A DECOUPLING INTERPRETATION OF AN OLD ARGUMENT FOR VINOGRADOV'S MEAN VALUE THEOREM

BRIAN COOK, KEVIN HUGHES, ZANE KUN LI, AKSHAT MUDGAL, OLIVIER ROBERT, AND PO-LAM YUNG

ABSTRACT. We interpret into decoupling language a refinement of a 1973 argument due to Karatsuba on Vinogradov's mean value theorem. The main goal of our argument is to answer what precisely solution counting in older partial progress on Vinogradov's mean value theorem corresponds to in Fourier decoupling theory.

1. INTRODUCTION

1.1. **Motivation.** Let $s \geqslant 1$ and $k \geqslant 2$ be integers. For $X \geqslant 1$, let $J_{s,k}(X)$ be the number of solutions to the degree k Vinogradov system in $2s$ variables:

$$x_1^j + x_2^j + \cdots + x_s^j = y_1^j + y_2^j + \cdots + y_s^j, \quad 1 \leqslant j \leqslant k \tag{1.1}$$

where all variables $x_1, \ldots, x_s, y_1, \ldots, y_s \in [1, X] \cap \mathbb{N}$. Nontrivial upper bounds for $J_{s,k}(X)$ were first studied by Vinogradov in 1935 [32] and such results are collectively referred to as Vinogradov's Mean Value Theorem (VMVT) in the literature. The main conjecture in VMVT, now a theorem as of 2015, was that for every $\varepsilon > 0$ and $s, k \in \mathbb{N}$, one has

$$J_{s,k}(X) \lesssim_{s,k,\varepsilon} X^\varepsilon (X^s + X^{2s - \frac{k(k+1)}{2}}) \tag{1.2}$$

for all $X \geqslant 1$. It is not hard to see that $J_{s,k}(X) \gtrsim_{s,k} X^s + X^{2s - k(k+1)/2}$ and applying Hölder's inequality, we may deduce (1.2) for all $s \in \mathbb{N}$ from the $s = k(k+1)/2$ case. VMVT plays an important role in understanding Waring's problem and the Riemann zeta function, see for example [11, 12, 19, 34]. When $k = 2$, the main conjecture in VMVT is classical. In 2014, Wooley [35] proved the $k = 3$ case of VMVT using the method of efficient congruencing (see also [20] for a shorter proof due to Heath-Brown). In 2015, the $k \geqslant 2$ case was proven by Bourgain, Demeter, and Guth in [3] using Fourier decoupling for the degree k moment curve from which VMVT followed as a corollary. Finally, in 2017, Wooley [36], gave an alternative proof of (1.2) for all $k \geqslant 2$ using nested efficient congruencing.

After the proofs of VMVT using the Fourier method of decoupling [3] and the number theoretic method of efficient congruencing [36], it has been an interesting question to determine how these two methods are related and whether a "dictionary" between the two methods could be obtained. The study of this dictionary has led to new proofs of Fourier decoupling for the parabola [23], cubic moment curve [15], and the degree k moment curve [16]; these having been inspired from the efficient congruencing arguments in [26, Section 4], [20], and [36], respectively. Additionally, a decoupling interpretation of the study of VMVT over ellipsephic sets [1] led to a proof of Fourier decoupling for fractal sets on the parabola [5].

In this article, we revisit a particular classical VMVT which states that

$$J_{s,k}(X) \lesssim_{s,k} X^{2s - \frac{k(k+1)}{2} + \frac{1}{2}k^2(1 - \frac{1}{k})^{s/k}} \tag{1.3}$$

for all $X \geqslant 1$ and $s = kl$ with $l \in \mathbb{N}$. This result should be compared to the supercritical $s \geqslant k(k+1)/2$ case in (1.2). For s very large compared to k, we have an extra term $\frac{1}{2}k^2(1 - \frac{1}{k})^{s/k}$ in the exponent, which decays exponentially in s for every fixed value of k, instead of an ε. The estimate

(1.3) appears (for example) in Vaughan's book [31, Chapter 5] and is a refinement of an argument of Karatsuba [22] from 1973 (see also Stechkin [27] from 1975). The loss of the $X^{\frac{1}{2}k^2(1-\frac{1}{k})^{s/k}}$ comes from combining the subcritical estimate $J_{k,k}(X) \lesssim_k X^k$, which follows from the Newton-Girard identities, along with an iterative argument to derive estimates for $J_{s,k}(X)$ when s is supercritical.

The main purpose of this paper is to illustrate how this refined argument of Karatsuba can be adapted to give a proof of a non-sharp Fourier decoupling inequality for the degree k moment curve in the supercritical regime. The key difficulty that prevents the direct use of ideas from [15, 16, 23] is the heavy reliance on solution counting in (1.3). One of the main points of this article is to clarify the role of such solution counting arguments in the study of Fourier decoupling. The mechanism driving the solution counting arguments will allow us to prove the key Lemma 4.4 below, which concerns the geometry of Fourier supports of the functions appearing in our main Theorem 1.1.

Since our goal is to clarify the role of solution counting in Fourier decoupling and Bourgain, Demeter, and Guth have already given the sharpest possible moment curve decoupling theorem in [3], we will work over $\mathbb{Q}_q$ rather than over $\mathbb{R}$. This will allow us to present the argument in the cleanest possible manner, free of technical difficulties arising from the inconvenience of the uncertainty principle in $\mathbb{R}^k$. See also [14] for another decoupling paper that works over $\mathbb{Q}_q$ rather than $\mathbb{R}$, there however, the authors use the observation that decoupling over $\mathbb{Q}_q$ is quantitatively more efficient than decoupling over $\mathbb{R}$ in terms of exponential sum estimates.

Notation. As k will be fixed, we will allow all constants to depend on k. Given two positive expressions X and Y, we write $X \lesssim Y$ if $X \leqslant CY$ for some constant C that is allowed to depend on k. If C depends on some additional parameter A, then we write $X \lesssim_A Y$. We write $X \sim Y$ if $X \lesssim Y$ and $Y \lesssim X$. By writing $f(x) = O(g(x))$, we mean $|f(x)| \lesssim g(x)$. We say that f has Fourier support in a set Ω if its Fourier transform $\hat{f}$ is supported in Ω.

To prepare the reader for the myriad of intervals that will occur later in Sections 4 and 5, there will be three types of interval lengths: intervals named with a "K" will be associated to the smallest scale δ, intervals named with a "J" will be associated to the intermediate scale $\nu \approx \delta^{1/k}$, and intervals named with an "I" will be associated to the largest scale $\kappa \approx \delta^\varepsilon$ (though on a first reading, it might be easier to set $\kappa = 1/q$). Finally, in the context of the decoupling constant $\mathfrak{D}_p(\delta)$, defined in (1.5) below, we call p subcritical if $p < k(k+1)$ and p supercritical if $p \geqslant k(k+1)$ (rather than the more accurate but slightly more clumsy "not subcritical").

1.2. **Analysis over $\mathbb{Q}_q$ and decoupling.** Fix a degree $k \geqslant 2$ and a prime number q with $q > k$. We reserve the letter p for the Lebesgue exponent in the main Theorem 1.1. We very briefly review the harmonic analysis over $\mathbb{Q}_q$ needed to set up the statement of decoupling. See also Section 2 and [14, Section 2] for further discussion surrounding the harmonic analysis and basic geometric facts over $\mathbb{Q}_q$ that are useful in decoupling. Additionally see Chapters 1 and 2 of [28] and Chapter 1 (in particular Sections 1 and 4) of [33] for a more complete discussion of analysis on $\mathbb{Q}_q$.

The field $\mathbb{Q}_q$ is the completion of $\mathbb{Q}$ under the q-adic norm, defined by $|0| = 0$ and $|q^a b/c| = q^{-a}$ if $a \in \mathbb{Z}$, $b, c \in \mathbb{Z}\backslash\{0\}$ and q is relatively prime to both b and c. Then $\mathbb{Q}_q$ can be identified (bijectively) with the set of all formal series

$$\mathbb{Q}_q = \Big\{ \sum_{j=k}^{\infty} a_j q^j : k \in \mathbb{Z}, a_j \in \{0, 1, \ldots, q-1\} \text{ for every } j \geqslant k \Big\},$$

and the q-adic norm on $\mathbb{Q}_q$ satisfies $|\sum_{j=k}^{\infty} a_j q^j| = q^{-k}$ if $a_k \neq 0$. Strictly speaking we should be writing $|\cdot|_q$ instead of $|\cdot|$, but we omit this dependence as q is fixed. The q-adic norm on $\mathbb{Q}_q$ induces a norm on $\mathbb{Q}_q^k$, which we denote also by $|\cdot|$ by abuse of notation, via $|(\xi_1, \ldots, \xi_k)| := \max_{1 \leqslant i \leqslant k} |\xi_i|$. Of particular importance is the ultrametric inequality: $|\xi + \eta| \leqslant \max\{|\xi|, |\eta|\}$ with equality if $|\xi| \neq |\eta|$. An interval in $\mathbb{Q}_q$ is then a set of the form $\{\xi \in \mathbb{Q}_q : |\xi - a| \leqslant r\}$, where $a \in \mathbb{Q}_q$ and $r \geqslant 0$; r will then be called the length of the interval. We also will use $|I|$ to denote the length of

an interval I. The ring of integers $\mathbb{Z}_q$ coincides with the unit interval $\{\xi \in \mathbb{Q}_q : |\xi| \leqslant 1\}$. A cube in $\mathbb{Q}_q^k$ of side length r is then a product of k intervals in $\mathbb{Q}_q$ of lengths r. We will work with Schwartz functions defined on $\mathbb{Q}_q^k$ (i.e. finite linear combinations of characteristic functions of cubes in $\mathbb{Q}_q^k$). The Fourier transform of such a function f will be given by

$$\widehat{f}(\xi) := \int_{\mathbb{Q}_q^k} f(x)\chi(-x \cdot \xi)dx$$

where χ is a fixed element in the Pontryagin dual $\widehat{\mathbb{Q}_q}$ of $\mathbb{Q}_q$ that restricts to the principal character on the additive subgroup $\mathbb{Z}_q$ and restricts to a non-principal character on the additive subgroup $q^{-1}\mathbb{Z}_q$, $x \cdot \xi = \sum_{i=1}^k x_i\xi_i$ if $x = (x_1, \ldots, x_k)$ and $\xi = (\xi_1, \ldots, \xi_k)$, and dx is the Haar measure on the additive group $\mathbb{Q}_q^k$ normalized so that $\int_{\mathbb{Z}_q^k} dx = 1$. One key property of the Fourier transform that we will use is that $\widehat{1_{\mathbb{Z}_q}} = 1_{\mathbb{Z}_q}$, that is, the Fourier transform of the unit ball is the unit ball, see [33, p.42] for a proof.

We are interested in the unit moment curve

$$\gamma(t) := (t, t^2, \ldots, t^k), \quad |t| \leqslant 1.$$

For $\delta \in q^{-\mathbb{N}}$ and any interval $I \subset \mathbb{Q}_q$ with length $\geqslant \delta$, let $P_\delta(I)$ be a partition of I into intervals of length δ. Write P_δ for $P_\delta(\mathbb{Z}_q)$. To each interval $I \subset \mathbb{Z}_q$, one associates a parallelepiped

$$\theta_I := \left\{ \gamma(a) + \sum_{j=1}^k t_j\gamma^{(j)}(a) \in \mathbb{Q}_q^k : |t_j| \leqslant |I|^j \text{ for all } 1 \leqslant j \leqslant k \right\}$$

of dimensions $|I| \times |I|^2 \times \cdots \times |I|^k$ where $a \in I$; this parallelepiped is independent of the choice of $a \in I$. Note that $\bigcup_{K \in P_\delta} \theta_K$ is a covering of a δ^k neighborhood of the unit moment curve (in fact it covers a suitable anisotropic neighborhood of that curve). One also associates to each $K \in P_\delta$ a cube

$$\tau_K := \{(\xi_1, \ldots, \xi_k) \in \mathbb{Q}_q^k : |\xi_j - a^j| \leqslant \delta \text{ for all } 1 \leqslant j \leqslant k\} \tag{1.4}$$

of side length δ, where $a \in K$; again this is independent of the choice of $a \in K$. Note that for each $K \subset P_\delta$, the ultrametric inequality gives that $\theta_K \subset \tau_K$.

For an interval $I \subset \mathbb{Z}_q$, let f_I be defined such that $\widehat{f}_I := \widehat{f} \cdot 1_{I \times \mathbb{Q}_q^{k-1}}$. For $p \geqslant 2$ and $\delta \in q^{-\mathbb{N}}$, let $\mathfrak{D}_p(\delta)$ be the smallest constant such that the inequality

$$\|f\|_{L^p(\mathbb{Q}_q^k)} \leqslant \mathfrak{D}_p(\delta) \Big(\sum_{K \in P_\delta} \|f_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{1/2} \tag{1.5}$$

holds for every Schwartz function f on $\mathbb{Q}_q^k$ with its Fourier transform $\widehat{f}$ supported on $\bigcup_{K \in P_\delta} \theta_K$. Note that $f = \sum_{K \in P_\delta} f_K$. Bourgain, Demeter, and Guth [3] showed that

$$\mathfrak{D}_p(\delta) \lesssim_{\varepsilon, p, q} \delta^{-\varepsilon} \big(1 + \delta^{-(\frac{1}{2} - \frac{k(k+1)}{2p})} \big), \tag{1.6}$$

and this estimate is sharp. Strictly speaking [3] proves a decoupling theorem over $\mathbb{R}$ rather than over $\mathbb{Q}_q$, but the same proof can be used to derive (1.6). Choosing f to be a sum of Dirac deltas immediately implies (1.2).

1.3. **The main result.** By interpreting the refinement of Karatsuba's argument for (1.3) into decoupling language, our main result is then the following Fourier decoupling analogue of (1.3). In the same way (1.3) is a weaker partial result towards (1.2), Theorem 1.1 and Corollary 1.2 should be viewed as the analogous weaker counterpart of the sharp bound (1.6).

Theorem 1.1. *Let $p_0 \in 2\mathbb{N}$ be an even integer and let $c(p_0) \geqslant 0$ be such that*

$$\mathfrak{D}_{p_0}(\delta) \leqslant C_1 \delta^{-(\frac{1}{2} - \frac{k(k+1)}{2p_0}) - \frac{c(p_0)}{p_0}(1 - \frac{1}{k})^{p_0/(2k)}} \quad \text{for all } \delta \in q^{-\mathbb{N}} \tag{1.7}$$

where C_1 is independent of δ. If $p \in p_0 + 2k\mathbb{N}$ and $0 < \varepsilon < 1$, then

$$\mathfrak{D}_p(\delta) \lesssim_{p,\varepsilon,C_1} q^{a(p,p_0)/p} \delta^{-(\frac{1}{2} - \frac{k(k+1)}{2p}) - \frac{c(p_0)}{p}(1 - \frac{1}{k})^{p/(2k)} - \varepsilon} \quad \text{for all } \delta \in q^{-\mathbb{N}} \tag{1.8}$$

where

$$a(p, p_0) := (\frac{p - p_0}{2k})(\frac{p_0}{2} + \frac{k^2 + 7k - 4}{2}) + \frac{k}{2}(\frac{p - p_0}{2k})(\frac{p - p_0}{2k} + 1). \tag{1.9}$$

Since $\mathfrak{D}_p(\delta) \geqslant 1$ for all p, (1.7) implies that $c(p_0)$, k, and p_0 are such that

$$\frac{1}{2} - \frac{k(k+1)}{2p_0} + \frac{c(p_0)}{p_0}(1 - \frac{1}{k})^{\frac{p_0}{2k}} \geqslant 0. \tag{1.10}$$

It is also known that $\mathfrak{D}_{2k}(\delta) \lesssim_\varepsilon \delta^{-\varepsilon}$ for any $\varepsilon > 0$, see for example [8, Exercise 11.19] for the Euclidean case; we provide a proof for the case over $\mathbb{Q}_q$ in the appendix for the convenience of the reader. We also remark that [21] proved, in the case of local fields, a related square function estimate with a bound independent of δ if the f_K's are Fourier supported in a δ^k neighborhood of $\gamma(K)$; see also [13] and [2] for similar estimates. Choosing $p_0 = 2k$ and $c(p_0) = k^2/2 + \varepsilon$ for any $\varepsilon > 0$ in applying Theorem 1.1 we obtain:

Corollary 1.2. *Let $p \in 2k\mathbb{N}$ and $0 < \varepsilon < 1$. Then*

$$\mathfrak{D}_p(\delta) \lesssim_{p,\varepsilon} q^{O(k+p/k)} \delta^{-(\frac{1}{2} - \frac{k(k+1)}{2p}) - \frac{k^2}{2p}(1 - \frac{1}{k})^{p/(2k)} - \varepsilon} \quad \text{for all } \delta \in q^{-\mathbb{N}}$$

where the implied constant in the exponent of q is absolute (and independent of k).

The exponent of q in Corollary 1.2 is more precisely $\frac{a(p,2k)}{p} = (\frac{1}{2k} - \frac{1}{p})\frac{k^2 + 9k - 4}{2} + \frac{1}{4}(\frac{p}{2k} - 1)$, but we opt to write it as above since it more clearly illustrates what the main terms are. Note that the hypothesis in Theorem 1.1 is always satisfied if p_0 is any fixed exponent $\geqslant 2$ and $c(p_0)$ is chosen large enough. One can view Theorem 1.1 as a way of upgrading trivial $l^2 L^{p_0}$ decoupling at say some subcritical p to $l^2 L^p$ decoupling for all large p with only a loss that decreases exponentially as $p \to +\infty$. Of course, if one already knew the sharp estimate in the critical $p_0 = k(k+1)$ case, then Theorem 1.1 implies that we know the sharp decoupling estimate for all $p \in k(k+1) + 2k\mathbb{N}$. However this already follows from interpolating the critical estimate with the trivial $l^2 L^\infty$ decoupling estimate.

Though Corollary 1.2 implies (1.3) with an extra X^ε that comes from the $\delta^{-\varepsilon}$ factor in Corollary 1.2, Corollary 1.2 is more general and this extra $\delta^{-\varepsilon}$ term comes from needing some additional uniformity in the case of the general f Fourier supported in $\bigcup_{K \in P_\delta} \theta_K$ and an application of the broad-narrow argument to get around the use of the Prime Number Theorem in the proof of (1.3) (see Section 4.1.1). See Sections 3.5 and 5.1 for some more discussion comparing the VMVT case and the general f decoupling case.

We end with some discussion about how the proof of Corollary 1.2 (and Theorem 1.1) contrasts with modern decoupling proofs of degree k moment curve decoupling [3, 16] which prove (1.6). Unlike the argument in [3, 16], we are missing any lower dimensional decoupling input and while we do use induction on scales, the iteration itself is unique in that it iterates on the p in $l^2 L^p$ decoupling. Schematically, the iteration to prove Theorem 1.1 controls $l^2 L^p$ decoupling by $l^2 L^{p-2k}$ decoupling at a larger scale. After $O(p/k)$ steps, we are reduced to $l^2 L^{2k}$ decoupling for the degree k moment curve which follows (essentially) from the Newton-Girard identities. The iteration is surprisingly efficient when it controls $l^2 L^p$ decoupling by $l^2 L^{p-2k}$ decoupling as long as both p and $p - 2k$ are supercritical. However after about $\frac{1}{2k}(p - \frac{k(k+1)}{2})$ steps, we enter the subcritical regime

for which the iteration becomes inefficient and this is why we accrue an additional $\delta^{-\frac{k^2}{2p}(1-\frac{1}{k})^{p/(2k)}}$ term. When $k = 2$, the argument for Corollary 1.2 uses $O(p)$ steps to prove a weak non-sharp $l^2 L^p$ decoupling estimate. This is to be compared to the modern proof of decoupling for the parabola where to prove the sharp critical $l^2 L^6$ decoupling, one uses $O(\varepsilon^{-1})$ many steps (see for example the proof of [23, Lemma 2.12]). In the harmonic analysis literature, iterating on p is not a new idea as such an argument was already used by Drury [9] to prove cubic moment curve restriction, though we believe this is the first time such an argument has appeared in the decoupling literature. See also [25] by the fourth author for a similar idea in the additive combinatorics literature which was recently used to obtain diameter free estimates for the quadratic VMVT.

Additionally, at each iterative step, three scales are key: the smallest scale δ, the intermediate scale $\delta^{1/k}$, and the largest scale 1 (though strictly speaking in our proof the largest scale is actually δ^ε rather than 1 for technical reasons). This can be compared to [3, 16] which uses scales δ, δ^ε and 1.

This paper is organized as follows: In Section 2, we review some basic geometric and harmonic analysis facts in $\mathbb{Q}_q$ that will be used throughout this paper. In Section 3, we review the refinement of the 1973 argument of Karatsuba at a high level. In Section 4, we prove Lemma 4.2 which is the main lemma that is used to prove Theorem 1.1. This is accomplished via combining a standard broad-narrow argument in Section 4.1.1 and some geometric properties of the moment curve that use the Newton-Girard identites, see Lemma 4.4. In Section 5, we dyadically pigeonhole to obtain some uniformity in our estimates and prove Theorem 1.1 and Corollary 1.2. Finally, in the appendix, we include a proof of $\mathfrak{D}_{2k}(\delta) \lesssim_\varepsilon \delta^{-\varepsilon}$ for completeness.

2. Wavepacket decomposition and some basic geometric facts

Throughout this paper, we will make use of wavepacket decomposition which allows us to decompose a function f, which is Fourier supported in some θ_K, into linear combinations of indicator functions of translates of the parallelpiped "dual" to θ_K. That the q-adic character χ is trivial on $\mathbb{Z}_q$ gives a much cleaner wavepacket decomposition when working over $\mathbb{Q}_q$ than over $\mathbb{R}$. See [30, Section 3] or [17, Section 2.4] for some discussion about wavepacket decomposition over $\mathbb{R}$ in the context of the paraboloid (though the same ideas apply for the degree k moment curve).

Fix $\delta \in q^{-\mathbb{N}}$. It will be convenient to introduce the shorthand

$$\theta_\delta := \delta \mathbb{Z}_q \times \delta^2 \mathbb{Z}_q \times \cdots \times \delta^k \mathbb{Z}_q$$

and

$$T_\delta := \delta^{-1} \mathbb{Z}_q \times \delta^{-2} \mathbb{Z}_q \times \cdots \times \delta^{-k} \mathbb{Z}_q.$$

They are dual to each other in the sense that

$$T_\delta = \{x \in \mathbb{Q}_q^k : |x \cdot \xi| \leqslant 1 \text{ for all } \xi \in \theta_\delta\}.$$

Since for any $1 \leqslant j \leqslant k$, any interval in $\mathbb{Q}_q$ of length δ^j is the disjoint union of $\delta^{-(k-j)}$ many intervals of length δ^k, it follows that θ_δ is the disjoint union of $\delta^{-\frac{k(k-1)}{2}}$ many cubes of side lengths δ^k in $\mathbb{Q}_q^k$. Similarly, any cube in $\mathbb{Q}_q^k$ of side length δ^{-k} is a disjoint union of $\delta^{-\frac{k(k-1)}{2}}$ many translates of T_δ.

Now for $a \in \mathbb{Z}_q$, let M_a be the $k \times k$ lower-triangular matrix given by

$$M_a = (\gamma'(a) \ \gamma''(a) \ \cdots \ \gamma^{(k)}(a))$$

where we view $\gamma^{(j)}(a)$ as a column vector. Then for any $K \in P_\delta$, we have

$$\theta_K = \gamma(a) + M_a \theta_\delta \tag{2.1}$$

for any $a \in K$. In fact, the right hand side is independent of $a \in K$ since if $b \in K$, then

$$\gamma(b) = \gamma(a) + \sum_{j=1}^{k} (j!)^{-1} \gamma^{(j)}(a)(b-a)^j \in \gamma(a) + M_a \theta_\delta,$$

and

$$M_a = M_b \begin{pmatrix} 1 & 0 & \cdots & 0 \\ (1!)^{-1}(b-a) & 1 & \cdots & 0 \\ (2!)^{-1}(b-a)^2 & (1!)^{-1}(b-a) & \cdots & 0 \\ \vdots & & \ddots & \\ ((k-1)!)^{-1}(b-a)^{k-1} & ((k-2)!)^{-1}(b-a)^{k-2} & \cdots & 1 \end{pmatrix} \tag{2.2}$$

where the second matrix on the right hand side preserves $\theta_\delta = \delta \mathbb{Z}_q \times \delta^2 \mathbb{Z}_q \times \cdots \times \delta^k \mathbb{Z}_q$ (here we have used the fact that $|k!| = 1$ in $\mathbb{Q}_q$ since $q > k$).

For $K \in P_\delta$ and any $a \in K$, let $T_{0,K}$ be the dual parallelepiped to θ_K centered at the origin given by

$$T_{0,K} = \{x \in \mathbb{Q}_q^k : |x \cdot (\xi - \gamma(a))| \leqslant 1 \text{ for all } \xi \in \theta_K\}.$$

Using (2.1), it is not hard to see that

$$\begin{aligned} T_{0,K} &= \{x \in \mathbb{Q}_q^k : |x \cdot \gamma^{(j)}(a)| \leqslant \delta^{-j} \text{ for all } 1 \leqslant j \leqslant k\} \\ &= \{x \in \mathbb{Q}_q^k : M_a^T x \in T_\delta\} = M_a^{-T} T_\delta \end{aligned}$$

for any $a \in K$. This parallelepiped depends only on K but not on the choice of $a \in K$, since (2.2) shows that

$$M_a^{-T} = M_b^{-T} \begin{pmatrix} 1 & O(\delta) & O(\delta^2) & \cdots & O(\delta^{k-1}) \\ 0 & 1 & O(\delta) & \cdots & O(\delta^{k-2}) \\ 0 & 0 & 1 & \cdots & O(\delta^{k-3}) \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

where $O(\delta^j)$ is some number in $\mathbb{Q}_q$ with norm $\leqslant \delta^j$, and the second matrix on the right hand side is a bijection that preserves T_δ by the ultrametric inequality.

Lemma 2.1. *Let $\delta \in q^{-\mathbb{N}}$ and fix $K \in P_\delta$. Then*

(i) $\theta_K - \theta_K$ is the disjoint union of $\delta^{-\frac{k(k-1)}{2}}$ cubes of side lengths δ^k, and

(ii) every cube of side length δ^{-k} in $\mathbb{Q}_q^k$ is the disjoint union of $\delta^{-\frac{k(k-1)}{2}}$ many translates of $T_{0,K}$.

Proof. (i) Recall that θ_δ is the disjoint union of $\delta^{-\frac{k(k-1)}{2}}$ cubes of side lengths δ^k. Since M_a is a bijection that maps cubes of side length δ^k to cubes of side length δ^k for any $a \in K$, and $\theta_K - \theta_K = M_a\theta_\delta$ for any $a \in K$, the assertion follows. Note that $\theta_K - \theta_K$ is just a translation of θ_K to the origin.

(ii) Recall that any cube in $\mathbb{Q}_q^k$ of side length δ^{-k} is a disjoint union of $\delta^{-\frac{k(k-1)}{2}}$ many translates of T_δ. Since M_a^{-T} is a bijection that maps cubes of side length δ^{-k} to cubes of side length δ^{-k} for any $a \in K$, and $T_{0,K} = M_a^{-T}T_\delta$ for any $a \in K$, the assertion follows.

$\square$

From Lemma 2.1(ii), we may deduce that translates of $T_{0,K}$ tile $\mathbb{Q}_q^k$; we denote the collection of such translates by $\mathbb{T}(K)$. We are now ready to state the version of wavepacket decomposition that we will use.

Lemma 2.2 (Wavepacket decomposition). *Let $\delta \in q^{-\mathbb{N}}$ and fix $K \in P_\delta$. Let g be a Schwartz function with Fourier transform supported in θ_K. Then $|g|$ is constant on every $T \in \mathbb{T}(K)$, and $\widehat{g1_T}$ is supported on θ_K for every $T \in \mathbb{T}(K)$. Hence it is natural to write*

$$g = \sum_{T \in \mathbb{T}(K)} g1_T, \qquad (2.3)$$

where each term $g1_T$ (which we will call a "wavepacket") is Fourier supported on θ_K and has constant modulus on every $T \in \mathbb{T}(K)$. It also follows that if $\mathcal{T}$ is any subset of $\mathbb{T}(K)$, then $\sum_{T \in \mathcal{T}} g1_T$ is Fourier supported in θ_K.

Proof. First, to prove that $|g|$ is constant on any translates of $T_{0,K}$, one only needs to prove the case when $\delta = 1$, $K = \mathbb{Z}_q$, and then apply a change of variables, but we opt for a more explicit proof. We will show that $|g(x)|$ is constant for all $x \in A + T_{0,K}$ for any $A \in \mathbb{Q}_q^k$. By Fourier inversion we have that

$$|g(x)| = |\int_{\theta_K} \widehat{g}(\xi)\chi(\xi \cdot x)\, d\xi|$$

$$= |\int_{|t_1| \leqslant \delta, \ldots, |t_k| \leqslant \delta^k} \widehat{g}(\gamma(a) + \sum_{j=1}^k t_j\gamma^{(j)}(a))\chi([\gamma(a) + \sum_{j=1}^k t_j\gamma^{(j)}(a)] \cdot x)\, dt|$$

$$= |\int_{|t_1| \leqslant \delta, \ldots, |t_k| \leqslant \delta^k} \widehat{g}(\gamma(a) + M_a t)\chi(M_a^T x \cdot t)\, dt|.$$

For $x \in A + T_{0,K}$, we write $x = A + M_a^{-T}y'$ where $|y_j'| \leqslant \delta^{-j}$ for $j = 1, 2, \ldots, k$. Therefore

$$|g(x)| = |\int_{|t_1| \leqslant \delta, \ldots, |t_k| \leqslant \delta^k} \widehat{g}(\gamma(a) + M_a t)\chi(M_a^T A \cdot t)\chi(y' \cdot t)\, dt|$$

$$= |\int_{|t_1| \leqslant \delta, \ldots, |t_k| \leqslant \delta^k} \widehat{g}(\gamma(a) + M_a t)\chi(M_a^T A \cdot t)\, dt|$$

where we have used that $y' \cdot t \in \mathbb{Z}_q$, and so $\chi(y' \cdot t) = 1$. The right hand side is then independent of y' and so the above equality is true for all $x \in A + T_{0,K}$. In particular this shows that $|g|$ is a constant on $A + T_{0,K}$. This constant depends on K, g and A, but is a constant nonetheless.

Next, to prove that $\widehat{g1_T}$ is supported on θ_K, it suffices to observe that $\widehat{g1_T} = \widehat{g} * \widehat{1_T}$, and that $\widehat{1_T}$ is supported on $\theta_K - \theta_K$ for every $T \in \mathbb{T}(K)$: in fact, for every $T \in \mathbb{T}(K)$, $\widehat{1_T}$ is a modulation

of $\widehat{1_{T_{0,K}}}$, and if a is any point in K, then $T_{0,K} = M_a^{-T} T_\delta$. It follows that

$$\widehat{1_{T_{0,K}}}(\xi) = \int_{M_a^{-T} T_\delta} \chi(-x \cdot \xi) dx$$

$$= \det(M_a)^{-1} \int_{T_\delta} \chi(-M_a^{-T} y \cdot \xi) dy = \det(M_a)^{-1} \delta^{-k(k+1)/2} 1_{\theta_\delta}(M_a^{-1}\xi)$$

is supported on $M_a \theta_\delta = \theta_K - \theta_K$. Finally, the decomposition (2.3) follows since parallelepipeds in $\mathbb{T}(K)$ tile $\mathbb{Q}_q^k$. This completes the proof of the lemma. $\qquad\square$

3. Sketch of the Karatsuba argument

Before we dive into the proof of Theorem 1.1, we review the proof of (1.3) with an eye towards interpreting each step into decoupling language. See also, for example, [31, Section 5.1] or [29, Theorem 13 - Lemma 21] for more details of the number theoretic argument. Just for this section, we revert back to calling p a prime so as to best match these references.

3.1. **Step 1: Introducing some p-adic separation.** Given $X \geqslant 1$, one finds, using the Prime Number Theorem, a prime $p \sim X^{1/k}$ such that $J_{s,k}(X)$ is controlled by $J_{s,k}(X,p)$, where $J_{s,k}(X,p)$ is defined to be the number of solutions $(x_1, \ldots, x_s, y_1, \ldots, y_s) \in ([1,X] \cap \mathbb{N})^{2s}$ to (1.1) with the additional condition that $x_1, \ldots, x_k$ are pairwise distinct mod p and $y_1, \ldots, y_k$ are pairwise distinct mod p. Since p is rather large, this is a rather mild condition and so we heuristically should still expect $J_{s,k}(X) \approx J_{s,k}(X,p)$. The benefit of this extra p-adic separation (transversality) in these $2k$ variables is that we will get to apply Linnik's Lemma (in Step 3, (3.3) below) which will up to permutation uniquely determine these variables.

3.2. **Step 2: Applying the union bound/Hölder.** We now write $J_{s,k}(X,p)$ as

$$\int_{[0,1]^{2s}} \Big| \sum_{\substack{a_1,\ldots,a_k \pmod{p} \\ a_i \text{ pairwise distinct}}} \prod_{j=1}^k \sum_{\substack{n_j \equiv a_j \pmod{p} \\ 1 \leqslant n_j \leqslant X}} e(n_j \alpha_1 + \cdots + n_j^k \alpha_k)\Big|^2 \Big| \sum_{1 \leqslant n \leqslant X} e(n\alpha_1 + \cdots + n^k \alpha_k)\Big|^{2s-2k} d\alpha.$$

Write $|\sum_{1 \leqslant n \leqslant X}|^{2s-2k} = |\sum_{a \pmod p} \sum_{n \equiv a \pmod p}|^{2s-2k}$ and apply Hölder's inequality to control the above by

$$p^{2s-2k} \max_{a \pmod p} \int_{[0,1]^{2s}} \Big| \sum_{\substack{a_1,\ldots,a_k \pmod{p} \\ a_i \text{ pairwise distinct}}} \prod_{j=1}^k \sum_{\substack{n_j \equiv a_j \pmod{p} \\ 1 \leqslant n_j \leqslant X}} e(n_j \alpha_1 + \cdots + n_j^k \alpha_k)\Big|^2 \times$$

$$\Big| \sum_{\substack{n \equiv a \pmod{p} \\ 1 \leqslant n \leqslant X}} e(n\alpha_1 + \cdots + n^k \alpha_k)\Big|^{2s-2k} d\alpha. \tag{3.1}$$

Denote the integral above to be $J_{s,k}(X,p,a)$. This expression counts the number of solutions $(x_1, \ldots, x_s, y_1, \ldots, y_s) \in ([1,X] \cap \mathbb{N})^{2s}$ to (1.1) with $x_1, \ldots, x_k$ pairwise distinct mod p, $y_1, \ldots, y_k$ pairwise distinct mod p, and $x_{k+1} \equiv \cdots \equiv x_s \equiv y_{k+1} \equiv \cdots \equiv y_s \equiv a \pmod{p}$.

3.3. **Step 3: Solution counting.** Translation invariance of the Vinogradov system implies that we may bound $J_{s,k}(X,p,a)$ by $J_{s,k}(X,p,0)$. Rearrange the Vinogradov system (1.1) as

$$x_{k+1}^j + \cdots + x_s^j - y_{k+1}^j - \cdots - y_s^j = y_1^j + \cdots + y_k^j - x_1^j - \cdots - x_k^j, \quad 1 \leqslant j \leqslant k \tag{3.2}$$

where $x_1, \ldots, x_k$ are distinct mod p and $y_1, \ldots, y_k$ are distinct mod p and since we are considering $J_{s,k}(X,p,0)$, we have that $x_{k+1}, \ldots, x_s, y_{k+1}, \ldots, y_s \equiv 0 \pmod{p}$. Each choice of $x_1, \ldots, x_k, y_1, \ldots, y_k$ gives $\leqslant J_{s-k,k}(X/p)$ many solutions to $(x_{k+1}, \ldots, x_s, y_{k+1}, \ldots, y_s)$. To see this, write the count for

(3.2) as an integral and use the triangle inequality; the basic idea being that shifts of the Vinogradov system can only give fewer solutions.

Next, fixing one of the at most $J_{s-k,k}(X/p)$ many tuples $(x_{k+1}, \ldots, x_s, y_{k+1}, \ldots, y_s)$, how many valid $x_1, \ldots, x_k, y_1, \ldots, y_k$ are there? Since requiring $y_1, \ldots, y_k$ to be distinct mod p is a rather mild condition, there are $\leqslant X^k$ such $(y_1, \ldots, y_k)$. Any valid $(x_1, \ldots, x_k) \in ([1, X] \cap \mathbb{N})^k$ must satisfy

$$x_1^j + \cdots + x_k^j \equiv H_j \ (\mathrm{mod} \ p^j), \quad 1 \leqslant j \leqslant k$$

where the x_i are pairwise disjoint mod p for some H_j that depends on $(y_1, \ldots, y_k)$ (of which there are $\leqslant X^k$ many possibilities) and $(x_{k+1}, \ldots, x_s, y_{k+1}, \ldots, y_s)$ (of which there are $\leqslant J_{s-k,k}(X/p)$ many possibilities). Since $p^k \sim X$, instead of counting integers between 1 and X, we can count the x_i mod p^k. Thus it remains to count the number of residue classes $(x_1 \ (\mathrm{mod} \ p^k), \ldots, x_k \ (\mathrm{mod} \ p^k))$ such that

$$x_1^j + \cdots + x_k^j \equiv H_j \ (\mathrm{mod} \ p^j), \quad 1 \leqslant j \leqslant k \tag{3.3}$$

and $x_i \ (\mathrm{mod} \ p^k)$ are pairwise distinct mod p. Linnik's Lemma [24] then says that there are at most $k! p^{k(k-1)/2}$ many such k-tuples of residue classes and the proof follows from first upgrading all residue classes mod p^j in (3.3) to mod p^k (by paying a cost of $p^{k(k-1)/2}$) and then using the Newton-Girard identities which essentially uniquely determine the $x_1, \ldots, x_k$ (up to permutation). This bound is efficient since probabilistic heuristics suggest that we should expect $\approx (p^k)^k / p^{k(k+1)/2} = p^{k(k-1)/2}$ many solutions. Thus we have that

$$J_{s,k}(X, p, 0) \lesssim_k J_{s-k,k}(X/p) X^k p^{k(k-1)/2}. \tag{3.4}$$

3.4. **Step 4: Iteration.** Putting Steps 1 to 3 together we obtain the iteration that

$$J_{s,k}(X) \lesssim_k p^{2s-2k} J_{s-k,k}(X/p) X^k p^{k(k-1)/2}. \tag{3.5}$$

Running this iteration about $O(s/k)$ many steps reduces to an estimate on $J_{k,k}(X)$ from which one can easily compute there are $O(X^k)$ many solutions by the Newton-Girard identities. The iteration (3.5) is sharp if both s and $s-k$ are supercritical. If they are, then heuristically, we expect $J_{s,k}(X) \approx X^{2s-k(k+1)/2}$ and $J_{s-k,k}(X/p) \approx (X/p)^{2(s-k)-k(k+1)/2}$. Then the right hand side of (3.5) becomes $X^{2s} X^{-3k/2-k^2/2} p^{k^2}$ which is equal to $X^{2s-k(k+1)/2}$ since $p \sim X^{1/k}$. However, both sides are not the same if one of s or $s-k$ is subcritical. This is where the inefficiency of $X^{\frac{k^2}{2}(1-\frac{1}{k})^{s/k}}$ comes from.

3.5. **Interpreting Steps 1-4 into decoupling.** Having briefly summarized the number theoretic argument into four steps, we now briefly sketch the main points to interpret into decoupling. First we discuss the scales needed in the proof. From Steps 1 and 3, there are three scales: the largest scale X, the intermediate scale $p \sim X^{1/k}$, and the smallest scale 1. Correspondingly in our proof, we use three scales: the smallest scale δ, the intermediate scale $\nu := q^{\lfloor \log_q \delta^{1/k} \rfloor} \sim \delta^{1/k}$, and the largest scale 1. For some technical reasons surrounding the broad-narrow reduction, in lieu of the scale 1, we will actually use the scale $\kappa := q^{\lfloor \log_q \delta^\varepsilon \rfloor}$ where ε is as in (1.8).

Next, we discuss the reduction to the decoupling analogue of (3.1). In Step 1, two residue classes being distinct mod p means they are p-adically separated by a distance 1 and so this should correspond to two intervals which are 1-separated. To get around the use of the Prime Number Theorem, we make use instead of broad-narrow reduction due to Bourgain and Guth in [4] which will allow us to reduce to controlling a multilinear decoupling expression.

Third, the loss of p^{2s-2k} in Step 2 above deserves some mention. This loss comes from essentially having applied the union bound

$$|\sum_{1 \leqslant n \leqslant X} e(n\alpha_1 + \cdots + n^k\alpha_k)| = |\sum_{a \pmod p} \sum_{\substack{n \equiv a \pmod p \\ 1 \leqslant n \leqslant X}} e(n\alpha_1 + \cdots + n^k\alpha_k)|$$

$$\leqslant p \max_{a \pmod p} |\sum_{\substack{n \equiv a \pmod p \\ 1 \leqslant n \leqslant X}} e(n\alpha_1 + \cdots + n^k\alpha_k)|.$$

Heuristically we expect this inequality to be efficient since each $\sum_{n \equiv a \pmod p}$ contributes equally to the entire sum as the exponential sum should not bias one residue class mod p over another. This however is not necessarily true in the decoupling case and will require us to obtain some extra uniformity via dyadic pigeonholing, see Section 5.1, later.

Finally, to interpret the solution counting Step 3, we make use of the simple identity

$$\int_{\mathbb{Q}_q^k} f(x)\,dx = \widehat{f}(0)$$

which converts the integral of f into a question of whether 0 is contained in the support of $\widehat{f}$. This is done in Lemma 4.4 below and the proof relies on the Newton-Girard identities, much like in the proof of Linnik's Lemma. This part of the argument requires that p is even and is reminiscent of a Córdoba-Fefferman argument (see for example [8, Section 3.2] or [6, 7, 10]).

4. THE MAIN LEMMA

One standard property about the moment curve decoupling constant that we use is affine rescaling. This property plays the analogue of translation-dilation invariance of the Vinogradov system (1.1).

Lemma 4.1 (Affine rescaling). *Let g be a Schwartz function on $\mathbb{Q}_q^k$ Fourier supported in $\bigcup_{K \in P_\delta} \theta_K$. Then for any interval $I \subset \mathbb{Z}_q$ of length $\kappa \geqslant \delta$, we have*

$$\|g_I\|_{L^p(\mathbb{Q}_q^k)} \leqslant \mathfrak{D}_p(\frac{\delta}{\kappa})(\sum_{K \in P_\delta(I)} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2)^{1/2}.$$

Proof. This proof is standard and follows from a change of variables which can be found for example in [8, Section 11.2]. $\qquad\square$

Our main lemma in proving Theorem 1.1 is the following:

Lemma 4.2. *Let $p \in 2k + 2\mathbb{N}$, $\delta \in q^{-\mathbb{N}}$ and $\kappa \in q^{-\mathbb{N}} \cap [\delta, 1)$. Let $\nu = q^{\lfloor \log_q \delta^{1/k} \rfloor} \in q^{-\mathbb{N}}$ so that $\nu \leqslant \delta^{1/k}$. If g is a Schwartz function with Fourier support in $\bigcup_{K \in P_\delta} \theta_K$, then we have*

$$\int_{\mathbb{Q}_q^k} |g|^p \leqslant C\mathfrak{D}_p(\frac{\delta}{\kappa})^p(\sum_{K \in P_\delta} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2)^{p/2} + Cq^{-k(k-1)}\kappa^{-(k^2+4k-2)}\nu^{-k(k-1)/2}N^{p-2k} \times$$

$$\mathfrak{D}_{p-2k}(\frac{\delta}{\nu})^{p-2k} \max_{K \in P_\delta} \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}^k (\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)})^k \max_{J \in P_\nu}(\sum_{K' \in P_\delta(J)} \|g_{K'}\|_{L^{p-2k}(\mathbb{Q}_q^k)}^2)^{(p-2k)/2}$$

where N is the number of $J \in P_\nu$ for which $g_J \neq 0$ and C depends only on k and p.

Here κ is a somewhat technical parameter that is chosen to be roughly δ^ε later in Section 5. However, on a first reading, it might be more convenient for the reader to take $\kappa = 1/q$ to better grasp the moving parts of the argument. The somewhat non-standard decoupling right hand side in Lemma 4.2 is reminiscent of the right hand side used in Theorem 1.2 of [18]. To give more context to the above lemma, the following estimate is true:

Lemma 4.3. *For any $p > 2k$, we have*

$$\Big(\sum_{K \in P_\delta} \|g_K\|^2_{L^p(\mathbb{Q}_q^k)} \Big)^{p/2}$$

$$\leqslant N^{(p-2k)/2} \max_{K \in P_\delta} \|g_K\|^k_{L^\infty(\mathbb{Q}_q^k)} \Big(\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \Big)^k \max_{J \in P_\kappa} \Big(\sum_{K \in P_\delta(J)} \|g_K\|^2_{L^{p-2k}(\mathbb{Q}_q^k)} \Big)^{(p-2k)/2}$$

where N is as defined in Lemma 4.2.

Proof. Hölder's inequality gives us

$$\|g_K\|_{L^p(\mathbb{Q}_q^k)} \leqslant \|g_K\|^{\frac{2k}{p}}_{L^\infty(\mathbb{Q}_q^k)} \|g_K\|^{1-\frac{2k}{p}}_{L^{p-2k}(\mathbb{Q}_q^k)},$$

and so, applying

$$\Big(\sum_K a_K^{\frac{2k}{p}} b_K^{\frac{2k}{p}} c_K^{2(1-\frac{2k}{p})} \Big)^{\frac{p}{2}} \leqslant (\max_K a_K)^k \Big(\sum_K b_K \Big)^k \Big(\sum_K c_K^2 \Big)^{\frac{p-2k}{2}}$$

with $a_K = b_K = \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}$ and $c_K = \|g_K\|_{L^{p-2k}(\mathbb{Q}_q^k)}$, we get

$$\Big(\sum_{K \in P_\delta} \|g_K\|^2_{L^p(\mathbb{Q}_q^k)} \Big)^{p/2} \leqslant \max_{K \in P_\delta} \|g_K\|^k_{L^\infty(\mathbb{Q}_q^k)} \Big(\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \Big)^k \Big(\sum_{K \in P_\delta} \|g_K\|^2_{L^{p-2k}(\mathbb{Q}_q^k)} \Big)^{(p-2k)/2}.$$

It remains to observe that

$$\Big(\sum_{K \in P_\delta} \|g_K\|^2_{L^{p-2k}(\mathbb{Q}_q^k)} \Big)^{(p-2k)/2} \leqslant N^{(p-2k)/2} \max_{J \in P_\kappa} \Big(\sum_{K \in P_\delta(J)} \|g_K\|^2_{L^{p-2k}(\mathbb{Q}_q^k)} \Big)^{(p-2k)/2}.$$

$\square$

Suppose for a moment that in Lemma 4.3, we had an equality instead of an inequality. This is indeed the case when $g(x)$ is the exponential sum $X^{-100k^2} 1_{|x| \leqslant X^{100k}} \sum_{j=1}^X e(\gamma(j) \cdot x)$ that arises in using decoupling to estimate the number of solutions in (1.3). As $N \leqslant \nu^{-1}$ (and taking, for convenience, $\kappa = 1/q$), Lemma 4.2 would give us

$$\mathfrak{D}_p(\delta)^p \leqslant C \, \mathfrak{D}_p(q\delta)^p + C q^{5k-2} \nu^{-\frac{p}{2}+k-\frac{k(k-1)}{2}} \mathfrak{D}_{p-2k} \Big(\frac{\delta}{\nu} \Big)^{p-2k} \tag{4.1}$$

Heuristically, we expect this iteration to be efficient as long as $p - 2k$ (and so also p) is supercritical. To see this, if r is supercritical, then we heuristically expect that $\mathfrak{D}_r(\delta)^r \approx \delta^{-\frac{r}{2}+\frac{k(k+1)}{2}}$ for all δ. Thus the iteration should be efficient if with this assumption on the size of $\mathfrak{D}_r(\delta)^r$, both sides of (4.1) are the same. The right hand side of (4.1) is then

$$\sim_q \big(\delta^{-\frac{1}{k}} \big)^{\frac{p}{2}+k-\frac{k(k-1)}{2}} \big(\delta^{-1+\frac{1}{k}} \big)^{\frac{p-2k}{2}-\frac{k(k+1)}{2}} = \delta^{-(\frac{p}{2}-\frac{k(k+1)}{2})}$$

which is comparable to the left hand side of (4.1). A similar calculation shows that this iteration is not efficient if at least one of p or $p - 2k$ is subcritical.

Unfortunately the reverse inequality in Lemma 4.3 fails to hold for general g. This is because we lack the uniformity in the exponential sum that one considers when one counts solutions to the Vinogradov system. This uniformity can be restored by pigeonholing, which only produces $\delta^{-\varepsilon}$ losses. This pigeonholing must be done before one applies induction on scales and iterates on the Lebesgue exponent p. The full argument is carried out in detail in Section 5.

4.1. Proof of Lemma 4.2.

The proof of Lemma 4.2 uses a broad/narrow dichotomy, due to Bourgain and Guth [4] combined with some basic geometric geometric properties of the moment curve. See also for example [8, Chapter 7].

4.1.1. *The broad-narrow argument.* First, we have the pointwise bound $|g(x)| \leqslant \sum_{I \in P_\kappa} |g_I(x)|$. At every point $x \in \mathbb{Q}_q^k$, let $\mathcal{I}_x$ be the set of all intervals $I' \in P_\kappa$ such that $|g_{I'}(x)| \geqslant \kappa \max_{I \in P_\kappa} |g_I(x)|$. Suppose first $\mathcal{I}_x$ contains at least k (disjoint) intervals, say $I_1', \ldots, I_k'$ (all dependent on x) of length κ and $|g_{I_1'}(x)| = \max_{I \in P_\kappa} |g_I(x)|$: in this case we have

$$|g(x)| \leqslant \kappa^{-1} \max_{I \in P_\kappa} |g_I(x)| \leqslant \kappa^{-1} \kappa^{-(k-1)/k} |g_{I_1'}(x) \ldots g_{I_k'}(x)|^{1/k}$$

$$\leqslant \kappa^{-2+1/k} \max_{\substack{I_1, \ldots, I_k \in P_\kappa \\ d(I_i, I_j) > \kappa \, \forall i \neq j}} |g_{I_1}(x) \ldots g_{I_k}(x)|^{1/k}.$$

Here we used that in $\mathbb{Q}_q^k$, two distinct intervals of the same length are separated by at least that length. Alternatively, $\mathcal{I}_x$ contains at most $k - 1$ intervals, in which case

$$|g(x)| \leqslant \sum_{I \in \mathcal{I}_x} |g_I(x)| + \sum_{I \in P_\kappa \backslash \mathcal{I}_x} |g_I(x)| < (k-1) \max_{I \in P_\kappa} |g_I(x)| + \sum_{I \in P_\kappa \backslash \mathcal{I}_x} \kappa \max_{I \in P_\kappa} |g_I(x)| < k \max_{I \in P_\kappa} |g_I(x)|.$$

As a result, we obtain the pointwise bound that for each $x \in \mathbb{Q}_q^k$, we have

$$|g(x)| \leqslant k \max_{I \in P_\kappa} |g_I(x)| + \kappa^{-2+1/k} \max_{\substack{I_1, \ldots, I_k \in P_\kappa \\ d(I_i, I_j) > \kappa \, \forall i \neq j}} |g_{I_1}(x) \ldots g_{I_k}(x)|^{1/k}$$

which, upon raising both sides to power $2k$ and applying $(A + B)^{2k} \leqslant 2^{2k-1}(A^{2k} + B^{2k})$ (a consequence of the convexity of $x \mapsto x^{2k}$), yields

$$|g(x)|^{2k} \leqslant 2^{2k-1} k^{2k} \max_{I \in P_\kappa} |g_I(x)|^{2k} + 2^{2k-1} \kappa^{-(4k-2)} \max_{\substack{I_1, \ldots, I_k \in P_\kappa \\ d(I_i, I_j) > \kappa \, \forall i \neq j}} |g_{I_1}(x) \ldots g_{I_k}(x)|^2. \qquad (4.2)$$

Using this pointwise bound while integrating we find that

$$\int_{\mathbb{Q}_q^k} |g|^p = \int_{\mathbb{Q}_q^k} |g|^{2k} |g|^{p-2k}$$

$$\leqslant C \int_{\mathbb{Q}_q^k} (\max_{I \in P_\kappa} |g_I|^2)^k |g|^{p-2k} + C \kappa^{-(4k-2)} \int_{\mathbb{Q}_q^k} \max_{\substack{I_1, \ldots, I_k \in P_\kappa \\ d(I_i, I_j) > \kappa \, \forall i \neq j}} |g_{I_1} \ldots g_{I_k}|^2 |g|^{p-2k}$$

$$\leqslant C \int_{\mathbb{Q}_q^k} (\sum_{I \in P_\kappa} |g_I|^2)^k |g|^{p-2k} + C \kappa^{-(4k-2)} \sum_{\substack{I_1, \ldots, I_k \in P_\kappa \\ d(I_i, I_j) > \kappa \, \forall i \neq j}} \int_{\mathbb{Q}_q^k} |g_{I_1} \ldots g_{I_k}|^2 |g|^{p-2k}$$

$$\leqslant C \int_{\mathbb{Q}_q^k} (\sum_{I \in P_\kappa} |g_I|^2)^k |g|^{p-2k} + C \kappa^{-(4k-2)} \kappa^{-k} \max_{\substack{I_1, \ldots, I_k \in P_\kappa \\ d(I_i, I_j) > \kappa \, \forall i \neq j}} \int_{\mathbb{Q}_q^k} |g_{I_1} \ldots g_{I_k}|^2 |g|^{p-2k}$$

for some C depending on k. Hölder's inequality followed by Minkowski's inequality implies that the first term satisfies

$$C \int_{\mathbb{Q}_q^k} (\sum_{I \in P_\kappa} |g_I|^2)^k |g|^{p-2k} \leqslant C \Big(\int_{\mathbb{Q}_q^k} (\sum_{I \in P_\kappa} |g_I|^2)^{p/2} \Big)^{2k/p} \Big(\int_{\mathbb{Q}_q^k} |g|^p \Big)^{(p-2k)/p}$$

$$\leqslant C \Big(\sum_{I \in P_\kappa} \|g_I\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^k \Big(\int_{\mathbb{Q}_q^k} |g|^p \Big)^{(p-2k)/p}$$

$$\leqslant \frac{1}{2} \int_{\mathbb{Q}_q^k} |g|^p + C' \Big(\sum_{I \in P_\kappa} \|g_I\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2}$$

for some C' that depends on k and p. The last inequality uses Young's inequality and the fact that $p \geqslant 2k$. Therefore,

$$\int_{\mathbb{Q}_q^k} |g|^p \lesssim_p \Big(\sum_{I \in P_\kappa} \|g_I\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2} + \kappa^{-(5k-2)} \max_{\substack{I_1,\dots,I_k \in P_\kappa \\ d(I_i,I_j) > \kappa \, \forall i \neq j}} \int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g|^{p-2k}.$$

Using affine rescaling (Lemma 4.1) and applying the definition (1.5) of our decoupling constant, we deduce that

$$\Big(\sum_{I \in P_\kappa} \|g_I\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2} \leqslant \mathfrak{D}_p(\frac{\delta}{\kappa})^p \Big(\sum_{K \in P_\delta} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2}.$$

Plugging this into the above yields

$$\int_{\mathbb{Q}_q^k} |g|^p \lesssim_p \mathfrak{D}_p(\frac{\delta}{\kappa})^p \Big(\sum_{K \in P_\delta} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2} + \kappa^{-(5k-2)} \max_{\substack{I_1,\dots,I_k \in P_\kappa \\ d(I_i,I_j) > \kappa \, \forall i \neq j}} \int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g|^{p-2k}. \qquad (4.3)$$

This inequality (4.3) is the analogue of Step 1 in Section 3.1. The requirement that we analyze solutions to the Vinogradov system with $x_1, \dots, x_s$ and $y_1, \dots, y_s$ being distinct mod p corresponds to the requirement that we analyze $\int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g|^{p-2k}$ with $d(I_i, I_j) > \kappa$ for all $1 \leqslant i \neq j \leqslant k$ with $\kappa = 1/q$.

Next, we mimic Step 2 in Section 3.2. Recalling our definition of N in the statement of Lemma 4.2, Hölder's inequality gives

$$|g|^{p-2k} \leqslant N^{p-2k-1} \sum_{J \in P_\nu} |g_J|^{p-2k}.$$

Applying this in the second term in (4.3), we get

$$\int_{\mathbb{Q}_q^k} |g|^p$$
$$\lesssim_p \mathfrak{D}_p(\frac{\delta}{\kappa})^p \Big(\sum_{K \in P_\delta} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2} + \kappa^{-(5k-2)} N^{p-2k-1} \max_{\substack{I_1,\dots,I_k \in P_\kappa \\ d(I_i,I_j) > \kappa \, \forall i \neq j}} \sum_{J \in P_\nu} \int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g_J|^{p-2k}$$
$$\lesssim_p \mathfrak{D}_p(\frac{\delta}{\kappa})^p \Big(\sum_{K \in P_\delta} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2} + \kappa^{-(5k-2)} N^{p-2k} \max_{\substack{I_1,\dots,I_k \in P_\kappa \\ d(I_i,I_j) > \kappa \, \forall i \neq j}} \max_{J \in P_\nu} \int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g_J|^{p-2k} \quad (4.4)$$

which is the analogue of Step 2 in Section 3.2.

To analyze the second term in (4.4), we fix $I_1, \dots, I_k \in P_\kappa$ with $d(I_i, I_j) > \kappa$ for all $i \neq j$, and fix $J \in P_\nu$ with $g_J \neq 0$. To estimate the integral $\int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g_J|^{p-2k}$, first note that the Fourier transform of $|g_J|^2 = g_J \overline{g_J}$ is supported in the parallelepiped $\theta_J - \theta_J$, of dimension $\nu \times \nu^2 \times \cdots \times \nu^k$. Since our hypothesis guarantees that $p - 2k$ is an even positive integer, the same is true for the Fourier transform of $|g_J|^{p-2k}$. Lemma 2.1(i) applied to $J \in P_\nu$ instead of $K \in P_\delta$ shows that the Fourier support of $|g_J|^{p-2k}$ is the disjoint union of $\nu^{-k(k-1)/2}$ many cubes of side lengths ν^k, and we denote this collection of cubes by $\{\square\}$. This corresponds to the fact that we have a k-tuple of residue classes $(H_1 \ (\mathrm{mod}\ p), H_2 \ (\mathrm{mod}\ p^2), \dots, H_k \ (\mathrm{mod}\ p^k))$ which we can upgrade to $p^{k(k-1)/2}$ many k-tuples of the form $(H_1' \ (\mathrm{mod}\ p^k), H_2' \ (\mathrm{mod}\ p^k), \dots, H_k' \ (\mathrm{mod}\ p^k))$. Note that the side length ν^k of the cubes $\square$ is $\leqslant \delta$.

We now apply Fourier inversion and turn products into convolutions. We have

$$
\int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g_J|^{p-2k} = \sum_{\substack{K_i \in P_\delta(I_i) \\ i=1,\ldots,k}} \sum_{\substack{\bar{K}_j \in P_\delta(I_j) \\ j=1,\ldots,k}} \int_{\mathbb{Q}_q^k} g_{K_1} \cdots g_{K_k} \overline{g_{\bar{K}_1}} \cdots \overline{g_{\bar{K}_k}} |g_J|^{p-2k}
$$

$$
= \sum_{\square} \sum_{\substack{K_i \in P_\delta(I_i) \\ i=1,\ldots,k}} \sum_{\substack{\bar{K}_j \in P_\delta(I_j) \\ j=1,\ldots,k}} \widehat{g_{K_1}} * \cdots * \widehat{g_{K_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}} * (\widehat{|g_J|^{p-2k} 1_\square})(0).
$$

For each fixed $\square$ and $\bar{K}_1 \in P_\delta(I_1), \ldots, \bar{K}_k \in P_\delta(I_k)$, let $S(\bar{K}_1, \ldots, \bar{K}_k, \square)$ be the set of all $(K_1, \ldots, K_k)$ with $K_i \in P_\delta(I_i)$ such that

$$
0 \in \operatorname{supp}(\widehat{g_{K_1}} * \cdots * \widehat{g_{K_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}} * (\widehat{|g_J|^{p-2k} 1_\square})). \tag{4.5}
$$

We will prove in Lemma 4.4 below that $\#S(\bar{K}_1, \ldots, \bar{K}_k, \square) \leqslant (q\kappa)^{-k(k-1)}$. If we think of the model case when $\kappa = 1/q$, this would say that the $\bar{K}_i$ and $\square$ uniquely determine the K_i in (4.5). This analogous to the situation in Linnik's lemma where once we upgrade (3.3) to residue classes mod p^k, the remaining variables are essentially uniquely determined.

We now write

$$
\int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g_J|^{p-2k}
$$

$$
= \Bigg| \sum_{\square} \sum_{\substack{\bar{K}_j \in P_\delta(I_j) \\ j=1,\ldots,k}} \sum_{\substack{(K_1,\ldots,K_k) \in \\ S(\bar{K}_1,\ldots,\bar{K}_k,\square)}} \widehat{g_{K_1}} * \cdots * \widehat{g_{K_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}} * (\widehat{|g_J|^{p-2k} 1_\square})(0) \Bigg|
$$

$$
\leqslant \sum_{\square} \sum_{\substack{\bar{K}_j \in P_\delta(I_j) \\ j=1,\ldots,k}} \sum_{\substack{(K_1,\ldots,K_k) \in \\ S(\bar{K}_1,\ldots,\bar{K}_k,\square)}} \int_{\mathbb{Q}_q^k} |g_{K_1} \cdots g_{K_k} g_{\bar{K}_1} \cdots g_{\bar{K}_k}| |g_J|^{p-2k} * |\widecheck{1_\square}|
$$

$$
\leqslant \sum_{\square} \Big(\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \Big)^k (q\kappa)^{-k(k-1)} \max_{K \in P_\delta} \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}^k \int_{\mathbb{Q}_q^k} |g_J|^{p-2k} * |\widecheck{1_\square}|.
$$

Since $\int_{\mathbb{Q}_q^k} |\widecheck{1_\square}| = 1$ and the number of $\square$ is $\nu^{-k(k-1)/2}$, this gives

$$
\int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 |g_J|^{p-2k} \leqslant \nu^{-\frac{k(k-1)}{2}} (q\kappa)^{-k(k-1)} \max_{K \in P_\delta} \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}^k \Big(\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \Big)^k \int_{\mathbb{Q}_q^k} |g_J|^{p-2k}.
$$

Applying affine rescaling shows that this is

$$
\leqslant \nu^{-\frac{k(k-1)}{2}} (q\kappa)^{-k(k-1)} \max_{K \in P_\delta} \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}^k \times
$$

$$
\Big(\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \Big)^k \mathfrak{D}_{p-2k} \Big(\frac{\delta}{\nu}\Big)^{p-2k} \Big(\sum_{K' \in P_\delta(J)} \|g_{K'}\|_{L^{p-2k}(\mathbb{Q}_q^k)}^2 \Big)^{\frac{p-2k}{2}}. \tag{4.6}
$$

One can think of (4.6) as the analogue of (3.4) in Section 3.3 in the following way: the term $\nu^{-k(k-1)/2}(q\kappa)^{-k(k-1)} \max_{K \in P_\delta} \|g_K\|_\infty^k$ plays the role of $p^{k(k-1)/2}$ from Linnik's lemma, the term $(\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_\infty)^k$ plays the role of X^k, and finally the term $\mathfrak{D}_{p-2k}(\frac{\delta}{\nu})^{p-2k}(\sum_{K' \in P_\delta(J)} \|g_{K'}\|_{p-2k}^2)^{\frac{p-2k}{2}}$ plays the role of the $J_{s-k,k}(X/p)$.

Plugging (4.6) back to (4.4), we then obtain

$$\int_{\mathbb{Q}_q^k} |g|^p \lesssim_p \mathfrak{D}_p(\frac{\delta}{\kappa})^p \Big(\sum_{K \in P_\delta} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2} + q^{-k(k-1)} \kappa^{-(k^2+4k-2)} \nu^{-\frac{k(k-1)}{2}} N^{p-2k} \times$$

$$\mathfrak{D}_{p-2k}(\frac{\delta}{\nu})^{p-2k} \max_{K \in P_\delta} \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}^k \Big(\sum_{\bar{K} \in P_\delta} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \Big)^k \Big(\sum_{K' \in P_\delta(J)} \|g_{K'}\|_{L^{p-2k}(\mathbb{Q}_q^k)}^2 \Big)^{\frac{p-2k}{2}}.$$

4.1.2. *Geometry of the moment curve.* The proof of Lemma 4.2 is now complete modulo the proof of the following lemma, which provides the key geometric input that enables one to count $\#S(\bar{K}_1, \ldots, \bar{K}_k, \square)$. This is the analogue of Linnik's Lemma ([29, Corollary 17] and the estimate for $\mathbf{B}(\mathbf{g})$ in the proof of [31, Lemma 5.1]); see also [13, Proposition 1.3] and [2, Proposition 3.1]. Both proofs use the Newton-Girard identities in essentially the same way. The hypothesis that $q > k$, where q is the characteristic of our base field $\mathbb{Q}_q$ and k is the degree of the moment curve, plays a role in the following lemma.

Lemma 4.4. *Let* $p \in 2k + 2\mathbb{N}$, $\delta \in q^{-\mathbb{N}}$, $\kappa \in q^{-\mathbb{N}} \cap [\delta, 1)$, *and* $\nu = q^{\lceil \log_q \delta^{1/k} \rceil} \in q^{-\mathbb{N}}$ *so that* $\nu \leq \delta^{1/k}$. *Suppose that* $I_1, \ldots, I_k \in P_\kappa$ *with* $d(I_i, I_j) > \kappa$ *for all* $i \neq j$. *Let* $\square$ *be a cube of side length* ν^k *and* $\bar{K}_1 \in P_\delta(I_1), \ldots, \bar{K}_k \in P_\delta(I_k)$. *Define* $S(\bar{K}_1, \ldots, \bar{K}_k, \square)$ *be the set of all ordered* k-tuples $(K_1, \ldots, K_k)$ *with* $K_i \in P_\delta(I_i)$ *such that*

$$0 \in \mathrm{supp}(\widehat{g_{K_1}} * \cdots * \widehat{g_{K_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}} * (\widehat{|g_J|^{p-2k} 1_\square})).$$

Then

$$\#S(\bar{K}_1, \ldots, \bar{K}_k, \square) \leq (q\kappa)^{-k(k-1)}.$$

Proof. Assume for the sake of contradiction that $\#S(\bar{K}_1, \ldots, \bar{K}_k, \square) > (q\kappa)^{-k(k-1)} \geq 1$. We can find two k-tuples of intervals $(A_1, \ldots, A_k)$ and $(B_1, \ldots, B_k)$ with each $A_i, B_i \in P_\delta(I_i)$ such that

$$0 \in \mathrm{supp}(\widehat{g_{A_1}} * \cdots * \widehat{g_{A_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}} * (\widehat{|g_J|^{p-2k} 1_\square})), \tag{4.7}$$

$$0 \in \mathrm{supp}(\widehat{g_{B_1}} * \cdots * \widehat{g_{B_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}} * (\widehat{|g_J|^{p-2k} 1_\square})), \tag{4.8}$$

and such that there exists an i_0 with $d(A_{i_0}, B_{i_0}) > (q\kappa)^{-(k-1)}\delta$. Indeed, if not, picking an arbitrary $(C_1, \ldots, C_k) \in S(\bar{K}_1, \ldots, \bar{K}_k, \square)$, shows that any other $(D_1, \ldots, D_k) \in S(\bar{K}_1, \ldots, \bar{K}_k, \square)$ must satisfy $d(C_i, D_i) \leq (q\kappa)^{-(k-1)}\delta$. This gives at most $(q\kappa)^{-k(k-1)}$ many k-tuples which violates our initial assumption that $\#S(\bar{K}_1, \ldots, \bar{K}_k, \square) > (q\kappa)^{-k(k-1)}$. Without loss of generality, we may assume that $i_0 = 1$.

Since for each $i = 1, 2, \ldots, k$, we have $A_i, B_i \subset I_i$ and $d(I_i, I_j) > \kappa$ for all $i \neq j$, this implies

$$d(A_i, A_j) \geqslant q\kappa, \quad d(B_i, B_j) \geqslant q\kappa, \quad d(A_i, B_j) \geqslant q\kappa \quad \text{whenever } j \neq i \tag{4.9}$$

(thus the only distances we do not have any control over are the ones of the form $d(A_i, B_i)$, $i \neq 1$). By (4.7) and (4.8), we have that

$$0 \in \Big(\sum_{i=1}^k \tau_{A_i} - \sum_{i=1}^k \tau_{\bar{K}_i} + \square\Big) \cap \Big(\sum_{i=1}^k \tau_{B_i} - \sum_{i=1}^k \tau_{\bar{K}_i} + \square\Big)$$

where here we recall the definition of τ_K in (1.4). Each τ_{A_i}, τ_{B_i}, and $\tau_{\bar{K}_i}$ are cubes in $\mathbb{Q}_q^k$ of side length δ and $\square$ is a cube in $\mathbb{Q}_q^k$ of side length $\nu^k \leqslant \delta$. Thus by the ultrametric inequality, both $\sum_{i=1}^k \tau_{A_i} - \sum_{i=1}^k \tau_{\bar{K}_i} + \square$ and $\sum_{i=1}^k \tau_{B_i} - \sum_{i=1}^k \tau_{\bar{K}_i} + \square$ are cubes in $\mathbb{Q}_q^k$ of side length δ. Furthermore, by the ultrametric inequality, since two cubes of side length δ are either completely

disjoint or exactly the same, we must have

$$\sum_{i=1}^{k}\tau_{A_i} - \sum_{i=1}^{k}\tau_{\bar{K}_i} + \square = \sum_{i=1}^{k}\tau_{B_i} - \sum_{i=1}^{k}\tau_{\bar{K}_i} + \square$$

and hence

$$\sum_{i=1}^{k}\tau_{A_i} - \sum_{i=1}^{k}\tau_{B_i} = B(0,\delta).$$

Therefore (after another application of the ultrametric inequality) there exists $\xi_{A_i} \in A_i$ and $\xi_{B_i} \in B_i$ such that

$$|\sum_{i=1}^{k}\xi_{A_i}^{j} - \sum_{i=1}^{k}\xi_{B_i}^{j}| \leqslant \delta \tag{4.10}$$

for $j = 1, 2, \ldots, k$.

We now use the Newton-Girard identities to derive a contradiction. For $j = 1, 2, \ldots, k$, define the power sums $p_j(x_1,\ldots,x_k) := x_1^j + \cdots + x_k^j$. Next for $j = 1, 2, \ldots, k$, define the elementary symmetric polynomials $e_j(x_1,\ldots,x_k) := \sum_{1 \leqslant i_1 < \cdots < i_j \leqslant k} x_{i_1} \cdots x_{i_j}$. Additionally, let $e_0(x_1,\ldots,x_k) := 1$. Then we have the two identities:

$$(X - x_1)(X - x_2)\cdots(X - x_k) = \sum_{j=0}^{k}(-1)^j e_j(x_1,\ldots,x_k)X^{k-j} \tag{4.11}$$

and for $j = 1, 2, \ldots, k$, we have

$$je_j(x_1,\ldots,x_k) = \sum_{i=0}^{j-1}(-1)^i e_{j-i-1}(x_1,\ldots,x_k)p_{i+1}(x_1,\ldots,x_k). \tag{4.12}$$

See, for example, [29, Lemma 15] for a proof.

Let $e_j(A) := e_j(\xi_{A_1},\ldots,\xi_{A_k})$ and $p_j(A) := p_j(\xi_{A_1},\ldots,\xi_{A_k})$. Similarly define $e_j(B)$ and $p_j(B)$. By (4.11), we then have

$$(\xi_{A_1} - \xi_{B_1})\cdots(\xi_{A_1} - \xi_{B_k}) = \sum_{j=0}^{k}(-1)^j e_j(B)\xi_{A_1}^{k-j} \tag{4.13}$$

and

$$0 = (\xi_{A_1} - \xi_{A_1})\cdots(\xi_{A_1} - \xi_{A_k}) = \sum_{j=0}^{k}(-1)^j e_j(A)\xi_{A_1}^{k-j}. \tag{4.14}$$

Subtracting (4.14) from (4.13) and using that $|\xi_{A_1} - \xi_{B_j}| \geqslant q\kappa$ for any $j \neq 1$ (which follows from (4.9)) shows that

$$(q\kappa)^{k-1}|\xi_{A_1} - \xi_{B_1}| \leqslant |\sum_{j=0}^{k}(-1)^j(e_j(B) - e_j(A))\xi_{A_1}^{k-j}| \leqslant \max_j |e_j(B) - e_j(A)|. \tag{4.15}$$

Next we claim that $|e_j(B) - e_j(A)| \leqslant \delta$ for all $j = 1, 2, \ldots, k$. We prove this by induction. Since $e_1 = p_1$, $|e_1(B) - e_1(A)| \leqslant \delta$ by the $j = 1$ case of (4.10). Now assume that for some $J = 1, 2, \ldots, k-1$ we had $|e_j(B) - e_j(A)| \leqslant \delta$ for all $j = 1, 2, \ldots, J$. Then by (4.12),

$$|(J+1)e_{J+1}(B) - (J+1)e_{J+1}(A)| = |\sum_{i=0}^{J}(-1)^i(e_{J-i}(B)p_{i+1}(B) - e_{J-i}(A)p_{i+1}(A))|$$

$$\leqslant \max_{0 \leqslant i \leqslant J}|e_{J-i}(B)p_{i+1}(B) - e_{J-i}(A)p_{i+1}(A)|.$$

Observe that

$$|e_{J-i}(B)p_{i+1}(B) - e_{J-i}(A)p_{i+1}(A)| = |e_{J-i}(B)(p_{i+1}(B) - p_{i+1}(A)) + p_{i+1}(A)(e_{J-i}(B) - e_{J-i}(A))|$$
$$\leqslant \max(|p_{i+1}(B) - p_{i+1}(A)|, |e_{J-i}(B) - e_{J-i}(A)|) \leqslant \delta$$

by the inductive hypothesis and (4.10). Since q is a prime $> k$, it follows that $|e_{J+1}(B) - e_{J+1}(A)| \leqslant \delta$.

Applying this conclusion to (4.15) then yields that $(q\kappa)^{k-1}|\xi_{A_1} - \xi_{B_1}| \leqslant \delta$. But this contradicts the fact that $d(A_1, B_1) > (q\kappa)^{-(k-1)}\delta$. Therefore we must have $\#S(\bar{K}_1, \ldots, \bar{K}_k, \square) \leqslant (q\kappa)^{-k(k-1)}$ which completes the proof of the lemma. $\qquad\square$

5. Proof of Theorem 1.1 and Corollary 1.2

5.1. Dyadic pigeonholing. It is more convenient to bound

$$D_p(\delta) := \sup_{\delta_0 \in q^{-\mathbb{N}} \cap [\delta,1]} \mathfrak{D}_p(\delta_0) \tag{5.1}$$

instead of $\mathfrak{D}_p(\delta)$ as $D_p(\delta)$ is defined for all real $\delta \in (0,1]$ (rather than just for $\delta \in q^{-\mathbb{N}}$) and is monotonic, that is, $D_p(\delta_L) \leqslant D_p(\delta_S)$ if $\delta_L \geqslant \delta_S$.

Proposition 5.1. *For even integers $p > 2k$, there exists a constant $C > 0$, depending only on k and p, such that for every $0 < \varepsilon < 1$, we have*

$$D_p(\delta)^p \leqslant C(\log \delta^{-1})^{3p}\left[D_p(\delta^{1-\varepsilon})^p + q^{\frac{p}{2} + \frac{k^2+7k-4}{2}}\delta^{-(k^2+4k-2)\varepsilon}\delta^{-\frac{1}{k}(\frac{p}{2} + \frac{k(k-3)}{2})}D_{p-2k}(\delta^{1-\frac{1}{k}})^{p-2k}\right] \tag{5.2}$$

for all $0 < \delta < 1$.

Proof. To bound $D_p(\delta)^p$, suppose $0 < \delta < 1$ and $\delta_0 \in q^{\mathbb{Z}}$ with $\delta_0 \in [\delta, 1]$. We need to bound $\mathfrak{D}_p(\delta_0)^p$ by decoupling down to frequency scale δ_0.

Let f be a Schwartz function on $\mathbb{Q}_q^k$ with Fourier support in $\bigcup_{K \in P_{\delta_0}} \theta_K$. Then $f = \sum_{K \in P_{\delta_0}} f_K$ where $\widehat{f_K} := \widehat{f} 1_{K \times \mathbb{Q}_q^{k-1}}$. We want to prove the existence of $C > 0$ so that for any $0 < \varepsilon < 1$,

$$\int_{\mathbb{Q}_q^k} |f|^p \leqslant C(\log \delta^{-1})^{3p} \times$$

$$\left[D_p(\delta^{1-\varepsilon})^p + q^{\frac{p}{2} + \frac{k^2+7k-4}{2}}\delta^{-(k^2+4k-2)\varepsilon}\delta^{-\frac{1}{k}(\frac{p}{2} + \frac{k(k-3)}{2})}D_{p-2k}(\delta^{1-\frac{1}{k}})^{p-2k}\right]\left(\sum_{K \in P_{\delta_0}} \|f_K\|_{L^p(\mathbb{Q}_q^k)}^2\right)^{p/2}. \tag{5.3}$$

In fact, we will prove that for any translate Q of $B_{\delta_0^{-k}} := \{x \in \mathbb{Q}_q^k : |x| \leqslant \delta_0^{-k}\}$, we have

$$\int_Q |f|^p \leqslant C(\log \delta^{-1})^{3p} \times$$

$$\left[D_p(\delta^{1-\varepsilon})^p + q^{\frac{p}{2} + \frac{k^2+7k-4}{2}}\delta^{-(k^2+4k-2)\varepsilon}\delta^{-\frac{1}{k}(\frac{p}{2} + \frac{k(k-3)}{2})}D_{p-2k}(\delta^{1-\frac{1}{k}})^{p-2k}\right]\left(\sum_{K \in P_{\delta_0}} \|f_K\|_{L^p(Q)}^2\right)^{p/2}. \tag{5.4}$$

The estimate (5.3) then follows by summing over all such Q's that tile $\mathbb{Q}_q^k$, and applying Minkowski's inequality to bring an $\ell^{p/2}$ norm over Q on the right hand side into the sum over $K \in P_{\delta_0}$.

Thus we now turn to the proof of (5.4). Note that for any translate Q of $B_{\delta_0^{-k}}$, we have that $\widehat{1_Q}$ is supported in $B_{\delta_0^k}$. Therefore $f1_Q$ is still Fourier supported in $\bigcup_{K \in P_{\delta_0}} \theta_K$ since $\theta_K + B_{\delta_0^k} = \theta_K$ for all $K \in P_{\delta_0}$. Next, we have $(f1_Q)_K = f_K 1_Q$; indeed

$$\widehat{(f1_Q)_K} = \widehat{f1_Q} 1_{K \times \mathbb{Q}_q^{k-1}} = (\widehat{f} * \widehat{1_Q}) 1_{K \times \mathbb{Q}_q^{k-1}} = (\widehat{f} 1_{K \times \mathbb{Q}_q^{k-1}}) * \widehat{1_Q} = \widehat{f_K} * \widehat{1_Q}.$$

As a result, to prove (5.4), it suffices to prove (5.3) under the additional assumption that f is supported on Q. Since f is an arbitrary Schwartz function with Fourier support in $\bigcup_{K \in P_{\delta_0}} \theta_K$, we may assume $Q = B_{\delta_0^{-k}}$. Thus from now on, we assume additionally that f and all the f_K are supported on $B_{\delta_0^{-k}}$ and prove (5.3).

We first dyadically pigeonhole f by wavepacket height. Write $H^* = \max_{K \in P_{\delta_0}} \|f_K\|_{L^\infty(\mathbb{Q}_q^k)}$. For $K \in P_{\delta_0}$ and $H \in 2^{\mathbb{Z}} H^* \cap (\delta_0^{1+\frac{k(k-1)}{2p}} H^*, H^*]$, let

$$f_K^{(H)} = f_K 1_{H/2 < |f_K| \leqslant H}$$

where here $f_K : \mathbb{Q}_q^k \to \mathbb{C}$, $|f_K|$ is the absolute value of f_K, and the last characteristic function is meant to be the indicator function of the set $\{x \in \mathbb{Q}_q^k : H/2 < |f_K(x)| \leqslant H\}$. Since f_K is supported on $B_{\delta_0^{-k}}$, so is $f_K^{(H)}$. By Lemma 2.2, since f_K is Fourier supported in θ_K, we then have

$$f_K^{(H)} = (\sum_{T \in \mathbb{T}(K)} f_K 1_T) 1_{H/2 < |f_K| \leqslant H} = \sum_{T \in \mathbb{T}(K)} (f_K 1_T) 1_{H/2 < |f_K 1_T| \leqslant H} \tag{5.5}$$

where the last equality is because $|f_K 1_T|$ constant on every $T \in \mathbb{T}(K)$. Again by Lemma 2.2, note that $f_K^{(H)}$ is Fourier supported in θ_K. Using the terminology of Lemma 2.2, the nonzero wavepackets that make up $f_K^{(H)}$ are all of height $\sim H$. Then

$$\|f - \sum_{K \in P_{\delta_0}} \sum_{H \in 2^{\mathbb{Z}} H^* \cap (\delta_0^{1+\frac{k(k-1)}{2p}} H^*, H^*]} f_K^{(H)}\|_{L^\infty(\mathbb{Q}_q^k)} \leqslant \sum_{K \in P_{\delta_0}} \|f_K 1_{|f_K| \leqslant \delta_0^{1+\frac{k(k-1)}{2p}} H^*}\|_{L^\infty(\mathbb{Q}_q^k)}$$

$$\leqslant \delta_0^{-1}(\delta_0^{1+\frac{k(k-1)}{2p}} H^*) = \delta_0^{\frac{k(k-1)}{2p}} H^*$$

so since f and $f_K^{(H)}$ are supported on $B_{\delta_0^{-k}}$,

$$\|f - \sum_{K \in P_{\delta_0}} \sum_{H \in 2^{\mathbb{Z}} H^* \cap (\delta_0^{1+\frac{k(k-1)}{2p}} H^*, H^*]} f_K^{(H)}\|_{L^p(\mathbb{Q}_q^k)} \leqslant (\delta_0^{\frac{k(k-1)}{2p}} H^*)|B_{\delta_0^{-k}}|^{\frac{1}{p}} = H^* \delta_0^{-\frac{k(k+1)}{2p}}$$

$$\leqslant \max_{K \in P_{\delta_0}} \|f_K\|_{L^p(\mathbb{Q}_q^k)} \leqslant (\sum_{K \in P_{\delta_0}} \|f_K\|_{L^p(\mathbb{Q}_q^k)}^2)^{1/2}$$

where the second inequality follows from writing $f_K = f_K * \check{1}_{\theta_K}$ and applying Young's inequality $\|f_K\|_{L^\infty(\mathbb{Q}_q^k)} \leqslant \|f_K\|_{L^p(\mathbb{Q}_q^k)} \|\check{1}_{\theta_K}\|_{L^{p'}(\mathbb{Q}_q^k)}$. This shows

$$\|f\|_{L^p(\mathbb{Q}_q^k)} \leqslant \sum_{H \in 2^{\mathbb{Z}} H^* \cap (\delta_0^{1+\frac{k(k-1)}{2p}} H^*, H^*]} \|\sum_{K \in P_{\delta_0}} f_K^{(H)}\|_{L^p(\mathbb{Q}_q^k)} + (\sum_{K \in P_{\delta_0}} \|f_K\|_{L^p(\mathbb{Q}_q^k)}^2)^{1/2}.$$

Next we dyadically pigeonhole so that each relevant $f_K^{(H)}$ is made up of about the same number of wavepackets. Let now $\nu = q^{\lfloor \log_q \delta_0^{1/k} \rfloor} \leqslant \delta_0^{1/k}$. From (5.5), $f_K^{(H)}$ is Fourier supported in θ_K and supported in $B_{\delta_0^{-k}}$. Since a $T \in \mathbb{T}(K)$ is either completely contained in or completely disjoint from $B_{\delta_0^{-k}}$, we then can write

$$f_K^{(H)} = \sum_{T \in \mathbb{T}(K), T \subset B_{\delta_0^{-k}}} (f_K 1_T) 1_{H/2 < |f_K 1_T| \leqslant H}. \tag{5.6}$$

Furthermore, the $T \in \mathbb{T}(K)$ which are contained in $B_{\delta_0^{-k}}$ perfectly partition $B_{\delta_0^{-k}}$ into $\delta_0^{-k(k-1)/2}$ many translates of $T_{0,k}$. Thus (5.6) has at most $\delta_0^{-k(k-1)/2}$ many nonzero terms. Therefore for $\alpha \in 2^{\mathbb{N}} \cap [1, \delta_0^{-k(k-1)/2}]$, let

$$f_K^{(H,\alpha)} := f_K^{(H)}$$

if the number of nonzero terms in (5.6) (that is, the number of nonzero wavepackets in $f_K^{(H)}$) is in $(\alpha/2, \alpha]$, and 0 otherwise. Thus now we have that

$$f_K^{(H)} = \sum_{\alpha \in 2^{\mathbb{N}} \cap [1, \delta_0^{-k(k-1)/2}]} f_K^{(H,\alpha)} \tag{5.7}$$

and each $f_K^{(H,\alpha)}$ is a function which is supported in $B_{\delta_0^{-k}}$ and Fourier supported in θ_K which has $\sim \alpha$ many nonzero wavepackets of height $\sim H$.

Finally, we dyadically pigeonhole so that given a K, the parent interval J of length ν has about the same number of children K' of length δ_0 such that $f_{K'}^{(H,\alpha)} \neq 0$. To be more precise, fix a K and let J be the unique parent interval of length ν containing K. This parent J contains ν/δ_0 many intervals K' of length δ_0 and hence J has at most ν/δ_0 many children K' such that $f_{K'}^{(H,\alpha)} \neq 0$. For $K \subset J$ and $\beta \in 2^{\mathbb{N}} \cap [1, \nu/\delta_0]$, let

$$f_K^{(H,\alpha,\beta)} := f_K^{(H,\alpha)}$$

if the number of children K'' of J with $f_{K''}^{(H,\alpha)} \neq 0$ is in $(\beta/2, \beta]$, that is, if $\#\{K'' \in P_{\delta_0}(J) : f_{K''}^{(H,\alpha)} \neq 0\} \in (\beta/2, \beta]$, and 0 otherwise. Thus we now have

$$f_K^{(H,\alpha)} = \sum_{\beta \in 2^{\mathbb{N}} \cap [1, \nu/\delta_0]} f_K^{(H,\alpha,\beta)} \tag{5.8}$$

and each $f_K^{(H,\alpha,\beta)}$ is a function which is supported in $B_{\delta_0^{-k}}$, Fourier supported in θ_K, has $\sim \alpha$ many nonzero wavepackets of height $\sim H$, and K's parent J has $\sim \beta$ children each of which also are supported in $B_{\delta_0^{-k}}$, Fourier supported in θ_K, and have $\sim \alpha$ many nonzero wavepackets of height $\sim H$.

Thus combining (5.7) and (5.8) gives

$$f_K^{(H)} = \sum_{\alpha \in 2^{\mathbb{N}} \cap [1, \delta_0^{-k(k-1)/2}]} \sum_{\beta \in 2^{\mathbb{N}} \cap [1, \nu/\delta_0]} f_K^{(H,\alpha,\beta)}$$

which implies

$$\|f\|_{L^p(\mathbb{Q}_q^k)} \lesssim \sum_{H \in 2^{\mathbb{Z}} H^* \cap (\delta_0^{1 + \frac{k(k-1)}{2p}} H^*, H^*]} \sum_{\alpha \in 2^{\mathbb{N}} \cap [1, \delta_0^{-k(k-1)/2}]} \sum_{\beta \in 2^{\mathbb{N}} \cap [1, \nu/\delta_0]} \| \sum_{K \in P_{\delta_0}} f_K^{(H,\alpha,\beta)} \|_{L^p(\mathbb{Q}_q^k)}$$

$$+ \Big(\sum_{K \in P_{\delta_0}} \|f_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{1/2}.$$

Fix now $\varepsilon > 0$. For each of the $\lesssim (\log \delta_0^{-1})^3$ choices of (H, α, β), we apply Lemma 4.2 with δ replaced by δ_0, to

$$g := \sum_{K \in P_{\delta_0}} f_K^{(H,\alpha,\beta)} = \sum_{J \in P_\nu^{(H,\alpha,\beta)}} \sum_{K \in P_{\delta_0}(J)} f_K^{(H,\alpha)} \tag{5.9}$$

where

$$P_\nu^{(H,\alpha,\beta)} = \{J \in P_\nu : \#\{K'' \in P_{\delta_0}(J) : f_{K''}^{(H,\alpha)} \neq 0\} \in (\beta/2, \beta]\}$$

and $\kappa := q^{\lfloor \log_q \delta_0^\varepsilon \rfloor} \leqslant \delta_0^\varepsilon$. Note that this implies

$$g_J = 1_{P_\nu^{(H,\alpha,\beta)}}(J) \sum_{K \in P_{\delta_0}(J)} f_K^{(H,\alpha)} \tag{5.10}$$

and $g_K = f_K^{(H,\alpha)}$ if K's parent J is contained in $P_\nu^{(H,\alpha,\beta)}$ and 0 otherwise. Write N for the number of $J \in P_\nu$ for which $g_J \neq 0$ as in Lemma 4.2, and so $N = \# P_\nu^{(H,\alpha,\beta)}$. Note that by assumption the number of nonzero terms in the $\sum_K$ in (5.10) is $\sim \beta$.

With this, we then first compute

$$\max_{K \in P_{\delta_0}} \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}^k \sim H^k \tag{5.11}$$

since $g_K = f_K^{(H,\alpha,\beta)}$ which has height $\sim H$. Next, we have

$$\big(\sum_{\bar{K} \in P_{\delta_0}} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \big)^k = \big(\sum_{J \in P_\nu} \sum_{\bar{K} \in P_{\delta_0}(J)} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \big)^k$$

$$= \big(\sum_{J \in P_\nu^{(H,\alpha,\beta)}} \sum_{\bar{K} \in P_{\delta_0}(J)} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \big)^k \sim (N\beta H)^k \tag{5.12}$$

since there are N such J for which $g_J \neq 0$ and by how g is defined, each of these J's that contribute has $\sim \beta$ children $\bar{K}$ such that $g_{\bar{K}} = f_{\bar{K}}^{(H,\alpha)} \neq 0$. We can finish this estimate once again by using that $g_{\bar{K}}$ has height $\sim H$. Third,

$$\max_{J \in P_\nu} \big(\sum_{K \in P_{\delta_0}(J)} \|g_K\|_{L^{p-2k}(\mathbb{Q}_q^k)}^2 \big)^{(p-2k)/2} = \max_{J \in P_\nu} \big(\sum_{K \in P_{\delta_0}(J)} \|g_K\|_{L^{p-2k}(B_{\delta_0^{-k}})}^2 \big)^{(p-2k)/2}$$

$$\sim_{p,k} \beta^{(p-2k)/2} H^{p-2k} \alpha \delta_0^{-k(k+1)/2} \tag{5.13}$$

since by how g is defined, the $\sum_{K \in P_{\delta_0}(J)}$ has $\sim \beta$ terms and each term is made up of $\sim \alpha$ wavepackets of height $\sim H$. Note here we made use that each $T \in \mathbb{T}(K)$ has volume $\delta_0^{-k(k+1)/2}$ and g_K is supported on $B_{\delta_0^{-k}}$. Finally, a similar computation gives that

$$\big(\sum_{K \in P_{\delta_0}} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \big)^{p/2} = \big(\sum_{J \in P_\nu} \sum_{K \in P_{\delta_0}(J)} \|g_K\|_{L^p(B_{\delta_0^{-k}})}^2 \big)^{p/2} \sim_{p,k} (N\beta)^{p/2} H^p \alpha \delta_0^{-k(k+1)/2}. \tag{5.14}$$

Combining (5.11)-(5.14) gives that

$$\max_{K \in P_{\delta_0}} \|g_K\|_{L^\infty(\mathbb{Q}_q^k)}^k \big(\sum_{\bar{K} \in P_{\delta_0}} \|g_{\bar{K}}\|_{L^\infty(\mathbb{Q}_q^k)} \big)^k \max_{J \in P_\nu} \big(\sum_{K \in P_{\delta_0}(J)} \|g_K\|_{L^{p-2k}(\mathbb{Q}_q^k)}^2 \big)^{(p-2k)/2}$$

$$\sim_{p,k} N^{-(p-2k)/2} \big(\sum_{K \in P_{\delta_0}} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \big)^{p/2}.$$

Using this with Lemma 4.2 where g is as given in (5.9), then shows that

$$\int_{\mathbb{Q}_q^k} |g|^p \leqslant C \mathfrak{D}_p\big(\tfrac{\delta_0}{\kappa}\big)^p \Big(\sum_{K \in P_{\delta_0}} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2}$$

$$+ C q^{-k(k-1)} \kappa^{-(k^2+4k-2)} \nu^{-\frac{k(k-1)}{2}} N^{\frac{p-2k}{2}} \mathfrak{D}_{p-2k}\big(\tfrac{\delta_0}{\nu}\big)^{p-2k} \Big(\sum_{K \in P_{\delta_0}} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2}.$$

Note $\frac{\delta_0}{\kappa} \geqslant \frac{\delta_0}{\delta_0^\varepsilon} = \delta_0^{1-\varepsilon} \geqslant \delta^{1-\varepsilon}$, so $\mathfrak{D}_p\big(\tfrac{\delta_0}{\kappa}\big) \leqslant D_p\big(\tfrac{\delta_0}{\kappa}\big) \leqslant D_p(\delta^{1-\varepsilon})$ where in the second inequality we have used monotonicity. Similarly, $\frac{\delta_0}{\nu} \geqslant \frac{\delta_0}{\delta_0^{1/k}} = \delta_0^{1-\frac{1}{k}} \geqslant \delta^{1-\frac{1}{k}}$, so $\mathfrak{D}_{p-2k}\big(\tfrac{\delta_0}{\nu}\big) \leqslant D_{p-2k}(\delta^{1-\frac{1}{k}})$. As a

result,

$$\int_{\mathbb{Q}_q^k} |g|^p \leqslant CD_p(\delta^{1-\varepsilon})^p \Big(\sum_{K \in P_{\delta_0}} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2}$$

$$+ Cq^{-k(k-1)} \kappa^{-(k^2+4k-2)} \nu^{-\frac{k(k-1)}{2}} N^{\frac{p-2k}{2}} D_{p-2k}(\delta^{1-\frac{1}{k}})^{p-2k} \Big(\sum_{K \in P_{\delta_0}} \|g_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2}.$$

Now use $N \leqslant \nu^{-1}$ and $\|g_K\|_{L^p(\mathbb{Q}_q^k)} = \|f_K^{(H,\alpha,\beta)}\|_{L^p(\mathbb{Q}_q^k)} \leqslant \|f_K\|_{L^p(\mathbb{Q}_q^k)}$. Thus

$$\int_{\mathbb{Q}_q^k} |f|^p \leqslant C(\log \delta^{-1})^{3p} \times$$

$$\Big[D_p(\delta^{1-\varepsilon})^p + q^{-k(k-1)} \kappa^{-(k^2+4k-2)} \nu^{-(\frac{p}{2}+\frac{k(k-3)}{2})} D_{p-2k}(\delta^{1-\frac{1}{k}})^{p-2k} \Big] \Big(\sum_{K \in P_{\delta_0}} \|f_K\|_{L^p(\mathbb{Q}_q^k)}^2 \Big)^{p/2}.$$

But $\nu^{-1} \leqslant q\delta_0^{-1/k} \leqslant q\delta^{-1/k}$ and $\kappa^{-1} \leqslant q\delta_0^{-\varepsilon} \leqslant q\delta^{-\varepsilon}$. This completes the proof of (5.2). $\qquad \square$

5.2. Proof of Theorem 1.1.

We now finish the proof of Theorem 1.1. It suffices to iterate (5.2) by using an induction on p and induction on δ. Applying the definition of $D_p(\delta)$ from (5.1) and the hypothesis of Theorem 1.1 gives that

$$D_{p_0}(\delta)^{p_0} \leqslant C_1 \delta^{-(\frac{p_0}{2}-\frac{k(k+1)}{2})-c(p_0)(1-\frac{1}{k})^{p_0/(2k)}}$$

for all $\delta \in (0,1)$ and some $c(p_0) \geqslant 0$ such that the power of δ^{-1} is nonnegative. Note that from (1.9), $a(p,p_0) = a(p-2k,p_0) + \frac{p}{2} + \frac{k^2+7k-4}{2}$ and $a(p_0,p_0) = 0$. Additionally, (1.10) gives $b(p_0) \geqslant 0$ where

$$b(p) := (p - k(k+1))(1 - \frac{1}{k})^{-\frac{p}{2k}} + 2c(p_0),$$

and $b(p)$ is an increasing function of p on $[2,\infty)$: indeed, for $p \geqslant 2$, we have

$$b'(p) = (1 - \frac{1}{k})^{-\frac{p}{2k}} \Big[1 + \frac{p - k(k+1)}{2k} \log(1 - \frac{1}{k})^{-1} \Big]$$

$$= (1 - \frac{1}{k})^{-\frac{p}{2k}} \Big[1 + \frac{p - k(k+1)}{2k} (\log k - \log(k-1)) \Big]$$

$$\geqslant (1 - \frac{1}{k})^{-\frac{p}{2k}} \Big[1 + \frac{2 - k(k+1)}{2k} (\log k - \log(k-1)) \Big]$$

$$\geqslant (1 - \frac{1}{k})^{-\frac{p}{2k}} \Big[1 + \frac{2 - k(k+1)}{2k} \frac{1}{k-1} \Big],$$

where we used $p \geqslant 2$ in the first inequality, and used $\log k - \log(k-1) \leqslant \frac{1}{k-1}$ with $2 - k(k+1) \leqslant 0$ in the second inequality. This gives

$$b'(p) \geqslant (1 - \frac{1}{k})^{-\frac{p}{2k}} \Big[1 + \frac{2 - k - k^2}{2k(k-1)} \Big] = (1 - \frac{1}{k})^{-\frac{p}{2k}} \Big[1 - \frac{k+2}{2k} \Big] \geqslant 0$$

since $k \geqslant 2$, proving that $b(p)$ is an increasing function of p on $[2,\infty)$. As a result, from $b(p_0) \geqslant 0$, we see that $b(p) \geqslant 0$ for all $p \geqslant p_0$, and hence

$$\frac{p}{2} - \frac{k(k+1)}{2} + c(p_0)(1 - \frac{1}{k})^{\frac{p}{2k}} \geqslant 0 \tag{5.15}$$

for all $p \geqslant p_0$.

Assume for every $0 < \varepsilon < 1$ and all $\delta \in (0,1)$ we know

$$D_{p-2k}(\delta)^{p-2k} \leqslant C_{p-2k,\varepsilon} q^{a(p-2k,p_0)} \delta^{-(\frac{p-2k}{2}-\frac{k(k+1)}{2})-c(p_0)(1-\frac{1}{k})^{\frac{p-2k}{2k}}-\varepsilon}$$

for some $p \in p_0 + 2k\mathbb{N}$ (this is true for $p = p_0 + 2k$) and $C_{p-2k,\varepsilon}$ is allowed to depend on C_1. Then (5.2) gives

$$D_p(\delta)^p \leqslant C(\log \delta^{-1})^{3p}\Big[D_p(\delta^{1-\varepsilon})^p$$

$$+ C_{p-2k,\varepsilon}q^{a(p,p_0)}\delta^{-(k^2+4k-2)\varepsilon}\delta^{-\frac{1}{k}(\frac{p}{2}+\frac{k(k-3)}{2})}\delta^{-(1-\frac{1}{k})(\frac{p-2k}{2}-\frac{k(k+1)}{2})-c(p_0)(1-\frac{1}{k})^{\frac{p-2k}{2k}+1}-\varepsilon}\Big]$$

$$= C(\log \delta^{-1})^{3p}D_p(\delta^{1-\varepsilon})^p + CC_{p-2k,\varepsilon}q^{a(p,p_0)}\delta^{-(\frac{p}{2}-\frac{k(k+1)}{2})-c(p_0)(1-\frac{1}{k})^{\frac{p}{2k}}-(k^2+4k)\varepsilon}$$

for all $\delta, \varepsilon \in (0,1)$ where C here depends only on k and p. Iterating this inequality M times with M to be chosen later gives that

$$D_p(\delta)^p \leqslant C^M(\log \delta^{-1})^{3Mp}D_p(\delta^{(1-\varepsilon)^M})^p$$

$$+ CC_{p-2k,\varepsilon}q^{a(p,p_0)}\delta^{-(k^2+4k)\varepsilon}\sum_{j=0}^{M-1}C^j(\log\delta^{-1})^{3pj}\delta^{-(1-\varepsilon)^j[(\frac{p}{2}-\frac{k(k+1)}{2})+c(p_0)(1-\frac{1}{k})^{\frac{p}{2k}}]}.$$

Trivially, we have $D_p(\delta^{(1-\varepsilon)^M}) \leqslant \delta^{-(1-\varepsilon)^M/2}$. Thus

$$D_p(\delta)^p \leqslant C^M(\log\delta^{-1})^{3Mp}\delta^{-(1-\varepsilon)^Mp/2}$$

$$+ CC_{p-2k,\varepsilon}q^{a(p,p_0)}\delta^{-(k^2+4k)\varepsilon}\sum_{j=0}^{M-1}C^j(\log\delta^{-1})^{3pj}\delta^{-(1-\varepsilon)^j[(\frac{p}{2}-\frac{k(k+1)}{2})+c(p_0)(1-\frac{1}{k})^{\frac{p}{2k}}]}. \qquad (5.16)$$

By (5.15), the power of δ^{-1} in (5.16) is positive and so using that $(1-\varepsilon)^j \leqslant 1$, the sum can be controlled by

$$MC^M(\log\delta^{-1})^{3Mp}\delta^{-(\frac{p}{2}-\frac{k(k+1)}{2})-c(p_0)(1-\frac{1}{k})^{\frac{p}{2k}}}.$$

Inserting this into (5.16) and choosing M be the least integer such that $(1-\varepsilon)^M \leqslant \varepsilon$ (and so $M = \lceil\frac{\log\varepsilon^{-1}}{\log(1-\varepsilon)^{-1}}\rceil$) then shows that

$$D_p(\delta) \lesssim_{p,\varepsilon,C_1} q^{a(p,p_0)/p}(\log\delta^{-1})^{3M}\delta^{-(\frac{1}{2}-\frac{k(k+1)}{2p})-\frac{c(p_0)}{p}(1-\frac{1}{k})^{\frac{p}{2k}}-\frac{(k^2+4k)\varepsilon}{p}}$$

for all $\delta, \varepsilon \in (0,1)$. Since $(\log\delta^{-1})^{3M} \lesssim_\varepsilon \delta^{-\varepsilon}$, by redefining ε we have

$$D_p(\delta) \lesssim_{p,\varepsilon,C_1} q^{a(p,p_0)/p}\delta^{-(\frac{1}{2}-\frac{k(k+1)}{2p})-\frac{c(p_0)}{p}(1-\frac{1}{k})^{\frac{p}{2k}}-\varepsilon}.$$

APPENDIX A. PROOF OF $\mathfrak{D}_{2k}(\delta) \lesssim_\varepsilon \delta^{-\varepsilon}$

Fix $k \in \mathbb{N}$ and a prime $q > k$. For $\delta \in q^{-\mathbb{N}}$, let $S(\delta)$ be the smallest constant such that the reverse square function estimate

$$\int_{\mathbb{Q}_q^k}|g|^{2k} \leqslant S(\delta)^{2k}\int_{\mathbb{Q}_q^k}\big(\sum_{K\in P_\delta}|g_K|^2\big)^k$$

holds for every Schwartz function g on $\mathbb{Q}_q^k$ with Fourier transform supported in $\bigcup_{K\in P_\delta}\theta_K$. We will prove that

$$S(\delta) \lesssim_\varepsilon \delta^{-\varepsilon}$$

for every $\varepsilon > 0$, which by Minkowski's inequality is stronger than the assertion $\mathfrak{D}_{2k}(\delta) \lesssim_\varepsilon \delta^{-\varepsilon}$.

Let $\delta \in q^{-\mathbb{N}}$, g be as above, and $\kappa \in q^{-\mathbb{N}} \cap [\delta, 1]$. The broad/narrow dichotomy given by the pointwise estimate (4.2) implies

$$\int_{\mathbb{Q}_q^k}|g|^{2k} \leqslant 2^{2k-1}k^{2k}\sum_{I\in P_\kappa}\int_{\mathbb{Q}_q^k}|g_I|^{2k} + 2^{2k-1}\kappa^{-(4k-2)}\sum_{\substack{I_1,\ldots,I_k\in P_\kappa \\ d(I_i,I_j)>\kappa\,\forall i\neq j}}\int_{\mathbb{Q}_q^k}|g_{I_1}\ldots g_{I_k}|^2 \qquad (A.1)$$

Furthermore, by a rescaling argument similar to that in Lemma 4.1, we have

$$\sum_{I \in P_\kappa} \int_{\mathbb{Q}_q^k} |g_I|^{2k} \leqslant S(\frac{\delta}{\kappa})^{2k} \sum_{I \in P_\kappa} \int_{\mathbb{Q}_q^k} \Big(\sum_{K \in P_\delta(I)} |g_K|^2 \Big)^k \leqslant S(\frac{\delta}{\kappa})^{2k} \int_{\mathbb{Q}_q^k} \Big(\sum_{K \in P_\delta} |g_K|^2 \Big)^k \qquad \text{(A.2)}$$

where we used the pointwise inequality $\sum_{I \in P_\kappa} (\sum_{K \in P_\delta(I)} |g_K|^2)^k \leqslant (\sum_{K \in P_\delta} |g_K|^2)^k$ in the last inequality. To proceed further, fix now $I_1, \dots, I_k \in P_\kappa$ with $d(I_i, I_j) > \kappa$ for all $i \neq j$. We expand

$$\int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 = \sum_{\substack{K_i \in P_\delta(I_i) \\ i=1,\dots,k}} \sum_{\substack{\bar{K}_j \in P_\delta(I_j) \\ j=1,\dots,k}} \int_{\mathbb{Q}_q^k} g_{K_1} \cdots g_{K_k} \overline{g_{\bar{K}_1}} \cdots \overline{g_{\bar{K}_k}}$$

and write

$$\int_{\mathbb{Q}_q^k} g_{K_1} \cdots g_{K_k} \overline{g_{\bar{K}_1}} \cdots \overline{g_{\bar{K}_k}} = [\widehat{g_{K_1}} * \cdots * \widehat{g_{K_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}}](0).$$

For each $\bar{K}_1 \in P_\delta(I_1), \dots, \bar{K}_k \in P_\delta(I_k)$, we count the number of ordered k-tuples $(K_1, \dots, K_k)$ with $K_i \in P_\delta(I_i)$ for $i = 1, \dots, k$ and $0 \in \text{supp}(\widehat{g_{K_1}} * \cdots * \widehat{g_{K_k}} * \widehat{\overline{g_{\bar{K}_1}}} * \cdots * \widehat{\overline{g_{\bar{K}_k}}})$. The proof of Lemma 4.4 shows that the number of such ordered k-tuples is $\leqslant (q\kappa)^{-k(k-1)}$ (in fact, here we only need that $\widehat{g_{K_j}}$ is supported in the cube τ_{K_j} rather than the smaller parallelepiped θ_{K_j}). So using Cauchy-Schwarz,

$$\sum_{\substack{K_i \in P_\delta(I_i) \\ i=1,\dots,k}} \sum_{\substack{\bar{K}_j \in P_\delta(I_j) \\ j=1,\dots,k}} \int_{\mathbb{Q}_q^k} g_{K_1} \cdots g_{K_k} \overline{g_{\bar{K}_1}} \cdots \overline{g_{\bar{K}_k}} \leqslant (q\kappa)^{-k(k-1)} \sum_{\substack{K_i \in P_\delta(I_i) \\ i=1,\dots,k}} \int_{\mathbb{Q}_q^k} |g_{K_1} \cdots g_{K_k}|^2.$$

It follows that

$$\sum_{\substack{I_1,\dots,I_k \in P_\kappa \\ d(I_i,I_j)>\kappa \,\forall i \neq j}} \int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 \leqslant (q\kappa)^{-k(k-1)} \int_{\mathbb{Q}_q^k} \Big(\sum_{K \in P_\delta} |g_K|^2 \Big)^k. \qquad \text{(A.3)}$$

Alternatively, multilinear restriction estimate and L^2 orthogonality says that for any ball $B_{\delta^{-1}}$ of radius δ^{-1} in $\mathbb{Q}_q^k$, one has

$$\int_{B_{\delta^{-1}}} |g_{I_1} \cdots g_{I_k}|^2 \lesssim_\kappa (\delta^{k-1})^k \prod_{j=1}^{k} \int_{B_{\delta^{-1}}} |g_{I_j}|^2 = |B_{\delta^{-1}}|^{-(k-1)} \prod_{j=1}^{k} \int_{B_{\delta^{-1}}} \Big(\sum_{K_j \in P_\delta(I_j)} |g_{K_j}|^2 \Big),$$

and since each $|g_{K_j}|$ is constant on B_δ^{-1}, we have

$$|B_{\delta^{-1}}|^{-(k-1)} \prod_{j=1}^{k} \int_{B_{\delta^{-1}}} \Big(\sum_{K_j \in P_\delta(I_j)} |g_{K_j}|^2 \Big) = \int_{B_{\delta^{-1}}} \prod_{j=1}^{k} \Big(\sum_{K_j \in P_\delta(I_j)} |g_{K_j}|^2 \Big).$$

Summing over all $B_{\delta^{-1}} \subset \mathbb{Q}_q^k$ and all $I_1, \dots, I_k \in P_\kappa$, we have

$$\sum_{\substack{I_1,\dots,I_k \in P_\kappa \\ d(I_i,I_j)>\kappa \,\forall i \neq j}} \int_{\mathbb{Q}_q^k} |g_{I_1} \cdots g_{I_k}|^2 \lesssim_\kappa \int_{\mathbb{Q}_q^k} \Big(\sum_{K \in P_\delta} |g_K|^2 \Big)^k,$$

which for the purposes below is as good as (A.3). Putting (A.2) and (A.3) back into (A.1), we have

$$S(\delta)^{2k} \leqslant 2^{2k-1} k^{2k} S(\frac{\delta}{\kappa})^{2k} + 2^{2k-1} \kappa^{-(4k-2)} (q\kappa)^{-k(k-1)}.$$

Iterating this gives

$$S(\delta)^{2k} \leqslant (2^{2k-1} k^{2k})^N S(\frac{\delta}{\kappa^N})^{2k} + N 2^{2k-1} \kappa^{-(4k-2)} (q\kappa)^{-k(k-1)}$$

for all positive integers N for which $\kappa^N \geqslant \delta$; in particular, applying this with $N = \lfloor \frac{\log \delta^{-1}}{\log \kappa^{-1}} \rfloor$, and noting that $S(\delta/\kappa^N) \leqslant (\delta/\kappa^N)^{-1/2} \leqslant \kappa^{-1/2}$, we have

$$S(\delta)^{2k} \leqslant \delta^{-\frac{\log(2^{2k-1}k^{2k})}{\log \kappa^{-1}}} \kappa^{-k} + \frac{\log \delta^{-1}}{\log \kappa^{-1}} 2^{2k-1} \kappa^{-(4k-2)} (q\kappa)^{-k(k-1)}.$$

By choosing $\kappa = \kappa(\varepsilon)$ sufficiently small so that $\frac{\log(2^{2k-1}k^{2k})}{\log \kappa^{-1}} \leqslant 2k\varepsilon$, one obtains $S(\delta) \lesssim_{\varepsilon} \delta^{-\varepsilon}$, as desired.

References

1. Kirsti D. Biggs, *Efficient congruencing in ellipsephic sets: the quadratic case*, Acta Arith. **200** (2021), no. 4, 331–348.
2. Kirsti D. Biggs, Julia Brandes, and Kevin Hughes, *Reinforcing a Philosophy: A counting approach to square functions over local fields*, arXiv:2201.09649.
3. Jean Bourgain, Ciprian Demeter, and Larry Guth, *Proof of the main conjecture in Vinogradov's mean value theorem for degrees higher than three*, Ann. of Math. (2) **184** (2016), no. 2, 633–682.
4. Jean Bourgain and Larry Guth, *Bounds on oscillatory integral operators based on multilinear estimates*, Geom. Funct. Anal. **21** (2011), no. 6, 1239–1295.
5. Alan Chang, Jaume de Dios Pont, Rachel Greenfeld, Asgar Jamneshan, Zane Kun Li, and José Madrid, *Decoupling for fractal subsets of the parabola*, Mathematische Zeitschrift **301** (2022), 1851–1879.
6. Antonio Córdoba, *The Kakeya maximal function and the spherical summation multipliers*, Amer. J. Math. **99** (1977), no. 1, 1–22.
7. ———, *Geometric Fourier analysis*, Ann. Inst. Fourier (Grenoble) **32** (1982), no. 3, vii, 215–226.
8. Ciprian Demeter, *Fourier restriction, decoupling, and applications*, Cambridge Studies in Advanced Mathematics, vol. 184, Cambridge University Press, Cambridge, 2020.
9. Stephen W. Drury, *Restrictions of Fourier transforms to curves*, Ann. Inst. Fourier (Grenoble) **35** (1985), no. 1, 117–123.
10. Charles Fefferman, *A note on spherical summation multipliers*, Israel J. Math. **15** (1973), 44–52.
11. Kevin Ford, *Vinogradov's integral and bounds for the Riemann zeta function*, Proc. London Math. Soc. (3) **85** (2002), no. 3, 565–633.
12. ———, *Zero-free regions for the Riemann zeta function*, Number theory for the millennium, II (Urbana, IL, 2000), A K Peters, Natick, MA, 2002, pp. 25–56.
13. Philip T. Gressman, Shaoming Guo, Lillian B. Pierce, Joris Roos, and Po-Lam Yung, *Reversing a philosophy: from counting to square functions and decoupling*, J. Geom. Anal. **31** (2021), no. 7, 7075–7095.
14. Shaoming Guo, Zane Kun Li, and Po-Lam Yung, *Improved discrete restriction for the parabola*, arXiv:2103.09795, to appear in *Mathematical Research Letters*.
15. ———, *A bilinear proof of decoupling for the cubic moment curve*, Trans. Amer. Math. Soc. **374** (2021), no. 8, 5405–5432.
16. Shaoming Guo, Zane Kun Li, Po-Lam Yung, and Pavel Zorin-Kranich, *A short proof of ℓ^2 decoupling for the moment curve*, American J. Math. **143** (2021), no. 6, 1983–1998.
17. Larry Guth, *A restriction estimate using polynomial partitioning*, J. Amer. Math. Soc. **29** (2016), no. 2, 371–413.
18. Larry Guth, Dominique Maldague, and Hong Wang, *Improved decoupling for the parabola*, arXiv:2009.07953, to appear in the *Journal of the European Mathematical Society*.
19. D. R. Heath-Brown, *A new kth derivative estimate for exponential sums via Vinogradov's mean value*, Proceedings of the Steklov Institute of Mathematics **296** (2017), 88–103.
20. ———, *The cubic case of Vinogradov's mean value theorem – a simplified approach to Wooley's "efficient congruencing"*, Essential Number Theory **1** (2022), no. 1, 1–12.
21. Jonathan Hickman and James Wright, *A non-archimedean variant of Littlewood–Paley theory for curves*, The Journal of Geometric Analysis **33** (2023), no. 104.
22. A. A. Karatsuba, *Mean value of the modulus of a trigonometric sum*, Izv. Akad. Nauk SSSR Ser. Mat. **37** (1973), 1203–1227.
23. Zane Kun Li, *An l^2 decoupling interpretation of efficient congruencing: the parabola*, Rev. Mat. Iberoam. **37** (2021), no. 5, 1761–1802.
24. U. V. Linnik, *On Weyl's sums*, Rec. Math. [Mat. Sbornik] N.S. **12(54)** (1943), 28–39.
25. Akshat Mudgal, *Diameter free estimates for the quadratic Vinogradov mean value theorem*, Proceedings of the London Mathematical Society **126** (2023), no. 1, 76–128.

26. Lillian B. Pierce, *The Vinogradov mean value theorem [after Wooley, and Bourgain, Demeter and Guth]*, Astérisque Exposés Bourbaki **407** (2019), 479–564.
27. S. B. Stečkin, *Mean values of the modulus of a trigonometric sum*, Trudy Mat. Inst. Steklov. **134** (1975), 283–309.
28. M. H. Taibleson, *Fourier analysis on local fields*, Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1975.
29. Terence Tao, *254A, Notes 5: Bounding exponential sums and the zeta function*, https://terrytao.wordpress.com/2015/02/07/254a-notes-5-bounding-exponential-sums-and-the-zeta-function/.
30. _____, *Recent progress on the restriction conjecture*, arXiv:math/0311181.
31. R. C. Vaughan, *The Hardy-Littlewood method*, second ed., Cambridge Tracts in Mathematics, vol. 125, Cambridge University Press, Cambridge, 1997.
32. I.M. Vinogradov, *New estimates for Weyl sums*, Dokl. Akad. Nauk SSSR **8** (1935), 195–198.
33. V. S. Vladimirov, I. V. Volovich, and E. I. Zelenov, *p-adic analysis and mathematical physics*, Series on Soviet and East European Mathematics, vol. 1, World Scientific Publishing Co., Inc., River Edge, NJ, 1994.
34. Trevor D. Wooley, *Translation invariance, exponential sums, and Waring's problem*, Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. II, Kyung Moon Sa, Seoul, 2014, pp. 505–529.
35. _____, *The cubic case of the main conjecture in Vinogradov's mean value theorem*, Adv. Math. **294** (2016), 532–561.
36. _____, *Nested efficient congruencing and relatives of Vinogradov's mean value theorem*, Proceedings of the London Mathematical Society **118** (2019), no. 4, 942–1016.

Department of Mathematics, Virginia Tech, Blacksburg, VA 24061-0123, USA
Email address: briancookmath@gmail.com

School of Mathematics, The University of Bristol, Bristol BS8 1UG; and the Heilbronn Insitute for Mathematical Research, Bristol, UK
Email address: khughes.math@gmail.com

Department of Mathematics, North Carolina State University, Raleigh, NC 27695, USA
Email address: zkli@ncsu.edu

Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK
Email address: Akshat.Mudgal@maths.ox.ac.uk

Université de Lyon, Université de Saint-Étienne, CNRS UMR 5208, Institut Camille Jordan, F-42000 Saint-Étienne, France.
Email address: olivier.robert@univ-st-etienne.fr

Mathematical Sciences Institute, Australian National University, Canberra ACT 2601, Australia
Email address: PoLam.Yung@anu.edu.au