Annals of Mathematics **192** (2020), 47–163
https://doi.org/10.4007/annals.2020.192.1.2

The Polynomial Carleson operator

By Victor Lie

Dedicated to Elias Stein on the occasion of his 80^{th} birthday celebration.

Abstract

We prove affirmatively the one-dimensional case of a conjecture of Stein regarding the L^p-boundedness of the Polynomial Carleson operator for $1 < p < \infty$.

Our proof relies on two new ideas: (i) we develop a framework for *higher-order wave-packet analysis* that is consistent with the time-frequency analysis of the (generalized) Carleson operator, and (ii) we introduce a *local analysis* adapted to the concepts of mass and counting function, which yields a new tile discretization of the time-frequency plane that has the major consequence of eliminating the exceptional sets from the analysis of the Carleson operator. As a further consequence, we are able to deliver the full L^p-boundedness range and prove directly—without interpolation techniques—the strong L^2 bound for the (generalized) Carleson operator, answering a question raised by C. Fefferman.

1. Introduction

In this paper we will discuss the following conjecture of E. Stein regarding the behavior of the so-called Polynomial Carleson operator:

Conjecture ([118], [121]). *Let G denote either $\mathbb{T}$ or $\mathbb{R}$ with $G^n := \prod_{j=1}^{n} G$, $n \in \mathbb{N}$. Further, let $\mathcal{Q}_{d,n}$ be the class of all real-coefficient polynomials in n variables with no constant term and of degree less than or equal to d, $d \in \mathbb{N}$, and let K be a suitable Calderón–Zygmund kernel on G^n. Then the Polynomial*

Keywords: Time-frequency analysis, Carleson's theorem, Polynomial Carleson operator, higher-order wave-packet analysis

AMS Classification: Primary: 42A20, 42A50.

The author was supported by the National Science Foundation under Grant No. DMS-1500958. The current paper was initially revised while the author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2017 semester. The most recent revision was performed while the author was supported by the National Science Foundation under Grant No. DMS-1900801.

Carleson operator defined as[1]

$$C_{d,n}f(x) := \sup_{Q\in\mathcal{Q}_{d,n}} \left| \int_{G^n} e^{i\,Q(y)}\, K(y)\, f(x-y)\, dy \right| \tag{1}$$

obeys the bound

$$\|C_{d,n}f\|_{L^p(G^n)} \lesssim \|f\|_{L^p(G^n)} \tag{2}$$

for any $1 < p < \infty$.

The main result of our paper is

Main Theorem. *The above conjecture holds for* $n = 1$.

The problem concerning the boundedness properties of the Polynomial Carleson operator has a very rich history with numerous connections to other branches of analysis. Many of these will be extensively discussed in Section 1.3 below. In what follows, though, we prefer to start with a very concise outline of the key results that shaped the above conjecture and our present result, saving for later a more detailed account of the context in which these were developed.

1.1. *Previous directly related results*: *a brief overview.* Stein's conjecture on the Polynomial Carleson operator can be regarded

- in the case $d = 1$ and $n = 1$ as an extension of the celebrated Carleson–Hunt Theorem ([15], [56]) asserting that $C_{1,1}$ is bounded from L^p to L^p as long as $1 < p < \infty$; and
- in the case $d = 1$ and general n as an extension of Sjölin's result ([110]); see also [102], [47] and [68].

Under the crucial limiting assumption that the supremum in (1) be taken over polynomials with *no linear term*, a special case of the conjecture was established in work of Stein ([118] for dimension $n = 1$ and quadratic polynomials) and Stein–Wainger ([121] for general dimensions n and polynomials of arbitrary degree d). Due to the absence of linear terms in the phase of the kernel, the Stein and Stein–Wainger results do not contain the Carleson–Hunt Theorem.[2]

Finally, in [81], we made a significant advance by proving the L^2-weak boundedness of the full Quadratic Carleson operator $C_{2,1}$—incorporating polynomials with linear terms—in dimension one. For this, we developed a new

[1] Throughout the paper, for notational simplicity, we will omit the principal value symbol from all the mathematical displays.

[2] The operators considered by Stein and Wainger have no (generalized) modulation symmetry; thus, based on the symmetry complexity heuristic discussed in Section 1.4 below, one expects the analysis of such operators to be significantly simpler than that of the one corresponding to the Carleson operator.

approach to the time-frequency analysis of the quadratic phase, relying on the so-called *relational* perspective introduced in Section 2 of that paper. To this framework for quadratic wave-packet analysis, we adapted the ideas presented by Fefferman in his reproof of Carleson's theorem ([36]).

1.2. *Insights in our proof.* Passing now to the mathematical aspects of the present paper, we mention here the two main ideas on which our proof is based:

- Development of the *proper framework for the higher-order wave-packet theory* that in our context needs to be adapted to the time-frequency analysis of the (Polynomial) Carleson operator.
- A *new discretization* of the family of time-frequency tiles arising in the decomposition of our operator, which is a manifestation of the *local analysis* methodology that we develop around the newly introduced concept of mass adapted to a spatial region and local counting function.[3] This discretization has as a major implication the elimination of exceptional sets from the analysis of the Carleson operator. This latter fact has in turn two main consequences: (i) it yields boundedness for the complete range of exponents for the one-dimensional case of Stein's conjecture, and (ii) it provides for the first time a direct proof—without recourse to interpolation—of the L^2-boundedness of the Carleson operator, thus answering an open question raised by C. Fefferman in [36].

Another interesting aspect worth mentioning here is that, based on a novel tile selection algorithm involving a single dyadic grid discretization of the time-frequency coordinate axes,[4] our proof develops and treats an *exact* discretization of our Polynomial Carleson operator, unlike the previous approaches to the Carleson operator ([36], [73]) that relied on taking averages over suitable model operators. This aspect becomes quintessential when studying the topic of the boundedness of the Carleson operator near L^1; for the latter, please see the last item in Section 11 as well as Section 12 in [85].

Beyond these facts, there will be several other points in our approach (see, e.g., Section 8) that extend the intuition and methods developed in [81] for treating the particular case $d,\, p = 2$. These latter methods were further influenced by the powerful geometric and combinatorial ideas presented in [36].

This being said, we briefly elaborate on the two main ideas mentioned earlier:

---

[3]For more on this as well as on a philosophical outline of the proof of our main result, please see Section 5.

[4]See Section 7.2 and also Observation 12 therein.

Regarding the higher-order wave-packet framework, we develop a tile decomposition of the time-frequency plane into Heisenberg well-localized "curved regions" representing area-one neighborhoods of polynomials in the class $\mathcal{Q}_{d,1}$. The precise geometry of the tiles appears as a manifestation of the so-called *relational* perspective introduced in [81] and is directly related with a good control over the inner product—see, e.g., equation (56) below—of the "smaller pieces" (operators) into which $\mathcal{Q}_{d,1}$ is decomposed. Indeed, as the name suggests, this perspective stresses the importance of *interactions* between objects rather than simply treating such objects independently only in terms of their L^∞ size localization in time and frequency; for further details, see Section 2 in [81]. Our time-frequency representation of the tiles recovers, from a completely different angle, the more general uncertainty principle for differential operators developed by C. Fefferman in [37].

With respect to tile discretization, we design a new, spatially localized procedure of partitioning the family of tiles, which relies on a refined definition of the concept of mass of a tile, recursive stopping-time arguments, and a delicate combinatorial procedure. Within this process a special role is played by the local counting functions associated with suitable geometric configurations of tiles called "trees." All previously known estimates for controlling unions of such trees involved the L^∞ size of global counting functions, which in turn required one to excise the sets on which the L^∞ norms are too large. In particular, these "exceptional" sets caused a series of technical difficulties in all the earlier works regarding or related with the L^p-boundedness of the Carleson operator; these difficulties accounted for the lack—until now—of a direct approach to providing strong L^2 bounds. In the present paper one of the key insights is that we relate, via the localized mass parameter, the structure (i.e., the spatial location) of the trees to the local behavior of suitable counting functions, thereby enabling us to replace the previous L^∞-norm estimates with weaker BMO-norm-type estimates and thus eliminate the presence of the above mentioned exceptional sets.

1.3. *Historical background and motivation.* Before explaining the underlying motivation for Stein's conjecture, let us rewrite the expression (1) for the Polynomial Carleson operator in two equivalent forms that will put matters in proper perspective. Throughout this section, for simplicity, we will consider the case of $G=\mathbb{R}$.

First, notice that we can express

$$C_{d,n}f(x)=\sup_{\lambda}\left|T_\lambda f(x)\right|, \tag{3}$$

with

$$T_\lambda f(x):=\int_{\mathbb{R}^n} e^{i\,Q_\lambda(y)}\,K(y)\,f(x-y)\,dy, \tag{4}$$

where here $Q_\lambda(y) = \sum_{1\le|\beta|\le d} \lambda_\beta\, y^\beta \in \mathcal{Q}_{d,n}$ is a general real-coefficient polynomial with no constant term in n variables of degree at most d, with $\beta = (\beta_1, \ldots, \beta_n)$ a multi-index in[5] $\mathbb{N}^n$ and $\lambda = (\lambda_\beta)_\beta$ the sequence of coefficients of Q_λ.

Now, by making the change of variable $y \mapsto x - y$, we notice that the operator $C_{d,n}$ is part of a larger class of maximal operators of the type

$$T_* f(x) := \sup_\lambda \left| \int_{\mathbb{R}^n} e^{i\, Q_\lambda(x,y)}\, K(x,y)\, f(y)\, dy \right|, \tag{5}$$

where Q_λ, $K : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{C}$ are such that the phase function Q_λ is smooth and real-valued while K is a suitable integral kernel that is smooth away from the main diagonal $x = y$.

Second, we note that it is possible to recast the problem of boundedness for $C_{d,n}$ without the parameter λ (and thus, of course, without the corresponding supremum), at the price of losing smoothness of the phase in the x-parameter of Q_λ in (5). Indeed, by applying the Kolmogorov–Seliverstov–Plessner linearization argument ([129]), one sees that the L^p-boundedness of $C_{d,n}$ follows from the corresponding L^p bounds for an operator of the form

$$\int_{\mathbb{R}^n} e^{i\, Q(x,y)}\, K(x,y)\, f(y)\, dy, \tag{6}$$

where in the specific situation of $C_{d,n}$ we have $Q(x,\cdot) \in \mathcal{Q}_{d,n}$ a real polynomial whose coefficients are measurable functions of x, and $K(x,y) = K(x-y)$ with K a suitable Calderón–Zygmund kernel on $\mathbb{R}^n$.

The interest in studying the Polynomial Carleson operator comes from several different directions, and with these alternative formulations of the operator in hand, we can now turn to discuss the motivations for such a study.

A. *Maximal singular oscillatory integrals in the Euclidean setting.* The key prototypical example of a *maximal* singular oscillatory integral is the so-called Carleson operator (presented as Example A.1 below). This operator arises naturally in the study of the almost-everywhere convergence of Fourier Series. This latter topic originates in the effort of nineteenth-century mathematics to provide a rigorous foundation for the theory of Fourier Series initiated by J. Fourier in [38]. As a very brief historical overview, we mention the following landmark results:

Dirichlet ([76]), who was a student of Fourier, established the convergence at all points of Fourier Series for *differentiable* functions, while Du Bois Reymond ([32]) subsequently showed the existence of *continuous* functions whose

[5] Throughout the paper we use the convention $\mathbb{N} := \{0, 1, 2, \ldots\}$.

Fourier Series diverge at a point (and in fact at any rational point).[6] Once H. Lebesgue ([75]) had established his theory of measure and integration—which provided the correct framework to understand the previous divergence pathologies as behavior on "negligible" sets—N. Luzin ([89]) conjectured in 1913 that the Fourier Series of any $f \in L^2(\mathbb{T})$ converges to f almost everywhere. A decade later, Luzin's Ph.D. student A. Kolmogorov ([61], [62]) showed surprisingly that there are functions in $L^1(\mathbb{T})$ whose Fourier series diverge (*almost*) *everywhere*. After decades of misbelief in light of Kolmogorov's result, L. Carleson proved in 1966 that Luzin's conjecture is in fact true ([15]), thereby setting the foundation for what is known today as time-frequency analysis.

By analogy with the approach to proving Lebesgue's differentiation theorem for L^1 functions via the L^1-weak bounds for the Hardy–Littlewood maximal function, Carleson established the almost-everywhere convergence of Fourier Series of L^2 functions by providing L^2-weak bounds for the corresponding maximal operator $\sup_{n\in\mathbb{N}} |S_n f(x)|$ derived from the sequence of partial Fourier sums $S_n f$ attached to f, which, up to admissible error terms, represents nothing other than the aforementioned Carleson operator $C := C_{1,1}$.

At this point, we can present several significant examples of operators in the literature that fit within the framework of either (5) or (6). These in turn will lead us naturally to consider the Polynomial Carleson operator:

Example A.1. Consider an operator as in (5), with $n = 1$, $Q(x, y) = \lambda\,(x - y)$, and $K(x, y) = K(x - y) = \frac{1}{x-y}$. Equivalently, in (6), set $n = 1$, $Q(x, y) = a(x) \cdot y$ with a measurable, and $K(x - y) = \frac{1}{x-y}$.

In this context,[7] (5) or (6) represents the Carleson operator over $\mathbb{R}$ whose L^2-weak boundedness implies and, based on Stein's maximal principle ([115]), is in fact equivalent to the affirmative answer to Luzin's conjecture.

The L^p bounds, $1 < p < \infty$, for the Carleson operator were established by R. Hunt in [56].

Example A.2. In (5), set $n \geq 1$, $Q_\lambda(x, y) = \lambda \cdot (x - y)$ and $K(x, y) = K(x-y)$ a Calderón–Zygmund kernel. Equivalently, in (6), set $n \geq 1$, $Q(x, y) = a(x) \cdot y$ with $a = (a_1, \ldots a_n)$ measurable, and K as before.

[6]It is worth mentioning that the surprising result in [32] generated a deep interest in understanding the set of divergence for a Fourier Series of a continuous function, which further was the main catalyst for the development of a new theory of measure and integration; for this, see Lebesgue's study [74].

[7]We mention here that by applying a general transference principle due to Marcinkiewicz and Zygmund one can show that L^p-bounds for the (generalized) Carleson operator over $\mathbb{R}$ or $\mathbb{T}$ are equivalent.

This situation corresponds to the n-dimensional Carleson operator for which full L^p bounds, $1 < p < \infty$, were provided by P. Sjölin in [110] and later reproved by different means in [102].

Example A.3. In (5), set $n = 1$, $Q_\lambda(x,y) = \lambda \cdot (x-y)^2$, and $K(x,y) = K(x-y) = \frac{1}{x-y}$, with the obvious analogue in (6): $n = 1$, $Q(x,y) = a(x) \cdot (x-y)^2$, and K as before.

This case was proposed and treated by E. Stein ([118]). Unlike Carleson's theorem in [15], whose proof relies on wave-packet analysis, this result is based on more standard Fourier analysis techniques, namely on obtaining a good asymptotic formula for the Fourier transform of the expression $e^{i\lambda y^2}/y$ followed by an application of TT^* methods.

Example A.4. In (5), set $n \geq 1$, $Q_\lambda(x,y) = \sum_{2\leq|\beta|\leq d} \lambda_\beta\,(x-y)^\beta \in \mathcal{Q}_{d,n}$ with $d \geq 2$, and $K_\lambda(x,y) = K(x-y)$ with K a standard Calderón–Zygmund kernel, again with the obvious analogue in (6).

This situation extends the previous setting from A.3 and was investigated by Stein and Wainger in [121]. Notice that this latter setting does not include Carleson's or Sjölin's results, since *no* linear term is allowed in Q_λ. The Stein–Wainger proof is based on Van der Corput estimates and again TT^* methods.

Convergent point of interests. A very natural motivating theme arises: to find a common path connecting the methods of proof and the results presented in Examples A.3 and A.4 (i.e., Stein ([118]) and Stein–Wainger ([121])) with those of Examples A.1 and A.2 (i.e., Carleson–Hunt ([15], [56]) and Sjölin ([110])). We thus arrive naturally at the definition of the Polynomial Carleson operator in (1) and the formulation of Stein's conjecture regarding its L^p bounds.

B. *Singular oscillatory integrals on nilpotent groups.* In an extensive study regarding harmonic analysis on nilpotent Lie groups, [104], [105], [106], Ricci and Stein proved that, under the assumptions that Q is a real polynomial in both variables $(x,y) \in \mathbb{R}^n \times \mathbb{R}^n$ and $K(x,y) = K(x-y)$ with K a standard Calderón–Zygmund kernel, the operator represented by (6) is bounded on $L^p(\mathbb{R}^n)$ for $1 < p < \infty$. This of course can be regarded as *a model case for our conjecture above*, in the situation in which the "stopping times" represented by the coefficients of the monomials in y in the expression $Q(x,\cdot) \in \mathcal{Q}_{d,n}$ are themselves polynomials in x.

Stein and Ricci's motivation in considering this problem relies on the fact that such operators appear naturally in three distinct but interrelated contexts:

- singular integrals on lower-dimensional varieties in $\mathbb{R}^n$ (see, e.g., [111], [120], [122]);

- twisted convolution on the Heisenberg group and extensions to other nilpotent groups (see, e.g., [42], [91], [92]); and
- Radon transforms and their application to the study of the $\bar{\partial}$-Neumann problem (see, e.g., [100], [101], [50], [17], [18]).

For more on this, we refer the interested reader to the specific examples corresponding to each of these topics and appearing in [105] (see also Chapters XI, XII and XIII in [117]).

C. *Connections with Radon-like transforms.* With $n \in \mathbb{N}$ as before, let $\gamma : \mathbb{R} \to \mathbb{R}^n$ be a suitable (smooth) curve. We define two operators on functions over $\mathbb{R}^n$ as follows:[8]

- the maximal function along γ given by

$$\mathcal{M}_\gamma f(x) := \sup_{0<\epsilon<1} \frac{1}{2\epsilon} \int_{|t|<\epsilon} |f(x-\gamma(t))|\, dt; \tag{7}$$

- the Hilbert transform along γ given by

$$\mathcal{H}_\gamma f(x) := \int_{|t|<1} f(x-\gamma(t))\, \frac{dt}{t}. \tag{8}$$

The theory of singular integral operators of type (8) arose naturally in the study of solutions of constant-coefficient parabolic differential operators; see the works of [57], [35], and [34]. One specific example is the L^2-boundedness of (8) obtained by Fabes ([34]) in the case $n = 2$ and $\gamma(t) = (t, t^2)$, by applying the method of rotations to a singular integral associated with the heat equation. This was later extended by several authors, e.g., in [119], [1], [60], [54].

The study of maximal operators of type (7) was hinted at by use of the method of rotations in connection with Poisson integrals on symmetric spaces ([116]). The first L^p results were obtained by Nagel, Rivière, and Wainger in [96], [97], while a more general Euclidean-translation-invariant theory was developed by Stein and Wainger ([120]) in the case of one-dimensional submanifolds. All of the above results relied fundamentally on (1) Fourier methods via the Plancherel formula and (2) a suitable non-degeneracy curvature condition on γ via the method of stationary phase.

With these settled, the interest naturally shifted to the case of "variable" curves[9] γ. Thus, in this new context, one is concerned with operators of the form

$$\mathcal{M}_{(\gamma)} f(x) := \sup_{0<\epsilon<1} \frac{1}{2\epsilon} \int_{|t|<\epsilon} |f(x-\gamma(x,t))|\, dt \tag{9}$$

[8]The function f here is assumed to be in $L^1_{loc}(\mathbb{R}^n)$.

[9]Or, more generally, submanifolds.

and the associated singular integral form

$$\mathcal{H}_{(\gamma)} f(x) := \int_{|t|<1} f(x - \gamma(x,t)) \, \frac{dt}{t}. \tag{10}$$

This more general situation brings many complications for which new methods needed to be developed; in particular, one finds oneself in a non-translation-invariant setting, suggesting that one needs to go beyond Fourier-analytic tools. A first step in this direction was made by Nagel, Stein, and Wainger in [98], where they obtained an L^2 result in the special case of some smooth variable curves γ.[10] Their approach relied on TT^* methods. This work was greatly extended in the deep study of [18]. Other extensions to more general contexts such as nilpotent Lie groups or integral operators arising from the study of boundary-value problems in connection with the $\bar{\partial}$-Neumann problem for strongly pseudo-convex domains were already discussed in the "Singular oscillatory integrals on nilpotent groups" subsection above. All of these results relied on various curvature and smoothness conditions.

In an effort to unify and extend many of the above themes one could aim to

(i) require minimal or no smoothness in the x-parameter, or
(ii) preserve smoothness but drop the curvature condition in the t-parameter.

The various possible combinations of the presence of one or both of the above items give rise to a new class of problems, which generally are significantly more involved than the problem described above and for which there is presently no satisfactory answer. To understand the relevance and difficulty of some of these classes of problems we list here several important examples; for simplicity, we focus only on the case $n = 2$ and hence $x = (x_1, x_2) \in \mathbb{R}^2$:

Example C.1: *The case* $\gamma(x,t) = (t, v(x)\, t)$. This is one of the most striking examples. Let us assume first that we only have (ii) above and thus presume v is sufficiently smooth. For v *analytic*, Bourgain ([12]) proved L^p bounds on (9),[11] while the analogous result for the Hilbert transform was proved in a slightly more general context by Stein and Street in [112].

Assuming now that both (ii) and (i) are present, the story is as follows: If $v(x) = v(x_1, x_2) = v(x_1)$ is a function of only *one variable* and only assumed to be *measurable*, then the L^2-boundedness of (10) is *equivalent* to Carleson's theorem on the pointwise convergence of Fourier Series discussed in Example A.1. (This equivalence is almost immediate and can be found for example in [70].) More general L^p bounds—but still not within the fully expected

[10] Note that here it is essential that $\gamma(x,t)$ be smooth not just in the t-parameter but also in the x-parameter.

[11] Strictly speaking his result is for $p = 2$, but the extension to the case $1 < p \leq \infty$ is more or less standard.

range $1 < p < \infty$—were only recently obtained in [5] and [6]. Other similar but slightly more general results (i.e., addressing a Lipschitz perturbation of a single variable vector field) can be found in [51] and [52].

Regarding the general setting of genuinely two-variable vector fields v, it is a well-known fact that mere measurability, or even α-Hölder continuity with any $\alpha < 1$, is not enough to guarantee any L^p bounds[12] for either (9) or (10). The difficult and long-standing open problem of whether or not Lipschitz regularity[13] of v is enough to imply any non-trivial L^p bounds for (9) is often referred to as the *Zygmund conjecture*. The analogous problem for the Hilbert transform (10) was raised by Stein and is currently also widely open. As of today, the best general[14] regularity result is due to Lacey and Li ([69]), who via time-frequency analysis proved—using only measurability assumptions on v—L^p control for $p > 2$ over the Hilbert transform restricted to annuli. As a consequence of this last result, in [70], they proved a conditional result of the following flavor: if a suitable Kakeya-type maximal operator obeys appropriate bounds then, assuming that v has $C^{1+\epsilon}$ regularity, one has that $\mathcal{H}_{(\gamma)}$ is bounded on $L^2(\mathbb{R}^2)$. For more on all these, we invite the reader to consult [69], [70] and [6] and the bibliographies therein.

Example C.2: *The case* $\gamma(x,t) = (t,\, v(x)\, t^2)$. In this situation we completely remove item (ii), reimposing a non-trivial curvature in t. If v is only assumed to be measurable, then L^p bounds with $2 < p \leq \infty$ are known to be true for (9) ([90]) and to fail for (10) ([58]). If v is Lipschitz, then L^p bounds for the full range $1 < p \leq \infty$ hold for both $\mathcal{M}_{(\gamma)}$ ([53]) and $\mathcal{H}_{(\gamma)}$ ([28]). Notice again that if $v(x) = v(x_1)$ is a measurable function of only *one variable*, then the L^2-boundedness of (10) is equivalent to Stein's result ([118]) discussed in Example A.3. above.

Example C.3: *The case* $\gamma(x,t) = (t \sum_{1\leq\beta\leq d} v_\beta(x_1)\, t^\beta)$ *with* $d \in \mathbb{N}$, $d \geq 2$ *and* v_β *measurable functions.* This represents a natural attempt to unify Examples C.1 and C.2 in terms of the t-variable behavior, at the price of restricting the x-dependence of the v_β's to only the first variable. Based on our comments above, one can easily see now that the L^2 bounds of (10) in this setting are in fact *equivalent* to the Polynoimal Carleson conjecture stated the beginning of our paper for the case $n = 1$ and $p = 2$.

The case $\gamma(x,t) = (t,\, \sum_{2\leq\beta\leq d} v_\beta(x_1)\, t^\beta)$ for both (9) and (10) within the maximal range $1 < p < \infty$ was very recently solved by the author in [87]. In the same extensive study we provide a new method for approaching

[12] Excepting of course the trivial case $p = \infty$ for the operator defined by (9).

[13] With suitable smallness condition on $\|v\|_{Lip}$.

[14] *I.e.*, with no extra assumption that v be essentially a Lipschitz perturbation of a single-variable vector field.

several classes of singular integral operators in particular including the one-dimensional Polynomial Carleson with no linear phase case addressed by Stein and Wainger in [121] and discussed earlier within Example A.4.

1.4. *Further motivation: a wave-packet analysis perspective.* Our discussion in this subsection aims to classify/group various families of operators depending on their behavior relative to the following fundamental classes of symmetries:[15]

- Translations:

$$\tau_y f(x) := f(x-y), \quad a \in \mathbb{R}; \tag{11}$$

- Dilations:

$$D_\lambda f(x) := \lambda^{\frac{1}{2}} f(\lambda x), \quad \lambda \in \mathbb{R}_+; \tag{12}$$

- Generalized modulations of order j, $j \in \mathbb{N}$:

$$M_{j,a_j} f(x) := e^{i a_j x^j} f(x), \quad a_j \in \mathbb{R}. \tag{13}$$

We end this introductory commentary by stating the following guiding heuristic:

A HEURISTIC SYMMETRY PRINCIPLE: *The classes of symmetries of an operator are responsible for the nature of the approach/techniques to be involved in the analysis of its boundedness properties.* Following this line, loosely speaking one observes a correlation between the richness of the class of symmetries obeyed by an operator and the difficulty of proving its boundedness properties, due to the fact that one's approach must be invariant under the operator's symmetries.

Below, by gradually increasing the complexity of our objects, we discuss several fundamental classes of operators.

1.4.1. *Calderón–Zygmund theory (wavelets): Hilbert transform.* As is well known, the classical Hilbert transform over $\mathbb{R}$, defined as

$$Hf(x) := \int_{\mathbb{R}} f(x-y)\, \frac{dy}{y}, \tag{14}$$

is the only L^2-bounded linear operator (up to linear combinations with the identity operator) that commutes with translations and dilations, that is,

(1) $H\,\tau_y = \tau_y H$;
(2) $H D_\lambda = D_\lambda H$.

The L^p-boundedness, $1 < p < \infty$, of the Hilbert transform is due to M. Riesz ([107]), and over the years several other proofs have been found. A particularly suggestive approach studies the action of the Hilbert transform over a wavelet system, using as an intermediate step the existence of wavelet

[15] Below one may consider $f \in L^1_{\text{loc}}(\mathbb{R})$.

systems that form bases for $L^2(\mathbb{R})$. Recall that a wavelet system may be generated by the discrete action of dilation and translation symmetries on a single function, that is, $\{D_{2^j}\tau_k\varphi\}_{k,j\in\mathbb{Z}}$ with φ a suitable smooth function on $\mathbb{R}$.

It is precisely this symmetry of the Hilbert transform with respect to the translation and dilation actions generating wavelet bases that gives symbolic value to the wavelet-based study of the Hilbert transform, thus confirming the principle stated above. In this sense, one can view the wavelet theory as a dyadic framework for Calderón–Zygmund theory.

1.4.2. *Standard/linear wave packet analysis* (*modulation invariance*): (1) *Carleson operator*. Over $\mathbb{R}$, the Carleson operator

$$Cf(x) = C_{1,1}f(x) := \sup_{a\in\mathbb{R}}\left|\int_{\mathbb{R}} e^{i\,a\,y}\,\frac{1}{y}\,f(x-y)\,dy\right| \tag{15}$$

can be rewritten as[16]

$$Cf(x) = \sup_{a\in\mathbb{R}}\left|M_a^*\,H\,M_a\,f(x)\right|, \tag{16}$$

where throughout the paper we denote the adjoint of an operator T by T^*.

It is now easily observed that the Carleson operator is a maximal (sublinear) operator that commutes with translations and dilations and is invariant under modulations. That is, beyond commuting with translations and dilations as in (1) and (2) above (with H replaced by C), the Carleson operator obeys the further symmetry

(3) $CM_a = C$.

Based on our symmetry principle, this suggests that any method one chooses to prove the L^2 (weak) boundedness of C should remain invariant under such symmetries, in particular, under the standard modulation symmetry. This was indeed the case in [15], where Carleson developed—in disguise—a time-frequency analysis of the "adapted" Fourier coefficients of the input function of C.

The same heuristic principle was later utilized explicitly by C. Fefferman in his influential new proof of Carleson's result ([36]), where he introduced the wave-packet discretization of the Carleson operator. Mirroring the wavelet approach in the Hilbert transform setting, Fefferman used elementary building blocks consisting of wave-packets, that is, objects of the form $\{M_{a2^{-j}}D_{2^j}\tau_k\varphi\}_{a,k,j\in\mathbb{Z}}$, where again φ is a suitable smooth function over $\mathbb{R}$.

Consequently, from the above description of the work in [15] and [36], we see that the natural framework for standard time-frequency analysis lies

[16] For notational simplicity we refer to modulations of order one as simply "modulations," and instead of $M_{1,a}$ we simply write M_a.

within wave-packet theory, which in turn relies on the action of the three relevant symmetries: dilations, translations and (standard) modulations.

1.4.3. *Standard/linear wave packet analysis* (*modulation invariance*): (2) *Bilinear Hilbert transform.* Consider the Bilinear Hilbert transform B, defined a priori for Schwartz functions $f,\, g \in \mathcal{S}(\mathbb{R})$ by

$$B(f,g)(x) := \int_{\mathbb{R}} f(x-t)\, g(x+t)\, \frac{dt}{t}. \tag{17}$$

This bilinear operator appeared in Calderón's study of the Cauchy integral on Lipschitz curves ([14]). In this context, Calderón conjectured that B maps boundedly $L^p \times L^q \to L^r$ whenever $1 < p,\, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$.

One of the key insights in approaching this problem is to realize that the Bilinear Hilbert transform shares many similarities with the Carleson operator above. Indeed, in addition to the by now standard symmetries of commutation with translation and dilation, one also has the modulation symmetry given by

$$B(M_a f, M_a g) = M_{2a} B(f,g). \tag{18}$$

Applying thus the heuristic principle from above, one expects wave-packet analysis to play a key role in this problem. The confirmation of this fact came in [71] and [72], where Lacey and Thiele proved that Calderón's conjecture holds under the supplementary restriction $r > \frac{2}{3}$. (Despite sustained effort, the remaining case $\frac{1}{2} < r \leq \frac{2}{3}$ is still open.) Using the time-frequency tools developed in these papers, they were able to give a third, concise, proof of Carleson's theorem ([73]).

1.4.4. *Higher order wave packet analysis* (*generalized modulation invariance*): (1) *Trilinear Hilbert transform.* After the Lacey–Thiele breakthrough, a series of papers (e.g., [94], [95], [93], [24], [23]) extended the modern time-frequency framework to many other classes of multilinear operators motivated by applications to (mainly) ergodic theory and non-linear scattering theory. However, in all these papers, the underlying common feature is that any of the treated operators are at least "morally" invariant under translations, dilations, and linear modulations. For this reason, these problems could be successfully addressed by the standard wave-packet theory developed for treating the Carleson operator and later the Bilinear Hilbert transform.

However, the situation changes if one proposes to investigate the boundedness of the so-called Trilinear Hilbert transform

$$T(f,g,h)(x) := \int_{\mathbb{R}} f(x+t)\, g(x+2t)\, h(x+3t) \frac{dt}{t}. \tag{19}$$

The motivation for considering this object goes well beyond naturally generalizing the operators introduced in Sections 1.4.1 and 1.4.3, as can be witnessed by the deep connections with the fields of number theory, additive combinatorics, and ergodic theory discussed in Section 1.4.7 below.

The main question for the Trilinear Hilbert transform (19) is whether T maps $L^p \times L^q \times L^r \to L^s$ boundedly with the expected Hölder condition $\frac{1}{p}+\frac{1}{q}+\frac{1}{r}=\frac{1}{s}$ and $1<p,\, q,\, r<\infty$ with, say, $s\geq 1$. Nothing is known about this question except that, if one drops the condition $s\geq 1$, there exist negative examples for suitable choices of p, q, r, and s as shown in [22].

A primary source of difficulty for this question is that T has *more symmetries* than those already mentioned: in addition to the translation, dilation, and linear modulation symmetries, T also obeys a *quadratic* modulation symmetry; that is, for $a\in\mathbb{R}$,

$$T(M_{2,3a}f, M_{2,-3a}g, M_{2,a}h)(x) := M_{2,a}\, T(f,\, g, h)(x). \tag{20}$$

Thus, according to our symmetry principle, the standard wave-packet theory is not efficient in this setting since the (linear) wave-packet framework is not invariant under quadratic modulations. Indeed, all previous attempts to approach this problem with only linear wave-packet theory have failed. Thus, developing a higher-order wave-packet theory that, in particular, includes quadratic wave packets, seems a natural enterprise toward a better understanding of this problem.

1.4.5. *Higher order wave packet analysis* (*generalized modulation invariance*): (2) *Polynomial Carleson operator.* Recall the one-dimensional Polynomial Carleson operator (of degree $d\in\mathbb{N}$)

$$C_{d,1}f(x) := \sup_{Q\in\mathcal{Q}_{d,1}} \left|\int_{\mathbb{R}} e^{i\,Q(y)}\,\frac{1}{y}\,f(x-y)\,dy\right|. \tag{21}$$

We immediately notice that our Polynomial Carleson operator enjoys translation, dilation, and linear modulation invariance, thus obeying all the preliminary conditions that point towards a wave-packet methodology in the treatment of this operator. However, one further notices that if $d\geq 2$, then beyond the previous symmetries, the Polynomial Carleson operator $C_{d,1}$ is further invariant under the action of higher-order modulations (see (13)) given by $\{M_{j,a_j}\}_{j\in\{2,\dots,d\}}$. Thus, based on our earlier considerations, it seems natural that a successful approach to Stein's conjecture on the L^p-boundedness of the Polynomial Carleson operator should involve higher-order wave-packet theory. As we will see, this is indeed the case; in our proof of the one-dimensional case of this conjecture we develop a new way of representing and understanding the time-frequency representation and interaction of higher-order wave packets.

We stress that our analysis of the Polynomial Carleson operator (including here the Quadratic Carleson operator partially treated in [81]) represents the first step in the present literature in passing from the (standard) linear to the higher-order wave-packet approach. With respect to the hierarchy of *symmetry complexity*, the Polynomial Carleson operator is one level up relative to the standard Carleson operator or the Bilinear Hilbert transform, while, if we fix $d = 2$ ($n = 1$), the Quadratic Carleson operator $C_{2,1}$ obeys similar symmetry invariances with the Trilinear Hilbert transform.

Finally, a word of caution: while the *symmetry complexity* paradigm serves as a helpful heuristic in understanding the level of difficulty and the nature of the approach involved in bounding certain operators, this hierarchy need not be taken ad litteram. Indeed, the deeper structure of a given operator may reveal several other subtleties that significantly impact the difficulty of addressing the operator's boundedness. For example, such subtleties likely render the problem of the boundedness Trilinear Hilbert transform extremely difficult and, in particular, likely more challenging than the boundedness problem solved in this paper for the Polynomial Carleson operator $C_{d,1}$. This is the case even though, for large $d \in \mathbb{N}$, the Polynomial Carleson operator has *more symmetries* than the Trilinear Hilbert transform. As a consequence, the Polynomial Carleson operator could be regarded as an intermediate milestone between our understanding of the Bilinear and the Trilinear Hilbert transform.

1.4.6. *Wave packet analysis in higher dimensions*: *Triangular Hilbert transform*. In this section we address the topic of multi-dimensional wave packet analysis in the context of "highly" singular integral operators. This is a very recent direction in the area of time-frequency analysis that complements the aforementioned discussion revolving around singular multilinear operators in one dimension. While time-frequency (wave-packet) analysis has been previously applied to higher-dimensional cases—see for example the discussion on the Carleson operator in $\mathbb{R}^n$ detailed in example A.2. of Section 1.3—these applications treated only integral operators having kernels with low-dimensional singularity (e.g., higher-dimensional Calderón–Zygmund kernels), for which one can essentially apply or extend one-dimensional time-frequency techniques with no major difficulties.

To clarify the above, we take as a prototype for our discussion the following family of $(n+1)$-linear forms (dual to n-dimensional n-linear Hilbert transforms):

$$\Lambda_{\vec{\beta}_0,\ldots,\vec{\beta}_n}(F_0,\ldots,F_n) := \int_{\mathbb{R}^n}\int_{\mathbb{R}} \prod_{j=0}^{n} F_j(\vec{x}-\vec{\beta}_j\, t)\,\frac{dt}{t}\,d\vec{x}, \tag{22}$$

where here $n \in \mathbb{N}$, $n \geq 1$, $\vec{\beta}_j \in \mathbb{R}^n$, and $F_j \in \mathcal{S}(\mathbb{R}^n)$.

The main question in this context is whether there exists

$$1 < p_j < \infty \quad \text{with} \quad \sum_{j=0}^{n} \frac{1}{p_j} = 1 \tag{23}$$

for which the following holds:

$$|\Lambda_{\vec{\beta}_0,\ldots,\vec{\beta}_n}(F_0,\ldots,F_n)| \lesssim_{\{\vec{\beta}_j\}_{j=0}^n} \prod_{j=0}^n \|F_j\|_{L^{p_j}(\mathbb{R}^n)}. \tag{24}$$

Formulated in [66], this is a remarkable, difficult problem about which little is known, and nothing outside very special cases; see [25], [33], and [66]. In discussing its relevance, we will follow [33] and [66] and show that the question of the validity of the estimate (24) encompasses all the classes of problems discussed within Sections 1.4.1–1.4.5. Further motivation for considering the problem (24) comes from ergodic theory and will be discussed in the next section.

The extremely broad nature of this problem can be seen in the following observations:

- The case $n = 1$ trivially covers the boundedness of the Hilbert transform discussed in Section 1.4.1.
- The case $n = 2$ (for β_j's in general position) is wide open and, if proved affirmatively, recovers the uniform bounds for the one-dimensional bilinear Hilbert transform defined in (17) and treated in [80] (see also [45]).
- For the same case $n = 2$, it turns out that the bound in (24) is equivalent—with the same constant—with the corresponding bound for the so-called *Triangular Hilbert transform* defined by

$$\Lambda_{\triangle}(F_0,F_1,F_2) := \int_{\mathbb{R}^3} F_0(x,y)\, F_1(y,z)\, F_2(z,x)\, \frac{d(x,y,z)}{x+y+z}, \tag{25}$$

  which further implies suitable L^p bounds for the Carleson operator. At this point it is worth saying that controlling the L^p bounds of $\Lambda_{\triangle}$ gives, via the method of rotations, further control over the less singular expression studied in [25] and defined by

$$\tilde{\Lambda}^K_{B_0,B_1,B_2}(F_0,F_1,F_2) := \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} \prod_{j=0}^2 F_j(\vec{x} - B_j\,\vec{t})\, K(\vec{t})\, d\vec{t}\, d\vec{x}, \tag{26}$$

  where in this case $B_j \in \mathcal{M}_2(\mathbb{R})$ and K is a 2-dimensional Calderón–Zygmund kernel.
- The case $n = 3$ encapsulates as a particular case the Trilinear Hilbert transform introduced in Section 1.4.4.
- Finally, for general $n \in \mathbb{N}$ and suitable particular choices of $\{F_j\}_{j=0}^n$, the bounds in (24) are equivalent to the corresponding L^p bounds for the Polynomial Carleson operator.

1.4.7. *Wave-packet analysis*: *connections with number theory*, *additive combinatorics*, *and ergodic theory*. We end this extensive motivational chapter of our paper with a concise discussion of the deep interconnections between wave-packet (time-frequency) analysis and fields such as number theory, additive combinatorics, and ergodic theory. More precisely, our discussion will focus on the relationships between

- on the one hand—as part of (higher order) wave-packet analysis—the problems of the boundedness of the Bilinear and Trilinear Hilbert transforms (see Sections 1.4.3 and 1.4.4) and that of the Triangular Hilbert transform (see Section 1.4.6); and
- on the other hand—as part of number theory/additive combinatorics—problems of counting additive patterns in subsets of integers (Roth's and Szemerédi's theorems on arithmetic progressions) and the deeply interrelated problems—as part of the ergodic theory area—of the behavior of Furstenberg's non-conventional averages and of the pointwise convergence of bilinear Birkhoff averages for two commuting measure-preserving transformations.

I. *Counting additive patterns in sets.* We start our journey with the following famous theorem of Roth ([108]): if $A \subseteq \mathbb{Z}$ has positive upper density, then A contains infinitely many arithmetic progressions of length 3. This can be seen as a particular case of Szemerédi's theorem ([123], [124]) asserting that under the same hypothesis A contains infinitely many arithmetic progressions of length k for any $k \geq 3$.

Now both Roth's and Szemerédi's theorems rely on a certain dichotomy between (pseudo-)randomness and structure. Indeed, assuming without loss of generality that $A \subseteq \mathbb{Z}_N := \{1, \ldots, N\}$ with $|A| = \delta N$ for some fixed $\delta \in (0,1)$, the schematic idea is as follows: On the one hand, if A is "close" to a random set, then one expects to have roughly $\delta^k N^2$ k-term arithmetic progressions contained in A. On the other hand, somehow resembling in spirit Freiman's celebrated sum-set theorem ([39], [40]), a non-random set is expected to have a certain amount of additive structure. In this situation, one can exploit this structure and restrict the discussion to a long arithmetic subprogression of $\mathbb{Z}_N$ on which A has higher density. The desired conclusion is obtained by iterating this argument.

In order to better expose some interesting connections with wave-packet theory, we will now very briefly outline the above δ-density argument via a Fourier-analytic approach that also reveals the parallelism between the proofs of Roth's and Szemerédi's theorems.

We start by mentioning that both proofs rely on the study of the k-linear form

$$\Lambda_k(f_1, f_2, \ldots, f_k) := \frac{1}{N^2} \sum_{x,r \in \mathbb{Z}_N} f_1(x)\, f_2(x+r)\, \ldots\, f_k(x+(k-1)r), \tag{27}$$

where for each $1 \le j \le k$, $f_j : \mathbb{Z}_{\mathbb{N}} \mapsto \mathbb{C}$. Observe[17] that for $f_j = 1_A$, the form $\Lambda_k(1_A, \ldots, 1_A)$ represents the N^2-normalized number of k-term arithmetic progressions in A.

Now coming closer to time-frequency themes, notice that (27) is related to the single-scale discretizated version of the dualized form attached to the Bilinear Hilbert transform in (17) for $k = 3$ and to that of the Trilinear Hilbert transform in (19) for $k = 4$.

Assume first that $k = 3$. Then, via a standard computation, one obtains

$$\Lambda_3(1_A, 1_A, 1_A) = \widehat{1}_A(0)^3 + \sum_{\xi \neq 0} \widehat{1}_A(\xi)^2\, \widehat{1}_A(-2\xi) =: I + II. \tag{28}$$

As it will soon turn out, I represents the dominant term and II an error term. It can now be easily seen that $I = \delta^3$, thus matching the heuristic expected number of 3-progressions for a random set. In treating the second term we have two possible scenarios:

- The pseudo-random scenario: for all $\xi \neq 0$, one has $|\widehat{1}_A(\xi)| \le \frac{\delta^2}{2}$. In this situation via Parseval's identity one immediately gets that $II \le \frac{\delta^3}{2}$, from which one deduces $|\Lambda_3(1_A, 1_A, 1_A)| \ge \frac{\delta^3}{2}$, thus obtaining the desired conclusion.
- The structured scenario: there exists $\xi \neq 0$ such that $|\widehat{1}_A(\xi)| > \frac{\delta^2}{2}$. In this case we have that the inner product of 1_A against the *linear* wave-packet $e^{\frac{2\pi i x\xi}{N}}$ is large, or equivalently that 1_A correlates with the character $e^{\frac{2\pi i x\xi}{N}}$. As a consequence A must be biased along a suitable arithmetic progression of length $\approx \delta^2\sqrt{N}$ on which A has density at least $\delta + \frac{\delta^2}{100}$, thus yielding a $\frac{\delta^2}{100}$ density increment of our set.

The proof of Roth's theorem is completed by iterating the above algorithm: if at some point we land in the random case, we are done; otherwise, the density must increase successively, eventually reaching 1, at which point the statement of the theorem is trivially satisfied.

We pass now to presenting a brief outline of Gowers's proof of Szemerédi's theorem ([43], [44]), so $k \ge 4$. As in the discussion above, setting $f_A := 1_A - \delta 1_{\mathbb{Z}_N}$, we proceed by estimating

$$\Lambda_k(1_A, \ldots, 1_A) = \Lambda_k(\delta 1_{\mathbb{Z}_N}, \ldots, \delta 1_{\mathbb{Z}_N}) + \text{Other Terms} = I + II. \tag{29}$$

We notice that $I = \delta^k$ while II consists of $2^k - 1$ terms involving at least one copy of f_A. The main challenge at this point is to find the right substitute for the dichotomy presented in the case $k = 3$ that relies on (the lack of) correlation with a linear wave-packet. This was solved in [43] and [44] by introducing the

[17] In the reasonings immediately below we denote by $\hat{f}$ the discrete Fourier transform of f on $\mathbb{Z}_N$ and by 1_A the characteristic function of A.

key concept of (Gowers) d-uniformity norms $\|\cdot\|_{U^d}$, the definition of which we omit for brevity. Following this, one has that II can be bounded by an expression involving $\|f_A\|_{U^{k-1}}$. With these observations, one has

- The pseudo-random scenario (or equivalently A is $k-1$-uniform): $\|f_A\|_{U^{k-1}} < c(\delta)$. In this situation, for $c(\delta)$ small enough, one gets that $|II| \leq \frac{\delta^k}{2}$ and thus $\Lambda_k(1_A, \ldots, 1_A) \geq \frac{\delta^k}{2}$, finishing the argument.
- The structured scenario: $\|f_A\|_{U^{k-1}} \geq c(\delta)$. This is a much harder situation to analyze; unlike the $k = 3$ case, here one cannot derive a bias of A against a linear character, but rather one hopes that A correlates with a *higher-order* wave-packet.[18] This latter claim may be seen as a manifestation of the so-called inverse conjecture for the Gowers norms (see [48], [49]). As it turns out, Gowers's original approach requires only a weaker form of the above conjecture still suffiecient to isolate polynomial patterns, which, via an exponential sum argument, produce the desired density increment on a subprogression of Z_N, thereby finishing the inductive step.

II. *Furstenberg's non-conventional averages.* The aim of this subsection is to give a very succinct account of some of the interesting interconnections between wave-packet analysis and ergodic theory. We start our discussion by introducing a parallelism between the following two categories of objects:

- Multilinear Euclidean averages:

$$T_{\vec{a},\mathbb{R},r,n}(f_1, \ldots, f_n)(x) := \frac{1}{2r} \int_{|t|\leq r} \prod_{j=1}^{n} f_j(x + a_j\, t)\, dt, \tag{30}$$

  where here $n \in \mathbb{N}$, $\vec{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, $x,\, t \in \mathbb{R}$, $r > 0$, and $f_j \in L^\infty(\mathbb{R})$ for $1 \leq j \leq n$.
- Multilinear ergodic averages:

$$T_{\vec{a},X,N,n}(f_1, \ldots, f_n)(x) := \frac{1}{2N+1} \sum_{|l|\leq N} \prod_{j=1}^{n} f_j(S^{a_j\, l} x), \tag{31}$$

  where here (X, σ, μ, S) is a dynamical system,[19], $\vec{a} = (a_1, \ldots, a_n) \in \mathbb{Z}^n$, $x \in X$, $N,\, n \in \mathbb{N}$, and $f_j \in L^\infty(X)$ for $1 \leq j \leq n$.

As it turns out, the problems regarding the behavior of (31) as $N \to \infty$, specifically norm or almost-everywhere convergence, are related to the boundedness properties of the maximal operator associated to (30), that is, $T^*_{\vec{a},\mathbb{R},n} =$

[18]Notice here the similarity with the increased complexity of the symmetry behavior of the Trilinear Hilbert transform as compared with that of the Bilinear Hilbert transform, discussed in Sections 1.4.3 and 1.4.4.

[19]That is, (X, σ, μ) is a complete probability space with an invertible bimeasurable transformation $S:\ X \to X$ such that $\mu S^{-1} = \mu$.

$\sup_{r>0}|T_{\vec{a},\mathbb{R},r,n}|$. Indeed, standard transference results establish an equivalence between the boundedness properties of $T^*_{\vec{a},\mathbb{R},n}$ and those corresponding to $T^*_{\vec{a},\mathbb{X},n}$.

Now in the Euclidean setting, Stein's maximal principle ([115]) reveals that in many relevant instances the almost-everywhere convergence of (30) is in fact equivalent to obtaining suitable weak $L^p(\mathbb{R})$ bounds for the maximal operator $T^*_{\vec{a},\mathbb{R},n}$. Consequently, the almost-everywhere convergence of (30) for suitable $f_j \in L^{p_j}(\mathbb{R})$ is usually derived by applying the following strategy: one first proves weak $L^p(\mathbb{R})$ bounds for $T^*_{\vec{a},\mathbb{R},n}(f_1,\ldots,f_n)$ for a proper choice of p and $f_j \in C_0^\infty(\mathbb{R})$ (with the norm dependent only on $\|f_j\|_{L^{p_j}(\mathbb{R})}$), followed by a trivial density argument.

In the ergodic theoretical setting, though, the situation is significantly more delicate since there is no direct analogue for the dense subclass $C_0^\infty(\mathbb{R})$. Thus, in this instance, one first needs to prove an almost-everywhere convergence result for $T_{\vec{a},X,N,n}(f_1,\ldots,f_n)$ for $f_j \in L^\infty(X)$, and only after that extend the convergence result to $f_j \in L^{p_j}(X)$, once one shows that $T^*_{\vec{a},X,n}(f_1,\ldots,f_n)$ (or equivalently $T^*_{\vec{a},\mathbb{R},n}(f_1,\ldots,f_n)$) obeys suitable $L^{p_1}\times\cdots\times L^{p_n}$ to weak L^p bounds for a proper choice of $p>0$.

Finally, it is worth mentioning that one can circumvent the lack of the density argument counterpart in the ergodic theoretical setting if one is able to replace the weak L^p bound of $T^*_{\vec{a},\mathbb{R},n}(f_1,\ldots,f_n)$ by a suitable L^q variational norm estimate for $T_{\vec{a},\mathbb{R},r,n}$.

Below we analyze several interesting cases:

- $n=1$ (linear averages): In this situation, the almost-everywhere convergence of (30) as $r\to 0$ is equivalent to Lebesgue's differentiation theorem, and it follows directly (as well as the norm convergence result) from the Hardy–Littlewood maximal theorem, which states that the maximal operator $T^*_{a,\mathbb{R},1}$ is bounded from $L^p(\mathbb{R})$ to $L^p(\mathbb{R})$ for any $1<p\le\infty$ and for $p=1$, maps $L^1(\mathbb{R})$ into $L^{1,\infty}(\mathbb{R})$. On the ergodic theoretic side, the L^2 norm convergence of (31) is the celebrated von Neumann mean ergodic theorem, [99], while the almost-everywhere convergence result is known as Birkhoff's pointwise ergodic theorem ([10]). As mentioned in the introductory discussion, the almost-everywhere convergence of (31) needs first to be established for $L^\infty(X)$ functions, followed by the bounds on $T^*_{a,X,1}$ obtained via transference from the Euclidean counterpart.
- $n=2$ (bilinear averages): In the Euclidean setting, the central operator that governs the almost-everywhere (and trivially the norm) convergence of (30) is given by the (Sub)Bilinear Maximal operator

$$
\begin{aligned}
M_2(f_1,f_2)(x) &:= T^*_{\vec{a},\mathbb{R},2}(f_1,f_2)(x)\\
&:= \sup_{r>0}\frac{1}{2r}\left|\int_{|t|\le r} f_1(x+a_1\,t)\,f_2(x+a_2\,t)\,dt\right|,
\end{aligned}
\tag{32}
$$

which is simply the maximal analogue—for arbitrary $\vec{a} \in \mathbb{R}^2$ (with $a_1 \neq a_2$)—of the Bilinear Hilbert transform defined in (17). One can obtain via standard results for the classical Hardy–Littlewood maximal operator and interpolation that M_2 maps $L^{p_1}(\mathbb{R}) \times L^{p_2}(\mathbb{R}) \to L^p(\mathbb{R})$ for $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} < 1$. Using subtle wave-packet techniques in the spirit of those implemented for its singular integral counterpart (17), M. Lacey proved in [67] that the previous range of M_2 can be extended to the situation $\frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} < \frac{3}{2}$, $1 < p_1, p_2 \leq \infty$. As in the case of the Bilinear Hilbert transform, the complete possible range $\frac{3}{2} \leq \frac{1}{p} = \frac{1}{p_1} + \frac{1}{p_2} < 2$ remains an interesting difficult open question.

In the ergodic setting the (L^2-)norm convergence can be seen as a very particular[20] case of the result obtained by Conze and Lesigne in [19], while the almost-everywhere convergence result of the averages in (31) was first obtained by Bourgain in [13] for $f_1, f_2 \in L^\infty(X)$, which as usual can now be extended to more general L^p spaces based on the bounds obtained for M_2.

Finally, it is worth mentioning that both Bourgain's result in [13] as well as its singular ergodic variant—namely, the almost-everywhere convergence of the so-called Ergodic Bilinear Hilbert transform—can be proved via wave-packet analysis intertwined with suitable variational estimates, as shown by C. Demeter in [21].

- General $n \geq 2$ (Furstenberg averages): In the Euclidean setting, it is shown in [24] that the maximal operator $T^*_{\vec{a},\mathbb{R},n}$ is bounded from $L^{p_1} \times \cdots \times L^{p_n}$ to L^p for $\frac{1}{p} = \frac{1}{p_1} + \cdots + \frac{1}{p_n} < \frac{1}{2}$ and $1 < p_i \leq \infty$. In the ergodic realm, letting $a_j = j$, the averages in (31) are referred to as Furstenberg averages and appeared naturally in the context of Furstenberg's recurrence theorem ([41]), which essentially states that for any $f \neq 0$, $f \geq 0$ almost everywhere, one has

$$\liminf_{N\to\infty} \int_X T_{\vec{a},X,N,n}(f,\ldots,f)\, f\, d\mu > 0. \tag{33}$$

  This was used to provide a different proof of Szemerédi's theorem on arithmetic progressions, ((([123], [124])), discussed in the previous section. The L^2-norm convergence of (31) (also extendable to L^p, $1 \leq p < \infty$) was proved independently by Host and Kra ([55]) and Ziegler ([127]). The corresponding problem for almost-everywhere convergence is still wide open for $n \geq 3$.

We end this subsection by mentioning that there are many other interesting works that further exploit the deep connections between wave-packet analysis and ergodic theory, including the so-called return times theorem ([11], [23]), as well as extensions of the expressions in (30) and (31) to averages along cubes (see [55] and [24]).

[20]For more on this, see the next subsection.

III. *The problem of two commuting transformations.* Building upon the above discussion of multiple ergodic averages, one can switch the focus to the situation in which the averages involve more than one transformation. Concretely, taking $n = 2$, and assuming $S, U : X \to X$ are two μ-preserving transformations that commute, i.e., $US = SU$, one is prompted to study the behavior of

$$T_{X,N,2,S,U}(f, g)(x) := \frac{1}{N} \sum_{l=0}^{N-1} f(S^l x)\, g(U^l x) \tag{34}$$

for f, $g : X \to \mathbb{C}$ measurable, $n \in \mathbb{N}$, and $x \in X$.

Conze and Lesigne ([19]) proved the L^2-norm convergence of (34) for f, $g \in L^\infty(X)$; their result can be extended to the L^p case for $f \in L^{p_1}(X)$, $g \in L^{p_2}(X)$, and $\frac{1}{p} \geq \frac{1}{p_1} + \frac{1}{p_2}$ with $1 \leq p < \infty$. Higher-order variants of (34) are also known to converge in norm due to the work in [125], [4], and [126].

The other direction, concerning the almost-everywhere convergence of (34), constitutes one of the fundamental open problems in the area. Only very particular results are known, such as the special case $U = S^{-1}$ (with S invertible) due to Bourgain in [13]. This last result was later reproved (and strengthened quantitatively) in [30] using wave-packet (time-frequency) techniques in order to control a suitable variational norm of $T_{\mathbb{R},N,2,S,S^{-1}}$.

Returning to the general norm-convergence setting, very recently, the authors in [33] were able to obtain a parallelism with part of the results in [19], by proving quantitative bounds for a suitable variational norm involving the L^2 norm differences for $T_{X,N,2,S,U}$. Interestingly enough, via Calderón's transference principle, the latter problem is reduced to the task of obtaining the analogous variational norm estimates for an averaging operator that shares many similarities with the Triangular Hilbert transform in (25). It is worth mentioning that the proof of this result is based not on wave-packet analysis but rather on techniques that resemble the energy methods in partial differential equations and that involve integration by parts, positivity arguments, and the Cauchy–Schwarz inequality.

Finally, notice that, via the Triangular Hilbert transform, one is able to establish an interesting link between two major open problems: on the ergodic theoretical side, the problem of the almost-everywhere convergence of the bilinear averages for two commuting transformations and, on the harmonic analysis side, the problem of the boundedness of the Trilinear Hilbert transform.

1.5. *Structure of the paper.* Next, we briefly outline the structure of our paper:

- In Section 2 we establish notation and present the general procedure of constructing our tiles.
- In Section 3 we elaborate on the discretization of our operator $C_{d,1}$.

- Section 4 is dedicated to the study of the interaction between tiles where some key concepts—the density and the geometric factor of a tile/pair of tiles—are introduced.
- In Section 5 we develop a local analysis methodology that is foundational for our approach and present a philosophical outline of the proof of our main result. The reader who is interested in capturing the essence of our argument with minimal technicalities should thus consult this section.
- The new discretization algorithm of the family of tiles is worked out in detail in Section 6.
- Next, in Section 7, we present the main definitions and reduce the Main Proposition to two auxiliary propositions, Proposition 1 and Proposition 2.
- Section 8 deals with the proof of Propositions 1.
- Section 9—the most technical one—prepares the ground for the proof of Proposition 2.
- Section 10 addresses the proof of Proposition 2 while Section 11 is dedicated to some final remarks.
- In the appendix we include several useful results regarding the distribution and growth of polynomials.

Finally, given that in many respects [36] and [81] can be regarded as a foundation for this paper, when possible, we have chosen to preserve here the notation, definitions, and general structure of those earlier works.

Acknowledgements. First and foremost, I am grateful that during my formative undergraduate years I had the chance to interact with and learn from a gifted group of professors and researchers revolving around the "Simion Stoilow" Institute of Mathematics of the Romanian Academy. A special role in this group is played by Nicolae Popa—a man whose friendship helped shape my mathematical development and with whom I have always resonated on a romantic view of mathematics. I would like to express my deep gratitude to Charlie Fefferman for his continued support, as well as for reading parts of the manuscript and providing useful feedback. I would like to thank my Ph.D. adviser Christoph Thiele, who first introduced me to the question regarding the behavior of the Quadratic (Polynomial) Carleson operator and who inspired me to visualize the texture of time-frequency analysis from a geometric perspective. In this latter direction I would also like to extend my thanks to John Garnett and Terence Tao for cultivating, throughout my graduate student years at UCLA, an intuitive geometric approach to analysis that profoundly impacted my views of the field. I am indebted to Jim Wright for useful advice about the history of the present problem and to my colleague Zubin Gautam for his help in improving the English presentation of the paper. I also thank Pavel Zorin-Kranich for pointing out some minor typos in an earlier version of this paper. Finally, I thank the anonymous referees whose comments improved the presentation of this paper.

2. **Notation and construction of the tiles**

We start by introducing the corresponding canonical dyadic grids on[21] $[0,1)=\mathbb{T}$ and on $\mathbb{R}$. Throughout the paper the letters I and J refer to dyadic intervals corresponding to the grid associated with $\mathbb{T}$ while the Greek letters $\alpha^1,\ldots,\alpha^d$, with $d\in\mathbb{N}$ a fixed parameter, stand for dyadic intervals associated with the grid in $\mathbb{R}$. All the dyadic intervals considered in this paper are of the form $[k2^{-j},\,(k+1)2^{-j})$ for appropriate $k,\,j\in\mathbb{Z}$.

A *tile* P is a $(d+1)$-tuple of dyadic intervals, i.e.,

$$P=[\alpha^1,\alpha^2,\ldots,\alpha^d,I],\ \text{ s.t. } |\alpha^j|=|I|^{-1},\ \ j\in\{1,\ldots,d\}. \tag{35}$$

For notational simplicity, we will often refer to $P=[\alpha^1,\alpha^2,\ldots,\alpha^d,I]$ as $P=[\vec{\alpha},I]$, where here $\vec{\alpha}=(\alpha^1,\alpha^2,\ldots,\alpha^d)$.

The collection of all tiles P will be denoted by $\mathbb{P}$.

Now, for each tile $P=[\vec{\alpha},I]$, we will associate a *geometric time-frequency representation*, denoted with $\hat{P}$. The exact procedure is described in several steps:

- For I above, we set $x_I=(x_I^1,x_I^2,\ldots,x_I^d)\in\mathbb{T}^d$ to be the $d-$tuple defined inductively as follows: x_I^1,x_I^2 are the endpoints of the interval I, and then, for larger d, the remaining points are successive midpoints of the intervals resulted at the previous step(s). That is, if $d\geq 3$, then $x_I^3=\frac{x_I^1+x_I^2}{2}$ is the mid-point of I; next, if $d\geq 4$, $x_I^4=\frac{x_I^1+x_I^3}{2}$ is the mid-point of the left half of I; next, if $d\geq 5$, $x_I^5=\frac{x_I^3+x_I^2}{2}$ is the mid-point of the right half of I and so on until we reach the d-th coordinate.
- Recalling that $\mathcal{Q}_d$ stands for the class of all real polynomials of degree at most d, we make the following conventions: If not specified, q will always designate an element of $\mathcal{Q}_{d-1}$, while Q will refer to an element of $\mathcal{Q}_d$. When appearing together in a proof, q will designate the derivative of Q.
- We define

$$\mathcal{Q}_{d-1}(P):=\{q\in\mathcal{Q}_{d-1}\,|\,q(x_I^j)\in\alpha^j\quad\forall\ j\in\{1,\ldots,d\}\}$$

and set the notation

$$q\in P\ \text{ if and only if }\ q\in\mathcal{Q}_{d-1}(P). \tag{36}$$

[21] Depending on our convenience the symbol $\mathbb{T}$ stands for either $[-\frac{1}{2},\,\frac{1}{2})$—when appearing in the definition of the Polynomial Carleson operator, or $[0,1)$—when referring to the discretization of our time-frequency plane.

- With all these done, we define

$$\hat{P} := \{(x, q(x)) \mid x \in I \text{ and } q \in P\}. \tag{37}$$

The collection of all geometric tiles $\hat{P}$ will be denoted by $\hat{\mathbb{P}}$.

For each tile $P = [\vec{\alpha}, I] = [\alpha^1, \alpha^2, \ldots, \alpha^d, I] \in \mathbb{P}$, we associate the "*central polynomial*" $q_P \in \mathcal{Q}_{d-1}$ given by the Lagrange interpolation polynomial:

$$q_P(y) := \sum_{j=1}^{d} \frac{\prod_{\substack{k=1\\k\neq j}}^{d}(y - x_I^k)}{\prod_{\substack{k=1\\k\neq j}}^{d}(x_I^j - x_I^k)}\, c(\alpha^j). \tag{38}$$

Observation 1. Remark that, due to Lemma C in the appendix, one may think of $\hat{P}$ as roughly being the $|I|^{-1}$ neighborhood of the graph of the "central polynomial" q_P restricted to the spatial interval I.

Now, if I is any (dyadic) interval, we denote by $c(I)$ the center of I. Let I_r be the "right brother" of I, that is, the interval having the properties $c(I_r) = c(I) + |I|$ and $|I_r| = |I|$; similarly, the "left brother" of I will be denoted I_l with $c(I_l) = c(I) - |I|$ and $|I_l| = |I|$. If $a > 0$ is some real number, by aI we mean the interval with the same center $c(I)$ and with length $|aI| = a|I|$; the same conventions apply to intervals $\{\alpha^k\}_k$.

In the following we will also work with dilates of our tiles: for $a > 0$ and $P = [\alpha^1, \alpha^2, \ldots, \alpha^d, I]$, we set $aP := [a\alpha^1, a\alpha^2, \ldots, a\alpha^d, I]$. Similarly, we write

$$a\hat{P} := \widehat{aP} = \{(x, q(x)) \mid x \in I \text{ and } q \in \mathcal{Q}_{d-1}(aP)\}.$$

Also, if $\mathcal{P} \subseteq \mathbb{P}$, then by convention $a\mathcal{P} := \{aP \mid P \in \mathcal{P}\}$; similarly, if $\hat{\mathcal{P}} \subseteq \hat{\mathbb{P}}$, then $\widehat{a\mathcal{P}} := \left\{\widehat{aP} \mid P \in \mathcal{P}\right\}$.

For $P = [\vec{\alpha}, I]$, we denote the collection of its neighbors by

$$N(P) := \{P' = [\vec{\alpha'}, I] \mid \alpha'^k \in \{\alpha^k,\ \alpha^k_r,\ \alpha^k_l\} \ \ \forall\, k \in \{1, \ldots, d\}\}.$$

Assume $P = [\vec{\alpha}, I_P]$. We define

$$I_{P^*} := \left[c(I_P) - \frac{17}{2}|I_P|,\, c(I_P) - \frac{3}{2}|I_P|\right) \cup \left[c(I_P) + \frac{3}{2}|I_P|,\, c(I_P) + \frac{17}{2}|I_P|\right) \tag{39}$$

and let

$$I_{P^*} =: \bigcup_{j=1}^{14} I_{P^*}^j \tag{40}$$

be the partition of I_{P^*} into dyadic intervals of length $|I_P|$.

Also we let

$$\tilde{I}_P := 17 I_P. \tag{41}$$

In many of the situations, for notational simplicity, we will abuse notation and identify[22] $P = [\vec{\alpha}, I_P] \in \mathbb{P}$ with its correspondent representation $\widehat{P} \in \widehat{\mathbb{P}}$. Similarly, we will often identify P^* with its geometric representation $\widehat{P}^*$, where

$$\widehat{P}^* := \{(x, q(x)) \mid x \in I_{P^*} \text{ and } q \in P\}. \tag{42}$$

Throughout the paper p will be the index of the Lebesgue space L^p and, unless otherwise mentioned, will obey $1 < p < \infty$. Also, p' will be its Hölder conjugate (i.e., $\frac{1}{p} + \frac{1}{p'} = 1$), while $p^* := \min(p, p')$.

For $f \in L^p(\mathbb{T})$, we denote by

$$Mf(x) := \sup_{x \in I} \frac{1}{|I|} \int_I |f|$$

the Hardy-Littlewood maximal function associated to f.

As in [81], for $x \in \mathbb{R}$, we set $\lceil x \rceil := \frac{1}{1+|x|}$.

For A, $B > 0$, we say $A \lesssim B$ (respectively $A \gtrsim B$) if there exists an absolute constant $C > 0$ such that $A < CB$ (respectively $A > CB$); if the constant C depends on some quantity $\delta > 0$, then we may write $A \lesssim_\delta B$. If $C^{-1}A < B < CA$ for some (positive) absolute constant C, then we write $A \approx B$. Also we write $A \ll B$ if and only if there exists a large constant, say $C > 100^{100}$, such that $A < CB$. Similarly, we write $A \ll_d B$ if and only if there exists $c(d) > (100d)^{100d}$ such that $A < c(d)B$.

Throughout the paper, the parameters $\eta = \eta(d)$, $c(d)$ designate positive numbers depending on d while c stands for a large positive number; all these are allowed to change from line to line.

3. **Discretization**

Our goal in this section is to present the discretization of our Polynomial Carleson operator on the torus[23] defined by

$$C_{d,1} f(x) := \sup_{Q \in \mathcal{Q}_{d,1}} \left| \int_{\mathbb{T}} e^{i\,Q(y)} \cot(\pi y)\, f(x-y)\, dy \right|. \tag{43}$$

In what follows, for notational simplicity, we will refer to the operator $C_{d,1}$ as T.

In direct connection with the discussion in Section 1.4, we start by emphasizing the group symmetries of T as displayed in the following relation:

(44)

$$Tf(x) = \sup_{a_1, \ldots, a_d \in \mathbb{R}} |M_{1,a_1} \ldots, M_{d,a_d} H M^*_{1,a_1} \ldots, M^*_{d,a_d} f(x)| = \sup_{Q \in \mathcal{Q}_d} |T_Q f(x)|,$$

[22]There should be no confusion as the precise meaning should be clear from the context.

[23]Both for continuity of historical lineage (see [15] and [36]) as well as for clarity of exposition we present our proof on the torus as opposed to the real line. However, the latter situation follows similarly with no significant modifications.

where $\{M_{j,a_j}\}_{j\in\{1,\ldots,d\}}$ is the family of generalized modulations, defined in (13), H is the periodic Hilbert transform, and[24]

$$T_Q f(x) := \int_{\mathbb{T}} e^{i\,(Q(x)-Q(x-y))}\,\cot(\pi y)\,f(x-y)\,dy, \tag{45}$$

with $Q \in \mathcal{Q}_d$ given by $Q(y) = \sum_{j=1}^{d} a_j\,y^j$.

Next, up to easily controlled smooth error terms, (45) can be written in an equivalent form as[25]

$$T_Q f(x) \approx \int_{\mathbb{T}} e^{i\,(\int_y^x q)}\,\frac{1}{x-y}\,f(y)\,dy, \tag{46}$$

where we recall that in the above q stands for the derivative of Q.

Now linearizing the supremum in T, we write

$$Tf(x) = T_{Q_x} f(x) = \int_{\mathbb{T}} e^{i\,(\int_y^x q_x)}\,\frac{1}{x-y}\,f(y)\,dy, \tag{47}$$

where $Q_x(y) := \sum_{j=1}^{d} a_j(x)\,y^j$ with $\{a_j(\cdot)\}_{j\in\{1,\ldots d\}}$ measurable functions and q_x is the derivative of Q_x, that is, $q_x(t) = \frac{d}{dt}Q_x(t)$ with $\int_y^x q_x = \int_y^x q_x(t)\,dt$.

Further, proceeding as in [36] and [81], we define ψ to be an odd C^∞ function such that

$$\text{supp }\psi \subseteq \{y \in \mathbb{R} \mid 2 < |y| < 8\} \tag{48}$$

and

$$\frac{1}{y} = \sum_{k\geq 0} \psi_k(y) \qquad \forall\ 0 < |y| < 1,$$

where by definition $\psi_k(y) := 2^k\psi(2^k y)$ with $k \in \mathbb{N}$.

Using this, we deduce that

$$Tf(x) = \sum_{k\geq 0} T_k f(x) := \sum_{k\geq 0} \int_{\mathbb{T}} e^{i\,(\int_y^x q_x)}\,\psi_k(x-y)\,f(y)\,dy. \tag{49}$$

Now recalling (36), for each $P = [\vec{\alpha}, I] \in \mathbb{P}$, we set

$$E(P) := \{x \in I \mid q_x \in P\}\,. \tag{50}$$

Also, if $|I| = 2^{-k}$ $(k \geq 0)$, we define the operators T_P on $L^2(\mathbb{T})$ by

$$T_P f(x) := \left\{\int_{\mathbb{T}} e^{i\,(\int_y^x q_x)}\,\psi_k(x-y)\,f(y)\,dy\right\}\chi_{E(P)}(x). \tag{51}$$

[24]Given the existence of the supremum in the identity (44), we can replace the phase function $Q(y)$ appearing in (43) by $Q(x)-Q(x-y)$. The convenience for this change becomes transparent in (46) and extends throughout our later reasonings.

[25]Throughout the paper, we ignore possible absolute constants multiplying the kernel of our operators.

Notice that if $\mathbb{P}_{(k)} := \left\{P = [\vec{\alpha}, I] \in \mathbb{P} \mid |I| = 2^{-k}\right\}$, then for fixed $k \in \mathbb{N}$, the set represented by

$$\{E(P)\}_{P \in \mathbb{P}_{(k)}}$$

forms a partition of $[0,1)$, and so

$$T_k f(x) = \sum_{P \in \mathbb{P}_{(k)}} T_P f(x).$$

Consequently, we have the *exact* discretization[26]

$$Tf(x) = \sum_{k \geq 0} T_k f(x) = \sum_{P \in \mathbb{P}} T_P f(x). \tag{52}$$

This ends our decomposition.

We finish this section with the following

Observation 2. Here we record two facts that will be very useful in our later reasonings:

- For a tile $P = [\vec{\alpha}, I_P]$, based on (39), (48) and (51), we deduce that

$$\text{supp } T_P \subseteq I_P \quad \text{and} \quad \text{supp } T_P^* \subseteq I_{P^*}, \tag{53}$$

  where here T_P^* denotes the adjoint of T_P.
- Taking D to be the smallest integer larger than $100d \log_2(100d)$ and splitting

$$\mathbb{P} = \bigcup_{j=0}^{D-1} \bigcup_{k \geq 0} \mathbb{P}_{(kD+j)},$$

  we can assume from now on—after conveniently identifying $\mathbb{P}$ with one of $\mathbb{P}_{(kD+j)}$—that the following *scale separation condition* holds:

(54) if $P_j = [\vec{\alpha}_j, I_j] \in \mathbb{P}$ with $j \in \{1, 2\}$ such that $|I_1| \neq |I_2|$, then either $|I_1| \leq 2^{-D} |I_2|$ or $|I_2| \leq 2^{-D} |I_1|$.

4. Some preliminary elements about tiles

A key task in this paper will be to prove the L^2-boundedness of T, which, via standard reasonings, amounts to understanding the operator norm $\|T^*\|_{L^2 \mapsto L^2}$.[27] Appealing to (52), we notice that

$$\|T^* f\|_2^2 = \sum_{P_1, P_2 \in \mathbb{P}} \left\langle T_{P_1}^* f, T_{P_2}^* f \right\rangle. \tag{55}$$

[26]Without loss of generality, throughout this paper we assume that $\mathbb{P}$ is finite.

[27]As often encountered in analysis, taking the adjoint of an integral operator offers sometimes the advantage of working with a smoother kernel—this being precisely the motivation behind using here the identity $\|T\|_{L^2 \mapsto L^2} = \|T^*\|_{L^2 \mapsto L^2}$.

In this way, we are naturally led to considering the interactions

$$\left\langle T^*_{P_1} f, T^*_{P_2} f \right\rangle, \tag{56}$$

thus motivating our present section.

Our main purpose will be to show that the operator discretization in Section 3—which is fundamentally based on the *relational perspective* introduced in [81]—is designed such that the interaction in (56) is controlled by a suitable defined density of the sets $E(P_1)$ and $E(P_2)$, and by the appropriately defined normalized distance between the geometric representations of our tiles $\hat{P}_1$ and $\hat{P}_2$. The concrete realization of this is encapsulated in the main result of this section given by Lemma 1 below.

To that end, we will first need to introduce some quantitative concepts that are adapted to the information offered by the localization of $\{T_{P_j}\}_{j\in\{1,2\}}$.

4.1. *Properties of T_P and T^*_P.* In this section we very briefly record the time-frequency localization properties of our elementary building blocks, which should be regarded as a weighted[28] generalized wave-packet decomposition of our operator T.

For $P=[\vec{\alpha}, I]\in\mathbb{P}$ with $|I|=2^{-k}$, $k\in\mathbb{N}$, we have

$$\begin{aligned} T_P f(x) &= \left(\int_{\mathbb{T}} e^{i\,(\int_y^x q_x)}\,\psi_k(x-y)\,f(y)\,dy\right)\chi_{E(P)}(x),\\ T^*_P f(x) &= \int_{\mathbb{T}} e^{-i\,(\int_x^y q_y)}\,\psi_k(y-x)\,\left(\chi_{E(P)} f\right)(y)\,dy. \end{aligned} \tag{57}$$

As will be soon revealed, as a consequence of Lemma 1 below,[29] we have that[30]

(58)
- the time-frequency localization of T_P is "morally" given by the geometric representation $\hat{P}$;
- the time-frequency localization of T^*_P is "morally" given by the geometric representation $\widehat{P^*}$.

4.2. *Factors of a tile.* In this section we introduce two important concepts that will impact our understanding of the interaction in (56).

For a tile $P=[\vec{\alpha}, I]$, we define two quantities:

[28]Relative to the information carried by the sets $E(P)$, $P\in\mathbb{P}$.

[29]This is the essence of the relational perspective introduced in [81], namely to understand the time-frequency localization of a function/operator in $L^2(\mathbb{R})$—in our context $T^*_P f$—depending on how it interacts via the scalar product with similar nature exterior objects, i.e., by analyzing the size of the interaction $\langle T^*_{P_1} f, T^*_{P_2} f\rangle$. This contrasts with (while in the standard situations recovers) the classical approach that determines the time-frequency portrait of $T^*_P f$ by simply studying the L^∞-size distribution of $T^*_P f$ and $\widehat{T^*_P f}$, respectively.

[30]At this point it is worth recalling the geometric interpretation of $\hat{P}$ as given by Observation 1.

(a) An *absolute* one of *analytic* nature.[31]

We define the *density factor of* P to be the expression

$$A_0(P) := \frac{|E(P)|}{|I|}. \tag{59}$$

This definition is motivated by the fact that $A_0(P)$ determines the L^2 operator norm of T_P. Indeed, one immediately has that

$$\|T_P f\|_2 \leq |A_0(P)|^{\frac{1}{2}} \, \|f\|_2, \tag{60}$$

and moreover, that $\|T_P\|_{2\mapsto 2} \approx |A_0(P)|^{\frac{1}{2}}$.

(b) A *relative* one of *geometric* nature, as follows.

Suppose first that we are given $q \in \mathcal{Q}_{d-1}$ and J an interval (not necessarily dyadic); we introduce the quantity

$$\Delta_q(J) := \frac{\text{dist}^J(q,0)}{|J|^{-1}}, \tag{61}$$

where, for $q_1, q_2 \in \mathcal{Q}_{d-1}$, we use the notation

$$\text{dist}^A(q_1,q_2) := \sup_{y\in A}\{\text{dist}_y(q_1,q_2)\} \quad \text{and} \quad \text{dist}_y(q_1,q_2) := |q_1(y) - q_2(y)|\,.$$

Observe that we have the following *monotonicity property*:

$$J_1 \subseteq J_2 \ \text{ implies } \ \Delta_q(J_1) \leq \Delta_q(J_2). \tag{62}$$

Now we define the *geometric factor of* P *with respect to* q as[32]

$$\lceil \Delta_q(P) \rceil\,, \tag{63}$$

where

$$\Delta_q(P) := \inf_{q_1\in P} \Delta_{q-q_1}(I_P). \tag{64}$$

The motivation for this last concept arises from the geometric control—obtained via a Van der Corput type lemma—offered by

$$|K_{q,P}(x)| \lesssim_d \lceil \Delta_q(P) \rceil^{\frac{1}{d}}\,,$$

where here $K_{q,P}(x) := \left(\int e^{i\,[\int_y^x q_x - \int_y^x q]}\,\psi_k(x-y)\,dy\right)\chi_{E(P)}(x)$ with $|I_P| = 2^{-k}$, is a fundamental expression directly linked[33] with the study of (56).

[31] This is inspired by the mass quantity introduced by C. Fefferman in [36].

[32] Recall that given $x \in \mathbb{R}$, we let $\lceil x \rceil := \frac{1}{1+|x|}$.

[33] For the explicit relationship, see the proof of Lemma 6 and (274) therein.

4.3. *Normalized distance between two tiles.* Our aim in this subsection is to properly quantify within the time-frequency plane the relative distance between two given tiles P_1^* and P_2^*. This will be encoded into a concept that naturally extends the geometric factor introduced in (63) above and whose motivation stays on similar grounds.

In what follows, we will only consider the non-trivial case $I_{P_1}^* \cap I_{P_2}^* \neq \emptyset$; also, throughout this section, for notational simplicity, we set $I_{P_1} =: I_1$, $I_{P_2} =: I_2$.

Definition 1 (*Geometric factor associated to a pair of tiles*). Given two tiles P_1 and P_2, we define the *geometric factor of the pair* (P_1, P_2) by

$$\lceil \Delta(P_1, P_2) \rceil ,$$

where[34]

$$\Delta(P_1, P_2) := \frac{\inf_{\substack{q_1 \in P_1 \\ q_2 \in P_2}} \left(\sup_{y \in \tilde{I}_1 \cap \tilde{I}_2} \operatorname{dist}_y(q_1, q_2) \right)}{|\tilde{I}_1 \cap \tilde{I}_2|^{-1}} .$$

Definition 2 (*Interaction polynomial*). For P_1 and P_2 as above, we let the $(P_1,\, P_2)$-*interaction polynomial* be

$$q_{1,2} := q_{P_1} - q_{P_2} . \tag{65}$$

Observation 3. With this notation, using the results in the appendix and assuming without loss of generality that $|I_1| \geq |I_2|$, we have that

$$\lceil \Delta(P_1, P_2) \rceil \approx_d \max \left\{ \left\lceil \Delta_{q_{P_1}}(P_2) \right\rceil ,\ \left\lceil \Delta_{q_{P_2}}(P_1) \right\rceil \right\} \approx_d \lceil \Delta_{q_{1,2}}(I_2) \rceil .$$

Once at this point we are ready to state the main result of Section 4.

4.4. *Control over the interaction between two tiles.* With the previous notation and definitions, we have

LEMMA 1 (Tile interaction control: preliminary version). *Let* $P_1,\, P_2 \in \mathbb{P}$. *Then, we have*

$$|\left\langle T_{P_1}^* \, f, T_{P_2}^* \, g \right\rangle| \lesssim_d \lceil \Delta(P_1, P_2) \rceil^{\frac{1}{d}} \, \frac{\int_{E(P_1)} |f| \int_{E(P_2)} |g|}{\max\left(|I_1|, |I_2|\right)} \tag{66}$$

or, equivalently,[35]

$$\left\| T_{P_1} T_{P_2}^* \right\|_{2 \mapsto 2}^2 \lesssim_d \min \left\{ \frac{|I_2|}{|I_1|}, \frac{|I_1|}{|I_2|} \right\} \lceil \Delta(P_1, P_2) \rceil^{\frac{2}{d}} \, A_0(P_1) \, A_0(P_2). \tag{67}$$

We will postpone the proof of this lemma for Section 9.2. In fact there we will prove a more refined version of Lemma 1, namely, Lemma 6.

[34]Recall here the following notation: $\tilde{I}_j = \tilde{I}_{P_j} := 17 I_{P_j}$, where $j \in \{1,\, 2\}$.

[35]Notice the central role played by the newly introduced concepts of density and geometric factors.

5. **Philosophy and outline of the proof: a paradigm shift—from global to local analysis.**

This section explains the general philosophy behind the proof of our Main Theorem.

5.1. *Reduction to the Main Proposition*: *informal*. Our main goal in this paper is to prove the boundedness of the expression

$$\|T\|_{L^p\to L^p} = \|\sum_{P\in\mathbb{P}} T_P\|_{L^p\to L^p} = \|\sum_{P\in\mathbb{P}} T_P^*\|_{L^{p'}\to L^{p'}} = \|T^*\|_{L^{p'}\to L^{p'}} \tag{68}$$

for any $1<p<\infty$, where here we used (52).

For expository reasons, throughout most of this section we set $p=2$. As a first step in evaluating $\|T^*\|_{L^2\to L^2}$ it becomes natural to group the terms $\langle T_{P_1}^* f, T_{P_2}^* f\rangle$ according to their magnitude. By inspecting (66) and (67) we deduce that this magnitude depends on two types of parameters:

- the density factors $A_0(P_1)$ and $A_0(P_2)$, which, as mentioned in Section 4.2, are of an absolute nature reflecting strictly the magnitude of the associated operator norms $\{\|T_{P_j}^*\|_{L^2\mapsto L^2}\}_{j\in\{1,2\}}$; and
- the geometric factor $\lceil\Delta(P_1,P_2)\rceil$, which quantifies the relative geometric position of the tiles P_1 and P_2.

Accordingly, we are invited to decompose the family

$$\mathbb{P} = \bigcup_n \mathbb{P}_n = \bigcup_{n,r} \mathcal{P}_n^r \tag{69}$$

such that, *informally*, one has that each $\mathbb{P}_n$

(70)
- consists of tiles with uniform density factor $\approx 2^{-n}$;[36]
- can be written as $\bigcup_r \mathcal{P}_n^r$ where, for each $\mathcal{P}_n^r$, there exists $q_{n,r}\in\mathcal{Q}_{d-1}$ such that $\lceil\Delta_{q_{n,r}}(P)\rceil\approx 1$ for all $P\in\mathcal{P}_n^r$. Notice that for any $P_1,\,P_2\in\mathcal{P}_n^r$ with $I_{P_1}\cap I_{P_2}\neq\emptyset$, we have $\lceil\Delta(P_1,P_2)\rceil\approx 1$.[37]

Assuming now that we have been able to rigorously shape both (69) and (70), our goal will be to prove the following:

Main Proposition. *Fix* $n\in\mathbb{N}$. *Then there exists a constant* $\eta=\eta(d)\in(0,\frac{1}{2})$ *depending only on* d *such that*

$$\left\|T^{\mathbb{P}_n} f\right\|_p \lesssim_{p,d} 2^{-n\,\eta(1-\frac{1}{p^*})}\,\|f\|_p \tag{71}$$

for all $f\in L^p(\mathbb{T})$ *and* $1<p<\infty$.

[37]This accounts for the absolute nature of $A_0(P)$, which does not interact with the environment exterior to P.

[38]Each such family $\mathcal{P}_n^r$ is essentially a tree—a tile-geometric structure that is the elementary building block in any modulation-invariant problem.

If we believe this for the moment, then our Main Theorem immediately follows. Indeed, applying (52), (69), (71) and the triangle inequality, we have

$$\|Tf\|_p \leq \sum_n \left\|T^{\mathbb{P}_n} f\right\|_p \lesssim_{p,d} \sum_n 2^{-n\,\eta(1-\frac{1}{p^*})} \, \|f\|_p \lesssim_{p,d} \, \|f\|_p \, .$$

However, it turns out that *rigorously* realizing the decomposition (69) is quite subtle. In particular, each of the desired properties in (70) presents its own challenges:

(72)
- First, the concept of the density factor of a tile is too rough. Indeed, there is no connection between two tiles being geometrically "close" (i.e., $\lceil\Delta(P_1, P_2)\rceil \approx 1$) and the relative magnitude of their density factors. This affects the geometry of the trees $\mathcal{P}_n^r \subseteq \mathbb{P}_n$ since we cannot guarantee the convexity[39] of each $\mathcal{P}_n^r$, which turns out to be essential. The solution to this is the introduction of a smoother, weighted density factor called the *mass* of a tile.
- Second, we have no a priori control on the structure of $\mathbb{P}_n$, that is, on how many trees $\{\mathcal{P}_n^r\}_r$ can sit above a given spatial location. This latter aspect is encoded in the behavior of the so-called *counting function* $\mathcal{N}_{\mathbb{P}_n}$ whose growth control is directly related with a good bound on $\left\|T^{\mathbb{P}_n} f\right\|_p$ as desired in (71).

Thus, in order to speak meaningfully about the partition (69) and about the ideas behind the proof of the Main Proposition, we need to take a detour and elaborate more on the concepts of mass of a tile and the counting function.

5.2. *Establishing a perspective*: *from global to local properties*. In this section we take a comparative look at how the concerns raised in (72) have been addressed over time. As we will see, this brief account is suggestive of how the evolution over the concepts of mass and counting function inform both the features of the decomposition (69) and the treatment of $T^{\mathbb{P}_n} f$.

5.2.1. *Historical context*. We start our analysis with Fefferman's approach, in which our present methods are rooted.[40] In [36], Fefferman smooths out the term $A_0(P)$ by introducing the mass $A(P)$ as follows: if $P = [\alpha, I] \in \mathbb{P}$, then[41]

$$A(P) := \sup_{\substack{P'=[\alpha',I'] \in \mathbb{P} \\ I \subseteq I'}} \frac{|E(P')|}{|I'|} \, \left\lceil \Delta(P, \, P') \right\rceil^N . \tag{73}$$

[39] For the explicit definition of convexity, please see the third item in Definition 8.

[40] Throughout this discussion we use the terminology introduced in this paper and consider only the linear-phase case, that is, $d = 1$.

[41] Here $N \in \mathbb{N}$ is some fixed large natural number.

This definition addresses the first challenge of (72); indeed, one can now see the desired smoothing effect: if $P_1 \leq P_2$, then $A(P_1) \geq A(P_2)$.

With this problem settled, one can now go ahead and partition

$$\mathbb{P} = \bigcup_n \mathbb{P}_n \quad \text{with} \quad \mathbb{P}_n := \{P \in \mathbb{P} \,|\, A(P) \in (2^{-n}, 2^{-n+1}]\}. \tag{74}$$

The crux of Fefferman's argument in estimating the quantity $\|T^{\mathbb{P}_n}\|_{L^2[0,1]}$ (see Main Lemma, Section 6 in [36]) relies on the following implicit underlying estimate[42]

$$\|T^{\mathbb{P}_n} f\|_{L^2([0,1])} \lesssim \log(1 + \|\mathcal{N}_{\mathbb{P}_n}\|_{L^\infty([0,1])}) \left(\sup_{P\in\mathbb{P}_n} A(P)\right)^{\frac{1}{2}} \|f\|_{L^2([0,1])}, \tag{75}$$

where here $\mathcal{N}_{\mathbb{P}_n} := \sum_{P\in\mathbb{P}_n^{\max}} \chi_{I_P}(x)$ stands for the counting function associated to the maximal tiles[43] in $\mathbb{P}_n$ with the density factor $\geq 2^{-n}$. Notice that (75) clarifies precisely the meaning of the second item in (72). In order to exploit (75) we have to control $\|\mathcal{N}_{\mathbb{P}_n}\|_{L^\infty}$ by excising the "exceptional" set A_n on which $\mathcal{N}_{\mathbb{P}_n}$ is large. Thus, instead of (71) (for $p = 2$), the corresponding estimate in [36] reads

$$\left\|T^{\mathbb{P}_n} f\right\|_{L^2([0,1]\setminus A_n)} \lesssim (\log K)\, n\, 2^{-\frac{n}{2}} \; \|f\|_{L^2[0,1]}\,, \tag{76}$$

where here for an arbitrary $K > 1$ one sets $A_n := \{x \in [0,1] \,|\, \mathcal{N}_{\mathbb{P}_n}(x) > K\, 2^{2n}\}$ and uses the trivial bound $\|\mathcal{N}_{\mathbb{P}_n}\|_{L^1[0,1]} \lesssim 2^n$ in order to obtain the smallness property $|A_n| \lesssim \frac{1}{K}\, 2^{-n}$. From this, taking $E := \bigcup_n A_n$, we deduce $|E| \lesssim \frac{1}{K}$, while from (76), the triangle inequality, and Chebyshev's inequality, for any $\beta > 0$, we conclude

$$|\{x \in \mathbb{T} \,|\, |Tf(x)| > \beta\}| \lesssim |E| + \frac{(\log K)^2}{\beta^2} \; \|f\|^2_{L^2(\mathbb{T})}\,. \tag{77}$$

Finally, a variational argument relative to K gives $\|T\|_{L^2\to L^1} < \infty$, which together with an application of Stein's maximal principle ([115]) implies the weak-L^2-boundedness of T. The other known approaches to the Carleson operator ([15] and [73]) similarly require the removal of exceptional sets, and thus they are also limited to only establishing L^2 to $L^{2,\infty}$ bounds for T.

We end by saying that the same types of difficulties (and more) are encountered in establishing general L^p-boundedness for $1 < p < \infty$. Indeed, all the previous treatments of the Carleson operator had to account for the presence and then, at the expense of extra technicalities, the removal of such exceptional sets in order to conclude that T maps L^p into $L^{p,\infty}$. Accordingly, it is natural to ask whether one can establish the strong-type boundedness

[42]For more on this estimate, the reader is invited to consult the fourth item in Section 11.

[43]For the unfamiliar reader, please see Definitions 4 and 5 in Section 7.

of T from L^p to L^p directly without appealing to weak-type estimates as an intermediary and with minimal recourse to interpolation arguments.

Motivated by the discussion above, in what follows we introduce a methodology whose aim is to

- provide sharp estimates adapted to spatial locations;
- eliminate the usage of exceptional sets; and
- provide in a unified manner the full strong-type boundedness L^p range, $1 < p < \infty$, for the (Polynomial) Carleson operator.

5.2.2. *Local analysis*: *intuition.* The above contextualization sets the scene for the new perspective introduced in the present paper, which develops a local analysis tailored to the two key concepts of mass and counting function. The story below unravels the intuitive trajectory followed by the author in building the proof of the Main Proposition and should be viewed as a brief and heuristic preview of the elements that will be introduced with successively increasing degrees of rigor within Sections 5.3, 5.4, and 6.2.

We begin by noticing that the approach reflected by (75) has, via the mass and the counting function defined in [36], a *global character*:

- $A(P)$ picks the largest density factor of P' weighted relative to P as P' runs through the set of all tiles at smaller scales $|I'|$ obeying the natural condition $I' \supseteq I$. Consequently, (73) cannot offer any precise information about the spatial location that contributes the most to the magnitude of $A(P)$.
- $\mathcal{N}_{\mathbb{P}_n}$ collects the data from all the maximal tiles within $\mathbb{P}_n$, thus again offering no information about where in space the magnitude of $\mathcal{N}_{\mathbb{P}_n}$ becomes suitably large. Such information, however, is a key element in the hope of circumventing the exceptional set analysis.

In this light, we adopt an approach of child-like scrutiny: "What if in (75), instead of ending our reasoning by merely discarding the set A_n where $\mathcal{N}_{\mathbb{P}_n}$ is large, we were to continue our analysis by (1) first examining the structure of the family $\mathbb{P}_n(A_n)$ of tiles that have their spatial intervals within A_n, then (2) properly designing a local version of (75) adapted to the location A_n, and then (3) iterating this process until we exhausted the whole family of tiles $\mathbb{P}_n$?" Heuristically, this approach accounts for a stopping time algorithm that at each level $l \in \mathbb{N}$ takes as its input a union $A_n^l \subseteq \mathbb{T}$ of disjoint intervals and a family $\mathbb{P}_n(A_n^l)$ of tiles with spatial projection within A_n^l, and outputs a new "exceptional" set $A_n^{l+1} \subset A_n^l$ together with the corresponding collection of tiles $\mathbb{P}_n(A_n^{l+1})$ and an updated $\mathbb{P}_n(A_n^l) := \mathbb{P}_n^{\text{old}}(A_n^l) \setminus \mathbb{P}_n(A_n^{l+1})$. The hope is that this iterative process will yield a collection of localized operators $\{T^{\mathbb{P}_n(A_n^l)}\}_l$ such that (1) their sum recovers the full operator $T^{\mathbb{P}_n}$; (2) each piece $T^{\mathbb{P}_n(A_n^l)}$ satisfies a "local version" of (75) incorporating a corresponding local counting function whose L^∞ norm is suitably controlled; and (3) the size of $\|T^{\mathbb{P}_n(A_n^l)}\|_{L^2(A_n^{l+1})}$ is "much smaller" than $\|T^{\mathbb{P}_n(A_n^l)}\|_{L^2(A_n^l)}$, thus indicating that $\{T^{\mathbb{P}_n(A_n^l)}\}_l$ forms a family of almost orthogonal operators.

In order to advance toward formalizing this approach, it becomes natural to search for ways of expressing the *local* behavior of the counting function $\mathcal{N}_{\mathbb{P}_n}$ and the mass $A(P)$. Turning first towards $\mathcal{N}_{\mathbb{P}_n}$, we decrypt its local properties from the structure of its level sets. A simple but important insight is that

$$|\{x\,|\,\mathcal{N}_{\mathbb{P}_n}(x) > (K+L)\,2^n\}| \leq \frac{1}{2}\,|\{x\,|\,\mathcal{N}_{\mathbb{P}_n}(x) > K\,2^n\}| \tag{78}$$

for any $K,\, L \geq 10$, which, via the John–Nirenberg inequality, can be seen as a direct manifestation of the fact that $\mathcal{N}_{\mathbb{P}_n}$ belongs to dyadic BMO with $\|\mathcal{N}_{\mathbb{P}_n}\|_{\mathrm{BMO}_D} \leq 2^n$. Moving a step further in this direction, one is thus invited to run a stopping time algorithm adapted to the level sets of $\mathcal{N}_{\mathbb{P}_n}$, which produces the collection of sets $\{\mathcal{I}_n^l\}_l$ with $\mathcal{I}_n^l$ the maximal interval representation of $A_n^l = \{x\,|\,\mathcal{N}_{\mathbb{P}_n}(x) > l\,n\,2^n\}$. (This is an informal and heuristic description of the creation of the intervals in (80), whose rigorous construction is more involved and will be introduced in Section 6.2 below; see also Observations 4 and 9.) From (78) one infers that $\{\mathcal{I}_n^l\}_l$ forms a *sparse* collection of intervals (this corresponds to relation (80b))—which will be crucial in achieving the almost orthogonality mentioned above.

The intervals $\mathfrak{I} \in \mathcal{I}_n^l$ will serve as the relevant spatial locations in our local analysis. Once at this point, we introduce the concept of mass adapted to $\mathfrak{I}$, denoted $A_{\mathfrak{I}}(P)$, and then design the family of tiles $\mathbb{P}_n(\mathfrak{I})$ such that (1) for each tile $P \in \mathbb{P}_n(\mathfrak{I})$, we have $A_{\mathfrak{I}}(P) \approx 2^{-n}$; (2) we achieve an effective L^∞ control over the corresponding local counting function $\mathcal{N}_{\mathbb{P}_n(\mathfrak{I})}$; and (3) we end up with a partition of the whole family of tiles $\mathbb{P} = \bigcup_n \mathbb{P}_n = \bigcup_n \bigcup_{l,\,\mathfrak{I}\in\mathcal{I}_n^l} \mathbb{P}_n(\mathfrak{I})$. (These itemized features are encapsulated in (82)–(84) below.) Within this framework we obtain the following localized version of (75):

$$\begin{aligned}\|T^{\mathbb{P}_n(\mathfrak{I})} f\|_{L^2(\mathfrak{I})} &\lesssim \log(1+\|\mathcal{N}_{\mathbb{P}_n(\mathfrak{I})}\|_{L^\infty}) \left(\sup_{P\in\mathbb{P}_n(\mathfrak{I})} A(P)\right)^{\frac{1}{2}} \|f\|_{L^2(\tilde{\mathfrak{I}})} \\ &\lesssim n\,2^{-\frac{n}{2}}\,\|f\|_{L^2(\tilde{\mathfrak{I}})},\end{aligned} \tag{79}$$

which provides the desired control over the components $\{T^{\mathbb{P}_n(\mathfrak{I})}\}_{\mathfrak{I}}$. This is the point that marks the trade-off between the process of localization and the excision of the exceptional sets.

Finally, using the aforementioned sparsity of the collection $\{\mathcal{I}_n^l\}_l$ together with the local control obtained in (79), one can show the alluded almost orthogonality of the pieces $\{T^{\mathbb{P}_n(A_n^l)}\}_l$, where $T^{\mathbb{P}_n(A_n^l)} := \sum_{\mathfrak{I}\in\mathcal{I}_n^l} T^{\mathbb{P}_n(\mathfrak{I})}$. This in turn implies the strong L^2 global estimate for T.

In light of the above, our proof for the case $d=1, p=2$ can be heuristically thought of as a *local* modeling of Fefferman's approach in [36].

5.3. *Polynomial Carleson: key elements in our local analysis.* With the high-level preview of our approach completed in the previous section, we increase our level of rigor and introduce in a more formal manner the five key ingredients on which the proof of our Main Proposition will be based:

- The collections of stopping-time intervals $\{\mathcal{I}_n^l\}_{l,n\in\mathbb{N}}$ obeying
  - for each $l,\, n \in \mathbb{N}$ the family $\mathcal{I}_n^l$ consists of (maximal) disjoint dyadic intervals such that for any $l_0,\, n_0 \in \mathbb{N}$ one has

$$\bigcup_{n<n_0}\bigcup_{\mathfrak{I}\in\mathcal{I}_n^l}\mathfrak{I}\subseteq\bigcup_{\mathfrak{J}\in\mathcal{I}_{n_0}^l}\mathfrak{J}\ \text{ and }\ \bigcup_{l\geq l_0}\bigcup_{\mathfrak{I}\in\mathcal{I}_n^l}\mathfrak{I}\subseteq\bigcup_{\mathfrak{J}\in\mathcal{I}_n^{l_0}}\mathfrak{J};\tag{80a}$$

(80)
  - each interval in $\mathcal{I}_n^{l+1}$ is a child of some interval in $\mathcal{I}_n^l$;
  - the collections $\{\mathcal{I}_n^l\}_{l,n\in\mathbb{N}}$ obey a *Carleson packing condition*, or equivalently, form a *sparse* family of intervals, i.e.,

$$\forall\,\mathfrak{J}\in\bigcup_{l,n\in\mathbb{N}}\mathcal{I}_n^l\ \text{ one has }\ \sum_{\substack{\mathfrak{I}\subsetneq\mathfrak{J}\\ \mathfrak{I}\in\bigcup_{l,n\in\mathbb{N}}\mathcal{I}_n^l}}|\mathfrak{I}|\leq\frac{1}{2}\,|\mathfrak{J}|.\tag{80b}$$

- The concept of *mass adapted to an interval*: for $\mathfrak{J}\subseteq\mathbb{T}$ a dyadic interval and $P=[\vec{\alpha},I]\in\mathbb{P}$ with $I\subseteq\mathfrak{J}$, we set

$$A_{\mathfrak{J}}(P):=\sup_{\substack{P'=[\vec{\alpha}',I']\in\mathbb{P}\\ I\subseteq I'\subseteq\mathfrak{J}}}\frac{|E(P')|}{|I'|}\,\lceil\Delta(10P,\,10P')\rceil^{N},\tag{81}$$

  where here $N\in\mathbb{N}$ is a fixed large natural number.
- The associated families of tiles $\{\mathbb{P}_n(\mathcal{I}_n^l)\}_{l,n\in\mathbb{N}}$ such that
  - by definition $\mathbb{P}_n(\mathcal{I}_n^l):=\bigcup_{\mathfrak{J}\in\mathcal{I}_n^l}\mathbb{P}_n(\mathfrak{J})$ with

$$\mathbb{P}_n(\mathfrak{J}):=\left\{P=[\vec{\alpha}_P,I_P]\in\mathbb{P}\ \middle|\ \begin{array}{c} I_P\subseteq\mathfrak{J},\, I_P\nsubseteq\bigcup_{\mathfrak{I}\in\mathcal{I}_n^{l+1}}\mathfrak{I}\\ A_{\mathfrak{J}}(P)\in(2^{-n},2^{-n+1}]\end{array}\right\};\tag{82}$$

  - the associated *local counting function* is under good control, i.e.,

$$\forall\,x\in\mathbb{T}\quad\mathcal{N}_{\mathbb{P}_n(\mathfrak{J})}(x):=\sum_{P=[\vec{\alpha}_P,I_P]\in\mathbb{P}_n^{\max}(\mathfrak{J})}\chi_{I_P}(x)\lesssim n\,2^n,\tag{83}$$

    where here $\mathbb{P}_n^{\max}(\mathfrak{J})$ stands for the collection of tiles that are maximal within $\mathbb{P}_n(\mathfrak{J})$;
  - the families $\{\mathbb{P}_n(\mathcal{I}_n^l)\}_{l,n\in\mathbb{N}}$ form a *partition* of $\mathbb{P}$ with

$$\mathbb{P}_n:=\bigcup_{l\in\mathbb{N}}\mathbb{P}_n(\mathcal{I}_n^l)=\bigcup_{l\in\mathbb{N}}\bigcup_{\mathfrak{J}\in\mathcal{I}_n^l}\mathbb{P}_n(\mathfrak{J})\quad\text{and}\quad\mathbb{P}=\bigcup_{n\in\mathbb{N}}\mathbb{P}_n.\tag{84}$$

- *Good local control*: for any $l,\, n\in\mathbb{N}$, $\mathfrak{I}\in\mathcal{I}_n^l$ and $1<p<\infty$ the following holds:[44]

$$\operatorname{supp} T^{\mathbb{P}_n(\mathfrak{I})}\subseteq\mathfrak{I}\quad\text{and}\quad\operatorname{supp} T^{\mathbb{P}_n(\mathfrak{I})^*}\subseteq\mathfrak{I}\tag{85}$$

and

$$\|T^{\mathbb{P}_n(\mathfrak{I})}f\|_{L^p(\mathfrak{I})}\lesssim_p\left(\sup_{P\in\mathbb{P}_n(\mathfrak{I})}A_{\mathfrak{I}}(P)\right)^{1-\frac{1}{p^*}}\|f\|_{L^p(\mathfrak{I})}.\tag{86}$$

- The *localized support estimates* (local non-concentration): for any $\mathfrak{T}\in\mathcal{I}_n^r$, $\mathfrak{I}\in\mathcal{I}_n^l$ with $l\le r$ and $1<p<\infty$, we have

$$\|T^{\mathbb{P}_n(\mathfrak{I})}f\|_{L^p(\mathfrak{T})}\lesssim_p\|\mathcal{N}_{\mathbb{P}_n(\mathfrak{I})}\|_{L^\infty(\mathfrak{I})}\left(\frac{|\mathfrak{T}|}{|\mathfrak{I}|}\right)^{\frac{1}{p}}\|f\|_{L^p(\mathfrak{I})},\tag{87}$$

and

$$\|T^{\mathbb{P}_n(\mathfrak{I})^*}f\|_{L^{p'}(\mathfrak{T})}\lesssim_p\|\mathcal{N}_{\mathbb{P}_n(\mathfrak{I})}\|_{L^\infty(\mathfrak{I})}\left(\frac{|\mathfrak{T}|}{|\mathfrak{I}|}\right)^{\frac{1}{p'}}\|f\|_{L^{p'}(\mathfrak{I})}.\tag{88}$$

At this point, based on (52) and (84), we remark that

$$T^{\mathbb{P}_n}f=\sum_{l\in\mathbb{N}}\sum_{\mathfrak{I}\in\mathcal{I}_n^l}T^{\mathbb{P}_n(\mathfrak{I})}f.\tag{89}$$

The above ingredients constitute the key foundational elements for the entire edifice of our paper. In the remaining subsection we will integrate these elements into a simple but powerful framework within which the proof of our Main Proposition opens up.

5.4. *A Carleson Embedding principle for linear operators.* In this subsection we present a general principle that might be of independent interest and that provides a natural framework for the proof of our Main Proposition:

Theorem (A Carleson Embedding principle for operators). *Let S be a linear operator from*[45] *$L^0(\mathbb{T})$ to $L^0(\mathbb{T})$ and $1<p<\infty$ a fixed parameter. Assume that the following hold*:

---

[44]Strictly speaking, in the initial decomposition we only have the tamer relations $\operatorname{supp} T^{\mathbb{P}_n(\mathfrak{I})^*}\subseteq\tilde{\mathfrak{I}}$ and $\|T^{\mathbb{P}_n(\mathfrak{I})}f\|_{L^p(\mathfrak{I})}\lesssim_p 2^{-n\,\eta\,(1-\frac{1}{p^*})}\,\|f\|_{L^p(\mathfrak{I})}$ for some $\eta>0$. However, one can partition each $\mathbb{P}_n(\mathfrak{I})$ into $\lesssim n$ terms of the form $\mathbb{P}_{n,nm}(\mathfrak{I})\cup\mathbb{P}_{n,bd}(\mathfrak{I})$ such that the *normal* component $T^{\mathbb{P}_{n,nm}(\mathfrak{I})}$—later called an ($2^n$-separated) L^∞-forest (see Definition 10)—indeed verifies the sharp desired relations, while the *boundary* component $T^{\mathbb{P}_{n,bd}(\mathfrak{I})}$—referred to as a sparse L^∞-forest (see same definition)—may be regarded as an error term. Thus, for expository reasons, for the remainder of this section we will assume without loss of generality that for every $\mathfrak{I}\in\mathcal{I}_n^l$, $\mathbb{P}_n(\mathfrak{I})=\mathbb{P}_{n,nm}(\mathfrak{I})$.

[45]Here $L^0(\mathbb{T})$ stands for the space of Lebesgue measurable functions acting from $\mathbb{T}$ to C.

- (sparse collection of intervals) *there exists a collection* $\mathcal{I}$ *of* (*dyadic*) *intervals in* $[0,1]$ *obeying the Carleson packing condition*

$$\forall\, \mathfrak{I}\in\mathcal{I} \quad \sum_{\substack{\mathfrak{T}\subsetneq\mathfrak{I}\\ \mathfrak{T}\in\mathcal{I}}} |\mathfrak{T}| \le \frac{1}{2}\,|\mathfrak{I}|; \tag{90}$$

- (discretization subordinated to $\mathcal{I}$) *there exists a collection of linear operators* $\{S_{\mathfrak{I}}\}_{\mathfrak{I}\in\mathcal{I}}$ *for which the following hold*:
  - (good local control) *there exists* $\bar{C}>0$ *such that for any* $\mathfrak{I}\in\mathcal{I}$,

$$\|S_{\mathfrak{I}}f\|_{L^p} \le \bar{C}\,\|f\|_{L^p(\mathfrak{I})}, \tag{91}$$

and

$$\text{supp } S_{\mathfrak{I}}\subset\mathfrak{I} \quad \textit{and} \quad \text{supp } S^*_{\mathfrak{I}}\subset\mathfrak{I}; \tag{92}$$

  - (local non-concentration) *there exists* $\bar{D}>0$ *and* $\mu\in(0,1]$ *such that for any* $\mathfrak{T}$, $\mathfrak{I}\in\mathcal{I}$ *with* $\mathfrak{T}\subseteq\mathfrak{I}$, *the following holds*:

$$\|S_{\mathfrak{I}}f\|_{L^p(\mathfrak{T})} \le \bar{D}\left(\frac{|\mathfrak{T}|}{|\mathfrak{I}|}\right)^{\mu}\|f\|_{L^p(\mathfrak{I})} \tag{93}$$

and

$$\|S^*_{\mathfrak{I}}f\|_{L^{p'}(\mathfrak{T})} \le \bar{D}\left(\frac{|\mathfrak{T}|}{|\mathfrak{I}|}\right)^{\mu}\|f\|_{L^{p'}(\mathfrak{I})}; \tag{94}$$

  - (sparse representation) *the operator* S *may be written as*

$$S=\sum_{\mathfrak{I}\in\mathcal{I}} S_{\mathfrak{I}}. \tag{95}$$

Then

$$\|Sf\|_{L^p} \lesssim_p \frac{\bar{C}}{\mu}\,\log(1+\frac{\bar{D}}{\bar{C}})\,\|f\|_{L^p}. \tag{96}$$

Corollary. (1) *Let* $\mathcal{I}$ *be any given sparse family of* (*dyadic*) *intervals within* $[0,1]$ *and* $1<p<\infty$. *For* $f\in L^p(\mathbb{T})$ *and* $\mathfrak{I}\in\mathcal{I}$, *we set* $S_{\mathfrak{I}}f := \frac{\int_{\mathfrak{I}} f}{|\mathfrak{I}|}\,\chi_{\mathfrak{I}}$ *and simply define* $Sf:=\sum_{\mathfrak{I}\in\mathcal{I}} S_{\mathfrak{I}}f$. *One can easily check that all the hypotheses above are verified, while the conclusion can be seen as a variant of the classical Carleson Embedding theorem—hence the name of our principle.*
(2) *The main model serving as a prototype for the above principle is of course the Polynomial Carleson operator. Indeed, one can easily notice that each of the elements introduced in Section* 5.3 (*representing the quintessence of Sections* 9 *and* 10 *below*) *has a direct correspondent in the conditions imposed in the hypothesis of our principle. Indeed, fixing* $n\in\mathbb{N}$ *and* $1<p<\infty$, *and simply taking*

$$Sf := T^{\mathbb{P}_n} f, \tag{97}$$

we see that (90) *is satisfied for* $\mathcal{I}=\bigcup_l \mathcal{I}_n^l$ (*recall* (80)), *and with* $S_{\mathfrak{I}}(f) := T^{\mathbb{P}_n(\mathfrak{I})}f$, *one further has that* (91) *is a reformulation of* (86) *with* $\bar{C}\approx 2^{-n(1-\frac{1}{p^*})}$, *(93) and (94) are equivalent with (87) and (88) for $\bar{D} \approx n2^n$ and $\mu = \frac{1}{\max\{p,p'\}}$, and (95) is a consequence of (89). We thus conclude that our Main Proposition holds.*

Informal proof. For expository reasons and given our main focus, we will provide an informal proof of our principle in the specific case of the Polynomial Carleson operator, so by assuming that S verifies (97). All of the below are made rigorous in Section 10.2. The adaptation to the general case stated in the above principle is straightforward.

We start with a simple observation: By applying the pigeonhole principle, we can restrict our discussion to the subfamily of sets $\{\mathcal{I}_n^{\frac{10}{\mu}ln}\}_l$ and thus, for notational convenience, re-denoting $\mathcal{I}_n^{\frac{10}{\mu}ln}$ by $\mathcal{I}_n^l$ we obtain a specialized version of (80b) that reads

$$\forall\, l \in \mathbb{N} \text{ and } \mathfrak{I} \in \mathcal{I}_n^l, \quad \text{one has} \sum_{\substack{\mathfrak{T}\subset\mathfrak{I}\\ \mathfrak{T}\in\mathcal{I}_n^{l+1}}} |\mathfrak{T}| \leq \frac{1}{2^{\frac{10}{\mu}n}}\,|\mathfrak{I}|. \tag{98}$$

Assume now for simplicity that $p = 2$. Then we claim that

$$\{T^{\mathbb{P}_n(\mathfrak{I})}\}_{\mathfrak{I}\in\bigcup_{l\in\mathbb{N}}\mathcal{I}_n^l} \text{ forms a family of almost orthogonal operators.} \tag{99}$$

Indeed, (99) is a direct consequence of

- the local estimates (86) (for $p = 2$), which, put more simply, read[46]

$$\forall\, l,\, n \in \mathbb{N},\, \mathfrak{I} \in \mathcal{I}_n^l, \qquad \|T^{\mathbb{P}_n(\mathfrak{I})} f\|_{L^2} \lesssim 2^{-\frac{n}{2}}\,\|f\|_{L^2}; \tag{100}$$

- the local non-concentration conditions (87) and (88), which via the sparsity of the collection $\{\mathcal{I}_n^l\}_{l\in\mathbb{N}}$ encoded in (98) and an application of Cauchy-Schwarz inequality, reads as follows: for any $\mathfrak{T} \in \mathcal{I}_n^r$, $\mathfrak{I} \in \mathcal{I}_n^l$ with $l \leq r$, either $\mathfrak{T} \cap \mathfrak{I} = \emptyset$ or

$$\begin{aligned}\left|\left\langle T^{\mathbb{P}_n(\mathfrak{I})} f,\, T^{\mathbb{P}_n(\mathfrak{T})} g\right\rangle\right| &\lesssim_d \left(\frac{|\mathfrak{T}|}{|\mathfrak{I}|}\right)^{\frac{1}{2}} \|f\|_{L^2(\mathfrak{I})}\,\|g\|_{L^2(\mathfrak{T})}\\ &\lesssim 2^{-(r-l)\,n}\,\|f\|_{L^2(\mathfrak{I})}\,\|g\|_{L^2(\mathfrak{T})}\end{aligned} \tag{101}$$

and

$$\begin{aligned}\left|\left\langle T^{\mathbb{P}_n(\mathfrak{I})^*} f,\, T^{\mathbb{P}_n(\mathfrak{T})^*} g\right\rangle\right| &\lesssim_d \left(\frac{|\mathfrak{T}|}{|\mathfrak{I}|}\right)^{\frac{1}{2}} \|f\|_{L^2(\mathfrak{I})}\,\|g\|_{L^2(\mathfrak{T})}\\ &\lesssim 2^{-(r-l)\,n}\,\|f\|_{L^2(\mathfrak{I})}\,\|g\|_{L^2(\mathfrak{T})}.\end{aligned} \tag{102}$$

[46]Notice that due to (92) for $S_{\mathfrak{I}} = T^{\mathbb{P}_n(\mathfrak{I})}$, one can drop the restriction over the region of integration on both sides of the inequality in (86).

In effect, one can circumvent the almost-orthogonality argument from above and use an approach that works directly for general p. Indeed, we apply the following heuristic: (87) asserts the smallness of $\|T^{\mathbb{P}_n(\mathfrak{I})}f\|_{L^p(\mathfrak{T})}$ subject to the smallness of the output spatial support $\mathfrak{T}$ relative to $\mathfrak{I}$, while (88) asserts the same principle for the adjoint operator $T^{\mathbb{P}_n(\mathfrak{I})^*}$. Since the output support of the adjoint operator is equivalent with the input support of the original operator, combining (87) and (88) and defining for each $\mathfrak{I} \in \mathcal{I}_n^l$ the set $\mathfrak{I}_0 := \mathfrak{I} \setminus \bigcup_{\substack{\mathfrak{T}\in\mathcal{I}_n^{l+1}\\ \mathfrak{T}\subsetneq\mathfrak{I}}} \mathfrak{T}$, we deduce that for any $l \in \mathbb{N}$ and $\mathfrak{I} \in \mathcal{I}_n^l$, we have

$$\|T^{\mathbb{P}_n(\mathfrak{I})}f\|_{L^p} \approx_p \|\chi_{\mathfrak{I}}\, T^{\mathbb{P}_n(\mathfrak{I})}\chi_{\mathfrak{I}_0}f\|_{L^p} \;+ \text{Error Term} \tag{103}$$

and similarly

$$\|T^{\mathbb{P}_n(\mathfrak{I})^*}f\|_{L^{p'}} \approx_p \|\chi_{\mathfrak{I}}\, T^{\mathbb{P}_n(\mathfrak{I})^*}\chi_{\mathfrak{I}_0}f\|_{L^{p'}} \;+ \text{Error Term}. \tag{104}$$

Now fix $1 < p \leq 2$, and assume without loss of generality that $p' \in 2\mathbb{N}$. (The more general situation follows a similar strategy via some minor but unavoidable technicalities.) Now using $\|T^{\mathbb{P}_n}\|_{p\to p} = \|T^{\mathbb{P}_n{}^*}\|_{p'\to p'}$, we appeal to (84)–(89), (103), and (104) to deduce

$$\begin{aligned}
\|T^{\mathbb{P}_n{}^*}f\|_{p'}^{p'} &= \sum_{l\in\mathbb{N}}\sum_{\mathfrak{I}\in\mathcal{I}_n^l} \|T^{\mathbb{P}_n(\mathfrak{I})^*}\|_{p'}^{p'} \;+ \text{Error Term}\\
&\approx \sum_{l\in\mathbb{N}}\sum_{\mathfrak{I}\in\mathcal{I}_n^l} \|\chi_{\mathfrak{I}}\, T^{\mathbb{P}_n(\mathfrak{I})^*}\chi_{\mathfrak{I}_0}f\|_{p'}^{p'} \;+ \text{Error Term}\\
&\lesssim_{(86)} \sum_{l\in\mathbb{N}}\sum_{\mathfrak{I}\in\mathcal{I}_n^l} \left(\sup_{P\in\mathbb{P}_n(\mathfrak{I})} A_{\mathfrak{I}}(P)\right) \|f\chi_{\mathfrak{I}_0}\|_{p'}^{p'} \lesssim 2^{-n}\,\|f\|_{p'}^{p'}.
\end{aligned} \tag{105}$$

The argument for the case $2 < p < \infty$ is quite similar, the only difference being that instead of $\|T^{\mathbb{P}_n{}^*}f\|_{p'}^{p'}$ one works with $\|T^{\mathbb{P}_n}f\|_p^p$. This finishes the heuristic, conceptual approach to our Main Proposition. □

The grand scheme in approaching our Polynomial Carleson operator introduced in Sections 5.3 and 5.4 is part of what we call a *local analysis* methodology, which together with the Carleson Embedding principle stated above provides a general framework for treating operators within time-frequency area that, loosely speaking, (1) admit a discretization into smaller, localized components that are subordinated to a family of intervals obeying a Carleson packing-type condition, and (2) on each such component one can exert good local control.

As it turns out, this methodology has recently found other closely related forms of expression with various interesting applications; for more on this, please see the first comment in Section 11.

Observation 4 (Road map to our proof).

(1) We start by mentioning that the present Section 5 was added only after the paper was accepted for publication, in order both to provide a better intuition of the philosophy behind our proof and to properly formalize the local (time-frequency) analysis methodology which—without being formally stated in the initial version of our paper—was foundational for our approach.

(2) In order to remain consistent with the original form of our paper, we have preserved its presentation throughout the remaining sections. We stress that aside from some different notation and terminology, the only distinction between the informal proof above and the formal one presented in the remaining sections is that in the latter we regroup the stopping time intervals $\{\mathcal{I}_n^l\}_{l,n\in\mathbb{N}}$ at each level l into stopping time sets $\{A_n^r\}_{r,n\in\mathbb{N}}$, with the latter representing faithfully the geometry of the level sets of some suitable local counting functions.[47] Consequently, all the definitions and reasonings that follow are adapted to sets representing unions of disjoint intervals rather than simply intervals. That being said, here is the road map that lays out the correspondences between the conceptual outline provided above in this section and the rigorous proof contained within the next sections:

(3) The refined mass concept introduced in (81) corresponds to Definition 3.

(4) The algorithm for constructing the analogue of the collections of stopping-time intervals $\{\mathcal{I}_n^l\}_{l,n\in\mathbb{N}}$ and the associated families of tiles $\{\mathbb{P}_n(\mathcal{I}_n^l)\}_{l,n\in\mathbb{N}}$ is presented in Section 6.2. The properties displayed in (80b)–(83) have their analogues introduced—via Definition 7—in Section 6.2, see the description given at Stages $n.1$ and $n.2$. Additionally, (82) corresponds to (150), (83) corresponds to (148), while (84) is equivalent with (151)–(153).

(5) The local estimate (86) essentially establishes L^p control over an L^∞-forest (see Definition 10), which is referred to as the Main Lemma for $p=2$ (see Section 10.1) and as relation (359) for general $1<p<2$ with the obvious correspondent for the remaining case $p>2$ (see Section 10.2.2).

(6) Relation (99) is the content of the proof of Proposition 2 for $p=2$ treated in Section 10.2.1, with (101) and (102) corresponding to (345) and (344), respectively.

(7) The localized support estimates (87) and (88) are a direct realization of Lemmas 3 and 4 via an application of triangle inequality to the canonical row decomposition of an L^∞-forest.

(8) Finally, the reasonings presented in (103)–(105) are rigorously formalized in Section 10.2.2.

[47] For the exact correspondence, the reader is invited to consult Observation 9.

In what follows we will provide the rigorous argumentation of the heuristic proof of our Main Proposition above.

6. The partition of $\mathbb{P}$

6.1. *Preliminaries.* Before describing our tile-partition process, we need to introduce several concepts:

Definition 3 (*Mass of a tile adapted to a given environment*). Let $\mathcal{A}$ be a (finite) union of dyadic intervals in $[0,1]$ and $\mathcal{P}$ be a finite family of tiles. For $P=[\vec{\alpha},I]\in\mathcal{P}$ with $I\subseteq\mathcal{A}$, we define the *mass* of P relative to the set of tiles $\mathcal{P}$ and the set of spatial locations $\mathcal{A}$ as

$$A_{\mathcal{P},\mathcal{A}}(P):=\sup_{\substack{P'=[\vec{\alpha}',I']\in\mathcal{P}\\ I\subseteq I'\subseteq\mathcal{A}}}\frac{|E(P')|}{|I'|}\,\left\lceil\Delta(10P,\,10P')\right\rceil^{N}, \tag{106}$$

where here $N\in\mathbb{N}$ is a fixed large natural number.

Next, we introduce a qualitative concept that characterizes the overlapping relation between tiles.

Definition 4 (*Aiming for "orderings"*). Let $P_j=[\vec{\alpha}_j,I_j]\in\mathbb{P}$ with $j\in\{1,2\}$. We write

- $P_1\leq P_2$ if and only if $I_1\subseteq I_2$ and there exists $q\in P_2$ such that $q\in P_1$;
- $P_1\trianglelefteq P_2$ if and only if $I_1\subseteq I_2$ and for all $q\in P_2$, we have $q\in P_1$.

Additionally, we write $P_1<P_2$ if $P_1\leq P_2$ and $|I_1|<|I_2|$, and analogously for $\lhd$.

Observation 5. Notice that $\leq$ is not a partial order relation while $\trianglelefteq$ is. Also, based on (54) and Lemma C in the appendix, we deduce that $P_1<P_2$ implies $2P_1\lhd 2P_2$.

In the following two definitions we elaborate on the "pseudo-ordering" $\leq$:

Definition 5 (*Maximal/minimal tiles within a given family*).

(1) If $\mathcal{P}\subseteq\mathbb{P}$ is a family of tiles with some prescribed properties, we say that $P\in\mathcal{P}$ is *maximal* (relative to $\mathcal{P}$) if and only if

$$\forall\,P'\in\mathcal{P}\ \text{ s.t. }\ P\leq P',\ \text{ we have }\ P=P'. \tag{107}$$

(2) Similarly, if $\mathcal{P}\subseteq\mathbb{P}$ is a family of tiles, we say that $P\in\mathcal{P}$ is *minimal* (relative to $\mathcal{P}$) if and only if

$$\forall\,P'\in\mathcal{P}\ \text{ s.t. }\ P\geq P',\ \text{ we have }\ P=P'. \tag{108}$$

Definition 6 (*Incomparable and negligible families of tiles*). We say that $\mathcal{P} \subset \mathbb{P}$ is an *incomparable* family of tiles if and only if

$$\forall\, P_1,\, P_2 \in \mathcal{P} \text{ distinct, we have } P_1 \not\leq P_2 \text{ and } P_2 \not\leq P_1. \tag{109}$$

Also we call $\mathcal{P} \subset \mathbb{P}$ *negligible* if $\mathcal{P}$ can be written as a union of at most $c(d)$ incomparable families of tiles.

We end this subsection with the following observation, which connects the qualitative statement $P_1 \leq P_2$ with the geometric factor of the pair (P_1, P_2)—which is of quantitative nature—and with the analytic behavior of the polynomials belonging to $\{P_j\}_j$, respectively:

Observation 6. Let $P_1 = [\vec{\alpha}_1, I_{P_1}]$, $P_2 = [\vec{\alpha}_2, I_{P_2}] \in \mathbb{P}$. Then, the following hold:

(1) If $P_1 \leq P_2$, then $\Delta(P_1, P_2) = 0$. Conversely, if $\Delta(P_1, P_2) = 0$, then $2P_1 \leq 2P_2$.
(2) If $P_1 \leq P_2$ then, as a consequence of Definition 4 and Lemma C in the appendix, there exists $c(d) \in (0,\, (100\, d)^d]$ such that
  - there exists $q_1 \in P_1$ with

$$\sup_{q_2 \in P_2} \|q_2 - q_1\|_{L^\infty(\tilde{I}_{P_2})} \leq c(d)\, |I_{P_2}|^{-1}; \tag{110}$$

  - for all $q_1 \in P_1$, we have

$$\sup_{q_2 \in P_2} \|q_2 - q_1\|_{L^\infty(\tilde{I}_{P_1})} \leq c(d)\, |I_{P_1}|^{-1}. \tag{111}$$

6.2. *Tile partitioning*: *the stopping-time inductive algorithm*. In this section we present an inductive algorithm for partitioning our set of tiles into

$$\mathbb{P} = \bigcup_{n \in \mathbb{N}} \mathbb{P}_n, \tag{112}$$

with each $\mathbb{P}_n$ being a set of tiles of mass n relative to certain space regions. Our algorithm will be based on a stopping time process involving the John–Nirenberg inequality that is correlated with the level set analysis of various counting functions. This process is constructive and is based on an ascending induction over n.

Now before initiating our construction, we need to introduce the following:

Definition 7. Let $\mathcal{A} = \bigcup \mathcal{A}_j$ and $\mathcal{B} = \bigcup_k \mathcal{B}_k$ be two sets such that both $\{\mathcal{A}_j\}_j$ and $\{\mathcal{B}_k\}_k$ are collections of maximal (disjoint) dyadic intervals.

We write

$$\mathcal{A} \prec \mathcal{B} \tag{113}$$

if and only if each $\mathcal{A}_j$ is contained in some $\mathcal{B}_k$.

Moreover, given an absolute constant $c>0$, we write

$$\mathcal{A} \prec_c \mathcal{B} \tag{114}$$

if and only if $\mathcal{A} \prec \mathcal{B}$ and for every $\mathcal{B}_k$, the following holds:

$$|\bigcup_{\mathcal{A}_j\subseteq\mathcal{B}_k} \mathcal{A}_j| \leq 2^{-c}\,|\mathcal{B}_k|. \tag{115}$$

Step 1. *Construction of the family* $\mathbb{P}_1$. This construction will proceed in two stages:

- *Stage* 1.1 We define a sequence of nested sets $\{A_1^k\}_{k\in\mathbb{N}}$ such that
  - A_1^k is a finite union of maximal disjoint dyadic intervals,
  - $A_1^k \prec_c A_1^{k-1}$ for some $c>1$ and any $k\geq 1$, and
  - the L^∞ norm of a suitable "counting function of order one" adapted to A_1^k is under control.
- *Stage* 1.2 For each set A_1^k, we define a corresponding family of tiles $\mathcal{P}_1[A_1^k]$ with the following two key properties:
  - $\mathcal{P}_1[A_1^k]$ is a *convex* family of tiles—that is, if $P_1\leq P\leq P_2$ with $P_1,\,P_2\in\mathcal{P}_1[A_1^k]$, then $P\in\mathcal{P}_1[A_1^k]$;
  - each tile $P=[\vec{\alpha}_P,I_P]\in\mathcal{P}_1[A_1^k]$ has the properties

$$\begin{gathered} I_P\subseteq A_1^k \;\text{ and }\; I_P\nsubseteq A_1^{k+1},\\ 2^{-1}<A_{\mathbb{P},A_1^k}(P)\leq 1. \end{gathered} \tag{116}$$

This being said, we are ready to initiate the following:

Stage 1.1: *Construction of the sets* $\{A_1^k\}_{k\geq 0}$. As mentioned earlier, we apply an inductive argument.

1.1.1. *Step* $k=0$. We define the set A_1^0 as

$$A_1^0 := [0,1]. \tag{117}$$

1.1.2. *Step* $k\geq 1$. Since the first step was already verified, we assume that as the byproduct of the step $k-1$ we obtained a set

$$A_1^{k-1},$$

which can be represented as a finite union of disjoint dyadic intervals.

- We start by identifying the collection of maximal tiles

$$\mathcal{P}_1^{\max}[A_1^{k-1}] := \left\{P=[\vec{\alpha}_P,I_P]\in\mathbb{P}\,\Big|\;\begin{matrix} P\text{ maximal}\\ I_P\subseteq A_1^{k-1}\end{matrix}\;\text{ and }\;\frac{|E(P)|}{|I_P|}>2^{-1}\right\}. \tag{118}$$

Remark that $\mathcal{P}_1^{\max}[A_1^{k-1}]$ consists of incomparable tiles.

- Next, we define

$$\mathcal{C}_1[A_1^{k-1}] := \sum_{P\in\mathcal{P}_1^{\max}[A_1^{k-1}]} \chi_{E(P)} \tag{119}$$

and notice, based on the remark above, that

$$\forall\, x\in[0,1] \quad \mathcal{C}_1[A_1^{k-1}](x)\le 1. \tag{120}$$

- Define the *counting function of order one adapted to* A_1^{k-1} as

$$\mathcal{N}_1[A_1^{k-1}] := \sum_{P\in\mathcal{P}_1^{\max}[A_1^{k-1}]} \chi_{I_P}. \tag{121}$$

From (120) we notice that $\mathcal{N}_1[A_1^{k-1}]$ satisfies the relation

$$\|\mathcal{N}_1[A_1^{k-1}]\|_{\mathrm{BMO}_C} := \sup_{\substack{J\mathrm{dyadic}\\ J\subseteq[0,1]}} \frac{\sum_{\substack{I_P\subseteq J\\ P\in\mathcal{P}_1^{\max}[A_1^{k-1}]}} |I_P|}{|J|} \le 2. \tag{122}$$

- Now setting

$$\|\mathcal{N}_1[A_1^{k-1}]\|_{\mathrm{BMO}_D} := \sup_{\substack{J\mathrm{dyadic}\\ J\subseteq[0,1]}} \frac{1}{|J|}\int_J \left|\mathcal{N}_1[A_1^{k-1}] - \frac{\int_J \mathcal{N}_1[A_1^{k-1}]}{|J|}\right|, \tag{123}$$

we deduce that

$$\|\mathcal{N}_1[A_1^{k-1}]\|_{\mathrm{BMO}_D} \le 2\|\mathcal{N}_1[A_1^{k-1}]\|_{\mathrm{BMO}_C}. \tag{124}$$

- Applying the John–Nirenberg inequality, for $\gamma > c\,\|\mathcal{N}_1[A_1^{k-1}]\|_{\mathrm{BMO}_C}$, we have[48]

$$|\{x\in J\,|\sum_{\substack{I_P\subseteq J\\ P\in\mathcal{P}_1^{\max}[A_1^{k-1}]}} \chi_{I_P}(x)>\gamma\}| \lesssim 2^{-c}\,|J|. \tag{125}$$

- Conclude that the set

$$A_1^k := \{x\in[0,1]\,|\,\mathcal{N}_1[A_1^{k-1}](x) > c\,\|\mathcal{N}_1[A_1^{k-1}]\|_{\mathrm{BMO}_C}\} \tag{126}$$

can be written as a finite union of disjoint dyadic intervals with

$$A_1^k \prec_c A_1^{k-1}. \tag{127}$$

This process will end in a finite number of steps since the family $\mathbb{P}$ is finite.

[48]Throughout the section the constant $c\gg 1$ is an absolute constant that is allowed to change from line to line.

Observation 7. (1) Define $\mathcal{P}_1^{\max} := \bigcup_{k\geq 0}\mathcal{P}_1^{\max}[A_1^k]$, and let the *global counting function of order one* be

$$\mathcal{N}_1 := \sum_{P\in\mathcal{P}_1^{\max}} \chi_{I_P}. \tag{128}$$

Notice that as a consequence of the above construction, we have

$$\sup_k \|\mathcal{N}_1[A_1^k]\|_{\mathrm{BMO}_C} \lesssim \|\mathcal{N}_1\|_{\mathrm{BMO}_C} \lesssim \sup_k \|\mathcal{N}_1[A_1^k]\|_{L^\infty(A_1^k\setminus A_1^{k+1})} \lesssim 1. \tag{129}$$

Thus deduce that if we set $\bar{\mathcal{P}}_1^{\max}[A_1^k] := \mathcal{P}_1^{\max}[A_1^k]\setminus\mathcal{P}_1^{\max}[A_1^{k+1}]$ and define $\bar{\mathcal{N}}_1[A_1^k] := \sum_{P\in\bar{\mathcal{P}}_1^{\max}[A_1^k]}\chi_{I_P}$, then we have

$$\sup_k \|\bar{\mathcal{N}}_1[A_1^k]\|_{L^\infty} \lesssim 1. \tag{130}$$

(2) For any $0\leq l\leq k$, we have that $A_1^k\subseteq A_1^l$ with

$$A_1^k \prec_{(k-l)\,c} A_1^l. \tag{131}$$

Stage 1.2: *Construction of the sets* $\{\mathcal{P}_1[A_1^k]\}_{k\geq 1}$. As mentioned above, we will associate to each of the sets within

$$\{A_1^k\}_{k\geq 0},$$

constructed at Stage 1.1, a corresponding collection of tiles $\mathcal{P}_1[A_1^k]$.

Our construction process follows an ascending induction pattern.

- For $k=0$, define

$$\mathcal{P}_1[A_1^0] := \{P=[\vec{\alpha}_P, I_P]\in\mathbb{P}\,|\, I_P\nsubseteq A_1^1 \ \text{ and }\ A_{\mathbb{P},A_1^0}(P)\in(2^{-1},2^0]\}. \tag{132}$$

- For general $k\in\mathbb{N}$, we set

$$\mathcal{P}_1[A_1^k] := \left\{P=[\vec{\alpha}_P, I_P]\in\mathbb{P}\,|\ \begin{array}{l} I_P\subseteq A_1^k, I_P\nsubseteq A_1^{k+1} \\ A_{\mathbb{P},A_1^k}(P)\in(2^{-1},2^0] \end{array}\right\}. \tag{133}$$

- Finally, we now define *the collection of tiles of mass* (*of order*) 1 as

$$\mathbb{P}_1 := \bigcup_{k\geq 0}\mathcal{P}_1[A_1^k]. \tag{134}$$

Here the construction of the 1-mass set ends.

Step n. *Constructing the family* $\mathbb{P}_n$, $n\geq 2$. Following the above inductive algorithm, suppose now that for $j_1, j_2, \ldots j_{n-1}\in\mathbb{N}$ and $n\geq 2$, we have constructed the sets[49]

$$A_{n-1}^{j_{n-1}}[A_{n-2}^{j_{n-2}},\ldots,A_1^{j_1}]. \tag{135}$$

As before, for the n^{th} step we will have two stages:

[49]Throughout this section, if $n=2$, then we decree $A_{n-1}^{j_{n-1}}[A_{n-2}^{j_{n-2}},\ldots A_1^{j_1}] := A_1^{j_1}$ with the obvious adaptations/correspondences for the other defined concepts.

- *Stage n.1*: Define a finite sequence of nested sets

$$\left\{A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]\right\}_k$$

such that
  - $A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]$ is a finite union of maximal (disjoint) dyadic intervals;
  - $A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]\prec_{nc}A_{n-1}^{j_{n-1}}[A_{n-2}^{j_{n-2}},\ldots,A_1^{j_1}]$;
  - for any[50] $J\subseteq A_{n-s}^{j_{n-s}+1}[A_{n-s-1}^{j_{n-s-1}},\ldots,A_1^{j_1}]$ dyadic interval and any $s\in\{1,\ldots,n-1\}$, we either have

$$J\subset A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]$$

or

$$J\cap A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]=\emptyset.$$

- *Stage n.2*: For each set $A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]$, construct a corresponding family of tiles

$$\mathcal{P}_n(A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}])$$

such that
  - $\mathcal{P}_n(A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}])$ is a convex family;
  - each tile $P=[\vec{\alpha}_P,I_P]\in\mathcal{P}_n(A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}])$ has the properties

$$\begin{aligned}&I_P\subseteq A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}],\\&I_P\nsubseteq A_n^{k+1}[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]\text{ and also}\\&I_P\nsubseteq A_{n-s}^{j_{n-s}+1}[A_{n-s-1}^{j_{n-s-1}}\ \ldots,A_1^{j_1}]\ \forall\ s\in\{1,\ldots,n-1\},\\&2^{-n}<A_{\mathbb{P},A_n^k[A_{n-1}^{j_{n-1}}\ \ldots,A_1^{j_1}]}(P)\le 2^{-n+1}.\end{aligned}\tag{136}$$

Stage n.1: Construction of the sets $\left\{A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]\right\}_{k\in\mathbb{N}}$.

- As at Step 1, we will proceed by induction:
  - When $k=0$, we simply set

$$A_n^0[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]:=A_{n-1}^{j_{n-1}}[A_{n-2}^{j_{n-2}},\ldots,A_1^{j_1}].\tag{137}$$

  - For $k\ge 1$, we assume we have constructed

$$A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}].$$

[50]If $s=n-1$, we set $A_{n-s}^{j_{n-s}+1}[A_{n-s-1}^{j_{n-s-1}},\ldots,A_1^{j_1}]:=A_1^{j_1+1}$.

- Next we identify the collection of maximal tiles

$$\mathcal{P}_n^{\max}[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]:=$$

(138)
$$\left\{ P=[\vec{\alpha},I_P]\in\mathbb{P}\,\middle|\, \begin{array}{c} I_P\subseteq A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}] \\ I_P\nsubseteq A_{n-s}^{j_{n-s}+1}[A_{n-s-1}^{j_{n-s-1}},\ldots,\, A_1^{j_1}]\ \ \forall\, s<n \\ P \text{ maximal and } \frac{|E(P)|}{|I_P|}>2^{-n} \end{array} \right\}.$$

- Define

(139)
$$\mathcal{C}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]:=\sum\nolimits_{P\in\mathcal{P}_n^{\max}[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]}\chi_{E(P)},$$

and deduce that

(140)
$$\forall\, x\in[0,1],\qquad \mathcal{C}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]](x)\le 2^n.$$

- Define the *counting function of order n adapted to*

$$A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]$$

as

(141)
$$\mathcal{N}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]:=\sum\nolimits_{P\in\mathcal{P}_n^{\max}[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]}\chi_{I_P},$$

and from (140) notice that

$$\mathcal{N}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]\in \mathrm{BMO}_D$$

with

(142)
$$\|\mathcal{N}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]\|_{\mathrm{BMO}_C}\le 2^n.$$

- Now define the set

(143)
$$\begin{aligned} A_n^{k}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]:=\big\{x\in[0,1]\,\big|\,&\mathcal{N}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]](x)\\ &>n\,c\,\|\mathcal{N}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]\|_{\mathrm{BMO}_C}\big\}, \end{aligned}$$

and notice that via a John–Nirenberg argument similar with that in (125) one gets that $A_n^{k}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]$ is a finite union of maximal disjoint dyadic intervals with

(144)
$$A_n^{k}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]\prec_{n\,c} A_n^{k-1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}].$$

This process will end in a finite number of steps.

Observation 8. (1) Define

$$\mathcal{P}_n^{\max}:=\bigcup_k\ \bigcup_{j_1,\ldots j_{n-1}\in\mathbb{N}}\mathcal{P}_n^{\max}[A_n^{k}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]],$$

and let the *global counting function of order n* be

(145)
$$\mathcal{N}_n:=\sum_{P\in\mathcal{P}_n^{\max}}\chi_{I_P}.$$

Notice that as a consequence of the above construction we have

$$\sup_{k,\, j_1,\, \ldots,\, j_{n-1}} \|\mathcal{N}_n[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]]\|_{\mathrm{BMO}_C} \lesssim \|\mathcal{N}_n\|_{\mathrm{BMO}_C} \lesssim 2^n \tag{146}$$

and, moreover, defining the set of tiles

$$\begin{aligned}&\bar{\mathcal{P}}_n^{\max}[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]]\\ &\qquad := \mathcal{P}_n^{\max}[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]] \setminus \mathcal{P}_n^{\max}[A_n^{k+1}[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]],\end{aligned}$$

with the associated local counting function

$$\bar{\mathcal{N}}_n[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]] := \sum_{P \in \bar{\mathcal{P}}_n^{\max}[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]]} \chi_{I_P}, \tag{147}$$

we have

$$\sup_{k,\, j_1,\, \ldots,\, j_{n-1}} \|\bar{\mathcal{N}}_n[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]]\|_{L^\infty} \lesssim n\, 2^n. \tag{148}$$

(2) For any $k, l, j_1, \ldots, j_{n-1}$ with $k \geq l$, we have that

$$|A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]| \prec_{(k-l)\, n\, c} |A_n^l[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]|. \tag{149}$$

Stage n.2: Construction of the sets $\left\{\mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]]\right\}_{k\in\mathbb{N}}$.

- For $k \in \mathbb{N}$, we define

$$\begin{aligned}&\mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]]\\ &:= \left\{ \begin{array}{c} P \in \mathbb{P} \setminus \bigcup_{j=1}^{n-1} \mathbb{P}_j \\ P = [\vec{\alpha}_P, I_P] \end{array} \;\middle|\; \begin{array}{r} I_P \subseteq A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}] \\ I_P \nsubseteq A_n^{k+1}[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}] \\ \forall\, s < n \quad I_P \nsubseteq A_{n-s}^{j_{n-s}+1}[A_{n-s-1}^{j_{n-s-1}}, \ldots,\, A_1^{j_1}] \\ A_{\mathbb{P}, A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]}(P) \in (2^{-n}, 2^{-n+1}) \end{array} \right\}.\end{aligned} \tag{150}$$

- Next, we set

$$\mathbb{P}_n[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}] := \bigcup_k \mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}}, \ldots,\, A_1^{j_1}]]. \tag{151}$$

• Finally, we define *the collection of tiles of mass* (*of order*) n

$$\mathbb{P}_n := \bigcup_{j_1,\dots j_{n-1}\in\mathbb{N}} \mathbb{P}_n[A_{n-1}^{j_{n-1}},\dots, A_1^{j_1}]. \tag{152}$$

Here the construction of the n-mass set ends.

Remark that, from the above algorithm, we have

$$\mathbb{P} = \bigcup_{n\geq 0} \mathbb{P}_n. \tag{153}$$

This ends the partition of our set $\mathbb{P}$.

Observation 9. In light of the second item in Observation 4 one can immediately see the correspondences between the stopping time intervals $\{\mathcal{I}_n^l\}_{l,n\in\mathbb{N}}$ with their associated families of tiles $\{\mathbb{P}_n(\mathcal{I}_n^l)\}_{l,n\in\mathbb{N}}$ in Section 5.3 and the sets $A[\cdot]$ with their associated family of tiles constructed in the present section. Indeed, recall that (82) corresponds to (150), (83) corresponds to (148) while the collections of intervals $\{\mathcal{I}_n^l\}_{l,n\in\mathbb{N}}$ may now be obtained from the sets $A[\cdot]$ by the following procedure: Fix $n\in\mathbb{N}$, and denote by $\tilde{A}_n^{j_n}[A_{n-1}^{j_{n-1}},\dots,A_1^{j_1}]$ the collection of all maximal intervals in the decomposition of the set $A_n^{j_n}[A_{n-1}^{j_{n-1}},\dots,A_1^{j_1}]$. Then we set $\mathcal{I}_n^0 := [0,1]$ and inductively, for general $l\geq 1$, we define $\mathcal{I}_n^l$ as the collection of maximal intervals within

$$\bigcup_{j_s} \tilde{A}_n^{j_n}[A_{n-1}^{j_{n-1}},\dots,A_1^{j_1}] \setminus \bigcup_{r=0}^{l-1} \mathcal{I}_n^l.$$

7. Reduction of the Main Proposition

In this section, we will show that the understanding of the operator associated to the family $\mathbb{P}_n$ (the subject of our Main Proposition) can be reduced to the analysis of the operator associated to a *forest*—a better structured set of tiles that consists of a suitable collection of separated trees.

7.1. *Preliminaries.* We first introduce the main concepts that will play the central role in the analysis performed on the collections of tiles $\mathbb{P}_n$, $n\in\mathbb{N}$. Some of these have a direct analogue in the work of Fefferman [36].

Definition 8 (*Tree*). We say that a set of tiles $\mathcal{P}\subset\mathbb{P}$ is a *tree* (relative to $\leq$) with *top* P_0 if the following conditions[51] are satisfied:

(1) If $P\in\mathcal{P}$, then $\frac{3}{2}P \lneq 10P_0$.
(2) If $P\in\mathcal{P}$ and $P'\in N(P)$ such that $\frac{4}{3}P' \lneq 10P_0$, then $P'\in\mathcal{P}$.
(3) If $P_1,\ P_2\ \in\mathcal{P}$ and $P_1\leq P\leq P_2$, then $P\in\mathcal{P}$.

[51] To avoid the boundary problems arising from the use of a single dyadic grid and from the definition of our tiles, our reasonings will often involve a dilation factor of the tiles.

Definition 9 (*Sparse tree*). Let $C>0$ be an absolute constant. We say that a set of tiles $\mathcal{P}\subset\mathbb{P}$ is a C-*sparse tree* if $\mathcal{P}$ is a tree and for any $P=[\vec{\alpha},I]\in\mathcal{P}$, we have

$$\sum_{\substack{P'=[\vec{\alpha}',I']\in\mathcal{P}\\ I'\subseteq I}}|I'|\leq C\,|I|. \tag{154}$$

In our later reasonings, the specific value of the constant C will be of no relevance[52] and thus we will simply refer to a C-sparse tree as a *sparse tree*.

Definition 10. [L^∞-*forest*] Fix $n\in\mathbb{N}$. We say that $\mathcal{P}\subseteq\mathbb{P}_n$ is an L^∞-*forest* of generation n if and only if the following two conditions hold:

(1) $\mathcal{P}$ is a collection of *separated* trees; i.e.,

$$\mathcal{P}=\bigcup_{j\in\mathbb{N}}\mathcal{P}_j \tag{155}$$

with each $\mathcal{P}_j$ a tree with top $P_j=[\vec{\alpha}_j,I_j]$ and such that

$$\forall\ k\neq j\ \text{ and }\ \forall\ P\in\mathcal{P}_j,\qquad 2P\nleq 2P_k. \tag{156}$$

(2) The $\mathcal{P}$-*counting function*

$$\mathcal{N}_{\mathcal{P}}(x):=\sum_j\chi_{I_j}(x) \tag{157}$$

obeys the estimate $\|\mathcal{N}_{\mathcal{P}}\|_{L^\infty}\lesssim 2^n$.

Further, if $\mathcal{P}\subseteq\mathbb{P}_n$ consists only of sparse separated trees, then we refer to $\mathcal{P}$ as a *sparse* L^∞-*forest*.

Definition 11 (BMO-*forest*). A set $\mathcal{P}\subseteq\mathbb{P}_n$ is called a BMO-*forest* of generation n or just simply a *forest*[53] if and only if the following hold:

(1) $\mathcal{P}$ may be written as

$$\mathcal{P}=\bigcup_{j\in\mathbb{N}}\mathcal{P}_j, \tag{158}$$

with each $\mathcal{P}_j$ an L^∞-*forest* (of generation n);

(2) for any $P\in\mathcal{P}_j$ and $P'\in\mathcal{P}_k$ with $j,k\in\mathbb{N}$, $j<k$, we either have $I_P\cap I_{P'}=\emptyset$ or[54]

$$|I_{P'}|\leq 2^{j-k}\,|I_P|. \tag{159}$$

[52] All the constants C appearing in this context will be bounded by a positive absolute constant possibly depending only on d.

[53] When the context is clear we may no longer specify the order of the generation.

[54] The base 2 here has no relevance. One could replace it with any $c>1$ so that (159) transforms into $|I_{P'}|\leq c^{j-k}\,|I_P|$. More generally, it is in fact enough for the collection $\{I_{P_k}\}_k$ to obey a Carleson packing condition.

As before, if $\mathcal{P} \subseteq \mathbb{P}_n$ consists only of sparse L^∞-forests, then, we refer at $\mathcal{P}$ as a *sparse forest*.

Observation 10. Notice that if $\mathcal{P} \subseteq \mathbb{P}_n$ is a forest then, due to (159) above, the counting function

$$\mathcal{N}_{\mathcal{P}} := \sum_j \mathcal{N}_{\mathcal{P}_j} \tag{160}$$

obeys the estimate

$$\|\mathcal{N}_{\mathcal{P}}\|_{\mathrm{BMO}_C} \lesssim 2^n, \tag{161}$$

whence the alternative name of the BMO-forest.

Also notice that if $\mathcal{P} \subseteq \mathbb{P}_n$ is a collection of separated trees, then $\mathcal{P}$ is automatically a (BMO)-forest.

We can now provide the key statements on which the proof of our Main Proposition relies. Their proofs will be the focus of the remaining part of the paper.

PROPOSITION 1 (Control over a sparse forest). *Let $\mathcal{P} \subseteq \mathbb{P}_n$ be a sparse forest. Then there exists $\eta = \eta(d) \in (0, \frac{1}{2})$, depending only on the degree d, such that for any $1 < p < \infty$, we have*

$$\left\|T^{\mathcal{P}}\right\|_p \lesssim_{p,d} 2^{-n\,\eta\,(1-\frac{1}{p^*})}. \tag{162}$$

PROPOSITION 2 (Control over a general forest). *Let $\mathcal{P} \subseteq \mathbb{P}_n$ be a forest. Then there exists $\eta = \eta(d) \in (0, \frac{1}{2})$, depending only on the degree d, such that for any $1 < p < \infty$, we have*

$$\left\|T^{\mathcal{P}}\right\|_p \lesssim_{p,d} 2^{-n\,\eta\,(1-\frac{1}{p^*})}. \tag{163}$$

7.2. *Reduction of the Main Proposition to Proposition* 2.

Aim: In this section we intend to show that, for a fixed $n \in \mathbb{N}$, the set $\mathbb{P}_n$ can be roughly decomposed into a union of $c\,n$ forests for some suitable $c \in \mathbb{N}$.

We start by recalling (150)–(152), from which we deduce

$$\mathbb{P}_n := \bigcup_{k\in\mathbb{N}} \bigcup_{j_1,\dots j_{n-1}\in\mathbb{N}} \mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}}, \dots, A_1^{j_1}]]. \tag{164}$$

We now make the following

Claim 1. For each $j_1, \dots, j_{n-1}, k \in \mathbb{N}$, the set

$$\mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}}, \dots, A_1^{j_1}]] \tag{165}$$

can be decomposed as a union of at most $c\,n$ L^∞-forests (of generation n)

$$\{\mathcal{P}_n^s[A_n^k[A_{n-1}^{j_{n-1}}, \dots, A_1^{j_1}]]\}_{s\in\{1,\dots,c\,n\}}, \tag{166}$$

where here $c \in \mathbb{N}$ is some absolute constant.

Observation 11. Notice that if we believe our claim for the moment, then denoting

$$\mathbb{P}_n^s := \bigcup_{k\in\mathbb{N}} \bigcup_{j_1,\ldots j_{n-1}\in\mathbb{N}} \mathcal{P}_n^s[A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]], \tag{167}$$

we have that $\mathbb{P}_n^s$ is a BMO-forest. Indeed, this follows from the key condition (144) in our construction of tiles and from the fact that, as a consequence of Claim 1, each $\mathcal{P}_n^s[A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]$ is an L^∞-forest.

Thus, since

$$\mathbb{P}_n := \bigcup_{s=0}^{c\,n} \mathbb{P}_n^s,$$

we conclude that $\mathbb{P}_n$ can be written as a union of at most $c\,n$ forests as desired.

We start by recalling the construction from Step n and Observation 8 in our previous section. Maintaining the same notation, assume we are given the following:

- the set $A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]$;
- the set of tiles $\mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]$, and
- the set of maximal tiles $\bar{\mathcal{P}}_n^{\max}[A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]$.

We first notice that $\bar{\mathcal{P}}_n^{\max}[A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]$ represents the set of all the maximal elements $P\in\mathcal{P}_n^{\max}[A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]$ such that $\frac{|E(P)|}{|I_P|}>2^{-n}$, $I_P\subset A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]$, and $I_P\nsubseteq A_n^{k+1}[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]$.

Next we recall the good L^∞-control on the associated counting function $\bar{\mathcal{N}}_n[A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]]$ as directed by (148).

Throughout this section fix the values of $k,\, j_1,\ldots,j_{n-1}, n\in\mathbb{N}$. In what follows, for notational simplicity, we will omit the dependence on the expression $A_n^k[A_{n-1}^{j_{n-1}},\ldots,\, A_1^{j_1}]$.

Observation 12. The forest decomposition of $\mathbb{P}_n$—our main task in the present section—relies on an algorithm that involves dilation of tiles. This latter fact is required by the presence of the so-called "boundary effect" emerging from the construction of our tiles and whose manifestation can be seen in two key instances: (1) designing the convex structure of a tree, and (2) creating "spaces" (i.e., separation) among trees inside our family $\mathcal{P}_n$.

The presence of the boundary effect in the context of the first item is a direct consequence of the higher-order wave-packets involved in our problem, which impose on our tiles a multi-interval frequency location, i.e., $P=[\vec{\alpha}, I]$ with $\vec{\alpha}=(\alpha^1,\alpha^2,\ldots,\alpha^d)$. As a side note, in the linear case $d=1$ addressing the classical Carleson operator (and hence standard wave-packet theory), the first item does not raise any issue since in this situation " $\leq$ " is a partial order relation—unlike in the general case treated here (see Observation 5).

Regarding the second item, the situation is as follows: Previously, in Fefferman's paper on the classical Carleson operator ([36]) this second item was treated by taking *averages* of suitable model operators over a *continuum* of shifted dyadic grids. A similar strategy was later pursued in [73]. In the present paper, we introduce a novel feature by designing an algorithm that allows the usage of a *single dyadic grid* [55] in the discretization of our operator—thus eliminating the necessity of considering averages over model operators—at the cost of appealing in our reasonings to tile dilations.[56] This new algorithm, to which the remainder of this section is devoted, incorporates with minor adaptations Fefferman's approach in [36] (covered here within (168)–(179)) and was first developed by the author in [81].

As already mentioned in the above observation, the main challenge in proving our claim is to form "spaces" among the trees. But for this, we will need first to create the tree structures. Thus, our first step is to "stick" every tile $P \in \mathcal{P}_n$ to a top (a maximal tile with respect to " $\leq$ "). More precisely, we will proceed as follows:

Let $\bar{\mathcal{P}}_n^{\max} = \left\{\bar{P}_j\right\}_j$. As in [81], we define

$$\bar{\mathcal{P}}_n := \left\{P \in \mathcal{P}_n \mid \exists \;\; j \in \mathbb{N} \;\; s.t. \;\; 4P \lhd \bar{P}_j\right\} \tag{168}$$

and further define the set

$$\mathcal{C}_n := \left\{P \in \mathcal{P}_n \mid \text{there are no chains } P \lneq P_1 \lneq \cdots \lneq P_n \;\&\; \{P_j\}_{j=1}^n \subseteq \mathcal{P}_n \right\}. \tag{169}$$

With this done, we claim that

$$\mathcal{P}_n \setminus \mathcal{C}_n \subseteq \bar{\mathcal{P}}_n. \tag{170}$$

Indeed, assume that $P \in \mathcal{P}_n \setminus \mathcal{C}_n$. Then from (169) we have that there exist $\{P_j\}_{j=1}^n \subseteq \mathcal{P}_n$ such that

$$P \lneq P_1 \lneq \cdots \lneq P_n. \tag{171}$$

Then, since $P_n \in \mathcal{P}_n$, we must have

$$A_{\mathbb{P}, A_n^k[A_{n-1}^{j_{n-1}}, \ldots, A_1^{j_1}]}(P_n) \in (2^{-n},\, 2^{-n+1}]$$

and hence, from the definition of $\bar{\mathcal{P}}_n^{\max}$ and Definition 3, we have

$$\exists\, \bar{P} \in \bar{\mathcal{P}}_n^{\max} \text{ s.t. } \Delta(10P_n,\, 10\bar{P}) < 2^{\frac{n}{N}}, \tag{172}$$

[55] This feature becomes of key importance in the study of the boundedness of the Carleson operator near L^1 (for more on this, see Section 12 in [85]) with the latter theme being discussed in the last item of Section 11.

[56] Hence for this second item, the presence of the boundary effect has to do with our single dyadic grid discretization choice and not with the degree d involved in the higher order wave-packet analysis.

which, based on Lemma C in the appendix, implies

$$\sup_{\substack{q_n\in P_n\\ \bar{q}\in\bar{P}}} \|q_n-\bar{q}\|_{L^\infty(\tilde{I}_{P_n})} \leq 20\,(100\,d)^d\,|I_{P_n}|^{-1}\,2^{\frac{n}{N}}. \tag{173}$$

This last relation together with (171) and (111) gives us

$$\sup_{\substack{q_1\in P_1\\ \bar{q}\in\bar{P}}} \|q_1-\bar{q}\|_{L^\infty(\tilde{I}_{P_1})} \leq 50\,(100\,d)^d\,|I_{P_1}|^{-1}. \tag{174}$$

Appealing now to (110) and making essential use of the second item in Observation 2, we have that

$$\exists\, q\in P \text{ s.t. } \sup_{\bar{q}\in\bar{P}} \|q-\bar{q}\|_{L^\infty(\tilde{I}_{P_1})} \leq 50\,(100\,d)^d\,|I_{P_1}|^{-1} < |I_P|^{-1}, \tag{175}$$

which now implies that $4P\lhd\bar{P}$, thus proving (170).

Now, defining the set $\mathcal{D}_n\subseteq\mathcal{C}_n$ with the property $\mathcal{P}_n\setminus\mathcal{D}_n=\bar{\mathcal{P}}_n$, we remark that $\mathcal{D}_n$ breaks up as a disjoint union of at most n sets $\mathcal{D}_n^1\cup\mathcal{D}_n^2\cup\cdots\cup\mathcal{D}_n^n$ with each $\mathcal{D}_n^j$ being (recall Definition 6) an incomparable family of tiles. As a consequence, $\mathcal{D}_n$ may be written as a union of at most n sparse L^∞-forests and hence, assuming that Proposition 1 holds,[57] we can erase this set from $\mathcal{P}_n$ without affecting our claim.

Thus, in what follows, it will be enough to limit ourselves to the set of tiles $\bar{\mathcal{P}}_n$ which for convenience we will re-denote by $\mathcal{P}_n$.

Returning to our Claim 1, our aim is to show that

$$\mathcal{P}_n=\bigcup_{j=1}^{c\,n}\mathcal{S}_{nj}, \tag{176}$$

with each $\mathcal{S}_{nj}$ an L^∞-forest of generation n.

Now set

$$B(P):=\#\left\{j\mid 4P\trianglelefteq\bar{P}_j\right\}\qquad\forall\ P\in\mathcal{P}_n. \tag{177}$$

Notice that based on (148), (168), and (177) we have that

$$\mathcal{P}_n=\bigcup_{j=1}^{c\,n}\mathcal{P}_{nj}, \tag{178}$$

with[58]

$$\begin{aligned}\mathcal{P}_{nj}:=\left\{P\in\mathcal{P}_n\mid 2^{j-1}\leq B(P)<2^j\right\}&\\ \text{and then } j \text{ runs through the set } \{1,\ldots,cn\}.&\end{aligned} \tag{179}$$

---

[57] Notice that Proposition 1 is just a very particular case of Proposition 2.

[58] Strictly speaking, from the construction of $\mathcal{P}_n$ one deduces that $\mathcal{P}_{nj}=\emptyset$ if $2^j>n\,2^n$.

In what follows, we will show that each set $\mathcal{P}_{nj}$ can be written as

$$\mathcal{P}_{nj} = \mathcal{S}_{nj} \cup \mathcal{R}_{nj}, \tag{180}$$

such that

- $\mathcal{S}_{nj}$ is an L^∞-forest of generation n;
- $\mathcal{R}_{nj}$ is a *negligible* collection of tiles.

Observation 13. Strictly speaking $\mathcal{S}_{nj}$ as defined above is guaranteed to be an L^∞-forest only if $j \gtrsim \log n$. For small(er) values of j, the counting function associated with the separated trees within $\mathcal{S}_{nj}$ obeys only $\|\mathcal{N}_{\mathcal{S}_{nj}}\|_{L^\infty} \lesssim n\, 2^{n-j}$ instead of the desired estimate $\|\mathcal{N}_{\mathcal{S}_{nj}}\|_{L^\infty} \lesssim 2^n$ that should be required according to Definition 10. However the latter can be easily achieved by further subdividing $\mathcal{S}_{nj}$ into at most $c\, \frac{n}{2^j}$ subfamilies with each being a genuine L^∞-forest. Alternatively, one can simply modify Definition 10 by allowing an extra harmless logarithmic factor in the bound of the counting function defined in (157), i.e., $\|\mathcal{N}_{\mathcal{P}}\|_{L^\infty} \lesssim n\, 2^n$.

Now fix a family $\mathcal{P}_{nj}$.

Step 1. *Identifying the candidates for the tops of the future trees.* For this, we let

$$\mathcal{P}_{nj}^{\max} := \{P^r = [\vec{\alpha}_r,\, I_r]\}_{r\in\{1,\dots,s\}} \subseteq \mathcal{P}_{nj} \tag{181}$$

be the set of tiles with the property that

$$4P^r \text{ is maximal with respect to } \leq \text{ inside the set } 4\mathcal{P}_{nj}. \tag{182}$$

Now, in many of the further reasonings we will use the following:

Four key properties.

(A) $4P^l \leq 4P^m$ implies $I_l = I_m$;
(B) for all $P \in \mathcal{P}_{nj}$, there exists P^l such that $12P \trianglelefteq 4P^l$;
(C) if $P \in \mathcal{P}_{nj}$ such that there exists $m \neq l$ with $\begin{cases} 4P \trianglelefteq 4P^l, \\ 4P \trianglelefteq 4P^m \end{cases}$, then $\begin{cases} 4P^m \leq 4P^l, \\ 4P^l \leq 4P^m \end{cases}$;
(D) if $P_j = [\vec{\alpha}, I_j] \in \mathbb{P}$ with $j \in \{1,2\}$ such that $|I_1| \neq |I_2|$, then $|I_1| \leq 2^{-D}\,|I_2|$ or $|I_2| \leq 2^{-D}\,|I_1|$, where we recall here that $D \in \mathbb{N}$ such that $D \geq 100d\log_2(100d)$.

The four properties explained.

(A) This is an immediate consequence of (182) and Definition 5.
(B) From (182) we have that for any $P \in \mathcal{P}_{nj}$, there exists P^l such that $4P \leq 4P^l$; now (B) is essentially a consequence of Observation 5.
(C) This follows from a contrapositive reasoning: if $4P^l$ and $4P^m$ are incomparable, then, using the fact that " $\trianglelefteq$ " is a partial order relation (see

Observation 5), we deduce that $B(P) \geq 2^{j+1}$, thus contradicting the fact that $P \in \mathcal{P}_{nj}$;

(D) This simply restates (54); see Observation 2.

Step 2. *Isolating the negligible family of tiles $\mathcal{R}_{nj}$.* Our aim here is to properly trim the set $\mathcal{P}_{nj}$ so that the resulting family will have all the desired properties of an L^∞-forest of generation n.

In order to do so, we define the following three sets:

- $\mathcal{R}^1_{nj}$: the family of tiles that are "far away" from $\mathcal{P}^{\max}_{nj}$:

$$\mathcal{R}^1_{nj} := \left\{P \in \mathcal{P}_{nj} \,|\, \forall\ P^l \text{ one has } \frac{3}{2}P \not\leq P^l\right\}. \tag{183}$$

- $\mathcal{R}^2_{nj}$: the family of neighboring maximal tiles:

$$\mathcal{R}^2_{nj} := \left\{P \in \mathcal{P}_{nj} \,|\, \exists\, P^l \text{ such that } P \in N(P^l)\right\}. \tag{184}$$

- $\mathcal{R}^3_{nj}$: the family of neighboring minimal tiles:

$$\mathcal{R}^3_{nj} := \{P \in \mathcal{P}_{nj} \,\big|\, \exists\, P' \in \mathcal{P}_{nj} \setminus (\mathcal{R}^1_{nj} \cup \mathcal{R}^2_{nj}) \text{ minimal such that } P \in N(P')\}. \tag{185}$$

With this, we define

$$\mathcal{R}_{nj} := \mathcal{R}^1_{nj} \cup \mathcal{R}^2_{nj} \cup \mathcal{R}^3_{nj}. \tag{186}$$

Claim 2. The set $\mathcal{R}_{nj}$ is a negligible family of tiles.

The proof of this claim will be done via contradiction:

- *For the set* $\mathcal{R}^1_{nj}$: Assume that there exist P_1, $P_2 \in \mathcal{R}^1_{nj}$ such that $P_1 \lneq P_2$. Now applying (B) we have that there exist P^{l_1}, $P^{l_2} \in \mathcal{P}^{\max}_{nj}$ such that $12P_i \trianglelefteq 4P^{l_i}$ with $i \in \{1, 2\}$. Now using (D) we must have $\frac{3}{2}P_1 \trianglelefteq 12P_2 \trianglelefteq 4P^{l_2}$, thus contradicting the assumption that $P_1 \in \mathcal{R}^1_{nj}$.
- *For the set* $\mathcal{R}^2_{nj}$: Assume that there exist P_1, $P_2 \in \mathcal{R}^2_{nj}$ such that $P_1 \lneq P_2$ and hence there exist P^{l_1}, $P^{l_2} \in \mathcal{P}^{\max}_{nj}$ such that $|I_{P_1}| = |I_{P^{l_1}}| < |I_{P^{l_2}}| = |I_{P_2}|$ and $\frac{3}{2}P_i \leq P^{l_i}$ for $i \in \{1, 2\}$. Applying (D) we have that $4P^{l_1} \lneq 4P^{l_2}$, contradicting the maximality assumption.
- *For the set* $\mathcal{R}^3_{nj}$: Assume that there exist P_1, P_2, $P_3 \in \mathcal{R}^3_{nj}$ such that $P_1 \lneq P_2 \lneq P_3$. The key observation is that any $P \in N(P_2)$ must belong to $\mathcal{P}_{nj}$ since $4P_1 \triangleleft 4P \triangleleft 4P_3$. This however implies that for any $P' \in N(P_3)$, there exists $P \in N(P_2)$ such that $P \lneq P'$ thus contradicting the assumption that $P_3 \in \mathcal{R}^3_{nj}$.

Step 3. *Verifying that the set $\mathcal{S}_{nj} := \mathcal{P}_{nj} \setminus \mathcal{R}_{nj}$ is an L^∞-forest.* For the remaining set $\mathcal{S}_{nj}$, we proceed as follows:

(1) Set

$$S_m := \left\{ P \in \mathcal{S}_{nj} \mid \frac{3}{2}P \leq P^m \right\}.$$

In what follows we consider only those sets S_m that are non-empty. Without loss of generality we may suppose that

$$\mathcal{S}_{nj} = \bigcup_{m=1}^{s} S_m.$$

(2) Introduce the "clustering" relation among the sets $\{S_m\}_m$:

$$S_m \propto S_l$$

if and only if there exists $P_1 \in S_m$ and there exists $P_2 \in S_l$ such that $2P_1 \leq 2P^l$ or $2P_2 \leq 2P^m$.

(3) Define a second relation on $\{S_m\}_m$ given by

$$S_m \backsim S_l$$

if and only if $4P^m \leq 4P^l$ or equivalently $4P^l \leq 4P^m$.

(4) Deduce that $S_m \propto S_l$ implies $S_m \backsim S_l$, and making use of property (C) conclude that " $\backsim$ " is an equivalence relation.

(5) Let $\hat{m} := \{l \mid S_l \backsim S_m\}$; then the cardinality of $\hat{m}$ is at most $c(d)$, and for

$$\hat{S}_m := \bigcup_{m' \in \hat{m}} S_{m'},$$

one has that $\hat{S}_m$ is a tree having as a top any P^l with $l \in \hat{m}$.

Let us justify (1)–(5). Relations $(1), (2)$, and (3) are simply definitions, and thus we need only verify (4) and (5).

We start with item (4). Assume that $S_m \propto S_l$ with $m \neq l$, and thus without loss of generality that there exists $P_1 \in S_m$ such that $2P_1 \leq 2P^l$. Notice first that $|I_{P_1}| < |I_{P^l}|$, as otherwise we must have $P_1 \in \mathcal{R}^2_{nj}$, which is not allowed. Thus $4P_1 \trianglelefteq 4P^l$ and $4P_1 \trianglelefteq 4P^m$, and hence from (C) we conclude that $S_m \backsim S_l$. Next we need to show that " $\backsim$ " is an equivalence relation. The only non-trivial part is to check transitivity. Thus assume that $S_m \backsim S_l$ and $S_l \backsim S_r$. Since $S_l \neq \emptyset$, we have that there exists P with $|I_P| < |I_{P^l}|$ and $\frac{3}{2}P \leq P^l$. Since from (A) we must have $|I_{P^m}| = |I_{P^l}| = |I_{P^r}|$, deduce that $4P \trianglelefteq 4P^m$ and $4P \trianglelefteq 4P^r$. Thus, from (C) we further have $4P^m \leq 4P^r$, which implies $S_m \backsim S_r$.

We pass now to proving item (5). The fact that any equivalence class $\hat{m}$ has at most $c(d)$ elements is a direct consequence of item (3). We will now focus on proving that $\hat{S}_m$ is a tree by verifying all of the three items in Definition 8:

- If $P \in \hat{S}_m$, then $\frac{3}{2}P \leq 10P_0$ for any $P_0 \in \{P^l\}_{l \in \hat{m}}$.

This is a direct consequence of the fact that for all $P \in \hat{S}_m$, there exists P^l with $l \in \hat{m}$ such that $\frac{3}{2}P \leq P^l$. However for any other P^r with $r \in \hat{m}$, we have that $4P^r \leq 4P^l$ and $4P^l \leq 4P^r$ and hence $\frac{3}{2}P \leq 10P^r$.

- If $P \in \hat{S}_m$ and $P' \in N(P)$ such that $\frac{4}{3}P' \leq 10P_0$, then $P' \in \hat{S}_m$.

  Let us take $P \in \hat{S}_m$. Our task is to first show a milder fact: any $P' \in N(P)$ is available in the family $\mathcal{S}_{nj}$. Now recall two important facts: $\mathcal{S}_{nj} \subset \mathcal{P}_{nj}$ with $\mathcal{P}_{nj}$ defined in (179), and $\mathcal{S}_{nj}$ does not contain any neighboring minimal elements in $\mathcal{P}_{nj}$ since $\mathcal{S}_{nj} \subseteq \mathcal{P}_{nj} \setminus \mathcal{R}^3_{nj}$. Now from $P \in \hat{S}_m$ we deduce

(187)
  – there exists $l \in \hat{m}$ such that $\frac{3}{2}P \lneq P^l$;
  – there exist $P^{\min} \in \mathcal{P}_{nj} \setminus \mathcal{S}_{nj}$ and a chain $\{P_i\}_{i=1}^{M} \subseteq \mathcal{P}_{nj}$ with $M \in \mathbb{N}$, $M \geq 2$ such that $P_M \lneq P_{M-1} \lneq \cdots \lneq P_1$ and $P_M = P^{\min}$ and $P_1 = P$.

  This immediately implies the following key relation:

$$4P^{\min} \lhd 4P' \lhd 4P^l. \tag{188}$$

  From Definition 3 and relation (179) we conclude that

$$\text{if } P' \in N(P), \text{ then } P' \in \mathcal{P}_{nj}.$$

  Recalling (185), we notice that $P' \notin \mathcal{R}^3_{nj}$ since otherwise this would contradict our hypothesis $P \in \hat{S}_m$. With these, deduce that if $P' \in N(P)$ such that $\frac{4}{3}P' \lneq 10P^l$, then $\frac{3}{2}P' \lneq P^l$ implying $P' \in \mathcal{S}_{nj}$ and hence $P' \in \hat{S}_m$.

- If $P_1,\ P_2 \in \hat{S}_m$ and $P_1 \leq P \leq P_2$, then $P \in \hat{S}_m$.

  We may assume without loss of generality that $P_1 \lneq P \lneq P_2$. Notice that in this case we immediately have that $\frac{3}{2}P_1 \lhd \frac{3}{2}P \lhd \frac{3}{2}P_2$, which immediately implies both that $P \in \mathcal{P}_{nj}$ and that $\frac{3}{2}P \lhd P^m$, and hence $P \in \hat{S}_m$.

  This proves our item (5), saying that $\hat{S}_m$ is a tree with top $P_0 \in \{P^l\}_{l\in\hat{m}}$.

  Finally, since for any two distinct $\hat{S}_m$ and $\hat{S}_l$ we have that taking any correspondent $S_{m'} \in \hat{S}_m$ and $S_{l'} \in \hat{S}_l$ the relation $S_{m'} \propto S_{l'}$ does *not* hold, we conclude that the set

$$\mathcal{S}_{nj} = \bigcup_m \hat{S}_m \tag{189}$$

is an L^∞-forest as in Definition 10.

We end this section with the following

Observation 14. For the remaining part of the paper, whenever dealing with $\mathcal{P}$ a generic (sparse) forest—as, for example, in the statements of Propositions 1 and 2—one should think at $\mathcal{P}$ as

$$\bigcup_{k,j\in\mathbb{N}} \bigcup_{j_1,\dots j_{n-1}\in\mathbb{N}} \mathcal{S}_{nj}[A_n^k[A_{n-1}^{j_{n-1}}, \dots,\ A_1^{j_1}]], \tag{190}$$

with each

$$\mathcal{S}_{nj}[A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]]\subset\mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}},\ldots,A_1^{j_1}]]$$

constructed in a similar fashion as S_{nj} in (189).

8. **Proof of Proposition 1**

This section will be divided in three subsections: the first one deals with the concept of dyadic Calderón–Zygmund decomposition of a set subordinated to a family of dyadic intervals, a useful prerequisite, for the next two subsections in which we will treat the L^p-boundedness, $1<p<\infty$, of the operator associated with a generic sparse forest.

8.1. *Dyadic Calderón–Zygmund set decomposition relative to a family of dyadic intervals.* Assume we are given a set $S=\bigcup_j S_j\subseteq[0,1]$ consisting of a finite union of disjoint dyadic intervals $\{S_j\}_j$ and a family $\mathcal{I}$ of dyadic intervals, not necessarily disjoint, with the property that $\bigcup_{I\in\mathcal{I}} I\subseteq S$ and that for any S_j, there exists $I\in\mathcal{I}$ such that $I\subseteq S_j$. Now define

$$\mathcal{I}_{\min}:=\{I\in\mathcal{I}\,|\,I\text{ minimal interval relative to inclusion}\}. \tag{191}$$

Next, let

$$\breve{\mathcal{I}}(S):=\left\{\begin{array}{c}J\subset S\\ J\text{ dyadic}\end{array}\;\middle|\;\begin{array}{c}\text{precisely one of the left or right children}\\ \text{of } J\text{ contains an element of }\mathcal{I}_{\min}\end{array}\right\}$$

and

$$\breve{\mathcal{I}}_c(S):=\left\{\begin{array}{c}J_0\text{ children of }J\\ J\in\breve{\mathcal{I}}(S)\end{array}\;\middle|\;J_0\cap\bigcup_{I\in\mathcal{I}_{\min}}I=\emptyset\right\}.$$

We now define the dyadic Calderón–Zygmund S-decomposition subordinated to the family $\mathcal{I}$ as

$$CZ_S(\mathcal{I}):=\mathcal{I}_{\min}\cup\breve{\mathcal{I}}_c(S). \tag{192}$$

Observation 15. (1) Notice that the collection of dyadic intervals represented by $CZ_S(\mathcal{I})$ forms a partition of S. Moreover, for any $J\in CZ_S(\mathcal{I})$ and $I\in\mathcal{I}$, one has that either $J\cap I=\emptyset$ or $J\subseteq I$.

(2) If $\mathcal{P}$ is a collection of tiles, we define $\mathcal{I}_{\mathcal{P}}:=\{I\,|\,\exists\,P=[\vec{\alpha},I]\in\mathcal{P}\}$. Then, for any $S\subseteq[0,1]$ finite union of dyadic intervals with the property $\bigcup_{I\in\mathcal{I}_{\mathcal{P}}}I\subseteq S$, we define

$$CZ_S(\mathcal{P}):=CZ_S(\mathcal{I}_{\mathcal{P}}). \tag{193}$$

Similarly, recalling (40) and letting $\mathcal{I}_{\mathcal{P}^*}:=\{I_{P^*}^j\}_{\substack{P\in\mathcal{P}\\1\le j\le 14}}$ if $\bigcup_{I\in\mathcal{I}_{\mathcal{P}^*}}I\subseteq S$, we set $CZ_S(\mathcal{P}^*):=CZ_S(\mathcal{I}_{\mathcal{P}^*})$.

(3) A key feature of the Calderón–Zygmund construction presented above is given by the following: let $\mathcal{P} \subseteq \mathbb{P}$ be a tree with top $P_0 = [\vec{\alpha}_0, I_0]$, and assume that

$$A_{\mathbb{P},I_0}(P) \lesssim \delta \quad \forall \quad P \in \mathcal{P} \tag{194}$$

for some $\delta \in (0,1]$.

Then, for any $P' = [\vec{\alpha}', I'] \in \mathbb{P}$ with the properties that $I' \in CZ_{I_0}(\mathcal{P})$ and $10P' \leq 10P$, one still has

$$A_{\mathbb{P},I_0}(P') \lesssim \delta. \tag{195}$$

This is a direct consequence of the smoothening effect encoded in the mass definition (106).

(4) More generally, let

$$\mathcal{P} \subseteq \mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}}, \ldots, A_1^{j_1}]] \text{ and } S = A_n^k[A_{n-1}^{j_{n-1}}, \ldots, A_1^{j_1}].$$

Assume that there exists $q_0 \in \mathcal{Q}_{d-1}$ and $\mu \in [0,1]$ such that for any $P \in \mathcal{P}$ one has $\lceil \Delta_{q_0}(P) \rceil \leq \mu$.

Then for any $P' = [\vec{\alpha}', I'] \in \mathbb{P}$ with the properties $I' \in CZ_S(\mathcal{P})$ and $\lceil \Delta_{q_0}(P') \rceil \leq \mu$, one has

$$A_{\mathbb{P},S}(P') \lesssim 2^{-n} \, \mu^{-N}, \tag{196}$$

where here N is the parameter appearing in (106).

8.2. *Sparse forest*: *the* L^2 *bound.* We begin by restating the result that we need to prove:

Proposition 1. *Let $\mathcal{P} \subseteq \mathbb{P}_n$ be a sparse forest. Then there exists $\eta \in (0, 1/2)$, depending only on the degree d, such that for $1 < p < \infty$ we have*

$$\left\| T^{\mathcal{P}} \right\|_p \lesssim_{p,d} 2^{-n\,\eta\,(1-\frac{1}{p^*})}.$$

Throughout the remaining reasonings we will assume without loss of generality that

$$n \geq c(d) \geq (100d)^{100d}.$$

Assume $P = [\vec{\alpha}, I]$ and $P' = [\vec{\alpha}', I']$ with $|I| \leq |I'|$. As directed by (66) in Lemma 1 (for specific details, one is invited to consult the proof of (270) in Lemma 6), we have that the following holds:

$$|T_{P'} T_P^* f(x)| \lesssim \lceil \Delta(P, P') \rceil^{1/d} \frac{\int_{E(P)} |f|}{|I'|} \chi_{E(P')}(x). \tag{197}$$

Now, proceeding as in the corresponding proof of Proposition 1 in [81] (further inspired by the approach of Lemma 2 in [36]), we have

$$\begin{aligned}\int_{\mathbb{T}}\left|\left(T^{\mathcal{P}}\right)^{*}f(x)\right|^{2}dx &\lesssim \left|\sum_{\substack{P'\in\mathcal{P}\\ P'=[\vec{\alpha}',I']}}\int_{\mathbb{T}}f(x)\left\{\sum_{\substack{P=[\vec{\alpha},I]\in\mathcal{P}\\ |I|\leq|I'|}}\overline{T_{P'}T_{P}^{*}f}(x)\right\}dx\right|\\ &\lesssim \sum_{P'\in\mathcal{P}}\int_{E(P')}|f|\left\{\sum_{P\in a(P')}\left\lceil\Delta(P,P')\right\rceil^{1/d}\frac{\int_{E(P)}|f|}{|I'|}\right\}\\ &\quad+\sum_{P'\in\mathcal{P}}\int_{E(P')}|f|\left\{\sum_{P\in b(P')}\left\lceil\Delta(P,P')\right\rceil^{1/d}\frac{\int_{E(P)}|f|}{|I'|}\right\}\\ &\stackrel{\text{def}}{=} A\,+\,B,\end{aligned}$$

where here we have used the following notation:

$$a(P'):=\left\{P=[\vec{\alpha},I]\in\mathcal{P},\ |I|\leq|I'|\ \text{ and }\ I^{*}\cap I'^{*}\neq\emptyset\ \mid\ \Delta(P,P')\leq 2^{n\epsilon}\right\},$$

$$b(P'):=\left\{P=[\vec{\alpha},I]\in\mathcal{P},\ |I|\leq|I'|\ \text{ and }\ I^{*}\cap I'^{*}\neq\emptyset\ \mid\ \Delta(P,P')\geq 2^{n\epsilon}\right\},$$

with $\epsilon\in(0,1)$ sufficiently small (e.g., $\epsilon=\frac{1}{100\,(N+d)}$ with N defined in (106)).

Further, we have

$$A\lesssim\sum_{P'\in\mathcal{P}}\int_{E(P')}|f(x)|\left\{\frac{1}{|I'|}\sum_{P\in a(P')}\int_{E(P)}|f|\right\}dx=\int|f|\,V_a(|f|),$$

where by definition

$$V_a(f):=\sum_{P'=[\vec{\alpha}',I']\in\mathcal{P}}\frac{\chi_{E(P')}}{|I'|}\sum_{P\in a(P')}\int_{E(P)}f. \tag{198}$$

Similarly, using the definition of $b(P')$ we deduce

$$\begin{aligned}B&\lesssim\sum_{P'=[\vec{\alpha}',I']\in\mathcal{P}}\int_{E(P')}|f(x)|\left\{\frac{2^{-n\,\frac{\epsilon}{d}}}{|I'|}\sum_{P\in b(P')}\int_{E(P)}|f|\right\}dx\\ &=2^{-n\,\frac{\epsilon}{d}}\int|f|\,V_b(|f|),\end{aligned}$$

where by definition

$$V_b(f):=\sum_{P'=[\vec{\alpha}',I']\in\mathcal{P}}\frac{\chi_{E(P')}}{|I'|}\sum_{P\in b(P')}\int_{E(P)}f. \tag{199}$$

We will now focus on providing L^2-bounds on $V_a(f)$.

Fix $1<r<2$, and let r' be the Hölder conjugate of r. Suppose without loss of generality that $f\geq 0$. Then

$$V_a(f)\leq \sum_{P'=[\vec{\alpha}',I']\in\mathcal{P}} \chi_{E(P')}\left(\frac{\int_{\tilde{I}'} f^r}{|I'|}\right)^{\frac{1}{r}} \frac{\|\sum_{P\in a(P')}\chi_{E(P)}\|_{r'}}{|I'|^{\frac{1}{r'}}}.$$

The first key observation derived from the structure of the set $\mathcal{P}$ and the definition of $a(P')$ is

Claim 3. The following Carleson measure type condition holds:

$$\Big\|\sum_{P\in a(P')}\chi_{E(P)}\Big\|_{r'}\lesssim_r 2^{-\frac{n}{r'}(1-5d\epsilon-5N\epsilon)}\,|I'|^{\frac{1}{r'}}. \tag{200}$$

Here N is the parameter used in the definition of mass in (106).

As a consequence of (159) in Definition 11, it is enough to show (200) for $\mathcal{P}$ a sparse L^∞-forest.

Step 1. *A toy-model*: *The incomparable set of tiles case.*

Claim 4 (Adaptation of Claim 3). If $\mathcal{R}\subset\mathcal{P}$ is such that $\mathcal{R}$ is an *incomparable* collection of tiles, then the restriction of (200) to $\mathcal{R}$ holds; that is,

$$\Big\|\sum_{\substack{P\in a(P')\\ P\in\mathcal{R}}}\chi_{E(P)}\Big\|_{r'}\lesssim_r 2^{-\frac{n}{r'}(1-5d\epsilon-5N\epsilon)}\,|I'|^{\frac{1}{r'}}. \tag{201}$$

In order to show (201), we first claim that

$$\Big\|\sum_{\substack{P\in a(P')\\ P\in\mathcal{R}}}\chi_{E(P)}\Big\|_{1}\lesssim 2^{-n(1-5d\epsilon-5N\epsilon)}\,|I'|. \tag{202}$$

Note first that it is enough to prove (202) for $\mathcal{R}\subset\mathcal{P}\subset\mathcal{P}_n[A_n^k[A_{n-1}^{j_{n-1}},\dots,A_1^{j_1}]]$ with $a(P')\cap\mathcal{R}\neq\emptyset$. Taking now $S=A_n^k[A_{n-1}^{j_{n-1}},\dots,A_1^{j_1}]\cap 50I_{P'}$ and using the first and fourth item in Observation 15, we deduce

(203)
$$\Big\|\sum_{\substack{P\in a(P')\\ P\in\mathcal{R}}}\chi_{E(P)}\Big\|_{1}\leq\Big\|\sum_{\substack{P=[\vec{\alpha},I],\ \Delta(P,P')\leq 2^{n\epsilon}\\ I\in CZ_S(a(P')\cap\mathcal{R})}}\chi_{E(P)}\Big\|_{1}$$
$$\lesssim 2^{-n}\,2^{5\epsilon nN}\,2^{5\epsilon nd}\sum_{I\in CZ_S(a(P')\cap\mathcal{R})}|I|\lesssim 2^{-n(1-5\epsilon N-5\epsilon d)}\,|I'|.$$

The L^∞ bound follows trivially since $\mathcal{R}$ is an incomparable family of tiles:

$$\Big\|\sum_{P\in\mathcal{R}}\chi_{E(P)}\Big\|_{\infty}\leq 1. \tag{204}$$

By interpolating (or equivalently applying Hölder) between (202) and (204) we deduce that (201) holds.

Step 2. *The general sparse forest case.* By Definition 10, we have that $\mathcal{P} \cap a(P') = \bigcup_j \mathcal{P}_j$ with $\{\mathcal{P}_j\}_j$ sparse separated trees.

Further, set top $\mathcal{P}_j = P_j$ and let

$$\mathcal{P}_j^1 = \{P \in \mathcal{P}_j \,|\, \text{there is no chain } \; P < P^1 < \cdots < P^n = P_j \; \text{s.t.} \; P^k \in \mathcal{P}_j\}$$

and

$$\mathcal{P}_j^2 := \mathcal{P}_j \setminus \mathcal{P}_j^1.$$

In the above setting, by appealing to maximal chain decompositions, we notice that $\mathcal{P} \cap a(P')$ can be written as

$$\left(\bigcup_{l=1}^{n} \mathcal{A}_l\right) \cup \left(\bigcup_j \mathcal{P}_j^2\right), \tag{205}$$

such that

- each $\mathcal{A}_l$ is a set of incomparable tiles; and
- the second component satisfies

$$\sum_{j,\, \mathcal{P}_j^2 \neq \emptyset} \chi_{I_{P_j}} \leq 1. \tag{206}$$

Indeed, to see this we notice that if P_i and P_j are the tops of two separated trees such that $\mathcal{P}_i^2$, $\mathcal{P}_j^2 \neq \emptyset$, then either $I_{P_i} \cap I_{P_j} = \emptyset$ or we must have $\Delta(P_i, P_j) \gtrsim 2^n \max\{|\omega_{P_i}|,\, |\omega_{P_j}|\}$. However, only the first scenario is possible since the condition P_i, $P_j \in a(P')$ requires $\Delta(P_i, P_j) \lesssim_d 2^{\epsilon n} \max\{|\omega_{P_i}|,\, |\omega_{P_j}|\}$.

Finally, from Step 1, we know that (201) holds for each $\mathcal{R} = \mathcal{A}_l$, while from the fact that each $\mathcal{P}_j$ is a sparse tree we deduce that

$$\Big\| \sum_{P \in \mathcal{P}_j^2} \chi_{E(P)} \Big\|_{r'} \lesssim_r 2^{-n\frac{1}{r'}} \, |I_{P_j}|^{\frac{1}{r'}}. \tag{207}$$

Thus combining Step 1 with (206) and (207), we conclude that Claim 3 is true.

Now, in order to control the term A, it remains to show the following:

Claim 5. With the previous notation, defining

$$\mathcal{V}f := \sum_{P = [\vec{\alpha}, I] \in \mathcal{P}} \chi_{E(P)} \left(\frac{\int_{\tilde{I}} f^r}{|I|} \right)^{\frac{1}{r}}, \tag{208}$$

we have

$$\|\mathcal{V}f\|_2 \lesssim_r \Big\| \sum_{P \in \mathcal{P}} \chi_{E(P)} \Big\|_{\mathrm{BMO}_C} \|f\|_2 \lesssim \|f\|_2. \tag{209}$$

Now set $\mathcal{I} := \{I \,|\, \exists\, P = [\vec{\alpha}, I] \in \mathcal{P}\}$ and $E(I) := \bigcup_{\substack{P=[\vec{\alpha},I_P]\in\mathcal{P}\\ I_P=I}} E(P)$. Rewrite $\mathcal{V}$ as follows:

$$\mathcal{V}f = \sum_{I\in\mathcal{I}} \chi_{E(I)} \left(\frac{\int_{\tilde{I}} f^r}{|I|}\right)^{\frac{1}{r}}.$$

Denote $\mathcal{I}_m := \{I \in \mathcal{I} \,|\, \frac{\int_{\tilde{I}} f^r}{|I|} \approx 2^m\}$, and notice that $\mathcal{I} = \bigcup_{m\in\mathbb{Z}} \mathcal{I}_m$. Also denote by $\mathcal{I}_m^{\max}$ the set of maximal intervals (with respect of inclusion) in $\mathcal{I}_m$. Assume without loss of generality that $\int_{\mathbb{T}} f^r \approx 2^{m_0}$ for some $m_0 \in \mathbb{Z}$. Now, for each $m \geq m_0$, notice then that $\mathcal{I}_m^{\max}$ consists of pairwise disjoint intervals.

Then we have

$$\mathcal{V}f = \sum_{m\in\mathbb{Z}} \sum_{I\in\mathcal{I}_m} \chi_{E(I)} \left(\frac{\int_{\tilde{I}} f^r}{|I|}\right)^{\frac{1}{r}} \lesssim \left(\int_{\mathbb{T}} f^r\right)^{\frac{1}{r}} + \sum_{m\geq m_0} \sum_{J\in\mathcal{I}_m^{\max}} \sum_{\substack{I\subseteq J\\ I\in\mathcal{I}_m}} 2^{\frac{m}{r}} \chi_{E(I)},$$

and thus, ignoring the L^r norm of f, one has

$$\|\mathcal{V}f\|_2^2 \approx \sum_{m,m'\geq m_0} \sum_{\substack{J\in\mathcal{I}_m^{\max}\\ J'\in\mathcal{I}_{m'}^{\max}}} 2^{\frac{m+m'}{r}} \int \Big(\sum_{\substack{I\subseteq J\\ I\in\mathcal{I}_m}} \chi_{E(I)}\Big)\Big(\sum_{\substack{I'\subseteq J'\\ I'\in\mathcal{I}_{m'}}} \chi_{E(I')}\Big)$$
$$\approx \sum_{m\geq m_0} \sum_{m'\geq m} \sum_{J\in\mathcal{I}_m^{\max}} \sum_{\substack{J'\subseteq J\\ J'\in\mathcal{I}_{m'}^{\max}}} 2^{\frac{m+m'}{r}} \int \Big(\sum_{\substack{I\subseteq J\\ I\in\mathcal{I}_m}} \chi_{E(I)}\Big)\Big(\sum_{\substack{I'\subseteq J'\\ I'\in\mathcal{I}_{m'}}} \chi_{E(I')}\Big).$$

Let $1 \leq q < \infty$ and $J \subseteq [0, 1]$ fixed.

Applying the John–Nirenberg inequality to

$$\| \sum_{I\in\mathcal{I}} \chi_{E(I)} \|_{\mathrm{BMO}_D} \lesssim 1, \tag{210}$$

we deduce the Carleson packing condition

$$\| \sum_{\substack{I\subseteq J\\ I,\,J\in\mathcal{I}}} \chi_{E(I)} \|_q^q \lesssim_q |J|. \tag{211}$$

Now, from (211) and Cauchy–Schwarz, for $1 < p < r < 2$, we further have

$$\|\mathcal{V}f\|_2^2 \lesssim \sum_m \sum_{m'\geq m} 2^{\frac{m+m'}{r}} \sum_{J\in\mathcal{I}_m^{\max}} \Big\| \sum_{I\subseteq J} \chi_{E(I)} \Big\|_{p'} \Big\| \sum_{\substack{J'\subseteq J\\ J'\in\mathcal{I}_{m'}^{\max}}} \sum_{I'\subseteq J'} \chi_{E(I')} \Big\|_p$$
$$\lesssim \sum_m \sum_{J\in\mathcal{I}_m^{\max}} 2^{\frac{m}{r}} |J|^{\frac{1}{p'}} \sum_{m'\geq m} 2^{\frac{m'}{r}} \Big(\sum_{\substack{J'\subseteq J\\ J'\in\mathcal{I}_{m'}^{\max}}} |J'| \Big)^{\frac{1}{p}}$$
$$\lesssim \sum_m \sum_{J\in\mathcal{I}_m^{\max}} 2^{\frac{m}{r}} |J|^{\frac{1}{p'}} \sum_{m'\geq m} 2^{\frac{m'}{r}} 2^{-\frac{m'}{p}} \Big(\int_{\tilde{J}} f^r \Big)^{\frac{1}{p}}$$

$$\lesssim \sum_m \sum_{J\in\mathcal{I}_m^{\max}} 2^{\frac{2m}{r}}\,|J| \lesssim \sum_m 2^{\frac{2m}{r}}\,2^{-m}\int_{(M_r f)^r\gtrsim 2^m}(M_r f)^r$$
$$\lesssim \int (M_r f)^2 \lesssim_r \int f^2,$$

where here we denoted $M_r f(x) := \left(\sup_{x\in I}\frac{\int_{\tilde{I}}|f|^r}{|I|}\right)^{\frac{1}{r}}$.

This ends the proof of Claim 5. □

Thus, now combining (200) and (209), for an appropriate choice of ϵ, we conclude that

$$A \lesssim_{r,d} 2^{-\frac{n}{r'}(1-5d\epsilon-5N\epsilon)}\,\|f\|_2^2 \lesssim 2^{-\frac{n}{2r'}}\,\|f\|_2^2.$$

The B term can be similarly treated if one replaces (200) with just

(212) $$\|\sum_{P\in b(P')}\chi_{E(P)}\|_{r'} \lesssim_r |I'|^{\frac{1}{r'}},$$

thus obtaining

$$B \lesssim_d 2^{-n\,\frac{\epsilon}{d}}\,\|f\|_2^2.$$

Now, properly choosing r and ϵ, we conclude that there exists $\eta=\eta(d)\in(0,1)$ such that

(213) $$\left\|T^{\mathcal{P}}f\right\|_2 \lesssim_d 2^{-n\,\frac{\eta}{2}}\,\|f\|_2.$$

This ends our proof of Proposition 1 for $p=2$. □

8.3. *Sparse forest: the L^p bound*, $1<2\neq p<\infty$. Suppose first that $2<p<\infty$. For any $f\in L^1(\mathbb{T})$, we define the operator

(214) $$L^{\mathcal{P}}f(x) := \sum_{P=[\vec{\alpha},I_P]\in\mathcal{P}}\frac{\int_{I_{P^*}}f}{|I_P|}\chi_{E(P)}.$$

Now, on the one hand, repeating the reasonings from the case $p=2$, one has

$$\|L^{\mathcal{P}}f\|_2 \lesssim_d \|f\|_2.$$

On the other hand,

$$\left\|L^{\mathcal{P}}\right\|_{\infty\to\mathrm{BMO}_D} \lesssim 1.$$

Interpolating now between the $L^2\to L^2$ and $L^\infty\to\mathrm{BMO}_D$ estimates, we obtain that

(215) $$\|L^{\mathcal{P}}\|_p \lesssim_{p,d} 1.$$

Consequently, based on the straightforward relation

$$|T^{\mathcal{P}}f| \lesssim L^{\mathcal{P}}|f|,$$

we also get that for any $2\le p<\infty$, one has

(216) $$\|T^{\mathcal{P}}\|_p \lesssim_{p,d} 1.$$

Interpolating now between (213) and (216) one obtains the desired conclusion (possibly by changing the exponent η by a small factor).

For the case $1<p<2$, we need to focus on the behavior of $T^{\mathcal{P}*}$.

Indeed, on the one hand, we know that

$$\left\|T^{\mathcal{P}*}\right\|_{2\to 2}=\left\|T^{\mathcal{P}}\right\|_{2\to 2}\lesssim_d 2^{-n\frac{\eta}{2}}.$$

On the other hand, for $f\in L^{\infty}$, we have

$$\|L^{\mathcal{P}*}f\|_{\mathrm{BMO}_D}=\Big\|\sum_{P=[\vec{\alpha},I_P]\in\mathcal{P}}\frac{\int_{E(P)}f}{|I_P|}\chi_{I_{P^*}}\Big\|_{\mathrm{BMO}_D}\lesssim\|f\|_{\infty}\,.$$

Thus, as before, for any $2\le q=p'<\infty$, one has

$$\|T^{\mathcal{P}*}f\|_q\lesssim\|L^{\mathcal{P}*}|f|\|_q\lesssim_{q,d}1.$$

The claim now follows by interpolation.[59] □

9. Preparations for the proof of Proposition 2

This is one of the most technical sections of the paper, and, as the name suggests, is meant for "preparing the ground" for the proof of Proposition 2. It splits in three subsections, with several results therein being in close analogy with the similar ones in [36] and/or [81].

9.1. *Single tree estimates.* In this section, we prove two local L^p estimates addressing the tree structure.

Lemma 2 (L^p-uniform mass tree estimate). *Fix $\delta\in(0,1]$, and let $\mathcal{P}\subseteq\mathbb{P}$ be a tree with spatial support I_0 such that*

$$A_{\mathbb{P},I_0}(P)\le\delta\quad\forall\quad P\in\mathcal{P}. \tag{217}$$

Then, for $1<p<\infty$, we have

$$\left\|T^{\mathcal{P}}\right\|_p\lesssim_{p,d}\delta^{\frac{1}{p}}. \tag{218}$$

Observation 16. The L^p-tree lemma above follows directly from the lemma below. Indeed, with the notation in Lemma 3, by simply setting $A=[0,1]$, one has that (217) immediately implies

$$\frac{E_A^{\mathcal{P}}(J)}{|J|}\lesssim\delta\qquad\forall\,J\in CZ_{I_0}(\mathcal{P}). \tag{219}$$

[59] We use here the fact that $\left\|T^{\mathcal{P}*}\right\|_{p'\to p'}=\left\|T^{\mathcal{P}}\right\|_{p\to p}$.

LEMMA 3 (Spatially localized L^p-tree estimates). *Let $\mathcal{P}$ be a tree with top $P_0=[\vec{\alpha_0}, I_0]$, and let $A\subseteq[0,1]$ be a measurable set. Define*

$$E_A^{\mathcal{P}}(J):=\bigcup_{P\in\mathcal{P}}(A\cap E(P)\cap J)\qquad\forall\, J\in CZ_{I_0}(\mathcal{P}). \tag{220}$$

Then, for $1<p<\infty$, we have

$$\left\|\chi_A\, T^{\mathcal{P}} f\right\|_p\lesssim_{p,d}\left(\sup_{J\in CZ_{I_0}(\mathcal{P})}\frac{|E_A^{\mathcal{P}}(J)|}{|J|}\right)^{\frac{1}{p}}\|f\|_p\,. \tag{221}$$

Proof. The proof of this result is in the same spirit as the L^2-tree estimate lemma provided in [36].

Let $\underline{q}_0\in\mathcal{Q}_{d-1}$ be the central polynomial of P_0, and let $\underline{Q}_0\in\mathcal{Q}_d$ be the unique polynomial with $\frac{d}{dx}\,\underline{Q}_0(x)=\underline{q}_0$ and $\underline{Q}_0(0)=0$. Assume without loss of generality that

$$\underline{Q}_0(y)=\sum_{j=1}^{d}a_j^0\,y^j. \tag{222}$$

Appealing to generalized modulation symmetries, we make use of a suitable transformation with the aim of moving our discussion near the real axis of the time-frequency plane; that is, we define

$$\mathcal{T}^{\mathcal{P}}:=\left(\prod_{j=1}^{d}M^*_{j,a_j^0}\right)T^{\mathcal{P}}\left(\prod_{j=1}^{d}M_{j,a_j^0}\right) \tag{223}$$

and $g(x):=\prod_{j=1}^{d}M^*_{j,a_j^0}f(x)$, and we notice that

$$\left\|\chi_A\, T^{\mathcal{P}} f\right\|_p=\left\|\chi_A\,\mathcal{T}^{\mathcal{P}} g\right\|_p.$$

For a fixed $x\in\mathbb{T}$, we further define[60]

$$k_0(x):=\inf\{k\in D\,\mathbb{N}\,|\,\exists\, P\in\mathcal{P}\ \text{s.t.}\ |I_P|=2^{-k}\ \&\ \chi_{E(P)}(x)\neq 0\},$$
$$k_1(x):=\sup\{k\in D\,\mathbb{N}\,|\,\exists\, P\in\mathcal{P}\ \text{s.t.}\ |I_P|=2^{-k}\ \&\ \chi_{E(P)}(x)\neq 0\}$$

and notice that

$$\mathcal{T}^{\mathcal{P}}g(x):=\sum_{\substack{k=k_0(x)\\ k\in D\,\mathbb{N}}}^{k_1(x)}\int_{\mathbb{T}}\psi_k(y)\,e^{i\,(\int_{x-y}^{x}q_x-\int_{x-y}^{x}\underline{q}_0)}\,g(x-y)\,dy. \tag{224}$$

[60] Based on Observation 2, we can assume without loss of generality that $\mathcal{P}\subset\bigcup_{k\in\mathbb{N}}\mathbb{P}_{kD}$ finite.

Observation 17. Relation (224) above is the key place where we use the convexity of the tree and thus the very reason for which we need to remove the possible boundary effect by requiring item (2) in Definition 8.

With this, we notice that

$$|\chi_A(x)\,\mathcal{T}^{\mathcal{P}}g(x)| \leq \chi_A(x)\sum_{\substack{k=k_0(x)\\ k\in D\,\mathbb{N}}}^{k_1(x)}\int_{\mathbb{T}}|\psi_k(y)|\,|e^{i\int_{x-y}^{x}(q_x-\underline{q}_0)}-1|\,|g(x-y)|\,dy$$
$$+\,\chi_A(x)\left|\sum_{\substack{k=k_0(x)\\ k\in D\,\mathbb{N}}}^{k_1(x)}\int_{\mathbb{T}}\psi_k(y)\,g(x-y)\,dy\right| =: \mathcal{A}(x)\,+\,\mathcal{B}(x).$$

At the heuristic level the $\mathcal{A}$-term should be thought as an error term, while the $\mathcal{B}$-term as a variant of a maximal Hilbert transform. With these we pass now to the actual estimates for the two terms.

For the first term, applying Lemma C in the appendix, we deduce

$$\begin{aligned} |\mathcal{A}(x)| &\lesssim \chi_A(x)\,\|q_x-\underline{q}_0\|_{L^\infty(x+[-2^{-k_0(x)+5},\,2^{-k_0(x)+5}])}\\ &\quad\times\sum_{\substack{k=k_0(x)\\ k\in D\,\mathbb{N}}}^{k_1(x)}\int_{\mathbb{T}}|y\,\psi_k(y)|\,|g(x-y)|\,dy \\ &\lesssim_d \chi_A(x)\,2^{k_0(x)}\int_{x-2^{-k_0(x)+5}}^{x+2^{-k_0(x)+5}}|g(y)|\,dy. \end{aligned} \tag{225}$$

Now defining

$$M_A^{\mathcal{P}}g(x) := \begin{cases}\sup_{I\supset J}\frac{1}{|I|}\int_{100I}|g| & \text{if } x\in E_A^{\mathcal{P}}(J) \text{ and } J\in CZ_{I_0}(\mathcal{P}),\\ 0 & \text{otherwise,}\end{cases} \tag{226}$$

we immediately deduce

$$\mathcal{A}(x)\lesssim_d M_A^{\mathcal{P}}g(x). \tag{227}$$

For the second term, we proceed as follows: first we define

$$\mathcal{H}(y) := \sum_{k\in D\mathbb{N}}\psi_k(y)$$

and, for any given fixed $K\in\mathbb{N}$, set

$$\mathcal{H}_K(y) := \begin{cases} 2^{K+9}\int_{-2^{-K-10}}^{2^{-K-10}}\mathcal{H}(y-s)\,ds & \text{if } |y|\geq 2^{-K},\\ \\ 0 \quad \text{otherwise.}\end{cases}. \tag{228}$$

With these, we have the following control over the K-truncated sums:

$$\Big|\sum_{\substack{k\in D\mathbb{N}\\ k\le K}}\psi_k(y)-\mathcal{H}_K(y)\Big|\lesssim\frac{2^{-K}}{y^2+2^{-2K}}\quad\forall\,y\in\mathbb{T}. \tag{229}$$

Now using (229), we deduce that

$$\begin{aligned}\mathcal{B}(x)\lesssim\chi_A(x)\sup_{K\ge k_0(x)}2^K\int_{2^{-K}}^{2^{-K}}|\mathcal{H}*g(x-s)|\,ds\\ +\chi_A(x)\sup_{K\ge k_0(x)}2^K\int_{-2^{-K}}^{2^{-K}}|g(x-y)|\,dy.\end{aligned} \tag{230}$$

Now combining (226) with (230), we have that

$$\mathcal{B}(x)\lesssim M_A^{\mathcal{P}}(\mathcal{H}*g)(x)\,+\,M_A^{\mathcal{P}}(g)(x). \tag{231}$$

Putting together (227) with (231), it only remains to notice that

$$\left\|M_A^{\mathcal{P}}g\right\|_p\lesssim\left(\sup_{J\in CZ_{I_0}(\mathcal{P})}\frac{|E_A^{\mathcal{P}}(J)|}{|J|}\right)^{\frac{1}{p}}\|Mg\|_p\lesssim_p\left(\sup_{J\in CZ_{I_0}(\mathcal{P})}\frac{|E_A^{\mathcal{P}}(J)|}{|J|}\right)^{\frac{1}{p}}\|g\|_p \tag{232}$$

and

$$\|\mathcal{H}*g\|_p\lesssim_p\|g\|_p\,. \tag{233}$$

Thus, we conclude that (221) holds. □

Lemma 4 (Spatially localized L^p-adjoint tree estimates). *Let $\mathcal{P}$ be a tree with spatial support I_0. Assume we are given $A\subseteq[0,1]$ a measurable set and for $J\in CZ_{20I_0}(\mathcal{P}^*)$, we define*[61]

$$E_A^{\mathcal{P}^*}(J):=A\cap J. \tag{234}$$

Then, for $1<p<\infty$, we have[62]

$$\left\|\chi_A T^{\mathcal{P}^*}f\right\|_p\lesssim_{p,d}\left(\sup_{J\in CZ_{20I_0}(\mathcal{P}^*)}\frac{|E_A^{\mathcal{P}^*}(J)|}{|J|}\right)^{\frac{1}{p}}\sup_{P\in\mathcal{P}}A_{\mathbb{P},I_0}(P)^{\frac{1}{p'}}\;\|f\|_p\,. \tag{235}$$

[61]This lemma remains true if instead of $CZ_{20I_0}(\mathcal{P}^*)$ we use a standard Calderón–Zygmund decomposition of the support of $T^{\mathcal{P}^*}$ with respect to the minimal intervals relative to inclusion belonging to $\mathcal{I}_{\mathcal{P}^*}$.

[62]Notice that if $\mathcal{P}$ is a normal tree, then the conclusion of our lemma holds for $CZ_{20I_0}(\mathcal{P}^*)$ replaced by $CZ_{I_0}(\mathcal{P}^*)$ in (235).

Proof. We first notice that by applying the same reasonings as in the previous proof, specifically relying on (223), one can assume without loss of generality that our tree lives at frequency zero, or in other words that $Q_0 \equiv 0$ in (222).[63]

Now fix $x \in \mathbb{T}$ and notice that with the previous notation one has

$$\chi_A T^{\mathcal{P}^*} f(x) = \sum_{J \in CZ_{20 I_0}(\mathcal{I}_{\mathcal{P}^*})} \chi_{J \cap A} \left\{ \sum_{P \in \mathcal{P}} T_P{}^* f(x) \right\}. \tag{236}$$

Next, fixing $J \in CZ_{20 I_0}(\mathcal{P}^*)$ and assuming without loss of generality that $x \in J$, we deduce

$$\begin{aligned}
&\left| T^{\mathcal{P}^*} f(x) - \frac{1}{|J|} \int_J T^{\mathcal{P}^*} f(s) ds \right| \\
&\quad = \left| \frac{1}{|J|} \int_J \left\{ \sum_{\substack{P \in \mathcal{P} \\ 2^{-k} = |I_P| \geq |J|}} \int_{\mathbb{T}} \left[\varphi_k(x-y) - \varphi_k(s-y) \right] f(y) \chi_{E(P)}(y) dy \right\} ds \right| \\
&\lesssim \sum_{\substack{P \in \mathcal{P} \\ 2^{-k} = |I_P| \\ I_{P^*} \supseteq J}} 2^k \, |J| \, \frac{\int_{E(P)} |f|}{|I_P|}.
\end{aligned} \tag{237}$$

Thus, from (237), we deduce that

$$\begin{aligned}
&\sum_{J \in CZ_{20 I_0}(\mathcal{P}^*)} \chi_{J \cap A} \left| T^{\mathcal{P}^*} f(x) - \frac{1}{|J|} \int_J T^{\mathcal{P}^*} f(s) ds \right| \\
&\qquad \lesssim \sum_{J \in CZ_{20 I_0}(\mathcal{P}^*)} \chi_{J \cap A} \sum_{\substack{P \in \mathcal{P} \\ I_{P^*} \supseteq J}} \frac{|J|}{|I_P|} \frac{\int_{E(P)} |f|}{|I_P|}.
\end{aligned} \tag{238}$$

Denote

$$M_{\mathcal{P}^*} f(x) := \sum_{J \in CZ_{20 I_0}(\mathcal{P}^*)} \chi_J(x) \sup_{I \supseteq J} \frac{1}{|I|} \int_I |f|(s) ds. \tag{239}$$

---

[63] Alternatively one can apply the reasonings in our proof to the operator $\chi_A \mathcal{T}^{\mathcal{P}^*}$, where here $\mathcal{T}^{\mathcal{P}}$ is defined by (223).

Deduce from (238) and (239) that

(240)

$$\|\chi_A T^{\mathcal{P}^*} f\|_p \lesssim \left(\sup_{J\in CZ_{20I_0}(\mathcal{P}^*)} \frac{|E_A^{\mathcal{P}^*}(J)|}{|J|}\right)^{\frac{1}{p}} \|M_{\mathcal{P}^*}(T^{\mathcal{P}^*} f)\|_p$$

$$+ \left(\sup_{J\in CZ_{20I_0}(\mathcal{P}^*)} \frac{|E_A^{\mathcal{P}^*}(J)|}{|J|}\right)^{\frac{1}{p}} \left\|\sum_{J\in CZ_{20I_0}(\mathcal{P}^*)} \chi_J \sum_{\substack{P\in\mathcal{P}\\ I_{P^*}\supseteq J}} \frac{|J|}{|I_P|} \frac{\int_{E(P)}|f|}{|I_P|}\right\|_p .$$

Now we notice that

$$\left\|\sum_{J\in CZ_{20I_0}(\mathcal{P}^*)} \chi_J \sum_{\substack{P\in\mathcal{P}\\ I_{P^*}\supseteq J}} \frac{|J|}{|I_P|} \frac{\int_{E(P)}|f|}{|I_P|}\right\|_{1,\infty} \lesssim \|Mf\|_{1,\infty} \lesssim \|f\|_1 \tag{241}$$

and

$$\left\|\sum_{J\in CZ_{20I_0}(\mathcal{P}^*)} \chi_J \sum_{\substack{P\in\mathcal{P}\\ I_{P^*}\supseteq J}} \frac{|J|}{|I_P|} \frac{\int_{E(P)}|f|}{|I_P|}\right\|_{\infty} \lesssim \left(\sup_{P\in\mathcal{P}} A_0(P)\right) \|f\|_\infty . \tag{242}$$

Interpolating now between (241) and (242) we get that for $1 < p < \infty$,

$$\left\|\sum_{J\in CZ_{20I_0}(\mathcal{P}^*)} \chi_J \sum_{\substack{P\in\mathcal{P}\\ I_{P^*}\supseteq J}} \frac{|J|}{|I_P|} \frac{\int_{E(P)}|f|}{|I_P|}\right\|_{p} \lesssim \left(\sup_{P\in\mathcal{P}} A_0(P)\right)^{\frac{1}{p'}} \|f\|_p . \tag{243}$$

Also, we trivially have that

(244)

$$\|M_{\mathcal{P}^*}(T^{\mathcal{P}^*} f)\|_p \lesssim \|M(T^{\mathcal{P}^*} f)\|_p \lesssim_p \|T^{\mathcal{P}^*} f\|_p \lesssim_{p,d} \sup_{P\in\mathcal{P}} A_{\mathbb{P},I_0}(P)^{\frac{1}{p'}} \|f\|_p .$$

Now combining (240), (243) and (244), we conclude that (235) holds. □

9.2. *Interaction between two tiles: a refined version.* In this section we develop the discussion in Section 4 with the aim of obtaining a more refined estimate of the generic interaction

$$\left\langle T_{P_1}^* f, T_{P_2}^* f\right\rangle . \tag{245}$$

Indeed, assuming without loss of generality that $I_{P_1}^* \cap I_{P_2}^* \neq \emptyset$ and that $\Delta(P_1, P_2) \gg_d 1$, we will show that, given any $\epsilon_0 \in (0,1)$, there exists a set $I_{12} \subset \tilde{I}_{P_1} \cap \tilde{I}_{P_2}$ of small relative measure, i.e.,

$$\frac{|I_{12}|}{|\tilde{I}_{P_1} \cap \tilde{I}_{P_2}|} \lesssim_d \lceil \Delta(P_1, P_2)\rceil^{\frac{1-\epsilon_0}{d}}$$

such that for any $n \in \mathbb{N}$, one has

$$|\langle T^*_{P_1} f, T^*_{P_2} f\rangle| \lesssim_{n,d,\epsilon_0} \lceil \Delta(P_1,P_2)\rceil^n A_0(P_1)^{\frac{1}{2}} A_0(P_2)^{\frac{1}{2}} \|f\|^2_{L^2(I_{P_1}\cup I_{P_2})} + \|T^*_{P_1} f\|_{L^2(I_{12})} \|T^*_{P_2} f\|_{L^2(I_{12})}. \tag{246}$$

This relation may be seen as a refined version of Van der Corput lemma and will further serve as a key model for the discussion in the next section on the interaction between two separated trees/rows.

9.2.1. *Spatial Calderón–Zygmund decompositions adapted to a polynomial.* In this subsection we develop a general algorithm for partitioning a given interval $J \subset \mathbb{T}$ into a union of dyadic intervals having suitable, "good" properties relative to a given polynomial $q \in \mathcal{Q}_{d-1}$. This decomposition will be quintessential in the proof of (246). Our precise statement and description of the algorithm is given below.

LEMMA 5 (q-"good" decomposition of an interval J). *Let $J \subset \mathbb{T}$ be an interval that can be decomposed into a finite union of dyadic intervals $\bigcup_m J^m$ with each $|J^m| \geq \frac{|J|}{100}$. Also let $q \in \mathcal{Q}_{d-1}$ with $d \in \mathbb{N}$, $d \geq 2$ be a polynomial such that*

$$q \notin \mathcal{Q}_0 \tag{247}$$

and

$$0 < \lambda \leq \Delta_q(J). \tag{248}$$

Then there exist a partition

$$J = J_s(q,\lambda) \cup J_l(q,\lambda) \tag{249}$$

and $c_1(d)$, $c_2(d) > 0$ such that

- *the "(q,λ)-small" component $J_s(q,\lambda)$ can be written as a union of at most $9d$ dyadic intervals having the same length*[64]

$$w(J,q,\lambda) := c_1(d)\, \lambda^{\frac{1}{d}}\, \Delta_q(J)^{-\frac{1}{d}}\, |J|, \tag{250}$$

  and hence

$$|J_s(q,\lambda)| \leq 9\, d\, w(J,q,\lambda); \tag{251}$$

- *defining*

$$\eta(J,q,\lambda) := c_2(d)\, \lambda^{\frac{d-1}{d}}\, \Delta_q(J)^{\frac{1}{d}}\, |J|^{-1}, \tag{252}$$

  one has

$$\{x \in J \,|\, |q(x)| < \eta(J,q,\lambda)\} \subseteq J_s(q,\lambda); \tag{253}$$

[64]Throughout this section, our choice of $c_1(d) \gg c_2(d)$ will be made such that the quantities $w(J,q,\lambda)$ and $\eta(J,q,\lambda)$ are dyadic numbers.

- *the "$(q,\,\lambda)$-large" component $J_l(q,\lambda)$ can be itself partitioned into finitely many dyadic intervals*

$$J_l(q,\lambda)=\bigcup_{W\in CZ_{(q,\lambda)}(J)} W, \tag{254}$$

where here we define $CZ_{(q,\lambda)}(J)$ as the $(q,\,\lambda)$-Calderón–Zygmund decomposition of J, that is, the standard Calderón–Zygmund interval decomposition[65] *of the set J relative to the set $J_s(q,\lambda)$ from which we excise the intervals belonging to $J_s(q,\lambda)$;*

- *for each $W\in CZ_{(q,\lambda)}(J)$, the following key properties hold:*[66]

$$\inf_{x\in W}|q(x)|\gtrsim_d \sup_{x\in W}|q(x)|\gtrsim_d \eta(J,q,\lambda), \tag{255}$$

$$|W|\geq c(d)\,w(J,q,\lambda), \tag{256}$$

$$\Delta_q(W)\geq c(d)\,\lambda, \tag{257}$$

and

$$\left\|\frac{q^{(s)}}{q}\right\|_{L^\infty(W)}\leq c(d)\,\frac{1}{|W|^s}\qquad \forall\, s\in\{0,\ldots,\,d-1\}. \tag{258}$$

Proof. Let us first define

$$\mathcal{M}_q(J):=\{x\in J\mid x \text{ is a local minimum for } |q|\}.$$

From (247) we can assume that $\mathcal{M}_q(J)=\{x^m\}_m$ is finite.[67] Notice that the cardinal of $\mathcal{M}_q(J)$ obeys

$$r:=\#\mathcal{M}_q(J)\leq 3\,d. \tag{259}$$

With the previous notation, we define the $(q,\,\lambda)$-small component of J_s as given by

$$J_s(q,\lambda):=\bigcup_{j=1}^{l} I^j, \tag{260}$$

[65]We recall here that given a collection $\mathcal{A}$ of disjoint dyadic intervals inside a given (finite union of) dyadic interval(s) $S\subseteq[0,1]$, we refer to the standard Calderón–Zygmund decomposition of the interval S with respect to $\mathcal{A}$ as the collection $\mathcal{A}\cup\mathcal{B}$, where here $\mathcal{B}$ is the set of all dyadic intervals contained in S obeying the following: (1) $\mathcal{A}\cup\mathcal{B}$ forms a partition of S; (2) for any $I\in\mathcal{A}$ and $J\in\mathcal{B}$, one has $2I\not\supseteq J$ and $2J\not\supseteq I$; (3) $\mathcal{B}$ is a collection of maximal dyadic intervals obeying (1) and (2).

[66]Recall that throughout the paper the constant $c(d)>0$ is allowed to change from line to line.

[67]In particular, the graph of $|q|$ is not a straight line parallel with the real axis, as otherwise the above lemma is trivial.

where the family of dyadic intervals $\{I^j\}_{j\in\{1,\ldots,l\}}$ forms a (maximal) covering of $\mathcal{M}_q(J)$ with the following properties:

- $|I^j| = w(J,q,\lambda) \quad \forall\, j\in\{1,\ldots,l\}$;
- $3\,I^j\cap\mathcal{M}_q(J)\neq\emptyset$.

Observe here that based on (259) and the definition of $J_s(q,\lambda)$ one has $l\leq 9\,d$. Also, from our hypothesis about J, for a proper choice of $c_1(d)$, we have that

$$\text{either } \operatorname{dist}(J_s(q,\lambda),\,\partial J)=0 \text{ or } \operatorname{dist}(J_s(q,\lambda),\,\partial J)\geq w(J,q,\lambda). \tag{261}$$

Next, setting

$$\mathcal{L}_q^{\eta(J,q,\lambda)}(J):=\{x\in J\,|\ |q|(x)<\eta(J,q,\lambda)\},$$

we apply Lemma B (see the appendix section below) with $I=J$ and $\eta=\eta(J,q,\lambda)$, and together with (250) and (252) (for an appropriate choice of $c_2(d)$ in (252)) we deduce

$$|\mathcal{L}_q^{\eta(J,q,\lambda)}(J)|\leq w(J,q,\lambda), \tag{262}$$

thus proving property (253).

We now pass to the analysis of the $(q,\,\lambda)$-large component $J_l(q,\lambda)$.

First, we notice that based on observation (261) definition (254) makes sense.

Next, from definitions (260) and (254) we notice that given any $W\in CZ_{(q,\lambda)}(J)$, the following hold:[68]

- There exist unique consecutive points $x^m,\,x^{m+1}\in\mathcal{M}_q(J)$ and $x^m<x^{m+1}$ such that

$$W=:[a,b]\subseteq[x^m+w(J,q,\lambda),\,x^{m+1}-w(J,q,\lambda)]. \tag{263}$$

- The interval $[x^m,\,x^{m+1}]$ can be decomposed into two intervals $L_1:=[x^m,\,y^m]$ and $L_2:=[y^m,\,x^{m+1}]$ such that

$$\begin{aligned}&|q| \text{ restricted to } L_1 \text{ is monotone increasing,}\\ &|q| \text{ restricted to } L_2 \text{ is monotone decreasing.}\end{aligned} \tag{264}$$

Now, from (264) we further deduce that

$$\inf_{x\in W}|q(x)|=\min\{|q|(a),\,|q|(b)\}. \tag{265}$$

[68] Below, we assume without loss of generality that $W\subset[x^1,\,x^r]$ and that for any m one has $|x^{m+1}-x^m|\geq 10\,w(J,q,\lambda)$. Otherwise, the required adaptations are quite straightforward.

Assume without loss of generality that $\inf_{x\in W}|q(x)| = |q|(a)$. Then letting $R_1 := [x^m,\, a]$ and $R_2 := [x^m,\, b]$, we have

(266)
- $\|q\|_{L^\infty(R_1)} = |q|(a) = \inf_{x\in W}|q(x)|$;
- $\|q\|_{L^\infty(R_2)} = \|q\|_{L^\infty(W)}$; and
- $1 \le \frac{|R_2|}{|R_1|} \le 5$.

Now (255) follows from (266) and an application of Lemma A in the appendix.

Relation (256) follows directly from the definition of the Calderón–Zygmund decomposition $CZ_{(q,\lambda)}(J)$.

Next, (257) follows from

$$\Delta_q(W) = \frac{\operatorname{dist}^W(q,0)}{|W|^{-1}} = \frac{\operatorname{dist}^{R_2}(q,0)}{|W|^{-1}} \ge \frac{1}{5}\,\Delta_q(R_2)$$
$$\ge \frac{1}{5}\,\Delta_q([x^m,\, x^m + w(J,q,\lambda)]) \ge \frac{\eta(J,q,\lambda)}{5\,w(J,q,\lambda)^{-1}} = c(d)\,\lambda.$$

Finally, (258) is a direct consequence of (255) and the Lagrange interpolation formula applied to $I := W$; see (390) in the proof of Lemma A in the appendix. □

9.2.2. *The tile-interaction lemma.* We conclude this subsection by describing how the concepts and definitions introduced in Section 9.2.1 and in Section 4 merge into providing a precise description of (246).

In what follows, we will only consider the non-trivial case $I^*_{P_1} \cap I^*_{P_2} \neq \emptyset$; also, throughout this section, for notational simplicity, we simply set $I_{P_1} = I_1$, $I_{P_2} = I_2$ and we suppose without loss of generality that $|I_1| \ge |I_2|$.

Definition 12 (*Critical intersection set*). Now let $\epsilon_0 \in (0,1)$. With the notation and conventions from Lemma 5, we define the (ϵ_0-)*critical intersection set* $I_{1,2}$ of the pair (P_1, P_2) as

$$I_{1,2} := J_s(q_{1,2}, \lambda) \tag{267}$$

for the particular choices
- $J := \tilde{I}_1 \cap \tilde{I}_2$;
- $\lambda := \Delta(P_1,P_2)\,\lceil \Delta(P_1,P_2)\rceil^{1-\epsilon_0}$.

Now using Lemma 5 together with the principle of (non-)stationary phase, one deduces the following:

Lemma 6 (Tile interaction control: refined version). *Let* $P_1,\ P_2 \in \mathbb{P}$. *Then, with the above notation and conventions, we have*

$$\left|\int \tilde{\chi}_{I^c_{1,2}} T^*_{P_1} f\, \overline{T^*_{P_2} g}\right| \lesssim_{n,\,d,\,\epsilon_0} \lceil \Delta(P_1,P_2)\rceil^{n}\, \frac{\int_{E(P_1)}|f| \int_{E(P_2)}|g|}{\max\left(|I_1|, |I_2|\right)} \quad \forall\, n \in \mathbb{N}, \tag{268}$$

$$\int_{I_{1,2}} |T^*_{P_1} f\, \overline{T^*_{P_2} g}| \lesssim_{d,\epsilon_0} \lceil\Delta(P_1,P_2)\rceil^{\frac{1-\epsilon_0}{d}} \frac{\int_{E(P_1)}|f| \int_{E(P_2)}|g|}{\max(|I_1|,|I_2|)}, \tag{269}$$

where $\tilde{\chi}_{I^c_{1,2}}$ *is a smooth variant of the corresponding cut-off.*

Moreover, we also have

$$\left\|T_{P_1} T^*_{P_2}\right\|_2^2 \lesssim_d \min\left\{\frac{|I_2|}{|I_1|}, \frac{|I_1|}{|I_2|}\right\} \lceil\Delta(P_1,P_2)\rceil^{\frac{2}{d}} A_0(P_1)\, A_0(P_2). \tag{270}$$

Proof. Assume throughout the proof that $\Delta(P_1,P_2) \gg_d 1$, as otherwise the above statements are trivial.

Next, notice that relation (269) is straightforward based on (251), Definition 12, and on the fact that

$$|T^*_{P_j} f| \lesssim \frac{\int_{E(P_j)}|f|}{|I_j|}\chi_{I^*_j} \qquad \forall\ j\in\{1,2\},$$

which in turn is a consequence of (57).

We now turn our attention towards (268).

Apply the algorithm described in Lemma 5 for the following parameters: $J=\tilde{I}_1\cap\tilde{I}_2$ and $\lambda := \Delta(P_1,P_2)^{\epsilon_0}$. We then obtain the collection $CZ_{(q,\lambda)}(J) := \{W_r\}_r$ representing the $(q,\,\lambda)$-Calderón–Zygmund decomposition of $J\setminus J_s(q,\lambda)$ relative to the set $J_s(q,\lambda)$.

Let φ be a smooth cutoff of $\chi_{I^c_{1,2}}$ such that $\varphi\geq 0$ and

$$\varphi|_{J\setminus I^c_{1,2}} = 1 \text{ and } \varphi|_{\frac{3}{4}I_{1,2}\cup(\frac{5}{4}J)^c} = 0. \tag{271}$$

Now take any smooth partition of unity adapted to the collection $CZ_{(q,\lambda)}(J)$ that is identically zero on the set $\frac{3}{4}I_{1,2}\cup(\frac{5}{4}J)^c$. Thus without loss of generality we may assume that

$$\varphi = \sum_r \varphi_{W_r}, \tag{272}$$

where here

(273)
- $\varphi_{W_r}\in C^\infty_0$ is adapted to W_r with $0\leq\varphi_{W_r}\leq 1$;
- $\|\varphi_{W_r}\|_{C^s}\lesssim |W_r|^{-s}$ for any $s\in\mathbb{N}$;
- $\varphi_{W_r}=1$ on W_r; and
- $\varphi_{W_r}=0$ on $[0,1]\setminus\frac{5}{4}W_r$.

Observation 18. It is important to notice that for appropriate choices of the d-dependent constants in Lemma 5, and based on the results in the appendix, we have that the properties of the interaction polynomial $q_{1,2}$ on each of the W_r (see (255)–(258)) are transferable with no modifications (up to further d-dependent constants) to any difference polynomial of the form q_1-q_2 with $q_1\in P_1$ and $q_2\in P_2$.

With this we have

$$\int \varphi\, T_{P_1}^{*} f\, \overline{T_{P_2}^{*} g} = \int f\, \overline{T_{P_1}(\varphi\, T_{P_2}^{*} g)}$$
$$= \int\int (f\chi_{E(P_1)})(x)\,(\bar{g}\chi_{E(P_2)})(s)\,\mathcal{K}(x,s)\,dx\,ds,$$

where

$$\mathcal{K}(x,s) := \int e^{i\,[\int_y^s q_s - \int_y^x q_x]}\ \psi_{k_1}(x-y)\,\varphi(y)\,\psi_{k_2}(s-y)\,dy. \tag{274}$$

Here we have used the convention $|I_1| = 2^{-k_1}$, $|I_2| = 2^{-k_2}$ with $k_2 \geq k_1$ positive integers.

Let us set $\tilde{Q}(y) := \int_y^s q_s - \int_y^x q_x$, $\tilde{q} = \tilde{Q}'$ and $u(y) := \psi_{k_1}(x-y)\,\psi_{k_2}(s-y)$. Then, writing $e^{i\,\tilde{Q}(y)} = \left(\frac{1}{i\,\tilde{q}(y)}\,\frac{d}{dy}\right)(e^{i\,\tilde{Q}(y)})$ and integrating by parts n times in the expression

$$\mathcal{K}(x,s) = \int \left[\left(\frac{1}{i\,\tilde{q}(y)}\,\frac{d}{dy}\right)^n (e^{i\,\tilde{Q}(y)})\right]\ \varphi(y)\,u(y)\,dy, \tag{275}$$

we obtain

(276)

$$|\mathcal{K}(x,s)| \lesssim_n \int \sum_{\substack{a_1+\dots+a_{n+1}=n\\ b_1+\dots+b_n=n\\ a_j+b_j\leq n+1\\ a_j,\,b_j\in\mathbb{N}}} \prod_{j=1}^{n} \left|\left(\frac{d}{dy}\right)^{a_j}\left(\frac{1}{\tilde{q}(y)^{b_j}}\right)\right|\ \left|\left(\frac{d}{dy}\right)^{a_{n+1}}(\varphi(y)\,u(y))\right|\,dy.$$

Now for generic a, b, $c \in \mathbb{N}$, making use of Observation 18, we have

$$\left|\left(\frac{d}{dy}\right)^{a}\left(\frac{1}{\tilde{q}^b(y)}\right)\right| \lesssim_{a,b,d} \sup_{\substack{s\leq a\\ n_1+\dots+n_s=a\\ n_1,\,n_2,\,\dots,n_s\in\mathbb{N}}} \frac{|\tilde{q}^{(n_1)}\dots\tilde{q}^{(n_s)}|}{|\tilde{q}|^{s+b}},$$

$$\left|\left(\frac{d}{dy}\right)^{b}\varphi(y)\right| \lesssim \sum_{W_r\in\mathcal{W}} \frac{1}{|W_r|^b}\,|\tilde{\varphi}_{W_r}(y)|,$$

and

$$\left|\left(\frac{d}{dy}\right)^{c}u(y)\right| \lesssim \frac{1}{|I_2|^c}\,|\psi_{k_1}(x-y)\,\tilde{\psi}_{k_2}(y-s)|,$$

where here $\tilde{\varphi}_{W_r}$ and $\tilde{\psi}_{k_2}$ are functions with the same localization/smoothness properties as φ_{W_r} and ψ_{k_2} respectively.

Now using (255)–(258), we get

$$|\mathcal{K}(x,s)| \lesssim_n \frac{1}{|I_1|}\,\frac{1}{|I_2|}\sum_{W_r}\frac{|W_r|}{\Delta_q(W_r)^n} \lesssim \frac{1}{|I_1|}\,\lceil\Delta(P_1,P_2)\rceil^{n\,\epsilon_0}, \tag{277}$$

which proves (268).

For (270), we repeat the previous argument but now in the setting $\epsilon_0 = 0$ and $n = 1$, and once we reach the first inequality in (277) we appeal to (263), (264) and (255) in order to obtain the estimate

$$\sum_{W_r} \frac{|W_r|}{\Delta_q(W_r)} = \sum_{W_r} \frac{1}{\|q\|_{L^\infty(W_r)}} \lesssim_d \frac{1}{\eta(J,q,1)} \approx_d \lceil \Delta(P_1,P_2) \rceil^{\frac{1}{d}} \, |I_2|.$$

This ends the proof of our lemma. □

9.3. *Interaction between two separated trees/rows.* In this final subsection, we will show that relation (246) can be extended to situations that allow the interactions of a more complex structured family of tiles. That is, instead of single tiles, we will control the interaction of two δ^{-1}-separated trees and moreover of two rows—concepts that will be defined immediately below. Both Lemmas 7 and 8 in this subsection have a direct correspondent in [36] (see Lemmas 4 and 5) and further on in [81] (see Lemmas 2 and 3 therein), respectively.

We start by introducing a concept that, heuristically, quantifies the almost disjointness of the frequency locations of two trees.

Definition 13 (*Separated trees*). Fix a number $\delta \in (0,1]$. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two trees with tops $P_1 = [\vec{\alpha}_1, I_1]$ and $P_2 = [\vec{\alpha}_2, I_2]$ respectively. We say that $\mathcal{P}_1$ and $\mathcal{P}_2$ are δ^{-1}-*separated* if either $I_1 \cap I_2 = \emptyset$ or else

- $P = [\vec{\alpha}, I] \in \mathcal{P}_1$ and $I \subseteq I_2 \implies \lceil \Delta(P,P_2) \rceil < \delta$;
- $P = [\vec{\alpha}, I] \in \mathcal{P}_2$ and $I \subseteq I_1 \implies \lceil \Delta(P,P_1) \rceil < \delta$.

Next, motivated by its better spatial localization properties (see Observation 19 below) we introduce a special notion of tree:

Definition 14 (*Normal tree*). A tree $\mathcal{P}$ with top $P_0 = [\vec{\alpha}_0, I_0]$ is called *normal* if for any $P = [\vec{\alpha}, I] \in \mathcal{P}$, we have $100I \cap (I_0)^c = \emptyset$.

Observation 19. If $\mathcal{P}$ is a normal tree as above, then

$$\text{supp}\, T^{\mathcal{P}^*} := \bigcup_{P\in\mathcal{P}} \text{supp}\, T_P^* \subseteq I_0.$$

Notice that any tree can be written as a union between a normal tree and a sparse tree.

Definition 15 (*Separation and critical sets*). Fix $\delta \in (0,1)$ small[69] and $\epsilon_0 \in (0,1)$. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two δ^{-1}-separated trees as in Definition 13 with $I_2 \subseteq I_1$. Also let q_j be the central polynomial of P_j, $j \in \{1,2\}$, and

[69]For the remainder of the present section, we assume without loss of generality that $\delta \ll 1$, as otherwise all the results within this subsection are trivial.

$q_{1,2}=q_1-q_2$ the (P_1,P_2)-interaction polynomial.[70] Recalling the construction in Lemma 5, we define

- $I[s]$: the *separation set* of $\mathcal{P}_1$ and $\mathcal{P}_2$ by

$$I[s]:=J_s\left(q_{1,2},\,c_0(d)\,\delta^{-1}\right) \tag{278}$$

  for $J:=\tilde{I}_1\cap\tilde{I}_2$ and $c_0(d)>0$ properly[71]chosen;
- $I[c]$: the (ϵ_0-)*critical intersection set* by

$$I[c]:=J_s\left(q_{1,2},\,c_0(d)\,\delta^{-\epsilon_0}\right) \tag{279}$$

  for $J:=\tilde{I}_1\cap\tilde{I}_2$.

If one adds the extra-assumption $\mathcal{P}_2$ normal tree, then one can take in the above $J:=I_2$.

Observation 20.

(1) In what follows we will choose $c_0(d)$ in (278) such that if $\{I^j\}_{j=1}^l$ is the decomposition of $I[s]$ analogous to (260), then for any $j\in\{1,\dots,l\}$,

$$\text{if } I^j\cap\tilde{I}_P\neq\emptyset,\ \text{ then } |I_P|>|I^j|\quad\ \forall\ P=[\vec{\alpha},I_P]\in\mathcal{P}_1\cup\mathcal{P}_2. \tag{280}$$

Deduce that, in particular, we must have that for any $j\in\{1,\dots,l\}$,

$$\Delta_{q_{1,2}}(I^j)\gtrsim_d\delta^{-1}. \tag{281}$$

(2) From relations (254)–(257) of Lemma 5, we further deduce that for any dyadic $I\subset J$ such that $I[s]\cap\frac{3}{2}I=\emptyset$, we have

$$\inf_{x\in I}|q_{1,2}(x)|\leq\sup_{x\in I}|q_{1,2}(x)|\leq c(d)\inf_{x\in I}|q_{1,2}(x)|, \tag{282}$$

where here $c(d)\leq(100d)^d$, and

$$\Delta_{q_{1,2}}(I)\gtrsim_d\delta^{-1}. \tag{283}$$

Moreover, one has

for all $P=[\vec{\alpha},I_P]\in\mathcal{P}_1$ such that $I[s]\cap\tilde{I}_P=\emptyset$ and $I_P\subset I_2$, we have

$$\mathrm{Graph}(q_2)\cap\left(c(d)\delta^{-1}\right)\widehat{P}=\emptyset.$$

Of course, the same is true for the symmetric relation, i.e., replacing the index 1 with 2 and vice versa.

(3) Again based on Lemma 5, we deduce

$$\text{for all } P=[\vec{\alpha},I_P]\in\mathcal{P}_1\cup\mathcal{P}_2,\ \text{ we have } |\tilde{I}_P\cap I[c]|\lesssim_d\delta^{\frac{1-\epsilon_0}{d}}|I_P|. \tag{284}$$

---

[70] Throughout this section we assume without loss of generality that $q_{1,2}\notin\mathcal{Q}_0$ as otherwise all of our results below become a restatement of the corresponding results from the classical Carleson operator (linear polynomial) case; see [36].

[71] See Observation 20 below.

(4) For the remainder of the section, for convenience, we will set

$$\epsilon_0 := \frac{1}{2}. \tag{285}$$

The results below, though, hold for any choice of $\epsilon_0 \in (0,1)$, where of course the implicit bounds in " $\lesssim$ " depend on ϵ_0.

LEMMA 7 (Interaction of normal separated trees). *Let $\mathcal{P}_j$ for $j \in \{1,2\}$ be two normal and δ^{-1}-separated trees with tops $P_j = [\vec{\alpha}_j, I_j]$. Then, for any $f,\, g \in L^2(\mathbb{T})$ and $n \in \mathbb{N}$, we have*

- *if $I_1 = I_2$, then*

$$\begin{aligned} \left|\left\langle T^{\mathcal{P}_1\,*} f,\, T^{\mathcal{P}_2\,*} g\right\rangle\right| \lesssim_{n,d} \delta^n \left\|f\right\|_{L^2(I_1)} \left\|g\right\|_{L^2(I_2)} \\ + \left\|\chi_{I[c]} T^{\mathcal{P}_1\,*} f\right\|_2 \left\|\chi_{I[c]} T^{\mathcal{P}_2\,*} g\right\|_2 . \end{aligned} \tag{286}$$

- *More generally, if $I_2 \subseteq I_1$ then*

$$\begin{aligned} \left|\left\langle T^{\mathcal{P}_1\,*} f, T^{\mathcal{P}_2\,*} g\right\rangle\right| \lesssim_{n,d} \delta^n \left(\left\|Mf\right\|_{L_2(I_2)} + \left\|M(T^{\mathcal{P}_1\,*} f)\right\|_{L_2(I_2)}\right) \left\|g\right\|_{L_2(I_2)} \\ + \left\|\chi_{I[c]} T^{\mathcal{P}_1\,*} f\right\|_2 \left\|\chi_{I[c]} T^{\mathcal{P}_2\,*} g\right\|_2 . \end{aligned} \tag{287}$$

Observation 21. (i) All of the work performed in Section 9.2, culminating in (268) and (269) in Lemma 6, was designed in order to achieve (286). The specific form of (286) creates a nice parallelism with the part of Lemma 4 in [36] addressing a similar estimate. Indeed, the first term on the right-hand side of (286) corresponds to the situation in which the interacting tiles can be essentially assimilated as *linear* wave-packets, behaving thus as in the classical Carleson case treated by Fefferman. The second term is the one that truly captures the (non-linear) polynomial nature of our problem, in which two tiles $P^j \in \mathcal{P}_j$, $j \in \{1,2\}$—despite having a small geometric factor $\lceil \Delta(P^1, P^2) \rceil$—can still intersect, thus preventing a very fast decay of their interaction $\langle T^*_{P^1} f, T^*_{P^2} g\rangle$.

If one is willing to sacrifice this parallelism, one can circumvent Section 9.2 completely and simply appeal to Van der Corput estimates in order to obtain

$$\left|\left\langle T^{\mathcal{P}_1\,*} f,\, T^{\mathcal{P}_2\,*} g\right\rangle\right| \lesssim_d \delta^{\frac{1}{d}} \left\|f\right\|_{L^2(I_1)} \left\|g\right\|_{L^2(I_2)},$$

which is still enough to prove Proposition 2. For more on this, one can consult the second item in Section 11 and the work in [84].

(ii) The assumption for our trees to be normal is not required for the case $I_1 = I_2$. However, in the case $I_2 \subsetneq I_1$, one needs to assume at the very least that $\mathcal{P}_2$ is normal, since, otherwise it could happen that there exist $P^j \in \mathcal{P}_j$, $j \in \{1,2\}$, with $I_{P^1} \cap I_{P^2} = \emptyset$ and $\tilde{I}_{P^1} \cap \tilde{I}_{P^2} \neq \emptyset$ such that $\{q \in \mathcal{Q}_{d-1} \,|\, q \in P^1$ and $q \in P^2\} \neq \emptyset$, thus invalidating (287).

Proof. In what follows, we will only address the more general relation (287), which asks for an extra-localization of our estimates within the smaller interval I_2.[72]

Step 1. *Decomposition of the tree-interaction into linear-like and higher order interactions.* We start our proof by translating the notions introduced in Lemma 5 to our context; in particular, (254) now becomes

$$J_l(q,\lambda) = \bigcup_{W\in CZ_{(q,\lambda)}(J)} W \tag{288}$$

for $J := I_2$, $q := q_{1,2}$ and $\lambda := c_0(d)\,\delta^{-1}$ chosen as in (278).

Notice that with the above notation and conventions, we have

$$J_l(q,\lambda) = I_2 \setminus I[s]. \tag{289}$$

Next, consider a partition of unity adapted to $\{W\}_{W\in CZ_{(q,\lambda)}(J)} \cup \{I[s]\}$ such that

$$\sum_{W\in CZ_{(q,\lambda)}(J)} \varphi_W^2(x) + \underline{\varphi}_{I[s]}^2(x) = 1 \quad \text{for} \quad x\in I_2 \tag{290}$$

and such that the following hold:

- $\{\varphi_W\}$ obey (273);
- $0\le \underline{\varphi}_{I[s]} \le 1$ is smooth, adapted to $I[s]$ with $\underline{\varphi}_{I[s]}(x) = 0$ for $x\in I_2\setminus I[s]$ and $\underline{\varphi}_{I[s]}(x) = 1$ for $x\in \frac{3}{4}I[s]$;
- $0\le \varphi_{I[s]}, \varphi_{I[c]} \le 1$ smooth with $\varphi_{I[c]}(x) = 0$ for $x\in I_2\setminus I[c]$ and $\varphi_{I[c]}(x) = 1$ for $x\in \frac{3}{4}I[c]$ and

$$\underline{\varphi}_{I[s]}^2 = \varphi_{I[s]}^2 + \varphi_{I[c]}^2. \tag{291}$$

For $j\in\{1,2\}$, we further define the following tile-sets:

$$\mathcal{P}_j(I[s]) := \{P = [\vec{\alpha}, I_P] \in \mathcal{P}_j \mid \tilde{I}_P \cap (\operatorname{supp} T^{\mathcal{P}_2 *}) \cap \frac{3}{2} I[s] \ne \emptyset\} \tag{292}$$

and, for each $W\in J_l(q,\lambda)$, set

$$\mathcal{P}_j(W) := \{P = [\vec{\alpha}, I_P] \in \mathcal{P}_j \mid \tilde{I}_P \cap (\operatorname{supp} T^{\mathcal{P}_2 *}) \cap \frac{3}{2} W \ne \emptyset\}. \tag{293}$$

[72] From the hypothesis $\mathcal{P}_2$ normal tree one derives that either $I[c]\subseteq I_2$ or else $I[c]=\emptyset$.

With these we have

$$\left\langle T^{\mathcal{P}_1{}^*}f, T^{\mathcal{P}_2{}^*}g\right\rangle = \sum_{W\in J_l(q,\lambda)} \left\langle \varphi_W T^{\mathcal{P}_1{}^*}f, \varphi_W T^{\mathcal{P}_2{}^*}g\right\rangle + \left\langle \underline{\varphi}_{I[s]} T^{\mathcal{P}_1{}^*}f, \underline{\varphi}_{I[s]} T^{\mathcal{P}_2{}^*}g\right\rangle$$
$$= \sum_{W\in J_l(q,\lambda)} \left\langle \varphi_W T^{\mathcal{P}_1(W)^*}f, \varphi_W T^{\mathcal{P}_2(W)^*}g\right\rangle + \left\langle \underline{\varphi}_{I[s]} T^{\mathcal{P}_1(I[s])^*}f, \underline{\varphi}_{I[s]} T^{\mathcal{P}_2(I[s])^*}g\right\rangle. \tag{294}$$

This decomposition of the tree interaction $\left\langle T^{\mathcal{P}_1{}^*}f, T^{\mathcal{P}_2{}^*}g\right\rangle$ reflects the philosophy described in Observation 21:

- The first term encapsulates the *linear-like behavior* of the interacting wave-packets that are spatially localized in a region that lives far away from the separation set $I[s]$. Notice now that (282) and (283) in Observation 20 hold for any $I := \tilde{I}_P \cap \frac{3}{2}W$ with $P = [\vec{\alpha}, I_P] \in \mathcal{P}_1(W) \cup \mathcal{P}_2(W)$. As a consequence, we expect all the interactions of the form $\left\langle \varphi_W T^{\mathcal{P}_1{}^*}f, \varphi_W T^{\mathcal{P}_2{}^*}g\right\rangle$ to be small, in the sense expressed by (296) below.
- The second term is a manifestation of the interaction between *higher order wave-packets* that are now spatially localized around the separation set $I[s]$. Using

$$\left\langle \underline{\varphi}_{I[s]} T^{\mathcal{P}_1{}^*}f, \underline{\varphi}_{I[s]} T^{\mathcal{P}_2{}^*}g\right\rangle = \left\langle \varphi_{I[s]} T^{\mathcal{P}_1(I[s])^*}f, \varphi_{I[s]} T^{\mathcal{P}_2(I[s])^*}g\right\rangle + \left\langle \varphi_{I[c]} T^{\mathcal{P}_1{}^*}f, \varphi_{I[c]} T^{\mathcal{P}_2{}^*}g\right\rangle =: I + II, \tag{295}$$

  we will show that the first term can be treated—based on the first item in Observation 20—in a similar fashion with the linear-like interaction discussed earlier, while the second term will be simply controlled by the L^2 norms of $T^{\mathcal{P}_1{}^*}f$ and $T^{\mathcal{P}_2{}^*}g$ restricted to the critical set $I[c]$; see relations (296) and (297).

With these being said, our goal will be to prove the following relation:

$$\left|\left\langle \varphi_W T^{\mathcal{P}_1{}^*}f, \varphi_W T^{\mathcal{P}_2{}^*}g\right\rangle\right| + \left|\left\langle \varphi_{I[s]} T^{\mathcal{P}_1{}^*}f, \varphi_{I[s]} T^{\mathcal{P}_2{}^*}g\right\rangle\right| \lesssim_{n,d} \delta^n \left(\|Mf\|_{L_2(I_2)} + \left\|M(T^{\mathcal{P}_1{}^*}f)\right\|_{L_2(I_2)}\right)\|g\|_{L_2(I_2)}. \tag{296}$$

Note that the critical set interaction is trivially controlled via Cauchy-Schwarz:

$$\left|\left\langle \varphi_{I[c]} T^{\mathcal{P}_1{}^*}f, \varphi_{I[c]} T^{\mathcal{P}_2{}^*}g\right\rangle\right| \lesssim_d \left\|\chi_{I[c]} T^{\mathcal{P}_1{}^*}f\right\|_2 \left\|\chi_{I[c]} T^{\mathcal{P}_2{}^*}g\right\|_2. \tag{297}$$

Assuming for the moment (296), we have that (287) follows immediately from (294), (296), (297) and the simple observation $\#CZ_{(q,\lambda)}(J) \lesssim_d \log\frac{1}{\delta}$.

Step 2. *A short-time Fourier transform-type decomposition.* We start by defining a real-valued function $\phi \in C_0^{\infty}(\mathbb{R})$ with the following properties:

(298)
- $\operatorname{supp}\phi \subset \left\{\frac{1}{4} \leq |x| \leq \frac{1}{2}\right\}$;
- ϕ is even;
- $|\hat{\phi}(\xi) - 1| \lesssim_n |\xi|^{n+1} \quad \forall\, |\xi| \leq 1$ and $n \in \mathbb{N}$;
- $|\hat{\phi}(\xi)| \lesssim_n |\xi|^{-n-1} \quad \forall\, |\xi| \geq 1$.

Next, with the previous notation, we let

$$d_W := \min\{|W|,\, |I_P| \,|\, P = [\vec{\alpha}, I_P] \in \mathcal{P}_1(W) \cup \mathcal{P}_2(W)\} \tag{299}$$

and

$$d_{I[s]} := c(d)\, w(I_2, q_{12}, \delta^{-1}). \tag{300}$$

Thus deduce that for an appropriate choice of $c(d)$, one has $|I[s]| \approx_d d_{I[s]}$ and moreover $|I[c]| \approx_d \delta^{\frac{1}{2d}}\, d_{I[s]}$.

Let $j \in \{1, 2\}$ and K be a label that stands for either $I[s]$ or W. With this, we define

$$\phi_K(x) := (\delta^{\frac{1}{2d}} d_K)^{-1} \phi((\delta^{\frac{1}{2d}} d_K)^{-1} x) \tag{301}$$

and the corresponding operators

$$\tilde{\phi}_K \ : L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R}) \quad \textit{by} \ \ \tilde{\phi}_K f := \phi_K * f \tag{302}$$

and

$$\Phi_{j,K} \ : L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R}) \quad \textit{by} \ \ \Phi_{j,K} := \left(\prod_{l=1}^{d} M_{l,a_l^j}\right) \tilde{\phi}_K \left(\prod_{l=1}^{d} M_{l,a_l^j}^{*}\right). \tag{303}$$

In the last line we assumed that $q_j \in \mathcal{Q}_{d-1}$ is the central polynomial corresponding to $P_j = [\vec{\alpha}_j, I_0]$ and that

$$Q_j(y) = \sum_{l=1}^{d} a_l^j\, y^l \tag{304}$$

is the unique polynomial in $\mathcal{Q}_d$ such that $\frac{d}{dx} Q_j(x) = q_j$ and $Q_j(0) = 0$.

For $x \in I_2$ and $j \in \{1, 2\}$, we now let

$$\varphi_K T^{\mathcal{P}_j(K)^*} f(x) =: \Phi_{j,K} \varphi_K T^{\mathcal{P}_j(K)^*} f(x) \ + \ \Omega_{j,K} f(x). \tag{305}$$

Notice that the first term can be thought at the heuristic level as a smooth truncation whose frequency representation identifies with $(\varphi_K T^{\mathcal{P}_j(K)^*} f)\widehat{\ }$ on a $(\delta^{\frac{1}{2d}} d_K)^{-1}$-neighborhood around the moral support of $(\varphi_K T^{\mathcal{P}_j(K)^*} f)\widehat{\ }$. As a consequence, we expect the reminder term represented by $\Omega_{j,K} f$ to behave as an error term; indeed, this will be a direct consequence of (307) below.

Our intention is to prove that for any $h \in L^2(\mathbb{T})$, $j \in \{1,2\}$ and K as above, one has[73]

$$|\Phi_{2,K}\, \Phi_{1,K}\, \varphi_K\, h(x)| \,+\, |\Phi_{1,K}\, \Phi_{2,K}\, \varphi_K\, h(x)| \lesssim_{n,d} \delta^{\frac{n}{2d}}\, Mh(x) \tag{306}$$

and

$$\left|\Omega^*_{j,K} h(x)\right| \lesssim_{n,d} \left(\mathcal{R}_K * |h|\right)(x) \quad \text{with} \quad \|\mathcal{R}_K\|_1 \lesssim_{n,d} \delta^{\frac{n}{2d}}, \tag{307}$$

where

$$\mathcal{R}_K(y) := \sum_{2^k \leq (d_K)^{-1}} (\delta^{\frac{1}{2d}} 2^k d_K)^n\, 2^k\, \chi_{[-2^{-k+10}, 2^{-k+10}]}(y). \tag{308}$$

Step 3. *The linear-like interactions*: *proof of* (296). We start by assuming for the moment that (306) and (307) hold and show (296). Making use of the fact that

$$T^{\mathcal{P}_j *} f(x) = T^{\mathcal{P}_j(K)^*} f(x) \quad \text{for } x \in \frac{3}{2} K \cap \operatorname{supp} T^{\mathcal{P}_2 *}, \tag{309}$$

we have

$$\begin{aligned}
&\left\langle \varphi_K T^{\mathcal{P}_1 *} f, \varphi_K T^{\mathcal{P}_2 *} g \right\rangle \\
&\quad = \left\langle \Phi_{1,K} \varphi_K T^{\mathcal{P}_1 *} f, \Phi_{2,K} \varphi_K T^{\mathcal{P}_2 *} g \right\rangle \,+\, \left\langle \Phi_{1,K} \varphi_K T^{\mathcal{P}_1 *} f, \Omega_{2,K}\, g \right\rangle \\
&\quad\quad + \left\langle \Omega_{1,K} f, \varphi_K T^{\mathcal{P}_2 *} g \right\rangle \\
&\quad = \left\langle \Phi_{2,K} \Phi_{1,K} \varphi_K T^{\mathcal{P}_1 *} f, \varphi_K T^{\mathcal{P}_2 *} g \right\rangle \,+\, \left\langle \Omega^*_{2,K} \Phi_{1,K} \varphi_K T^{\mathcal{P}_1 *} f, g \chi_{I_2} \right\rangle \\
&\quad\quad + \left\langle f, \Omega^*_{1,K} \varphi_K T^{\mathcal{P}_2 *} g \right\rangle =: A \,+\, B \,+\, C.
\end{aligned} \tag{310}$$

Now using (306), (307), the fact that $\mathcal{P}_2$ is normal and Lemma 2 (or Lemma 4), one has

$$\begin{aligned}
|A| &\lesssim_{n,d} \delta^{\frac{n}{2d}} \left\langle M(T^{\mathcal{P}_1 *} f), |T^{\mathcal{P}_2 *} g| \right\rangle \\
&\lesssim_{n,d} \delta^{\frac{n}{2d}} \left\| M(T^{\mathcal{P}_1 *} f) \right\|_{L_2(I_2)} \|g\|_{L_2(I_2)},
\end{aligned} \tag{311}$$

$$\begin{aligned}
|B| &\lesssim_d \left\langle \mathcal{R}_K * \left(|\Phi_{1,K} \varphi_K T^{\mathcal{P}_1 *} f| \right), |g| \chi_{I_2} \right\rangle \\
&\lesssim_{n,d} \delta^{\frac{n}{2d}} \left\| M(T^{\mathcal{P}_1 *} f) \right\|_{L_2(I_2)} \|g\|_{L_2(I_2)},
\end{aligned} \tag{312}$$

and

$$|C| \lesssim_d \left\langle |f|, \mathcal{R}_K * \left(|T^{\mathcal{P}_2 *} g| \right) \right\rangle \lesssim_{n,d} \delta^{\frac{n}{2d}} \, \|Mf\|_{L_2(I_2)} \, \|g\|_{L_2(I_2)}. \tag{313}$$

Now putting together (310)–(313) we conclude that (296) holds.

[73]The estimates below are given for a fixed large $n \in \mathbb{N}$ in (298). Since (298) holds for any n, in the final estimates one can replace the exponent $\frac{n}{2d}$ by simply n.

We return now to proving (306) and (307) and focus first on (306):

$$
\begin{aligned}
&|\Phi_{2,K}\,\Phi_{1,K}\,\varphi_K\,u(x)| \\
&\quad\lesssim \int_{\mathbb{R}} |\varphi_K(s)\,u(s)| \left|\int_{\mathbb{R}} \phi_K(x-y)\,\phi_K(y-s)\,e^{i\,(\int_y^x q_2 - \int_y^s q_1)}\,dy\right| ds \qquad (314)\\
&\quad=: \int_{\mathbb{R}} |\varphi_K(s)\,u(s)|\,|\mathcal{K}_K(x,s)|\,ds.
\end{aligned}
$$

Now setting $2^{\underline{k}} := (\delta^{\frac{1}{2d}} d_K)^{-1}$ we further notice that, up to conjugation, $\mathcal{K}_K$ is a variant of the kernel $\mathcal{K}$ defined in (274). Moreover, from the definition of $\underline{k}$ and condition $s \in I_K := \operatorname{supp}\varphi_K$ we deduce that the integrand variable obeys $y \in \underline{I}_K := \operatorname{supp}\varphi_K\, +\, \operatorname{supp}\phi_K$ and that

$$2^{-\underline{k}}\,\|q_{1,2}\|_{L^\infty(\underline{I}_K)} \gtrsim \delta^{-\frac{1}{2}}.$$

With these said, one can now follow step by step the reasonings displayed for the proof of (268) in Lemma 6, specifically (275)–(277), to conclude

$$|\mathcal{K}_K(x,s)| \lesssim_{n,d} \delta^{\frac{n}{2}}\,2^{\underline{k}}\,\chi_{\{|t|\lesssim 2^{-\underline{k}}\}}(x-s), \qquad (315)$$

which immediately implies (306).

We now turn our attention towards (307) and notice that it is enough to show that

$$\mathcal{E}_{j,K}h := \left(\prod_{l=1}^{d} M^*_{l,a_l^j}\right)\,\Omega^*_{j,K}\left(\prod_{l=1}^{d} M_{l,a_l^j}\right)h, \qquad h \in L^2(\mathbb{T}). \qquad (316)$$

obeys the estimate

$$|\mathcal{E}_{j,K}h(x)| \lesssim \int |h(x-y)|\,\mathcal{R}_K(y), \qquad (317)$$

with $\mathcal{R}_K$ verifying (307) and (308).

We first notice that

$$\mathcal{E}_{j,K} = \sum_{P\in\mathcal{P}_j(K)} \mathcal{E}_{j,P}, \qquad (318)$$

where

$$
\begin{aligned}
\mathcal{E}_{j,P}h :=& \left(\prod_{l=1}^{d} M^*_{l,a_l^j}\right)\,T_P\,\varphi_K\left(\prod_{l=1}^{d} M_{l,a_l^j}\right)h \qquad (319)\\
&- \left(\prod_{l=1}^{d} M^*_{l,a_l^j}\right)\,T_P\,\varphi_K\left(\prod_{l=1}^{d} M_{l,a_l^j}\right)\tilde{\phi}_K\,h.
\end{aligned}
$$

Now fix $P \in \mathcal{P}_j(K)$, and assume $|I_P| = 2^{-k}$, $k \in \mathbb{N}$; then, for any $x \in E(P)$, we further have

$$|\mathcal{E}_{j,P}h(x)| \lesssim \int |(h\,\varphi_K)(x-y)|\,|r_x^P(y) - (r_x^P * \phi_K)(y)|\,dy, \qquad (320)$$

where

$$r_x^P(y) := e^{i\int_{x-y}^{x}(q_x-q_j)}\,\psi_k(y). \tag{321}$$

Appealing now to Lemma C in the appendix, we deduce that for any $x \in E(P)$, one has $\|q_x - q_j\|_{L^\infty(\tilde{I}_P)} \lesssim_d |I_P|^{-1} = 2^k$, from which

$$|\widehat{r_x^P}(\xi)| \lesssim_n \left(1+\frac{|\xi|}{2^k}\right)^{-n-1}, \quad n \in \mathbb{N}. \tag{322}$$

From (298) and (322) we obtain

$$\begin{aligned}&\int |\widehat{r_x^P}(\xi)|\,|1-\widehat{\phi_K}(\xi)|\,d\xi\\ &\lesssim \int_{|\xi|\le(\delta^{\frac{1}{2d}}\,d_K)^{-1}}\left(1+\frac{|\xi|}{2^k}\right)^{-n-1}(\delta^{\frac{1}{2d}}\,d_K)^{n+1}\,|\xi|^{n+1}\,d\xi\\ &\quad+\int_{|\xi|>(\delta^{\frac{1}{2d}}\,d_K)^{-1}}\left(1+\frac{|\xi|}{2^k}\right)^{-n-1}d\xi \lesssim_n 2^k\,(2^k\,\delta^{\frac{1}{2d}}\,d_K)^n.\end{aligned} \tag{323}$$

Putting together (308), (318)–(320) and (323) one can now easily verify that (317) and hence (307) hold.

Step 4: The higher order interaction: control over $\langle \underline{\varphi}_{I[s]}T^{\mathcal{P}_1*}f, \underline{\varphi}_{I[s]}T^{\mathcal{P}_2*}g\rangle$. The upper bound on the higher order interaction component follows now from (295), (296) and (297). □

We conclude this section by recording the following natural extension of the previous result:

Definition 16 (*Row*). A *row* is a collection $\mathcal{P} = \bigcup_{j\in\mathbb{N}}\mathcal{P}^j$ of normal trees $\mathcal{P}^j$ with tops $P_0^j = [\vec{\alpha}_0^j, I_0^j]$ such that the $\left\{I_0^j\right\}$ are pairwise disjoint.

Lemma 8 (Row-tree interaction). *Let $\mathcal{P}$ be a row as above, let $\mathcal{P}'$ be a normal tree with top $P_0' = [\vec{\alpha}_0', I_0']$, and suppose that for all $j \in \mathbb{N}$, $I_0^j \subseteq I_0'$ and $\mathcal{P}^j$, $\mathcal{P}'$ are δ^{-1}separated trees; denote by $I^j[c]$ the critical intersection set between each $\mathcal{P}^j$ and $\mathcal{P}'$.*

Then for any f, $g \in L^2(\mathbb{T})$ and $n \in \mathbb{N}$, we have that

$$\left|\left\langle T^{\mathcal{P}'*}f, T^{\mathcal{P}*}g\right\rangle\right| \lesssim_{n,d} \delta^n\,\|f\|_2\,\|g\|_2 + \left\|\sum_j \chi_{I^j[c]}T^{\mathcal{P}'*}f\right\|_2\left\|\sum_j \chi_{I^j[c]}T^{\mathcal{P}^j*}g\right\|_2. \tag{324}$$

Proof. For a fixed j, we apply Lemma 7 to the tree $\mathcal{P}_1 := \mathcal{P}'$ and the normal tree $\mathcal{P}_2 := \mathcal{P}^j$. Then, rewriting (287), we have

$$\begin{aligned}\left|\left\langle T^{\mathcal{P}'^*} f, T^{\mathcal{P}^{j*}} g\right\rangle\right| \lesssim_{n,d} \delta^n \left(\|Mf\|_{L_2(I_0^j)} + \left\|M(T^{\mathcal{P}'^*} f)\right\|_{L_2(I_0^j)}\right) \|g\|_{L_2(I_0^j)} \\ + \left\|\chi_{I^j[c]} T^{\mathcal{P}'^*} f\right\|_2 \left\|\chi_{I^j[c]} T^{\mathcal{P}^{j*}} g\right\|_2 .\end{aligned} \tag{325}$$

Now (324) follows from a simple application of Cauchy-Schwarz followed by Lemma 4 (specialized to the easier case $A = [0, 1]$ and $p = 2$). □

10. **Proof of Proposition 2**

In this section we will complete the last step required for finalizing the proof of our Main Theorem, namely, to present the proof of Proposition 2. This will be done in two stages: in the first subsection we will treat the case of an L^∞-forest, while in the second subsection we will approach the general case of a BMO-forest.

10.1. *A fundamental case*: *L^2-control over an L^∞-forest.* In this subsection we prove the following result:

MAIN LEMMA. (L^2-control over an L^∞-forest). *Let $\mathcal{P} \subset \mathbb{P}_n$ be an L^∞-forest of generation n. Then there exists $\eta = \eta(d) \in (0, 1)$ such that*

$$\|T^{\mathcal{P}} f\|_2 \lesssim_d 2^{-\frac{n}{2}\eta} \|f\|_2. \tag{326}$$

Moreover, if $\mathcal{P}$ is normal and $2^{100\,n\,d}$-separated,[74] *then—decomposing $\mathcal{P}$ canonically into a union of rows $\{\mathcal{R}_j\}$—one has*

$$\|T^{\mathcal{P}^*} f\|_2^2 \lesssim \sum_j \|T^{\mathcal{R}_j{}^*} f\|_2^2 + 2^{-5\,n} \|f\|_2^2, \tag{327}$$

from which one deduces the improved bound

$$\|T^{\mathcal{P}} f\|_2 \lesssim_d 2^{-\frac{n}{2}} \|f\|_2. \tag{328}$$

Observation 22. Relation (327) should be regarded as a strong almost-orthogonality relation arising from the good geometric properties imposed on $\mathcal{P}$. It essentially states that, up to a negligible term, one expects

$$\|T^{\mathcal{P}^*} f\|_2^2 \lesssim \sum_j \|T^{\mathcal{R}_j{}^*} f\|_2^2.$$

[74] As expected, an L^∞-forest $\mathcal{P}$ is called normal if all the trees inside are normal; the same principle applies for the δ^{-1}-separateness condition.

Proof. Step 1. *Row decomposition*: *preliminaries.* Recalling Definition 10, specifically (155)–(157), we have that $\mathcal{P}$ can be decomposed into a collection of separated trees

$$\mathcal{P}=\bigcup_{j\in\mathbb{N}}\mathcal{P}_j \tag{329}$$

with the property that the $\mathcal{P}$-counting function obeys

$$\mathcal{N}_{\mathcal{P}}(x)=\sum_j \chi_{I_j}(x)\leq c\,2^n \qquad \forall\, x\in\mathbb{T}, \tag{330}$$

where here $c>0$ is an absolute constant and for each j, $P_j=[\vec{\alpha}_j, I_j]$ is a top of the tree $\mathcal{P}_j$.

Now in order to select the "germs" of our future rows we proceed as follows: we first identify the collection of maximal intervals $\{I_{1k}\}_k$ among the collection of top time-intervals $\{I_j\}_j$. Then for each such I_{1k}, we select among the $P_j's$ one top tile, say P_{1k}, having as the time-interval I_{1k}. We let $\mathcal{R}_1$ be the collection of all the trees whose tops belong to the selected set $\{P_{1k}\}_k$. We then erase this entire collection of trees from our forest $\mathcal{P}$ and apply the same process as above to our remaining set of trees within $\mathcal{P}$. This way we continue to form $\mathcal{R}_2$, $\mathcal{R}_3$ and so on until we run out of trees. Finally, condition (330) guarantees that this process will end in at most $c\,2^n$ steps, and hence

$$\mathcal{P}=\bigcup_{j=1}^{c\,2^n}\mathcal{R}_j, \tag{331}$$

with each $\mathcal{R}_j$ being a maximal collection of spatially disjoint maximal trees $\{\mathcal{T}_{j,k}\}_k$ with $\bigcup_{j,k}\mathcal{T}_{j,k}=\bigcup_j\mathcal{P}_j$.

Step 2. *Row decomposition*: *isolating the normal tree component.* If I_{jk} stands for the time-interval of the top of $\mathcal{T}_{j,k}$, define the boundary component

$$\mathcal{T}_{j,k}^{bd}:=\{P\in\mathcal{T}_{j,k}\,|\,100 I_P\cap(I_{jk})^c\neq\emptyset\}$$

and let

$$\mathcal{P}^{bd}:=\bigcup_{j=1}^{c\,2^n}\mathcal{R}_j^{bd}\quad\text{with}\quad \mathcal{R}_j^{bd}:=\bigcup_k\mathcal{T}_{j,k}^{bd}.$$

With these done, we notice that $\mathcal{P}^{bd}$ is a sparse L^∞-forest and thus, via Proposition 1, one immediately gets the analogue of (326) for this boundary component.

As a consequence, we can erase $\mathcal{P}^{bd}$ from our initial forest $\mathcal{P}$ and deduce that for the new $\mathcal{P}$, the updated version of (331) represents its desired (canonical) row-decomposition.

Step 3. *Creating large separation among trees.* For each of the (remaining) trees $\mathcal{T}_{j,k}$, we construct inductively the set of tiles $\tilde{\mathcal{T}}_{j,k}$ as follows: at the first stage, we initialize $\tilde{\mathcal{T}}_{j,k}$ as the empty set, select the collection of all minimal tiles in $\mathcal{T}_{j,k}$, then move it to $\tilde{\mathcal{T}}_{j,k}$ and finally remove $\tilde{\mathcal{T}}_{j,k}$ from $\mathcal{T}_{j,k}$; Then, updating at each stage $\tilde{\mathcal{T}}_{j,k}$ and $\mathcal{T}_{j,k}$, we repeat this algorithm for $100\,n\,d$ times at which moment we stop. The resulting collection of tiles have the following properties:

- $\bigcup_{j,k} \tilde{\mathcal{T}}_{j,k}$ can be decomposed into at most $c(d)\,n$ negligible sets and thus this collection of tiles can be easily treated by Propositon 1; and
- $\mathcal{P} = \bigcup_{j=1}^{c\,2^n} \mathcal{R}_j = \bigcup_{j,k} \mathcal{T}_{j,k}$ is a normal and $2^{100\,n\,d}$-separated L^∞-forest.

Step 4. *The proofs of* (327) *and* (328). Both (327) and (328) follow easily from the following key observation: the operators $\{T^{\mathcal{R}_j}\}_{j=1}^{cn}$ are strongly almost orthogonal. More precisely, for $k \neq j$, we have that

- $\|T^{\mathcal{R}_k\,*}\,T^{\mathcal{R}_j}\|_{2\mapsto 2} = 0$;
- $\|T^{\mathcal{R}_k}\,T^{\mathcal{R}_j\,*}\|_{2\mapsto 2} \lesssim 2^{-10\,n}$.

Indeed, the first item is a direct consequence of the pairwise disjointness of the sets $\{\operatorname{supp} T^{\mathcal{R}_j}\}_j$. For the second item, one makes use of the strong $(2^{100\,n\,d})$-separateness hypothesis, relation (284) and Lemmas 8 and 4. $\square$

10.2. *The general BMO-forest case.* We start by restating the result that we need to prove

Proposition 2. *Let $\mathcal{P} \subseteq \mathbb{P}_n$ be a forest. Then there exists $\eta \in (0, 1/2)$, depending only on the degree d, such that for $1 < p < \infty$, we have*

$$\left\|T^{\mathcal{P}}\right\|_p \lesssim_{p,d} 2^{-n\,\eta\,(1-\frac{1}{p^*})}.$$

10.2.1. *The L^2 bound.* We start by recalling the setting described in Observation 14 as well as Definition 11. Appealing to a standard pigeonhole principle, from now on we can assume without loss of generality that

- the family $\mathcal{P}$ can be written as

$$\mathcal{P} = \bigcup_{k\geq 0} \mathcal{P}_n^k, \tag{332}$$

  with $\mathcal{P}$ a (BMO)-forest of generation n such that for each $k \geq 0$

$$\mathcal{P}_n^k \subset \mathbb{P}_n \text{ is an } L^\infty\text{-forest of generation } n; \tag{333}$$

- the spatial support of the tiles in $\mathcal{P}_n^k$ is contained in a set A_n^k that can be represented as a finite union of maximal (disjoint) dyadic intervals;
- there exists $c_0 \geq 10$ such that for each $k \in \mathbb{N}$, one has

$$A_n^k \prec_{n\,c_0} A_n^{k+1}; \text{ and} \tag{334}$$

- if $P = [\vec{\alpha}, I_P] \in \mathcal{P}_n^k$, then

$$\begin{aligned} & - I_P \subseteq A_n^k, \\ (335) \quad & - I_P \nsubseteq A_n^{k+1}, \\ & - A_{\mathbb{P}, A_n^k}(P) \in (2^{-n},\, 2^{-n+1}]. \end{aligned}$$

Observation 23. Following similar reasonings with those described in Section 10.1 (see Step 3 in the proof of the Main Lemma) we define

(336)
$$\check{\mathcal{C}}_n^k := \left\{ P \in \mathcal{P}_n^k \mid \text{there are no chains } P \lneq P_1 \lneq \cdots \lneq P_n \;\&\; \{P_j\}_{j=1}^n \subseteq \mathcal{P}_n^k \right\}$$

and notice that

- The set $\check{\mathcal{C}}_n := \bigcup_{k\geq 0} \check{\mathcal{C}}_n^k$ can be decomposed into a union of at most n sparse forests; applying Propositon 1 to each of the resulting sparse forests we have that the associated operator $T^{\check{\mathcal{C}}_n}$ is under control.
- Erasing from each $\mathcal{P}_n^k$ the corresponding set $\check{\mathcal{C}}_{100nd}^k$ one has that

$$\mathcal{P} := \bigcup_{k\geq 0} \mathcal{P}_n^k$$

is a BMO-forest of generation n such that each $\mathcal{P}_n^k$ is an L^∞-forest with the property that any two trees inside $\mathcal{P}_n^k$ are $2^{100\,n\,d}$-separated.

Now let

$$(337) \qquad \mathcal{P}_{n,bd}^{k,e} := \left\{ \begin{array}{c} P = [\vec{\alpha}, I] \in \mathcal{P}_n^k \\ (\text{hence } I \subseteq A_n^k) \end{array} \;\middle|\; \begin{array}{c} \exists\, J \subseteq A_n^{k+1} \text{ s.t. } 100I \cap J^c \neq \emptyset \\ \text{and } |I| < |J| \end{array} \right\}.$$

Next, for $\mathcal{P}_n^{k,\max}$ the collection of maximal tiles in $\mathcal{P}_n^k$, we set

$$(338) \quad \mathcal{P}_{n,bd}^{k,i} := \{P \in \mathcal{P}_n^k \mid \exists P_{kj} \in \mathcal{P}_n^{k,\max} \text{s.t.} P \leq P_{kj} \text{ and } 100 I_P \cap (I_{P_{kj}})^c \neq \emptyset\}.$$

Then, for each $\mathcal{P}_n^k$, we define its *boundary* forest component as

$$(339) \qquad \mathcal{P}_{n,bd}^k = \mathcal{P}_{n,bd}^{k,i} \cup \mathcal{P}_{n,bd}^{k,e}.$$

The *normal* forest component is defined as

$$(340) \qquad \mathcal{P}_{n,nm}^k := \mathcal{P}_n^k \setminus \mathcal{P}_{n,bd}^k.$$

Finally, we set

$$(341) \qquad \mathcal{P}_{bd} := \bigcup_{k\in\mathbb{N}} \mathcal{P}_{n,bd}^k$$

and

$$(342) \qquad \mathcal{P}_{nm} := \bigcup_{k\in\mathbb{N}} \mathcal{P}_{n,nm}^k.$$

Now, following the outline in Section 5.4, our plan is as follows:

- to estimate the L^2-bound of the operator $T^{\mathcal{P}_{nm}}$ we will show that the family $\{T^{\mathcal{P}^k_{n,nm}}\}_k$ consists of almost orthogonal operators;
- to treat the operator $T^{\mathcal{P}_{bd}}$ one simply notices that $\mathcal{P}_{bd}$ is a sparse forest and hence one can directly apply Propositon 1.

Claim 6. With the above notation, for $\eta=\eta(d)\in(0,1)$, one has

$$\left\|T^{\mathcal{P}_{nm}}\right\|_2\lesssim 2^{-\frac{n}{2}\eta}. \tag{343}$$

In order to prove the above claim, using the TT^*-method and recalling (334), it is enough to show that there exists absolute $c>0$ such that for any $k,\,k'\in\mathbb{N}$, one has

$$\|T^{\mathcal{P}^k_{n,nm}}\,T^{\mathcal{P}^{k'}_{n,nm}}{}^*\|_2\lesssim 2^{-c\,|k-k'|\,n}, \tag{344}$$

$$\|T^{\mathcal{P}^k_{n,nm}}{}^*\,T^{\mathcal{P}^{k'}_{n,nm}}\|_2\lesssim 2^{-c\,|k-k'|\,n}. \tag{345}$$

Indeed, (343) will then easily follow, since for any $k\in\mathbb{N}$,

$$\|T^{\mathcal{P}^k_{n,nm}}f\|_2\lesssim 2^{-\frac{n}{2}\eta}\,\|f\|_2\,. \tag{346}$$

Notice that (346) is a direct consequence of the second item in Observation 23 and the Main Lemma.

This being said, let us start by proving (344).

Without loss of generality, we suppose $k'>k+1$. Now applying Cauchy–Schwarz we have

$$\left|\left\langle T^{\mathcal{P}^k_{n,nm}}{}^*f,\,T^{\mathcal{P}^{k'}_{n,nm}}{}^*g\right\rangle\right|\le\|\chi_{A^{k'}_n}T^{\mathcal{P}^k_{n,nm}}{}^*f\|_2\,\|T^{\mathcal{P}^{k'}_{n,nm}}{}^*g\|_2.$$

Here we have used that $\mathcal{P}^{k'}_{n,nm}$ is normal and thus supp $T^{\mathcal{P}^{k'}_{n,nm}}{}^*\subseteq A^{k'}_n$.

Next, from the way in which we have constructed $\mathcal{P}^k_{n,nm}$ and A^{k+1}_n, we have that if $A^{k+1}_n=\bigcup J$ is the decomposition of A^{k+1}_n into maximal (disjoint) intervals, then

$$\forall\,P\in\mathcal{P}^k_{n,nm}\ \text{and}\ \forall\,J\ \text{s.t.}\ \tilde{I}_P\cap J\neq\emptyset,\quad\text{we have}\ |I_P|\ge|J|. \tag{347}$$

Thus, for any $P\in\mathcal{P}^k_{n,nm}$, we either have $\tilde{I}_P\cap A^{k+1}_n=\emptyset$ or, using (334) and (347), the following holds:

$$\frac{|\tilde{I}_P\cap A^{k'}_n|}{|\tilde{I}_P|}\le\frac{|\tilde{I}_P\cap A^{k'}_n|}{|\tilde{I}_P\cap A^{k+1}_n|}\lesssim 2^{-c_0\,|k'-k-1|\,n}. \tag{348}$$

Reaching this point, we recall that $\mathcal{P}^k_{n,nm}$ is a *normal* L^∞-forest of n^{th} generation and hence, applying the same reasoning as in the proof of the Main

Lemma, we have that

$$\mathcal{P}^{k}_{n,nm} = \bigcup_{j=1}^{c\,2^n} \mathcal{R}^{k}_{j}, \tag{349}$$

with each $\mathcal{R}^{k}_{j}$ a row.

Then, using (348) and applying Lemma 4 for $A := A^{k'}_{n}$, we obtain[75]

$$\left\| \chi_{A^{k'}_{n}}\, T^{\mathcal{P}^{k}_{n,nm}}{}^{*} f \right\|_2 \lesssim \sum_{j=1}^{c\,2^n} \left\| \chi_{A^{k'}_{n}}\, T^{\mathcal{R}^{k}_{j}}{}^{*} f \right\|_2 \lesssim 2^{-c\,|k-k'|\,n}\, \|f\|_2\,, \tag{350}$$

which proves (344).

We will now move on to the proof of (345).

As before, we start by first applying Cauchy–Schwarz

$$\left| \left\langle T^{\mathcal{P}^{k}_{n,nm}} f,\, T^{\mathcal{P}^{k'}_{n,nm}} g \right\rangle \right| \leq \|\chi_{A^{k'}_{n}}\, T^{\mathcal{P}^{k}_{n,nm}} f\|_2\, \|T^{\mathcal{P}^{k'}_{n,nm}} g\|_2.$$

Based on (349) and the fact that the operators $\{T^{\mathcal{R}^{k}_{j}}\}_j$ have disjoint supports, we have

$$\|\chi_{A^{k'}_{n}}\, T^{\mathcal{P}^{k}_{n,nm}} f\|_2^2 = \sum_j \|\chi_{A^{k'}_{n}}\, T^{\mathcal{R}^{k}_{j}} f\|_2^2 \lesssim 2^n \sup_j \|\chi_{A^{k'}_{n}}\, T^{\mathcal{R}^{k}_{j}} f\|_2^2. \tag{351}$$

Now, applying Lemma 3 to our row $\mathcal{R}^{k}_{j}$ (with the obvious replacement of the partition $CZ_{I_0}(\mathcal{P})$ by $CZ_{I_{\mathcal{R}^{k}_{j}}}(\mathcal{R}^{k}_{j})$ with $I_{\mathcal{R}^{k}_{j}} \subseteq A^{k}_{n}$ the spatial support of the row $\mathcal{R}^{k}_{j}$), we have

$$\|\chi_{A^{k'}_{n}}\, T^{\mathcal{R}^{k}_{j}} f\|_2 \lesssim \left(\sup_{J \in CZ_{I_{\mathcal{R}^{k}_{j}}}(\mathcal{R}^{k}_{j})} \frac{|E_{A^{k'}_{n}}(J)|}{|J|} \right)^{\frac{1}{2}} \|f\|_2\,. \tag{352}$$

Now, based on (347), (348), the construction of $\mathcal{P}^{k}_{n,nm}$ and the assumption $k' > k+1$, we have

$$\sup_{J \in CZ_{I_{\mathcal{R}^{k}_{j}}}(\mathcal{R}^{k}_{j})} \frac{|E_{A^{k'}_{n}}(J)|}{|J|} \lesssim 2^{-c_0\,|k'-k-1|\,n}.$$

Thus, combining this last observation with (351) and (352), we deduce

$$\|\chi_{A^{k'}_{n}}\, T^{\mathcal{P}^{k}_{n,nm}} f\|_2 \lesssim 2^{-c\,|k-k'|\,n}\, \|f\|_2\,,$$

which together with (346) implies (345).

[75] Recall that throughout the paper the constant $c > 0$ is allowed to change from line to line.

10.2.2. *The L^p bound*, $1<2\neq p<\infty$. In this section, based on assumptions (332)–(335) and Observation 23, we will show that

$$\left\|T^{\mathcal{P}_{nm}}\right\|_p \lesssim_p 2^{-n\,\eta\,(1-\frac{1}{p^*})}. \tag{353}$$

Our proof will be split into two cases:

Case 1: *Assume* $1<p<2$. In this situation we notice that $p^*=p$, and thus (353) is equivalent to

$$\left\|T^{\mathcal{P}_{nm}\,*}\right\|_{p'} \lesssim_{p'} 2^{-\frac{n\eta}{p'}}. \tag{354}$$

Firstly we notice—based on elementary interpolation techniques—that it is enough to prove (354) only for $p'\in 2\mathbb{N}$ with $p'\geq 2$.

In this context, at the heuristic level, our goal is to show that

$$\|\sum_k T^{\mathcal{P}^k_{n,nm}\,*} f\|_{p'}^{p'} \lesssim_{p'} \sum_k \|T^{\mathcal{P}^k_{n,nm}\,*} f\|_{p'}^{p'} + \text{Error}, \tag{355}$$

where the "Error" term above is appropriately small and will be made precise in what follows.

Indeed, for this we first notice that (up to conjugation), we have

$$\Big\|\sum_k T^{\mathcal{P}^k_{n,nm}\,*} f\Big\|_{p'}^{p'} \approx_{p'} \sum_{\substack{(k_1,\ldots,k_{p'}),\,(r_1,\ldots,r_{p'})\in\mathbb{N}^{p'}\\ r_1+\cdots+r_{p'}=p'}} \int (T^{\mathcal{P}^{k_1}_{n,nm}\,*} f)^{r_1}\ldots(T^{\mathcal{P}^{k_{p'}}_{n,nm}\,*} f)^{r_{p'}}$$

and after applying the Hölder and Jensen inequalities we further have

$$\begin{aligned}\Big\|\sum_k T^{\mathcal{P}^k_{n,nm}\,*} f\Big\|_{p'}^{p'} &\lesssim_{p'} \sum_{\substack{(k_1,\ldots,k_{p'}),\,(r_1,\ldots,r_{p'})\in\mathbb{N}^{p'}\\ r_1+\cdots+r_{p'}=p'}} \left(\int_{\bigcap_{j=1}^{p'} A_n^{k_j}} \left|T^{\mathcal{P}^{k_1}_{n,nm}\,*} f\right|^{p'}\right)^{\frac{r_1}{p'}}\\ &\qquad\ldots\left(\int_{\bigcap_{j=1}^{p'} A_n^{k_j}} \left|T^{\mathcal{P}^{k_{p'}}_{n,nm}\,*} f\right|^{p'}\right)^{\frac{r_{p'}}{p'}}\\ &\lesssim_{p'} \sum_k \sum_{m\in\mathbb{N}} |m+1|^{100p'} \int_{A_n^{k+m}} \left|T^{\mathcal{P}^k_{n,nm}\,*} f\right|^{p'}.\end{aligned}$$

Thus, we have just proved that for $p'\in 2\mathbb{N}$, with $p'>1$, we have that

$$\Big\|\sum_k T^{\mathcal{P}^k_{n,nm}\,*} f\Big\|_{p'}^{p'} \lesssim_{p'} \sum_k \sum_{m\in\mathbb{N}} |m+1|^{100p'} \int_{A_n^{k+m}} \left|T^{\mathcal{P}^k_{n,nm}\,*} f\right|^{p'}, \tag{356}$$

which trivially translates into

$$\Big\|\sum_k T^{\mathcal{P}^k_{n,nm}{}^*} f\Big\|_{p'}^{p'} \lesssim_{p'} \sum_k \|T^{\mathcal{P}^k_{n,nm}{}^*} f\|_{p'}^{p'} + \sum_{\substack{m\geq 10p'\\ m\in\mathbb{N}}} \sum_k m^{100p'} \int_{A_n^{k+m}} \left|T^{\mathcal{P}^k_{n,nm}{}^*} f\right|^{p'}. \tag{357}$$

Notice that (357) is the precise formulation of the heuristic described in relation (355).

The next step is to treat the main term

$$A=\sum_k \|T^{\mathcal{P}^k_{n,nm}{}^*} f\|_{p'}^{p'}. \tag{358}$$

We first prove that it is enough to show that (354) holds for $\mathcal{P}^k_{n,nm}$ (uniformly in k), that is,

$$\left\|T^{\mathcal{P}^k_{n,nm}{}^*} f\right\|_{p'} \lesssim_{p'} 2^{-\frac{n\eta}{p'}}\,\|f\|_{p'}. \tag{359}$$

Indeed, assume for the moment that (359) holds.

Then, we first split the input of $T^{\mathcal{P}^k_{n,nm}{}^*}$ into disjoint sets $\{\chi_{A_n^{k+l}\setminus A_n^{k+l+1}}\}_{l\in\mathbb{N}}$ and notice that, based on (359), for any $l\in\mathbb{N}$, one has

$$\left\|T^{\mathcal{P}^k_{nm}{}^*}\chi_{A_n^{k+l}\setminus A_n^{k+l+1}} f\right\|_{p'} \lesssim_{p'} 2^{-\frac{n\eta}{p'}}\,\|\chi_{A_n^{k+l}\setminus A_n^{k+l+1}} f\|_{p'}. \tag{360}$$

For $l\geq 2$, however, we can do better; for this, we first apply the standard Hölder inequality relative to the row decomposition of $\mathcal{P}^k_{n,nm}$:

$$\|T^{\mathcal{P}^k_{n,nm}{}^*} f\|_{p'} \lesssim_{p'} (2^n)^{\frac{1}{p}} \left\{\sum_{j=1}^{c2^n} \|T^{\mathcal{R}^k_j{}^*} f\|_{p'}^{p'}\right\}^{\frac{1}{p'}}. \tag{361}$$

With these, from (348) and Lemma 3, we deduce

$$\|T^{\mathcal{R}^k_j{}^*}(\chi_{A_n^{k+l}\setminus A_n^{k+l+1}}\,\cdot)\|_{p'} = \|\chi_{A_n^{k+l}\setminus A_n^{k+l+1}}\,T^{\mathcal{R}^k_j}\|_p \lesssim_p 2^{-\frac{l\,n}{p}}. \tag{362}$$

Now denoting $E^k_j := \bigcup_{P\in\mathcal{R}^k_j} E(P)$ and using that $\operatorname{supp} T^{\mathcal{R}^k_j} \subseteq E^k_j$ with $\{E^k_j\}_j$ pairwise disjoint, we have from (361) and (362) that

$$\|T^{\mathcal{P}^k_{n,nm}{}^*}\chi_{A_n^{k+l}\setminus A_n^{k+l+1}} f\|_{p'} \lesssim_{p'} 2^{-\frac{n(l-1)}{p}}\,\|\chi_{A_n^{k+l}\setminus A_n^{k+l+1}} f\|_{p'}. \tag{363}$$

Deduce from (360), (363), and Hölder's inequality that

$$\|T^{\mathcal{P}^k_{n,nm}{}^*} f\|_{p'}^{p'} \lesssim_{p'} \sum_{l\in\mathbb{N}} (l+1)^{p'}\,2^{-n\eta}\,\min\{1, 2^{-\frac{n(l-1-\eta)p'}{p}}\}\,\|\chi_{A_n^{k+l}\setminus A_n^{k+l+1}} f\|_{p'}^{p'}. \tag{364}$$

Now replacing (364) in (358) and summing over k we conclude that

$$A \lesssim_{p'} 2^{-n\eta}\, \|f\|_{p'}^{p'}. \tag{365}$$

Returning now to the proof of (359), the simplest approach[76] is provided by the following short argument, which holds uniformly in k:

- For the case $p = p' = 2$, we already know that (359) holds from the Main Lemma (or equivalently from (346)).
- For $p' \in 2\mathbb{N} \setminus \{0\}$, one can simply apply (361) together with Lemma 4 for $A := \mathbb{T}$ in order to deduce the trivial bound

$$\|T^{\mathcal{P}_{n,nm}^{k}{}^{*}} f\|_{p'} \lesssim_{p'} \|f\|_{p'}. \tag{366}$$

Conclude from the above using standard interpolation that (359) holds.

We pass now to the error term

$$B := \sum_{m \geq 10\,p'} \sum_{k} m^{100p'} \int_{A_n^{k+m}} \left| T^{\mathcal{P}_{n,nm}^{k}{}^{*}} f \right|^{p'}.$$

We first notice that

$$\int_{A_n^{k+m}} \left| T^{\mathcal{P}_{n,nm}^{k}{}^{*}} f \right|^{p'} \lesssim_{p'} (2^n)^{p'} \sum_{j=1}^{c2^n} \int_{A_n^{k+m}} \left| T^{\mathcal{R}_j^k{}^{*}} f \right|^{p'}. \tag{367}$$

Now, based on (348) and Lemma 4, we deduce that for each j, we have

$$\int_{A_n^{k+m}} \left| T^{\mathcal{R}_j^k{}^{*}} f \right|^{p'} \lesssim 2^{-m\,n}\, \|f\|_{p'}^{p'}. \tag{368}$$

Combining (362) with (368) we further have, for $m,\, l \geq 0$,

$$\int_{A_n^{k+m}} \left| T^{\mathcal{R}_j^k{}^{*}} \chi_{A_n^{k+l} \setminus A_n^{k+l+1}}\, f \right|^{p'} \lesssim_{p'} 2^{-\frac{l\,n\,p'}{2p}}\, 2^{-\frac{m\,n}{2}}\, \|\chi_{A_n^{k+l} \setminus A_n^{k+l+1}}\, f\|_{p'}^{p'}. \tag{369}$$

Next, proceeding in a similar fashion with (364), we have

$$\int_{A_n^{k+m}} \left| T^{\mathcal{R}_j^k{}^{*}}\, f \right|^{p'} \lesssim_{p'} 2^{-\frac{m\,n}{2}} \sum_{l \in \mathbb{N}} (l+1)^{p'}\, 2^{-\frac{l\,n\,p'}{2p}}\, \|\chi_{A_n^{k+l} \setminus A_n^{k+l+1}}\, \chi_{E_j^k} f\|_{p'}^{p'}. \tag{370}$$

Putting together (367) and (370) we deduce that

$$B \lesssim_{p'} \sum_{m \geq 10\,p'} m^{100p'}\, 2^{-\frac{m\,n}{2}}\, 2^{n\,p'} \sum_{k} \sum_{l \in \mathbb{N}} (l+1)^{p'}\, 2^{-\frac{l\,n\,p'}{2p}}\, \|\chi_{A_n^{k+l} \setminus A_n^{k+l+1}}\, f\|_{p'}^{p'},$$

and hence

$$B \lesssim_{p'} 2^{-n}\, \|f\|_{p'}^{p'}. \tag{371}$$

Finally, from (365) and (371), we conclude that (353) holds.

[76]For a different approach, please see the fifth remark in Section 11.

Case 2: *Assume* $2 < p < \infty$. In this situation we have that $p^* = p'$, and hence (353) is equivalent with

$$\left\|T^{\mathcal{P}_{nm}}\right\|_p \lesssim_p 2^{-\frac{n\,\eta}{p}}. \tag{372}$$

Once at this point, we notice that we can follow line by line the same arguments as in Case 1 by simply dropping the adjoint symbol in the corresponding proof. The key aspect that allows us to work with only this simple modification is that all the L^∞-forests appearing in the reasonings from Case 1 consist of *normal* trees and hence $\operatorname{supp} T^{\mathcal{P}^k_{n,nm}} \subseteq A^k_n$. □

11. **Final remarks**

In this final section of our paper we focus on several themes related with the main topic of our paper among which the most relevant are

- the local analysis methodology as developed in Section 5 and some of its more recent connections and applications;
- the resolution of the general n-dimensional case of the conjecture on the Polynomial Carleson operator; and
- the long standing open problem on the convergence of Fourier Series near L^1.

(1) As revealed in Section 5, the local analysis emerging from the study of the local properties of the concepts of mass and counting function played an essential role in our tile discretization, elimination of exceptional sets, and finally, in the slick argument that realizes the passage from the $p = 2$ case to the complete L^p range, $1 < p < \infty$. Its motivation and genesis were explained in Section 5.2. In this remark, we will place this local analysis into a historical context and mention some of its more recent related manifestations together with their applications.

As discussed in extenso in the introduction, the resolution of Luzin's conjecture on the pointwise convergence of Fourier Series for square integrable functions can be equivalently rephrased in terms of the L^2-weak boundedness of the Carleson operator. There are three known proofs of this result, each with its own particularities and subtleties:

- First is the proof of Carleson ([15]), which appeals to a fine analysis of the structure of the input function f by carefully studying the orthogonality and magnitude properties of its properly rescaled Fourier coefficients. While not formally stated as such, Carleson's approach is based in effect on an analysis and stopping time algorithm relative to the concept of *size* of f—informally, an l^2 average of local Fourier coefficients measuring essentially the BMO information carried by f in a certain region of the time-frequency plane.
- Secondly, the proof of Fefferman ([36]) moves the focus from the input f to the operator T itself by building up "a partial sum operator from simpler

pieces." This proof brings into play several fundamental concepts such as wave-packet decompositions and their geometric representation, structured family organization, and tile orderings. Further on, Fefferman's approach involves a stopping time-algorithm adapted to the concept of *mass*—with the latter measuring the amount of the graph of the linearizing phase function within each given tile.

- Finally, the third proof conceived three decades later by Lacey and Thiele ([73]) uses the modern language they developed for proving the boundedness of the Bilinear Hilbert transform ([71], [72]). Their approach is essentially a synthesis of the first two proofs discussed above, relying on a double stopping time argument involving simultaneously the concepts of size and mass.

At this point it is worth saying that beyond some subtle features, and despite carrying different numbers of stopping time applications, all three of these approaches are morally of equal strength within the canonical boundedness range $1 < p < \infty$.[77] This should not be surprising, since the mass and size are—morally speaking—in a duality relation.

The reader is by now familiar that all of the above proofs directly provide only the L^2 to $L^{2,\infty}$-boundedness of the Carleson operator. As already revealed in Section 5.2, the explanation for this common limitation is due to the fact that all these proofs require the removal of so-called exceptional sets. The deep reason for this is that both the mass and size stopping times used in any of [15], [36] or [73] do not exhibit a *localized-output character*: that is, if one fixes the magnitude of either the size or mass of a collection of tiles within any of these algorithms, one is not able to specify a non-trivial spatial region within which all the time intervals of these tiles cluster.[78]

With this picture in mind, we can now better understand the paradigm shift introduced in the present paper that moves the focus from the *global* to the *local* properties of the relevant concepts—which in the case of the present approach are exemplified by the concepts of mass and counting functions. This motivates the reference to the *local analysis* methodology outlined in Section 5 and developed throughout the paper.

As it turns out, several years after the original version of this paper was made available on arXiv in 2011, the type of local analysis developed here found, in closely related forms, interesting applications:

[77]However, as p approaches 1, meaningful, non-trivial differences arise among these three methods, as shown by the author in [85] (see Section 12 therein). For more on the difficult open problem of the pointwise convergence of Fourier Series near L^1, please read the last remark.

[78]For more on this, please consult Section 5.2.2.

- In [20] and [27] the authors use a local analysis formulated in the language of outer measures (with the latter introduced in [31])—referred to as "localized outer L^p embeddings"—in order to prove sparse domination[79] results for a class of multilinear singular integrals that include the Bilinear Hilbert transform and variational Carleson operators, respectively.
- In [7], [8], and [9] the authors develop the so-called "helicoidal method," which has as its very heart the concept of "localized size estimates". Indeed, this method relies on an iterative algorithm that—in the language used by the authors—constructs depth-$(m+1)$ localization results from their depth-m correspondents. As it turns out, the "depth-0 localization size-estimates" are, departing from the case of the Carleson operator, a direct analogue of the "localized mass estimates" (86) developed here but now phrased in the language used by Lacey and Thiele in [73]. Indeed, by simply dualizing and writing $\Lambda(f,g) := \langle T^{\mathbb{P}} f,\, g\rangle$, with $T = T^{\mathbb{P}}$ say the Carleson operator, at an informal level
  (i) one first introduces a concept of localized size that corresponds to the "localized mass" Definition 81 here (see, e.g., Definition 14 in [9], Definition 9 in [7], or Definition 19 in [8]);
  (ii) one proves a generic *localized estimate*—analogous to (86) here—for the form $\Lambda_{\mathfrak{I}}(f,g) := \left\langle T^{\mathbb{P}(\mathfrak{I})} f,\, g\right\rangle$, where here the location $\mathfrak{I}$ is some (dyadic) interval and $\mathbb{P}(\mathfrak{I})$ are the tiles supported within $\mathfrak{I}$ (see, e.g., Theorem 19 (for $n=1$) or Lemma 27 in [9] or Lemma 21 in [8]); and
  (iii) one runs a stopping time algorithm in order to create a sparse family of intervals $\{\mathcal{I}_l\}_l$ having similar[80] construction and properties as those in (80) (see, e.g., the proof of Theorem 12 in [9] or the proof of Theorem 7 in [7]).

  At the end of this process one obtains the analogue of (89), i.e., $\Lambda(f,g) = \sum_l \sum_{\mathfrak{I}\in\mathcal{I}^l} \Lambda_{\mathfrak{I}}(f,g)$. This latter decomposition, in particular, immediately implies sparse domination results for the scalar case under discussion. That this approach can be extended to treat many other (multi)linear and maximal operators T within the realm of time-frequency analysis is an insight of Benea and Muscalu. Moreover, via the helicoidal method, they provide a general and robust framework for transforming the localized size estimates

---

[79] See [77], [78] and [79] for the origin of sparse domination theory.

[80] One can adapt the stopping time construction of $\{\mathcal{I}_l\}_l$ to other relevant concepts that depend on the nature of the problem and one's favorite approach. In our paper the stopping time procedure is conducted relative to the level sets of suitable counting functions, while in [7], [8], and [9] the stopping time algorithm is performed relative to the level sets of the maximal Hardy-Littlewood operator associated to the input functions f and g. This latter is the natural analogous approach in the Lacey-Thiele ([73]) mass-size framework.

for a given operator into corresponding sparse form estimates and weighted (multiple) vector-valued extensions.

(2) During the years that have elapsed since the first version of the present paper was available on arXiv, the full n-dimensional conjecture on the Polynomial Carleson operator $C_{d,n}$ was solved in a few stages: in 2017, building on the ideas and methods in the present paper, Zorin-Kranich ([128]) proved the L^p-boundedness of $C_{d,n}$ for general $n \geq 1$ and $2 \leq p < \infty$ and for Calderón–Zygmund kernels in (6) that are not necessarily translation-invariant. A month later, based on the observation in [128] that one can apply directly the Van der Corput estimates proved in [121] in order to estimate the L^2-interaction of two trees (see Observation 21 in the present paper), we showed in [84] that the one-dimensional techniques developed in the present paper can easily be extended to the higher dimensional case in order to provide the full L^p-boundedness range, $1 < p < \infty$, for $C_{d,n}$ as defined by (1), thus providing an approach dealing only with translation-invariant Calderón–Zygmund kernels. Soon thereafter, Zorin-Kranich issued a new version of [128] that extended his initial result in order to cover the full L^p range for the more general non-translation invariant case.

(3) This remark is a consequence of a fruitful conversation that the author had with C. Thiele and M. Bateman, and it refers to a vector-valued variant of the Carleson Theorem. More precisely, using the tile-partitioning algorithm developed in Section 6.2, we devised an alternative proof that for any $1 < p, q < \infty$, one has[81]

$$\Big\|\Big(\sum_k |Cf_k|^q\Big)^{\frac{1}{q}}\Big\|_p \lesssim_{p,q} \Big\|\Big(\sum_k |f_k|^q\Big)^{\frac{1}{q}}\Big\|_p, \tag{373}$$

an inequality that had been proven in [46] using weighted and extrapolation theory. However, as a consequence of the Main Theorem presented in this paper, one can extend (373) to the situation when $C = C_{1,1}$ is replaced by $C_{d,1}$.

(4) It is worth mentioning that via the local analysis employed in the present paper, we obtain the following (informal) upgrade of (75):

If $\mathcal{P} = \bigcup_k \mathcal{P}_k \subseteq \mathbb{P}_n$ is a collection of separated trees, and $\mathcal{N}_{\mathcal{P}}$ stands for the usual counting function associated with $\mathcal{P}$, then

$$\|\sum_k T^{\mathcal{P}_k{}^*} f\|_2 \lesssim \log(1 + \|\mathcal{N}_{\mathcal{P}}\|_{\mathrm{BMO}_C}) \left(\sum_k \|T^{\mathcal{P}_k{}^*} f\|_2^2\right)^{\frac{1}{2}}. \tag{374}$$

This is in contrast with both [24] and [36], where (374) is only present with $\|\mathcal{N}_{\mathcal{P}}\|_{\mathrm{BMO}_C}$ replaced by $\|\mathcal{N}_{\mathcal{P}}\|_{L^\infty}$. Based on the discussions in Section 5.2 and

[81] Here we use the notation from Section 1.

the first remark in this section, we deduce that (374) implies (and is essentially equivalent with) the elimination of exceptional sets.

(5) A different, more involved, but direct approach to (359) (i.e., not appealing to formal interpolation), was presented in an earlier version of this paper and was based on the following heuristic hinted at by an informal interpolation argument:

Recalling (327) in the Main Lemma, we know—ignoring the error term—that

$$\|T^{\mathcal{P}^k_{n,nm}}{}^*f\|_2 \lesssim \left(\sum_{j=1}^{c2^n}\|T^{\mathcal{R}^k_j}{}^*f\|_2^2\right)^{\frac{1}{2}}. \tag{375}$$

We also trivially have

$$\|T^{\mathcal{P}^k_{n,nm}}{}^*f\|_\infty \lesssim \sum_{j=1}^{c2^n}\|T^{\mathcal{R}^k_j}{}^*f\|_\infty.$$

Thus, at the heuristic level, we expect for any $2 \leq p' < \infty$ to have

$$\|T^{\mathcal{P}^k_{n,nm}}{}^*f\|_{p'} \lesssim_{p'} \left\{\sum_{j=1}^{c2^n}\|T^{\mathcal{R}^k_j}{}^*f\|_{p'}^{p}\right\}^{\frac{1}{p}} \lesssim (2^n)^{\frac{1}{p}-\frac{1}{p'}}\left\{\sum_{j=1}^{c2^n}\|T^{\mathcal{R}^k_j}{}^*f\|_{p'}^{p'}\right\}^{\frac{1}{p'}}. \tag{376}$$

The key message here is that one is able to decouple the information carried by the rows of a forest with a gain of $(2^n)^{-\frac{1}{p'}}$ over the Hölder bound. Notice at this point that any gain over the trivial bound $(2^n)^{\frac{1}{p}}$ would be enough for our claim (359).

The precise form of this decoupling argument is given by the following:

Observation 24. Let $p' \in 2\mathbb{N}$, $p' \geq 2$, and $\mathcal{P}^k_{n,nm}$ be an L^∞-forest of generation n whose standard decomposition into rows is given by $\{\mathcal{R}^k_j\}_{j=1}^{c2^n}$. Assume that any two distinct trees within this row decomposition are normal and $2^{100\,n\,d\,p'}$-separated. Then there exists $\eta \in (0,1)$ such that the following holds:

$$\|T^{\mathcal{P}^k_{n,nm}}{}^*f\|_{p'} \lesssim_{p'} (2^n)^{\frac{1}{p}-\frac{\eta}{p'}}\left\{\sum_{j=1}^{c2^n}\|T^{\mathcal{R}^k_j}{}^*f\|_{p'}^{p'}\right\}^{\frac{1}{p'}} + 2^{-\frac{10n}{p'}}\|f\|_{p'}. \tag{377}$$

The proof of this statement relies crucially on the separateness assumption of the trees, which is further reflected in the time-frequency localization properties of each of the maximal trees belonging to the forest. The appeal of this approach is that it provides the desired $L^{p'}$-decay (for $p' \in 2\mathbb{N}$, $p' \geq 2$) in a direct fashion with no actual usage of interpolation methods.

(6) This remark is dedicated to an important feature of the behavior of the counting functions $\{\mathcal{N}_n\}_n$ as defined in (145). In [85], the author characterized the L^1-weak behavior of the so-called lacunary Carleson operator (see the

next remark). A key idea in that study was the understanding of the newly introduced concept of a *grand maximal function*, which in our current context is defined as follows:

Fix $j \in \mathbb{N}$, and set

$$\mathcal{N}(j) := \frac{1}{2^{j-1}} \sum_{n=2^{j-1}+1}^{2^j} \frac{1}{2^{n-1}} \mathcal{N}_n. \tag{378}$$

Define the grand maximal counting function of order $l \in \mathbb{N}$, $(l \geq 2)$ by

$$\mathcal{N}^{[l]} := \sup_{j \leq l} \mathcal{N}(j). \tag{379}$$

With these we have the following key property:

$$\|\mathcal{N}^{[l]}\|_{1,\infty} \lesssim \log l, \tag{380}$$

with the right-hand side bound being *sharp*.

In particular, the existence of extremal configurations of tiles that realize the reverse inequality

$$\|\mathcal{N}^{[l]}\|_{1,\infty} \gtrsim \log l \tag{381}$$

is responsible for the iterative stopping time moment chosen in our tile construction algorithm. That is, in order to achieve (80b) (and later (98)), one is required in (143) to cut the level set at height

$$c\, n\, \|\mathcal{N}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}}, \ldots, A_1^{j_1}]]\|_{\mathrm{BMO}_C}$$

instead of the simpler

$$c\, \|\mathcal{N}_n[A_n^{k-1}[A_{n-1}^{j_{n-1}}, \ldots, A_1^{j_1}]]\|_{\mathrm{BMO}_C}.$$

(7) Finally, the previous remark connects with the behavior of the Carleson operator C near L^1, a deep and fundamental theme in harmonic analysis.

At the foundation of this theme resides the following heuristic question:[82]

What is the behavior of the (almost-everywhere) pointwise convergence of Fourier Series between the two known cases for the Lebesgue-scale spaces $L^p(\mathbb{T})$:

- $p = 1$, *divergence of Fourier Series (Kolmogorov)*;
- $p > 1$, *convergence of Fourier Series (Carleson–Hunt)*?

Now using the fact that the pointwise convergence of Fourier Series is directly related to the $L^{1,\infty}$ behavior of the Carleson operator, one can reformulate the above vague question into a precise problem:

[82] In what follows we adopt the formalism from our paper [85].

Problem: (1) Let $Y \subseteq L^1(\mathbb{T})$ be a rearrangement-invariant (quasi-)Banach space. Provide necessary and sufficient conditions for Y to be a C-space, that is, there exists $c > 0$ such that

$$\|Cf\|_{1,\infty} \leq c\,\|f\|_Y. \tag{382}$$

(2) In Lorentz space terminology, the above can be expressed as follows: Give a satisfactory description of the Lorentz spaces $Y \subseteq L^1(\mathbb{T})$ that are also C-spaces. If such exists, describe the maximal Lorentz $\mathcal{C}$-space Y_0. In terms

of known results we have two possible directions:

- On the *negative* side (i.e., aiming for successively smaller Banach rearrangement-invariant spaces that are not $\mathcal{C}$-spaces): As mentioned above, the history of this direction starts with the result of Kolmogorov, showing that $L^1(\mathbb{T})$ is not a C-space. The next results are due to Chen ([16]), Prohorenko ([103]), and Körner ([65]). The best result to date belongs to Konyagin ([63], [64]), who proved that for $\phi(u) = o(u\sqrt{\frac{\log u}{\log\log u}})$, as $u \to \infty$ the space $X = \phi(L)$ does not admit pointwise convergence.
- On the *positive* side (i.e., identifying increasingly large C-spaces Y): Historically, the topic starts with the results of Carleson and Hunt for $Y = L^p(\mathbb{T})$, $p > 1$. Next, Sjölin ([109]) showed that one can take $Y = L\log L\log\log L$, while F. Soria ([113], [114]) increased Y to a rearrangement-invariant quasi-Banach space denoted B^*_φ. The best current results belong to Antonov ([2]) for the Lorentz space $Y = L\log L\log\log\log L$ and to J. Arias de Reyna ([3]) for the quasi-Banach space $Y = QA$ (a rearrangement-invariant quasi-Banach space that essentially has as its largest possible Lorentz subspace precisely Antonov's space).

In this context, there are several points worth mentioning:

- The proofs of all of the positive results mentioned above rely on extrapolation methods.
- In [83] the author reproved all of the above positive results via a unified method relying solely on time-frequency tools.
- Currently, all the positive results can be explained entirely based on the behavior of the grand maximal counting function (of order l) (379), more precisely on the fact that there are configurations of tiles for which inequality (381) holds (extremizers). In order to explain our claim, at least at the heuristic level, we use duality and write our Carleson operator as a bilinear form given by

$$\Lambda(f,g) := < Cf,\, g > \approx \sum_n \sum_{\substack{\mathcal{P}_k \subset \mathbb{P}_n \\ \mathcal{P}_k\ L^\infty-\text{forest}}} \sum_{\substack{\mathcal{P} \subseteq \mathcal{P}_k \\ \mathcal{P}\ \text{maximal tree}}} \langle C^{\mathcal{P}} f,\, g\rangle. \tag{383}$$

The key issue is that currently there are no methods *near* L^1 to distinguish[83] between the absolute and the conditional summation in the right-hand side of (383). One can argue that if one assumes absolute summation in the right-hand side of (383), then Antonov's result is the best possible result and, further, it is a direct consequence of the logarithmic divergence of the $L^{1,\infty}$ norm of the grand maximal counting function (of order l).

- There exists an old model problem for the problem stated above that has its own history. This model problem regards the almost-everywhere convergence of lacunary sequences of partial Fourier sums and goes back to the early twentieth century in works of Kolmogorov (see [59]), Littlewood and Paley ([88]), and Zygmund ([129]). In the quest to identify the largest possible Lorentz space (or rearrangement-invariant quasi-Banach space) for which one has almost-everywhere convergence along *lacunary* sequences of partial Fourier sums, some partial progress has been made; see the works of Chen ([16]), Prohorenko ([103]), Körner ([65]), and later Konyagin ([63], [64]). Finally, a related investigation in the Walsh-Fourier setting was performed in [29]; see also [26].

  More recently, in [82] and [85], the author succeeded in giving a definitive answer to this model problem. Indeed, by defining the lacunary Carleson operator associated with an (arbitrary) lacunary sequence $\{n_j\}_j$ as

$$C_{\text{lac}}^{\{n_j\}_j} f(x) := \sup_{j\in\mathbb{N}} \left| \int_{\mathbb{T}} e^{i\,2\pi\, n_j\,(x-y)}\, \cot(\pi\,(x-y))\, f(y)\, dy \right|, \tag{384}$$

  we have that there exists $C_1 = C_1(\{n_j\}_j) > 0$ such that

$$\|C_{\text{lac}}^{\{n_j\}_j} f\|_{1,\infty} \le C_1\, \|f\|_{L\log\log L\log\log\log\log L}\,, \tag{385}$$

  and moreover that this result is essentially *sharp*. The proof relies in a key fashion on the properties of the grand maximal counting function explained above.

  Also, very recently, in [86], we provided the sharp result regarding the strong L^1 bound for the lacunary Carleson operator, that is, there exists $C_2 = C_2(\{n_j\}_j) > 0$ such that

$$\|C_{\text{lac}}^{\{n_j\}_j} f\|_{1} \le C_2\, \|f\|_{L\log L}\,. \tag{386}$$

Returning now to the original problem of the pointwise convergence of the full sequence of partial sums, we mention that the recent works [85] and [86] revealed certain subtle key points:

[83] There is a similar problem regarding the maximal boundedness range for the Bilinear Hilbert transform; see Section 1.4.3.

- The *structure of the frequencies* of the trees involved in the time-frequency decomposition of the Carleson operator plays a fundamental role in identifying larger classes of rearrangement-invariant Banach spaces for which we have pointwise convergence.
- The *structure of the input function* creates certain "resonances" with the aforementioned structure of the frequencies.

As a consequence, we expect that structural theorems from additive combinatorics will play a fundamental role in any relevant advancement on the problem.

We end by listing the three relevant main conjectures in this subject:

Conjecture 1 ($L^{1,\infty}$ behavior). *The largest Lorentz space* $Y_0 \subseteq L^1(\mathbb{T})$ *such that there exists* $c = c_{Y_0} > 0$ *with*

$$\|Cf\|_{1,\infty} \leq c\,\|f\|_{Y_0} \tag{387}$$

is $Y_0 = L\sqrt{\log L}$.

Conjecture 2 (L^1 behavior). *The largest Lorentz space* $Y_1 \subseteq L^1(\mathbb{T})$ *such that there exists* $c = c_{Y_1} > 0$ *with*

$$\|Cf\|_{1} \leq c\,\|f\|_{Y_1} \tag{388}$$

is $Y_1 = L\log L$.

If true, Conjecture 1 is essentially sharp due to the result of Konyagin ([63], [64]), while if Conjecture 2 is true, then it is definitely sharp due to the fact that both the Hardy–Littlewood maximal operator and the Hilbert transform map $L\log L$ into L^1 sharply.

Finally, we present a last conjecture; although weaker than both conjectures above, its resolution would still be a major breakthrough in the field of time-frequency analysis due to the new methods that one would need to develop:

Conjecture 3 ($L^{1,\infty}$ intermediate behavior). *Prove that there exits a constant* $c > 0$ *such that*

$$\|Cf\|_{1,\infty} \leq c\,\|f\|_{L\log L}. \tag{389}$$

12. Appendix: Results on the L^∞-distribution of polynomials

In this last section we present few useful results about controlling the growth and the size of the level sets of a given polynomial of a fix degree. At the hearth of all our three lemmas below stays the classical Lagrange interpolation formula.

LEMMA A. *If* $q \in \mathcal{Q}_{d-1}$, $d \geq 1$, *and* I, $J \subseteq \mathbb{T}$ *are some non-degenerate*[84] *intervals* (*not necessarily dyadic*) *obeying* $I \supseteq J$, *then there exists a positive constant* $c(d) \leq (2d)^d$ *such that*

$$\|q\|_{L^\infty(I)} \leq c(d) \left(\frac{|I|}{|J|}\right)^{d-1} \|q\|_{L^\infty(J)} .$$

Proof. Let $\{x_J^k\}_{k\in\{1,\dots,d\}}$ be obtained as in the procedure described in Section 2. Then, since $q \in \mathcal{Q}_{d-1}$, for any $x \in I$ we have that

$$q(x) := \sum_{j=1}^{d} \frac{\prod_{\substack{k=1\\k\neq j}}^{d} (x - x_J^k)}{\prod_{\substack{k=1\\k\neq j}}^{d} (x_J^j - x_J^k)}\, q(x_J^j). \tag{390}$$

As a consequence,

$$\|q\|_{L^\infty(I)} \leq d\ \|q\|_{L^\infty(J)} \sup_{\substack{j\\x\in I}} \left| \frac{\prod_{\substack{k=1\\k\neq j}}^{d} (x - x_J^k)}{\prod_{\substack{k=1\\k\neq j}}^{d} (x_J^j - x_J^k)} \right| \leq d\ \|q\|_{L^\infty(J)}\ \frac{|I|^{d-1}}{(\frac{|J|}{2d})^{d-1}}. \qquad \square$$

LEMMA B. *If* $q \in \mathcal{Q}_{d-1}$, $d \geq 2$, $\eta > 0$, *and* $I \subseteq \mathbb{T}$ *is some* (*dyadic*) *interval*, *then there exists* $0 < c(d) \leq 100d^2$ *such that*

$$|\{y \in I \mid |q(y)| < \eta\}| \leq c(d) \left(\frac{\eta}{\|q\|_{L^\infty(I)}}\right)^{\frac{1}{d-1}} |I|. \tag{391}$$

Proof. The set $A_\eta = \{y \in I \mid |q(y)| < \eta\}$ is the pre-image of $(-\eta, \eta)$ under a polynomial of degree $d-1$, so it can be written as

$$A_\eta = \bigcup_{k=1}^{r} J_k(\eta),$$

where $r \in \mathbb{N}$, $r \leq d-1$, and $\{J_k(\eta)\}_k$ are open intervals. Now all that remains is to apply Lemma A with $J = J_k(\eta)$ for each k. $\square$

LEMMA C. *If* $P = [\alpha^1, \alpha^2, \dots, \alpha^d, I] \in \mathbb{P}$ *and* $q \in P$ *with* $d \geq 1$, *then there exists* $0 < c(d) \leq (100\, d)^d$ *such that*

$$\|q - q_P\|_{L^\infty(\tilde{I})} \leq c(d)\, |I|^{-1}.$$

Proof. Set $u := q - q_P$; then, since both q, $q_P \in P$, we deduce that for all $k \in \{1, \dots, d\}$, one has

$$u(x_I^k) \in [-|I|^{-1}, |I|^{-1}].$$

[84] That is, $|I|$, $|J| > 0$.

On the other hand,

$$u(x) := \sum_{j=1}^{d} \frac{\prod_{\substack{k=1\\k\neq j}}^{d} (x - x_I^k)}{\prod_{\substack{k=1\\k\neq j}}^{d} (x_I^j - x_I^k)}\, u(x_I^j) \qquad \forall\ x \in \tilde{I}.$$

Then, proceeding as in Lemma A, we conclude

$$\|u\|_{L^\infty(\tilde{I})} \leq d\,|I|^{-1}\,\frac{|\tilde{I}|^{d-1}}{(\frac{|I|}{2\,d})^{d-1}} \leq (100\,d)^d\,|I|^{-1}. \qquad \square$$

References

[1] L. Alpár, Sur certaines transformées des séries de puissance absolument convergentes sur la frontière de leur cercle de convergence, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 287–316. MR 0152640. Zbl 0115.28601.

[2] N. Y. Antonov, Convergence of Fourier series, in *Proceedings of the XX Workshop on Function Theory* (*Moscow*, 1995), **2** East J. Approx. no. 2, 1996, pp. 187–196. MR 1407066. Zbl 0919.42004.

[3] J. Arias-de-Reyna, Pointwise convergence of Fourier series, *J. London Math. Soc.* (2) **65** no. 1 (2002), 139–153. MR 1875141. Zbl 1017.42002. https://doi.org/10.1112/S0024610701002824.

[4] T. Austin, On the norm convergence of non-conventional ergodic averages, *Ergodic Theory Dynam. Systems* **30** no. 2 (2010), 321–338. MR 2599882. Zbl 1206.37003. https://doi.org/10.1017/S014338570900011X.

[5] M. Bateman, Single annulus L^p estimates for Hilbert transforms along vector fields, *Rev. Mat. Iberoam.* **29** no. 3 (2013), 1021–1069. MR 3090145. Zbl 1283.42019. https://doi.org/10.4171/RMI/748.

[6] M. Bateman and C. Thiele, L^p estimates for the Hilbert transforms along a one-variable vector field, *Anal. PDE* **6** no. 7 (2013), 1577–1600. MR 3148061. Zbl 1285.42014. https://doi.org/10.2140/apde.2013.6.1577.

[7] C. Benea and C. Muscalu, Multiple vector-valued inequalities via the helicoidal method, *Anal. PDE* **9** no. 8 (2016), 1931–1988. MR 3599522. Zbl 1361.42005. https://doi.org/10.2140/apde.2016.9.1931.

[8] C. Benea and C. Muscalu, Quasi-Banach valued inequalities via the helicoidal method, *J. Funct. Anal.* **273** no. 4 (2017), 1295–1353. MR 3661402. Zbl 1392.42024. https://doi.org/10.1016/j.jfa.2017.04.014.

[9] C. Benea and C. Muscalu, Sparse domination via the helicoidal method, 2017. arXiv 1707.05484.

[10] G. D. Birkhoff, Proof of the ergodic theorem, *Proc. Nat. Acad. Sci. U.S.A.* **17** no. 12 (1931), 656–660. Zbl 57.1011.02. https://doi.org/10.1073/pnas.17.2.656.

[11] J. Bourgain, Temps de retour pour les systèmes dynamiques (return times of dynamical systems), *C. R. Acad. Sci. Paris Sér. I Math.* **306** no. 12 (1988), 483–485. MR 0939436. Zbl 0655.58022.

[12] J. Bourgain, A remark on the maximal function associated to an analytic vector field, in *Analysis at Urbana, Vol. I* (Urbana, IL, 1986–1987), *London Math. Soc. Lecture Note Ser.* **137**, Cambridge Univ. Press, Cambridge, 1989, pp. 111–132. MR 1009171. Zbl 0692.42006. https://doi.org/10.1017/CBO9780511662294.006.

[13] J. Bourgain, Double recurrence and almost sure convergence, *J. Reine Angew. Math.* **404** (1990), 140–161. MR 1037434. Zbl 0685.28008. https://doi.org/10.1515/crll.1990.404.140.

[14] A.-P. Calderón, Cauchy integrals on Lipschitz curves and related operators, *Proc. Nat. Acad. Sci. U.S.A.* **74** no. 4 (1977), 1324–1327. MR 0466568. Zbl 0373.44003. https://doi.org/10.1073/pnas.74.4.1324.

[15] L. Carleson, On convergence and growth of partial sums of Fourier series, *Acta Math.* **116** (1966), 135–157. MR 0199631. Zbl 0144.06402. https://doi.org/10.1007/BF02392815.

[16] Y. M. Chen, An almost everywhere divergent Fourier series of the class $L(\log^+ \log^+ L)^{1-\varepsilon}$, *J. London Math. Soc.* **44** (1969), 643–654. MR 0240539. Zbl 0169.39703. https://doi.org/10.1112/jlms/s1-44.1.643.

[17] M. Christ, Hilbert transforms along curves. I. Nilpotent groups, *Ann. of Math.* (2) **122** no. 3 (1985), 575–596. MR 0819558. Zbl 0593.43011. https://doi.org/10.2307/1971330.

[18] M. Christ, A. Nagel, E. M. Stein, and S. Wainger, Singular and maximal Radon transforms: analysis and geometry, *Ann. of Math.* (2) **150** no. 2 (1999), 489–577. MR 1726701. Zbl 0960.44001. https://doi.org/10.2307/121088.

[19] J.-P. Conze and E. Lesigne, Théorèmes ergodiques pour des mesures diagonales, *Bull. Soc. Math. France* **112** no. 2 (1984), 143–175. MR 0788966. Zbl 0595.28018. https://doi.org/10.24033/bsmf.2003.

[20] A. Culiuc, F. Di Plinio, and Y. Ou, Domination of multilinear singular integrals by positive sparse forms, *J. Lond. Math. Soc.* (2) **98** no. 2 (2018), 369–392. MR 3873113. Zbl 1402.42013. https://doi.org/10.1112/jlms.12139.

[21] C. Demeter, Pointwise convergence of the ergodic bilinear Hilbert transform, *Illinois J. Math.* **51** no. 4 (2007), 1123–1158. MR 2417419. Zbl 1175.37013. https://doi.org/10.1215/ijm/1258138536.

[22] C. Demeter, Divergence of combinatorial averages and the unboundedness of the trilinear Hilbert transform, *Ergodic Theory Dynam. Systems* **28** no. 5 (2008), 1453–1464. MR 2449537. Zbl 1168.42007. https://doi.org/10.1017/S0143385707001101.

[23] C. Demeter, M. T. Lacey, T. Tao, and C. Thiele, Breaking the duality in the return times theorem, *Duke Math. J.* **143** no. 2 (2008), 281–355. MR 2420509. Zbl 1213.42064. https://doi.org/10.1215/00127094-2008-020.

[24] C. Demeter, T. Tao, and C. Thiele, Maximal multilinear operators, *Trans. Amer. Math. Soc.* **360** no. 9 (2008), 4989–5042. MR 2403711. Zbl 1268.42034. https://doi.org/10.1090/S0002-9947-08-04474-7.

[25] C. Demeter and C. Thiele, On the two-dimensional bilinear Hilbert transform, *Amer. J. Math.* **132** no. 1 (2010), 201–256. MR 2597511. Zbl 1206.42010. https://doi.org/10.1353/ajm.0.0101.

[26] F. Di Plinio, Lacunary Fourier and Walsh-Fourier series near L^1, *Collect. Math.* **65** no. 2 (2014), 219–232. MR 3189278. Zbl 1311.42008. https://doi.org/10.1007/s13348-013-0094-3.

[27] F. Di Plinio, Y. Q. Do, and G. N. Uraltsev, Positive sparse domination of variational Carleson operators, *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **18** no. 4 (2018), 1443–1458. MR 3829751. Zbl 1403.42010. https://doi.org/10.2422/2036-2145.201612_009.

[28] F. Di Plinio, S. Guo, C. Thiele, and P. Zorin-Kranich, Square functions for bi-Lipschitz maps and directional operators, *J. Funct. Anal.* **275** no. 8 (2018), 2015–2058. MR 3841536. Zbl 1400.42023. https://doi.org/10.1016/j.jfa.2018.07.005.

[29] Y. Do and M. T. Lacey, On the convergence of lacunacy Walsh-Fourier series, *Bull. Lond. Math. Soc.* **44** no. 2 (2012), 241–254. MR 2914604. Zbl 1241.42005. https://doi.org/10.1112/blms/bdr088.

[30] Y. Do, R. Oberlin, and E. A. Palsson, Variation-norm and fluctuation estimates for ergodic bilinear averages, *Indiana Univ. Math. J.* **66** no. 1 (2017), 55–99. MR 3623404. Zbl 1368.42005. https://doi.org/10.1512/iumj.2017.66.5983.

[31] Y. Do and C. Thiele, L^p theory for outer measures and two themes of Lennart Carleson united, *Bull. Amer. Math. Soc.* (*N.S.*) **52** no. 2 (2015), 249–296. MR 3312633. Zbl 1318.42016. https://doi.org/10.1090/S0273-0979-2014-01474-0.

[32] P. du Bois-Reymond, Untersuchungen über die Convergenz und Divergenz der Fourierschen Darstellungsformeln., *Münch. Abh.* **12** no. II (1876), 1–102. Zbl 09.0298.01.

[33] P. Durcik, V. Kovač, K. A. Škreb, and C. Thiele, Norm variation of ergodic averages with respect to two commuting transformations, *Ergodic Theory Dynam. Systems* **39** no. 3 (2019), 658–688. MR 3904183. Zbl 1416.37007. https://doi.org/10.1017/etds.2017.48.

[34] E. B. Fabes, Singular integrals and partial differential equations of parabolic type, *Studia Math.* **28** (1966/67), 81–131. MR 0213744. Zbl 0144.35002. https://doi.org/10.4064/sm-28-1-81-131.

[35] E. B. Fabes and N. M. Rivière, Singular integrals with mixed homogeneity, *Studia Math.* **27** (1966), 19–38. MR 0209787. Zbl 0161.32403. https://doi.org/10.4064/sm-27-1-19-38.

[36] C. Fefferman, Pointwise convergence of Fourier series, *Ann. of Math.* (2) **98** (1973), 551–571. MR 0340926. Zbl 0268.42009. https://doi.org/10.2307/1970917.

[37] C. L. Fefferman, The uncertainty principle, *Bull. Amer. Math. Soc.* (*N.S.*) **9** no. 2 (1983), 129–206. MR 0707957. Zbl 0526.35080. https://doi.org/10.1090/S0273-0979-1983-15154-6.

[38] J. Fourier, *Théorie Analytique de la Chaleur*, Éditions Jacques Gabay, Paris, 1988, reprint of the 1822 original. MR 1414430.

[39] G. A. Freĭman, *Foundations of a Structural Theory of Set Addition*, (in Russian), Kazan Gos. Ped. Inst., Kazan, 1966. MR 0360495. Zbl 0203.35305.

[40] G. A. Freĭman, *Foundations of a Structural Theory of Set Addition*, *Trans. Math. Monogr.* **37**, Amer. Math. Soc., Providence, R. I., 1973, translated from the Russian. MR 0360496. Zbl 0271.10044. https://doi.org/10.1090/mmono/037.

[41] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256. MR 0498471. Zbl 0347.28016. https://doi.org/10.1007/BF02813304.

[42] D. Geller and E. M. Stein, Estimates for singular convolution operators on the Heisenberg group, *Math. Ann.* **267** no. 1 (1984), 1–15. MR 0737332. Zbl 0537.43005. https://doi.org/10.1007/BF01458467.

[43] W. T. Gowers, A new proof of Szemerédi's theorem for arithmetic progressions of length four, *Geom. Funct. Anal.* **8** no. 3 (1998), 529–551. MR 1631259. Zbl 0907.11005. https://doi.org/10.1007/s000390050065.

[44] W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** no. 3 (2001), 465–588. MR 1844079. Zbl 1028.11005. https://doi.org/10.1007/s00039-001-0332-9.

[45] L. Grafakos and X. Li, Uniform bounds for the bilinear Hilbert transforms. I, *Ann. of Math.* (2) **159** no. 3 (2004), 889–933. MR 2113017. Zbl 1071.44004. https://doi.org/10.4007/annals.2004.159.889.

[46] L. Grafakos, J. M. Martell, and F. Soria, Weighted norm inequalities for maximally modulated singular integral operators, *Math. Ann.* **331** no. 2 (2005), 359–394. MR 2115460. Zbl 1133.42019. https://doi.org/10.1007/s00208-004-0586-2.

[47] L. Grafakos, T. Tao, and E. Terwilleger, L^p bounds for a maximal dyadic sum operator, *Math. Z.* **246** no. 1-2 (2004), 321–337. MR 2031458. Zbl 1076.42013. https://doi.org/10.1007/s00209-003-0601-4.

[48] B. Green and T. Tao, An inverse theorem for the Gowers $U^3(G)$ norm, *Proc. Edinb. Math. Soc.* (2) **51** no. 1 (2008), 73–153. MR 2391635. Zbl 1202.11013. https://doi.org/10.1017/S0013091505000325.

[49] B. Green, T. Tao, and T. Ziegler, An inverse theorem for the Gowers $U^{s+1}[N]$-norm, *Ann. of Math.* (2) **176** no. 2 (2012), 1231–1372. MR 2950773. Zbl 1282.11007. https://doi.org/10.4007/annals.2012.176.2.11.

[50] A. Greenleaf and G. Uhlmann, Estimates for singular Radon transforms and pseudodifferential operators with singular symbols, *J. Funct. Anal.* **89** no. 1 (1990), 202–232. MR 1040963. Zbl 0717.44001. https://doi.org/10.1016/0022-1236(90)90011-9.

[51] S. Guo, Hilbert transform along measurable vector fields constant on Lipschitz curves: L^2 boundedness, *Anal. PDE* **8** no. 5 (2015), 1263–1288. MR 3393679. Zbl 1323.42013. https://doi.org/10.2140/apde.2015.8.1263.

[52] S. Guo, Hilbert transform along measurable vector fields constant on Lipschitz curves: L^p boundedness, *Trans. Amer. Math. Soc.* **369** no. 4 (2017), 2493–2519. MR 3592519. Zbl 1356.42007. https://doi.org/10.1090/tran/6750.

[53] S. Guo, J. Hickman, V. Lie, and J. Roos, Maximal operators and Hilbert transforms along variable non-flat homogeneous curves, *Proc. Lond. Math. Soc.* (3) **115** no. 1 (2017), 177–219. MR 3669936. Zbl 1388.42039. https://doi.org/10.1112/plms.12037.

[54] G. Halász, On a theorem of L. Alpár concerning Fourier series of powers of certain functions, *Studia Sci. Math. Hungar.* **2** (1967), 67–72. MR 0211184. Zbl 0214.32001.

[55] B. Host and B. Kra, Nonconventional ergodic averages and nilmanifolds, *Ann. of Math.* (2) **161** no. 1 (2005), 397–488. MR 2150389. Zbl 1077.37002. https://doi.org/10.4007/annals.2005.161.397.

[56] R. A. Hunt, On the convergence of Fourier series, in *Orthogonal Expansions and their Continuous Analogues* (*Proc. Conf.*, *Edwardsville*, *Ill.*, 1967), Southern Illinois Univ. Press, Carbondale, Ill., 1968, pp. 235–255. MR 0238019. Zbl 0159.35701.

[57] B. F. Jones, Jr., A class of singular integrals, *Amer. J. Math.* **86** (1964), 441–462. MR 0161099. Zbl 0123.08501. https://doi.org/10.2307/2373175.

[58] G. A. Karagulyan, On unboundedness of maximal operators for directional Hilbert transforms, *Proc. Amer. Math. Soc.* **135** no. 10 (2007), 3133–3141. MR 2322743. Zbl 1162.42009. https://doi.org/10.1090/S0002-9939-07-08731-X.

[59] Y. Katznelson, *An introduction to harmonic analysis*, third ed., *Cambridge Mathematical Library*, Cambridge University Press, Cambridge, 2004. MR 2039503. Zbl 1055.43001. https://doi.org/10.1017/CBO9781139165372.

[60] R. Kaufman, Uniform convergence of Fourier series in harmonic analysis, *Studia Sci. Math. Hungar.* **10** no. 1-2 (1975), 81–83. MR 0435710. Zbl 0354.42004.

[61] A. N. Kolmogorov, Une serie de Fourier-Lebesgue divergente presque partout, *Fund. Math.* **4** (1923), 324–328. Zbl 49.0205.02. https://doi.org/10.4064/fm-4-1-324-328.

[62] A. N. Kolmogorov, Une serie de Fourier-Lebesgue divergente partout, *C. R. Acad. Sci. Paris* **183** (1926), 1327–1329. Zbl 52.0269.02.

[63] S. V. Konyagin, On divergence of trigonometric Fourier series everywhere, *C. R. Acad. Sci. Paris Sér. I Math.* **329** no. 8 (1999), 693–697. MR 1724074. Zbl 0942.42002. https://doi.org/10.1016/S0764-4442(00)88219-1.

[64] S. V. Konyagin, On the divergence everywhere of trigonometric Fourier series, *Mat. Sb.* **191** no. 1 (2000), 103–126. MR 1753494. Zbl 0967.42004. https://doi.org/10.1070/SM2000v191n01ABEH000449.

[65] T. W. Körner, Everywhere divergent Fourier series, *Colloq. Math.* **45** no. 1 (1981), 103–118 (1982). MR 0652607. Zbl 0491.42011. https://doi.org/10.4064/cm-45-1-103-118.

[66] V. Kovač, C. Thiele, and P. Zorin-Kranich, Dyadic triangular Hilbert transform of two general functions and one not too general function, *Forum*

Math. Sigma **3** (2015), e25, 27. MR 3482272. Zbl 1335.42009. https://doi.org/10.1017/fms.2015.25.

[67] M. Lacey, The bilinear maximal functions map into L^p for $2/3 < p \leq 1$, *Ann. of Math.* (2) **151** no. 1 (2000), 35–57. MR 1745019. Zbl 0967.47031. https://doi.org/10.2307/121111.

[68] M. Lacey, Carleson's theorem: proof, complements, variations, *Publ. Mat.* **48** no. 2 (2004), 251–307. MR 2091007. Zbl 1066.42003. https://doi.org/10.5565/PUBLMAT_48204_01.

[69] M. Lacey and X. Li, Maximal theorems for the directional Hilbert transform on the plane, *Trans. Amer. Math. Soc.* **358** no. 9 (2006), 4099–4117. MR 2219012. Zbl 1095.42010. https://doi.org/10.1090/S0002-9947-06-03869-4.

[70] M. Lacey and X. Li, On a conjecture of E. M. Stein on the Hilbert transform on vector fields, *Mem. Amer. Math. Soc.* **205** no. 965 (2010), viii+72. MR 2654385. Zbl 1190.42005. https://doi.org/10.1090/S0065-9266-10-00572-7.

[71] M. Lacey and C. Thiele, L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$, *Ann. of Math.* (2) **146** no. 3 (1997), 693–724. MR 1491450. Zbl 0914.46034. https://doi.org/10.2307/2952458.

[72] M. Lacey and C. Thiele, On Calderón's conjecture, *Ann. of Math.* (2) **149** no. 2 (1999), 475–496. MR 1689336. Zbl 0934.42012. https://doi.org/10.2307/120971.

[73] M. Lacey and C. Thiele, A proof of boundedness of the Carleson operator, *Math. Res. Lett.* **7** no. 4 (2000), 361–370. MR 1783613. Zbl 0966.42009. https://doi.org/10.4310/MRL.2000.v7.n4.a1.

[74] H. Lebesgue, Sur la divergence et la convergence non uniforme des séries de fourier, *C. R. Acad. Sci. Paris* **141** (1906), 875–877. Zbl 36.0331.02.

[75] H. Lebesgue, Sur les intégrales singulières, *Ann. Fac. Sci. Toulouse Sci. Math. Sci. Phys.* (3) **1** (1909), 25–117. MR 1508308. Zbl 41.0327.02. Available at http://www.numdam.org/item?id=AFST_1909_3_1__25_0.

[76] G. Lejeune Dirichlet, Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données, *J. Reine Angew. Math.* **4** (1829), 157–169. MR 1577724. Zbl 004.0148cj. https://doi.org/10.1515/crll.1829.4.157.

[77] A. K. Lerner, A pointwise estimate for the local sharp maximal function with applications to singular integrals, *Bull. Lond. Math. Soc.* **42** no. 5 (2010), 843–856. MR 2721744. Zbl 1203.42023. https://doi.org/10.1112/blms/bdq042.

[78] A. K. Lerner, On an estimate of Calderón-Zygmund operators by dyadic positive operators, *J. Anal. Math.* **121** (2013), 141–161. MR 3127380. Zbl 1285.42015. https://doi.org/10.1007/s11854-013-0030-1.

[79] A. K. Lerner, A simple proof of the A_2 conjecture, *Int. Math. Res. Not. IMRN* no. 14 (2013), 3159–3170. MR 3085756. Zbl 1318.42018. https://doi.org/10.1093/imrn/rns145.

[80] X. Li, Uniform bounds for the bilinear Hilbert transforms. II, *Rev. Mat. Iberoam.* **22** no. 3 (2006), 1069–1126. MR 2320411. Zbl 1133.42022. https://doi.org/10.4171/RMI/483.

[81] V. Lie, The (weak-L^2) boundedness of the quadratic Carleson operator, *Geom. Funct. Anal.* **19** no. 2 (2009), 457–497. MR 2545246. Zbl 1178.42007. https://doi.org/10.1007/s00039-009-0010-x.

[82] V. Lie, On the pointwise convergence of the sequence of partial Fourier sums along lacunary subsequences, *J. Funct. Anal.* **263** no. 11 (2012), 3391–3411. MR 2984070. Zbl 1259.42005. https://doi.org/10.1016/j.jfa.2012.08.013.

[83] V. Lie, On the boundedness of the Carleson operator near L^1, *Rev. Mat. Iberoam.* **29** no. 4 (2013), 1239–1262. MR 3148602. Zbl 1297.42011. https://doi.org/10.4171/RMI/755.

[84] V. Lie, A note on the polynomial Carleson operator in higher dimensions, 2017. arXiv 1712.03092.

[85] V. Lie, Pointwise convergence of Fourier series (I). On a conjecture of Konyagin, *J. Eur. Math. Soc.* (*JEMS*) **19** no. 6 (2017), 1655–1728. MR 3646873. Zbl 1372.42001. https://doi.org/10.4171/JEMS/703.

[86] V. Lie, The pointwise convergence of Fourier series (II). Strong L^1 case for the lacunary Carleson operator, *Adv. Math.* **357** (2019), 106831, 84. MR 4017921. Zbl 1429.42003. https://doi.org/10.1016/j.aim.2019.106831.

[87] V. Lie, A unified approach to three themes in harmonic analysis (I & II), 2019. arXiv 1902.03807.

[88] J. E. Littlewood and R. E. A. C. Paley, Theorems on Fourier series and power series, *J. London Math. Soc.* **6** no. 3 (1931), 230–233. MR 1574750. Zbl 0002.18803. https://doi.org/10.1112/jlms/s1-6.3.230.

[89] N. N. Luzin, *Integral i trigonometričeskiĭ ryad*, Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow-Leningrad, 1951, editing and commentary by N. K. Bari and D. E. Men′šov. MR 0048364. Zbl 0045.33101.

[90] G. Marletta and F. Ricci, Two-parameter maximal functions associated with homogeneous surfaces in $\mathbf{R}^n$, *Studia Math.* **130** no. 1 (1998), 53–65. MR 1623004. Zbl 0921.42014. Available at http://matwbn.icm.edu.pl/ksiazki/sm/sm130/sm13014.pdf.

[91] G. Mauceri, M. A. Picardello, and F. Ricci, A Hardy space associated with twisted convolution, *Adv. in Math.* **39** no. 3 (1981), 270–288. MR 0614164. Zbl 0503.46037. https://doi.org/10.1016/0001-8708(81)90004-9.

[92] D. Müller, Singular kernels supported by homogeneous submanifolds, *J. Reine Angew. Math.* **356** (1985), 90–118. MR 0779377. Zbl 0551.43005. https://doi.org/10.1515/crll.1985.356.90.

[93] C. Muscalu, J. Pipher, T. Tao, and C. Thiele, Bi-parameter paraproducts, *Acta Math.* **193** no. 2 (2004), 269–296. MR 2134868. Zbl 1087.42016. https://doi.org/10.1007/BF02392566.

[94] C. Muscalu, T. Tao, and C. Thiele, L^p estimates for the biest. I. The Walsh case, *Math. Ann.* **329** no. 3 (2004), 401–426. MR 2127984. Zbl 1073.42009. https://doi.org/10.1007/s00208-004-0518-1.

[95] C. Muscalu, T. Tao, and C. Thiele, L^p estimates for the biest. II. The Fourier case, *Math. Ann.* **329** no. 3 (2004), 427–461. MR 2127985. Zbl 1073.42010. https://doi.org/10.1007/s00208-003-0508-8.

[96] A. Nagel, N. Rivière, and S. Wainger, On Hilbert transforms along curves, *Bull. Amer. Math. Soc.* **80** (1974), 106–108. MR 0450899. Zbl 0293.44002. https://doi.org/10.1090/S0002-9904-1974-13374-4.

[97] A. Nagel, N. M. Rivière, and S. Wainger, On Hilbert transforms along curves. II, *Amer. J. Math.* **98** no. 2 (1976), 395–403. MR 0450900. Zbl 0334.44012. https://doi.org/10.2307/2373893.

[98] A. Nagel, E. M. Stein, and S. Wainger, Hilbert transforms and maximal functions related to variable curves, in *Harmonic Analysis in Euclidean Spaces* (Proc. Sympos. Pure Math., Williams Coll., Williamstown, Mass., 1978, Part 1), *Proc. Sympos. Pure Math., XXXV, Part*, Amer. Math. Soc., Providence, R.I., 1979, pp. 95–98. MR 0545242. Zbl 0463.42008. https://doi.org/10.1090/pspum/035.1.

[99] J. von Neumann, Proof of the quasi-ergodic hypothesis, *Proc. Nat. Acad. Sci. U.S.A.* **18** no. 1 (1932), 70–82. Zbl 58.1271.03. https://doi.org/10.1073/pnas.18.1.70.

[100] D. H. Phong and E. M. Stein, Singular integrals related to the Radon transform and boundary value problems, *Proc. Nat. Acad. Sci. U.S.A.* **80** no. 24, , Phys. Sci. (1983), 7697–7701. MR 0728667. Zbl 0567.42010. https://doi.org/10.1073/pnas.80.24.7697.

[101] D. H. Phong and E. M. Stein, Hilbert integrals, singular integrals, and Radon transforms. I, *Acta Math.* **157** no. 1-2 (1986), 99–157. MR 0857680. Zbl 0622.42011. https://doi.org/10.1007/BF02392592.

[102] M. Pramanik and E. Terwilleger, A weak L^2 estimate for a maximal dyadic sum operator on $\mathbb{R}^n$, *Illinois J. Math.* **47** no. 3 (2003), 775–813. MR 2007237. Zbl 1040.42014. https://doi.org/10.1215/ijm/1258138194.

[103] V. I. Prohorenko, Divergent Fourier series, *Mat. Sb.* (*N.S.*) **75 (117)** (1968), 185–198. MR 0223815. https://doi.org/10.1070/SM1968v004n02ABEH002786.

[104] F. Ricci and E. M. Stein, Oscillatory singular integrals and harmonic analysis on nilpotent groups, *Proc. Nat. Acad. Sci. U.S.A.* **83** no. 1 (1986), 1–3. MR 0822187. Zbl 0583.43010. https://doi.org/10.1073/pnas.83.1.1.

[105] F. Ricci and E. M. Stein, Harmonic analysis on nilpotent groups and singular integrals. I. Oscillatory integrals, *J. Funct. Anal.* **73** no. 1 (1987), 179–194. MR 0890662. Zbl 0622.42010. https://doi.org/10.1016/0022-1236(87)90064-4.

[106] F. Ricci and E. M. Stein, Harmonic analysis on nilpotent groups and singular integrals. III. Fractional integration along manifolds, *J. Funct. Anal.* **86** no. 2 (1989), 360–389. MR 1021141. Zbl 0684.22006. https://doi.org/10.1016/0022-1236(89)90057-8.

[107] M. Riesz, Sur les fonctions conjuguées, *Math. Z.* **27** no. 1 (1928), 218–244. MR 1544909. Zbl 53.0259.02. https://doi.org/10.1007/BF01171098.

[108] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109. MR 0051853. Zbl 0050.04002. https://doi.org/10.1112/jlms/s1-28.1.104.

[109] P. Sjölin, An inequality of Paley and convergence a.e. of Walsh-Fourier series, *Ark. Mat.* **7** (1969), 551–570. MR 0241885. Zbl 0169.08203. https://doi.org/10.1007/BF02590894.

[110] P. Sjölin, Convergence almost everywhere of certain singular integrals and multiple Fourier series, *Ark. Mat.* **9** (1971), 65–90. MR 0336222. Zbl 0212.41703. https://doi.org/10.1007/BF02383638.

[111] P. Sjölin, Convolution with oscillating kernels, *Indiana Univ. Math. J.* **30** no. 1 (1981), 47–55. MR 0600031. Zbl 0419.47020. https://doi.org/10.1512/iumj.1981.30.30004.

[112] P. Sjölin and F. Soria, Remarks on a theorem by N. Yu. Antonov, *Studia Math.* **158** no. 1 (2003), 79–97. MR 2014553. Zbl 1053.42005. https://doi.org/10.4064/sm158-1-7.

[113] F. Soria, Note on differentiation of integrals and the halo conjecture, *Studia Math.* **81** no. 1 (1985), 29–36. MR 0818168. Zbl 0526.28003. https://doi.org/10.4064/sm-81-1-29-36.

[114] F. Soria, On an extrapolation theorem of Carleson-Sjölin with applications to a.e. convergence of Fourier series, *Studia Math.* **94** no. 3 (1989), 235–244. MR 1019791. Zbl 0684.42001. https://doi.org/10.4064/sm-94-3-235-244.

[115] E. M. Stein, On limits of sequences of operators, *Ann. of Math.* (2) **74** (1961), 140–170. MR 0125392. Zbl 0103.08903. https://doi.org/10.2307/1970308.

[116] E. M. Stein, Maximal functions. I. Spherical means, *Proc. Nat. Acad. Sci. U.S.A.* **73** no. 7 (1976), 2174–2175. MR 0420116. Zbl 0332.42018. https://doi.org/10.1073/pnas.73.7.2174.

[117] E. M. Stein, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*, *Princeton Mathematical Series* **43**, Princeton University Press, Princeton, NJ, 1993, With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III. MR 1232192. Zbl 0821.42001. https://doi.org/10.1515/9781400883929.

[118] E. M. Stein, Oscillatory integrals related to Radon-like transforms, in *Proceedings of the Conference in Honor of Jean-Pierre Kahane* (Orsay, 1993), no. Special Issue, 1995, pp. 535–551. MR 1364908. Zbl 0971.42009.

[119] E. M. Stein and S. Wainger, The estimation of an integral arising in multiplier transformations, *Studia Math.* **35** (1970), 101–104. MR 0265995. Zbl 0202.12401. https://doi.org/10.4064/sm-35-1-101-104.

[120] E. M. Stein and S. Wainger, Problems in harmonic analysis related to curvature, *Bull. Amer. Math. Soc.* **84** no. 6 (1978), 1239–1295. MR 0508453. Zbl 0393.42010. https://doi.org/10.1090/S0002-9904-1978-14554-6.

[121] E. M. Stein and S. Wainger, Oscillatory integrals related to Carleson's theorem, *Math. Res. Lett.* **8** no. 5-6 (2001), 789–800. MR 1879821. Zbl 0998.42007. https://doi.org/10.4310/MRL.2001.v8.n6.a9.

[122] R. S. Strichartz, Singular integrals supported on submanifolds, *Studia Math.* **74** no. 2 (1982), 137–151. MR 0679830. Zbl 0501.43007. https://doi.org/10.4064/sm-74-2-137-151.

[123] E. Szemerédi, On sets of integers containing no four elements in arithmetic progression, *Acta Math. Acad. Sci. Hungar.* **20** (1969), 89–104. MR 0245555. Zbl 0175.04301. https://doi.org/10.1007/BF01894569.

[124] E. SZEMERÉDI, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245. MR 0369312. Zbl 0303.10056. https://doi.org/10.4064/aa-27-1-199-245.

[125] T. TAO, Norm convergence of multiple ergodic averages for commuting transformations, *Ergodic Theory Dynam. Systems* **28** no. 2 (2008), 657–688. MR 2408398. Zbl 1181.37004. https://doi.org/10.1017/S0143385708000011.

[126] M. N. WALSH, Norm convergence of nilpotent ergodic averages, *Ann. of Math.* (2) **175** no. 3 (2012), 1667–1688. MR 2912715. Zbl 1248.37008. https://doi.org/10.4007/annals.2012.175.3.15.

[127] T. ZIEGLER, Universal characteristic factors and Furstenberg averages, *J. Amer. Math. Soc.* **20** no. 1 (2007), 53–97. MR 2257397. Zbl 1198.37014. https://doi.org/10.1090/S0894-0347-06-00532-7.

[128] P. ZORIN-KRANICH, Maximal polynomial modulations of singular integrals, 2017. arXiv 1711.03524.

[129] A. ZYGMUND, *Trigonometric Series. Vol. I, II*, Third ed., *Cambridge Math. Library*, Cambridge Univ. Press, Cambridge, 2002, with a foreword by Robert A. Fefferman. MR 1963498. Zbl 1084.42003.

(Received: August 1, 2017)
(Revised: May 14, 2020)

INSTITUTE OF MATHEMATICS OF THE ROMANIAN ACADEMY, BUCHAREST, ROMANIA
Current address: PURDUE UNIVERSITY, PURDUE, IN
E-mail: vlie@purdue.edu