

A GUIDE TO CARLESON'S THEOREM

CIPRIAN DEMETER

ABSTRACT. This paper is meant to be a gentle introduction to Carleson's theorem on pointwise convergence of Fourier series.

1. Introduction. Let

$$S_n f(x) = \sum_{k=-n}^n \widehat{f}(k) e^{2\pi i k x},$$

be the partial Fourier series of the L^1 function f on $[0, 1]$. In 1966, Carleson proved the following long standing conjecture.

Theorem 1.1 [5]. *For each $f \in L^2([0, 1])$, the Fourier series $S_n f$ converges almost everywhere to f .*

Soon after that, a slight modification of Carleson's method allowed Hunt [11] to extend the result to L^p functions, for $p > 1$.

Theorem 1.1 has since received many proofs, most notably by Fefferman [10] and by Lacey and Thiele [15]. The impact of Carleson's theorem has increased in recent years thanks to its connections with scattering theory [19], ergodic theory [8, 9], the theory of directional singular integrals in the plane [1, 2, 6, 7, 13, 14] and the theory of operators with quadratic modulations [17, 18]. A more detailed description can be found in [12]. These connections have motivated the discovery of various new arguments for Theorem 1.1. While these arguments share some similarities, each of them has a distinct personality. Along these lines, it is interesting to note that, for almost every specific application of Carleson's theorem in the aforementioned fields, only one of the arguments will do the job.

2010 AMS *Mathematics subject classification.* Primary 42A20, Secondary 42A45.

Keywords and phrases. Fourier series, singular integrals, Carleson's theorem.
Received by the editors on December 21, 2012.

DOI:10.1216/RMJ-2015-45-1-169 Copyright ©2015 Rocky Mountain Mathematics Consortium

All the arguments for Theorem 1.1 are technical. To present the main ideas in a transparent way, we will instead analyze the closely related Walsh-Fourier series, which we recall below.

For $n \geq 0$ the n th Walsh function w_n is defined recursively by the formula

$$\begin{aligned} w_0 &= 1_{[0,1)} \\ w_{2n} &= w_n(2x) + w_n(2x - 1) \\ w_{2n+1} &= w_n(2x) - w_n(2x - 1). \end{aligned}$$

Given $f : [0, 1] \rightarrow \mathbf{C}$ we recall the partial Walsh-Fourier series of f

$$S_n^W f(x) = \sum_{k=0}^n \langle f, w_k \rangle w_k(x).$$

The following theorem was proved by Billard, by adapting Carleson's methods.

Theorem 1.2 [4]. *For each $1 < p \leq \infty$ and each $f \in L^p([0, 1])$, the series $S_n^W f(x)$ converges almost everywhere to $f(x)$.*

We present a few proofs of Theorem 1.2 which are translations of their Fourier analogues. In each case the translation can be done in more than one way; the proofs presented reflect the author's taste. While a few of the features of the original proofs from the Fourier case will be lost in translation, the main line of thought will be essentially preserved intact in the Walsh case. Very little originality is claimed by the author.

This paper is by no means a complete guide to Carleson's theorem; in particular, we shall make no attempt to describe in detail any of its aforementioned applications. The main goal is to give a self contained but concise survey of some of the main arguments in the literature.

2. The Walsh phase plane. It turns out that there is a multiscale description for $S_n^W f$. Let \mathcal{D}_+ denote the collection of all dyadic intervals which are subsets of $\mathbf{R}_+ = [0, \infty)$. We call $\mathbf{R}_+ \times \mathbf{R}_+$ the Walsh phase plane.

Definition 2.1. A tile $p = I_p \times \omega_p$ is a rectangle of area one, such that $I_p, \omega_p \in \mathcal{D}_+$. A bitile $P = I_P \times \omega_P$ is a rectangle of area two, such that $I_P, \omega_P \in \mathcal{D}_+$. Let $\omega_{P_l}, \omega_{P_u}$ be the left (or lower) and right (or upper) halves of ω_P . We will denote by $P_l = I_P \times \omega_{P_l}$ and $P_u = I_P \times \omega_{P_u}$ the lower and upper tiles of P . We denote by \mathbf{P}_{all} the collection of all bitiles.

Given a tile $p = [2^j m, 2^j(m + 1)] \times [2^{-j} n, 2^{-j}(n + 1)]$ we define the associated Walsh wave packet

$$W_p(x) = 2^{-j/2} w_n(2^{-j} x - m).$$

To understand the relevance of the Walsh phase plane, we recall a few tools from [22], see also [21].

Every $x \in \mathbf{R}_+ = [0, \infty)$ can be identified uniquely with a doubly-infinite set of binary digits $a_n = a_n(x)$ such that

$$x = \sum_{n \in \mathbf{Z}} a_n 2^n$$

where $a_n \in \{0, 1\}$ and $\liminf_{n \rightarrow -\infty} a_n = 0$. Note that a_n is eventually zero as $n \rightarrow \infty$. We define two operations on \mathbf{R}_+ . First,

$$x \oplus y := \sum_n b_n 2^n$$

where

$$b_n := a_n(x) + a_n(y) \text{ mod } 2.$$

We caution the reader that b_n is not always the same as $a_n(x \oplus y)$, as the example $x = \sum_{n < 0: n \text{ odd}} 2^n$, $y = \sum_{n < 0: n \text{ even}} 2^n$ shows. Also, since $1 = x \oplus y = x \oplus z$ where $z = 1 + \sum_{n < 0: n \text{ odd}} 2^n$, (\mathbf{R}_+, \oplus) is not, strictly speaking a group. We will be content with observing that for all practical purposes (\mathbf{R}_+, \oplus) can be thought of as being a group, in the sense that \oplus behaves like a genuine group operation if we exclude pairs x, y of zero product Lebesgue measure.

Define the second operation by

$$x \otimes y := \sum_n c_n 2^n$$

where

$$c_n := \sum_{m \in \mathbf{Z}} a_m(x) a_{n-m}(y) \bmod 2.$$

We note that this sum is always finite. If we neglect zero measure sets, $(\mathbf{R}_+, \otimes, \oplus)$ can be thought of as being a field with characteristic two. It will be implicitly assumed that various equalities to follow hold outside zero measure sets.

Define the function $e_W : \mathbf{R}_+ \rightarrow \{-1, 1\}$ such that $e_W(x) = 1$ when $a_{-1}(x) = 0$ and $e_W(x) = -1$ when $a_{-1}(x) = 1$. This 1-periodic function is the Walsh analogue of $e^{2\pi i x}$. It is easy to check that

$$w_n(x) = e_W(x \otimes n) 1_{[0,1]}(x);$$

thus, w_n can be thought of as being the Walsh analogue of $e^{2\pi i n x}$. Also, for each tile $p = I_p \times \omega_p$, we have

$$W_p(x) = \frac{1}{|I_p|^{1/2}} w_0 \left(\frac{x - l(I_p)}{|I_p|} \right) e_W(x \otimes l(\omega_p)),$$

where $l(J)$ denotes the left endpoint of J . A simple computation shows that for each bitile P

$$(1) \quad W_{P_l}(x) = W_{P_u}(x) e_{I_P}(x)$$

where $e_I(x) = 1$ on I_l and $e_I(x) = -1$ on I_r .

The collection of all wave packets W_p where p ranges over all tiles with fixed scale forms a complete orthonormal system in $L^2(\mathbf{R}_+)$, and thus

$$f = \sum_{p: |I_p|=2^j} \langle f, W_p \rangle W_p.$$

We introduce the *Walsh* (also called *Walsh-Fourier*) transform of a function $f : \mathbf{R}_+ \rightarrow \mathbf{C}$ to be

$$\mathcal{F}_W f(\xi) = \widehat{f}(\xi) := \int e(x \otimes \xi) f(x) dx.$$

It is easy to see that its inverse \mathcal{F}_W^{-1} coincides with \mathcal{F}_W .

Arguably, the most important feature that makes the Walsh phase plane technically simpler than its Fourier counterpart is the absence of the strong form of the ‘‘Uncertainty Principle.’’ This allows the existence of functions compactly supported in both time and frequency. The best example is $1_{[0,1]}$, which equals its Walsh transform. A quick computation shows that, for each interval $I \in \mathcal{D}_+$

$$(2) \quad \widehat{1}_I(\xi) = |I|1_{[0,|I|^{-1}]}(\xi)e(\xi \otimes x_I)$$

where x_I is an arbitrary element of I . Similarly,

$$\widehat{W}_p(\xi) = \frac{1}{|\omega_p|^{1/2}}w_0\left(\frac{\xi - l(\omega_p)}{|\omega_p|}\right)e_W(\xi \otimes l(I_p)).$$

Thus, W_p is spatially supported in I_p , while its Walsh transform is supported in ω_p . An application of Plancherel’s theorem shows that

$$(3) \quad \langle W_p, W_{p'} \rangle = 0$$

whenever the tiles p and p' do not intersect.

The following partial relation of order was introduced by Fefferman [10].

Definition 2.2 (Order). For two tiles or bitiles P, P' , we write $P \leq P'$ if $I_P \subset I_{P'}$ and $\omega_{P'} \subset \omega_P$.

Note that P and P' are comparable under \leq if and only if they intersect as sets. We will refer to maximal (or minimal) tiles (or bitiles) with respect to \leq as simply being maximal (or minimal).

Definition 2.3 (Convexity). A collection \mathbf{P} of bitiles is called convex if whenever $P, P'' \in \mathbf{P}$, $P' \in \mathbf{P}_{\text{all}}$ and $P \leq P' \leq P''$, we must also have $P' \in \mathbf{P}$.

For a collection \mathbf{p} of tiles or bitiles we denote by $A(\mathbf{p}) = \bigcup_{p \in \mathbf{p}} I_p \times \omega_p$ the region in \mathbf{R}_+^2 covered by them.

We will use the fact (see [22, Lemma 2.5]) that, for each convex set of bitiles \mathbf{P} , the region $A(\mathbf{P})$ can be written (not necessarily in a unique way) as a disjoint union of tiles \mathbf{p}

$$(4) \quad A(\mathbf{P}) = A(\mathbf{p}).$$

We can identify any region in the Walsh phase plane which is a finite union of pairwise disjoint tiles $p \in \mathbf{p}$ with the subspace of $L^2(\mathbf{R}_+)$ spanned by $(W_p)_{p \in \mathbf{p}}$. Indeed, it turns out that if two such collections \mathbf{p} and \mathbf{p}' of tiles cover the same area in the phase plane, then $(W_p)_{p \in \mathbf{p}}$ and $(W_p)_{p \in \mathbf{p}'}$ span the same vector space in $L^2(\mathbf{R}_+)$, see [22, Corollary 2.7]. In particular,

$$(5) \quad \sum_{p \in \mathbf{p}} \langle f, W_p \rangle W_p(x) = \sum_{p \in \mathbf{p}'} \langle f, W_p \rangle W_p(x).$$

The projection operator onto this subspace

$$\Pi_{\mathbf{p}} f(x) = \sum_{p \in \mathbf{p}} \langle f, W_p \rangle W_p(x)$$

will be referred to as the *phase space projection* onto \mathbf{p} . Equation (3) guarantees that $(W_p)_{p \in \mathbf{p}}$ forms an orthonormal basis of the range of $\Pi_{\mathbf{p}}$. In particular, if \mathbf{P} is a convex union of bitiles and \mathbf{p} satisfies (4), then we abuse notation and define

$$\Pi_{\mathbf{P}} f := \Pi_{\mathbf{p}} f.$$

An easy induction argument proves

Lemma 2.4 [22, Corollary 2.4]. *Let \mathbf{p}, \mathbf{p}' be a finite collection of pairwise disjoint tiles such that $A(\mathbf{p}') \subset A(\mathbf{p})$. Then there exists a collection \mathbf{p}'' of pairwise disjoint tiles which includes all the tiles in \mathbf{p}' such that $A(\mathbf{p}) = A(\mathbf{p}'')$.*

Lemma 2.4 and (3) will imply that, for each convex collection \mathbf{P} of bitiles and for each $P \in \mathbf{P}$,

$$(6) \quad \langle f, W_p \rangle = \langle \Pi_{\mathbf{P}} f, W_p \rangle, \quad p \in \{P_u, P_l\}.$$

Remark 2.5. We mention that the relation of order as well as concepts such as convexity and phase space projections can be extended naturally to the two- (or higher) dimensional case. This will be explored in Section 8.

Fix now $n \geq 0$. Note that $S_n^W f(x) = \Pi_{\mathbf{p}_n} f(x)$, where \mathbf{p}_n is the collection of the tiles $[0, 1] \times [k, k + 1]$, $0 \leq k \leq n$. We will partition $A(\mathbf{p}_n) = [0, 1] \times [0, n + 1]$ in a different way. Namely, for each point $(x, \xi) \in [0, 1] \times [0, n + 1]$, there exists a unique bitile P such that $(x, \xi) \in P_l$ and $(x, n + 1) \in P_u$. This bitile is precisely the minimal one such that $n + 1, \xi \in \omega_P$ and $x \in I_P$. Note that the tiles P_l corresponding to all these P will partition $[0, 1] \times [0, n + 1]$. But then (5) will imply that

$$S_n^W f(x) = \sum_{P \in \mathbf{P}_{\text{all}}: n+1 \in \omega_{P_u}} \langle f, W_{P_l} \rangle W_{P_l}(x);$$

in particular,

$$\sup_{n \geq 0} |S_n^W f(x)| = \left| \sum_{P \in \mathbf{P}_{\text{all}}} \langle f, W_{P_l} \rangle W_{P_l}(x) 1_{\omega_{P_u}}(N(x)) \right|$$

for a suitable function $N : \mathbf{R}_+ \rightarrow \mathbf{N}$. The roles of P_u and P_l can be interchanged, without altering the nature or the difficulty. For pedagogical reasons, we choose to work with the model sums

$$C_{\mathbf{P}} f(x) = \sum_{P \in \mathbf{P}} \langle f, W_{P_u} \rangle W_{P_u}(x) 1_{\omega_{P_l}}(N(x)),$$

where $\mathbf{P} \subset \mathbf{P}_{\text{all}}$ and $f : \mathbf{R}_+ \rightarrow \mathbf{C}$. Using the standard approximation argument combined with the almost everywhere convergence of $S_n^W f(x)$ for characteristic functions of intervals, Theorem 1.2 will follow from the following inequality.

Theorem 2.6. *We have, for each $1 < p < \infty$, $N : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ and $f \in L^p(\mathbf{R}_+)$,*

$$\|C_{\mathbf{P}_{\text{all}}} f\|_p \leq C_p \|f\|_p.$$

The constant C_p does not depend on f and N .

Note that, to recover Theorem 1.2, we could restrict attention to functions f on $[0, 1]$ and to the bitiles spatially supported in $[0, 1]$. We will do so in some, but not all, the proofs to follow. We will always allow the choice function N to take any value in \mathbf{R}_+ , not just integers.

3. Estimates for a single tree. Throughout the paper, we will denote by Mf the Hardy-Littlewood maximal function of f . The

various implicit constants hidden in the notation \lesssim will typically be allowed to depend on the Hölder exponents p, p_i, s_i , etc.

Fefferman [10] organized the bitiles in structures that he named *trees* and *forests*. The restriction C_T to a tree will be a Calderón-Zygmund object which can be investigated with classical methods. The contribution of forests is controlled by using various forms of orthogonality between the tree operators C_T . All of the approaches described in the following sections will rely on this strategy.

Definition 3.1. Let $I_T \in \mathcal{D}_+$ and $\xi_T \in \mathbf{R}_+ \setminus \{n2^{-k} : n, k \in \mathbf{Z}\}$. A tree T with top data (I_T, ξ_T) is a collection of bitiles such that $I_P \subset I_T$ and $\xi_T \in \omega_P$ for each $P \in T$. If $P_T \in \mathbf{P}_{\text{all}}$ is such that $P \leq P_T$ for each $P \in T$, we call P_T a top bitile for T . Note that such a P_T is not unique and T need not contain a top bitile.

A tree is called *overlapping* if the tiles $\{P_u : P \in T\}$ intersect. A tree is called *lacunary* if the tiles $\{P_l : P \in T\}$ intersect.

Each tree can be decomposed as $T = T_l \cup T_0$ where

$$\begin{aligned} T_l &= \{P \in T : \xi_T \in \omega_{P_l}\} \\ T_0 &= \{P \in T : \xi_T \subset \omega_{P_u}\}. \end{aligned}$$

Note that T_l is lacunary while T_0 is overlapping. Moreover, if T is convex then so are the trees T_l and T_0 . These observations will allow us to always assume the tree we deal with is either overlapping or lacunary.

The classical example of lacunary tree is the *Littlewood-Paley* tree consisting of all bitiles of the form $P^I := I \times [0, |I|^{-1}]$, $I \in \mathcal{D}_+$. Note that, for each P^I in the Littlewood-Paley tree, we have $W_{P^I}(x) = h_I(x)$, where h_I is the L^2 normalized Haar function equal to $|I|^{-1/2}$ on the left half I_l and to $-|I|^{-1/2}$ on the right half I_u . Recall that, if I is a subset of the Littlewood-Paley tree, we have

$$(7) \quad \left\| \sum_{P^I \in I} \langle f, h_I \rangle h_I \right\|_p \lesssim \|f\|_p, \quad 1 < p < \infty.$$

For a lacunary tree T , we denote

$$O_T f(x) = \sum_{P \in T} \langle f, W_{P_u} \rangle W_{P_u}(x).$$

We have the following generalization of (7).

Lemma 3.2 (Single tree estimate: singular integral). *Let T be a lacunary tree. Then, for each $1 < p < \infty$*

$$\|O_T f\|_p \lesssim \|f\|_p.$$

Proof. Recall that $\widehat{W_{P_u}}$ is supported in ω_{P_u} . Call Ω the collection of all intervals ω_{P_u} , $P \in T$ and note that they are pairwise disjoint and sit within a distance smaller than their length from ξ_T . By the Walsh version of the Littlewood-Paley theorem applied to $O_T f$ and then to f we have

$$\begin{aligned} \|O_T f\|_p &\lesssim \left\| \left(\sum_{\omega \in \Omega} |\mathcal{F}_W^{-1}(1_\omega \mathcal{F}_W(O_T f))|^2 \right)^{1/2} \right\|_p \\ &= \left\| \left(\sum_{\omega \in \Omega} \left| \sum_{P \in T: |\omega_{P_u}|=|\omega|} \langle f, W_{P_u} \rangle W_{P_u} \right|^2 \right)^{1/2} \right\|_p \\ &= \left\| \left(\sum_{P \in T} \frac{|\langle f, W_{P_u} \rangle|^2}{|I_P|} 1_{I_P} \right)^{1/2} \right\|_p \lesssim \|f\|_p. \quad \square \end{aligned}$$

The next lemma shows that the operator C_T restricted to a tree T is a maximal function, if T is overlapping and a maximal truncation of a discrete Calderón-Zygmund operator, if T is lacunary.

Lemma 3.3 (Single tree estimate: maximal truncations). *Let T be a tree. Then, for each $1 < p < \infty$,*

$$\left\| \sum_{P \in T} \langle f, W_{P_u} \rangle W_{P_u}(x) 1_{\omega_{P_l}}(N(x)) \right\|_{L^p(\mathbf{R}_+)} \lesssim \|f\|_p.$$

Proof. It suffices to prove the lemma when T is either lacunary or overlapping. We start with the lacunary case. Note that, for each x , there is a $k = k(x)$ such that

$$\sum_{P \in T} \langle f, W_{P_u} \rangle W_{P_u}(x) 1_{\omega_{P_l}}(N(x)) = \sum_{\substack{P \in T \\ |\omega_P| \geq 2^k}} \langle f, W_{P_u} \rangle W_{P_u}(x).$$

Note that, if $P, P' \in T$ and $|\omega_P| > |\omega_{P'}|$, then ω_{P_u} is disjoint from and sits on the right of $\omega_{P'_u}$. Thus, an interval $\omega = \omega(x)$ exists which contains all ω_{P_u} with $|\omega_P| \geq 2^k$ and which will have empty intersection with all ω_{P_u} satisfying $|\omega_P| < 2^k$. We can thus write

$$\sum_{\substack{P \in T \\ |\omega_P| \geq 2^k}} \langle f, W_{P_u} \rangle W_{P_u}(x) = \mathcal{F}_W^{-1}[1_\omega \mathcal{F}_W[O_T f]](x).$$

Using (2) and the fact that $x \oplus [0, |\omega|^{-1}]$ is an interval of length $|\omega|^{-1}$ containing x , we get

$$\left| \sum_{\substack{P \in T \\ |\omega_P| \geq 2^k}} \langle f, W_{P_u} \rangle W_{P_u}(x) \right| \leq |\omega| \int_{x \oplus [0, |\omega|^{-1}]} |O_T f(y)| dy \lesssim M(O_T f)(x).$$

It now suffices to apply Lemma 3.2.

Assume next that T is overlapping. This case is immediate by noting that, for each x ,

$$\left| \sum_{P \in T} \langle f, W_{P_u} \rangle W_{P_u}(x) 1_{\omega_{P_l}}(N(x)) \right| = |\langle f, W_{P_u} \rangle W_{P_u}(x)| \lesssim M(f)(x),$$

where P is the unique (possibly nonexistent) bitile with $(x, N(x)) \in P_l$. \square

We remark that (6) implies that, whenever T is convex and $P \in T$, we have $\langle f, W_{P_u} \rangle = \langle \Pi_T f, W_{P_u} \rangle$. Thus, the result of Lemma 3.3 can also be written in a localized form

$$(8) \quad \|C_T f\|_p \lesssim \|\Pi_T f\|_p.$$

We close this section by proving L^p estimates for the phase space projection associated with a tree.

Proposition 3.4. *For each convex tree T and each $1 < p \leq \infty$,*

$$\|\Pi_T f\|_p \lesssim \|f\|_p.$$

Proof. We first observe that, for each $x \in I_T$, $\Pi_T f(x) = \Pi_P f(x)$ where P is the minimal bitile in T with $x \in I_P$. Indeed, according to Lemma 2.4, there exists a collection \mathbf{p}'' of pairwise disjoint tiles which includes P_u and P_l such that

$$\Pi_T f(x) = \Pi_P f(x) + \sum_{p \in \mathbf{p}'' \setminus \{P_u, P_l\}} \Pi_p f(x).$$

But P is minimal, hence $\{x\} \cap I_p \cap I_P = \emptyset$ for each $p \in \mathbf{p}'' \setminus \{P_u, P_l\}$.

Finally, note that, for each bitile P and each $x \in I_P$, we have

$$|\Pi_P f(x)| \leq |I_P|^{-1/2} (|\langle f, W_{P_u} \rangle| + |\langle f, W_{P_l} \rangle|) \leq \frac{2}{|I_P|} \int_{I_P} |f| \lesssim Mf(x). \quad \square$$

4. Size and pointwise estimates outside exceptional sets. We are now ready to see the first proof of Theorem 2.6. This argument bears some resemblance to the original argument of Carleson [5]. A form of this argument has been used in [19] in the Walsh case, while the proof of the Fourier case is hidden in [8]. This type of argument proved instrumental in applications to the Return Times theorem [8] and the directional Hilbert transform in the plane [7].

The main tool is the *size* of a collection of bitiles, a concept introduced by Lacey and Thiele [16] in the Fourier case and by Thiele [22] (it is therein referred to as *density*) in the Walsh case.

Definition 4.1. The size of a collection \mathbf{P} of bitiles with respect to a function $f : \mathbf{R}_+ \rightarrow \mathcal{C}$ is defined as

$$\text{size}_f(\mathbf{P}) = \sup_{P \in \mathbf{P}} \frac{\|\Pi_P f\|_2}{|I_P|^{1/2}}.$$

The next two propositions record some of the key features of $\text{size}_f(\mathbf{P})$.

Proposition 4.2. *Let T be a convex tree. Then, for each $1 \leq p \leq \infty$,*

$$\|\Pi_T f\|_p \lesssim \text{size}_f(T) |I_T|^{1/p}.$$

Proof. It suffices to prove that $\|\Pi_T f\|_\infty \lesssim \text{size}_f(T)$. The proof of Proposition 3.4 shows that, for each $x \in I_T$, $\Pi_T f(x) = \Pi_P f(x)$ where P is the minimal bitile in T with $x \in I_P$. To close the argument, observe that, for each bitile P , we have

$$\begin{aligned} \|\Pi_P f\|_\infty &\leq |I_P|^{-1/2} (|\langle f, W_{P_u} \rangle| + |\langle f, W_{P_l} \rangle|) \\ &\leq |I_P|^{-1/2} \sqrt{2} (|\langle f, W_{P_u} \rangle|^2 + |\langle f, W_{P_l} \rangle|^2)^{1/2} \\ &= |I_P|^{-1/2} \sqrt{2} \|\Pi_P f\|_2. \quad \square \end{aligned}$$

Proposition 4.3. *For each \mathbf{P} and $f : \mathbf{R}_+ \rightarrow \mathbf{C}$,*

$$\text{size}_f(\mathbf{P}) \lesssim \sup_{P \in \mathbf{P}} \inf_{x \in I_P} M(f)(x).$$

Proof. This is immediate since, for each P ,

$$\|\Pi_P f\|_2 = (|\langle f, W_{P_u} \rangle|^2 + |\langle f, W_{P_l} \rangle|^2)^{1/2} \leq \frac{\sqrt{2}}{|I_P|^{1/2}} \int_{I_P} |f|. \quad \square$$

Definition 4.4. A forest \mathcal{F} is a finite collection of bitiles which consists of a disjoint union of convex trees.

We note that each finite convex collection \mathbf{P} of bitiles can be turned into a forest, possibly in more than one way. Indeed, start with a maximal element P from \mathbf{P} , and construct the maximal tree in \mathbf{P} with top bitile P . Remove this tree T from \mathbf{P} and note that $\mathbf{P} \setminus T$ remains convex. Repeat the procedure with $\mathbf{P} \setminus T$ replacing \mathbf{P} , to select the next tree. Iterate this until all the bitiles from the original \mathbf{P} are selected.

We will sometimes use the notation $N_{\mathcal{F}}$ for the counting function of a forest \mathcal{F}

$$N_{\mathcal{F}}(x) = \sum_{T \in \mathcal{F}} 1_{I_T}(x).$$

A key idea in many of the approaches to Carleson's theorem is to split the bitiles into forests with a certain size.

Lemma 4.5. *Let \mathbf{P} be a finite convex collection of bitiles, and let $f : \mathbf{R}_+ \rightarrow \mathbf{C}$. Then $\mathbf{P} = \mathbf{P}_{\text{hi}} \cup \mathbf{P}_{\text{lo}}$, such that*

- both \mathbf{P}_{lo} and \mathbf{P}_{hi} are convex;
- $\text{size}_f(\mathbf{P}_{\text{lo}}) \leq 1/2 \text{size}_f(\mathbf{P})$;
- \mathbf{P}_{hi} is a convex forest with trees $T \in \mathcal{F}$ satisfying

$$\sum_{T \in \mathcal{F}} |I_T| \lesssim \text{size}_f(\mathbf{P})^{-2} \|f\|_2^2.$$

Proof. This is a recursive procedure. Set $\mathbf{P}_{\text{stock}} := \mathbf{P}$ and $\mathcal{F} = \emptyset$. Select a maximal bitile $t \in \mathbf{P}_{\text{stock}}$ such that

$$\|\Pi_t f\|_2 |I_t|^{-1/2} > \frac{\text{size}_f(\mathbf{P})}{2}.$$

Define

$$T(t) = \{P \in \mathbf{P}_{\text{stock}} : P \leq t\},$$

and note that since $\mathbf{P}_{\text{stock}}$ is convex, both $\mathbf{P}_{\text{stock}} \setminus T(t)$ and the tree $T(t)$ will be convex. Add $T(t)$ to the family \mathcal{F} . Reset $\mathbf{P}_{\text{stock}} := \mathbf{P}_{\text{stock}} \setminus T(t)$, and restart the procedure.

The algorithm is over when there is no t to be selected. Then define \mathbf{P}_{hi} to consist of the union of all bitiles in all the trees from \mathcal{F} , and let $\mathbf{P}_{\text{lo}} = \mathbf{P} \setminus \mathbf{P}_{\text{hi}}$.

The first two needed properties as well as the convexity of \mathcal{F} are quite immediate. By maximality, the selected bitiles t are pairwise disjoint, and thus the functions $\Pi_t f$ are pairwise orthogonal, thanks to (3). It follows that

$$\sum_{T \in \mathcal{F}} |I_T| = \sum_t |I_t| \leq 4 \text{size}_f(\mathbf{P})^{-2} \sum_t \|\Pi_t f\|_2^2 \leq 4 \text{size}_f(\mathbf{P})^{-2} \|f\|_2^2. \quad \square$$

We can iterate the lemma to obtain

Proposition 4.6 (Size decomposition). *Let \mathbf{P} be a finite convex collection of bitiles. Then*

$$\mathbf{P} = \bigcup_{2^{-n} \leq \text{size}_f(\mathbf{P})} \mathbf{P}_n \cup \mathbf{P}_{\text{null}},$$

such that

- $\text{size}_f(\mathbf{P}_n) \leq 2^{-n}$;
- \mathbf{P}_n is a convex forest with trees $T \in \mathcal{F}_n$ satisfying

$$(10) \quad \sum_{T \in \mathcal{F}_n} |I_T| \lesssim 2^{2n} \|f\|_2^2,$$

- $\Pi_P f \equiv 0$ for each $P \in \mathbf{P}_{\text{null}}$.

We are now ready for the main line of the argument.

Proof of Theorem 2.6. By using restricted type interpolation and a limiting argument, it suffices to prove

$$(11) \quad |\{x : |C_{\mathbf{P}} f(x)| \gtrsim \lambda\}| \lesssim \frac{1}{\lambda^p} |E|$$

for each $E \subset \mathbf{R}_+$ with finite measure, each $|f| \leq 1_E$, each finite convex $\mathbf{P} \subset \mathbf{P}_{\text{all}}$, for each $\lambda > 1$, $4 < p < \infty$ and also for each $0 < \lambda \leq 1$, $1 < p < \infty$. The implicit constant in the inequality (11) will only depend on p .

We start with the simpler case $\lambda > 1$, $4 < p < \infty$. Note that the bound $\text{size}_f(\mathbf{P}) \lesssim 1$ follows from Proposition 4.6. Let $\mathbf{P}_n, \mathcal{F}_n$ with $2^{-n} \lesssim 1$, be the collections from Proposition 4.6. For each $T \in \mathcal{F}_n$ with top bitile $P_T \in T$, define the saturation $T^* = \{P \in \mathbf{P}_n : P \leq P_T\}$. Note that the trees T^* remain convex, but in general they are not pairwise disjoint. Call \mathcal{F}_n^* the collection of the trees T^* , and note that $I_{T^*} = I_T$.

Define the exceptional set

$$F = \bigcup_{2^{-n} \lesssim 1} \bigcup_{T^* \in \mathcal{F}_n^*} \{x : |C_{T^*} f(x)| > \lambda 2^{-n/2}\}.$$

Note that, by (8) and (9) we have

$$|\{x : |C_{T^*} f(x)| > \lambda 2^{-n/2}\}| \lesssim |I_T| (\lambda 2^{n/2})^{-p}.$$

Combining this with (10) and $p > 4$, we obtain $|F| \lesssim 1/\lambda^p |E|$.

Thus, it remains to prove that $|C_{\mathbf{P}}f(x)| \lesssim \lambda$ on F^c . The crucial observation behind this approach to Theorem 2.6 is that, for each n , the contribution to each x comes from a single tree $T^* \in \mathcal{F}_n^*$. Indeed, note that the contributing bitiles $P \in \mathbf{P}_n$ satisfy $(x, N(x)) \in I_P \times \omega_{P_l}$. Since all these bitiles P contain $(x, N(x))$, they will be pairwise comparable under \leq . Call P_x the unique maximal bitile among them. Let T_x^* be one of the trees in \mathcal{F}_n^* containing P_x . It follows that all the contributing bitiles belong to T_x^* . Thus, if $x \notin F$,

$$|C_{\mathbf{P}_n}f(x)| = |C_{T_x^*}f(x)| \leq \lambda 2^{-n/2}.$$

It further follows by linearity that

$$|C_{\mathbf{P}}f(x)| \leq \sum_{2^{-n} \lesssim 1} |C_{\mathbf{P}_n}f(x)| \lesssim \lambda.$$

The case $0 < \lambda \leq 1$ is very similar but we need an additional exceptional set

$$G := \{x : M(f)(x) > \lambda^p\}.$$

Since $|G| \lesssim 1/\lambda^p|E|$ and since W_{P_l} is supported on I_P , it is enough to prove (11) with \mathbf{P} restricted to those bitiles such that $P \not\subset G$. Another application of Proposition 4.3 shows that $\text{size}_f(\mathbf{P}) \lesssim \lambda^p$. Let $\mathbf{P}_n, \mathcal{F}_n$ with $2^{-n} \lesssim \lambda^p$ be the collections from Proposition 4.6 corresponding to our new \mathbf{P} . Define

$$F = \bigcup_{2^{-n} \lesssim \lambda^p} \bigcup_{T^* \in \mathcal{F}_n^*} \{x : |C_{T^*}f(x)| > \lambda^{1/2} 2^{-\frac{n}{2p}}\}.$$

As before, (8) and (9) imply

$$|\{x : |C_{T^*}f(x)| > \lambda^{1/2} 2^{-n/2p}\}| \lesssim |I_T| \left(\frac{2^{-n}}{\lambda^{1/2} 2^{-n/2p}} \right)^{p/p-1}.$$

We immediately get that $|F| \lesssim |E| \leq 1/\lambda^p|E|$. Note also that, if $x \notin F$,

$$|C_{\mathbf{P}}f(x)| \leq \sum_{2^{-n} \lesssim \lambda^p} |C_{\mathbf{P}_n}f(x)| \leq \sum_{2^{-n} \lesssim \lambda^p} \lambda^{1/2} 2^{-n/2p} \lesssim \lambda. \quad \square$$

5. Mass and Fefferman's argument. The argument in this section is a translation to the Walsh case of Fefferman's proof [10]. The key tool used in this proof is *mass*.

Definition 5.1. The mass of a convex collection \mathbf{P} of bitiles is defined as

$$\text{mass}(\mathbf{P}) = \sup_{P \in \mathbf{P}} \frac{|E(P)|}{|I_P|},$$

where $E(P) = I_P \cap N^{-1}(\omega_P)$.

In some sense the mass of a single bitile P measures (or better said, it puts an upper bound on, since $\omega_{P_i} \subset \omega_P$) how much P contributes to $C_P f(x)$. Indeed, it suffices to note that, for $1 \leq p \leq \infty$, we have $\|C_P f\|_p \leq \text{mass}(P)^{1/p} \|f\|_p$. We will next extend this inequality to the case of trees and then to a special type of forests.

We have the following analogue of Lemma 4.5. We will restrict attention to the bitiles spatially supported in $[0, 1]$

$$\mathbf{P}_{[0,1]} := \{P \in \mathbf{P}_{\text{all}} : I_P \subset [0, 1]\}.$$

Lemma 5.2. *Let \mathbf{P} be a finite convex collection of bitiles in $\mathbf{P}_{[0,1]}$. Then $\mathbf{P} = \mathbf{P}_{\text{hi}} \cup \mathbf{P}_{\text{lo}}$, such that*

- both \mathbf{P}_{lo} and \mathbf{P}_{hi} are convex;
- $\text{mass}(\mathbf{P}_{\text{lo}}) \leq 1/2 \text{mass}(\mathbf{P})$;
- \mathbf{P}_{hi} is a convex forest with trees $T \in \mathcal{F}$ satisfying

$$\sum_{T \in \mathcal{F}} |I_T| \lesssim \text{mass}(\mathbf{P})^{-1}.$$

Proof. This is another recursive procedure. Set $\mathbf{P}_{\text{stock}} := \mathbf{P}$. Select a maximal bitile $t \in \mathbf{P}_{\text{stock}}$ such that

$$\text{mass}(t) > \frac{\text{mass}(\mathbf{P})}{2}.$$

Define

$$T(t) = \{P \in \mathbf{P}_{\text{stock}} : P \leq t\}.$$

and note that, since $\mathbf{P}_{\text{stock}}$ is convex, both $\mathbf{P}_{\text{stock}} \setminus T(t)$ and the tree $T(t)$ will be convex. Add $T(t)$ to the family \mathcal{F} . Reset the new value $\mathbf{P}_{\text{stock}} := \mathbf{P}_{\text{stock}} \setminus T(t)$, and restart the procedure.

The algorithm is over when there is no t to be selected. Then define \mathbf{P}_{hi} to consist of the union of all bitiles in all the trees from \mathcal{F} , and let $\mathbf{P}_{\text{lo}} = \mathbf{P} \setminus \mathbf{P}_{\text{hi}}$.

The first two needed properties as well as the convexity of \mathcal{F} are quite immediate. By maximality the selected bitiles t are pairwise disjoint, and thus the sets $E(t)$ will also be pairwise disjoint. Since $E(t) \subset [0, 1]$, it follows that

$$\sum_{T \in \mathcal{F}} |I_T| = \sum_t |I_t| \leq 2 \text{mass}(\mathbf{P})^{-1} \sum_t |E(t)| \leq 2 \text{mass}(\mathbf{P})^{-1}. \quad \square$$

Note that the mass of any collection is trivially bounded by 1. We can iterate the lemma to obtain the following analogue of Proposition 4.6.

Proposition 5.3 (Mass decomposition). *Let $\mathbf{P} \subset \mathbf{P}_{[0,1]}$ be a finite convex collection of bitiles. Then*

$$\mathbf{P} = \bigcup_{n \geq 0} \mathbf{P}_n \cup \mathbf{P}_{\text{null}},$$

such that

- $\text{mass}(\mathbf{P}_n) \leq 2^{-n}$;
- \mathbf{P}_n is a convex forest with trees $T \in \mathcal{F}_n$ satisfying

$$\sum_{T \in \mathcal{F}_n} |I_T| \lesssim 2^n,$$

- $C_P f \equiv 0$ for each $P \in \mathbf{P}_{\text{null}}$ and each f .

It turns out that we have not only L^1 but also dyadic BMO_Δ control for $N_{\mathcal{F}_n}$. Recall that

$$\|f\|_{\text{BMO}_\Delta} = \sup_{I \text{ dyadic}} \frac{1}{|I|} \int_I \left| f(x) - \frac{1}{|I|} \int_I f \right| dx.$$

Proposition 5.4. *We have*

$$\|N_{\mathcal{F}_n}\|_{\text{BMO}_\Delta} \lesssim 2^n.$$

In particular, for each $I \in \mathcal{D}_+$ and $\lambda > 0$

$$\left| \left\{ x : \sum_{T \in \mathcal{F}_n: I_T \subset I} 1_{I_T}(x) \geq C\lambda 2^n \right\} \right| \leq e^{-\lambda}|I|,$$

where C is large enough.

Proof. The first inequality follows since one can easily check that

$$\|N_{\mathcal{F}_n}\|_{\text{BMO}_\Delta} \lesssim \sup_{I \in \mathcal{D}_+} \frac{\sum_{T \in \mathcal{F}_n: I_T \subset I} |I_T|}{|I|}.$$

The second one is just a consequence of John-Nirenberg’s inequality. \square

Given a convex tree T with top data (ξ_T, I_T) , define \mathbf{J}_T to be the collection of all maximal dyadic intervals $J \subset I_T$ such that J contains no I_P with $P \in T$. For $J \in \mathbf{J}_T$, define

$$G_J = J \cap \left(\bigcup_{P \in T: J \subset I_P} E(P) \right).$$

We first observe that the intervals of \mathbf{J}_T form a partition of I_T . Note also that, if J intersects some I_P with $P \in T$, then the convexity of T forces the dyadic parent of each $J \in \mathbf{J}_T$ to equal $I_{P(J)}$, for some $P(J) \in T$. Thus, $G_J \subset E(P(J))$, which implies the following crucial Carleson measure type estimate

$$(12) \quad |G_J| \leq 2\text{mass}(T)|J|,$$

for each $J \in \mathbf{J}_T$. We also need to observe that $C_T f$ is supported on $\bigcup_{J \in \mathbf{J}_T} G_J$.

Proposition 5.5 (Tree estimate). *Let T be a convex tree. Then, for each $1 < p < \infty$,*

$$\|C_T f\|_p \lesssim (\text{mass}(T))^{1/p} \|f\|_p.$$

Proof. Assume first that T is overlapping. Note that, for each $J \in \mathbf{J}_T$ and $x \in J$, if $C_T f(x)$ is nonzero, then $|C_T f(x)| = 1/|I_P|^{1/2} |\langle f, W_{P_u} \rangle|$ for some $P \in T$ with $J \subset I_P$. Thus, $|C_T f(x)| \leq \inf_{y \in J} M(f)(y)$. We conclude that

$$\begin{aligned} \int |C_T f(x)|^p dx &= \sum_{J \in \mathbf{J}_T} \int_{G_J} |C_T f(x)|^p dx \\ &\lesssim \sum_{J \in \mathbf{J}_T} |G_J| \inf_{y \in J} [M(f)(y)]^p \\ &\leq 2 \operatorname{mass}(T) \sum_{J \in \mathbf{J}_T} \int_J |Mf(x)|^p dx \\ &= 2 \operatorname{mass}(T) \|Mf\|_{L^p(I_T)}^p \\ &\lesssim \operatorname{mass}(T) \|f\|_p^p. \end{aligned}$$

On the other hand, if T is lacunary, we reason as in the proof of Lemma 3.3 to write

$$C_T f(x) = \mathcal{F}_W^{-1}(1_\omega \mathcal{F}_W(O_T f))(x),$$

where the interval ω is any interval which contains all ω_{P_u} with $x \in I_P$. Recall, however, that $|I_P| \geq 2|J|$ for all such P and, hence, since $\xi_T \in \omega_{P_l}$, all ω_{P_u} will be contained in the interval $[\xi_T, \xi_T + |J|^{-1}]$. Thus,

$$|C_T f(x)| \leq \frac{1}{|J|} \int_{x \oplus [0, |J|]} |O_T f(y)| dy \lesssim \inf_{y \in J} M(O_T f)(y).$$

A repetition of the argument from the overlapping case combined with Lemma 3.2 ends the proof. \square

The proof of this proposition shows that, for each $J \in \mathbf{J}_T$,

$$(13) \quad \|C_T f\|_{L^\infty(J)} \leq \inf_{x \in J} M(V_T f)(x),$$

where $V_T f = \Pi_T f$ if T is overlapping and $V_T f = O_T(\Pi_T f)$ if T is lacunary.

Definition 5.6. We call \mathcal{F} a Fefferman forest if two bitiles from any two distinct trees in \mathcal{F} are pairwise disjoint.

It turns out that, outside a small exceptional set, each forest \mathcal{F} can be written as a disjoint union of a small number of Fefferman forests.

Lemma 5.7 (The Fefferman trick). *Assume that \mathbf{P} is a convex set in $\mathbf{P}_{[0,1]}$, which is organized as a forest \mathcal{F} . For each $K > 1$, there is an exceptional set F with $|F| \leq e^{-K}$ such that*

•

$$(14) \quad \left\| \sum_{T \in \mathcal{F}: I_T \not\subset F} 1_{I_T} \right\|_{\infty} \lesssim K \|N_{\mathcal{F}}\|_{\text{BMO}_{\Delta}};$$

• *the bitiles in $\mathbf{P}_F := \bigcup_{T \in \mathcal{F}: I_T \not\subset F} \{P : P \in T\}$ can be partitioned into $O(\log(K \|N_{\mathcal{F}}\|_{\text{BMO}_{\Delta}}))$ Fefferman forests.*

Proof. Define $F = \{x : \sum_{T \in \mathcal{F}} 1_{I_T}(x) > CK \|N_{\mathcal{F}}\|_{\text{BMO}_{\Delta}}\}$, for large enough C . Note that $|F| \leq e^{-K}$ by John-Nirenberg’s inequality. Call $\mathcal{F}_F := \{T \in \mathcal{F} : I_T \not\subset F\}$. Then (14) is immediate. Let P_T be any top bitile for T . Define, for each $l \in \mathbf{N}$ with $1 \leq 2^l \leq CK \|N_{\mathcal{F}}\|_{\text{BMO}_{\Delta}}$,

$$\mathbf{P}_F^l := \{P \in \mathbf{P}_F : 2^l \leq \#\{T \in \mathcal{F}_F : P \leq P_T\} < 2^{l+1}\}.$$

It remains to prove that \mathbf{P}_F^l is a Fefferman forest. Note that \mathbf{P}_F^l is convex, in part because \mathbf{P} is convex. For each maximal element $t \in \mathbf{P}_F^l$, let $T(t) = \{P \in \mathbf{P}_F^l : P \leq t\}$. Obviously, each tree $T(t)$ is convex. Assume, for contradiction, $P \leq P'$ for some $P \in T(t)$, $P' \in T(t')$ with $t \neq t'$. Then $P \leq t'$, in addition to $P \leq t$. Thus, I_t intersects $I_{t'}$. But then, since t and t' are pairwise incomparable under \leq , the sets $\{T \in \mathcal{F}_F : t \leq P_T\}$ and $\{T \in \mathcal{F}_F : t' \leq P_T\}$ must be pairwise disjoint. Note that

$$\{T \in \mathcal{F}_F : t \leq P_T\} \cup \{T \in \mathcal{F}_F : t' \leq P_T\} \subset \{T \in \mathcal{F}_F : P \leq P_T\},$$

and this will force the contradiction $2^l + 2^l < 2^{l+1}$. \square

We will repeatedly use the fact that, if T, T' are trees in a Fefferman forest, then $C_T f$ and $C_{T'} f$ are disjointly supported while $\Pi_T f$ and $\Pi_{T'} f$ are orthogonal.

Proposition 5.8 (Forest estimate). *Let \mathcal{F} be a Fefferman forest. Then, for each $1 < p < \infty$, there exists $\delta(p) < 1/p$ such that*

$$\|C_{\mathcal{F}}f\|_p \lesssim (\text{mass}(\mathcal{F}))^{1/p} (\|N_{\mathcal{F}}\|_{\infty})^{\delta(p)} \|f\|_p.$$

When $p \geq 2$, we can take $\delta(p) = 0$.

Proof. First let $p \geq 2$. Consider the vector-valued operator $Vf = (\Pi_T f)_{T \in \mathcal{F}}$. Proposition 3.4 shows that $\|V\|_{L^{\infty} \rightarrow l^{\infty}(L^{\infty})} \lesssim 1$. On the other hand, the pairwise orthogonality of $\Pi_T f$ implies the bound $\|V\|_{L^2 \rightarrow l^2(L^2)} \lesssim 1$. Interpolation [3] now gives

$$(15) \quad \|V\|_{L^p \rightarrow l^p(L^p)} \lesssim 1.$$

Using Proposition 5.5, we get

$$\|C_{\mathcal{F}}f\|_p^p = \sum_{T \in \mathcal{F}} \|C_T f\|_p^p \lesssim \sum_{T \in \mathcal{F}} \text{mass}(T) \|\Pi_T f\|_p^p \leq \text{mass}(\mathcal{F}) \|f\|_p^p.$$

Assume next that $p < 2$. Split \mathcal{F} into $\|N_{\mathcal{F}}\|_{\infty}$ forests \mathcal{F}_k so that, for each k and each $T, T' \in \mathcal{F}_k$, we have $I_T \cap I_{T'} = \emptyset$. Note that we have, as before,

$$\begin{aligned} \|C_{\mathcal{F}}f\|_p^p &= \sum_k \sum_{T \in \mathcal{F}_k} \|C_T f\|_p^p \lesssim \sum_k \sum_{T \in \mathcal{F}_k} \text{mass}(T) \|f 1_{I_T}\|_p^p \\ &\leq \text{mass}(\mathcal{F}) \sum_k \|f\|_p^p = \text{mass}(\mathcal{F}) \|N_{\mathcal{F}}\|_{\infty} \|f\|_p^p. \end{aligned}$$

To close the argument, interpolate the L^p and L^2 bounds for the operator $f \mapsto C_{\mathcal{F}}f$. \square

We are now ready to prove the following variant of Theorem 2.6. Note that this weaker version is enough to prove Theorem 1.2 via the standard approximation argument. Then, in the case $1 < p \leq 2$, one can actually use Stein's continuity principle [20] to show that Theorem 1.2 implies the stronger Theorem 2.6. The author is not aware of any substitute argument which proves the same implication for $p > 2$. However, in Section 7, we present a recent delicate refinement of Fefferman's argument due to Lie [18] which closes this gap.

Theorem 5.9. *For each $1 < q < p < \infty$, we have*

$$\|C_{\mathbf{P}_{[0,1]}} f\|_q \lesssim \|f\|_p.$$

Proof. By invoking a limiting argument, it suffices to prove the estimate with $\mathbf{P}_{[0,1]}$ replaced by a finite convex collection $\mathbf{P} \subset \mathbf{P}_{[0,1]}$, as long as the implicit constant in the inequality is independent of \mathbf{P} . Apply Proposition 5.3 to \mathbf{P} to get $\mathbf{P}_n, \mathcal{F}_n$. Apply Lemma 5.7 to each \mathcal{F}_n with $K = K_n := (n+1)L$, $L \geq 2$ to be determined later. We get exceptional sets $|F_n| \leq e^{-L(n+1)}$ and the partitions of $\mathcal{F}_{F_n, n} := \{T \in \mathcal{F}_n : I_T \not\subset F_n\}$ into $O(n + \log L)$ Fefferman forests $\mathcal{F}_{n,k}$. Define $F = \cup_n F_n$ and note that $|F| \lesssim e^{-L}$. Also let \mathbf{P}^* be the bitiles in the trees from $\cup_n \mathcal{F}_{F_n, n}$. Note that $C_{\mathbf{P}} f(x) = C_{\mathbf{P}^*} f(x)$ on F^c . By Proposition 5.8 and linearity, we get

$$\begin{aligned} \|C_{\mathbf{P}} f\|_{L^p(F^c)} &\leq \sum_n \sum_k \|C_{\mathcal{F}_{n,k}} f\|_{L^p([0,1])} \\ &\lesssim \sum_n 2^{-n/p} (n + \log L) (n2^n L)^{\delta(p)} \|f\|_p \\ &\lesssim L^{1/p} \|f\|_p. \end{aligned}$$

Thus, for each $\lambda > 1$,

$$|\{x \in [0, 1] : |C_{\mathbf{P}} f(x)| > \lambda\}| \lesssim L \left(\frac{\|f\|_p}{\lambda} \right)^p + e^{-L}.$$

By optimizing L , we get that, for each $s < p$,

$$|\{x \in [0, 1] : |C_{\mathbf{P}} f(x)| > \lambda\}| \lesssim \left(\frac{\|f\|_p}{\lambda} \right)^s.$$

A simple integration argument finishes the proof of the theorem. \square

6. Combining mass and size: The Lacey-Thiele argument.

The Fourier version of this argument is due to Lacey and Thiele [15]. It uses both mass and size and thus it combines elements of the two proofs we have seen in earlier sections. The interplay between mass and size made this approach particularly well suited for applications to

the problem of singular integrals along vector fields [1, 2, 13] since it opened the door to the use of Kakeya type maximal functions.

The definition of $\text{size}(\mathbf{P})$ will remain the same as in Section 4. Let F be a finite measure subset of \mathbf{R}_+ . We slightly modify the definition of mass.

Definition 6.1. The mass of a convex collection \mathbf{P} of bitiles relative to F is defined as

$$\text{mass}_F(\mathbf{P}) = \sup_{P \in \mathbf{P}} \frac{|E_F(P)|}{|I_P|}$$

where now $E_F(P) = F \cap I_P \cap N^{-1}(\omega_P)$.

We have the following version of Proposition 5.3.

Proposition 6.2 (Mass decomposition). *Let $\mathbf{P} \subset \mathbf{P}_{\text{all}}$ be a finite convex collection of bitiles. Then*

$$\mathbf{P} = \bigcup_{m \geq 0} \mathbf{P}_m^* \cup \mathbf{P}_{\text{null}}^*$$

such that

- $\text{mass}_F(\mathbf{P}_m^*) \leq 2^{-m}$;
- \mathbf{P}_m^* is a convex forest with trees $T \in \mathcal{F}_m^*$ satisfying

$$\sum_{T \in \mathcal{F}_m^*} |I_T| \lesssim 2^m |F|;$$

- $C_P f \equiv 0$ for each $P \in \mathbf{P}_{\text{null}}^*$ and each $f : \mathbf{R}_+ \rightarrow \mathbf{C}$.

Proposition 6.3 (Tree estimate). *Let T be a convex tree. Then, for each $f \in L^2(\mathbf{R}_+)$,*

$$|\langle C_T f, 1_F \rangle| \lesssim |I_T| \text{mass}_F(T) \text{size}_f(T).$$

Proof. For $J \in \mathbf{J}_T$, define as before $G_J = J \cap (\bigcup_{P \in T: J \subset I_P} E_F(P))$. Recall that $C_T f|_{1_F}$ restricted to J is supported on G_J . We then estimate as in the proof of Proposition 5.5,

$$\begin{aligned} |\langle C_T f, 1_F \rangle| &\leq \sum_{J \in \mathbf{J}_T} \int_{G_J} |C_T f| \\ &\leq \text{mass}_F(T) \sum_{J \in \mathbf{J}_T} \int_J M(V_T(f)) \\ &\leq \text{mass}_F(T) \sum_{J \in \mathbf{J}_T} |J|^{1/2} \left(\int_J |M(V_T(f))|^2 \right)^{1/2} \\ &\leq \text{mass}_F(T) |I_T|^{1/2} \|V_T(f)\|_2, \end{aligned}$$

where $V_T f = \Pi_T f$ if T is overlapping and $V_T = O_T(\Pi_T f)$ is T is lacunary. The result now follows from (9) and Lemma 3.2. \square

Proof of Theorem 2.6. By using a restricted type interpolation and a limiting argument, it suffices to prove that, given $1 < p < \infty$, $p \neq 2$, with dual exponent p' and, given finite measure subsets E, G of \mathbf{R}_+ , there exists an $F \subset G$ with $|F| \geq |G|/2$ such that

$$(16) \quad |\langle C_{\mathbf{P}} f, 1_F \rangle| \lesssim |E|^{1/p} |F|^{1/p'}$$

for each $|f| \leq 1_E$ and each finite convex $\mathbf{P} \subset \mathbf{P}_{\text{all}}$.

We first analyze the case $p > 2$, when we can take $F = G$. Note that $\text{size}_f(\mathbf{P}) \lesssim 1$. Apply to \mathbf{P} Propositions 4.6 and 6.2 to get $\mathbf{P}_n, \mathbf{P}_m^*, \mathcal{F}_n, \mathcal{F}_m^*$ for $m \geq 0$ and $2^{-n} \lesssim 1$. Define $\mathbf{P}_{n,m} := \mathbf{P}_n \cap \mathbf{P}_m^*$. We organize the bitiles in $\mathbf{P}_{n,m}$ into a forest in two different ways. Call $\mathcal{F}_{n,m}$ the trees $T \cap \mathbf{P}_m^*$ with $T \in \mathcal{F}_n$, and call $\mathcal{F}_{n,m}^*$ the trees $T \cap \mathbf{P}_n$ with $T \in \mathcal{F}_m^*$. Note that the resulting trees are convex and

$$\begin{aligned} \sum_{T \in \mathcal{F}_{n,m}} |I_T| &\lesssim 2^{2n} |E| \\ \sum_{T \in \mathcal{F}_{n,m}^*} |I_T| &\lesssim 2^m |F|. \end{aligned}$$

Thus, using Proposition 6.3 with the two partitions above, we get

$$|\langle C_{\mathbf{P}_{n,m}} f, 1_F \rangle| \lesssim 2^{-n} 2^{-m} \min(2^{2n} |E|, 2^m |F|).$$

We conclude that

$$\begin{aligned} |\langle C_{\mathbf{P}}f, 1_F \rangle| &\leq \sum_{2^{-n} \lesssim 1, m \geq 0} |\langle C_{\mathbf{P}_{n,m}}f, 1_F \rangle| \\ &\lesssim \sum_{n,m} 2^{-n} 2^{-m} (2^{2n}|E|)^{1/p} (2^m|F|)^{1/p'} \\ &\lesssim |E|^{1/p} |F|^{1/p'}. \end{aligned}$$

Now consider the case $p < 2$. We define

$$F := G \setminus \left\{ x : M(1_E)(x) > 10 \frac{|E|}{|G|} \right\}.$$

Note that $|F| \geq |G|/2$. Call $\mathbf{P}^* = \{P \in \mathbf{P} : I_P \not\subset F^c\}$, and note that $\langle C_{\mathbf{P}}f, 1_F \rangle = \langle C_{\mathbf{P}^*}f, 1_F \rangle$. Proposition 4.3 implies the bound $\text{size}_f(\mathbf{P}^*) \lesssim |E|/|F|$.

Proceeding as before, with the two partitions this time for \mathbf{P}^* , we obtain

$$\begin{aligned} |\langle C_{\mathbf{P}^*}f, 1_F \rangle| &\leq \sum_{\substack{2^{-n} \lesssim |E|/|F| \\ m \geq 0}} |\langle C_{\mathbf{P}_{n,m}}f, 1_F \rangle| \\ &\lesssim \sum_{\substack{2^{-n} \lesssim |E|/|F| \\ m \geq 0}} 2^{-n} 2^{-m} (2^{2n}|E|)^{1/p'} (2^m|F|)^{1/p} \\ &\lesssim |E|^{1/p} |F|^{1/p'}. \quad \square \end{aligned}$$

7. A direct proof of strong L^p bounds. All three arguments presented earlier rely on proving weak type bounds and using interpolation. Moreover, as observed in Section 5, Fefferman's argument does not imply strong type L^p bounds for the operator $C_{\mathbf{P}_{[0,1]}}$ when $p \geq 2$. The recent argument of Lie [18] proves that

$$(17) \quad \|C_{\mathbf{P}_{[0,1]}}f\|_p \lesssim \|f\|_{L^p[0,1]}$$

directly, without any interpolation that restricts f to special classes of functions (such as sub-characteristic). The case $p = 2$ will require no interpolation in the argument, while the case $p \neq 2$ will rely on

vector-valued interpolation for various operators relevant to the proof. We will restrict attention to $p \geq 2$.

This argument, while also interesting in itself, has been developed in [18] in order to solve a conjecture of Stein on the boundedness of the quadratic Carleson operator.

We will keep the notation for mass from Section 5 and restrict attention to the bitiles in $\mathbf{P}_{[0,1]}$. Recall (see Lemma 5.7) that Fefferman’s approach consisted of decomposing every forest into a small number of Fefferman forests, outside a small exceptional set. The lemma below shows how to iterate the Fefferman trick inside the exceptional set, until all bitiles are exhausted. Each stage of the iteration creates more layers of Fefferman forests, but the size of their spatial support will get exponentially smaller.

Lemma 7.1. *Let $\mathbf{P} \subset \mathbf{P}_{[0,1]}$ be a finite convex collection of bitiles. Then we have the following partitions*

$$(18) \quad \mathbf{P} = \bigcup_{k \geq 0} \mathbf{P}_k$$

$$(19) \quad \mathbf{P}_k = \bigcup_{m \geq 1} \mathbf{Q}_k^m$$

$$(20) \quad \mathbf{Q}_k^m = \bigcup_{1 \leq n \lesssim k} \mathcal{F}_k^{m,n}$$

where each $\mathcal{F}_k^{m,n}$ is a Fefferman forest such that

$$(21) \quad \|N_{\mathcal{F}_k^{m,n}}\|_\infty \lesssim k2^k$$

and

$$(22) \quad \text{mass} \left(\mathbf{P} \setminus \bigcup_{i=0}^{k-1} \mathbf{P}_i \right) \leq 2^{1-k}.$$

Moreover, there are sets $(E_k^m)_{m,k \geq 0}$ which are finite disjoint unions \mathbf{J}_k^m of dyadic intervals such that

$$(23) \quad E_k^{m+1} \subset E_k^m \quad \text{and} \quad |E_k^{m+1} \cap J| \leq e^{-10k}|J|, \text{ for } J \in \mathbf{J}_k^m$$

$$(24) \quad P \in \mathbf{Q}_k^{m+1} \implies I_P \subset E_k^m, I_P \not\subset E_k^{m+1}$$

$$(25) \quad P \in \mathbf{Q}_k^{m+1} \implies |I_P \cap E_k^{m+2}| \leq e^{-10k}|I_P|.$$

Proof. Set $\mathbf{P}_0 := \emptyset$. Assume that, given $k > 0$, each \mathbf{P}_i has been defined for each $0 \leq i \leq k - 1$ so that it satisfies (19)–(25).

Let us see how to define \mathbf{P}_k . For the rest of the argument, set $\mathbf{P}_{\text{stock}} = \mathbf{P} \setminus (\cup_{0 \leq i \leq k-1} \mathbf{P}_i)$. Let $\mathbf{P}_{k,\max}^0$ be the maximal bitiles in the set

$$\{P \in \mathbf{P}_{\text{stock}} : \text{mass}(P) \geq 2^{-k}\}.$$

Define $E_k^0 := [0, 1]$,

$$\mathbf{P}_k^0 = \{P \in \mathbf{P}_{\text{stock}} : P \leq P' \text{ for some } P' \in \mathbf{P}_{k,\max}^0\}$$

$$N_k^0(x) = \sum_{P \in \mathbf{P}_{k,\max}^0} 1_{I_P}(x),$$

and the exceptional set

$$E_k^1 = \{x : N_k^0(x) \geq C_1 k 2^k\}$$

where C_1 is a large enough constant to be determined later. Note that, since the sets E_P are pairwise disjoint for $P \in \mathbf{P}_{k,\max}^0$, it follows that $\|N_k^0\|_{\text{BMO}_\Delta} \leq 2^k$. This and John-Nirenberg in turn imply (23) for $m = 0$. Define

$$\mathcal{Q}_k^1 = \{P \in \mathbf{P}_k^0 : I_P \not\subset E_k^1\},$$

and note that (24) holds for $m = 0$.

We continue the construction of the sets E_k^m, \mathbf{Q}_k^m inductively. Fix $m_0 \geq 1$. Assume $E_k^0, \dots, E_k^{m_0}$ and $\mathbf{Q}_k^1, \dots, \mathbf{Q}_k^{m_0}$ have been constructed so that (23)–(24) hold for each $0 \leq m \leq m_0 - 1$ and so that (25) holds for each $m \leq m_0 - 2$. Define $\mathbf{P}_{k,\max}^{m_0}$ to be the maximal bitiles in the set

$$\{P \in \mathbf{P}_{\text{stock}} : I_P \subset E_k^{m_0}, \text{mass}(P) \geq 2^{-k}\}.$$

Also let

$$\mathbf{P}_k^{m_0} = \{P \in \mathbf{P}_{\text{stock}} : P \leq P' \text{ for some } P' \in \mathbf{P}_{k,\max}^{m_0}\},$$

$$N_k^{m_0}(x) = \sum_{P \in \mathbf{P}_{k,\max}^{m_0}} 1_{I_P}(x),$$

and the exceptional set

$$E_k^{m_0+1} = \{x : N_k^{m_0}(x) \geq C_1 k 2^k\}.$$

Finally, define

$$\mathbf{Q}_k^{m_0+1} = \{P \in \mathbf{P}_k^{m_0} : I_P \not\subset E_k^{m_0+1}\}.$$

Note that (23)–(24) hold when $m = m_0$, for the same reasons as before. To check (25) for $m = m_0 - 1$, let $P \in \mathbf{Q}_k^{m_0}$. Since $I_P \not\subset E_k^{m_0}$, we have that either $I_P \cap E_k^{m_0} = \emptyset$, or $I_P \cap E_k^{m_0}$ is a finite disjoint union of dyadic intervals $J \in \mathbf{J}_k^{m_0}$. Note that John-Nirenberg’s inequality guarantees that, for each $J \in \mathbf{J}_k^{m_0}$ we have

$$|J \cap E_k^{m_0+1}| \leq e^{-10k} |J|,$$

which immediately implies (25) for $m = m_0 - 1$.

It is easy to see that each \mathbf{Q}_k^m is convex. It can be organized into a forest of trees with tops in

$$\left\{ P \in \mathbf{P}_{k,\max}^{m-1} : I_P \not\subset E_k^m \right\}.$$

Note that the counting function has a favorable L^∞ bound

$$\left\| \sum_{P \in \mathbf{P}_{k,\max}^{m-1} : I_P \not\subset E_k^m} 1_{I_P} \right\|_\infty \leq C_1 k 2^k.$$

The forests $\mathcal{F}_k^{m,n}$ are now obtained via the Fefferman trick.

It is immediate that the collections \mathbf{Q}_k^m are pairwise disjoint for fixed k , because of (24). Note that the algorithm (for fixed k) will end with a finite value of m . We set \mathbf{Q}_k^m to be empty for all larger values of m and define $\mathbf{P}_k = \bigcup_{m \geq 1} \mathbf{Q}_k^m$. Finally, observe that, when the algorithm for fixed k is over, there cannot be any P left in $\in \mathbf{P}_{\text{stock}} \setminus \mathbf{P}_k$ with $\text{mass}(P) \geq 2^{-k}$. Indeed, note that $I_P \subset E_k^0$ trivially. While $I_P \subset E_k^m$, the algorithm will continue to run. But, by (23), there should exist an m such that $I_P \subset E_k^m, I_P \not\subset E_k^{m+1}$. Thus, if $\text{mass}(P) \geq 2^{-k}$, then P gets automatically selected in \mathbf{Q}_k^{m+1} . \square

A distinct feature of the approach in this section is the almost orthogonality between a function with small support and a function locally constant on this support.

Lemma 7.2 (Almost orthogonality). *Let \mathbf{J} be a collection of pairwise disjoint intervals, and let $f : \cup_{J \in \mathbf{J}} J \rightarrow \mathbf{C}$ such that $|f|$ is constant on each $J \in \mathbf{J}$. Assume that $E_J \subset J$ satisfies $|E_J| \leq \alpha|J|$ for each $J \in \mathbf{J}$. Then, for each $g : \cup_{J \in \mathbf{J}} E_J \rightarrow \mathbf{C}$, we have*

$$|\langle f, g \rangle| \leq \alpha^{1/2} \|f\|_2 \|g\|_2.$$

Proof.

$$\begin{aligned} |\langle f, g \rangle| &\leq \sum_J \int_{E_J} |f| |g| \\ &\leq \sum_J |E_J|^{1/2} \sup_{x \in J} |f(x)| \|g\|_{L^2(J)} \\ &\leq \alpha^{1/2} \sum_J \|f\|_{L^2(J)} \|g\|_{L^2(J)} \\ &\leq \alpha^{1/2} \|f\|_2 \|g\|_2. \quad \square \end{aligned}$$

Proposition 7.3 (Almost orthogonality between forests). *Fix k . For each $m \geq 0$, let \mathcal{F}^m denote one of the forests $\mathcal{F}_k^{m,n}$ from Lemma 7.1. Define*

$$\Pi_{\mathcal{F}^m} f = \sum_{T \in \mathcal{F}^m} \Pi_T f.$$

Then, for each $m' \geq m + 2$,

$$|\langle \Pi_{\mathcal{F}^m} f, \Pi_{\mathcal{F}^{m'}} f \rangle| \lesssim e^{-(m'-m-2)k/2} \|f\|_2^2.$$

Proof. Note that, due to (21), we can split \mathcal{F}^m into layers $\mathcal{F}^m = \bigcup_{l \lesssim k2^k} \mathcal{F}^m(l)$, so that for each l , the trees T in $\mathcal{F}^m(l)$ have pairwise disjoint intervals I_T .

Arguing as in the proof of Proposition 3.4 we first observe that, for each x , we have $\sum_{T \in \mathcal{F}^m(l)} \Pi_T f(x) = \Pi_P f(x)$, where P is the minimal

bitile in $\cup_{T \in \mathcal{F}^m(l)} T$ such that $x \in I_P$. Second, by (24), for each interval $J \in \mathbf{J}_k^m$ and each $P \in \cup_{T \in \mathcal{F}^m(l)} T$, J cannot contain either the right or the left halves of I_P . Third, (1) guarantees that each $|\Pi_P f|$ is constant on both the right and the left halves of I_P . We now conclude that the function $|\sum_{T \in \mathcal{F}^m(l)} \Pi_T f|$ is constant on each $J \in \mathbf{J}_k^m$. Applying Lemma 7.2 to $(\sum_{T \in \mathcal{F}^m(l)} \Pi_T f)1_{E_k^m}$ and $\Pi_{\mathcal{F}^{m'}} f$, by virtue of (23), we get

$$\begin{aligned} |\langle \Pi_{\mathcal{F}^m} f, \Pi_{\mathcal{F}^{m'}} f \rangle| &\leq \sum_l |\langle \Pi_{\mathcal{F}^m(l)} f, \Pi_{\mathcal{F}^{m'}} f \rangle| \\ &\lesssim k2^k e^{-10(m'-m-1)k/2} \|\Pi_{\mathcal{F}^{m'}} f\|_2 \|\Pi_{\mathcal{F}^m} f\|_2 \\ &\lesssim e^{-(m'-m-2)k/2} \|f\|_2^2. \quad \square \end{aligned}$$

Lemma 7.4 (Schur's test). *Let T_m be a sequence of operators on $L^2([0, 1])$ with adjoints T_m^* such that, for each m, m' ,*

$$\|T_{m'} T_m^*\|_{2 \rightarrow 2} \leq c_{|m'-m|}$$

where

$$\sum_{n \geq 0} c_n = c < \infty.$$

Then

$$\sum_m \|T_m f\|_2^2 \leq c \|f\|_2^2.$$

Proof. We have

$$\begin{aligned} \sum_m \|T_m f\|_2^2 &= \sum_m \langle T_m^* T_m f, f \rangle \\ &\leq \left\| \sum_m T_m^* T_m f \right\|_2 \|f\|_2 \\ &= \left(\sum_{m, m'} \langle T_m^* T_m f, T_{m'}^* T_{m'} f \rangle \right)^{1/2} \|f\|_2 \\ &\leq \left(\sum_{m, m'} \|T_{m'} T_m^*\|_{2 \rightarrow 2} \|T_m f\|_2 \|T_{m'} f\|_2 \right)^{1/2} \|f\|_2. \end{aligned}$$

It now suffices to note that this can further be bounded via Cauchy-Schwartz by

$$c^{1/2} \left(\sum_m \|T_m f\|_2^2 \right)^{1/2} \|f\|_2. \quad \square$$

Corollary 7.5. *With the notation in Proposition 7.3, we have, for each $2 \leq p < \infty$,*

$$\sum_m \|\Pi_{\mathcal{F}^m} f\|_p^p \lesssim (k2^k)^{p-2} \|f\|_p^p.$$

Proof. To get the $p = 2$ case apply Schur's test to the families of operators $T_m = \Pi_{\mathcal{F}^{3m+i}}$ with $i \in \{0, 1, 2\}$. Note that $\Pi_{\mathcal{F}}^* f = \overline{\Pi_{\mathcal{F}}(\overline{f})}$ for each forest \mathcal{F} .

We next get an L^∞ bound. Since by (21) each forest consists of $O(k2^k)$ layers, Proposition 3.4 implies the bound $\|\Pi_{\mathcal{F}^m} f\|_\infty \lesssim k2^k \|f\|_\infty$. Then we use interpolation [3] for the vector-valued operator $Vf = (\Pi_{\mathcal{F}^m} f)_{m \geq 1}$ as in the proof of Proposition 5.8. \square

Lemma 7.6. *With the notation in Proposition 7.3, we have for each $2 \leq p < \infty$,*

$$\sum_m \|C_{\mathcal{F}^m} f\|_p^p \lesssim 2^{-k} \|f\|_p^p.$$

Proof. Using that \mathcal{F}^m is a Fefferman forest and the tree estimate Proposition 5.5 we get

$$\|C_{\mathcal{F}^m} f\|_p^p = \sum_{T \in \mathcal{F}^m} \|C_T f\|_p^p \lesssim 2^{-k} \sum_{T \in \mathcal{F}^m} \|\Pi_T f\|_p^p.$$

Consider the vector-valued operator $Vf = (\Pi_T f)_{T \in \cup_m \mathcal{F}^m}$. Note that

$$\|Vf\|_{L^2(\ell^2)} = \left(\sum_m \|\Pi_{\mathcal{F}^m} f\|_2^2 \right)^{1/2} \lesssim \|f\|_2,$$

by Corollary 7.5, and also that

$$\|Vf\|_{L^\infty(I^\infty)} = \sup_{T \in \cup_m \mathcal{F}^m} \|\Pi_T f\|_\infty \lesssim \|f\|_\infty,$$

by virtue of Proposition 3.4. Interpolation [3] finishes the proof. \square

The proof of (17) when $p \geq 2$ will now immediately follow from the next estimate, by invoking the triangle inequality.

Proposition 7.7. *Let $2 \leq p < \infty$. For each $f \in L^p([0, 1])$, we have*

$$\|C_{\mathbf{P}_k} f\|_p \lesssim 2^{-Lk} \|f\|_p,$$

for some $L = L(p) > 0$ independent of k .

Proof. Fix k . We can assume p is an integer and then use interpolation. The value of L will change throughout the argument. For each $m \geq 0$, let \mathcal{F}^m denote one of the forests $\mathcal{F}_k^{m,n}$ from Lemma 7.1. By the triangle inequality, it will suffice to prove that

$$\sum_m \|C_{\mathcal{F}^m} f\|_p^p + \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ p^2 + m_1 < m_p}} \int |C_{\mathcal{F}^{m_1}} f| \prod_{i=2}^p |C_{\mathcal{F}^{m_i}} f| \lesssim 2^{-Lk} \|f\|_p^p.$$

Note that the first term is taken care of by Lemma 7.6. We next focus on the second term. The restriction $p^2 + m_1 < m_p$ in the summation can be achieved by splitting the integers in classes of residues modulo p^2 and using the triangle inequality. This separation will be used to achieve extra decay.

For each $T \in \mathcal{F}^{m_1}$, let, as before, \mathbf{J}_T be the collection of all maximal dyadic intervals $J \subset I_T$ such that J contains no I_P with $P \in T$. Denote by S_T the support

$$\bigcup_{P \in T} (I_P \cap N^{-1}(\omega_P))$$

of $C_T f$. Recall that the sets S_T are pairwise disjoint.

Since, for each $P \in T$, we have $I_P \subset E_k^{m_1-1}$, (23)–(25) will imply that

$$|I_P \cap (\text{supp } C_{\mathcal{F}^{m_p}} f)| \leq e^{-(m_p - m_1 - 1)k} |I_P|.$$

Since the dyadic parent of J must equal I_P for some $P \in T$, we conclude also that

$$(26) \quad |J \cap (\text{supp } C_{\mathcal{F}^{m_p}} f)| \leq e^{-(m_p - m_1 - 1)k} |J|.$$

Thus,

$$\begin{aligned} & \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} \int |C_{\mathcal{F}^{m_1}} f| \prod_{i=2}^p |C_{\mathcal{F}^{m_i}} f| \\ & \leq \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} \sum_{T \in \mathcal{F}^{m_1}} \sum_{J \in \mathbf{J}_T} \int_J |C_T f| \prod_{i=2}^p |C_{\mathcal{F}^{m_i}} f| \\ & = \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} \sum_{T \in \mathcal{F}^{m_1}} \sum_{J \in \mathbf{J}_T} \int_{J \cap (\text{supp } C_{\mathcal{F}^{m_p}} f)} |C_T f| \prod_{i=2}^p |1_{S_T} C_{\mathcal{F}^{m_i}} f|. \end{aligned}$$

By invoking (13), (26), Lemma 3.2 and Hölder's inequality this can further be bounded by

$$\begin{aligned} & \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} \sum_{T \in \mathcal{F}^{m_1}} \sum_{J \in \mathbf{J}_T} |J \cap (\text{supp } C_{\mathcal{F}^{m_p}} f)|^{1/p} \\ & \quad \times \inf_{x \in J} M(V_T f)(x) \prod_{i=2}^p \left(\int_J |1_{S_T} C_{\mathcal{F}^{m_i}} f|^p \right)^{1/p} \\ & \leq \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} e^{-(m_p - m_1 - 1)k/p} \sum_{T \in \mathcal{F}^{m_1}} \sum_{J \in \mathbf{J}_T} \left(\int_J [M(V_T f)]^p \right)^{1/p} \\ & \quad \times \prod_{i=2}^p \left(\int_J |1_{S_T} C_{\mathcal{F}^{m_i}} f|^p \right)^{1/p} \\ & \lesssim \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} e^{-(m_p - m_1 - 1)k/p} \sum_{T \in \mathcal{F}^{m_1}} \|\Pi_T f\|_p \\ & \quad \times \prod_{i=2}^p \left(\int |1_{S_T} C_{\mathcal{F}^{m_i}} f|^p \right)^{1/p} \end{aligned}$$

$$\leq \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} e^{-(m_p - m_1 - 1)k/p} \left(\sum_{T \in \mathcal{F}^{m_1}} \|\Pi_T f\|_p^p \right)^{1/p} \prod_{i=2}^p \|C_{\mathcal{F}^{m_i}} f\|_p.$$

In the last line we have used the orthogonality of $\Pi_T f$ and the disjointness of S_T .

Using (15), the forest estimate in Proposition 5.8, Hölder’s inequality and the fact that

$$C_{\mathcal{F}^m} f = C_{\mathcal{F}^m} (\Pi_{\mathcal{F}^m} f),$$

we upgrade the last estimate to

$$\begin{aligned} & \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} \int |C_{\mathcal{F}^{m_1}} f| \prod_{i=2}^p |C_{\mathcal{F}^{m_i}} f| \\ & \lesssim 2^{-[k(p-1)]/p} \sum_{\substack{m_1 \leq m_2 \leq \dots \leq m_p \\ m_1 + p^2 < m_p}} e^{-(m_p - m_1 - 1)k/p} \prod_{i=1}^p \|\Pi_{\mathcal{F}^{m_i}} f\|_p \\ & \lesssim e^{-kp} \sum_m \|\Pi_{\mathcal{F}^m} f\|_p^p. \end{aligned}$$

Proposition 7.7 now follows from Corollary 7.5. □

8. A proof without any appeal to the choice function Let ψ, ϕ be Schwartz functions supported on $\mathbf{T} = [-1/2, 1/2)$ such that $\int \psi = 1, \int \phi = 0$. Define the kernel

$$K(t, s) = \sum_{k < 0} \psi_k(t) \phi_k(s),$$

where $\psi_k(t) = 2^{-k} \psi(t2^{-k})$ and $\phi_k(s) = 2^{-k} \phi(s2^{-k})$. Consider the trilinear form

$$\Lambda(F_1, F_2, F_3) = \int_{\mathbf{T}^4} F_1(x+t, y+s) F_2(x+s, y) F_3(x, y) K(t, s) \, dx \, dy \, dt \, ds,$$

where $+$ denotes the summation modulo one on the torus. This is an example of a dualized two dimensional bilinear Hilbert transform.

Motivated in part by questions from Ergodic Theory, in [9] it is proved that

$$(27) \quad |\Lambda(F_1, F_2, F_3)| \lesssim \|F_1\|_{p_1} \|F_2\|_{p_2} \|F_3\|_{p_3},$$

whenever $1/p_1 + 1/p_2 + 1/p_3 = 1$ and $2 < p_i < \infty$.

Interestingly, this implies the $L^p, p > 2$ boundedness of the Carleson operator defined as

$$Cf(x) = \sup_{N \in \mathbf{N}} \left| \int_{\mathbf{T}} f(x+s) e^{iNs} \frac{ds}{s} \right|,$$

and, in particular, the almost everywhere convergence of the Fourier series $S_n f(x)$ in the $p > 2$ regime. To see this, it suffices to apply (27) to $F_1(x, y) = f(y)$, $F_2(x, y) = e^{ixN(y)}g(y)$, $F_3(x, y) = e^{-ixN(y)}h(y)$ with $\|g\|_{p_2} = \|h\|_{p_3} = 1$ and to an appropriate ϕ such that $\sum_{k < 0} \phi_k(s) = 1/s$ for $s \in [-1/4, 1/4] \setminus \{0\}$.

We will prove the Walsh model analogue of (27) for the dyadic kernel

$$K^W(t, s) = \sum_{k \geq 0} 2^{3k/2} 1_{I_k}(t) h_{I_k}(s),$$

where $I_k = [0, 2^{-k}]$ and h_{I_k} is the L^2 normalized Haar function. Now define

$$\begin{aligned} & \Lambda^W(F_1, F_2, F_3) \\ &= \int_{[0,1]^4} F_1(x \oplus t, y \oplus s) F_2(x \oplus s, y) F_3(x, y) K^W(t, s) dx dy dt ds. \end{aligned}$$

Theorem 8.1.

$$|\Lambda^W(F_1, F_2, F_3)| \lesssim \|F_1\|_{p_1} \|F_2\|_{p_2} \|F_3\|_{p_3},$$

whenever $1/p_1 + 1/p_2 + 1/p_3 = 1$ and $2 < p_i < \infty$.

The definition of $\Lambda(F_1, F_2, F_3)$ involves no choice function $N(x)$, so we recover a proof of the boundedness of the Carleson operator which

makes no mention of it! The proof of Theorem 8.1 is very close in spirit to the proof of the boundedness of the Walsh model of the bilinear Hilbert transform [16]. This shows once more the deep connection between these two operators.

Define the projection operators

$$\begin{aligned} \pi_\omega^1 F(x, y) &= \mathcal{F}_W^{-1}[1_\omega(\xi)\mathcal{F}_W(F)(\xi, \eta)](x, y) \\ \pi_\omega^2 F(x, y) &= \mathcal{F}_W^{-1}[1_\omega(\eta)\mathcal{F}_W(F)(\xi, \eta)](x, y). \end{aligned}$$

It can be easily checked that

$$\begin{aligned} \Lambda^W(F_1, F_2, F_3) &= \int_{[0,1]^2} \sum_{\omega \in \mathcal{D}_+ : |\omega| \geq 2} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 \\ &\quad \times F_1(x, y) \pi_{\omega_u}^1 F_2(x, y) \pi_{\omega_u}^1 F_3(x, y) \, dx \, dy \\ &\quad + \int_{[0,1]^2} \sum_{\omega \in \mathcal{D}_+ : |\omega| \geq 2} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_u}^2 \\ &\quad \times F_1(x, y) \pi_{\omega_l}^1 F_2(x, y) \pi_{\omega_l}^1 F_3(x, y) \, dx \, dy, \end{aligned}$$

where ω_l, ω_u denote the left and right halves of ω .

By symmetry, it will suffice to estimate the first integral, which we denote by $\Lambda^{W,1}(F_1, F_2, F_3)$. Let Ω be a collection of intervals in \mathcal{D}_+ . Define

$$\begin{aligned} \Lambda_\Omega(F_1, F_2, F_3) &= \int_{[0,1]^2} \sum_{\omega \in \Omega} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 F_1(x, y) \pi_{\omega_u}^1 F_2(x, y) \pi_{\omega_u}^1 F_3(x, y) \, dx \, dy. \end{aligned}$$

The first case of interest is when the intervals in Ω are nested. This will be a precursor of the tree estimate in Lemma 8.6.

Proposition 8.2. *Let Ω be a collection of intervals in \mathcal{D}_+ which contain a point ξ . Then*

$$|\Lambda_\Omega(F_1, F_2, F_3)| \lesssim \|F_1\|_{p_1} \|F_2\|_{p_2} \|F_3\|_{p_3},$$

whenever $1/p_1 + 1/p_2 + 1/p_3 = 1$ and $1 < p_i < \infty$.

Proof. By splitting Ω in two parts, it suffices to assume that either $\xi \in \omega_l$ for all ω or $\xi \in \omega_u$ for all ω . In the first case, note that the intervals ω_u are pairwise disjoint. We estimate using Hölder's inequality and the boundedness of the square function S associated with the intervals ω_u on the first component

$$\begin{aligned} |\Lambda_\Omega(F_1, F_2, F_3)| &\lesssim \int_{[0,1]^2} \sup_\omega |\pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 F_1(x, y)| \\ &\quad \times \left(\sum_{\omega \in \Omega} |\pi_{\omega_u}^1 F_2(x, y)|^2 \right)^{1/2} \\ &\quad \times \left(\sum_{\omega \in \Omega} |\pi_{\omega_u}^1 F_3(x, y)|^2 \right)^{1/2} dx dy \\ &\leq \|M(F_1)\|_{p_1} \|SF_2\|_{p_2} \|SF_3\|_{p_3} \\ &\lesssim \|F_1\|_{p_1} \|F_2\|_{p_2} \|F_3\|_{p_3}. \end{aligned}$$

In the second case, note that ω_l are pairwise disjoint. We will run a standard telescoping argument. Denote by $\tilde{\Omega}$ the collection of all set differences between intervals ω_u of consecutive length. Note that each $\tilde{\omega} = \omega'_u \setminus \omega_u \in \tilde{\Omega}$ is the union of at most two intervals whose length is at least $|\omega_u|$. We call ω^{ext} the complement (in \mathbf{R}_+) of the largest interval ω_u , $\omega \in \Omega$, and $F_i^{\text{ext}} = \pi_{\omega^{\text{ext}}}^1 F_i$ for $i = 2, 3$. Note that, trivially, for each $\omega \in \Omega$,

$$\pi_{\omega_u}^1 F_i = F_i^{\text{ext}} - \sum_{\tilde{\omega} \in \tilde{\Omega}: \tilde{\omega} \not\subset \omega_u} \pi_{\tilde{\omega}}^1 F_i, \quad i = 2, 3.$$

We can now write

$$\Lambda_\Omega(F_1, F_2, F_3)$$

(28)

$$= \int_{[0,1]^2} \sum_{\omega \in \Omega} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 F_1(x, y) F_2^{\text{ext}}(x, y) F_3^{\text{ext}}(x, y) dx dy$$

(29)

$$- \int_{[0,1]^2} \sum_{\omega \in \Omega} \sum_{\tilde{\omega} \in \tilde{\Omega}: \tilde{\omega} \not\subset \omega_u} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 F_1(x, y) F_2^{\text{ext}}(x, y) \pi_{\tilde{\omega}}^1 F_3(x, y) dx dy$$

$$(30) \quad - \int_{[0,1]^2} \sum_{\omega \in \Omega} \sum_{\tilde{\omega} \in \tilde{\Omega}: \tilde{\omega} \not\sim \omega_u} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 F_1(x, y) \pi_{\tilde{\omega}}^1 F_2(x, y) F_3^{\text{ext}}(x, y) dx dy$$

$$(31) \quad + \int_{[0,1]^2} \sum_{\omega \in \Omega} \sum_{\tilde{\omega} \in \tilde{\Omega}: \tilde{\omega} \not\sim \omega_u} \sum_{\tilde{\omega}' \in \tilde{\Omega}: \tilde{\omega}' \not\sim \omega_u} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 F_1(x, y) \pi_{\tilde{\omega}}^1 F_2(x, y) \pi_{\tilde{\omega}'}^1 F_3(x, y) dx dy.$$

The term (28) is controlled by Hölder’s inequality and the boundedness of the two-dimensional singular integral operator $TF = \sum_{\omega \in \Omega} \pi_{[0,|\omega|/2]}^1 \times \pi_{\omega_l}^2 F$. To control (31) we note that the term corresponding to a triple $\omega, \tilde{\omega}, \tilde{\omega}'$ is nonzero only when $\tilde{\omega}, \tilde{\omega}'$ are adjacent to each other. This is a simple computation involving the Walsh transform that will diagonalize the sum. Then use Cauchy-Schwartz and the boundedness of the square function $(\sum_{\tilde{\omega} \in \tilde{\Omega}} |\pi_{\tilde{\omega}}^1 F|^2)^{1/2}$ and of the maximal truncation operator

$$T^*F = \sup_{\delta > 0} \left| \sum_{\omega \in \Omega: |\omega| > \delta} \pi_{[0,|\omega|/2]}^1 \pi_{\omega_l}^2 F \right|.$$

A similar argument will take care of terms (29) and (30); the details are left to the reader. \square

The next step is to discretize the form Λ in both time and frequency.

Definition 8.3. A one and a half dimensional bitile (tile) $P = R_P \times \omega_P$, for short 3/2-bitile (tile), is a product of a dyadic square $R_P = I_P \times J_P \in \mathcal{D}_+^2$ and a dyadic interval $\omega_P \in \mathcal{D}_+$ with $|\omega_P| = 2|I_P|^{-1}$ ($|\omega_P| = |I_P|^{-1}$). If P is a 3/2-bitile, we define as before the upper and lower 3/2-tiles $P_u = R_P \times \omega_{P_u}, P_l = R_P \times \omega_{P_l}$.

Call $\mathbf{P}_{\text{all}}^{3/2}$ the entire collection of 3/2-bitiles. There is a natural partial order on $\mathbf{P}_{\text{all}}^{3/2}$ which, by abusing earlier notation will be denoted by \leq . The reader can correctly anticipate that $P \leq P'$ will stand for $R_P \subset R_{P'}$ and $\omega_{P'} \subset \omega_P$. Convexity on $\mathbf{P}_{\text{all}}^{3/2}$ will be understood with respect to this order.

Definition 8.4. Let $R_T \subset \mathbf{D}_+^2$ be a dyadic square, and let $\xi_T \in \mathbf{R}_+$. A 3/2-tree T with top data (R_T, ξ_T) is a collection of 3/2-bitiles in $\mathbf{P}_{\text{all}}^{3/2}$ such that $R_P \subset R_T$ and $\xi_T \in \omega_P$ for each $P \in T$.

With each 3/2-bitile P we will identify two regions in the phase space $\mathbf{R}_+^4 = \mathbf{R}_+^2 \times \widehat{\mathbf{R}_+^2}$. One is the region covered by the two-dimensional bitile $R_P \times [0, |\omega_P|] \times \omega_P$. The other one is $R_P \times \omega_P \times \mathbf{R}_+$, which is the union of two dimensional bitiles of the form $R_P \times \omega_P \times \omega$. We will denote by \mathcal{C}_P the collection of all these two-dimensional bitiles. It is easy to check that if $\mathbf{P} \subset \mathbf{P}_{\text{all}}^{3/2}$ is convex, then both collections of bitiles $\{R_P \times [0, |\omega_P|] \times \omega_P : P \in \mathbf{P}\}$ and $\cup_{P \in \mathbf{P}} \mathcal{C}_P$ are convex with respect to the two-dimensional order. We will denote by $\Pi_{\mathbf{P}}^2 F$ and $\Pi_{\mathbf{P}}^{3/2} F$ the phase space projections associated with the two collections (cf. Remark 2.5).

Note that $\Pi_{\mathbf{P}}^{3/2} F$ can be thought of as being a 1-1/2-dimensional projection, since this operator produces no localization on the second frequency component. For example, if p is a 3/2-tile, then

$$\begin{aligned} \Pi_p^{3/2} F(x, y) &= \left[\int F(x', y) W_{I_p \times \omega_p}(x') dx' \right] W_{I_p \times \omega_p}(x) 1_{J_p}(y) \\ &= \sum_{|\omega|=|\omega_p|} \langle F, W_{I_p \times \omega_p} W_{J_p \times \omega} \rangle W_{I_p \times \omega_p}(x) W_{J_p \times \omega}(y). \end{aligned}$$

For a convex 3/2-tree T , define

$$\begin{aligned} \Lambda_T(F_1, F_2, F_3) &= \int_{[0,1]^2} \sum_{P \in T} 1_{R_P}(x, y) \pi_{[0, |\omega_P|/2]}^1 \pi_{\omega_{P_l}}^2 \\ &\quad \times F_1(x, y) \pi_{\omega_{P_u}}^1 F_2(x, y) \pi_{\omega_{P_u}}^1 F_3(x, y) dx dy, \end{aligned}$$

and observe that variants of (6) will imply that

$$\Lambda_T(F_1, F_2, F_3) = \Lambda_T(\Pi_T^2(F_1), \Pi_T^{3/2}(F_2), \Pi_T^{3/2}(F_3)).$$

An argument very similar to the one in Proposition 8.2 will prove

$$|\Lambda_T(F_1, F_2, F_3)| \lesssim \|F_1\|_{s_1} \|F_2\|_{s_2} \|F_3\|_{s_3},$$

whenever $1/s_1 + 1/s_2 + 1/s_3 = 1$ and $1 < s_i < \infty$. Combining these, we get

$$(32) \quad |\Lambda_T(F_1, F_2, F_3)| \lesssim \|\Pi_T^2(F_1)\|_{s_1} \|\Pi_T^{3/2}(F_2)\|_{s_2} \|\Pi_T^{3/2}(F_3)\|_{s_3}.$$

Definition 8.5. Let $\mathbf{P} \subset \mathbf{P}_{\text{all}}^{3/2}$ be a collection of 3/2-bitiles. For $F : \mathbf{R}_+^2 \rightarrow \mathbf{C}$, define

$$\begin{aligned} \text{size}_{F,2}(\mathbf{P}) &= \sup_{P \in \mathbf{P}} \frac{\|\Pi_P^2 F\|_2}{|R_P|^{1/2}} \\ \text{size}_{F,3/2}(\mathbf{P}) &= \sup_{P \in \mathbf{P}} \frac{\|\Pi_P^{3/2} F\|_2}{|R_P|^{1/2}}. \end{aligned}$$

Lemma 8.6 (Tree estimate). *Let $T \subset \mathbf{P}_{\text{all}}^{3/2}$ be a convex 3/2-tree. Then*

$$\|\Pi_T^2 F\|_p \lesssim |R_T|^{1/p} \text{size}_{F,2}(T), \quad \text{for each } 1 \leq p \leq \infty.$$

If, in addition, $\|F\|_\infty \leq 1$, then

$$\|\Pi_T^{3/2} F\|_p \lesssim |R_T|^{1/p} (\text{size}_{F,3/2}(T))^{2/p}, \quad \text{for each } 2 \leq p \leq \infty.$$

Proof. The crucial observation in both cases is that

$$(33) \quad \Pi_T^i F(x, y) = \Pi_P^i F(x, y),$$

for each $i \in \{3/2, 2\}$, where P is the minimal 3/2-bitile in T such that $(x, y) \in R_P$. This follows as in the proof of Proposition 3.4. The first inequality then follows by noting that $\|\Pi_P^2 F\|_\infty \lesssim \|\Pi_P^2 F\|_2$, as in the one-dimensional case.

Let us now analyze the second inequality. Let \mathbf{P}_T be the collection of all 3/2-bitiles in T such that, for each $P \in \mathbf{P}_T$, there is (x, y) such that P is the minimal 3/2-bitile in T with $(x, y) \in R_P$. It is easy to see that

$$(34) \quad \sum_{P \in \mathbf{P}_T} |R_P| \lesssim |R_T|.$$

Indeed, if $P \in \mathbf{P}_T$, by the convexity of T at least one of the four dyadic children of R_P will contain no $R_{P'}$ with $P' \in \mathbf{P}_T$.

Next we observe that, for each 3/2-bitile P ,

$$(35) \quad \|\Pi_P^{3/2} F\|_\infty \leq \|F\|_\infty,$$

which together with Hölder's inequality gives, if $\|F\|_\infty \leq 1$ and $2 \leq p \leq \infty$

$$\|\Pi_P^{3/2} F\|_p \leq |R_P|^{1/p} (\text{size}_{F,3/2}(P))^{2/p}.$$

The result now follows by combining this with (33) and (34). \square

The proof of Proposition 4.6 applies with essentially no modification to prove the following variant.

Proposition 8.7 (Size decomposition). *Let i be either 2 or 3/2. Let $F : \mathbf{R}_+^2 \rightarrow \mathbf{C}$. Let \mathbf{P} be a finite convex collection of 3/2-bitiles.*

Then

$$\mathbf{P} = \bigcup_{2^{-n} \leq \text{size}_{F,i}(\mathbf{P})} \mathbf{P}_n \cup \mathbf{P}_{\text{null}},$$

such that

- $\text{size}_{F,i}(\mathbf{P}_n) \leq 2^{-n}$;
- \mathbf{P}_n is a convex forest with convex 3/2-trees $T \in \mathcal{F}_n$ satisfying

$$\sum_{T \in \mathcal{F}_n} |R_T| \lesssim 2^{2n} \|F\|_2^2;$$

- $\Pi_P^i F \equiv 0$ for each $P \in \mathbf{P}_{\text{null}}$.

Proof of Theorem 8.1. By multilinear restricted type interpolation it suffices to prove

$$|\Lambda^{W,1}(F_1, F_2, F_3)| \lesssim |E_1|^{1/p_1} |E_2|^{1/p_2} |E_3|^{1/p_3}$$

whenever $|F_i| \leq 1_{E_i}$ for finite measure subsets E_i of \mathbf{R}_+^2 . Fix $2 < p_i < \infty$ for the rest of the proof. We begin by observing that

$$\Lambda^{W,1}(F_1, F_2, F_3) = \sum_{P \in \mathbf{P}_{\text{all}}^{3/2}} \Lambda_P(F_1, F_2, F_3),$$

and thus it further suffices to prove

$$\left| \sum_{P \in \mathbf{P}} \Lambda_P(F_1, F_2, F_3) \right| \lesssim |E_1|^{1/p_1} |E_2|^{1/p_2} |E_3|^{1/p_3},$$

for all convex, finite $\mathbf{P} \subset \mathbf{P}_{\text{all}}^{3/2}$.

Note that by (35) and the natural two-dimensional extension of Proposition 4.3, we have

$$\text{size}_{F_1,2}(\mathbf{P}), \text{size}_{F_2,3/2}(\mathbf{P}), \text{size}_{F_3,3/2}(\mathbf{P}) \lesssim 1.$$

Let $\mathbf{P}_n^{(1)}$, $\mathbf{P}_n^{(2)}$ and $\mathbf{P}_n^{(3)}$ be the collections provided by Proposition 8.7, relative to $\text{size}_{F_1,2}(\mathbf{P})$, $\text{size}_{F_2,3/2}(\mathbf{P})$ and $\text{size}_{F_3,3/2}(\mathbf{P})$, respectively. Define $\mathbf{P}_{n_1, n_2, n_3} = \mathbf{P}_{n_1}^{(1)} \cap \mathbf{P}_{n_2}^{(2)} \cap \mathbf{P}_{n_3}^{(3)}$. Note that

$$(36) \quad \sum_{P \in \mathbf{P}} \Lambda_P(F_1, F_2, F_3) = \sum_{2^{-n_1}, 2^{-n_2}, 2^{-n_3} \lesssim 1} \sum_{P \in \mathbf{P}_{n_1, n_2, n_3}} \Lambda_P(F_1, F_2, F_3).$$

Reasoning as in the proof of Theorem 2.6 from Section 6, we organize $\mathbf{P}_{n_1, n_2, n_3}$ as a forest in three different ways, with the L^1 norm of the counting function of the tops bounded by $2^{2n_1}|E_1|$, $2^{2n_2}|E_2|$ and $2^{2n_3}|E_3|$, respectively.

Now pick $1 > \alpha > \max\{2/p_2, 2/p_3\}$. Using (36) and (32) with $s_2 = s_3 = 2/\alpha$, along with Lemma 8.6, we get

$$\begin{aligned} \left| \sum_{P \in \mathbf{P}} \Lambda_P(F_1, F_2, F_3) \right| &= \sum_{2^{-n_1}, 2^{-n_2}, 2^{-n_3} \lesssim 1} 2^{-n_1 - n_2\alpha - n_3\alpha} \\ &\quad \times \min(2^{2n_1}|E_1|, 2^{2n_2}|E_2|, 2^{2n_3}|E_3|) \\ &\leq \sum_{2^{-n_1}, 2^{-n_2}, 2^{-n_3} \lesssim 1} 2^{-n_1 - n_2\alpha - n_3\alpha} (2^{2n_1}|E_1|)^{1/p_1} \\ &\quad \times (2^{2n_2}|E_2|)^{1/p_2} (2^{2n_3}|E_3|)^{1/p_3} \\ &\lesssim |E_1|^{1/p_1} |E_2|^{1/p_2} |E_3|^{1/p_3}. \end{aligned}$$

The argument is now complete. \square

Acknowledgments. The author would like to thank Christoph Thiele for clarifying discussions on the argument from [18] and to

Fangye Shi for carefully reading the original manuscript and pointing out a few typos. Part of the material in this paper was organized while teaching a class on harmonic analysis. The author is grateful to his students Francesco Di Plinio and Prabath Silva for motivating him and to his collaborators for enriching his understanding of time-frequency analysis.

REFERENCES

1. M. Bateman, *Single annulus L^p estimates for Hilbert transforms along vector fields*, preprint.
2. M. Bateman and C. Thiele, *L^p estimates for the Hilbert transforms along a one variable vector field*, preprint.
3. A. Benedek and R. Panzone, *The space L^p , with mixed norm*, Duke Math. J. **28** (1961), 301–324.
4. P. Billard, *Sur la convergence presque partout des séries de Fourier-Walsh des fonctions de l'espace $L^2(0, 1)$* , Stud. Math. **28** (1966/1967), 363–388.
5. L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 137–157.
6. C. Demeter, *Singular integrals along N directions in \mathbf{R}^2* , Proc. Amer. Math. Soc. **138** (2010), 4433–4442.
7. C. Demeter and F. Di Plinio, *Logarithmic L^p bounds for maximal directional singular integrals in the plane*, J. Geomet. Anal. **24** (2014), 375–416.
8. C. Demeter, M. Lacey, T. Tao and C. Thiele, *Breaking the duality in the return times theorem*, Duke Math. J. **143** (2008), 281–355.
9. C. Demeter and C. Thiele, *On the two-dimensional bilinear Hilbert transform*, Amer. J. Math. **132** (2010), 201–256.
10. C. Fefferman, *Pointwise convergence of Fourier series*, Ann. Math. **98** (1973), 551–571.
11. Richard A. Hunt, *On the convergence of Fourier series*, 1968 *Orthogonal expansions and their continuous analogues*, (Proc. Conf., Edwardsville, Ill., 1967), Southern Ill. Univ. Press, Carbondale, IL.
12. M. Lacey, *Carleson's theorem: Proof, complements, variations*, Publ. Mat. **48** (2004), 251–307.
13. M. Lacey and X. Li, *On a conjecture of E.M. Stein on the Hilbert transform on vector fields*, Mem. Amer. Math. Soc. **205** (2010), viii+72 pp.
14. ———, *Maximal theorems for the directional Hilbert transform on the plane*, Trans. Amer. Math. Soc. **358** (2006), 4099–4117.
15. M. Lacey and C. Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. **7** (2000), 361–370.
16. ———, *L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$* , Ann. Math. **146** (1997), 693–724.

17. V. Lie, *The (weak- L^2) boundedness of the quadratic Carleson operator*, *Geom. Funct. Anal.* **19** (2009), 457–497.
18. ———, *The polynomial Carleson operator*, preprint.
19. C. Muscalu, T. Tao and C. Thiele, *A Carleson theorem for a Cantor group model of the scattering transform*, *Nonlinearity* **16** (2003), 219–246.
20. E.M. Stein, *On limits of sequences of operators*, *Ann. Math.* **74** (1961), 140–170.
21. T. Tao, *Lecture notes for harmonic analysis in the phase plane*, available at <http://www.math.ucla.edu/tao>.
22. C. Thiele, *The quartile operator and pointwise convergence of Walsh series*, *Trans. Amer. Math. Soc.* **352** (2000), 5745–5766.

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY, BLOOMINGTON, IN 47405
Email address: demeterc@indiana.edu