

RESEARCH ARTICLE

# Squarefree Integers in Arithmetic Progressions to Smooth Moduli

Alexander P. Mangerel 

Centre de Recherches Mathématiques, Université de Montréal, Pavillon André-Aisenstadt, 2920 Chemin de la Tour, Montréal, Québec H3T 1J4, Canada

Department of Mathematical Sciences, Durham University, Upper Mountjoy Campus, Stockton Road, Durham DH1 3LE;

E-mail: [smangerel@gmail.com](mailto:smangerel@gmail.com).

**Received:** 7 September 2020; **Revised:** 13 September 2021; **Accepted:** 2 October 2021

**2020 Mathematics Subject Classification:** *Primary* – 11N25; *Secondary* – 11N37, 11N69

**Abstract**

Let  $\varepsilon > 0$  be sufficiently small and let  $0 < \eta < 1/522$ . We show that if  $X$  is large enough in terms of  $\varepsilon$ , then for any squarefree integer  $q \leq X^{196/261-\varepsilon}$  that is  $X^\eta$ -smooth one can obtain an asymptotic formula with power-saving error term for the number of squarefree integers in an arithmetic progression  $a \pmod{q}$ , with  $(a, q) = 1$ . In the case of squarefree, smooth moduli this improves upon previous work of Nunes, in which  $196/261 = 0.75096\dots$  was replaced by  $25/36 = 0.69\bar{4}$ . This also establishes a level of distribution for a positive density set of moduli that improves upon a result of Hooley. We show more generally that one can break the  $X^{3/4}$ -barrier for a density 1 set of  $X^\eta$ -smooth moduli  $q$  (without the squarefree condition).

Our proof appeals to the  $q$ -analogue of the van der Corput method of exponential sums, due to Heath-Brown, to reduce the task to estimating correlations of certain Kloosterman-type complete exponential sums modulo prime powers. In the prime case we obtain a power-saving bound via a cohomological treatment of these complete sums, while in the higher prime power case we establish savings of this kind using  $p$ -adic methods.

**Contents**

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Proof Strategy . . . . .	4
<b>2</b>	<b>Setting up the Key Estimate</b>	<b>6</b>
2.1	First reductions . . . . .	6
2.2	Bounding incomplete exponential sums on average . . . . .	9
<b>3</b>	<b>Correlations of <math>K_2</math> Sums to Prime Moduli: Cohomological Methods</b>	<b>15</b>
3.1	Cohomological Interpretation of $K_2$ and (14) . . . . .	16
3.2	Applying the trace formula . . . . .	17
<b>4</b>	<b>Correlations of <math>K_2</math>-sums to Prime Power Moduli: Stationary Phase Methods</b>	<b>19</b>
4.1	Preparation . . . . .	20
4.2	The $\epsilon = 0$ Terms . . . . .	21
4.3	Bounds for Large $n$ . . . . .	25
4.4	Bounds for Small $n$ and Large $p$ . . . . .	29
<b>5</b>	<b>Applying the <math>K_2</math> Correlations Bounds: Proof of Proposition 2.4</b>	<b>34</b>

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

**6 Proof of Theorems 1.1 and 1.5** **41**

6.1 First Result:  $q \leq X^{3/4-\varepsilon}$  42

6.2 Squarefree  $X^{3/4-\varepsilon} < q \leq X^{3/4+1/1044-\varepsilon}$  43

6.3 Non-squarefree  $q > X^{3/4-\varepsilon}$  45

**1. Introduction**

It is a classical problem in analytic number theory to study the distribution of well-known sequences in arithmetic progressions. The example of interest to us here is the sequence of squarefree integers; that is, integers  $n$  such that  $p^2 \nmid n$  for all primes  $p$ . Writing the indicator function for this sequence as  $\mu^2(n)$  for all  $n \in \mathbb{N}$ , where  $\mu$  denotes the Möbius function,<sup>1</sup> it is known that for a given modulus  $q$ , the asymptotic equidistribution estimate

$$\sum_{\substack{n \leq X \\ n \equiv a \pmod{q}}} \mu^2(n) = \frac{1}{\phi(q/(q, a))} \sum_{\substack{n \leq X \\ (n, q) = (a, q)}} \mu^2(n) + o(X/q) \tag{1}$$

for the count of squarefree integers in a progression  $a \pmod{q}$  holds if  $X$  is sufficiently large relative to  $q$ . It is a challenging question to determine the optimal constant  $\theta \in [0, 1)$  such that (1) holds as soon as  $q \leq X^\theta$ . It is widely believed that any  $\theta < 1$  should be admissible, but this is far from proven in general. It has been known for some time (see [19]) that any  $\theta < 2/3$  is admissible. At present, the best result that is available for all moduli  $q$  is that any  $\theta < 25/36 = 0.69\bar{4}$  is admissible, which is a recent result of Nunes [16].

If we impose constraints on the modulus  $q$ , one might expect that it is possible to improve this range of  $\theta$ . For instance, Hooley [8] showed that  $\theta$  can be taken in the interval  $[2/3, 3/4)$  provided that  $q$  has a moderately large<sup>2</sup> prime factor  $p$ , the size of which depends on  $\theta$ . As a consequence, he deduced that for a positive proportion of moduli  $q \leq X^{3/4-\varepsilon}$  the asymptotic (1) holds (the proportion depends on  $\varepsilon$  and tends toward 0 as  $\varepsilon \rightarrow 0^+$ ). Recently, Liu, Shparlinski and Zhang [14] improved upon this result by proving a Bombieri–Vinogradov-type theorem for squarefree integers, showing that the required asymptotic formula holds for all residue classes  $(a, q) = 1$  for almost all  $q \leq X^{3/4-\varepsilon}$ .

In this article, we shall improve upon Nunes’s result for a different, natural collection of moduli, specifically those  $q$  that are  $X^\eta$ -smooth (otherwise called  $X^\eta$ -friable); that is, such that all prime factors  $p$  of  $q$  satisfy  $p \leq X^\eta$  for  $\eta > 0$  small. For many such moduli, including all squarefree  $X^\eta$ -smooth moduli, we shall in fact be able to improve upon the range of admissibility  $\theta < 3/4$  from the works of Hooley and of Liu, Shparlinski and Zhang.

More precisely, our first main result is as follows.

**Theorem 1.1.** *Let  $\varepsilon > 0$  be small and let  $0 < \eta < 1/522$ . Let  $X$  be large in terms of  $\eta, \varepsilon$  and let  $q \leq X^{196/261-\varepsilon}$  be  $X^\eta$ -smooth and squarefree. Then there is a  $\delta > 0$ , depending only on  $\eta$  and  $\varepsilon$ , such that for any  $a \pmod{q}$  with  $(a, q) \leq X^\varepsilon$ ,*

$$\sum_{\substack{n \leq X \\ n \equiv a \pmod{q}}} \mu^2(n) = \frac{1}{\phi(q/(q, a))} \sum_{\substack{n \leq X \\ (n, q/(q, a)) = 1}} \mu^2(n) + O_\varepsilon\left(\frac{X^{1-\delta}}{q}\right).$$

**Remark 1.2.** Our proof actually shows that for  $q \leq X^{3/4-\varepsilon}$  one may take as the smoothness exponent any  $\eta \leq 6/25$ , whereas for  $X^{3/4-\varepsilon} < q \leq X^{196/261-\varepsilon}$  the admissible range is  $\eta < 1/522$ ; it is probable that this range in  $\eta$  may be improved somewhat, though it was not our primary objective to obtain an optimal such range.

<sup>1</sup>Recall that the Möbius function is the arithmetic function satisfying  $\mu(n) = 0$  if  $n$  is not squarefree and otherwise  $\mu(p_1 \cdots p_k) = (-1)^k$  if  $p_1, \dots, p_k$  are distinct primes.

<sup>2</sup>For instance, when  $\theta = 3/4 - \varepsilon$  for  $\varepsilon > 0$  small,  $p$  must have size  $X^{2/3-\varepsilon}$ .

Our result has the following trivial corollaries. The first improves upon the range<sup>3</sup> of moduli that is accessible in the density result of Hooley quoted earlier.

**Corollary 1.3.** *For any  $\varepsilon > 0$ , a positive proportion of the moduli  $q \leq X^{196/261-\varepsilon}$  satisfies*

$$\max_{(a,q)=1} \left| \sum_{\substack{n \leq X \\ n \equiv a \pmod{q}}} \mu^2(n) - \frac{1}{\phi(q)} \sum_{\substack{n \leq X \\ (n,q)=1}} \mu^2(n) \right| = o_\varepsilon(X/q).$$

**Corollary 1.4.** *Let  $\varepsilon > 0$  be small,  $0 < \eta < 1/522$  and let  $X$  be sufficiently large in terms of  $\eta$  and  $\varepsilon$ . Let  $q$  be squarefree and  $X^\eta$ -smooth and let  $a$  be a coprime residue class modulo  $q$ . Then provided  $X \geq q^{261/196+\varepsilon}$ , there is a squarefree integer in the set  $\{n \leq X : n \equiv a \pmod{q}\}$ .*

It is desirable to remove the assumption that  $q$  is squarefree, so that the above results continue to hold in some form for  $q$  merely  $X^\eta$ -smooth. While we are not able to derive this conclusion for all such moduli, we do show that almost all  $X^\eta$ -smooth moduli do have this property.

**Theorem 1.5.** *There is a  $\lambda > 0$  such that the following is true. Let  $\eta > 0$  and set*

$$\mathcal{Q}_\lambda(X) := \{q \leq X^{3/4+\lambda} : P^+(q) \leq X^\eta\}.$$

*Then for all but  $O_\eta(|\mathcal{Q}_\lambda(X)|/\log X)$  moduli  $q \in \mathcal{Q}_\lambda(X)$ , we have*

$$\max_{(a,q)=1} \left| \sum_{\substack{n \leq X \\ n \equiv a \pmod{q}}} \mu^2(n) - \frac{1}{\phi(q)} \sum_{\substack{n \leq X \\ (n,q)=1}} \mu^2(n) \right| = o_\eta(X/q).$$

Given  $y \geq 2$  we will say that a positive integer  $q$  is  $y$ -ultrasmooth if, whenever  $p^n || q$  we have  $p^n \leq y$ . What we will actually show is that the conclusion of Theorem 1.5 applies to all elements of

$$\mathcal{Q}'_\lambda(X) := \{q \in \mathcal{Q}_\lambda(X) : q \text{ is } X^\eta\text{-ultrasmooth}\}.$$

It is easy to see that by the union bound,

$$\begin{aligned} |\mathcal{Q}_\lambda(X) \setminus \mathcal{Q}'_\lambda(X)| &\leq \sum_{p \leq X^\eta} |\{q \in \mathcal{Q}_\lambda(X) : p^\nu || q, p^\nu > X^\eta\}| \leq \sum_{p \leq X^\eta} \sum_{q \leq X^{3/4+\lambda-\eta}} 1_{P^+(q) \leq X^\eta} \\ &\ll_\eta \frac{X^\eta}{\log X} |\mathcal{Q}_\lambda(X)| X^{-\eta} \ll \frac{|\mathcal{Q}_\lambda(X)|}{\log X}, \end{aligned}$$

which implies the claimed bound on the exceptional set when  $X \geq X_0(\eta)$ .

**Remark 1.6.** To be precise, there are two reasons why we do not treat all smooth moduli in this article. The first is that our method does not directly treat moduli  $q$  such that  $q$  has a large power of 2 or 3 as a divisor. It is likely that it could be modified to treat this case but at the cost of adding a nontrivial amount of pages to this already long article.

A more serious issue arises from the fact that in the course of the proof we need to be able to factor our modulus  $q$  (or, more precisely, a suitably chosen divisor of it) into coprime parts with well-controlled sizes (here we allow perturbations of size  $X^\eta$  only). If  $q$  were merely  $X^\eta$ -smooth but not  $X^\eta$ -ultrasmooth – for example, if  $q$  were divisible by a prime power  $p^\nu > X^{1/100}$ , say, with  $p \leq X^\eta$  – then this prime power would have to arise in one of the factors, biasing its size in a manner that might be incompatible with the bounds we obtain. The  $X^\eta$ -ultrasmoothness condition is in place to prevent such a bias in size from occurring.

<sup>3</sup>We emphasise that  $196/261 = 3/4 + 1/1044 > 3/4$ .

**Remark 1.7.** It is worth noting that our proof actually shows that we may asymptotically estimate the number of squarefree integers in a progression modulo  $q$  when  $q \leq X^{3/4-\varepsilon}$  which is  $X^\eta$ -smooth, without further assumptions (see Subsection 6.1 for a proof). However, this by itself is not sufficient to improve upon Hooley’s positive density result.

**1.1. Proof Strategy**

To prove Theorems 1.1 and 1.5, we will use a method of Heath-Brown that was used by Irving [9] to study the distribution of the divisor function in arithmetic progressions. To motivate our argument, it is useful to see where obstructions occur in the classical treatment of counting squarefree integers in progressions, as found in [19].

Assume in what follows that  $(a, q) = 1$ , for convenience. Using the classical identity

$$\mu^2(n) = \sum_{kl^2=n} \mu(l)$$

and decoupling the parameters  $k$  and  $l$  by localising them in short intervals (as we do in Section 2), it can be shown that the error term in (1) is controlled by averaged incomplete exponential sums<sup>4</sup>

$$\sum_{k \pmod q}^* \frac{1}{k} \left| \sum_{\substack{d \in I \\ (d,q)=1}} e_q \left( ka\bar{d}^2 \right) \right|, \tag{2}$$

where  $I$  is some interval of size  $< q$ . Using the completion method, one is led to bound averages of ‘quadratic Kloosterman sums’

$$K_2(A, B; q) := \sum_{x \pmod q}^* e_q \left( A\bar{x}^2 + Bx \right),$$

where  $A \in (\mathbb{Z}/q\mathbb{Z})^\times$  and  $B \in \mathbb{Z}/q\mathbb{Z}$ . Using the Chinese remainder theorem, we may, of course, factor  $K_2$  as a product of complete sums modulo prime powers  $p^n \parallel q$ . Each such factor can be estimated pointwise by  $O(p^{n/2})$ , using either:

- (i) a trivial application of the Bombieri–Dwork–Weil bound (see Lemma 2.2) when  $n = 1$  or
- (ii) the  $p$ -adic method of stationary phase for  $n \geq 2$  (see Lemma 2.1).

The resulting bound  $O_\varepsilon(X^\varepsilon q^{1/2})$  for the complete sum modulo  $q$  is of size  $o(X/q)$  as required, as long as  $q \leq X^{2/3-\varepsilon}$ .

One way to go beyond the  $X^{2/3}$  barrier is to try to exploit the averaging in  $k$  in (2), rather than employing a pointwise bound. Indeed, when  $q$  factors sufficiently nicely, the  $q$ -analogue of the van der Corput method of exponential sums, developed by Ringrose [21] and Heath-Brown [7, Theorem 2], enables one to reduce the problem above to estimating correlations of complete exponential sums (to be discussed momentarily). Using (a variant of) a result of Fouvry et al. [5] to estimate such correlations, Irving is able to treat  $X^\varepsilon$ -smooth and squarefree moduli of size  $q \leq X^{2/3+1/246-\varepsilon}$ . Aside from the fact that the sums  $K_2$  entering the picture differ in behaviour from genuine Kloosterman sums (so that the results of [5] do not apply), this strategy by itself is insufficient in the case of squarefree integers, even falling short of Nunes’s result (see Subsection 2.2.2 for further details).

To do better, we incorporate an additional idea. The key difficulty in understanding the distribution of squarefree integers in progressions is to estimate the count of points on the curve  $xy^2 \equiv a \pmod q$ , for  $x$  and  $y$  lying in certain ranges. Clearly, if  $I, J \subset \mathbb{Z}/q\mathbb{Z}$  are intervals and  $q'$  is any divisor of  $q$ , then

$$|\{(x, y) \in I \times J : xy^2 \equiv a \pmod q\}| \leq |\{(x, y) \in I \times J : xy^2 \equiv a \pmod{q'}\}|;$$

<sup>4</sup>Henceforth, given  $x \in \mathbb{R}$  and  $q \in \mathbb{N}$  we shall write  $e(x) := e^{2\pi ix}$  and  $e_q(x) := e(x/q)$ .

moreover, observe that the analogous bound for (2) with  $q'$  (of suitable size) in place of  $q$  – that is,  $O(X^\varepsilon(q')^{1/2})$  – can be of size  $o(X/q)$  for larger choices of  $q$  than  $X^{2/3-\varepsilon}$ . There is therefore an advantage in working with exponential sums modulo a suitably sized divisor of  $q$ , whenever such a divisor can be found; this was observed as well in [8], but its implementation differs from what we do here. As we shall show below, the fact that  $q$  is  $X^\eta$ -smooth means we can find divisors of  $q$  of any prescribed size (up to factors of size  $X^\eta$ ), in particular the size required to use the trick above. We therefore end up applying Irving’s method to treat Kloosterman-type sums modulo  $q'$  instead, which, when combined with the analysis of correlations described above, results in a gain not over the range  $q \leq X^{2/3-\varepsilon}$  but over a larger range of  $q$  instead.

Let us describe more precisely a key feature of our argument. A crucial step in Irving’s method (again for the divisor function) involves giving nontrivial estimates for the correlations

$$\sum_{b \in I} \text{Kl}(a, b + h_1; p) \cdots \text{Kl}(a, b + h_k; p),$$

where  $p \mid q, h_1, \dots, h_k \in \mathbb{F}_p, a \in \mathbb{F}_p^\times, I \subset \mathbb{F}_p$  is an interval and

$$\text{Kl}(A, B; p) := \sum_{x \in \mathbb{F}_p^\times} e_p(A\bar{x} + Bx) \text{ for } A, B \in \mathbb{F}_p$$

is the classical Kloosterman sum modulo  $p$ . In our circumstances, we treat  $K_2$  sums to both prime and prime power moduli, each case of which requires a separate analysis.

In the prime case, the corresponding correlation sum that we need to treat is of the form<sup>5</sup>

$$\sum_{b \in I} K_2(a, b + h_1; p) \cdots K_2(a, b + h_k; p) \overline{K}_2(a, b + h'_1; p) \cdots \overline{K}_2(a, b + h'_l; p),$$

where  $k+l \geq 1$ . We treat (completions of) such sums, which are the subject of Theorem 3.1, in Section 3 of this article, using cohomological methods, in particular, the sheaf-theoretic Fourier transform of Deligne.

Specifically, we view  $K_2$  as the  $\ell$ -adic Fourier transform of the trace function of an Artin–Schreier sheaf ( $\ell \neq p$  being an auxiliary prime), which is itself a trace function, pointwise pure of weight 1. Treating its correlations amounts to identifying cases in which tensor products of the underlying Galois representations are, or are not, geometrically trivial, a task facilitated by the Goursat–Kolchin–Ribet criterion (see [6] for an array of example applications of this method).

In the prime power case, our work is simplified (in terms of the required theoretical preliminaries but not in the amount of technical details) by the fact that these  $K_2$  sums can be explicitly computed using the  $p$ -adic stationary phase method. For instance, when  $n \geq 2$  and  $p > 3$  is prime, we have (see Lemma 2.1)

$$K_2(A, B; p^n) = p^{n/2} \left( \frac{3A}{p^n} \right) \varepsilon_{p,n} \sum_{\substack{y \pmod{p^{[n/2]}} \\ y^3 \equiv \overline{2AB} \pmod{p^{[n/2]}}} } e_{p^n}(3Ay^2),$$

for any  $A \in (\mathbb{Z}/p^n\mathbb{Z})^\times$  and  $B \in \mathbb{Z}/p^n\mathbb{Z}$ , where  $\varepsilon_{p,n} \in S^1$ . The correlation problem then revolves around bounding exponential sums over variables from a variety determined by the polynomials  $\{X^3 - \overline{2a}(b + h_i)\}_i$  and  $\{X^3 - \overline{2a}(b + h'_j)\}_j$ , as  $b$  varies.

An important work treating such correlations, with  $K_2$  sums again replaced by Kloosterman sums, was undertaken by Ricotta and Royer [20]. They used their estimates to establish distribution theorems for Kloosterman sums modulo  $p^n$ , as  $p \rightarrow \infty$ . However, their method is efficient mainly when  $n$  is fixed, as it relies on treating the correlation sum (via explicit formulae for Kloosterman sums to prime powers) as an exponential sum whose phase function turns out to be a polynomial modulo  $p^n$  of degree essentially as large as  $n$  and then applying the standard van der Corput–Weyl method. We also employ

<sup>5</sup>We emphasise that unlike classical Kloosterman sums, the  $K_2$  sums are not real-valued in general.

this strategy but instead use Vinogradov’s method (and the recent solution to the Vinogradov main conjecture, due to Bourgain–Demeter–Guth [2] and Wooley [25]) independently in place of van der Corput’s method, which leads to a stronger result when  $n$  is of bounded size as  $p$  grows.

In our circumstances we will also require a treatment that is effective for moduli  $p^n$  with rather large values of  $n$  and  $p$  possibly fixed. Fortunately, very recent work of Milićević and Zhang [15] introduced a method that suits this situation, in which one iteratively applies the stationary phase method to recover from the correlation sum a sum over a ‘generically trivial’ variety, up to small error (at least when  $n$  is large enough), rather than using Weyl sum estimates. A combination of the arguments<sup>6</sup> in each of these regimes will suit our needs.

## 2. Setting up the Key Estimate

### 2.1. First reductions

Let  $\varepsilon, \eta > 0$  be sufficiently small, let  $X$  be large relative to  $\varepsilon$  and  $\eta$  and let  $X^{2/3-\varepsilon} \leq q \leq X^{9/10-\varepsilon}$  be  $X^\eta$ -smooth. Given  $a$  a residue class modulo  $q$ , an arithmetic function  $g : \mathbb{N} \rightarrow \mathbb{C}$  and a set  $E \subset \mathbb{N}$ , define

$$\Delta_g(E; q, a) := \sum_{\substack{n \in E \\ n \equiv a \pmod{q}}} g(n) - \frac{1}{\phi(q/(a, q))} \sum_{\substack{n \in E \\ (n, q) = (a, q)}} g(n).$$

We shall also use the shorthand  $\Delta_g(X; q, a) := \Delta_g([1, X] \cap \mathbb{N}; q, a)$ . For the remainder of this section we will assume that  $(a, q) = 1$ .

Take  $g = \mu^2$ , the indicator function for the squarefree integers. Using the classical identity  $\mu^2(n) = \sum_{md^2=n} \mu(d)$  we obtain

$$\Delta_{\mu^2}(X; q, a) = \sum_{\substack{d \leq \sqrt{X} \\ (d, q) = 1}} \mu(d) \Delta_1(X/d^2; q, a\bar{d}^2),$$

where  $\bar{d}$  denotes the residue class modulo  $q$  with  $d\bar{d} \equiv 1 \pmod{q}$ . Let  $\delta \in (0, 1/20)$  and let

$$X^{\delta+\varepsilon} \leq V_0 \leq X^{1-\delta}/q$$

be a parameter to be chosen. For  $m \leq Y$  and  $1 \leq b \leq m$ , note the trivial bound

$$\Delta_1(Y; m, b) = \left\lfloor \frac{Y-b}{m} \right\rfloor - \frac{1}{\phi(m)} \left( \frac{\phi(m)}{m} Y + O(\tau(m)) \right) = O(1),$$

whence it follows that

$$\Delta_{\mu^2}(X; q, a) = \sum_{\substack{V_0 < d \leq \sqrt{X} \\ (d, q) = 1}} \mu(d) \Delta_1(X/d^2; q, a\bar{d}^2) + O(V_0).$$

It will be convenient in what follows later to remove the coefficient  $\mu(d)$ . Decomposing dyadically the sum in  $d$  and applying the triangle inequality, we find that there is a  $V_0 < V \leq \sqrt{X}$  such that

$$\begin{aligned} \Delta_{\mu^2}(X; q, a) &\ll (\log X) \left| \sum_{\substack{d \sim V \\ (d, q) = 1}} \mu(d) \Delta_1(X/d^2; q, a\bar{d}^2) \right| + V_0 \\ &\leq (\log X) \sum_{\substack{d \sim V \\ (d, q) = 1}} \left| \Delta_1(X/d^2; q, a\bar{d}^2) \right| + V_0. \end{aligned}$$

<sup>6</sup>Unfortunately, the bounds for prime power moduli  $p^n$  with  $n \geq 2$  are rather poorer than the bounds for prime moduli. This is due, in part, to the lack of rigid algebraic data. As a result of these weaker conclusions, we have chosen to leave the exponent  $\delta$  in Theorem 1.5, which is necessarily weaker than what is obtainable in Theorem 1.1, inexplicit.

Next, we further subdivide the range of  $m \leq X/d^2$  into dyadic subintervals, leading to the existence of  $U$  satisfying  $UV^2 \leq X$ , such that

$$\Delta_{\mu^2}(X; q, a) \ll (\log X)^2 \sum_{\substack{d \sim V \\ (d, q)=1}} \left| \sum_{\substack{m \sim U \\ md^2 \leq X \\ m \equiv a\bar{d}^2 \pmod{q}}} 1 - \frac{1}{\phi(q)} \sum_{\substack{m \sim U \\ md^2 \leq X \\ (m, q)=1}} 1 \right| + V_0.$$

To remove the condition  $md^2 \leq X$ , we split  $(U, 2U]$  and  $(V, 2V]$  into subintervals of respective lengths  $UV_0^{-1}$  and  $VV_0^{-1}$ , of which there are  $\ll V_0^2$  in total. We thus find that there are  $U_1 \in (U, 2U], V_1 \in (V, 2V]$  with  $U_1V_1^2 \leq 8X$  such that

$$\Delta_{\mu^2}(X; q, a) \ll V_0^2(\log X)^2 \sum_{\substack{d \in I(V_1) \\ (d, q)=1}} \left| \Delta_1(I(U_1); q, a\bar{d}^2) \right| + V_0 + \mathcal{E},$$

where

$$I(U_1) = (U_1, U_1 + UV_0^{-1}], \text{ and } I(V_1) = (V_1, V_1 + VV_0^{-1}]$$

and  $\mathcal{E}$  counts the number of pairs  $(m, d)$  such that

- (i)  $md^2 > X$ ,
- (ii)  $md^2 \equiv a \pmod{q}$  and
- (iii)  $d \in I(V_1), m \in I(U_1)$  with  $U_1' \in (U, 2U], V_1' \in (V, 2V]$  satisfying  $U_1'(V_1')^2 \leq X$ .

We easily see that

$$\begin{aligned} X < md^2 &\leq (U_1' + U'V_0^{-1})(V_1' + V'V_0^{-1})^2 \\ &\leq U_1'(V_1')^2 + U'V_1'V_0^{-1} + 3U_1'V_1'V'V_0^{-1} + 3U'V'V_1'V_0^{-2} \\ &\leq X + O(XV_0^{-1}). \end{aligned}$$

As  $X/V_0 \geq q^{1+\delta}$ , by Shiu's theorem [22] the contribution from these pairs is

$$\mathcal{E} \ll \sum_{\substack{X < n \leq X + O(X/V_0) \\ n \equiv a \pmod{q}}} \tau(n) + \frac{1}{\phi(q)} \sum_{\substack{X < n \leq X + O(X/V_0) \\ (n, q)=1}} \tau(n) \ll_{\delta} \frac{X \log X}{qV_0}.$$

We thus find that

$$\Delta_{\mu^2}(X; q, a) \ll_{\delta} V_0^2(\log X)^2 \sum_{\substack{d \in I(V_1) \\ (d, q)=1}} \left| \Delta_1(I(U_1); q, a\bar{d}^2) \right| + V_0 + X(\log X)/(qV_0), \tag{3}$$

again with  $U_1V_1^2 \leq 8X$ . Note that we may assume that  $V_1 > X^{1-\delta-\epsilon}/(qV_0)$ , since otherwise we immediately obtain

$$\Delta_{\mu^2}(X; q, a) \ll_{\epsilon, \delta} V_0^2 X^{\epsilon} |I(V_1)| + V_0 + X(\log X)/(qV_0) \ll X^{1-\delta}/q. \tag{4}$$

Let  $\tilde{q}$  be a divisor of  $q$  to be chosen later (the smoothness assumption on  $q$  will be useful in this selection). Using the fact that when  $U_1V_1^2 \leq 8X$ ,

$$\sum_{d \in I(V_1)} \frac{1}{\phi(q')} \sum_{m \in I(U_1)} 1_{(m, q')=1} = \frac{|I(V_1)|}{\phi(q')} \left( \frac{\phi(q')}{q'} |I(U_1)| + O(\tau(q')) \right) \ll \frac{X}{q'V_1V_0^2} + X^{\epsilon} \frac{V_1}{q'V_0}$$

for  $q' \in \{q, \tilde{q}\}$ , we obtain

$$\begin{aligned} \sum_{\substack{d \in I(V_1) \\ (d, q)=1}} \left| \Delta_1(I(U_1); q, a\bar{d}^2) \right| &\leq \sum_{\substack{d \in I(V_1) \\ (d, q)=1}} \left( \sum_{\substack{m \in I(U_1) \\ m \equiv a\bar{d}^2 \pmod{q}}} 1 + \frac{1}{\phi(q)} \sum_{\substack{m \in I(U_1) \\ (m, q)=1}} 1 \right) \\ &\leq \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} \left( \sum_{\substack{m \in I(U_1) \\ m \equiv a\bar{d}^2 \pmod{\tilde{q}}} 1 + \frac{1}{\phi(\tilde{q})} \sum_{\substack{m \in I(U_1) \\ (m, \tilde{q})=1}} 1 \right) + O\left(\frac{X}{\tilde{q}V_0^2V_1} + X^\varepsilon \frac{V_1}{\tilde{q}V_0}\right) \\ &= \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} \Delta_1(I(U_1); \tilde{q}, a\bar{d}^2) + O\left(\frac{X}{\tilde{q}V_1V_0^2} + \frac{X^\varepsilon V_1}{V_0\tilde{q}}\right). \end{aligned}$$

Hence, from (3) and  $V_1 \ll X^{1/2}$ ,

$$\Delta_{\mu^2}(X; q, a) \ll_\varepsilon V_0^2 (\log X)^2 \left| \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} \Delta_1(I(U_1); \tilde{q}, a\bar{d}^2) \right| + V_0 + \frac{X^{1+\varepsilon}}{qV_0} + \frac{X^{1+\varepsilon}}{\tilde{q}V_1} + \frac{V_0X^{1/2+\varepsilon}}{\tilde{q}}$$

for  $\tilde{q} \mid q, X^{\delta+\varepsilon} \leq V_0 \leq X^{1-\delta}/q$  and  $V_1 \geq \max\{V_0, X^{1-\delta-\varepsilon}/(qV_0)\}$ .

Having decoupled the variables  $d$  and  $m$  and removed the weight  $\mu(d)$ , we now introduce additive characters into the fold. By orthogonality, we have

$$\begin{aligned} \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} \left( \sum_{\substack{m \in I(U_1) \\ m \equiv a\bar{d}^2 \pmod{\tilde{q}}} 1 - \frac{1}{\phi(\tilde{q})} \sum_{\substack{m \in I(U_1) \\ (m, \tilde{q})=1}} 1 \right) &= \frac{1}{\tilde{q}} \sum_{k \pmod{\tilde{q}}} \left( \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} e_{\tilde{q}}(-ka\bar{d}^2) \right) \left( \sum_{m \in I(U_1)} e_{\tilde{q}}(km) \right) \\ &\quad - \frac{1}{\phi(\tilde{q})} \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} \sum_{\substack{m \in I(U_1) \\ (m, \tilde{q})=1}} 1. \end{aligned}$$

By the sieve,

$$\sum_{\substack{m \in I(U_1) \\ (m, \tilde{q})=1}} 1 = \frac{\phi(\tilde{q})}{\tilde{q}} |I(U_1)| + O(\tau(\tilde{q})),$$

the main term of which cancels the sum with  $k = 0$  above and so we obtain

$$\begin{aligned} \Delta_{\mu^2}(X; q, a) &\ll_\varepsilon \frac{V_0^2 (\log X)^2}{\tilde{q}} \sum_{\substack{k \pmod{\tilde{q}} \\ k \neq 0}} \left| \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} e_{\tilde{q}}(-ka\bar{d}^2) \sum_{m \in I(U_1)} e_{\tilde{q}}(km) \right| \\ &\quad + V_0 + \frac{X^{1+\varepsilon}}{qV_0} + \frac{X^{1+\varepsilon}}{\tilde{q}V_1} + \frac{V_0X^{1/2+\varepsilon}}{\tilde{q}} \\ &\ll_\varepsilon \frac{V_0^2 (\log X)^2}{\tilde{q}} \sum_{1 \leq |k| \leq \tilde{q}/2} \left| \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} e_{\tilde{q}}(ka\bar{d}^2) \right| \left| \sum_{m \in I(U_1)} e_{\tilde{q}}(-km) \right| \\ &\quad + V_0 \left( 1 + \frac{X^{1/2+\varepsilon}}{\tilde{q}} \right) + \frac{X^{1+\varepsilon}}{\tilde{q}V_1} + \frac{X^{1+\varepsilon}}{qV_0}. \end{aligned}$$

Applying the geometric series estimate

$$\sum_{m \in I(U_1)} e_{\tilde{q}}(km) \ll \min\{|I(U_1)|, \|k/\tilde{q}\|^{-1}\} \ll \tilde{q}/k$$

for  $1 \leq k \leq \tilde{q}/2$ , we thus obtain

$$\Delta_{\mu^2}(X; q, a) \ll_{\varepsilon} V_0^2 (\log X)^2 \sum_{1 \leq k \leq \tilde{q}/2} \frac{1}{k} \left| \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} e_{\tilde{q}}(kad\bar{d}^2) \right| + \frac{X^{1+\varepsilon}}{\tilde{q}V_1} + \frac{X \log X}{qV_0} + V_0.$$

**2.2. Bounding incomplete exponential sums on average**

Our main objective from here on is to get a suitable estimate on average over  $k$  for

$$S_{\tilde{q},a}(k; V_1) := \sum_{\substack{d \in I(V_1) \\ (d, \tilde{q})=1}} e_{\tilde{q}}(kad\bar{d}^2).$$

To simplify matters further, we separate the range of  $k$  according to  $(k, \tilde{q}) = f$ , giving

$$\sum_{1 \leq k \leq \tilde{q}-1} \frac{|S_{\tilde{q},a}(k; V_1)|}{k} = \sum_{\substack{f|\tilde{q} \\ f < \tilde{q}}} \frac{1}{f} \sum_{k' \pmod{\tilde{q}/f}}^* \frac{|S_{\tilde{q}/f,a}(k'; V_1)|}{k'}. \tag{5}$$

Fix  $f \mid \tilde{q}$  with  $f < \tilde{q}$  for the moment and put  $q' := \tilde{q}/f$ . Completing the sum, we obtain

$$\begin{aligned} S_{q',a}(k'; V_1) &= \sum_{l \pmod{q'}}^* e_{q'}(k'a\bar{l}^2) \sum_{\substack{d \in I(V_1) \\ d \equiv l \pmod{q'}}} 1 \\ &= \frac{1}{q'} \sum_{r \pmod{q'}} \sum_{l \pmod{q'}}^* e_{q'}(k'a\bar{l}^2 + rl) \sum_{d \in I(V_1)} e_{q'}(-rd) \\ &= \frac{1}{q'} \sum_{r=1}^{q'} e_{q'}(-rV_1) K_2(k'a, r; q') g_{q'}(r), \end{aligned} \tag{6}$$

where for  $Q \geq 2$  we have set

$$K_2(A, B; Q) := \sum_{x \pmod{Q}}^* e_Q(A\bar{x}^2 + Bx) \quad (A \in (\mathbb{Z}/Q\mathbb{Z})^\times, B \in \mathbb{Z}/Q\mathbb{Z}),$$

the complete exponential sum defined in the introduction, as well as

$$g_{q'}(r) := e_{q'}(rV_1) \sum_{d \in I(V_1)} e_{q'}(-rd) = \sum_{1 \leq d \leq V_1/V_0} e_{q'}(-rd) \ll \min\{V_1/V_0, q'/r\}. \tag{7}$$

We would like to use partial summation to remove the weight  $g_{q'}$  in (6); however, the long sum in  $r$  will make this inefficient in the sequel if we do not split the interval into shorter segments. To this end, let  $1 \leq K \leq q' - 1$  be a parameter that we will choose later. We split

$$S_{q',a}(k'; V_1) = \frac{1}{q'} \sum_{1 \leq m \leq q'/K} \sum_{K(m-1) < r \leq Km} e_{q'}(-rV_1) K_2(k'a, r; q') g_{q'}(r).$$

Given  $1 \leq m \leq q'/K$ , set

$$\kappa(m; k'a, q') := \max_{1 \leq R \leq K} \left| \sum_{r=K(m-1)+1}^{K(m-1)+R} e_{q'}(-rV_1) K_2(k'a, r; q') \right|.$$

Applying partial summation to estimate the derivative  $g'_{q'}$  of  $g_{q'}$ , we get

$$|g_{q'}(r+1) - g_{q'}(r)| = \left| \int_r^{r+1} g'_{q'}(t) dt \right| \leq \max_{r \leq t < r+1} |g'_{q'}(t)| \ll \frac{V_1}{q'V_0} \min\{V_1/V_0, q'/r\}.$$

Combining this, (7) and partial summation once again, we obtain

$$\begin{aligned} S_{q',a}(k'; V_1) &\ll \frac{1}{q'} \sum_{1 \leq m \leq q'/K} \kappa(m; k'a, q') \cdot K \max_{(m-1)K < r \leq mK} |g_{q'}(r+1) - g_{q'}(r)| \\ &\ll \frac{V_1}{q'V_0} \sum_{1 \leq m \leq q'/K} \frac{\kappa(m; k'a, q')}{m} + K \left( \frac{V_1}{q'V_0} \right)^2 |K_2(k'a, 0; q')|. \end{aligned} \tag{8}$$

We may control the terms in (5) with large  $f$  directly using a square-root cancelling bound for  $K_2(A, B, Q)$ , which we derive below.

**2.2.1. Point-wise bounds**

**Lemma 2.1** (*p*-adic stationary phase method). *For any  $n \geq 2$ ,  $a \in (\mathbb{Z}/p^n\mathbb{Z})^\times$  and  $b \in \mathbb{Z}/p^n\mathbb{Z}$ , we have*

$$K_2(a, b; p^n) = \left( \frac{3a}{p^n} \right) \varepsilon_{p,n} p^{n/2} \sum_{\substack{y \pmod{p^{\lfloor n/2 \rfloor}} \\ y^3 \equiv \overline{2ab} \pmod{p^{\lfloor n/2 \rfloor}}}^* e_{p^n}(3ay^2), \tag{9}$$

where  $\varepsilon_{p,n} = 1$  if  $n$  is even and  $\varepsilon_{p,n} = i^{(p-1)^2/4}$  if  $n$  is odd.

*Proof.* Applying Lemmas 12.2 and 12.3 of [10], we have that if  $A \in (\mathbb{Z}/p^n\mathbb{Z})^\times$  and  $B \in \mathbb{Z}/p^n\mathbb{Z}$ , then

$$K_2(A, B; p^n) = p^{n/2} \sum_{\substack{y \pmod{p^{\lfloor n/2 \rfloor}} \\ y^3 \equiv \overline{2AB} \pmod{p^{\lfloor n/2 \rfloor}}}^* e_{p^n}(Ay^2 + B\bar{y}) \theta_{p^n}(y; A, B),$$

where for  $y$  satisfying  $y^3 \equiv \overline{2AB} \pmod{p^{\lfloor n/2 \rfloor}}$  we have set

$$\theta_{p^n}(y; A, B) := \begin{cases} 1 & \text{if } 2 \mid n \\ p^{-1/2} \sum_{z \pmod{p}} e_p \left( 3Az^2 + z \left( (2Ay - B\bar{y}^2)/p^{(n-1)/2} \right) \right) & \text{if } 2 \nmid n. \end{cases}$$

A key point here is that the set of critical points – that is, solutions to  $y^3 \equiv \overline{2AB} \pmod{p^{\lfloor n/2 \rfloor}}$  – is invariant under translations by  $p^{n-\lfloor n/2 \rfloor} \mathbb{Z}/p^n\mathbb{Z}$  (see, e.g., [15, Lemma 1]); in particular, by choosing a lift of such critical points to solutions to  $y^3 \equiv \overline{2AB} \pmod{p^n}$  via Hensel’s lemma, we may rewrite

$$Ay^2 + B\bar{y} \equiv 3Ay^2 \pmod{p^n}.$$

When  $n = 2m$ , with  $m \geq 1$ , we simply have (with  $A = a$  and  $B = b$ )

$$K_2(a, b; p^{2m}) = p^m \sum_{\substack{y \pmod{p^m} \\ y^3 \equiv 2ab \pmod{p^m}}}^* e_{p^{2m}}(3ay^2),$$

which implies the claim in this case. On the other hand, completing the square and using the explicit computation of Gauss sums modulo  $p$  when  $n = 2m + 1$ , we get

$$\begin{aligned} \theta_{p^{2m+1}}(y; A, B) &= p^{-1/2} \sum_{z \pmod{p}} e_p \left( 3A \left( z + \frac{2Ay - B\bar{y}^2}{6A} \right)^2 - \frac{(2Ay - B\bar{y}^2)^2}{12A} \right) \\ &= \left( \frac{3A}{p} \right) i^{(p-1)^2/4} e_{p^{2m+1}} \left( -\bar{3}Ay^2(1 - \bar{2}AB\bar{y}^3)^2 \right). \end{aligned}$$

Again, using the  $p^{m+1}\mathbb{Z}/p^{2m+1}\mathbb{Z}$ -translation invariance of the solutions to  $y^3 \equiv \bar{2}AB \pmod{p^m}$ , at critical points the exponential here is simply 1 and we obtain (when  $A = a$  and  $B = b$ )

$$K_2(a, b; p^{2m+1}) = \left( \frac{3a}{p} \right) i^{(p-1)^2/4} p^{(2m+1)/2} \sum_{\substack{y \pmod{p^m} \\ y^3 \equiv 2ab \pmod{p^m}}}^* e_{p^{2m+1}}(3ay^2).$$

The claim is proved. □

**Lemma 2.2.** *Let  $Q \geq 2$ . Then*

$$\max_{\substack{A \in (\mathbb{Z}/Q\mathbb{Z})^\times \\ B \in \mathbb{Z}/Q\mathbb{Z}}} |K_2(A, B; Q)| \ll_\varepsilon Q^{1/2+\varepsilon}.$$

*Proof.* We reduce to the case of prime power moduli using the Chinese remainder theorem. Indeed, if  $Q = m_1 m_2$  where  $m_1$  and  $m_2$  are coprime, then selecting  $r_1, r_2 \in \mathbb{Z}$  such that  $m_1 r_1 + m_2 r_2 = 1$ , we have

$$K_2(A, B; Q) = K_2(r_2 A, r_2 B; m_1) K_2(r_1 A, r_1 B; m_2). \tag{10}$$

Applying this inductively over the prime power divisors of  $Q$  and taking maxima over  $A$  and  $B$ , we obtain

$$\max_{\substack{A \in (\mathbb{Z}/Q\mathbb{Z})^\times \\ B \in \mathbb{Z}/Q\mathbb{Z}}} |K_2(A, B; Q)| \leq \prod_{p^n || Q} \max_{\substack{A \in (\mathbb{Z}/p^n\mathbb{Z})^\times \\ B \in \mathbb{Z}/p^n\mathbb{Z}}} |K_2(A, B; p^n)|.$$

Now, observe that for each  $p || Q$  prime and  $A, B \in \mathbb{Z}/p\mathbb{Z}$  we can represent

$$K_2(A, B; p) = \sum_{x \in \mathbb{F}_p^\times} e_p(Ax^2 + Bx) = \sum_{x \in \mathbb{F}_p^\times} e_p(f_{A,B}(x)),$$

where  $f_{A,B}(x) = A/x^2 + Bx \in \mathbb{F}_p(x)$ . By the Bombieri–Dwork–Weil bound [1, Theorem 6] (see also [3, Section 3.5]),

$$\max_{A \in \mathbb{F}_p^\times, B \in \mathbb{F}_p} |K_2(A, B; p)| \leq C' \sqrt{p} \tag{11}$$

for  $C' > 0$  an absolute constant. On the other hand, if  $p^n || Q$  with  $n \geq 2$ , then by Lemma 2.1 we have

$$|K_2(A, B; p^n)| \leq p^{n/2} |\{y \pmod{p^{\lfloor n/2 \rfloor}} : y^3 \equiv \bar{2}AB \pmod{p^{\lfloor n/2 \rfloor}}\}|.$$

The solutions to  $y^3 \equiv \overline{2AB} \pmod{p}$  lift uniquely to solutions modulo  $p^{\lfloor n/2 \rfloor}$  by Hensel’s lemma, so the cardinality in the previous line is  $\leq 3$ . In particular,

$$\max_{\substack{A \in (\mathbb{Z}/p^n\mathbb{Z})^\times \\ B \in \mathbb{Z}/p^n\mathbb{Z}}} |K_2(A, B; p^n)| \leq Cp^{n/2}$$

for  $C := \max\{C', 3\}$ .

We therefore conclude that

$$\max_{\substack{A \in (\mathbb{Z}/Q\mathbb{Z})^\times \\ B \in \mathbb{Z}/Q\mathbb{Z}}} |K_2(A, B; Q)| \leq C^{\omega(Q)} Q^{1/2} \ll_\varepsilon Q^{1/2+\varepsilon}$$

as claimed. □

Combining Lemma 2.2 with (6) and  $V_1 \ll X^{1/2}$  gives

$$\begin{aligned} S_{q',a}(k'; V_1) &\ll_\varepsilon (\log q) \max_{\substack{r \pmod{q'} \\ r \neq 0}} |K_2(k'a, r; q')| + \frac{|g_{q'}(0)| |K_2(k'a, 0; q')|}{q'} \\ &\ll_\varepsilon (q')^{1/2+\varepsilon} + X^\varepsilon (X/q')^{1/2} V_0^{-1}. \end{aligned}$$

Let  $Z \geq 2$  be a parameter to be chosen later. Applying this with  $q' = \tilde{q}/f$  with  $f > Z$  in particular, it follows immediately that

$$\begin{aligned} \sum_{\substack{f|q \\ f > Z}} \frac{1}{f} \sum_{1 \leq k \leq \tilde{q}/(2f)} \frac{|S_{\tilde{q}/f,a}(k'; V_1)|}{k'} &\ll_\varepsilon \tilde{q}^\varepsilon \sum_{\substack{f|\tilde{q} \\ f > Z}} \frac{1}{f} \left( (\tilde{q}/f)^{1/2+\varepsilon} + X^\varepsilon (X/\tilde{q})^{1/2} f^{1/2} V_0^{-1} \right) \\ &\ll X^\varepsilon \left( \tilde{q}^{1/2} Z^{-3/2} + Z^{-1/2} (X/\tilde{q})^{1/2} V_0^{-1} \right), \end{aligned}$$

which, in combination with (8), thus gives

$$\begin{aligned} \Delta_{\mu^2}(X; q, a) &\ll_\varepsilon X^\varepsilon \frac{V_0 V_1}{\tilde{q}} \sum_{\substack{f|\tilde{q} \\ f \leq Z}} \sum_{1 \leq k' \leq \tilde{q}/(2f)} \sum_{m=1}^{\tilde{q}/(fK)} \frac{\kappa(m; k'a, \tilde{q}/f)}{k'm} \\ &\quad + V_0 \left( 1 + X^\varepsilon \left( \frac{X}{Z\tilde{q}} \right)^{1/2} \right) + X^\varepsilon \left( \frac{V_0^2 \tilde{q}^{1/2}}{Z^{3/2}} + \frac{X}{qV_0} + \frac{X}{\tilde{q}V_1} + KV_1^2 \left( \frac{Z}{\tilde{q}} \right)^{3/2} \right), \end{aligned} \tag{12}$$

provided  $K \leq \tilde{q}/Z$ .

**2.2.2. Bounds on average**

It turns out (see Subsection 6.1) that merely applying Lemma 2.2 directly to  $K_2$  (after replacing  $q$  by  $\tilde{q}$  as we have done above) results in a power-saving upper bound for  $\Delta_{\mu^2}(X; q, a)$  for any  $q \leq X^{3/4-\varepsilon}$ , provided  $\tilde{q}$  can be chosen with appropriate size. This is independent of squarefreeness considerations and is modeled after Hooley’s idea in [8].

Ignoring the effect of the sum over divisors  $f$  (which will have little influence in the sequel), in order to do better we will need to find a power savings in  $X$  over the pointwise estimate from Lemma 2.2; that is,

$$\max_{1 \leq m \leq \tilde{q}/K-1} \max_{1 \leq k' \leq \tilde{q}-1} |\kappa(m; k'a, \tilde{q})| \leq K \max_{\substack{A \in (\mathbb{Z}/\tilde{q}\mathbb{Z})^\times \\ B \in \mathbb{Z}/\tilde{q}\mathbb{Z}}} |K_2(A, B; \tilde{q})| \ll_\varepsilon K \tilde{q}^{1/2+\varepsilon},$$

by utilising the averaging of  $K_2$  sums implicit in the definition of  $\kappa(m; k'a, \tilde{q})$ . To this end, we will apply the  $q$ -van der Corput method of Heath-Brown, as formulated by Irving in [9]. The following is proved *mutatis mutandis* by the arguments in [9].

**Proposition 2.3.** ([9, Lemma 4.3]). Let  $K, L \geq 1$  and  $M \in \mathbb{Z}$ . Suppose  $Q \geq 4$  factors as  $Q = Q_0 \cdots Q_L$ , with  $Q_j \geq 2$  for each  $0 \leq j \leq L$ . Let  $J$  be an interval with  $|J| \leq K$  and set

$$T(b, M) := \sum_{k \in J} e_Q(-Mk) K_2(b, k; Q),$$

where  $b$  is a coprime residue class modulo  $Q$ . If  $K \geq \max\{Q_1, \dots, Q_L\}$ , then

$$|T(b, M)| \ll_{\varepsilon, L} Q^{1/2+\varepsilon} K \left( \sum_{j=1}^L \left( \frac{Q_{L-j+1}}{K} \right)^{2^{L-j}} + \frac{Q}{K^{(L+1)} Q_0^{2^{L-1}+1}} \sum_{1 \leq |h_1| \leq K/Q_1} \cdots \sum_{1 \leq |h_L| \leq K/Q_L} |T(\mathbf{h})| \right)^{2^{-L}}$$

where for each  $\mathbf{h} \in \mathbb{Z}^L$  with  $1 \leq |h_j| \leq K/Q_j$  there is an interval  $J(\mathbf{h})$  of size  $\leq K$  and  $b'$  coprime to  $q$  such that

$$T(\mathbf{h}) := \sum_{k \in J(\mathbf{h})} \prod_{I \subseteq \{1, \dots, L\}} \mathcal{C}^{|I|} K_2 \left( b', k + \sum_{i \in I} Q_i h_i, Q_0 \right),$$

$\mathcal{C}(z) := \bar{z}$  being the complex conjugation map.

Proposition 2.3 indicates that our main point of focus for the remainder of the argument will be the estimation of  $|T(\mathbf{h})|$ , for  $\mathbf{h}$  satisfying  $Q_j |h_j| \in [1, K]$  for all  $1 \leq j \leq L$ . We will estimate these terms pointwise in  $\mathbf{h}$ , the key point being that, outside of a sparse set of  $\mathbf{h}$ , we will obtain significant cancellation. This will result in the following.

**Proposition 2.4.** Adopt the notation of Proposition 2.3.

i) Assume that  $Q = Q_0 \cdots Q_L$  is squarefree. Then

$$\sum_{1 \leq |h_1| \leq K/Q_1} \cdots \sum_{1 \leq |h_L| \leq K/Q_L} |T(\mathbf{h})| \ll_{\varepsilon, L} \frac{K^L}{Q} Q_0^{2^{L-1}+3/2+\varepsilon} \left( \frac{K}{Q_0} + 1 \right).$$

ii) Then there is a  $\delta' = \delta'(L) \in (0, 2^{-2^L}]$  such that the following holds. Suppose  $Q = Q_0 \cdots Q_L$  is such that  $(Q_i, Q_j) = 1$  for all  $0 \leq i < j \leq L$  and  $(Q_0, 6) = 1$ . Assume also that  $K/Q_j > Q_0^{2\delta'}$  for all  $p^v || Q_0$  and all  $1 \leq j \leq L$ . Then

$$\sum_{1 \leq |h_1| \leq K/Q_1} \cdots \sum_{1 \leq |h_L| \leq K/Q_L} |T(\mathbf{h})| \ll_{\varepsilon, L} \frac{K^L}{Q} Q_0^{2^{L-1}+2-\delta'+\varepsilon}.$$

Proposition 2.4 will be proved in the next two sections. As a first step, we shall replace  $T(\mathbf{h})$  by analogous complete sums (13) modulo prime powers with an additional additive phase using the following.

**Lemma 2.5.** Let  $I \subset \mathbb{Z}/Q_0\mathbb{Z}$  be an interval and let  $A_0 \in (\mathbb{Z}/Q_0\mathbb{Z})^\times$ ,  $B_0 \in \mathbb{Z}/Q_0\mathbb{Z}$ . Let  $N, M \geq 0$  with  $N + M \geq 1$  and let  $\mathbf{h} \in \mathbb{Z}^N$ ,  $\mathbf{h}' \in \mathbb{Z}^M$ . Write  $Q_0 = \prod_{1 \leq j \leq k} p_j^{\alpha_j}$  for distinct primes  $p_j$  and  $k = \omega(Q_0)$ . Then there exist  $\mathbf{A}, \mathbf{B} \in \prod_{1 \leq j \leq k} \mathbb{Z}/p_j^{\alpha_j}\mathbb{Z}$  such that

$$\begin{aligned} & \sum_{B \in I} \prod_{1 \leq i \leq N} K_2(A_0, B_0 + h_i; Q_0) \prod_{1 \leq j \leq M} \bar{K}_2(A, B + h'_j; Q_0) \\ & \ll \sum_{C \pmod{Q_0}} \min \left\{ \frac{|I|}{Q_0}, \frac{1}{Q_0 \|C/Q_0\|} \right\} \prod_{j=1}^k |T(A_j, B_j, C, \mathbf{h}, \mathbf{h}'; p_j^{\alpha_j})|, \end{aligned}$$

where for each  $Q \mid Q_0$  we set

$$T(A, B, C, \mathbf{h}, \mathbf{h}'; Q) := \sum_b \sum_{(\text{mod } Q)} e_Q(CBb) \cdot \prod_{1 \leq i \leq N} K_2(A, b + h_i; Q) \prod_{1 \leq j \leq M} \overline{K}_2(A, b + h'_j; Q). \tag{13}$$

*Proof.* For  $B \in \mathbb{Z}/Q_0\mathbb{Z}$  put

$$g(B) := \prod_{1 \leq i \leq N} K_2(A, B + h_i; Q_0) \prod_{1 \leq j \leq M} \overline{K}_2(A, B + h'_j; Q_0).$$

Completing the sum over  $B$  modulo  $Q_0$ , the left-hand side is

$$\begin{aligned} \sum_B \sum_{(\text{mod } Q_0)} 1_I(B)g(B) &= \frac{1}{Q_0} \sum_C \sum_{(\text{mod } Q_0)} \left( \sum_{D \in I} e_{Q_0}(-CD) \right) \left( \sum_B \sum_{(\text{mod } Q_0)} g(B)e_{Q_0}(CB) \right) \\ &\ll \frac{1}{Q_0} \sum_C \sum_{(\text{mod } Q_0)} \min\{|I|, \|C/Q_0\|^{-1}\} |T(A, B, C, \mathbf{h}, \mathbf{h}'; Q_0)|. \end{aligned}$$

It thus suffices to show the existence of  $A_j, B_j \pmod{p_j^{\alpha_j}}$  for each  $1 \leq j \leq k$ , such that

$$T(A, B, C, \mathbf{h}, \mathbf{h}'; Q_0) = \prod_{1 \leq j \leq k} T(A_j, B_j, C, \mathbf{h}, \mathbf{h}'; p_j^{\alpha_j}).$$

We prove this by induction on  $k$ , the number of distinct prime factors of  $Q_0$ . When  $k = 1$  there is nothing to prove. Assume this works for any  $Q_0$  having  $k$  distinct prime factors and now suppose that  $Q_0$  has  $k + 1$  such factors. Write  $Q_0 = Q'_0 p^\alpha$ , where  $p \nmid Q'_0$  and  $\alpha \geq 1$ . Let  $r, s \in \mathbb{Z}$  be chosen such that  $rQ'_0 + sp^\alpha = 1$ . By the Chinese remainder theorem, every  $B \pmod{Q_0}$  can be written uniquely as  $B = uQ'_0 + vp^\alpha$ , where  $0 \leq u \leq p^\alpha - 1$  and  $0 \leq v \leq Q'_0 - 1$ . Thus,

$$\begin{aligned} T(A, B, C, \mathbf{h}, \mathbf{h}'; Q_0) &= \sum_u \sum_{(\text{mod } p^\alpha)} e_{p^\alpha}(Cu) \sum_v \sum_{(\text{mod } Q'_0)} e_{Q'_0}(Cv) \\ &\cdot \prod_{1 \leq i \leq N} K_2(A, uQ'_0 + vp^\alpha + h_i; p^\alpha Q'_0) \prod_{1 \leq j \leq M} \overline{K}_2(A, uQ'_0 + vp^\alpha + h'_j; p^\alpha Q'_0). \end{aligned}$$

Applying (10) and the symmetry  $K_2(\alpha, \gamma\beta; q') = K_2(\gamma^2\alpha, \beta; q')$  for any  $\gamma \in (\mathbb{Z}/q'\mathbb{Z})^\times$ , we see that the products are

$$\begin{aligned} &\left( \prod_{i=1}^N K_2(s^3 A, uQ'_0 + h_i; p^\alpha) \prod_{j=1}^M \overline{K}_2(s^3 A, uQ'_0 + h'_j; p^\alpha) \right) \\ &\cdot \left( \prod_{i=1}^N K_2(r^3 A, vp^\alpha + h_i; Q'_0) \prod_{j=1}^M \overline{K}_2(r^3 A, vp^\alpha + h'_j; Q'_0) \right). \end{aligned}$$

Making the change of variables  $u \mapsto uQ'_0$  and  $v \mapsto vp^\alpha$ , we thus get

$$T(A, B, C, \mathbf{h}, \mathbf{h}'; Q_0) = T(s^3 A, \overline{Q}'_0, C, \mathbf{h}, \mathbf{h}'; p^\alpha) \cdot T(r^3 A, \overline{p}^\alpha, C, \mathbf{h}, \mathbf{h}'; Q'_0),$$

where the inverses are taken modulo  $p^\alpha$  and  $Q'_0$  respectively.

By induction, we can factor the second bracketed term similarly into  $k$  products and the claim follows.  $\square$

In Irving’s work [9], where squarefree moduli specifically are treated, an application of a (variant of a) result of Fouvry et al. on correlations of Kloosterman sums ([5, Proposition 3.2]) is used to control the complete sums to prime moduli arising in the factorisation in Lemma 2.5. To achieve the same goal, we will prove in full detail a similar result for correlations of the complete exponential sums  $K_2$  in the next section. In Section 4, we treat the same problem for prime power moduli, using rather different techniques.

### 3. Correlations of $K_2$ Sums to Prime Moduli: Cohomological Methods

The goal of this section is to provide an estimate for sums like (13). We will prove, in full detail, the following analogue of [5, Proposition 3.2] and [9, Section 4.3] (the latter of which cites private communications for the corresponding result for Kloosterman sums). This will be the main input to the proof of Proposition 2.4 i).

Throughout this section, fix a prime  $p$ . Given  $N, M \geq 1$  and tuples  $\mathbf{h} \in \mathbb{F}_p^N$  and  $\mathbf{h}' \in \mathbb{F}_p^M$ , we define

$$T = T_{\mathbf{h}, \mathbf{h}'} := \{h_1, \dots, h_N, h'_1, \dots, h'_M\},$$

and for each  $\tau \in T$  define

$$\begin{aligned} \mu(\tau) &= \mu_{\mathbf{h}}(\tau) := |\{1 \leq j \leq N : h_j = \tau\}| \\ \nu(\tau) &= \nu_{\mathbf{h}'}(\tau) := |\{1 \leq j \leq M : h'_j = \tau\}|. \end{aligned}$$

**Theorem 3.1.** For  $A \in \mathbb{F}_p^\times$ ,  $\psi \in \widehat{\mathbb{F}_p}$  a possibly trivial additive character and  $h_1, \dots, h_N, h'_1, \dots, h'_M \in \mathbb{F}_p$  (where  $N + M \geq 1$ ), we have

$$\sum_{B \in \mathbb{F}_p} \psi(B) \prod_{i=1}^N K_2(A, B + h_i; p) \prod_{j=1}^M \overline{K}_2(A, B + h'_j; p) \ll (N + M) 3^{N+M} p^{\frac{N+M+1}{2}} \tag{14}$$

unless  $\psi$  is trivial and  $\mu(\tau) \equiv \nu(\tau) \pmod{3}$  for all  $\tau \in T$ . The implied constant is absolute (e.g., does not depend on  $N$  or  $M$ ).

**Remark 3.2.** We can rewrite the left-hand side of (14) as

$$\sum_{B \in \mathbb{F}_p} \psi(B) \prod_{\tau \in T} K_2(A, B + \tau; p)^{\mu(\tau)} \overline{K}_2(A, B + \tau; p)^{\nu(\tau)}.$$

Note that when  $\psi$  is trivial and  $\mu(\tau) = \nu(\tau)$  for all  $\tau \in T$ , the summands are all nonnegative. We should therefore not expect to find any improvement over the trivial bound in general (outside of the possibility that the absolute values of  $p^{-1/2} K_2(A, B + \tau; p)$  are small, which is typical for large  $p$  given that these normalised sums become equidistributed according to Haar measure on  $SU_3(\mathbb{C})$  as  $B$  varies modulo  $p$  and  $p \rightarrow \infty$ ). It will transpire from the proof of Theorem 3.1 that our characterisation of when savings are available for the above correlation sums is essentially sharp.

To prove Theorem 3.1 we will employ the method surveyed in [6]. Highlights of this method include

- o the interpretation of  $K_2$  as the trace function of an  $\ell$ -adic sheaf, pure of weight 1 and the computation of its geometric monodromy group (following Katz [12]);
- o the determination of geometric isomorphisms between shifts of these sheaves to find the monodromy of a product sheaf via the Goursat–Kolchin–Ribet criterion; and

- the application of the Grothendieck–Lefschetz trace formula and the Grothendieck–Ogg–Shafarevich Euler–Poincaré formula.

We will use the statements from [18] that also track the dependencies in the number of factors (i.e.,  $N$  and  $M$ ), for completeness. For the sake of concision, we refer the reader to these (and the references therein) for additional definitions, notations and other details.

### 3.1. Cohomological Interpretation of $K_2$ and (14)

In the language of [3], the pointwise bound on  $K_2$  modulo primes from Lemma 2.2 (see (11)) can be seen as the outcome of applying the Grothendieck–Lefschetz trace formula and Deligne’s Riemann hypothesis over finite fields [4] to the one-dimensional Artin–Schreier  $\ell$ -adic sheaf  $\mathcal{L}_{e(f_{A,B}/p)}$  on  $\mathbb{G}_m \times \mathbb{F}_p$ :

$$\begin{aligned} \left| \sum_{x \in \mathbb{F}_p^\times} e(f_{A,B}(x)/p) \right| &= \left| \sum_{x \in \mathbb{F}_p^\times} \iota \operatorname{tr}(\operatorname{Frob}_{x,p} \mid \mathcal{L}_{e(f_{A,B}/p)}) \right| \\ &= \left| \sum_{i=0}^2 (-1)^i \iota \operatorname{tr}(\operatorname{Frob}_p \mid H_c^i(\mathbb{G}_m \times \overline{\mathbb{F}}_p, \mathcal{L}_{e(f_{A,B}/p)})) \right| \\ &\leq \sqrt{p} \cdot \dim H_c^1(\mathbb{G}_m \times \overline{\mathbb{F}}_p, \mathcal{L}_{e(f_{A,B}/p)}), \end{aligned}$$

where  $\ell \neq p$  is an auxiliary prime,  $\iota : \overline{\mathbb{Q}}_\ell \rightarrow \mathbb{C}$  is a compatible<sup>7</sup> embedding,  $\operatorname{Frob}_{x,p} \in \pi_1^{\text{geom}}(\mathbb{G}_m, \overline{\eta})$  is the geometric Frobenius class at  $x \in \mathbb{G}_m(\mathbb{F}_p)$  for  $\overline{\eta}$  a geometric generic point and  $\operatorname{Frob}_p \in \operatorname{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$  is the geometric Frobenius (acting on the cohomology groups). The dimension of the first cohomology group is bounded independently from  $p$  by the Grothendieck–Ogg–Shafarevich formula, giving the constant  $C'$  in (11) above.

To handle sums of  $K_2$  sums in Theorem 3.1, we adopt a different perspective and consider  $K_2$  as an  $\ell$ -adic trace function itself, using the  $\ell$ -adic Fourier transform of Deligne. Some properties of the sheaf thus produced are given in the following lemma.

**Lemma 3.3.** *Assume that  $p$  is an odd prime, fix an auxiliary prime  $\ell \neq p$ , a compatible embedding  $\iota : \overline{\mathbb{Q}}_\ell \rightarrow \mathbb{C}$  and let  $A \in \mathbb{F}_p^\times$ . There exists a middle-extension  $\overline{\mathbb{Q}}_\ell$ -sheaf of  $\overline{\mathbb{Q}}_\ell$ -modules  $\mathcal{G}_A$  that is lisse on  $\mathbb{G}_m \times \mathbb{F}_p$  and pointwise pure of weight 0, as well as  $\gamma \in \mathbb{C}$  of modulus one, such that for every  $B \in \mathbb{G}_m(\mathbb{F}_p)$ ,*

$$\iota \operatorname{tr}(\operatorname{Frob}_{B,p} \mid \mathcal{G}_A) = \frac{-\gamma K_2(A, B; p)}{\sqrt{p}}. \tag{15}$$

Moreover:

1.  $\mathcal{G}_A$  has rank 3, with a unique  $\infty$ -break at  $2/3$ ;
2.  $\operatorname{Swan}_\infty(\mathcal{G}_A) = 2$ ;
3.  $\mathcal{G}_A$  is tamely ramified at 0 and its local monodromy is a unipotent pseudoreflection;
4. if  $p > 7$ , the arithmetic and geometric monodromy groups  $G_{\text{arith}}$ ,  $G_{\text{geom}}$  (that is, the Zariski closures of the images of the arithmetic/geometric fundamental groups by the corresponding representation composed with  $\iota$ ) are equal and isomorphic to  $\operatorname{SL}_3(\mathbb{C})$ .

*Proof.* A sheaf  $\mathcal{G}'_A$  satisfying all of the above properties but (4) is given by the  $\ell$ -adic Fourier transform of the Artin–Schreier sheaf  $\mathcal{L}_{e(f(X)/p)}$ , where  $f(X) = A/X^2 \in \mathbb{F}_p(X)$ . The resulting trace function (the left-hand side of (15)) is indeed the discrete Fourier transform of  $e(f(x)/p)$ , namely,  $K_2$ . The

<sup>7</sup>By this we mean that  $\iota$  must take  $\ell$ -adic additive characters  $\lambda : \mathbb{F}_p \rightarrow \overline{\mathbb{Q}}_\ell$ , implicit in the definition of the  $\ell$ -adic Fourier transform defined below, to usual additive characters  $\iota \circ \lambda : \mathbb{F}_p \rightarrow \mathbb{C}$ ; for example,  $\iota(\lambda(x)) := e(x/p)$  for each  $x \in \mathbb{F}_p$ .

statements of the lemma are all contained in [12, Section 7.3, Section 7.12 (SL-Example(3)), Theorem 7.12.3.1], with  $h = f$ ,  $\alpha = 0$  and  $n_0 = \text{ord}_0(f) = 2$ . In particular, the rank is  $1 + n_0 = 3$ .

Regarding the last property, since  $f(x) + f(-x)$  is nonconstant, the geometric monodromy group is isomorphic to  $\text{SL}_3(\mathbb{C})$  for  $p > 7$  by the theorem [12, Theorem 7.12.3.1]. By [12, Section 7.12 (SL-Example(3))],  $\det \mathcal{G}'_A$  is geometrically trivial, whence arithmetically isomorphic to  $\beta \otimes \overline{\mathbb{Q}}_\ell$ , where  $\beta$  is a  $p$ -Weil number of weight 0. Therefore, setting  $\mathcal{G}_A = \beta^{-1/3} \otimes \mathcal{G}'_A$  yields

$$\text{SL}_3(\mathbb{C}) = G_{\text{geom}}(\mathcal{G}_A) = G_{\text{geom}}(\mathcal{G}'_A) \leq G_{\text{arith}}(\mathcal{G}_A) \leq \text{SL}_3(\mathbb{C}),$$

with  $\mathcal{G}_A$  still verifying the previous properties. The theorem then holds with the twist  $\gamma = \iota\beta^{-1/3} \in \mathbb{C}$ .  $\square$

Note that the purity and dimensionality statements encompass the bound (11) from Lemma 2.2:

$$|K_2(A, B; p)| \leq \dim(\mathcal{G}_A)\sqrt{p} = 3\sqrt{p} \tag{16}$$

(including at  $B = 0$ , when  $K_2(A, 0; p)$  is, up to a unimodular factor, quadratic Gauss sum with modulus exactly  $\sqrt{p}$ ).

The bound in (14) is trivial if  $p \leq 7$ , so we may assume in what follows that  $p > 7$ . Using Lemma 3.3, we may bound the modulus of the left-hand side of (14) as

$$\ll p^{\frac{N+M}{2}} \left( \left| \sum_{B \in U(\mathbb{F}_p)} \iota \text{tr}(\text{Frob}_{B,p} \mid \mathcal{H}_{A,h,\psi}) \right| + 3^{N+M}(N+M) \right), \tag{17}$$

setting

$$\mathcal{H}_{A,h,\psi} := \mathcal{L}_\psi \otimes \left( \bigotimes_{\tau \in T_{h,h'}} [+ \tau]^* \mathcal{G}_A^{\otimes \mu(\tau)} \otimes [+ \tau]^* D(\mathcal{G}_A)^{\otimes \nu(\tau)} \right)$$

and letting  $U(\mathbb{F}_p)$  denote its set of lisse points. Here, the tensor product  $\otimes$  and dual  $D$  are understood at the level of the corresponding  $\ell$ -adic representations (as in [11, 12]). The second term in brackets in (17) arises by applying (16) to the set of ramification points of the tensor product, of which there are  $\leq N+M$ .

The trace formula will now be applied to (17).

### 3.2. Applying the trace formula

By the Grothendieck–Lefschetz trace formula and the Grothendieck–Ogg–Shafarevich formula (see the references in [18, Theorem 2.5]),

$$\sum_{B \in U(\mathbb{F}_p)} \text{tr}(\text{Frob}_{B,p} \mid \mathcal{H}_{A,h,\psi}) = p \cdot \text{tr}(\text{Frob}_p \mid (\mathcal{H}_{A,h,\psi})_{\pi_{1,p}^{\text{geom}}(U, \overline{\eta})}) + O(E(\mathcal{H}_{A,h,\psi})\sqrt{p}) \tag{18}$$

with an absolute implied constant, where

$$\begin{aligned} E(\mathcal{H}_{A,h,\psi}) &= \text{rank}(\mathcal{H}_{A,h,\psi}) \left( |\text{Sing}(\mathcal{H}_{A,h,\psi})| + \sum_{x \in \text{Sing}(\mathcal{H}_{A,h,\psi})} \text{Swan}_x(\mathcal{H}_{A,h,\psi}) \right) \\ &\leq 3^{N+M} (N+M + (N+M) \cdot 2 + 1) \leq 4(N+M)3^{N+M}, \end{aligned}$$

the inequality coming from Lemma 3.3 and the ramification properties of the tensor product recalled in [11, Lemma 1.3]. Let

$$\mathcal{H}_{A,h,\psi}^+ = \mathcal{L}_\psi \oplus \left( \bigoplus_{\tau \in T_{h,h'}} [+ \tau]^* \mathcal{G}_A \right)$$

and note that

$$\begin{aligned} G_{\text{geom}} \left( \mathcal{H}_{A,h,\psi}^+ \right) &\leq G_{\text{arith}} \left( \mathcal{H}_{A,h,\psi}^+ \right) \ni \text{Frob}_p \\ &\leq G_{\text{geom}}(\mathcal{L}_\psi) \times G_{\text{geom}}(\mathcal{G}_A)^H = \begin{cases} \mathbb{Z}/p \times \text{SL}_3(\mathbb{C})^H & : \psi \neq 0 \\ \text{SL}_3(\mathbb{C})^H & : \psi = 0, \end{cases} \end{aligned}$$

where  $H := |T_{h,h'}|$ . A very convenient case arises when these inclusions are equalities: in this event,  $\text{Frob}_p$  acts trivially on the coinvariant space in (18) and the right-hand side becomes

$$\begin{aligned} &p \cdot \dim(\mathcal{H}_{A,h,\psi})_G + O \left( E(\mathcal{H}_{A,h,\psi})\sqrt{p} \right) \\ &= \delta_{\psi=0} \prod_{\tau \in T} \text{mult}_1 \left( \text{Std}^{\otimes \mu(\tau)} \otimes D(\text{Std})^{\otimes \nu(\tau)} \right) + O((N + M)3^{N+M} p^{1/2}), \end{aligned}$$

by Schur’s lemma (see [17, Lemma 4.6, Proposition 4.8]), where  $\text{mult}_1$  is the multiplicity of the trivial representation on a representation  $\rho$ . Finally, the multiplicities are zero if and only if,  $3 | (\mu(\tau) - \nu(\tau))$  for all  $\tau \in T_{h,h'}$ : this is precisely the content of [13, Prop. 4.5(2)], in the case  $N = 3$  (in the notation there,  $\text{mult}_1 \left( \text{Std}^{\otimes \mu(\tau)} \otimes D(\text{Std})^{\otimes \nu(\tau)} \right) = A_{\nu(\tau),\mu(\tau)}$ , for each  $\tau$ ).

Therefore, it only remains to prove:

**Lemma 3.4.** *The arithmetic and geometric monodromy groups of  $\mathcal{H}_{A,h,\psi}$  coincide and are isomorphic to  $\mathbb{Z}/p \times \text{SL}_3(\mathbb{C})^H$  if  $\psi$  is nontrivial and  $\text{SL}_3(\mathbb{C})^H$  otherwise, where  $H = |T_{h,h'}|$ .*

*Proof.* Let us first consider the case where  $\psi$  is trivial. By the Goursat–Kolchin–Ribet criterion [12, Section 1.8], since the pair  $(\text{SL}_3(\mathbb{C}), \text{Std})_{i=1,2}$  is Goursat-adapted (in the language given there), it suffices to show that there exists no geometric isomorphism of the form

$$[+h]^* \mathcal{G}_A \cong [+h']^* \mathcal{G}_A \otimes \mathcal{L}, \quad h \neq h',$$

for a one-dimensional sheaf  $\mathcal{L}$ . Without loss of generality,  $h' = 0$ . Let us assume by contradiction that  $h \neq 0$ . Since the left-hand side is ramified exactly at  $-h$  and  $\infty$ ,  $\mathcal{L}$  must be ramified at  $0$  and  $-h$  (and possibly at  $\infty$ ). It follows that

$$\mathcal{G}_A^{I_0} \cong ([+h]^* \mathcal{G}_A)^{I-h} \cong (\mathcal{G}_A \otimes \mathcal{L})^{I-h} \cong \mathcal{G}_A \otimes \mathcal{L}^{I-h} = 0.$$

However, this contradicts Lemma 3.3(3), which states that the stalk at  $0$  of  $\mathcal{G}_A$  is a rank 2 pseudoreflection.

If  $\psi$  is nontrivial, let us write  $\mathcal{H}_{A,h,\psi}^+ = \mathcal{L}_\psi \oplus \mathcal{F}$ , where we may assume by the above that the arithmetic and geometric monodromy groups of  $\mathcal{F}$  coincide and are isomorphic to  $\text{SL}_3(\mathbb{C})^H$ . We have

$$G_{\text{geom}} \left( \mathcal{H}_{A,h,\psi}^+ \right) \leq G_{\text{arith}} \left( \mathcal{H}_{A,h,\psi}^+ \right) \leq \mathbb{F}_p \times \text{SL}_3(\mathbb{C})^H,$$

and  $G_{\text{geom}}$  surjects onto both  $\mathbb{F}_p$  and  $\text{SL}_3(\mathbb{C})^H$ . The surjection onto  $\mathbb{F}_p$  implies that  $G_{\text{geom}}$  has at least  $p$  connected components, of which the component at the identity contains  $\text{SL}_3(\mathbb{C})^H$ . It follows that  $G_{\text{geom}} = \mathbb{F}_p \times \text{SL}_3(\mathbb{C})^H$ , as claimed. □

This finishes altogether the proof of Theorem 3.1.

### 4. Correlations of $K_2$ -sums to Prime Power Moduli: Stationary Phase Methods

In this section, we complement the results of the previous section, which applied to correlations of  $K_2$  sums to prime moduli, with a treatment of  $K_2$  sums to higher prime power moduli.

Before stating Theorem 4.1, we need a few elements of notation. Fix  $n \geq 2$  and  $p > 3$ . Let  $N, M \geq 0$  with  $K := N + M \geq 1$  and suppose  $\mathbf{h} \in \mathbb{Z}^N$  and  $\mathbf{h}' \in \mathbb{Z}^M$ . Similar to the previous section, given  $\tau \in \mathbb{Z}/p^n\mathbb{Z}$  we define

$$\begin{aligned} \mu(\tau) &= \mu_{\mathbf{h}}(\tau) := |\{1 \leq j \leq N : h_j \equiv \tau \pmod{p^n}\}| \\ \nu(\tau) &= \nu_{\mathbf{h}'}(\tau) := |\{1 \leq j \leq M : h'_j \equiv \tau \pmod{p^n}\}|, \end{aligned}$$

and we define

$$T = T_{\mu, \nu} := \{\tau \in \mathbb{Z}/p^n\mathbb{Z} : \mu(\tau) + \nu(\tau) \geq 1\}.$$

We will prove the following estimates.

**Theorem 4.1.** *Let  $p > 3$  be prime and let  $n \geq 2$ . Let  $N, M \geq 0$  with  $N + M \geq 1$  and let  $\mathbf{h} \in \mathbb{Z}^N$  and  $\mathbf{h}' \in \mathbb{Z}^M$ . Let  $c \in \mathbb{Z}/p^n\mathbb{Z}$ ,  $a \in (\mathbb{Z}/p^n\mathbb{Z})^\times$  and let  $\mu = \mu_{\mathbf{h}}$ ,  $\nu = \nu_{\mathbf{h}'}$  and  $T = T_{\mathbf{h}, \mathbf{h}'}$  be defined as above. Further, put*

$$\rho = 1_{p \leq 3|T|/2-1} + \left\lceil \frac{\log(20(N + M)^3)}{\log p} \right\rceil,$$

and if  $n > (N + M)^3 2^{N+M}$ , assume in addition that

$$\min\{|\tau - \tau'|_p : \tau, \tau' \in T, \tau \neq \tau'\} \geq p^{-2|T|^{-2}(\lfloor 2^{-|T|}n \rfloor - \rho)},$$

where  $|x|_p$  denotes the  $p$ -adic absolute value of  $x \in \mathbb{Q}$ . Then

$$\begin{aligned} & \sum_{b \pmod{p^n}} e_{p^n}(cb) \prod_{1 \leq i \leq N} K_2(a, b + h_i; p^n) \prod_{1 \leq j \leq M} \overline{K}_2(a, b + h'_j; p^n) \\ & \ll_{N, M, \varepsilon} p^{(N+M+2)n/2} \begin{cases} p^{-\frac{1}{n^2(n-1)^2 + \varepsilon}} & \text{if } n \leq (N + M)^3 2^{N+M} \\ p^{-n2^{-N-M} + 1} & \text{if } n > (N + M)^3 2^{N+M}, \end{cases} \end{aligned}$$

unless either

- (i)  $p \equiv 2 \pmod{3}$ ,  $c = 0$  and  $\mu(\tau) = \nu(\tau)$  for all  $\tau \in T$ ,
- (ii)  $p \equiv 1 \pmod{3}$ ,  $c = 0$  and  $3 | (\mu(\tau) - \nu(\tau))$  for all  $\tau \in T$  or
- (iii)  $p \equiv 1 \pmod{3}$ ,  $p^{n-1} | c$  and  $3 | (\mu(\tau) - \nu(\tau))$ ; in this case, the bound is  $O_{N, M}(p^{(N+M+2)n/2-1/2})$ .

**Remark 4.2.** The condition  $3 | (\mu(\tau) - \nu(\tau))$  in (ii) of the above statement is the same condition that arose in connection with the trivial bound in Theorem 3.1. This condition, along with (i) and (iii), arises in this context for the following reason. Modulo prime powers  $p^n$  with  $n \geq 2$ , a complete sum  $K_2(a, b; p^n)$  may be explicitly evaluated as a sum over the set of critical points of a certain  $p$ -adic phase function. The correlations of the  $K_2$  sums therefore expand as a linear combination of exponential sums modulo  $p^n$ , each with a fixed frequency. The trivial bound cannot be improved whenever one of these frequencies is zero, and this occurs precisely in the degenerate cases listed in the statement of Theorem 4.1 (this is treated in Proposition 4.3).

The remainder of the section is devoted to proving this theorem.

4.1. Preparation

Fix  $a \in (\mathbb{Z}/p^n\mathbb{Z})^\times$ ,  $c \in \mathbb{Z}/p^n\mathbb{Z}$ ,  $\mathbf{h} \in \mathbb{Z}^N$  and  $\mathbf{h}' \in \mathbb{Z}^M$  and define  $\mu, \nu$  and  $T$  as above. Set

$$S_{p^n}(\mathbf{h}, \mathbf{h}'; c, a) := \sum_b \sum_{(\text{mod } p^n)} e_{p^n}(cb) \prod_{1 \leq i \leq N} K_2(a, b + h_i; p^n) \prod_{1 \leq j \leq M} \overline{K}_2(a, b + h'_j; p^n).$$

We can rewrite this expression as

$$\begin{aligned} S_{p^n}(\mathbf{h}, \mathbf{h}'; c, a) &= \sum_d \sum_{(\text{mod } p)} \sum_{\substack{b \pmod{p^n} \\ b \equiv d \pmod{p}}} e_{p^n}(cb) \prod_{\tau \in T} K_2(a, b + \tau; p^n)^{\mu(\tau)} \overline{K}_2(a, b + \tau; p^n)^{\nu(\tau)} \\ &=: \sum_d \sum_{(\text{mod } p)} \mathcal{S}_{p^n}(c, \mu, \nu; a, d). \end{aligned}$$

We deduce from Lemma 2.1 that unless  $4a^2(b + \tau) \equiv 4a^2(d + \tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3}$  (i.e., the cube of a residue class prime to  $p$ ) for all  $\tau \in T$  we get that  $\mathcal{S}_{p^n}(c, \mu, \nu; a, d) = 0$ . We will henceforth assume the condition

$$(\diamond) : 4a^2(d + \tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \text{ for all } \tau \in T,$$

which, as  $T$  and  $a$  are fixed, depends only on  $d$ .

Now, depending on whether  $p \equiv 1 \pmod{3}$  or  $p \equiv 2 \pmod{3}$ , if  $d$  satisfies  $(\diamond)$ , then there are either

- exactly 3 critical points, when  $p \equiv 1 \pmod{3}$ , or
- exactly one critical point, when  $p \equiv 2 \pmod{3}$ .

Letting  $u_0$  be a primitive cube root modulo  $p^{\lfloor n/2 \rfloor}$  (by Hensel’s lemma, this is necessarily a lift of a primitive cube root modulo  $p$ ), we may define a fixed branch of the cube root  $r \mapsto s(r)$  such that  $s(r)^3 \equiv r \pmod{p}$  whenever  $r \in (\mathbb{Z}/p\mathbb{Z})^{\times 3}$  and lift this branch to a branch of cube root modulo  $p^{\lfloor n/2 \rfloor}$  as well; then, if  $x^3 \equiv r \pmod{p^{\lfloor n/2 \rfloor}}$ , we obtain that

$$x \equiv s(r)u_0^j \pmod{p^m} \text{ for some } 0 \leq j \leq d_p - 1,$$

setting  $d_p = 1$  if  $p \equiv 2 \pmod{3}$  and  $d_p = 3$  if  $p \equiv 1 \pmod{3}$ . In this way, we can write (in the notation of Lemma 2.1)

$$K_2(a, b + \tau; p^n) = \left(\frac{3a}{p^n}\right) \varepsilon_{p,n} p^{n/2} \sum_{0 \leq j \leq d_p - 1} e_{p^n}(3as(\overline{2a}(b + \tau))^2 u_0^j).$$

Fix  $d$  satisfying  $(\diamond)$ . For ease of notation, in the sequel we will write

$$\mathcal{S}_{p^n}^*(c, \mu, \nu; a, d) = \left(\frac{3a}{p^n}\right)^{N+M} \frac{1}{\varepsilon_{p,n}^{N+M}} \mathcal{S}_{p^n}(c, \mu, \nu; a, d)$$

and study  $\mathcal{S}_{p^n}^*$  as  $d$  varies.

For each  $\tau \in T$  set

$$U(\tau) := \{0, \dots, d_p - 1\}^{\mu(\tau)}, \quad V(\tau) := \{0, \dots, d_p - 1\}^{\nu(\tau)},$$

as well as

$$\mathcal{U} := \prod_{\tau \in T} U(\tau), \quad \mathcal{V} := \prod_{\tau \in T} V(\tau).$$

With these notations, we can write

$$\begin{aligned} \mathcal{S}_{p^n}^*(c, \mu, \nu, a, d) &= p^{(N+M)n/2} \sum_{\substack{j=(j(\tau))_{\tau \in T} \in \mathcal{U} \\ j'=(j'(\tau))_{\tau \in T} \in \mathcal{V}}} \sum_{\substack{b \pmod{p^n} \\ b \equiv d \pmod{p}}} e_{p^n} \left( bc + 3a \sum_{\tau \in T} s(\overline{2a}(b + \tau))^2 \right) \\ &\quad \times \left( \sum_{1 \leq i \leq \mu(\tau)} u_0^{j_i(\tau)} - \sum_{1 \leq i' \leq \nu(\tau)} u_0^{j'_{i'}(\tau)} \right). \end{aligned}$$

For notational convenience, we simplify the above expression further as follows. Given  $\epsilon = (\epsilon_\tau)_{\tau \in T} \in (\mathbb{Z}/p^n\mathbb{Z})^{|T|}$  and  $b \in \mathbb{Z}/p^n\mathbb{Z}$ , define

$$f_{T, \epsilon}(b) := bc + 3a \sum_{\tau \in T} \epsilon_\tau s(\overline{2a}(b + \tau))^2. \tag{19}$$

Provided  $d$  satisfies  $(\diamond)$ , we may thus write

$$\mathcal{S}_{p^n}^*(c, \mu, \nu; a, d) = p^{(N+M)n/2} \sum_{\epsilon \in (\mathbb{Z}/p^n\mathbb{Z})^{|T|}} \phi(\epsilon) \sum_{\substack{b \pmod{p^n} \\ b \equiv d \pmod{p}}} e_{p^n}(f_{T, \epsilon}(b)),$$

where we have set

$$\phi(\epsilon) := \left| \left\{ (j, j') \in \mathcal{U} \times \mathcal{V} : \epsilon_\tau = \sum_i u_0^{j_i(\tau)} - \sum_{i'} u_0^{j'_{i'}(\tau)} \text{ for all } \tau \in T \right\} \right| \geq 0,$$

for each  $\epsilon \in (\mathbb{Z}/p^n\mathbb{Z})^{|T|}$ . Note that

$$\sum_{\epsilon \in (\mathbb{Z}/p^n\mathbb{Z})^{|T|}} \phi(\epsilon) = |\mathcal{U}| |\mathcal{V}| \ll 3^{N+M},$$

and if  $p \equiv 2 \pmod{3}$ , then  $\phi(\epsilon) = 1$  if  $\epsilon_\tau = \mu(\tau) - \nu(\tau)$  for all  $\tau \in T$  and 0 otherwise.

We now separate the  $\epsilon = \mathbf{0}$  term from the remaining  $\epsilon$ , so that<sup>8</sup>

$$\begin{aligned} &\sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^*(c, \mu, \nu; a, d) \\ &= \sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*, \neq \mathbf{0}}(c, \mu, \nu; a, d) + \phi(\mathbf{0}) p^{(N+M)n/2} \sum_{d \pmod{p}}^\diamond \sum_{\substack{b \pmod{p^n} \\ b \equiv d \pmod{p}}} e_{p^n}(f_{T, \mathbf{0}}(b)) \\ &=: \sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*, \neq \mathbf{0}}(c, \mu, \nu; a, d) + \sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*, \mathbf{0}}(c, \mu, \nu; a, d). \end{aligned}$$

### 4.2. The $\epsilon = \mathbf{0}$ Terms

The contribution to  $\mathcal{S}_{p^n}(h, h'; c, a)$  from  $\epsilon = \mathbf{0}$  can be estimated as follows.

**Proposition 4.3.** *With the above notation, we have*

$$\sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*, \mathbf{0}}(c, \mu, \nu; a, d) = 0$$

<sup>8</sup>We use the notation  $\sum_{d \pmod{p}}^\diamond$  to denote a sum over  $d \pmod{p}$  satisfying  $(\diamond)$  (for  $a$  and  $T$  fixed).

unless  $p^{n-1} | c$  and  $3 | (\mu(\tau) - \nu(\tau))$  for all  $\tau \in T$ . In this latter case, the following nontrivial bounds hold:

$$\sum_{d \pmod{p}}^{\diamond} \mathcal{S}_{p^n}^{*,0}(c, \mu, \nu; a, d) \ll_{N,M} p^{(N+M+2)n/2} \cdot \begin{cases} p^{-1/2} & : p \equiv 1 \pmod{3}, c \neq 0, 3 | (\mu(\tau) - \nu(\tau)) \forall \tau \in T \\ p^{-1} & : p \equiv 2 \pmod{3}, c \neq 0, \mu(\tau) = \nu(\tau) \forall \tau \in T. \end{cases}$$

To prove this result, we need several lemmas.

**Lemma 4.4.** *Let  $p \equiv 1 \pmod{3}$  be prime and let  $n \geq 2$  and  $D \geq 1$ . Let  $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{Z}$  with  $\max\{|\alpha_1|, |\alpha_2|, |\alpha_3|\} \leq D$ . Assume that  $p^{n/2} > 20D^3$  and let  $1 \leq u_0 \leq p^{\lfloor n/2 \rfloor} - 1$  be a primitive cube root modulo  $p^{\lfloor n/2 \rfloor}$ . Then either  $\alpha_1 = \alpha_2 = \alpha_3$  or else*

$$v_p(\alpha_1 + \alpha_2 u_0 + \alpha_3 u_0^2) < \left\lceil \frac{\log(20D^3)}{\log p} \right\rceil.$$

*Proof.* Set  $r := \left\lceil \frac{\log(20D^3)}{\log p} \right\rceil \leq n/2$ . Assume the claim is false, so that  $p^r | (\alpha_1 + \alpha_2 u_0 + \alpha_3 u_0^2)$  and  $\alpha_1, \alpha_2, \alpha_3$  are not all the same. As  $p \equiv 1 \pmod{3}$  and  $u_0$  is a primitive cube root, we must have  $u_0 \not\equiv 1 \pmod{p^{\lfloor n/2 \rfloor}}$ . By assumption, at least one of  $\alpha_1, \alpha_2, \alpha_3$  does not vanish and up to multiplication by  $u_0^j$  for  $j \in \{0, 1, 2\}$  (which does not change the  $p$ -adic valuation), we may suppose  $\alpha_3$  does not. As  $u_0^3 \equiv 1 \pmod{p^r}$ , we have

$$\begin{aligned} \alpha_1 + \alpha_2 u_0 + \alpha_3 u_0^2 &= (\alpha_1 - \alpha_3) + (\alpha_2 - \alpha_3)u_0 + \alpha_3(1 + u_0 + u_0^2) \\ &\equiv (\alpha_1 - \alpha_3) + (\alpha_2 - \alpha_3)u_0 \pmod{p^r}. \end{aligned}$$

Now, on one hand this implies that

$$\alpha_1 - \alpha_3 \equiv -u_0(\alpha_2 - \alpha_3) \pmod{p^r}. \tag{20}$$

On the other, taking cubes, we obtain that

$$(\alpha_1 - \alpha_3)^3 + (\alpha_2 - \alpha_3)^3 \equiv 0 \pmod{p^r}.$$

However, as

$$|(\alpha_1 - \alpha_3)^3 + (\alpha_2 - \alpha_3)^3| \leq 2 \cdot (2D)^3 < 20D^3 \leq p^r,$$

it follows that  $(\alpha_1 - \alpha_3)^3 = -(\alpha_2 - \alpha_3)^3$  and therefore that  $\alpha_1 - \alpha_3 = -(\alpha_2 - \alpha_3)$ . Plugging this into the earlier congruence implies that either  $\alpha_1 - \alpha_3 = \alpha_2 - \alpha_3 = 0$ , or else  $u_0 \equiv 1 \pmod{p^r}$ . Since the  $\alpha_j$  are not all the same by assumption, the first of these is impossible. To see that the second fails as well, note that if  $u_0 \not\equiv 1 \pmod{p^{\lfloor n/2 \rfloor}}$  but  $u_0 = 1 + mp^r$  for some  $m \in \mathbb{Z}$ , then of course  $p^{\lfloor n/2 \rfloor - r} \nmid m$ . On the other hand, since  $u_0$  is a cube root of unity modulo  $p^{\lfloor n/2 \rfloor}$ ,

$$u_0^3 = (1 + mp^r)^3 = 1 + mp^r(3 + 3p^r m + m^2 p^{2r}) \equiv 1 \pmod{p^{\lfloor n/2 \rfloor}},$$

which is only possible for  $p \neq 3$  if  $v_p(m) \geq \lfloor n/2 \rfloor - r$ , a contradiction. The claim follows. □

**Lemma 4.5.** *Let  $T \subseteq \mathbb{Z}/p^n\mathbb{Z}$  and suppose  $\phi(\mathbf{0}) \neq 0$ . Assume furthermore that  $p^n > 20(N + M)^3$ .*

- a) *If  $p \equiv 2 \pmod{3}$ , then  $\mu(\tau) = \nu(\tau)$  for all  $\tau \in T$ .*
- b) *If  $p \equiv 1 \pmod{3}$ , then  $\mu(\tau) \equiv \nu(\tau) \pmod{3}$  for all  $\tau \in T$ .*

*Proof.* In each case,  $\phi(\mathbf{0}) \neq 0$  if and only if there exist tuples  $\mathbf{j} \in \mathcal{U}, \mathbf{j}' \in \mathcal{V}$  such that

$$0 = \sum_{1 \leq i \leq \mu(\tau)} u_0^{j_i(\tau)} - \sum_{1 \leq i' \leq \nu(\tau)} u_0^{j'_{i'}(\tau)} \text{ for all } \tau \in T.$$

- a) When  $p \equiv 2 \pmod{3}$ , we may (trivially) take  $u_0 = 1$  and  $j_i(\tau) = j'_{i'}(\tau) = 0$  for all  $i, i'$  and  $\tau$ , which leads immediately to the conclusion  $\mu(\tau) = \nu(\tau)$  for all  $\tau \in T$ .
- b) When  $p \equiv 1 \pmod{3}$ , we write the above expression as

$$0 = \left(m_\tau + n_\tau u_0 + p_\tau u_0^2\right) - \left(m'_\tau + n'_\tau u_0 + p'_\tau u_0^2\right) = (m_\tau - m'_\tau) + (n_\tau - n'_\tau)u_0 + (p_\tau - p'_\tau)u_0^2,$$

where  $m_\tau, n_\tau, p_\tau$  and  $m'_\tau, n'_\tau, p'_\tau$  are nonnegative integers satisfying

$$m_\tau + n_\tau + p_\tau = \mu(\tau) \qquad m'_\tau + n'_\tau + p'_\tau = \nu(\tau),$$

for all  $\tau \in T$ . By Lemma 4.4, it follows that  $m_\tau - m'_\tau = n_\tau - n'_\tau = p_\tau - p'_\tau = \ell$ , say. We conclude that

$$\mu(\tau) - \nu(\tau) = (m_\tau + n_\tau + p_\tau) - (m'_\tau + n'_\tau + p'_\tau) = 3\ell,$$

which implies the claim. □

**Lemma 4.6.** *Let  $p > 3$  be prime and let  $C \in \mathbb{Z}/p\mathbb{Z}$ . Let  $A \in (\mathbb{Z}/p\mathbb{Z})^\times$  and let  $\tilde{T} \subseteq \mathbb{Z}/p\mathbb{Z}$ .*

- a) *If  $p \equiv 2 \pmod{3}$ , then*

$$\sum_{\substack{d \pmod{p} \\ 4A^2(d+\tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \\ \forall \tau \in \tilde{T}}} e_p(dC) = p1_{C \equiv 0 \pmod{p}} + O(|\tilde{T}|).$$

- b) *If  $p \equiv 1 \pmod{3}$ , then*

$$\sum_{\substack{d \pmod{p} \\ 4A^2(d+\tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \\ \forall \tau \in \tilde{T}}} e_p(dC) = 3^{-|\tilde{T}|} p1_{C \equiv 0 \pmod{p}} + O\left(|\tilde{T}|^2 p^{1/2}\right).$$

*Proof.* a) When  $p \equiv 2 \pmod{3}$  every residue class modulo  $p$  is a cube. Thus, if  $d$  satisfies the condition in the sum on the left-hand side, then as  $p > 3$  this is equivalent to  $\prod_{\tau \in \tilde{T}} (d + \tau) \in (\mathbb{Z}/p\mathbb{Z})^\times$ , which is satisfied for all but  $O(|\tilde{T}|)$  residue classes  $d$  modulo  $p$ . Thus, the sum in question is simply

$$\sum_{d \pmod{p}} e_p(dC) + O(|\tilde{T}|) = p1_{C \equiv 0 \pmod{p}} + O(|\tilde{T}|),$$

as required.

b) Let  $\Xi_3(p) := \{\chi \pmod{p} : \chi^3 = \chi_0\}$ , where  $\chi_0$  denotes the trivial multiplicative character modulo  $p$ . Since the set of multiplicative characters modulo  $p$  is a cyclic group of order  $p - 1$  and  $3|(p - 1)$ , we can write

$$\Xi_3(p) = \{\xi^j : j \in \{-1, 0, 1\}\}, \text{ where } \xi := \chi_1^{(p-1)/3}$$

for some fixed generator  $\chi_1$  for the group of characters mod  $p$ . We note that for any  $b \in \mathbb{Z}/p\mathbb{Z}$ ,

$$\sum_{-1 \leq j \leq 1} \xi^j(b) = \begin{cases} 3 & \text{if } b \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \\ 0 & \text{otherwise} \end{cases},$$

and so the exponential sum in question is

$$\begin{aligned} & \sum_{d \pmod p} e_p(dC) \prod_{\tau \in \tilde{T}} 1_{4A^2(d+\tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3}} \\ &= 3^{-|\tilde{T}|} \sum_{j \in \{-1,0,1\}^{|\tilde{T}|}} \xi(4A^2)^{t(j)} \sum_{d \pmod p} \xi \left( \prod_{\tau \in \tilde{T}} (d+\tau)^{j_\tau} \right) e_p(dC), \end{aligned} \tag{21}$$

where we have written  $t(j) := \sum_{\tau \in \tilde{T}} j_\tau$ .

When  $j = \mathbf{0}$  we get

$$\begin{aligned} & 3^{-|\tilde{T}|} \sum_{d \pmod p} \chi_0 \left( \prod_{\tau \in \tilde{T}} (d+\tau) \right) e_p(dC) \\ &= 3^{-|\tilde{T}|} \sum_{d \pmod p} e_p(dC) + O(|\tilde{T}|3^{-|\tilde{T}|}) = 3^{-|\tilde{T}|} (p1_{C \equiv 0 \pmod p} + O(|\tilde{T}|)). \end{aligned} \tag{22}$$

For  $j \neq \mathbf{0}$ , we define

$$g_j(d) := \prod_{\substack{\tau \in \tilde{T} \\ j_\tau = 1}} (d+\tau), \quad h_j(d) := \prod_{\substack{\tau \in \tilde{T} \\ j_\tau = -1}} (d+\tau),$$

and what remains to be estimated is the expression

$$3^{-|\tilde{T}|} \sum_{\substack{j \in \{-1,0,1\}^{|\tilde{T}|} \\ j \neq \mathbf{0}}} \sum_{\substack{d \pmod p \\ p \nmid (d+\tau) \forall \tau \in \tilde{T}}} \xi(g_j(d)\bar{h}_j(d))e_p(dC) \ll \max_{j \neq \mathbf{0}} \left| \sum_{\substack{d \pmod p \\ p \nmid (d+\tau) \forall \tau \in \tilde{T}}} \xi(g_j(d)\bar{h}_j(d))e_p(dC) \right|.$$

We interpret the latter sum as a sum over  $\mathbb{F}_p$  and estimate the sum using the language of  $\ell$ -adic trace functions, for an auxiliary prime  $\ell \neq p$ . For each  $j \neq \mathbf{0}$  let  $\mathcal{K}_j$  denote the Kummer sheaf attached to  $\xi \circ (g_j/h_j)$  and let  $\mathcal{L}_C$  be the Artin–Schreier sheaf attached  $d \mapsto e_p(dC)$ . Then the sum we must estimate is

$$\sum_{d \in U_j(\mathbb{F}_p)} t_{\mathcal{K}_j}(d)t_{\mathcal{L}_C}(d) = \sum_{d \in U_j(\mathbb{F}_p)} t_{\mathcal{F}}(d),$$

where  $\mathcal{F} := \mathcal{K}_j \otimes \mathcal{L}_C$  and  $U_j(\mathbb{F}_p) := \{d \in \mathbb{F}_p : d \neq -\tau \forall \tau \in \tilde{T}\}$ . By Corollary 2.31 of [17], this is

$$p1_{\mathcal{F} \text{ geom. trivial}} + O(\text{cond}(\mathcal{F})^2 p^{1/2}).$$

When  $j \neq \mathbf{0}$ , we claim that  $\mathcal{F}$  is not geometrically trivial; that is, that  $\mathcal{K}_j$  and  $D(\mathcal{L}_C)$  are not geometrically isomorphic. Indeed, on one hand, if  $C \neq 0$ , then  $\mathcal{L}_C$  (and thus  $D(\mathcal{L}_C)$ ) has a lone wild ramification point at  $\infty$  (with  $\text{Swan}_\infty(\mathcal{L}_C) = 1$ ), while in contrast  $\mathcal{K}_j$  has only tame ramification points at the zeros of  $g_j h_j$  (and, in particular,  $\text{Swan}_\infty(\mathcal{K}_j) = 0$ ). If  $C = 0$ , then because  $g_j(d)h_j(d)$  has only distinct roots and is nonconstant, it is not the cube of a polynomial. Thus,  $\mathcal{K}_j$  is ramified in at least one point and hence not geometrically trivial in this case as well. Finally,

$$\text{cond}(\mathcal{F}) = \text{cond}(\mathcal{K}_j)\text{cond}(\mathcal{L}_C) \leq 3(\deg g_j + \deg h_j + 1) \ll |\tilde{T}|,$$

so that indeed

$$\sum_{\substack{d \pmod{p} \\ p \nmid g_j(d)h_j(d) \\ p \nmid (d+\tau) \forall \tau \in T}} \xi(g_j(d)\bar{h}_j(d))e_p(dC) \ll |\tilde{T}|^2 p^{1/2},$$

for all  $j \neq 0$ . Inserting this and (22) into (21) and summing over  $j$ , we reach the claim. □

*Proof of Proposition 4.3.* Suppose  $\epsilon = 0$ . Since  $f_{T,0}(b) = cb$ , we have

$$\begin{aligned} \sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*,0}(c, \mu, \nu; a, d) &= \phi(0)p^{(N+M)n/2} \sum_{d \pmod{p}}^\diamond \sum_{t \pmod{p^{n-1}}} e_{p^n}(c(pt+d)) \\ &= \phi(0)p^{(N+M+2)n/2-1} 1_{p^{n-1}|c} \sum_{d \pmod{p}}^\diamond e_p(dc'), \end{aligned}$$

where  $c' = 0$  if  $p^{n-1} \nmid c$  and otherwise  $c \equiv c'p^{n-1} \pmod{p^n}$ . By Lemma 4.6, we have

$$\sum_{d \pmod{p}}^\diamond e_p(dc') = 3^{-|\tilde{T}|} 1_{p \equiv 1 \pmod{3}} p 1_{c' \equiv 0 \pmod{p}} + O_{|\tilde{T}|}(p^{1/2} 1_{p \equiv 1 \pmod{3}} + 1_{p \equiv 2 \pmod{3}}),$$

where  $\tilde{T} := \{\tau \pmod{p} : \tau \in T\}$ . As  $|\tilde{T}| \leq |T| \leq N + M$ , the above expression is thus

$$\ll_{N,M} p^{(N+M+2)n/2} \left( 1_{c \equiv 0 \pmod{p^n}} + 1_{p^{n-1}|c} \left( p^{-1/2} 1_{p \equiv 1 \pmod{3}} + p^{-1} 1_{c=0} 1_{p \equiv 2 \pmod{3}} \right) \right).$$

By Lemma 4.5, this gives

$$\begin{aligned} \sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*,0}(c, \mu, \nu; a, d) &\ll_{N,M} p^{(N+M+2)n/2} 1_{p \equiv 1 \pmod{3}} \left( 1_{c \equiv 0 \pmod{p^n}} + p^{-1/2} 1_{p^{n-1}|c} \right) \prod_{\tau \in \tilde{T}} 1_{3|(\mu(\tau)-\nu(\tau))} \\ &\quad + p^{(N+M+2)n/2} 1_{p \equiv 2 \pmod{3}} \left( 1_{c \equiv 0 \pmod{p^n}} + p^{-1} 1_{p^{n-1}|c} \right) \prod_{\tau \in \tilde{T}} 1_{\mu(\tau)=\nu(\tau)}, \end{aligned}$$

which implies the claim. □

In the next two subsections, we treat the estimation of

$$\sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*,\neq 0}(c, \mu, \nu; a, d)$$

in two ways: first, we will prove an estimate that is efficient when  $n$  is large and subsequently a different estimate that is most efficient when  $n$  is not large (but  $p$  is).

### 4.3. Bounds for Large $n$

The main result of this subsection is the following.

**Proposition 4.7.** *Let  $p > 3$  be prime. Assume that  $n \geq 2^{N+M}(N+M)^3$ . Then we have*

$$\sum_{d \pmod{p}}^\diamond \mathcal{S}_{p^n}^{*,\neq 0}(c, \mu, \nu; a, d) \ll_{N,M} p^{(N+M+2-2\delta_1)n/2}$$

with  $\delta_1 := 2^{-N-M}$ , unless there are at least two distinct  $\tau, \tau' \in T$  such that

$$\tau \equiv \tau' \pmod{p^{r_p(n)}},$$

where we define

$$r_p(n) := \left\lfloor \frac{2}{(N+M)(N+M-1)} \left( \lfloor 2^{-N-M} n \rfloor - 1_{p \leq 3(N+M)/2-1} - \left\lceil \frac{\log(20(N+M)^3)}{\log p} \right\rceil \right) \right\rfloor.$$

The above proposition will be proved using a method of Milićević and Zhang [15, Section 4], a consequence of which is the following proposition concerning exponential sums with argument function  $f_{T,\epsilon}$  (see (19) for the definition).

**Proposition 4.8.** *There exist constants  $\delta_i = \delta_i(|T|) > 0$  for  $i = 1, 2, 3$  and  $\rho = \rho(N, M, |T|) > 0$  such that the following holds.*

*For any positive integer  $n \geq 2$  such that  $p^{n/2} > 20(N+M)^3$ , any  $d \in \mathbb{Z}/p\mathbb{Z}$  and any nonzero  $\epsilon \in (\mathbb{Z}/p^n\mathbb{Z})^{|T|}$  with  $\phi(\epsilon) \neq 0$ , either*

o *the estimate*

$$\sum_{\substack{b \in \mathbb{Z}/p^n\mathbb{Z} \\ b \equiv d \pmod{p}}} e_{p^n}(f_{T,\epsilon}(b)) \ll_{|T|} p^{(1-\delta_1)n}$$

*holds or else*

o *there are at least two distinct  $\tau, \tau' \in T$  such that  $\epsilon_\tau, \epsilon_{\tau'} \not\equiv 0 \pmod{p^n}$  and*

$$\tau \equiv \tau' \pmod{p^{\lfloor \delta_3(\lfloor \delta_2 n \rfloor - \rho) \rfloor}}.$$

*In particular, the values  $\delta_1 = \delta_2 = 2^{-|T|}$ ,  $\delta_3 = \binom{|T|}{2}^{-1}$  and*

$$\rho = 1_{p \leq 3|T|/2-1} + \left\lceil \frac{\log(20(N+M)^3)}{\log p} \right\rceil$$

*are admissible.*

We need the following simple lemma about  $p$ -adic valuations of generalised binomial coefficients.

**Lemma 4.9.** *Let  $p > 3$ . Then for any  $k \geq 1$ ,*

$$0 \leq \max_{0 \leq j \leq k} v_p \left( \binom{2/3}{j} \right) \leq 1_{p \leq 3k/2-1}.$$

*Proof.* Notice that

$$\binom{2/3}{j} = (j!)^{-1} \prod_{0 \leq l \leq j-1} (2/3 - l) = \frac{(-1)^{j-1} 2}{3^j j!} \prod_{1 \leq l \leq j-1} (3l - 2),$$

and so as  $p > 3$ ,

$$v_p \left( \binom{2/3}{j} \right) = v_p \left( \prod_{1 \leq l \leq j-1} (3l - 2) \right) - v_p(j!).$$

Note that

$$\max_{0 \leq j \leq k} v_p \left( \binom{2/3}{j} \right) \geq v_p \left( \binom{2/3}{0} \right) \geq 0.$$

By Legendre’s formula for  $p$ -adic valuations of factorials, it suffices to show that

$$v_p \left( \prod_{1 \leq l \leq j-1} (3l - 2) \right) \leq \sum_{r \geq 1} \left\lfloor \frac{j}{p^r} \right\rfloor + 1_{p \leq 3j/2 - 1},$$

for each  $1 \leq j \leq k$ . To check this inequality, we observe first that if  $1 \leq l_0 \leq j$  is minimal such that  $p|(3l_0 - 2)$  and  $l > l_0$  also has this property then  $p|(l - l_0)$ . Clearly, by minimality we must have  $v_p(3l_0 - 2) = 1$ : if  $p \equiv 1 \pmod{3}$  then  $l_0$  satisfies  $p = 3l_0 - 2$ ; if  $p \equiv 2 \pmod{3}$  then  $2p = 3l_0 - 2$ . In particular, if  $p > 3j/2 - 1$ , then  $l_0$  does not exist.

Consider next when  $p \leq 3j/2 - 1$ . Then we have

$$v_p \left( \prod_{1 \leq l \leq j-1} (3l - 2) \right) = \sum_{r \geq 1} |\{1 \leq l \leq j - 1 : l \equiv 2\bar{3} \pmod{p^r}\}| \leq \sum_{r \geq 1} \left\lfloor \frac{j - a_r}{p^r} \right\rfloor,$$

where  $0 \leq a_r \leq p^r - 1$  is a minimal representative of the residue class  $2\bar{3} \pmod{p^r}$ . We clearly have  $\lfloor (j - a_r)/p^r \rfloor \leq \lfloor j/p^r \rfloor$ , so that, summing over  $r$ , we obtain the desired bound.  $\square$

The following further observations will be key. If  $b \in (\mathbb{Z}/p^n\mathbb{Z})^\times$ ,  $\kappa \in \mathbb{N}$  and  $t \in \mathbb{Z}/p^n\mathbb{Z}$ , we note that

$$\begin{aligned} s(\overline{2a}(b + \tau + p^\kappa t))^2 &= s(\overline{2a}(b + \tau)(1 + p^\kappa t/(b + \tau)))^2 \\ &= s(\overline{2a}(b + \tau))^2 \sum_{l \geq 0} \binom{2/3}{l} (p^\kappa t)^l (b + \tau)^{-l} = \sum_{l \geq 0} \binom{2/3}{l} (p^\kappa t)^l s(\overline{2a}(b + \tau))^{2-3l} \\ &\equiv s(\overline{2a}(b + \tau))^2 + \frac{2}{3} s(\overline{2a}(b + \tau))^{-1} p^\kappa t \pmod{p^{\min\{n, 2\kappa\}}}, \end{aligned}$$

where the second equality is owed to the convergence in the  $p$ -adic topology of the power series

$$(1 + x)^{2/3} = \sum_{l \geq 0} \binom{2/3}{l} x^l, \tag{23}$$

as long as  $|x|_p < 1$ . It follows from this that

$$\begin{aligned} f_{T, \epsilon}(b + p^\kappa t) &\equiv \left( bc + 3a \sum_{\tau \in T} \epsilon_\tau s(\overline{2a}(b + \tau))^2 \right) + p^\kappa t \left( c + 2a \sum_{\tau \in T} \epsilon_\tau s(\overline{2a}(b + \tau))^{-1} \right) \pmod{p^{\min\{2\kappa, n\}}}. \end{aligned} \tag{24}$$

Analogously, we also define the ‘derivative’ functions

$$f_{T, \epsilon}^{(j)}(b) := b^{1-j} c 1_{j \in \{0, 1\}} + 3a \binom{2/3}{j} \sum_{\tau \in T} \epsilon_\tau s(\overline{2a}(b + \tau))^{2-3j}.$$

*Proof of Proposition 4.8.* We begin by noting that if  $|\{\tau \in T : \epsilon_\tau \not\equiv 0 \pmod{p^n}\}| \leq 1$ , then since  $\epsilon$  is nonzero the  $p$ -adic stationary phase method (see, e.g., Lemma 1 (1) of [15]) implies that for some  $\tau_0 \in T$ ,

$$\begin{aligned} \sum_{\substack{b \in \mathbb{Z}/p^n\mathbb{Z} \\ b \equiv d \pmod{d}}} e_{p^n}(f_{T,\epsilon}(b)) &= \sum_{\substack{b \in \mathbb{Z}/p^n\mathbb{Z} \\ b \equiv d \pmod{d}}} e_{p^n}(bc + \epsilon_{\tau_0}s(\overline{2a}(b + \tau_0))^2) \\ &= p^{n/2} \sum_{\substack{b \in \mathbb{Z}/p^n\mathbb{Z} \\ b \equiv d \pmod{p} \\ \epsilon_{\tau_0}s(\overline{2a}(b + \tau_0))^2 \equiv -c \pmod{p^{\lfloor n/2 \rfloor}}} } e_{p^n}(bc + \epsilon_{\tau_0}s(\overline{2a}(b + \tau_0))^2) \\ &\ll p^{n/2}, \end{aligned}$$

since the inner sum has at most a single summand. As  $N + M \geq 1$  this satisfies the first claim, so henceforth we may assume that  $|\{\tau \in T : \epsilon_\tau \not\equiv 0 \pmod{p^n}\}| \geq 2$ .

The second condition is now vacuously satisfied if  $0 < \delta_2 n < 1$ , so we may assume that  $\delta_2 n \geq 1$ . We put  $X := \{b \pmod{p^n} : b \equiv d \pmod{p}\}$ , noting that this is  $p^{\lfloor \delta_2 n \rfloor} \mathbb{Z}/p^n \mathbb{Z}$ -invariant. Putting

$$\mathcal{E}_{|T|} := \left\{ b \in \mathbb{Z}/p^n\mathbb{Z} : f^{(j)}(b) \equiv 0 \pmod{p^{\lfloor \delta_2 n \rfloor}} \text{ for all } 1 \leq j \leq |T| \right\},$$

the proof of Proposition 8 of [15] shows that

$$\sum_{b \in X} e_{p^n}(f_{T,\epsilon}(b)) = \sum_{b \in X \cap \mathcal{E}_{|T|}} e_{p^n}(f_{T,\epsilon}(b)) + O_{|T|} \left( p^{(1-\delta_1)n} \right), \tag{25}$$

as long as (writing  $f_{T,\epsilon}^{(0)} = f_{T,\epsilon}$ ) the relations

$$f_{T,\epsilon}^{(j)}(b + p^\kappa t) \equiv f_{T,\epsilon}^{(j)}(b) + p^\kappa t f_{T,\epsilon}^{(j+1)}(b) \pmod{p^{\min\{2\kappa, n\}}}$$

hold for  $0 \leq j \leq |T| - 1$ ; this is guaranteed by a calculation analogous to (24). The proof of Proposition 9 there shows that if

$$\rho_0 := \max_{1 \leq j \leq |T|} \nu_p \left( \binom{\lfloor 2/3 \rfloor}{j} \right) + \min_{\tau \in T} \nu_p(\epsilon_\tau),$$

then  $\mathcal{E}_T = \emptyset$  whenever

$$\min\{\tau - \tau'_p : \tau, \tau' \in T, \tau \neq \tau', \epsilon_\tau, \epsilon_{\tau'} \not\equiv 0 \pmod{p^n}\} \geq p^{-\delta_3(\lfloor \delta_2 n \rfloor - \rho_0)}$$

or, equivalently, whenever

$$\tau \not\equiv \tau' \pmod{p^{\lfloor \delta_3(\lfloor \delta_2 n \rfloor - \rho) \rfloor}} \text{ for all distinct } \tau, \tau' \text{ with nonzero } \epsilon_\tau, \epsilon_{\tau'}.$$

Let us assume the latter condition. Then (25) yields the estimate

$$\sum_{b \in X} e_{p^n}(f_{T,\epsilon}(b)) \ll_{|T|} p^{(1-\delta_1)n},$$

and it remains to check that the required constraints on the constants hold. As noted above the proof of Proposition 8 in [15], we may take  $\delta_1 = \delta_2 = 2^{-|T|}$  and  $\delta_3 = \binom{|T|}{2}^{-1}$ . Now, if  $p \equiv 2 \pmod{3}$ , then  $\epsilon_\tau$  can be identified with an integer in  $[-N, M]$  and we deduce that  $\nu_p(\epsilon_\tau) \leq \lceil \log(2(N + M))/\log p \rceil$ , which is acceptable. Thus, consider when  $p \equiv 1 \pmod{3}$ . We may then write

$$\epsilon_\tau = \alpha_\tau + \beta_\tau u_0 + \gamma_\tau u_0^2,$$

and as  $\phi(\epsilon) \neq 0$  such a representation exists with  $|\alpha_\tau|, |\beta_\tau|, |\gamma_\tau| \leq N + M$  for all  $\tau \in T$ . Since  $\epsilon$  is nonzero we may find  $\tau' \in T$  such that  $\epsilon_{\tau'} \not\equiv 0 \pmod{p^n}$  and therefore  $\alpha_{\tau'}, \beta_{\tau'}$  and  $\gamma_{\tau'}$  are not the same. By Lemma 4.4 we get

$$\min_{\tau \in T} v_p(\epsilon_\tau) \leq v_p(\epsilon_{\tau'}) \leq \left\lceil \frac{\log(20(N + M)^3)}{\log p} \right\rceil.$$

Finally, by Lemma 4.9, we have

$$\max_{1 \leq j \leq |T|} v_p \left( \binom{2/3}{j} \right) \leq 1_{p \leq 3|T|/2-1}.$$

We thus obtain

$$\rho_0 \leq 1_{p \leq 3|T|/2-1} + \left\lceil \frac{\log(20(N + M)^3)}{\log p} \right\rceil =: \rho.$$

The claim then follows by replacing  $\rho_0$  with its upper bound  $\rho$  in the second alternative of the proposition, which relaxes the condition there. □

*Proof of Proposition 4.7.* Note that by the assumed lower bound for  $n$ , we must have  $p^{n=2} > 20(N+M)^3$ . We may partition the nonzero  $\epsilon$  into the sets

$$\mathcal{A}_T := \left\{ \epsilon \in (\mathbb{Z}/p^n\mathbb{Z})^{|T|} \setminus \{\mathbf{0}\} : \min_{\substack{\tau \neq \tau' \\ \epsilon_\tau, \epsilon_{\tau'} \neq 0 \pmod{p^n}}} |\tau - \tau'|_p \geq p^{-\delta_3(\lfloor \delta_2 n \rfloor - \rho)} \right\}$$

$$\mathcal{B}_T := (\mathbb{Z}/p^n\mathbb{Z})^{|T|} \setminus \{\mathbf{0}\} \setminus \mathcal{A}_T.$$

Noting that  $|T| \leq N + M$ , we apply Proposition 4.8 to bound

$$p^{(N+M)n/2} \sum_{d \pmod{p}}^\diamond \sum_{\epsilon \in \mathcal{A}_T} \phi(\epsilon) \sum_{\substack{b \pmod{p^n} \\ b \equiv d \pmod{p}}} e_{p^n}(f_T, \epsilon(b)) \ll_{N, M} p^{(N+M)n/2+1} \cdot p^{(1-\delta_1)n}$$

$$\ll p^{(N+M+2-2\delta_1)n/2+1}.$$

On the other hand, if  $\mathcal{B}_T \neq \emptyset$ , then the second alternative of the proposition holds. This implies the claim. □

#### 4.4. Bounds for Small $n$ and Large $p$

The above estimates are useful for  $n$  sufficiently large in terms of  $N, M$ . In this subsection we provide an estimate which is more efficient for  $n \ll_{N, M} 1$  but with  $p \gg_{N, M} 1$ . As before, we write

$$\sum_{d \pmod{p}}^\diamond S_{p^n}^{*, \neq 0}(c, \mu, \nu; a, d) = \sum_{d \pmod{p}}^\diamond \sum_{\substack{\epsilon \in (\mathbb{Z}/p^n\mathbb{Z})^{|T|} \\ \epsilon \neq \mathbf{0}}} \phi(\epsilon) \sum_{\substack{b \pmod{p^n} \\ b \equiv d \pmod{p}}} e_{p^n}(f_T, \epsilon(b)),$$

where  $T = T_{h, h'}$  is as above. The estimate we prove in this case is as follows.

**Proposition 4.10.** *Let  $\epsilon \neq \mathbf{0}$ . Assume that  $n \leq (N + M)^3 2^{N+M}$ . Then*

$$\sum_{d \pmod{p}}^\diamond \sum_{\substack{b \pmod{p^n} \\ b \equiv d \pmod{p}}} e_{p^n}(f_T, \epsilon(b)) \ll_{\epsilon, N, M} p^{n - \frac{1}{n^2(n-1)^2} + \epsilon}.$$

**Remark 4.11.** Note that the upper bound is trivially satisfied for  $p = O_{N,M}(1)$ , by choosing a suitable constant. We make no attempt to specify this dependence in the sequel, though this could be done in principle provided that one had an effective bound in the critical case  $s = \frac{1}{2}k(k + 1)$  in Vinogradov’s mean value theorem (see Theorem 4.13). For some work in this direction for  $s$  slightly larger, see [23].

In preparation for the proof of Proposition 4.10, note that from the identity (23) for  $|x|_p < 1$ , given  $b = d + pt$  with  $t \in \mathbb{Z}/p^{n-1}\mathbb{Z}$  we can write

$$\begin{aligned} f_{T,\epsilon}(d + pt) &= (d + pt)c + \sum_{\tau \in \tilde{T}} \epsilon_\tau s (\overline{2a}(d + \tau + pt))^2 \equiv \sum_{l=0}^{n-1} t^l a_l(\epsilon, d) \pmod{p^n} \\ &=: P_{\epsilon,d}(t) \pmod{p^n}, \end{aligned}$$

where we have put

$$a_l(\epsilon, d) := p^l \left( \binom{2/3}{l} (2a)^l \sum_{\tau \in \tilde{T}} \epsilon_\tau s (\overline{2a}(d + \tau))^{2-3l} + cd^{1-l} 1_{l \in \{0,1\}} \right).$$

Note that  $P_{\epsilon,d}(t)$  is a polynomial in  $t$  modulo  $p^{n-1}$ . To evaluate the exponential sum over  $b$  (and thus over  $t \pmod{p^{n-1}}$ ), we will, roughly speaking, split the set of  $d$  according to the degree (modulo  $p^n$ ) of  $P_{\epsilon,d}(t)$  and use bounds for Weyl sums of that degree. As we will not be able to extract cancellation when  $P_{\epsilon,d}$  has degree 0 or 1, we will need to check that the number of  $d$  (satisfying  $(\diamond)$ ) for which this happens is small. To this end, we need the following lemma, which is a modification of Proposition 4.8 of [20].

**Lemma 4.12.** *Let  $w \in \mathbb{Z}/p\mathbb{Z}$  and  $\tilde{T} \subset \mathbb{Z}/p\mathbb{Z}$ . Let also  $\tilde{\epsilon} \in (\mathbb{Z}/p\mathbb{Z})^{|\tilde{T}|} \setminus \{\mathbf{0}\}$ . Then for  $j = 1, 2$ ,*

$$\left\{ \{d \pmod{p} : 4a^2(d + \tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \forall \tau \in \tilde{T} \text{ and } \sum_{\tau \in \tilde{T}} \tilde{\epsilon}_\tau s (\overline{2a}(d + \tau))^{2-3j} \equiv w \pmod{p} \} \right\}$$

is  $O_{|\tilde{T}|}(1)$ .

*Proof.* The proof when  $j = 2$  is completely similar to that of  $j = 1$ , so we focus only on the latter case.

As in the proof of Lemma 4.6, we translate the problem to  $\mathbb{F}_p$ , so that, for example,  $\tilde{T}$  is identified with a subset of  $\mathbb{F}_p$ . Put  $\mathbb{F} := \mathbb{F}_p$  if  $p \equiv 1 \pmod{3}$  and  $\mathbb{F} := \mathbb{F}_p[X]/(X^2 + X + 1)\mathbb{F}_p[X]$  when  $p \equiv 2 \pmod{3}$ . Define also

$$N_{\tilde{T}}(w) := \left| \{d \in \mathbb{F} : 4a^2(d + \tau) \in \mathbb{F}^{\times 3} \forall \tau \in \tilde{T} \text{ and } \sum_{\tau \in \tilde{T}} \tilde{\epsilon}_\tau s ((2a)^{-1}(d + \tau))^{-1} = w \} \right|.$$

When  $p \equiv 1 \pmod{3}$ ,  $N_{\tilde{T}}(w)$  is precisely the count on the left-hand side in the statement of the lemma; when  $p \equiv 2 \pmod{3}$ ,  $N_{\tilde{T}}(w)$  is an upper bound for the desired quantity, since  $\mathbb{F}_p \subseteq \mathbb{F}$ . In what follows we let  $U_0$  denote a primitive cube root of unity in  $\mathbb{F}$ ; naturally, this satisfies  $1 + U_0 + U_0^2 = 0$ .

Let  $\mathbf{a} \in \mathbb{F}_p^{|\tilde{T}|}$ . Consider the product

$$Q_w(\mathbf{a}) := \prod_{j \in \{-1, 0, 1\}^{|\tilde{T}|}} \left( w - \sum_{\tau \in \tilde{T}} U_0^{j\tau} a_\tau \right),$$

which is a polynomial of total degree  $\leq 3^{|\tilde{T}|}$  in the variables  $(a_\tau)_{\tau \in \tilde{T}} \in \mathbb{F}_p^{|\tilde{T}|}$ . Note that  $Q_w(\mathbf{a}) = Q_w((a_\tau U_0^{j\tau})_{\tau \in \tilde{T}})$  for all  $j \in \{-1, 0, 1\}^{|\tilde{T}|}$ . This implies that we can find a polynomial  $\tilde{Q}_w$ , defined over

$\mathbb{F}$ , such that<sup>9</sup>  $Q_w(\mathbf{a}) = \tilde{Q}_w((a_\tau^3)_{\tau \in \tilde{T}})$ . In particular, we can write

$$Q_w(\mathbf{a}) = \sum_{\substack{r_1, \dots, r_{|\tilde{T}|} \geq 0 \\ r_1 + \dots + r_{|\tilde{T}|} \leq 3^{|\tilde{T}|-1}}} b_{r_1, \dots, r_{|\tilde{T}|}}(w) a_{\tau_1}^{3r_1} \dots a_{\tau_{|\tilde{T}|}}^{3r_{|\tilde{T}|}},$$

where  $\{\tau_1, \dots, \tau_{|\tilde{T}|}\}$  is an enumeration of  $\tilde{T}$ . Now, define

$$\begin{aligned} \tilde{R}_w(Y) &:= \left( \prod_{\tau \in \tilde{T}} (Y + \tau)^{3^{|\tilde{T}|-1}} \right) \cdot Q_w((\tilde{\epsilon}_\tau s((2a)^{-1}(Y + \tau))^{-1})_{\tau \in \tilde{T}}) \\ &= \sum_{\substack{r_1, \dots, r_{|\tilde{T}|} \geq 0 \\ r_1 + \dots + r_{|\tilde{T}|} \leq 3^{|\tilde{T}|-1}}} b_{r_1, \dots, r_{|\tilde{T}|}}(w) \prod_{1 \leq j \leq |\tilde{T}|} (2a\tilde{\epsilon}_{\tau_j}^3)^{r_j} \cdot \prod_{1 \leq j \leq |\tilde{T}|} (Y + \tau_j)^{3^{|\tilde{T}|-1} - r_j}, \end{aligned}$$

which is a polynomial in  $Y$  of degree  $\leq |\tilde{T}|3^{|\tilde{T}|-1} \ll_{|\tilde{T}|} 1$  over  $\mathbb{F}$ . Every  $d \in \mathbb{F}$  counted by  $N_{\tilde{T}}(w)$  is a root of  $\tilde{R}_w(Y)$  over  $\mathbb{F}$ , since it is a root of  $Q_w((\tilde{\epsilon}_\tau s((2a)^{-1}(Y + \tau))^{-1})_{\tau \in \tilde{T}})$ . It follows that, provided  $\tilde{R}_w(Y)$  is a nonzero polynomial, we get  $N_{\tilde{T}}(w) \ll_{|\tilde{T}|} 1$  as required. It therefore suffices to show that  $\tilde{R}_w(Y)$  is nonzero.

Assume first that  $w \neq 0$ . The leading coefficient (in  $Y$ ) of  $R_w(Y)$  is  $b_{0, \dots, 0}(w) = w^{3^{|\tilde{T}|}} \neq 0$ , so that  $\tilde{R}_w(Y)$  is necessarily nonzero in this case.

Suppose next that  $w = 0$ . Note that by hypothesis there is a  $\tau_1 \in \tilde{T}$  such that  $\tilde{\epsilon}_{\tau_1} \neq 0$ . Setting  $Y = -\tau_1$  and expanding the product we see that only the term with  $r_1 = 3^{|\tilde{T}|-1}$  (and  $r_j = 0$  for  $j \neq 1$ ) survives, arising as the monomial in  $a_{\tau_1}^{3^{|\tilde{T}|}}$  in  $Q_0(\mathbf{a})$ . Explicitly, this term has coefficient

$$[a_{\tau_1}^{3^{|\tilde{T}|}}]Q_0(\mathbf{a}) = \prod_{j \in \{-1, 0, 1\}} (-U_0^j)^{r_1} = -1.$$

It follows that  $b_{3^{|\tilde{T}|-1}, 0, \dots, 0} = -(2a\tilde{\epsilon}_{\tau_1}^3)^{3^{|\tilde{T}|-1}}$  and thus

$$\tilde{R}_0(-\tau_1) = -((2a)\tilde{\epsilon}_{\tau_1}^3)^{3^{|\tilde{T}|-1}} \in \mathbb{F}^\times.$$

The claim thus follows when  $w \neq 0$  as well. □

For larger degree polynomials we need bounds for Weyl sums. In contrast to the work in [20], where Weyl differencing is used to obtain cancellation, we will instead apply Vinogradov’s method in order to obtain a stronger Weyl sum estimate. For this, we recall the Vinogradov main conjecture, proved in the groundbreaking work of Bourgain, Demeter and Guth [2] (and independently in the work of Wooley [25]).

**Theorem 4.13.** (Theorem 1.1 of [2]). *Let  $k \geq 1, P \geq 1$ . Given  $\mathbf{x} \in [0, 1]^k$ , put*

$$f_k(\mathbf{x}, P) := \sum_{1 \leq n \leq P} e(x_1 n + x_2 n^2 + \dots + x_k n^k).$$

<sup>9</sup>This can be seen, for example, by noting that

$$3Q_w(\mathbf{a}) = Q_w((a_\tau)_\tau) + Q_w((a_\tau U_0^{1\tau'=\tau})_\tau) + Q_w((a_\tau U_0^{21\tau'=\tau})_\tau)$$

and then expanding the product and noting that only terms in  $a_{\tau'}^3$  survive, for each  $\tau' \in T$ .

Furthermore, for  $s \in \mathbb{N}$  define

$$J_{s,k}(P) := \int_{[0,1]^k} |f_k(\mathbf{x}, P)|^{2s} d\mathbf{x}.$$

Then

$$J_{s,k}(P) \ll_{\varepsilon} P^{\varepsilon} \left( P^s + P^{2s-k(k+1)/2} \right).$$

*Proof of Proposition 4.10.* Recall that

$$P_{\epsilon,d}(t) = \sum_{0 \leq j \leq n-1} a_j(\epsilon, d)t^j \text{ for } 0 \leq t \leq p^{n-1} - 1.$$

We wish to estimate

$$\sum_{d \pmod{p}}^{\diamond} \sum_{t \pmod{p^{n-1}}} e_{p^n}(P_{\epsilon,d}(t)).$$

Let  $R \geq 1, N := \frac{1}{2}p^{n-1}$ . Also, put

$$P_{\epsilon,d,t}(z) := P_{\epsilon,d}(t+z) = \sum_{0 \leq i \leq n-1} z^i \sum_{i \leq j \leq n-1} \binom{j}{i} t^{j-i} a_j(\epsilon, d) = \sum_{0 \leq i \leq n-1} A_i(\epsilon, d, t) z^i,$$

where we have defined

$$A_i(\epsilon, d, t) := p^i \sum_{i \leq j \leq n-1} (pt)^{j-i} \binom{j}{i} \left( \binom{2/3}{i} (2a)^j \sum_{\tau \in T} \epsilon_{\tau} s(\overline{2a}(d+\tau))^{2-3j} + cd^{1-j} 1_{j \in \{0,1\}} \right).$$

We define

$$F_{\epsilon}(d) := R^{-2} \sum_{t \pmod{p^n}} S_{\epsilon,d,t}(R) := R^{-2} \sum_{|t| \leq N} \sum_{1 \leq y, z \leq R} e_{p^n}(P_{\epsilon,d,t}(yz)) + O(R^2).$$

Set  $\ell := n(n-1)/2$ . We invoke Vinogradov’s method, as exposed in Subsection 8.5 of [10]. Given  $\alpha \in \mathbb{R}$  and  $Y \geq 1$ , define

$$D(\alpha, Y) := Y^{-2} \sum_{|m| \leq Y} \left| \sum_{|n| \leq Y} e(\alpha mn) \right|.$$

Now, for each  $|t| \leq \frac{1}{2}p^{n-1}$ , we obtain (cf. [10, equation (8.76)])

$$|S_{\epsilon,d,t}(R)| \leq \left( \ell^{2(n-1)} R^{4\ell(\ell-1)+n(n-1)} J_{\ell,n-1}^2(R) \Delta(t) \right)^{\frac{1}{2\ell^2}}, \tag{26}$$

where we have written

$$\Delta(t) := \prod_{1 \leq h \leq n-1} D(A_h(\epsilon, d, t)/p^n, \ell R^h).$$

Our goal is to extract savings over the trivial bound  $\Delta(t) \ll 1$ . To this end, we consider  $D(\alpha, Y)$  for  $\alpha$  a

$p$ -adic rational. Put  $\alpha = a/p^s$ , for  $(a, p) = 1$  and  $s \geq 1$ . We obtain

$$\begin{aligned} D(a/p^s, Y) &\ll Y^{-2} \sum_{|m| \leq Y} \min\{Y, \|am/p^s\|^{-1}\} \\ &= Y^{-2} \sum_{0 \leq r \leq s} \sum_{\substack{1 \leq |u| \leq \max\{1, \frac{1}{2} p^{s-r}\} \\ p \nmid u}} \sum_{\substack{|m| \leq Y \\ am \equiv u p^r \pmod{p^s}}} \min\{Y, p^{s-r}/|u|\} \\ &= Y^{-2} \sum_{0 \leq r \leq s-1} p^{s-r} \sum_{\substack{1 \leq |u| \leq \frac{1}{2} p^{s-r} \\ p \nmid u}} \frac{1}{|u|} \sum_{\substack{|m| \leq Y \\ am \equiv u p^r \pmod{p^s}}} 1 + Y^{-1} \sum_{\substack{|m| \leq Y \\ p^s | m}} 1 \\ &\ll Y^{-1} (p^s/Y + 1) \log(p^s) + p^{-s} \end{aligned}$$

in this case. We specialise  $Y = \ell R^h$  and

$$\alpha = A_h(\epsilon, d, t)/p^n = \tilde{A}_h/p^{n-\theta_h},$$

where  $\theta_h = \theta_h(\epsilon, d, t) := v_p(A_h(\epsilon, d, t))$  and  $p \nmid \tilde{A}_h$ , for each  $1 \leq h \leq n-1$ . Note that  $\theta_h(\epsilon, d, t) \geq h$ , with equality if and only if we have

$$(2a)^h \binom{2/3}{h} \sum_{\tau \in T} \epsilon_\tau s(\overline{2a}(d + \tau))^{2-3h} \not\equiv -c 1_{h=1} \pmod{p},$$

this condition being independent of  $t$ .

For  $\epsilon, d$  and  $t$  fixed, let

$$\mathfrak{D} = \mathfrak{D}(\epsilon, d, t) := \max\{1 \leq j \leq n-1 : \theta_j(\epsilon, d, t) < n\},$$

if this maximum is defined, and let  $\mathfrak{D}(\epsilon, d, t) = 0$  otherwise. If  $\mathfrak{D} \geq 1$ , then we obtain

$$\Delta(t) \ll_n R^{-\mathfrak{D}} (p^{n-\mathfrak{D}} R^{-\mathfrak{D}} + 1) \log p + p^{\theta_{\mathfrak{D}}-n}.$$

Suppose now that  $\mathfrak{D} \geq 2$ . Applying Theorem 4.13, we obtain from (26) that

$$\begin{aligned} |S_{\epsilon, d, t}(R)| &\ll_{\epsilon, n} \left( R^{4\ell(\ell-1)+n(n-1)+\epsilon} \left( R^{2\ell} + R^{4\ell-n(n-1)} \right) \left( R^{-\mathfrak{D}} (p^{n-\mathfrak{D}} R^{-\mathfrak{D}} + 1) + p^{\theta_{\mathfrak{D}}-n} \right) \right)^{\frac{1}{2\ell^2}} \\ &\ll R^{2+\epsilon} \left( R^{-\mathfrak{D}/2\ell^2} \left( (p^{n-\mathfrak{D}}/R^{\mathfrak{D}})^{1/2\ell^2} + 1 \right) + p^{-(n-\theta_{\mathfrak{D}})/2\ell^2} \right), \end{aligned}$$

recalling that  $2\ell = n(n-1)$ . Taking  $R = (N/\sqrt{p})^{1/2} \asymp p^{n/2-3/4}$ , we deduce that

$$\begin{aligned} F_\epsilon(d) &\ll_{\epsilon, d, t} |\{t \pmod{p^{n-1}} : \exists d : 4a^2(d + \tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \forall \tau \in T \text{ and } \mathfrak{D}(\epsilon, d, t) \in \{0, 1\}\}| \\ &\quad + p^\epsilon \sum_{\substack{|t| \leq N \\ \mathfrak{D} = \mathfrak{D}(\epsilon, d, t) \geq 2}} \left( (N/\sqrt{p})^{-\mathfrak{D}/4\ell^2} + (p^{n-\mathfrak{D}}(N/\sqrt{p})^{-\mathfrak{D}})^{1/2\ell^2} + p^{-(n-\theta_{\mathfrak{D}})/2\ell^2} \right) + (N/\sqrt{p})^2 \\ &\ll_\epsilon |\{t \pmod{p^{n-1}} : \exists d : 4a^2(d + \tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \forall \tau \in T \text{ and } \mathfrak{D}(\epsilon, d, t) \in \{0, 1\}\}| \\ &\quad + p^{n-1 - \frac{1}{n^2(n-1)^2} + \epsilon}. \end{aligned}$$

In this way, we obtain

$$\begin{aligned}
 & \sum_{d \pmod{p}}^{\diamond} F_{\epsilon}(d) \\
 & \ll_{\epsilon, n} \sum_{t \pmod{p^{n-1}}} |\{d \pmod{p} : 4a^2(d + \tau) \in (\mathbb{Z}/p\mathbb{Z})^{\times 3} \forall \tau \in T \text{ and } \mathcal{D}(\epsilon, d, t) \in \{0, 1\}\}| \\
 & + p^{n - \frac{1}{n^2(n-1)^2} + \epsilon}.
 \end{aligned}$$

It remains to treat the contribution from pairs  $(t, d)$  for which  $\mathcal{D}(\epsilon, d, t) \in \{0, 1\}$ . By construction, if  $\mathcal{D}(\epsilon, d, t) = 0$ , then  $\theta_1(\epsilon, d, t) \geq n > 1$ . Similarly, if  $\mathcal{D}(\epsilon, d, t) = 1$  and  $n \geq 3$ , then  $\theta_2(d, t) \geq n > 2$ . In either of these cases, provided  $p$  is large enough in terms of  $n$ , we get

$$\sum_{\tau \in T} \epsilon_{\tau} s(\overline{2a}(d + \tau))^{2-3(\mathcal{D}+1)} \equiv 0 \pmod{p}.$$

Applying Lemma 4.12, we find that for each  $t \pmod{p^{n-1}}$  the number of  $d$  satisfying  $(\diamond)$  such that this latter congruence holds is  $\ll_{N, M} 1$ . Since  $n \ll_{N, M} 1$ , we deduce that

$$\sum_{d \pmod{p}}^{\diamond} F_{\epsilon}(d) \ll_{\epsilon, N, M} p^{n - \frac{1}{n^2(n-1)^2} + \epsilon} + p^{n-1} \ll p^{n - \frac{1}{n^2(n-1)^2} + \epsilon}.$$

It remains to consider those  $d$  for which  $\mathcal{D}(\epsilon, d, t) = 1$  and  $n = 2$ . In this case, we know that  $p^2 \nmid A_1(\epsilon, d, t)$  and, moreover, that  $A_1(\epsilon, d, t) = a_1(\epsilon, d)$ ; that is,  $A_1(\epsilon, d, t)$  is independent of  $t$ . The sum we wish to bound is therefore

$$\sum_{\substack{d \pmod{p} \\ p^2 \nmid a_1(\epsilon, d)}}^{\diamond} e_{p^2}(a_0(\epsilon, d)) \sum_{t \pmod{p}} e_p(t(a_1(\epsilon, d)/p)) = 0,$$

so the bound required is satisfied in this case as well. □

### 5. Applying the $K_2$ Correlations Bounds: Proof of Proposition 2.4

In this section we will prove Proposition 2.4. Suppose  $Q \geq 2$  factors as  $Q = Q_0 \cdots Q_L$ , where the  $Q_i$  are mutually coprime. Let  $p^\nu \parallel Q_0$  with  $\nu \geq 1$  and  $p > 3$  prime. Let  $C \in \mathbb{Z}/p^\nu\mathbb{Z}$  and  $\mathbf{h} \in \mathbb{Z}^L$  and define

$$\begin{aligned}
 M_{p^\nu}(C, \mathbf{h}) := & \max_{A, B \in (\mathbb{Z}/p^\nu\mathbb{Z})^\times} \left| \sum_{b \pmod{p^\nu}} e_{p^\nu}(CBb) \prod_{\substack{I \subseteq \{1, \dots, L\} \\ |I|=0 \pmod{2}}} K_2(A, b + H_I; p^\nu) \right. \\
 & \left. \cdot \prod_{\substack{J \subseteq \{1, \dots, L\} \\ |J|=1 \pmod{2}}} \overline{K}_2(A, b + H_J; p^\nu) \right|,
 \end{aligned}$$

where for each  $I \subseteq \{1, \dots, L\}$  we set  $H_I := \sum_{i \in I} Q_i h_i$ . Writing  $N$  to denote the number of subsets of  $\{1, \dots, L\}$  of even cardinality and  $M$  the number of subsets with odd cardinality, the total number of  $K_2$  factors in the sum is clearly  $N + M = 2^L$ .

Define now  $T_{\mathbf{h}} := \{H_I : I \subseteq \{1, \dots, L\}\}$ , as well as

$$\begin{aligned}
 \mu_{\mathbf{h}}(\tau) & := |\{I \subseteq \{1, \dots, L\} : |I| \equiv 0 \pmod{2}, H_I \equiv \tau \pmod{p^\nu}\}| \\
 \nu_{\mathbf{h}}(\tau) & := |\{I \subseteq \{1, \dots, L\} : |I| \equiv 1 \pmod{2}, H_I \equiv \tau \pmod{p^\nu}\}|.
 \end{aligned}$$

Further, define

$$\mathcal{T}_{p^\nu} := \{\mathbf{h} \in \mathbb{Z}^L : \mu_{\mathbf{h}}(\tau) \equiv \nu_{\mathbf{h}}(\tau) \pmod{3} \text{ for all } \tau \in T_{\mathbf{h}}\}.$$

The following lemma, which is analogous to [9, Lemma 4.5], allows us to control how frequently either  $\mathbf{h} \in \mathcal{T}_{p^\nu}$  or  $\mu_{\mathbf{h}}(\tau) \equiv \nu_{\mathbf{h}}(\tau) \pmod{3}$ .

**Lemma 5.1.** *Let  $p > 3$ ,  $p \mid Q_0$  and let  $\mathbf{h} \in \mathbb{Z}^L$ . If  $\mu_{\mathbf{h}}(\tau) \equiv \nu_{\mathbf{h}}(\tau) \pmod{3}$  for all  $\tau \in T_{\mathbf{h}}$ , then there is  $1 \leq i \leq L$  such that  $p \mid h_i$ . In particular, if  $\mathbf{h} \in \mathcal{T}_{p^\nu}$  then  $p \mid \prod_{1 \leq i \leq L} h_i$ .*

*Proof.* If  $\mathbf{h} \in \mathcal{T}_{p^\nu}$ , then by definition  $\mu_{\mathbf{h}}(\tau) \equiv \nu_{\mathbf{h}}(\tau) \pmod{3}$  for all  $\tau \in T$  and the second assertion immediately follows from the first. Thus, it suffices to prove that  $p \mid \prod_{1 \leq i \leq L} h_i$  whenever  $3 \mid (\mu_{\mathbf{h}}(\tau) - \nu_{\mathbf{h}}(\tau))$  for all  $\tau \in T$ .

We denote

$$E(\mathbf{h}) := \prod_{1 \leq i \leq L} (1 - e_p(Q_i h_i)).$$

Let  $\Phi_p(z) := \sum_{0 \leq j \leq p-1} z^j = \prod_{a \in (\mathbb{Z}/p\mathbb{Z})^\times} (z - e_p(a))$  the cyclotomic polynomial of order  $p$ . Assume for the sake of contradiction that  $p \nmid h_i \forall i$ . Then

$$\prod_{a \in (\mathbb{Z}/p\mathbb{Z})^\times} E(a\mathbf{h}) = \prod_{1 \leq i \leq L} \prod_{a \in (\mathbb{Z}/p\mathbb{Z})^\times} (1 - e_p(a h_i Q_i)) = \prod_{b \in (\mathbb{Z}/p\mathbb{Z})^\times} (1 - e_p(b))^L = \Phi_p(1)^L = p^L,$$

using the fact  $(Q_i, Q_0) = 1$  and thus that  $p \nmid a h_i Q_i$  whenever  $p \nmid a$ , for all  $1 \leq i \leq L$ .

On the other hand, we have

$$E(a\mathbf{h}) = \prod_{1 \leq i \leq L} (1 - e_p(a h_i Q_i)) = \sum_{I \subseteq \{1, \dots, L\}} (-1)^{|I|} e_p(a H_I) = \sum_{\tau \in T} (\mu_{\mathbf{h}}(\tau) - \nu_{\mathbf{h}}(\tau)) e_p(a \tau),$$

and since  $3 \mid (\mu_{\mathbf{h}}(\tau) - \nu_{\mathbf{h}}(\tau))$  for each  $\tau \in T$ , we can find  $m_\tau \in \mathbb{Z}$  such that

$$E(a\mathbf{h}) = 3 \sum_{\tau \in T} m_\tau e_p(a \tau).$$

It follows that

$$\begin{aligned} p^L &= \prod_{a \in (\mathbb{Z}/p\mathbb{Z})^\times} E(a\mathbf{h}) = 3^{p-1} \prod_{\sigma \in \text{Gal}(\mathbb{Q}(e_p(1))/\mathbb{Q})} \sigma \left( \sum_{\tau \in T} m_\tau e_p(\tau) \right) \\ &= 3^{p-1} N_{\mathbb{Q}(e_p(1))/\mathbb{Q}} \left( \sum_{\tau \in T} m_\tau e_p(\tau) \right). \end{aligned}$$

Note that the bracketed sum over  $\tau \in T$  on the right-hand side is an algebraic integer in  $\mathbb{Q}(e_p(1))$ , so its norm is a rational integer. This implies that  $3 \mid p^L$ , which is false since  $p > 3$ . This contradiction implies that there must be an  $i$  for which  $p \mid h_i$ , as claimed.  $\square$

Using Lemma 5.1, we can finally give the following bounds on the correlation sums  $M_{p^\nu}(C, \mathbf{h})$ .

**Proposition 5.2.** *Let  $L \geq 2$ . Fix parameters  $Y = Y(L) = 2^{3L+5}$  and  $Z = Z(L) = 2^{3L} 2^{2L}$ . Let  $p^\nu \parallel Q_0$  and let  $C \in \mathbb{Z}/p^\nu \mathbb{Z}$ . Assume additionally that*

$$\min_{I, J \subseteq \{1, \dots, L\}} \{|H_I - H_J|_p : H_I \neq H_J\} \geq \begin{cases} p^{-r(\nu)} & \text{if } p > Y \text{ and } \nu > Z \\ p^{-\lfloor r(\nu)/2 \rfloor} & \text{if } 3 < p \leq Y, \nu > Z, \end{cases} \tag{27}$$

where  $r(\nu) := 2^{1-2L} \lfloor \nu 2^{-2L} \rfloor - 1$ . Then we have

$$\frac{M_{p^\nu}(C, \mathbf{h})}{p^{\nu(2^{L-1}+1)}} \ll_{\varepsilon, L} (p^{-1/2} + 1_{C=0}) 1_{p^{\nu-1}|C} 1_{p|\prod_i h_i} + \begin{cases} p^{-1/2} & \text{if } \nu = 1 \\ p^{-\nu/2^{2L}} & \text{if } p > Y \text{ and } \nu > Z \\ p^{-\nu/2^{2L}} & \text{if } 3 < p \leq Y \text{ and } \nu > Z \\ p^{-1/\nu^2(\nu-1)^2+\varepsilon} & \text{if } p > Y \text{ and } 2 \leq \nu \leq Z. \end{cases}$$

*Proof.* Write  $Q_0 = q_1 q_2 q_3 q_4 q_5$ , where the  $q_j$  satisfy  $(q_i, q_j) = 1$  for  $i \neq j$ , according to the following rules:

- (i)  $p|Q_0 \Rightarrow p|q_1$ .
- (ii)  $p^\nu|Q_0, p > Y$  and  $\nu > Z \Rightarrow p^\nu|q_2$ .
- (iii)  $p^\nu|Q_0, p \leq Y$  and  $\nu > Z \Rightarrow p^\nu|q_3$ .
- (iv)  $p^\nu|Q_0, p > Y$  and  $\nu \leq Z \Rightarrow p^\nu|q_4$ .
- (v)  $q_5 = Q_0/(q_1 q_2 q_3 q_4)$ .

We summarise (in an effective form) the bounds that the previous sections imply for each prime power divisor of  $Q_0$ , according to which of the  $q_i$  they divide.

(I) If  $p|q_1$ , Theorem 3.1 yields

$$M_p(C, \mathbf{h}) \ll 3^{2L} p^{2^{L-1}} \left( 2^L \sqrt{p} + p 1_{p|C} 1_{\mathcal{T}_p}(\mathbf{h}) \right). \tag{28}$$

By Lemma 5.1, if  $\mathbf{h} \in \mathcal{T}_p$ , then  $p|\prod_i h_i$ , so we may conclude that

$$M_p(C, \mathbf{h}) \ll_L p^{2^{L-1}+1} \left( p^{-1/2} + 1_{p|C} 1_{p|\prod_i h_i} \right), \tag{29}$$

which implies the bound in the first case.

(II) Suppose  $p^\nu|q_2$  and  $p^{\nu-1} \nmid C$  then as  $|T| \leq 2^L$ . With a view towards applying Theorem 4.1, recall the definition

$$\rho = 1_{p \leq 3 \cdot 2^{L-1}-1} + \left\lceil \frac{\log(20 \cdot 2^{3L})}{\log p} \right\rceil.$$

Since  $p > 2^{3L+5}$ , we have  $\rho = 1$  and so

$$2|T|^{-2} \left( \lfloor \nu 2^{-|T|} \rfloor - \rho \right) = 2^{1-L} \left( \lfloor 2^{-2L} \nu \rfloor - 1 \right) = r(\nu).$$

In light of (27), Theorem 4.1 yields

$$M_{p^\nu}(C, \mathbf{h}) \ll_L p^{\nu(2^{L-1}+1)-\nu/2^{2L}}.$$

If, in addition  $p^{\nu-1}|C$ , then Lemma 5.1 implies that

$$M_{p^\nu}(C, \mathbf{h}) \ll_L \max_a \pmod{p^\nu} p^{\nu(2^{L-1}+1)} \left( p^{-\nu/2^{2L}} + 1_{p|\prod_i h_i} \left( p^{-1/2} + 1_{C=0} \right) \right), \tag{30}$$

and the bound in the second case is proved.

(III) Suppose  $p^\nu \parallel q_3$ ; then  $3 < p \leq 2^{3L+5}$ . We thus have the uniform bound

$$\rho \leq 1 + \left\lceil \frac{\log(2^{3L+5})}{\log 5} \right\rceil \leq 1 + 2(L + 2).$$

As in case (II), we have

$$M_{p^\nu}(C, \mathbf{h}) \ll_L p^{\nu(2^{L-1}+1)} \left( p^{-\nu/2^{2L}} + 1_{p \mid \prod_i h_i} \left( p^{-1/2} + 1_{C=0} \right) \right) \tag{31}$$

where we used the fact that since  $Z$  is chosen suitably,

$$2^{1-2L} \left( \left\lfloor \nu 2^{-2L} \right\rfloor - 2L - 5 \right) \geq \nu 2^{-2L-2L} \geq r(\nu)/2$$

whenever  $L \geq 2$ .

(IV) If  $p^\nu \parallel q_4$ , then, again by Theorem 4.1,

$$M_{p^\nu}(C, h) \ll_{\varepsilon, L} p^{\nu(2^{L-1}+1)} \left( p^{-\frac{1}{\nu^2(\nu-1)^2} + \varepsilon} + 1_{p^{n-1} \mid C} \left( p^{-1/2} + 1_{C=0} \right) 1_{p \mid \prod_i h_i} \right), \tag{32}$$

as required.

(V) For all of the primes powers  $p^\nu \parallel q_5$  the bound given is trivial. □

The following simple lemma allows us to bound the number of tuples  $\mathbf{h}$  where (27) fails.

**Lemma 5.3.** *Let  $c \in (0, 2^{-2L}]$  and let  $d \mid Q_0$ . Assume that  $K/Q_j \geq Q_0^{2c}$  for all  $1 \leq j \leq L$ . Then the number of tuples  $\mathbf{h} \in \mathbb{Z}^L$  with  $1 \leq |h_j| \leq K/Q_j$  for all  $j$ , such that for each  $p^\nu \parallel d$*

$$\min_{\substack{I, J \subseteq \{1, \dots, L\} \\ H_I \neq H_J}} |H_I - H_J|_p < p^{-c\nu}$$

is  $\ll 2^L \tau(d)^{2L-1} d^{-c} K^L / (Q_1 \cdots Q_L)$ .

*Proof.* Write  $d = \prod_{1 \leq i \leq m} p_i^{\nu_i}$ , so that  $m = \omega(d)$ . For each  $1 \leq i \leq m$ , choose a pair of disjoint subsets  $I_i, J_i \subseteq \{1, \dots, L\}$ , such that  $I_i \cup J_i \neq \emptyset$ . We may define a matrix  $A_{I, J} = \{a_{i, j}\}_{i, j}$  with integer entries via

$$a_{i, j} := \begin{cases} Q_j & \text{if } j \in I_i, \\ -Q_j & \text{if } j \in J_i, \\ 0 & \text{otherwise.} \end{cases}$$

By composing  $A_{I, J}$  with projections, we may view it as a homomorphism  $A_{I, J} : \mathbb{Z}^L \rightarrow \prod_{1 \leq i \leq m} (\mathbb{Z}/p_i^{\lceil c\nu_i \rceil} \mathbb{Z})$ , such that for each  $1 \leq i \leq m$  the  $i$ th entry of  $A_{I, J} \mathbf{h}$  is

$$(A_{I, J} \mathbf{h})_i := \sum_{1 \leq j \leq L} a_{i, j} h_j \pmod{p_i^{\lceil c\nu_i \rceil}} = \sum_{l \in I_i} Q_l h_l - \sum_{l \in J_i} Q_l h_l \pmod{p_i^{\lceil c\nu_i \rceil}},$$

whenever  $\mathbf{h} \in \mathbb{Z}^L$ . Note that  $\ker(A_{I, J})$  is a lattice in  $\mathbb{Z}^L$  with covolume  $\tilde{d} := \prod_{1 \leq i \leq m} p_i^{\lceil c\nu_i \rceil} \geq d^c$  and trivially  $A_{I, J}$  is injective on the quotient  $\mathbb{Z}^L / \ker(A_{I, J})$ , with the unique reduced zero class being  $\mathbf{0}$ . By lattice periodicity (and  $\tilde{d} \leq Q_0^{2c} \leq K/Q_j$  for each  $j$ ), we deduce that

$$\begin{aligned} |\{\mathbf{h} \in \ker(A_{I, J}) : 1 \leq |h_j| \leq K/Q_j \forall 1 \leq j \leq L\}| &\ll \tilde{d}^{-1} (2K)^L / (Q_1 \cdots Q_L) \\ &\ll d^{-c} (2K)^L / (Q_1 \cdots Q_L). \end{aligned} \tag{33}$$

Given these remarks, we may prove the lemma as follows. Let  $\mathcal{H}_d$  be the set of tuples  $\mathbf{h}$  described in the statement of the lemma. Given  $\mathbf{h} \in \mathcal{H}_d$ , for each prime  $p^\nu \mid d$  there are distinct subsets  $I_p = I_p(\mathbf{h})$ ,  $J_p = J_p(\mathbf{h})$  of  $\{1, \dots, L\}$  such that  $|H_{I_p} - H_{J_p}|_p \leq p^{-\lceil c\nu \rceil}$ . Now, put  $I'_p := I_p \setminus (I_p \cap J_p)$  and  $J'_p := J_p \setminus (I_p \cap J_p)$ , so that  $I'_p \cup J'_p \neq \emptyset$ ,  $I'_p$  and  $J'_p$  are disjoint and, moreover,  $H_{I_p} - H_{J_p} = H_{I'_p} - H_{J'_p}$ . We thus have, for each  $p^\nu \mid d$ ,

$$\sum_{l \in I'_p} Q_l h_l - \sum_{l \in J'_p} Q_l h_l \equiv 0 \pmod{p^{\lceil c\nu \rceil}}.$$

By the previous definitions, it follows that  $\mathbf{h} \in \ker(A_{\{I'_p\}_p, \{J'_p\}_p})$ . Hence, we obtain the upper bound

$$|\mathcal{H}_d| \leq \sum_{\substack{\mathbf{I}, \mathbf{J} \subseteq \{1, \dots, L\}^m \\ I_i \cap J_i = \emptyset \forall i \\ I_i \cup J_i \neq \emptyset \forall i}} |\{\mathbf{h} \in \ker(A_{\mathbf{I}, \mathbf{J}}) : 1 \leq |h_j| \leq K/Q_j \forall 1 \leq j \leq L\}|.$$

The number of pairs of tuples of sets  $\mathbf{I}, \mathbf{J}$  is  $\leq \binom{2L}{2}^m \leq 2^{(2L-1)\omega(d)} \leq \tau(d)^{2L-1}$ , so by (33) we obtain

$$|\mathcal{H}_d| \ll 2^L \tau(d)^{2L-1} d^{-c} K^L / (Q_1 \cdots Q_L),$$

as claimed. □

*Proof of Proposition 2.4.* i) Assume  $Q_0, \dots, Q_L$  are all coprime and squarefree. Case (I) of Proposition 5.2 gives

$$M_p(C, \mathbf{h}) \ll_L p^{2^{L-1}+1/2} \min\{(p, C)^{1/2}, (p, \prod_i h_i)^{1/2}\},$$

for each  $p \mid Q_0$ . Combining this with Lemma 2.5, we obtain that

$$\begin{aligned} T(\mathbf{h}) &= \sum_{k \in J(\mathbf{h})} \prod_{I \subseteq \{1, \dots, L\}} c^{|I|} K_2 \left( b, k + \sum_{i \in I} Q_i h_i; Q_0 \right) \\ &\ll \sum_{C \in \mathbb{Z}/(Q_0\mathbb{Z})} \min \left\{ \frac{|J(\mathbf{h})|}{Q_0}, \frac{1}{Q_0 \|C/Q_0\|} \right\} \prod_{p \mid Q_0} M_p(C, \mathbf{h}) \\ &\ll_{\varepsilon, L} Q_0^{2^{L-1}+1/2+\varepsilon} \sum_{C \in \mathbb{Z}/(Q_0\mathbb{Z})} \min \left\{ \frac{|J(\mathbf{h})|}{Q_0}, \frac{1}{Q_0 \|C/Q_0\|} \right\} \prod_{p \mid Q_0} \min\{(p, C)^{1/2}, (p, \prod_i h_i)^{1/2}\} \\ &\ll_{\varepsilon} Q_0^{2^{L-1}+1/2+\varepsilon} \sum_{\substack{d \mid Q_0 \\ d < Q_0}} \frac{(d, \prod_i h_i)^{1/2}}{d} \sum_{C' \pmod{Q_0/d}}^* \frac{1}{C'} + \frac{K}{Q_0} (Q_0, \prod_i h_i)^{1/2} \\ &\ll_{\varepsilon} (1 + K/Q_0) Q_0^{2^{L-1}+1/2+\varepsilon} (Q_0, \prod_i h_i)^{1/2}, \end{aligned}$$

using  $|J(\mathbf{h})| \leq K$  and the obvious inequality  $(d, \prod_i h_i) \leq (Q_0, \prod_i h_i)$  for any  $d \mid Q_0$ . Employing  $(Q_0, \prod_i h_i)^{1/2} \leq \prod_i (Q_0, h_i)^{1/2}$  and summing in  $h_1, \dots, h_L$  yields

$$\begin{aligned}
 & \sum_{1 \leq |h_1| \leq K/Q_1} \cdots \sum_{1 \leq |h_L| \leq K/Q_L} |T(\mathbf{h})| \\
 & \ll_{\varepsilon, L} (1 + K/Q_0) Q_0^{2^{L-1} + 1/2 + \varepsilon} \prod_{1 \leq j \leq L} \sum_{1 \leq |h_j| \leq K/Q_j} (Q_0, h_j)^{1/2} \\
 & \ll (1 + K/Q_0) Q_0^{2^{L-1} + 1/2 + \varepsilon} \sum_{e_1, \dots, e_L | Q_0} (e_1 \cdots e_L)^{1/2} \prod_{1 \leq j \leq L} \sum_{\substack{1 \leq |h'_j| \leq K/(e_j Q_j) \\ (h'_j, Q_0) = 1}} 1 \\
 & \ll_L (1 + K/Q_0) Q_0^{2^{L-1} + 1/2 + \varepsilon} \frac{K^L}{Q_1 \cdots Q_L} \left( \sum_{e | Q_0} e^{-1/2} \right)^L \\
 & \ll_{\varepsilon, L} (1 + K/Q_0) Q_0^{2^{L-1} + 1/2 + \varepsilon} \frac{K^L}{Q_1 \cdots Q_L} = (1 + K/Q_0) Q_0^{2^{L-1} + 3/2 + \varepsilon} K^L / Q,
 \end{aligned}$$

using  $Q/Q_0 = Q_1 \cdots Q_L$  in the last line. This implies i).

ii) As above, we would like to estimate

$$\mathcal{T} := \sum_{1 \leq |h_1| \leq K/Q_1} \cdots \sum_{1 \leq |h_L| \leq K/Q_L} \sum_C \sum_{(\text{mod } Q_0)} \min \left\{ \frac{K}{Q_0}, \frac{1}{Q_0 \|C/Q_0\|} \right\} \prod_{p^v || Q_0} M_{p^v}(C, \mathbf{h}).$$

Here, we are assuming that  $(Q_0, 6) = 1$ . We factor  $Q_0 = q_1 q_2 q_3 q_4 q_5$ , where

$$q_1 = \prod_{p | Q_0} p, \quad q_2 = \prod_{\substack{p^v || Q_0 \\ p > Y, v > Z}} p^v, \quad q_3 = \prod_{\substack{p^v || Q_0 \\ 3 < p \leq Y, v > Z}} p^v, \quad q_4 = \prod_{\substack{p^v || Q_0 \\ p > Y, 2 \leq v \leq Z}} p^v \text{ and } q_5 = \prod_{\substack{p^v || Q_0 \\ p \leq Y, 2 \leq v \leq Z}} p^v.$$

By construction,  $q_5 \ll_L 1$ , so that

$$\prod_{p^v || q_5} M_{p^v}(C, \mathbf{h}) \ll q_5^{(N+M+2)/2} \ll_L 1.$$

We focus next on the prime power divisors of  $Q'_0 := Q_0/q_5$ . As above, define  $r(v) := \left\lfloor 2^{1-2L} \left\lfloor v 2^{-2L-2L} \right\rfloor \right\rfloor$ . For each  $1 \leq i \leq 4$  and each  $p^v || q_i$ , we write

$$M_{p^v}(C, \mathbf{h}) = p^{(2^{L-1} + 1)v} (M_{p^v, 1}(C, \mathbf{h}) + M_{p^v, 2} + M_{p^v, 3}(\mathbf{h})),$$

where we have defined

$$M_{p^v, 1}(C, \mathbf{h}) = \begin{cases} 1_{p^{v-1} | C} 1_{p | \prod_j h_j} (p^{-1/2} + 1_{p^v | C}) & \text{if } i = 2, 3, 4 \\ 1_{p | C} 1_{p | \prod_j h_j} & \text{if } i = 1, \end{cases}$$

as well as

$$M_{p^v, 2} = \begin{cases} p^{-1/2} & \text{if } i = 1 \\ p^{-v/2^{2L}} & \text{if } i = 2, 3, \\ p^{-v/Z^5 + \varepsilon} & \text{if } i = 4 \end{cases}, \quad M_{p^v, 3}(\mathbf{h}) = \begin{cases} 0 & \text{if } i = 1, 4 \\ 1 & \text{if } i = 2, 3 \text{ and (27) fails.} \end{cases}$$

Note that  $M_{p^v,1}$  depends only on  $(C, Q'_0) = \prod_{1 \leq i \leq 4} (C, q_i)$ . We factor  $C = C'd$  and  $d = d_1 d_2 d_3 d_4 d_5$ , where  $d_j = (C, q_j)$ , for each  $1 \leq j \leq 5$  and get

$$\begin{aligned}
 \mathcal{T} &\ll_{\varepsilon,L} \sum_{\substack{d|Q_0 \\ d=d_1 d_2 d_3 d_4 d_5}} \left( \sum_{C' \pmod{Q_0/d}}^* \min \left\{ \frac{K}{Q_0}, \frac{1}{Q_0 \|C'd/Q_0\|} \right\} \right) \\
 &\cdot \sum_{1 \leq |h_1| \leq K/Q_1} \cdots \sum_{1 \leq |h_L| \leq K/Q_L} \prod_{1 \leq i \leq 4} \prod_{p^v | q_i} p^{(2^{L-1}+1)v} (M_{p^v,1}(d_i, \mathbf{h}) + M_{p^v,2} + M_{p^v,3}(\mathbf{h})) \\
 &\ll_{\varepsilon,L} Q_0^{2^{L-1}+1} \sum_{e_1 f_1 = q_1}^* f_1^{-1/2} \sum_{e_4 f_4 = q_4}^* f_4^{-Z^{-5}+\varepsilon} \sum_{e_2 f_2 g_2 = q_2}^* f_2^{-2^{-2L}} \sum_{e_3 f_3 g_3 = q_3}^* f_3^{-2^{-2L}} \\
 &\cdot \sum_{\substack{d|Q_0 \\ d=d_1 d_2 d_3 d_4 d_5 \\ p^v || e_j \Rightarrow p^{v-1} | d_j \\ 1 \leq j \leq 4}} \left( \sum_{C' \pmod{Q_0/d}}^* \min \left\{ K/Q_0, \frac{1}{Q_0 \|C'd/Q_0\|} \right\} \right) \left( \prod_{1 \leq j \leq 4} \prod_{p|e_j} (p^{-1/2} 1_{j \neq 1} + 1_{v_p(e_j)=v_p(d_j)}) \right) \\
 &\cdot \sum_{\substack{1 \leq |h_1| \leq K/Q_1 \\ p | g_2 g_3 \Rightarrow \exists j: M_{p^v,3} \neq 0 \\ \text{rad}(e_1 e_2 e_3 e_4) | \prod_j h_j}} \cdots \sum_{1 \leq |h_L| \leq K/Q_L} 1, \tag{34}
 \end{aligned}$$

where the summation symbol  $\sum_{uvw=x}^*$  indicates that  $(u, v) = (v, w) = (w, u) = 1$  (and, analogously,  $\sum_{uv=x}^*$  indicates that  $(u, v) = 1$ ), so that if  $p|u$ , say, then  $v_p(u) = v_p(x)$ .

By dropping the constraint on  $e_1, e_2, e_3, e_4$ , the innermost sum over the tuples  $\mathbf{h}$  can be bounded above by

$$\{ \mathbf{h} \in \mathbb{Z}^L : 1 \leq |h_j| \leq K/Q_j \forall j, p | g_2 g_3 \Rightarrow \min_{\substack{I, J \subseteq \{1, \dots, L\} \\ H_I \neq H_J}} |H_I - H_J| < p^{-r(v)/2} \}.$$

If we define

$$\delta' := \min \left\{ 2^{-2L}, Z^{-5}, \min_{v > Z} \frac{\lfloor r(v)/2 \rfloor}{v} \right\},$$

then by applying Lemma 5.3 with  $c = \delta'$  and  $d = g_2 g_3$  (which is a divisor of  $Q_0$ ) and using the crude bound  $\tau(d)^{2L-1} \ll_{\varepsilon,L} d^\varepsilon$  we may bound this cardinality by

$$\ll_{\varepsilon,L} X^\varepsilon \frac{K^L}{(g_2 g_3)^{\delta'} Q_1 \cdots Q_L} = X^\varepsilon \frac{K^L Q_0}{(g_2 g_3)^{\delta'} Q}.$$

Inserting this into our earlier upper bound for  $\mathcal{T}$ , we obtain

$$\begin{aligned}
 \mathcal{T} &\ll_{\varepsilon,L} \frac{X^\varepsilon K^L Q_0^{2^{L-1}+2}}{Q} \sum_{\substack{e_1 f_1 = q_1 \\ e_4 f_4 = q_4}}^* \frac{1}{f_1^{1/2} f_4^{-Z^{-5}}} \sum_{\substack{e_2 f_2 g_2 = q_2 \\ e_3 f_3 g_3 = q_3}}^* \frac{1}{(f_2 f_3)^{2^{-2L}} (g_2 g_3)^{\delta'}} \\
 &\cdot \sum_{\substack{d|Q_0 \\ d=e_1 d_2 d_3 d_4 d_5 \\ v_p(d_j) \geq v_p(e_j)-1 \\ 2 \leq j \leq 4}} \left( \sum_{C' \pmod{Q_0/d}}^* \min \left\{ K/Q_0, \frac{1}{Q_0 \|C'd/Q_0\|} \right\} \right) \left( \prod_{2 \leq j \leq 4} \prod_{p|e_j} (p^{-1/2} + 1_{v_p(e_j)=v_p(d_j)}) \right).
 \end{aligned}$$

We parametrise  $d = e_1 d_2 d_3 d_4 d_5$  with  $d_j = e_j D_j$ , where  $D_j | \text{rad}(e_j)$ , for each  $2 \leq j \leq 4$ . Using  $\delta' \leq 2^{-2L}$  and  $\delta' \leq Z^{-5}$ , we find

$$\mathcal{T} \ll_{\varepsilon, L} \frac{X^\varepsilon K^L Q_0^{2^{L-1}+2}}{Q} \sum_{\substack{e_1 f_1 = q_1 \\ e_4 f_4 = q_4}}^* \frac{1}{f_1^{1/2} f_4^{\delta'}} \sum_{\substack{e_2 f_2 g_2 = q_2 \\ e_3 f_3 g_3 = q_3}}^* \frac{1}{(f_2 f_3)^{\delta'} (g_2 g_3)^{\delta'}} \cdot \sum_{\substack{D_j | \text{rad}(e_j) \\ 2 \leq j \leq 4}} (D_2 D_3 D_4)^{-1/2} \left( C' \sum_{(\text{mod } Q_0 / (e_1 e_2 D_2 e_3 D_3 e_4 D_4))}^* \min \left\{ K/Q_0, \frac{1}{Q_0 \|C' e_1 e_2 D_2 e_3 D_3 e_4 D_4 / Q_0\|} \right\} \right).$$

If we bound the sum over  $C'$  trivially, we obtain  $\ll \frac{\log Q_0}{e_1 e_2 e_3 e_4 D_2 D_3 D_4}$ , which finally leads to

$$\begin{aligned} \mathcal{T} &\ll_{\varepsilon, L} \frac{X^\varepsilon K^L Q_0^{2^{L-1}+2}}{Q} \sum_{\substack{e_1 f_1 = q_1 \\ e_4 f_4 = q_4}}^* \frac{1}{(f_1 f_4)^{\delta'}} e_1 e_4 \sum_{\substack{e_2 f_2 g_2 = q_2 \\ e_3 f_3 g_3 = q_3}}^* \frac{1}{(f_2 f_3)^{\delta'} (g_2 g_3)^{\delta'}} e_2 e_3 \\ &\ll_{\varepsilon} \frac{X^\varepsilon K^L Q_0^{2^{L-1}+2}}{Q} (q_1 q_2 q_3 q_4)^{-\delta'} \ll_L X^\varepsilon K^L Q_0^{2^{L-1}+2-\delta'} / Q, \end{aligned}$$

again using  $q_5 \ll_L 1$  and  $Q_0 = q_1 q_2 q_3 q_4 q_5$ . This proves claim ii). □

**6. Proof of Theorems 1.1 and 1.5**

This section is devoted to the proofs of Theorems 1.1 and 1.5.

Let  $\varepsilon > 0$  be sufficiently small and let  $0 < \eta < 1/522$ . Let  $X$  be sufficiently large in terms of  $\varepsilon$  and let  $X^{2/3-\varepsilon} < q \leq X^{3/4+1/1044-\varepsilon}$  be  $X^\eta$ -smooth. We wish to show that there is a  $\delta = \delta(\varepsilon, \eta) > 0$  such that

$$\Delta_{\mu^2}(X; q, a) = \sum_{\substack{n \leq X \\ n \equiv a \pmod{q}}} \mu^2(n) - \frac{1}{\phi(q/(a, q))} \sum_{\substack{n \leq X \\ (n, q) = (a, q)}} \mu^2(n) \ll X^{1-\delta} / q,$$

for any residue class  $a \pmod{q}$  with  $(a, q) \leq X^\varepsilon$ , with further size constraints on  $q$  depending on whether or not  $q$  is squarefree.

Replacing  $X, q$  and  $a$  by  $X/(a, q), q/(q, a)$  and  $a/(a, q)$  if necessary, we may reduce our problem to the case in which  $a$  is coprime to  $q$ : indeed,  $q/(q, a)$  is still  $X^\eta$ -smooth and as  $(a, q) \leq X^\varepsilon$  the bound above is implied by the case  $(a, q) = 1$  (upon taking  $\delta$  slightly smaller). In what follows, we will focus solely on when the residue class  $a$  is coprime to  $q$ .

We begin by summarising the analysis in Section 2, which is valid when  $(a, q) = 1$ . Let  $\delta > 0$  be fixed, to be selected later. We fixed a scale  $V_0 \geq X^{\delta+\varepsilon}$  and found a second scale  $\max\{X^{1-\delta-\varepsilon}/(qV_0), V_0\} \leq V_1 \ll X^{1/2}$  and an interval  $I(V_1) = (V_1, V_1 + V_1/V_0]$  such that (see (12))

$$\begin{aligned} \Delta_{\mu^2}(X; q, a) &\ll_{\varepsilon} X^\varepsilon \frac{V_0 V_1}{\tilde{q}} \sum_{\substack{f|q \\ f \leq Z}} \sum_{\substack{f|k \\ (\text{mod } \tilde{q}/f)}}^* \sum_{\substack{m \\ (\text{mod } \tilde{q}/f) \\ m \neq 0}} \frac{\kappa(m; ka, \tilde{q}/f)}{km} + V_0 \left( 1 + X^\varepsilon V_0 \tilde{q}^{1/2} Z^{-3/2} \right) \\ &+ X^\varepsilon \left( \frac{X}{qV_0} + \frac{X}{\tilde{q}V_1} + \left( \frac{X}{Z\tilde{q}} \right)^{1/2} + KV_1^2 \left( \frac{Z}{\tilde{q}} \right)^{3/2} \right), \end{aligned} \tag{35}$$

for any divisor  $\tilde{q}$  of  $q$  and any  $Z \geq 1$ . We recall here that, given  $Q \geq 1$  and  $K \geq 1$ , we have set

$$\kappa(M, N; Q) := \max_{1 \leq R \leq K} \left| \sum_{K(M-1) < r \leq K(M-1)+R} e_Q(-rV_1) K_2(N, r; Q) \right|,$$

for each  $M \geq 1$  and  $N \in \mathbb{Z}$  coprime to  $Q$ .

To proceed, we need to be able to choose  $\tilde{q}$  of suitable size in order to obtain the desired  $O_\varepsilon(X^{1-\delta}/q)$  bound. The following lemma will be useful in this vein, especially in order to apply the results of the last few sections.

**Lemma 6.1.** *Let  $\eta > 0$  and suppose  $X \geq 3$  is large enough relative to  $\eta$ . Let  $q \in \mathbb{N}$  be  $X^\eta$ -smooth.*

- a) *If  $0 < v < 1$ , then there is a divisor  $q'$  of  $q$  such that  $q' \in (q^v, X^\eta q^v]$ .*
- b) *Assume furthermore that  $q \in (X, 2X]$  is  $X^\eta$ -ultrasmooth. Let  $k \geq 4$  and let  $u_1, \dots, u_k \in (\eta, 1 - \eta)$  with  $u_1 + \dots + u_k = 1$ . Then we can find  $q_1, \dots, q_k$ , mutually coprime with  $(q_k, 6) = 1$ , such that  $q_1 \in (X^{u_1}, X^{u_1+\eta} 2^{v_2(q)} 3^{v_3(q)})$ ,  $q_2 \in (X^{u_2}, X^{u_2+\eta}]$ ,  $q_j \in (X^{u_j}, X^{u_j+\eta}]$  for each  $3 \leq j \leq k - 1$  and  $q_k \in (X^{u_k-(k-1)\eta} 2^{-v_2(q)} 3^{-v_3(q)}, 2X^{u_k}]$  such that  $q = q_1 \cdots q_k$ .*

*Proof.* a) Enumerate the prime factors of  $q$  in ascending order as  $p_1 \leq p_2 \leq \dots \leq p_{\omega(q)}$ . Let  $r$  be chosen maximally such that  $p_1 \cdots p_r \leq q^v$ . Then  $q' := p_1 \cdots p_r p_{r+1} > q^v$  by maximality and, moreover,  $q' \leq q^v p_{r+1} \leq q^v X^\eta$ . This establishes the first claim.

b) Arguing similarly as in a), order the prime power factors of  $q$  as  $p_1^{\alpha_1} < \dots < p_{\omega(q)}^{\alpha_{\omega(q)}} \leq X^\eta$ . Pick  $1 \leq N_1 \leq N$  to be the minimal integer such that  $\prod_{1 \leq j \leq N_1} p_j^{\alpha_j} > X^{u_1}$  and set  $q_1 := \prod_{1 \leq j \leq N_1} p_j^{\alpha_j}$ . By minimality,  $q_1/p_{N_1}^{\alpha_{N_1}} \leq X^{u_1}$ , whence  $q_1 \in (X^{u_1}, X^{u_1+\eta}]$ .

We set  $X' := X/q_1 \in [X^{1-u_1-\eta}, X^{1-u_1})$  and  $q' := q/q_1$ . As  $u_2 > 0$ ,  $N_1 < N$ . We then select  $N_1 < N_2 \leq N$  such that  $q_2 := \prod_{N_1 < j \leq N_2} p_j^{\alpha_j} > X^{u_2}$ , so that  $q_2 \in (X^{u_2}, X^{u_2+\eta}]$ .

Repeating this process  $k - 1$  times, we obtain integers  $q_1, \dots, q_{k-1}$  such that  $q_j \in (X^{u_j}, X^{u_j+\eta}]$ . We replace  $q_1$  by  $q_1 2^{v_2(q)} 3^{v_3(q)}$  if the powers of 2 and 3 dividing  $q$  are not already divisors of  $q_1, \dots, q_{k-1}$ . The factors  $q_j$  are mutually coprime by construction. Putting  $q_k := q/(q_1 \cdots q_{k-1})$ , the above construction and  $q \in (X, 2X]$  forces  $(q_k, 6) = 1$  and

$$q_k \in (X^{1-u_1-\dots-u_{k-1}-(k-1)\eta} 2^{-v_2(q)} 3^{-v_3(q)}, 2X^{1-u_1-\dots-u_{k-1}}],$$

which implies the claim since  $u_k = 1 - u_1 - \dots - u_{k-1}$  by definition. □

**6.1. First Result:**  $q \leq X^{3/4-\varepsilon}$

We begin by considering the easier range  $X^{2/3-\varepsilon} < q \leq X^{3/4-\varepsilon}$ , in which we need not assume that  $q$  is squarefree. Let  $\eta/12 \leq \delta < 1/50$ . By Lemma 6.1 a), we can choose a divisor  $\tilde{q}$  of  $q$  with  $\tilde{q} \in (X^{1/2+12\delta}, X^{1/2+12\delta+\eta}]$ . Take  $V_0 = K = X^{\delta+\varepsilon}$  and  $Z = X^{10\delta}$ . Using Lemma 2.2, we can bound

$$\max_{q'|\tilde{q}} \max_{\substack{1 \leq k' \leq q' \\ (k', q')=1}} \max_{1 \leq m \leq q'/K} \kappa(m; k'a, q') \ll K \tilde{q}^{-1/2+\varepsilon} \ll X^{\delta+\varepsilon} X^{1/4+6\delta+\eta/2}.$$

From (35) and the above parameter choices, as well as the lower bound  $V_1 \geq X^{1-\delta-\varepsilon}/(qV_0) \geq X^{1/4-2\delta-3\varepsilon}$  that we may assume (see (4) above), we obtain

$$\begin{aligned} \Delta_{\mu^2}(X; q, a) \ll_\varepsilon & \frac{X^{2\delta+2\varepsilon+1/2}}{\tilde{q}^{1/2}} + X^{2\delta+3\varepsilon-15\delta} \tilde{q}^{1/2} + X^{\delta+\varepsilon} + \frac{X^{1-\delta}}{q} + \frac{X^{1+\varepsilon}}{\tilde{q}V_1} \\ & + X^{\delta+\varepsilon} X^{1/2(1-10\delta)} \tilde{q}^{-1/2} + X^{1+\delta+\varepsilon+15\delta} \tilde{q}^{-3/2} \end{aligned}$$

$$\begin{aligned} &\ll_{\varepsilon} X^{1/4-4\delta+2\varepsilon} + X^{1/4-7\delta+\eta/2+\varepsilon} + X^{1/4-\delta+\varepsilon} + \frac{X^{1-\delta}}{q} + X^{1/4-10\delta+2\varepsilon} \\ &+ X^{1/4-10\delta+4\varepsilon} + X^{1/4-2\delta+\varepsilon} \\ &\ll X^{1-\delta}/q, \end{aligned}$$

using the fact that  $\eta \leq 12\delta$  in the last line. This implies the claim for  $q \leq X^{3/4-\varepsilon}$ .

**Remark 6.2.** The above proof shows that we can obtain power-savings in the range  $q \leq X^{3/4-\varepsilon}$  for  $q$  that are  $X^\eta$ -smooth with any  $\eta < 6/25$ , since in this case the choice  $\delta = \eta/12$  is admissible.

**6.2. Squarefree  $X^{3/4-\varepsilon} < q \leq X^{3/4+1/1044-\varepsilon}$**

In this range of the modulus  $q$  we will invoke Propositions 2.3 and 2.4 i), and to this end we need a suitable divisor  $\tilde{q}$  of  $q$  as well as a suitable factorisation for  $\tilde{q}$ . We begin this subsection by outlining the parameter choices needed to this end.

Let  $L \geq 2$ . Fix  $0 < \delta < 1/10$  to be chosen later and let  $\theta := 3/4 + \lambda$ , where  $\eta/2 < \lambda < \min\{1/20, 1/(2L)\}$ . We will determine constraints on  $\lambda$  momentarily, from which we will conclude that any  $\lambda < 1/1044$  will be admissible.

Set  $\gamma := 2\delta + \lambda + \varepsilon$  for  $\varepsilon > 0$  small and put

$$\sigma := \frac{1}{L} + \frac{2(2^{L+2} + L)}{L}\gamma,$$

assuming that  $\gamma$  and  $L$  are chosen so that this is  $< 1/4 + \lambda$ . Also, put  $K := \lfloor X^{\sigma/2-\lambda} \rfloor$ . Suppose that  $q \in (X^\theta, 2X^\theta]$  is squarefree and  $X^\eta$ -smooth. As  $\sigma < 1/4 + \lambda$ , by Lemma 6.1 a) we may choose  $\tilde{q} \in (X^{1/2+\sigma}, X^{1/2+\sigma+\eta}]$ .

Of course,  $\tilde{q}$  is also  $X^\eta$ -smooth. Applying Lemma 6.1 b), we can find a factorisation  $\tilde{q} = \tilde{q}_0\tilde{q}_1 \cdots \tilde{q}_L$ , such that

$$\begin{aligned} \tilde{q}_{L-j+1} &\in (X^{\sigma/2-(2^j+1)\gamma-\eta}, X^{\sigma/2-(2^j+1)\gamma}] \text{ for all } 1 \leq j \leq L-1, \\ \tilde{q}_1 &\in (X^{\sigma/2-(2^L+1)\gamma-\eta}, 6X^{\sigma/2-(2^L+1)\gamma}], \\ \tilde{q}_0 &\in [X^{\sigma-(2^{L+1}+2)\gamma}/6, 2X^{\sigma-(2^{L+1}+2)\gamma+L\eta}]. \end{aligned}$$

This is indeed possible since  $\tilde{q} \in (X^{1/2+\sigma}, X^{1/2+\sigma+\eta}]$  and

$$\begin{aligned} &\sigma - (2^{L+1} + 2)\gamma + L\eta + \sum_{1 \leq j \leq L} \left( \frac{\sigma}{2} - (2^j + 1)\gamma - \eta \right) = (L/2 + 1)\sigma - (2^{L+2} + L)\gamma \\ &= \frac{1}{2} + \sigma + \frac{L}{2} \left( \sigma - \frac{1}{L} - \frac{2(2^{L+2} + L)}{L}\gamma \right) = \frac{1}{2} + \sigma. \end{aligned}$$

Note that, by construction, we have  $K \geq \max\{\tilde{q}_1, \dots, \tilde{q}_L\}$  and, moreover,

$$\max \left\{ \left( \frac{\tilde{q}_0^{1/2}}{K} \right)^{2^{-L}}, \left( \frac{\tilde{q}_{L-j+1}}{K} \right)^{2^{-j}} \right\} \ll X^{-\gamma} \text{ for all } 1 \leq j \leq L. \tag{36}$$

In addition, since  $\gamma \leq \sigma/2$ ,

$$\tilde{q}_0^{-2^{-(L+1)}} \ll X^{-2^{-(L+1)}\sigma + (1+2^{-L})\gamma} \ll X^{-\gamma}. \tag{37}$$

Next, select  $V_0 = X^{\delta+\varepsilon}$  and  $Z = X^{\frac{2}{3}(\sigma/2+4\delta+2\lambda-2\varepsilon)}$ , which satisfies  $Z \leq \tilde{q}/K$ . We apply all of these parameter choices in (35), as well as the inequalities

- (i)  $\eta/2 < \lambda < 1/20$  and  $\delta < 1/10$ ,
- (ii)  $X^{1-\delta}/q \geq \frac{1}{2}X^{1/4-\delta-\lambda}$ ,
- (iii)  $\sigma > 10\gamma > 10(\delta + \lambda)$  and
- (iv)  $X^{1-\delta-\varepsilon}/(qV_0) \leq V_1 \ll X^{1/2}$ ,

to obtain

$$\begin{aligned} \Delta_{\mu^2}(X; q, a) &\ll_{\varepsilon} X^{\frac{1}{2}+\delta+2\varepsilon} \tilde{q}^{-1} \max_{\substack{f|\tilde{q} \\ f \leq Z}} \sum_{k \pmod{\tilde{q}/f}}^* \sum_{\substack{m \pmod{\tilde{q}/f} \\ m \neq 0}} \frac{|\kappa(m; ka, \tilde{q}/f)|}{km} \\ &\quad + X^{2\delta-(\sigma/2+4\delta+2\lambda-2\varepsilon)+3\varepsilon} \tilde{q}^{1/2} \\ &\quad + \frac{X^{1-\delta}}{q} + \frac{X^{1+\varepsilon}}{\tilde{q}V_1} + X^{1/2-\frac{1}{3}(\sigma/2+4\delta+2\lambda-2\varepsilon)+\varepsilon} \tilde{q}^{-1/2} + X^{1+\sigma/2+(\sigma/2+4\delta+2\lambda-2\varepsilon)-\lambda+\varepsilon} \tilde{q}^{-3/2} \\ &= X^{-\sigma+\delta+2\varepsilon} \max_{\substack{f|\tilde{q} \\ f \leq Z}} \sum_{k \pmod{\tilde{q}/f}}^* \sum_{\substack{m \pmod{\tilde{q}/f} \\ m \neq 0}} \frac{|\kappa(m; ka, \tilde{q}/f)|}{km} + \frac{X^{1-\delta}}{q} \cdot X^{-\delta-\lambda+\eta/2+5\varepsilon} \\ &\quad + \frac{X^{1-\delta}}{q} \left( 1 + X^{-\sigma+3\delta+2\lambda+2\varepsilon} + X^{-\frac{2\sigma}{3}-\delta/3+\lambda/3+5\varepsilon/3} + X^{-\sigma/2+5\delta+\lambda+\eta-\varepsilon} \right) \\ &\ll_{\varepsilon} X^{-\sigma+\delta+2\varepsilon} \max_{\substack{f|\tilde{q} \\ f \leq Z}} \sum_{k \pmod{\tilde{q}/f}}^* \sum_{\substack{m \pmod{\tilde{q}/f} \\ m \neq 0}} \frac{|\kappa(m; ka, \tilde{q}/f)|}{km} + \frac{X^{1-\delta}}{q}, \end{aligned} \tag{38}$$

provided that  $\varepsilon > 0$  is sufficiently small.

Fix  $f_0 \mid \tilde{q}$  with  $f_0 \leq Z$  that maximises the sum over  $f$  in (35) and let  $q' := \tilde{q}/f_0$ . We can factor  $q' = q'_0 \cdots q'_L$  as in Proposition 2.3 by writing  $\tilde{q} = \tilde{q}_0 \cdots \tilde{q}_L$  and setting  $q'_j := \tilde{q}_j / (\tilde{q}_j, f_0)$ , for each  $j$ . Putting  $Q = q'$  and recalling that  $K \geq \max_{1 \leq j \leq L} \tilde{q}_j$ , we may combine Proposition 2.4 i) with Proposition 2.3 to get that for each  $1 \leq m \leq q' - 1$  and each  $1 \leq k \leq q' - 1$  coprime to  $q'$ ,

$$\begin{aligned} &|\kappa(m; ka, q')| \\ &\ll_{\varepsilon, L} (q')^{1/2+\varepsilon} K \left( \sum_{1 \leq j \leq L} \left( \frac{q'_{L-j+1}}{K} \right)^{2^{-j}} + \left( \frac{q'(q')^{\varepsilon} (q'_0)^{2^{L-1}+3/2} (2K)^L}{K^{L+1} (q'_0)^{2^{L-1}+1}} \frac{1}{q'} (K/q'_0 + 1) \right)^{2^{-L}} \right) \\ &\ll_L (\tilde{q}/f_0)^{1/2+\varepsilon} K \left( \sum_{1 \leq j \leq L} \left( \frac{\tilde{q}_{L-j+1}}{(f_0, \tilde{q}_{L-j+1})K} \right)^{2^{-j}} + \left( \frac{\sqrt{\tilde{q}_0}/(\tilde{q}_0, f_0)}{K} \right)^{2^{-L}} + (\tilde{q}_0/(f_0, \tilde{q}_0))^{-2^{-L-1}} \right) \\ &\ll K \tilde{q}^{1/2} X^{-\gamma} \ll X^{\frac{1}{4}+\sigma+\eta/2-\gamma-\lambda}, \end{aligned}$$

using  $f_0^{-1/2+\varepsilon} (\tilde{q}_0, f_0)^{2^{-L-1}} \ll_{\varepsilon} 1$  whenever  $L \geq 2$ . Inserting this bound into (38) and summing  $m$  and  $k$ , when  $\varepsilon$  is sufficiently small we obtain

$$\begin{aligned} \Delta_{\mu^2}(X; q, a) &\ll_{\varepsilon} X^{-\sigma+\delta+3\varepsilon} \cdot X^{\frac{1}{4}+\sigma+\eta/2-\gamma-\lambda} + \frac{X^{1-\delta}}{q} \\ &= \frac{X^{1-\delta}}{q} \left( 1 + X^{-\lambda+\eta/2+3\varepsilon} \right) \ll \frac{X^{1-\delta}}{q}. \end{aligned}$$

It remains to show that any  $\lambda < 1/1044$  is admissible. With the parameter choices made earlier, we have assumed that

$$\sigma = \frac{1}{L} + \frac{2(2^{L+2} + L)\gamma}{L} < \frac{1}{4} + \lambda < \frac{1}{4} + \gamma,$$

which forces  $L \geq 5$  and

$$\gamma \leq \frac{L/4 - 1}{2^{L+3} + L}.$$

Since the bounds are decreasing in  $L$ , we deduce by setting  $L = 5$  that

$$\lambda < \gamma \leq \frac{1}{1044}.$$

Furthermore, as  $\gamma = 2\delta + \lambda$ , we may always choose  $\delta$  small enough so that any  $\lambda > 1/1044 - \varepsilon'$  is obtainable, for any  $\varepsilon' > 0$ . We thus deduce that for  $\eta > 0$  sufficiently small (in particular, smaller than  $2\lambda$ ) we can find  $\delta = \delta(\eta, \varepsilon)$  such that if  $q \leq X^{3/4+1/1044-\varepsilon}$  is smooth and squarefree, then

$$\Delta_{\mu^2}(X; q, a) \ll_{\varepsilon} X^{1-\delta}/q,$$

for any residue class  $a$  modulo  $q$ .

**6.3. Non-squarefree  $q > X^{3/4-\varepsilon}$**

The proof follows similar lines to that of Theorem 1.1 but invoking Proposition 2.4 ii) rather than i). The choice of parameters can be rigged up similarly as in the previous proof, save that the factors  $\tilde{q}_0, \dots, \tilde{q}_L$  must be chosen to satisfy  $K \geq \max\{\tilde{q}_1, \dots, \tilde{q}_L\}$  and also

$$\max \left\{ \left( \frac{\tilde{q}_0^{1-\delta'}}{K} \right)^{2^{-L}}, \left( \frac{\tilde{q}_{L-j+1}}{K} \right)^{2^{-j}} \right\} \ll X^{-\gamma},$$

in analogy to (36), which amounts to replacing the condition  $\tilde{q}_0^{-2^{-(L+1)}} \ll X^{-\gamma}$  by  $\tilde{q}_0^{-(1-\delta')2^{-L}} \ll X^{\gamma}$ . We also must choose  $\tilde{q}_0$  to be coprime to 2 and 3 in order to apply the results of Section 4, but by the  $X^{\eta}$ -ultra-smooth condition this can be done at a cost of  $X^{2\eta}$  in precision in the choice of  $\tilde{q}_1$  (as is done explicitly in Lemma 6.1 b).

Finally, in order to apply Lemma 5.3 we must assume  $K/\tilde{q}_j \geq \tilde{q}_0^{\delta'}$  for all  $1 \leq j \leq L$ , where  $\delta' = \delta'(L) \in (0, 2^{-2^L}]$  arises in Proposition 2.4. This can be assured in light of the choice  $\sigma$  from the previous subsection: up to  $X^{\eta}$  factors (where  $\eta$  can be chosen as small as desired), we have, uniformly in  $1 \leq j \leq L$ ,

$$K/\tilde{q}_j \geq \frac{1}{2} X^{2^j \gamma + 2\delta} \geq \frac{1}{2} X^{4\gamma + 2\delta},$$

with  $\gamma = 2\delta + \lambda$ , whereas

$$\tilde{q}_0^{\delta'} \leq 2X^{2\sigma\delta' + \delta'\eta L} \leq X^{\delta'(2/L + \eta L) + (2^{L+3} + 2L)\gamma\delta'/L},$$

so that as  $\delta' \leq 2^{-2^L}$  it suffices, for instance, to have  $\gamma + \delta > \delta'(1/L + \eta L)$ . We leave to the interested reader the determination of an explicit choice of  $\lambda$  and  $\delta$  (both of which necessarily depend on  $\delta'$ , which could be reduced if necessary) in which the range  $q \leq X^{3/4+\lambda}$  is admissible with power savings  $X^{1-\delta}/q$ .

*Proof of Corollary 1.3.* Put  $Y := X^{196/261-\varepsilon}$  and set  $u := (\log Y)/(\eta \log X)$ . It suffices to show that

$$|\{q \leq Y : P^+(q) \leq X^\eta, \mu^2(q) = 1\}| \gg_\eta Y.$$

Of course, we have

$$\sum_{q \leq Y} \mu^2(q) 1_{P^+(q) \leq X^\eta} = \sum_{d \leq Y^{1/2}} \mu(d) 1_{P^+(d) \leq X^\eta} \sum_{m \leq Y/d^2} 1_{P^+(m) \leq X^\eta}.$$

Put  $D := (\log X)^{1/2}$ . Then we trivially bound

$$\sum_{D < d \leq Y^{1/2}} \mu(d) 1_{P^+(d) \leq X^\eta} \sum_{m \leq Y/d^2} 1_{P^+(m) \leq X^\eta} \ll Y \sum_{d > D} d^{-2} \ll Y (\log X)^{-1/2}.$$

On the other hand, standard estimates for smooth numbers (see, e.g., Theorem III.5.8 of [24]) yield

$$\begin{aligned} & \sum_{d \leq D} \mu(d) 1_{P^+(d) \leq X^\eta} \sum_{m \leq Y/d^2} 1_{P^+(m) \leq X^\eta} \\ &= Y \sum_{d \leq D} \frac{\mu(d) \rho(\log(Y/d^2)/\eta \log X)}{d^2} 1_{P^+(d) \leq X^\eta} + O_\eta(YD/\log X) \\ &= \frac{6}{\pi^2} \rho(u) Y + O_\eta \left( Y \sum_{d \leq D} \frac{|\rho(u) - \rho(u - v_d)|}{d^2} + Y (\log X)^{-1/2} \right), \end{aligned}$$

where  $\rho$  denotes the Dickman function and for each  $d \leq D$  we set  $v_d := \frac{2(\log d)}{\eta \log X}$ . As

$$w \rho'(w) = -\rho(w - 1)$$

for  $w > 1$ , we observe that for each  $d \leq D$ ,

$$\begin{aligned} |\rho(u - v_d) - \rho(u)| &= \left| \int_{u-v_d}^u \rho'(t) dt \right| \leq v_d \max_{0 \leq t \leq v_d} |\rho'(u - t)| \\ &\ll_\eta \frac{\log \log X}{\log X} \frac{\rho(u - 2)}{u} \ll_\eta \frac{\log \log X}{\log X}, \end{aligned}$$

provided  $X$  is large enough in terms of  $\eta$ . We thus deduce that

$$|\{q \leq Y : P^+(q) \leq X^\eta, \mu^2(q) = 1\}| = \frac{6}{\pi^2} \rho(u) Y \left( 1 + O_\eta(1/\sqrt{\log X}) \right) \gg_\eta Y,$$

as  $u \ll_\eta 1$ , and the claim follows. □

**Acknowledgements.** The author is grateful to the anonymous referees for a careful reading of the article, for several corrections to the content of Section 3 and for many recommendations that helped in improving the exposition. The author warmly thanks Corentin Perret-Gentil for invaluable discussions and suggested references that were crucial to the proof of Theorem 3.1, as well as many helpful comments contributing to better readability. The author would also like to thank Aled Walker for a useful suggestion leading to an improvement to the arguments in the proof of Theorem 4.1. This article began while the author was visiting the California Institute of Technology in February 2020. The author would like to thank Caltech for the excellent working conditions, as well as Maksym Radziwiłł for the invitation and for bringing the work [16] to his attention.

**Conflicts of Interest:** None.

**References**

[1] E. Bombieri, ‘On exponential sums in finite fields’, *Am. J. Math.* **88**(1) (1966), 71–105.  
 [2] J. Bourgain, C. Demeter and L. Guth, ‘Proof of the main conjecture in Vinogradov’s mean value theorem for degrees higher than three’, *Ann. Math.* **184**(2) (2016), 633–682.

- [3] P. Deligne, *Cohomologie étale*, Vol. **569** of *Lecture Notes in Mathematics* (Springer, Berlin, 1977). Séminaire de géométrie algébrique du Bois-Marie SGA 4 $\frac{1}{2}$ .
- [4] P. Deligne, 'La conjecture de Weil. II', *Publ. Math. Inst. Hautes Études Sci.* **52**(1) (1980), 137–252.
- [5] E. Fouvry, S. Ganguly, E. Kowalski and P. Michel, 'Gaussian distribution for the divisor function and Hecke eigenvalues in arithmetic progressions', *Comment. Math. Helv.* **89**(4) (2014), 979–1014.
- [6] É. Fouvry, E. Kowalski and P. Michel, 'A study in sums of products', *Philos. Trans. A* **373**(2040) (2015), 1–126.
- [7] D. R. Heath-Brown, 'The largest prime factor of  $x^3 + 2$ ', *Proc. London Math. Soc.* **82**(3) (2001), 554–596.
- [8] C. Hooley, 'A note on square-free numbers in arithmetic progressions', *Bull. Lond. Math. Soc.* **7**(2) (1975), 133–138.
- [9] A. J. Irving, 'The divisor function in arithmetic progressions to smooth moduli', *Int. Math. Res. Not.* **15** (2015), 6675–6698.
- [10] H. Iwaniec and E. Kowalski, *Analytic Number Theory* (Colloquium Publications, American Mathematical Society, Providence, RI, 2004).
- [11] N. M. Katz, *Gauss Sums, Kloosterman Sums, and Monodromy Groups*, Vol. **116** of *Annals of Mathematics Studies* (Princeton University Press, Princeton, NJ, 1988).
- [12] N. M. Katz, *Exponential Sums and Differential Equations*, Vol. **124** of *Annals of Mathematics Studies* (Princeton University Press, Princeton, NJ, 1990).
- [13] E. Kowalski and G. Ricotta, Fourier coefficients of  $GL(N)$  automorphic forms in arithmetic progressions, *Geom. Funct. Anal.* **24** (2014), 1229–1297.
- [14] K. Liu, I. E. Shparlinski and T. Zhang, 'Average distribution of  $k$ -free integers in arithmetic progressions', *Math. Nach.* **293**(8), 1505–1514.
- [15] D. Miličević and S. Zhang, 'Distribution of Kloosterman paths to high prime power moduli', Preprint, 2020, [arXiv:2005.08865v1\[math.NT\]](https://arxiv.org/abs/2005.08865v1).
- [16] R. M. Nunes, 'On the least squarefree number in an arithmetic progression', *Mathematika* **63**(2) (2017), 483–498.
- [17] C. Perret-Gentil, *Probabilistic Aspects of Short Sums of Trace Functions over Finite Fields*, PhD thesis (ETH Zürich, 2016).
- [18] C. Perret-Gentil, 'Gaussian distribution of short sums of trace functions over finite fields', *Math. Proc. Cambridge Philos. Soc.* **163**(3) (2017), 385–422.
- [19] K. Prachar, 'Über die kleinste quadratfreie zahl einer arithmetischen reihe', *Monatsh. Math.* **62** (1958), 173–176.
- [20] G. Ricotta and E. Royer, 'Kloosterman paths of prime power moduli', *Comm. Math. Helv.* **93** (2018), 493–532.
- [21] C. J. Ringrose, The  $q$ -analogue of van der Corput's method, DPhil Thesis, Oxford, 1985.
- [22] P. Shiu, 'A Brun–Titchmarsh theorem for multiplicative functions', *J. für die Reine und Angew. Math.* **313** (1980), 161–170.
- [23] R. S. Steiner, 'Effective Vinogradov's mean value theorem via efficient boxing', *J. Number Theory* **204** (2019), 354–404.
- [24] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, 3rd ed., Vol. **163** of *Graduate Studies in Mathematics* (American Mathematical Society, Providence, RI, 2015).
- [25] T. D. Wooley, 'Nested efficient congruencing and relatives of Vinogradov's mean value theorem', *Proc. London Math. Soc.* **118**(4) (2019), 942–1016.