Bounds for anisotropic Carleson operators

Citation for published version:
Roos, J 2019, 'Bounds for anisotropic Carleson operators', *Journal of Fourier Analysis and Applications*, vol. 25, pp. 2324–2355. https://doi.org/10.1007/s00041-018-09657-7

Digital Object Identifier (DOI):
10.1007/s00041-018-09657-7

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of Fourier Analysis and Applications

BOUNDS FOR ANISOTROPIC CARLESON OPERATORS

JORIS ROOS

Abstract. We prove weak $(2,2)$ bounds for maximally modulated anisotropically homogeneous smooth multipliers on $\mathbb{R}^n$. These can be understood as generalizing the classical one-dimensional Carleson operator. For the proof we extend the time-frequency method by Lacey and Thiele to the anistropic setting. We also discuss a related open problem concerning Carleson operators along monomial curves.

1. Introduction

Let us consider $\mathbb{R}^n$ equipped with the anisotropic dilations given by

$$\delta_\lambda(x) = (\lambda^{\alpha_1} x_1, \dots, \lambda^{\alpha_n} x_n), \tag{1.1}$$

where $\alpha = (\alpha_1, \dots, \alpha_n)$ with $\alpha_i \geq 1$ for $i = 1, \dots, n$. We write $|\alpha| = \sum_{i=1}^n \alpha_i$ and fix the anisotropic norm

$$\rho(x) = \max\{|x_i|^{\frac{1}{\alpha_i}} \,:\, i = 1, \dots, n\}.$$

For an integer $\nu \geq 0$, we say that a function m on $\mathbb{R}^n$ is in the class $\mathscr{M}^\nu$ if

(a) m is bounded and contained in $C^\nu(\mathbb{R}^n \setminus \{0\})$, and
(b) $m(\delta_\lambda(\xi)) = m(\xi)$ for all $\xi \neq 0$ and $\lambda > 0$.

Let us denote

$$\|m\|_{\mathscr{M}^\nu} = \sup_{|\beta| \leq \nu} \sup_{\rho(\xi)=1} |\partial^\beta m(\xi)|.$$

Define the Carleson operator associated with the multiplier m as

$$\mathscr{C}_m f(x) = \sup_{N \in \mathbb{R}^n} \left| \int_{\mathbb{R}^n} \widehat{f}(\xi) e^{ix\xi} m(\xi - N) d\xi \right|.$$

Replacing α by $c\alpha$ for some scalar c does not modify the classes $\mathcal{M}^\nu$. Thus we make the assumption $\alpha_1 = 1$ for normalization. For technical reasons we also assume that the α_i are positive integers.

Then we can state our main result as follows.

Theorem 1.1. *Let $\nu_0 \geq 3|\alpha| + 2$ be an integer. There exists $C > 0$ depending only on α and ν_0 such that for all $m \in \mathscr{M}^{\nu_0}$ we have*

$$\|\mathscr{C}_m f\|_{2,\infty} \leq C \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2. \tag{1.2}$$

Date: July 4, 2018.

2010 *Mathematics Subject Classification.* 42B20, 42B25, 44A12.

Key words and phrases. Carleson operators, anisotropic dilations, time-frequency analysis, singular integrals.

The proof of this theorem is based on the time-frequency techniques of Lacey and Thiele [LT00]. In the one-dimensional case $n = 1, \alpha = 1$ we recover the weak $(2,2)$ bound for Carleson's operator, which immediately implies Carleson's theorem on almost everywhere convergence of Fourier series [Car66] (up to a standard transference argument, see [KT80]). In the isotropic case $\alpha = (1, \dots, 1)$ the theorem follows from a result of Sjölin [Sjö71]. Pramanik and Terwilleger [PT03] study weak $(2,2)$ bounds for the isotropic case in $\mathbb{R}^n$ using the method of Lacey and Thiele. This is extended to strong (p,p) for $1 < p < \infty$ by Grafakos, Tao and Terwilleger in [GTT04]. We speculate that Theorem 1.1 could also be extended to strong (p,p) for $1 < p < \infty$ using the methods from [GTT04]. However we don't pursue this here to keep the exposition simple.

It would be interesting to know if (1.2) holds for all $\nu_0 > \frac{|\alpha|}{2}$, which is a natural lower bound suggested by the anisotropic Hörmander-Mikhlin theorem (see [FR66]). The same question is also open in the isotropic case. It seems plausible that (1.2) should at least hold for all $\nu_0 \geq |\alpha|+1$, because curiously, the only place in the current proof that requires more than that is a tail estimate in the single tree lemma (see (7.5)). All the main terms can be bounded using only $\nu_0 \geq |\alpha|+1$.

The method of Lacey and Thiele involves several ingredients. The first step is a reduction to a discrete dyadic model operator that involves summation over certain regions in phase space which are called tiles. This is detailed in Section 3. In this step we encounter a complication which is caused by the absence of rotation invariance in the anisotropic case. We resolve this using an anisotropic cone decomposition (see Lemma 3.3). The next step is a certain procedure of combinatorial nature the purpose of which is to organize the tiles into certain collections (which are called trees) each of which is associated with a component of the operator that behaves more like a classical singular integral operator (see Lemma 4.4). The combinatorial part of the argument requires only little modification compared to the original procedure in [LT00] (see Sections 4, 5, 6). The core component and most difficult part of the proof is the single tree estimate (see Section 7). Due to an extra dependence on the linearizing function, the estimate is more technical than the corresponding estimate in [LT00]. It is complicated further by the presence of anisotropic dilations.

In Section 2 we discuss a related open problem on Carleson operators along monomial curves, which served as a main motivation for this work. We demonstrate how Theorem 1.1 can be applied to a certain family of rougher multipliers that can be seen as a toy model for the Carleson operators along monomial curves. We also discuss the particular case of the parabolic Carleson operator (2.4), which exhibits some additional symmetries and is related to Lie's quadratic Carleson operator [Lie09]. Some partial progress on Carleson operators along monomial curves was obtained in [GPRY17]. Also see the related work [PY15] on Carleson operators along paraboloids in dimensions $n \geq 3$.

Acknowledgments. The author thanks his doctoral advisor Christoph Thiele for encouragement and countless valuable comments and discussions on this project. He is also deeply indebted to Po-Lam Yung for many important and fruitful conversations. This work was carried out while the author was supported by the German Academic Scholarship Foundation.

2. Carleson operators along monomial curves

For a positive integer $d \geq 2$, let us consider the multiplier of the Hilbert transform along the curve (t, t^d) in the form

$$m_d(\xi,\eta) = p.v. \int_{\mathbb{R}} e^{i\xi t - i\eta t^d} \frac{dt}{t}, \ (\xi,\eta) \in \mathbb{R}^2. \tag{2.1}$$

It is currently an open problem to decide whether $\mathscr{C}_{m_d}$ satisfies any L^p bounds. The multiplier m_d satisfies the anisotropic dilation symmetry

$$m_d(\lambda\xi, \lambda^d\eta) = m_d(\xi,\eta)$$

for $\lambda > 0$ and $(\xi,\eta) \neq 0$. However, Theorem 1.1 does not apply because m_d is too rough to be in the class $\mathscr{M}^\nu$ for any positive integer ν.

Next we discuss a family of toy model operators. In the following discussion we focus on the intersection of the quadrant $\{\xi \geq 0, \eta \geq 0\}$ with the region $\eta^{\frac{1}{d}} \leq 2\xi$. The other quadrants can be treated similarly (though depending on the parity of d the phase might not have a critical point in each quadrant; this is an inconsequential subtlety that we will ignore). Our restriction to the region $\eta^{\frac{1}{d}} \leq 2\xi$ is natural because stationary phase considerations show that m_d is smooth away from the axis $\eta = 0$. For $\xi > 0$ and $\eta > 0$ we define

$$m_{d,1}(\xi,\eta) = \left(\eta^{\frac{1}{d}}\xi^{-1}\right)^{\frac{d'}{2}} e^{i\left(\eta^{\frac{1}{d}}\xi^{-1}\right)^{-d'}} \psi(\eta^{\frac{1}{d}}\xi^{-1}), \tag{2.2}$$

where $\frac{1}{d} + \frac{1}{d'} = 1$ and ψ is a smooth cutoff function supported in $[-2, 2]$ and equal to one on a slightly smaller interval. We extend $m_{d,1}$ continuously by setting it equal to zero on the remainder of $\mathbb{R}^2$.

From a standard computation using the stationary phase principle (see [Ste93, Ch. VIII.1, Prop. 3]) we can see that, up to a negligible constant rescaling, this term constitutes the main contribution to the oscillatory integral in (2.1). The remainder term from stationary phase is smoother in the variables ξ and η and is therefore simpler to handle. We will ignore it for the purpose of this discussion.

From the definition we see directly that $m_{d,1}$ (and therefore also m_d) is only Hölder continuous of class $C^{\frac{1}{2(d-1)}}$ along the axis $\eta = 0$ while it is infinitely differentiable away from that axis.

Let us from here on denote

$$\zeta = \zeta(\xi,\eta) = \eta^{\frac{1}{d}}\xi^{-1}$$

for $\xi > 0, \eta > 0$. Since we do not know how to handle $m_{d,1}$ we introduce a family of modified, less oscillatory multipliers which is defined on the quadrant

$\{\xi > 0, \eta > 0\}$ by

$$m_{d,\delta}(\xi,\eta) = \zeta^{\frac{d'}{2}} e^{i\zeta^{-d'\delta}} \psi(\zeta), \tag{2.3}$$

where $0 \leq \delta \leq 1$ is a parameter (and we again extend $m_{d,\delta}$ to the rest of $\mathbb{R}^2$ by zero). The multiplier $m_{d,\delta}$ still fails to be in $\mathscr{M}^\nu$ for every positive integer ν and $\delta > 0$. However, we can nevertheless apply Theorem 1.1 to bound $\mathscr{C}_{m_{d,\delta}}$ for small enough δ. For this purpose we assume that ψ takes the form

$$\psi = \sum_{j \leq 0} \varphi_j,$$

where $\varphi_j(x) = \varphi(2^{-j}x)$ and φ is a smooth bump function supported in $[1/2, 2]$ and satisfying $\sum_{j\in\mathbb{Z}} \varphi_j(x) = 1$ for all $x \neq 0$. Then we have the following consequence of Theorem 1.1.

Corollary 2.1. *There exists $\delta_0 > 0$ such that for all $0 \leq \delta < \delta_0$ we have*

$$\|\mathscr{C}_{m_{d,\delta}} f\|_{2,\infty} \lesssim \|f\|_2.$$

Proof. Let us write

$$m_{d,\delta,j}(\xi,\eta) = \zeta^{\frac{d'}{2}} e^{i\zeta^{-d'\delta}} \varphi_j(\zeta) \quad (\text{for } \xi > 0, \eta > 0),$$

$$T_j f(x,y) = \int_{\mathbb{R}^2} \widehat{f}(\xi,\eta) m_{d,\delta,j}(\xi,\eta) e^{ix\xi + iy\eta} d(\xi,\eta).$$

By the change of variables $\eta \mapsto 2^{jd}\eta$, we see that

$$T_j f(x,y) = 2^{j\frac{d'}{2}} 2^{jd} \int_{\mathbb{R}^2} \widehat{f}(\xi, 2^{jd}\eta) \widetilde{m}_{d,\delta,j}(\xi,\eta) e^{ix\xi + i2^{jd}y\eta} d(\xi,\eta) = 2^{j\frac{d'}{2}} \mathrm{D}_{2^{jd}} \widetilde{T}_j \mathrm{D}_{2^{-jd}} f(x,y),$$

where $\mathrm{D}_\lambda f(x,y) = f(x, \lambda y)$ and

$$\widetilde{m}_{d,\delta,j}(\xi,\eta) = \zeta^{\frac{d'}{2}} e^{i2^{-jd'\delta}\zeta^{-d'\delta}} \varphi(\zeta) \quad (\text{for } \xi > 0, \eta > 0),$$

$$\widetilde{T}_j f(x,y) = \int_{\mathbb{R}^2} \widehat{f}(\xi,\eta) \widetilde{m}_{d,\delta,j}(\xi,\eta) e^{ix\xi + iy\eta} d(\xi,\eta).$$

We have

$$\|\widetilde{m}_{d,\delta,j}\|_{\mathscr{M}^\nu} \lesssim 2^{-jd'\delta\nu}$$

for every integer $\nu \geq 0$ (where the implied constant depends on ν, d, δ and φ). Using Theorem 1.1 we therefore obtain

$$\|\mathscr{C}_{m_{d,\delta}}\|_{L^2 \to L^{2,\infty}} \lesssim \sum_{j \leq 0} 2^{j\frac{d'}{2}} \|\mathscr{C}_{\widetilde{m}_{d,\delta,j}}\|_{L^2 \to L^{2,\infty}} \lesssim \sum_{j \leq 0} 2^{jd'(\frac{1}{2} - \delta\nu_0)}.$$

Thus, setting $\delta_0 = \frac{1}{2}\nu_0^{-1}$ yields the claim. $\square$

An improvement for the bound on ν_0 in Theorem 1.1 will give an improvement for δ_0. However, even if we could show Theorem 1.1 for all $\nu_0 > |\alpha|/2$, we would still have $\delta_0 < \frac{1}{|\alpha|} < 1$. Therefore additional insight is likely to be required to bound the operator $\mathscr{C}_{m_{d,1}}$ (and thus $\mathscr{C}_{m_d}$).

For the case of the parabola, $d = 2$, there are some additional obstructions in bounding $\mathscr{C}_{m_2}$. Let us write the parabolic Carleson operator as

$$\mathscr{C}^{\mathrm{par}} f(x,y) = \sup_{N\in\mathbb{R}^2} \left| p.v. \int_{\mathbb{R}} f(x-t, y-t^2) e^{iN_1 t + iN_2 t^2} \frac{dt}{t} \right|. \tag{2.4}$$

Apart from the linear modulation symmetries given by

$$\mathscr{C}^{\mathrm{par}} f = \mathscr{C}^{\mathrm{par}} \mathrm{M}_N f$$

for $N \in \mathbb{R}^2$, there are additional modulation symmetries. For a polynomial in two variables, $P = P(x, y)$, we write the corresponding polynomial modulation as

$$\mathrm{M}_P f(x,y) = e^{iP(x,y)} f(x,y).$$

Then we have that

$$\begin{aligned} \mathscr{C}^{\mathrm{par}} \mathrm{M}_{Nx^2} f &= \mathscr{C}^{\mathrm{par}} f, \\ \mathscr{C}^{\mathrm{par}} \mathrm{M}_{Nx(y+x^2)} f &= \mathscr{C}^{\mathrm{par}} f, \\ \mathscr{C}^{\mathrm{par}} \mathrm{M}_{N(y+x^2)^2} f &= \mathscr{C}^{\mathrm{par}} f \end{aligned} \tag{2.5}$$

hold for all $N \in \mathbb{R}$. The quadratic modulation symmetry (2.5) suggests a connection to Lie's quadratic Carleson operator [Lie09]. Indeed, even a certain partial L^2 bound for $\mathscr{C}^{\mathrm{par}}$ would immediately imply an L^2 bound for the quadratic Carleson operator (see [GPRY17]). These are all the polynomial modulation symmetries of the operator $\mathscr{C}^{\mathrm{par}}$ (up to linear combination). To see that, introduce the change of variables $\tau(x, y) = (x, y + x^2)$ and observe that

$$\mathscr{C}^{\mathrm{par}} f = \mathscr{C}^{\mathrm{sh}}(f \circ \tau^{-1}) \circ \tau,$$

where

$$\mathscr{C}^{\mathrm{sh}} f(x,y) = \sup_{N\in\mathbb{R}^2} \left| p.v. \int_{\mathbb{R}} f(x-t, y-2xt) e^{iN_1 t + iN_2 t^2} \frac{dt}{t} \right|.$$

It is easy to check that we have

$$\mathscr{C}^{\mathrm{sh}} \mathrm{M}_P f = \mathscr{C}^{\mathrm{sh}} f$$

for a polynomial P if and only if P is of degree at most two. This shows that the list of polynomial modulation symmetries that we gave for $\mathscr{C}^{\mathrm{par}}$ is complete. Also since τ is measure preserving, L^p bounds for $\mathscr{C}^{\mathrm{par}}$ and $\mathscr{C}^{\mathrm{sh}}$ are equivalent.

3. Reduction to a model operator

Before we begin we need to introduce some more notation and definitions. Denote

$$\mathrm{dist}_\alpha(A, B) = \inf_{x\in A, y\in B} \rho(x-y)$$

and $\mathrm{dist}_\alpha(A, x) = \mathrm{dist}_\alpha(A, \{x\})$. For $a, b \in \mathbb{R}^n$ we write

$$[a, b] = \prod_{i=1}^{n} [a_i, b_i]$$

and similarly $(a, b), [a, b)$. We will refer to all such sets as *rectangles*. For a rectangle $I \subset \mathbb{R}^n$ we define $c(I)$ to be its center. By an *anisotropic cube* we mean a rectangle $[a,b]$ such that $b_i - a_i = \lambda^{\alpha_i}$ holds for all $i = 1,\dots,n$ and some $\lambda > 0$. We define the collection of *anisotropic dyadic cubes* by

$$\mathcal{D}^\alpha = \{[\delta_{2^k}(\ell), \delta_{2^k}(\ell+1))\,:\, \ell \in \mathbb{Z}^n,\, k \in \mathbb{Z}\}\,.$$

Every two anisotropic dyadic cubes have the property that they are either disjoint or contained in one another. Moreover, for every $I \in \mathcal{D}^\alpha$ there exists a unique dyadic cube $I^+ \in \mathcal{D}^\alpha$ such that $|I^+| = 2^{|\alpha|}|I|$ and $I \subset I^+$. We call I^+ the *parent* of I and say that I is a *child* of I^+.

Definition 3.1. A *tile* P is a rectangle in $\mathbb{R}^n \times \mathbb{R}^n$ of the form

$$P = I_P \times \omega_P,$$

where $I_P, \omega_P \in \mathcal{D}^\alpha$ and $|I_P| \cdot |\omega_P| = 1$.

The set of tiles is denoted by $\overline{\mathcal{P}}$. Given a tile P we denote its *scale* by $k_P = |I_P|^{1/|\alpha|}$. For $r \in \{0,1\}^n$ and a tile P with $\omega_P = [\delta_{2^{-k_P}}(\ell), \delta_{2^{-k_P}}(\ell+1)]$ we define the *semi-tile* $P(r)$ by

$$P(r) = I_P \times \omega_{P(r)},\ \text{where } \omega_{P(r)} = \Big[\delta_{2^{-k_P}}(\ell + \frac{1}{2}r), \delta_{2^{-k_P}}(\ell + \frac{1}{2}(r+1))\Big].$$

The model operator is built up using a large family of wave packets adapted to tiles. It is convenient to generate this family by letting the symmetry group of our operator act on a single bump function. For this purpose, let ϕ be a Schwartz function on $\mathbb{R}^n$ such that $0 \le \widehat{\phi} \le 1$ with $\widehat{\phi}$ being supported in $[-\frac{b_0}{2}, \frac{b_0}{2}]^n$ and equal to 1 on $[-\frac{b_1}{2}, \frac{b_1}{2}]^n$, where $0 < b_1 < b_0 \ll 1$ are some fixed, small numbers whose ratio is not too large (it becomes clear what precisely is required in Section 7). For example, we may set $b_0 = \frac{1}{10}$, $b_1 = \frac{9}{100}$. We denote translation, modulation and dilation of a function f by

$$\mathrm{T}_y f(x) = f(x-y), \qquad (y \in \mathbb{R}^n)$$
$$\mathrm{M}_\xi f(x) = e^{ix\xi} f(x), \qquad (\xi \in \mathbb{R}^n)$$
$$\mathrm{D}^p_\lambda f(x) = \lambda^{-\frac{|\alpha|}{p}} f\left(\delta_{\lambda^{-1}}(x)\right), \qquad (\lambda, p > 0),$$

where $|\alpha| = \sum_{i=1}^n \alpha_i$.

Given a tile P and $N \in \mathbb{R}^n$ we define the wave packets ϕ_P, ψ_P^N on $\mathbb{R}^n$ by

$$\phi_P(x) = \mathrm{M}_{c(\omega_{P(0)})} \mathrm{T}_{c(I_P)} \mathrm{D}^2_{2^{k(P)}} \phi(x) \tag{3.1}$$

$$\widehat{\psi_P^N}(\xi) = \mathrm{T}_N m(\xi) \cdot \widehat{\phi_P}(\xi) \tag{3.2}$$

We think of ϕ_P as being essentially time-frequency supported in the semi-tile $P(0)$. More precisely, we have that $\widehat{\phi_P}$ is compactly supported in (a small cube centrally contained in) $\omega_{P(0)}$ and $|\phi_P|$ decays rapidly outside of I_P.

For $N \in \mathbb{R}^n$ and $r \neq 0$ we introduce the dyadic model sum operator

$$A_N^{r,m} f(x) = \sum_{P \in \overline{\mathcal{P}}} \langle f, \phi_P \rangle \psi_P^N(x) \mathbf{1}_{\omega_{P(r)}}(N). \tag{3.3}$$

This reduces to the model sum of Lacey and Thiele [LT00] in the case $n = \alpha = 1$ and to that of Pramanik and Terwilleger [PT03] in the isotropic case $\alpha = (1,\dots,1)$.

Theorem 3.2. *For every large enough integer ν_0 there exists $C > 0$ depending only on ν_0, α and the choice of ϕ such that for all multipliers $m \in \mathscr{M}^{\nu_0}$ we have*

$$\| \sup_{N\in\mathbb{R}^n} |A_N^{r,m} f| \|_{2,\infty} \le C \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2. \tag{3.4}$$

The proof of the theorem is contained in Sections 4, 5, 6, 7. We conclude this section by showing that Theorem 3.2 implies Theorem 1.1. For this purpose we employ the averaging procedure of Lacey and Thiele [LT00] combined with an anisotropic cone decomposition of the multiplier m. By an *anisotropic cone* we mean a subset $\Theta \subsetneq \mathbb{R}^n$ of the form

$$\Theta = \{\delta_t(\xi) \,:\, t > 0,\, \xi \in Q\}$$

for some cube $Q \subset \mathbb{R}^n$. Let us denote $\mathcal{B}_s = \{x \,:\, \rho(x) \le s\}$. Let

$$\mathbf{A}^{r,m} f(x) = \lim_{R\to\infty} \frac{1}{R^{2|\alpha|}} \int_{\mathcal{B}_R} \int_{\mathcal{B}_R} \int_0^1 \mathrm{M}_{-\eta} \mathrm{T}_{-y} \mathrm{D}^2_{2^{-s}} A^{r,m}_{2^{-s}\eta} \mathrm{D}^2_{2^s} \mathrm{T}_y \mathrm{M}_\eta f(x) ds dy d\eta. \tag{3.5}$$

Lemma 3.3. *For every $r \in \{0,1\}^n$ and every test function f, the function $\mathbf{A}^{r,m} f(x)$ is well-defined and also a test function. We have*

$$\widehat{\mathbf{A}^{r,m} f}(\xi) = \theta_r(\xi) m(\xi) \widehat{f}(\xi)$$

for some smooth function θ_r that is independent of m. Moreover, there exists a constant $\varepsilon_0 > 0$ and an anisotropic cone Θ_r such that

$$\theta_r(\xi) > \varepsilon_0 \quad \textit{for all } \xi \in \Theta_r.$$

and

$$(-\infty, \varepsilon_0]^n \subset \bigcup_{r\in\{0,1\}^n\setminus\{0\}} \Theta_r. \tag{3.6}$$

Proof. By expanding definitions we see that

$$(\mathrm{M}_{-\eta} \mathrm{T}_{-y} \mathrm{D}^2_{2^{-s}} A^{r,m}_{2^{-s}\eta} \mathrm{D}^2_{2^s} \mathrm{T}_y \mathrm{M}_\eta f)^\wedge(\xi)$$

is equal to a universal constant times

$$m(\xi) \sum_{P\in\overline{\mathcal{P}}} \langle \widehat{f}, \mathrm{T}_{-\eta+\delta_{2^s}(c(\omega_{P(0)}))} \mathrm{M}_{y-\delta_{2^{-s}}(c(I_P))} \mathrm{D}^2_{2^{s-k_P}} \widehat{\phi} \rangle$$
$$\times\, \mathrm{T}_{-\eta+\delta_{2^s}(c(\omega_{P(0)}))} \mathrm{M}_{y-\delta_{2^{-s}}(c(I_P))} \mathrm{D}^2_{2^{s-k_P}} \widehat{\phi}(\xi) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)),$$

where we have used that $m(\delta_{2^{-s}}(\xi)) = m(\xi)$. The previous display equals

$$m(\xi) \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}^n} \sum_{u\in\mathbb{Z}^n} 2^{-|\alpha|(s-k)} \int_{\mathbb{R}^n} \widehat{f}(\zeta) e^{i(y-\delta_{2^{-s+k}}(u+\frac12))(\xi-\zeta)} \overline{\widehat{\phi}}\Big(\delta_{2^{-s+k}}(\zeta+\eta) - \Big(\ell + \frac14\Big)\Big) d\zeta$$
$$\times\, \widehat{\phi}\Big(\delta_{2^{-s+k}}(\xi+\eta) - \Big(\ell + \frac14\Big)\Big) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)).$$

Applying the Poisson summation formula to the summation in u and using the Fourier support information of the function ϕ we see that the previous display equals (up to a universal constant)

$$m(\xi) \widehat{f}(\xi) \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}^n} |\widehat{\phi}|^2 \Big(\delta_{2^{-s+k}}(\xi+\eta) - \Big(\ell + \frac14\Big)\Big) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)).$$

Observe that the expression no longer depends on the variable y. It remains to compute the function $\theta_r(\xi) = c \cdot \lim_{R\to\infty} I_R(\xi)$, where c is a universal constant and

$$I_R(\xi) = \frac{1}{R^{|\alpha|}} \int_{\mathcal{B}_R} \int_0^1 \sum_{k\in\mathbb{Z}} \sum_{\ell\in\mathbb{Z}^n} |\widehat{\phi}|^2 \Big(\delta_{2^{-s+k}}(\xi+\eta) - \Big(\ell + \frac{1}{4}\Big)\Big) \mathbf{1}_{\omega_{P(r)}}(\delta_{2^{-s}}(\eta)) ds d\eta.$$

Note the formula

$$\int_0^1 \sum_{k\in\mathbb{Z}} F(2^{k-s}) ds = \frac{1}{\log 2} \int_0^\infty F(t) \frac{dt}{t},$$

which follows from a change of variables $2^{k-s} \to t$. Using this we have

$$I_R(\xi) = \frac{c}{R^{|\alpha|}} \int_{\mathcal{B}_R} \int_0^\infty \sum_{\ell\in\mathbb{Z}^n} |\widehat{\phi}|^2 \Big(\delta_t(\xi+\eta) - \Big(\ell + \frac{1}{4}\Big)\Big) \mathbf{1}_{Q_r}(\delta_t(\eta) - \ell) \frac{dt}{t} d\eta,$$

where $Q_r = \left[\frac{1}{2}r, \frac{1}{2}(r+1)\right] = \prod_{i=1}^n \left[\frac{1}{2}r_i, \frac{1}{2}(r_i+1)\right]$ and $c = (\log 2)^{-1}$ (c may change from line to line in this proof). To simplify our expression further we perform the change of variables

$$\delta_t(\xi+\eta) - \ell \to \zeta$$

in the integration in η. This yields

$$I_R(\xi) = c \int_{\mathbb{R}^n} \int_0^\infty \chi(\zeta) \mathbf{1}_{Q_r}(\zeta - \delta_t(\xi)) \left(\sum_{\ell\in\mathbb{Z}^n} \frac{\mathbf{1}_{\rho(\zeta+\ell-\delta_t(\xi))\le tR}}{(tR)^{|\alpha|}} \right) \frac{dt}{t} d\zeta \tag{3.7}$$

where we have set

$$\chi(\zeta) = |\widehat{\phi}|^2 \Big(\zeta - \frac{1}{4}\Big).$$

Observe that the integrand in (3.7) is supported in a compact subset of $\mathbb{R}^n \times (0,\infty)$ (which depends on ξ). By counting the ℓ for which the summand is non-zero we see that for every fixed $\zeta, \xi \in \mathbb{R}^n$ and $t > 0$ the sum

$$\sum_{\ell\in\mathbb{Z}^n} \frac{\mathbf{1}_{\rho(\zeta+\ell-\delta_t(\xi))\le tR}}{(tR)^{|\alpha|}}$$

converges to a universal constant as $R \to \infty$. Thus, from Lebesgue's dominated convergence theorem we conclude that

$$\theta_r(\xi) = c \int_{\mathbb{R}^n} \int_0^\infty \chi(\zeta) \mathbf{1}_{Q_r}(\zeta - \delta_t(\xi)) \frac{dt}{t} d\zeta. \tag{3.8}$$

Evidently we have $\theta_r(\delta_t(\xi)) = \theta_r(\xi)$ for every $t > 0$ and $\xi \in \mathbb{R}^n$. From our choice of ϕ we get that χ is supported on $Q^{(0)}$ and equal to one on $Q^{(1)}$, where

$$Q^{(j)} = \Big[\frac{1}{4} - \frac{b_j}{2}, \frac{1}{4} + \frac{b_j}{2}\Big]$$

for $j = 0, 1$. Let us set

$$\Theta_r = \{\delta_t(\xi) \,:\, \xi \in Q^{(1)} - Q_r\}.$$

Then we can read off (3.8) that θ_r is greater than some positive constant on $\Theta_r^{(1)}$. Note that

$$Q^{(j)} - Q_r = \Big[-\frac{1}{2}r - \Big(\frac{1}{4} + \frac{b_j}{2}\Big), -\frac{1}{2}r + \Big(\frac{1}{4} + \frac{b_j}{2}\Big)\Big].$$

Looking at the anisotropic cone generated by each of the regions $Q^{(1)} - Q_r$ we see that (3.6) is satisfied for sufficiently small ε_0. □

In the isotropic case $\alpha = (1, \ldots, 1)$ we can assume without loss of generality that the multiplier m is supported in some arbitrarily chosen cone. Due to the lack of rotation invariance this assumption becomes invalid in the anisotropic setting.

Proof of Theorem 1.1. Let $m \in \mathscr{M}^{\nu_0}$. Without loss of generality we may assume that m is supported in the "quadrant" $(-\infty, 0]^n$. By (3.6) we can choose smooth functions $(\varrho_r)_r$ such that ϱ_r is supported in Θ_r and

$$\sum_{r \in \{0,1\}^n \setminus \{0\}} \varrho_r(\xi) = 1$$

for $\xi \in (-\infty, 0]^n$. By the triangle inequality and Lemma 3.3, we have

$$\|\mathscr{C}_m f\|_{2,\infty} \le \sum_{r \in \{0,1\}^n \setminus \{0\}} \| \sup_{N \in \mathbb{R}^n} |\mathbf{A}^{r, \theta_r^{-1} \varrho_r m} \mathrm{M}_N f| \|_{2,\infty}.$$

Here θ_r^{-1} refers to the function $\xi \mapsto (\theta_r(\xi))^{-1}$, which is bounded on Θ_r. By (3.5) and Minkowski's integral inequality, the previous is no greater than

$$\sum_{r \in \{0,1\}^n \setminus \{0\}} \limsup_{R \to \infty} \frac{1}{R^{2|\alpha|}} \int_{\mathcal{B}_R} \int_{\mathcal{B}_R} \int_0^1 \| \sup_{N \in \mathbb{R}^n} |A_N^{r, \theta_r^{-1} \varrho_r m} \mathrm{D}_{2^s}^2 \mathrm{T}_y \mathrm{M}_\eta f| \|_{2,\infty} ds dy d\eta,$$

which by Theorem 3.2 is bounded by

$$C \sum_{r \in \{0,1\}^n \setminus \{0\}} \|\theta_r^{-1} \varrho_r m\|_{\mathscr{M}^{\nu_0}} \|f\|_2 \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2.$$

□

4. Boundedness of the model operator

In this section we describe the proof of Theorem 3.2. We follow [LT00]. First, we perform some preliminary reductions. Given a measurable function $N : \mathbb{R}^n \to \mathbb{R}^n$ we define

$$Tf(x) = A^r_{N(x)} f(x).$$

Note that the estimate (3.4) is equivalent to showing

$$\|Tf\|_{2,\infty} \le C \|m\|_{\mathscr{M}^{\nu_0}} \|f\|_2$$

with C not depending on the choice of the measurable function N. By duality, it is equivalent to show

$$|\langle Tf, \mathbf{1}_E \rangle| \lesssim \|m\|_{\mathscr{M}^{\nu_0}} |E|^{\frac{1}{2}} \|f\|_2,$$

where E is an arbitrary measurable set. By scaling, we may assume without loss of generality that $\|f\|_2 = 1$ and $|E| \le 1$. Thus, by the triangle inequality, it suffices to show that

$$\sum_{P\in\mathcal{P}} |\langle f, \phi_P\rangle\langle \mathbf{1}_{E\cap N^{-1}(\omega_{P(r)})}, \psi_P^{N(\cdot)}\rangle| \lesssim \|m\|_{\mathscr{M}^{\nu_0}}, \tag{4.1}$$

for all finite sets of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$, with the implied constant being independent of $f, E, N, \mathcal{P}$. Throughout this and the following sections we fix $r \in \{0,1\}^n \setminus \{0\}$. Before we continue we need to introduce certain collections of tiles called trees. There is a partial order on tiles defined by

$$P \le P' \qquad \text{if} \qquad I_P \subset I_{P'} \quad \text{and} \quad c(\omega_{P'}) \in \omega_P.$$

Observe that two tiles are comparable with respect to $\le$ if and only if they have a non-empty intersection.

Definition 4.1. A finite collection $\mathbf{T} \subset \overline{\mathcal{P}}$ of tiles is called a *tree* if there exists $P \in \mathbf{T}$ such that $P' \le P$ for every $P' \in \mathbf{T}$. In that case, P is uniquely determined and referred to as the *top* of the tree $\mathbf{T}$. We denote the top of a tree $\mathbf{T}$ by $P_{\mathbf{T}} = I_{\mathbf{T}} \times \omega_{\mathbf{T}}$ and write $k_{\mathbf{T}} = |I_{\mathbf{T}}|^{1/|\alpha|}$.
A tree $\mathbf{T}$ is called a 1–*tree* if $c(\omega_{\mathbf{T}}) \notin \omega_{P(r)}$ for all $P \in \mathbf{T}$ and it is called a 2–*tree* if $c(\omega_{\mathbf{T}}) \in \omega_{P(r)}$ for all $P \in \mathbf{T}$. These names are due to historical reasons (see [LT00]).

The notion of a tree was first introduced in this context by C. Fefferman [Fef73]. For a tile $P \in \overline{\mathcal{P}}$ we write

$$E_P = E \cap N^{-1}(\omega_P) \text{ and } E_{P(r)} = E \cap N^{-1}(\omega_{P(r)}).$$

The *mass* of a single tile P is defined as

$$\mathcal{M}(P) = \sup_{P' \ge P} \int_{E_{P'}} w_{P'}^{\nu_1}(x)dx, \tag{4.2}$$

where ν_1 is a fixed large positive number depending only on $|\alpha|$ that is to be determined later and

$$w_P^{\nu}(x) = \mathrm{T}_{c(I_P)}\mathrm{D}^1_{2^{k(P)}} w^{\nu}(x),$$

where the weight w^ν takes the form

$$w^\nu(x) = (1 + \rho(x))^{-\nu}.$$

For convenience we also write $w_P = w_P^{\nu_1}$. For a collection of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$ we define their mass as

$$\mathcal{M}(\mathcal{P}) = \sup_{P\in\mathcal{P}} \mathcal{M}(P). \tag{4.3}$$

The *energy* of a collection of tiles $\mathcal{P}$ is defined as

$$\mathcal{E}(\mathcal{P}) = \sup_{\mathbf{T}\subset\mathcal{P}\ 2\text{–tree}} \left(\frac{1}{|I_T|} \sum_{P\in\mathbf{T}} |\langle f, \phi_P\rangle|^2 \right)^{1/2}. \tag{4.4}$$

These quantities and the following lemmas originate in [LT00].

Lemma 4.2 (Mass lemma). *If $\nu_1 > |\alpha| + 1$, then there exists $C > 0$ depending only on α such that for every finite set of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$ there is a decomposition $\mathcal{P} = \mathcal{P}_{\text{light}} \cup \mathcal{P}_{\text{heavy}}$ such that*

$$\mathcal{M}(\mathcal{P}_{\text{light}}) \leq 2^{-2}\, \mathcal{M}(\mathcal{P}) \tag{4.5}$$

and $\mathcal{P}_{\text{heavy}}$ is a union of a set $\mathcal{T}$ of trees such that

$$\sum_{\mathbf{T}\in\mathcal{T}} |I_{\mathbf{T}}| \leq \frac{C}{\mathcal{M}(\mathcal{P})}. \tag{4.6}$$

Lemma 4.3 (Energy lemma). *There exists $C > 0$ depending only on α such that for every finite set of tiles $\mathcal{P} \subset \overline{\mathcal{P}}$ there is a decomposition $\mathcal{P} = \mathcal{P}_{\text{low}} \cup \mathcal{P}_{\text{high}}$ such that*

$$\mathcal{E}(\mathcal{P}_{\text{low}}) \leq 2^{-1}\, \mathcal{E}(\mathcal{P}) \tag{4.7}$$

and $\mathcal{P}_{\text{high}}$ is a union of a set $\mathcal{T}$ of trees such that

$$\sum_{\mathbf{T}\in\mathcal{T}} |I_{\mathbf{T}}| \leq \frac{C}{\mathcal{E}(\mathcal{P})^2}. \tag{4.8}$$

Lemma 4.4 (Tree estimate). *There exists $C > 0$ depending only on α such that if $m \in \mathscr{M}^{\nu_0}$, then the following inequality holds for every tree $\mathbf{T}$:*

$$\sum_{P\in\mathbf{T}} |\langle f, \phi_P\rangle \langle \psi_P^{N(\cdot)}, \mathbf{1}_{E_{P(r)}}\rangle| \leq C\|m\|_{\mathscr{M}^{\nu_0}} |I_{\mathbf{T}}| \mathcal{E}(\mathbf{T})\mathcal{M}(\mathbf{T}) \tag{4.9}$$

The proofs of these lemmas are contained in Sections 5, 6, 7, respectively. By iterated application of these lemmas we obtain a proof of (4.1). This argument is literally the same as in [LT00], but we include it here for convenience of the reader. Let $\mathcal{P}$ be a finite collection of tiles. We will decompose $\mathcal{P}$ into disjoint sets $(\mathcal{P}_\ell)_{\ell\in\mathcal{N}}$ (where $\mathcal{N}$ is some finite set of integers) such that each $\mathcal{P}_\ell$ satisfies

$$\mathcal{M}(\mathcal{P}_\ell) \leq 2^{2\ell} \quad \text{and} \quad \mathcal{E}(\mathcal{P}_\ell) \leq 2^{\ell} \tag{4.10}$$

and is equal to the union of a set of trees $\mathcal{T}_\ell$ such that

$$\sum_{\mathbf{T}\in\mathcal{T}_\ell} |I_{\mathbf{T}}| \leq C 2^{-2\ell}. \tag{4.11}$$

This is achieved by the following procedure:

(1) Initialize $\mathcal{P}^{\text{stock}} := \mathcal{P}$ and choose an initial ℓ that is large enough such that

$$\mathcal{M}(\mathcal{P}^{\text{stock}}) \leq 2^{2\ell} \quad \text{and} \quad \mathcal{E}(\mathcal{P}^{\text{stock}}) \leq 2^{\ell}. \tag{4.12}$$

(2) If $\mathcal{M}(\mathcal{P}^{\text{stock}}) > 2^{2(\ell-1)}$, then apply Lemma 4.2 to decompose $\mathcal{P}^{\text{stock}}$ into $\mathcal{P}_{\text{light}}$ and $\mathcal{P}_{\text{heavy}}$. We add[1] $\mathcal{P}_{\text{heavy}}$ to $\mathcal{P}_\ell$ and update $\mathcal{P}^{\text{stock}} := \mathcal{P}_{\text{light}}$ (thus, we now have $\mathcal{M}(\mathcal{P}^{\text{stock}}) \leq 2^{2(\ell-1)}$).
(3) If $\mathcal{E}(\mathcal{P}^{\text{stock}}) > 2^{\ell-1}$, then apply Lemma 4.3 to decompose $\mathcal{P}^{\text{stock}}$ into $\mathcal{P}_{\text{low}}$ and $\mathcal{P}_{\text{high}}$. We add $\mathcal{P}_{\text{high}}$ to $\mathcal{P}_\ell$ and update $\mathcal{P}^{\text{stock}} := \mathcal{P}_{\text{low}}$ (thus, we now have $\mathcal{E}(\mathcal{P}^{\text{stock}}) \leq 2^{\ell-1}$).
(4) If $\mathcal{P}^{\text{stock}}$ is not empty, then replace ℓ by $\ell - 1$ and go to Step (2).

[1] We can think of all the $\mathcal{P}_\ell$ as being initialized by the empty set.

Then we can finish the proof of (4.1) by using (4.10), (4.11), (4.9) and keeping in mind that we always have $\mathcal{M}(\mathcal{P}) \leq \|w^{\nu_1}\|_1$:

$$\sum_{P\in\mathcal{P}} |\langle f,\phi_P\rangle\langle \mathbf{1}_{E\cap N^{-1}(\omega_{P(r)})},\psi_P^{N(\cdot)}\rangle| = \sum_{\ell\in\mathcal{N}}\sum_{\mathbf{T}\in\mathcal{T}_\ell}\sum_{P\in\mathbf{T}} |\langle f,\phi_P\rangle\langle \mathbf{1}_{E\cap N^{-1}(\omega_{P(r)})},\psi_P^{N(\cdot)}\rangle|$$

$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} \sum_{\ell\in\mathcal{N}} 2^\ell \min(1,2^{2\ell}) \sum_{\mathbf{T}\in\mathcal{T}_\ell} |I_{\mathbf{T}}| \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \sum_{\ell\in\mathbb{Z}} 2^{-\ell}\min(1,2^{2\ell}) \lesssim \|m\|_{\mathscr{M}^{\nu_0}}.$$

To conclude this section we collect several standard auxiliary estimates for m, K, ϕ_P, ψ_P^N which are used during the remainder of the proof. First, from the definition of $\mathscr{M}^\nu$ we have the symbol estimate

$$|\partial_i^\nu m(\xi)| \leq \|m\|_{\mathscr{M}^\nu}\rho(\xi)^{-\nu\alpha_i} \tag{4.13}$$

for every integer $\nu \leq \nu_0$ and $i = 1,\dots,n$. If we let K denote the corresponding kernel (that is, $\widehat{K} = m$), we have

$$|K(x)| \lesssim \|m\|_{\mathscr{M}^{\lfloor\frac{|\alpha|}{2}\rfloor+1}}\rho(x)^{-|\alpha|} \tag{4.14}$$

for $x \neq 0$. This is a consequence of the anisotropic Hörmander-Mikhlin theorem (see [FR66]). For every integer $\nu \geq 0$ and $N \notin \omega_{P(0)}$ we have

$$|\psi_P^N(x)| \lesssim \|m\|_{\mathscr{M}^\nu}|I_P|^{1/2}w_P^\nu(x), \tag{4.15}$$

where the implicit constant depends only on ν, α and the choice of ϕ. We defer the proof of this estimate to Section 8.

The next estimates concern the interaction of two wave packets associated with distinct tiles. Let $P, P' \in \overline{\mathcal{P}}$ be tiles. The idea is that if P, P' are disjoint (or equivalently, incomparable with respect to $\leq$) then their associated wave packets are almost orthogonal, i.e. $\langle\phi_P,\phi_{P'}\rangle$ is negligibly small. Indeed, if ω_P and $\omega_{P'}$ are disjoint, then we even have $\langle\phi_P,\phi_{P'}\rangle = 0$. However, as an artifact of the Heisenberg uncertainty principle, in the case that only I_P and $I_{P'}$ are disjoint, we need to deal with tails. The precise estimate we need is as follows. Assume that $|I_P| \geq |I_{P'}|$. Then for every integer $\nu \geq 0$ we have that

$$|\langle\phi_P,\phi_{P'}\rangle| \lesssim |I_P|^{-\frac12}|I_{P'}|^{\frac12}(1+2^{-k_P}\rho(c(I_P)-c(I_{P'})))^{-\nu}, \tag{4.16}$$

where the implicit constant depends only on ν and ϕ. See [Thi06, Lemma 2.1] for the version of this estimate for one-dimensional wave packets. Similarly, we have

$$|\langle\psi_P^N,\psi_{P'}^N\rangle| \lesssim \|m\|_{\mathscr{M}^\nu}^2|I_P|^{-\frac12}|I_{P'}|^{\frac12}(1+2^{-k_P}\rho(c(I_P)-c(I_{P'})))^{-\nu}, \tag{4.17}$$

for every integer $\nu \geq 0$ provided that $N \notin \omega_{P(0)} \cup \omega_{P'(0)}$. We prove (4.16) and (4.17) in Section 8.

5. Proof of the mass lemma

In this section we prove Lemma 4.2. The proof is in essence the same as in [LT00, Prop. 3.1]. Let $\mathcal{P}$ be a finite set of tiles and set $\mu = \mathcal{M}(\mathcal{P})$. We define the set of heavy tiles by

$$\mathcal{P}_{\text{heavy}} = \Big\{P\in\mathcal{P} \,:\, \mathcal{M}(P) > \frac{\mu}{4}\Big\}$$

and accordingly $\mathcal{P}_{\text{light}} = \mathcal{P} \backslash \mathcal{P}_{\text{heavy}}$. Then (4.5) is automatically satisfied. It remains to show (4.6). By the definition of mass (4.2) we know that for every $P \in \mathcal{P}_{\text{heavy}}$ there exists a $P' = P'(P) \in \overline{\mathcal{P}}$ with $P' \geq P$ such that

$$\int_{E_{P'}} w_{P'}(x)dx > \frac{\mu}{4} \tag{5.1}$$

Note that P' need not be in $\mathcal{P}$. Let $\mathcal{P}'$ be the maximal elements in

$$\{P'(P) : P \in \mathcal{P}_{\text{heavy}}\}$$

with respect to the partial order $\leq$ of tiles. Then $\mathcal{P}_{\text{heavy}}$ is a union of trees with tops in $\mathcal{P}'$. Therefore it suffices to show

$$\sum_{P \in \mathcal{P}'} |I_P| \leq \frac{C}{\mu} \tag{5.2}$$

First we rewrite (5.1) as

$$\sum_{j=0}^{\infty} \int_{E_P \cap (\delta_{2^j}(I_P) \backslash \delta_{2^{j-1}}(I_P))} w_P(x)dx > C\mu \sum_{j=0}^{\infty} 2^{-j}. \tag{5.3}$$

where we adopt the temporary convention that $\delta_{2^{-1}}(I_P) = \emptyset$ and for $j \geq 0$,

$$\delta_{2^j}(I_P) = \prod_{i=1}^{n} \Big[c(I_P)_i - 2^{(k_P+j)\alpha_i - 1}, c(I_P)_i + 2^{(k_P+j)\alpha_i - 1}\Big).$$

Thus, for every $P \in \mathcal{P}'$ there exists a $j \geq 0$ such that

$$\int_{E_P \cap (\delta_{2^j}(I_P) \backslash \delta_{2^{j-1}}(I_P))} \frac{dx}{(1 + 2^{-k_P}\rho(x - c(I_P)))^{\nu_1}} > C|I_P|\mu 2^{-j}. \tag{5.4}$$

Note that for $x \in \delta_{2^j}(I_P) \backslash \delta_{2^{j-1}}(I_P)$ we have

$$1 + 2^{-k_P}\rho(x - c(I_P)) \geq C2^j.$$

Using this we obtain from (5.4),

$$|I_P| < C\mu^{-1}|E_P \cap \delta_{2^j}(I_P)|2^{-(\nu_1-1)j}. \tag{5.5}$$

Summarizing, we have shown that for every $P \in \mathcal{P}'$ there exists $j \geq 0$ such that (5.5) holds. This leads us to define for every $j \geq 0$, a set of tiles $\mathcal{P}_j$ by

$$\mathcal{P}_j = \{P \in \mathcal{P}' : |I_P| < C\mu^{-1}|E_P \cap \delta_{2^j}(I_P)|2^{-j(\nu_1-1)}\}.$$

The estimate (5.2) will follow by summing over j if we can show that

$$\sum_{P \in \mathcal{P}_j} |I_P| \leq C2^{-j}\mu^{-1} \tag{5.6}$$

for all $j \geq 0$. To show (5.6) we use a covering argument reminiscent of Vitali's covering lemma. Fix $j \geq 0$. For every tile $P = I_P \times \omega_P$ we have an enlarged tile $\delta_{2^j}(I_P) \times \omega_P$ (this is not a tile anymore). We inductively choose $P_i \in \mathcal{P}_j$ such that $|I_{P_i}|$ is maximal among the $P \in \mathcal{P}_j \backslash \{P_0, \dots, P_{i-1}\}$ and the enlarged tile of P_i is disjoint from the enlarged tiles of $P_0, \dots, P_{i-1}$. Since $\mathcal{P}_j$ is finite, this process terminates after finitely many steps, so that we have selected a subset $\mathcal{P}_j' = \{P_0, P_1, \dots\} \subset \mathcal{P}_j$ of tiles whose enlarged tiles are pairwise disjoint. By

construction, for every $P \in \mathcal{P}_j$ there exists a unique $P' \in \mathcal{P}'_j$ such that $|I_P| \le |I_{P'}|$ and the enlarged tiles of P and P' intersect. We call P *associated* with P'.
Now the claim is that if two tiles $P, Q \in \mathcal{P}_j$ are associated with the same $P' \in \mathcal{P}'_j$, then I_P and I_Q are disjoint. To see this note that ω_P intersects $\omega_{P'}$ by definition. Thus, since $|I_P| \le |I_{P'}|$, we have $\omega_{P'} \subset \omega_P$. The same holds for Q. Therefore we have $\omega_{P'} \subset \omega_P \cap \omega_Q$. But $P, Q \in \mathcal{P}_j \subset \mathcal{P}'$ are disjoint tiles, so we must have $I_P \cap I_Q = \emptyset$. Moreover, all tiles P associated with P' satisfy $I_P \subset \delta_{2^{j+2}}(I_{P'})$. Therefore we get

$$\begin{aligned}\sum_{P \in \mathcal{P}_j} |I_P| &= \sum_{P' \in \mathcal{P}'_j} \sum_{\substack{P \in \mathcal{P}_j \\ \text{assoc. with } P'}} |I_P| = \sum_{P' \in \mathcal{P}'_j} \left| \bigcup_{\substack{P \in \mathcal{P}_j \\ \text{assoc. with } P'}} I_P \right| \\ &\le \sum_{P' \in \mathcal{P}'_j} 2^{(j+2)|\alpha|} |I_{P'}| \le C\mu^{-1} 2^{-j(\nu_1 - |\alpha| - 1)} \sum_{P' \in \mathcal{P}'_j} |E \cap N^{-1}(\omega_{P'}) \cap \delta_{2^j}(I_{P'})| \\ &\le C 2^{-j} \mu^{-1},\end{aligned}$$

using that $\nu_1 > |\alpha| + 1$. The penultimate inequality is a consequence of (5.5) and the last inequality follows, because the sets $N^{-1}(\omega_{P'}) \cap \delta_{2^j}(I_{P'})$ are disjoint and $|E| \le 1$.

6. Proof of the energy lemma

In this section we prove Lemma 4.3. We adapt the argument of Lacey and Thiele [LT00, Prop. 3.2]. The tree selection algorithm of Lacey and Thiele relies on the natural ordering of real numbers. In our situation this can be replaced by any functional on $\mathbb{R}^n$ that separates $\omega_{P(0)}$ from $\omega_{P(r)}$ for every tile $P \in \overline{\mathcal{P}}$ (this was already observed in [PT03]). Let i_0 be such that $r_{i_0} = 1$ (exists because $r \ne 0$). Let us introduce the projection to the i_0th coordinate: $\pi_0 : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto x_{i_0}$. Then we have that

$$\pi_0(\xi) < \pi_0(\eta) \tag{6.1}$$

holds for every $\xi \in \omega_{P(0)}, \eta \in \omega_{P(r)}, P \in \overline{\mathcal{P}}$.
Let $\varepsilon = \mathcal{E}(\mathcal{P})$. For a 2–tree $\mathbf{T}_2$ we define

$$\Delta(\mathbf{T}_2) = \left(\frac{1}{|I_{\mathbf{T}_2}|} \sum_{P \in \mathbf{T}_2} |\langle f, \phi_P \rangle|^2 \right)^{1/2}.$$

We will now describe an algorithm to choose the desired collection of trees $\mathcal{T}$ and also an auxiliary collection of 2–trees $\mathcal{T}_2$:

(1) Initialize $\mathcal{T} := \mathcal{T}_2 := \emptyset$ and $\mathcal{P}^{\text{stock}} := \mathcal{P}$.
(2) Choose a 2–tree $\mathbf{T}_2 \subset \mathcal{P}^{\text{stock}}$ such that
  (a) $\Delta(\mathbf{T}_2) \ge \varepsilon/2$, and
  (b) $\pi_0(c(\omega_{\mathbf{T}_2}))$ is minimal among all the 2–trees in $\mathcal{P}^{\text{stock}}$ satisfying (a).
  If no such $\mathbf{T}_2$ exists, then terminate.
(3) Let $\mathbf{T}$ be the maximal tree in $\mathcal{P}^{\text{stock}}$ with top $P_{\mathbf{T}_2}$ (with respect to set inclusion).

(4) Add $\mathbf{T}$ to $\mathcal{T}$ and $\mathbf{T}_2$ to $\mathcal{T}_2$. Also, remove all the elements of $\mathbf{T}$ from $\mathcal{P}^{\text{stock}}$. Then continue again with Step (2).

Since $\mathcal{P}$ is finite it is clear that the algorithm terminates after finitely many steps. Also note for every $\mathbf{T} \in \mathcal{T}$ there exists a unique $\mathbf{T}_2 \in \mathcal{T}_2$ with $\mathbf{T}_2 \subset \mathbf{T}$, and vice versa. After the algorithm terminates we set $\mathcal{P}_{\text{low}} = \mathcal{P}^{\text{stock}}$ and $\mathcal{P}_{\text{high}}$ to be the union of the trees in $\mathcal{T}$. Then, (4.7) is automatically satisfied and it only remains to show

$$\sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}| \lesssim \varepsilon^{-2}. \tag{6.2}$$

Before we do that we establish a geometric property of the selected trees that will be crucial in the following.

Lemma 6.1. *Let $\mathbf{T}_2 \neq \mathbf{T}_2' \in \mathcal{T}_2$ and $P \in \mathbf{T}_2, P' \in \mathbf{T}_2'$. If $\omega_P \subset \omega_{1P'}$, then $I_{P'} \cap I_{\mathbf{T}_2} = \emptyset$.*

Proof. Note that $c(\omega_{\mathbf{T}_2}) \in \omega_P \subset \omega_{P'(0)}$ while $c(\omega_{\mathbf{T}_2'}) \in \omega_{P'(r)}$. By (6.1) and condition (b) in Step (2) we therefore conclude that $\mathbf{T}_2$ was chosen before $\mathbf{T}_2'$ during the above algorithm. Let $\mathbf{T}$ be the tree in $\mathcal{T}$ such that $\mathbf{T}_2 \subset \mathbf{T}$. Thus, if $I_{P'}$ was not disjoint from $I_{\mathbf{T}_2} = I_{\mathbf{T}}$, then it would be contained in $I_{\mathbf{T}}$ and therefore $P' \leq P_{\mathbf{T}}$ which means it would have been included into $\mathbf{T}$ during Step (3). That is a contradiction. □

The sum in (6.2) equals

$$\sum_{T_2 \in \mathcal{T}_2} \Delta(\mathbf{T}_2)^{-2} \sum_{P \in \mathbf{T}_2} |\langle f, \phi_P \rangle| \leq 4\varepsilon^{-2} \sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2,$$

where $\bigcup \mathcal{T}_2 = \bigcup_{\mathbf{T}_2 \in \mathcal{T}_2} \mathbf{T}_2$. Let us write

$$\sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2 = \Big\langle \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P, f \Big\rangle \tag{6.3}$$

and use the Cauchy-Schwarz inequality to estimate this by

$$\Big\| \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P \Big\|_2, \tag{6.4}$$

where we used that $\|f\|_2 = 1$. So far we have shown that

$$\varepsilon^2 \sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}| \lesssim \Big\| \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P \Big\|_2. \tag{6.5}$$

Thus if we can show that

$$\Big\| \sum_{P \in \bigcup \mathcal{T}_2} \langle f, \phi_P \rangle \phi_P \Big\|_2^2 \lesssim \varepsilon^2 \sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}|, \tag{6.6}$$

then (6.2) follows. Expanding the L^2 norm in (6.6) we get that the left hand side is bounded by

$$\sum_{\substack{P,P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} |\langle f, \phi_P \rangle \langle f, \phi_{P'} \rangle \langle \phi_P, \phi_{P'} \rangle| + 2 \sum_{\substack{P,P' \in \bigcup \mathcal{T}_2, \\ \omega_P \subset \omega_{P'(0)}}} |\langle f, \phi_P \rangle \langle f, \phi_{P'} \rangle \langle \phi_P, \phi_{P'} \rangle|. \tag{6.7}$$

Here we have used that $\langle \phi_P, \phi'_P \rangle = 0$ if $\omega_{P(0)} \cap \omega_{P'(0)} = \emptyset$ and therefore either $\omega_P = \omega_{P'}$, $\omega_P \subset \omega_{P'(0)}$, or $\omega_{P'} \subset \omega_{P(0)}$ (the last two cases are symmetric). We treat both sums in this term separately. Estimating the smaller one of $|\langle f, \phi_P \rangle|$ and $|\langle f, \phi_{P'} \rangle$ by the larger one, we obtain that the first sum in (6.7) is

$$\lesssim \sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2 \sum_{\substack{P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} |\langle \phi_P, \phi_{P'} \rangle|.$$

Using (4.16) we estimate this by

$$\sum_{P \in \bigcup \mathcal{T}_2} |\langle f, \phi_P \rangle|^2 \sum_{\substack{P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu}. \tag{6.8}$$

Notice that $I_P \cap I_{P'} = \emptyset$ for $P \neq P'$ in the inner sum. This implies

$$\sum_{\substack{P' \in \bigcup \mathcal{T}_2, \\ \omega_P = \omega_{P'}}} (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu} \lesssim \int_{\mathbb{R}^n} (1 + \rho(x))^{-\nu} dx \lesssim 1,$$

provided that $\nu > |\alpha|$. Therefore (6.8) is

$$\lesssim \sum_{\mathbf{T}_2 \in \mathcal{T}_2} \sum_{P \in \mathbf{T}_2} |\langle f, \phi_P \rangle|^2 \leq \varepsilon^2 \sum_{\mathbf{T}_2 \in \mathcal{T}_2} |I_{\mathbf{T}_2}|, \tag{6.9}$$

as desired. It remains to estimate the second sum in (6.7). To that end it suffices to show that

$$\sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} |\langle f, \phi_P \rangle \langle f, \phi_{P'} \rangle \langle \phi_P, \phi_{P'} \rangle| \lesssim \varepsilon^2 |I_{\mathbf{T}_2}|, \tag{6.10}$$

for every $\mathbf{T}_2 \in \mathcal{T}_2$, where

$$\mathcal{S}_P = \Big\{ P' \in \bigcup \mathcal{T}_2 \,:\, \omega_P \subset \omega_{P'(0)} \Big\}.$$

Here we follow the argument given in [Lac04]. Observe that if $P \in \mathbf{T}_2$, then $\mathcal{S}_P \cap \mathbf{T}_2 = \emptyset$. Interpreting the singleton $\{P\}$ as a 2–tree we obtain

$$|\langle f, \phi_P \rangle| \leq \varepsilon |I_P|^{1/2} \tag{6.11}$$

for all $P \in \mathcal{P}$. Combining this with (4.16) we can estimate the left hand side of (6.10) by

$$\varepsilon^2 \sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} |I_{P'}| (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu}. \tag{6.12}$$

Indeed, Lemma 6.1 implies that $I_{\mathbf{T}_2} \cap I_{P'} = \emptyset$ for every $P \in \mathbf{T}_2, P' \in \mathcal{S}_P$. Moreover, it also implies that for $P' \neq P'' \in \mathcal{S}_P$ we have $I_{P'} \cap I_{P''} = \emptyset$. These facts facilitate the following estimate:

$$\sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} |I_{P'}| (1 + 2^{-k_P} \rho(c(I_P) - c(I_{P'})))^{-\nu} \lesssim \sum_{P \in \mathbf{T}_2} \sum_{P' \in \mathcal{S}_P} \int_{I_{P'}} (1 + 2^{-k_P} \rho(c(I_P) - x))^{-\nu} dx$$

$$\lesssim \sum_{P \in \mathbf{T}_2} \int_{(I_{\mathbf{T}_2})^c} (1 + 2^{-k_P} \rho(c(I_P) - x))^{-\nu}.$$

Since $\mathbf{T}_2$ is a tree, the last quantity can be estimated by

$$\sum_{k\le k_{\mathbf{T}_2}}\sum_{u\in Q_k\cap(\mathbb{Z}^n+\frac12)}\int_{(I_{\mathbf{T}_2})^c}(1+\rho(u-\delta_{2^{-k}}(x)))^{-\nu}dx,$$

where $Q_k\in\mathcal{D}^\alpha$ is an anisotropic dyadic rectangle of scale $k_{\mathbf{T}_2}-k$ that is given by a rescaling of $I_{\mathbf{T}_2}$. The previous display is no greater than a constant times

(6.13)

$$\sum_{k\le k_{\mathbf{T}_2}}2^{k|\alpha|}\left(\sum_{u\in Q_k\cap(\mathbb{Z}^n+\frac12)}(1+\mathrm{dist}_\alpha((Q_k)^c,u)^{-|\alpha|-\gamma}\right)\left(\int_{\mathbb{R}^n}(1+\rho(x))^{-(\nu-|\alpha|-\gamma)}dx\right),$$

where $\nu>2|\alpha|$ and γ is a fixed and sufficiently small positive constant. The integral over x in the previous display is bounded by a constant depending on $\nu-|\alpha|-\gamma>|\alpha|$. To estimate the sum over u we note that for every u in the indicated range there exists a lattice point $v\in\partial Q_k\cap\mathbb{Z}^n$ such that $\mathrm{dist}_\alpha((Q_k)^c,u)\ge\frac12\rho(v-u)$. Thus we may bound the sum over u by

$$\sum_{v\in\partial Q_k\cap\mathbb{Z}^n}\sum_{u\in\mathbb{Z}^n+\frac12}(1+\rho(v-u))^{-|\alpha|-\gamma}\lesssim|\partial Q_k\cap\mathbb{Z}^n|\lesssim 2^{(k_{\mathbf{T}_2}-k)|\alpha|_\infty}.$$

Thus, (6.13) is bounded by a constant times

$$2^{k_{\mathbf{T}_2}|\alpha|_\infty}\sum_{k\le k_{\mathbf{T}_2}}2^{k(|\alpha|-|\alpha|_\infty)}\lesssim 2^{k_{\mathbf{T}_2}|\alpha|}=|I_{\mathbf{T}_2}|.$$

This proves (6.10).

7. Proof of the tree estimate

In this section we prove Lemma 4.4. This is the core of the proof. For a rectangle $I=\prod_{i=1}^n I_i\in\mathcal{D}^\alpha$ we denote by $\widetilde{I}$ the enlarged rectangle defined by

$$\widetilde{I}=\prod_{i=1}^n(2^{\alpha_i+1}-1)I_i.$$

Here λI_i is the interval of length $\lambda|I_i|$ with the same center as I_i. Let $\mathcal{J}$ be the partition of $\mathbb{R}^n$ that is given by the collection of maximal anisotropic dyadic rectangles $J\in\mathcal{D}^\alpha$ such that $\widetilde{J}$ does not contain any I_P with $P\in\mathbf{T}$ (maximal with respect to inclusion). Set $\varepsilon=\mathcal{E}(\mathbf{T})$ and $\mu=\mathcal{M}(\mathbf{T})$. Choose phase factors $(\epsilon_P)_P$ of modulus 1 such that

$$\begin{aligned}\sum_{P\in\mathbf{T}}|\langle f,\phi_P\rangle\langle\psi_P^{N(\cdot)},\mathbf{1}_{E_{P(r)}}\rangle|&=\int_{\mathbb{R}^n}\sum_{P\in\mathbf{T}}\epsilon_P\langle f,\phi_P\rangle\psi_P^{N(x)}(x)\mathbf{1}_{E_{P(r)}}(x)dx\\&\le\left\|\sum_{P\in\mathbf{T}}\epsilon_P\langle f,\phi_P\rangle\psi_P^N\mathbf{1}_{E_{P(r)}}\right\|_1\le\mathcal{K}_1+\mathcal{K}_2,\end{aligned}$$

where

$$\mathcal{K}_1=\sum_{J\in\mathcal{J}}\sum_{P\in\mathbf{T},|I_P|\le|J^+|}\|\langle f,\phi_P\rangle\psi_P^{N(\cdot)}\mathbf{1}_{E_{P(r)}}\|_{L^1(J)},$$

$$\mathcal{K}_2 = \sum_{J\in\mathcal{J}} \Big\| \sum_{P\in\mathbf{T}, |I_P|>|J^+|} \epsilon_P \langle f, \phi_P\rangle \psi_P^{N(\cdot)} \mathbf{1}_{E_{P(r)}} \Big\|_{L^1(J)}.$$

We first estimate $\mathcal{K}_1$. This is the easy part, since in the sum defining $\mathcal{K}_1$ we have that I_P is disjoint from $\widetilde{J}$. Again, interpreting the singleton $\{P\}$ as a 2–tree we see that (6.11) holds for all $P \in \mathbf{T}$. This gives

$$\mathcal{K}_1 \leq \varepsilon \sum_{J\in\mathcal{J}} \sum_{\substack{P\in\mathbf{T}\\ |I_P|\leq|J^+|}} 2^{|\alpha|k_P/2} \|\psi_P^{N(\cdot)} \mathbf{1}_{E_{P(r)}}\|_{L^1(J)}.$$

Using (4.15) the previous display is seen to be no larger than a constant times

$$\|m\|_{\mathscr{M}^{\nu_0}} \varepsilon \sum_{J\in\mathcal{J}} \sum_{\substack{P\in\mathbf{T}\\ |I_P|\leq|J^+|}} \int_{J\cap E_{P(r)}} w^{\nu_0}(2^{-k_P}(x - c(I_P)))dx$$

$$\leq \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon\mu \sum_{J\in\mathcal{J}} \sum_{\substack{P\in\mathbf{T}\\ |I_P|\leq|J^+|}} 2^{|\alpha|k_P} \sup_{x\in J} w^{2|\alpha|+\frac{2}{3}}(2^{-k_P}(x - c(I_P))), \tag{7.1}$$

where we have set $\nu_1 = |\alpha| + \frac{4}{3}$. Since I_P is disjoint from $\widetilde{J}$ we can estimate (7.1) as

$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon\mu \sum_{J\in\mathcal{J}} \sum_{\substack{k\in\mathbb{Z},\\ 2^{k|\alpha|}\leq|J^+|}} 2^{|\alpha|k} \sum_{\substack{P\in\mathbf{T},\\ k_P=k}} w^{2|\alpha|+\frac{2}{3}}(2^{-k}\mathrm{dist}_\alpha(J, I_P)). \tag{7.2}$$

Before we proceed, we claim that for every $\nu > |\alpha|$, $k \in \mathbb{Z}$ and fixed $J \in \mathcal{J}$ with $2^{k|\alpha|} \leq |J^+|$ we have

$$\sum_{\substack{P\in\mathbf{T},\\ k_P=k}} w^{\nu}(2^{-k}\mathrm{dist}_\alpha(J, I_P)) \lesssim 1, \tag{7.3}$$

where the implicit constant blows up as ν approaches $|\alpha|$. To verify the claim, let us assume for simpler notation that J is centered at the origin. Then by disjointness of I_P and $\widetilde{J}$ we have

$$\mathrm{dist}_\alpha(J, I_P) \gtrsim \mathrm{dist}_\alpha(0, I_P) \gtrsim 2^k\rho(m),$$

where $m = (m_1, \ldots, m_n) \in \mathbb{Z}^n$ is such that $I_P = \prod_{i=1}^n [2^{k\alpha_i}m_i, 2^{k\alpha_i}(m_i + 1))$. Thus the sum in (7.3) is

$$\lesssim \sum_{m\in\mathbb{Z}^n} (1 + \rho(m))^{-\nu},$$

which implies the claim.

Estimate (7.2) by

$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon\mu \sum_{J\in\mathcal{J}} w^{|\alpha|+\frac{1}{3}}(2^{-k_\mathbf{T}}\mathrm{dist}_\alpha(J, I_T)) \sum_{\substack{k\in\mathbb{Z},\\ 2^{k|\alpha|}\leq|J^+|}} 2^{|\alpha|k}, \tag{7.4}$$

Here $k_\mathbf{T}$ is the scale of $I_\mathbf{T}$ and we have used (7.3) and

$$2^{-k}\mathrm{dist}_\alpha(J, I_P) \geq 2^{-k_\mathbf{T}}\mathrm{dist}_\alpha(J, I_\mathbf{T}).$$

Summing the geometric series, (7.4) is

$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon\mu \sum_{J\in\mathcal{J}} w^{|\alpha|+\frac{1}{3}}(2^{-k_{\mathbf{T}}}\mathrm{dist}_\alpha(J,I_P))|J|.$$

The sum in that expression can be estimated as follows:

$$\sum_{J\in\mathcal{J}} w^{|\alpha|+\frac{1}{3}}(2^{-k_{\mathbf{T}}}\mathrm{dist}_\alpha(J,I_P))|J| \lesssim \sum_{J\in\mathcal{J}} \int_J (1+2^{-k_{\mathbf{T}}}\rho(x-c(I_{\mathbf{T}})))^{-(|\alpha|+\frac{1}{3})}dx.$$

By disjointness of the J we can bound this by

$$\int_{\mathbb{R}^n} (1+2^{-k_{\mathbf{T}}}\rho(x-c(I_{\mathbf{T}})))^{-(|\alpha|+\frac{1}{3})}dx = |I_{\mathbf{T}}| \int_{\mathbb{R}^n} (1+\rho(x))^{-(|\alpha|+\frac{1}{3})}dx \lesssim |I_{\mathbf{T}}|.$$

To summarize, we showed that

$$\mathcal{K}_1 \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon\mu |I_{\mathbf{T}}|, \tag{7.5}$$

using that $\nu_0 \geq 3|\alpha|+2$.

Let us proceed to estimating $\mathcal{K}_2$. This is more difficult. We may assume that the sum runs only over those J for which there is a $P\in\mathbf{T}$ such that $|I_P|>|J^+|$. Then $|I_{\mathbf{T}}|>|J^+|$ and $J\subset\widetilde{I_{\mathbf{T}}}$. From now on let such a J be fixed. Define

$$G_J = J \cap \bigcup_{P\in\mathbf{T},|I_P|>|J^+|} E_{P(r)} \tag{7.6}$$

Before proceeding we prove the following.

Lemma 7.1. *There exists a constant $C>0$ independent of J such that*

$$|G_J| \leq C\mu|J| \tag{7.7}$$

Proof. By definition of J, there exists $P_0\in\mathbf{T}$ such that I_{P_0} is contained in $\widetilde{J^+}$. We claim that there exists a tile $P_0<P'<P_{\mathbf{T}}$ such that $|I_{P'}|=|J^{++}|$. Indeed, note $|I_{P_0}|\leq|J^{++}|$. If there is equality, we simply take $P'=P_0$. Otherwise we take $I_{P'}\in\mathcal{D}^\alpha$ to be the unique dyadic ancestor of I_{P_0} such that $|I_{P'}|=|J^{++}|$ and choose $\omega_{P'}$ accordingly such that it contains $c(\omega_{\mathbf{T}})$. Now we have

$$|\omega_P| = |I_P|^{-1} \leq |J^{++}|^{-1} = |I_{P'}|^{-1} = |\omega_{P'}|$$

for every tile $P\in\mathbf{T}$ with $|I_P|>|J^+|$. This implies $\omega_P\subset\omega_{P'}$ and thus

$$G_J \subset J\cap E_{P'}.$$

As a consequence,

$$|G_J| \leq \int_{E_{P'}} \mathbf{1}_J(x)dx \lesssim |I_{P'}| \int_{E_{P'}} w_{P'}(x)dx \lesssim \mu|J|.$$

□

Let us define

$$F_J = \sum_{\substack{P\in\mathbf{T},\\ |I_P|>|J^+|}} \epsilon_P\langle f,\phi_P\rangle \psi_P^{N(\cdot)} \mathbf{1}_{E_{P(r)}}. \tag{7.8}$$

Since every tree can be written as the union of a 1–tree and a 2–tree, we may treat each of these cases separately.

7.1. **The case of 1–trees.** Assume that $\mathbf{T}$ is a 1–tree. This is the easier case. The reason is that for every $P, P' \in \mathbf{T}$, $\omega_P \neq \omega_{P'}$ we have that $\omega_{P(r)}$ and $\omega_{P'(r)}$ are disjoint and thus we have good orthogonality of the summands in (7.8). Using (6.11) and (4.15) we see that

$$|F_J(x)| \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon \sum_{\substack{P\in\mathbf{T},\\ |I_P|>|J^+|}} (1+2^{-k_P}\rho(x-c(I_P)))^{-\nu_0} \mathbf{1}_{E_{P(r)}}(x).$$

Using disjointness of the $E_{P(r)}$ this can be estimated by

$$\|m\|_{\mathscr{M}^{\nu_0}} \varepsilon \cdot \sup_{k\in\mathbb{Z}} \sum_{m\in\mathbb{Z}^n+\frac{1}{2}} (1+2^{-k}\rho(x-\delta_{2^k}(m)))^{-\nu_0}.$$

By an index shift we see that

$$\sum_{m\in\mathbb{Z}^n+\frac{1}{2}} (1+2^{-k}\rho(x-\delta_{2^k}(m)))^{-\nu_0} = \sum_{m\in\mathbb{Z}^n+\frac{1}{2}} (1+\rho(m+\gamma))^{-\nu_0},$$

where $\gamma \in [0,1]^n$ depends on k and x. The last sum is $\lesssim 1$ independently of γ. Thus we proved the pointwise estimate

$$|F_J(x)| \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon. \tag{7.9}$$

Combining this with the support estimate (7.7) we obtain

$$\|F_J\|_{L^1(J)} \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon \mu |J|. \tag{7.10}$$

Summing over the pairwise disjoint $J \subset \widetilde{I_{\mathbf{T}}}$ we obtain

$$\mathcal{K}_2 \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \varepsilon \mu |I_{\mathbf{T}}|$$

as desired. Note that we only needed $\nu_0 > |\alpha|$ to obtain this estimate.

7.2. **The case of 2–trees.** Here we assume that $\mathbf{T}$ is a 2–tree. The additional x-dependence present in the wave packets $\psi_P^{N(x)}$ makes this part more difficult than the congruent argument in [LT00]. This problem arises already in the isotropic case [PT03]. The goal is again to obtain a pointwise estimate for F_J. In the following we fix $x \in J$ such that $F_J(x) \neq 0$. Observe that the $\omega_{P(r)}$, $P \in \mathbf{T}$ are nested. Let us denote the smallest (resp. largest) $\omega_{P(r)}$ (resp. ω_P) such that $x \in N^{-1}(\omega_{P(r)}) \cap E$ by ω_- (resp. ω_+). Let $k_+ \in \mathbb{Z}$ be such that $|\omega_+| = 2^{k_+|\alpha|}$ and $k_- \in \mathbb{Z}$ such that $|\omega_-| = 2^{-k_-|\alpha|-n}$ (note from the definition that $\omega_- \notin \mathcal{D}^\alpha$ if $\alpha \neq (1,\dots,1)$). Then the nestedness property implies

$$F_J(x) = \sum_{\substack{P\in\mathbf{T},\\ k_+\le k_P\le k_-}} \epsilon_P \langle f, \phi_P\rangle \psi_P^{N(x)}(x)$$

Define

$$h_x = \mathrm{M}_{c(\omega_+)} \mathrm{D}^1_{2^{k_+}} \phi_+ - M_{c(\omega_-)} D^1_{2^{k_-}} \phi_-,$$

where $\phi_+(x) = b_1^{-n}\phi(b_1^{-1}x)$ and ϕ_- is a Schwartz function satisfying $0 \le \widehat{\phi_-} \le 1$ such that $\widehat{\phi_-}$ is supported on $[-\frac{b_2}{2}, \frac{b_2}{2}]$ and equals to one on $[-\frac{b_3}{2}, \frac{b_3}{2}]$, where $b_{j+2} = \frac{1}{2} + b_j$ for $j = 0, 1$. From the definition we see that $\widehat{h_x}$ is supported on $b_0 b_1^{-1}\omega_+ \cap (2b_3\omega_-)^c$ and equal to one on $\omega_+ \cap (2b_2\omega_-)^c$. In particular, $\widehat{h_x}(\xi)$ equals to one if $\xi \in \operatorname{supp} \phi_P$ and $k_+ \leq k_P \leq k_-$ and vanishes if k_P is outside this range. For technical reasons that become clear further below we need the support of $\widehat{h_x}$ to keep a certain distance to ω_-. We obtain

$$F_J(x) = \sum_{P \in \mathbf{T}} \epsilon_P \langle f, \phi_P \rangle (\psi_P^{N(x)} * h_x)(x).$$

Fix $\xi_0 \in \omega_{\mathbf{T}}$. We decompose

$$F_J(x) = \sum_{P \in \mathbf{T}} \epsilon_P \langle f, \phi_P \rangle (\psi_P^{\xi_0} * h_x)(x) + \sum_{P \in \mathbf{T}} \epsilon_P \langle f, \phi_P \rangle (\psi_P^{N(x)} - \psi_P^{\xi_0}) * h_x)(x) \tag{7.11}$$

$$= G * \mathrm{M}_{\xi_0} K * h_x(x) + G * (\mathrm{M}_{N(x)} K - \mathrm{M}_{\xi_0} K) * h_x(x) \tag{7.12}$$

where

$$G = \sum_{P \in \mathbf{T}} \epsilon_P \langle f, \phi_P \rangle \phi_P. \tag{7.13}$$

Before proceeding with the proof we record the following simple variant of a standard fact about maximal functions (see [Duo01]).

Lemma 7.2. *Let $\lambda > 0$ and w be an integrable function on $\mathbb{R}^n$ which is constant on $\{\rho(y) \leq \lambda\}$ and radial and decreasing with respect to ρ, i.e.*

$$w(x) \leq w(y)$$

if $\rho(x) \geq \rho(y)$, with equality if $\rho(x) = \rho(y)$. Let $x \in \mathbb{R}^n$ and $J \subset \mathbb{R}^n$ be such that $J \subset \{y : \rho(x-y) \leq \lambda\}$. Then we have

$$|F * w|(x) \leq \|w\|_1 \sup_{J \subset I} \frac{1}{|I|} \int_I |F(y)| dy,$$

where the supremum is taken over all anisotropic cubes $I \subset \mathbb{R}^n$.

Proof. First we assume that w is a step function. That is,

$$w(y) = \sum_{j=1}^{\infty} c_j \mathbf{1}_{\rho(y) \leq r_j}$$

with $\lambda \leq r_1 < r_2 < \cdots$. Then we have

$$F * w(x) = \sum_j r_j^{|\alpha|} c_j \frac{1}{r_j^{|\alpha|}} \int_{\rho(x-y) \leq r_j} |F(y)| dy \leq \|w\|_1 \sup_{J \subset I} \frac{1}{|I|} \int_I |F(y)| dy.$$

The general case follows by approximation of w by step functions and an application of Lebesgue's dominated convergence theorem. □

Since

$$|h_x(y)| \lesssim 2^{-k_+|\alpha|} |\phi_+|(\delta_{2^{-k_+}}(y)) + 2^{-k_-|\alpha|} |\phi_-|(\delta_{2^{-k_-}}(y)) \tag{7.14}$$

and $x \in J$, $|J| \leq 2^{k_+|\alpha|} \leq 2^{k_-|\alpha|}$ we have from Lemma 7.2 that

$$|G * \mathrm{M}_{\xi_0} K * h_x(x)| \lesssim \sup_{J \subset I} \frac{1}{|I|} \int_I |G * \mathrm{M}_{\xi_0} K(y)| dy. \tag{7.15}$$

Let us assume for the moment that we also have the estimate

$$|G * (\mathrm{M}_{N(x)}K - \mathrm{M}_{\xi_0}K) * h_x(x)| \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \sup_{J\subset I} \frac{1}{|I|}\int_I |G(y)|dy. \tag{7.16}$$

We will first show how to finish the proof from here. At the end of the section we will then show that (7.16) indeed holds.
From (7.12), (7.15), (7.16) and Lemma 7.1 we see that

$$\sum_{\substack{J\in\mathcal{J},\\ J\subset \widetilde{I_{\mathbf{T}}}}} \|F_J\|_{L^1(J)} \lesssim \mu \sum_{\substack{J\in\mathcal{J},\\ J\subset \widetilde{I_{\mathbf{T}}}}} |J| \left(\sup_{J\subset I}\frac{1}{|I|}\int_I |G * \mathrm{M}_{\xi_0}K(y)|dy + \|m\|_{\mathscr{M}^{\nu_0}} \sup_{J\subset I}\frac{1}{|I|}\int_I |G(y)|dy \right)$$

By disjointness of the $J\in\mathcal{J}$ this is no greater than

$$\mu\left(\|\mathcal{M}(G*\mathrm{M}_{\xi_0}K)\|_{L^1(\widetilde{I_{\mathbf{T}}})} + \|m\|_{\mathscr{M}^{\nu_0}} \|\mathcal{M}(G)\|_{L^1(\widetilde{I_{\mathbf{T}}})}\right), \tag{7.17}$$

where $\mathcal{M}$ denotes the maximal function defined by

$$\mathcal{M}F(y) = \sup_{y\in I}\frac{1}{|I|}\int_I |F|,$$

where the supremum runs over all anisotropic cubes $I\subset\mathbb{R}^n$. Clearly, $\mathcal{M}$ is a bounded operator $L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^n)$, because it is bounded pointwise by a composition of one-dimensional Hardy-Littlewood maximal functions applied in each component.

Applying the Cauchy-Schwarz inequality and the L^2 boundedness of $\mathcal{M}$ we see that (7.17) is

$$\lesssim \mu |I_T|^{\frac12}\left(\|G*\mathrm{M}_{\xi_0}K\|_2 + \|m\|_{\mathscr{M}^{\nu_0}}\|G\|_2\right).$$

By repeating the arguments that lead to the proof of (6.6), using (4.16) or (4.17), respectively, we obtain that

$$\|G*\mathrm{M}_{\xi_0}K\|_2 + \|m\|_{\mathscr{M}^{\nu_0}}\|G\|_2 \lesssim \|m\|_{\mathscr{M}^{\nu_0}}\varepsilon |I_{\mathbf{T}}|^{\frac12}.$$

This concludes the proof. It remains to prove (7.16). Let us write

$$R(y) = (\mathrm{M}_{N(x)}K - \mathrm{M}_{\xi_0}K) * h_x(y).$$

We will give two different estimates for R. The first one is only effective if $\rho(y)$ is large and the second one if $\rho(y)$ is small. Let us start with the first estimate. By Fourier inversion, we can write $R(y)$ (up to a constant) as

$$\int_{\mathbb{R}^n} (m(\xi - N(x)) - m(\xi-\xi_0))\widehat{h_x}(\xi)e^{i\xi y}d\xi. \tag{7.18}$$

Fix y and let i be such that $\rho(y) = |y_i|^{1/\alpha_i}$. Then we integrate by parts in the ith component to see that (7.18) is bounded by

$$\lesssim \rho(y)^{-\nu'\alpha_i}\int_{\mathbb{R}^n}\left|\partial_{\xi_i}^{\nu'}\left[(m(\xi-N(x)) - m(\xi-\xi_0))\widehat{h_x}(\xi)\right]\right|d\xi \tag{7.19}$$

for integer $\nu'\ge 0$, where we have used that $\rho(y)\ge 2^{k_-}$ to estimate $|\delta_{2^{-k_-}}(y)|\ge 2^{-k_-}\rho(y)$.

Let $\ell \le \nu_0$ be a non-negative integer. Using (4.13) we obtain

$$\left|\partial_{\xi_i}^{\ell}\Big[m(\xi-N(x))-m(\xi-\xi_0)\Big]\right| \le \|m\|_{\mathscr{M}^\ell}(\rho(\xi-N(x))^{-\ell\alpha_i}+\rho(\xi-\xi_0)^{-\ell\alpha_i}). \tag{7.20}$$

Recall that ξ_0 and $N(x)$ are contained in ω_- and the integrand of (7.19) is supported on $b_0b_1^{-1}\omega_+ \cap (2b_3\omega_-)^c$. Also, there exist $\omega_1,\dots,\omega_M \in \mathcal{D}^\alpha$ such that

$$\omega_- \subsetneq \omega_1 \subsetneq \cdots \subsetneq \omega_M = \omega_+$$

and $|\omega_j| = 2^{-k_j|\alpha|}$ with $k_1 = k_-$ and $k_{j+1} = k_j - 1$. If $\xi \in (2b_3\omega_-)^c$ we have

$$\min(\rho(\xi-N(x)),\rho(\xi-\xi_0)) \gtrsim 2^{-k_-}. \tag{7.21}$$

On the other hand, if $\xi \in (b_0b_1^{-1}\omega_j)\cap\omega_{j-1}^c$ for $j=2,\dots,M$, then

$$\min(\rho(\xi-N(x)),\rho(\xi-\xi_0)) \gtrsim 2^{-k_j}. \tag{7.22}$$

Combining (7.20) and (7.21), (7.22) we get

$$|\partial_{\xi_i}^{\ell}\Big[m(\xi-N(x))-m(\xi-\xi_0)\Big]| \lesssim \|m\|_{\mathscr{M}^\ell}\sum_{j=1}^{M} 2^{k_j\ell\alpha_i}\mathbf{1}_{b_0b_1^{-1}\omega_j}(\xi). \tag{7.23}$$

We also have

$$|\partial_{\xi_i}^{\ell}\widehat{h_x}(\xi)| \lesssim 2^{k_+\ell\alpha_i}\mathbf{1}_{b_0b_1^{-1}\omega_+}(\xi)+2^{k_-\ell\alpha_i}\mathbf{1}_{2b_3\omega_-}(\xi). \tag{7.24}$$

Thus we see from (7.23) and (7.24) that for all $i=1,\dots,n$ and $0\le\ell\le\nu'$ we obtain

$$\int_{\mathbb{R}^n}\left|\partial_{\xi_i}^{\ell}\Big[m(\xi-N(x))-m(\xi-\xi_0)\Big]\partial_{\xi_i}^{\nu'-\ell}\widehat{h_x}(\xi)\right|d\xi \lesssim \|m\|_{\mathscr{M}^{\nu'}}2^{k_-(\nu'\alpha_i-|\alpha|)},$$

provided that $\nu'\alpha_i \ge |\alpha|$. Setting $\nu' = \lceil\frac{\nu_0}{\alpha_i}\rceil$, we have shown that

$$|R(y)| \lesssim \|m\|_{\mathscr{M}^{\nu_0}}2^{-k_-|\alpha|}(2^{-k_-}\rho(y))^{-\nu_0}. \tag{7.25}$$

It remains to find a good estimate for $R(y)$ when $\rho(y)$ is small. Let us estimate

$$|R(y)| \le R_+(y)+R_-(y),$$

where

$$R_\pm = |(\mathrm{M}_{N(x)}K-\mathrm{M}_{\xi_0}K)*\mathrm{D}^1_{2^{k_\pm}}\phi_\pm|.$$

The first claim is that if $\rho(y)\le 2^{k_\pm+1}$, then

$$R_\pm(y) \lesssim \|m\|_{\mathscr{M}^{\nu_0}}2^{-k_\pm|\alpha|}. \tag{7.26}$$

(Here and throughout the proof of this claim $\pm$ always stands for a fixed choice of sign, either $+$ or $-$.) To see this, we first estimate $R_\pm(y)$ by

$$2^{-k_\pm|\alpha|}\int_{\mathbb{R}^n}|(e^{i(N(x)-\xi_0)z}-1)K(z)\phi_\pm(2^{-k_\pm}(y-z))|dz \lesssim 2^{-k_\pm|\alpha|}(\mathbf{I}+\mathbf{II}),$$

where

$$\mathbf{I} = \int_{\rho(z)\le 2^{k_\pm+2}}|(e^{i(N(x)-\xi_0)z}-1)K(z)\phi_\pm(2^{-k_\pm}(y-z))|dz, \quad \text{and}$$

$$\mathbf{II} = \sum_{j=2}^{\infty}\int_{2^{k_\pm+j}\le\rho(z)\le 2^{k_\pm+j+1}}|K(z)\phi_\pm(2^{-k_\pm}(y-z))|dz.$$

We first estimate **I**. Changing variables $z \mapsto \delta_{2^{k_\pm+2}}(z)$ we see that
$$\mathbf{I} \lesssim \int_{\rho(z)\le 1} |(e^{i\delta_{2^{k_\pm+2}}(N(x)-\xi_0)z} - 1)K(z)|dz.$$
Using (4.14), the previous display is
$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} |\delta_{2^{k_\pm+2}}(N(x)-\xi_0)| \int_{\rho(z)\le 1} |z|\rho(z)^{-|\alpha|}dz.$$
Using $\rho(\delta_{2^{k_\pm+2}}(N(x)-\xi_0)) = 2^{k_\pm+2}\rho(N(x)-\xi_0) \lesssim 1$ we can bound this further as
$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} \int_{\rho(z)\le 1} \rho(z)^{1-|\alpha|}dz \lesssim \|m\|_{\mathscr{M}^{\nu_0}}.$$
This proves that $\mathbf{I} \lesssim \|m\|_{\mathscr{M}^{\nu_0}}$. It remains to treat **II**. Here we make use of the fact that we have $\rho(y-z) \ge 2^{k_\pm+j-1}$ in the integrand of **II**, because of our assumption $\rho(y) \le 2^{k_\pm}$. Using the decay of $\phi_\pm$ we obtain
$$\mathbf{II} \lesssim \|m\|_{\mathscr{M}^{\nu_0}} \sum_{j=2}^\infty 2^{-j} \int_{2^{k_\pm+j}\le\rho(z)\le 2^{k_\pm+j+1}} \rho(z)^{-|\alpha|}dz \lesssim \|m\|_{\mathscr{M}^{\nu_0}}.$$
Thus we have proved (7.26). The only further ingredient which we need in order to verify (7.16) is a good estimate for $R_+(y)$ when $2^{k_++1} \le \rho(y) \le 2^{k_-}$. In order to do this we need to do a slightly more careful decomposition. Let us write
$$Q_\ell = [\delta_{2^{k_+}}(\ell), \delta_{2^{k_+}}(\ell+1)) = \prod_{i=1}^n [2^{k_+\alpha_i}\ell_i, 2^{k_+\alpha_i}(\ell_i+1))$$
for $\ell \in \mathbb{Z}^n$. Assume that $y \in Q_\ell$ with $1 \le |\ell|_\infty < 2^{k_- - k_+}$. We have
$$R_+(y) \le 2^{-k_+|\alpha|} \sum_{s\in\mathbb{Z}^n} \int_{Q_s} |(e^{i(N(x)-\xi_0)z} - 1)K(z)\phi_+(2^{-k_+}(y-z))|dz. \tag{7.27}$$
Moreover, the same estimates that were used to prove (7.26) yield
$$\int_{Q_s} |(e^{i(N(x)-\xi_0)z} - 1)K(z)\phi_+(2^{-k_+}(y-z))|dz$$
$$\lesssim \|m\|_{\mathscr{M}^{\nu_0}} 2^{k_+-k_-}(1+\rho(s-\ell))^{-|\alpha|-1}(1+\rho(s))^{1-|\alpha|}$$
Plugging this inequality into (7.27) we obtain
$$\begin{aligned} R_+(y) &\lesssim \|m\|_{\mathscr{M}^{\nu_0}} 2^{-k_-}2^{k_+(1-|\alpha|)} \sum_{s\in\mathbb{Z}^n}(1+\rho(s-\ell))^{-|\alpha|-1}(1+\rho(s))^{1-|\alpha|} \\ &\lesssim \|m\|_{\mathscr{M}^{\nu_0}} 2^{-k_-}(2^{k_+}\rho(\ell))^{1-|\alpha|}, \end{aligned}$$
where the last inequality requires ν to be large enough. Therefore,
$$R_+(y) \lesssim \|m\|_{\mathscr{M}^{\nu_0}} 2^{-k_-}\rho(y)^{1-|\alpha|}. \tag{7.28}$$
Finally, summarizing (7.25), (7.26) and (7.28) we have shown that
$$|R(y)| \lesssim \|m\|_{\mathscr{M}^{\nu_0}}(w_0(y) + w_+(y) + w_-(y) + w_1(y)),$$
where
$$w_0(y) = 2^{-k_-|\alpha|}(2^{-k_-}\rho(y))^{-|\alpha|-1}\mathbf{1}_{\rho(y)\ge 2^{k_-}},$$

$$w_{\pm}(y) = 2^{-k_{\pm}|\alpha|}\mathbf{1}_{\rho(y)\le 2^{k_{\pm}+1}},$$

$$w_1(y) = 2^{-k_-}\rho(y)^{1-|\alpha|}\mathbf{1}_{2^{k_+ +1}\le\rho(y)\le 2^{k_-}}.$$

Each of these functions is integrable with an $L^1(\mathbb{R}^n)$ norm not depending on k_-, k_+, radial and decreasing with respect to ρ in the sense of Lemma 7.2 and constant on $\{\rho(y)\le 2^{k_+}\}$ or $\{\rho(y)\le 2^{k_-}\}$. Thus, applying Lemma 7.2 to each of these functions yields (7.16). Note that to prove (7.16) we only required that $\nu_0\ge|\alpha|+1$.

8. Proofs of auxiliary estimates

In this section we prove (4.15), (4.16) and (4.17).

Proof of (4.15). Expanding definitions and using Fourier inversion we see that, up to a universal constant, $|\psi_P^N(x)|$ is equal to

$$2^{k_P|\alpha|/2}\left|\int_{\mathbb{R}^n} e^{i\xi(x-c(I_P))}m(\xi-N)\widehat{\phi}(\delta_{2^{k_P}}(\xi-c(\omega_{P(0)})))d\xi\right|.$$

Via a change of variables $\delta_{2^{k_P}}(\xi-c(\omega_{P(0)}))\to\zeta$ and using that $m(\xi)=m(\delta_{2^{k_P}}(\xi))$ this becomes

$$2^{-k_P|\alpha|/2}\left|\int_{\mathbb{R}^n} e^{i\zeta\delta_{2^{-k_P}}(x-c(I_P))}m(\zeta+\delta_{2^{k_P}}(c(\omega_{P(0)})-N))\widehat{\phi}(\zeta)d\zeta\right|. \tag{8.1}$$

Let us fix x and P and take i to be such that $\rho(x-c(I_P))=|x_i-c(I_P)_i|^{1/\alpha_i}$. From repeated integration by parts we see that (8.1) is bounded by

$$2^{-k_P|\alpha|/2}(2^{-k_P}\rho(x-c(I_P)))^{-\nu'\alpha_i}\int_{\mathbb{R}^n}\left|\partial_{\zeta_i}^{\nu'}(m(\zeta+\delta_{2^{k_P}}(c(\omega_{P(0)})-N))\widehat{\phi}(\zeta))\right|d\zeta,$$

for every integer $\nu'\ge 0$. We set $\nu'=\lceil\nu/\alpha_i\rceil\le\nu$. Since $N\notin\omega_{P(0)}$ we have $|\zeta+\delta_{2^{k_P}}(c(\omega_{P(0)})-N)|\gtrsim 1$. Therefore,

$$\int_{\mathbb{R}^n}\left|\partial_{\zeta_i}^{\nu'}(m(\zeta+\delta_{2^{k_P}}(c(\omega_{P(0)})-N))\widehat{\phi}(\zeta))\right|d\zeta\lesssim\|m\|_{\mathscr{M}^{\nu'}}\le\|m\|_{\mathscr{M}^{\nu}}.$$

This concludes the proof of (4.15) in the case that $\rho(x-c(I_P))\ge 1$. In the case $\rho(x-c(I_P))\le 1$ we simply use the triangle inequality on (8.1). ☐

Proof of (4.16) *and* (4.17). If $c(I_P)=c(I_{P'})$, the estimates are trivial. Thus we may assume $c(I_P)\ne c(I_{P'})$. We have

$$|\langle\phi_P,\phi_{P'}\rangle|\le|I_P|^{-\frac12}|I_{P'}|^{-\frac12}\int_{\mathbb{R}^n}|\phi(\delta_{2^{k_P}}(x-c(I_P)))\phi(\delta_{2^{k_{P'}}}(x-c(I_{P'})))|dx. \tag{8.2}$$

Since

$$\rho(c(I_P)-c(I_{P'}))\le\rho(x-c(I_P))+\rho(x-c(I_{P'})),$$

at least one of $\rho(x-c(I_P)),\rho(x-c(I_{P'}))$ is $\ge\frac12\rho(c(I_P)-c(I_{P'}))$. Thus, splitting the integral over x accordingly and using rapid decay of ϕ, the right hand side of (8.2) is no greater than a constant times

$$|I_P|^{-\frac12}|I_{P'}|^{\frac12}(1+2^{-k_P}\rho(c(I_P)-c(I_{P'})))^{-\nu}+|I_P|^{\frac12}|I_{P'}|^{-\frac12}(1+2^{-k_{P'}}\rho(c(I_P)-c(I_{P'})))^{-\nu}.$$

Recalling that we assumed $|I_P| \geq |I_{P'}|$ we see that the previous display is bounded by a constant times

$$|I_P|^{-\frac{1}{2}}|I_{P'}|^{\frac{1}{2}}(1 + 2^{-k_P}\rho(c(I_P) - c(I_{P'})))^{-\nu}.$$

This proves (4.16). The estimate (4.17) can be proven in the same way, by using the decay estimate (4.15). ☐

References

[Car66] L. Carleson. On convergence and growth of partial sums of Fourier series. *Acta Math.*, 116:135–157, 1966.

[Duo01] J. Duoandikoetxea. *Fourier analysis*, volume 29 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001. Translated and revised from the 1995 Spanish original by D. Cruz-Uribe.

[Fef73] C. Fefferman. Pointwise convergence of Fourier series. *Ann. of Math. (2)*, 98:551–571, 1973.

[FR66] E. Fabes and N. Rivière. Singular integrals with mixed homogeneity. *Stud. Mat.*, 27(1):19–38, 1966.

[GPRY17] S. Guo, L. B. Pierce, J. Roos, and P.-L. Yung. Polynomial Carleson operators along monomial curves in the plane. *J. Geom. Anal.*, 27(4):2977–3012, 2017.

[GTT04] L. Grafakos, T. Tao, and E. Terwilleger. L^p bounds for a maximal dyadic sum operator. *Math. Z.*, 246(1-2):321–337, 2004.

[KT80] C. Kenig and P. A. Tomas. Maximal operators defined by Fourier multipliers. *Studia Math.*, 68:79–83, 1980.

[Lac04] M. T. Lacey. Carleson's Theorem: Proof, Complements, Variations. *Publ. Mat.*, 48(2):251–307, 2004.

[Lie09] V. Lie. The (weak–L^2) Boundedness of the Quadratic Carleson Operator. *Geom. Funct. Anal.*, 19:457–497, 2009.

[LT00] M. T. Lacey and C. Thiele. A proof of boundedness of the Carleson operator. *Math. Res. Lett.*, 7(4):361–370, 2000.

[PT03] M. Pramanik and E. Terwilleger. A weak L^2 estimate for a maximal dyadic sum operator on $\mathbb{R}^n$. *Illinois J. Math.*, 47(3):775–813, 2003.

[PY15] L. B. Pierce and P.-L. Yung. A polynomial Carleson operator along the paraboloid. 2015. arXiv:1505.03882. To appear in Rev. Mat. Iberoam.

[Sjö71] P. Sjölin. Convergence almost everywhere of certain singular integrals and multiple Fourier series. *Ark. Mat.*, 9:65–90, 1971.

[Ste93] E. M. Stein. *Harmonic Analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of *Princeton Mathematical Series*. Princeton University Press, 1993.

[Thi06] C. Thiele. *Wave Packet Analysis*. Number 105 in CBMS Regional Conference Series in Mathematics. AMS, 2006.

Mathematical Institute, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany. *Current Address:* Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Dr, Madison, WI-53703, USA

E-mail address: jroos@math.wisc.edu