Discrete analogues of maximally modulated singular integrals of Stein-Wainger type: 2(Zn) bounds

DISCRETE ANALOGUES OF MAXIMALLY MODULATED SINGULAR INTEGRALS OF STEIN–WAINGER TYPE: $\ell^2(\mathbb{Z}^n)$ BOUNDS

JORIS ROOS

Abstract. Consider the maximal operator

$$\mathscr{C}f(x)=\sup_{\lambda\in\mathbb{R}}\Big|\sum_{y\in\mathbb{Z}^n\setminus\{0\}}f(x-y)e(\lambda|y|^{2d})K(y)\Big|,\quad (x\in\mathbb{Z}^n),$$

where d is a positive integer, K an appropriate Calderón–Zygmund kernel and $n\geq 1$. This is a discrete analogue of a real–variable operator studied by Stein and Wainger. The nonlinearity of the phase introduces a variety of new difficulties that are not present in the real–variable setting. We prove $\ell^2(\mathbb{Z}^n)$–bounds for $\mathscr{C}$. Our arguments are inspired by the recent seminal paper of Krause covering the case $n=1$, earlier key partial progress by Krause and Lacey, and Bourgain's classical maximal multi–frequency lemma in combination with variation–norm estimates from recent joint work of the author with S. Guo and P.-L. Yung.

1. Introduction

Let d and n be positive integers and K a homogeneous Calderón-Zygmund kernel on $\mathbb{R}^n$, taking the form

$$K(x)=\text{p.v.}\frac{\Omega(x)}{|x|^n},$$

where Ω is a smooth function on $\mathbb{R}^n\setminus\{0\}$ that is homogeneous of degree zero. We also assume that $\int_{\mathbb{S}^{n-1}}\Omega(x)d\sigma(x)=0$, where σ denotes the surface measure on the sphere $\mathbb{S}^{n-1}\subset\mathbb{R}^n$. Consider the following operator acting on functions $f:\mathbb{Z}^n\to\mathbb{C}$,

$$\mathscr{C}f(x)=\sup_{\lambda\in\mathbb{R}}\Big|\sum_{y\in\mathbb{Z}^n\setminus\{0\}}f(x-y)e(\lambda|y|^{2d})K(y)\Big|,\quad (x\in\mathbb{Z}^n),\tag{1.1}$$

where $|y|=(y_1^2+\cdots+y_n^2)^{1/2}$ and $e(x)=e^{2\pi ix}$. This is a discrete analogue of a maximal operator studied by Stein and Wainger [12]. We also refer to $\mathscr{C}$ as a discrete Carleson operator. This is motivated by the formal resemblance to Carleson's operator given by the presence of a supremum over the modulation parameters λ. However, we stress that the (substantial) difficulties encountered in the analysis of the present operator are of a fundamentally different nature than those encountered in the analysis of Carleson's operator. The nonlinearity of the phase causes a number of new challenges arising from a curious fusion of number–theoretic and analytic phenomena which are not present in the real–variable case. We refer to [7], [8] for further discussion of the history motivating the study of the present operator and to [9], [10] for background and recent progress on some other related discrete analogues in harmonic analysis. The following is our main result.

Theorem 1.1. *There is a constant $C\in(0,\infty)$ such that*

$$\|\mathscr{C}f\|_{\ell^2(\mathbb{Z}^n)}\leq C\|f\|_{\ell^2(\mathbb{Z}^n)}.\tag{1.2}$$

The constant C only depends on d,n and K.

Date: July 2, 2019.
2010 *Mathematics Subject Classification.* 42B15, 42B20, 42B25.

The case $n = 1$ was one of the main results in Krause's seminal paper [7], where notably he also gives $\ell^p(\mathbb{Z})$ bounds for a range of p including $[2, \infty)$. Important partial progress on the case $n = d = 1$, $p = 2$ was made earlier by Krause and Lacey [8]. They considered a maximal operator with the supremum over $\lambda \in \mathbb{R}$ replaced by a restricted supremum over $\lambda \in \Lambda$, imposing certain arithmetic as well as analytic conditions on the set Λ. The case $n = d = 1$, $p = 2$ was also the subject of a question posed by Lillian Pierce during a 2015 workshop at the American Institute of Mathematics.
The purpose of this paper is to give a somewhat different proof for $n = 1$ in the special case of ℓ^2. At the same time we demonstrate that it generalizes to higher dimensions $n > 1$ without significant additional effort. The specific choice of the phase in (1.1) and the assumptions made on the kernel K are imposed primarily in favor of simplicity. The specific choice of phase function enables a fairly direct application of the variation–norm estimates from [4], while the choice of the even integer exponent $2d$ is made in order not to introduce certain artificial additional number–theoretic issues. Various interesting extensions for other phase functions could be topics for further investigation.

A key ingredient to our proof is given by appropriate variation–norm estimates for the associated real–variable operators acting on functions $f : \mathbb{R}^n \to \mathbb{C}$,

$$\mathcal{H}^{(\lambda)} f(x) = \int_{\mathbb{R}^n} f(x-y)e(\lambda|y|^{2d})K(y)dy, \quad (x \in \mathbb{R}^n).$$

The following variation–norm estimate is proven in [4, Theorem 1.1]:

$$\|V^r\{\mathcal{H}^{(\lambda)} f \,:\, \lambda > 0\}\|_{L^2(\mathbb{R}^n)} \leq C_{d,n}(r-2)^{-1}\|f\|_{L^2(\mathbb{R}^n)}, \tag{1.3}$$

valid for all $r \in (2, 3)$. Here $C_{d,n} \in (0, \infty)$ is a constant only depending on d, n and the r–variation of a family $\mathbf{a} = (a_\lambda)_{\lambda \in \mathcal{J}} \subset \mathbb{C}$ (with $\mathcal{J} \subset \mathbb{R}$) of complex numbers is defined by

$$V^r(\mathbf{a}) = \sup_{\lambda_0 < \cdots < \lambda_N} \Big(\sum_{j=1}^N |a_{\lambda_j} - a_{\lambda_{j-1}}|^r \Big)^{1/r},$$

where the supremum is taken over all finite subsets $\{\lambda_0 < \cdots < \lambda_N\} \subset \mathcal{J}$. The dependence on r is not stated explicitly in [4], but can be extracted from the proof (also see the comment following Corollary 9.4 there for more details in the case $n = 1$).

In the proof of (1.2) we will make use of a slight variant of the estimate (1.3). The way in which the precise estimate that we require differs from (1.3) should be regarded as a harmless technicality (the precise estimate we shall need is (7.4) below). While that estimate does not seem to be a formal consequence of (1.3), its validity is easily demonstrated by following certain preliminary steps in the proof of (1.3) in [4] (see §7 for details). On a related note, in [7], the variation–norm estimates from [4] form a key ingredient in the proof of certain pointwise ergodic theorems, which we do not discuss here.

Variation–norm estimates enter the proof of (1.2) through a variant of the well–known maximal multi–frequency lemma due to Bourgain [2], which we state as Theorem 2.2 below. Dependence on Bourgain's lemma is also the main reason for the limitation of the present result to $\ell^2(\mathbb{Z}^n)$. Indeed, all other components of the argument generalize readily to $\ell^p(\mathbb{Z}^n)$ with $p \in (1, \infty)$. More precisely, the Propositions 3.1, 3.2, 3.3 below hold with ℓ^2 replaced by ℓ^p for all $p \in (1, \infty)$ (by standard interpolation arguments; with the respective decay rates γ additionally depending on p). In contrast to this, it seems that the ℓ^p analogue of Proposition 3.4 below would necessarily feature an exponential loss in s (this is a well–known limitation of the multi–frequency lemma). Thus, the most we can reasonably expect from this particular route of attack are ℓ^p–bounds for p in a sufficiently small neighborhood of 2. Not to distract the reader we do not pursue this in this article. To obtain ℓ^p–bounds for large $p > 2$, it is more opportune to adapt the subtle square–function techniques of Ionescu and Wainger [5]. This is also the route taken in [7]. Another interesting open problem is to determine whether $\mathscr{C}$ (even for $n = d = 1$) is also bounded on ℓ^p for all small $p > 1$.

Structure of the paper. In §2 we introduce several tools used throughout the proof. The most substantial of these are certain known exponential sum estimates from [11] and a variant of Bourgain's multi–frequency lemma [2].
In §3 we give the proof of Theorem 1.1. The basic strategy follows that of [8], splitting the multiplier into a number–theoretic approximate ('major arcs') and an error term ('minor arcs'). This approach goes back to Bourgain [2] and can be viewed as an instance of the Hardy–Littlewood circle method. The proof involves four distinct components, which (with a slight abuse of terminology) we refer to as 'Minor arcs I/II' and 'Major arcs I/II'. The main differences of this paper compared to the previous works [7], [8] appear in the major arcs.
In §4 ('Minor arcs I') we perform a preliminary TT^* argument to reduce the set of modulation parameters λ. This was one of the key novelties from [7], which we reproduce in a slightly simplified form.
In §5 ('Minor arcs II') we estimate the error terms from a number–theoretic approximation of the multipliers. This is a standard argument using the fundamental theorem of calculus (which only becomes possible after the crucial reduction from §4). This is already featured in [8].
In §6 ('Major arcs I') we handle the number–theoretic component of the main contribution to the multiplier by exploiting exponential sum estimates. A somewhat unanticipated dichotomy appears here between the cases $d = 1$ and $d \geq 2$.
In §7 ('Major arcs II') we handle the analytic component of the main contribution. This is where we reduce to Bourgain's multi–frequency lemma and also give the proof of the required variation–norm estimate by using results from [4]. This is the only part of the argument that distinguishes between $n = 1$ and $n \geq 2$.
Finally, §8 contains the proof of our version of Bourgain's lemma. This is quite standard, following Bourgain's original argument [2].

Acknowledgement. The author is grateful to Ben Krause for many helpful discussions and for introducing him to discrete harmonic analysis and in particular to this topic.

2. Preliminaries

We write $A \lesssim B$ to denote existence of a constant C such that $A \leq C \cdot B$, where the admissible dependencies of the constant C will be specified, or clear from context. Throughout the text we allow constants to depend on the ambient dimension n, the degree d and the kernel K. Similarly, $A \approx B$ signifies that both, $A \lesssim B$ and $B \lesssim A$. The notation $A = B + O(X)$ stands for $|A - B| \lesssim X$.

2.1. Fourier transforms on $\mathbb{Z}^n$, $\mathbb{T}^n$, $\mathbb{R}^n$ and transference.

For Fourier transforms of functions $f : \mathbb{Z}^n \to \mathbb{C}$, $g : \mathbb{T}^n \to \mathbb{C}$ we use the notations

$$\widehat{f}(\xi) = \mathcal{F}_{\mathbb{Z}^n} f(\xi) = \sum_{x \in \mathbb{Z}^n} e(-\xi \cdot x) f(x) \quad \text{and}$$

$$\mathcal{F}^{-1}[g](x) = \mathcal{F}_{\mathbb{Z}^n}^{-1}[g](x) = \int_{\mathbb{T}^n} e(\xi \cdot x) \widehat{g}(\xi) d\xi.$$

Here $\mathbb{T}^n = (\mathbb{R}/\mathbb{Z})^n$. A function $g : \mathbb{R}^n \to \mathbb{C}$ that satisfies $g(x + z) = g(x)$ for all $z \in \mathbb{Z}^n$ will be called *periodic* and be silently identified with the corresponding function on $\mathbb{T}^n$.
For a function $h : \mathbb{R}^n \to \mathbb{C}$ we write

$$\widehat{h}(\xi) = \mathcal{F}_{\mathbb{R}^n} h(\xi) = \int_{\mathbb{R}^n} e(-\xi \cdot x) h(x) dx \quad \text{and}$$

$$\mathcal{F}^{-1}[h](x) = \mathcal{F}_{\mathbb{R}^n}^{-1}[h](x) = \widehat{h}(-x).$$

In particular, Fourier transforms on $\mathbb{Z}^n$ or $\mathbb{R}^n$ will be denoted by the same symbols unless the distinction is not clear from context, or is emphasized for other reasons.

For a bounded periodic function $m : \mathbb{R}^n \to \mathbb{C}$ we denote by $m(\mathrm{D})$, the associated Fourier multiplier acting on $\mathbb{Z}^n$, defined as

$$m(\mathrm{D})f(x) = \mathcal{F}_{\mathbb{Z}^n}^{-1}[m \cdot \mathcal{F}_{\mathbb{Z}^n} f](x), \quad (x \in \mathbb{Z}^n).$$

We slightly abuse notation and also write $m(\mathrm{D})$ for the Fourier multiplier acting on $\mathbb{R}^n$, defined as

$$m(\mathrm{D})h(x) = \mathcal{F}_{\mathbb{R}^n}^{-1}[m \cdot \mathcal{F}_{\mathbb{R}^n} f](x), \quad (x \in \mathbb{R}^n).$$

Let $(m_\lambda)_{\lambda \in \Lambda}$ be a family of bounded functions supported on a fundamental domain of $\mathbb{T}^n$ (such as a translate of the unit cube $[0,1)^n$) and denote their periodizations by

$$\mathbf{m}_\lambda(\xi) = \sum_{z \in \mathbb{Z}^n} m_\lambda(\xi + z), \quad (\xi \in \mathbb{R}^n).$$

We will make use of the following transference principle.

Lemma 2.1. *Suppose that for some constant $A > 0$,*

$$\| \sup_{\lambda \in \Lambda} |m_\lambda(\mathrm{D}) f| \|_{L^2(\mathbb{R}^n)} \le A \|f\|_{L^2(\mathbb{R}^n)}.$$

Then

$$\| \sup_{\lambda \in \Lambda} |\mathbf{m}_\lambda(\mathrm{D}) f| \|_{\ell^2(\mathbb{Z}^n)} \lesssim_n A \|f\|_{\ell^2(\mathbb{Z}^n)},$$

where the implicit constant only depends on n.

The proof of this fact is standard (see [2, Lemma 4.4]; there in the case $n = 1$, but the argument also works also for $n \ge 2$).

2.2. **Some notation and TT^*.** For a function $\mathcal{K} : \mathbb{Z}^n \times \mathbb{Z}^n \to \mathbb{C}$ we denote by $T_{\mathcal{K}}$ the operator defined formally by

$$T_{\mathcal{K}} f(x) = \sum_{y \in \mathbb{Z}^n} \mathcal{K}(x,y) f(y). \tag{2.1}$$

Then the operator $T_{\mathcal{K}} T_{\mathcal{K}}^*$ is formally given by $T_{\mathcal{K}} T_{\mathcal{K}}^* = T_{\mathcal{K}^\sharp}$, where the kernel $\mathcal{K}^\sharp$ is

$$\mathcal{K}^\sharp(x,y) = \sum_{z \in \mathbb{Z}^n} \mathcal{K}(x,z) \overline{\mathcal{K}(y,z)}.$$

2.3. **Kernel decomposition.** Let ψ be a smooth function on $\mathbb{R}^n$ supported in $\{1/2 \le |x| \le 2\}$ with $0 \le \psi \le 1$ and $\sum_{j \in \mathbb{Z}} \psi_j(x) = 1$ for every $x \ne 0$, where $\psi_j(x) = \psi(2^{-j}x)$. Decompose

$$K(x) = \sum_{j \ge 1} K_j(x),$$

with $K_1(x) = \sum_{j \le 1} \psi_j(x) K(x)$ and $K_j(x) = \psi_j(x) K(x)$ for $j \ge 2$. Then for all $j \ge 1$ and all $x \in \mathbb{R}^n \setminus 0$,

$$|K_j(x)| \lesssim 2^{-jn}, \quad |\nabla K_j(x)| \lesssim 2^{-j(n+1)}, \quad \operatorname{supp} K_j \subset \{x : |x| \le 2^{j+1}\}. \tag{2.2}$$

2.4. **Variation–norm estimates imply multi–frequency L^2 estimates.** Here we record a variant of Bourgain's multi–frequency lemma [2, Lemma 4.13]. To expose the general principle, we adopt a slightly more general perspective than required.

The setting is as follows. Fix a real number $\sigma > 0$. We are given a family of bounded functions $(m_\lambda)_{\lambda\in\Lambda}$ on $\mathbb{R}^n$, indexed by a subset Λ of the real numbers such that

$$\operatorname{supp} m_\lambda \subset B_{\sigma/2} \quad \text{for every } \lambda \in \Lambda, \text{ and} \tag{2.3}$$

$$\sup_{\lambda\in\Lambda} \|m_\lambda\|_{L^\infty(\mathbb{R}^n)} \leq c_0 \tag{2.4}$$

for some constant $c_0 \in (0,\infty)$. Here $B_r \subset \mathbb{R}^n$ denotes the open ball of radius r centered at the origin. We also fix a large integer $M \geq 2$ and a vector $\theta = (\theta_1, \dots, \theta_M) \in (\mathbb{R}^n)^M$ with the separation condition

$$|\theta_j - \theta_k| \geq \sigma \quad \text{for every } 1 \leq j < k \leq M. \tag{2.5}$$

The key assumption is that the Fourier multipliers $m_\lambda(\mathrm{D})$ obey the following variation–norm estimate:

$$\|V^r\{m_\lambda(\mathrm{D})f \,:\, \lambda \in \Lambda\}\|_{L^2(\mathbb{R}^n)} \leq c_0(r-2)^{-\gamma}\|f\|_{L^2(\mathbb{R}^n)} \tag{2.6}$$

for some $\gamma > 0$ and all $r \in (2,3)$. Observe that (2.4) and (2.6) imply

$$\|\sup_{\lambda\in\Lambda} |m_\lambda(\mathrm{D})f|\|_{L^2(\mathbb{R}^n)} \leq 2c_0\|f\|_{L^2(\mathbb{R}^n)}. \tag{2.7}$$

We define the maximal function $\mathfrak{M}^\theta F$, acting on functions $F : \mathbb{R}^n \to \mathbb{C}^M$ by

$$\mathfrak{M}^\theta F(x) = \sup_{\lambda\in\Lambda} \Big| \sum_{j=1}^{M} e(\theta_j \cdot x)[m_\lambda(\mathrm{D})F_j](x) \Big|, \quad (x \in \mathbb{R}^n).$$

For $a \in \mathbb{C}^M$ we fix the Euclidean norm $|a| = (|a_1|^2 + \cdots + |a_M|^2)^{1/2}$.

Theorem 2.2. *Assume that θ and $(m_\lambda)_{\lambda\in\Lambda}$ are such that* (2.3), (2.4), (2.5), (2.6) *hold. Then*

$$\|\mathfrak{M}^\theta F\|_{L^2(\mathbb{R}^n)} \leq C_n \cdot c_0(\log M)^{1+\gamma}\|F\|_{L^2(\mathbb{R}^n;\mathbb{C}^M)}. \tag{2.8}$$

Here C_n is a constant only depending on n.

This is a variant of [2, Lemma 4.13] (also see [8, Theorem 3.3, Theorem 3.5], [4, Theorem 9.1]). The proof is in essence identical to that of Bourgain. For convenience and future reference, we provide the details in §8.

Remarks. 1. By scaling invariance it suffices to consider the case $\sigma = 1$. We keep the parameter σ only for later notational convenience.

2. Upon replacing the assumption (2.6) by a suitable jump–norm endpoint inequality (such as in [4, Theorem 1.1]), the constant $(\log M)^{1+\gamma}$ can be improved to $\log M$. For $n \geq 2$, such jump–norm inequalities are available for the family of multipliers that we have in mind, but they seem to be open for $n = 1$. Since the power of the logarithm is of no consequence in our arguments here, we work with the slightly weaker assumption (2.6) which we can verify for our multipliers in all dimensions $n \geq 1$ using the results from [4].

2.5. **Exponential sum estimates.** Given integers $x_1, x_2, \dots, x_m$ at least one of which is non–zero we often use the notation $(x_1, x_2, \dots, x_m)$ for the greatest common divisor of $x_1, \dots, x_m$. It will be clear from context whether $(x_1, \dots, x_m)$ refers to the greatest common divisor, or the vector of the integers $x_1, \dots, x_m$. For a positive integer q we use the notation

$$[q] = \mathbb{Z} \cap [0, q).$$

The letter q always denotes a positive integer throughout the text. By a *reduced rational* we mean a fraction $\frac{a}{q}$ with $a \in \mathbb{Z}$ and $(a,q)=1$. For a positive integer $D \geq 2$, $x \in \mathbb{R}^n$ and real coefficients $\xi = (\xi_\alpha)_{1 \leq |\alpha| \leq D}$ we define the polynomial

$$P(\xi; x) = \sum_{1 \leq |\alpha| \leq D} \xi_\alpha x^\alpha,$$

where $\alpha \in \mathbb{N}_0^n$ denotes a multiindex. A key ingredient will be the following exponential sum estimate, due to Stein and Wainger [11, Proposition 3].

Proposition 2.3. *Let φ be a smooth function on $\mathbb{R}^n$ such that $|\varphi(x)| \leq 1$ and $|\nabla \varphi(x)| \leq (1+|x|)^{-1}$ for all $x \in \mathbb{R}^n$. For every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon, n, D) > 0$ such that the following holds: for every $N \geq 1$ and every ξ with the property that for some α_0 with $1 \leq |\alpha_0| \leq D$ there exists a reduced rational $\frac{a}{q} \in \mathbb{Q}$ such that*

$$|\xi_{\alpha_0} - \tfrac{a}{q}| \leq \tfrac{1}{q^2} \quad \text{and} \quad N^\varepsilon \leq q \leq N^{|\alpha_0| - \varepsilon},$$

we have

$$\Big| \sum_{x \in \mathbb{Z}^n, |x| \leq N} e(P(\xi; x)) \varphi(x) \Big| \leq C N^{n-\delta}.$$

The constant C may only depend on ε, n, D.

2.6. **Approximation of the multipliers.** For $j \geq 1$, $\lambda \in \mathbb{R}$ and $\xi \in \mathbb{R}^n$ we define the multipliers

$$m_{j,\lambda}(\xi) = \sum_{y \in \mathbb{Z}^n} e(\lambda |y|^{2d} + \xi \cdot y) K_j(y). \tag{2.9}$$

This defines a periodic function both in λ and ξ. Following Bourgain [2], the starting point for our arguments is an appropriate approximation for the value of $m_{j,\lambda}(\xi)$ when ξ and λ are close to rationals with small denominator. To formulate the result, we define the exponential sums

$$S(\tfrac{a}{q}, \tfrac{\mathbf{b}}{q}) = \frac{1}{q^n} \sum_{r \in [q]^n} e(\tfrac{a}{q} |r|^{2d} + \tfrac{\mathbf{b}}{q} \cdot r) \tag{2.10}$$

for rationals $\frac{a}{q} \in \mathbb{Q}$, $\frac{\mathbf{b}}{q} \in \mathbb{Q}^n$ with $(a, \mathbf{b}, q) = 1$ (note that this condition makes $S(\frac{a}{q}, \frac{\mathbf{b}}{q})$ well–defined).

By a well–known exponential sum estimate (see [1, Theorem 2.6]) we have for every $\varepsilon > 0$,

$$|S(\tfrac{a}{q}, \tfrac{\mathbf{b}}{q})| \lesssim_{\varepsilon, d, n} q^{-1/(2d) + \varepsilon}.$$

A less precise version of this estimate can also be obtained from Proposition 2.3. While we will not, strictly speaking, make use of this particular estimate, it does provide useful intuition. The following observation will be crucial at various points in the proof of Theorem 1.1.

Lemma 2.4. *Suppose that $\frac{a}{q} \in \mathbb{Q}$, $\frac{\mathbf{b}}{q} \in \mathbb{Q}^n$, $(a, \mathbf{b}, q) = 1$ and $(a, q) > 1$. Then $S(\frac{a}{q}, \frac{\mathbf{b}}{q}) = 0$.*

We postpone the standard proof of this to the end of this section. Next, we define the real–variable versions of the multipliers $m_{j,\lambda}(\xi)$ by

$$\Phi_{j,\lambda}(\xi) = \int_{\mathbb{R}^n} e(\lambda |y|^{2d} + \xi \cdot y) K_j(y) dy. \tag{2.11}$$

At this point we record the following standard oscillatory integral decay estimate in the spirit of van der Corput's lemma:

$$|\Phi_{j,\lambda}(\xi)| \lesssim (1 + 2^{2dj} |\lambda| + 2^j |\xi|)^{-\frac{1}{2d}}. \tag{2.12}$$

For the proof we refer to [12, Proposition 2.1]. This estimate does not enter in the proof of the approximation result in this section, but will be important later on. Our basic approximation result for the multipliers $m_{j,\lambda}(\xi)$ now reads as follows.

Lemma 2.5. *Let j, q be positive integers with $q \le 2^{j-2}$. Let $a \in \mathbb{Z}, \mathbf{b} \in \mathbb{Z}^n$ with $(a, \mathbf{b}, q) = 1$. Further, assume that $\lambda \in \mathbb{R}$, $\xi \in \mathbb{R}^n$ are such that*

$$|\lambda - \tfrac{a}{q}| \le \delta 2^{-(2d-1)j} \quad \text{and} \quad |\xi - \tfrac{\mathbf{b}}{q}| \le \delta, \tag{2.13}$$

where $\delta \in (2^{-j}, 1)$. Then

$$m_{j,\lambda}(\xi) = S(\tfrac{a}{q}, \tfrac{\mathbf{b}}{q})\Phi_{j,\lambda-\frac{a}{q}}(\xi - \tfrac{\mathbf{b}}{q}) + O(q\delta), \tag{2.14}$$

where the implicit constant depends only on d, n, K.

The proof is similar to that of the corresponding statement in [2] (see Lemma 5.12 there).

Proof of Lemma 2.5. Writing $y = uq + r$ with $u \in \mathbb{Z}^n$, $r \in [q]^n$, we can express $m_{j,\lambda}(\xi)$ as

$$q^{-n} \sum_{r \in [q]^n} e(\tfrac{a}{q}|r|^{2d} + \tfrac{\mathbf{b}}{q} \cdot r) I_{q,r}(\lambda - \tfrac{a}{q}, \xi - \tfrac{\mathbf{b}}{q}),$$

where

$$I_{q,r}(\nu, \eta) = q^n \sum_{u \in \mathbb{Z}^n} e(\nu|uq + r|^{2d} + \eta \cdot (uq + r)) K_j(uq + r).$$

It suffices to show that for every $r \in [q]^n$ and every $(\nu, \eta) \in \mathbb{R} \times \mathbb{R}^n$ with

$$|\nu| \le \delta 2^{-(2d-1)j}, \quad |\eta| \le \delta$$

we have the relation

$$I_{q,r}(\nu, \eta) = \int_{\mathbb{R}^n} e(\nu|t|^{2d} + \eta \cdot t) K_j(t) dt + O(\delta q). \tag{2.15}$$

The integral on the right–hand side of (2.15) equals

$$q^n \int_{\mathbb{R}^n} e(\nu|tq + r|^{2d} + \eta(tq + r)) K_j(tq + r) dt,$$

which in turn can be split as

$$q^n \sum_{u \in \mathbb{Z}^n} \int_{[0,1]^n} e(\nu|uq + r + tq|^{2d} + \eta \cdot (uq + r + tq)) K_j(uq + r + tq) dt. \tag{2.16}$$

In this display it holds that

$$|\nu|uq + r + tq|^{2d} - \nu|uq + r|^{2d}| \lesssim \delta q$$

since $|r| \le q$, $|uq + r + qt| \approx |uq + r| \approx 2^j$ and $\nu \le \delta 2^{-(2d-1)j}$. Similarly,

$$|\eta \cdot (uq + r + tq) - \eta \cdot (uq + r)| \lesssim \delta q.$$

Using also that $\int_{\mathbb{R}^n} |K_j(t)| dt \approx 1$, this yields that (2.16) is

$$q^n \sum_{u \in \mathbb{Z}^n} \int_{[0,1]^n} e(\nu|uq + r|^{2d} + \eta \cdot (uq + r)) K_j(uq + r + tq) dt + O(\delta q). \tag{2.17}$$

Finally, note from (2.2) that

$$|K_j(uq + r + tq) - K_j(uq + r)| \lesssim 2^{-j(n+1)} q \le 2^{-jn} \delta q.$$

Then we see that (2.17) can be written as

$$q^n \sum_{u \in \mathbb{Z}^n} e(\nu|uq + r|^{2d} + \eta \cdot (uq + r)) K_j(uq + r) + O(\delta q),$$

which establishes (2.15). □

Proof of Lemma 2.4. Let $(a,q)=v>1$. Write $a=a'v$ and $q=q'v$. Then

$$q^n S(\tfrac{a}{q},\tfrac{\mathbf{b}}{q}) = \sum_{u\in[v]^n}\sum_{r\in[q']^n} e(\tfrac{a'}{q'}|uq'+r|^{2d}+\tfrac{\mathbf{b}}{q}\cdot(uq'+r))$$

$$= \left[\sum_{r\in[q']^n} e(\tfrac{a'}{q'}|r|^{2d}+\tfrac{\mathbf{b}}{q}\cdot r)\right]\prod_{i=1}^{n}\sum_{u_i\in[v]} e(\tfrac{\mathbf{b}_i}{v}\cdot u_i)$$

Since $(a,\mathbf{b},q)=1$ and $v>1$, there must exist i_0 such that $\mathbf{b}_{i_0}$ is not divisible by v. But that implies $\sum_{\ell\in[v]} e(\frac{\mathbf{b}_{i_0}}{v}\ell)=0$. □

3. Proof of Theorem 1.1

To prove the theorem, we need to obtain an $\ell^2(\mathbb{Z}^n)$ bound for the maximal operator

$$\sup_{\lambda\in\mathbb{R}}\left|\sum_{j\geq 1} m_{j,\lambda}(\mathrm{D})f\right|,$$

where $m_{j,\lambda}$ is defined in (2.9). A first observation is that for each fixed j,

$$\|\sup_{\lambda\in\mathbb{R}}|m_{j,\lambda}(\mathrm{D})f|\|_{\ell^2(\mathbb{Z}^n)}\lesssim\|f\|_{\ell^2(\mathbb{Z}^n)},$$

by the triangle inequality, Young's convolution inequality and (2.2). As a consequence, we may in the following assume that $j\geq j_0$, where j_0 is a sufficiently large constant depending on d and n.

Before we proceed, we give a rough description of what will be done. For this purpose, we will be deliberately vague when using the terms 'small' and 'close'. At this point, the reader should imagine these terms as being relative to appropriate fractional powers of 2^j, which might differ at each occurrence and will have to be chosen carefully in the sequel. Roughly speaking, the approximation (2.14) tells us what $m_{j,\lambda}(\xi)$ is when λ and ξ are close to rationals with small denominator. On the other hand, Proposition 2.3 tells us that $|m_{j,\lambda}(\xi)|$ is small if any of $\lambda,\xi_1,\dots,\xi_n$ is not close to a rational with small denominator. This naturally leads to a decomposition of $m_{j,\lambda}$ into two new functions. The first arises from summing the main contributions $S(\frac{a}{q},\frac{\mathbf{b}}{q})\Phi_{j,\lambda-\frac{a}{q}}(\xi-\frac{\mathbf{b}}{q})$ over a suitable collection of rational $(\frac{a}{q},\frac{\mathbf{b}}{q})$ with small q. In the terminology of the Hardy–Littlewood circle method, these are the *major arcs*. The second function is an error term, which will subsume both the approximation error from (2.14) and the *minor arcs*, i.e. the cases when at least one of $\lambda,\xi_1,\dots,\xi_n$ is not close to one of the chosen rationals. This decomposition is stated below as (3.6). Following this approach naively already leads to a fundamental problem: the error term crucially depends on λ, but we know only little more about except that its absolute value is small. This leaves us with few strategies to handle the maximal operator corresponding to the error term. This was one of the reasons for the restriction on the parameters λ imposed in [8]. A key insight of Krause [7] was that by a preliminary TT^* argument on the multiplier $m_{j,\lambda}(\xi)$, we may discard 'most' parameters λ: as long as we restrict to λ sufficiently close to a rational with sufficiently small denominator, the TT^* argument yields summable decay in j (see Proposition 3.1 below). For each j, this only leaves λ contained in a union of a few small intervals (see (3.2) below). This allows us to bound the remaining maximal operator for the error term by a standard argument using the fundamental theorem of calculus, the crucial size information on the error and a crude λ–derivative estimate (see Proposition 3.2 below). We proceed with the precise estimates.

3.1. **Decomposition of the multiplier and minor arcs.** Define

$$\mathfrak{A}_j = \{\tfrac{a}{q} \in \mathbb{Q} : (a,q)=1,\, q \in \mathbb{Z} \cap [1, 2^{\lfloor j\varepsilon_1 \rfloor})\}, \tag{3.1}$$

$$X_j = \bigcup_{\alpha \in \mathfrak{A}_j} \{\lambda \in \mathbb{R} : |\lambda - \alpha| \le 2^{-2dj+\varepsilon_1 j}\}, \tag{3.2}$$

where $\varepsilon_1 \in (0, 2^{-5})$ is a small fixed number that will be determined depending on d. Observe that the union in (3.2) is disjoint. The TT^* argument alluded to above yields the following result.

Proposition 3.1. *There exists $\gamma > 0$ only depending on d and n such that for all $j \ge 1$,*

$$\| \sup_{\lambda \notin X_j} |m_{j,\lambda}(\mathrm{D}) f| \|_{\ell^2(\mathbb{Z}^n)} \lesssim 2^{-j\gamma} \|f\|_{\ell^2(\mathbb{Z}^n)}.$$

The proof can be seen as somewhat parallel to that of Stein–Wainger [12] and is given in §4. From now on we can restrict our attention to the multipliers $m_{j,\lambda}(\xi)\mathbf{1}_{X_j}(\lambda)$. In order to define the major arc approximations we need to set up some notation. For a positive integer s define

$$\mathcal{R}_s = \{(\tfrac{a}{q}, \tfrac{\mathbf{b}}{q}) \in \mathbb{Q} \times \mathbb{Q}^n : (a, \mathbf{b}, q) = 1,\, q \in \mathbb{Z} \cap [2^{s-1}, 2^s)\}.$$

Fix a smooth function χ on $\mathbb{R}^n$ with $0 \le \chi \le 1$ that is supported in $\{|\xi| \le 1/2\}$ and equal to one on $[-1/4, 1/4]$. For $s \ge 1$ and $\xi \in \mathbb{R}^n$ we write $\chi_s(\xi) = \chi(2^{10s}\xi)$. Further define for s with $s \le \varepsilon_1 j$,

$$L^s_{j,\lambda}(\xi) = \sum_{(\alpha,\beta) \in \mathcal{R}_s} S(\alpha,\beta) \Phi^*_{j,\lambda-\alpha}(\xi - \beta) \chi_s(\xi - \beta), \tag{3.3}$$

where $\Phi^*_{j,\nu}$ is given by

$$\Phi^*_{j,\nu} = \Phi_{j,\nu} \cdot \mathbf{1}_{|\nu| \le 2^{-2dj+\varepsilon_1 j}}. \tag{3.4}$$

From the definition of $\mathcal{R}_s$ it is clear that $L^s_{j,\lambda}(\xi)$ is periodic in λ and ξ. Also note that if $L^s_{j,\lambda}(\xi) \ne 0$ (where $s \le \varepsilon_1 j$), then $\lambda \in X_j$. Define

$$L_{j,\lambda} = \sum_{1 \le s \le \varepsilon_1 j} L^s_{j,\lambda}. \tag{3.5}$$

Next, the function $E_{j,\lambda}$ is defined as the difference of $m_{j,\lambda}\mathbf{1}_{X_j}(\lambda)$ and $L_{j,\lambda}$ so that

$$m_{j,\lambda} \cdot \mathbf{1}_{X_j}(\lambda) = L_{j,\lambda} + E_{j,\lambda}. \tag{3.6}$$

From the definitions, $L_{j,\lambda}(\xi)$ and $E_{j,\lambda}(\xi)$ are periodic in λ and ξ and vanish unless $\lambda \in X_j$.

Proposition 3.2. *The constant ε_1 can be chosen small enough depending on d so that there exists $\gamma > 0$ only depending on d and n such that for all $j \ge 1$,*

$$\| \sup_{\lambda \in X_j} |E_{j,\lambda}(\mathrm{D}) f| \|_{\ell^2(\mathbb{Z}^n)} \lesssim 2^{-j\gamma} \|f\|_{\ell^2(\mathbb{Z}^n)}.$$

The proof is given in §5. The basic idea is that the absolute value of $E_{j,\lambda}$ should be small (two reasons to believe this are Lemma 2.5 and Proposition 2.3) and its λ–derivatives are not too large. The structure of X_j then allows us to effectively deploy the fundamental theorem of calculus to deal with the supremum over λ.

3.2. **Major arcs.** It now remains to bound the maximal operator associated with the multiplier

$$\sum_{j\geq 1} L_{j,\lambda} = \sum_{j\geq 1}\sum_{1\leq s\leq \varepsilon_1 j} L^s_{j,\lambda} = \sum_{s\geq 1} L^s_\lambda,$$

where we have set

$$L^s_\lambda = \sum_{j\geq \varepsilon_1^{-1}s} L^s_{j,\lambda}. \tag{3.7}$$

The proof of Theorem 1.1 will be completed if we can exhibit $\gamma > 0$ such that for all $s\geq 1$,

$$\|\sup_{\lambda\in\mathbb{R}} |L^s_\lambda(\mathrm{D})f|\|_{\ell^2(\mathbb{Z}^n)} \lesssim_{d,n} 2^{-\gamma s}\|f\|_{\ell^2(\mathbb{Z}^n)}. \tag{3.8}$$

Before proceeding we give a brief informal explanation of the proof strategy. The first step is an appropriate factorization of the multipliers $L^s_\lambda(\xi)$, exploiting that for each fixed (λ,ξ) at most one summand in the sum over the rationals (α,β) is non–zero (this follows directly from the definitions). The contributing α is determined by λ alone, but β depends on both λ and ξ. This dependence is not compatible with the delicate multi–frequency theory such as presented in §8. We will resolve this difficulty by factoring into two multipliers: one that carries the exponential sums $S(\alpha,\beta)$ and preserves the exact dependence of β and one that carries the real–variable oscillatory integrals Φ_j but which we will sum over a *larger* set of frequencies β, removing the dependence on λ. Each of the factors will then be estimated separately, using very different techniques.

Remark. In view of this complication it seems natural to attempt to consider the multipliers in question for each denominator q separately (thus sacrificing potential cancellation among different q with $2^{s-1}\leq q < 2^s$). While this would remove the troublesome dependence of β on λ and simplify the arguments below, the resulting estimates for each q would not be summable in q. This means that cancellation among different denominators is crucial.

We now begin the proof of (3.8) with the definition of some auxiliary sets of rationals:

$$\begin{aligned}
\mathcal{A}_s &= \{\alpha\in\mathbb{Q} \,:\, (\alpha,\beta)\in\mathcal{R}_s \text{ for some } \beta\},\\
\mathcal{B}_s(\alpha) &= \{\beta\in\mathbb{Q}^n \,:\, (\alpha,\beta)\in\mathcal{R}_s\},\\
\mathcal{B}^\sharp_s &= \{\tfrac{\mathbf{b}}{q} \,:\, \mathbf{b}\in\mathbb{Z}^n,\, q\in\mathbb{Z}\cap[2^{s-1},2^s)\}.
\end{aligned} \tag{3.9}$$

By definition,

$$(\alpha,\beta)\in\mathcal{R}_s \iff \alpha\in\mathcal{A}_s,\ \beta\in\mathcal{B}_s(\alpha)$$

and

$$\mathcal{B}_s(\alpha)\subset\mathcal{B}^\sharp_s \quad \text{for all } \alpha.$$

Also note that $\mathcal{B}_s(\alpha)=\emptyset$ if $\alpha\notin\mathcal{A}_s$. Fix a smooth function $\widetilde{\chi}$ with $0\leq\widetilde{\chi}\leq 1$ that equals to one on $\{|\xi|\leq 1/2\}$ (and hence on the support of χ) and is supported in $\{|\xi|\leq 1\}$. Set $\widetilde{\chi}_s(\xi)=\widetilde{\chi}(2^{10s})$. We proceed to define $L^{s,1}_\lambda(\xi)$, $L^{s,2}_\alpha(\xi)$ by

$$L^{s,1}_\lambda(\xi) = \sum_{\beta\in\mathcal{B}^\sharp_s} \Phi^s_\lambda(\xi-\beta)\chi_s(\xi-\beta), \tag{3.10}$$

$$L^{s,2}_\alpha(\xi) = \sum_{\beta\in\mathcal{B}_s(\alpha)} S(\alpha,\beta)\widetilde{\chi}_s(\xi-\beta), \tag{3.11}$$

where we have set

$$\Phi^s_\lambda = \sum_{j\geq\varepsilon_1^{-1}s} \Phi^*_{j,\lambda}. \tag{3.12}$$

We claim that the proof of (3.8) reduces to the following two estimates.

Proposition 3.3. *There exists $\gamma > 0$ depending on d, n such that for every $s \geq 1$*

$$\| \sup_{\alpha \in \mathcal{A}_s} |L^{s,2}_\alpha(\mathrm{D}) f| \|_{\ell^2(\mathbb{Z}^n)} \lesssim 2^{-\gamma s} \|f\|_{\ell^2(\mathbb{Z}^n)}. \tag{3.13}$$

This will be proven in §6 by making use of exponential sum estimates.

Proposition 3.4. *For every $s \geq 1$,*

$$\| \sup_{\lambda \in \mathbb{R}} |L^{s,1}_\lambda(\mathrm{D}) f| \|_{\ell^2(\mathbb{Z}^n)} \lesssim s^2 \|f\|_{\ell^2(\mathbb{Z}^n)}. \tag{3.14}$$

This is an application of Theorem 2.2, which in turn relies on a variation–norm estimate that is similar to (1.3). We give the details in §7.

We finish this section by showing how Propositions 3.3 and 3.4 imply (3.8). By Kolmogorov–Seliverstov linearization it suffices to show that for every function $\lambda : \mathbb{Z}^n \to \mathbb{R}$, the $\ell^2(\mathbb{Z}^n) \to \ell^2(\mathbb{Z}^n)$ operator norm of the linear operator

$$f \mapsto (x \mapsto L^s_{\lambda(x)}(\mathrm{D}) f(x))$$

is $\lesssim 2^{-\gamma s}$ for some $\gamma > 0$, where both γ and the implicit constant only depend on d, n, K (and thus in particular, are independent of the function $\lambda(\cdot)$). From (3.3), (3.4), (3.7), (3.12) we deduce the identity

$$L^s_{\lambda(x)}(\xi) = \sum_{\beta \in \mathcal{B}_s(\alpha(x))} S(\alpha(x), \beta) \Phi^s_{\lambda(x) - \alpha(x)}(\xi - \beta) \chi_s(\xi - \beta), \tag{3.15}$$

where for each $x \in \mathbb{Z}^n$, $\alpha(x)$ is defined as the unique $\alpha \in \mathcal{A}_s$ such that $|\lambda(x) - \alpha| \leq 2^{-3s}$ (say), or as an arbitrary value from the complement of $\mathcal{A}_s$ if no such α exists (in this case, $L^s_{\lambda(x)}(\xi) = 0$). From (3.15), (3.10), (3.11) and disjointness of the supports of $\chi_s(\cdot - \beta)$ for different $\beta \in \mathcal{B}^\sharp_s$,

$$L^s_{\lambda(x)}(\xi) = L^{s,1}_{\lambda(x) - \alpha(x)}(\xi) L^{s,2}_{\alpha(x)}(\xi).$$

Applying the linearizations of (3.13) and (3.14) this yields (3.8) .

4. Minor arcs I: Proof of Proposition 3.1

Since the output $m_{j,\lambda}(\mathrm{D}) f(x)$ only depends on the values of f in a 2^{j+1}–neighborhood of the point x, a standard localization argument allows us to assume that f is supported in the set $B_j = \{y \in \mathbb{Z}^n : |y| \leq 2^j\}$. Fix an arbitrary function $\lambda : \mathbb{Z}^n \to \mathbb{R} \setminus X_j$ and write

$$T_{j,\lambda} f(x) = m_{j,\lambda(x)}(\mathrm{D})(f \mathbf{1}_{B_j})(x) = \sum_{y \in \mathbb{Z}^n} f(y) \mathcal{K}_{j,\lambda}(x, y),$$

where

$$\mathcal{K}_{j,\lambda}(x, y) = e(\lambda(x) |x - y|^{2d}) K_j(x - y) \mathbf{1}_{B_j}(y).$$

Then the kernel of $T_{j,\lambda} T^*_{j,\lambda}$ is given by

$$\mathcal{K}^\sharp_{j,\lambda}(x, y) = \sum_{z \in \mathbb{Z}^n} e(\lambda(x) |z|^{2d} - \lambda(y) |y - x + z|^{2d}) \times K_j(z) \overline{K_j(y - x + z)} \mathbf{1}_{B_j}(x - z). \tag{4.1}$$

Note that $\mathcal{K}^\sharp_{j,\lambda}(x, y) = 0$ unless

$$|x| \leq 2^{j+2} \quad \text{and} \quad |y| \leq 2^{j+2}. \tag{4.2}$$

Let $\delta_0 > 0$ and $c_0 > 0$ be determined later and define

$$E_{j,\lambda} = \{(x, y) \in \mathbb{Z}^n \times \mathbb{Z}^n : |\mathcal{K}^\sharp_{j,\lambda}(x, y)| \geq c_0 2^{-j(n + \delta_0)}\}.$$

Lemma 4.1. *The constants c_0 and δ_0 can be chosen depending on d, n such that for every $j \geq 1$ it holds that*
$$|E_{j,\lambda}| \lesssim 2^{2nj - \frac{1}{11}\varepsilon_1 j}. \tag{4.3}$$
where ε_1 is as in (3.1), (3.2).

Before proving this statement we show how it can be used to finish the proof of Proposition 3.1. By definition of $E_{j,\lambda}$,
$$|\mathcal{K}^\sharp_{j,\lambda}(x,y)| \lesssim 2^{-nj-\delta_0 j}\mathbf{1}_{B_{j+2}\times B_{j+2}}(x,y) + 2^{-nj}\mathbf{1}_{E_{j,\lambda}}(x,y).$$
With (4.3) this implies
$$\|\mathcal{K}^\sharp_{j,\lambda}\|_{\ell^2(\mathbb{Z}^n\times\mathbb{Z}^n)} \lesssim 2^{-\delta_0 j} + 2^{-\frac{1}{22}\varepsilon_1 j}. \tag{4.4}$$
By the Cauchy–Schwarz inequality we have
$$|\langle T_{\mathcal{K}^\sharp_{j,\lambda}} f, g\rangle| \leq \sum_{x\in\mathbb{Z}^n}\sum_{y\in\mathbb{Z}^n} |g(x)||f(y)||\mathcal{K}^\sharp_{j,\lambda}(x,y)| \leq \|f\|_{\ell^2(\mathbb{Z}^n)}\|g\|_{\ell^2(\mathbb{Z}^n)}\|\mathcal{K}^\sharp_{j,\lambda}\|_{\ell^2(\mathbb{Z}^n\times\mathbb{Z}^n)},$$
which by (4.4) and ℓ^2 duality leads to
$$\|T_{j,\lambda}\|_{\ell^2\to\ell^2} = \|T_{\mathcal{K}^\sharp_{j,\lambda}}\|^{1/2}_{\ell^2\to\ell^2} \lesssim 2^{-\gamma j}$$
with $\gamma = \min(\frac{1}{2}\delta_0, \frac{1}{44}\varepsilon_1)$. It remains to prove Lemma 4.1.

In fact we will prove something stronger: the claim is that after choosing c_0 and δ_0 suitably, we have for every fixed $(x', y^*) \in \mathbb{Z}^{n-1}\times\mathbb{Z}^n$ that
$$|\{x_1 \in \mathbb{Z} : (x_1, x', y^*) \in E_{j,\lambda}\}| \lesssim 2^{j-\frac{1}{11}\varepsilon_1 j}. \tag{4.5}$$
In other words, each (x', y^*)–slice of $E_{j,\lambda}$ has small cardinality. By Fubini's theorem and (4.2) this implies the claimed inequality (4.3).

Our argument is based on [7, §6]. For future reference, we will be more careful with explicit constants than strictly necessary in this proof. The reader can safely ignore all constants only depending on d in the estimates that follow. Fixing $(x', y^*) \in \mathbb{Z}^{n-1}\times\mathbb{Z}^n$, we define
$$\mathcal{E} = \{x_1 \in \mathbb{Z} : (x_1, x', y^*) \in E_{j,\lambda}\}.$$
Set $\varepsilon_0 = \frac{1}{11}\varepsilon_1$.

Claim. The numbers c_0 and δ_0 can be chosen such that the following holds: for every $u \in \mathcal{E}$ there exists a reduced rational $\frac{a}{q}$ with $q \leq 2^{\varepsilon_0 j+1}d$ such that
$$|(u - y_1^*)\lambda(y^*) - \tfrac{a}{q}| \leq 2^{-j(2d-1)+\varepsilon_0 j} \tag{4.6}$$

Proof. Note that the coefficient of z_1^{2d-1} in the phase of (4.1) is equal to $2d(x_1 - y_1)\lambda(y)$. By Dirichlet's approximation theorem, there exists a reduced rational $\frac{a}{q}$ with $q \leq 2^{j(2d-1)-\varepsilon_0 j}$ such that
$$|2d(u - y_1^*)\lambda(y^*) - \tfrac{a}{q}| \leq q^{-1}2^{-j(2d-1)+\varepsilon_0 j} \leq \tfrac{1}{q^2}.$$
Applying Proposition 2.3 (with $N = 2^j$) we may choose c_0 and δ_0 (depending on the choice of ε_0) so that $q \leq 2^{\varepsilon_0 j}$ (because $|\mathcal{K}^\sharp_{j,\lambda}(u, x', y^*)| \geq c_0 2^{-j(n+\delta_0)}$). Dividing through by $2d$ yields the claim. □

From now on we fix c_0 and δ_0 to make the statement in the claim valid. We will also assume $j \geq j_0$, where j_0 is a large constant depending only on d that will be determined later. Our goal is now to show that $|\mathcal{E}| \leq 2^{j-\varepsilon_0 j}$. Arguing by contradiction, we assume that
$$|\mathcal{E}| > 2^{j-\varepsilon_0 j}. \tag{4.7}$$
It is clear that
$$\mathcal{E} \subset [-2^{j+2}, 2^{j+2}]. \tag{4.8}$$

We now exploit the three properties (4.6), (4.7), (4.8) to prove that $\lambda(y^*) \in X_j$, which establishes the required contradiction. First, we claim that there exist $u_1, u_2 \in \mathcal{E}$ such that

$$1 \le u_2 - u_1 \le 2^{\varepsilon_0 j+5}, \tag{4.9}$$

Indeed, suppose that all elements of $\mathcal{E}$ were pairwise separated by at least $2^{\varepsilon_0 j+5}$. Then, by (4.8) we would have $|\mathcal{E}| \le 2^{j-\varepsilon_0 j-1}$, which contradicts (4.7). Consequently, there must exist $u_1, u_2 \in \mathcal{E}$ such that (4.9) holds. By (4.6) there exist reduced rationals $\frac{a}{q}, \frac{a'}{q'}$ with $\max(q, q') \le 2^{\varepsilon_0 j+1}d$ and

$$|(u_1 - y_1^*)\lambda(y^*) - \tfrac{a}{q}| \le 2^{-j(2d-1)+\varepsilon_0 j},$$

$$|(u_2 - y_1^*)\lambda(y^*) - \tfrac{a'}{q'}| \le 2^{-j(2d-1)+\varepsilon_0 j}.$$

Then,

$$|\lambda(y^*) - \tfrac{a^*}{q^*}| \le 2^{-j(2d-1)+\varepsilon_0 j+1}, \tag{4.10}$$

where $\frac{a^*}{q^*} = (u_2 - u_1)^{-1}(\frac{a'}{q'} - \frac{a}{q})$ is a reduced rational with

$$q^* \le qq'(u_2 - u_1) \le 2^{3\varepsilon_0 j+7}d^2. \tag{4.11}$$

With (4.10) we have already obtained a somewhat decent rational approximation for $\lambda(y^*)$. However, to conclude $\lambda(y^*) \in X_j$, we need to show that the approximation is actually tighter by almost another factor of 2^{-j} on the right–hand side (see (3.2)). Denote the set of reduced rationals $\frac{a}{q} \in [0, 1)$ with $q \le 2^{\varepsilon_0 j+1}d$ and $a \in [q]$ by $\mathscr{A}$. Then for each $\alpha \in \mathscr{A}$ we define

$$\mathscr{F}_\alpha = \{u \in \mathcal{E} : |(u - y_1^*)\lambda(y^*) - \alpha|_{\mathbb{T}} \le 2^{-(2d-1)j+\varepsilon_0 j}\},$$

where $|\xi|_{\mathbb{T}} = \min_{z\in\mathbb{Z}} |\xi + z| \le |\xi|$. By (4.6), we have $\mathcal{E} = \cup_{\alpha\in\mathscr{A}} \mathscr{F}_\alpha$. Since also $|\mathscr{A}| \le d^2 2^{2\varepsilon_0 j+1}$, the pigeonhole principle and (4.7) imply that there exists $\alpha_0 = \frac{a_0}{q_0} \in \mathscr{A}$ such that

$$|\mathscr{F}_{\alpha_0}| \ge 2^{j-3\varepsilon_0 j-1}d^{-2}.$$

Next we invoke the following general fact: for positive integers N and k with $Nk^{-2} \ge 2$ and a subset $A \subset [-N/2, N/2] \cap \mathbb{Z}$ with $|A| \ge Nk^{-1}$ there exist $a, b \in A$ with $Nk^{-3} \le b - a \le Nk^{-2}$. (Indeed, covering $[-N/2, N/2]$ with k^2 intervals of size Nk^{-2}, there must exist at least one interval I with $|A \cap I| \ge Nk^{-3}$. Now choose $b = \max A \cap I$ and $a = \min A \cap I$.) Applying this observation (with $N = 2^{j+3}$, $k = \lceil 2^{3\varepsilon_0 j+4}d^2\rceil$, $A = \mathscr{F}_{\alpha_0}$) we obtain $v_1, v_2 \in \mathscr{F}_{\alpha_0}$ such that

$$2^{j-9\varepsilon_0 j-10}d^{-6} \le v_2 - v_1 \le 2^{j-6\varepsilon_0 j-5}d^{-4}. \tag{4.12}$$

By definition of $\mathscr{F}_{\alpha_0}$ there exist integers ℓ_1, ℓ_2 such that

$$|(v_1 - y_1^*)\lambda(y^*) - (\alpha_0 + \ell_1)| \le 2^{-(2d-1)j+\varepsilon_0 j},$$

$$|(v_2 - y_1^*)\lambda(y^*) - (\alpha_0 + \ell_2)| \le 2^{-(2d-1)j+\varepsilon_0 j}.$$

This implies, using the lower bound in (4.12), that

$$|\lambda(y^*) - \tfrac{\ell_2-\ell_1}{v_2-v_1}| \le 2^{-2dj+10\varepsilon_0 j+11}d^6. \tag{4.13}$$

We claim that

$$\tfrac{\ell_2-\ell_1}{v_2-v_1} = \tfrac{a^*}{q^*}. \tag{4.14}$$

Indeed, suppose not. Then, from (4.12) and (4.11) we have

$$|\tfrac{\ell_2-\ell_1}{v_2-v_1} - \tfrac{a^*}{q^*}| \ge \tfrac{1}{(v_2-v_1)q^*} \ge 2^{-j+3\varepsilon_0 j-2}d^2.$$

On the other hand, from (4.10) and (4.13),

$$|\tfrac{\ell_2-\ell_1}{v_2-v_1} - \tfrac{a^*}{q^*}| \le 2^{-(2d-1)j+\varepsilon_0 j+2},$$

for $j \geq j_0$ large enough. This yields a contradiction (again, for $j \geq j_0$ large enough). Thus, (4.14) holds. Summarizing, we have proven that

$$|\lambda(y^*) - \tfrac{a^*}{q^*}| \leq 2^{-2dj+11\varepsilon_0 j}$$

for $j \geq j_0$ large enough (from (4.14) and (4.13)). Further, $(a^*, q^*) = 1$ and $q^* \leq d^2 2^{3\varepsilon_0 j+7} \leq 2^{\lfloor 11\varepsilon_0 j \rfloor}$ for large enough $j \geq j_0$. Recalling that we set $\varepsilon_0 = \frac{1}{11}\varepsilon_1$, this means precisely that $\lambda(y^*) \in X_j$.

Remarks. 1. The argument simplifies slightly in the case $d > 1$: in place of the upper bound in (4.12), the trivial upper bound 2^{j+3} would be sufficient.

2. From the proof it is clear that the factor $\frac{1}{11}$ appearing in (4.3) is not sharp. However, this is not relevant for our discussion.

5. Minor arcs II: Proof of Proposition 3.2

We will make use of the following fact.

Lemma 5.1. *Let $\Lambda \subset \mathbb{R}$ be a disjoint union of intervals $(I_j)_{1\leq j\leq N}$ with $|I_j| \leq \delta$, and $(m_\lambda)_{\lambda\in\Lambda}$ a family of bounded periodic functions on $\mathbb{R}^n$ such that*

$$\sup_{\lambda\in\Lambda} \|m_\lambda\|_{L^\infty(\mathbb{T}^n)} \leq A, \tag{5.1}$$

the function $I_j \to \mathbb{C}$, $\lambda \mapsto m_\lambda(\xi)$ is absolutely continuous for a.e. $\xi \in \mathbb{R}^n$ and every $j = 1, \ldots, N$, and

$$\sup_{\lambda\in\Lambda} \|\partial_\lambda m_\lambda\|_{L^\infty(\mathbb{T}^n)} \leq B, \tag{5.2}$$

Then

$$\|\sup_{\lambda\in\Lambda} |m_\lambda(\mathrm{D})f|\|_{\ell^2(\mathbb{Z}^n)} \leq (N^{1/2}A + (2NAB\delta)^{1/2})\|f\|_{\ell^2(\mathbb{Z}^n)}.$$

The proof is via a standard argument using the fundamental theorem of calculus which we postpone to the end of this section. In order to apply Lemma 5.1 to the multipliers $(E_{j,\lambda})_{\lambda\in X_j}$ we will prove that

$$|E_{j,\lambda}(\xi)| \lesssim 2^{-\gamma j} \tag{5.3}$$

for some $\gamma > 0$ only depending on d, n (in particular, not depending on the choice of ε_1) and all $\lambda \in \mathbb{R}, \xi \in \mathbb{R}^n$, $j \geq 1$. Moreover, we have directly from the definitions (3.6), (3.3), (2.11), (2.9) that for a.e. $\lambda \in \mathbb{R}, \xi \in \mathbb{R}^n$ and every $j \geq 1$,

$$|\partial_\lambda E_{j,\lambda}(\xi)| \lesssim 2^{2dj}. \tag{5.4}$$

Then Lemma 5.1 (with $\Lambda = X_j \cap [0,1)$, $m_\lambda = E_{j,\lambda}$, $N = |\mathfrak{A}_j| \leq 2^{2\varepsilon_1 j}$, $\delta \leq 2^{-2dj+\varepsilon_1 j+1}$) gives

$$\|\sup_{\lambda\in X_j} |E_{j,\lambda}(\mathrm{D})f|\|_{\ell^2(\mathbb{Z}^n)} \lesssim 2^{\frac{1}{2}(3\varepsilon_1-\gamma)j}\|f\|_{\ell^2(\mathbb{Z}^n)}. \tag{5.5}$$

Thus we obtain the claimed decay in j as long as $\varepsilon_1 < \frac{1}{3}\gamma$. We turn our attention to proving (5.3).

Assume $\lambda \in X_j$ (otherwise $E_{j,\lambda}(\xi) = 0$). Fix $\varepsilon_2 = 2^{-5}$ (this can be replaced by any sufficiently small absolute constant with $\varepsilon_2 > \varepsilon_1$). We define the *major arcs*

$$\mathfrak{M}_j = \bigcup_{\substack{(\alpha,\beta)\in\mathcal{R}_s,\\ 1\leq s\leq\varepsilon_2 j}} \mathfrak{M}_j(\alpha,\beta), \text{ where}$$

$$\mathfrak{M}_j(\alpha,\beta) = \{(\lambda,\xi) \in \mathbb{R}\times\mathbb{R}^n : |\lambda-\alpha| \leq 2^{-2dj+\varepsilon_2 j}, |\xi-\beta| \leq 2^{-j+\varepsilon_2 j}\}.$$

We need the following disjointness statement for the neighborhoods of the rationals involved in the sum defining $L_{j,\lambda}(\xi)$.

Lemma 5.2. *For each $(\lambda, \xi) \in \mathbb{R} \times \mathbb{R}^n$ there exists at most one (α, β) with $(\alpha, \beta) \in \mathcal{R}_s$ for some $1 \le s \le \varepsilon_2 j$ such that*

$$S(\alpha, \beta)\Phi^*_{j,\lambda-\alpha}(\xi - \beta)\chi_s(\xi - \beta) \neq 0. \tag{5.6}$$

If that is the case and also $s \le \varepsilon_1 j$, then

$$L_{j,\lambda}(\xi) = L^s_{j,\lambda}(\xi) = S(\alpha, \beta)\Phi^*_{j,\lambda-\alpha}(\xi - \beta)\chi_s(\xi - \beta).$$

(Otherwise, $L_{j,\lambda}(\xi) = 0$.)

Proof. Fix $(\lambda, \xi) \in \mathbb{R} \times \mathbb{R}^n$. Take $(\alpha, \beta) \in \mathcal{R}_s, (\alpha', \beta') \in \mathcal{R}_{s'}$ such that (5.6) holds. Suppose that $\alpha \neq \alpha'$. Then

$$2^{-2\varepsilon_2 j} \le 2^{-(s+s')} \le |\alpha - \alpha'| \le 2^{-2dj+\varepsilon_1 j+1}.$$

This is a contradiction. Thus, $\alpha = \alpha'$. Write $(\alpha, \beta) = (\frac{a}{q}, \frac{\mathbf{b}}{q})$, $(\alpha', \beta') = (\frac{a'}{q'}, \frac{\mathbf{b}'}{q'})$ with $(a, \mathbf{b}, q) = (a', \mathbf{b}', q') = 1$ and $2^{s-1} \le q < 2^s$, $2^{s'-1} \le q' < 2^{s'}$. By Lemma 2.4 and (5.6) we have $(a, q) = 1$ and $(a', q') = 1$. But since $\alpha = \alpha'$, this implies $q = q'$ and thus $s = s'$. Taking another look at (5.6) we see that $\beta = \beta'$ (by inspecting the support of $\chi_s = \chi_{s'}$). The claim about $L_{j,\lambda}(\xi)$ follows from the claim we just proved and (3.5), (3.3). $\square$

The proof of (5.3) naturally splits into several cases.

Case 1: $(\lambda, \xi) \in \mathfrak{M}_j$. Then there exist $1 \le s_0 \le \varepsilon_2 j$ and $(\alpha_0, \beta_0) \in \mathcal{R}_{s_0}$ such that $(\lambda, \xi) \in \mathfrak{M}_j(\alpha_0, \beta_0)$. From Lemma 2.5 (with $\delta = 2^{-j+\varepsilon_2 j}$, $q \le 2^{\varepsilon_2 j}$) we gather that

$$m_{j,\lambda}(\xi) = S(\alpha_0, \beta_0)\Phi_{j,\lambda-\alpha_0}(\xi - \beta_0) + O(2^{-j+2\varepsilon_2 j}) \tag{5.7}$$

We distinguish two further cases.

Case 1.1: $1 \le s_0 \le \varepsilon_1 j$. From Lemma 5.2 we deduce

$$L_{j,\lambda}(\xi) = L^{s_0}_{j,\lambda}(\xi) = S(\alpha_0, \beta_0)\Phi_{j,\lambda-\alpha_0}(\xi - \beta_0).$$

With (5.7) this gives

$$|E_{j,\lambda}(\xi)| = |m_{j,\lambda}(\xi) - L_{j,\lambda}(\xi)| \lesssim 2^{-j+2\varepsilon_2 j}.$$

Case 1.2: $\varepsilon_1 j < s_0 \le \varepsilon_2 j$. We may write $\alpha_0 = \frac{a_0}{q_0}$, $\beta_0 = \frac{\mathbf{b}_0}{q_0}$ with $(a_0, \mathbf{b}_0, q_0) = 1$, $2^{s_0-1} \le q_0 < 2^{s_0}$. In particular, $q_0 \ge 2^{\lfloor \varepsilon_1 j \rfloor}$.

We claim that we must have $(a_0, q_0) > 1$. Indeed, suppose $(a_0, q_0) = 1$. Since $\lambda \in X_j$, there exists a reduced rational $\frac{a_1}{q_1}$ with $q_1 < 2^{\lfloor \varepsilon_1 j \rfloor}$ and

$$|\tfrac{a_1}{q_1} - \lambda| \le 2^{-2dj+\varepsilon_1 j}$$

Since $q_0 > q_1$, the reduced rationals $\frac{a_1}{q_1}$ and $\frac{a_0}{q_0}$ do not coincide. Therefore,

$$2^{-(\varepsilon_1+\varepsilon_2)j} \le \tfrac{1}{q_0 q_1} \le |\tfrac{a_1}{q_1} - \tfrac{a_0}{q_0}| \le 2^{-2dj+\varepsilon_2 j+1}.$$

This is a contradiction. Thus we must have $(a_0, q_0) > 1$ and so $S(\alpha_0, \beta_0) = 0$ by Lemma 2.4. In particular, $|m_{j,\lambda}(\xi)| \lesssim 2^{-j+2\varepsilon_2 j}$ by (5.7). Also, from Lemma 5.2 we see that $L_{j,\lambda}(\xi) = 0$.

Case 2: $(\lambda,\xi)\notin\mathfrak{M}_j$**.** In this case we bound

$$|E_{j,\lambda}(\xi)|\le|m_{j,\lambda}(\xi)|+|L_{j,\lambda}(\xi)|$$

and estimate the two terms on the right–hand side separately.

Fix $\epsilon<\frac{\varepsilon_2}{n+1}$ and set $N=2^j$. By Dirichlet's approximation theorem there exist reduced fractions $\frac{a}{q},\frac{b_1}{r_1},\dots,\frac{b_n}{r_n}$ with $q\le N^{2d-\epsilon}$, $\max(r_1,\dots,r_n)\le N^{1-\epsilon}$ and

$$|\lambda-\tfrac{a}{q}|\le\tfrac{1}{q}N^{-2d+\epsilon},\ |\xi_k-\tfrac{b_k}{r_k}|\le\tfrac{1}{r_k}N^{-1+\epsilon}\text{ for }k=1,\dots,n.$$

Setting $q_*=\mathrm{lcm}(q,r_1,\dots,r_n)$, we must have $q_*\ge 2^{\lfloor\varepsilon_2 j\rfloor}$ because $(\lambda,\xi)\notin\mathfrak{M}_j$. Thus at least one of $q,r_1,\dots,r_n$ must be $\ge 2^{\epsilon j}$ (otherwise $q_*\le 2^{\epsilon(n+1)j}$ which is a contradiction because $\epsilon<\frac{\varepsilon_2}{n+1}$). By Proposition 2.3 we then obtain

$$|m_{j,\lambda}(\xi)|\lesssim 2^{-\delta j}.$$

It remains to estimate $|L_{j,\lambda}(\xi)|$. Suppose that $L_{j,\lambda}(\xi)\neq 0$. Then, by Lemma 5.2 there exists $(\alpha,\beta)\in\mathcal{R}_s$ for some $1\le s\le\varepsilon_1 j$ such that

$$L_{j,\lambda}(\xi)=S(\alpha,\beta)\Phi^*_{j,\lambda-\alpha}(\xi-\beta)\chi_s(\xi-\beta). \tag{5.8}$$

Then $|\lambda-\alpha|\le 2^{-2dj+\varepsilon_2 j}$. Since $(\lambda,\xi)\notin\mathfrak{M}_j$,

$$|\xi-\beta|\ge 2^{-j+\varepsilon_2 j}.$$

With (2.12) and (5.8), this implies

$$|L_{j,\lambda}(\xi)|\le|\Phi_{j,\lambda-\alpha}(\xi-\beta)|\lesssim 2^{-\frac{\varepsilon_2}{2d}j}.$$

Proof of Lemma 5.1. By the fundamental theorem of calculus, we have for absolutely continuous $g:[a,b]\to\mathbb{C}$,

$$\sup_{\lambda\in[a,b]}|g(\lambda)|^2\le|g(a)|^2+2\int_a^b|g(t)||g'(t)|dt.$$

Hence,

$$\|\sup_{\lambda\in\Lambda}|m_\lambda(\mathrm{D})f|\|^2_{\ell^2(\mathbb{Z}^n)}\le\sum_{j=1}^N\|m_{\inf\, I_j}(\mathrm{D})f\|^2_{\ell^2(\mathbb{Z}^n)} \tag{5.9}$$

$$+\quad 2\sum_{j=1}^N\sum_{x\in\mathbb{Z}^n}\int_{I_j}|m_t(\mathrm{D})f(x)||\partial_\lambda m_t(\mathrm{D})f(x)|dt.$$

By the Cauchy–Schwarz inequality and Fubini's theorem,

$$\sum_{x\in\mathbb{Z}^n}\int_{I_j}|m_t(\mathrm{D})f(x)||\partial_\lambda m_t(\mathrm{D})f(x)|dt$$

$$\le\Big(\int_{I_j}\|m_t(\mathrm{D})f\|^2_{\ell^2(\mathbb{Z}^n)}dt\Big)^{1/2}\Big(\int_{I_j}\|\partial_\lambda m_t(\mathrm{D})f\|^2_{\ell^2(\mathbb{Z}^n)}dt\Big)^{1/2}.$$

Combining this with (5.9) and using Plancherel's theorem with the assumptions (5.1), (5.2) we obtain the claim.

Remark. Observe that the same argument works for ℓ^p with $p\neq 2$ and more general families of operators. $\square$

6. Major arcs I: Proof of Proposition 3.3

Note that since $L^{s,2}_{s,\alpha}(\xi) = L^{s,2}_{s,\alpha+1}(\xi)$, we may restrict the supremum to $\alpha \in \mathcal{A}_s \cap [0,1)$, without loss of generality. Next, from (3.11) we have for $y \in \mathbb{Z}^n$,

$$\mathcal{F}^{-1}[L^{s,2}_{\alpha}](y) = \sum_{\beta \in \mathcal{B}_s(\alpha)} S(\alpha,\beta) \int_{[0,1]^n} e(\xi \cdot y) \widetilde{\chi}_s(\xi - \beta) d\xi$$

$$= \sum_{\beta \in \mathcal{B}_s(\alpha) \cap [0,1)^n} S(\alpha,\beta) e(\beta \cdot y) \phi_s(y),$$

where $\phi_s = \mathcal{F}^{-1}_{\mathbb{R}^n}[\widetilde{\chi}_s]$. Note that $\|\phi_s\|_{L^1(\mathbb{R}^n)} \approx 1$.

For an arbitrary function $\alpha : \mathbb{Z}^n \to \mathcal{A}_s \cap [0,1)$ we define

$$\mathfrak{K}_{s,\alpha}(x,y) = \mathcal{F}^{-1}[L^{s,2}_{\alpha(x)}](x-y). \tag{6.1}$$

Then Proposition 3.3 is a consequence of the following.

Proposition 6.1. *There exists $\gamma > 0$ depending only on d, n such that*

$$\|T_{\mathfrak{K}_{s,\alpha}} f\|_{\ell^2(\mathbb{Z}^n)} \lesssim 2^{-\gamma s} \|f\|_{\ell^2(\mathbb{Z}^n)},$$

with the implicit constant only depending on d and n, but not on the functions α, f. (The notation $T_{\mathfrak{K}_{s,\alpha}}$ is defined in (2.1)*.)*

Remark. The proof shows that the same result holds with ℓ^2 replaced by ℓ^p for every $p \in (1,\infty)$ (with decay rate depending on p).

For every $x \in \mathbb{Z}^n$ there exist $\mathrm{q}(x) \in \mathbb{Z} \cap [2^{s-1}, 2^s)$ and $\mathrm{a}(x) \in [\mathrm{q}(x)]$ with $(\mathrm{a}(x), \mathrm{q}(x)) = 1$ such that

$$\alpha(x) = \tfrac{\mathrm{a}(x)}{\mathrm{q}(x)}.$$

For the proof we will employ a TT^*–argument. We begin by computing the kernel of TT^*. Note that $T_{\mathfrak{K}_{s,\alpha}} T^*_{\mathfrak{K}_{s,\alpha}} = T_{\mathfrak{K}^\sharp_{s,\alpha}}$, where

$$\mathfrak{K}^\sharp_{s,\alpha}(x,y) = \sum_{z \in \mathbb{Z}^n} \mathfrak{K}_{s,\alpha}(x,z) \overline{\mathfrak{K}_{s,\alpha}(y,z)}.$$

From (6.1),

$$\mathfrak{K}^\sharp_{s,\alpha}(x,y) = \sum_{\substack{\beta \in \mathcal{B}_s(\alpha(x)) \cap [0,1)^n, \\ \beta' \in \mathcal{B}_s(\alpha(y)) \cap [0,1)^n}} S(\alpha(x),\beta) e(x \cdot \beta) \overline{S(\alpha(y),\beta')} e(-y \cdot \beta')$$

$$\times \left[\sum_{z \in \mathbb{Z}^n} \phi_s(x-z) \overline{\phi_s(y-z)} e(z \cdot (\beta' - \beta)) \right]$$

Next we claim that for every $\beta, \beta' \in \mathcal{B}^\sharp_s \cap [0,1)^n$ with $\beta \neq \beta'$ it holds that

$$\sum_{z \in \mathbb{Z}^n} \phi_s(x-z) \overline{\phi_s(y-z)} e(z \cdot (\beta' - \beta)) = 0. \tag{6.2}$$

To see this, define a Schwartz function on $\mathbb{R}^n$ by

$$\Xi(t) = \phi_s(x-t) \overline{\phi_s(y-t)} e(t(\beta' - \beta)), \quad (t \in \mathbb{R}^n).$$

Then

$$\widehat{\Xi}(\xi) = [\mathrm{M}_{-x} \widetilde{\chi}_s * \mathrm{M}_{-y} \widetilde{\chi}_s](\xi + \beta - \beta'),$$

where we used the notation $\mathrm{M}_u g(x) = e(u \cdot x)g(x)$. From the definitions of $\widetilde{\chi}_s$ and $\mathcal{B}_s^\sharp$ we then have for $\xi \in \mathbb{Z}^n$ that $\widehat{\Xi}(\xi) = 0$ unless $\xi + \beta - \beta' = 0$. However, $\beta, \beta' \in [0,1)^n$ and $\beta \neq \beta'$ imply $\beta - \beta' \notin \mathbb{Z}^n$. Hence, by the Poisson summation formula the left–hand side of (6.2) is equal to

$$\sum_{z \in \mathbb{Z}^n} \Xi(z) = \sum_{\xi \in \mathbb{Z}^n} \widehat{\Xi}(\xi) = 0.$$

As a consequence,

$$\mathfrak{K}_{s,\alpha}^\sharp(x,y) = \overline{\kappa_{s,\alpha}(x,y)} \cdot [\phi_s * \overline{\phi_s}](x-y), \tag{6.3}$$

where we set

$$\kappa_{s,\alpha}(x,y) = \sum_{\beta \in \mathcal{B}_s(\alpha(x)) \cap \mathcal{B}_s(\alpha(y)) \cap [0,1)^n} S(\alpha(y),\beta)\overline{S(\alpha(x),\beta)} e((y-x)\cdot\beta). \tag{6.4}$$

For the following computation we fix $(x,y) \in \mathbb{Z}^n \times \mathbb{Z}^n$ and write

$$a = \mathrm{a}(y), \quad q = \mathrm{q}(y), \quad a' = \mathrm{a}(x), \quad q' = \mathrm{q}(x)$$

for short. As a consequence of Lemma 2.4, we may assume $(a,q) = (a',q') = 1$ and read the sum over β in (6.4) as running over the set

$$\{\tfrac{\mathbf{b}}{q} \,:\, \mathbf{b} \in [q]^n\} \cap \{\tfrac{\mathbf{b}}{q'} \,:\, \mathbf{b} \in [q']^n\},$$

which is equal to

$$\{\tfrac{\mathbf{b}}{q_\flat} \,:\, \mathbf{b} \in [q_\flat]^n\},$$

where we have set $q_\flat = (q,q')$. Thus,

$$\kappa_{s,\alpha}(x,y) = \sum_{\mathbf{b} \in [q_\flat]^n} S(\tfrac{a}{q}, \tfrac{\mathbf{b}}{q_\flat})\overline{S(\tfrac{a'}{q'}, \tfrac{\mathbf{b}}{q_\flat})} e((y-x)\cdot\tfrac{\mathbf{b}}{q_\flat}).$$

Expanding the exponential sums by (2.10), we can rewrite this as

$$(qq')^{-n} \sum_{r \in [q]^n, r' \in [q']^n} e(\tfrac{a}{q}|r|^{2d} - \tfrac{a'}{q'}|r'|^{2d}) \left[\sum_{\mathbf{b} \in [q_\flat]^n} e(\tfrac{\mathbf{b}}{q_\flat} \cdot (r - r' + y - x)) \right],$$

which, in view of the relation $N^{-1} \sum_{l \in [N]} e(\frac{lz}{N}) = \mathbf{1}_{z \equiv 0 \ (\mathrm{mod}\ N)}$, is equal to

$$(\tfrac{q'}{q_\flat})^{-n} \sum_{u \in [\frac{q'}{q_\flat}]^n} q^{-n} \sum_{r \in [q]^n} e(\tfrac{a}{q}|r|^{2d} - \tfrac{a'}{q'}|r + y - x + u\cdot q_\flat|^{2d}). \tag{6.5}$$

Inspection of this exponential sum reveals several scenarios in which no cancellation can be expected. For instance, a typical case where (6.5) exhibits no cancellation is when $a = a'$, $q = q'$ and $y - x$ is divisible by q (then $\kappa_{s,\alpha}(x,y) = 1$). Additional degeneracies arise in the case $d = 1$, requiring a more careful analysis. For $w \in \mathbb{Z}^n$ we define

$$S_{x,y}(w) = q^{-n} \sum_{r \in [q]^n} e(\tfrac{a}{q}|r|^{2d} - \tfrac{a'}{q'}|r + w|^{2d}). \tag{6.6}$$

In the case $d \geq 2$ it will suffice to exploit cancellation from the exponential sum (6.6), whereas in the case $d = 1$ we will sometimes need to make use of cancellation from the sum over u in (6.5).

The case $d \geq 2$**.** Viewing the phase in (6.6) as a polynomial in r, the coefficient of r_1^{2d-1} is equal to $\frac{-2da'w_1}{q'}$. This leads us to define

$$\mathcal{E}_x = \{w \in \mathbb{Z}^n \,:\, (2dw_1, \mathrm{q}(x)) \geq 2^{s/2}\}.$$

By sorting modulo $\mathrm{q}(x)$ we see that for $z \in \mathbb{Z}^n, N \geq 2^s$,

$$N^{-n}|\mathcal{E}_x \cap (z + [N]^n)| \lesssim 2^{-s/2}. \tag{6.7}$$

If $w \notin \mathcal{E}_x$, then Proposition 2.3 yields (here we are crucially using $d \geq 2$)

$$|S_{x,y}(w)| \lesssim 2^{-\gamma s}$$

for some sufficiently small $\gamma \in (0, \frac{1}{2})$ depending on d and n. Using the triangle inequality on the sum over u in (6.5) leads to the estimation

$$|\kappa_{s,\alpha}(x,y)| \lesssim 2^{-\gamma s} + \sum_{\nu | \mathrm{q}(x)} (\tfrac{\mathrm{q}(x)}{\nu})^{-n} \sum_{u \in [\mathrm{q}(x)/\nu]^n} \mathbf{1}_{\mathcal{E}_x}(y - x + u \cdot \nu), \tag{6.8}$$

where we have removed the (x, y)–dependence of $q_\flat = (\mathrm{q}(x), \mathrm{q}(y))$ by summing over all divisors of $\mathrm{q}(x)$. Hence, recalling (6.3), we see for every $x \in \mathbb{Z}^n$ that

$$\sum_{y \in \mathbb{Z}^n} |\mathfrak{K}^\sharp_{s,\alpha}(x,y)| \lesssim 2^{-\gamma s} + \tau(q(x)) \sup_{u \in \mathbb{Z}^n} \sum_{y \in \mathbb{Z}^n} \mathbf{1}_{\mathcal{E}_x}(y - x + u) |\phi_s * \phi_s|(x - y),$$

where $\tau(q)$ denotes the number of divisors of q. Using the standard divisor bound $\tau(q) \lesssim_\varepsilon q^\varepsilon$, (6.7) and rapid decay of $\phi_s * \phi_s$, we obtain

$$\sum_{y \in \mathbb{Z}^n} |\mathfrak{K}^\sharp_{s,\alpha}(x,y)| \lesssim 2^{-\gamma s}$$

for every $x \in \mathbb{Z}^n$. Since also $\mathfrak{K}^\sharp_{s,\alpha}(x,y) = \overline{\mathfrak{K}^\sharp_{s,\alpha}(y,x)}$, we infer from Schur's test that

$$\|T_{\mathfrak{K}^\sharp_{s,\alpha}}\|_{\ell^2(\mathbb{Z}^n) \to \ell^2(\mathbb{Z}^n)} \lesssim 2^{-\gamma s}.$$

This concludes the proof of Proposition 6.1.

The case $d = 1$**.** First assume that $q_\flat = (\mathrm{q}(x), \mathrm{q}(y)) \leq 2^{s/3}$. Then $\frac{q'}{q_\flat} \geq 2^{2s/3-1}$. Viewing the phase in (6.5) as a polynomial in u, the coefficient of u_1^2 is $-\frac{a' q_\flat^2}{q'}$ which equals a reduced rational with denominator in $[2^{s/3}, 2^s] \cap \mathbb{Z}$. Thus, applying Proposition 2.3 to the exponential sum over u yields

$$|\kappa_{s,\alpha}(x,y)| \lesssim 2^{-\gamma s}$$

for a small enough $\gamma > 0$. Next we handle the case that $q_\flat \geq 2^{s/3}$. We will exploit cancellation from the summation over r in (6.5). The exponential sum on the right–hand side of (6.6) factors into n one–dimensional sums. It will be enough to estimate the first factor, which is given by

$$I = q^{-1} \sum_{r_1 \in [q]} e(\tfrac{A}{Q} r_1^2 - \tfrac{2a'w_1}{q'} r_1),$$

where $\frac{A}{Q} = \frac{a}{q} - \frac{a'}{q'}$ with $(A, Q) = 1$. We are led to distinguish two cases. Suppose that $Q \geq 2^{s/3}$. Then, since also $Q \leq \frac{qq'}{q_\flat} \leq 2^{5s/3}$, we may apply Proposition 2.3 to obtain

$$|I| \lesssim 2^{-\delta s} \tag{6.9}$$

for some small enough $\delta > 0$. On the other hand, assume $Q \le 2^{s/3}$. Then, by reorganizing the summation modulo Q,

$$I = q^{-1}\big[\sum_{s\in[Q]} e(\tfrac{A}{Q}s^2 - \tfrac{2a'w_1}{q'}s)\big]\cdot\big[\sum_{u\in[M]} e(-\tfrac{2a'w_1Q}{q'}u)\big] + O(2^{-\frac{2}{3}s}),$$

where $M = \lfloor \frac{q}{Q} \rfloor$. Summing the geometric sum over u and using the triangle inequality on the sum over s we get

$$|I| \lesssim 2^{-\frac{2}{3}s}|1 - e(\tfrac{2a'w_1Q}{q'})|^{-1} \lesssim 2^{-\frac{2}{3}s}|\tfrac{2a'w_1Q}{q'}|_{\mathbb{T}}^{-1}, \tag{6.10}$$

where $|\xi|_{\mathbb{T}} = \min_{z\in\mathbb{Z}} |\xi + z|$. Note that Q depends on both x and y. To remove the dependence on y we define for a positive integer $\upsilon \le 2^{s/3}$ the set

$$E_x^{(\upsilon)} = \{w_1 \in \mathbb{Z} \,:\, |\tfrac{2a'w_1\upsilon}{q'}|_{\mathbb{T}} \le 2^{-s/2}\}. \tag{6.11}$$

Let $\iota = (2\upsilon, q') \lesssim 2^{s/3}$ and $\mathfrak{q} = \frac{q'}{\iota}$. Let $\mathcal{R} \subset \mathbb{Z}$ be a complete residue system modulo $\mathfrak{q}$. Then $a'\frac{2\upsilon}{\iota}\mathcal{R}$ is also a complete residue system modulo $\mathfrak{q}$. Thus

$$|E_x^{(\upsilon)} \cap \mathcal{R}| = |\{\ell \in [\mathfrak{q}] \,:\, |\ell/\mathfrak{q}|_{\mathbb{T}} \le 2^{-s/2}\}| \lesssim \mathfrak{q}2^{-s/2}.$$

Since $\mathfrak{q} \lesssim 2^{\frac{2}{3}s} < 2^s$, we then have for every $N \ge 2^s$ and $z \in \mathbb{Z}$,

$$N^{-1}|E_x^{(\upsilon)} \cap (z + [N])| \lesssim 2^{-s/2}.$$

Define

$$\mathcal{E}_x = \bigcup_{\upsilon\le 2^{s/3}} \{w \in \mathbb{Z}^n \,:\, w_1 \in E_x^{(\upsilon)}\}. \tag{6.12}$$

Then if $w \notin \mathcal{E}_x$, we gather from (6.6), (6.9), (6.10), (6.11) that

$$|S_{x,y}(w)| \le |I| \lesssim \max(2^{-\frac{1}{6}s}, 2^{-\delta s})$$

and for every $z \in \mathbb{Z}^n$ and $N \ge 2^s$,

$$N^{-n}|\mathcal{E}_x \cap (z + [N]^n)| \lesssim 2^{-\frac{1}{6}s}.$$

The fact that we have chosen the exceptional set $\mathcal{E}_x$ only depending on x (as opposed to both x and y) allows us to recycle the crude argument using Schur's test seen in the case $d \ge 2$. Indeed, summarizing the above we have shown that (6.8) again holds for all $(x, y) \in \mathbb{Z}^n \times \mathbb{Z}^n$ with $\mathcal{E}_x$ defined as in (6.12) (and $\gamma > 0$ small enough, possibly different from above). This completes the proof of Proposition 6.1.

7. Major arcs II: Proof of Proposition 3.4

Define

$$\mathcal{L}_\lambda^{s,1}(\xi) = \sum_{\beta\in\mathcal{B}_s^\sharp\cap[0,1)^n} \Phi_\lambda^s(\xi - \beta)\chi_s(\xi - \beta). \tag{7.1}$$

By the definitions of χ_s and $\mathcal{B}_s^\sharp$ (see (3.9)), the function $\mathcal{L}_\lambda^{s,1}$ is supported within a cube of sidelength 1 (that is, within a fundamental domain of $\mathbb{T}^n = (\mathbb{R}/\mathbb{Z})^n$). Moreover, from (3.10) we identify $L_\lambda^{s,1}(\xi)$ as the periodization of $\mathcal{L}_\lambda^{s,1}$, i.e.

$$L_\lambda^{s,1}(\xi) = \sum_{w\in\mathbb{Z}^n} \mathcal{L}_\lambda^{s,1}(\xi + w).$$

We will prove that

$$\|\sup_{\lambda\in\mathbb{R}} |\mathcal{L}_\lambda^{s,1}(\mathrm{D})f|\|_{L^2(\mathbb{R}^n)} \lesssim_{d,n} s^2\|f\|_{L^2(\mathbb{R}^n)}. \tag{7.2}$$

This implies the required estimate (3.14) by Lemma 2.1. The proof of (7.2) is by Theorem 2.2. Indeed, letting $\mu_{\lambda,s}(\xi) = \Phi_\lambda^s(\xi)\chi_s(\xi)$ and $\Theta_s = \mathcal{B}_s^\sharp \cap [0,1)^n$, we may write

$$\mathcal{L}_\lambda^{s,1}(\xi) = \sum_{\theta\in\Theta_s} \mu_{\lambda,s}(\xi-\theta).$$

The corresponding Fourier multiplier operator can be written as

$$\mathcal{L}_\lambda^{s,1}(\mathrm{D})f(x) = \sum_{\theta\in\Theta_s} e(x\cdot\theta)[\mu_{\lambda,s}(\mathrm{D})F_\theta](x),$$

where we have set $F_\theta = \mathcal{F}^{-1}[\widetilde{\chi}_s \cdot \widehat{f}(\cdot+\theta)]$. The claimed inequality (7.2) then follows from Theorem 2.2 and an application of Plancherel's theorem once we establish the following key variation–norm estimate: for $r \in (2,3)$,

$$\|V^r\{\mu_{\lambda,s}(\mathrm{D})f \,:\, \lambda \in (0,1]\}\|_{L^2(\mathbb{R}^n)} \lesssim_{d,n} (r-2)^{-1}\|f\|_{L^2(\mathbb{R}^n)}, \tag{7.3}$$

with constant independent of s.

Proof of the variation–norm estimate (7.3)**.** Here we find it convenient to adopt notation matching that of [4]. We will show a result that is slightly stronger than required, but follows with no additional effort. Let $\ell_0 \in \mathbb{Z}$ and $\ell^* : (0,\infty) \to \mathbb{R}$ a real–valued function. Define

$$\mathcal{H}_*^{(u)}f(x) = \int_{\mathbb{R}^n} f(x-t)e(u|t|^\alpha)\big[\sum_{\ell_0\le\ell\le\ell^*(u)} K_\ell(t)\big]dt,$$

for real $\alpha > 1$. We will assume that ℓ^* has the following properties:

(1) there exists $\beta \in [0,1]$ such that $\ell^*(2^{j\alpha}) = -j\beta$ for all $j \in \mathbb{Z}$, and
(2) ℓ^* is non–increasing.

Proposition 7.1. *Let $\ell_0 \in \mathbb{Z}$ and ℓ^* such that properties (1) and (2) hold. Then for all $r \in (2,3)$ we have*

$$\|V^r\{\mathcal{H}_*^{(u)}f \,:\, u > 0\}\|_{L^2(\mathbb{R}^n)} \lesssim (r-2)^{-1}\|f\|_{L^2(\mathbb{R}^n)} \tag{7.4}$$

with the implicit constant depending only on α, n, K, but not on ℓ_0, ℓ_*, r.

Remark. The restriction to L^2 is only due to the application we have in mind.

We first show how (7.4) implies (7.3). From the definitions (3.12), (3.4), (2.11),

$$\Phi_u^s(\xi) = \sum_{\ell\ge\varepsilon_1^{-1}s} \Phi_{\ell,u}(\xi)\mathbf{1}_{|u|\le 2^{-2d\ell+\varepsilon_1\ell}} = \int_{\mathbb{R}^n} e(u|y|^{2d} + \xi\cdot y)\big[\sum_{\ell_0\le\ell\le\ell^*(u)} K_\ell(y)\big]dy = \widehat{\mathcal{H}_*^{(u)}f}(\xi),$$

where we have set $\ell_0 = \varepsilon_1^{-1}s$ and

$$\ell^*(u) = (2d-\varepsilon_1)^{-1}\log_2(u^{-1}).$$

Observe that ℓ^* satisfies the properties (1) and (2) above with $\alpha = 2d$ and $\beta = \frac{2d}{2d-\varepsilon_1}$. Finally, from Minkowski's inequality and Young's convolution inequality we obtain

$$\|V^r\{\mu_{u,s}(\mathrm{D})f \,:\, u\in(0,1]\}\|_{L^2(\mathbb{R}^n)} \le \|\mathcal{F}^{-1}[\chi_s]\|_{L^1(\mathbb{R}^n)}\|V^r\{\mathcal{H}_*^{(u)}f \,:\, u>0\}\|_{L^2(\mathbb{R}^n)}.$$

Since $\|\mathcal{F}^{-1}[\chi_s]\|_{L^1(\mathbb{R}^n)} \approx 1$, we derive (7.3) as a direct consequence of (7.4).

The proof of Proposition 7.1 is a slight modification of the arguments used to prove Theorem 1.1 of [4]. As we will see, the required modifications do not affect the core of the argument in [4]. To give a precise explanation of the necessary changes, we need to introduce some more notation. For further reading on variation–norm estimates, we refer the reader to Jones–Seeger–Wright [6]. For $\lambda > 0$ and a family of complex numbers $\mathbf{a} = (a_u)_{u\in\mathcal{J}}$ where $\mathcal{J} \subset \mathbb{R}$, define the *$\lambda$–jump function* $N_\lambda(\mathbf{a})$ as the supremum of all positive integers N such that there exist indices $s_1 < t_1 < \cdots < s_N < t_N$ in $\mathcal{J}$ with $|a_{t_j} - a_{s_j}| > \lambda$ for all $j = 1, \dots, N$. Moreover, for $j \in \mathbb{Z}$ and $\mathbf{a} = (a_u)_{u>0}$ we define the *short variations*

$$V_j^r(\mathbf{a}) = V^r\{a_u \,:\, u \in [2^{\alpha j}, 2^{\alpha(j+1)}]\}, \quad S_r(\mathbf{a}) = \big(\sum_{j\in\mathbb{Z}} |V_j^r(\mathbf{a})|^r\big)^{1/r}.$$

We will also write $\mathcal{H}f(x) = [\mathcal{H}^{(u)}f(x)]_{u>0}$ and similarly $\mathcal{H}_*f(x) = [\mathcal{H}_*^{(u)}f(x)]_{u>0}$. For $p \in [4/3, 4]$ we claim that

$$\sup_{\lambda>0} \|\lambda[N_\lambda\{\mathcal{H}_*^{(2^{j\alpha})}f \,:\, j \in \mathbb{Z}\}]^{1/2}\|_{L^p(\mathbb{R}^n)} \lesssim \|f\|_{L^p(\mathbb{R}^n)} \tag{7.5}$$

with constant not depending on $p \in [4/3, 4]$. Moreover, for $r \in (2, 3)$ the claim is that

$$\|S_r(\mathcal{H}_*f)\|_{L^2(\mathbb{R}^n)} \lesssim c_r\|f\|_{L^2(\mathbb{R}^n)}, \tag{7.6}$$

with the implicit constant independent of p, r and $c_r = (r-2)^{-1}$ if $n = 1$ and $c_r = 1$ if $n \geq 2$. In particular, the estimate also holds for $r = 2$ if $n \geq 2$.

Suppose for the moment that we have established (7.5) and (7.6). By (7.5) and [6, Lemma 2.1] we obtain for $r \in (2, 3)$,

$$\|V^r\{\mathcal{H}_*^{2^{j\alpha}} \,:\, j \in \mathbb{Z}\}\|_{L^2(\mathbb{R}^n)} \lesssim (r-2)^{-1}\|f\|_{L^2(\mathbb{R}^n)}. \tag{7.7}$$

By sorting into long and short jumps as in [6, Lemma 1.3] we further have the pointwise inequality

$$V^r(\mathcal{H}_*f) \lesssim V^r\{\mathcal{H}_*^{2^{j\alpha}} \,:\, j \in \mathbb{Z}\} + S_r(\mathcal{H}_*f),$$

which implies (7.4) via (7.6) and (7.7). It remains to verify (7.5) and (7.6). We write

$$\mathcal{H}_\ell^{(u)}f(x) = \int_{\mathbb{R}^n} f(x-t)e(u|t|^\alpha)K_\ell(t)dt.$$

Then our operator is by definition given as

$$\mathcal{H}_*^{(u)} = \sum_{\ell\in\mathbb{Z}} \mathcal{H}_\ell^{(u)}\mathbf{1}_{\ell_0\leq\ell\leq\ell^*(u)},$$

while the operator considered in [4] is

$$\mathcal{H}^{(u)} = \sum_{\ell\in\mathbb{Z}} \mathcal{H}_\ell^{(u)}.$$

We also define $\mathcal{H}_{*,\ell}^{(u)}f(x) = \mathcal{H}_\ell^{(u)}f(x)\mathbf{1}_{\ell_0\leq\ell\leq\ell^*(u)}$.

Long jumps: Proof of (7.5)**.** The arguments required to prove (7.5) are contained in §3 of [4]. We will not reproduce the full details here, but only comment on how the proof changes. Note that

$$\mathcal{H}_{*,\ell}^{(2^{k\alpha})}f = [\mathcal{H}_\ell^{(2^{k\alpha})}f]\mathbf{1}_{\ell_0\leq\ell\leq-k\beta}.$$

Following [4, §3] we decompose

$$\mathcal{H}_*^{(2^{k\alpha})} = \widetilde{\mathcal{H}}_{*,k,0} + \sum_{l\geq0} \mathcal{H}_{*,k,l}\mathbf{1}_{l\leq(1-\beta)k},$$

where

$$\widetilde{\mathcal{H}}_{*,k,0}f(x) = \int_{\mathbb{R}^n} f(x-t)\big[\sum_{\ell_0\leq\ell\leq-k} K_\ell(t)\big]dt,$$

$$\mathcal{H}_{*,k,0}f(x) = \int_{\mathbb{R}^n} f(x-t)(e(2^{k\alpha}|t|^\alpha)-1)\big[\sum_{\ell_0\le\ell\le -k} K_\ell(t)\big]dt,$$

$$\mathcal{H}_{*,k,l}f(x) = \int_{\mathbb{R}^n} f(x-t)e(2^{k\alpha}|t|^\alpha)K_{-k+l}(t)dt, \quad (l>0).$$

Note that $\widetilde{\mathcal{H}}_{*,k,0}$ is a standard truncated singular integral, so the required jump norm estimate for $\widetilde{\mathcal{H}}_{*,k,0}$ follows from [3, Theorem A] (as in [4, Lemma 3.2]). The jump norm estimates for $\mathcal{H}_{*,k,\ell}$, $\ell \ge 0$ follow by standard arguments reducing to Lépingle's inequality (more precisely, to Proposition 2.1 in [4] which is proven in [6]). These arguments are detailed in the remainder of §3 of [4] and are not affected by the truncations. Indeed, the truncation $l \le (1-\beta)k$ is harmless, because it cannot increase the value of (the appropriate version of) the square–function on the left–hand side of display (3.4) in [4]. The truncation from $\ell_0 \le \ell$ only affects the operator $\mathcal{H}_{*,k,0}$ (the operators $\mathcal{H}_{*,k,l}$ for $l>0$ match the $\mathcal{H}_{k,l}$ from [4, §3] up to the presence of smooth cutoffs in our formulation, as opposed to the rough cutoffs present in [4]). Moreover, the truncation in $\mathcal{H}_{*,k,0}$ effects in essence only a substitution of the kernel K by the fixed truncated kernel $\sum_{\ell_0\le\ell} K_\ell$, which can only improve the corresponding estimates in [4, §3].

Short jumps: Proof of (7.6)**.** We begin by reproducing display (4.6) from [4], which reads

$$\sum_{\ell\in\mathbb{Z}} \Big\| \Big(\sum_{j\in\mathbb{Z}} |V_j^r(\mathcal{H}_{\ell-j}f)|^r\Big)^{1/r}\Big\|_{L^p(\mathbb{R}^n)} \lesssim C_{p,r}\|f\|_{L^p(\mathbb{R}^n)}. \tag{7.8}$$

This estimate implies (7.6) with $\mathcal{H}$ in place of $\mathcal{H}_*$. The proof of this estimate for all $p \in (\frac{2n}{2n-1},\infty)$ and $r\in(2,\infty)$ constitutes the main content of [4] and is contained in §4 and §5 there. The key observation is now that, since ℓ^* is non–increasing and varies at most by one on each interval $[2^{\alpha j}, 2^{\alpha(j+1)}]$ (indeed, $\ell^*(2^{\alpha j}) - \ell^*(2^{\alpha(j+1)}) = \beta \le 1$), it holds that

$$V_j^r(\mathcal{H}_{*,\ell-j}f(x)) \le V_j^r(\mathcal{H}_{\ell-j}f(x)) + \sup_{u\in[2^{\alpha j},2^{\alpha(j+1)}]} |\mathcal{H}_{\ell-j}^{(u)}f(x)|. \tag{7.9}$$

By Stein and Wainger's argument (more precisely, by an application of [12, Theorem 1] if α is an even integer; and a similar argument for other α),

$$\sum_{\ell\in\mathbb{Z}} \| \sup_{u\in[2^{\alpha j},2^{\alpha(j+1)}]} |\mathcal{H}_{\ell-j}^{(u)}f|\|_{L^p(\mathbb{R}^n)} \lesssim C_p\|f\|_{L^p(\mathbb{R}^n)}$$

for all $p\in(1,\infty)$, where by complex interpolation we deduce that the constant C_p stays bounded for p contained in any fixed compact subinterval of $(1,\infty)$. Together with (7.8) and (7.9) this yields

$$\sum_{\ell\in\mathbb{Z}} \Big\| \Big(\sum_{j\in\mathbb{Z}} |V_j^r(\mathcal{H}_{*,\ell-j}f)|^r\Big)^{1/r}\Big\|_{L^p(\mathbb{R}^n)} \lesssim (C_p + C_{p,r})\|f\|_{L^p(\mathbb{R}^n)}. \tag{7.10}$$

Suppose that $n\ge 2$. Then (7.8) holds for $p=2$ and as a consequence of complex interpolation we see that $\sup_{r\in[2,3]} C_{2,r} < \infty$. Hence the claimed inequality (7.6) is a direct consequence of (7.10). On the other hand, if $n=1$ then we do not obtain the estimate (7.8) at $p=2$ from [4]. Instead, we follow the same reasoning as in Corollary 9.4 from [4]. Indeed, by (9.9) in [4] we have (7.8) for $p=r\in(2,3)$ with $C_{p,p}\lesssim (p-2)^{-1}$, whence by (7.10),

$$\|S_p(\mathcal{H}_*f)\|_{L^p(\mathbb{R}^n)} \lesssim (p-2)^{-1}\|f\|_{L^p(\mathbb{R}^n)} \tag{7.11}$$

for $p\in(2,3)$. Next, Stein and Wainger's argument (again, reducing to [12, Theorem 1] or a variant thereof) yields

$$\|\sup_{u>0}|\mathcal{H}_*^{(u)}f|\|_{L^q(\mathbb{R}^n)} \lesssim \|f\|_{L^q(\mathbb{R}^n)} \tag{7.12}$$

for every $q \in (1,\infty)$. Interpolating between (7.11) with $p = \frac{r}{4} + \frac{3}{2}$ and (7.12) with $q = \frac{3}{2}$ we obtain the required inequality (7.6).

8. Appendix: Proof of Theorem 2.2

Recall that we are concerned with the multi–frequency maximal function

$$\mathfrak{M}^\theta F(x) = \sup_{\lambda\in\Lambda} \Big| \sum_{j=1}^{M} e(\theta_j \cdot x)[m_\lambda(\mathrm{D})F_j](x)\Big|, \quad (x \in \mathbb{R}^n),$$

acting on vector–valued functions $F : \mathbb{R}^n \to \mathbb{C}^M$. First we assume without loss of generality that $\sigma = 1$ (using scaling invariance of the inequality). Let $B(M)$ denote the best constant in the inequality (2.8). An application of the Cauchy–Schwarz inequality and (2.7) yield the *a priori* estimate $B(M) \leq 2c_0 M^{\frac{1}{2}} < \infty$.

The first step is to exploit the uncertainty principle: by the support assumption (2.3), the function $m_\lambda(\mathrm{D})F_j$ is essentially locally constant at scale 1. We will use this to argue that in order to prove $B(M) \leq K$, it suffices to show

$$\Big\|\Big(\fint_{B_{1/2}} \sup_{\lambda>0} \Big|\sum_{j=1}^{M} e(\theta_j \cdot x)[m_\lambda(\mathrm{D})F_j](x+u)\Big|^2 du\Big)^{1/2}\Big\|_{L^2(\mathbb{R}^n)} \leq \tfrac{1}{2}K\|F\|_{L^2(\mathbb{R}^n;\mathbb{C}^M)}. \tag{8.1}$$

To show this let ψ be a smooth function such that $0 \leq \psi \leq 1$ and ψ is equal to 1 on $B_{1/2}$ and supported in B_1. Then by the support assumption (2.3),

$$m_\lambda(\mathrm{D})F_j = m_\lambda(\mathrm{D})\psi(\mathrm{D})F_j.$$

Thus,

$$m_\lambda(\mathrm{D})F_j(x+u) - m_\lambda(\mathrm{D})F_j(x) = m_\lambda(\mathrm{D})[\psi(D)(e(u\cdot\mathrm{D})-1)F_j](x)$$

Moreover, for every $\xi \in \mathbb{R}^d$ and $u \in B_{1/2}$,

$$|\psi(\xi)(e(\xi\cdot u)-1)| \leq \tfrac{1}{2}.$$

As a consequence, for every fixed $u \in B_{1/2}$ we obtain

$$\|\mathfrak{M}^\theta F\|_{L^2(\mathbb{R}^n)} \leq \Big\|\sup_{\lambda>0}\Big|\sum_{j=1}^{M} e(\theta_j\cdot x)[m_\lambda(\mathrm{D})F_j](x+u)\Big|\Big\|_{L^2(x\in\mathbb{R}^n)} + \tfrac{1}{2}B(M)\|F\|_{L^2(\mathbb{R}^n;\mathbb{C}^M)}.$$

Averaging over $u \in B_{1/2}$ we see that (8.1) implies $B(M) \leq \frac{1}{2}K + \frac{1}{2}B(M)$ which implies $B(M) \leq K$ since we already know that $B(M) < \infty$. Thus we are now reduced to estimating the left–hand side of (8.1). By Fubini's theorem and the change of variables $x \mapsto x - u$ we see that the left–hand side of (8.1) is equal to the $L^2(\mathbb{R}^n)$ norm of

$$\Big(\fint_{B_{1/2}} \sup_{\lambda>0}\Big|\sum_{j=1}^{M} e(-\theta_j\cdot u)e(\theta_j\cdot x)[m_\lambda(\mathrm{D})F_j](x)\Big|^2 du\Big)^{1/2} \tag{8.2}$$

We now continue estimating this quantity for a fixed $x \in \mathbb{R}^n$. We rewrite (8.2) as

$$C\Big\|\sup_{a\in A_x}\Big|\sum_{j=1}^{M} e(\theta_j\cdot u)a_j\Big|\Big\|_{L^2(u\in B_{1/2})}, \tag{8.3}$$

where we have set

$$A_x = \{(e(\theta_j x)[m_\lambda(\mathrm{D})F_j](x))_{j=1,\ldots,M} \,:\, \lambda\in\Lambda\} \subset \mathbb{C}^M.$$

For a set $A \subset \mathbb{C}^M$ and $t \in (0, \infty)$ we define the t–entropy number $E(A,t)$ as the minimum number of t-balls required to cover A whenever $t \le \operatorname{diam}(A)$ and $E(A,t) = 0$ if $t > \operatorname{diam}(A)$. The following is the key estimate.

Lemma 8.1. *Let $A \subset \mathbb{C}^M$. Then for every $a^* \in A$,*

$$\big\| \sup_{a\in A} \big| \sum_{j=1}^M e(\theta_j \cdot u) a_j \big| \big\|_{L^2(u\in B_{1/2})} \le C\big(|a^*| + \int_0^\infty \min(M^{\frac12}, E(A,t)^{\frac12}) dt\big), \tag{8.4}$$

where $C \in (0,\infty)$ is a constant only depending on n. (Note that while the right–hand side is well–defined for all $A \subset \mathbb{C}^M$, it equals infinity unless the diameter of A is finite.)

We postpone the proof of this lemma to the end of this section and first show how it can be used to finish the proof of (2.8). Using (8.4) we can estimate (8.3) by

$$\le C\big(\big(\sum_{j=1}^M |m_{\lambda_0}(\mathrm{D})F_j(x)|^2\big)^{1/2} + \int_0^{R_x} \min(M^{\frac12}, E(A_x,t)^{\frac12}) dt\big),$$

where we have fixed an arbitrary $\lambda_0 \in \Lambda$ and defined

$$R_x = 2 \sup_{a\in A_x} |a| = 2 \sup_{\lambda\in\Lambda} \big(\sum_{j=1}^M |m_\lambda(\mathrm{D})F_j(x)|^2\big)^{1/2}.$$

Note that R_x is finite for almost every $x \in \mathbb{R}^n$ by (2.7) and $E(A_x,t) = 0$ for $t > R_x$ by definition since the diameter of A_x is at most R_x. We now estimate

$$\int_0^{R_x} \min(M^{\frac12}, E(A_x,t)^{\frac12}) dt \le \int_0^{R_x M^{-1/2}} M^{1/2} dt + \int_{R_x M^{-1/2}}^{R_x} \min(M^{\frac12}, E(A_x,t)^{\frac12}) dt. \tag{8.5}$$

Now choose $r > 2$ close enough to 2 so that $1/2 \le M^{\frac12 - \frac1r} \le 1$. Then we use a geometric mean to estimate

$$\min(M^{\frac12}, E(A_x,t)^{\frac12}) \le M^{\frac12-\frac1r} E(A_x,t)^{\frac1r} \le E(A_x,t)^{\frac1r}.$$

Consequently, (8.5) may be estimated as

$$\le R_x + [\sup_{t>0} tE(A_x,t)^{\frac1r}] \int_{R_x M^{-1/2}}^{R_x} \frac{dt}{t} \le R_x + \tfrac12 (\log M)[\sup_{t>0} tE(A_x,t)^{\frac1r}].$$

Note that by (2.7) it holds that

$$\|R_x\|_{L^2(x\in\mathbb{R}^n)} \le 4c_0 \|F\|_{L^2(\mathbb{R}^n;\mathbb{C}^M)}.$$

It remains to observe that for every $t > 0$ and $x \in \mathbb{R}^n$,

$$tE(A_x,t)^{\frac1r} \le \big(\sum_{j=1}^M |V^r\{m_\lambda(\mathrm{D})F_j(x) \,:\, \lambda\in\Lambda\}|^2\big)^{1/2}.$$

Thus (2.6) yields

$$\|\sup_{t>0} tE(A_x,t)^{\frac1r}\|_{L^2(x\in\mathbb{R}^n)} \le c_0 (r-2)^{-\gamma} \|F\|_{L^2(\mathbb{R}^n;\mathbb{C}^M)}. \tag{8.6}$$

By choice of r we have $(r-2)^{-\gamma} \le C(\log M)^\gamma$. Altogether we proved that the left–hand side of (8.1) is

$$\le C \cdot c_0 (\log M)^{1+\gamma} \|F\|_{L^2(\mathbb{R}^n;\mathbb{C}^M)},$$

as claimed.

It remains to prove Lemma 8.1. We first prove the following easier estimate.

Lemma 8.2. *For every $A \subset \mathbb{C}^M$,*

$$\Big\| \sup_{a\in A} \Big| \sum_{j=1}^{M} e(\theta_j \cdot u) a_j \Big| \Big\|_{L^2(u\in B_{1/2})} \leq C \min(M^{\frac{1}{2}}, (\#A)^{\frac{1}{2}}) \sup_{a\in A} |a|, \tag{8.7}$$

where $\#A$ denotes the cardinality of A (which may equal infinity) and C is a constant only depending on n.

Proof. First, an application of the Cauchy–Schwarz inequality yields

$$\Big| \sum_{j=1}^{M} e(\theta_j \cdot u) a_j \Big| \leq M^{\frac{1}{2}} |a|$$

for every $a \in \mathbb{C}^M$. Next, choose a non–negative smooth function χ such that $\mathbf{1}_{B_{1/2}} \lesssim \chi$ and $\widehat{\chi}$ is supported in $B_{1/2}$. Then the left–hand side of (8.7) is

$$\lesssim \Big(\sum_{a\in A} \Big\| \sum_{j=1}^{M} e(\theta_j \cdot u) \chi(u) a_j \Big\|_{L^2(u\in\mathbb{R}^n)}^2 \Big)^{1/2}.$$

By Plancherel's theorem and the separation condition (2.5),

$$\Big\| \sum_{j=1}^{M} e(\theta_j \cdot u) \chi(u) a_j \Big\|_{L^2(u\in\mathbb{R}^n)} = \Big\| \sum_{j=1}^{M} \widehat{\chi}(u - \theta_j) a_j \Big\|_{L^2(u\in\mathbb{R}^n)} \lesssim |a|.$$

□

Proof of Lemma 8.1. It is no loss of generality to assume that A has finite diameter, since otherwise there is nothing to prove. Fix an arbitrary $a^* \in A$. For every $\ell \in \mathbb{Z}$ there exists a finite set $B^{(\ell)} \subset A - A$ with $\#B^{(\ell)} \leq E(A, 2^\ell)$ such that $|b| \leq 2^{\ell+1}$ for every $b \in B^{(\ell)}$ and every $a \in A$ can be written as

$$a = a^* + \sum_{\ell\in\mathbb{Z}} b^{(\ell)}$$

with $b^{(\ell)} \in B^{(\ell)}$. (To construct the sets $B^{(\ell)}$, choose for each small enough ℓ, $E(A, 2^\ell)$ many balls covering A. Then to each chosen ball of generation ℓ, associate a ball of generation $\ell + 1$ that intersects it and let $B^{(\ell)}$ consist of the differences between the centers of balls of generation ℓ and their associated balls of generation $\ell + 1$.) Then

$$\sum_{j=1}^{M} e(\theta_j \cdot u) a_j = \sum_{j=1}^{M} e(\theta_j \cdot u) a_j^* + \sum_{\ell\in\mathbb{Z}} \sum_{j=1}^{M} e(\theta_j \cdot u) b_j^{(\ell)}$$

From this and (8.7) we obtain

$$\Big\| \sup_{a\in A} \Big| \sum_{j=1}^{M} e(\theta_j \cdot u) a_j \Big| \Big\|_{L^2(u\in[0,1]^n)} \leq C \big(|a^*| + \sum_{\ell\in\mathbb{Z}} 2^\ell \min(M^{\frac{1}{2}}, E(A, 2^\ell)^{\frac{1}{2}} \big).$$

Using monotonicity of $r \mapsto E(A, r)$ this implies (8.4). □

References

[1] G. I. Arkhipov, V. N. Chubarikov, A. A. Karatsuba. *Trigonometric sums in number theory and analysis.* Translated from the 1987 Russian original. De Gruyter Expositions in Mathematics, 39. Berlin, 2004.

[2] J. Bourgain. Pointwise ergodic theorems for arithmetic sets. *Inst. Hautes Études Sci. Publ. Math.*, (69):5–45, 1989. With an appendix by the author, Harry Furstenberg, Yitzhak Katznelson and Donald S. Ornstein.

[3] J. T. Campbell, R. L. Jones, K. Reinhold, M. Wierdl. Oscillation and variation for singular integrals in higher dimensions. *Trans. Amer. Math. Soc.*, 355(5):2115–2137, 2003.

[4] S. Guo, J. Roos, P.-L. Yung. Sharp variation-norm estimates for oscillatory integrals related to Carleson's theorem. Preprint, arXiv:1710.10988, to appear in *Anal. PDE.*

[5] A. D. Ionescu, S. Wainger. L^p boundedness of discrete singular Radon transforms. *J. Amer. Math. Soc.*, 19(2):357–383, 2006.

[6] R. L. Jones, A. Seeger, J. Wright. Strong variational and jump inequalities in harmonic analysis. *Trans. Amer. Math. Soc.*, 360(12):6711–6742, 2008.

[7] B. Krause. Discrete Analogues in Harmonic Analysis: Maximally Monomially Modulated Singular Integrals Related to Carleson's Theorem. Preprint, arXiv:1803.09431v3.

[8] B. Krause, M. Lacey. A discrete quadratic Carleson theorem on ℓ^2 with a restricted supremum. *Int. Math. Res. Notices*, (10):3180–3208, 2017.

[9] M. Mirek, E. M. Stein, B. Trojan. $\ell^p(\mathbb{Z}^d)$–estimates for discrete operators of Radon type: Maximal functions and vector-valued estimates. *J. Funct. Anal.* (published online), 2018.

[10] ———. $\ell^p(\mathbb{Z}^d)$–estimates for discrete operators of Radon type: variational estimates. *Invent. Math.*, 209(3):665–748, 2017.

[11] E. M. Stein, S. Wainger. Discrete analogues in harmonic analysis. I. ℓ^2 estimates for singular Radon transforms. *Amer. J. Math.*, 121(6):1291–1336, 1999.

[12] ———. Oscillatory integrals related to Carleson's theorem. *Math. Res. Lett.*, 8(5-6):789–800, 2001.

Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Dr, Madison, WI-53706, USA
E-mail address: `jroos@math.wisc.edu`