VARIATION-NORM AND FLUCTUATION ESTIMATES FOR ERGODIC BILINEAR AVERAGES

YEN DO   RICHARD OBERLIN   EYVINDUR A. PALSSON

ABSTRACT. For any dynamical system, we show that higher variation-norms for the sequence of ergodic bilinear averages of two functions satisfy a large range of bilinear L^p estimates. It follows that, with probability one, the number of fluctuations along this sequence may grow at most polynomially with respect to (the growth of) the underlying scale. These results strengthen previous works of Lacey and Bourgain where almost surely convergence of the sequence was proved (which is equivalent to the qualitative statement that the number of fluctuations is finite at each scale). Via transference, the proof reduces to establishing new bilinear L^p bounds for variation-norms of truncated bilinear operators on $\mathbb{R}$, and the main new ingredient of the proof of these bounds is a variation-norm extension of maximal Bessel inequalities of Lacey and Demeter–Tao–Thiele.

1. INTRODUCTION

Let T be an invertible bi-measurable measure-preserving transformation on a complete probability space (X, Ω, μ). Given two measurable functions f_1, f_2 on X, we consider their ergodic bilinear averages, namely

$$M_k[f_1, f_2](x) \;=\; \frac{1}{k} \sum_{n=0}^{k-1} f_1(T^n x) f_2(T^{-n} x) \;\; (k = 1, 2, \dots) \;.$$

It was shown by Bourgain in [2] that if $f_1, f_2 \in L^\infty(X)$ then $(M_k[f_1, f_2](x))_{k \geqslant 1}$ is convergent for μ-almost every $x \in X$. Thanks to a bilinear maximal function estimate of Lacey [14], Bourgain's result remains valid for $(f_1, f_2) \in L^{p_1} \times L^{p_2}$ for every (p_1, p_2, q) satisfying

$$(1) \qquad \frac{1}{q} = \frac{1}{p_1} + \frac{1}{p_2}, \; \frac{2}{3} < q < \infty, \;\; 1 < p_1, p_2 \leqslant \infty \;,$$

and this has been regarded as a bilinear analogue of the classical Birkhoff ergodic theorem. A similar result also holds for a variant of M_k (namely the ergodic bilinear Hilbert transform), see Demeter [3] and Demeter–Tao–Thiele [7].

Our aim in this paper is to further demonstrate that the sequence $M_k[f_1, f_2](x)$, $k \geqslant 1$, converges rapidly. To formulate a consequence of our estimates, we recall the notion of fluctuations of a given sequence $(a_1, a_2, \dots)$. Given a scale $\lambda > 0$, the number of fluctuations in (a_k) with respect to this scale is the largest number ℓ such that there exists ℓ disjoint intervals

$$[n_1, m_1), [n_2, m_2), \dots [n_\ell, m_\ell)$$

with the following properties: for every $1 \leqslant j \leqslant \ell$ it holds that $|a_{m_j} - a_{n_j}| \geqslant 1/\lambda$. It follows from the Cauchy criteria that (a_k) is convergent if any only if it has a finite number of fluctuations at every (finite) scale. Thus results of [2, 14] could be interpreted as saying that: for almost every $x \in X$, at every scale, the number of fluctuations along $M_k(f, g)(x)$ is finite. It turns out that this number grows at most polynomially as $\lambda \to \infty$.

Theorem 1.1. *Assume that p_1, p_2, q satisfying (1). Then there exists $R < \infty$ such that for every $f_1 \in L^{p_1}$ and $f_2 \in L^{p_2}$ the following holds: for almost every $x \in X$ the number of fluctuations in the sequence $(M_k[f_1, f_2](x))_{k \geqslant 1}$ at any scale $\lambda > 0$ is bounded above by $O(\lambda^R)$, where the implicit constant is uniform over λ but could depends on x and f_1, f_2.*

For an interesting discussion about applications of fluctuation estimates in ergodic theory, we refer the readers to Avigad–Rute [1] (cf. Kovac [11]).

Theorem 1.1 is an immediate consequence of Theorem 1.2 below, which provides a more quantitative estimate. To formulate this result, we recall the notion of variation-norm. Given $\Omega \subset \mathbb{R}$ and $a : \Omega \to \mathbb{C}$, let its r-variation norm be

$$\|a(t)\|_{V_t^r(\Omega)} \quad := \quad \sup_{n, N_0 < \cdots < N_n} (|a(N_0)|^r + \sum_{j=1}^n |a(N_j) - a(N_{j-1})|^r)^{1/r} \quad ,$$

in the sup we require $N_j \in \Omega$ for every j. We also use the semi-norm variant $\widetilde{V}^r$ defined similarly without the first term $|a(N_0)|^r$.

Theorem 1.2. *Assume that p_1, p_2, q satisfying (1). Then there exists $R < \infty$ such that the following holds for every $r > R$:*

$$\|M_k[f_1, f_2](x)\|_{L_x^q(V_k^r)} \quad \lesssim \quad \|f_1\|_{p_1} \|f_2\|_{p_2}$$

Via a modification of standard transference arguments (which we will detail in Section 2), Theorem 1.2 follows from L^p estimates for bilinear singular integrals, Theorem 1.3 below. To formulate the result, we fix some notations.

Given $K : \mathbb{R} \to \mathbb{C}$ sufficiently nice, consider the bilinear operator with kernel K

$$(2) \qquad B[f_1, f_2](x) = \int_{\mathbb{R}} f_1(x + y) f_2(x - y) K(y) \, dy \quad ,$$

which is *a priori* well-defined for Schwarz functions f_1 and f_2. For any $t > 0$ let B_t be the bilinear operator with kernel $t^{-1} K(t^{-1} y)$.

We will be interested in $K : \mathbb{R} \to \mathbb{C}$ such that the following properties hold uniformly over $\xi \neq 0$:

$$(3) \qquad |\widehat{K}(\xi)| \quad \lesssim \quad \min(1, \frac{1}{|\xi|}) \quad ,$$

$$(4) \qquad |\frac{d^n}{d\xi^n} \widehat{K}(\xi)| \quad \lesssim_n \quad \min(\frac{1}{|\xi|^{n-1}}, \frac{1}{|\xi|^{n+1}}), \qquad n \geqslant 1 \quad .$$

We will in fact work with K where (4) holds for $1 \leqslant n \leqslant n_0$, here n_0 is some given large number; now the implicit constants are allowed to depend on n_0. In this case, we will say that K satisfies (3) and (4) up to order n_0.

Theorem 1.3. *Assume that p_1, p_2, q satisfies* (1) *and $r > 2$. Then there exists n_0 finite such that if K satisfies* (3) *and* (4) *up to order n_0 then*

$$\|B_t(f_1, f_2)(x)\|_{L_x^q(V_t^r)} \lesssim \|f_1\|_{L^{p_1}(\mathbb{R})} \|f_2\|_{L^{p_2}(\mathbb{R})} \quad,$$

where the implicit constant may depend on n_0 and on the implicit constants of (3) *and* (4) *for $1 \leqslant n \leqslant n_0$.*

Comparing Theorem 1.3 with Theorem 1.2, it can be seen that there is a discrepancy between the two ranges $r > 2$ and $r > R$. With the current transference techniques, it seems that to get the range $r > 2$ for Theorem 1.2 one would need a version of Theorem 1.3 that accommodates rougher K's, such as $K(y) = 1_{|y| \leqslant 1}$, which would be an interesting open problem left for future studies. In fact, in our transference argument we also prove a weaker version of Theorem 1.3 for this particular K where instead of $r > 2$ we only have $r > R$ for some finite R, see Theorem 2.1.

Our proof of Theorem 1.3 could be viewed as a variation-norm extension of Lacey's proof of the boundedness of the bilinear maximal function in [14], although we will follow more closely the expositions in Demeter–Tao–Thiele [7] and Demeter [3]. The main new ingredient of the proof (compared to [14, 3, 7]) is a variation-norm extension of maximal Bessel inequality for phase plane projections, which in turn relies on variation-norm estimates for Fourier projection operators associated with a collection of frequencies. Maximal estimates for these multi-frequency projection operators were introduced in Bourgain [2], and variation-norm estimates for smooth multi-frequency Fourier projections were also considered in [17]. In our context, it turns out that we need variation-norm estimates for sharp multi-frequency Fourier projections, similar to the original settings considered by Bourgain. On the other hand, L^2 bounds would be sufficient for our purpose, and these estimates are proved in Theorem 8.1 by adapting an argument in [17].

We mention some closely related works in addition to [2, 14, 7, 3]. A dyadic version of Theorem 1.3 was considered in our previous work [9] (which in turn is an adaptation of Thiele [18] to the variation-norm setting). The method of proof in Demeter [3] relies on a weaker version of Theorem 1.3 where the variation-norms are replaced by finitary oscillation norms, which were also used by Demeter–Lacey–Tao–Thiele [6] (see also Demeter [4, 5], Nazarov–Oberlin–Thiele [17]) to improve the L^p ranges in the Bourgain return time theorem. For a nice introduction to variation-norm estimates in harmonic analysis, see Jones–Seeger–Wright [10]. The time-frequency analysis framework used in our proof originated from Lacey–Thiele's proof of the boundedness of the bilinear Hilbert transform [12, 13].

1.1. **Outline of the paper.** In Section 2 we detail the transference argument that deduces Theorem 1.2 from Theorem 1.3. In Section 3 we discuss how a short-long decomposition of the variation-norm leads to a reduction of Theorem 1.3 to two sub-theorems, which respectively treat the contribution of the long-jumps and the contribution of the short-jumps. The proof of these Theorems will use restricted weak-type interpolation methods, which we recall in Section 4.1. In Section 5 we recall standard terminologies in time-frequency analysis, which will be used in Section 6 to describe some wave packet representation for the operators underlying the long-jump and short-jump contributions. Some old and new auxiliary estimates will be recalled and proved in Section 7, Section 8, Section 9. In Section 11

we prove a new variation-norm extension of the maximal Bessel inequalities of Lacey [14] and Demeter–Tao–Thiele [7], which will be used in Section 10 and Section 12 to prove the desired estimates for the contribution of the long-jumps. In Section 13 we briefly discuss the needed cosmetic changes that could be applied (to the treatment of the long-jump contribution) to get the desired estimates for the short-jump contributions.

1.2. **Notational convention.** Given an interval I, we let $c(I)$ denote the center of the interval, and for each constant $C > 0$ we let CI denote the dilate of I around its center by the factor C. We will use $\mathbf{e}$ and $\mathbf{i}$ to refer to the numbers $\exp(1)$ and $\sqrt{-1}$ respectively, leaving their non-boldfaced counterparts free for other purposes.

For every interval I let $\widetilde{\chi}_I(x) = (1 + (\frac{x-c(I)}{|I|})^2)^{-2}$.

For each $s \geqslant 1$ we let $\mathcal{M}^s$ denote the L^s Hardy-Littlewood maximal operator

$$\mathcal{M}^s[f](x) \;:=\; \sup_R \left(\frac{1}{2R} \int_{x-R}^{x+R} |f(y)|^s \, dy \right)^{1/s}$$

and abbreviate $\mathcal{M} := \mathcal{M}^1$.

Throughout the paper we let $\mathcal{F}$ denote the Fourier transform

$$(5) \qquad \widehat{h}(\xi) \;\equiv\; \mathcal{F}[h(.)](\xi) \;:=\; \int_{\mathbb{R}} \mathbf{e}^{-\mathbf{i}2\pi\xi x} h(x) dx \;\;.$$

Note that with this normalization we have

$$h(x) \;=\; \int_{\mathbb{R}} \mathbf{e}^{\mathbf{i}2\pi x\xi} \widehat{h}(\xi) d\xi \;\;.$$

2. The transference argument

In this section we deduce Theorem 1.2 from Theorem 1.3 using a variant of standard transference arguments in [2, 7]. Our first step is to show that the continuous version Theorem 1.2 holds, namely

Theorem 2.1. *For every $t > 0$ let S_t denote the following operator*

$$S_t[f_1, f_2](x) \;=\; \frac{1}{t} \int_0^t f_1(x+t) f_2(x-t) dt \;\;.$$

Then for every (p_1, p_2, q) satisfying (1) there exists $R < \infty$ such that for every $r > R$ it holds that

$$(6) \qquad \|S_t[f_1, f_2](x)\|_{L_x^q(V_t^r)} \;\lesssim\; \|f_1\|_{p_1} \|f_2\|_{p_2} \;\;.$$

Proof of Theorem 2.1. If (p_1, p_2, q) that satisfies (1) we let $n_0 = n_0(p_1, p_2)$ be the constant required in Theorem 1.3.

Fix r below. We divide the proof into two steps.

Step 1: Let $R_0 = 2(1 + (n_0 + 1)\frac{u_0}{u_0-1})$ where $u_0 = \min(p_1, p_2, 2q) > 1$. We first show that for $r > R_0$ it holds that

$$(7) \qquad \|\sup_{\lambda>0} \lambda N(S_t, \lambda)^{\frac{1}{r}}\|_{L_x^q} \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2}$$

Clearly, we may find $1 < u < u_0$ and $r_0 > 2$ such that $r > r_0(1 + (n_0 + 1)\frac{u}{u-1})$. For brevity, let $n_1 = (n_0 + 1)\frac{u}{u-1}$.

Now, for each $0 < \alpha \leqslant 1/2$ let K_α be a C^∞ function supported in $[0, 1]$ such that $1_{\alpha \leqslant y \leqslant 1-\alpha} \leqslant K_\alpha(y) \leqslant 1$, we may construct K such that $|K_\alpha^{(n)}| \lesssim \alpha^{-n}$ for any $n \geqslant 1$.

It is clear that for any $n \geqslant 0$ and $k \geqslant 0$ we have

$$\frac{d^n}{d\xi^n} \widehat{K}_\alpha(\xi) \lesssim_{n,k} \alpha^{-k}(1 + |\xi|)^{-k} .$$

Therefore $\alpha^{n_0+1} K_\alpha$ satisfies the assumptions (3) and (4) up to order n_0 (we emphasize that the implicit constants are independent of α). Let $B_{t,\alpha}$ denote the bilinear operator with kernel $\frac{1}{t} K_\alpha(\frac{y}{t})$. It follows that for any $r_0 > 2$ we have

$$(8) \qquad \alpha^{n_0+1} \|B_{\alpha,t}[f_1, f_2](x)\|_{L_x^q(V_t^{r_0})} \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2} .$$

Let S^* denote the positive maximal version of S_t, namely

$$S^*[f_1, f_2](x) = \sup_{t>0} S_t[|f_1|, |f_2|](x) .$$

By the bilinear maximal estimate of Lacey, it holds that

$$\|S^*[f_1, f_2]\|_{L^q} \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2} .$$

Let $u > 1$ be such that $u < \min(p_1, p_2, 2q)$, then applying the above estimate for the triple $(\frac{p_1}{u}, \frac{p_2}{u}, \frac{q}{u})$ we obtain

$$(9) \qquad \|S^*[|f_1|^u, |f_2|^u]^{1/u}\|_{L^q} \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2}$$

Now, for brevity in the following we understand that $S_t = S_t[f_1, f_2](x)$, $B_{\alpha,t} = B_{\alpha,t}[f_1, f_2](x)$, $S^* = S^*[f_1, f_2](x)$, and $S^{*,u} = S^*[|f_1|^u, |f_2|^u](x)^{1/u}$.

Given any sequence (or functions) $\{a(t), t \in \Omega\}$, let $N(a, \lambda)$ be the number of fluctuations with respect to scale $1/\lambda$, i.e. the largest k such that there exists a sequence of k disjoint intervals $[N_0, N_1), \ldots, [N_{k-1}, N_k)$, where each $N_j \in \Omega$ and furthermore $|a_{N_j} - a_{N_{j-1}}| > \lambda$ for every $1 \leqslant j \leqslant k$.

For any $t > 0$, using Holder's inequality we have

$$|S_t - B_{\alpha,t}| \leqslant (2\alpha)^{(u-1)/u} S^{*,u} .$$

Let $\beta := (2\alpha)^{(u-1)/u}$, we have

$$(10) \qquad N(S_t, 3\beta S^{*,u}) \leqslant N(B_{\alpha,t}, \beta S^{*,u})$$

here the fluctuation counts are used with respect to the t variable. Using the basic estimate $\lambda N(a, \lambda)^{1/r_0} \lesssim \|a\|_{V^{r_0}}$ and using (8), for every $r_0 > 2$ we have

$$\left\| \beta^{1+n_1} S^{*,u} \cdot N\left(S_t, 3\beta S^{*,u}\right)^{1/r_0} \right\|_{L_x^q} =$$

$$= (2\alpha)^{1+n_0} \left\| \beta S^{*,u} \cdot N\left(S_t, 3\beta S^{*,u}\right)^{1/r_0} \right\|_{L_x^q} \lesssim$$

$$\lesssim \alpha^{1+n_0} \|B_{\alpha,t}[f_1, f_2](x)\|_{L_x^q(V_t^{r_0})} \lesssim \|f_1\|_{p_1} \|f_2\|_{p_2} .$$

Using the Holder inequality and (9), it follows that

$$\left\| \beta S^{*,u} \cdot N(S_t, 3\beta S^{*,u})^{\frac{1}{r_0(1+n_1)}} \right\|_{L_x^q} \lesssim$$

$$\lesssim \|S^{*,u}\|_{L_x^q}^{\frac{n_1}{1+n_1}} \cdot \left\| \beta^{1+n_1} S^{*,u} \cdot N\left(S_t, 3\beta S^{*,u}\right)^{\frac{1}{r_0}} \right\|_{L_x^q}^{\frac{1}{1+n_1}}$$

therefore

$$(11) \qquad \|\beta S^{*,u} \cdot N(S_t, 3\beta S^{*,u})^{\frac{1}{r_0(1+n_1)}}\|_{L_x^q} \quad \lesssim \quad \|f_1\|_{p_1} \|f_2\|_{p_2}$$

We note that this estimate holds for any $0 \leqslant \beta \leqslant 1$. Letting $\beta = 2^{-k}/3$, $k \geqslant 0$, and using the triangle inequality it follows that

$$\|\sum_{k \geqslant 0} 2^{-(1+\epsilon)k} S^{*,u} N\left(S_t, 2^{-k} S^{*,u}\right)^{\frac{1}{r_0(1+n_1)}}\|_{L_x^q} \quad \lesssim \quad \|f_1\|_{p_1} \|f_2\|_{p_2} \quad .$$

Since $N(S_t, \lambda) = 0$ for $\lambda \geqslant 2S^*$, and since $S^{*,u} \geqslant S^*$, it follows that

$$\|(S^{*,u})^{-\epsilon} \sup_{\lambda > 0} \lambda^{1+\epsilon} N(S_t, \lambda)^{\frac{1}{r_0(1+n_1)}}\|_{L_x^q} \quad \lesssim \quad \|f_1\|_{p_1} \|f_2\|_{p_2} \quad .$$

Using Holder's inequality and using (9), we obtain

$$\|\sup_{\lambda > 0} \lambda N(S_t, \lambda)^{\frac{1}{r_0(1+n_1)(1+\epsilon)}}\|_{L_x^q} \quad \lesssim \quad \|f_1\|_{p_1} \|f_2\|_{p_2}$$

therefore by choosing ϵ small so that $r > (1 + \epsilon)r_0(1 + n_1)$ we obtain (7).

Step 2: We now prove (6); the argument below is similar to an argument in [8]. We plan to use bilinear Marcinkiewicz interpolation: given each (p_1, p_2, q) satisfying (1) we may let R to be the largest R_0 of the exponents associated with any four rectangular weak-type endpoints. Let $r > R$, then we could use (7) at all of these weak-type endpoints. By monotone convergence it suffices to show that for any increasing sequence of measurable functions (N_k) it holds that

$$\|(\sum_k |S_{N_k} - S_{N_{k-1}}|^r)^{1/r}\|_q \quad \lesssim \quad \|f_1\|_{p_1} \|f_2\|_{p_2} \quad .$$

Let $T[f_1, f_2] = (\sum_k |S_{N_k}[f_1, f_2] - S_{N_{k-1}}[f_1, f_2]|^r)^{1/r}$. By bilinear interpolation it suffices to prove the weak-type estimate

$$\lambda |\{x : T[f_1, f_2](x) > \lambda\}|^{\frac{1}{q}} \quad \lesssim \quad \|f_1\|_{p_1} \|f_2\|_{p_2}$$

with uniform implicit constants over $\lambda > 0$. By scaling symmetries and dilation symmetry of S_t, we may assume $\lambda = \|f_1\|_{p_1} = \|f_2\|_{p_2} = 1$. Let

$$E \quad = \quad \{x : \sup_k |S_{N_k} - S_{N_{k-1}}| > 1\}$$

Clearly $|E| \leqslant \|N(S_t, 1)\|_q \lesssim 1$. For $x \notin E$, we estimate $T[f_1, f_2](x)$ by considering level sets for $|S_{N_k} - S_{N_{k-1}}|$ (as a function of k) and obtain:

$$T[f_1, f_2](x)^q \quad \lesssim \quad \left(\sum_{j \geqslant 0} 2^{-jr} N(S_t, 2^{-j})\right)^{q/r}$$

$$\lesssim_{\epsilon,q} \quad \sum_{j \geqslant 0} 2^{-(1-\epsilon)qj} N(S_t, 2^{-j})^{q/r}$$

Therefore by the Chebysheff inequality we obtain

$$|\{x : T[f_1, f_2](x) > 1\}|^{1/q} \quad \lesssim \quad |E|^{1/q} + |\{x \notin E : T[f_1, f_2](x) > 1\}|^{1/q}$$

$$\lesssim \quad 1 + (\int \sum_{j \geqslant 0} 2^{-(1-\epsilon)qj} N(S_t, 2^{-j})^{q/r} dx)^{1/q} \quad .$$

Using (7) for $\widetilde{r} = \frac{r}{1+\delta}$ where $\delta > 0$ is sufficiently small so that $\widetilde{r} > R$, we have

$$\int N(S_t, 2^{-j})^{q/r} dx = \|N(S_t, 2^{-j})^{1/\widetilde{r}}\|_{L^{q/(1+\delta))_x}}^{q/(1+\delta)} \lesssim 2^{jq/(1+\delta)}$$

Therefore

$$|\{x : T[f_1, f_2](x) > 1\}|^{1/q} \lesssim 1 + (\sum_{j \geqslant 0} 2^{-(1-\epsilon)qj} 2^{jq/(1+\delta)})^{1/q} \lesssim 1$$

by choosing $\epsilon > 0$ sufficiently small depending on δ (which in turn depends on r and R). This completes the proof of (6). $\square$

We now transfer Theorem 2.1 to the integers. Fix $r > R$. We'll show that for any two sequences $f_1(n)$ and $f_2(n)$ indexed by $\mathbb{Z}$ it holds that

$$\|\widetilde{M}_k[f_1, f_2](n)\|_{L_n^q(V_k^r)} \lesssim \|f_1\|_{\ell^{p_1}(\mathbb{Z})} \|f_2\|_{\ell^{p_2}(\mathbb{Z})} ,$$

$$\widetilde{M}_k[f_1, f_2](n) := \frac{1}{k} \sum_{m=0}^{k-1} f_1(n+m) f_2(n-m) .$$

To see this, let S_t be the bilinear operator defined in (2) with kernel $t^{-1} 1_{0<y<t}$, where $t > 0$. We extend f_1 and f_2 from $\mathbb{Z}$ to $\mathbb{R}$ by letting:

(i) $F_1(x) = f_1(n)$ if there exists $n \in \mathbb{Z}$ such that $|x - (n + 1/2)| < 1/3$, and $F_1(x) = 0$ otherwise;

(ii) $F_2(x) = f_2(n)$ if there exists $n \in \mathbb{Z}$ such that $|x - (n - 1/2)| < 1/3$, and $F_2(x) = 0$ otherwise.

Let $n \in \mathbb{Z}$ and $x \in [n - \frac{1}{6}, n + \frac{1}{6}]$. Then for any $m \in \mathbb{Z}$ it holds that

$$\int_{m<y<m+1} F_1(x+y) F_2(x-y) dy = \frac{2}{3} f_1(n+m) f_2(n-m)$$

Thus for any $k \geqslant 0$ we have

$$S_k(F_1, F_2)(x) = \frac{1}{k} \int_{0<y<k} F_1(x+y) F_2(x-y) dy = \frac{2}{3} \widetilde{M}_k[f_1, f_2](n) ,$$

and consequently

$$\|\widetilde{M}_k[f_1, f_2](n)\|_{V_k^r} \lesssim \inf_{x \in [n - \frac{1}{6}, n + \frac{1}{6}]} \|S_k(F_1, F_2)(x)\|_{V_k^r} .$$

It follows that

$$\|\widetilde{M}_k[f_1, f_2](n)\|_{L_n^q(V_k^r)} \lesssim \|S_t(F_1, F_2)(x)\|_{L_x^q(V_t^r)} ,$$

and using Theorem 2.1 we can bound the right hand side by

$$\lesssim \|F_1\|_{L^{p_1}(\mathbb{R})} \|F_2\|_{L^{p_2}(\mathbb{R})} = C' \|f_1\|_{\ell^{p_1}(\mathbb{Z})} \|f_2\|_{\ell^{p_2}(\mathbb{Z})} .$$

Our next step is to transfer the result on $\mathbb{Z}$ to a more general setting. Let T be a measure-preserving transformation on a complete probability space (X, Ω, μ). Let f and g be given.

Fix a large integer N, which we will send to ∞ later. All implicit constants below are independent of N and x.

For fixed x, let $M(f, g, N)(x)$ be the r-variation norm of the finite sequence indexed by $0 \leqslant k \leqslant N$:

$$\frac{1}{k} \sum_{0 \leqslant m \leqslant k-1} f(T^m x) g(T^{-m} x) \qquad .$$

Note that for every $0 \leqslant n \leqslant N$ the value of $M(f, g, N)(T^n x)$ depends only on $f(T^m x)$ and $g(T^m x)$ with $|m| \leqslant 2N$. Thus, using the $\mathbb{Z}$-result, it follows that

$$\sum_{|n| \leqslant N} |M(f, g, N)(T^n x)|^q \quad \lesssim \quad \Big(\sum_{|m| \leqslant 2N} |f(T^m x)|^{p_1} \Big)^{q/p_1} \Big(\sum_{|m| \leqslant 2N} |g(T^m x)|^{p_2} \Big)^{q/p_2}$$

Integrating over $x \in X$ and using the Hölder inequality, we obtain

$$\sum_{|n| \leqslant N} \int_X |M(f, g, N)(T^n x)|^q d\mu(x) \quad \lesssim$$

$$\lesssim \quad \Big(\sum_{|m| \leqslant 2N} \int_X |f(T^m x)|^{p_1} d\mu(x) \Big)^{q/p_1} \Big(\sum_{|m| \leqslant 2N} \int_X |g(T^m x)|^{p_2} d\mu(x) \Big)^{q/p_2} \quad .$$

Using the fact that T is bi-measure preserving on (X, μ), we obtain

$$\|M(f, g, N)\|_{L^q(X,\mu)} \quad \lesssim \quad \|f\|_{L^{p_1}(X,\mu)} \|g\|_{L^{p_2}(X,\mu)} \quad ,$$

and by sending $N \to \infty$ we obtain the conclusion of Theorem 1.2. This completes the transference argument, and the rest of the paper is devoted to the proof of Theorem 1.3. We'll assume that K satisfies (3) and (4) up to some large order that may depend on p_1, p_2, q, r. We will also free the symbol S_t which could be used in the future for different purposes.

3. Separation of short and long jumps

For any function $a(t)$ on $\mathbb{R}$ it is not hard to see that

$$\|a(t)\|_{V_t^r} \quad \lesssim \quad \|a(t)\|_{S_t} + \|a(2^n)\|_{V_n^r(\mathbb{Z})} \quad ,$$

$$\|a(t)\|_{S_t} \quad := \quad \Big(\sum_{n \in \mathbb{Z}} \|a(t)\|_{\widetilde{V}_t^2([2^n, 2^{n+1}])}^2 \Big)^{1/2}$$

Applying this estimate to $a(t) = B_t[f_1, f_2](x)$, the proof of Theorem 1.3 is divided into two parts: the first part handle the long-jumps (i.e. $\|a(2^n)\|_{V_n^r(\mathbb{Z})}$) and the second part handles the short jumps (i.e. $\|a(t)\|_{S_t}$).

Theorem 3.1. *For any $r > 2$ and p_1, p_2, q satisfying (1) it holds that*

$$\|B_{2^n}(f_1, f_2)(x)\|_{L^q_{x \in \mathbb{R}}(V_{n \in \mathbb{Z}}^r)} \quad \lesssim_{p_1, p_2, r} \quad \|f_1\|_{p_1} \|f_2\|_{p_2}$$

Theorem 3.2. *Assume that p_1, p_2, q satisfy (1) and $r > 2$. Assume that $K_s, 1 \leqslant s \leqslant 2$ is a family of kernels such that K_s satisfies (3), (4) up to a high order $n_0 = n_0(p_1, p_2, q, r)$, and furthermore*

$$(12) \qquad \qquad \widehat{K}_s(\xi) \lesssim |\xi| \quad , \quad \xi \neq 0 \quad ,$$

and the implicit constants are uniform over $1 \leqslant s \leqslant 2$. Then it holds that

$$\Big\| \int_1^2 \Big(\sum_{n \in \mathbb{Z}} \Big| \int f_1(\cdot + y) f_2(\cdot - y) 2^{-n} K_s(2^{-n} y) dy \Big|^2 \Big)^{1/2} ds \Big\|_q \quad \lesssim \quad \|f_1\|_{L^{p_1}(\mathbb{R})} \|f_2\|_{L^{p_2}(\mathbb{R})} \quad .$$

Theorem 3.1 immediately takes care of the long-jump component of the variation norm $\|B_t[f,g]\|_{V_t^r}$. Below we deduce the desired estimate for the short jump component from Theorem 3.2.

We first note that if $a(t)$ is differentiable then using $\|.\|_{\widetilde{V}^2} \leqslant \|.\|_{\widetilde{V}^1}$ we obtain

$$\|a(t)\|_{\widetilde{V}_t^2[2^n,2^{n+1}]} \lesssim \|a'(t)\|_{L_t^1[2^n,2^{n+1}]} = 2^n\|a'(2^n s)\|_{L_s^1[1,2]} .$$

Therefore using Minkowski's inequality we have

$$\|a(t)\|_{S_t} \lesssim \|(\sum_{n\in\mathbb{Z}} 2^{2n}|a'(2^n s)|^2)^{1/2}\|_{L_s^1[1,2]}$$

$$\lesssim \|(\sum_{n\in\mathbb{Z}} |(2^n s)a'(2^n s)|^2)^{1/2}\|_{L_s^1[1,2]}$$

We plan to apply the estimate to $a(t) = B_t[f_1,f_2](x)$ where x is fixed. Let $h(y) = -(K(y) + yK'(y))$, or equivalently $\widehat{h}(\xi) = \xi\frac{d}{d\xi}\widehat{K}(\xi)$. Let H_t be the bilinear singular integral with kernel $t^{-1}h(t^{-1}y)$. Then

$$t\frac{d}{dt}B_t[f_1,f_2](x) = H_t[f_1,f_2](x) ,$$

therefore

$$\|B_t[f_1,f_2](x)\|_{S_t} \lesssim \|(\sum_{n\in\mathbb{Z}} |H_{2^n s}[f_1,f_2](x)|^2)^{1/2}\|_{L_s^1[1,2]}$$

We may write $H_{2^n s}[f_1,f_2](x) = \int f_1(x+y)f_2(x-y)2^{-n}K_s(2^{-n}y)dy$ with $K_s(y) := s^{-1}h(s^{-1}y)$, and it is not hard to see that K_s satisfies (3), (4), (12) uniformly in $s \in [1,2]$. Thus, the desired estimates for the short jump component of $B_t[f_1,f_2]$ follows from Theorem 3.2.

In the rest of the paper we prove Theorem 3.1 and Theorem 3.2. We will use the restricted weak-type interpolation approach of [15], which will be discussed in the next section.

4. Linearization and interpolation

4.1. **Linearization.** For each x consider a measurable function $L : \mathbb{R} \to \mathbb{Z}_+$ the set of positive integers, and two sequences of measurable functions: a non-decreasing integer valued sequence $(k_n(x))_{n=0}^{L(x)}$ and a sequence $(a_n(x))_{n=1}^{L(x)}$ such that $\sum_{n\geqslant 0}|a_n(x)|^{r'} \leqslant 1$. Then an appropriate choice of L and such sequences guarantees that

$$\|B_{2^k}(f_1,f_2)(x)\|_{V_k^r(\mathbb{Z})} \leqslant 2\sum_{n=1}^{L(x)}\Big(B_{2^{k_n}}[f_1,f_2](x) - B_{2^{k_n-1}}[f_1,f_2](x)\Big)a_n(x) .$$

Similarly, for each $s \in [1,2]$ we may find a sequence of measurable functions $(d_n(s,x))_{n=-\infty}^{\infty}$ such that $\sum_n |d_n(s,x)|^2 \leqslant 1$, and

$$(\sum_{n\in\mathbb{Z}} |\int f_1(x+y)f_2(x-y)2^n K_s(2^n y)dy|^2)^{1/2}$$

$$\leqslant 2\sum_{n\in\mathbb{Z}} \int f_1(x+y)f_2(x-y)2^n K_s(2^n y)dy\, d_n(s,x)$$

The desired estimates in Theorem 3.1 and Theorem 3.2 follow from certain restricted-weak type estimates for the following tri-linear forms, which we will discuss in the next section.

$$\Lambda_{long}(f_1, f_2, f_3) = \left\langle \sum_{n=1}^{L} \left(B_{2^{k_n}}[f_1, f_2] - B_{2^{k_{n-1}}}[f_1, f_2] \right) a_n, f_3 \right\rangle$$

$$\Lambda_{short}(f_1, f_2, f_3) = \int_1^2 \Lambda_{short,s}(f_1, f_2, f_3) ds$$

$$\Lambda_{short,s}(f_1, f_2, f_3) = \left\langle \sum_{n \in \mathbb{Z}} \int f_1(\cdot + y) f_2(\cdot - y) 2^{-n} K_s(2^{-n}y) dy d_n(s, \cdot), f_3 \right\rangle .$$

4.2. **Restricted weak-type interpolation.** For any $G \subset \mathbb{R}$ with finite Lebesgue measure, we say that $H \subset G$ is a minor subset if $|H| \leqslant |G|/2$.

Let $\alpha = (\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3$ be such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and at most one α_j could be negative. We say that a tri-linear functional $\Lambda(f_1, f_2, f_3)$ satisfies restricted weak-type estimates with exponents α if the following holds.

Case 1: $\min(\alpha_1, \alpha_2, \alpha_3) \geqslant 0$. Then we require existence of $j_0 \in \{1, 2, 3\}$ with the following property: for every triple (F_1, F_2, F_3) of finite Lebesgue measurable subsets of $\mathbb{R}$ we could find $B \subset F_{j_0}$ minor subset such that

$$(13) \qquad\qquad \Lambda(f_1, f_2, f_3) \lesssim |F_1|^{\alpha_1} |F_2|^{\alpha_2} |F_3|^{\alpha_3}$$

for any f_1, f_2, f_3 with the following property: $|f_j| \leqslant 1_{F_j}$ if $j \neq j_0$, $|f_{j_0}| \leqslant 1_{F_{j_0} - B}$.

Case 2: $\min(\alpha_1, \alpha_2, \alpha_3) < 0$. Let k be such that $\alpha_k < 0$. By assumptions on α the other α_j's are nonnegative. Then we require the above property with $j_0 = k$.

Let A be the hexagon on the plane $L = \{\alpha_1 + \alpha_2 + \alpha_3 = 1\}$ with vertices

$$(14) \qquad A_1(-\frac{1}{2}, \frac{1}{2}, 1), \quad A_2(\frac{1}{2}, -\frac{1}{2}, 1), \quad A_3(\frac{1}{2}, 1, -\frac{1}{2}),$$
$$A_4(-\frac{1}{2}, 1, \frac{1}{2}), \quad A_5(1, -\frac{1}{2}, \frac{1}{2}), \quad A_6(1, \frac{1}{2}, -\frac{1}{2})$$

By the interpolation argument of [15], to show Theorem 3.1 and Theorem 3.2 it suffices to prove that in any given neighborhood (in the plane L) of any vertex of A we could find α such that $\Lambda_{long}(f_1, f_2, f_3)$ and $\Lambda_{short}(f_1, f_2, f_3)$ satisfy restricted weak-type estimates with exponents α. (Note that when α is near a vertex of A it is automatic that at most one coordinate of α could be negative.)

It will be clear from our proof (of the restricted weak-type estimates for all involved trilinear forms) that the index j_0 and the exceptional set B depend only on α and F_1, F_2, F_3. Also, in the proof the choice of α (inside any small neighborhoods of any given vertices of H) will not depend on the underlying trilinear form.

Therefore, *a posteriori*, to show the restricted weak-type estimates for Λ_{short} it suffices to obtain the same estimate for $\Lambda_{short,s}$ (with the same set of exponents), provided that the implicit constants are uniform over $s \in [1, 2]$. This uniformity in turn is a consequent of the fact that the implicit constants in the assumptions for K_s are uniform over $s \in [1, 2]$.

Similarly, in the proof for Λ_{long} we'll decompose it into a weighted sum of simpler trilinear forms, and it suffices to obtain the restricted weak-type estimates for

each of the new forms (with the same set of exponents) provided that the implicit constants are uniform.

5. TERMINOLOGY OF TILES AND TREES

In this section we recall some terminologies from [14, 7, 3] that will be used in the proof. By a *grid* we mean a collection of intervals whose lengths are integral powers of 2 such that if I, I' are two intersecting elements then $I \subset I'$ or $I' \subset I$. In addition to the standard grid $\mathcal{G}_0$ of dyadic intervals $2^i[m, m+1)$, we will use the grids

$$\mathcal{G}_{\ell,t} = \left\{ \left[2^i(m + \frac{\ell}{5}), 2^i(m + \frac{\ell}{5} + 1) \right) : i = t \ (mod \ 4), m \in \mathbb{Z} \right\}$$

where ℓ and t are integers, clearly $G_{\ell,t}$ depends only on $\ell \ (mod \ 5)$ and $t \ (mod \ 4)$. We will also make use of the grids

$$\mathcal{G}_1 = \left\{ \left[2^i(m + \frac{(-1)^i}{3}), 2^i(m + \frac{(-1)^i}{3} + 1) \right) : m \in \mathbb{Z} \right\}$$

$$\mathcal{G}_2 = \left\{ \left[2^i(m - \frac{(-1)^i}{3}), 2^i(m - \frac{(-1)^i}{3} + 1) \right) : m \in \mathbb{Z} \right\}.$$

It is clear that for every (not necessarily dyadic) interval I there is a $d \in \{0, 1, 2\}$ and a $J \in \mathcal{G}_d$ such that $I \subset J$ and $J \subset 3I$; we then say that I is *d-regular*.

A *tile* p is a rectangle $I_p \times \omega_p \subset \mathbb{R}^2$ of area 1 such that I_p is dyadic. A *tri-tile* P will consist of a quadruplet of intervals $(I_P, \omega_{P_1}, \omega_{P_2}, \omega_{P_3})$ where I_P is dyadic and $|I_P||\omega_{P_i}| = 1$ for each i. Associated to the tri-tile P are the three tiles $P_i = I_P \times \omega_{P_i}$ which justify the notation that is implicit in the previous sentence.

For each quadruplet of integers $\nu = (j_1, j_2, e, i)$ such that $0 \leqslant j_1, j_2 \leqslant 4$ and $498 \leqslant |e| \leqslant 4002$ and $0 \leqslant i < 4000$, consider the collection of tri-tiles

$$\begin{aligned} \mathbf{P}_\nu \quad = \quad \{ \big(\quad & [2^{-i'}m, 2^{-i'}(m+1)), \\ & [2^{i'}(n + \frac{j_1}{5}), 2^{i'}(n + \frac{j_1}{5} + 1)), \\ & [2^{i'}(n + e + \frac{j_2}{5}), 2^{i'}(n + e + \frac{j_2}{5} + 1)), \\ & [2^{i'}(2n + e + \frac{j_1 + j_2}{5} + 1), 2^{i'}(2n + e + \frac{j_1 + j_2}{5} + 2)) \big) \\ & : m, n, i' \in \mathbb{Z}, i' = i \ (mod \ 4000) \} \end{aligned}$$

(15)

Above, we clearly have $\omega_{P_1} \in \mathcal{G}_{j_1,i}$, $\omega_{P_2} \in \mathcal{G}_{j_2,i}$, and $\omega_{P_3} \in \mathcal{G}_{j_1+j_2,i}$.

Fixing ν for the remainder of the section (some definitions below depend on ν), we now recall, from [3] (cf. [15]), some notions of order for tiles.

Definition 5.1. *For two tiles p, p' we write*

- *$p' < p$ if $I_{p'} \subsetneq I_p$ and $3\omega_p \subsetneq 3\omega_{p'}$*
- *$p' \leqslant p$ if $p' < p$ or $p' = p$*
- *$p' \lesssim p$ if $I_{p'} \subset I_p$ and $\omega_p \subset 10|e|\omega_{p'}$*
- *$p' \lesssim' p$ if $p' \lesssim p$ and $10\omega_{p'} \cap 10\omega_p = \varnothing$*

It is not hard to see that if $P, P' \in \mathbf{P}_\nu$ are two tri-tiles with $P'_i < P_i$ for some $i \in \{1, 2, 3\}$ then $P'_j \lesssim' P_j$ for each $j \in \{1, 2, 3\}\backslash\{i\}$.

The ordering above gives rise to the concept of a tree, which we recall below:

Definition 5.2. *Let $i \in \{1, 2, 3\}$. An i-overlapping tree is a collection of tri-tiles $T \subset \mathbf{P}_\nu$ together with a top tri-tile $P_T \in \mathbf{P}_\nu$ which satisfies*

$$P_i < (P_T)_i \text{ for all } P \in T\backslash\{P_T\}.$$

We say that T is a tree if it is an i-overlapping tree for some $i \in \{1, 2, 3\}$. We say that T is a tree with top if $P_T \in T$.

A tree T is called j-lacunary if

$$P_j \lesssim' (P_T)_j \text{ for all } P \in T\backslash\{P_T\}.$$

It follows that a tree is j-lacunary if and only if it is i-overlapping for some $i \in \{1, 2, 3\}\backslash\{j\}$, furthermore for each $P \in T$ we have $\operatorname{sgn}(c(\omega_{P_j}) - c(\omega_{(P_T)_j})) = \epsilon_{i,j}$ where we define $\epsilon_{i,j} = \operatorname{sgn}(e)$ for $(i, j) \in \{(1, 2), (1, 3), (3, 2)\}$ and $\epsilon_{i,j} = -\operatorname{sgn}(e)$ for $(i, j) \in \{(2, 1), (3, 1), (2, 3)\}$. We will abbreviate $I_T := I_{P_T}$.

Definition 5.3. *We will say that a collection of trees $\mathbf{T}$ is* strongly j-disjoint *for some $j \in \{1, 2, 3\}$ if*

(1) *Each $T \in \mathbf{T}$ is j-lacunary*
(2) *If $T, T' \in \mathbf{T}$ and $T \neq T'$ then $T \cap T' = \varnothing$*
(3) *If $T, T' \in \mathbf{T}$, $T \neq T'$, $P \in T$, $P' \in T'$, and $\omega_{P_j} \subsetneq \omega_{P'_j}$ then $I_{P'} \cap I_T = \varnothing$*
(4) *If $T, T' \in \mathbf{T}$, $T \neq T'$, and $P' \in T'$ then $P'_j \nleq (P_T)_j$*

Note that, due to our choice of order on tiles, the condition (4) above is somewhat nonstandard (in comparison with, say, [3]). Also note that conditions (2) and (3) imply that if $T, T' \in \mathbf{T}$, $T \neq T'$, $P \in T$, and $P' \in T'$ then $P_j \cap P'_j = \varnothing$.

6. Discretization

In this section, we discuss discretization, i.e. wavelet representation, for Λ_{long} and $\Lambda_{short,s}$. We'll largely follow [3]. We'll discuss in details the process for Λ_{long}, the discretization for $\Lambda_{short,s}$ will be similar and discussed at the end of the section.

6.1. **Cancellation between dilates.** The point of conditions (3), (4) is that they allow one to decompose K into much simpler kernels:

Lemma 6.1. *If K satisfies (3) and (4) then*

$$(16) \qquad \widehat{K}(\xi) \;=\; \sum_{j \in \mathbb{Z}}^{\infty} c_j \widehat{K}_j(2^j \xi) \;, \quad \forall\, \xi \neq 0 \;,$$

where $\{c_j\}_j \in \ell^1(\mathbb{Z})$ and each $\widehat{K}_j$ could be furthermore written as the sum of dilates of a single generating function $\widehat{K}_j(\cdot) = \sum_{\ell \geqslant 0} \widehat{\phi}_j(2^\ell \cdot)$, where $supp(\widehat{\phi}_j) \subset \{500 \leqslant |\xi| \leqslant 4000\}$, and it holds uniformly in j that

$$(17) \qquad |\phi_j^{(n)}(x)| \;\leqslant\; C_{m,n}(1 + |x|)^{-m}$$

for every $m, n \geqslant 0$. If K satisfies (3) and (4) up to some high order then (17) holds for $m, n \leqslant M$ with M comparably large.

The utility of this approach lies in the following cancellation between dilates of K_j: for every integers $k_1 \leqslant k_2$ we have $\widehat{K}_j(2^{k_2}\xi) - \widehat{K}_j(2^{k_1}\xi) = \sum_{k_1 \leqslant \ell < k_2} \widehat{\phi}_j(2^\ell \xi)$, which turns out to be convenient for reducing Λ_{long} to wavelet operators. Namely, by pulling out the sum in j, we thus see that the consideration of Λ_{long} reduces to considerations of Λ_j, defined by:

$$\Lambda_j(f_1, f_2, f_3) \;\; := \;\; \left\langle \sum_{n=1}^{L} \Big(\sum_{k_{n-1} \leqslant \ell < k_n} B_{\phi_j, \ell}[f_1, f_2] \Big) a_n, f_3 \right\rangle \;\; ,$$

$$B_{\phi_j, \ell}[f_1, f_2](x) \;\; := \;\; \int f_1(x+y) f_2(x-y) 2^{-\ell} \phi_j(2^{-\ell} y) \; dy \;\; ,$$

and $B_{\phi_j, \ell}$ could be decomposed into a finite number of discrete wavelet operators at scale ℓ; this decomposition will be discussed in Section 6.2.

Proof of Lemma 6.1. Observing that the given assumptions on $\widehat{K}$ implies the existence of $\widehat{K}(0+)$ and $\widehat{K}(0-)$. We consider two cases. We'll only consider the setting when (3) and (4) hold for all orders; the finite order case could be achieved by the same argument.

  <u>Case 1:</u> Suppose that $\widehat{K}(0+) = \widehat{K}(0-) = 0$, then using the given assumptions on K it follows that for every $n \geqslant 0$ it holds that

$$\Big| \frac{d^n}{d\xi^n} \widehat{K}(\xi) \Big| \;\; \lesssim \;\; |\xi|^{-n} \min(|\xi|, 1/|\xi|)$$

(the improvement is at $n = 0$). Let η be a nonnegative C^∞ bump function on $\{1000 \leqslant |\xi| \leqslant 2000\}$ such that $\sum_j \eta(2^j \xi) = 1$ for every $\xi \neq 0$. Let

$$\widehat{K}_j(\xi) \;\; = \;\; 2^{|j|} \widehat{K}(2^{-j}\xi) \eta(\xi) \;\; ,$$

which is supported in $\{1000 \leqslant |\xi| \leqslant 2000\}$, it is routine to check that (16) holds with $c_j = 2^{-|j|}$, and $\frac{d^n}{d\xi^n} \widehat{K}_j(\xi) \lesssim_n 1$ for all $n \geqslant 0$. Let $\widehat{\phi}_j(\xi) = \widehat{K}_j(\xi) - \widehat{K}_j(2\xi)$, then $\widehat{\phi}_j$ has the desired properties.

  <u>Case 2:</u> $(\widehat{K}(0+), \widehat{K}(0-)) \neq (0,0)$. Let φ be such that $\widehat{\varphi}$ is supported on $[-2000, 2000]$ and is in $C^\infty(\mathbb{R} - \{0\})$, and $\widehat{\varphi}(\xi) = \widehat{K}(0+)$ for $\xi \in [0, 1000]$ and $\varphi(\xi) = \widehat{K}(0-)$ for $\xi \in [-1000, 0)$. Then by writing $\widehat{K} = \widehat{\phi} + (\widehat{K} - \widehat{\phi})$ and applying the analysis in Case 1 for $K - \phi$, we are left with $\widehat{\varphi}$, for which we will decompose directly into the sum of dilates of a single generating function. Namely, let $\widehat{\phi}(\xi) = \widehat{\varphi}(\xi) - \widehat{\varphi}(2\xi)$, it is clear that $\widehat{\phi}$ satisfies the desired properties. $\square$

6.2. Wave packet representation.

Below we will decompose $B_{\phi_j, \ell}[f_1, f_2]$ into wavelet sums. For convenience of notation, we will suppress the variable j, namely below $\phi = \phi_j$ whose Fourier transform is supported on $\{500 \leqslant |\xi| \leqslant 2000\}$ and ϕ satisfies (17) up to sufficiently high order.

Definition 6.2. *We say that ψ is an L^2 wave packet of order M adapted to a tile $p = I \times \omega$ if $\widehat{\psi}$ is supported in ω and the following estimate holds for all $0 \leqslant m, n \leqslant M$:*

$$(18) \qquad \frac{d^n}{dx^n} \Big(e^{-2\pi i c(\omega) x} \psi(x) \Big) \;\; \leqslant \;\; C_{M,N,n,m} \frac{1}{|I|^{n+\frac{1}{2}}} \widetilde{\chi}_I(x)^m$$

Lemma 6.3. *Given any $M, N > 0$, if ϕ satisfies (17) up to sufficiently high order and $\widehat{\phi}$ is supported in $\{500 \leqslant |\xi| \leqslant 4000\}$ then for each $\ell \in \mathbb{Z}$ $B_{\phi,\ell}[f_1, f_2](x)$ can be written as the sum over $\nu = (j_1, j_2, e)$, $0 \leqslant j_1, j_2 \leqslant 4$ and $498 \leqslant |e| \leqslant 4002$ integers, of*

$$(19) \qquad \sum_{j \in \mathbb{Z}} 2^{-N|j|} \sum_{P \in \mathbf{P}_\nu : |\omega| = 2^{-\ell}} |I_P|^{-1/2} \langle f_1, \psi_{j,P,1} \rangle \langle f_2, \psi_{j,P,2} \rangle \psi_{j,P,3}(x) \ ,$$

where (uniform over tri-tiles $P \in \mathbf{P}_\nu$, $i = 1, 2, 3$, and $j \geqslant 0$) $\psi_{j,P,i}$ is an L^2 wave packet adapted to P_i of order M (the constants in (18) may depend on M, N, ν).

Proof. We consider $M = \infty$ below, the finite case is similar. Recall that the Fourier transform $\mathcal{F}$ is defined by (5). We first make several remarks about $B_{\phi,\ell}$. Suppose that the supports of $\widehat{\psi}_1$ and $\widehat{\psi}_2$ are contained in intervals ω_1, ω_2 respectively. Then the identity

$$\mathcal{F}\left[B_{\phi,\ell}[\psi_1, \psi_2]\right](\xi) \ = \ \int_{\mathbb{R}} \widehat{\psi}_1(\xi - \eta) \widehat{\psi}_2(\eta) \widehat{\phi}(2^l(2\eta - \xi)) \, d\eta$$

gives rise to two observations. First, if $B_{\phi,\ell}(\psi_1, \psi_2)$ does not vanish then

$$(20) \qquad (\omega_1 - \omega_2) \cap \{500 \cdot 2^{-\ell} \ \leqslant \ |\xi| \leqslant 4000 \cdot 2^{-\ell}\} \neq \varnothing.$$

Second, the support of Fourier transform of $B_{\phi,\ell}(\psi_1, \psi_2)$ is contained in

$$(21) \qquad \qquad \qquad \omega_1 + \omega_2.$$

Now, turning to spatial localization, we have

$$(22) \qquad \mathbf{e}^{-2\pi \mathbf{i} 2 \xi x} B_{\phi,\ell}[\psi_1, \psi_2](x) \ = \ B_{\phi,\ell}\left[\mathbf{e}^{-2\pi \mathbf{i} \xi(\cdot)} \psi_1, \mathbf{e}^{-2\pi \mathbf{i} \xi(\cdot)} \psi_2\right](x) \ .$$

Since ϕ satisfies (17), it follows that if for some ξ, x_1, x_2 we have

$$\frac{d^n}{dx^n}\left(\mathbf{e}^{-2\pi \mathbf{i} \xi x} \psi_i(x)\right) \ \leqslant \ C_{n,m} 2^{-\ell(n+\frac{1}{2})}\left(1 + \frac{|x - x_i|}{2^\ell}\right)^{-m}$$

for each $i = 1, 2$, and for every $n, m \geqslant 0$, then

$$(23) \qquad \frac{d^n}{dx^n}\left(\mathbf{e}^{-2\pi \mathbf{i} 2 \xi x} B_{\phi,\ell}[\psi_1, \psi_2](x)\right) \ \leqslant$$

$$\leqslant \ C_{n,m} 2^{-\ell(n+1)}\left(1 + \frac{|x_1 - x_2|}{2^\ell}\right)^{-m}\left(1 + \frac{|x - (x_1 + x_2)/2|}{2^\ell}\right)^{-m}$$

for each $n, m \geqslant 0$. (Here we emphasize that $C_{m,n}$'s are independent of ξ.)

Fix a Schwartz function $\widehat{\psi}$ supported on $[0, 2/5)$ such that $\sum_{j \in \mathbb{Z}} |\widehat{\psi}(\cdot - \frac{j}{5})|^2 = 1$. For each pair of intervals (I, ω) with $|\omega||I| = 1$ let

$$\widehat{\psi}_{I,\omega} = \mathbf{e}^{-2\pi \mathbf{i} c(I) \xi} |\omega|^{-1/2} \widehat{\psi}(|\omega|^{-1}(\xi - c(\omega)))$$

which is supported inside the right half of ω. Using a Fourier sampling theorem, for any Schwartz function f it holds that

$$f = \sum_{j=0}^{4} \sum_{\substack{\omega \in \mathcal{G}_0 \\ |\omega| = 2^{-\ell}}} \sum_{\substack{I \in \mathcal{G}_0 \\ |I| = 2^\ell}} \left\langle f, \psi_{I, \omega + \frac{j}{5}|\omega|} \right\rangle \psi_{I, \omega + \frac{j}{5}|\omega|} \qquad .$$

Let $\omega_1 = \omega + \frac{j_1}{5}|\omega|$ and $\omega_2 = \omega + (\frac{j_2}{5} + e)|\omega|$. By (20), it follows that $B_{\phi,\ell}[f_1, f_2](x)$ can be written as the sum over triplets of integers (j_1, j_2, e), with $0 \leqslant j_1, j_2 \leqslant 4$ and $498 \leqslant |e| \leqslant 4002$, of

$$(24) \qquad \sum_{\substack{\omega \in \mathcal{G}_0 \\ |\omega|=2^{-\ell}}} \sum_{\substack{I_1 \in \mathcal{G}_0 \\ |I_1|=2^\ell}} \sum_{\substack{I_2 \in \mathcal{G}_0 \\ |I_2|=2^\ell}} \langle f_1, \psi_{I_1,\omega_1} \rangle \langle f_2, \psi_{I_2,\omega_2} \rangle \, \varphi_{I_1,I_2,\omega,j_1,j_2,e}(x) \quad,$$

$$\varphi_{I_1,I_2,\omega,j_1,j_2,e}(x) \quad := \quad B_{\phi,\ell}[\psi_{I_1,\omega+\frac{j_1}{5}|\omega|}, \psi_{I_2,\omega+(\frac{j_2}{5}+e)|\omega|}](x) \quad.$$

By (21), $\widehat{\varphi}_{I_1,I_2,\omega,j_1,j_2,e}$ is supported on $\omega_3 := [0,|\omega|) + 2c(\omega) + (\frac{j_1+j_2}{5} + e)|\omega|$ and, by (23), satisfies

$$(25) \qquad \frac{d^n}{dx^n}\Big(\mathbf{e}^{-2\pi\mathbf{i}(2c(\omega)+(\frac{j_1+j_2}{5}+e)|\omega|)x}\varphi_{I_1,I_2,\omega,j_1,j_2,e}(x)\Big) \leqslant$$

$$\leqslant \quad C_{n,m}2^{-\ell(n+1)}(1 + \frac{|c(I_1) - c(I_2)|}{2^\ell})^{-m}(1 + \frac{|x - (c(I_1)+c(I_2))/2|}{2^\ell})^{-m}$$

$$\leqslant \quad C_{n,m}2^{-\ell(n+1)}(1 + \frac{|c(I_1) - c(I_2)|}{2^\ell})^{-m}(1 + \frac{|x - c(I_1)|}{2^\ell})^{-m}$$

for each $n, m \geqslant 0$. (Note that j_1, j_2, e are bounded.)

We now fix I_1 and further divide the right hand side of (24) according to $j := 2^{-\ell}(c(I_1) - c(I_2))$. For $j = 0$ i.e. for terms in the sum (24) where $I_1 = I_2 =: I$ we may define the tri-tile $P = (I, \omega_1, \omega_2, \omega_3)$ and the corresponding wave packets naturally

$$\psi_{0,P,1} \quad = \quad \psi_{I_1,\omega_1} \quad, \quad \psi_{0,P,2} \quad = \quad \psi_{I_2,\omega_2} \quad, \quad \psi_{0,P,3} \quad = \quad |I|^{1/2}\psi_{I_1,I_2,\omega,j_1,j_2,e} \quad.$$

The remaining terms can be dealt with by using the rapid decay in $|c(I_1)-c(I_2)|$, (25): we still define $\psi_{j,P,1} = \psi_{I_1,\omega_1}$, however to shift the localization of ψ_{I_2,ω_2} to I_1 we define

$$\psi_{j,P,2} \quad := \quad (1 + \frac{|c(I_1) - c(I_2)|}{2^\ell})^{-L} \cdot \psi_{I_2,\omega_2} \quad,$$

$$\psi_{j,P,3} \quad := \quad 2^{Nj}(1 + \frac{|c(I_1) - c(I_2)|}{2^\ell})^L \big(|I|^{1/2}\psi_{I_1,I_2,\omega,j_1,j_2,e}\big)$$

for some large L. (The rapid decay in (25) takes care of the extra factors in $\psi_{j,P,3}$.)

Finally, we split (24) up one more time so that whenever $|I| > |I'|$ we have $|I| = 2^{4000k}|I'|$ for some positive integer k. Note that while this splitting gives rise to the sparseness required by $\mathbf{P}_\nu$, it also means that we need to relabel and rescale the $\phi_{j,P,i}$ slightly to maintain the sequence of weights $2^{-N|j|}$. $\square$

It follows that to prove restricted weak-type estimates for Λ_{long}, we are left with showing the following theorem. In the theorem, $\nu = (j_1, j_2, e, i)$ is any quadruplet of integers such that $0 \leqslant j_1, j_2 \leqslant 4$, $498 \leqslant |e| \leqslant 4002$, $0 \leqslant i < 4000$, and $\mathbf{P}_\nu$ is defined by (15).

Theorem 6.4. *Let $r > 2$ and p_1, p_2, q satisfy (1).*

Suppose that (uniformly over tri-tiles $P \in \mathbf{P}_\nu$, $i = 1, 2, 3$), $\psi_{P,i}$ is an L^2 normalized wave packet adapted to P_i up to order M sufficiently large (the required M may depend on p_1, p_2, q).

Then the trilinear form

$$\left\langle \sum_{n=1}^{L} \sum_{P \in \mathbf{P}_\nu: \ 2^{k_n-1} \leqslant |I_P| < 2^{k_n}} |I_P|^{-1/2} \langle f_1, \psi_{P,1} \rangle \langle f_2, \psi_{P,2} \rangle \psi_{P,3} \ a_n, f_3 \right\rangle$$

satisfies restricted weak-type estimates with exponents α arbitrarily close to any given vertex of A defined by (14).

Recall that $\sum_{n=1}^{L(x)} |a_n(x)|^{r'} \leqslant 1$. For convenience, let $a_P(x) := a_m(x)$ if $m = m(P,x)$ is the unique integer in $\{1, \ldots, L(x)\}$ satisfying $2^{k_m-1}(x) \leqslant |I_P| < 2^{k_m(x)}$, and $a_P(x) = 0$ if such m does not exist. We also let $\phi_{P,3}(x) = a_p(x)\psi_{P,3}(x)$ and $\phi_{P,i} = \psi_{P,i}$ for $i = 1, 2$. Let $\mathbf{P}$ be a finite subset of $\mathbf{P}_\nu$. It suffices to demonstrate that the trilinear form

$$(26) \qquad \Lambda_{\mathbf{P}}(f_1, f_2, f_3) = \left\langle \sum_{P \in \mathbf{P}} |I_P|^{-1/2} \langle f_1, \phi_{P,1} \rangle \langle f_2, \phi_{P,2} \rangle \phi_{P,3}, f_3 \right\rangle$$

is of restricted weak type with exponents α, with $\mathbf{P}$-uniform implicit constants.

6.3. **Discretization for $\Lambda_{short,s}$.** Recall that

$$\Lambda_{short,s}(f_1, f_2, f_3) \ = \ \left\langle \sum_{n \in \mathbb{Z}} \int f_1(\cdot + y) f_2(\cdot - y) 2^{-n} K_s(2^{-n}y) dy d_n(s, \cdot), f_3 \right\rangle$$

Using $\widehat{K}_s(\xi) \lesssim |\xi|$ we could proceed as in Case 1 of the proof of Lemma 6.1 and obtain a decomposition

$$\widehat{K}_s(\xi) \ = \ \sum_{j \in \mathbb{Z}} 2^{-|j|} \widehat{K}_{s,j}(2^j \xi)$$

where $\widehat{K}_{s,j}$ is supported in $\{1000 \leqslant |\xi| \leqslant 2000\}$ and $\frac{d^n}{\xi^n} \widehat{K}_{s,j}(\xi) \lesssim_n 1$ for all $n \geqslant 0$. Thus it suffices to consider restricted weak-type estimates for

$$\left\langle \sum_{n \in \mathbb{Z}} \int f_1(\cdot + y) f_2(\cdot - y) 2^{-(n+j)} K_{s,j}(2^{-(n+j)}y) dy \ d_n(s, \cdot), f_3 \right\rangle$$

$$= \ \left\langle \sum_{n \in \mathbb{Z}} B_{K_{s,j},n}[f_1, f_2](\cdot) d_n(s, \cdot), f_3 \right\rangle \quad .$$

Now using Lemma 6.3 with $K_{s,j}$ playing the role of ϕ, it follows that to obtain the desired restricted weak-type estimates for Λ_{short} we are left with showing the following theorem. Below, $\nu = (j_1, j_2, e, i)$ is any quadruplet of integers such that $0 \leqslant j_1, j_2 \leqslant 4$, $498 \leqslant |e| \leqslant 4002$, $0 \leqslant i < 4000$, and $\mathbf{P}_\nu$ is defined by (15).

Theorem 6.5. *Let $r > 2$ and p_1, p_2, q satisfy (1).*

Suppose that (uniformly over tri-tiles $P \in \mathbf{P}_\nu$, $i = 1, 2, 3$), $\psi_{P,i}$ is an L^2 wave packet adapted to P_i up to order M sufficiently large.

Let $(d_n)_{n \in \mathbb{Z}}$ be a sequence of measurable functions such that $\sum_n |d_n(x)|^2 \leqslant 1$.

Then the trilinear form

$$\left\langle \sum_{n \in \mathbb{Z}} \sum_{P \in \mathbf{P}_\nu: \ |I_P| = 2^n} |I_P|^{-1/2} \langle f_1, \psi_{P,1} \rangle \langle f_2, \psi_{P,2} \rangle \psi_{P,3} \ d_n, f_3 \right\rangle$$

satisfies restricted weak-type estimates with exponents α arbitrarily close to any vertex of A defined by (14).

7. AUXILIARY ESTIMATES

The following bound follows from the Lépingle inequality and a square function argument, see [10] for details.

Lemma 7.1. *Let ζ be any Schwartz function. Let $\zeta_k(\cdot) = 2^{-k}\zeta(2^{-k}\cdot)$. Then for $r > 2$ and $1 < s < \infty$ it holds that*

$$\|(\zeta_k * f)(x)\|_{L_x^s(V_k^r(\mathbb{Z}))} \lesssim_{r,s} \|f\|_{L^s}$$

Next, we have a Rademacher-Menshov type lemma:

Lemma 7.2. *Let $f_1, \ldots, f_N$ be functions on a measure space X such that for every sequence of signs $\epsilon_1, \ldots, \epsilon_N \in \{1, -1\}$ it holds that*

$$(27) \qquad \|\epsilon_1 f_1 + \epsilon_2 f_2 + \cdots + \epsilon_N f_N\|_{L^2} \leqslant B.$$

Then

$$\|\sum_{j=1}^n f_j(x)\|_{L_x^2(V_n^2)} \lesssim (1 + \log(N))B.$$

Proof. We rewrite $\sum_{j=1}^n f_j(x) = f_n(x) + \sum_{1 \leqslant j < n} f_j(x)$. Estimating the V^2 norm by the ℓ^2 norm, it is clear that

$$\|f_n(x)\|_{L_x^2(V_n^2)} \lesssim \left(\sum_{j=1}^N \|f_j\|_{L^2}^2\right)^{1/2} \lesssim \mathbf{E}_{\epsilon_1,\ldots,\epsilon_n}\|\sum_j \epsilon_j f_j\|_{L^2} \lesssim B .$$

It remains to consider the contribution from $\sum_{1 \leqslant j < n} f_j$. For each $n \in \{1, \ldots, N\}$ we will decompose $[1, n)$ into disjoint subintervals,

$$[0, n) = \bigcup_{m \leqslant \log_2(N)} \omega_{n,m} ,$$

as follows: Let I be the dyadic interval of length 2^{m+1} that contains n. If n is on the left half of I then let $\omega_{n,m} = \varnothing$. If n is on the right half of I then let $\omega_{n,m}$ be the left half of I. It follows that

$$\|\sum_{j < n} f_j(x)\|_{L_x^2(V_n^2)} \leqslant \sum_{0 \leqslant m \leqslant \log_2(N)} \|\sum_{j \in \omega_{n,m}} f_j(x)\|_{L_x^2(V_n^2)}.$$

Since for each m, $\omega_{n,m}$ is constant (in n) on dyadic intervals of length 2^m, we have

$$\|\sum_{j \in \omega_{n,m}} f_j(x)\|_{L_x^2(V_n^2)} \lesssim \|(\sum_{\omega \ dyadic: \ |\omega|=2^m} |\sum_{j \in \omega} f_j(x)|^2)^{1/2}\|_{L_x^2}$$
$$\leqslant B.$$

here the final inequality follows from another appeal to (27) using sequences (ϵ_j) that are constant (as functions of j) on dyadic intervals of length 2^m. $\square$

We'll also use a Bessel inequality, Lemma 7.3. For a proof see e.g. [7, Proposition 13.1]. Below recall that $\psi_{P,j}$ are (unmodified) L^2-normalized Fourier wave packet up to sufficiently high orders:

Lemma 7.3. *Let $j \in \{1, 2, 3\}$. Let $\mathbf{T}$ be a collection of strongly j-disjoint trees, let $\mathbf{Q} = \bigcup_{T \in \mathbf{T}} T$, and suppose that $\|\sum_{T \in \mathbf{T}} 1_{I_T}\|_{L^\infty} \leqslant L$. Then for any sequence of coefficients $\{b_P\}_{P \in \mathbf{Q}}$*

$$\|\sum_{P \in \mathbf{Q}} b_P \psi_{P,j}\|_{L^2} \;\lesssim\; \log(1 + L)\|b_P\|_{\ell^2(\mathbf{Q})}.$$

8. A VARIATION-NORM MULTIPLIER ESTIMATE

In this section we consider a variation-norm version of Bourgain [2, Lemma 4.11], namely Theorem 8.1 below. In the following, let $\xi_1 < \ldots < \xi_N$ be real numbers. For each integer k we denote the sharp multi-frequency projection at scale k onto $\xi_1, \ldots, \xi_N$ by $\Pi_k[f] = \mathcal{F}^{-1}[1_{R_k}\widehat{f}]$, where $R_k = \bigcup_{\xi \in \Omega}(\xi - 2^{-k}, \xi + 2^{-k})$.

Theorem 8.1. *For every $r > 2$ and $\epsilon > 0$ it holds that*

$$\|\Pi_k[f](x)\|_{L_x^2(V_k^r)} \;\leqslant\; C_{r,\epsilon} N^\epsilon \|f\|_{L^2} \quad .$$

A variant of Theorem 8.1 with smooth multi-frequency projections was considered in [17], where a range of L^p estimates was obtained; for the current paper we need sharp frequency projections, but L^2 is sufficient.

The starting point of our proof is Lemma 3.2 from [17]:

Lemma 8.2. *Suppose that $\{c_k\}_{k=0}^\infty$ is a sequence in $\mathbb{R}^N$, and $2 < q < r$. Then*

$$\|\sum_{j=1}^N c_{k,j} \mathbf{e}^{2\pi \mathbf{i} \xi_j y}\|_{L_{y \in [0,1]}^2(V_{k \geqslant 0}^r)} \;\leqslant\; C N^{(\frac{1}{2} - \frac{1}{q})\frac{r}{r-2}} \|c_k\|_{V_{k \geqslant 0}^q(\ell^2(\mathbb{R}^N))}$$

where C may depend on r, q and $\min_j |\xi_j - \xi_{j-1}|$.

Through a standard averaging argument (see e.g. the proof of Proposition 4.1 in [17]), the lemma above gives

Proposition 8.1. *Let χ be a smooth function such that $\widehat{\chi}$ is identically one on $[-0.9, 0.9]$ and supported on $[-1, 1]$. Assume that $\xi_{j+1} \geqslant \xi_j + 1$ for each j, and let $\chi_{k,j}$ be defined by $\widehat{\chi}_{k,j}(\xi) = \chi(2^k(\xi - \xi_j))$. Then for $r > 2$ and $\epsilon > 0$ it holds that*

$$\|\sum_{1 \leqslant j \leqslant N} (\chi_{k,j} * f)(x)\|_{L_x^2(V_{k \geqslant 0}^r)} \;\leqslant\; C_{r,\epsilon,\chi} N^\epsilon \|f\|_{L^2}$$

Proof. To keep the paper self-contained, we sketch the averaging argument. Let $f_j(x) = \mathcal{F}^{-1}[1_{|\xi - \xi_j| \leqslant 1}\widehat{f}(\xi)](x)$. Since ξ_j's are separated, we have $\|f\|_2 \approx \|(f_j)\|_{L^2(\ell_j^2)}$. Let M be the best constant such that if $supp(\widehat{g}_j) \subset [\xi_j - 1, \xi_j + 1]$ for all j then

$$\|\sum_{1 \leqslant j \leqslant N} (\chi_{k,j} * g_j)(x)\|_{L_x^2(V_{k \geqslant 0}^r)} \;\leqslant\; M\|(g_j)\|_{L_x^2(\ell_j^2)} \quad,$$

by the triangle inequality and Lemma 7.2 it is clear that $M = O_{\chi,r}(N) < \infty$. Our aim is to show that $M = O_{\epsilon,r,\chi}(N^\epsilon)$. Since $\widehat{g}_j$ is supported on $[\xi_j - 1, \xi_j + 1]$, for $|y|$ small we have

$$\|g_j(x) - \mathbf{e}^{2\pi \mathbf{i} \xi_j y} g_j(x - y)\|_{L_x^2(\ell_j^2)} \;\lesssim\; |y|\|g_j\|_{L_x^2(\ell_j^2)} \quad .$$

Averaging over $0 \leqslant y \leqslant \delta$ with $1 \lesssim \delta < 1$ sufficiently small, we obtain

$$\| \sum_{1 \leqslant j \leqslant N} (\chi_{k,j} * g_j)(x) \|_{L^2_x(V^r_{k \geqslant 0})}$$

$$\leqslant C_\delta \| \sum_{1 \leqslant j \leqslant N} \mathbf{e}^{2\pi \mathbf{i} \xi_j y} (\chi_{k,j} * g_j)(x - y) \|_{L^2_{y \in [0,1]}(L^2_x(V^r_{k \geqslant 0}))} + \frac{M}{2} \|g_j\|_{L^2_x(\ell^2_j)}$$

$$= C_\delta \| \sum_{1 \leqslant j \leqslant N} \mathbf{e}^{2\pi \mathbf{i} \xi_j y} (\chi_{k,j} * g_j)(x) \|_{L^2_x(L^2_{y \in [0,1]}(V^r_{k \geqslant 0}))} + \frac{M}{2} \|g_j\|_{L^2_x(\ell^2_j)}$$

using translation invariant and Fubini. Using Lemma 8.2 for each fixed x it follows that for $q > 2$ sufficiently close to 2 (depending on $\epsilon > 0$) we have

$$\| \sum_{1 \leqslant j \leqslant N} (\chi_{k,j} * g_j)(x) \|_{L^2_x(V^r_{k \geqslant 0})} \leqslant C_{\epsilon,q,r} N^\epsilon \|(\chi_{k,j} * g_j)(x)\|_{L^2_x(V^q_k(\ell^2_j))} + \frac{M}{2} \|g_j\|_{L^2_x(\ell^2_j)}$$

$$\leqslant C_{\epsilon,q,r} N^\epsilon \|(\chi_{k,j} * g_j)(x)\|_{\ell^2_j(L^2_x(V^q_k))} + \frac{M}{2} \|g_j\|_{L^2_x(\ell^2_j)}$$

$$\leqslant C_{\epsilon,q,r,\chi} N^\epsilon \|g_j\|_{\ell^2_j(L^2_x)} + \frac{M}{2} \|g_j\|_{L^2_x(\ell^2_j)} \quad ,$$

in the second estimate we used $q > 2$ and in the last estimate we used Lemma 7.1. Since this holds for arbitrary (g_j) satisfying $supp(\widehat{g_j}) \subset [\xi_j - 1, \xi_j + 1]$, by definition of M we obtain $M \leqslant C_{\chi,\epsilon,q,r} N^\epsilon + \frac{M}{2}$, therefore $M = O_{\epsilon,r,\chi}(N^\epsilon)$ as desired. $\square$

Now, using Proposition 8.1 and a simple square function argument, we obtain a frequency separated version of Theorem 8.1. Namely, if $\xi_{j+1} \geqslant \xi_j + 1$ for each j then for $r > 2$ and $\epsilon > 0$ it holds that

$$\|\Pi_k[f]\|_{L^2_x(V^r_{k \geqslant 0})} \leqslant C_{r,\epsilon} N^\epsilon \|f\|_{L^2} \quad .$$

To remove the frequency separation requirement $\xi_{j+1} \geqslant \xi_j + 1$, we will need the following estimate, which will be proved using Lemma 7.2.

Proposition 8.2. *Suppose that S is a finite set of integers. Then*

(28) $$\|\Pi_k[f]\|_{L^2_x(V^2_{k \in S})} \leqslant C(1 + \log(|S|)) \|f\|_{L^2}$$

Proof. Let $n = |S|$. Let $s_1 < \ldots < s_n$ be elements of S. For $j = 1, \ldots, n-1$ write

$$f_j = \mathcal{F}^{-1}[(1_{R_{s_j}} - 1_{R_{s_{j+1}}})\widehat{f}]$$

and let $f_n = \mathcal{F}^{-1}[1_{R_{s_n}} \widehat{f}]$. Then, the f_j are orthogonal in $L^2(\mathbb{R})$, and $\Pi_{s_k}[f] = \sum_{j \geqslant k} f_j$, so Lemma 7.2 gives (28). $\square$

Proof of Theorem 8.1. By monotone convergence, it suffices to prove

$$\|\Pi_k[f](x)\|_{L^2_x(V^r_{k \in [a,b]})} \leqslant C_{r,\epsilon} N^\epsilon \|f\|_{L^2}$$

for every finite interval $[a, b]$, provided that the constant is independent of $[a, b]$.

Now, we may choose $\{k_j\}_{j=0}^N$ with

$$a = k_0 \leqslant \ldots \leqslant k_N = b$$

so that the number of connected components of R_k is constant on each interval $[k_j, k_{j+1})$. Then, for each x

$$\|\Pi_k[f](x)\|_{V^r_{k \in [a,b]}} \quad \leqslant \quad C\|\Pi_{k_j}[f](x)\|_{V^r_{j \in [0,N]}} + C \left(\sum_{j=1}^{N} \|\Pi_k[f](x)\|^2_{V^r_{k \in [k_{j-1}, k_j)}} \right)^{1/2}.$$

The contribution to the L^2_x norm of the first term on the right above is acceptable by Proposition 8.2. Furthermore, for each $k, k' \in [k_{j-1}, k_j)$ we have

$$\Pi_k[f](x) - \Pi_{k'}[f](x) = \Pi_k[f_j](x) - \Pi_{k'}[f_j](x)$$

where $f_j = \mathcal{F}^{-1}[(1_{R_{k_{j-1}}} - 1_{R_{k_j}})\widehat{f}]$. Thus, using the orthogonality of the f_j it suffices to show that, for each $1 \leqslant j \leqslant N$,

$$\|\Pi_k[f](x)\|_{L^2_x(V^r_{k \in [k_{j-1}, k_j)})} \quad \leqslant \quad C_{r,\epsilon} N^\epsilon \|f\|_{L^2}.$$

Fix j and let M be the (constant) number of connected components of R_k for $k \in [k_{j-1}, k_j)$, clearly $M \leqslant N$. For $k \in [k_{j-1}, k_j)$ we can write R_k as the disjoint union of open intervals

$$R_k = \bigcup_{1 \leqslant \ell \leqslant M} I_{\ell,k}$$

where $I_{\ell,k'} \subset I_{\ell,k}$ for $k' > k$. Let $\Omega = \{\xi_1, \ldots, \xi_N\}$ and define

$$L^\sharp := \{\ell \in [1, M] : |I_{\ell,k_{j-1}} \cap \Omega| = 1\} \quad, \quad \Pi_k^\sharp[f] := \sum_{\ell \in L^\sharp} \mathcal{F}^{-1}[1_{I_{\ell,k}} \widehat{f}]$$

and let $L^\flat = [1, M] - L^\sharp$ and define $\Pi^\flat$ analogously. Clearly $\Pi_k = \Pi_k^\sharp + \Pi_k^\flat$. Rescaling by a factor of $2^{k_{j-1}}$, an application of the known frequency-separated case immediately gives

$$\|\Pi_k^\sharp[f](x)\|_{L^2_x(V^r_{k \in [k_{j-1}, k_j)})} \quad \leqslant \quad C_{r,\epsilon} |L^\sharp|^\epsilon \|f\|_{L^2}$$

and so it remains to consider $\Pi^\flat$. For each $\ell \in L^\flat$ let

$$\zeta_\ell = \min(I_{\ell,k_{j-1}} \cap \Omega) \quad, \quad \rho_\ell = \max(I_{\ell,k_{j-1}} \cap \Omega)$$

and $I'_\ell = (\zeta_\ell, \rho_\ell)$, so that $I_{\ell,k} = (\zeta_\ell - 2^{-k}, \zeta_\ell] \cup I'_\ell \cup [\rho_\ell, \rho_\ell + 2^{-k})$. Now, define

$$\Pi_k^{\flat,1}[f] = \sum_{\ell \in L^\flat} \mathcal{F}^{-1}[1_{(\zeta_\ell - 2^{-k}, \zeta_\ell + 2^{-k})} \widehat{f}]$$

$$\Pi_k^{\flat,2}[f] = \sum_{\ell \in L^\flat} \mathcal{F}^{-1}[1_{(\rho_\ell - 2^{-k}, \rho_\ell + 2^{-k})} \widehat{f}].$$

For $k \in [k_{j-1}, k_j)$ we obtain the decomposition

$$\Pi_k^\flat[f] = g + \Pi_k^{\flat,1}[h_1] + \Pi_k^{\flat,2}[h_2]$$

where $g = \mathcal{F}^{-1}[\sum_{\ell \in L^\flat} I'_\ell \widehat{f}]$ (which stays the same under $\Pi^\flat$) and

$$h_1 = \mathcal{F}^{-1}[\sum_{\ell \in L^\flat} 1_{(\zeta_\ell - 2^{-k_{j-1}}, \zeta_\ell]} \widehat{f}] \quad, \quad h_2 = \mathcal{F}^{-1}[\sum_{\ell \in L^\flat} 1_{[\rho_\ell, \rho_\ell + 2^{-k_{j-1}})} \widehat{f}]$$

Rescaling by a factor of 2^{k_j-1}, an application of the known frequency-separated case then gives

$$\|\Pi_k^{\flat,i}[h_i](x)\|_{L^2_x(V^r_{k\in[k_{j-1},k_j]})} \;\leqslant\; C_{r,\epsilon}|L^\flat|^\epsilon\|h_i\|_{L^2}$$

for $i = 1, 2$, finishing the proof. $\square$

9. Size and a variation-norm size bound

We will use the following standard notion of size:

Definition 9.1. *Let $j \in \{1,2,3\}$, $\mathbf{P} \subset \mathbf{P}_\nu$, and f be a function on $\mathbb{R}$. Then*

$$\text{size}_j(\mathbf{P}, f) = \sup_{T \subset \mathbf{P}} \left(\frac{1}{|I_T|} \sum_{P \in T} |\langle f, \phi_{P,j} \rangle|^2 \right)^{1/2}$$

where the supremum is over all j-lacunary trees contained in $\mathbf{P}$ and where the functions $\phi_{P,j}$ are defined in Section 6.

The aim of this section is to prove:

Proposition 9.1. *Let $s > 1$, $r > 2$, and $\mathbf{P} \subset \mathbf{P}_\nu$. Then for $j \in \{1,2,3\}$*

$$\text{size}_j(\mathbf{P}, f) \;\lesssim_{r,s}\; \sup_{P,P'\in\mathbf{P}} \sup_{I_P \subset I \subset I_{P'}} \left(\frac{1}{|I|} \int |f(x)|^s \widetilde{\chi}_I(x)^2 \, dx \right)^{1/s}$$

where the inside supremum is over dyadic intervals.

We will make use of a John-Nirenberg type lemma, proven in [16]

Lemma 9.2. *Let $\{c_P\}_{P\in\mathbf{P}}$ be a collection of coefficients. Let $j \in \{1,2,3\}$. For $1 \leqslant p < \infty$ let*

$$B_p \;=\; \sup_{T \subset \mathbf{P}} \frac{1}{|I_T|^{1/p}} \left\| \left(\sum_{P\in\mathbf{T}} |c_P|^2 \frac{1_{I_P}}{|I_P|} \right)^{1/2} \right\|_{L^p}$$

where the sup is over all j-lacunary trees, and define $B_{1,\infty}$ analogously. Then

$$B_2 \;\leqslant\; CB_{1,\infty}$$

Recall that $\nu = (j_1, j_2, e, i)$ with $0 \leqslant j_1, j_2 \leqslant 4$, $498 \leqslant |e| \leqslant 4002$, and $0 \leqslant i < 4000$. We will also need the following lemma:

Lemma 9.3. *There is a Schwartz function ζ such that for each j-lacunary tree $T \subset \mathbf{P}_\nu$, $j \in \{1,2,3\}$, each sequence of coefficients $\{c_P\}_{P\in T}$ and each integer k with $2^k \leqslant |I_T|$ and $-k = i \mod 4000$ we have*

$$\sum_{P\in T: \; |I_P|\geqslant 2^k} c_P\phi_{P,j}(x) = e^{2\pi i c(\omega_{(P_T)_j})x} \left(2^{-k}\zeta(2^{-k}\cdot) * \left[e^{-2\pi i c(\omega_{(P_T)_j})\cdot} \sum_{P\in T} c_P\phi_{P,j} \right] \right)(x)$$

Proof. One can check that for each P with $|I_P| < |I_T|$, we have

$$\omega_{P,j} \subset c(\omega_{(P_T)_j}) \pm (10|I_P|^{-1}, 10000|I_P|^{-1})$$

(the sign depends on e and j.) Choosing ζ with $\widehat{\zeta} = 1$ on $(-10000, 10000)$ and $\widehat{\zeta}$ supported on $(-10001, 10001)$, the fact that for $P \in T$ we have $-\log_2(|I_P|) = i \mod 4000$ then gives the lemma since $2^{4000} \cdot 10 > 10001$. $\square$

Proof of Proposition 9.1. The case $j = 1, 2$ of the desired conclusion is standard and we could actually get $s = 1$, so the argument below (while applicable for all j) is only needed for $j = 3$. By Lemma 9.2 it suffices to fix a j-lacunary tree $T \subset \mathbf{P}$ and show that

$$
(29) \qquad \frac{1}{|I_T|^{1/s}} \Big\| \Big(\sum_{P \in T} |\langle f, \phi_{P,j} \rangle|^2 \frac{1_{I_P}}{|I_P|} \Big)^{1/2} \Big\|_{L^s}
$$

$$
\lesssim_{r,s} \quad \sup_{P,P' \in T} \sup_{I_P \subset I \subset I_{P'}} \Big(\frac{1}{|I|} \int |f(x)|^s \widetilde{\chi}_I(x)^2 \, dx \Big)^{1/s}.
$$

By dividing T into maximal subtrees with top, we may assume that $P_T \in T$. Let R denote the right side of (29). If $supp(f) \subset \mathbb{R} \backslash 2I_T$ then for each $P \in T$

$$
(30) \qquad |\langle f, \phi_{P,j} \rangle| \frac{1}{|I_P|^{1/2}} \quad \leqslant \quad C \Big(\frac{|I_P|}{|I_T|} \Big)^{M-2} \frac{1}{|I_P|} \int |f(x)| \widetilde{\chi}_{I_P}(x)^2 \, dx
$$

$$
\leqslant \quad C \Big(\frac{|I_P|}{|I_T|} \Big)^{M-2} R \ .
$$

With M sufficiently large (say $M > 3$), it follows that the left side of (29) is bounded above by

$$
\leqslant \quad C \frac{1}{|I_T|^{1/s}} \sum_{P \in T} \Big(\frac{|I_P|}{|I_T|} \Big)^{M-2} |I_P|^{1/s} R
$$

$$
\leqslant \quad CR
$$

Thus, it remains to prove (29) for functions supported on $2I_T$. From this support assumption, we see that it suffices (by choosing $I = I_T$) to show

$$
(31) \qquad \Big\| \Big(\sum_{P \in T} |\langle f, \phi_{P,j} \rangle|^2 \frac{1_{I_P}}{|I_P|} \Big)^{1/2} \Big\|_{L^s} \quad \leqslant \quad C_{r,s} \quad \|f\|_{L^s}.
$$

By the usual Rademacher function argument, the left side of (31) is

$$
\leqslant \quad \sup_{\{b_P\}_{P \in T}} \Big\| \sum_{P \in T} b_P \langle f, \phi_{P,j} \rangle h_{I_P} \Big\|_{L^s}
$$

where the supremum is over all sequences $\{b_P\}$ of ± 1's on T and h_{I_P} is the L^2 normalized Haar function adapted to I_P. After fixing such a sequence and using duality, we are then reduced to showing the bound

$$
(32) \qquad \Big\| \sum_{P \in T} b_P \langle g, h_{I_P} \rangle \phi_{P,j} \Big\|_{L^{s'}} \quad \leqslant \quad C_{r,s} \|g\|_{L^{s'}}
$$

where $s' = s/(s-1)$. Recalling the definition of $\phi_{P,j}$ the left side of (32) is

$$
(33) \qquad \leqslant \quad \Big\| \sum_{\substack{P \in T \\ |I_P| \geqslant 2^k}} b_P \langle g, h_{I_P} \rangle \psi_{P,j}(x) \Big\|_{L_x^{s'}(V_k^r)}.
$$

By Lemma 9.3[1] the display above is

$$\leq \ \left\| 2^{-k}\zeta(2^{-k}\cdot) * \left[\mathbf{e}^{-2\pi \mathbf{i}c(\omega_{(P_T)_j})\cdot} \sum_{P\in T} b_P \left\langle g, h_{I_P} \right\rangle \psi_{P,j} \right](x) \right\|_{L_x^{s'}(V_k^r)}.$$

By Lemma 7.1, the display above is

$$\leq \ C_{r,s} \ \| \sum_{P\in T} b_P \left\langle g, h_{I_P} \right\rangle \psi_{P,j} \|_{L^{s'}}$$

$$\lesssim \ \|g\|_{L^{s'}(\mathbb{R})}$$

the second estimate follows from standard Calderón-Zygmund theory. $\square$

10. A VARIATION-NORM SIZE INCREMENT LEMMA

Proposition 10.1. *Let* $\mathbf{P}\subset\mathbf{P}_\nu$ *be a finite collection of tri-tiles,* $\delta > 0$, $r > 2$, *and* $j \in \{1,2,3\}$. *Suppose that* M *(from the hypotheses of Theorem 6.4) is sufficiently large depending on* δ. *Then for each* α *satisfying*

$$\text{size}_j(\mathbf{P}, f) \ \leq \ \alpha$$

we can find a collection of trees $\mathbf{T}$, *each contained in* $\mathbf{P}$, *satisfying*

$$(34) \qquad\qquad \text{size}_j(\mathbf{P}\backslash \bigcup_{T\in\mathbf{T}} T, f) \ \leq \ \frac{1}{2}\alpha,$$

$$(35) \qquad\qquad \sum_{T\in\mathbf{T}} |I_T| \ \lesssim \ \left(\frac{\|f\|_{L^\infty}}{\alpha} \right)^\delta \alpha^{-2}\|f\|_{L^2}^2 \ .$$

Below, we show how Proposition 10.1 follows from the variation-norm Bessel inequality, Theorem 11.1. The proof uses a standard stopping time argument, which we recall in order to note that our condition (4) in the definition of strong j-disjointness is satisfied.

Below recall that if T is i-overlapping then for each $j \in \{1,2,3\}\backslash\{i\}$ the sign $\epsilon_{i,j} := \text{sgn}(c(\omega_{P_j}) - c(\omega_{(P_T)_j}))$ depends only on i, j, e, and not on T (for details see the discussion after Definition 5.2).

Proof. (reduction to Bessel inequality). By scaling f we may asssume that $\alpha = 1$.

It suffices to show that for each $i \in \{1,2,3\}\backslash\{j\}$ we could find $\mathbf{T}$ satisfying (35) such that for each i-overlapping tree $T \subset \mathbf{P}\backslash\bigcup_{T'\in\mathbf{T}} T'$ we have

$$(36) \qquad\qquad \frac{1}{|I_T|} \sum_{P\in T} |\left\langle f, \phi_{P,j} \right\rangle|^2 \ \leq \ \frac{1}{4}.$$

Let $T_0 = S_0 = \varnothing$. Suppose $T_0, \ldots, T_n$ and $S_0, \ldots, S_n$ have been chosen and set

$$\mathbf{P}_n = \mathbf{P}\backslash(\bigcup_{k=0}^n T_k \cup S_k)$$

If there are no i-overlapping trees $T \subset \mathbf{P}_n$ violating (36) then we finish by setting

$$\mathbf{T} = \{T_k\}_{k=1}^n \cup \{S_k\}_{k=1}^n.$$

---

[1]Here we use the fact that the variation over all k in (33) is the same as the variation restricted to $2^k \leq |I_T|$ and $-k = i \mod 4000$ which is the same as the restricted variation for convolution which is bounded by the variation over all k for the convolution.

Otherwise, if $\mathbf{P}_n$ contains an i-overlapping tree violating (36) then we may choose such a tree $\tilde{T}_{n+1}$ so that $\epsilon_{i,j}c(\omega_{(P_{\tilde{T}_{n+1}})_j})$ is maximal. We then let T_{n+1} be the maximal (with respect to inclusion) i-overlapping tree contained in $\mathbf{P}_n$ which satisfies $P_{T_{n+1}} = P_{\tilde{T}_{n+1}}$. Let S_{n+1} be the maximal (with respect to inclusion) j-overlapping tree contained in $\mathbf{P}_n \backslash T_{n+1}$ which satisfies $P_{S_{n+1}} = P_{\tilde{T}_{n+1}}$

Since $\mathbf{P}$ is finite and $T_n \neq \varnothing$, this process will eventually terminate, yielding some

$$\mathbf{T} := \{T_k\}_{k=1}^N \cup \{S_k\}_{k=1}^N.$$

We claim that the collection $\{T_k\}_{k=1}^N$ is strongly j-disjoint (recall that this is defined in Definition 5.3), and so Proposition 10.1 follows from Theorem 11.1. It suffices to verify condition (3) and condition (4) of Definition 5.3.

In the following, let $k \neq k'$, $P \in T_k$, and $P' \in T_{k'}$.

For (3), assume that $\omega_{P_j} \subsetneq \omega_{P'_j}$. Then $|\omega_{P'_j}| \geqslant 2^{4000}|\omega_{P_j}|$ which implies that $\epsilon_{i,j}c(\omega_{(P_{T_k})_j}) > \epsilon_{i,j}c(\omega_{(P_{T_{k'}})_j})$ and so $k < k'$. But, since $3\omega_{(P_{T_k})_j} \subset 30|e|\omega_{P_j} \subset 3\omega_{P'_j}$ and $P' \notin S_k$ we must have $I_{P'} \cap I_{T_k} = \varnothing$.

Now, to see condition (4), by symmetry it suffices to show that $P'_j \nleqslant (P_{T_k})_j$. First suppose that $P'_j = (P_{T_k})_j$, or equivalently $P' = P_{T_k}$. Then for each $P \in T_k$ we have $P_i \leqslant P'_i \leqslant (P_{T_{k'}})_i$ and so we must have $k < k'$ or else every element of T_k would have already been chosen in $T_{k'}$. But, if $k < k'$ then we would have $P' \in T_k$, contradicting $P' \in T_{k'}$. Now, suppose that $P'_j < (P_{T_k})_j$. Then, as in the verification of (3), $\epsilon_{i,j}c(\omega_{(P_{T_k})_j}) > \epsilon_{i,j}c(\omega_{(P_{T_{k'}})_j})$ and so $k < k'$. But, the fact that $P' \notin S_k$ contradicts $P'_j \leqslant (P_{T_k})_j$. $\square$

11. A variation-norm Bessel inequality

In this section, we fix $\sigma > 0$ and assume that the order M of the wave packets (from the hypotheses of Theorem 6.4) is sufficiently large depending on σ. Our goal here is to prove the following variation-norm Bessel inequality:

Theorem 11.1. *Let $\mathbf{T}$ be a collection of strongly j-disjoint trees, such that*

$$(37) \quad \sup_{I \; dyadic} \left(\frac{1}{|I|} \sum_{\substack{P \in T \\ I_P \subset I}} |\langle f, \phi_{P,j}\rangle|^2\right)^{1/2} \quad \leqslant \quad 1 \quad \leqslant \quad 2\left(\frac{1}{|I_T|} \sum_{P \in T} |\langle f, \phi_{P,j}\rangle|^2\right)^{1/2} \quad ,$$

for each $T \in \mathbf{T}$. Let $\mathbf{P} = \bigcup_{T \in \mathbf{T}} T$. Then

$$\sum_{P \in \mathbf{P}} |\langle f, \phi_{P,j}\rangle|^2 \quad \lesssim_\sigma \|f\|_{L^\infty}^\sigma \|f\|_{L^2}^2.$$

As in [7], we prove Theorem 11.1 via a sequence of reductions.

11.1. **Proof of Theorem 11.1, reduction 1.** Thanks to Lemma 11.2 below, Theorem 11.1 follows immediately from the following proposition:

Proposition 11.1. *Assume $\mathbf{P}$ and $\mathbf{T}$ as in Theorem 11.1. Then for all $\delta > 0$*

$$(38) \quad \sum_{P \in \mathbf{P}} |\langle f, \phi_{P,j}\rangle|^2 \quad \lesssim_\delta \quad \|N_\mathbf{T}\|_{L^\infty}^\delta \int |f(x)|^2 \widetilde{\chi}_I(x)^{10} \; dx \quad ,$$

if $I_T \subset I$ dyadic for all $T \in \mathbf{T}$.

Lemma 11.2 below in turn is a result from [7] where it was proved using a series of interesting Lemmas. To keep the current paper self-contained, we'll sketch a direct proof, which simplifies some arguments in [7]. To formulate the lemma, we first fix some notations. For $\mathbf{S} \subset \mathbf{T}$ let $N_{\mathbf{S}}$ denote $\sum_{T \in \mathbf{S}} 1_{I_T}$, and define

$$\|\mathbf{S}\|_{BMO} := \sup_{I \ dyadic} \frac{1}{|I|} \sum_{T \in \mathbf{S}: I_T \subset I} |I_T| \ .$$

Lemma 11.2. *Let $A, B > 0$ and $0 < \delta < 1$. Let $\mathbf{T}$ be a collection of trees. If for every subset $\mathbf{S}$ of $\mathbf{T}$ it holds that $\|N_{\mathbf{S}}\|_1 \leqslant A\|N_{\mathbf{S}}\|_\infty^\delta$ and $\|\mathbf{S}\|_{BMO} \leqslant B\|N_{\mathbf{S}}\|_\infty^\delta$ then*

$$\|\mathbf{T}\|_{BMO} \lesssim_\delta B^{1/(1-\delta)} \quad , \quad \|N_{\mathbf{T}}\|_1 \lesssim_\delta AB^{\delta/(1-\delta)}$$

Proof of Lemma 11.2. We first show that $\|\mathbf{T}\|_{BMO} \leqslant (3B)^{1/(1-\delta)}$. It suffices to show that for every dyadic interval I_0 it holds that

$$(39) \qquad \frac{1}{|I_0|} \sum_{T \in \mathbf{T}: I_T \subset I_0} |I_T| \leqslant \frac{1}{2}\|\mathbf{T}\|_{BMO} + 3^{\delta/(1-\delta)} B^{1/(1-\delta)}$$

Fix I_0. Let $\mathbf{S}$ contains all elements $T \in \mathbf{T}$ such that $I_T \subset I_0$ and the set $\{S \in \mathbf{T} : I_T \subset I_S \subset I_0\}$ contains at most $(3B)^{1/(1-\delta)}$ elements. Clearly $\|N_{\mathbf{S}}\|_\infty \leqslant (3B)^{1/(1-\delta)}$, therefore by the given assumption we have

$$(40) \qquad \|\mathbf{S}\|_{BMO} \leqslant B(3B)^{\delta/(1-\delta)} = 3^{\delta/(1-\delta)} B^{1/(1-\delta)} \ .$$

Let $\mathbf{J}$ be the set of maximal dyadic intervals $J \subset I_0$ such that the set $\{S \in \mathbf{T} : J \subset I_S \subset I_0\}$ contains more than $(3B)^{1/(1-\delta)}$ elements. Clearly, for every $T \in \mathbf{T}\backslash\mathbf{S}$ such that $I_T \subset I_0$, I_T is contained in one of these J's. It follows that

$$(41) \qquad \frac{1}{|I_0|} \sum_{\substack{T \in \mathbf{T}\backslash\mathbf{S} \\ I_T \subset I_0}} |I_T| \quad \leqslant \quad \frac{1}{|I_0|} \sum_{J \in \mathbf{J}} \sum_{\substack{T \in \mathbf{T} \\ I_T \subset J}} |I_T| \quad \leqslant \quad \frac{\|\mathbf{T}\|_{BMO}}{|I_0|} \sum_{J \in \mathbf{J}} |J| \ .$$

By maximality of J, there exists $T \in \mathbf{T}$ such that $I_T = J$. Let $\mathbf{S}_J$ denote the collection of such T, then $\|\mathbf{S}_J\|_{BMO} = \|N_{\mathbf{S}_J}\|_\infty = |\mathbf{S}_J|$, therefore using the given assumption we obtain $\|N_{\mathbf{S}_J}\|_\infty \leqslant B^{1/(1-\delta)}$. For every $x \in J$ it follows that

$$N_{\mathbf{S}}(x) \quad \geqslant \sum_{T \in \mathbf{T}: J \subsetneq I_T \subset I_0} 1_{I_T}(x) \quad \geqslant \quad \Big(\sum_{T \in \mathbf{T}: J \subset I_T \subset I_0} 1_{I_T}(x) \Big) - B^{1/(1-\delta)}$$

$$\geqslant \quad (3^{1/(1-\delta)} - 1)B^{1/(1-\delta)} \quad > \quad 2 \cdot 3^{\delta/(1-\delta)} B^{1/(1-\delta)} \ .$$

Together with (41), we obtain

$$\frac{1}{|I_0|} \sum_{T \in \mathbf{T}\backslash\mathbf{S}: I_T \subset I_0} |I_T| \quad < \quad \frac{\|\mathbf{T}\|_{BMO}}{|I_0|} \frac{\|N_{\mathbf{S}}\|_{L^1(I_0)}}{2 \cdot 3^{\delta/(1-\delta)}} \quad \leqslant \quad \|\mathbf{T}\|_{BMO} \frac{\|\mathbf{S}\|_{BMO}}{2 \cdot 3^{\delta/(1-\delta)}} \ .$$

Using (40), (39) immediately follows, completing the proof of $\|\mathbf{T}\|_{BMO} \lesssim_\delta B^{1/(1-\delta)}$.

We now free the variables I_0, $\mathbf{S}$, $\mathbf{J}$ to be used for other purposes below.

Fix a large constant $C > 0$ to be chosen later. Let $\mathbf{S}$ contain all $T \in \mathbf{T}$ such that I_T is not a subset of $\{x : N_{\mathbf{T}}(x) > CB^{1/(1-\delta)}\}$. It is clear that $\|N_{\mathbf{S}}\|_\infty \leqslant CB^{1/(1-\delta)}$, so by the given hypothesis $\|N_{\mathbf{S}}\|_1 \leqslant C^\delta AB^{\delta/(1-\delta)}$. It suffices to show that

$$(42) \qquad \|N_{\mathbf{T}\backslash\mathbf{S}}\|_1 = O_\delta(C^{-1}\|N_{\mathbf{T}}\|_1) \ .$$

Indeed, from (42) by choosing C large we obtain $\|N_{\mathbf{T}\backslash\mathbf{S}}\|_1 \leqslant 1/2\|N_{\mathbf{T}}\|_1$, thus $\|N_{\mathbf{T}}\|_1 \leqslant 2\|N_{\mathbf{S}}\|_1$ which implies the desired estimate.

Let $\mathbf{J}$ be the collection of maximal dyadic subintervals of $\{x : N_{\mathbf{T}}(x) > CB^{1/(1-\delta)}\}$. It follows that if $T \in \mathbf{T}\backslash\mathbf{S}$ then I_T is a subset of some element of $\mathbf{J}$. Therefore

$$\|N_{\mathbf{T}\backslash\mathbf{S}}\|_1 \;\leqslant\; \sum_{J\in\mathbf{J}}\sum_{T\in\mathbf{T}:I_T\subset J}|I_T| \;\;\leqslant\;\; \sum_{J\in\mathbf{J}}|J|\|\mathbf{T}\|_{BMO}$$

$$\leqslant\;\; \|\mathbf{T}\|_{BMO}|\{x : N_{\mathbf{T}}(x) > CB^{1/(1-\delta)}\}| \;\;\leqslant\;\; \|\mathbf{T}\|_{BMO}\frac{\|N_{\mathbf{T}}\|_1}{CB^{1/(1-\delta)}} \;\;.$$

Since $\|\mathbf{T}\|_{BMO} = O_\delta(B^{1/(1-\delta)})$, we obtain $\|N_{\mathbf{T}\backslash\mathbf{S}}\|_1 = O_\delta(C^{-1}\|N_{\mathbf{T}}\|_1)$, as desired. $\square$

11.2. **Proof of Theorem 11.1, reduction 2.** We first note that (38) follows from the unweighted version where the factor $\widetilde{\chi}_I^{10}$ is not on the right hand side. Indeed, writing $\langle f,\phi_{P,j}\rangle = \langle f\widetilde{\chi}_I^{10}, \widetilde{\chi}_I^{-10}\phi_{P,j}\rangle$ and using the fact that $\widetilde{\chi}_I^{-10}(x)$ is also a polynomial in x (which implies that $\widetilde{\chi}_I^{-10}\psi_{P,j}$ is still a wave packet adapted to P_j of order sufficiently large, recall also that $\phi_{P,j}$ and $\psi_{P,j}$ are the same if $j = 1, 2$ and related by a variational factor if $j = 3$), (38) follows from applying the unweighted version to $f\widetilde{\chi}_I^{10}$ and the rescaled wave packets.

We now show that the unweighted (38) follows from the following proposition.

Proposition 11.2. *Let $\mathbf{T}$ be strongly j-disjoint. Let $\mathbf{P} = \bigcup_{T\in\mathbf{T}}T$. Then for every $L \geqslant \|N_{\mathbf{T}}\|_\infty$ there exists $\mathbf{P^*} \subset \mathbf{P}$ with the following two properties:*

$$(43) \qquad\qquad \sum_{P\in\mathbf{P}\backslash\mathbf{P^*}}|\langle f,\phi_{P,j}\rangle|^2 \;\lesssim_\delta\; L^\delta\|f\|_2^2 \;\;,$$

$$(44) \qquad\qquad \Big|\bigcup_{P\in\mathbf{P^*}}I_P\Big| \;\lesssim_\delta\; L^{-1}\sum_{T\in\mathbf{T}}|I_T| \;\;.$$

Indeed, apply Proposition 11.2 with $L = C\|N_T\|_\infty$ for a sufficiently large δ-dependent constant C. Now, to get (the unweighted) (38) it suffices to show

$$(45) \qquad\qquad \sum_{P\in\mathbf{P^*}}|\langle f,\phi_{P,j}\rangle|^2 \leqslant \frac{1}{2}\sum_{P\in\mathbf{P}}|\langle f,\phi_{P,j}\rangle|^2 \;\;.$$

Let I_{P_0} be a maximal interval in $\{I_P, P \in \mathbf{P^*}\}$, and remove from $\mathbf{P^*}$ all tri-tiles P such that P is in the same tree as P_0 and $I_P \subset I_{P_0}$. We repeat this process with what is left of $\mathbf{P^*}$. This algorithm gives a collection of tree $\mathbf{T^*}$ such that $\{I_T, T \in \mathbf{T^*}\}$ cover $\mathbf{P^*}$ while $\sum_{T\in\mathbf{T^*}}1_{I_T} \leqslant \sum_{T\in\mathbf{T}}1_{I_T}$. Now, using (37) and (44), (45) follows from the following sequence of estimates and choosing C large:

$$\sum_{P\in\mathbf{P^*}}|\langle f,\phi_{P,j}\rangle|^2 \;\leqslant\; \sum_{T\in\mathbf{T^*}}|I_T| \;\;\leqslant\;\; \|N_{\mathbf{T^*}}\|_\infty\Big|\bigcup_{P\in\mathbf{P^*}}I_P\Big| \;\;\lesssim$$

$$\lesssim_\delta\; (L/C)L^{-1}\sum_{T\in\mathbf{T}}|I_T| \;\;\leqslant\;\; \frac{2}{C}\sum_{P\in\mathbf{P}}|\langle f,\phi_{P,j}\rangle|^2 \;\;.$$

11.3. **Proof of Theorem 11.1, reduction 3.** In this section we reduce Proposition 11.2 to the following more technical result. We first fix some notations. Given $A > 1$ and $d \in \{0, 1, 2\}$, a collection of intervals $\mathcal{I} \subset \mathcal{G}_0$ is (A, d)-sparse if

- for each $I \in \mathcal{I}$, AI is d-regular (see Section 5);
- for each $I, I' \in \mathcal{I}$ with $|I| > |I'|$, we have $|I| \geqslant 2^{100A}|I'|$;
- for each $I, I' \in \mathcal{I}$ with $|I| = |I'|$, we have $\text{dist}(I, I') \geqslant 100A|I'|$.

Proposition 11.3. *Let $A, L, \eta \in (1, \infty)$ and $\epsilon > 0$. Let $\mathbf{T}$ be a collection of strongly j-disjoint trees with $\|N_{\mathbf{T}}\|_{L^\infty} \leqslant L$. Let $\mathbf{P} = \bigcup_{T \in \mathbf{T}} T$, and suppose that $\{I_P : P \in \mathbf{P}\} \cup \{I_T : T \in \mathbf{T}\}$ is (A, d) sparse. Then, there exists $\mathbf{P}^* \subset \mathbf{P}$ such that*

$$|\bigcup_{P \in \mathbf{P}^*} I_P| \lesssim_\eta (A^{-\eta} + L^{-\eta}) \sum_{T \in \mathbf{T}} |I_T| , \quad and$$

$$\sum_{P \in \mathbf{P} \backslash \mathbf{P}^*} |\langle f, \phi_{P,j} \rangle|^2 \lesssim_{\eta,\epsilon} ((AL)^\epsilon + L^8 A^{3-\eta})^2 \|f\|_{L^2}^2.$$

Assuming Proposition 11.3 we will prove Proposition 11.2.

Let $\mathbf{T}$ be j-strongly disjoint and $\mathbf{P} = \bigcup_{T \in \mathbf{T}} T$.

By a simple pigeonholing argument, given any $A > 1$ we may partition $\mathbf{P}$ into subsets $\mathbf{P}_1, \ldots, \mathbf{P}_L$ where $L = O(A^2)$, with the following property: for each $1 \leqslant k \leqslant L$ there exists $d \in \{1, 2, 3\}$ such that $\{I_P, P \in \mathbf{P}_k\}$ is (A, d)-sparse.

This partition also lead to a partition of each $T \in \mathbf{T}$, therefore $\mathbf{P}_k$ is also the union of a collection $\mathbf{T}_k$ of j-strongly disjoint trees with $\|N_{\mathbf{T}_k}\|_\infty \leqslant \|N_{\mathbf{T}}\|_\infty \leqslant L$. Each tree in $\mathbf{T}_k$ could be further decomposed into subtrees such that: each of the new subtrees contains its own top, and the top intervals of the subtrees are disjoint. We obtain $\mathbf{T}_k'$ a collection of trees with top, which is still j-strongly disjoint, furthermore $\|N_{\mathbf{T}_k'}\|_\infty \leqslant \|N_{\mathbf{T}_k}\|_\infty \leqslant L$.

We are now in a position to apply Proposition 11.3 for $\mathbf{T}_k'$, producing $\mathbf{P}_k^* \subset \mathbf{P}_k$. Letting $\mathbf{P}^* = \bigcup_{1 \leqslant k \leqslant L} \mathbf{P}_k^*$ and using $L = O(A^2)$ it follows that

$$|\bigcup_{P \in \mathbf{P}^*} I_P| \lesssim_\eta (A^{2-\eta} + A^2 L^{-\eta}) \sum_{T \in \mathbf{T}} |I_T|$$

$$\sum_{P \in \mathbf{P} \backslash \mathbf{P}^*} |\langle f, \phi_{P,j} \rangle|^2 \lesssim_{\eta,\epsilon} (A^{1+\epsilon} L^\epsilon + L^8 A^{4-\eta})^2 \|f\|_2^2$$

The desired estimates for $\mathbf{P}^*$ follows by letting $\epsilon = \delta/2$, $A = L^{\epsilon/(1+\epsilon)}$, and η large.

11.4. **Proof of Theorem 11.1, reduction 4.** Let $\mathbf{I} = \{I_T : T \in \mathbf{T}\}$. Let $D_\eta > 0$ to be chosen later (depending only on η). For each $I \in \mathbf{I}$ consider the $D_\eta(A^{-\eta} + L^{-\eta})|I|$ neighborhood of its endpoints, i.e. the set of x such that $\text{dist}(x, \partial I) \leqslant D_\eta(A^{-\eta} + L^{-\eta})|I|$. Let E_1 be the union of these neighborhoods over $I \in \mathbf{I}$, and let $E_2 = \{\sum_{I \in \mathbf{I}} (\mathcal{M} 1_I)^2 > L^2\}$. We then let $\mathbf{P}^*$ be the set of all $P \in \mathbf{P}$ such that $I_P \subset E_1 \cup E_2$. Using the Fefferman–Stein maximal inequality, it follows that

$$|\bigcup_{P \in \mathbf{P}^*} I_P| \lesssim_\eta (A^{-\eta} + L^{-\eta}) \sum_{I \in \mathbf{I}} |I| + L^{-(4+2\eta)} \|\sum_{I \in \mathbf{I}} (\mathcal{M} 1_I)^2\|_{2+\eta}^{2+\eta}$$

$$\lesssim_\eta (A^{-\eta} + L^{-\eta}) \|N_{\mathbf{T}}\|_1 .$$

Let $\mathbf{T}_2 = \{T \in \mathbf{T} : I_T \nsubseteq E_1 \cup E_2\}$ and let $\mathbf{I}_2$ denote the set of top intervals of $\mathbf{T}_2$. We now show that $\|\sum_{T \in \mathbf{T}_2} (\mathcal{M}[1_{I_T}])^2\|_\infty \lesssim_\eta L^4$, and this will allow us to reduce Proposition 11.3 to Proposition 11.4 below. Since for each $I \in \mathbf{I}_2$ there are at most L elements of $\mathbf{T}$ with $I_T = I$, it suffices to show that

$$\sum_{I \in \mathbf{I}_2} (\mathcal{M}[1_I])(x)^2 \lesssim_\eta L^3$$

uniform over $x \in \mathbb{R}$, which we fix below. By further dividing $\mathbf{I}_2$ it suffices to prove that $\sum_{I \in \mathbf{I}_3} (\mathcal{M}[1_I])(x)^2 \lesssim_\eta L^2$ for every $\mathbf{I}_3 \subset \mathbf{I}_2$ with the following property: if $I, I' \in \mathbf{I}_3$ and $|I'| < |I|$ then $|I'| \leqslant 2^{-L}|I|$. By further dividing $\mathbf{I}_3$ we may assume

that one of the following situations occur: (i) $x \in I$ for all $I \in \mathbf{I}_3$; (ii) x is on the left of I for all $I \in \mathbf{I}_3$; (iii) x is on the right of I for all $I \in \mathbf{I}_3$.

Now, the desired estimate is clear for (i), so by symmetry we only consider situation (ii). By monotonicity we may assume further that x is the left endpoint of some $J \in \mathbf{I}_3$. By definition of E_1 it follows that for every $I \in \mathbf{I}_3 - \{J\}$ we have $\mathrm{dist}(x, I) \gtrsim L^{-\eta} \max(|J|, |I|)$. Using the (A, d) sparseness of $\mathbf{I}_3$, it follows that

$$
\begin{aligned}
\sum_{I \in \mathbf{I}_3} (\mathcal{M}[1_I])(x)^2 \quad &\lesssim \quad \sum_{|I|=|J|} (\mathcal{M}1_I)(x)^2 \quad + \sum_{|I| \leqslant 2^{-L}|J|} (\mathcal{M}1_I)(x)^2 + \sum_{|I| \geqslant 2^L |J|} (\mathcal{M}1_I)(x)^2 \\
&\lesssim_\eta \quad 1 \quad + \quad L^\eta 2^{-L} \quad + \quad \inf_{x \in J} \sum_{I \in \mathbf{I}_3 : |I| \geqslant 2^L |J|} (\mathcal{M}1_I)(x)^2 \\
&\lesssim \quad L^2 \qquad \text{(using the definition of } E_2\text{)}.
\end{aligned}
$$

Proposition 11.4. *Let $A, L, \eta > 1$ and $\epsilon > 0$. Let $\mathbf{T}$ be strongly j-disjoint with*

$$
(46) \qquad \qquad \|\sum_{T \in \mathbf{T}} (\mathcal{M}[1_{I_T}])^2\|_{L^\infty} \quad \leqslant \quad L \quad .
$$

Let $\mathbf{P} = \bigcup_{T \in \mathbf{T}} T$. Assume that $\{I_P : P \in \mathbf{P}\} \cup \{I_T : T \in \mathbf{T}\}$ is (A, d) sparse, and

$$
(47) \qquad \qquad \sup_{x \in I_P} \mathrm{dist}(x, \partial I_T) \geqslant D_\eta A^{-\eta} |I_T|
$$

for all $P \in \mathbf{P}$, $T \in \mathbf{T}$. Then for D_η sufficiently large depending on η it holds that

$$
\sum_{P \in \mathbf{P}} |\langle f, \phi_{P,j}\rangle|^2 \quad \lesssim_{\eta,\epsilon} \quad ((AL)^\epsilon + L^2 A^{3-\eta})^2 \|f\|_{L^2}^2.
$$

11.5. Proof of Proposition 11.4. For convenience of notation, assume without loss of generality that $A^2 L$ is an integer. By duality, it suffices to show

$$
(48) \qquad \qquad \|\sum_{P \in \mathbf{P}} b_P \phi_{P,j}\|_{L^2} \quad \lesssim_{\eta,\epsilon} \quad (AL)^\epsilon + L^2 A^{3-\eta}
$$

for every sequence $\{b_P\}_{P \in \mathbf{P}}$ such that $\|b\|_{\ell^2(\mathbf{P})} = 1$, which we will fix below.

Let $\mathcal{J} = \{I_T : T \in \mathbf{T}\}$ and let $\mathcal{J}_A = \{(I_T)_A : T \in \mathbf{T}\}$ where $(I_T)_A$ is an interval in $\mathcal{G}_d$ (guaranteed by (A, d) sparsity) such that $AI_T \subset (I_T)_A \subset 3AI_T$.

From the (A, d) sparsity, it is clear that the map from $I_T \to (I_T)_A$ is bijective from $\mathcal{J}$ to $\mathcal{J}_A$ and that if $I_T \subsetneq I_{T'}$ then $(I_T)_A \subsetneq (I_{T'})_A$.

We now decompose $\mathcal{J}_A$ into "layers". Let $\mathcal{J}_{A,1}$ be the set of maximal intervals in $\mathcal{J}_A$ and for $m \geqslant 1$ let $\mathcal{J}_{A,m+1}$ be the set of maximal intervals in $\mathcal{J}_A \backslash \bigcup_{n=1}^m \mathcal{J}_{A,n}$.

Now, since $(I_T)_A \subset 3AI_T$ for each T, using (46) we have

$$
\|\sum_{J \in \mathcal{J}_A} 1_J\|_{L^\infty} \quad \leqslant \quad 16A^2 \|\sum_T (\mathcal{M}[1_{I_T}])^2\|_\infty \quad \leqslant \quad 16A^2 L \quad .
$$

Thus, $\mathcal{J}_{A,1}, \ldots, \mathcal{J}_{A,16A^2 L}$ partition $\mathcal{J}_A$. Letting $\mathcal{J}_m = \{J \in \mathcal{J} : (J)_A \in \mathcal{J}_{A,m}\}$ it follows that $\mathcal{J}_1, \ldots \mathcal{J}_{16A^2 L}$ partition $\mathcal{J}$. Thanks to (A, d) sparsity of $\mathcal{J}$ again, this partition is consistent with the usual set inclusion ordering in $\mathcal{J}$, in the sense that if $J \in \mathcal{J}_m$, $J' \in \mathcal{J}_n$, and $J \subsetneq J'$ then $m > n$.

For $J \in \mathcal{J}$ let m be such that $J \in \mathcal{J}_m$, and define

$$
\mathbf{P}_J \quad := \quad \{P \in \mathbf{P} : I_P = J\}
$$

$$
\mathbf{P}_{<J} \quad := \quad \{P \in \mathbf{P} : I_P \subsetneq J \text{ but } I_P \nsubseteq J' \text{ for all } J' \in \bigcup_{m'>m} \mathcal{J}_{m'}\}.
$$

We obtain the following partition of $\mathbf{P}$:

$$(49) \qquad \mathbf{P} \;=\; \bigcup_{1 \leqslant m \leqslant 16 A^2 L} \bigcup_{J \in \mathcal{J}_m} \mathbf{P}_J \cup \mathbf{P}_{<J} \;\;.$$

By definition of $\phi_{P,j}$, it is clear that (48) will follow from the following estimates

$$(50) \qquad \|\sum_{J \in \mathcal{J}} \sum_{P \in \mathbf{P}_J: \; |I_P| \geqslant 2^k} b_P \psi_{P,j}(x)\|_{L_x^2(V_k^r)} \;\; \lesssim_\eta \;\; 1 + \log^2(AL) + A^{-\eta} L$$

$$(51) \qquad \|\sum_{J \in \mathcal{J}} \sum_{P \in \mathbf{P}_{<J}: \; |I_P| \geqslant 2^k} b_P \psi_{P,j}(x)\|_{L_x^2(V_k^r)} \;\; \lesssim_{\eta,\epsilon} \;\; (AL)^\epsilon + L^2 A^{3-\eta} \;\;.$$

11.6. **Proof of** (50). Recall that $\|b\|_{\ell^2(\mathbf{P})} = 1$. Recall that $\psi_{P,j}$ is a wave function of order M, which is assumed sufficiently large compared to η. We first estimate the error term

$$E(x) \;:=\; \sum_{J \in \mathcal{J}: \; x \notin (J)_A} \sum_{P \in \mathbf{P}_J} |b_P \psi_{P,j}(x)| \;.$$

Lemma 11.3. *It holds that*

$$\|\sum_{J \in \mathcal{J}} \sum_{P \in \mathbf{P}_J: \; |I_P| \geqslant 2^k} b_P \psi_{P,j}(x)\|_{V_k^r} \;\; \leqslant \;\; \|\sum_{m=1}^n \sum_{J \in \mathcal{J}_m} \sum_{P \in \mathbf{P}_J} b_P \psi_{P,j}(x)\|_{V_n^r} + 2 E(x) \;.$$

Proof. Using the triangle inequality and the definition of $E(x)$, the left hand side of the desired estimate is bounded above by

$$\leqslant \;\; \|\sum_{J \in \mathcal{J}: \; x \in (J)_A} \sum_{P \in \mathbf{P}_J: \; |I_P| \geqslant 2^k} b_P \psi_{P,j}(x)\|_{V_k^r} + E(x)$$

$$= \;\; \|\sum_{J \in \mathcal{J}: \; x \in (J)_A, \; |J| \geqslant 2^k} \sum_{P \in \mathbf{P}_J} b_P \psi_{P,j}(x)\|_{V_k^r} + E(x) \;\;.$$

Now, the intervals $(J)_A$ that contain x are nested, with larger interval belongs to some $\mathcal{J}_{A,m}$ with smaller m, thus we could bound the last display by

$$\leqslant \;\; \|\sum_{m=1}^n \sum_{J \in \mathcal{J}_m: x \in (J)_A} \sum_{P \in \mathbf{P}_J} b_P \psi_{P,j}(x)\|_{V_n^r} \;\; + \;\; E(x)$$

$$\leqslant \;\; \|\sum_{m=1}^n \sum_{J \in \mathcal{J}_m} \sum_{P \in \mathbf{P}_J} b_P \psi_{P,j}(x)\|_{V_n^r} \;\; + \;\; 2 E(x)$$

finishing the proof. $\square$

Lemma 11.4. *It holds that*

$$\|E(x)\|_{L^2} \;\; \lesssim_\eta \;\; A^{-\eta} L \;.$$

Proof. We note that any $T \in \mathbf{T}$ contributes at most $O(1)$ tri-tiles to each $\mathbf{P}_J$ and such a contribution would necessitate $J \subset I_T$. Thus, $|\mathbf{P}_J| \lesssim \|N_{\mathbf{T}}\|_{L^\infty} \leqslant L$, so

$$\|\sum_{J \in \mathcal{J}} \sum_{P \in \mathbf{P}_J} |I_P|^{1/2} |\psi_{P,j}|\|_{L^\infty} \;\; \lesssim \;\; L \|\sum_{J \in \mathcal{J}} (\mathcal{M}[1_J])^2\|_{L^\infty} \;\; \lesssim \;\; L^2 \;\;.$$

We also have

$$\|1_{\mathbb{R} \setminus A I_P} \psi_{P,j}\|_{L^1} \;\; \lesssim_M \;\; A^{1-M} |I_P|^{1/2}$$

which gives

$$\|\sum_{J\in\mathcal{J}:\ x\notin(J)_A}\sum_{P\in\mathbf{P}_J}|I_P|^{-1/2}|b_P^2\psi_{P,j}(x)|\|_{L^1}\ \lesssim_M\ A^{1-M}\|b\|_{\ell^2(\mathbf{P})}^2\ =\ A^{1-M}\ .$$

Choosing M sufficiently large, depending on η, the desired bound for $\|E\|_2$ follows by an application of Cauchy-Schwarz. $\square$

Applying Lemma 7.3, we see that for any sequence $\{\epsilon_m\}_{m=1}^{16A^2L}\subset\{1,-1\}$ we have

$$\|\sum_{m=1}^{16A^2L}\sum_{J\in\mathcal{J}_m}\sum_{P\in\mathbf{P}_J}\epsilon_m b_P\psi_{P,j}\|_{L^2}\ \lesssim\ (1+\log(L))$$

therefore, by Lemma 7.2 we have

$$\|\sum_{m=1}^{n}\sum_{J\in\mathcal{J}_m}\sum_{P\in\mathbf{P}_J}b_P\psi_{P,j}(x)\|_{L_x^2(V_n^r)}\ \lesssim\ [1+\log(16A^2L)]\cdot[1+\log(L)]$$

which, combined with Lemma 11.4 and Lemma 11.3, gives (50).

11.7. **Proof of** (51). Here, we use two error terms

$$E_1(x)\ =\ \sum_{m=1}^{16A^2L}\sum_{\substack{J\in\mathcal{J}_m\\x\notin J}}\sum_{P\in\mathbf{P}_{<J}}|b_P\psi_{P,j}(x)|$$

$$E_2(x)\ =\ \sum_{m=2}^{16A^2L}\sum_{\substack{J\in\mathcal{J}_m\\x\in J}}\sum_{m'<m}\sum_{\substack{P\in\mathbf{P}_{<J^{m'}}\\|I_P|<|J|}}|b_P\psi_{P,j}(x)|$$

where, if $J\in\mathcal{J}_m$ and $m'<m$ then we let $J^{m'}$ denote the unique element of $\mathcal{J}_{m'}$ such that $(J)_A\subset(J^{m'})_A$.

Lemma 11.5. *It holds that*

$$\|\sum_{J\in\mathcal{J}}\sum_{\substack{P\in\mathbf{P}_{<J}\\|I_P|\geqslant 2^k}}b_P\psi_{P,j}(x)\|_{V_k^r}\ \lesssim\ E_1(x)+E_2(x)+\|\sum_{m=1}^{n}\sum_{J\in\mathcal{J}_m}\sum_{P\in\mathbf{P}_{<J}}b_P\psi_{P,j}(x)\|_{V_n^r}$$

$$+\ \Big(\sum_{m,J\in\mathcal{J}_m}\|\sum_{P\in\mathbf{P}_{<J}:\ |I_P|\geqslant 2^k}b_P\psi_{P,j}(x)\|_{V_k^r}^2\Big)^{1/2}\ .$$

Remark: A simpler analogue of Lemma 11.5 was considered in [7, Lemma 12.2]. Our Lemma 11.5 (and the following Lemma 11.7) in fact fills in a small gap in [7, Lemma 12.2], where an error term similar to E_2 was not treated.

Proof. By the triangle inequality

$$\|\sum_{J\in\mathcal{J}}\sum_{\substack{P\in\mathbf{P}_{<J}\\|I_P|\geqslant 2^k}}b_P\psi_{P,j}(x)\|_{V_k^r}\ \leqslant\ \|\sum_{\substack{J\in\mathcal{J}\\x\in J}}\sum_{\substack{P\in\mathbf{P}_{<J}\\|I_P|\geqslant 2^k}}b_P\psi_{P,j}(x)\|_{V_k^r}+E_1(x)$$

Let $J_1\subsetneqq\ldots\subsetneqq J_N$ be the (nested) intervals in $\mathcal{J}$ that contain x. Choose $k_1,\ldots,k_N$ so that $2^{k_l}=|J_l|$, which (together with N) are functions of x. Then,

the first term on the right of the last display could be rewritten as

$$= \ \|\sum_{1\leqslant\ell\leqslant N}\sum_{\substack{P\in\mathbf{P}_{<J_\ell}\\|I_P|\geqslant 2^k}}b_P\psi_{P,j}(x)\|_{V_k^r} \ \leqslant \ \|\sum_{\ell}\sum_{\substack{P\in\mathbf{P}_{<J_\ell}\\|I_P|\geqslant 2^k,\,|I_P|\geqslant|J_{\ell+1}|}}b_P\psi_{P,j}(x)\|_{V_k^r}+E_2(x)$$

where, for the inequality above, we use the fact that $J_{\ell+1}\subsetneqq J_\ell$ and so $(J_{\ell+1})_A\subsetneqq (J_\ell)_A$. Using a long jump/short jump decomposition of the variation-norm, the first term on the right side of the inequality above is $\leqslant \mathcal{A}_1+\mathcal{A}_2$, where

$$\mathcal{A}_1 \ = \ \|\sum_{\ell}\sum_{P\in\mathbf{P}_{<J_\ell}:\,|I_P|\geqslant 2^{k_n},\,|I_P|\geqslant 2^{k_{\ell+1}}}b_P\psi_{P,j}(x)\|_{V_n^r}$$

$$\mathcal{A}_2 \ = \ 2\,(\sum_n\|\sum_\ell\sum_{\substack{P\in\mathbf{P}_{<J_\ell}\\|I_P|\geqslant 2^k,\,|I_P|\geqslant 2^{k_{\ell+1}}}}b_P\psi_{P,j}(x)\|_{V_{k_{n+1}\leqslant k<k_n}^r}^2)^{1/2}.$$

It is clear that

$$\mathcal{A}_1 \ = \ \|\sum_{\ell<n}\sum_{\substack{P\in\mathbf{P}_{<J_\ell}\\|I_P|\geqslant 2^{k_{\ell+1}}}}b_P\psi_{P,j}(x)\|_{V_n^r} \ \leqslant \ \|\sum_{\ell<n}\sum_{P\in\mathbf{P}_{<J_\ell}}b_P\psi_{P,j}(x)\|_{V_n^r}+E_2(x)$$

$$= \ \|\sum_{m=1}^n\sum_{\substack{J\in\mathcal{J}_m\\x\in J}}\sum_{P\in\mathbf{P}_{<J}}b_P\psi_{P,j}(x)\|_{V_n^r}+E_2(x)$$

$$\leqslant \ \|\sum_{m=1}^n\sum_{J\in\mathcal{J}_m}\sum_{P\in\mathbf{P}_{<J}}b_P\psi_{P,j}(x)\|_{V_n^r}+E_1(x)+E_2(x) \ ,$$

$$\mathcal{A}_2 \ = \ 2\Big(\sum_n\Big\|\sum_{\substack{P\in\mathbf{P}_{<J_n}\\|I_P|\geqslant 2^k}}b_P\psi_{P,j}(x)+\sum_{\ell<n}\sum_{\substack{P\in\mathbf{P}_{<J_\ell}\\|I_P|\geqslant 2^{k_{\ell+1}}}}b_P\psi_{P,j}(x)\Big\|_{V_{k_{n+1}\leqslant k<k_n}^r}^2\Big)^{1/2}$$

$$= \ 2\Big(\sum_n\|\sum_{\substack{P\in\mathbf{P}_{<J_n}\\|I_P|\geqslant 2^k}}b_P\psi_{P,j}(x)\|_{V_{k_{n+1}\leqslant k<k_n}^r}^2\Big)^{1/2}$$

$$\leqslant \ 2\Big(\sum_{m,J\in\mathcal{J}_m}\|\sum_{\substack{P\in\mathbf{P}_{<J}\\|I_P|\geqslant 2^k}}b_P\psi_{P,j}(x)\|_{V_k^r}^2\Big)^{1/2}.$$

□

Lemma 11.6. *It holds that*

$$\|E_1(x)\|_{L^2} \ \lesssim_\eta \ LA^{1-\eta} \ .$$

Proof. Using Cauchy-Schwartz it suffices to show that for each m

$$\|\sum_{J\in\mathcal{J}_m:\,x\notin J}\sum_{P\in\mathbf{P}_{<J}}|b_P\psi_{P,j}(x)|\|_{L^2} \ \lesssim_\eta \ L^{1/2}A^{-\eta}\|b_P\|_{\ell^2(\bigcup_{J\in\mathcal{J}_m}\mathbf{P}_{<J})} \ .$$

Using the Fefferman-Stein maximal inequality and the fact that the intervals in $\mathcal{J}_m$ are disjoint, it suffices to prove that if $J\in\mathcal{J}_m$ and $x\notin J$ then

$$\sum_{P\in\mathbf{P}_{<J}}|b_P\psi_{P,j}(x)| \ \lesssim_\eta \ L^{1/2}A^{-\eta}\|b_P\|_{\ell^2(\mathbf{P}_{<J})}|J|^{-1/2}(\mathcal{M}[1_J](x))^2.$$

Now, if T intersects $P_{<J}$ then $J \subset I_T$, therefore using (46) we see that at most L trees in $\mathbf{T}$ contribute a given $\mathbf{P}_{<J}$. Thus, using Cauchy Schwarz it suffices to show that, for each $T \in \mathbf{T}$,

$$(52) \qquad \sum_{P \in T \cap \mathbf{P}_{<J}} |\psi_{P,j}(x)|^2 \quad \lesssim_\eta \quad A^{-2\eta}|J|^{-1}(\mathcal{M}[1_J](x))^4.$$

Choosing D_η large enough, the (A, d) sparsity and (47) imply that for each P in the sum above

$$\sup_{y \in I_P} \operatorname{dist}(y, \partial J) \quad \geqslant \quad D_\eta A^{-\eta} 2^{50A}|J|^{1/2}|I_P|^{1/2}$$

$$\geqslant \quad 2^{49A}|J|^{1/2}|I_P|^{1/2} \quad .$$

Recall that M is the order of the wave packet $\psi_{P,j}$. Thus, for $x \notin J$, choosing M large enough we obtain

$$|\psi_{P,j}(x)|^2 \quad \lesssim_M \quad 2^{-49A(M-4)}\left(\frac{|J|}{|I_P|}\right)^{-(M-4)}|I_P|^{-1}(\mathcal{M}[1_{I_P}](x))^4$$

$$\lesssim_\eta \quad A^{-2\eta}\left(\frac{|J|}{|I_P|}\right)^{-2}|J|^{-1}(\mathcal{M}[1_J](x))^4.$$

Summing over $P \in \mathbf{P}_{<J} \cap T$ we obtain (52). $\square$

Lemma 11.7. *It holds that*

$$\|E_2(x)\|_{L^2_x} \quad \lesssim_\eta \quad L^2 A^{3-\eta}$$

Proof. By Cauchy-Schwarz, it suffices to show that for $1 \leqslant m' < m$ we have

$$\|\sum_{\substack{J \in \mathcal{J}_m \\ x \in J}} \sum_{\substack{P \in \mathbf{P}_{<Jm'} \\ |I_P| < |J|}} |b_P \psi_{P,j}(x)|\|_{L^2_x} \quad \lesssim_\eta \quad L^{1/2}A^{-\eta}\|b_P\|_{\ell^2(\bigcup_{J \in \mathcal{J}_{m'}} \mathbf{P}_{<J})}$$

The above bound will follow, by Cauchy-Schwarz, from the following two estimates

$$(53) \qquad \|\sum_{\substack{J \in \mathcal{J}_m \\ x \in J}} \sum_{\substack{P \in \mathbf{P}_{<Jm'} \\ |I_P| < |J|}} |I_P|^{1/2}|\psi_{P,j}(x)|\|_{L^\infty} \quad \lesssim_\eta \quad LA^{-2\eta}$$

$$(54) \qquad \|\sum_{\substack{J \in \mathcal{J}_m \\ x \in J}} \sum_{\substack{P \in \mathbf{P}_{<Jm'} \\ |I_P| < |J|}} |I_P|^{-1/2}|b_P^2 \psi_{P,j}(x)|\|_{L^1} \quad \lesssim \quad \|b_P\|^2_{\ell^2(\bigcup_{J \in \mathcal{J}_{m'}} \mathbf{P}_{<J})}.$$

To see (53) fix x and choose the unique $J \in \mathcal{J}_m$ with $x \in J$. As in Lemma 11.6, it suffices to show that for each $T \in \mathbf{T}$

$$(55) \qquad \sum_{P \in \mathbf{P}_{<Jm'} \cap T: \, |I_P| < |J|} |I_P|^{1/2}|\psi_{P,j}(x)| \quad \lesssim_\eta \quad A^{-2\eta}.$$

Choosing D_η large, it follows (as in Lemma 11.6) that the following holds for every P in the sums above:

$$\inf_{y \in I_P} \operatorname{dist}(y, \partial J) \quad \geqslant \quad 2^{49A}|J|^{1/2}|I_P|^{1/2} \quad .$$

Since $|I_P| < |J|$ and $P \in \mathbf{P}_{<Jm'}$, it follows that $I_P \cap J = \varnothing$. Since $x \in J$, using (A, d) sparseness and (47) we obtain

$$\operatorname{dist}(x, I_P) \quad \geqslant \quad 2^{49A}|J|^{1/2}|I_P|^{1/2} \quad .$$

Due to the restriction of the sum to tiles in a single tree, each dyadic interval is the time interval of at most $O(1)$ tri-tiles, and so for each $k > 0$

$$\sum_{P \in \mathbf{P}_{<Jm'} \cap T: \ |I_P| = 2^{-k}|J|} |I_P|^{1/2} |\psi_{P,j}(x)| \ \lesssim_M \ 2^{-49A(M-1)} 2^{-k(M-1)/2} \ ,$$

and summing over k gives (55).

To see (54) simply use the fact that the intervals in $\mathcal{J}_m$ are pairwise disjoint to estimate the left side by

$$\leq \ \| \sum_{P \in \bigcup_{J \in \mathcal{J}_{m'}} \mathbf{P}_{<J}} |I_P|^{-1/2} |b_P^2 \psi_{P,j}(x)| \|_{L^1} \ \lesssim \ \|b\|_{\ell^2(\bigcup_{J \in \mathcal{J}_{m'}} \mathbf{P}_{<J})}^2$$

□

Applying Lemma 7.2 and Lemma 7.3 as in the proof of (50) we have

$$\| \sum_{m=1}^{n} \sum_{J \in \mathcal{J}_m} \sum_{P \in \mathbf{P}_{<J}} b_P \psi_{P,j}(x) \|_{L_x^2(V_n^r)} \ \lesssim \ 1 + \log^2(AL)$$

Thus, using Lemma 11.5, Lemma 11.6, and Lemma 11.7, to finish the proof of (51) it suffices to establish the following inequality (for each m and $J \in \bigcup_m \mathcal{J}_L$):

$$(56) \qquad \| \sum_{P \in \mathbf{P}_{<J}: \ |I_P| \geq 2^k} b_P \psi_{P,j}(x) \|_{L_x^2(V_k^r)} \ \lesssim_\epsilon \ L^\epsilon \|b_P\|_{\ell^2(\mathbf{P}_{<J})}.$$

Let $\mathbf{T}_J$ be the collection of trees in $\mathbf{T}$ which contribute to $\mathbf{P}_{<J}$. As above, we have $|\mathbf{T}_J| \leq L$. For each $T \in \mathbf{T}_J$ let $\xi_T = c(\omega_{(P_T)_j})$. Then, for each $P \in T \in \mathbf{T}_J$ we have

$$\omega_{P_j} \ \subset \ (\xi_T - 10|e| \cdot |\omega_{P_j}|, \xi_T + 10|e| \cdot |\omega_{P_j}|).$$

Furthermore, from condition (4) in the definition of strong j-disjointness and the fact that $J \subset I_T$ for each $T \in \mathbf{T}_J$, we have that

$$\operatorname{dist}(\xi_T, \omega_{P_j}) \ \geq \ |\omega_{P_j}|/4$$

for each $P \in \mathbf{P}_{<J}$ and each $T \in \mathbf{T}_J$. Therefore, if we let

$$R_k \ = \ \bigcup_{T \in \mathbf{T}_j} (\xi_T - 10|e|2^{-k}, \xi_T + 10|e|2^{-k})$$

and Π_k be the Fourier projection operator $\Pi_k[f] = \mathcal{F}^{-1}[1_{R_k} \hat{f}]$ then, for each $k = -i \mod 4000$ we have

$$(57) \qquad \sum_{\substack{P \in \mathbf{P}_{<J} \\ |I_P| \geq 2^k}} b_P \psi_{P,j} \ = \ \Pi_k[\sum_{P \in \mathbf{P}_{<J}} b_P \psi_{P,j}].$$

Thus, by Theorem 8.1 and Lemma 7.3 we have (56).

12. Concluding the proof of Theorem 6.4

Let $\mathbf{P}$ be a finite subset of $\mathbf{P}_\nu$. Our aim is to prove that the trilinear form

$$(58) \qquad \Lambda_\mathbf{P}(f_1, f_2, f_3) \ = \ \left\langle \sum_{P \in \mathbf{P}} |I_P|^{-1/2} \langle f_1, \phi_{P,1} \rangle \langle f_2, \phi_{P,2} \rangle \phi_{P,3}, f_3 \right\rangle$$

satisfies restricted weak-type estimates with exponents α arbitrarily close to any vertex of A define by (14), with implicit constants uniform over $\mathbf{P}$. We'll consider neighborhoods of $A_1(-\frac{1}{2}, \frac{1}{2}, 1)$, the other vertices could be treated similarly.

By (dyadic) dilation symmetry we can assume $|F_1| \in [1/2, 1)$. Fix $s > 1$ close to 1 to be chosen later, and choose

$$B \;=\; \bigcup_{j=1}^{3} \{\mathcal{M}^s[1_{F_j}] \geqslant C|F_j|^{1/s}\} \;,$$

with C sufficiently large so that $|B| \leqslant \frac{1}{4}$. Let $|f_1| \leqslant 1_{F_1 - B}$ and $|f_2| \leqslant 1_{F_2}$ and $|f_3| \leqslant 1_{F_3}$. Decompose $\mathbf{P} = \bigcup_{k \geqslant 0} \mathbf{P}_k$ where

$$\mathbf{P}_k = \{P \in \mathbf{P} : 2^k \leqslant 1 + \mathrm{dist}(I_P, B^c)/|I_P| < 2^{k+1}\}.$$

For $P \in \mathbf{P}_k$ we have

$$
\begin{aligned}
\sup_{I_P \subset I} \left(\frac{1}{|I|} \int |f_j(x)|^s \widetilde{\chi}_I(x)^2 \, dx \right)^{1/s} &\lesssim\; 2^{k/s} \sup_{I_P \subset I} \left(\frac{1}{|2^k I|} \int |f_j(x)|^s \widetilde{\chi}_{2^k I}(x)^2 \, dx \right)^{1/s} \\
&\lesssim\; 2^{k/s} \inf_{x \in 2^k I_P} \mathcal{M}^s[1_{F_j}](x) \\
&\lesssim\; 2^{k/s} |F_j|^{1/s}
\end{aligned}
$$

Therefore, by Proposition 9.1, for $j = 2, 3$ we have

$$(59) \qquad\qquad S_j \;:=\; \mathrm{size}_j(\mathbf{P}_k, f_j) \;\lesssim\; 2^{k/s}|F_j|^{1/s}.$$

Here (and below) the implicit constants may depend on r, s, and β_i (defined below). Now, when $j = 1$ we will obtain the improved estimate

$$(60) \qquad\qquad S_1 \;:=\; \mathrm{size}_1(\mathbf{P}_k, 1_{B^c} f_1) \leqslant C 2^{-(M-2)k}$$

by exploiting the fact that the interval I in the last sup has to be contained inside another $I_{P'}$ for some $P' \in \mathbf{P}_k$.

Now, applying Proposition 10.1 repeatedly, we obtain a decomposition of $\mathbf{P}_k$ into collections of trees $(\mathbf{T}_n)_{n \in \mathbb{Z}}$ with

$$(61) \qquad\qquad \sum_{T \in \mathbf{T}_n} |I_T| \;\lesssim\; 2^n,$$

such that for any $T \in \mathbf{T}_n$ we have

$$(62) \qquad\qquad \mathrm{size}_i(T, f_i) \;\lesssim\; 2^{-n/(2s)}|F_i|^{1/(2s)}.$$

Now, for any tree T we have

$$(63) \qquad\qquad \sum_{P \in T} |I_P|^{-1/2} \prod_{i=1}^{3} |\langle f_i, \phi_{P,i} \rangle| \;\leqslant\; 3|I_T| \prod_{i=1}^{3} \mathrm{size}_i(T, f_i).$$

To see (63), by further decomposing T if needed we may assume that T is i-overlapping for some $i \in \{1, 2, 3\}$. Then estimating

$$|I_P|^{-1/2}|\langle f_i, \phi_{P,i} \rangle| \;\leqslant\; \mathrm{size}_i(T, f_i)$$

and applying Cauchy-Schwarz to estimate the remaining bilinear sum by

$$\lesssim\; |I_T| \prod_{j \in \{1,2,3\} \setminus \{i\}} \mathrm{size}_j(T, f_j)$$

one obtains (63).

Applying (61), (62), (63), we obtain

$$|\Lambda_k(f_1, f_2, f_3)| \;\lesssim\; \sum_n 2^n \prod_{i=1}^{3} \min(S_i, 2^{-n/(2s)}|F_i|^{1/(2s)}) \;,$$

$$|\Lambda_k(f_1, f_2, f_3)| \;:=\; \sum_{P \in \mathbf{P}_k} |I_P|^{-1/2} \prod_{i=1}^{3} |\langle f_i, \phi_{P,i} \rangle|.$$

For any $\beta_1, \beta_2, \beta_3 \in [0, 1]$, we obtain

$$|\Lambda_k(f_1, f_2, f_3)| \;\lesssim\; S_1 S_2 S_3 \sum_n 2^n \min\left(1, 2^{-n\frac{\beta_1 + \beta_2 + \beta_3}{2s}} \prod_{i=1}^{3} |F_i|^{\frac{\beta_i}{2s}} S_i^{-\beta_i}\right) \;.$$

The above estimate is a two sided geometric series if we choose β_i's such that $\beta_1 + \beta_2 + \beta_3 > 2s$ (which is possible for s close to 1). Letting $\gamma_i := 2s\beta_i/(\beta_1 + \beta_2 + \beta_3)$ we obtain

$$|\Lambda_k(f_1, f_2, f_3)| \;\lesssim\; \prod_{i=1}^{3} S_i^{1-\gamma_i} |F_i|^{\gamma_i/(2s)}$$

$$\lesssim\; 2^{\frac{k}{s}(2 - \gamma_2 - \gamma_3 - s(M-2)(1-\gamma_1))} \left(\prod_{i=1}^{3} |F_i|^{1-\frac{\gamma_i}{2}}\right)^{1/s} \qquad \text{(using (59), (60))}.$$

Again assuming that $\beta_1 + \beta_2 + \beta_3 > 2s$ we are guaranteed $\gamma_1 < 1$ and so, choosing M large enough depending on β we may sum in k to conclude

$$|\Lambda(f_1, f_2, f_3)| \;\lesssim\; \left(\prod_{i=1}^{3} |F_i|^{1-\frac{\gamma_i}{2}}\right)^{1/s}$$

Since $|F_1| \sim 1$, we can ignore its contribution in the above estimate. Now, by sending $(s, \beta_1, \beta_2, \beta_3)$ to $(1, 1, 1, 0)$ inside the region $\{\beta_1 + \beta_2 + \beta_3 > 2s\} \cap \{0 \leqslant \beta_1, \beta_2, \beta_3 \leqslant 1 < s\}$, we obtain the desired claim.

13. Proof of Theorem 6.5

The proof of Theorem 6.5 is entirely similar to the proof of Theorem 6.4, essentially the main difference is that variation-norm estimates such as the continuous Lépingle inequality (see Lemma 7.1) is replaced by the classical Littlewood–Paley square function estimate. We briefly discuss the cosmetic changes, the details are left to the reader. We may define $\phi_{P,j} = \psi_{P,j}$ for $j = 1, 2$, and $\phi_{P,3} = \psi_{P,3} d_n$ if $|I_P| = 2^n$ and 0 otherwise.

Now, the sizes are defined exactly as before, and to get the size estimates for $\text{size}_3(\mathbf{P}, f)$ (as in Proposition 9.1) we use the same proof, the only difference is near the end we appeal to the classical L^p estimates for the Littlewood–Paley square functions associated with scales of the underlying tree, instead of the continuous Lépingle inequality.

Now, to get the size increment estimate (as in Proposition 10.1) we use the same reduction to a Bessel inequality as in Theorem 11.1. To prove this Bessel estimate for the new $\phi_{P,3}$, we follow the same sequence of reductions and the proof reduces to proving Proposition 11.4 with the new modified wave packets. We perform the same partition of $\mathbf{P}$ as in (49), and it suffices to show the following two analogues

of (50) and (51). Below we let S_k denote the ℓ_2 sum of a sequence indexed by k and (b_P) is a sequence on $\mathbf{P}$ with normalized $\ell^2(\mathbf{P})$ norm.

$$(64) \qquad \|\sum_{J \in \mathcal{J}} \sum_{P \in \mathbf{P}_J:\; |I_P| = 2^k} b_P \psi_{P,3}(x)\|_{L_x^2(S_k)} \quad \lesssim_\eta \quad 1 + \log^2(AL) + A^{-\eta} L$$

$$(65) \qquad \|\sum_{J \in \mathcal{J}} \sum_{P \in \mathbf{P}_{<J}:\; |I_P| = 2^k} b_P \psi_{P,3}(x)\|_{L_x^2(S_k)} \quad \lesssim_{\eta,\epsilon} \quad (AL)^\epsilon + L^2 A^{3-\eta} \quad .$$

The proofs of these two estimates are similar. We'll use the same error terms $E(x)$, $E_1(x)$, and $E_2(x)$, and using analogues of Lemma 11.3 and Lemma 11.5 the proofs of (64) and (65) reduce to proving

$$(66) \qquad \|(\sum_{m=1} (\sum_{J \in \mathcal{J}_m} \sum_{P \in \mathbf{P}_J} b_P \psi_{P,3})^2)^{1/2}\|_2 \quad \lesssim \quad [1 + \log(16 A^2 L)][1 + \log(L)]$$

$$(67) \qquad \|(\sum_{m=1} (\sum_{J \in \mathcal{J}_m} \sum_{P \in \mathbf{P}_{<J}} b_P \psi_{P,3})^2)^{1/2}\|_2 \quad \lesssim \quad [1 + \log(16 A^2 L)][1 + \log(L)]$$

$$(68) \qquad \|(\sum_k (\sum_{P \in \mathbf{P}_{<J}:\; |I_P| = 2^k} b_P \psi_{P,3})^2)^{1/2}\|_2 \quad \lesssim_\epsilon \quad L^\epsilon \|b_P\|_{\ell^2(\mathbf{P}_{<J})}$$

We note that the square norm is bounded above by the 2-variation norm. Thus, using Lemma 7.2 the estimates (66) and (67) follow from Lemma 7.3. Similarly, using Proposition 8.2 and the Fourier projection representation (57), the estimate (68) follows from Lemma 7.3.

References

[1] Jeremy Avigad and Jason Rute. Oscillation and the mean ergodic theorem for uniformly convex banach spaces. *Ergodic Theory and Dynamical Systems*, to appear.

[2] J. Bourgain. Double recurrence and almost sure convergence. *J. Reine Angew. Math.*, 404:140–161, 1990.

[3] Ciprian Demeter. Pointwise convergence of the ergodic bilinear Hilbert transform. *Illinois J. Math.*, 51(4):1123–1158, 2007.

[4] Ciprian Demeter. On some maximal multipliers in L^p. *Rev. Mat. Ibero.*, 26(3):947–964, 2010.

[5] Ciprian Demeter. Improved range in the return times theorem. *Canad. Math. Bull.*, 55(4):708–722, 2012.

[6] Ciprian Demeter, Michael T. Lacey, Terence Tao, and Christoph Thiele. Breaking the duality in the return times theorem. *Duke Math. J.*, 143(2):281–355, 2008.

[7] Ciprian Demeter, Terence Tao, and Christoph Thiele. Maximal multilinear operators. *Trans. Amer. Math. Soc.*, 360(9):4989–5042, 2008.

[8] Yen Do, Camil Muscalu, and Christoph Thiele. Variational estimates for paraproducts. *Revista Mat. Ibero.*, 28(3):857–878, 2012.

[9] Yen Do, Richard Oberlin, and Eyvindur Palsson. Variational bounds for a dyadic model of the bilinear hilbert transform. *Illinois J. Math.*, 53(2):105–119, 2013.

[10] Roger L. Jones, Andreas Seeger, and James Wright. Strong variational and jump inequalities in harmonic analysis. *Trans. Amer. Math. Soc.*, 360(12):6711–6742, 2008.

[11] Vjeko Kovac. Quantitative norm convergence of double ergodic averages associated with two commuting group actions. *Ergodic Theory Dyn. Syst.*, to appear, 2014.

[12] Michael Lacey and Christoph Thiele. On Calderón's conjecture. *Ann. of Math. (2)*, 149(2):475–496, 1999.

[13] Michael Lacey and Christoph Thiele. A proof of boundedness of the Carleson operator. *Math. Res. Lett.*, 7(4):361–370, 2000.

[14] Michael T. Lacey. The bilinear maximal functions map into L^p for $2/3 < p \leqslant 1$. *Ann. of Math. (2)*, 151(1):35–57, 2000.

[15] Camil Muscalu, Terence Tao, and Christoph Thiele. Multi-linear operators given by singular multipliers. *J. Amer. Math. Soc.*, 15(2):469–496 (electronic), 2002.

[16] Camil Muscalu, Terence Tao, and Christoph Thiele. L^p estimates for the biest. II. The Fourier case. *Math. Ann.*, 329(3):427–461, 2004.

[17] Fedor Nazarov, Richard Oberlin, and Christoph Thiele. A Calderón Zygmund decomposition for multiple frequencies and an application to an extension of a lemma of Bourgain. *Math. Res. Lett.*, 17(2-3):529–545, 2010.

[18] Christoph Thiele. The maximal quartile operator. *Rev. Mat. Iberoamericana*, 17(1):107–135, 2001.

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF VIRGINIA, CHARLOTTESVILLE, VA 22904-4137, USA
E-mail address: yendo@virginia.edu

DEPARTMENT OF MATHEMATICS, FLORIDA STATE UNIVERSITY, TALLAHASSEE, FL 32306-4510, USA
E-mail address: roberlin@math.fsu.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267, USA
E-mail address: eap2@williams.edu