TWO WEIGHT INEQUALITIES FOR THE CAUCHY TRANSFORM FROM $\mathbb{R}$ TO $\mathbb{C}_+$

MICHAEL T. LACEY$^1$, ERIC T. SAWYER$^2$, CHUN-YEN SHEN$^3$, IGNACIO URIARTE-TUERO$^4$, AND BRETT D. WICK$^5$

ABSTRACT. We characterize those pairs of weights $\sigma$ on $\mathbb{R}$ and $\tau$ on $\mathbb{C}_+$ for which the Cauchy transform $C_{\sigma}f(z) \equiv \int_{\mathbb{R}} \frac{f(t)}{t-z} \sigma(dt)$, $z \in \mathbb{C}_+$, is bounded from $L^2(\mathbb{R};\sigma)$ to $L^2(\mathbb{C}_+;\tau)$. The characterization is in terms of an $A_2$ condition on the pair of weights and testing conditions for the transform, extending the recent solution of the two weight inequality for the Hilbert transform. As corollaries of this result we derive (1) a characterization of embedding measures for the model space $K_\vartheta$, for arbitrary inner function $\vartheta$, and (2) a characterization of the (essential) norm of composition operators mapping $K_\vartheta$ into a general class of Hardy and Bergman spaces.

CONTENTS

1. Introduction 1
2. Dyadic Grids, Good and Bad Decomposition 8
3. Necessary Conditions 12
4. Global to Local Reduction, I 20
5. Local Estimates, I 27
6. The Form $B_{\text{below}}$ 36
7. Elementary Estimates 37
References 41

1. INTRODUCTION

In this paper we characterize the boundedness for the Cauchy transform:

$$C_{\sigma}f(z) \equiv \int_{\mathbb{R}} \frac{f(t)}{z-t} \sigma(dt)$$

2000 Mathematics Subject Classification. Primary: 42B20 Secondary: 42B35.

Key words and phrases. Two weight inequalities, Cauchy transform, Riesz transform, Poisson operator, Carleson Measure, Model Space.

1. Research supported in part by National Science Foundation DMS grants # 0968499 and # 1265570, a grant from the Simons Foundation (#229596 to Michael Lacey), and the Australian Research Council through grant ARC-DP120100399.
2. Research supported in part by NSERC.
3. Research supported in part by the NSC, through grant NSC102-2115-M-008-015-MY2.
4. Research supported in part by a National Science Foundation DMS grants # 0901524 and # 1056965, MTM2010-16232, MTM2009-14694-C02-01 (Spain), and a Sloan Foundation Fellowship.
5. Research supported in part by a National Science Foundation DMS grant # 0955432.
as a map between $L^2(\mathbb{R}; \sigma)$ and $L^2(\mathbb{R}^+; \tau)$, where $\sigma$ and $\tau$ are two arbitrary weights, i.e. locally finite positive Borel measures. The characterization is in terms of a joint Poisson $A_2$ condition and a set of testing conditions.

We are motivated by the study of the model space $K_\vartheta = H^2 \ominus \vartheta H^2$, where $\vartheta$ is an inner function. These spaces are essential to the Nagy-Foias model for contractions on a Hilbert space. Function theoretic properties of the space $K_\vartheta$ are therefore of significant interest, and for the basics we point the reader to the text [22] and survey [23] and the references therein for a guide to this intricate literature.

In Theorem 1.10, we characterize the Carleson measures for $K_\vartheta$ spaces, a question posed by Cohn [6]. Previously, Aleksandrov [1] characterized isometric Carleson measures, and otherwise definitive results have only been proved in the so-called ‘one component’ case [6, 28].

We also characterize the norm of a composition operator from a $K_\vartheta$ space to any one of a general class of analytic function spaces, which include Hardy and the entire scale of Bergman spaces. See Theorem 1.11. The operator-theoretic properties of composition operators have been of intense interest for 60 years, see for instance the text [7], but the only result concerning composition operators on an arbitrary model space is the elegant characterization of compactness from $K_\vartheta$ to $H^2$ obtained by Lyubarskii-Malinnikova [16].

Our methods to attack the characterization question use real variable techniques, and so it is more convenient to describe the main results in terms of the Riesz transform. Let $R$ denote the one-dimensional Riesz transform acting on functions in the following manner. For $x \in \mathbb{R}^2_+$ and a signed measure $\nu$ on $\mathbb{R}$ we are interested in the family of $\mathbb{R}^2$-valued operators given by

$$R\nu(x) \equiv \int \frac{x - t}{|x - t|^2} \nu(dt), \quad x \in \mathbb{R}^2_+.$$ 

We write the coordinates of this operator as $(R^1, R^2)$. The second coordinate $R^2$ is the Poisson transform $P$, and the Cauchy transform is

$$C\nu \equiv R^1\nu + iR^2\nu.$$ 

Let $\sigma$ denote a weight on $\mathbb{R}$ and $\tau$ denote a measure on the upper half plane $\mathbb{R}^2_+$. Finding necessary and sufficient conditions on the pair of measures $\sigma$ and $\tau$ so that the estimate below holds is the two weight problem for the Riesz transform

$$\|R f\|_{L^2(\mathbb{R}^2_+; \tau)} \leq A \|f\|_{L^2(\mathbb{R}; \sigma)}.$$ 

This Theorem characterizes the two weight inequality, under restrictions on the supports of the two weights.

**Theorem 1.2.** Let $\sigma$ be a weight on $\mathbb{R}$ and $\tau$ a weight on the closed upper half-plane $\mathbb{R}^2_+$. The two weight inequality (1.1) holds if and only if these conditions hold uniformly over all intervals $I \subset \mathbb{R}$, and Carleson cubes $Q_I = I \times [0, |I|]$. An $A_2$ condition holds: For a finite positive constant $A_2$,

$$\frac{\tau(Q_I)}{|I|} \cdot \int_{\mathbb{R} \setminus I} \frac{|I|}{(|I| + \text{dist}(t, I))^2} \sigma(dt) \leq A_2,$$

$$\frac{\sigma(I)}{|I|} \cdot \int_{\mathbb{R}^2_+ \setminus Q_I} \frac{|I|}{(|I| + \text{dist}(x, Q_I))^2} \tau(dx) \leq A_2,$$
and, these testing inequalities hold: For a finite positive constant \( T \),

\[
\int_{Q_i} |R_\sigma 1_I(x)|^2 \tau(dx) \leq T^2 \sigma(I),
\]

(1.4)

\[
\int_I |R_\tau^* 1_{Q_i}(t)|^2 \sigma(dt) \leq T^2 \tau(Q_i).
\]

(1.5)

Moreover, if \( \mathcal{T} \) and \( \mathcal{A}_2 \) are the best constants in these inequalities, then \( \mathcal{N} \simeq \mathcal{A} \equiv \mathcal{A}_2^{1/2} + \mathcal{I} \).

Restricting \( \tau \) to be supported on \( \mathbb{R} \times \{0\} \) would reduce to the Hilbert transform case. Then the Theorem above is the foundational result of Lacey-Sawyer-Shen-Uriarte-Tuero [15] and Lacey [11], with the further refinements of Hytönen [9], answering a conjecture of Nazarov-Treil-Volberg [20, 29]. The necessity of the conditions above, given the norm condition are obvious for the testing inequalities (1.4)—(1.5), and the necessity of the \( A_2 \) condition is known. Thus the main content is the sufficiency of the conditions above for the norm inequality.

A version of the above theorem holds in a more general context. In any dimension, as long as one weight is supported on the real line, the Cauchy transform can be replaced by an arbitrary fractional Riesz transform. See [26].

1.1. Other Results. We now present applications of the main result. Because of the close connection with analytic functions and the Cauchy transform our applications are drawn from this area.

1.1.1. Setting on the Disk, Compactness. There are two forms of a Cauchy transform on the disk. For \( \sigma \) a weight on \( \mathbb{T} = \partial \mathbb{D} \), and \( f \in L^2(\mathbb{T}, \sigma) \), the transform could be either

\[
\int_{\mathbb{T}} \frac{f(w)}{z-w} \sigma(dA(w)), \quad \text{or} \quad \int_{\mathbb{T}} \frac{f(w)}{1-wz} \sigma(dA(w)).
\]

The two are unitarily equivalent, via the map \( f \mapsto f(w)/w \). We prefer the second formulation, and denote it by \( C_\sigma f(z) \). The main theorem, formulated on the disk, is below.

**Theorem 1.6.** Let \( \sigma \) be a weight on \( \mathbb{T} = \partial \mathbb{D} \), and \( \tau \) a weight on \( \overline{\mathbb{D}} \). The inequality below holds, for some finite positive \( \mathcal{C} \),

\[
\| C_\sigma \sigma f \|_{L^2(\mathbb{T}, \tau)} \leq \mathcal{C} \| f \|_{L^2(\mathbb{T}, \sigma)},
\]

if and only if these constants are finite: For the Poisson extension operator \( P \) on the disk,

\[
\sigma(\mathbb{T}) \cdot \tau(\overline{\mathbb{D}}) + \sup_{z \in \overline{\mathbb{D}}} \{ P(\sigma 1_{\mathbb{T} \setminus I})(z)P\tau(z) + P\sigma(z)P(\tau 1_{\overline{\mathbb{D}} \setminus B_I})(z) \} \equiv \mathcal{A}_2,
\]

(1.7)

\[
\sup_I \sigma(I)^{-1} \int_{B_I} |C_\sigma 1_I(z)|^2 \tau(dA(z)) \equiv \mathcal{T}^2,
\]

\[
\sup_I \tau(B_I)^{-1} \int_I |C_\tau^* 1_{B_I}(w)|^2 \sigma(dw) \equiv \mathcal{T}^2,
\]

where these conventions hold. The last two inequalities are uniform over all intervals \( I \subset \mathbb{T} \), with \( |I| \leq \frac{1}{2} \), \( B_I \equiv \{ z \in \overline{\mathbb{D}} \colon z = re^{i\theta}, |1-r| \leq |I|, e^{i\theta} \in I \} \) is the Carleson box over \( I \).

Finally, for the best constants \( \mathcal{C}, \mathcal{A}_2, \mathcal{T} \) in these inequalities, we have \( \mathcal{C} \simeq \mathcal{A}_2^{1/2} + \mathcal{T} \).
We are not aware of how to derive this theorem from the results on the upper half-plane, rather the proof must be repeated, taking into account a few minor complications. First, it is elementary to see that the kernel of the Cauchy transform satisfies

\[ \frac{2}{1 - \overline{z}w} = 1 + P_z(w) + iQ_z(w) \]

where the right side consists of the Poisson and conjugate Poisson kernels, namely for \( z \in \mathbb{D} \) and \( w \in \mathbb{T}, \)

\[ P_z(w) = \frac{1 - |z|^2}{|w - z|^2}, \quad Q_z(w) = \frac{2 \text{Im}(z \overline{w})}{|w - z|^2}. \]

The leading constant term on the right in (1.8) leads to the global term \( \sigma(\mathbb{T}) \tau(\mathbb{D}) \) in the \( A_2 \) condition (1.7). The necessity of the other conditions is then seen in a manner similar to the setting of the upper half-plane. In the sufficient condition, the main point is to prove boundedness of the transform

\[ \int_{\mathbb{T}} \{P_z(w) + iQ_z(w)\} f(w) \sigma(dw). \]

This is then in a form closely matching the vector Riesz transform.

One reduces to the case when the weight \( \sigma \) is supported on a fixed small arc of \( \mathbb{T}. \) This is accomplished by showing that a version of the weak-boundedness principle holds so that one can see that the testing inequalities still hold for the restricted weight \( \sigma. \) One should also reduce to the case of the weight \( \tau \) is supported in a narrow annulus close to the boundary of the arc that supports \( \sigma. \) The random grids on the circle \( \mathbb{T} \) are constructed by a rotation of the standard lattice on \( \mathbb{T}. \) The corresponding grid in \( \mathbb{D} \) is then constructed in the analogous manner, resulting in the standard Bergman tree. Modifications of energy and monotonicity are handled in an analogous fashion, with derivative calculations being done in radial and angular coordinates on relative to the arc. The remainder of the proof then follows similarly to what is done in the case discussed in detail in the rest of the paper. In the interest of brevity, we leave the modifications to the interested reader.

We now turn to compactness of the Cauchy transform.

**Theorem 1.9.** Under the assumptions of Theorem 1.6, the operator \( C_\sigma \) is compact if and only if \( C_\sigma : L^2(\mathbb{T}; \sigma) \to L^2(\mathbb{D}; \tau) \) is bounded and these conditions hold:

\[
\limsup_{r \uparrow 1} \sup_{|z| = r} \left\{ P(\sigma 1_{\mathbb{T}\setminus I})(z) P\tau(z) + P\sigma(z) P(\tau 1_{\mathbb{D}\setminus B_I})(z) \right\} = 0,
\]

\[
\limsup_{|I| < \epsilon} \sup_{|z| < r} \left| C_\sigma 1_I(z) \right|^2 \tau(dA(z)) = 0,
\]

\[
\limsup_{|I| < \epsilon} \left| C_\cdot^T(1_{B_I})(t) \right|^2 \sigma(dt) = 0.
\]

The details of the proof will not be given; they are a bit easier than those of [12, §9]. With this result in hand, one can state characterizations of compact analogs of the two theorems that follow.

1.1.2. **Carleson Measures for the Space \( K_\vartheta.** Let \( H^2(\mathbb{D}) \) denote the Hardy space of analytic functions on the unit disk \( \mathbb{D}. \) Let \( \vartheta \) be an inner function on \( \mathbb{D}, \) namely an analytic function
Moreover, we have that 

\[ \vartheta(z) = B_\Lambda(z) \exp \left( - \int_\mathbb{T} \frac{\xi + z}{\xi - z} \nu(d\xi) \right), \]

where \( B_\Lambda \) is a Blaschke product with zero set given by \( \Lambda \subset \mathbb{D} \) and \( \nu \) is a measure on \( \mathbb{T} \) singular with respect to Lebesgue measure. The space \( K_\vartheta \equiv H^2(\mathbb{D}) \ominus \vartheta H^2(\mathbb{D}) \) is called the model space associated to \( \vartheta \). Functions in this space admit an analytic continuation through the set \( \mathbb{T} \setminus \Sigma(\vartheta) \), where the latter set is the spectrum of \( \vartheta \), defined to be the closed set

\[ \Sigma(\vartheta) \equiv \text{clos}(\Lambda \cup \supp(\nu)) = \left\{ \zeta \in \mathbb{D} : \liminf_{z \to \zeta} \vartheta(z) = 0 \right\}. \]

Function theoretic properties of the space \( K_\vartheta \) are of significant interest, and we concentrate here on Carleson measures for the space. Recall that a measure \( \mu \) is a \( K_\vartheta \)-Carleson measure if we have the following estimate holding:

\[ \int_\mathbb{D} |f(z)|^2 d\mu(z) \leq C(\mu)^2 \|f\|_{K_\vartheta}^2 \quad \forall f \in K_\vartheta. \]

Since \( K_\vartheta \) is a subspace of \( H^2 \), every Carleson measure for \( H^2 \) is also one for \( K_\vartheta \), but its norm may be significantly smaller. And, a Carleson measure for \( K_\vartheta \) need not be one for \( H^2 \).

This problem has been intensely studied by numerous authors, with the question of characterization posed by Cohn [6] in 1982. An attractive special case when \( \vartheta \) satisfies the ‘one-component’, or ‘connected level set’ condition, namely that the enlargement of the spectrum, given by

\[ \Omega(\epsilon) \equiv \{ z \in \mathbb{D} : |\vartheta(z)| < \epsilon \}, \quad \epsilon > 0 \]

is connected for some \( \epsilon > 0 \). In this case, Cohn op. cit. and Treil and Volberg [28], showed that \( \mu \) is \( K_\vartheta \)-Carleson if and only if the Carleson condition \( \mu(B_I) \lesssim |I| \) holds for all intervals \( I \) such that the Carleson box \( B_I \) intersects \( \Omega(\epsilon) \). See also the alternate proof obtained by Aleksandrov in [2]. For more general \( \vartheta \), see however the counterexample of Nazarov-Volberg [21], based on the famous counterexample of Nazarov [19] to the Sarason conjecture. Apparently, there are very few results known for general \( \vartheta \), with one of these being the remarkable results of Aleksandrov [1] characterizing those \( \mu \) for which \( K_\vartheta \) isometrically embeds into \( L^2(\mathbb{D}; \mu) \), under the natural embedding map.

In the case when the measure \( \mu \) is supported on \( \mathbb{T} \) a characterization of the \( K_\vartheta \)-Carleson measures is a corollary of the two weight Hilbert inequality obtained in [11, 12]. However, for measures with more general supports, we need the full characterization obtained in this paper. We now translate this problem to one about weighted estimates for the Cauchy transform by following the exposition of Nazarov and Volberg [21]. Let \( \sigma \) denote the Clark measure on \( \mathbb{T} \). Then we have that \( L^2(\mathbb{T}; \sigma) \) is unitarily equivalent to \( K_\vartheta \) via a unitary \( U \). Moreover, we have that \( U^* : L^2(\mathbb{T}; \sigma) \to K_\vartheta \) has the integral representation given by

\[ U^* f(z) \equiv (1 - \vartheta(z)) \int_\mathbb{T} \frac{f(\xi)}{1 - \xi z} \sigma(d\xi). \]

Note that \( U^* \) is, up to a multiplication operator, the Cauchy transform \( C \) of \( f \) with respect to the measure \( \sigma \). For the inner function \( \vartheta \) and measure \( \mu \), define a new measure \( \nu_{\vartheta, \mu} \equiv |1 - \vartheta|^2 \mu \).
Then we have that $\mu$ is a Carleson measure for $K_\varphi$ if and only if $C : L^2(\mathbb{T}; \sigma) \to L^2(\mathbb{D}; \nu_{\varphi, \mu})$ is bounded. Indeed, suppose that $C : L^2(\mathbb{T}; \sigma) \to L^2(\mathbb{D}; \nu_{\varphi, \mu})$ is bounded. Then we have that

$$\int_{\mathbb{T}} \left| U^* f(z) \right|^2 d\mu(z) = \int_{\mathbb{T}} \left| (1 - \vartheta(z)) C(f\sigma)(z) \right|^2 d\mu(z) = \int_{\mathbb{T}} \left| C(f\sigma)(z) \right|^2 d\nu_{\varphi, \mu}(z) \leq C^2 \|f\|^2_{L^2(\mathbb{T}; \sigma)}.$$ 

For $g \in K_\varphi$ let $f = U_g$, then the above inequality gives

$$\int_{\mathbb{T}} |g(z)|^2 d\mu(z) \leq C^2 \|Uf\|^2_{L^2(\mathbb{T}; \sigma)} = \|g\|^2_{K_\varphi}.$$ 

Thus, we have that $C(\mu)^2 \leq C^2$. However, this argument is completely reversible, and we in fact arrive at $C(\mu) = C$. So understanding the Carleson measures for $K_\varphi$ is equivalent to deducing the boundedness of $C : L^2(\mathbb{T}; \sigma) \to L^2(\mathbb{D}; \nu_{\varphi, \mu})$ (a similar argument applies to the Hardy space of the upper half plane $\mathbb{C}_+$. Our characterization of the Carleson measures for $K_\varphi$ is given by the following theorem.

**Theorem 1.10.** Let $\mu$ be a non-negative Borel measure supported on $\overline{\mathbb{D}}$ and let $\vartheta$ be an inner function on $\mathbb{D}$ with Clark measure $\sigma$. Set $\nu_{\mu, \vartheta} = |1 - \vartheta|^2 \mu$. The following are equivalent:

(i) $\mu$ is a Carleson measure for $K_\varphi$, namely,

$$\int_{\mathbb{T}} |f(z)|^2 d\mu(z) \leq C(\mu)^2 \|f\|^2_{K_\varphi} \quad \forall f \in K_\varphi;$$

(ii) The Cauchy transform $C$ is a bounded map between $L^2(\mathbb{T}; \sigma)$ and $L^2(\overline{\mathbb{D}}; \nu_{\mu, \vartheta})$, i.e., $C : L^2(\mathbb{T}; \sigma) \to L^2(\overline{\mathbb{D}}; \nu_{\mu, \vartheta})$ is bounded;

(iii) The three conditions in (1.7) hold for the pair of measures $\sigma$ and $\nu_{\mu, \vartheta}$. Moreover,

$$C(\mu) \simeq \|C\|_{L^2(\mathbb{T}; \sigma) \to L^2(\overline{\mathbb{D}}; \nu_{\mu, \vartheta})} \simeq A_2^{1/2} + \mathcal{F}.$$ 

The equivalence between (i) and (ii) is sketched before the theorem, while the equivalence between (ii) and (iii) follows from interpreting Theorem 1.6 in the context at hand. The Clark measure $\sigma$ is a singular measure supported on the set $\{z : \vartheta(z) = 1\}$, hence it and $\nu_{\mu, \vartheta} = |1 - \vartheta|^2 \mu$ do not have common point masses, so that the $A_2$ conditions could be phrased more simply.

There is a variant of Theorem 1.9 that holds, characterizing those measures $\mu$ such that $K_\varphi$ embeds compactly into $L^2(\overline{\mathbb{D}}; \mu)$. The interested reader can combine Theorems 1.10 and 1.9 to formulate it. Variants of these results hold with the disk replaced by $\mathbb{C}_+$, and we again leave the details to the reader.

1.1.3. **Composition Operators on $K_\varphi$.** Let $\varphi : \mathbb{D} \to \mathbb{D}$ be holomorphic. The composition operator with symbol $\varphi$ is $C_\varphi f = f \circ \varphi$. The Littlewood subordination principle implies that these operators are bounded from $H^2(\mathbb{D})$ to $H^2(\mathbb{D})$, and to the Bergman spaces $A_\alpha(\mathbb{D})$ given by the norm

$$\|f\|_{A_\alpha(\mathbb{D})}^2 \equiv \int_{\mathbb{D}} |f(z)|^2 (1 - |z|^2)^\alpha dA(z), \quad -1 < \alpha < \infty.$$ 

However, the upper bound supplied by the subordination argument will not be sharp in general. The subject at hand is to describe finer operator theoretic properties in terms of the properties of $\varphi$. Since $K_\varphi$ is a subspace of $H^2$, it follows that $C_\varphi$ is bounded as a map from $K_\varphi$ to $H^2$, or any of the Bergman spaces above. In this setting, we can characterize three properties of $C_\varphi$: the norm of $C_\varphi$, its essential norm, and compactness. There is an
Let \( \tau \) be a weight on \( \overline{\mathbb{D}} \), and define a Hilbert space of analytic functions by taking the closure of \( H^\infty(\mathbb{D}) \) with respect to the norm for \( L^2(\overline{\mathbb{D}}; \tau) \). Call the resulting space \( H^2_\tau \). Thus, if \( \tau \) is Lebesgue measure on \( \mathbb{T} \), the space \( H^2_\tau \) is the Hardy space, and if \( \tau(dA(z)) = (1-|z|^2)^\sigma dA(z) \), it is the Bergman space.

To the function \( \varphi \) and weight \( \tau \) we associate the pullback measure \( \tau_\varphi \) defined as a measure on \( \overline{\mathbb{D}} \), as \( \tau_\varphi(Eices) = \tau(\varphi^{-1}(E)) \). Then

\[
\|C_\varphi f\|_{H^2_\tau}^2 = \int_{\overline{\mathbb{D}}} |f \circ \varphi(z)|^2 \tau(dA(z)) = \int_{\overline{\mathbb{D}}} |f(z)|^2 \tau_\varphi(dA(z)).
\]

That is, \( C_\varphi : K_\vartheta \rightarrow H^2_\tau \) is unitarily equivalent to the embedding operator \( I_{\tau_\varphi} : K_\vartheta \rightarrow L^2(\overline{\mathbb{D}}; \tau_\varphi) \). Thus, we have that the boundedness of the composition operator \( C_\varphi : K_\vartheta \rightarrow H^2_\tau(\mathbb{D}) \) is equivalent to determining when \( \tau_\varphi \) is a Carleson measure for \( K_\vartheta \). By Theorem 1.6 we have the following answer.

**Theorem 1.11.** Let \( \vartheta \) be an inner function. Let \( \varphi : \mathbb{D} \rightarrow \mathbb{D} \) be analytic and let \( \tau_\varphi \) denote the pullback measure associated to \( \varphi \). The following are equivalent:

1. \( C_\varphi : K_\vartheta \rightarrow H^2_\tau \) is bounded;
2. \( \tau_\varphi \) is a Carleson measure for \( K_\vartheta \), namely,
   \[
   \int_{\overline{\mathbb{D}}} |f(z)|^2 \tau_\varphi(dA(z)) \leq C(\tau_\varphi)^2 \|f\|^2_{K_\vartheta} \quad \forall f \in K_\vartheta;
   \]
3. The conditions of Theorem 1.6 (iii) hold for the pair of weights \( \sigma \) on \( \mathbb{T} \) and \( \nu_{\tau_\varphi, \vartheta} = |1-\vartheta|^2 \tau_\varphi \).

A corresponding characterization of compactness can be obtained, in terms of the limits in Theorem 1.9 being zero. If one is interested in the essential norm, the limits in Theorem 1.9 should be taken to be limit superiors. A theorem of this level of generality is new even when \( K_\vartheta \) is replaced by the Hardy space \( H^2 \).

In the setting of composition operators from \( H^2 \) to \( H^2 \), MacCluer [17] characterized the compact operators in terms of the the measure \( |\mathbb{T} \cap \varphi^{-1}(E)| \) being a vanishing Carleson measure. Shapiro [27] calculates the essential norm of the same operators in terms of the Nevanlinna counting function \( N_\varphi \). Specializing his result to compactness, the characterization states that \( N_\varphi(w) = o(1-|w|) \) as \( |w| \rightarrow 1 \). The connection between the two approaches is analyzed in Lefèvre-Li-Queffélec [10]. In the setting of the Theorem above, one weight is the pullback measure, and the other is Lebesgue measure on the unit circle. If the pullback measure is a (vanishing) Carleson measure the conditions above can be verified by ad hoc means. The survey [23] includes additional points of view and references related to Shapiro’s results.

In the setting of composition operators from \( K_\vartheta \) to \( H^2 \), there is an elegant characterization of compactness due to Lyubarskii-Malinnikova [16, Theorem 1] expressed in terms of the Nevanlinna counting function of \( \vartheta \), namely that

\[
N_\varphi(w) \frac{1-|\vartheta(w)|^2}{1-|w|^2} = o(1-|w|), \quad |w| \rightarrow 1.
\]
Roughly speaking, this condition only imposes the Shapiro condition as \( w \) approaches the spectrum \( \Sigma(\vartheta) \subset T \). We are not aware any other results at this level of generality for composition operators on \( K_\vartheta \) spaces.

1.2. Proof and Organization. We concern ourselves with the proof of Theorem 1.2. The testing inequalities (1.4) and (1.5) are obviously necessary, and we provide the known argument for necessity of the \( A_2 \) condition (1.3) below. The bulk of the argument concerns the sufficiency of the \( A_2 \) condition and testing conditions for the norm estimate. For this we follow the model of the argument derived from the beautiful strategy of Nazarov-Treil-Volberg [20]. This strategy works however for any Calderón-Zygmund operator, whereas delicate properties of the operator at hand must inform the proof. These additional elaborations were provided for the Hilbert transform in [9,11,14,15].

Central here is the notion of monotonicity and energy inequalities, which control subtle off-diagonal terms in the proof. These conditions are asymmetric with respect to the role of the weights, and so these two conditions are different in the current setting. One of them involves geometric arguments that are not present in the setting of the Hilbert transform, and we present that case first below. The other uses both components of the Riesz transform in order to control only part of the energy. Both versions of the energy inequality require more sophisticated formulations than those for the Hilbert transform.

Following the development of the energy inequalities, they must be bootstrapped to more complicated inequalities, in two different (highly non-obvious) ways. In addition, one must incorporate stopping data from the functions on which one is testing the norm of the Cauchy transform. Again the arguments are asymmetric with respect to the weights, but share many commonalities with the case of the Hilbert transform. One of these bootstrapped inequalities, the functional energy inequalities, does not require much adjustment, but the second size lemma, requires a more careful analysis, due to the more sophisticated formulation of the energy inequality. Accordingly, we present only the more novel of the two cases in full. The first of these bootstrapped results is a functional form of the energy inequality, deduced from a two weight inequality for a Bergman-like operator. It is used to reduce the global inequality to a series of local inequalities. The second bootstrapped result is a subtle recursion to prove the local inequality. Here, the distinctions between the energy inequality and its simpler variants for the Hilbert transform creates additional complications in the proof.

The next section has the standard random dyadic grid construction of [20]; section §3 is the essence of the matter, deriving the energy inequalities. Following that, the more robust parts of the argument are presented, with §4 focusing on the global to local reductions, §5 studying the local estimates, §6 studying one of the bilinear forms that arise in §4. Finally, in §7 we focus on the elementary estimates for bilinear forms arising from the analysis in the proof.

2. Dyadic Grids, Good and Bad Decomposition

Let \( \mathcal{D} \) denote the standard dyadic grid in \( \mathbb{R} \). A random dyadic grid \( \mathcal{D} \) is specified by \( \xi \in \{0,1\}^\mathbb{Z} \) and choice of \( 1 \leq \lambda \leq 2 \). The elements of \( \mathcal{D} \) are given by

\[
I \equiv \hat{I} + \xi = \lambda \left\{ \hat{I} + \sum_{n : 2^{-n} < |\hat{I}|} 2^{-n} \xi_n \right\}.
\]
Place the uniform probability measure $\mathbb{P}$ on $\xi \in \{0,1\}^\mathbb{Z}$, and choose $\lambda$ with respect to normalized measure on $[1,2]$ with measure $\frac{d\lambda}{\lambda}$.

Fix $0 < \epsilon < 1$ and $r \in \mathbb{N}$. An interval $I \in \mathcal{D}$ is said to be $(\epsilon, r)$-bad if there is an interval $J \in \mathcal{D}$ such that $|J| > 2^r|I|$ and $\text{dist}(I, \partial J) < |J|^\epsilon |J|^{1-\epsilon}$. Otherwise, an interval $I$ will be called $(\epsilon, r)$-good. We have the following well-known properties associated to the random dyadic grid $\mathcal{D}$.

**Proposition 2.1.** The following properties hold:

1. The property of $I = \hat{I} + \xi$ being $(\epsilon, r)$-good depends only on $\xi$ and $|I|$;
2. $p_{\text{good}} \equiv \mathbb{P}(I \text{ is } (\epsilon, r) \text{-good})$ is independent of $I$;
3. $p_{\text{bad}} \equiv 1 - p_{\text{good}} \lesssim \epsilon^{-1}2^{-cr}$.

We now indicate how, associated to a dyadic lattice $\mathcal{D}$ on $\mathbb{R}$, we create a “dyadic lattice” on $\mathbb{R}_+^2$. For any interval $I \in \mathcal{D}$, we define $Q_I$, the Carleson square over the interval $I$, as the set

$$Q_I \equiv I \times [0,|I|].$$

We then set $\mathcal{D}_+ \equiv \{Q_I : I \in \mathcal{D}\}$. $\mathcal{D}_+$ will denote the Carleson cubes associated to the standard dyadic lattice $\mathcal{D}$ on $\mathbb{R}$. Note that if we write $Q \in \mathcal{D}_+$, then there is a corresponding $I \in \mathcal{D}$ such that $Q = Q_I$. For a cube $Q$ we let $\ell(Q)$ denote the side length of the cube, i.e, for $Q = Q_I$, we have that $\ell(Q) = \ell(Q_I) = |I| = |Q|^\frac{1}{2}$.

The collection $\mathcal{D}_+$ is extended to a dyadic grid $\mathcal{D}^2$ on $\mathbb{R}_+^2$, defined to be all cubes of the form $I \times |I|((0,1)+n)$, for $n \in \mathbb{N}$. Similar to above, we have the following notion of good and bad cubes. Fix $0 < \epsilon < 1$ and $r \in \mathbb{N}$. A cube $Q \in \mathcal{D}^2$ is said to be $(\epsilon, r)$-bad if there is a cube $Q' \in \mathcal{D}^2$ such that $\ell(Q') > 2^r\ell(Q)$ and $\text{dist}(Q, \partial Q') < \ell(Q)^\epsilon \ell(Q')^{1-\epsilon}$. Otherwise, a cube $Q$ will be called $(\epsilon, r)$-good. Similar to Proposition 2.1 we have the following properties holding.

**Proposition 2.2.** The following properties hold:

1. The property of $Q = \hat{Q} + (\xi,0)$ being $(\epsilon, r)$-good depends only on $\xi$ and $\ell(Q)$;
2. $p_{\text{good}} \equiv \mathbb{P}(Q \text{ is } (\epsilon, r) \text{-good})$ is independent of $Q$;
3. $p_{\text{bad}} \equiv 1 - p_{\text{good}} \lesssim \epsilon^{-1}2^{-cr}$.

We introduce the Haar basis adapted to the weights $\sigma$ on $\mathbb{R}$ and $\tau$ on $\mathbb{R}_+^2$, and we do so with the tacit assumption that $\sigma$ assigns positive measure to each dyadic interval and $\tau$ to each dyadic cube. While this is not true in general, standard modifications handle the general case. To avoid cumbersome notation later on, for a set $E$ we will identify the set with its indicator function, i.e. $E \equiv 1_E$. For $I \in \mathcal{D}$, if $\sigma$ assigns positive mass to both children of $I$, define

$$h_I^\sigma \equiv \sqrt{\frac{\sigma(I_+)\sigma(I_-)}{\sigma(I)}} \left( \frac{I_+}{\sigma(I_+)} - \frac{I_-}{\sigma(I_-)} \right).$$

Otherwise, set $h_I^\sigma = 0$. Note that this is a $L^2(\mathbb{R};\sigma)$ normalized function and has integral 0 with respect to $\sigma$. Also $\{h_I^\sigma : I \in \mathcal{D}, h_I^\sigma \neq 0\}$ is an orthonormal basis of $L^2(\mathbb{R};\sigma)$. We will also let $\hat{f}_\sigma(I) \equiv \langle f, h_I^\sigma \rangle_\sigma$ and will let

$$\Delta_I^\sigma f \equiv \langle f, h_I^\sigma \rangle_\sigma h_I^\sigma = I_+E_{I_+}^\sigma f + I_-E_{I_-}^\sigma f - IE_{I}^\sigma f.$$
Otherwise, we set $\Delta^\sigma f \equiv 0$. We have the following identity from this formula

$$E^\sigma f = \sum_{I \subset J} E^\sigma_I \Delta^\sigma_I f.$$ 

We next discuss the Haar basis on $\mathbb{R}^2_+$. Given a cube $Q \in \mathcal{D}^2$, with $\tau(Q') > 0$, for at least two children $Q'$ of $Q$, set

$$\Delta_Q^\tau g \equiv \sum_{Q' \text{ child of } Q, \tau(Q') > 0} E^\tau_{Q'} g \cdot Q' - E^\tau_Q g \cdot Q.$$ 

This is the standard martingale difference.

Say that $\mathcal{D}$ is admissible for $\sigma$ and $\tau$ if and only if $\sigma$ does not have a point mass at the endpoint of any interval $I \in \mathcal{D}$ and $\tau$ does not assign positive mass to the boundary of any $Q \in \mathcal{D}^2$. With probability one the random grid $\mathcal{D}$ is admissible, and we always assume this below. This is so, due to the incorporation of the dilation factor into the definition of the random grid, though throughout, we will assume that the dilation factor is 1. Thus, with probability one, we can define the Haar bases $\{\Delta^\sigma_Q\}_{Q \in \mathcal{D}^2}$ and $\{h^\sigma_I\}_{I \in \mathcal{D}}$ as above. Now, write the identity operator in $L^2(\mathbb{R}; \sigma)$ as

$$f = P^\sigma_{\text{good}} f + P^\sigma_{\text{bad}} f \quad \text{where } P^\sigma_{\text{good}} f \equiv \sum_{I \in \mathcal{D} : I \text{ is } (\epsilon,\tau) - \text{good}} \Delta^\sigma_I f.$$ 

Similar notation applies for the identity operator on $L^2(\mathbb{R}^2_+; \tau)$. For the remainder of the paper we let $\| \cdot \|_{\sigma}$ and $\langle \cdot, \cdot \rangle_{\sigma}$ denote the norm and inner product in $L^2(\mathbb{R}; \sigma)$. Identical notation applies for $\| \cdot \|_{\tau}$ and $\langle \cdot, \cdot \rangle_{\tau}$ in $L^2(\mathbb{R}^2_+; \tau)$.

We have the following proposition.

**Proposition 2.3.** The following estimate holds:

$$E\|P^\sigma_{\text{bad}} f\|_{\sigma}^2 \lesssim \epsilon^{-1} 2^{-e\tau \|f\|_{\sigma}^2}.$$ 

An identical estimate is true for the function $g$ and the weight $\tau$.

Using the ideas of good and bad dyadic cubes in non-homogeneous harmonic analysis, one can deduce that it suffices to study only good functions. This reduction is very familiar, and its requirements well-known. To make this reduction we will make these standing assumptions: First, that there is an a priori inequality

$$\|Rf\|_{\tau} \leq N \|f\|_{\sigma}.$$ 

And, second, for any pair of functions $f \in L^2(\mathbb{R}, \sigma)$ and $g \in L^2(\mathbb{R}^2_+, \tau)$ with disjoint compact supports, there holds

$$(2.4) \quad \langle Rf, g \rangle_{\tau} = \int_{\mathbb{R}} \int_{\mathbb{R}^2_+} f(y) g(x) \frac{x - y}{|x - y|^2} \tau(dx) \sigma(dy).$$ 

We will refer to this as the canonical value of $\langle Rf, g \rangle_{\tau}$.

This lemma lets one reduce to only considering good functions in the bilinear form associated to $Rf$. 

Lemma 2.5. Let $\sigma$ and $\tau$ be a pair of weights for which the a priori inequality (1.1) holds, and $\mathcal{R} \equiv A^{1/2}_2 + \mathcal{T}$ is finite. Suppose that for a choice of $0 < \varepsilon < 1$, all $r \in \mathbb{N}$, and all admissible dyadic grids $\mathcal{D}$ and $\mathcal{D}^2$

\begin{equation}
(2.6) \quad \left| \langle R_{\sigma}(P_{\text{good}}^r f), P_{\text{good}}^r g \rangle_\tau \right| \leq C_{\varepsilon, r} \mathcal{R} \|f\|_\sigma \|g\|_\tau.
\end{equation}

Then (1.1) holds with $\mathcal{N} \lesssim \mathcal{R}$.

Proof of Lemma 2.5. Let $f \in L^2(\mathbb{R}; \sigma)$ and $g \in L^2(\mathbb{R}^2_+; \tau)$ be arbitrary functions. Then we have

\[ \left| \langle R_{\sigma} f, g \rangle_\tau \right| \leq \left| \langle R_{\sigma} P_{\text{good}}^r f, P_{\text{good}}^r g \rangle_\tau \right| + \left| \langle R_{\sigma} P_{\text{good}}^r f, P_{\text{bad}}^r g \rangle_\tau \right| + \left| \langle R_{\sigma} P_{\text{bad}}^r f, g \rangle_\tau \right|. \]

Now using (2.6) and the a priori inequality (1.1),

\[ \left| \langle R_{\sigma} f, g \rangle_\tau \right| \leq C_{r, \mathcal{R}} \|f\|_\sigma \|g\|_\tau + \mathcal{N} (\|f\|_\sigma \|P_{\text{bad}}^r g\|_\tau + \|g\|_\tau \|P_{\text{bad}}^r f\|_\tau). \]

Taking expectation over the choice of random grid $\mathcal{D}$ and $\mathcal{D}^2$ we have by Proposition 2.3 that

\[ \left| \langle R_{\sigma} f, g \rangle_\tau \right| \lesssim C_{r, \mathcal{R}} \|f\|_\sigma \|g\|_\tau + \varepsilon^{-1} 2^{-r} \mathcal{N} \|f\|_\sigma \|g\|_\tau. \]

But, for appropriate selection of $f \in L^2(\mathbb{R}; \sigma)$ and $g \in L^2(\mathbb{R}^2_+; \tau)$ we have

\[ \|f\|_\sigma \|g\|_\tau \mathcal{N} \lesssim \left| \langle R_{\sigma} f, g \rangle_\tau \right| \lesssim C_{r, \mathcal{R}} \|f\|_\sigma \|g\|_\tau + \varepsilon^{-1} 2^{-r} \mathcal{N} \|f\|_\sigma \|g\|_\tau. \]

Choosing $r$ sufficiently large for a given $\varepsilon$ then lets one conclude that $\mathcal{N} \lesssim \mathcal{R}$.

One way to gain the a priori inequality is to impose a standard truncation on the singular integral. They are defined as follows, for all $0 < \alpha < \beta$

\[ R_{\alpha, \beta}(\sigma f)(x) \equiv \int_{\mathbb{R}} K_{\alpha, \beta}(x, t) f(t) \sigma(dt), \]

where \[ |x - t| \cdot |K_{\alpha, \beta}(x, t)| + |x - t|^2 \cdot |\nabla K_{\alpha, \beta}(x, t)| \leq C 1_{[a/2, 2\beta]}(|t|), \]

and \[ K_{\alpha, \beta}(x, t) = \frac{x - t}{|x - t|^2}, \quad \text{if } \alpha < |t| < \beta, \ t \in \mathbb{R}^2_+. \]

Thus, the choice of kernel $K_{\alpha, \beta}$ is compactly supported, satisfies a size and gradient condition. Finally, it agrees with the Riesz transform kernel for most values where it is not zero. With a choice of standard truncations one can then define a uniform norm over certain truncations to be the best constant $\mathcal{N}_{\alpha_0, \beta_0}$ in the inequality below, in which $0 < \alpha_0 < \beta_0$

\[ \sup_{\alpha_0 < \alpha < \beta < \beta_0} \|R_{\alpha, \beta}(\sigma f)\|_\tau \leq \mathcal{N}_{\alpha_0, \beta_0} \|f\|_\sigma. \]

It is elementary to see that $\mathcal{N}_{\alpha_0, \beta_0} \leq C_{\alpha_0, \beta_0} A_2$. That is, assuming the $A_2$ condition, there is always an a priori inequality for standard truncations. Moreover, under the assumptions of the main theorem, one can show that the limit below will be finite:

\[ \lim_{\alpha \downarrow 0} \lim_{\beta \uparrow \infty} \left| \langle R_{\alpha, \beta}(\sigma f), g \rangle_\tau \right|, \quad f \in L^2(\mathbb{R}, \sigma), \ g \in L^2(\mathbb{R}^2_+, \tau). \]

This is left to the reader, as the additional complications needed to prove this do not require any ideas that go beyond the scope of this paper.
3. Necessary Conditions

We begin with some conventions.

- For two dyadic intervals $I, J$ and integer $s$ we write $J \subset_s I$, and say ‘$J$ is $s$-strongly contained in $I$’ if $J \subset I$ and $2^s|J| \leq |I|$. We are interested in the cases of $s = r$, the integer associated with goodness, and $s = 4r$. The value of $4r$ is used in §4, and the value of $r$ is used in this section and §5.
- The center of interval the $J$ is denoted $t_J$ and $x_Q$ is the center of cube $Q$. The Poisson average at interval $I$ is $P(f, I) \equiv P(f, |I|)$, and we frequently appeal to the approximation

$$P(f, I) = \int_{\mathbb{R}} \frac{|I|}{|I|^2 + \text{dist}(t, I)^2} f(t) \, dt \simeq \int_{\mathbb{R}} \frac{|I|}{(|I| + \text{dist}(t, I))^2} f(t) \, dt.$$ 

The same approximation holds for the Poisson average on $\mathbb{R}^2_+$.
- We are working with operators that carry $L^2(\mathbb{R}; \sigma)$ into $L^2(\mathbb{R}_+^2; \tau)$, as well as their duals. We will use $f, \phi \in L^2(\mathbb{R}; \sigma)$ and $g, \varphi \in L^2(\mathbb{R}_+^2; \tau)$ to denote the functions being acted on by the operators in question.
- There is opportunity for confusion about ranges of integration. Generically, we will have the integration variable $t \in \mathbb{R}$, and $x \in \mathbb{R}^2$. But, if we write $x - t$, we are viewing $t \in \mathbb{R} \subset \mathbb{R}^2$, with the natural inclusion. If the role of the dimensions are important for $x$, we will write $x = (x_1, x_2)$ or $z = x_1 + ix_2$ as integrating variables.
- Letters in sans-serif denote operators, for example $\mathcal{R}$ denotes the Riesz transforms, and we will have other related variants, such as the Poisson and a Bergman-like operator.
- The ‘hard’ case is the analysis of the operators in the direction from $L^2(\mathbb{R}_+^2; \tau)$ into $L^2(\mathbb{R}; \sigma)$. This case is treated first, followed by the reverse direction. The two cases are not dual. Herein, we develop the monotonicity principle, and the energy principle, both in parts I and II.
- For a dyadic interval $I \in \mathcal{D}$, we set $\pi I$ to be its parent in $\mathcal{D}$, the minimal interval in $\mathcal{D}$ which strictly contains $I$. For a dyadic subtree $\mathcal{F} \subset \mathcal{D}$, we set $\pi_{\mathcal{F}} I$ to be the minimal element $F \in \mathcal{F}$ with $I \subset F$ (so $\pi_{\mathcal{F}} F = F$). At a key point, we will use the notation $\pi_{\mathcal{F}}^2 I$ to be the minimal element of $\mathcal{F}$ which strictly contains $\pi_{\mathcal{F}} I$ (it is the $\mathcal{F}$-grandparent of $I$).

3.1. $A_2$ Condition. We show that the assumed norm inequality implies the $A_2$ condition.

**Proposition 3.1.** There holds $\mathcal{A}_2 \lesssim \mathcal{N}^2$.

**Proof.** Fix an interval $I$, and note that

$$|\mathcal{R}^* Q_I(t)| \gtrsim \frac{\tau(Q_I)}{|I| + \text{dist}(t, I)}, \quad t \notin I.$$ 

Squaring, and integrating against the measure $\sigma$, the assumed norm inequality gives us

$$\frac{\tau(Q_I)^2}{|I|} \int_{\mathbb{R} \setminus I} \frac{|I|}{(|I| + \text{dist}(t, I))^2} \sigma(dt) \lesssim \mathcal{N}^2 \tau(Q_I).$$

Dividing out by $\tau(Q_I)$ will complete the first half of the $A_2$ bound.
For the second half, note that
\[ |R_\sigma I(x)| \gtrsim \frac{|\sigma(I)|}{|I| + \text{dist}(x, Q_I)}, \quad x \notin Q_I. \]

And then the argument is completed as before.

\[ \square \]

3.2. Monotonicity, I. We begin our discussion of the critical off-diagonal considerations, where we will use the condition
\[ \left| \nabla^j \frac{x - t}{|x - t|^2} \right| \lesssim \frac{1}{|x - t|^{1+j}}, \quad j = 1, 2. \]

Monotonicity refers to the domination of off-diagonal inner products by positive operators. In this direction, we can establish the monotonicity principle, using the operators and sets given by
\[ T_\tau g(x) = \int_{\mathbb{R}^2_+} \frac{g(y) \cdot y}{y_2 + (y_1 - x_1)^2 + x_2^2} \tau(dy), \]
\[ V_I = \bigcup_{t \in I} \{ x = (x_1, x_2) : |x_1 - t| < x_2 \}. \]

Lemma 3.4. [Monotonicity, I] If \( \varphi \in L^2(\mathbb{R}^2_+; \tau) \) is non-negative and compactly supported on the complement of the set \( V_I \), and \( 10J \subset I \), then
\[ |\langle R_\tau^* \varphi, h^\sigma_{J'} \rangle_\sigma| \lesssim T_\tau \varphi(x_{Q_J}) \cdot \left\langle t, h^\sigma_{J'} \right\rangle_\sigma, \quad J' \subset J. \]

Assume that \( \varphi \in L^2(\mathbb{R}^2_+; \tau) \) is supported on the complement of the set \( Q_I \), \( 10J \subset I \) and \( f \in L^2(\mathbb{R}; \sigma) \), supported on \( J \), has \( \sigma \)-integral zero. Then these two estimates hold:
\[ |\langle R_\tau^* \varphi, h^\sigma_{J'} \rangle_\sigma| \lesssim T_\tau |\varphi|(x_{Q_J}) \cdot \left\langle t, h^\sigma_{J'} \right\rangle_\sigma, \quad J' \subset J, \]
\[ |\langle R_\tau^* \varphi, f \rangle_\sigma| \lesssim T_\tau |\varphi|(x_{Q_J}) \cdot \int_J |f(t)| \sigma(dt). \]

Proof. We are in a situation where (2.4) holds. The basic property is that
\[ \langle t, h^\sigma_{J'} \rangle_\sigma = \int_J (t - t_J) h^\sigma_{J'}(t) \sigma(dt) \]
and that the integrand is non-negative. From this, and a standard application of the kernel estimates in (3.2), the estimates (3.6) and (3.7) follow.

We turn to the critical equivalence (3.5). From the integral zero property of the Haar function, we have the equality below:
\[
\langle R^*_t \varphi, h^*_J \rangle_\sigma = \int \int_{\mathbb{R}^2_+ \setminus V_I} \varphi(x) h^*_J(t) \frac{x - t}{|x - t|^2} \sigma(dt) \tau(dx) \\
= \int \int_{\mathbb{R}^2_+ \setminus V_I} \varphi(x) h^*_J(t) \left\{ \frac{x - t}{|x - t|^2} - \frac{x - t_J}{|x - t_J|^2} \right\} \sigma(dt) \tau(dx).
\]

Now, the first, and main, point is that the first coordinate of the term in braces above satisfies for all \( t \in J \), and \( x \in \mathbb{R}^2_+ \setminus V_I \),
\[
(3.9) \quad \text{sgn}(t - t_J) \left\{ \frac{x_1 - t}{|x - t|^2} - \frac{x_1 - t_J}{|x - t_J|^2} \right\} \geq c \frac{|t - t_J|}{|J|^2 + \text{dist}(x_1, J)^2}.
\]
Indeed, for \( x \notin V_I \), we have \(|x_1 - t| \geq x_2\). And, therefore, it is simple to see that the difference below satisfies (3.9), in which we change the denominator of the first fraction,
\[
\text{sgn}(t - t_J) \left\{ \frac{x_1 - t}{|x - t|^2} - \frac{x_1 - t_J}{|x - t_J|^2} \right\}.
\]
Thus, it remains to see, after a simple derivative calculation, that
\[
\left| \frac{x_1 - t}{|x - t_J|^2} - \frac{x_1 - t_J}{|x - t_J|^2} \right| \leq C \frac{|t - t_J|}{|x_1 - t_J|^3} \\
\lesssim 2^{-r} \frac{|t - t_J|}{|J|^2 + \text{dist}(x_1, J)^2}.
\]
Since we are free to take the integer \( r \) as large as we wish, this completes the proof of (3.9). Combining (3.8) and (3.9) will complete the proof of (3.5).

3.3. **Energy Inequality, I.** Essential for the control of the off-diagonal terms is the energy inequality. We need two definitions. For a interval \( I \), set the energy of \( I \) to be
\[
(3.10) \quad E(\sigma, I)^2 \equiv \sigma(I)^{-1} \sum_{J: \sigma(J) \leq \sigma(I)} \left\langle \frac{t}{|I|}, h^*_J \right\rangle_\sigma^2.
\]
Since the Haar functions have mean value zero with respect to \( \sigma \), we are free to replace \( t \) above by \( t - t_I \), so that \( E(\sigma, I)_I \lesssim 1 \).

Define \( WI \) to be the partition of the interval \( I \) consisting of those maximal intervals \( K \subset I \) with \( \text{dist}(K, \partial I) \geq |K|^{\epsilon} |I|^{1-\epsilon} \). So these are the maximal intervals that are ‘good with respect to \( I \.’ We refer to the collection \( WI \) as the ‘Whitney’ collection of \( I \). These intervals need not be good, but they do satisfy the following.

**Proposition 3.11.** For any interval \( I \), and any good \( J \subset I \), there is a \( K \in WI \) which contains it. Moreover,
\[
(3.12) \quad \left\| \sum_{K \in WI} 2^r K(x) \right\|_\infty \lesssim 1.
\]
The reason for this definition is that the maximal good intervals contained inside of $I$ need not be Whitney, they can for instance have accumulation points strictly contained in $I$.

**Proof.** Any good interval $J \in \mathcal{I}$ must satisfy a stronger set of conditions than those that define $\mathcal{W}I$, so there must be an interval $K \in \mathcal{W}I$ that contains it. Concerning the second claim, suppose $K_1, K_2 \in \mathcal{W}I$ with $2^r|K_1| < |K_2| < 2^{r+2}|I|$, and $2^r K_1 \cap 2^r K_2 \neq \emptyset$. Then

$$\text{dist}(\pi K_1, \partial I) \geq \text{dist}(K_1, \partial I) - |K_1|$$

$$\geq \text{dist}(K_2, \partial I) - 2^r |K_2| - (2^r + 1)|K_1|$$

$$\geq |K_2|^\epsilon |I|^{1-\epsilon} - 2^{r+1}|K_2|$$

$$\geq \frac{1}{2} |K_2|^\epsilon |I|^{1-\epsilon} \geq |\pi K_1|^\epsilon |I|^{1-\epsilon}.$$  

That is, $\pi K_1$ meets the criteria for membership in $\mathcal{W}I$, which is a contradiction to maximality. \hfill \Box

We now state and prove the first energy inequality.

**Lemma 3.13.** [Energy Inequality, I] For all intervals $I_0$ and partitions $\mathcal{I}$ of $I_0$ into (not necessarily good) dyadic intervals,

$$\sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W}I} T_\tau(QI_0 \setminus QK)(x_{QK})^2 E(\sigma, K)^2 \sigma(K) \lesssim \mathcal{R}^2 \tau(QI_0).$$  

(3.14)

The argument is broken into two Lemmas, of which the main one is a version of the inequality above, but with holes in the argument of $T_\tau$.

**Lemma 3.15.** For all intervals $I_0$ and partitions $\mathcal{I}$ of $I_0$ into dyadic intervals, we have

$$\sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W}I} T_\tau(QI_0 - V_{2^r K})(x_{QK})^2 E(\sigma, K)^2 \sigma(K) \lesssim \mathcal{R}^2 \tau(QI_0).$$

Here $V_I$ is defined in (3.3), and $s = r/2$.

**Proof.** The sum over the left is over positive quantities, and so it suffices to consider some finite sub-sum, and establish the bound for it. But, with a finite sum, we can then invoke (3.5), so that for each term in the finite sum,

$$T_\tau(QI_0 - V_{2^r K})(x_{QK})^2 E(\sigma, K)^2 \sigma(K) \lesssim \sum_{J : J \subset K} \langle R_\tau^*(QI_0 - V_{2^r K}), h_J^{\sigma} \rangle^2.$$

It is sufficient to show that

$$\sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W}I} \sum_{J : J \subset K} \langle R_\tau^*(QI_0 - V_{2^r K}), h_J^{\sigma} \rangle^2 \lesssim \mathcal{R}^2 \tau(QI_0).$$

(3.16)

Now, turn to the assumption of testing $R_\tau^*$ on Carleson cubes. Obviously,

$$\sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W}I} \| R_\tau^* Q_{2^r K} \|_{L^2(K, \sigma)}^2 \lesssim \mathcal{R}^2 \sum_{I \in \mathcal{I}} \tau(QI_0) \lesssim \mathcal{R}^2 \tau(QI_0),$$

$$\sum_{I \in \mathcal{I}} \| R_\tau^* Q_{I_0} \|_{L^2(I, \sigma)}^2 \lesssim \mathcal{R}^2 \tau(QI_0).$$
The first inequality also depends upon (3.12), and the fact that $2^s K \subset I$. Taking the difference, we see that

$$
\sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W} I} \sum_{J \subset K} \| R^*_\tau(Q_{I_0} - Q_{2^s K}) \|_{L^2(I, \sigma)}^2 \lesssim R^2 \tau(Q_{I_0}).
$$

And so, (3.16) follows from the estimate below,

$$
\sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W} I} \sum_{J \subset K} \langle R^*_\tau(V_{2^s K} - Q_{2^s K}), h^\sigma_J \rangle \lesssim R^2 \tau(Q_{I_0}).
$$

For the proof of this last estimate, it is convenient to use duality. For $K \in \mathcal{W} I$, and functions $\varphi_K \in L^2(K, \sigma)$, of integral zero, it follows that $\varphi_K$ is in the linear span of the Haar functions $\{h^\sigma_J : J \subset K, J \text{ is good}\}$. We should show that

$$
\sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W} I} \sum_{J \subset K} \langle R^*_\tau(V_{2^s K} - Q_{2^s K}), \varphi_K \rangle \lesssim \left\| \sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W} I} \varphi_K \right\|_{\sigma}. \tag{3.17}
$$

First, apply monotonicity from (3.7). Thus, the inner product above is at most

$$
T^\tau(V_{2^s K} - Q_{2^s K})(x_{Q_K}) \int_K |\varphi_K(t)| \sigma(dt). \tag{3.18}
$$

Write $V_{2^s K} - Q_{2^s K}$ as the disjoint union of $V_{K}^{\text{top}} \cup V_{k}^{\text{bottom}}$, where $V_{K}^{\text{top}} \equiv \{(x, y) \in V_{2^s K} - Q_{2^s K} : 8y \geq |2^s K|\}$, that is, $V_{K}^{\text{top}}$ is ‘away’ from the $x$-axis, see Figure 2. For the top, we have, restricting the integration to $V_{K}^{\text{top}}$

$$
T^\tau(V_{K}^{\text{top}})(x_{Q_K}) \lesssim \inf_{x \in K} P^\tau_{\tau} Q_{I_0}(x).
$$

Therefore, it follows that

$$
\text{LHS}^{\text{top}}(3.17) \lesssim \sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W} I} \inf_{x \in K} P^\tau_{\tau} Q_{I_0}(x) \int_K |\varphi_K(t)| \sigma(dt)
\lesssim \sum_{I \in \mathcal{I}} \sum_{K \in \mathcal{W} I} \inf_{x \in K} P^\tau_{\tau} Q_{I_0}(x) \sigma(K)^{1/2} \|\varphi_K\|_{\sigma}.
$$

An application of Cauchy–Schwarz, and using the testing conditions for the Poisson operator will complete the proof.

For the set $V_{K}^{\text{bottom}}$, we just use the $A_2$ condition in this form:

$$
T^\tau V_{K}^{\text{bottom}}(x_{Q_K}) \frac{\sigma(K)}{|K|} \lesssim \mathcal{A}_2.
$$
Thus,

$$\text{LHS}_{\text{bottom}}^{(3.17)} \lesssim \sum_{I \in \mathcal{I}} \sum_{K \in W} T \tau V_{K}^{\text{bottom}}(x_{Q_{K}}) \int_{K} |\varphi_{K}(t)| \sigma(dt)$$

$$\lesssim \sum_{I \in \mathcal{I}} \sum_{K \in W} T \tau V_{K}^{\text{bottom}}(x_{Q_{K}}) \sqrt{\sigma(K)} \|\varphi_{K}\|_{\sigma}$$

$$\lesssim \mathcal{R} \sum_{I \in \mathcal{I}} \sum_{K \in W} \sqrt{|K|} \cdot T \tau V_{K}^{\text{bottom}}(x_{Q_{K}}) \|\varphi_{K}\|_{\sigma}$$

$$\lesssim \mathcal{R} \tau(Q_{I_{0}})^{1/2} \left\| \sum_{K \in W} \varphi_{K} \right\|_{\sigma}.$$ 

Use the bounded overlap property of the $V_{2s}^{\text{bottom}}$ proved in Proposition 3.19 to get the last estimate. Our proof is complete. \hfill \Box

**Proposition 3.19.** Let $\mathcal{I}$ be a partition of interval $I_{0}$ into dyadic intervals. We have

$$\sum_{I \in \mathcal{I}} V_{I}^{\text{bottom}}(x, y) \leq 2.$$

**Proof.** See Figure 2 for an illustration. The collection $\mathcal{I}$ is a partition of dyadic intervals. Assume that it is a sub-partition, with no two dyadic intervals sharing a common endpoint. Then, the sets $V_{I}^{\text{bottom}}$ are pairwise disjoint, as we argue. Suppose that $|I_{1}| \geq |I_{2}|$, and $I_{2}$ lies to the right of $I_{1}$. The dyadic property implies that $\text{dist}(I_{1}, I_{2}) \geq |I_{2}|$. By translation and dilation invariance, we can assume that the right hand endpoint of $I_{1}$ is the origin, and that the left hand endpoint of $I_{2}$ is equal to one, which is the length of $I_{2}$. Then, the right boundary of $V_{I_{1}}^{\text{bottom}}$ and the left boundary of $V_{I_{2}}^{\text{bottom}}$ are the two lines

$$\left\{(t, 2t) : 0 \leq t \leq \frac{|I_{1}|}{16}\right\}, \quad \left\{(1 - s, 2s) : 0 \leq s \leq \frac{1}{16}\right\}.$$ 

These two line segments do not intersect, so the proposition is proved. \hfill \Box

**Lemma 3.20.** For all intervals $I_{0}$, and partition $\mathcal{I}$ of $I_{0}$ into dyadic intervals,

$$\sum_{I \in \mathcal{I}} \sum_{K \in W} T \tau(V_{2s}^{\text{bottom}} \setminus Q_{K})(x_{Q_{K}})E(\sigma, K)^{2} \sigma(K) \lesssim \mathcal{R} \tau(Q_{I_{0}}).$$

**Proof.** Observe that the estimate below

$$\sum_{I \in \mathcal{I}} \sum_{K \in W} T \tau(V_{2s}^{\text{bottom}} - Q_{2s}^{\text{bottom}})(x_{Q_{K}})^{2} E(\sigma, K)^{2} \sigma(K) \lesssim \mathcal{R} \tau(Q_{I_{0}})$$

FIGURE 2. Two sets $V_{I}^{\text{bottom}}$ are indicated in light gray.
follows from the argument above, beginning at (3.17), since our first step was to apply
monotonicity in the form of (3.18). Now, using the \( \mathscr{A}_2 \) condition, and the bounded overlap
property of Proposition 3.11, one easily sees that
\[
\sum_{l \in \mathcal{I}} \sum_{K \in \mathcal{W}_l} T_l(Q_{2^l K} \setminus Q_K)(x_{Q_K})^2 \sigma(K) \lesssim \mathcal{R}^2(\mathcal{Q}_{I_0}).
\]
That completes the proof. \( \square \)

Proof of (3.14). This is an immediate combination of Lemma 3.15 and 3.20. \( \square \)

3.4. Monotonicity, II. Below, we will phrase the monotonicity estimate in terms of the
\( L_0^2(Q_J; \tau) \) norm, which is the norm for the subspace of \( L^2(Q_J; \tau) \) which is orthogonal to
constants. One should note that we could have used this type of definition for \( E(\sigma, I) \) in
(3.10). But, also note that the \( L_0^2(Q_J; \tau) \) norm equals
\[
2 \|g\|_{L_0^2(Q_J; \tau)}^2 = \mathcal{E}_{Q_J}^* \int_Q |g(x) - g(x')|^2 \tau(dx).
\]

Lemma 3.22. [Monotonicity Property, II] Let \( I \) be an interval, and suppose that \( f \in L^2(\mathbb{R}; \sigma) \) is not supported on \( I \). Then, for intervals \( 10J \subset I \),
\[
\|R_\sigma f\|_{L_0^2(Q_J; \tau)} \lesssim \mathcal{P}_\sigma(|f|, J) \left\| \frac{x}{|J|} \right\|_{L_0^2(Q_J; \tau)}.
\]
Moreover, if \( f \geq 0 \),
\[
\mathcal{P}_\sigma(f, |J|) \left\| \frac{x_1}{|J|} \right\|_{L_0^2(Q_J; \tau)} \lesssim \|R_\sigma f\|_{L_0^2(Q_J; \tau)}.
\]

Notice that the second inequality reverses the first, but the important details of the in-
equality are that only \( x_1 \) appears on the left and the full vector of transforms appears on
the right.

Proof. The canonical value of the inner products (2.4) is used. The first inequality (3.23) is
simple. For coordinates \( j = 1, 2 \), the function \( R_\sigma^j f \) is \( C^2 \) and real-valued on \( Q_J \). It follows
from (3.2), and the mean value theorem that for any \( x, y = (y_1, y_2) \in Q_J \), there is an
\( z = (z_1, z_2) \), on the line between \( x \) and \( y \) so that
\[
R_\sigma^j f(x) - R_\sigma^j f(y) = (x - y) \cdot \nabla R_\sigma^j(z).
\]
By inspection,
\[
|\nabla R_\sigma^j f(z)| \lesssim \frac{1}{|J|} \cdot \mathcal{P}_\sigma(|f|, J).
\]
So (3.23) follows from (3.21).

For the inequality (3.24), it suffices to prove the following. For \( x \in Q_J \) and \( y = (y_1, y_2) \in Q_J \), there holds
\[
\frac{|x_1 - y_1|}{|J|} \mathcal{P}_\sigma(f, J) \lesssim |R_\sigma f(x) - R_\sigma f(y)|.
\]
Assuming (3.25) holds, we simply square and integrate with respect to \( Q_J \tau(dx) \) to conclude
(3.24).
The proof of (3.25) depends upon this elementary calculation: For \( \{j, k\} = \{1, 2\} \),
\[
\inf_{x \in Q_j} |\partial_x R^j f(x)| \geq |J|^{-1} P_0(f, J),
\]
\[
\sup_{x \in Q_j} |\partial_x R^j f(x)| \leq c_r |J|^{-1} P_0(f, J),
\]
where the first constant is absolute, and the second can be made arbitrarily small, with large \( r \). In the first line, we are taking the partial of \( R^j \) in the \( j \)th direction, and in the second, in the complementary direction.

To deduce (3.25), assuming that \( |x - y| < 4|x_1 - y_1| \), use the differential inequalities above with \( j = 1 \) and \( k = 2 \). We only need the first coordinate of the Riesz transforms to verify (3.25).

And, otherwise, \( x \) and \( y \) have very similar first coordinate, and differ mostly in the second coordinate, so one can use the differential inequalities with \( j = 2 \) and \( k = 1 \). We use the second coordinate of the Riesz transforms to verify (3.25).

\[ \square \]

3.5. **Energy Inequality, II.** We focus on the energy inequality in the dual setting. For an interval \( I \), we define the energy in a different, but equivalent, way than before,

\[
E(\tau, I)^2 \equiv \tau(Q_I)^{-1} \|x I\|_{L^2_0(Q_I, \tau)}^2.
\]

Keep in mind that \( x \in \mathbb{R}^2_+ \). Here \( L^2_0(Q_I, \tau) \) denotes the norm of the function, less its mean.

**Lemma 3.27.** [Energy Inequality, II] For any interval \( I_0 \) and partition \( \mathcal{P} \) of \( I_0 \) into dyadic intervals,

\[
\sum_{I \in \mathcal{P}} \sum_{K \in \mathcal{W}I} P_0(I_0 \setminus K, K)^2 E(\tau, K)^2 \tau(Q_K) \leq \mathcal{A}^2 \sigma(Q_{I_0}).
\]

**Proof.** Using the \( A_2 \) inequality, we can enlarge the holes, namely,

\[
\sum_{I \in \mathcal{P}} \sum_{K \in \mathcal{W}I} P(\sigma \cdot 2^r K \setminus K, K)^2 \tau(Q_K) \leq \mathcal{A}^2 \sum_{I \in \mathcal{P}} \sum_{K \in \mathcal{W}I} |K| \cdot P(\sigma \cdot 2^r K, K)
\]

\[
\leq \sum_{I \in \mathcal{P}} \sum_{K \in \mathcal{W}I} \sigma(2^r K) \leq \mathcal{A}^2 \sigma(I_0).
\]

Note that this depends critically on the bounded overlap property (3.12).

It remains to consider the sum with the Poisson term being \( P(\sigma \cdot (I_0 - 2^r K), K) \). We first consider the contribution to energy coming only from the first coordinate of \( x \). It suffices to prove the estimate with \( \mathcal{P} \) a finite sub-partition of \( I_0 \), and the assumption that each \( \mathcal{W}I \) is also finite. The constant will be independent of this assumption. The monotonicity property (3.24) applies, so that it suffices to estimate

\[
\sum_{I \in \mathcal{P}} \sum_{K \in \mathcal{W}I} \|R_\sigma(I_0 - 2^r K)\|^2_{L^2(Q_K; \tau)}
\]

\[
\leq \sum_{I \in \mathcal{P}} \sum_{K \in \mathcal{W}I} \|R_\sigma I_0\|^2_{L^2(Q_K; \tau)} + \|R_\sigma(2^r K)\|^2_{L^2(Q_K; \tau)}
\]

\[
\leq \|R_\sigma I_0\|^2_{L^2(Q_{I_0}; \tau)} + \sum_{I \in \mathcal{P}} \sum_{K \in \mathcal{W}I} \|R_\sigma(2^r K)\|^2_{L^2(Q_K; \tau)}.
\]
And these two terms are controlled by the testing inequalities and (3.12).

It remains to consider the contribution to energy coming from the second coordinate, which is controlled by testing inequalities, but only applied in the second coordinate of the Riesz, namely the Poisson integral. Observe that

\[
\mathcal{P}(\sigma \cdot (I_0 - 2rK), K)^2 \geq \mathcal{P}(\sigma \cdot I_0, K)^2 \int_{Q_K} x_2^2 \tau(dx) \\
\leq \int_{Q_K} (R_\sigma^2 I_0)^2 \tau(dx).
\]

Then this case is controlled by the testing inequalities. \qed

4. Global to Local Reduction, I

4.1. Initial Reductions. We can assume that \(f\) is supported on a (large) interval \(I^0\), and \(g\) is supported on \(Q_{I_0}\). By trivial application of the testing inequalities, we can further assume that \(f\) has \(\sigma\)-integral zero, and \(g\) has \(\tau\)-integral zero. Thus, \(f, g\) are in the span of good adapted Haar functions. And we can assume that \(|I^0| \geq 2^r |J|\) for all \(J\) in the Haar support of \(f\), and similarly \(|I^0| \geq 2^r |Q|^{1/2}\) for all cubes \(Q\) in the Haar support of \(g\).

Further restrictions on the Haar supports of \(f\) and \(g\) are made, these restrictions are phrased in terms of \(r\) and \(\epsilon\). The values needed for \(r\) and \(\epsilon\) are derived from the elementary estimates. For an integer \(0 \leq s_f < 4r\) (which plays no further role in the argument), assume that

\[
f = \sum_{I \in D \cap I^0 : \log_2 |I| \in 4rZ + s_f + 1} \Delta^f_I f,
\]

and let \(\mathcal{D}^r_f \equiv \{I \in D : I \subset I^0 \text{ and } \log_2 |I| \in 4rZ + s_f\}\). Thus, \(\mathcal{D}^r_f\) is a grid containing all the children of intervals in the Haar support of \(f\) – and many more intervals. The constant \(4r\) in (4.1) is imposed due to considerations in §5. Likewise, for an integer \(0 \leq s_g < 4r\), assume that

\[
g = \sum_{Q \in \mathcal{D}^2 : Q \subset Q_{I^0} \text{ and } \log_2 |Q| \in 4rZ + s_g + 1} \Delta^g_Q g,
\]

and let \(\mathcal{D}^r_g \equiv \{I \in D : I \subset I^0 \text{ and } \log_2 |I| \in 4rZ + s_g\}\). Our specificity about the martingale difference support for \(f\) and \(g\) has the purpose of easing the technical burdens at different points in the proof below. Consequently, we will reference the grids \(\mathcal{D}^r_f\) and \(\mathcal{D}^r_g\) when appropriate.

Let \(P_{\text{Car}}^r g \equiv \sum_{I \in \mathcal{D}^r_f} \Delta^r_Q I g\), where the sum is only over Carleson cubes. It suffices to consider (good) functions in the range of this projection. This proposition is proved in §7.

**Proposition 4.2.** The following estimate holds:

\[
\|R^*_\tau (g - P_{\text{Car}}^r g)\|_\sigma \lesssim 2^{1/2} \|g\|_\tau.
\]

Thus, we assume throughout that \(g\) is a good function, with \(P_{\text{Car}}^r g = g\), as well as satisfying the further restriction on the Haar supports described above. We define the two triangular
forms

\begin{align}
B^{\text{above}}(f, g) &\equiv \sum_{I, J : J \in \mathcal{D}_r g} \mathbb{E}^\tau_{Q_I} \Delta^\tau_{Q_I} g \cdot \langle R^\tau_{Q_{IJ} \setminus Q_I} g, \Delta^\tau_{Q_{IJ} \setminus Q_I} \rangle_{\sigma}, \\
B^{\text{below}}(f, g) &\equiv \sum_{I, J : I \in \mathcal{D}_r g} \mathbb{E}^\tau_{Q_I} \Delta^\tau_{Q_I} f \cdot \langle \Delta^\tau_{Q_{IJ} \setminus Q_I} g, R_{\sigma} J_I \rangle_{\tau},
\end{align}

where \( J_I \) is the child of \( I \) that contains \( J \), and \( Q_{IJ} \) is the child of \( Q_I \) that contains \( Q_J \). See Figure 3. These two forms are dual to one another, but their analysis is different, due to the assumptions on the supports of \( \sigma \) and \( \tau \).

**Proposition 4.5.** The following estimate is true:

\[ |\langle R^\tau_{Q_I} g, f \rangle_{\sigma} - B^{\text{above}}(f, g) - B^{\text{below}}(f, g)| \lesssim \mathcal{R} \|f\|_{\tau} \|g\|_{\sigma}. \]

The proof of Proposition 4.5 appears in Section 7. We concentrate on the ‘above’ form in the remainder of this section.

### 4.2. The Stopping Data

The function \( g \) is in the linear span of the martingale differences associated with good Carleson cubes and is supported on the cube \( Q_{I_0} \). Construct stopping intervals for \( g \), which is a collection of dyadic intervals \( \mathcal{F} \). Initialize \( \mathcal{F} \) to be the maximal elements of \( \mathcal{D}_r g \) contained in \( I_0 \). In the inductive stage, if \( F \in \mathcal{F} \) is minimal, add to \( \mathcal{F} \) the maximal children \( I \in \mathcal{D}_r g \) of \( F \) that meet either of these conditions:

1. (A large average) \( \mathbb{E}^\tau_{Q_I} |g| \geq 10 \mathbb{E}^\tau_{Q_F} |g| \);
2. (Energy Stopping) \( \sum_{K \in \mathcal{W}} \mathbb{T}_{\mathcal{R}(Q_K \setminus Q_K)}(x_{Q_K})^2 E(\sigma, K) \sigma(K) \geq C_0 \mathcal{R}^2 \mathcal{T}(Q_I) \).

Then, if \( \mathbb{E}^\tau_{Q_I} |g| \geq 2 \mathbb{E}^\tau_{Q_F} |g| \), set \( \alpha_g(I) \equiv \mathbb{E}^\tau_{Q_I} |g| \), otherwise set \( \alpha_g(I) \equiv \mathbb{E}^\tau_{Q_F} |g| \). This completes the construction of the stopping data for \( g \).

It is a consequence of the energy inequality that \( \mathcal{F} \) satisfies a \( \tau \)-Carleson condition for a sufficiently large constant \( C_0 \) for energy stopping. Namely,

\[ \sum_{F' \in \mathcal{F} : F' \subseteq F} \tau(Q_{F'}) \leq \frac{1}{2} \tau(Q_F), \quad F \in \mathcal{F}. \]

It is also immediate from construction that the stopping intervals control the averages of \( g \) in the following sense: For all intervals \( I \in \mathcal{D}_g, I \subseteq I_0 \), we have

\[ |\mathbb{E}^\tau_{Q_I} g| = \left| \sum_{K : K \supseteq I} \mathbb{E}^\tau_{Q_K} \Delta^\tau_{Q_K} g \right| \lesssim \alpha_g(\pi_F I). \]
(The notation $\pi_F J$ is defined at the beginning of §3.)

We make this brief remark about the collections $F$ and $\{WF : F \in F\}$. For each $F \in F$, and good $J \in F$, there is a $K \in WF$ with $J \subset K$. For intervals $F \in F$, define Haar projections by

$$H_F^r g = \sum_{I : \pi F I = F} \Delta_{Q_I}^r g,$$
$$H_F^\sigma f = \sum_{J : \pi F J = F} \Delta_J^\sigma f.$$

In the second line, we take $\pi_F J$ to be the smallest member of $F$ so that $J \in 4\tau \pi F$. For the projection on $L^2(\mathbb{R}^+_4; \tau)$, we include the parent of $F$. We call this inequality the quasi-orthogonality bound; it is basic to the proof:

$$\sum_{F \in F} \{ \| \Delta_{Q_F}^r g \| \tau \cdot (Q_F)^{1/2} + \| H_F^r g \|_r \} \| H_F^\sigma f \|_\sigma \lesssim \| f \|_r \| g \|_\sigma. \tag{4.6}$$

This follows from the $\tau$-Carleson property of $F$ and the quasi-orthogonality of the Haar projections. It will appear below with different choices of these orthogonal projections.

Observe that we have

$$B_{\text{above}}(f, g) = \sum_{F \in F} \sum_{F' : F \supset F} B_{\text{above}}(\tilde{H}_F^\sigma f, \tilde{H}_{F'} g). \tag{4.7}$$

Indeed, the definition of $B_{\text{above}}$ is over a sum of pairs of intervals $J, I$ with $J \in 4\tau I$ and say $\pi_F J = F$. We necessarily have $J \in 4\tau (\pi_F I)$, hence $F \subset \pi_F I$. Then it is clear that this pair of intervals appear exactly once on the right in the top line of (4.7).

We can now turn to the global to local reduction for the form $B_{\text{above}}$. In the sum below we are taking that part of the right side of (4.7) which is 'separated' by $F$. Namely, in the lemma below, we form the sum only over pairs of intervals $I, J$ that do not have the same $F$-parent.

**Lemma 4.8.** [Global to Local Reduction, I] The following estimate holds:

$$\left| \sum_{I} \sum_{J : \pi F J \subset I} \sum_{J \in 4\tau I} \mathbb{E}_{Q_{I/J}} \Delta_{Q_I}^r g \cdot \langle R_{I/J}^r Q_{I/J} \Delta_J f \rangle_\sigma \right| \lesssim \mathcal{R} \| g \|_r \| f \|_\sigma. \tag{4.9}$$

The proof is a consequence of the functional energy inequalities that we present below.

**Proof.** We invoke, for the first time, the exchange argument, namely exchanging the inequality concerning a singular integral for one involving a purely positive operator. This entails (a) controlling the sums of martingale differences by the stopping values; (b) replacing the argument of the singular integral by a stopping interval; (c) appealing to interval, or Carleson cube, testing and quasi-orthogonality to complete the bound in this case; (d) for the complementary argument in the singular integral, appeal to monotonicity, to get a positive operator; (e) appeal to energy inequality and quasi-orthogonality or the functional energy inequality to complete the bound in this case.

The details in this case are as follows. We are to bound the sum

$$\sum_{F \in F} \sum_{I : F \subset I} \sum_{J : \pi F J = F} \mathbb{E}_{Q_{I/J}} \Delta_{Q_I}^r g \cdot \langle R_{I/J}^r Q_{I/J} \Delta_J f \rangle_\sigma.$$


Write the argument of the Riesz transform as \( Q_{1F} = Q_F + (Q_{1F} - Q_F) \). In the case that the argument is \( Q_F \), observe that by construction of the stopping data, that

\[
\left| \sum_{I : F \subseteq I} E_{Q_{1I}}^r A_{Q_I}^r g \right| \lesssim E_{Q_F}^r |g|.
\]

Therefore, we can estimate using the testing inequality for the Riesz transform:

\[
\left| \sum_{I : F \subseteq I} \sum_{J : \hat{\pi}_F J = F} E_{Q_{1I}}^r A_{Q_I}^r g \cdot (R_i^r Q_F, \Delta_j^\sigma f)_\sigma \right| \\
\lesssim E_{Q_F}^r |g| \left| \sum_{J : \hat{\pi}_F J = F} (R_i^r Q_F, \Delta_j^\sigma f)_\sigma \right| \\
\lesssim A E_{Q_F}^r |g| \cdot \tau(Q_F)^{1/2} \| \tilde{H}_f^\sigma \|_\sigma^2.
\]

The sum over \( F \in \mathcal{F} \) of this last expression is controlled by quasi-orthogonality, (4.6).

In the case of the argument of the Riesz transform being \( Q_{1F} - Q_F \), first note that, again by construction of stopping data, that

\[
\left| \sum_{I : F \subseteq I} E_{Q_{1I}}^r A_{Q_I}^r g \cdot (Q_{1F} - Q_F) \right| \lesssim Mg,
\]

where \( Mg \) is the dyadic \( \tau \)-adapted maximal function of \( g \). Of course, \( \| Mg \|_\tau \lesssim \| g \|_\tau \). Now, since \( J \subseteq_\tau F \), from the monotonicity principle, (3.6), we see that

\[
\sum_{F \in \mathcal{F}} \left| \sum_{I : F \subseteq I} \sum_{J : \hat{\pi}_F J = F} E_{Q_{1I}}^r A_{Q_I}^r g \cdot (R_i^r (Q_{1F} - Q_F), \Delta_j^\sigma f)_\sigma \right| \\
\lesssim \sum_{F \in \mathcal{F}} \sum_{K \in W F} \tau(Mg \cdot Q_F^c)(x_{Q_K}) \sum_{J : \hat{\pi}_F J = F} \left| \frac{t}{|K|} h_{J_\sigma} \right| \tilde{f}_\sigma(J). \\
(4.10)
\]

The last line follows from Cauchy–Schwarz and the definition of the measure \( \mu \) and \( \tilde{\tau} \) that we give here, in which a scale of \( K \) is taken out.

\[
\tilde{\tau} g(K) \equiv \int_{\mathbb{R}^+} \frac{g(y)}{y^2 + \text{dist}(y_1, K)^2} \tau(dy),
\]

\[
\mu_K \equiv \sum_{J : \hat{\pi}_F J = F} \langle t, h_{J_\sigma}^\sigma \rangle^2, \quad F \in \mathcal{F}, \ K \in W F.
\]

That the term above involving \( Mg \) in (4.10) is bounded by \( A \| g \|_\tau \) is the conclusion of Lemma 4.13, the functional energy inequality.

4.3. Functional Energy Inequality, I.

Lemma 4.13. [Functional Energy Inequality, I] The operator $\hat{T}_\tau$, as defined in (4.11), satisfies this two weight inequality:

\begin{equation}
\sum_{F \in \mathcal{F}} \sum_{K \in W_F} \hat{T}_\tau(\phi \cdot Q_F^c)(K)^2 \mu_K \lesssim \mathcal{R}^2 \|\phi\|^2.
\end{equation}

The plan of the proof, following Hytönen [9], is to derive a dyadic approximate to the terms $\hat{T}_\tau(\phi \cdot Q_F^c)$. The latter will then satisfy a two weight theorem in terms of an $A_2$ condition, and a pair of testing conditions. We can then verify these latter conditions to complete the proof of the result above.

4.4. Reduction to Dyadic Operators. Recall that the dyadic grid $\mathcal{D}$ is in $\mathbb{R}$, which is associated to a dyadic grid $\mathcal{D}^2$ on the $\mathbb{R}^2_+$. The collection of intervals $\{3I : I \in \mathcal{D}\}$ is a union of three collections $\mathcal{C}_u$, $1 \leq u \leq 3$, each of which is like the dyadic grid with respect to nesting and covering properties. This property is easy to check and well-known: For any non-dyadic interval $I$, there are at least two choices of $1 \leq u \leq 3$, and an interval $J \in \mathcal{C}_u$ such that $3I \subset J$ and $|J| = 12|I|$. Denote this interval by $\pi_u I$, if it exists. If there is no such interval, set $\pi_u I = \emptyset$.

It is then follows that for $I \in \mathcal{D}$ and each $j \geq 1$, there is a choice of $1 \leq u = u_{I,j} \leq 3$ so that $3 \cdot 2^j I \subset \pi_{u} (\pi_{u} I) \equiv (\pi_{u} I)^{(j)}$. From this, we have

Proposition 4.15. [9, Prop 6.6] For $I \in \mathcal{D}$, and function $\phi \geq 0$, there holds

$$
\tilde{T}_\tau(\phi, I) \simeq \sum_{u : \pi_{u} I \neq \emptyset} Q_u(\phi, \pi_{u} I)
$$

where $Q_u(\phi, J) \equiv \sum_{j \geq 0} \frac{1}{|J(j)|^2} \int_{Q_{j+1}(j)} \phi \, d\tau$, $J \in \mathcal{C}_u$.

In the last definition, $J^{(j)}$ is the $j$-fold ancestor of $J \in \mathcal{C}_u$.

Proof. It follows from the definition (4.11) that

$$
\tilde{T}_\tau(\phi, I) \simeq \sum_{j=1}^{\infty} \frac{1}{|J(j)|^2} \int_{Q_{j+1}(j)} \phi \, d\tau.
$$

For each $j$, there is a choice of $1 \leq u \leq 3$ with $2^j I \subset (\pi_{u} I)^{(j)}$, and since $12|I| = |\pi_{u} I|$, we also have $(\pi_{u} I)^{(j)} \subset 12 \cdot 2^j I$. This proves

$$
\tilde{T}_\tau(\phi, I) \lesssim \sum_{u : \pi_{u} I \neq \emptyset} Q_u(\phi, \pi_{u} I).
$$

The reverse inequality is easier. \hfill \square

It follows that the inequality (4.14) would follow from stronger dyadic inequalities, namely

$$
\sum_{K \in \mathcal{K}_u} Q_{u,\tau}(\phi \cdot Q_F^c, K)^2 \mu_{u,K} \lesssim \mathcal{R}^2 \|\phi\|_{L^2(\mathbb{R}_+^2, \tau)},
$$

where $\mathcal{K}_u \equiv \bigcup_{F \in \mathcal{F}} \{\pi_{u} K : K \in W_F : \mu_{K} \neq 0, \pi_{u} K \neq \emptyset\}$,
and \( \mu_{u,K} = \mu_{u^{-1}K} \). This inequality is dualized to

\[
\sum_{K \in \mathcal{K}} \mathcal{Q}_{u,\tau}(\phi \cdot K^c, K) \int_{W_K} g \, d\mu \lesssim \mathcal{R} \| \phi \|_{L^2(\mathbb{R}^2_+, \tau)} \| g \|_{L^2(\mathbb{R}^2_+)}.
\]

where \( W_K \equiv K \times [\frac{1}{2}|K|, |K|] \), the top half of the Carleson box over \( K \), and by abuse of notation, \( \mu \equiv \sum_{F \in \mathcal{F}} \sum_{K \in \mathcal{K} \cap \mathcal{F}} \mu_{u,K} \delta_{x_{W_K}} \). The inequality above has necessary and sufficient conditions for boundedness in terms of an \( A_2 \) condition and two testing inequalities.

**Lemma 4.17.** The inequality (4.16) holds if and only if these conditions hold uniformly in \( I \in \mathcal{C}_u \):

\[
\left( \frac{\tau(Q_{j(1)} \setminus Q_I)}{|I|} \right) \mu(Q_I) \| I \|^2 \lesssim \mathcal{R}^2,
\]

\[
\left| \int_{Q_I} \left[ \sum_{J : J \subseteq I} \left( \frac{\tau(Q_{j(1)} \setminus Q_J)}{|J|^2} \cdot Q_J \right) \right] \right|^2 d\mu \lesssim \mathcal{R}^2 \tau(Q_I),
\]

\[
\left| \int_{Q_I} \left[ \sum_{J : J \subseteq I} \left( \frac{\mu(Q_J)}{|J|^2} \cdot (Q_{j(1)} \setminus Q_J) \right) \right] \right|^2 d\tau \lesssim \mathcal{R}^2 \mu(Q_I).
\]

**Proof.** Write the left side of (4.16) as follows

\[
LHS(4.16) = \sum_{K \in \mathcal{K}} \sum_{J : J \not\subseteq K} |J|^{-2} \int_{Q_{j(1)} \setminus Q_K} g \, d\tau \cdot \int_{W_K} g \, d\mu
\]

\[
\simeq \sum_{K \in \mathcal{K}} \sum_{J : J \not\subseteq K} |J|^{-2} \int_{Q_{j(1)} \setminus Q_J} g \, d\tau \cdot \int_{W_K} g \, d\mu
\]

\[
= \sum_{J} |J|^{-2} \int_{Q_{j(1)} \setminus Q_J} g \, d\tau \cdot \int_{Q_J} g \, d\mu.
\]

The necessity of the three conditions is then clear. The sufficiency then follows from Hytönen [9, Theorem 6.8].

It remains to verify the three conditions (4.18)—(4.20). The first \( A_2 \) condition follows immediately from the \( A_2 \) condition (with holes) and the bound \( \mu(Q_I) \lesssim \sigma(I) \cdot |I|^2 \), which is used below.

### 4.4.1. Forward Testing Inequality for \( \mathcal{Q} \)

We verify the testing inequality (4.19), using a recursion along the stopping collection \( \mathcal{F} \). Let \( I_0 \in \mathcal{C}_u \) be as in (4.19), and let \( \mathcal{K}_{u,0} \) be those \( K \in \mathcal{K}_u \) which are strictly contained in \( I_0 \) and \( \mu_{u,K} \not= 0 \), moreover, \( \pi_u^{-1}K \in \mathcal{W}F \) for some \( F \in \mathcal{F} \) which is not contained in \( I_0 \). Let \( \mathcal{K}_{u,1} \) be the \( K \in \mathcal{K}_u \), so that \( \pi_u^{-1}K \in \mathcal{W}F \), where \( F \in \mathcal{F} \) is a maximal interval strictly contained in \( I_0 \). These two inequalities are a consequence of the energy inequality: For \( k = 0, 1 \),

\[
N_k(I_0) \equiv \left\| \sum_{K \in \mathcal{K}_{u,k}} \sum_{J : J \not\subseteq K} \frac{\tau(Q_{j(1)} \setminus Q_J)}{|J|^2} \cdot Q_J \right\|_\mu \lesssim \mathcal{R} \tau(Q_{I_0})^{1/2}.
\]

To recurse along the stopping collection, let \( \mathcal{F}_1 \) be the maximal elements of \( \mathcal{F} \) strictly contained in \( I_0 \), and inductively set \( \mathcal{F}_{j+1} \) to be the maximal elements of \( \mathcal{F} \) which are strictly
contained in some \( F \in \mathcal{F}_j \). Using a standard Cauchy–Schwarz estimate, and the geometric decay property of the stopping collection \( \mathcal{F} \) in \( \tau \)-measure, we can estimate

\[
\left\| \sum_{J : J \subseteq I_0} \frac{\tau(Q_{J(1)} \setminus Q_J)}{|J|^2} \cdot Q_J \right\| \mu \lesssim \sum_{k=0,1} N_k(I_0)^2 + \sum_{j=1}^{\infty} j^2 \sum_{F \in \mathcal{F}_j} \sum_{k=0,1} N_k(F)^2 \\
\lesssim \mathcal{R} \left\{ \tau(Q_{I_0}) + \sum_{j=1}^{\infty} j^2 \sum_{F \in \mathcal{F}_j} \tau(Q_F) \right\} \\
\lesssim \mathcal{R} \tau(Q_{I_0}).
\]

It remains to verify (4.21). The case of \( k = 1 \) in (4.21) is a special case of the energy inequality. By Proposition 4.15, and the definition of \( \mu \) in (4.12),

\[
N_1(I_0)^2 \lesssim \sum_{F \in \mathcal{F}_1} \sum_{L \in \mathcal{K}_{a,1}} \tilde{T}_\tau(Q_{I_0} \setminus Q_L)(L)^2 \mu(Q_L) \\
\lesssim \sum_{F \in \mathcal{F}_1} \sum_{K \in \mathcal{V}_F} T_\tau(Q_{I_0} \setminus Q_K)(x_{Q_K}, |K|)^2 E(\sigma, K) \lesssim \mathcal{R}^2 \tau(Q_{I_0}).
\]

The case of \( k = 0 \) is also an instance of the energy inequality, once we observe the following. Write \( I_0 = I_{0,0} \cup I_{0,1} \cup I_{0,2} \), where the latter three intervals are in \( \mathcal{D} \), and have equal length. For \( I_{0,s} \), let \( F_{s,1} \subset F_{s,2} \subset F_{s,3} \subset F_{s,4} \) be the four immediate ancestors of \( I_{0,s} \) in the \( \mathcal{F} \)-stopping tree. Then, for each good interval \( J \subset I_{0,s} \), we have \( \tilde{\pi}_F J \subset F_{s,4} \), since the scales in \( \mathcal{F} \) are separated by \( r \). It follows that we can partition the intervals \( K \in \mathcal{K}_{a,0} \) also contained in \( I_{0,s} \) into four collections, \( \mathcal{K}_{s,t} \), \( 1 \leq t \leq 4 \), for which \( \pi_u^{-1} K \in \mathcal{W}_{s,t} \), for all \( K \in \mathcal{K}_{s,t} \). Then, this inequality is an instance of the energy inequality.

\[
\sum_{K \in \mathcal{K}_{s,t}} \tilde{T}_\tau(Q_{I_0} \setminus Q_K)(K)^2 \mu(W_{Q_K}) \lesssim \mathcal{R}^2 \tau(Q_{I_0}).
\]

This completes the proof, after summing over \( 0 \leq s \leq 2 \) and \( 1 \leq t \leq 4 \).

4.4.2. The backwards testing inequality. We prove (4.20). There are two cases, based upon the decomposition of the set \( Q_{J(1)} \setminus Q_J \). Set \( W^k_j \equiv \int_{J} J \times [2^{-k-1}, 2^{-k}]|J|, \) for integers \( k \geq 0 \), so that \( Q_J = \bigcup_{k=0}^{\infty} W^k_j \). From the definition of \( \mu \), see (4.12), there holds \( \mu(W^k_j) \lesssim 2^{-2k}|J|^2 \sigma(J) \).

Let \( I \) be as in (4.20). The sets \( W^0_{J(1)} \) are the top halves of \( Q_{J(1)} \), and they are pairwise disjoint as \( J \subseteq I \) varies. Then, it is clear that for each integer \( k \geq 0 \),

\[
\int_{Q_I} \left[ \sum_{J : J \subseteq I} \frac{\mu(W^k_j)}{|J|^2} \cdot W^0_{J(1)} \right]^2 d\tau = \sum_{J : J \subseteq I} \frac{\mu(W^k_j)^2}{|J|^4} \tau(W^0_{J(1)}) \\
\lesssim 2^{-2k} s_{L_2} \sum_{J : J \subseteq I} \mu(W^k_j) \lesssim 2^{-2k} s_{L_2} \mu(Q_I).
\]

This is summable in \( k \geq 0 \).
Let \( \tilde{J} = J^{(1)} \setminus J \) be the sibling to \( J \), so that the cubes \( Q_{\tilde{J}} \) overlap. For integers \( k \geq 0 \), estimate

\[
\int_{Q_{\tilde{J}}} \left[ \sum_{J : J \subseteq I} \frac{\mu(W_j^k)}{|J|^2} \cdot Q_j \right]^2 \, d\tau = \sum_{J : J \subseteq I} \frac{\mu(W_j^k)}{|J|^2} \sum_{K : K \subseteq J} |K|^2 \tau(Q_K) 
\approx 2^{-2k} \|\mu\|_2 \sum_{J : J \subseteq I} \frac{\mu(W_j^k)}{|J|^2} \sum_{K : K \subseteq J} |K|^2
\approx 2^{-2k} \|\mu\|_2 \sum_{J : J \subseteq I} \mu(W_j^k) \lesssim 2^{-2k} \|\mu\|_2 \mu(Q_I).
\]

5. Local Estimates, I

The bound for the form \( B^{\text{above}}(f, g) \), defined in (4.3), satisfies the global to local estimate (4.9). Therefore, it remains to prove the following estimate. In this Lemma, the collection \( \mathcal{F} \), and the notation is as in §4.

**Lemma 5.1.** For each \( F \in \mathcal{F} \), we have

\[
|B^{\text{above}}(\tilde{H}_F^r f, H_F^r g)| \lesssim \mathcal{R} \left\{ \mathbb{E}^r_{Q,F} |g| \cdot \tau(Q_F)^{1/2} + \|H_F^r g\| \right\} \|\tilde{H}_F^r f\|_\sigma.
\]

In view of the global to local reduction in Lemma 4.8, and the quasi-orthogonality inequality (4.6), this lemma completes the control of the form \( B^{\text{above}}(f, g) \).

The estimate is homogeneous in \( f \), but not \( g \). The following step is a simple appeal to the Carleson cube testing hypothesis. For each \( J \), define \( \varepsilon_J \) by the formula

\[
\varepsilon_J \mathbb{E}^r_{Q,F} |g| = \sum_{I : \pi_F I = F} \mathbb{E}^r_{Q_I} \Delta^r_{Q_I} g.
\]

Since we restrict to \( J \) with \( \pi_F J = F \), it follows from the construction of the stopping tree \( \mathcal{F} \), that \( |\varepsilon_J| \lesssim 1 \). We perform part of the exchange argument. In the bilinear form, the argument of the Riesz transform is \( Q_J = Q_F - (Q_F - Q_{IJ}) \). With just \( Q_F \), we have

\[
\sum_{I : \pi_F I = F} \mathbb{E}^r_{Q_{IJ}} \Delta^r_{Q_I} g \langle R_\tau Q_F, \Delta^r_{J} f \rangle_\sigma = \mathbb{E}^r_{Q_F} |g| \langle R_\tau Q_F, \sum_{J : \pi_F J = F} \varepsilon_J \Delta^\sigma_{J} f \rangle_\sigma
\]

\[
\leq \mathbb{E}^r_{Q_F} |g| \|Q_F R_\tau Q_F\|_\sigma \left\| \sum_{J : \pi_F J = F} \varepsilon_J \Delta^\sigma_{J} f \right\|_\sigma
\]

\[
\leq \mathcal{R} \mathbb{E}^r_{Q_F} |g| \cdot \tau(Q_F)^{1/2} \|\tilde{H}_F^r f\|_\sigma.
\]

This uses the testing inequality and orthogonality of the martingale differences. Quasi-orthogonality is used to sum this last estimate over all \( F \in \mathcal{F} \).

It remains to bound the term where the argument of the Riesz transform is \( (Q_F - Q_{IJ}) \), that is the stopping form

\[
B^{\text{stop}}_F(f, g) \equiv \sum_{I : I \subseteq F} \mathbb{E}^r_{Q_{IJ}} \Delta^r_{Q_I} g \cdot \sum_{J : \pi_F J = F} \langle R^r_\tau (Q_F - Q_{IJ}), \Delta^\sigma_{J} f \rangle_\sigma.
\]

No naive application of the remaining part of the exchange argument will be successful. Instead, we will use a sophisticated recursion to prove the following lemma.
Lemma 5.2. The following estimate is true:
(5.3) \[ |B^\text{stop}_F(f, g)| \lesssim \mathcal{A} \|g\|_{\tau} \|f\|_\sigma. \]

One then can invoke quasi-orthogonality to sum the estimate above since it is applied to the corresponding Haar projections of \( f \) and \( g \), namely, to \( H^\#_F f \) and \( H^\#_F g \) respectively.

5.1. The Initial Definitions and the Size Lemma. We regard the interval \( F \in \mathcal{F} \) as fixed. We need an elaborate decomposition of the bilinear form \( B^\text{stop}_F(f, g) \), by dividing up the intervals \( I, J \) over which the sum is formed. The distinction between the \( 4r \) and \( r \) in (4.1) is now important. Recall that
\[ D^r_F \equiv \{ I \in \mathcal{D} : I \subset I^0 \ \log_2 |I| \in rZ + s_f \}, \]
and that the Haar support of \( f \) is in the good intervals in
\[ D^{4r}_F \equiv \{ I \in \mathcal{D} : I \subset I^0 \ \log_2 |I| \in 4rZ + s_f + 1 \}. \]

We define \( D^r_g \) and \( D^{4r}_g \) similarly.

Let \( \mathcal{P} \) be a collection of pairs \( (I, J) \in D^{4r}_F \times D^{4r}_F \) of intervals \( J \in 4r, I \subset F \). Denote a generic element by \( (P_1, P_2) \in \mathcal{P} \). Set \( \mathcal{P}_j \equiv \{ P_j : (P_1, P_2) \in \mathcal{P} \} \) to be the projection onto the respective coordinate. Also, let \( \tilde{\mathcal{P}}_1 \equiv \{ (P_1)_{P_2} : (P_1, P_2) \in \mathcal{P} \} \).

We say that \( \mathcal{P} \) is admissible if
(1) For \( (P_1, P_2) \in \mathcal{P} \), there holds \( P_2 \in 4r, P_1 \) and both \( P_1 \) and \( P_2 \) are good.
(2) Let \( F' \) be an energy stopping interval of \( F \), as defined at the beginning of §4.2. No interval in \( \tilde{\mathcal{P}}_1 \) is contained in \( F' \), and no interval \( P_2 \in \mathcal{P}_2 \) satisfies \( P_2 \in 4r, F' \).
(3) For each fixed \( Q \), the collection \( \{ P_1 : (P_1, Q) \in \mathcal{P} \} \) is convex. Namely, if \( P_1 \subset P_2 \subset P_3 \) and \( P_1, P_2, P_3 \in \mathcal{P}_1 \), and \( P_2 \in D^{4r}_g \) is good, then \( P_2 \in \mathcal{P}_1 \).

Set
\[ B_\mathcal{P}(f, g) \equiv \sum_{(P_1, P_2) \in \mathcal{P}} \mathbb{E}^r_{P_1} \Delta^r_{P_1} g \cdot \langle R_r(QF, 2P_1), \Delta^r_{P_2} f \rangle_\sigma, \]
and let \( \mathcal{N}_\mathcal{P} \) be the norm of the bilinear form \( B_\mathcal{P} \), that is the best constant in the inequality
\[ |B_\mathcal{P}(f, g)| \leq \mathcal{N}_\mathcal{P} \|g\|_{\tau} \|f\|_\sigma. \]
Thus, it suffices to show that \( \mathcal{N}_\mathcal{P} \lesssim \mathcal{A} \) for all admissible \( \mathcal{P} \).

With an abuse of language, we say that admissible collections \( \mathcal{P}^t, t \geq 0 \), are orthogonal if for any \( s \neq t \), then \( \mathcal{P}_s^t \cap \mathcal{P}_t^s = \emptyset \) and \( \mathcal{P}_s^t \cap \mathcal{P}_t^t = \emptyset \). A simple lemma is then an estimate of the norm of a form which is the union of orthogonal collections.

Lemma 5.4. Given orthogonal admissible collections \( \mathcal{P}^t, t \geq 0 \),
\[ \mathcal{N}_{\bigcup_t \mathcal{P}^t} \leq \sqrt{2} \sup_t \mathcal{N}_{\mathcal{P}^t}. \]

Proof. Let \( \Pi^r_i \) be the Haar projection in \( L^2(\mathbb{R}; \sigma) \) onto the span of the functions \( \{ h^r_J : J \in \mathcal{P}_s^2 \} \). And, let \( \Pi^r_i \) be the Haar projection in \( L^2(\mathbb{R}^+; \tau) \) onto the span of the functions \( \{ h^r_I : I \in \mathcal{P}_s^1 \} \).

The projections \( \Pi^r_i \) are indeed orthogonal in \( L^2(\mathbb{R}; \sigma) \). Concerning the projections \( \Pi^r_i \), note that an interval \( I \) has two children, and could be in a distinct collection \( \mathcal{P}^t_1 \) for two distinct choices of \( t \). Therefore, we have
\[ \sum_t \|\Pi^r_i g\|_\sigma^2 \leq 2\|g\|_\sigma^2. \]
The Lemma is completed by estimating
\[ |B_{\mathcal{P}, \mathcal{P}}(f, g)| = \left| \sum_t B_{\mathcal{P}, \mathcal{P}}(f, g) \right| \]
\[ \leq \sum_t |B_{\mathcal{P}, \mathcal{P}}(\Pi^g_t f, \Pi^g_t g)| \]
\[ \leq \sum_t \mathcal{N}_{\mathcal{P}} \|\Pi^g_t g\|_r \|\Pi^g_t f\|_\sigma \]
\[ \leq \sup_t \mathcal{N}_{\mathcal{P}} \times \left[ \sum_t \|\Pi^g_t g\|^2_r \times \sum_t \|\Pi^g_t f\|^2_\sigma \right]^{1/2} \]
\[ \leq \sqrt{2} \sup_t \mathcal{N}_{\mathcal{P}} \cdot \|g\|_r \|f\|_\sigma. \]
\[ \Box \]

The critical notion of size serves as a crude approximation to the norm of the bilinear form \( B_{\mathcal{P}} \), and it has to be defined with some care. We begin with a supplemental definition. For an interval \( I \in \mathcal{D}_g^4 \), set \( \mathcal{K}_I \) to be the maximal cubes \( K \in \mathcal{D}_g^4 \) such that \( 10K \subset I \). The seemingly too severe restrictions on the Haar supports of \( f \) and \( g \) and the distinction between \( \mathcal{D}_g^4 \) and \( \mathcal{D}_g^{4r} \) have been imposed to achieve this proposition.

**Proposition 5.5.** Let \( I \in \mathcal{D}_g^{4r} \) and \( J \in \mathcal{D}_g^{4r} \). Suppose that \( J \in_{4r} I \), and \( J \) is good. Then there is a unique \( K \in \mathcal{K}_{I_J} \) with \( J \subset \mathcal{K} \).

**Proof.** Since \( \mathcal{K}_{I_J} \) is a partition of \( I_J \), there is some \( K \in \mathcal{K}_{I_J} \) such that \( K \cap J \neq \emptyset \). Now, \( J \in_{4r} I \) implies \( 2^{4r}|J| \leq |I| \), so the conclusion is obvious if \( 2^r|J| \leq |K| \).

But, aside from these considerations, goodness of \( J \) implies that \( J \) is contained in some \( K \in \mathcal{K}_{I_J} \), since goodness is a more stringent condition than membership in \( \mathcal{K}_{I_J} \). And, by the above, we need only consider the case of \( |K| \leq 2^{-3r+1}|I| \). But then,
\[ |J|/|K|^{1-\epsilon} \leq \text{dist}(J, \partial I) \leq |K| + \text{dist}(K, \partial I) \leq (2^r + 1)|K|, \]
which yields
\[ |J|/|K|^{1-\epsilon} \leq (2^r + 1)[|K|/|I|]^{1-\epsilon} \leq (2^r + 1)2^{-3r+1}. \]
And, from this the proposition follows. \[ \Box \]

These definitions are essential,
\[ \lambda = \lambda_\mathcal{P} \equiv \sum_{P_2 \in \mathcal{P}_2} \langle t, h_{P_2}^\sigma \rangle^2 \delta_x \eta_{P_2}, \]
\[ \text{size}(\mathcal{P})^2 \equiv \sup_{t \in T_{\mathcal{P}} : \tau(Q_I) > 0} \mathcal{T}_\tau(Q_F \setminus Q_I)(x_{Q_I})^2 \frac{\lambda(\text{Saw}_P I)}{\tau(Q_I)|I|^2}, \]
\[ \text{Saw}_P I \equiv \bigcup \{ x_{Q_{P_2}} : P_2 \in \mathcal{P}_2 : P_2 \subset \mathcal{K}_I \}, \]
\[ \mathcal{T}_\mathcal{P} \equiv \bigcup \{ \mathcal{K}_I : I \in \mathcal{P}_1 \}. \]

The measure \( \lambda \) is derived from the energy terms, but we do not have scale modifications above. It is used in the definition of size, which also uses a (not quite standard) sawtooth-type definition associated to Carleson measure estimates. Besides the sawtooth region being a discrete set, we caution the reader that (a) we do not have Carleson measures in this
argument, and (b) the distinction between this definition and the standard definition is important, as it stems from the formulation of the energy stopping condition. Finally, the definition of size is a supremum over the collection \( \mathcal{T} \), which is again motivated by the energy stopping conditions, as evidenced by the next proposition which permits the base step in our recursion.

**Proposition 5.6.** There holds for admissible \( \mathcal{P} \), size(\( \mathcal{P} \)) \( \lesssim R \).

**Proof.** Consider an \( I \in \mathcal{T} \) with \( \pi_F I \neq F \). Then, \( \lambda(\text{Saw}_ retard I) = 0 \), by the definition of admissibility, and there is nothing to test. Otherwise, the interval \( I \) must fail the Energy Stopping condition in \( \S 4.2 \). This latter condition is phrased in terms of the Whitney collection \( \mathcal{W}I \). Now, notice that each good \( J \subsetneq I \) is contained in some \( K \in \mathcal{W}I \). By the definition of the sawtooth region,

\[
\tau(Q_F \setminus Q_I)(x_{Q_I})^2 \frac{\lambda(\text{Saw}_ retard I)}{|I|^2} \lesssim \int_{\mathbb{R}^2} \frac{(Q_F \setminus Q_I)(x)(dx)}{|I| + \text{dist}(x, Q_I)^2} \leq \sum_{K \in \mathcal{W}I} |K|^2 E(\sigma, K)^2 \sigma(K) \lesssim \mathcal{R}^2 \tau(Q_I).
\]

The first inequality is trivial. The second follows from the uniform estimate below, for all \( K \in \mathcal{W}I \),

\[
|K|^2 + \text{dist}(x, Q_K)^2 \lesssim |I|^2 + \text{dist}(x, Q_I)^2, \quad x \in Q_F \setminus Q_I.
\]

The last inequality is by the Energy Stopping rule. Hence size(\( \mathcal{P} \)) \( \lesssim \mathcal{R} \). \( \square \)

This is then the main lemma.

**Lemma 5.7.** [Size Lemma] Any admissible collection \( \mathcal{P} \) admits a decomposition into collections \( \mathcal{P}^{\text{big}} \cup \mathcal{P}^{\text{small}} \), so that on the one hand, \( \mathcal{N}_{\mathcal{P}^{\text{big}}} \lesssim \text{size}(\mathcal{P}) \), and on the other, \( \mathcal{P}^{\text{small}} \) is the union of orthogonal collections \( \bigcup_{t \geq 0} \mathcal{P}_t^{\text{small}} \), with

\[
(5.8) \quad \sup_{t \geq 0} \text{size}(\mathcal{P}_t^{\text{small}}) \leq \frac{1}{4} \text{size}(\mathcal{P}).
\]

**Proof of (5.3).** The form \( B_{\text{stop}}(f, g) = B_{\mathcal{P}_0}(f, g) \), for an admissible collection \( \mathcal{P}_0 \) which then satisfies size(\( \mathcal{P}_0 \)) \( \lesssim \mathcal{R} \). By recursive application of the Size Lemma, we have that \( \mathcal{P}_0 \) is the union of collections \( \mathcal{P}_{n,t} \), where (a) \( n, t \geq 1 \), (b) for fixed \( n \geq 1 \) the collections \( \{ \mathcal{P}_{n,t} : t \geq 1 \} \) are admissible and orthogonal, and (c) we have the inequality

\[
\mathcal{N}_{\mathcal{P}_{n,t}} \lesssim 2^{-2n} \text{size}(\mathcal{P}_0), \quad n, t \geq 0.
\]

In view of Lemma 5.4, this gives us a clearly summable estimate in \( n \). \( \square \)

### 5.2. Main Construction

The tool to achieve the decomposition of the Size Lemma is the collection \( \mathcal{L} = \bigcup_{t \geq 0} \mathcal{L}_t \), the latter defined recursively. Set \( S \equiv \text{size}(\mathcal{P}) \), and take \( \mathcal{L}_0 \) to be the minimal intervals \( L \in \mathcal{P}_0 \) such that

\[
(5.9) \quad \tau(Q_F \setminus Q_L)(x_{Q_L})^2 \frac{\lambda(\text{Saw}_ retard L)}{|L|^2} \geq c^2 S^2 \tau(Q_L).
\]
Here, $0 < c < 1$ will be sufficiently small, but absolute. There must be such intervals, by the definition of size. Then, for $t > 0$, inductively define $\mathcal{L}_t$ to be the minimal intervals $L$ such that

\begin{equation}
\lambda(\text{Saw}_P L) \geq (1 + c^2) \sum_{\substack{L' \in \mathcal{L}_{t-1} \\ L' \subset L}} \lambda(\text{Saw}_P L').
\end{equation}

The collection $\mathcal{P}_{\text{small}}$ is given as follows. First, set

$$ \mathcal{P}_{\text{small}}^0 \equiv \left\{ (P_1, P_2) \in \mathcal{P} : \tilde{P}_1 \text{ does not have a parent in } \mathcal{L} \right\}. $$

And, for $L \in \mathcal{L}$, set

$$ \mathcal{P}_{\text{small},L} \equiv \left\{ (P_1, P_2) : \tilde{\pi}_L P_2 = \pi_L \tilde{P}_1 = L, \tilde{P}_1 \not\subset L \right\}. $$

Here, we are defining $\tilde{\pi}_L P_2$ to be the minimal element of $L \in \mathcal{L}$ with $P_2 \not\subset L$. Note that (a) if $P_2 \subset L$, but $P_2 \not\subset \pi_L L$, then the point $x_{Q_{P_2}} \not\in L$, and (b) for $(P_1, P_2) \in \mathcal{P}$, the condition $\pi_L \tilde{P}_1 = L$ implies $P_2 \not\subset \pi_L P$. Momentarily, we will need this additional notation. Set $\tilde{\pi}^1_L P_2 \equiv \tilde{\pi}_L P_3$ and inductively define $\tilde{\pi}^{t+1}_L P_2$ to be the minimal member of $\mathcal{L}$ that strictly contains $\tilde{\pi}^t_L P_2$.

We check that the ‘small’ collection meets the required conditions.

**Proof of (5.8).** Admissibility requires conditions of convexity in $\mathcal{P}_1$, holding $P_2$ fixed, which follow from the definitions above, and the corresponding properties of $\mathcal{P}$. Clearly, the collections defined above are orthogonal in $\mathcal{L}$. A given interval either does not have a parent in $\mathcal{L}$, or has a unique parent in $\mathcal{L}$.

We verify that the collections above have small size. For $\mathcal{P}_{\text{small}}$, since each interval $\tilde{P}_1$ must fail (5.9), the size of this collection is smaller by the factor $c^2$. We turn to $\mathcal{P}_{\text{small},L}$. In the case that $L \in \mathcal{L}_0$, this follows immediately from the definition of $\mathcal{L}_0$ in (5.9). In the case that $L \in \mathcal{L}$ is not minimal, then $L$ has $\mathcal{L}$-children, which will appear below. We have for $I \in \mathcal{T}_{\mathcal{P}_{\text{small},L}}$; and $\mathcal{L}_I \equiv \left\{ L' \in \mathcal{L} : \pi^1_L L' = L, L' \subset I \right\}$ that

$$ \frac{\mathcal{T}_r(Q_F \setminus Q_I)(x_{Q_I})^2}{|I|^2} \lambda(\text{Saw}_{\mathcal{P}_{\text{small},L}} I) \leq \frac{\mathcal{T}_r(Q_F \setminus Q_I)(x_{Q_I})^2}{|I|^2} \lambda(\text{Saw}_{\mathcal{P}} I \setminus \bigcup_{L' \in \mathcal{L}_I} \text{Saw}_{\mathcal{P}} L') \leq c^2 \frac{\mathcal{T}_r(Q_F \setminus Q_I)(x_{Q_I})^2}{|I|^2} \sum_{L' \in \mathcal{L}_I} \lambda(\text{Saw}_{\mathcal{P}} L') \leq c^2 \mathcal{T}_r(Q_F \setminus Q_I)(x_{Q_I})^2 |I|^2 \lambda(\text{Saw}_{\mathcal{P}} I) \leq c^2 S^2 \tau(Q_I). $$

Here, note that we dominated $\lambda(\text{Saw}_{\mathcal{P}} I)$, and then use (5.10), the selection rule for $\mathcal{L}$.

\[ \square \]

5.3. **The Big Collections.** To understand the collection $\mathcal{P}_{\text{big}} \equiv \mathcal{P} \setminus \mathcal{P}_{\text{small}}$, note that if $(P_1, P_2) \in \mathcal{P}$ is such that $\tilde{P}_1$ has parent $L \in \mathcal{L}$, then we must have $P_2 \not\subset \pi_L P$. Thus, $\tilde{\pi}_L P_2$ is well-defined, and for some integer $t \geq 1$, necessarily $\tilde{\pi}^{t+1}_L P_2 = \pi_L \tilde{P}_1 = L$.

The decomposition of $\mathcal{P} \setminus \mathcal{P}_{\text{small}}$ proceeds as follows. Consider first

$$ \mathcal{P}^1_L \equiv \left\{ (P_1, P_2) \in \mathcal{P}_L : \tilde{\pi}^1_L P_2 = \tilde{P}_1 = L \right\}. $$

Clearly, these collections are mutually orthogonal. And, we have
Lemma 5.11. There holds

\[ |B_{P_1}(f, g)| \lesssim S \|f\|_\sigma \|g\|_\tau. \]

Proof. Define \( M_L \) by \( M_L \equiv \sup_{I \supset L} |E_{Q_I} g| \). Then, \( M_L^2 \tau(Q_L) \lesssim \|g\|_\tau^2 \). For each \( K \in \mathcal{K}_L \), we can estimate by the definition of \( \mathcal{K}_L \), and monotonicity (3.6), that

\[
\left| \sum_{(P_1, P_2) \in \mathcal{P}^L_1} \mathbb{E}^\sigma_{Q_{P_1}} g \cdot \langle R_\tau^*(Q_F - Q_{P_1}), \Delta_{P_2} f \rangle_\tau \right|
\leq \sum_{P_2 \in (\mathcal{P}^L_1)_2} \sum_{P_1 \in (\mathcal{P}^L_1)_1} \mathbb{E}^\sigma_{Q_{P_1}} g \cdot \langle R_\tau^*(Q_F - Q_{P_1}), \Delta_{P_2} f \rangle_\tau 
\leq M_L T_\tau(Q_F \setminus Q_K) (x_{Q_L}) \sum_{P_2 : P_2 \subset K} \left| \frac{t}{|K|} h_{P_2}^\sigma \right| |\hat{f}_\sigma(P_2)| 
\leq SM_L \tau(Q_K)^{1/2} \left[ \sum_{P_2 : P_2 \subset K} \left| \hat{f}_\sigma(P_2) \right|^2 \right]^{1/2} 
\lesssim SM_L \cdot \tau(Q_L)^{1/2} \|f\|_\sigma \lesssim S \|g\|_\tau \|f\|_\sigma.
\]

Notice that the convexity in \( P_1 \) imposed by admissibility is critical to replace the martingale differences on \( g \) by the term \( M_L \), in the application of the monotonicity principle. Then, we use Cauchy–Schwarz in \( P_2 \), and importantly, the definition of size to gain the factor \( S \tau(Q_K)^{1/2} \) above. It is clear that quasi-orthogonality permits the bound

\[
M_L \sum_{K \in \mathcal{K}_L} \tau(Q_K)^{1/2} \left[ \sum_{P_2 : P_2 \subset K} \left| \hat{f}_\sigma(P_2) \right|^2 \right]^{1/2} \lesssim SM_L \cdot \tau(Q_L)^{1/2} \|f\|_\sigma \lesssim S \|g\|_\tau \|f\|_\sigma.
\]

The proof is complete.

The second contribution is the collections which are ‘separated’ by \( \mathcal{L} \). For integers \( t \geq 1 \),

\[ \mathcal{P}^L_{t,L} \equiv \{(P_1, P_2) \in \mathcal{P} : \pi_{\mathcal{L}}^{t+1} P_2 = \pi_{\mathcal{L}}(\tilde{P}_1) = L \}. \]

Here, we require that \( P_2 \) have a parent which is \( t \)-steps below that of \( \tilde{P}_1 \) in the \( \mathcal{L} \)-tree. It is clear that these are orthogonal in \( L \in \mathcal{L} \). Finally, the norms of the collections has exponential decay in \( t \).

Lemma 5.12. For \( L \in \mathcal{L} \) and \( t \geq 2 \),

\[ \mathcal{N}_{\mathcal{P}^L_{2,t}} \lesssim (1 + c^2)^{-t/2} S. \]

It follows immediately that

\[ \sum_{t=1}^{\infty} \mathcal{N}_{\cup_{L \in \mathcal{L}} \mathcal{P}^L_{2,t}} \lesssim S \sum_{t=1}^{\infty} (1 + c^2)^{-t/2} \lesssim S. \]

The Lemma above depends upon properties of the collection \( \mathcal{T}_\mathcal{P} \). Namely, that the testing intervals for \( \mathcal{P}^L_{2,t,L} \) are approximately at the same level as \( L \) in the \( \mathcal{L} \)-tree.

Proposition 5.13. For each interval \( K \in \mathcal{T}_{\mathcal{P}^L_{2,t,L}} \), there holds \( L \subset \pi_{\mathcal{L}}^3 K \).
Proof. Suppose that $K \in \mathcal{K}_{\tilde{P}_1}$, where $\pi_P \tilde{P}_1 = L$. The primary points to exploit are that the cubes $\tilde{P}_1$ are the children of good cubes, and that
\[
10|K| \leq \text{dist}(K, \partial \tilde{P}_1) \leq 2^r 10|K|,
\]
as follows from the definition of $\mathcal{K}_{\tilde{P}_1}$. Proceeding by contradiction, suppose that there are $L_1, L_2 \in \mathcal{L}$ with
\[
(5.14) \quad K \subsetneq L_1 \subsetneq L_2 \subsetneq \tilde{P}_1.
\]
Suppose that $L_s \in \tilde{P}_1$, for either $s = 1, 2$. Since these intervals are from a grid with scales separated by $4r$, we then have $L_s \in_{4r} \tilde{P}_1$, and then goodness of the parent of $L_s$ implies
\[
|L_s|^{\varepsilon} |\tilde{P}_1|^{1-\varepsilon} \leq \text{dist}(L_s, \partial \tilde{P}_1) \leq \text{dist}(K, \partial \tilde{P}_1) \leq 10 \cdot 2^r |K|
\]
which is a contradiction, since it implies $2^{(1-\varepsilon) r} \leq 2^r$.

Thus, we must have $L_s \not\subsetneq \tilde{P}_1$, for $1 \leq s \leq 2$. Let $I_s \in \tilde{P}_1$ be such that $L_s \in \mathcal{K}_{I_s}$. If $I_1 \subsetneq \tilde{P}_1$, then $I_1 \in_{4r} P_1$, and we again see a contradiction to $K \in \mathcal{K}_{I_1}$, namely
\[
|I_1|^{\varepsilon} |\tilde{P}_1|^{1-\varepsilon} \leq \text{dist}(I_1, \partial \tilde{P}_1) \leq \text{dist}(K, \partial \tilde{P}_1) \leq 10 \cdot 2^r |K|.
\]

Assume that $I_1 \not\subsetneq_{4r} I$. Equality $I_1 = I_2$ cannot hold, since $K \subsetneq L_1$. We must have $\tilde{P}_1 \in_{4r} I_1 \in_{4r} I_2$, and that means
\[
|\tilde{P}_1|^{\varepsilon} |I_2|^{1-\varepsilon} \leq \text{dist}(\tilde{P}_1, \partial I_2) \leq \text{dist}(L_2, \partial I_2) \leq 10 2^r |\tilde{P}_1|.
\]

And this final contradiction proves that (5.14) cannot hold, and this proves the proposition. \hfill \Box

Proof of Lemma 5.12. We verify that size($\mathcal{P}_{L,t}^2$) has exponential decay in $t$. It suffices to consider the case of $t \geq 9$. For $I \in \mathcal{D}_y^r$, with $I \subset F$,
\[
\lambda(\text{Saw}_{\mathcal{P}_{L,t}^2} I) \leq \sum_{L' \in \mathcal{L} : L' \subset L} \lambda(\text{Saw}_{\mathcal{P}} L')
\]
\[
\leq (1 + c^2)^{-1} \sum_{L' \in \mathcal{L} : L' \subset L} \lambda(\text{Saw}_{\mathcal{P}} L')
\]
\[
: \leq (1 + c^2)^{-t+4} \sum_{L' \in \mathcal{L} : L' \subset L} \lambda(\text{Saw}_{\mathcal{P}} L') \leq (1 + c^2)^{-t+4} \lambda(\text{Saw}_{\mathcal{P}} I).
\]

One should be careful to note that Proposition 5.13 was used in the first inequality, to impose the restriction $\pi_{L-t-3} L' = L$ on the right. Then, the construction of $L$, namely (5.10), and induction, is used to gain the factors of $(1 + c^2)^{-1}$ above. Therefore, for $t \geq 9$,
\[
(1 + c^2)^t T_\tau(Q_F \setminus Q_I)(x_Q)^2 \frac{\lambda(\text{Saw}_{\mathcal{P}_{L,t}^2} I)}{|I|^2} \lesssim T_\tau(Q_F \setminus Q_I)(x_Q)^2 \frac{\lambda(\text{Saw}_{\mathcal{P}} I)}{|I|^2} \lesssim S^2 \tau(Q_I).
\]

And this shows in particular that
\[
(5.15) \quad \text{size}(\mathcal{P}_{L,t}^2) \lesssim (1 + c^2)^{-t/2} S.
\]
We turn to the estimation of the norm of \( B_{\mathcal{P}^2_{L,t}} \), namely that it is no more than \((1+c^2)^{-t/2}S\). The key point is that letting \( \mathcal{S} \) be the \( \mathcal{L} \)-children of \( L \), there holds \( P_2 \in_r S \subset \tilde{P}_1 \) for all \((P_1,P_2) \in \mathcal{P}^2_{L,t} \equiv Q\).

Assume that the Haar support of \( g \) is contained in \( Q_1 \). We construct stopping data for \( g \) with some specificity. The root of \( G \) is \( F \), and then in the recursive step, if \( I \in \mathcal{G} \) is minimal, add in the maximal children \( I' \subset I \) such that \( I \in \mathcal{D}_g, \pi I \) is good and \( \mathbb{E}^g_{\mathcal{Q}_I} |g| \geq 10 \mathbb{E}^r_{\mathcal{Q}_I} |g| \). The unusual requirement here is that \( \pi I \) is good, but this can be done because \( g \) is good.

Define \( \bar{g}P_2 \) to be the minimal element \( G \in \mathcal{G} \) with \( P_2 \subset G \) and \( P_2 \in_r \pi G \), and let \( \Pi'_G \) be the projection onto the Haar coefficients \( P_2 \in Q_2 \) with \( \bar{g}P_2 = G \). This basic estimate, which we combine with monotonicity, holds for all intervals \( P_2 \)

\[
(5.16) \quad \left| \sum_{P_1:(P_1,P_2) \in Q} \mathbb{E}^\sigma_{Q_{P_1}} g \cdot (Q_F - Q_{P_1}) \right| \lesssim \mathbb{E}^\tau_{Q_{P_2}} |g| \cdot Q_F.
\]

It is critical that admissibility for \( Q \) implies convexity in the \( P_1 \) coordinate, holding \( P_2 \) fixed, so that the stopping average appears above.

There are three sums to consider, where we sum over the pairs:

**Case A:** Those pairs \( P_2 \in_r \bar{g}P_2 \subset \tilde{P}_1 \). So, in particular, the interval \( \bar{g}P_2 \) separates \( P_2 \) and \( \tilde{P}_1 \).

**Case B:** Those pairs \( P_2 \in_{A^r} P_1 \) and \( \tilde{P}_1 \subset \bar{g}P_2 \). We further split this case into (1) \( \pi S P_2 \in Q_1 \), and (2) \( \pi S P_2 \notin Q_1 \).

Call the resulting collections of pairs \( Q_A \), \( Q_{B,1} \) and \( Q_{B,2} \).

**Case A.** Hold an interval \( G \in \mathcal{G} \) fixed. Below, let \( \Pi'_G \equiv \sum_{P_2: \bar{g}P_2 = G} \Delta^\sigma_{P_2} \), and similarly define \( \Pi'_K \) for \( K \in \mathcal{K}_G \). Then,

\[
\Xi_A(G) \equiv \left| \sum_{P_1,P_2 \in Q_A} \mathbb{E}^\sigma_{Q_{P_1}} g \cdot \left\langle R^x_{\tau}(Q_F - Q_{P_1}), \Delta^\sigma_{P_2} f \right\rangle_{\tau} \right|
\]

\[
= \left| \sum_{K \in \mathcal{K}_G} \mathbb{E}^\sigma_{Q_{P_1}} g \cdot \left\langle R^x_{\tau}(Q_F - Q_{P_1}), \Delta^\sigma_{P_2} f \right\rangle_{\tau} \right|
\]

\[
\lesssim \mathbb{E}^\tau_{Q_{P_2}} |g| \cdot \sum_{K \in \mathcal{K}_G} \mathbb{E}^\tau_{Q_K} |Q_F \setminus Q_K(x_{Q_K})| \sum_{P_2: \bar{g}P_2 = G} \left\langle \frac{t}{|K|}, h^\sigma_{P_2} \right\rangle_{\sigma} |\hat{f}_{\sigma}(P_2)|
\]

\[
\lesssim \text{size}(Q) \mathbb{E}^\tau_{Q_{P_2}} |g| \cdot \sum_{K \in \mathcal{K}_G} \tau(Q_K)^{1/2} \|\Pi'_K f\|_{\sigma}
\]

\[
(5.17) \quad \lesssim \text{size}(Q) \mathbb{E}^\tau_{Q_{P_2}} |g| \cdot \tau(Q_G)^{1/2} \|\Pi'_G f\|_{\sigma}.
\]

The first equality is important: Each \( P_2 \) is contained in some \( K \in \mathcal{K}_G \). We use (5.16), to get the stopping value and the monotonicity principle (3.6). Then, Cauchy–Schwarz in \( K \in \mathcal{K}_G \), noting that the elements of \( \mathcal{K}_G \) appear in the supremum defining \( \text{size}(Q) \), whence we have
the inequality
\[
T_T(Q_F \setminus Q_I)(x_{Q_K}) \left[ \sum_{P_2 : \ell P_2 \in G} \left| \frac{t}{|I|} , h_{P_2}^\sigma \right|_\sigma^2 \right]^{1/2} \lesssim \text{size}(Q) \tau(Q_K)^{1/2}.
\]

Finally, use the quasi-orthogonality bound (4.6) and (5.15) to see that
\[
\sum_{G \in \mathcal{G}} \Xi_A(G) \lesssim (1 + c^2)^{-t/2} S \|g\|_{\sigma} \|f\|_{\sigma}.
\]

**Case B1.** Hold a $G \in \mathcal{G}$ fixed, and consider $S \in \mathcal{S} \cap \mathcal{Q}_1$ which has $\mathcal{G}$-parent $G$. Then, we can repeat, with only small notational changes, the argument that lead to (5.17) to conclude that
\[
\Xi_{B,1}(G, S) \equiv \sum_{(P_1, P_2) \in \mathcal{Q}_{B,1}, \#_G P_1 = S} \mathbb{E}_{Q_{P_1}} g \cdot \left( R_T^\sigma(Q_F - Q_{P_1}), \Delta_{P_2}^\sigma f \right)_\tau \lesssim \text{size}(Q) \mathbb{E}_{G} |g| \tau(Q_S)^{1/2}\|\Pi^g_{S} f\|_{\sigma}.
\]
Using quasi-orthogonality, (5.15), and the fact that the intervals $S \in \mathcal{S}$ are disjoint, we can sum
\[
\sum_{G \in \mathcal{G}} \sum_{S \in \mathcal{S} \cap \mathcal{Q}_1} \Xi_{B,1}(G, S) \lesssim (1 + c^2)^{-t/2} S \|f\|_{\sigma} \|g\|_{\tau}.
\]

**Case B2.** We are concerned with $S \in \mathcal{S}$ which is not contained in $\mathcal{Q}_1$. Then, there is an interval $I$ which is the $\mathcal{Q}_1$-parent of $S$. We claim that $10S \subset I$. Now, $S \in \mathcal{K}_I$, for some $I' \in \mathcal{P}_1$. If $I' \subset I$, our claim is obvious. So, we must have $I \subsetneq I'$. But then, $2^{4r}|I| \leq |I'|$, and there holds, since both $I$ and $I'$ are children of good intervals,
\[
\frac{1}{2} |I'|^{1-\varepsilon} \leq \text{dist}(\pi I, \partial I') \leq 10 \cdot 2^r |S| \leq 10 \cdot 2^r |I|.
\]
And, this is a contradiction.

Using notation similar to Case A, we have
\[
\Xi_{B,2}(G, S) \equiv \sum_{(P_1, P_2) \in \mathcal{Q}_{B,2}, P_2 \in S} \mathbb{E}_{Q_{P_1}} g \cdot \left( R_T^\sigma(Q_F - Q_{P_1}), \Delta_{P_2}^\sigma f \right)_\tau \lesssim \mathbb{E}_{Q_G} |g| \tau(Q_F \setminus Q_S)(x_{Q_S}) \sum_{P_2 : P_2 \in S} \left| \frac{t}{|S|} , h_{P_2}^\sigma \right|_\sigma \|f_\sigma(P_2)\| \lesssim (1 + c^2)^{-t/2} S \cdot \mathbb{E}_{Q_G} |g| \tau(Q_S)^{1/2}\|\Pi^g_{S} f\|_{\sigma}.
\]
One should be careful to note that the convexity in $P_1$ is used in the first inequality, followed by (5.16), and the monotonicity principle (3.6). The essential points are (a) we have enough
separation in the argument of the Riesz transform to apply the monotonicity principle and (b) that the proof of (5.15) shows that

\[ T_\tau Q_F(x_{QS})^2 \sum_{P_2 : P_2 \subset S} \left\langle \frac{t}{|S'|}, h_{P_2}^\sigma \right\rangle^2 \lesssim (1 + c^2)^{-1} S^2 \tau(Q_S). \]

The disjointness of the intervals $S \in \mathcal{S}$, and quasi-orthogonality clearly allow us to deduce

\[ \sum_{G \in \mathcal{G}} \sum_{S \in \mathcal{S}} \Xi_{B,2}(G, S) \lesssim (1 + c^2)^{-1/2} \|f\|_\sigma \|g\|_\tau. \]

\[ \square \]

6. The Form $B_{\text{below}}$

The analysis of the form $B_{\text{below}}(f, g)$ defined in (4.4) has an analysis that is similar to that of the form $B_{\text{above}}$, and even more similar to that of the analogous forms in the two weight inequality for the Hilbert transform [11, 15]. We will state the highlights of the analysis.

The first step is the construction of stopping $f$ intervals $\mathcal{F}$. Recall the formula (4.1), concerning the Haar support of $f$, and the definition of $\mathcal{D}_f$. The Haar support of $f$ are intervals $\pi I$, $I \in \mathcal{D}_f$ such that $\pi I$ is good.

In the initial stage, we take the maximal elements of $f$ to be the children of the maximal intervals in the Haar support of $f$, and for $I \in \mathcal{F}$. In the inductive stage, if $F \in \mathcal{F}$ is minimal, we add to $\mathcal{F}$ the maximal sub-children $F' \subset F$, with $F' \in \mathcal{D}_f$, such that either

1. (A large average) $E_{F'}|f| \geq 10E_F|f|$, 
2. (Energy Stopping) $\sum_{K \in W} F'E_{F'}(\sigma \cdot (F \setminus K), K)^2 E(\tau, Q_K)^2 \tau(Q_K) \geq C_0 R^2 \sigma(F')$.

Recall (3.26), and Lemma 3.27 and that $C_0$ is sufficiently large constant.

Define Haar projections by

\[ H^\sigma_f \equiv \Delta_{\pi F} f + \sum_{J : \pi J = F} \Delta_J f, \]
\[ \tilde{H}^\tau_f g \equiv \sum_{I : \pi I = F} \Delta_{Q_I} g. \]

We have the quasi-orthogonality inequality, compare to (4.6),

\[ \sum_{F \in \mathcal{F}} \{ E_{F'}|f| \cdot \sigma(F')^{1/2} + \|H^\sigma_f\|_\sigma \} \|\tilde{H}^\tau_f g\|_\tau \lesssim \|f\|_\sigma \|g\|_\tau. \]

As before, we will use this inequality as written, and with different choices of the orthogonal projections $\tilde{H}^\tau_f$.

This is the global-to-local reduction for $B_{\text{below}}(f, g)$. The sum is over intervals that are ‘separated’ by $\mathcal{F}$.

**Lemma 6.1.** [Global to Local Reduction, II] The following estimate holds:

\[ \left| \sum_{J} \sum_{I : I \subseteq \pi J, I \in \mathcal{D}_f} E_{J_I} \Delta_J f \cdot \langle R_{\sigma} J_I, \Delta_{Q_I} g \rangle_\tau \right| \lesssim \mathcal{R} \|f\|_\sigma \|g\|_\tau. \]
Using the exchange argument, the main step in the proof of this estimate is the functional energy inequality that we state next.

6.1. **Functional Energy Inequality, II.** We turn to the functional energy inequality, which is a two weight inequality for the Poisson operator. The latter is defined in terms of this measure on \(\mathbb{R}^2_+\) derived from \(\tau\). Define

\[
\mu_K \equiv \sum_{I : \pi^2_F I \supseteq F} \left\| \frac{\Delta^r_{Q_I}}{t} \right\|^2_{K}, \quad F \in \mathcal{F}, \ K \in \mathcal{W}_F.
\]

This is the functional energy inequality in this case.

**Lemma 6.2.** [Functional Energy Inequality, II] The following inequality holds

\[
\sum_{F \in \mathcal{F}} \sum_{K \in \mathcal{W}_F} P_{\sigma}(\phi \cdot F^c)(x_{Q_K})^2 \mu_K \lesssim \mathcal{R}^2 \|\phi\|^2_{L^2(\mathbb{R}, \sigma)}.
\]

The proof is a modification of that for the first functional energy inequality.

6.2. **The Local Estimate.** The last step is to control the local form, namely, to prove the following.

**Lemma 6.3.** For each \(F \in \mathcal{F}\), we have

\[
|B_{\text{stop}}^F(f, g)| \lesssim \mathcal{R} \|f\|_{\sigma} \|g\|_{\tau}.
\]

After a second application of the exchange argument, it remains to consider the stopping form:

\[
B_{\text{stop}}^F(f, g) \equiv \sum_{I : \pi_F I = F} \mathbb{E}^\sigma_{\tau} \Delta^r_{f} \cdot \sum_{J : \pi_{4r} J = F} \langle R_{\sigma}(F-I_J), \Delta^r_{g} \rangle_{\sigma}.
\]

**Lemma 6.4.** The following estimate is true:

\[
|B_{\text{stop}}^F(f, g)| \lesssim \mathcal{R} \|f\|_{\sigma} \|g\|_{\tau}.
\]

This argument is a variant of the proof of Lemma 5.2.

7. **Elementary Estimates**

This section collects some considerations which, while not completely elementary, rely upon the \(A_2\) condition. Then, reductions are proved, namely the Carleson cube projection in Proposition 4.2, and the reduction to the ‘above’ and ‘below’ projections in Proposition 4.5.

7.1. **Weak-Boundedness.** By the weak-boundedness condition, we mean this estimate.

**Proposition 7.1.** Let \(\sigma\) and \(\tau\) satisfy the \(A_2\) condition (1.3). Then, for any two intervals \(I, J\), intersecting only at their boundaries, we have

\[
|\langle R_{\sigma} f \cdot I, g \cdot Q_J \rangle_{\tau}| \lesssim \mathcal{A}_2^{1/2} \|f\|_{\sigma} \|g\|_{\tau}.
\]

For the proof, we will need the classical result of Muckenhoupt [18], characterizing the two weight Hardy inequality, on the line.
Theorem A. For weights \( \hat{w} \) and \( \hat{\sigma} \) supported on \( \mathbb{R}_+ \),
\[
\left\| \int_0^x f \hat{\sigma}(dy) \right\|_{\hat{w}} \leq \mathcal{B}\|f\|_{\hat{\sigma}},
\]
(7.2)
where \( \mathcal{B}^2 \simeq \sup_{0<r} \int_r^\infty \hat{w}(dx) \times \int_0^r \hat{\sigma}(dy) \).

Proof. We abandon the possibility of cancellation, taking the absolute value of the kernel of the Riesz transform. It is clear that we can assume that \( I \) and \( J \) share a common endpoint, which after a translation, can be taken to be 0. Assume that \( I \) lies to the left of \( J \).

The point of the next steps is to pass to dual formulations of a Hardy inequality with respect to changed measures. Write \( \hat{\sigma}(dx) \equiv \sigma(-dx) \cdot 1_{[0,\infty)} \), and also derive from \( \tau \) a one dimensional weight by setting
\[
\hat{\tau}(s, t) \equiv \int_{s < |x| < t} d\tau, \quad 0 < s < t.
\]
Likewise for \( g \in L^2(Q_J; \tau) \), set \( \hat{g} \) to be the non-negative function with
\[
\int_s^t \hat{g}(u) \ d\tau = \int_{s < |x| < t} |g(x)| \ d\tau, \quad 0 < s < t.
\]
Then, by a conditional expectation argument, \( \|\hat{g}\|_{\hat{\tau}} \leq \|g\|_{\tau} \). Finally, turning to the Riesz transform, we have
\[
|\langle R_\sigma f, g \rangle_{\tau} | = \left| \int_{Q_J} \int_0^\infty \frac{x-t}{|x-t|^2} f(t) g(x) \tau(dx) \sigma(dt) \right|
\]
\[
= \left| \int_0^\infty \int_{Q_J} \frac{x+t}{|x+t|^2} f(-t) g(x) \tau(dx) \hat{\sigma}(dt) \right|
\]
\[
\leq \int_0^\infty \int_0^\infty \frac{1}{s+t} \hat{f}(t) \hat{g}(s) \hat{\tau}(ds) \hat{\sigma}(dt).
\]
Here, \( \hat{f}(t) = f(-t) \). This last form is divided into dual forms.

The first is integration over the region \( \{(s, t) : 0 < t < s\} \), on which \( \frac{1}{s+t} \leq \frac{1}{t} \), and so it suffices to bound the form
\[
\int_0^\infty \hat{f}(t) \int_0^t \frac{g(s) \hat{\tau}(ds) \hat{\sigma}(dt)}{t} \leq \|\hat{f}\|_{\hat{\sigma}} \left[ \int_0^\infty \left[ \int_0^t \frac{\hat{g}(s) \hat{\tau}(ds)}{s} \right]^2 \frac{\hat{\sigma}(dt)}{t^2} \right]^{1/2}.
\]
The last expression is the Hardy inequality with the weight \( \frac{\sigma(dt)}{t^2} \). Whence it suffices to bound the constant \( \mathcal{B}^2 \), given the expression in (7.2). It is a supremum over \( r > 0 \) of
\[
\int_r^\infty \hat{\sigma}(dt) \times \int_r^r \hat{\tau}(ds) \leq \int_{-r}^\infty \frac{r}{t^2} \sigma(dt) \times \frac{1}{r} \int_{Q_{[0,r]}} \tau(dx),
\]
as follows by inspection. But, the last expression is clearly dominated by our \( A_2 \) assumption.

The second expression is integration over the region \( \{(s, t) : 0 < s \leq t\} \), on which \( \frac{1}{s+t} \leq \frac{1}{s} \), and so it suffices to bound the form
\[
\int_0^\infty \frac{\hat{g}(s)}{s} \int_0^s \frac{\hat{f}(t) \hat{\sigma}(dt)}{t} \hat{\tau}(ds) \leq \|\hat{g}\|_{\hat{\tau}} \left[ \int_0^\infty \left[ \int_0^s \frac{\hat{f}(t) \hat{\sigma}(dt)}{s} \right]^2 \frac{\hat{\tau}(ds)}{s^2} \right]^{1/2}.
\]
In this case, the expression in (7.2) is the supremum over \( r > 0 \) of
\[
\int_0^\infty \frac{\sigma(ds)}{s^2} \times \int_r^0 \tilde{\sigma}(dt) \leq \int_{|x| \geq r} \frac{r}{s^2} \tilde{\sigma}(ds) \times \frac{1}{r} \int_0^r \sigma(dt) \lesssim \mathcal{A}_2.
\]
The proof is complete. \( \square \)

**7.2. Proof of Proposition 4.2.** The basic fact is this lemma on martingale differences in \( L^2(\mathbb{R}^2_+; \tau) \) where the dyadic cube \( Q \) does not touch the boundary.

**Lemma 7.3.** Let \( Q \) be a dyadic cube in \( \mathbb{R}^2_+ \), and let \( J \) be a dyadic interval in \( \mathbb{R} \). Under the condition
\[
\text{dist}(Q, \mathbb{R}) \geq \min\{|J|, |Q|^{1/2}\},
\]
then
\[
\left| \langle R_\sigma \Delta^\sigma_J f, \Delta^\sigma_Q g \rangle \right| \lesssim \frac{\min\{|J|, |Q|^{1/2}\}}{(|Q|^{1/2} + \text{dist}(Q, J))^2} \sigma(J)\tau(Q) \| \Delta^\sigma_Q g \|_\tau \cdot |\hat{f}_\sigma(J)|.
\]

**Proof.** If \(|J| \leq |Q|^{1/2}\), we write
\[
\langle R_\sigma \Delta^\sigma_J f, \Delta^\sigma_Q g \rangle = \langle \Delta^\sigma_J f, R^*_\sigma \Delta^\sigma_Q g \rangle_{\sigma} = \int \int_Q \Delta^\sigma_Q g(t - x) \left\{ \frac{x - t}{|x - t|^2} - \frac{x - t_J}{|x - t_J|^2} \right\} \tau(dx) \cdot \Delta^\sigma_J f \sigma(dt).
\]
Here, \( t_J \) is the center of \( J \). By (3.2), this is at most
\[
\frac{|J|}{(|Q|^{1/2} + \text{dist}(Q, J))^2} \| \Delta^\sigma_Q g \|_{L^1(\mathbb{R}^2_+; \tau)} \| \Delta^\sigma_J f \|_{L^1(\mathbb{R}; \sigma)} \lesssim \frac{|J|}{(|Q|^{1/2} + \text{dist}(Q, J))^2} \sigma(J)\tau(Q) \| \Delta^\sigma_Q g \|_\tau \cdot |\hat{f}_\sigma(J)|.
\]
The dual estimate is entirely similar. \( \square \)

We turn to the proof of Proposition 4.2. In fact, we will prove something a little stronger.

**Lemma 7.4.** The estimate below holds:
\[
\sum_{Q} \sum_{J : \text{dist}(Q, \mathbb{R}) \leq \min\{|J|, |Q|^{1/2}\}} \left| \langle R_\sigma \Delta^\sigma_J f, \Delta^\sigma_Q g \rangle \right| \lesssim \mathcal{A}_2^{1/2} \| f \| \| g \|_\tau.
\]

If \( Q \) is not a Carleson cube, then automatically, \( \text{dist}(Q, \mathbb{R}) \geq \min\{|J|, |Q|^{1/2}\} \).

**Proof.** For integers \( s, t \in \mathbb{N} \), and a dyadic cube \( Q \), let \( J(Q, j, k) \) be the dyadic intervals in \( \mathcal{R} \) with \( 2^k |J| = |Q|^{1/2} \), and
\[
2^{k-1} |Q|^{1/2} \leq \text{dist}(J, Q) < 2^k |Q|^{1/2}.
\]
Note that with \( j, k \) fixed, each \( J \) is in \( \mathcal{J}(Q,j,k) \) for a unique \( Q \),
\[
\left[ \sum_{J \in \mathcal{J}(Q,j,k)} \frac{|J|}{\text{dist}(Q,J)^2} \sqrt{\sigma(J)} \hat{f}_\sigma(J) \right]^2 \leq \sum_{J \in \mathcal{J}(Q,j,k)} \frac{|J| \cdot \hat{f}_\sigma(J)^2}{\text{dist}(Q,J)^2} \times \sum_{J \in \mathcal{J}(Q,j,k)} \frac{|J| \cdot \sigma(J)}{\text{dist}(Q,J)^2}
\lesssim 2^{-2j-2k}|Q|^{-1/2} \mathcal{P}(\sigma, I_Q) \sum_{J \in \mathcal{J}(Q,j,k)} \left| \hat{f}_\sigma(J) \right|^2.
\]

Here, in the Poisson term, \( I_Q \) is the projection of \( Q \) onto the \( x_1 \)-axis.

From this, and Cauchy–Schwarz, it follows that
\[
\sum_{Q} \sum_{J \in \mathcal{J}(Q,j,k)} \left| \langle R_\sigma \Delta^\tau f, \Delta^\tau Qg \rangle_\tau \right| \lesssim 2^{-j-k} \sum_{Q} \frac{\tau(Q)^{1/2}}{|Q|^{1/4}} \mathcal{P}(\sigma, I_Q)^{1/2} \| \Delta_{Q}g \|_\tau \cdot \left[ \sum_{J \in \mathcal{J}(Q,j,k)} \hat{f}_\sigma(J)^2 \right]^{1/2}
\lesssim 2^{-j-k} \mathcal{P}_{1/2}(\| f \|_\sigma \| g \|_\tau).
\]

There is the complementary case, which begins this way. For integers \( s, t \in \mathbb{N} \), and a dyadic interval \( J \), set \( Q(J,j,k) \) to be those cubes \( Q \) with \( 2^{2j}|J|^2 = |Q| \), and
\[
2^{k-1}|J| \leq \text{dist}(Q,J) < 2^k|J|.
\]

Note that with \( j, k \) fixed, each \( Q \) is in \( \mathcal{Q}(J,j,k) \) for a unique \( Q \),
\[
\left[ \sum_{Q \in \mathcal{Q}(J,j,k)} \frac{|Q|^{1/2}}{\text{dist}(Q,J)^2} \sqrt{\tau(Q)} \| \Delta^\tau Qg \|_\tau \right]^2 \leq \sum_{Q \in \mathcal{Q}(J,j,k)} \frac{|Q|^{1/2}}{\text{dist}(Q,J)^2} \| \Delta_{Q}g \|_\tau^2 \times \sum_{Q \in \mathcal{Q}(J,j,k)} \frac{|Q|^{1/2} \cdot \tau(Q)}{\text{dist}(Q,J)^2}
\lesssim 2^{-j-2k}|J|^{-1} \mathcal{P}(\tau, J) \sum_{Q \in \mathcal{Q}(J,j,k)} \| \Delta_{Q}g \|_\tau^2.
\]

The conclusion of the argument is then just like above and we omit the details. \( \square \)

7.3. **Proof of Proposition 4.5.** In the proof of Proposition 4.5, it suffices to consider \( g \) in the span of (good) martingale differences \( \Delta_{Q,j} \), namely associated with Carleson cubes. There are quite a few subcases of the proof, all controlled by goodness, and the \( A_2 \) condition.

Define a sub-bilinear form, and several collections of pairs of intervals as follows.
\[
B^P(f,g) \equiv \sum_{(I,J) \in P} \left| \langle R^\tau_\sigma \Delta^\tau_{Q,j} g, \Delta^\tau_{Q} f \rangle_\sigma \right|,
\]
\[
\mathcal{P}_{\text{diagonal}} \equiv \{(I,J) : 3I \cap 3J \neq \emptyset, 2^{-4r}|I| \leq |J| \leq 2^{4r}|I| \},
\]
\[
\mathcal{P}_{\text{far}} \equiv \{(I,J) : 3I \cap 3J = \emptyset \},
\]
\[
\mathcal{P}_{\text{near}} \equiv \{(I,J) : J \subset 3I \setminus I \text{ or } I \subset 3J \setminus J \}.
\]
Then the first cases are as follows:

**Lemma 7.5.** For \( \star \in \{ \text{diagonal, far, near} \} \), the following holds

\[
B^{P_\star}(f, g) \lesssim \mathcal{R} \| f \|_\sigma \| g \|_\tau.
\]

The details of these cases are straightforward, appealing to cancellation from the ‘small’ martingale difference, and the \( A_2 \) condition. Complete details on these cases have appeared in [12, 14, 20].

The next two cases concern the reduction of the ‘large’ martingale difference to the child that contains the ‘small’ martingale difference. There are two different estimates which are as follows.

**Lemma 7.6.** Both of these bilinear forms are bounded by \( \mathcal{R} \| f \|_\sigma \| g \|_\tau \),

\[
\sum_{(I,J) : I \supset J} \left| \left\langle R^\tau_\sigma (\Delta^\tau_{Q,J} g \cdot 1_{Q,J \setminus Q,I}), \Delta^\tau_{Q,I} f \right\rangle \sigma \right|,
\]

\[
\sum_{(I,J) : J \supset I} \left| \left\langle R^\tau_\sigma (\Delta^\tau_{Q,J} f \cdot 1_{I \setminus J}), \Delta^\tau_{Q,J} g \right\rangle \tau \right|.
\]

Recall that \( I_J \) is the child of \( I \) that contains \( J \), and \( Q_{IJ} \) is the dyadic cube, child of \( Q_J \) that contains \( Q_I \). Moreover, in the first estimate, we have the full martingale difference on \( g \).

The proof of this estimate is very similar to that of \( P_{\text{near}} \), and we again omit the details.

**References**

[1] A. B. Aleksandrov, *Isometric embeddings of co-invariant subspaces of the shift operator*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **232** (1996), no. Issled. po Linein. Oper. i Teor. Funktsii. 24, 5–15, 213 (Russian, with English and Russian summaries); English transl., J. Math. Sci. (New York) **92** (1998), no. 1, 3543–3549.

[2] A. B. Aleksandrov, *A simple proof of the Volberg-Treil theorem on the embedding of covariant subspaces of the shift operator*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **217** (1994), no. Issled. po Linein. Oper. i Teor. Funktsii. 22, 26–35, 218 (Russian, with English and Russian summaries); English transl., J. Math. Sci. (New York) **85** (1997), no. 2, 1773–1778.

[3] A. D. Baranov, *Embeddings of model subspaces of the Hardy class: compactness and Schatten–von Neumann ideals*, Izv. Ross. Akad. Nauk Ser. Mat. **73** (2009), no. 6, 3–28 (Russian, with Russian summary); English transl., Izv. Math. **73** (2009), no. 6, 1077–1100.

[4] Lennart Carleson, *Interpolations by bounded analytic functions and the corona problem*, Ann. of Math. (2) **76** (1962), 547–559.

[5] William S. Cohn, *Carleson measures and operators on star-invariant subspaces*, J. Operator Theory **15** (1986), no. 1, 181–202.

[6] W. S. Cohn, *Carleson measures for functions orthogonal to invariant subspaces*, Pacific J. Math. **103** (1982), no. 2, 347–364.

[7] Carl C. Cowen and Barbara D. MacCluer, *Composition operators on spaces of analytic functions*, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL, 1995.

[8] Tuomas P. Hytönen, Michael T. Lacey, and Carlos Pérez, *Sharp weighted bounds for the \( q \)-variation of singular integrals*, Bull. Lond. Math. Soc. **45** (2013), no. 3, 529–540, DOI 10.1112/blms/bds114. MR3065022

[9] Tuomas P. Hytönen, *The two weight inequality for the Hilbert transform with general measures* (2013), available at [http://www.arxiv.org/abs/1312.0843](http://www.arxiv.org/abs/1312.0843).

[10] Pascal Lefèvre, Daniel Li, Hervé Queffélec, and Luis Rodríguez-Piazza, *Nevanlinna counting function and Carleson function of analytic maps*, Math. Ann. **351** (2011), no. 2, 305–326.

[11] Michael T. Lacey, *Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization*, II, Duke MJ, to appear (2013), available at [http://www.arxiv.org/abs/1301.4663](http://www.arxiv.org/abs/1301.4663).
[12] ______, *The Two Weight Inequality for the Hilbert Transform: A Primer*, Submitted (2013), available at http://www.arxiv.org/abs/1304.5004.

[13] Michael Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, *A characterization of two weight norm inequalities for maximal singular integrals with one doubling measure*, Anal. PDE 5 (2012), no. 1, 1–60.

[14] Michael T. Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, *A Two Weight Inequality for the Hilbert transform Assuming an Energy Hypothesis*, J. Funct. Anal. 263 (2012), 305–363.

[15] Michael T. Lacey, Eric T. Sawyer, Ignacio Uriarte-Tuero, and Chun-Yen Shen, *Two Weight Inequality for the Hilbert Transform: A Real Variable Characterization, I*, Duke MJ, to appear, available at http://www.arxiv.org/abs/1201.4319.

[16] Yu. Lyubarskii and E. Malinnikova, *Composition Operators on Model Spaces* (2013), available at http://arxiv.org/abs/1205.5172.

[17] Barbara D. MacCluer, *Compact composition operators on $H^p(B_N)$*, Michigan Math. J. 32 (1985), no. 2, 237–248.

[18] Benjamin Muckenhoupt, *Hardy’s inequality with weights*, Studia Math. 44 (1972), 31–38. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I.

[19] F. Nazarov, *A counterexample to Sarason’s conjecture*, Preprint, MSU (1997), available at http://www.math.msu.edu/~fedja/prepr.html.

[20] F. Nazarov, S. Treil, and A. Volberg, *Two weight estimate for the Hilbert transform and Corona decomposition for non-doubling measures* (2004), available at http://arxiv.org/abs/1003.1596.

[21] F. Nazarov and A. Volberg, *The Bellman function, the two weight Hilbert transform, and embeddings of the model spaces $K_\theta$*, J. Anal. Math. 87 (2002), 385–414. Dedicated to the memory of Thomas H. Wolff.

[22] N. K. Nikol’skii, *Treatise on the shift operator*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 273, Springer-Verlag, Berlin, 1986. Spectral function theory; With an appendix by S. V. Hrushev [S. V. Khrushchëv] and V. V. Peller; Translated from the Russian by Jaak Peetre.

[23] Alexei Poltoratski and Donald Sarason, *Aleksandrov-Clark measures*, Recent advances in operator-related function theory, Contemp. Math., vol. 393, Amer. Math. Soc., Providence, RI, 2006, pp. 1–14.

[24] Eric T. Sawyer, *A characterization of a two weight norm inequality for maximal operators*, Studia Math. 75 (1982), no. 1, 1–11.

[25] ______, *A characterization of two weight norm inequalities for fractional and Poisson integrals*, Trans. Amer. Math. Soc. 308 (1988), no. 2, 533–545.

[26] Eric T. Sawyer, C.-Y. Shen, and I. Uriarte-Tuero, *A Geometric Condition, Necessity of Energy, and Two Weight Boundedness of Fractional Riesz Transforms*, available at http://arxiv.org/abs/1310.4484.

[27] Joel H. Shapiro, *The essential norm of a composition operator*, Ann. of Math. (2) 125 (1987), no. 2, 375–404.

[28] S. R. Treil and A. L. Vol’berg, *Embedding theorems for invariant subspaces of the inverse shift operator*, Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 149 (1986), no. Issled. Linein. Teor. Funktsii. XV, 38–51, 186–187 (Russian, with English summary); English transl., J. Soviet Math. 42 (1988), no. 2, 1562–1572.

[29] A. Volberg, *Calderón-Zygmund capacities and operators on nonhomogeneous spaces*, CBMS Regional Conference Series in Mathematics, vol. 100, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2003.
Michael T. Lacey, School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta GA USA 30332-0160
E-mail address: lacey@math.gatech.edu

Eric T. Sawyer, Department of Mathematics & Statistics, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4K1
E-mail address: sawyer@mcmaster.ca

Chun-Yen Shen, Department of Mathematics, National Central University, Chungli, Taiwan 32054
E-mail address: chunyshen@gmail.com

Ignacio Uriarte-Tuero, Department of Mathematics, Michigan State University, East Lansing MI USA 48824
E-mail address: ignacio@math.msu.edu

Brett D. Wick, School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA USA 30332-0160
E-mail address: wick@math.gatech.edu