

THE MULTILINEAR CIRCLE METHOD AND A QUESTION OF BERGELSON

DARIUSZ KOSZ, MARIUSZ MIREK, SARAH PELUSE, RENHUI WAN, AND JAMES WRIGHT

Dedicated to the memory of John L. Wright.

ABSTRACT. Let $k \in \mathbb{Z}_+$ and $(X, \mathcal{B}(X), \mu)$ be a probability space equipped with a family of commuting invertible measure-preserving transformations $T_1, \dots, T_k: X \rightarrow X$. Let $P_1, \dots, P_k \in \mathbb{Z}[n]$ be polynomials with integer coefficients and distinct degrees. We establish pointwise almost everywhere convergence of the multilinear polynomial ergodic averages

$$\frac{1}{N} \sum_{n=1}^N f_1(T_1^{P_1(n)}x) \cdots f_k(T_k^{P_k(n)}x), \quad x \in X,$$

as $N \rightarrow \infty$ for any functions $f_1, \dots, f_k \in L^\infty(X)$. Besides a couple of results in the bilinear setting $k = 2$, and then only in the single transformation case $T_1 = T_2$, this is the first pointwise result for general polynomial multilinear ergodic averages in arbitrary measure-preserving systems. This answers a question of Bergelson from 1996 in the affirmative for any polynomials with distinct degrees, and makes progress on the Furstenberg–Bergelson–Leibman conjecture.

In this paper, we build a versatile *multilinear circle method* by developing the Ionescu–Wainger multiplier theorem for the set of canonical fractions, which gives a positive answer to a question of Ionescu and Wainger from 2005. We also establish multilinear L^p -improving bounds and an inverse theorem in higher order Fourier analysis for averages over polynomial corner configurations, which we use to establish a multilinear analogue of Weyl’s inequality and its real counterpart, a Sobolev smoothing estimate.

CONTENTS

1. Introduction	2
1.1. A brief history	2
1.2. Norm convergence	4
1.3. Pointwise convergence	5
1.4. Statement of the main results	6
1.5. The Furstenberg–Bergelson–Leibman conjecture	9
1.6. Reduction to the integer shift system	10

Date: September 3, 2026.

Dariusz Kosz was partially supported by Basque Government grant BERC 2022-2025, by Spanish State Research Agency grant CEX2021-001142-S, and by National Science Centre of Poland grant SONATA BIS 2022/46/E/ST1/00036. Mariusz Mirek was partially supported by NSF grant DMS-2154712 and by NSF CAREER grant DMS-2236493. Sarah Peluse was partially supported by NSF grant DMS-2516641 and by a Sloan Research Fellowship. James Wright was partially supported by a Leverhulme Research Fellowship RF-2023-709\9.

1.7. The multilinear circle method	12
1.8. Final remarks and open problems	21
Acknowledgments	22
2. Notation and useful tools	23
2.1. Basic notation	23
2.2. Euclidean spaces	24
2.3. Function spaces	24
2.4. Convolutions	25
2.5. Fourier transform	25
2.6. Variational norms	27
3. Ionescu–Wainger multiplier theorem for canonical fractions	28
3.1. Magyar–Stein–Wainger sampling method	28
3.2. A refined variant of the Ionescu–Wainger multiplier theorem	29
3.3. A seminorm variant of the Ionescu–Wainger theorem	37
4. Inverse theorems	38
4.1. Basic definitions and the statement of the inverse theorem	38
4.2. Gowers norms and their basic properties	41
4.3. The main technical reduction	43
4.4. Gowers norm control	46
4.5. Degree lowering	49
5. Multilinear L^p -improving inequalities	64
5.1. Vinogradov mean value theorem	65
5.2. Lifting procedure	66
5.3. Refinements	68
5.4. Parameter towers and flows	69
5.5. The integer case	71
5.6. The real case	72
6. Multilinear Weyl inequality and Sobolev smoothing theorem	74
6.1. Multilinear smoothing inequalities	74
6.2. Structural theorem for adjoints	75
6.3. Proof of multilinear Weyl inequality	76
7. Multilinear circle method: proof of pointwise ergodic theorem	79
7.1. Dyadic Ionescu–Wainger projections	80
7.2. Minor arc estimates	81
7.3. Major arc estimates	82
References	101

1. INTRODUCTION

1.1. A brief history. A fundamental problem in ergodic theory is to understand the convergence, both in norm and pointwise almost everywhere, of multilinear polynomial ergodic averages. This line of inquiry started in the early 1930s with von Neumann’s mean ergodic

theorem [78] and Birkhoff's pointwise ergodic theorem [11]. Significant advances, which we will summarize shortly, have been made in this area of research over the last century.

In 1996, Bergelson [6, Question 9] formulated the following question.

Question 1.1 (Bergelson, 1996). Let $k \in \mathbb{Z}_+$ and $(X, \mathcal{B}(X), \mu)$ be a probability space endowed with a family of commuting invertible measure-preserving transformations $T_1, \dots, T_k : X \rightarrow X$. Let $P_1, \dots, P_k \in \mathbb{Z}[n]$ be any polynomials with integer coefficients. Is it true that for any functions $f_1, \dots, f_k \in L^\infty(X)$ the multilinear polynomial ergodic averages

$$A_{N; X, T_1, \dots, T_k}^{P_1, \dots, P_k}(f_1, \dots, f_k)(x) := \mathbb{E}_{n \in [N]} \prod_{i \in [k]} f_i(T_i^{P_i(n)} x), \quad x \in X, \quad (1.2)$$

converge pointwise almost everywhere on X with respect to μ as $N \rightarrow \infty$?

Here and throughout the paper we use the notation $[N] := (0, N] \cap \mathbb{Z}$ for any real number $N \geq 1$ and $\mathbb{E}_{y \in Y} f(y) := \frac{1}{\#Y} \sum_{y \in Y} f(y)$ for any finite set $Y \neq \emptyset$ and any function $f : Y \rightarrow \mathbb{C}$.

One of the main results of this paper is the following theorem.

Theorem 1.3. *The answer to Question 1.1 is yes for any $k \in \mathbb{Z}_+$ and any polynomials $P_1, \dots, P_k \in \mathbb{Z}[n]$ with distinct degrees.*

In order to understand the origins of Bergelson's question, one has to go back to 1977, when Furstenberg [36] gave an ergodic theoretic proof of Szemerédi's theorem [89], which asserts that every subset of the integers with positive upper density must contain arbitrarily long arithmetic progressions. In Furstenberg's approach [36], multilinear averages of the form (1.2) with $P_1(n) = n, \dots, P_k(n) = kn$ and $T_1 = \dots = T_k$ served as a natural tool to detect recurrent points and, consequently, arithmetic progressions in subsets of integers with positive upper density. This gave rise to a new field now known as ergodic Ramsey theory.

Not long afterwards, Bergelson had the great insight to initiate a challenging program with the goal of establishing a polynomial extension of Szemerédi's theorem by studying the asymptotic behavior of the corresponding multilinear polynomial ergodic averages. This led to his foundational paper [5] on weakly mixing PET (polynomial ergodic theorem), in which the van der Corput differencing technique was cemented as one of the primary tools in the field and was used to establish $L^2(X)$ norm convergence of multilinear averages (1.2) to the product of integrals $\int_X f_1 \cdots \int_X f_k$ provided that $T_1 = \dots = T_k = T$ is a weakly mixing measure-preserving transformation on X and $P_1, \dots, P_k$ are pairwise essentially distinct, i.e., $P_i - P_j$ is nonconstant whenever $i \neq j$.

Almost two decades after Furstenberg's paper [36], Bergelson and Leibman [8] achieved the main goal of Bergelson's program by establishing (among other results — see the discussion below Conjecture 1.15 and inequality (1.17)) the following far-reaching polynomial extension of the classical Poincaré recurrence theorem and the multidimensional Szemerédi theorem of Furstenberg and Katznelson [39].

Theorem 1.4 (The Bergelson–Leibman polynomial Szemerédi theorem, 1996). *Let $k \in \mathbb{Z}_+$ and $(X, \mathcal{B}(X), \mu)$ be a probability space endowed with a family of commuting invertible measure-preserving transformations $T_1, \dots, T_k : X \rightarrow X$. Let $P_1, \dots, P_k \in \mathbb{Z}[n]$ be polynomials with vanishing constant terms. Then, for any*

$A \in \mathcal{B}(X)$ satisfying $\mu(A) > 0$, one has

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \mu \left(T_1^{-P_1(n)}(A) \cap \cdots \cap T_k^{-P_k(n)}(A) \right) > 0. \quad (1.5)$$

Theorem 1.4 is merely a special case of a more general multidimensional polynomial Szemerédi theorem by Bergelson and Leibman [8], which is formulated in (1.17) below. This theorem sparked interest in understanding the asymptotic behavior of multilinear polynomial ergodic averages (1.2) as $N \rightarrow \infty$; see Conjecture 1.15 and inequality (1.17). Therefore, Theorem 1.3 naturally contributes to Bergelson’s program of understanding the asymptotic behavior of multilinear polynomial ergodic averages.

1.2. Norm convergence. Bergelson’s Question 1.1 was initially about the convergence of (1.2) in both the $L^2(X)$ norm and pointwise almost everywhere. Nowadays, the $L^2(X)$ norm convergence of (1.2) is fairly well understood due to groundbreaking work of Walsh [95]. Prior to Walsh’s work, there was an extensive body of research towards establishing $L^2(X)$ norm convergence for (1.2) in the single transformation case $T_1 = \cdots = T_k = T$. This includes fundamental work for linear polynomials due to Host and Kra [51] and, independently, Ziegler [98]. We also refer to Leibman’s appendix in [7] where the constructions of characteristic factors from the papers by Host and Kra [51] and Ziegler [98] are compared and shown to be equivalent. The case of general polynomials was addressed in work by Leibman [67], as well as by Frantzikinakis and Kra [33], and Host and Kra [50]. In the single transformation case, in contrast to Walsh’s paper [95], one can identify the limiting function in (1.2) thanks to the theory of Host–Kra factors [51] and equidistribution on nilmanifolds [67]. For instance, the limit in the single transformation case for linearly independent polynomials was established in [33] (see also [34, 50]), whereas the cases for three polynomials or multiples of one polynomial were addressed in [31]. We also refer to [32], where precise references and additional information on this topic can be found.

The case of arbitrary commuting measure-preserving transformations is much more complicated and for linear polynomials was subsequently studied by Tao [90], Austin [3], and Host [49]. In [23], Chu, Frantzikinakis, and Host established $L^2(X)$ norm convergence for averages of the form (1.2) when the polynomials have distinct degrees. Finally, Walsh [95] established norm convergence of (1.2) in the general case, even handling noncommutative transformations $T_1, \dots, T_k$ generating a nilpotent group. For more on this topic, we refer the interested reader to the discussion below Conjecture 1.15, as well as the articles [2, 99] for alternative proofs and generalizations of Walsh’s result and the survey articles [6, 7, 32] that include comprehensive historical background and an extensive literature on the subject of norm convergence and its combinatorial applications.

Although Walsh’s result establishes $L^2(X)$ norm convergence for (1.2), the question of identifying the limit for arbitrary transformations and polynomials remains widely unanswered, save the case of a single transformation and linear polynomials [51, 98] or a single transformation and linearly independent polynomials [33, 34, 50, 31].

Identifying the limit for general polynomial ergodic averages is a well-known open problem in ergodic theory, even in the linear case. The $L^2(X)$ limit for (1.2) in the case of commuting transformations and linearly independent polynomials has very recently been identified in a breakthrough paper by Frantzikinakis and Kuca [35]. They proved the following.

Theorem 1.6 (Frantzikinakis–Kuca, 2022). *Let $k \in \mathbb{Z}_+$ and $(X, \mathcal{B}(X), \mu)$ be a probability space endowed with a family of commuting invertible measure-preserving transformations $T_1, \dots, T_k: X \rightarrow X$. Let $P_1, \dots, P_k \in \mathbb{Z}[n]$ be linearly independent polynomials. Then the rational Kronecker factor is characteristic for the averages (1.2). In particular, if $T_1, \dots, T_k$ are totally ergodic, i.e., T_i^n is ergodic for any $n \in \mathbb{Z}_+$ and $i \in [k]$, then we have*

$$\lim_{N \rightarrow \infty} \left\| A_{N; X, T_1, \dots, T_k}^{P_1, \dots, P_k}(f_1, \dots, f_k) - \prod_{i \in [k]} \int_X f_i(y) d\mu(y) \right\|_{L^2(X)} = 0 \quad (1.7)$$

for any $f_1, \dots, f_k \in L^\infty(X)$.

Combining (1.7) with our Theorem 1.3 shows that the time averages (1.2) also converge pointwise almost everywhere to the product of space averages for totally ergodic transformations. More precisely, we have the following corollary.

Corollary 1.8. *Let $k \in \mathbb{Z}_+$ and $(X, \mathcal{B}(X), \mu)$ be a probability space endowed with a family of commuting invertible measure-preserving transformations $T_1, \dots, T_k: X \rightarrow X$ that are totally ergodic. Let $P_1, \dots, P_k \in \mathbb{Z}[n]$ be polynomials with distinct degrees. Then we have*

$$\lim_{N \rightarrow \infty} A_{N; X, T_1, \dots, T_k}^{P_1, \dots, P_k}(f_1, \dots, f_k)(x) = \prod_{i \in [k]} \int_X f_i(y) d\mu(y)$$

for any $f_1, \dots, f_k \in L^\infty(X)$ and for almost every $x \in X$.

As we have seen in the brief overview above, there has been more than four decades of tremendous effort from both the ergodic and combinatorial perspectives to understand the asymptotic nature of multilinear polynomial ergodic averages (1.2) in the $L^2(X)$ norm.

1.3. Pointwise convergence. The state of knowledge is dramatically worse for pointwise almost everywhere convergence of multilinear polynomial ergodic averages. The phenomenon of pointwise convergence, while the most natural mode of convergence, is very subtle and can differ significantly from norm convergence, making the study of pointwise convergence problems quite challenging. To begin to see the differences, let us consider the so-called uniform averages, which are given by (1.2) with shifted averages $\mathbb{E}_{n \in [N] \setminus [M]} := \frac{1}{N-M} \sum_{n=M+1}^N$ in place of the ordinary averages $\mathbb{E}_{n \in [N]} := \frac{1}{N} \sum_{n=1}^N$. These two averages coincide for $M = 0$. The asymptotic behavior of uniform averages in the $L^2(X)$ norm is pivotal in combinatorial applications; we refer to [6, 7, 18] for comprehensive motivations and an extensive literature.

By the dominated convergence theorem, the pointwise almost everywhere convergence of the uniform averages $\mathbb{E}_{n \in [N] \setminus [M]} \prod_{i \in [k]} f_i(T_i^{P_i(n)} x)$ as $N - M \rightarrow \infty$ implies their norm convergence for all $f_1, \dots, f_k \in L^\infty(X)$ on a probability space $(X, \mathcal{B}(X), \mu)$. Interestingly, pointwise almost everywhere convergence for uniform averages can fail, even for linear averages, as shown by del Junco and Rosenblatt [59]. This stands in sharp contrast to $L^2(X)$ norm convergence of uniform averages, which, for instance, follows from Zorin-Kranich's paper [99] in a very general setting.

However, returning to the averages in (1.2), as indicated in Bergelson's question (see also Conjecture 1.15 below), it is conjectured that pointwise convergence holds, highlighting a significant distinction between ordinary averages $\mathbb{E}_{n \in [N]}$ and uniform averages $\mathbb{E}_{n \in [N] \setminus [M]}$.

Pointwise almost everywhere convergence has only been established for a few special cases of averages $A_{N;X,T_1,\dots,T_k}^{P_1,\dots,P_k}(f_1,\dots,f_k) = \mathbb{E}_{n \in [N]} \prod_{i \in [k]} f_i(T_i^{P_i(n)}x)$.

1. The case $k = 1$ with $P_1(n) = n$ corresponds to the classical Birkhoff ergodic theorem [11].
2. The case $k = 1$ with an arbitrary polynomial $P_1 \in \mathbb{Z}[n]$ was an open problem of Bellow [4] and Furstenberg [38], and was solved by Bourgain in a series of fundamental papers [12, 13, 14] in the late 1980s.
3. Soon afterwards, Bourgain [15] also proved pointwise almost everywhere convergence in the bilinear setting when $k = 2$ with $T_1 = T_2$, $P_1(n) = an$, and $P_2(n) = bn$ for any $a, b \in \mathbb{Z}$, providing a positive answer to a question of Furstenberg [37].
4. Recently, the second author along with Krause and Tao [62] established pointwise almost everywhere convergence when $k = 2$ with $T_1 = T_2$, $P_1(n) = n$, and an arbitrary polynomial $P_2 \in \mathbb{Z}[n]$ of degree at least two. Throughout the paper, we will call

$$\mathbb{E}_{n \in [N]} f_1(T^n x) f_2(T^{P_2(n)} x)$$

the *Furstenberg–Weiss averages*, even though Furstenberg and Weiss studied $L^2(X)$ norm convergence of such averages in [40] only for quadratic polynomials P_2 .

5. Pointwise convergence of multilinear polynomial averages was established for some special classes of measure-preserving systems, such as exact endomorphisms and K -automorphisms [25] and nilsystems [68]. The context of commuting transformations along linear orbits for distal systems was studied in [53] and [29].

Aside from these results, nothing further is known about pointwise almost everywhere convergence for multilinear polynomial averages (1.2) in general measure-preserving systems.

Our paper forges new ground on three fronts: we treat averages with an arbitrary degree of multilinearity k , with any commuting transformations, and with all polynomials allowed to be nonlinear. All prior work in general measure-preserving systems treated only the linear or bilinear case $k \leq 2$, a single transformation, and at most one nonlinear polynomial. We obtain, for example, the first pointwise almost everywhere convergence result for averages with commuting transformations $S, T: X \rightarrow X$ corresponding to the “sqorners” configuration

$$\mathbb{E}_{n \in [N]} f_1(S^n x) f_2(T^{n^2} x), \quad x \in X, \tag{1.9}$$

recently considered by the third author along with Prendiville and Shao [83]. The methods used in our paper are robust enough to suggest that we can handle more general classes of ergodic averages than those considered here. We will elaborate on this later in the introduction, when we describe the *multilinear circle method* that we develop in this article.

1.4. Statement of the main results. Throughout this paper, the triple $(X, \mathcal{B}(X), \mu)$ denotes a σ -finite measure space and $\mathbb{K}$ denotes either $\mathbb{Z}$ or $\mathbb{R}$. Correspondingly, $\mathbb{K}[n]$ denotes the space of all formal polynomials $P(n)$ with coefficients and indeterminate n in $\mathbb{K}$. Each polynomial $P \in \mathbb{K}[n]$ is always identified with a map $\mathbb{K} \ni n \mapsto P(n) \in \mathbb{K}$.

Let $k \in \mathbb{Z}_+$. Given a family $\mathcal{T} = \{T_1, \dots, T_k\}$ of invertible commuting measure-preserving transformations on X , complex-valued measurable functions $f_1, \dots, f_k$ on X , a family of

polynomials $\mathcal{P} = \{P_1, \dots, P_k\} \subset \mathbb{Z}[\mathbb{n}]$, and a real number $N \geq 1$, we define, as in (1.2), the corresponding multilinear polynomial ergodic average by

$$A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)(x) := \mathbb{E}_{n \in [N]} \prod_{i \in [k]} f_i(T_i^{P_i(n)} x), \quad x \in X, \quad (1.10)$$

and its truncated version by

$$\tilde{A}_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)(x) := \mathbb{E}_{n \in [N] \setminus [N/2]} \prod_{i \in [k]} f_i(T_i^{P_i(n)} x), \quad x \in X. \quad (1.11)$$

We will often abbreviate $A_{N;X,\mathcal{T}}^{\mathcal{P}}$ to $A_{N;X}^{\mathcal{P}}$ and $\tilde{A}_{N;X,\mathcal{T}}^{\mathcal{P}}$ to $\tilde{A}_{N;X}^{\mathcal{P}}$ when the transformations are understood. Depending on how explicit we want to be, in some instances, we will write out the averages

$$\begin{aligned} A_{N;X}^{\mathcal{P}} &= A_{N;X}^{P_1, \dots, P_k} \quad \text{or} \quad A_{N;X,\mathcal{T}}^{\mathcal{P}} = A_{N;X,T_1, \dots, T_k}^{P_1, \dots, P_k}, \\ \tilde{A}_{N;X}^{\mathcal{P}} &= \tilde{A}_{N;X}^{P_1, \dots, P_k} \quad \text{or} \quad \tilde{A}_{N;X,\mathcal{T}}^{\mathcal{P}} = \tilde{A}_{N;X,T_1, \dots, T_k}^{P_1, \dots, P_k}. \end{aligned}$$

Using the asymptotic notation from Section 2 and the definition of r -variational norm $\mathbf{V}^r$ from (2.9), we now state the main result of this article, i.e., the following quantitative ergodic theorem that implies Theorem 1.3 and, consequently, gives an affirmative answer to Bergelson's Question 1.1 for polynomials with distinct degrees.

Theorem 1.12. *Let $k \in \mathbb{Z}_+$ and $(X, \mathcal{B}(X), \mu)$ be a σ -finite measure space equipped with a family $\mathcal{T}$ of invertible measure-preserving transformations $T_1, \dots, T_k: X \rightarrow X$. Suppose that $\mathcal{P} = \{P_1, \dots, P_k\} \subset \mathbb{Z}[\mathbb{n}]$ is a family of polynomials with distinct degrees. Let $f_i \in L^{p_i}(X)$ for $i \in [k]$ and some $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$, and let $A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)$ be the averages defined in (1.10). Then the following statements hold.*

- (i) (Mean ergodic theorem) *The functions $A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)$ converge in the $L^p(X)$ norm as $N \rightarrow \infty$.*
- (ii) (Pointwise ergodic theorem) *The functions $A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)$ converge pointwise almost everywhere as $N \rightarrow \infty$.*
- (iii) (Maximal ergodic theorem) *One has*

$$\left\| \sup_{N \in \mathbb{Z}_+} |A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)| \right\|_{L^p(X)} \lesssim \prod_{i \in [k]} \|f_i\|_{L^{p_i}(X)}. \quad (1.13)$$

- (iv) (Variational ergodic theorem) *Fix $r > 2$ and $\lambda > 1$. Then one has*

$$\|\mathbf{V}^r(A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k) : N \in \mathbb{D})\|_{L^p(X)} \lesssim \prod_{i \in [k]} \|f_i\|_{L^{p_i}(X)} \quad (1.14)$$

if $\mathbb{D} = \{\lambda_n : n \in \mathbb{N}\} \subset [1, +\infty)$ is λ -lacunary, i.e., $\inf_{n \in \mathbb{N}} \frac{\lambda_{n+1}}{\lambda_n} \geq \lambda$.

The same results hold with the truncated averages $\tilde{A}_{N;X,\mathcal{T}}^{\mathcal{P}}$ in place of $A_{N;X,\mathcal{T}}^{\mathcal{P}}$.

We now give some remarks about Theorem 1.12 and its consequences.

1. On a probability space, bounded functions are L^p functions and so the conclusion from part (ii) implies Theorem 1.3.

2. When $k = 1$ and $P_1(n) = n$, parts (i)–(iii) follow from works of von Neumann [78], Birkhoff [11], and Hopf [48]. It is also known that the maximal function is of weak-type $(1, 1)$, which can be derived from the corresponding bounds for the Hardy–Littlewood maximal function [87] by using the Calderón transference principle [19]. Part (iv) follows from [57], and weak-type $(1, 1)$ estimates also hold.
3. When $k = 1$ and $P_1 \in \mathbb{Z}[n]$ is arbitrary, part (i) was established by Furstenberg [37], parts (ii) and (iii) were established by Bourgain [12, 13, 14], and part (iv) follows, for instance, from [75]. Here, if $\deg P_1 \geq 2$, then pointwise almost everywhere convergence may fail for $p_1 = 1$, as was shown in [18, 66].
4. When $k = 2$, $P_1(n) = an$ and $P_2(n) = bn$ with $a, b \in \mathbb{Z}$, and $T_1 = T_2$, part (i) was established by Furstenberg [37] and part (ii) for bounded functions on a probability space was established by Bourgain [15]. Part (iii) follows from work of Lacey [65]; see also [24]. Finally, part (iv) was established by Do, Oberlin, and Palsson [27] for large $r > 2$.
5. When $k = 2$, $P_1(n) = n$ and $P_2 \in \mathbb{Z}[n]$ is arbitrary with $\deg P_2 \geq 2$, and $T_1 = T_2$, part (i) was originally established by Furstenberg and Weiss [40] in the case $P_2(n) = n^2$, whereas for general polynomials $P_2 \in \mathbb{Z}[n]$ with $\deg P_2 \geq 2$, it follows from the papers of Host and Kra [50] and Frantzikinakis and Kra [33]. Parts (ii)–(iv) were recently established by the second author with Krause and Tao in [62].
6. The implicit constant in (1.14) is allowed to depend on the parameters $p_1, \dots, p_k, r, \lambda$ and on the polynomials $P_1, \dots, P_k$.
7. If (1.14) is established for some $r > 2$ and, for example, the 2-lacunary set $\{2^n : n \in \mathbb{N}\}$, then (1.13) holds for the same $p_1, \dots, p_k$ with an implicit constant that depends only on $p_1, \dots, p_k, P_1, \dots, P_k$. The maximal inequality (1.13) also addresses a question raised by Christ, Durcik, and Roos in [22, Section 6, Problem 4].
8. If (1.14) holds with some $r > 2$ for all $\lambda > 1$ and all λ -lacunary sets $\mathbb{D} \subset [1, +\infty)$, then the limit of the averages $A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)(x)$ exists for μ -almost every $x \in X$ as $\mathbb{D} \ni N \rightarrow \infty$. This, in turn, implies pointwise almost everywhere convergence by taking the $2^{1/s}$ -lacunary set $\{2^{n/s} : n \in \mathbb{N}\}$ for every $s \in \mathbb{Z}_+$; see [62, Proposition 3.2(i)].
9. Combining (1.13) with the pointwise almost everywhere convergence of the averages $A_{N;X,\mathcal{T}}^{\mathcal{P}}(f_1, \dots, f_k)$ and the dominated convergence theorem yields the norm convergence, which was originally proven by Chu, Frantzikinakis, and Host [23], using different methods.
10. Taking into account the last three items above, it suffices to establish inequality (1.14). We can assume that $N \geq C_0$ for some fixed $C_0 \in \mathbb{Z}_+$.
11. The condition $r > 2$ in Theorem 1.12 is necessary, as no variational estimate is possible for $r \leq 2$; see [62, Corollary 12.4]. Thus, the range in (1.14) is sharp.
12. We are also able to “break duality” in Theorem 1.12 by handling some ranges of exponents $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} > 1$, at the cost of increasing r slightly in (1.14).
13. Estimates (1.13) and (1.14) imply the analogous estimates with $\tilde{A}_{N;X,\mathcal{T}}^{\mathcal{P}}$ in place of $A_{N;X,\mathcal{T}}^{\mathcal{P}}$ and vice versa. This follows by a simple telescoping argument; see [62, Proposition 3.2(iii)]. We will work with $\tilde{A}_{N;X,\mathcal{T}}^{\mathcal{P}}$ to avoid some technicalities.

14. In the proof of Theorem 1.12, we will use the Calderón transference principle [19], which reduces the estimates (1.13) and (1.14) in abstract measure-preserving systems $(X, \mathcal{B}(X), \mu)$ to those for the integer shift system, which is σ -finite; see Example 1.18. This is the reason why we formulate Theorem 1.12 for σ -finite measure spaces. However, from the point of view of applications in combinatorics or elsewhere, where the statistical properties of the ergodic averages (1.10) for totally ergodic systems matter, the general setting of σ -finite measure spaces is not interesting (since the limits are zero when $\mu(X) = \infty$) and only finite measure spaces are important.
15. A continuous analogue of the “sqorners” averages (1.9) was studied by Christ, Durcik, Kovač, and Roos in [21] and pointwise convergence for these averages was established. The key estimate is a Sobolev smoothing bound from an earlier paper of Christ, Durcik, and Roos [22, Theorem 5], which is a special instance of our multilinear Weyl inequality Theorem 6.1 in the real setting $\mathbb{K} = \mathbb{R}$. Theorem 6.1 establishes this key smoothing inequality for general k -linear polynomial corner averages in the distinct degree case. We also refer to a recent paper by Hsu and Yu-Hsiang Lin [52, Proposition 1.2], which provides a short and elementary proof of the smoothing inequality from [22, Theorem 5]. In [22], Christ, Durcik, and Roos use their smoothing inequality to prove maximal estimates and quantitative nonlinear Roth-type theorems for sqorn configurations in $\mathbb{R}^2$. The same applications for distinct degree polynomial corner configurations in $\mathbb{R}^k$ can be derived from our general smoothing inequality Theorem 6.1.

1.5. The Furstenberg–Bergelson–Leibman conjecture. Both Theorem 1.3 and Theorem 1.12 yield progress on the Furstenberg–Bergelson–Leibman conjecture, which asserts the following.

Conjecture 1.15 (The Furstenberg–Bergelson–Leibman conjecture). *Let $d, k \in \mathbb{Z}_+$ and $(X, \mathcal{B}(X), \mu)$ be a probability space equipped with a family of invertible measure-preserving transformations $T_1, \dots, T_d: X \rightarrow X$ generating a nilpotent group. Consider any polynomials $P_{1,1}, \dots, P_{i,j}, \dots, P_{d,k} \in \mathbb{Z}[n]$ with integer coefficients. Then, for any functions $f_1, \dots, f_k \in L^\infty(X)$, the multilinear polynomial ergodic averages*

$$\mathbb{E}_{n \in [N]} \prod_{j \in [k]} f_j \left(T_1^{P_{1,j}(n)} \dots T_d^{P_{d,j}(n)} x \right), \quad x \in X, \quad (1.16)$$

converge pointwise almost everywhere as $N \rightarrow \infty$.

We give a few remarks about Conjecture 1.15.

1. This conjecture is a very challenging problem in pointwise ergodic theory and modern harmonic analysis. It had been promoted by Bergelson in [6, Question 9, p. 52] and in [7, Section 6, p. 838], and in person by Furstenberg (see Austin’s article [2, p. 6662]) before it was published by Bergelson and Leibman [9, Section 5.5, p. 468].
2. Bergelson’s Question 1.1 is a special case of Conjecture 1.15. Indeed, if $d = k$, $P_{j,j} = P_j \in \mathbb{Z}[n]$, $P_{i,j} \equiv 0$ whenever $i \neq j$, and the transformations $T_1, \dots, T_k$ commute, then the averages from (1.16) coincide with the averages in (1.2). Hence, Theorem 1.3 constitutes progress on Conjecture 1.15 in the commutative case for arbitrary polynomials with distinct degrees.

3. The polynomial Szemerédi theorem of Bergelson and Leibman [8] was proved with the averages (1.16) in place of (1.5) for invertible commuting measure-preserving transformations $T_1, \dots, T_d$. To be more precise, if $P_{1,1}, \dots, P_{i,j}, \dots, P_{d,k} \in \mathbb{Z}[n]$ are polynomials with vanishing constant terms, then for any $A \in \mathcal{B}(X)$ satisfying $\mu(A) > 0$, they show that

$$\liminf_{N \rightarrow \infty} \mathbb{E}_{n \in [N]} \mu \left(\bigcap_{j \in [k]} T_1^{-P_{1,j}(n)} \dots T_d^{-P_{d,j}(n)}(A) \right) > 0. \quad (1.17)$$

The most general polynomial Szemerédi theorem for commuting transformations and IP-sets to date is due to Bergelson and McCutcheon [10]. We also refer to [18, Theorem 6.13] for a variant of (1.17) with the uniform averages $\mathbb{E}_{n \in [N] \setminus [M]}$ in place of $\mathbb{E}_{n \in [N]}$. In [9], Bergelson and Leibman began the study of recurrence theorems in the nilpotent setting, which triggered the formulation of Conjecture 1.15.

4. The Furstenberg–Bergelson–Leibman conjecture initially asked about convergence for (1.16) both in $L^2(X)$ and pointwise almost everywhere. As mentioned previously, Conjecture 1.15 in the context of $L^2(X)$ norm convergence for (1.16) was established by Walsh [95]. Zorin-Kranich [99] also established $L^2(X)$ norm convergence for (1.16) with the uniform averages $\mathbb{E}_{n \in [N] \setminus [M]}$ in place of $\mathbb{E}_{n \in [N]}$.
5. Bergelson and Leibman [9] showed that convergence for (1.16) may fail if $T_1, \dots, T_d$ generate a solvable group that is not nilpotent. This suggests that the nilpotent setting is the appropriate one for this conjecture.
6. Conjecture 1.15 in a genuinely nilpotent (i.e., of step at least two) setting is widely open, except one case. Recently, the second author along with Ionescu, Magyar, and Szarek [54] proved this conjecture with $d \in \mathbb{Z}_+$ and $k = 1$ for arbitrary polynomials $P_{1,1}, \dots, P_{d,1} \in \mathbb{Z}[n]$ and arbitrary invertible measure-preserving transformations $T_1, \dots, T_d$ on a σ -finite measure space $(X, \mathcal{B}(X), \mu)$ that generate a nilpotent group of step two; see also [55, 72] for some special cases of [54].

1.6. Reduction to the integer shift system. In pointwise convergence problems, the most important dynamical system is the integer shift system.

Example 1.18. Consider the k -dimensional lattice $(\mathbb{Z}^k, \mathcal{B}(\mathbb{Z}^k), \mu_{\mathbb{Z}^k})$, where $\mathcal{B}(\mathbb{Z}^k)$ is the σ -algebra of all subsets of $\mathbb{Z}^k$ and $\mu_{\mathbb{Z}^k}$ is counting measure on $\mathbb{Z}^k$, equipped with the standard family of shifts $S_1, \dots, S_k: \mathbb{Z}^k \rightarrow \mathbb{Z}^k$. More precisely, let $S_i(x) := x - e_i$ for all $x \in \mathbb{Z}^k$, where e_i is the i -th standard basis vector for each $i \in [k]$. The average $A_{N;X,T_1,\dots,T_k}^{P_1,\dots,P_k}$ from (1.10) with the family $\{S_1, \dots, S_k\}$ in place of $\{T_1, \dots, T_k\}$ can be rewritten as

$$A_{N;\mathbb{Z}^k}^{P_1,\dots,P_k}(f_1, \dots, f_k)(x) = \mathbb{E}_{n \in [N]} \prod_{i \in [k]} f_i(x - P_i(n)e_i), \quad x \in \mathbb{Z}^k. \quad (1.19)$$

Its truncated variant (1.11) takes the form

$$\tilde{A}_{N;\mathbb{Z}^k}^{P_1,\dots,P_k}(f_1, \dots, f_k)(x) = \mathbb{E}_{n \in [N] \setminus [N/2]} \prod_{i \in [k]} f_i(x - P_i(n)e_i), \quad x \in \mathbb{Z}^k. \quad (1.20)$$

We will often abbreviate $A_{N;\mathbb{Z}^k}^{P_1,\dots,P_k}$ to $A_{N;\mathbb{Z}^k}^{\mathcal{P}}$ and $\tilde{A}_{N;\mathbb{Z}^k}^{P_1,\dots,P_k}$ to $\tilde{A}_{N;\mathbb{Z}^k}^{\mathcal{P}}$.

In view of the Calderón transference principle [19] (or, more precisely, following the argument from [62, Proposition 3.2(ii)] or [60, Theorem 1.6]), it will suffice to work with the integer shift system and establish (1.14) with $A_{N;\mathbb{Z}^k}^{\mathcal{P}}$ in place of $A_{N;X,\mathcal{T}}^{\mathcal{P}}$. This will allow us to utilize the algebraic structure of $\mathbb{Z}^k$ and employ Fourier methods on $\mathbb{Z}^k$, which are not available in abstract measure preserving systems in general.

The Calderón transference principle allows us to transfer the quantitative estimates (1.13) and (1.14) from the integer shift system to corresponding estimates for $A_{N;X,\mathcal{T}}^{\mathcal{P}}$ in abstract measure-preserving systems $(X, \mathcal{B}(X), \mu)$. We emphasize that the Calderón transference principle [19] only transfers quantitative bounds that imply pointwise almost everywhere convergence, but does not transfer pointwise almost everywhere convergence itself. In fact, in the integer shift system, pointwise convergence is implied by norm convergence, since the $\ell^\infty(\mathbb{Z}^k)$ norm is dominated by the $\ell^2(\mathbb{Z}^k)$ norm. Hence, we will only be concerned with proving quantitative bounds for $A_{N;\mathbb{Z}^k}^{\mathcal{P}}$ or $\tilde{A}_{N;\mathbb{Z}^k}^{\mathcal{P}}$ in the integer shift system, not pointwise convergence on $\mathbb{Z}^k$. For technical reasons we will only work with the truncated averages $\tilde{A}_{N;\mathbb{Z}^k}^{\mathcal{P}}$.

After these reductions, our main result reads as follows.

Theorem 1.21. *Let $k \in \mathbb{Z}_+$ and $\mathcal{P} = \{P_1, \dots, P_k\} \subset \mathbb{Z}[n]$ be a family of polynomials with distinct degrees. Let $f_i \in \ell^{p_i}(\mathbb{Z}^k)$ for $i \in [k]$ and some $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$, and let $\tilde{A}_{N;\mathbb{Z}^k}^{\mathcal{P}}(f_1, \dots, f_k)$ be the averages defined in (1.20). Fix $r > 2$ and $\lambda > 1$. Then one has*

$$\left\| \mathbf{V}^r \left(\tilde{A}_{N;\mathbb{Z}^k}^{\mathcal{P}}(f_1, \dots, f_k) : N \in \mathbb{D} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim \prod_{i \in [k]} \|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)}, \quad (1.22)$$

if $\mathbb{D} = \{\lambda_n : n \in \mathbb{N}\} \subset [1, +\infty)$ is λ -lacunary. The same result holds with the averages $A_{N;\mathbb{Z}^k}^{\mathcal{P}}(f_1, \dots, f_k)$, defined in (1.19), in place of $\tilde{A}_{N;\mathbb{Z}^k}^{\mathcal{P}}(f_1, \dots, f_k)$.

Arguing as in [62, Proposition 3.2(ii)–(iii)], one sees that Theorem 1.21 implies Theorem 1.12. Thus, it suffices to prove Theorem 1.21.

To establish the r -variational inequality in (1.22), we will develop a new robust method that we call the *multilinear circle method*. This method can be viewed as a classical Hardy–Littlewood–Ramanujan circle method in $\ell^p(\mathbb{Z}^k)$ spaces. A bilinear variant of the circle method was recently developed by the second author with Krause and Tao [62] in the context of pointwise convergence for the Furstenberg–Weiss averages $\mathbb{E}_{n \in [N]} f_1(T^n x) f_2(T^{P(n)} x)$ for any $P \in \mathbb{Z}[n]$ with $\deg P \geq 2$. The arguments from [62] are limited to the Furstenberg–Weiss averages. Here we can handle genuinely multilinear averages involving polynomials with distinct degrees and arbitrary commuting transformations. Although the arguments in this paper are inspired by those in [62], we will present a conceptually different approach to overcome several new difficulties that arise in the more general setting. The key tools that we develop to prove inequality (1.22) that make up our multilinear circle method are the following.

1. An Ionescu–Wainger multiplier theorem for the set of canonical fractions; see Theorem 3.3. This is a multifrequency multiplier theorem, which we prove in Section 3 for the set of *canonical fractions* (see (1.23) for their definition), giving a positive

- answer to a question of Ionescu and Wainger from [56, Remark 3, p. 361]. This new tool will enable us to implement the circle method directly in $\ell^p(\mathbb{Z}^k)$ spaces.
2. *An inverse theorem for averages over distinct degree polynomial corner configurations*; see Theorem 4.14. The inverse theorem is a powerful result in additive combinatorics, offering important structural information that reveals the minor and major arc structure for the multilinear operators $\hat{A}_{N;\mathbb{Z}^k}^{\mathcal{P}}(f_1, \dots, f_k)$. We prove it in Section 4 in both the integer and real settings, and it constitutes a joint generalization of an inverse theorem of the third author [80] for averages over distinct degree polynomial progressions and of the third author along with Prendiville and Shao [83] for averages over the two-dimensional “sqorners” configuration $(x_1, x_2), (x_1 + n, x_2), (x_1, x_2 + n^2)$.
 3. *A multilinear $\ell^p(\mathbb{Z}^k)$ -improving inequality*; see Theorem 5.2. This is a new inequality in the polynomial corners setting that will allow us to relax the $\ell^\infty(\mathbb{Z}^k)$ bounds that arise in the inverse theorem to $\ell^p(\mathbb{Z}^k)$ bounds, and to handle genuinely multilinear cases for $k \geq 3$. We prove this inequality in Section 5 in both the integer and real settings.
 4. *Multilinear Weyl and Sobolev smoothing inequalities*; see Theorem 6.1. The multilinear Weyl inequality (1.36) is a key tool to control the minor arc contribution in our multilinear circle method. The multilinear Sobolev smoothing inequality will be used to understand the major arc contribution. These inequalities are proved in Section 6, and are sometimes referred to as *smoothing estimates* in the literature. A bilinear Weyl inequality was recently proved in [62] as a consequence of the inverse theorem of the third author [80]. However, the bounds obtained in [62] are logarithmic in scale, which is insufficient for the methods in this paper. Here, we make quantitative improvements to the bilinear estimates and derive a multilinear Weyl inequality with polynomial bounds that are consistent with the bounds in the classical Weyl inequality for exponential sums. This is possible thanks to our Ionescu–Wainger multiplier theorem for the set of canonical fractions, inverse theorem for polynomial corner configurations, and multilinear $\ell^p(\mathbb{Z}^k)$ -improving inequality.

Finally, in Section 7, we will use these tools to develop the multilinear circle method in the context of Theorem 1.21. An important new feature of our argument is that we do not require p -adic methods, which were employed in [62]. This answers a question of Magyar [70], who asked whether the use of p -adic and adelic harmonic analysis is necessary in [62].

1.7. The multilinear circle method. The details of the multilinear circle method in the context of Theorem 1.21 will be presented in Section 7. We now briefly describe its key features, starting with a few basic concepts from the classical circle method.

1.7.1. A first glimpse of the circle method. The use of the classical circle method to attack pointwise convergence problems for linear ergodic averages with polynomial orbits originates in Bourgain’s papers [12, 13, 14], and can be summarized as follows:

- (a) To control the minor arc contribution, we apply Plancherel’s theorem and Weyl’s inequality for exponential sums.
- (b) To control the major arc contribution, we use multifrequency harmonic analysis in the spirit of the Ionescu–Wainger multiplier theorem; see Theorem 3.3 and Theorem 3.26.

We now illustrate the ideas from (a) and (b) in the context of r -variational estimates (1.22) in the linear case $k = 1$ for finitely supported $f \in \ell^2(\mathbb{Z})$. This will highlight the differences between the classical circle method and the multilinear circle method.

Let $P \in \mathbb{Z}[n]$ be a polynomial of degree $d \geq 2$. For the sake of discussion we will work with $A_{N;\mathbb{Z}}^P$ instead of its truncation. Using the Fourier transform, we note that

$$\mathcal{F}_{\mathbb{Z}} A_{N;\mathbb{Z}}^P(f)(\xi) = m_N(\xi) \mathcal{F}_{\mathbb{Z}} f(\xi), \quad \xi \in \mathbb{T},$$

where the multiplier m_N is the exponential sum

$$m_N(\xi) := \mathbb{E}_{n \in [N]} e(\xi P(n)).$$

The classical circle method can be used to understand the nature of the multiplier m_N . This will require the concepts of canonical fractions and their corresponding major arcs. Given $N_1, N_2 \in \mathbb{R}_+$, we define the set of *canonical fractions* by

$$\mathcal{R}_{\leq N_1} := \left\{ \frac{a}{q} \in \mathbb{Q}/\mathbb{Z} : q \in [N_1] \text{ and } \gcd(a, q) = 1 \right\}, \quad (1.23)$$

where $\frac{a}{q}$ represents the equivalence class $\frac{a}{q} + \mathbb{Z}$, and the corresponding set of *major arcs* by

$$\mathfrak{M}_{\leq N_2}(\mathcal{R}_{\leq N_1}) := \bigcup_{\theta \in \mathcal{R}_{\leq N_1}} [\theta - N_2, \theta + N_2].$$

The set of *minor arcs* is then defined as the complement of the set of major arcs in $\mathbb{T}$. We formulate Weyl's estimate for the multiplier $m_N(\xi)$ as follows: for every $C \in \mathbb{R}_+$ there exists a small constant $c \in (0, 1)$ such that we have

$$\xi \in \mathbb{T} \setminus \mathfrak{M}_{\leq N-d\delta^{-C}}(\mathcal{R}_{\leq \delta^{-C}}) \implies |m_N(\xi)| \leq c^{-1}(\delta^c + N^{-c}) \quad (1.24)$$

for all $N \geq 1$ and $\delta \in (0, 1]$. In fact, (1.24) is the classical Weyl sum estimate for normalized exponential sums; see, for example, [91, Exercise 1.1.21, p. 16].

If $N \geq 1$ is sufficiently large in terms of δ , say $N > 3\delta^{-3C}$, then the intervals that comprise the set of major arcs are narrow and disjoint in $\mathbb{T}$. Taking a smooth even cutoff function $\eta: \mathbb{R} \rightarrow [0, 1]$ such that $\mathbb{1}_{[-1/4, 1/4]} \leq \eta \leq \mathbb{1}_{(-1/2, 1/2)}$, we define for $N_1, N_2 \in \mathbb{R}_+$ the smooth projection operators $\Pi[\leq N_1, \leq N_2]: \ell^2(\mathbb{Z}) \rightarrow \ell^2(\mathbb{Z})$ by setting

$$\mathcal{F}_{\mathbb{Z}}(\Pi[\leq N_1, \leq N_2]f)(\xi) := \sum_{\theta \in \mathcal{R}_{\leq N_1}} \eta(N_2^{-1}(\xi - \theta)) \mathcal{F}_{\mathbb{Z}} f(\xi). \quad (1.25)$$

These projections will be called the Ionescu–Wainger projections and will allow us to effectively localize to the major arcs. Their boundedness properties will be extensively studied in Section 3 in a more general context. By (1.25) and Plancherel's theorem, the operator $\Pi[\leq \delta^{-C}, \leq N^{-d}\delta^{-C}]$ is a contraction on $\ell^2(\mathbb{Z})$ provided that the arcs in $\mathfrak{M}_{\leq N-d\delta^{-C}}(\mathcal{R}_{\leq \delta^{-C}})$ are disjoint. We also see that $\Pi[\leq \delta^{-C}, \leq N^{-d}\delta^{-C}]$ is bounded on $\ell^p(\mathbb{Z})$ for all $p \in [1, \infty]$, since

$$\left\| \Pi[\leq \delta^{-C}, \leq N^{-d}\delta^{-C}]f \right\|_{\ell^p(\mathbb{Z})} \lesssim \#\mathcal{R}_{\leq \delta^{-C}} \|f\|_{\ell^p(\mathbb{Z})} \leq \delta^{-2C} \|f\|_{\ell^p(\mathbb{Z})}. \quad (1.26)$$

This bound is not very useful, but much better bounds will follow from the Ionescu–Wainger multiplier theorem proved in Section 3; see Theorem 3.3.

Taking $\delta = 4^{1/C} N^{-\varepsilon}$ with $\varepsilon \in (0, 1)$ sufficiently small, say $\varepsilon C < 1/4$, and N sufficiently large (so that, in particular, $\delta \leq 1$), we gain a negative power of N in (1.24) and this, combined with Plancherel's theorem, yields

$$\left\| A_{N;\mathbb{Z}}^P \left(f - \Pi \left[\leq N^{C\varepsilon}, \leq N^{-d+C\varepsilon} \right] f \right) \right\|_{\ell^2(\mathbb{Z})} \lesssim N^{-c\varepsilon} \|f\|_{\ell^2(\mathbb{Z})}, \quad (1.27)$$

since $\mathcal{F}_{\mathbb{Z}}(f - \Pi[\leq N^{C\varepsilon}, \leq N^{-d+C\varepsilon}]f)$ vanishes on the major arcs

$$\mathfrak{M}_{\leq N^{-d+C\varepsilon}/4}(\mathcal{R}_{\leq N^{C\varepsilon}}) \supseteq \mathfrak{M}_{\leq N^{-d}\delta^{-C}}(\mathcal{R}_{\leq \delta^{-C}}).$$

If we use (1.27) and the fact that $N \in \mathbb{D}$, where $\mathbb{D}$ is lacunary, inequality (1.22) for $k = 1$ and $p = 2$ is reduced to proving the following inequality

$$\left\| \mathbf{V}^r \left(A_{N;\mathbb{Z}}^P \left(\Pi \left[\leq N^{C\varepsilon}, \leq N^{-d+C\varepsilon} \right] f \right) : N \in \mathbb{D} \right) \right\|_{\ell^2(\mathbb{Z})} \lesssim \|f\|_{\ell^2(\mathbb{Z})}. \quad (1.28)$$

To estimate (1.28), we split $\Pi[\leq N^{C\varepsilon}, \leq N^{-d+C\varepsilon}]$ dyadically into pieces corresponding to reduced fractions $\theta = \frac{a}{q}$ with $q \simeq 2^l$. Then, for a major arc frequency $\xi \in \mathbb{T}$ such that $\eta(N^{d-C\varepsilon}(\xi - \theta)) \neq 0$, we approximate $m_N(\xi)$ by a product $G(\theta)\mathfrak{m}_N(\xi - \theta)$ of arithmetic and continuous symbols given by

$$G(\theta) := \mathbb{E}_{n \in [q]} e(\theta P(n)) \quad \text{for } \theta = \frac{a}{q} \text{ with } q \simeq 2^l,$$

and

$$\mathfrak{m}_N(\xi) := \int_0^1 e(\xi P(Nt)) dt \quad \text{for } \xi \in \mathbb{T},$$

respectively. After a further factorization, the arithmetic part is summable over $l \in \mathbb{N}$ because $|G(\theta)| \lesssim q^{-c} \lesssim 2^{-cl}$, whereas the continuous part can be handled by Theorem 3.26. This completes the outline of the proof of (1.28).

1.7.2. Weyl's inequality in $\ell^p(\mathbb{Z})$ spaces and basic Ionescu–Wainger theory. The argument presented above for estimating (1.22) when $k = 1$ and $f \in \ell^2(\mathbb{Z})$ can be extended to $f \in \ell^p(\mathbb{Z})$ for all $p \in (1, \infty)$. However, there are various challenges: for example, it is not clear how to make use of Weyl's inequality, as Plancherel's theorem is not available in $\ell^p(\mathbb{Z})$ when $p \neq 2$. A way to overcome this difficulty is to proceed as follows:

- (i) One works directly with the Ionescu–Wainger projections from (1.25), as they are bounded on $\ell^p(\mathbb{Z})$ for $p \in (1, \infty)$ and localized to the major arcs.
- (ii) Instead of working with the exponential sum m_N itself, one works with the corresponding averaging operator $A_{N;\mathbb{Z}}^P$.

The $\ell^p(\mathbb{Z})$ norms of the Ionescu–Wainger projections $\Pi[\leq \delta^{-C}, \leq N^{-d}\delta^{-C}]$ have reasonably good growth in terms of the size of the set of canonical fractions $\mathcal{R}_{\leq \delta^{-C}}$ due to our Ionescu–Wainger multiplier theorem. Namely, for any $p \in (1, \infty)$ and $\rho \in (0, 1)$, whenever

$$N > C_p \delta^{-C_p} \quad (1.29)$$

for some large $C_p \in \mathbb{R}_+$, it follows from Theorem 3.3 that

$$\left\| \Pi \left[\leq \delta^{-C}, \leq N^{-d}\delta^{-C} \right] f \right\|_{\ell^p(\mathbb{Z})} \lesssim_{p,\rho} \delta^{-C\rho} \|f\|_{\ell^p(\mathbb{Z})}. \quad (1.30)$$

This bound represents a significant quantitative improvement compared to (1.26) and will be proved in Section 3 as a part of our Ionescu–Wainger multifrequency multiplier theory.

Below are a few comments about this inequality.

1. In an impactful and influential paper, Ionescu and Wainger [56] established a deep multiplier theorem for the set of so-called *Ionescu–Wainger fractions* (see (1.31) below for their definition), in order to establish $\ell^p(\mathbb{Z}^k)$ bounds for discrete singular integral Radon transforms. The Ionescu–Wainger multiplier theorem quickly became the main tool in the study of discrete analogues in harmonic analysis [73, 75].
2. The set of Ionescu–Wainger fractions is defined as

$$\tilde{\mathcal{R}}_{\leq N} := \left\{ \frac{a}{q} \in \mathbb{Q}/\mathbb{Z} : q \in P_{\leq N} \text{ and } \gcd(a, q) = 1 \right\}, \quad (1.31)$$

where $P_{\leq N}$ is a subtle set of natural numbers with certain prime power factorizations. In particular, the original Ionescu–Wainger theory [56] implies inequality (1.30), as well as the conclusion of Theorem 3.3, for the projections defined as in (1.25), with the set of Ionescu–Wainger fractions $\tilde{\mathcal{R}}_{\leq \delta^{-C}}$ in place of the set of canonical fractions $\mathcal{R}_{\leq \delta^{-C}}$, provided that, instead of condition (1.29), we assume that

$$\log N \gtrsim \delta^{-C\gamma} \quad (1.32)$$

holds for some arbitrarily small $\gamma \in (0, 1)$.

3. Ionescu–Wainger theory [56] was originally developed for scalar-valued multipliers. An important aspect of the theory was that the $\ell^p(\mathbb{Z})$ norm of the Ionescu–Wainger multipliers corresponding to the set of Ionescu–Wainger fractions $\tilde{\mathcal{R}}_{\leq \delta^{-C}}$ was controlled by a multiple of $\log(\delta^{-C} + 1)^D$, where $D := \lfloor 2\gamma^{-1} \rfloor + 1$. The proof in [56] is based on an intricate inductive argument that takes advantage of superorthogonality phenomena. A slightly different proof with the factor $\log(\delta^{-C} + 1)$ in place of $\log(\delta^{-C} + 1)^D$ was given in [73]. The latter proof relied on certain recursive arguments instead of the induction used in [56]. This approach helped clarify the role of the underlying square functions and orthogonalities; see also [75, Section 2]. Ionescu–Wainger theory, among other topics, was discussed by Pierce [84] in the context of superorthogonality phenomena. Finally, we refer the reader to the recent paper of Tao [92], where a uniform bound in place of $\log(\delta^{-C} + 1)$ was obtained.
4. The central question about Ionescu–Wainger theory for the set of canonical fractions from (1.23) had remained unanswered since it was first formulated in [56, Remark 3, p. 361]. In Section 3, the question is answered affirmatively.

For any $p \in (1, \infty)$ and $\rho \in (0, 1)$, if $N > C_p \delta^{-C_p}$, then (1.30) implies

$$\left\| A_{N;\mathbb{Z}}^P \left(f - \Pi \left[\leq \delta^{-C}, \leq N^{-d} \delta^{-C} \right] f \right) \right\|_{\ell^p(\mathbb{Z})} \lesssim_{p,\rho} \delta^{-C\rho} \|f\|_{\ell^p(\mathbb{Z})}. \quad (1.33)$$

For $p = 2$, by Plancherel’s theorem and (1.24), we have the stronger bound

$$\left\| A_{N;\mathbb{Z}}^P \left(f - \Pi \left[\leq \delta^{-C}, \leq N^{-d} \delta^{-C} \right] f \right) \right\|_{\ell^2(\mathbb{Z})} \lesssim (\delta^c + N^{-c}) \|f\|_{\ell^2(\mathbb{Z})}. \quad (1.34)$$

Thus, for any $p \in (1, \infty)$, by interpolating (1.33) and (1.34), we obtain that there exist constants $C'_p \in \mathbb{Z}_+$ and $c_p \in (0, 1)$ such that if $N > C'_p \delta^{-C'_p}$, then

$$\left\| A_{N;\mathbb{Z}}^P \left(f - \Pi \left[\leq \delta^{-C}, \leq N^{-d} \delta^{-C} \right] f \right) \right\|_{\ell^p(\mathbb{Z})} \lesssim_p (\delta^{c_p} + N^{-c_p}) \|f\|_{\ell^p(\mathbb{Z})}. \quad (1.35)$$

Hence, inequality (1.35) can be thought of as a generalization of Weyl's inequality for exponential sums (1.24) to $\ell^p(\mathbb{Z})$ spaces for all $p \in (1, \infty)$. Here, it is important that the bounds in (1.35) are consistent with the bounds in the classical Weyl inequality for exponential sums (1.24). An essential new tool that makes these estimates available is the Ionescu–Wainger multiplier theorem for the set of canonical fractions. Thus, Ionescu–Wainger theory should be understood as a tool that enables us to interpret exponential sum estimates in terms of purely functional analytic language. This interpretation is very useful in the context of the multilinear Weyl inequality stated in (1.36) below.

1.7.3. The multilinear Weyl inequality and the minor arc contribution. In the multilinear setting (when $k \geq 2$), Plancherel's theorem and Weyl's inequality (1.24) are no longer viable to effectively control the contribution from the minor arcs. As a result, it is not possible to implement the classical circle method as we discussed above.

However, in view of the discussion about the generalization of Weyl's inequality to $\ell^p(\mathbb{Z})$ spaces, it is natural to formulate the multilinear version using functional analysis. Namely, we expect that if $\mathcal{P} = \{P_1, \dots, P_k\} \subset \mathbb{Z}[n]$ is such that the degrees $d_i := \deg P_i$ are distinct, and if $1 < p_1, \dots, p_k < \infty$ are such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$, then for each fixed $C \in \mathbb{R}_+$ the following holds with some small $c \in (0, 1)$ possibly depending on $k, \mathcal{P}, p_1, \dots, p_k, p, C$. For every $f_i \in \ell^{p_i}(\mathbb{Z}^k)$ for all $i \in [k]$ and for all $N \geq 1$ and $\delta \in (0, 1]$, if $f_j \in \ell^2(\mathbb{Z}^k) \cap \ell^{p_j}(\mathbb{Z}^k)$ for some $j \in [k]$ and the j -th Fourier transform $\mathcal{F}_{j,\mathbb{Z}^k} f_j$ vanishes on the major arcs $\mathfrak{M}_{\leq N^{-d_j} \delta^{-C}}^j(\mathcal{R}_{\leq \delta^{-C}})$, then

$$\left\| A_{N;\mathbb{Z}^k}^{P_1, \dots, P_k}(f_1, \dots, f_k) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim (\delta^c + N^{-c}) \prod_{i \in [k]} \|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)}. \quad (1.36)$$

Here, $\mathcal{F}_{j,\mathbb{Z}^k}$ and $\mathfrak{M}_{\leq N^{-d_j} \delta^{-C}}^j(\mathcal{R}_{\leq \delta^{-C}})$ denote, respectively, the Fourier transform and the set of major arcs defined with respect to the j -th variable. Inequality (1.36) is a multilinear variant of inequality (1.35) and will be referred to as the *multilinear Weyl inequality* throughout the paper. It will play a key role in facilitating our *multilinear circle method*, namely:

multilinear Weyl's inequality $\longleftrightarrow$ control of the minor arc contribution.

The proof of inequality (1.36), as well as its real counterpart, the multilinear Sobolev smoothing inequality, is given in Section 6; see Theorem 6.1.

We now describe the key components of our *multilinear circle method* in the context of establishing Theorem 1.21. We begin by explaining how the multilinear Weyl inequality from (1.36) controls the contribution from the minor arcs. If we take $\delta \simeq N^{-\varepsilon}$ with $\varepsilon \in (0, 1)$ sufficiently small, say $\varepsilon C < 1/4$, and N sufficiently large, then we gain a negative power of

N in (1.36). Hence, if we split $f_i = f_{i,N}^0 + f_{i,N}^1$ for each $i \in [k]$, where

$$\begin{aligned} f_{i,N}^0 &:= f_i - \Pi \left[\leq N^{C\varepsilon}, \leq N^{-d_i+C\varepsilon} \right] f_i, \\ f_{i,N}^1 &:= \Pi \left[\leq N^{C\varepsilon}, \leq N^{-d_i+C\varepsilon} \right] f_i, \end{aligned}$$

then, in view of (1.36) and from the fact that $N \in \mathbb{D}$, where $\mathbb{D}$ is lacunary, inequality (1.22) is reduced to proving that

$$\left\| \mathbf{V}^r \left(A_{N;\mathbb{Z}^k}^{P_1, \dots, P_k} (f_{1,N}^1, \dots, f_{k,N}^1) : N \in \mathbb{D} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim \prod_{i \in [k]} \|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)}, \quad (1.37)$$

where $f_{1,N}^1, \dots, f_{k,N}^1$ are supported on major arcs. Here, in this informal discussion, we work with the original average (1.19) instead of its truncation (1.20). The idea of the proof of inequality (1.37) will be discussed momentarily.

Our multilinear Weyl inequality with polynomial bounds (1.36) is a new tool that we will also use in the analysis of the major arc contribution. A bilinear Weyl inequality was recently proved in [62] for the Furstenberg–Weiss averages as a consequence of an inverse theorem of the third author [80]. Namely, under suitable assumptions, for some sufficiently large $\tilde{C} \in \mathbb{Z}_+$, we have

$$\left\| \mathbb{E}_{n \in [N]} f_1(x - n) f_2(x - P(n)) \right\|_{\ell^p(\mathbb{Z})} \lesssim (\delta^c + \langle \log N \rangle^{-c\tilde{C}}) \prod_{i \in [2]} \|f_i\|_{\ell^{p_i}(\mathbb{Z})} \quad (1.38)$$

for any polynomial $P \in \mathbb{Z}[n]$ with degree at least two. However, these bounds are logarithmic in scale, and thus insufficient for the purpose of this paper. The logarithmic factor in the inequality (1.38) is a result of applying Ionescu–Wainger theory with the Ionescu–Wainger fractions (1.31). The construction of the Ionescu–Wainger fractions forces condition (1.32), which subsequently leads to the presence of the logarithmic factor in (1.38). In our approach, by applying Ionescu–Wainger theory with the set of canonical fractions (1.23) as developed in Section 3, we can quantitatively improve the bilinear estimates from (1.38) and obtain a multilinear Weyl inequality with polynomial bounds in (1.36) that match the bounds from the classical Weyl inequality for exponential sums (1.24). It is critical here that Ionescu–Wainger theory with the set of canonical fractions holds under condition (1.29), which enables us to obtain a negative power of N in (1.36).

Apart from Ionescu–Wainger theory for the set of canonical fractions, other essential tools needed to derive (1.36) include an inverse theorem for averages over polynomial corner configurations with polynomial bounds, and a multilinear $\ell^p(\mathbb{Z}^k)$ -improving inequality; see Theorem 4.14 and Theorem 5.2, respectively. In order to prove inequality (1.36), we first apply the inverse theorem, which reveals the major arc structure corresponding to the set of canonical fractions. By combining this theorem with the Hahn–Banach separation theorem, we can also reveal the major and minor arc structure for the adjoint operators corresponding to $A_{N;\mathbb{Z}^k}^{P_1, \dots, P_k}(f_1, \dots, f_k)$. Next, we use our version of Ionescu–Wainger theory to maintain the distinction between major and minor arcs and to preserve the polynomial bounds in terms of δ obtained from the inverse theorem throughout this process. Finally, the $\ell^p(\mathbb{Z}^k)$ -improving inequality is applied to relax the $\ell^\infty(\mathbb{Z}^k)$ bounds arising in the inverse theorem to $\ell^p(\mathbb{Z}^k)$

bounds that we need here. This ultimately implies the multilinear Weyl inequality with polynomial bounds (1.36), as desired.

1.7.4. The inverse theorem for averages over polynomial corner configurations. In 2019, the third author developed a new technique in higher order Fourier analysis that she used to prove power-saving bounds in the finite field version of the polynomial Szemerédi theorem for linearly independent polynomial progressions [79], answering a question of Bourgain and Chang [16]. This method, now called *degree lowering*, was improved and adapted to the integer setting by the third author and Prendiville [82], proving the first quantitative bounds for the size of subsets of $[N]$ lacking nontrivial copies of the *nonlinear Roth configuration* $x, x+n, x+n^2$. The third author then extended this result to all distinct degree polynomial progressions [80]. The key technical result of [80], from which quantitative bounds in the polynomial Szemerédi theorem follow by a standard density increment iteration, is an inverse theorem asserting that if averages of 1-bounded functions over fixed distinct degree polynomial progressions are large, then the underlying 1-bounded functions must correlate with functions whose Fourier transforms are supported on major arcs.

Theorem 1.39 (Peluse, 2020). *Let $k \in \mathbb{Z}_+$ and $P_1, \dots, P_k \in \mathbb{Z}[n]$ be polynomials with vanishing constant terms and distinct degrees $d_1 := \deg P_1 < \dots < d_k := \deg P_k$. Then there exist large constants $C_1, C_2 \in \mathbb{Z}_+$ such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C_1 \delta^{-C_1}$. If $f_0, f_1, \dots, f_k \in \ell^\infty(\mathbb{Z})$ are 1-bounded functions supported on $[\pm O(N^{d_k})]$ such that*

$$\left| \left\langle f_0, \mathbb{E}_{n \in [N]} \prod_{i \in [k]} f_i(x - P_i(n)) \right\rangle \right| \geq \delta N^{d_k},$$

then, using the projections from (1.25), one has

$$\left| \left\langle f_1, \Pi[\leq C_2 \delta^{-C_2}, \leq C_2 \delta^{-C_2} N^{-d_1}] f_1 \right\rangle \right| \geq C_1^{-1} \delta^{C_1} N^{d_k}.$$

When $k = 2$, one can obtain strong structural information about f_0 and f_2 from Theorem 1.39; see [81] for details. This structural information was then used by the second author along with Krause and Tao in [62] to prove their bilinear Weyl inequality for Furstenberg–Weiss averages (1.38). When $k > 2$, the argument from [81] breaks down, and so structural information cannot be deduced for $f_2, \dots, f_k$; see the discussion in the last section of [81].

In order to prove our multilinear Weyl inequality, we require an extension of Theorem 1.39 to higher dimensions that simultaneously produces strong structural information about each function appearing in the average. Such an inverse theorem is proved in Section 4 in both the integer and real settings; see Theorem 4.14. This inverse theorem simultaneously generalizes work of the third author [80] proving Theorem 1.39 and of the third author along with Prendiville and Shao proving an inverse theorem for averages over sqorners

$$(x_1, x_2), (x_1 + n, x_2), (x_1, x_2 + n^2).$$

1.7.5. The multilinear $\ell^p(\mathbb{Z}^k)$ -improving inequality. In Section 5, we establish the following multilinear $\ell^p(\mathbb{Z}^k)$ -improving inequality. For $k \geq 2$ and $j \in [k]$, there are exponents $1 < p <$

$q < 2$ and $1 < p_1, \dots, p_k < \infty$ satisfying $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p}$ and $p_j = 2$ such that the following scale-invariant bound

$$\|A_{N;\mathbb{Z}^k}^{\mathcal{P}}(f_1, \dots, f_k)\|_{\ell^q(\mathbb{Z}^k)} \lesssim_{\mathcal{P}} N^{-D(\frac{1}{p}-\frac{1}{q})} \prod_{i \in [k]} \|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)} \quad (1.40)$$

holds with $D := d_1 + \dots + d_k$. The bound (1.40) plays a key role in relaxing the $\ell^\infty(\mathbb{Z}^k)$ bounds required to apply the inverse theorem to $\ell^p(\mathbb{Z}^k)$ bounds. It is essential that our multilinear $\ell^p(\mathbb{Z}^k)$ -improving inequality is obtained without any ε -loss in the scale N . Otherwise, we could not derive the multilinear Weyl inequality with polynomial bounds in (1.36).

Such improving bounds have a long history in the continuous setting (with $\mathbb{R}$ in place of $\mathbb{Z}$) but it is only recently that they were established in the integer setting and then only for linear averages. In [62], the bounds (1.40) were needed for particular bilinear averages and were deduced from the linear case of Han–Kovač–Lacey–Madrid–Yang in [47] by passing to the dual operator. In the general case considered here, genuinely multilinear $\ell^p(\mathbb{Z}^k)$ -improving estimates are needed, and it is no longer possible to reduce matters to the linear case.

To prove (1.40) with no ε -loss, we marry Christ's *refinements method* [20] (which has proved very successful for L^p -improving estimates in the continuous setting; see [88], where multilinear L^p improving bounds are established for Radon-like transforms), adapted to the multilinear setting, with the Vinogradov mean value theorem [17, 96].

1.7.6. Major arcs contribution: paraproduct-type decomposition. The objective now is to understand estimate (1.37). Our aim is to reduce (1.37) to estimates for certain model operators, which will be more manageable. We may normalize each f_i so that $\|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)} = 1$. We proceed in two steps:

- (i) We begin with trimming the size of denominators. This means that each function $f_{i,N}^1 = \Pi[\leq N^{C\varepsilon}, \leq N^{-d_i+C\varepsilon}]f_i$ can be split into finitely many pieces $f_{i,N}^{1,l_i}$, where each piece corresponds to projections onto major arcs centered at fractions in $\mathcal{R}_{\leq N^{C\varepsilon}}$ whose denominators have size of order 2^{l_i} for some $l_i \in \mathbb{N}$ such that $2^{l_i} \leq N^{C\varepsilon}$. We have seen a similar reduction in the linear case above.
- (ii) Next, we perform a certain paraproduct decomposition of each $f_{i,N}^{1,l_i}$ to match the scales of the bumps in the definition of the Ionescu–Wainger projections. In other words, we replace $\eta(N^{d_i-C\varepsilon}(\xi-\theta))$ in the definition of $f_{i,N}^{1,l_i}$ by $\eta(N^{d_i}(\xi-\theta))$ for $\theta = \frac{a}{q}$ with $q \simeq 2^{l_i}$. This can be done by a simple telescoping argument, where we further split $f_{i,N}^{1,l_i}$ into finitely many highly oscillatory and nonoscillatory pieces $f_{i,N}^{1,l_i,s_i}$ for some $s_i \in \mathbb{N}$ such that $2^{s_i} \leq N^{C\varepsilon}$. If $s_i > 0$, then $f_{i,N}^{1,l_i,s_i}$ is highly oscillatory, which means that its i -th Fourier transform is supported on an annulus of size $N^{d_i}2^{-s_i}$ centered at $\theta = \frac{a}{q}$. Otherwise, if $s_i = 0$, then $f_{i,N}^{1,l_i,0}$ is nonoscillatory and corresponds to the bumps $\eta(N^{d_i}(\xi-\theta))$ for $\theta = \frac{a}{q}$ with $q \simeq 2^{l_i}$.

This reduction is standard, and it was also performed in [62]. By using this decomposition, inequality (1.37) will follow if we can show that, for all $l_1, s_1, \dots, l_k, s_k \in \mathbb{N}$ and for some

$c \in (0, 1)$, the following inequality holds

$$\left\| \mathbf{V}^r \left(A_{N; \mathbb{Z}^k}^{P_1, \dots, P_k} \left(f_{1,N}^{1, l_1, s_1}, \dots, f_{k,N}^{1, l_k, s_k} \right) : N \in \mathbb{D} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-c(l+s)}, \quad (1.41)$$

with $l := \max\{l_i : i \in [k]\}$ and $s := \max\{s_i : i \in [k]\}$, where each f_i is normalized so that $\|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)} = 1$ and $f_{i,N}^{1, l_i, s_i} \equiv 0$ if $N^{C\varepsilon} < 2^{\max\{l_i, s_i\}}$. Here our approach deviates significantly from [62]. We distinguish two cases:

- (i) the *high frequency* case $s \gtrsim l$;
- (ii) the *low frequency* case $l \gtrsim s$.

By standard approximations on the major arcs, inequality (1.41) can be reduced to showing

$$\left\| \mathbf{V}^r \left(A_{2^u; \mathbb{Z}^k}^{P_1, \dots, P_k} \left(f_{1,N}^{1, l_1, s_1}, \dots, f_{k,N}^{1, l_k, s_k} \right) : N \in \mathbb{D} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-c(l+s)}, \quad (1.42)$$

where $u := 100k(s+1)$ in the high frequency case, and $u := 100k(l+1)$ in the low frequency case. An important feature of this reduction is that the operator $A_{2^u; \mathbb{Z}^k}^{P_1, \dots, P_k}$ is independent of the parameter $N \in \mathbb{D}$.

1.7.7. The major arc contribution: high frequency case. In the high frequency case, the multifrequency problem becomes, in fact, a single frequency problem. By using Taylor expansions, we can separate the nonoscillatory pieces from the highly oscillatory pieces. The nonoscillatory pieces are handled by a seminorm variant of the Ionescu–Wainger theorem; see Theorem 3.26. This theorem is interesting in its own right, as it eliminates the need to consider small and large scales as done previously. The highly oscillatory pieces are controlled by the multilinear Sobolev smoothing inequality, which is a real variant of the multilinear Weyl inequality. This inequality allows us to gain a negative power of 2^s for the highly oscillatory pieces, which yields the desired decay in (1.42) in the high frequency case thanks to $s \gtrsim l$.

This is a novel aspect of our argument and shows that the Sobolev smoothing inequality and the multilinear Weyl inequality, as well as the inverse theorem in both the integer and real settings, should always be proved in tandem.

1.7.8. The major arc contribution: low frequency case. To control the low frequency case, we will use a multiparameter norm interchanging inequality (7.64) combined with the multilinear Weyl inequality (1.36). This is an important part of our argument, where the multilinear Weyl inequality is also useful for controlling the contribution from the major arcs. Precisely, it will allow us to gain a negative power of 2^l , which yields the desired decay in (1.42) in the low frequency case thanks to $l \gtrsim s$.

A multiparameter norm interchanging inequality, to a certain extent, linearizes maximal functions in a way that makes it possible to apply, for instance, the multilinear Weyl inequality or similar tools. By a simple interpolation, we can assume that $r = \infty$ in (1.42) and focus on maximal functions. Our approach, which relies on the multiparameter norm interchanging inequality, also eliminates the need to consider small and large scales.

Finally, we highlight that our argument heavily depends on the polynomial bounds in the multilinear Weyl inequality (1.36). In the bilinear case treated in [62], the parameter u from (1.42) is exponential in l , while in our case it is linear. Furthermore, the Weyl

inequality from [62] was proved with decay $\delta^c + \langle \log N \rangle^{-c\tilde{C}}$, which includes a logarithmic term that requires u to be exponential in l . However, thanks to our improvement in the Ionescu–Wainger theorem, our Weyl inequality holds with $\delta^c + N^{-c}$, allowing us to have a linear dependence between l and u . This is crucial for handling the low frequency case and arranging the sizes of supports in certain model operators in terms of u in such a way that their contribution to the norm of these model operators becomes negligible compared to the size of cl with c arising in (1.36). If u were exponential in l , then our method would not work.

1.8. Final remarks and open problems. We have given an outline of the proof of inequality (1.22), discussing the crucial tools and steps of our multilinear circle method. We wrap up the introduction by giving a few concluding remarks and open problems.

1. The multiparameter norm interchanging inequality approach allows us to avoid the p -adic methods used in [62] and significantly simplifies the argument. Moreover, the p -adic approach from [62] (especially sharp L^p -improving inequalities in the p -adic setting), even if adjusted to our context, would not be sufficient to handle the major arc estimates in the low frequency setting.

Problem 1.43. Is a p -adic analogue of Theorem 1.21 (or its maximal counterpart, which should be simpler) true?

2. Our Ionescu–Wainger theory is not uniform with respect to the size of the set of canonical fractions (1.23). However, based on Tao’s recent paper [92], we know that the Ionescu–Wainger multiplier theorem for the set of Ionescu–Wainger fractions (1.31) holds with uniform bounds.

Problem 1.44. Does our Ionescu–Wainger multiplier theorem hold with bounds independent of the size of the set of canonical fractions?

3. The methods outlined in this paper are robust and fairly versatile. Specifically, the input from harmonic analysis, such as the multilinear $\ell^p(\mathbb{Z}^k)$ -improving inequality or the proof of Theorem 1.21, does not rely on the assumption that the polynomials have distinct degrees. This assumption is only used in the proof of the inverse theorem and is not used anywhere else in the paper. In other words, if we treated Theorem 4.14 as a black box, we could derive all other conclusions from this paper without needing to assume that our polynomials have distinct degrees. With this in mind, we hope that the methods outlined in the paper will be used in other contexts to address similar questions.

Problem 1.45. Is Theorem 1.12 for the weighted operators

$$\sum_{n \in [\pm N] \setminus \{0\}} \frac{1}{n} \prod_{i \in [k]} f_i(T_i^{P_i(n)} x) \quad (1.46)$$

in place of the averages $A_{N;X,\mathcal{T}}^{\mathcal{P}}$ from (1.10) true? The operators from (1.46) are multilinear truncated singular integral variants of $A_{N;X,\mathcal{T}}^{\mathcal{P}}$. It is also interesting to know any $\ell^p(\mathbb{Z}^k)$ bounds for the full multilinear singular integral operators, i.e., the

operators corresponding to (1.46) with summation over $[\pm N] \setminus \{0\}$ replaced with $\mathbb{Z} \setminus \{0\}$.

4. The assumption that our polynomials have distinct degrees is only used in the proof of the inverse theorem, as it is essential to run the degree lowering argument in Section 4. However, the proof of Theorem 4.14 also demonstrates that the assumption that the polynomials have distinct degrees in the multilinear setting may have a similar impact as curvature does in the linear theory of maximal and singular Radon transforms in the discrete [73, 75] or Euclidean [87] settings. This is apparent when we start with large multilinear averages (4.15) and apply the PET induction scheme and the degree lowering argument and end up with an expression as in (4.76) that corresponds to modulated linear polynomial averages that can be handled by linear methods, which exploit the curvature. This interesting phenomenon requires further investigation. In view of this remark, it is natural to ask the following.

Problem 1.47. Is it possible to establish an inverse theorem like in Theorem 4.14 for linearly independent polynomials with polynomial bounds in δ ? That would be an important step toward proving Theorem 1.3 for linearly independent polynomials.

5. A well-known open problem in ergodic theory is to establish pointwise almost everywhere convergence of bilinear averages

$$\mathbb{E}_{n \in [N]} f_1(S^n x) f_2(T^n x), \quad x \in X,$$

corresponding to “corner” configurations, for two commuting measure-preserving transformations $S, T: X \rightarrow X$; see [32, Problem 19]. In the single transformation setting, the pointwise almost everywhere convergence of the trilinear averages

$$\mathbb{E}_{n \in [N]} f_1(T^n x) f_2(T^{2n} x) f_3(T^{3n} x), \quad x \in X,$$

corresponding to the arithmetic progressions of length four, is another well-known open problem; see [32, Problem 11]. Our methods break down in these cases, as the minor arc contributions are no longer negligible, and we are therefore unable to say anything in this direction. In the trilinear case, identities such as

$$e(\alpha x^2) \cdot e(-3\alpha(x-n)^2) \cdot e(3\alpha(x-2n)^2) \cdot e(-\alpha(x-3n)^2) = 1$$

speak to the presence of a *quadratic* modulation invariance, which would most likely require developing quadratic time–frequency analysis in the spirit of [65, 26]; this is out of reach at the moment. However, we think that our methods represent a significant step in understanding the Furstenberg–Bergelson–Leibman conjecture.

Problem 1.48. How useful are the methods presented in this paper in understanding Question 1.1 for linearly dependent polynomials?

ACKNOWLEDGMENTS

We warmly thank Vitaly Bergelson, Nikos Frantzikinakis, Ben Green, Noah Kravitz, Borys Kuca, Tomasz Z. Szarek, and Terence Tao for helpful comments on an earlier draft of this paper. The first author gratefully acknowledges the support of the Basque Center for Applied Mathematics, where he carried out his postdoctoral research under the supervision

of Renato Lucà and Luz Roncal, as well as the support of the Wrocław University of Science and Technology through the Academia Iuvenum membership. The second and third authors gratefully acknowledge the support and hospitality of the Institute for Advanced Study, which helped facilitate this collaboration. Finally, we thank the referees for careful reading of the manuscript and useful remarks that led to the improvement of the presentation.

2. NOTATION AND USEFUL TOOLS

We now set up notation that will be used throughout the paper.

2.1. Basic notation. The set of positive integers and the set of nonnegative integers will be denoted, respectively, by $\mathbb{Z}_+ := \{1, 2, \dots\}$ and $\mathbb{N} := \{0, 1, 2, \dots\}$. For $d \in \mathbb{Z}_+$ the sets $\mathbb{Z}^d$, $\mathbb{R}^d$, $\mathbb{C}^d$ and $\mathbb{T}^d := (\mathbb{R}/\mathbb{Z})^d$ have their standard meaning. We also denote $\mathbb{R}_+ := (0, \infty)$.

For $a = (a_1, \dots, a_d) \in \mathbb{Z}^d$ and $q \in \mathbb{Z}_+$, we denote by $\gcd(a, q)$ the greatest common divisor of a and q , i.e., the largest $n \in \mathbb{Z}_+$ that divides q and all components $a_1, \dots, a_d$. Clearly, any vector in $\mathbb{Q}^d$ has a unique representation as $\frac{a}{q}$ with $q \in \mathbb{Z}_+$, $a \in \mathbb{Z}^d$, and $\gcd(a, q) = 1$.

We use $\mathbb{1}_A$ to denote the indicator function of a set A . For a statement S we write $\mathbb{1}_S$ to denote its indicator, equal to 1 if S is true and 0 if S is false. In particular, we have $\mathbb{1}_A(x) = \mathbb{1}_{x \in A}$.

Throughout the paper $C > 0$ is an absolute constant that may change from occurrence to occurrence. For two nonnegative quantities A, B we write $A \lesssim B$ if there is an absolute constant $C > 0$ such that $A \leq CB$. We also write $A \simeq B$ when $A \lesssim B \lesssim A$. We will write $\lesssim_\delta$ or $\simeq_\delta$ to emphasize that the implicit constant depends on δ . For a function $f: X \rightarrow \mathbb{C}$ and a positive-valued function $g: X \rightarrow (0, \infty)$, we write $f = O(g)$ if there exists a constant $C > 0$ such that $|f(x)| \leq Cg(x)$ for all $x \in X$. We will also write $f = O_\delta(g)$ if the implicit constant depends on δ .

We will use the Japanese bracket notation

$$\langle z \rangle := (1 + |z|^2)^{1/2} \simeq 1 + |z|, \quad z \in \mathbb{C}.$$

For any $x \in \mathbb{R}$ we will use the floor and fractional part functions

$$\lfloor x \rfloor := \max\{n \in \mathbb{Z} : n \leq x\} \quad \text{and} \quad \{x\} := x - \lfloor x \rfloor.$$

All logarithms in this paper are taken base 2, and for any $N \geq 1$ we define

$$\text{Log } N := \lfloor \log N \rfloor, \tag{2.1}$$

i.e., the unique natural number such that $2^{\text{Log } N} \leq N < 2^{\text{Log } N + 1}$.

For every $N \in \mathbb{R}_+$ we set

$$\begin{aligned} [N]_{\mathbb{Z}} &:= (0, N] \cap \mathbb{Z}, & [\pm N]_{\mathbb{Z}} &:= [-N, N] \cap \mathbb{Z}, \\ [N]_{\mathbb{R}} &:= (0, N], & [\pm N]_{\mathbb{R}} &:= [-N, N]. \end{aligned} \tag{2.2}$$

We shall abbreviate $[N]_{\mathbb{Z}}$ to $[N]$ and $[\pm N]_{\mathbb{Z}}$ to $[\pm N]$. We will also write

$$\begin{aligned} \mathbb{N}_{\leq N} &:= [0, N] \cap \mathbb{N}, & \mathbb{N}_{\geq N} &:= [N, \infty) \cap \mathbb{N}, \\ \mathbb{N}_{< N} &:= [0, N) \cap \mathbb{N}, & \mathbb{N}_{> N} &:= (N, \infty) \cap \mathbb{N}. \end{aligned}$$

For any $\lambda > 1$ we will say that the set

$$\mathbb{D} = \{\lambda_n : n \in \mathbb{N}\} \subset [1, +\infty) \tag{2.3}$$

is λ -lacunary if $\inf_{n \in \mathbb{N}} \frac{\lambda_{n+1}}{\lambda_n} \geq \lambda$.

Finally, $\mathcal{C}: \mathbb{C} \rightarrow \mathbb{C}$ will be the complex conjugation, i.e., $\mathcal{C}z := \bar{z}$ for $z \in \mathbb{C}$.

2.2. Euclidean spaces. For $d \in \mathbb{Z}_+$ let $\{e_i \in \mathbb{R}^d : i \in [d]\}$ be the standard basis in $\mathbb{R}^d$. The standard inner product and the corresponding Euclidean norm on $\mathbb{R}^d$ are denoted, respectively, by

$$x \cdot \xi := \sum_{i \in [d]} x_i \xi_i \quad \text{and} \quad |x| := |x|_2 := \sqrt{x \cdot x}$$

for every $x = (x_1, \dots, x_d)$ and $\xi = (\xi_1, \dots, \xi_d) \in \mathbb{R}^d$.

Throughout the paper the d -dimensional torus $\mathbb{T}^d$, which unless otherwise stated will be identified with $[-1/2, 1/2)^d$, is endowed with the periodic norm

$$\|\xi\| := \left(\sum_{i \in [d]} \|\xi_i\|^2 \right)^{1/2}, \quad \xi = (\xi_1, \dots, \xi_d) \in \mathbb{T}^d,$$

where $\|\xi_i\| := \text{dist}(\xi_i, \mathbb{Z})$ for every $\xi_i \in \mathbb{T}$ and $i \in [d]$. However, identifying $\mathbb{T}^d$ with $[-1/2, 1/2)^d$, we see that the norm $\|\cdot\|$ coincides with the Euclidean norm $|\cdot|$ restricted to $[-1/2, 1/2)^d$.

Throughout this paper we work with a fixed cutoff function $\eta: \mathbb{R} \rightarrow [0, 1]$ that is a smooth and even function such that

$$\mathbb{1}_{[-1/4, 1/4]} \leq \eta \leq \mathbb{1}_{(-1/2, 1/2)}. \quad (2.4)$$

All constants are permitted to depend on η .

2.3. Function spaces. All vector spaces in this paper will be defined over the complex numbers $\mathbb{C}$. The triple $(X, \mathcal{B}(X), \mu)$ is a measure space X with σ -algebra $\mathcal{B}(X)$ and σ -finite measure μ . The space of all μ -measurable complex-valued functions defined on X will be denoted by $L^0(X, \mu)$. The space of all functions in $L^0(X, \mu)$ whose modulus is integrable with p -th power is denoted by $L^p(X, \mu)$ for $p \in (0, \infty)$, whereas $L^\infty(X, \mu)$ denotes the space of all essentially bounded functions in $L^0(X, \mu)$. These notions can be extended to functions taking values in a finite-dimensional normed vector space $(B, \|\cdot\|_B)$. For instance, we define

$$L^p(X, \mu; B) := \left\{ F \in L^0(X, \mu; B) : \|F\|_{L^p(X, \mu; B)} := \|\|F\|_B\|_{L^p(X, \mu)} < \infty \right\},$$

where $L^0(X, \mu; B)$ denotes the space of measurable functions from X to B (up to almost everywhere equivalence). One can extend these notions to infinite-dimensional B , at least when B is separable. In this paper, however, we will always be able to work in finite-dimensional settings by appealing to standard approximation arguments. For ease of notation, we shall abbreviate $L^p(X, \mu)$ to $L^p(X)$ and $L^p(X, \mu; B)$ to $L^p(X; B)$ for $p \in [0, \infty]$.

In our case we will mainly work with $X = \mathbb{R}^d$ or $X = \mathbb{T}^d$ endowed with Lebesgue measure, or with $X = \mathbb{Z}^d$ endowed with counting measure. If X is endowed with counting measure, then we abbreviate $L^p(X)$ to $\ell^p(X)$ and $L^p(X; B)$ to $\ell^p(X; B)$.

For a continuous linear map $T: B_1 \rightarrow B_2$ between two normed vector spaces B_1 and B_2 , we use $\|T\|_{B_1 \rightarrow B_2}$ to denote its operator norm.

We say that $f \in L^0(X)$ is 1-bounded if $f \in L^\infty(X)$ and $\|f\|_{L^\infty(X)} \leq 1$. We will mainly work with 1-bounded functions in Sections 4 and 6.

For any $F \in L^0(X, \mu)$ and a measurable set $A \subseteq X$ with $0 < \mu(A) < \infty$ we will use the convenient notation

$$\mathbb{E}_{t \in A}^\mu F(t) := \frac{1}{\mu(A)} \int_A F(t) d\mu(t). \quad (2.5)$$

If X is endowed with counting measure $\mu(A) := \#A$, then for $\emptyset \neq A \subseteq X$ we will abbreviate $\mathbb{E}_{t \in A}^\mu F(t)$ to $\mathbb{E}_{t \in A} F(t)$, which is the average of F over A , i.e., $\mathbb{E}_{t \in A} F(t) := \frac{1}{\#A} \sum_{t \in A} F(t)$. Sometimes, to simplify notation, we will write $|A|$ instead of $\#A$ for the cardinality of a set A ; see, for example, Section 4.

2.4. Convolutions. Let $(\mathbb{G}, +)$ be a locally compact abelian group (LCA group) equipped with a Haar measure $\lambda_{\mathbb{G}}$. For two functions $f, g \in L^1(\mathbb{G})$ we define their convolution by

$$f *_{\mathbb{G}} g(x) := f * g(x) := \int_{\mathbb{G}} f(x - y) g(y) d\lambda_{\mathbb{G}}(y).$$

For the sake of simplicity, we will usually abbreviate $*_{\mathbb{G}}$ to $*$; the meaning of this notation will always be clear from the context and will not cause any confusion. Since $\mathbb{G}$ is abelian, we readily see that $f * g = g * f$.

Fix $k \in \mathbb{Z}_+$ and, for each $i \in [k]$, let $(\mathbb{G}_i, +)$ be a locally compact abelian group equipped with a Haar measure $\lambda_{\mathbb{G}_i}$. We consider their product $\mathbb{G} = \mathbb{G}_1 \times \cdots \times \mathbb{G}_k$ with the product measure $\lambda_{\mathbb{G}} = \lambda_{\mathbb{G}_1} \times \cdots \times \lambda_{\mathbb{G}_k}$. Given $i \in [k]$, for any $f \in L^1(\mathbb{G})$ and $g \in L^1(\mathbb{G}_i)$ we may define their i -th convolution by

$$f *_{\mathbb{G}_i} g(x) := f *_i g(x) := \int_{\mathbb{G}_i} f(x_1, \dots, x_{i-1}, x_i - y, x_{i+1}, \dots, x_k) g(y) d\lambda_{\mathbb{G}_i}(y),$$

for $x = (x_1, \dots, x_k) \in \mathbb{G}$. In other words, $f *_i g(x)$ coincides with the function f except the i -th variable where it is defined as a convolution with the function g . We will also write $f *_{\mathbb{G}_i} g(x) = f *_{\mathbb{G}, i} g(x)$ if $\mathbb{G}_1 = \cdots = \mathbb{G}_k = \mathbb{G}$. We will mainly work with $\mathbb{G} = \mathbb{R}^d$ endowed with Lebesgue measure or with $\mathbb{G} = \mathbb{Z}^d$ endowed with counting measure.

2.5. Fourier transform. Although we will use Fourier analysis only on $\mathbb{R}^d$, $\mathbb{T}^d$, and $\mathbb{Z}^d$, it will be convenient to set out some abstract harmonic analysis notation to perform this analysis in a unified fashion and to avoid repetition. We shall write $e(z) := e^{2\pi i z}$ for every $z \in \mathbb{C}$, where $i^2 = -1$.

As before, let $(\mathbb{G}, +)$ be an LCA group equipped with a Haar measure $\lambda_{\mathbb{G}}$. It is well known (see, for instance, [86]) that every LCA group $\mathbb{G}$ has a Pontryagin dual $\hat{\mathbb{G}} = (\hat{\mathbb{G}}, +)$, an LCA group with a Haar measure $\lambda_{\hat{\mathbb{G}}}$ and a pairing, i.e., a continuous bihomomorphism $\mathbb{G} \times \hat{\mathbb{G}} \ni (x, \xi) \mapsto \langle x, \xi \rangle \in \mathbb{T}$, such that the Fourier transform $\mathcal{F}_{\mathbb{G}}: L^1(\mathbb{G}) \rightarrow C(\hat{\mathbb{G}})$ given by

$$\mathcal{F}_{\mathbb{G}} f(\xi) := \int_{\mathbb{G}} f(x) e(\langle x, \xi \rangle) d\lambda_{\mathbb{G}}(x), \quad \xi \in \hat{\mathbb{G}},$$

extends to a unitary map from $L^2(\mathbb{G})$ to $L^2(\hat{\mathbb{G}})$. We have Plancherel's identity

$$\|\mathcal{F}_{\mathbb{G}} f\|_{L^2(\hat{\mathbb{G}})} = \|f\|_{L^2(\mathbb{G})}, \quad f \in L^2(\mathbb{G}),$$

and the inverse Fourier transform $\mathcal{F}_{\mathbb{G}}^{-1}: L^2(\hat{\mathbb{G}}) \rightarrow L^2(\mathbb{G})$ is given by the formula

$$\mathcal{F}_{\mathbb{G}}^{-1}f(x) = \int_{\hat{\mathbb{G}}} f(\xi) e(-\langle x, \xi \rangle) d\lambda_{\hat{\mathbb{G}}}(\xi), \quad x \in \mathbb{G},$$

whenever $f \in L^1(\hat{\mathbb{G}}) \cap L^2(\hat{\mathbb{G}})$. We will mainly work with concrete pairs $(\mathbb{G}, \hat{\mathbb{G}})$ of Pontryagin dual LCA groups:

- (i) Consider $\mathbb{G} = \mathbb{R}^d$ with Lebesgue measure $\lambda_{\mathbb{R}^d} = dx$. Then $\hat{\mathbb{G}} = \mathbb{R}^d$ with Lebesgue measure $\lambda_{\mathbb{R}^d} = d\xi$ is a Pontryagin dual, with pairing $\langle x, \xi \rangle := x \cdot \xi$. Moreover, for any $f \in L^1(\mathbb{R}^d)$ we have

$$\mathcal{F}_{\mathbb{R}^d}f(\xi) := \int_{\mathbb{R}^d} f(x) e(x \cdot \xi) dx, \quad \xi \in \mathbb{R}^d.$$

- (ii) Consider $\mathbb{G} = \mathbb{Z}^d$ with counting measure $\lambda_{\mathbb{Z}^d}$. Then $\hat{\mathbb{G}} = \mathbb{T}^d$ with Lebesgue measure $\lambda_{\mathbb{T}^d} = d\xi$ is a Pontryagin dual, with pairing $\langle x, \xi \rangle := x \cdot \xi$. Moreover, for any $f \in \ell^1(\mathbb{Z}^d)$ we have

$$\mathcal{F}_{\mathbb{Z}^d}f(\xi) := \sum_{x \in \mathbb{Z}^d} f(x) e(x \cdot \xi), \quad \xi \in \mathbb{T}^d.$$

- (iii) Consider LCA groups $\mathbb{G}_1, \dots, \mathbb{G}_k$ with Pontryagin duals $\hat{\mathbb{G}}_1, \dots, \hat{\mathbb{G}}_k$. Then the product $\mathbb{G} = \mathbb{G}_1 \times \dots \times \mathbb{G}_k$ with product Haar measure $\lambda_{\mathbb{G}} = \lambda_{\mathbb{G}_1} \times \dots \times \lambda_{\mathbb{G}_k}$ is an LCA group with Pontryagin dual $\hat{\mathbb{G}} = \hat{\mathbb{G}}_1 \times \dots \times \hat{\mathbb{G}}_k$ with product Haar measure $\lambda_{\hat{\mathbb{G}}} = \lambda_{\hat{\mathbb{G}}_1} \times \dots \times \lambda_{\hat{\mathbb{G}}_k}$ and pairing $\langle (x_1, \dots, x_k), (\xi_1, \dots, \xi_k) \rangle := \langle x_1, \xi_1 \rangle + \dots + \langle x_k, \xi_k \rangle$.

For LCA product groups $\mathbb{G} = \mathbb{G}_1 \times \dots \times \mathbb{G}_k$ we consider partial Fourier transforms. Given $i \in [k]$, for $x = (x_1, \dots, x_k) \in \mathbb{G}$ and $\xi_i \in \hat{\mathbb{G}}_i$ we set

$$x_{(\xi_i)} := (x_1, \dots, x_{i-1}, \xi_i, x_{i+1}, \dots, x_k).$$

Then, for any $f \in L^1(\mathbb{G}) \cap L^2(\mathbb{G})$, the i -th Fourier transform is given by

$$\mathcal{F}_{i,\mathbb{G}}f(x_{(\xi_i)}) := \int_{\mathbb{G}_i} f(x) e(\langle x_i, \xi_i \rangle) d\lambda_{\mathbb{G}_i}(x_i),$$

for $x = (x_1, \dots, x_k) \in \mathbb{G}$ and $\xi_i \in \hat{\mathbb{G}}_i$. In a similar manner, we define the i -th inverse Fourier transform.

For any bounded measurable function $\mathbf{m}: \hat{\mathbb{G}} \rightarrow \mathbb{C}$ and any test function $f: \mathbb{G} \rightarrow \mathbb{C}$ we define the Fourier multiplier operator by

$$T_{\mathbb{G}}[\mathbf{m}]f(x) := \int_{\hat{\mathbb{G}}} e(-\langle x, \xi \rangle) \mathbf{m}(\xi) \mathcal{F}_{\mathbb{G}}f(\xi) d\lambda_{\hat{\mathbb{G}}}(\xi), \quad x \in \mathbb{G}. \quad (2.6)$$

One may think of $f: \mathbb{G} \rightarrow \mathbb{C}$ as a function on $\mathbb{G}$ that is compactly supported (and smooth if $\mathbb{G} = \mathbb{R}^d$), or as any other function for which (2.6) makes sense. We call $\mathbf{m}$ a Fourier multiplier on $\hat{\mathbb{G}}$.

When $\mathbb{G} = \mathbb{R}^d$ and $\mathbf{m}: \mathbb{R}^d \rightarrow \mathbb{C}$ is a Fourier multiplier on $\mathbb{R}^d$ supported on $[-1/2, 1/2)^d$, we at times abuse notation and identify $\mathbf{m}$ with its periodization

$$\mathbf{m}_{\text{per}}(\xi) := \sum_{n \in \mathbb{Z}^d} \mathbf{m}(\xi + n),$$

thereby also viewing $\mathbf{m}$ as a Fourier multiplier on $\mathbb{T}^d$; see [71, Section 2] for a detailed comparison of $T_{\mathbb{R}^d}[\mathbf{m}]$ and $T_{\mathbb{Z}^d}[\mathbf{m}]$.

If $\mathbb{G} = \mathbb{G}_1 \times \cdots \times \mathbb{G}_k$ is a product of LCA groups and $\mathbf{m}: \hat{\mathbb{G}}_i \rightarrow \mathbb{C}$ is a bounded measurable function acting only on $\hat{\mathbb{G}}_i$ for some $i \in [k]$, then for any $f \in L^1(\mathbb{G}) \cap L^2(\mathbb{G})$ it makes sense to consider i -th Fourier multiplier operators, which can be defined by

$$T_{i,\mathbb{G}}[\mathbf{m}]f(x) := \int_{\hat{\mathbb{G}}_i} e(-\langle x_i, \xi_i \rangle) \mathbf{m}(\xi_i) \mathcal{F}_{i,\mathbb{G}} f(x_{(\xi_i)}) d\lambda_{\hat{\mathbb{G}}_i}(\xi_i),$$

for $x = (x_1, \dots, x_k) \in \mathbb{G}$.

Finally, for any finite set of frequencies $\Sigma \subseteq \hat{\mathbb{G}}$ we define

$$T_{\mathbb{G}}^{\Sigma}[\mathbf{m}]f(x) := T_{\mathbb{G}} \left[\sum_{\theta \in \Sigma} \tau_{\theta} \mathbf{m} \right] f(x), \quad x \in \mathbb{G}, \quad (2.7)$$

where $\tau_{\theta} \mathbf{m}(\xi) := \mathbf{m}(\xi - \theta)$ for $\xi \in \hat{\mathbb{G}}$. If $\mathbb{G} = \mathbb{G}_1 \times \cdots \times \mathbb{G}_k$ is a product of LCA groups, then we similarly define $T_{i,\mathbb{G}}^{\Sigma}[\mathbf{m}]f$ for any finite set $\Sigma \subseteq \hat{\mathbb{G}}_i$ and $i \in [k]$.

2.6. Variational norms. For any $\mathbb{I} \subseteq \mathbb{R}$, any family $(\mathbf{a}_t)_{t \in \mathbb{I}} = (\mathbf{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C}$, and any exponent $r \in [1, \infty)$, the r -variation seminorm is defined to be

$$V^r(\mathbf{a}_t : t \in \mathbb{I}) := \sup_{J \in \mathbb{Z}_+} \sup_{\substack{t_0 < \cdots < t_J \\ t_j \in \mathbb{I}}} \left(\sum_{j \in [J]} |\mathbf{a}_{t_j} - \mathbf{a}_{t_{j-1}}|^r \right)^{1/r}, \quad (2.8)$$

where the supremum is taken over all finite increasing sequences in $\mathbb{I}$, and is set by convention to equal zero if $\#\mathbb{I} \leq 1$. Taking limits as $r \rightarrow \infty$ we also adopt the convention

$$V^{\infty}(\mathbf{a}_t : t \in \mathbb{I}) := \sup_{t, t' \in \mathbb{I}} |\mathbf{a}_t - \mathbf{a}_{t'}|.$$

The r -variation norm for $r \in [1, \infty)$ is defined by

$$\mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I}) := \sup_{t \in \mathbb{I}} |\mathbf{a}_t| + V^r(\mathbf{a}_t : t \in \mathbb{I}). \quad (2.9)$$

Note that the $\mathbf{V}^r$ norm is nonincreasing in r , and comparable to the ℓ^{∞} norm when $r = \infty$. We also observe the simple triangle inequality

$$\mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I}) \lesssim \mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I}_1) + \mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I}_2)$$

whenever $\mathbb{I} = \mathbb{I}_1 \sqcup \mathbb{I}_2$ is an ordered partition of $\mathbb{I}$, which means that $t_1 < t_2$ for all $t_1 \in \mathbb{I}_1, t_2 \in \mathbb{I}_2$. If $\mathbb{I} \subseteq \mathbb{R}$ is countable, then we have

$$\mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I}) \lesssim \left(\sum_{t \in \mathbb{I}} |\mathbf{a}_t|^r \right)^{1/r}.$$

One easily establishes the algebra property

$$\mathbf{V}^r(\mathbf{a}_t \mathbf{b}_t : t \in \mathbb{I}) \lesssim \mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I}) \mathbf{V}^r(\mathbf{b}_t : t \in \mathbb{I}) \quad (2.10)$$

for any scalar sequences $(\mathbf{a}_t)_{t \in \mathbb{I}}, (\mathbf{b}_t)_{t \in \mathbb{I}}$.

Finally, the r -variational spaces will be defined by

$$\mathbf{V}^r(\mathbb{I}) := \{(\mathbf{a}_t : t \in \mathbb{I}) \subseteq \mathbb{C} : \mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I}) < \infty\},$$

and we will also write $(\mathbf{a}_t)_{t \in \mathbb{I}} \in L^p(X; \mathbf{V}^r(\mathbb{I}))$ if and only if

$$\|(\mathbf{a}_t)_{t \in \mathbb{I}}\|_{L^p(X; \mathbf{V}^r(\mathbb{I}))} = \|(\mathbf{a}_t : t \in \mathbb{I})\|_{L^p(X; \mathbf{V}^r(\mathbb{I}))} = \|\mathbf{V}^r(\mathbf{a}_t : t \in \mathbb{I})\|_{L^p(X)} < \infty.$$

We will use a similar convention for the spaces $V^r(\mathbb{I})$.

We also recall the well known Rademacher–Menshov inequality, which asserts that for any $n_0, m \in \mathbb{N}$ with $n_0 < 2^m$ and any sequence of complex numbers $(\mathbf{a}_n : n \in \mathbb{N})$ we have

$$V^2(\mathbf{a}_n : n \in U) \leq \sqrt{2} \sum_{i \in \mathbb{N}_{\leq m}} \left(\sum_{j \in [2^{m-i}]} \left| \mathbf{1}_{U_j^i \subseteq U} \sum_{n \in U_j^i} (\mathbf{a}_{n+1} - \mathbf{a}_n) \right|^2 \right)^{1/2}, \quad (2.11)$$

where $U := U(n_0, m) := [n_0, 2^m) \cap \mathbb{Z}$ and $U_j^i := [(j-1)2^i, j2^i) \cap \mathbb{Z}$ for $i, j \in \mathbb{Z}$; see [74, Lemma 2.5, p. 534] for the proof.

3. IONESCU–WAINGER MULTIPLIER THEOREM FOR CANONICAL FRACTIONS

The results of this section will be stated for future reference for arbitrary dimension $d \in \mathbb{Z}_+$, although only $d = 1$ will be used throughout this paper.

3.1. Magyar–Stein–Wainger sampling method. An important tool in discrete analogues in harmonic analysis is a quantitative variant of the Shannon sampling theorem from a seminal paper of Magyar, Stein, and Wainger [71]. Recalling the definition of $T_{\mathbb{Z}^d}^\Sigma[\mathbf{m}]$ from (2.7), their sampling result can be subsumed under the following useful proposition.

Proposition 3.1. *For $d \in \mathbb{Z}_+$ there is a constant $\mathbf{C}_{\text{MSW}}(d) \in \mathbb{R}_+$ such that the following holds. Let $p \in [1, \infty]$ and $q \in \mathbb{Z}_+$. If $\mathbf{m} : \mathbb{R}^d \rightarrow L(B_1, B_2)$ is a measurable function supported on $[-1/2, 1/2)^d/q$, whose values are bounded linear operators between two finite-dimensional Banach spaces B_1 and B_2 , then*

$$\left\| T_{\mathbb{Z}^d}^{(\frac{1}{q}\mathbb{Z}/\mathbb{Z})^d}[\mathbf{m}] \right\|_{\ell^p(\mathbb{Z}^d; B_1) \rightarrow \ell^p(\mathbb{Z}^d; B_2)} \leq \mathbf{C}_{\text{MSW}}(d) \|T_{\mathbb{R}^d}[\mathbf{m}]\|_{L^p(\mathbb{R}^d; B_1) \rightarrow L^p(\mathbb{R}^d; B_2)}.$$

We now make a few remarks about Proposition 3.1.

1. The sampling principle developed by Magyar, Stein, and Wainger serves as a “bridge” between discrete analogues in harmonic analysis and classical harmonic analysis. It enables the deduction of discrete problems from their continuous counterparts.
2. The proof of Proposition 3.1 can be found in [71, Corollary 2.1, p. 196]. We also refer to [76] for a generalization of Proposition 3.1 to real interpolation spaces.
3. All Banach spaces that arise in our problems are infinite-dimensional and separable. Despite Proposition 3.1 being formulated for finite-dimensional Banach spaces, it can still be applied in our situation using standard approximation arguments that reduce infinite-dimensional settings to finite-dimensional settings.
4. We emphasize that B_1 and B_2 are general (finite-dimensional) Banach spaces and thus, in view of the previous remark, Proposition 3.1 includes maximal functions and can also accommodate r -variation seminorms as well as many other seminorms.
5. A fundamental question in discrete harmonic analysis is whether it is possible to replace the set $(\frac{1}{q}\mathbb{Z}/\mathbb{Z})^d$ appearing in Proposition 3.1 with other sets of fractions Σ , while still maintaining mapping properties of $T_{\mathbb{Z}^d}^\Sigma[\mathbf{m}]$ that depend in a good way on the size of Σ , under a suitable support condition for $\mathbf{m}$. This was achieved for the

first time in a groundbreaking paper of Ionescu and Wainger [56] for a certain family of reduced fractions Σ whose denominators have special factorization properties.

In this paper, our aim is to prove a variant of the Ionescu–Wainger multiplier theorem for the set of canonical fractions that arises from the classical circle method. This will be critical in our proof of the multilinear Weyl inequality; see Theorem 6.1.

3.2. A refined variant of the Ionescu–Wainger multiplier theorem. For $d \in \mathbb{Z}_+$ and $N \geq 1$ define 1-periodic sets of the so-called *canonical fractions* by

$$\mathcal{R}_{\leq N}^d := \left\{ \frac{a}{q} \in (\mathbb{Q}/\mathbb{Z})^d : q \in [N] \text{ and } \gcd(a, q) = 1 \right\}, \quad (3.2)$$

where $\gcd(a, q) := \gcd(a_1, \dots, a_d, q)$ denotes the greatest common divisor of $a_1, \dots, a_d, q$. To be precise, the elements of $\mathcal{R}_{\leq N}^d$ are equivalence classes $\theta = \frac{a}{q} + \mathbb{Z}^d$ in $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$, with a well-defined denominator q such that every representative $\frac{b}{q} \in \theta$ satisfies $\gcd(b, q) = 1$. We often identify $\frac{a}{q} \in \mathbb{Q}^d$ with the equivalence class $\frac{a}{q} + \mathbb{Z}^d$.

Using the definition of $T_{\mathbb{Z}^d}^\Sigma[\mathbf{m}]$ from (2.7) we now formulate the main result of this section. Here, $p'_0 := p_0/(p_0 - 1)$ denotes the conjugate exponent of p_0 .

Theorem 3.3. *Let $p'_0 \leq p \leq p_0$ for some $p_0 \in 2\mathbb{Z}_+$. Then there exists a constant $\mathbf{C}_{\text{IW}}(p_0) \in \mathbb{R}_+$ such that for every $d \in \mathbb{Z}_+$ and every $N \in \mathbb{N}_{\geq 100}$ the following is true. Assume that*

$$0 < \vartheta \leq (2p_0 N^{p_0})^{-1}$$

and let $\mathbf{m}: \mathbb{R}^d \rightarrow L(H_1, H_2)$ be a measurable function supported on $[-\vartheta, \vartheta]^d$, whose values are bounded linear operators between two separable Hilbert spaces H_1 and H_2 . Denote $\rho_N := \rho_N(p_0) := \mathbf{C}_{\text{IW}}(p_0) \frac{\log \log \log N}{\log \log N}$. Then

$$\left\| T_{\mathbb{Z}^d}^{\mathcal{R}_{\leq N}^d}[\mathbf{m}] \right\|_{\ell^p(\mathbb{Z}^d; H_1) \rightarrow \ell^p(\mathbb{Z}^d; H_2)} \lesssim_{d, p_0} N^{\rho_N} \|T_{\mathbb{R}^d}[\mathbf{m}]\|_{L^{p_0}(\mathbb{R}^d; H_1) \rightarrow L^{p_0}(\mathbb{R}^d; H_2)}. \quad (3.4)$$

A few remarks about Theorem 3.3.

1. Theorem 3.3 gives an affirmative answer to a question of Ionescu and Wainger from [56, Remark 3, below Theorem 1.5, p. 361], which asks about (3.4) in the scalar setting $H_1 = H_2 = \mathbb{C}$ with $C_{d, p_0, \rho} N^\rho$ for arbitrary $\rho \in \mathbb{R}_+$ in place of N^{ρ_N} . Obviously, (3.4) yields the bound postulated in [56], since $\lim_{N \rightarrow \infty} \rho_N = 0$.
2. Although Proposition 3.1 plays an important role in the proof of Theorem 3.3, the latter result is only applicable to multipliers defined in the Hilbert space setting. However, in the next section, we will show that Theorem 3.3 can be applied to some Banach space settings. This can be done for a broad range of Banach spaces that, in some manner, exhibit similarities to Hilbert spaces.
3. The advantage of using Theorem 3.3 is that square function estimates in the discrete setting can be directly deduced from the corresponding continuous square function estimates, which, as we will soon see, is very useful in many applications.
4. The norm on the right hand side of (3.4), unlike the support hypothesis, is scale-invariant. To be more precise, if $\mathbf{m}$ is replaced with $\mathbf{m}(A \cdot)$ for any invertible linear transformation $A: \mathbb{R}^d \rightarrow \mathbb{R}^d$, then the norm in (3.4) remains unchanged.

5. It would be interesting to know whether inequality (3.4) holds with an absolute constant independent of N . Although in most applications inequality (3.4) is sufficient, having (3.4) with a constant independent of N would simplify some arguments.
6. In the proof of Theorem 3.3 we will use variants of two properties introduced in [56]: a uniqueness property and the so-called $\mathcal{O}$ property. These two properties efficiently detect strong orthogonalities between the Ionescu–Wainger fractions, which are rationals whose denominators have certain factorization properties. The novelty of our approach lies in using the set of canonical fractions (3.2) instead of the Ionescu–Wainger fractions [56], which are quite complex.

3.2.1. Proof of Theorem 3.3, simple reductions. The case $p_0 = 2$ holds by Plancherel’s theorem. By duality and interpolation we can assume that $p = p_0 = 2r$ for $r \in \mathbb{N}_{\geq 2}$. Replacing $\mathfrak{m}$ by $\tilde{\mathfrak{m}}: \mathbb{R}^d \rightarrow L(H_1 \oplus H_2, H_1 \oplus H_2)$, where $\tilde{\mathfrak{m}}(x)(h_1, h_2) := (0, \mathfrak{m}(x)(h_1))$ for $x \in \mathbb{R}^d$, we can take $H_1 = H_2 = H$. We normalize $\|T_{\mathbb{R}^d}[\mathfrak{m}]\|_{L^{2r}(\mathbb{R}^d; H) \rightarrow L^{2r}(\mathbb{R}^d; H)} = 1$. Finally, we can assume that $N \geq N_0$ for a large integer $N_0 \in \mathbb{N}_{\geq 10^{10}}$, since $\mathcal{R}_{\leq N}^d$ produces at most N^{d+1} operators, each having a norm bounded by $\mathbf{C}_{\text{MSW}}(d)$ in view of Proposition 3.1.

3.2.2. Uniqueness property and property $\mathcal{O}_N$. We follow [56] in using the uniqueness property and property $\mathcal{O}$, two important concepts introduced therein for detecting orthogonality.

Definition 3.5 (Uniqueness property). A sequence $(v_1, \dots, v_n)$ has the *uniqueness property* if there is $i \in [n]$ such that $v_i \neq v_j$ for all $j \in [n] \setminus \{i\}$. In other words, the element v_i occurs in the sequence $(v_1, \dots, v_n)$ exactly once. This definition will primarily be used for sequences of rational numbers, but it is also applicable to more general sequences.

Using Definition 3.5 we can formulate an abstract orthogonality principle.

Proposition 3.6 (Abstract orthogonality principle). *Given $r \in \mathbb{Z}_+$, there exists a constant $\mathbf{C}_{\text{UP}}(r) \in \mathbb{R}_+$ such that the following is true. Let $(X, \mathcal{B}(X), \mu)$ be a measure space, let H be a separable Hilbert space, and let $V_1, \dots, V_r \subset \mathbb{Z}_+$ be finite sets, not necessarily disjoint. For every pair (i, v) with $v \in V_i$ let $F_{i,v} \in L^{2r}(X; H)$ and suppose that for each $2r$ -tuple*

$$(v_{1,1}, v_{1,2}, \dots, v_{r,1}, v_{r,2}) \in V_1^2 \times \dots \times V_r^2$$

with the uniqueness property the following condition holds

$$\int_X \prod_{i \in [r]} \langle F_{i,v_{i,1}}(x), F_{i,v_{i,2}}(x) \rangle_H d\mu(x) = 0. \quad (3.7)$$

Then

$$\int_X \prod_{i \in [r]} \left\| \sum_{v \in V_i} F_{i,v}(x) \right\|_H^2 d\mu(x) \leq \mathbf{C}_{\text{UP}}(r) \int_X \prod_{i \in [r]} \left(\sum_{v \in V_i} \|F_{i,v}(x)\|_H^2 \right) d\mu(x).$$

Proof. We refer to [75, Corollary 2.24] for a detailed proof; see also [56, Lemma 2.2, p. 363] for a scalar version of this principle. $\square$

Our aim will be to split the set of canonical fractions $\mathcal{R}_{\leq N}^d$ into a controlled number of pieces for which the uniqueness property implies condition (3.7).

Here our approach deviates from [56] and [75]. In the latter two papers an essential role was played by the Ionescu–Wainger rationals whose denominators were factored into two parts, one being a relatively large product of powers of small prime divisors, and the other consisting of a product of a small number of powers of large prime divisors.

In contrast to [56] and [75], instead of working with the Ionescu–Wainger rationals, we shall work directly with the set of canonical fractions $\mathcal{R}_{\leq N}^d$ from (3.2). We introduce a new concept of N -lifted composites, which will allow us to effectively split $\mathcal{R}_{\leq N}^d$.

Definition 3.8 (N -lifted composites). Let $\mathbb{P}$ be the set of prime numbers. For $N \in \mathbb{Z}_+$ an N -lifted composite Q_q of an integer $q \in [N]$ is defined by setting

$$Q_q := \prod_{p \in \mathbb{P} : q \in p\mathbb{Z}} p^{\lfloor \log_p N \rfloor}$$

with the convention $Q_1 := 1$. In other words, q and Q_q are divisible by exactly the same prime numbers, while Q_q is a product of integers of the form $p^{\lfloor \log_p N \rfloor}$ for some $p \in \mathbb{P} \cap \mathbb{N}_{\leq N}$. Obviously, q divides Q_q for each $q \in [N]$.

We need to adapt the concept of property $\mathcal{O}$ from [56] to fit our needs.

Definition 3.9 (Property $\mathcal{O}_N$). For $N \in \mathbb{N}_{\geq N_0}$ we shall say that $\Lambda \subset \mathbb{Z}_+$ has *property $\mathcal{O}_N$* if $\Lambda \subseteq S_1 \cdots S_k := \{s_1 \cdots s_k : s_1 \in S_1, \dots, s_k \in S_k\}$ for some $k \in [\frac{5 \log N}{\log \log N}]$, where $S_1, \dots, S_k$ are disjoint subsets of $\{p^{\lfloor \log_p N \rfloor} : p \in \mathbb{P} \cap \mathbb{N}_{\leq N}\}$.

The following lemma explains the restriction $k \in [\frac{5 \log N}{\log \log N}]$.

Lemma 3.10. *If N_0 is sufficiently large, then for every $N \in \mathbb{N}_{\geq N_0}$ and distinct $p_1, \dots, p_{k_N} \in \mathbb{P}$ with $k_N := \lfloor \frac{5 \log N}{\log \log N} \rfloor$ we have $p_1 \cdots p_{k_N} > N$.*

Proof. For $n \in \mathbb{Z}_+$, its primorial $n\#$ is the product of all $p \in \mathbb{P} \cap \mathbb{N}_{\leq n}$. By the prime number theorem, we have $\lim_{n \rightarrow \infty} \sqrt[n]{n\#} = e$. Take $N \in \mathbb{Z}_+$ such that $n_N := \lfloor \log N \rfloor + 1$ is large. Then $n_N\# > 2^{n_N} > N$ and $\#(\mathbb{P} \cap \mathbb{N}_{\leq n_N}) \leq \lfloor \frac{5 \log N}{\log \log N} \rfloor = k_N$. The latter implies $p_1 \cdots p_{k_N} \geq n_N\#$, as desired. $\square$

3.2.3. Partitioning of the set of canonical fractions. For $N \in \mathbb{N}_{\geq N_0}$ and $\frac{a}{q} \in \mathcal{R}_{\leq N}^d$ the corresponding composite Q_q has at most $\lfloor \frac{5 \log N}{\log \log N} \rfloor$ factors $p^{\lfloor \log_p N \rfloor}$, while each p is one of the first $\lfloor \frac{2N}{\log N} \rfloor$ primes. This motivates the following counting lemma originating in [73].

Lemma 3.11. *There exists a constant $\mathbf{C}_{\text{sur}} \in \mathbb{Z}_+$ such that the following holds. For $N \in \mathbb{N}_{\geq N_0}$ let V be a set of size $l \in [\frac{2N}{\log N}]$ and let $k \in [\frac{5 \log N}{\log \log N}]$ satisfy $k \leq l$. Then, denoting $J := \lfloor \mathbf{C}_{\text{sur}} 2^{\mathbf{C}_{\text{sur}} \log N / \log \log N} \rfloor$, we can find J surjective (not necessarily distinct) functions $g_1, \dots, g_J : V \rightarrow [k]$ such that $\max\{\#g_j(E) : j \in [J]\} = k$ for every $E \subseteq V$ with $\#E = k$.*

Proof. We refer to [75, Lemma 2.16]. $\square$

Proceeding in two steps, we split $\mathcal{R}_{\leq N}^d$ into at most $N^{\frac{C}{\log \log N}}$ pieces, each satisfying property $\mathcal{O}_N$.

Step 1. For a large integer $N \in \mathbb{Z}_+$, by Lemma 3.10, we can split $\mathcal{R}_{\leq N}^d$ into disjoint subsets $\mathcal{R}_{N,0}^d := \mathcal{R}_{\leq 1}^d$ and $\mathcal{R}_{N,k}^d$ for $k \in [\frac{5 \log N}{\log \log N}]$, consisting of fractions $\frac{a}{q} \in \mathcal{R}_{\leq N}^d$ whose denominators q have exactly k prime divisors. We may assume, without loss of generality, that $k > 0$, since the case $k = 0$ may be readily handled by Proposition 3.1. We also assume that $\mathcal{R}_{N,k}^d$ is nonempty.

Step 2. Next, we fix $k \in [\frac{5 \log N}{\log \log N}]$ as above and apply Lemma 3.11 with

$$V := \{p^{\lfloor \log_p N \rfloor} : p \in \mathbb{P} \cap \mathbb{N}_{\leq N}\} = \{Q_p : p \in \mathbb{P} \cap \mathbb{N}_{\leq N}\},$$

obtaining surjections $g_1, \dots, g_J : V \rightarrow [k]$ with $J := \lfloor \mathbf{C}_{\text{sur}} 2^{\mathbf{C}_{\text{sur}} \log N / \log \log N} \rfloor$. Given $j \in [J]$, we define $S_i^j := g_j[\{i\}]^{-1}$ for $i \in [k]$ and, consequently, we set

$$\mathcal{R}_{N,k,j}^d := \left\{ \frac{a}{q} \in \mathcal{R}_{N,k}^d : Q_q \in S_1^j \cdots S_k^j \setminus \bigcup_{i \in [j-1]} S_1^i \cdots S_k^i \right\}, \quad (3.12)$$

where if $j = 1$, then the union above is understood to be empty. For each $j \in [J]$ the set $\{Q_q : \frac{a}{q} \in \mathcal{R}_{N,k,j}^d\} \subset \mathbb{Z}_+$ has property $\mathcal{O}_N$. Moreover, $\mathcal{R}_{N,k}^d = \bigcup_{j \in [J]} \mathcal{R}_{N,k,j}^d$, which immediately follows from Lemma 3.11.

Our task is to prove (3.4) with any nonempty $\mathcal{R}_{N,k,j}^d$ in place of $\mathcal{R}_{\leq N}^d$. To facilitate matters, we introduce some useful notation and terminology.

3.2.4. $\mathbb{P}$ -irreducibility and factorization. We now present a new concept of $\mathbb{P}$ -irreducible fractions, which will be crucial in verifying condition (3.7).

Definition 3.13 (Quotients, reduced fractions and $\mathbb{P}$ -irreducible fractions). Fix $d, q \in \mathbb{Z}_+$. Identifying $\frac{a}{q}$ with $\frac{a}{q} + \mathbb{Z}^d$, we define the following sets.

1. The set of *quotients* or *q-quotients* is defined by $\mathcal{Q}(q) := \frac{1}{q} \mathbb{N}_{< q}^d$.
2. The set of *reduced fractions* is defined by $\mathcal{R}(1) := \mathcal{Q}(1)$ or

$$\mathcal{R}(q) := \left\{ \frac{a}{q} \in \mathcal{Q}(q) : \gcd(a, q) = 1 \right\} \quad \text{for } q > 1.$$

3. The set of $\mathbb{P}$ -irreducible fractions is defined by $\mathcal{I}(1) := \mathcal{Q}(1)$ or

$$\mathcal{I}(q) := \left\{ \frac{a}{q} \in \mathcal{Q}(q) : \frac{q}{\gcd(a, q)} \text{ has the same prime divisors as } q \right\} \quad \text{for } q > 1.$$

It is easy to see that $\mathcal{I}(q) \supseteq \mathcal{R}(q)$ and that $\mathcal{I}(q)$ is a disjoint union of the sets $\mathcal{R}(\tilde{q})$, where $\tilde{q}$ runs over the set of all positive divisors of q having exactly the same prime divisors as q .

Example 3.14. For $d = 1$ and $q = 12$, identifying $\frac{a}{q}$ with $\frac{a}{q} + \mathbb{Z}$, we obtain

$$\mathcal{R}(12) = \frac{1}{12} \{1, 5, 7, 11\} \quad \text{and} \quad \mathcal{I}(12) = \frac{1}{12} \{1, 2, 5, 7, 10, 11\}.$$

For $\Lambda \subset \mathbb{Z}_+$ define

$$\mathcal{R}(\Lambda) := \bigcup_{q \in \Lambda} \mathcal{R}(q) \quad \text{and} \quad \mathcal{I}(\Lambda) := \bigcup_{q \in \Lambda} \mathcal{I}(q).$$

Obviously, we have $\mathcal{R}(\Lambda) \subseteq \mathcal{I}(\Lambda)$ and the set of canonical fractions (3.2) can be written as $\mathcal{R}_{\leq N}^d = \mathcal{R}([N])$ by using this definition. Moreover, an important feature of $\mathcal{R}(\Lambda)$ used in [56] and [75] is the factorization property asserting that if $\gcd(q, q') = 1$ for all $(q, q') \in \Lambda \times \Lambda'$, then

$$\mathcal{R}(\Lambda\Lambda') = \mathcal{R}(\Lambda) \oplus \mathcal{R}(\Lambda').$$

The direct sum $\oplus$ indicates that each element in the sumset $\mathcal{R}(\Lambda) + \mathcal{R}(\Lambda')$ has a unique representation $\frac{a}{q} + \frac{a'}{q'}$ as an element of $\mathbb{T}^d$. The identity above is a simple consequence of the Chinese remainder theorem. We now strengthen this factorization property to the $\mathbb{P}$ -irreducible fractions $\mathcal{I}(\Lambda)$.

Definition 3.15 ($\mathbb{P}$ -separated sets). A set $\Lambda \subseteq \mathbb{Z}_+$ is $\mathbb{P}$ -separated if for any distinct elements $q_1, q_2 \in \Lambda$ there is $p \in \mathbb{P}$ that divides exactly one of them.

We now prove the factorization property for $\mathcal{I}(\Lambda)$.

Lemma 3.16. Assume that $\Lambda, \Lambda' \subseteq \mathbb{Z}_+$ are $\mathbb{P}$ -separated and $\gcd(q, q') = 1$ for all $(q, q') \in \Lambda \times \Lambda'$. Then the map $\Phi: \mathcal{I}(\Lambda) \times \mathcal{I}(\Lambda') \rightarrow \mathcal{I}(\Lambda\Lambda')$ given by

$$\Phi\left(\frac{a}{q}, \frac{a'}{q'}\right) := \frac{a}{q} + \frac{a'}{q'} \quad \text{for} \quad \left(\frac{a}{q}, \frac{a'}{q'}\right) \in \mathcal{I}(\Lambda) \times \mathcal{I}(\Lambda')$$

is a bijection. In particular, we have $\mathcal{I}(\Lambda\Lambda') = \mathcal{I}(\Lambda) \oplus \mathcal{I}(\Lambda')$.

Proof. By the Chinese remainder theorem, it follows that Φ is surjective, and from Definition 3.15 we deduce that Φ is injective. The second part of the lemma now follows easily from the first part. $\square$

To establish (3.4) in Theorem 3.3, it suffices to prove

$$\left\| T_{\mathbb{Z}^d}^{\mathcal{R}_{N,k,j}^d}[\mathbf{m}]f \right\|_{\ell^{2r}(\mathbb{Z}^d; H)} \lesssim_{d,p_0} N^{\mathbf{C}_{\text{IW}}(p_0) \frac{\log \log \log N}{\log \log N}} \|f\|_{\ell^{2r}(\mathbb{Z}^d; H)} \quad (3.17)$$

since we have normalized $\|T_{\mathbb{R}^d}[\mathbf{m}]\|_{L^{2r}(\mathbb{R}^d; H) \rightarrow L^{2r}(\mathbb{R}^d; H)} = 1$. Indeed summing (3.17) over $k \in [\frac{5 \log N}{\log \log N}]$ and $j \in [J]$ incurs an acceptable, additional factor

$$\frac{5 \log N}{\log \log N} \cdot \mathbf{C}_{\text{sur}} 2^{\mathbf{C}_{\text{sur}} \log N / \log \log N}.$$

We fix k and j . The multiplier corresponding to $T_{\mathbb{Z}^d}^{\mathcal{R}_{N,k,j}^d}[\mathbf{m}]$ is given by

$$\Delta_{N,k,j}(\xi) := \sum_{q \in S_1 \cdots S_k} \sum_{\frac{a}{q} \in \mathcal{I}(q)} \mathbb{1}_{N,k,j}\left(\frac{a}{q}\right) \mathbf{m}\left(\xi - \frac{a}{q}\right),$$

where, as in (3.12), the sets $S_1 := S_1^j := g_j[\{1\}]^{-1}, \dots, S_k := S_k^j := g_j[\{k\}]^{-1}$ form a partition of V determined by the surjection g_j , and $\mathbb{1}_{N,k,j}(\frac{a}{q}) := 1$ if $\frac{a}{q}$ is $\mathbb{P}$ -irreducible and upon reduction belongs to $\mathcal{R}_{N,k,j}^d$, and $\mathbb{1}_{N,k,j}(\frac{a}{q}) := 0$ otherwise. We note that the sets of $\mathbb{P}$ -irreducible fractions $\mathcal{I}(q)$ are pairwise disjoint as q varies over the product set $S_1 \cdots S_k$.

3.2.5. Denominators. Now we exploit orthogonality between denominators. In the argument below we will use Lemma 3.16 repeatedly. Note that

$$\left\| T_{\mathbb{Z}^d}^{\mathcal{R}_{N,k,j}^d}[\mathbf{m}]f \right\|_{\ell^{2r}(\mathbb{Z}^d; H)}^{2r} = \sum_{x \in \mathbb{Z}^d} \left\| \sum_{q \in S_1 \cdots S_k} \sum_{u \in \mathcal{I}(q)} f_u(x) \right\|_H^{2r},$$

where $f_u := \mathcal{F}_{\mathbb{Z}^d}^{-1}(\mathbb{1}_{N,k,j}(u) \mathbf{m}(\cdot - u) \mathcal{F}_{\mathbb{Z}^d} f)$. Let $\mathcal{S}_\emptyset := \{1\}$ and $\mathcal{S}_L := \prod_{l \in L} S_l$ for every nonempty $L \subseteq [k]$. We observe that

$$\sum_{x \in \mathbb{Z}^d} \left\| \sum_{q \in S_1 \cdots S_k} \sum_{u \in \mathcal{I}(q)} f_u(x) \right\|_H^{2r} = \sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in \mathcal{S}_\emptyset} \left\| F_{q,\emptyset}(x) \right\|_H^2 \right)^r,$$

where $F_{q,L} := \sum_{q' \in \mathcal{S}_{[k] \setminus L}} \sum_{u \in \mathcal{I}(qq')} f_u$ for every $q \in \mathcal{S}_L$. Now, proceeding as in [75, Section 2.4], we can show for each $l \in [k]$ that

$$\sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in \mathcal{S}_{[l-1]}} \left\| F_{q,[l-1]}(x) \right\|_H^2 \right)^r \leq \mathbf{C}_{\text{UP}}(r) \sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in \mathcal{S}_{[l]}} \left\| F_{q,[l]}(x) \right\|_H^2 \right)^r \quad (3.18)$$

with $\mathbf{C}_{\text{UP}}(r)$ from Proposition 3.6. After k iterations of (3.18), we obtain

$$\sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in \mathcal{S}_\emptyset} \left\| F_{q,\emptyset}(x) \right\|_H^2 \right)^r \leq \mathbf{C}_{\text{UP}}(r)^k \sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in \mathcal{S}_{[k]}} \left\| F_{q,[k]}(x) \right\|_H^2 \right)^r$$

or, equivalently,

$$\left\| T_{\mathbb{Z}^d}^{\mathcal{R}_{N,k,j}^d}[\mathbf{m}]f \right\|_{\ell^{2r}(\mathbb{Z}^d; H)}^{2r} \leq \mathbf{C}_{\text{UP}}(r)^k \sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in S_1 \cdots S_k} \left\| \sum_{u \in \mathcal{I}(q)} f_u(x) \right\|_H^2 \right)^r. \quad (3.19)$$

The constant $\mathbf{C}_{\text{UP}}(r)^{k/2r} \leq C_r 2^{C_r \log N / \log \log N}$ is acceptable, since $k \leq \frac{5 \log N}{\log \log N}$.

3.2.6. Numerators. Now we exploit orthogonalities between numerators to bound the right-hand side of (3.19). In the argument below, as before, we will use Lemma 3.16 repeatedly. The right-hand side of (3.19) is equal to

$$\sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in S_1 \cdots S_k} \left\| \sum_{u \in \mathcal{I}(q)} f_u(x) \right\|_H^2 \right)^r = \sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in \mathcal{S}_{[k]}} \left\| F_{q,[k]}(x) \right\|_H^2 \right)^r.$$

Set $F_{q,q',q''} := \sum_{u'' \in \mathcal{I}(q'')} \left\| \sum_{u' \in \mathcal{I}(q')} \sum_{u \in \mathcal{I}(q)} f_{u+u'+u''} \right\|_H^2$ and observe that

$$\sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in \mathcal{S}_{[k]}} \left\| F_{q,[k]}(x) \right\|_H^2 \right)^r = \sum_{x \in \mathbb{Z}^d} \sum_{q \in \mathcal{S}_\emptyset} \left(\sum_{q' \in \mathcal{S}_{[k]}} \sum_{q'' \in \mathcal{S}_\emptyset} F_{q,q',q''}(x) \right)^r.$$

Denote $\mathcal{F}_{L,L'} := \sum_{q \in \mathcal{S}_L} \left(\sum_{q' \in \mathcal{S}_{L' \setminus L}} \sum_{q'' \in \mathcal{S}_{[k] \setminus L'}} F_{q,q',q''} \right)^r$ for $L \subseteq L' \subseteq [k]$. We can rewrite the right-hand side above as $\|\mathcal{F}_{\emptyset,[k]}\|_{\ell^1(\mathbb{Z}^d)}$. Thus, we obtain

$$\sum_{x \in \mathbb{Z}^d} \left(\sum_{q \in S_1 \cdots S_k} \left\| \sum_{u \in \mathcal{I}(q)} f_u(x) \right\|_H^2 \right)^r = \|\mathcal{F}_{\emptyset,[k]}\|_{\ell^1(\mathbb{Z}^d)}. \quad (3.20)$$

Assuming $L \subsetneq L' \subseteq [k]$ and writing $L_+ = L \cup \{l\}$ and $L'_- = L' \setminus \{l\}$ with $l = \min L' \setminus L$, we proceed as in [75, Section 2.5] to show that

$$\|\mathcal{F}_{L,L'}\|_{\ell^1(\mathbb{Z}^d)} \leq C_r (\|\mathcal{F}_{L_+,L'}\|_{\ell^1(\mathbb{Z}^d)} + \|\mathcal{F}_{L,L'_-}\|_{\ell^1(\mathbb{Z}^d)}) \quad (3.21)$$

with $C_r \simeq_r 1$ by using Proposition 3.6. After k iterations of (3.21), we obtain

$$\|\mathcal{F}_{\emptyset,[k]}\|_{\ell^1(\mathbb{Z}^d)} \leq C_r^k \sum_{L \subseteq [k]} \|\mathcal{F}_{L,L}\|_{\ell^1(\mathbb{Z}^d)}. \quad (3.22)$$

The implied constant in (3.22) is acceptable, since $k \in [\frac{5 \log N}{\log \log N}]$.

3.2.7. Square function estimates. Gathering the estimates and identities (3.19), (3.20), and (3.22), we reduce matters to proving

$$\sum_{L \subseteq [k]} \|\mathcal{F}_{L,L}\|_{\ell^1(\mathbb{Z}^d)} \lesssim_{r,d} N^{\frac{2rC \log \log \log N}{\log \log N}} \|f\|_{\ell^{2r}(\mathbb{Z}^d; H)}^{2r}.$$

Since there are 2^k subsets $L \subseteq [k]$, we may treat each summand above separately at the expense of introducing an acceptable factor $C_r 2^{C_r \log N / \log \log N}$.

We fix L . Set $\mathcal{I}' := \mathcal{I}(\mathcal{S}_{[k] \setminus L}) \cap \mathcal{R}_{\leq N}^d$. Since $\mathbb{1}_{N,k,j}(u + u') = 0$ for every $u \in \mathcal{I}(\mathcal{S}_L)$ and $u' \in \mathcal{I}(\mathcal{S}_{[k] \setminus L}) \setminus \mathcal{R}_{\leq N}^d$, we are reduced to proving the inequality

$$\sum_{x \in \mathbb{Z}^d} \sum_{q \in \mathcal{S}_L} \left(\sum_{u' \in \mathcal{I}'} \left\| \sum_{u \in \mathcal{I}(q)} f_{u+u'}(x) \right\|_H^2 \right)^r \leq C_{r,d}^{2r} N^{\frac{2rC \log \log \log N}{\log \log N}} \|f\|_{\ell^{2r}(\mathbb{Z}^d; H)}^{2r} \quad (3.23)$$

that will be interpreted as a linear operator norm bound from $\ell^{2r}(\mathbb{Z}^d; H)$ to $\ell^{2r}(\mathbb{Z}^d \times \mathcal{S}_L; \ell^2(\mathcal{I}'; H))$. The reduction to (3.23) uses the fact that the sets $\mathcal{I}(q) \cap \mathcal{I}(q')$ are empty for any distinct $q, q' \in \mathcal{S}_{[k] \setminus L}$.

The following lemma is key to establishing (3.23). This is the only place where the expression $\frac{\log \log \log N}{\log \log N}$ appears in the exponent. It would be interesting to determine whether $\frac{1}{\log \log N}$ could be used instead.

Lemma 3.24. *There exists a constant $\mathbf{C}_{\text{div}} \in \mathbb{R}_+$ such that for all $q \in \mathbb{Z}_+$ and $N \in \mathbb{N}_{\geq N_0}$, if q has $k \in [\frac{5 \log N}{\log \log N}]$ prime divisors, then q has at most $\lfloor \mathbf{C}_{\text{div}} N^{\mathbf{C}_{\text{div}} \frac{\log \log \log N}{\log \log N}} \rfloor$ positive divisors not greater than N .*

Proof. Let $\{p_1, \dots, p_k\} \subset \mathbb{P}$ be the set of prime divisors of q . Our aim will be to count of all possible tuples $(\alpha_1, \dots, \alpha_k) \in \mathbb{N}^k$ such that $p_1^{\alpha_1} \cdots p_k^{\alpha_k} \leq N$. We may assume that $\alpha_1 + \dots + \alpha_k \leq \lfloor 2 \log N \rfloor$, since otherwise $p_1^{\alpha_1} \cdots p_k^{\alpha_k} \geq 2^{\lfloor 2 \log N \rfloor + 1} \geq 4^{\log N} > N$. Thus, we have $\max\{\alpha_1, \dots, \alpha_k\} \leq \lfloor 2 \log N \rfloor$. Note that $\mathbb{N}_{\leq 2 \log N}$ can be split into a finite number of intervals of the form

$$I_i := \mathbb{N}_{\leq 2 \log N} \cap [(i-1) \log \log N, i \log \log N),$$

each having about $\log \log N$ elements except the last one, which may be shorter. Now to each tuple $(\alpha_1, \dots, \alpha_k)$ we assign a unique tuple of integers $(i_1, \dots, i_k)$ such that $(\alpha_1, \dots, \alpha_k) \in I_{i_1} \times \dots \times I_{i_k}$. If $(\alpha_1, \dots, \alpha_k) \in I_{i_1} \times \dots \times I_{i_k}$, then

$$(i_1 + \dots + i_k - k) \frac{\log \log N}{2} \leq \alpha_1 + \dots + \alpha_k \leq \lfloor 2 \log N \rfloor,$$

which implies that $i_1 + \dots + i_k \leq k + \frac{4 \log N}{\log \log N} \leq \lfloor \frac{10 \log N}{\log \log N} \rfloor$. Taking

$$\mathcal{A} := \left\{ (\alpha_1, \dots, \alpha_k) \in \mathbb{N}_{\leq 2 \log N}^k : p_1^{\alpha_1} \dots p_k^{\alpha_k} \leq N \right\}$$

we obtain the desired bound

$$\#\mathcal{A} \leq \binom{\lfloor 10 \log N / \log \log N \rfloor + k}{k} [2 \log \log N]^k \leq CN^{C \frac{\log \log \log N}{\log \log N}}.$$

Indeed, the number of tuples $(i_1, \dots, i_k) \in \mathbb{N}^k$ with $i_1 + \dots + i_k \leq \lfloor \frac{10 \log N}{\log \log N} \rfloor$ is at most $\binom{\lfloor 10 \log N / \log \log N \rfloor + k}{k} \leq 2^{15 \log N / \log \log N}$, and for each fixed $(i_1, \dots, i_k)$ the size of $I_{i_1} \times \dots \times I_{i_k}$ is at most $[2 \log \log N]^k \leq CN^{C \frac{\log \log \log N}{\log \log N}}$. $\square$

We say that $\tilde{q} \in \mathbb{Z}_+$ is an *admissible divisor* of q if it divides q , satisfies $\tilde{q} \in [N]$, and has exactly the same prime divisors as q . By Lemma 3.24 we can split $\mathcal{I}(q) \cap \mathcal{R}_{\leq N}^d$ into disjoint sets $\mathcal{R}(q_1), \dots, \mathcal{R}(q_m)$ for some $m \in [\mathbf{C}_{\text{div}} N^{\mathbf{C}_{\text{div}} \frac{\log \log \log N}{\log \log N}}]$, where $q_1, \dots, q_m$ are admissible divisors of q . By the triangle inequality followed by Hölder's inequality, it suffices to show that

$$\sum_{x \in \mathbb{Z}^d} \sum_{q \in \mathcal{S}_L} \left(\sum_{u' \in \mathcal{I}'} \left\| \sum_{u \in \mathcal{R}(\tilde{q})} \tilde{f}_{u+u'}(x) \right\|_H^2 \right)^r \leq C_{r,d}^{2r} N^{2rC \frac{\log \log \log N}{\log \log N}} \|f\|_{\ell^{2r}(\mathbb{Z}^d; H)}^{2r}$$

holds uniformly for all possible choices of admissible divisors $\tilde{q}$ of q , where the function $\tilde{f}_{u+u'}$ is a variant of $f_{u+u'}$ without the indicator function $\mathbb{1}_{N,k,j}$. Indeed, this follows because $\mathbb{1}_{N,k,j}(u+u')$ is constant on each $\mathcal{R}(\tilde{q})$ when $u' \in \mathcal{I}'$ is fixed, and equal to 0 if $u \in \mathcal{I}(q) \setminus \mathcal{R}_{\leq N}^d$.

Let $\phi: \mathbb{R}^d \rightarrow [0, 1]$ be smooth and such that $\mathbb{1}_{[-\frac{6}{5}, \frac{6}{5}]^d} \leq \phi \leq \mathbb{1}_{[-\frac{9}{5}, \frac{9}{5}]^d}$. Let $\psi: \mathbb{R}^d \rightarrow [0, \infty)$ be smooth, supported on $[-\frac{1}{5}, \frac{1}{5}]^d$, and such that $\|\psi\|_{L^1(\mathbb{R}^d)} = 1$. Define $\zeta := \phi * \psi$ so that $\mathbb{1}_{[-1,1]^d} \leq \zeta \leq \mathbb{1}_{[-2,2]^d}$. Proceeding as in [75, Section 2.6], for each $q \in \mathcal{S}_L$ and each admissible divisors $\tilde{q}$ of q , we obtain

$$\sum_{x \in \mathbb{Z}^d} \left(\sum_{u' \in \mathcal{I}'} \left\| \sum_{u \in \mathcal{R}(\tilde{q})} \tilde{f}_{u+u'}(x) \right\|_H^2 \right)^r \leq C_{r,d}^{2r} \sum_{x \in \mathbb{Z}^d} \left(\sum_{u' \in \mathcal{I}'} \left\| \sum_{u \in \mathcal{R}(\tilde{q})} f_{u+u'}^\zeta(x) \right\|_H^2 \right)^r,$$

where $f_{u+u'}^\zeta$ is a variant of $\tilde{f}_{u+u'}$ with $\mathbf{m}$ replaced by a nice bump multiplier ζ .

3.2.8. Square function estimates for the nice bump multiplier. Finally, for each fixed choice of admissible divisors $\tilde{q}$ of $q \in \mathcal{S}_L$, and every $t \in [1, \infty]$, we shall show the following inequality

$$\left(\sum_{x \in \mathbb{Z}^d} \sum_{q \in \mathcal{S}_L} \left(\sum_{u' \in \mathcal{I}'} \left\| \sum_{u \in \mathcal{R}(\tilde{q})} f_{u+u'}^\zeta(x) \right\|_H^2 \right)^t \right)^{\frac{1}{2t}} \leq C_d N^{C \frac{\log \log \log N}{\log \log N}} \|f\|_{\ell^{2t}(\mathbb{Z}^d; H)}$$

that will be interpreted as a linear operator norm bound from $\ell^{2t}(\mathbb{Z}^d; H)$ to $\ell^{2t}(\mathbb{Z}^d \times \mathcal{S}_L; \ell^2(\mathcal{I}'; H))$. By interpolation it suffices to consider $t \in \{1, \infty\}$.

The case $t = 1$ holds with an absolute constant independent of N by Plancherel's theorem. Indeed, the functions $\mathcal{F}_{\mathbb{Z}^d} f_{u+u'}^\zeta$ have disjoint supports, since $\tilde{q} \leq N$, $r \geq 2$, and

$\vartheta \leq (4rN^{2r})^{-1} \leq (8N^4)^{-1}$, and for distinct $q, q' \in \mathcal{S}_L$ the corresponding admissible divisors $\tilde{q}, \tilde{q}'$ have distinct sets of prime divisors.

In the remaining case $t = \infty$, it suffices to show that

$$\sup_{x \in \mathbb{Z}^d} \sup_{q \in \mathcal{S}_L} \sum_{u' \in \mathcal{I}'} \left\| \sum_{u \in \mathcal{Q}(\tilde{q})} f_{u+u'}^\zeta(x) \right\|_H^2 \leq C_d^2 \|f\|_{\ell^\infty(\mathbb{Z}^d; H)}^2 \quad (3.25)$$

holds uniformly for all $\tilde{q} \in [N]$ that are (not necessarily admissible) divisors of q , and once (3.25) is established, the desired inequality with $t = \infty$ follows. Indeed, it suffices to apply the inclusion–exclusion formula involving the Möbius function μ , which asserts that

$$\sum_{u \in \mathcal{R}(\tilde{q})} F(u) = \sum_{b \in \mathcal{D}(\tilde{q})} \mu\left(\frac{\tilde{q}}{b}\right) \sum_{u \in \mathcal{Q}(b)} F(u),$$

where $\mathcal{D}(\tilde{q})$ is the set of all positive divisors of $\tilde{q}$, and to use (3.25) together with Lemma 3.24, which implies that the number of all positive divisors of a given admissible divisor $\tilde{q} \in [N]$ is at most $\lfloor \mathbf{C}_{\text{div}} N^{\mathbf{C}_{\text{div}} \frac{\log \log \log N}{\log \log N}} \rfloor$. In order to prove (3.25), we proceed as in [75, Section 2.7].

3.3. A seminorm variant of the Ionescu–Wainger theorem. The goal of this subsection is to obtain an r -variation variant of Theorem 3.3; see (2.8).

Theorem 3.26. *Let $p'_0 \leq p \leq p_0$ for some $p_0 \in 2\mathbb{Z}_+$ and let $r \in (2, \infty]$. Then there exists a constant $\mathbf{C}_{\text{IW}}(p_0, r) \in \mathbb{R}_+$ such that for every $d \in \mathbb{Z}_+$ and every $N \in \mathbb{N}_{\geq 100}$ the following is true. Assume that $\mathbb{D} \subseteq [1, \infty)$ is λ -lacunary for some $\lambda > 1$; see (2.3). For each $n \in \mathbb{N}$, let*

$$0 < \vartheta_n \leq \min\{(2p_0 N^{p_0})^{-1}, \lambda_n^{-1}\},$$

and let $\mathbf{m}_n: \mathbb{R}^d \rightarrow \mathbb{C}$ be measurable and supported on $[-\vartheta_n, \vartheta_n]^d$. Moreover, set

$$\begin{aligned} \mathbf{A}_{p,r} &:= \|(T_{\mathbb{R}^d}[\mathbf{m}_n])_{n \in \mathbb{N}}\|_{L^p(\mathbb{R}^d) \rightarrow L^p(\mathbb{R}^d; \mathbf{V}^r(\mathbb{N}))}, \\ \mathbf{B}_{p_0} &:= \sup_{(\omega_n)_{n \in \mathbb{N}} \in \{-1, 1\}^{\mathbb{N}}} \left\| \sum_{n \in \mathbb{N}} \omega_n T_{\mathbb{R}^d}[\mathbf{m}_{n+1} - \mathbf{m}_n] \right\|_{L^{p_0}(\mathbb{R}^d) \rightarrow L^{p_0}(\mathbb{R}^d)}. \end{aligned} \quad (3.27)$$

Denote $\rho'_N := \rho'_N(p_0, r) := \mathbf{C}_{\text{IW}}(p_0, r) \frac{\log \log \log N}{\log \log N}$. Then

$$\left\| (T_{\mathbb{Z}^d}^{\mathcal{R}_{\leq N}^d}[\mathbf{m}_n])_{n \in \mathbb{N}} \right\|_{\ell^p(\mathbb{Z}^d) \rightarrow \ell^p(\mathbb{Z}^d; \mathbf{V}^r(\mathbb{N}))} \lesssim_{\lambda, d, p_0, r} N^{\rho'_N} (\mathbf{A}_{p,r} + \mathbf{B}_{p_0}). \quad (3.28)$$

Some remarks about Theorem 3.26 are in order.

1. Theorem 3.26 is a seminorm variant of the Ionescu–Wainger theorem stated at the beginning of this section. We present a fairly general argument, which can be used to deduce similar results for jumps or oscillations (as well as for norms corresponding to real interpolation spaces) in place of r -variations. We refer to [76] for the definitions.
2. The restriction $r \in (2, \infty]$ arises from the fact that the key tool used in our proof is the Rademacher–Menshov inequality from (2.11). It is highly tempting to ask whether Theorem 3.26 remains true for arbitrary seminorms in place of r -variations like the Magyar–Stein–Wainger sampling principle; see Proposition 3.1.

3. In our applications we will have to verify that the quantities $\mathbf{A}_{p,r}$ and $\mathbf{B}_{p_0}$ are finite. It might be a difficult task in general, but we will be working with multipliers $\mathbf{m}_n$ that can be handled using techniques from classical harmonic analysis [87]. To prove that $\mathbf{A}_{p,r}$ is finite we will use a square function argument from [58]. This argument will allow us to replace $(T_{\mathbb{R}^d}[\mathbf{m}_n])_{n \in \mathbb{N}}$ with a dyadic martingale. The latter will be handled using the r -variational inequality of Lépingle for martingales with $r > 2$; see, for instance, [69, 76]. We note that the restriction to $r \in (2, \infty]$ is necessary in this case. To verify that $\mathbf{B}_{p_0}$ is finite we will use classical square function methods in the spirit of Littlewood and Paley; see, for instance, [87, 85, 74].

Proof of Theorem 3.26. Fix a λ -lacunary set $\mathbb{D} \subseteq [1, \infty)$ for some $\lambda > 1$, and let κ_N satisfy $2^{\kappa_N} = 100p_0(\log \lambda)^{-1}N$. We show for all $f \in \ell^p(\mathbb{Z}^d)$ that

$$\left\| \mathbf{V}^r(T_{\mathbb{Z}^d}^{\mathcal{R}_{\leq N}^d}[\mathbf{m}_n]f : n \in \mathbb{N}_{<2^{\kappa_N}}) \right\|_{\ell^p(\mathbb{Z}^d)} \lesssim_{\lambda,d,p_0,r} N^{\rho'_N} \mathbf{B}_{p_0} \|f\|_{\ell^p(\mathbb{Z}^d)}, \quad (3.29)$$

$$\left\| \mathbf{V}^r(T_{\mathbb{Z}^d}^{\mathcal{R}_{\leq N}^d}[\mathbf{m}_n]f : n \in \mathbb{N}_{\geq 2^{\kappa_N}}) \right\|_{\ell^p(\mathbb{Z}^d)} \lesssim_{\lambda,d,p_0,r} N^{\rho'_N} \mathbf{A}_{p,r} \|f\|_{\ell^p(\mathbb{Z}^d)}. \quad (3.30)$$

Combining (3.29) and (3.30), we see that (3.28) readily follows. By using the Rademacher–Menshov inequality from (2.11) followed by Khinchine’s inequality and Theorem 3.3 with $\mathbf{B}_{p_0}$ from (3.27), we obtain (3.29), since $\kappa_N \simeq \log N$. Invoking Proposition 3.1 followed by Theorem 3.3, we obtain (3.30) with $\mathbf{A}_{p,r}$ from (3.27), and the proof is complete. $\square$

4. INVERSE THEOREMS

In this section, we simultaneously prove an inverse theorem for averages over distinct degree corner configurations in both the discrete and continuous settings. This will be essential to establish a multilinear Weyl inequality and a Sobolev smoothing theorem in Section 6. We begin with fixing necessary notation and terminology.

4.1. Basic definitions and the statement of the inverse theorem. Throughout, $\mathbb{K}$ is either the set of integers $\mathbb{Z}$ or the set of real numbers $\mathbb{R}$. If $\mathbb{K} = \mathbb{Z}$, then $\lambda_{\mathbb{K}}$ is counting measure on $\mathbb{Z}$. If $\mathbb{K} = \mathbb{R}$, then $\lambda_{\mathbb{K}}$ is Lebesgue measure on $\mathbb{R}$. Let $\mathbb{K}_+ := \{x \in \mathbb{K} : x > 0\}$. We write $\lambda_{\mathbb{K}^k} := \lambda_{\mathbb{K}}^{\otimes k}$ for the product measure on $\mathbb{K}^k$. For a measurable set $E \subseteq \mathbb{K}^k$, we abbreviate $\lambda_{\mathbb{K}^k}(E)$ to $|E|_{\mathbb{K}^k}$.

Fix $k \in \mathbb{Z}_+$, and consider a polynomial mapping

$$\mathcal{P} := (P_1, \dots, P_k) : \mathbb{K} \rightarrow \mathbb{K}^k, \quad (4.1)$$

where $P_1, \dots, P_k \in \mathbb{K}[n]$ are polynomials with distinct degrees such that

$$d_1 := \deg P_1 < \dots < d_k := \deg P_k, \quad (4.2)$$

and also define

$$D := D_k := d_1 + \dots + d_k. \quad (4.3)$$

For $N \geq 1$, recalling the definitions of $[N]_{\mathbb{K}}$ from (2.2) and $\mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}}$ from (2.5), we define multilinear averages for $f_1, \dots, f_k \in L^0(\mathbb{K}^k, \lambda_{\mathbb{K}^k})$ by

$$A_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i \in [k]} f_i(x - P_i(t)e_i), \quad x \in \mathbb{K}^k, \quad (4.4)$$

and their truncations by

$$\tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}} \setminus [N/2]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i \in [k]} f_i(x - P_i(t)e_i), \quad x \in \mathbb{K}^k. \quad (4.5)$$

For example, in the integer case $\mathbb{K} = \mathbb{Z}$, we have

$$A_{N; \mathbb{Z}^k}^{\mathcal{P}}(f_1, \dots, f_k)(x) = \lfloor N \rfloor^{-1} \sum_{n \in [N]} \prod_{i \in [k]} f_i(x - P_i(n)e_i),$$

whereas, in the real case $\mathbb{K} = \mathbb{R}$, we have

$$A_{N; \mathbb{R}^k}^{\mathcal{P}}(f_1, \dots, f_k)(x) = N^{-1} \int_0^N \prod_{i \in [k]} f_i(x - P_i(t)e_i) dt.$$

Let $\mathcal{C}z := \bar{z}$ for all $z \in \mathbb{C}$. For $f, g \in L^2(\mathbb{K}^k, \lambda_{\mathbb{K}^k})$, define the inner product

$$\langle f, g \rangle := \langle f, g \rangle_{\lambda_{\mathbb{K}^k}} := \int_{\mathbb{K}^k} f(x) \overline{g(x)} d\lambda_{\mathbb{K}^k}(x) = \int_{\mathbb{K}^k} f(x) \mathcal{C}g(x) d\lambda_{\mathbb{K}^k}(x).$$

For compactly supported $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k, \lambda_{\mathbb{K}^k})$, note the identities

$$\left\langle A_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle = \left\langle f_j, A_{N; \mathbb{K}^k}^{\mathcal{P}, *j}(f_1, \dots, f_{j-1}, f_0, f_{j+1}, \dots, f_k) \right\rangle \quad (4.6)$$

and

$$\left\langle \tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle = \left\langle f_j, \tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}, *j}(f_1, \dots, f_{j-1}, f_0, f_{j+1}, \dots, f_k) \right\rangle \quad (4.7)$$

for all $j \in [k]$, where the adjoint operators $A_{N; \mathbb{K}^k}^{\mathcal{P}, *j}$ and $\tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}, *j}$ are given by

$$A_{N; \mathbb{K}^k}^{\mathcal{P}, *j}(g_1, \dots, g_k)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i \in [k]} g_i^{*j}(x - \mathbf{P}_i^{*j}(t)) \quad (4.8)$$

and

$$\tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}, *j}(g_1, \dots, g_k)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}} \setminus [N/2]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i \in [k]} g_i^{*j}(x - \mathbf{P}_i^{*j}(t)) \quad (4.9)$$

with $g_i^{*j} := \mathcal{C}^{1_{i \neq j}} g_i$ and $\mathbf{P}_i^{*j}(t) := 1_{i \neq j} P_i(t)e_i - P_j(t)e_j$. We often abbreviate $A_{N; \mathbb{K}^k}^{\mathcal{P}}$ to $A_{N; \mathbb{K}^k}$ and $\tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}}$ to $\tilde{A}_{N; \mathbb{K}^k}$. In some instances, we will write out the averages $A_{N; \mathbb{K}^k}^{\mathcal{P}} = A_{N; \mathbb{K}^k}^{P_1, \dots, P_k}$ and $\tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}} = \tilde{A}_{N; \mathbb{K}^k}^{P_1, \dots, P_k}$ depending on how explicit we need to be, and likewise with $A_{N; \mathbb{K}^k}^{\mathcal{P}, *j}$ and $\tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}, *j}$ for $j \in [k]$.

We also fix some notation and terminology from the classical circle method. For a finite set $\Sigma \subset \mathbb{K}$ and any $N \in \mathbb{R}_+$, we define the set of major arcs by

$$\mathfrak{M}_{\leq N}(\Sigma) := \bigcup_{\theta \in \Sigma} [\theta - N, \theta + N]. \quad (4.10)$$

Using (4.10), we define the set of major arcs in the j -th component by

$$\mathfrak{M}_{\leq N}^j(\Sigma) := \mathbb{K}^{j-1} \times \mathfrak{M}_{\leq N}(\Sigma) \times \mathbb{K}^{k-j}. \quad (4.11)$$

Recall the definition of the set of canonical fractions $\mathcal{R}_{\leq N}^d$ from (3.2). For every $N \geq 1$, when $d = 1$, we shall abbreviate $\mathcal{R}_{\leq N}^d$ to

$$\mathcal{R}_{\leq N} := \left\{ \frac{a}{q} \in \mathbb{Q}/\mathbb{Z} : q \in [N] \text{ and } \gcd(a, q) = 1 \right\}. \quad (4.12)$$

We shall also write $\mathcal{R}_{\leq N}^{\mathbb{K}} := \mathcal{R}_{\leq N}$ if $\mathbb{K} = \mathbb{Z}$ and $\mathcal{R}_{\leq N}^{\mathbb{K}} := \{0\}$ if $\mathbb{K} = \mathbb{R}$.

Let $M_1 \geq 1$ and $M_2 \in \mathbb{R}_+$, and recall the definition of $T_{i, \mathbb{K}^k}^{\Sigma}[\mathfrak{m}]$ from (2.7). Using this definition, we introduce the Ionescu–Wainger projections by setting

$$\Pi_{\mathbb{K}}^i[\leq M_1, \leq M_2]f(x) := T_{i, \mathbb{K}^k}^{\mathcal{R}_{\leq M_1}^{\mathbb{K}}}[\eta_{[\leq M_2]}]f(x) \quad \text{for } i \in [k], x \in \mathbb{K}^k, \quad (4.13)$$

where $\eta: \mathbb{R} \rightarrow [0, 1]$ is a smooth and even function satisfying (2.4), while

$$\eta_{[\leq M_2]}(x) := \eta(M_2^{-1}x) \quad \text{for } x \in \mathbb{R}.$$

Note that the projections from (4.13) localize the major arcs $\mathfrak{M}_{\leq M_2}^i(\mathcal{R}_{\leq M_1}^{\mathbb{K}})$. For every $p \in [1, \infty]$ and $f \in L^p(\mathbb{K}^k)$, by the triangle inequality, we have

$$\|\Pi_{\mathbb{K}}^i[\leq M_1, \leq M_2]f\|_{L^p(\mathbb{K}^k)} \lesssim \#\mathcal{R}_{\leq M_1}^{\mathbb{K}} \|f\|_{L^p(\mathbb{K}^k)}.$$

This bound is much worse than the bounds obtained in Section 3, but it holds for all $p \in [1, \infty]$ and will be sufficient for the purpose of this section.

The main result of this section is the following inverse theorem.

Theorem 4.14. *Let $\mathbb{K}$ be either $\mathbb{Z}$ or $\mathbb{R}$. Fix $C_0 \in \mathbb{Z}_+$ and $k \in \mathbb{Z}_+$, and let $\mathcal{P}$ be a polynomial mapping satisfying (4.1)–(4.3). Then there exist large constants $C_1, C_2 \in \mathbb{Z}_+$ depending only on C_0 and $\mathcal{P}$ such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C_1\delta^{-C_1}$, and fix $j \in [k]$. If $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ are 1-bounded functions supported on $I := \prod_{i \in [k]} [\pm C_0 N^{d_i}]_{\mathbb{K}}$ such that*

$$\left| \left\langle \tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle \right| \geq \delta N^D, \quad (4.15)$$

then there exists a 1-bounded function $h_j \in L^\infty(\mathbb{K}^k)$ supported on I such that

$$\left| \left\langle f_j, \Pi_{\mathbb{K}}^j[\leq M_1, \leq M_2]h_j \right\rangle \right| \geq C_1^{-1} \delta^{C_1} N^D,$$

where $M_1 := C_2 \delta^{-C_2}$ and $M_2 := C_2 \delta^{-C_2} N^{-d_j}$.

We will prove this theorem in the next few subsections by extending the methods from [80, 81, 83]. In particular, we adapt the strategy from [81, Subsection 7.1] to the multidimensional setting in order to deduce a structural result for each of the 1-bounded functions $f_1, \dots, f_k$ in Theorem 4.14. A new difficulty specific to the multidimensional setting arises during the degree lowering part of the argument, coming from the dependence of phases produced by the U^2 -inverse theorem on variables besides the differencing parameters, which prevents one from simply applying the major arc lemma as in [80]. In [83], the issue could be sidestepped by a simple application of the pigeonhole principle, but this trick is limited to the case

of configurations of the form $(x_1, x_2), (x_1 + P_1(n), x_2), (x_1, x_2 + P_2(n))$ with P_1 linear. To overcome the new difficulty in the general distinct degree case, we perform a more elaborate pigeonholing argument and introduce another application of PET induction combined with concatenation inside the main degree lowering argument.

4.2. Gowers norms and their basic properties. We continue fixing necessary notation. We also gather important tools that will be used later.

4.2.1. Uniform and Fejér measures. Let $k \in \mathbb{Z}_+$. For any measurable $I \subseteq \mathbb{K}^k$ such that $0 < |I|_{\mathbb{K}^k} < \infty$, we define the uniform and Fejér measures by

$$d\lambda_{\mathbb{K}^k, I}(x) := |I|_{\mathbb{K}^k}^{-1} \cdot \mathbf{1}_I(x) d\lambda_{\mathbb{K}^k}(x) \quad \text{and} \quad d\sigma_{\mathbb{K}^k, I}(x) := \kappa_{\mathbb{K}^k, I}(x) d\lambda_{\mathbb{K}^k}(x),$$

respectively, where $\kappa_{\mathbb{K}^k, I}(x)$ is the Fejér kernel defined by

$$\kappa_{\mathbb{K}^k, I}(x) := |I|_{\mathbb{K}^k}^{-2} \cdot \mathbf{1}_I * \mathbf{1}_{-I}(x) = |I|_{\mathbb{K}^k}^{-2} |\{y \in I : x + y \in I\}|_{\mathbb{K}^k}. \quad (4.16)$$

We observe that $\lambda_{\mathbb{K}^k, I}(\mathbb{K}^k) = 1$ and $\sigma_{\mathbb{K}^k, I}(\mathbb{K}^k) = 1$. We will mainly work with $k = 1$, and we shall write $\mathbb{K}$ in place of $\mathbb{K}^1$ in the above formulae. If $\mathbb{K} = \mathbb{R}$, $N \in \mathbb{R}_+$, and $I = [N]_{\mathbb{R}}$, then, by (4.16), one can easily check that the Fejér kernel $\kappa_{\mathbb{R}, I}(x)$ equals $N^{-1}(1 - N^{-1}|x|)$ for $|x| \leq N$, and 0 otherwise.

4.2.2. Multiplicative discrete derivatives. Let $k \in \mathbb{Z}_+$. For every $f \in L^0(\mathbb{K}^k)$ and $x, h, h' \in \mathbb{K}^k$, we define the multiplicative discrete derivatives by

$$\Delta_h f(x) := f(x) \overline{f(x+h)} \quad \text{and} \quad \Delta'_{(h, h')} f(x) := \overline{f(x+h)} f(x+h').$$

For $s \in \mathbb{Z}_+$ and vectors $\mathbf{h} = (h_1, \dots, h_s) \in (\mathbb{K}^k)^s$ and $\mathbf{h}' = (h'_1, \dots, h'_s) \in (\mathbb{K}^k)^s$, we also define the iterated multiplicative discrete derivatives by

$$\begin{aligned} \Delta_{\mathbf{h}} f(x) &:= \Delta_{h_1, \dots, h_s} f(x) := \Delta_{h_1}(\cdots (\Delta_{h_s} f(x)) \cdots), \\ \Delta'_{(\mathbf{h}, \mathbf{h}')} f(x) &:= \Delta'_{(h_1, h'_1), \dots, (h_s, h'_s)} f(x) := \Delta'_{(h_1, h'_1)}(\cdots (\Delta'_{(h_s, h'_s)} f(x)) \cdots). \end{aligned}$$

Remark 4.17. The above formulae may be rewritten more concisely as

$$\begin{aligned} \Delta_{\mathbf{h}} f(x) &= \prod_{\omega \in \{0,1\}^s} \mathcal{C}^{|\omega|} f(x + \omega \circ \mathbf{h}), \\ \Delta'_{(\mathbf{h}, \mathbf{h}')} f(x) &= \prod_{\omega \in \{0,1\}^s} \mathcal{C}^{|\omega|} f(x + \omega \circ \mathbf{h} + (\mathbf{1} - \omega) \circ \mathbf{h}'), \end{aligned} \quad (4.18)$$

where we denote $\mathbf{1} := (1, \dots, 1) \in \{0,1\}^s$ and define $\omega \circ \mathbf{h} := \sum_{i \in [s]} \omega_i h_i$ for every $\omega = (\omega_1, \dots, \omega_s) \in \{0,1\}^s$ and $\mathbf{h} = (h_1, \dots, h_s) \in (\mathbb{K}^k)^s$. If $k = 1$, then $\omega \circ \mathbf{h}$ coincides with the usual inner product $\omega \cdot \mathbf{h}$ on $\mathbb{K}^s$.

4.2.3. The Gowers box and uniformity norms. We use multiplicative discrete derivatives to define the (localized) Gowers box and uniformity norms.

Definition 4.19 (Gowers box and uniformity norms). For $k, s \in \mathbb{Z}_+$ and $n \in [k]$, let $V \equiv \mathbb{K}^n$ be an n -dimensional linear subspace (when $\mathbb{K} = \mathbb{R}$) or sublattice (when $\mathbb{K} = \mathbb{Z}$) of $\mathbb{K}^k$ endowed with the natural measure $\lambda_{\mathbb{K}^n}$. Let I be a subset of $\mathbb{K}^k$ with finite and positive measure with respect to $\lambda_{\mathbb{K}^k}$ and $H_1, \dots, H_s$ be subsets of V with finite and positive measure with respect to $\lambda_{\mathbb{K}^n}$. Setting $\mathbf{H} := \prod_{i \in [s]} H_i$, the normalized Gowers box norm of $f \in L^\infty(\mathbb{K}^k)$ supported on I with respect to the sets of translates $H_1, \dots, H_s$ is given by

$$\|f\|_{\square_{H_1, \dots, H_s}^s(I)} := \left(|I|_{\mathbb{K}^k}^{-1} \int_{\mathbb{K}^k} \mathbb{E}_{\mathbf{h}, \mathbf{h}' \in \mathbf{H}}^{\lambda_{\mathbb{K}^n}^{\otimes 2s}} \Delta'_{(\mathbf{h}, \mathbf{h}')} f(x) d\lambda_{\mathbb{K}^k}(x) \right)^{2^{-s}}. \quad (4.20)$$

After the change of variables $x \mapsto x - h'_1 - \dots - h'_s$, we can rewrite (4.20) as

$$\|f\|_{\square_{H_1, \dots, H_s}^s(I)} := \left(|I|_{\mathbb{K}^k}^{-1} \int_{\mathbb{K}^k} \mathbb{E}_{\mathbf{h}, \mathbf{h}' \in \mathbf{H}}^{\lambda_{\mathbb{K}^n}^{\otimes 2s}} \Delta_{\mathbf{h} - \mathbf{h}'} f(x) d\lambda_{\mathbb{K}^k}(x) \right)^{2^{-s}}.$$

For a subset $J \subset V$ with finite and positive measure with respect to $\lambda_{\mathbb{K}^n}$, the Gowers U^s -norm of $f \in L^\infty(\mathbb{K}^k)$ supported on I with respect to J is given by

$$\|f\|_{U_J^s(I)} := \|f\|_{\square_{J, \dots, J}^s(I)}. \quad (4.21)$$

We make some remarks about Definition 4.19.

1. Although it is not immediately obvious, the quantity on the right-hand side of (4.20) is nonnegative and the box norms satisfy the triangle inequality. Proofs of these standard facts and all other statements in this subsection can be found in [44, Appendix B]; see also [28] for generalizations.
2. After the change of variables $x \mapsto x - h'_1 - \dots - h'_s$ in (4.20), we can rewrite the 2^s -th power of the local Gowers box norm as

$$\|f\|_{\square_{H_1, \dots, H_s}^{2^s}(I)} = |I|_{\mathbb{K}^k}^{-1} \int_{\mathbb{K}^k} \mathbb{E}_{h_1 \in V}^{\sigma_{\mathbb{K}^n, H_1}} \dots \mathbb{E}_{h_s \in V}^{\sigma_{\mathbb{K}^n, H_s}} \Delta_{h_1, \dots, h_s} f(x) d\lambda_{\mathbb{K}^k}(x). \quad (4.22)$$

In fact, our arguments will more often feature (4.22) than (4.20).

3. For every $m \in [s]$, from (4.18) and (4.22) we immediately see that

$$\|f\|_{\square_{H_1, \dots, H_s}^{2^s}(I)} = \mathbb{E}_{h_{m+1} \in V}^{\sigma_{\mathbb{K}^n, H_{m+1}}} \dots \mathbb{E}_{h_s \in V}^{\sigma_{\mathbb{K}^n, H_s}} \|\Delta_{h_{m+1}, \dots, h_s} f\|_{\square_{H_1, \dots, H_m}^{2^m}(I)}. \quad (4.23)$$

We similarly relate the Gowers U^{s+1} -norm to the Gowers U^s -norm.

4. Finally, we emphasize that we will not usually need the full generality of the multidimensional definitions of (4.20) or (4.21). We will mainly work with the case $k \in \mathbb{Z}_+$ and $n = 1$. To be more specific, we will use (4.20) or (4.21) with $V = \mathbb{K}e_j$ for some $j \in [k]$, where $\mathbb{K}e_j := \{te_j : t \in \mathbb{K}\}$ and $\{e_i : i \in [k]\}$ is the standard basis of $\mathbb{K}^k$.

We can similarly define the Gowers box inner product.

Definition 4.24 (Gowers box inner products). For $k, s \in \mathbb{Z}_+$ and $n \in [k]$, let $V, I, H_1, \dots, H_s$ be as in Definition 4.19. For every $\omega \in \{0, 1\}^s$ let $f_\omega \in L^\infty(\mathbb{K}^k)$ be supported on a set I_ω with

$|I_\omega|_{\mathbb{K}^k} = |I|_{\mathbb{K}^k}$. The Gowers box inner product of $(f_\omega)_{\omega \in \{0,1\}^s}$ with respect to $H_1, \dots, H_s$ is given by

$$\langle (f_\omega)_{\omega \in \{0,1\}^s} \rangle_{H_1, \dots, H_s}^I := |I|_{\mathbb{K}^k}^{-1} \int_{\mathbb{K}^k} \mathbb{E}_{\mathbf{h}, \mathbf{h}' \in \mathbf{H}}^{\lambda_{\mathbb{K}^n}^{\otimes 2s}} \prod_{\omega \in \{0,1\}^s} \mathcal{C}^{|\omega|} f_\omega(x + [\mathbf{h}, \mathbf{h}']_\omega) d\lambda_{\mathbb{K}^k}(x),$$

where $\mathbf{H} := \prod_{i \in [s]} H_i$ and $[\mathbf{h}, \mathbf{h}']_\omega := \omega \circ \mathbf{h} + (1 - \omega) \circ \mathbf{h}'$ for all $\mathbf{h}, \mathbf{h}' \in \mathbf{H}$ and $\omega \in \{0,1\}^s$. After the change of variables $x \mapsto x - h'_1 - \dots - h'_s$, we can rewrite this as

$$|I|_{\mathbb{K}^k}^{-1} \int_{\mathbb{K}^k} \mathbb{E}_{\mathbf{h} \in V^s}^{\sigma_{\mathbb{K}^n, H_1} \otimes \dots \otimes \sigma_{\mathbb{K}^n, H_s}} \prod_{\omega \in \{0,1\}^s} \mathcal{C}^{|\omega|} f_\omega(x + \omega \circ \mathbf{h}) d\lambda_{\mathbb{K}^k}(x).$$

Iterating the Cauchy–Schwarz inequality yields the following inequality, which is a local version of [93, Equation 11.6].

Proposition 4.25 (The Gowers–Cauchy–Schwarz inequality). *Let $V, I, H_1, \dots, H_s, \mathbf{H}$ and $(f_\omega)_{\omega \in \{0,1\}^s}, (I_\omega)_{\omega \in \{0,1\}^s}$ be as in Definition 4.24. Then we have*

$$\left| \langle (f_\omega)_{\omega \in \{0,1\}^s} \rangle_{H_1, \dots, H_s}^I \right| \leq \prod_{\omega \in \{0,1\}^s} \|f_\omega\|_{\square_{H_1, \dots, H_s}^s(I_\omega)}. \quad (4.26)$$

Finally, we will require a general version of the U^2 -inverse theorem. For $k \in \mathbb{Z}_+$ and the Pontryagin dual $\hat{\mathbb{K}}^k$ (i.e., $\mathbb{R}^k = \hat{\mathbb{R}}^k$ or $\mathbb{T}^k = \hat{\mathbb{Z}}^k$), let $\lambda_{\hat{\mathbb{K}}^k}$ be Lebesgue measure on $\hat{\mathbb{K}}^k$. A proof of the following standard result can be found in [61, Lemma 5.1] in the case $\mathbb{K} = \mathbb{R}$, and a proof of a quantitatively weaker version can be found in [80, Lemma 2.4] in the case $\mathbb{K} = \mathbb{Z}$.

Proposition 4.27 (U^2 -inverse theorem). *Let $I, H \subset \mathbb{K}$ be intervals with $0 < |I|_{\mathbb{K}}, |H|_{\mathbb{K}} < \infty$. If $f \in L^\infty(\mathbb{K})$ is 1-bounded and supported on I , then*

$$\|f\|_{U_H^2(I)}^4 \leq |H|_{\mathbb{K}}^{-2} \|\mathcal{F}_{\mathbb{K}} f\|_{L^\infty(\hat{\mathbb{K}}, \lambda_{\hat{\mathbb{K}}})}^2.$$

4.3. The main technical reduction. The purpose of this subsection is to formulate a slightly stronger variant of Theorem 4.14, and then to show that the conclusion of the theorem for $j = 1$ implies the same conclusion for general $j \in [k]$. To do this, we will need the notion of admissible polynomials.

Definition 4.28. Fix $0 < \delta \leq 1$, $N \in [1, \infty)$, and $d \in \mathbb{Z}_+$. Let $Q(t) = \sum_{i=0}^d c_i(Q) t^i \in \mathbb{K}[t]$ be a polynomial of degree d with the leading coefficient $\ell(Q) := c_d(Q)$. We say that Q is (d, δ, N) -admissible with tolerance $A \in [1, \infty)$ if its coefficients satisfy

$$A^{-1} \delta^A \leq |\ell(Q)| \leq A \delta^{-A} \quad \text{and} \quad |c_i(Q)| \leq A \delta^{-A} N^{d-i} \quad \text{for } i \in \mathbb{N}_{< d}. \quad (4.29)$$

When the value of A is understood, we just say that polynomials Q that satisfy (4.29) are (d, δ, N) -admissible. Now, we formulate a quantitatively uniform extension of Theorem 4.14 for (d, δ, N) -admissible polynomials.

Theorem 4.30. Fix $A \in [1, \infty)$, $C_0 \in \mathbb{Z}_+$, and $k \in \mathbb{Z}_+$, and let $\mathcal{P}$ be a polynomial mapping satisfying (4.1)–(4.3). Then there exist large constants $C_1, C_2 \in \mathbb{Z}_+$ depending on A, C_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C_1 \delta^{-C_1}$, and fix $j \in [k]$. If

each P_i is (d_i, δ, N) -admissible with tolerance A , and $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ are 1-bounded functions supported on $I := \prod_{i \in [k]} [\pm C_0 N^{d_i}]_{\mathbb{K}}$ such that

$$\left| \left\langle A_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle \right| \geq \delta N^D, \quad (4.31)$$

then there exists a 1-bounded function $h_j \in L^\infty(\mathbb{K}^k)$ supported on I such that

$$\left| \left\langle f_j, \Pi_{\mathbb{K}}^j[\leq M_1, \leq M_2] h_j \right\rangle \right| \geq C_1^{-1} \delta^{C_1} N^D,$$

where $M_1 := C_2 \delta^{-C_2}$ and $M_2 := C_2 \delta^{-C_2} N^{-d_j}$.

A few comments about this theorem are in order.

1. A fixed polynomial $Q \in \mathbb{K}[t]$ of degree $d \in \mathbb{Z}_+$ is (d, δ, N) -admissible with tolerance $\max\{|\ell(Q)|, |\ell(Q)|^{-1}\}$ for all $0 < \delta \leq 1$ and $N \geq C_1 \delta^{-C_1}$ if C_1 is sufficiently large in terms of the coefficients of Q . Thus, the entries P_i of $\mathcal{P}$ from (4.1) are (d_i, δ, N) -admissible with tolerance $O_{\mathcal{P}}(1)$.
2. In view of the previous remark, Theorem 4.30 implies Theorem 4.14. Indeed, condition (4.15) for the truncated averages $\tilde{A}_{N; \mathbb{K}^k}^{\mathcal{P}}$ implies that

$$\left| \left\langle A_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle \right| \gtrsim \delta N^D \quad \text{or} \quad \left| \left\langle A_{N/2; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle \right| \gtrsim \delta N^D$$

holds, which, in turn, gives (4.31) with comparable N and δ . Theorem 4.30 then yields the desired conclusion of Theorem 4.14.

3. The subsections that follow are devoted to proving Theorem 4.30 in the special case $j = 1$. The aim of this subsection is to show that the general case $j \in [k]$ can be derived inductively from this special case.

Next, we introduce some convenient notation and important terminology that will facilitate working in high dimensions.

Definition 4.32. For $l, m, n \in \mathbb{Z}_+$ and an l -dimensional vector $x \in \mathbb{K}^l$, let

$$\begin{aligned} x_{\leq l} &:= x_{< l+1} := (x_1, \dots, x_l) \in \mathbb{K}^l, \\ x_{> m} &:= x_{\geq m+1} := (x_{m+1}, \dots, x_{m+l}) \in \mathbb{K}^l, \\ x_{(m, n]} &:= (x_{m+1}, \dots, x_n) \in \mathbb{K}^l \quad \text{if } n - m = l, \\ x_{(m, n]^c} &:= (x_1, \dots, x_m, x_{n+1}, \dots, x_{n+l-m}) \in \mathbb{K}^l \quad \text{if } m < \min\{l, n+1\}. \end{aligned}$$

We use this notation whenever we need to emphasize that the entries of the underlying vector $x \in \mathbb{K}^l$, whose dimension l is clear from the context, are indexed differently from the standard convention $x = (x_1, \dots, x_l)$. In particular, for $m \in [l+1]$, we define $x_{\{m\}^c}$ by setting

$$x_{\{m\}^c} := x_{(m-1, m]^c} := (x_1, \dots, x_{m-1}, x_{m+1}, \dots, x_{l+1}) \in \mathbb{K}^l,$$

with obvious modifications of indexing parameters when $m \in \{1, l+1\}$.

4.3.1. **Proof of Theorem 4.30 assuming Theorem 4.30 for $j = 1$.** We now demonstrate that it is enough to prove Theorem 4.30 for $j = 1$, as the remaining cases $j \in [k] \setminus \{1\}$ can be derived from this case. The statement of Theorem 4.30 for $j = 1$ is as follows.

Theorem 4.33. *Fix $A \in [1, \infty)$, $C_0 \in \mathbb{Z}_+$, and $k \in \mathbb{Z}_+$, and let $\mathcal{P}$ be a polynomial mapping satisfying (4.1)–(4.3). Then there exist large constants $C_{1,1}, C_{1,2} \in \mathbb{Z}_+$ depending on A, C_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C_{1,1}\delta^{-C_{1,1}}$. If each P_i is (d_i, δ, N) -admissible with tolerance A , and $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ are 1-bounded functions supported on $I := \prod_{i \in [k]} [\pm C_0 N^{d_i}]_{\mathbb{K}}$ such that*

$$\left| \left\langle A_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle \right| \geq \delta N^D, \quad (4.34)$$

then we have

$$\left| \left\langle f_1, \Pi_{\mathbb{K}}^1[\leq M_1, \leq M_2] f_1 \right\rangle \right| \geq C_{1,1}^{-1} \delta^{C_{1,1}} N^D, \quad (4.35)$$

where $M_1 := C_{1,2} \delta^{-C_{1,2}}$ and $M_2 := C_{1,2} \delta^{-C_{1,2}} N^{-d_1}$.

Remark 4.36. Theorem 4.33 remains true if we assume that $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ are 1-bounded functions supported on $I := \prod_{i \in [k]} [\pm C_0 \delta^{-C_0} N^{d_i}]_{\mathbb{K}}$.

Assuming momentarily that Theorem 4.33 holds, we derive Theorem 4.30. The proof of Theorem 4.33 will be presented in the next two subsections.

Proof of Theorem 4.30. Assume that Theorem 4.33 holds for all $k \in \mathbb{Z}_+$. Fix $k \in \mathbb{Z}_+$ and $j \in [k]$ as in Theorem 4.30. By Theorem 4.33, we can assume that $k > 1$ and $j > 1$. Set $D_i := d_1 + \dots + d_i$ for all $i \in [k]$. Furthermore, fix 1-bounded functions $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ supported on I . We will proceed in two steps, assuming that $N \geq C_* \delta^{-C_*}$ for some large $C_* \in \mathbb{Z}_+$.

Step 1. Applying (4.31), (4.6), the Cauchy–Schwarz inequality, and then (4.6) again, we obtain

$$\delta N^D \leq \left| \left\langle f_1, A_{N; \mathbb{K}^k}^{\mathcal{P}, *1}(f_0, f_2, \dots, f_k) \right\rangle \right| \lesssim N^{D/2} \left| \left\langle f_0, A_{N; \mathbb{K}^k}^{\mathcal{P}}(F_1, f_2, \dots, f_k) \right\rangle \right|^{1/2}$$

with $F_1 := A_{N; \mathbb{K}^k}^{\mathcal{P}, *1}(f_0, f_2, \dots, f_k)$, given explicitly by the formula

$$F_1(x) := \mathbb{E}_{y \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} f_0(x + P_1(y)e_1) \prod_{i \in [k] \setminus \{1\}} C f_i(x - P_i(y)e_i + P_1(y)e_1)$$

for all $x \in \mathbb{K}^k$. Now, by Theorem 4.33, there exist large $C_{1,1}, C_{1,2} \in \mathbb{Z}_+$ such that, whenever $N \geq C_{1,1} \delta^{-C_{1,1}}$, we have

$$\left| \left\langle F_1, \Pi_{\mathbb{K}}^1[\leq M_1, \leq M_2] F_1 \right\rangle \right| \geq C_{1,1}^{-1} \delta^{C_{1,1}} N^D,$$

where $M_1 := C_{1,2} \delta^{-C_{1,2}}$ and $M_2 := C_{1,2} \delta^{-C_{1,2}} N^{-d_1}$, which is equivalent to

$$\left| \left\langle f_0, A_{N; \mathbb{K}^k}^{\mathcal{P}}(\Pi_{\mathbb{K}}^1[\leq M_1, \leq M_2] F_1, f_2, \dots, f_k) \right\rangle \right| \geq C_{1,1}^{-1} \delta^{C_{1,1}} N^D. \quad (4.37)$$

Step 2. By the pigeonhole principle, at the expense of worsening the lower bound in (4.37) by multiplying it by a large power of δ , we can consider the averages over shorter intervals in (4.37). Then, by the mean value theorem, one can replace the function $\Pi_{\mathbb{K}}^1[\leq M_1, \leq M_2]F_1$ with a constant function. It follows that there exists a measurable set $X_1 \subseteq [C_0 N^{d_1}]_{\mathbb{K}}$ with $\lambda_{\mathbb{K}}(X_1) \gtrsim \delta^{O(1)} N^{d_1}$ such that, for every $x_1 \in X_1$, we have

$$\left| \left\langle f_{0,x_1}^1, A_{N_1, \mathbb{K}^{k-1}}^{\mathcal{Q}_{>1}}(f_{2,x_1}, \dots, f_{k,x_1}) \right\rangle \right| \gtrsim \delta^{O(1)} N_1^{D_k - D_1},$$

where $\delta^{O(1)} N \lesssim N_1 \lesssim \delta^{-O(1)} N$ and $\mathcal{Q}_{>1} := \{Q_2, \dots, Q_k\}$ is a family of polynomials Q_i which are (d_i, δ, N) -admissible with tolerance $A_i \lesssim_{A, C_{1,1}, d_k} 1$, with $f_{0,x_1}^1(x_{>1}) := f_0^1(x_1, x_{>1})$ and $f_{i,x_1}(x_{>1}) := f_i(x_1, x_{>1})$ for all $i \in [k] \setminus \{1\}$ and $x_{>1} \in \mathbb{K}^{k-1}$. If $j = 2$, then we apply Theorem 4.33 for each $x_1 \in X_1$ (or, more precisely, we apply the result given in Remark 4.36, since the functions $f_{0,x_1}^1, f_{2,x_1}, \dots, f_{k,x_1}$ are supported on $\prod_{i \in [k]} [\pm C_{2,0} \delta^{-C_{2,0}} N_1^{d_i}]_{\mathbb{K}}$ for some $C_{2,0} \in \mathbb{Z}_+$) to conclude that there exist constants $C_{2,1}, C_{2,2} \in \mathbb{Z}_+$ depending on A, C_0, d_k such that, whenever $N \geq C_{2,1} \delta^{-C_{2,1}}$, we have

$$\left| \left\langle f_{2,x_1}, \Pi_{\mathbb{K}}^2[\leq M_1, \leq M_2] f_{2,x_1} \right\rangle \right| \geq C_{2,1}^{-1} \delta^{C_{2,1}} N^{D_k - D_1},$$

where $M_1 := C_{2,2} \delta^{-C_{2,2}}$ and $M_2 := C_{2,2} \delta^{-C_{2,2}} N^{-d_2}$. Integrating the above inequality over $x_1 \in X_1$ and using duality, we obtain

$$\left| \left\langle f_2, \Pi_{\mathbb{K}}^2[\leq M_1, \leq M_2] h_2 \right\rangle \right| \gtrsim \delta^{O(1)} N^{D_k},$$

where $h_2(x) := f_2(x) g_2(x_1)$ for some 1-bounded $g_2 \in L^\infty(\mathbb{K})$. This shows that Theorem 4.30 holds for $j = 2$, and iterating this argument, we conclude that it holds for all $j \in [k]$ with constants $C_{j,1}, C_{j,2} \in \mathbb{Z}_+$ in place of C_1 and C_2 , respectively. Finally, the implied constants $C_{j,1}, C_{j,2}$ for each $j \in [k]$ can be taken to be uniform in $j \in [k]$. This completes the proof of Theorem 4.30. $\square$

4.4. Gowers norm control. In this subsection, we will begin the preparation to prove Theorem 4.33. The first step involves the PET induction technique, which is a method going back to the work of Bergelson and Leibman [8], with a coefficient tracking scheme as in [80]. The primary tool of PET induction is the following variant of van der Corput's inequality.

Lemma 4.38. *Fix $\mathfrak{g} \in L^1(\mathbb{K})$ and an interval $J \subset \mathbb{K}$. Then*

$$\left| \mathbb{E}_{y \in J}^{\lambda_{\mathbb{K}}} \mathfrak{g}(y) \right|^2 \leq 2 \mathbb{E}_{y \in J}^{\lambda_{\mathbb{K}}} \mathbb{E}_{h \in \mathbb{K}}^{\sigma_{\mathbb{K}, [H]_{\mathbb{K}}}} \mathbf{1}_{J-h}(y) \Delta_h \mathfrak{g}(y)$$

for every $H \in \mathbb{R}_+$ such that $1 \leq H \leq |J|_{\mathbb{K}}$.

Proof. We first insert an extra average in $h \in [H]_{\mathbb{K}}$ and then apply the Cauchy–Schwarz inequality. $\square$

An important consequence of Lemma 4.38 is the following simple lemma.

Lemma 4.39. *Fix an interval $J \subset \mathbb{K}$ and $H \in \mathbb{R}_+$ such that $1 \leq H \leq |J|_{\mathbb{K}}$. If $\mathfrak{g}_1 \in L^\infty(\mathbb{K}^k)$ and $\mathfrak{g}_2 \in L^\infty(\mathbb{K}^{k+1})$ are 1-bounded functions such that*

$$\|\mathfrak{g}_1\|_{L^1(\mathbb{K}^k)} \leq \alpha \quad \text{and} \quad \sup_{y \in \mathbb{K}} \|\mathfrak{g}_2(\cdot, y)\|_{L^1(\mathbb{K}^k)} \leq \alpha \quad (4.40)$$

for some $\alpha \in \mathbb{R}_+$, then one can find $\eta \in \mathbb{C}$ such that $|\eta| \leq 2H|J|_{\mathbb{K}}^{-1}$ and

$$\left| \frac{1}{\alpha} \int_{\mathbb{K}^k} \mathbf{g}_1(x) \mathbb{E}_{y \in J}^{\lambda_{\mathbb{K}}} \mathbf{g}_2(x, y) d\lambda_{\mathbb{K}^k}(x) \right|^2$$

does not exceed the following nonnegative expression

$$\frac{2}{\alpha} \int_{\mathbb{K}^k} \mathbb{E}_{y \in J}^{\lambda_{\mathbb{K}}} \mathbb{E}_{h \in \mathbb{K}}^{\sigma_{\mathbb{K}, [H]_{\mathbb{K}}}} \mathbf{g}_2(x, y) \overline{\mathbf{g}_2(x, y + h)} d\lambda_{\mathbb{K}^k}(x) + \eta.$$

Proof. We first apply the Cauchy–Schwarz inequality and Lemma 4.38, and then observe that $\mathbb{1}_{J-h}$ can be replaced by $\mathbb{1}_J$ in the resulting bound at the cost of an error term $\eta \in \mathbb{C}$ satisfying $|\eta| \leq 2H|J|_{\mathbb{K}}^{-1}$. $\square$

Fix $L \in \mathbb{Z}_+$ and a family $\{\mathbf{Q}_l : l \in [L]\}$ of polynomial vectors $\mathbf{Q}_l := (Q_{l,1}, \dots, Q_{l,k})$. We say that $\mathbf{Q}_l$ is *constant* if each component $Q_{l,i}$ is constant.

Proposition 4.41. *Fix $N \in [1, \infty)$, $k \in \mathbb{Z}_+$, $L \in \mathbb{Z}_+$, $l_0 \in [L]$, $\alpha \in \mathbb{R}_+$, and a family $\{\mathbf{Q}_l : l \in [L]\}$ of polynomial vectors that are not constant. If $\mathbf{f}_0, \mathbf{f}_1, \dots, \mathbf{f}_L \in L^0(\mathbb{K}^k)$ are 1-bounded, with $L^1(\mathbb{K}^k)$ norms bounded by α , and*

$$\left| \frac{1}{\alpha} \int_{\mathbb{K}^k} \mathbf{f}_0(x) \mathbb{E}_{y \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{l \in [L]} \mathbf{f}_l(x - \mathbf{Q}_l(y)) d\lambda_{\mathbb{K}^k}(x) \right| \geq \varepsilon$$

for some $\varepsilon \in (0, 1]$, then for every $\varepsilon_0 \in (0, \varepsilon^2/10]$ with $\varepsilon_0 \varepsilon^2 N \geq 10^3$, we have

$$\left| \frac{1}{\alpha} \int_{\mathbb{K}^k} \mathbb{E}_{y \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{l \in [L]} \mathbf{f}_l(x - \mathbf{Q}_l(y)) \overline{\mathbf{f}_l(x - \mathbf{Q}_l(y + h))} d\lambda_{\mathbb{K}^k}(x) \right| \geq \frac{\varepsilon^2}{10} \quad (4.42)$$

for all h belonging to a measurable set $\mathcal{H} \subseteq [\pm \varepsilon_0 N]_{\mathbb{K}} \setminus [\pm \varepsilon_0 \varepsilon^2 N/10^2]_{\mathbb{K}}$ such that $|\mathcal{H}|_{\mathbb{K}} \geq \varepsilon_0 \varepsilon^2 N/10^2$. Thus, for all $h \in \mathcal{H}$, substituting $x \mapsto x + \mathbf{Q}_{l_0}(y)$ yields

$$\left| \frac{1}{\alpha} \int_{\mathbb{K}^k} \mathbf{f}'_0(x) \mathbb{E}_{y \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{l' \in [L']} \mathbf{f}'_{l'}(x - \mathbf{Q}'_{l'}(y)) d\lambda_{\mathbb{K}^k}(x) \right| \geq \frac{\varepsilon^2}{10}$$

for some integer $L' \leq 2L - 1$, where $\mathbf{f}'_0(x)$ is the product of all factors independent of y , the new family $\{\mathbf{Q}'_{l'} : l' \in [L']\}$ of polynomial vectors consists exactly of those $\mathbf{Q}_l - \mathbf{Q}_{l_0}$ and $\mathbf{Q}_l(\cdot + h) - \mathbf{Q}_{l_0}$ that are not constant, $\mathbf{f}'_{l'}$ is either $\mathbf{f}_l$ when $\mathbf{Q}'_{l'} = \mathbf{Q}_l - \mathbf{Q}_{l_0}$ or $\overline{\mathbf{f}_l}$ when $\mathbf{Q}'_{l'} = \mathbf{Q}_l(\cdot + h) - \mathbf{Q}_{l_0}$, and, consequently, $\mathbf{f}'_0, \mathbf{f}'_1, \dots, \mathbf{f}'_{L'}$ are 1-bounded, with $L^1(\mathbb{K}^k)$ norms bounded by α .

Proof. It suffices to apply Lemma 4.39 with $\mathbf{g}_1(x) = \mathbf{f}_0(x)$ and $\mathbf{g}_2(x, y) = \prod_{l \in [L]} \mathbf{f}_l(x - \mathbf{Q}_l(y))$, noting that $\mathbf{g}_1, \mathbf{g}_2$ are 1-bounded and satisfy (4.40). $\square$

Repeated applications of Proposition 4.41 according to the PET induction scheme allow us to control the left-hand side of (4.34) by Gowers box norms.

Lemma 4.43. *Under the assumptions of Theorem 4.33, we have*

$$\mathbb{E}_{\mathbf{h} \in \mathbb{K}^T}^{\mu_{\mathbb{K}^T, H_T}} \|f_k\|_{\square_{c_1(\mathbf{h})[\pm H']_{\mathbb{K}^{e_k}}, \dots, c_s(\mathbf{h})[\pm H']_{\mathbb{K}^{e_k}}}}^{2^s}(I) \geq C^{-1} \delta^C \quad (4.44)$$

for some integers $s, T, C \geq 2$ bounded in terms of A, C_0, d_k , some nonzero multilinear polynomials $c_1(\mathbf{h}), \dots, c_s(\mathbf{h})$ of degree $d_k - 1$ with coefficients in $\mathbb{K}$ that are bounded in terms of A and d_k , some $H' \in \mathbb{K}$ such that $H' \geq C^{-1}\delta^C N$, and $\mu_{\mathbb{K}^T, \mathbf{H}_T} := \lambda_{\mathbb{K}, \mathcal{H}_1} \otimes \dots \otimes \lambda_{\mathbb{K}, \mathcal{H}_T}$, where each $\mathcal{H}_i$ is an interval of the form $[\pm H]_{\mathbb{K}}$ for some $H \in \mathbb{K}$ such that $H \geq C^{-1}\delta^C N$.

In fact, Kravitz, Kuca, and Leng [63] have recently proven this exact Gowers box norm control result in the case $\mathbb{K} = \mathbb{Z}$ when $P_1, \dots, P_k$ have bounded coefficients (as a special case of their Proposition 4.8). Their argument goes through with no change when $P_1, \dots, P_k$ are merely admissible, and thus Lemma 4.43 holds for this more general class of polynomials we consider here, in both cases $\mathbb{K} = \mathbb{Z}$ and $\mathbb{K} = \mathbb{R}$. We omit the details.

When $\mathbb{K} = \mathbb{Z}$, averages of Gowers box norms such as those appearing on the left-hand side of (4.44) can be bounded in terms of U^s -norms using *concatenation theorems*. The first concatenation theorem was proven by Tao and Ziegler in [94] for the purpose of obtaining asymptotics for polynomial progressions in the primes. Although it was applicable in very broad generality, their concatenation theorem was purely qualitative, as the proof of Tao and Ziegler produced no explicit bounds. Peluse and Prendiville [82] (for the averages of Gowers box norms arising from the nonlinear Roth configuration) and Peluse [80] (for more general averages of Gowers box norms) proved the first quantitative concatenation theorems, which had polynomial bounds. Applying an even more general quantitative concatenation theorem of Kravitz, Kuca, and Leng [63, Theorem 1.10] (see also [64]) with polynomial bounds proves the following key theorem in the case $\mathbb{K} = \mathbb{Z}$.

Theorem 4.45. *Fix $A \in [1, \infty)$, $C_0 \in \mathbb{Z}_+$, and $k \in \mathbb{Z}_+$, and let $\mathcal{P}$ be a polynomial mapping satisfying (4.1)–(4.3). Then there exists a large constant $C_* \in \mathbb{Z}_+$ depending only on A, C_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C_*\delta^{-C_*}$. If each P_i is (d_i, δ, N) -admissible with tolerance A , and $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ are 1-bounded functions supported on $I := \prod_{i \in [k]} [\pm C_0\delta^{-C_0}N^{d_i}]_{\mathbb{K}}$, then we have the following implication*

$$\left| \left\langle A_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k), f_0 \right\rangle \right| \geq \delta N^D \implies \|f_k\|_{U_{[\pm H_*]_{\mathbb{K}} e_k}^s(I)} \geq C_*^{-1} \delta^{C_*}, \quad (4.46)$$

with $H_* := 7C_0\delta^{-C_0}N^{d_k}$ and an integer $s \geq 2$ bounded in terms of A, C_0, d_k .

To deduce Theorem 4.45 from Lemma 4.43 in the case $\mathbb{K} = \mathbb{R}$, one can simply use the pigeonhole principle to find $\mathbf{h}$ for which $c_1(\mathbf{h}), \dots, c_s(\mathbf{h}) \gtrsim_{A, d_k} \delta^{O_{A, d_k}(1)} N^{d_k-1}$ and

$$\|f_k\|_{\square_{c_1(\mathbf{h})[\pm H']_{\mathbb{K}} e_k, \dots, c_s(\mathbf{h})[\pm H']_{\mathbb{K}} e_k}^{2^s(I)} \gtrsim_{A, C_0, d_k} \delta^{O_{A, d_k}(1)},$$

apply suitable changes of variables to replace $c_i(\mathbf{h})[\pm H']_{\mathbb{K}}$ by $[\pm H_i]_{\mathbb{K}}$, where $H_i \simeq_{A, C_0, d_k} \delta^{O_{A, d_k}(1)} N^{d_k}$ for each $i \in [s]$, and then use the Gowers–Cauchy–Schwarz inequality (4.26) and the support of f_k to replace each $[\pm H_i]_{\mathbb{K}}$ by two intervals $[\pm H_*]_{\mathbb{K}}$ of uniform length, thereby also replacing s by $s+1$. The U^s -norm appearing on the left-hand side of (4.46) can be interpreted as an average of one-dimensional U^s -norms with respect to the last variable:

$$\|f_k\|_{U_{[\pm H_*]_{\mathbb{K}} e_k}^s(I)}^{2^s} = |I_{<k}|_{\mathbb{K}^{k-1}}^{-1} \int_{\mathbb{K}^{k-1}} \|f_k(x_{<k}, \cdot)\|_{U_{[\pm H_*]_{\mathbb{K}}}^s(I_k)}^{2^s} d\lambda_{\mathbb{K}^{k-1}}(x_{<k}).$$

4.5. Degree lowering. In this subsection, we carry out the degree lowering argument, following the broad outline of [80, Sections 7 and 8] while overcoming new difficulties that arise in the multidimensional setting.

As before, $\mathbb{K}$ is either $\mathbb{Z}$ or $\mathbb{R}$ with the corresponding dual groups $\hat{\mathbb{Z}} = \mathbb{T}$ (identified with $[-1/2, 1/2)$) and $\hat{\mathbb{R}} = \mathbb{R}$. The dual group $\hat{\mathbb{K}}$ is endowed with a canonical norm: $\|\xi\|_{\hat{\mathbb{K}}} := \text{dist}(\xi, \mathbb{Z})$ when $\mathbb{K} = \mathbb{Z}$ and $\|\xi\|_{\hat{\mathbb{K}}} := |\xi|$ when $\mathbb{K} = \mathbb{R}$.

4.5.1. Degree lowering preparation. First, we collect the tools used in the degree lowering argument. The following result is the inverse form of Weyl's inequality when $\mathbb{K} = \mathbb{Z}$, and of van der Corput's lemma when $\mathbb{K} = \mathbb{R}$.

Proposition 4.47. *Fix $P \in \mathbb{K}[t]$ of degree $d \in \mathbb{Z}_+$, of the form $P(t) = \xi_d t^d + \dots + \xi_1 t + \xi_0$. Then there exists a large constant $C \in \mathbb{Z}_+$ depending only on d such that, for every $\varepsilon \in (0, 1]$ and $N \geq C\varepsilon^{-C}$, the following holds. If*

$$\left| \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} e(P(t)) \right| \geq \varepsilon,$$

then there exists a positive integer $q \in [C\varepsilon^{-C}]$ such that

$$\|q\xi_i\|_{\hat{\mathbb{K}}} \leq C\varepsilon^{-C} N^{-i}$$

for all $i \in [d]$. If $\mathbb{K} = \mathbb{R}$, then we take $q = 1$.

The proof for $\mathbb{K} = \mathbb{Z}$ can be found in [45, Proposition 4.3], whereas the case $\mathbb{K} = \mathbb{R}$ is the classical van der Corput lemma for oscillatory integrals with polynomial phases; see [87, Proposition 2, p. 332] or [1, Theorem 1.1, p. 13].

The following simple lemma will also be needed.

Lemma 4.48. *Fix $\alpha \in \mathbb{R}$ and $\gamma \in \mathbb{R}_+$. If $|\alpha - u| \leq \gamma$ for some $u \in \mathbb{Q}$, then, for any $M \in [1, \infty)$, one can find $m \in [\pm M]$ and $\theta \in [-1, 1]$ such that*

$$\alpha = u + m \frac{\gamma}{M} + \theta \frac{\gamma}{M}.$$

Proof. We take $m := \lfloor M(\alpha - u)\gamma^{-1} \rfloor$ and $\theta := M(\alpha - u)\gamma^{-1} - m$. □

To state the next lemma, we will need some notation from [80]. For any $s \in \mathbb{Z}_+$ and $X \subseteq \mathbb{K}^{2s}$, let $\square_s(X)$ denote the set of $3s$ -tuples

$$\left(k_1^{(1)}, \dots, k_s^{(1)}, k_1^{(2)}, \dots, k_s^{(2)}, k_1^{(3)}, \dots, k_s^{(3)} \right) \in \mathbb{K}^{3s}$$

such that

$$\left(k_1^{(1)}, \dots, k_s^{(1)}, k_1^{(\omega_1+2)}, \dots, k_s^{(\omega_s+2)} \right) \in X$$

for all $\omega \in \{0, 1\}^s$. The following lemma is analogous to [80, Lemma 7.4], and is proved in the same way.

Lemma 4.49 (Dual-difference interchange). *Fix $C_0, s, d \in \mathbb{Z}_+$. Then there exists a constant $C \in \mathbb{Z}_+$ depending only on C_0, s, d such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C\delta^{-C}$. Let $C_0^{-1}\delta^{C_0}N^d \leq H \leq C_0\delta^{-C_0}N^d$ and let $\mathfrak{H} \subseteq [H]_{\mathbb{K}}^{2s}$ be a measurable set such that $\lambda_{\mathbb{K}}^{\otimes 2s}(\mathfrak{H}) \geq C_0^{-1}\delta^{C_0}H^{2s}$. Consider $F(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} F_t(x)$, where the map*

$$\mathbb{K} \times [N]_{\mathbb{K}} \ni (x, t) \mapsto F_t(x) \in \mathbb{C}$$

is measurable and, for each $t \in [N]_{\mathbb{K}}$, the map $\mathbb{K} \ni x \mapsto F_t(x)$ is 1-bounded and supported on $J := [\pm C_0 N^d]_{\mathbb{K}}$. If

$$\mathbb{E}_{(\mathbf{h}, \mathbf{h}') \in \mathfrak{H}}^{\lambda_{\mathbb{K}}^{\otimes 2s}} \left| |J|_{\mathbb{K}}^{-1} \int_{\mathbb{K}} [\Delta'_{(h_i, h'_i)_{i \in [s]}} F(x)] e(\phi(\mathbf{h}, \mathbf{h}')x) d\lambda_{\mathbb{K}}(x) \right|^2 \geq C_0^{-1} \delta^{C_0}$$

for some measurable function $\phi: \mathfrak{H} \rightarrow \hat{\mathbb{K}}$, then

$$\mathbb{E}_{\mathbf{k} \in [H]_{\mathbb{K}}^{3s}}^{\lambda_{\mathbb{K}}^{\otimes 3s}} \mathbb{1}_{\square_s(\mathfrak{H})}(\mathbf{k}) \left| |J|_{\mathbb{K}}^{-1} \int_{\mathbb{K}} G_{\mathbf{k}}(x) e(\psi(\mathbf{k})x) d\lambda_{\mathbb{K}}(x) \right|^2 \geq C^{-1} \delta^C,$$

where, for all $\mathbf{k} = (k_1^{(1)}, \dots, k_s^{(1)}, k_1^{(2)}, \dots, k_s^{(2)}, k_1^{(3)}, \dots, k_s^{(3)}) \in \square_s(\mathfrak{H})$, we have

$$G_{\mathbf{k}}(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \Delta'_{(k_i^{(2)}, k_i^{(3)})_{i \in [s]}} F_t(x)$$

and $\psi: \square_s(\mathfrak{H}) \rightarrow \hat{\mathbb{K}}$ is given by

$$\psi(\mathbf{k}) := \sum_{\omega \in \{0,1\}^s} (-1)^{|\omega|} \phi(k_1^{(1)}, \dots, k_s^{(1)}, k_1^{(\omega_1+2)}, \dots, k_s^{(\omega_s+2)}).$$

We will also need the following variant of [80, Lemma 7.5].

Lemma 4.50. Fix $s, C_0 \in \mathbb{Z}_+$. Then there exists a constant $C \in \mathbb{Z}_+$ depending only on s, C_0 such that the following holds. Let $M, H \in [1, \infty)$, $\varepsilon \in (0, 1]$, and, for each $i \in [s]$, let $\phi_i: \mathbb{K}^{2s} \rightarrow \hat{\mathbb{K}}$ be a measurable function independent of the $(s+i)$ -th variable. If $C_0^{-1} \varepsilon^{C_0} M \leq H \leq C_0 \varepsilon^{-C_0} M$ and $f \in L^\infty(\mathbb{K})$ is a 1-bounded function supported on $J := [\pm M]_{\mathbb{K}}$ and such that

$$\mathbb{E}_{\mathbf{h}, \mathbf{h}' \in [H]_{\mathbb{K}}^{2s}}^{\lambda_{\mathbb{K}}^{\otimes 2s}} \left| |J|_{\mathbb{K}}^{-1} \int_{\mathbb{K}} \Delta'_{(\mathbf{h}, \mathbf{h}')} f(x) e\left(\sum_{i \in [s]} \phi_i(\mathbf{h}, \mathbf{h}')x\right) d\lambda_{\mathbb{K}}(x) \right|^2 \geq \varepsilon,$$

then we have $\|f\|_{U_{[H]_{\mathbb{K}}}^{s+1}(J)} \geq C^{-1} \varepsilon^C$.

The next lemma has no analogue in [80], and addresses the added complexity of executing a pigeonholing argument in the multidimensional setting.

Lemma 4.51. Fix $s \in \mathbb{Z}_+$. There exists a constant $C \in \mathbb{Z}_+$ depending only on s such that the following holds. Assume that $\delta \in (0, 1]$ and $H \geq C\delta^{-C}$, and set $B := [H]_{\mathbb{K}}^{2s}$. Then one can find a measurable function $\phi: B \rightarrow \hat{\mathbb{K}}$ and an exceptional set $K_\phi \subseteq \square_s(B)$ with $|K_\phi|_{\mathbb{K}^{3s}} \leq 2C\delta^{-2}H^{-1}|\square_s(B)|_{\mathbb{K}^{3s}}$ such that if $\mathbf{k} \in \square_s(B) \setminus K_\phi$, then

$$\|q\psi(\mathbf{k})\|_{\hat{\mathbb{K}}} > \delta^{-1}H^{-1}$$

holds for all $q \in [\delta^{-1}]$, where $\psi(\mathbf{k})$ is defined as in Lemma 4.49.

Proof. We will consider the cases $\mathbb{K} = \mathbb{Z}$ and $\mathbb{K} = \mathbb{R}$ separately.

Case $\mathbb{K} = \mathbb{Z}$. Define a random function $\phi: B \rightarrow \mathbb{T}$ by choosing each value $\phi(\mathbf{h}, \mathbf{h}')$ independently and uniformly at random from $\mathbb{T}$. Set

$$\mathfrak{M} := \{\xi \in \mathbb{T} : \|q\xi\|_{\mathbb{T}} \leq \delta^{-1}H^{-1} \text{ for some positive integer } q \leq \delta^{-1}\}.$$

Note that $\lambda_{\mathbb{T}}(\mathfrak{M}) \leq C\delta^{-2}H^{-1}$. Moreover, we have $\square_s(B) = [H]^{3s}$, and for any $\varepsilon_1, \dots, \varepsilon_{2s} \in \{-1, 1\}$ and distinct pairs $(\mathbf{h}_1, \mathbf{h}'_1), \dots, (\mathbf{h}_{2s}, \mathbf{h}'_{2s}) \in B$, the function $\sum_{i \in [2s]} \varepsilon_i \phi(\mathbf{h}_i, \mathbf{h}'_i)$ is also uniformly distributed on $\mathbb{T}$. We split $\square_s(B) = K_1 \cup K_2$, where

$$K_1 := \{\mathbf{k} \in [H]^{3s} : k_i^{(2)} \neq k_i^{(3)} \text{ for all } i \in [s]\}$$

and $K_2 := [H]^{3s} \setminus K_1$. Then $\mathbb{E}_{\phi}[\mathbb{1}_{\mathfrak{M}}(\psi(\mathbf{k}))] = \lambda_{\mathbb{T}}(\mathfrak{M})$ for $\mathbf{k} \in K_1$, where $\mathbb{E}_{\phi}$ is the expected value over the possible choices of ϕ . Consequently, we have

$$\mathbb{E}_{\phi} \left[\sum_{\mathbf{k} \in [H]^{3s}} \mathbb{1}_{\mathfrak{M}}(\psi(\mathbf{k})) \right] \leq \sum_{\mathbf{k} \in K_1} \lambda_{\mathbb{T}}(\mathfrak{M}) + \sum_{\mathbf{k} \in K_2} 1 \leq C\delta^{-2}H^{-1} \cdot H^{3s} + CH^{3s-1}$$

by $|K_1|_{\mathbb{Z}^{3s}} \leq H^{3s}$ and $|K_2|_{\mathbb{Z}^{3s}} \leq CH^{3s-1}$. Thus, for some $\phi: B \rightarrow \mathbb{T}$, we have

$$\sum_{\mathbf{k} \in [H]^{3s}} \mathbb{1}_{\mathfrak{M}}(\psi(\mathbf{k})) \leq 2C\delta^{-2}H^{3s-1},$$

and this ϕ has the desired properties.

Case $\mathbb{K} = \mathbb{R}$. Fix a large integer $n \in \mathbb{Z}_+$ to be determined later, and split $[H]_{\mathbb{R}}^s = \bigcup_{J \in \mathcal{J}} J$, where $\mathcal{J} := \{J_i \subseteq [H]_{\mathbb{R}}^s : i \in [2^{ns}]\}$ is a family of 2^{ns} dyadic cubes, each having volume $2^{-ns}H^s$. The cubes in $\mathcal{J}$ are essentially disjoint, i.e., their interiors are disjoint. Define $\phi(\mathbf{h}, \mathbf{h}') := \sum_{i \in [2^{ns}]} 2^{i+1}\delta^{-1}H^{-1}\mathbb{1}_{J_i}(\mathbf{h}')$. Let $\mathcal{J}_0$ consist of all pairs $(J, J') \in \mathcal{J}^2$ of cubes with essentially disjoint sides. We split $\square_s(B) = K_1 \cup K_2$, where

$$K_1 := \{\mathbf{k} \in [H]_{\mathbb{R}}^{3s} : \mathbf{k}_{(s, 2s]} \in J, \mathbf{k}_{(2s, 3s]} \in J' \text{ for some } (J, J') \in \mathcal{J}_0\}$$

and $K_2 := [H]_{\mathbb{R}}^{3s} \setminus K_1$. Note that $|\psi(\mathbf{k})| > \delta^{-1}H^{-1}$ for almost all $\mathbf{k} \in K_1$. If $\mathbf{k} \in K_2$, then $\mathbf{k}_{(s, 2s]} \in J$ and $\mathbf{k}_{(2s, 3s]} \in J'$ for some $J, J' \in \mathcal{J}$, which have at least one side in common. This gives $|K_2|_{\mathbb{R}^{3s}} \leq s2^{-n}H^{3s} \leq H^{3s-1}$ provided that n is sufficiently large in terms of H and s . Hence,

$$\int_{[H]_{\mathbb{R}}^{3s}} \mathbb{1}_{[\pm\delta^{-1}H^{-1}]_{\mathbb{R}}}(\psi(\mathbf{k})) d\lambda_{\mathbb{R}^{3s}}(\mathbf{k}) = \int_{K_2} \mathbb{1}_{[\pm\delta^{-1}H^{-1}]_{\mathbb{R}}}(\psi(\mathbf{k})) d\lambda_{\mathbb{R}^{3s}}(\mathbf{k}) \leq H^{3s-1}$$

and ϕ has the desired properties. $\square$

4.5.2. Conditional degree lowering argument. We will now carry out a conditional variant of the degree lowering argument following [80, Section 8].

We will use the convention that $\mathbb{K}^0 := \{0\}$ denotes the trivial vector space endowed with the Dirac delta measure $\lambda_{\mathbb{K}^0}$ at zero. As in Definition 4.32, for any polynomial mapping $\mathcal{P} = (P_1, \dots, P_k): \mathbb{K} \rightarrow \mathbb{K}^k$, we will write $\mathcal{P}_{\leq l} := (P_1, \dots, P_l): \mathbb{K} \rightarrow \mathbb{K}^l$ and $\mathcal{P}_{> l} := (P_{l+1}, \dots, P_k): \mathbb{K} \rightarrow \mathbb{K}^{k-l}$ so that $\mathcal{P} = (\mathcal{P}_{\leq l}, \mathcal{P}_{> l})$ for any $l \in [k-1]$. Moreover, let $\mathcal{P}_{> 0} := \mathcal{P}_{\leq k} := \mathcal{P}$. We analogously define $\mathcal{P}_{< l} := \mathcal{P}_{\leq l-1}$, $\mathcal{P}_{\geq l} := \mathcal{P}_{> l-1}$, and $\mathcal{P}_{(m, l]}, \mathcal{P}_{(m, l]^c}$ for $m, l \in [k]$ with $m \leq l$.

Our main result here is the following.

Lemma 4.52 (Conditional degree lowering lemma). *Fix an integer $s \geq 3$, $A \in [1, \infty)$, $C_0 \in \mathbb{Z}_+$, $k, l \in \mathbb{Z}_+ \setminus \{1\}$, $m \in \mathbb{Z}_+$ with $m \leq l \leq k$, and $d_1, \dots, d_k \in \mathbb{Z}_+$ with $d_1 < \dots < d_k$. Then there exists a large constant $C \in \mathbb{Z}_+$ depending only on s, A, C_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C\delta^{-C}$. Let $\mathcal{P} = (P_1, \dots, P_k)$ be a polynomial mapping as in (4.1), where each $P_i \in \mathbb{K}[t]$ has degree d_i and is (d_i, δ, N) -admissible with tolerance A . Assume that $C_0^{-1}\delta^{C_0}N^{d_m} \leq H \leq C_0\delta^{-C_0}N^{d_m}$. Let $f_0, f_1, \dots, f_{m-1} \in L^\infty(\mathbb{K}^m)$ be 1-bounded measurable functions supported on $I_{\leq m} := \prod_{i \in [m]} [\pm C_0\delta^{-C_0}N^{d_i}]_{\mathbb{K}}$. Given a measurable mapping $\xi := \xi_{(m,l)} := (\xi_{m+1}, \dots, \xi_l): \mathbb{K}^m \rightarrow \hat{\mathbb{K}}^{l-m}$ and a frequency vector $\eta := \eta_{(l,k)} := (\eta_{l+1}, \dots, \eta_k) \in \hat{\mathbb{K}}^{k-l}$, define*

$$F_m(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} f_0(x + \mathbf{P}_m(t)) \left(\prod_{i=1}^{m-1} f_i(x - \mathbf{P}_i(t) + \mathbf{P}_m(t)) \right) \cdot (\xi, \eta)(x, t),$$

where $\mathbf{P}_i(t) := P_i(t)e_i$ for $i \in [m]$ and

$$(\xi, \eta)(x, t) := e(\xi(x + \mathbf{P}_m(t)) \cdot \mathcal{P}_{(m,l)}(t) + \eta \cdot \mathcal{P}_{>l}(t))$$

for $x \in \mathbb{K}^m$ and $t \in [N]_{\mathbb{K}}$. We have the following degree lowering implication

$$\|F_m\|_{U_{[H]_{\mathbb{K}}e_m}^s(I_{\leq m})} \geq C_0^{-1}\delta^{C_0} \implies \|F_m\|_{U_{[H]_{\mathbb{K}}e_m}^{s-1}(I_{\leq m})} \geq C^{-1}\delta^C$$

provided that Condition 4.53 formulated below is satisfied.

Condition 4.53 ((m, l) -major arc condition). *For $A, k, l, m, d_1, \dots, d_k, \mathcal{P}$ as above and $c_0 \in \mathbb{Z}_+$, there exists a large constant $c \in \mathbb{Z}_+$ depending only on A, c_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq c\delta^{-c}$, and that each $P_i \in \mathbb{K}[t]$ is (d_i, δ, N) -admissible with tolerance A . Let $g_0, g_1, \dots, g_{m-1} \in L^\infty(\mathbb{K}^{m-1})$ be 1-bounded measurable functions supported on $J_{<m}$, where $J_{<1} := \{0\} = \mathbb{K}^0$ and $J_{<m} := \prod_{i \in [m-1]} [\pm c_0\delta^{-c_0}N^{d_i}]_{\mathbb{K}}$ if $m \geq 2$. Given a measurable mapping $\zeta := \zeta_{(m,l)} := (\zeta_m, \dots, \zeta_l): \mathbb{K}^{m-1} \rightarrow \hat{\mathbb{K}}^{l-m+1}$ and a frequency vector $\theta := \theta_{(l,k)} := (\theta_{l+1}, \dots, \theta_k) \in \hat{\mathbb{K}}^{k-l}$, define $G_m \in \mathbb{C}$ by*

$$\int_{\mathbb{K}^{m-1}} g_0(y) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{m-1} g_i(y - \mathbf{P}_i(t)) \right) e(\zeta(y) \cdot \mathcal{P}_{(m,l)}(t) + \theta \cdot \mathcal{P}_{>l}(t)) d\lambda_{\mathbb{K}^{m-1}}(y)$$

with $\mathbf{P}_i(t) := P_i(y)e_i$. Set $D_n := d_1 + \dots + d_n$ for $n \in [k]$ and $D_0 := 0$. If

$$|G_m| \geq c_0^{-1}\delta^{c_0}N^{D_{m-1}}, \quad (4.54)$$

then we can find an integer $q \in [c\delta^{-c}]$ and a measurable set $Y \subseteq J_{<m}$ such that $|Y|_{\mathbb{K}^{m-1}} \geq c^{-1}\delta^c N^{D_{m-1}}$ and, for every $y \in Y$, we have

$$\sum_{j=m}^l N^{d_j} \|q\zeta_j(y)\|_{\hat{\mathbb{K}}} + \sum_{j=l+1}^k N^{d_j} \|q\theta_j\|_{\hat{\mathbb{K}}} \leq c\delta^{-c}. \quad (4.55)$$

If $\mathbb{K} = \mathbb{R}$, then we take $q = 1$.

Note that, in the special case $m = 1$, condition (4.54) can be rewritten as

$$\left| g_0(0) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} e(\zeta(0) \cdot \mathcal{P}_{(m,l)}(t) + \theta \cdot \mathcal{P}_{>l}(t)) \right| \geq c_0^{-1}\delta^{c_0}.$$

This condition implies (4.55) by Lemma 4.47, since the polynomials $P_1, \dots, P_k$ have distinct degrees. Thus, the major arc condition for $m = 1$ holds. In order to verify Condition 4.53 for $m \geq 2$, we will adapt an induction argument from [80] using the conditional degree lowering argument from Lemma 4.52.

Proof of Lemma 4.52. The proof is broadly similar to the analogous argument in [80], with the main difference being the need for a more elaborate pigeonholing argument to deal with the dependence of phases on multilinear fixed variables. We will proceed in a few steps and, if necessary, we will distinguish between the cases $m \geq 2$ and $m = 1$ in each step. In order to ensure consistency in our arguments and notation, when $m = 1$, we will identify $\mathbb{K}^1$ with $\mathbb{K}^0 \times \mathbb{K}$ by setting $\mathbb{K} \ni x \mapsto (0, x) \in \mathbb{K}^1$. We notice that the spaces of measurable functions $L^0(\mathbb{K})$ and $L^0(\mathbb{K}^1)$ can also be identified by declaring $L^0(\mathbb{K}) \ni f \mapsto \mathbb{1}_{\{0\}} \otimes f \in L^0(\mathbb{K}^1)$. From now on, when $m = 1$, we will make use of these identifications and assume that all functions $f \in L^0(\mathbb{K})$ are defined on the product space $\{0\} \times \mathbb{K}$ with the product measure $\lambda_{\mathbb{K}^0} \otimes \lambda_{\mathbb{K}}$. We will also abbreviate F_m and G_m to F and G , respectively.

Step 1. By (4.20), we obtain

$$\mathbb{E}_{\mathbf{h}, \mathbf{h}' \in [H]_{\mathbb{K}}^{s-2}}^{\lambda_{\mathbb{K}}^{\otimes 2(s-2)}} \|\Delta'_{(\mathbf{h} \circ e_m, \mathbf{h}' \circ e_m)} F\|_{U^2_{[H]_{\mathbb{K}} e_m}(I_{\leq m})}^4 \geq (C_0^{-1} \delta^{C_0})^{2^s},$$

where $\mathbf{h} \circ e_m := (h_1 e_m, \dots, h_{s-2} e_m) \in (\mathbb{K}^m)^{s-2}$ for $\mathbf{h} = (h_1, \dots, h_{s-2})$.

For any $x_{<m} = (x_1, \dots, x_{m-1}) \in \mathbb{K}^{m-1}$, where $x_{<1} := 0$, we set

$$F_{x_{<m}}(x_m) := F(x_{<m}, x_m) := F(x_1, \dots, x_{m-1}, x_m), \quad x_m \in \mathbb{K}.$$

We observe that, for some constant $C_1 \in \mathbb{Z}_+$, the function $F_{x_{<m}}$ is supported on the set $J_m := [\pm C_1 \delta^{-C_1} N^{d_m}]_{\mathbb{K}}$ for any $x_{<m} \in J_{<m}$, and $F_{x_{<m}} \equiv 0$ for any $x_{<m} \notin J_{<m}$, where $J_{<m}$ is as in Condition 4.53 with C_1 in place of c_0 .

By the popularity principle and Proposition 4.27, for some constant $C_2 \in \mathbb{Z}_+$, there exists a subset $X_{<m} \subseteq J_{<m}$ with $|X_{<m}|_{\mathbb{K}^{m-1}} \geq C_2^{-1} \delta^{C_2} |J_{<m}|_{\mathbb{K}^{m-1}}$ such that if $x_{<m} \in X_{<m}$, then

$$(x_m, t) \mapsto f_0(x + \mathbf{P}_m(t)) \left(\prod_{i=1}^{m-1} f_i(x - \mathbf{P}_i(t) + \mathbf{P}_m(t)) \right) \cdot (\xi, \eta)(x, t),$$

where $x = (x_{<m}, x_m)$, is a measurable function from $\mathbb{K} \times [N]_{\mathbb{K}}$ to $\mathbb{C}$, and there is a measurable set $H_{x_{<m}} \subseteq [H]_{\mathbb{K}}^{2(s-2)}$ with

$$|H_{x_{<m}}|_{\mathbb{K}^{2(s-2)}} \geq C_2^{-1} \delta^{C_2} \left| [H]_{\mathbb{K}}^{2(s-2)} \right|_{\mathbb{K}^{2(s-2)}}$$

such that, for every $(\mathbf{h}, \mathbf{h}') \in H_{x_{<m}}$, we have

$$|\mathcal{F}(x_{<m}, \mathbf{h}, \mathbf{h}'; \xi')|^2 \geq C_2^{-1} \delta^{C_2} N^{2d_m} \quad (4.56)$$

for some $\xi' \in \hat{\mathbb{K}}'$, where $\hat{\mathbb{K}}' \subseteq \hat{\mathbb{K}}$ is a finite set independent of $x_{<m} \in X_{<m}$, and

$$\mathcal{F}(x_{<m}, \mathbf{h}, \mathbf{h}'; \xi') := \int_{\mathbb{K}} \Delta'_{(\mathbf{h} \circ e_m, \mathbf{h}' \circ e_m)} F_{x_{<m}}(x_m) e(\xi' x_m) d\lambda_{\mathbb{K}}(x_m).$$

We can take $\hat{\mathbb{K}}'$ finite, since $\hat{\mathbb{K}} \ni \xi' \mapsto \mathcal{F}(x_{<m}, \mathbf{h}, \mathbf{h}'; \xi') \in \mathbb{C}$ is continuous.

Moreover, by (4.56), for every $x_{<m} \in X_{<m}$ and $(\mathbf{h}, \mathbf{h}') \in H_{x_{<m}}$, we have

$$|\mathcal{F}(x_{<m}, \mathbf{h}, \mathbf{h}'; \phi_{x_{<m}}(\mathbf{h}, \mathbf{h}'))|^2 \geq C_2^{-1} \delta^{C_2} N^{2d_m} \quad (4.57)$$

for some measurable function

$$J_{<m} \times [H]_{\mathbb{K}}^{s-2} \times [H]_{\mathbb{K}}^{s-2} \ni (x_{<m}, \mathbf{h}, \mathbf{h}') \mapsto \phi_{x_{<m}}(\mathbf{h}, \mathbf{h}') \in \hat{\mathbb{K}}$$

such that $\phi_{x_{<m}} : [H]_{\mathbb{K}}^{2(s-2)} \rightarrow \hat{\mathbb{K}}$ is measurable for every $x_{<m} \in J_{<m}$.

Step 2. For every $\mathbf{k} = (k_1^{(1)}, \dots, k_{s-2}^{(1)}, k_1^{(2)}, \dots, k_{s-2}^{(2)}, k_1^{(3)}, \dots, k_{s-2}^{(3)}) \in \mathbb{K}^{3(s-2)}$, we define the corresponding projections $\mathbf{k}^{(\iota)} := (k_1^{(\iota)}, \dots, k_{s-2}^{(\iota)})$ for $\iota \in \{2, 3\}$ and difference operators $\Delta'_{\mathbf{k}; e_m} := \Delta'_{(\mathbf{k}^{(2)} \circ e_m, \mathbf{k}^{(3)} \circ e_m)}$.

Fix $x_{<m} \in X_{<m}$. By Lemma 4.49 applied to (4.57), we can estimate

$$\mathbb{E}_{\mathbf{k} \in [H]_{\mathbb{K}}^{3(s-2)}}^{\lambda_{\mathbb{K}}^{\otimes 3(s-2)}} \mathbf{1}_{\square_{s-2}(H_{x_{<m}})}(\mathbf{k}) \left| \int_{\mathbb{K}} G_{x_{<m}, \mathbf{k}}(x_m) e(\psi_{x_{<m}}(\mathbf{k}) x_m) d\lambda_{\mathbb{K}}(x_m) \right|^2$$

from below by a constant multiple of $\delta^{O(1)} N^{2d_m}$, where

$$\psi_{x_{<m}}(\mathbf{k}) := \sum_{\omega \in \{0,1\}^{s-2}} (-1)^{|\omega|} \phi_{x_{<m}}(k_1^{(1)}, \dots, k_{s-2}^{(1)}, k_1^{(\omega_1+2)}, \dots, k_{s-2}^{(\omega_{s-2}+2)})$$

and, denoting $x := (x_{<m}, x_m) \in \mathbb{K}^m$ for $x_m \in \mathbb{K}$, we define $G_{x_{<m}, \mathbf{k}}(x_m)$ by

$$\mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \Delta'_{\mathbf{k}; e_m} f_0(x + \mathbf{P}_m(t)) \left(\prod_{i=1}^{m-1} \Delta'_{\mathbf{k}; e_m} f_i(x - \mathbf{P}_i(t) + \mathbf{P}_m(t)) \right) \Delta'_{\mathbf{k}; e_m}(\xi, \eta)(x, t).$$

By the popularity principle, for some constant $C_3 \in \mathbb{Z}_+$, if $x_{<m} \in X_{<m}$, then the set $K_{x_{<m}}$, defined as the set of all $\mathbf{k} \in \square_{s-2}(H_{x_{<m}})$ such that

$$\left| \int_{\mathbb{K}} G_{x_{<m}, \mathbf{k}}(x_m) e(\psi_{x_{<m}}(\mathbf{k}) x_m) d\lambda_{\mathbb{K}}(x_m) \right| \geq C_3^{-1} \delta^{C_3} N^{d_m},$$

is measurable and satisfies $|K_{x_{<m}}|_{\mathbb{K}^{3(s-2)}} \geq C_3^{-1} \delta^{C_3} |[H]_{\mathbb{K}}^{3(s-2)}|_{\mathbb{K}^{3(s-2)}}$.

Step 3. Consider the following measurable set

$$Z := \left\{ (x_{<m}, \mathbf{k}) \in J_{<m} \times \square_{s-2}([H]_{\mathbb{K}}^{2(s-2)}) : x_{<m} \in X_{<m} \text{ and } \mathbf{k} \in K_{x_{<m}} \right\}$$

and its sections $Z_{\mathbf{k}} := \{x_{<m} \in J_{<m} : (x_{<m}, \mathbf{k}) \in Z\}$. By Fubini's theorem, for some constant $C_4 \in \mathbb{Z}_+$, there is a measurable set $K \subseteq \square_{s-2}([H]_{\mathbb{K}}^{2(s-2)})$ with $|K|_{\mathbb{K}^{3(s-2)}} \geq C_4^{-1} \delta^{C_4} |\square_{s-2}([H]_{\mathbb{K}}^{2(s-2)})|_{\mathbb{K}^{3(s-2)}}$ such that if $\mathbf{k} \in K$, then

$$|Z_{\mathbf{k}}|_{\mathbb{K}^{m-1}} \geq C_4^{-1} \delta^{C_4} |J_{<m}|_{\mathbb{K}^{m-1}} \quad (4.58)$$

and $Z_{\mathbf{k}}$ has the property that $\mathbf{k} \in K$ and $x_{<m} \in Z_{\mathbf{k}}$ together imply $\mathbf{k} \in K_{x_{<m}}$.

Step 4. Let $C_5 \in \mathbb{Z}_+$ be a large constant to be specified later. By Lemma 4.51, we can find a measurable function $\phi: [H]_{\mathbb{K}}^{2(s-2)} \rightarrow \hat{\mathbb{K}}$ and an exceptional set $E_\phi \subseteq \square_{s-2}([H]_{\mathbb{K}}^{2(s-2)})$ with

$$|E_\phi|_{\mathbb{K}^{3(s-2)}} \leq C(s)C_5^2\delta^{-2C_5}H^{-1} \left| \square_{s-2}([H]_{\mathbb{K}}^{2(s-2)}) \right|_{\mathbb{K}^{3(s-2)}}, \quad (4.59)$$

with a constant $C(s)$ depending only on s , such that if $\mathbf{k} \in \square_{s-2}([H]_{\mathbb{K}}^{2(s-2)}) \setminus E_\phi$, then $\|q\psi(\mathbf{k})\|_{\hat{\mathbb{K}}} > C_5\delta^{-C_5}H^{-1}$ for all $q \in [C_5\delta^{-C_5}]$ and $\psi(\mathbf{k})$ as in Lemma 4.49. In the case $m \geq 2$ we adjust the definition of the function $\psi_{x_{<m}}$ by setting $\psi_{x_{<m}}^*(\mathbf{k}) := \psi_{x_{<m}}(\mathbf{k})$ if $(x_{<m}, \mathbf{k}) \in Z$ and $\psi_{x_{<m}}^*(\mathbf{k}) := \psi(\mathbf{k})$ otherwise. Let K^* be the set of all $\mathbf{k} \in K$ such that the map $x_{<m} \mapsto \psi_{x_{<m}}^*(\mathbf{k})$ is measurable. We have $|K^*|_{\mathbb{K}^{3(s-2)}} = |K|_{\mathbb{K}^{3(s-2)}}$. From now on, we will not distinguish between $\psi_{x_{<m}}$ and $\psi_{x_{<m}}^*$, and $\psi_{x_{<m}}$ will refer to the adjusted version. Similarly, K will refer to K^* . For all $\mathbf{k} \in K$, by the definition of $K_{x_{<m}}$ and (4.58), we have

$$\int_{\mathbb{K}^{m-1}} \left| \int_{\mathbb{K}} G_{x_{<m}, \mathbf{k}}(x_m) e(\psi_{x_{<m}}(\mathbf{k})x_m) d\lambda_{\mathbb{K}}(x_m) \right| d\lambda_{\mathbb{K}^{m-1}}(x_{<m}) \gtrsim \delta^{O(1)} N^{D_m}$$

with some implicit constants that are chosen independently of C_5 .

Step 5. Next, we will show that Condition 4.53 can, essentially, be applied to the inequality above. By duality, we have a lower bound of the same order for

$$\int_{\mathbb{K}^{m-1}} \int_{\mathbb{K}} g'_0(x_{<m}) G_{x_{<m}, \mathbf{k}}(x_m) e(\psi_{x_{<m}}(\mathbf{k})x_m) d\lambda_{\mathbb{K}}(x_m) d\lambda_{\mathbb{K}^{m-1}}(x_{<m}) \quad (4.60)$$

with some 1-bounded $g'_0 \in L^\infty(\mathbb{K}^{m-1})$. Since $\Delta'_{\mathbf{k}; e_m}(\xi, \eta)(x, t)$ has the form

$$e\left(\sum_{\omega \in \{0,1\}^{s-2}} (-1)^{|\omega|} \xi(x + \mathbf{P}_m(t) + (\omega \cdot \mathbf{k}^{(2)})e_m + ((\mathbf{1} - \omega) \cdot \mathbf{k}^{(3)})e_m) \cdot \mathcal{P}_{[m,l]}(t)\right),$$

after the change of variables $x_m \mapsto x_m - \mathbf{P}_m(t)$ we can rewrite (4.60) as

$$\int_{\mathbb{K}^m} f'_0(x) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{m-1} \Delta'_{\mathbf{k}; e_m} f_i(x - \mathbf{P}_i(t)) \right) e(\zeta'_{[m,l]}(x) \cdot \mathcal{P}_{[m,l]}(t)) d\lambda_{\mathbb{K}^m}(x),$$

where $f'_0(x) := g'_0(x_{<m}) \Delta'_{\mathbf{k}; e_m} f_0(x) e(\psi_{x_{<m}}(\mathbf{k})x_m)$, and the measurable mapping $\zeta'_{[m,l]} := (\zeta'_m, \dots, \zeta'_l): \mathbb{K}^m \rightarrow \hat{\mathbb{K}}^{l-m}$ is defined by setting $\zeta'_m(x) := -\psi_{x_{<m}}(\mathbf{k})$ and

$$\zeta'_i(x) := \sum_{\omega \in \{0,1\}^{s-2}} (-1)^{|\omega|} \xi_i(x + (\omega \cdot \mathbf{k}^{(2)})e_m + ((\mathbf{1} - \omega) \cdot \mathbf{k}^{(3)})e_m) \quad \text{for } i \in [l] \setminus [m].$$

By the pigeonhole principle, we can find $x_m \in I_m$ for which the integral

$$\int_{\mathbb{K}^{m-1}} g_0(y) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{m-1} g_i(y - \mathbf{P}_i(t)) \right) e(\zeta_{[m,l]}(y) \cdot \mathcal{P}_{[m,l]}(t)) d\lambda_{\mathbb{K}^{m-1}}(y)$$

admits a lower bound of order $\delta^{O(1)} N^{D_{m-1}}$, and the functions above are the corresponding sections of the previous ones, i.e., we have $g_0(y) := f'_0(y, x_m)$ and $g_i(y) := \Delta'_{\mathbf{k}; e_m} f_i(y, x_m)$ for $i \in [m-1]$, and $\zeta_i(y) := \zeta'_i(y, x_m)$ for $i \in [l] \setminus [m-1]$ so that $\zeta_{[m,l]} := (\zeta_m, \dots, \zeta_l): \mathbb{K}^{m-1} \rightarrow \hat{\mathbb{K}}^{l-m+1}$.

Now, the (m, l) -major arc condition can be applied to the last integral, yielding that the following holds for some large $c \in \mathbb{Z}_+$ independent of C_5 . For each $\mathbf{k} \in K$, we can find a

positive integer $q_{\mathbf{k}} \in [c\delta^{-c}]$ and a measurable set $Y_{<m, \mathbf{k}} \subseteq J_{<m}$ such that $|Y_{<m, \mathbf{k}}|_{\mathbb{K}^{m-1}} \geq c^{-1}\delta^c |J_{<m}|_{\mathbb{K}^{m-1}}$ and

$$\|q_{\mathbf{k}}\psi_{x_{<m}}(\mathbf{k})\|_{\mathbb{K}} \leq c\delta^{-c}N^{-d_m}, \quad x_{<m} \in Y_{<m, \mathbf{k}}.$$

Then, by the pigeonhole principle, we find a constant $C_6 \in \mathbb{Z}_+$ (depending on c and C_4 , but not on C_5), a measurable set $K' \subseteq K$ with $|K'|_{\mathbb{K}^{3(s-2)}} \geq C_6^{-1}\delta^{C_6}|\square_{s-2}([H]_{\mathbb{K}}^{2(s-2)})|_{\mathbb{K}^{3(s-2)}}$, and $q \in [C_6\delta^{-C_6}]$ such that

$$\|q\psi_{x_{<m}}(\mathbf{k})\|_{\mathbb{K}} \leq C_6\delta^{-C_6}N^{-d_m}, \quad \mathbf{k} \in K', x_{<m} \in Y_{<m, \mathbf{k}}.$$

We can assume, without loss of generality, that

$$Y_{<m, \mathbf{k}} := \left\{ x_{<m} \in J_{<m} : \|q\psi_{x_{<m}}(\mathbf{k})\|_{\mathbb{K}} \leq C_6\delta^{-C_6}N^{-d_m} \right\},$$

and $|Y_{<m, \mathbf{k}}|_{\mathbb{K}^{m-1}} \geq C_6^{-1}\delta^{C_6}|J_{<m}|_{\mathbb{K}^{m-1}}$. If $\mathbb{K} = \mathbb{R}$, then we simply take $K' = K$, which is possible because $q_{\mathbf{k}} = 1$ for all $\mathbf{k} \in K$.

Step 6. We now specify $C_5 > 2C_0C_6$ to conclude that, for every $\mathbf{k} \in K'$, if $x_{<m} \in J_{<m} \setminus Z_{\mathbf{k}}$ and $x_{<m} \in Y_{<m, \mathbf{k}}$, then, necessarily, $\mathbf{k} \in E_{\phi}$. By using (4.59), together with the assumption that $N \geq C\delta^{-C}$ for a sufficiently large constant $C \in \mathbb{Z}_+$, and by Fubini's theorem, we obtain that, for some constant $C_7 \in \mathbb{Z}_+$, there is a measurable set $X'_{<m} \subseteq X_{<m}$ with

$$|X'_{<m}|_{\mathbb{K}^{m-1}} \geq C_7^{-1}\delta^{C_7}|J_{<m}|_{\mathbb{K}^{m-1}}$$

such that if $x_{<m} \in X'_{<m}$, then there is a measurable set $U_{x_{<m}} \subseteq K'$ with

$$|U_{x_{<m}}|_{\mathbb{K}^{3(s-2)}} \geq C_7^{-1}\delta^{C_7} \left| \square_{s-2}([H]_{\mathbb{K}}^{2(s-2)}) \right|_{\mathbb{K}^{3(s-2)}}$$

and with the property that $x_{<m} \in X'_{<m}$ and $\mathbf{k} \in U_{x_{<m}}$ together imply $x_{<m} \in Y_{<m, \mathbf{k}} \cap Z_{\mathbf{k}}$ and $\mathbf{k} \in K_{x_{<m}}$.

Step 7. Let $M_0 := C_6^{-1}C_8\delta^{C_6-C_8}$ for a large $C_8 \in \mathbb{Z}_+$ to be specified later. By Lemma 4.48 and the pigeonhole principle, there are measurable sets $X''_{<m} \subseteq X'_{<m}$ with $|X''_{<m}|_{\mathbb{K}^{m-1}} \gtrsim \delta^{O(1)}|J_{<m}|_{\mathbb{K}^{m-1}}$ and $V_{x_{<m}} \subseteq U_{x_{<m}}$ with

$$|V_{x_{<m}}|_{\mathbb{K}^{3(s-2)}} \gtrsim \delta^{O(1)} \left| \square_{s-2}([H]_{\mathbb{K}}^{2(s-2)}) \right|_{\mathbb{K}^{3(s-2)}}$$

such that, for some $a, b \in \mathbb{Z}$ satisfying $|a| + |b| \lesssim \delta^{-O(1)}$, if $x_{<m} \in X''_{<m}$, then

$$\psi_{x_{<m}}(\mathbf{k}) = \frac{a}{q} + \frac{b}{M_0N^{d_m}} + \frac{\theta_{x_{<m}}(\mathbf{k})}{M_0N^{d_m}} \quad \text{with} \quad |\theta_{x_{<m}}(\mathbf{k})| \leq 1$$

for all $\mathbf{k} \in V_{x_{<m}}$. Here, it is crucial that $\psi_{x_{<m}}(\mathbf{k})$ is the original function from Step 2, since $x_{<m} \in X''_{<m}$ and $\mathbf{k} \in V_{x_{<m}}$ together imply $(x_{<m}, \mathbf{k}) \in Z$.

Set $\Omega_1 := \{\omega \in \{0, 1\}^{s-2} : \omega_1 = 1\}$. For $x_{<m} \in X''_{<m}$, define $\eta_{x_{<m}}^{(1)}(\mathbf{k})$ by

$$- \sum_{\omega \in \Omega_1} (-1)^{|\omega|} \phi_{x_{<m}} \left(k_1^{(1)}, \dots, k_{s-2}^{(1)}, k_1^{(\omega_1+2)}, \dots, k_{s-2}^{(\omega_{s-2}+2)} \right) + \frac{a}{q} + \frac{b}{M_0N^{d_m}}$$

so that $\eta_{x_{<m}}^{(1)}$ does not depend on $k_1^{(2)}$. Similarly, for each $i \in [s-2] \setminus \{1\}$, set $\Omega_i := \{\omega \in \{0, 1\}^{s-2} : \omega_1 = \dots = \omega_{i-1} = 0, \omega_i = 1\}$ and define $\eta_{x_{<m}}^{(i)}(\mathbf{k})$ by

$$- \sum_{\omega \in \Omega_i} (-1)^{|\omega|} \phi_{x_{<m}} \left(k_1^{(1)}, \dots, k_{s-2}^{(1)}, k_1^{(\omega_1+2)}, \dots, k_{s-2}^{(\omega_{s-2}+2)} \right)$$

so that $\eta_{x_{<m}}^{(i)}$ does not depend on $k_i^{(2)}$. Note that if $\mathbf{k} \in V_{x_{<m}}$, then

$$\left| \phi_{x_{<m}}(k_1^{(1)}, \dots, k_{s-2}^{(1)}, k_1^{(2)}, \dots, k_{s-2}^{(2)}) - \sum_{i \in [s-2]} \eta_{x_{<m}}^{(i)}(\mathbf{k}) \right| = \left| \frac{\theta_{x_{<m}}(\mathbf{k})}{M_0 N^{d_m}} \right| \leq \frac{1}{M_0 N^{d_m}}.$$

Step 8. By the pigeonhole principle, if $x_{<m} \in X''_{<m}$, then there exists a parameter $\mathbf{h}''_{x_{<m}} \in [H]_{\mathbb{K}}^{s-2}$ for which the corresponding set

$$H'_{x_{<m}} := \{(\mathbf{h}, \mathbf{h}') \in H_{x_{<m}} : (\mathbf{h}, \mathbf{h}', \mathbf{h}''_{x_{<m}}) \in V_{x_{<m}}\}$$

satisfies $|H'_{x_{<m}}|_{\mathbb{K}^{2(s-2)}} \gtrsim \delta^{O(1)} |[H]_{\mathbb{K}}^{2(s-2)}|_{\mathbb{K}^{2(s-2)}}$. If $C_8 \in \mathbb{Z}_+$ is large, then

$$\mathbb{E}_{(\mathbf{h}, \mathbf{h}') \in H'_{x_{<m}}}^{\lambda_{\mathbb{K}}^{\otimes 2(s-2)}} \left| \int_{\mathbb{K}} \Delta'_{(\mathbf{h} \circ e_m, \mathbf{h}' \circ e_m)} F_{x_{<m}}(x_m) \prod_{i=1}^{s-2} e(\eta_{x_{<m}}^{(i)}(\mathbf{h}, \mathbf{h}', \mathbf{h}''_{x_{<m}}) x_m) d\lambda_{\mathbb{K}}(x_m) \right|^2$$

admits a lower bound of order $\delta^{O(1)} N^{2d_m}$ thanks to (4.57).

Applying Lemma 4.50 for each fixed $x_{<m} \in X''_{<m}$ and integrating over all $x_{<m} \in X''_{<m}$, we conclude, by positivity, that $\|F\|_{U_{[H]_{\mathbb{K}} e_m}^{s-1}(I_{\leq m})}} \gtrsim \delta^{O(1)}$. $\square$

4.5.3. The major arc and degree lowering lemmas. Our goal is now to prove an unconditional version of the degree lowering lemma. To do this, we will need to verify that Condition 4.53 always holds.

Lemma 4.61 ($((m, l)$ -major arc lemma). *Fix $A \in [1, \infty)$, $C_0 \in \mathbb{Z}_+$, $k, l \in \mathbb{Z}_+ \setminus \{1\}$, $m \in \mathbb{Z}_+$ with $m \leq l \leq k$, and $d_1, \dots, d_k \in \mathbb{Z}_+$ with $d_1 < \dots < d_k$. Then there exists a large constant $C \in \mathbb{Z}_+$ depending only on A, C_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C\delta^{-C}$. Let $\mathcal{P} = (P_1, \dots, P_k)$ be a polynomial mapping as in (4.1), where each $P_i \in \mathbb{K}[t]$ has degree d_i and is (d_i, δ, N) -admissible with tolerance A . Let $g_0, g_1, \dots, g_{m-1} \in L^\infty(\mathbb{K}^{m-1})$ be 1-bounded measurable functions supported on $I_{<m}$, where $I_{<1} := \{0\} = \mathbb{K}^0$ and $I_{<m} := \prod_{i \in [m-1]} [\pm C_0 \delta^{-C_0} N^{d_i}]_{\mathbb{K}}$ if $m \geq 2$. Given a measurable mapping $\zeta := \zeta_{[m, l]} := (\zeta_m, \dots, \zeta_l) : \mathbb{K}^{m-1} \rightarrow \hat{\mathbb{K}}^{l-m+1}$ and a frequency vector $\theta := \theta_{[l, k]} := (\theta_{l+1}, \dots, \theta_k) \in \hat{\mathbb{K}}^{k-l}$, define $G_m \in \mathbb{C}$ by*

$$\int_{\mathbb{K}^{m-1}} g_0(y) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{m-1} g_i(y - \mathbf{P}_i(t)) \right) e(\zeta(y) \cdot \mathcal{P}_{[m, l]}(t) + \theta \cdot \mathcal{P}_{>l}(t)) d\lambda_{\mathbb{K}^{m-1}}(y)$$

with $\mathbf{P}_i(t) := P_i(y) e_i$. Set $D_n := d_1 + \dots + d_n$ for $n \in [k]$ and $D_0 := 0$. If

$$|G_m| \geq C_0^{-1} \delta^{C_0} N^{D_{m-1}}, \quad (4.62)$$

then we can find an integer $q \in [C\delta^{-C}]$ and a measurable set $Y \subseteq I_{<m}$ such that $|Y|_{\mathbb{K}^{m-1}} \geq C^{-1}\delta^C N^{D_{m-1}}$ and, for every $y \in Y$, we have

$$\sum_{j=m}^l N^{d_j} \|q\zeta_j(y)\|_{\mathbb{K}} + \sum_{j=l+1}^k N^{d_j} \|q\theta_j\|_{\mathbb{K}} \leq C\delta^{-C}. \quad (4.63)$$

If $\mathbb{K} = \mathbb{R}$, then we take $q = 1$.

Observe that inequality (4.62) is exactly inequality (4.54) with $c_0 = C_0$. Thus, Lemma 4.61 verifies the (m, l) -major arc condition, thereby establishing the degree lowering lemma unconditionally; see Lemma 4.70 below.

Proof of Lemma 4.61. In the discussion following Lemma 4.52, we have already observed that Lemma 4.61 holds for $m = 1$ and any $k, l \in \mathbb{Z}_+ \setminus \{1\}$ such that $m \leq l \leq k$. In other words, the $(1, l)$ -major arc lemma is true, establishing the base case of an induction argument. Let $k, l \in \mathbb{Z}_+ \setminus \{1\}$ be fixed and let $m \in \mathbb{Z}_+$ be such that $m+1 \leq l$. Assuming that the (m, l) -major arc lemma is true, we proceed inductively to show that the $(m+1, l)$ -major arc lemma is also true. Let $g_0, g_1, \dots, g_m \in L^\infty(\mathbb{K}^m)$ be 1-bounded measurable functions supported on $I_{\leq m} := \prod_{i \in [m]} [\pm C_0 \delta^{-C_0} N^{d_i}]_{\mathbb{K}}$. Let $\zeta_{[m+1, l]} := (\zeta_{m+1}, \dots, \zeta_l): \mathbb{K}^m \rightarrow \mathbb{K}^{l-m}$ be a measurable mapping and $\theta := \theta_{(l, k]} := (\theta_{l+1}, \dots, \theta_k) \in \hat{\mathbb{K}}^{k-l}$. Suppose that

$$|G_{m+1}| \geq C_0^{-1} \delta^{C_0} N^{D_m}. \quad (4.64)$$

We now prove that (4.64) implies condition (4.63) with $m+1$ in place of m .

Step 1. By making the change of variables $y \mapsto y + \mathbf{P}_m(t)$ in G_{m+1} , applying the Cauchy–Schwarz inequality to double the t variable, and undoing the change of variables, we obtain that

$$\left| \int_{\mathbb{K}^m} g_0(y) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{m-1} g_i(y - \mathbf{P}_i(t)) \right) \mathcal{C}F_m(y - \mathbf{P}_m(t)) \cdot (\xi, \theta)(y, t) d\lambda_{\mathbb{K}^m}(y) \right|$$

admits a lower bound of order $\delta^{O(1)} N^{D_m}$, where $F_m(y)$ is given by

$$\mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} g_0(y + \mathbf{P}_m(t)) \left(\prod_{i=1}^{m-1} g_i(y + \mathbf{P}_m(t) - \mathbf{P}_i(t)) \right) \cdot (\xi, \theta)(y + \mathbf{P}_m(t), t)$$

for $y \in \mathbb{K}^m$, and we also define

$$(\xi, \theta)(y, t) := e(\zeta_{(m, l]}(y) \cdot \mathcal{P}_{(m, l]}(t) + \theta_{(l, k]} \cdot \mathcal{P}_{(l, k]}(t)) \quad \text{for } t \in [N]_{\mathbb{K}}, y \in \mathbb{K}^m.$$

Step 2. We claim that the estimate from Step 1 implies that

$$\|\mathcal{C}F_m\|_{U_{[H]_{\mathbb{K}} e_m}^s(I_{\leq m})} \gtrsim \delta^{O(1)} \quad (4.65)$$

for some $\delta^{O(1)} N^{d_m} \lesssim H \lesssim \delta^{-O(1)} N^{d_m}$ and an integer $s \geq 2$ bounded in terms of A, C_0, d_k . For this purpose, we need to eliminate the factor $(\zeta, \theta)(y, t)$ by applying van der Corput differencing d_k+1 times. More precisely, we apply Proposition 4.41 with the conclusion (4.42)

for the expression from Step 1 repeatedly $d_k + 1$ times and obtain a lower bound of the same order for

$$\mathbb{E}_{\mathbf{h} \in \mathbf{H}}^{\lambda_{\mathbb{K}}^{\otimes d_k+1}} \left| \int_{\mathbb{K}^m} \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i=1}^m \prod_{\omega \in \{0,1\}^{d_k+1}} \mathcal{C}^{|\omega|} g'_i(x - \mathbf{P}_i(t + \omega \cdot \mathbf{h})) d\lambda_{\mathbb{K}^m}(x) \right|$$

with $g'_m := \mathcal{C}F_m$ and $g'_i := g_i$ for $i \in [m-1]$, and with some suitably chosen $\mathbf{H} \subseteq [\pm N]_{\mathbb{K}}^{d_k+1}$ such that $|\mathbf{H}|_{\mathbb{K}^{d_k+1}} \gtrsim \delta^{O(1)} |[\pm N]_{\mathbb{K}}^{d_k+1}|_{\mathbb{K}^{d_k+1}}$. We observe that, for each $\mathbf{h} \in \mathbf{H}$, the family

$$\{P_i(\cdot + \omega \cdot \mathbf{h}) : i \in [m], \omega \in \{0,1\}^{d_k+1}\}$$

consists of (d_m, δ, N) -admissible with a new tolerance constant $A' \lesssim_{A, C_0, d_k} 1$.

Now, we can apply Theorem 4.45, or more precisely, repeat the arguments used to derive it. Namely, (4.65) follows by first applying [63, Proposition 4.8] (which, again, also holds for admissible polynomials and when $\mathbb{K} = \mathbb{R}$) for each $\mathbf{h} \in \mathbf{H}$ and then, after averaging over $\mathbf{h}$, either applying [63, Theorem 1.10] when $\mathbb{K} = \mathbb{Z}$, or performing a change of variables and applying the Gowers–Cauchy–Schwarz inequality (4.26) when $\mathbb{K} = \mathbb{R}$.

Step 3. By the induction hypothesis, the (m, l) -major arc condition holds. Thus, using (4.65) and Lemma 4.52 repeatedly $s - 2$ times, we obtain that

$$\|\mathcal{C}F_m\|_{U_{[H]_{\mathbb{K}}^{e_m}}^2(I_{\leq m})}^4 = |I_{< m}|_{\mathbb{K}^{m-1}}^{-1} \int_{\mathbb{K}^{m-1}} \|\mathcal{C}F_m(x_{< m}, \cdot)\|_{U_{[H]_{\mathbb{K}}}^2(I_m)}^4 d\lambda_{\mathbb{K}^{m-1}}(x_{< m})$$

admits a lower bound of order $\delta^{O(1)}$. By the popularity principle and Proposition 4.27, we find a measurable function $\zeta_m : \mathbb{K}^{m-1} \rightarrow \hat{\mathbb{K}}$ such that

$$\int_{\mathbb{K}^{m-1}} \left| \int_{\mathbb{K}} F_m(x) e(\zeta_m(x_{< m}) x_m) d\lambda_{\mathbb{K}}(x_m) \right| d\lambda_{\mathbb{K}^{m-1}}(x_{< m}) \gtrsim \delta^{O(1)} N^{D_m}. \quad (4.66)$$

Dualizing the left-hand side of (4.66), expanding F_m , and making the change of variables $x \mapsto x - \mathbf{P}_m(t)$, we obtain that (4.66) can be rewritten as

$$\int_{\mathbb{K}^m} f'_0(x) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{m-1} f'_i(x - \mathbf{P}_i(t)) \right) \cdot (\xi, \theta)(x, t) d\lambda_{\mathbb{K}^m}(x) \gtrsim \delta^{O(1)} N^{D_m},$$

where $f'_0(x) := g_0(x) g''_0(x_{< m}) e(\zeta_m(x_{< m}) x_m)$ for some 1-bounded function $g''_0 \in L^\infty(\mathbb{K}^{m-1})$ such that the integrand above is nonnegative, $f'_i := g_i$ for each $i \in [m-1]$, and $\xi := \xi_{[m, l]} := (\xi_m, \dots, \xi_l) : \mathbb{K}^m \rightarrow \hat{\mathbb{K}}^{l-m+1}$ with $\xi_m(x) := -\zeta_m(x_{< m})$ and $\xi_i(x) := \zeta_i(x)$ for $i \in [l] \setminus [m]$ and $x \in \mathbb{K}^m$.

By the popularity principle applied to the last integral, there exists a measurable set $Y_m \subseteq I_m$ such that $|Y_m|_{\mathbb{K}} \gtrsim \delta^{O(1)} |I_m|_{\mathbb{K}}$ and, setting

$$\begin{aligned} h_i(x) &:= f'_i(x, x_m) \quad \text{for } x \in \mathbb{K}^{m-1}, i \in \mathbb{N}_{< m}, \\ \zeta'_i(x) &:= \xi_i(x, x_m) \quad \text{for } x \in \mathbb{K}^{m-1}, i \in \mathbb{N}_{\leq l} \setminus \mathbb{N}_{< m}, \end{aligned}$$

for every $x_m \in Y_m$, we obtain that the integral

$$\int_{\mathbb{K}^{m-1}} h_0(x) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{m-1} h_i(x - \mathbf{P}_i(t)) \right) \cdot (\zeta', \theta)(x, t) d\lambda_{\mathbb{K}^{m-1}}(x)$$

admits a lower bound of order $\delta^{O(1)}N^{D_{m-1}}$. By the induction hypothesis, this estimate implies that, for every $x_m \in Y_m$, we can find a positive integer $q_{x_m} \lesssim \delta^{-O(1)}$ and a measurable set $Y_{x_m} \subseteq I_{<m}$ satisfying $|Y_{x_m}|_{\mathbb{K}^{m-1}} \gtrsim \delta^{O(1)}N^{D_{m-1}}$ such that, for every $y \in Y_{x_m}$, we have

$$\sum_{j=m}^l N^{d_j} \|q_{x_m} \zeta'_j(y)\|_{\mathbb{K}} + \sum_{j=l+1}^k N^{d_j} \|q_{x_m} \theta_j\|_{\mathbb{K}} \lesssim \delta^{-O(1)}. \quad (4.67)$$

By the pigeonhole principle, we can find a positive integer $q \lesssim \delta^{-O(1)}$ and a measurable set $Y'_m \subseteq Y_m$ satisfying $|Y'_m|_{\mathbb{K}} \gtrsim \delta^{O(1)}|I_m|_{\mathbb{K}}$ for which the following holds: for every $x_m \in Y'_m$ there exists a measurable set $Y_{x_m} \subseteq I_{<m}$ satisfying $|Y_{x_m}|_{\mathbb{K}^{m-1}} \gtrsim \delta^{O(1)}N^{D_{m-1}}$ such that for every $y \in Y_{x_m}$ inequality (4.67) holds with q in place of q_{x_m} . If $\mathbb{K} = \mathbb{R}$, then we simply take $Y'_m = Y_m$, which is possible, since $q_{x_m} = 1$ for all $x_m \in Y_m$. Moreover, we can assume, without loss of generality, that

$$Y_{x_m} := \left\{ y \in I_{<m} : \sum_{j=m+1}^l N^{d_j} \|q \zeta_j(y)\|_{\mathbb{K}} + \sum_{j=l+1}^k N^{d_j} \|q \theta_j\|_{\mathbb{K}} \leq C_* \delta^{-C_*} \right\}$$

for some $1 \leq q \lesssim \delta^{-O(1)}$ and $C_* \in \mathbb{Z}_+$, and for all $x_m \in Y'_m$. Then the set Y_{x_m} is measurable and (4.63) holds with $m+1$ in place of m by taking

$$Y := \{(y, x_m) \in I_{<m} \times I_m : x_m \in Y'_m \text{ and } y \in Y_{x_m}\},$$

which is measurable and satisfies $|Y|_{\mathbb{K}^m} \gtrsim \delta^{O(1)}N^{D_m}$ by Fubini's theorem. $\square$

Having proved the (m, l) -major arc lemma, we can now establish the unconditional degree lowering lemma. For $l \in [k]$, a frequency vector $\xi := \xi_{>l} := (\xi_{l+1}, \dots, \xi_k) \in \hat{\mathbb{K}}^{k-l}$, and functions $f_1, \dots, f_l \in L^\infty(\mathbb{K}^k, \lambda_{\mathbb{K}^k})$ we define a modulated variant of the multilinear average from (4.4) by setting

$$A_{N, \xi; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{>l}}(f_1, \dots, f_l)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i \in [l]} f_i(x - \mathbf{P}_i(t)) \right) e(\xi \cdot \mathcal{P}_{>l}(t)) \quad (4.68)$$

for $x \in \mathbb{K}^k$, with $\mathbf{P}_i(t) := P_i(t)e_i$. From (4.68) we have $A_{N, \xi; \mathbb{K}^k}^{\mathcal{P}_{\leq k}, \mathcal{P}_{>k}} = A_{N; \mathbb{K}^k}^{\mathcal{P}}$ for any $\xi \in \mathbb{K}^k$. The adjoint operators $A_{N, \xi; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{>l}, *j}$ for $j \in [l]$ are given by

$$A_{N, \xi; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{>l}, *j}(g_1, \dots, g_l)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i \in [l]} g_i^{*j}(x - \mathbf{P}_i^{*j}(t)) \right) \overline{e(\xi \cdot \mathcal{P}_{>l}(t))} \quad (4.69)$$

for $x \in \mathbb{K}^k$, with $g_i^{*j} := \mathcal{C}^{\mathbb{1}_{i \neq j}} g_i$ and $\mathbf{P}_i^{*j}(t) := \mathbb{1}_{i \neq j} P_i(t)e_i - P_j(t)e_j$; see (4.6) and (4.8) for analogous formulae for $A_{N; \mathbb{K}^k}^{\mathcal{P}, *j}$.

Our concluding result here is as follows.

Lemma 4.70 (Degree lowering lemma). *Fix an integer $s \geq 3$, $A \in [1, \infty)$, $C_0 \in \mathbb{Z}_+$, $k, l \in \mathbb{Z}_+ \setminus \{1\}$ with $l \leq k$, and $d_1, \dots, d_k \in \mathbb{Z}_+$ with $d_1 < \dots < d_k$. Then there exists a large constant $C \in \mathbb{Z}_+$ depending only on s, A, C_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C\delta^{-C}$. Let $\mathcal{P} = (P_1, \dots, P_k)$ be a polynomial mapping as in (4.1), where each $P_i \in \mathbb{K}[t]$ has degree d_i and is (d_i, δ, N) -admissible with tolerance A . Assume that $C_0^{-1}\delta^{C_0}N^{d_l} \leq H \leq C_0\delta^{-C_0}N^{d_l}$. Let $f_0, f_1, \dots, f_{l-1} \in L^\infty(\mathbb{K}^k)$ be 1-bounded*

measurable functions supported on $I := \prod_{i \in [k]} [\pm C_0 \delta^{-C_0} N^{d_i}]$. For a frequency vector $\eta := (\eta_{l+1}, \dots, \eta_k) \in \hat{\mathbb{K}}^{k-l}$, define

$$F^\eta(x) := A_{N, \eta; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{> l}, *l}(f_1, \dots, f_{l-1}, f_0)(x) \quad \text{for } x \in \mathbb{K}^k$$

with $A_{N, \eta; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{> l}, *l}$ from (4.69). Then the following implication holds:

$$\|F^\eta\|_{U_{[H]\mathbb{K}^{e_l}}^s(I)} \geq C_0^{-1} \delta^{C_0} \implies \|F^\eta\|_{U_{[H]\mathbb{K}^{e_l}}^{s-1}(I)} \geq C^{-1} \delta^C.$$

Proof. We first apply the popularity principle to reduce the matter from $\mathbb{K}^k$ to $\mathbb{K}^l$, and then appeal to the (l, l) -major arc lemma and Lemma 4.52. $\square$

4.5.4. All together: proof of Theorem 4.33. Finally, we can complete the proof of Theorem 4.33. We begin with the following intermediate result.

Lemma 4.71. Fix $A \in [1, \infty)$, $C_0 \in \mathbb{Z}_+$, $k \in \mathbb{Z}_+ \setminus \{1\}$, and $d_1, \dots, d_k \in \mathbb{Z}_+$ with $d_1 < \dots < d_k$. Then there exists a large constant $C \in \mathbb{Z}_+$ depending only on A, C_0, d_k such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq C\delta^{-C}$. Let $\mathcal{P} = (P_1, \dots, P_k)$ be a polynomial mapping as in (4.1), where each $P_i \in \mathbb{K}[t]$ has degree d_i and is (d_i, δ, N) -admissible with tolerance A . Let $f_0, f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ be 1-bounded measurable functions supported on $I := \prod_{i \in [k]} [\pm C_0 \delta^{-C_0} N^{d_i}]_{\mathbb{K}}$. If condition (4.34) from Theorem 4.33 holds, then

$$\left| \left\langle A_{N, \zeta_{>1}; \mathbb{K}^k}^{\mathcal{P}_{\leq 1}, \mathcal{P}_{> 1}}(f_1), f'_0 \right\rangle \right| \geq C^{-1} \delta^C N^{D_k} \quad (4.72)$$

for some 1-bounded measurable function $f'_0 \in L^\infty(\mathbb{K}^k)$ satisfying $|f'_0| = |f_0|$ and some frequency vector $\zeta_{>1} := (\zeta_2, \dots, \zeta_k) \in \hat{\mathbb{K}}^{k-1}$.

Proof. Set $D_l := d_1 + \dots + d_l$ for all $l \in [k]$. Given $l \in [k]$, we will prove that there exists a frequency vector $\zeta_{>l} := (\zeta_{l+1}, \dots, \zeta_k) \in \hat{\mathbb{K}}^{k-l}$ such that

$$\left| \left\langle A_{N, \zeta_{>l}; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{> l}}(f_1, \dots, f_l), f_{0,l} \right\rangle \right| \gtrsim \delta^{O(1)} N^{D_k}, \quad (4.73)$$

for some $f_{0,l} \in L^\infty(\mathbb{K}^k)$ such that $|f_{0,l}| = |f_0|$. We prove (4.73) using backward induction on $l \in [k]$. The base case $l = k$, with $\zeta_{>k}$ and $\mathcal{P}_{>k}$ being empty vectors, follows from (4.34) because $A_{N, \zeta_{>k}; \mathbb{K}^k}^{\mathcal{P}_{\leq k}, \mathcal{P}_{> k}} = A_{N; \mathbb{K}^k}^{\mathcal{P}}$. We now assume that (4.73) holds for some $l \in [k] \setminus \{1\}$ and verify (4.73) for $l-1$ in place of l . Once the final case $l = 1$ is verified, Lemma 4.71 follows.

Step 1. Define $F_l(x) := A_{N, \zeta_{>l}; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{> l}, *l}(f_1, \dots, f_{l-1}, f_{0,l})(x)$. By duality and the Cauchy-Schwarz inequality, we may replace f_l with F_l in (4.73) and obtain

$$\left| \left\langle A_{N, \zeta_{>l}; \mathbb{K}^k}^{\mathcal{P}_{\leq l}, \mathcal{P}_{> l}}(f_1, \dots, f_{l-1}, F_l), f_{0,l} \right\rangle \right| \gtrsim \delta^{O(1)} N^{D_k}.$$

As in Step 2 of the proof of Lemma 4.61, we obtain $\|F_l\|_{U_{[H]\mathbb{K}^{e_l}}^s(I)} \gtrsim \delta^{O(1)}$ for some suitably bounded integer $s \geq 2$ and $\delta^{O(1)} N^{d_l} \lesssim H \lesssim \delta^{-O(1)} N^{d_l}$. Then

$$\|F_l\|_{U_{[H]\mathbb{K}^{e_l}}^2(I)}^4 = |I_{\{l\}^c}|_{\mathbb{K}^{k-1}}^{-1} \int_{\mathbb{K}^{k-1}} \|F_l(x_{\{l\}^c}, \cdot)\|_{U_{[H]\mathbb{K}}^2(I_l)}^4 d\lambda_{\mathbb{K}^{k-1}}(x_{\{l\}^c}) \gtrsim \delta^{O(1)}$$

by Lemma 4.70, where we define $x_{\{l\}^c} := (x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_k) \in \mathbb{K}^{k-1}$ and identify $(x_{\{l\}^c}, x_l)$ with $(x_1, \dots, x_k) \in \mathbb{K}^k$, and similarly define $I_{\{l\}^c} := \prod_{i \in [k] \setminus \{l\}} [\pm C_0 \delta^{-C_0} N^{d_i}]_{\mathbb{K}}$ and $I_l := [\pm C_0 \delta^{-C_0} N^{d_l}]_{\mathbb{K}}$. We will also write $x_{\{l\}^c} = (x_{<l}, x_{>l})$ and $I_{\{l\}^c} = I_{<l} \times I_{>l}$ for $x_{<l}, x_{>l}, I_{<l}, I_{>l}$ defined analogously.

By the popularity principle and Proposition 4.27, we find a measurable function $\xi: \mathbb{K}^{k-1} \rightarrow \hat{\mathbb{K}}$ and a measurable set $X_{>l} \subseteq I_{>l}$ satisfying $|X_{>l}|_{\mathbb{K}^{k-l}} \gtrsim \delta^{O(1)} |I_{>l}|_{\mathbb{K}^{k-l}}$ such that, for every $x_{>l} \in X_{>l}$, there exists a measurable set $X_{<l}^{x_{>l}} \subseteq I_{<l}$ satisfying $|X_{<l}^{x_{>l}}|_{\mathbb{K}^{l-1}} \gtrsim \delta^{O(1)} |I_{<l}|_{\mathbb{K}^{l-1}}$ with the following property: whenever $x_{>l} \in X_{>l}$ and $x_{<l} \in X_{<l}^{x_{>l}}$, we have

$$\left| \int_{\mathbb{K}} F_l(x_{<l}, x_l, x_{>l}) e(\xi(x_{<l}, x_{>l}) x_l) d\lambda_{\mathbb{K}}(x_l) \right| \gtrsim \delta^{O(1)} N^{d_l}. \quad (4.74)$$

Step 2. For some large $C_1 \in \mathbb{Z}_+$ to be specified later, we can find a frequency $\xi_0 \in \hat{\mathbb{K}}$ such that $\|q\xi_0\|_{\hat{\mathbb{K}}} > C_1 \delta^{-C_1} N^{-d_l}$ for all $q \in [C_1 \delta^{-C_1}]$. Indeed, set

$$\mathfrak{M} := \left\{ \zeta \in \hat{\mathbb{K}} : \|q\zeta\|_{\hat{\mathbb{K}}} \leq C_1 \delta^{-C_1} N^{-d_l} \text{ for some positive integer } q \leq C_1 \delta^{-C_1} \right\}.$$

The measure of $\mathfrak{M}$ does not exceed $10C_1^2 \delta^{-2C_1} N^{-d_l}$, which is less than $1/2$ provided that $N \geq C \delta^{-C}$ and C is sufficiently large in terms of C_1 . Let ξ_0 be any element of $\hat{\mathbb{K}} \setminus \mathfrak{M}$. We define $\xi_l: \mathbb{K}^{k-1} \rightarrow \hat{\mathbb{K}}$ by $\xi_l(x_{<l}, x_{>l}) := -\xi(x_{<l}, x_{>l})$ if $x_{>l} \in X_{>l}$ and $x_{<l} \in X_{<l}^{x_{>l}}$, and by $\xi_l(x_{<l}, x_{>l}) := -\xi_0$ otherwise. Then, with this measurable function ξ_l , we obtain for every $x_{>l} \in X_{>l}$ that

$$\int_{\mathbb{K}^{l-1}} \left| \int_{\mathbb{K}} F_l(x_{<l}, x_l, x_{>l}) e(-\xi_l(x_{<l}, x_{>l}) x_l) d\lambda_{\mathbb{K}}(x_l) \right| d\lambda_{\mathbb{K}^{l-1}}(x_{<l}) \gtrsim \delta^{O(1)} N^{D_l}.$$

Step 3. For each $x_{>l} \in X_{>l}$, arguing as in Step 3 of the proof of Lemma 4.61, we find a function $g_0 \in L^\infty(\mathbb{K}^k)$ with $|g_0| = |f_{0,l}| = |f_0|$ and, by the pigeonhole principle, an element $x_l \in I_l$ such that the integral

$$\int_{\mathbb{K}^{l-1}} g_0(x) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{l-1} f_i(x - \mathbf{P}_i(t)) \right) e((\xi_l(x_{<l}, x_{>l}), \zeta_{>l}) \cdot \mathcal{P}_{\geq l}(t)) d\lambda_{\mathbb{K}^{l-1}}(x_{<l})$$

admits a lower bound of order $\delta^{O(1)} N^{D_{l-1}}$.

Applying the (l, l) -major arc lemma and the pigeonhole principle, we find a positive integer $q \lesssim \delta^{-O(1)}$ (if $\mathbb{K} = \mathbb{R}$, then we take $q = 1$) and a measurable set $Y_{<l}^{x_{>l}} \subseteq I_{<l}$ such that $|Y_{<l}^{x_{>l}}|_{\mathbb{K}^{l-1}} \gtrsim \delta^{O(1)} |I_{<l}|_{\mathbb{K}^{l-1}}$ and, if $x_{<l} \in Y_{<l}^{x_{>l}}$, then $\|q\xi_l(x_{<l}, x_{>l})\|_{\hat{\mathbb{K}}} \leq C_2 \delta^{-C_2} N^{-d_l}$ for some absolute constant $C_2 \in \mathbb{Z}_+$ independent of C_1 . If $C_1 > C_2$, then

$$|Y_{<l}^{x_{>l}} \cap X_{<l}^{x_{>l}}|_{\mathbb{K}^{l-1}} = |Y_{<l}^{x_{>l}}|_{\mathbb{K}^{l-1}} \gtrsim \delta^{O(1)} |I_{<l}|_{\mathbb{K}^{l-1}}$$

by the definition of ξ_l and our choice of $\xi_0 \in \hat{\mathbb{K}} \setminus \mathfrak{M}$. Let $M_0 := C_2^{-1} C_3 \delta^{C_2 - C_3}$ for a large constant $C_3 \in \mathbb{Z}_+$ to be specified later. We use Lemma 4.48, followed by the pigeonhole principle, to find a frequency $\zeta_l \in \hat{\mathbb{K}}$ and measurable sets $Z_{>l} \subseteq X_{>l}$ satisfying $|Z_{>l}|_{\mathbb{K}^{l-1}} \gtrsim \delta^{O(1)} |I_{>l}|_{\mathbb{K}^{l-1}}$ and $Z_{<l}^{x_{>l}} \subseteq Y_{<l}^{x_{>l}}$ satisfying $|Z_{<l}^{x_{>l}}|_{\mathbb{K}^{l-1}} \gtrsim \delta^{O(1)} |I_{<l}|_{\mathbb{K}^{l-1}}$ for every $x_{>l} \in Z_{>l}$ and such that

$$|\xi_l(x_{<l}, x_{>l}) - \zeta_l| \leq M_0^{-1} N^{-d_l}. \quad (4.75)$$

Such a choice of $\zeta_l \in \hat{\mathbb{K}}$ is possible, since the function $\xi_l: \mathbb{K}^{k-1} \rightarrow \hat{\mathbb{K}}$ is measurable and we have at most $O(\delta^{-O(1)})$ possibilities to choose $\zeta_l \in \hat{\mathbb{K}}$, which is clear from Lemma 4.48. Now inequality (4.74) implies that

$$\left| \int_{\mathbb{K}} g'_0(x) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \left(\prod_{i=1}^{l-1} f_i(x - P_i(t)) \right) e((\xi_l(x_{<l}, x_{>l}), \zeta_{>l}) \cdot \mathcal{P}_{\geq l}(t)) d\lambda_{\mathbb{K}}(x_l) \right|$$

admits a lower bound of order $\delta^{O(1)} N^{d_l}$, where $g'_0 \in L^\infty(\mathbb{K}^k)$ is a 1-bounded function satisfying $|g'_0| = |f_0|$. If C_3 is sufficiently large, then, using (4.75), we can replace $\xi_l(x_{<l}, x_{>l})$ with ζ_l in the integral above and obtain (4.73) with $l-1$ in place of l and with a 1-bounded function $f_{0,l-1} \in L^\infty(\mathbb{K}^k)$ such that $|f_{0,l-1}| = |f_{0,l}| = |f_0|$. This completes the proof of the lemma. $\square$

Now, we can finally prove Theorem 4.33.

Proof of Theorem 4.33. Recall that $D := d_1 + \dots + d_k$. For $k > 1$, we apply Lemma 4.71 to find a frequency vector $\zeta_{>1} \in \hat{\mathbb{K}}^{k-1}$ such that (4.72) holds. By the popularity principle, we find a measurable set $X \subseteq I_{>1}$ satisfying $|X|_{\mathbb{K}^{k-1}} \gtrsim \delta^{O(1)} |I_{>1}|_{\mathbb{K}^{k-1}}$ and such that for every $x_{>1} \in X$ we have

$$\left| \int_{\mathbb{K}} g_{x_{>1}}(x) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} f'(x - P_1(t)) e(\zeta_{>1} \cdot \mathcal{P}_{>1}(t)) d\lambda_{\mathbb{K}}(x) \right| \gtrsim \delta^{O(1)} N^{d_1}, \quad (4.76)$$

whenever $N \gtrsim \delta^{-O(1)}$, where $f'(x) := f_1(x, x_{>1})$ and $g_{x_{>1}} \in L^\infty(\mathbb{K})$ satisfies $|g_{x_{>1}}| = |f_0|$. For $k = 1$, we have $e(\zeta_{>1} \cdot \mathcal{P}_{>1}(t)) := 1$ above and so (4.76) holds. We will proceed in a few steps, now fixing $x_{>1} \in X$ and assuming that $N \geq C_* \delta^{-C_*}$ for some large $C_* \in \mathbb{Z}_+$.

Step 1. Applying the Cauchy–Schwarz inequality to (4.76), we obtain

$$\|F_{x_{>1}}\|_{L^2(\mathbb{K})} \gtrsim \delta^{O(1)} N^{d_1/2}, \quad (4.77)$$

with $F_{x_{>1}}(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} f'(x - P_1(t)) e(\zeta_{>1} \cdot \mathcal{P}_{>1}(t))$, since $\|g_{x_{>1}}\|_{L^2(\mathbb{K})} \lesssim N^{d_1/2}$. By (4.77) and Plancherel's theorem, we obtain

$$\delta^{O(1)} N^{d_1} \lesssim \|F_{x_{>1}}\|_{L^2(\mathbb{K})}^2 = \|\mathcal{F}_{\mathbb{K}}(F_{x_{>1}})\|_{L^2(\hat{\mathbb{K}})}^2, \quad (4.78)$$

where, upon making the change of variables $x \mapsto x + P_1(t)$, we have

$$\mathcal{F}_{\mathbb{K}}(F_{x_{>1}})(\xi) = \mathcal{F}_{1, \mathbb{K}^k} f_1(\xi, x_{>1}) \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} e(\xi P_1(t) + \zeta_{>1} \cdot \mathcal{P}_{>1}(t)).$$

Step 2. Let $M \in \mathbb{Z}_+$ be a large constant to be specified later. By Proposition 4.47, there exists a large constant $C_1 \in \mathbb{Z}_+$ depending only on A and d_k such that if $N \geq C_1 M^{C_1} \delta^{-C_1 M}$ and

$$\left| \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} e(\xi P_1(t) + \zeta_{>1} \cdot \mathcal{P}_{>1}(t)) \right| \geq M^{-1} \delta^M, \quad (4.79)$$

then there exists an integer $q \in [C_1 M^{C_1} \delta^{-C_1 M}]$ (if $\mathbb{K} = \mathbb{R}$, then we take $q = 1$) such that

$$\|q\xi\|_{\hat{\mathbb{K}}} \leq C_1 M^{C_1} \delta^{-C_1 M} N^{-d_1}.$$

Here, it is crucial that the polynomials $P_1, \dots, P_k$ have distinct degrees.

Set $C' := 4C_1M^{C_1}$. Fix $M_1 := C'\delta^{-C'}$, $M_2 := C'\delta^{-C'}N^{-d_1}$, and a smooth and even function $\eta: \mathbb{R} \rightarrow [0, 1]$ satisfying (2.4), and set $\eta_{[\leq M_2]}(\xi) := \eta(M_2^{-1}\xi)$ for $\xi \in \mathbb{R}$. In order to distinguish between major and minor arcs, let us define

$$\Xi_{\mathbb{K}}(\xi) := \sum_{\theta \in \mathcal{R}_{\leq M_1}^{\mathbb{K}}} \eta_{[\leq M_2]}(\xi - \theta)^{1/2}$$

with $\mathcal{R}_{\leq M_1}^{\mathbb{Z}} := \mathcal{R}_{\leq M_1}$ and $\mathcal{R}_{\leq M_1}^{\mathbb{R}} := \{0\}$, where $\mathcal{R}_{\leq M_1}$ is given in (4.12). If $\mathbb{K} = \mathbb{Z}$, then $\Xi_{\mathbb{K}}^2$ is the Fourier transform of the Ionescu–Wainger projection from (4.13), since $\Xi_{\mathbb{K}}(\xi)^2 = \sum_{\theta \in \mathcal{R}_{\leq M_1}^{\mathbb{K}}} \eta_{[\leq M_2]}(\xi - \theta)$ by the disjointness of the supports of $\eta_{[\leq M_2]}(\xi - \theta)^{1/2}$ when $\theta \in \mathcal{R}_{\leq M_1}^{\mathbb{K}}$. To ensure disjointness, we used the condition $N \geq C_*\delta^{-C_*}$, assuming that $C_* \in \mathbb{Z}_+$ is large with respect to C' .

Step 3. Using $\Xi_{\mathbb{K}}$, we can write $\mathcal{F}_{\mathbb{K}}(F_{x_{>1}})(\xi) = G_1(\xi, x_{>1}) + G_2(\xi, x_{>1})$, where

$$\begin{aligned} G_1(\xi, x_{>1}) &:= \mathcal{F}_{1, \mathbb{K}^k} f_1(\xi, x_{>1}) \cdot \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} e(\xi P_1(t) + \zeta_{>1} \cdot \mathcal{P}_{>1}(t)) \cdot (1 - \Xi_{\mathbb{K}}(\xi)), \\ G_2(\xi, x_{>1}) &:= \mathcal{F}_{1, \mathbb{K}^k} f_1(\xi, x_{>1}) \cdot \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} e(\xi P_1(t) + \zeta_{>1} \cdot \mathcal{P}_{>1}(t)) \cdot \Xi_{\mathbb{K}}(\xi). \end{aligned}$$

By (4.79), we have $|G_1(\xi, x_{>1})| \leq M^{-1}\delta^M |\mathcal{F}_{1, \mathbb{K}^k} f_1(\xi, x_{>1})|$. Thus, we have

$$\|G_1(\cdot, x_{>1})\|_{L^2(\mathbb{K})}^2 \lesssim M^{-1}\delta^M N^{d_1}$$

by Plancherel's theorem, since $f_1(\cdot, x_{>1})$ is 1-bounded and supported on a set of size controlled by N^{d_1} . If M is large, then the above bound and (4.78) yield

$$\|\mathcal{F}_{1, \mathbb{K}^k} f_1(\cdot, x_{>1}) \cdot \Xi_{\mathbb{K}}(\cdot)\|_{L^2(\mathbb{K})}^2 \gtrsim \delta^{O(1)} N^{d_1}$$

by $|G_2(\xi, x_{>1})| \leq |\mathcal{F}_{1, \mathbb{K}^k} f_1(\xi, x_{>1}) \cdot \Xi_{\mathbb{K}}(\xi)|$. Integrating over X , we obtain that

$$\left| \int_{\mathbb{K}^{k-1}} \int_{\mathbb{K}} \mathcal{F}_{1, \mathbb{K}^k}(f_1)(\xi, x_{>1}) \overline{\mathcal{F}_{1, \mathbb{K}^k}(f_1)(\xi, x_{>1})} \Xi_{\mathbb{K}}^2(\xi) d\lambda_{\mathbb{K}}(\xi) d\lambda_{\mathbb{K}^{k-1}}(x_{>1}) \right|$$

admits a lower bound of order $\delta^{O(1)} N^D$. By Plancherel's theorem, we have

$$|\langle f_1, \Pi_{\mathbb{K}}^1[\leq M_1, \leq M_2] f_1 \rangle| \gtrsim \delta^{O(1)} N^D.$$

This establishes (4.35) and completes the proof of Theorem 4.33. $\square$

5. MULTILINEAR L^p -IMPROVING INEQUALITIES

In this section, we continue to use the notation introduced in Section 4. Our objective is to establish multilinear L^p -improving inequalities on $\mathbb{Z}^k$ and $\mathbb{R}^k$ for the multilinear averages

$$A_{N; \mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i \in [k]} f_i(x - P_i(t)e_i), \quad x \in \mathbb{K}^k,$$

defined as in (4.4) for any polynomial mapping $\mathcal{P} = (P_1, \dots, P_k): \mathbb{K} \rightarrow \mathbb{K}^k$ with $\deg(P_i) \in \mathbb{Z}_+$ for $i \in [k]$. In contrast to the previous section, the assumption that the degrees of $\mathcal{P}$ are distinct is not needed for the results in this section.

We will be interested in the L^p -improving bounds

$$\left\| A_{N;\mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k) \right\|_{L^q(\mathbb{K}^k)} \lesssim_{\mathcal{P}} N^{-D(\frac{1}{p}-\frac{1}{q})} \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^k)} \quad (5.1)$$

when $p < q$ with $\frac{1}{p} = \frac{1}{p_1} + \dots + \frac{1}{p_k}$, and $D = d_1 + \dots + d_k$ with $d_i = \deg(P_i)$ for $i \in [k]$. These bounds will apply equally well to the truncated averages (4.5).

Our main result here is the following multilinear L^p -improving inequality.

Theorem 5.2. *Fix $k \in \mathbb{N}_{\geq 2}$ and $j \in [k]$. Then we have (5.1) for some $1 < p < q < 2$ and $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p}$ and $p_j = 2$.*

Recall the multilinear adjoint operators $A_{N;\mathbb{K}^k}^{\mathcal{P},*j}$ defined by

$$A_{N;\mathbb{K}^k}^{\mathcal{P},*j}(f_1, \dots, f_k)(x) := \mathbb{E}_{t \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i \in [k]} \mathcal{C}^{\mathbf{1}_{i \neq j}} f_i(x - \mathbf{1}_{i \neq j} P_i(t) e_i + P_j(t) e_j).$$

Corollary 5.3. *Fix $k \in \mathbb{Z}_+$ and $j \in [k]$. Then there exist exponents $1 < r < 2$ and $1 < r_1, \dots, r_k < \infty$ such that $\frac{1}{r_1} + \dots + \frac{1}{r_k} = \frac{1}{r}$ and*

$$\left\| A_{N;\mathbb{K}^k}^{\mathcal{P},*j}(f_1, \dots, f_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim_{\mathcal{P}} N^{-D(\frac{1}{r}-\frac{1}{2})} \prod_{i \in [k]} \|f_i\|_{L^{r_i}(\mathbb{K}^k)}. \quad (5.4)$$

Proof of Corollary 5.3. When $k = 1$, we can use duality to replace the adjoint operator by $A_{N;\mathbb{K}}^{\mathcal{P}}$. If $d_1 = 1$, then (5.4) holds by Young's convolution inequality for $r_1 = \frac{4}{3}$, say. If $\mathbb{K} = \mathbb{Z}$ and $d_1 \geq 2$, then (5.4) holds by a result of Han–Kovač–Lacey–Madrid–Yang in [47]; see also [62, Proposition 6.21]. We refer to [20] and [46] for details in the remaining case $\mathbb{K} = \mathbb{R}$ and $d_1 \geq 2$.

When $k > 1$, let $p_1, \dots, p_k, p, q$ be the exponents from Theorem 5.2 with $p_j = 2$. By duality, Hölder's inequality, and (5.1), we obtain

$$\left\| A_{N;\mathbb{K}^k}^{\mathcal{P},*j}(f_1, \dots, f_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim_{\mathcal{P}} N^{-D(\frac{1}{p}-\frac{1}{q})} \|f_j\|_{L^{q'}(\mathbb{K}^k)} \prod_{i \in [k] \setminus \{j\}} \|f_i\|_{L^{p_i}(\mathbb{K}^k)},$$

where $q' := q/(q-1)$ is the conjugate exponent of q . Taking $r_j = q'$, $r_i = p_i$ for $i \neq j$, and r such that $\frac{1}{r_1} + \dots + \frac{1}{r_k} = \frac{1}{r}$, we obtain $\frac{1}{p} - \frac{1}{q} = \frac{1}{r} - \frac{1}{2}$ so that (5.4) holds. From $1 < p < q < 2$ and $\frac{1}{p} - \frac{1}{q} = \frac{1}{r} - \frac{1}{2}$, we see that $1 < r < 2$. $\square$

5.1. Vinogradov mean value theorem. The proof of Theorem 5.2 in the case $\mathbb{K} = \mathbb{Z}$ uses sharp bounds in the Vinogradov mean value theorem [17]. For fixed $s, d \in \mathbb{Z}_+$, consider the following system of equations

$$\begin{cases} (n_1 + \dots + n_s) - (m_1 + \dots + m_s) = 0, \\ (n_1^2 + \dots + n_s^2) - (m_1^2 + \dots + m_s^2) = 0, \\ \vdots \\ (n_1^d + \dots + n_s^d) - (m_1^d + \dots + m_s^d) = 0. \end{cases}$$

Let $J_{s,d}(N)$ be the number of its integer solutions $(n_1, \dots, n_s, m_1, \dots, m_s) \in [N]^{2s}$ for each fixed $N \in \mathbb{Z}_+$. The Vinogradov mean value theorem states that

$$J_{s,d}(N) \lesssim_\epsilon N^\epsilon (N^s + N^{2s-d(d+1)/2})$$

holds for all $\epsilon > 0$ and, moreover, the N^ϵ term may be removed in the range $s > d(d+1)/2$. This longstanding conjecture has recently been resolved by Wooley [97] in the cubic case and by Bourgain, Demeter, and Guth in [17] in the general case. In [96], Wooley used his efficient congruencing method to give an alternative proof of this conjecture.

In the 1930's, Vinogradov verified the conjecture in the range $s \gtrsim d^2 \log d$. By the usual integral representation of $J_{s,d}(N)$ one has $J_{s,d,\xi}(N) \leq J_{s,d}(N)$ for each $\xi \in \mathbb{Z}^d$, where $J_{s,d,\xi}(N)$ refers to the inhomogeneous system with the vector $(\xi_1, \dots, \xi_d)$ instead of $(0, \dots, 0)$ on the right-hand side.

Theorem 5.2 is a consequence of the following result.

Proposition 5.5. *Fix $k \in \mathbb{N}_{\geq 2}$ and $s_i \in \mathbb{N}_{\geq D^*}$ for all $i \in [k]$, where $D^* := \sum_{i \in [k]} \frac{d_i(d_i+1)}{2}$, and set $S := \sum_{i \in [k]} s_i$. Then, for any distinct $j, j' \in [k]$, we have the following restricted weak-type bound*

$$\left\| A_{N;\mathbb{K}^k}^{\mathcal{P}}(f_1, \dots, f_k) \right\|_{L^{q,\infty}(\mathbb{K}^k)} \lesssim_{\mathcal{P}} N^{-D(\frac{1}{p}-\frac{1}{q})} \prod_{i \in [k]} \|f_i\|_{L^{p_i,1}(\mathbb{K}^k)} \quad (5.6)$$

with $p_j = 2$, $p_{j'} = \frac{2S}{s_{j'}+1}$, and $p_i = \frac{2S}{s_i}$ for $i \notin \{j, j'\}$, where $1 < p < q < 2$ are chosen such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p}$ and $\frac{1}{q} = \frac{2S-s_j}{2S}$.

In order to deduce Theorem 5.2 from Proposition 5.5, we use a multilinear Marcinkiewicz interpolation argument in the spirit of [77, Appendix]. Precisely, we interpolate (5.6) with $k+1$ trivial bounds of the form (5.1) with $p = q$ and no decay in N , which hold for the tuples $(p_1, \dots, p_k)$ belonging to the set

$$\{(1, \infty, \dots, \infty), (\infty, 1, \infty, \dots, \infty), \dots, (\infty, \dots, \infty, 1), (\infty, \dots, \infty)\}.$$

Then Theorem 5.2 holds for some tuple $(p_1, \dots, p_k)$ and $1 < p < q < 2$ such that $p_j = 2$ and the vector $(1/p_1, \dots, 1/p_k, 1 - 1/q)$ lies in the interior of the convex hull of the $k+2$ analogous vectors determined by the corresponding exponents used in the interpolation.

Remark 5.7. If $\mathbb{K} = \mathbb{Z}$, then (5.6) holds with $s_i \in \mathbb{N}_{> d_i(d_i+1)/2}$ for $i \in [k]$.

5.2. Lifting procedure. To deduce (5.6) from Proposition 5.5, we may assume, without loss of generality, that each P_i has no constant term. Furthermore, the bound for $A_{N;\mathbb{K}^k}^{\mathcal{P}}$ will follow from the corresponding bound for the *lifted* multilinear averaging operator

$$A_{N;\mathbb{K}^k}^{\mathcal{P}, \text{lift}}(f_1, \dots, f_k)(x) := \mathbb{E}_{u \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i \in [k]} f_i(x - \Gamma_i(u)), \quad x \in \mathbb{K}^D,$$

where now $f_i: \mathbb{K}^D \rightarrow \mathbb{C}$ for $i \in [k]$. Here Γ_i is defined by

$$\Gamma_i(u) := (\underbrace{0, \dots, 0}_{D_{< i} \text{ times}}, u, u^2, \dots, u^{d_i}, \underbrace{0, \dots, 0}_{D_{> i} \text{ times}}) \in \mathbb{K}^D, \quad (5.8)$$

where $D_{< i} := d_1 + \dots + d_{i-1}$ and $D_{> i} := d_{i+1} + \dots + d_k$ for $i \in [k]$.

Below we extend an argument from [47, Section 5].

Proposition 5.9. *Assume that $P_1(0) = \dots = P_k(0) = 0$ and recall that $D^* = \sum_{i \in [k]} \frac{d_i(d_i+1)}{2}$. Then (5.1) follows from the estimate*

$$\left\| A_{N; \mathbb{K}^k}^{\mathcal{P}, \text{lift}}(f_1, \dots, f_k) \right\|_{L^q(\mathbb{K}^D)} \lesssim_{\mathcal{P}} N^{-D^*(\frac{1}{p} - \frac{1}{q})} \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^D)}, \quad (5.10)$$

and an analogous transference of restricted weak-type bounds also holds.

Proof. For each $x \in \mathbb{K}^D$ we use the notation

$$x = (x_{(1)}, \dots, x_{(k)}), \quad \text{where } x_{(i)} := (x_{D_{<i}+1}, \dots, x_{D_{<i+1}}) \in \mathbb{K}^{d_i},$$

or, alternatively, $x = (x', x'')$ referring to $x'' := (x_{D_{<2}}, \dots, x_{D_{<k+1}}) \in \mathbb{K}^k$ and

$$x' = (x'_{(1)}, \dots, x'_{(k)}), \quad \text{where } x'_{(i)} := (x_{D_{<i}+1}, \dots, x_{D_{<i+1}-1}) \in \mathbb{K}^{d_i-1}.$$

Step 1. We lift functions $g: \mathbb{K}^k \rightarrow \mathbb{C}$ to functions $f: \mathbb{K}^D \rightarrow \mathbb{C}$. The definition differs slightly whether $\mathbb{K} = \mathbb{Z}$ or $\mathbb{K} = \mathbb{R}$ due to the lack of dilation structure on $\mathbb{Z}$. Let $P_i(u) = a_1^i u + \dots + a_{d_i}^i u^{d_i}$ with $a_{d_i}^i \neq 0$. Given $g: \mathbb{Z}^k \rightarrow \mathbb{C}$ and $r_i \in R_i := \mathbb{N}_{<a_{d_i}^i}$, we define the lift $f = f_{g, r_1, \dots, r_k}: \mathbb{Z}^D \rightarrow \mathbb{C}$ of g by

$$f(x) := g(Q_1(x_{(1)}) + r_1, \dots, Q_k(x_{(k)}) + r_k) \mathbf{1}_{E_1}(x'_{(1)}) \cdots \mathbf{1}_{E_k}(x'_{(k)}),$$

where $E_i := [\pm N] \times \dots \times [\pm N^{d_i-1}]$ and

$$Q_i(x_{(i)}) := Q_i(x'_{(i)}, x_{D_{<i+1}}) := a_1^i x_{D_{<i}+1} + \dots + a_{d_i}^i x_{D_{<i+1}}.$$

In the real case, given $g: \mathbb{R}^k \rightarrow \mathbb{C}$, we define the lift $f = f_g: \mathbb{R}^D \rightarrow \mathbb{C}$ of g by

$$f(x) := g(Q_1(x_{(1)}), \dots, Q_k(x_{(k)})) \mathbf{1}_{E_1}(x'_{(1)}) \cdots \mathbf{1}_{E_k}(x'_{(k)}),$$

where $E_i := [\pm N]_{\mathbb{R}} \times \dots \times [\pm N^{d_i-1}]_{\mathbb{R}}$ and Q_i is as before.

Step 2. Fix $p_0 \in [1, \infty)$. Then

$$\|f\|_{L^{p_0}(\mathbb{K}^D)}^{p_0} \lesssim N^{D'} \|g\|_{L^{p_0}(\mathbb{K}^k)}^{p_0} \quad (5.11)$$

holds with $D' := \sum_{i \in [k]} \frac{(d_i-1)d_i}{2}$. Moreover, if $x \in \mathbb{K}^D$ is such that $x'_{(i)} \in E'_i$ for all $i \in [k]$, where $E'_i := [N]_{\mathbb{K}} \times \dots \times [N^{d_i-1}]_{\mathbb{K}}$, then we have the equality

$$A_{N; \mathbb{K}^k}^{\mathcal{P}, \text{lift}}(f_1, \dots, f_k)(x) = A_{N; \mathbb{K}^k}^{\mathcal{P}}(g_1, \dots, g_k)(Q_1(x_{(1)}) + r_1, \dots, Q_k(x_{(k)}) + r_k),$$

with $r_1 = \dots = r_k = 0$ if $\mathbb{K} = \mathbb{R}$.

Step 3. Suppose that (5.10) holds. By the above equality, we have

$$N^{D'} \left\| A_{N; \mathbb{K}^k}^{\mathcal{P}}(g_1, \dots, g_k) \right\|_{L^q(\mathbb{K}^k)}^q \lesssim_{\mathcal{P}} \left\| A_{N; \mathbb{K}^k}^{\mathcal{P}, \text{lift}}(f_1, \dots, f_k) \right\|_{L^q(\mathbb{K}^D)}^q.$$

Combining this with (5.10) and (5.11), we obtain

$$N^{\frac{D'}{q}} \left\| A_{N; \mathbb{K}^k}^{\mathcal{P}}(g_1, \dots, g_k) \right\|_{L^q(\mathbb{K}^k)} \lesssim_{\mathcal{P}} N^{-D^*(\frac{1}{p} - \frac{1}{q})} N^{\frac{D'}{p}} \prod_{i \in [k]} \|g_i\|_{L^{p_i}(\mathbb{K}^k)},$$

which in view of $D^* - D' = \sum_{i \in [k]} d_i = D$ gives (5.1), as desired.

The restricted weak-type estimates are handled analogously. $\square$

5.3. Refinements. By Proposition 5.9 it remains to prove

$$\left\| A_{N;\mathbb{K}^k}^{\mathcal{P}, \text{lift}}(f_1, \dots, f_k) \right\|_{L^{q,\infty}(\mathbb{K}^D)} \lesssim_{\mathcal{P}} N^{-D^*(\frac{1}{p}-\frac{1}{q})} \prod_{i \in [k]} \|f_i\|_{L^{p_i,1}(\mathbb{K}^D)} \quad (5.12)$$

for $p_1, \dots, p_k, p, q$ as in (5.6). We shall use the *refinement method* developed by Christ in [20]. From now on, we abbreviate $A_{N;\mathbb{K}^k}^{\mathcal{P}, \text{lift}}$ to $A_{N;\mathbb{K}}$. We denote the corresponding adjoint operators by $A_{N;\mathbb{K}}^{*i}$.

The bound (5.12) equivalently means that

$$\langle \mathbb{1}_{E_0}, A_{N;\mathbb{K}}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_k}) \rangle \lesssim N^{-D^*(\frac{1}{p}-\frac{1}{q})} |E_0|_{\mathbb{K}^D}^{\frac{1}{q'}} |E_1|_{\mathbb{K}^D}^{\frac{1}{p_1}} \cdots |E_k|_{\mathbb{K}^D}^{\frac{1}{p_k}}$$

holds for all measurable subsets $E_0, E_1, \dots, E_k \subseteq \mathbb{K}^D$. Denoting the left-hand side by K , we can rewrite the above inequality as

$$K^{2S} \lesssim N^{-D^*} |E_0|_{\mathbb{K}^D}^{s_j} |E_j|_{\mathbb{K}^D}^S |E_{j'}|_{\mathbb{K}^D}^{s_{j'}+1} \prod_{i \in [k] \setminus \{j, j'\}} |E_i|_{\mathbb{K}^D}^{s_i}, \quad (5.13)$$

since $\frac{1}{p} = \frac{2S+1-s_j}{2S}$, $\frac{1}{q} = \frac{2S-s_j}{2S}$, $p_j = 2$, $p_{j'} = \frac{2S}{s_{j'}+1}$, and $p_i = \frac{2S}{s_i}$ for $i \notin \{j, j'\}$.

Let $\alpha_0, \alpha_1, \dots, \alpha_k \in [0, 1]$ be parameters such that

$$K = \alpha_0 |E_0|_{\mathbb{K}^D} = \alpha_1 |E_1|_{\mathbb{K}^D} = \cdots = \alpha_k |E_k|_{\mathbb{K}^D}.$$

Note that, for each $i \in [k]$, we have

$$\langle \mathbb{1}_{E_0}, A_{N;\mathbb{K}}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_k}) \rangle = \langle \mathbb{1}_{E_i}, A_{N;\mathbb{K}}^{*i}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_{i-1}}, \mathbb{1}_{E_0}, \mathbb{1}_{E_{i+1}}, \dots, \mathbb{1}_{E_k}) \rangle$$

where, referring to the mapping (5.8), we define

$$A_{N;\mathbb{K}}^{*i}(g_1, \dots, g_k)(x) := \mathbb{E}_{u \in [N]_{\mathbb{K}}}^{\lambda_{\mathbb{K}}} \prod_{i' \in [k]} C^{\mathbb{1}_{i' \neq i}} g_{i'}(x - \mathbb{1}_{i' \neq i} \Gamma_{i'}(u) + \Gamma_i(u)).$$

We observe that the average value of $A_{N;\mathbb{K}}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_k})$ on E_0 is equal to

$$|E_0|_{\mathbb{K}^D}^{-1} \langle A_{N;\mathbb{K}}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_k}), \mathbb{1}_{E_0} \rangle = \alpha_0,$$

and the average value of $A_{N;\mathbb{K}}^{*i}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_{i-1}}, \mathbb{1}_{E_0}, \mathbb{1}_{E_{i+1}}, \dots, \mathbb{1}_{E_k})$ on E_i is α_i . Let $E_i^0 := E_i$.

We recursively define the *refinements* of E_i of level $r \in \mathbb{Z}_+$ by

$$\begin{aligned} E_0^r &:= \left\{ x \in E_0^{r-1} : A_{N;\mathbb{K}}(\mathbb{1}_{E_1^{r-1}}, \mathbb{1}_{E_2^{r-1}}, \dots, \mathbb{1}_{E_{k-1}^{r-1}}, \mathbb{1}_{E_k^{r-1}})(x) \geq c_{r,0} \alpha_0 \right\}, \\ E_1^r &:= \left\{ x \in E_1^{r-1} : A_{N;\mathbb{K}}^{*1}(\mathbb{1}_{E_0^r}, \mathbb{1}_{E_2^{r-1}}, \dots, \mathbb{1}_{E_{k-1}^{r-1}}, \mathbb{1}_{E_k^{r-1}})(x) \geq c_{r,1} \alpha_1 \right\}, \\ &\vdots \\ E_k^r &:= \left\{ x \in E_k^{r-1} : A_{N;\mathbb{K}}^{*k}(\mathbb{1}_{E_1^r}, \mathbb{1}_{E_2^r}, \dots, \mathbb{1}_{E_{k-1}^r}, \mathbb{1}_{E_0^r})(x) \geq c_{r,k} \alpha_k \right\}, \end{aligned}$$

where $(c_{r,0}, c_{r,1}, \dots, c_{r,k}) := (\frac{1}{2^{(r-1)(k+1)+1}}, \frac{1}{2^{(r-1)(k+1)+2}}, \dots, \frac{1}{2^{r(k+1)}})$.

The next result shows that passing to refinements is essentially lossless.

Lemma 5.14. *For all $r \in \mathbb{Z}_+$ and $i \in [k]$, we have*

$$\left\langle \mathbb{1}_{E_0^r}, A_{N;\mathbb{K}}(\mathbb{1}_{E_1^{r-1}}, \dots, \mathbb{1}_{E_k^{r-1}}) \right\rangle \geq c_{r,0}K \quad (5.15)$$

and

$$\left\langle \mathbb{1}_{E_i^r}, A_{N;\mathbb{K}}^*(\mathbb{1}_{E_1^r}, \dots, \mathbb{1}_{E_{i-1}^r}, \mathbb{1}_{E_0^r}, \mathbb{1}_{E_{i+1}^{r-1}}, \dots, \mathbb{1}_{E_k^{r-1}}) \right\rangle \geq c_{r,i}K. \quad (5.16)$$

Proof. Let $r = 1$. We estimate the left-hand side of (5.15) by

$$\left\langle \mathbb{1}_{E_0}, A_{N;\mathbb{K}}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_k}) \right\rangle - \left\langle \mathbb{1}_{E_0 \setminus E_0^1}, A_{N;\mathbb{K}}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_k}) \right\rangle \geq K - K/2 = K/2,$$

splitting $\mathbb{1}_{E_0^1} = \mathbb{1}_{E_0} - \mathbb{1}_{E_0 \setminus E_0^1}$ and then using the definition of E_0^1 . Similarly, for $i = 1$, we rewrite the left-hand side of (5.16) as

$$\left\langle \mathbb{1}_{E_1}, A_{N;\mathbb{K}}^*(\mathbb{1}_{E_0^1}, \mathbb{1}_{E_2}, \dots, \mathbb{1}_{E_k}) \right\rangle - \left\langle \mathbb{1}_{E_1 \setminus E_1^1}, A_{N;\mathbb{K}}^*(\mathbb{1}_{E_0^1}, \mathbb{1}_{E_2}, \dots, \mathbb{1}_{E_k}) \right\rangle$$

which by duality, the previous step, and the definition of E_1^1 is estimated by

$$\left\langle \mathbb{1}_{E_0^1}, A_{N;\mathbb{K}}(\mathbb{1}_{E_1}, \dots, \mathbb{1}_{E_k}) \right\rangle - \left\langle \mathbb{1}_{E_1 \setminus E_1^1}, A_{N;\mathbb{K}}^*(\mathbb{1}_{E_0^1}, \mathbb{1}_{E_2}, \dots, \mathbb{1}_{E_k}) \right\rangle \geq K/2 - K/4.$$

Analogously, for every $i \in [k]$, we have

$$\left\langle \mathbb{1}_{E_i^1}, A_{N;\mathbb{K}}^*(\mathbb{1}_{E_1^1}, \dots, \mathbb{1}_{E_{i-1}^1}, \mathbb{1}_{E_0^1}, \mathbb{1}_{E_{i+1}}, \dots, \mathbb{1}_{E_k}) \right\rangle \geq K/2^{i+1}.$$

By induction, the same argument yields (5.15) and (5.16) for every $r \in \mathbb{Z}_+$. $\square$

5.4. Parameter towers and flows. From now on, by symmetry, we may assume that $j = 1$ and $j' = k$. To establish (5.13) we may assume that $K > 0$ so that by Lemma 5.14 every refinement E_i^r is nonempty. Using α_i as before we can rewrite (5.13) as

$$\alpha_0^{s_1} \alpha_1^{s_2} \alpha_2^{s_3} \dots \alpha_k^{s_k} \lesssim N^{-D^*} |E_k|_{\mathbb{K}^D}. \quad (5.17)$$

We fix some element in one of the refinements, say $z \in E_1^r$, and obtain

$$A_{N;\mathbb{K}}^*(\mathbb{1}_{E_0^r}, \mathbb{1}_{E_2^{r-1}}, \dots, \mathbb{1}_{E_k^{r-1}})(z) \gtrsim_{k,r} \alpha_1.$$

Hence, we obtain that

$$|\{u \in [N]_{\mathbb{K}} : z + \Gamma_1(u) \in E_0^r \text{ and } z + \Gamma_1(u) - \Gamma_i(u) \in E_i^{r-1} \text{ for all } i \in [k] \setminus \{1\}\}|_{\mathbb{K}}$$

admits a lower bound of order $\alpha_1 N$ and, in particular, the parameter set

$$I^z := \{u \in [N]_{\mathbb{K}} : z + \Gamma_1(u) \in E_0^r\}$$

satisfies $|I^z|_{\mathbb{K}} \gtrsim_{k,r} \alpha_1 N$. Next, for each $u \in I^z$, we have $z + \Gamma_1(u) \in E_0^r$. Thus,

$$|\{v \in [N]_{\mathbb{K}} : z + \Gamma_1(u) - \Gamma_i(v) \in E_i^{r-1} \text{ for all } i \in [k]\}|_{\mathbb{K}} \gtrsim_{k,r} \alpha_0 N$$

and, in particular, the parameter set

$$I^{z,u} := \{v \in [N]_{\mathbb{K}} : z + \Gamma_1(u) - \Gamma_1(v) \in E_1^{r-1}\}$$

satisfies $|I^{z,u}|_{\mathbb{K}} \gtrsim_{k,r} \alpha_0 N$. This begins a construction of a parameter tower

$$\mathbf{P}_{(1,0,1)} := \{(u, v) \in [N]_{\mathbb{K}}^2 : u \in I^z, v \in I^{z,u}\}$$

such that $|\mathbf{P}_{(1,0,1)}|_{\mathbb{K}^2} \gtrsim_{k,r} \alpha_1 \alpha_0 N^2$ and a flow

$$\Phi_{(1,0,1)}(u, v) := z + \Gamma_1(u) - \Gamma_1(v) = z + (u - v, \dots, u^{d_1} - v^{d_1}, 0, \dots, 0)$$

such that $\Phi_{(1,0,1)}(\mathbf{P}_{(1,0,1)}) \subseteq E_1^{r-1}$. We express this as $E_1^r \rightarrow E_0^r \rightarrow E_1^{r-1}$.

When we flow from one set E_i into another $E_{i'}$ via the above process, there is no ambiguity which level of refinement we are using:

- (i) if $i < i'$, then necessarily we drop a level, so $E_i^r \rightarrow E_{i'}^{r-1}$;
- (ii) if $i' < i$, then necessarily we stay at the same level, so $E_i^r \rightarrow E_{i'}^r$.

Our flows will always consist of a uniformly controlled finite number of iterations; hence, there is no need to keep track of the level of refinement as long as we start with $z \in E_i^{r_0}$ for a fixed level $r_0 \lesssim 1$ which is sufficiently large to ensure that the last level is positive.

From now on, to simplify notation, we suppress the refinement level, dropping the superscript r . For the flow above, for example, we write $E_1 \rightarrow E_0 \rightarrow E_1$ and $\Phi_{(1,0,1)}(\mathbf{P}_{(1,0,1)}) \subseteq E_1$. Therefore,

$$|E_1|_{\mathbb{K}^D} \geq |\Phi_{(1,0,1)}(\mathbf{P}_{(1,0,1)})|_{\mathbb{K}^D} = \int_{\Phi_{(1,0,1)}(\mathbf{P}_{(1,0,1)})} d\lambda_{\mathbb{K}}^{\otimes d_1}(\xi_{(1)})$$

and the idea is to make the change of variables $\xi_{(1)} = \Phi_{(1,0,1)}(u, v)$.

Example 5.18. For this illustration, let us drop the subscript $(1, 0, 1)$ and suppose for simplicity that $d_1 = D = 2$. When $\mathbb{K} = \mathbb{R}$, we can use the usual change of variables formula involving the Jacobian J_{Φ} . Namely, we have

$$|E_1|_{\mathbb{R}^2} \geq \iint_{\mathbf{P}} |J_{\Phi}(t_1, t_2)| dt_1 dt_2 = \int_{I^z} \int_{I^{z, t_1}} 2|t_1 - t_2| dt_2 dt_1, \quad (5.19)$$

which reduces matters to obtaining an appropriate lower bound for a weighted measure of the parameter set $\mathbf{P}$. The latter can be effectively achieved using sublevel set bounds for J_{Φ} . When $\mathbb{K} = \mathbb{Z}$, we have

$$|E_1|_{\mathbb{Z}^2} \geq \sum_{\xi_{(1)} \in \Phi(\mathbf{P})} 1 = \sum_{\xi_{(1)} \in \Phi(\mathbf{P})} J_{\xi_{(1)}}^{-1} \sum_{(n, m) \in \mathbf{P}} \mathbf{1}_{\Phi(n, m) = \xi_{(1)}},$$

where $J_{\xi_{(1)}} := |\{(n, m) \in \mathbf{P} : \Phi(n, m) = \xi_{(1)}\}|_{\mathbb{Z}^2}$. Interchanging the sums gives

$$|E_1|_{\mathbb{Z}^2} \geq |\Phi(\mathbf{P})| = \sum_{(n, m) \in \mathbf{P}} J_{\Phi(n, m)}^{-1}, \quad (5.20)$$

reducing matters to obtaining a uniform lower bound for the “Jacobian” $J_{\xi_{(1)}}$, where $J_{\xi_{(1)}}$ counts the number of integer solutions $(n, m) \in [N]^2$ to the system of Diophantine equations

$$\begin{cases} z_1 + n - m = \xi_1, \\ z_2 + n^2 - m^2 = \xi_2. \end{cases}$$

Returning to the general case, in both settings the ultimate flow will be a concatenation of k subflows $(\overline{E_1 E_i})$ or $(\overline{E_0 E_i})$, where each $(\overline{E_{i'} E_i})$ is itself a flow consisting of some number of $E_{i'}$ and E_i alternating. When $\mathbb{K} = \mathbb{Z}$, the ultimate flow has the form

$$(\overline{E_1 E_0}) \rightarrow (\overline{E_1 E_2}) \rightarrow (\overline{E_1 E_3}) \rightarrow \dots \rightarrow (\overline{E_1 E_k}). \quad (5.21)$$

When $\mathbb{K} = \mathbb{R}$, the ultimate flow is slightly different and has the form

$$(\overline{E_0 E_1}) \rightarrow (\overline{E_0 E_2}) \rightarrow \cdots \rightarrow (\overline{E_0 E_k}). \quad (5.22)$$

For example, the flow $E_1 \rightarrow E_0 \rightarrow E_1 \rightarrow E_2 \rightarrow E_1$ produces a parameter tower

$$\mathbf{P}_{(1,0,1,2,1)} := \{(u, v, u', v') \in [N]_{\mathbb{K}}^4 : u \in I^z, v \in I^{z,u}, u' \in I^{z,u,v}, v' \in I^{z,u,v,u'}\}$$

with $|\mathbf{P}_{(1,0,1,2,1)}|_{\mathbb{K}^4} \gtrsim_{k,r} \alpha_1 \alpha_2 \alpha_1 \alpha_0 N^4$ and $\Phi_{(1,0,1,2,1)}(\mathbf{P}_{(1,0,1,2,1)}) \subseteq E_1$, where

$$\Phi_{(1,0,1,2,1)}(u, v, u', v') := z + \Gamma_1(u) - \Gamma_1(v) + \Gamma_1(u') - \Gamma_2(u') - \Gamma_1(v') + \Gamma_2(v').$$

5.5. The integer case. We now prove Proposition 5.5 in the integer setting.

Proof of Proposition 5.5 for $\mathbb{K} = \mathbb{Z}$. Suppose that the parameters i from (5.21) determine a sequence $\mathcal{I}$ of the following form

$$\underbrace{(1, 0, \dots, 1, 0)}_{2s_1 \text{ elements}}, \underbrace{(1, 2, \dots, 1, 2)}_{2s_2 \text{ elements}}, \underbrace{(1, 3, \dots, 1, 3)}_{2s_3 \text{ elements}}, \dots, \underbrace{(1, k-1, \dots, 1, k-1)}_{2s_{k-1} \text{ elements}}, \underbrace{(k, 1, \dots, 1, k)}_{2s_k+1 \text{ elements}}$$

with a slightly different final block, where we used the transition $E_{k-1} \rightarrow E_k$.

Step 1. Unpacking all this gives a parameter tower

$$\mathbf{P}_{\mathcal{I}} := \{(\mathbf{n}, \mathbf{m}) \in [N]^{2S} : \mathbf{n}_1 \in I^z, \mathbf{m}_1 \in I^{z, \mathbf{n}_1}, \dots, \mathbf{m}_S \in I^{z, \mathbf{n}_1, \mathbf{m}_1, \dots, \mathbf{n}_S}\},$$

where $r_0 := 10S$ and $z \in E_1^{r_0} \subseteq \mathbb{Z}^D$ is fixed. Thus, we have

$$|\mathbf{P}_{\mathcal{I}}|_{\mathbb{Z}^D} \gtrsim_{\mathcal{P}} \alpha_0^{s_1} \alpha_1^S \alpha_2^{s_2} \cdots \alpha_k^{s_k} N^{2S}.$$

The corresponding flow map $\Phi_{\mathcal{I}}(\mathbf{n}, \mathbf{m})$ can be expressed as

$$z - (\xi_{\mathbf{n}, \mathbf{m}}^1 - \Phi_{\mathcal{I}}^{(1)}(\mathbf{n}, \mathbf{m}), \Phi_{\mathcal{I}}^{(2)}(\mathbf{n}, \mathbf{m}), \dots, \Phi_{\mathcal{I}}^{(k-1)}(\mathbf{n}, \mathbf{m}), \xi_{\mathbf{n}, \mathbf{m}}^k - \Phi_{\mathcal{I}}^{(k)}(\mathbf{n}, \mathbf{m})),$$

where, writing $\mathbf{n} = (\mathbf{n}_{(1)}, \dots, \mathbf{n}_{(k)})$, $\mathbf{m} = (\mathbf{m}_{(1)}, \dots, \mathbf{m}_{(k)}) \in [N]^{s_1} \times \cdots \times [N]^{s_k}$, we let $\Phi_{\mathcal{I}}^{(i)}(\mathbf{n}, \mathbf{m}) := \Phi_i(\mathbf{n}_{(i)}, \mathbf{m}_{(i)})$ for all $i \in [k]$, with $\Phi_i(u, v)$ defined by

$$((u_1 + \cdots + u_{s_i}) - (v_1 + \cdots + v_{s_i}), \dots, (u_1^{d_i} + \cdots + u_{s_i}^{d_i}) - (v_1^{d_i} + \cdots + v_{s_i}^{d_i}))$$

for $u, v \in [N]^{s_i}$, and $\xi_{\mathbf{n}, \mathbf{m}}^1 \in \mathbb{Z}^{d_1}$ is independent of $\mathbf{n}_{(1)}, \mathbf{m}_{(1)}$, while $\xi_{\mathbf{n}, \mathbf{m}}^k \in \mathbb{Z}^{d_k}$ depends solely on $\mathbf{n}_{(k-1)}, \mathbf{m}_{(k-1)}$ if $k \geq 3$ or simply vanishes if $k = 2$.

Step 2. We drop the subscript $\mathcal{I}$. Since $\Phi(\mathbf{P}) \subseteq E_k$, we obtain

$$|E_k|_{\mathbb{Z}^D} \geq \sum_{(\mathbf{n}, \mathbf{m}) \in \mathbf{P}} J_{\Phi(\mathbf{n}, \mathbf{m})}^{-1},$$

proceeding as in (5.20). Next, we will find a uniform upper bound for J_{ξ} , where $\xi \in \mathbb{Z}^D$.

We will count all integer solutions $(\mathbf{n}, \mathbf{m}) \in [N]^{2S}$ to the system

$$\begin{cases} \xi_{\mathbf{n}, \mathbf{m}}^1 - \Phi_1(\mathbf{n}_{(1)}, \mathbf{m}_{(1)}) = z_{(1)} - \xi_{(1)}, \\ \Phi_2(\mathbf{n}_{(2)}, \mathbf{m}_{(2)}) = z_{(2)} - \xi_{(2)}, \\ \vdots \\ \Phi_{k-1}(\mathbf{n}_{(k-1)}, \mathbf{m}_{(k-1)}) = z_{(k-1)} - \xi_{(k-1)}, \\ \xi_{\mathbf{n}, \mathbf{m}}^k - \Phi_k(\mathbf{n}_{(k)}, \mathbf{m}_{(k)}) = z_{(k)} - \xi_{(k)}, \end{cases}$$

identifying $z, \xi \in \mathbb{Z}^D$ with $(z_{(1)}, \dots, z_{(k)}), \xi = (\xi_{(1)}, \dots, \xi_{(k)}) \in \mathbb{Z}^{d_1} \times \dots \times \mathbb{Z}^{d_k}$.

Step 3. Using $s_i > \frac{d_i(d_i+1)}{2}$, the definition of Φ_i , and the Vinogradov mean value theorem, we conclude that the subsystem consisting of all but the first and last equations has at most

$$J_{s_2, d_2}(N) \cdots J_{s_{k-1}, d_{k-1}}(N) \lesssim N^{2(s_2 + \dots + s_{k-1}) - (d_2(d_2+1) + \dots + d_{k-1}(d_{k-1}+1))/2}$$

solutions. Once these solutions have been counted, the variables $\mathbf{n}_{(k-1)}, \mathbf{m}_{(k-1)}$ are determined and we can bound the number of solutions to the last equation by $J_{s_k, d_k}(N) \lesssim N^{2s_k - d_k(d_k+1)/2}$. Once all variables $\mathbf{n}_{(i)}, \mathbf{m}_{(i)}$ with $i > 1$ have been determined, we can bound the number of solutions to the first equation by $J_{s_1, d_1}(N) \lesssim N^{2s_1 - d_1(d_1+1)/2}$. Altogether, we have $J_\xi \lesssim N^{2S-D^*}$. Thus,

$$|E_k|_{\mathbb{Z}^D} \gtrsim N^{-2S+D^*} |\mathbf{P}|_{\mathbb{Z}^D} \gtrsim N^{-2S+D^*} \alpha_0^{s_1} \alpha_1^S \alpha_2^{s_2} \cdots \alpha_k^{s_k} N^{2S}$$

and (5.17) holds. This completes the proof of Proposition 5.5 for $\mathbb{K} = \mathbb{Z}$. $\square$

5.6. The real case. We now prove Proposition 5.5 in the real setting.

Proof of Proposition 5.5 for $\mathbb{K} = \mathbb{R}$. In this case, we will produce exactly D equations in order to follow the Jacobian approach (5.19). In particular, the flow (5.22) now depends on the parities of the degrees d_i of the polynomials P_i . Precisely, each block $(\overline{E_0} \ E_i)$ in (5.22) ends with E_i and is of length d_i , except the last one which ends with E_k but is of length $d_k + 1$. Therefore, the parameters i from (5.22) determine a sequence $\mathcal{I}$ of the following form

$$(\omega_1^1, \dots, \omega_{d_1}^1, \omega_1^2, \dots, \omega_{d_2}^2, \dots, \omega_1^{k-1}, \dots, \omega_{d_{k-1}}^{k-1}, \omega_1^k, \dots, \omega_{d_k+1}^k),$$

where

$$\begin{aligned} (\omega_1^l, \dots, \omega_{d_l}^l) &:= \begin{cases} (0, l, \dots, 0, l) & \text{if } d_l \text{ is even,} \\ (l, 0, \dots, l, 0, l) & \text{if } d_l \text{ is odd,} \end{cases} \quad \text{for } l \in [k-1], \\ (\omega_1^k, \dots, \omega_{d_k+1}^k) &:= \begin{cases} (k, 0, \dots, k, 0, k) & \text{if } d_k \text{ is even,} \\ (0, k, \dots, 0, k) & \text{if } d_k \text{ is odd.} \end{cases} \end{aligned}$$

For example, if all d_i are even, then $\mathcal{I}$ takes the form

$$\underbrace{(0, 1, \dots, 0, 1)}_{d_1 \text{ elements}}, \underbrace{(0, 2, \dots, 0, 2)}_{d_2 \text{ elements}}, \dots, \underbrace{(0, k-1, \dots, 0, k-1)}_{d_{k-1} \text{ elements}}, \underbrace{(k, 0, \dots, k, 0, k)}_{d_k+1 \text{ elements}},$$

whereas if all d_i are odd, then $\mathcal{I}$ takes the form

$$\underbrace{(1, 0, \dots, 1, 0, 1)}_{d_1 \text{ elements}}, \underbrace{(2, 0, \dots, 2, 0, 2)}_{d_2 \text{ elements}}, \dots, \underbrace{(k-1, 0, \dots, k-1, 0, k-1)}_{d_{k-1} \text{ elements}}, \underbrace{(0, k, \dots, 0, k)}_{d_k+1 \text{ elements}}.$$

Step 1. Set $r_0 := 10D$. Unpacking all this gives a parameter tower

$$\mathbf{P}_{\mathcal{I}} := \{t \in [N]_{\mathbb{R}}^D : t_1 \in I^z, t_2 \in I^{z, t_1}, \dots, t_D \in I^{z, t_1, t_2, \dots, t_{D-1}}\},$$

for some fixed $z \in E_0^{r_0}$ if d_1 is even or $z \in E_1^{r_0}$ if d_1 is odd, and the flow map

$$\Phi_{\mathcal{I}}(\mathbf{t}) := z - (\xi_t^1 + \Phi_1(\mathbf{t}_{(1)}), \xi_t^2 + \Phi_2(\mathbf{t}_{(2)}), \dots, \xi_t^k + \Phi_k(\mathbf{t}_{(k)})),$$

where, writing $\mathbf{t} = (\mathbf{t}_{(1)}, \dots, \mathbf{t}_{(k)})$, $\xi_{\mathbf{t}} = (\xi_{\mathbf{t}}^1, \dots, \xi_{\mathbf{t}}^k) \in [N]_{\mathbb{R}}^{d_1} \times \dots \times [N]_{\mathbb{R}}^{d_k}$ as before, for each $i \in [k]$ such that d_i is even we have

$$\Phi_i(u_1, \dots, u_{d_i}) := (u_1 - u_2 + \dots + u_{d_i-1} - u_{d_i}, \dots, u_1^{d_i} - u_2^{d_i} + \dots + u_{d_i-1}^{d_i} - u_{d_i}^{d_i}),$$

whereas for each $i \in [k]$ such that d_i is odd we have

$$\Phi_i(u_1, \dots, u_{d_i}) := -(u_1 - u_2 + \dots - u_{d_i-1} + u_{d_i}, \dots, u_1^{d_i} - u_2^{d_i} + \dots - u_{d_i-1}^{d_i} + u_{d_i}^{d_i}),$$

and $\xi_{\mathbf{t}}^1$ vanishes, whereas if $i \neq 1$, then $\mathbf{t} \mapsto \xi_{\mathbf{t}}^i$ is a polynomial that may depend on the coordinates $\mathbf{t}_{(i')}$ with $i' < i$ but is independent of $\mathbf{t}_{(i')}$ with $i' \geq i$.

Step 2. We drop the subscript $\mathcal{I}$. Since $\Phi(\mathbf{P}) \subseteq E_k$, we obtain

$$|E_k|_{\mathbb{R}^D} \geq |\Phi(\mathbf{P})|_{\mathbb{R}^D} \gtrsim \int_{\mathbf{P}} |J_{\Phi}(\mathbf{t})| d\mathbf{t}, \quad (5.23)$$

and the Jacobian matrix Φ' , consisting of the first order partial derivatives of the polynomial map Φ , takes the following form

$$\begin{pmatrix} \mathbf{W}_1 & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{M}_{2,1} & \mathbf{W}_2 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{M}_{3,1} & \mathbf{M}_{3,2} & \mathbf{W}_3 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{k,1} & \mathbf{M}_{k,2} & \mathbf{M}_{k,3} & \dots & \mathbf{W}_k \end{pmatrix}$$

and for each block $\mathbf{W}_i$, after removing signs from every second row, we have

$$|\det \mathbf{W}_i(\mathbf{t})| = \left| \det \begin{pmatrix} 1 & 2\mathbf{t}_{D_{<i}+1} & 3\mathbf{t}_{D_{<i}+1}^2 & \dots & d_i \mathbf{t}_{D_{<i}+1}^{d_i-1} \\ 1 & 2\mathbf{t}_{D_{<i}+2} & 3\mathbf{t}_{D_{<i}+2}^2 & \dots & d_i \mathbf{t}_{D_{<i}+2}^{d_i-1} \\ 1 & 2\mathbf{t}_{D_{<i}+3} & 3\mathbf{t}_{D_{<i}+3}^2 & \dots & d_i \mathbf{t}_{D_{<i}+3}^{d_i-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2\mathbf{t}_{D_{<i}+1} & 3\mathbf{t}_{D_{<i}+1}^2 & \dots & d_i \mathbf{t}_{D_{<i}+1}^{d_i-1} \end{pmatrix} \right|$$

and therefore, using properties of the Vandermonde matrix, we conclude that

$$|J_{\Phi}(\mathbf{t})| = |\det \Phi'(\mathbf{t})| = \prod_{i \in [k]} \left(d_i! \prod_{\substack{j_i, j'_i \in [D_{<i+1}] \setminus [D_{<i}] \\ j_i < j'_i}} |\mathbf{t}_{j_i} - \mathbf{t}_{j'_i}| \right).$$

Step 3. Consider the innermost integral in (5.23). For fixed $\mathbf{t}_1, \dots, \mathbf{t}_{D-1}$ we integrate $|J_{\Phi}(\mathbf{t})|$ over $\mathbf{t}_D \in I^{z, \mathbf{t}_1, \dots, \mathbf{t}_{D-1}} \subseteq [0, N]$, where $|I^{z, \mathbf{t}_1, \dots, \mathbf{t}_{D-1}}|_{\mathbb{R}} \gtrsim_{k, r_0} \alpha_k N$. Let $I \subseteq I^{z, \mathbf{t}_1, \dots, \mathbf{t}_{D-1}}$ be such that $|I|_{\mathbb{R}} \gtrsim_{\mathcal{P}} \alpha_k N$ and $|\mathbf{t}_j - \mathbf{t}_D| \gtrsim_{\mathcal{P}} \alpha_k N$ for all $j \in [D-1] \setminus [D-d_k]$ and $\mathbf{t}_D \in I$. We pull out all factors independent of $\mathbf{t}_D$ and estimate the remaining part as follows

$$\int_{I^{z, \mathbf{t}_1, \dots, \mathbf{t}_{D-1}}} |\mathbf{t}_{D-d_k+1} - \mathbf{t}_D| \dots |\mathbf{t}_{D-1} - \mathbf{t}_D| d\mathbf{t}_D \gtrsim_{\mathcal{P}} |I|_{\mathbb{R}} (\alpha_k N)^{d_k-1} \gtrsim_{\mathcal{P}} (\alpha_k N)^{d_k}$$

with some constant that is uniform in $\mathbf{t}_1, \dots, \mathbf{t}_{D-1}$. Continuing this way, we obtain the uniform lower bound $(\alpha_0 N)^{d_k-1}$ for the analogous integral over $\mathbf{t}_{D-1}$ and after d_k such steps, using $\alpha_i \in (0, 1]$, we end up with the integrand

$$\prod_{i \in [k-1]} \left(d_i! \prod_{\substack{j_i, j'_i \in [D_{<i+1}] \setminus [D_{<i}] \\ j_i < j'_i}} |\mathbf{t}_{j_i} - \mathbf{t}_{j'_i}| \right) (\alpha_0 \alpha_1 \cdots \alpha_k N)^{\frac{d_k(d_k+1)}{2}}.$$

Repeating this for other blocks leads to $|E_k|_{\mathbb{R}^D} \gtrsim_{\mathcal{P}} (\alpha_0 \alpha_1 \cdots \alpha_k N)^{D^*}$ and, since $s_i \geq D^*$ for every $i \in [k]$, this implies $|E_k|_{\mathbb{R}^D} \gtrsim_{\mathcal{P}} \alpha_0^{s_1} \alpha_1^{s_2} \cdots \alpha_k^{s_k} N^{D^*}$ so that (5.17) holds. This completes the proof of Proposition 5.5 for $\mathbb{K} = \mathbb{R}$. $\square$

6. MULTILINEAR WEYL INEQUALITY AND SOBOLEV SMOOTHING THEOREM

In this section, we will formulate and prove a multilinear Weyl inequality, which asserts that the averaging operator $\tilde{A}_{N; \mathbb{Z}^k}^{P_1(\mathbf{n}), \dots, P_k(\mathbf{n})}(f_1, \dots, f_k)$ is negligible when the j -th Fourier transform of f_j , for at least one $j \in [k]$, vanishes on appropriate major arcs. We will also need a multilinear Sobolev smoothing inequality, which can be thought of as a continuous variant of the multilinear Weyl inequality for the averaging operators $\tilde{A}_{N; \mathbb{R}^k}^{P_1(\mathbf{t}), \dots, P_k(\mathbf{t})}(f_1, \dots, f_k)$.

6.1. Multilinear smoothing inequalities. Using notation from Section 4, in particular (4.6)–(4.12), we formulate the multilinear Weyl inequality and the multilinear Sobolev smoothing inequality as the following unified theorem.

Theorem 6.1. *Let $\mathbb{K}$ be either $\mathbb{Z}$ or $\mathbb{R}$. Fix $C_1, C_2 \in \mathbb{R}_+$ and $k \in \mathbb{Z}_+$, and let $\mathcal{P} = (P_1, \dots, P_k)$ be a polynomial mapping satisfying (4.1)–(4.3). Let $1 < p_1, \dots, p_k < \infty$ be exponents such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$. Then there exists a small constant $c \in (0, 1)$, depending only on $k, \mathcal{P}, p_1, \dots, p_k, p, C_1, C_2$, such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq 1$. Let $f_i \in L^{p_i}(\mathbb{K}^k)$ for all $i \in [k]$. If $f_j \in L^2(\mathbb{K}^k)$ for some $j \in [k]$ and the j -th Fourier transform $\mathcal{F}_{j, \mathbb{K}^k} f_j$ vanishes on the major arcs $\mathfrak{M}_{\leq N^{-d_j} \delta - C_2}^j(\mathcal{R}_{\leq \delta - C_1}^{\mathbb{K}})$, then*

$$\left\| \tilde{A}_{N; \mathbb{K}^k}^{P_1, \dots, P_k}(f_1, \dots, f_k) \right\|_{L^p(\mathbb{K}^k)} \leq c^{-1} (\delta^c + N^{-c}) \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^k)}. \quad (6.2)$$

The same conclusion holds for $A_{N; \mathbb{K}^k}^{P_1, \dots, P_k}$ in place of $\tilde{A}_{N; \mathbb{K}^k}^{P_1, \dots, P_k}$ in (6.2).

If $\mathbb{K} = \mathbb{Z}$, then (6.2) is the multilinear Weyl inequality. If $\mathbb{K} = \mathbb{R}$, then (6.2) is the multilinear Sobolev smoothing inequality. Theorem 6.1 plays an essential role in our arguments and will be used repeatedly. Here, in contrast to [62], we need this theorem for $\mathbb{K} = \mathbb{Z}$ and $\mathbb{K} = \mathbb{R}$. In [62], a bilinear variant of (6.2) for the Furstenberg–Weiss averages and $\mathbb{K} = \mathbb{Z}$ was established with $\delta^c + \langle \text{Log } N \rangle^{-c\tilde{C}}$ in place of $\delta^c + N^{-c}$; see (1.38). Although in [62] the decay $\delta^c + \langle \text{Log } N \rangle^{-c\tilde{C}}$ was sufficient, in our situation we need $\delta^c + N^{-c}$ in (6.2). This bound can be derived using Theorem 3.3. We abbreviate $\tilde{A}_{N; \mathbb{K}^k}^{P_1, \dots, P_k}$ to $\tilde{A}_{N; \mathbb{K}^k}$.

6.2. Structural theorem for adjoints. We now prove a structural theorem for the adjoint multilinear operators $\tilde{A}_{N;\mathbb{K}^k}^{*j}(f_1, \dots, f_k)$ from (4.9). The key tool will be the following variant of the Hahn–Banach theorem.

Lemma 6.3. *Let $(X, \mathcal{B}(X), \mu)$ be a σ -finite measure space. Given $A, B > 0$, $G \in L^2(X)$, and $\Phi \subseteq L^2(X)$, suppose that the following inverse theorem holds: if $f \in L^2(X)$ satisfies $\|f\|_{L^\infty(X)} \leq 1$ and $|\langle f, G \rangle| > A$, then $|\langle f, \phi \rangle| > B$ for some $\phi \in \Phi$. Then G lies in the closed convex hull of the set*

$$V := \{\lambda\phi \in L^2(X) : \phi \in \Phi \text{ and } |\lambda| \leq A/B\} \cup \{h \in L^2(X) : \|h\|_{L^1(X)} \leq A\}.$$

Proof. We refer to [62, Lemma 6.9]. $\square$

Lemma 6.3 combined with the inverse theorem from Section 4 will reveal a major and minor arc structure for the multilinear operators $\tilde{A}_{N;\mathbb{K}^k}^{*j}(f_1, \dots, f_k)$. In the case $\mathbb{K} = \mathbb{R}$, it still makes sense to refer to the major and minor arcs, even though we have exactly one major arc around 0. The following theorem will play an essential role in the proof of Theorem 6.1.

Theorem 6.4. *Let $\mathbb{K}$ be either $\mathbb{Z}$ or $\mathbb{R}$. Fix $C'_0 \in \mathbb{Z}_+$ and $k \in \mathbb{Z}_+$, and let $\mathcal{P} = (P_1, \dots, P_k)$ be a polynomial mapping satisfying (4.1)–(4.3). Then there exists a large constant $C' \in \mathbb{Z}_+$, depending only on $\mathcal{P}$ and C'_0 , such that the following holds. Assume that $\delta_0 \in (0, 1]$ and $N \geq C'\delta_0^{-C'}$, and fix $j \in [k]$. Then for all 1-bounded functions $f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ supported on $\prod_{i \in [k]} [\pm N_i]_{\mathbb{K}}$, where $N_i := C'_0 N^{d_i}$ for each $i \in [k]$, one can decompose*

$$\tilde{A}_{N;\mathbb{K}^k}^{*j}(f_1, \dots, f_k) = F + E, \tag{6.5}$$

where $F \in L^2(\mathbb{K}^k)$ and $\mathcal{F}_{j,\mathbb{K}^k} F$ is supported on the set

$$\prod_{i \in [j-1]} [\pm C' N_i]_{\mathbb{K}} \times \mathfrak{M}_{\leq C'\delta_0^{-C'} N^{-d_j}} \left(\mathcal{R}_{\leq C'\delta_0^{-C'}}^{\mathbb{K}} \right) \times \prod_{i \in [k] \setminus [j]} [\pm C' N_i]_{\mathbb{K}}. \tag{6.6}$$

Furthermore, we have

$$\|F\|_{L^\infty(\mathbb{K}^k)} \leq C'\delta_0^{-C'} \quad \text{and} \quad \|F\|_{L^1(\mathbb{K}^k)} \leq C'\delta_0^{-C'} N^D, \tag{6.7}$$

where $D := d_1 + \dots + d_k$, while the error term satisfies

$$\|E\|_{L^1(\mathbb{K}^k)} \leq C'_1 \delta_0 N^D. \tag{6.8}$$

The same conclusion holds for $A_{N;\mathbb{K}^k}^{*j}$ in place of $\tilde{A}_{N;\mathbb{K}^k}^{*j}$ in (6.5).

In the case $\mathbb{K} = \mathbb{Z}$, by (6.6), the j -th Fourier transform of F in the sum on the right-hand side of (6.5) is supported on the major arcs corresponding to the canonical fractions $\mathcal{R}_{\leq N_1}$ with $N_1 = C'\delta_0^{-C'}$. This part can be thought of as the major arcs part of $\tilde{A}_{N;\mathbb{Z}^k}^{*j}(f_1, \dots, f_k)$, whereas the error function E from (6.5) is negligible thanks to (6.8). From Theorem 6.13 below it will become clear that E can be thought of as the minor arcs part of $\tilde{A}_{N;\mathbb{Z}^k}^{*j}(f_1, \dots, f_k)$.

Proof of Theorem 6.4. The proof is a combination of Lemma 6.3 and Theorem 4.14; see [62, Corollary 6.10] or [61, Corollary 7.13] for details. $\square$

6.3. Proof of multilinear Weyl inequality. Here we will use the Ionescu–Wainger projections $\Pi_{\mathbb{K}}^j[\leq M_1, \leq M_2]$ from (4.13) with $M_1 := \delta^{-C}$ and $M_2 := \delta^{-C} N^{-d_j}$, where $\delta \in (0, 1)$ is small and $C \in \mathbb{R}_+$ is large depending on C_1, C_2 from Theorem 6.1. We make the following important remark.

Remark 6.9. Given a large integer $p_0 \in 2\mathbb{Z}$, there exists a large constant $C_{p_0} \geq 1$, possibly depending on C but independent of N and δ , such that if

$$N \geq C_{p_0} \delta^{-C_{p_0}}, \quad (6.10)$$

then for each fixed $\varepsilon \in (0, 1)$ and each $f \in L^q(\mathbb{K}^k)$ with $q \in [p'_0, p_0]$, we have

$$\left\| \Pi_{\mathbb{K}}^j \left[\leq \delta^{-C}, \leq N^{-d_j} \delta^{-C} \right] f \right\|_{L^q(\mathbb{K}^k)} \lesssim \delta^{-\varepsilon C} \|f\|_{L^q(\mathbb{K}^k)}. \quad (6.11)$$

- (i) If $\mathbb{K} = \mathbb{R}$, then $\mathcal{R}_{\leq M_1}^{\mathbb{R}} = \{0\}$ and so (6.11) holds for all $q \in [1, \infty]$ by Minkowski's convolution inequality with $\|\eta\|_{L^1(\mathbb{R})}$ in place of $\delta^{-\varepsilon C}$.
- (ii) If $\mathbb{K} = \mathbb{Z}$, then the situation is more difficult, but condition (6.10) allows us to use Theorem 3.3 to derive inequality (6.11).

Moreover, for intervals $I, J \subset \mathbb{R}$ of finite length and f as before, we have

$$\left\| \Pi_{\mathbb{K}}^j \left[\leq \delta^{-C}, \leq N^{-d_j} \delta^{-C} \right] f \right\|_{L^q(J_j)} \lesssim \delta^{-\varepsilon C} \left\langle \frac{\text{dist}(I, J)}{N^{d_j} \delta^C} \right\rangle^{-10} \|f\|_{L^q(I_j)}, \quad (6.12)$$

where $I_j := \mathbb{K}^{j-1} \times (I \cap \mathbb{K}) \times \mathbb{K}^{k-j}$ and $J_j := \mathbb{K}^{j-1} \times (J \cap \mathbb{K}) \times \mathbb{K}^{k-j}$.

- (iii) If $\mathbb{K} = \mathbb{R}$, then (6.12) follows, since $|\mathcal{F}_{\mathbb{R}}^{-1} \eta_{[\leq M_2]}(x)| \lesssim_{\eta} M_2 \langle M_2 x \rangle^{-20}$.
- (iv) If $\mathbb{K} = \mathbb{Z}$, then (6.12) follows by repeating the proof of [62, Lemma 5.17, p. 1045], with Theorem 3.3 in place of [62, Theorem 5.7, p. 1041].

The next theorem combined with the Ionescu–Wainger projections (4.13) will explain why the error term E from (6.5) satisfying (6.8) can be thought of as the minor arcs piece of the adjoint operator $\tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k)$. In what follows, we choose a very large integer $p_0 \in 2\mathbb{Z}$, depending on $k, \mathcal{P}, p_1, \dots, p_k, p, C$ but independent of N, δ , such that all interpolation arguments using (6.11) or (6.12) from Remark 6.9 will refer only to $q \in [p'_0, p_0]$.

Theorem 6.13. *Let $\mathbb{K}$ be either $\mathbb{Z}$ or $\mathbb{R}$. Fix $C \in \mathbb{R}_+$ and $k \in \mathbb{Z}_+$, and let $\mathcal{P} = (P_1, \dots, P_k)$ be a polynomial mapping satisfying (4.1)–(4.3). Let $1 < p_1, \dots, p_k < \infty$ be exponents such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} < 1$. Then there exists a small constant $c_0 \in (0, 1)$, depending only on $k, \mathcal{P}, p_1, \dots, p_k, p, C$, such that the following holds. Assume that $\delta \in (0, 1]$ and $N \geq 1$. Let $f_i \in L^{p_i}(\mathbb{K}^k)$ for all $i \in [k]$. If (6.10) holds, then for each $j \in [k]$ we have*

$$\left\| (\mathbb{I} - \Pi_{\mathbb{K}}^j) \tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k) \right\|_{L^p(\mathbb{K}^k)} \leq c_0^{-1} \delta^{c_0} \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^k)}, \quad (6.14)$$

where $\mathbb{I}$ and $\Pi_{\mathbb{K}}^j$ are the identity operator and $\Pi_{\mathbb{K}}^j[\leq \delta^{-C}, \leq N^{-d_j} \delta^{-C}]$, respectively. The same conclusion holds for $A_{N; \mathbb{K}^k}^{*j}$ in place of $\tilde{A}_{N; \mathbb{K}^k}^{*j}$ in (6.14).

Proof. The proof is fairly involved. We will adapt the arguments from [62]. Due to subtle differences between our proof and the proof of [62, Theorem 5.12(i)] we present the key details. We assume that δ is small; otherwise (6.14) follows from (6.11) and Hölder's inequality. We allow explicit and implicit constants to depend on $k, \mathcal{P}, p_1, \dots, p_k, p, C$ but not on N, δ . **Step 1.** Fix $C''_0 \in \mathbb{Z}_+$. We show, with a small constant $c_1 \in (0, 1)$, the estimate

$$\left\| (\mathbb{I} - \Pi_{\mathbb{K}}^j) \tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim \delta^{c_1} N^{D/2} \prod_{i \in [k]} \|f_i\|_{L^\infty(\mathbb{K}^k)} \quad (6.15)$$

for any 1-bounded $f_1, \dots, f_k \in L^\infty(\mathbb{K}^k)$ supported on $\prod_{i \in [k]} [\pm C''_0 N^{d_i}]_{\mathbb{K}}$.

Consider C' from Theorem 6.4 applied with $C'_0 = C''_0$, and choose $\delta_0 = \delta^\gamma$ for some $\gamma \in (0, 1)$ such that $\gamma C' < C/2$, where δ, C are as in (6.15). Then

$$\tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k) = F + E, \quad (6.16)$$

where F, E are as in Theorem 6.4 and satisfy (6.6)–(6.8) with δ_0, C' as above. If δ is small, then $\gamma C' < C/2$ implies $4C'\delta^{-\gamma C'} < \delta^{-C}$ so that by the support properties of $\mathcal{F}_{j, \mathbb{K}^k} F$, see (6.6), we have $(\mathbb{I} - \Pi_{\mathbb{K}}^j)F \equiv 0$. Using (6.16) and the latter properties of F we conclude that the proof of (6.15) can be reduced to showing the following estimate

$$\|(\mathbb{I} - \Pi_{\mathbb{K}}^j)E\|_{L^2(\mathbb{K}^k)} \lesssim \delta^{c_1} N^{D/2}. \quad (6.17)$$

Note that $\|\tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k)\|_{L^\infty(\mathbb{K}^k)} \lesssim 1$. By (6.5), (6.7), and (6.8) we have

$$\|E\|_{L^\infty(\mathbb{K}^k)} \lesssim \delta^{-C'\gamma} \quad \text{and} \quad \|E\|_{L^1(\mathbb{K}^k)} \lesssim \delta^\gamma N^D. \quad (6.18)$$

Now if $p_0 \in 2\mathbb{Z}$ is very large, so that $p'_0 \in (1, 2)$ is sufficiently close to 1, then by interpolating the bounds in (6.18), we obtain

$$\|E\|_{L^{p'_0}(\mathbb{K}^k)} \lesssim \delta^{\gamma/2} N^{D/p'_0}. \quad (6.19)$$

By our choice of δ in (6.10) and (6.11) with some fixed $\varepsilon \in (0, 1)$, we have

$$\left\| (\mathbb{I} - \Pi_{\mathbb{K}}^j) \tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k) \right\|_{L^{p_0}(\mathbb{K}^k)} \lesssim \delta^{-\varepsilon C} N^{D/p_0},$$

since $\tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k)$ vanishes outside $\prod_{i \in [k]} [\pm O(N_i)]_{\mathbb{K}}$. Thus, we have

$$\|(\mathbb{I} - \Pi_{\mathbb{K}}^j)E\|_{L^{p_0}(\mathbb{K}^k)} \lesssim \delta^{-\varepsilon C} N^{D/p_0}. \quad (6.20)$$

Using (6.11) again, this time together with (6.19), we obtain

$$\|(\mathbb{I} - \Pi_{\mathbb{K}}^j)E\|_{L^{p'_0}(\mathbb{K}^k)} \lesssim \delta^{-\varepsilon C} \delta^{\gamma/2} N^{D/p'_0}. \quad (6.21)$$

Interpolating (6.20) with (6.21) and adjusting $\varepsilon \in (0, 1)$, we establish (6.17) with $c_1 = \frac{\gamma}{5}$. Consequently, (6.15) follows.

Step 2. We now relax the $L^\infty(\mathbb{K}^k)$ control in (6.15) to the control on $L^p(\mathbb{K}^k)$ spaces. More precisely, assuming (6.10), for fixed $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{2}$ we show, with a small constant $c_2 \in (0, 1)$, the estimate

$$\left\| (\mathbb{I} - \mathbf{\Pi}_{\mathbb{K}}^j) \tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim \delta^{c_2} \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^k)} \quad (6.22)$$

for any $f_i \in L^{p_i}(\mathbb{K}^k)$ supported on $\prod_{i \in [k]} [\pm C_0'' N^{d_i}]_{\mathbb{K}}$ for all $i \in [k]$.

The main tool in proving (6.22) is the multilinear $L^p(\mathbb{K}^k)$ -improving inequality proved in Section 5, which establishes the estimate

$$\left\| \tilde{A}_{N; \mathbb{K}^k}^{*j}(g_1, \dots, g_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim N^{-D(\frac{1}{r} - \frac{1}{2})} \prod_{i \in [k]} \|g_i\|_{L^{r_i}(\mathbb{K}^k)}$$

and therefore

$$\left\| (\mathbb{I} - \mathbf{\Pi}_{\mathbb{K}}^j) \tilde{A}_{N; \mathbb{K}^k}^{*j}(g_1, \dots, g_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim N^{-D(\frac{1}{r} - \frac{1}{2})} \prod_{i \in [k]} \|g_i\|_{L^{r_i}(\mathbb{K}^k)} \quad (6.23)$$

holds for some $1 < r < 2$ and $1 < r_1, \dots, r_k < \infty$ such that $\frac{1}{r_1} + \dots + \frac{1}{r_k} = \frac{1}{r}$; see Corollary 5.3. Here, it is crucial that $1 < r < 2$. The idea of utilizing $L^p(\mathbb{K}^k)$ -improving inequalities in this context was initiated in [62]. Although the bilinear case $k = 2$ of (6.23), used in [62], can be deduced from the linear $L^p(\mathbb{K}^k)$ -improving inequality [47], the general case $k \geq 3$ requires a genuine multilinear $L^p(\mathbb{K}^k)$ -improving inequality established in Section 5.

We interpolate the bounds in (6.15) and (6.23) taking $\theta := \frac{r}{2} < 1$ and q_i such that $\frac{1}{q_i} = \frac{\theta}{r_i} + \frac{1-\theta}{\infty}$ for each $i \in [k]$ so that $\frac{1}{q_1} + \dots + \frac{1}{q_k} = \frac{\theta}{r} = \frac{1}{2}$. Since $-D(\frac{1}{r} - \frac{1}{2})\theta + \frac{D}{2}(1-\theta) = 0$, multilinear interpolation yields

$$\left\| (\mathbb{I} - \mathbf{\Pi}_{\mathbb{K}}^j) \tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim \delta^{c_1(1-\theta)} \prod_{i \in [k]} \|f_i\|_{L^{q_i}(\mathbb{K}^k)}.$$

By Plancherel's theorem and Hölder's inequality, for all $1 \leq p_1, \dots, p_k \leq \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{2}$ we also have

$$\left\| (\mathbb{I} - \mathbf{\Pi}_{\mathbb{K}}^j) \tilde{A}_{N; \mathbb{K}^k}^{*j}(f_1, \dots, f_k) \right\|_{L^2(\mathbb{K}^k)} \lesssim \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^k)}.$$

Interpolating the last two estimates yields (6.22), as desired. To remove the support condition in inequality (6.22) we use the off-diagonal decay estimate (6.12) and proceed much the same way as in [62, Corollary 6.24]. $\square$

Finally, we establish the multilinear Weyl inequality and the Sobolev smoothing inequality stated at the beginning of this section. The key tool will be Theorem 6.13, which will give us the desired decay in δ and N .

Proof of Theorem 6.1. We can assume that δ^{-1} and N are large; otherwise (6.2) holds by Hölder's inequality. We consider separately $p > 1$ and $p = 1$.

Step 1. Let $p > 1$. Let $\mathcal{F}_{j,\mathbb{K}^k} f_j$ vanish on the major arcs $\mathfrak{M}_{\leq N^{-d_j} \delta^{-C_2}}^j(\mathcal{R}_{\leq \delta^{-C_1}}^{\mathbb{K}})$ for some $C_1, C_2 \in \mathbb{R}_+$. Since $N^{-d_j} \delta^{-C} \leq N^{-d_j} \delta^{-C_2}$ and $\delta^{-C} \leq \delta^{-C_1}$ for $C := \min\{C_1, C_2\}$, we see that $\mathcal{F}_{j,\mathbb{K}^k} f_j$ vanishes on $\mathfrak{M}_{\leq N^{-d_j} \delta^{-C}}^j(\mathcal{R}_{\leq \delta^{-C}}^{\mathbb{K}})$.

Case 1. Consider $N \geq C_{p_0} \delta^{-C_{p_0}}$ as in (6.10). By duality it suffices to prove

$$|\langle \tilde{A}_{N;\mathbb{K}^k}(f_1, \dots, f_k), h \rangle| \lesssim \delta^c \|h\|_{L^{p'}(\mathbb{K}^k)} \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^k)}$$

for $h \in L^{p'}(\mathbb{K}^k)$ and $f_i \in L^{p_i}(\mathbb{K}^k)$ for all $i \in [k]$, with $f_j \in L^{p_j}(\mathbb{K}^k) \cap L^2(\mathbb{K}^k)$ such that $\mathcal{F}_{j,\mathbb{K}^k} f_j$ vanishes on $\mathfrak{M}_{\leq N^{-d_j} \delta^{-C}}^j(\mathcal{R}_{\leq \delta^{-C}}^{\mathbb{K}})$. As before, we abbreviate $\Pi_{\mathbb{K}}^j[\leq \delta^{-C}, \leq N^{-d_j} \delta^{-C}]$ to $\Pi_{\mathbb{K}}^j$. By (4.6) and properties of $\mathcal{F}_{j,\mathbb{K}^k} f_j$ we have

$$\left| \left\langle \tilde{A}_{N;\mathbb{K}^k}(f_1, \dots, f_k), h \right\rangle \right| = \left| \left\langle (\mathbb{I} - \Pi_{\mathbb{K}}^j) \tilde{A}_{N;\mathbb{K}^k}^{*j}(f_1, \dots, f_{j-1}, h, f_{j+1}, \dots, f_k), f_j \right\rangle \right|$$

and the desired bound holds with $c = c_0$ by Hölder's inequality and (6.14).

Case 2. Consider $N \leq C_{p_0} \delta^{-C_{p_0}}$. The proof follows from the previous case, with δ_0 such that $\delta_0^{-1} = (N/C_{p_0})^{1/C_{p_0}} \leq \delta^{-1}$, giving Theorem 6.1 for $p > 1$.

Step 2. Let $p = 1$. In this case linear averages (with $k = 1$) do not arise so we have $k \geq 2$. We assume that $\mathcal{F}_{j,\mathbb{K}^k} f_j$ vanishes on $\mathfrak{M}_{\leq N^{-d_j} \delta^{-C}}^j(\mathcal{R}_{\leq \delta^{-C}}^{\mathbb{K}})$ for some suitable $C \in \mathbb{R}_+$, as in Step 1. The key is to show that

$$\|\tilde{A}_{N;\mathbb{K}^k}(f_1, \dots, f_k)\|_{L^q(\mathbb{K}^k)} \lesssim \prod_{i \in [k]} \|f_i\|_{L^{q_i}(\mathbb{K}^k)} \quad (6.24)$$

for some choice of $1 < q_1, \dots, q_k < \infty$ with $q_j = 2$ and $\frac{1}{q_1} + \dots + \frac{1}{q_k} = \frac{1}{q}$ for $q < 1$ close to 1. Of course, if $q \geq 1$, then (6.24) holds trivially by Hölder's inequality but we need this for $L^q(\mathbb{K}^k)$ with $q < 1$, which is not a Banach space. Interestingly, inequality (6.24) follows by an application of the L^p -improving bound (5.4) from Corollary 5.3. Using the multilinear interpolation result for $q < 1$ in [43, Theorem 1], we can interpolate the bound (6.24) with the bounds (6.2) in Theorem 6.1 for $p > 1$ established in Step 1 to conclude

$$\|\tilde{A}_{N;\mathbb{K}^k}(f_1, \dots, f_k)\|_{L^1(\mathbb{K}^k)} \lesssim (\delta^c + N^{-c}) \prod_{i \in [k]} \|f_i\|_{L^{p_i}(\mathbb{K}^k)}$$

for some $c > 0$ and for all $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = 1$, as desired. This completes the proof of Theorem 6.1. $\square$

7. MULTILINEAR CIRCLE METHOD: PROOF OF POINTWISE ERGODIC THEOREM

In this section, we develop a multilinear circle method in the context of quantitative multilinear pointwise ergodic theorems. Although our method is robust enough to handle oscillation, or even full r -variational estimates, we only illustrate its strength in the context of the long r -variational multilinear pointwise ergodic bound (1.22). This will be sufficient to deduce the Furstenberg–Bergelson–Leibman conjecture for commuting transformations along polynomials with distinct degrees; see Theorem 1.21 as well as Theorem 1.12.

The bilinear circle method was developed for the first time in [62]. Here we shall present a novel different approach relying on the following tools:

1. We use the Ionescu–Wainger projections from (4.13) to make a distinction between minor and major arcs. These projections now correspond to the set of canonical fractions (4.12), which is a significant new input.
2. An essential ingredient in obtaining the minor arc estimates is the multilinear Weyl inequality (6.2) with a polynomial decay of the form $\delta^c + N^{-c}$; see Theorem 6.1. This result represents a significant improvement over the logarithmic decay $\delta^c + \langle \log N \rangle^{-c\tilde{C}}$ for the bilinear Furstenberg–Weiss averages from [62]. The key tools we use to derive (6.2) with a polynomial decay, which were not available in [62], are our Ionescu–Wainger multiplier theorem for the set of canonical fractions and multilinear L^p -improving inequality; see Theorem 3.3 and Theorem 5.2, respectively.
3. Another important ingredient is the Sobolev smoothing inequality, which is a continuous variant of the multilinear Weyl inequality; see Theorem 6.1. The multilinear Sobolev smoothing inequality is utilized to obtain the major arc estimates in the high frequency case. In the bilinear situation the major arc estimates involving high frequency cases can be handled by appealing to the bilinear Weyl inequality (6.2) and using a simple integration by parts argument. However, the general multilinear case necessitates the full strength of the Sobolev smoothing inequality as the integration by parts trick from [62] is limited to the bilinear averages. This is a new phenomenon which was not apparent previously.
4. The most challenging part is obtaining estimates for major arcs in the low frequency regime. A novel tool that we propose for controlling maximal functions is a powerful multiparameter norm interchanging inequality (7.64). It is important to note that polynomial decay in the multilinear Weyl inequality (6.2) is crucial for dealing with this case; see Remark 7.94 below. A one-parameter norm interchanging inequality originates in [62] as a consequence of the Rademacher–Menshov inequality (2.11). Our multiparameter norm interchanging inequality allows us to replace the p -adic methods used in [62] and significantly simplifies the argument. Moreover, we emphasize that the p -adic approach (especially sharp L^p -improving inequalities in the p -adic setting), even if adjusted to our context, would not be sufficient to handle the major arc estimates in the low frequency regime.

Given $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$, we will prove

$$\|\mathbf{V}^r(\tilde{A}_N(f_1, \dots, f_k) : N \in \mathbb{D})\|_{\ell^p(\mathbb{Z}^k)} \lesssim 1 \quad (7.1)$$

for all finitely supported $f_1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{p_k}(\mathbb{Z}^k)$, which are normalized so that $\|f_1\|_{\ell^{p_1}(\mathbb{Z}^k)} = \dots = \|f_k\|_{\ell^{p_k}(\mathbb{Z}^k)} = 1$. Then, by scaling and density arguments combined with the monotone convergence theorem, we readily obtain inequality (1.22) from Theorem 1.21 for all $f_1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{p_k}(\mathbb{Z}^k)$. The proof of (7.1) is long and intricate. We begin with making a distinction between minor and major arcs and for this purpose we will use the Ionescu–Wainger projections from (4.13) with dyadic parameters.

7.1. Dyadic Ionescu–Wainger projections. Let $\eta: \mathbb{R} \rightarrow [0, 1]$ be a smooth and even function satisfying (2.4). For any $n, \xi \in \mathbb{R}$, we set

$$\eta_{\leq n}(\xi) := \eta_{[\leq 2^n]}(\xi) = \eta(2^{-n}\xi).$$

For any $l \in \mathbb{N}$, using definition of canonical fractions (4.12), we set

$$\Sigma_{\leq l} := \mathcal{R}_{\leq 2^l} \quad \text{and} \quad \Sigma_l := \Sigma_{\leq l} \setminus \Sigma_{\leq l-1}$$

with $\Sigma_{\leq -1}$ denoting the empty set. Then we have

$$\#\Sigma_{\leq l} \leq 2^{2l}. \quad (7.2)$$

Similarly, for any $i \in [k]$, $l \in \mathbb{N}$, and $m \in \mathbb{Z}$, using definitions (4.10) and (4.11), we introduce the dyadic “major arc” by

$$\mathcal{M}_{\leq l, \leq m} := \mathfrak{M}_{\leq 2^m}(\Sigma_{\leq l}) \quad \text{and} \quad \mathcal{M}_{\leq l, \leq m}^i := \mathfrak{M}_{\leq 2^m}^i(\Sigma_{\leq l}).$$

We note that $\mathcal{M}_{\leq l, \leq m} \subseteq \mathcal{M}_{\leq l', \leq m'}$ whenever both $l \leq l'$ and $m \leq m'$ hold. Moreover, if $m \leq -2l - 2$, then the arcs $[\theta - 2^m, \theta + 2^m]$ that comprise $\mathcal{M}_{\leq l, \leq m}$ are pairwise disjoint. We also define

$$\mathcal{M}_{l, \leq m} := \mathcal{M}_{\leq l, \leq m} \setminus \mathcal{M}_{\leq l-1, \leq m} \quad \text{and} \quad \mathcal{M}_{l, m} := \mathcal{M}_{l, \leq m} \setminus \mathcal{M}_{l, \leq m-1}.$$

Referring to (4.13), we define the dyadic Ionescu–Wainger projections $\Pi_{\leq l, \leq m}^i: \ell^2(\mathbb{Z}^k) \rightarrow \ell^2(\mathbb{Z}^k)$ by setting

$$\Pi_{\leq l, \leq m}^i f(x) := \Pi_{\mathbb{Z}}^i[\leq 2^l, \leq 2^m]f(x) = T_{i, \mathbb{Z}^k}^{\Sigma_{\leq l}}[\eta_{\leq m}]f(x), \quad x \in \mathbb{Z}^k, \quad (7.3)$$

for each $f \in \ell^2(\mathbb{Z}^k)$. In particular, $\Pi_{\leq l, \leq m}^i$ is a self-adjoint (and real symmetric) operator, which is the identity operator except in the i -th variable.

Remark 7.4. The following properties are clear from (7.3).

- (i) The function $\mathcal{F}_{i, \mathbb{Z}^k}(\Pi_{\leq l, \leq m}^i f)$ is supported on the set $\mathcal{M}_{\leq l, \leq m}^i$, and if $\mathcal{F}_{i, \mathbb{Z}^k} f$ vanishes on that set, then $\Pi_{\leq l, \leq m}^i f \equiv 0$.
- (ii) If $m \leq -2l - 2$ and $\mathcal{F}_{i, \mathbb{Z}^k} f$ is supported on $\mathcal{M}_{\leq l, \leq m-2}^i$, then $\Pi_{\leq l, \leq m}^i f = f$.
- (iii) If $m \leq -2l - 2$, then $\Pi_{\leq l, \leq m}^i$ is a contraction on $\ell^2(\mathbb{Z}^k)$.

A consequence of Theorem 3.3 is the following important bound.

Proposition 7.5. *For every $k \in \mathbb{Z}_+$ and $p \in (1, \infty)$ there exists a constant $C_p \in \mathbb{R}_+$ such that for every $l \in \mathbb{N}$, $m \in \mathbb{Z}$, and $i \in [k]$, if*

$$m \leq -6 \max\{p, p'\}(l+1), \quad (7.6)$$

then

$$\|\Pi_{\leq l, \leq m}^i f\|_{\ell^p(\mathbb{Z}^k)} \leq C_p \left(2^{C_p \frac{l \log \log l}{\log l}} \mathbf{1}_{l \geq 10} + \mathbf{1}_{l < 10} \right) \|f\|_{\ell^p(\mathbb{Z}^k)}. \quad (7.7)$$

Proof. The proof of (7.7) is a direct consequence of Theorem 3.3. $\square$

7.2. Minor arc estimates. As in Section 4, we fix a polynomial mapping

$$\mathcal{P} := (P_1, \dots, P_k): \mathbb{R} \rightarrow \mathbb{R}^k,$$

with $P_1, \dots, P_k \in \mathbb{Z}[n]$ such that $d_1 := \deg P_1 < \dots < d_k := \deg P_k$ and $D := d_1 + \dots + d_k$; see (4.1)–(4.3). We also fix $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$ and choose $p_0 \in 2\mathbb{Z}_+$ satisfying

$$1 < p'_0 < p_1, \dots, p_k < p_0 < \infty. \quad (7.8)$$

The parameter p_0 from (7.8) will serve as an upper bound for $\max\{p, p'\}$ from condition (7.6) to ensure that Proposition 7.5 can be applied. One can think that the parameter p_0 from (7.8) is a sufficiently large even integer, which, if necessary, may be further adjusted depending on the interpolation arguments that will be used throughout this section.

Given a small fixed parameter

$$0 < \alpha < (10^6 k d_k p_0)^{-1}$$

and a large constant $C_0 \in \mathbb{Z}_+$ depending on $p_0, \mathcal{P}, \alpha$, we define the quantities

$$l_{(N)} := \text{Log } N^\alpha \quad \text{and} \quad L_{(N)} := \text{Log } N - l_{(N)}$$

for any $N \geq C_0$, using the log-scale notation (2.1). We take $C_0 \geq 2^{5/\alpha}$ so that $l_{(N)} \in \mathbb{N}_{\geq 5}$ for $N \geq C_0$. For each $i \in [k]$ and $\varepsilon_i \in \{0, 1\}$, define

$$\Pi_{\leq l_{(N)}, \leq -d_i L_{(N)}}^{i, \varepsilon_i} f_i := \begin{cases} \Pi_{\leq l_{(N)}, \leq -d_i L_{(N)}}^i f_i & \text{if } \varepsilon_i = 1, \\ f_i - \Pi_{\leq l_{(N)}, \leq -d_i L_{(N)}}^i f_i & \text{if } \varepsilon_i = 0. \end{cases}$$

Observe that there exists a small absolute constant $c \in (0, 1)$ such that

$$\|\tilde{A}_N(\Pi_{\leq l_{(N)}, \leq -d_1 L_{(N)}}^{1, \varepsilon_1} f_1, \dots, \Pi_{\leq l_{(N)}, \leq -d_k L_{(N)}}^{k, \varepsilon_k} f_k)\|_{\ell^p(\mathbb{Z}^k)} \lesssim N^{-c} \quad (7.9)$$

whenever $(\varepsilon_1, \dots, \varepsilon_k) \neq (1, \dots, 1)$. Indeed, assuming $\varepsilon_j = 0$ for some $j \in [k]$ we apply Theorem 6.1 with $C_1 = 1$, $C_2 = d_j$, and $\delta = 2^{5-l_{(N)}} \in (0, 1]$, and then (7.9) follows from Proposition 7.5. Writing $\tilde{A}_N(f_1, \dots, f_k)$ as

$$\sum_{(\varepsilon_1, \dots, \varepsilon_k) \in \{0, 1\}^k} \tilde{A}_N \left(\Pi_{\leq l_{(N)}, \leq -d_1 L_{(N)}}^{1, \varepsilon_1} f_1, \dots, \Pi_{\leq l_{(N)}, \leq -d_k L_{(N)}}^{k, \varepsilon_k} f_k \right)$$

we see that, in view of (7.9), the proof of (7.1) is reduced to estimating

$$\left\| \mathbf{V}^r \left(\tilde{A}_N \left(\Pi_{\leq l_{(N)}, \leq -d_1 L_{(N)}}^1 f_1, \dots, \Pi_{\leq l_{(N)}, \leq -d_k L_{(N)}}^k f_k \right) : N \in \mathbb{D} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \quad (7.10)$$

for all lacunary sequences $\mathbb{D} \subset [C_0, \infty)$ and normalized functions $f_1, \dots, f_k$, where the functions $\Pi_{\leq l_{(N)}, \leq -d_1 L_{(N)}}^1 f_1, \dots, \Pi_{\leq l_{(N)}, \leq -d_k L_{(N)}}^k f_k$ are all restricted to major arcs.

7.3. Major arc estimates. Our aim is to utilize certain model operators, which will be more manageable. We begin with trimming the size of denominators. To this end, we rewrite the average from (7.10) as

$$\sum_{l_1 \in \mathbb{N}_{\leq l_{(N)}}} \cdots \sum_{l_k \in \mathbb{N}_{\leq l_{(N)}}} \tilde{A}_N \left(\Pi_{l_1, \leq -d_1 L_{(N)}}^1 f_1, \dots, \Pi_{l_k, \leq -d_k L_{(N)}}^k f_k \right), \quad (7.11)$$

where $\Pi_{l_i, \leq -d_i L_{(N)}}^i := T_{i, \mathbb{Z}^k}^{\Sigma l_i} [\eta_{\leq -d_i L_{(N)}}]$ for every $i \in [k]$. Next, we match the scales of the bumps $\eta_{\leq -d_i \text{Log } N + d_i l_{(N)}}$ appearing in (7.11) by splitting them into nonoscillatory and highly oscillatory pieces

$$\eta_{\leq -d_i \text{Log } N} + \sum_{s_i \in [l_{(N)}]} (\eta_{\leq -d_i (\text{Log } N - s_i)} - \eta_{\leq -d_i (\text{Log } N - s_i + 1)}).$$

For $i \in [k]$ we define

$$\eta_N^{i,s_i} := \begin{cases} \eta_{\leq -d_i(\log N - s_i)} - \eta_{\leq -d_i(\log N - s_i + 1)} & \text{if } s_i > 0, \\ \eta_{\leq -d_i \log N} & \text{if } s_i = 0, \end{cases} \quad (7.12)$$

and set

$$\tilde{\Pi}_{l_i, s_i}^{i, N} := T_{i, \mathbb{Z}^k}^{\Sigma_{l_i}} [\eta_N^{i, s_i}]. \quad (7.13)$$

Using (7.12), (7.13), and the previous identities, we write (7.11) as

$$\sum_{l_1, s_1 \in \mathbb{N}_{\leq l(N)}} \cdots \sum_{l_k, s_k \in \mathbb{N}_{\leq l(N)}} \tilde{A}_N \left(\tilde{\Pi}_{l_1, s_1}^{1, N} f_1, \dots, \tilde{\Pi}_{l_k, s_k}^{k, N} f_k \right).$$

Set $\mathbb{D}_{l,s} := \{N \in \mathbb{D} : N \geq C_0 \text{ and } l(N) \geq \max\{l, s\}\}$, where

$$l := \max\{l_i : i \in [k]\} \quad \text{and} \quad s := \max\{s_i : i \in [k]\}. \quad (7.14)$$

It suffices to show, for any $s_1, l_1, \dots, s_k, l_k \in \mathbb{N}$, that

$$\left\| \mathbf{V}^r \left(\tilde{A}_N \left(\tilde{\Pi}_{l_1, s_1}^{1, N} f_1, \dots, \tilde{\Pi}_{l_k, s_k}^{k, N} f_k \right) : N \in \mathbb{D}_{l,s} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-c(l+s)}, \quad (7.15)$$

since summing (7.15) over $s_1, l_1, \dots, s_k, l_k$ yields the desired bound for (7.10).

7.3.1. Major arc approximations. We now take advantage of the fact that each function in (7.15) is restricted to major arcs, and approximate the average in (7.15) by a certain model operator. We first fix some notation.

For $N \geq 1$ we define the exponential sum

$$m_N(\xi) := \frac{1}{|(N/2, N] \cap \mathbb{Z}|} \sum_{n \in (N/2, N] \cap \mathbb{Z}} e(\xi \cdot \mathcal{P}(n)), \quad \xi \in \mathbb{T}^k, \quad (7.16)$$

and its continuous counterpart

$$\mathbf{m}_N(\xi) := 2 \int_{1/2}^1 e(\xi \cdot \mathcal{P}(Nt)) dt, \quad \xi \in \mathbb{R}^k. \quad (7.17)$$

Furthermore, given $a \in \mathbb{Z}^k$ and $q \in \mathbb{Z}_+$ such that $\gcd(a, q) = 1$, we consider the corresponding complete exponential sum

$$G\left(\frac{a}{q}\right) := \frac{1}{q} \sum_{n \in [q]} e\left(\frac{a}{q} \cdot \mathcal{P}(n)\right). \quad (7.18)$$

We also define the set of rational fractions

$$\Sigma_{l_1, \dots, l_k} := \left\{ \frac{a}{q} \in (\mathbb{Q}/\mathbb{Z})^k : \gcd(a, q) = 1 \text{ and } \frac{a}{q} \in \Sigma_{l_1} \times \cdots \times \Sigma_{l_k} \right\}, \quad (7.19)$$

which will be a natural domain for G in our applications.

Lemma 7.20. *Let $\mathcal{P}: \mathbb{R} \rightarrow \mathbb{R}^k$ be a polynomial mapping as in (4.1) whose components are polynomials with integer coefficients and satisfy (4.2). Then there exists a constant $C_{\mathcal{P}} \in \mathbb{R}_+$ such that for every $N \geq 1$, $M_1, \dots, M_k \in \mathbb{R}_+$, and $l_1, \dots, l_k \in \mathbb{N}$ the following holds. For every $\xi \in \mathbb{T}^k$ and $\theta \in \Sigma_{l_1, \dots, l_k}$ such that $|\xi_i - \theta_i| \leq M_i^{-1}$ for all $i \in [k]$, one has*

$$|m_N(\xi) - G(\theta)\mathbf{m}_N(\xi - \theta)| \leq C_{\mathcal{P}} 2^{kl} (\max\{M_i^{-1} N^{d_i-1} : i \in [k]\} + N^{-1}) \quad (7.21)$$

with $m_N, \mathbf{m}_N, G, \Sigma_{l_1, \dots, l_k}$ defined in (7.16)–(7.19).

Proof. It suffices to prove (7.21) for $N \geq 2^{kl}$, because the opposite case $N < 2^{kl}$ is trivial. Fix $\theta = \frac{a}{q} \in \Sigma_{l_1, \dots, l_k}$ and observe that

$$m_N(\xi) = \frac{1}{q} \sum_{r \in [q]} e(\theta \cdot \mathcal{P}(r)) \cdot q \sum_{n \in \mathbb{Z}} e((\xi - \theta) \cdot \mathcal{P}(qn + r)) \chi_N(qn + r),$$

where $\chi_N(n) := |(N/2, N] \cap \mathbb{Z}|^{-1} \mathbf{1}_{(N/2, N]}(n)$. We note that

$$\left| \sum_{n \in \mathbb{Z}} e((\xi - \theta) \cdot \mathcal{P}(qn + r)) \chi_N(qn + r) - \int_{\mathbb{R}} e((\xi - \theta) \cdot \mathcal{P}(qt + r)) \chi_N(q[t] + r) dt \right|$$

is bounded by $C_{\mathcal{P}} \max\{M_i^{-1} N^{d_i} : i \in [k]\}$ by the mean value theorem. Moreover, we have $\int_{\mathbb{R}} |\chi_N(q[t] + r) - \chi_N(qt + r)| dt \leq 10N^{-1}$. Using this and the restriction $q \leq 2^{kl}$, we obtain (7.21). $\square$

Next, given a finite subset $\Theta \subset (\mathbb{Q}/\mathbb{Z})^k$ and functions $m: \mathbb{T}^k \rightarrow \mathbb{C}$ and $S: (\mathbb{Q}/\mathbb{Z})^k \rightarrow \mathbb{C}$, we will be working with multilinear operators of the form

$$B^{\Theta}[S; m](g_1, \dots, g_k)(x) := \sum_{\theta \in \Theta} S(\theta) \sum_{y \in \mathbb{Z}^k} K_{\tau_{\theta} m}(y) \prod_{i \in [k]} g_i(x - (y \cdot e_i) e_i) \quad (7.22)$$

for $x \in \mathbb{Z}^k$, where $\tau_{\theta} m(\xi) := m(\xi - \theta)$ and

$$K_m(y) := \int_{\mathbb{T}^k} m(\xi) e(-\xi \cdot y) d\xi, \quad (7.23)$$

and, as before, $\{e_i : i \in [k]\}$ is the standard basis in $\mathbb{R}^k$.

Using (7.22) and (7.23), we prove the following technical but very useful lemma, which will often be applied in tandem with Lemma 7.20.

Lemma 7.24. *Let $N \geq 1$, $d_1, \dots, d_k \in \mathbb{Z}_+$, and $M_1, \dots, M_k \in [3, \infty)$, and let $\mathbf{m}_N: \mathbb{T}^k \rightarrow \mathbb{C}$ be a smooth function supported on $\prod_{i \in [k]} [\pm M_i^{-1}]$ such that*

$$\left\| \partial_1^{\beta_1} \dots \partial_k^{\beta_k} \mathbf{m}_N \right\|_{L^{\infty}(\mathbb{T}^k)} \lesssim_{\beta_1, \dots, \beta_k} N^{\beta_1 d_1 + \dots + \beta_k d_k} \quad (7.25)$$

for all $(\beta_1, \dots, \beta_k) \in \mathbb{N}^k$. Then, for every $\gamma \in (0, 1)$ and $\beta \in \mathbb{Z}_+$, we have

$$|K_{\mathbf{m}_N}(y)| \leq C_{\beta, \gamma} M_1^{-1} \dots M_k^{-1} \|\mathbf{m}_N\|_{L^{\infty}(\mathbb{T}^k)}^{1-\gamma} \prod_{i \in [k]} \langle y_i / N^{d_i} \rangle^{-\gamma \beta} \quad (7.26)$$

with some constant $C_{\beta, \gamma} \in \mathbb{R}_+$. In particular, if $\gamma \beta > 1$, then (7.26) implies

$$\|K_{\mathbf{m}_N}\|_{\ell^1(\mathbb{Z}^k)} \lesssim_{\beta, \gamma} (N^{d_1} M_1^{-1}) \dots (N^{d_k} M_k^{-1}) \|\mathbf{m}_N\|_{L^{\infty}(\mathbb{T}^k)}^{1-\gamma}. \quad (7.27)$$

Proof. Using (7.25), the support condition for $\mathbf{m}_N$, and integration by parts (sufficiently many times, say $\beta \in \mathbb{N}$, variable by variable) in the integral in (7.23) with $m = \mathbf{m}_N$, we obtain (7.26) with $\gamma = 1$. A convex combination of this bound and the trivial bound $|K_{\mathbf{m}_N}(y)| \lesssim_\beta M_1^{-1} \cdots M_k^{-1} \|\mathbf{m}_N\|_{L^\infty(\mathbb{T}^k)}$ yields (7.26) for $\gamma \in (0, 1)$, and (7.27) follows whenever $\gamma\beta > 1$. $\square$

Now we can formulate our first approximation result.

Proposition 7.28. *For every $N \geq 1$ and $l_1, s_1, \dots, l_k, s_k \in \mathbb{N}$ such that $l_{(N)} \geq \max\{l, s\}$ and for all $f_1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{p_k}(\mathbb{Z}^k)$ with $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$ the following estimate holds*

$$\|E_{l,s}^N(f_1, \dots, f_k)\|_{\ell^p(\mathbb{Z}^k)} \lesssim N^{-\frac{9}{10}} \prod_{i \in [k]} \|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)} \quad (7.29)$$

with $E_{l,s}^N(f_1, \dots, f_k)$ denoting the error term

$$\tilde{A}_N \left(\tilde{\Pi}_{l_1, s_1}^{1, N} f_1, \dots, \tilde{\Pi}_{l_k, s_k}^{k, N} f_k \right) - B^{\Sigma_{l_1, \dots, l_k}}[G; \mathbf{m}_N w_N^{s_1, \dots, s_k}](f_1, \dots, f_k),$$

where $w_N^{s_1, \dots, s_k} := \bigotimes_{i \in [k]} \eta_N^{i, s_i}$ and η_N^{i, s_i} is defined in (7.12).

Proof. By homogeneity, we can assume that $\|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)} = 1$ for all $i \in [k]$.

Step 1. First, we rewrite $\tilde{A}_N(\tilde{\Pi}_{l_1, s_1}^{1, N} f_1, \dots, \tilde{\Pi}_{l_k, s_k}^{k, N} f_k)(x)$ as

$$\sum_{\theta \in \Sigma_{l_1, \dots, l_k}} \sum_{y \in \mathbb{Z}^k} K_{m_N \tau_\theta w_N^{s_1, \dots, s_k}}(y) \prod_{i \in [k]} f_i(x - (y \cdot e_i) e_i).$$

By the triangle inequality and Minkowski's integral inequality followed by Hölder's inequality, the left-hand side of (7.29) is controlled by

$$\sum_{\theta \in \Sigma_{l_1, \dots, l_k}} \left\| K_{m_N \tau_\theta w_N^{s_1, \dots, s_k}} - G(\theta) K_{\tau_\theta(\mathbf{m}_N w_N^{s_1, \dots, s_k})} \right\|_{\ell^1(\mathbb{Z}^k)}.$$

Step 2. Fix $\theta = \frac{a}{q} \in \Sigma_{l_1, \dots, l_k}$ and expand the above norm as follows

$$\sum_{y \in \mathbb{Z}^k} \left| \int_{\mathbb{T}^k} (m_N(\xi) - G(\theta) \mathbf{m}_N(\xi - \theta)) w_N^{s_1, \dots, s_k}(\xi - \theta) e(-\xi \cdot y) d\xi \right|.$$

Using Lemma 7.20 with $M_i \simeq N^{d_i} 2^{-d_i s_i}$, together with the support condition for $w_N^{s_1, \dots, s_k}$ and the restriction $l_{(N)} \geq \max\{l, s\}$, we obtain

$$\|m_N \tau_\theta w_N^{s_1, \dots, s_k} - G(\theta) \tau_\theta(\mathbf{m}_N w_N^{s_1, \dots, s_k})\|_{L^\infty(\mathbb{T}^k)} \lesssim 2^{2kd_k l_{(N)}} N^{-1}. \quad (7.30)$$

Step 3. Now by (7.30) and Lemma 7.24, for any fixed $\gamma \in (0, 1)$, we have

$$\left\| K_{m_N \tau_\theta w_N^{s_1, \dots, s_k}} - G(\theta) K_{\tau_\theta(\mathbf{m}_N w_N^{s_1, \dots, s_k})} \right\|_{\ell^1(\mathbb{Z}^k)} \lesssim 2^{3kd_k l_{(N)}} N^{-(1-\gamma)}.$$

Using (7.2) we obtain that $\#\Sigma_{l_1, \dots, l_k} \lesssim 2^{2kl}$, which, combined with the last estimate, yields (7.29) upon choosing the parameter γ sufficiently small. This completes the proof of Proposition 7.28. $\square$

Now, in light of Proposition 7.28, instead of (7.15) it suffices to prove

$$\left\| \mathbf{V}^r \left(B^{\Sigma_{l_1, \dots, l_k}} [G; \mathbf{m}_N w_N^{s_1, \dots, s_k}] (f_1, \dots, f_k) : N \in \mathbb{D}_{l, s} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-c(l+s)} \quad (7.31)$$

for all normalized functions $f_i \in \ell^{p_i}(\mathbb{Z}^k)$ and all $s_1, l_1, \dots, s_k, l_k \in \mathbb{N}$.

7.3.2. Model operators. One difficulty in proving (7.31) is the need to obtain the decay in both parameters l and s from (7.14). For this purpose, we distinguish two cases, which will be handled separately later on:

- (i) the *high frequency* case $s > Cl$;
- (ii) the *low frequency* case $Cl \geq s$;

with some large $C \geq 1$ to be specified later. For each fixed $t \in [\frac{1}{2}, 1]$, using the bumps η_N^{i, s_i} from (7.12), we define

$$\eta_{N, t}^{i, s_i}(\xi_i) := e(\xi_i P_i(Nt)) \eta_N^{i, s_i}(\xi_i), \quad i \in [k]. \quad (7.32)$$

Using (7.32), we obtain the following identity

$$\mathbf{m}_N(\xi) w_N^{s_1, \dots, s_k}(\xi) = 2 \int_{1/2}^1 \left(\bigotimes_{i \in [k]} \eta_{N, t}^{i, s_i} \right) (\xi) dt.$$

Define

$$u := \begin{cases} 100k(s+1) & \text{if } s > Cl, \\ 100k(l+1) & \text{if } Cl \geq s. \end{cases} \quad (7.33)$$

Note that for any $N \in \mathbb{D}_{l, s}$ we have $N \geq \max\{2^{\max\{l, s\}/\alpha}, C_0\}$, which implies

$$N \geq 2^{10^5 k d_k p_0 \max\{l, s\}} C_0^{1/2} \geq 2^{100 p_0 d_k u}, \quad (7.34)$$

since $C_0 \geq 2^{5/\alpha} \geq 2^{10^6 k d_k p_0}$. By (7.34), we rewrite $B^{\Sigma_{l_1, \dots, l_k}} [G; \mathbf{m}_N w_N^{s_1, \dots, s_k}] (f_1, \dots, f_k)(x)$ as

$$2 \int_{1/2}^1 B^{\Sigma_{l_1, \dots, l_k}} [G; \eta_u^*] \left(T_{1, \mathbb{Z}^k}^{\Sigma_{l_1}} [\eta_{N, t}^{1, s_1}] f_1, \dots, T_{k, \mathbb{Z}^k}^{\Sigma_{l_k}} [\eta_{N, t}^{k, s_k}] f_k \right) (x) dt, \quad (7.35)$$

where $\eta_u^* := \bigotimes_{i \in [k]} \eta_{\leq -p_0 d_i u}$. Introducing new functions

$$F_{N, t}^{i, l_i, s_i} := F_{N, t}^{i, l_i, s_i}(f_i) := T_{i, \mathbb{Z}^k}^{\Sigma_{l_i}} [\eta_{N, t}^{i, s_i}] f_i, \quad i \in [k], \quad (7.36)$$

we will show that the multilinear operator

$$B^{\Sigma_{l_1, \dots, l_k}} [G; \eta_u^*] (F_{N, t}^{1, l_1, s_1}, \dots, F_{N, t}^{k, l_k, s_k})$$

from (7.35) can be replaced by the following model multilinear operator

$$\tilde{A}_{2^u} \left(\Pi_{l_1, \leq -p_0 d_1 u}^1 (F_{N, t}^{1, l_1, s_1}), \dots, \Pi_{l_k, \leq -p_0 d_k u}^k (F_{N, t}^{k, l_k, s_k}) \right),$$

with a satisfactory error term.

Proposition 7.37. Fix $l_1, s_1, \dots, l_k, s_k \in \mathbb{N}$ and $u \in \mathbb{Z}_+$ as in (7.33). Then for every $N \geq 1$ and for all $f_N^1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_N^k \in \ell^{p_k}(\mathbb{Z}^k)$ with $1 < p_1, \dots, p_k < \infty$ satisfying $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$ the following estimate holds

$$\|\mathbf{V}^r(\tilde{E}_{l,s}^u(f_N^1, \dots, f_N^k) : N \in \mathbb{D}_{l,s})\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-\frac{9u}{10}} \prod_{i \in [k]} \|\mathbf{V}^r(f_N^i : N \in \mathbb{D}_{l,s})\|_{\ell^{p_i}(\mathbb{Z}^k)}$$

with $\tilde{E}_{l,s}^u(f_N^1, \dots, f_N^k)$ denoting the error term

$$\tilde{A}_{2^u}(\Pi_{l_1, \leq -p_0 d_1 u}^1(f_N^1), \dots, \Pi_{l_k, \leq -p_0 d_k u}^k(f_N^k)) - B^{\Sigma_{l_1, \dots, l_k}}[G; \eta_u^*](f_N^1, \dots, f_N^k).$$

Proof. The proof is much the same as the proof of Proposition 7.28. $\square$

7.3.3. Basic r -variational estimates. We now gather certain simple r -variational estimates that will be useful for our arguments. The key tools will be the Ionescu–Wainger and r -variational Ionescu–Wainger theorems; see Theorem 3.3 and Theorem 3.26, respectively.

The following result is a simple exercise that uses the properties of the Fourier transform.

Lemma 7.38. Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be a compactly supported function with continuous derivative. Then there exists a constant $C_\phi \in \mathbb{R}_+$ depending on ϕ such that for all $n \in \mathbb{N}$ we have

$$\|\partial^n(\mathcal{F}_{\mathbb{R}}^{-1}\phi)\|_{L^1(\mathbb{R})} \leq C_\phi^{n+1}.$$

Lemma 7.39. Let $1 < p_1, \dots, p_k < \infty$ and $r \in (2, \infty]$. Then we have

$$\begin{aligned} \|\mathbf{V}^r(\Pi_{l_i, \leq -p_0 d_i u}^i(F_{N,t}^{i, l_i, s_i}(f_i) : N \in \mathbb{D}_{l,s})\|_{\ell^{p_i}(\mathbb{Z}^k)} &\lesssim_{p_i, r} C_{p_i}(l, s_i), \\ \|\mathbf{V}^r(F_{N,t}^{i, l_i, s_i}(f_i) : N \in \mathbb{D}_{l,s})\|_{\ell^{p_i}(\mathbb{Z}^k)} &\lesssim_{p_i, r} C_{p_i}(l, s_i), \end{aligned} \quad (7.40)$$

for all $i \in [k]$ and all choices of $l_1, s_1, \dots, l_k, s_k \in \mathbb{N}$, uniformly in $t \in [\frac{1}{2}, 1]$, with normalized $f_i \in \ell^{p_i}(\mathbb{Z}^k)$ and $F_{N,t}^{i, l_i, s_i}(f_i)$ as in (7.36), where

$$C_{p_i}(l, s_i) := (s_i + 1)(2^{O_{p_i}(\frac{l \log \log l}{\log l})} \mathbf{1}_{l \geq 10} + \mathbf{1}_{l < 10}).$$

Proof. By the definition of u in (7.33), one has $6(l_i + 1) \max\{p_i, p'_i\} \leq p_0 d_i u$ for each $i \in [k]$ so that (7.6) is satisfied. Consequently, (7.7) implies

$$\|\Pi_{l_i, \leq -p_0 d_i u}^i f\|_{\ell^{p_i}(\mathbb{Z}^k)} \lesssim 2^{O_{p_i}(\frac{l \log \log l}{\log l})} \mathbf{1}_{l \geq 10} + \mathbf{1}_{l < 10}. \quad (7.41)$$

Therefore, by (7.41) it suffices to show the second inequality in (7.40), since

$$\Pi_{l_i, \leq -p_0 d_i u}^i(F_{N,t}^{i, l_i, s_i}(f_i)) = F_{N,t}^{i, l_i, s_i}(\Pi_{l_i, \leq -p_0 d_i u}^i f_i).$$

If $s_i > 0$, then by (7.12) and (7.32) we dominate the r -variational norm by the square function (in fact, the shifted square function; see [62, Theorem B.1, p. 1097]) and the conclusion in (7.40) follows by invoking Theorem 3.3.

If $s_i = 0$, then we replace $\eta_{N,t}^{i, s_i}(\xi_i) = e(\xi_i P_i(Nt)) \eta_N^{i, s_i}(\xi_i)$ with $\eta_N^{i, s_i}(\xi_i)$ and the error term is controlled by the standard Littlewood–Paley arguments (see, for example, [30] or [74]) by Lemma 7.38, whereas the r -variational estimates for the latter multiplier follow from the Lépingle’s inequality [69] (see also [76]) combined with Theorem 3.26 and Theorem 3.3. $\square$

7.3.4. Major arcs estimates: high frequency case $s > Cl$. We can assume, without loss of generality, that $(s_1, \dots, s_k) \neq (0, \dots, 0)$; otherwise we are in the low frequency case. In the high frequency case, since $s > Cl$, it only suffices to obtain (7.31) with a decay of the form 2^{-cs} . Our aim will be to establish the following maximal theorem.

Theorem 7.42. *Fix $k \in \mathbb{Z}_+$ and let the exponents $2 \leq p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$ be given. Then there exists a small constant $c \in (0, 1)$ such that for all normalized $f_1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{p_k}(\mathbb{Z}^k)$ we have*

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} \left| B^{\Sigma_{l_1, \dots, l_k}}[G; \mathbf{m}_N w_N^{s_1, \dots, s_k}](f_1, \dots, f_k) \right| \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-cs} \quad (7.43)$$

for all choices of $s_1, l_1, \dots, s_k, l_k \in \mathbb{N}$ such that $s > Cl$ for some large $C \in \mathbb{R}_+$.

In contrast to the bilinear variant of Theorem 7.42 from [62], where an integration by parts argument can be employed, the general multilinear case, due to its combinatorial nature, is more delicate. The new robust tool used here is the Sobolev smoothing inequality (6.2) from Theorem 6.1 for $\mathbb{K} = \mathbb{R}$. Once Theorem 7.42 is proved, it can be readily used to obtain (7.31). Indeed, we claim that for each fixed $\rho \in (2, \infty)$ and $1 < p_1, \dots, p_k < \infty$ one has

$$\left\| \mathbf{V}^\rho \left(B^{\Sigma_{l_1, \dots, l_k}}[G; \mathbf{m}_N w_N^{s_1, \dots, s_k}](f_1, \dots, f_k) : N \in \mathbb{D}_{l,s} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{o(s)} \quad (7.44)$$

with any normalized $f_1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{p_k}(\mathbb{Z}^k)$. To prove this claim, we note that $u = 100k(s+1)$ in the high frequency case, and using (7.35) and Proposition 7.37, we reduce inequality (7.44) to establishing

$$\left\| \mathbf{V}^\rho \left(\tilde{A}_{2^u} \left(F_{N,t,u}^{1,l_1,s_1}, \dots, F_{N,t,u}^{k,l_k,s_k} \right) : N \in \mathbb{D}_{l,s} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{o(s)},$$

where

$$F_{N,t,u}^{i,l_i,s_i} := \Pi_{l_i, \leq -p_0 d_i u}^i (F_{N,t}^{i,l_i,s_i}(f_i)), \quad i \in [k]. \quad (7.45)$$

The latter inequality follows from (2.10), Hölder's inequality, and Lemma 7.39. By interpolating (7.44) and (7.43), or rather their unnormalized versions, and renormalizing the resulting bound, we obtain the desired claim in (7.31).

The key ingredient in proving (7.43) will be Theorem 6.1 for $\mathbb{K} = \mathbb{R}$. In order to apply Theorem 6.1, we have to pass to the continuous setting. To this end, for $p \in [1, \infty)$, we define an extension operator $E: \ell^p(\mathbb{Z}^k) \rightarrow L^p(\mathbb{R}^k)$ by

$$Ef(x) := \sum_{y \in \mathbb{Z}^k} \mathbb{1}_{\mathcal{C}^k}(x-y) f(y), \quad x \in \mathbb{R}^k, \quad (7.46)$$

for each $f \in \ell^p(\mathbb{Z}^k)$, where $\mathcal{C}^k := [-1/2, 1/2]^k$. Observe that $Ef \equiv f$ on $\mathbb{Z}^k$ and $\|Ef\|_{L^p(\mathbb{R}^k)} = \|f\|_{\ell^p(\mathbb{Z}^k)}$. Thus, in particular, $Ef \in L^p(\mathbb{R}^k)$. Moreover, $Ef(x+t) = Ef(x)$ holds for all $x \in \mathbb{Z}^k$ and $t \in \mathcal{C}^k$.

Proof of Theorem 7.42. The proof is complicated and for the sake of clarity it will be divided into a few simpler blocks. We claim that it suffices to find a small constant $c' \in (0, 1)$ such that for every $\theta \in \Sigma_{l_1, \dots, l_k}$ we have

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} \left| I_N^\theta(f_1, \dots, f_k) \right| \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-c's}, \quad (7.47)$$

where

$$I_N^\theta(f_1, \dots, f_k)(x) := \sum_{y \in \mathbb{Z}^k} K_{\tau_\theta(\mathbf{m}_N w_N^{s_1, \dots, s_k})}(y) \prod_{i \in [k]} f_i(x - (y \cdot e_i) e_i), \quad x \in \mathbb{Z}^k.$$

Indeed, since $s > Cl$ and $C \in \mathbb{R}_+$ may be as large as we need, by the triangle inequality and the estimate $\#\Sigma_{l_1, \dots, l_k} \lesssim 2^{2kl}$ we obtain

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} \left| B^{\Sigma_{l_1, \dots, l_k}}[G; \mathbf{m}_N w_N^{s_1, \dots, s_k}](f_1, \dots, f_k) \right| \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{2kl} 2^{-c's} \lesssim 2^{-cs}.$$

Step 1. It suffices to prove (7.47). Fix $\theta \in \Sigma_{l_1, \dots, l_k}$ and observe that

$$I_N^\theta(f_1, \dots, f_k)(x) = e(-k\theta \cdot x) I_N^0(M_\theta f_1, \dots, M_\theta f_k)(x),$$

where M_θ is the modulation operator $M_\theta f(x) := e(\theta \cdot x) f(x)$.

In view of this identity, the proof of (7.47) is reduced to showing

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} \left| I_N^0(g_1, \dots, g_k) \right| \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-c's} \quad (7.48)$$

for any normalized $g_i \in \ell^{p_i}(\mathbb{Z}^k)$. We abbreviate $K_{\mathbf{m}_N w_N^{s_1, \dots, s_k}}$ to K_N and recall the convenient notation $x_{(y_i)} := (x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_k)$ for all $x, y \in \mathbb{R}^k$. Using the extension operator E from (7.46), we can rewrite $I_N^0(g_1, \dots, g_k)$ as $I_N^0(Eg_1, \dots, Eg_k)$. Now, the p -th power of the norm in (7.48) takes the form

$$\sum_{x \in \mathbb{Z}^k} \int_{\mathcal{C}^k} \sup_{N \in \mathbb{D}_{l,s}} \left| \sum_{y \in \mathbb{Z}^k} \int_{\mathcal{C}^k} K_N(x - y) \prod_{i \in [k]} Eg_i((x + t)_{((y+t')_i)}) dt' \right|^p dt, \quad (7.49)$$

since $Eg_i(x_{(y_i)}) = Eg_i((x + t)_{((y+t')_i)})$ for all $t, t' \in \mathcal{C}^k$ and $x, y \in \mathbb{Z}^k$.

As in Proposition 7.28, using Lemma 7.24 we can show the bound

$$|K_N(x + z) - K_N(x)| \lesssim N^{-1/2} N^{-D} \prod_{i \in [k]} \langle x_i / N^{d_i} \rangle^{-2} \quad (7.50)$$

for all $x \in \mathbb{Z}^k$ and $z \in 2\mathcal{C}^k$. Therefore, denoting

$$\tilde{K}_N(x, y, t, t') := K_N((x + t) - (y + t')) - K_N(x - y),$$

and appealing to (7.50), we readily obtain

$$\sum_{x \in \mathbb{Z}^k} \int_{\mathcal{C}^k} \sup_{N \in \mathbb{D}_{l,s}} \left| \sum_{y \in \mathbb{Z}^k} \int_{\mathcal{C}^k} \tilde{K}_N(x, y, t, t') \prod_{i \in [k]} Eg_i((x + t)_{((y+t')_i)}) dt' \right|^p dt \lesssim 2^{-pc's}.$$

Indeed, the above bound holds by Minkowski's integral inequality followed by Hölder's inequality, as in Proposition 7.28. Next, define

$$\tilde{I}_N^0(G_1, \dots, G_k)(x) := \int_{\mathbb{R}^k} K_N(y) \prod_{i \in [k]} G_i(x - (y \cdot e_i)e_i) dy, \quad x \in \mathbb{R}^k.$$

After a change of variables $\|\sup_{N \in \mathbb{D}_{l,s}} \|\tilde{I}_N^0(Eg_1, \dots, Eg_k)\|_{L^p(\mathbb{R}^k)}^p$ is equal to

$$\sum_{x \in \mathbb{Z}^k} \int_{\mathcal{C}^k} \sup_{N \in \mathbb{D}_{l,s}} \left| \sum_{y \in \mathbb{Z}^k} \int_{\mathcal{C}^k} K_N((x+t) - (y+t')) \prod_{i \in [k]} Eg_i((x+t)_{((y+t')_i)}) dt' \right|^p dt.$$

Using (7.49) and gathering everything together, we conclude that

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} |I_N^0(g_1, \dots, g_k)| \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim 2^{-c's} + \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{I}_N^0(Eg_1, \dots, Eg_k)| \right\|_{L^p(\mathbb{R}^k)}.$$

At this stage, Theorem 6.1 can be applied, and it remains to establish

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{I}_N^0(G_1, \dots, G_k)| \right\|_{L^p(\mathbb{R}^k)} \lesssim 2^{-c's} \quad (7.51)$$

for any normalized $G_i \in L^{p_i}(\mathbb{R}^k)$, since $\|Eg_i\|_{L^{p_i}(\mathbb{R}^k)} = \|g_i\|_{\ell^{p_i}(\mathbb{Z}^k)} = 1$.

Step 2. Define $V := \{i \in [k] : s_i = 0\}$ and $v := \#V < k$, and write

$$V = \{i_1, \dots, i_v\} \quad \text{and} \quad V^c = \{i_{v+1}, \dots, i_k\}.$$

We interpret the polynomial mapping $\mathcal{P}$ from (4.1) as a vector in $\mathbb{R}^k$ writing

$$\mathcal{P}(t) = \mathcal{P}_V(t) + \mathcal{P}_{V^c}(t), \quad t \in \mathbb{R},$$

where $\mathcal{P}_V(t) := \sum_{i \in V} P_i(t)e_i \in \mathbb{R}^k$. For each $i_m \in V$, we can further write

$$e(\xi_{i_m} P_{i_m}(Nt)) = \sum_{n_m \in \mathbb{N}} \frac{(2\pi i N^{d_{i_m}} \xi_{i_m})^{n_m}}{n_m!} \frac{P_{i_m}(Nt)^{n_m}}{N^{d_{i_m} n_m}},$$

using the Taylor expansion of the function $x \mapsto e(x)$. Since

$$K_N(y) = \int_{\mathbb{R}^k} \mathbf{m}_N(\xi) w_N^{s_1, \dots, s_k}(\xi) e(-\xi \cdot y) d\xi,$$

we write $K_N(y)$ as

$$\sum_{(n_1, \dots, n_v) \in \mathbb{N}^v} \int_{\mathbb{R}^k} \left(\prod_{m \in [v]} \frac{\tilde{\eta}_{N, n_m}^{i_m, 0}(\xi \cdot e_{i_m})}{n_m!} \right) \mathbf{m}_{N, V^c}^{n_1, \dots, n_v}(\xi) w_N^{s_1, \dots, s_k}(\xi) e(-\xi \cdot y) d\xi,$$

where for $\zeta \in \mathbb{R}$, $\xi \in \mathbb{R}^k$, and $i_m \in V$ we define

$$\begin{aligned}\tilde{\eta}_{N,n_m}^{i_m,0}(\zeta) &:= (2\pi i N^{d_{i_m}} \zeta)^{n_m} \eta_N^{i_m,0}(\zeta), \\ \mathfrak{m}_{N,V^c}^{n_1,\dots,n_v}(\xi) &:= 2 \int_{1/2}^1 e(\xi \cdot \mathcal{P}_{V^c}(Nt)) B_N^t dt, \\ w_{N,V^c}^{s_1,\dots,s_k}(\xi) &:= \prod_{i \in V^c} \eta_N^{i,s_i}(\xi \cdot e_i),\end{aligned}$$

with $B_N^t = B_N^t(V; n_1, \dots, n_v)$ given by the formula

$$B_N^t := \prod_{m \in [v]} \frac{P_{i_m}(Nt)^{n_m}}{N^{d_{i_m} n_m}}.$$

These Taylor expansions reveal that

$$\tilde{I}_N^0(G_1, \dots, G_k)(x) = \int_{\mathbb{R}^k} K_N(y) \prod_{i \in [k]} G_i(x - (y \cdot e_i) e_i) dy,$$

which we can further rewrite as

$$\sum_{(n_1, \dots, n_v) \in \mathbb{N}^v} \left(\prod_{m \in [v]} \frac{\tilde{G}_{N,n_m}^{i_m,0}(x)}{n_m!} \right) A_{N,V^c}^{n_1, \dots, n_v}(G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}})(x), \quad (7.52)$$

where

$$\tilde{G}_{N,n}^{i,0} := T_{i, \mathbb{R}^k}[\tilde{\eta}_{N,n}^{i,0}] G_i \quad \text{and} \quad G_N^{i,s_i} := T_{i, \mathbb{R}^k}[\eta_N^{i,s_i}] G_i,$$

and

$$A_{N,V^c}^{n_1, \dots, n_v}(G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}})(x) := 2 \int_{1/2}^1 \left(\prod_{i \in V^c} G_N^{i,s_i}(x - P_i(Nt) e_i) \right) B_N^t dt.$$

Step 3. Integrating the right-hand side above by parts, we obtain

$$A_{N,V^c}^{n_1, \dots, n_v}(G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}})(x) = A_N^1(x) B_N^1 - \int_{1/2}^1 A_N^t(x) \frac{\partial B_N^t}{\partial t} dt \quad (7.53)$$

with $A_N^t = A_N^t(V^c; G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}})$ given by the formula

$$A_N^t(x) := 2 \int_{1/2}^t \prod_{i \in V^c} G_N^{i,s_i}(x - P_i(Ny) e_i) dy.$$

Let $1 \leq q_V, q_{V^c} < \infty$ be defined by $\frac{1}{q_V} = \sum_{i \in V} \frac{1}{p_i}$ and $\frac{1}{q_{V^c}} = \sum_{i \in V^c} \frac{1}{p_i}$ (if $V = \emptyset$, then we define only q_{V^c}). Now by the Sobolev smoothing inequality from Theorem 6.1 we obtain

$$\left\| A_{N; \mathbb{R}^k}^t(G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}}) \right\|_{L^{q_{V^c}}(\mathbb{R}^k)} \lesssim 2^{-c's} \prod_{i \in V^c} \|G_N^{i,s_i}\|_{L^{p_i}(\mathbb{R}^k)} \quad (7.54)$$

uniformly in $t \in [1/2, 1]$; here, it is crucial that V^c is nonempty when $s > Cl$. By (7.53) and (7.54), we obtain

$$\left\| A_{N,V^c}^{n_1, \dots, n_v} (G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}}) \right\|_{L^{q_{V^c}}(\mathbb{R}^k)} \lesssim C_{n_1, \dots, n_v}(s) \prod_{i \in V^c} \|G_N^{i, s_i}\|_{L^{p_i}(\mathbb{R}^k)}$$

with $C_{n_1, \dots, n_v}(s) := C_{i_1}^{n_1} \dots C_{i_v}^{n_v} 2^{-c's}$ for some large constants $C_{i_m} \in \mathbb{R}_+$.

Step 4. Now, we can bound $\left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{I}_N^0(G_1, \dots, G_k)| \right\|_{L^p(\mathbb{R}^k)}$ by

$$\begin{aligned} & \sum_{(n_1, \dots, n_v) \in \mathbb{N}^v} \frac{1}{n_1! \dots n_v!} \left\| \sup_{N \in \mathbb{D}_{l,s}} \left\| \prod_{m \in [v]} \tilde{G}_{N, n_m}^{i_m, 0} \right\|_{L^{q_V}(\mathbb{R}^k)} \right\|_{L^{q_{V^c}}(\mathbb{R}^k)} \\ & \times \left\| \sup_{N \in \mathbb{D}_{l,s}} \left\| A_{N, V^c}^{n_1, \dots, n_v} (G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}}) \right\|_{L^{q_{V^c}}(\mathbb{R}^k)} \right\|_{L^{q_V}(\mathbb{R}^k)} \end{aligned} \quad (7.55)$$

using (7.52) and Hölder's inequality; here, the first line in (7.55) does not appear when $V = \emptyset$. Using Lemma 7.38 and the properties of η , we obtain

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} \left\| \tilde{G}_{N, n_m}^{i_m, 0} \right\|_{L^{p_{i_m}}(\mathbb{R}^k)} \right\|_{L^{p_{i_m}}(\mathbb{R}^k)} \lesssim C_{i_m}^{n_m} \quad (7.56)$$

for every $m \in [v]$. By (7.56) and Hölder's inequality the first norm in (7.55) is controlled by $C_{i_1}^{n_1} \dots C_{i_v}^{n_v}$. By the final estimate from Step 3 and Hölder's inequality, the second norm in (7.55) does not exceed

$$\begin{aligned} & \left(\sum_{N \in \mathbb{D}_{l,s}} \left\| A_{N, V^c}^{n_1, \dots, n_v} (G_N^{i_{v+1}, s_{i_{v+1}}}, \dots, G_N^{i_k, s_{i_k}}) \right\|_{L^{q_{V^c}}(\mathbb{R}^k)} \right)^{1/q_{V^c}} \\ & \lesssim C_{n_1, \dots, n_v}(s) \left(\sum_{N \in \mathbb{D}_{l,s}} \prod_{i \in V^c} \|G_N^{i, s_i}\|_{L^{p_i}(\mathbb{R}^k)}^{q_{V^c}} \right)^{1/q_{V^c}} \\ & \leq C_{n_1, \dots, n_v}(s) \prod_{i \in V^c} \left\| \left(\sum_{N \in \mathbb{D}_{l,s}} |G_N^{i, s_i}|^{p_i} \right)^{1/p_i} \right\|_{L^{p_i}(\mathbb{R}^k)} \lesssim C_{n_1, \dots, n_v}(s), \end{aligned} \quad (7.57)$$

where, in the last line, we have used a standard square function argument [30] (see also [74]), which is applicable because $p_i \geq 2$. Finally, using (7.55)–(7.57) we deduce (7.51), since the factors $(C_{i_1}^{n_1} \dots C_{i_v}^{n_v})^2 (n_1! \dots n_v!)^{-1}$ are summable over $(n_1, \dots, n_v) \in \mathbb{N}^v$. This completes the proof of Theorem 7.42. $\square$

7.3.5. Major arcs estimates: low frequency case $Cl \geq s$. By (7.33) in the low frequency case we have

$$u := 100k(l+1). \quad (7.58)$$

In this case, since $Cl \geq s$, the proof of (7.31) follows from (7.35) if for each fixed $r \in (2, \infty)$ we show the estimate

$$\left\| \mathbf{V}^r \left(B^{\Sigma_{l_1, \dots, l_k}} [G; \eta_u^*] (F_{N,t}^{1, l_1, s_1}, \dots, F_{N,t}^{k, l_k, s_k}) : N \in \mathbb{D}_{l,s} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim (s+1)^k 2^{-cl}$$

for all normalized $f_i \in \ell^{p_i}(\mathbb{Z}^k)$ and all choices of $s_1, l_1, \dots, s_k, l_k \in \mathbb{N}$, uniformly in $t \in [\frac{1}{2}, 1]$, with $F_{N,t}^{i,l_i,s_i} = F_{N,t}^{i,l_i,s_i}(f_i)$ given in (7.36). In view of (7.45), Proposition 7.37, and Lemma 7.39, the above task reduces to showing

$$\left\| \mathbf{V}^r \left(\tilde{A}_{2^u} (F_{N,t,u}^{1,l_1,s_1}, \dots, F_{N,t,u}^{k,l_k,s_k}) : N \in \mathbb{D}_{l,s} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim (s+1)^k 2^{-cl}, \quad (7.59)$$

again for all normalized $f_i \in \ell^{p_i}(\mathbb{Z}^k)$ and all choices of $l_1, s_1, \dots, l_k, s_k \in \mathbb{N}$, uniformly in $t \in [\frac{1}{2}, 1]$, with $F_{N,t,u}^{i,l_i,s_i}$ given in (7.45).

The advantage of working with the average $\tilde{A}_{2^u}$ in (7.59) is that we can exploit the multilinear Weyl inequality (6.2) from Theorem 6.1 to produce an exponential decay in l as long as the underlying functions have Fourier transforms vanishing on suitable major arcs, like in (7.59). Observe that for each fixed $\rho \in (2, \infty)$, by Hölder's inequality and Lemma 7.39, for all normalized $f_i \in \ell^{p_i}(\mathbb{Z}^k)$ and all choices of $l_1, s_1, \dots, l_k, s_k \in \mathbb{N}$ one has

$$\begin{aligned} & \sup_{1/2 \leq t \leq 1} \left\| \mathbf{V}^\rho \left(\tilde{A}_{2^u} (F_{N,t,u}^{1,l_1,s_1}, \dots, F_{N,t,u}^{k,l_k,s_k}) : N \in \mathbb{D}_{l,s} \right) \right\|_{\ell^p(\mathbb{Z}^k)} \\ & \lesssim \prod_{i \in [k]} \sup_{1/2 \leq t \leq 1} \left\| \mathbf{V}^\rho \left(F_{N,t,u}^{i,l_i,s_i} : N \in \mathbb{D}_{l,s} \right) \right\|_{\ell^{p_i}(\mathbb{Z}^k)} \\ & \lesssim (s+1)^k \left(2^{O_{p_1, \dots, p_k}(\frac{l \log \log l}{\log t})} \mathbb{1}_{l \geq 10} + \mathbb{1}_{l < 10} \right). \end{aligned} \quad (7.60)$$

Our aim will be to establish the following maximal theorem.

Theorem 7.61. *Fix $k \in \mathbb{Z}_+$ and let the exponents $1 < p_1, \dots, p_k < \infty$ such that $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$ be given. Then there exists a small constant $\delta \in (0, 1)$ such that for all normalized $f_1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{p_k}(\mathbb{Z}^k)$ we have*

$$\sup_{1/2 \leq t \leq 1} \left\| \sup_{N \in \mathbb{D}_{l,s}} \left| \tilde{A}_{2^u} (F_{N,t,u}^{1,l_1,s_1}, \dots, F_{N,t,u}^{k,l_k,s_k}) \right| \right\|_{\ell^p(\mathbb{Z}^k)} \lesssim (s+1)^k 2^{-\delta l}$$

for all choices of $l_1, s_1, \dots, k_k, s_k \in \mathbb{N}$, with $F_{N,t,u}^{i,l_i,s_i}$ given in (7.45).

Fixing $r \in (2, \infty)$ and $\rho \in (2, r)$, interpolating the unnormalized $\mathbf{V}^\rho$ bound from (7.60) with the unnormalized maximal function bound from Theorem 7.61, and renormalizing, we readily deduce (7.59), as desired.

7.3.6. Multiparameter norm interchanging inequality. Here we formulate a multiparameter variant of the norm interchanging inequality from [62, Lemma 9.5, p. 1079]. The norm interchanging inequality from [62] asserts that for any measure space $(X, \mathcal{B}(X), \mu)$, any $K \in \mathbb{Z}_+$ and $1 \leq R < r \leq \infty$, and for any sequence $(f_k)_{k \in [K]} \subseteq L^r(X)$ we have

$$\|V^r(f_k : k \in [K])\|_{L^r(X)} \lesssim_{r,R} \|(f_k)_{k \in [K]}\|_{V^R([K]; L^r(X))}, \quad (7.62)$$

where

$$\|(f_k)_{k \in [K]}\|_{V^R([K]; L^r(X))} := \sup_{J \in \mathbb{Z}_+} \sup_{k_0 < \dots < k_J} \left(\sum_{j=0}^{J-1} \|f_{k_{j+1}} - f_{k_j}\|_{L^r(X)}^R \right)^{1/R}$$

with the inner supremum taken over all increasing sequences from $[K]$ of length $J + 1$. Inequality (7.62) states that we can insert the $L^r(X)$ norm under the r -variation norm, at the cost of worsening the r -variation parameter to $R < r$.

We begin with fixing some notation and terminology. Let $\mathbb{J} \subseteq \mathbb{R}_+$ be an at most countable set such that $\min \mathbb{J} \in \mathbb{J}$. Let $\mathbf{I}(\mathbb{J})$ denote the set of all finite increasing sequences from $\mathbb{J}$. The elements of $\mathbf{I}(\mathbb{J})$ will be denoted by bold letters $\mathbf{N}_{<J} = (N_j)_{j \in \mathbb{N}_{<J}}$ for some $J \in \mathbb{Z}_+$.

Fix $k \in \mathbb{Z}_+$ and let $W_1, \dots, W_k$ be linear normed spaces. Let $d \in \mathbb{Z}_+$ and $R_1, R_2 \in [1, \infty]$. Suppose that $B: W_1 \times \dots \times W_k \rightarrow \ell^{R_1}(\mathbb{Z}^d)$ is a multisublinear operator. By this we mean that for every $f_1, g_1 \in W_1, \dots, f_k, g_k \in W_k$ and any $j \in [k]$ we have the pointwise inequality

$$\begin{aligned} & \left| |B(f_1, \dots, f_{j-1}, f_j, f_{j+1}, \dots, f_k)| - |B(f_1, \dots, f_{j-1}, g_j, f_{j+1}, \dots, f_k)| \right| \\ & \leq |B(f_1, \dots, f_{j-1}, f_j - g_j, f_{j+1}, \dots, f_k)|. \end{aligned}$$

For any $(J_1, \dots, J_k) \in \mathbb{Z}_+^k$ and $\sigma = (\sigma_1, \dots, \sigma_k) \in \{0, 1\}^k$ define

$$\mathbf{N}_{<J_i}^{\sigma_i} := \begin{cases} \mathbf{N}_{<J_i} & \text{if } \sigma_i = 1, \\ \{-1\} & \text{if } \sigma_i = 0. \end{cases}$$

Moreover, for every $h_i = (h_{i,N})_{N \in \mathbb{J}} \subseteq W_i$ and $\mathbf{N}_{<J_i}^i = (N_j^i)_{j \in \mathbb{N}_{<J_i}} \in \mathbf{I}(\mathbb{J})$ for $i \in [k]$ define $\Lambda_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, R_1, R_2}(h_1, \dots, h_k)$ by

$$\left(\sum_{j_1 \in \mathbb{N}_{<J_1}^{\sigma_1}, \dots, j_k \in \mathbb{N}_{<J_k}^{\sigma_k}} \|B(\Delta h_{1, N_{j_1}^1}^{\sigma_1}, \dots, \Delta h_{k, N_{j_k}^k}^{\sigma_k})\|_{\ell^{R_1}(\mathbb{Z}^d)}^{R_2} \right)^{1/R_2},$$

where we set $N_{-1}^i := \min \mathbb{J} \in \mathbb{J}$ and

$$\Delta h_{i, N_{j_i}^i}^{\sigma_i} := \begin{cases} h_{i, N_{j_i}^i} - h_{i, N_{j_i-1}^i} & \text{if } \sigma_i = 1, \\ h_{i, N_{-1}^i} & \text{if } \sigma_i = 0. \end{cases}$$

Proposition 7.63. *Let $\mathbb{J} \subset \mathbb{R}_+$ be at most countable with $N_{-1} := \min \mathbb{J} \in \mathbb{J}$. Fix $k \in \mathbb{Z}_+$ and linear normed spaces $W_1, \dots, W_k$. Let $d \in \mathbb{Z}_+$ and $1 \leq R < r \leq \infty$. Suppose that $B: W_1 \times \dots \times W_k \rightarrow \ell^R(\mathbb{Z}^d)$ is a multisublinear operator. Then for every $h_i = (h_{i,N})_{N \in \mathbb{J}} \subseteq W_i$ for $i \in [k]$, we have*

$$\begin{aligned} & \left\| \sup_{N_1, \dots, N_k \in \mathbb{J}} |B(h_{1, N_1}, \dots, h_{k, N_k})| \right\|_{\ell^r(\mathbb{Z}^d)} \\ & \lesssim_{r, R} \sum_{\sigma \in \{0, 1\}^k} \sup_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k \in \mathbf{I}(\mathbb{J})} \Lambda_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, R, R}(h_1, \dots, h_k). \end{aligned} \tag{7.64}$$

Proof. The proof follows from simple iterative applications of (7.62). $\square$

The advantage of using Proposition 7.63 is that we do not distinguish between small and large scales, which was an essential argument in [62]. This is another important simplification of our method. Proposition 7.63 will be the key tool in our approach and will be applied with $d = k$, $B = \hat{A}_{2^u}$, and $W_i = \ell^{R_i}(\mathbb{Z}^k)$ for $i \in [k]$, where $\frac{1}{R} = \frac{1}{R_1} + \dots + \frac{1}{R_k}$.

7.3.7. Uncertainty principle. We will need a smooth partition of unity $(\phi_I : I \in \mathcal{I}_h)$ adapted to the partition $\mathcal{I}_h := \{[lh, (l+1)h) : l \in \mathbb{Z}\}$ of $\mathbb{R}$ into disjoint intervals I of equal length $h \in \mathbb{R}_+$. In fact, we can take

$$\phi_I(x) := \phi(h^{-1}(x - c(I))) \quad (7.65)$$

for each $I_l := [lh, (l+1)h) \in \mathcal{I}_h$, where $c(I_l) := (l + \frac{1}{2})h$ is the center of I_l , and $\phi : \mathbb{R} \rightarrow [0, 1]$ is a suitable smooth bump function. Using (7.65) we define

$$\phi_I^i(x) := \phi_I(x \cdot e_i), \quad x \in \mathbb{Z}^k, \quad (7.66)$$

for every $i \in [k]$. We will need two technical lemmas. Here, the symbol δ_0 stands for the Dirac delta distribution at the origin.

Lemma 7.67. *Fix $i \in [k]$, $\varepsilon \in (0, 1)$, and an interval I of length comparable to $2^{d_i u}$. Then, for $M \in \mathbb{N}$ and $x \in \mathbb{Z}^k$, one has*

$$|\phi_I^i * (\delta_0 - \mathcal{F}_{\mathbb{R}}^{-1} \eta_{\leq -(d_i - \varepsilon)u})(x)| \lesssim_{\varepsilon, M} 2^{-Mu} \quad (7.68)$$

and, for $c(I)$ being the center of I , one also has

$$|\phi_I^i * \mathcal{F}_{\mathbb{R}}^{-1} \eta_{\leq -(d_i - \varepsilon)u}(x)| \lesssim_M \langle 2^{-d_i u} |x \cdot e_i - c(I)| \rangle^{-M}. \quad (7.69)$$

In particular, combining (7.68) and (7.69) one obtains

$$|\phi_I^i * (\delta_0 - \mathcal{F}_{\mathbb{R}}^{-1} \eta_{\leq -(d_i - \varepsilon)u})(x)| \lesssim_{\varepsilon, M} 2^{-Mu} \langle 2^{-d_i u} |x \cdot e_i - c(I)| \rangle^{-M}. \quad (7.70)$$

Proof. Regarding (7.68), by the Poisson summation formula, we have

$$|\mathcal{F}_{\mathbb{Z}} \phi_I(\xi)| = \left| \sum_{n \in \mathbb{Z}} \mathcal{F}_{\mathbb{R}}^{-1} \phi_I(\xi + n) \right| \lesssim_N \frac{2^{d_i u}}{(2^{d_i u} \|\xi\|)^N}$$

for every $N \in \mathbb{N}$, with $\|\xi\| := \min_{n \in \mathbb{Z}} |\xi + n|$. Observe that

$$\sup_{x \in \mathbb{Z}^k} |\phi_I^i * (\delta_0 - \mathcal{F}_{\mathbb{R}}^{-1} \eta_{\leq -(d_i - \varepsilon)u})(x)| \leq \int_{\mathbb{T}} |\mathcal{F}_{\mathbb{Z}} \phi_I(\xi)| |1 - \eta_{\leq -(d_i - \varepsilon)u}(\xi)| d\xi.$$

Since $2^{d_i u} \|\xi\| \gtrsim 2^{\varepsilon u}$ whenever $1 - \eta_{\leq -(d_i - \varepsilon)u}(\xi) \neq 0$, applying the bound for $|\mathcal{F}_{\mathbb{Z}} \phi_I(\xi)|$ with some fixed $N \simeq (M + d_i)/\varepsilon$, we can control the integral above by 2^{-Mu} , as desired. Regarding (7.69), we can control its left-hand side by

$$\sum_{y \in \mathbb{Z}} \left\langle 2^{-d_i u} |x \cdot e_i - y - c(I)| \right\rangle^{-M} 2^{-(d_i - \varepsilon)u} \left\langle 2^{-(d_i - \varepsilon)u} |y| \right\rangle^{-2M}$$

and use the fact that $|x \cdot e_i - c(I)| \leq |x \cdot e_i - y - c(I)| + |y|$ for each $y \in \mathbb{Z}$. $\square$

Lemma 7.71. *Fix $i \in [k]$, $\varepsilon \in (0, 1)$, $f \in \ell^2(\mathbb{Z}^k)$, and an interval I . Then*

$$(\phi_I^i * \mathcal{F}_{\mathbb{R}}^{-1} \eta_{\leq -(d_i - \varepsilon)u}) \Pi_{l_i, \leq -p_0 d_i u}^i(f) \in \ell^2(\mathbb{Z}^k) \quad (7.72)$$

has its i -th Fourier transform supported on the major arcs $\mathcal{M}_{l_i, \leq -(d_i - \varepsilon)u+1}^i$.

Proof. Our claim readily follows, since the integral

$$\int_{\mathbb{T}} \sum_{\theta \in \Sigma_{l_i}} \eta_{\leq -p_0 d_i u}(\xi - \zeta - \theta) \mathcal{F}_{i, \mathbb{Z}^k} f(x_{(\xi - \zeta)}) \mathcal{F}_{\mathbb{Z}} \phi_I(\zeta) \eta_{\leq -(d_i - \varepsilon)u}(\zeta) d\zeta$$

is the i -th Fourier transform of the function from (7.72), evaluated at $x_{(\xi)}$. $\square$

7.3.8. All together: proof of Theorem 7.61. We now focus our attention on proving Theorem 7.61. To simplify notation, we write

$$F_N^i(f_i) := F_{N, t, u}^{i, l_i, s_i}(f_i) = \Pi_{l_i, \leq -p_0 d_i u}^i(F_{N, t}^{i, l_i, s_i}(f_i)), \quad i \in [k], \quad (7.73)$$

following (7.45). By Lemma 7.39, we have

$$\| \sup_{N \in \mathbb{D}_{l, s}} |F_N^i(f_i)| \|_{\ell^{p_i}(\mathbb{Z}^k)} \lesssim (s+1) 2^{o_{p_i}(l)} \quad (7.74)$$

for all $i \in [k]$ and all normalized $f_i \in \ell^{p_i}(\mathbb{Z}^k)$.

Proof of Theorem 7.61. Using notation from (7.73), we see that it suffices to prove that, with some constant $\delta_p \in (0, 1)$, we have

$$\sup_{1/2 \leq t \leq 1} \| \sup_{N \in \mathbb{D}_{l, s}} |\tilde{A}_{2^u}(F_N^1(f_1), \dots, F_N^k(f_k))| \|_{\ell^p(\mathbb{Z}^k)} \lesssim (s+1)^k 2^{-\delta_p l} \quad (7.75)$$

for all normalized $f_1 \in \ell^{p_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{p_k}(\mathbb{Z}^k)$. Since our argument is fairly intricate, it will be divided into several steps.

Step 1. It suffices to prove that, with some constant $\gamma \in (0, 1)$, we have

$$\sup_{1/2 \leq t \leq 1} \| \sup_{N \in \mathbb{D}_{l, s}} |\tilde{A}_{2^u}(F_N^1(f_1), \dots, F_N^k(f_k))| \|_{\ell^4(\mathbb{Z}^k)} \lesssim (s+1)^k 2^{-\gamma l} \quad (7.76)$$

for all normalized $f_1, \dots, f_k \in \ell^{4k}(\mathbb{Z}^k)$, with $F_N^i(f_i)$ as in (7.73) for all $i \in [k]$. Any other parameter $p \in (2, \infty)$ in place of 4 in inequality (7.76) would also work. However, using the number 4 would reduce the number of parameters we need to declare later, making our argument clearer. Note that inequality (7.76) implies inequality (7.75). To see this, suppose that (7.76) holds. Combining inequality (6.24) from the proof of Theorem 6.1 with (7.74), we obtain

$$\left\| \sup_{N \in \mathbb{D}_{l, s}} |\tilde{A}_{2^u}(F_N^1(f_1), \dots, F_N^k(f_k))| \right\|_{\ell^v(\mathbb{Z}^k)} \lesssim (s+1)^k 2^{o_{v_1, \dots, v_k, v}(l)} \quad (7.77)$$

for all normalized $f_1 \in \ell^{v_1}(\mathbb{Z}^k), \dots, f_k \in \ell^{v_k}(\mathbb{Z}^k)$, where $1 < v_1, \dots, v_k < \infty$ satisfy $\frac{1}{v_1} + \dots + \frac{1}{v_k} = \frac{1}{v}$ with $v < 1$ close to 1. Interpolating the unnormalized bounds from (7.77) and (7.76), using the multilinear interpolation result from [43, Theorem 1], and renormalizing yields (7.75) for all $1 < p_1, \dots, p_k < \infty$ from Theorem 7.61 satisfying $\frac{1}{p_1} + \dots + \frac{1}{p_k} = \frac{1}{p} \leq 1$.

By a simple density argument, it suffices to prove (7.76) for all normalized $f_1, \dots, f_k \in \ell^{4k}(\mathbb{Z}^k)$ that are also compactly supported. This additional assumption ensures that all functions also belong to $\ell^2(\mathbb{Z}^k)$. Hence, from now on we assume that $f_1, \dots, f_k \in \ell^{4k}(\mathbb{Z}^k)$ are compactly supported.

Step 2. By inequality (6.2) from Theorem 6.1, there exists a constant $c := c_k \in (0, 1)$ such that for all $g_1, \dots, g_k \in \ell^{2k}(\mathbb{Z}^k)$, if $g_i \in \ell^2(\mathbb{Z}^k)$ for some $i \in [k]$ and the i -th Fourier transform of g_i vanishes on $\mathcal{M}_{\leq w_1, \leq -d_i u + w_2}^i$, then

$$\|\tilde{A}_{2^u}(g_1, \dots, g_k)\|_{\ell^2(\mathbb{Z}^k)} \lesssim (2^{-c \min\{w_1, w_2\}} + 2^{-cu}) \prod_{i \in [k]} \|g_i\|_{\ell^{2k}(\mathbb{Z}^k)}. \quad (7.78)$$

Thus, our aim is to establish (7.76) by reducing the matter to (7.78). For this purpose, we choose $r \in (4, \infty)$ such that

$$100kD \left(\frac{1}{4} - \frac{1}{r} \right) < \frac{c}{1000}, \quad (7.79)$$

where $c := c_k \in (0, 1)$ is the exponent from inequality (7.78).

Step 3. For each $i \in [k]$ we now fix a partition of unity $(\phi_{I^i} : I^i \in \mathcal{I}^i)$ as in (7.65), where $\mathcal{I}^i := \mathcal{I}_{2^{d_i u}}$ is the partition of $\mathbb{R}$ into disjoint intervals I^i of equal length $2^{d_i u}$. Next, define $\mathcal{I} := \mathcal{I}^1 \times \dots \times \mathcal{I}^k$ and observe that for every $I = I^1 \times \dots \times I^k \in \mathcal{I}$ we have

$$\tilde{A}_{2^u}(\phi_{I^1}^1 F_N^1(f_1), \dots, \phi_{I^k}^k F_N^k(f_k)) = \tilde{A}_{2^u}(\phi_{I^1}^1 F_N^1(f_1^I), \dots, \phi_{I^k}^k F_N^k(f_k^I)) \mathbf{1}_{\tilde{I}}, \quad (7.80)$$

with $\phi_{I^i}^i$ defined in (7.66), $\tilde{I} := \tilde{I}^1 \times \dots \times \tilde{I}^k \subseteq \mathbb{R}^k$, and

$$f_i^I := (\mathbf{1}_{\tilde{I}^1} \otimes \dots \otimes \mathbf{1}_{\tilde{I}^{i-1}} \otimes \mathbf{1}_{\mathbb{R}} \otimes \mathbf{1}_{\tilde{I}^{i+1}} \otimes \dots \otimes \mathbf{1}_{\tilde{I}^k}) f_i, \quad i \in [k], \quad (7.81)$$

where $\tilde{I}^i$ is an interval centered at $c(I^i)$ of length $O(2^{d_i u})$ such that, whenever $\phi_{I^i}(x_i - P_i(n)) \neq 0$ for some $x \in \mathbb{R}^k$ and $n \in [2^u]$, the arguments $x_i - P_i(n)$ and x_i both belong to $\tilde{I}^i$. Indeed, with this choice we have $\phi_{I^i} \mathbf{1}_{\tilde{I}^i} = \phi_{I^i}$ and we can also localize the operator to $\tilde{I}$. Finally, note that for each $i \in [k]$, the family of intervals $\{\tilde{I}^i : I^i \in \mathcal{I}^i\}$ has bounded overlap.

For each $i \in [k]$, define

$$G_{N,I}^i := G_{N,I}^i(f_i) := \phi_{I^i}^i F_N^i(f_i^I) \quad (7.82)$$

with f_i^I as in (7.81). Moreover, for each $\omega = (\omega_1, \dots, \omega_k) \in \{0, 1\}^k$, define

$$G_{N,I}^{i, \omega_i} := G_{N,I}^{i, \omega_i}(f_i) := \begin{cases} (\phi_{I^i}^i * \mathcal{F}_{\mathbb{R}}^{-1} \eta_{\leq -(d_i - \varepsilon)u}) F_N^i(f_i^I) & \text{if } \omega_i = 1, \\ (\phi_{I^i}^i * (\delta_0 - \mathcal{F}_{\mathbb{R}}^{-1} \eta_{\leq -(d_i - \varepsilon)u})) F_N^i(f_i^I) & \text{if } \omega_i = 0. \end{cases}$$

Note that

$$G_{N,I}^i = G_{N,I}^{i,0} + G_{N,I}^{i,1}, \quad i \in [k].$$

Using (7.80) and (7.82), and the bounded overlap of $\{\tilde{I} : I \in \mathcal{I}\}$, we may write

$$\begin{aligned} & \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(F_N^1(f_1), \dots, F_N^k(f_k))| \right\|_{\ell^4(\mathbb{Z}^k)}^4 \\ & \lesssim \sum_{\omega \in \{0,1\}^k} \sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(G_{N,I}^{1, \omega_1}, \dots, G_{N,I}^{k, \omega_k})| \right\|_{\ell^4(\tilde{I} \cap \mathbb{Z}^k)}^4. \end{aligned} \quad (7.83)$$

Step 4. We need a few properties of the functions $G_{N,I}^{i,\omega_i}$. We fix $i \in [k]$ and $p_i \in (1, \infty)$. Since $G_{N,I}^{i,1} = G_{N,I}^i - G_{N,I}^{i,0}$, we have

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^{i,1}| \right\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i} \lesssim \left\| \sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^i| \right\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i} + \left\| \sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^{i,0}| \right\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i}.$$

Observe that $\sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^i| = \phi_{I^i}^i \sup_{N \in \mathbb{D}_{l,s}} |F_N^i(f_i^I)|$ by (7.82). Moreover, $\{\tilde{I} : I \in \mathcal{I}\}$ has bounded overlap. Thus, using Lemma 7.39 we obtain

$$\left\| \sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^i| \right\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i} \lesssim C'_{p_i}(l, s_i) \|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i} \quad (7.84)$$

with a constant $C'_{p_i}(l, s_i) \in \mathbb{R}_+$ of the form $(s_i + 1)^{p_i} (2^{O_{p_i}(\frac{l \log \log l}{\log l})} \mathbf{1}_{l \geq 10} + \mathbf{1}_{l < 10})$.

Next, taking $M = 1$ in (7.70) from Lemma 7.67, we have

$$\sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^{i,0}(x)|^{p_i} \lesssim_\varepsilon 2^{-p_i u} w_i(x \cdot e_i - c(I^i)) \sup_{N \in \mathbb{D}_{l,s}} |F_N^i(f_i^I)(x)|^{p_i}, \quad x \in \mathbb{Z}^k,$$

where $w_i(y) := \langle 2^{-d_i u} |y| \rangle^{-p_i}$ for $y \in \mathbb{Z}$. Since $w_i(x \cdot e_i - c(I^i))$ are summable over $I^i \in \mathcal{I}^i$, uniformly in $x \in \mathbb{Z}^k$, we see that

$$\sum_{I^i \in \mathcal{I}^i} \left\| \sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^{i,0}(f_i)| \right\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i} \lesssim 2^{-p_i u} \left\| \sup_{N \in \mathbb{D}_{l,s}} |F_N^i(f_i^I)| \right\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i}$$

and, since $\{\tilde{I} : I \in \mathcal{I}\}$ has bounded overlap, using Lemma 7.39 we obtain

$$\sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |G_{N,I}^{i,0}| \right\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i} \lesssim 2^{-p_i u} C'_{p_i}(l, s_i) \|f_i\|_{\ell^{p_i}(\mathbb{Z}^k)}^{p_i}. \quad (7.85)$$

Finally, by Lemma 7.71 with $\varepsilon = \frac{1}{10}$ for each $i \in [k]$ the i -th Fourier transform of the function $G_{N,I}^{i,1}(f_i)$ is supported on the major arcs $\mathcal{M}_{l_i, \leq -(d_i - \varepsilon)u + 1}^i$. Thus, it vanishes on the major arcs $\mathcal{M}_{\leq l_i - 1, \leq -d_i u + \varepsilon(l_i - 1)}^i$.

Step 5. Fix $\omega \in \{0, 1\}^k \setminus \{1\}^k$. By the previous discussion, we have

$$\sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(G_{N,I}^{1,\omega_1}, \dots, G_{N,I}^{k,\omega_k})| \right\|_{\ell^4(\tilde{I} \cap \mathbb{Z}^k)}^4 \lesssim 2^{-4u} (s+1)^{4k} 2^{o_k(l)} \quad (7.86)$$

for all normalized $f_1, \dots, f_k \in \ell^{4k}(\mathbb{Z}^k)$. Indeed, to prove inequality (7.86), we use the boundedness of $\tilde{A}_{2^u}$ and Hölder's inequality, together with inequalities (7.84) and (7.85) applied with $p = 4$ and $p_i = 4k$ for all $i \in [k]$.

Step 6. It remains to consider $\omega = (1, \dots, 1)$. We shall prove the bound

$$\sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(G_{N,I}^{1,1}, \dots, G_{N,I}^{k,1})| \right\|_{\ell^4(\tilde{I} \cap \mathbb{Z}^k)}^4 \lesssim (s+1)^{4k} 2^{o_k(l)} 2^{-\frac{4cl}{40}} \quad (7.87)$$

for all normalized $f_1, \dots, f_k \in \ell^{4k}(\mathbb{Z}^k)$. This is the main part of the argument, and once inequality (7.87) is verified, the proof follows. Indeed, it suffices to combine inequalities (7.83) and (7.86) for all $\omega \in \{0, 1\}^k \setminus \{1\}^k$ (recall that now $u = 100k(l+1)$ by (7.58)), and (7.87) to deduce (7.76) with $\gamma := c/50$.

By Hölder's inequality (with $r > 4$ as in condition (7.79)), we obtain

$$\begin{aligned} \sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(G_{N,I}^{1,1}, \dots, G_{N,I}^{k,1})| \right\|_{\ell^4(\tilde{I} \cap \mathbb{Z}^k)}^4 \\ \lesssim 2^{4uD(\frac{1}{4} - \frac{1}{r})} \sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(G_{N,I}^{1,1}, \dots, G_{N,I}^{k,1})| \right\|_{\ell^r(\mathbb{Z}^k)}^4. \end{aligned} \quad (7.88)$$

In view of condition (7.79), the bound (7.87) follows from (7.88) and

$$\sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(G_{N,I}^{1,1}, \dots, G_{N,I}^{k,1})| \right\|_{\ell^r(\mathbb{Z}^k)}^4 \lesssim (s+1)^{4k} 2^{o_k(l)} 2^{-\frac{4cl}{30}}. \quad (7.89)$$

Step 7. From now on we use the notation introduced above Proposition 7.63. By the norm interchanging inequality from Proposition 7.63 with $R = 4 < r$ applied to the $\ell^r(\mathbb{Z}^k)$ norm in (7.89), followed by Hölder's inequality, we have

$$\begin{aligned} \sum_{I \in \mathcal{I}} \left\| \sup_{N \in \mathbb{D}_{l,s}} |\tilde{A}_{2^u}(G_{N,I}^{1,1}, \dots, G_{N,I}^{k,1})| \right\|_{\ell^r(\mathbb{Z}^k)}^4 \\ \lesssim_r \sum_{\sigma \in \{0,1\}^k} \sum_{I \in \mathcal{I}} \left(\sup_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k \in \mathbf{I}(\mathbb{D}_{l,s})} \Lambda_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{I, \sigma, 4, 4}(f_1, \dots, f_k) \right)^4, \end{aligned} \quad (7.90)$$

where for $\sigma = (\sigma_1, \dots, \sigma_k) \in \{0, 1\}^k$, we define $\Lambda_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{I, \sigma, 4, 4}(f_1, \dots, f_k)$ by

$$\left(\sum_{(j_1, \dots, j_k) \in \mathbb{X}_\sigma} \|\tilde{A}_{2^u}(\Delta G_{N_{j_1}^1, I}^{1,1, \sigma_1}, \dots, \Delta G_{N_{j_k}^k, I}^{k,1, \sigma_k})\|_{\ell^4(\mathbb{Z}^k)}^4 \right)^{1/4},$$

and for each $i \in [k]$, setting $N_{-1}^i := \min \mathbb{D}_{l,s}$, we have

$$\mathbb{X}_\sigma := \prod_{i \in [k]} \mathbb{N}_{<J_i}^{\sigma_i} \quad \text{with} \quad \mathbb{N}_{<J_i}^{\sigma_i} := \begin{cases} \mathbb{N}_{<J_i} & \text{if } \sigma_i = 1, \\ \{-1\} & \text{if } \sigma_i = 0, \end{cases}$$

and

$$\Delta G_{N_{j_i}^i, I}^{i,1, \sigma_i} := \Delta G_{N_{j_i}^i, I}^{i,1, \sigma_i}(f_i) := \begin{cases} G_{N_{j_i}^i, I}^{i,1}(f_i) - G_{N_{j_i-1}^i, I}^{i,1}(f_i) & \text{if } \sigma_i = 1, \\ G_{N_{-1}^i, I}^{i,1}(f_i) & \text{if } \sigma_i = 0. \end{cases}$$

In view of the final remark in Step 4, for each $i \in [k]$, the i -th Fourier transform of $\Delta G_{N_{j_i}^i, I}^{i,1, \sigma_i}$ is supported on the major arcs $\mathcal{M}_{l_i, \leq -(d_i - \varepsilon)u+1}^i$ with $\varepsilon = \frac{1}{10}$. Thus, it vanishes on $\mathcal{M}_{\leq l_i-1, \leq -d_i u + \varepsilon(l_i-1)}^i$ and we may write

$$\Delta G_{N_{j_i}^i, I}^{i,1, \sigma_i} = \Delta G_{N_{j_i}^i, I}^{i,1, \sigma_i} - \Pi_{\leq l_i-1, \leq -d_i u + \varepsilon(l_i-1)}^i \Delta G_{N_{j_i}^i, I}^{i,1, \sigma_i}.$$

Step 8. Fix $\sigma \in \{0, 1\}^k$ and $\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k \in \mathbf{I}(\mathbb{D}_{l,s})$, and let $\mathfrak{S}(\sigma) := \{i \in [k] : \sigma_i = 1\}$. We may assume, without loss of generality, that $\mathfrak{S}(\sigma) = [m]$ for some $m \in \mathbb{N}_{\leq k}$. In order to simplify notation, for each $i \in [k]$, we define

$$H_i(f) := f - \Pi_{\leq l_i-1, \leq -d_i u + \varepsilon(l_i-1)}^i f,$$

and, assuming momentarily that $I \in \mathcal{I}$ is fixed, we set

$$g_{j_i}^i := g_{j_i}^i(f_i) := \Delta G_{N_{j_i}^i, I}^{i,1,1}(f_i) \quad \text{and} \quad g_i := g_i(f_i) := G_{N_{-1}^i, I}^{i,1}(f_i).$$

Given $p \in [1, \infty]$, define $\Theta_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, p}(\mathbf{h}_1, \dots, \mathbf{h}_m, h_{m+1}, \dots, h_k)$ by setting

$$\left(\sum_{(j_1, \dots, j_k) \in \mathbb{X}_\sigma} \|\tilde{A}_{2^u}(H_1(h_{j_1}^1), \dots, H_m(h_{j_m}^m), H_{m+1}(h_{m+1}), \dots, H_k(h_k))\|_{\ell^p(\mathbb{Z}^k)}^p \right)^{1/p}$$

for all sequences of functions $\mathbf{h}_1 = (h_{j_1}^1)_{j_1 \in \mathbb{N}_{<J_1}}, \dots, \mathbf{h}_m = (h_{j_m}^m)_{j_m \in \mathbb{N}_{<J_m}}$ and all functions $h_{m+1}, \dots, h_k$ for which this formula makes sense. Note that

$$\Theta_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{(0, \dots, 0), p}(h_1, \dots, h_k) = \|\tilde{A}_{2^u}(H_1(h_1), \dots, H_k(h_k))\|_{\ell^p(\mathbb{Z}^k)}.$$

We also observe that $\Lambda_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{I, \sigma, 4, 4}(f_1, \dots, f_k) = \Theta_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, 4}(\mathbf{g}_1, \dots, \mathbf{g}_m, g_{m+1}, \dots, g_k)$ for $\mathbf{g}_1 = (g_{j_1}^1)_{j_1 \in \mathbb{N}_{<J_1}}, \dots, \mathbf{g}_m = (g_{j_m}^m)_{j_m \in \mathbb{N}_{<J_m}}$ and $g_{m+1}, \dots, g_k$ as before.

Given $\sigma \in \{0, 1\}^k$ and $\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k \in \mathbf{I}(\mathbb{D}_{l,s})$, we will prove that

$$\begin{aligned} & \Theta_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, 4}(\mathbf{h}_1, \dots, \mathbf{h}_m, h_{m+1}, \dots, h_k) \\ & \lesssim 2^{o_k(l)} 2^{-\frac{cl}{30}} \left(\prod_{i \in \mathfrak{S}(\sigma)} \left\| \left(\sum_{j_i \in \mathbb{N}} |h_{j_i}^i|^4 \right)^{1/4} \right\|_{\ell^{4k}(\mathbb{Z}^k)} \right) \left(\prod_{i \in \mathfrak{S}(\sigma)^c} \|h_i\|_{\ell^{4k}(\mathbb{Z}^k)} \right). \end{aligned} \quad (7.91)$$

for every $\mathbf{h}_i = (h_{j_i}^i)_{j_i \in \mathbb{N}} \in \ell^{4k}(\mathbb{Z}^k; \ell^4(\mathbb{N}))$ for $i \in [m]$ and every $h_i \in \ell^{4k}(\mathbb{Z}^k)$ for $i \in [k] \setminus [m]$.

If (7.91) holds, then we can control $\Lambda_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{I, \sigma, 4, 4}(f_1, \dots, f_k)$ by

$$2^{o_k(l)} 2^{-\frac{cl}{30}} \left(\prod_{i \in \mathfrak{S}(\sigma)} \left\| \left(\sum_{j_i \in \mathbb{N}_{<J_i}} |\Delta G_{N_{j_i}^i, I}^{i,1,1}|^4 \right)^{1/4} \right\|_{\ell^{4k}(\mathbb{Z}^k)} \right) \left(\prod_{i \in \mathfrak{S}(\sigma)^c} \|G_{N_{-1}^i, I}^{i,1}\|_{\ell^{4k}(\mathbb{Z}^k)} \right)$$

so that $\sup_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k \in \mathbf{I}(\mathbb{D}_{l,s})} \Lambda_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{I, \sigma, 4, 4}(f_1, \dots, f_k)$ is controlled by

$$2^{o_k(l)} 2^{-\frac{cl}{30}} \left(\prod_{i \in \mathfrak{S}(\sigma)} \|V^4(G_{N, I}^{i,1} : N \in \mathbb{D}_{l,s})\|_{\ell^{4k}(\mathbb{Z}^k)} \right) \left(\prod_{i \in \mathfrak{S}(\sigma)^c} \|G_{N_{-1}^i, I}^{i,1}\|_{\ell^{4k}(\mathbb{Z}^k)} \right).$$

Inserting this bound into inequality (7.90) and applying Hölder's inequality, we sum over $I \in \mathcal{I}$. Since $\mathcal{I}$ has bounded overlap, we can use Lemma 7.67 and proceed as in Step 4. Additionally, invoking Lemma 7.39, we conclude that inequality (7.89) holds, as desired.

Step 9. It remains to establish (7.91). For this purpose, we first prove that

$$\begin{aligned} & \Theta_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, 2}(\mathbf{h}_1, \dots, \mathbf{h}_m, h_{m+1}, \dots, h_k) \\ & \lesssim 2^{o_k(l)} 2^{-\frac{cl}{10}} \left(\prod_{i \in \mathfrak{S}(\sigma)} \left\| \left(\sum_{j_i \in \mathbb{N}} |h_{j_i}^i|^2 \right)^{1/2} \right\|_{\ell^{2k}(\mathbb{Z}^k)} \right) \left(\prod_{i \in \mathfrak{S}(\sigma)^c} \|h_i\|_{\ell^{2k}(\mathbb{Z}^k)} \right) \end{aligned} \quad (7.92)$$

for every $\mathbf{h}_i = (h_{j_i}^i)_{j_i \in \mathbb{N}} \in \ell^{2k}(\mathbb{Z}^k; \ell^2(\mathbb{N}))$ for $i \in [m]$ and every $h_i \in \ell^{2k}(\mathbb{Z}^k)$ for $i \in [k] \setminus [m]$. To this end, we apply Khinchine's inequality (see, for instance, [41, Section C.5, p. 589]), which allows us to linearize the $\ell^2(\bar{\mathbb{X}}_\sigma)$ square function that appears in the definition of

$\Theta_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, 2}(\mathbf{h}_1, \dots, \mathbf{h}_m, h_{m+1}, \dots, h_k)$. Inserting the summation over $\bar{\mathbb{X}}_\sigma$ under the operator $\hat{A}_{2^u}$ and applying Weyl's inequality (7.78) with $w_1 = l - 1$ and $w_2 = \varepsilon(l - 1)$ for $\varepsilon = \frac{1}{10}$, we obtain (7.92).

Similarly, we prove that

$$\begin{aligned} & \Theta_{\mathbf{N}_{<J_1}^1, \dots, \mathbf{N}_{<J_k}^k}^{\sigma, 8}(\mathbf{h}_1, \dots, \mathbf{h}_m, h_{m+1}, \dots, h_k) \\ & \lesssim 2^{o_k(l)} \left(\prod_{i \in \mathfrak{S}(\sigma)} \left\| \left(\sum_{j_i \in \mathbb{N}} |h_{j_i}^i|^8 \right)^{1/8} \right\|_{\ell^{8k}(\mathbb{Z}^k)} \right) \left(\prod_{i \in \mathfrak{S}(\sigma)^c} \|h_i\|_{\ell^{8k}(\mathbb{Z}^k)} \right) \end{aligned} \quad (7.93)$$

for every $\mathbf{h}_i = (h_{j_i}^i)_{j_i \in \mathbb{N}} \in \ell^{8k}(\mathbb{Z}^k; \ell^8(\mathbb{N}))$ for $i \in [m]$ and every $h_i \in \ell^{8k}(\mathbb{Z}^k)$ for $i \in [k] \setminus [m]$. This inequality readily follows by applying Minkowski's integral inequality and the boundedness of the operator $\hat{A}_{2^u}$.

Then, noting that $\frac{1}{4} = \frac{\theta}{2} + \frac{1-\theta}{8}$ and $\frac{1}{4k} = \frac{\theta}{2k} + \frac{1-\theta}{8k}$ for $\theta = \frac{1}{3}$, we can interpolate (7.92) and (7.93), using a multilinear vector-valued interpolation similar to a Riesz–Thorin argument, to derive inequality (7.91). Specifically, we can adapt the proof of the multilinear complex interpolation theorem [42, Theorem 7.2.9, p. 514] to the multilinear vector-valued setting, following the approach in [41, Exercise 5.5.2, p. 339].

Now, the proof of Theorem 7.61 follows. $\square$

Remark 7.94. Finally, we note that if one wants to use the method from [62], then the argument in Step 6 would fail badly. In [62] the parameter u is exponential in l , while in our case it is linear; see (7.58). This is because in [62] the bilinear Weyl inequality was proved with the decay $\delta^c + \langle \log N \rangle^{-c\tilde{C}}$, which includes a logarithmic term that requires exponential $u = 2^{O(l)}$. However, thanks to the improvement in the Ionescu–Wainger theorem, our Weyl's inequality holds with $\delta^c + N^{-c}$, allowing us to have a linear dependence between l and u . This is crucial in deriving inequality (7.87), which follows from choosing a parameter r in condition (7.79). This choice renders the contribution of u in (7.88) negligible compared to the size of c arising in (7.78). If u were exponential in l , our approach in Step 6 would not be feasible at all.

REFERENCES

- [1] G.I. ARKHIPOV, V.N. CHUBARIKOV, A.A. KARATSUBA. Trigonometric sums in Number Theory and Analysis, de Gruyter Expositions in Mathematics **39**, Walter de Gruyter (2004).
- [2] T. AUSTIN. A proof of Walsh's convergence theorem using couplings. Int. Math. Res. Not. IMRN **15** (2015), pp. 6661–6674.
- [3] T. AUSTIN. On the norm convergence of non-conventional ergodic averages. Ergodic Theory Dynam. Systems **30** (2010), pp. 321–338.
- [4] A. BELLOW. Measure Theory Oberwolfach 1981. Proceedings of the Conference held at Oberwolfach, June 21–27, 1981. Lecture Notes in Mathematics **945**, editors D. Kölzow and D. Maharam-Stone. Springer-Verlag Berlin Heidelberg (1982). Section: Two problems submitted by A. Bellow, pp. 429–431.
- [5] V. BERGELSON. Weakly mixing PET. Ergodic Theory Dynam. Systems **7** (1987), no. 3, pp. 337–349.
- [6] V. BERGELSON. Ergodic Ramsey Theory – an update, Ergodic Theory of $\mathbb{Z}^d$ -actions (edited by M. Pollicott and K. Schmidt). London Math. Soc. Lecture Note Series **228** (1996), pp. 1–61.

- [7] V. BERGELSON. Combinatorial and Diophantine Applications of Ergodic Theory (with appendices by A. Leibman and by A. Quas and M. Wierdl). Handbook of Dynamical Systems, Vol. 1B, B. Hasselblatt and A. Katok, eds., Elsevier (2006), pp. 745–841.
- [8] V. BERGELSON, A. LEIBMAN. Polynomial extensions of van der Waerden’s and Szemerédi’s theorems. J. Amer. Math. Soc. **9** (1996), pp. 725–753.
- [9] V. BERGELSON, A. LEIBMAN. A nilpotent Roth theorem. Invent. Math. **147** (2002), pp. 429–470.
- [10] V. BERGELSON, R. MCCUTCHEON. An ergodic IP polynomial Szemerédi theorem. Memoirs Amer. Math. Soc. **146** (2000), viii+106pp.
- [11] G. BIRKHOFF. Proof of the ergodic theorem. Proc. Natl. Acad. Sci. USA **17** (1931), no. 12, pp. 656–660.
- [12] J. BOURGAIN. On the maximal ergodic theorem for certain subsets of the integers. Israel J. Math. **61** (1988), pp. 39–72.
- [13] J. BOURGAIN. On the pointwise ergodic theorem on L^p for arithmetic sets. Israel J. Math. **61** (1988), pp. 73–84.
- [14] J. BOURGAIN. Pointwise ergodic theorems for arithmetic sets. With an appendix by the author, H. Furstenberg, Y. Katznelson, and D.S. Ornstein. Inst. Hautes Etudes Sci. Publ. Math. **69** (1989), pp. 5–45.
- [15] J. BOURGAIN. Double recurrence and almost sure convergence. J. Reine Angew. Math. **404** (1990), pp. 140–161.
- [16] J. BOURGAIN, M.-C. CHANG. Nonlinear Roth type theorems in finite fields. Israel J. Math. **221** (2017), no. 2, pp. 853–867.
- [17] J. BOURGAIN, C. DEMETER, L. GUTH. Proof of the main conjecture in Vinogradov’s Mean Value Theorem for degrees higher than three. Ann. of Math. **184** (2016), no. 2, pp. 633–682.
- [18] Z. BUCZOLICH, R.D. MAULDIN. Divergent square averages. Ann. of Math. **171** (2010), no. 3, pp. 1479–1530.
- [19] A. CALDERÓN. Ergodic theory and translation invariant operators. Proc. Natl. Acad. Sci. USA **59** (1968), pp. 349–353.
- [20] M. CHRIST. Convolution, curvature, and combinatorics: a case study. Internat. Math. Res. Notices **19** (1998), pp. 1033–1048.
- [21] M. CHRIST, P. DURCIK, V. KOVAČ, J. ROOS. Pointwise convergence of certain continuous-time double ergodic averages. Ergodic Theory Dynam. Systems. **42** (2022), no. 7, pp. 2270–2280.
- [22] M. CHRIST, P. DURCIK, J. ROOS. Trilinear smoothing inequalities and a variant of the triangular Hilbert transform. Adv. Math. **390** (2021), article no. 107863.
- [23] Q. CHU, N. FRANTZIKINAKIS, B. HOST. Ergodic averages of commuting transformations with distinct degree polynomial iterates. Proc. Lond. Math. Soc. **102** (2011), no. 5, pp. 801–842.
- [24] C. DEMETER. Pointwise convergence of the ergodic bilinear Hilbert transform. Illinois J. Math. **51** (2007), no. 4, pp. 1123–1158.
- [25] J.-M. DERRIEN, E. LESIGNE. Un théorème ergodique polynomial ponctuel pour les endomorphismes exacts et les K-systèmes. Ann. Inst. H. Poincaré Probab. Statist. **32** (1996), no. 6, pp. 765–778.
- [26] C. DEMETER, M. LACEY, T. TAO, C. THIELE. Breaking the duality in the return times theorem. Duke Math. J. **143** (2008), no. 2, pp. 281–355.
- [27] Y. DO, R. OBERLIN, E.A. PALSSON. Variational bounds for a dyadic model of the bilinear Hilbert transform. Illinois J. Math. **57** (2013), no. 1, pp. 105–120.
- [28] P. DODOS, V. KANELLOPOULOS, T. KARAGEORGOS. L_p -regular sparse hypergraphs: box norms. Fundam. Math. **248** (2020), pp. 49–77.
- [29] S. DONOSO, W. SUN. Pointwise convergence of some multiple ergodic averages. Adv. Math. **330** (2018), pp. 946–996.
- [30] J. DUOANDIKOETXEA, J.L. RUBIO DE FRANCIA. Maximal and singular integral operators via Fourier transform estimates. Invent. Math. **84** (1986), no. 3, pp. 541–561.
- [31] N. FRANTZIKINAKIS. Multiple ergodic averages for three polynomials and applications. Trans. Amer. Math. Soc. **360** (2008), no. 10, pp. 5435–5475.
- [32] N. FRANTZIKINAKIS. Some open problems on multiple ergodic averages. Bull. Hellenic Math. Soc. **60** (2016), pp. 41–90.

- [33] N. FRANTZIKINAKIS, B. KRA. Polynomial averages converge to the product of integrals. *Israel J. Math.* **148** (2005), pp. 267–276.
- [34] N. FRANTZIKINAKIS, B. KRA. Ergodic averages for independent polynomials and applications. *J. London Math. Soc.* **74** (2006), no. 1, pp. 131–142.
- [35] N. FRANTZIKINAKIS, B. KUČA. Joint ergodicity for commuting transformations and applications to polynomial sequences. *Invent. Math.* **239** (2025), pp. 621–706.
- [36] H. FURSTENBERG. Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions. *J. Anal. Math.* **31** (1977), pp. 204–256.
- [37] H. FURSTENBERG. *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press (1981).
- [38] H. FURSTENBERG. Problems Session, Conference on Ergodic Theory and Applications. University of New Hampshire, Durham, NH, June 1982.
- [39] H. FURSTENBERG, Y. KATZNELSON. An ergodic Szemerédi theorem for commuting transformations. *J. Anal. Math.* **34** (1978), pp. 275–291.
- [40] H. FURSTENBERG, B. WEISS. A mean ergodic theorem for $\frac{1}{N} \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$. *Convergence in ergodic theory and probability* (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ. **5**, de Gruyter, Berlin (1996), pp. 193–227.
- [41] L. GRAFAKOS. *Classical Fourier Analysis*. Graduate Texts in Mathematics (GTM), volume 249. Springer New York, NY, (2014).
- [42] L. GRAFAKOS. *Modern Fourier Analysis*. Graduate Texts in Mathematics (GTM), volume 250. Springer New York, NY, (2014).
- [43] L. GRAFAKOS, T. TAO. Multilinear interpolation between adjoint operators. *J. Funct. Anal.* **199** (2003), no. 2, pp. 379–385.
- [44] B. GREEN, T. TAO. Linear equations in primes. *Ann. of Math.* **171** (2010), no. 3, pp. 1753–1850.
- [45] B. GREEN, T. TAO. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math.* **175** (2012), pp. 465–540.
- [46] P. GRESSMAN. L^p -improving properties of averages along polynomial curves and related integral estimates. *Math. Res. Lett.* **16** (2009), no. 16, pp. 971–989.
- [47] R. HAN, V. KOVAČ, M. LACEY, J. MADRID, F. YANG. Improving estimates for discrete polynomial averages. *J. Fourier Anal. Appl.* **16** (2020), article no. 42.
- [48] E. HOPF. The general temporally discrete Markoff process. *J. Rational Mech. Anal.* **3** (1954), pp. 13–45.
- [49] B. HOST. Ergodic seminorms for commuting transformations and applications. *Studia Math.* **195** (2009), pp. 31–49.
- [50] B. HOST, B. KRA. Convergence of polynomial ergodic averages. *Israel J. Math.* **149** (2005), pp. 1–19.
- [51] B. HOST, B. KRA. Non-conventional ergodic averages and nilmanifolds. *Ann. of Math.* **161** (2005), pp. 397–488.
- [52] M. HSU, F. YU-HSIANG LIN. A short proof on the boundedness of triangular Hilbert transform along curves. Available at [arXiv:2410.15791](https://arxiv.org/abs/2410.15791).
- [53] W. HUANG, S. SHAO, X. YE. Pointwise convergence of multiple ergodic averages and strictly ergodic models. *J. Anal. Math.* **139** (2019), pp. 265–305.
- [54] A.D. IONESCU, Á. MAGYAR, M. MIREK, T.Z. SZAREK. Polynomial averages and pointwise ergodic theorems on nilpotent groups. *Invent. Math.* **231** (2023), pp. 1023–1140.
- [55] A.D. IONESCU, A. MAGYAR, E.M. STEIN, S. WAINGER. Discrete Radon transforms and applications to ergodic theory. *Acta Math.* **198** (2007), pp. 231–298.
- [56] A.D. IONESCU, S. WAINGER. L^p boundedness of discrete singular Radon transforms. *J. Amer. Math. Soc.* **19** (2005), no. 2, pp. 357–383.
- [57] R.L. JONES, R. KAUFMAN, J. ROSENBLATT, M. WIERDL. Oscillation in ergodic theory. *Ergodic Theory Dynam. Systems* **18** (1998), no. 4, pp. 889–935.
- [58] R.L. JONES, A. SEEGER, J. WRIGHT. Strong variational and jump inequalities in harmonic analysis. *Trans. Amer. Math. Soc.* **360** (2008), no. 12, pp. 6711–6742.
- [59] A. DEL JUNCO, J. ROSENBLATT. Counterexamples in ergodic theory and number theory. *Math. Ann.* **245** (1979), pp. 185–197.

- [60] D. KOSZ. Sharp constants in inequalities admitting the Calderón transference principle. *Ergodic Theory Dynam. Systems* **44** (2024), pp. 1597–1608.
- [61] B. KRAUSE, M. MIREK, S. PELUSE, J. WRIGHT. Polynomial progressions in topological fields. *Forum Math. Sigma*. **12** (2024), article no. e106.
- [62] B. KRAUSE, M. MIREK, T. TAO. Pointwise ergodic theorems for non-conventional bilinear polynomial averages. *Ann. of Math.* **195** (2022), no. 3, pp. 997–1109.
- [63] N. KRAVITZ, B. KUČA, J. LENG. Quantitative concatenation for polynomial box norms. Available at [arXiv:2407.08636](https://arxiv.org/abs/2407.08636).
- [64] B. KUČA. Multidimensional polynomial patterns over finite fields: bounds, counting estimates and Gowers norm control. *Adv. Math.* **448** (2024), article no. 109700.
- [65] M. LACEY. The bilinear maximal functions map into L^p for $2/3 < p \leq 1$. *Ann. of Math.* **151** (2000), no. 1, pp. 35–57.
- [66] P. LAVICTOIRE. Universally L^1 -Bad Arithmetic Sequences. *J. Anal. Math.* **113** (2011), no. 1, pp. 241–263.
- [67] A. LEIBMAN. Convergence of multiple ergodic averages along polynomials of several variables. *Israel J. Math.* **146** (2005), pp. 303–315.
- [68] A. LEIBMAN. Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold. *Ergodic Theory Dynam. Systems* **25** (2005), no. 1, pp. 201–213.
- [69] D. LÉPINGLE. La variation d’ordre p des semi-martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*. **36** (1976), no. 4, pp. 295–316.
- [70] Á. MAGYAR. Private communication (2021).
- [71] Á. MAGYAR, E.M. STEIN, S. WAINGER. Discrete analogues in harmonic analysis: spherical averages. *Ann. of Math.* **155** (2002), pp. 189–208.
- [72] Á. MAGYAR, E.M. STEIN, S. WAINGER. Maximal operators associated to discrete subgroups of nilpotent Lie groups. *J. Anal. Math.* **101** (2007), pp. 257–312.
- [73] M. MIREK. $\ell^p(\mathbb{Z}^d)$ -estimates for discrete Radon transform: square function estimates. *Anal. PDE* **11** (2018), no. 3, pp. 583–608.
- [74] M. MIREK, E.M. STEIN, P. ZORIN-KRANICH. A bootstrapping approach to jump inequalities and their applications. *Anal. PDE* **13** (2020), no. 2, pp. 527–558.
- [75] M. MIREK, E.M. STEIN, P. ZORIN-KRANICH. Jump inequalities for translation-invariant operators of Radon type on $\mathbb{Z}^d$. *Adv. Math.* **365** (2020), article no. 107065.
- [76] M. MIREK, E.M. STEIN, P. ZORIN-KRANICH. Jump inequalities via real interpolation. *Math. Ann.* **376** (2020), no. 1–2, pp. 797–819.
- [77] C. MUSCALU, W. SCHLAG. *Classical and multilinear harmonic analysis. Vol. II. Cambridge Studies in Advanced Mathematics*, 138. Cambridge University Press, Cambridge, 2013.
- [78] J. VON NEUMANN. Proof of the quasi-ergodic hypothesis. *Proc. Natl. Acad. Sci. USA* **18** (1932), pp. 70–82.
- [79] S. PELUSE. On the polynomial Szemerédi theorem in finite fields. *Duke Math. J.* **168** (2019), no. 5, pp. 749–774.
- [80] S. PELUSE. Bounds for sets with no polynomial progressions. *Forum Math. Pi.* (2020), article no. e16.
- [81] S. PELUSE, S. PRENDIVILLE. A polylogarithmic bound in the nonlinear Roth theorem. *Int. Math. Res. Not.* **1** (2022), pp. 5658–5684.
- [82] S. PELUSE, S. PRENDIVILLE. Quantitative bounds in the non-linear Roth theorem. *Invent. Math.* **238** (2024), pp. 865–903.
- [83] S. PELUSE, S. PRENDIVILLE, X. SHAO. Bounds in a popular multidimensional nonlinear Roth theorem. *J. Lond. Math. Soc.* **110** (2024), article no. e70019.
- [84] L. PIERCE. On superorthogonality. With an appendix by Emmanuel Kowalski. *J. Geom. Anal.* **31** (2020), pp. 7096–7183.
- [85] J.L. RUBIO DE FRANCIA. A Littlewood–Paley inequality for arbitrary intervals. *Rev. Mat. Iberoam.* **1** (1985), pp. 1–14.
- [86] W. RUDIN. *Fourier analysis on groups. Interscience Tracts in Pure and Applied Mathematics*, No. 12. Interscience Publishers, New York–London 1962.

- [87] E.M. STEIN. Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals. Princeton University Press (1993).
- [88] B. STOVALL. L^p improving multilinear Radon-like transforms. *Rev. Mat. Iberoam.* **27** (2011), no. 3, pp. 1059–1085.
- [89] E. SZEMERÉDI. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), pp. 199–245.
- [90] T. TAO. Norm convergence of multiple ergodic averages for commuting transformations. *Ergodic Theory Dynam. Systems* **28** (2008), pp. 657–688.
- [91] T. TAO. Higher order Fourier analysis. *Graduate Studies in Mathematics* **142**, American Mathematical Society (2012).
- [92] T. TAO. The Ionescu–Wainger multiplier theorem and the adeles. *Mathematika* **67** (2021), no. 3, pp. 647–677.
- [93] T. TAO, V. VU. Additive Combinatorics. Cambridge University Press (2006).
- [94] T. TAO, T. ZIEGLER. Concatenation theorems for anti-Gowers-uniform functions and Host-Kra characteristic factors. *Discrete Anal.* (2016), article no. 13.
- [95] M. WALSH. Norm convergence of nilpotent ergodic averages. *Ann. of Math.* **175** (2012), no. 3, pp. 1667–1688.
- [96] T. WOOLEY. Vinogradov’s mean value theorem via efficient congruencing. *Ann. of Math.* **175** (2012), pp. 1575–1627.
- [97] T. WOOLEY. The cubic case of the main conjecture in Vinogradov’s mean value theorem. *Adv. Math.* **294** (2016), pp. 532–561.
- [98] T. ZIEGLER. Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20** (2007), pp. 53–97.
- [99] P. ZORIN-KRANICH. Norm convergence of multiple ergodic averages on amenable groups. *J. Anal. Math.* **130** (2016), pp. 219–241.

FACULTY OF PURE AND APPLIED MATHEMATICS, WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY,
WYBRZEŻE STANISŁAWA WYSPIAŃSKIEGO 27, 50-370 WROCLAW, POLAND

Email address: `dariusz.kosz@pwr.edu.pl`

DEPARTMENT OF MATHEMATICS, RUTGERS UNIVERSITY, PISCATAWAY, NJ 08854-8019, USA & INSTYTUT MATEMATYCZNY, UNIwersYTET WROCLAWSKI, PLAC GRUNWALDZKI 2/4, 50-384 WROCLAW, POLAND

Email address: `mariusz.mirek@rutgers.edu`

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY, 450 JANE STANFORD WAY, BUILDING 380, STANFORD, CA 94305, USA

Email address: `speluse@stanford.edu`

MINISTRY OF EDUCATION KEY LABORATORY OF NSLSCS, SCHOOL OF MATHEMATICAL SCIENCES, NANJING NORMAL UNIVERSITY, NANJING 210023, PEOPLE’S REPUBLIC OF CHINA

Email address: `wrh@njnu.edu.cn`

SCHOOL OF MATHEMATICS AND MAXWELL INSTITUTE FOR MATHEMATICAL SCIENCE, JAMES CLERK MAXWELL BUILDING, THE KING’S BUILDINGS, PETER GUTHRIE TAIT ROAD, EDINBURGH, EH9 3FD, UK

Email address: `J.R.Wright@ed.ac.uk`