TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 368, Number 6, June 2016, Pages 3931–3972
http://dx.doi.org/10.1090/tran/6548
Article electronically published on November 20, 2015

ENDPOINT BOUNDS FOR THE BILINEAR HILBERT TRANSFORM

FRANCESCO DI PLINIO AND CHRISTOPH THIELE

ABSTRACT. We study the behavior of the bilinear Hilbert transform (BHT) at the boundary of the known boundedness region $\mathcal{H}$. A sample of our results is the estimate

$$|\langle \mathrm{BHT}(f_1, f_2), f_3\rangle| \le C|F_1|^{\frac34}|F_2|^{\frac34}|F_3|^{-\frac12} \log\log\left(\mathrm{e}^{\mathrm{e}} + \frac{|F_3|}{\min\{|F_1|,|F_2|\}}\right),$$

valid for all tuples of sets $F_j \subset \mathbb{R}$ of finite measure and functions f_j such that $|f_j| \le \mathbf{1}_{F_j}$, $j = 1, 2, 3$, with the additional restriction that f_3 be supported on a major subset F_3' of F_3 that depends on $\{F_j : j = 1, 2, 3\}$. The use of subindicator functions in this fashion is standard in the given context. The double logarithmic term improves over the single logarithmic term obtained by D. Bilyk and L. Grafakos. Whether the double logarithmic term can be removed entirely, as is the case for the quartile operator, remains open.

We employ our endpoint results to describe the blow-up rate of weak-type and strong-type estimates for BHT as the tuple $\vec\alpha$ approaches the boundary of $\mathcal{H}$. We also discuss bounds on Lorentz-Orlicz spaces near $L^{\frac23}$, improving on results of M. Carro et al. The main technical novelty in our article is an enhanced version of the multi-frequency Calderón-Zygmund decomposition.

1. Introduction and main results

Recall that the classical Hilbert transform is bounded in L^p for $1 < p < \infty$. At the endpoint $p = 1$ one has several types of estimates such as Hardy space estimates or Lorentz-Orlicz space estimates. Most relevant for our discussion is the classical weak-type bound in L^1. The language of generalized restricted type estimates allows us to formulate a corresponding dual estimate at L^∞, and the two endpoint estimates suffice to recover L^p bounds by interpolation.

Somewhat analogously, it was shown in [16] that the Coifman-Meyer bilinear singular integrals

$$T(f_1, f_2)(x) = \mathrm{p.v.}\int_{\mathbb{R}^2} f_1(x-t_1)f_2(x-t_2)K(t_1,t_2)\,\mathrm{d}t_1\mathrm{d}t_2, \qquad x \in \mathbb{R}\ ,$$

where K is a homogeneous Calderón-Zygmund kernel in $\mathbb{R}^2$, obey the weak endpoint bound $T : L^1(\mathbb{R}) \times L^1(\mathbb{R}) \to L^{\frac12,\infty}(\mathbb{R})$. In the language of generalized restricted type, this is one of a triple of symmetric endpoint estimates which one may interpolate to obtain bounds in the entire allowed region for L^p estimates established by

Received by the editors March 24, 2014.
2010 *Mathematics Subject Classification.* Primary 42B20.
Key words and phrases. Bilinear Hilbert transform, multi-frequency Calderón-Zygmund decomposition, endpoint bounds.
The first author was partially supported by the National Science Foundation under the grant NSF-DMS-1206438.
The second author was partially supported by the grant NSF-DMS-1001535.

Coifman and Meyer in [6]; see also [13] for an extension to the non-homogeneous case.

This article is concerned with endpoint bounds for the more singular family of bilinear operators known as bilinear Hilbert transforms. Such endpoint bounds have been previously investigated in [1, 2, 5, 7]. The region of known L^p estimates for a bilinear Hilbert transform constitutes an open hexagon depicted in Figure 1. The six extremal points of the hexagon are all symmetric in the language of generalized restricted type estimates, and each of them corresponds to a hypothetic $L^1 \times L^2 \to L^{2/3}$ estimate for some dual of the bilinear Hilbert transform. The shape of this region already suggests that the bilinear Hilbert transform has a more colorful endpoint theory than the bilinear Coifman-Meyer operators discussed above, whose open region of boundedness is the entire open triangle depicted in the figure.

The additional thresholds provided by the short sides of the hexagon, which correspond to hypothetic estimates of the bilinear Hilbert transform mapping into L^p with $p = \frac{2}{3}$, are an important structural feature of modulation-invariant singular integrals. While it is not known whether or not bounds for the bilinear Hilbert transform can be extended past this threshold, such an extension does not hold for bilinear operators very closely related to the bilinear Hilbert transform, such as those obtained from allowing bounded coefficients in model sums representing the bilinear Hilbert transform, as explained in [19]. Using the same effect, it is possible to construct a trilinear modulation-invariant multiplier form which satisfies no bounds beyond these thresholds (C. Muscalu, personal communication). This motivates the study of endpoint estimates for the bilinear Hilbert transform at the boundary of the hexagon. Similar endpoint questions for Carleson's operator, the other stalwart of time-frequency analysis, have enjoyed some popularity as well in recent years [8–10, 22, 23].

Consider the family of trilinear forms with parameter $\vec{\beta} \in \mathbb{R}^3$ defined, for Schwartz functions $f_1, f_2, f_3 : \mathbb{R} \to \mathbb{C}$, by the principal value integral

$$\Lambda_{\vec{\beta}}(f_1, f_2, f_3) = \int_{\mathbb{R}} \text{p.v.} \int_{\mathbb{R}} f_1(x - \beta_1 t) f_2(x - \beta_2 t) f_3(x - \beta_3 t) \, \frac{\mathrm{d}t}{t} \, \mathrm{d}x. \tag{1.1}$$

By scaling and translation invariance we can restrict to vectors $\vec{\beta}$ of unit length and perpendicular to $(1, 1, 1)$. In effect this reduces $\Lambda_{\vec{\beta}}$ to a one-parameter family. The trilinear forms $\Lambda_{\vec{\beta}}$ arise as duals to the family of bilinear operators known as *bilinear Hilbert transforms*, written in singular integral form as

$$\text{BHT}_{\vec{b}}(f_1, f_2)(x) = \text{p.v.} \int_{\mathbb{R}} f_1(x - b_1 t) f_2(x - b_2 t) \, \frac{\mathrm{d}t}{t}, \qquad x \in \mathbb{R}.$$

Indeed, $\Lambda_{\vec{\beta}}(f_1, f_2, f_3) = \langle \text{BHT}_{\vec{b}}(f_1, f_2), \overline{f_3} \rangle$, with $\vec{\beta}$ and $\vec{b}$ related by $\beta_1 - \beta_3 = b_1, \beta_2 - \beta_3 = b_2$.

In a pair of articles by Lacey and the second author [17, 18], it is proved that in the non-degenerate case, meaning no two components of $\vec{\beta}$ are equal,

$$\|\text{BHT}_{\vec{b}}(f_1, f_2)\|_{\frac{p_1 p_2}{p_1 + p_2}} \leq C_{\vec{\beta}} C_{p_1, p_2} \|f_1\|_{p_1} \|f_2\|_{p_2} \tag{1.2}$$

for all $1 < p_1, p_2 \leq \infty$ with $2/3 < \frac{p_1 p_2}{p_1 + p_2} < \infty$. These bounds can be obtained via the interpolation procedure described e.g. in [25, 29] as a consequence of the family

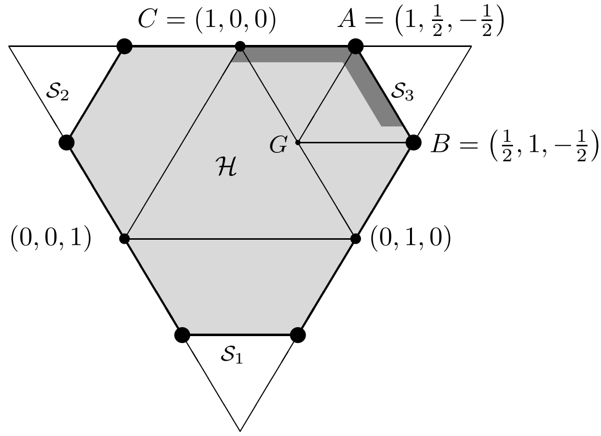

FIGURE 1. The hexagon $\mathcal{H}$ of (1.4). The darker shade indicates the approach region of (8.1).

of *generalized restricted weak-type* (GRWT) estimates

$$\big|\Lambda_{\vec\beta}(f_1, f_2, f_3)\big| \le C_{\vec\beta} C_{\vec\alpha} \prod_{j=1}^{3} |F_j|^{\alpha_j}, \qquad \forall \vec\alpha \in \operatorname{int}\mathcal{H}, \tag{1.3}$$

where

$$\mathcal{H} = \Big\{\vec\alpha = (\alpha_1, \alpha_2, \alpha_3) : \alpha_1 + \alpha_2 + \alpha_3 = 1,\ \max_j \alpha_j \le 1,\ \min_j \alpha_j \ge -\tfrac12\Big\} \tag{1.4}$$

is the shaded hexagon in Figure 1. The diction GRWT stands for (1.3) holding for all tuples of sets $F_j \subset \mathbb{R}$, $j = 1, 2, 3$, of finite measure and for all functions $|f_j| \le \mathbf{1}_{F_j}$, with the additional restriction that if j_* is the maximal index j such that $\alpha_j = \min_k\{\alpha_k\}$, then $|f_{j_*}| \le \mathbf{1}_{F'_{j_*}}$ for some subset $F'_{j_*} \subset F_{j_*}$, which is major in the sense $|F_{j_*}| \le 2|F'_{j_*}|$ and which may depend on F_1, F_2, F_3. Allowing the passage to a major subset is crucial if one of the parameters α_j is less than or equal to 0. Note that at every point of the open region discussed, at most one parameter is less than or equal to 0 and this is the one with index j_*. Assuming for example that $j_* = 3$, estimate (1.3) is equivalent to the weak-type bound

$$\|\mathrm{BHT}_{\vec b}(f_1, f_2)\|_{\frac{1}{1-\alpha_3},\infty} \le C_{\vec\beta} C_{\vec\alpha} \prod_{j=1,2} |F_j|^{\alpha_j}, \qquad \forall\, |f_j| \le \mathbf{1}_{F_j},\ j = 1, 2.$$

The symmetric nature of the form $\Lambda_{\vec\beta}$ and the notion of GRWT show that specializing $j_* = 3$ is no loss of generality.

In the degenerate case, that is, if two components of $\vec\beta$ are equal, questions on bounds for the bilinear Hilbert transform trivialize. The bilinear Hilbert transform then degenerates to a combination of the classical Hilbert transform and a pointwise product. There are three different degenerate cases, depending on which two components of $\vec\beta$ are equal. In each case, the region of GRWT estimates is no longer symmetric under permutation of the three indices, and neither contains the above hexagon $\mathcal{H}$ nor it is contained in the hexagon. This leads to an interesting array of questions concerning uniformity of bounds for the non-degenerate case in the vicinity of the degenerate case. Such questions have been addressed for example in [12, 20, 27, 28]. The present paper focuses only on the non-degenerate case and ignores the above uniformity questions by allowing constants $C_{\vec{\beta}} > 0$ that depend on $\vec{\beta}$ in an unspecified manner. In general the constants will blow up as $\Delta_{\vec{\beta}}$, the distance from $\vec{\beta}$ to the union of the three hyperplanes $\beta_j = \beta_k$, $k \neq j$, tends to 0.

A folklore conjecture is that the generalized restricted weak-type estimate (1.3) for the non-degenerate bilinear Hilbert transform extends to the region $\mathcal{S}$ defined by

$$\mathcal{S} = \bigcup_{j=1}^{3} \mathcal{S}_j, \qquad \mathcal{S}_j = \left\{\vec{\alpha} : \alpha_1 + \alpha_2 + \alpha_3 = 1,\ \alpha_j = -\tfrac{1}{2},\ \max\{\alpha_k\} < 1\right\},$$

that is, the union of the short open segments of the boundary of the hexagon.

Conjecture 1. *The GRWT estimate* (1.3) *holds for all tuples* $\vec{\alpha} \in \mathcal{S}$.

A main theme of the present paper is that additional insight can be obtained by lifting the subindicator condition on a careful choice of the functions f_j, that is, to allow for general L^p functions f_j rather than functions dominated by an indicator function. This is analogous to the classical case of the linear Hilbert transform discussed above, where the crucial boundary estimate is a weak-type estimate which allows the input function to be a general L^1 function, while the test function that one pairs with to obtain a bilinear form has to be a subindicator function, supported on a major subset as elaborated in the GRWT definition above. In the case of the trilinear form $\Lambda_{\vec{\beta}}$ one has a choice of three functions on which to lift the subindicator condition, yielding a relatively more diverse set of possible estimates.

At the typical corner A of the hexagon, the second coordinate $\alpha_2 = 1/2$ stands for the Hilbert space $L^2(\mathbb{R})$. In the vicinity of that corner it is therefore particularly efficient to lift the subindicator condition on the function f_2, since one has the full Hilbert space technique at hand. This was already observed in [7] for the quartile operator.

A side product of our investigations is a fairly straightforward adaption of the strategy of [7] to the present case of the bilinear Hilbert transform to obtain the following endpoint estimate at the corner A with a logarithmic correction term.

Theorem 1. *Let $f_2 \in L^2(\mathbb{R})$ and sets $F_1, F_3 \subset \mathbb{R}$ of finite measure be given. Then, there exists a major subset F_3' of F_3, depending on f_2, F_1, F_3, such that*

$$\left|\Lambda_{\vec{\beta}}(f_1, f_2, f_3)\right| \leq C_{\vec{\beta}} |F_1| \|f_2\|_2 |F_3|^{-\frac{1}{2}} \log\left(\mathrm{e} + \tfrac{|F_3|}{|F_1|}\right) \qquad \forall |f_1| \leq \mathbf{1}_{F_2},\ |f_3| \leq \mathbf{1}_{F_3'}.$$

This is our only estimate directly at the corner A of the hexagon. It is a strengthening of a result of [1, 2], where the same estimate is shown to hold under the further assumption that f_2 is a subindicator function as well. One may view the logarithmic correction term as a fallout of being on the edge $\overline{AC}$, which corresponds to the space L^1 for the first function.

Of course a symmetric estimate holds at the other six corners of the hexagon, and specializing again to three subindicator functions as in [1, 2] one obtains by interpolation GRWT estimates everywhere in the open hexagon. However, the interpolated estimates one obtains in this way are not as efficient in the vicinity of the boundary of the hexagon as what one obtains using the next two theorems.

To motivate the next theorem, consider the symmetric estimate to Theorem 1 at corner B, which puts the function f_1 in L^2. We would like to prove sharp estimates on the edge $\overline{AB}$. There the function f_1 is in L^p with p between 1 and 2. It is therefore natural to seek a Calderón-Zygmund decomposition of the function f_1, using Hilbert space technique on the good portion and some additional localization information on the bad portion. The Calderón-Zygmund decomposition has to respect a number of frequencies as does the multi-frequency Calderón-Zygmund decomposition (MFCZ) developed in [26]. A main point of the present paper is that in order to be successful on the edge of the hexagon we need a very sharp form of this MFCZ. Developing this MFCZ and applying it are the main technical advances of the present paper.

Theorem 2. *We write* $(t)_* = (1+t)(\log(\mathrm{e}+t))^3$. *Let* $\vec{\alpha} \in \mathcal{S}_3$, $f_1 \in L^{\frac{1}{\alpha_1}}(\mathbb{R})$, *and sets* $F_2, F_3 \subset \mathbb{R}$ *of finite measure be given. Then, there exists a major subset* F_3' *of* F_3, *depending on* f_1, F_2, F_3, *such that the estimate*

$$\begin{aligned} \left|\Lambda_{\vec{\beta}}(f_1,f_2,f_3)\right| \leq \frac{C_{\vec{\beta}}}{(1-\alpha_1)(1-\alpha_2)} \|f_1\|_{\frac{1}{\alpha_1}} |F_2|^{\alpha_2} |F_3|^{-\frac{1}{2}} \\ \times \Big(\max\big\{ \log\big(\tfrac{|F_3|}{|F_2|}\big), \tfrac{1}{1-\alpha_1} \big\} \Big)_*^{2\alpha_1 - 1} \end{aligned} \tag{1.5}$$

holds for all $|f_2| \leq \mathbf{1}_{F_2}$, $|f_3| \leq \mathbf{1}_{F_3'}$.

This theorem is analogous to [7, Proposition 2.1] for the quartile operator. Thanks to perfect localization of Walsh wave packets, an even sharper but trivial form of MFCZ is true in the discrete setting, and hence [7] obtains an estimate without the starred correction term, which is in fact a stronger form of Conjecture 1 for the quartile operator. The exponent of the starred term tends to 0 at the corner B and to 1 at the corner A, showing that the correction term caused by MFCZ becomes worse as one moves away from the Hilbert space. Theorem 2 is a phenomenon on the open edge $\overline{AB}$; we do not see how to obtain the theorem by interpolation from any estimates at the corners A and B, in particular not by interpolation with Theorem 1. The constant in Theorem 2 blows up as we approach either corner.

We return to Theorem 1 at the corner A as motivation for the following theorem. We fix the exponent $\alpha_2 = 1/2$ and the general function $f_2 \in L^2$, which puts us on the bisecting line $\overline{AG}$. This time we lift a second subindicator condition, namely on the function f_1, to obtain two unconstrained functions. On the edge $\overline{AG}$ the function f_1 is in L^p with $1 < p < 2$, and one can apply again the MFCZ to this function.

Theorem 3. *For all* $0 \leq \alpha_1 < 1$, $f_1 \in L^{\frac{1}{\alpha_1}}(\mathbb{R}), f_2 \in L^2(\mathbb{R}), F_3 \subset \mathbb{R}$, *there exists a major subset* F_3' *of* F_3, *depending on* f_1, f_2, F_3, *such that*

$$\left|\Lambda_{\vec{\beta}}(f_1,f_2,f_3)\right| \leq C_{\vec{\beta}} \frac{1}{1-\alpha_1} \Big(\tfrac{1}{1-\alpha_1}\Big)_*^{2\alpha_1-1} \|f_1\|_{\frac{1}{\alpha_1}} \|f_2\|_2 |F_3|^{\frac{1}{2}-\alpha_1} \qquad \forall |f_3| \leq \mathbf{1}_{F_3'}.$$

Similar estimates as in Theorem 3 with worse growth of the constant as one approaches the corner A can be obtained by standard interpolation methods from Theorem 1 and its symmetric counterparts. Namely, observe that the estimate of Theorem 1 is equivalent to the bound

$$\left|\Lambda_{\vec{\beta}}(f_1,f_2,f_3)\right| \leq \frac{C_{\vec{\beta}}}{1-\alpha_1} |F_1|^{\alpha_1} \|f_2\|_2 |F_3|^{\frac{1}{2}-\alpha_1} \tag{1.6}$$

for functions f_1, f_3 restricted as in the statement of the theorem. Marcinkiewicz-type interpolation as in Lemma 9.1 deduces the same type of estimate as in Theorem 3 from (1.6), albeit with a blow-up rate of $(1-\alpha_1)^{-5/2}$ as one approaches the corner A. On the other hand, one notes that (1.6) is stronger than what is obtained by specializing the estimate of Theorem 3 to subindicator functions f_1. Therefore, neither Theorem 1 nor Theorem 3 implies the other in full strength by the obvious deduction methods. Again, a sharper analogue of Theorem 3 for the quartile operator has been proved in [7, Proposition 2.3], lacking the starred correction term thanks to the perfect discrete MFCZ.

Restricting Theorem 2 to subindicator functions f_1 and interpolating with the symmetric version under interchanging the corners A and B yields the following punchline result.

Corollary 4. *For all tuples $\vec{\alpha} \in \mathcal{S}_j$, $j = 1, 2, 3$, we have the GRWT estimate*

$$\begin{aligned} \left|\Lambda_{\vec{\beta}}(f_1, f_2, f_3)\right| \le C_{\vec{\beta}} \Big(\prod_{k=1}^{3} |F_k|^{\alpha_k} \Big) \\ \times \max\left\{ \frac{1}{\min_{k\neq j}\{1-\alpha_k\}}, \log\log\Big(\mathrm{e}^{\mathrm{e}} + \frac{|F_j|}{\min_{k\neq j}|F_k|}\Big) \right\}. \end{aligned} \tag{1.7}$$

The special case of this result at the midpoint of $\overline{AB}$ has been highlighted in the abstract of this paper. This theorem is a weaker form of Conjecture 1 by the double logarithmic correction term. This estimate cannot be obtained by interpolation of Theorem 1 and its symmetric counterparts, which only yields the single logarithmic estimate that was observed in [1,2]. This highlights again that Theorem 2 encodes additional information relative to Theorem 1. Clarifying whether the double logarithmic term can be removed in the corollary is one of the more intriguing open questions on endpoint bounds for the bilinear Hilbert transform. Obviously we do not see how to do this with the present technology.

We conclude this discussion with a few remarks on our strengthening of the multi-frequency Calderón-Zygmund decomposition, Proposition 3.2. The bad portion of the MFCZ is the sum of functions b_I localized to intervals I and having mean zero with respect to a number N of bad frequencies relevant on the interval I. The main issue lies with estimating the interaction of this bad function b_I with wave packets which are frequency localized in a compact interval near a bad frequency, and which are spatially localized away from but not too far away from I. To make this interaction sufficiently small for our needs we work on the one hand with wave packets which have better than mere Schwartz function decay. We use an optimal, almost exponential, decay following a construction by Ingham. To utilize this decay we have to prepare the bad function b_I of the MFCZ to have mean zero not only against the dominant bad frequency, but also against approximately $\log(N)$ many equidistant frequencies in the vicinity of the dominating bad frequency. The price of all this is the appearance of the extra terms $(\cdot)_*$ occurring in Theorems 2 and 3. This is in contrast to the discrete analogues of [7], where of course one has wave packets which are compactly supported both in frequency and in space, and the interaction terms in question are simply zero. We stress that the use of almost exponential-type wave packets has no precedent in the context of time-frequency analysis. It is unnecessary for deeply interior estimates in the open hexagon, but it appears to be relevant for the sharp estimates at and near the boundary of the hexagon that we investigate.

It is our opinion that Proposition 3.2, or variants thereof, could be employed as well in the translation to the continuous case of the arguments of [8–10] on Carleson-type operators and of [27] on uniform estimates for the family of Walsh models of the bilinear Hilbert transforms.

Outline of the article. Sections 2 and 3 are concerned with the multi-frequency Calderón-Zygmund decomposition in general. Section 2 contains technical preliminaries on functions with compact frequency support and almost exponential decay rate. Our sharp version of the multi-frequency Calderón-Zygmund decomposition is introduced in Section 3 and its properties are discussed, most notably in Proposition 3.2.

We then turn to the bilinear Hilbert transform. In Sections 4 and 5 we rephrase the construction of the model sums for $\Lambda_{\vec{\beta}}$ and some classical results of time-frequency analysis. In Section 6 we apply Proposition 3.2 to obtain an estimate for the model sums of $\Lambda_{\vec{\beta}}$ restricted to a single forest with appropriate L^1 and L^∞ bounds on the counting function. The main steps of the proof of Theorems 1, 2, and 3, as well as the proof of Corollary 4, are given in Section 7.

Finally, in Section 8, we present several corollaries of our main results, elaborating on alternative formulations of the behavior of the bilinear Hilbert transform near the boundary. The first group of corollaries is concerned with the blow-up rates of the eight possible types of estimates for $\mathrm{BHT}_{\vec{b}}$, corresponding to different choices of sets of unrestricted functions, as the exponents approach the boundary of the hexagon $\mathcal{H}$ in Figure 1. These estimates are summarized in Table 2. The second group of corollaries, in the spirit of the article [5], is devoted to the boundedness properties of $\mathrm{BHT}_{\vec{b}}$ on Lorentz-Orlicz spaces near Hölder tuples $\vec{\alpha}$ on the open segment $\overline{AB}$ and at the corner A. These corollaries are proved in Section 9.

Notational remarks. The vector $\vec{\beta}$ is always non-degenerate, and all explicit and implicit constants in this paper may depend on $\Delta_{\vec{\beta}}$, the distance from $\vec{\beta}$ to the degenerate case as discussed above. Let $I \subset \mathbb{R}$ be an interval; $c(I)$ will denote the midpoint of I, and, for $C > 0$, by CI we refer to the interval with center $c(I)$ and length $C|I|$. We also write $x + I$ for the interval $\{x + y : y \in I\}$. We set

$$\|f\|_{L^p(I)} := \left(\int_I |f(x)|^p \tfrac{\mathrm{d}x}{|I|} \right)^{\frac{1}{p}}, \ 1 \le p < \infty, \qquad \|f\|_{L^\infty(I)} := \operatorname*{ess\,sup}_{x \in I} |f(x)|.$$

For $1 \le p < \infty$, the p-th Hardy-Littlewood maximal function is defined as

$$\mathsf{M}_p f(x) = \sup_{I \ni x} \|f\|_{L^p(I)}.$$

With $\mathcal{D}$, we indicate a generic grid on $\mathbb{R}$, that is, a collection of intervals such that $I \cap I' \in \{I, I', \emptyset\}$ for each $I, I' \in \mathcal{D}$. We write $\mathcal{D}_0$ for the standard dyadic grid on $\mathbb{R}$, while the notation $\mathcal{D}(I)$ refers to the standard dyadic grid on an interval $I \subset \mathbb{R}$. Finally, the constants $C > 0$, as well as the constants implied by the almost inequality sign $\lesssim$, may vary at each occurrence without explicit mention and are meant to be absolute, once $\vec{\beta}$ has been fixed, unless otherwise specified.

2. Rapidly decaying functions with compact frequency support

Throughout the article, u will be a positive, increasing and convex function on $[0,\infty)$ satisfying the normalized Osgood condition

$$\int_0^\infty \frac{1}{u(t)}\,\mathrm{d}t = 1 \tag{2.1}$$

and such that

$$B_u(\tau) := \sup_{t\geq 0}\Big((1+|u(t)|)\mathrm{e}^{-\tau t}\Big) < \infty \tag{2.2}$$

for all $\tau > 0$. Condition (2.2) holds, for instance, when $u(t) \leq C(1+t)^C$ for some $C > 1$. We will use the (evenly extended) inverse function of u,

$$U : \mathbb{R} \to [0,\infty), \qquad U(x) = \begin{cases} u^{-1}(|x|), & |x| \geq u(0), \\ 0, & |x| < u(0), \end{cases}$$

which is increasing on $[0,\infty)$ and satisfies

$$U(x) \leq |x| \qquad \forall\, x \in \mathbb{R}, \tag{2.3}$$

$$U(\vartheta x) \geq \vartheta U(x) - u(0) \qquad \forall\, x \in \mathbb{R}, \vartheta \in [0,1]. \tag{2.4}$$

The first estimate above is a consequence of (2.1), while (2.4) follows from (2.3) and concavity of U on $\{x \geq u(0)\}$.

Significant examples of functions u as such are given by the family

$$u_\lambda(t) = \frac{1}{\lambda}(t+\mathrm{e})\big(\log(t+\mathrm{e})\big)^{1+\lambda}, \qquad \lambda > 0. \tag{2.5}$$

The upcoming Lemma 2.1 is a reformulation of a result of Ingham [15]. In words, given any u satisfying the above assumptions, one obtains a smooth function υ with compact frequency support and with exponential decay rate given by a constant times U.

Lemma 2.1. *Let u be as above. Then, there exists a smooth non-negative function $\upsilon : \mathbb{R} \to [0,+\infty)$ with the properties that*

$$\mathbf{1}_{[-\frac{1}{6},\frac{1}{6}]}(\xi) \leq \widehat{\upsilon}(\xi) \leq \mathbf{1}_{[-\frac{1}{2},\frac{1}{2}]}(\xi), \qquad \forall \xi \in \mathbb{R}, \tag{2.6}$$

$$\big|\upsilon^{(j)}(x)\big| \leq C_{D,u}\mathrm{e}^{-\frac{1}{100}U(x)} \qquad \forall x \in \mathbb{R},\ j \in \{0,\ldots,D\}. \tag{2.7}$$

The positive constants $C_{D,u}$ in (2.7) depend on D and on $B_u(2D)$ from (2.2).

Proof. Below, the constant $C_D > 0$ depends only on D and may vary from line to line. Consider the sequence of functions $v_k : \mathbb{R} \to [0,\infty)$, defined by the recurrence

$$v_1 := u(1)\mathbf{1}_{[0,(u(1))^{-1}]}, \qquad v_k := v_{k-1} * \big(u(k)\mathbf{1}_{[0,(u(k)^{-1}]}\big), \quad k > 1.$$

It is easy to see that $\int v_k = 1$, that $\sup v_k \leq \sup v_1 = u(1)$, that $v_k \in \mathcal{C}^{k-2}(\mathbb{R})$, $k \geq 2$, and finally that $\operatorname{supp} v_k \subset \big[0,(u(1))^{-1}+\cdots+(u(k))^{-1}\big] \subset [0,1]$. The ($\mathcal{C}^m$, for each m) uniform limit v_0 of the v_k is therefore a smooth non-negative function with $\int v_0 = 1$, $\operatorname{supp} v_0 \subset [0,1]$. Moreover, v_0 is strictly positive in $(0,1)$ and satisfies the bounds (see [14, Theorem 1.3.5] for details)

$$\sup|v_0^{(k)}| \leq 2^k\big(\textstyle\prod_{j=1}^{k+1}u(j)\big) \leq (2u(k+1))^k, \qquad k = 0,1,\ldots, \tag{2.8}$$

where the last inequality comes from u being increasing. We set

$$v(\xi) := \int_{-1}^{1} v_0\big(3\xi - t + \tfrac{1}{2}\big)\,\mathrm{d}t, \qquad \upsilon(x) := \int_{\mathbb{R}} v(\xi)\mathrm{e}^{2\pi i x\xi}\,\mathrm{d}\xi = \int_{-\frac{1}{2}}^{\frac{1}{2}} v(\xi)\mathrm{e}^{2\pi i x\xi}\,\mathrm{d}\xi.$$

Since $\int v_0 = 1$, (2.6) follows by construction. For all $x \in \mathbb{R}$, $k \geq 0$, $0 \leq j \leq D$, we have that

$$\begin{aligned} |x|^k |\upsilon^{(j)}(x)| &\leq (2\pi)^{j-k} \sup_{|\xi|\leq \frac{1}{2}} \Big| \big(\tfrac{\mathrm{d}}{\mathrm{d}\xi}\big)^k \big(\xi^j v(\xi)\big) \Big| \\ &\leq (2\pi)^{D-k} \sum_{n=0}^{j} \tbinom{k}{n} \tfrac{j!}{(j-n)!} \sup_{|\xi|\leq\frac{1}{2}} \big(|\xi|^{j-n} |v^{(k-n)}(\xi)|\big) \leq C_D k^D \big(6u(k+1)\big)^k. \end{aligned}$$

In the last step above, we employed the crude bound $\binom{k}{n} \leq k^n \leq k^D$, and subsequently (2.8) coupled with the obvious fact that $\sup |v^{(k)}| \leq 2{\cdot}3^k \sup |v_0^{(k)}|$. For each $|x| \geq 6\mathrm{e}u(1)$, let $k(x)$ be the greatest integer $k \geq 0$ such that $6u(k+1)|x|^{-1} \leq \mathrm{e}^{-1}$. Thus $k(x) + 1 \leq U(x/6\mathrm{e}) \leq k(x) + 2$, and the above display for $k = k(x)$ reads

$$\begin{aligned} |\upsilon^{(j)}(x)| &\leq C_D \big(k(x)\big)^D \mathrm{e}^{-k(x)} \leq C_D \big(U(x/6\mathrm{e})\big)^D \mathrm{e}^{-U(x/6\mathrm{e})} \leq C_D \big(1 + \tfrac{|x|}{6\mathrm{e}}\big)^D \mathrm{e}^{-U(x/6\mathrm{e})} \\ &\leq C_D B_u(2D) \mathrm{e}^{-\frac{1}{2}U(x/6\mathrm{e})} \leq C_{D,u} \mathrm{e}^{-\frac{1}{100}U(x)}. \end{aligned}$$

We have relied on (2.3) for the third inequality, on (2.2) to pass to the second line and on (2.4) for the last step. We have thus obtained (2.7) for $|x| \geq 6\mathrm{e}u(1)$, with $C_{D,u} = C_D B_u(2D)$. To argue for $|x| \leq 6\mathrm{e}u(1)$, note that the bound $\sup |\upsilon^{(j)}| \leq C_D u(1)$ can be inferred as a particular case of the above discussion. In the range $|x| \leq 6\mathrm{e}u(1)$, this entails (2.7) with $C_{D,u} := C_D u(1)\mathrm{e}^{100U(6\mathrm{e}u(1))} \leq C_D u(1)\mathrm{e}^{600\mathrm{e}u(1)}$, which depends only on D and $B_u(1)$, and is thus of the required form. This concludes the proof of the lemma. □

Remark 2.2. The existence of an exponentially decaying smooth function with compactly supported Fourier transform is forbidden by the Paley-Wiener theorem. In [15], it is pointed out that if u is such that the integral in (2.1) diverges, there exists no such function decaying like (2.7). For instance, there is no smooth function f with $\widehat{f}$ compactly supported and decaying like $|f(x)| \lesssim \exp\big(-c|x|/\log(\mathrm{e}+|x|)\big)$.

For the remainder of the section, we write $I_0 := [-\frac{1}{2}, \frac{1}{2}]$. In the next two lemmata, we devise a splitting of a smooth function with spatial decay rate aU and frequency supported on I_0 into a part having spatial support contained in the $u(K)$-dilate of I_0 and Fourier transform exponentially small in K away from I_0, plus an exponentially small remainder.

Lemma 2.3. *Let φ be a Schwartz function with* $\operatorname{supp}\widehat{\varphi} \subset I_0$ *and satisfying the bound*

$$\big|\varphi^{(j)}(x)\big| \leq A\mathrm{e}^{-aU(x)} \qquad \forall x \in \mathbb{R},\ j = 0, 1, \tag{2.9}$$

for some constants $A > 0, 0 < a \leq \frac{1}{100}$. For each $K \geq 1$, $N > 0$ there exists a decomposition

$$\varphi := \phi + \mathrm{e}^{-\frac{a}{12}K}\psi \tag{2.10}$$

with the following properties:

$$\operatorname{supp}\phi \subset u(K)I_0, \tag{2.11}$$

$$|\widehat{\phi}(\zeta)| \lesssim \mathrm{e}^{-\frac{a}{12}K}(1+|\zeta|)^{-N} \qquad \forall |\zeta| \geq 2, \tag{2.12}$$

$$\left|\phi^{(j)}(x)\right| \lesssim (1+|x|)^{-N} \qquad \forall x \in \mathbb{R},\, j \in \{0,1\}, \tag{2.13}$$

$$\left|\psi^{(j)}(x)\right| \lesssim (1+|x|)^{-N} \qquad \forall x \in \mathbb{R},\, j \in \{0,1\}. \tag{2.14}$$

The implicit constants in (2.12)-(2.14) *depend only on* A, a, N *and* u.

Proof. Let v be a smooth function such that

$$\mathbf{1}_{\frac{u(K)}{3}I_0} \leq v \leq \mathbf{1}_{u(K)I_0}, \qquad |\widehat{v}(\zeta)| \lesssim u(K)\mathrm{e}^{-\frac{1}{100}U(u(K)\zeta)} \quad \forall\, \zeta \in \mathbb{R}; \tag{2.15}$$

such a function exists by Lemma 2.1. We realize the decomposition (2.10) by setting

$$\phi = \varphi v, \qquad \psi(x) := \mathrm{e}^{\frac{a}{12}K}\varphi(\mathbf{1}_{\mathbb{R}} - v).$$

Then (2.11) holds by construction. Furthermore, we obtain (2.14) from the bound

$$\left|\big((\mathbf{1}_{\mathbb{R}} - v)\varphi\big)^{(j)}(x)\right| \lesssim \mathrm{e}^{-\frac{a}{2}U(u(K)/6)}\mathrm{e}^{-\frac{a}{2}U(x)} \lesssim \mathrm{e}^{-\frac{a}{12}K}(1+|x|)^{-N}, \tag{2.16}$$

for $x \in \mathbb{R}$ and $j = 0, 1$, which follows by restricting to $|x| \geq u(K)/6$ via support considerations, then relying on (2.7), and finally using (2.4) and (2.2). Then, (2.13) is derived by comparison with (2.14). We are left with proving (2.12), that is, estimating $\widehat{\varphi v}(\zeta)$ for $|\zeta| \geq 2$. To do so, we use $|\widehat{\varphi}| \lesssim \mathbf{1}_{I_0}$ and later (2.15), so that

$$\left|(\widehat{\varphi} * \widehat{v})(\zeta)\right| \lesssim \int_{I_0} |v(\zeta - \eta)|\,\mathrm{d}\eta \lesssim \sup_{\eta \in I_0} |v(\zeta-\eta)| \lesssim u(K)\mathrm{e}^{-\frac{1}{100}U(u(K)(|\zeta|-1))}.$$

By repeatedly making use of (2.2), it is easy to see that, when $|\zeta| \geq 2$, the last right hand side is bounded by $\exp(-aK/12)(1+|\zeta|)^{-N}$ times a constant depending on u and N only. This concludes the proof of the lemma. □

The next decomposition, which is similar to the one of Lemma 2.3, but preserves mean zero with respect to a fixed frequency outside I_0, was partly inspired by [24, Lemma 3.1].

Lemma 2.4. *Let* φ, K, N, *be as in Lemma* 2.3, $R > 1$, $\xi_0 \in RI_0 \backslash I_0$. *There exists a decomposition*

$$\varphi := \phi + \mathrm{e}^{-\frac{a}{12}K}\psi \tag{2.17}$$

depending on ξ_0, *such that* (2.11)-(2.13) *hold for* ϕ, (2.14) *holds for* ψ, *and in addition*

$$\int_{\mathbb{R}} \phi(x)\,\mathrm{e}^{-2\pi i \xi_0 x}\,\mathrm{d}x = \int_{\mathbb{R}} \psi(x)\,\mathrm{e}^{-2\pi i \xi_0 x}\,\mathrm{d}x = 0. \tag{2.18}$$

The implicit constants in (2.12)-(2.14) *depend only on* A, a, R, N *and* u.

Proof. Set $w(\cdot) := \varphi(\cdot)\mathrm{e}^{-2\pi i \xi_0 \cdot}$. Then $0 \notin \operatorname{supp}\widehat{w} = I_0 - \xi_0$, that is, $\int w = 0$. Let v be as in (2.19). In view of the support condition on $\mathbf{1}_{\mathbb{R}} - v$ and later relying on (2.9), we preliminarily observe that

$$\begin{aligned} \Big|\int_{\mathbb{R}} wv\,\mathrm{d}x\Big| &= \Big|\int_{\mathbb{R}} w(\mathbf{1}_{\mathbb{R}} - v)\,\mathrm{d}x\Big| \lesssim \int_{|x| \geq \frac{u(K)}{6}} |w(x)|\,\mathrm{d}x \\ &\leq A\int_{|x| \geq \frac{u(K)}{6}} \mathrm{e}^{-aU(x)}\mathrm{d}x \lesssim \mathrm{e}^{-\frac{a}{2}U(u(K)/6)}\int_{\mathbb{R}} \mathrm{e}^{-\frac{a}{2}U(x)}\mathrm{d}x \lesssim \mathrm{e}^{-\frac{a}{12}K}. \end{aligned} \tag{2.19}$$

For the next to last inequality above, we used that U is increasing. Then, in the last step, we employed (2.4) for the first factor and (2.2) to estimate the integral. Since $w(\cdot) = \varphi(\cdot)\mathrm{e}^{-2\pi i \xi_0 \cdot}$, (2.17) is fulfilled if we set

$$\phi(x) := \left(w(x)v(x) - \frac{\int wv}{\int v} v(x) \right) \mathrm{e}^{2\pi i \xi_0 x},$$
$$\psi(x) := \mathrm{e}^{\frac{a}{12}K} \left(\frac{\int wv}{\int v} v(x) + (\mathbf{1}_{\mathbb{R}} - v)(x)w(x) \right) \mathrm{e}^{2\pi i \xi_0 x}.$$

With these definitions, the mean zero condition (2.18) for ϕ holds by construction. Then, (2.18) for ψ follows by difference, again in view of $\int w = 0$. By construction as well, $\operatorname{supp} \phi \subset u(K)I_0$, and we have earned (2.11).

Next, we prove (2.13) and (2.14). Recalling that the implicit constants are allowed to depend on R and that $|\xi_0| \le R$, we can ignore the modulation factor of ψ, and (2.14) is a consequence of the bounds

$$\left| \frac{\int_{\mathbb{R}} (wv)}{\int_{\mathbb{R}} v} v^{(j)}(x) \right| \lesssim \frac{\mathrm{e}^{-\frac{a}{12}K}}{u(K)} (1 + u(K))^N (1 + |x|)^{-N} \lesssim \mathrm{e}^{-\frac{a}{12}K} (1 + |x|)^{-N},$$
$$\left| \big((\mathbf{1}_{\mathbb{R}} - v)w \big)^{(j)}(x) \right| \lesssim \mathrm{e}^{-\frac{a}{2}U(u(K)/6)} \mathrm{e}^{-\frac{a}{2}U(x)} \lesssim \mathrm{e}^{-\frac{a}{12}K} (1 + |x|)^{-N},$$

for $x \in \mathbb{R}$ and $j = 0, 1$. For the first line of the last display, we have used inequality (2.19) and that $v^{(j)}$ is supported on $|x| \le u(K)/2$, and subsequently (2.2). The second line follows via the same argument we used for (2.16).

We now turn to (2.12). The term involving $\widehat{v}$ is easily bounded, taking (2.19), (2.15) and (2.2) into account, by

$$\mathrm{e}^{-\frac{a}{12}K} |\widehat{v}(\zeta - \xi_0)| \lesssim \mathrm{e}^{-\frac{a}{12}K} \mathrm{e}^{-\frac{1}{100}U(u(K)(\zeta - \xi_0))} \lesssim \mathrm{e}^{-\frac{a}{12}K} (1 + u(K)|\zeta - \xi_0|)^{-N}$$
$$\lesssim \mathrm{e}^{-\frac{a}{12}K} (1 + |\zeta|)^{-N}.$$

The above estimate actually holds for all $\zeta \in \mathbb{R}$, and the last almost inequality sign hides the constant $(1 + |\xi_0|)^N \le (1 + R)^N$, which we ignore. Finally, the term $(\widehat{w} * \widehat{v})(\zeta - \xi_0)$ is handled in exactly the same fashion as (2.17). The proof is complete. □

3. A multi-frequency Calderón-Zygmund decomposition

Throughout this section, our definitions depend on a fixed choice of the function u, of its extended inverse U, as in Section 2, and of parameters $R \ge 1, 0 < \varepsilon \le 2^{-8}$. The almost inequality signs appearing in the sequel hide implicit constants which are allowed to possibly depend on u and R only.

3.1. Top data and adapted functions with fast spatial decay.

We call *top datum* a pair (I, ξ), where $I \subset \mathbb{R}$ is a spatial interval and $\xi \in \mathbb{R}$ is a frequency. We say that a smooth function φ is u-adapted,[1] with adaptation rate $a > 0$, to the top datum (I, ξ) if

$$\text{(3.1)} \quad \left| \big(\mathrm{e}^{-2\pi i \xi \cdot} \varphi(\cdot) \big)^{(j)}(x) \right| \lesssim |I|^{-\frac{1}{2} - j} \exp\left(-aU\left(\tfrac{|x - c(I)|}{|I|} \right) \right) \qquad \forall x \in \mathbb{R},\ j \in \{0, 1\}.$$

Note that, by virtue of property (2.2) of u, (3.1) is stronger than the usual notion of adaptation of e.g. (3.9) below. In what follows, the adaptation rate a will be an

[1]Here, and in the remainder of the article, we adopt an L^2 normalization for our adapted functions.

absolute constant, which may be different at each occurrence. When we speak about collections of u-adapted functions, we assume, without explicit mention, uniformity of the adaptation rate a and of the implicit constants in the almost inequality sign of (3.1). Using the property of U (2.4), we see that for any pair of top data $(I,\xi),(J,\zeta)$ such that $|J|\le|I|\le 2|J|$, $|c(I)-c(J)|\le|J|$ and $|\zeta-\xi|\le R|I|^{-1}$, any φ u-adapted to (I,ξ) is also u-adapted to (J,ζ), with adaptation rate $a'\ge a/2$. Hence, there is no loss in generality with assuming that the spatial intervals I of our top data belong to the standard dyadic grid $\mathcal{D}_0$.

Remark 3.1. For each $0<\varepsilon\le 1$, examples of u-adapted functions to (I,ξ), with adaptation rate $a=\varepsilon/100$, are given by

$$\upsilon_{(I,\xi),\varepsilon} := \mathrm{Mod}_\xi \mathrm{Dil}^2_{\varepsilon^{-1}|I|}\mathrm{Tr}_{c(I)}\upsilon,$$

where υ is the output of Lemma 2.1 corresponding to u. In view of (2.6), $\widehat{\upsilon_{(I,\xi),\varepsilon}}$ is supported on the interval of length $\varepsilon|I|^{-1}$ centered at ξ.

3.2. **A multi-frequency Calderón-Zygmund decomposition with respect to top data.** Let $\varepsilon>0$ be a fixed small parameter and (I,ξ) be a top datum. A set of functions $\boldsymbol{\varphi}_{(I,\xi)}=\{\varphi_J : J\in\mathcal{D}(I)\}$, indexed by the dyadic subintervals of I, is a collection of *u-adapted wave packets* if

- each φ_J is smooth and u-adapted to (J,ξ),
- the support of each $\widehat{\varphi_J}$ is contained on an interval of length $\varepsilon|J|^{-1}$ centered at $\xi_J\in\mathbb{R}$, and $|\xi_J-\xi|\le R|J|^{-1}$.

If, furthermore, for each $\varphi_J\in\boldsymbol{\varphi}_{(I,\xi)}$ we also have that $|\xi_J-\xi|\ge|J|^{-1}$, so that consequently $\xi\notin\operatorname{supp}\widehat{\varphi_J}$, and in particular $\widehat{\varphi_J}(\xi)=0$, we say that $\boldsymbol{\varphi}_{(I,\xi)}$ is a collection of u-adapted wave packets *with mean zero* with respect to the top datum (I,ξ).

In Proposition 3.2, we devise a multi-frequency Calderón-Zygmund decomposition $f=g+b$ of $f\in L^p(\mathbb{R}), 1\le p<2$, adapted to a set of top data $(I,\xi)\in\mathcal{T}$. The L^2 norm of the *good part* g will depend on the L^1 norm of the counting function associated to the spatial intervals of $\mathcal{T}$ and, via u, on a parameter k related to the L^∞ norms of the counting function. The bad part b is such that the Carleson measure norms of the coefficients $|\langle b,\varphi_J\rangle|^2$, associated to collections $\boldsymbol{\varphi}_{(I,\xi)}$ of u-adapted wave packets with mean zero, are *simultaneously* exponentially small in k. The constant C appearing in the statement can be taken equal to 10^3a^{-1}, where a is the uniform adaptation rate of the $\boldsymbol{\varphi}_{(I,\xi)}$'s.

Proposition 3.2. *Let $k\ge 1$ and $\mathcal{T}=\{(I,\xi)\}$ be a collection of top data satisfying*

$$\Big\|\sum_{(I,\xi)\in\mathcal{T}}\mathbf{1}_{3u(Ck)I}\Big\|_\infty\le 2^k. \tag{3.2}$$

Let $f\in L^p(\mathbb{R})$, $1\le p<2$, be given. For each $\lambda>0$, denote by $E_\lambda=\{x\in\mathbb{R}: \mathsf{M}_pf(x)>\lambda\}$. Then, there exists a decomposition $f=g+b$ such that

$$\|g\|_2\lesssim\lambda^{2-p}\|f\|_p^{p-1}\Big(u(Ck)^2\log u(Ck)\Big\|\sum_{(I,\xi)\in\mathcal{T}}\mathbf{1}_I\Big\|_1\Big)^{\frac1p-\frac12} \tag{3.3}$$

and such that, for each $(I,\xi)\in\mathcal{T}$,

$$\sup_{J\not\subset E_\lambda}\frac{|\langle b,\varphi_J\rangle|}{|J|^{\frac12}}\lesssim 2^{-4k}\lambda, \tag{3.4}$$

whenever $\boldsymbol{\varphi}_{(I,\xi)}$ *is a collection of* u*-adapted wave packets with top datum* (I,ξ)*, and*

$$\left(\frac{1}{|J_0|}\sum_{J\not\subset E_\lambda, J\subset J_0}|\langle b,\varphi_J\rangle|^2\right)^{\frac12}\lesssim 2^{-4k}\lambda \tag{3.5}$$

whenever $\boldsymbol{\varphi}_{(I,\xi)}$ *is a collection of* u*-adapted wave packets with mean zero with respect to* (I,ξ)*, and* $J_0\in\mathcal{D}(I), J_0\not\subset E_\lambda$.

The proof is given in Subsection 3.4. In the next subsection, we develop some technical preliminaries involving u-adapted functions.

Remark 3.3 (Dyadic structure of $\operatorname{supp}\widehat{\varphi_J}$). It will be useful to give some sort of dyadic structure to the frequency supports of $\varphi_J\in\boldsymbol{\varphi}_{(I,\xi)}$ as well, working with the standard dyadic grid $\mathcal{D}_0$ and its translates $\mathcal{D}_j:=\{\omega+j|\omega|/3:\omega\in\mathcal{D}_0\}$, $j=1,2$. We do so by selecting for each φ_J the unique interval $\omega\in\mathcal{D}_0\cup\mathcal{D}_1\cup\mathcal{D}_2$ of length $|J|^{-1}$ with minimal $c(\omega)$ such that $(\xi_J-2\varepsilon|J|^{-1},\xi_J+2\varepsilon|J|^{-1})\subset\omega$, which we denote by ω_J. Noting that $\xi\in R\omega_J$ by definition, we realize that the collection $\{\omega_J:|J|=2^\ell\}$ has at most $3R$ elements. By pigeonholing $\boldsymbol{\varphi}_{(I,\xi)}$, at the cost of an additional $\lesssim R$ factor in our estimates, we can assume that all ω_J's come from the same dyadic grid $\mathcal{D}_0$ and that $\omega_J=\omega_{J'}$ whenever $|J|=|J'|$.

3.3. **A splitting of the u-adapted wave packets.** Throughout let $\varphi_J\in\boldsymbol{\varphi}_{(I,\xi)}$ be a collection of u-adapted wave packets. The functions

$$\varphi_J^{(0)}:=\Big(\mathrm{Tr}_{-c(J)}\mathrm{Dil}^2_{|J|^{-1}}\mathrm{Mod}_{-\xi_J}\varphi\Big),\qquad \varphi_J^{(0)}(x)=\varphi\big(c(J)+|J|x\big)\mathrm{e}^{-2\pi i\xi_J(c(J)+|J|x)},$$

whose frequency support lies in $I_0=[-\frac12,\frac12]$, satisfy the decay assumption (2.9) of Lemma 2.3 with a uniform choice of constants A,a. Applying Lemma 2.3 to each $\varphi_J^{(0)}$, for a fixed parameter $K\ge1$, we split $\varphi_J^{(0)}=\phi+\exp(-aK/12)\psi$. The resulting decomposition,

$$\begin{aligned}&\varphi_J=\phi_J+\mathrm{e}^{-\frac{a}{12}K}\psi_J,\\&\phi_J:=\mathrm{Mod}_{\xi_J}\mathrm{Dil}^2_{|J|}\mathrm{Tr}_{c(J)}\phi,\qquad \psi_J:=\mathrm{Mod}_{\xi_J}\mathrm{Dil}^2_{|J|}\mathrm{Tr}_{c(J)}\psi,\end{aligned}\tag{3.6}$$

inherits, from (2.11), (2.12) and (2.13)-(2.14) respectively, the properties

$$\operatorname{supp}\phi_J\subset u(K)J,\tag{3.7}$$

$$|\widehat{\phi_J}(\zeta)|\lesssim \mathrm{e}^{-\frac{a}{12}K}|J|^{\frac12}\big(|J||\zeta-\xi_J|\big)^{-N}\qquad\forall|\zeta-\xi_J|\ge2|J|^{-1},\tag{3.8}$$

$$\left|\big(\mathrm{e}^{-2\pi i\zeta\cdot}\gamma_J(\cdot)\big)^{(j)}(x)\right|\lesssim|J|^{-\frac12-j}\Big(1+\tfrac{|x-c(J)|}{|J|}\Big)^{-N}\quad\forall x\in\mathbb{R},\ j\in\{0,1\},\tag{3.9}$$

the last property holding for either $\gamma_J\in\{\phi_J,\psi_J\}$ and for all $|\zeta-\xi_J|\lesssim R|J|^{-1}$.

If, in addition, $\boldsymbol{\varphi}_{(I,\xi)}$ are u-adapted wave packets *with mean zero* with respect to (I,ξ), we can apply Lemma 2.4 to $\varphi_J^{(0)}$ instead, with choice of frequency $\xi_0:=|J|^{-1}(\xi-\xi_J)\in RI_0\backslash I_0$, and obtain a decomposition (3.6) such that, together with the above properties, there holds

$$\widehat{\phi_J}(\xi)=\widehat{\psi_J}(\xi)=0.\tag{3.10}$$

In the remainder of this subsection, we present several results involving the functions ϕ_J,ψ_J arising from the above decompositions of $\varphi_J\in\boldsymbol{\varphi}_{(I,\xi)}$. We start with the following well-known observation, whose proof relies on integration by parts; see [29, Lemma 2.3].

Lemma 3.4. *Let γ_J be a smooth function satisfying* (3.9) *for $\zeta = \xi \in \mathbb{R}$. Let $H \subset \mathbb{R}$ be an interval, $h \in L^1(\mathbb{R})$ with* $\operatorname{supp} h \subset H$ *and such that*

$$\int_{\mathbb{R}} h(x)\mathrm{e}^{-2\pi i \zeta x}\, \mathrm{d}x = 0 \tag{3.11}$$

for some $\zeta \in \mathbb{R}$ with $|\zeta - \xi| \lesssim R|J|^{-1}$. Then, there holds

$$|\langle h, |J|^{\frac12}\gamma_J\rangle| \lesssim \min\left\{\tfrac{|H|}{|J|}, 1\right\}\left(1 + \tfrac{|c(H)-c(I)|}{|I|}\right)^{-N}\|h\|_1.$$

Using Lemma 3.4 above, as well as the classical Calderón-Zygmund decomposition of h, one obtains the Carleson measure-type estimate of the next lemma, which, aside from notation, is the same as e.g. [29, Proposition 2.4.1].

Lemma 3.5. *Let $\{\gamma_J : J \in \mathcal{D}(I)\}$ be such that each γ_J satisfies* (3.9) *with $\zeta = \xi$, and in addition $\widehat{\gamma_J}(\xi) = 0$. For all $h \in L^1_{\mathrm{loc}}(\mathbb{R})$, $\lambda > 0$, $J_0 \in \mathcal{D}(I)$, $J_0 \not\subset \{\mathsf{M}_1 h > \lambda\}$,*

$$\left(\frac{1}{|J_0|}\sum_{J\not\subset\{\mathsf{M}_1h>\lambda\}, J\subset J_0} |\langle h, \gamma_J\rangle|^2\right)^{\frac12} \lesssim \lambda. \tag{3.12}$$

Of course, one can take $\gamma_J = \varphi_J, \phi_J, \psi_J$ in the above lemmata. Below, we rely on ϕ_J being a smooth function on the torus $3u(K)J$, with exponentially small Fourier coefficients outside the frequency band

$$\Xi(J) := 5\omega_J \cap \frac{\mathbb{Z}}{3u(K)|J|}, \tag{3.13}$$

to show that if an integrable function h is supported on $3u(K)J$ and has mean zero with respect to each $\zeta \in \Xi(J)$, then its integral against $|J|^{\frac12}\phi_J$ is exponentially small.

Lemma 3.6. *Let ϕ_J be a smooth function satisfying* (3.7) *and* (3.8)*. Let $h \in L^1(\mathbb{R})$ with* $\operatorname{supp} h \subset 3u(K)J$ *and such that*

$$\int_{\mathbb{R}} h(x)\mathrm{e}^{-2\pi i \zeta x}\, \mathrm{d}x = 0 \qquad \forall\, \zeta \in \Xi(J). \tag{3.14}$$

Then, there holds

$$|\langle h, |J|^{\frac12}\phi_J\rangle| \lesssim \mathrm{e}^{-\frac{a}{12}K}\|h\|_1. \tag{3.15}$$

Proof. We write $\mu = 3u(K)|J|$ for brevity. In view of the support condition (3.7) and of the decay (3.8), $|J|^{\frac12}\phi_J$ has Fourier coefficients on the torus $3u(K)J$ given by

$$c_\zeta = \frac{1}{\mu}\int_{3u(K)J} |J|^{\frac12}\phi_J(x)\mathrm{e}^{-2\pi i\zeta x} = \frac{|J|^{\frac12}}{\mu}\widehat{\phi_J}(\zeta), \qquad \zeta \in \mu^{-1}\mathbb{Z}.$$

From (3.8), we learn that

$$|c_\zeta| \lesssim \mathrm{e}^{-\frac{a}{12}K}\frac{|J|}{\mu}\big(|J||\zeta - \xi_J|\big)^{-N}, \quad \forall\, \zeta \notin \Xi(J).$$

Using assumption (3.14) for the second equality, and later the last display,

$$\langle h, |J|^{\frac{1}{2}}\phi_J\rangle = \int_{\mathbb{R}} h(x)\Big(\sum_{\zeta\in\mu^{-1}\mathbb{Z}} \overline{c_\zeta}\mathrm{e}^{-2\pi i\zeta x}\Big)\,\mathrm{d}x = \sum_{\substack{\zeta\in\mu^{-1}\mathbb{Z}\\ |J||\zeta-\xi_J|>2}} \overline{c_\zeta}\langle h, \mathrm{e}^{2\pi i\zeta\cdot}\rangle$$

$$\leq \|h\|_1 \sum_{\substack{\zeta\in\mu^{-1}\mathbb{Z}\\ |J||\zeta-\xi_J|>2}} |c_\zeta| \lesssim \mathrm{e}^{-\frac{a}{12}K}\|h\|_1 \sum_{\substack{\zeta\in\mu^{-1}\mathbb{Z}\\ |J||\zeta-\xi_J|>2}} \frac{|J|}{\mu}\big(|J||\zeta-\xi_J|\big)^{-N} \lesssim \mathrm{e}^{-\frac{a}{12}K}\|h\|_1$$

as claimed. The final inequality is obtained by interpreting the last summation over ζ as a Riemann sum. □

With the above lemmata in hand, we devise an exponentially small estimate for the discrete square function involving the ϕ_J's associated to the top datum (I,ξ) when $h\in L^1(\mathbb{R})$ is supported on H and has zero average against a set of $\lesssim Ku(K)$ frequencies near ξ, defined in (3.16).

Lemma 3.7. *For a top datum (I,ξ) and an interval $H\subset\mathbb{R}$ with $I\not\subset 3H$, $H\cap 3u(K)I\neq\emptyset$, define*
(3.16)

$$\Xi_H(I,\xi) := \{\xi\}\cup\bigcup\big\{\Xi(J) : J\in\mathcal{D}(I), J\not\subset 3H,\, H\subset 3u(K)J, |J|\leq 2^{10K}|H|\big\}.$$

Let $h\in L^1(\mathbb{R})$ with $\operatorname{supp} h\subset H$ and such that

$$\int_{\mathbb{R}} h(x)\mathrm{e}^{-2\pi i\zeta x}\,\mathrm{d}x = 0 \qquad \forall\zeta\in\Xi_H(I,\xi). \tag{3.17}$$

Let $\{\phi_J : J\in\mathcal{D}(I)\}$ be a collection of functions each satisfying (3.7) *and* (3.8). *Then*

$$\left\|\left(\sum_{J\not\subset 3H} |\langle h,\phi_J\rangle|^2\frac{\mathbf{1}_J}{|J|}\right)^{\frac{1}{2}}\right\|_1 \lesssim \mathrm{e}^{-\frac{a}{24}K}\|h\|_1. \tag{3.18}$$

Remark 3.8. Before entering the proof, observe that, if $H\cap 3u(K)I=\emptyset$ or $I\subset 3H$, then the square function in (3.18) is identically zero. Furthermore, as a consequence of the definition and of the discussion in Remark 3.3, $\Xi(J)=\Xi(J')$ whenever $J,J'\in\mathcal{D}(I), |J|=|J'|=2^{-\ell}|I|$. Below, we refer to this common discrete interval, depending only on ℓ and (I,ξ), by $\Xi(I,\xi,\ell)$, and we record the following observations.

- If $H\cap 3u(K)I\neq\emptyset, H\cap 9I=\emptyset$, then $\#\Xi_H(I,\xi)\lesssim u(K)\log u(K)$. Indeed, $3u(K)J\cap H\neq\emptyset$ only if $|I|\lesssim u(K)|J|$. Thus $\Xi_H(I,\xi)$ is contained in the union over $0\leq\ell\lesssim\log u(K)$ of the $\Xi(I,\xi,\ell)$, each of which has $\lesssim u(K)$ elements.
- Otherwise, within the assumptions of the lemma, it must be that $H\subset 9I$. In this case, reasoning as above yields that $\Xi_H(I,\xi)\lesssim Ku(K)$.

Proof of Lemma 3.7. By virtue of the support condition (3.7), the left hand side of (3.18) is bounded by

$$\sum_{\substack{J\not\subset 3H\\ u(K)J\cap H\neq\emptyset}} |\langle h,|J|^{\frac{1}{2}}\phi_J\rangle| = \sum_{\substack{J\not\subset 3H\\ H\subset 3u(K)J}} |\langle h,|J|^{\frac{1}{2}}\phi_J\rangle| = \sum_{\ell\gtrsim-\log u(K)}\sum_{J\in\boldsymbol{J}_\ell} |\langle h,|J|^{\frac{1}{2}}\phi_J\rangle|$$

where $\boldsymbol{J}_\ell := \{J \not\subset 3H, H \subset 3u(K)J, |J| \sim 2^\ell |H|\}$. Note that $\#\boldsymbol{J}_\ell \lesssim u(K)\log u(K)$. Now, for the $\lesssim K$ values $-\log u(K) \lesssim \ell \leq 10K$, in view of assumption (3.17), we apply Lemma 3.6 and estimate $|\langle h, |J|^{\frac12}\phi_J\rangle| \lesssim \mathrm{e}^{-\frac{a}{12}K}\|h\|_1$. On the other hand, when $\ell \geq 10K$, we use Lemma 3.4 and estimate $|\langle h, |J|^{\frac12}\phi_J\rangle| \lesssim |H|\|h\|_1/|J| \lesssim 2^{-\ell}\|h\|_1$. Summarizing, the last display is bounded by

$$u(K)\log u(K)\Big(K\mathrm{e}^{-\frac{a}{12}K} + \sum_{\ell\geq 10K} 2^{-\ell}\Big)\|h\|_1 \lesssim \mathrm{e}^{-\frac{a}{24}K}\|h\|_1,$$

where the last inequality follows from (2.2). The proof is complete. □

3.4. **Proof of Proposition 3.2.** By linearity and dilation invariance of assumptions and conclusions, we can assume $\|f\|_p = 1 = \lambda$. In the proof, we write $K := Ck$, with $C = 10^3 a^{-1}$. Let $Q \in \boldsymbol{Q}$ be the collection of maximal dyadic intervals such that $9Q \subset E_\lambda$. Then

$$\sup_{Q\in\boldsymbol{Q}} \|f\|_{L^p(Q)} \lesssim 1, \qquad \sup_{x\notin\bigcup_{Q\in\boldsymbol{Q}} Q} f(x) \lesssim 1, \qquad \sum_{Q\in\boldsymbol{Q}} |Q| \leq |E_\lambda| \lesssim 1, \tag{3.19}$$

and the intervals $\{3Q : Q \in \boldsymbol{Q}\}$ have finite overlap. By virtue of the second property in (3.19), there is no loss in generality with assuming that f is supported on $\bigcup_{Q\in\boldsymbol{Q}} Q$. Also, we erase from $\mathcal{T}$ those (I,ξ) with $I \subset 9Q$ for some $Q \in \boldsymbol{Q}$, since (3.5) is zero for such an (I,ξ).

Construction of g and b. For each $Q \in \boldsymbol{Q}$, referring to the notation in (3.16) with the choice of $H = 3Q$, define

$$\Xi_Q = \bigcup_{3Q\cap 3u(K)I\neq\emptyset} \Xi_{3Q}(I,\xi) \tag{3.20}$$

and denote

$$\mathcal{T}_{\mathrm{far}}(Q) := \big\{(I,\xi)\in\mathcal{T} : 3Q\cap 9I = \emptyset, 3Q \subset 3u(K)I\big\},$$
$$\mathcal{T}_{\mathrm{near}}(Q) := \big\{(I,\xi)\in\mathcal{T} : 3Q \subset 9I\big\}.$$

In view of the discussion in Remark 3.8 and using assumption (3.2), we have that

$$\begin{aligned}\#\Xi_Q &\lesssim u(K)\log(u(K))\big(\#\mathcal{T}_{\mathrm{far}}(Q)\big) + (u(K))^2\big(\#\mathcal{T}_{\mathrm{near}}(Q)\big)\\ &\lesssim u(K)^2 \inf_{3Q}\Big(\sum_{(I,\xi)} \mathbf{1}_{3u(K)I}\Big) \lesssim u(K)^2 2^k.\end{aligned} \tag{3.21}$$

As in [26, Theorem 1.1], for each $Q \in \boldsymbol{Q}$ we define g_Q to be the Riesz projection of $f_Q := f\mathbf{1}_Q$ on the finite-dimensional subspace of $L^2(3Q)$ spanned by $\big\{\exp(2\pi i\zeta x) : \zeta \in \Xi_Q\big\}$ and set $b_Q := f_Q - g_Q$, so that each g_Q, b_Q is supported inside $3Q$ and

$$\int_{\mathbb{R}} b_Q(x)\mathrm{e}^{-2\pi i\zeta x}\,\mathrm{d}x = 0 \qquad \forall\zeta\in\Xi_Q. \tag{3.22}$$

The elegant argument by Borwein and Erdelyi [3] (see [26] for a proof) gives the estimate

$$\|g_Q\|_{L^2(3Q)} + \|b_Q\|_{L^p(3Q)} \lesssim (\#\Xi_Q)^{\frac1p-\frac12}\|f_Q\|_{L^p(Q)} \lesssim (\#\Xi_Q)^{\frac1p-\frac12}. \tag{3.23}$$

We finally set $g = \sum_{Q\in\boldsymbol{Q}} g_Q, b = \sum_{Q\in\boldsymbol{Q}} b_Q$.

Consequences of the construction. A preliminary observation is that

$$\{\mathsf{M}_1 b \gtrsim 2^k\} \subset E_\lambda. \tag{3.24}$$

Indeed, for each interval $Z \not\subset E_\lambda$,

$$\|b\|_{L^1(Z)} \le \|b\|_{L^p(Z)} \lesssim \Big(\sum_{\substack{Q\in\boldsymbol{Q}\\ 3Q\subset 3Z}} \tfrac{|Q|}{|Z|} \|b_Q\|^p_{L^p(3Q)} \Big)^{\frac1p} \lesssim \big(2^k u(K)^2\big)^{\frac1p-\frac12} \lesssim 2^k,$$

whence the claim. We took into account that b is supported on the union of $\{3Q \in \boldsymbol{Q}\}$, having finite overlap, and later (3.23) coupled with (3.21).

The next estimate explains the choice of the mean zero frequencies Ξ_Q for b_Q which was done in (3.20). Indeed, we claim that whenever $\{\phi_J : J \in \mathcal{D}(I)\}$ is a collection of functions each satisfying (3.7) and (3.8),

$$\Bigg\| \Bigg(\sum_{J\not\subset 9Q} |\langle b_Q, \phi_J\rangle|^2 \frac{\mathbf{1}_J}{|J|} \Bigg)^{\frac12} \Bigg\|_1 \lesssim \mathrm{e}^{-\frac{a}{24}K} \|b_Q\|_1 \lesssim 2^{-10k}|Q|, \tag{3.25}$$

for all $Q \in \boldsymbol{Q}$. The last step simply follows from (3.23) and (3.21). To obtain the first inequality, we have used that if the left hand side of (3.25) is nontrivial, b_Q has mean zero with respect to all the frequencies $\Xi_{3Q}(I,\xi) \subset \Xi_Q$ appearing in (3.17), and consequently appealed to Lemma 3.7.

Proof of (3.3). The function

$$F := u(K)\log(u(K)) \sum_{(I,\xi)\in\mathcal{T}} \mathbf{1}_{3u(K)I} + u(K)^2 \sum_{(I,\xi)\in\mathcal{T}} \mathbf{1}_{9I}$$

satisfies

$$\|F\|_1 \lesssim u(K)^2 \log(u(K)) \Big\| \sum_{(I,\xi)\in\mathcal{T}} \mathbf{1}_I \Big\|_1, \qquad \Xi_Q \le \inf_{x\in 3Q} F(x),$$

the second inequality coming from the first line of (3.21). Making use of the finite overlap of the $3Q$'s and relying on (3.19),

$$\begin{aligned} \|g\|_2^2 &\lesssim \sum_{Q\in\mathbf{Q}} \|g_Q\|_2^2 \lesssim \sum_{Q\in\mathbf{Q}} |Q|(\#\Xi_Q)^{\frac2p-1} \le \Big(\sum_{Q\in\mathbf{Q}} |Q|\Big)^{2-\frac2p} \|F\|_1^{\frac2p-1} \\ &\lesssim \Big(u(K)^2 \log(u(K)) \Big\| \sum_{(I,\xi)\in\mathcal{T}} \mathbf{1}_I \Big\|_1 \Big)^{\frac2p-1}, \end{aligned}$$

as claimed. ∎

Proof of (3.4). Fix $(I,\xi) \in \mathcal{T}$, a collection of u-adapted wave packets $\boldsymbol{\varphi}_{(I,\xi)}$ and $\varphi_J \in \boldsymbol{\varphi}_{(I,\xi)}$ with $J \not\subset E_\lambda$. To bound $|J|^{-\frac12}|\langle b, \varphi_J\rangle|$, we rely on the decomposition (3.6) of φ_J and on properties (3.7)-(3.9) of ϕ_J, ψ_J. It is easy to see that

$$\mathrm{e}^{-\frac{a}{12}K} \frac{|\langle b, \psi_J\rangle|}{|J|^{\frac12}} \lesssim \mathrm{e}^{-\frac{a}{12}K} \inf_{x\in J} \mathsf{M}_1 b(x) \lesssim 2^{-4k}, \tag{3.26}$$

where the last inequality follows from (3.24) and $J \not\subset E_\lambda$. We then use (3.25) to estimate

$$\frac{|\langle b_Q, \phi_J\rangle|}{|J|^{\frac12}} \le \frac{|Q|}{|J|} \Bigg\| \Bigg(\sum_{J\not\subset 9Q} |\langle b_Q, \phi_J\rangle|^2 \frac{\mathbf{1}_J}{|J|} \Bigg)^{\frac12} \Bigg\|_1 \lesssim \frac{|Q|}{|J|} 2^{-10k}. \tag{3.27}$$

With this in hand, by virtue of the support condition (3.7),

$$\frac{|\langle b,\phi_J\rangle|}{|J|^{\frac12}} \le \sum_{3Q\subset 3u(K)J} \frac{|\langle b_Q,\phi_J\rangle|}{|J|^{\frac12}} \lesssim 2^{-10k} \sum_{3Q\subset 3u(K)J} \frac{|Q|}{|J|} \lesssim 2^{-4k},$$

and (3.4) follows by combining the last display with (3.26), in view of (3.6). □

Proof of (3.5). Let $\varphi_{(I,\xi)}$ be a collection of u-adapted wave packets *with mean zero* with respect to $(I,\xi)\in\mathcal{T}$. Here, we rely on the decomposition (3.6) of φ_J, with ϕ_J, ψ_J satisfying, in addition to (3.7)-(3.9), property (3.10), so that in particular $\widehat{\psi_J}(\xi)=0$ for all J. In view of this decomposition, (3.5) will follow from estimating, for all $J_0\in\mathcal{D}(I)$, $J_0\not\subset E_\lambda$,

$$\left(\frac{1}{|J_0|}\sum_{J\not\subset E_\lambda, J\subset J_0} |\langle b,\phi_J\rangle|^2\right)^{\frac12} \lesssim 2^{-4k}, \tag{3.28}$$

$$e^{-\frac{a}{12}K}\left(\frac{1}{|J_0|}\sum_{J\not\subset E_\lambda, J\subset J_0} |\langle b,\psi_J\rangle|^2\right)^{\frac12} \lesssim 2^{-4k}. \tag{3.29}$$

We first prove (3.29), which is easier. We have

$$e^{-\frac{a}{12}K}\left(\frac{1}{|J_0|}\sum_{\substack{J\not\subset E_\lambda\\ J\subset J_0}} |\langle b,\psi_J\rangle|^2\right)^{\frac12} \le e^{-\frac{a}{12}K}\left(\frac{1}{|J_0|}\sum_{\substack{J\not\subset \{\mathsf{M}_1 b\gtrsim\lambda 2^k\}\\ J\subset J_0}} |\langle b,\psi_J\rangle|^2\right)^{\frac12} \lesssim 2^{-4k},$$

employing (3.24) for the first step and then applying Lemma 3.5 for the second inequality, in view of (3.9) and of $\widehat{\psi_J}(\xi)=0$ for all J.

Once we establish the inequality

$$\frac{1}{|J_0|}\left\|\left(\sum_{\substack{J\not\subset E_\lambda\\ J\subset J_0}} |\langle b,\phi_J\rangle|^2\frac{\mathbf{1}_J}{|J|}\right)^{\frac12}\right\|_1 \lesssim 2^{-4k} \qquad \forall J_0\in\mathcal{D}(I), J_0\not\subset E_\lambda, \tag{3.30}$$

estimate (3.28), which is what is left to show in order to be done with (3.5), can be reached via the same John–Nirenberg-type argument used in the proof of [29, Proposition 2.4.1]. We prove (3.30) via the chain of inequalities

$$\begin{aligned}
\left\|\left(\sum_{\substack{J\not\subset E_\lambda\\ J\subset J_0}} |\langle b,\phi_J\rangle|^2\frac{\mathbf{1}_J}{|J|}\right)^{\frac12}\right\|_1 &= \left\|\left(\sum_{\substack{J\not\subset E_\lambda\\ J\subset J_0}} \Big|\Big\langle \sum_{Q:3Q\subset 3u(K)J_0} b_Q,\phi_J\Big\rangle\Big|^2\frac{\mathbf{1}_J}{|J|}\right)^{\frac12}\right\|_1\\
&\le \sum_{Q:3Q\subset 3u(K)J_0}\left\|\left(\sum_{\substack{J\not\subset 9Q\\ J\subset J_0}} |\langle b_Q,\phi_J\rangle|^2\frac{\mathbf{1}_J}{|J|}\right)^{\frac12}\right\|_1\\
&\lesssim 2^{-10k}\sum_{Q:3Q\subset 3u(K)J_0}|Q| \lesssim 2^{-10k}u(K)|J_0| \lesssim 2^{-4k}|J_0|,
\end{aligned}$$

where (3.25) has been used, for each b_Q, for the third step. This concludes the proof of (3.5) and in turn of Proposition 3.2. □

4. The model sums for $\Lambda_{\vec\beta}$

This section is dedicated to the discretization of the trilinear forms $\Lambda_{\vec\beta}$ into the model sums of (4.3) below. As in Section 3, to which we refer, our definitions depend on a choice of function u complying with the assumptions of Section 2. We keep this dependence implicit in the notation. A concrete choice of u will be made in Section 7.

4.1. **Tiles and u-wave packets.**

We call *tile* $t = I_t \times \omega_t$ the cartesian product of two intervals $I_t, \omega_t \subset \mathbb{R}$ with $|I_t||\omega_t| = 1$. We say that a Schwartz function φ_t is a u-wave packet adapted to the tile t if φ_t is u-adapted to $(I_t, c(\omega_t))$ in the sense of (3.1) and in addition $\operatorname{supp}\widehat{\varphi}_t \subset \omega_t$.

Referring to Remark 3.1, a u-wave packet adapted to the tile t is given by

$$\upsilon_t := \upsilon_{(I_t,c(\omega_t)),\varepsilon} = \mathrm{Mod}_\xi \mathrm{Dil}^2_{\varepsilon^{-1}|I|}\mathrm{Tr}_{c(I)}\upsilon. \tag{4.1}$$

The u-wave packets υ_t will be employed in the constructions of our model sums, with a suitable choice of ε. Again from Remark 3.1, we infer that the implicit constants for υ_t in (3.1) will depend only on our choice of u, and that the adaptation rate a will be a positive absolute constant.

4.2. **Tritiles and model sums.**

A *tritile* s is a triplet of tiles $s_j = I_{s_j} \times \omega_{s_j}$, $j = 1, 2, 3$, with $I_{s_1} = I_{s_2} = I_{s_3} =: I_s$. A collection of tritiles $s \in \mathbf{S}$ is well-discretized (with constant $R > 10$) if the following properties are satisfied:

- the collections $\{I_s : s \in \mathbf{S}\}$ and $\{10\omega_{s_j} : s \in \mathbf{S}\}$, $j \in \{1, 2, 3\}$, form grids;
- if $s, s' \in \mathbf{S}$ with $|I_s| < |I_{s'}|$, then $|I_s| \le R^{-10}|I_{s'}|$ (separation of scales);
- if $s \ne s' \in \mathbf{S}$ are such that $I_s = I_{s'}$, then $\omega_{s_j} \cap \omega_{s'_j} = \emptyset$ for each $j \in \{1, 2, 3\}$;
- if $s, s' \in \mathbf{S}$ with $|I_{s'}| \le |I_s|$ and $2\omega_{s_j} \cap 2\omega_{s'_j} \ne \emptyset$ for some $j \in \{1, 2, 3\}$, there holds

$$10\omega_{s_k} \cap 10\omega_{s'_k} = \emptyset \quad \forall k \ne j, \qquad R\omega_{s_k} \subset R\omega_{s'_k},\ \forall k \in \{1, 2, 3\}. \tag{4.2}$$

For a tritile s and $j = 1, 2, 3$, we denote by υ_{s_j} the s_j-adapted function obtained from (4.1) with $t = s_j$. Consider the model sums

$$\mathsf{\Lambda}_{\mathbf{S}}(f_1, f_2, f_3) := \sum_{s\in\mathbf{S}} \varepsilon_s |I_s|^{-\frac12} \prod_{j=1}^{3} \langle f_j, \upsilon_{s_j}\rangle, \qquad |\varepsilon_s| \le 1, \tag{4.3}$$

where $\mathbf{S}$ is an arbitrary finite collection of well-discretized tritiles. In the next subsection, we show how estimates for the trilinear form $\Lambda_{\vec\beta}$ are obtained from the corresponding, uniform over all finite collections of well-discretized tritiles $\mathbf{S}$, bounds for the model sums (4.3). Note that we are allowing for bounded coefficients ε_s, so that the finiteness assumption on $\mathbf{S}$ can be removed by a standard limiting argument.

4.3. **Discretization of the trilinear forms $\Lambda_{\vec\beta}$.**

To each triple $(\tau, x, \xi) \in (0,\infty)\times\mathbb{R}^2$, we associate the tile

$$t(\tau, x, \xi) := \big(x - \tfrac{\tau}{2}, x + \tfrac{\tau}{2}\big) \times \big(\xi - \tfrac{1}{2\tau}, \xi + \tfrac{1}{2\tau}\big).$$

For a unit vector $\vec\beta \in \mathbb{R}^3$ with $\vec\beta \cdot (1, 1, 1) = 0$, let $\vec\gamma \in \mathbb{R}^3$ be the unique unit vector such that $\vec\gamma, \vec\beta, (1, 1, 1)$ form a positively oriented orthogonal basis of $\mathbb{R}^3$. Choosing

$\varepsilon = 2^{-16}$, writing $\lambda = 1+\varepsilon$ for brevity and recalling that $\widehat{\upsilon}(0)$ is positive, we quote from [11, Section 6] the equality

$$\begin{aligned}(4.4)\qquad &\Lambda_\beta(f_1,f_2,f_3)\\ &= c_1\int_{\mathbb{R}^3}\lambda^{-\frac{\sigma}{2}}\Big(\prod_{j=1}^3\big\langle f_j,\upsilon_{t(\lambda^\sigma,x,\gamma_j\xi+\beta_j\lambda^{-\sigma})}\big\rangle\Big)\,\mathrm{d}\sigma\mathrm{d}x\mathrm{d}\xi + c_2\int_{\mathbb{R}}\Big(\prod_{j=1}^3 f_j(y)\Big)\,\mathrm{d}y\end{aligned}$$

holding for any three Schwartz functions f_j, where c_1, c_2 are non-zero constants depending on υ, ε only. The second summand on the right hand side of (4.4) is bounded by Hölder's inequality, so that estimates on Λ_β can be deduced from corresponding bounds on the triple integral. For triples $\boldsymbol{m} = (m_\sigma, m_x, m_\xi) \in \mathbb{Z}^3, \boldsymbol{\vartheta} := (\vartheta_\sigma,\vartheta_x,\vartheta_\xi) \in [0,1)^3$, we define the tritile $s^{(\boldsymbol{m},\boldsymbol{\vartheta})}$ by

$$s_j^{(\boldsymbol{m},\boldsymbol{\vartheta})} = \Big(\lambda^{m_\sigma+\vartheta_\sigma}(m_x+\vartheta_x) \pm \frac{\lambda^{m_\sigma+\vartheta_\sigma}}{2}\Big)\times\Big(\frac{\gamma_j(m_\xi+\vartheta_\xi)+\beta_j}{\lambda^{m_\sigma+\vartheta_\sigma}} \pm \frac{\lambda^{-(m_\sigma+\vartheta_\sigma)}}{2}\Big),$$

for $j = 1,2,3$. By suitably splitting the integration regions, the integral over $\mathbb{R}^3$ in (4.4) is equal to

$$(4.5)\qquad \int_{[0,1)^3}\Bigg(\sum_{\boldsymbol{m}\in\mathbb{Z}^3}|I_{s^{(\boldsymbol{m},\boldsymbol{\vartheta})}}|^{-\frac12}\prod_{j=1}^3\langle f_j,\upsilon_{s_j^{(\boldsymbol{m},\boldsymbol{\vartheta})}}\rangle\Bigg)\,\mathrm{d}\boldsymbol{\vartheta}.$$

Arguing exactly as in [29, pp. 50-51], for each fixed $\boldsymbol{\vartheta}$, $\{s^{(\boldsymbol{m},\boldsymbol{\vartheta})} : \boldsymbol{m}\in\mathbb{Z}^3\}$ can be decomposed into a finite union of well-discretized collections of tritiles, provided that the constant $R > 10$ appearing in the definition is chosen large enough, depending on the distance $\Delta_{\vec\beta}$ from the degenerate case. Therefore, L^p bounds for the triple integral in (4.5) and, in turn, for $\Lambda_{\vec\beta}$ follow by averaging the corresponding inequalities for the model sums (4.3).

Remark 4.1. By the same token, estimates of the type

$$\Lambda_{\vec\beta}(f_1,f_2,f_3) \le Q(\|f_1\|_{\frac{1}{\alpha_1}}, \|f_2\|_{\frac{1}{\alpha_2}}, |F_3|^{\alpha_3})$$

where Q is a certain positive function of its arguments, holding for possibly restricted f_1, f_2 and for all sets $F_3\subset\mathbb{R}$, with f_3 restricted to a suitable major subset F_3' of F_3, are obtained by averaging the bound of the same type for (4.3), *provided that F_3' can be chosen independently of the model sum.* This will be the case in the proofs of our theorems.

Remark 4.2 (Scale invariance of the model sums). Let $\vec\alpha$ be a Hölder triplet. For each $\mu > 0$, we have the equality

$$\Lambda_{\mathbf{S}}(f_1,f_2,f_3) = \sum_{s\in\mathbf{S}}\varepsilon_s|I_{s^\mu}|^{-\frac12}\prod_{j=1}^3\langle \mathrm{Dil}_\mu^{\frac{1}{\alpha_j}} f_j, \upsilon_{(s^\mu)_j}\rangle = \Lambda_{\mathbf{S}^\mu}(\mathrm{Dil}_\mu^{\frac{1}{\alpha_1}}f_1, \mathrm{Dil}_\mu^{\frac{1}{\alpha_2}}f_2, \mathrm{Dil}_\mu^{\frac{1}{\alpha_3}}f_3)$$

where $\mathbf{S}^\mu := \{s^\mu : s\in\mathbf{S}\}$ and each tritile s^μ is given by $I_{s^\mu} = T_\mu I_s$, $\omega_{(s^\mu)_j} = (T_\mu)^{-1}\omega_{s_j}$. Here, T_μ is the linear transformation $x\to\mu x$; note that, in general, μI_s and $T_\mu I_s$ are not the same. When μ is a power of R, the collection $\mathbf{S}^\mu$ is again a well-discretized collection of tritiles according to the definition of Subsection 4.2, so that the family of trilinear forms (4.3) is invariant under dyadic Hölder-type scaling.

5. Trees, size, and single tree estimates

We summarize the main definitions and technical tools needed to manufacture bounds for the model sums (4.3) in the framework of [17]. Our treatment deviates from the classical one in that we work with model sums involving uniformly u-adapted wave packets $\{\varphi_{s_j} : j = 1, 2, 3\}$, indexed over a generic well-discretized collection of tritiles $s \in \boldsymbol{S}$.

5.1. **Trees and size.** In our context, a *tree of tiles* $\boldsymbol{t}$ with *top datum* $(I_{\boldsymbol{t}}, \xi_{\boldsymbol{t}})$ is a finite collection of tiles such that $I_t \subset I_{\boldsymbol{t}}$, and $\xi_{\boldsymbol{t}} \in R\omega_t$ for each $t \in \boldsymbol{t}$. For our scopes, the technical requirement that $\{I_t : t \in \boldsymbol{t}\}$ is a grid will always be satisfied. If $\xi_{\boldsymbol{t}} \in R\omega_t \backslash 2\omega_t$ for all $t \in \boldsymbol{t}$, the tree $\boldsymbol{t}$ is called *lacunary.* We associate to each lacunary tree $\boldsymbol{t}$ and each $f \in L^1(\mathbb{R}) + L^\infty(\mathbb{R})$ the quantity

$$\operatorname{size}(f; \boldsymbol{t}) := \sup_{\{\varphi_t : t \in \boldsymbol{t}\}} \Big(\frac{1}{|I_{\boldsymbol{t}}|} \sum_{t \in \boldsymbol{t}} |\langle f, \varphi_t \rangle|^2 \Big)^{\frac{1}{2}},$$

the supremum being taken over all collections $\{\varphi_t : t \in \boldsymbol{t}\}$ of uniformly adapted u-wave packets. Arguing along the lines of Remark 3.3, each such collection can be written as a union of at most $\lesssim R$ collections of u-adapted wave packets with top datum (I, ξ), as defined in Section 3. We can thus reformulate the conclusion of Lemma 3.5 into the estimate

$$\operatorname{size}(f; \boldsymbol{t}) \lesssim \sup_{t \in \boldsymbol{t}} \inf_{x \in I_t} \mathsf{M}_1 f(x). \tag{5.1}$$

We give related definitions for tritiles. A *tree* of tritiles $\mathbf{T}$ *of type* $j \in \{1, 2, 3\}$ *with top datum* $(I_{\mathbf{T}}, \xi_{\mathbf{T}})$ (in short, j-tree) is a finite well-discretized collection of tritiles such that $I_s \subset I_{\mathbf{T}}$, $\xi_{\mathbf{T}} \in 2\omega_{s_j}$ for each $s \in \mathbf{T}$. A consequence of (4.2) is that if $\mathbf{T}$ is a j-tree, for $k \neq j$ the intervals $\{10\omega_{s_k} : s \in \mathbf{T}\}$ are pairwise disjoint while $\{R\omega_{s_k} : s \in \mathbf{T}\}$ are nested. It follows that there exists a frequency $\xi_{\mathbf{T},k}$ such that $\xi_{\mathbf{T},k} \in R\omega_{s_k} \backslash 2\omega_{s_k}$ for all $s \in \mathbf{T}$. In other words, for $k \neq j$ the collection $\mathbf{T}^{(k)} := \{s_k : s \in \mathbf{T}\}$ is a lacunary tree of tiles with top datum $(I_{\mathbf{T}}, \xi_{\mathbf{T},k})$.

For each finite, well-discretized collection of tritiles $\mathbf{S}$, each $f \in L^1(\mathbb{R}) + L^\infty(\mathbb{R})$, and each $j = 1, 2, 3$ we define

$$\operatorname{size}_j(f; \mathbf{S}) := \sup_{\substack{\mathbf{T} \subset \mathbf{S} \text{ is a tree} \\ \mathbf{T}^{(j)} \text{ is a lacunary tree}}} \operatorname{size}(f; \mathbf{T}^{(j)}), \tag{5.2}$$

inheriting from (5.1) the bound

$$\operatorname{size}_j(f; \mathbf{S}) \lesssim \sup_{s \in \mathbf{S}} \inf_{x \in I_s} \mathsf{M}_1 f(x). \tag{5.3}$$

The quantities size_j enter the following *single tree estimate* for model sums of the type (4.3).

Lemma 5.1. *Let* $\mathbf{T}$ *be a tree of tritiles. Then*

$$\sum_{s \in \mathbf{T}} |I_s|^{-\frac{1}{2}} \prod_{k=1}^{3} |\langle f_k, \varphi_{s_k} \rangle| \lesssim |I_{\mathbf{T}}| \prod_{k=1}^{3} \operatorname{size}_k(f; \mathbf{T}).$$

Proof. If $\mathbf{T}$ is a j-tree, the $\ell^\infty \times \ell^2 \times \ell^2$ Hölder inequality yields

$$\sum_{s \in \mathbf{T}} |I_s|^{-\frac{1}{2}} \prod_{k=1}^{3} |\langle f_k, \varphi_{s_k} \rangle| \leq \Big(\sup_{s \in \mathbf{T}} \frac{|\langle f_j, \varphi_{s_j} \rangle|}{|I_s|^{\frac{1}{2}}} \Big) \prod_{k \neq j} \Big(\sum_{s \in \mathbf{T}} |\langle f_k, \varphi_{s_k} \rangle|^2 \Big)^{\frac{1}{2}}.$$

Viewing $\{s_j\}$ as a lacunary tree, the first factor is $\lesssim \mathrm{size}_j(f;\mathbf{T})$. The second and third factors are each $\lesssim \mathrm{size}_k(f;\mathbf{T})|I_{\mathbf{T}}|^{\frac12}$, since $\mathbf{T}^{(k)}$ is a lacunary tree for $k \neq j$. This concludes the proof of the lemma. □

5.2. **The size lemma and a forest estimate.** The following lemma, known as the *size lemma*, is used to iteratively decompose a collection $\mathbf{S}$ into subcollections which are unions of trees of definite size (also known as *forests*). There is an abundance of analogous results in the literature: for the proof, we refer to [29, Lemma 5.3] and [25, Lemma 7.7].

Lemma 5.2. *Let* $\mathbf{S}$ *be a finite, well-discretized collection of tritiles and* $f \in L^2(\mathbb{R})$ *such that* $\mathrm{size}_j(f;\mathbf{S}) \le \sigma$. *Then* $\mathbf{S} = \mathbf{S}_{\mathrm{lo}} \cup \mathbf{S}_{\mathrm{hi}}$, *where*

$$\mathrm{size}_j(f;\mathbf{S}_{\mathrm{lo}}) \le \tfrac{\sigma}{2}, \tag{5.4}$$

$$\mathbf{S}_{\mathrm{hi}} \text{ is a disjoint union of trees } \mathbf{T}\in\mathcal{F} \text{ with } \Big(\sum_{\mathbf{T}\in\mathcal{F}} |I_{\mathbf{T}}|\Big)^{\frac12} \lesssim \frac{\|f\|_2}{\sigma}. \tag{5.5}$$

If in addition $f \in L^\infty(\mathbb{R})$, *for each interval* $I \subset \mathbb{R}$ *we have*

$$\frac{1}{|I|}\sum_{\mathbf{T}\in\mathcal{F}: I_{\mathbf{T}}\subset I} |I_{\mathbf{T}}| \lesssim \frac{\|f\|_\infty^2}{\sigma^2}. \tag{5.6}$$

It is convenient to combine Lemmata 5.1 and 5.2 into an estimate for model sums restricted to a union of trees satisfying a certain relation between size and counting functions. This result is analogous to [7, Lemma 4.4], but we include the proof for convenience.

Lemma 5.3. *Let* $f_3 \in L^2(\mathbb{R})$ *be given and* $\mathbf{S}$ *be a finite, well-discretized collection of tritiles. Assume that* $\mathbf{S}$ *can be written as a disjoint union of trees* $\mathbf{T}\in\mathcal{F}$ *satisfying, for some* $A > 0$,

$$\Big(\sum_{\mathbf{T}\in\mathcal{F}} |I_{\mathbf{T}}|\Big)^{\frac12} \le \frac{A}{\mathrm{size}_3(f_3;\mathbf{S})}. \tag{5.7}$$

Then, for all $f_1 \in L^2(\mathbb{R})$, $f_2 \in L^1(\mathbb{R}) + L^\infty(\mathbb{R})$,

$$\sum_{s\in\mathbf{S}} |I_s|^{-\frac12}\prod_{j=1}^3 |\langle f_j,\varphi_{s_j}\rangle| \lesssim A\|f_1\|_2 \mathrm{size}_2(f_2;\mathbf{S}). \tag{5.8}$$

Proof. Denote $\sigma_j = \mathrm{size}_j(f_j;\mathbf{S})$ for $j = 1,2,3$. By linearity in f_1, f_2 we can assume $\|f_1\|_2 = 1, \sigma_2 = 1$. Let $n_0 = \lceil \log\sigma_1 - \log\sigma_3 + \log A\rceil$. There are two cases: if $n_0 \le 0$, in other words $A\sigma_1 \le \sigma_3$, the left hand side of (5.8) is bounded, using Lemma 5.1, by

$$\sum_{\mathbf{T}\in\mathcal{F}}\sum_{s\in\mathbf{T}} |I_s|^{-\frac12}\prod_{j=1}^3 |\langle f_j,\varphi_{s_j}\rangle| \lesssim \prod_{j=1}^3 \mathrm{size}_j(f_j;\mathbf{S}) \sum_{\mathbf{T}\in\mathcal{F}} |I_{\mathbf{T}}| \lesssim A^2\sigma_1\sigma_3^{-1} \le A,$$

which is what we had to prove. Otherwise, we decompose $\mathbf{S}$ into collections $\mathbf{S}_n, n = 0,\ldots,n_0$, each being a disjoint union of trees $\mathbf{T}\in\mathcal{F}_n$ such that

$$\sum_{\mathbf{T}\in\mathcal{F}_n} |I_{\mathbf{T}}| \lesssim 2^{2n}(\sigma_1)^{-2},$$

$$\mathrm{size}_1(f_1;\mathbf{S}_n) \le 2^{-n}\sigma_1, \quad \mathrm{size}_2(f_2;\mathbf{S}_n) \le 1, \quad \mathrm{size}_3(f_3;\mathbf{S}_n) \le \sigma_3,$$

by iteratively applying the size lemma with $f = f_1$ for $n = 0, \ldots, n_0 - 1$, and by organizing the leftover collection $\mathbf{S}_{n_0}$ into a disjoint union of trees $\mathcal{F}_{n_0} := \{\mathbf{T} \cap \mathbf{S}_{n_0} : \mathbf{T} \in \mathcal{F}\}$. For this last collection, the first bound in the last display is inherited from (5.7). Applying again the single tree estimate for each $\mathbf{T} \in \mathcal{F}_n$, the left hand side of (5.8) is bounded by

$$\sum_{n=0}^{n_0} \sum_{\mathbf{T} \in \mathcal{F}_n} |I_{\mathbf{T}}| \prod_{j=1}^{3} \operatorname{size}_j(f_j; \mathbf{S}_n) \lesssim \sum_{n=0}^{n_0} 2^{n \frac{\sigma_3}{\sigma_1}} \lesssim A,$$

which finishes the proof. □

6. Forest estimates

Let $\mathbf{S}$ be a finite, well-discretized collection of tritiles and $\{\varphi_{s_j} : s \in \mathbf{S}, j = 1, 2, 3\}$ be a collection of uniformly u-adapted wave packets. We will provide several estimates on

$$\Lambda_{\mathbf{S}}(f_1, f_2, f_3) = \sum_{s \in \mathbf{S}} \varepsilon_s |I_s|^{-\frac{1}{2}} \prod_{j=1}^{3} \langle f, \varphi_{s_j} \rangle$$

when f_3 is bounded by 1 and supported on $F_3 \subset \mathbb{R}$ of finite measure, or on a suitable major subset of F_3. Note that the model sum (4.3) is a particular instance of the above display. In what follows, for any $\mathbf{S}' \subset \mathbf{S}$, we are indicating with $\Lambda_{\mathbf{S}'}$ that we are summing over $s \in \mathbf{S}'$.

6.1. **Estimates inside exceptional sets.** The first estimate deals with the case of $\mathbf{S}$ being localized inside the superlevel sets of the maximal functions of h_1, h_2.

Proposition 6.1. *Let $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ be a Hölder tuple with $0 \le \alpha_1, \alpha_2 \le 1, \alpha_3 \ge -\frac{1}{2}$. For functions $h_j \in L^{\frac{1}{\alpha_j}}(\mathbb{R})$, $j = 1, 2$, and a set $F_3 \subset \mathbb{R}$ of finite measure, define*

$$E^{\vec{\alpha}}_{h_1,h_2,F_3} := \bigcup_{j=1,2} \Big\{ \mathsf{M}_{\frac{1}{\alpha_j}} h_j > C \tfrac{\|h_j\|_{1/\alpha_j}}{|F_3|^{\alpha_j}} \Big\}, \tag{6.1}$$

$$\bar{E}^{\vec{\alpha}}_{h_1,h_2,F_3} = \bigcup \big\{ 3Q : Q \text{ max. dyad. int. } \subset E^{\vec{\alpha}}_{h_1,h_2,F_3} \big\},$$

$$F_3(\vec{\alpha}, h_1, h_2) := F_3 \backslash \bar{E}^{\vec{\alpha}}_{h_1,h_2,F_3}, \tag{6.2}$$

with C chosen large enough so that $|F_3| \le 4|F_3(\vec{\alpha}, h_1, h_2)|$. Assume that $I_s \subset E^{\vec{\alpha}}_{h_1,h_2,F_3}$ for all $s \in \mathbf{S}$. Then, for all functions $|f_1| \le |h_1|, |f_2| \le |h_2|$,

$$\Lambda_{\mathbf{S}}(f_1, f_2, f_3) \lesssim \|h_1\|_{\frac{1}{\alpha_1}} \|h_2\|_{\frac{1}{\alpha_2}} |F_3|^{\alpha_3} \qquad \forall\, |f_3| \le \mathbf{1}_{F_3(\vec{\alpha}, h_1, h_2)}.$$

Proof. We will rely on the following estimate, which is proved in the same way as, for instance, [1, Lemma 3.1]. For an interval J, $A > 1$, $|f_3| \le \mathbf{1}_{\mathbb{R} \backslash AJ}$ there holds

$$\Lambda_{\mathbf{S}(J)}(f_1, f_2, f_3) \lesssim A^{-100} |J| \Big(\inf_{x \in J} \mathsf{M}_{\frac{1}{\alpha_1}} f_1(x) \Big) \Big(\inf_{x \in J} \mathsf{M}_{\frac{1}{\alpha_2}} f_2(x) \Big),$$

where $\mathbf{S}(J) = \{s \in \mathbf{S} : I_s = J\}$. Now, for each interval J, let $k(J)$ be the minimal integer such that $2^{k+1} J \not\subset E_{h_1,h_2,F_3}$. Then one can take $A = 2^k$ in the above estimate, and, since $|f_j| \le |h_j|$,

$$\sum_{J:k(J)=k} \Lambda_{\mathbf{S}(J)}(f_1,f_2,f_3) \lesssim 2^{-100k} \sum_{J:k(J)=k} |J| \Big(\inf_{x\in J} \mathsf{M}_{\frac{1}{\alpha_1}} h_1(x)\Big)\Big(\inf_{x\in J} \mathsf{M}_{\frac{1}{\alpha_2}} h_2(x)\Big)$$
$$\lesssim 2^{-98k}\|h_1\|_{\frac{1}{\alpha_1}}\|h_2\|_{\frac{1}{\alpha_2}}|F_3|^{-(\alpha_1+\alpha_2)} \sum_{J:k(J)=k} |J| \lesssim 2^{-97k}\|h_1\|_{\frac{1}{\alpha_1}}\|h_2\|_{\frac{1}{\alpha_2}}|F_3|^{\alpha_3}$$

since the intervals $\{J : k(J) = k\}$ have at most 2^{k+1} overlap and are contained in $E^{\vec\alpha}_{h_1,h_2,F_3}$. The proof of the lemma is then finished by summing up over k. □

6.2. **The f_3-decomposition of S and forest estimates.** Throughout this subsection, fix a set $F_3 \subset \mathbb{R}$ and $|f_3| \le \mathbf{1}_{F_3}$. In view on the dyadic scaling invariance of the family of model sums (see Remark 4.2), we lose no generality by working with $|F_3| \sim 1$ in what follows. Note that any finite collection $\mathbf{S}$ admits the decomposition

$$\mathbf{S} = \bigcup_{k=0,1,\ldots} \mathbf{S}_k, \tag{6.3}$$

$$\mathrm{size}_3(f_3;\mathbf{S}_k) \lesssim 2^{-k}, \tag{6.4}$$

$$\mathbf{S}_k = \bigcup_{\mathbf{T}\in\mathcal{F}_k} \mathbf{T}, \quad \text{each } \mathbf{T} \text{ is a tree}, \qquad \sum_{\mathbf{T}\in\mathcal{F}} |I_\mathbf{T}| \lesssim 2^{2k}|F_3| \sim 2^{2k}, \tag{6.5}$$

$$\sup_{I\subset\mathbb{R}} \frac{1}{|I|} \sum_{\mathbf{T}\in\mathcal{F}_k : I_\mathbf{T}\subset I} |I_\mathbf{T}| \lesssim 2^{2k}, \tag{6.6}$$

which we call the f_3-decomposition of $\mathbf{S}$ into forests (unions of trees) $\mathbf{S}_k$. The decomposition is obtained by iteratively applying the size lemma with $f = f_3$, starting from $\sigma = \mathrm{size}_3(f_3;\mathbf{S}) \lesssim 1$. Since $\mathbf{S}$ is finite, the iteration terminates in finitely many steps.

The next two propositions provide estimates for the model sums restricted to $\mathbf{S}_k$ when either $f_2 \in L^2(\mathbb{R})$ or $|f_2| \le \mathbf{1}_{F_2}$ for some $F_2 \subset \mathbb{R}$ of finite measure. To unify notation, we write $h_2 = f_2$ if $f_2 \in L^2(\mathbb{R})$ is unrestricted and $h_2 = \mathbf{1}_{F_2}$ if f_2 is restricted to F_2.

Proposition 6.2. *Let $f_1 \in L^2(\mathbb{R})$ be given and f_2, F_3, f_3 as above. Assume that $\mathbf{S}_k$ satisfies* (6.4)-(6.5) *and that in addition*

$$\mathrm{size}_2(f_2;\mathbf{S}_k) \lesssim 2^{-n_0}\|h_2\|_2,$$

for some $n_0 \ge 0$. Then

$$|\Lambda_{\mathbf{S}_k}(f_1,f_2,f_3)| \lesssim 2^{-n_0}\min\Big\{1,(k-n_0)2^{-(k-n_0)}\Big\}\|f_1\|_2\|h_2\|_2.$$

Proof. By linearity of assumptions and conclusions in f_1 we can assume $\|f_1\|_2 = 1$. We split the proof into two cases, the first being when $k \le n_0$. In this case, we straightforwardly apply Lemma 5.3, whose assumptions are satisfied in view of (6.4)-(6.5), to the triple (f_1,f_2,f_3) with $A = |F_3|^{\frac12} \sim 1$. This yields

$$|\Lambda_{\mathbf{S}_k}(f_1,f_2,f_3)| \lesssim |F_3|^{\frac12}\mathrm{size}_2(f_2;\mathbf{S}_k) \lesssim 2^{-n_0}\|h_2\|_2, \tag{6.7}$$

which is what we had to prove. Let us deal with the case $k > n_0$. We begin by decomposing $\mathbf{S}$ into collections $\mathbf{S}_{k,n}, n = n_0,\ldots,k$, each being a disjoint union of trees $\mathbf{T} \in \mathcal{F}_{k,n}$ such that

$$\sum_{\mathbf{T}\in\mathcal{F}_{k,n}} |I_\mathbf{T}| \lesssim 2^{2n} \qquad \mathrm{size}_2(f_2;\mathbf{S}_{k,n}) \le 2^{-n}\|h_2\|_2,\ \mathrm{size}_3(f_3;\mathbf{S}_{k,n}) \le 2^{-k}, \tag{6.8}$$

by iteratively applying the size Lemma 5.2 to $f = f_2$ and by organizing the leftover collection $\mathbf{S}_{k,k}$ into a disjoint union of trees $\mathcal{F}_{k,k} := \{\mathbf{T} \cap \mathbf{S}_{k,k} : \mathbf{T} \in \mathcal{F}_k\}$. We are now allowed to apply Lemma 5.3 to the collections $\mathbf{S}_{k,n}$ with the roles of f_2, f_3 interchanged and with $A = \|h_2\|_2$, so that

$$|\Lambda_{\mathbf{S}_k}(f_1, f_2, f_3)| \le \sum_{n=n_0}^{k} |\Lambda_{\mathbf{S}_{k,n}}(f_1, f_2, f_3)| \lesssim (k-n_0)|\|h_2\|_2 \mathrm{size}_3(f_3; \mathbf{S}_{k,n}) \\ \lesssim (k-n_0) 2^{-k} \|h_2\|_2.$$

Putting together (6.7) with the last display, the proposition is proved. □

This proposition is a version of the previous one, differing in that the function f_1 is not locally L^2. The proof makes use of the multi-frequency decomposition of Proposition 3.2.

Proposition 6.3. *Let $f_1 \in L^{\frac{1}{\alpha_1}}(\mathbb{R})$, $1/2 < \alpha_1 \le 1$, be given and $f_2, F_3, f_3 \subset \mathbb{R}$ as above. Assume that $\mathbf{S}_k$ satisfies* (6.4)-(6.6) *and that in addition*

$$I_s \not\subset \left\{ \mathsf{M}_{\frac{1}{\alpha_1}} f_1(x) \gtrsim \|f_1\|_{\frac{1}{\alpha_1}} \right\} \qquad \forall s \in \mathbf{S}_k, \tag{6.9}$$

$$\mathrm{size}_2(f_2; \mathbf{S}_k) \lesssim 2^{-n_0} \|h_2\|_2, \tag{6.10}$$

for some $n_0 \ge 0$. Then,

$$|\Lambda_{\mathbf{S}_k}(f_1, f_2, f_3)| \lesssim \|f_1\|_{\frac{1}{\alpha_1}} \|h_2\|_2 \big(u(Ck)^2 \log u(Ck)\big)^{\alpha_1 - \frac{1}{2}} \\ \times \begin{cases} 2^{-n_0} 2^{(2\alpha_1 - 1)k}, & k \le n_0, \\ \frac{1}{2\alpha_1 - 1} 2^{-2(1-\alpha_1)k}, & k > n_0. \end{cases}$$

Proof. By linearity, we can assume $\|f_1\|_{1/\alpha_1} = 1$. We claim that $\mathbf{S}_k = \mathbf{S}'_k \cup \mathbf{S}''_k$, respectively written as a disjoint union of trees from the collections $\mathcal{F}'_k, \mathcal{F}''_k$ satisfying

$$\sum_{\mathbf{T} \in \mathcal{F}''_k} |I_{\mathbf{T}}| \lesssim 2^{-10k}, \tag{6.11}$$

$$\Big\| \sum_{\mathbf{T} \in \mathcal{F}'_k} \mathbf{1}_{3u(Ck) I_{\mathbf{T}}} \Big\|_\infty \lesssim 2^{5k}. \tag{6.12}$$

The proof of the claim is standard but technical, and we postpone it to the end of the section. Accordingly, we split $\Lambda_{\mathbf{S}_k} = \Lambda_{\mathbf{S}'_k} + \Lambda_{\mathbf{S}''_k}$ and estimate each summand separately.

The summand involving $\mathbf{S}''_k$ is an error term. Relying on the tree estimate of Lemma 5.1, we estimate

$$\prod_{j=1}^{3} \mathrm{size}_j(f_j; \mathbf{S}''_k) \sum_{\mathbf{T} \in \mathcal{F}''_k} |I_{\mathbf{T}}| \lesssim 2^{-n_0} 2^{-10k} \|h_2\|_2. \tag{6.13}$$

We have relied on (6.9) and inequality (5.3) to obtain that $\mathrm{size}_1(f_1; \mathbf{S}) \lesssim 1$, on assumption (6.10), and later employed (6.11).

We turn to the $\Lambda_{\mathbf{S}'_k}$ summand and first deal with the case $k > n_0$, which is the harder one. The first step consists again of decomposing $\mathbf{S}'_k$ into collections

$\mathbf{S}'_{k,n}, n = n_0, \ldots, k$, each being a disjoint union of trees $\mathbf{T} \in \mathcal{F}'_{k,n}$ such that

$$\sum_{\mathbf{T}\in\mathcal{F}'_{k,n}} |I_\mathbf{T}| \lesssim 2^{2n}, \qquad \text{size}_2(f_2;\mathbf{S}'_{k,n}) \le 2^{-n}\|h_2\|_2,\ \text{size}_3(f_3;\mathbf{S}'_{k,n}) \le 2^{-k}, \tag{6.14}$$

and in addition

$$\Big\| \sum_{\mathbf{T}\in\mathcal{F}'_{k,n}} \mathbf{1}_{3u(Ck)I_\mathbf{T}} \Big\|_\infty \lesssim 2^{5k}. \tag{6.15}$$

The three properties of (6.14) are obtained by the same argument used in Proposition 6.2 for (6.8), while (6.15) is carried over from (6.12) for $\mathcal{F}'_k$ by means of a reshuffling argument. Details are given at the end of the section. The next step is the definition of a set of top data $\mathcal{T}$ which is suitable for Proposition 3.2. Noting that for each 1-tree (resp. j-tree, $j \neq 1$) $\mathbf{T} \in \mathcal{F}'_{k,n}$, $\{\varphi_{s_1} : s \in \mathbf{T}\}$ is a collection of u-adapted wave packets (resp. u-adapted wave packets with mean zero) with respect to the top datum $(I, \xi_\mathbf{T})$ (resp. $(I, \xi_{\mathbf{T},1})$), according to the terminology of Section 3, we are led to define

$$\mathcal{T} := \{(I_\mathbf{T}, \xi_\mathbf{T}) : \mathbf{T} \in \mathcal{F}'_{k,n} \text{ is a 1-tree}\} \cup \{(I_\mathbf{T}, \xi_{\mathbf{T},1}) : \mathbf{T} \in \mathcal{F}'_{k,n} \text{ is a } j\text{-tree}, j \neq 1\}.$$

With this definition, by virtue of (6.15), we may appeal to Proposition 3.2 with $p = 1/\alpha_1$, $f = f_1$, $\lambda \sim 1 = \|f_1\|_{1/\alpha_1}$ and k replaced by $5k$. Writing $A_k = u(Ck)^2 \log u(Ck)$ for brevity, we obtain the decomposition $f_1 = g_n + b_n$, with

$$\|g_n\|_2 \lesssim \Big(A_k \sum_{\mathbf{T}\in\mathcal{F}'_{k,n}} |I_\mathbf{T}|\Big)^{\alpha_1 - \frac12} \lesssim 2^{(2\alpha_1-1)n}\big(A_k\big)^{\alpha_1-\frac12}, \tag{6.16}$$

$$\sup_{\mathbf{T}\in\mathcal{F}'_{k,n}\text{1-tree}} \sup_{s\in\mathbf{T}} \frac{|\langle b_n, \varphi_{s_1}\rangle|}{|I_s|^{\frac12}} \lesssim 2^{-16k}, \qquad \sup_{\mathbf{T}\in\mathcal{F}'_{k,n}\text{j-tree}} \text{size}(b_n;\mathbf{T}^{(1)}) \lesssim 2^{-16k}. \tag{6.17}$$

We used conclusion (3.3) of Proposition 3.2 for the first bound of (6.16) and (6.14) for the second step, while the inequalities of (6.17) follow respectively from conclusions (3.4) and (3.5). Repeating the proof of the tree Lemma 5.1 and using (6.17) yield the estimate

$$\begin{aligned} |\Lambda_{\mathbf{S}'_{k,n}}(b_n, f_2, f_3)| &\le \sum_{\mathbf{T}\in\mathcal{F}'_{k,n}} \sum_{s\in\mathbf{T}} |I_s|^{-\frac12} |\langle b, \varphi_{s_1}\rangle| \prod_{j=2}^{3} |\langle f_j, \varphi_{s_j}\rangle| \\ &\lesssim 2^{-16k} \prod_{j=2,3} \text{size}_j(f_j;\mathbf{S}_{k,n}) \Big(\sum_{\mathbf{T}\in\mathcal{F}'_{k,n}} |I_T|\Big) \lesssim 2^{-n_0} 2^{-14k}\|h_2\|_2. \end{aligned} \tag{6.18}$$

Now, in view of (6.14), we can apply Lemma 5.3 to $\mathbf{S}'_{k,n}$, with tuple (g_n, f_3, f_2) and $A = \|h_2\|_2$. Note that the roles of f_3 and f_2 are interchanged. This leads to

$$|\Lambda_{\mathbf{S}'_{k,n}}(g_n, f_2, f_3)| \lesssim \|g_n\|_2 \|h_2\|_2 \text{size}_3(f_3;\mathbf{S}_{k,n}) \lesssim 2^{(2\alpha_1-1)n}\big(A_k\big)^{\alpha_1-\frac12}\|h_2\|2^{-k}.$$

Note that the last right hand side of (6.18) is always smaller than the second member of the above display. Therefore, we estimate

$$|\Lambda_{\mathbf{S}'_k}(f_1,f_2,f_3)| \le \sum_{n=n_0}^{k} \big(|\Lambda_{\mathbf{S}'_k}(b_n,f_2,f_3)| + |\Lambda_{\mathbf{S}'_k}(g_n,f_2,f_3)|\big)$$

$$\lesssim \|h_2\|_2 \big(A_k\big)^{\alpha_1-\frac12} 2^{-k} \sum_{n=n_0}^{k} 2^{(2\alpha_1-1)n} \lesssim \|h_2\|_2 \big(A_k\big)^{\alpha_1-\frac12} \tfrac{1}{2\alpha_1-1} 2^{-2k(1-\alpha_1)}.$$

Collecting (6.13) and the last display, we have proved the required estimate when $k > n_0$.

In the case $k < n_0$, there is no need for the additional decomposition of $\mathbf{S}'_k$. We appeal directly to Proposition 3.2 along the same lines as above, this time using the trees of $\mathcal{F}'_k$ as our top data, and obtain a decomposition $f_1 = g + b$, with

$$\|g\|_2 \lesssim \Big(A_k \sum_{\mathbf{T}\in\mathcal{F}'_{k,}} |I_{\mathbf{T}}|\Big)^{\alpha_1-\frac12} \lesssim 2^{(2\alpha_1-1)k}\big(A_k\big)^{\alpha_1-\frac12},$$

$$\sup_{\mathbf{T}\in\mathcal{F}'_k \text{1-tree}} \sup_{s\in\mathbf{T}} \frac{|\langle b,\varphi_{s_1}\rangle|}{|I_s|^{\frac12}} \lesssim 2^{-16k}, \qquad \sup_{\mathbf{T}\in\mathcal{F}'_k \text{j-tree}} \mathrm{size}(b;\mathbf{T}^{(1)}) \lesssim 2^{-16k}.$$

We then apply Lemma 5.3 to $\mathbf{S}'_k$ with tuple (g,f_2,f_3) and $A = |F_3|^{\frac12} \sim 1$, yielding

$$|\Lambda_{\mathbf{S}'_k}(g,f_2,f_3)| \lesssim \|g\|_2 \mathrm{size}_2(f_2;\mathbf{S}'_k) \lesssim 2^{-n_0} 2^{(2\alpha_1-1)k}\big(A_k\big)^{\alpha_1-\frac12}\|h_2\|_2.$$

Arguing exactly as in the previous case, one sees that the $\Lambda_{\mathbf{S}'_k}(b,f_2,f_3)$ summand and the error term (6.13) are again smaller than the right hand side of the above estimate. This completes the proof of the proposition. □

Proof of the decomposition (6.11)-(6.12). We begin with some notation and preliminaries. We write $\mu := 3u(Ck)$ and $J_{\mathbf{T}} := \mu I_{\mathbf{T}}$ for $\mathbf{T}\in\mathcal{F}_k$. Note that $\mu \lesssim 2^k$. Furthermore, for any $\mathcal{G}\subset\mathcal{F}_k$ and for each interval $J\subset\mathbb{R}$, we denote

$$N_{\mathcal{G}}(x) = \sum_{\mathbf{T}\in\mathcal{G}} \mathbf{1}_{J_{\mathbf{T}}}(x), \qquad N_{\mathcal{G}',J}(x) = \sum_{\mathbf{T}\in\mathcal{G}:J\subsetneq J_{\mathbf{T}}} \mathbf{1}_{J_{\mathbf{T}}}(x).$$

We claim that (6.11)-(6.12) will follow once we show that any $\mathcal{G}\subset\mathcal{F}_k$ such that the dilated intervals $\{J = J_{\mathbf{T}} : \mathbf{T}\in\mathcal{G}\}$ belong to a fixed grid $\mathcal{D}$ admits the decomposition

$$\mathcal{G} = \mathcal{G}'\cup\mathcal{G}'', \qquad \|N_{\mathcal{G}'}\|_\infty \lesssim 2^{4k}, \qquad \|N_{\mathcal{G}''}\|_1 \lesssim 2^{-100k}. \tag{6.19}$$

The claim simply follows by decomposing $\mathcal{F}_k$ into $\lesssim \mu\log\mu \lesssim 2^k$ such $\mathcal{G}$'s, which is possible since the intervals $\{I = I_{\mathbf{T}} : \mathbf{T}\in\mathcal{F}_k\}$ belong to a finite union of dyadic grids.

We begin the proof of (6.19), fixing one such $\mathcal{G}$. Let $J\in\boldsymbol{J}^0$ be the collection of maximal intervals of $\{J = J_{\mathbf{T}} : \mathbf{T}\in\mathcal{G}\}\subset\mathcal{D}$. We inherit from (6.5) the inequality

$$\sum_{J\in\boldsymbol{J}_0} |J| \le \mu \sum_{\mathbf{T}\in\mathcal{F}_k} |I_{\mathbf{T}}| \lesssim \mu 2^{2k} \lesssim 2^{3k}. \tag{6.20}$$

Moreover, a consequence of (6.6) is that

$$\sum_{\mathbf{T}\in\mathcal{G}:J_{\mathbf{T}}\subset J} |J_{\mathbf{T}}| \lesssim \mu 2^{2k}|J| \lesssim 2^{3k}|J|, \qquad \forall J\subset\mathbb{R}.$$

Observing that for each $J \in \mathcal{D}$ $N_{\mathcal{G},J}(x) = N_{\mathcal{G},J}$ is constant on J, the last display implies that

$$\big|\{x \in J : N_{\mathcal{G}}(x) - N_{\mathcal{G},J} > C\lambda 2^{3k}\}\big| \leq 2^{-\lambda}|J|, \qquad \forall \lambda > 0, \tag{6.21}$$

if $J \in \mathcal{D}$ and the constant C is chosen large enough; this is John-Nirenberg's inequality.

We now construct $\mathcal{G}', \mathcal{G}''$. The set $\mathbb{J}^1 = \{N_{\mathcal{G}} > Ck2^{3k}\}$ is the union of its maximal intervals $J \in \mathcal{D}$. We call $\boldsymbol{J}^1$ the collection of such intervals. Setting $\mathcal{G}' := \{\mathbf{T} \in \mathcal{G} : J_{\mathbf{T}} \not\subset \mathbb{J}^1\}$, $\mathcal{G}'' := \mathcal{G} \backslash \mathcal{G}'$, it is easy to see that

$$\|N_{\mathcal{G}'}\|_\infty \lesssim k2^{3k} \lesssim 2^{4k}, \qquad \operatorname{supp} N_{\mathcal{G}''} \subset \mathbb{J}^1. \tag{6.22}$$

Furthermore, using (6.21) in the second step and (6.20) for the final inequality, we have the estimate

$$|\mathbb{J}^1| = \sum_{J \in \boldsymbol{J}^0} |\{x \in J : N_{\mathcal{G}}(x) > Ck2^{3k}\}| \leq \sum_{J \in \boldsymbol{J}^0} 2^{-400k}|J| \lesssim 2^{-200k}. \tag{6.23}$$

We will show that $\mathcal{G}''$ satisfies (6.19) by means of an iterative procedure. Assume that, at the j-th step, we have written $\mathcal{G}'' = \mathcal{G}''_{\text{now}} \cup \mathcal{G}_{\text{stock}}$, where $\|N_{\mathcal{G}''_{\text{now}}}\|_1 \lesssim 2^{-100k}$, and $N_{\mathcal{G}_{\text{stock}}}$ is supported on the set $\mathbb{J}^j$, which is a union of disjoint intervals $J \in \boldsymbol{J}^j \subset \mathcal{D}$ such that $|\mathbb{J}^j| \lesssim 2^{-200kj}$. In (6.23), we have the base case $j = 1$, with $\mathcal{G}''_{\text{now}} = \emptyset$, $\mathcal{G}_{\text{stock}} = \mathcal{G}''$. The $(j+1)$-th inductive step is as follows. We define $\mathbb{J}^{j+1} := \{N_{\mathcal{G}_{\text{stock}}} > Ck2^{3k}\}$, which is a union of maximal intervals $J' \in \boldsymbol{J}^{j+1} \subset \mathcal{D}$. Setting $\mathcal{G}_* := \{\mathbf{T} \in \mathcal{G}_{\text{stock}} : J_{\mathbf{T}} \not\subset \mathbb{J}^{j+1}\}$, we observe that

$$\|N_{\mathcal{G}_*}\|_\infty \lesssim k2^{3k} \lesssim 2^{4k}$$

so that

$$\|N_{\mathcal{G}_*}\|_1 \leq \|N_{\mathcal{G}_*}\|_\infty |\operatorname{supp} N_{\mathcal{G}_*}| \lesssim 2^{4k}|\mathbb{J}^j| \lesssim 2^{-100kj},$$

by the inductive assumption on $\mathbb{J}^j$. Also, relying on (6.21) to pass to the second line,

$$\begin{aligned} |\mathbb{J}^{j+1}| &\leq \sum_{J \in \mathbb{J}^j} \big|\{x \in J : N_{\mathcal{G}_{\text{stock}}}(x) > Ck2^{3k}\}\big| \\ &\leq \sum_{J \in \mathbb{J}^j} \big|\{x \in J : N_{\mathcal{G}}(x) - N_{\mathcal{G},J} > Ck2^{3k}\}\big| \lesssim 2^{-400k} \sum_{J \in \mathbb{J}^j} |J| \lesssim 2^{-200k(j+1)}. \end{aligned}$$

The inductive step is completed by updating $\mathcal{G}''_{\text{now}} := \mathcal{G}''_{\text{now}} \cup \mathcal{G}_*, \mathcal{G}_{\text{stock}} := \mathcal{G}_{\text{stock}} \backslash \mathcal{G}_*$. We iterate until $\mathcal{G}_{\text{stock}}$ is empty, which happens after finitely many steps, since $\mathcal{G}$ is a finite collection. At this point, $\mathcal{G}'' = \mathcal{G}''_{\text{now}}$ satisfies (6.19). This completes the proof of the claim. □

Details of the construction (6.14)-(6.15). The same argument employed in Proposition 6.2 for (6.8) yields the decomposition of $\mathbf{S}'_k$ into subcollections $\mathbf{S}'_{k,n}$, $n = n_0, \ldots, k-1$, each partitioned into a union of trees $\mathbf{T} \in \mathcal{G}_{k,n}$ satisfying (6.14) with $\mathcal{G}_{k,n}$ in place of $\mathcal{F}'_{k,n}$. The remaining collection $\mathbf{S}'_{k,k} := \mathbf{S}'_k \backslash (\mathbf{S}'_{k,n_0} \cup \cdots \cup \mathbf{S}'_{k,k-1})$, which has $\operatorname{size}_2(f_2; \mathbf{S}'_{k,k}) \lesssim 2^{-k}\|h_2\|_2 |F_3|^{-\frac{1}{2}}$, is partitioned into trees by $\mathcal{F}'_{k,k} = \{\mathbf{T}' := \mathbf{T} \cap \mathbf{S}'_{k,k} : \mathbf{T} \in \mathcal{F}'_k\}$, and the remaining claims of (6.14)-(6.15) are immediately inherited from (6.4), (6.5), and (6.12).

We now show how to construct a new partition $\mathcal{F}'_{k,n}$ of $\mathbf{S}'_{k,n}$ inheriting (6.15) from $\mathcal{F}'_k$ as well as retaining (6.14). By partitioning $\mathbf{S}'_{k,n}$, $\mathcal{G}_{k,n}$ into three subcollections, we can reduce to the case where all trees of $\mathcal{G}_{k,n}$ are 1-trees. Let **tops** be the collection of maximal tritiles in $\mathbf{S}'_{k,n}$ with respect to the following order relation: $s \ll_1 s'$ when $I_s \subset I_{s'}$ and $2\omega_{s'_1} \subset 2\omega_{s_1}$. Note that the boxes $\{I_s \times 2\omega_{s_1} : s \in \mathbf{tops}\}$ are pairwise disjoint. For each $s \in \mathbf{tops}$, form the tree $\mathbf{T}(s) = \{s\}$ with top data $(I_{\mathbf{T}}, \xi_T) = (I_s, c(\omega_{s_1}))$. Now, each $s' \in \mathbf{S}'_{k,n}$ is added to $\mathbf{T}(\bar{s})$ where $\bar{s}$ is the tritile with minimal $c(\omega_{s_1})$ among those $s \in \mathbf{tops}$ with $s' \ll_1 s$. We call $\mathcal{F}'_{k,n} := \{\mathbf{T} = \mathbf{T}(s) : s \in \mathbf{tops}\}$ the resulting partition of $\mathbf{S}'_{k,n}$.

To prove (6.14) for $\mathcal{F}'_{k,n}$, recall that each tritile $s \in \mathbf{tops}$ belonged to a unique tree $\tilde{\mathbf{T}}(s) \in \mathcal{G}_{k,n}$. Observing that $\{\omega_{s_1} : s \in \mathbf{tops}, \tilde{\mathbf{T}}(s) = \tilde{\mathbf{T}}\}$ have non-empty intersection, the intervals $\{I_s : s \in \mathbf{tops}, \tilde{\mathbf{T}}(s) = \tilde{\mathbf{T}}\}$, all contained in $I_{\tilde{\mathbf{T}}}$, must be pairwise disjoint. Hence,

$$\sum_{\mathbf{T} \in \mathcal{F}'_{k,n}} |I_{\mathbf{T}}| = \sum_{s \in \mathbf{tops}} |I_s| = \sum_{\tilde{\mathbf{T}} \in \mathcal{G}_{k,n}} \sum_{s \in \mathbf{tops}: \tilde{\mathbf{T}} = \tilde{\mathbf{T}}(s)} |I_s| \le \sum_{\tilde{\mathbf{T}} \in \mathcal{G}_{k,n}} |I_{\tilde{\mathbf{T}}}| \lesssim 2^{2n},$$

and we have verified (6.14) for $\mathcal{F}'_{k,n}$. The argument for (6.15) is similar, with trees from the forest $\mathcal{F}'_k$ playing the role of the $\tilde{\mathbf{T}}$'s above. This concludes our decomposition. □

7. Proofs of the main results

7.1. **Proofs of Theorems 1 to 3.** We will obtain our restricted type estimates on $\Lambda_{\vec{\beta}}$ via the reduction to the model sums (4.3), in particular, relying on Remark 4.1. At this time, we make our choice of generating function u and, consequently, of our mother function v in the definition (4.1) of v_{s_j}, taking $u := u_1$ from the family (2.5). Any other choice of the parameter $\lambda > 0$ in (2.5) is legal throughout our arguments. We invite the willing reader to check that alternative choices of λ (or of u altogether) do not bring essential improvements to the estimate of Theorems 2 and 3 and bring no improvements at all to Corollary 4.

Therefore, Theorems 1 to 3 will respectively follow from the corresponding versions for the model sums $\Lambda_{\mathbf{S}}$ below. We stress that the implicit constants appearing in the statements are uniform over all finite well-discretized collections of tritiles $\mathbf{S}$, and the major set F'_3 is explicitly chosen independently of $\mathbf{S}$.

Theorem 1′. *Let $\vec{\alpha} = (\frac{1}{2}, 1, -\frac{1}{2})$. For[2] $f_1 \in L^2(\mathbb{R})$, $|f_2| \le \mathbf{1}_{F_2}$, and $F_3 \subset \mathbb{R}$ of finite measure, let F'_3 be the major subset of F_3 defined via (6.2) by $F'_3 := F_3(\vec{\alpha}, f_1, \mathbf{1}_{F_2})$. Then, for all $|f_3| \le \mathbf{1}_{F'_3}$,*

$$|\Lambda_{\mathbf{S}}(f_1, f_2, f_3)| \lesssim \|f_1\|_2 |F_2| |F_3|^{-\frac{1}{2}} \log\left(\mathrm{e} + \tfrac{|F_3|}{|F_2|}\right).$$

Theorem 2′. *Let $\vec{\alpha} = (\alpha_1, \alpha_2, -\frac{1}{2}) \in \mathcal{S}_3$. For $f_1 \in L^{\frac{1}{\alpha_1}}(\mathbb{R})$, $|f_2| \le \mathbf{1}_{F_2}$, and $F_3 \subset \mathbb{R}$ of finite measure, let F'_3 be the major subset of F_3 defined via (6.2) by $F'_3 := F_3(\vec{\alpha}, f_1, \mathbf{1}_{F_2})$. Then for all $|f_3| \le \mathbf{1}_{F'_3}$, we have the estimate*

$$|\Lambda_{\mathbf{S}}(f_1, f_2, f_3)| \lesssim \tfrac{1}{(1-\alpha_1)(1-\alpha_2)} \|f_1\|_{\frac{1}{\alpha_1}} |F_2|^{\alpha_2} |F_3|^{-\frac{1}{2}} \times \left(\max\left\{(1-\alpha_1)^{-1}, \log\left(\tfrac{|F_3|}{|F_2|}\right)\right\}\right)_*^{2(1-\alpha_2)}.$$

[2]Note that to unify notation in the proofs below, we have switched herein the roles of the first and second arguments with respect to the statement of Theorem 1.

Theorem 3′. *Let $0 \le \alpha_1 < 1$, $\vec{\alpha} = (\alpha_1, \frac{1}{2}, \frac{1}{2} - \alpha_1)$. For $f_1 \in L^{\frac{1}{\alpha_1}}(\mathbb{R})$, $f_2 \in L^2(\mathbb{R})$, and $F_3 \subset \mathbb{R}$ of finite measure, let F_3' be the major subset of F_3 defined via (6.2) by $F_3' := F_3(\vec{\alpha}, f_1, f_2)$. Then, for all $|f_3| \le \mathbf{1}_{F_3'}$,*

$$|\Lambda_{\mathbf{S}}(f_1, f_2, f_3)| \lesssim \tfrac{1}{1-\alpha_1}\Big(\tfrac{1}{1-\alpha_1}\Big)_*^{2\alpha_1 - 1} \|f_1\|_{\frac{1}{\alpha_1}} \|f_2\|_2 |F_3|^{\frac{1}{2}-\alpha_1}.$$

Remark 7.1. By the dyadic Hölder scaling invariance of the family $\Lambda_{\mathbf{S}}$ pointed out in Remark 4.2, we may assume that $|F_3| \sim 1$ in our proofs. Also, linearity in f_1 of assumptions and conclusions for Theorem 1′, 2′ and in both f_1, f_2 for Theorem 3′ allows us to work, in these cases, with f_1, f_2 of unit norm in the respective spaces. We will work in the range $\alpha_1 > 3/4$ (say) in our proof of Theorem 3′, since the bounds in the complementary region are well-known (and uniform in α_1) from (1.2). Noting that the estimate of Theorem 3′ is stronger than the one of Theorem 2′ when $|F_2| \ge |F_3|$, we may conveniently restrict to $|F_2| \le |F_3| \sim 1$ when proving Theorem 2′. Finally, to unify notation, we write $h_2 = f_2$ if f_2 is unrestricted and $h_2 = \mathbf{1}_{F_2}$ if f_2 is restricted to F_2.

The first two steps of the proof are shared among the three theorems. Recalling from (6.1) the definition

$$E^{\vec{\alpha}}_{f_1,h_2,F_3} = \Big\{ \mathsf{M}_{\frac{1}{\alpha_1}} f_1 \gtrsim \frac{1}{|F_3|^{\alpha_1}} \Big\} \cup \Big\{ \mathsf{M}_{\frac{1}{\alpha_2}} h_2 \gtrsim \frac{\|h_2\|_{1/\alpha_2}}{|F_3|^{\alpha_2}} \Big\},$$

we decompose

$$\mathbf{S} = \mathbf{S}^{\text{bad}} \cup \mathbf{S}^1, \qquad \mathbf{S}^{\text{bad}} = \big\{ s \in \mathbf{S} : I_s \subset E^{\alpha}_{f_1,h_2,F_3} \big\}, \quad \mathbf{S}^1 = \mathbf{S} \backslash \mathbf{S}^{\text{bad}}. \tag{7.1}$$

Clearly $|\Lambda_{\mathbf{S}}| \le |\Lambda_{\mathbf{S}^{\text{bad}}}| + |\Lambda_{\mathbf{S}^1}|$. We handle the $\Lambda_{\mathbf{S}^{\text{bad}}}$ term by a straightforward application of Proposition 6.1, which gives

$$|\Lambda_{\mathbf{S}^{\text{bad}}}(f_1, f_2, f_3)| \lesssim \|f_1\|_{\frac{1}{\alpha_1}} \|h_2\|_{\frac{1}{\alpha_2}} |F_3|^{\alpha_3}, \qquad \forall |f_3| \le \mathbf{1}_{F_3'}. \tag{7.2}$$

Note that (7.2) complies with the required estimate for $\Lambda_{\mathbf{S}}$ in all three cases.

We now fix $|f_3| \le \mathbf{1}_{F_3}$ and perform the f_3-decomposition of $\mathbf{S}^1$ of Subsection 6.2 into collections $\mathbf{S}_k$ complying with (6.4) to (6.6), and in addition inheriting from $\mathbf{S}^1$ the property

$$I_s \not\subset E^{\vec{\alpha}}_{f_1,h_2,F_3} \qquad \forall s \in \mathbf{S}_k. \tag{7.3}$$

The remaining part of the proof, consisting of the estimation of the right hand side of

$$|\Lambda_{\mathbf{S}^1}(f_1, f_2, f_3)| \le \sum_{k \ge 0} |\Lambda_{\mathbf{S}_k}(f_1, f_2, f_3)|$$

is specific to each theorem.

Conclusion of the proof of Theorem 1′. Recall that f_2 is restricted, thus $h_2 = \mathbf{1}_{F_2}$, and that we are assuming $\|f_1\|_2 = 1$, $|F_3| \sim 1$. A consequence of (5.1) and (7.3) is that

$$\text{size}_2(f_2; \mathbf{S}_k) \lesssim \sup_{s \in \mathbf{S}^1} \inf_{x \in I_s} \mathsf{M}_1 f_2(x) \le \sup_{s \in \mathbf{S}^1} \inf_{x \in I_s} \mathsf{M}_1 h_2(x) \lesssim \min\big\{1, |F_2|\big\} \lesssim 2^{-n_0} |F_2|^{\frac{1}{2}},$$

where we have set $n_0 = \frac{1}{2}\big|\log|F_2|\big|$. The first bound after the second almost inequality sign is actually due to $|h_2| \le 1$. At this point, we apply Proposition 6.2

to each $\mathbf{S}_k$ and bound

$$\sum_{k\geq 0} |\Lambda_{\mathbf{S}_k}(f_1,f_2,f_3)| \lesssim 2^{-n_0}\|f_1\|_2\|h_2\|_2 \sum_{k\geq 0} \min\{1,(k-n_0)2^{-(k-n_0)}\}$$
$$\lesssim \min\Big\{1, \log\big(\tfrac{1}{|F_2|}\big)\Big\}|F_2|,$$

which, combined with (7.2), finishes the proof of Theorem 1′. □

Conclusion of the proof of Theorem 3′. For this theorem, f_2 is unrestricted, thus $h_2 = f_2$, and we are assuming $\|f_1\|_{\frac{1}{\alpha_1}} = \|f_2\|_2 = 1$. Again, from (7.3) and Lemma 5.1, we learn that

$$\mathrm{size}_2(f_2;\mathbf{S}_k) \leq \mathrm{size}_2(f_2;\mathbf{S}^1) \lesssim \sup_{s\in\mathbf{S}^1}\inf_{x\in I_s} \mathsf{M}_2 f_2(x) \lesssim |F_3|^{-\frac{1}{2}} \sim 1.$$

Also in view of (7.3), the assumption (6.9) of Proposition 6.3 is satisfied. Applying the proposition to each $\mathbf{S}_k$, with $n_0 = 0$, observing that $2\alpha_1 - 1$ is bounded away from zero in our range $\alpha_1 > 3/4$, and recalling $u(t) \lesssim t(\log(\mathrm{e}+t))^2$ and the notation $t_* = t(\log(\mathrm{e}+t))^3$, we find that

$$\sum_{k\geq 0} |\Lambda_{\mathbf{S}_k}(f_1,f_2,f_3)| \lesssim \|f_1\|_{\frac{1}{\alpha_1}}\|f_2\|_2 \sum_{k\geq 0} \big(u(Ck)^2\log u(Ck)\big)^{\alpha_1-\frac{1}{2}} 2^{-2(1-\alpha_1)k}$$
$$\lesssim \sum_{k\geq 0} \big(k\big)_*^{2\alpha_1-1} 2^{-2(1-\alpha_1)k} \lesssim \tfrac{1}{1-\alpha_1}\Big(\tfrac{1}{1-\alpha_1}\Big)_*^{2\alpha_1-1}.$$

The proof of Theorem 3′ is finished by combining the last display with (7.2). □

Conclusion of the proof of Theorem 2′. Here f_2 is restricted, thus $h_2 = \mathbf{1}_{F_2}$, and we are assuming $\|f_1\|_{\frac{1}{\alpha_1}} = 1$. Also, we only need to treat the case $|F_2| \leq |F_3| \sim 1$. As in the previous proofs, we take advantage of (7.3) and of Lemma 5.1 to obtain the inequality

$$\mathrm{size}_2(f_2;\mathbf{S}_k) \lesssim \sup_{s\in\mathbf{S}^1}\inf_{x\in I_s} \mathsf{M}_1 f_2(x) \leq \sup_{s\in\mathbf{S}^1}\inf_{x\in I_s} \mathsf{M}_1 h_2(x)$$
$$= \Big(\sup_{s\in\mathbf{S}^1}\inf_{x\in I_s} \mathsf{M}_{\frac{1}{\alpha_2}}\mathbf{1}_{F_2}(x)\Big)^{\frac{1}{\alpha_2}} \lesssim 2^{-n_0}|F_2|^{\frac{1}{2}},$$

where we have set $n_0 = -\frac{1}{2}\log|F_2| \geq 0$. We make use of (7.3) to verify the remaining assumption (6.9) of Proposition 6.3 and apply the proposition to each $\mathbf{S}_k$, estimating

$$\frac{\sum_{k\geq 0}|\Lambda_{\mathbf{S}_k}(f_1,f_2,f_3)|}{|F_2|^{\alpha_2}}$$
$$\lesssim \sum_{k=0}^{n_0} \big(k\big)_*^{2\alpha_1-1} 2^{(2\alpha_1-1)(k-n_0)} + \tfrac{1}{2\alpha_1-1}\sum_{k>n_0} \big(k\big)_*^{2\alpha_1-1} 2^{-2(1-\alpha_1)(k-n_0)}$$
$$\lesssim \tfrac{1}{1-\alpha_2}\big(n_0\big)_*^{2\alpha_1-1} + \tfrac{1}{(1-\alpha_1)(1-\alpha_2)}\Big(\max\Big\{\tfrac{1}{1-\alpha_1}, n_0\Big\}\Big)_*^{2\alpha_1-1}.$$

The bound of the last display, together with (7.2), yields Theorem 2′. □

7.2. **Proof of Corollary 4.** Using symmetry, we can work with tuples $\vec\alpha \in \mathcal{S}_3$ and treat the case $\alpha_2 \geq \alpha_1$. For tuples $\vec\alpha$ as such, specializing (1.5) to $|f_1| \leq \mathbf{1}_{F_1}$ yields the GRWT estimate

$$\big|\Lambda_{\vec\beta}(f_1,f_2,f_3)\big| \lesssim (1-\alpha_2)^{-1}|F_1|^{\alpha_1}|F_2|^{\alpha_2}|F_3|^{-\frac12}\Big(\log\big(\tfrac{|F_3|}{|F_2|}\big)\Big)_*^{2(1-\alpha_2)} \qquad \forall\vec\alpha\in\mathcal{S}_3. \tag{7.4}$$

Fix an $\vec\alpha$ as above and a triple of sets F_1, F_2, F_3, and let $0 < \varepsilon \leq 2(1-\alpha_2)$ be chosen later. Let $\vec a = (a_1, a_2, -\frac12) \in \mathcal{S}_3$ be the tuple with $a_2 = 1 - \varepsilon/2$: given f_1, f_2 restricted respectively to F_1, F_2, we may apply (7.4) with tuple $\vec a$ to bound

$$\big|\Lambda_{\vec\beta}(f_1,f_2,f_3)\big| \lesssim \varepsilon^{-1}|F_1|^{a_1}|F_2|^{a_2}|F_3|^{-\frac12}\Big(\log\big(\tfrac{|F_3|}{|F_2|}\big)\Big)_*^{\varepsilon} \tag{7.5}$$

for all functions f_3 restricted to a major subset $F_3' \subset F_3$. Switching the order of F_1, F_2 and replacing F_3 with F_3', we apply (7.4), again with tuple $\vec a$, to $\Lambda_{\sigma_{12}(\vec\beta)}$ instead, yielding

$$\begin{aligned}\big|\Lambda_{\vec\beta}(f_1,f_2,f_3)\big| = \big|\Lambda_{\sigma_{12}(\vec\beta)}(f_2,f_1,f_3)\big| &\lesssim \varepsilon^{-1}|F_2|^{a_1}|F_1|^{a_2}|F_3'|^{-\frac12}\Big(\log\big(\tfrac{|F_3'|}{|F_1|}\big)\Big)_*^{\varepsilon}\\ &\lesssim \varepsilon^{-1}|F_1|^{a_2}|F_2|^{a_1}|F_3|^{-\frac12}\Big(\log\big(\tfrac{|F_3|}{|F_1|}\big)\Big)_*^{\varepsilon}\end{aligned} \tag{7.6}$$

for all functions f_3 restricted to a major subset $F_3'' \subset F_3'$, which (with different constant) is also a major subset of F_3. Taking the ϑ-geometric mean of (7.5) and (7.6), for $\frac12 \leq \vartheta \leq 1$ such that $\alpha_1 = \vartheta a_1 + (1-\vartheta)a_2, \alpha_2 = \vartheta a_2 + (1-\vartheta)a_1$, we obtain that for all $|f_3| \leq \mathbf{1}_{F_3''}$,

$$|\Lambda_{\vec\beta}(f_1,f_2,f_3)| \lesssim \varepsilon^{-1}A^{\varepsilon}|F_1|^{\alpha}|F_2|^{\alpha_\star}|F_3|^{-\frac12}, \qquad A := \Big(\log\big(\tfrac{|F_3|}{\min\{|F_1|,|F_2|\}}\big)\Big)_*;$$

estimate (1.7) then follows by taking $\varepsilon = \min\{2(1-\alpha_2), (\log A)^{-1}\}$.

8. Interior estimates and Lorentz-Orlicz bounds for $\mathrm{BHT}_{\vec b}$

In this section, we list a number of corollaries following from our main theorems. The proofs are given in Section 9.

8.1. **Blow-up rates of interior estimates.** The endpoint bounds of our main results can be equivalently reformulated as estimates, of the appropriate type, for tuples $\vec\alpha \in \operatorname{int}\mathcal{H}$ with controlled dependence of the constants on the distances from $\vec\alpha$ to each side of $\partial\mathcal{H}$. We parametrize our tuples by

$$\begin{aligned}&\vec\alpha(\varrho,\delta) = \big(1-\varrho, \tfrac12+\varrho-\delta, -\tfrac12+\ \delta\big), \quad q(\delta) = \big(\tfrac32-\delta\big)^{-1},\\ &0<\varrho\leq\tfrac14+2^{-5}, \quad 0<\delta\leq\tfrac12+\varrho, \quad \min\{\varrho,\delta\}\leq 2^{-10}.\end{aligned} \tag{8.1}$$

The restrictions on ϱ, δ correspond to approaching $\partial\mathcal{H}$ within the darker shaded region in Figure 1. Estimates for other tuples near $\partial\mathcal{H}$ can be recovered by symmetry considerations.

The first corollary is devoted to $L^{q(\delta),\infty}$ estimates.

Corollary 5. *Let $\vec{\alpha} = \vec{\alpha}(\varrho, \delta), q = q(\delta)$ be as in* (8.1). *Then, we have the following bounds for $\|\mathrm{BHT}_{\vec{b}}(f_1, f_2)\|_{q,\infty}$:*

$$\le C_{\vec{\beta}} \max\left\{\varrho^{-1}, |\log\delta|\right\}|F_1|^{\alpha_1}|F_2|^{\alpha_2}, \qquad \forall\, |f_1| \le \mathbf{1}_{F_1},\ |f_2| \le \mathbf{1}_{F_2}; \tag{8.2}$$

$$\le C_{\vec{\beta}}\varrho^{-1}\left(\max\left\{\varrho^{-1}, \delta^{-1}\right\}\right)_*^{1-2\varrho}\|f_1\|_{\frac{1}{\alpha_1}}|F_2|^{\alpha_2}, \qquad \forall\, |f_2| \le \mathbf{1}_{F_2}; \tag{8.3}$$

$$\le C_{\vec{\beta}}\varrho^{-1}\max\left\{1, \left(\tfrac{\varrho}{\delta}\right)_*^{2\varrho}\right\}|F_1|^{\alpha_1}\|f_2\|_{\frac{1}{\alpha_2}} \qquad \forall\, |f_1| \le \mathbf{1}_{F_1}; \tag{8.4}$$

$$\le C_{\vec{\beta}}\|f_1\|_{\frac{1}{\alpha_1}}\|f_2\|_{\frac{1}{\alpha_2}}\varrho^{-1}\begin{cases}(\varrho^{-1})_*^{1-2\varrho}, & \varrho \le \delta,\\ \delta^{-\frac{1}{q}}\left(\frac{\varrho}{\delta}\right)^{2\varrho}, & \varrho > \delta.\end{cases} \tag{8.5}$$

The second deals with strong-type estimates.

Corollary 6. *Let $\vec{\alpha} = \vec{\alpha}(\varrho, \delta), q = q(\delta)$ be as in* (8.1). *Then, we have the following bounds for $\|\mathrm{BHT}_{\vec{b}}(f_1, f_2)\|_{q}$:*

$$\le C_{\vec{\beta}}\frac{\max\{\varrho^{-1}, |\log\delta|\}}{(\min\{\varrho, \delta\})^{\max\{1,\frac{1}{q}\}}}|F_1|^{\alpha_1}|F_2|^{\alpha_2}, \qquad \forall|f_1| \le \mathbf{1}_{F_1},\ |f_2| \le \mathbf{1}_{F_2}; \tag{8.6}$$

$$\le C_{\vec{\beta}}\frac{\max\{\varrho^{-1}, |\log\delta|\}}{(\min\{\varrho, \delta\})^{2\max\{1,\frac{1}{q}\}}}\|f_1\|_{\frac{1}{\alpha_1}}\|f_2\|_{\frac{1}{\alpha_2}}. \tag{8.7}$$

In Table 2, we summarize the blow-up rates of eight possible types of interior estimates as the tuple $\vec{\alpha}$ approaches the segments $\overline{CA}$, $\overline{AB}$ away from the endpoint A on the boundary of the shaded hexagon $\mathcal{H}$ in Figure 1. We use the results of

TABLE 2. Blow-up rates near the $\overline{CA}$ and $\overline{AB}$ sides of $\mathcal{H}$ away from the endpoint A. We recall that $(t)_* = (1+t)(\log(\mathrm{e}+t))^3$. The first four rows come from Corollary 5. Rows five and eight are obtained from Corollary 6. The sixth and seventh rows are obtained by specializing the corresponding estimate of line eight.

$\vec{\alpha} = \vec{\alpha}(\varrho, \delta)$, $q = q(\delta)$; see (8.1)	$\varrho = \mathrm{dist}(\vec{\alpha}, \overline{CA}) \to 0$		$\delta = \mathrm{dist}(\vec{\alpha}, \overline{AB}) \to 0$		
	BHT	Q[7]	BHT	BHT[1]	Q[7]
$L^{\frac{1}{\alpha_1},\mathrm{rst}} \times L^{\frac{1}{\alpha_2},\mathrm{rst}} \to L^{q,\infty}$, (8.2)	ϱ^{-1}	ϱ^{-1}	$\|\log\delta\|$	δ^{-1}	1
$L^{\frac{1}{\alpha_1}} \times L^{\frac{1}{\alpha_2},\mathrm{rst}} \to L^{q,\infty}$, (8.3)	$\varrho^{-1}(\varrho^{-1})_*^{1-2\varrho}$	ϱ^{-1}	$(\delta^{-1})_*^{1-2\varrho}$	N/A	1
$L^{\frac{1}{\alpha_1},\mathrm{rst}} \times L^{\frac{1}{\alpha_2}} \to L^{r,\infty}$, (8.4)	ϱ^{-1}		$(\delta^{-1})_*^{2\varrho}$	N/A	1
$L^{\frac{1}{\alpha_1}} \times L^{\frac{1}{\alpha_2}} \to L^{q,\infty}$, (8.5)	$\varrho^{-1}(\varrho^{-1})_*^{1-2\varrho}$	ϱ^{-1}	$\delta^{-\frac{1}{q}}\|\log\delta\|$	$\delta^{-(1+\frac{1}{q})}$	$\delta^{-\frac{1}{q}}$
$L^{\frac{1}{\alpha_1},\mathrm{rst}} \times L^{\frac{1}{\alpha_2},\mathrm{rst}} \to L^{q}$, (8.6)	$\varrho^{-(1+\frac{1}{q})}$		$\delta^{-\frac{1}{q}}\|\log\delta\|$	$\delta^{-(1+\frac{1}{q})}$	$\delta^{-\frac{1}{q}}$
$L^{\frac{1}{\alpha_1}} \times L^{\frac{1}{\alpha_2},\mathrm{rst}} \to L^{q}$	$\varrho^{-(1+\frac{2}{q})}$		$\delta^{-\frac{2}{q}}\|\log\delta\|$	$\delta^{-(1+\frac{2}{q})}$	$\delta^{-\frac{2}{q}}$
$L^{\frac{1}{\alpha_1},\mathrm{rst}} \times L^{\frac{1}{\alpha_2}} \to L^{q}$	$\varrho^{-(1+\frac{2}{q})}$		$\delta^{-\frac{2}{q}}\|\log\delta\|$	$\delta^{-(1+\frac{2}{q})}$	$\delta^{-\frac{2}{q}}$
$L^{\frac{1}{\alpha_1}} \times L^{\frac{1}{\alpha_2}} \to L^{q}$, (8.7)	$\varrho^{-(1+\frac{2}{q})}$		$\delta^{-\frac{2}{q}}\|\log\delta\|$	$\delta^{-(1+\frac{2}{q})}$	$\delta^{-\frac{2}{q}}$

Corollaries 5 and 6, including for comparison the corresponding estimates following, with the same methods, from the endpoint results of [1], and [7] for the Walsh case, mentioned in the introduction. Note that the behavior of the estimates of Corollaries 5 and 6 near the corners A (where both parameters ϱ, δ can go to zero at the same time) and C (where δ is away from zero) can be read directly from the corollaries.

8.2. **Lorentz-Orlicz space estimates.** In the same spirit of the article [5], we detail several Lorentz-Orlicz space bounds near Hölder tuples $(p_1, p_2, \frac{2}{3})$. The first one is obtained from Corollary 4, improving the logarithmic bumps in [5, Section 4.1] to doubly logarithmic ones.

Corollary 7. *Define the Lorentz-Orlicz quasinorms*

$$\|f\|_{L^{\frac{2}{3},\infty}(\log\log L)^{-1}(\mathbb{R})} := \sup_{t>0} \frac{t^{\frac{3}{2}} f^\star(t)}{\log\log(\mathrm{e}^{\mathrm{e}}+t)},$$
$$\|f\|_{L^{p,\frac{2}{3}}(\log\log L)^{\frac{2}{3}}(\mathbb{R})} := \left\| t^{\frac{1}{p}} \log\log\left(\mathrm{e}+\tfrac{1}{t}\right) f^\star(t) \right\|_{L^{\frac{2}{3}}(\mathbb{R};\frac{\mathrm{d}t}{t})}, \qquad 0<p<\infty.$$

Let $1<p_1,p_2<2$, $\frac{1}{p_1}+\frac{1}{p_2}=\frac{3}{2}$. *We have the estimate*

$$\|\mathrm{BHT}_{\vec{b}}(f_1,f_2)\|_{L^{\frac{2}{3},\infty}(\log\log L)^{-1}(\mathbb{R})} \le C_{\vec{\beta}} \prod_{j=1}^{2} (p_j)' \|f_j\|_{L^{p_j,\frac{2}{3}}(\log\log L)^{\frac{2}{3}}(\mathbb{R})}.$$

In the second corollary, which stems from Theorem 2, the first functional argument has no Lorentz-Orlicz bumps. This is also an improvement over [5, Section 4.1], which, unlike the results therein, does not rely on extrapolation theory.

Corollary 8. *For* $\varepsilon \ge 0$, *define the Lorentz-Orlicz quasinorms*

$$\|f\|_{L^{\frac{2}{3},\infty}(\log L)^{-\varepsilon}(\mathbb{R})} := \sup_{t>0} \frac{t^{\frac{3}{2}} f^\star(t)}{\big(\log(\mathrm{e}+t)\big)^{\varepsilon}},$$
$$\|f\|_{L^{p,\frac{2}{3}}(\log L)^{\varepsilon}(\mathbb{R})} := \left\| t^{\frac{1}{p}} \big(\log\left(\mathrm{e}+\tfrac{1}{t}\right)\big)^{\frac{3\varepsilon}{2}} f^\star(t) \right\|_{L^{\frac{2}{3}}(\mathbb{R};\frac{\mathrm{d}t}{t})}, \qquad 0<p<\infty.$$

Let $1<p_1,p_2<2$, $\frac{1}{p_1}+\frac{1}{p_2}=\frac{3}{2}$. *Then, for each* $\varepsilon > \frac{2}{(p_2)'}$, *there exists* $C_{p_1,\varepsilon}>0$ *such that*

$$\|\mathrm{BHT}_{\vec{b}}(f_1,f_2)\|_{L^{\frac{2}{3},\infty}(\log L)^{-\varepsilon}(\mathbb{R})} \le C_{\vec{\beta}} C_{p_1,\varepsilon} \|f_1\|_{p_1} \|f_2\|_{L^{p_2,\frac{2}{3}}(\log L)^{\varepsilon}(\mathbb{R})}.$$

Finally, Theorem 1 has as corollaries the following bounds near $L^1 \times L^2$, improving on the results of [5, Section 4.2]. Notice that the L^2 component, unlike in [5], has no Lorentz-Orlicz bumps.

Corollary 9. *We have the bounds*

$$\mathrm{BHT}_{\vec{b}} : L^{1,\frac{2}{3}}(\log L)^{\frac{2}{3}}(\mathbb{R}) \times L^2(\mathbb{R}) \to L^{\frac{2}{3},\infty}(\log L)^{-1}(\mathbb{R}),$$
$$\mathrm{BHT}_{\vec{b}} : L^1 \log L \log\log L(\mathbb{R}) \times L^2(\mathbb{R}) \to L^{\frac{2}{3},\infty}(\log L)^{-1}(\mathbb{R}).$$

9. Proofs of the corollaries of Section 8

9.1. **Proof of Corollary 5.** Recall that $\vec{\alpha} = (1-\varrho, \frac12 + \varrho - \delta, -\frac12 + \delta)$. In view of the equivalence

$$\|f\|_{p,\infty} \sim \sup_{\substack{F\subset\mathbb{R}\\ |F|<\infty}} \inf_{\substack{F'\subset F\\ C|F'|\geq |F|}} |F|^{\frac1p - 1}\big\langle f, \exp(i\arg f(\cdot))\mathbf{1}_{F'}\big\rangle, \tag{9.1}$$

all the estimates of the corollary will be proved by showing that, for all f_1,f_2 as specified and for all $F_3 \subset \mathbb{R}$, there exists a major subset F_3' of F_3 such that

$$\sup_{|f_3|\leq \mathbf{1}_{F_3'}} |F_3|^{-\alpha_3}\big|\Lambda_{\vec{\beta}}(f_1,f_2,f_3)\big| \tag{9.2}$$

is bounded by the corresponding right hand side.

Interpolation preliminaries. Before the actual proofs, we derive three abstract off-diagonal weak-type interpolation lemmata which will be extensively relied upon. Below, let T be a sublinear operator on $\mathbb{R}$ mapping Schwartz functions to locally integrable functions. We indicate by T^* the formal adjoint of T. What we have in mind is the linear operator

$$f_2 \mapsto T_{f_1}(f_2) := \mathrm{BHT}_{\vec{b}}(f_1, f_2),$$

where f_1 is a fixed Schwartz function. Observe that by essential self-adjointness of $\mathrm{BHT}_{\vec{b}}$, we have that $(T_{f_1})^*(f_3) = \mathrm{BHT}_{\vec{b}'}(f_1, f_3)$, where $\vec{b}'$ shares the same degeneracy constant $\Delta_{\vec{\beta}}$ associated to $\vec{b}$.

The first lemma is a variant of the usual off-diagonal Marcinkiewicz interpolation theorem; see e.g. [21]. We sketch the proof to emphasize the dependence of the constants.

Lemma 9.1. *Let there be given*

$$\tfrac12 \leq p_0 < p_1 \leq \infty, \quad \tfrac12 \leq q_0, q_1 \leq \infty,\ q_0, q_1 \neq 1, \qquad \tfrac{1}{p_0} - \tfrac{1}{p_1} = \tfrac{1}{q_0} - \tfrac{1}{q_1} = \mu > 0.$$

Assume that T satisfies the bounds

$$\|Tg\|_{q_j,\infty} \leq K_j |G|^{\frac{1}{p_j}} \qquad \forall |g| \leq \mathbf{1}_G. \tag{9.3}$$

Let $b_j = \min\{q_j, 1\}$. Then, for all $0<\vartheta<1$,

$$\|Tg\|_{q(\vartheta),\infty} \leq C(\vartheta)\|g\|_{p(\vartheta)}, \qquad \tfrac{1}{p(\vartheta)} = \tfrac{(1-\vartheta)}{p_0} + \tfrac{\vartheta}{p_1}, \quad \tfrac{1}{q(\vartheta)} = \tfrac{(1-\vartheta)}{q_0} + \tfrac{\vartheta}{q_1},$$

where

$$C(\vartheta) = C\big(K_0\gamma_0(\mu\vartheta)^{-\frac{1}{b_0}}\big)^{1-\vartheta}\big(K_1\gamma_1(\mu(1-\vartheta))^{-\frac{1}{b_1}}\big)^{\vartheta}, \qquad \gamma_j := \frac{q_j}{|1-q_j|},\ j = 0,1.$$

Proof. By eventually replacing p_1 with $p((\vartheta+1)/2)$, we can assume that $p_1 < \infty$. We preliminarily observe that the assumptions (9.3), coupled with the ℓ^q triangle inequality on $L^{q,\infty}(\mathbb{R})$ [4, Section 3] upgrade to the bounds

$$\|Tg\|_{q_j,\infty(\mathbb{R})} \lesssim \gamma_j K_j \|g\|_{p_j,b_j}, \qquad j=0,1,$$

where $\|\cdot\|_{\pi,\mu}$ denotes the Lorentz quasinorm on the Lorentz space $L^{\pi,\mu}(\mathbb{R})$. We begin the actual proof: by a rearrangement argument, we can assume that $g = g^\star$.

Let $\delta > 0$ be a parameter to be chosen later. For $t > 0$, we define $g^t(x) = g(x)\mathbf{1}_{g(x)>g(\delta t)}$, $g_t = g - g^t$. Using the above display, we see that

$$t^{\frac{1}{q}}(Tg^{2t})^\star(t) \lesssim \gamma_0 K_0 t^{-(\frac{1}{p_0}-\frac{1}{p(\vartheta)})}\|g^{2t}\|_{p_0,b_0},$$
$$t^{\frac{1}{q}}(Tg_{2t})^\star(t) \lesssim \gamma_1 K_1 t^{-(\frac{1}{p_1}-\frac{1}{p(\vartheta)})}\|g_{2t}\|_{p_1,b_1}.$$

The lemma then follows from the estimates

$$\sup_{t>0} t^{-(\frac{1}{p_0}-\frac{1}{p(\vartheta)})}\|g^t\|_{p_0,b_0} \lesssim \delta^{\frac{1}{p_0}-\frac{1}{p(\vartheta)}}(\mu\vartheta)^{-\frac{1}{b_0}}\|g\|_p,$$
$$\sup_{t>0} t^{-(\frac{1}{p_1}-\frac{1}{p(\vartheta)})}\|g_t\|_{p_1,b_1} \lesssim \delta^{\frac{1}{p_1}-\frac{1}{p(\vartheta)}}(\mu(1-\vartheta))^{-\frac{1}{b_1}}\|g\|_p,$$

which are obtained by means of Hölder's inequality and finally by optimizing δ. □

We will also use a version which does not upgrade the type of the estimate. Notice that the constant in (9.4) does not blow up as $\vartheta \to 0$ or 1. The proof is simple and we omit it.

Lemma 9.2. *Let there be given*

$$\tfrac{1}{2} < p_0 < p_1 \leq \infty, \quad \tfrac{1}{2} \leq q_0, q_1 \leq \infty, \qquad \tfrac{1}{p_0} - \tfrac{1}{p_1} = \tfrac{1}{q_0} - \tfrac{1}{q_1} = \mu > 0.$$

Assume that T satisfies the bounds

$$\|Tg\|_{q_j,\infty} \leq K_j\|g\|_{p_j}.$$

Then, for all $0 < \vartheta < 1$,

$$\|Tg\|_{q(\vartheta),\infty} \lesssim (K_0)^{1-\vartheta}(K_1)^\vartheta\|g\|_{p(\vartheta)}, \quad \tfrac{1}{p(\vartheta)} = \tfrac{(1-\vartheta)}{p_0} + \tfrac{\vartheta}{p_1},\ \tfrac{1}{q(\vartheta)} = \tfrac{(1-\vartheta)}{q_0} + \tfrac{\vartheta}{q_1}. \tag{9.4}$$

The next lemma exploits the equivalence (9.1) to interpolate between the $L^{p,\infty}$ estimates of T and T^*.

Lemma 9.3. *Let $0 < \alpha < 1$, $-1 < \beta < 0$, and assume that for all $G \subset \mathbb{R}$ of finite measure*

$$\|Tg\|_{\frac{1}{1-\beta},\infty},\ \|T^*g\|_{\frac{1}{1-\beta},\infty} \leq K|G|^\alpha \qquad \forall |g| \leq \mathbf{1}_G.$$

Then, for all $0 < t \leq \alpha$ and for all $G \subset \mathbb{R}$ of finite measure, we have the estimate

$$\|Tg\|_{\frac{1}{1-\beta-t},\infty},\ \|T^*g\|_{\frac{1}{1-\beta-t},\infty} \lesssim \frac{K}{|\beta|}|G|^{\alpha-t} \qquad \forall |g| \leq \mathbf{1}_G.$$

Proof. By symmetry, it suffices to carry the proof for T. Fixing $G \subset \mathbb{R}$ of finite measure, $|g| \leq \mathbf{1}_G$, and using (9.1), it suffices to show that for all $0 < t \leq \alpha$, $F \subset \mathbb{R}$ with $|F| < \infty$, there exists a set $F' \subset F$ with $|F| \leq C|F'|$ such that, for all $|f| \leq \mathbf{1}_{F'}$,

$$|\langle Tg, f\rangle| \lesssim K|\beta|^{-1}|G|^\alpha|F|^\beta\big(\tfrac{|F|}{|G|}\big)^t. \tag{9.5}$$

Since inequality (9.5) holds by assumption for $t = 0$, with no need for $|\beta|^{-1}$, there is nothing to prove if $|G| \leq |F|$. Assume $|G| > |F|$ and let $n = \lceil \log|G| - \log|F| \rceil$. We apply our assumption for T^* instead, so that the roles of F and G are reversed, and, via (9.1), we obtain the existence of a set $H^{(1)} \subset G := G^{(0)}$, with $|H^{(1)}| \geq C^{-1}|G|$, such that

$$|\langle g\mathbf{1}_{H^{(1)}}, T^*f\rangle| \lesssim K|F|^\alpha|G|^\beta \qquad \forall |f| \leq \mathbf{1}_F.$$

Iterating, we define a sequence $G^{(k+1)} = G^{(k)} \backslash H^{(k+1)}$, stopping when $|G^{(\bar{k})}| \leq |F|$. Note that $|G^{(k)}| \leq \mathrm{e}^{-ck}|G|$, so that $\bar{k} \leq n/c$, where $c = \log C - \log(C-1)$. Finally, the assumption provides a set $F' \subset F$ with $|F'| \geq C^{-1}|F|$ such that

$$|\langle T(g\mathbf{1}_{G^{(\bar{k})}}), f\mathbf{1}_{F'}\rangle| \lesssim K|G^{(\bar{k})}|^{\alpha}|F|^{\beta} \leq K|F|^{\alpha+\beta} \qquad \forall |g| \leq \mathbf{1}_G.$$

Observing that $G = H^{(1)} \cup \cdots \cup H^{(\bar{k}-1)} \cup G^{(\bar{k})}$ leads to the estimate

$$\begin{aligned} |\langle Tg, f\mathbf{1}_{F'}\rangle| &\leq |\langle T(g\mathbf{1}_{G^{(\bar{k})}}), f\mathbf{1}_{F'}\rangle| + \sum_{k=1}^{\bar{k}} |\langle g\mathbf{1}_{H^{(k)}}, T^*(f\mathbf{1}_{F'})\rangle| \\ &\lesssim K|F|^{\alpha+\beta} + K|F|^{\alpha}|G|^{\beta}\sum_{k=1}^{n/c} \mathrm{e}^{-\beta c k} \\ &\lesssim K(1+c^{-1}|\beta|^{-1})|F|^{\alpha+\beta} \lesssim K|\beta|^{-1}|G|^{\alpha-t}|F|^{\beta+t}, \end{aligned}$$

which completes the proof of (9.5) and, in turn, of the lemma. □

Proof of (8.2). The proof is split into two cases.

Case $0 < \delta \leq \varrho/2$. The bound (9.2) $\lesssim |F_1|^{\alpha_1}|F_2|^{\alpha_2}|F_3|^{\alpha_3}$ when $|F_3| \leq |F_1|$ follows from the (uniform) strong-type bounds (1.2) on the line segment $(1/4, a, 3/4-a)$, $0 \leq a \leq 3/4$. When $|F_1| \leq |F_3|$ instead, we apply (1.7) with tuple $\vec{a} := (1-\varrho+\delta, \alpha_2, -\frac{1}{2}) \in \mathcal{S}_3$, which has $1-a_1 \geq \varrho/2$, $1-a_2 \gtrsim 1$, yielding a major subset F_3' of F_3 such that, for all suitably restricted f_j,

$$\left|\Lambda_{\vec{\beta}}(f_1,f_2,f_3)\right| \lesssim \left[\left(\tfrac{|F_1|}{|F_3|}\right)^{\delta} \max\left\{\varrho^{-1}, \log\log\left(\mathrm{e}^{\mathrm{e}} + \tfrac{|F_3|}{|F_1|}\right)\right\}\right] |F_1|^{\alpha_1}|F_2|^{\alpha_2}|F_3|^{\alpha_3}.$$

By optimizing $t \mapsto t^{\delta}\log\log(\mathrm{e}^{\mathrm{e}}+t^{-1})$ on $t \in [0,1]$, one sees that the square bracketed term is bounded by $C\max\{\varrho^{-1}, |\log\delta|\}$, as claimed in (8.2), concluding the proof of this case.

Case $\delta > \varrho/2$. In this range $\max\{\varrho^{-1}, |\log\delta|\} = \varrho^{-1}$. We fix F_1, $|f_1| \leq \mathbf{1}_{F_1}$ and define the linear operator $g \mapsto T_{f_1}(g) := \mathrm{BHT}_{\vec{b}}(f_1, g)$. The previous case and essential self-adjointness yield

$$\|T_{f_1}(g)\|_{q(\varrho/2),\infty}, \|(T_{f_1})^*(g)\|_{q(\varrho/2),\infty} \lesssim \varrho^{-1}|F_1|^{\alpha_1}|G|^{\frac{1}{2}+\frac{\varrho}{2}} \qquad \forall |g| \leq \mathbf{1}_G,$$

so that an application of Lemma 9.3 with $t = \delta - \varrho/2$ entails

$$\sup_{G, |g| \leq \mathbf{1}_G} |G|^{-\alpha_2}\|T_{f_1}(g)\|_{q(\delta),\infty} \lesssim \sup_{G, |g|\leq \mathbf{1}_G} |G|^{-(\frac{1}{2}+\frac{\varrho}{2})}\|T_{f_1}(g)\|_{q(\varrho/2),\infty} \lesssim \varrho^{-1}|F_1|^{\alpha_1},$$

which is what is required in (8.2). □

Proof of (8.3). We fix $f_1 \in L^{\frac{1}{\alpha_1}}(\mathbb{R})$ of unit norm. The proof is again split into two cases.

Case $\varrho \geq \delta$. We show that for any given F_2, F_3 there is a major subset F_3' of F_3 such that

$$\left|\Lambda_{\vec{\beta}}(f_1,f_2,f_3)\right| \lesssim \varrho^{-1}\left(\max\left\{\varrho^{-1}, \delta^{-1}\right\}\right)_*^{1-2\varrho}|F_2|^{\alpha_2}|F_3|^{\alpha_3} \tag{9.6}$$

for all $|f_2| \leq \mathbf{1}_{F_2}$, $|f_3| \leq \mathbf{1}_{F_3'}$, which implies (8.3) via the usual equivalence. Assume first $|F_2| \geq |F_3|$. We apply Theorem 3 to the obvious choice of f_1 and to f_2 restricted to F_2, obtaining a major subset F_3' of F_3 such that

$$\begin{aligned}\left|\Lambda_{\vec\beta}(f_1,f_2,f_3)\right| &\lesssim \varrho^{-1}\left(\varrho^{-1}\right)_*^{1-2\varrho}\|f_2\|_2|F_3|^{\frac12-\alpha_1}\\ &\le \varrho^{-1}\left(\varrho^{-1}\right)_*^{1-2\varrho}|F_2|^{\alpha_2}|F_3|^{\alpha_3}\left(\frac{|F_3|}{|F_2|}\right)^{\alpha_2-\frac12},\end{aligned}$$

which, since $\alpha_2 \ge \frac12$, is stronger than (9.6). If instead $|F_3|\ge|F_2|$, we apply estimate (1.5) with tuple $\vec a=(\alpha_1,\alpha_2+\delta,-1/2)\in\mathcal S_3$, obtaining, for any pair of suitably restricted functions f_2,f_3,

$$\left|\Lambda_{\vec\beta}(f_1,f_2,f_3)\right| \lesssim \varrho^{-1}|F_2|^{\alpha_2}|F_3|^{\alpha_3}\left[\left(\max\left\{\varrho^{-1},\log\left(\tfrac{|F_3|}{|F_2|}\right)\right\}\right)_*^{1-2\varrho}\left(\tfrac{|F_2|}{|F_3|}\right)^{\delta}\right];$$

noting that the term in square brackets is $\lesssim(\max\{\varrho^{-1},\delta^{-1}\})_*^{1-2\varrho}$ leads to (9.6).

Case $\delta>\varrho$. Analogously to what we did in the previous proof, we define the linear operator $g\mapsto T_{f_1}(g):=\mathrm{BHT}_{\vec b}(f_1,g)$. Again by Theorem 3 (note that $\delta=\varrho$ corresponds to $\alpha_2=\frac12$)

$$\|T_{f_1}(g)\|_{q(\varrho),\infty},\|(T_{f_1})^*(g)\|_{q(\varrho),\infty}\lesssim \varrho^{-1}\left(\varrho^{-1}\right)_*^{1-2\varrho}\|f_1\|_{\frac{1}{\alpha_1}}|G|^{\frac12}\qquad \forall|g|\le \mathbf 1_G,$$

and Lemma 9.3 with $t=\delta-\varrho$ entails

$$\sup_{G,|g|\le\mathbf 1_G}|G|^{-\alpha_2}\|T_{f_1}(g)\|_{q(\delta)}\lesssim \sup_{G,|g|\le\mathbf 1_G}|G|^{-\frac12}\|T_{f_1}(g)\|_{q(\varrho)}\lesssim \varrho^{-1}\left(\varrho^{-1}\right)_*^{1-2\varrho}\|f_1\|_{\frac{1}{\alpha_1}},$$

which is what we had to prove in (8.3). □

Proof of (8.4). We separate two cases.

Case $\varrho\ge\delta$. We fix $f_2\in L^{1/\alpha_2}(\mathbb R)$ of unit norm and prove that for any given F_1,F_3 there is a major subset F_3' of F_3 such that

$$\left|\Lambda_{\vec\beta}(f_1,f_2,f_3)\right|\lesssim\varrho^{-1}\left(\tfrac{\varrho}{\delta}\right)_*^{2\varrho}|F_1|^{\alpha_1}|F_3|^{\alpha_3}\qquad\forall|f_1|\le\mathbf 1_{F_1},\ |f_3|\le\mathbf 1_{F_3'}. \tag{9.7}$$

As in the proof of (8.2), the case $|F_3|\le|F_1|$ follows from the known strong-type bounds (1.2). When $|F_1|\le|F_3|$, we apply estimate (1.5) with tuple $\vec a=(\alpha_2,\alpha_1+\delta,-1/2)\in\mathcal S_3$ and switching the roles of f_1,f_2. We obtain, for suitably restricted f_1,f_3,

$$\left|\Lambda_{\vec\beta}(f_1,f_2,f_3)\right|\lesssim\varrho^{-1}|F_1|^{\alpha_1}|F_3|^{\alpha_3}\left[\left(\log\left(\tfrac{|F_3|}{|F_1|}\right)\right)_*^{2(\varrho-\delta)}\left(\tfrac{|F_1|}{|F_3|}\right)^{\delta}\right];$$

estimating the term in square brackets by $C(\frac{\varrho}{\delta})_*^{2\varrho}$ leads to (9.7) and finishes the proof.

Case $\varrho\le\delta$. For this case, we fix F_1 and $|f_1|\le\mathbf 1_{F_1}$ and interpolate through Lemma 9.2 the estimates

$$\|T_{f_1}(g)\|_{q(\varrho),\infty}\lesssim\varrho^{-1}|F|^{\alpha_1}\|g\|_2,\qquad \|T_{f_1}(g)\|_{q(\varrho+1/2),\infty}\lesssim\varrho^{-1}|F|^{\alpha_1}\|g\|_\infty$$

for the linear operator $g\mapsto T_{f_1}(g):=\mathrm{BHT}_{\vec b}(f_1,g)$, the first of which is obtained in the previous case, while the second can be read from (8.2) when $\alpha_2=0$. □

Proof of (8.5). The proof is split into two cases, both relying on interpolation.

Case $\varrho \le \delta$. For this case, we fix $f_1 \in L^{1/\alpha_1}(\mathbb{R})$ and interpolate through Lemma 9.2 the estimates

$$\|T_{f_1}(g)\|_{q(\varrho),\infty} \lesssim \varrho^{-1}(\varrho^{-1})_*^{1-2\varrho}\|f_1\|_{\frac{1}{\alpha_1}}\|g\|_2,$$
$$\|T_{f_1}(g)\|_{q(\varrho+1/2),\infty} \lesssim \varrho^{-1}(\varrho^{-1})_*^{1-2\varrho}\|f_1\|_{\frac{1}{\alpha_1}}\|g\|_\infty$$

for $T_{f_1}(g) := \mathrm{BHT}_{\vec{b}}(f_1, g)$. The first estimate above is exactly a reformulation of Theorem 3, while the second one can be read from the case $\alpha_2 = 0$ of (8.3).

Case $\varrho > \delta$. We fix $f_2 \in L^{1/\alpha_2}(\mathbb{R})$. We learn from (8.4) that the linear operator $T_{f_2}(g) := \mathrm{BHT}_{\vec{b}}(g, f_2)$ satisfies the estimates

$$\|T_{f_2}(g)\|_{q(\frac{\delta}{2}),\infty} \lesssim \varrho^{-1}\Big(\frac{\varrho}{\delta}\Big)_*^{2\varrho}\|f_2\|_{\frac{1}{\alpha_2}}|F_1|^{\alpha_1+\frac{\delta}{2}},$$
$$\|T_{f_2}(g)\|_{q(\frac{3\delta}{2}),\infty} \lesssim \varrho^{-1}\Big(\frac{\varrho}{\delta}\Big)_*^{2\varrho}\|f_2\|_{\frac{1}{\alpha_2}}|F_1|^{\alpha_1-\frac{\delta}{2}},$$

respectively corresponding to (8.4) for tuples $(1-\rho \pm \delta/2, \alpha_2, -\frac{1}{2}+\delta \mp \delta/2)$. We now use Lemma 9.1 on T, with $\vartheta = \frac{1}{2}$, $p_0 = \frac{1}{\alpha_1+\delta/2}$, $p_1 = \frac{1}{\alpha_1-\delta/2}$, $q_0 = q(3\delta/2)$, $q_1 = q(\delta/2)$, and observe that $\mu = \delta$, γ_0, γ_1 therein are uniformly bounded, and $1/(2b_0) + 1/(2b_1) = 1/q(\delta)$, which entails the estimate

$$\|T_{f_2}(f_1)\|_{q(\delta),\infty} \lesssim \delta^{-\frac{1}{q(\delta)}}\varrho^{-1}\Big(\tfrac{\varrho}{\delta}\Big)_*^{2\varrho}\|f_1\|_{\frac{1}{\alpha_1}}\|f_2\|_{\frac{1}{\alpha_2}}.$$

This completes the proof. □

9.2. **Proof of Corollary 6.** In the proof, we will need the following more precise form of the interpolation result [25, Lemma 3.11] (see also [29, Theorem 3.8]).

Lemma 9.4. *Let T be a bisublinear operator on $\mathbb{R}$ mapping pairs of Schwartz functions into measurable functions. Let $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3) \in \operatorname{int}\mathcal{H}$ be a Hölder tuple with $\alpha_3 = \min\{\alpha_j\} \le 0$. Suppose that for a given $\varepsilon > 0$, $O_\varepsilon(\vec{\alpha}) := \{\vec{a} : a_1+a_2+a_3 = 1, \max_j |\alpha_j - a_j| \le \varepsilon\} \subset \operatorname{int}\mathcal{H}$, and there exists $K > 0$ such that the estimate*

$$\|T(f_1,f_2)\|_{\frac{1}{1-a_3},\infty} \le K\prod_{j=1}^{2}|F_j|^{a_j} \qquad \forall\, |f_1| \le \mathbf{1}_{F_1},\, |f_2| \le \mathbf{1}_{F_2} \tag{9.8}$$

holds for all tuples $\vec{a} \in O_\varepsilon(\vec{\alpha})$. Then,

$$\|T(f_1,f_2)\|_{\frac{1}{1-\alpha_3}} \lesssim \varepsilon^{-(1-\alpha_3)}K|F_1|^{\alpha_1}|F_2|^{\alpha_2}, \qquad \forall |f_1| \le \mathbf{1}_{F_1},\ |f_2| \le \mathbf{1}_{F_2},$$
$$\|T(f_1,f_2)\|_{\frac{1}{1-\alpha_3}} \lesssim \varepsilon^{-2(1-\alpha_3)}K\|f_1\|_{\frac{1}{\alpha_1}}\|f_2\|_{\frac{1}{\alpha_2}}.$$

Proof. We sketch the proof of the second estimate; the proof of the first estimate is implicit in the argument for the second one. Let $q = (1-\alpha_3)^{-1}$. By rearrangement, we can assume that f_1, f_2 are non-negative, supported on $(0,\infty)$, and non-decreasing. For $j = 1,2$, $k_j \in \mathbb{Z}$, define $f_j^{k_j} = f_j\mathbf{1}_{[2^{k_j},2^{k_j+1})}$. Arguing as in [25], we exploit uniformity in $O_\varepsilon(\vec{\alpha})$ of (9.8) to obtain the estimate

$$\|T(f_1^{k_1}, f_2^{k_2})\|_q^q \lesssim \varepsilon^{-1}K^r 2^{-\varepsilon|k_1-k_2|}\big(\textstyle\prod_{j=1}^2 f_j(2^{k_j})2^{\alpha_j k_j}\big)^q,$$

the implicit constant being absolute. Since

$$\|T(f_1,f_2)\|_q^q \le \sum_{k_1,k_2\in\mathbb{Z}}\|T(f_1^{k_1}, f_2^{k_2})\|_q^q$$

the second estimate of the lemma follows by bounding the resulting sum as in [25]. □

We first prove Corollary 6 for tuples $\vec\alpha$ outside the reflexive Banach triangle, that is, with $q(\delta) \le 1$. For such a tuple, referring to (8.1), set $\varepsilon = \min\{\varrho, \delta\}/2$. We read from (8.2) that condition (9.8) of the interpolation Lemma 9.4 holds for all tuples in $O_\varepsilon(\vec\alpha)$ with constant $K \lesssim \max\{\varrho^{-1}, |\log\delta\}$, so that the estimates of Corollary 6 in this range follow by a straightforward application of the lemma.

We now deal with the case of $\vec\alpha$ inside the reflexive Banach triangle: by symmetry and duality, we can restrict to proving the case $\alpha_2 \ge \alpha_3$. Note that, according to (8.1), $\varrho \le 2^{-5}$. We can then write $\vec\alpha$ as a convex combination of the tuple $\vec\omega = (1/2, 1/6, 1/3)$ and of a tuple $\vec\gamma$ with $\gamma_3 = \min\gamma_j = 0$, $1-\gamma_1 \ge \varrho/16$. Therefore, estimates (8.6), (8.7) for $\vec\alpha$ follow from complex interpolation of the corresponding estimates for $\vec\gamma$, established in the previous step, with those, well-known, for $\vec\omega$. This concludes the proof of Corollary 6.

9.3. **Proofs of Corollaries 7 to 9.** To prove Corollary 7, we observe that estimate (1.7) of Corollary 4 can be rewritten as

$$\frac{t^{\frac{3}{2}}}{\log\log(\mathrm{e}^{\mathrm{e}}+t)}\big(\mathrm{BHT}_{\vec b}(\mathbf{1}_{F_1}, \mathbf{1}_{F_2})\big)^{\star}(t) \lesssim \prod_{j=1}^{2}(p_j)'|F_j|^{\frac{1}{p_j}}\log\log\big(\mathrm{e}^{\mathrm{e}}+|F_j|^{-1}\big) \quad \forall t>0.$$

Then, the corollary follows from the above display by arguing along the lines of [5, Section 4.1]; we omit the details.

To obtain Corollary 8 from Theorem 2, for each $\varepsilon = \frac{2}{(p_2)'} + \zeta > \frac{2}{(p_2)'}$ we take $\alpha_1 = \frac{1}{p_1}, \alpha_2 = \frac{1}{p_2}$ in (1.5) and estimate, using that $1-\alpha_2 = \frac{1}{(p_2)'}$,

$$\begin{aligned}\big|\langle \mathrm{BHT}_{\vec b}(f_1,f_2), f_3\rangle\big| &\lesssim\ C_{p_1}\|f_1\|_{p_1}|F_2|^{\frac{1}{p_2}}|F_3|^{-\frac12}\Big(\log\big(\tfrac{|F_3|}{|F_2|}\big)\Big)_*^{\frac{2}{(p_2)'}}\\ &\lesssim \zeta^{-3}C_{p_1}\|f_1\|_{p_1}|F_2|^{\frac{1}{p_2}}|F_3|^{-\frac12}\Big(\log\Big(\mathrm{e}+\tfrac{|F_3|}{|F_2|}\Big)\Big)^{\varepsilon};\end{aligned}$$

we can take $C_{p_1} := (p_1')^3(p_2)'$. Setting $C_{p_1,\varepsilon} := \zeta^{-3}C_{p_1}$, this can be rearranged into

$$\|\mathrm{BHT}_{\vec b}(f_1,f_2)\|_{L^{\frac23,\infty}(\log L)^{-\varepsilon}(\mathbb{R})} \lesssim C_{p_1,\varepsilon}\|f_1\|_{p_1}|F_2|^{\frac{1}{p_2}}\Big(\log\big(\mathrm{e}+\tfrac{1}{|F_2|}\big)\Big)^{\varepsilon}$$

for all $|f_2| \le \mathbf{1}_{F_2}$. Corollary 8 follows from the last display by recalling that (see [4])

$$\|f\|_{L^{\frac23,\infty}(\log L)^{-\varepsilon}(\mathbb{R})} \sim \inf\Big\{\|\gamma_k\|_{\ell^{\frac23}} : f = \sum_k \gamma_k f_k,\ \|f_k\|_{L^{\frac23,\infty}(\log L)^{-\varepsilon}(\mathbb{R})} \le 1\Big\},$$

and subsequently performing the elementary procedure described in [7, Section 2].

For the details of the derivation of Corollary 9 from Theorem 1, we refer to [7, Section 2]. We only mention that an intermediate step towards the second estimate is the strengthening of Theorem 1:

$$\|\mathrm{BHT}_{\vec b}(f_1,f_2)\|_{L^{\frac23,\infty}(\log L)^{-1}(\mathbb{R})} \le C_{\vec\beta}\|f_1\|_1\|f_2\|_2\log\big(\tfrac{\|f_1\|_\infty}{\|f_1\|_1}\big).$$

The above inequality follows from Theorem 1 via, for instance, the theory of [4] (see also [5, Theorem 3.3]).

Acknowledgements

The first author wants to express his gratitude to the Hausdorff Center for Mathematics at the University of Bonn, Germany, and in particular to Diogo Oliveira e Silva for the kind hospitality during his May 2013 research visit, when this article was initiated.

References

[1] Dmitriy Bilyk and Loukas Grafakos, *Distributional estimates for the bilinear Hilbert transform*, J. Geom. Anal. **16** (2006), no. 4, 563–584, DOI 10.1007/BF02922131. MR2271944 (2007j:46033b)

[2] Dmitriy Bilyk and Loukas Grafakos, *A new way of looking at distributional estimates; applications for the bilinear Hilbert transform*, Collect. Math. **Vol. Extra** (2006), 141–169. MR2264208 (2007j:46033a)

[3] Peter Borwein and Tamás Erdélyi, *Nikolskii-type inequalities for shift invariant function spaces*, Proc. Amer. Math. Soc. **134** (2006), no. 11, 3243–3246, DOI 10.1090/S0002-9939-06-08533-9. MR2231907 (2007e:41014)

[4] María Carro, Leonardo Colzani, and Gord Sinnamon, *From restricted type to strong type estimates on quasi-Banach rearrangement invariant spaces*, Studia Math. **182** (2007), no. 1, 1–27, DOI 10.4064/sm182-1-1. MR2326489 (2008f:46033)

[5] María Jesús Carro, Loukas Grafakos, José María Martell, and Fernando Soria, *Multilinear extrapolation and applications to the bilinear Hilbert transform*, J. Math. Anal. Appl. **357** (2009), no. 2, 479–497, DOI 10.1016/j.jmaa.2009.04.021. MR2557660 (2010k:44008)

[6] R. R. Coifman and Yves Meyer, *On commutators of singular integrals and bilinear singular integrals*, Trans. Amer. Math. Soc. **212** (1975), 315–331. MR0380244 (52 #1144)

[7] Ciprian Demeter and Francesco Di Plinio, *Endpoint bounds for the quartile operator*, J. Fourier Anal. Appl. **19** (2013), no. 4, 836–856, DOI 10.1007/s00041-013-9275-4. MR3089425

[8] Francesco Di Plinio, *Lacunary Fourier and Walsh-Fourier series near L^1*, Collect. Math. **65** (2014), no. 2, 219–232, DOI 10.1007/s13348-013-0094-3. MR3189278

[9] Francesco Di Plinio, *Weak-L^p bounds for the Carleson and Walsh-Carleson operators* (English, with English and French summaries), C. R. Math. Acad. Sci. Paris **352** (2014), no. 4, 327–331, DOI 10.1016/j.crma.2014.02.005. MR3186922

[10] Yen Q. Do and Michael T. Lacey, *On the convergence of lacunacy Walsh-Fourier series*, Bull. Lond. Math. Soc. **44** (2012), no. 2, 241–254, DOI 10.1112/blms/bdr088. MR2914604

[11] Yen Do and Christoph Thiele, *L^p theory for outer measures and two themes of Lennart Carleson united*, Bull. Amer. Math. Soc. (N.S.) **52** (2015), no. 2, 249–296, DOI 10.1090/S0273-0979-2014-01474-0. MR3312633

[12] Loukas Grafakos and Xiaochun Li, *Uniform bounds for the bilinear Hilbert transforms. I*, Ann. of Math. (2) **159** (2004), no. 3, 889–933, DOI 10.4007/annals.2004.159.889. MR2113017 (2006e:42011)

[13] Loukas Grafakos and Rodolfo H. Torres, *Multilinear Calderón-Zygmund theory*, Adv. Math. **165** (2002), no. 1, 124–164, DOI 10.1006/aima.2001.2028. MR1880324 (2002j:42029)

[14] Lars Hörmander, *The analysis of linear partial differential operators. I*, Distribution theory and Fourier analysis, Classics in Mathematics, Springer-Verlag, Berlin, 2003. Reprint of the second (1990) edition [Springer, Berlin; MR1065993 (91m:35001a)]. MR1996773

[15] A. E. Ingham, *A Note on Fourier Transforms*, J. London Math. Soc. **S1-9**, no. 1, 29. MR1574706

[16] Carlos E. Kenig and Elias M. Stein, *Multilinear estimates and fractional integration*, Math. Res. Lett. **6** (1999), no. 1, 1–15, DOI 10.4310/MRL.1999.v6.n1.a1. MR1682725 (2000k:42023a)

[17] Michael Lacey and Christoph Thiele, *L^p estimates on the bilinear Hilbert transform for* $2 < p < \infty$, Ann. of Math. (2) **146** (1997), no. 3, 693–724, DOI 10.2307/2952458. MR1491450 (99b:42014)

[18] Michael Lacey and Christoph Thiele, *On Calderón's conjecture*, Ann. of Math. (2) **149** (1999), no. 2, 475–496, DOI 10.2307/120971. MR1689336 (2000d:42003)

[19] Michael T. Lacey, *The bilinear maximal functions map into L^p for* $2/3 < p \leq 1$, Ann. of Math. (2) **151** (2000), no. 1, 35–57, DOI 10.2307/121111. MR1745019 (2001b:42015)

[20] Xiaochun Li, *Uniform bounds for the bilinear Hilbert transforms. II*, Rev. Mat. Iberoam. **22** (2006), no. 3, 1069–1126, DOI 10.4171/RMI/483. MR2320411 (2008c:42014)

[21] Yi Yu Liang, Li Guang Liu, and Da Chun Yang, *An off-diagonal Marcinkiewicz interpolation theorem on Lorentz spaces*, Acta Math. Sin. (Engl. Ser.) **27** (2011), no. 8, 1477–1488, DOI 10.1007/s10114-011-0287-1. MR2822822 (2012f:46055)

[22] Victor Lie, *On the pointwise convergence of the sequence of partial Fourier sums along lacunary subsequences*, J. Funct. Anal. **263** (2012), no. 11, 3391–3411, DOI 10.1016/j.jfa.2012.08.013. MR2984070

[23] Victor Lie, *On the boundedness of the Carleson operator near* L^1, Rev. Mat. Iberoam. **29** (2013), no. 4, 1239–1262, DOI 10.4171/RMI/755. MR3148602

[24] Camil Muscalu, Jill Pipher, Terence Tao, and Christoph Thiele, *Multi-parameter paraproducts*, Rev. Mat. Iberoam. **22** (2006), no. 3, 963–976, DOI 10.4171/RMI/480. MR2320408 (2008b:42037)

[25] Camil Muscalu, Terence Tao, and Christoph Thiele, *Multi-linear operators given by singular multipliers*, J. Amer. Math. Soc. **15** (2002), no. 2, 469–496, DOI 10.1090/S0894-0347-01-00379-4. MR1887641 (2003b:42017)

[26] Fedor Nazarov, Richard Oberlin, and Christoph Thiele, *A Calderón-Zygmund decomposition for multiple frequencies and an application to an extension of a lemma of Bourgain*, Math. Res. Lett. **17** (2010), no. 3, 529–545, DOI 10.4310/MRL.2010.v17.n3.a11. MR2653686 (2011d:42047)

[27] Richard Oberlin and Christoph Thiele, *New uniform bounds for a Walsh model of the bilinear Hilbert transform*, Indiana Univ. Math. J. **60** (2011), no. 5, 1693–1712, DOI 10.1512/iumj.2011.60.4445. MR2997005

[28] Christoph Thiele, *A uniform estimate*, Ann. of Math. (2) **156** (2002), no. 2, 519–563, DOI 10.2307/3597197. MR1933076 (2003i:47036)

[29] Christoph Thiele, *Wave packet analysis*, CBMS Regional Conference Series in Mathematics, vol. 105, published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006. MR2199086 (2006m:42073)

Department of Mathematics, Brown University, Box 1917, Providence, Rhode Island 02912

E-mail address: fradipli@math.brown.edu

Hausdorff Institute for Mathematics, Universität Bonn, Endenicher Allee 60, D-53115 Bonn, Germany — and — Department of Mathematics, University of California Los Angeles, Los Angeles, California 90095-1555

E-mail address: thiele@math.uni-bonn.de