Invent. math. (2023) 231:1023–1140
https://doi.org/10.1007/s00222-022-01159-0

Polynomial averages and pointwise ergodic theorems on nilpotent groups

Alexandru D. Ionescu[1] · **Ákos Magyar**[2] · **Mariusz Mirek**[3,4] · **Tomasz Z. Szarek**[4,5]

Dedicated to the memory of Elias M. Stein our friend and advisor.

Received: 10 December 2021 / Accepted: 29 August 2022 /
Published online: 27 September 2022

Abstract We establish pointwise almost everywhere convergence for ergodic averages along polynomial sequences in nilpotent groups of step two of

The first, second and third authors were supported in part by NSF Grants DMS-2007008 and DMS-1600840 and DMS-2154712 respectively. The third author was also partially supported by the Department of Mathematics at Rutgers University and by the National Science Centre in Poland, Grant Opus 2018/31/B/ST1/00204. The fourth author was partially supported by the National Science Centre of Poland, Grant Opus 2017/27/B/ST1/01623, the Juan de la Cierva Incorporación 2019, Grant Number IJC2019-039661-I, the Agencia Estatal de Investigación, Grant PID2020-113156GB-I00/AEI/10.13039/501100011033, the Basque Government through the BERC 2022-2025 program, and by the Spanish Ministry of Sciences, Innovation and Universities: BCAM Severo Ochoa accreditation SEV-2017-0718.

✉ Mariusz Mirek
mariusz.mirek@rutgers.edu

Alexandru D. Ionescu
aionescu@math.princeton.edu

Ákos Magyar
magyar@math.uga.edu

Tomasz Z. Szarek
tzszarek@bcamath.org

[1] Princeton University, Princeton, USA

[2] University of Georgia, Athens, USA

[3] Rutgers University, New Brunswick, USA

[4] Instytut Matematyczny, Uniwersytet Wrocławski, Wrocław, Poland

[5] BCAM - Basque Center for Applied Mathematics, Bilbao, Spain

measure-preserving transformations on σ-finite measure spaces. We also establish corresponding maximal inequalities on L^p for $1 < p \leq \infty$ and ρ-variational inequalities on L^2 for $2 < \rho < \infty$. This gives an affirmative answer to the Furstenberg–Bergelson–Leibman conjecture in the linear case for all polynomial ergodic averages in discrete nilpotent groups of step two. Our proof is based on almost-orthogonality techniques that go far beyond Fourier transform tools, which are not available in the non-commutative, nilpotent setting. In particular, we develop what we call a *nilpotent circle method* that allows us to adapt some of the ideas of the classical circle method to the setting of nilpotent groups.

Contents

1 Introduction

1.1 The Furstenberg–Bergelson–Leibman conjecture

Assume that $(X, \mathcal{B}(X), \mu)$ denotes a σ-finite measure space. Let $\mathbb{Z}[\mathrm{n}]$ denote the space of all polynomials $P(\mathrm{n})$ with one indeterminate n and integer coefficients. Given any family of invertible measure-preserving transformations $T_1, \ldots, T_d : X \to X, d \geq 1$, a measurable function $f \in L^p(X)$, $p \geq 1$, polynomials $P_1, \ldots, P_d \in \mathbb{Z}[\mathrm{n}]$, and an integer $N \geq 1$, we define the polynomial ergodic averages

$$A^{P_1,\ldots,P_d}_{N;X,T_1,\ldots,T_d}(f)(x) := \frac{1}{|[-N,N]\cap\mathbb{Z}|}\sum_{n\in[-N,N]\cap\mathbb{Z}} f(T_1^{P_1(n)}\cdots T_d^{P_d(n)}x), \quad x \in X. \tag{1.1}$$

A fundamental problem in ergodic theory is to establish convergence in norm and pointwise almost everywhere for the polynomial ergodic averages

(1.1) as $N \to \infty$ for functions $f \in L^p(X)$, $1 \leq p \leq \infty$. The problem goes back to at least the early 1930's with von Neumann's mean ergodic theorem [55] and Birkhoff's pointwise ergodic theorem [10] and led to profound extensions such as Bourgain's polynomial pointwise ergodic theorem [11–13] and Furstenberg's ergodic proof [24] of Szemerédi's theorem [53] in particular. Furstenberg's proof was also the starting point of ergodic Ramsey theory, which resulted in many natural generalizations of Szemerédi's theorem, including a polynomial Szemerédi theorem of Bergelson and Leibman [7] that motivates the following far reaching conjecture:

Conjecture 1.1 (Furstenberg–Bergelson–Leibman conjecture [8, Section 5.5, p. 468]) *Given integers $d, k, m, N \in \mathbb{Z}_+$, let $T_1, \ldots, T_d : X \to X$ be a family of invertible measure-preserving transformations of a probability measure space $(X, \mathcal{B}(X), \mu)$ that generates a nilpotent group of step k. Assume that $P_{1,1}, \ldots, P_{i,j}, \ldots, P_{d,m} \in \mathbb{Z}[\mathrm{n}]$ are such that $P_{i,j}(0) = 0$. Then for any $f_1, \ldots, f_m \in L^\infty(X)$, the non-conventional multiple polynomial averages*

$$A^{P_{1,1},\ldots,P_{d,m}}_{N;X,T_1,\ldots,T_d}(f_1, \ldots, f_m)(x) = \frac{1}{|[-N, N] \cap \mathbb{Z}|} \sum_{n \in [-N,N] \cap \mathbb{Z}} \prod_{j=1}^{m} f_j(T_1^{P_{1,j}(n)} \cdots T_d^{P_{d,j}(n)} x) \tag{1.2}$$

converge for μ-almost every $x \in X$ as $N \to \infty$.

Conjecture 1.1 is a major open problem in ergodic theory that was promoted in person by Furstenberg, see [1, p. 6662] and [36], before being published in [8]. Bergelson–Leibman [8] showed that convergence may fail if the transformations $T_1, \ldots, T_d$ generate a solvable group, so the nilpotent setting is probably the appropriate setting for Conjecture 1.1. Our main goal in this paper is to establish this conjecture in the linear $m = 1$ setting in the case when $T_1, \ldots, T_d$ generate a nilpotent group of step two.

A few remarks about this conjecture and the current state of the art are in order.

1. The averages (1.2) are multilinear generalizations of the averages (1.1) in the case $m = 1$ and $P_{j,1} = P_j$ for all $j \in \{1, \ldots, d\}$. The basic case $d = k = m = 1$ with $P_{1,1}(n) = n$ follows from Birkhoff's ergodic theorem [10].
2. The case $d = k = m = 1$ with an arbitrary polynomial $P_{1,1} \in \mathbb{Z}[\mathrm{n}]$ was a famous open problem of Bellow [3] and Furstenberg [25] solved by Bourgain in his breakthrough papers [11–13].
3. Some particular examples of averages (1.2) with $m = 1$ and polynomial mappings with degree at most two in the step two nilpotent setting were studied in [32,43].

4. The multilinear theory, in contrast to the commutative linear theory, is widely open. Only a few results in the bilinear $m = 2$ and commutative $d = k = 1$ setting are known. Bourgain [14] proved pointwise convergence when $P_{1,1}(n) = an$ and $P_{1,2}(n) = bn$, $a, b \in \mathbb{Z}$. More recently, the third author with Krause and Tao [38] established pointwise convergence for the polynomial Furstenberg–Weiss averages [26,27] corresponding to $P_{1,1}(n) = n$ and $P_{1,2}(n) = P(n)$, $\deg P \geq 2$.
5. Except for these few cases, there are no other results concerning pointwise convergence for the averages (1.2). The situation is completely different, however, for the question of norm convergence. A breakthrough paper of Walsh [56] (see also [1]) gives a complete picture of $L^2(X)$ norm convergence of the averages (1.2) for any $T_1, \ldots, T_d \in \mathbb{G}$ where $\mathbb{G}$ is a nilpotent group of transformations of a probability space. Prior to this, there was an extensive body of research towards establishing $L^2(X)$ norm convergence, including groundbreaking works of Host–Kra [28], Ziegler [57], Bergelson [4], and Leibman [40]. See also [2,20,23,29,54] and the survey articles [5,6,22] for more details and references, including a comprehensive historical background.

1.2 Statement of the main results

We can now state the main result of this paper.

Theorem 1.2 (Main result) *Let $d_1 \in \mathbb{Z}_+$ be given and let $T_1, \ldots, T_{d_1} : X \to X$ be a family of invertible measure-preserving transformations of a σ-finite measure space $(X, \mathcal{B}(X), \mu)$ that generates a nilpotent group of step two. Assume that $P_1, \ldots, P_{d_1} \in \mathbb{Z}[\mathrm{n}]$ are such that $P_j(0) = 0$, $1 \leq j \leq d_1$, and let $d_2 := \max\{\deg P_j : j \in \{1, \ldots, d_1\}\}$. Assume $f \in L^p(X)$, $1 \leq p \leq \infty$, and let $A_{N;X}^{P_1,\ldots,P_{d_1}}(f) = A_{N;X,T_1,\ldots,T_{d_1}}^{P_1,\ldots,P_{d_1}}(f)$ be the averages defined in (1.1).*

(i) *(Mean ergodic theorem) If $1 < p < \infty$, then the averages $A_{N;X}^{P_1,\ldots,P_{d_1}}(f)$ converge in the $L^p(X)$ norm as $N \to \infty$.*
(ii) *(Pointwise ergodic theorem) If $1 < p < \infty$, then the averages $A_{N;X}^{P_1,\ldots,P_{d_1}}(f)$ converge pointwise almost everywhere as $N \to \infty$.*
(iii) *(Maximal ergodic theorem) If $1 < p \leq \infty$, then one has*

$$\Big\| \sup_{N \in \mathbb{Z}_+} |A_{N;X}^{P_1,\ldots,P_{d_1}}(f)| \Big\|_{L^p(X)} \lesssim_{d_1,d_2,p} \|f\|_{L^p(X)}. \tag{1.3}$$

The implicit constant in (1.3) may depend on d_1, d_2, and p, but is independent of the coefficients of the underlying polynomials.

The restriction $p > 1$ is necessary in the case of nonlinear polynomials as was shown in [15,39]. We provide now a few remarks about Theorem 1.2.

1. Parts (ii) and (iii) of Theorem 1.2 are completely new even in the case $p = 2$ and extend Bourgain's polynomial ergodic theorems [11–13] to the non-commutative nilpotent setting. In particular, Theorem 1.2 (ii) gives an affirmative answer to Conjecture 1.1 for all polynomials $P_1, \dots, P_{d_1} \in \mathbb{Z}[n]$ and all measure-preserving transformations $T_1, \dots, T_{d_1} : X \to X$ generating a nilpotent group of step two. Moreover, Theorem 1.2 gives affirmative answers to [33, Problems 1, 2] for nilpotent groups of step two.
2. If $(X, \mathcal{B}(X), \mu)$ is a probability space and the family of measure preserving transformations $(T_1, \dots, T_{d_1})$ is totally ergodic, then Theorem 1.2(ii) implies that

$$\lim_{N\to\infty} A_{N;X}^{P_1,\dots,P_{d_1}}(f)(x) = \int_X f(y)d\mu(y) \tag{1.4}$$

μ-almost everywhere on X. We recall that a family of measure preserving transformations $(T_1, \dots, T_{d_1})$ is called *ergodic* on X if $T_j^{-1}(B) = B$ for all $j \in \{1, \dots, d_1\}$ implies $\mu(B) = 0$ or $\mu(B) = 1$ and is called *totally ergodic* if the family $(T_1^n, \dots, T_{d_1}^n)$ is ergodic for all $n \in \mathbb{Z}_+$. In view of (1.4), we see that the polynomial orbits

$$\mathcal{O}_x := \left\{T_1^{P_1(n)} \cdots T_{d_1}^{P_{d_1}(n)} x : \ n \in \mathbb{Z}\right\}$$

have a limiting distribution and, in fact, are uniformly distributed for μ-almost every $x \in X$ when the family $(T_1, \dots, T_{d_1})$ is totally ergodic.
3. The conclusion of the mean ergodic Theorem 1.2(i) follows from [56] if $(X, \mathcal{B}(X), \mu)$ has finite measure, but our proof allows one to deal with the more general σ-finite setting.

1.3 The universal step-two group $\mathbb{G}_0$

The proof of Theorem 1.2 will follow from our second main result, Theorem 1.3 below, for averages on universal nilpotent groups of step two. We start with some definitions. For integers $d \geq 1$, we define

$$Y_d := \{(l_1, l_2) \in \mathbb{Z} \times \mathbb{Z} : 0 \leq l_2 < l_1 \leq d\}$$

and the "universal" step-two nilpotent Lie groups $\mathbb{G}_0^{\#} = \mathbb{G}_0^{\#}(d)$

$$\mathbb{G}_0^{\#} := \{(x_{l_1 l_2})_{(l_1,l_2)\in Y_d} : x_{l_1 l_2} \in \mathbb{R}\}, \tag{1.5}$$

with the group multiplication law

$$[x \cdot y]_{l_1 l_2} := \begin{cases} x_{l_1 0} + y_{l_1 0} & \text{if } l_1 \in \{1, \dots, d\} \text{ and } l_2 = 0, \\ x_{l_1 l_2} + y_{l_1 l_2} + x_{l_1 0} y_{l_2 0} & \text{if } l_1 \in \{1, \dots, d\} \text{ and } l_2 \in \{1, \dots, l_1 - 1\}. \end{cases} \tag{1.6}$$

Alternatively, we can also define the group $\mathbb{G}_0^{\#}$ as the set of elements

$$g = (g^{(1)}, g^{(2)}), \quad g^{(1)} = (g_{l_1 0})_{l_1 \in \{1,\dots,d\}} \in \mathbb{R}^d, \\ g^{(2)} = (g_{l_1 l_2})_{(l_1, l_2) \in Y_d'} \in \mathbb{R}^{d'}, \tag{1.7}$$

where $d' := d(d-1)/2$ and $Y_d' := \{(l_1, l_2) \in Y_d : l_2 \geq 1\}$. Letting

$$R_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d'} \quad \text{denote the bilinear form} \quad [R_0(x, y)]_{l_1 l_2} := x_{l_1 0} y_{l_2 0}, \tag{1.8}$$

we notice that the product rule in the group $\mathbb{G}_0^{\#}$ is given by

$$[g \cdot h]^{(1)} := g^{(1)} + h^{(1)}, \quad [g \cdot h]^{(2)} := g^{(2)} + h^{(2)} + R_0(g^{(1)}, h^{(1)}) \tag{1.9}$$

if $g = (g^{(1)}, g^{(2)})$ and $h = (h^{(1)}, h^{(2)})$. For any $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^{\#}$, its inverse is given by

$$g^{-1} = \big(- g^{(1)}, -g^{(2)} + R_0(g^{(1)}, g^{(1)}) \big).$$

The second variable of $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^{\#}$ is called the central variable. Based on the product structure (1.9) of the group $\mathbb{G}_0^{\#}$, it is not difficult to see that $g \cdot h = h \cdot g$ for any $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0^{\#}$ and $h = (0, h^{(2)}) \in \mathbb{G}_0^{\#}$.

Let $\mathbb{G}_0 = \mathbb{G}_0(d)$ denote the discrete subgroup

$$\mathbb{G}_0 := \mathbb{G}_0^{\#} \cap \mathbb{Z}^{|Y_d|}. \tag{1.10}$$

Let $A_0 : \mathbb{R} \to \mathbb{G}_0^{\#}$ denote the canonical polynomial map (or the moment curve on $\mathbb{G}_0^{\#}$)

$$[A_0(x)]_{l_1 l_2} := \begin{cases} x^{l_1} & \text{if } l_2 = 0, \\ 0 & \text{if } l_2 \neq 0, \end{cases} \tag{1.11}$$

and notice that $A_0(\mathbb{Z}) \subseteq \mathbb{G}_0$. For $x = (x_{l_1 l_2})_{(l_1,l_2)\in Y_d} \in \mathbb{G}_0^{\#}$ and $\Lambda \in (0,\infty)$, we define

$$\Lambda \circ x := (\Lambda^{l_1+l_2} x_{l_1 l_2})_{(l_1,l_2)\in Y_d} \in \mathbb{G}_0^{\#}. \tag{1.12}$$

Notice that the dilations $\Lambda\circ$ are group homomorphisms on the group $\mathbb{G}_0$ that are compatible with the map A_0, i.e. $\Lambda \circ A_0(x) = A_0(\Lambda x)$.

Let $\chi : \mathbb{R} \to [0,1]$ be a smooth function supported on the interval $[-2,2]$. Given any real number $N \geq 1$ and a finitely supported function $f : \mathbb{G}_0 \to \mathbb{C}$, we can define a smoothed average along the moment curve A_0 by the formula

$$M_N^{\chi}(f)(x) := \sum_{n\in\mathbb{Z}} N^{-1}\chi(N^{-1}n) f(A_0(n)^{-1}\cdot x), \qquad x \in \mathbb{G}_0. \tag{1.13}$$

The main advantage of working on the group $\mathbb{G}_0$ with the polynomial map A_0 is the presence of the compatible dilations $\Lambda\circ$ defined in (1.12), which lead to a natural family of associated balls. This can be efficiently exploited by noting that M_N^{χ} is a convolution operator on $\mathbb{G}_0$.

The convolution of functions on the group $\mathbb{G}_0$ is defined by the formula

$$(f * g)(x) := \sum_{y\in\mathbb{G}_0} f(y^{-1}\cdot x) g(y) = \sum_{z\in\mathbb{G}_0} f(z) g(x\cdot z^{-1}). \tag{1.14}$$

Then it is not difficult to see that $M_N^{\chi}(f)(x) = f * G_N^{\chi}(x)$, where

$$G_N^{\chi}(x) := \sum_{n\in\mathbb{Z}} N^{-1}\chi(N^{-1}n)\mathbb{1}_{\{A_0(n)\}}(x), \quad x \in \mathbb{G}_0. \tag{1.15}$$

We are now ready to state our second main result.

Theorem 1.3 (Boundedness on $\mathbb{G}_0$) *Let $\mathbb{G}_0 = \mathbb{G}_0(d)$, $d \geq 1$, be the discrete nilpotent group defined in (1.10). For any $f \in \ell^p(\mathbb{G}_0)$, $1 \leq p \leq \infty$, let $M_N^{\chi}(f)$ be the average defined in (1.13) with a smooth function $\chi : \mathbb{R} \to [0,1]$ supported on the interval $[-2,2]$.*

(i) *(Maximal estimates) If $1 < p \leq \infty$, then one has*

$$\big\| \sup_{N\geq 1} |M_N^{\chi}(f)| \big\|_{\ell^p(\mathbb{G}_0)} \lesssim_{d,p,\chi} \|f\|_{\ell^p(\mathbb{G}_0)}. \tag{1.16}$$

(ii) *(Long variational estimates) If $1 < p < \infty$, $\rho > \max\{p, \frac{p}{p-1}\}$, and $\tau \in (1,2]$, then*

$$\big\| V^{\rho}\big(M_N^{\chi}(f) : N \in \mathbb{D}_{\tau}\big) \big\|_{\ell^p(\mathbb{G}_0)} \lesssim_{d,p,\rho,\tau,\chi} \|f\|_{\ell^p(\mathbb{G}_0)}, \tag{1.17}$$

where $\mathbb{D}_\tau := \{\tau^n : n \in \mathbb{N}\}$. *See* (2.3) *for the definition of the* ρ*-variation seminorms* V^ρ.

Some comments are in order.

1. Theorem 1.3 will be used to prove Theorem 1.2. The main tool in this reduction will be the Calderón transference principle [16], and the details will be given in Sect. 3.
2. Theorem 1.3 extends the results of [44,47] to the non-commutative, nilpotent setting. Its conclusions remain true for rough averages, i.e. when $\chi = \mathbb{1}_{[-1,1]}$ in (1.13), but it is more convenient to work with smooth averages.
3. The restriction $p > 1$ in Theorem 1.3 is sharp due to [15,39]. However, the range of $\rho > \max\{p, \frac{p}{p-1}\}$ is only sharp when $p = 2$ due to Lépingle's inequality [41]. One could hope to improve this to the full range $\rho > 2$ for exponents $p \neq 2$, but only at the expense of additional complexity in the proof. We do not address this here since the limited range $\rho > \max\{p, \frac{p}{p-1}\}$ is already sufficient for us to establish Theorem 1.2.

1.4 Overview of the proof

We will show in Sect. 3 that Theorem 1.2 is a consequence of Theorem 1.3 upon performing lifting arguments and adapting the Calderón transference principle. Our main goal therefore is to prove Theorem 1.3, which takes up the bulk of this paper.

Bourgain's seminal papers [11–13] generated a large amount of research and progress in the field. Many other discrete operators have been analyzed by many authors motivated by problems in Analysis and Ergodic Theory. See, for example, [15,32,34,37–39,42–44,46,47,49,50,52] for some results of this type and more references. A common feature of all of these results, which plays a crucial role in the proofs, is that one can use Fourier analysis techniques, in particular, the powerful framework of the classical circle method, to perform the analysis.

Our situation in Theorem 1.3 is different as new difficulties arise. The main issue is that there is no good Fourier transform on nilpotent groups that is compatible with the structure of the underlying convolution operators and at the level of analytical precision of the classical circle method. The second obstacle is the absence of a good delta function compatible with the group multiplication on $(\mathbb{G}_0, \cdot)$ (defined in (1.6)). This prevents us from using a naive implementation of the circle method. The classical delta function

$$\mathbb{1}_{\{0\}}(x^{-1}\cdot y)=\int_{\mathbb{T}^d\times\mathbb{T}^{d'}}\mathfrak{e}((y^{(1)}-x^{(1)}).\theta^{(1)})\mathfrak{e}((y^{(2)}-x^{(2)}).\theta^{(2)})\,d\theta^{(1)}d\theta^{(2)}, \tag{1.18}$$

does not detect the group multiplication correctly, see Sect. 2 and (2.2) for notation.

These two issues lead to very significant difficulties in the proof and require substantial new ideas. We developed the following tools to circumvent these problems:

(i) Classical Fourier techniques will be replaced with almost-orthogonality methods based on exploiting high order TT^* arguments for operators defined on the discrete group $\mathbb{G}_0$ which arise in the proof of Theorem 1.3. Studying high powers of TT^* (i.e. $(TT^*)^r$ for a large $r\in\mathbb{Z}_+$) allows for a simple heuristic lying behind the proof of Waring-type problems to be used efficiently (and rigorously) in the context of our proof. This heuristic says that, the more variables that occur in the Waring-type equation, the easier is to find a solution. Manipulating the parameter r (usually taking r to be very large), we can always decide how many variables we have at our disposal, making the operators in our questions "smoother and smoother".

(ii) Our main new construction in this paper is what we call a *nilpotent circle method*, an iterative procedure, starting from the center of the group and moving down along its central series, that allows us to use some of the ideas of the classical circle method recursively at every stage. In our case of nilpotent groups of step two, the procedure consists of two basic iterations and one additional step corresponding to "major arcs". The key feature of this approach is that it is adapted to the classical delta function as in (1.18). The minor arcs analysis needs two types of Weyl's inequalities: the classical one as well as the nilpotent one in the spirit of Davenport [21] and Birch [9], which was proved in [33]. The major arcs analysis brings into play some tools that combine continuous harmonic analysis on groups $\mathbb{G}_0^{\#}$ with arithmetic harmonic analysis over finite integer rings modulo $Q\in\mathbb{Z}_+$.

We outline the argument in Sect. 1.4.1 below.

1.4.1 A nilpotent circle method and ℓ^2 theory

To illustrate our main iterative procedure, it suffices to consider the boundedness of the maximal function M_N^{χ} on $\ell^2(\mathbb{G}_0)$. We would like to prove that

$$\Big\|\sup_{k\geq 0}|f*G_{2^k}^{\chi}|\Big\|_{\ell^2(\mathbb{G}_0)}\lesssim\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{1.19}$$

Inequality (1.19) involves a genuinely sublinear operator, preventing a naive implementation of high order TT^* arguments. This contrasts sharply with the situation of singular integral operators studied in [33]. We begin with a delicate decomposition of the kernels $G^{\chi}_{2^k}$ adjusted to the nilpotent structure of the underlying group $\mathbb{G}_0$. Notice that these kernels have a product structure

$$\begin{aligned} G^{\chi}_{2^k}(g) &:= L_k(g^{(1)})\mathbb{1}_{\{0\}}(g^{(2)}),\\ L_k(g^{(1)}) &:= \sum_{n\in\mathbb{Z}} 2^{-k}\chi(2^{-k}n)\mathbb{1}_{\{0\}}(g^{(1)} - A^{(1)}_0(n)), \end{aligned} \tag{1.20}$$

where $A^{(1)}_0(n) := (n, \dots, n^d) \in \mathbb{Z}^d$ and $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0$ as in (1.7).

First stage. We start by decomposing the kernels $G^{\chi}_{2^k}$ in the central variable. For any integers $s \geq 0$ and $m \geq 1$, we define the set of rational fractions

$$\begin{aligned} \mathcal{R}^m_s := \ & \{a/q : a = (a_1, \dots, a_m) \in \mathbb{Z}^m,\\ & q \in [2^s, 2^{s+1} - 1] \cap \mathbb{Z},\ \gcd(a_1, \dots, a_m, q) = 1\}. \end{aligned} \tag{1.21}$$

We define also $\mathcal{R}^m_{\leq a} := \bigcup_{0\leq s\leq a} \mathcal{R}^m_s$. For $x^{(1)} = (x^{(1)}_{l_1 0}) \in \mathbb{R}^d$, $x^{(2)} = (x^{(2)}_{l_1 l_2}) \in \mathbb{R}^{d'}$ and $\Lambda \in (0, \infty)$, we define the partial dilations

$$\begin{aligned} \Lambda \circ x^{(1)} &= (\Lambda^{l_1} x^{(1)}_{l_1 0})_{l_1\in\{1,\dots,d\}} \in \mathbb{R}^d,\\ \Lambda \circ x^{(2)} &= (\Lambda^{l_1+l_2} x^{(2)}_{l_1 l_2})_{(l_1,l_2)\in Y'_d} \in \mathbb{R}^{d'}, \end{aligned} \tag{1.22}$$

which are induced by the group-dilations defined in (1.12). We fix a small constant $\delta = \delta(d) \ll 1$, a large constant $D = D(d) \gg \delta^{-8}$, and a smooth even cutoff function $\eta_0 : \mathbb{R} \to [0, 1]$ such that $\mathbb{1}_{[-1,1]} \leq \eta_0 \leq \mathbb{1}_{[-2,2]}$. For $k \geq D^2$ and $s \leq \delta k$, we define the periodic Fourier multipliers

$$\begin{aligned} \Xi_{k,s}(\xi^{(2)}) &:= \sum_{a/q\in\mathcal{R}^{d'}_s} \eta_{\leq \delta k}(2^k \circ (\xi^{(2)} - a/q)),\\ \Xi^c_k &:= 1 - \sum_{s\in[0,\delta k]} \Xi_{k,s}, \end{aligned} \tag{1.23}$$

where $\eta_{\leq\Lambda}(x) := \eta_0(|x|/2^{\lfloor\Lambda\rfloor})$ and $\lfloor\Lambda\rfloor := \max\{n \in \mathbb{Z} : n \leq \Lambda\}$. Then we decompose

$$\begin{aligned} \mathbf{1}_{\{0\}}(g^{(2)}) = & \sum_{s\in[0,\delta k]} \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi_{k,s}(\xi^{(2)})\, d\xi^{(2)}\\ & + \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi^c_k(\xi^{(2)})\, d\xi^{(2)}, \end{aligned} \tag{1.24}$$

where $g^{(2)}.\xi^{(2)}$ denotes the usual scalar product of vectors in $\mathbb{R}^{d'}$ and $\mathfrak{e}(z) := e^{2\pi i z}$. This induces our first stage decomposition $G_{2^k}^{\chi} = K_k^c + \sum_{s\in[0,\delta k]} K_{k,s}$, where, with the notation in (1.20),

$$K_{k,s}(g) := L_k(g^{(1)})N_{k,s}(g^{(2)}), \quad K_k^c(g) := L_k(g^{(1)})N_k^c(g^{(2)}), \tag{1.25}$$

and

$$\begin{aligned} N_{k,s}(g^{(2)}) &:= \eta_{\le\delta k}(2^{-k}\circ g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_{k,s}(\xi^{(2)})\,d\xi^{(2)}, \\ N_k^c(g^{(2)}) &:= \eta_{\le\delta k}(2^{-k}\circ g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_k^c(\xi^{(2)})\,d\xi^{(2)}. \end{aligned} \tag{1.26}$$

The main bounds we prove in the first stage are the *first minor arcs estimate*,

$$\|f * K_k^c\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-k/D^2}\|f\|_{\ell^2(\mathbb{G}_0)} \tag{1.27}$$

for any $k \ge D^2$ and $f \in \ell^2(\mathbb{G}_0)$, and the *first transition estimate*,

$$\Big\| \sup_{\max(D^2,s/\delta)\le k\le\kappa_s} |f * K_{k,s}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)} \tag{1.28}$$

for any $s \ge 0$ and $f \in \ell^2(\mathbb{G}_0)$, $\kappa_s := 2^{2D(s+1)^2}$.

In the commutative setting, minor arcs estimates such as (1.27) follow using Weyl estimates and the Plancherel theorem. As we do not have a useful Fourier transform on the group $\mathbb{G}_0$, our main tool to prove the bounds (1.27) is a high order T^*T argument. More precisely, we analyze the kernel of the convolution operator $\{(\mathcal{K}_k^c)^*\mathcal{K}_k^c\}^r$, where $\mathcal{K}_k^c f := f * K_k^c$ and r is sufficiently large, and show that its $\ell^1(\mathbb{G}_0)$ norm is $\lesssim 2^{-k}$. The main ingredient in this proof is the non-commutative Weyl estimate in Proposition 2.3 (i), which was proved earlier in [33].

To prove the transition estimates (1.28), we apply the Rademacher–Menshov inequality (2.7) with a logarithmic loss to reduce to proving the inequality

$$\Big\| \sum_{k\in[J,2J]} \varkappa_k (f * H_{k,s}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-4s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)} \tag{1.29}$$

for any $J \ge \max(D^2, s/\delta)$ and any coefficients $\varkappa_k \in [-1, 1]$, where $H_{k,s} := K_{k+1,s} - K_{k,s}$. For this, we use a high order version of the Cotlar–Stein lemma, which relies again on precise analysis of the kernel of the convolution operator $\{(\mathcal{H}_{k,s})^*\mathcal{H}_{k,s}\}^r$, where $\mathcal{H}_{k,s} f := f * H_{k,s}$ and r is sufficiently large. The key

exponential gain of $2^{-4s/D^2}$ in (1.29) is due to a non-commutative Gauss sums estimate, see Proposition 2.3 (ii).

Second stage. In view of (1.27)–(1.28) it remains to prove that

$$\Big\| \sup_{k\geq \kappa_s} |f * K_{k,s}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)} \tag{1.30}$$

for any fixed integer $s \geq 0$. For this, we have to decompose the kernels $K_{k,s}$ in the non-central variables. We examine the kernels $L_k(g^{(1)})$ in (1.20) and rewrite them as

$$L_k(g^{(1)}) = \eta_{\leq \delta k}(2^{-k} \circ g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)}) S_k(\xi^{(1)})\, d\xi^{(1)} \tag{1.31}$$

where $g^{(1)}.\xi^{(1)}$ denotes the usual scalar product of vectors in $\mathbb{R}^d$ and

$$S_k(\xi^{(1)}) := \sum_{n\in\mathbb{Z}} 2^{-k} \chi(2^{-k}n) \mathfrak{e}(-A_0^{(1)}(n).\xi^{(1)}). \tag{1.32}$$

For any integers $Q \geq 1$ and $m \geq 1$, we define the set of fractions

$$\widetilde{\mathcal{R}}_Q^m := \{a/Q : a = (a_1, \ldots, a_m) \in \mathbb{Z}^m\}. \tag{1.33}$$

We fix a large denominator $Q_s := (2^{Ds+D})! = 1 \cdot 2 \cdot \ldots \cdot 2^{Ds+D}$ and define the periodic multipliers

$$\begin{aligned}
\Psi_{k,s}^{\mathrm{low}}(\xi^{(1)}) &:= \sum_{a/q\in\widetilde{\mathcal{R}}_{Q_s}^d} \eta_{\leq \delta' k}(2^k \circ (\xi^{(1)} - a/q)),\\
\Psi_{k,s,t}(\xi^{(1)}) &:= \sum_{a/q\in\mathcal{R}_t^d\setminus\widetilde{\mathcal{R}}_{Q_s}^d} \eta_{\leq \delta' k}(2^k \circ (\xi^{(1)} - a/q)),\\
\Psi_k^c(\xi^{(1)}) &:= 1 - \Psi_{k,s}^{\mathrm{low}} - \sum_{t\in[0,\delta' k]} \Psi_{k,s,t}\\
&= 1 - \sum_{a/q\in\mathcal{R}_{\leq \delta' k}^d} \eta_{\leq \delta' k}(2^k \circ (\xi^{(1)} - a/q)),
\end{aligned} \tag{1.34}$$

where $\delta' > \delta$ is a suitable constant and the sets $\mathcal{R}_t^d$ are as in (1.21). Since $k \geq \kappa_s = 2^{2D(s+1)^2}$, it is easy to see that the cutoff functions $\eta_{\leq \delta' k}(2^k \circ (\xi^{(1)} - a/q))$ have disjoint supports and the multipliers $\Psi_{k,s}^{\mathrm{low}}$, $\Psi_{k,s,t}$, Ψ_k^c take values in the interval $[0, 1]$.

We then define the kernels $L^{\mathrm{low}}_{k,s}, L_{k,s,t}, L^{c}_{k} : \mathbb{Z}^d \to \mathbb{C}$ by

$$L_*(g^{(1)}) = \phi_k^{(1)}(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)}) S_k(\xi^{(1)}) \Psi_*(\xi^{(1)})\, d\xi^{(1)}, \quad (1.35)$$

where $\Psi_* \in \{\Psi^{\mathrm{low}}_{k,s}, \Psi_{k,s,t}, \Psi^{c}_{k}\}$, and, finally, our main kernels $G^{\mathrm{low}}_{k,s}, G_{k,s,t}$, $G^{c}_{k,s} : \mathbb{Z}^d \to \mathbb{C}$ by

$$G_*(g) := L_*(g^{(1)}) N_{k,s}(g^{(2)}). \quad (1.36)$$

The estimates we prove at this stage are the *second minor arcs estimate*,

$$\| f * G^{c}_{k,s} \|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-k/D^2} \| f \|_{\ell^2(\mathbb{G}_0)} \quad (1.37)$$

for any $s \geq 0$, $k \geq 2^{2D(s+1)^2}$, and $f \in \ell^2(\mathbb{G}_0)$, and the *second transition estimate*,

$$\Big\| \sup_{\max(\kappa_s, t/\delta) \leq k \leq \kappa_t} |f * G_{k,s,t}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-t/D^2} \| f \|_{\ell^2(\mathbb{G}_0)} \quad (1.38)$$

for any $s \geq 0$, $t \geq Ds + D$, and $f \in \ell^2(\mathbb{G}_0)$, where $\kappa_t := 2^{2D(t+1)^2}$.

The proofs of these estimates are similar to the proofs of the corresponding first stage estimates (1.27)–(1.28), using high order T^*T arguments. Surprisingly, instead of using the non-commutative oscillatory sums estimates in Proposition 2.3, we only use the classical ones from Proposition 2.2 here. We emphasize, however, that the underlying nilpotent structure is very important and that these estimates are only possible after performing the two reductions in the first stage, namely, the restriction to major arcs corresponding to denominators $\simeq 2^s$ and the restriction to parameters $k \geq \kappa_s$. We finally remark that, if we applied the circle method simultaneously to both central and non-central variables, we would encounter serious difficulties that do not allow for an efficient control of the phase functions arising in the corresponding exponential sums and oscillatory integrals, especially on major arcs.

Final stage. After these reductions, it remains to bound the contributions of the "major arcs" in both the central and the non-central variables. More precisely, we prove the bounds

$$\begin{aligned} \Big\| \sup_{k \geq \kappa_s} |f * G^{\mathrm{low}}_{k,s}| \Big\|_{\ell^2(\mathbb{G}_0)} &\lesssim 2^{-s/D^2} \| f \|_{\ell^2(\mathbb{G}_0)}, \\ \Big\| \sup_{k \geq \kappa_t} |f * G_{k,s,t}| \Big\|_{\ell^2(\mathbb{G}_0)} &\lesssim 2^{-t/D^2} \| f \|_{\ell^2(\mathbb{G}_0)}, \end{aligned} \quad (1.39)$$

for any $s \geq 0$, $t \geq Ds + D$, and $f \in \ell^2(\mathbb{G}_0)$.

The main idea here is different: we write the kernels $G^{\text{low}}_{k,s}$ and $G_{k,s,t}$ as tensor products of two components up to acceptable errors. One of these components is essentially a maximal average operator on a continuous group, which can be analyzed using the classical method of Christ [17]. The other component is an arithmetic operator-valued analogue of the classical Gauss sums, which generates the key exponential factors $2^{-s/D^2}$ and $2^{-t/D^2}$ in (1.39).

1.4.2 ℓ^p theory and variation norms

The problem of passing from ℓ^2 estimates to ℓ^p estimates in the context of discrete polynomial averages has been investigated extensively in recent years (see, for example, [44] and the references therein), and we will be somewhat brief on this.

The full $\ell^p(\mathbb{G}_0)$ bounds in Theorem 1.3 rely on first proving $\ell^2(\mathbb{G}_0)$ bounds. In fact, we first establish (1.17) for $p = 2$ and $\rho > 2$, by following essentially the steps described above. Then we use the positivity of the operators M^{χ}_N (i.e. $M^{\chi}_N(f) \geq 0$ if $f \geq 0$) to prove the maximal operator bounds (1.16) for all $p \in (1, \infty]$. Finally, we use vector-valued interpolation between the bounds (1.17) with $p = 2$ and $\rho > 2$ and (1.16) with $p \in (1, \infty]$ to complete the proof of Theorem 1.3.

A new ingredient, which is interesting in its own right, is Proposition 9.4, which provides $\ell^p(\mathbb{H}_Q)$ bounds for the so-called shifted maximal inequality, see [51, Section 5.10, p. 78] as well as [48, Section 4.2.4, p. 148] for similar results in the commutative setting. Tools of these kinds are not apparent in the commutative theory as the delta function (1.18) correctly detects the underlying convolution structure. In our case, as we mentioned above, there is no delta function that would be compatible with the convolution structure on $\mathbb{G}_0$. This is a serious obstruction, which forced us to establish Proposition 9.4. This completes the outline of the proof of Theorem 1.3.

1.4.3 General nilpotent groups

The primary goal is, of course, to establish the full Conjecture 1.1 in the linear $m = 1$ case for arbitrary invertible measure-preserving transformations $T_1, \ldots, T_d$ that generate a nilpotent group of any step $k \geq 2$. The iterative argument we have outlined in Sect. 1.4.1 could, in principle, be extended to higher step groups, at least as long as the group and the polynomial sequence have suitable "universal"-type structure, as one could try to go down along the central series of the group and prove minor arcs and transition estimates at every stage.

However, this is only possible if one can prove suitable analogues of the nilpotent Weyl's inequalities in Proposition 2.3 on general nilpotent groups of step $k \geq 3$. The point is to have a small (not necessarily optimal, but nontrivial) gain for bounds on oscillatory sums over many variables, corresponding to the kernels of high power $(T^*T)^r$ operators, whenever frequencies are restricted to the minor arcs. In our case, the formulas are explicit, see the identities (2.23), and we can use ideas of Davenport [21] and Birch [9] for Diophantine forms in many variables to control the induced oscillatory sums, but the analysis seems to be more complicated for the higher step nilpotent groups. This is an interesting problem in its own right, corresponding to Waring-type problems on nilpotent groups, which may be interpreted as a question about solutions of suitable systems of Diophantine equations induced by the moment curve on $\mathbb{G}_0$. A qualitative variant of the Waring problem in the context of nilpotent groups was recently investigated in [30,31], see also the references given there.

Nevertheless, we hope that the methods of the proof of Theorem 1.3 will be useful to establish a quantitative variant of the Waring problem on $\mathbb{G}_0$ in the spirit of the asymptotic formula of Hardy and Littlewood as in the classical Waring problem. We plan to investigate this question as well as its connections with Conjecture 1.1 in the near future.

1.5 Organization

In Sect. 2, we summarize our main notation and collect some important lemmas. In Sect. 3, we show how to use the conclusions of Theorem 1.3 to prove Theorem 1.2. In Sect. 4, we outline the main $\ell^2(\mathbb{G}_0)$ argument in the proof of Theorem 1.3 and divide this argument into five lemmas. In Sects. 5, 6, 7, and 8, we prove these lemmas, starting with the minor arcs estimates in Lemmas 4.2 and 4.4, the major arcs estimates in Lemma 4.5, and the (more difficult) transition estimates in Lemmas 4.3 and 4.6 . In Sect. 9, we prove the maximal $\ell^p(\mathbb{G}_0)$ estimates (1.16), $p \in (1, \infty)$, using some of the more technical estimates in Appendices A and B.

2 Notation and preliminaries

In this section we set up most of our notation and state some important lemmas that will be used in the rest of the paper.

2.1 Basic notation

The sets of positive integers and nonnegative integers will be denoted by $\mathbb{Z}_+ := \{1, 2, \ldots\}$ and $\mathbb{N} := \{0, 1, 2, \ldots\}$. For $d \in \mathbb{Z}_+$ the sets $\mathbb{Z}^d$, $\mathbb{R}^d$, $\mathbb{C}^d$

and $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ have standard meaning. We denote $\mathbb{R}_+ := (0,\infty)$ and $\mathbb{Z}_q := \{1,\dots,q\}$ for $q \in \mathbb{Z}_+$.

For any $x \in \mathbb{R}$ we let $\lfloor x \rfloor$ denote its integer part, $\lfloor x \rfloor := \max\{n \in \mathbb{Z} : n \le x\}$. For any $a \in \mathbb{C}^d$ we will use the Japanese bracket notation $\langle a\rangle := (1+|a|^2)^{1/2}$. For any sequence $(a_k)_{k\in\mathbb{Z}}$ of complex numbers we define the difference operator by

$$\Delta_k a_k := a_{k+1} - a_k. \tag{2.1}$$

We use $\mathbb{1}_A$ to denote the indicator function of a set A. We let $C > 0$ denote general constants which 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 $A \le CB$. We will write $A \simeq B$ when $A \lesssim B \lesssim A$. For two quantities A, B we will use $A \ll B$ to indicate that there is a small constant $C > 0$ such that $|A| \le CB$. We will write $\lesssim_\delta$ or $\simeq_\delta$ or $\ll_\delta$ to emphasize that the implicit constants may depend on the parameter δ.

2.1.1 Function spaces

For an open set $U \subseteq \mathbb{R}^d$ let $C(U)$ denote the space of continuous functions $f : U \to \mathbb{C}$. Let $C^n(U) \subset C(U)$ denote the space of continuous functions f on U whose partial derivatives of order $\le n \in \mathbb{Z}_+$ all exist and are continuous, and $C^\infty(U) := \bigcap_{n\in\mathbb{Z}_+} C^n(U)$. The partial derivatives of a function $f : \mathbb{R}^d \to \mathbb{C}$ will be denoted by $\partial_{x_j} f = \partial_j f$; for any multi-index $\alpha \in \mathbb{N}^d$ let $\partial^\alpha f$ denote the derivative operator $\partial_1^{\alpha_1}\dots\partial_d^{\alpha_d} f$ of total order $|\alpha| := \alpha_1 + \cdots + \alpha_d$.

Given a measure space Y we let $L^p(Y)$, $p \in [1,\infty]$, denote the standard Lebesgue spaces of complex-valued functions on Y. These spaces can be extended to functions taking values in a finite dimensional normed vector space $(B, \|\cdot\|_B)$,

$$L^p(Y;B) := \big\{F : Y \to B \text{ measurable} : \|F\|_{L^p(Y;B)} := \big\|\|F\|_B\big\|_{L^p(Y)} < \infty\big\}.$$

In our case we will usually have $X = \mathbb{G}_0^\#$ or $X = \mathbb{R}^d$ or $X = \mathbb{T}^d$ equipped with the Lebesgue measure, and $X = \mathbb{G}_0$ or $X = \mathbb{Z}^d$ endowed with the counting measure. If X is endowed with counting measure we will shorten $L^p(X)$ to $\ell^p(X)$ and $L^p(X;B)$ to $\ell^p(X;B)$.

2.1.2 The Fourier transform

The standard inner product on $\mathbb{R}^m$, $m \geq 1$, is denoted by

$$x.\xi := \sum_{k=1}^{m} x_k \xi_k \tag{2.2}$$

for every $x = (x_1, \ldots, x_m)$, $\xi = (\xi_1, \ldots, \xi_m) \in \mathbb{R}^m$. Letting $\mathfrak{e}(z) := e^{2\pi i z}$, $z \in \mathbb{C}$, the (Euclidean) Fourier transform and inverse Fourier transform of functions $f \in L^1(\mathbb{R}^m)$ will be denoted by

$$\mathcal{F}_{\mathbb{R}^m} f(\xi) := \int_{\mathbb{R}^m} f(x)\mathfrak{e}(-x.\xi)\, dx, \quad \mathcal{F}^{-1}_{\mathbb{R}^m} f(x) := \int_{\mathbb{R}^m} f(\xi)\mathfrak{e}(x.\xi)\, d\xi.$$

We shall also abbreviate $\widehat{f} = \mathcal{F}_{\mathbb{R}^m} f$.

2.2 ρ-Variations

For any family $(a_t : t \in \mathbb{I})$ of elements of $\mathbb{C}$ indexed by a totally ordered set $\mathbb{I}$, and any exponent $1 \leq \rho < \infty$, the ρ-variation seminorm is defined by

$$V^\rho(a_t)_{t\in\mathbb{I}} = V^\rho(a_t : t \in \mathbb{I}) := \sup_{J\in\mathbb{Z}_+} \sup_{\substack{t_0<\cdots<t_J \\ t_j\in\mathbb{I}}} \Big(\sum_{j=0}^{J-1} |a(t_{j+1}) - a(t_j)|^\rho \Big)^{1/\rho}, \tag{2.3}$$

where the supremum is taken over all finite increasing sequences in $\mathbb{I}$.

It is easy to see that $\rho \mapsto V^\rho$ is non-increasing, and for every $t_0 \in \mathbb{I}$ one has

$$\begin{aligned} \sup_{t\in\mathbb{I}} |a_t| &\leq |a_{t_0}| + V^\rho(a_t : t \in \mathbb{I}) \leq \sup_{t\in\mathbb{I}} |a_t| + V^\rho(a_t : t \in \mathbb{I}) \\ &=: \widetilde{V}^\rho(a_t : t \in \mathbb{I}). \end{aligned} \tag{2.4}$$

Notice that $\widetilde{V}^\rho$ clearly defines a norm on the space of functions from $\mathbb{I}$ to $\mathbb{C}$. Moreover

$$\widetilde{V}^\rho(a_t : t \in \mathbb{I}) \lesssim \widetilde{V}^\rho(a_t : t \in \mathbb{I}_1) + \widetilde{V}^\rho(a_t : t \in \mathbb{I}_2) \tag{2.5}$$

whenever $\mathbb{I} = \mathbb{I}_1 \cup \mathbb{I}_2$ is an ordered partition of $\mathbb{I}$, that is $\max \mathbb{I}_1 = \min \mathbb{I}_2$. Finally, if $\mathbb{I}$ is at most countable, then

$$\widetilde{V}^\rho(a_t : t \in \mathbb{I}) \lesssim \Big(\sum_{t\in\mathbb{I}} |a_t|^\rho \Big)^{1/\rho}. \tag{2.6}$$

We also recall from [46, Lemma 2.5] the Rademacher–Menshov inequality, which asserts that for any $2 \le \rho < \infty$ and $j_0, m \in \mathbb{N}$ so that $j_0 < 2^m$ and any sequence of complex numbers $(\mathfrak{a}_k : k \in \mathbb{N})$ we have

$$V^\rho(\mathfrak{a}_j : j_0 \le j \le 2^m) \le \sqrt{2} \sum_{i=0}^{m} \Big(\sum_{j\in[j_0 2^{-i}, 2^{m-i}-1]\cap\mathbb{Z}} \big|\mathfrak{a}_{(j+1)2^i} - \mathfrak{a}_{j2^i})\big|^2 \Big)^{1/2}. \tag{2.7}$$

Finally, for every family of measurable functions $(a_t : t \in \mathbb{I}) \subseteq \mathbb{C}$ by a slight abuse of notation we continue to write

$$\|V^\rho(a_t : t \in \mathbb{I})\|_{L^p(X)} = \|(a_t)_{t\in\mathbb{I}}\|_{L^p(X;V^\rho)}.$$

2.3 Products and convolutions on the group $\mathbb{G}_0$

We now establish formulas that will be repeatedly used in the proof of Theorem 1.3.

Many of our $\ell^2(\mathbb{G}_0)$ estimates will be based on high order T^*T arguments. Let $S_1, T_1, \dots, S_r, T_r : \ell^2(\mathbb{G}_0) \to \ell^2(G_0)$ be convolution operators defined by some $\ell^1(\mathbb{G}_0)$ kernels $L_1, K_1, \dots, L_r, K_r : \mathbb{G}_0 \to \mathbb{C}$, i.e. $S_j f = f * L_j$ and $T_j f = f * K_j$ for $j \in \{1, \dots, r\}$. Then the adjoint operators $S_1^*, \dots, S_r^*$ are also convolution operators, defined by the kernels $L_1^*, \dots, L_r^*$ given by $L_j^*(g) := \overline{L_j(g^{-1})}$. Moreover, using (1.14), for any $f \in \ell^2(\mathbb{G}_0)$ and $x \in \mathbb{G}_0$, we have

$$\begin{aligned}&(S_1^* T_1 \dots S_r^* T_r f)(x)\\ &= \sum_{h_1,g_1,\dots,h_r,g_r\in\mathbb{G}_0} \Big\{ \prod_{j=1}^{r} L_j^*(h_j) K_j(g_j) \Big\} f(g_r^{-1}\cdot h_r^{-1}\cdot \dots \cdot g_1^{-1}\cdot h_1^{-1}\cdot x).\end{aligned} \tag{2.8}$$

In other words $(S_1^* T_1 \dots S_r^* T_r f)(x) = (f * A^r)(x)$, where the kernel A^r is given by

$$A^r(y) := \sum_{h_1,g_1,\dots,h_r,g_r\in\mathbb{G}_0} \Big\{ \prod_{j=1}^{r} \overline{L_j(h_j)} K_j(g_j) \Big\} \mathbb{1}_{\{0\}}(g_r^{-1}\cdot h_r \cdot \dots \cdot g_1^{-1}\cdot h_1 \cdot y). \tag{2.9}$$

To use these formulas we decompose $h_j = (h_j^{(1)}, h_j^{(2)})$, $g_j = (g_j^{(1)}, g_j^{(2)})$ as in (1.7). Then

$$[h_1^{-1} \cdot g_1 \cdot \ldots \cdot h_r^{-1} \cdot g_r]^{(1)} = \sum_{1 \le j \le r} (-h_j^{(1)} + g_j^{(1)}), \tag{2.10}$$

$$\begin{aligned}[h_1^{-1} \cdot g_1 \cdot \ldots \cdot h_r^{-1} \cdot g_r]^{(2)} = &\sum_{1 \le j \le r} \big\{ -(h_j^{(2)} - g_j^{(2)}) + R_0(h_j^{(1)}, h_j^{(1)} - g_j^{(1)}) \big\} \\ &+ \sum_{1 \le l < j \le r} R_0(-h_l^{(1)} + g_l^{(1)}, -h_j^{(1)} + g_j^{(1)}),\end{aligned} \tag{2.11}$$

as a consequence of applying (1.9) inductively. In most of our applications the operators $S_1, T_1, \ldots, S_r, T_r$ are equal and defined by a kernel K that has product structure, i.e.

$$\begin{aligned}S_1 f = T_1 f = \cdots = S_r f = T_r f = f * K,\\ K(g) = K(g^{(1)}, g^{(2)}) = K^{(1)}(g^{(1)}) K^{(2)}(g^{(2)}).\end{aligned} \tag{2.12}$$

In this case we can derive an additional formula for the kernel A^r. We use the identity

$$\mathbb{1}_{\{0\}}(x^{-1} \cdot y) = \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \mathfrak{e}((y^{(1)} - x^{(1)}).\theta^{(1)}) \mathfrak{e}((y^{(2)} - x^{(2)}).\theta^{(2)})\, d\theta^{(1)} d\theta^{(2)}$$

and the formula (2.9) to write

$$A^r(y) = \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \mathfrak{e}\big(y^{(1)}.\theta^{(1)}\big) \mathfrak{e}\big(y^{(2)}.\theta^{(2)}\big) \Sigma^r\big(\theta^{(1)}, \theta^{(2)}\big)\, d\theta^{(1)} d\theta^{(2)}, \tag{2.13}$$

where

$$\begin{aligned}\Sigma^r\big(\theta^{(1)}, \theta^{(2)}\big) := &\sum_{h_j, g_j \in \mathbb{G}_0} \Big\{ \prod_{j=1}^{r} \overline{K(h_j)} K(g_j) \Big\} \\ &\prod_{i=1}^{2} \mathfrak{e}\big(- [h_1^{-1} \cdot g_1 \cdot \ldots \cdot h_r^{-1} \cdot g_r]^{(i)}.\theta^{(i)} \big).\end{aligned}$$

Recalling the product formula (2.12) we can write

$$\Sigma^r\big(\theta^{(1)}, \theta^{(2)}\big) = \Pi^r\big(\theta^{(1)}, \theta^{(2)}\big) \Omega^r\big(\theta^{(2)}\big), \tag{2.14}$$

for any $(\theta^{(1)}, \theta^{(2)}) \in \mathbb{T}^d \times \mathbb{T}^{d'}$, where

$$\begin{aligned}\Pi^r\big(\theta^{(1)}, \theta^{(2)}\big) := \sum_{h_j^{(1)}, g_j^{(1)} \in \mathbb{Z}^d} & \Big\{ \prod_{j=1}^{r} \overline{K^{(1)}(h_j^{(1)})} K^{(1)}(g_j^{(1)}) \Big\} \\ & \times \mathfrak{e}\big(\theta^{(1)}. \sum_{1 \le j \le r} (h_j^{(1)} - g_j^{(1)})\big) \\ & \times \mathfrak{e}\Big(- \theta^{(2)}.\Big\{ \sum_{1 \le j \le r} R_0(h_j^{(1)}, h_j^{(1)} - g_j^{(1)}) \\ & \quad + \sum_{1 \le l < j \le r} R_0(-h_l^{(1)} + g_l^{(1)}, -h_j^{(1)} + g_j^{(1)}) \Big\}\Big)\end{aligned} \tag{2.15}$$

and

$$\begin{aligned}\Omega^r\big(\theta^{(2)}\big) &:= \sum_{h_j^{(2)}, g_j^{(2)} \in \mathbb{Z}^{d'}} \Big\{ \prod_{j=1}^{r} \overline{K^{(2)}(h_j^{(2)})} K^{(2)}(g_j^{(2)}) \Big\} \\ &\quad \times \mathfrak{e}\big(\theta^{(2)}. \sum_{1 \le j \le r} (h_j^{(2)} - g_j^{(2)})\big) \\ &= \Big| \sum_{g^{(2)} \in \mathbb{Z}^{d'}} K^{(2)}(g^{(2)}) \mathfrak{e}\big(- \theta^{(2)}.g^{(2)}\big) \Big|^{2r}.\end{aligned} \tag{2.16}$$

2.4 Exponential sums and oscillatory integrals

We will often use the following estimates, which follow easily using the Poisson summation formula and integration by parts.

Lemma 2.1 *Assume that $m, M \in \mathbb{Z}_+$ satisfy $M \ge m + 1$, and $f : \mathbb{R}^m \to \mathbb{C}$ is a $C^M(\mathbb{R})$ compactly supported function. Then, for any $\xi \in [-1/2, 1/2]^m$, we have*

$$\Big| \sum_{n \in \mathbb{Z}^m} f(n) \mathfrak{e}(n.\xi) - \int_{\mathbb{R}^m} f(x) \mathfrak{e}(x.\xi)\, dx \Big| \lesssim_M \int_{\mathbb{R}^m} \sum_{n=1}^{m} |\partial_n^M f(x)|\, dx. \tag{2.17}$$

As a consequence, for any $j \in \{1, \dots, m\}$ we have

$$\Big|\sum_{x\in\mathbb{Z}^m} f(x)\mathfrak{e}(x.\xi)\Big| \lesssim_M |\xi_j|^{-M}\int_{\mathbb{R}^m}|\partial_j^M f(x)|\,dx + \int_{\mathbb{R}^m}\sum_{n=1}^m|\partial_n^M f(x)|\,dx. \tag{2.18}$$

Many of our arguments will rely on estimates of exponential sums and oscillatory integrals involving polynomial phases. We record first some classical Weyl-type estimates, which are proved for example in [52, Proposition 1]:

Proposition 2.2 (i) *Assume that $P \geq 1$ is an integer and $\phi_P : \mathbb{R} \to \mathbb{R}$ is a $C^1(\mathbb{R})$ function satisfying*

$$|\phi_P| \leq \mathbb{1}_{[-P,P]}, \quad \int_{\mathbb{R}} \big|\phi_P'(x)\big|\,dx \leq 1. \tag{2.19}$$

Assume that $\varepsilon > 0$ and $\theta = (\theta_1, \dots, \theta_d) \in \mathbb{R}^d$ has the property that there is $l \in \{1, \dots, d\}$ and an irreducible fraction $a/q \in \mathbb{Q}$ with $q \in \mathbb{Z}_+$, such that

$$|\theta_l - a/q| \leq 1/q^2 \ \text{ and } \ q \in [P^\varepsilon, P^{l-\varepsilon}]. \tag{2.20}$$

Then there is a constant $\overline{C} = \overline{C}_d \geq 1$ such that

$$\Big|\sum_{n\in\mathbb{Z}} \phi_P(n)\mathfrak{e}\big(-(\theta_1 n + \cdots + \theta_d n^d)\big)\Big| \lesssim_\varepsilon P^{1-\varepsilon/\overline{C}}. \tag{2.21}$$

(ii) *For any irreducible fraction $\theta = a/q \in (\mathbb{Z}/q)^d$, $a = (a_1, \dots, a_d) \in \mathbb{Z}^d$, $q \in \mathbb{Z}_+$, we have*

$$\Big|q^{-1}\sum_{n\in\mathbb{Z}_q} \mathfrak{e}\big(-(\theta_1 n + \cdots + \theta_d n^d)\big)\Big| \lesssim q^{-1/\overline{C}}. \tag{2.22}$$

We will also need non-commutative versions of these Weyl estimates. With the notation in Sect. 1.3, for $r \in \mathbb{Z}_+$ let $D, \widetilde{D} : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{G}_0^\#$, given by

$$\begin{aligned} D((n_1,\dots,n_r),(m_1,\dots,m_r)) &:= A_0(n_1)^{-1}\cdot A_0(m_1)\cdot\ldots\cdot A_0(n_r)^{-1}\cdot A_0(m_r),\\ \widetilde{D}((n_1,\dots,n_r),(m_1,\dots,m_r)) &:= A_0(n_1)\cdot A_0(m_1)^{-1}\cdot\ldots\cdot A_0(n_r)\cdot A_0(m_r)^{-1}. \end{aligned} \tag{2.23}$$

By definition, we have

$$[A_0(n)]_{l_1l_2} = \begin{cases} n^{l_1} & \text{if } l_2 = 0,\\ 0 & \text{if } l_2 \geq 1, \end{cases} \quad [A_0(n)^{-1}]_{l_1l_2} = \begin{cases} -n^{l_1} & \text{if } l_2 = 0,\\ n^{l_1+l_2} & \text{if } l_2 \geq 1. \end{cases}$$

Thus, using (2.10) and (2.11), for $x = (x_1, \dots, x_r) \in \mathbb{R}^r$ and $y = (y_1, \dots, y_r) \in \mathbb{R}^r$ one has

$$[D(x,y)]_{l_1 l_2} = \begin{cases} \sum_{j=1}^{r} (y_j^{l_1} - x_j^{l_1}) & \text{if } l_2 = 0, \\ \sum_{1 \le j_1 < j_2 \le r} (y_{j_1}^{l_1} - x_{j_1}^{l_1})(y_{j_2}^{l_2} - x_{j_2}^{l_2}) + \sum_{j=1}^{r} (x_j^{l_1+l_2} - x_j^{l_1} y_j^{l_2}) & \text{if } l_2 \ge 1, \end{cases} \tag{2.24}$$

and

$$[\widetilde{D}(x,y)]_{l_1 l_2} = \begin{cases} \sum_{j=1}^{r} (x_j^{l_1} - y_j^{l_1}) & \text{if } l_2 = 0, \\ \sum_{1 \le j_1 < j_2 \le r} (x_{j_1}^{l_1} - y_{j_1}^{l_1})(x_{j_2}^{l_2} - y_{j_2}^{l_2}) + \sum_{j=1}^{r} (y_j^{l_1+l_2} - x_j^{l_1} y_j^{l_2}) & \text{if } l_2 \ge 1. \end{cases} \tag{2.25}$$

For $P \in \mathbb{Z}_+$ assume $\phi_P^{(j)}, \psi_P^{(j)} : \mathbb{R} \to \mathbb{R}$, $j \in \{1, \dots, r\}$, are $C^1(\mathbb{R})$ functions with the properties

$$\begin{aligned} &\sup_{1 \le j \le r} \big[|\phi_P^{(j)}| + |\psi_P^{(j)}| \big] \le \mathbb{1}_{[-P,P]}, \\ &\sup_{1 \le j \le r} \int_{\mathbb{R}} |[\phi_P^{(j)}]'(x)| + |[\psi_P^{(j)}]'(x)| \, dx \le 1. \end{aligned} \tag{2.26}$$

For $\theta = (\theta_{l_1 l_2})_{(l_1,l_2) \in Y_d} \in \mathbb{R}^{|Y_d|}$, $r \in \mathbb{Z}_+$, and $P \in \mathbb{Z}_+$ let

$$S_{P,r}(\theta) = \sum_{n,m \in \mathbb{Z}^r} \mathfrak{e}(-D(n,m).\theta) \Big\{ \prod_{j=1}^{r} \phi_P^{(j)}(n_j) \psi_P^{(j)}(m_j) \Big\}$$

and

$$\widetilde{S}_{P,r}(\theta) = \sum_{n,m \in \mathbb{Z}^r} \mathfrak{e}(-\widetilde{D}(n,m).\theta) \Big\{ \prod_{j=1}^{r} \phi_P^{(j)}(n_j) \psi_P^{(j)}(m_j) \Big\},$$

where D and $\widetilde{D}$ are defined as in (2.24)–(2.25).

The following key estimates are proved in [33, Proposition 5.1 and Lemma 3.1]:

Proposition 2.3 (i) *For any $\varepsilon > 0$ there is $r = r(\varepsilon, d) \in \mathbb{Z}_+$ sufficiently large such that for all $P \in \mathbb{Z}_+$ we have*

$$|S_{P,r}(\theta)| + |\widetilde{S}_{P,r}(\theta)| \lesssim_\varepsilon P^{2r} P^{-1/\varepsilon}, \tag{2.27}$$

provided that there is $(l_1, l_2) \in Y_d$ and an irreducible fraction $a/q \in \mathbb{Q}$, $q \in \mathbb{Z}_+$, such that

$$|\theta_{l_1 l_2} - a/q| \le 1/q^2 \quad \text{and} \quad q \in [P^\varepsilon, P^{l_1+l_2-\varepsilon}]. \tag{2.28}$$

(ii) *For any irreducible fraction $a/q \in \mathbb{Q}$, $a = (a_{l_1 l_2})_{(l_1,l_2)\in Y_d} \in \mathbb{Z}^{|Y_d|}$, $q \in \mathbb{Z}_+$, we define the arithmetic coefficients*

$$\begin{aligned} G(a/q) &:= q^{-2r} \sum_{v,w \in \mathbb{Z}_q^r} \mathfrak{e}\big(-D(v,w).(a/q)\big), \\ \widetilde{G}(a/q) &:= q^{-2r} \sum_{v,w \in \mathbb{Z}_q^r} \mathfrak{e}\big(-\widetilde{D}(v,w).(a/q)\big). \end{aligned} \tag{2.29}$$

Then for any $\varepsilon > 0$ there is $r = r(\varepsilon, d) \in \mathbb{Z}_+$ sufficiently large such that

$$|G(a/q)| + |\widetilde{G}(a/q)| \lesssim_\varepsilon q^{-1/\varepsilon}. \tag{2.30}$$

We will also use a related integral estimate, see Lemma 5.4 in [33]:

Proposition 2.4 *Given $\varepsilon > 0$ there is $r = r(\varepsilon, d)$ sufficiently large as in Proposition 2.3 such that*

$$\begin{aligned} &\Big| \int_{\mathbb{R}^r \times \mathbb{R}^r} \Big\{ \prod_{j=1}^r \phi_j(x_j)\psi_j(y_j) \Big\} \mathfrak{e}(-D(x,y).\beta)\, dxdy \Big| \lesssim \langle \beta \rangle^{-1/\varepsilon}, \\ &\Big| \int_{\mathbb{R}^r \times \mathbb{R}^r} \Big\{ \prod_{j=1}^r \phi_j(x_j)\psi_j(y_j) \Big\} \mathfrak{e}(-\widetilde{D}(x,y).\beta)\, dxdy \Big| \lesssim \langle \beta \rangle^{-1/\varepsilon}, \end{aligned} \tag{2.31}$$

for any $\beta \in \mathbb{R}^{|Y_d|}$ and for any $C^1(\mathbb{R})$ functions $\phi_1, \psi_1, \dots, \phi_r, \psi_r : \mathbb{R} \to \mathbb{C}$ satisfying, for any $j \in \{1, \dots, r\}$, the following bound

$$|\phi_j(x)| + |\partial_x \phi_j(x)| + |\psi_j(x)| + |\partial_x \psi_j(x)| \lesssim \mathbb{1}_{[-1,1]}(x).$$

3 Ergodic theorems: Proof of Theorem 1.2

Assuming momentarily that Theorem 1.3 has been proved we will illustrate how to use it to establish Theorem 1.2. For this purpose we introduce a smoothed variant of average (1.1).

Let $d_1 \in \mathbb{Z}_+$. Given any family $T_1, \dots, T_{d_1} : X \to X$ of invertible measure-preserving transformations, a measurable function $f \in L^p(X)$, $p \in [1, \infty]$, polynomials $P_1, \dots, P_{d_1} \in \mathbb{Z}[n]$, a real number $N \geq 1$, and a smooth function $\chi : \mathbb{R} \to [0, 1]$ supported on the interval $[-2, 2]$ we can define a smoothed polynomial ergodic average $A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f) \in L^p(X)$ by the formula

$$A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f)(x) := \sum_{n\in\mathbb{Z}} N^{-1}\chi(N^{-1}n) f(T_1^{P_1(n)} \cdots T_{d_1}^{P_{d_1}(n)} x), \quad x \in X. \tag{3.1}$$

3.1 Calderón transference principle

We now establish a variant of the Calderón transference principle [16], which will allow us to deduce maximal and ρ-variational estimates for smoothed averages (3.1) from the corresponding estimates for the averages M^{χ}_N along the moment curve A_0 on the group $\mathbb{G}_0$, see Theorem 1.3.

Proposition 3.1 *Let $d_1 \in \mathbb{Z}_+$ be given and let $T_1, \dots, T_{d_1} : X \to X$ be a family of invertible measure-preserving transformations of a σ-finite measure space $(X, \mathcal{B}(X), \mu)$ that generates a nilpotent group of step two. Let $P_1, \dots, P_{d_1} \in \mathbb{Z}[n]$ be such that $P_j(0) = 0$, $1 \leq j \leq d_1$, and let $d_2 := \max\{\deg P_j : j \in \{1, \dots, d_1\}\}$. Assume $f \in L^p(X)$ for some $1 \leq p \leq \infty$, and let $A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f)$ be the average defined in (3.1) corresponding to a smooth function $\chi : \mathbb{R} \to [0, 1]$ supported on the interval $[-2, 2]$. Let M^{χ}_N be the average from Theorem 1.3.*

(i) *If M^{χ}_N satisfies (1.16) for some $1 < p \leq \infty$ then*

$$\Big\| \sup_{N\in\mathbb{Z}_+} |A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f)| \Big\|_{L^p(X)} \lesssim_{d_1,d_2,p,\chi} \|f\|_{L^p(X)}. \tag{3.2}$$

(ii) *If M^{χ}_N satisfies (1.17) for some $1 < p < \infty$, $\rho > \max\{p, \frac{p}{p-1}\}$ and $\tau \in (1, 2]$, then*

$$\big\| V^{\rho}\big(A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f) : N \in \mathbb{D}_{\tau}\big) \big\|_{L^p(X)} \lesssim_{d_1,d_2,p,\rho,\tau,\chi} \|f\|_{L^p(X)}, \tag{3.3}$$

where $\mathbb{D}_{\tau} = \{\tau^n : n \in \mathbb{N}\}$.

Proof We proceed in two steps. We perform first a lifting procedure, which allows us to replace the polynomials $P_1, \dots, P_{d_1}$ with the moment curve A_0 from (1.11). Then we can employ the ideas from the transference principle [16] to complete the proof.

Step 1. Let $\mathbb{G} := \mathbb{G}[T_1, \dots, T_{d_1}]$ be a nilpotent group of step two generated by $T_1, \dots, T_{d_1}$, so

$$[[T_i, T_j], T_l] = \mathrm{Id}, \quad \text{for all} \quad i, j, l \in \{1, \dots, d_1\}, \tag{3.4}$$

where $[S, T] := S^{-1}T^{-1}ST$ denotes the commutator of any two invertible maps $S, T : X \to X$. Define $S_{ij} := [T_i, T_j] = T_i^{-1}T_j^{-1}T_iT_j$, for $i, j \in \{1, \dots, d_1\}$, then by (3.4) note that $T_iT_j = T_jT_iS_{ij}$, and $T_kS_{ij} = S_{ij}T_k$ for all $i, j, k \in \{1, \dots, d_1\}$. Hence

$$\prod_{i=1}^{d_1} T_i^{m_i} \prod_{j=1}^{d_1} T_j^{n_j} = \prod_{j=1}^{d_1} T_j^{m_j+n_j} \prod_{1\le i<j\le d_1} S_{ji}^{m_jn_i}. \tag{3.5}$$

Formula (3.5) gives rise to a homomorphism $T : \mathbb{G}_0(d_1) \to \mathbb{G}$ defined by

$$T(g) := \prod_{l_1=1}^{d_1} T_{l_1}^{m_{l_1 0}} \prod_{1\le l_2<l_1\le d_1} S_{l_1l_2}^{m_{l_1l_2}}, \quad \text{for any} \quad g = (m_{l_1l_2})_{(l_1,l_2)\in Y_{d_1}} \in \mathbb{G}_0(d_1).$$

Let $A : \mathbb{Z} \to \mathbb{G}_0(d_1)$ be defined by $A(n) := (P_1(n), \dots, P_{d_1}(n), 0, \dots, 0)^{-1}$ and note that

$$T(A(n)^{-1}) = T_1^{P_1(n)} \cdots T_{d_1}^{P_{d_1}(n)}. \tag{3.6}$$

In view of [33, Lemma 2.2] there exists $d \in \mathbb{Z}_+$ depending only on the integers $d_1, d_2 \in \mathbb{Z}_+$, and a homomorphism $\Phi : \mathbb{G}_0(d) \to \mathbb{G}_0^{\#}(d_1)$ such that for all $n \in \mathbb{Z}$ one has

$$A(n) = \Phi(A_0(n)). \tag{3.7}$$

From the proof of [33, Lemma 2.2] one can easily deduce that for every $g \in \mathbb{G}_0(d)$ we have $\Phi(g) \in \mathbb{Z}^{d_1} \times (\mathbb{Z}/2)^{d_1'}$. Combining (3.6) with (3.7) we see that the group $\Phi^{-1}[\mathbb{G}_0(d_1)]$ acts on X via $\Phi^{-1}[\mathbb{G}_0(d_1)] \times X \ni (g, x) \mapsto g \odot x \in X$ defined by $g \odot x = (T \circ \Phi(g))x$, which allows us to write

$$A_{N;X,\chi}^{P_1,\dots,P_{d_1}}(f)(x) = \sum_{n\in\mathbb{Z}} N^{-1}\chi(N^{-1}n) f(A_0(n)^{-1} \odot x). \tag{3.8}$$

Step 2. We now prove (3.2) and (3.3). We will only prove (3.3), since the proof of (3.2) is similar and we omit the details. Define $f_L^x(g) := f(g \odot x)\mathbb{1}_{[-1,1]^{d+d'}}(L^{-1} \circ g)\mathbb{1}_{\Phi^{-1}[\mathbb{G}_0(d_1)](g)}$ for $L > 0$, $x \in X$ and $g \in \mathbb{G}_0(d)$.

Using (3.8) and the fact that $g^2 \in \Phi^{-1}[\mathbb{G}_0(d_1)]$, $g \in \mathbb{G}_0(d)$, observe that for $g \in \mathbb{G}_0(d)$ obeying $L^{-1} \circ g \in [-1, 1]^{d+d'}$ one has

$$V^\rho\big(A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f)(g^2 \odot x) : N \in \mathbb{D}_\tau \cap [1, L]\big) = V^\rho\big(M^\chi_N(f^x_{CL})(g^2) : N \in \mathbb{D}_\tau \cap [1, L]\big)$$

for some large absolute constant $C > 0$ depending only on d.

Summing over all $g \in \mathbb{G}_0(d)$ obeying $L^{-1} \circ g \in [-1, 1]^{d+d'}$, and integrating over X, we have

$$\begin{aligned}\Big(\prod_{(l_1,l_2)\in Y_d} L^{l_1+l_2}\Big)\big\|V^\rho\big(A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f) : N \in \mathbb{D}_\tau \cap [1, L]\big)\big\|^p_{L^p(X)} &\lesssim \int_X \big\|V^\rho\big(M^\chi_N(f^x_{CL}) : N \in \mathbb{D}_\tau\big)\big\|^p_{\ell^p(\mathbb{G}_0)} d\mu(x)\\ &\lesssim \int_X \|f^x_{CL}\|^p_{\ell^p(\mathbb{G}_0)} d\mu(x)\\ &\lesssim \Big(\prod_{(l_1,l_2)\in Y_d} L^{l_1+l_2}\Big)\|f\|^p_{L^p(X)},\end{aligned} \tag{3.9}$$

using also (1.17) in the second estimate. Dividing both sides of (3.9) by $\prod_{(l_1,l_2)\in Y_d} L^{l_1+l_2}$ and letting $L \to \infty$ we obtain (3.3). □

Having proven estimates (3.2) and (3.3) we can easily complete the proof of Theorem 1.2.

3.2 Proof of Theorem 1.2(iii)

Let $\chi : \mathbb{R} \to [0, 1]$ be a smooth function such that $\mathbb{1}_{[-1,1]} \le \chi \le \mathbb{1}_{[-2,2]}$. Note that

$$\sup_{N\in\mathbb{Z}_+} |A^{P_1,\dots,P_{d_1}}_{N;X}(f)(x)| \le \sup_{N\in\mathbb{Z}_+} A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(|f|)(x).$$

Appealing to (3.2) we conclude (1.3). □

3.3 Proof of Theorem 1.2(ii)

By a simple density argument, using the maximal inequality (1.3), it suffices to establish pointwise convergence for $f \in L^p(X) \cap L^\infty(X)$ with $1 < p < \infty$. Invoking ρ-variational inequality (3.3) one has

$$\lim_{\mathbb{D}_\tau \ni M,N\to\infty} |A^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f)(x) - A^{P_1,\dots,P_{d_1}}_{M;X,\chi}(f)(x)| = 0$$

μ-almost everywhere on X. The same is true for the operators

$$\tilde{A}^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f)(x) := \frac{1}{|[-N,N]\cap\mathbb{Z}|}\sum_{n\in\mathbb{Z}}\chi(N^{-1}n)f(T_1^{P_1(n)}\dots T_{d_1}^{P_{d_1}(n)}x), \quad x\in X.$$

Let $\varepsilon > 0$ and pick a smooth function $\chi : \mathbb{R} \to [0,1]$ such that $\|\mathbb{1}_{[-1,1]} - \chi\|_{L^1(\mathbb{R})} < \varepsilon$. Fix $f \in L^p(X)\cap L^\infty(X)$ such that $\|f\|_{L^\infty(X)} = 1$ and $f \geq 0$, and note that

$$\begin{aligned}
&\limsup_{\mathbb{D}_\tau\ni M,N\to\infty} |A^{P_1,\dots,P_{d_1}}_{N;X}(f)(x) - A^{P_1,\dots,P_{d_1}}_{M;X}(f)(x)|\\
&\leq 2\limsup_{\mathbb{D}_\tau\ni N\to\infty} |A^{P_1,\dots,P_{d_1}}_{N;X}(f)(x) - \tilde{A}^{P_1,\dots,P_{d_1}}_{N;X,\chi}(f)(x)|\\
&\lesssim \limsup_{\mathbb{D}_\tau\ni N\to\infty}\frac{1}{|[-N,N]\cap\mathbb{Z}|}\sum_{n\in\mathbb{Z}}\big|\chi(N^{-1}n) - \mathbb{1}_{[-1,1]}(N^{-1}n)\big|\\
&\lesssim \|\mathbb{1}_{[-1,1]} - \chi\|_{L^1(\mathbb{R})}\\
&\lesssim \varepsilon,
\end{aligned} \tag{3.10}$$

for μ-almost all $x \in X$. Letting $\varepsilon \to 0^+$ we obtain that the limit

$$\lim_{\mathbb{D}_\tau\ni N\to\infty} A^{P_1,\dots,P_{d_1}}_{N;X}(f)(x)$$

exists μ-almost everywhere on X for every $\tau \in (1,2]$. Using this with $\tau = 2^{1/s}$ for $s \in \mathbb{Z}_+$ we obtain that there exists a function $f_s^* \in L^p(X)$ such that

$$\lim_{n\to\infty} A^{P_1,\dots,P_{d_1}}_{2^{n/s};X}(f)(x) = f_s^*(x) \tag{3.11}$$

μ-almost everywhere on X for every $s \in \mathbb{Z}_+$. Since $\mathbb{D}_2 \subseteq \mathbb{D}_{2^{1/s}}$ we conclude that $f_1^* = f_s^*$ for all $s \in \mathbb{Z}_+$. Now for each $s \in \mathbb{Z}_+$ and each $N \in \mathbb{Z}_+$ let $(n_m)_{m\in\mathbb{N}} \subseteq \mathbb{N}$ be a sequence such that $2^{n_N/s} \leq N < 2^{(n_N+1)/s}$. Then by (3.11) for $f \geq 0$ we have

$$2^{-1/s} f_1^*(x) \leq \liminf_{N\to\infty} A^{P_1,\dots,P_{d_1}}_{N;X}(f)(x) \leq \limsup_{N\to\infty} A^{P_1,\dots,P_{d_1}}_{N;X}(f)(x) \leq 2^{1/s} f_1^*(x).$$

Letting $s \to \infty$ we obtain

$$\lim_{N\to\infty} A_{N;X}^{P_1,\dots,P_{d_1}}(f)(x) = f_1^*(x)$$

μ-almost everywhere on X. This completes the proof of Theorem 1.2(ii). □

3.4 Proof of Theorem 1.2(i)

Finally pointwise convergence from Theorem 1.2(ii) combined with maximal inequality (1.3) and dominated convergence theorem gives norm convergence for any $f \in L^p(X)$ with $1 < p < \infty$ and the proof of Theorem 1.2 is completed. □

4 Maximal and variational estimates on $\mathbb{G}_0$: ℓ^2 theory

In this section we discuss the nilpotent circle method on the discrete group $\mathbb{G}_0$, and outline the proof of the key ρ-variational inequality (1.17) for $p = 2$ and $2 < \rho < \infty$.

Assume that $\tau \in (1, 2]$ is a fixed parameter. The basic case is $\tau = 2$, but we need slightly stronger bounds for the ergodic theory application, see (3.10). We also fix a smooth function $\chi : \mathbb{R} \to [0, 1]$ supported on $[-2, 2]$. For simplicity of notation, for $k \in \mathbb{N}$ and $x \in \mathbb{G}_0$, let

$$\begin{aligned} \mathcal{M}_k f(x) &:= M_{\tau^k}^{\chi} f(x) = \sum_{n\in\mathbb{Z}} \tau^{-k}\chi(\tau^{-k}n) f(A_0(n)^{-1}\cdot x) = (f * K_k)(x), \\ K_k(x) &:= G_{\tau^k}^{\chi}(x) = \sum_{n\in\mathbb{Z}} \tau^{-k}\chi(\tau^{-k}n)\mathbb{1}_{\{A_0(n)\}}(x), \end{aligned} \tag{4.1}$$

see (1.13) and (1.15) for the definitions M_N^{χ} and G_N^{χ} respectively.

Our aim is to establish (1.17) for $p = 2$ and $2 < \rho < \infty$, which with the new notation can be rewritten as follows:

Theorem 4.1 *Let $\tau \in (1, 2]$ and $2 < \rho < \infty$ be given. Then for any $f \in \ell^2(\mathbb{G}_0)$ one has*

$$\big\| V^{\rho}\big(\mathcal{M}_k(f) : k \geq 0\big)\big\|_{\ell^2(\mathbb{G}_0)} \lesssim_{d,\rho,\tau,\chi} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.2}$$

In particular, one also has

$$\big\| \sup_{k\geq 0} |\mathcal{M}_k f|\big\|_{\ell^2(\mathbb{G}_0)} \lesssim_{d,\tau,\chi} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.3}$$

The proof of Theorem 4.1 will take up Sect. 4, 5, 6, 7, and 8. For simplicity of notation, all the implied constants in this proof are allowed to depend on d, τ, χ, ρ.

We fix $\eta_0 : \mathbb{R} \to [0, 1]$ a smooth even function such that $\mathbb{1}_{[-1,1]} \leq \eta_0 \leq \mathbb{1}_{[-2,2]}$. For $t \in \mathbb{R}$ and integers $j \geq 1$ we define

$$\eta_j(t) := \eta_0(\tau^{-j}t) - \eta_0(\tau^{-j+1}t), \quad 1 = \sum_{j=0}^{\infty} \eta_j. \tag{4.4}$$

For any $A \in [0, \infty)$ we define

$$\eta_{\leq A} := \sum_{j \in [0,A] \cap \mathbb{Z}} \eta_j. \tag{4.5}$$

By a slight abuse of notation we also let η_j and $\eta_{\leq A}$ denote the smooth radial functions on $\mathbb{R}^m, m \in \mathbb{Z}_+$, defined by $\eta_j(x) = \eta_j(|x|)$ and $\eta_{\leq A}(x) = \eta_{\leq A}(|x|)$

To prove Theorem 4.1 we need to decompose the kernels defining the operators $\mathcal{M}_k$. The kernels K_k have product structure

$$\begin{aligned} K_k(g) &:= L_k(g^{(1)})\mathbb{1}_{\{0\}}(g^{(2)}), \\ L_k(g^{(1)}) &:= \sum_{n \in \mathbb{Z}} \tau^{-k}\chi(\tau^{-k}n)\mathbb{1}_{\{0\}}(g^{(1)} - A_0^{(1)}(n)), \end{aligned} \tag{4.6}$$

where $A_0^{(1)}(n) := (n, \dots, n^d) \in \mathbb{Z}^d$ and $g = (g^{(1)}, g^{(2)}) \in \mathbb{G}_0$ as in (1.7).

4.1 The main decomposition

We first decompose the singular kernel $\mathbb{1}_{\{0\}}(g^{(2)})$ in the central variable $g^{(2)}$ into smoother kernels. For any $s \in \mathbb{N}$ and $m \in \mathbb{Z}_+$ we define the set of rational fractions

$$\begin{aligned} \mathcal{R}_s^m := \ & \{a/q : a = (a_1, \dots, a_m) \in \mathbb{Z}^m, \\ & q \in [\tau^s, \tau^{s+1}) \cap \mathbb{Z}, \ \gcd(a_1, \dots, a_m, q) = 1\}. \end{aligned} \tag{4.7}$$

We define also $\mathcal{R}_{\leq a}^m := \bigcup_{0 \leq s \leq a} \mathcal{R}_s^m$. For $x^{(1)} = (x_{l_1 0}^{(1)})_{l_1 \in \{1,\dots,d\}} \in \mathbb{R}^d$, $x^{(2)} = (x_{l_1 l_2}^{(2)})_{(l_1,l_2) \in Y_d'} \in \mathbb{R}^{d'}$ and $\Lambda \in (0, \infty)$ we define the partial dilations

$$\Lambda \circ x^{(1)} = (\Lambda^{l_1} x^{(1)}_{l_1 0})_{l_1 \in \{1,\ldots,d\}} \in \mathbb{R}^d, \quad \Lambda \circ x^{(2)} = (\Lambda^{l_1+l_2} x^{(2)}_{l_1 l_2})_{(l_1,l_2)\in Y'_d} \in \mathbb{R}^{d'}, \tag{4.8}$$

which are induced by the group-dilations defined in (1.12).

We fix two small constants $\delta = \delta(d) \ll \delta' = \delta'(d)$ such that $\delta' \in (0, (10d)^{-10}]$ and $\delta \in (0, (\delta')^4]$, and a large constant $D = D(d) \gg \delta^{-8}$. These constants depend on arithmetic properties of the polynomial sequence A_0, more precisely on the structural constants in Propositions 2.2–2.4. For example, we could take $\delta' = (10d)^{-10}$, then take $\delta = \delta'/\overline{C}_1$, where $\overline{C}_1$ is a large constant depending on the constant $\overline{C}$ in Proposition 2.2. Then we fix an integer $r = r(\delta) \geq \delta^{-4}$ such that the bounds in Propositions 2.3–2.4 hold with $\varepsilon = \delta^4$, and then take $D := \lfloor r\delta^{-4} \rfloor + 1$. To summarize

$$1 \ll 1/\delta' \ll 1/\delta \ll r \ll D. \tag{4.9}$$

For $k \geq (D/\ln\tau)^2$ we fix two cutoff functions $\phi^{(1)}_k : \mathbb{R}^d \to [0,1]$, $\phi^{(2)}_k : \mathbb{R}^{d'} \to [0,1]$, such that

$$\phi^{(1)}_k(g^{(1)}) := \eta_{\leq \delta k}(\tau^{-k} \circ g^{(1)}), \quad \phi^{(2)}_k(g^{(2)}) := \eta_{\leq \delta k}(\tau^{-k} \circ g^{(2)}). \tag{4.10}$$

For $k, w \in \mathbb{N}$ so that $k \geq (D/\ln\tau)^2$ and $0 \leq w \leq k$ and for any 1-periodic sets of rationals $\mathcal{A} \subseteq \mathbb{Q}^d$, $\mathcal{B} \subseteq \mathbb{Q}^{d'}$ we define the periodic Fourier multipliers by

$$\begin{aligned} \Psi_{k,w,\mathcal{A}}(\xi^{(1)}) &:= \sum_{a/q\in\mathcal{A}} \eta_{\leq \delta' w}(\tau^k \circ (\xi^{(1)} - a/q)), \quad \xi^{(1)} \in \mathbb{T}^d, \\ \Xi_{k,w,\mathcal{B}}(\xi^{(2)}) &:= \sum_{b/q\in\mathcal{B}} \eta_{\leq \delta w}(\tau^k \circ (\xi^{(2)} - b/q)), \quad \xi^{(2)} \in \mathbb{T}^{d'}. \end{aligned} \tag{4.11}$$

For $k \geq (D/\ln\tau)^2$ and $s \in [0, \delta k] \cap \mathbb{Z}$ we define the periodic Fourier multipliers $\Xi_{k,s} : \mathbb{R}^{d'} \to [0,1]$,

$$\Xi_{k,s}(\xi^{(2)}) := \Xi_{k,k,\mathcal{R}^{d'}_s}(\xi^{(2)}) = \sum_{a/q\in\mathcal{R}^{d'}_s} \eta_{\leq \delta k}(\tau^k \circ (\xi^{(2)} - a/q)). \tag{4.12}$$

For $k \geq (D/\ln\tau)^2$ we write

$$\begin{aligned}\mathbb{1}_{\{0\}}(g^{(2)}) &= \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\, d\xi^{(2)} \\ &= \sum_{s\in[0,\delta k]\cap\mathbb{Z}} \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_{k,s}(\xi^{(2)})\, d\xi^{(2)} \\ &\quad + \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_k^c(\xi^{(2)})\, d\xi^{(2)},\end{aligned} \tag{4.13}$$

where $g^{(2)}.\xi^{(2)}$ denotes the usual scalar product of vectors in $\mathbb{R}^{d'}$ and

$$\Xi_k^c := 1 - \sum_{s\in[0,\delta k]\cap\mathbb{Z}} \Xi_{k,s}. \tag{4.14}$$

Then we decompose $K_k = K_k^c + \sum_{s\in[0,\delta k]\cap\mathbb{Z}} K_{k,s}$, where, with the notation in (4.6), we have

$$K_{k,s}(g) := L_k(g^{(1)})N_{k,s}(g^{(2)}), \quad K_k^c(g) := L_k(g^{(1)})N_k^c(g^{(2)}), \tag{4.15}$$

and

$$\begin{aligned}N_{k,s}(g^{(2)}) &:= \phi_k^{(2)}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_{k,s}(\xi^{(2)})\, d\xi^{(2)}, \\ N_k^c(g^{(2)}) &:= \phi_k^{(2)}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_k^c(\xi^{(2)})\, d\xi^{(2)}.\end{aligned} \tag{4.16}$$

We first show that we can bound the contributions of the minor arcs in the central variables:

Lemma 4.2 *For any integer $k \geq (D/\ln\tau)^2$ and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\|f * K_k^c\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-k/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.17}$$

This is proved in Sect. 5.1 below.

We now turn to the operators $K_{k,s}$, and show first that we can bound the contributions corresponding to scales $k \geq 0$ being not very large. More precisely, for any $s \geq 0$ we define

$$\kappa_s := 2^{(D/\ln\tau)(s+1)^2}. \tag{4.18}$$

Lemma 4.3 *For any integer $s \geq 0$ and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\begin{aligned} &\big\| V^\rho(f * K_{k,s} : \max((D/\ln\tau)^2, s/\delta) \leq k < 2\kappa_s)\big\|_{\ell^2(\mathbb{G}_0)} \\ &\quad \lesssim \tau^{-s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)} \end{aligned} \tag{4.19}$$

and

$$\Big\| \sup_{\max((D/\ln\tau)^2, s/\delta)\leq k<2\kappa_s} |f * K_{k,s}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.20}$$

This is proved in Sect. 7 below.

After these reductions, it remains to prove that

$$\big\| V^\rho(f * K_{k,s} : k \geq \kappa_s)\big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)} \quad \text{for any integer } s \geq 0. \tag{4.21}$$

The kernels $K_{k,s}$ are now reasonably well adapted to a natural family of non-isotropic balls in the central variables, at least when $\tau^s \simeq 1$, and we need to start decomposing in the non-central variables. We examine the kernels $L_k(g^{(1)})$ defined in (4.6), and rewrite them in the form

$$\begin{aligned} L_k(g^{(1)}) &= \sum_{n\in\mathbb{Z}} \tau^{-k}\chi(\tau^{-k}n)\mathbb{1}_{\{0\}}(-A_0^{(1)}(n) + g^{(1)}) \\ &= \phi_k^{(1)}(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)}) S_k(\xi^{(1)})\, d\xi^{(1)}, \end{aligned} \tag{4.22}$$

where $g^{(1)}.\xi^{(1)}$ denotes the usual scalar product of vectors in $\mathbb{R}^d$, and

$$S_k(\xi^{(1)}) := \sum_{n\in\mathbb{Z}} \tau^{-k}\chi(\tau^{-k}n)\mathfrak{e}(-A_0^{(1)}(n).\xi^{(1)}). \tag{4.23}$$

For any integers $Q \in \mathbb{Z}_+$ and $m \in \mathbb{Z}_+$ we define the set of fractions

$$\widetilde{\mathcal{R}}_Q^m := \{a/Q : a = (a_1, \ldots, a_m) \in \mathbb{Z}^m\}. \tag{4.24}$$

For any integer $s \geq 0$ we fix a large denominator

$$Q_s := \big(\lfloor \tau^{D(s+1)} \rfloor\big)! = 1 \cdot 2 \cdot \ldots \cdot \lfloor \tau^{D(s+1)} \rfloor, \tag{4.25}$$

and using (4.11) define the periodic multipliers

$$\Psi^{\mathrm{low}}_{k,s}(\xi^{(1)}) := \Psi_{k,k,\widetilde{\mathcal{R}}^d_{Q_s}}(\xi^{(1)}) = \sum_{a/q\in\widetilde{\mathcal{R}}^d_{Q_s}} \eta_{\le\delta' k}(\tau^k\circ(\xi^{(1)}-a/q)),$$

$$\Psi_{k,s,t}(\xi^{(1)}) := \Psi_{k,k,\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s}}(\xi^{(1)}) = \sum_{a/q\in\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s}} \eta_{\le\delta' k}(\tau^k\circ(\xi^{(1)}-a/q)),$$

$$\begin{aligned}\Psi^c_k(\xi^{(1)}) &:= 1-\Psi^{\mathrm{low}}_{k,s}(\xi^{(1)}) - \sum_{t\in[0,\delta' k]\cap\mathbb{Z}} \Psi_{k,s,t}(\xi^{(1)})\\ &= 1-\sum_{a/q\in\mathcal{R}^d_{\le\delta' k}} \eta_{\le\delta' k}(\tau^k\circ(\xi^{(1)}-a/q)).\end{aligned} \tag{4.26}$$

Since $k\ge\kappa_s = 2^{(D/\ln\tau)(s+1)^2}$ we see that $Q_s\le\tau^{\delta^2 k}$. Therefore the supports of the cutoff functions $\eta_{\le\delta' k}(\tau^k\circ(\xi^{(1)}-a/q))$ are all disjoint and the multipliers $\Psi^{\mathrm{low}}_{k,s}$, $\Psi_{k,s,t}$, Ψ^c_k take values in the interval $[0,1]$. Notice also that $\Psi_{k,s,t}\equiv 0$ unless $t\ge D(s+1)$, and that the cutoffs used in these definitions depend on $\delta' k$ not on δk as in the case of the central variables.

We examine the formula (4.22) and define the kernels $L^{\mathrm{low}}_{k,s}, L_{k,s,t}, L^c_k : \mathbb{Z}^d\to\mathbb{C}$ by

$$L_*(g^{(1)}) = \phi^{(1)}_k(g^{(1)})\int_{\mathbb{T}^d}\mathfrak{e}(g^{(1)}.\xi^{(1)})S_k(\xi^{(1)})\Psi_*(\xi^{(1)})\,d\xi^{(1)}, \tag{4.27}$$

where $(L_*,\Psi_*)\in\{(L^{\mathrm{low}}_{k,s},\Psi^{\mathrm{low}}_{k,s}),(L_{k,s,t},\Psi_{k,s,t}),(L^c_k,\Psi^c_k)\}$. For any $k\ge\kappa_s$ we obtain $K_{k,s} = G^{\mathrm{low}}_{k,s} + \sum_{t\le\delta' k} G_{k,s,t} + G^c_{k,s}$, where the kernels $G^{\mathrm{low}}_{k,s}, G_{k,s,t}, G^c_{k,s} : \mathbb{Z}^{|Y_d|}\to\mathbb{C}$ are defined by

$$\begin{aligned}G^{\mathrm{low}}_{k,s}(g) &:= L^{\mathrm{low}}_{k,s}(g^{(1)})N_{k,s}(g^{(2)}),\\ G_{k,s,t}(g) &:= L_{k,s,t}(g^{(1)})N_{k,s}(g^{(2)}),\\ G^c_{k,s}(g) &:= L^c_k(g^{(1)})N_{k,s}(g^{(2)}).\end{aligned} \tag{4.28}$$

To prove (4.21) we need to establish Lemmas 4.4–4.6.

Our next lemma shows that the contribution of the minor arcs can be suitably bounded:

Lemma 4.4 *For any integers $s\ge 0$ and $k\ge\kappa_s$, and for any $f\in\ell^2(\mathbb{G}_0)$ we have*

$$\|f * G^c_{k,s}\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-k/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.29}$$

It remains to bound the contributions of the major arcs in both the central and the non-central variables. We start with the contributions corresponding to averages over large k.

Lemma 4.5 (i) *For any integer $s \geq 0$ and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\big\| V^\rho(f * G^{\mathrm{low}}_{k,s} : k \geq \kappa_s) \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.30}$$

In particular, we have

$$\big\| \sup_{k \geq \kappa_s} |f * G^{\mathrm{low}}_{k,s}| \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-s/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.31}$$

(ii) *For any integers $s \geq 0$, $t \geq D(s+1)$, and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\big\| V^\rho(f * G_{k,s,t} : k \geq \kappa_t) \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.32}$$

where $\kappa_t := 2^{(D/\ln \tau)(t+1)^2}$ as in (4.18). In particular, we have

$$\big\| \sup_{k \geq \kappa_t} |f * G_{k,s,t}| \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.33}$$

Finally, we deal with the operators defined by the kernels $G_{k,s,t}$ for intermediate values of k.

Lemma 4.6 *For any integers $s \geq 0$, and $t \geq D(s+1)$, and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\big\| V^\rho(f * G_{k,s,t} : \max(\kappa_s, t/\delta') \leq k < 2\kappa_t) \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.34}$$

In particular, we have

$$\big\| \sup_{\max(\kappa_s, t/\delta') \leq k < 2\kappa_t} |f * G_{k,s,t}| \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{4.35}$$

We will prove these lemmas in Sects. 5–8. Theorem 4.1 follows from Lemmas 4.2–4.6.

For later use in the ℓ^p theory, we will sometimes need to work with slightly more general kernels on $\mathbb{G}_0$. Given two 1-periodic sets of rationals $\mathcal{A} \subseteq \mathbb{Q}^d$ and $\mathcal{B} \subseteq \mathbb{Q}^{d'}$, we define

$$\begin{aligned} K_{k,w,\mathcal{A},\mathcal{B}}(g) &:= L_{k,w,\mathcal{A}}(g^{(1)}) N_{k,w,\mathcal{B}}(g^{(2)}), \\ K'_{k,w,\mathcal{A},\mathcal{B}}(g) &:= L'_{k,w,\mathcal{A}}(g^{(1)}) N_{k,w,\mathcal{B}}(g^{(2)}), \end{aligned} \tag{4.36}$$

where

$$L_{k,w,\mathcal{A}}(g^{(1)}) := \phi_k^{(1)}(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)}) \Psi_{k,w,\mathcal{A}}(\xi^{(1)}) S_k(\xi^{(1)})\, d\xi^{(1)},$$
$$L'_{k,w,\mathcal{A}}(g^{(1)}) := \phi_k^{(1)}(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)}) \Psi_{k,w,\mathcal{A}}(\xi^{(1)}) [\Delta_k S_k](\xi^{(1)})\, d\xi^{(1)},$$
$$N_{k,w,\mathcal{B}}(g^{(2)}) := \phi_k^{(2)}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)}) \Xi_{k,w,\mathcal{B}}(\xi^{(2)})\, d\xi^{(2)}. \tag{4.37}$$

The multipliers $\Psi_{k,w,\mathcal{A}}$ and $\Xi_{k,w,\mathcal{B}}$ are defined in (4.11) and $\Delta_k S_k = S_{k+1} - S_k$ as in (2.1). Using the definitions, it is easy to see, for example, that $L^{\mathrm{low}}_{k,s}(g^{(1)}) = L_{k,k,\widetilde{\mathcal{R}}^d_{Q_s}}(g^{(1)})$, $L_{k,s,t}(g^{(1)}) = L_{k,k,\mathcal{R}^d_t \setminus \widetilde{\mathcal{R}}^d_{Q_s}}(g^{(1)})$, and $N_{k,s}(g^{(2)}) = N_{k,k,\mathcal{R}^{d'}_s}(g^{(2)})$ as in (4.28).

5 Minor arcs contributions: Proofs of Lemmas 4.2 and 4.4

In this section we use high order T^*T arguments to bound the minor arcs contributions.

5.1 Proof of Lemma 4.2

We proceed in two steps:

Step 1. We define the operators $\mathcal{K}^c_k f := f * K^c_k$. Set $\varepsilon = \delta^4$ and fix a positive integer $r = r(d)$ large enough such that the bounds as in Propositions 2.3 and 2.4 hold. Then

$$\{(\mathcal{K}^c_k)^* \mathcal{K}^c_k\}^r f(x) = (f * A^{c,r}_k)(x),$$

where, using the formulas (2.13)–(2.16) and (4.15), one has

$$A^{c,r}_k(y) = \eta_{\le 3\delta k}(\tau^{-k} \circ y) \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \mathfrak{e}\big(y.\theta\big) \Pi^{c,r}_k\big(\theta^{(1)}, \theta^{(2)}\big) \Omega^{c,r}_k\big(\theta^{(2)}\big)\, d\theta^{(1)} d\theta^{(2)}, \tag{5.1}$$

where

$$\Pi^{c,r}_k\big(\theta^{(1)}, \theta^{(2)}\big) := \sum_{h^{(1)}_j, g^{(1)}_j \in \mathbb{Z}^d} \Big\{ \prod_{j=1}^{r} L_k(h^{(1)}_j) L_k(g^{(1)}_j) \Big\} \times \mathfrak{e}\Big(\theta^{(1)}. \sum_{1 \le j \le r} (h^{(1)}_j - g^{(1)}_j)\Big)$$

$$\times \mathfrak{e}\Big(-\theta^{(2)}.\Big\{ \sum_{1\le j\le r} R_0(h_j^{(1)}, h_j^{(1)} - g_j^{(1)}) + \sum_{1\le l<j\le r} R_0(-h_l^{(1)} + g_l^{(1)}, -h_j^{(1)} + g_j^{(1)})\Big\}\Big)$$

and

$$\Omega_k^{c,r}\big(\theta^{(2)}\big) := \Big| \sum_{g^{(2)}\in\mathbb{Z}^{d'}} N_k^c(g^{(2)})\mathfrak{e}\big(-\theta^{(2)}.g^{(2)}\big)\Big|^{2r}.$$

Using the defining formula (4.6) we can write

$$\begin{aligned}\Pi_k^{c,r}(\theta) = \tau^{-2rk} \sum_{n_j,m_j\in\mathbb{Z}} \Big\{\prod_{j=1}^{r} \chi(\tau^{-k}n_j)\chi(\tau^{-k}m_j)\Big\}\\ \times \mathfrak{e}\Big(-\theta^{(1)}.\sum_{1\le j\le r}(A_0^{(1)}(m_j) - A_0^{(1)}(n_j))\Big)\\ \times \mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le j\le r} R_0(A_0^{(1)}(n_j), A_0^{(1)}(n_j) - A_0^{(1)}(m_j))\Big\}\Big)\\ \times \mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le l<j\le r} R_0(A_0^{(1)}(n_l) - A_0^{(1)}(m_l),\\ A_0^{(1)}(n_j) - A_0^{(1)}(m_j))\Big\}\Big).\end{aligned}$$

Using (2.24) it is easy to see that

$$\begin{aligned}&\theta^{(1)}.\sum_{1\le j\le r}(A_0^{(1)}(m_j) - A_0^{(1)}(n_j))\\ &+\theta^{(2)}.\Big\{\sum_{1\le j\le r} R_0(A_0^{(1)}(n_j), A_0^{(1)}(n_j) - A_0^{(1)}(m_j))\Big\}\\ &+\theta^{(2)}.\Big\{\sum_{1\le l<j\le r} R_0(A_0^{(1)}(n_l) - A_0^{(1)}(m_l), A_0^{(1)}(n_j) - A_0^{(1)}(m_j))\Big\} = \theta.D(n,m).\end{aligned}$$

Therefore

$$\Pi_k^{c,r}(\theta) = \tau^{-2kr} \sum_{n,m\in\mathbb{Z}^r} \Big\{\prod_{j=1}^{r} \chi(\tau^{-k}n_j)\chi(\tau^{-k}m_j)\Big\}\mathfrak{e}\big(-\theta.D(n,m)\big). \tag{5.2}$$

We can also derive a good formula for the kernel $\Omega_k^{c,r}$. Letting

$$F_k(\beta^{(2)}) := \sum_{g^{(2)}\in\mathbb{Z}^{d'}} \eta_{\le \delta k}(\tau^{-k}\circ g^{(2)})\mathfrak{e}(-g^{(2)}.\beta^{(2)}) \tag{5.3}$$

and recalling the definition in (4.16), we have

$$\Omega_k^{c,r}(\theta^{(2)}) = \Big|\int_{\mathbb{T}^{d'}} F_k(\theta^{(2)}-\xi^{(2)})\Xi_k^c(\xi^{(2)})\,d\xi^{(2)}\Big|^{2r}. \tag{5.4}$$

Step 2. We now prove that $\|A_k^{c,r}\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-k}$. Using also the formula (5.1) for this it suffices to prove that if $k \ge (D/\ln\tau)^2$ then

$$\big|\Pi_k^{c,r}\big(\theta^{(1)},\theta^{(2)}\big)\Omega_k^{c,r}\big(\theta^{(2)}\big)\big| \lesssim \tau^{-k/\delta} \quad \text{for any } (\theta^{(1)},\theta^{(2)})\in\mathbb{T}^d\times\mathbb{T}^{d'}. \tag{5.5}$$

We examine the formula (5.3) and apply Lemma 2.1 with $M\in\mathbb{Z}_+$ sufficiently large to conclude that, for any $\beta^{(2)}\in[-1/2,1/2]^{d'}$, we have

$$|F_k(\beta^{(2)})| \lesssim_M \prod_{(l_1,l_2)\in Y_d'}\Big\{\tau^{k(l_1+l_2+\delta)}\big(1+|\beta_{l_1l_2}^{(2)}|\tau^{k(l_1+l_2+\delta)}\big)^{-M}\Big\}. \tag{5.6}$$

To prove (5.5) we use the formulas (5.2) and (5.4), and consider two cases depending on the location of $\theta^{(2)}$. Assume first that $\theta^{(2)}$ is far from the support of Ξ_k^c, i.e.

$$\begin{aligned}&\text{there is an irreducible fraction } a/q \text{ with } q\le\tau^{\delta k-4} \text{ and } a=(a_{l_1l_2})_{(l_1,l_2)\in Y_d'}\\ &\text{such that } |\theta_{l_1l_2}^{(2)}-a_{l_1l_2}/q|\le\tau^{\delta k/2}\tau^{-k(l_1+l_2)} \text{ for any } (l_1,l_2)\in Y_d'.\end{aligned} \tag{5.7}$$

In view of the definitions (4.12) and (4.14) it follows that for any $\xi^{(2)}$ in the support of the function Ξ_k^c there is $(l_1,l_2)\in Y_d'$ such that $|\xi_{l_1l_2}^{(2)}-\theta_{l_1l_2}^{(2)}|\ge\tau^{\delta k/2}\tau^{-k(l_1+l_2)}$. Then $|F_k(\theta^{(2)}-\xi^{(2)})|\lesssim\tau^{-2k/\delta}$ if $\xi^{(2)}$ is in the support of Ξ_k^c, as a consequence of (5.6). The bounds (5.5) follow using (5.4) if $\theta^{(2)}$ satisfies (5.7).

On the other hand, assume that $\theta^{(2)}$ does not satisfy (5.7). By the Dirichlet principle, for any $(l_1,l_2)\in Y_d'$ there is an irreducible fraction $a_{l_1l_2}/q_{l_1l_2}$ such that

$$\Big|\theta_{l_1l_2}^{(2)}-\frac{a_{l_1l_2}}{q_{l_1l_2}}\Big|\le\frac{1}{q_{l_1l_2}\tau^{k(l_1+l_2)-\delta^2k}} \quad\text{and}\quad q_{l_1l_2}\in[1,\tau^{k(l_1+l_2)-\delta^2k}]\cap\mathbb{Z}.$$

Since $\theta^{(2)}$ does not satisfy the property (5.7), it follows that at least one of the denominators $q_{l_1 l_2}$ is larger than $\tau^{\delta^2 k}$. In particular, the property (2.28) is verified if $P \simeq \tau^k$. Recalling the formula (5.2), we can apply Proposition 2.3 (i) to conclude that $\big|\Pi_k^{c,r}(\theta^{(1)}, \theta^{(2)})\big| \lesssim \tau^{-2k/\delta}$. Moreover, $\|F_k\|_{L^1(\mathbb{T}^{d'})} \lesssim 1$ due to (5.6), therefore $\big|\Omega_k^{c,r}(\theta^{(2)})\big| \lesssim 1$ as a consequence of (5.4). The desired bounds (5.5) follow in this case as well, which completes the proof of Lemma 4.2. □

5.2 Proof of Lemma 4.4

For later use we prove a slightly more general version of Lemma 4.4. For 1-periodic set of rationals $\mathcal{B} \subseteq \mathcal{R}^{d'}_{\le \delta k}$, we define new kernels

$$G^c_{k,\mathcal{B}}(g) := L^c_k(g^{(1)}) N_{k,k,\mathcal{B}}(g^{(2)}), \tag{5.8}$$

where $N_{k,k,\mathcal{B}}$ is defined in (4.37). We now prove the following lemma:

Lemma 5.1 *For any 1-periodic set of rationals $\mathcal{B} \subseteq \mathcal{R}^{d'}_{\le \delta k}$, $k \ge (D/\ln \tau)^2$, and any $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\|f * G^c_{k,\mathcal{B}}\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-k/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}. \tag{5.9}$$

Since $G^c_{k,\mathcal{R}^{d'}_s} = G^c_{k,s}$, see (4.28), Lemma 4.4 follows from Lemma 5.1.

Proof of Lemma 5.1 As before, we shall proceed in several steps.

Step 1. We define the operators $\mathcal{G}^c_{k,\mathcal{B}} f := f * G^c_{k,\mathcal{B}}$. Since $G^c_{k,\mathcal{B}}(x) = L^c_k(x^{(1)}) N_{k,k,\mathcal{B}}(x^{(2)})$ we have

$$\{(\mathcal{G}^c_{k,\mathcal{B}})^* \mathcal{G}^c_{k,\mathcal{B}}\}^r f(x) = (f * A^r_{k,\mathcal{B}})(x),$$

where

$$\begin{aligned} A^r_{k,\mathcal{B}}(y) &= \eta_{\le 3\delta k}(\tau^{-k} \circ y) \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \mathfrak{e}\big(y.\theta\big) \Pi^r_k\big(\theta^{(1)}, \theta^{(2)}\big) \Omega^r_{k,\mathcal{B}}\big(\theta^{(2)}\big)\, d\theta^{(1)} d\theta^{(2)}, \\ \Pi^r_k\big(\theta^{(1)}, \theta^{(2)}\big) &:= \sum_{h^{(1)}_j, g^{(1)}_j \in \mathbb{Z}^d} \Big\{ \prod_{j=1}^r \overline{L^c_k(h^{(1)}_j)} L^c_k(g^{(1)}_j) \Big\} \mathfrak{e}\big(\theta^{(1)}. \sum_{1 \le j \le r} (h^{(1)}_j - g^{(1)}_j)\big) \\ &\quad \times \mathfrak{e}\Big(- \theta^{(2)}. \Big\{ \sum_{1 \le j \le r} R_0(h^{(1)}_j, h^{(1)}_j - g^{(1)}_j) \\ &\quad + \sum_{1 \le l < j \le r} R_0(-h^{(1)}_l + g^{(1)}_l, -h^{(1)}_j + g^{(1)}_j) \Big\} \Big) \end{aligned} \tag{5.10}$$

and, with F_k defined as in (5.3), we may write

$$\Omega^r_{k,\mathcal{B}}\big(\theta^{(2)}\big) := \Big| \int_{\mathbb{T}^{d'}} F_k(\theta^{(2)} - \xi^{(2)})\, \Xi_{k,k,\mathcal{B}}(\xi^{(2)})\, d\xi^{(2)} \Big|^{2r}. \tag{5.11}$$

To prove Lemma 5.1 it suffices to show that for any $(\theta^{(1)}, \theta^{(2)}) \in \mathbb{T}^d \times \mathbb{T}^{d'}$ we have

$$\Big| \Pi^r_k\big(\theta^{(1)}, \theta^{(2)}\big) \Omega^r_{k,\mathcal{B}}\big(\theta^{(2)}\big) \Big| \lesssim \tau^{-k/\delta'}. \tag{5.12}$$

Step 2. Assume first that $\theta^{(2)}$ is far from the support of $\Xi_{k,k,\mathcal{B}}$, in the sense that

$$|\tau^k \circ (\theta^{(2)} - a/Q)| \geq \tau^{2\delta k} \quad \text{for any} \quad a/Q \in \mathcal{B} \subseteq \mathcal{R}^{d'}_{\leq \delta k}.$$

Using (5.6) it follows that $\big|\Omega^r_{k,\mathcal{B}}\big(\theta^{(2)}\big)\big| \lesssim \tau^{-2r^2 k}$. Moreover

$$\big|\Pi^r_k\big(\theta^{(1)}, \theta^{(2)}\big)\big| \lesssim \|L^c_k\|^{2r}_{\ell^1(\mathbb{Z}^d)} \lesssim \Big\{ \prod_{1 \leq l \leq d} \tau^{k(l+\delta)} \Big\}^{2r},$$

and the desired bounds (5.12) follow in this case.

Step 3. On the other hand, assume that

$$\begin{aligned} &|\tau^k \circ (\theta^{(2)} - a/Q)| \leq \tau^{2\delta k} \quad \text{for some irreducible fraction} \\ &\quad a/Q \in \mathcal{B} \subseteq \mathcal{R}^{d'}_{\leq \delta k}. \end{aligned} \tag{5.13}$$

In this case we prove the stronger bounds

$$\Big| \Pi^r_k\big(\theta^{(1)}, \theta^{(2)}\big) \Big| \lesssim \tau^{-k/\delta'} \quad \text{for any} \quad \theta^{(1)} \in \mathbb{T}^{d'}. \tag{5.14}$$

We examine the formulas (5.10) and (4.27) to rewrite

$$\begin{aligned} \Pi^r_k\big(\theta^{(1)}, \theta^{(2)}\big) = \int_{(\mathbb{T}^d)^{2r}} & \mathcal{V}^r_k(\theta^{(1)}, \theta^{(2)}; \zeta^{(1)}_1, \xi^{(1)}_1, \dots, \zeta^{(1)}_r, \xi^{(1)}_r) \\ & \times \prod_{1 \leq j \leq r} \big\{ \overline{S_k(\zeta^{(1)}_j)}\, \overline{\Psi^c_k(\zeta^{(1)}_j)} S_k(\xi^{(1)}_j) \Psi^c_k(\xi^{(1)}_j) \big\} \\ & \times d\zeta^{(1)}_1 d\xi^{(1)}_1 \dots d\zeta^{(1)}_r d\xi^{(1)}_r, \end{aligned} \tag{5.15}$$

where $\zeta_1^{(1)}, \xi_1^{(1)}, \ldots, \zeta_r^{(1)}, \xi_r^{(1)} \in \mathbb{T}^d$ and

$$\begin{aligned}\mathcal{V}_k^r(\theta^{(1)}, \theta^{(2)}; \zeta_1^{(1)}, \xi_1^{(1)}, \ldots, \zeta_r^{(1)}, \xi_r^{(1)}) \\ := \sum_{h_j, g_j \in \mathbb{Z}^d} \prod_{1\le j\le r} \Big\{ \overline{\phi_k^{(1)}(h_j)} \\ \mathfrak{e}\big((\theta^{(1)} - \zeta_j^{(1)}).h_j\big)\phi_k^{(1)}(g_j)\mathfrak{e}\big(-(\theta^{(1)} - \xi_j^{(1)}).g_j\big)\Big\} \\ \times \mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le j\le r} R_0(h_j, h_j - g_j) \\ + \sum_{1\le l<j\le r} R_0(-h_l + g_l, -h_j + g_j)\Big\}\Big).\end{aligned} \tag{5.16}$$

We will show below that

$$\big|S_k(\beta^{(1)})\Psi_k^c(\beta^{(1)})\big| \lesssim \tau^{-k\delta'/(2d\overline{C})} \quad \text{for any } \beta^{(1)} \in \mathbb{T}^d, \tag{5.17}$$

where $\overline{C}$ is a constant from Proposition 2.2. We will also show that

$$\begin{aligned}\big|\mathcal{V}_k^r(\theta^{(1)}, \theta^{(2)}; \zeta_1^{(1)}, \xi_1^{(1)}, \ldots, \zeta_r^{(1)}, \xi_r^{(1)})\big| \\ \lesssim \Big\{\prod_{1\le l\le d} \tau^{k(l+\delta)}\Big\}^{2r} \min_{\substack{1\le j\le r\\ 1\le l\le d}} \big[1 + \tau^{k(l-8\delta)}\|\theta_l^{(1)} - \zeta_{j,l}^{(1)}\|_Q \\ + \tau^{k(l-8\delta)}\|\theta_l^{(1)} - \xi_{j,l}^{(1)}\|_Q\big]^{-D^2},\end{aligned} \tag{5.18}$$

for any $\theta^{(1)} = (\theta_l^{(1)})_{l\in\{1,\ldots,d\}} \in \mathbb{T}^d$, $\zeta_j^{(1)} = (\zeta_{j,l}^{(1)})_{l\in\{1,\ldots,d\}} \in \mathbb{T}^d$, and $\xi_j^{(1)} = (\xi_{j,l}^{(1)})_{l\in\{1,\ldots,d\}} \in \mathbb{T}^d$. Here $Q \le \tau^{\delta k+1}$ and $\theta^{(2)}$ are as in (5.13), and

$$\|\mu\|_Q := \inf_{m\in\mathbb{Z}} |\mu - m/Q| \quad \text{for any } \mu \in \mathbb{R}. \tag{5.19}$$

The desired estimates (5.14) would clearly follow from these bounds and the identity in (5.15). Here the assumption $\delta \ll \delta'$ in (4.9) plays an important role.

Step 4. The bound in (5.18) follows from the more precise formulas in Lemma 5.2 below, using repeated integration by parts in the variables x_j, y_j to prove bounds on the function $\mathcal{Z}_k^r$ defined in (5.25) and using the trivial bound $|\mathcal{W}_Q^r| \le 1$ for the function defined in (5.24). We prove now the bounds (5.17). Assume $\beta^{(1)} = (\beta_l^{(1)})_{l\in\{1,\ldots,d\}}$. By the Dirichlet principle for any $l \in$

$\{1, \dots, d\}$ there is an irreducible fraction a_l/q_l such that

$$\left|\beta_l^{(1)} - a_l/q_l\right| \le \frac{1}{q_l \tau^{lk-\delta' k/2}} \quad \text{and} \quad q_l \in [1, \tau^{lk-\delta' k/2}] \cap \mathbb{Z}. \tag{5.20}$$

If $q_l \le \tau^{\delta' k/(2d)}$ for all $l \in \{1, \dots, d\}$ then $\Psi_k^c(\beta^{(1)}) = 0$ due to the definition (4.26). On the other hand, if $q_l \in [\tau^{\delta' k/(2d)}, \tau^{lk-\delta' k/2}] \cap \mathbb{Z}$ for some $l \in \{1, \dots, d\}$ then we apply Proposition 2.2 with $P \simeq \tau^k$ and $\varepsilon = \delta'/(2d)$. Recalling the definition (4.23) it follows that $\left|S_k(\beta^{(1)})\right| \lesssim \tau^{-k\delta'/(2d\overline{C})}$, and the desired bound in (5.17) follow. □

For later use, in Sect. 8, we prove an approximate formula for the multiplier $\mathcal{V}_k^r$.

Lemma 5.2 *Assume that $k \ge D/\ln\tau$ and $1 \le Q \le \tau^{2\delta k}$. Assume also that*

$$\theta^{(2)} = a^{(2)}/Q + \alpha^{(2)}, \quad a^{(2)} \in \mathbb{Z}^{d'}, \quad |\tau^k \circ \alpha^{(2)}| \le \tau^{4\delta k} \tag{5.21}$$

and

$$\begin{aligned} \theta^{(1)} - \xi_j^{(1)} &= b_j/Q + \beta_j,\ \theta^{(1)} - \zeta_j^{(1)} = c_j/Q + \gamma_j, \quad b_j, c_j \in \mathbb{Z}^d, \\ & Q\beta_j, Q\gamma_j \in [-1/2, 1/2]^d, \end{aligned} \tag{5.22}$$

for any $j \in \{1, \dots, r\}$. Then we have the approximate identity

$$\begin{aligned} &\mathcal{V}_k^r(\theta^{(1)}, \theta^{(2)}; \zeta_1^{(1)}, \xi_1^{(1)}, \dots, \zeta_r^{(1)}, \xi_r^{(1)}) \\ &\quad = \mathcal{W}_Q^r(a^{(2)}; b_1, c_1, \dots, b_r, c_r) \cdot \mathcal{Z}_k^r(\alpha^{(2)}; \beta_1, \gamma_1, \dots, \beta_r, \gamma_r) + O(\tau^{-D^3 k}), \end{aligned} \tag{5.23}$$

where

$$\begin{aligned} &\mathcal{W}_Q^r(a^{(2)}; b_1, c_1, \dots, b_r, c_r) \\ &\quad := \Big\{ Q^{-2rd} \sum_{\mu_j, \nu_j \in \mathbb{Z}_Q^d} \Big(\prod_{1 \le j \le r} \mathfrak{e}\big(-(b_j/Q).\mu_j\big)\mathfrak{e}\big((c_j/Q).\nu_j\big)\Big) \\ &\qquad \times \mathfrak{e}\Big(-(a^{(2)}/Q).\Big\{ \sum_{1 \le j \le r} R_0(\nu_j, \nu_j - \mu_j) \\ &\qquad + \sum_{1 \le l < j \le r} R_0(-\nu_l + \mu_l, -\nu_j + \mu_j)\Big\}\Big)\Big\}, \end{aligned} \tag{5.24}$$

and

$$\begin{aligned}\mathcal{Z}_k^r(\alpha^{(2)};\beta_1,\gamma_1,\dots,\beta_r,\gamma_r) := \int_{\mathbb{R}^{2rd}} \Big\{\prod_{1\le l\le d}\tau^{kl}\Big\}^{2r}\\ \times\prod_{1\le j\le r}\Big\{\eta_{\le\delta k}(x_j)\mathfrak{e}\big(-(\tau^k\circ\beta_j).x_j\big)\eta_{\le\delta k}(y_j)\mathfrak{e}\big((\tau^k\circ\gamma_j).y_j\big)\Big\}\\ \times\mathfrak{e}\Big(-(\tau^k\circ\alpha^{(2)}).\Big\{\sum_{1\le j\le r}R_0(y_j,y_j-x_j)\\ +\sum_{1\le l<j\le r}R_0(-y_l+x_l,-y_j+x_j)\Big\}\Big)\,dx_j dy_j.\end{aligned} \tag{5.25}$$

Proof We decompose $g_j = Qm_j+\mu_j$, $h_j = Qn_j+\nu_j$, $m_j,n_j\in\mathbb{Z}^d$, $\mu_j,\nu_j\in\mathbb{Z}^d_Q$ and then rewrite the formula (5.16) in the form

$$\begin{aligned}&\mathcal{V}_k^r(\theta^{(1)},\theta^{(2)};\zeta_1^{(1)},\xi_1^{(1)},\dots,\zeta_r^{(1)},\xi_r^{(1)})\\ &=\sum_{\mu_j,\nu_j\in\mathbb{Z}^d_Q}\sum_{n_j,m_j\in\mathbb{Z}^d}\prod_{1\le j\le r}\Big\{\eta_{\le\delta k}(\tau^{-k}\circ(Qn_j+\nu_j))\\ &\quad\times\mathfrak{e}\big(\gamma_j.(Qn_j+\nu_j)\big)\mathfrak{e}\big((c_j/Q).\nu_j\big)\\ &\quad\times\eta_{\le\delta k}(\tau^{-k}\circ(Qm_j+\mu_j))\mathfrak{e}\big(-\beta_j.(Qm_j+\mu_j)\big)\mathfrak{e}\big(-(b_j/Q).\mu_j\big)\Big\}\\ &\quad\times\mathfrak{e}\Big(-\alpha^{(2)}.\Big\{\sum_{1\le j\le r}R_0(h_j,h_j-g_j)+\sum_{1\le l<j\le r}R_0(-h_l+g_l,-h_j+g_j)\Big\}\Big)\\ &\quad\times\mathfrak{e}\Big(-(a^{(2)}/Q).\Big\{\sum_{1\le j\le r}R_0(\nu_j,\nu_j-\mu_j)\\ &\quad+\sum_{1\le l<j\le r}R_0(-\nu_l+\mu_l,-\nu_j+\mu_j)\Big\}\Big).\end{aligned}$$

We fix the variables μ_j, ν_j and use the Poisson summation formula to replace the sum over m_j,n_j with integrals. Using (2.17) with $\xi=(-Q\beta,Q\gamma)$ and M large we see that the difference is rapidly decreasing in τ^k, due to the assumptions (5.21)–(5.22). Therefore

$$\begin{aligned}\mathcal{V}_k^r&(\theta^{(1)},\theta^{(2)};\zeta_1^{(1)},\xi_1^{(1)},\dots,\zeta_r^{(1)},\xi_r^{(1)})\\ &=\sum_{\mu_j,\nu_j\in\mathbb{Z}^d_Q}\Big\{\prod_{1\le j\le r}\mathfrak{e}\big(-(b_j/Q).\mu_j\big)\mathfrak{e}\big((c_j/Q).\nu_j\big)\Big\}\\ &\quad\times\mathfrak{e}\Big(-(a^{(2)}/Q).\Big\{\sum_{1\le j\le r}R_0(\nu_j,\nu_j-\mu_j)\\ &\quad+\sum_{1\le l<j\le r}R_0(-\nu_l+\mu_l,-\nu_j+\mu_j)\Big\}\Big)\end{aligned}$$

$$\times \int_{\mathbb{R}^{2rd}} \prod_{1\le j\le r} \Big\{ \eta_{\le \delta k}(\tau^{-k}\circ(Qn_j+\nu_j))\mathfrak{e}\big(\gamma_j.(Qn_j+\nu_j)\big)$$
$$\times \eta_{\le \delta k}(\tau^{-k}\circ(Qm_j+\mu_j))\mathfrak{e}\big(-\beta_j.(Qm_j+\mu_j)\big)\Big\}$$
$$\times \mathfrak{e}\Big(-\alpha^{(2)}.\Big\{\sum_{1\le j\le r} R_0(h_j, h_j-g_j)$$
$$+ \sum_{1\le l<j\le r} R_0(-h_l+g_l, -h_j+g_j)\Big\}\Big)\, dm_j dn_j + O(\tau^{-D^3k}),$$

where $h_j = Qn_j + \nu_j$ and $g_j = Qm_j + \mu_j$ in the last line. We make the changes of variables $x_j = \tau^{-k}\circ(Qm_j+\mu_j)$, $y_j = \tau^{-k}\circ(Qn_j+\nu_j)$, and the desired formulas (5.23)–(5.25) follow. □

6 Major arcs contributions: Proof of Lemma 4.5

Our primary goal in this section is to prove the bounds (4.30)–(4.33). For later use in the ℓ^p theory, we will prove in fact slightly stronger bounds at several stages.

6.1 Arithmetic decompositions

We will write the kernels $G^{\text{low}}_{k,s}$ and $G_{k,s,t}$ as tensor products plus error terms. For any integer $Q \in \mathbb{Z}_+$ we define the subgroup

$$\mathbb{H}_Q := \{h = (Qh_{l_1l_2})_{(l_1,l_2)\in Y_d} \in \mathbb{G}_0 : h_{l_1,l_2} \in \mathbb{Z}\}. \tag{6.1}$$

Clearly $\mathbb{H}_Q \subseteq \mathbb{G}_0$ is a normal subgroup. Let $\mathbb{J}_Q$ denote the coset

$$\mathbb{J}_Q := \{b = (b_{l_1l_2})_{(l_1,l_2)\in Y_d} \in \mathbb{G}_0 : b_{l_1,l_2} \in \mathbb{Z}\cap[0, Q-1]\}, \tag{6.2}$$

with the natural induced group structure. Notice that

$$\text{the map } (b,h)\mapsto b\cdot h \text{ defines a bijection from } \mathbb{J}_Q\times\mathbb{H}_Q \text{ to } \mathbb{G}_0. \tag{6.3}$$

Assume that $Q \in \mathbb{Z}_+$ and $\tau^k \ge Q$. For any $a \in \mathbb{Z}^d$ and $\xi \in \mathbb{R}^d$ let

$$J_k(\xi) := \tau^{-k}\int_{\mathbb{R}} \chi(\tau^{-k}x)\mathfrak{e}[-A_0^{(1)}(x).\xi]\,dx = \int_{\mathbb{R}} \chi(y)\mathfrak{e}[-A_0^{(1)}(y).(\tau^k\circ\xi)]\,dy,$$
$$J'_k(\xi) := \tau^{-k}\int_{\mathbb{R}} \chi'(\tau^{-k}x)\mathfrak{e}[-A_0^{(1)}(x).\xi]\,dx = \int_{\mathbb{R}} \chi'(y)\mathfrak{e}[-A_0^{(1)}(y).(\tau^k\circ\xi)]\,dy,$$
$$S(a/Q) := Q^{-1}\sum_{n\in\mathbb{Z}_Q} \mathfrak{e}[-A_0^{(1)}(n).a/Q], \tag{6.4}$$

where $\chi'(x) := (1/\tau)\chi(x/\tau) - \chi(x)$. For any $\iota \in \{0, 1\}$ we also let

$$S_k^\iota := \begin{cases} S_k & \text{if } \iota = 0, \\ \Delta_k S_k & \text{if } \iota = 1, \end{cases} \quad \chi^\iota := \begin{cases} \chi & \text{if } \iota = 0, \\ \chi' & \text{if } \iota = 1, \end{cases} \quad J_k^\iota := \begin{cases} J_k & \text{if } \iota = 0, \\ J_k' & \text{if } \iota = 1. \end{cases} \tag{6.5}$$

where $S_k : \mathbb{R}^d \to \mathbb{R}$ are defined as in (4.23). We first prove an approximation formula for the functions S_k^ι.

Lemma 6.1 *If $k \geq D/\ln\tau$, $|\tau^k \circ \xi| \leq \tau^{k/4}$, $1 \leq Q \leq \tau^{k/4}$, $a \in \mathbb{Z}^d$, and $\iota \in \{0, 1\}$ then*

$$|S_k^\iota(a/Q + \xi) - S(a/Q)J_k^\iota(\xi)| \lesssim \tau^{-Dk}. \tag{6.6}$$

Proof We write

$$\begin{aligned} S_k^\iota(a/Q + \xi) &= \sum_{n\in\mathbb{Z},\, m\in\mathbb{Z}_Q} \tau^{-k}\chi^\iota(\tau^{-k}(Qn+m))\mathfrak{e}[-A_0^{(1)}(Qn+m).(a/Q+\xi)] \\ &= \sum_{m\in\mathbb{Z}_Q} \mathfrak{e}[-A_0^{(1)}(m).a/Q]\Big\{\sum_{n\in\mathbb{Z}} \tau^{-k}\chi^\iota(\tau^{-k}(Qn+m)) \\ &\quad \mathfrak{e}[-A_0^{(1)}(Qn+m).\xi]\Big\}. \end{aligned}$$

For any $m \in \mathbb{Z}_Q$ we apply the estimates (2.17) (with $m = 1$, $\xi = 0$, and M large) to replace the sum over n with the corresponding integral, at the expense of an acceptable error. The desired approximate identity (6.6) follows by a linear change of variables. □

We now prove an approximate formula for the kernels $K_{k,w,\mathcal{A},\mathcal{B}}$ from (4.36).

Lemma 6.2 *Assume that $k, w \in \mathbb{N}$, $k \geq D/\ln\tau$, $0 \leq w \leq k$ and let $1 \leq Q \leq \tau^{\delta k}$. Let $\mathcal{A} \subseteq \widetilde{\mathcal{R}}_Q^d$ and $\mathcal{B} \subseteq \widetilde{\mathcal{R}}_Q^{d'}$ be 1-periodic sets of rationals. If $h \in \mathbb{H}_Q$ and $b_1, b_2 \in \mathbb{G}_0$ satisfy $|b_j| \leq Q^4$, $j \in \{1, 2\}$, then we can decompose*

$$K_{k,w,\mathcal{A},\mathcal{B}}(b_1 \cdot h \cdot b_2) = W_{k,w,Q}(h)V_{\mathcal{A},\mathcal{B},Q}(b_1 \cdot b_2) + E_{k,w,\mathcal{A},\mathcal{B}}(h, b_1, b_2), \tag{6.7}$$

where, for any $h = (h^{(1)}, h^{(2)}) \in \mathbb{H}_Q$ and $b = (b^{(1)}, b^{(2)}) \in \mathbb{G}_0$, one has

$$\begin{aligned} W_{k,w,Q}(h) := \; & Q^{d+d'}\phi_k(h)\int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \eta_{\leq\delta' w}(\tau^k \circ \xi)\eta_{\leq\delta w}(\tau^k \circ \theta) \\ & \times \mathfrak{e}(h.(\xi, \theta))J_k(\xi)\, d\xi d\theta, \end{aligned} \tag{6.8}$$

$$V_{\mathcal{A},\mathcal{B},Q}(b) := Q^{-d-d'}\Big\{\sum_{\sigma^{(1)}\in\mathcal{A}\cap[0,1)^d} S(\sigma^{(1)})\mathfrak{e}[b^{(1)}.(\sigma^{(1)})]\Big\} \times\Big\{\sum_{\sigma^{(2)}\in\mathcal{B}\cap[0,1)^{d'}} \mathfrak{e}[b^{(2)}.(\sigma^{(2)})]\Big\}. \tag{6.9}$$

Here $\phi_k(h) := \phi_k^{(1)}(h^{(1)})\phi_k^{(2)}(h^{(2)})$ *and the error terms* $E_{k,\mathcal{A},\mathcal{B}}$ *satisfy the bounds*

$$\big|E_{k,w,\mathcal{A},\mathcal{B}}(h,b_1,b_2)\big| \lesssim \tau^{-k/2}\Big\{\prod_{(l_1,l_2)\in Y_d}\tau^{-(l_1+l_2)k}\Big\} \times\eta_{\le 2\delta k}(\tau^{-k}\circ h^{(1)})\eta_{\le 2\delta k}(\tau^{-k}\circ h^{(2)}). \tag{6.10}$$

Proof We start from the formula $K_{k,w,\mathcal{A},\mathcal{B}}(g) = L_{k,w,\mathcal{A}}(g^{(1)})N_{k,w,\mathcal{B}}(g^{(2)})$, and recall the definitions (4.11) and (4.37). Letting $b_1 = (b_1^{(1)}, b_1^{(2)})$, $b_2 = (b_2^{(1)}, b_2^{(2)})$, $h = (h^{(1)}, h^{(2)})$ we have

$$\begin{aligned} &b_1\cdot h\cdot b_2 = (g^{(1)}, g^{(2)}),\\ &g^{(1)} := h^{(1)} + b_1^{(1)} + b_2^{(1)},\\ &g^{(2)} := h^{(2)} + b_1^{(2)} + b_2^{(2)} + R_0(b_1^{(1)}, h^{(1)}) + R_0(h^{(1)} + b_1^{(1)}, b_2^{(1)}). \end{aligned} \tag{6.11}$$

Using (4.11) and (4.37) we have

$$\begin{aligned} L_{k,w,\mathcal{A}}(g^{(1)}) &= \phi_k^{(1)}(g^{(1)})\int_{\mathbb{T}^d}\mathfrak{e}(g^{(1)}.\xi^{(1)})S_k(\xi^{(1)})\Psi_{k,w,\mathcal{A}}(\xi^{(1)})\,d\xi^{(1)}\\ &= \eta_{\le\delta k}(\tau^{-k}\circ g^{(1)})\sum_{\sigma^{(1)}\in\mathcal{A}\cap[0,1)^d}\int_{\mathbb{R}^d}\eta_{\le\delta' w}(\tau^k\circ\xi)S_k(\sigma^{(1)}+\xi)\\ &\quad\times\mathfrak{e}[(h^{(1)}+b_1^{(1)}+b_2^{(1)}).(\sigma^{(1)}+\xi)]\,d\xi, \end{aligned}$$

and

$$\begin{aligned} N_{k,w,\mathcal{B}}(g^{(2)}) &= \phi_k^{(2)}(g^{(2)})\int_{\mathbb{T}^{d'}}\mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi_{k,w,\mathcal{B}}(\xi^{(2)})\,d\xi^{(2)}\\ &= \eta_{\le\delta k}(\tau^{-k}\circ g^{(2)})\sum_{\sigma^{(2)}\in\mathcal{B}\cap[0,1)^{d'}}\int_{\mathbb{R}^{d'}}\eta_{\le\delta w}(\tau^k\circ\theta)\\ &\quad\times\mathfrak{e}\big\{[h^{(2)}+b_1^{(2)}+b_2^{(2)}+R_0(b_1^{(1)},h^{(1)})\\ &\quad+R_0(h^{(1)}+b_1^{(1)},b_2^{(1)})].(\sigma^{(2)}+\theta)\big\}\,d\theta. \end{aligned}$$

We notice that if $h^{(1)} \in (Q\mathbb{Z})^d$, $b_1^{(1)}, b_2^{(1)} \in \mathbb{Z}^d$, $\sigma^{(1)} \in \mathcal{A} \subseteq \widetilde{\mathcal{R}}_Q^d$, $|\tau^{-k} \circ h^{(1)}| \lesssim \tau^{\delta k}$, $|b_1^{(1)}| + |b_2^{(1)}| \lesssim Q^4$, $Q \leq \tau^{\delta k}$, $\xi \in \mathbb{R}^d$, and $|\tau^k \circ \xi| \lesssim \tau^{\delta' k}$ then

$$\eta_{\leq \delta k}(\tau^{-k} \circ g^{(1)}) = \eta_{\leq \delta k}(\tau^{-k} \circ h^{(1)}) + O(\tau^{-3k/4}), \tag{6.12}$$

$$\begin{aligned}\mathfrak{e}[(h^{(1)} + b_1^{(1)} + b_2^{(1)}).(\sigma^{(1)} + \xi)] &= \mathfrak{e}[(b_1^{(1)} + b_2^{(1)}).(\sigma^{(1)})]\mathfrak{e}[(h^{(1)} + b_1^{(1)} + b_2^{(1)}).\xi] \\ &= \mathfrak{e}[(b_1^{(1)} + b_2^{(1)}).(\sigma^{(1)})]\mathfrak{e}(h^{(1)}.\xi) + O(\tau^{-3k/4}).\end{aligned} \tag{6.13}$$

Using also Lemma 6.1 we have

$$\begin{aligned}\Big| L_{k,w,\mathcal{A}}(g^{(1)}) - \eta_{\leq \delta k}(\tau^{-k} \circ h^{(1)}) \sum_{\sigma^{(1)} \in \mathcal{A} \cap [0,1)^d} S(\sigma^{(1)})\mathfrak{e}[(b_1^{(1)} + b_2^{(1)}).(\sigma^{(1)})] \\ \times \int_{\mathbb{R}^d} \eta_{\leq \delta' w}(\tau^k \circ \xi) J_k(\xi)\mathfrak{e}(h^{(1)}.\xi)\, d\xi \Big| \lesssim \tau^{-2k/3} \prod_{1 \leq l_1 \leq d} \tau^{-l_1 k}.\end{aligned} \tag{6.14}$$

Moreover, assuming also that $h^{(2)} \in (Q\mathbb{Z})^{d'}$, $b_1^{(2)}, b_2^{(2)} \in \mathbb{Z}^{d'}$, $\sigma^{(2)} \in \mathcal{B} \subseteq \widetilde{\mathcal{R}}_Q^{d'}$, $|\tau^{-k} \circ h^{(2)}| \lesssim \tau^{\delta k}$, $|b_1^{(2)}| + |b_2^{(2)}| \lesssim Q^4$, $\theta \in \mathbb{R}^{d'}$, and $|\tau^k \circ \theta| \lesssim \tau^{\delta k}$, we have

$$\eta_{\leq \delta k}(\tau^{-k} \circ g^{(2)}) = \eta_{\leq \delta k}(\tau^{-k} \circ h^{(2)}) + O(\tau^{-3k/4}), \tag{6.15}$$

$$\begin{aligned}&\mathfrak{e}\big\{[h^{(2)} + b_1^{(2)} + b_2^{(2)} + R_0(b_1^{(1)}, h^{(1)}) + R_0(h^{(1)} + b_1^{(1)}, b_2^{(1)})].(\sigma^{(2)} + \theta)\big\} \\ &\quad = \mathfrak{e}\big\{[b_1^{(2)} + b_2^{(2)} + R_0(b_1^{(1)}, b_2^{(1)})].(\sigma^{(2)})\big\}\mathfrak{e}(h^{(2)}.\theta) + O(\tau^{-3k/4}).\end{aligned} \tag{6.16}$$

Therefore

$$\begin{aligned}\Big| N_{k,w,\mathcal{B}}(g^{(2)}) - \eta_{\leq \delta k}(\tau^{-k} \circ h^{(2)}) \\ \times \sum_{\sigma^{(2)} \in \mathcal{B} \cap [0,1)^{d'}} \mathfrak{e}\big\{[b_1^{(2)} + b_2^{(2)} + R_0(b_1^{(1)}, b_2^{(1)})].(\sigma^{(2)})\big\} \\ \times \int_{\mathbb{R}^{d'}} \eta_{\leq \delta w}(\tau^k \circ \theta)\mathfrak{e}(h^{(2)}.\theta)\, d\theta \Big| \lesssim \tau^{-2k/3} \prod_{(l_1,l_2) \in Y_d'} \tau^{-(l_1+l_2)k}.\end{aligned} \tag{6.17}$$

The conclusion of the lemma follows from (6.14) and (6.17). □

6.2 Gauss sums operators

We consider now the convolution operators defined by the kernels $V_{\mathcal{A},\mathcal{B},Q}$ on the quotient groups $\mathbb{J}_Q$ (see (6.2)). The convolution of two functions on the

group $\mathbb{J}_Q$ is defined by a formula similar to (1.14), namely

$$(f *_{\mathbb{J}_Q} g)(x) := \sum_{y\in\mathbb{J}_Q} g(y) f(y^{-1}\cdot x) = \sum_{y\in\mathbb{J}_Q} g(x\cdot y^{-1}) f(y). \quad (6.18)$$

Lemma 6.3 *Assume that $Q \in \mathbb{Z}_+$ and $\mathcal{A} \subseteq \widetilde{\mathcal{R}}^d_Q$ and $\mathcal{B} \subseteq \widetilde{\mathcal{R}}^{d'}_Q$ are 1-periodic sets of rationals and let $V_{\mathcal{A},\mathcal{B},Q}$ be the kernels defined in (6.9).*

(i) *Let $q_{\mathcal{B}} := \min\{q \in \mathbb{Z}_+ : a/q \in \mathcal{B}$ and $\gcd(a_1,\dots,a_{d'},q) = 1\}$, then for $f \in \ell^2(\mathbb{J}_Q)$ we have*

$$\big\| f *_{\mathbb{J}_Q} V_{\widetilde{\mathcal{R}}^d_Q,\mathcal{B},Q}\big\|_{\ell^2(\mathbb{J}_Q)} \lesssim q_{\mathcal{B}}^{-1/D}\|f\|_{\ell^2(\mathbb{J}_Q)}. \quad (6.19)$$

In particular, if $s \geq 0$ then for $V_s^{\mathrm{low}} := V_{\widetilde{\mathcal{R}}^d_{Q_s},\mathcal{R}^{d'}_s,Q_s}$ inequality (6.19) ensures

$$\big\| f *_{\mathbb{J}_{Q_s}} V_s^{\mathrm{low}}\big\|_{\ell^2(\mathbb{J}_{Q_s})} \lesssim \tau^{-s/D}\|f\|_{\ell^2(\mathbb{J}_{Q_s})}. \quad (6.20)$$

(ii) *Let $q_{\mathcal{A}} := \min\{q \in \mathbb{Z}_+ : a/q \in \mathcal{A}$ and $\gcd(a_1,\dots,a_d,q) = 1\}$. If $q_1 \simeq q_{\mathcal{A}}$ for every $a_1/q_1 \in \mathcal{A}$, and $1 \leq q_2 \lesssim q_{\mathcal{A}}^{1/D}$ for every $a_2/q_2 \in \mathcal{B}$, then for $f \in \ell^2(\mathbb{J}_Q)$ we have*

$$\big\| f *_{\mathbb{J}_Q} V_{\mathcal{A},\mathcal{B},Q}\big\|_{\ell^2(\mathbb{J}_Q)} \lesssim q_{\mathcal{A}}^{-1/D}\|f\|_{\ell^2(\mathbb{J}_Q)}. \quad (6.21)$$

In particular, if $s \geq 0$, $t \geq D(s+1)$, then for $V_{s,t} := V_{\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s},\mathcal{R}^{d'}_s,Q_t}$ we have

$$\big\| f *_{\mathbb{J}_{Q_t}} V_{s,t}\big\|_{\ell^2(\mathbb{J}_{Q_t})} \lesssim \tau^{-t/D}\|f\|_{\ell^2(\mathbb{J}_{Q_t})}. \quad (6.22)$$

Proof As in Sect. 5 we will use a high order T^*T argument.

Step 1. Define the operator $\mathcal{V}_{\mathcal{A},\mathcal{B},Q} f := f *_{\mathbb{J}_Q} V_{\mathcal{A},\mathcal{B},Q}$. For the integer $r = r(d)$ as before we have

$$\{(\mathcal{V}_{\mathcal{A},\mathcal{B},Q})^*\mathcal{V}_{\mathcal{A},\mathcal{B},Q}\}^r f(x) = (f *_{\mathbb{J}_Q} V^r_{\mathcal{A},\mathcal{B},Q})(x),$$

where, as in Sect. 5, we have

$$V^r_{\mathcal{A},\mathcal{B},Q}(y) := \sum_{h_1,g_1,\dots,h_r,g_r\in\mathbb{J}_Q} \Big\{\prod_{j=1}^r \overline{V_{\mathcal{A},\mathcal{B},Q}(h_j)} V_{\mathcal{A},\mathcal{B},Q}(g_j)\Big\} \times \mathbb{1}_{\{0\}}(g_r^{-1}\cdot h_r\cdot\ldots\cdot g_1^{-1}\cdot h_1\cdot y). \quad (6.23)$$

Using the formula

$$\begin{aligned}&\mathbb{1}_{\{0\}}(x^{-1}\cdot y)\\&\quad=Q^{-(d+d')}\sum_{a\in\mathbb{Z}^d_Q\times\mathbb{Z}^{d'}_Q}\mathfrak{e}\big[(y^{(1)}-x^{(1)}).(a^{(1)}/Q)\big]\mathfrak{e}\big[(y^{(2)}-x^{(2)}).(a^{(2)}/Q)\big]\end{aligned}$$

and the definition (6.23) we obtain

$$\begin{aligned}V^r_{\mathcal{A},\mathcal{B},Q}(y)\;&=Q^{-(d+d')}\sum_{a\in\mathbb{Z}^d_Q\times\mathbb{Z}^{d'}_Q}\mathfrak{e}\big[y^{(1)}.(a^{(1)}/Q)\big]\\&\quad\times\mathfrak{e}\big[y^{(2)}.(a^{(2)}/Q)\big]\Upsilon^r_{\mathcal{A},\mathcal{B},Q}\big(a^{(1)}/Q,a^{(2)}/Q\big),\end{aligned}$$

where

$$\begin{aligned}\Upsilon^r_{\mathcal{A},\mathcal{B},Q}\big(\theta^{(1)},\theta^{(2)}\big)&:=\sum_{h_1,g_1,\ldots,h_r,g_r\in\mathbb{J}_Q}\Big\{\prod_{j=1}^r\overline{V_{\mathcal{A},\mathcal{B},Q}(h_j)}V_{\mathcal{A},\mathcal{B},Q}(g_j)\Big\}\\&\times\mathfrak{e}\big(-[h_1^{-1}\cdot g_1\cdot\ldots\cdot h_r^{-1}\cdot g_r]^{(1)}.\theta^{(1)}\big)\mathfrak{e}\big(-[h_1^{-1}\cdot g_1\cdot\ldots\cdot h_r^{-1}\cdot g_r]^{(2)}.\theta^{(2)}\big).\end{aligned}\tag{6.24}$$

Step 2. Taking into account (6.9) we may write

$$\begin{aligned}V_{\mathcal{A},\mathcal{B},Q}(y^{(1)},y^{(2)}):=\;&Q^{-(d+d')}\sum_{\alpha^{(1)}\in(\mathbb{Z}_Q/Q)^d,\,\alpha^{(2)}\in(\mathbb{Z}_Q/Q)^{d'}}m_{\mathcal{A}}(\alpha^{(1)})m_{\mathcal{B}}(\alpha^{(2)})\\&\times\mathfrak{e}[y^{(1)}.\alpha^{(1)}]\mathfrak{e}[y^{(2)}.\alpha^{(2)}],\end{aligned}$$

where $m_{\mathcal{A}}(\alpha^{(1)}):=S(\alpha^{(1)})\mathbb{1}_{\mathcal{A}\cap[0,1)^d}(\alpha^{(1)})$ and $m_{\mathcal{B}}(\alpha^{(2)}):=\mathbb{1}_{\mathcal{B}\cap[0,1)^{d'}}(\alpha^{(2)})$. Using formulas (2.10)–(2.11) we may simplify (6.24). We notice that the sum over the variables $h^{(2)}_j,g^{(2)}_j$, $j\in\{1,\ldots,r\}$ leads to δ-functions in the variables $\theta^{(2)}-\beta^{(2)}_j$ and $\theta^{(2)}-\alpha^{(2)}_j$. Thus

$$\begin{aligned}&\Upsilon^r_{\mathcal{A},\mathcal{B},Q}\big(\theta^{(1)},\theta^{(2)}\big)=\big|m_{\mathcal{B}}(\theta^{(2)})\big|^{2r}Q^{-2rd}\\&\Big\{\sum_{\beta^{(1)}_1,\alpha^{(1)}_1,\ldots,\beta^{(1)}_r,\alpha^{(1)}_r\in(\mathbb{Z}_Q/Q)^d}\;\sum_{h^{(1)}_1,g^{(1)}_1,\ldots,h^{(1)}_r,g^{(1)}_r\in\mathbb{Z}^d_Q}\\&\times\prod_{j=1}^r\big\{\overline{m_{\mathcal{A}}(\beta^{(1)}_j)}\mathfrak{e}\big[h^{(1)}_j.(\theta^{(1)}-\beta^{(1)}_j)\big]\cdot m_{\mathcal{A}}(\alpha^{(1)}_j)\mathfrak{e}\big[-g^{(1)}_j.(\theta^{(1)}-\alpha^{(1)}_j)\big]\big\}\\&\times\mathfrak{e}\Big[-\theta^{(2)}.\Big(\sum_{1\le j\le r}R_0(h^{(1)}_j,h^{(1)}_j-g^{(1)}_j)\end{aligned}$$

$$+\sum_{1\le l<j\le r} R_0(-h_l^{(1)}+g_l^{(1)},-h_j^{(1)}+g_j^{(1)})\Big)\Big]\Big\}. \tag{6.25}$$

Step 3. Our aim now is to show that

$$\big\|V^r_{\widetilde{\mathcal{R}}^d_Q,\mathcal{B},Q}\big\|_{\ell^1(\mathbb{J}_Q)}\lesssim q_{\mathcal{B}}^{-1}. \tag{6.26}$$

This will establish (6.19) and (6.20), by taking $Q=Q_s$ and $\mathcal{B}=\mathcal{R}^{d'}_s$. To prove (6.26) it suffices to show

$$\big|\Upsilon^r_{\widetilde{\mathcal{R}}^d_Q,\mathcal{B},Q}\big(\theta^{(1)},\theta^{(2)}\big)\big|\lesssim (q_1+q_2)^{-1/\delta^4}\mathbb{1}_{\mathcal{B}\cap[0,1)^{d'}}(\theta^{(2)}), \tag{6.27}$$

where $q_1|Q$, $q_2|Q$ are the denominators of the irreducible representation of the fractions $\theta^{(1)}$ and $\theta^{(2)}$ respectively.

Inserting the formula $S(\gamma^{(1)})=Q^{-1}\sum_{n\in\mathbb{Z}_Q}\mathfrak{e}[-A_0^{(1)}(n).\gamma^{(1)}]$, see (6.4), into the identity (6.25) with $\mathcal{A}=\widetilde{\mathcal{R}}^d_Q$, we notice that the sums over the variables $\alpha_j^{(1)}$ and $\beta_j^{(1)}$ lead to δ-functions. More precisely,

$$\begin{aligned}\Upsilon^r_{\widetilde{\mathcal{R}}^d_Q,\mathcal{B},Q}\big(\theta^{(1)},\theta^{(2)}\big)&=\mathbb{1}_{\mathcal{B}\cap[0,1)^{d'}}(\theta^{(2)})Q^{-2r}\\&\times\Big\{\sum_{n_j,m_j\in\mathbb{Z}_Q}\mathfrak{e}\Big[\theta^{(1)}.\Big(\sum_{1\le j\le r}A_0^{(1)}(n_j)-A_0^{(1)}(m_j)\Big)\Big]\\&\times\mathfrak{e}\Big[-\theta^{(2)}.\Big(\sum_{1\le j\le r}R_0\big(A_0^{(1)}(n_j),A_0^{(1)}(n_j)-A_0^{(1)}(m_j)\big)\\&+\sum_{1\le l<j\le r}R_0\big(A_0^{(1)}(m_l)-A_0^{(1)}(n_l),A_0^{(1)}(m_j)-A_0^{(1)}(n_j)\big)\Big)\Big]\Big\}\\&=\mathbb{1}_{\mathcal{B}\cap[0,1)^{d'}}(\theta^{(2)})Q^{-2r}\sum_{n,m\in\mathbb{Z}_Q^r}\mathfrak{e}\big[-D(n,m).(\theta^{(1)},\theta^{(2)})\big],\end{aligned}$$

where $D(n,m)$ is defined in (2.24). Using Proposition 2.3 (ii) we obtain (6.27) as desired.

Step 4. To prove (6.21) as well as (6.22) with $Q=Q_t$ and $\mathcal{A}=\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s}$, $\mathcal{B}=\mathcal{R}^{d'}_s$ we show

$$\big\|V^r_{\mathcal{A},\mathcal{B},Q}\big\|_{\ell^1(\mathbb{J}_Q)}\lesssim q_{\mathcal{A}}^{-1}. \tag{6.28}$$

We still use the formula (6.25), with $\mathcal{A}\subseteq\widetilde{\mathcal{R}}^d_Q$ and $\mathcal{B}\subseteq\widetilde{\mathcal{R}}^{d'}_Q$ satisfying $q_1\simeq q_{\mathcal{A}}$ for every $a_1/q_1\in\mathcal{A}$, and $1\le q_2\lesssim q_{\mathcal{A}}^{1/D}$ for every $a_2/q_2\in\mathcal{B}$. We would like to first evaluate the sums over the variables $h_j^{(1)}$ and $g_j^{(1)}$; these sums would lead to δ-functions if $\theta^{(2)}=0$, but there is an obstruction for other

values of $\theta^{(2)}$. However, we can exploit the fact that the denominators of fractions $\theta^{(2)}$ are small. Indeed, assume that $a^{(2)}/q_2 = \theta^{(2)}$ is the irreducible representation of the fraction $\theta^{(2)}$, where $1 \le q_2 \lesssim q_{\mathcal{A}}^{1/D}$ and q_2 divides Q. For $j \in \{1, \dots, r\}$ we decompose $h_j^{(1)} = q_2 y_j + y_j'$, $g_j^{(1)} = q_2 x_j + x_j'$, $y_j', x_j' \in \mathbb{Z}^d_{q_2}$, $y_j, x_j \in \mathbb{Z}^d_{Q/q_2}$. Then we notice that

$$(Q/q_2)^{-2rd} \sum_{y_1, x_1, \dots, y_r, x_r \in \mathbb{Z}^d_{Q/q_2}} \prod_{j=1}^{r} \mathfrak{e}\big[q_2 y_j.(\theta^{(1)} - \beta_j^{(1)})\big]\mathfrak{e}\big[-q_2 x_j.(\theta^{(1)} - \alpha_j^{(1)})\big]$$
$$= \prod_{j=1}^{r} \mathbb{1}_{\mathbb{Z}^d}\big[q_2(\theta^{(1)} - \beta_j^{(1)})\big]\mathbb{1}_{\mathbb{Z}^d}\big[q_2(\theta^{(1)} - \alpha_j^{(1)})\big].$$

Therefore, using formula (6.25), one sees

$$\big|\Upsilon^r_{\mathcal{A},\mathcal{B},Q}\big(\theta^{(1)}, \theta^{(2)}\big)\big| \le \mathbb{1}_{\mathcal{B}\cap[0,1)^{d'}}(\theta^{(2)})\Big\{ \sum_{\beta_1^{(1)}, \alpha_1^{(1)}, \dots, \beta_r^{(1)}, \alpha_r^{(1)} \in (\mathbb{Z}_Q/Q)^d} \prod_{j=1}^{r} \mathbb{1}_{\mathbb{Z}^d}\big[q_2(\theta^{(1)} - \beta_j^{(1)})\big]$$
$$\times \mathbb{1}_{\mathbb{Z}^d}\big[q_2(\theta^{(1)} - \alpha_j^{(1)})\big]|m_{\mathcal{A}}(\beta_j^{(1)})||m_{\mathcal{A}}(\alpha_j^{(1)})|\Big\}.$$

Recall that $m_{\mathcal{A}}(\gamma) = S(\gamma)\mathbb{1}_{\mathcal{A}\cap[0,1)^d}(\gamma)$. It follows from Proposition 2.2 (ii) that for any $\gamma \in \mathcal{A}$ we have $|m_{\mathcal{A}}(\gamma)| \lesssim q_{\mathcal{A}}^{-1/\overline{C}}$, since $q_1 \simeq q_{\mathcal{A}}$ for every $a_1/q_1 \in \mathcal{A}$. Therefore

$$\big|\Upsilon^r_{\mathcal{A},\mathcal{B},Q}\big(\theta^{(1)}, \theta^{(2)}\big)\big| \lesssim \mathbb{1}_{\mathcal{B}\cap[0,1)^{d'}}(\theta^{(2)})\mathbb{1}_{(\mathcal{A}+(\mathbb{Z}/q_2)^d)\cap[0,1)^d}(\theta^{(1)})q_{\mathcal{A}}^{-2r/\overline{C}}q_2^{2dr},$$

where $\mathcal{A} + (\mathbb{Z}/q_2)^d := \{a/q_2 + \theta : \theta \in \mathcal{A}, a \in \mathbb{Z}^d\}$. The desired bound (6.28) follows since $1 \le q_2 \lesssim q_{\mathcal{A}}^{1/D}$ for every $a_2/q_2 \in \mathcal{B}$, and $r \in \mathbb{Z}_+$ is sufficiently large. □

6.3 Maximal and variational operators on the group $\mathbb{H}_Q$

The main result of this subsection is the following lemma:

Lemma 6.4 *Assume that $2 < \rho < \infty$, $\tau \in (1, 2]$, and $k, k_0, w \in \mathbb{N}$ satisfy $0 \le w \le k$ and $k \ge k_0 \ge D/\ln\tau$. Assume that $1 \le Q \le \tau^{\delta k}$ and let $W_{k,w,Q} : \mathbb{H}_Q \to \mathbb{C}$ be defined as in (6.8). Then, for any $f \in \ell^2(\mathbb{H}_Q)$ and*

$\mathbb{D} \subseteq \mathbb{N}$ *one has*

$$\left\| V^{\rho}(f *_{\mathbb{H}_Q} W_{k,k,Q} : k \in \mathbb{D}_{k_0,Q}) \right\|_{\ell^2(\mathbb{H}_Q)} \lesssim \|f\|_{\ell^2(\mathbb{H}_Q)}, \tag{6.29}$$

uniformly in Q, *where* $\mathbb{D}_{k_0,Q} := \{k \in \mathbb{D} : k \geq k_0,\ \tau^{\delta k} \geq Q\}$.

Moreover, for every $w \in \mathbb{N}$ *and every sequence* $\{\varkappa_k\}_{k\in\mathbb{N}} \subseteq \mathbb{C}$ *satisfying* $\sup_{k\in\mathbb{N}} |\varkappa_k| \leq 1$,

$$\Big\| \sum_{k\in\mathbb{D}_{k_0,Q},\, k>w} \varkappa_k f *_{\mathbb{H}_Q} (W_{k,w+1,Q} - W_{k,w,Q}) \Big\|_{\ell^2(\mathbb{H}_Q)} \lesssim \tau^{-w/D^2} \|f\|_{\ell^2(\mathbb{H}_Q)}, \tag{6.30}$$

for any $f \in \ell^2(\mathbb{H}_Q)$, *uniformly in* Q.

The main idea to prove (6.29)–(6.30) is to compare our operators with suitable operators on the Lie group $\mathbb{G}_0^{\#}$. More precisely for $0 \leq w \leq k$ we define the kernels $\widetilde{W}_{k,w} : \mathbb{G}_0^{\#} \to \mathbb{C}$ by

$$\widetilde{W}_{k,w}(x) := \phi_k(x) \int_{\mathbb{R}^d \times \mathbb{R}^{d'}} \eta_{\leq \delta' w}(\tau^k \circ \xi) \eta_{\leq \delta w}(\tau^k \circ \theta) \mathfrak{e}(x.(\xi,\theta)) J_k(\xi)\, d\xi d\theta, \tag{6.31}$$

where $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^d \times \mathbb{R}^{d'} = \mathbb{G}_0^{\#}$ and $\phi_k(x) = \phi_k^{(1)}(x^{(1)})\phi_k^{(2)}(x^{(2)})$.

Then we have a continuous version of Lemma 6.4:

Proposition 6.5 *Assume that* $2 < \rho < \infty$, $\tau \in (1,2]$, *and* $k, w \in \mathbb{N}$ *satisfy* $0 \leq w \leq k$. *With* $\widetilde{W}_{k,w} : \mathbb{G}_0^{\#} \to \mathbb{C}$ *defined as in* (6.31), *for any* $f \in L^2(\mathbb{G}_0^{\#})$ *one has*

$$\left\| V^{\rho}(f *_{\mathbb{G}_0^{\#}} \widetilde{W}_{k,k} : k \geq 0) \right\|_{L^2(\mathbb{G}_0^{\#})} \lesssim \|f\|_{L^2(\mathbb{G}_0^{\#})}. \tag{6.32}$$

In particular, one has

$$\left\| \sup_{k\geq 0} |f *_{\mathbb{G}_0^{\#}} \widetilde{W}_{k,k}| \right\|_{L^2(\mathbb{G}_0^{\#})} \lesssim \|f\|_{L^2(\mathbb{G}_0^{\#})}. \tag{6.33}$$

Moreover, for any $w \in \mathbb{N}$, *any sequence* $\{\varkappa_k\}_{k\in\mathbb{N}} \subseteq \mathbb{C}$ *satisfying* $\sup_{k\in\mathbb{N}} |\varkappa_k| \leq 1$, *and any* $f \in L^2(\mathbb{G}_0^{\#})$ *one has*

$$\Big\| \sum_{k>w} \varkappa_k f *_{\mathbb{G}_0^{\#}} (\widetilde{W}_{k,w+1} - \widetilde{W}_{k,w}) \Big\|_{L^2(\mathbb{G}_0^{\#})} \lesssim \tau^{-w/D} \|f\|_{L^2(\mathbb{G}_0^{\#})}. \tag{6.34}$$

Continuous maximal operators such as (6.33) have been extensively studied, see for example the conclusive work of Christ–Nagel–Stein–Wainger [19]. However, the variational estimates in the nilpotent setting in the spirit of [19] appear to be new. For the convenience of the reader we provide a self-contained proof of Proposition 6.5 in Appendix A. Assuming that Proposition 6.5 holds, we show how to use it to deduce Lemma 6.4.

Proof of Lemma 6.4 We define the Q-cubes

$$\mathcal{C}_Q := [0, Q)^d \times [0, Q)^{d'} \subseteq \mathbb{G}_0^{\#}, \tag{6.35}$$

and notice that the map $(\mu, h) \mapsto \mu \cdot h$ defines a measure-preserving bijection from $\mathcal{C}_Q \times \mathbb{H}_Q$ to $\mathbb{G}_0^{\#}$. Let $1 \le p < \infty$. Given $f \in \ell^p(\mathbb{H}_Q)$ we define

$$\begin{aligned} &f^{\#}(\mu \cdot h) := f(h) \text{ for any } (\mu, h) \in \mathcal{C}_Q \times \mathbb{H}_Q, \\ &f^{\#} \in L^p(\mathbb{G}_0^{\#}), \quad \|f^{\#}\|_{L^p(\mathbb{G}_0^{\#})} = Q^{(d+d')/p} \|f\|_{\ell^p(\mathbb{H}_Q)}. \end{aligned} \tag{6.36}$$

We now prove the following bounds: for any $1 \le p < \infty$ and $2 < \rho < \infty$ we have

$$\begin{aligned} \big\| V^{\rho}(f *_{\mathbb{H}_Q} W_{k,k,Q} : k \in \mathbb{D}_{k_0,Q}) \big\|_{\ell^p(\mathbb{H}_Q)} \\ \lesssim Q^{-(d+d')/p} \big\| V^{\rho}(f^{\#} *_{\mathbb{G}_0^{\#}} \widetilde{W}_{k,k} : k \ge 0) \big\|_{L^p(\mathbb{G}_0^{\#})} \\ + \|f\|_{\ell^p(\mathbb{H}_Q)}, \end{aligned} \tag{6.37}$$

and

$$\begin{aligned} &\Big\| \sum_{k \in \mathbb{D}_{k_0,Q},\, k > w} \varkappa_k f *_{\mathbb{H}_Q} (W_{k,w+1,Q} - W_{k,w,Q}) \Big\|_{\ell^p(\mathbb{H}_Q)} \\ &\lesssim Q^{-(d+d')/p} \Big\| \sum_{k \in \mathbb{D}_{k_0,Q},\, k > w} \varkappa_k f^{\#} *_{\mathbb{G}_0^{\#}} (\widetilde{W}_{k,w+1} - \widetilde{W}_{k,w}) \Big\|_{L^p(\mathbb{G}_0^{\#})} \\ &\quad + \tau^{-w/8} \|f\|_{\ell^p(\mathbb{H}_Q)}. \end{aligned} \tag{6.38}$$

It is easy to see that the inequalities (6.37)–(6.38) with $p = 2$ can be combined with (6.32), (6.34), and (6.36) to complete the proof of Lemma 6.4.

It remains to prove the bounds (6.37)–(6.38). For this we compare the functions $f *_{\mathbb{H}_Q} W_{k,w,Q} : \mathbb{H}_Q \to \mathbb{C}$ and $f^{\#} *_{\mathbb{G}_0^{\#}} \widetilde{W}_{k,w} : \mathbb{G}_0^{\#} \to \mathbb{C}$. By (6.8) and (6.31), we have $Q^{d+d'} \widetilde{W}_{k,w}(h) = W_{k,w,Q}(h)$ for any $h \in \mathbb{H}_Q$. Moreover, by (6.31) notice that

$$\big|\widetilde{W}_{k,w}(\mu_1 \cdot h \cdot \mu_2) - \widetilde{W}_{k,w}(h)\big| \lesssim E_k(h), \tag{6.39}$$

where

$$E_k(h) := \tau^{-k/2}\Big\{\prod_{(l_1,l_2)\in Y_d}\tau^{-(l_1+l_2)k}\Big\}\eta_{\le 2\delta k}(\tau^{-k}\circ h^{(1)})\eta_{\le 2\delta k}(\tau^{-k}\circ h^{(2)}),$$

for any $h, \mu_1, \mu_2 \in \mathbb{G}_0^{\#}$ with $|\mu_1| + |\mu_2| \lesssim Q^4$, provided that $k \ge D/\ln\tau$, $0 \le w \le k$ and $1 \le Q \le \tau^{\delta k}$. Thus

$$\begin{aligned}(f^{\#} *_{\mathbb{G}_0^{\#}} \widetilde{W}_{k,w})(\mu\cdot h) &= \sum_{h_1\in\mathbb{H}_Q}\int_{\mathcal{C}_Q} f^{\#}(\mu_1\cdot h_1)\widetilde{W}_{k,w}(\mu\cdot h\cdot h_1^{-1}\cdot\mu_1^{-1})\,d\mu_1\\ &= \sum_{h_1\in\mathbb{H}_Q} f(h_1)\int_{\mathcal{C}_Q}\widetilde{W}_{k,w}(\mu\cdot h\cdot h_1^{-1}\cdot\mu_1^{-1})\,d\mu_1,\end{aligned}$$

for any $(\mu, h) \in \mathcal{C}_Q \times \mathbb{H}_Q$. Using (6.39) we have

$$\Big|\int_{\mathcal{C}_Q}\widetilde{W}_{k,w}(\mu\cdot h\cdot h_1^{-1}\cdot\mu_1^{-1})\,d\mu_1 - W_{k,w,Q}(h\cdot h_1^{-1})\Big| \lesssim E_k(h\cdot h_1^{-1})Q^{d+d'}.$$

Therefore, for any $f \in \ell^p(\mathbb{H}_Q)$, $h \in \mathbb{H}_Q$ and $\mu \in \mathcal{C}_Q$, one has

$$(f *_{\mathbb{H}_Q} W_{k,w,Q})(h) = (f^{\#} *_{\mathbb{G}_0^{\#}} \widetilde{W}_{k,w})(\mu\cdot h) + O\big(\tau^{k/4}(|f| *_{\mathbb{H}_Q} E_k)(h)\big),$$

provided that $k \ge D/\ln\tau$, $0 \le w \le k$ and $1 \le Q \le \tau^{\delta k}$. The desired bounds (6.37) and (6.38) follow from the last identity and the observation that $\sum_{k\ge w}\tau^{k/4}\|E_k\|_{\ell^1(\mathbb{H}_Q)} \lesssim \tau^{-w/8}$ for any $w \in \mathbb{N}$. □

6.4 Proof of Lemma 4.5

We begin with a transference lemma which will be used repeatedly.

Lemma 6.6 *As in Lemma 6.4, assume that $2 < \rho < \infty$, $\tau \in (1, 2]$, $k \ge k_0$, and $1 \le Q \le \tau^{\delta k}$. Assume that $K_k^{\mathbb{G}_0} : \mathbb{G}_0 \to \mathbb{C}$ are given kernels such that*

$$K_k^{\mathbb{G}_0}(b_1\cdot h\cdot b_2) := W_k^{\mathbb{H}_Q}(h)V^{\mathbb{J}_Q}(b_1\cdot b_2) + E_k(h, b_1, b_2), \qquad (6.40)$$

for any $h \in \mathbb{H}_Q$ *and* $b_1, b_2 \in \mathbb{G}_0$ *satisfying* $|b_j| \leq Q^4$, $j \in \{1, 2\}$, *for some kernels* $W_k^{\mathbb{H}_Q} : \mathbb{H}_Q \to \mathbb{C}$ *and* $V^{\mathbb{J}_Q} : \mathbb{J}_Q \to \mathbb{C}$, *where the error terms satisfy the estimates*

$$\sup_{|b_1|,|b_2|\leq Q^4} \|E_k(\cdot, b_1, b_2)\|_{\ell^1(\mathbb{H}_Q)} \lesssim \tau^{-k/3}. \tag{6.41}$$

Let $\mathbb{D} \subseteq \mathbb{N}$ *and* $\mathbb{D}_{k_0,Q} = \{k \in \mathbb{D} : k \geq k_0,\ \tau^{\delta k} \geq Q\}$ *as in Lemma 6.4. Let also* $\mathcal{K}_k^{\mathbb{G}_0} f := f *_{\mathbb{G}_0} K_k^{\mathbb{G}_0}$, *and* $\mathcal{W}_k^{\mathbb{H}_Q} g := g *_{\mathbb{H}_Q} W_k^{\mathbb{H}_Q}$, *and* $\mathcal{V}^{\mathbb{J}_Q} h := h *_{\mathbb{J}_Q} V^{\mathbb{J}_Q}$ *denote the convolution operators corresponding to the kernels* $K_k^{\mathbb{G}_0}$, $W_k^{\mathbb{H}_Q}$, *and* $V^{\mathbb{J}_Q}$.

Then for any $1 \leq p < \infty$ *and either* $B = V^\rho$ *or* $B = \ell^\infty$

$$\begin{aligned}
&\big\|\big(\mathcal{K}_k^{\mathbb{G}_0}\big)_{k\in\mathbb{D}_{k_0,Q}}\big\|_{\ell^p(\mathbb{G}_0)\to\ell^p(\mathbb{G}_0;B)}\\
&\quad\lesssim \big\|\big(\mathcal{W}_k^{\mathbb{H}_Q}\big)_{k\in\mathbb{D}_{k_0,Q}}\big\|_{\ell^p(\mathbb{H}_Q)\to\ell^p(\mathbb{H}_Q;B)}\big\|\mathcal{V}^{\mathbb{J}_Q}\big\|_{\ell^p(\mathbb{J}_Q)\to\ell^p(\mathbb{J}_Q)}\\
&\quad+\tau^{-k_0/8}Q^{-1/(8\delta)}.
\end{aligned} \tag{6.42}$$

Moreover, for any sequence $\{\varkappa_k\}_{k\in\mathbb{N}} \subseteq \mathbb{C}$ *satisfying* $\sup_{k\in\mathbb{N}} |\varkappa_k| \leq 1$

$$\begin{aligned}
\Big\|\sum_{k\in\mathbb{D}_{k_0,Q}} \varkappa_k \mathcal{K}_k^{\mathbb{G}_0}\Big\|_{\ell^p(\mathbb{G}_0)\to\ell^p(\mathbb{G}_0)} \lesssim &\Big\|\sum_{k\in\mathbb{D}_{k_0,Q}} \varkappa_k \mathcal{W}_k^{\mathbb{H}_Q}\Big\|_{\ell^p(\mathbb{H}_Q)\to\ell^p(\mathbb{H}_Q)}\\
&\times \big\|\mathcal{V}^{\mathbb{J}_Q}\big\|_{\ell^p(\mathbb{J}_Q)\to\ell^p(\mathbb{J}_Q)}\\
&+\tau^{-k_0/8}Q^{-1/(8\delta)}.
\end{aligned} \tag{6.43}$$

Proof Using (6.40) for $b \in \mathbb{J}_Q$ and $h \in \mathbb{H}_Q$ we may write

$$\begin{aligned}
(f * K_k^{\mathbb{G}_0})(b\cdot h) &= \sum_{h_1\in\mathbb{H}_Q,\, b_1\in\mathbb{J}_Q} f(b_1\cdot h_1) K_k^{\mathbb{G}_0}(b\cdot h\cdot h_1^{-1}\cdot b_1^{-1})\\
&= \sum_{h_1\in\mathbb{H}_Q,\, b_1\in\mathbb{J}_Q} f(b_1\cdot h_1)\big\{W_k^{\mathbb{H}_Q}(h\cdot h_1^{-1})V^{\mathbb{J}_Q}(b\cdot b_1^{-1})\\
&\quad + E_k(h\cdot h_1^{-1}, b, b_1^{-1})\big\}.
\end{aligned} \tag{6.44}$$

For any $h' \in \mathbb{H}_Q$ and $b \in \mathbb{J}_Q$ let $F_{\mathbb{J}_Q}(b, h') := \sum_{b_1\in\mathbb{J}_Q} f(b_1\cdot h')V^{\mathbb{J}_Q}(b\cdot b_1^{-1})$. We also take

$$F_k(h,b) := \sum_{h_1 \in \mathbb{H}_Q} F_{\mathbb{J}_Q}(b,h_1) W_k^{\mathbb{H}_Q}(h \cdot h_1^{-1})$$
$$G_k(h,b,b_1) := \sum_{h_1 \in \mathbb{H}_Q} |f(b_1 \cdot h_1) E_k(h \cdot h_1^{-1}, b, b_1^{-1})|.$$

Then by (6.44) we have

$$\begin{aligned} &\big\| V^\rho(f * K_k^{\mathbb{G}_0} : k \in \mathbb{D}_{k_0,Q}) \big\|_{\ell^p(\mathbb{G}_0)} \\ &\le \Big(\sum_{h \in \mathbb{H}_Q,\, b \in \mathbb{J}_Q} V^\rho\big(F_k(h,b) : k \in \mathbb{D}_{k_0,Q}\big)^p \Big)^{1/p} \\ &\quad + 2 \sum_{b,b_1 \in \mathbb{J}_Q} \Big(\sum_{h \in \mathbb{H}_Q} \Big(\sum_{k \in \mathbb{D}_{k_0,Q}} |G_k(h,b,b_1)|^\rho \Big)^{p/\rho} \Big)^{1/p} \\ &=: I_1 + I_2. \end{aligned} \tag{6.45}$$

For the first sum in (6.45) we now see that

$$I_1 \le \big\| \big(\mathcal{W}_k^{\mathbb{H}_Q}\big)_{k \in \mathbb{D}_{k_0,Q}} \big\|_{\ell^p(\mathbb{H}_Q) \to \ell^p(\mathbb{H}_Q; V^\rho)} \big\| \mathcal{V}^{\mathbb{J}_Q} \big\|_{\ell^p(\mathbb{J}_Q) \to \ell^p(\mathbb{J}_Q)} \|f\|_{\ell^p(\mathbb{G}_0)},$$

whereas for the second one we use (6.41) to conclude that $I_2 \lesssim \tau^{-k_0/8} Q^{-1/(8\delta)} \|f\|_{\ell^p(\mathbb{G}_0)}$. This proves (6.42) when $B = V^\rho$. The remaining conclusions of the lemma follow in a similar way. □

We now establish a slightly more general result for the kernels $K_{k,w,\mathcal{A},\mathcal{B}} : \mathbb{G}_0 \to \mathbb{C}$ as in (6.7). Let $\mathcal{V}_{\mathcal{A},\mathcal{B},Q} f := f *_{\mathbb{J}_Q} V_{\mathcal{A},\mathcal{B},Q}$ denote the convolution operator corresponding to the kernel $V_{\mathcal{A},\mathcal{B},Q} : \mathbb{J}_Q \to \mathbb{C}$ from (6.9).

Lemma 6.7 *As in Lemma 6.4, assume that $\rho \in (2,\infty)$, $\tau \in (1,2]$, and $k, k_0, w, Q \in \mathbb{N}$ satisfy $0 \le w \le k$, $k \ge k_0 \ge D/\ln \tau$, and $1 \le Q \le \tau^{\delta k}$. Assume that $\mathcal{A} \subseteq \widetilde{\mathcal{R}}_Q^d$ and $\mathcal{B} \subseteq \widetilde{\mathcal{R}}_Q^{d'}$ are 1-periodic sets of rationals. Then, for any $f \in \ell^2(\mathbb{G}_0)$ and $\mathbb{D} \subseteq \mathbb{N}$ we have*

$$\begin{aligned} &\big\| V^\rho(f * K_{k,k,\mathcal{A},\mathcal{B}} : k \in \mathbb{D}_{k_0,Q}) \big\|_{\ell^2(\mathbb{G}_0)} \\ &\lesssim \big(\|\mathcal{V}_{\mathcal{A},\mathcal{B},Q}\|_{\ell^2(\mathbb{J}_Q) \to \ell^2(\mathbb{J}_Q)} + \tau^{-k_0/8} Q^{-1/(8\delta)} \big) \|f\|_{\ell^2(\mathbb{G}_0)}, \end{aligned} \tag{6.46}$$

uniformly in Q and $k_0 \ge D/\ln \tau$, where as before $\mathbb{D}_{k_0,Q} = \{k \in \mathbb{D} : k \ge k_0,\ \tau^{\delta k} \ge Q\}$. Moreover for any sequence $\{\varkappa_k\}_{k \in \mathbb{N}} \subseteq \mathbb{C}$ satisfying $\sup_{k \in \mathbb{N}} |\varkappa_k| \le 1$, any $f \in \ell^2(\mathbb{G}_0)$, and any $Q \in \mathbb{Z}_+$, $w \in \mathbb{N}$ we have

$$\begin{aligned}\Big\| \sum_{k\in\mathbb{D}_{k_0,Q},\, k>w} \varkappa_k f * (K_{k,w+1,\mathcal{A},\mathcal{B}} - K_{k,w,\mathcal{A},\mathcal{B}})\Big\|_{\ell^2(\mathbb{G}_0)} \\ \lesssim \big(\tau^{-w/D^2}\|\mathcal{V}_{\mathcal{A},\mathcal{B},Q}\|_{\ell^2(\mathbb{J}_Q)\to\ell^2(\mathbb{J}_Q)} \\ + \tau^{-\max(k_0,w)/8} Q^{-1/(8\delta)}\big)\|f\|_{\ell^2(\mathbb{G}_0)}.\end{aligned} \tag{6.47}$$

Proof To prove (6.46) we use Lemma 6.6 with $K_k^{\mathbb{G}_0} = K_{k,k,\mathcal{A},\mathcal{B}}$, $W_k^{\mathbb{H}_Q} = W_{k,k,Q}$ and $V^{\mathbb{J}_Q} = V_{\mathcal{A},\mathcal{B},Q}$ as in Lemma 6.2. The assumptions (6.40)–(6.41) in Lemma 6.6 follow from (6.7) and (6.10). The bounds (6.46) follow from (6.42) with $p = 2$ and (6.29).

On the other hand, taking $K_k^{\mathbb{G}_0} = K_{k,w+1,\mathcal{A},\mathcal{B}} - K_{k,w,\mathcal{A},\mathcal{B}}$, $W_k^{\mathbb{H}_Q} = W_{k,w+1,Q} - W_{k,w,Q}$, and $V^{\mathbb{J}_Q} = V_{\mathcal{A},\mathcal{B},Q}$, the bounds (6.47) follow from (6.43) and (6.30). □

We are now finally ready to complete the proof of Lemma 4.5.

Proof of Lemma 4.5 Notice that $G_{k,s}^{\mathrm{low}} = K_{k,k,\widetilde{\mathcal{R}}^d_{Q_s},\mathcal{R}^{d'}_s}$. We use (6.46) with $Q = Q_s$ and $k_0 = \kappa_s$; in view of (6.20) we have $\|V_{\widetilde{\mathcal{R}}^d_{Q_s},\mathcal{R}^{d'}_s,Q_s}\|_{\ell^2(\mathbb{J}_{Q_s})\to\ell^2(\mathbb{J}_{Q_s})} \lesssim \tau^{-s/D}$, and the bounds (4.30)–(4.31) follow from (6.46).

Assuming that $s \geq 0, t \geq D(s+1)$ and taking $\mathcal{A} \subseteq \mathcal{R}^d_t \setminus \widetilde{\mathcal{R}}^d_{Q_s}$ and $\mathcal{B} \subseteq \mathcal{R}^{d'}_{\leq s}$ we conclude, using (6.21) and (6.46) with $Q = Q_t$ and $k_0 = \kappa_t$, that

$$\big\|V^\rho(f * K_{k,k,\mathcal{A},\mathcal{B}} : k \geq \kappa_t)\big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D}\|f\|_{\ell^2(\mathbb{G}_0)} \tag{6.48}$$

for any $2 < \rho < \infty$, as well as

$$\Big\| \sup_{k\geq\kappa_t} |f * K_{k,k,\mathcal{A},\mathcal{B}}|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{6.49}$$

The desired bounds (4.32)–(4.33) follow since $G_{k,s,t} = K_{k,k,\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s},\mathcal{R}^{d'}_s}$. □

7 Transition estimates I: Proof of Lemma 4.3

In this section we prove the bounds (4.19)–(4.20). Let $H_{k,s} := K_{k+1,s} - K_{k,s}$ for $k \geq j_0 := \max((D/\ln\tau)^2, s/\delta)$ and apply the Rademacher–Menshov inequality (2.7) with $m = \lfloor (D/\ln\tau)(s+1)^2\rfloor + 4$. For (4.19) it suffices to prove for any fixed $i \in [0, m]$ that

$$\Big\|\Big(\sum_{j\in[j_0 2^{-i}, 2^{m-i}-1]} \Big|\sum_{k\in[j2^i,(j+1)2^i-1]} f * H_{k,s}\Big|^2\Big)^{1/2}\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-2s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}.$$

Using Khintchine's inequality and dividing again dyadically, for (4.19) it suffices to prove that

$$\Big\| \sum_{k\in[J,2J]} \varkappa_k (f * H_{k,s}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-4s/D^2} \big\| f \big\|_{\ell^2(\mathbb{G}_0)} \tag{7.1}$$

for any $J \geq \max((D/\ln\tau)^2, s/\delta)$ and any coefficients $\varkappa_k \in [-1, 1]$.

To prove (7.1) we examine the definition (4.15) and the further decompose

$$\begin{aligned} H_{k,s} &= H^1_{k,s} + H^2_{k,s} + H^3_{k,s},\\ H^1_{k,s}(g) &:= [\Delta_k L_k](g^{(1)})\phi^{(2)}_k(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_{k,s}(\xi^{(2)})\,d\xi^{(2)},\\ H^2_{k,s}(g) &:= L_{k+1}(g^{(1)})\Delta_k[\phi^{(2)}_k](g^{(2)})\big\} \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\,\Xi_{k,s}(\xi^{(2)})\,d\xi^{(2)},\\ H^3_{k,s}(g) &:= L_{k+1}(g^{(1)})\phi^{(2)}_{k+1}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})[\Delta_k\,\Xi_{k,s}](\xi^{(2)})\,d\xi^{(2)}. \end{aligned} \tag{7.2}$$

We will prove that, for any $k \geq \max((D/\ln\tau)^2, s/\delta)$ and $\iota \in \{2, 3\}$,

$$\big\| f * H^\iota_{k,s} \big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-k/D} \big\| f \big\|_{\ell^2(\mathbb{G}_0)}. \tag{7.3}$$

We will also prove that

$$\Big\| \sum_{k\in[J,2J]} \varkappa_k (f * H^1_{k,s}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-s/D} \big\| f \big\|_{\ell^2(\mathbb{G}_0)} \tag{7.4}$$

for any $J \geq \max((D/\ln\tau)^2, s/\delta)$ and any coefficients $\varkappa_k \in [-1, 1]$. These two bounds would clearly imply the bounds (7.1).

7.1 Proof of (7.3)

Step 1. Assume first that $\iota = 2$ and recall the definition of the functions $\phi^{(2)}_k$ in (4.10). Notice that if $g = (g^{(1)}, g^{(2)})$ is in the support of the kernel $H^2_{k,s}$ then there is $(l_1, l_2) \in Y'_d$ such that $|g^{(2)}_{l_1 l_2}| \gtrsim \tau^{k(l_1+l_2)}$. Therefore we can integrate by parts many times in the variable $\xi^{(2)}_{l_1 l_2}$ (recall the definition (4.12)) to prove that the kernels $H^2_{k,s}$ decay rapidly in k, i.e. $|H^2_{k,s}(g)| \lesssim \tau^{-k/\delta}$ for any $g \in \mathbb{G}_0$. The desired bounds (7.3) follow.

Step 2. Assume now that $\iota = 3$. In this case we use a high order T^*T argument as in Sect. 5. Notice that the kernels $H^3_{k,s}$ have product structure, so we can apply the identities (2.13)–(2.16). With r being a sufficiently large

integer such that the bounds in Propositions 2.3 and 2.4 hold with $\varepsilon = \delta^4$, it suffices to prove that

$$\big|\Pi_{k+1}^{c,r}\big(\theta^{(1)},\theta^{(2)}\big)\Gamma_{k,s}^{r}\big(\theta^{(2)}\big)\big| \lesssim \tau^{-k/\delta} \quad \text{for any} \quad (\theta^{(1)},\theta^{(2)}) \in \mathbb{T}^d \times \mathbb{T}^{d'}, \quad k \geq (D/\ln\tau)^2, \tag{7.5}$$

where $\Pi_k^{c,r}$ is as in (5.2) and

$$\Gamma_{k,s}^{r}(\theta^{(2)}) = \Big| \int_{\mathbb{T}^{d'}} F_{k+1}(\theta^{(2)} - \xi^{(2)})\{\Xi_{k+1,s}(\xi^{(2)}) - \Xi_{k,s}(\xi^{(2)})\}\, d\xi^{(2)}\Big|^{2r}. \tag{7.6}$$

The functions $F_k : \mathbb{T}^{d'} \to \mathbb{C}$ are defined in (5.3) and satisfy the bounds (5.6).

The proof of (7.5) is similar to the proof of (5.5). Indeed, if $\theta^{(2)}$ is close to a fraction with small denominator, in the sense of (5.7), then $|F_{k+1}(\theta^{(2)} - \xi^{(2)})| \lesssim \tau^{-2k/\delta}$ if $\xi^{(2)}$ is in the support of $\Xi_{k+1,s} - \Xi_{k,s}$, due to (5.6). The bounds (7.5) follow in this case. Otherwise, if $\theta^{(2)}$ does not satisfy (5.7), then there is $(l_1, l_2) \in Y_d'$ and an irreducible fraction $a_{l_1l_2}/q_{l_1l_2}$ such that

$$\Big|\theta_{l_1l_2}^{(2)} - \frac{a_{l_1l_2}}{q_{l_1l_2}}\Big| \leq \frac{1}{q_{l_1l_2}\tau^{k(l_1+l_2)-\delta^2 k}} \quad \text{and} \quad q_{l_1l_2} \in [\tau^{\delta^2 k}, \tau^{k(l_1+l_2)-\delta^2 k}] \cap \mathbb{Z}.$$

Using Proposition 2.3 with $P \simeq \tau^k$ we conclude that $\big|\Pi_{k+1}^{c,r}(\theta^{(1)},\theta^{(2)})\big| \lesssim \tau^{-2k/\delta}$. The desired bounds (7.5) follow in this case as well.

7.2 Proof of (7.4)

To prove the more difficult bounds (7.4) we will use a high order almost orthogonality argument. For this we need a good description of the operators $\{(\mathcal{H}_{k,s}^1)^*\mathcal{H}_{k,s}^1\}^r$ and $\{\mathcal{H}_{k,s}^1(\mathcal{H}_{k,s}^1)^*\}^r$, where $\mathcal{H}_{k,s}^1 f := f * H_{k,s}^1$ and, as before, $r \in \mathbb{Z}_+$ is a sufficiently large integer such that the bounds in Propositions 2.3 and 2.4 hold with $\varepsilon = \delta^4$. More precisely:

Lemma 7.1 *For any $k \geq \max((D/\ln\tau)^2, s/\delta)$ and $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\{(\mathcal{H}_{k,s}^1)^*\mathcal{H}_{k,s}^1\}^r f = f * \{B_k^r + E_k^r\}, \quad \{\mathcal{H}_{k,s}^1(\mathcal{H}_{k,s}^1)^*\}^r f = f * \{\widetilde{B}_k^r + \widetilde{E}_k^r\}, \tag{7.7}$$

where

$$B_k^r(h) := \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)} \Big\} \times \Big\{ \sum_{a/Q=(a^{(1)}/q_1,a^{(2)}/q_2)\in \mathcal{R}^d_{\le \delta k}\times \mathcal{R}^{d'}_s \cap [0,1)^{d+d'}} \mathfrak{e}(h.a/Q) G(a/Q) \Big\} \times \eta_{\le 3\delta k}(\tau^{-k}\circ h) \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \eta_{\le \delta k/2}(\zeta^{(1)})\eta_{\le \delta k/2}(\zeta^{(2)}) P'(\zeta)\mathfrak{e}[(\tau^{-k}\circ h).\zeta]\,d\zeta, \tag{7.8}$$

$$\widetilde{B}_k^r(h) := \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)} \Big\} \times \Big\{ \sum_{a/Q=(a^{(1)}/q_1,a^{(2)}/q_2)\in \mathcal{R}^d_{\le \delta k}\times \mathcal{R}^{d'}_s \cap [0,1)^{d+d'}} \mathfrak{e}(h.a/Q) \widetilde{G}(a/Q) \Big\} \times \eta_{\le 3\delta k}(\tau^{-k}\circ h) \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \eta_{\le \delta k/2}(\zeta^{(1)})\eta_{\le \delta k/2}(\zeta^{(2)}) \widetilde{P}'(\zeta)\mathfrak{e}[(\tau^{-k}\circ h).\zeta]\,d\zeta, \tag{7.9}$$

and

$$\|E_k^r\|_{\ell^1(\mathbb{G}_0)} + \|\widetilde{E}_k^r\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-k/4}. \tag{7.10}$$

Here $G(a/Q)$ and $\widetilde{G}(a/Q)$ are as in (2.29), $\chi'(x) = (1/\tau)\chi(x/\tau) - \chi(x)$, and

$$\begin{aligned} P'(\zeta) &:= \int_{\mathbb{R}^r\times\mathbb{R}^r} \Big\{ \prod_{1\le j\le r} \chi'(w_j)\chi'(y_j) \Big\} \mathfrak{e}[-\zeta.D(w,y)]\,dwdy, \\ \widetilde{P}'(\zeta) &:= \int_{\mathbb{R}^r\times\mathbb{R}^r} \Big\{ \prod_{1\le j\le r} \chi'(w_j)\chi'(y_j) \Big\} \mathfrak{e}[-\zeta.\widetilde{D}(w,y)]\,dwdy. \end{aligned} \tag{7.11}$$

For later use we also define the functions $P(\zeta)$ and $\widetilde{P}(\zeta)$ as in (7.11), using however the cutoff function $\chi(w_j)\chi(y_j)$ instead of $\chi'(w_j)\chi'(y_j)$. For $\iota \in \{0, 1\}$ we also let

$$P^\iota := \begin{cases} P & \text{if } \iota = 0, \\ P' & \text{if } \iota = 1, \end{cases} \quad \widetilde{P}^\iota := \begin{cases} \widetilde{P} & \text{if } \iota = 0, \\ \widetilde{P}' & \text{if } \iota = 1. \end{cases} \tag{7.12}$$

Using Proposition 2.4 we may estimate

$$\big|D_\zeta^\alpha P^\iota(\zeta)\big| + \big|D_\zeta^\alpha \widetilde{P}^\iota(\zeta)\big| \lesssim_{|\alpha|} \langle\zeta\rangle^{-1/\delta^2} \tag{7.13}$$

for any $\zeta \in \mathbb{R}^d \times \mathbb{R}^{d'}$, any multi-index $\alpha \in \mathbb{N}^{d+d'}$, and any $\iota \in \{0, 1\}$.

Proof of Lemma 7.1 We only prove in detail the claims for the operators $\{(\mathcal{H}^1_{k,s})^*\mathcal{H}^1_{k,s}\}^r$, since the claims for the operators $\{\mathcal{H}^1_{k,s}(\mathcal{H}^1_{k,s})^*\}^r$ follow by analogous arguments. In view of (2.13)–(2.16) we have

$$\{(\mathcal{H}^1_{k,s})^*\mathcal{H}^1_{k,s}\}^r = f * H^r_{k,s}$$

where

$$\begin{aligned} H^r_{k,s}(y) := \ & \eta_{\le 3\delta k}(\tau^{-k}\circ y)\int_{\mathbb{T}^d\times\mathbb{T}^{d'}} \\ & \mathfrak{e}\big(y.\theta\big)\Pi^{r,1}_k\big(\theta^{(1)},\theta^{(2)}\big)\Omega^{r,2}_{k,s}\big(\theta^{(2)}\big)\,d\theta^{(1)}d\theta^{(2)}. \end{aligned} \tag{7.14}$$

The multipliers $\Pi^{r,1}_k$ and $\Omega^{r,2}_{k,s}$ can be calculated as in the proof of Lemma 4.2. Namely,

$$\Pi^{r,1}_k\big(\theta\big) = \tau^{-2kr}\sum_{n,m\in\mathbb{Z}^r}\Big\{\prod_{1\le j\le r}\chi'(\tau^{-k}n_j)\chi'(\tau^{-k}m_j)\Big\}\mathfrak{e}\big(-\theta.D(n,m)\big), \tag{7.15}$$

and, with F_k defined as in (5.3), one has

$$\Omega^{r,2}_{k,s}(\theta^{(2)}) = \Big|\int_{\mathbb{T}^{d'}} F_k(\theta^{(2)}-\xi^{(2)})\,\Xi_{k,s}(\xi^{(2)})\,d\xi^{(2)}\Big|^{2r}. \tag{7.16}$$

We now show that the kernels $H^r_{k,s}$ are equivalent to the kernels B^r_k defined in (7.8) up to acceptable ℓ^1 errors satisfying (7.10). We accomplish this in several steps:

Step 1. We first replace the multiplier $\Omega^{r,2}_{k,s}(\theta^{(2)})$ with $\Xi_{k,s}(\theta^{(2)})$, at the expense of acceptable ℓ^1 errors. For this we show that

$$\begin{aligned} &\big|\Omega^{r,2}_{k,s}(\theta^{(2)}) - \Xi_{k,s}(\theta^{(2)})\big| \\ &\quad\lesssim \begin{cases} 1 & \text{if there is } a/q\in\mathcal{R}^{d'}_s \text{ such that } |\tau^k\circ(\theta^{(2)}-a/q)|\in[\tau^{\delta k/2},\tau^{2\delta k}],\\ \tau^{-k/\delta} & \text{otherwise.}\end{cases} \end{aligned} \tag{7.17}$$

Indeed, since the functions F_k satisfy the bounds (5.6), we have $\|F_k\|_{L^1(\mathbb{T}^{d'})}\lesssim 1$, so $\big|\Omega^{r,2}_{k,s}(\theta^{(2)})\big| + \big|\Xi_{k,s}(\theta^{(2)})\big| \lesssim 1$ for any $\theta^{(2)}\in\mathbb{T}^{d'}$. On the other hand, if $|\tau^k\circ(\theta^{(2)}-a/q)|\le\tau^{\delta k/2}$ for some $a/q\in\mathcal{R}^{d'}_s$ then $\Xi_{k,s}(\theta^{(2)})=1$ and, in fact, $\Xi_{k,s}(\xi^{(2)})=1$ for all $\xi^{(2)}\in\mathbb{T}^{d'}$ with $|\tau^k\circ(\theta^{(2)}-\xi^{(2)})|\le\tau^{\delta k/2}$.

Therefore, using (5.6) with M large enough and the definition (5.3) we have

$$\Big|\int_{\mathbb{T}^{d'}} F_k(\theta^{(2)}-\xi^{(2)})\,\Xi_{k,s}(\xi^{(2)})\,d\xi^{(2)}-1\Big| \lesssim \tau^{-k/\delta}+\Big|\int_{\mathbb{T}^{d'}} F_k(\theta^{(2)}-\xi^{(2)})\,d\xi^{(2)}-1\Big| = \tau^{-k/\delta}.$$

Thus $\big|\Omega^{r,2}_{k,s}(\theta^{(2)})-\Xi_{k,s}(\theta^{(2)})\big|\lesssim \tau^{-k/\delta}$, as claimed in (7.17).

Finally, if $|\tau^k\circ(\theta^{(2)}-a/q)|\gtrsim \tau^{2\delta k}$ for all $a/q\in\mathcal{R}^{d'}_s$ then $\Xi_{k,s}(\theta^{(2)})=0$ and, in fact, $\Xi_{k,s}(\xi^{(2)})=0$ for all $\xi^{(2)}\in\mathbb{T}^{d'}$ with $|\tau^k\circ(\theta^{(2)}-\xi^{(2)})|\le\tau^{\delta k/2}$. The desired bounds (7.17) follow as before in this case.

Given (7.17) we can define

$$H^{r,1}_{k,s}(y) := \eta_{\le 3\delta k}(\tau^{-k}\circ y)\int_{\mathbb{T}^d\times\mathbb{T}^{d'}} \mathfrak{e}\big(y.\theta\big)\Pi^{r,1}_k\big(\theta^{(1)},\theta^{(2)}\big) \times\Xi_{k,s}\big(\theta^{(2)}\big)\,d\theta^{(1)}d\theta^{(2)}, \tag{7.18}$$

and the difference $H^r_{k,s}-H^{r,1}_{k,s}$ is an acceptable ℓ^1 error.

Step 2. We now restrict to major arcs in the variable $\theta^{(1)}$, so we define

$$H^{r,2}_{k,s}(y) := \eta_{\le 3\delta k}(\tau^{-k}\circ y)\int_{\mathbb{T}^d\times\mathbb{T}^{d'}} \mathfrak{e}\big(y.\theta\big)\Pi^{r,1}_k\big(\theta^{(1)},\theta^{(2)}\big) \times\Psi_{k,\le\delta k}\big(\theta^{(1)}\big)\,\Xi_{k,s}\big(\theta^{(2)}\big)\,d\theta^{(1)}d\theta^{(2)}, \tag{7.19}$$

where

$$\Psi_{k,\le\delta k}\big(\theta^{(1)}\big) := \sum_{a/q\in\mathcal{R}^d_{\le\delta k}} \eta_{\le\delta k}(\tau^k\circ(\theta^{(1)}-a/q)). \tag{7.20}$$

We will show that $\|H^{r,1}_{k,s}-H^{r,2}_{k,s}\|_{\ell^1(\mathbb{G}_0)}\lesssim\tau^{-k}$. Indeed, if $\theta^{(1)}$ is in the support of $1-\Psi_{k,\le\delta k}$ then we apply Dirichlet's principle to find an irreducible fraction $(a_{l0}/q_{l0})_{l\in\{1,\dots,d\}}$ such that

$$\Big|\theta^{(1)}_{l0}-\frac{a_{l0}}{q_{l0}}\Big|\le\frac{1}{q_{l0}\tau^{kl-\delta^2k}} \quad\text{and}\quad q_{l0}\in[1,\tau^{kl-\delta^2k}]\cap\mathbb{Z},$$

and at least one of the denominators q_{l0} is larger than τ^{δ^2k}. But then we examine the definition (7.15) and apply Proposition 2.3 (i) to conclude that $\big|\Pi^{r,1}_k\big(\theta^{(1)},\theta^{(2)}\big)\big|\lesssim\tau^{-k/\delta}$. The desired error bounds follow.

Step 3. We now approximate the sum in the definition of $\Pi^{r,1}_k$. Assume that $\theta=\big(\theta^{(1)},\theta^{(2)}\big)$ is a point in $\mathbb{R}^{|Y_d|}$ and $a/Q\in\mathbb{Q}^{|Y_d|}$ is an irreducible fraction

such that

$$\big|\tau^k \circ \big(\theta - a/Q\big)\big| \le 2\tau^{\delta k+4}, \quad Q \le \tau^{2\delta k+2}. \tag{7.21}$$

We examine the sum in the formula (7.15). For any $j \in \{1, \dots, r\}$ we decompose $n_j = Qw_j + x_j$, $m_j = Qy_j + z_j$, $x_j, z_j \in \{0, \dots, Q-1\}$, $w_j, y_j \in \mathbb{Z}$. Letting $\beta = \theta - a/Q$ we notice that

$$\mathfrak{e}\big(-\theta.D(n,m)\big) = \mathfrak{e}\big(-\beta.D(Qw+x, Qy+z)\big)\mathfrak{e}\big(-(a/Q).D(x,z)\big).$$

Moreover, if $\big|\tau^k \circ \beta\big| \lesssim \tau^{\delta k}$ and $|Qw| + |Qy| \lesssim \tau^k$ then

$$\begin{aligned}\beta.D(Qw+x, Qy+z) &= \beta.D(Qw, Qy) + O(Q\tau^{-k+\delta k}) \\ &= (Q \circ \beta).D(w,y) + O(Q\tau^{-k+\delta k}),\end{aligned}$$

as one can see easily from the formula (2.24). In addition

$$\prod_{1\le j\le r} \chi'(\tau^{-k}n_j)\chi'(\tau^{-k}m_j) = \prod_{1\le j\le r} \chi'(\tau^{-k}Qw_j)\chi'(\tau^{-k}Qy_j) + O(Q\tau^{-k}).$$

Therefore

$$\begin{aligned}\Pi_k^{r,1}(\theta) = \tau^{-2kr}\Big\{ &\sum_{|w|,|y|\lesssim \tau^k/Q} \Big\{ \prod_{1\le j\le r} \chi'(\tau^{-k}Qw_j)\chi'(\tau^{-k}Qy_j)\Big\} \\ &\times \mathfrak{e}\big[-(Q\circ\beta).D(w,y)\big]\Big\} \\ &\times \Big\{ \sum_{x,z\in\mathbb{Z}^r_Q} \mathfrak{e}\big(-(a/Q).D(x,z)\big)\Big\} + O(Q\tau^{-k+\delta k}).\end{aligned} \tag{7.22}$$

Recall the definition (2.29). Using the Poisson summation formula we may replace the sum over $w, y \in \mathbb{Z}^r$ with the corresponding integral, at the expense of $O(\tau^{-2k})$ errors, and then change variables to reach the formula (7.11). Therefore

$$\Pi_k^{r,1}(\theta) = P'(\tau^k \circ \beta)G(a/Q) + O(\tau^{-k+8\delta k}). \tag{7.23}$$

The contribution of the error term can be incorporated into the kernel E_k^r, while the main term can be substituted into the formula (7.19), leading to the desired formula (7.8) after changes of variables. We have established (7.8) and (7.9) with $\eta_{\le\delta k}(\zeta^{(1)})\eta_{\le\delta k}(\zeta^{(2)})$ in place of $\eta_{\le\delta k/2}(\zeta^{(1)})\eta_{\le\delta k/2}(\zeta^{(2)})$. Finally we can use (7.13) to replace cutoff functions $\eta_{\le\delta k}(\zeta^{(1)})\eta_{\le\delta k}(\zeta^{(2)})$ with $\eta_{\le\delta k/2}(\zeta^{(1)})\eta_{\le\delta k/2}(\zeta^{(2)})$. This completes the proof of the lemma. □

We return now to the proof of the main bounds (7.4). In view of the Cotlar–Stein lemma it suffices to prove the following:

Lemma 7.2 *If* $k, j \geq \max((D/\ln\tau)^2, s/\delta)$ *and* $j \in [k/2, k]$ *then*

$$\begin{aligned}\|\mathcal{H}^1_{j,s}(\mathcal{H}^1_{k,s})^*\|_{\ell^2(\mathbb{G}_0)\to\ell^2(\mathbb{G}_0)} &+\|(\mathcal{H}^1_{j,s})^*\mathcal{H}^1_{k,s}\|_{\ell^2(\mathbb{G}_0)\to\ell^2(\mathbb{G}_0)}\\ &\lesssim \tau^{-2s/D}\tau^{-2|j-k|/D}.\end{aligned} \tag{7.24}$$

Proof **Step 1.** We prove these bounds first when $j = k$, so we prove that the operators $\mathcal{H}^1_{k,s}$ are suitably bounded on $\ell^2(\mathbb{G}_0)$. In view of Lemma 7.1, it suffices to prove that

$$\|B^r_k\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-2rs/D}. \tag{7.25}$$

We notice that

$$\Big|\sum_{a/Q=(a^{(1)}/q_1,a^{(2)}/q_2)\in\mathcal{R}^d_{\leq\delta k}\times\mathcal{R}^{d'}_s\cap[0,1)^{d+d'}} \mathfrak{e}(h.a/Q)G(a/Q)\Big| \lesssim \tau^{-s}$$

for any $h \in \mathbb{G}_0$, as a consequence of Proposition 2.3 (ii). For $\iota \in \{0, 1\}$ we let

$$\begin{aligned}X^{\iota,r}_k(h) :=&\Big\{\prod_{(l_1,l_2)\in Y_d}\tau^{-k(l_1+l_2)}\Big\}\eta_{\leq 3\delta k}(\tau^{-k}\circ h)\\ &\times\int_{\mathbb{R}^d\times\mathbb{R}^{d'}}\eta_{\leq\delta k/2}(\zeta^{(1)})\eta_{\leq\delta k/2}(\zeta^{(2)})P^\iota(\zeta)\mathfrak{e}[(\tau^{-k}\circ h).\zeta]\,d\zeta.\end{aligned} \tag{7.26}$$

Notice that

$$\|X^{\iota,r}_k\|_{\ell^1(\mathbb{G}_0)} \lesssim 1 \quad \text{for any } k \in \mathbb{N}. \tag{7.27}$$

Indeed, invoking (7.13) and integrating by parts in (7.26) we conclude that

$$|X^{\iota,r}_k(h)| \lesssim \Big\{\prod_{(l_1,l_2)\in Y_d}\tau^{-k(l_1+l_2)}\Big\}(1+|\tau^{-k}\circ h|)^{-1/\delta}$$

for any $h \in \mathbb{G}_0$. Now we see that inequality (7.25) follows from (7.27) with $\iota = 1$.

Step 2. Since we have already proved that $\|(\mathcal{H}^1_{j,s})^*\|_{\ell^2\to\ell^2} \lesssim \tau^{-s/D} \lesssim 1$, we can estimate

$$\begin{aligned}\|\mathcal{H}^1_{j,s}(\mathcal{H}^1_{k,s})^*\|_{\ell^2\to\ell^2} &= \|\mathcal{H}^1_{j,s}(\mathcal{H}^1_{k,s})^*\mathcal{H}^1_{k,s}(\mathcal{H}^1_{j,s})^*\|^{1/2}_{\ell^2\to\ell^2}\\ &\lesssim \|\mathcal{H}^1_{j,s}[(\mathcal{H}^1_{k,s})^*\mathcal{H}^1_{k,s}]\|^{1/2}_{\ell^2\to\ell^2}\\ &\lesssim \|\mathcal{H}^1_{j,s}[(\mathcal{H}^1_{k,s})^*\mathcal{H}^1_{k,s}]^2\|^{1/4}_{\ell^2\to\ell^2}\\ &\lesssim \ldots \lesssim \|\mathcal{H}^1_{j,s}[(\mathcal{H}^1_{k,s})^*\mathcal{H}^1_{k,s}]^{2^a}\|^{1/2^{a+1}}_{\ell^2\to\ell^2},\end{aligned} \tag{7.28}$$

for any $j \le k$, where 2^a is the smallest dyadic number $\ge r$. The norm $\|(\mathcal{H}^1_{j,s})^*\mathcal{H}^1_{k,s}\|_{\ell^2\to\ell^2}$ can be estimated in the same way, so it suffices to prove that for any $j \in [k/2, k]$ such that $k, j \ge \max((D/\ln\tau)^2, s/\delta)$ we have

$$\begin{aligned}&\|\mathcal{H}^1_{j,s}[(\mathcal{H}^1_{k,s})^*\mathcal{H}^1_{k,s}]^r\|_{\ell^2\to\ell^2} + \|(\mathcal{H}^1_{j,s})^*[\mathcal{H}^1_{k,s}(\mathcal{H}^1_{k,s})^*]^r\|_{\ell^2\to\ell^2}\\ &\lesssim \tau^{-8rs/D}\tau^{-8r|j-k|/D}.\end{aligned} \tag{7.29}$$

The bounds on the two terms in the left-hand side of (7.29) are similar, and we only provide the proof for the first term. We use Lemma 7.1. The contribution of the error kernel E^r_k is bounded by $C\tau^{-k/4}$, due to (7.10), which is better than needed. It remains to prove that

$$\big\|B^r_k * H^1_{j,s}\big\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-8rs/D}\tau^{-8r|j-k|/D}. \tag{7.30}$$

We examine the formula (7.8) and decompose the kernel B^r_k

$$\begin{aligned}B^r_k &= \sum_{a/Q=(a^{(1)}/q_1,a^{(2)}/q_2)\in\mathcal{R}^d_{\le\delta k}\times\mathcal{R}^{d'}_s\cap[0,1)^{d+d'}} G(a/Q)X^r_{k,a/Q},\\ X^r_{k,a/Q}(h) &:= X^r_k(h)\mathfrak{e}(h.a/Q),\end{aligned} \tag{7.31}$$

where the kernels $X^r_k := X^{1,r}_k$ have been defined in (7.26). In view of the rapid decay of the coefficients $G(a/Q)$ (see (2.30)), for (7.30) it suffices to prove that

$$\big\|X^r_{k,a/Q} * H^1_{j,s}\big\|_{\ell^1(\mathbb{G}_0)} \lesssim Q^{8/\delta}\tau^{-8r|j-k|/D} \tag{7.32}$$

for any irreducible fraction $a/Q \in \mathbb{Q}^{|Y_d|}$ with denominator $Q \in [\tau^s, \tau^{2\delta k+2}]$.

We examine now the definition (7.2) and decompose

$$\begin{aligned} H^1_{j,s}(g) &= \sum_{b^{(2)}/q_2 \in \mathcal{R}_s^{d'} \cap [0,1)^{d'}} H_j^{1,b^{(2)}/q_2}(g) \\ &= \sum_{b^{(2)}/q_2 \in \mathcal{R}_s^{d'} \cap [0,1)^{d'}} [\Delta_j L_j](g^{(1)}) \mathfrak{e}(g^{(2)}.b^{(2)}/q_2) Y_j(g^{(2)}), \qquad (7.33) \\ Y_j(g^{(2)}) &:= \phi_j^{(2)}(g^{(2)}) \int_{\mathbb{R}^{d'}} \mathfrak{e}(g^{(2)}.\beta^{(2)}) \eta_{\le \delta j}(\tau^j \circ \beta^{(2)})\, d\beta^{(2)}, \end{aligned}$$

For (7.32) it suffices to prove that

$$\big\| X^r_{k,a/Q} * H_j^{1,b^{(2)}/q_2} \big\|_{\ell^1(\mathbb{G}_0)} \lesssim Q^{4/\delta} \tau^{-8r|j-k|/D} \qquad (7.34)$$

for any $b^{(2)}/q_2 \in \mathcal{R}_s^{d'}$, as the sum over $b^{(2)}/q_2$ contains at most $\tau^{s/\delta}$ terms and $Q \ge \tau^s$.

Step 3. Using the definitions we estimate

$$\begin{aligned} &\big\| X^r_{k,a/Q} * H_j^{1,b^{(2)}/q_2} \big\|_{\ell^1(\mathbb{G}_0)} \\ &= \sum_{h=(h^{(1)},h^{(2)}) \in \mathbb{G}_0} \Big| \sum_{g=(g^{(1)},g^{(2)}) \in \mathbb{G}_0} H_j^{1,b^{(2)}/q_2}(g) X^r_{k,a/Q}(g^{-1} \cdot h) \Big| \\ &\le \sum_{h=(h^{(1)},h^{(2)}) \in \mathbb{G}_0,\, g^{(2)} \in \mathbb{Z}^{d'}} |Y_j(g^{(2)})| \Big| \sum_{g^{(1)} \in \mathbb{Z}^d} [\Delta_j L_j](g^{(1)}) X^r_k(g^{-1} \cdot h) \\ &\quad \mathfrak{e}\big[(g^{-1} \cdot h).(a/Q)\big] \Big|. \end{aligned} \qquad (7.35)$$

To get decay in $|k-j|$ the main point is to bound efficiently the sum over $g^{(1)}$ in the expression above, using the cancellation of the kernel $\Delta_j L_j$. We rewrite this sum in the form

$$\begin{aligned} &\Big| \sum_{n \in \mathbb{Z}} \tau^{-j} \chi'(\tau^{-j} n) X^r_k\big((A_0^{(1)}(n), g^{(2)})^{-1} \cdot h\big) \\ &\qquad \times \mathfrak{e}\big[- A_0^{(1)}(n).(a^{(1)}/Q) + R_0(A_0^{(1)}(n), A_0^{(1)}(n) - h^{(1)}).(a^{(2)}/Q) \big] \Big|, \end{aligned}$$

for any $h = (h^{(1)}, h^{(2)}) \in \mathbb{G}_0$ and $g^{(2)} \in \mathbb{Z}^{d'}$, where $\chi'(x) = (1/\tau)\chi(x/\tau) - \chi(x)$ as before. It follows easily from the definition (7.33) that $\|Y_j\|_{\ell^1(\mathbb{Z}^{d'})} \lesssim 1$ uniformly in $j \in \mathbb{Z}$. Therefore, for (7.34) it suffices to prove that

$$\sum_{h=(h^{(1)},h^{(2)})\in\mathbb{G}_0}\Big|\sum_{n\in\mathbb{Z}}\tau^{-j}\chi'(\tau^{-j}n)X_k^r\big(h^{(1)}-A_0^{(1)}(n),\\ h^{(2)}+R_0(A_0^{(1)}(n),A_0^{(1)}(n)-h^{(1)})\big)\\ \times\mathfrak{e}\big[-A_0^{(1)}(n).(a^{(1)}/Q)+R_0(A_0^{(1)}(n),\\ A_0^{(1)}(n)-h^{(1)}).(a^{(2)}/Q)\big]\Big|\\ \lesssim Q^{4/\delta}\tau^{-8r|j-k|/D}. \tag{7.36}$$

We further decompose $n=mQ+\rho$, $m\in\mathbb{Z}$, $\rho\in[0,Q-1]\cap\mathbb{Z}$, and notice that the oscillatory factor in the sum above does not depend on m. For (7.36) it suffices to prove that

$$\sum_{h\in\mathbb{G}_0}\Big|\sum_{m\in\mathbb{Z}}\tau^{-j}\chi'(\tau^{-j}(mQ+\rho))X_k^r\big(A_0(mQ+\rho)^{-1}\cdot h\big)\Big|\lesssim\tau^{-8r|j-k|/D} \tag{7.37}$$

for any $Q\in[1,\tau^{2\delta k+2}]$ and $\rho\in[0,Q-1]\cap\mathbb{Z}$, as the sum over ρ contains only Q terms.

Finally, we examine the kernels X_k^r. These kernels can be extended to the continuous Lie group $\mathbb{G}_0^{\#}\simeq\mathbb{R}^{|Y_d|}$, according to the defining formula (7.26). Using (7.13) and integration by parts it follows that

$$|X_k^r(h)|+\sum_{(l_1,l_2)\in Y_d}\tau^{k(l_1+l_2)}\big|(\partial_{h_{l_1l_2}}X_k^r)(h)\big|\\ \lesssim\Big\{\prod_{(l_1,l_2)\in Y_d}\tau^{-k(l_1+l_2)}\Big\}(1+|\tau^{-k}\circ h|)^{-2/\delta} \tag{7.38}$$

for any $h\in\mathbb{R}^{|Y_d|}$. Therefore, for any $g\in\mathbb{G}_0^{\#}$ with $|\tau^{-j}\circ g|\lesssim 1$, we have

$$|X_k^r(h)-X_k^r(g^{-1}\cdot h)|\lesssim\tau^{j-k}\Big\{\prod_{(l_1,l_2)\in Y_d}\tau^{-k(l_1+l_2)}\Big\}(1+|\tau^{-k}\circ h|)^{-1/\delta}. \tag{7.39}$$

Therefore

$$\sum_{h\in\mathbb{G}_0}\Big|\sum_{m\in\mathbb{Z}}\tau^{-j}\chi'(\tau^{-j}(mQ+\rho))\big[X_k^r\big(A_0(mQ+\rho)^{-1}\cdot h\big)-X_k^r(h)\big]\Big|\lesssim\tau^{j-k}.$$

Moreover, since $\int_{\mathbb{R}} \chi'(x)\,dx = 0$, we have

$$\sum_{h\in\mathbb{G}_0}\Big|\sum_{m\in\mathbb{Z}}\tau^{-j}\chi'(\tau^{-j}(mQ+\rho))X_k^r(h)\Big| = \Big(\sum_{h\in\mathbb{G}_0}\big|X_k^r(h)\big|\Big)\Big|\sum_{m\in\mathbb{Z}}\tau^{-j}\chi'(\tau^{-j}(mQ+\rho))\Big| \lesssim Q\tau^{-j}.$$

The desired bounds (7.37) follow since $j \in [k/2, k]$ and $Q \le \tau^{2\delta k+2}$. This completes the proof of the lemma. □

7.3 Proof of (4.20)

Given that we already proved the variational inequality (4.19), in view of (2.4) it suffices to prove that

$$\|f * K_{k_0,s}\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{7.40}$$

where k_0 in an integer satisfying $|k_0 - 3\kappa_s/2| \le 1$. We decompose $K_{k_0,s} = G^{\mathrm{low}}_{k_0,s} + \sum_{t\le\delta' k_0} G_{k_0,s,t} + G^c_{k_0,s}$ as in (4.28). The contributions of the operators defined by the kernels $G^{\mathrm{low}}_{k_0,s}$ and $G^c_{k_0,s}$ are suitably bounded due to Lemma 4.4 and Lemma 4.5 (i) proved in the previous sections. The contributions of the operators defined by the kernels $G_{k_0,s,t}$ are bounded due to Lemma 4.5 (ii) and Lemma 4.6 proved in Sect. 8 below. The bounds (7.40) follow.

8 Transition estimates II: Proof of Lemma 4.6

In this section we prove bounds (4.34) and (4.35). In fact we establish a stronger result which will be used in $\ell^p(\mathbb{G}_0)$ theory in Sect. 9.

Lemma 8.1 *Assume that $s \ge 0$, and $t \ge D(s+1)$, and let $\mathcal{A} \subseteq \mathcal{R}^d_t \setminus \widetilde{\mathcal{R}}^d_{Q_s}$, $\mathcal{B} \subseteq \mathcal{R}^{d'}_{\le s}$ be 1-periodic sets of rationals. Then for any $2 < \rho < \infty$ and for any $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\big\|V^\rho(f * K_{k,k,\mathcal{A},\mathcal{B}} : \max(\kappa_s, t/\delta') \le k < 2\kappa_t)\big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{8.1}$$

where $K_{k,w,\mathcal{A},\mathcal{B}}$ is the kernel defined in (4.36). In particular, we have

$$\Big\|\sup_{\max(\kappa_s,t/\delta')\le k<2\kappa_t}|f * K_{k,k,\mathcal{A},\mathcal{B}}|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{8.2}$$

The estimates (8.1)–(8.2) imply (4.34)–(4.35), since $G_{k,s,t} = K_{k,k,\mathcal{R}_t^d \setminus \widetilde{\mathcal{R}}_{Q_s}^d, \mathcal{R}_s^{d'}}$. Moreover, the bounds (8.2) follow from (4.35) and (6.49). Thus our main goal is to prove the bounds (8.1).

As in Sect. 7 we let $G_{k,k,\mathcal{A},\mathcal{B}} := \Delta_k K_{k,k,\mathcal{A},\mathcal{B}} = K_{k+1,k+1,\mathcal{A},\mathcal{B}} - K_{k,k,\mathcal{A},\mathcal{B}}$ for $k \geq \max(\kappa_s, t/\delta')$, apply the Rademacher–Menshov inequality (2.7) and then Khintchine's inequality. As in Sect. 7, for (8.1) it suffices to prove that

$$\Big\| \sum_{k\in[J,2J]} \varkappa_k (f * G_{k,k,\mathcal{A},\mathcal{B}}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-4t/D^2} \|f\|_{\ell^2(\mathbb{G}_0)} \tag{8.3}$$

for any $J \geq \max(\kappa_s, t/\delta')$ and any coefficients $\varkappa_k \in [-1, 1]$.

We examine the definitions (4.36) and (4.37), and further decompose

$$\begin{aligned}
G_{k,k,\mathcal{A},\mathcal{B}} &= G^1_{k,k,\mathcal{A},\mathcal{B}} + G^2_{k,k,\mathcal{A},\mathcal{B}} + G^3_{k,k,\mathcal{A},\mathcal{B}} + G^4_{k,k,\mathcal{A},\mathcal{B}},\\
G^1_{k,k,\mathcal{A},\mathcal{B}}(g) &:= \phi_k(g) \int_{\mathbb{T}^d\times\mathbb{T}^{d'}} \mathfrak{e}(g.\xi)\Psi_{k,k,\mathcal{A}}(\xi^{(1)})\\
&\qquad \times \Xi_{k,k,\mathcal{B}}(\xi^{(2)})[\Delta_k S_k](\xi^{(1)})\, d\xi^{(1)} d\xi^{(2)},\\
G^2_{k,k,\mathcal{A},\mathcal{B}}(g) &:= [\Delta_k\phi_k](g) \int_{\mathbb{T}^d\times\mathbb{T}^{d'}} \mathfrak{e}(g.\xi)\Psi_{k,k,\mathcal{A}}(\xi^{(1)})\\
&\qquad \times \Xi_{k,k,\mathcal{B}}(\xi^{(2)}) S_{k+1}(\xi^{(1)})\, d\xi^{(1)} d\xi^{(2)},\\
G^3_{k,k,\mathcal{A},\mathcal{B}}(g) &:= \phi_{k+1}(g) \int_{\mathbb{T}^d\times\mathbb{T}^{d'}} \mathfrak{e}(g.\xi)\Psi_{k,k,\mathcal{A}}(\xi^{(1)})\\
&\qquad \times [\Delta_k \Xi_{k,k,\mathcal{B}}](\xi^{(2)})] S_{k+1}(\xi^{(1)})\, d\xi^{(1)} d\xi^{(2)},\\
G^4_{k,k,\mathcal{A},\mathcal{B}}(g) &:= \phi_{k+1}(g) \int_{\mathbb{T}^d\times\mathbb{T}^{d'}} \mathfrak{e}(g.\xi)[\Delta_k\Psi_{k,k,\mathcal{A}}](\xi^{(1)})\\
&\qquad \times \Xi_{k+1,k+1,\mathcal{B}}(\xi^{(2)}) S_{k+1}(\xi^{(1)})\, d\xi^{(1)} d\xi^{(2)},
\end{aligned} \tag{8.4}$$

where $\phi_k(g) = \phi_k^{(1)}(g^{(1)})\phi_k^{(2)}(g^{(2)})$ as before. As in Sect. 7 we will prove that

$$\Big\| \sum_{k\in[J,2J]} \varkappa_k (f * G^1_{k,k,\mathcal{A},\mathcal{B}}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/D} \|f\|_{\ell^2(\mathbb{G}_0)}, \tag{8.5}$$

for any $J \geq \max(\kappa_s, t/\delta')$ and any coefficients $\varkappa_k \in [-1, 1]$. We will also prove that

$$\|f * G^\iota_{k,k,\mathcal{A},\mathcal{B}}\|_{\ell^2(\mathbb{G}_0)} \lesssim \tau^{-k/D} \|f\|_{\ell^2(\mathbb{G}_0)}, \tag{8.6}$$

for any $k \geq \max(\kappa_s, t/\delta')$ and $\iota \in \{2, 3, 4\}$. These two estimates would clearly imply the bounds (8.3), thus completing the proof of Lemma 8.1.

8.1 Proof of the bounds (8.5)

As in Sect. 7, we will use a high order almost orthogonality argument. For this purpose we need a good description of the operators $\{(\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}})^*\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}}\}^r$ and $\{\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}}(\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}})^*\}^r$, where $\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}}f := f * G^1_{k,k,\mathcal{A},\mathcal{B}}$. We note that $G^1_{k,k,\mathcal{A},\mathcal{B}} = K'_{k,k,\mathcal{A},\mathcal{B}}$, see the definitions in (4.36) and (4.37). For $\iota \in \{0,1\}$ let

$$K^\iota_{k,w,\mathcal{A},\mathcal{B}} := \begin{cases} K_{k,w,\mathcal{A},\mathcal{B}} & \text{if } \iota = 0, \\ K'_{k,w,\mathcal{A},\mathcal{B}} & \text{if } \iota = 1, \end{cases} \quad L^\iota_{k,w,\mathcal{A}} := \begin{cases} L_{k,w,\mathcal{A}} & \text{if } \iota = 0, \\ L'_{k,w,\mathcal{A}} & \text{if } \iota = 1. \end{cases} \tag{8.7}$$

For later use we consider both operators $\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}}f := f * K^\iota_{k,k,\mathcal{A},\mathcal{B}}, \iota \in \{0,1\}$.

Lemma 8.2 *Assume that $\mathcal{A} \subseteq \mathbb{Q}^d$, $\mathcal{B} \subseteq \mathbb{Q}^{d'}$ are 1-periodic sets and assume that $\{q \in \mathbb{Z}_+ : a/q \in \mathcal{A}$ and $\gcd(a_1,\dots,a_d,q) = 1\} \subseteq [q_\mathcal{A}, 4q_\mathcal{A}]$ for some $q_\mathcal{A} \in \mathbb{Z}_+$. Assume that $q_\mathcal{A} \geq Q^D$ for any irreducible fraction $a/Q \in \mathcal{B}$, and $k \geq (D/\ln\tau)^2$ satisfies $\tau^{\delta' k} \geq q_\mathcal{A}$. If $r \in \mathbb{Z}_+$ is sufficiently large then for every $f \in \ell^2(\mathbb{G}_0)$ we have*

$$\begin{aligned} \{(\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}})^*\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}}\}^r f &= f * \{F^{\iota,r}_k + O^{\iota,r}_k\}, \\ \{\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}}(\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}})^*\}^r f &= f * \{\widetilde{F}^{\iota,r}_k + \widetilde{O}^{\iota,r}_k\}, \end{aligned} \tag{8.8}$$

where

$$\begin{aligned} F^{\iota,r}_k(h) := \Big\{ &\sum_{a^{(2)}/Q \in \mathcal{B}\cap[0,1)^{d'}} \sum_{\sigma \in [\mathcal{A}+(\mathbb{Z}/Q)^d]\cap[0,1)^d} \mathcal{C}(a^{(2)}/Q,\sigma) \\ &\times \mathfrak{e}(h^{(1)}.\sigma)\mathfrak{e}\big(h^{(2)}.(a^{(2)}/Q)\big)\Big\} \\ &\times \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)} \Big\} \eta_{\leq 3\delta k}(\tau^{-k}\circ h) \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \Big\{ \prod_{i=1}^2 \eta_{\leq \delta k/2}(\zeta^{(i)}) \Big\} \\ &\times P^\iota(\zeta)\mathfrak{e}[(\tau^{-k}\circ h).\zeta]\,d\zeta, \end{aligned} \tag{8.9}$$

$$\begin{aligned} \widetilde{F}^{\iota,r}_k(h) := \Big\{ &\sum_{a^{(2)}/Q \in \mathcal{B}\cap[0,1)^{d'}} \sum_{\sigma \in [\mathcal{A}+(\mathbb{Z}/Q)^d]\cap[0,1)^d} \widetilde{\mathcal{C}}(a^{(2)}/Q,\sigma) \\ &\times \mathfrak{e}(h^{(1)}.\sigma)\mathfrak{e}\big(h^{(2)}.(a^{(2)}/Q)\big)\Big\} \\ &\times \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)} \Big\} \eta_{\leq 3\delta k}(\tau^{-k}\circ h) \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \Big\{ \prod_{i=1}^2 \eta_{\leq \delta k/2}(\zeta^{(i)}) \Big\} \\ &\times \widetilde{P}^\iota(\zeta)\mathfrak{e}[(\tau^{-k}\circ h).\zeta]\,d\zeta, \end{aligned} \tag{8.10}$$

and

$$\|O^{\iota,r}_k\|_{\ell^1(\mathbb{G}_0)} + \|\widetilde{O}^{\iota,r}_k\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-k}. \tag{8.11}$$

The functions P^ι and $\widetilde{P}^\iota$ are as in (7.12) (see also (7.11)), and the coefficients $\mathcal{C}$ and $\widetilde{\mathcal{C}}$ satisfy

$$|\mathcal{C}(a^{(2)}/Q,\sigma)| + |\widetilde{\mathcal{C}}(a^{(2)}/Q,\sigma)| \lesssim q_{\mathcal{A}}^{-1/\delta} \tag{8.12}$$

for any $a^{(2)}/Q \in \mathcal{B} \cap [0,1)^{d'}$ and $\sigma \in [\mathcal{A} + (\mathbb{Z}/Q)^d] \cap [0,1)^d$.

Proof We only prove in detail the claims for the operators $\{(\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}})^* \mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}}\}^r$, since the claims for the operators $\{\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}}(\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}})^*\}^r$, follow by analogous arguments.

Step 1. By (4.36) notice that the kernels $K^\iota_{k,k,\mathcal{A},\mathcal{B}}$ have product structure. Thus in view of (2.13)–(2.16) we have

$$\{(\mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}})^* \mathcal{K}^\iota_{k,k,\mathcal{A},\mathcal{B}}\}^r f = f * K^{\iota,r}_{k,k,\mathcal{A},\mathcal{B}},$$

where

$$\begin{aligned} K^{\iota,r}_{k,k,\mathcal{A},\mathcal{B}}(y) := \ & \eta_{\le 3\delta k}(\tau^{-k} \circ y) \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \mathfrak{e}\big(y.\theta\big) \Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big) \\ & \times \Omega^r_{k,k,\mathcal{B}}\big(\theta^{(2)}\big)\, d\theta^{(1)} d\theta^{(2)}, \end{aligned} \tag{8.13}$$

and the multipliers $\Pi^{\iota,r}_{k,k,\mathcal{A}}$ and $\Omega^r_{k,k,\mathcal{B}}$ are given by

$$\begin{aligned} \Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big) := & \sum_{h^{(1)}_j, g^{(1)}_j \in \mathbb{Z}^d} \Big\{ \prod_{j=1}^r \overline{L^\iota_{k,k,\mathcal{A}}(h^{(1)}_j)} L^\iota_{k,k,\mathcal{A}}(g^{(1)}_j) \Big\} \\ & \mathfrak{e}\big(\theta^{(1)}.\sum_{1\le j\le r}(h^{(1)}_j - g^{(1)}_j)\big) \\ & \times \mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le j\le r} R_0(h^{(1)}_j, h^{(1)}_j - g^{(1)}_j) \\ & + \sum_{1\le l<j\le r} R_0(-h^{(1)}_l + g^{(1)}_l, -h^{(1)}_j + g^{(1)}_j)\Big\}\Big) \end{aligned} \tag{8.14}$$

and, with F_k defined as in (5.3),

$$\Omega^r_{k,k,\mathcal{B}}\big(\theta^{(2)}\big) := \Big| \int_{\mathbb{T}^{d'}} F_k(\theta^{(2)} - \xi^{(2)})\, \Xi_{k,k,\mathcal{B}}(\xi^{(2)})\, d\xi^{(2)} \Big|^{2r}. \tag{8.15}$$

As in the proof of Lemma 7.1, our goal is to show that the kernels $K^{\iota,r}_{k,k,\mathcal{A},\mathcal{B}}$ are equivalent to the kernels $F^{\iota,r}_k$ in (8.9), up to acceptable ℓ^1 errors. For this

we need to replace the multipliers $\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big)\Omega^{r}_{k,k,\mathcal{B}}\big(\theta^{(2)}\big)$ with more explicit multipliers, at the expense of acceptable errors.

Step 2. We will follow the ides from Sects. 5–7. As in (5.15) we may write

$$\begin{aligned}\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big) &= \int_{(\mathbb{T}^d)^{2r}} \mathcal{V}^r_k(\theta^{(1)},\theta^{(2)};\zeta^{(1)}_1,\xi^{(1)}_1,\dots,\zeta^{(1)}_r,\xi^{(1)}_r)\\ &\times \prod_{1\le j\le r}\Big\{\overline{S^{\iota}_k(\zeta^{(1)}_j)}\,\overline{\Psi_{k,k,\mathcal{A}}(\zeta^{(1)}_j)}S^{\iota}_k(\xi^{(1)}_j)\Psi_{k,k,\mathcal{A}}(\xi^{(1)}_j)\Big\}\\ &\quad d\xi^{(1)}_1 d\zeta^{(1)}_1\dots d\xi^{(1)}_r d\zeta^{(1)}_r, \end{aligned}\tag{8.16}$$

where (see also in (5.16)) we have

$$\begin{aligned}&\mathcal{V}^r_k(\theta^{(1)},\theta^{(2)};\zeta^{(1)}_1,\xi^{(1)}_1,\dots,\zeta^{(1)}_r,\xi^{(1)}_r)\\ &\quad=\sum_{h_j,g_j\in\mathbb{Z}^d}\prod_{1\le j\le r}\Big\{\overline{\phi^{(1)}_k(h_j)}\mathfrak{e}\big((\theta^{(1)}-\zeta^{(1)}_j).h_j\big)\phi^{(1)}_k(g_j)\mathfrak{e}\big(-(\theta^{(1)}-\xi^{(1)}_j).g_j\big)\Big\}\\ &\qquad\times\mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le j\le r}R_0(h_j,h_j-g_j)+\sum_{1\le l<j\le r}R_0(-h_l+g_l,-h_j+g_j)\Big\}\Big).\end{aligned}$$

In view of (5.6) we have a rapid decay $|\Omega^r_{k,k,\mathcal{B}}(\theta^{(2)})|\lesssim\tau^{-Dk}$ unless $|\tau^k\circ(\theta^{(2)}-a^{(2)}/Q)|\le\tau^{2\delta k}$ for some $a^{(2)}/Q\in\mathcal{B}$. Hence, we may assume that $\theta^{(2)}=\alpha^{(2)}+a^{(2)}/Q$ for some $a^{(2)}/Q\in\mathcal{B}$ and $|\tau^k\circ\alpha^{(2)}|\le\tau^{2\delta k}$. The condition (5.21) is then verified so we can use Lemma 5.2.

We now define new projections

$$\Phi_{k,\mathcal{A}+(\mathbb{Z}/Q)^d}(\theta^{(1)}):=\sum_{\sigma\in\mathcal{A}+(\mathbb{Z}/Q)^d}\eta_{\le 2\delta' k}(\tau^k\circ(\theta^{(1)}-\sigma)),$$

where $\mathcal{A}+(\mathbb{Z}/Q)^d:=\{\sigma+a/Q:\sigma\in\mathcal{A},a\in\mathbb{Z}^d\}$. Examining (5.18) we conclude that $\mathcal{V}^r_k$ decays rapidly unless $\tau^{kl}\|\theta^{(1)}_l-\xi^{(1)}_{j,l}\|_Q\le\tau^{10\delta k}$ and $\tau^{kl}\|\theta^{(1)}_l-\zeta^{(1)}_{j,l}\|_Q\le\tau^{10\delta k}$ for all $j\in\{1,\dots,r\}$ and $l\in\{1,\dots,d\}$, thus we may replace $\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big)$ with $\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big)\Phi_{k,\mathcal{A}+(\mathbb{Z}/Q)^d}\big(\theta^{(1)}\big)$ at the expense of $O(\tau^{-Dk})$ error term.

Expanding the cutoff functions $\Psi_{k,k,\mathcal{A}}$, invoking rapid decay from (5.18) of $\mathcal{V}^r_k$ as above and using Lemma 5.2 we may replace $\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big)\Phi_{k,\mathcal{A}+(\mathbb{Z}/Q)^d}\big(\theta^{(1)}\big)$ with

$$\sum_{\sigma\in\mathcal{A}+(\mathbb{Z}/Q)^d}\sum_{\underline{b},\underline{c}\in(\mathbb{Z}^d_Q)^r}\iota_Q(\sigma;\underline{b},\underline{c})$$
$$\times\mathcal{W}^r_Q(a^{(2)},b_1,c_1,\dots,b_r,c_r)\eta_{\le 2\delta' k}(\tau^k\circ(\theta^{(1)}-\sigma))$$
$$\times\int_{(\mathbb{R}^d)^{2r}}\Big\{\prod_{j=1}^r\eta_{\le\delta' k}(\tau^k\circ(\xi_j^{(1)}-\sigma+b_j/Q))$$
$$\times\eta_{\le\delta' k}(\tau^k\circ(\zeta_j^{(1)}-\sigma+c_j/Q))\Big\}$$
$$\times\mathcal{Z}^r_k(\alpha^{(2)};\theta^{(1)}-\xi_1^{(1)}-b_1/Q,\theta^{(1)}-\zeta_1^{(1)}-c_1/Q,\dots,$$
$$\theta^{(1)}-\xi_r^{(1)}-b_r/Q,\theta^{(1)}-\zeta_r^{(1)}-c_r/Q)$$
$$\times\prod_{1\le j\le r}\{\overline{S^\iota_k(\zeta_j^{(1)})}S^\iota_k(\xi_j^{(1)})\}d\xi_1^{(1)}d\zeta_1^{(1)}\dots d\xi_r^{(1)}d\zeta_r^{(1)}$$

at the expenses of $O(\tau^{-Dk/2})$ errors, where $\mathcal{Z}^r_k$ and $\mathcal{W}^r_Q$ have been defined in (5.24)–(5.25), $\underline{b}=(b_1,\dots,b_r)\in(\mathbb{Z}^d_Q)^r$, $\underline{c}=(c_1,\dots,c_r)\in(\mathbb{Z}^d_Q)^r$, and the coefficients ι_Q are defined by

$$\iota_Q(\sigma;\underline{b},\underline{c}):=\begin{cases}1 & \text{if } \sigma-b_j/Q,\sigma-c_j/Q\in\mathcal{A}\text{ for any } j\in\{1,\dots,r\};\\ 0 & \text{otherwise.}\end{cases}\tag{8.17}$$

We make the changes of variables $\xi_j^{(1)}=\beta_j+\sigma-b_j/Q$ and $\zeta_j^{(1)}=\gamma_j+\sigma-c_j/Q$ in the latter integral. In view of Lemma 6.1 we can also replace $S^\iota_k(\xi_j^{(1)})$ and $S^\iota_k(\zeta_j^{(1)})$ with $S(\sigma-b_j/Q)J^\iota_k(\beta_j)$ and $S(\sigma-c_j/Q)J^\iota_k(\gamma_j)$, at the expense of acceptable errors. Therefore, the integral formula above shows that if $\theta^{(2)}=\alpha^{(2)}+a^{(2)}/Q$ for some $a^{(2)}/Q\in\mathcal{B}$ and $|\tau^k\circ\alpha^{(2)}|\le 2\tau^{2\delta k}$, then

$$\Pi^{\iota,r}_{k,k,\mathcal{A}}(\theta^{(1)},\theta^{(2)})=\sum_{\sigma\in\mathcal{A}+(\mathbb{Z}/Q)^d}\sum_{\underline{b},\underline{c}\in(\mathbb{Z}^d_Q)^r}\iota_Q(\sigma;\underline{b},\underline{c})$$
$$\times\mathcal{W}^r_Q(a^{(2)},\underline{b},\underline{c})\eta_{\le 2\delta' k}(\tau^k\circ(\theta^{(1)}-\sigma))$$
$$\times\prod_{1\le j\le r}\{S(\sigma-b_j/Q)\overline{S(\sigma-c_j/Q)}\}$$
$$\times\int_{\mathbb{R}^{2rd}}\Big\{\prod_{1\le j\le r}\eta_{\le\delta' k}(\tau^k\circ\beta_j)\eta_{\le\delta' k}(\tau^k\circ\gamma_j)J^\iota_k(\beta_j)\overline{J^\iota_k(\gamma_j)}\Big\}$$
$$\times\mathcal{Z}^r_k(\alpha^{(2)};\theta^{(1)}-\sigma-\beta_1,\theta^{(1)}-\sigma-\gamma_1,\dots,$$
$$\theta^{(1)}-\sigma-\beta_r,\theta^{(1)}-\sigma-\gamma_r)\,d\beta_1d\gamma_1\dots d\beta_rd\gamma_r$$

$$+ O(\tau^{-Dk/3}), \tag{8.18}$$

where $\mathcal{W}^r_Q(a^{(2)}, \underline{b}, \underline{c}) = \mathcal{W}^r_Q(a^{(2)}, b_1, c_1, \dots, b_r, c_r)$.

Step 3. Using the definitions (6.4)–(6.5) and (5.25), the integral over β_j, γ_j in (8.18) is equal to

$$\begin{aligned}
&\int_{\mathbb{R}^{2rd}} \int_{\mathbb{R}^{2r}} \mathfrak{e}\Big(- (\tau^k \circ \alpha^{(2)}).\Big\{ \sum_{1\le j\le r} R_0(y_j, y_j - x_j) \\
&\qquad + \sum_{1\le l<j\le r} R_0(-y_l + x_l, -y_j + x_j)\Big\}\Big) \\
&\qquad \times \prod_{1\le j\le r} \Big\{ \eta_{\le \delta k}(x_j)\mathfrak{e}\Big(- (\tau^k \circ \alpha^{(1)}).x_j\Big)\eta_{\le \delta k}(y_j)\mathfrak{e}\Big((\tau^k \circ \alpha^{(1)}).y_j\Big) \\
&\qquad \times \chi^\iota(u_j)\chi^\iota(v_j)\widehat{\eta_{\le \delta' k}}(A_0^{(1)}(u_j) - x_j)\widehat{\eta_{\le \delta' k}}(-A_0^{(1)}(v_j) + y_j)\Big\}\, du_j dv_j dx_j dy_j,
\end{aligned} \tag{8.19}$$

where $\widehat{\eta_{\le \delta' k}}$ denotes the Euclidean Fourier transform of $\eta_{\le \delta' k}$ and $\alpha^{(1)} := \theta^{(1)} - \sigma$.

We notice that we may replace the factors $\eta_{\le \delta k}(x_j)$ and $\eta_{\le \delta k}(y_j)$ with 1 in the formula (8.19), at the expense of $O(\tau^{-Dk})$ errors, due to the stronger localizations induced by the factors in the last line. Then we make the changes of variables $x_j = A_0^{(1)}(u_j) + x'_j$, $y_j = A_0^{(1)}(v_j) + y'_j$ to rewrite the remaining integral in the form

$$\begin{aligned}
I^\iota_k(\alpha^{(1)}, \alpha^{(2)}) := \int_{\mathbb{R}^{2rd}} \prod_{1\le j\le r} \big\{\widehat{\eta_{\le \delta' k}}(-x'_j)\widehat{\eta_{\le \delta' k}}(y'_j)\mathfrak{e}\big(- (\tau^k \circ \alpha^{(1)}).(x'_j - y'_j)\big)\big\} \\
\times \Big\{ \int_{\mathbb{R}^{2r}} \mathfrak{e}\big(- (\tau^k \circ \alpha^{(2)}).T(\underline{x}', \underline{y}', \underline{u}, \underline{v})\big) \prod_{1\le j\le r} \big\{\chi^\iota(u_j)\chi^\iota(v_j)\big\} \\
\times \mathfrak{e}\big(- (\tau^k \circ \alpha).D(\underline{v}, \underline{u})\big)\, d\underline{u} d\underline{v}\Big\} d\underline{x}' d\underline{y}',
\end{aligned} \tag{8.20}$$

where $\alpha = (\alpha^{(1)}, \alpha^{(2)})$, the function $D : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}$ is defined as in (2.24), and

$$\begin{aligned}
&T(\underline{x}', \underline{y}', \underline{u}, \underline{v}) := T_1(\underline{x}', \underline{y}', \underline{u}, \underline{v}) + T_2(\underline{x}', \underline{y}'), \\
&T_1(\underline{x}', \underline{y}', \underline{u}, \underline{v}) := \sum_{1\le j\le r} \big[R_0(A_0^{(1)}(v_j), y'_j - x'_j) + R_0(y'_j, A_0^{(1)}(v_j) - A_0^{(1)}(u_j))\big] \\
&\quad + \sum_{1\le l<j\le r} \big[R_0(A_0^{(1)}(u_l) - A_0^{(1)}(v_l), x'_j - y'_j) + R_0(x'_l - y'_l, A_0^{(1)}(u_j) - A_0^{(1)}(v_j))\big], \\
&T_2(\underline{x}', \underline{y}') := \sum_{1\le j\le r} R_0(y'_j, y'_j - x'_j) + \sum_{1\le l<j\le r} R_0(x'_l - y'_l, x'_j - y'_j).
\end{aligned} \tag{8.21}$$

To summarize, we have proved that if $\theta^{(2)} = \alpha^{(2)} + a^{(2)}/Q$ for some $a^{(2)}/Q \in \mathcal{B}$ and $|\tau^k \circ \alpha^{(2)}| \le 2\tau^{2\delta k}$, then

$$\Big|\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big) - \sum_{\sigma\in\mathcal{A}+(\mathbb{Z}/Q)^d} \mathcal{C}(a^{(2)}/Q,\sigma)\eta_{\le 2\delta' k}(\tau^k\circ(\theta^{(1)}-\sigma)) \times I^{\iota}_k(\theta^{(1)}-\sigma,\alpha^{(2)})\Big| \lesssim \tau^{-Dk/3}, \tag{8.22}$$

where the multipliers I^{ι}_k are defined as in (8.20), and

$$\mathcal{C}(a^{(2)}/Q,\sigma) := \sum_{\underline{b},\underline{c}\in(\mathbb{Z}^d_Q)^r} \iota_Q(\sigma;\underline{b},\underline{c})\mathcal{W}^r_Q(a^{(2)},\underline{b},\underline{c}) \times \prod_{1\le j\le r}\big\{S(\sigma-b_j/Q)\overline{S(\sigma-c_j/Q)}\big\}. \tag{8.23}$$

Notice that the coefficients $\mathcal{C}(a^{(2)}/Q,\sigma)$ satisfy the desired bounds (8.12) because $Q^D \le q_{\mathcal{A}}$, and $\big|S(\varrho)\big| \lesssim q_{\mathcal{A}}^{-\delta}$ for any $\varrho \in \mathcal{A}$, as a consequence of (2.22).

Step 4. We now show that if $|\tau^k \circ \alpha^{(1)}| + |\tau^k \circ \alpha^{(2)}| \ge \tau^{\delta k/2}$ then

$$|I^{\iota}_k(\alpha^{(1)},\alpha^{(2)})| \lesssim \tau^{-k/\delta}. \tag{8.24}$$

We shall apply Proposition 2.4. For this we rewrite

$$\mathfrak{e}\big(-(\tau^k\circ\alpha^{(2)}).T_1(\underline{x}',\underline{y}',\underline{u},\underline{v})\big)\prod_{1\le j\le r}\big\{\chi^{\iota}(u_j)\chi^{\iota}(v_j)\big\} = \prod_{1\le j\le r}\big\{\psi_j(u_j)\phi_j(v_j)\big\},$$

where, using the formulas (8.21), we obtain

$$\begin{aligned}\psi_j(u_j) &:= \chi^{\iota}(u_j)\mathfrak{e}\Big\{-(\tau^k\circ\alpha^{(2)}).\big[-R_0(y'_j,A^{(1)}_0(u_j)) \\ &\quad + \sum_{j<l\le r} R_0(A^{(1)}_0(u_j),x'_l-y'_l) + \sum_{1\le l<j} R_0(x'_l-y'_l,A^{(1)}_0(u_j))\big]\Big\},\\ \phi_j(v_j) &:= \chi^{\iota}(v_j)\mathfrak{e}\Big\{-(\tau^k\circ\alpha^{(2)}).\big[R_0(A^{(1)}_0(v_j),y'_j-x'_j)+R_0(y'_j,A^{(1)}_0(v_j)) \\ &\quad - \sum_{j<l\le r} R_0(A^{(1)}_0(v_j),x'_l-y'_l) - \sum_{1\le l<j} R_0(x'_l-y'_l,A^{(1)}_0(v_j))\big]\Big\}.\end{aligned}$$

Then we notice that the contribution to the integral in (8.20) coming from the points $(\underline{x}',\underline{y}')$ outside the ball $B_r := \{(\underline{x}',\underline{y}') \in \mathbb{R}^{dr}\times\mathbb{R}^{dr} : |\underline{x}'|+|\underline{y}'| \le \tau^{-\delta' k/2}\}$ is negligible, due to the rapid decay of the function $\widehat{\eta_{\le\delta' k}}$. On the other hand, if $|\underline{x}'|+|\underline{y}'| \le \tau^{-\delta' k/2}$ and $|\tau^k\circ\alpha^{(2)}| \le 2\tau^{2\delta k}$, then the functions ψ_j and ϕ_j defined above have bounded $C^1(\mathbb{R})$ norms, $\|\psi_j\|_{C^1} + \|\phi_j\|_{C^1} \lesssim 1$, so

we can apply Proposition 2.4 for any $(x', y') \in B_r$. The desired bounds (8.24) follow.

On the other hand, if $|\tau^k \circ \alpha^{(1)}| + |\tau^k \circ \alpha^{(2)}| \lesssim \tau^{\delta k/2}$ then we observe that

$$\int_{\mathbb{R}^d} \widehat{\eta_{\le \delta' k}}(z) z^\beta \, dz = 0, \tag{8.25}$$

for any multi-index $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{N}^d \setminus \{0\}$. Since $T(\underline{x}', \underline{y}', \underline{u}, \underline{v})$ is a polynomial in the variables x_j, y_j, we can use a Taylor expansion to see that

$$\Big| \int_{\mathbb{R}^{2rd}} \prod_{1\le j\le r} \{\widehat{\eta_{\le \delta' k}}(-x'_j)\widehat{\eta_{\le \delta' k}}(y'_j)\} \Big[\mathfrak{e}\Big(- (\tau^k \circ \alpha^{(1)}). \sum_{1\le j\le r} (x'_j - y'_j)\Big) \\ \times \mathfrak{e}\big(- (\tau^k \circ \alpha^{(2)}).T(\underline{x}', \underline{y}', \underline{u}, \underline{v})\big) - 1 \Big] d\underline{x}' d\underline{y}' \Big| \lesssim \tau^{-Dk},$$

provided that $|\tau^k \circ \alpha^{(1)}| + |\tau^k \circ \alpha^{(2)}| \lesssim \tau^{\delta k/2}$ and $|\underline{u}| + |\underline{v}| \lesssim 1$. Recalling also the definition (7.11), we have the approximate identity

$$I^\iota_k(\alpha) = P^\iota(\tau^k \circ \alpha)\eta_{\le \delta k/2}(\tau^k \circ \alpha^{(1)})\eta_{\le \delta k/2}(\tau^k \circ \alpha^{(2)}) + O(\tau^{-k/\delta}). \tag{8.26}$$

Step 5. We examine the functions $\Omega^r_{k,k,\mathcal{B}}$ defined in (8.15). Using (5.6) it is easy to see that

$$\big|\Omega^r_{k,k,\mathcal{B}}(a^{(2)}/Q + \alpha^{(2)}) - 1\big| \lesssim \tau^{-Dk} \quad \text{if } |\tau^k \circ \alpha^{(2)}| \le 2\tau^{\delta k/2} \text{ and } a^{(2)}/Q \in \mathcal{B}. \tag{8.27}$$

Compare (8.27) with the bounds from (7.17). Combining this with (8.22), (8.24), and (8.26) we derive our main approximate identity for multipliers,

$$\Big| \Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)}, \theta^{(2)}\big)\Omega^r_{k,k,B}(\theta^{(2)}) - \sum_{a^{(2)}/Q\in\mathcal{B}} \sum_{\sigma\in\mathcal{A}+(\mathbb{Z}/Q)^d} \mathcal{C}(a^{(2)}/Q, \sigma) \\ \times \eta_{\le \delta k/2}(\tau^k \circ (\theta^{(1)} - \sigma))\eta_{\le \delta k/2}(\tau^k \circ (\theta^{(2)} - a^{(2)}/Q)) \\ \times P^\iota(\tau^k \circ (\theta^{(1)} - \sigma, \theta^{(2)} - a^{(2)}/Q))\Big| \lesssim \tau^{-k/\delta}. \tag{8.28}$$

The desired conclusions (8.8)–(8.9) follow using the identity (8.13). □

We now return to the proof of the bounds (8.5). In view of the Cotlar–Stein lemma it suffices to prove the following:

Lemma 8.3 *Assume that $s \geq 0$, $t \geq D(s+1)$, and let $\mathcal{A} \subseteq \mathcal{R}^d_t \setminus \widetilde{\mathcal{R}}^d_{Q_s}$, $\mathcal{B} \subseteq \mathcal{R}^{d'}_{\leq s}$ be 1-periodic sets of rationals. If $k, j \geq \max(\kappa_s, t/\delta')$ and $j \in [k/2, k]$, then*

$$\begin{aligned}\|\mathcal{G}^1_{j,j,\mathcal{A},\mathcal{B}}(\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}})^*\|_{\ell^2(\mathbb{G}_0)\to\ell^2(\mathbb{G}_0)} &+\|(\mathcal{G}^1_{j,j,\mathcal{A},\mathcal{B}})^*\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}}\|_{\ell^2(\mathbb{G}_0)\to\ell^2(\mathbb{G}_0)}\\ &\lesssim \tau^{-2t/D}\tau^{-2|j-k|/D}. \end{aligned} \tag{8.29}$$

Proof We will use Lemma 8.2 with $\iota = 1$, since $G^1_{k,k,\mathcal{A},\mathcal{B}} = K'_{k,k,\mathcal{A},\mathcal{B}}$. The proof will proceed in several steps as the proof of Lemma 7.2.

Step 1. We will abbreviate $F^{1,r}_k(h)$ to $F^r_k(h)$, where

$$\begin{aligned}F^r_k(h) := \Big\{ \sum_{a^{(2)}/Q\in\mathcal{B}\cap[0,1)^{d'}} \sum_{\sigma\in[\mathcal{A}+(\mathbb{Z}/Q)^d]\cap[0,1)^d} &\mathcal{C}(a^{(2)}/Q,\sigma)\\ \times\mathfrak{e}(h^{(1)}.\sigma)\mathfrak{e}\big(h^{(2)}.(a^{(2)}/Q)\big)\Big\} &X^r_k(h),\end{aligned}$$

where $X^r_k := X^{1,r}_k$ are the kernels defined in (7.26). In view of (7.27) and (8.12) we have

$$\|F^r_k\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-t/(2\delta)}.$$

This shows that $\|\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}}\|_{\ell^2(\mathbb{G}_0)\to\ell^2(\mathbb{G}_0)} \lesssim \tau^{-t/r}$, and bound (8.29) follows if $j = k$.

To prove the bounds (8.29) in the general case $j \leq k$ we use first a high order T^*T argument, as in (7.28), so it suffices to prove that

$$\begin{aligned}\|\mathcal{G}^1_{j,j,\mathcal{A},\mathcal{B}}[(\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}})^*\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}}]^r\|_{\ell^2\to\ell^2} + \|(\mathcal{G}^1_{j,j,\mathcal{A},\mathcal{B}})^*[\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}}(\mathcal{G}^1_{k,k,\mathcal{A},\mathcal{B}})^*]^r\|_{\ell^2\to\ell^2}\\ \lesssim \tau^{-8rt/D}\tau^{-8r|j-k|/D},\end{aligned} \tag{8.30}$$

for any $j \in [k/2, k]$ such that $k, j \geq \max(\kappa_s, t/\delta')$. The two bounds are similar, so we will focus on bounding the first term. We use Lemma 8.2, and notice that the contribution of the error kernel $O^{1,r}_k$ is controlled by $O(\tau^{-k})$, which is better than needed. It remains to prove that

$$\big\|F^r_k * G^1_{j,j,\mathcal{A},\mathcal{B}}\big\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-8rt/D}\tau^{-8r|j-k|/D}. \tag{8.31}$$

Step 2. Using Lemma 6.1 the kernels $G^1_{j,j,\mathcal{A},\mathcal{B}} = K'_{j,j,\mathcal{A},\mathcal{B}} = L'_{j,j,\mathcal{A}}N_{j,j,\mathcal{B}}$, can be rewritten as

$$\sum_{b^{(1)}/q_1\in\mathcal{A}\cap[0,1)^d,\, b^{(2)}/q_2\in\mathcal{B}\cap[0,1)^{d'}} S(b^{(1)}/q_1)\mathfrak{e}(g^{(1)}.(b^{(1)}/q_1))\mathfrak{e}(g^{(2)}.(b^{(2)}/q_2))Y_j(g),$$
$$Y_j(g) := \phi_j(g)\int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \mathfrak{e}(g.\zeta)J_j'(\zeta^{(1)})\eta_{\le\delta' j}(\tau^j\circ\zeta^{(1)})\eta_{\le\delta j}(\tau^j\circ\zeta^{(2)})\,d\zeta^{(1)}d\zeta^{(2)}, \tag{8.32}$$

up to rapidly decreasing errors. Here $\phi_j(g) = \phi_j^{(1)}(g^{(1)})\phi_j^{(2)}(g^{(2)})$ as before, and the functions J_j' are defined as in (6.4).

As in (7.31), we define $X^r_{k,a/q}(h) = X^r_k(h)\mathfrak{e}(h.a/q)$. We define also $Y_{j,a/q}(g) = Y_j(g)\mathfrak{e}(g.a/q)$, with Y_j as in (8.32). By the definition of F^r_k and the rapid exponential decay $|\mathcal{C}(a^{(2)}/Q,\sigma)| \lesssim \tau^{-t/\delta}$ (see (8.12) with $\mathcal{A}\subseteq\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_s}$ and $\mathcal{B}\subseteq\mathcal{R}^{d'}_{\le s}$), for (8.31) it suffices to prove that

$$\big\|X^r_{k,a/q} * Y_{j,a'/q'}\big\|_{\ell^1(\mathbb{G}_0)} \lesssim \tau^{-8r|j-k|/D} \tag{8.33}$$

for any irreducible fractions $a/q, a'/q' \in \mathbb{Q}^{d+d'}$ with denominators $q, q' \le \tau^{2t+2}$.

Step 3. Let $Q = qq' \in [1,\tau^{4t+4}]$ and recall the definitions (6.1)–(6.2). Since $\mathfrak{e}((g\cdot h\cdot g').a/q) = \mathfrak{e}((g\cdot g').a/q)$ and $\mathfrak{e}((g\cdot h\cdot g').a'/q') = \mathfrak{e}((g\cdot g').a'/q')$ if $h\in\mathbb{H}_Q$ and $g, g'\in\mathbb{G}_0$, we have

$$\begin{aligned}
&\big\|X^r_{k,a/q} * Y_{j,a'/q'}\big\|_{\ell^1(\mathbb{G}_0)}\\
&\quad= \sum_{\mu\in\mathbb{J}_Q,\, h\in\mathbb{H}_Q}\Big|\sum_{\mu_1\in\mathbb{J}_Q,\, h_1\in\mathbb{H}_Q} X^r_{k,a/q}(\mu_1^{-1}\cdot h_1^{-1}\cdot h\cdot\mu)Y_{j,a'/q'}(h_1\cdot\mu_1)\Big|\\
&\quad\le \sum_{\mu,\mu_1\in\mathbb{J}_Q,\, h\in\mathbb{H}_Q}\Big|\sum_{h_1\in\mathbb{H}_Q} X^r_k(\mu_1^{-1}\cdot h_1^{-1}\cdot h\cdot\mu)Y_j(h_1\cdot\mu_1)\Big|.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\big\|X^r_{k,a/q} * Y_{j,a'/q'}\big\|_{\ell^1(\mathbb{G}_0)}\\
&\quad\lesssim \sum_{\mu,\mu_1\in\mathbb{J}_Q,\, h,h_1\in\mathbb{H}_Q}\big|X^r_k(\mu_1^{-1}\cdot h_1^{-1}\cdot h\cdot\mu) - X^r_k(h\cdot\mu)\big|\big|Y_j(h_1\cdot\mu_1)\big|\\
&\qquad+\sum_{\mu,\mu_1\in\mathbb{J}_Q,\, h\in\mathbb{H}_Q}|X^r_k(h\cdot\mu)|\Big|\sum_{h_1\in\mathbb{H}_Q} Y_j(h_1\cdot\mu_1)\Big|.
\end{aligned} \tag{8.34}$$

Using (7.38), for any $g, g_1\in\mathbb{G}_0$ we have

$$\begin{aligned}
|X^r_k(g_1^{-1}\cdot g) - X^r_k(g)| \lesssim\ & \tau^{j-k}(1+|\tau^{-j}\circ g_1|)^{2/\delta+2}\\
&\times\Big\{\prod_{(l_1,l_2)\in Y_d}\tau^{-k(l_1+l_2)}\Big\}(1+|\tau^{-k}\circ g|)^{-1/\delta+1},
\end{aligned}$$

which is a stronger version of (7.39). Moreover, using the definition of the kernel Y_j in (8.32),

$$|Y_j(g_1)| \lesssim \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-j(l_1+l_2-\delta_{l_1l_2})} \Big\} \times \int_{\mathbb{R}} |\chi'(u)| \Big(1 + \big|\tau^{\widetilde{\delta} j}\big(A_0(u) - \tau^{-j} \circ g_1\big)\big|\Big)^{-4/\delta} du,$$

uniformly in $g_1 \in \mathbb{G}_0$. Here $\widetilde{\delta} = (\delta_{l_1l_2})_{(l_1,l_2)\in Y_d}$ and $\delta_{l_1l_2} = \delta$ if $(l_1, l_2) \in Y_d'$ and $\delta_{l_1l_2} = \delta'$ otherwise. Since

$$1 + |\tau^{-j} \circ g_1| \lesssim 1 + \big|\tau^{\widetilde{\delta} j}\big(A_0(u) - \tau^{-j} \circ g_1\big)\big|,$$

we obtain the desired bound for the first term in the right-hand side of (8.34).

Next, we focus on the second term in the right-hand side of (8.34). Notice that using (2.17) we are able to prove that

$$\Big| \sum_{h_1\in\mathbb{H}_Q} \phi_j(h_1 \cdot \mu_1)\mathfrak{e}\big((h_1 \cdot \mu_1).\zeta\big)\Big| \lesssim Q^{-d-d'} \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{j(l_1+l_2+\delta)} \Big\} \Big(1 + \tau^{\delta j}|\tau^j \circ \zeta|\Big)^{-D},$$

uniformly in $|\tau^j \circ \zeta| \lesssim \tau^{j/4}$, $Q \lesssim \tau^{j/8}$ and $\mu_1 \in \mathbb{J}_Q$. Further, since $J_j'(0) = 0$, it follows from the definition of J_j' (see (6.4)) that $|J_j'(\zeta^{(1)})| \lesssim \min(1, |\tau^j \circ \zeta^{(1)}|)$ for any $\zeta^{(1)} \in \mathbb{R}^d$. Combining the above with (7.38) we bound the second term in the right-hand side of (8.34) by

$$\begin{aligned}
&\sum_{\mu_1\in\mathbb{J}_Q} \Big| \sum_{h_1\in\mathbb{H}_Q} Y_j(h_1 \cdot \mu_1)\Big| \\
&\le \sum_{\mu_1\in\mathbb{J}_Q} \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \Big| \sum_{h_1\in\mathbb{H}_Q} \phi_j(h_1 \cdot \mu_1)\mathfrak{e}\big((h_1 \cdot \mu_1).\zeta\big)\Big| \big|J_j'(\zeta^{(1)})\big| \\
&\quad \times \big|\eta_{\le\delta' j}(\tau^j \circ \zeta^{(1)})\eta_{\le\delta j}(\tau^j \circ \zeta^{(2)})\big|\, d\zeta^{(1)}d\zeta^{(2)} \\
&\lesssim \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{j(l_1+l_2+\delta)} \Big\} \Big(1 + \tau^{\delta j}|\tau^j \circ \zeta|\Big)^{-D} \\
&\quad \times \big|\tau^j \circ \zeta^{(1)}\big|\, d\zeta^{(1)}d\zeta^{(2)} \\
&\lesssim \tau^{-\delta j}.
\end{aligned}$$

Recalling that $j \in [k/2, k]$ we see that the desired estimates (8.33) follow. □

8.2 Proof of the bounds (8.6) for $\iota = 2$

Notice that if g is in the support of the kernel $G^2_{k,k,\mathcal{A},\mathcal{B}}$ then there is $(l_1, l_2) \in Y_d$ such that $|g_{l_1 l_2}| \gtrsim \tau^{\delta k} \tau^{k(l_1+l_2)}$. Therefore we can integrate by parts many times in the variable $\xi_{l_1 l_2}$ to prove that the kernels $G^2_{k,k,\mathcal{A},\mathcal{B}}$ have rapid decay, i.e. $|G^2_{k,k,\mathcal{A},\mathcal{B}}(g)| \lesssim \tau^{-k/\delta}$ for any $g \in \mathbb{G}_0$. The desired bounds (8.6) follow. □

8.3 Proofs of the bounds (8.6) for $\iota = 3$ and $\iota = 4$

As before, we use a high order T^*T argument. Notice that the kernels $G^3_{k,k,\mathcal{A},\mathcal{B}}$ and $G^4_{k,k,\mathcal{A},\mathcal{B}}$ defined in (8.4) have product structure

$$\begin{aligned}
&G^3_{k,k,\mathcal{A},\mathcal{B}}(g) = I^3_{k,k,\mathcal{A}}(g^{(1)}) J^3_{k,k,\mathcal{B}}(g^{(2)}),\\
&I^3_{k,k,\mathcal{A}}(g^{(1)}) := \phi^{(1)}_{k+1}(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)}) \Psi_{k,k,\mathcal{A}}(\xi^{(1)}) S_{k+1}(\xi^{(1)})\, d\xi^{(1)},\\
&J^3_{k,k,\mathcal{B}}(g^{(2)}) := \phi^{(2)}_{k+1}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)}) [\Delta_k \Xi_{k,k,\mathcal{B}}](\xi^{(2)})\, d\xi^{(2)}, \quad (8.35)
\end{aligned}$$

and

$$\begin{aligned}
&G^4_{k,k,\mathcal{A},\mathcal{B}}(g) = I^4_{k,k,\mathcal{A}}(g^{(1)}) J^4_{k,k,\mathcal{B}}(g^{(2)}),\\
&I^4_{k,k,\mathcal{A}}(g^{(1)}) := \phi^{(1)}_{k+1}(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)}) [\Delta_k \Psi_{k,k,\mathcal{A}}](\xi^{(1)}) S_{k+1}(\xi^{(1)})\, d\xi^{(1)},\\
&J^4_{k,k,\mathcal{B}}(g^{(2)}) := \phi^{(2)}_{k+1}(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)}) \Xi_{k+1,k+1,\mathcal{B}}(\xi^{(2)})\, d\xi^{(2)}. \quad (8.36)
\end{aligned}$$

We define the operators $\mathcal{G}^\iota_{k,k,\mathcal{A},\mathcal{B}}$ by $\mathcal{G}^\iota_{k,k,\mathcal{A},\mathcal{B}} f := f * G^\iota_{k,k,\mathcal{A},\mathcal{B}}$, $\iota \in \{3, 4\}$. Using (2.13)–(2.16) we have

$$\big\{(\mathcal{G}^\iota_{k,k,\mathcal{A},\mathcal{B}})^* \mathcal{G}^\iota_{k,k,\mathcal{A},\mathcal{B}}\big\}^r f = f * G^{\iota,r}_{k,k,\mathcal{A},\mathcal{B}},$$

for a sufficiently large integer $r \in \mathbb{Z}_+$ and $\iota \in \{3, 4\}$, where the kernels $G^{\iota,r}_{k,k,\mathcal{A},\mathcal{B}}$ are given by

$$\begin{aligned}
&G^{\iota,r}_{k,k,\mathcal{A},\mathcal{B}}(y)\\
&\quad := \eta_{\le 3\delta k}(\tau^{-k} \circ y) \int_{\mathbb{T}^d \times \mathbb{T}^{d'}} \mathfrak{e}\big(y.\theta\big) \Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)}, \theta^{(2)}\big) \Omega^{\iota,r}_{k,k,\mathcal{B}}\big(\theta^{(2)}\big)\, d\theta^{(1)} d\theta^{(2)}.
\end{aligned}$$

The multipliers $\Pi^{\iota,r}_{k,k,\mathcal{A}}$ are given by

$$\begin{aligned}\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big) := \sum_{h_j^{(1)},g_j^{(1)}\in\mathbb{Z}^d} \Big\{\prod_{j=1}^{r}\overline{I^{\iota}_{k,k,\mathcal{A}}(h_j^{(1)})}I^{\iota}_{k,k,\mathcal{A}}(g_j^{(1)})\Big\} \\ \times \mathfrak{e}\big(\theta^{(1)}\cdot\sum_{1\le j\le r}(h_j^{(1)}-g_j^{(1)})\big) \\ \times \mathfrak{e}\Big(-\theta^{(2)}\cdot\Big\{\sum_{1\le j\le r}R_0(h_j^{(1)},h_j^{(1)}-g_j^{(1)}) \\ +\sum_{1\le l<j\le r}R_0(-h_l^{(1)}+g_l^{(1)},-h_j^{(1)}+g_j^{(1)})\Big\}\Big).\end{aligned} \tag{8.37}$$

Moreover, with F_{k+1} defined as in (5.3), the multipliers $\Omega^{\iota,r}_{k,k,\mathcal{B}}$ are given by

$$\begin{aligned}\Omega^{3,r}_{k,k,\mathcal{B}}\big(\theta^{(2)}\big) &:= \Big|\int_{\mathbb{T}^{d'}}F_{k+1}(\theta^{(2)}-\xi^{(2)})[\Delta_k\Xi_{k,k,\mathcal{B}}](\xi^{(2)})\,d\xi^{(2)}\Big|^{2r},\\ \Omega^{4,r}_{k,k,\mathcal{B}}\big(\theta^{(2)}\big) &:= \Big|\int_{\mathbb{T}^{d'}}F_{k+1}(\theta^{(2)}-\xi^{(2)})\Xi_{k+1,k+1,\mathcal{B}}(\xi^{(2)})\,d\xi^{(2)}\Big|^{2r}.\end{aligned} \tag{8.38}$$

For (8.6) it suffices to prove that for $\iota\in\{3,4\}$ we have the multiplier bounds

$$\Big|\Pi^{\iota,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big)\Omega^{\iota,r}_{k,k,\mathcal{B}}\big(\theta^{(2)}\big)\Big| \lesssim \tau^{-k/\delta} \quad \text{for any } (\theta^{(1)},\theta^{(2)})\in\mathbb{T}^d\times\mathbb{T}^{d'}. \tag{8.39}$$

The proof of (8.39) follows by similar arguments as in Lemma 8.2. We consider two cases:

Case 1. Assume first that $\iota = 3$. Notice that we have rapid decay $|\Omega^{3,r}_{k,k,\mathcal{B}}(\theta^{(2)})| \lesssim \tau^{-Dk}$ unless $|\tau^k\circ(\theta^{(2)}-a^{(2)}/Q)| \le \tau^{2\delta k}$ for some $a^{(2)}/Q\in\mathcal{B}$. In this case the symbols $\Pi^{3,r}_{k,k,\mathcal{A}}$ satisfy similar bounds as the symbols $\Pi^{0,r}_{k,k,\mathcal{A}}$ analyzed in the proof of Lemma 8.2. In particular, we have

$$\begin{aligned}\Big|\Pi^{3,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big) - \sum_{\sigma\in\mathcal{A}+(\mathbb{Z}/Q)^d}\mathcal{C}(a^{(2)}/Q,\sigma)\eta_{\le 2\delta' k}(\tau^k\circ(\theta^{(1)}-\sigma)) \\ \times I^3_k(\theta^{(1)}-\sigma,\alpha^{(2)})\Big| \lesssim \tau^{-Dk/3},\end{aligned}$$

which is analogous to the approximate identity (8.22). The coefficients $\mathcal{C}(a^{(2)}/Q,\sigma)$ are as in (8.23), while the functions I^3_k are similar to the functions I^{ι}_k defined in (8.20) (with the factor $\chi^{\iota}(u_j)\chi^{\iota}(v_j)$ replaced by $\chi(u_j/2)\chi(v_j/2)/4$). We still have the key bounds

$$|I_k^3(\alpha^{(1)},\alpha^{(2)})| \lesssim \tau^{-k/\delta} \quad \text{if } |\tau^k\circ\alpha^{(1)}|+|\tau^k\circ\alpha^{(2)}|\geq \tau^{\delta k/2},$$

which are similar to (8.24). The main difference is that the bounds (8.27) are replaced by

$$\big|\Omega^{3,r}_{k,k,\mathcal{B}}(a^{(2)}/Q+\alpha^{(2)})\big| \lesssim \tau^{-Dk} \quad \text{if } |\tau^k\circ\alpha^{(2)}|\leq 2\tau^{\delta k/2} \text{ and } a^{(2)}/Q\in\mathcal{B},$$

due to the presence of the difference factor $[\Delta_k\Xi_{k,k,\mathcal{B}}](\xi^{(2)})$ in the definition (8.38) of the multipliers $\Omega^{3,r}_{k,k,\mathcal{B}}$. The desired estimate (8.39) for $\iota=3$ follows from the last three bounds.

Case 2. Assume now that $\iota=4$. As in (8.16) we rewrite

$$\Pi^{4,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big)=\int_{(\mathbb{T}^d)^{2r}}\mathcal{V}^r_{k+1}(\theta^{(1)},\theta^{(2)};\zeta_1^{(1)},\xi_1^{(1)},\dots,\zeta_r^{(1)},\xi_r^{(1)}) \tag{8.40}$$
$$\times\prod_{1\leq j\leq r}\Big\{\overline{S_{k+1}(\zeta_j^{(1)})}\,\overline{[\Delta_k\Psi_{k,k,\mathcal{A}}](\zeta_j^{(1)})}S_{k+1}(\xi_j^{(1)})$$
$$\times[\Delta_k\Psi_{k,k,\mathcal{A}}](\xi_j^{(1)})\Big\}\,d\xi_1^{(1)}d\zeta_1^{(1)}\dots d\xi_r^{(1)}d\zeta_r^{(1)},$$

where $\mathcal{V}^r_{k+1}(\theta^{(1)},\theta^{(2)};\zeta_1^{(1)},\xi_1^{(1)},\dots,\zeta_r^{(1)},\xi_r^{(1)})$ is as in (5.16).

In view of (5.6) we have a rapid decay $|\Omega^{4,r}_{k,k,\mathcal{B}}(\theta^{(2)})|\lesssim\tau^{-Dk}$ unless $|\tau^k\circ(\theta^{(2)}-a^{(2)}/Q)|\leq\tau^{2\delta k}$ for some $a^{(2)}/Q\in\mathcal{B}$. On the other hand, in this case we can use similar arguments as in the proof of Lemma 8.2 to simplify the multipliers $\Pi^{4,r}_{k,k,\mathcal{A}}$, at the expense of acceptable errors. After several reductions we derive an approximate formula similar to (8.22), namely

$$\Big|\Pi^{4,r}_{k,k,\mathcal{A}}\big(\theta^{(1)},\theta^{(2)}\big)-\sum_{\sigma\in\mathcal{A}+(\mathbb{Z}/Q)^d}\mathcal{C}(a^{(2)}/Q,\sigma)\eta_{\leq 2\delta' k}(\tau^k\circ(\theta^{(1)}-\sigma))$$
$$\times I_k^4(\theta^{(1)}-\sigma,\alpha^{(2)})\Big|\lesssim\tau^{-Dk/3}, \tag{8.41}$$

provided that $\theta^{(2)}=\alpha^{(2)}+a^{(2)}/Q$ for some $a^{(2)}/Q\in\mathcal{B}$ and $|\tau^k\circ\alpha^{(2)}|\leq\tau^{2\delta k}$. The coefficients $\mathcal{C}(a^{(2)}/Q,\sigma)$ are the same as in (8.23), and I_k^4 is defined as in (8.20), namely

$$I_k^4(\alpha^{(1)},\alpha^{(2)}):=\int_{\mathbb{R}^{2rd}}\prod_{1\leq j\leq r}\big\{\widehat{\eta'_{\leq\delta' k}}(-x'_j)\widehat{\eta'_{\leq\delta' k}}(y'_j)\mathfrak{e}\big(-(\tau^k\circ\alpha^{(1)}).(x'_j-y'_j)\big)\big\}$$
$$\times\Big\{\int_{\mathbb{R}^{2r}}\mathfrak{e}\big(-(\tau^k\circ\alpha^{(2)}).T(\underline{x}',\underline{y}',\underline{u},\underline{v})\big)\prod_{1\leq j\leq r}\big\{\chi(u_j)\chi(v_j)\big\}$$
$$\times\mathfrak{e}\big(-(\tau^k\circ\alpha).D(2\underline{v},2\underline{u})\big)\,d\underline{u}d\underline{v}\Big\}d\underline{x}'d\underline{y}', \tag{8.42}$$

where $\eta'_{\leq\delta' k}(z) := \eta_{\leq\delta'(k+1)}(\tau\circ z) - \eta_{\leq\delta' k}(z)$, and $\widehat{\eta'_{\leq\delta' k}}$ denotes the Fourier transform of the function $\eta'_{\leq\delta' k}$, and the function T is defined as in (8.21).

The functions I_k^4 still satisfy the bounds $|I_k^4(\alpha^{(1)},\alpha^{(2)})| \lesssim \tau^{-k/\delta}$ if $|\tau^k\circ\alpha^{(1)}|+|\tau^k\circ\alpha^{(2)}|\geq\tau^{\delta k/2}$, which are similar to (8.24). The main difference is that the identities (8.25) are replaced by the stronger identities

$$\int_{\mathbb{R}^d}\widehat{\eta'_{\leq\delta' k}}(z)z^\beta\,dz = 0,$$

for any multi-index β, including $\beta = 0$. Therefore we can use a Taylor expansion (as in the proof of (8.26)) to see that $|I_k^4(\alpha^{(1)},\alpha^{(2)})| \lesssim \tau^{-Dk}$ if $|\tau^k\circ\alpha^{(1)}|+|\tau^k\circ\alpha^{(2)}|\lesssim\tau^{\delta k/2}$. The desired bound in (8.39) follows for $\iota = 4$. □

9 Maximal estimates on $\ell^p(\mathbb{G}_0)$: Proof of Theorem 1.3

In this section we complete the proof of the ℓ^p theory in Theorem 1.3.

Theorem 9.1 *With $\mathcal{M}_k$ defined as in (4.1) for $\tau = 2$, and for any $p\in(1,\infty]$we have*

$$\Big\|\sup_{k\geq 0}|\mathcal{M}_k f|\Big\|_{\ell^p(\mathbb{G}_0)} \lesssim_p \|f\|_{\ell^p(\mathbb{G}_0)},\quad f\in\ell^p(\mathbb{G}_0). \tag{9.1}$$

Notice that the maximal inequality (9.1) for $\tau = 2$ implies the full maximal inequality for any $\tau > 1$. By interpolation with the variational ℓ^2 estimates in Theorem 4.1, this completes the proof of the main Theorem 1.3.

To prove Theorem 9.1 we will use Lemma 9.2 and Propositions 9.3 and 9.4 below.

Lemma 9.2 *Assume that there is a constant $\gamma > 0$ such that for every $u\in(1,2]$, $\rho\in(0,1)$, and $\lambda > 0$ there is a sequence of linear operators $(A_k^{\lambda,\rho})_{k\geq 0}$ such that*

$$\Big\|\sup_{k\geq 0}|A_k^{\lambda,\rho}f|\Big\|_{\ell^u(\mathbb{G}_0)} \lesssim_{\rho,u} \lambda^\rho\|f\|_{\ell^u(\mathbb{G}_0)},\quad \text{for any } f\in\ell^u(\mathbb{G}_0), \tag{9.2}$$

and

$$\Big\|\sup_{k\geq 0}|\mathcal{M}_k f - A_k^{\lambda,\rho}f|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim_\rho \lambda^{-\gamma}\|f\|_{\ell^2(\mathbb{G}_0)},\quad \text{for any } f\in\ell^2(\mathbb{G}_0). \tag{9.3}$$

Then the estimate (9.1) holds true for every $p > 1$.

Proof This is a general interpolation result. See for example [34, Lemma 7.1] or [32, Lemma 4.4] for proofs of such results. □

We will need the following logarithmic maximal estimates.

Proposition 9.3 *For every $p \in (1,\infty)$, $f \in \ell^p(\mathbb{G}_0)$, and $J \in \mathbb{N}$ we have*

$$\Big\| \sup_{j\in[J+1,2J]} |\mathcal{M}_j f| \Big\|_{\ell^p(\mathbb{G}_0)} \lesssim_p \log(J+2)\|f\|_{\ell^p(\mathbb{G}_0)}.$$

Proposition 9.3 will be proved in Sect. 9.2. The idea of using restricted $\ell^p(\mathbb{G}_0)$ estimates as in Proposition 9.3 together with $\ell^2(\mathbb{G}_0)$ bounds to prove the full $\ell^p(\mathbb{G}_0)$ estimates (9.1) originates in Bourgain's paper [13].

Finally, we will also need the following shifted maximal inequality for the kernels $W_{k,w,Q}$ with $0 \le w \le k$ defined in (6.8).

Proposition 9.4 *For any $p \in (1,\infty)$, $Q \ge 1$, and $w \in \mathbb{N}$ we have*

$$\Big\| \sup_{2^{k/4}\ge Q,\, k\ge w} \big|f *_{\mathbb{H}_Q} W_{k,w,Q}\big| \Big\|_{\ell^p(\mathbb{H}_Q)} \lesssim_p (w+1)\|f\|_{\ell^p(\mathbb{H}_Q)}, \quad f \in \ell^p(\mathbb{H}_Q).$$

We prove Proposition 9.4 in Appendix B. For now we show how to use the conclusions of Propositions 9.3 and 9.4 to complete the proof of Theorem 9.1.

9.1 Proof of Theorem 9.1

We divide the proof in several steps:

Step 1. In view Lemma 9.2, in order to prove (9.1) it suffices to find a sequence of linear operators $(A_k^{\lambda,\rho})_{k\in\mathbb{N}}$, $\rho \in (0,1)$ and $\lambda > 0$ satisfying (9.2) and (9.3). For $\lambda \le e^D$ we can just set $A_k^{\lambda,\rho} \equiv 0$ and the bounds (9.3) follow from the already established $\ell^2(\mathbb{G}_0)$ theory for the maximal operator $\sup_{k\ge 0}|\mathcal{M}_k f|$.

Therefore from now on we may focus only on $\lambda \ge e^D$. Let us define

$$S := \lfloor \ln \lambda \rfloor \ge D. \tag{9.4}$$

Recall from (4.18) and (4.25) that for S as in (9.4) we have respectively

$$\kappa_S = 2^{(D/\ln 2)(S+1)^2} \quad \text{and} \quad Q_S = (2^{D(S+1)})!.$$

If $\lambda \ge e^D$ and $k \le \kappa_S$ then we just define $A_k^{\lambda,\rho} = \mathcal{M}_k$. The bounds (9.3) are trivial, whereas the bounds (9.2) follow from Proposition 9.3. Indeed, since $S^4 \simeq (\ln\lambda)^4$ we have

$$\Big\| \sup_{1\le k\le \kappa_S} |\mathcal{M}_k f| \Big\|_{\ell^u(\mathbb{G}_0)} \le \sum_{j=1}^{2D(S+1)^2} \Big\| \sup_{2^{j-1}\le k\le 2^j} |\mathcal{M}_k f| \Big\|_{\ell^u(\mathbb{G}_0)}$$
$$\lesssim \sum_{j=0}^{2D(S+1)^2} (j+1)\|f\|_{\ell^u(\mathbb{G}_0)} \lesssim (\log\lambda)^4 \|f\|_{\ell^u(\mathbb{G}_0)}.$$

Step 2. Assume now that $\lambda \ge e^D$ and $k \ge \kappa_S$. We set $A_k^{\lambda,\rho} f = f * K_{k,S,\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'}}$, where the kernels $K_{k,w,\mathcal{A},\mathcal{B}}$ are defined as in (4.36). In view of Lemma 9.2 it suffices to show that

$$\Big\| \sup_{k\ge\kappa_S} |f * K_{k,S,\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'}}| \Big\|_{\ell^p(\mathbb{G}_0)} \lesssim_p (\ln\lambda)\|f\|_{\ell^p(\mathbb{G}_0)}, \quad f\in\ell^p(\mathbb{G}_0) \tag{9.5}$$

$$\Big\| \sup_{k\ge\kappa_S} |\mathcal{M}_k f - f * K_{k,S,\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'}}| \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-\delta/D^3}\|f\|_{\ell^2(\mathbb{G}_0)}, \quad f\in\ell^2(\mathbb{G}_0). \tag{9.6}$$

for every $p\in(1,2]$.

Let $\mathcal{K}_{k,w,\mathcal{A},\mathcal{B}}$, $\mathcal{W}_{k,w,Q}$ and $\mathcal{V}_{\mathcal{A},\mathcal{B},Q}$ denote the convolution operators corresponding respectively to the kernels $K_{k,w,\mathcal{A},\mathcal{B}}$, $W_{k,w,Q}$ and $V_{\mathcal{A},\mathcal{B},Q}$ defined in Lemma 6.2. Let $Q = Q_S$, $\mathcal{A} = \widetilde{\mathcal{R}}_{Q_S}^d$, $\mathcal{B} = \widetilde{\mathcal{R}}_{Q_S}^{d'}$, $k_0 = \lfloor \kappa_S \rfloor$, and $w = S$. Notice that $1 \le Q_S \le 2^{\delta k_0}$, so the decomposition (6.7) and the error term estimate (6.10) of Lemma 6.2 hold.

We prove first the bounds (9.5). We apply Lemma 6.6 with $\mathcal{K}_k^{\mathbb{G}_0} = \mathcal{K}_{k,S,\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'}}$, $\mathcal{W}_k^{\mathbb{H}_{Q_S}} = \mathcal{W}_{k,S,Q_S}$ and $\mathcal{V}^{\mathbb{J}_{Q_S}} = \mathcal{V}_{\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'},Q_S}$ and conclude from (6.42) (with $B=\ell^\infty$) that

$$\Big\| \sup_{k\ge\kappa_S} |f * K_{k,S,\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'}}| \Big\|_{\ell^p(\mathbb{G}_0)} \lesssim \|(\mathcal{W}_{k,S,Q_S})_{k\ge\kappa_S}\|_{\ell^p(\mathbb{H}_{Q_S})\to\ell^p(\mathbb{H}_{Q_S};\ell^\infty)}$$
$$\times \|\mathcal{V}_{\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'},Q_S}\|_{\ell^p(\mathbb{J}_{Q_S})\to\ell^p(\mathbb{J}_{Q_S})}\|f\|_{\ell^p(\mathbb{G}_0)} + 2^{-\kappa_S/8}\|f\|_{\ell^p(\mathbb{G}_0)}. \tag{9.7}$$

From Proposition 9.4 we know that

$$\|(\mathcal{W}_{k,S,Q_S})_{k\ge\kappa_S}\|_{\ell^p(\mathbb{H}_{Q_S})\to\ell^p(\mathbb{H}_{Q_S};\ell^\infty)} \lesssim S. \tag{9.8}$$

We also know that

$$\|\mathcal{V}_{\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'},Q_S}\|_{\ell^p(\mathbb{J}_{Q_S})\to\ell^p(\mathbb{J}_{Q_S})} \le \big\|V_{\widetilde{\mathcal{R}}_{Q_S}^d,\widetilde{\mathcal{R}}_{Q_S}^{d'},Q_S}\big\|_{\ell^1(\mathbb{J}_{Q_S})} \lesssim 1, \tag{9.9}$$

which follows from the direct computation

$$\begin{aligned} &V_{\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S},Q_S}(b)\\ &= Q_S^{-d-d'}\Big\{\sum_{a\in\mathbb{Z}^d_{Q_S}} S(a/Q_S)\mathfrak{e}(b^{(1)}.a/Q_S)\Big\}\Big\{\sum_{c\in\mathbb{Z}^{d'}_{Q_S}} \mathfrak{e}(b^{(2)}.c/Q_S)\Big\}\\ &= Q_S^{-1}\sum_{n\in\mathbb{Z}_{Q_S}} \mathbb{1}_{\{A_0(n)\}}(b). \end{aligned}$$

The bounds (9.5) follow from (9.7)–(9.9).

Step 3. Finally, we prove the bounds (9.6). Observe that for $k \geq \kappa_S$ we have the following decomposition, with the notation in Sect. 4,

$$\begin{aligned} \mathcal{M}_k f - f * K_{k,S,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}} &= \mathcal{M}_k f - f * \Big[\sum_{s\in[0,\delta k]} K_{k,s}\Big]\\ &\quad + f * \Big[\sum_{s\in(\delta S,\delta k]} K_{k,s}\Big] + f * \Big[\sum_{s\in[0,\delta S]} G^c_{k,s}\Big]\\ &\quad + f * K_{k,k,\mathcal{R}^d_{\leq\delta' k}\setminus\widetilde{\mathcal{R}}^d_{Q_S},\mathcal{R}^{d'}_{\leq\delta S}} - f * K_{k,k,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}\setminus\mathcal{R}^{d'}_{\leq\delta S}}\\ &\quad + f * \Big[K_{k,k,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}} - K_{k,S,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}}\Big]. \end{aligned}$$

Therefore, to prove (9.6) it is enough to show that for every $\lambda \geq e^D$ and $f \in \ell^2(\mathbb{G}_0)$

$$\Big\| \sup_{k\geq\kappa_S} \Big|\mathcal{M}_k f - f * \Big[\sum_{s\in[0,\delta k]} K_{k,s}\Big]\Big|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-1}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.10}$$

$$\Big\| \sup_{k\geq\kappa_S} \Big|f * \Big[\sum_{s\in(\delta S,\delta k]} K_{k,s}\Big]\Big|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-\delta/D^3}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.11}$$

$$\Big\| \sup_{k\geq\kappa_S} \Big|f * \Big[\sum_{s\in[0,\delta S]} G^c_{k,s}\Big]\Big|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-1}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.12}$$

$$\Big\| \sup_{k\geq\kappa_S} |f * K_{k,k,\mathcal{R}^d_{\leq\delta' k}\setminus\widetilde{\mathcal{R}}^d_{Q_S},\mathcal{R}^{d'}_{\leq\delta S}}|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-1/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.13}$$

$$\Big\| \sup_{k\geq\kappa_S} |f * K_{k,k,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}\setminus\mathcal{R}^{d'}_{\leq\delta S}}|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-\delta/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.14}$$

$$\Big\| \sup_{k\geq\kappa_S} \Big|f * \Big[K_{k,k,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}} - K_{k,S,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}}\Big]\Big|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-\delta/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{9.15}$$

Step 4. We now establish inequalities (9.10)–(9.15). Notice that $\mathcal{M}_k f - f * \Big[\sum_{s\in[0,\delta k]} K_{k,s}\Big] = f * K_k^c$, and the bounds (9.10) follow from Lemma 4.2. Similarly, the bounds (9.12) follow from Lemma 5.1 with $\mathcal{B} = \mathcal{R}^{d'}_{\le \delta S}$. In addition, combining (4.20) with (4.21) we obtain

$$\Big\| \sup_{k\ge\kappa_S} \Big| f * \Big[\sum_{s\in(\delta S,\delta k]} K_{k,s}\Big]\Big|\Big\|_{\ell^2(\mathbb{G}_0)} \le \sum_{s>\delta S} \Big\| \sup_{k\ge\max(\kappa_S,s/\delta)} |f * K_{k,s}|\Big\|_{\ell^2(\mathbb{G}_0)}$$
$$\lesssim \sum_{s>\delta S} 2^{-s/D^2}\|f\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-\delta/D^3}\|f\|_{\ell^2(\mathbb{G}_0)}.$$

This proves (9.11). Moreover, using (6.49) and (8.2) with $\mathcal{A} = \mathcal{R}^d_t \setminus \widetilde{\mathcal{R}}^d_{Q_S}$ and $\mathcal{B} = \mathcal{R}^{d'}_{\le\delta S}$,

$$\Big\| \sup_{k\ge\kappa_S} |f * K_{k,k,\mathcal{R}^d_{\le\delta' k}\setminus\widetilde{\mathcal{R}}^d_{Q_S},\mathcal{R}^{d'}_{\le\delta S}}|\Big\|_{\ell^2(\mathbb{G}_0)}$$
$$\le \sum_{t\ge D(S+1)} \Big\| \sup_{k\ge\max(\kappa_S,t/\delta')} |f * K_{k,k,\mathcal{R}^d_t\setminus\widetilde{\mathcal{R}}^d_{Q_S},\mathcal{R}^{d'}_{\le\delta S}}|\Big\|_{\ell^2(\mathbb{G}_0)}$$
$$\lesssim \sum_{t\ge D(S+1)} 2^{-t/D^2}\|f\|_{\ell^2(\mathbb{G}_0)} \lesssim \lambda^{-1/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}.$$

This completes the proof of (9.13).

We prove now the bounds (9.14). We apply Lemma 6.2 with $Q = Q_S$, $\mathcal{A} = \widetilde{\mathcal{R}}^d_{Q_S}$, $\mathcal{B} = \widetilde{\mathcal{R}}^{d'}_{Q_S} \setminus \mathcal{R}^{d'}_{\le\delta S}$, $k_0 = \lfloor\kappa_S\rfloor$ and $w = k$. Then we apply Lemma 6.6 and conclude from (6.42) that

$$\Big\| \sup_{k\ge\kappa_S} |f * K_{k,k,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}\setminus\mathcal{R}^{d'}_{\le\delta S}}|\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim \|(\mathcal{W}_{k,k,Q_S})_{k\ge\kappa_S}\|_{\ell^2(\mathbb{H}_{Q_S})\to\ell^2(\mathbb{H}_{Q_S};\ell^\infty)}$$
$$\times \big\|\mathcal{V}_{\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}\setminus\mathcal{R}^{d'}_{\le\delta S},Q_S}\big\|_{\ell^2(\mathbb{J}_{Q_S})\to\ell^2(\mathbb{J}_{Q_S})}\|f\|_{\ell^2(\mathbb{G}_0)} + 2^{-\lambda/8}\|f\|_{\ell^2(\mathbb{G}_0)}.$$

By (6.19) we may conclude that

$$\big\|\mathcal{V}_{\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}\setminus\mathcal{R}^{d'}_{\le\delta S},Q_S}\big\|_{\ell^2(\mathbb{J}_{Q_S})\to\ell^2(\mathbb{J}_{Q_S})} \lesssim 2^{-\delta S/D}\|f\|_{\ell^2(\mathbb{J}_{Q_S})}.$$

The bounds (9.14) follow using also Lemma 6.4.

Finally, we prove the bounds (9.15). By a simple square function argument and Khinchine's inequality it suffices to prove that for every $w \ge S$, every sequence $(\varkappa_k)_{k\in\mathbb{N}} \subseteq [-1,1]$ and any $f \in \ell^2(\mathbb{G}_0)$ we have

$$\Big\| \sum_{k\ge\max\{\kappa_S,w+1\}} \varkappa_k f * \big[K_{k,w+1,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}} - K_{k,w,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}}\big]\Big\|_{\ell^2(\mathbb{G}_0)}$$

$$\lesssim 2^{-w/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{9.16}$$

We apply again Lemma 6.2 with $Q = Q_S$, $\mathcal{A} = \widetilde{\mathcal{R}}^d_{Q_S}$, $\mathcal{B} = \widetilde{\mathcal{R}}^{d'}_{Q_S}$ and $w \geq S$. Then we apply Lemma 6.6 with $k_0 = \max\{\kappa_S, w+1\}$, $K_k^{\mathbb{G}_0} = K_{k,w+1,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}} - K_{k,w,\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S}}$, $W_k^{\mathbb{H}_{Q_S}} = W_{k,w+1,Q_S} - W_{k,w,Q_S}$ and $V^{\mathbb{J}_{Q_S}} = V_{\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S},Q_S}$ and conclude from (6.43) that the left-hand side of (9.16) is controlled by

$$\Big\|\sum_{k\geq\max\{\kappa_S,w+1\}} \varkappa_k\big[\mathcal{W}_{k,w+1,Q_S} - \mathcal{W}_{k,w,Q_S}\big]\Big\|_{\ell^2(\mathbb{H}_{Q_S})\to\ell^2(\mathbb{H}_{Q_S})}\|f\|_{\ell^2(\mathbb{G}_0)} + 2^{-w/8}\|f\|_{\ell^2(\mathbb{G}_0)},$$

since $\|\mathcal{V}_{\widetilde{\mathcal{R}}^d_{Q_S},\widetilde{\mathcal{R}}^{d'}_{Q_S},Q_S}\|_{\ell^2(\mathbb{J}_{Q_S})\to\ell^2(\mathbb{J}_{Q_S})} \lesssim 1$ by (9.9). Finally, using (6.30) we obtain

$$\Big\|\sum_{k\geq\max\{\kappa_S,w+1\}} \varkappa_k\big[\mathcal{W}_{k,w+1,Q_S} - \mathcal{W}_{k,w,Q_S}\big]\Big\|_{\ell^2(\mathbb{H}_{Q_S})\to\ell^2(\mathbb{H}_{Q_S})} \lesssim 2^{-w/D^2}$$

as desired and the proof of (9.15) is finished. This also completes the proof of Theorem 9.1.

9.2 Proof of Proposition 9.3

To prove Proposition 9.3 we exploit the positivity of the operator $\mathcal{M}_k f$, i.e., $\mathcal{M}_k f \geq 0$ whenever $f \geq 0$. We will extend the ideas of Bourgain [13, Lemma 7.32] (see also [32, Lemmas 4.2 and 4.3]) to the nilpotent setting. We will need the following technical result, to approximate the original operator.

Lemma 9.5 *For every $\mu \in \mathbb{Z}_+$ there is a constant $C_\mu > 0$ such that for every $f \in \ell^2(\mathbb{G}_0)$ the following inequality*

$$\big\|\mathcal{M}_k f - f * U_{k,J,S,\mu}\big\|_{\ell^2(\mathbb{G}_0)} \leq C_\mu S^{-1/D^2}\|f\|_{\ell^2(\mathbb{G}_0)},$$

holds uniformly in $1 \le J \le k \le 2J$, $1 \le S \le 2^{\delta k}$ *satisfying* $S^D \le 2^{\delta' k}$ *and* $S \le J^{\mu}$. *Here*

$$\begin{aligned} &U_{k,J,S,\mu}(g) \\ &:= \phi_k(g) \sum_{\sigma^{(1)} \in \mathcal{R}^d_{\le D\log_2 S} \cap [0,1)^d} \sum_{\sigma^{(2)} \in \mathcal{R}^{d'}_{\le \log_2 S} \cap [0,1)^{d'}} \mathfrak{e}\big(g.(\sigma^{(1)}, \sigma^{(2)})\big) S(\sigma^{(1)}) \\ &\quad \times \Big\{ \prod_{(l_1,l_2)\in Y_d} 2^{-k(l_1+l_2)} \Big\} \\ &\quad \times \int_{\mathbb{R}^{d+d'}} \eta_{\le \delta' D\mu \lfloor \log_2 J \rfloor}(\xi^{(1)}) \eta_{\le \delta D\mu \lfloor \log_2 J \rfloor}(\xi^{(2)}) J_k(2^{-k} \circ \xi^{(1)}) \\ &\quad \times \mathfrak{e}[(2^{-k} \circ g).\xi]\, d\xi. \end{aligned} \tag{9.17}$$

We show first how to use Lemma 9.5 to prove Proposition 9.3. We proceed in several steps.

Step 1. Since the result is clear for $p = \infty$ it suffices to consider only $p \in (1, 2]$ and nonnegative functions $f : \mathbb{G}_0 \to [0, \infty)$. Let $\widetilde{K}_j(x) = K_j(x^{-1})$. By a general abstract argument, involving duality and a separation in scales j (see [13] and [32, Lemma 4.2]), it suffices to show that

$$\Big\| \sum_{j\in\mathcal{S}} h_j * \widetilde{K}_j \Big\|_{\ell^R(\mathbb{G}_0)} \lesssim_R |F|^{1/R}, \tag{9.18}$$

for any even integer $R \ge 2$, any subset $F \subseteq \mathbb{G}_0$, any functions h_j satisfying

$$h_j = g_j \mathbb{1}_F, \quad g_j : \mathbb{G}_0 \to [0, 1], \quad \sum_{j\in\mathcal{S}} g_j(x) \le 1 \text{ for any } x \in \mathbb{G}_0, \tag{9.19}$$

and any subset $\mathcal{S} \subseteq [J + 1, 2J]$ satisfying the sparseness property $|l - l'| \ge D\mu \log_2 J$ if $l \ne l' \in \mathcal{S}$. Here $\mu = \mu(R)$ is a sufficiently large constant to be determined later (in (9.22)).

Indeed, by a duality argument there are functions $0 \le g_j \le 1$ for $J < j \le 2J$, such that $\sum_{J<j\le 2J} g_j(x) = 1$, $x \in \mathbb{G}_0$, and

$$\sup_{J<j\le 2J} |f * K_j(x)| = \sum_{J<j\le 2J} f * K_j(x) g_j(x), \quad x \in \mathbb{G}_0, \quad J \ge 1.$$

Then, we have

$$\Big\| \sup_{J<j\le 2J} |f * K_j| \Big\|_{\ell^p(\mathbb{G}_0)} = \Big\| \sum_{J<j\le 2J} (f * K_j) g_j \Big\|_{\ell^p(\mathbb{G}_0)}$$
$$\le \sup_{\|h\|_{\ell^{p'}(\mathbb{G}_0)}\le 1} \Big\| \sum_{J<j\le 2J} (hg_j) * \widetilde{K}_j \Big\|_{\ell^{p'}(\mathbb{G}_0)} \|f\|_{\ell^p(\mathbb{G}_0)}.$$

Using interpolation it suffices to show that the latter operator is of restricted weak type (R, R) for any integer $R \ge 2$, with norm $\lesssim_R \log(J+2)$. This means that we need to show that for every fixed integer $R \ge 2$, every finite subset $F \subseteq \mathbb{G}_0$ and every $J \ge 1$ we have

$$\Big\| \sum_{J<j\le 2J} h_j * \widetilde{K}_j \Big\|_{\ell^R(\mathbb{G}_0)} \lesssim_R \log(J+2)|F|^{1/R},$$

where $h_j = g_j \mathbb{1}_F$ for every $J < j \le 2J$. Finally, we partition the set $(J, 2J]$ into at most $D\mu \log_2 J + 1$ subsets $\mathcal{S}$ with the sparseness property mentioned above. Therefore, we reduced our task to showing (9.18). We prove (9.18) by induction over R. The case $R = 2$ follows from the $\ell^2(\mathbb{G}_0)$ boundedness of the maximal function $\sup_{j\ge 0} |\mathcal{M}_j|$. The case of general R can be reduced to proving that

$$\Big\| \Big(\prod_{n=2}^{R} h_{j_n} * \widetilde{K}_{j_n} \Big) * (K_{j_1} - K_{j_0}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim_R J^{-R} |F|^{1/2}, \tag{9.20}$$

uniformly in $J = j_0 < j_1 < \cdots < j_R \le 2J$ satisfying

$$j_{n+1} - j_n \ge D\mu \log_2 J, \quad 1 \le n \le R-1. \tag{9.21}$$

See [32, Lemma 4.2] for the details of this reduction, which apply in our case as well.

Step 2. To prove (9.20) we first define some constants

$$A := D^4 + R, \quad \mu := D^2 A^R + R, \quad S_n := J^{A^n}, \quad 1 \le n \le R. \tag{9.22}$$

We may assume that $J \gtrsim_\mu 1$, so $1 \le S_n \le 2^{\delta J/2}$, $S_n^D \le 2^{\delta' J/2}$ and $S_n \le J^\mu$, $1 \le n \le R$. For simplicity of notation, in the rest of this subsection the implicit constants are allowed to depend on R. Using Lemma 9.5 we obtain for every $f \in \ell^2(\mathbb{G}_0)$ that

$$\big\| f * \widetilde{K}_{j_n} - f * \widetilde{U}_{n,J,\mu} \big\|_{\ell^2(\mathbb{G}_0)} \lesssim S_n^{-1/D^2} \|f\|_{\ell^2(\mathbb{G}_0)}, \quad 1 \le n \le R, \quad J \ge J_0, \tag{9.23}$$

where $\widetilde{U}_{n,J,\mu}(x) := U_{j_n,J,S_n,\mu}(x^{-1})$, see (9.17). Here we use the fact that if $Tf = f * K$ and $\widetilde{T}f = f * \widetilde{K}$, then $\|T\|_{\ell^2(\mathbb{G}_0)\to\ell^2(\mathbb{G}_0)} = \|\widetilde{T}\|_{\ell^2(\mathbb{G}_0)\to\ell^2(\mathbb{G}_0)}$.

We show that

$$\Big\| \prod_{n=2}^{R} h_{j_n} * \widetilde{K}_{j_n} - \prod_{n=2}^{R} h_{j_n} * \widetilde{U}_{n,J,\mu} \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim J^{-R} |F|^{1/2}, \tag{9.24}$$

uniformly in $J = j_0 < j_1 < \cdots < j_R \le 2J$ satisfying (9.21). Indeed, notice that

$$\|h_{j_n} * \widetilde{U}_{n,J,\mu}\|_{\ell^\infty(\mathbb{G}_0)} \le \|\widetilde{U}_{n,J,\mu}\|_{\ell^1(\mathbb{G}_0)} \|h_{j_n}\|_{\ell^\infty(\mathbb{G}_0)} \lesssim S_n^{2D(d+d')}, \quad 1 \le n \le R. \tag{9.25}$$

Since $U_{n,J,\mu} = U_{j_n,J,S_n,\mu}$, see (9.17), this follows from the identity

$$\begin{aligned} U_{n,J,\mu}(g) = \phi_{j_n}(g) \sum_{\sigma^{(1)} \in \mathcal{R}^d_{\le D\log_2 S_n} \cap [0,1)^d} \; \sum_{\sigma^{(2)} \in \mathcal{R}^{d'}_{\le \log_2 S_n} \cap [0,1)^{d'}} \\ \mathfrak{e}\big(g.(\sigma^{(1)}, \sigma^{(2)})\big) S(\sigma^{(1)}) \\ \times \int_{\mathbb{R}} \chi(u) \Big\{ \prod_{(l_1,l_2)\in Y_d} 2^{-j_n(l_1+l_2)} \Big\} \widehat{\eta_{\le \delta' D\mu \lfloor \log_2 J \rfloor}}\big(A_0^{(1)}(u) - 2^{-j_n} \circ g^{(1)}\big) \\ \times \widehat{\eta_{\le \delta D\mu \lfloor \log_2 J \rfloor}}\big(- 2^{-j_n} \circ g^{(2)}\big)\, du, \end{aligned} \tag{9.26}$$

see also (6.4). Using (9.23) and (9.25) we can estimate the left-hand side of (9.24) by

$$\begin{aligned} C \sum_{n=2}^{R} \Big(\prod_{k=2}^{n-1} \|h_{j_k} * \widetilde{U}_{k,J,\mu}\|_{\ell^\infty(\mathbb{G}_0)} \Big) \Big(\prod_{k=n+1}^{R} \|h_{j_k} * \widetilde{K}_{j_k}\|_{\ell^\infty(\mathbb{G}_0)} \Big) \\ \times \big\| h_{j_n} * \widetilde{K}_{j_n} - h_{j_n} * \widetilde{U}_{n,J,\mu} \big\|_{\ell^2(\mathbb{G}_0)} \\ \lesssim \sum_{n=2}^{R} \Big(\prod_{k=2}^{n-1} S_k^{2D(d+d')} \Big) S_n^{-1/D^2} |F|^{1/2} \\ \lesssim \sum_{n=2}^{R} J^{4D(d+d')A^{n-1} - A^n D^{-2}} |F|^{1/2} \lesssim J^{-R} |F|^{1/2}, \end{aligned}$$

since $4D(d+d')A^{n-1} - A^n D^{-2} \le -A^{n-1} \le -R$, see (9.22). The bounds (9.24) follow.

Step 3. In view of (9.24), for (9.20) it is enough to prove that

$$\Big\| \Big(\prod_{n=2}^{R} h_{j_n} * \widetilde{U}_{n,J,\mu} \Big) * (K_{j_1} - K_{j_0}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim J^{-R} |F|^{1/2}, \tag{9.27}$$

uniformly in $J = j_0 < j_1 < \cdots < j_R \le 2J$ satisfying (9.21). Let us define

$$\begin{aligned} X_{n,J,\mu}(g) := \phi_{j_n}(g) \int_{\mathbb{R}} \chi(u) \Big\{ \prod_{(l_1,l_2)\in Y_d} 2^{-j_n(l_1+l_2)} \Big\} \\ \widehat{\eta_{\le \delta' D\mu \lfloor \log_2 J \rfloor}}\big(A_0^{(1)}(u) - 2^{-j_n} \circ g^{(1)}\big) \\ \times \widehat{\eta_{\le \delta D\mu \lfloor \log_2 J \rfloor}}\big(- 2^{-j_n} \circ g^{(2)}\big)\, du, \end{aligned} \tag{9.28}$$

$$X_{n,J,\mu,\sigma}(g) := X_{n,J,\mu}(g) \mathfrak{e}(g.\sigma). \tag{9.29}$$

Using (9.26) we have

$$h_{j_n} * \widetilde{U}_{n,J,\mu} = \sum_{\sigma_n^{(1)} \in \mathcal{R}^d_{\le D \log_2 S_n} \cap [0,1)^d,\, \sigma_n^{(2)} \in \mathcal{R}^{d'}_{\le \log_2 S_n} \cap [0,1)^{d'}} S(\sigma_n^{(1)}) \cdot h_{j_n} * \widetilde{X}_{n,J,\mu,\sigma_n}.$$

In view of (9.22), for (9.27) it suffices to show that

$$\Big\| \Big(\prod_{n=2}^{R} h_{j_n} * \widetilde{X}_{n,J,\mu,\sigma_n} \Big) * (K_{j_1} - K_{j_0}) \Big\|_{\ell^2(\mathbb{G}_0)} \lesssim J^{-2\mu} |F|^{1/2}, \tag{9.30}$$

for any $\sigma_n^{(1)} \in \mathcal{R}^d_{\le D \log_2 S_n} \cap [0,1)^d$, $\sigma_n^{(2)} \in \mathcal{R}^{d'}_{\le \log_2 S_n} \cap [0,1)^{d'}$, $2 \le n \le R$.

Observe that

$$f * (K_{j_1} - K_{j_0})(g) = \sum_{u \in \mathbb{Z}} \chi_{j_0,j_1}(u) f(A_0(u)^{-1} \cdot g),$$

where $\chi_{j_0,j_1}(v) = 2^{-j_1}\chi(2^{-j_1}v) - 2^{-j_0}\chi(2^{-j_0}v)$. Notice that

$$\Big| \sum_{v \in \mathbb{Z}} \chi_{j_0,j_1}(Qv + b) \Big| \lesssim 2^{-j_0}, \quad Q \in \mathbb{Z}_+, \quad b \in \mathbb{Z}_Q. \tag{9.31}$$

Therefore we have

$$\Big(\prod_{n=2}^{R} h_{j_n} * \widetilde{X}_{n,J,\mu,\sigma_n}\Big) * (K_{j_1} - K_{j_0})(g)$$
$$= \sum_{v\in\mathbb{Z}} \chi_{j_0,j_1}(v) \sum_{y_2,\dots,y_R\in\mathbb{G}_0} \Big(\prod_{n=2}^{R} X_{n,J,\mu,\sigma_n}\big(y_n\cdot g^{-1}\cdot A_0(v)\big)h_{j_n}(y_n)\Big)$$
$$= \sum_{y_2,\dots,y_R\in\mathbb{G}_0} \Big(\prod_{n=2}^{R} h_{j_n}(y_n)\Big) H(y_2\cdot g^{-1},\dots,y_R\cdot g^{-1}), \tag{9.32}$$

where

$$H(y_2,\dots,y_R) := \sum_{v\in\mathbb{Z}} \chi_{j_0,j_1}(v)\Big(\prod_{n=2}^{R} X_{n,J,\mu,\sigma_n}\big(y_n\cdot A_0(v)\big)\Big). \tag{9.33}$$

For (9.30) it suffices to show that there are functions $H_n = H_{n,J,\mu} \geq 0$, $2\leq n\leq R$, such that

$$\|H_n\|_{\ell^1(\mathbb{G}_0)} \lesssim 1 \text{ for } 2\leq n\leq R \quad \text{and} \quad |H(y_2,\dots,y_R)| \lesssim J^{-2\mu}\prod_{n=2}^{R} H_n(y_n). \tag{9.34}$$

Indeed, assuming (9.34) and using (9.32) we can bound the left-hand side of (9.30) by

$$CJ^{-2\mu}\Big\|\prod_{n=2}^{R} h_{j_n} * \widetilde{H}_n\Big\|_{\ell^2(\mathbb{G}_0)} \leq CJ^{-2\mu}\prod_{n=2}^{R}\|h_{j_n} * \widetilde{H}_n\|_{\ell^{2(R-1)}(\mathbb{G}_0)} \lesssim J^{-2\mu}|F|^{1/2}.$$

Step 4. It remains to prove (9.34). Let q_n be the denominator of σ_n. By (9.22) one has

$$Q := \prod_{n=2}^{R} q_n \lesssim \prod_{n=2}^{R} S_n^{2Dd'} \leq J^{\mu}. \tag{9.35}$$

Splitting the summation in v in (9.33) into classes modulo Q and using (9.29) we obtain

$$|H(y_2,\dots,y_R)| \leq \sum_{b\in\mathbb{Z}_Q}\Big|\sum_{v\in\mathbb{Z}}\chi_{j_0,j_1}(Qv+b)\Big(\prod_{n=2}^{R} X_{n,J,\mu}\big(y_n\cdot A_0(Qv+b)\big)\Big)\Big|$$
$$\lesssim \sum_{b\in\mathbb{Z}_Q}\Big|\sum_{v\in\mathbb{Z}}\chi_{j_0,j_1}(Qv+b)\Big(\prod_{n=2}^{R} X_{n,J,\mu}\big(y_n\cdot A_0(Qv+b)\big) - \prod_{n=2}^{R} X_{n,J,\mu}(y_n)\Big)\Big|$$

$$+\sum_{b\in\mathbb{Z}_Q}\Big|\sum_{v\in\mathbb{Z}}\chi_{j_0,j_1}(Qv+b)\Big(\prod_{n=2}^{R}X_{n,J,\mu}(y_n)\Big)\Big|=:I_1+I_2. \tag{9.36}$$

Using the definition (9.28) it is easy to see that for every $y\in\mathbb{G}_0^{\#}$ and $2\le n\le R$ one has

$$\begin{aligned}&|X_{n,J,\mu}(y)|+\sum_{(l_1,l_2)\in Y_d}2^{j_n(l_1+l_2)}J^{-\delta_{l_1l_2}D\mu}\big|(\partial_{y_{l_1l_2}}X_{n,J,\mu})(y)\big|\\&\lesssim\Big\{\prod_{(l_1,l_2)\in Y_d}2^{-j_n(l_1+l_2)}J^{\delta_{l_1l_2}D\mu}\Big\}\int_{\mathbb{R}}\chi(u)\Big\langle J^{\widetilde{\delta}D\mu}\big(A_0(u)-2^{-j_n}\circ y\big)\Big\rangle^{-2D}du,\end{aligned} \tag{9.37}$$

where $\widetilde{\delta}=(\delta_{l_1l_2})_{(l_1,l_2)\in Y_d}$ and $\delta_{l_1l_2}=\delta$ if $(l_1,l_2)\in Y_d'$ and $\delta_{l_1l_2}=\delta'$ otherwise. Since $2^{j_1-j_n}\lesssim J^{-D\mu}$ (the separation condition (9.21)), for every $y,h\in\mathbb{G}_0$ satisfying $|2^{-j_1}\circ h|\lesssim 1$ we have

$$\begin{aligned}&|X_{n,J,\mu}(y\cdot h)-X_{n,J,\mu}(y)|\\&\lesssim J^{-3\mu}\Big\{\prod_{(l_1,l_2)\in Y_d}2^{-j_n(l_1+l_2)}J^{\delta_{l_1l_2}D\mu}\Big\}\\&\quad\times\int_{\mathbb{R}}\chi(u)\Big\langle J^{\widetilde{\delta}D\mu}\big(A_0(u)-2^{-j_n}\circ y\big)\Big\rangle^{-D}du.\end{aligned} \tag{9.38}$$

Using (9.37)–(9.38) if $|Qv+b|\lesssim 2^{j_1}$ then we have

$$\begin{aligned}&\Big|\prod_{n=2}^{R}X_{n,J,\mu}\big(y_n\cdot A_0(Qv+b)\big)-\prod_{n=2}^{R}X_{n,J,\mu}(y_n)\Big|\\&\lesssim J^{-3\mu}\prod_{n=2}^{R}\Big(\int_{\mathbb{R}}\chi(u_n)\Big\{\prod_{(l_1,l_2)\in Y_d}2^{-j_n(l_1+l_2)}J^{\delta_{l_1l_2}D\mu}\Big\}\\&\quad\times\Big\langle J^{\widetilde{\delta}D\mu}\big(A_0(u_n)-2^{-j_n}\circ y_n\big)\Big\rangle^{-D}du_n\Big).\end{aligned}$$

Since $\sum_{b\in\mathbb{Z}_Q}\sum_{v\in\mathbb{Z}}|\chi_{j_0,j_1}(Qv+b)|\lesssim 1$, we see that the required decomposition (9.34) for the first term I_1 in (9.36) follows. The decomposition for I_2 also follows using (9.31), (9.35) and (9.37). This completes the proof of Proposition 9.3.

9.3 Proof of Lemma 9.5

Observe that we may assume that $k \geq D^2\mu$, otherwise the conclusion is trivial. Observe that we have a decomposition

$$\begin{aligned}\mathcal{M}_k f - f * U_{k,J,S,\mu} = \mathcal{M}_k f - f * \Big[\sum_{s\in[0,\delta k]} K_{k,s}\Big]\\ + f * \Big[\sum_{s\in(\log_2 S,\delta k]} K_{k,s}\Big] + f * \Big[\sum_{s\in[0,\log_2 S]} G^c_{k,s}\Big]\\ + f * K_{k,k,\mathcal{R}^d_{\leq\delta' k}\setminus\mathcal{R}^d_{\leq D\log_2 S},\mathcal{R}^{d'}_{\leq\log_2 S}}\\ + f * \Big[K_{k,k,\mathcal{R}^d_{\leq D\log_2 S},\mathcal{R}^{d'}_{\leq\log_2 S}} - U_{k,J,S,\mu}\Big].\end{aligned}$$

To prove Lemma 9.5 it remains to show that for any $f \in \ell^2(\mathbb{G}_0)$, $k \geq D^2\mu$, $J \leq k \leq 2J$, and $S \leq J^\mu$ we have the following estimates:

$$\Big\|\mathcal{M}_k f - f * \Big[\sum_{s\in[0,\delta k]} K_{k,s}\Big]\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-k/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.39}$$

$$\Big\|f * \Big[\sum_{s\in(\log_2 S,\delta k]} K_{k,s}\Big]\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim S^{-1/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.40}$$

$$\big\|f * G^c_{k,\mathcal{R}^{d'}_{\leq\log_2 S}}\big\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-k/D^2}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.41}$$

$$\big\|f * K_{k,k,\mathcal{R}^d_{\leq\delta' k}\setminus\mathcal{R}^d_{\leq D\log_2 S},\mathcal{R}^{d'}_{\leq\log_2 S}}\big\|_{\ell^2(\mathbb{G}_0)} \lesssim S^{-1/D}\|f\|_{\ell^2(\mathbb{G}_0)}, \tag{9.42}$$

$$\Big\|f * \Big[K_{k,k,\mathcal{R}^d_{\leq D\log_2 S},\mathcal{R}^{d'}_{\leq\log_2 S}} - U_{k,J,S,\mu}\Big]\Big\|_{\ell^2(\mathbb{G}_0)} \lesssim S^{-1}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{9.43}$$

Here and in the rest of this subsection the implicit constants are allowed to depend on μ. The bounds (9.39) follow from Lemma 4.2. The bounds (9.40) follow from (4.20)–(4.21). The bounds (9.41) follow from Lemma 5.1 with $\mathcal{B} = \mathcal{R}^{d'}_{\leq\log_2 S}$.

To prove the bounds (9.42) we use Lemma 8.2 with $\iota = 0$, so we have the decomposition

$$\begin{aligned}\{(\mathcal{K}_{k,k,\mathcal{R}^d_p\setminus\mathcal{R}^d_{\leq D\log_2 S},\mathcal{R}^{d'}_{\leq\log_2 S}})^*\mathcal{K}_{k,k,\mathcal{R}^d_p\setminus\mathcal{R}^d_{\leq D\log_2 S},\mathcal{R}^{d'}_{\leq\log_2 S}}\}^r f\\ = f * \{F^{0,r}_k + O^{0,r}_k\},\end{aligned} \tag{9.44}$$

for any $p \in (D\log_2 S, \delta' k]$. Here $\|O_k^{0,r}\|_{\ell^1(\mathbb{G}_0)} \lesssim 2^{-k}$, $\mathcal{A} = \mathcal{R}_p^d \setminus \mathcal{R}_{\le D\log_2 S}^d$, $\mathcal{B} = \mathcal{R}_{\le \log_2 S}^{d'}$, and

$$\begin{aligned} F_k^{0,r}(h) := \Big\{ \sum_{a^{(2)}/Q \in \mathcal{B}\cap[0,1)^{d'}} \sum_{\sigma \in [\mathcal{A}+(\mathbb{Z}_Q/Q)^d]\cap[0,1)^d} & \mathcal{C}(a^{(2)}/Q,\sigma) \\ & \times \mathfrak{e}(h^{(1)}.\sigma)\mathfrak{e}\big(h^{(2)}.(a^{(2)}/Q)\big)\Big\} \\ \times \Big\{ \prod_{(l_1,l_2)\in Y_d} 2^{-k(l_1+l_2)} \Big\} & \eta_{\le 3\delta k}(2^{-k}\circ h) \\ \times \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} & \eta_{\le \delta k/2}(\zeta^{(1)})\eta_{\le \delta k/2}(\zeta^{(2)})P(\zeta)\mathfrak{e}[(2^{-k}\circ h).\zeta]\,d\zeta. \end{aligned}$$

The function P was defined in (7.12), and the coefficients $\mathcal{C}$ satisfy the bounds

$$|\mathcal{C}(a^{(2)}/Q,\sigma)| \lesssim 2^{-p/\delta},$$

for any $a^{(2)}/Q \in \mathcal{R}_{\le \log_2 S}^{d'} \cap [0,1)^{d'}$ and $\sigma \in [\mathcal{R}_p^d \setminus \mathcal{R}_{\le D\log_2 S}^d + (\mathbb{Z}_Q/Q)^d] \cap [0,1)^d$. Using this estimate and (7.27) (with $\iota = 0$), we see that $\|F_k^{0,r}\|_{\ell^1(\mathbb{G}_0)} \lesssim 2^{-p/(2\delta)}$. The desired bounds (9.42) follow by summation over $p \ge D\log_2 S$.

Finally, to prove the bounds (9.43) we use first Lemma 6.1 to see that

$$\|K_{k,D\mu\lfloor\log_2 J\rfloor,\mathcal{R}_{\le D\log_2 S}^d,\mathcal{R}_{\le\log_2 S}^{d'}} - U_{k,J,S,\mu}\|_{\ell^1(\mathbb{G}_0)} \lesssim 2^{-k}.$$

Therefore it remains to establish the following:

Lemma 9.6 *Assume $\mu \ge 1$, $k \ge D^2\mu$, $J \le k \le 2J$, and $S \le J^\mu$. Then for any $f \in \ell^2(\mathbb{G}_0)$,*

$$\begin{aligned} \Big\| f * \Big[K_{k,k,\mathcal{R}_{\le D\log_2 S}^d,\mathcal{R}_{\le\log_2 S}^{d'}} - K_{k,D\mu\lfloor\log_2 J\rfloor,\mathcal{R}_{\le D\log_2 S}^d,\mathcal{R}_{\le\log_2 S}^{d'}} \Big] \Big\|_{\ell^2(\mathbb{G}_0)} \\ \lesssim S^{-1}\|f\|_{\ell^2(\mathbb{G}_0)}. \end{aligned} \tag{9.45}$$

Proof For $w \in \mathbb{N}$ and $\mathcal{I} \subseteq \{1,2\}$ we define the auxiliary functions

$$\begin{aligned} \Upsilon_{w,\mathcal{I}}^{(1)} &= \begin{cases} \eta_{\le\delta'(w+1)} - \eta_{\le\delta' w} & \text{if } 1 \in \mathcal{I}, \\ \eta_{\le\delta' w} & \text{if } 1 \notin \mathcal{I}, \end{cases} \\ \Upsilon_{w,\mathcal{I}}^{(2)} &= \begin{cases} \eta_{\le\delta(w+1)} - \eta_{\le\delta w} & \text{if } 2 \in \mathcal{I}, \\ \eta_{\le\delta w} & \text{if } 2 \notin \mathcal{I}. \end{cases} \end{aligned} \tag{9.46}$$

Then we define the projections $\Psi_{k,w,\mathcal{A},\mathcal{I}}$ and $\Xi_{k,w,\mathcal{B},\mathcal{I}}$ as in (4.11),

$$\Psi_{k,w,\mathcal{A},\mathcal{I}}(\xi^{(1)}) := \sum_{a/q\in\mathcal{A}} \Upsilon^{(1)}_{w,\mathcal{I}}(\tau^k \circ (\xi^{(1)} - a/q)),$$
$$\Xi_{k,w,\mathcal{B},\mathcal{I}}(\xi^{(2)}) := \sum_{b/q\in\mathcal{B}} \Upsilon^{(1)}_{w,\mathcal{I}}(\tau^k \circ (\xi^{(2)} - b/q)),$$

where $\mathcal{A}\subseteq\mathbb{Q}^d$ and $\mathcal{B}\subseteq\mathbb{Q}^{d'}$ are 1-periodic sets. Then we define the associated kernels

$$K_{k,w,\mathcal{A},\mathcal{B},\mathcal{I}}(g) = L_{k,w,\mathcal{A},\mathcal{I}}(g^{(1)})N_{k,w,\mathcal{B},\mathcal{I}}(g^{(2)}),$$
$$L_{k,w,\mathcal{A},\mathcal{I}}(g^{(1)}) := \phi^{(1)}_k(g^{(1)}) \int_{\mathbb{T}^d} \mathfrak{e}(g^{(1)}.\xi^{(1)})\Psi_{k,w,\mathcal{A},\mathcal{I}}(\xi^{(1)})S_k(\xi^{(1)})\,d\xi^{(1)},$$
$$N_{k,w,\mathcal{B},\mathcal{I}}(g^{(2)}) := \phi^{(2)}_k(g^{(2)}) \int_{\mathbb{T}^{d'}} \mathfrak{e}(g^{(2)}.\xi^{(2)})\Xi_{k,w,\mathcal{B},\mathcal{I}}(\xi^{(2)})\,d\xi^{(2)}.$$

Let $w_0 := D\mu\lfloor\log_2 J\rfloor$ and observe that

$$\begin{aligned}
&K_{k,k,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S}} - K_{k,w_0,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S}}\\
&\quad= \sum_{w=w_0}^{k-1} (K_{k,w+1,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S}} - K_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S}})\\
&\quad= \sum_{w=w_0}^{k-1} \sum_{\emptyset\neq\mathcal{I}\subseteq\{1,2\}} K_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S},\mathcal{I}}.
\end{aligned}$$

Therefore (9.45) is reduced to prove that for any $w\in[w_0,k-1]$ and $\mathcal{I}\neq\emptyset$

$$\|f * K_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S},\mathcal{I}}\|_{\ell^2(\mathbb{G}_0)} \lesssim 2^{-w/D}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{9.47}$$

We examine the definition of the kernels $K_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S},\mathcal{I}}$ and notice that we can replace the cutoff function ϕ_k by the cutoff function

$$\phi_{k,0}(g) = \eta_{\le D}(2^{-k}\circ g^{(1)})\eta_{\le D}(2^{-k}\circ g^{(2)}).$$

Indeed, letting $K_{k,w,S,\mathcal{I}}$ denote the corresponding kernel we have

$$\begin{aligned}
&K_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S},\mathcal{I}}(g) - K_{k,w,S,\mathcal{I}}(g)\\
&\quad= \big(\phi_k(g) - \phi_{k,0}(g)\big) \sum_{\sigma\in\mathcal{R}^d_{\le D\log_2 S}\cap[0,1)^d\times\mathcal{R}^{d'}_{\le\log_2 S}\cap[0,1)^{d'}}
\end{aligned}$$

$$\times\, \mathfrak{e}(g.\sigma)2^{-k}\sum_{n\in\mathbb{Z}}\mathfrak{e}\big(-A_0(n).\sigma\big)\chi(2^{-k}n)\Big\{\prod_{(l_1,l_2)\in Y_d}2^{-k(l_1+l_2)}\Big\}$$
$$\times\, \widehat{\Upsilon^{(1)}_{w,\mathcal{I}}}(A^{(1)}_0(2^{-k}n)-2^{-k}\circ g^{(1)})\widehat{\Upsilon^{(2)}_{w,\mathcal{I}}}(-2^{-k}\circ g^{(2)}),$$

which shows that

$$\|K_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{R}^{d'}_{\le\log_2 S},\mathcal{I}}-K_{k,w,S,\mathcal{I}}\|_{\ell^1(\mathbb{G}_0)}\lesssim S^{D^2}2^{-D^2w}\lesssim 2^{-w}.$$

To bound the operators defined by the kernels $K_{k,w,S,\mathcal{I}}$ we use again a high order T^*T argument, so it suffices to prove that

$$\|\{(\mathcal{K}_{k,w,S,\mathcal{I}})^*\mathcal{K}_{k,w,S,\mathcal{I}}\}^r f\|_{\ell^2(\mathbb{G}_0)}\lesssim 2^{-w}\|f\|_{\ell^2(\mathbb{G}_0)}. \tag{9.48}$$

The proof of (9.48) proceeds along the same lines as the proof of Lemma 8.2. However, there are some subtle differences arising from the fact that we can only hope for a rapid decay with respect to w, which might be much smaller than k. In particular, this is the reason why we had to replace the function ϕ_k by $\phi_{k,0}$. For the convenience of the reader we shall provide the details.

In view of (2.13)–(2.16) we have

$$\{(\mathcal{K}_{k,w,S,\mathcal{I}})^*\mathcal{K}_{k,w,S,\mathcal{I}}\}^r f=f*K^r_{k,w,S,\mathcal{I}},$$

where

$$K^r_{k,w,S,\mathcal{I}}(y):=\mathbb{1}_{|2^{-k}\circ y|\lesssim 1}\int_{\mathbb{T}^d\times\mathbb{T}^{d'}}\mathfrak{e}\big(y.\theta\big)\Pi^r_{k,w,S,\mathcal{I}}\big(\theta\big)\Omega^r_{k,w,S,\mathcal{I}}\big(\theta^{(2)}\big)\,d\theta^{(1)}d\theta^{(2)},$$

and the multipliers $\Pi^r_{k,w,S,\mathcal{I}}$ and $\Omega^r_{k,w,S,\mathcal{I}}$ are given by

$$\Pi^r_{k,w,S,\mathcal{I}}\big(\theta\big):=\sum_{h^{(1)}_j,g^{(1)}_j\in\mathbb{Z}^d}\Big\{\prod_{j=1}^r\overline{L_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{I},0}(h^{(1)}_j)}L_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{I},0}(g^{(1)}_j)\Big\}$$
$$\times\,\mathfrak{e}\big(\theta^{(1)}.\sum_{1\le j\le r}(h^{(1)}_j-g^{(1)}_j)\big)$$
$$\times\,\mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le j\le r}R_0(h^{(1)}_j,h^{(1)}_j-g^{(1)}_j)$$
$$+\sum_{1\le l<j\le r}R_0(-h^{(1)}_l+g^{(1)}_l,-h^{(1)}_j+g^{(1)}_j)\Big\}\Big),$$

where $L_{k,w,\mathcal{A},\mathcal{I},0}$ is defined as $L_{k,w,\mathcal{A},\mathcal{I}}$ except that we replace $\phi^{(1)}_k$ by $\phi^{(1)}_{k,0}$. With $F_{k,0}$ defined in a similar way as in (5.3) (we replace $\eta_{\le\delta k}$ by $\eta_{\le D}$) we have

$$\Omega^r_{k,w,S,\mathcal{I}}\big(\theta^{(2)}\big) := \Big| \int_{\mathbb{T}^{d'}} F_{k,0}(\theta^{(2)} - \xi^{(2)})\, \Xi_{k,w,\mathcal{R}^{d'}_{\le \log_2 S},\mathcal{I}}(\xi^{(2)})\, d\xi^{(2)} \Big|^{2r}.$$

We first analyze the kernel $\Omega^r_{k,w,S,\mathcal{I}}$. Note that

$$\begin{aligned}
&\int_{\mathbb{T}^{d'}} F_{k,0}(\theta^{(2)} - \xi^{(2)})\, \Xi_{k,w,\mathcal{R}^{d'}_{\le \log_2 S},\mathcal{I}}(\xi^{(2)})\, d\xi^{(2)} \\
&= \sum_{a^{(2)}/Q \in \mathcal{R}^{d'}_{\le \log_2 S} \cap [0,1)^{d'}} \sum_{g^{(2)} \in \mathbb{Z}^{d'}} \eta_{\le D}(2^{-k} \circ g^{(2)})\, \mathfrak{e}\big(-g^{(2)}.(\theta^{(2)} - a^{(2)}/Q)\big) \\
&\quad \times \Big\{ \prod_{(l_1,l_2)\in Y'_d} 2^{-k(l_1+l_2)} \Big\} \widehat{\Upsilon^{(2)}_{w,\mathcal{I}}}(-2^{-k} \circ g^{(2)}).
\end{aligned}$$

Notice that we may replace the factor $\eta_{\le D}(2^{-k} \circ g^{(2)})$ by 1 above, at the expence of ℓ^1 error term $O(S^{2rD^2} 2^{-D^2 w}) \lesssim 2^{-w}$ (here we have used the fact that integration with respect to θ produces a delta and trivializes summation in y). After this replacement we can use the Poisson summation formula and we end up with

$$\sum_{a^{(2)}/Q \in \mathcal{R}^{d'}_{\le \log_2 S} \cap [0,1)^{d'}} \sum_{M \in \mathbb{Z}^{d'}} \Upsilon^{(2)}_{w,\mathcal{I}}\big(2^k \circ (\theta^{(2)} - a^{(2)}/Q - M)\big).$$

This means that we can deal with a simpler kernel

$$\begin{aligned}
K^{r,1}_{k,w,S,\mathcal{I}}(y) := \mathbb{1}_{|2^{-k}\circ y| \lesssim 1} \sum_{a^{(2)}/Q \in \mathcal{R}^{d'}_{\le \log_2 S} \cap [0,1)^{d'}} \mathfrak{e}\big(y^{(2)}.a^{(2)}/Q\big) \int_{\mathbb{T}^d \times \mathbb{R}^{d'}} \mathfrak{e}\big(y.\theta\big) \\
\times\, \Pi^r_{k,w,S,\mathcal{I}}\big(\theta^{(1)}, \theta^{(2)} + a^{(2)}/Q\big) \big(\Upsilon^{(2)}_{w,\mathcal{I}}\big(2^k \circ \theta^{(2)}\big)\big)^{2r}\, d\theta^{(1)} d\theta^{(2)}.
\end{aligned}$$

We now focus on $\Pi^r_{k,w,S,\mathcal{I}}$. As in (5.15)–(5.16) we may write

$$\begin{aligned}
\Pi^r_{k,w,S,\mathcal{I}}\big(\theta^{(1)}, \theta^{(2)}\big) &= \int_{(\mathbb{T}^d)^{2r}} \mathcal{V}^r_{k,0}(\theta^{(1)}, \theta^{(2)}; \zeta^{(1)}_1, \xi^{(1)}_1, \dots, \zeta^{(1)}_r, \xi^{(1)}_r) \\
&\times \prod_{1\le j\le r} \big\{ \overline{S_k(\zeta^{(1)}_j)}\, \overline{\Psi_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{I}}(\zeta^{(1)}_j)}\, S_k(\xi^{(1)}_j) \Psi_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{I}}(\xi^{(1)}_j) \big\} \\
&\times d\xi^{(1)}_1 d\zeta^{(1)}_1 \dots d\xi^{(1)}_r d\zeta^{(1)}_r, \qquad (9.49)
\end{aligned}$$

where

$$\begin{aligned}\mathcal{V}^r_{k,0}(\theta^{(1)},\theta^{(2)};\zeta^{(1)}_1,\xi^{(1)}_1,\dots,\zeta^{(1)}_r,\xi^{(1)}_r)&\\ =\sum_{h_j,g_j\in\mathbb{Z}^d}\prod_{1\le j\le r}&\Big\{\overline{\phi^{(1)}_{k,0}(h_j)}\mathfrak{e}\big((\theta^{(1)}-\zeta^{(1)}_j).h_j\big)\\ &\times\phi^{(1)}_{k,0}(g_j)\mathfrak{e}\big(-(\theta^{(1)}-\xi^{(1)}_j).g_j\big)\Big\}\\ &\times\mathfrak{e}\Big(-\theta^{(2)}.\Big\{\sum_{1\le j\le r}R_0(h_j,h_j-g_j)\\ &\quad+\sum_{1\le l<j\le r}R_0(-h_l+g_l,-h_j+g_j)\Big\}\Big).\end{aligned}$$

Further, proceeding as in the proof of Lemma 5.2 we see that for $|2^k\circ\theta^{(2)}|\lesssim 2^{\delta w}$ and $a^{(2)}/Q\in\mathcal{R}^{d'}_{\le\log_2 S}\cap[0,1)^{d'}$ we have

$$\begin{aligned}&\mathcal{V}^r_{k,0}(\theta^{(1)},\theta^{(2)}+a^{(2)}/Q;\zeta^{(1)}_1,\xi^{(1)}_1,\dots,\zeta^{(1)}_r,\xi^{(1)}_r)\\ &\quad=\mathcal{W}^r_Q(a^{(2)};\underline{b},\underline{c})\mathcal{Z}^r_{k,0}(\theta^{(2)};\beta_1,\gamma_1,\dots,\beta_r,\gamma_r)+O(2^{-D^3k}),\end{aligned}$$

where $\underline{b},\underline{c}\in\mathbb{Z}^{rd}$ and $\beta_j,\gamma_j\in[-1/(2Q),1/(2Q)]^d$ are defined in (5.22). Here $\mathcal{W}^r_Q(a^{(2)};b,c)$ is defined in (5.24) and $\mathcal{Z}^r_{k,0}$ is a modification of (5.25), i.e.

$$\begin{aligned}\mathcal{Z}^r_{k,0}(\theta^{(2)};\beta_1,\gamma_1,\dots,\beta_r,\gamma_r):=\int_{\mathbb{R}^{2rd}}&\Big\{\prod_{l=1}^d 2^{kl}\Big\}^{2r}\\ &\times\prod_{1\le j\le r}\Big\{\eta_{\le D}(x_j)\mathfrak{e}\big(-(2^k\circ\beta_j).x_j\big)\eta_{\le D}(y_j)\mathfrak{e}\big((2^k\circ\gamma_j).y_j\big)\Big\}\\ &\times\mathfrak{e}\Big(-(2^k\circ\theta^{(2)}).\Big\{\sum_{1\le j\le r}R_0(y_j,y_j-x_j)\\ &\quad+\sum_{1\le l<j\le r}R_0(-y_l+x_l,-y_j+x_j)\Big\}\Big)\,dx_jdy_j.\end{aligned}$$

Further, we have an analogue of (5.18), namely

$$\begin{aligned}&\big|\mathcal{V}^r_{k,0}(\theta^{(1)},\theta^{(2)}+a^{(2)}/Q;\zeta^{(1)}_1,\xi^{(1)}_1,\dots,\zeta^{(1)}_r,\xi^{(1)}_r)\big|\\ &\lesssim\Big\{\prod_{1\le l\le d}2^{kl}\Big\}^{2r}\min_{\substack{1\le j\le r\\1\le l\le d}}\big[1+2^{kl-\delta w}\|\theta^{(1)}_l-\zeta^{(1)}_{j,l}\|_Q+2^{kl-\delta w}\|\theta^{(1)}_l-\xi^{(1)}_{j,l}\|_Q\big]^{-D},\end{aligned}$$

for any $\theta^{(1)} = (\theta_l^{(1)})_{l\in\{1,\ldots,d\}} \in \mathbb{T}^d$, $\zeta_j^{(1)} = (\zeta_{j,l}^{(1)})_{l\in\{1,\ldots,d\}} \in \mathbb{T}^d$, and $\xi_j^{(1)} = (\xi_{j,l}^{(1)})_{l\in\{1,\ldots,d\}} \in \mathbb{T}^d$. Using this we proceed as in Step 2 of the proof of Lemma 8.2. Having a rapid decay unless $2^{kl}\|\theta_l^{(1)} - \xi_{j,l}^{(1)}\|_Q \le 2^{2\delta w}$ and $2^{kl}\|\theta_l^{(1)} - \zeta_{j,l}^{(1)}\|_Q \le 2^{2\delta w}$ for all $j \in \{1,\ldots,r\}$ and $l \in \{1,\ldots,d\}$ we expand the cutoff functions $\Psi_{k,w,\mathcal{R}^d_{\le D\log_2 S},\mathcal{I}}$ in (9.49) and we use Lemma 6.1 to obtain

$$\|K^{r,1}_{k,w,S,\mathcal{I}} - K^{r,2}_{k,w,S,\mathcal{I}}\|_{\ell^1(\mathbb{G}_0)} \lesssim S^{4rD(d+d')}2^{-Dw/4} \lesssim 2^{-w},$$

where

$$\begin{aligned}
K^{r,2}_{k,w,S,\mathcal{I}}(y) := \mathbb{1}_{|2^{-k}\circ y|\lesssim 1} &\sum_{a^{(2)}/Q\in\mathcal{R}^{d'}_{\le\log_2 S}\cap[0,1)^{d'}} \\
&\sum_{\sigma\in[\mathcal{R}^d_{\le D\log_2 S}+(\mathbb{Z}_Q/Q)^d]\cap[0,1)^d} \mathcal{C}(a^{(2)}/Q,\sigma) \\
&\times \mathfrak{e}\big(y^{(2)}.a^{(2)}/Q\big)\int_{\mathbb{R}^{d+d'}} \mathfrak{e}\big(y.\theta\big)\eta_{\le 2\delta' w+D}(2^k\circ(\theta^{(1)}-\sigma)) \\
&\times \int_{\mathbb{R}^{2rd}} \mathcal{Z}^r_{k,0}\big(\theta^{(2)};\theta^{(1)}-\xi_j^{(1)}-\sigma,\theta^{(1)}-\zeta_j^{(1)}-\sigma\big) \\
&\times \Big\{\prod_{j=1}^r \Upsilon^{(1)}_{w,\mathcal{I}}(2^k\circ\xi_j^{(1)})\Upsilon^{(1)}_{w,\mathcal{I}}(2^k\circ\zeta_j^{(1)})J_k(\xi_j^{(1)})\overline{J_k(\zeta_j^{(1)})}\Big\} \\
&\times d\xi_1^{(1)}d\zeta_1^{(1)}\ldots d\xi_r^{(1)}d\zeta_r^{(1)} \\
&\times \big(\Upsilon^{(2)}_{w,\mathcal{I}}\big(2^k\circ\theta^{(2)}\big)\big)^{2r}\,d\theta^{(1)}d\theta^{(2)}.
\end{aligned}$$

Here $\mathcal{C}(a^{(2)}/Q,\sigma)$ is defined as in (8.23) with

$$\iota_Q(\sigma;\underline{b},\underline{c}) := \begin{cases} 1 & \text{if } \sigma - b_j/Q, \sigma - c_j/Q \in \mathcal{R}^d_{\le D\log_2 S} \text{ for any } j\in\{1,\ldots,r\};\\ 0 & \text{otherwise.}\end{cases}$$

Note that $\mathcal{C}(a^{(2)}/Q,\sigma)$ satisfies the estimate

$$|\mathcal{C}(a^{(2)}/Q,\sigma)| \lesssim Q^{3rd}Q_1^{-2r/\overline{C}} \lesssim Q^{3rd} \lesssim S^{3rd}, \tag{9.50}$$

for any $a^{(2)}/Q \in \mathcal{R}^{d'}_{\le\log_2 S}\cap[0,1)^{d'}$ and $\sigma \in [\mathcal{R}^d_{\le D\log_2 S}+(\mathbb{Z}_Q/Q)^d]\cap[0,1)^d$, where Q_1 is a denominator of the first component of σ and $\overline{C}$ is the constant from Proposition 2.2. Therefore it suffices to deal with the kernel $K^{r,2}_{k,w,S,\mathcal{I}}$. Next, we focus on the integral over $\xi_j^{(1)}$, $\zeta_j^{(1)}$ above. Proceeding as in Step 3 of

the proof of Lemma 8.2 we are able to prove that up to an error term $O(2^{-Dw})$ this integral is equal to $I_{k,w}(\theta^{(1)}-\sigma,\theta^{(2)})$, where

$$I_{k,w}(\theta^{(1)},\theta^{(2)}) := \int_{\mathbb{R}^{2rd}} \prod_{1\le j\le r} \big\{\widehat{\Upsilon^{(1)}_{w,\mathcal{I}}}(-x'_j)\widehat{\Upsilon^{(1)}_{w,\mathcal{I}}}(y'_j)\mathfrak{e}\big(-(2^k\circ\theta^{(1)}).(x'_j-y'_j)\big)\big\}$$
$$\times\Big\{\int_{\mathbb{R}^{2r}} \mathfrak{e}\big(-(2^k\circ\theta^{(2)}).T(\underline{x}',\underline{y}',\underline{u},\underline{v})\big)\prod_{1\le j\le r}\big\{\chi(u_j)\chi(v_j)\big\}$$
$$\mathfrak{e}\big(-(2^k\circ\theta).D(\underline{v},\underline{u})\big)\,d\underline{u}d\underline{v}\Big\}d\underline{x}'d\underline{y}'.$$

Therefore we have

$$\|K^{r,2}_{k,w,S,\mathcal{I}}-K^{r,3}_{k,w,S,\mathcal{I}}\|_{\ell^1(\mathbb{G}_0)}\lesssim S^{4rD(d+d')}2^{-Dw/2}\lesssim 2^{-w},$$

where

$$K^{r,3}_{k,w,S,\mathcal{I}}(y) := \mathbb{1}_{|2^{-k}\circ y|\lesssim 1}\sum_{a^{(2)}/Q\in\mathcal{R}^{d'}_{\le\log_2 S}\cap[0,1)^{d'}}\ \sum_{\sigma\in[\mathcal{R}^d_{\le D\log_2 S}+(\mathbb{Z}_Q/Q)^d]\cap[0,1)^d}\mathcal{C}(a^{(2)}/Q,\sigma)$$
$$\times\,\mathfrak{e}\big(y.(\sigma,a^{(2)}/Q)\big)\int_{\mathbb{R}^{d+d'}}\mathfrak{e}\big(y.\theta\big)\eta_{\le 2\delta' w+D}(2^k\circ\theta^{(1)})$$
$$\times\,\big(\Upsilon^{(2)}_{w,\mathcal{I}}\big(2^k\circ\theta^{(2)}\big)\big)^{2r}I_{k,w}(\theta)\,d\theta^{(1)}d\theta^{(2)}.$$

Next, proceeding as in Step 4 of the proof of Lemma 8.2 we conclude

$$\eta_{\le 2\delta' w+D}(2^k\circ\theta^{(1)})\big(\Upsilon^{(2)}_{w,\mathcal{I}}\big(2^k\circ\theta^{(2)}\big)\big)^{2r}I_{k,w}(\theta)$$
$$=\eta_{\le\delta w/2}(2^k\circ\theta^{(1)})\eta_{\le\delta w/2}(2^k\circ\theta^{(2)})\big(\Upsilon^{(2)}_{w,\mathcal{I}}\big(2^k\circ\theta^{(2)}\big)\big)^{2r}$$
$$\times\,\big(\Upsilon^{(1)}_{w,\mathcal{I}}(0)\big)^{2r}P(2^k\circ\theta)+O(2^{-w/\delta}),$$

where P is defined in (7.12). Therefore using (9.50) we obtain

$$\|K^{r,3}_{k,w,S,\mathcal{I}}-K^{r,4}_{k,w,S,\mathcal{I}}\|_{\ell^1(\mathbb{G}_0)}\lesssim 2^{-w/(2\delta)}S^{4r(d+d')}\lesssim 2^{-w},$$

where

$$K^{r,4}_{k,w,S,\mathcal{I}}(y) := \mathbb{1}_{|2^{-k}\circ y|\lesssim 1} \sum_{a^{(2)}/Q\in\mathcal{R}^{d'}_{\le\log_2 S}\cap[0,1)^{d'}} \sum_{\sigma\in[\mathcal{R}^{d}_{\le D\log_2 S}+(\mathbb{Z}_Q/Q)^d]\cap[0,1)^d} \mathcal{C}(a^{(2)}/Q,\sigma)$$
$$\times \mathfrak{e}\big(y.(\sigma,a^{(2)}/Q)\big)\Big\{\prod_{(l_1,l_2)\in Y_d} 2^{-k(l_1+l_2)}\Big\}$$
$$\times \int_{\mathbb{R}^{d+d'}} \mathfrak{e}\big[(2^{-k}\circ y).\theta\big]\eta_{\le\delta w/2}(\theta^{(1)})\eta_{\le\delta w/2}(\theta^{(2)})$$
$$\times \big(\Upsilon^{(2)}_{w,\mathcal{I}}(\theta^{(2)})\big)^{2r}\big(\Upsilon^{(1)}_{w,\mathcal{I}}(0)\big)^{2r} P(\theta)\,d\theta^{(1)}d\theta^{(2)}.$$

Finally, to prove (9.48) it suffices to show that

$$\|K^{r,4}_{k,w,S,\mathcal{I}}\|_{\ell^1(\mathbb{G}_0)} \lesssim 2^{-w}.$$

If $1\in\mathcal{I}$, then $\Upsilon^{(1)}_{w,\mathcal{I}}(0)=0$ and there is nothing to prove. Otherwise, since $\mathcal{I}\neq\emptyset$ we need to have $2\in\mathcal{I}$. This means that $|\theta^{(2)}|\gtrsim 2^{\delta w}$ and using (7.13) with $\iota=0$ together with (9.50) we have

$$\|K^{r,4}_{k,w,S,\mathcal{I}}\|_{\ell^1(\mathbb{G}_0)} \lesssim S^{4r(d+d')}2^{-w/(2\delta)} \lesssim 2^{-w}.$$

This proves (9.48) and consequently the proof of Lemma 9.6 is completed. □

Acknowledgements This work was started in collaboration with Steve Wainger. The authors would like to thank him for his mentorship and friendship over many years and for many inspiring discussions on this topic. We also thank Bartosz Langowski for reading the manuscript at the very early stages of our work. Finally, we thank the referees for careful reading of the manuscript and useful remarks that led to the improvement of the presentation.

Appendix A. Proof of Proposition 6.5

In this section we prove the estimates (6.32) and (6.34). We begin with proving (6.34), which will be needed in the proof of (6.32).

A.1. Proof of inequality (6.34)

We examine the definitions (6.31) and (6.4), and rewrite

$$\widetilde{W}_{k,w+1}(x)-\widetilde{W}_{k,w}(x)=\phi_k(x)\sum_{\emptyset\neq\mathcal{I}\subseteq\{1,2\}} S_{k,w,\mathcal{I}}(x),$$

where for $\mathcal{I} \subseteq \{1, 2\}$ we define

$$\begin{aligned} S_{k,w,\mathcal{I}}(x) &:= S^{(1)}_{k,w,\mathcal{I}}(x^{(1)}) S^{(2)}_{k,w,\mathcal{I}}(x^{(2)}), \\ S^{(1)}_{k,w,\mathcal{I}}(x^{(1)}) &:= \Big\{ \prod_{l\in\{1,\dots,d\}} \tau^{-kl} \Big\} \int_{\mathbb{R}} \chi(u) \widehat{\Upsilon^{(1)}_{w,\mathcal{I}}}(A^{(1)}_0(u) - \tau^{-k} \circ x^{(1)})\, du, \\ S^{(2)}_{k,w,\mathcal{I}}(x^{(2)}) &:= \Big\{ \prod_{(l_1,l_2)\in Y'_d} \tau^{-k(l_1+l_2)} \Big\} \widehat{\Upsilon^{(2)}_{w,\mathcal{I}}}(-\tau^{-k} \circ x^{(2)}), \end{aligned} \tag{A.1}$$

and $\Upsilon^{(1)}_{w,\mathcal{I}}$ and $\Upsilon^{(2)}_{w,\mathcal{I}}$ are defined in (9.46). Let $\mathcal{S}_{k,w,\mathcal{I}} f := f *_{\mathbb{G}^{\#}_0} S_{k,w,\mathcal{I}}$. Notice that

$$\|\phi_k S_{k,w,\mathcal{I}} - S_{k,w,\mathcal{I}}\|_{L^1(\mathbb{G}^{\#}_0)} \lesssim \tau^{-Dk}, \qquad \mathcal{I} \neq \emptyset, \quad 0 \le w < k.$$

Therefore, to prove (6.34) it suffices to show that if $w \ge 0$ and $\mathcal{I} \neq \emptyset$ then

$$\Big\| \sum_{k>w} \varkappa_k \mathcal{S}_{k,w,\mathcal{I}} f \Big\|_{L^2(\mathbb{G}^{\#}_0)} \lesssim \tau^{-w/D} \|f\|_{L^2(\mathbb{G}^{\#}_0)},$$

provided that $|\varkappa_k| \le 1$. In view of the Cotlar–Stein lemma it suffices to prove that

$$\begin{aligned} \|\mathcal{S}_{j,w,\mathcal{I}} \mathcal{S}^*_{k,w,\mathcal{I}}\|_{L^2(\mathbb{G}^{\#}_0)\to L^2(\mathbb{G}^{\#}_0)} + \|\mathcal{S}^*_{j,w,\mathcal{I}} \mathcal{S}_{k,w,\mathcal{I}}\|_{L^2(\mathbb{G}^{\#}_0)\to L^2(\mathbb{G}^{\#}_0)} \\ \lesssim \tau^{-2w/D} \tau^{-|k-j|/D}, \end{aligned} \tag{A.2}$$

uniformly in $0 \le w < j \le k$ and $\mathcal{I} \neq \emptyset$. We will prove the estimates only for the first term in the left-hand side above, since the second term can be treated in a similar way.

With $\widetilde{\delta} = (\delta_{l_1 l_2})_{(l_1,l_2)\in Y_d}$, $\delta_{l_1 l_2} = \delta$ if $(l_1, l_2) \in Y'_d$ and $\delta_{l_1 0} = \delta'$ as before, it is easy to see that

$$\begin{aligned} &|S_{k,w,\mathcal{I}}(x)| + \sum_{(l_1,l_2)\in Y_d} \tau^{k(l_1+l_2)-\delta_{l_1 l_2} w} \big|(\partial_{x_{l_1 l_2}} S_{k,w,\mathcal{I}})(x)\big| \\ &\lesssim \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)+\delta_{l_1 l_2} w} \Big\} \int_{\mathbb{R}} \chi(u) \Big\langle \tau^{\tilde{\delta} w} \big(A_0(u) - \tau^{-k} \circ x\big) \Big\rangle^{-D} du, \end{aligned} \tag{A.3}$$

uniformly in $x \in \mathbb{G}_0^{\#}$, $0 \le w < k$. Observe that for every $\theta \in \mathbb{R}^{d+d'}$ we also have

$$\widehat{S_{k,w,\mathcal{I}}}(\theta) = \Upsilon^{(1)}_{w,\mathcal{I}}(\tau^k \circ \theta^{(1)})\Upsilon^{(2)}_{w,\mathcal{I}}(\tau^k \circ \theta^{(2)}) \int_{\mathbb{R}} \chi(u)\mathfrak{e}\big(-\theta.A_0(\tau^k u)\big)\,du. \tag{A.4}$$

Step 1. We prove first the bounds (A.2) when $k - j \ge w$. Using (A.4) we have $\int_{\mathbb{G}_0^{\#}} S_{k,w,\mathcal{I}}(x)\,dx = 0$ for $\mathcal{I} \neq \emptyset$. Therefore the kernels $\mathbb{K}_{k,j}$ of $\mathcal{S}_{j,w,\mathcal{I}}\mathcal{S}^*_{k,w,\mathcal{I}}$ satisfy the bounds

$$\|\mathbb{K}_{k,j}\|_{L^1(\mathbb{G}_0^{\#})} \le \int_{\mathbb{G}_0^{\#}} |S_{j,w,\mathcal{I}}(y)| \int_{\mathbb{G}_0^{\#}} |S_{k,w,\mathcal{I}}(x \cdot y) - S_{k,w,\mathcal{I}}(x)|\,dx\,dy. \tag{A.5}$$

Using now the bounds (A.3) we obtain

$$\begin{aligned}|S_{k,w,\mathcal{I}}(x \cdot y) - S_{k,w,\mathcal{I}}(x)| &\lesssim \tau^{-(k-j)/2}\Big\{\prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)+\delta_{l_1 l_2} w}\Big\}\langle \tau^{-j} \circ y\rangle \\ &\quad \times \int_{\mathbb{R}} \chi(u)\Big\langle \tau^{\tilde{\delta} w}\big(A_0(u) - \tau^{-k} \circ x\big)\Big\rangle^{-D/8+1} du \\ &\quad \times \Big\langle \tau^{\tilde{\delta} w}\big(\tau^{-k} \circ y\big)\Big\rangle^{D/4},\end{aligned}$$

for any $x, y \in \mathbb{G}_0^{\#}$, provided that $k - j \ge w$. Therefore, using (A.5),

$$\begin{aligned}\|\mathbb{K}_{k,j}\|_{L^1(\mathbb{G}_0^{\#})} &\lesssim \tau^{-(k-j)/2} \int_{\mathbb{G}_0^{\#}} \Big\{\prod_{(l_1,l_2)\in Y_d} \tau^{-j(l_1+l_2)+\delta_{l_1 l_2} w}\Big\} \\ &\quad \times \int_{\mathbb{R}} \chi(v)\Big\langle \tau^{\tilde{\delta} w}\big(A_0(v) - \tau^{-j} \circ y\big)\Big\rangle^{-D/4} dv\,dy \lesssim \tau^{-(k-j)/2}.\end{aligned}$$

This proves (A.2) provided that $k - j \ge w$.

Step 2. Assume now that $k - j \le w$. Using a high order T^*T argument it suffices to prove that if $0 \le w < k$ and $\mathcal{I} \neq \emptyset$ then

$$\|(\mathcal{S}^*_{k,w,\mathcal{I}}\mathcal{S}_{k,w,\mathcal{I}})^r\|_{L^2(\mathbb{G}_0^{\#}) \to L^2(\mathbb{G}_0^{\#})} \lesssim \tau^{-w}. \tag{A.6}$$

Using the formulas (2.13)–(2.16) we see that $(\mathcal{S}^*_{k,w,\mathcal{I}}\mathcal{S}_{k,w,\mathcal{I}})^r f = f *_{\mathbb{G}_0^{\#}} \mathbb{K}^r_k$, where

$$\mathbb{K}^r_k(z) = \int_{\mathbb{R}^d \times \mathbb{R}^{d'}} \mathfrak{e}\big(\theta.z\big)\big(\Upsilon^{(2)}_{w,\mathcal{I}}(\tau^k \circ \theta^{(2)})\big)^{2r} I^r_{k,w,\mathcal{I}}(\theta^{(1)}, \theta^{(2)})\,d\theta, \tag{A.7}$$

and

$$I^r_{k,w,\mathcal{I}}(\theta) := \int_{\mathbb{R}^{2rd}} \Big\{ \prod_{i=1}^{r} \overline{S^{(1)}_{k,w,\mathcal{I}}(h^{(1)}_i)} S^{(1)}_{k,w,\mathcal{I}}(g^{(1)}_i) \Big\} \mathfrak{e}\big(\theta^{(1)}.\sum_{1\le i\le r}(h^{(1)}_i - g^{(1)}_i)\big)$$
$$\times \mathfrak{e}\Big(-\theta^{(2)}.\Big\{ \sum_{1\le i\le r} R_0(h^{(1)}_i, h^{(1)}_i - g^{(1)}_i)$$
$$+ \sum_{1\le p<i\le r} R_0(-h^{(1)}_p + g^{(1)}_p, -h^{(1)}_i + g^{(1)}_i)\Big\}\Big)\, dh^{(1)}_i dg^{(1)}_i.$$

Using the definitions (A.1), (8.21), and (2.24), and making the changes of variables $h^{(1)}_i = \tau^k \circ (A^{(1)}_0(v_i) + y_i)$, $g^{(1)}_i = \tau^k \circ (A^{(1)}_0(u_i) + x_i)$ we rewrite

$$I^r_{k,w,\mathcal{I}}(\theta) = \int_{\mathbb{R}^{2rd}} \prod_{i=1}^{r} \Big\{ \widehat{\Upsilon^{(1)}_{w,\mathcal{I}}}(y_i) \widehat{\Upsilon^{(1)}_{w,\mathcal{I}}}(-x_i) \mathfrak{e}\big(-(\tau^k \circ \theta^{(1)}).(x_i - y_i)\big)\Big\}$$
$$\times \Big\{ \int_{\mathbb{R}^{2r}} \mathfrak{e}\big(-(\tau^k \circ \theta^{(2)}).T(\underline{x}, \underline{y}, \underline{u}, \underline{v})\big) \prod_{i=1}^{r} \{\chi(u_j)\chi(v_j)\} \tag{A.8}$$
$$\times \mathfrak{e}\big(-(\tau^k \circ \theta).D(\underline{v}, \underline{u})\big)\, d\underline{u} d\underline{v}\Big\}\, d\underline{x} d\underline{y}.$$

In view of (A.3) we have

$$\| S_{k,w,\mathcal{I}}(x) \mathbb{1}_{|\tau^{-k}\circ x|\ge 10^d d^{10}} \|_{L^1(\mathbb{G}^{\#}_0)} \lesssim \tau^{-\delta D w/2}.$$

To prove (A.6) it suffices to show that for a large fixed constant $C_r \gg 1$ we have

$$\| \mathbb{K}^r_k(x) \mathbb{1}_{|\tau^{-k}\circ x|\le C_r} \|_{L^1(\mathbb{G}^{\#}_0)} \lesssim \tau^{-w}.$$

In view of (A.7), for this is suffices to show that for any $(\theta^{(1)}, \theta^{(2)}) \in \mathbb{R}^d \times \mathbb{R}^{d'}$ we have

$$\big|\big(\Upsilon^{(2)}_{w,\mathcal{I}}(\tau^k \circ \theta^{(2)})\big)^{2r} I^r_{k,w,\mathcal{I}}(\theta^{(1)}, \theta^{(2)})\big|$$
$$\lesssim \big|\Upsilon^{(2)}_{w,\mathcal{I}}(\tau^k \circ \theta^{(2)})\big|^{2r} \tau^{-4w} (1 + \tau^{-2\delta' w} |\tau^k \circ \theta^{(1)}|)^{-1/\delta}. \tag{A.9}$$

This is similar to the proof in Steps 3 and 4 of Lemma 8.2. Indeed, first we integrate by parts many times in x_i (or in y_i) in the identity (A.8) to see that

$$\big|\big(\Upsilon^{(2)}_{w,\mathcal{I}}(\tau^k \circ \theta^{(2)})\big)^{2r} I^r_{k,w,\mathcal{I}}(\theta^{(1)}, \theta^{(2)})\big|$$
$$\lesssim \big|\Upsilon^{(2)}_{w,\mathcal{I}}(\tau^k \circ \theta^{(2)})\big|^{2r} (1 + \tau^{-2\delta' w} |\tau^k \circ \theta^{(1)}|)^{-D}$$

for any $(\theta^{(1)}, \theta^{(2)}) \in \mathbb{R}^d \times \mathbb{R}^{d'}$. It remains to prove (A.9) if $|\tau^k \circ \theta^{(2)}| \le 2\tau^{\delta w+4}$ and $|\tau^k \circ \theta^{(1)}| \le \tau^{3\delta' w}$. In this case we can use Proposition 2.4 as in Step 4 in Lemma 8.2 to prove a suitable decay if $|\tau^k \circ \theta| \ge \tau^{\delta w-4}$. Finally, if $|\tau^k \circ \theta| \le \tau^{\delta w-4}$ then we may assume that $1 \in \mathcal{I}$, so

$$\int_{\mathbb{R}^d} \widehat{\Upsilon^{(1)}_{w,\mathcal{I}}}(x) x^\beta = 0$$

for any multi-index $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{N}^d$. This is similar to (8.25) and can be used to show that $\big|I^r_{k,\mathcal{I}}(\theta)\big| \lesssim \tau^{-Dw}$ if $|\tau^k \circ \theta| \le \tau^{\delta w-4}$. This finishes the proof of inequality (6.34). □

A.2. Proof of inequality (6.32)

The space $X = \mathbb{G}_0^\#$ endowed with the Lebesgue measure $\mu_{\mathbb{G}_0^\#} = |\cdot|$ and the quasi-metric

$$\mathfrak{q}_{\mathbb{G}_0^\#}(x, y) := \sup_{(l_1,l_2)\in Y_d} \Big(\big|[x \cdot y^{-1}]_{l_1 l_2}\big|^{1/(l_1+l_2)} \Big), \qquad x, y \in \mathbb{G}_0^\# \tag{A.10}$$

defines a space of homogeneous type $(\mathbb{G}_0^\#, \mathcal{B}(\mathbb{G}_0^\#), \mu_{\mathbb{G}_0^\#}, \mathfrak{q}_{\mathbb{G}_0^\#})$. This in turn allows us to associate a system of dyadic cubes for X in the sense of Christ [18, Theorem 11].

Following [35, Section 3 and 4, pp. 6721–6726] we can define the martingale sequence $\mathbb{E}_k f(x) = \mathbb{E}[f|\mathcal{F}_k](x)$ for $k \in \mathbb{Z}$, $x \in \mathbb{G}_0^\#$, and $f \in L^1_{\rm loc}(\mathbb{G}_0^\#)$, where $\mathcal{F}_k$ is the filtration corresponding to the system of Christ's dyadic cubes, see [35, formula (27), p. 6721] and [35, Lemma 3.1, p. 6721].

An important ingredient in the proof of inequality (6.32) will be Lépingle's inequality [41,45], which asserts that for every $1 < p < \infty$ and $2 < \rho < \infty$ and every $f \in L^p(\mathbb{G}_0^\#)$ one has

$$\|V^\rho(\mathbb{E}_k f : k \in \mathbb{Z})\|_{L^p(\mathbb{G}_0^\#)} \lesssim_{p,\rho} \|f\|_{L^p(\mathbb{G}_0^\#)}. \tag{A.11}$$

We now establish certain variational estimates necessary to prove (6.32). In a similar way as in (6.31), let us define new kernels on $\mathbb{G}_0^\#$ by setting

$$W_k(x) := \int_{\mathbb{R}^d\times\mathbb{R}^{d'}} \eta_0(\tau^k \circ \xi^{(1)})\eta_0(\tau^k \circ \xi^{(2)})\mathfrak{e}(x.\xi) J_k(\xi^{(1)})\, d\xi,$$
$$x \in \mathbb{G}_0^\#, \quad \tau > 1.$$

Observe that

$$W_k(x) = \int_{\mathbb{R}} \tau^{-k} \chi(\tau^{-k}u)\psi_k(x - A_0(u))\, du, \quad x \in \mathbb{G}_0^{\#},$$

where for $k \in \mathbb{Z}$ and $x \in \mathbb{G}_0^{\#}$ we set

$$\psi_k(x) := \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)} \Big\} \psi(\tau^{-k} \circ x), \quad \psi(x) := \widehat{\eta}_0(-x^{(1)})\widehat{\eta}_0(-x^{(2)}).$$

The main result of this subsection is stated below.

Lemma A.1 *Let $2 < \rho < \infty$ be given. Then for any $g \in L^2(\mathbb{G}_0^{\#})$ one has*

$$\|V^\rho(g *_{\mathbb{G}_0^{\#}} W_k : k \in \mathbb{Z})\|_{L^2(\mathbb{G}_0^{\#})} \lesssim_{\rho,\tau} \|g\|_{L^2(\mathbb{G}_0^{\#})}. \tag{A.12}$$

Proof We reduce the matters to Lépingle's inequality for bounded martingales (A.11).

Step 1. Let $\mu_0 := \int_{\mathbb{R}} \chi(x)dx$ and define

$$T_k g(x) := g *_{\mathbb{G}_0^{\#}} W_k(x) - g *_{\mathbb{G}_0^{\#}} (\mu_0\psi_k)(x) =: g *_{\mathbb{G}_0^{\#}} \mathbb{K}_k(x), \quad x \in \mathbb{G}_0^{\#}.$$

Observe that

$$\begin{aligned}\|V^\rho(g *_{\mathbb{G}_0^{\#}} W_k : k \in \mathbb{Z})\|_{L^2(\mathbb{G}_0^{\#})} \lesssim \|V^\rho(g *_{\mathbb{G}_0^{\#}} \psi_k : k \in \mathbb{Z})\|_{L^2(\mathbb{G}_0^{\#})} \\ + \Big\| \Big(\sum_{k\in\mathbb{Z}} |T_k g|^2 \Big)^{1/2} \Big\|_{L^2(\mathbb{G}_0^{\#})}.\end{aligned} \tag{A.13}$$

As in the Jones–Seeger–Wright paper [35] we can conclude that

$$\|V^\rho(g *_{\mathbb{G}_0^{\#}} \psi_k : k \in \mathbb{Z})\|_{L^2(\mathbb{G}_0^{\#})} \lesssim_{\rho,\tau} \|g\|_{L^2(\mathbb{G}_0^{\#})}. \tag{A.14}$$

Indeed, let $\mathbb{E}_k f$ denote the martingale sequence, as above, and define the martingale difference operator $\mathbb{D}_k = \mathbb{E}_k - \mathbb{E}_{k-1}$ and proceeding as in the proof of [35, Lemma 3.2, p. 6722] we are able to prove that there is a constant $\gamma > 0$ such that for any $f \in L^2(\mathbb{G}_0^{\#})$ the estimate

$$\|(\mathbb{D}_m f) *_{\mathbb{G}_0^{\#}} \psi_{M_0(k+m)+b} - \mathbb{E}_{k+m}\mathbb{D}_m f\|_{L^2(\mathbb{G}_0^{\#})} \lesssim \tau^{-\gamma|k|}\|\mathbb{D}_m f\|_{L^2(\mathbb{G}_0^{\#})},$$

holds uniformly in $k, m \in \mathbb{Z}$, and $b \in \mathbb{Z}_{M_0}$; here $M_0 \in \mathbb{N}$ is fixed but large constant such that $\delta = 2^{-M_0}$ in the construction of Christ's dyadic cubes, see [18, Theorem 11]. This estimate and a simple square function argument (see

[35, Section 4, p. 6724]) reduces (A.14) to Lépingle's inequality (A.11) and the claim follows.

Step 2. The proof will be completed if we estimate the square function from (A.13). By Khintchine's inequality it suffices to show that for every $f \in L^2(\mathbb{G}_0^\#)$ one has

$$\Big\| \sum_{k\in\mathbb{Z}} \varkappa_k T_k g \Big\|_{L^2(\mathbb{G}_0^\#)} \lesssim \|g\|_{L^2(\mathbb{G}_0^\#)},$$

for any coefficients $\varkappa_k \in [-1, 1]$. Using the Cotlar–Stein lemma it remains to prove that

$$\|\mathbb{K}_k^* *_{\mathbb{G}_0^\#} \mathbb{K}_j\|_{L^1(\mathbb{G}_0^\#)} + \|\mathbb{K}_j *_{\mathbb{G}_0^\#} \mathbb{K}_k^*\|_{L^1(\mathbb{G}_0^\#)} \lesssim \tau^{-|k-j|}, \qquad k \geq j. \tag{A.15}$$

We prove only the first estimate since the second one is analogous. Note that

$$|\mathbb{K}_k^* *_{\mathbb{G}_0^\#} \mathbb{K}_j(x)| \leq \int_{\mathbb{G}_0^\#} |\mathbb{K}_j(y)| \big|\mathbb{K}_k(x^{-1}\cdot y) - \mathbb{K}_k(x^{-1})\big|\, dy, \tag{A.16}$$

since we have $\int_{\mathbb{G}_0^\#} \mathbb{K}_j(x)dx = 0$. Further, using the estimate

$$\begin{aligned}|\psi_k(x\cdot y - z) - \psi_k(x - z)| \lesssim \tau^{-|k-j|}\Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)}\Big\} \\ \times \langle \tau^{-j}\circ y\rangle^{D+1} \langle \tau^{-k}\circ x\rangle^{-D/2+1},\end{aligned}$$

which holds uniformly in $k \geq j$, $|\tau^{-k}\circ z| \lesssim 1$, and $x, y \in \mathbb{G}_0^\#$, we obtain

$$\begin{aligned}&\big|\mathbb{K}_k(x\cdot y) - \mathbb{K}_k(x)\big| \\ &\quad\lesssim \tau^{-|k-j|}\Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)}\Big\} \langle \tau^{-j}\circ y\rangle^{D+1} \langle \tau^{-k}\circ x\rangle^{-D/2+1}.\end{aligned}$$

Combining this with (A.16) and a simple estimate

$$|\mathbb{K}_j(y)| \lesssim \Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-j(l_1+l_2)}\Big\} \langle \tau^{-j}\circ y\rangle^{-4D},$$

we conclude

$$|\mathbb{K}_k^* *_{\mathbb{G}_0^\#} \mathbb{K}_j(x)| \lesssim \tau^{-|k-j|}\Big\{ \prod_{(l_1,l_2)\in Y_d} \tau^{-k(l_1+l_2)}\Big\} \langle \tau^{-k}\circ x\rangle^{-D/8}, \qquad x \in \mathbb{G}_0^\#.$$

This shows (A.15) and the proof of Lemma A.1 is completed. □

We now prove inequality (6.32). Note that

$$\begin{aligned}
&\big\| V^\rho(f *_{\mathbb{G}_0^\#} \widetilde{W}_{k,k} : k \geq 0)\big\|_{L^2(\mathbb{G}_0^\#)} \\
&\quad \leq \big\| V^\rho(f *_{\mathbb{G}_0^\#} W_k : k \in \mathbb{Z})\big\|_{L^2(\mathbb{G}_0^\#)} \\
&\qquad + \Big\| \Big(\sum_{k\geq 0} |f *_{\mathbb{G}_0^\#} (\widetilde{W}_{k,0} - W_k)|^2 \Big)^{1/2} \Big\|_{L^2(\mathbb{G}_0^\#)} \\
&\qquad + \sum_{w\in\mathbb{N}} \Big\| \Big(\sum_{k>w} |f *_{\mathbb{G}_0^\#} (\widetilde{W}_{k,w+1} - \widetilde{W}_{k,w})|^2 \Big)^{1/2} \Big\|_{L^2(\mathbb{G}_0^\#)}.
\end{aligned}$$

The ρ-variations are bounded due to Lemma A.1. The first square function is bounded due to the following pointwise bound

$$|f *_{\mathbb{G}_0^\#} (\widetilde{W}_{k,0} - W_k)(x)| \lesssim \tau^{-k/2} |f| *_{\mathbb{G}_0^\#} E_k(x),$$

where

$$E_k(h) := \Big\{ \prod_{(l_1,l_2)\in Y_d} 2^{-k(l_1+l_2)} \Big\} \langle 2^{-k} \circ h \rangle^{-D}, \qquad h \in \mathbb{G}_0^\#.$$

Appealing to Khintchine's inequality and (6.34) we conclude that the second square function is bounded by a constant multiple of $2^{-w/D} \|f\|_{L^2(\mathbb{G}_0^\#)}$, which completes the proof of (6.32). □

Appendix B. Proof of Proposition 9.4: shifted maximal function

Using the definition of $J_k(\xi)$, (see (6.4)), and (6.8) we obtain

$$\begin{aligned}
W_{k,w,Q}(h) = \phi_k(h) \Big(\prod_{(l_1,l_2)\in Y_d} Q\beta_{l_1 l_2} 2^{-k(l_1+l_2)} \Big) \\
\times \int_{\mathbb{R}} \chi(x) \widehat{\eta}_0\big(\beta^{(1)}\big(2^{-k} \circ h^{(1)} - A_0^{(1)}(x)\big)\big) \\
\times \widehat{\eta}_0\big(\beta^{(2)}(2^{-k} \circ h^{(2)})\big)\, dx,
\end{aligned}$$

where $\beta = (\beta^{(1)}, \beta^{(2)}) = (\beta_{l_1 l_2}) \in \mathbb{R}^{d+d'}$, $\beta_{l_1 l_2} = 2^{\lfloor \delta w \rfloor}$ if $l_2 \neq 0$, $\beta_{l_1 l_2} = 2^{\lfloor \delta' w \rfloor}$ if $l_2 = 0$. We define the quasi-norm on $\mathfrak{q}_\beta : \mathbb{R}^{Y_d} \to [0, \infty)$ by

$$\mathfrak{q}_\beta(x) = \sup_{(l_1,l_2)\in Y_d} (\beta_{l_1 l_2} |x_{l_1 l_2}|)^{1/(l_1+l_2)}. \tag{B.1}$$

Since $\mathfrak{q}_\beta(\lambda \circ x) = \lambda \mathfrak{q}_\beta(x)$, we have

$$|W_{k,w,Q}(h)| \lesssim \int_{\mathbb{R}} \chi(u) \Big(\prod_{(l_1,l_2)\in Y_d} Q\beta_{l_1 l_2} 2^{-k(l_1+l_2)} \Big) \times \Big(1 + 2^{-k}\mathfrak{q}_\beta(h - A_0(2^k u))\Big)^{-D} du. \tag{B.2}$$

For $Q \in \mathbb{Z}_+$, $h \in \mathbb{H}_Q$, and $u \in [-2, 2]$ we define

$$M_{Q,w,u} f(h) := \sup_{k\in\mathbb{N},\, 2^{k/2}\geq 8Q2^{w/8}} \Big(\prod_{(l_1,l_2)\in Y_d} Q\beta_{l_1 l_2} 2^{-k(l_1+l_2)} \Big) \times \sum_{\{y\in\mathbb{H}_Q : \mathfrak{q}_\beta(h\cdot y^{-1} - A_0(2^k u)) < 2^k\}} |f(y)|, \tag{B.3}$$

and notice that, as a consequence of (B.2),

$$\big|f *_{\mathbb{H}_Q} W_{k,w,Q}(h)\big| \lesssim \sum_{n=0}^{\infty} 2^{-nD/2} \int_{-2}^{2} M_{Q,w,2^{-n}u} f(h)\, du,$$

for any $h \in \mathbb{H}_Q$, integer k satisfying $2^{k/2} \geq 8Q2^{w/8}$, and $f \in \ell^p(\mathbb{H}_Q)$, uniformly in Q and w. Therefore, for Proposition 9.4 it suffices to prove the following:

Theorem B.1 *For any $Q \in \mathbb{Z}_+$, $w \in \mathbb{N}$, and $u \in [-2, 2]$ we have*

$$\begin{aligned} &\|M_{Q,w,u}\|_{\ell^1(\mathbb{H}_Q)\to\ell^{1,\infty}(\mathbb{H}_Q)} \lesssim (w+1), \\ &\|M_{Q,w,u}\|_{\ell^p(\mathbb{H}_Q)\to\ell^p(\mathbb{H}_Q)} \lesssim_p (w+1), \qquad p \in (1,\infty]. \end{aligned} \tag{B.4}$$

B.1. Proof of Theorem B.1

We begin with some simple observations related to the quasi-distance $\mathfrak{q}_\beta$ and the associated quasi-balls $B_{\beta,\mathbb{H}_Q}(x, r)$ defined for any $x \in \mathbb{G}_0^{\#}$ and $r > 0$ by

$$\begin{aligned} &B_\beta(x, r) = \{y \in \mathbb{G}_0^{\#} \colon \mathfrak{q}_\beta(x \cdot y^{-1}) < r\}, \\ &B_{\beta,\mathbb{H}_Q}(x, r) = \{y \in \mathbb{H}_Q \colon \mathfrak{q}_\beta(x \cdot y^{-1}) < r\} = B_\beta(x, r) \cap \mathbb{H}_Q. \end{aligned} \tag{B.5}$$

We record first several simple properties, which follow directly from the definition (B.1) and the observation that $1 \leq \beta_{l_1 l_2} \leq \beta_{l_1' 0}$ for any $(l_1, l_2) \in Y_d$ and $l_1' \in \{1, \ldots, d\}$.

Lemma B.2 *The following relations holds uniformly for any $x, y \in \mathbb{G}_0^{\#}$:*

(a) $\mathfrak{q}_\beta(x) \geq 0$ *for every* $x \in \mathbb{G}_0^{\#}$ *and* $\mathfrak{q}_\beta(x) = 0$ *if and only if* $x = 0$,
(b) $\mathfrak{q}_\beta(x + y) + \mathfrak{q}_\beta(x \cdot y) \lesssim \mathfrak{q}_\beta(x) + \mathfrak{q}_\beta(y)$,
(c) $\mathfrak{q}_\beta(x^{-1}) \simeq \mathfrak{q}_\beta(x)$,
(d) $1 + \mathfrak{q}_\beta(x) \lesssim 1 + |\beta x| \lesssim (1 + \mathfrak{q}_\beta(x))^{2d}$, *where* $\beta x := (\beta_{l_1 l_2} x_{l_1 l_2})_{l_1 l_2}$.

We start with a simple lemma concerning the cardinality of the quasi-balls $B_{\beta,\mathbb{H}_Q}(x, r)$.

Lemma B.3 *For any* $x \in \mathbb{G}_0^{\#}$ *and* $r \geq 2Q2^{\delta' w}$ *we have*

$$|B_{\beta,\mathbb{H}_Q}(x, r)| \simeq \prod_{(l_1,l_2)\in Y_d} \frac{r^{l_1+l_2}}{Q\beta_{l_1 l_2}}.$$

Proof Observe that for $x, y \in \mathbb{G}_0^{\#}$ we have

$$(x \cdot y^{-1})^{(1)} = x^{(1)} - y^{(1)}, \quad (x \cdot y^{-1})^{(2)} = x^{(2)} - y^{(2)} + R_0(y^{(1)} - x^{(1)}, y^{(1)}). \tag{B.6}$$

Therefore

$$\begin{aligned} &B_{\beta,\mathbb{H}_Q}(x, r) = \big\{ y^{(1)} \in (Q\mathbb{Z})^d,\ y^{(2)} \in (Q\mathbb{Z})^{d'} : \\ &\beta_{l0}|x_{l0} - y_{l0}| < r^l \text{ for any } l \in \{1, \dots, d\} \\ &\text{and } \beta_{l_1 l_2}|x_{l_1 l_2} - y_{l_1 l_2} + R_0(y^{(1)} - x^{(1)}, y^{(1)})_{l_1 l_2}| < r^{l_1+l_2} \text{ for any } (l_1, l_2) \in Y_d' \big\}. \end{aligned} \tag{B.7}$$

This desired volume bounds follow. □

Next, we prove two facts concerning the quasi-norm $\mathfrak{q}_\beta$ and shifted balls.

Lemma B.4 *There exists a universal constant* $C_0 \geq 1$ *such that for any* $x \in \mathbb{H}_Q$, $u \in [-2, 2]$, *and any* $k \in \mathbb{N}$ *satisfying* $2^{k/2} \geq 2Q2^{\delta' w}$, *there is* $z \in \mathbb{H}_Q$ *such that*

$$\big\{ y \in \mathbb{H}_Q : \mathfrak{q}_\beta\big(x \cdot y^{-1} - A_0(2^k u)\big) < 2^k \big\} \subseteq B_{\beta,\mathbb{H}_Q}(z, C_0 2^k). \tag{B.8}$$

Proof We choose $z \in \mathbb{H}_Q$ satisfying the inequalities

$$\beta_{l_1 0}|z_{l_1 0} - x_{l_1 0} + (2^k u)^{l_1}| \leq 2^{k l_1}, \quad l_1 \in \{1, \dots, d\}, \tag{B.9}$$

$$\begin{aligned} &\beta_{l_1 l_2}|z_{l_1 l_2} - x_{l_1 l_2} + R_0(x^{(1)} - z^{(1)}, x^{(1)} - A_0^{(1)}(2^k u))_{l_1 l_2}| \\ &\leq 2^{k(l_1+l_2)}, \quad (l_1, l_2) \in Y_d'. \end{aligned} \tag{B.10}$$

This is indeed possible due to the assumption $Q2^{\delta' w+1} \le 2^{k/2}$. Using (B.6) we see that for any $y \in \mathbb{H}_Q$ satisfying $\mathfrak{q}_\beta\big(x \cdot y^{-1} - A_0(2^k u)\big) < 2^k$ we have

$$\beta_{l_1 0}|x_{l_1 0} - y_{l_1 0} - (2^k u)^{l_1}| < 2^{k l_1}, \quad l_1 \in \{1, \dots, d\}, \tag{B.11}$$

$$\beta_{l_1 l_2}|x_{l_1 l_2} - y_{l_1 l_2} + R_0(y^{(1)} - x^{(1)}, y^{(1)})_{l_1 l_2}| < 2^{k(l_1+l_2)}, \quad (l_1, l_2) \in Y'_d. \tag{B.12}$$

We want to show that $y \in B_{\beta,\mathbb{H}_Q}(z, C_0 2^k)$ for some large constant C_0. Using (B.9) and (B.11)

$$\beta_{l_1 0}|z_{l_1 0} - y_{l_1 0}| \le 2^{k l_1 + 1}, \quad 1 \le l_1 \le d.$$

To finish the proof of Lemma B.4 it is enough to show that

$$\beta_{l_1 l_2}|z_{l_1 l_2} - y_{l_1 l_2} + R_0(y^{(1)} - z^{(1)}, y^{(1)})_{l_1 l_2}| \lesssim 2^{k(l_1+l_2)}, \quad (l_1, l_2) \in Y'_d. \tag{B.13}$$

This follows by combining the bounds (B.9)–(B.12) and the identity

$$\begin{aligned} z_{l_1 l_2} - y_{l_1 l_2} + R_0(y^{(1)} - z^{(1)}, y^{(1)})_{l_1 l_2} &= x_{l_1 l_2} - y_{l_1 l_2} + R_0(y^{(1)} - x^{(1)}, y^{(1)})_{l_1 l_2} \\ &+ z_{l_1 l_2} - x_{l_1 l_2} + R_0(x^{(1)} - z^{(1)}, x^{(1)} - A_0^{(1)}(2^k u))_{l_1 l_2} \\ &+ R_0\big(x^{(1)} - z^{(1)} - A_0^{(1)}(2^k u) + A_0^{(1)}(2^k u), y^{(1)} - x^{(1)} + A_0^{(1)}(2^k u)\big)_{l_1 l_2}. \end{aligned}$$

This completes the proof of the lemma. □

Lemma B.5 *There is a constant $C_1 \ge 1$ such that for any $u \in [-2, 2]$, $x \in \mathbb{H}_Q$, and $n \in \mathbb{Z}$ satisfying $2^{n/2} \ge Q2^{\delta' w+3}$ there is a sequence of points $\{x_0, x_1, \dots, x_{w+10}\} \subseteq \mathbb{H}_Q$, $x = x_{w+10}$, with the following property: if $z \in \mathbb{H}_Q$, $k \le n$ satisfies $2^{k/2} \ge Q2^{\delta' w+1}$, and*

$$\big\{y \in \mathbb{H}_Q : \mathfrak{q}_\beta\big(z \cdot y^{-1} - A_0(2^k u)\big) < 2^k\big\} \subseteq B_{\beta,\mathbb{H}_Q}(x, 2^n), \tag{B.14}$$

then

$$B_{\beta,\mathbb{H}_Q}(z, 2^k) \subseteq \bigcup_{j \in \{0,\dots,w+10\}} B_{\beta,\mathbb{H}_Q}(x_j, C_1 2^n). \tag{B.15}$$

Proof For any $s \ge 0$ we define a point $x_s = \widetilde{x} \in \mathbb{H}_Q$ such that the inequalities

$$\begin{aligned} &\beta_{l0}|\widetilde{x}_{l0} - x_{l0} - (2^{n-s}u)^l| \le 2^{nl}, \\ &\beta_{l_1 l_2}\big|\widetilde{x}_{l_1 l_2} - x_{l_1 l_2} + R_0\big(x^{(1)} - \widetilde{x}^{(1)}, x^{(1)} + A_0^{(1)}(2^{n-s}u)\big)_{l_1 l_2} \\ &\quad + (2^{n-s}u)^{l_1+l_2}\big| \le 2^{n(l_1+l_2)}, \end{aligned} \tag{B.16}$$

for any $l \in \{1, \dots, d\}$ and any $(l_1, l_2) \in Y_d'$. Such a choice is possible because of the assumption $2^{n/2} \geq Q2^{\delta' w+4}$, and, in fact, we can set $x_s = x$ if $s \geq 10 + w$.

Given these points $\{x_0, \dots, x_{w+10}\}$, assume now that $k = n - s$, $s \geq 0$, is an integer and $z \in \mathbb{H}_Q$ is a point such that the inclusion (B.14) holds. With $\widetilde{x} = x_s$ we would like to show that $B_{\beta,\mathbb{H}_Q}(z, 2^k) \subseteq B_{\beta,\mathbb{H}_Q}(\widetilde{x}, C_1 2^n)$. In view of Lemma B.2 it suffices to show that

$$\mathfrak{q}_\beta(z \cdot \widetilde{x}^{-1}) \lesssim 2^n. \tag{B.17}$$

To see this we fix a point $y \in \mathbb{H}_Q$ such that $\mathfrak{q}_\beta\big(z \cdot y^{-1} - A_0(2^k u)\big) \leq 2^k$, and notice that $z \cdot \widetilde{x}^{-1} = E + I$, where $\mathfrak{q}_\beta(E) \lesssim 2^n$ and $I = A_0(2^k u) \cdot y \cdot \widetilde{x}^{-1}$ satisfies

$$\begin{aligned} I^{(1)} &= y^{(1)} - \widetilde{x}^{(1)} + A_0^{(1)}(2^k u), \\ I^{(2)} &= y^{(2)} - \widetilde{x}^{(2)} + R_0(\widetilde{x}^{(1)}, \widetilde{x}^{(1)}) + R_0(A_0^{(1)}(2^k u), y^{(1)}) \\ &\quad - R_0(A_0^{(1)}(2^k u) + y^{(1)}, \widetilde{x}^{(1)}). \end{aligned}$$

We would like to see that $\mathfrak{q}_\beta(I) \lesssim 2^n$. Since $y \in B_{\beta,\mathbb{H}_Q}(x, 2^n)$ we have

$$\begin{aligned} &\beta_{l0}|x_{l0} - y_{l0}| < 2^{nl} \quad l \in \{1, \dots, d\} \\ &\beta_{l_1 l_2}|x_{l_1 l_2} - y_{l_1 l_2} + R_0(y^{(1)} - x^{(1)}, y^{(1)})_{l_1 l_2}| < 2^{n(l_1+l_2)} \quad (l_1, l_2) \in Y_d', \end{aligned}$$

see (B.7). Combining these inequalities with (B.16) and recalling that $\beta_{l0} \gtrsim \beta_{l_1 l_2} \geq 1$ it follows easily that $\mathfrak{q}_\beta(I) \lesssim 2^n$, as desired. □

Now we are ready to complete the proof of Theorem B.1.

Proof of Theorem B.1 **Step 1.** We define an auxiliary maximal function

$$\begin{aligned} \widetilde{M}_{Q,w} f(h) := &\sup_{h \in B_{\beta,\mathbb{H}_Q}(g,2^k),\, 2^{k/2} \geq Q2^{w/8}} |B_{\beta,\mathbb{H}_Q}(g, 2^k)|^{-1} \\ &\times \sum_{y \in B_{\beta,\mathbb{H}_Q}(g,2^k)} |f(y)|, \quad h \in \mathbb{H}_Q, \end{aligned}$$

where the supremum is taken over all the quasi-balls $B_{\beta,\mathbb{H}_Q}(g, 2^k)$ that contain h. For any $f \in \ell^1(\mathbb{H}_Q)$ and $\lambda > 0$ we define the set

$$\mathcal{O}_\lambda := \{h \in \mathbb{H}_Q : \widetilde{M}_{Q,w} f(h) \geq \lambda\}. \tag{B.18}$$

By a standard Vitali covering argument (using also Lemma B.2 (b)) we can select a maximal finite family of disjoint balls $B^j_{\beta,\mathbb{H}_Q} = B_{\beta,\mathbb{H}_Q}(g_j, 2^{k_j})$, $2^{k_j/2} \geq Q2^{w/8}$, $j \in J(\lambda, f)$, such that

$$\begin{aligned} &|B^j_{\beta,\mathbb{H}_Q}|^{-1} \sum_{y\in B^j_{\beta,\mathbb{H}_Q}} |f(y)| \geq \lambda \qquad \text{for any } j \in J(\lambda, f),\\ &\bigcup_{j\in J(\lambda,f)} B^j_{\beta,\mathbb{H}_Q} \subseteq \mathcal{O}_\lambda \subseteq \bigcup_{j\in J(\lambda,f)} \widetilde{B}^j_{\beta,\mathbb{H}_Q}, \end{aligned} \tag{B.19}$$

where $\widetilde{B}^j_{\beta,\mathbb{H}_Q} = B^j_{\beta,\mathbb{H}_Q}(g_j, C_2 2^{k_j})$ is a fixed multiple of the quasi-ball $B^j_{\beta,\mathbb{H}_Q}$ for a suitable constant $C_2 \geq 1$. In particular,

$$|\mathcal{O}_\lambda| \simeq \sum_{j\in J(\lambda,f)} |\widetilde{B}^j_{\beta,\mathbb{H}_Q}| \simeq \sum_{j\in J(\lambda,f)} |B^j_{\beta,\mathbb{H}_Q}| \lesssim \|f\|_{\ell^1(\mathbb{H}_Q)}/\lambda, \tag{B.20}$$

so the operator $\widetilde{M}_{Q,w}$ is a bounded operator from $\ell^1(\mathbb{H}_Q)$ to $\ell^{1,\infty}(\mathbb{H}_Q)$, uniformly in Q and w.

Step 2. To complete the proof of the theorem it suffices to show that there is a constant $C_3 \geq 1$ sufficiently large such that

$$|\{h \in \mathbb{H}_Q\colon M_{Q,w,u} f(h) \geq C_3\lambda\}| \lesssim (1+w)|\{h \in \mathbb{H}_Q\colon \widetilde{M}_{Q,w} f(h) \geq \lambda\}|, \tag{B.21}$$

for every $\lambda > 0$. Using the definition (B.3), we see that if $M_{Q,w,u} f(z) \geq C_3\lambda$ then there is an integer k satisfying $2^{k/2} \geq 8Q2^{w/8}$ such that

$$\Big(\prod_{(l_1,l_2)\in Y_d} Q\beta_{l_1 l_2} 2^{-k(l_1+l_2)}\Big) \sum_{\{y\in\mathbb{H}_Q\colon \mathfrak{q}_\beta(z\cdot y^{-1} - A_0(2^k u)) < 2^k\}} |f(y)| \geq C_3\lambda. \tag{B.22}$$

Using Lemma B.4 we know that there is $\widetilde{z} \in \mathbb{H}_Q$ such that

$$\big\{y \in \mathbb{H}_Q\colon \mathfrak{q}_\beta(z\cdot y^{-1} - A_0(2^k u)) < 2^k\big\} \subseteq B_{\beta,\mathbb{H}_Q}(\widetilde{z}, C_0 2^k). \tag{B.23}$$

Using Lemma B.3 and (B.22), and assuming that C_3 is sufficiently large it follows that

$$|B_{\beta,\mathbb{H}_Q}(\widetilde{z}, 2^{k+a})|^{-1} \sum_{y\in B_{\beta,\mathbb{H}_Q}(\widetilde{z},2^{k+a})} |f(y)| \geq 2\lambda, \tag{B.24}$$

where a is the smallest integer with the property that $2^a \geq C_0$. Therefore $B_{\beta,\mathbb{H}_Q}(\widetilde{z}, 2^{k+a}) \subseteq \mathcal{O}_\lambda$ (see the definition (B.18)), so the ball $B_{\beta,\mathbb{H}_Q}(\widetilde{z}, 2^{k+a})$ intersects one of the selected balls $B^j_{\beta,\mathbb{H}_Q}$ for some $j \in J(\lambda, f)$. Therefore

$$B_{\beta,\mathbb{H}_Q}(\widetilde{z}, 2^{k+a}) \subseteq \widetilde{B}^j_{\beta,\mathbb{H}_Q} \subseteq B_{\beta,\mathbb{H}_Q}(g_j, 2^{k_j+b}) \qquad \text{for some } j \in J(\lambda, f), \tag{B.25}$$

where $b \in \mathbb{N}$ is a universal constant such that $C_2 \leq 2^b$ and $k + a \leq k_j + b$.

On the other hand, we use Lemma B.5 (with $n = k_j + b$ and $x = g_j$), starting from the inclusion (B.23), and (B.15), so

$$z \in \bigcup_{i\in\{0,\dots,w+10\}} B_{\beta,\mathbb{H}_Q}(g^i_j, C_1 2^{k_j+b}),$$

for suitable points $g^i_j \in \mathbb{H}_Q$ (that do not depend on k). Consequently we get

$$\{z \in \mathbb{H}_Q \colon M_{Q,w,u} f(z) \geq C_3\lambda\} \subseteq \bigcup_{j\in J(\lambda,f)} \bigcup_{i\in\{0,\dots,w+10\}} B_{\beta,\mathbb{H}_Q}(g^i_j, C_1 2^{k_j+b}),$$

The desired estimate (B.21) follows using also (B.20), which completes the proof of the theorem. □

References

1. Austin, T.: A proof of Walsh's convergence theorem using couplings. Int. Math. Res. Not. IMRN **15**, 6661–6674 (2015)
2. Austin, T.: On the norm convergence of non-conventional ergodic averages. Ergod. Theory Dyn. Syst. **30**, 321–338 (2010)
3. Bellow, A.: Measure theory Oberwolfach 1981. In: Kölzow, D., Maharam-Stone, D. (eds.) Proceedings of the Conference Held at Oberwolfach, June 21–27, 1981. Lecture Notes in Mathematics, vol. 945, pp. 429–431. Springer, Berlin Heidelberg (1982) . Section: Two problems submitted by A. Bellow
4. Bergelson, V.: Weakly mixing PET. Ergod. Theory Dyn Syst. **7**(3), 337–349 (1987)
5. Bergelson, V.: Ergodic Ramsey Theory—an update. In: Pollicott, M., Schmidt, K. (eds.) Ergodic Theory of $\mathbb{Z}^d$-Actions. London Math. Soc. Lecture Note Series, vol. 228, pp. 1–61 (1996)
6. Bergelson, V.: Combinatorial and diophantine applications of ergodic theory (with appendices by A. Leibman and by A. Quas and M. Wierdl). In: Hasselblatt, B., Katok, A. (eds.) Handbook of Dynamical Systems, vol. 1B, pp. 745–841. Elsevier, Amsterdam (2006)
7. Bergelson, V., Leibman, A.: Polynomial extensions of van der Waerden's and Szemerédi's theorems. J. Am. Math. Soc. **9**, 725–753 (1996)
8. Bergelson, V., Leibman, A.: A nilpotent Roth theorem. Invent. Math. **147**, 429–470 (2002)
9. Birch, B.J.: Forms in many variables. Proc. R. Soc. Lond. A **265**, 245–263 (1962)
10. Birkhoff, G.: Proof of the ergodic theorem. Proc. Natl. Acad. Sci. U.S.A. **17**(12), 656–660 (1931)

11. Bourgain, J.: On the maximal ergodic theorem for certain subsets of the integers. Isr. J. Math. **61**, 39–72 (1988)
12. Bourgain, J.: On the pointwise ergodic theorem on L^p for arithmetic sets. Isr. J. Math. **61**, 73–84 (1988)
13. Bourgain, J.: Pointwise ergodic theorems for arithmetic sets, with an appendix by the author, H. Furstenberg, Y. Katznelson and D.S. Ornstein. Inst. Hautes Études Sci. Publ. Math. **69**, 5–45 (1989)
14. Bourgain, J.: Double recurrence and almost sure convergence. J. Reine Angew. Math. **404**, 140–161 (1990)
15. Buczolich, Z., Mauldin, R.D.: Divergent square averages. Ann. Math. **171**(3), 1479–1530 (2010)
16. Calderón, A.: Ergodic theory and translation invariant operators. Proc. Natl. Acad. Sci. U.S.A. **59**, 349–353 (1968)
17. Christ, M.: Hilbert transforms along curves: I. Nilpotent groups. Ann. Math. **122**(3), 575–596 (1985)
18. Christ, M.: A $T(b)$ theorem with remarks on analytic capacity and the Cauchy integral. Colloq. Math. **60**(61), 601–628 (1990)
19. Christ, M., Nagel, A., Stein, E.M., Wainger, S.: Singular and maximal Radon transforms: analysis and geometry. Ann. Math. (2) **150**(2), 489–577 (1999)
20. Chu, Q., Frantzikinakis, N., Host, B.: Ergodic averages of commuting transformations with distinct degree polynomial iterates. Proc. Lond. Math. Soc. **102**(5), 801–842 (2011)
21. Davenport, H.: Cubic forms in thirty-two variables. Philos. Trans. R. Soc. Lond. A **251**, 193–232 (1959)
22. Frantzikinakis, N.: Some open problems on multiple ergodic averages. Bull. Hell. Math. Soc. **60**, 41–90 (2016)
23. Frantzikinakis, N., Kra, B.: Polynomial averages converge to the product of integrals. Isr. J. Math. **148**, 267–276 (2005)
24. Furstenberg, H.: Ergodic behavior of diagonal measures and a theorem of Szemeredi on arithmetic progressions. J. Anal. Math. **31**, 204–256 (1977)
25. Furstenberg, H.: Problems session. In: Conference on Ergodic Theory and Applications. University of New Hampshire, Durham (1982)
26. Furstenberg, H.: Nonconventional ergodic averages. In: The legacy of John von Neumann (Hempstead, NY, 1988), Proceedings of a Symposium in Pure Mathematics, vol. 50, pp. 43–56. American Mathematical Society, Providence (1990)
27. Furstenberg, H., Weiss, B.: A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$. In: Convergence in Ergodic Theory and Probability (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ., vol. 5, pp. 193–227. de Gruyter, Berlin (1996)
28. Host, B., Kra, B.: Non-conventional ergodic averages and nilmanifolds. Ann. Math. **161**, 397–488 (2005)
29. Host, B., Kra, B.: Convergence of polynomial ergodic averages. Isr. J. Math. **149**, 1–19 (2005)
30. Hu, Y.-Q.: Polynomial maps and polynomial sequences in groups. Available at arXiv:2105.08000
31. Hu, Y.-Q.: Waring's Problem For Locally Nilpotent Groups: The Case of Discrete Heisenberg Groups. Available at arXiv:2011.06683
32. Ionescu, A., Magyar, Á., Stein, E.M., Wainger, S.: Discrete Radon transforms and applications to ergodic theory. Acta Math. **198**, 231–298 (2007)
33. Ionescu, A., Magyar, Á., Wainger, S.: Averages along polynomial sequences in discrete nilpotent Lie groups: singular Radon transforms. In: Advances in Analysis: The Legacy of Elias M. Stein. Princeton Mathematical Series 50, , pp. 146–188. Princeton Univ. Press, Princeton (2014)

34. Ionescu, A.D., Wainger, S.: L^p boundedness of discrete singular Radon transforms. J. Am. Math. Soc. **19**(2), 357–383 (2005)
35. Jones, R.L., Seeger, A., Wright, J.: Strong variational and jump inequalities in harmonic analysis. Trans. Am. Math. Soc. **360**, 6711–6742 (2008)
36. Kra, B.: Private communication (2021)
37. Krause, B.: Discrete analogoues in harmonic analysis: maximally monomially modulated singular integrals related to Carleson's theorem. Available at arXiv:1803.09431
38. Krause, B., Mirek, M., Tao, T.: Pointwise ergodic theorems for non-conventional bilinear polynomial averages. Ann. Math. **195**(3), 997–1109 (2022)
39. LaVictoire, P.: Universally L^1-bad arithmetic sequences. J. Anal. Math. **113**(1), 241–263 (2011)
40. Leibman, A.: Convergence of multiple ergodic averages along polynomials of several variables. Isr. J. Math. **146**, 303–315 (2005)
41. Lépingle, D.: La variation d'ordre p des semi-martingales. Z. Wahrscheinlichkeitstheorie Verw. Gebiete **36**, 295–316 (1976)
42. Magyar, Á., Stein, E.M., Wainger, S.: Discrete analogues in harmonic analysis: spherical averages. Ann. Math. **155**, 189–208 (2002)
43. Magyar, Á., Stein, E.M., Wainger, S.: Maximal operators associated to discrete subgroups of nilpotent Lie groups. J. Anal. Math. **101**, 257–312 (2007)
44. Mirek, M., Stein, E.M., Trojan, B.: $\ell^p(\mathbb{Z}^d)$-estimates for discrete operators of Radon type: variational estimates. Invent. Math. **209**(3), 665–748 (2017)
45. Mirek, M., Stein, E.M., Zorin-Kranich, P.: Jump inequalities via real interpolation. Math. Ann. **376**(1–2), 797–819 (2020)
46. Mirek, M., Stein, E.M., Zorin-Kranich, P.: A bootstrapping approach to jump inequalities and their applications. Anal. PDE **13**(2), 527–558 (2020)
47. Mirek, M., Stein, E.M., Zorin-Kranich, P.: Jump inequalities for translation-invariant operators of Radon type on $\mathbb{Z}^d$. Adv. Math. **365**, 57 (2020)
48. Muscalu, C., Schlag, W.: Classical and Multilinear Harmonic Analysis. Vol. II. Cambridge Studies in Advanced Mathematics, vol. 138. Cambridge University Press, Cambridge (2013)
49. Pierce, L.: Discrete fractional Radon transforms and quadratic forms. Duke Math. J. **161**, 69–106 (2012)
50. Pierce, L., Yung, P.-L.: A polynomial Carleson operator along the paraboloid. Rev. Mat. Iberoam. **35**, 339–422 (2019)
51. Stein, E.M.: Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals. With the assistance of Timothy S. Murphy. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III, Princeton University Press, Princeton (1993)
52. Stein, E.M., Wainger, S.: Discrete analogues in harmonic analysis, I: ℓ^2 estimates for singular Radon transforms. Am. J. Math. **121**, 1291–1336 (1999)
53. Szemerédi, E.: On sets of integers containing no k elements in arithmetic progression. Acta Arith. **27**, 199–245 (1975)
54. Tao, T.: Norm convergence of multiple ergodic averages for commuting transformations. Ergod. Theory Dyn. Syst. **28**, 657–688 (2008)
55. von Neumann, J.: Proof of the quasi-ergodic hypothesis. Proc. Natl. Acad. Sci. U.S.A. **18**, 70–82 (1932)
56. Walsh, M.: Norm convergence of nilpotent ergodic averages. Ann. Math. **175**(3), 1667–1688 (2012)
57. Ziegler, T.: Universal characteristic factors and Furstenberg averages. J. Am. Math. Soc. **20**, 53–97 (2007)