STUDIA MATHEMATICA 211 (2) (2012)

Weighted bounds for variational Fourier series

by

YEN DO (New Haven, CT) and MICHAEL LACEY (Atlanta, GA)

Abstract. For $1 < p < \infty$ and for weight w in A_p, we show that the r-variation of the Fourier sums of any function f in $L^p(w)$ is finite a.e. for r larger than a finite constant depending on w and p. The fact that the variation exponent depends on w is necessary. This strengthens previous work of Hunt–Young and is a weighted extension of a variational Carleson theorem of Oberlin–Seeger–Tao–Thiele–Wright. The proof uses weighted adaptation of phase plane analysis and a weighted extension of a variational inequality of Lépingle.

1. Introduction. For a measurable function f on $[0,1]$, let Sf denote the maximal Fourier sum:

$$Sf(x) := \sup_n |(S_n f)(x)|, \qquad S_n f(x) := \sum_{|k|<n} \widehat{f}(k) e^{i2\pi kx}.$$

Here, $\widehat{f}(k) = \int_0^1 f(x) e^{-i2\pi kx}\,dx$ is the kth Fourier coefficient, and by convention, $S_n f = 0$ for $n \le 0$. (We use strict inequality $|k| < n$ in the definition of S_n for the convenience of the transference argument in Section 1.2.)

By the Carleson–Hunt theorem [C, H], S is bounded on L^p for $1 < p < \infty$, which leads to a.e. convergence of the Fourier series of functions in L^p. See also Sjölin [S] for the Walsh case, and [F, LT2] for alternative proofs. More quantitative information about the convergence rate of Fourier series has been obtained by Oberlin–Seeger–Tao–Thiele–Wright [OST+], via bounds on a strengthening of S. To formulate this strengthening of S, we first recall the r-variation norm of a sequence $(a_n)_{n\in\mathbb{Z}}$. If $0 < r < \infty$ then

$$\|(a_n)\|_{V^r} := \sup_{M,\, N_0<\cdots<N_M} \Big[|a_{N_0}|^r + \sum_{j=1}^{M} |a_{N_j} - a_{N_{j-1}}|^r \Big]^{1/r},$$

and for $r = \infty$ we have $\|(a_n)\|_{V^\infty} = \sup_n |a_n|$. It is clear that if $\|(a_n)\|_{V^r}$ is finite for some $r < \infty$ then (a_n) is a Cauchy sequence and therefore is convergent; the finiteness of $\|a\|_{V^r}$ may be considered as a quantitative

2010 *Mathematics Subject Classification*: Primary 42B20; Secondary 42B25, 42B35.
Key words and phrases: weights, Carleson, pointwise convergence, Fourier series, variation, Lépingle inequality.

DOI: 10.4064/sm211-2-4

measurement of the convergence rate of (a_n). The variational strengthening of S considered in [OST+] is the following operator:

$$S_{[r]}f(x) = \sup_{M,\, N_0<\cdots<N_M} \Big[\sum_{j=1}^{M} |S_{N_j}f(x) - S_{N_{j-1}}f(x)|^r\Big]^{1/r}, \tag{1.1}$$

and it was shown in [OST+] that, for $1 < p < \infty$, $S_{[r]}$ is bounded in $L^p([0,1])$ if $r > \max(2, p')$.

Convergence of Fourier series in non-Lebesgue settings was also considered in Hunt–Young [HY], where it was shown that S is bounded on $L^p(w)$ for any A_p weight w, $1 < p < \infty$. See also [GMS] for extensions to more generalized settings. Recall that a positive a.e. weight w is in A_p if uniformly over intervals I we have

$$[w]_{A_p} := \sup_I \frac{1}{|I|}\Big[\int_I w(x)\,dx\Big]\Big[\frac{1}{|I|}\int_I w(x)^{-1/(p-1)}\,dx\Big]^{p-1} < \infty.$$

Our aim in this paper is to strengthen the results of [HY] and [OST+] by considering weighted estimates for $S_{[r]}$.

THEOREM 1.1. *Let $1 < p < \infty$ and $w \in A_p$. Then there is an $R = R(p, [w]_{A_p}) < \infty$ such that for all $r \in (R, \infty]$ we have*

$$\|S_{[r]}f\|_{L^p([0,1],w)} \le C\|f\|_{L^p([0,1],w)} \tag{1.2}$$

for some constant C depending only on w, p, r.

As remarked above, Theorem 1.1 gives more quantitative information about the convergence of Fourier series than [HY] (which corresponds to the endpoint $r = \infty$). Theorem 1.1 follows from

THEOREM 1.2. *Let $1 < p < \infty$ and $w \in A_q$ for some $q \in [1, p)$. Then for $r > \max(2q, pq/(p-q))$ we have*

$$\|S_{[r]}f\|_{L^p([0,1],w)} \le C\|f\|_{L^p([0,1],w)} \tag{1.3}$$

for some constant C depending only on w, p, q, r.

To derive Theorem 1.1 from Theorem 1.2, let $1 < p < \infty$ and $w \in A_p$. Since the A_p condition is an open condition, we have $w \in A_q$ for some $1 < q < p$ (see e.g. [L2]). Then (1.2) follows from applying Theorem 1.2.

We would like to point out that, in the conclusion of Theorem 1.1, the variation exponent must depend upon $w \in A_p$. Indeed, suppose towards a contradiction that there is some $p \in (1, \infty)$ such that (1.2) holds for every $w \in A_p$ and for fixed $r \in (0, \infty)$. Using the fact that the variation-norm decreases as r increases, we may assume that $r > 1$. Then $S_{[r]}$ is sublinear, and an application of the Rubio de Francia extrapolation theorem shows that the same inequality (with the same r) would have to hold for w being the Lebesgue measure and *all* $p \in (1, \infty)$, contradicting an example in [OST+, Section 2]. We also remark that in the Lebesgue setting, when $w \equiv 1 \in A_1$, the range of r in Theorem 1.2 is sharp.

Our proof of Theorem 1.2 extends our previous work in [DL] on a Walsh–Fourier model of $S_{[r]}$ and at the same time is a weighted extension of [OST+]. The proof uses two new ingredients: weighted analysis on the Fourier phase plane, and a weighted extension of a classical variational inequality of Lépingle (Lemma 5.2). The weighted adaptation of analysis on the Fourier phase plane in our proof follows closely the adaptation in [DL], modulo (substantial) technicalities arising from the lack of perfect localization of Fourier wave packets. In particular, our approach is different from the elegant argument in [HY] where a good-λ argument was used to deduce weighted bounds for S from the Carleson–Hunt theorem. It is not hard to see that a naive adaptation of the good-λ approach in [HY] does not apply to the variation-norm Carleson operator. Our approach is inspired by an argument of Rubio de Francia [RdF], though it is easier to see this inspiration in the dyadic setting of [DL]. We anticipate that the weighted phase plane analysis in our proof will be useful in a variety of open problems involving weighted bounds for multilinear operators with oscillatory nature, where a naive adaptation of the approach in [HY] seems not applicable ([1]). It is interesting to compare our paper with that of Bennett–Harrison [BH].

1.1. Notational conventions. (i) Henceforth, we work on the real line $\mathbb{R}$, and set $\widehat{f}(\xi) = \int f(x)e^{-i2\pi x\xi}\,dx$.

(ii) For any $1 \le t < \infty$ we will denote by $\mathcal{M}_t f$ the L^t Hardy–Littlewood maximal function, and by $\mathcal{M}_{t,w} f$ the weighted L^t maximal function

$$\mathcal{M}_{t,w} f(x) = \sup_{I:\, x\in I} \left(\frac{1}{w(I)} \int_I |f(x)|^t w(x)\,dx\right)^{1/t}.$$

(iii) The dyadic intervals $\mathbf{D}$ will play a distinguished role. We denote by $f^\sharp$ the dyadic sharp maximal function of f, namely

$$f^\sharp(x) := \sup_{I\in\mathbf{D}} 1_I(x)|I|^{-1} \int_I \Big| f - |I|^{-1}\int_I f(y)\,dy\Big|\,dx.$$

All BMO norms, unless otherwise specified, are dyadic BMO norms, namely $\|f\|_{\mathrm{BMO}} = \|f^\sharp\|_\infty$. An important inequality for this paper is the familiar estimate

$$\|\phi\|_{L^p(w)} \simeq \|\phi^\sharp\|_{L^p(w)}, \qquad w \in A_p. \tag{1.4}$$

(iv) For any interval I and $c > 0$ we denote by cI the interval with length $c|I|$ and with the same center as I. This should not be confused with $c(I)$

([1]) We would like to point out that Xiaochun Li [L2] has some unpublished results about weighted estimates for the bilinear Hilbert transform.

which will denote the center of I. A standard property of an A_p weight w is that it is doubling: there exists $\gamma=\gamma(w)$ such that for any interval I and any $k\geq 0$,

$$w(2^kI)\leq 2^{\gamma k}w(I). \tag{1.5}$$

(v) For any set G we denote $w(G)=\int_G w(x)\,dx$.

1.2. Transference to a singular integral form. Using a weighted variant of a transference argument in [OST+, Appendix A], it is not hard to see that Theorem 1.2 follows from Theorem 1.3 stated below. Here, we define

$$C_{[r]}f(x):=\sup_{K,\,N_0<\cdots<N_K}\Big(\sum_{j=1}^{K}\Big|\int_{N_{j-1}}^{N_j}\widehat{f}(\xi)e^{i2\pi x\xi}\,d\xi\Big|^r\Big)^{1/r}. \tag{1.6}$$

Theorem 1.3. *Let $1<p<\infty$ and $w\in A_q$ for some $q\in[1,p)$. Then for $r>\max\big(2q,\frac{pq}{p-q}\big)$ we have*

$$\|C_{[r]}f\|_{L^p(\mathbb{R},w)}\leq C\|f\|_{L^p(\mathbb{R},w)} \tag{1.7}$$

for some constant C depending only on w, p, q, r.

For the reader's convenience, we include details of the transference argument.

For any $K\geq 1$ and $m\geq 1$, let $I_{m,K}$ be the set of all non-decreasing sequences of length $K+1$ in $\{0,\dots,m\}$. For each such sequence $\vec{N}=(N_0\leq\cdots\leq N_K)$ we construct the variation sum

$$S_{\vec{N}}f=\Big(\sum_{j=1}^{K}|S_{N_j}f-S_{N_{j-1}}f|^r\Big)^{1/r}. \tag{1.8}$$

Since the set $I_{m,K}$ is bigger when m or K is larger, by two applications of the monotone convergence theorem it suffices to show that

$$\Big\|\sup_{\vec{N}\in I_{m,K}}S_{\vec{N}}f\Big\|_{L^p([0,1],w)}\leq C\|f\|_{L^p([0,1],w)},$$

where the implicit constant is uniform over m and K. Let $\sigma=w^{1-p'}$. Then the above inequality has the following equivalent dual form: for f defined on $[0,1]$ and for g defined on $[0,1]\times I_m\times\{1,\dots,K\}$ (we will write $g_{\vec{N},j}(x)$ to denote $g(x,\vec{N},j)$),

$$\begin{aligned}\int_0^1 f(x)\sum_{\vec{N}\in I_{m,K}}\sum_{j=1}^{K}&[(S_{N_j}-S_{N_{j-1}})g_{\vec{N},j}](x)\,dx\\ &\leq C\|f\|_{L^p([0,1],w)}\Big\|\sum_{\vec{N}\in I_{m,K}}\Big(\sum_{j=1}^{K}|g_{\vec{N},j}|^{r'}\Big)^{1/r'}\Big\|_{L^{p'}([0,1],\sigma)}.\end{aligned} \tag{1.9}$$

To prove (1.9), we may assume without loss of generality that f and $g_{\vec{N},j}$ are trigonometric polynomials for any $\vec{N} \in I_m$ and $1 \le j \le K$.

For any $N \ge 0$ let C_N be the Fourier multiplier operator on $L^2(\mathbb{R})$ whose symbol is the characteristic function of $\{-(N-1/3) \le \xi \le N-1/3\}$ (by definition $C_N \equiv 0$ if $N < 1/3$). Let $\delta(x) = e^{-\pi x^2}$ and $\delta_M(x) = \delta(x/M)$.

By standard transference theory (see e.g. [SW, p. 261]), for any integer N and any 1-periodic trigonometric polynomials P, Q we have

$$\int_0^1 P(x) S_N Q(x)\, dx = \lim_{M\to\infty} \frac{1}{M} \int_{\mathbb{R}} P(x)\delta_{M/\alpha} C_N(\delta_{M/\beta} Q)(x)\, dx,$$

for any $\alpha, \beta \in (0,1)$ such that $\alpha^2 + \beta^2 = 1$. We take $\alpha = \beta = 1/\sqrt{2}$. It follows that the left hand side of (1.9) is the same as

$$\lim_{M\to\infty} \frac{1}{M} \int_{\mathbb{R}} f(x)\delta_{M/\alpha}(x) \sum_{\vec{N}\in I_m} \sum_{j=1}^{K} [(C_{N_j} - C_{N_{j-1}})(\delta_{M/\beta} g_{\vec{N},j})](x)\, dx.$$

It follows from Theorem 1.3 that the analogue of (1.9) for C_N's holds, thus the above limit is bounded above by

$$C \limsup_{M\to\infty} \frac{1}{M} \|f\delta_{M/\alpha}\|_{L^p(\mathbb{R},w)} \Big\| \delta_{M/\beta} \sum_{\vec{N}\in I_{m,K}} \Big(\sum_{j=1}^{K} |g_{\vec{N},j}|^{r'}\Big)^{1/r'} \Big\|_{L^{p'}(\mathbb{R},\sigma)}. \tag{1.10}$$

Since $w \in A_q \subset A_p$, we have $\sigma = w^{1-p'} \in A_{p'}$ and in particular both w and σ are doubling weights. On the other hand, it follows from exponential decay of δ that, for any doubling measure μ and any $1 < q < \infty$ and any 1-periodic function h,

$$\sup_{M\ge 1} \frac{1}{M^{1/q}} \|\delta_M h\|_{L^q(\mathbb{R},\mu)} \le C\|h\|_{L^q([0,1],\mu)}.$$

In view of this observation, (1.9) follows immediately from (1.10).

We take up the proof of Theorem 1.3 below.

2. Discretization. In this section we reduce the task of proving (1.7) to proving similar bounds on model operators. Consider absolute constants $C_2 \in [1,\infty)$ and $C_3 \in (0, C_2)$ and $C_{2,1}, C_{2,2}, C_1 \in [C_2, \infty)$. Constants with these properties are called *admissible.*

2.1. Tiles and bitiles. In this paper, a *tile* is a dyadic rectangle of area 1, which we will write $p = I_p \times \omega_p$ and refer to I_p as the spatial interval and ω_p as the frequency interval of p. By a *bitile* P we mean a rectangle $I_P \times \omega_P$ that contains (as subsets) two tiles P_1 and P_2 such that they share the same (dyadic) spatial interval I_P and

$$\sup C_2\omega_{P_1} \le \inf C_2\omega_{P_2}, \qquad |\omega_P| \le C_1(|\omega_{P_1}| + |\omega_{P_2}|),$$
$$\omega_P = \text{convex hull}(C_{2,1}\omega_{P_1} \cup C_{2,2}\omega_{P_2}).$$

The classical setting (see e.g. [LT2]) when a bitile is a dyadic rectangle of area 2 is the special case of our general setting when $C_2 = C_{2,1} = C_{2,2} = C_1 = 1$.

We say that two bitiles P and P' are *disjoint* if they are disjoint in the phase plane. Denote by $\widetilde{\omega}_P$ the convex hull of $C_2\omega_{P_1} \cup C_2\omega_{P_2}$, clearly $\widetilde{\omega}_P \subset \omega_P$. In this paper, whenever we talk about a bitile collection it shall be assumed that the implicit constants above are the same for any two bitiles.

2.2. Fourier wave packets. For every tile $p = I_p \times \omega_p$, a function ϕ_p is called a *Fourier packet adapted to* p if $\text{supp}(\widehat{\phi_p}) \subset C_3\omega_p$, furthermore for any $N > 0$ and $n \ge 0$ we have (for some $C_{N,n}$ depending only on N and n)

$$\left|\frac{d^n}{dx^n}\phi_p(x)\right| \le C_{N,n}\frac{1}{|I_p|^{1/2+n}}\left(1 + \frac{|x - c(I_p)|}{|I_p|}\right)^{-(N+n)}; \tag{2.1}$$

here recall that $c(I_p)$ denotes the center of I_p. In a family of Fourier packets, we will assume that the implicit constants involved are uniform.

2.3. Discretization and the model operators. For any $r \in [1,\infty)$ and any finite collection $\mathbf{P}$ of bitiles, let

$$C_{r,\mathbf{P}}f := \sup_{K,\, N_0<\cdots<N_K}\Big(\sum_{j=1}^{K}\Big|\sum_{P\in\mathbf{P}}\langle f, \phi_{P_1}\rangle\phi_{P_1}1_{\{N_{j-1}\notin\omega_P,\, N_j\in\omega_{P_2}\}}\Big|^r\Big)^{1/r}.$$

A symmetric variant of $C_{r,\mathbf{P}}$ can be obtained by changing the limiting condition involving N_j, N_{j-1} in the above definition to $\{N_{j-1} \in \omega_{P_1},\, N_j \notin \omega_P\}$.

Without loss of generality, we assume in the rest of the paper that $2q < r < \infty$ and $q \in (1,\infty)$. Via a discretization argument in [OST+], which we summarize below, Theorem 1.3 follows from the theorem below and its symmetric variant (whose proof is completely analogous).

THEOREM 2.1. *There is a constant $C < \infty$ independent of f and $\mathbf{P}$ such that*

$$\|C_{r,\mathbf{P}}f\|_{L^p(w)} \le C\|f\|_{L^p(w)} \tag{2.2}$$

for any finite collection $\mathbf{P}$ of bitiles and any $p \in (q,\infty)$ such that $1/p < 1/q - 1/r$.

Discretization. We sketch the main ideas of our weighted adaptation of the discretization argument in [OST+, Section 3]. For each interval (a, b)

with non-dyadic endpoints, let $\mathbf{J}$ be the collection of maximal dyadic intervals in (a,b) such that $\operatorname{dist}(J,a), \operatorname{dist}(J,b) \geq |J|$. It is not hard to see that $\mathbf{J}$ partitions (a,b), and the ratio between two adjacent elements of J is at most 2. By direct examination, it follows that there are $O(1)$ possible mutually exclusive scenarios involving relative locations of J inside (a,b), and these scenarios are characterized by the following information:

- whether J is the left or right child or its dyadic parent,
- the distance from a to J, which could be arbitrarily large,
- the distance from b to J, which could be arbitrarily large.

More specifically, we may divide $\mathbf{J}$ into $O(1)$ disjoint subsets of the following type: If m,n,k are bounded positive integers and "side" is "left" or "right" then we denote by $\mathbf{J}_{k,m,n,\mathrm{side}}$ the set of all dyadic intervals J such that J is the "side" child of its dyadic parent, and $a \in J_{\mathrm{low}}(k,m)$ and $b \in J_{\mathrm{high}}(k,n)$.

- If $k=1$ then $J_{\mathrm{low}} = J-(m+1)|J|$ and $J_{\mathrm{high}} = |J|+(n+1)|J|$.
- If $k=2$ then $J_{\mathrm{low}} = J-(m+1)|J|$ and $J_{\mathrm{high}} = [\sup J + n|J|, \infty)$.
- If $k=3$ then $J_{\mathrm{low}} = (-\infty, \inf J - m|J|]$ and $J_{\mathrm{high}} = |J|+(n+1)|J|$.

The following example of such a partition was given in [OST+]; we include this example for the convenience of the reader. Below are the values of (k,m,n,side):

$$\{(1,2,1,\mathrm{left}), (1,2,2,\mathrm{left}), (1,3,1,\mathrm{left}), (1,3,2,\mathrm{left}), (2,1,1,\mathrm{left}),\\ (2,1,1,\mathrm{right}), (2,2,1,\mathrm{right}), (3,4,1,\mathrm{left}), (3,3,1,\mathrm{right}), (3,4,2,\mathrm{left})\}.$$

Since the relative ratios between adjacent intervals in $\mathbf{J}$ are bounded by 2, we may construct nonnegative L^∞ normalized bump functions φ_J such that $1_{(a,b)}(\xi) = \sum_{J\in\mathbf{J}} \varphi_J(\xi)$, furthermore φ_J is supported inside a $1+c$ dilation of J for each $J \in \mathbf{J}$, where the absolute constant $c>0$ can be taken arbitrarily small. By using a standard Fourier sampling theorem for the Schwartz band-limited function $\mathcal{F}^{-1}(\widehat{f}(\xi)\sqrt{\varphi_J})$ (cf. [T2]) we can easily decompose

$$\widehat{f}(\xi)\varphi_J(\xi) = \sum_{|I|=1/(2^L|J|)} \langle f, \phi_{I\times J}\rangle \widehat{\phi}_{I\times J}(\xi)$$

for some positive integer $L = O(1)$ where $\widehat{\phi}_{I\times J}(\xi) := |I|^{1/2}\sqrt{\varphi_J(\xi)}e^{-2\pi i c(I)\xi}$. Note that the frequency support of $\phi_{I\times J}$ is inside a $1+c$ dilation of J with $c>0$ that can be chosen small. Furthermore, it is clear that the collections of functions $(\phi_{I\times J} : |I| = 2^{-L}|J|^{-1})$ can be decomposed ([2]) into $O(1)$ families of Fourier wave packets adapted to the tiles in the phase plane.

([2]) This decomposition ensures that there is only one wave packet associated with each dyadic rectangle of area 1.

Let $\mathbf{P}_{\text{side}}$ denote the collection of all dyadic rectangles of area 2^{-L} whose frequency interval is the "side" child of its parent. Then

$$\int_a^b e^{i2\pi x\xi}\widehat{f}(\xi)\,d\xi = \sum_{k=1}^{3}\sum_{m,n,\text{side}}\sum_{p\in\mathbf{P}_{\text{side}}}\langle f,\phi_p\rangle\phi_p(x)1_{\{a\in L_p(k,m),\,b\in U_p(k,n)\}},$$

where the intervals $L_p(k,m)$ and $U_p(k,n)$ are the J_{low} and J_{high} of $J=\omega_p$.

Now, under the assumption that f is Schwartz, it is no loss of generality to assume that the sequence $(N_0<\cdots<N_K)$ (used in the definition of $C_{[r]}$) does not contain endpoints of dyadic intervals. Performing the above partition on every (N_{j-1},N_j), it then follows from the triangle inequality that

$$C_{[r]}f \le \sum_{m,n,\text{side}} C_{1,m,n,\text{side}}f(x)+C_{2,m,n,\text{side}}f(x)+C_{3,m,n,\text{side}}f(x)$$

with

$$C_{k,m,n,\text{side}}f(x) := \sup_{K,(N_j)}\Big(\sum_{j=1}^{K}\Big|\sum_{p\in\mathbf{P}_{\text{side}}}\langle f,\phi_p\rangle\phi_p(x)1_{\{N_{j-1}\in L_p(k,m),\,N_j\in U_p(k,m)\}}\Big|^r\Big)^{1/r}$$

for $k=1,2,3$. It is not hard to see that for each $1\le m,n=O(1)$, we can bound $C_{3,m,n}f(x)$ by a sum of $O_L(1)$ operators of the same nature as $C_{r,\mathbf{P}}$, with appropriate choice of admissible constants C_1, C_2, $C_{2,1}$, $C_{2,2}$ and C_3. Similarly, $C_{2,m,n}f(x)$ can be bounded by a symmetric variant of $C_{r,\mathbf{P}}$. Since any interval $[a,b)$ can be written as $(-\infty,b)\setminus(-\infty,a)$, it is not hard to see that $C_{1,m,n}f(x)$ can be controlled by two operators of the same nature as $C_{3,m,n}f(x)$. Thus, Theorem 1.3 follows from Theorem 2.1. This completes the discretization step. ∎

Below we set up a linearized variant of $C_{r,\mathbf{P}}$. By duality in ℓ^r, to show (2.2) it suffices consider the following operator (we omit the dependence on r for simplicity):

$$(C_{\mathbf{P}}f)(x) = \sum_{j=1}^{K(x)}\sum_{P\in\mathbf{P}}\langle f,\phi_{P_1}\rangle\phi_{P_1}(x)1_{\{N_{j-1}(x)\notin\omega_P,\,N_j(x)\in\omega_{P_2}\}}d_j(x);$$

here $K:\mathbb{R}\to\mathbb{Z}_+$, $N_0(x)<\cdots<N_K(x)$ and $\{d_j\}$ are measurable functions, with

$$|d_1(x)|^{r'}+\cdots+|d_{K(x)}(x)|^{r'}=1.$$

For each bitile P, let $d_P(x)$ be 0 unless there exists a (clearly unique) j such that $N_{j-1}(x)\notin\omega_P$ and $N_j(x)\in\omega_{P_2}$, in which case we set $d_P(x)=d_j(x)$. For a function g, we note that $\langle C_{\mathbf{P}}f, gw\rangle = B_{\mathbf{P}}(f,g)$, where

$$B_{\mathbf{P}}(f,g) := \sum_{P\in\mathbf{P}}\langle f,\phi_{P_1}\rangle\langle\phi_{P_1}\,d_P,\,gw\rangle.$$

We say that $G' \subset G$ is a *major subset* if $w(G') > w(G)/2$ and we say G' has *full measure* if $w(G') = w(G)$. Via a standard restricted weak-type interpolation argument [MTT1, Section 2], Theorem 2.1 follows from the following proposition:

PROPOSITION 2.2. *Let F, G be such that $w(F), w(G) < \infty$. Then there are major subsets of F and G, denoted respectively by $\widetilde{F}$ and $\widetilde{G}$, such that:*

(i) *at least one subset has full measure, and*
(ii) *for any $|f| \leq 1_{\widetilde{F}}$ and $|g| \leq 1_{\widetilde{G}}$ and any finite collection of bitiles $\mathbf{P}$ we have*

$$B_{\mathbf{P}}(f, g) \leq C w(F)^{1/p} w(G)^{1-1/p} \tag{2.3}$$

for all $p \in (q, \infty)$ such that $1/p < 1/q - 1/r$.

In the rest of the paper, we will prove Proposition 2.2.

3. Decomposition of bitile collections. Without loss of generality we may assume the following separation conditions:

(S1) The ratio $\mathrm{dist}(\omega_{P_1}, \omega_{P_2})/|\omega_{P_1}|$ is constant over $P \in \mathbf{P}$
(S2) For any two bitiles P and P', if $\omega_P \cap \omega_{P'} \neq \emptyset$ and $|I_P| = |I_{P'}|$ then $\omega_P = \omega_{P'}$.
(S3) For any two bitiles P and P', if $|I_P| > |I_{P'}|$ then $|\omega_P| < |\omega_{P'_1}|/K_0$ for some large absolute constant K_0 that will be chosen in the proof. (The choice of K_0 is refined a bounded number of times below.)

REMARK 3.1. First, we will require that $K_0 > 2/(C_2 - C_3)$. This means that for any $1 \leq i \leq 2$, if $C_3\omega_{P_i} \cap C_3\omega_{P'_i} \neq \emptyset$ and $|I_P| > |I_{P'}|$ then $\omega_P \subset C_2\omega_{P'_i}$.

3.1. Trees. In this paper, a finite collection T of bitiles is a *tree* if there exists a dyadic interval I_T and a real number ξ_T such that for any $P \in T$ we have

$$I_P \subset I_T \quad \text{and} \quad \omega_T := \left[\xi_T - \frac{1}{2|I_T|}, \xi_T + \frac{1}{2|I_T|}\right) \subset \widetilde{\omega}_P.$$

We will refer to I_T as the *top interval* of T. Similarly, ξ_T and ω_T will be referred to as the *top frequency* and the *top frequency interval* of T.

We say that T is 2-*overlapping* if $\xi_T \in C_2\omega_{P_2}$ for every $P \in T$, and we say that T is 2-*lacunary* if $\xi_T \notin C_2\omega_{P_2}$ for every $P \in T$.

It is clear that any tree can be split into two trees, one of each type. Furthermore, the union of two trees with the same (I_T, ξ_T) is a tree and we may use the pair (I_T, ξ_T) for the new tree. If these two trees are 2-lacunary then the new tree is also 2-lacunary.

Remark 3.2. By further requiring that $K_0 > C_3/(2C_1+1)$ in the separation assumption (S3), we obtain the following properties (cf. Remark 3.1). Let T be a tree and let $P, P' \in T$ be two different bitiles.

- If $|I_P| = |I_{P'}|$ then $I_P \cap I_{P'} = \emptyset$.
- If T is 2-overlapping and $|I_P| > |I_{P'}|$ then $\omega_P \cap C_3\omega_{P'_1} = \emptyset$.
- If T is 2-lacunary and $|I_P| > |I_{P'}|$ then $\omega_P \cap C_3\omega_{P'_2} = \emptyset$.

Remark 3.3. If there is a dyadic interval J such that for every $P \in \mathbf{T}$ we have $I_P \subset J$ then we can decompose T into $O(1)$ subtrees, each tree has J as top interval (the top frequencies of these subtrees are not necessarily the same, but they are $O(1/|J|)$ away from the original ξ_T). Essentially, this is because we would have $|\widetilde{\omega}_P| \geq 2/|J|$ and then one can always partition T into two desired trees depending on the relative position of ξ_T in $\widetilde{\omega}_P$.

3.2. Tile norms. Below, for any collection $\mathbf{Q}$ of bitiles we denote

$$S_{\mathbf{Q}} f(x) := \left[\sum_{P \in \mathbf{Q}} \frac{|\langle f, \phi_{P_1}\rangle|^2}{|I_P|} 1_{I_P}\right]^{1/2}.$$

Definition 3.4 (Size). The *size* of a collection $\mathbf{P}$ of bitiles is

$$\operatorname{size}(\mathbf{P}) := \sup_{T \subset \mathbf{P}} w(I_T)^{-1/2} \|S_T f\|_{L^2(w)}.$$

The supremum is over all 2-overlapping trees $T \subset \mathbf{P}$.

It is clear that for $w \equiv 1$ one recovers the standard definition of size (cf. [LT1]). For any interval I, let

$$\widetilde{\chi}_I(x) = \left[1 + \left(\frac{x - c(I)}{|I|}\right)^2\right]^{-1/2}.$$

Note that if $J \subset I$ then $\widetilde{\chi}_J \leq C\widetilde{\chi}_I$, and this estimate will be used implicitly in future estimates.

Definition 3.5 (Density). Recall the definition of the functions d_j from (2.3). Fix a large constant $D \in (0,\infty)$. The *density* of a collection $\mathbf{P}$ of bitiles is defined to be

$$\operatorname{density}(\mathbf{P}) := \sup_T \left(\frac{1}{w(I_T)} \int \widetilde{\chi}_{I_T}^D |g|^{r'} \sum_{j: N_j \in \omega_T} |d_j|^{r'} w\right)^{1/r'},$$

where the supremum is over nonempty trees $T \subset \mathbf{P}$.

Choose D to be very large depending on w, p, q, r in the proof of Proposition 2.2 in Section 6 (see also the proof of Lemma 3.15). All the implicit constants are allowed to depend on D.

When the elements of $\mathbf{P}$ are disjoint in the phase plane, the following improved notion of density is more useful in future estimates; see also Lemma 4.2.

Definition 3.6 (Improved density). The *improved density* of a collection $\mathbf{P}$ of bitiles is defined to be

$$\widetilde{\mathrm{density}}(\mathbf{P}) := \sup_{P\in\mathbf{P}} \left(\frac{1}{w(I_P)}\int \widetilde{\chi}_{I_P}^{D}|g|^{r'} \sum_{j\colon N_j\in\omega_{P_2}} |d_j|^{r'} w\right)^{1/r'}.$$

It is clear that $\widetilde{\mathrm{density}}(\mathbf{P}) \le C\,\mathrm{density}(\mathbf{P})$ for any $\mathbf{P}$.

3.3. Decomposition by size. We have the following size bound:

Lemma 3.7. *Assume $w \in A_q$. Then for any $N > 0$ there is a constant $C = C(N,q,w) < \infty$ such that for any $\mathbf{P}$,*

$$\mathrm{size}(\mathbf{P}) \le C \sup_{P\in\mathbf{P}} \left(\frac{1}{w(I_P)}\int |f|^q \widetilde{\chi}_{I_P}^{N} w\right)^{1/q}.$$

The main ingredient in the proof of Lemma 3.7 is the following John–Nirenberg characterization of size, which is a standard result in the Lebesgue setting (see e.g. [MTT3]). The proof of the Lebesgue case of this characterization extends smoothly to the weighted setting (see [DL, Lemma 3.5]); we omit the details.

Lemma 3.8. *For any $1 < p < \infty$ and any collection $\mathbf{P}$ we have*

$$\sup_{T\subset\mathbf{P}} \frac{1}{w(I_T)^{1/p}}\|S_T f\|_{L^p(w)} \sim_p \sup_{T\subset\mathbf{P}} \frac{1}{w(I_T)}\|S_T f\|_{L^{1,\infty}(w)},$$

the suprema being over all 2-overlapping trees.

Proof of Lemma 3.7 using Lemma 3.8. By decomposing T into smaller subtrees (using Remark 3.3), we may assume that $I_T = I_P$ for some $P \in T$. Thus, it suffices to show that

$$\|S_T f\|_{L^q(w)} \le C\|f\widetilde{\chi}_{I_T}^{N}\|_{L^q(w)}.$$

But $w \in A_q$, hence $\|S_T f\|_{L^q(w)} \lesssim \|(S_T f)^\sharp\|_{L^q(w)}$. Therefore it suffices to show that for any $N < \infty$ we have

$$(S_T f)^\sharp \le C\mathcal{M}_1(f\widetilde{\chi}_{I_T}^{N}). \tag{3.1}$$

For any dyadic interval J let

$$c_J = \left(\sum_{P\in T: J\subset I_P} \frac{|\langle f,\phi_{P_1}\rangle|^2}{|I_P|}\right)^{1/2}.$$

Then

$$\frac{1}{|J|}\int_J |S_T f(x) - c_J|\,dx \le \left(\frac{1}{|J|}\int_J |S_T f(x)^2 - c_J^2|\,dx\right)^{1/2}$$
$$= \frac{1}{|J|^{1/2}}\bigg\|\bigg(\sum_{P\in T:\, I_P\subsetneq J} |\langle f,\phi_{P_1}\rangle|^2 \frac{1_{I_P}}{|I_P|}\bigg)^{1/2}\bigg\|_2.$$

Using the known Lebesgue case of Lemma 3.7 (see e.g. [MTT3, Lemma 6.8]), we obtain

$$\frac{1}{|J|}\int_J |S_T f(x) - c_J|\,dx \le C \sup_{P\in T:\, I_P\subsetneq J} \frac{1}{|I_P|}\int |f(x)|\widetilde{\chi}_{I_P}(x)^{N+4}\,dx$$
$$\le C \inf_{x\in J\cap I_T} \mathcal{M}_1(f\widetilde{\chi}_{I_T}^N)(x),$$

and (3.1) follows immediately. ∎

We remark that the following bound was proved in the above proof of Lemma 3.7:

Corollary 3.9. *Assume $w \in A_q$. Then for any 2-overlapping tree T and any $N > 0$ we have*

$$\|S_T f\|_{\mathrm{BMO}} \le C_N \inf_{x\in I_T} \mathcal{M}_1(f\widetilde{\chi}_{I_T}^N)(x);$$

here we use the dyadic BMO norm.

For convenience, in the rest of the paper we say that a collection $\mathbf{T}$ of 2-overlapping trees is *well-separated* if the following conditions are satisfied:

(i) If $T, T' \in \mathbf{T}$ are two different trees, and $P \in T$ and $P' \in T'$ and $|I_P| > |I_{P'}|$, then either $C_3\omega_{P_1} \cap C_3\omega_{P'_1} = \emptyset$ or $I_{P'} \cap I_T = \emptyset$.

(ii) If $P, P' \in \bigcup_{T\in\mathbf{T}} T$ are two different bitiles with $|I_P| = |I_{P'}|$ then $I_P \times C_3\omega_{P_1}$ and $I_{P'} \times C_3\omega_{P'_1}$ are disjoint.

Lemma 3.10. *Let $\mathbf{P}$ be a collection of bitiles with size bounded above by 2α for some $\alpha > 0$. Then we can find a collection $\mathbf{T}$ of trees such that:*

- *The bitile collection $\mathbf{P} - \bigcup_{T\in\mathbf{T}} T$ has size less than α.*
- *If another tree collection $\mathbf{T}'$ covers $\bigcup_{T\in\mathbf{T}} T$ then for some $C = C(w) < \infty$ we have*

$$\sum_{T\in\mathbf{T}} w(I_T) \le C \sum_{T'\in\mathbf{T}'} w(I_{T'}). \tag{3.2}$$

- *If $q_0 \in (q,\infty)$ then there exists $\beta = \beta(p,w,q,q_0) < \infty$ such that for any $k \ge 0$ and for any $1 \le p < \infty$ we have*

$$\Big\|\sum_{T\in\mathbf{T}} 1_{2^k T}\Big\|_{L^p(w)} \le C2^{\beta k}\alpha^{-2q_0}\|f\|_{L^{2pq_0}(w)}^{2q_0}. \tag{3.3}$$

Here $C = C(p,w,q,q_0) < \infty$.

Proof. For convenience let $a_P = \langle f, \phi_{P_1}\rangle$. We follow the standard algorithm from [LT2]. If $\mathrm{size}(\mathbf{P}) \geq \alpha$ then there exists a nonempty 2-overlapping tree $T_2 \subset \mathbf{P}$ such that $\|S_{T_2} f\|_{L^2(w)} \geq \alpha^2 w(I_{T_2})$. We select such a tree with minimal value of ξ_{T_2} (3), and let T be the maximal tree in $\mathbf{P}$ with top data (I_{T_2}, ξ_{T_2}). We then remove from $\mathbf{P}$ the bitiles in T and repeat this argument until the remaining collection of bitiles has size less than α. We obtain a collection $\mathbf{T}$ of trees such that

- $\mathbf{P} - \bigcup_{T\in\mathbf{T}} T$ has size less than α;
- each $T \in \mathbf{T}$ contains a 2-overlapping subtree T_2 such that

$$w(I_T) \leq C\alpha^{-2}\|S_{T_2} f\|^2_{L^2(w)} = C\alpha^{-2}\sum_{P\in T_2} |a_P|^2 \frac{w(I_P)}{|I_P|}. \tag{3.4}$$

It then follows from a standard geometrical consideration that the tree collection $\mathbf{T}_2 := \{T_2 : T \in \mathbf{T}\}$ is well-separated when the constant K_0 in (S3) is chosen sufficiently large (see also Remark 3.1). We omit the details.

Proof of (3.2). Assume that $\mathbf{T}'$ covers $\mathbf{Q} := \bigcup_{T\in\mathbf{T}} T$, without loss of generality we can assume $\bigcup_{T'\in\mathbf{T}'} T' = \mathbf{Q}$. Let $\mathbf{Q}_2 = \bigcup_{T\in\mathbf{T}} T_2$. It follows from (3.4) that

$$\sum_{T\in\mathbf{T}} w(I_T) \leq C\alpha^{-2}\sum_{P\in\mathbf{Q}_2} |a_P|^2 \frac{w(I_P)}{|I_P|}. \tag{3.5}$$

Now, divide each $T' \in \mathbf{T}'$ into three trees,

$$\begin{aligned} T'_0 &= \{P \in T' : \inf C_2\omega_{P_1} \leq \xi_{T'} < \sup C_3\omega_{P_1}\},\\ T'_1 &= \{P \in T' : \sup C_3\omega_{P_1} \leq \xi_{T'} < \inf C_2\omega_{P_2}\},\\ T'_2 &= \{P \in T' : \xi_{T'} \in C_2\omega_{P_2}\}. \end{aligned}$$

Clearly, T'_2 is 2-overlapping. Since $\mathrm{size}(\mathbf{P}) \leq C\alpha$, we have

$$\sum_{P\in T'_2} |\langle f, \phi_{P_1}\rangle|^2 \frac{w(I_P)}{|I_P|} = \|S_{T'_2} f\|^2_{L^2(w)} \leq C\alpha^2 w(I_{T'}). \tag{3.6}$$

On the other hand, since $\mathbf{T}_2$ is well-separated, the rectangles

$$I_P \times [\inf C_2\omega_{P_1}, \sup C_3\omega_{P_1}), \qquad P \in \mathbf{Q}_2,$$

are pairwise disjoint in the phase plane. This implies that the bitiles of $T'_0 \cap \mathbf{Q}_2$ are spatially disjoint (as their frequency intervals overlap). Thus,

$$\sum_{P\in T'_0\cap\mathbf{Q}_2} |\langle f, \phi_{P_1}\rangle|^2 \frac{w(I_P)}{|I_P|} \leq \sum_{P\in T'_0\cap\mathbf{Q}_2} \mathrm{size}(\{P\})^2 w(I_P) \leq C\alpha^2 w(I_{T'}). \tag{3.7}$$

(3) To be more careful, one can fix a top frequency for each of these trees, and then select one tree (there are only finitely many of them) whose top frequency is minimal.

Next, we show that $T_1' \cap \mathbf{Q}_2$ can be grouped into $O(1)$ collections of 2-overlapping trees whose top intervals are disjoint. Together with the given assumption on the size of $\mathbf{P}$, this would imply

$$\sum_{P \in T_1' \cap \mathbf{Q}_2} |\langle f, \phi_{P_1} \rangle|^2 \frac{w(I_P)}{|I_P|} \leq C\alpha^2 \omega(I_{T'}). \tag{3.8}$$

Let M be the set of elements of $T_1' \cap \mathbf{Q}_2$ with maximal spatial intervals. The grouping of elements in $T_1' \cap \mathbf{Q}_2$ can be done as follows:

- Any element $P \in M$ can be viewed as one 2-overlapping tree, and we place these single-element trees in to the first tree collection.
- For any $P \in M$, we show below that we can place every $P' \in T_1' \cap \mathbf{Q}_2$ such that $I_{P'} \subsetneq I_P$ in $O(1)$ trees sharing the top interval I_P.

Since the intervals I_P, $P \in M$, are disjoint, it remains to show that if $P' \in T_1' \cap \mathbf{Q}_2$ and $I_{P'} \subsetneq I_P$ then

$$\inf C_2\omega_{P_2'} < \sup C_2\omega_{P_2} < \sup C_2\omega_{P_2'}. \tag{3.9}$$

Indeed, since $|\omega_{P_2'}| = 1/|I_{P'}| \geq 2/|I_P|$ it follows from (3.8) that we may take $-1/(2|I_P|) + \sup C_2\omega_{P_2}$ or $1/(2|I_P|) + \sup C_2\omega_{P_2}$ as the top frequency for these trees.

To see the first inequality in (3.9), we assume (towards a contradiction) that $\sup C_2\omega_{P_2} \leq \inf C_2\omega_{P_2'}$. By the selection algorithm, the 2-overlapping tree $S \in \mathbf{T}_2$ that contains P must be selected before the 2-overlapping tree S' of P'. Now, by definition of T_1' we have

$$[\sup C_3\omega_{P_1}, \inf C_2\omega_{P_2}) \cap [\sup C_3\omega_{P_1'}, \inf C_2\omega_{P_2'}) \neq \emptyset$$

(they both contain $\xi_{T'}$). On the other hand, by ensuring the constant K_0 is sufficiently large in the separation assumption (S3), we achieve that $\omega_P \subset$ convex hull$(C_2\omega_{P_1'} \cap C_2\omega_{P_2'})$. But then P' must be cleared out as part of the maximal tree with the same top data as S, leading to a contradiction. This proves the first half of (3.9).

To see the second inequality in (3.9), as before exploit the fact that

$$[\sup C_3\omega_{P_1}, \inf C_2\omega_{P_2}) \cap [\sup C_3\omega_{P_1'}, \inf C_2\omega_{P_2'}) \neq \emptyset.$$

By ensuring the constant K_0 in the separation assumption (S3) is sufficiently large, we have $|\omega_{P_2'}| \geq |\omega_{P_2}|$. As a consequence, if $\sup C_2\omega_{P_2} \geq \sup C_2\omega_{P_2'}$ then the interval $[\sup C_3\omega_{P_1}, \inf C_2\omega_{P_2})$ will be above $\inf C_2\omega_{P_2'}$, contradicting the nonempty intersection. This completes the proof of (3.9) and hence (3.8).

Finally, collecting inequalities (3.6)–(3.8), we obtain

$$\sum_{P \in T'} |\langle f, \phi_{P_1} \rangle|^2 \frac{w(I_P)}{|I_P|} \leq C\alpha^2 w(I_{T'}).$$

Summing over $T' \in \mathbf{T}'$ and using (3.5), we obtain the desired estimate (3.2).

Proof of (3.3). Fix k and let

$$N^{[k]} := \sum_{T\in\mathbf{T}} 1_{2^k I_T}.$$

It suffices to show the following good lambda estimate: given any $L \in (0,\infty)$ there exists $c_0 \in (0,\infty)$ and $c \in (0,\infty)$ such that

$$w(\{N^{[k]} > \lambda\} \cap E_\lambda^{[k]}) \le \frac{1}{L} w(\{N^{[k]} > \lambda/4\}), \tag{3.10}$$

where

$$E_\lambda^{[k]} := \{\mathcal{M}_{2q,w} f \le c2^{-c_0 k}\alpha\lambda^{1/(2q_0)}\}. \tag{3.11}$$

Indeed, choosing L sufficiently large (depending on $p \in [1,\infty)$) and applying a standard bootstrapping argument, we obtain

$$\begin{aligned}\Big\| \sum_{T\in\mathbf{T}} 1_{I_T}\Big\|_{L^p(w)} &\le C2^{O(k)}\alpha^{-2q_0}\|\mathcal{M}_{2q,w}(f)\|^{2q_0}_{L^{2pq_0}(w)}\\ &\le C2^{O(k)}\alpha^{-2q_0}\|f\|^{2q_0}_{L^{2pq_0}(w)},\end{aligned}$$

as desired. Here we have used the fact that $\mathcal{M}_{1,w}$ is bounded from $L^t(w) \to L^t(w)$ for any $1 < t < \infty$ and any positive weight w; note that we always have $2pq_0 > 2q$.

To prove (3.10), we use the following estimate which follows from (the remark after) Lemma 3.11 below: for any dyadic interval I and $q_0 \in (q,\infty)$,

$$w(\{N_I^{[k]} > \lambda/4\}) \le C2^{O(k)}\alpha^{-2q}\lambda^{-q/q_0} w(I)\Big[\inf_{x\in I}\mathcal{M}_{2q,w}(f\widetilde{\chi}_I^N)(x)\Big]^{2q} \tag{3.12}$$

where

$$N_I^{[k]} := \sum_{T\in\mathbf{T}:\, I_T\subset I} 1_{I_T}.$$

Let $\mathbf{I}$ be the collection of all maximal dyadic intervals of $\{N^{[k]} > \lambda/4\}$. We apply (3.12) to elements of $\mathbf{I}$ that intersect $E_\lambda^{[k]}$. Let I be one such interval; then it follows from the maximality of I that $\{N^{[k]} > \lambda\} \cap I$ is a subset of $\{N_I^{[k]} > \lambda/4\}$. Thus,

$$w(\{N^{[k]} > \lambda\} \cap I) \le C2^{O(k)}[\alpha^{-2q}\lambda^{-q/q_0}w(I)][c2^{-c_0k}\alpha\lambda^{1/(2q_0)}]^{2q}$$

and by choosing c sufficiently small and c_0 sufficiently large we obtain

$$w(\{N^{[k]} > \lambda\} \cap I) \le Cc^{2q}w(I) \le \frac{w(I)}{L}.$$

Summing the above estimates over all $I \in \mathbf{I}$ that intersect $E_\lambda^{[k]}$, we arrive at (3.10):

$$w(\{N^{[k]} > \lambda\} \cap E_\lambda^{[k]}) \le \sum_{I \in \mathbf{I}\colon I \cap E_\lambda^{[k]} \neq \emptyset} w(\{N^{[k]} > \lambda\} \cap I) \le \frac{1}{L} \sum_{I \in \mathbf{I}} w(I) = \frac{1}{L} w(\{N^{[k]} > \lambda/4\}). \ \blacksquare$$

LEMMA 3.11. *Let I be an interval and let $\mathbf{T}$ be a well-separated collection of 2-overlapping trees such that for any $T \in \mathbf{T}$ we have $I_T \subset I$, and*

$$w(I_T) \le C\alpha^{-2} \|S_T f\|_{L^2(w)}. \tag{3.13}$$

Then for any $q_0 \in (q, \infty)$ and $N > 0$ there is $C = C(q_0, w, N) < \infty$ such that

$$w(\{N^{[k]} > \lambda\}) \le C 2^{O(k)} [\alpha^{-1} \lambda^{-1/(2q_0)} \|f \widetilde{\chi}_I^N\|_{L^{2q}(w)}]^{2q}, \tag{3.14}$$

where

$$N^{[k]} := \sum_{T \in \mathbf{T}} 1_{2^k I_T}. \tag{3.15}$$

The implicit constant in $O(k)$ depends on w and q.

REMARK 3.12. As a consequence of (3.14), we obtain

$$w(\{N^{[k]} > \lambda\}) \le C 2^{O(k)} w(I) \Big[\alpha^{-1} \lambda^{-1/(2q_0)} \inf_{x \in I} \mathcal{M}_{2q,w}(f \widetilde{\chi}_I^N)(x)\Big]^{2q}. \tag{3.16}$$

Proof of Lemma 3.11. Since $N^{[k]}$ is integer-valued, without loss of generality we may assume $\lambda \ge 1/2$. We estimate

$$w(\{N^{[k]} > \lambda\}) \le \sum_{l \ge 0} w(\{2^l \lambda < N^{[k]} \le 2^{l+1} \lambda\}) \tag{3.17}$$

and it is not hard to see that

$$w(\{2^l \lambda < N^{[k]} \le 2^{l+1} \lambda\}) \le w(\{N_l^{[k]} > 2^l \lambda\}) \tag{3.18}$$

where

$$N_l^{[k]} := \sum_{T \in \mathbf{T}_l} 1_{2^k I_T} \quad \text{and} \quad \mathbf{T}_l := \{T \in \mathbf{T} : 2^k I_T \not\subset \{N^{[k]} > 2^{l+1}\lambda\}\}.$$

Write N_l for $N_l^{[0]}$; clearly $N_l \le N_l^{[k]}$ for $k \ge 0$. We first show that

$$\|N_l\|_\infty \le 2^{l+1} \lambda. \tag{3.19}$$

Indeed, take any x, and let $\mathbf{T}_x = \{T \in \mathbf{T}_l : x \in I_T\}$. Clearly,

$$N_l(x) \le \sum_{T \in \mathbf{T}_x} 1_{2^k I_T}.$$

Since the collection of top intervals of elements of T_x is nested, there is one minimal element. Note that if $I_1 \subset I_2$ are two intervals then for $k \ge 0$ we have $2^k I_1 \subset 2^k I_2$. Therefore the intervals $2^k I_T$ with $T \in \mathbf{T}_x$ are also nested and the minimal of them contains a point $y \in \{N^{[k]} \le 2^{l+1}\lambda\}$ by definition of $\mathbf{T}_l$. Therefore,

$$N_l(x) \le N^{[k]}(y) \le 2^{l+1}\lambda,$$

completing the proof of (3.19).

Now, denote $\mathbf{P}_l = \bigcup_{T\in\mathbf{T}_l} T$ and as usual

$$S_{\mathbf{P}_l} f = \Big(\sum_{P\in\mathbf{P}_l} |\langle f, \phi_{P_1}\rangle|^2 \frac{1_{I_P}}{|I_P|}\Big)^{1/2}.$$

It follows from (3.13), (1.5), and Hölder's inequality that

$$\alpha^{2q}\|N_l^{[k]}\|_{L^1(w)} \le C2^{\gamma k}\|S_{\mathbf{P}_l} f\|_{L^{2q}(w)}^{2q}. \tag{3.20}$$

For N large let $f_I = f\widetilde{\chi}_I$. The key estimate in our proof of (3.14) is

Claim 3.13. *For any $s \in (0,1)$ and $\delta > 0$ there is $C = C(\epsilon, s, N) < \infty$ such that*

$$(S_{\mathbf{P}_l} f)^\sharp \le C\|N_l\|_\infty^\delta(\mathcal{M}_2 f_I + [\alpha\mathcal{M}_2(N_l^{1/2})]^s(\mathcal{M}_2 f_I)^{1-s}). \tag{3.21}$$

Now we show (3.14) using the above claim. It follows from (3.20), (3.21), and the assumption $w \in A_q$ that

$$\begin{aligned}\alpha\|N_l^{[k]}\|_{L^1(w)}^{1/(2q)} &\le C2^{O(k)}\|(S_{\mathbf{P}_l} f)^\sharp\|_{L^{2q}(w)}\\ &\le C2^{O(k)}\|N_l\|_\infty^\delta(\|f_I\|_{L^{2q}(w)} + [\alpha\|N_l^{1/2}\|_{L^{2q}(w)}]^s\|f_I\|_{L^{2q}(w)}^{1-s})\\ &\le C2^{O(k)}(\|N_l\|_\infty^\delta\|f_I\|_{L^{2q}(w)} + \|N_l\|_\infty^{\delta+s(1/2-1/2q)}\alpha^s\|N_l\|_{L^1(w)}^{s/(2q)}\|f_I\|_{L^{2q}(w)}^{1-s}).\end{aligned}$$

The numbers $\delta > 0$ and $s > 0$ will be chosen very close to 0. Consequently, after bootstrapping, it follows that for any $\epsilon > 0$,

$$\alpha\|N_l^{[k]}\|_{L^1(w)}^{1/(2q)} \le C2^{O(k)}\|N_l\|_\infty^{\epsilon/(2q)}\|f_I\|_{L^{2q}(w)}.$$

Therefore, it follows from the bound $\|N_l\|_\infty \le 2^{l+1}\lambda$ of (3.19) that

$$w(\{N_l^{[k]} > 2^l\lambda\}) \le C2^{O(k)}2^{-l(1-\epsilon)}\alpha^{-2q}\lambda^{-1+\epsilon}w(I)\Big[\inf_{x\in I}\mathcal{M}_{2q,w}f(x)\Big]^{2q}.$$

Choosing $\epsilon > 0$ very small allows for summation over $l \ge 0$ of the above estimate. Using (3.17) and (3.18), we obtain the desired estimate (3.14).

Proof of Claim 3.13. Fix any dyadic J. For any $T \in \mathbf{T}_l$ let $T_J := \{P \in T : I_P \subset J\}$, and by decomposing T_J into $O(1)$ subtrees we may assume that T_J is a tree with a new top interval $I_T \cap J$ for every $T \in \mathbf{T}_l$. It suffices to show that for any $x \in J$,

$$\frac{1}{|J|^{1/2}}\Big\|\Big(\sum_{T\in\mathbf{T}_l}|S_{T_J} f|^2\Big)^{1/2}\Big\|_{L^2} \le \text{the value at } x \text{ of RHS of (3.21).} \tag{3.22}$$

By Lemma 3.14 belov, for any $0 < s \le 1$ there is $C = C_s < \infty$ such that

$$\Big(\sum_{T \in \mathbf{T}_l} \|S_{T_J} f\|_2^2 \Big)^{1/2} \le C \|f\|_2 + C\alpha^s \|N_l^{1/2}\|_{L^2(J)}^s \|f\|_2^{1-s}. \tag{3.23}$$

Here we have used the fact that for any $P \in \mathbf{P}$,

$$\frac{|a_P|}{|I_P|^{1/2}} = \Big(\frac{1}{w(I_P)} \int |a_P|^2 \frac{1_{I_P}}{|I_P|} w(x)\, dx \Big)^{1/2} \le \alpha.$$

Since for any $P \in T_J$ we have $I_P \subset I \cap J$, it follows from Corollary 3.9 that

$$\|S_{T_J} f\|_{\mathrm{BMO}} \le C \inf_{x \in I \cap J} \mathcal{M}_1(f \widetilde{\chi}_{I \cap J}^N)(x). \tag{3.24}$$

Interpolate the estimates (3.23) and (3.24) to prove (3.22) using a now-standard localization argument (see e.g. [LT3]). The idea is to decompose $f = \sum_{k \ge 0} f_k$ where $f_0 = f 1_{I \cap J}$ and $f_k = f 1_{2^k(I \cap J) \setminus 2^{k-1}(I \cap J)}$ for $k \ge 1$ and apply (3.23) and (3.24) to f_k. More specifically, for $p \in (2, \infty)$ we have

$$\begin{aligned}
\Big\| \Big(\sum_{T \in \mathbf{T}_l} |S_{T_J} f_k|^p \Big)^{1/p} \Big\|_p &= \Big(\sum_{T \in \mathbf{T}_l} \|S_{T_J} f_k\|_p^p \Big)^{1/p} \\
&\le \Big(\sum_{T \in \mathbf{T}_l} \|S_{T_J} f_k\|_2^2 \Big)^{1/p} \sup_{T \in \mathbf{T}_l} \|S_{T_J} f_k\|_{\mathrm{BMO}}^{1-2/p}
\end{aligned}$$

$$\le C_{N,p} 2^{-Nk} |I \cap J|^{1/p} \inf_{x \in I \cap J} (\mathcal{M}_2 f_I(x) + [\alpha \mathcal{M}_2(N_l^{1/2})(x)]^{2s/p} [\mathcal{M}_2 f_I(x)]^{1-2s/p}).$$

Summing over $k \ge 0$ we obtain

$$\begin{aligned}
&\Big\| \Big(\sum_{T \in \mathbf{T}_l} |S_{T_J} f|^p \Big)^{1/p} \Big\|_p \\
&\qquad \le C |J|^{1/p} \inf_{x \in J} (\mathcal{M}_2 f_I(x) + [\alpha \mathcal{M}_2(N_l^{1/2})(x)]^{2s/p} \mathcal{M}_2 f_I(x)^{1-2s/p}).
\end{aligned} \tag{3.25}$$

On the other hand, using Hölder's inequality we find that

$$\Big\| \Big(\sum_{T \in \mathbf{T}_l} |S_{T_J} f_k|^2 \Big)^{1/2} \Big\|_p \le \|N_l\|_\infty^{1/2-1/p} \Big\| \Big(\sum_{T \in \mathbf{T}_l} |S_{T_J} f|^p \Big)^{1/p} \Big\|_p. \tag{3.26}$$

Combining (3.25), (3.26) and Hölder's inequality yields

$$\begin{aligned}
\frac{1}{|J|^{1/2}} \Big\| \Big(\sum_{T \in \mathbf{T}_l} |S_{T_J} f|^2 \Big)^{1/2} \Big\|_{L^2} &\le \frac{1}{|J|^{1/p}} \Big\| \Big(\sum_{T \in \mathbf{T}_l} |S_{T_J} f|^2 \Big)^{1/2} \Big\|_{L^p} \\
&\le C \|N_l\|_\infty^{1/2-1/p} [\mathcal{M}_2 f_I(x) + [\alpha \mathcal{M}_2(\sqrt{N_l})(x)]^{2s/p} [\mathcal{M}_2 f_I(x)]^{1-2s/p}.
\end{aligned}$$

Choosing $p > 2$ sufficiently close to 2 we get the desired estimate (3.22). ∎

The following lemma, needed for our proof of Claim 3.13, is contained implicitly in [T1], where in fact a stronger logarithmic variant was proved (see also [HL] for a vector- valued generalization).

LEMMA 3.14. *Let* $\mathbf{T}$ *be a well-separated collection of* 2*-overlapping trees and let* $\mathbf{P} = \bigcup_{T\in\mathbf{T}} T$. *Then for any* $0 < s \le 1$ *we have*

$$\Big(\sum_{P\in\mathbf{P}} |\langle f, \phi_{P_1}\rangle|^2\Big)^{1/2} \le C_s\bigg(\|f\|_2 + \bigg[\sup_{P\in\mathbf{P}} \frac{|\langle f, \phi_{P_1}\rangle|}{|I_P|^{1/2}}\Big(\sum_{T\in\mathbf{T}} |I_T|\Big)^{1/2}\bigg]^s \|f\|_2^{1-s}\bigg). \tag{3.27}$$

REMARK. While any $0 < s < 1$ would be enough for applications to the Lebesgue setting of Carleson theorems (see e.g. [LT2] and [OST+] where $s = 1/3$ is used), our applications to Claim 3.13 require arbitrarily small $s > 0$. We include a proof of (3.27) (following largely [T1]) below.

Proof of Lemma 3.14. Without loss of generality we may assume $\|f\|_2 = 1$. Denote

$$N = \sum_{T\in\mathbf{T}} 1_{I_T}, \quad a_P = \langle f, \phi_{P_1}\rangle, \quad A = \Big(\sum_{P\in\mathbf{P}} |a_P|^2\Big)^{1/2}, \quad B = \sup_{P\in\mathbf{P}} \frac{|a_P|}{|I_P|^{1/2}}.$$

We then divide $\mathbf{P}$ into subcollections $\mathbf{P}_k$, where for any $k \ge 0$ we have

$$\mathbf{P}_k = \Big\{P \in \mathbf{P} : 2^{-k-1}B < \frac{|a_P|}{|I_P|^{1/2}} \le 2^{-k}B\Big\},$$

and let $\mathbf{P}_{\ge k} = \bigcup_{j\ge k} \mathbf{P}_j$. Using the known special case $s = 1/3$ of (3.27) proved in [LT2] (see also [OST+] for a setting similar to the current paper) for the restriction to $\mathbf{P}_{\ge k}$ of the tree collection $\mathbf{T}$, we obtain

$$\Big(\sum_{P\in\mathbf{P}_{\ge k}} |a_P|^2\Big)^{1/2} \le C + C(2^{-k}B)^{1/3}\|N\|_1^{1/6};$$

in particular for $k \ge \max(0, \log_2[B(\sum_{T\in\mathbf{T}} |I_T|)^{1/2}])$ we have

$$\Big(\sum_{P\in\mathbf{P}_{\ge k}} |a_P|^2\Big)^{1/2} \le C. \tag{3.28}$$

On the other hand, it follows from the definition of $\mathbf{P}_k$ that

$$\Big(\sum_{P\in\mathbf{P}_k} |a_P|^2\Big)^{1/2} \sim 2^{-k}B\Big(\sum_{P\in\mathbf{P}_k} |I_P|\Big)^{1/2}. \tag{3.29}$$

We can also view $\mathbf{P}_k$ as a collection of single-bitile trees, which is clearly well-separated. Thus again by the known case $s = 1/3$ of (3.27), it follows that

$$\Big(\sum_{P\in\mathbf{P}_k} |a_P|^2\Big)^{1/2} \le C + C\Big[2^{-k}B\Big(\sum_{P\in\mathbf{P}_k}|I_P|\Big)^{1/2}\Big]^{1/3}. \tag{3.30}$$

Combining (3.29) and (3.30), we deduce that, for any $k \ge 0$,

$$\Big(\sum_{P\in\mathbf{P}_k} |a_P|^2\Big)^{1/2} \le C. \tag{3.31}$$

From (3.28) and (3.31) we obtain

$$\sum_{P\in\mathbf{P}} |a_P|^2 \le C + C\max(0, \log_2[B\|N\|_1^{1/2}]).$$

Using the trivial estimate $\max(0, \log x) \le x$ for any $x > 0$, we conclude that

$$\Big(\sum_{P\in\mathbf{P}} |a_P|^2\Big)^{1/2} \le C(1 + [B\|N\|_1^{1/2}]^s)$$

for any $0 < s \le 1$, as desired. ■

3.4. Decomposition by density. Since $|g| \le 1_G$, the density of any collection is bounded above by 1. For the result below, it is important that the constant D in the definition of density is sufficiently large, much bigger than the doubling exponent γ of w. We return to this point in the proof.

LEMMA 3.15. *For any collection $\mathbf{P}$ of bitiles and any $\alpha > 0$ we can find a collection $\mathbf{T}$ of trees such that the density of $\mathbf{P} - \bigcup_{T\in\mathbf{T}} T$ is bounded above by α and*

$$\sum_{T\in\mathbf{T}} w(I_T) \le C\alpha^{-r'} w(G)$$

here r is the variational exponent used in the definition of density.

REMARK. This is a weighted extension of [OST+, Proposition 4.4], and the proof below is adapted from [OST+], which is in turn a variational adaptation of the standard argument. The variant of Lemma $\widetilde{3.15}$ with improved density follows immediately, since for any $\mathbf{P}$ we have $\widetilde{\text{density}}(\mathbf{P}) \le C\,\text{density}(\mathbf{P})$.

Proof of Lemma 3.15. If density$(\mathbf{P}) > \alpha$ then there is a nonempty tree $T \subset \mathbf{P}$ such that

$$\omega(I_T) \le \alpha^{-r'} \int \widetilde{\chi}_{I_T}^D |g|^{r'} \sum_{j\colon N_j\in\omega_T} |d_j|^{r'} w. \tag{3.32}$$

We select T such that $|I_T|$ is maximal, and then by enlarging T (keeping I_T and ξ_T) if necessary we may assume that T is maximal in $\mathbf{P}$ with respect to set inclusion. Let T_+ and T_- be the maximal trees in $\mathbf{P}$ with the same top interval as T but with top frequencies $\xi_T - 1/(2|I_T|)$ and $\xi_T + 1/(2|I_T|)$

respectively. We then remove from $\mathbf{P}$ the union of T, T_+, T_-. Continuing this selection process, which will stop since $\mathbf{P}$ is assumed finite, we obtain a collection $\mathbf{T}$ of trees such that

$$\text{density}\Big(\mathbf{P} - \bigcup_{T\in\mathbf{T}} (T \cup T_- \cup T_+)\Big) \leq \alpha.$$

We now show that

$$\sum_{T\in\mathbf{T}} w(I_T) \leq C\alpha^{-r'} w(G).$$

By the selection algorithm, it is not hard to see that for $T \neq T'$ in $\mathbf{T}$ the rectangles $I_T \times \omega_T$ and $I_{T'} \times \omega_{T'}$ are disjoint. Now, it follows from (3.32) that for any $T \in \mathbf{T}$ there exists an integer $k = k(T) \geq 0$ such that

$$\omega(I_T) \leq C2^{-Dk}\alpha^{-r'} \int_{2^k I_T} |g|^{r'} \sum_{j\colon N_j\in\omega_T} |d_j|^{r'} w. \tag{3.33}$$

We then sort the trees in $\mathbf{T}$ according to the value of $k(T)$. More specifically for each $k \geq 0$ let $\mathbf{T}_k = \{T \in \mathbf{T} : k(T) = k\}$. It suffices to show that

$$\sum_{T\in\mathbf{T}_k} w(I_T) \leq C\alpha^{-r'} 2^{-k} w(G). \tag{3.34}$$

Fix k. Select a subcollection $\mathbf{S}_k \subset \mathbf{T}_k$ such that the rectangles $2^k I_S \times \omega_S$ with $S \in \mathbf{S}_k$ are pairwise disjoint, and such that

$$\sum_{T\in\mathbf{T}_k} \omega(I_T) \leq C \sum_{S\in\mathbf{S}_k} \omega(2^{k+2} I_S). \tag{3.35}$$

Note that this will imply the desired estimate (3.34). By choosing $D > \gamma+10$, where γ is the doubling exponent for w, it follows from (3.33) and (3.35) that

$$\begin{aligned}
\sum_{T\in\mathbf{T}_k} \omega(I_T) &\leq C2^{k\gamma} \sum_{S\in\mathbf{S}_k} \omega(I_S) \\
&\leq C2^{-k}\alpha^{-r'} \int \sum_j \sum_{S\in\mathbf{S}_k} 1_{\{(x,N_j(x))\in 2^k I_S\times\omega_S\}} |d_j|^{r'} |g|^{r'} w \\
&\leq C2^{-k}\alpha^{-r'} \int |g|^{r'} \sum_j |d_j|^{r'} w \leq C2^{-k} w(G).
\end{aligned}$$

It remains to select $\mathbf{S}_k$. Assume without loss of generality that $\mathbf{T}_k$ is nonempty. Then we choose $S \in \mathbf{T}_k$ such that $|I_S|$ is maximal and then remove all $T \in \mathbf{T}$ if

$$2^k I_T \times \omega_T \cap 2^k I_S \times \omega_S \neq \emptyset.$$

Starting from the remaining collection, we repeat the above selection procedure until no trees are left. We then let $\mathbf{S}_k$ be the collection of selected trees. For any $S \in \mathbf{S}_k$, let $\mathbf{T}_S$ denote the collection of trees in $\mathbf{T}$ that are removed after S is selected; to show (3.35) it suffices to show that

$$\sum_{T \in \mathbf{T}_S} 1_{I_T} \le C 1_{2^{k+2} I_S}. \tag{3.36}$$

Note that if $T \in \mathbf{T}_S$ then $|I_T| \le |I_S|$ and $2^k I_T \cap 2^k I_S \neq \emptyset$, so clearly $I_T \subset 2^{k+2} I_S$. Also $|\omega_T| \ge |\omega_S|$ and $\omega_T \cap \omega_S \neq \emptyset$, so out of any four trees in $\mathbf{T}_S$ at least two will have overlapping top frequency intervals. The desired estimate (3.36) then follows from the fact that the rectangles $I_T \times \omega_T$ (with $T \in \mathbf{T}_S$) are disjoint. ∎

4. The tree estimate. In this section we prove several estimates for the restriction of the (model) Carleson operator to a tree. Lemma 4.1 is applicable to any tree, while Lemma 4.2 improves the L^1 case of Lemma 4.1 when the elements of the underlying tree are disjoint in the phase plane.

Recall that for any bitile collection $\mathbf{Q}$ we denote

$$C_{\mathbf{Q}} f(x) = \sum_{P \in \mathbf{Q}} \langle f, \phi_{P_1} \rangle \phi_{P_1}(x) d_P(x)$$

with d_P defined as follows: First, $(d_k)_{k \ge 1}$ and N_k are two sequences of measurable functions of x such that

- For each x there is some integer $K = K(x) < \infty$ such that $d_k(x) = 0$ for $k > K$, and uniformly over x we have $\sum_{k \ge 0} |d_k(x)|^{r'} = 1$.
- For any x we have $N_0(x) < N_1(x) < \cdots$.

Then for each x define $d_P(x) = 0$ unless there exists an index k such that $N_{k-1} \notin \omega_P$ and $N_k \in \omega_{P_2}$, in which case such an index is unique and we define $d_P(x) := d_k(x)$. We note that if $P \in \mathbf{P}$ then

$$\int \widetilde{\chi}_{I_P}^D |g|^{r'} \sum_{k\colon N_k \in \omega_{P_2}} |d_k|^{r'} w \le C w(I_P) \operatorname{density}(\{P\})^r.$$

The above observation will be used implicitly below.

Lemma 4.1. *Let T be a tree. Assume $s \in [1, r']$. Then there exists some $C = C(s, w) < \infty$ such that*

$$\|1_{I_T} g C_T f\|_{L^s(w)} \le C w(I_T)^{1/s} \operatorname{size}(T) \operatorname{density}(T), \tag{4.1}$$

and furthermore for any $N > 0$ there exists $C = C(N, s, w) < \infty$ such that, for any $k \ge 0$,

$$\|1_{2^{k+1} I_T \setminus 2^k I_T} g C_T f\|_{L^s(w)} \le C 2^{-Nk} w(I_T)^{1/s} \operatorname{size}(T) \operatorname{density}(T). \tag{4.2}$$

Remark. As a consequence, for any $s \in [1, r']$ we obtain

$$\|g C_T f\|_{L^s(w)} \le C w(I_T)^{1/s} \operatorname{size}(T) \operatorname{density}(T). \tag{4.3}$$

Proof of Lemma 4.1. By Hölder's inequality and using the doubling property of w it suffices to show (4.1) and (4.2) for $s = r'$, which will be assumed in the rest of the proof. By dividing T into two subtrees if necessary, we can assume that the tree is either 2-overlapping or 2-lacunary. We will return to this distinction below.

Proof of (4.1). We will prove a stronger estimate, where the restriction 1_{I_T} is not required. Let $\mathbf{J}$ be the set of maximal dyadic intervals such that

$$I_P \not\subset 3J$$

for any $P \in T$. It is not hard to see that $\mathbf{J}$ partitions $\mathbb{R}$. Let

$$T_J := \{P \in T : |I_P| \leq C_4 |J|\} \tag{4.4}$$

for some absolute constant $C_4 \geq 4$ to be chosen later. The left hand side of (4.1) (with $s = r'$ now) is bounded above by $A + B$ where

$$A := \Big(\sum_{J\in\mathbf{J}} \int_J |gC_{T_J} f|^{r'} w\Big)^{1/r'}, \tag{4.5}$$

$$B := \Big(\sum_{J\in\mathbf{J}} \int_J |gC_{T\setminus T_J} f)|^{r'} w\Big)^{1/r'}. \tag{4.6}$$

To bound A, we fix $J \in \mathbf{J}$ and first estimate the contribution of each $P \in T_J$:

$$\begin{aligned}\Big(\int_J |gC_{\{P\}} f|^{r'} w\Big)^{1/r'} &\leq C_N \frac{|\langle f, \phi_{P_1}\rangle|}{|I_P|^{1/2}} \Big(\int |1_J g \widetilde{\chi}_{I_P}^{N+D} d_P|^{r'} w\Big)^{1/r'} \\ &\leq C w(I_P)^{1/r'} \operatorname{size}(\{P\}) \operatorname{density}(\{P\}) \sup_{y\in J} \widetilde{\chi}_{I_P}(y)^N.\end{aligned}$$

From the triangle inequality, it follows that

$$\begin{aligned}&\Big(\int_J |gC_{T_J} f|^{r'} w\Big)^{1/r'} \\ &\qquad \leq C \operatorname{size}(T) \operatorname{density}(T) \sum_{P\in T_J} w(I_P)^{1/r'} \Big(1 + \frac{\operatorname{dist}(J, I_P)}{|I_P|}\Big)^{-4N}.\end{aligned} \tag{4.7}$$

By the A_∞ property of w there exists a constant $\beta_0 > 0$ such that if $I \subset I'$ are two intervals then

$$\frac{w(I)}{w(I')} \leq C \Big(\frac{|I|}{|I'|}\Big)^{\beta_0}.$$

Without loss of generality, we may choose the doubling constant γ in (1.5) to be large enough such that $\gamma > \beta_0$.

For any $P \in T_J$ we can find an interval K of length comparable to $|I_P| + |J| + \operatorname{dist}(J, I_P)$ that contains both I_P and J. Since $|I_P| = O(|J|)$ we can choose K to be a dilation of J. We then have

$$\begin{aligned}\frac{w(I_P)}{w(J)} = \frac{w(I_P)}{w(K)}\,\frac{w(K)}{w(J)} &\le C\Big(\frac{|I_P|}{|K|}\Big)^{\beta_0}\Big(\frac{|K|}{|J|}\Big)^{\gamma} = C\Big(\frac{|I_P|}{|J|}\Big)^{\beta_0}\Big(\frac{|K|}{|J|}\Big)^{\gamma-\beta_0}\\ &\le C\Big(\frac{|I_P|}{|J|}\Big)^{\beta_0}\Big(1+\frac{\operatorname{dist}(J,I_P)}{|J|}\Big)^{\gamma-\beta_0}\\ &\le C\Big(\frac{|I_P|}{|J|}\Big)^{\beta_0}\Big(1+\frac{\operatorname{dist}(J,I_P)}{|I_P|}\Big)^{\gamma-\beta_0}.\end{aligned}$$

Therefore by choosing N sufficiently large it follows from (4.7) that

$$\begin{aligned}&\Big(\int_J |gC_{T_J}f|^{r'}w\Big)^{1/r'}\\ &\le C\,\mathrm{size}(T)\,\mathrm{density}(T)\sum_{P\in T_J}\Big(\frac{|I_P|}{|J|}\Big)^{\beta_0/r'} w(J)^{1/r'}\Big(1+\frac{\operatorname{dist}(J,I_P)}{|I_P|}\Big)^{-3N}\\ &= C\,\mathrm{size}(T)\,\mathrm{density}(T)w(J)^{1/r'}\sum_{k\ge -1}\sum_{|I_P|=2^{-k}|J|}2^{-k\beta_0/r'}\Big(1+\frac{\operatorname{dist}(J,I_P)}{|I_P|}\Big)^{-3N}.\end{aligned}$$

Using the fact that $3J$ does not contain any I_P, $P\in T_J$, and the fact that elements of T_J of the same size are spatially disjoint, it is not hard to bound the last display by

$$\le C\,\mathrm{size}(T)\,\mathrm{density}(T)w(J)^{1/r'}\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-2N}$$

Thus, we can bound A by

$$A\le C\,\mathrm{size}(T)\,\mathrm{density}(T)\Big(\sum_{J\in\mathbf{J}} w(J)\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-2Nr'}\Big)^{1/r'}.$$

Note that by definition $3J$ does not contain I_T. It follows that for any $x\in J$,

$$1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\sim 1+\frac{|x-c(I_T)|}{|I_T|}.$$

Choosing N large and using disjointness of J's, we obtain

$$\sum_{J\in\mathbf{J}} w(J)\Big(1+\frac{\operatorname{dist}(J,I_T)}{|I_T|}\Big)^{-2Nr'}\le C\int\widetilde{\chi}_{I_T}^N w\le Cw(I_T).$$

Consequently, we have

$$A\le Cw(I_T)^{1/r'}\,\mathrm{size}(T)\,\mathrm{density}(T).$$

To bound B, let $F_J=\bigcup_{T\in T\setminus T_J}\omega_{P_2}$. We first show that

$$\int_J |g|^{r'}\sum_{j:N_j\in F_J}|d_j|^{r'}w\le Cw(J)[\mathrm{density}(T)]^{r'}. \tag{4.8}$$

We construct $O(1)$ nonempty subtrees of T such that F_J is contained inside the union of the frequency intervals of these trees. The top interval of each such subtree will be of length $\sim |J|$ and will be contained in some $O(1)$ dilation of J. Clearly, (4.8) follows as a consequence of this construction.

To construct these trees, first we construct their (common) top interval J_0. Let $\pi(J)$ be the dyadic parent of J. Then we can find $Q \in T$ such that $I_Q \subset 3\pi(J)$, therefore we can select a dyadic interval J_0 such that

$$I_Q \subset J_0 \subset 3\pi(J), \qquad |J_0| \geq |J|.$$

Now, note that by dividing T into three trees if necessary, we may assume without loss of generality that only one of the following scenarios happens:

(i) $\xi_T \in \omega_{P_2}$ for every $P \in T$, or
(ii) $\xi_T < \inf \omega_{P_2}$ for every $P \in T$, or
(iii) $\xi_T \geq \sup \omega_{P_2}$ for every $P \in T$.

In each of these scenarios, one tree will be constructed. The desired tree has only one element Q and has top data (J_0, ω_0), and ω_0 is constructed below: it will be shown that

$$F_J \subset \omega_0 \subset \widetilde{\omega}_Q. \tag{4.9}$$

We note that by choosing C_4 large in the definition (4.4) we can ensure that for any $P \in T \setminus T_J$ we have $|\omega_P| < 1/|J_0|$. Furthermore, if $C_2 > 1$ we can also ensure that $|\omega_P| < (C_2 - 1)|J_0|/2$.

If (i) is satisfied, we let ω_0 be the dyadic interval of length $1/|J_0|$ containing ξ_T. It is clear that for any $P \in T \setminus T_J$ we have $\omega_{P_2} \subset \omega_0$ and $\omega_0 \subset \omega_{Q_2}$, and (4.9) follows immediately.

If (ii) is satisfied, we let $\omega_0 = [\xi_T, \xi_T + 1/|J_0|)$. Since for any $P \in T \setminus T_J$ we have $|\omega_P| < |\omega_0| < |\omega_{Q_2}|$, it follows that we always have $\omega_{P_2} \subset \omega_0 \subset \widetilde{\omega}_Q$, as desired.

If (iii) is satisfied, we let $\omega_0 = [\xi_T - 1/|J_0|, \xi_T)$, and argue as for (ii).

This completes the proof of (4.8). ∎

Now we return to our task of estimating B. We remark that any $J \in \mathbf{J}$ that contributes to B must satisfy $|J| < |I_T|/C_4 \leq |I_T|/4$, therefore $J \subset 3I_T$. We now consider two cases:

Case 1: *T is 2-lacunary.* By ensuring that the constant K_0 in the separation assumption (S3) is sufficiently large, it follows that for $P, P' \in T$ with $|I_P| > |I_{P'}|$ we have $\omega_{P_2} \subset \omega_{P'}$. Since $\{N_j(x)\}$ is an increasing sequence for every x, it follows from a geometrical consideration that for each x there is at most one m and such that $d_P(x) \neq 0$ for some $P \in T$ with $|I_P| = 2^m$. Here it is important that the limiting condition reads $\{N_{j-1} \notin \omega_P, N_j \in \omega_{P_2}\}$.

Now, uniformly over m we have

$$\sum_{P\in T:\,|I_P|=2^m}\left(1+\frac{|x-c(I_P)|}{|I_P|}\right)^{-2}=O(1).$$

It then follows from (4.8) that

$$\begin{aligned}\|1_J g C_{T\setminus T_J} f\|_{L^{r'}(w)} &\le C\sup_{P\in T}\frac{|\langle f,\phi_{P_1}\rangle|}{|I_P|^{1/2}}\Big(\int_J |g|^{r'}\sup_{k:\,N_k\in\omega_{T_0}}|d_j|^{r'}w\Big)^{1/r'}\\ &\le C\operatorname{size}(T)\operatorname{density}(T)w(J)^{1/r'}.\end{aligned}$$

Consequently, we obtain the desired estimate:

$$\begin{aligned}B &\le C\Big(\sum_{J\in\mathbf{J}}w(J)\Big)^{1/r'}\operatorname{size}(T)\operatorname{density}(T)\\ &\le Cw(I_T)^{1/r'}\operatorname{size}(T)\operatorname{density}(T).\end{aligned}$$

CASE 2: *T is 2-overlapping.* We estimate pointwise

$$\begin{aligned}&|C_{T\setminus T_J}f(x)|\\ &\le\Big(\sum_j\Big|\sum_{P\in T\setminus T_J:\,N_{j-1}\notin\omega_P,\,N_j\in\omega_{P_2}}\langle f,\phi_{P_1}\rangle\phi_{P_1}\Big|^r\Big)^{1/r}\Big(\sum_{j:\,N_j\in\omega_{T_0}}|d_k(x)|^{r'}\Big)^{1/r'}.\end{aligned}$$

Therefore

$$\begin{aligned}(4.10)\quad &\|1_J g C_{T\setminus T_J}f\|_{L^{r'}(w)}\\ &\le Cw(J)^{1/r'}\operatorname{density}(T)\sup_{x\in J}\Big(\sum_j\Big|\sum_{P\in T\setminus T_J:\,N_{j-1}\notin\omega_P,\,N_j\in\omega_{P_2}}\langle f,\phi_{P_1}\rangle\phi_{P_1}\Big|^r\Big)^{1/r}.\end{aligned}$$

Note that for any P the frequency support of ϕ_{P_1} is contained inside $C_3\omega_{P_1}=(1-c)C_2\omega_{P_1}$ for $c=1-C_3/C_2\in(0,1)$, which is uniform over P's. Recall that T is a 2-overlapping tree and the relative positions of the tiles in each bitile are uniform over $\mathbf{P}$.

Now, by choosing the constant K in the separation assumption (S3) to be sufficiently large, we can find a lacunary family of smooth Littlewood–Paley projection operators Π_n such that: Π_n is a smooth Fourier multiplier operator whose symbol is supported in $\{|\xi|=O(2^n)\}$, and furthermore (thanks to separation) $\Pi_n\Pi_k=\Pi_k$ for any $n<k$ and $\phi_{P_1}=(\Pi_n-\Pi_{n-1})\phi_{P_1}$ for $n=\log_2|I_P|$.

It follows that for any $x\in J$ we can bound

$$\begin{aligned}&\Big(\sum_k\Big|\sum_{P\in T\setminus T_J:\,N_{k-1}\notin\omega_P,\,N_k\in\omega_{P_2}}\langle f,\phi_{P_1}\rangle\phi_{P_1}\Big|^r\Big)^{1/r}\\ &\qquad\le\sup_{K,\,n_0<\cdots<n_K<O(1)-\log_2|J|}\Big(\sum_{j=1}^K|(\Pi_{n_j}-\Pi_{n_{j-1}})g_T|^r\Big)^{1/r}.\end{aligned}$$

where $g_T := \sum_{P\in T}\langle f,\phi_{P_1}\rangle\phi_{P_1}$. The last display can be continued as

$$= \sup_{K,\, n_0<\cdots<n_K<O(1)-\log_2|J|}\Big(\sum_{j=1}^{K}|\Pi_{\log_2|J|}(\Pi_{n_j}-\Pi_{n_{j-1}})g_T|^r\Big)^{1/r}$$
$$\le M_J\Big(\sup_{K,\, n_0<\cdots<n_K}\Big(\sum_{j=1}^{K}|\Pi_{n_j}g_T-\Pi_{n_{j-1}}g_T|^r\Big)^{1/r}\Big),$$

using Minkowski's inequality and standard arguments. Here, M_J denotes the following local maximal operator:

$$M_Jf = \sup_{I:\, J\subset I}\frac{1}{|I|}\int_I |f|.$$

For simplicity we denote by $\|g_T\|_{V^r}$ the variational expression inside M_J in the above estimate. Recall that all the J such that $T\setminus T_J$ are disjoint and contained in $3I_T$. Thus, it follows from (4.10) and the above estimate that

$$\begin{aligned}
B &\le C\,\mathrm{density}(T)\Big(\sum_{J\in\mathbf{J}}w(J)M_J(\|g_T\|_{V^r})^{r'}\Big)^{1/r'}\\
&\le C\,\mathrm{density}(T)\|1_{3I_T}M(\|g_T\|_{V^r})\|_{L^{r'}(w)}\\
&\le C\,\mathrm{density}(T)w(I_T)^{1/r'-1/(2q)}\|M(\|g_T\|_{V^r})\|_{L^{2q}(w)},
\end{aligned}$$

since $r'<2<2q$. Using $w\in A_q\subset A_{2q}$ and Lemma 5.2 below, we obtain

$$B\le C\,\mathrm{density}(T)w(I_T)^{1/r'-1/(2q)}\|g_T\|_{L^{2q}(w)}.$$

To show the desired bound for B it remains to show that

$$\|g_T\|_{L^{2q}(w)}\le Cw(I_T)^{1/(2q)}\,\mathrm{size}(T).$$

Take h to be any function in $L^{(2q)'}(w)$ where $(2q)'$ denotes the dual exponent of $2q$. Let $\sigma = w^{-(2q)'/(2q)}$; since $w\in A_q\subset A_{2q}$ it is clear that $\sigma\in A_{(2q)'}$. We have

$$\begin{aligned}
\langle g_T, wh\rangle &= \sum_{P\in T}\langle f,\phi_{P_1}\rangle\langle hw,\phi_{P_1}\rangle\\
&\le \int\Big(\sum_{P\in T}|\langle f,\phi_{P_1}\rangle|^2\frac{1_{I_P}}{|I_P|}\Big)^{1/2}\Big(\sum_{P\in T}|\langle hw,\phi_{P_1}\rangle|^2\frac{1_{I_P}}{|I_P|}\Big)^{1/2}dx\\
&\le \|S_Tf\|_{L^{2q}(w)}\|S_T(hw)\|_{L^{(2q)'}(\sigma)}.
\end{aligned}$$

Then using the John–Nirenberg characterization of size in Lemma 3.8 and the estimate (3.1), it is not hard to see that

$$\begin{aligned}
\langle g_T, wh\rangle &\le Cw(I_T)^{1/(2q)}\,\mathrm{size}(T)\|hw\|_{L^{(2q)'}(\sigma)}\\
&= Cw(I_T)^{1/(2q)}\,\mathrm{size}(T)\|h\|_{L^{(2q)'}(w)},
\end{aligned}$$

as desired.

Proof of (4.2). Let $\widetilde{g} = g1_{2^{k+1}I_T \setminus 2^k I_T}$. Note that it suffices to consider $k \geq 2$. One proceeds as in the above proof of (4.1) with $\widetilde{g}$ in place of g. It suffices to observe that in the proof of (4.1) we do not need to consider (4.6) for $k \geq 2$ since all the J that contribute to this term are contained in $3I_T$. Furthermore, any J that contributes to (4.5) satisfies

$$\frac{\operatorname{dist}(J, I_T)}{|I_T|} \geq C2^k,$$

therefore in the rest of the proof one could easily introduce a decaying factor. ∎

LEMMA 4.2. *Let T be a tree and suppose that any two bitiles of T are disjoint. Then there exists some $C = C(w) < \infty$ such that*

$$\|gC_T f\|_{L^1(w)} \leq Cw(I_T)\operatorname{size}(T)\,\widetilde{\operatorname{density}}(T). \tag{4.11}$$

Proof. Clearly the elements of T must be spatially disjoint using the separation assumption on $\mathbf{P}$ and the fact that T is a tree. Thus, by the triangle inequality it suffices to show (4.11) for any single-element tree; but the improved L^1 tree estimate is clear for these trees. ∎

5. Weighted variational inequalities for Littlewood–Paley families. In this section, we prove weighted extensions of a Lépingle inequality, namely a variational inequality for Littlewood–Paley families [L1, B, JSW, PX]. Note that the dyadic variant of Lemma 5.2 below was proved in [DL].

DEFINITION 5.1. Fix absolute constants $C \in (1,\infty)$ and $\{C_N : N \in \mathbb{N}\}$, $m \geq 1$. A sequence $(f_j)_{j\in\mathbb{Z}}$ of functions is a *Littlewood–Paley family* if each f_j has frequency support inside $\{C^{-1}2^{-j} < |\xi| < C2^{-j}\}$, and

$$\left|\frac{d^N}{dx^N} f_I(x)\right| \leq C_N 2^{-jN}[1 + |x|2^{-j}]^{-m}.$$

LEMMA 5.2. *Let $1 < p < \infty$, $w \in A_p$ and $r \neq 2$. Let $s = \min(r, 2)$. Then for any Littlewood–Paley family (f_j) we have*

$$\Big\|\sup_{K,\,N_0<\cdots<N_K}\Big(\sum_{k=1}^{K}\Big|\sum_{N_{k-1}<j\leq N_k} f_j\Big|^r\Big)^{1/r}\Big\|_{L^p(w)} \leq C\Big\|\Big(\sum_j |f_j|^s\Big)^{1/s}\Big\|_{L^p(w)}. \tag{5.1}$$

Proof. Let Δ_j be the Littlewood–Paley projection of f into an enlarged frequency range $\{(2C)^{-1}2^{-j} < |\xi| < 2C\cdot 2^{-j}\}$ such that $\Delta_j f_j = f_j$. It then suffices to show that for any $w \in A_p$ and any family of Littlewood–Paley

projections (Δ_j) and any vector-valued function $\mathbf{f} = (f_j)_{j\in\mathbb{Z}}$ we have
(5.2)

$$\Big\| \sup_{K,\, N_0<\cdots<N_K} \Big(\sum_{k=1}^{K} \Big| \sum_{N_{k-1}<j\le N_k} \Delta_j f_j \Big|^r \Big)^{1/r} \Big\|_{L^p(w)} \le C \Big\| \Big(\sum_j |f_j|^s \Big)^{1/s} \Big\|_{L^p(w)}.$$

Let $T\mathbf{f}$ denote the variational operator inside $\|\cdot\|_{L^p(w)}$ on the left hand side of (5.2). Then it suffices to show the following pointwise bound for the dyadic sharp maximal function of $T\mathbf{f}$: for any $1 < t < \infty$,

$$(T\mathbf{f})^\sharp(x) \le \mathcal{M}_t(\mathbf{f})(x), \qquad |\mathbf{f}| = \Big(\sum_j |f_j|^s \Big)^{1/s}, \tag{5.3}$$

Indeed, since $w \in A_p$ this will imply that

$$\|T\mathbf{f}\|_{L^p(w)} \le C \|(T\mathbf{f})^\sharp\|_{L^p(w)} \le C \|\mathcal{M}_t(\mathbf{f})\|_{L^p(w)}.$$

We now take $1 < t < p$ sufficiently small such that $w \in A_{p/t}$, and the desired estimate (5.2) then follows:

$$\|T\mathbf{f}\|_{L^p(w)} \le C \|\mathcal{M}_t(\mathbf{f})\|_{L^p(w)} \le C \|\mathbf{f}\|_{L^p(w)}.$$

It remains to show (5.3); we use an argument from [DMT]. Take any dyadic interval I containing x. Let c_j be a constant defined as follows:

$$c_j = \begin{cases} |I|^{-1} \int \phi_j * f_j & \text{if } 2^j < 1/|I|, \\ 0 & \text{otherwise,} \end{cases}$$

where ϕ_j is the corresponding convolution function of Δ_j. Further let

$$c_I = \sup_{K,\, N_0<\cdots<N_K} \Big(\sum_k \Big| \sum_{N_{k-1}<j\le N_k} c_j \Big|^r \Big)^{1/r}.$$

Then it is not hard to see that

$$\Big| \sup_{K,\, N_0<\cdots<N_K} \Big(\sum_k \Big| \sum_{N_{k-1}<j\le N_k} \Delta_j f_j \Big|^r \Big)^{1/r} - c_I \Big|$$
$$\le \sup_{K,\, N_0<\cdots<N_K} \Big(\sum_k \Big| \sum_{N_{k-1}<j\le N_k} (\Delta_j f_j - c_j) \Big|^r \Big)^{1/r}.$$

We then decompose

$$\Delta_j f_j - c_j = g_j + b_j$$

where

$$(g_j, b_j) = \begin{cases} (0, \Delta_j f_j - c_j) & \text{if } 2^j < 1/|I|, \\ (\Delta_j(f_j 1_{3I}), \Delta_j(f_j 1_{(3I)^c})) & \text{otherwise.} \end{cases}$$

It is not hard to see that for any $y \in I$ we have

$$|b_j(y)| \le C\mathcal{M}_1 f_j(x) \min[(2^j|I|)^\epsilon, (2^j|I|)^{-\epsilon}].$$

The parameter $\epsilon > 0$ here depends on the decay of ϕ_j and its derivative. Now, by Hölder's inequality and the known Lebesgue case [4] of (5.2), we have

$$\begin{aligned}
\frac{1}{|I|}\int_I \sup_{K,\,N_0<\cdots<N_K} \Big(\sum_k \Big|\sum_{N_{k-1}<j\le N_k} \Delta_j g_j\Big|^r\Big)^{1/r} \\
&\le \frac{1}{|I|^{1/t}}\Big\|\sup_{K,\,N_0<\cdots<N_K} \Big(\sum_k \Big|\sum_{N_{k-1}<j\le N_k} \Delta_j g_j\Big|^r\Big)^{1/r}\Big\|_t \\
&\le C\frac{1}{|I|^{1/t}}\Big\|\Big(\sum_j |g_j|^s\Big)^{1/s}\Big\|_t \le \frac{1}{|I|^{1/t}}\Big\|\Big(\sum_j |f_j 1_{3I}|^s\Big)^{1/s}\Big\|_t \le \mathcal{M}_t(\mathbf{f})(x).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\frac{1}{|I|}\int_I \sup_{K,\,N_0<\cdots<N_K} \Big(\sum_k \sum_{N_{k-1}<j\le N_k} |b_j|^r\Big)^{1/r} &\le \frac{1}{|I|}\int_I \sum_j |b_j(y)|\,dy \\
&\le C\sum_j \min[(2^j|I|)^{\epsilon}, (2^j|I|)^{-\epsilon}]\mathcal{M}_1 f_j(x) \\
&\le C\sup_j \mathcal{M}_1 f_j(x) \le C\mathcal{M}_1(\mathbf{f})(x) \le C\mathcal{M}_t(\mathbf{f})(x). \ \blacksquare
\end{aligned}$$

6. The main argument and proof of Proposition 2.2. Without loss of generality assume that $w(F) > 0$ and $w(G) > 0$ and

$$\max(w(F), w(G)) = 1.$$

Recall that our aim is to find major subsets of F and G respectively such that at least one of them has full measure, and if $|f|$ and $|g|$ are supported inside these sets and bounded above by 1 then

$$B_{\mathbf{P}}(f,g) \le Cw(F)^{1/p}w(G)^{1-1/p} \tag{6.1}$$

for all $p \in (q,\infty)$ such that $1/r > 1/q - 1/p$. The major subsets will be chosen using the weighted maximal function (see its definition in Section 1.1).

Case 1: $w(F) \le w(G)$. We choose $\widetilde{F} = F$ and $\widetilde{G} = G \setminus \Omega$ with

$$\Omega := \{\mathcal{M}_{1,w} 1_F > Cw(F)\}$$

and $C < \infty$ sufficiently large such that $w(\Omega) < 1/2$.

Fix $q_0 \in (q,\infty)$ very close to q. We use the following estimate whose (rather standard) proof is included later:

---

[4] Note that in the Lebesgue case, (5.2) is equivalent to (5.1) thanks to boundedness of the vector-valued maximal function; this was observed in [DMT].

LEMMA 6.1. *For any* $\eta \in (2q_0/r, 1)$ *there is a positive constant* $\epsilon = \epsilon(\eta, q_0, r)$ *such that*

$$B_{\mathbf{P}}(f,g) \le C\,\mathrm{size}(\mathbf{P})^{1-\eta}\,\mathrm{density}(\mathbf{P})^{\epsilon} w(F)^{\eta/(2q_0)}. \tag{6.2}$$

Furthermore, if the elements of $\mathbf{P}$ *are disjoint in the phase plane then a stronger variant of* (6.2) *holds where* $\widetilde{\mathrm{density}}(\mathbf{P})$ *is used in place of* $\mathrm{density}(\mathbf{P})$*.*

Below we show how Lemma 6.1 implies the desired estimate (6.1) using an argument from [MTT2, MTT3]. We decompose the original $\mathbf{P}$ as $\bigcup_{k\ge 0} \mathbf{P}^{[k]}$ where

$$\mathbf{P}^{[k]} = \Big\{ P \in \mathbf{P} : 2^k \le 1 + \frac{\mathrm{dist}(I_P, \Omega^c)}{|I_P|} < 2^{k+1} \Big\}.$$

Observe that if $P \in \mathbf{P}^{[k]}$ then $2^{k+2} I_P \cap \Omega^c \ne \emptyset$. Therefore, using Lemma 3.7 we obtain

$$\mathrm{size}(\mathbf{P}^{[k]}) \le C 2^{O(k)} w(F)^{1/q}. \tag{6.3}$$

On the other hand, it is not hard to see that

$$\widetilde{\mathrm{density}}(\mathbf{P}^{[k]}) \le C 2^{-Dk/2}.$$

Now, observe that if $k \ge 1$ then the collection $\mathbf{P}^{[k]}$ can be decomposed into $O(1)$ bitile subcollections, such that for any two $P \ne P'$ in a subcollection we have $I_P \times \omega_P \cap I_{P'} \times \omega_{P'} = \emptyset$. To see this, note that for $k \ge 1$ the length of any nested sequence in $\{I_P : P \in \mathbf{P}^{[k]}\}$ must be $O(1)$. It then follows that we can decompose $\mathbf{P}^{[k]}$ into $O(1)$ subcollections, in each collection the spatial intervals I_P of two bitiles are either the same or disjoint, and via another decomposition (to ensure that any two different bitiles sharing the same spatial interval are far from each other in frequency) we can obtain $O(1)$ subcollections with the desired properties.

Thus, for the purpose of proving (6.1) we may assume without loss of generality that for $k > k_0$ the elements of $\mathbf{P}^{[k]}$ are disjoint in the phase plane. For those k we have

$$\begin{aligned} B_{\mathbf{P}^{[k]}}(f,g) &\le C\,\mathrm{size}(\mathbf{P}^{[k]})^{1-\eta} [\widetilde{\mathrm{density}}(\mathbf{P}^{[k]})]^{\epsilon} w(F)^{\eta/(2q_0)} \\ &\le C 2^{-D\epsilon k/2}\,\mathrm{size}(\mathbf{P}^{[k]})^{1-\eta} w(F)^{\eta/(2q_0)} \qquad (\text{since } \mathrm{supp}(g) \subset \Omega^c) \\ &\le C 2^{-D\epsilon k/2} [2^{O(k)} w(F)^{1/q}]^{1-\eta} w(F)^{\eta/(2q_0)}. \end{aligned}$$

Choosing D large in the definition of density (certainly D depends on q, q_0, r, w) we obtain

$$B_{\mathbf{P}^{[k]}}(f,g) \le C 2^{-\epsilon k} w(F)^{(1-\eta)/q + \eta/(2q_0)}, \qquad k > k_0.$$

On the other hand for $0 \le k < k_0$ disjointness may not be available, and we only have $\mathrm{density}(\mathbf{P}^{[k]}) = O(1)$, but since $k_0 = O(1)$ we also have

$\mathrm{size}(\mathbf{P}^{[k]}) = O(w(F)^{1/q})$ from (6.3). Using a similar argument to that presented before, we obtain

$$B_{\mathbf{P}^{[k]}}(f,g) \le Cw(F)^{(1-\eta)/q+\eta/(2q_0)}, \qquad k \le k_0.$$

Thus, summing the above estimates over $k \ge 0$ we obtain

$$B_{\mathbf{P}}(f,g) \le Cw(F)^{(1-\eta)/q+\eta/(2q_0)}.$$

For any p such that

$$\frac{1}{p} < \frac{1}{q} - \frac{1}{r}$$

we can choose q_0 sufficiently close to q and η sufficiently close to $2q_0/r$ (keeping $1 > \eta > 2q_0/r$ and $q_0 > q$) such that

$$(1-\eta)/q + \eta/(2q_0) > 1/p.$$

The desired estimate (6.1) now follows immediately, using $w(F) \le 1$:

$$B_{\mathbf{P}}(f,g) \le Cw(F)^{1/p}.$$

Proof of Lemma 6.1. We only show the general case when $\mathbf{P}$ is arbitrary. An analogous argument is used in the case when any two elements of $\mathbf{P}$ are disjoint in the phase plane, and the estimate is in terms of the improved density. The main difference is the use of the improved tree estimate (Lemma 4.2) in place of the standard tree estimate (Lemma 4.1).

For convenience, we denote $S_1 = \mathrm{size}(\mathbf{P})$, $E_1 = w(F)^{1/(2q_0)}$ and $D_1 = \mathrm{density}(\mathbf{P})$. Using Lemma 3.10 and Lemma 3.15 we can decompose $\mathbf{P} = \bigcup_{n\in\mathbb{Z}} \mathbf{P}_n$ where each $\mathbf{P}_n$ is the union of trees from a tree collection $\mathbf{T}_n$ such that

$$\sum_{T\in\mathbf{T}_n} w(I_T) \le C2^n, \qquad \mathrm{size}(\mathbf{P}_n) \le C2^{-n/(2q_0)}E_1, \qquad \mathrm{density}(\mathbf{P}_n) \le 2^{-n/r'}.$$

The tree estimate (4.3) (applied with L^1 norm) then shows

$$\begin{aligned} B_{\mathbf{P}}(f,g) &\le C\sum_{n\in\mathbb{Z}}\sum_{T\in\mathbf{T}_n} w(I_T)\,\mathrm{size}(T)\,\mathrm{density}(T) \\ &\le C\sum_{n\in\mathbb{Z}} 2^n \min(S_1, 2^{-n/(2q_0)}E_1)\min(D_1, 2^{-n/r'}). \end{aligned}$$

It follows that for $\alpha, \beta \in [0,1]$ we have

$$\begin{aligned} B_{\mathbf{P}}(f,g) &\le CS_1D_1\sum_{n\in Z} \min(1, 2^{-n/(2q_0)}E_1S_1^{-1})^{\alpha}\min(1, 2^{-n/r'}D_1^{-1})^{\beta} \\ &\le CS_1D_1\sum_{n\in\mathbb{Z}} 2^n\min(1, 2^{-nK}(E_1/S_1)^{\alpha}D_1^{-\beta}), \end{aligned}$$

where

$$K := \frac{\alpha}{2q_0} + \frac{\beta}{r'}.$$

Under the assumption $r > 2q$ we can choose $q_0 > q$ such that $r > 2q_0$. Then we can find $\alpha, \beta \in [0, 1]$ such that

$$\frac{\alpha}{2q_0} + \frac{\beta}{r'} > 1. \tag{6.4}$$

We thus obtain a two-sided geometric series which is bounded above by its largest term. Thus

$$B_{\mathbf{P}}(f, g) \le CS_1 D_1 (E_1/S_1)^{\alpha/K} D_1^{-\beta/K} = CS_1^{1-\alpha/K} E_1^{\alpha/K} D_1^{1-\beta/K}.$$

Let $\eta = \alpha/K$; we have $\eta \in (2q_0/r, 1)$ and in fact varying $\alpha, \beta \in [0, 1]$ respecting the condition (6.4) we can obtain any value of η in $(2q_0/r, 1)$. Furthermore

$$\epsilon := 1 - \frac{\beta}{K} = 1 - r'\left(1 - \frac{\eta}{2q_0}\right) = \frac{r'}{2q_0}\left(\eta - \frac{2q_0}{r}\right) > 0,$$

giving the desired estimate (6.2). This completes the proof of Lemma 6.1. ∎

Case 2: $w(F) > w(G)$. We will choose $\widetilde{G} = G$ and $\widetilde{F} = F \setminus \Omega$ where

$$\Omega := \{\mathcal{M}_{1,w} 1_G > Cw(G)\}$$

with $C < \infty$ sufficiently large such that $w(\Omega) < 1/2$. We will use the following estimate, whose proof is included later:

Lemma 6.2. *Suppose that* density$(\mathbf{P}) \le Mw(G)^{1/r'}$ *for some* $M \ge 1$. *Then for any* $p < \infty$ *there exists a constant* $\delta = \delta(p, q, w, r) > 0$ *such that*

$$\mathbf{B}_{\mathbf{P}}(f, g) \le CM \operatorname{size}(\mathbf{P})^{\delta} w(G)^{1/p' - 1/r'}. \tag{6.5}$$

Now we show how Lemma 6.2 implies the desired estimate (6.1). Decompose $\mathbf{P}$ into $\bigcup_{h \ge 0} \mathbf{P}^{[h]}$ where

$$\mathbf{P}^{[h]} = \left\{P \in \mathbf{P} : 2^h \le 1 + \frac{\operatorname{dist}(I_P, \Omega^c)}{|I_P|} < 2^{h+1}\right\}.$$

We verify that

$$\operatorname{density}(\mathbf{P}^{[h]}) \le C2^{O(h)} \Big[\sup_{x \in \Omega^c} (\mathcal{M}_{1,w} 1_G)(x)\Big]^{1/r'} \le C2^{O(h)} w(G)^{1/r'}; \tag{6.6}$$

here the implicit constant in $O(h)$ depends on the doubling exponent γ of w. Indeed, let T be any nonempty tree in $\mathbf{P}^{[h]}$. It is clear that

$$1 + \frac{\operatorname{dist}(I_T, \Omega^c)}{|I_T|} \le 2^{h+1}.$$

We then enlarge I_T by a factor of $O(2^h)$ to obtain an interval J such that

$J \cap \Omega^c \neq \emptyset$; clearly $w(J) \leq C2^{O(h)} w(I_T)$ and therefore

$$\left(\frac{1}{w(I_T)} \int \widetilde{\chi}_{I_T}^D |g|^{r'} w\right)^{1/r'} \leq C2^{O(h)} \Big[\inf_{x \in J} \mathcal{M}_{1,w} 1_G(x)\Big]^{1/r'},$$

from which the estimate (6.6) follows immediately.

On the other hand, since $\operatorname{supp}(f) \subset \Omega^c$, it follows from Lemma 3.7 that

$$\operatorname{size}(\mathbf{P}^{[h]}) \leq C_N 2^{-Nh}$$

for any $N > 0$. Take N very large in the above estimate; then (6.5) and (6.6) imply that

$$\mathbf{B}_{\mathbf{P}^{[h]}}(f,g) \leq C2^{-h} w(G)^{1/r'} w(G)^{1/p'-1/r'} = C2^{-h} w(G)^{1/p'},$$

and (6.1) now follows from summing these estimates over $h \geq 0$.

Proof of Lemma 6.2. Fix $q_0 \in (q, \infty)$. Using Lemmas 3.10 and 3.15, we can decompose $\mathbf{P} = \bigcup_{n \in \mathbb{Z}} \mathbf{P}_n$ where $\mathbf{P}_n$ is the union of trees from a tree collection $\mathbf{T}_n$ such that

$$\sum_{T \in \mathbf{T}_n} w(I_T) \leq 2^n, \quad \operatorname{size}(\mathbf{P}_n) \leq C2^{-n/2q_0}, \quad \operatorname{density}(\mathbf{P}_n) \leq C2^{-n/r'} w(G)^{1/r'}.$$

We use Lemma 3.10 again and decompose $\mathbf{P}_n = \bigcup_{m \geq 0} \mathbf{P}_{n,m}$ where $\mathbf{P}_{n,m}$ is the union of trees from a tree collection $\mathbf{T}_{n,m}$ such that

$$\operatorname{size}(\mathbf{P}_{n,m}) \leq C2^{-(n+m)/(2q_0)},$$

$$\sum_{T \in \mathbf{T}_{n,m}} w(I_T) \leq C \sum_{T \in \mathbf{T}_n} w(I_T) \leq C2^n,$$

$$\Big\| \sum_{T \in \mathbf{T}_{n,m}} 1_{2^k I_T} \Big\|_{L^p(w)} \leq C2^{O(k)} 2^{n+m} w(F)^{1/p} = C2^{O(k)} 2^{n+m}.$$

In particular, it follows from the doubling property of w that

$$\Big\| \sum_{T \in \mathbf{T}_{n,m}} 1_{2^k I_T} \Big\|_{L^1(w)} \leq C2^{\gamma k} 2^n.$$

Consequently, by interpolation, for any $1 < p < \infty$ and any $\epsilon > 0$ we have

$$\Big\| \sum_{T \in \mathbf{T}_{n,m}} 1_{2^k I_T} \Big\|_{L^{p-\epsilon}(w)} \leq C2^{O(k)} 2^{m/p'} 2^n. \tag{6.7}$$

Here, the implicit constant in $O(k)$ may depend on p, ϵ, w. For convenience, for any $k \geq 0$ let $N_{n,m}^{[k]}$ denote the counting function

$$N_{n,m}^{[k]} = \sum_{T \in \mathbf{T}_{n,m}} 1_{2^k I_T}.$$

Decomposing $1 = 1_{I_T} + \sum_{k \geq 0} (1_{2^{k+1} I_T} - 1_{2^k I_T})$ for each T and applying

Hölder's inequality, we obtain

$$B_{\mathbf{P}_{n,m}}(f,g) \le C \sum_{k\ge -1} B_k(n,m)$$

where

$$B_{-1}(n,m) := \int (N^{[0]}_{n,m})^{1/r} \Big(\sum_{T\in \mathbf{T}_{n,m}} |1_{I_T} g C_T f|^{r'} \Big)^{1/r'} w\,dx,$$

$$B_k(n,m) := \int (N^{[k+1]}_{n,m} - N^{[k]}_{n,m})^{1/r} \Big(\sum_{T\in \mathbf{T}_{n,m}} |1_{2^{k+1}I_T \setminus 2^k I_T} g C_T f|^{r'} \Big)^{1/r'} w\,dx$$

for $k \ge 0$.

Estimate for $\sum_{n,m} B_{-1}(n,m)$. Fix $p < \infty$ very large and $\epsilon > 0$ very small, such that in particular $p-\epsilon > r$. Apply Hölder's inequality to obtain

$$B_{-1}(n,m) \le C \|(N^{[0]}_{n,m})^{1/r}\|_{L^{p-\epsilon}(w)} \Big\| \Big(\sum_{T\in \mathbf{T}_{n,m}} |1_{I_T} g C_T f|^{r'} \Big)^{1/r'} \Big\|_{L^{(p-\epsilon)'}(w)}.$$

By (6.7), the first factor can be rewritten as

$$\Big\| \sum_{T\in \mathbf{T}_{n,m}} 1_{I_T} \Big\|^{1/r}_{L^{(p-\epsilon)/r}(w)} \le C 2^{n/r} 2^{(1/r-1/p)m}$$

since $p-\epsilon > r$. The second factor is supported inside $\mathrm{supp}(g) \subset G$, thus it can be bounded above by

$$Cw(G)^{\frac{1}{(p-\epsilon)'} - \frac{1}{r'}} \Big\| \Big(\sum_{T\in \mathbf{T}_{n,m}} |1_{I_T} g C_T f|^{r'} \Big)^{1/r'} \Big\|_{L^{r'}(w)}$$
$$= Cw(G)^{\frac{1}{(p-\epsilon)'} - \frac{1}{r'}} \Big(\sum_{T\in \mathbf{T}_{n,m}} \|1_{I_T} g C_T f\|^{r'}_{L^{r'}(w)} \Big)^{1/r'}.$$

Using the tree estimate (4.1), we can bound the above expression by

$$Cw(G)^{\frac{1}{(p-\epsilon)'} - \frac{1}{r'}} \Big[\sum_{T\in \mathbf{T}_{n,m}} w(I_T) \Big]^{1/r'} \mathrm{size}(\mathbf{P}_{n,m})\, \mathrm{density}(\mathbf{P}_{n,m})$$
$$\le Cw(G)^{\frac{1}{(p-\epsilon)'} - \frac{1}{r'}} \big[2^{\frac{n}{r'}} \big] \big[2^{-(n+m)(\frac{1}{2q_0} - \delta)} \mathrm{size}(\mathbf{P})^{\delta} \big]$$
$$\times \min(2^{-n/r'} w(G)^{1/r'}, Mw(G)^{1/r'}).$$

Here $\delta \in (0, 1/(2q_0))$ is very small and will be specified later.

Since $M \geq 1$, it follows that

$$\sum_{m \geq 0} B_{-1}(m,n)$$
$$\leq CMw(G)^{\frac{1}{(p-\epsilon)'}} \operatorname{size}(\mathbf{P})^{\delta} \sum_{m \geq 0} 2^{\frac{n}{r}+(\frac{1}{r}-\frac{1}{p})m+\frac{n}{r'}-(n+m)(\frac{1}{2q_0}-\delta)} \min(2^{-n/r'},1).$$

Since $r > 2q$ we can always choose $q_0 > q$ such that $r > 2q_0$, and then choose $\delta > 0$ depending on q_0, r and such that

$$\frac{1}{r} - \left(\frac{1}{2q_0} - \delta\right) < 0,$$

which implies $1/r - 1/p - 1/(2q_0) + \delta < 0$. Therefore the above summation over $m \geq 0$ converges, and

$$\sum_n \sum_{m \geq 0} B_{-1}(m,n) \leq CMw(G)^{\frac{1}{(p-\epsilon)'}} \operatorname{size}(\mathbf{P})^{\delta} \sum_{n \in \mathbb{Z}} 2^{n(\frac{1}{r}-\frac{1}{2q_0}+\delta)} \min(1, 2^{\frac{n}{r'}}).$$

Since

$$\frac{1}{r} - \frac{1}{2q_0} < 0 < \frac{1}{r} - \frac{1}{2q_0} + \frac{1}{r'}$$

we can refine our previous choice of $\delta = \delta(q_0, r) > 0$ such that the above estimate of $\sum_n \sum_{m \geq 0} B_{-1}(m,n)$ remains a two-sided geometric series. It follows that

$$\sum_n \sum_{m \geq 0} B_{-1}(m,n) \leq CMw(G)^{\frac{1}{(p-\epsilon)'}} \operatorname{size}(\mathbf{P})^{\delta}.$$

Since we can choose $p < \infty$ arbitrarily large and since $w(G) \leq 1$, it follows that

$$\sum_n \sum_{m \geq 0} B_{-1}(m,n) \leq CMw(G)^{1/p'} \operatorname{size}(\mathbf{P})^{\delta}$$

for any $p < \infty$.

Estimate for $\sum_{n,m} B_k(n,m)$. The argument is similar to the above estimate for the sum of $B_{-1}(n,m)$, with the following difference: we will collect some power 2^k, and we will gain the decay factor 2^{-Nk} from the tree estimate (4.2) where N could be chosen arbitrarily large. We obtain, via a similar argument and by choosing N large enough, the estimate

$$\sum_n \sum_{m \geq 0} B_k(m,n) \leq C2^{-k} Mw(G)^{1/p'} \operatorname{size}(\mathbf{P})^{\delta}$$

for any $p < \infty$.

Summing over $k \geq -1$, we obtain the desired estimate (6.5). This completes the proof of Lemma 6.2. ∎

Acknowledgments. The first author was supported in part by grant NSF-DMS 1201456. The second author was supported in part by grant NSF-DMS 0968499 and a grant from the Simons Foundation (#229596 to Michael Lacey).

References

[BH] J. Bennett and S. Harrison, *Weighted norm inequalities for oscillatory integrals with finite type phases on the line*, Adv. Math. 229 (2012), 2159–2183.

[B] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, Inst. Hautes Études Sci. Publ. Math. 69 (1989), 5–45.

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

[DL] Y. Do and M. Lacey, *Weighted bounds for variational Walsh–Fourier series*, J. Fourier Anal. Appl., to appear.

[DMT] Y. Do, C. Muscalu, and C. Thiele, *Variational estimates for paraproducts*, Rev. Mat. Iberoamer. 28 (2012), 857–878.

[F] C. Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. (2) 98 (1973), 551–571; Erratum, ibid. 146 (1997), 239.

[GMS] L. Grafakos, J. M. Martell, and F. Soria, *Weighted norm inequalities for maximally modulated singular integral operators*, Math. Ann. 331 (2005), 359–394.

[H] R. A. Hunt, *On the convergence of Fourier series*, in: Orthogonal Expansions and their Continuous Analogues (Edwardsville, IL, 1967), Southern Illinois Univ. Press, Carbondale, IL, 1968, 235–255.

[HY] R. A. Hunt and W. S. Young, *A weighted norm inequality for Fourier series*, Bull. Amer. Math. Soc. 80 (1974), 274–277.

[HL] T. P. Hytönen and M. Lacey, *Pointwise convergence of vector-valued Fourier series*, arXiv:1205.0261.

[JSW] R. L. Jones, A. Seeger, and J. Wright, *Strong variational and jump inequalities in harmonic analysis*, Trans. Amer. Math. Soc. 360 (2008), 6711–6742.

[LT1] M. Lacey and C. Thiele, *On Calderón's conjecture*, Ann. of Math. (2) 149 (1999), 475–496.

[LT2] M. Lacey and C. Thiele, *A proof of boundedness of the Carleson operator*, Math. Res. Lett. 7 (2000), 361–370.

[LT3] M. Lacey and C. Thiele, *L^p estimates on the bilinear Hilbert transform for* $2 < p < \infty$, Ann. of Math. (2) 146 (1997), 693–724.

[L1] D. Lépingle, *La variation d'ordre p des semi-martingales*, Z. Wahrsch. Verw. Gebiete 36 (1976), 295–316.

[L2] X. Li, personal communication.

[L2] B. Muckenhoupt, *Weighted norm inequalities for the Hardy maximal function*, Trans. Amer. Math. Soc. 165 (1972), 207–226.

[MTT1] C. Muscalu, T. Tao, and C. Thiele, *Multi-linear operators given by singular multipliers*, J. Amer. Math. Soc. 15 (2002), 469–496.

[MTT2] C. Muscalu, T. Tao, and C. Thiele, *The bi-Carleson operator*, Geom. Funct. Anal. 16 (2006), 230–277.

[MTT3] C. Muscalu, T. Tao, and C. Thiele, *L^p estimates for the biest. II. The Fourier case*, Math. Ann. 329 (2004), 427–461.

[OST+] R. Oberlin, A. Seeger, T. Tao, C. Thiele, and J. Wright, *A variation norm Carleson theorem*, J. Eur. Math. Soc. 14 (2012), 421–464.

[PX] G. Pisier and Q. Xu, *The strong p-variation of martingales and orthogonal series*, Probab. Theory Related Fields 77 (1988), 497–514.

[RdF] L. Rubio de Francia, *A Littlewood–Paley inequality for arbitrary intervals*, Rev. Mat. Iberoamer. 1 (1985), no. 2, 1–14.

[S] P. Sjölin, *An inequality of Paley and convergence a.e. of Walsh–Fourier series*, Ark. Mat. 7 (1969), 551–570.

[SW] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton Math. Ser. 32, Princeton Univ. Press, Princeton, NJ, 1971.

[T1] C. Thiele, *On the bilinear Hilbert transform*, Habilitationsschrift, Kiel, 1998.

[T2] C. Thiele, *Wave Packet Analysis*, CBMS Reg. Conf. Ser. Math. 105, Amer. Math. Soc., Providence, RI, 2006.

Yen Do
Department of Mathematics
Yale University
New Haven, CT 06511, U.S.A.
E-mail: yenquang.do@yale.edu

Michael Lacey
School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.
E-mail: lacey@math.gatech.edu

Received July 19, 2012 (7575)