsilon-\gamma/2}}{(1 + N|\theta_r - \theta_{r'} - a/q|)^\gamma}.$$

Hence

$$\delta^{2\gamma} R^2 \ll \sum_{1 \leq r, r' \leq R} G(\theta_r - \theta_{r'}), \quad (5.9)$$

where

$$G(\alpha) = \sum_{q \leq Q} \sum_{a=0}^{q-1} \frac{q^{\varepsilon-\gamma/2}}{(1 + N|\sin(\alpha - a/q)|)^\gamma}.$$

The inequality (5.9) is very similar to [6, Eq. (4.16)], but with N^2 replaced by N . We have an additional factor of q^ε in the definition of $G(\alpha)$, and we save a γ th power in the denominator, whereas Bourgain saves only a $(\gamma/2)$ nd power. Bourgain's argument carries through, and we obtain (5.5).

We have shown that Lemma 5.2 implies Lemma 5.1. To prove Lemma 5.2, we begin by establishing the following ' ε -sharp' $(2t)$ th moment bound.

Lemma 5.3. *Let $\psi : \mathbb{Z} \rightarrow \mathbb{C}$ with $|\psi| \leq \mu$. Then*

$$\int_{\mathbb{T}} |\hat{\psi}(\alpha)|^{2t} d\alpha \ll N^{2t-1+\varepsilon}.$$

Proof. By orthogonality, we have

$$\begin{aligned} \int_{\mathbb{T}} |\hat{\psi}(\alpha)|^{2t} d\alpha &= \int_{\mathbb{T}} \sum_{\mathbf{n} \in \mathbb{Z}^{2t}} \psi(n_1) \cdots \psi(n_t) \overline{\psi(n_{t+1})} \cdots \overline{\psi(n_{2t})} \\ &\quad e(\alpha(n_1 + \cdots + n_t - n_{t+1} - \cdots - n_{2t})) d\alpha \\ &= \sum_{\substack{\mathbf{n}: \\ n_1 + \cdots + n_t = n_{t+1} + \cdots + n_{2t}}} \psi(n_1) \cdots \psi(n_t) \overline{\psi(n_{t+1})} \cdots \overline{\psi(n_{2t})}. \end{aligned}$$

The triangle inequality now gives

$$\begin{aligned} \int_{\mathbb{T}} |\hat{\psi}(\alpha)|^{2t} d\alpha &\leq \sum_{\substack{\mathbf{n}: \\ n_1 + \cdots + n_t = n_{t+1} + \cdots + n_{2t}}} \mu(n_1) \cdots \mu(n_{2t}) \\ &\ll X^{2t(d-1)} \sum_{\substack{\mathbf{z} \in [X]^{2t}: \\ z_1^d + \cdots + z_t^d = z_{t+1}^d + \cdots + z_{2t}^d}} 1. \end{aligned}$$

Our hypothesis on t now yields

$$\int_{\mathbb{T}} |\hat{\psi}(\alpha)|^{2t} d\alpha \ll X^{2t(d-1)} X^{2t-d+\varepsilon} = X^{2td-d+\varepsilon}$$

so, by (2.2) and (2.7), we finally have

$$\int_{\mathbb{T}} |\hat{\psi}(\alpha)|^{2t} d\alpha \ll (NL)^{2t-1} X^\varepsilon \ll N^{2t-1+\varepsilon}.$$

□

We now prove Lemma 5.2, again using Bourgain's strategy. We begin by obtaining pointwise estimates for $\hat{\mu}$. The triangle inequality gives

$$\begin{aligned} \sigma(b)\hat{\mu}(\theta) &= e(\theta b/W) \sum_{\substack{x \leq X: \\ x^d \equiv -b \pmod{W}}} dx^{d-1} e(\theta x^d/W) \\ &\ll \left| \sum_{\substack{x \leq X: \\ x^d \equiv -b \pmod{W}}} x^{d-1} e(\theta x^d/W) \right|. \end{aligned}$$

Hence, by partial summation, we obtain

$$\hat{\mu}(\theta) \ll X^{d/2} + X^{d-1} \sup_{X^{1/2} \leq P \leq X} |g(\theta; P)|, \quad (5.10)$$

where

$$g(\theta; P) = \sigma(b)^{-1} \sum_{\substack{x \leq P: \\ x^d \equiv -b \pmod{W}}} e(\theta x^d/W).$$

Moreover, by (2.4), we have

$$g(\theta; P) \ll \sup_{z \in [W]} \left| \sum_{\substack{x \leq P: \\ x \equiv z \pmod{W}}} e(\theta x^d/W) \right|. \quad (5.11)$$

Writing $x = Wy + z$, we find that

$$\sum_{\substack{x \leq P: \\ x \equiv z \pmod{W}}} e(\theta x^d/W) = \sum_{y \leq P/W} e(W^{d-1}\theta h(y)) + O(1), \quad (5.12)$$

for some monic polynomial h of degree d .

Let $P \in [X^{1/2}, X]$ and $z \in [W]$. The Weyl sums

$$g_1(\alpha) := \sum_{y \leq P/W} e(\alpha h(y))$$

are very classical, and are discussed in many texts. For reasons of economy, we employ Baker's estimates [2], as packaged in [12, §2]. The bounds apply to monic polynomials h of degree d , and are uniform in the other coefficients of h ; in particular, they are uniform in z . It is plain from the proof of [12, Lemma 2.3] that the quantity $\sigma(d)$ therein may be replaced by 2^{1-d} . We conclude thus.

Lemma 5.4. *If*

$$|g_1(\alpha)| > (P/W)^{1-2^{1-d}+\varepsilon}$$

then there exist relatively prime integers $r > 0$ and b such that

$$g_1(\alpha) \ll r^{\varepsilon-1/d} P W^{-1} (1 + (P/W)^d |\alpha - b/r|)^{-1/d}.$$

From Lemma 5.4, we deduce that if

$$|g_1(W^{d-1}\theta)| > X^{1-2^{1-d}+\varepsilon}$$

then there exist relatively prime integers $r > 0$ and b such that

$$g_1(W^{d-1}\theta) \ll r^{\varepsilon-1/d} X W^{-1} (1 + (X/W)^d |W^{d-1}\theta - b/r|)^{-1/d}.$$

In this case we can put

$$a = \frac{b}{(b, W^{d-1})}, \quad q = \frac{r W^{d-1}}{(b, W^{d-1})},$$

thus obtaining relatively prime integers $q > 0$ and a such that

$$g_1(W^{d-1}\theta) \ll X q^{\varepsilon-1/d} (1 + X^d W^{-1} |\theta - a/q|)^{-1/d}. \quad (5.13)$$

Write

$$\mathfrak{n} = \{\theta \in \mathbb{T} : |\hat{\mu}(\theta)| \leq X^{d-2^{-d}}\}. \quad (5.14)$$

In light of (2.2) and (2.7), we can collect (5.10), (5.11), (5.12) and (5.13) to obtain the following ‘major arc estimate’: if $\theta \in \mathbb{T} \setminus \mathfrak{n}$ then there exist relatively prime integers q and a such that $0 \leq a \leq q-1$ and

$$\hat{\mu}(\theta) \ll N L q^{\varepsilon-1/d} (1 + N |\theta - a/q|)^{-1/d}. \quad (5.15)$$

Now that we have made the necessary preparations, we complete the proof of Lemma 5.2. This parallels our proof that Lemma 5.1 follows from Lemma 5.2. Consider the large spectra

$$\mathcal{R}_\delta = \{\alpha \in \mathbb{T} : |\hat{\psi}(\alpha)| > \delta N L\},$$

noting from (2.2) and (2.7) the crude bound

$$\|\hat{\psi}\|_\infty \leq \|\psi\|_1 \leq \|\mu\|_1 \leq \sum_{x \leq X} dx^{d-1} \sim X^d \leq N L. \quad (5.16)$$

Similarly to before, it suffices to show that if $\varepsilon_0 > 0$ then we have

$$\text{meas}(\mathcal{R}_\delta) \ll_{\varepsilon_0} \frac{1}{\delta^{2t+\varepsilon_0} N}.$$

This time, we can use Lemma 5.3 to reduce consideration to δ in the range

$$N^{-\varepsilon} < \delta < 1, \quad (5.17)$$

wherein we recall our notational convention for ε .

Let $\theta_1, \dots, \theta_R$ be N^{-1} -spaced points in \mathcal{R}_δ . It now remains to show (5.5). Again with (5.6), we will find that

$$\delta^{2\gamma} N^\gamma L^\gamma R^2 \ll \sum_{1 \leq r, r' \leq R} |\hat{\mu}(\theta_r - \theta_{r'})|^\gamma. \quad (5.18)$$

We now verify this inequality by following the corresponding argument in the proof of [10, Lemma 6.3].

Let $a_n \in \mathbb{C}$ be such that $|a_n| \leq 1$ and $\psi(n) = a_n \mu(n)$, for $n \in [N]$. Furthermore, let $c_1, \dots, c_R \in \mathbb{C}$ be such that $|c_r| = 1$ and

$$c_r \hat{\psi}(\theta_r) = |\hat{\psi}(\theta_r)| \quad (1 \leq r \leq R).$$

It follows from the Cauchy–Schwarz inequality and (5.16) that

$$\begin{aligned} \delta^2 N^2 L^2 R^2 &\leq \left(\sum_{r \in [R]} |\hat{\psi}(\theta_r)| \right)^2 = \left(\sum_{r \in [R]} c_r \sum_n a_n \mu(n) e(n\theta_r) \right)^2 \\ &\leq \|\mu\|_1 \sum_n \mu(n) \left| \sum_{r \in [R]} c_r e(n\theta_r) \right|^2 \ll NL \sum_n \mu(n) \left| \sum_{r \in [R]} c_r e(n\theta_r) \right|^2, \end{aligned}$$

and so

$$\delta^2 NLR^2 \ll \sum_{1 \leq r, r' \leq R} |\hat{\mu}(\theta_r - \theta_{r'})|.$$

An application of Hölder’s inequality now harvests (5.18).

Consider $\theta = \theta_r - \theta_{r'}$ in the summand on the right hand side of (5.18). By (2.2), (2.7) and (5.14), the contribution from $\theta \in \mathfrak{n}$ to the right hand side of (5.18) is $O(R^2 N^{\gamma(1+\varepsilon-2^{-d}/d)})$. By (5.17), this is $o(\delta^{2\gamma} N^\gamma L^\gamma R^2)$. Hence

$$\delta^{2\gamma} N^\gamma L^\gamma R^2 \ll \sum_{\substack{1 \leq r, r' \leq R: \\ \theta \notin \mathfrak{n}}} |\hat{\mu}(\theta_r - \theta_{r'})|^\gamma. \quad (5.19)$$

Let $Q = C_6 + \delta^{-3d}$, with C_6 a large positive constant. By (5.15), the contribution to the right hand side of (5.19) from denominators $q > Q$ is bounded, up to a constant, by

$$R^2 N^\gamma L^\gamma Q^{\varepsilon-\gamma/d}.$$

This is negligible compared to the left hand side of (5.19), as C_6 is large. We thus conclude from (5.15) and (5.19) that

$$\delta^{2\gamma} R^2 \ll \sum_{q \leq Q} \sum_{\substack{a \bmod q \\ (a,q)=1}} \sum_{1 \leq r, r' \leq R} \frac{q^{\varepsilon-\gamma/d}}{(1 + N|\theta_r - \theta_{r'} - a/q|)^{\gamma/d}}.$$

Hence

$$\delta^{2\gamma} R^2 \ll \sum_{1 \leq r, r' \leq R} G_2(\theta_r - \theta_{r'}), \quad (5.20)$$

where

$$G_2(\alpha) = \sum_{q \leq Q} \sum_{a=0}^{q-1} \frac{q^{\varepsilon-\gamma/d}}{(1 + N|\sin(\alpha - a/q)|)^{\gamma/d}}.$$

The inequality (5.20) is very similar to [6, Eq. (4.16)], but with N^2 replaced by N , and with $Q \sim \delta^{-3d}$ rather than $Q \sim \delta^{-5}$. The exponents differ but, since $\gamma > d$, Bourgain’s argument carries through, and provides the desired bound (5.5) for R . This completes the proof of Lemma 5.2. We have established all of the results in this section. In particular, we know from Lemma 5.1 that ν satisfies a restriction estimate at any exponent $u > 2t$.

6. THE K -TRIVIAL COUNT

In this section we show that ν saves $1/s$ on K -trivial solutions. Let $t \in \mathbb{N}$ be such that the number of solutions $\mathbf{z} \in [X]^{2t}$ to (1.10) is $O_{t,d,\varepsilon}(X^{2t-d+\varepsilon})$, and assume $s > 2t$.

Lemma 6.1. *The number of $\mathbf{x} \in [X]^s$ with $(x_1^d, \dots, x_s^d) \in K$ is*

$$O_{k,s,d,\varepsilon}(X^{s-d-d/(s-1)+\varepsilon}).$$

Proof. The set K lies in the union of k subspaces of the form

$$\{\mathbf{y} \in \mathbb{Q}^s : \mathbf{c} \cdot \mathbf{y} = \mathbf{d} \cdot \mathbf{y} = 0\},$$

where $\mathbf{d} \in \mathbb{Q}^s$ is a fixed vector that is not proportional to \mathbf{c} . Our task, therefore, is to count solutions $\mathbf{x} \in [X]^s$ to the system

$$c_1 x_1^d + \dots + c_s x_s^d = d_1 x_1^d + \dots + d_s x_s^d = 0. \quad (6.1)$$

From (6.1) we obtain

$$e_1 x_1^d + \dots + e_{s-1} x_{s-1}^d = 0,$$

where $(e_1, \dots, e_{s-1}) \neq \mathbf{0}$, and by rescaling we may assume that

$$(e_1, \dots, e_{s-1}) \in \mathbb{Z}^{s-1}.$$

Let u be the number of nonzero e_i , and note that $1 \leq u \leq s-1$. Without loss of generality $e_1, \dots, e_u \in \mathbb{Z} \setminus \{0\}$ and $e_{u+1} = \dots = e_{s-1} = 0$, so that

$$e_1 x_1^d + \dots + e_u x_u^d = 0. \quad (6.2)$$

By orthogonality, the number \mathcal{M} of solutions $(x_1, \dots, x_u) \in [X]^u$ to (6.2) is

$$\int_{\mathbb{T}} \sum_{\mathbf{y} \in [X]^u} e\left(\alpha \sum_{i \leq u} e_i y_i^d\right) d\alpha = \int_{\mathbb{T}} \left(\prod_{i \leq u} \sum_{y_i \leq X} e(\alpha e_i y_i^d) \right) d\alpha.$$

By Hölder's inequality we now have, for some $i \in [u]$,

$$\mathcal{M} \ll \int_{\mathbb{T}} |f(e_i \alpha)|^u d\alpha,$$

where

$$f(\theta) = \sum_{x \leq X} e(\theta x^d).$$

Note that $e_i \neq 0$, since $i \in [u]$. By periodicity, a change of variables reveals that

$$\int_{\mathbb{T}} |f(e_i \alpha)|^u d\alpha = \int_{\mathbb{T}} |f(\alpha)|^u d\alpha.$$

Another application of Hölder's inequality now gives

$$\mathcal{M} \ll \int_{\mathbb{T}} |f(\alpha)|^u d\alpha \ll \left(\int_{\mathbb{T}} |f(\alpha)|^{s-1} d\alpha \right)^{u/(s-1)}.$$

As $s-1 \geq 2t$ and $|f(\alpha)| \leq X$, we now have

$$\mathcal{M} \ll \left(X^{s-1-2t} \int_{\mathbb{T}} |f(\alpha)|^{2t} d\alpha \right)^{u/(s-1)}. \quad (6.3)$$

From (6.1) we have

$$c_{u+1}x_{u+1}^d + \cdots + c_s x_s^d = -(c_1 x_1^d + \cdots + c_u x_u^d). \quad (6.4)$$

Given integers x_1, \dots, x_u , the number of solutions $(x_{u+1}, \dots, x_s) \in \mathbb{Z}^{s-u}$ to (6.4) is, by orthogonality, at most

$$\int_{\mathbb{T}} \left| \sum_{\mathbf{z} \in [X]^{s-u}} e\left(\alpha \sum_{i \leq s-u} c_{u+i} z_i^d\right) \right| d\alpha.$$

By following our calculation bounding \mathcal{M} , we deduce that this quantity is bounded by

$$\left(X^{s-1-2t} \int_{\mathbb{T}} |f(\alpha)|^{2t} d\alpha \right)^{(s-u)/(s-1)}.$$

Coupling this information with (6.3), we find that the number \mathcal{N} of solutions $\mathbf{x} \in [X]^s$ to (6.1) satisfies

$$\mathcal{N} \ll \left(X^{s-1-2t} \int_{\mathbb{T}} |f(\alpha)|^{2t} d\alpha \right)^{s/(s-1)}.$$

By orthogonality, the integral $\int_{\mathbb{T}} |f(\alpha)|^{2t} d\alpha$ equals the number of solutions $\mathbf{z} \in [X]^{2t}$ to (1.10) which, by hypothesis, is $O(X^{2t-d+\varepsilon})$. Hence

$$\mathcal{N} \ll (X^{s-1-2t} X^{2t-d+\varepsilon})^{s/(s-1)} \ll X^{(s-1-d)s/(s-1)+2\varepsilon} = X^{s-d-d/(s-1)+2\varepsilon},$$

which proves the lemma. \square

Corollary 6.2. *The majorant ν saves $1/s$ on K -trivial solutions.*

Proof. By (1.7), our task is to establish the inequality

$$\sum_{\mathbf{y} \in K} \prod_{i=1}^s \nu(y_i) \ll_{k,s,d} N^{s-1-1/s}.$$

By (2.5) and (2.10), we have

$$\sum_{\mathbf{y} \in K} \prod_{i=1}^s \nu(y_i) \ll (X^{d-1}L)^s \sum_{\substack{\mathbf{x} \in [X]^s: \\ (x_1^d+b, \dots, x_s^d+b)/W \in K}} 1.$$

By the definition of K , the condition $(x_1^d+b, \dots, x_s^d+b)/W \in K$ is equivalent to the condition

$$(x_1^d, \dots, x_s^d) \in K.$$

Now Lemma 6.1 yields

$$\sum_{\mathbf{y} \in K} \prod_{i=1}^s \nu(y_i) \ll (X^{d-1}L)^s X^{s-d-d/(s-1)+\varepsilon} \ll X^{d(s-1)-d/(s-1)+2\varepsilon},$$

so by (2.2) and (2.7) we have

$$\sum_{\mathbf{y} \in K} \prod_{i=1}^s \nu(y_i) \ll N^{s-1-1/(s-1)+\varepsilon} \ll N^{s-1-1/s}.$$

\square

7. THE DENSITY BOUND

Finally, we have all of the ingredients needed to prove Theorem 1.2. Recall (2.9). Note that \mathcal{A} has only K -trivial solutions to (1.3), in the sense that if $\mathbf{n} \in \mathcal{A}^s$ and $\mathbf{c} \cdot \mathbf{n} = 0$ then $\mathbf{n} \in K$. Indeed, suppose $n_1, \dots, n_s \in \mathcal{A}$ and $\mathbf{c} \cdot \mathbf{n} = 0$. Then, by (1.2) and (2.6), we have (1.5) with

$$x_i = (Wn_i - b)^{1/d} \in A \quad (1 \leq i \leq s).$$

Our hypothesis on A then tells us that

$$(Wn_1 - b, \dots, Wn_s - b) = (x_1^d, \dots, x_s^d) \in K.$$

From its construction, we see that K is invariant under translations and dilations, so we now have $\mathbf{n} \in K$, which confirms that \mathcal{A} has only K -trivial solutions to (1.3).

We apply [10, Proposition 2.8] to the majorant ν , noting that $\mathcal{A} \subseteq \text{supp}(\nu)$. We showed in §4 that ν has Fourier decay of level $O(w^{\varepsilon-1/2})$. We showed in §5 that ν satisfies a restriction estimate at the exponent $s - 1/2$. We showed in §6 that ν saves $1/s$ on trivial solutions. Hence

$$\sum_{n \in \mathcal{A}} \nu(n) \ll \frac{N}{\min\{\log \log(w^{1/2-\varepsilon}), \log N\}^{s-2-\varepsilon}},$$

so by (2.1) we have

$$\sum_{n \in \mathcal{A}} \nu(n) \ll N(\log \log \log \log X)^{2+\varepsilon-s}.$$

Coupling this with (1.9) yields

$$\delta^d N \ll N(\log \log \log \log X)^{2+\varepsilon-s},$$

and so $\delta \ll (\log \log \log \log X)^{\frac{2-s}{d}+\varepsilon}$. By (1.8), this gives (1.6), completing the proof of Theorem 1.2.

REFERENCES

- [1] T. Apostol, *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Heidelberg, 1976.
- [2] R. C. Baker, *Diophantine Inequalities*, London Math. Soc. Monographs (N.S.) **1**, Clarendon Press, Oxford, 1986.
- [3] V. Blomer and A. Granville, *Estimates for representation numbers of quadratic forms*, Duke Math. J. **135** (2006), 261–302.
- [4] T. F. Bloom, *Translation invariant equations and the method of Sanders*, Bull. Lond. Math. Soc. **44** (2012), 1050–1067.
- [5] T. F. Bloom, *A quantitative improvement for Roth’s theorem on arithmetic progressions*, J. London Math. Soc. (2) **93** (2016), 643–663.
- [6] J. Bourgain, *On $\Lambda(p)$ -subsets of squares*, Israel J. Math. **67** (1989), 291–311.
- [7] J. Bourgain, *Fourier transform restriction phenomena for certain lattice subsets and applications to nonlinear evolution equations, I. Schrödinger equations*, Geom. Funct. Anal. **3** (1993), 107–156.
- [8] J. Bourgain, *On the Vinogradov mean value*, arXiv:1601.08173.

- [9] J. Bourgain, C. Demeter and L. Guth, *Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three*, Ann. of Math. (2) **184** (2016), 633–682.
- [10] T. D. Browning and S. M. Prendiville, *A transference approach to a Roth-type theorem in the squares*, Int. Math. Res. Not. **2016**, 1–30.
- [11] J. Brüdern and T. D. Wooley, *Subconvexity for additive equations: pairs of undenary cubic forms*, J. Reine Angew. Math. **696** (2014), 31–67.
- [12] S. Chow, *Waring's problem with shifts*, Mathematika **62** (2016), 13–46.
- [13] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000.
- [14] B. J. Green, *Roth's theorem in the primes*, Ann. of Math. (2) **161** (2005), 1609–1636.
- [15] H. A. Helfgott, *The ternary Goldbach conjecture is true*, arXiv:1312.7748.
- [16] K. Henriot, *Logarithmic bounds for translation-invariant equations in squares*, Int. Math. Res. Not. **2015**, 12540–12562.
- [17] K. Henriot, *Additive equations in dense variables via truncated restriction estimates*, arXiv:1508.05923.
- [18] L.-K. Hua, *On the representation of numbers as the sums of the powers of primes*, Math. Z. **44** (1939), 335–346.
- [19] L.-K. Hua, *Additive Theory of Prime Numbers*, Transl. Math. Monogr. **13**, Amer. Math. Soc., Providence, 1965.
- [20] K. Kawada and T. D. Wooley, *On the Waring–Goldbach problem for fourth and fifth powers*, Proc. London Math. Soc. (3) **83** (2001), 1–50.
- [21] E. Keil, *On a diagonal quadric in dense variables*, Glasg. Math. J. **56** (2014), 601–628.
- [22] A. V. Kumchev and T. D. Wooley, *On the Waring–Goldbach problem for eighth and higher powers*, J. London Math. Soc. (2) **93** (2016), 811–824.
- [23] J. Y. Liu and T. Zhan, *New Developments in the Additive Theory of Prime Numbers*, World Scientific, Singapore, to appear.
- [24] S. M. Prendiville, *Four variants of the Fourier-analytic transference principle*, arXiv:1509.09200.
- [25] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.
- [26] M. L. Smith, *On solution-free sets for simultaneous quadratic and linear equations*, J. Lond. Math. Soc. (2) **79** (2009), 273–293.
- [27] T. Tao, *Some recent progress on the restriction conjecture*, Fourier analysis and convexity, Birkhäuser Boston, Boston, MA, 2004, pp. 217–243.
- [28] K. Thanigasalam, *Improvement on Davenport's iterative method and new results in additive number theory, III*, Acta Arith. **48** (1987), 97–116.
- [29] K. Thanigasalam, *On sums of positive integral powers and simple proof of $G(6) \leq 31$* , Bull. Calcutta Math. Soc. **81** (1989), 279–294.
- [30] R. C. Vaughan, *On Waring's problem for cubes*, J. Reine Angew. Math. **365** (1986), 122–170.
- [31] R. C. Vaughan, *The Hardy–Littlewood method*, 2nd edition, Cambridge University Press, Cambridge, 1997.
- [32] R. C. Vaughan and T. D. Wooley, *Waring's problem: a survey*, Number theory for the millennium III, 301–340, A. K. Peters, Natick, MA, 2002.
- [33] I. M. Vinogradov, *Representation of an odd number as the sum of three primes*, Dokl. Akad. Nauk SSSR **15** (1937), 291–294 (in Russian).
- [34] E. Waring, *Meditationes Arithmeticae*, 2nd edition, Archdeacon, Cambridge, 1770.
- [35] T. D. Wooley, *The asymptotic formula in Waring's problem*, Int. Math. Res. Not. **2012**, 1485–1504.
- [36] T. D. Wooley, *Multigrade efficient congruencing and Vinogradov's mean value theorem*, Proc. Lond. Math. Soc. (3) **111** (2015), 519–560.
- [37] T. D. Wooley, *On Waring's problem for intermediate powers*, Acta Arith. **176** (2016), 241–247.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF YORK, HESLINGTON, YORK, YO10
5DD, UNITED KINGDOM

E-mail address: `sam.chow@york.ac.uk`