

# The bilinear Hilbert transform is pointwise finite

Michael T. Lacey

*To my father, H. E. Lacey, on the occasion of his sixtieth birthday*

**Abstract.** Let  $f \in L^\infty$  and  $g \in L^2$  be supported on  $[0, 1]$ . Then the principal value integral below exists in  $L^1$ .

$$\text{p. v. } \int f(x+y) g(x-y) \frac{dy}{y} .$$

## 1. Introduction.

The bilinear Hilbert transform is

$$Hfg(x) = \int f(x+y) g(x-y) \frac{dy}{y} .$$

Two different rates of translation are incorporated into the integral, making it an extraordinarily subtle object. A beautiful conjecture, formulated first by A. Calderón in 1964,<sup>1</sup> is that  $H$  maps  $L^2 \times L^\infty$  into  $L^2$ .

This paper gives some concrete indication that the conjecture could be true.

---

<sup>1</sup> This date was supplied to me by R. Coifman. Independently, C. Fefferman posed the same conjecture in the early 1970's. The motivation for both was the Cauchy integral on Lipschitz curves.

**Theorem 1.1.** *For functions  $f$  and  $g$  supported on  $[0, 1]$ ,*

$$\left\| \int f(x+y) g(x-y) \frac{dy}{y} \right\|_1 \leq C \|f\|_\infty \|g\|_2 .$$

*In particular, the bilinear Hilbert transform is pointwise finite for  $f \in L^\infty$  and  $g \in L^2$ .*

There is only one natural antecedent to this conjecture of Calderón, namely Carleson's Theorem ([C]) on the pointwise convergence of Fourier series. And the starting point of this paper is C. Fefferman's proof ([F]) of that Theorem.

We then view one of the functions, say  $f$ , as a fixed bounded function. This allows us to use  $L^2$  techniques to bound  $Hfg$  for  $g \in L^2([-1, 1])$ . From this point, the large scale structure of our proof is borrowed from Fefferman. The operator is decomposed into pieces localized in space and frequency. These pieces have an intrinsic size, and they are grouped accordingly.

The proof depends upon combinatorial considerations to break up the small pieces of the decomposition into a relatively small number of orthogonal parts. A critical feature of the decomposition are certain Carleson measure estimates which arise from the decomposition of the function  $f$ .

The whole analysis is done solely on the function  $f$ , and only using the fact that  $f$  is bounded. The paper fully develops the line of reasoning, within these restrictions. Further progress on the conjectures can be made, by gaining a deeper understanding of the decompositions of the function  $f$  and their properties. As well, the structure of the function  $g$  must also be exploited, whereas in this paper  $g$  is essentially ignored. Both of these steps can be taken,<sup>2</sup> but will add significant complications to an already difficult paper. A discussion of this line of investigation will be postponed.

The proof occupies the rest of the paper, which is organized along the following lines. A strong working knowledge of Fefferman's proof [F] would certainly be an aid to the reader. Our approach uses especially, the combinatorial methods therein. But moreover, many of the details must be treated with more care in the present setting of the bilinear Hilbert transform.

---

<sup>2</sup> These words written in May of 1995 turned out to be prophetic. Together with C. Thiele [LT], a range of  $L^p$  bounds have been established for the bilinear Hilbert transform on the real line.

Section 2 gives the decomposition used in the proof, and a very crude outline of the proof. Section 3 treats some technical issues.

Section 4 initiates the real work. The first half of it takes up the most striking new element of the proof: the Carleson measure estimate of Lemma 4.4. It is the work-horse of the whole proof. And its proof employs the full strength of the combinatorial ideas of [F]. Section 6 is exclusively devoted to a second Carleson measure estimate. Its proof is less difficult than the first, but the formulation is far from obvious.

The rest of Section 4, as well as Section 5 are devoted building up large units of bounded operators from the decomposition. The units of Section 5 are trees, which are in fact Calderón-Zygmund operators. Lemma 5.1 explains succinctly a fundamental difficulty. This difficulty is the sole subject of Section 6.

Concerning notation, a capital  $C$  will denote a constant, perhaps one that changes from line to line; in contrast, a lower case  $c$  will denote the center of a relevant interval; the indicator function of a set  $A$  will be denoted by  $\mathbf{1}_A$ , or when  $A$  has a complicated description,  $\mathbf{1}[A]$ ; a similar notation will be used in summation

$$\sum_j [j \in A] f(j) = \sum_{j \in A} f(j).$$

## 2. The decomposition.

This section will describe a decomposition of the bilinear singular integrals. Define the Fourier transform by

$$\mathcal{F}f(\xi) = \int e^{-2\pi i x \xi} f(x) dx,$$

and set  $\langle f, g \rangle = \int f \bar{g} dx$ .

We regard the bounded function  $f$  as fixed and supported on  $[0, 1]$ , with  $g$  varying but supported on  $[0, 1]$ . Notice that then the bilinear Hilbert transform is compactly supported, and we can restrict  $1/y$  to  $|y| < 2$ . The proof is based upon a decomposition of  $Hf \cdot$  into a relatively small number of components orthogonal in either space or frequency.

The place to start the decomposition is the kernel  $1/y$ . Consider a

kernel  $K(y) = \sum_{j \geq 1} K_j(y)$ , where for each  $j$ ,

$$(2.1) \quad \text{supp}(\mathcal{F}K_j) \subset \left\{ \xi : \left( \frac{1}{3} - \frac{1}{100} \right) 3^j \leq \xi \leq \frac{1}{3} 3^j \right\},$$

$$(2.2) \quad |K_j(y) - K_j(y')| \leq C \frac{|y - y'|}{|y|^2}, \quad \text{if } 2|y - y'| \leq |y|,$$

as well as

$$(2.3) \quad |K_j(y)| \leq C 3^j (1 \wedge (3^j y)^{-M}), \quad \text{where } 30 < M < \infty.$$

The choice of  $M$  will be some large but fixed value. The negative values of  $j$  are irrelevant, as we are only concerned with  $|y| < 2$ .<sup>3</sup> We can write

$$\frac{1}{y} = J(y) + \sum_{v=0}^V K_+^v(y) + K_-^v(y), \quad 0 < |y| < 2,$$

where  $\|J(y)\|_1 < \infty$ . We take  $V = \min\{v : (1 + 1/56)^v \geq 3\}$ . For  $0 \leq v < V$ , the kernels

$$K_+^v \left( \left( 1 + \frac{1}{200} \right)^v y \right)$$

satisfies (2.1), (2.2) and (2.3). The kernels

$$\overline{K_-^v \left( \left( 1 + \frac{1}{200} \right)^v y \right)}$$

satisfy the same conditions. We show how to treat

$$Tfg(x) = \int f(x+y)g(x-y)K(y)dy,$$

and the kernel  $J$  is trivial.

---

<sup>3</sup> The Fourier transform of  $K$  will be supported on  $(0, \infty)$ . The proof accomodates  $\mathcal{F}K_j$  being supported on the positive and negative axes, but this slightly complicates some other combinatorial considerations.

**Pairs.** Fix triadic grids  $\mathcal{G}$  of  $\mathbb{R}$  and  $\widehat{\mathcal{G}}$  of  $\widehat{\mathbb{R}}$ .<sup>4</sup> Call  $p = [\omega, I]$  a *pair* if  $\omega \in \widehat{\mathcal{G}}$ , and  $I \in \mathcal{G}$  with  $|\omega| = |I|^{-1}$ . We write  $p = [\omega_p, I_p]$ . Fix a function  $\phi$  with  $L^2$  norm 1,  $\text{supp}(\widehat{\phi}) \subset [-1/100, 1/100]$ ,

$$(2.4) \quad |\phi(x)| \leq C_M (1 \wedge |x|^{-M}),$$

where  $30 < M < \infty$  is large, but fixed, and the collection of functions below is a tight frame in  $L^2(\mathbb{R})$ .

$$(2.5) \quad \{\phi^{m,n}(x) = e^{2\pi im/200} \phi(x - 100n) : n, m \in \mathbb{Z}\}.$$

This last condition means that

$$\sum_{m,n} |\langle f, \phi^{m,n} \rangle|^2 = 2 \|f\|_2^2,$$

for all square integrable  $f$ ; and so

$$(2.6) \quad f(x) = \frac{1}{2} \sum_{m,n} \phi^{m,n}(x) \langle f, \phi^{m,n} \rangle,$$

at least in a  $L^2$  sense. By considerations in [D, Section 3.4], this amounts to choosing  $\phi$  so that

$$\sum_n \left| \widehat{\phi}\left(\xi + \frac{n}{200}\right) \right|^2 \equiv \text{constant}.$$

All of these requirements can be met by choosing  $\widehat{\phi}$  to be symmetric, and for an increasing function  $\nu(\xi)$  on  $[0, 1/200]$  with  $\nu(0) = 0$  and  $\nu(1/200) = \pi/2$ , setting

$$\widehat{\phi}(\xi) = \lambda \begin{cases} \cos \nu(\xi), & 0 \leq \xi \leq \frac{1}{200}, \\ \sin \nu\left(\xi - \frac{1}{200}\right), & \frac{1}{200} \leq \xi \leq \frac{1}{100}, \\ 0, & \text{otherwise.} \end{cases}$$

where  $\lambda$  is a normalizing constant. For any such  $\nu$ , the function  $\phi$  will satisfy (2.7); and for a smooth choice of  $\nu$ ,  $\phi$  satisfies (2.4).

---

<sup>4</sup> That is,  $\mathcal{G}$  is a union of intervals whose lengths are powers of 3. The set  $\{I \in \mathcal{G} : |I|=3^{-j}\}$  partitions  $\mathbb{R}$ , and the partitions are refining as  $|I|$  decreases.

A further property of  $\phi$  is that

$$(2.8) \quad \{\phi(x - 200n) : n \in \mathbb{Z}\} \quad \text{are orthonormal.}$$

This is due to (2.7).

For pairs  $[\omega, I]$ , set

$$\phi_{[\omega, I]}(x) = |I|^{-1/2} e^{2\pi i c(\omega)x} \phi\left(\frac{x - c(I)}{|I|}\right).$$

Here,  $c(J)$  is the center of the interval  $J$ . We will also use the notation

$$(2.9) \quad \Phi(\omega) = c(\omega) + \frac{|\omega|}{27} \left[-\frac{1}{2}, \frac{1}{2}\right],$$

to denote the support of the Fourier transform of  $\phi_{[\omega, I]}$ . Much more commonly, we will use the notation

$$k_\omega(y) = k_j(y), \quad \text{where } |\omega| = 3^j.$$

Let  $f_{[\omega, I]}(x) = \phi_p(x) \langle f, \phi_p \rangle$ , which forms our decomposition of  $f$ .

The smallest unit in the decomposition is

$$(2.10) \quad T_p g(x) = T_{[\omega, I]} g(x) = \int k_\omega(y) f_p(x+y) g(x-y) dy.$$

The proof will bound  $T^0 g = \sum_{\text{all pairs}} T_p g$ , which is certainly not the full singular integral. However, as will be explained in the next section, it is enough to bound  $T^0$ .

Here are some simple properties of the  $T_p$ . First, how big is  $T_p$ ? Certainly,  $|T_p g(x)| \leq \|f_p\|_\infty |k_j| * |g|(x)$ , hence

$$(2.11) \quad \|T_p\|_2 \leq C \|f_p\|_\infty = C \frac{|\langle f, \phi_p \rangle|}{\sqrt{|I_p|}}.$$

We will use the notation

$$\text{size}(p) = \frac{|\langle f, \phi_p \rangle|}{\sqrt{|I_p|}}.$$

The operators  $T_p$  has good space and frequency localization. The frequency localization has been made quite precise, as will be described

by associating to every  $\omega$  two sets  $A(\omega)$  and  $A^*(\omega)$ . They are defined by

$$(2.12) \quad A(\omega) = c(\omega) + \frac{1}{3}|\omega| + \frac{|\omega|}{9} \left[ -\frac{1}{2}, \frac{1}{2} \right],$$

and

$$(2.13) \quad A^*(\omega) = 2c(\omega) + \frac{1}{3}|\omega| + \frac{|\omega|}{9} \left[ -\frac{1}{2}, \frac{1}{2} \right].$$

The point of these definitions is that

$$T_{[\omega, I]} g = T_{[\omega, I]} (\mathcal{F}^{-1} \mathbf{1}_{A(\omega)} \mathcal{F} g),$$

and

$$T_{[\omega, I]}^* g = T_{[\omega, I]}^* (\mathcal{F}^{-1} \mathbf{1}_{A^*(\omega)} \mathcal{F} g).$$

Or equivalently, by taking adjoints,

$$(2.14) \quad T_p g = \mathcal{F}^{-1} \mathbf{1}_{A^*(\omega)} \mathcal{F} T_p g \quad \text{and} \quad T_p^* g = \mathcal{F}^{-1} \mathbf{1}_{A(\omega)} \mathcal{F} T_p^* g.$$

To verify these equalities, write the ‘‘Fourier transform’’ of  $T_p$  as follows

$$\begin{aligned} T_{[\omega, I]} g(x) &= \mathcal{F}_\alpha^{-1} \left( \int e^{-2\pi i \alpha y} k_\omega(y) f_p(x+y) dy \hat{g}(\alpha) \right) (x) \\ &= \mathcal{F}_\alpha^{-1} \left( \mathcal{F}_\beta^{-1} (\widehat{k_\omega}(\alpha - \beta) f_{[\omega, I]}(\beta)) (x) \hat{g}(\alpha) \right) (x). \end{aligned}$$

Then, the second half of (2.14) follows by looking back at the definitions of  $k_\omega$  and  $\phi_{[\omega, I]}$ .

A similar calculation holds for  $T_p^*$ . Recalling (2.10), note that

$$\begin{aligned} T_p^* g(x) &= \int k_\omega(-y) f_p(x-2y) g(x-y) dy \\ &= \mathcal{F}_\alpha^{-1} (\mathcal{F}_\beta^{-1} (\widehat{k_\omega}(\alpha + 2\beta) \hat{f}_p(-\beta)) (x) \hat{g}(\alpha)) (x). \end{aligned}$$

From this, the other half of (2.14) follows.

Another simple property of the  $T_p$ ’s is that, up to modest changes in  $f$  and  $g$ , for any pair  $p = [\omega, I]$ ,  $\omega$  can be assumed to be centered at

the origin. Let us carry out the calculation. For a pair  $p = [\omega, I]$ , and  $c \in \hat{\mathbb{R}}$ , (not necessarily the center of  $\omega$ ) notice that

$$\begin{aligned} f_{[\omega, I]}(x) &= \phi_{[\omega, I]}(x) \langle f, \phi_{[\omega, I]} \rangle \\ &= e^{2\pi i x c} \phi_{[\omega - c, I]}(x) \langle e^{-2\pi i c \cdot} f(\cdot), \phi_{[\omega - c, I]}(\cdot) \rangle \\ &= e^{2\pi i x c} \tilde{f}_{[\omega - c, I]}(x), \end{aligned}$$

where  $\tilde{f}(x) = e^{-2\pi i c x} f(x)$ . From this it follows that

$$\begin{aligned} T_{[\omega, I]}g(x) &= \int k_\omega(y) f_{[\omega, I]}(x + y) g(x - y) dy \\ (2.15) \quad &= \int k_\omega(y) e^{2\pi i(x+y)c} \tilde{f}_{[\omega - c, I]}(x + y) g(x - y) dy \\ &= e^{4\pi i x c} \int k_\omega(y) \tilde{f}_{[\omega - c, I]}(x + y) (e^{-2\pi i c(x-y)} g(x - y)) dy. \end{aligned}$$

It is convenient to shift  $\omega$ 's back to the origin because for such pairs  $p$ ,  $f_p$  has derivative dominated by the scale of  $p$ . That is,  $|d/dx f_p(x)| \leq C/|I|$ .

**Partial Order for Pairs.** The last topic for this section is a natural partial order for pairs. Write  $p' = [\omega', I'] < p = [\omega, I]$  if  $\omega \subset \omega'$  and  $I' \subset I$ . This partial order encodes the orthogonality properties of the  $T_p$ , in the following sense: if  $p$  and  $p'$  are not comparable then  $T_p$  and  $T_{p'}$  are, roughly speaking, orthogonal.

C. Fefferman's approach focuses attention on three separate issues. The first is this: arbitrary collections of pairs must be controlled in terms of their maximal elements under the partial order ' $<$ '. The critical question concerns sets of pairs  $\mathcal{P}$  which are mutually incomparable under  $<$ . How big is  $\sum_{[\omega, I] \in \mathcal{P}} \mathbf{1}_I(x)$ ? The answer is contained in a Carleson measure estimate, the proof of which takes up the first half of Section 4 below.<sup>5</sup> The next issue concerns  $T^\mathcal{P} = \sum_{p \in \mathcal{P}} T_p$ , where  $\mathcal{P}$  is a set of pairs which are incomparable under  $<$ . With the Carleson measure estimate in place, and it's implicit orthogonality, one can check that

$$\|T^\mathcal{P}\|_2 \leq C \sqrt{\sup_{p \in \mathcal{P}} \text{size}(p)}.$$

---

<sup>5</sup> The corresponding estimate is required in [F] but is easily obtained.



The next important group of pairs are those which form a tree under the partial order  $<$ . They turn out to be Calderon-Zygmund operators, but now it can happen that

$$\|T^{\mathcal{P}}\|_2 \simeq 1,$$

regardless of how small the size of individual pairs in  $\mathcal{P}$ . This difficulty does not occur in [F]. Fortunately, it turns out that for trees,  $\mathcal{P}$ , the operator norm of  $T^{\mathcal{P}}$  has an explicit form in terms of the decomposition of  $f$ . We take advantage of this in Section 6 to prove a second Carleson measure estimate. This lemma provides a way to control those trees which have large norm, even though their constituent parts are small.

Employing certain combinatorial tricks, one can then show that for an arbitrary collection of pairs  $\mathcal{P}$ ,

$$\|T^{\mathcal{P}}\|_2 \simeq \sup \{ \|T^{\mathcal{P}'}\|_2 : \mathcal{P}' \subset \mathcal{P}, \mathcal{P}' \text{ a tree} \}.$$

The concluding step in the proof is then to decompose the set of all pairs into collections  $\mathcal{P}_n$  with  $\|T^{\mathcal{P}_n}\|_2 \simeq 2^{-n}$ . This estimate is then summed over  $n$ , completing the proof.

### 3. Technicalities.

This section serves as a catch-all; it includes all the steps that need to be explained, but would have hampered the flow of the previous section. The overall direction is to explain how to pass from  $T^0 = \sum_{\text{all pairs}} T_p$  to the integral in Theorem 1.1. But first,  $T^0$  will be further modified.

**Central Intervals and Admissible Pairs.** The modification of  $T^0$  is needed to gain a certain improvement in triadic intervals.<sup>6</sup> Say that  $I$  is *central* if it is the middle third of a triadic interval. Phrased differently,  $I$  is central if both  $I$  and  $3I$  are triadic. The “convenient property of central triadic intervals” is

$$(3.1) \quad I_1 \subsetneq I_2 \subsetneq \cdots \subsetneq I_m, \text{ all } I_m \text{ central implies } 3^{m-1}I_1 \subset I_m.$$

The proof is immediate.

---

<sup>6</sup> This notion doesn't enter into the proof until the very end.

Let us observe that for every  $\omega \in \hat{\mathcal{G}}$ ,  $A(\omega)$  is central in  $\hat{\mathcal{G}}$ . Indeed  $A(\omega)$  is the middle third of  $\hat{\mathcal{G}}$ . But to achieve the same result for  $A^*(\omega)$  we make a specific choice of grids. We take  $\hat{\mathcal{G}} = \{3^j[n - 1/2, n + 1/2) : n, j \in \mathbb{Z}\}$ . Then  $A^*(\omega)$  is central in  $\hat{\mathcal{G}}$ , as is easily checked. More generally if  $\hat{\mathcal{G}}^a = \{I + a : I \in \hat{\mathcal{G}}\}$  is the grid shifted by  $a$ , then for  $\omega \in \hat{\mathcal{G}}^a$  the interval  $A^*(\omega)$  is central in  $\hat{\mathcal{G}}^{2a}$ . These observations concerning  $A^*(\omega)$  do not hold for an arbitrary triadic grid, and the notion of centrality does not enter the proof until Section 6.

Call  $p = [\omega, I] \in \hat{\mathcal{G}}^b \times \mathcal{G}^a$  *admissible* if  $\omega$  is central in  $\hat{\mathcal{G}}^b$ , and  $I$  satisfies the following:  $c(I) \in a + 200|I|\mathbb{Z}$ , where  $a$  is fixed. This requirement on  $I$ , with (2.8), shows that the functions

$$(3.2) \quad \{\phi_{[\omega, I]} : [\omega, I] \text{ admissible; } \omega \text{ fixed}\} \text{ are orthogonal.}$$

This will be useful in Section 6.

The proof in the next four sections will bound  $T^1 = \sum_{p \text{ admissible}} T_p$ , and in the next section, “pair” will mean “admissible pair”.

**Lemma 3.3.** *The bound  $\|T^1 g\|_1 \leq C \|g\|_2$  implies the same bound for the full singular integral.*

PROOF. The proof averages the bound for  $T^1$  over space and frequency, with the central tool being the resolution of the identity below. For  $a, b \in \mathbb{R}$ , let

$$\Phi^{a,b} f(x) := e^{2\pi i a x} \phi(x - b) \langle f, e^{2\pi i a x} \phi(x - b) \rangle.$$

Then,

$$(3.4) \quad f(x) = C \iint (\Phi^{a,b} f)(x) da db.$$

One checks this by showing that for all  $f, g \in L^2$ ,

$$C^{-2} \langle f, g \rangle = \iiint (\Phi^{a,b} f)(x) \overline{(\Phi^{a,b} g)(x)} da db dx.$$

(In the language of wavelets,  $\Phi^{a,b} f$  is an example of a continuous windowed Fourier transform. See [D]). We apply this formula to  $f$ , which is bounded and supported on  $[0, 1]$ .

The definition of  $T^1$ , pair, central and admissible pair all depend upon the choice of grids. Thus, if  $\mathcal{G}$  is a grid, denote by  $\mathcal{G}^a$  it’s shift

by  $a$ . Notice that the central intervals are just shifted in  $\mathcal{G}^a$ , and that they have density  $1/3$  at all scales. That is, among all the intervals  $I \in \mathcal{G}$  with  $|I| = 2^j$ , every third one is central. Clearly, the admissible space intervals have density  $1/200$  at all scales. Then, let

$$f^{a,b,j}(x) = \sum_{[\omega, I] \in \hat{\mathcal{G}}^a \times \mathcal{G}^b} [p \text{ admissible, } |I| = 3^j] \phi_p(x) \langle f, \phi_p \rangle.$$

From Carleson’s Theorem, one sees that this sum is convergent for almost every  $x$ . It follows from (3.4) that

$$f(x) = \lim_{M \rightarrow \infty} \frac{C}{M^2} \int_0^M \int_0^M f^{a,b,j}(x) da db.$$

And so, if  $T^{a,b}$  is the operator  $T^1$  formed over the grid  $\hat{\mathcal{G}}^b \times \mathcal{G}^a$ , we see that

$$\sum_j \int k_j(y) f(x+y) g(x-y) dy = \lim_{M \rightarrow \infty} \frac{C}{M^2} \int_0^M \int_0^M T^{a,b} g(x) da db.$$

The assumption is that we have appropriate norm bounds on  $T^{a,b}$ , independent of  $a$  and  $b$ . These same bounds then clearly apply to the averages above. In this way, we can pass to the full singular integrals.

#### 4. No two pairs comparable.

The object of study herein are sets of pairs  $\mathcal{P}$  for which no two distinct pairs in  $\mathcal{P}$  are comparable with respect to “ $<$ ” Call such a set of pairs a *thicket*. A convenient way to reformulate the not comparable condition is this. Two pairs  $p = [\omega, I]$  and  $p' = [\omega', I']$  are not comparable under ‘ $<$ ’ if and only if the two rectangles in space-frequency plane  $I \times \omega$  and  $I' \times \omega'$  are disjoint.

Here are the three facts this section is devoted to proving.

**Lemma 4.1.** *Let  $0 < b < 1$ . If  $\mathcal{P}$  is a thicket and  $\text{size}(p) \geq b$  for all  $p \in \mathcal{P}$ , then for any  $\varepsilon > 0$ ,*

$$\left\| \sum_{[\omega, I] \in \mathcal{P}} \mathbf{1}_I \right\|_s \leq C_{s,\varepsilon} b^{-2-\varepsilon}, \quad 1 < s < \infty.$$

In particular the function above is in every  $L^r$  class, for  $r < \infty$ .

**Lemma 4.2.** *If  $\mathcal{P}$  is a thicket and  $\text{size}(p) \leq b$  for all  $p \in \mathcal{P}$  then for all  $1 < r \leq 2$ ,*

$$(4.3) \quad \|T^{\mathcal{P}}g\|_1 \leq C_r b^{\delta(r)} \|g\|_r, \quad \text{where } \delta(r) > 0.$$

Note that this is a  $L^r \rightarrow L^1$  estimate.

Both of these will follow from a study of  $f_p$ , which begins with the following Carleson measure lemma.

**Lemma 4.4.** *Let  $0 < b < 1$ . Let  $\mathcal{P}$  be a thicket with  $\text{size}(p) \geq b$  for all  $p \in \mathcal{P}$ . Then for all intervals  $U$ , and all  $\varepsilon > 0$ ,*

$$(4.5) \quad \sum_{[\omega, I] \in \mathcal{P}} [I \subset U] |\langle g, \phi_{[\omega, I]} \rangle|^2 \leq C_\varepsilon b^{-\varepsilon} |U| \|g\|_\infty^2.$$

This Lemma, crucial to the whole line of reasoning of this paper, has an intricate and combinatorial proof. In fact, it already reflects the large scale structure of Fefferman’s argument. The procedure is to identify nice subsets of thickets which satisfy (4.5), and decompose a thicket into a relatively small number of these nice subsets. In choosing our terminology, the thickets, spindly sets, shrubs and hedges of this section, we have chosen words, which in the American vernacular, pertain to the understory of a forest. This seems appropriate due to the close connection between thickets and the forests of Section 7.

The inequality is also probably in it’s optimal form; at any rate, the purely  $L^2$  estimate

$$\sum_{p \in \mathcal{P}} |\langle g, \phi_p \rangle|^2 \leq C b^{-10,000} \|g\|_2^2$$

is false. Indeed, one can use the Fourier transform in this last inequality, and then it is quickly seen to be false for  $\widehat{\phi}_p$  even if  $\widehat{g} = \mathbf{1}_{[-1,1]}$ . Put another way, our Lemma 4.4 relies critically upon  $\phi$  being compactly supported in the frequency variable.

Further, the proof given here will be applied to decompositions of other functions that arise later in this section. To this end, it is important to note that the only property of  $\phi$  or rather  $\phi_p$  used is this:

$$(4.6) \quad \omega \cap \omega' = \emptyset \quad \text{implies} \quad \text{supp}(\widehat{\phi}_\omega) \cap \text{supp}(\widehat{\phi}_{\omega'}) = \emptyset.$$

The other element of the proof is an estimate of the form

$$\sum_{[\omega, I] \in \mathcal{P}} [I \subset U] |I| \leq C b^{-3} |U|, \quad \text{for all intervals } U.$$

In the end, it is the desire to apply Lemma 4.4 to other “thickets” that justifies, even necessitates, our seemingly absurdly specific requirements on the decomposition given in Section 2. To emphasize this point, let us record here a second Lemma which we shall prove at the same time as the one above.

**Lemma 4.7.** *Let  $\varepsilon > 0$ . Let  $\mathcal{P}$  be a thicket and in addition assume we have the estimate*

$$(4.8) \quad \sum_{[\omega, I] \in \mathcal{P}} [I \subset U] |I| \leq C b^{-4} |U|, \quad \text{for all intervals } U.$$

Let  $\{\varphi_p : p \in \mathcal{P}\}$  be functions satisfying

$$(4.9) \quad \omega_p \cap \omega_{p'} = \emptyset \quad \text{implies} \quad \text{supp}(\widehat{\varphi}_p) \cap \text{supp}(\widehat{\varphi}_{p'}) = \emptyset,$$

$$(4.10) \quad |\varphi_p(x)| \leq C_M \left(1 + \frac{\text{dist}(x, I_p)}{|I_p|}\right)^{-M},$$

where  $M$  depends upon  $\varepsilon$ . Then for all intervals  $U$  and bounded functions  $g$ ,

$$(4.11) \quad \sum_{p \in \mathcal{P}} [I_p \subset U] |\langle g, \varphi_p \rangle|^2 \leq C_\varepsilon b^{-\varepsilon} |U| \|g\|_\infty^2.$$

In the Lemmas above and below,  $0 < \varepsilon < 1$  is fixed. It’s choice depends ultimately on the choice of  $1 < r < 2$  in (4.3). The value of  $\varepsilon$  forces a certain rate of decay on  $\phi$ . Namely in (2.4) and (2.3),  $M$  has to be chosen sufficiently large, but finite. The constants depending upon  $\varepsilon$  are independent of  $f$  and all pairs.

And, in the proof, it is convenient to assume these two conditions on pairs: for all pairs  $p = [\omega, I]$  and  $p' = [\omega', I']$

$$(4.12) \quad \omega = \omega' \text{ and } I \neq I' \quad \text{implies} \quad \text{dist}(I, I') \geq 400 b^{-\varepsilon} |I|.$$

$$(4.13) \quad \omega' \subsetneq \omega \quad \text{implies} \quad |I| \leq b^{1000} |I'|.$$

These conditions can be assumed by breaking up the set of all pairs into  $O(b^{-2\varepsilon})$  disjoint sets. The effect of this will be to multiply our bound in (4.5) by (a trivial amount) of  $b^{-2\varepsilon}$ .

**Spindly sets of pairs.** Introduce a new partial order on pairs. Call  $[\omega, I] \ll [\omega', I']$  if and only if  $\omega \supset \omega'$  and

$$(4.14) \quad |c(I') - c(I)| \leq \frac{1}{3} b^{-\varepsilon} |I'| \sum_{j \geq 0} [|I| < 3^{-j}|I'|] 3^{-j}.$$

One can check that this is indeed a partial order, namely if  $p \ll p'$  and  $p' \ll p''$  then  $p \ll p''$ .<sup>7</sup> (By (4.12), if  $p \ll p'$ , then  $\omega' \subsetneq \omega$ .) Call a set of pairs  $\mathcal{P}$  *spindly* if no two pairs in  $\mathcal{P}$  are comparable under  $\ll$ .

**Lemma 4.15** *If  $\mathcal{P}$  is spindly and  $\text{size}(p) \geq b$  for all  $p \in \mathcal{P}$ , then it satisfies (4.5).*

PROOF. Assume that  $\mathcal{P}$  is finite, and  $I \subset U$  for all  $[\omega, I] \in \mathcal{P}$ . Then let  $B$  denote the best constant in the inequality

$$(4.16) \quad \left\| \sum_{p \in \mathcal{P}} \varepsilon_p \phi_p(x) \langle g, \phi_p \rangle \right\|_2 \leq B \|g\|_\infty,$$

$g$  supported in  $b^{-\varepsilon}U$ ,  $\varepsilon_p \in \{\pm 1\}$ . Our intention is to provide an estimate for  $B$ . Averaging over all choices of signs  $\varepsilon_p$  will give a square function inequality which is weaker than what is claimed in the Lemma, in that a restriction is placed upon the support of  $g$ . We return to this point at the end of the proof.

We will have need of some trivial estimates below. Take  $g$  to be a function bounded by 1. Then  $|\langle g, \phi_{[\omega, I]} \rangle| \leq C \sqrt{|I|} \|g\|_\infty$ . Also

$$(4.17) \quad \begin{aligned} & |\langle \phi_{[\omega, I]}, \phi_{[\omega', I']} \rangle| \\ & \leq \begin{cases} 0, & \text{if } \omega \cap \omega' = \emptyset, \\ C_\varepsilon \sqrt{\frac{|I'|}{|I|}} \left(1 + \frac{\text{dist}(I, I')}{|I|}\right)^{-20/\varepsilon}, & \text{if } \omega \subset \omega'. \end{cases} \end{aligned}$$

---

<sup>7</sup> Roughly speaking,  $p \ll p'$  if  $\omega \supset \omega'$  and  $(3b^\varepsilon)^{-1}I \cap (3b^\varepsilon)^{-1}I' \neq \emptyset$ , and  $\ll$  is the transitive hull of this pairwise relation. The trouble with the simpler condition is that it does not define a partial order. And we have a need yet for certain intricate combinatorial ideas, which depend upon a partial order, as in Lemma 4.28.

Write out the left hand side of (4.16) as

$$\left\| \sum_{p \in \mathcal{P}} \varepsilon_p \phi_p(x) \langle g, \phi_p \rangle \right\|_2^2 = \mathcal{D} + \mathcal{O},$$

where  $\mathcal{D}$  denotes the diagonal term

$$\mathcal{D} = \left\langle \sum_{p \in \mathcal{P}} \phi_p(x) \langle \phi_p, \phi_p \rangle \langle g, \phi_p \rangle, g(x) \right\rangle$$

and the off-diagonal term is, because of (4.9),

$$\mathcal{O} = \sum_{p=[\omega, I] \in \mathcal{P}} \varepsilon_p \langle \phi_p(x), g(x) \rangle \sum_{p'=[\omega', I'] \in \mathcal{P}} [\omega \subsetneq \omega'] \varepsilon_{p'} \langle \phi_{p'}, \phi_p \rangle \langle g, \phi_{p'} \rangle.$$

The assumed inequality can be used on the diagonal term  $\mathcal{D}$ .

$$\mathcal{D} = \|g\|_2 \left\| \sum_{p \in \mathcal{P}} \phi_p(x) \langle g, \phi_p \rangle \right\|_2 \leq B \sqrt{b^{-\varepsilon} |U|} \|g\|_\infty^2.$$

For the off-diagonal term, fix a  $p = [\omega, I] \in \mathcal{P}$ . Then the sets  $\{I' : [\omega', I'] \in \mathcal{P}, \omega' \supset \omega, I' \neq I\}$  are pairwise disjoint, and do not intersect  $b^{-\varepsilon} I$ . Hence, by (4.17),

$$\begin{aligned} \mathcal{S}_p &= \left| \sum_{p' \in \mathcal{P}} [\omega \subsetneq \omega'] \varepsilon_{p'} \langle \phi_{p'}, \phi_p \rangle \langle g, \phi_{p'} \rangle \right| \\ &\leq C \|g\|_\infty \sum_{p' \in \mathcal{P}} [\omega \subsetneq \omega'] \frac{1}{\sqrt{|I|}} \left(1 + \frac{\text{dist}(I, I')}{|I|}\right)^{-20/\varepsilon} |I'| \\ &\leq C \|g\|_\infty \int_{(b^{-\varepsilon} I)^c} \frac{1}{\sqrt{|I|}} \left(1 + \frac{\text{dist}(I, x)}{|I|}\right)^{-20/\varepsilon} dx \\ &\leq C_\varepsilon b^{10} \|g\|_\infty \sqrt{|I|}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathcal{O} &\leq \left\| \sum_{p \in \mathcal{P}} |\phi_p(x)| \mathcal{S}_p \right\|_1 \|g\|_\infty \\ &\leq C_\varepsilon b^{10} \|g\|_\infty^2 \left\| \sum_{p=[\omega, I] \in \mathcal{P}} |\phi_p(x)| \sqrt{|I|} \right\|_1 \\ &\leq C_\varepsilon b^{10} \|g\|_\infty^2 \sum_{[\omega, I] \in \mathcal{P}} |I|. \end{aligned}$$

Putting these estimates together, we see that

$$(4.18) \quad B^2 \leq B\sqrt{b^{-\varepsilon}|U|} + C_\varepsilon b^{10} \sum_{[\omega, I] \in \mathcal{P}} |I|.$$

If  $B\sqrt{b^{-\varepsilon}|U|}$  is the larger of the two terms on the right, then  $B \leq 2\sqrt{b^{-\varepsilon}|U|}$  and we have proved (4.16). So we assume that this is not the case and we derive a contradiction by applying the inequality to  $f$ . Of course  $f$  is not supported on  $b^{-\varepsilon}U$ , nevertheless we have

$$(4.19) \quad \Delta_p^2 = |\langle f, \phi_p \rangle - \langle f \mathbf{1}_{b^{-\varepsilon}U}, \phi_p \rangle|^2 \leq C_\varepsilon b^{10} |I|.$$

Note that this bound only depends upon the  $L^\infty$  bound on  $f$ . Hence

$$\begin{aligned} \frac{1}{4} b^2 \sum_{[\omega, I] \in \mathcal{P}} |I| &\leq \sum_{p \in \mathcal{P}} |\langle f, \phi_p \rangle|^2 \\ &\leq 2 \sum_{[\omega, I] \in \mathcal{P}} |\langle f \mathbf{1}_{b^{-\varepsilon}U}, \phi_p \rangle|^2 + \Delta_p^2 \\ &\leq C_\varepsilon b^{10} \sum_{[\omega, I] \in \mathcal{P}} |I|. \end{aligned}$$

We therefore see a contradiction for small  $b$ , which is enough to prove (4.16), because we only assumed a lower bound on the size of  $p \in \mathcal{P}$ . To extend the square function inequality to all bounded functions, note that inequality (4.19) is valid for all such functions, and so shows that the condition that  $g$  be supported in  $b^{-\varepsilon}U$  is superfluous, thereby establishing the Lemma.

We observe that the proof above contains the following Lemma.

**Lemma 4.20.** *Lemma 4.7 holds under the additional assumption that  $\mathcal{P}$  is spindly.*

PROOF. We repeat the argument above, up the equation (4.18). Then we appeal to (4.8) to conclude that (4.16) holds, with best constant  $B \leq b^{-\varepsilon} \sqrt{|U|}$ . That is the Lemma holds up the restriction on the support of  $g$  in (4.16). But this restriction is removed just as above.

**Shrubs.** Call  $\mathcal{P}$  a *shrub* with top  $p^t = [\omega^t, I^t]$  if  $p \ll p^t$  for every  $p \in \mathcal{P}$ , but no two pairs in  $\mathcal{P}$  are comparable with respect to  $\ll$ . Because the



sets  $\{I : [\omega, I] \in \mathcal{P}\}$  are disjoint and contained in  $b^{-\varepsilon}I^t$ , the inequality (4.5) is trivial. What is more to the point is the following decomposition, which is essentially a corollary to the Fefferman-Stein maximal inequalities. This Lemma depends only on the combinatorics of pairs.

**Lemma 4.21.** *Let  $\mathcal{P}$  be a shrub with top  $p^t = [\omega^t, I^t]$ . Then there is a set  $\mathcal{P}' \subset \mathcal{P}$  and a set  $F \subset b^{-\varepsilon}I^t$  for which the following three conditions hold.*

- i)  $\mathcal{P}'$  can be written as a union of at most  $O(b^{-4\varepsilon})$  spindly sets.
- ii) For all  $[\omega, I] \in \mathcal{P} \setminus \mathcal{P}'$ ,  $b^{-\varepsilon}I \subset F$ .
- iii) And  $|F| \leq C_\varepsilon b^{100} |I^t|$ .

PROOF. Notice that the sets  $\{I : [\omega, I] \in \mathcal{P}\}$  are disjoint, for otherwise two pairs would be comparable under  $<$ . Set

$$F_0 = \left\{ x : \sum_{[\omega, I] \in \mathcal{P}} (M\mathbf{1}_I)^\alpha(x) > b^{-(3-\alpha)\varepsilon} \right\}, \quad \alpha = 1 + \frac{\varepsilon}{2}.$$

By the Fefferman-Stein maximal inequalities

$$\left\| \left( \sum_{[\omega, I] \in \mathcal{P}} (M\mathbf{1}_I)^\alpha(x) \right)^{1/\alpha} \right\|_\beta \leq C_{\alpha, \beta} (b^{-\varepsilon} |I^t|)^{1/\beta}, \quad 1 < \beta < \infty.$$

Using this with  $\beta$  large implies that  $|F_0| \leq C_\varepsilon b^{100} |I^t|$ .

Now, set  $\mathcal{P}'_0 = \{[\omega, I] \in \mathcal{P} : b^{-\varepsilon}I \not\subset F_0\}$ . Our claim is that

$$\left\| \sum_{[\omega, I] \in \mathcal{P}'_0} \mathbf{1}_{[4b^{-\varepsilon}I]}(x) \right\|_\infty \leq 64 b^{-3\varepsilon}.$$

Consider an  $x$  and intervals  $I_1, \dots, I_J$  with  $x \in 4b^{-\varepsilon}I_j$  for all  $1 \leq j \leq J$ , and  $[\omega_j, I_j] \in \mathcal{P}'_0$ . Suppose that  $|I_1| \leq |I_j|$  for all  $j$ . Then, for all  $y \in I_1$ ,

$$(M\mathbf{1}_{I_j}(y))^\alpha \geq 8^{-\alpha} b^{\varepsilon\alpha}, \quad 1 \leq j \leq J,$$

so that

$$\sum_{j=1}^J (M\mathbf{1}_{I_j}(y))^\alpha \geq J 8^{-\alpha} b^{\varepsilon\alpha}, \quad y \in I_1.$$

The right hand side above cannot be more than  $64 b^{-3\varepsilon}$ , or we see that  $[\omega_1, I_1] \notin \mathcal{P}'_0$ . This gives an upper bound on  $J$ .

The set  $\mathcal{P}'_0$  will have to have some more pairs deleted before it can be decomposed into spindly sets. To accomplish this, let

$$U_1 = \bigcup_{[\omega, I] \in \mathcal{P}'_0} b^{-\varepsilon} I.$$

Choose  $\bar{p}_j = [\bar{\omega}_j, \bar{I}_j] \in \mathcal{P}'_0$  so that that  $\{b^{-\varepsilon} \bar{I}_j : j \geq 1\}$  forms a minimal cover of the set  $U_1$ . Set  $\mathcal{P}'_1 = \{\bar{p}_j : j \geq 1\}$ . Delete these pairs from  $\mathcal{P}'_0$  and repeat this procedure. As a result, we have

$$\mathcal{P}'_0 = \mathcal{P}'_1 \cup \dots \cup \mathcal{P}'_J, \quad J \leq 64 b^{-3\varepsilon},$$

and for each  $\mathcal{P}'_j$ , the sets  $\{b^{-\varepsilon} I : [\omega, I] \in \mathcal{P}'_j\}$  form a minimal cover of their union.

Last of all, we claim that for each  $\mathcal{P}'_j$ ,  $1 \leq j \leq J$ , there is a set  $F_j \subset b^{-\varepsilon} I^t$  and a  $\mathcal{P}''_j \subset \mathcal{P}'_j$  so that

- i)  $\mathcal{P}''_j$  is a union of  $O(\log 1/b)$  spindly collections of pairs.
- ii) For all  $[\omega, I] \in \mathcal{P}'_j \setminus \mathcal{P}''_j$ ,  $b^{-\varepsilon} I \subset F_j$ .
- iii)  $|F_j| \leq C b^{200} |I^t|$ .

These last three conditions in fact follow from the Vitali Covering Lemma: from  $\{b^{-\varepsilon} I : [\omega, I] \in \mathcal{P}'_j\}$  select pairs  $[\bar{\omega}_v, \bar{I}_v] \in \mathcal{P}'_j$  so that the intervals  $b^{-\varepsilon} \bar{I}_v$  are disjoint in  $v$  and

$$\sum_v b^{-\varepsilon} |\bar{I}_v| \geq \frac{1}{2} \left| \bigcup_{[\omega, I] \in \mathcal{P}'_j} b^{-\varepsilon} I \right|.$$

The collection  $\{[\bar{\omega}_v, \bar{I}_v] : v \geq 1\}$  is clearly spindly. Repeating this procedure  $O(\log 1/b)$  times will prove these three conditions.

Last of all, we conclude the Lemma by taking  $\mathcal{P}' = \bigcup_{j=1}^J \mathcal{P}''_j$ , and  $F = \bigcup_{j=0}^J F_j$ .

In the first half of the proof just given we have made an observation which will be used below. Let us formulate it as a Lemma.

**Lemma 4.22.** *Let  $\mathcal{I}$  be a collection of intervals. For a choice of  $1 \leq \alpha < \infty$  set*

$$E = \left\{ x : \left( \sum_{I \in \mathcal{I}} (M \mathbf{1}_I(x))^\alpha \right)^{1/\alpha} > J \right\}.$$

Then for all  $x$

$$\sum_{I \in \mathcal{I}} [b^{-\varepsilon} I \not\subset E] \mathbf{1}_{b^{-\varepsilon} I}(x) \leq 2^\alpha (b^{-\varepsilon \alpha} J)^\alpha.$$

A second Lemma of a general nature is also relevant at several parts of our argument. It shows that it suffices to prove the Carleson measure estimate, up to exceptional sets.

**Lemma 4.23.** *Let  $\{a_p : p \in \mathcal{P}\}$  be non-negative numbers associated to pairs. Assume that for all intervals  $U$  there is an open set  $E \subset U$  so that  $|E| \leq |U|/2$  and*

$$\sum_{p \in \mathcal{P}} [I_p \not\subset U] |a_p| \leq C |U|.$$

Then, for all intervals  $U$ ,

$$(4.24) \quad \sum_{p \in \mathcal{P}} |a_p| \leq 2C |U|.$$

PROOF. Fix the interval  $U$  for which we want to prove (4.24). Set  $E_1$  to be the set  $E$  of the Lemma, and let  $\mathcal{P}'_1 = \{p \in \mathcal{P} : I_p \not\subset U\}$ . We apply the hypotheses of the Lemma to the components of the set  $E_1$ . Thus, let  $U_{1,k}$  be the open components of  $E_1$ , and let  $\mathcal{P}_{1,k} = \{p \in \mathcal{P} \setminus \mathcal{P}'_1 : I_p \subset U_{1,k}\}$ . Then for each  $k$  there is an open set  $E_{1,k} \subset U_{1,k}$ , with  $|E_{1,k}| \leq |U_{1,k}|/2$ , so that setting  $\mathcal{P}'_{1,k} = \{p \in \mathcal{P} : U_{1,k} \supset I_p, I_p \not\subset E_{1,k}\}$ , we have

$$\sum_{p \in \mathcal{P}'_{1,k}} |a_p| \leq C |U_{1,k}|.$$

But,  $\sum_k |U_{1,k}| \leq |E_1| \leq 1/2 |U|$ . And so,

$$\sum_k |E_{1,k}| \leq \frac{1}{2} \sum_k |U_{1,k}| \leq \frac{1}{2} |E_1| \leq \frac{1}{4} |U|.$$

Set  $\mathcal{P}'_2 = \bigcup_k \mathcal{P}'_{1,k}$ . We conclude that

$$\sum_{p \in \mathcal{P}'_1 \cup \mathcal{P}'_2} |a_p| \leq C \left(1 + \frac{1}{2}\right) |U|.$$

This argument can be continued inductively inside the components of  $E_2 = \bigcup_k E_{1,k}$ , thereby proving the Lemma.

**Hedges.** Call a set of pairs  $\mathcal{P}$  a *hedge* if no two pairs in  $\mathcal{P}$  are comparable under  $<$  and  $\mathcal{P}$  satisfies the following linearity or tree-like condition: for all  $p, p', p'' \in \mathcal{P}$  with  $p \ll p'$  and  $p \ll p''$ , either  $p' \ll p''$  or  $p'' \ll p'$ .

**Lemma 4.25.** *If  $\mathcal{P}$  is a hedge and  $\text{size}(p) \geq b$  for all  $p \in \mathcal{P}$ , then (4.5) holds.*

PROOF. Let  $\bar{p}_k$  be the pairs in  $\mathcal{P}$  maximal with respect to  $\ll$ . Set  $\mathcal{P}_k = \{p \in \mathcal{P} : p \ll \bar{p}_k\}$ . These sets are shrubs and are pairwise disjoint. Indeed, more is true: if  $p \in \mathcal{P}_k, p' \in \mathcal{P}_{k'}$  and  $p \ll p'$  then  $\mathcal{P}_k = \mathcal{P}_{k'}$ . The situation is this.  $p \ll p'$  and  $p \ll \bar{p}_k$ . Hence from the definition of a hedge,  $p' \ll \bar{p}_k$ . But also  $p' \ll \bar{p}_{k'}$ , so that  $\bar{p}_k$  and  $\bar{p}_{k'}$  are comparable under  $\ll$ . Maximality then forces  $\bar{p}_k = \bar{p}_{k'}$ .

A corollary of this is that if  $\mathcal{P}'_k \subset \mathcal{P}_k$  is spindly for all  $k$  then so is  $\bigcup_k \mathcal{P}'_k$ . But we know how to construct spindly sets from shrubs. And we can give a proof of (4.5) for a fixed interval  $U$ .

Assume that  $\bar{I}_k \subset U$  for each  $k$ . Apply Lemma 4.21 to each  $\mathcal{P}_k$ . This gives us sets  $\mathcal{P}'_k \subset \mathcal{P}_k$  and  $F_k \subset b^{-\varepsilon} \bar{I}_k$  satisfying i)-iii) of that Lemma. It follows from Lemma 4.15 that  $\bigcup \mathcal{P}'_k$  satisfies (4.5). Also from the fact that  $\bar{p}_k$  is spindly and  $\text{size}(p) \geq b$ , we see that (4.8) holds for  $\{\bar{p}_k\}$ . Hence

$$(4.26) \quad \sum_k |F_k| \leq C b^{100} \sum_k |\bar{I}_k| \leq C b^{90} |U|.$$

That is for  $b$  sufficiently small,  $C b^{90}$  will be no more than  $1/2$ , and then the assumptions of Lemma 4.23 are seen to hold, with  $a_p = |\langle g, \phi_p \rangle|^2$ . Therefore (4.5) holds. Again, it suffices to prove the Lemma for small  $b$  as only a lower bound on the size of pairs is assumed.

The next Lemma is a trivial adaptation of the previous proof.

**Lemma 4.27.** *Lemma 4.7 hold under the additional assumption that  $\mathcal{P}$  is a hedge.*

Last of all, we want to decompose thickets into a small number of hedges. This Lemma depends upon the combinatorics of the pairs, as well as the Carleson measure estimate.

**Lemma 4.28.** *Let  $\mathcal{P}$  be a thicket and assume that for some interval  $U$ ,  $I \subset U$  for all  $[\omega, I] \in \mathcal{P}$ . Assume that (4.8) holds for every hedge  $\mathcal{P}' \subset \mathcal{P}$ . Then there is a set  $E \subset U$  and a set  $\mathcal{P}' \subset \mathcal{P}$  so that*

- i)  $\mathcal{P}'$  is a union of at most  $O(\log 1/b)$  hedges.
- ii) For all  $[\omega, I] \in \mathcal{P} \setminus \mathcal{P}'$ ,  $I \subset E$ .
- iii) And  $|E| \leq C_\varepsilon b^{100} |U|$ .

This lemma, plus Lemma 4.23 and Lemma 4.25 will prove Lemma 4.4 and Lemma 4.7.

PROOF. Begin by letting  $\bar{p}_k^1 = [\bar{\omega}_k^1, \bar{I}_k^1]$  be the maximal pairs in  $\mathcal{P}$  with respect to  $\ll$ . The set  $\{\bar{p}_k^1\}$  is spindly, and so in particular is a hedge. It satisfies (4.8) by assumption. That is,

$$\sum_k |\bar{I}_k^1| \leq C_\varepsilon b^{-4} |U|.$$

This allows us to delete some pairs from  $\mathcal{P}$ . Set

$$F_0 = \left\{ x : \sum_k (M\mathbf{1}[\bar{I}_k^1](x))^2 > b^{-110} \right\}.$$

It follows that  $|F_0| \leq C_\varepsilon b^{100} |U|$ ; further let  $\bar{p}_j = [\bar{\omega}_j, \bar{I}_j]$  be an enumeration of those  $\bar{p}_k^1$  such that  $b^{-\varepsilon} \bar{I}_k^1 \not\subset F_0$ . As we have already seen in Lemma 4.22, we then have that

$$\sum_j \mathbf{1}[3b^{-\varepsilon} \bar{I}_j](x) \leq 6b^{-300}, \quad \text{for all } x \in b^{-\varepsilon} U.$$

Take  $\mathcal{P}^1 = \{p \in \mathcal{P} : p \ll \text{some } \bar{p}_j\}$ . Note that if  $[\omega, I] \in \mathcal{P} \setminus \mathcal{P}^1$ , then  $[\omega, I] \ll \text{some } \bar{p}_k^1$  with  $b^{-\varepsilon} \bar{I}_k^1 \subset F_0$ , and so  $I \subset F_0$  as well.

A few more pairs must be deleted from  $\mathcal{P}^1$  in order to gain a certain combinatorial advantage. We will proceed in an inductive fashion. Choose a pair  $p_1 = [\omega_1, I_1] \in \mathcal{P}^1$  for which  $|I_1|$  is maximal. Let

$$F_1 = \bigcup \left\{ I : |I| \leq b^{1000} |I_1|, \sum_{0 \leq j \leq 500 \log 1/b} 3^{-j} \leq \frac{|c(I) - c(I_1)|}{b^{-\varepsilon} |I_1|/3} \leq \sum_{j \geq 0} 3^{-j} \right\}.$$

Note that  $|F_1| \leq b^{400} |I_1|$ .<sup>8</sup> Choose a pair  $p_2 = [\omega_2, I_2] \in \mathcal{P}^1 \setminus \{p_1\}$ , with  $I_2 \not\subset F_1$  and  $|I_2|$  maximal. Then define  $F_2$  as above. It follows that  $I_1 \not\subset F_2$ . Continue this procedure until  $\mathcal{P}^1$  is exhausted. Then the set  $E$  of ii) and iii) above is  $E = F_0 \cup \bigcup_{j \geq 1} F_j$ . Let  $\mathcal{P}' = \{p_j : j \geq 1\}$ . Observe that

$$(4.29) \quad |E| \leq C_\epsilon b^{100} |U| + b^{400} \sum_{[\omega, I] \in \mathcal{P}'} |I|.$$

Of course ii) holds. It remains to verify i) and iii). And here observe that the last inequality and i), together with (4.5) for hedges, trivially give iii). So it remains to check i).

The advantage gained in passing from  $\mathcal{P}^1$  to  $\mathcal{P}'$  is this: if  $p = [\omega, I], p' = [\omega', I'] \in \mathcal{P}'$  and  $p \ll p'$  then as follows from the removal of the sets  $F_j, j \geq 1$ ,

$$|c(I) - c(I')| \leq \frac{1}{3} b^{-\epsilon} |I'| \sum_{j \leq 500 \log 1/b} 3^{-j}.$$

As a corollary, we see that  $\mathcal{P}'$  satisfies the following good combinatorial condition: if  $p_1 \ll p_2, p_3 \ll p_4$ , all  $p_i \in \mathcal{P}'$  then either  $p_2 \ll p_3$  or  $p_3 \ll p_2$ .<sup>9</sup> To see this, let  $p_i = [\omega_i, I_i]$ . We then have  $\omega_1 \supset \omega_2, \omega_3 \supset \omega_4$ . Thus, we must have *e.g.*  $\omega_2 \supset \omega_3$ . Under the assumption of (4.12),  $p_1 \ll p_2, p_3$ , and  $p_2 \neq p_3$  implies that we must have  $\omega_3 \subsetneq \omega_2$ . Then

$$\begin{aligned} |c(I_2) - c(I_3)| &\leq |c(I_3) - c(I_1)| + |c(I_1) - c(I_2)| \\ &\leq \frac{1}{3} b^{-\epsilon} |I_3| \left( \sum_{0 \leq j \leq 500 \log 1/b} 3^{-j} + \frac{|I_2|}{|I_3|} \sum_{0 \leq j \leq 500 \log 1/b} 3^{-j} \right) \\ &\leq \frac{1}{3} b^{-\epsilon} |I_3| \sum_{0 \leq j} [3^{-j} |I_3| > |I_2|] 3^{-j}. \end{aligned}$$

The last line follows from the fact that  $|I_2|/|I_3| \leq b^{1000}$ , (see (4.13)) and shows that  $p_2 \ll p_3$ .

We can now apply a combinatorial trick of Fefferman. Let  $\mathcal{B}(p) = \#\{\bar{p}_k : p \ll \bar{p}_k\}$ . For all  $p \in \mathcal{P}'$ , we have  $\mathcal{B}(p) \leq 6b^{-300}$ , for if

$$[\omega, I] \ll \bar{p}_{k(1)}, \dots, \bar{p}_{k(v)}, \quad v = \mathcal{B}([\omega, I]),$$

---

<sup>8</sup> Here we are simply deleting a small neighborhood of the boundary of  $(3b^\epsilon)^{-1} I_1$ . Recall that by (4.12), if  $[\omega, I] \ll [\omega_1, I_1]$ , then  $|I| < b^{1000} |I_1|$ .

<sup>9</sup> Note that for ' $<$ ' this is trivial.

then for  $x \in I$ ,  $\sum_k \mathbf{1}[3b^{-\varepsilon}\bar{I}_k](x) \geq v$ , but by construction this last sum can't be more than  $6b^{-300}$ . Furthermore,  $\mathcal{B}(p)$  has the following combinatorial property: if  $p \ll p'$ , and  $p \ll p''$ , but  $p'$  and  $p''$  not comparable under  $\ll$ , with all pairs in  $\mathcal{P}'$ , then  $\mathcal{B}(p) \geq \mathcal{B}(p') + \mathcal{B}(p'')$ .

Indeed, write  $p' \ll \bar{p}_{k(1)}, \dots, \bar{p}_{k(v)}$ , and  $p'' \ll \bar{p}_{j(1)}, \dots, \bar{p}_{j(w)}$  where  $v = \mathcal{B}(p')$  and  $w = \mathcal{B}(p'')$ . If some  $\bar{p}_{k(\sigma)}$  equals some  $\bar{p}_{j(\tau)}$  then one would have  $p' \ll p''' = \bar{p}_{k(\sigma)} = \bar{p}_{j(\tau)}$  and  $p'' \ll p'''$ . The situation is  $p \ll p', p'' \ll p'''$  which by the good combinatorial property of  $\mathcal{P}'$  forces  $p'$  and  $p''$  to be comparable under  $\ll$ . This is a contradiction which forces the inequality  $\mathcal{B}(p) \geq \mathcal{B}(p') + \mathcal{B}(p'')$ .

But then the hedges are easy to define, simply take  $\mathcal{H}_v = \{p \in \mathcal{P}' : 2^v \leq \mathcal{B}(p) < 2^{v+1}\}$  for  $v \leq 120 \log 1/b$ . The combinatorial property of  $\mathcal{B}(p)$  show that each  $\mathcal{H}_v$  is a hedge, finishing the proof.

We have completed the proof of the critical Carleson measure Lemma.

The next Lemma initiates the proof of the second estimate (Lemma 4.2) on the operator  $T^{\mathcal{P}}$ , but it's proof will also yield the first estimate (Lemma 4.1). We need an improvement of Lemma 4.4.

**Lemma 4.30.** *Let  $\mathcal{P}$  be a thicket with  $I_p \subset b^{-\varepsilon/2}[-1, 1]$  for all  $p \in \mathcal{P}$ . One then has the inequality below for any  $\varepsilon > 0$ .*

$$(4.31) \quad \left\| \left( \sum_{p \in \mathcal{P}} |f_p(x)|^2 \right)^{1/2} \right\|_r \leq C_{\varepsilon,r} b^{-\varepsilon} \|f\|_{\infty}, \quad 2 \leq r < \infty.$$

PROOF. To prove the Lemma, it suffices to establish that for all  $\varepsilon > 0$ ,

$$(4.32) \quad \left\| \left( \sum_{p=[\omega, I] \in \mathcal{P}} \left| \frac{\langle g, \phi_p \rangle}{\sqrt{|I|}} \mathbf{1}_I(x) \right|^2 \right)^{1/2} \right\|_{\text{BMO}} \leq C_{\varepsilon} b^{-\varepsilon} \|g\|_{\infty}.$$

For then properties of BMO give

$$\left\| \left( \sum_{p=[\omega, I] \in \mathcal{P}} \left| \frac{\langle g, \phi_p \rangle}{\sqrt{|I|}} \mathbf{1}_I(x) \right|^2 \right)^{1/2} \right\|_r \leq C_{\varepsilon} r b^{-\varepsilon} \|g\|_{\infty},$$

which with the Fefferman-Stein maximal inequalities gives the Lemma.

But also, if  $\mathcal{P}$  is a thicket and each pair  $p$  has size at least  $b/2$ , and so  $b \mathbf{1}_I(x) \leq C |f_p(x)|$ , it follows from the decay of  $\varphi_p$  that  $b^{-\varepsilon} I_p \subset [-1, 1]$  for all  $p \in \mathcal{P}$ . Hence

$$\left\| \sum_{[\omega, I] \in \mathcal{P}} \mathbf{1}_I(x) \right\|_r \leq C_\varepsilon r^2 b^{-2-\varepsilon},$$

which is the conclusion of Lemma 4.1.

To check (4.32) it is enough to show that for all triadic intervals  $U$ ,

$$\int_U \sum_{[\omega, I] \in \mathcal{P}} [I \not\subset U] \left| \langle g, \phi_{[\omega, I]} \rangle \frac{1}{\sqrt{|I|}} \mathbf{1}_I(x) \right|^2 dx \leq C_\varepsilon b^{-\varepsilon} |U| \|g\|_\infty^2,$$

where  $g$  is a function bounded by 1. But this is precisely (4.5) above.

To bound  $T^\mathcal{P}$  as in (4.3) we will dualize and provide a proof of the estimate

$$(4.33) \quad \|T^{\mathcal{P}*} g\|_s \leq C_s b^{\delta(s)} \|g\|_\infty,$$

where  $2 \leq s < \infty$  and  $\delta(s) > 0$ . We localize  $T^{\mathcal{P}*}$  in the space variable.

**Lemma 4.34.** *Define, for  $\varepsilon > 0$ ,*

$$T_{\#[\omega, I]}^* g(x) = \mathbf{1}[b^{-\varepsilon} I](x) T_{[\omega, I]}^* (\mathbf{1}[b^{-\varepsilon} I] g)(x).$$

*Let  $\mathcal{P}$  be a thicket with size  $(p) \geq b$  for all  $p \in \mathcal{P}$ . And let  $g$  be a function bounded by 1. Then*

$$\left\| \sum_{p \in \mathcal{P}} |T_p^* g(x) - T_{\#p}^* g(x)| \right\|_r \leq C_{r, \varepsilon} b^3, \quad 1 < r < \infty.$$

PROOF. We will do half of the proof, the other half being similar. Estimate ( $p = [\omega, I]$ )

$$\begin{aligned} & |T_p^* g(x) - \mathbf{1}[b^{-\varepsilon} I](x) T_p^* g(x)| \\ & \leq \mathbf{1}[(b^{-\varepsilon} I)^c](x) \int |k_\omega(-y) f_p(x - 2y) g(x - y)| dy \\ & = \mathbf{1}[(b^{-\varepsilon} I)^c](x) \left( \int_{|y| \leq \text{dist}(x, I)/3} + \int_{|y| \geq \text{dist}(x, I)/3} \dots dy \right) \\ & = A_p(x) + B_p(x). \end{aligned}$$



For the first term, use the decay of  $f_p$  away from  $I$ : from the definition,  $|f_p(z)| \leq C_\varepsilon (\text{dist}(z, I)/|I|)^{-6/\varepsilon}$ , which implies that

$$\begin{aligned} A_p(x) &\leq C \mathbf{1}[(b^{-\varepsilon} I)^c](x) \left( \frac{\text{dist}(x, I)}{|I|} \right)^{-6/\varepsilon} |k_\omega| * |g|(x) \\ &\leq C b^6 (M\mathbf{1}_I)^2(x), \end{aligned}$$

where  $M$  denotes the Hardy-Littlewood maximal function.

For the second term, use the decay of  $k_\omega$ :

$$\begin{aligned} |k_\omega(y)| &\leq C_\varepsilon |\omega| (|\omega| |y|)^{-6/\varepsilon}, \\ B_p(x) &\leq C_\varepsilon \mathbf{1}[(b^{-\varepsilon} I)^c](x) \left( \frac{\text{dist}(x, I)}{|I|} \right)^{-6/\varepsilon} \leq C_\varepsilon b^6 (M\mathbf{1}_I)^2(x). \end{aligned}$$

Thus, using Lemma 4.1 and the Fefferman-Stein maximal inequalities we see that

$$\begin{aligned} \left\| \sum_{p \in \mathcal{P}} A_p(x) + B_p(x) \right\|_r &\leq C_\varepsilon b^6 \left\| \sum_{[\omega, I] \in \mathcal{P}} |M\mathbf{1}_I(x)|^2 \right\|_r \\ &\leq C_\varepsilon b^6 \left\| \left( \sum_{[\omega, I] \in \mathcal{P}} |M\mathbf{1}_I(x)|^2 \right)^{1/2} \right\|_{2r}^2 \\ &\leq C_{r, \varepsilon} b^6 \left\| \sum_{[\omega, I] \in \mathcal{P}} \mathbf{1}_I(x) \right\|_r \\ &\leq C_{r, \varepsilon} b^3. \end{aligned}$$

This finishes the proof.

At this point, the top item on the agenda is a decomposition of  $g$  into functions analogous to  $f_p$ . But we have to abandon the luxury of reconstructing  $g$  after the fact, as is done for  $f$  in Lemma 3.3. This creates an extra problem, of an essentially technical nature.

Recall the definition of  $A^*(\omega)$ , (2.13). This interval is triadic in (a shift of)  $\hat{\mathcal{G}}$ . More can be said:  $A^*(\omega)$  is central so that  $3A^*(\omega)$  is triadic;  $c(A^*(\omega)) = 2c(\omega) + |\omega|/6$ ;  $|A^*(\omega)| = |\omega|/9$ ; and

$$(4.35) \quad \omega \cap \omega' = \emptyset \quad \text{implies} \quad 3A^*(\omega) \cap 3A^*(\omega') = \emptyset.$$

Indeed, if  $\omega \cap \omega' = \emptyset$ , then

$$|c(A^*(\omega)) - c(A^*(\omega'))| = 2|c(\omega) - c(\omega')| \geq |\omega| + |\omega'|,$$

which demonstrates the assertion. These observations inform the definitions below.

To a pair  $[\omega, I]$  associate the functions

$$(4.36) \quad \begin{aligned} & \chi_{\delta, [\omega, I]}(x) \\ &= |I|^{-1/2} \exp\left(2\pi i x \left(c(A^*(\omega)) + \delta \frac{|\omega|}{56}\right)\right) \phi\left(\frac{x - c(I)}{|I|}\right) \end{aligned}$$

where  $\delta = -56, -55, \dots, 55$ . For each pair, the collection of functions

$$\{\chi_{\delta, [\omega, I+28n|I|]} : -56 \leq \delta < 56, n \in \mathbb{Z}, \omega \in \hat{\mathcal{G}}\}$$

is just a rescaling of the collection in (2.5) and so satisfies (2.6). (The constant  $A$  in that equation is irrelevant, and so we will take it to be 1). For a pair  $p$ , set

$$(4.37) \quad Q_{[\omega, I]} = \sum_{|\delta| \leq 5} \sum_{|n| \leq (300b^\epsilon)^{-1}} \chi_{\delta, [\omega, I+28n|I|]} \otimes \chi_{\delta, [\omega, I+28n|I|]} .$$

The sum over  $\delta$  is restricted because if  $|\delta| \geq 6$ , then the support of  $(\chi_{\delta, [\omega, J]})^\wedge$  will not intersect  $A^*(\omega)$ . Therefore, one should have  $T_p^*g \simeq T_p^*Q_p g$ . This we will quantify in the next Lemma.

**Lemma 4.38.** *If  $\mathcal{P}$  is a thicket then*

$$\left\| \sum_{\mathcal{P}} |T_p^*g(x) - T_p^*Q_p g(x)| \right\|_r \leq C_r b^6 \|g\|_\infty, \quad r < \infty .$$

PROOF. Since a dilate of (2.6) is in force, it follows from (2.14) that

$$T_p^*g = T_p^* \left( \sum_{|\delta| \leq 5} \sum_{n=-\infty}^{\infty} \chi_{\delta, [\omega, I+28n|I|]} \langle g, \chi_{\delta, [\omega, I+28n|I|]} \rangle \right) .$$

The sum above is absolutely convergent. So we will give a pointwise estimate for

$$\langle g, \chi_{\delta, [\omega, J]} \rangle T_p^* \chi_{\delta, [\omega, J]}(x), \quad \text{where } \text{dist}(I, J) \geq b^{-\epsilon} |I| .$$

The  $\delta$  is unimportant, and so will be dropped from notation.

Set  $\Phi_I(x) = (1 \wedge (|x|/|I|)^{100/\varepsilon})$ . We need to consider the integral

$$\begin{aligned} & \left| \int k_\omega(-y) \phi_p(x - 2y) \chi_{[\omega, J]}(x - y) dy \right| \\ & \leq \frac{C_\varepsilon}{|I|} \int \Phi_I(y) \Phi_I(\text{dist}(x - 2y, I)) \Phi_I(\text{dist}(x - y, J)) \frac{dy}{|I|} \\ & := E(J, x). \end{aligned}$$

Sum this over  $J$  to get

$$\begin{aligned} & \sum_J [ |I| = |J|, \text{dist}(J, I) \geq b^{-\varepsilon}|I| ] E(J, x) \\ & \leq \frac{C_\varepsilon}{|I|} \int \Phi_I(y) \Phi_I(\text{dist}(x - 2y, I)) \Phi_I(\text{dist}(x - y, (b^{-\varepsilon}I)^c)) \frac{dy}{|I|} \\ & \leq \frac{C_\varepsilon}{|I|} b^{10} (M\mathbf{1}_I)^2(x). \end{aligned}$$

Remembering that both  $f$  and  $g$  are bounded by 1, so that

$$|\langle \bar{f}, \phi_p \rangle \langle \bar{g}, \chi_{\delta, [\omega, J]} \rangle| \leq C |I|,$$

it follows that

$$\left\| \sum_{\mathcal{P}} |T_p^*(I - Q_p)g(x)| \right\|_r \leq C_\varepsilon b^{10} \left\| \left( \sum_{[\omega, I] \in \mathcal{P}} (M\mathbf{1}_I(x))^2 \right)^{1/2} \right\|_{2r}^2 \leq C b^6,$$

by Lemma 4.1.

With this last Lemma, we have associated with  $\mathcal{P}$  a collection of pairs

$$\mathcal{P}^e = \{[\omega, J] : \text{for some } [\omega, I] \in \mathcal{P}, \text{dist}(J, I) \leq (300b^\varepsilon)^{-1}|I|\}.$$

Our claim is that  $\mathcal{P}^e$  essentially obeys Lemma 4.30.

**Lemma 4.39.** *If  $\mathcal{P}$  is a thicket with  $\text{size}(p) \geq b$  for all  $p \in \mathcal{P}$ , then the following inequalities hold for all  $\varepsilon > 0$ .*

$$\left\| \left( \sum_{|\delta| \leq 5} \sum_{[\omega, J] \in \mathcal{P}^e} \left| \frac{1}{\sqrt{|J|}} \mathbf{1}_J(x) \langle g, \chi_{\delta, [\omega, J]} \rangle \right|^2 \right)^{1/2} \right\|_r \leq C_{\varepsilon, r} b^{-2\varepsilon} \|g\|_\infty,$$

where  $r < \infty$ .

PROOF. It suffices to prove the estimate above with the BMO norm replacing the  $L^r$  norm. This in turn follows from Lemma 4.7.

We see how to apply that Lemma to the pairs  $\mathcal{P}^e$ , and the functions  $\{\chi_{\delta,p} : p \in \mathcal{P}^e\}$ , where  $|\delta| \leq 5$  is fixed. The assumption (4.8) holds, because  $\mathcal{P}$  is a thicket and we have proved Lemma 4.4. The condition (4.9) follows from the definition of the  $\chi_{\delta,p}$ . And (4.9) follows from (4.35).

The assumption of Lemma 4.7 that is not immediate is that  $\mathcal{P}^e$  be a thicket. Indeed, it will not be so, in general. Yet if  $\mathcal{P}$  is spindly, then  $\mathcal{P}^e$  is a thicket. Therefore our Lemma follows from Lemma 4.7 if  $\mathcal{P}$  is spindly. Continuing in this vein, we have Lemma 4.21. And therefore, by Lemma 4.23 our current Lemma holds under the more general assumption that  $\mathcal{P}$  be a hedge. Finally, we have Lemma 4.28, so that our Lemma is seen to hold under the sole assumption that  $\mathcal{P}$  is a thicket.

We can now give the final proof of this section.

PROOF OF LEMMA 4.2. We can assume that  $b/2 \leq \text{size}(p) \leq b$  for all  $p \in \mathcal{P}$ . For if not, we write  $\mathcal{P} = \bigcup_n \mathcal{P}_n$ , where  $\mathcal{P}_n = \{p \in \mathcal{P} : 2^{-n-1} \leq \text{size}(p) \leq 2^{-n}b\}$ . For each  $n$  we will prove the estimate

$$\|T^{\mathcal{P}_n}g\| \leq C_r (2^{-n}b)^{\delta(r)} \|g\|_r, \quad 1 < r < 2.$$

This is summed over  $n$  to conclude the Lemma.

Recall that we wish to establish the inequalities (4.33). The interesting part is to establish the bound

$$(4.40) \quad \|T^{\mathcal{P}^*}g\|_2 \leq C_\varepsilon b^{1/2-\varepsilon} \|g\|_\infty, \quad \varepsilon > 0.$$

Henceforth,  $g$  will be assumed to be bounded by 1.

Make the following definitions. For  $p = [\omega, I] \in \mathcal{P}$ , denote

$$\begin{aligned} \mathcal{P}(p) &= \{[\omega', I'] \in \mathcal{P} : \omega \subset \omega', \text{dist}(I', I) < b^{-\varepsilon}|I|\}, \\ \mathcal{P}^b(p) &= \{[\omega', I'] \in \mathcal{P} : \omega \subset \omega', \text{dist}(I', I) \geq b^{-\varepsilon}|I|\}. \end{aligned}$$

Expand

$$\begin{aligned} \|T^{\mathcal{P}^*}g\|_2^2 &\leq \sum_{p \in \mathcal{P}} \|T_p^*g\|_2^2 + \sum_{p \in \mathcal{P}} |\langle T_p^*g(x), T^{\mathcal{P}(p)^*}g(x) \rangle| \\ &\quad + \sum_{p \in \mathcal{P}} |\langle T_p^*g(x), T^{\mathcal{P}^b(p)^*}g(x) \rangle| \\ &= \mathcal{A} + \mathcal{B} + \mathcal{C}. \end{aligned}$$

This is justified since  $\mathcal{F}T_p^*$  is supported in  $A(\omega) \subset \omega$ , hence if  $\omega \cap \omega' = \emptyset$ ,  $\langle T_p^*g, T_{p'}^*g \rangle = 0$ .

For the diagonal term, by Lemma 4.38, we can write

$$\begin{aligned} \mathcal{A} &= \sum_{p \in \mathcal{P}} \|T_p^*g\|_2^2 \\ &\leq Cb^2 + \sum_{p \in \mathcal{P}} \|T_p^*Q_p g\|_2^2 \\ &\leq Cb^2 + Cb^2 \sum_{p \in \mathcal{P}} \|Q_p g\|_2^2 \\ &\leq Cb^2 + Cb^{2-4\epsilon} \sum_{p \in \mathcal{P}^e} \|\chi_p \langle g, \chi_p \rangle\|_2^2 \\ &\leq Cb^{2-10\epsilon}, \end{aligned}$$

where in the last line we invoke Lemma 4.39. (And again, the  $|\delta| < 5$  is dropped from notation).

The diagonal estimate also enters into the second term,  $\mathcal{B}$ .

$$\begin{aligned} \mathcal{B} &= \sum_{p \in \mathcal{P}} |\langle T_p^*g(x), T^{\mathcal{P}(p)^*}g(x) \rangle| \\ &\leq Cb \sum_{p \in \mathcal{P}} \left\langle |T_p^*g(x)|, \sum_{[\omega', I'] \in \mathcal{P}(p)} (M\mathbf{1}_{I'}(x))^2 \right\rangle \\ &\leq Cb \sum_{p \in \mathcal{P}} \|T_p^*g\|_2 \left\| \sum_{[\omega', I'] \in \mathcal{P}(p)} (M\mathbf{1}_{I'}(x))^2 \right\|_2 \end{aligned}$$

But the last term on the right is no more than  $C\sqrt{b^{-\epsilon}|I|}$ , since the sets  $\{I' : [\omega', I'] \in \mathcal{P}(p)\}$  are disjoint and contained in  $b^{-\epsilon}I$ . We continue as follows.

$$\mathcal{B} \leq Cb^{1-2\epsilon} \left( \sum_{p \in \mathcal{P}} \|T_p^*g\|_2^2 \right)^{1/2} \left( \sum_{[\omega, I] \in \mathcal{P}} b^{-\epsilon}|I| \right)^{1/2} \leq Cb^{1-15\epsilon}.$$

The last term,  $\mathcal{C}$ , is the least interesting. All the intervals  $\{I' : [\omega', I'] \in \mathcal{P}^b(p)\}$  are much smaller than  $I$ , and are pairwise disjoint. So if  $J$  is any interval of length  $|I|$ ,

$$\begin{aligned} \int_J |T_p^* g(x) T^{\mathcal{P}^b(p)*} g(x)| dx &\leq C_\varepsilon b \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^{-10/\varepsilon} \\ &\quad \cdot \int_J \sum_{[\omega', I'] \in \mathcal{P}^b(p)} (M \mathbf{1}_{I'})^2(x) dx \\ &\leq C_\varepsilon b^{10} \left(1 + \frac{\text{dist}(I, J)}{|I|}\right)^{-5} |J|. \end{aligned}$$

It follows that

$$\mathcal{C} \leq C_\varepsilon b^9 \sum_{[\omega, I] \in \mathcal{P}} |I| \leq C b^6.$$

This finishes the proof of the  $L^2$  estimate.

For the  $L^r$  bound,  $r > 2$ , use Lemma 4.38 and Lemma 4.39 to see that

$$\begin{aligned} \left\| \sum_{p \in \mathcal{P}} T_p^* g \right\|_r &\leq C b + \left\| \sum_{p \in \mathcal{P}} T_p^* (Q_p g)(x) \right\|_r \\ &\leq C b + \left\| \sum_{p \in \mathcal{P}} (|k_\omega| * |f_p|^2(x) |k_\omega| * |Q_p g|^2(x))^{1/2} \right\|_r \\ (4.41) \quad &\leq C_r b + \left\| \left( \sum_{p \in \mathcal{P}} |k_\omega| * |f_p|^2(x) \right)^{1/2} \right\|_r \\ &\quad \cdot \left\| \left( \sum_{p \in \mathcal{P}} |k_\omega| * |Q_p g|^2(x) \right)^{1/2} \right\|_r \\ &\leq C_{\varepsilon, r} b^{-3\varepsilon}. \end{aligned}$$

Hence the Lemma follows by interpolating with the better  $L^2$  bound.

### 5. Trees.

The emphasis in this section will be on sets of pairs  $\mathcal{P}$  which are trees under the partial order on pairs. Call a set of pairs  $\mathcal{P}$  a *tree with top*  $p^t = [\omega^t, I^t]$  if all  $p \in \mathcal{P}$  are less than  $p^t$ . (The top need not be in the tree, nor is the top unique.) It turns out that the  $T^{\mathcal{P}}$  are familiar

objects, namely Calderón-Zygmund operators. We need to estimate  $\|T^{\mathcal{P}}\|_2$  and quantify the orthogonality between trees.

**Lemma 5.1.** *Let  $\mathcal{P}$  be a tree then*

$$\|T^{\mathcal{P}}\|_r \leq C_r, \quad 1 < r < \infty.$$

The norm estimate here is  $O(1)$ , which means that we will have to identify that part of a tree which contributes to the large norm estimate. This is the purpose of the next definitions.

**Flavors of Trees.** Call a tree  $\mathcal{P}$  with top  $p^t = [\omega^t, I^t]$  an  $\alpha$ -tree if  $c(\omega^t) \in A(\omega)$  for all  $[\omega, I] \in \mathcal{P}$ . Call  $\mathcal{P}$  an  $\alpha^*$ -tree if  $2c(\omega^t) \in A^*(\omega)$  for all  $[\omega, I] \in \mathcal{P}$ . And call  $\mathcal{P}$  a  $\beta$ -tree if  $c(\omega^t) \notin A(\omega)$  and  $2c(\omega^t) \notin A^*(\omega)$  for all  $[\omega, I] \in \mathcal{P}$ .

Of these three flavors of trees,  $\beta$ -trees are the easiest, since they are especially nice Calderón-Zygmund operators. The other flavors of trees are essentially the paraproducts, as we shall see.

**PROOF OF LEMMA 5.1.** For  $\omega \supset \omega^t$ , let  $\mathcal{P}(\omega) = \{p \in \mathcal{P} : p = [\omega, I] \text{ for some } I\}$ . This is useful since  $\mathcal{P} = \bigcup_{\omega \supset \omega^t} \mathcal{P}(\omega)$ . That is, the relevant frequency intervals form an increasing sequence. Let  $b = \sup_{p \in \mathcal{P}} \text{size}(p)$ .

The top interval  $\omega^t$  can be assumed to be centered at the origin, by the considerations in (2.15). This means that

$$|f_p(x) - f_p(y)| \leq C b \frac{|x - y|}{|I|}, \quad \text{for all } p \in \mathcal{P}.$$

So, writing

$$T^{\mathcal{P}}g(x) = \int \sum_{p \in \mathcal{P}} k_{\omega}(x - y) f_p(2x - y) g(y) dy = \int K(x, y) g(y) dy,$$

one easily checks that  $K(x, y)$  is a standard Calderón-Zygmund kernel. In particular,  $K$  satisfies the conditions

$$(5.2) \quad |K(x, y)| \leq \frac{C b}{|x - y|},$$

and if  $2|x - z| \leq |x - y|$ ,

$$(5.3) \quad |K(x, y) - K(z, y)| + |K(y, x) - K(y, z)| \leq Cb \frac{|x - z|}{|x - y|^2}.$$

It therefore suffices to establish the boundedness of  $T^{\mathcal{P}}$  on  $L^2$ .

In case of a  $\beta$ -tree everything is easy; we see that  $T^{\mathcal{P}}\mathbf{1} = T^{\mathcal{P}*}\mathbf{1} = 0$ , so that the  $L^2$  boundedness follows from Cotlar's Lemma. In particular,  $\|T^{\mathcal{P}}\|_2 \leq Cb$ .

The two remaining two cases are duals of one another, so we only consider the case of an  $\alpha$ -tree. In that case  $T^{\mathcal{P}}\mathbf{1} \neq 0$  but  $T^{\mathcal{P}*}\mathbf{1} = 0$ . Indeed,  $T_p g$  acts on  $\mathcal{F}^{-1}\mathbf{1}[A(\omega)]\mathcal{F}g$ , by (2.14), and  $0 \notin A^*(\omega)$  as  $\mathcal{P}$  is an  $\alpha$ -tree. Thus  $T_p^*\mathbf{1} = 0$ . On the other hand,  $T_p\mathbf{1}$  need not be zero.

For pairs  $p \in \mathcal{P}$ , set

$$\frac{T_p\mathbf{1}(x)}{\langle f, \phi_p \rangle} = \alpha_p(x) = \int k_\omega(y) \phi_p(x + y) dy.$$

One easily checks that  $\text{supp}(\widehat{\alpha_{[\omega, I]}}) \subset A^*(\omega)$ , which are disjoint lacunary intervals, as  $\omega$  varies. One can also see that the inequality below holds.

$$(5.4) \quad \left\| \sum_p [p = [\omega, I] \text{ for some } I] a_p \alpha_p(x) \right\|_2^2 \leq C \sum_p |a_p|^2,$$

where  $\omega$  is fixed. See for instance [D, equation (3.4.4)]. Then

$$T^{\mathcal{P}}\mathbf{1}(x) = \sum_{p \in \mathcal{P}} \langle f, \phi_p \rangle \alpha_p(x),$$

and the BMO norm of this last term is easily seen to be

$$(5.5) \quad \|T^{\mathcal{P}}\mathbf{1}\|_{\text{BMO}} \leq C_* \sup_U \left( \frac{1}{|U|} \sum_{[\omega, I] \in \mathcal{P}} [I \subset U] |\langle f, \phi_p \rangle|^2 \right)^{1/2}.$$

The supremum is finite as  $f$  is bounded. This estimate is in general, sharp. Namely, for all  $\varepsilon > 0$ , one can construct a bounded function  $f$  and an  $\alpha$ -tree  $\mathcal{P}$  so that  $\text{size}(p) < \varepsilon$  for all  $p \in \mathcal{P}$ , and  $\|T^{\mathcal{P}}\mathbf{1}\|_{\text{BMO}} \simeq 1$ .

This lemma, and it's proof, demonstrate clearly a central reason why the bilinear Hilbert transform is so difficult to understand: regardless of the size of individual  $T_p$ 's their sum can (and will) have large



norm. And any estimate of the form  $T^{\mathcal{P}} : L^2 \rightarrow L^1$ , say will also be  $O(1)$ . Thus the weaker inequalities of the previous section are of no use for the paraproducts.

Despite the bad features of  $\alpha$ -trees, they do admit a certain control. Namely, there cannot be a great number of disjoint  $\alpha$ -trees, all of a fixed norm. This point is made precise with a Carleson measure estimate in the next section.

These considerations suggest the following definition for the intrinsic size of an arbitrary collection of pairs  $\mathcal{P}$ . Set

$$(5.6) \quad \text{size}(\mathcal{P}) = \sup \{ \|T^{\mathcal{P}'}\|_2 : \text{for all trees } \mathcal{P}' \subset \mathcal{P} \}.$$

We shall see, in Section 7, that this really is the appropriate notion for the size of  $\mathcal{P}$ .

The objective we take up in the remainder of this section is to provide a control for large unions of trees with small norm. And in the Lemmas, we will make no distinction on the flavor ( $\alpha$ ,  $\alpha^*$ , or  $\beta$ ) of the trees involved.

**Normal Trees.** Let  $0 < b < 1$ . ( $b$  is associated with the size of pairs and trees). Beginning with this definition, we will use a parameter  $A > 3$  which won't play a role until the end of proof. Call a tree  $\mathcal{P}$  with top  $p^t = [\omega^t, I^t]$  *normal* if these two conditions are satisfied. For all  $[\omega, I] \in \mathcal{P}$ , one has  $|I| \leq (b/A)^{40000} |I^t|$ ; and  $\text{dist}(I, \partial I^t) \geq (b/A)^{400} |I^t|$ .

For a normal tree we nearly have that

$$T^{\mathcal{P}} g \simeq \mathbf{1}[I^t] T^{\mathcal{P}}(g \mathbf{1}[I^t]).$$

We make this precise, and due to some Fourier calculations yet to be done, the truncation in the space variable needs to be done in a smooth way. Thus let  $\zeta(x)$  be a smooth function satisfying

$$(5.7) \quad \mathbf{1}_{[-1/4, 1/4]}(x) \leq \zeta(x) \leq \mathbf{1}_{[-1/2, 1/2]}(x),$$

and  $|\mathcal{F}\zeta(\xi)| \leq C |\xi|^{-100}$ . For an interval  $J$  let  $\zeta_J(x) = \zeta((x - c(J))/|J|)$ . Define for  $p = [\omega, I]$  in a normal tree  $\mathcal{P}$ ,

$$(5.8) \quad T_{\#p} g(x) = \zeta_{\mu(I)}(x) T_p(\zeta_{\mu(I)} g)(x),$$

where  $\mu(I)$  is  $(b/A)^{100} \sqrt{|I^t|/|I|} I$ . Normality guarantees that  $\mu(I) \subset I^t$ . Then let  $T_{\#}^{\mathcal{P}} = \sum_{p \in \mathcal{P}} T_{\#p}$ .

**Lemma 5.9.** *For a normal tree  $\mathcal{P}$ , we have the inequality*

$$(5.10) \quad |T_{\#}^{\mathcal{P}}g(x) - T^{\mathcal{P}}g(x)| \leq C \left(\frac{b}{A}\right)^{1000} (M\mathbf{1}_{I^t})^2(x) Mg(x).$$

*The same inequality holds for  $T_{\#}^{\mathcal{P}^*}$ . In particular,*

$$(5.11) \quad \|T_{\#}^{\mathcal{P}}\|_r \leq C_r \max \left\{ \text{size}(\mathcal{P}), \left(\frac{b}{A}\right)^{100} \right\}, \quad 1 < r < \infty.$$

PROOF. For  $j = 1, 2, \dots$  let

$$\mathcal{P}_j = \{[\omega, I] \in \mathcal{P} : |I| = 3^{-j}|I^t|\}.$$

(As  $\mathcal{P}$  is normal,  $\mathcal{P}_j$  is empty if  $j \leq 40000 \log A/b$ .) This set is linearly ordered in the  $I$  coordinate. We can estimate for  $[\omega, I] \in \mathcal{P}$

$$\begin{aligned} & |T_{\#[\omega, I]}g(x) - T_{[\omega, I]}g(x)| \\ & \leq \mathbf{1}\left[\frac{\mu(I)}{8}\right](x) \int_{x-y \notin \mu(I)/4} |k_{\omega}(y) f_p(x+y) g(x-y)| dy \\ & \quad + \mathbf{1}\left[\left(\frac{\mu(I)}{8}\right)^c\right](x) \int_{-\infty}^{\infty} |k_{\omega}(y) f_p(x+y) g(x-y)| dy \\ & \leq C Mg(x) \left( \left(\frac{|\mu(I)|}{|I|}\right)^{-20} \mathbf{1}\left[\frac{\mu(I)}{8}\right](x) \right. \\ & \quad \left. + \mathbf{1}\left[\left(\frac{\mu(I)}{8}\right)^c\right](x) \left(\frac{\text{dist}(x, \mu(I)/16)}{|I|}\right)^{-50} \right). \end{aligned}$$

This follows on the one hand by the decay estimate for  $k_{\omega}$ , and on the other by the decay of  $f_p$ .

Sum this estimate over  $p \in \mathcal{P}_j$ .

$$\begin{aligned} & \sum_{[\omega, I] \in \mathcal{P}_j} |T_{\#[\omega, I]}g(x) - T_p g(x)| \\ & \leq C \left(\frac{|\mu(I)|}{|I|}\right)^{-10} \left(1 + \frac{\text{dist}(x, I^t)}{|I^t|}\right)^{-2} Mg(x) \\ & \leq C \left(\frac{b}{A}\right)^{1000} 3^{-5j} \left(1 + \frac{\text{dist}(x, I^t)}{|I^t|}\right)^{-2} Mg(x). \end{aligned}$$

This is summed over  $j$  such that  $3^{-j} \leq (b/A)^{400}$ , and so yields the Lemma.

**Separation of Trees.** Two normal trees  $\mathcal{P}$  with top  $[\omega^t, I^t]$  and  $\mathcal{P}'$  with top  $[\omega^{t'}, I^{t'}]$  are *separated* if  $I^t \cap I^{t'} = \emptyset$ , or otherwise if

$$\alpha) p' = [\omega', I'] \in \mathcal{P}', \text{ and } I' \subset I^t \text{ implies } \text{dist}(\omega', \omega^t) > (A/b)^{3000} |\omega'|.$$

$$\beta) p = [\omega, I] \in \mathcal{P}, \text{ and } I \subset I^{t'} \text{ implies } \text{dist}(\omega, \omega^{t'}) > (A/b)^{3000} |\omega|.$$

$\gamma$ ) and finally, assuming  $I^{t'} \subset I^t$ , for all  $[\omega, I] \in \mathcal{P}$ ,  $|I| \leq (b/A)^{500} |I^{t'}|$  implies  $\text{dist}(I, \partial I^{t'}) \geq (b/A)^{400} |I^{t'}|$ .

This next Lemma quantifies the essential orthogonality between trees.

**Lemma 5.12.** *For separated trees as above, we have*

$$\|T_{\#}^{\mathcal{P}} T_{\#}^{\mathcal{P}'*}\|, \|T_{\#}^{\mathcal{P}*} T_{\#}^{\mathcal{P}'}\|_2 \leq C \left(\frac{b}{A}\right)^{500}.$$

*More importantly, we have the following local estimates on inner products. Assuming that  $I^{t'} \subset I^t$ ,*

$$(5.13) \quad |\langle T_{\#}^{\mathcal{P}} g, T_{\#}^{\mathcal{P}'} h \rangle| \leq C \left(\frac{b}{A}\right)^{500} \|Mg + |T_{\#}^{\mathcal{P}} g|\|_{L^2(I^{t'})} \|h\|_{L^2(I^{t'})}.$$

*A similar inequality holds for  $T_{\#}^{\mathcal{P}*}$  and  $T_{\#}^{\mathcal{P}'*}$ .*

In the special case of  $I^t = I^{t'}$ , we have  $T^{\mathcal{P}*} T^{\mathcal{P}'} = T^{\mathcal{P}} T^{\mathcal{P}'*} = 0$ , which is obvious from the Fourier transform side, and so the norm estimate above is easy. The local inner product estimate, which is essential, contains more information and so the proof is not as easy. But the underlying idea is no different.

PROOF. The interesting case is  $I^{t'} \subset I^t$ . Let

$$\lambda' = \min \{|I'| : [\omega', I'] \in \mathcal{P}'\},$$

and split  $\mathcal{P}$  into the trees  $\mathcal{P}^{\#} = \{[\omega, I] \in \mathcal{P} : |I| \geq \lambda'\}$  and  $\mathcal{P}^b = \{[\omega, I] \in \mathcal{P} : |I| < \lambda'\}$ . A technical point hinges on the fact that a pair  $[\omega, I] \in \mathcal{P}^b$  could satisfy  $I \cap I^{t'} = \emptyset$  and yet  $\mu(I) \supset I^{t'}$ . We treat the point by redefining  $T_{\#p}$ .

Recalling (5.8),  $T_{\#}^{\mathcal{P}}$  is defined in terms of the intervals  $\mu(I)$ . Instead, define

$$(5.14) \quad \check{\mu}(I) = \left(\frac{b}{A}\right)^{100} \sqrt{\frac{|I^{t'}|}{|I|}} I,$$

and

$$\check{T}^{\mathcal{P}^b} g = \sum_{p \in \mathcal{P}^b} \zeta_{\check{\mu}(I)} T_p(\zeta_{\check{\mu}(I)} g).$$

Then a trivial adaptation of Lemma 5.9 shows that

$$|T_{\#}^{\mathcal{P}^b} g(x) - \check{T}^{\mathcal{P}^b} g(x)| \leq C \left(\frac{b}{A}\right)^{500} Mg(x),$$

and so to prove the lemma it suffices to replace  $T_{\#}^{\mathcal{P}^b}$  by  $\check{T}^{\mathcal{P}^b}$ . Now observe that

$$\langle \check{T}^{\mathcal{P}^b} g, T_{\#}^{\mathcal{P}'} g \rangle = \sum_{[\omega, I] \in \mathcal{P}} \sum_{[\omega', I'] \in \mathcal{P}'} [\check{\mu}(I) \cap \mu(I') \neq \emptyset] \langle \check{T}_p g, T_{\#p'} h \rangle.$$

But now, for  $[\omega, I] \in \mathcal{P}$  and  $[\omega', I'] \in \mathcal{P}'$ , part  $\gamma$ ) of the definition of separated shows that if  $\check{\mu}(I) \cap \mu(I') \neq \emptyset$ , then  $I \subset I^{t'}$ . This is because  $|I| < |I'|$ , and so  $|\check{\mu}(I)| < |\mu(I')|$ . So

$$\begin{aligned} \text{dist}(I, I') &\leq 2|\mu(I')| \\ &\leq \left(\frac{b}{A}\right)^{100} \sqrt{|I^{t'}| |I|} \\ &\leq \left(\frac{b}{A}\right)^{20000} |I^{t'}| \\ &\leq \text{dist}(I', \partial I^{t'}), \end{aligned}$$

by the definition of normality. But then  $I \subset I^{t'}$ .

With this done, we can assume that for all  $[\omega, I] \in \mathcal{P}^{\#}$ ,  $\mu(I) \cap I^{t'} \neq \emptyset$ , and for all  $[\omega, I] \in \mathcal{P}^b$ , we have  $I \subset I^{t'}$ . Then, let

$$\lambda = \min \{|I'| : [\omega', I'] \in \mathcal{P}' \cup \mathcal{P}^b\}.$$

By an abuse of notation, we will write  $T^{\mathcal{P}} = T_{\#}^{\mathcal{P}^{\#}} + \check{T}^{\mathcal{P}^b}$  for the purposes of this proof.

From the definition of separated it follows that for all  $[\omega, I] \in \mathcal{P}$  and  $[\omega', I'] \in \mathcal{P}'$

$$\text{dist}(\omega, \omega') \geq \left(\frac{A}{b}\right)^{3000} \lambda^{-1} := \left(\frac{A}{b}\right)^{500} D.$$

Here  $D = (A/b)^{2500} / \lambda \geq (A/b)^{40000} |I^{t'}|$ . We shall see that the lemma is true because  $\widehat{T^{\mathcal{P}}}$  and  $\widehat{T_{\#}^{\mathcal{P}'}}$  live on disjoint sets.

Let  $c$  and  $c'$  denote respectively the centers of  $\omega^t$  and  $\omega^{t'}$ . Recalling that  $\widehat{T_p}$  is supported on  $A^*(\omega)$  which is centered at  $2c(\omega) + |\omega|/3$ , let  $\Phi(x)$  be a function with

$$(5.15) \quad \text{supp}(\Phi) \subset \left[-\frac{1}{D}, \frac{1}{D}\right],$$

$$(5.16) \quad |\check{\Phi}(\xi) - 1| \leq C \left(\frac{|\xi - 2c'|}{D}\right)^{-50},$$

$$(5.17) \quad \|\Phi(x)\|_1 \leq C,$$

and

$$(5.18) \quad \check{\Phi}(2c) = 0.$$

Since  $|c - c'|$  is so large, there is no problem accommodating this last condition.

Write

$$\mathcal{E}'h = T_{\#}^{\mathcal{P}'} h - \Phi * T_{\#}^{\mathcal{P}'} h.$$

Note that as  $D^{-1}$  is so small, the definition of normal and (5.15) imply that  $\mathcal{E}'h$  is supported in  $I^{t'}$ . Further, we can write

$$\langle T^{\mathcal{P}} g, T_{\#}^{\mathcal{P}'} h \rangle = \langle T^{\mathcal{P}} g, \Phi * T_{\#}^{\mathcal{P}'} h \rangle + \langle T^{\mathcal{P}} g, \mathcal{E}'h \rangle = \mathcal{A} + \mathcal{B}.$$

Our first claim is that

$$(5.19) \quad |T^{\mathcal{P}*}(\Phi * H)(x)| \leq C \left(\frac{b}{A}\right)^{500} \chi * |H|(x),$$

where  $\chi(x) = D(1 + D|x|)^{-3/2}$ . From this, it follows that

$$\begin{aligned} |\mathcal{A}| &= |\langle g, T^{\mathcal{P}*}(\Phi * T_{\#}^{\mathcal{P}'} h) \rangle| \\ &\leq C \left(\frac{b}{A}\right)^{500} \langle |g|, \chi * |T_{\#}^{\mathcal{P}'} h| \rangle \\ &= C \left(\frac{b}{A}\right)^{500} \langle \chi * |g|, |T_{\#}^{\mathcal{P}'} h| \rangle \\ &\leq C \left(\frac{b}{A}\right)^{500} \|Mg\|_{L^2(I^{t'})} \|h\|_{L^2(I^{t'})}. \end{aligned}$$

This is the principal estimate.

The second claim is that

$$(5.20) \quad \|\mathcal{E}'\|_2 \leq C \left(\frac{b}{A}\right)^{500},$$

hence, as  $\text{supp } \mathcal{E}'h \subset I^{t'}$ , and  $\mathcal{E}'h = \mathcal{E}'(h\mathbf{1}_{I^{t'}})$ ,

$$|\mathcal{B}| \leq \langle |T^{\mathcal{P}}g|, |\mathcal{E}'h| \rangle \leq C \left(\frac{b}{A}\right)^{500} \|T^{\mathcal{P}}g\|_{L^2(I^{t'})} \|h\|_{L^2(I^{t'})}.$$

The estimates on  $\mathcal{A}$  and  $\mathcal{B}$  prove the Lemma.

We turn to the proof of the two claims (5.19) and (5.20). There is no harm in assuming that  $\omega^t$  is centered at the origin. See (2.15). For the proof of (5.19), write

$$T^{\mathcal{P}*}H(x) = \int K(x, y) H(y) dy.$$

As the cut-off function  $\zeta$  in (5.7) is assumed to be smooth,  $K(x, y)$  is a generalized Calderón-Zygmund kernel, and in particular satisfies the gradient condition (5.3) above, with  $b$  in that inequality replaced by 1. (Recall that we are not making any assumption about the size of pairs in the current Lemma). Then  $g \rightarrow T^{\mathcal{P}*}(\Phi * g)$  has kernel

$$\int K(x, z) \Phi(z - y) dz.$$

Note that  $\int \Phi = 0$ , due to (5.18), and our assumption that  $c = 0$ ; also the support of  $\Phi$  is small, see (5.15). There are two estimates to be made. On the one hand, with the assumption  $|I| > \lambda = (A/b)^{2500}/D$  for all  $p \in \mathcal{P}$ , we see that

$$\left| \frac{d}{dz} K(x, z) \right| \leq C \lambda^{-2} = C \left(\frac{b}{A}\right)^{5000} D^2, \quad \text{for all } x \text{ and } z.$$

Consequently using (5.15) and  $\int \Phi dx = 0$ ,

$$\begin{aligned} \left| \int K(x, z) \Phi(z - y) dz \right| &= \left| \int_{|z-y| \leq 2/D} (K(x, z) - K(x, y)) \Phi(z - y) dz \right| \\ &\leq \left(\frac{b}{A}\right)^{2500} D. \end{aligned}$$

On the other hand, if  $|x - y| > 4/D$  use (5.3) and (5.17) to see that

$$\begin{aligned} \left| \int K(x, z) \Phi(z - y) dz \right| &= \left| \int (K(x, z) - K(x, y)) \Phi(z - y) dz \right| \\ &\leq C |x - y|^{-2} \int_{|z-y| \leq 2/D} |z - y| |\Phi(z - y)| dz \\ &\leq C D^{-1} |x - y|^{-2}. \end{aligned}$$

Notice that

$$\begin{aligned} D^{-1}|x|^{-2} &\leq \left(\frac{b}{A}\right) D (1 + D|x|)^{-3/2}, \quad \text{if } D|x| \geq \left(\frac{A}{b}\right)^{1000}, \\ D \left(\frac{b}{A}\right)^{2000} &= \min_{D|x| \leq (A/b)^{1000}} \left(\frac{b}{a}\right)^{500} D (1 + |x|)^{3/2}, \end{aligned}$$

so that (5.19) follows.

For the second claim (5.20), verify the dual inequality

$$\|T_{\#}^{\mathcal{P}'*}(h - \Phi * h)\|_2 \leq C \left(\frac{b}{A}\right)^{500} \|h\|_2.$$

But we can estimate by (5.10), Lemma 5.1, (2.14) and (5.16) to see that

$$\begin{aligned} \|T_{\#}^{\mathcal{P}'*}(h - \Phi * h)\|_2 &\leq C \left(\frac{b}{A}\right)^{1000} \|M(h - \Phi * h)\|_2 + \|T^{\mathcal{P}'*}(h - \Phi * h)\|_2 \\ &\leq C \left(\frac{b}{A}\right)^{1000} \|h\|_2 + C \left\| \mathbf{1} \left[ 2c' - \frac{3}{\lambda}, 2c' + \frac{3}{\lambda} \right] \mathcal{F}(h - \Phi * h) \right\|_2 \\ &\leq C \left(\frac{b}{A}\right)^{1000} \|h\|_2. \end{aligned}$$

The case of the adjoints being similar, we have completed the proof of the Lemma.

A point which will come up several times below is that the intervals  $I_p$  can range over the whole of the real line. (This was also an issue in the preceding section). Here, let us note that if  $\mathcal{P}$  is a tree with top  $t$ , then

$$(5.21) \quad \text{size}(\mathcal{P}) \geq b \quad \text{implies} \quad I^t \subset b^{-\varepsilon}[-1, 1],$$

as easily follows from the decay of  $\varphi_p$ .

The emphasis of the next group of Lemmas is on forming large unions of trees.

**Rows.** Define a *row* to be a union of normal trees  $\mathcal{P}_j$  with tops  $p_j^t = [\omega_j^t, I_j^t]$  for which the sets  $\{I_j^t\}$  are pairwise disjoint. Two rows  $\mathcal{R} = \bigcup_j \mathcal{P}_j$  with tops  $[\omega_j^t, I_j^t]$ , and  $\mathcal{R}' = \bigcup_j \mathcal{P}'_j$  with tops  $[\omega_{j'}^{t'}, I_{j'}^{t'}]$  are *separated* if each  $I_{j'}^{t'}$  is contained in some  $I_j^t$ , where  $\mathcal{P}_j$  and  $\mathcal{P}'_{j'}$  are separated.

**Lemma 5.22.** *Let  $\mathcal{R}$  be a row with  $\text{size}(\mathcal{R}) \leq b$ . Then*

$$(5.23) \quad \|T_{\#}^{\mathcal{R}}\|_r \leq C_r b, \quad 1 < r < \infty.$$

*If  $\mathcal{R}$  and  $\mathcal{R}'$  are separated rows then*

$$(5.24) \quad \|T_{\#}^{\mathcal{R}} T_{\#}^{\mathcal{R}'*}\|_2, \|T_{\#}^{\mathcal{R}*} T_{\#}^{\mathcal{R}'}\|_2 \leq C \left(\frac{b}{A}\right)^{500}.$$

PROOF. As the operators  $T_{\#}^{\mathcal{R}}$  act coordinatewise on  $\bigoplus_j L^r(I_j^t)$ , the first assertion follows from Lemma 5.9.

And for the second assertion, again due to the coordinatewise action of the operators, it suffices to consider the case where  $\mathcal{R}$  is in fact a tree  $\mathcal{P}$  with top  $[\omega^t, I^t]$ , and each  $I_{j'}^{t'}$  is contained in  $I^t$ . Then, using (5.13) and Cauchy-Schwartz,

$$\begin{aligned} |\langle T_{\#}^{\mathcal{P}} g, T_{\#}^{\mathcal{R}'} h \rangle| &= \sum_{j'} |\langle T_{\#}^{\mathcal{P}} g, T_{\#}^{\mathcal{R}'_{j'}} h \rangle| \\ &\leq C \left(\frac{b}{A}\right)^{500} \sum_{j'} \|Mg + |T_{\#}^{\mathcal{P}} g|\|_{L^2(I_{j'}^{t'})} \|h\|_{L^2(I_{j'}^{t'})} \\ &\leq C \left(\frac{b}{A}\right)^{500} \|g\|_{L^2(I^t)} \|h\|_{L^2(I^t)}. \end{aligned}$$

The last Lemma of this section provides an estimate for a large number of rows.

**Lemma 5.25.** *Let  $\mathcal{R} = \mathcal{R}_1 \cup \dots \cup \mathcal{R}_N$  be pairwise separated rows where the number of rows  $N$  is at most  $C(A/b)^{220}$ . If  $\text{size}(\mathcal{R}) \leq b$ , then, for  $1 < r < 2$ ,*

$$\|T^{\mathcal{R}} g\|_r \leq C_r b \|g\|_2.$$



PROOF. Cotlar’s Lemma, and Lemma 5.22 provide the estimate

$$\left\| \sum_{n=1}^N T_{\sharp}^{\mathcal{R}_n} g \right\|_2 \leq C b \|g\|_2 .$$

To finish the Lemma, we need to remove the sharp above. We do this by assuming that for each tree  $\mathcal{T}$  in a row, we have  $\text{size}(\mathcal{T}) \geq b/2$ , so that (5.21) is in force. Denote the top space intervals of the trees in the  $n$ -th row by  $\{I_{n,k} : k \geq 1\}$ . These intervals are disjoint in  $k$ . For a choice of  $1 < r < 2$ , let  $1/s + 1/2 = 1/r$ . We have by Lemma 5.9,

$$\begin{aligned} \left\| \sum_{n=1}^N |T_{\sharp}^{\mathcal{R}_{n^*}} g - T^{\mathcal{R}_{n^*}} g| \right\|_r &\leq C \left(\frac{b}{A}\right)^{1000} \left\| \sum_{n=1}^N \sum_{k=1}^{\infty} (M\mathbf{1}[I_{n,k}])^2 \right\|_s^2 \|Mg\|_2 \\ &\leq C_r \left(\frac{b}{A}\right)^{1000} \left\| \sum_{n=1}^N \sum_{k=1}^{\infty} \mathbf{1}[I_{n,k}] \right\|_s^2 \|g\|_2 \\ &\leq C_r \left(\frac{b}{A}\right)^{100} \|g\|_2 . \end{aligned}$$

**6. Orchards.**

An important point left uncovered in the previous section is the behavior of the  $\alpha$  and  $\alpha^*$  trees. These operators are in essence para-products, and can lead to operators of large norm, regardless of how small their constituent parts are. An adequate control of these objects requires a new Carleson measure estimate, which is the subject of this section.

Let  $\mathcal{P}_i, i \geq 1$ , be  $\alpha$ -trees with tops  $p_i^t = [\omega_i^t, I_i^t]$ . Assume

- i) For  $i \neq i', \mathcal{P}_i$  and  $\mathcal{P}_{i'}$  are disjoint.
- ii) For an absolute constant  $c_1 = (12 C_*)^{-1}$ , with  $C_*$  as in (5.5),

$$c_1 b^2 |I_i^t| \leq \sum_{p \in \mathcal{P}_i} |\langle f, \phi_p \rangle|^2 .$$

- iii) For all  $p \in \bigcup_i \mathcal{P}_i, \text{size}(p) \leq b^{30}$ .

Such a collection of  $\alpha$ -trees we will refer to as an  $\alpha$ -orchard. An analogous collection of  $\alpha^*$ -trees we will refer to as an  $\alpha^*$ -orchard.<sup>10</sup>

Of these three conditions, the first is a modest regularity condition. The second is easily seen to be related to the norm of the tree operator  $T^{\mathcal{P}^i}$ . In particular, for a  $\alpha$ -tree  $\mathcal{P}$ , let

$$\lambda^2 = \sup_U \frac{1}{|U|} \sum_{p=[\omega, I] \in \mathcal{P}} [I \subset U] |\langle f, \phi_p \rangle|^2.$$

Then, our analysis from the previous section, and in particular (5.5), shows that

$$d_1 \lambda \leq \|T^{\mathcal{P}}\|_2 \leq D_1 \lambda.$$

Thus the trees in an orchard have a minimal size. The last condition iii) is in contrast to the analysis of say Section 4, in which a lower bound on pairs is imposed. Yet, in the context of  $\alpha$ -trees, the interesting case is when the sizes of individual pairs are all quite small.

An essential fact about  $\alpha$ -orchards is that the top intervals  $I_i^t$  obey a Carleson measure estimate much like the one established for sets of incomparable pairs.

**Lemma 6.1.** *There is a  $b_0 > 0$  so that for all  $0 < b < b_0$ , and any  $\alpha$ -orchard  $\mathcal{O} = \bigcup_i \mathcal{P}_i$  as above, and all intervals  $U$ ,*

$$(6.2) \quad \sum_i [I_i^t \subset U] |I_i^t| \leq C b^{-2} |U|.$$

The rest of this section is taken up with the proof of this inequality. The general approach is to identify trees  $\mathcal{P}'_i \subset \mathcal{P}_i$ , which still satisfy ii) above, with the additional property that the functions  $\{\phi_p : p \in \bigcup_i \mathcal{P}'_i\}$  form a basis in  $L^2$ . Applying this basis to our function  $f(x)$  will complete the proof.

Recall the combinatorial structure of an  $\alpha$ -tree  $\mathcal{P}$  with top frequency interval  $\omega^t$ . Each pair  $[\omega, I] \in \mathcal{P}$  has  $c(\omega^t) \in A(\omega)$ . These intervals are central triadic intervals. From (2.12), we have  $c(\omega) \simeq c(\omega^t) - |\omega|/3$ . And so the sets  $\Phi(\omega)$  of (2.9) form a lacunary sequence of intervals, from which one would have Littlewood-Paley estimates.

---

<sup>10</sup> Because condition iii) is an upper bound, any sufficiently small value of  $c_2$  is acceptable.

(See (6.3) below). Also, recalling (3.2), we have that the functions  $\phi_{[\omega, I]}$  and  $\phi_{[\omega, J]}$  are orthogonal, provided  $I \neq J$ .

The interval  $U$  in (6.2) can be fixed, and we assume that  $I_i^t \subset U$  for all  $i$ . A formally weaker statement will imply (6.2). Namely, there is an open set  $F \subset U$  so that

- $\alpha)$   $|F| \leq |U|/4$ .
- $\beta)$   $\sum_i [I_i^t \not\subset F] |I_i^t| \leq C b^{-2} |U|$ .

We then appeal to Lemma 4.23 to prove (6.2). Yet, in proving this weaker statement, we can further impose the assumption

- $\delta)$   $\| \sum_i [I_i^t \subset U] \mathbf{1}_{I_i^t}(x) \|_\infty \leq b^{-22}$ .

To be quite specific, the statement to be proved is this: given  $c_1$  in ii), we can choose  $b_0$  so that for every  $0 < b < b_0$  and every orchard which satisfies  $\delta)$  also satisfies  $\alpha)$  (with  $1/4$  replaced by  $1/2$ ) and  $\beta)$  (with a possibly larger constant).

Let us now argue that there is no loss of generality in assuming  $\delta)$ . For this, we take an arbitrary orchard,  $\mathcal{O}$  and select appropriate subcollection of the trees  $\mathcal{P}_i$  which satisfy  $\delta)$ . The subcollection should be taken in this way. Denote by  $I_{j(1,v)}^t$ , for  $v \geq 1$ , the maximal elements from among  $\{I_i^t : i \geq 1\}$ . Remove the intervals  $I_{j(1,v)}^t$  from the list, and and again take the maximal intervals  $I_{j(2,v)}^t$ , for  $v \geq 1$ . Repeat this procedure until the orchard is exhausted. The  $\alpha$ -trees  $\{\mathcal{P}_{j(k,v)} : 1 \leq k \leq K = b^{-22}, v \geq 1\}$  satisfy  $\delta)$ . Assume  $\alpha)$  and  $\beta)$  also hold. Set

$$E_0 = \left\{ x : \sum_{k=1}^{\infty} \sum_{v=1}^{\infty} \mathbf{1}[I_{j(k,v)}^t](x) \geq (2b)^{-22} \right\},$$

that is, this set is the set on which  $\delta)$  fails. We see that

$$\begin{aligned} |E| &\leq \left| \left\{ x : \sum_{k=1}^K \sum_{v=1}^{\infty} \mathbf{1}[I_{j(k,v)}^t](x) = b^{-22} \right\} \right| \\ &\leq |F| + b^{22} \sum_{k=1}^K \sum_{v=1}^{\infty} [I_{j(k,v)}^t \not\subset F] |I_{j(k,v)}^t| \\ &\leq \left( \frac{1}{4} + C b^{-20} \right) |U|. \end{aligned}$$

This will be less than  $|U|/2$ , provided  $b$  is small enough. Thus, a slightly weaker form of  $\alpha)$  holds. Taking  $E = E_0 \cup F$ , we see that

$$\alpha') |E| \leq |U|/2$$

and that  $\beta$ ) holds for the orchard  $\mathcal{O}$  with the set  $F$  replaced by the set  $E$ . Again, by Lemma 4.23, this is enough to prove (6.2).

Assume  $\delta$ ). We turn our attention to the deletion of certain small sets of pairs. First we can assume each tree  $\mathcal{P}_i$  is finite, without violating our condition ii). Second, for any  $\alpha$ -tree  $\mathcal{P}$ , we have the fact that

$$(6.3) \quad \left( \sum_{p \in \mathcal{P}} |\phi_p \otimes \phi_p|^2 \right)^{1/2} : L^\infty \longrightarrow \text{BMO}.$$

By observing that the supports of the functions  $\phi_p$  in frequency form a lacunary disjoint sequence, the operation above is seen to be an ordinary Littlewood-Paley square function, albeit conjugated by an exponential to account for the location of the tree in frequency, and then the bound above is immediate.

With this observation, we can delete pairs in  $\mathcal{P}_i$  which fall close to the boundary of  $I_i^t$ . Specifically, set

$$(6.4) \quad \mathcal{P}_i^\partial = \{[\omega, I] \in \mathcal{P}_i : \text{dist}(I, \partial I_i^t) \leq b^{100} |I_i^t|\}.$$

It follows that

$$\sum_{p \in \mathcal{P}_i^\partial} |\langle f, \phi_p \rangle|^2 \leq C b^{100} |I_i^t|.$$

Therefore, after removal of this set of pairs ii) holds with a slightly smaller constant.

The top of  $\mathcal{P}_i$  must be removed. Namely, say that  $p \in \mathcal{P}_i^t$  if there is no chain

$$p = p_1 \prec p_2 \prec \cdots \prec p_M,$$

with  $M = 10,000 \log(b/\text{size}(p))$ , and all  $p_m \in \mathcal{P}_i$ . Then it follows that

$$\sum_{p \in \mathcal{P}_i^t} |\langle f, \phi_p \rangle|^2 \leq b^5 |I_i^t|,$$

due to the condition iii) above. Thus, for  $b$  sufficiently small, these pairs can also be removed from  $\mathcal{P}_i$  with only a minimal weakening of ii). Set  $\mathcal{P}^t = \bigcup_i \mathcal{P}_i^t$ .

The import of these last two conditions is that for  $[\omega, I] \in \mathcal{P}_i \setminus (\mathcal{P}_i^\partial \cup \mathcal{P}_i^t)$ ,

$$(6.5) \quad |I| \leq (b \text{size}(p))^{5000} |I_i^t|.$$

We arrive at a critical point, which centers on the relationship between distinct trees. Consider two top intervals  $I_i^t$  and  $I_{i'}^t$  which intersect. Consider the following subtree of  $\mathcal{P}_i$ .

$$(6.6) \quad \mathcal{Q}_{i,i'} = \{[\omega, I] \in \mathcal{P}_i : I \subset I_{i'}^t, \Phi(\omega') \subsetneq \Phi(\omega) \text{ for some } [\omega', I'] \in \mathcal{P}_{i'}\}.$$

Recall that our objective is to identify a highly orthogonal set of functions  $\phi_p$ . Since  $\text{supp}(\widehat{\phi_p}) \subset \Phi(\omega)$ , the sets of pairs above are certainly a cause of concern.<sup>11</sup> But our claim is that  $\mathcal{Q}_{i,i'}$  can only admit chains of bounded length in the partial order ' $<$ '. In particular, there are no ten pairs

$$p_1 \lesssim p_2 \lesssim \cdots \lesssim p_{10},$$

with all  $p_k = [\omega_k, I_k] \in \mathcal{Q}_{i,i'}$  for  $1 \leq k \leq 10$ . For assuming otherwise, the sets  $A(\omega_k)$  are central, and  $A(\omega_{10}) \subsetneq \cdots \subsetneq A(\omega_1)$ . By the good property of centrality, (3.1),

$$\text{dist}(A(\omega_{10}), \partial A(\omega_1)) \geq 3^9 |A(\omega_{10})|.$$

Recall that  $A(\omega_{10}) \subset \omega_{10}$ , and  $|A(\omega_{10})| = |\omega_{10}|/9$ . Hence,  $9\omega_{10} \subset A(\omega_1)$ . Yet,  $c(\omega_{i'}^t) \in p_{10} \in \mathcal{O}_{i,i'}$  and so for some  $p' \in \mathcal{P}_{i'}$ ,  $\Phi(\omega') \subsetneq \Phi(\omega_{10})$ , and so  $A(\omega') \subset \omega_{10} \subset A(\omega_1)$ . This means that  $p_1 \in \mathcal{P}_i$  is in the  $\alpha$ -tree with top  $[\omega_{i'}^t, I_{i'}^t]$ . But recall that in the  $\alpha$ -tree  $\mathcal{P}_{i'}$ , the intervals  $\{\Phi(\omega') : [\omega', I'] \in \mathcal{P}_{i'}\}$  are lacunary. As a consequence, for any  $[\omega', I'] \in \mathcal{P}_{i'}$ , the intervals  $\Phi(\omega')$  and  $\Phi(\omega_1)$  are either equal or disjoint. This contradicts the assumption that  $p_1 \in \mathcal{Q}_{i,i'}$ .

Thus, this set of pairs cannot contain chains of length ten, as claimed. Using the upper bound on the size of pairs, iii), and the tree structure of  $\mathcal{P}_i$ , we see that

$$\sum_{p \in \mathcal{Q}_{i,i'}} |\langle f, \phi_p \rangle|^2 \leq 10 b^{30} |I_{i'}^t \cap I_i^t|.$$

Set  $\mathcal{Q}_i = \bigcup_{i'} \mathcal{Q}_{i,i'}$ . Using the assumption  $\delta$ ), we see that

$$\sum_{p \in \mathcal{Q}_i} |\langle f, \phi_p \rangle|^2 \leq 10 b^8 |I_i^t|.$$

---

<sup>11</sup> Recall that  $\phi_{[\omega,I]}$  and  $\phi_{[\omega,J]}$  are orthogonal for  $I \neq J$ , due to (3.2). In  $\mathcal{Q}_{i,i'}$ , we need only concern ourselves with the case of  $\omega'$  being a strict subset of  $\omega$ .

This last estimate will be quite small. The pairs in  $\mathcal{Q}_i$  can be deleted without affecting ii).

To summarize, we can without loss of generality assume that the sets  $\mathcal{P}_i^\partial$ ,  $\mathcal{P}_i^t$ , and  $\mathcal{Q}_i$  are empty for all  $i$ . For if they are not, we remove the pairs in these sets, and (a trivial weakening of) ii) above continues to hold. The essential advantage gained by these manipulations is as follows. For  $p = [\omega, I] \in \mathcal{O}$ , set

$$(6.7) \quad \mathcal{O}(p) = \{p' = [\omega', I'] \in \mathcal{O} : \Phi(\omega) \subsetneq \Phi(\omega')\}.$$

That is  $\mathcal{O}(p)$  contains all pairs  $p'$  for which  $|I'| < |I|$  and  $\phi_{p'}$  is not orthogonal to  $\phi_p$ . Then

$$(6.8) \quad \begin{aligned} \{I' : [\omega', I'] \in \mathcal{O}(p)\} &\text{ are pairwise disjoint} \\ &\text{and contained in } [(b \text{ size}(p))^{-500} I]^c. \end{aligned}$$

To see this claim, it is enough to verify that if  $p' = [\omega', I'] \in \mathcal{O}$  is such that  $\Phi(\omega) \subsetneq \Phi(\omega')$ , then

$$(6.9) \quad \text{dist}(I, I') \geq (b \text{ size}(p))^{-500} |I|.$$

This clearly proves the last half of (6.8). Yet it also proves disjointness, as is easily seen: consider  $p_1 \neq p_2 \in \mathcal{O}(p)$ , with  $p_i = [\omega_i, I_i]$ , for  $i = 1, 2$ . We want to show that  $I_1 \cap I_2 = \emptyset$ . The frequency intervals both contain  $\Phi(\omega)$ , and hence  $\omega_1 \cap \omega_2$  is not empty. If  $\omega_1 = \omega_2$ , it is clear that  $I_1$  and  $I_2$  are disjoint. And otherwise, (6.9) shows that  $I_1$  and  $I_2$  are disjoint.

We establish (6.9). As  $\Phi(\omega) \subsetneq \Phi(\omega')$ , the  $\alpha$ -tree structure dictates that the two pairs are in distinct trees:  $p \in \mathcal{P}_i$ , and  $p' \in \mathcal{P}_{i'}$  with  $i \neq i'$ . The first case is  $I' \not\subset I_i^t$ . For then, it follows from the removal of the sets  $\mathcal{P}_i^\partial$  and  $\mathcal{P}_i^t$ , that

$$(6.10) \quad \begin{aligned} \text{dist}(I, I') &\geq \text{dist}(I, \partial I_i^t) \\ &\geq b^{100} \frac{|I_i^t|}{|I|} |I| \\ &\geq (b \text{ size}(p))^{-1000} |I|. \end{aligned}$$

Recall (6.4) and (6.5). This is stronger than (6.9).

Thus we can assume that  $I' \subset I_i^t$ . But then, reversing primes in (6.6),  $p' \in \mathcal{Q}_{i',i}$ , a set of pairs that has been removed. The verification of (6.9) is complete.

The combinatorial work of the proof is complete. We have a set of pairs  $\mathcal{O}$  which satisfy (6.8). This condition is a rather strong incomparability condition. Hence the final portion of the argument is a reprise of the techniques of Section 4, and in particular the proof of Lemma 4.4.

Our objective is to establish the following inequality. Fix an interval  $U$ , and assume  $I_i^t \subset U$  for all  $i$ . For all bounded functions  $g$  supported on  $2U$ , and choices of signs  $\{\varepsilon_p : p \in \mathcal{O}\}$ ,

$$(6.11) \quad \left\| \sum_{p \in \mathcal{O}} \varepsilon_p \phi_p(x) \langle g, \phi_p \rangle \right\|_2 \leq C \sqrt{|U|} \|g\|_\infty^2 .$$

To establish it, we can assume that  $\mathcal{O}$  is finite, and therefore the inequality must hold with some finite constant on the right hand side. Let  $B$  denote the best constant in this assumed inequality; an upper bound on  $B$  can be given.

Set  $S_p = \phi_p \otimes \phi_p$ , then

$$(6.12) \quad \left\| \sum_{p \in \mathcal{O}} \varepsilon_p \phi_p \langle g, \phi_p \rangle \right\|_2^2 = \left\| \sum_{p \in \mathcal{O}} \varepsilon_p S_p g \right\|_2^2 = \mathcal{D} + \mathcal{O} ,$$

where  $\mathcal{D}$  and  $\mathcal{O}$  are the diagonal and off-diagonal terms respectively.

In the diagonal, we can use the fact that  $S_p$  is a self-adjoint projection to write

$$(6.13) \quad \begin{aligned} \mathcal{D} &= \sum_{p \in \mathcal{O}} \langle S_p g, S_p g \rangle \\ &= \left\langle \sum_{p \in \mathcal{O}} S_p g, g \right\rangle \\ &\leq \left\| \sum_{p \in \mathcal{O}} S_p g \right\|_2 \|g\|_2 \\ &\leq B \sqrt{|U|} \|g\|_\infty^2 . \end{aligned}$$

This estimate employs the assumed inequality (6.11), with best constant, together with the fact that  $g$  is supported on  $2U$ .

In the off-diagonal term, we have, recalling the notation  $\mathcal{O}(p)$  of (6.7),

$$(6.14) \quad \begin{aligned} \mathcal{O} &\leq 2 \sum_{p \in \mathcal{O}} \sum_{p' \in \mathcal{O}(p)} |\langle S_p g, S_{p'} g \rangle| \\ &\leq 2 \sum_{p \in \mathcal{O}} |\langle g, \phi_p \rangle| \sum_{p' \in \mathcal{O}(p)} |\langle \phi_p, \phi_{p'} \rangle \langle g, \phi_{p'} \rangle| \end{aligned}$$

Denote the inner sum by  $\mathcal{S}_p$ . Recall the estimate

$$|\langle \phi_{[\omega, I]}, \phi_{[\omega', I']} \rangle| \leq \begin{cases} 0, & \text{if } \omega \cap \omega' = \emptyset, \\ C_\varepsilon \sqrt{\frac{|I'|}{|I|}} \left(1 + \frac{\text{dist}(I, I')}{|I|}\right)^{-20/\varepsilon}, & \text{if } \omega \subset \omega'. \end{cases}$$

We noted this in Section 4, and it is easy to verify. Use it in the estimate of  $\mathcal{S}_p$ .

$$\begin{aligned} \mathcal{S}_p &= \sum_{p' \in \mathcal{O}(p)} |\langle \phi_p, \phi_{p'} \rangle \langle g, \phi_{p'} \rangle| \\ &\leq C \|g\|_\infty \sum_{[\omega', I'] \in \mathcal{O}(p)} \frac{1}{\sqrt{|I'|}} \left(1 + \frac{\text{dist}(I, I')}{|I|}\right)^{-20} |I'|. \end{aligned}$$

At this point, the essential ingredient from the first half of the proof enters in. Namely, using (6.8), we can continue the estimate of  $\mathcal{S}_p$  as follows. The intervals  $\{I' : [\omega', I'] \in \mathcal{O}(p)\}$  are pairwise disjoint and contained in the complement of  $\bar{I} = (b \text{ size}(p))^{-100} I$ . Hence, the sum above can be dominated by

$$\begin{aligned} \mathcal{S}_p &\leq C \|g\|_\infty \int_{\bar{I}^c} \frac{1}{\sqrt{|I|}} \left(1 + \frac{\text{dist}(I, x)}{|I|}\right)^{-20} dx \\ &\leq C (b \text{ size}(p))^{100} \sqrt{|I|} \|g\|_\infty. \end{aligned}$$

Placing this estimate into that for the off-diagonal, (6.14), we get

$$\begin{aligned} \mathcal{O} &\leq C \|g\|_\infty \sum_{p=[\omega, I] \in \mathcal{O}} (b \text{ size}(p))^{100} |\langle g, \phi_p \rangle| \sqrt{|I|} \\ &\leq C \|g\|_\infty^2 b^{400} \sum_{p \in \mathcal{O}} |\langle f, \phi_p \rangle|^2, \end{aligned}$$

with the last line following from iii) at the beginning of this section.

We collect estimates. Namely, the last display together with (6.12) and (6.13), to see that

$$B^2 \leq C \left( B \sqrt{|U|} + b^{400} \sum_{p \in \mathcal{O}} |\langle f, \phi_p \rangle|^2 \right).$$



If the first term is the larger of the two on the right, then we see that the best constant  $B$  is no more than  $C\sqrt{|U|}$ , which is a perfectly adequate estimate.

If the second term is the larger of the two, a contradiction is seen. We can apply the inequality to  $f \mathbf{1}_{2U}$ . In particular,

$$(6.15) \quad \left\| \sum_{p \in \mathcal{O}} \varepsilon_p \phi_p \langle f \mathbf{1}_{2U}, \phi_p \rangle \right\|_2 \leq C \left( b^{400} \sum_{p \in \mathcal{O}} |\langle f, \phi_p \rangle|^2 \right)^{1/2},$$

since  $f$  is bounded by 1. It follows from the removal of the tops  $\mathcal{P}_i^t$ , and in particular (6.5), that

$$(6.16) \quad \begin{aligned} |\langle f, \phi_p \rangle - \langle f \mathbf{1}_{2U}, \phi_p \rangle| &\leq \|f\|_\infty \|\mathbf{1}[(2U)^c] \phi_p(x)\|_1 \\ &\leq (b \text{ size}(p))^{10} \sqrt{|I|} \\ &\leq \frac{1}{100} |\langle f, \phi_p \rangle| \sqrt{|I|}. \end{aligned}$$

Hence, we can average over choices of signs in (6.15), to see that

$$\sum_{p \in \mathcal{O}} |\langle f, \phi_p \rangle|^2 \leq C b^{400} \sum_{p \in \mathcal{O}} |\langle f, \phi_p \rangle|^2,$$

which can only hold if  $b \geq C'$ , an absurdity.

The final touch in the proof our Lemma is short and sweet. We have established (6.11); apply the inequality to  $f$ , averaging over choices of signs. Noting (6.16), we see that

$$\sum_{p \in \mathcal{O}} |\langle f, \phi_p \rangle|^2 \leq C |U|.$$

Yet, in the combinatorial half of the proof, we were careful to preserve the lower half of condition ii) at the beginning of this section, and thus, as the Lemma claims,

$$\sum_i [I_i^t \subset U] |I_i^t| \leq C b^{-2} |U|.$$

**7. Forests.**

In this section, we combine the principal estimates of the previous three sections, and complete the proof of the bound for the bilinear Hilbert transform. We begin with a definition of a collection of pairs, a forest.

Call a set of pairs  $\mathcal{F}$  a *forest* if

$\alpha$ )  $\text{size}(\mathcal{F}) \leq b$ . (See (5.6)).

$\beta$ ) If  $p, p', p'' \in \mathcal{P}$ , and  $p < p', p < p''$  then  $p'$  and  $p''$  are comparable, *e.g.*  $p' < p''$ .

$\gamma$ ) No point  $x$  is in more than  $J = O((A/b)^{210})$  intervals  $I_1, \dots, I_J$ , where the pairs  $[\omega_j, I_j]$  are in  $\mathcal{P}$ , and mutually incomparable under  $<$ .

$\delta$ )  $I_p \subset b^{-40}[-1, 1]$  for all  $p \in \mathcal{F}$ .

The first condition is a natural restraint on the size of the collection of pairs; the middle condition is a critical combinatorial condition imposing a tree-like structure on the forest; and the next to last condition is used to write a forest as a small number of rows. The last assumption will be satisfied by appealing to (5.21).

**Lemma 7.1.** *If  $\mathcal{F}$  is a forest, then there is a set  $E \subset (0, 1)$  of measure at most  $C(b/A)^{80}$  so that for all  $1 < r < 2$ , some  $\delta > 0$ ,*

$$\|T^{\mathcal{P}}g\|_{L^r(E^c)} \leq C_r b^\delta (\log A) \|g\|_2 .$$

Two preparatory Lemmas are in order. First of all, our various estimates break down on small subsets, and the next Lemma justifies the deletion of these bad, thin sets.

**Lemma 7.2.** *Suppose that  $T$  is an operator on a finite measure space  $(X, \mathcal{A}, \mu)$  so that for some  $0 < b < 1$ , and all  $A > 10$  there is a set  $F \subset X$  of measure at most  $C(b/A)^{20}$  so that*

$$\|Tg\|_{L^r(X \setminus F)} \leq C b A \|g\|_{L^2(X)} , \quad r < 2 .$$

*Then,*

$$\|Tg\|_r \leq C \mu(X)^{1/r-1/2} b^{1/2} \|g\|_2 .$$

PROOF. It suffices to assume that  $\mu(X) = 1$ , for the general case follows from this. Let  $g \in L^2(X)$  have norm 1, and  $\lambda > 0$ . We have the estimate

$$\mu(\{Tg > \lambda\}) \leq |F| + \lambda^{-r} \int_{X \setminus F} |Tg|^r du \leq C \left( \left(\frac{b}{A}\right)^{20} + (b A \lambda^{-1})^r \right).$$

Minimizing the estimate over  $A$  will prove the Lemma.

Deleting small subsets of  $(0, 1)$  also requires us to delete sets of pairs which live on these sets. This is the subject of the next Lemma.

**Lemma 7.3.** *Let  $\{I_j : j \geq 1\}$  be a collection of disjoint triadic intervals. Let  $\mathcal{B}$  be a set of pairs with  $\text{size}(p) \leq b$ , and for all  $[\omega, I] \in \mathcal{B}$ ,  $I \subset I_j$  for some  $j$ . Set  $E = \bigcup_j 2I_j$ . Then for a choice of  $\delta = \delta(r) > 0$ ,*

$$\|T^{\mathcal{B}}g\|_{L^r(E^c)} \leq C_r b^\delta \|g\|_2, \quad 1 < r < 2.$$

PROOF. In the proof, we can in addition assume that  $\text{size}(p) \geq b/2$ , for then we can sum the estimate obtained for the sets  $\mathcal{B}_n = \{p \in \mathcal{B} : 2^{-n}b \leq \text{size}(p) \leq 2^{-n+1}b\}$ , for  $n \geq 1$ , to get the Lemma as stated.

Let  $\bar{p}_k = [\bar{\omega}_k, \bar{I}_k]$  be the maximal pairs in  $\mathcal{B}$ . Remove the top from  $\mathcal{B}$ . Namely let  $\mathcal{B}^t$  be those pairs in  $p \in \mathcal{B}$  for which there is no chain

$$p = p_1 \prec p_2 \prec \cdots \prec p_m,$$

with  $m > M = 100 (\log 1/b)$  and  $p_1, p_2, \dots, p_m \in \mathcal{B}$ . Then  $\mathcal{B}^t$  can be written as a union of  $O(\log 1/b)$  sets which are not comparable under  $<$ . Hence, Lemma 4.2 implies

$$\|T^{\mathcal{B}^t}g\|_r \leq C_r b^\delta \|g\|_2, \quad 1 \leq r < 2,$$

which is stronger than our conclusion.

Let  $\mathcal{B}^\sharp = \mathcal{B} \setminus \mathcal{B}^t$ , and set  $\mathcal{P}_k = \{p \in \mathcal{B}^\sharp : p < \bar{p}_k\}$ . This is a tree, with top interval  $\bar{I}_k$  much larger than  $|I|$  for all  $[\omega, I] \in \mathcal{P}_k$ . Thus,  $T^{\mathcal{P}_k}$  will be quite small off of the set  $E = \bigcup_j 2I_j$ . In particular, choose  $j$  so that  $\bar{I}_k \subset I_j$ , which must exist. Then, one easily sees that

$$|T^{\mathcal{P}_k}g(x)| \leq C b^{100} (M \mathbf{1}_{\bar{I}_k}(x))^2 Mg(x), \quad \text{if } x \notin 2I_j.$$

The explicit calculation is much in the spirit of Lemma 5.9. Of course  $\mathcal{B}^\sharp = \bigcup_k \mathcal{P}_k$ , hence for  $r < 2$ , let  $1/r = 1/2 + 1/s$ , and write

$$\begin{aligned} \|T^{\mathcal{B}^\sharp} g\|_{L^r(E^c)} &\leq C b^{100} \left\| Mg \sum_k (M \mathbf{1}_{\bar{I}_k})^2 \right\|_r \\ &\leq C b^{100} \|Mg\|_2 \left\| \sum_k (M \mathbf{1}_{\bar{I}_k})^2 \right\|_s^2 \\ &\leq C b^{100} \|g\|_2 \left\| \sum_k \mathbf{1}_{\bar{I}_k} \right\|_s. \end{aligned}$$

Here the lower bound on the size of pairs enters in. The  $\bar{p}_k$ , being maximal, are incomparable under  $<$ . Hence Lemma 4.1 applies to show that

$$\left\| \sum_k \mathbf{1}_{\bar{I}_k} \right\|_s \leq C_s b^{-3},$$

which will finish the proof of the Lemma.

We turn to the proof of the bound for forests.

PROOF OF LEMMA 7.1. The first task is to rephrase the definition of a forest in terms of trees, which depends critically on the condition  $\beta$ ). Let  $\bar{p}_j = [\bar{\omega}_j, \bar{I}_j]$  be the maximal pairs in  $\mathcal{F}$ . Let  $\mathcal{P}_j = \{p \in \mathcal{F} : p < [\bar{\omega}_j, \bar{I}_j]\}$ . Each  $\mathcal{P}_j$  is a tree and  $\mathcal{F} = \bigcup_j \mathcal{P}_j$ . Moreover, if  $j \neq j'$ , no two pairs  $p \in \mathcal{P}_j$  and  $p' \in \mathcal{P}_{j'}$  are comparable. For if not, assuming  $p < p'$ , then one has  $p < \bar{p}_j$  as well as  $p < \bar{p}_{j'}$ . But these last two pairs being maximal, are incomparable, contradicting  $\beta$ ).

The last condition in the definition of a forest, condition  $\gamma$ ) implies that

$$\sum_j \mathbf{1}_{\bar{I}_j}(x) \leq C \left(\frac{A}{b}\right)^{210}, \quad \text{for all } x.$$

The next steps of the proof are made with the intent of extracting normal separated trees from the  $\mathcal{P}_j$ . The process starts by deleting the top and bottom from  $\mathcal{F}$ . First the top. Let  $\mathcal{F}^t$  be the set of pairs  $p \in \mathcal{F}$  for which there is no strictly ascending chain

$$p = p_1 \prec p_2 \prec \cdots \prec p_m,$$

with  $m > M = 40000 (\log A/b)$  and  $p_1, p_2, \dots, p_m \in \mathcal{F}$ . Then  $\mathcal{F}^t$  can be written as a union of  $M$  sets  $\mathcal{F}_m^t$  for which no two pairs in any  $\mathcal{F}_m^t$  are comparable. Thus, by Lemma 4.2.,

$$(7.4) \quad \|T^{\mathcal{F}^t} g\|_r \leq C b^{\delta(r)} (\log A) \|g\|_2, \quad 1 < r < 2.$$

This is more than what is claimed in the conclusion above.

Let  $\mathcal{F}^0 = \mathcal{F} \setminus \mathcal{F}^t$ , and now remove the bottom of  $\mathcal{F}^0$ . Let  $\mathcal{F}^b$  be the set of  $p \in \mathcal{F}^0$  for which there is no descending chain

$$p_1 \prec p_2 \prec \cdots \prec p_m = p,$$

with  $m > M = 40000 \log(A/b)$  and  $p_1, p_2, \dots, p_m \in \mathcal{F}^0$ . As before,

$$\|T^{\mathcal{F}^b} g\|_r \leq C b^{\delta(r)} (\log A) \|g\|_r, \quad 1 < r < 2.$$

Let  $\mathcal{F}^1 = \mathcal{F} \setminus (\mathcal{F}^t \cup \mathcal{F}^b)$  and  $\mathcal{P}_j^1 = \mathcal{F}^1 \cap \mathcal{P}_j$ .

The exceptional set enters in. Two sets are defined below to conform with the formulation of Lemma 7.3. Set

$$E_i = \bigcup_j \left\{ x : \text{dist}(x, \partial \bar{I}_j) \leq 2i \left(\frac{b}{A}\right)^{400} |\bar{I}_j| \right\} \quad \text{for } i = 1, 2.$$

The set  $E = E_2$  is the exceptional set of our Lemma. By part  $\gamma$ ) and  $\delta$ ) of the definition of a forest,  $|E| \leq C (b/A)^{90}$ . Next, we delete some pairs. Let  $\mathcal{B}$  denote those pairs in  $\mathcal{F}$  for which  $I \subset E_1$ . It follows from Lemma 7.3 that

$$\|T^{\mathcal{B}} g\|_{L^r(E^c)} \leq C_r b^\delta \|g\|_2, \quad 1 < r < 2, \delta > 0.$$

The exceptional set will play no other role in the proof.

We have identified all the pairs to remove from  $\mathcal{F}$ . Set  $\mathcal{F}^\sharp = \mathcal{F} \setminus (\mathcal{F}^t \cup \mathcal{F}^b \cup \mathcal{B})$ , and  $\mathcal{P}_j^\sharp = \mathcal{P}_j \cap \mathcal{F}^\sharp$ . The  $\mathcal{P}_j^\sharp$  are much nicer trees. In particular,

- i)  $\mathcal{P}_j^\sharp$  is a normal tree.
- ii) If  $j \neq j'$  then  $\mathcal{P}_j^\sharp$  and  $\mathcal{P}_{j'}^\sharp$  are separated trees.

Here is the verification of the first claim. There are two conditions to check. The second, that  $\text{dist}(I, \partial \bar{I}_j) > (b/A)^{400} |\bar{I}_j|$  follows from the removal of the set  $\mathcal{B}$ . Finally, by the removal of the top, for each  $[\omega, I] \in \mathcal{P}_j^\sharp$ , there is a chain of triadic intervals

$$I = I_1 \subsetneq I_2 \subsetneq \cdots \subsetneq I_M = \bar{I}_j, \quad M > 40000 \log \frac{A}{b}.$$

And so  $|I| \leq (b/A)^{40000} |\bar{I}_j|$ . This shows normality.

The second claim will follow from the removal of the bottom: for  $j \neq j'$ , recall that the top of  $\mathcal{P}_j^\sharp$  is  $[\bar{\omega}_j, \bar{I}_j]$ , and assume that  $\bar{I}_j \cap \bar{I}_{j'} \neq \emptyset$ . Let's check the condition  $\alpha$ ) in the definition of separated. For  $p = [\omega, I] \in \mathcal{P}_j^\sharp$  with  $I \subset \bar{I}_{j'}$ , as  $p \notin \mathcal{P}^b$ , there is a descending chain  $[\omega_1, I_1] = p_1 \precneq p_2 \precneq \dots \precneq p_M = p$  where  $M = 3000 \log A/b$ , and all of the  $p_m \in \mathcal{F}^0$ . The situation is that  $I_1 \subset I \subset \bar{I}_{j'}$ , and  $p_1 \in \mathcal{P}_j$ . But  $p_1$  and  $[\bar{\omega}_{j'}, \bar{I}_{j'}]$  cannot be comparable, by the condition  $\beta$ ) in the definition of forest. That is,  $\omega_1 \cap \bar{\omega}_{j'} = \emptyset$ . But all of the  $\omega_n$ 's are central, so by the convenient property of central dyadic intervals, (3.1),  $\omega$  sits well inside  $\omega_1$ . In particular,  $\text{dist}(\omega, \omega_{j'}^0) \geq \text{dist}(\omega, \partial\omega_1) \geq 3^M |\omega| \geq (A/b)^{3000} |\omega|$ . This verifies  $\alpha$ ) in the definition of separated, with the proof of  $\beta$ ) following by symmetry. The last condition in the definition of separated follows from the removal of the pairs in  $\mathcal{B}$ . This finishes ii) above.

The next task is to show that  $\mathcal{F}^\sharp$  can be written as a union of at most  $J = (A/b)^{210}$  separated rows  $\mathcal{R}_1, \dots, \mathcal{R}_J$ . To see this, let  $\{I^i\}$  be the maximal intervals from the  $\{\bar{I}_j\}$  with no repetitions. For each  $I^i$ , let  $[\bar{\omega}_{j(i)}, I^i]$  be one of the maximal pairs  $\bar{p}_{j(i)}$ . Then  $\mathcal{R} = \bigcup_i \mathcal{P}_{j(i)}^\sharp$ . Delete the maximal pairs  $[\bar{\omega}_{j(i)}, I^i]$  from the list of all maximal pairs, and repeat the procedure above. Condition  $\gamma$ ) guarantees that the procedure stops in at most  $(A/b)^{210}$  steps. Separability of the rows  $\mathcal{R}_j$  follows from the construction.

With this decomposition Lemma 5.25 concludes the proof is concluded by appealing to Lemma 5.25, Lemma 7.2 and part  $\delta$ ) of the definition of a forest.

A critical combinatorial trick will permit us to write much larger sets of pairs as a union of a small number of forests.

**Lemma 7.5.** *Let  $\mathcal{P}$  be a set pairs with  $\text{size}(\mathcal{P}) \leq b$ . Futher, letting  $\bar{p}_k = [\bar{\omega}_k, \bar{I}_k]$  denote the maximal pairs in  $\mathcal{P}$ , assume that they obey a Carleson measure estimate*

$$(7.6) \quad \sum_k [\bar{I}_k \subset U] |\bar{I}_k| \leq C b^{-140} |U|.$$

for all intervals  $U$ . In addition assume that  $I_k \subset b^{-40} [-1, 1]$  for all  $k$ . Then, for some  $\delta > 0$ ,

$$\|T^{\mathcal{P}} g\|_1 \leq C b^\delta \|g\|_2.$$

PROOF. The Carleson measure estimate implies that

$$\left\| \sum_k \mathbf{1}_{\bar{I}_k} \right\|_r \leq C_r b^{-180}, \quad 1 < r < \infty.$$

Hence, the two sets

$$E_i = \left\{ x : \sum_k \mathbf{1}[i\bar{I}_k](x) \geq \left(\frac{A}{b}\right)^{210} \right\}, \quad i = 1, 2,$$

have very small measure:  $|E_2| \leq C(b/A)^{100}$ . Use this set to delete some pairs. Set  $\mathcal{B} = \{[\omega, I] \in \mathcal{P} : I \subset E_1\}$ . By Lemma 7.3,

$$\|T^{\mathcal{B}}\|_{L^r(E_2^c)} \leq C_r b^\delta \|g\|_2, \quad 1 < r < 2.$$

For the set  $\mathcal{P}^\sharp = \mathcal{P} \setminus \mathcal{B}$ , we will show that there is a set  $F \subset (0, 1)$  of measure at most  $(b/A)^{20}$ , so that

$$\|T^{\mathcal{P}^\sharp}\|_{L^r(F^c)} \leq C_r b^\delta (\log A) \|g\|_2, \quad 1 < r < 2.$$

The estimate of this Lemma will then follow from Lemma 7.2.

We shall see that  $\mathcal{P}^\sharp$  can be decomposed into  $O(\log A/b)$  forests. Therefore, the estimate above follows from Lemma 7.1. The decomposition is accomplished by means of the following combinatorial trick, which has already been used in Section 4.

Let  $B(p) =$  the number of  $k$  for which  $p < \bar{p}_k$ . Simply define

$$\mathcal{F}_m = \{p \in \mathcal{P} : 2^{m-1} \leq B(p) < 2^m\}.$$

By construction, this set is empty if  $m > O(\log A/b)$ .

It remains to verify that each  $\mathcal{F}_m$  is a forest. The first condition in the definition is trivial, the next to last condition follows from the deletion of the set of pairs  $\mathcal{B}$  and the last condition follows from the hypothesis of the Lemma. The middle condition  $\beta)$  in the definition of a forest must be checked. But it is a consequence of the following combinatorial property of  $B(p)$ : for  $p, p', p'' \in \mathcal{P}$ , with  $p < p', p < p''$  but  $p'$  and  $p''$  not comparable implies that  $B(p) \geq B(p') + B(p'')$ . So if in addition  $p', p'' \in \mathcal{F}_m$ , then  $p \notin \mathcal{F}_m$ , proving that  $\mathcal{F}_m$  is a forest. To see the super-additive property, write  $p' < \bar{p}_{j(1)}, \bar{p}_{j(2)}, \dots, \bar{p}_{j(s)}$  where  $B(p') = s$ , and  $p'' < \bar{p}_{k(1)}, \bar{p}_{k(2)}, \dots, \bar{p}_{k(t)}$  where  $B(p'') = t$ . Now, if some  $\bar{p}_{j(u)}$  equals some  $\bar{p}_{k(v)}$ , the situation would be  $p < p', p'' <$

$p''' = \bar{p}_{j,u} = \bar{p}_{k,v}$ . But it is a simple property of triadic intervals that the last condition forces  $p'$  and  $p''$  to be comparable, a contradiction. Thus,  $p < p' < \bar{p}_{j(1)}, \bar{p}_{j(2)}, \dots, \bar{p}_{j(s)}, \bar{p}_{k(1)}, \bar{p}_{k(2)}, \dots, \bar{p}_{k(t)}$  all pairs being distinct, which means that  $B(p) \geq s + t$ .

The previous Lemma, with its reliance on the Carleson measure estimate, clearly implies the following two corollaries, which are stated for specificity. For orchards, the necessary Carleson measure estimate is Lemma 6.1. (Recall that this Lemma applies only for sufficiently small  $b$ ).

**Corollary 7.7.** *Let  $\mathcal{O}$  be an  $\alpha$  or an  $\alpha^*$ -orchard, with  $\text{size}(\mathcal{O}) \leq b$ , where  $0 < b < b_0$ . Assume that  $I_p \subset b^{-40}[-1, 1]$  for all  $p \in \mathcal{O}$ . Then, for some  $\delta > 0$ ,*

$$\|T^{\mathcal{O}}g\|_1 \leq C b^\delta \|g\|_2 .$$

And, assuming a lower bound on the size of pairs, the necessary Carleson measure estimate is Lemma 4.1.

**Corollary 7.8.** *Let  $\mathcal{P}$  be a set of pairs with  $\text{size}(\mathcal{P}) \leq b$ , and with  $\text{size}(p) \geq b^{30}$  for all  $p \in \mathcal{P}$ . Then,*

$$\|T^{\mathcal{P}}g\|_1 \leq C b^\delta \|g\|_2 .$$

The last Lemma describes the inductive procedure with which the set of all pairs can be broken up into sets to which the previous two corollaries can be applied.

**Lemma 7.9.** *Let  $0 < b < b_0$ . Let  $\mathcal{P}$  be a set of pairs satisfying*

$$\alpha) \text{ size}(p) \leq b^{30} \text{ for all } p \in \mathcal{P} .$$

$$\beta) \text{ size}(\mathcal{P}) \leq b .$$

*Then  $\mathcal{P} = \mathcal{P}^\# \cup \mathcal{P}^b$  where, for some  $\delta > 0$ ,*

$$\|T^{\mathcal{P}^b}g\|_1 \leq C b^\delta \|g\|_2 ,$$

*and  $\mathcal{P}^\#$  satisfies  $\alpha)$  and  $\beta)$  above, with  $b$  replaced by  $b/2$ .*



PROOF. Consider those pairs  $\bar{p} = [\bar{\omega}, \bar{I}]$  in  $\mathcal{P}$  for which there is an  $\alpha$ -tree  $\mathcal{T}$  with top  $\bar{p}$  so that

$$(7.10) \quad \sum_{p \in \mathcal{T}} |\langle f, \phi_p \rangle|^2 \geq c_1 b |\bar{I}|.$$

Here,  $c_1$  is constant which appears in the definition of an orchard. However, it follows from the hypothesis  $\beta$ ) that the sum above cannot be more than an absolute constant times  $b$ . In addition,  $\bar{I} \subset b^{-1}[-1, 1]$ , as follows from (5.21).

Let  $\bar{p}_1 = [\bar{\omega}_1, \bar{I}_1]$  be such a pair, so that the interval  $\bar{I}_1$  is maximal among all such pairs. Take  $\mathcal{T}_1^\alpha$  to be all pairs  $p = [\omega, I] \in \mathcal{P}$  so that  $A(\omega)$  contains the center of  $\bar{\omega}_1$ , and  $I \subset \bar{I}_1$ . That is,  $\mathcal{T}_1^\alpha$  is the largest  $\alpha$ -tree in  $\mathcal{P}$  with top  $\bar{p}_1$ . Repeat this procedure to the collection  $\mathcal{P} \setminus \mathcal{T}_1^\alpha$  to get an  $\alpha$ -tree  $\mathcal{T}_2^\alpha$  with top  $\bar{p}_2$ . Continue this procedure indefinitely, thereby obtaining a sequence of  $\alpha$ -trees  $\mathcal{T}_j^\alpha$  with tops  $\bar{p}_j$ .

We claim that  $\mathcal{O}^\alpha = \bigcup \mathcal{T}_j^\alpha$  is an  $\alpha$ -orchard. There are three conditions to check, yet each of these follows immediately from the construction. Clearly, conditions ii) and iii) hold. And condition i) follows from maximality of the  $\bar{I}_j$ . Therefore, Corollary 7.7 applies, showing that

$$\|T^{\mathcal{O}^\alpha} g\|_1 \leq C b^\delta \|g\|_2, \quad \delta > 0.$$

Remove the pairs  $\mathcal{T}^\alpha$  from  $\mathcal{P}$ , and call the resulting set  $\mathcal{P}^1$ . By our choice of the constant  $c_1$  in (7.10), which was made in the definition of an orchard, we see that for any  $\alpha$ -tree  $\mathcal{T}$  in  $\mathcal{P}^1$ ,

$$\|T^\mathcal{T}\|_2 \leq \frac{b}{6}.$$

Continue by removing  $\alpha^*$ -trees from  $\mathcal{P}^1$  in exactly the same manner as the  $\alpha$ -trees were removed. We get a set  $\mathcal{T}^{\alpha^*} \subset \mathcal{P}^1$  with

$$\|T^{\mathcal{T}^{\alpha^*}} g\|_1 \leq C b^\delta \|g\|_2, \quad \delta > 0.$$

Let  $\mathcal{P}^2$  be the collection of pairs obtained by removing those pairs in  $\mathcal{T}^{\alpha^*}$  from  $\mathcal{P}^1$ . It follows that any  $\alpha^*$ -tree in  $\mathcal{P}^2$  has norm at most  $b/6$ . Now, since  $\text{size}(p) < b^{30}$ , for all  $p$ , we see that any  $\beta$ -tree in  $\mathcal{P}^2$  has very small norm. Hence, for any tree  $\mathcal{T}$  in  $\mathcal{P}^2$ ,

$$\|T^\mathcal{T}\|_2 \leq \frac{b}{2},$$

which is seen by writing  $\mathcal{T}$  as a union of an  $\alpha$ -tree, a  $\alpha^*$ -tree and a  $\beta$ -tree. That is,  $\text{size}(\mathcal{P}^2) \leq b/2$ , and the set satisfies the condition  $\beta$  of the Lemma with  $b$  replaced by  $b/2$ .

We turn our attention to the condition  $\alpha$ ). Set

$$\mathcal{P}^b = \left\{ p \in \mathcal{P}^2 : \left(\frac{b}{2}\right)^{30} \leq \text{size}(p) \leq b^{30} \right\},$$

and

$$\mathcal{P}^\# = \left\{ p \in \mathcal{P}^2 : \text{size}(p) \leq \left(\frac{b}{2}\right)^{30} \right\}.$$

The second collection satisfies  $\alpha$ ) and  $\beta$ ) with  $b$  replaced by  $b/2$ . And it remains to see that the first collection of pairs leads to an operator with small norm. Yet, with the lower bound on the size of pairs, we are in a position to apply Corollary 7.8 to  $\mathcal{P}^b$ , and so the proof of the Lemma is complete.

A brief argument will finish the proof of the boundedness of the bilinear Hilbert transform. Recall from Section 3 that set  $\mathcal{P}$  to be the set of all (admissible) pairs, we need only prove

$$(7.11) \quad \|T^{\mathcal{P}}g\|_1 \leq C \|g\|_2$$

Let  $\mathcal{P}^b = \{p : \text{size}(p) \geq b_0^{30}\}$ , and  $\mathcal{P}^\#$  be the complementary set of pairs. Now,  $\text{size}(\mathcal{P}^b) \leq C$ , and so with the lower bound on the size of pairs, we can apply Corollary 7.8 to see that

$$\|T^{\mathcal{P}^b}g\|_1 \leq C_0 \|g\|_2.$$

Iteratively applying Lemma 7.9 to  $\mathcal{P}^\#$ , we can write this collection of pairs as a union of collections  $\mathcal{P}_n$ , with

$$\|T^{\mathcal{P}_n}g\|_1 \leq C 2^{-\delta n} \|g\|_2.$$

This estimate is summable in  $n$ , as  $\delta > 0$ , and so it proves (7.11), finishing the proof of the boundedness of the bilinear Hilbert transform.

## References.

- [1] Carleson, L., On the convergence and growth of partial sums of Fourier series. *Acta Math.* **116** (1966), 135-157.

- [2] Fefferman, C., Pointwise convergence of Fourier series. *Ann. Math.* **98** (1973), 551-571.
- [3] Daubechies. I., *Ten Lectures on Wavelets*. CBMS-NSF regional conference series in applied mathematics. SIAM (1992), Philadelphia.
- [4] Lacey, M., Thiele, C.,  $L^p$  bounds for the bilinear Hilbert transform. *Proc. Nat. Acad. Sci. USA* **94** (1997), 33-35.

*Recibido:* 29 de octubre de 1.996

Michael T. Lacey\*  
School of Mathematics  
Georgia Institute of Technology  
Atlanta, GA 30332  
U.S.A.  
lacey@math.gatech.edu

---

\* Research supported in part by an NSF Postdoctoral Fellowship, and NSF grants.