GENERALIZED CARLESON EMBEDDINGS INTO WEIGHTED OUTER MEASURE SPACES

YEN DO AND MARK LEWERS

ABSTRACT. We prove generalized Carleson embeddings for the continuous wave packet transform from $L^p(\mathbb{R}, w)$ into an outer L^p space over $\mathbb{R}\times\mathbb{R}\times(0,\infty)$ for $2 < p < \infty$ and weight $w \in \mathbb{A}_{p/2}$. This work is a weighted extension of the corresponding Lebesgue result in [12] and generalizes a similar result in [9]. The proof in this article relies on L^2 restriction estimates for the wave packet transform which are geometric and may be of independent interest.

1. INTRODUCTION

Lennart Carleson's influential paper [3] in 1966 resolved Lusin's Conjecture by proving the Fourier series of a function $f \in L^2[0,1]$ converges almost everywhere; see also the work of Hunt [14]. The techniques used in Carleson's proof, now referred to as *time-frequency analysis*, have since played an important role in analysis and serve as a tool in proving L^p estimates on modulation invariant integral operators. We highlight for example Fefferman's proof of Lusin's Conjecture in [13] along with Lacey and Thiele's use of time-frequency analysis in their work on the bilinear Hilbert transform [15, 16] and the Carleson operator [17].

The methods of time-frequency analysis usually pass the analysis on a multilinear form Λ to a model sum

$$\Lambda(f_1,...,f_n) \sim \sum_s c_s(\Lambda)\cdot a_{s,1}(f_1)\cdots a_{s,n}(f_n) \tag{1.1}$$

indexed over a discrete collection of rectangles s in the phase plane. The terms within each summand, localized to s, are either dependent on the form Λ or one of the input functions f_j. From there, the rectangles are grouped into specified collections on which the desired estimates are obtainable. As shown by the first author and Thiele in [12], this procedure follows an outer measure framework which focuses on two main steps in proving L^p estimates for multilinear forms. The first step is to estimate the multilinear form Λ by applying Hölder's inequality in the context of outer measures,

$$|\Lambda(f_1,...,f_n)| \leq C\prod_{j=1}^n \left\|F_j(f_j)\right\|_{\mathcal{L}^{p_j}(X,\sigma,\mathcal{S}_j)}. \tag{1.2}$$

Here, $\mathcal{L}^{p_j}(X,\sigma,\mathcal{S}_j)$ is an *outer* L^{p_j} *space* constructed over a suitable outer measure space $(X,\sigma,\mathcal{S}_j)$. The formulation of outer L^p spaces will be discussed in Section 2. The operators F_j are akin to the $a_{s,j}$ in (1.1) and represent a suitable projection of Λ over X. The final steps are then to establish outer measure L^p embeddings on each operator F_j in the form

$$\left\|F_j(f_j)\right\|_{\mathcal{L}^{p_j}(X,\sigma,\mathcal{S}_j)} \leq C(F_j,p_j)\|f_j\|_{L^{p_j}(\mathbb{R})}. \tag{1.3}$$

Estimates such as (1.3) are referred to, see [12, 8], as *generalized Carleson embeddings*.

Date: September 11, 2020.
2010 *Mathematics Subject Classification.* 42B20.
Y.D. partially supported by NSF grant DMS-1800855.

What the outer measure framework reveals is that the crux in establishing inequalities on modulation invariant operators is passed to proving Carleson embeddings of form (1.3). This is seen in several recent articles which focus on solving certain Carleson embeddings in order to obtain L^p estimates for specific operators. In the original article [12] which introduced the outer measure framework, the key result in reproving L^p estimates on the bilinear Hilbert transform with the same restricted range as [15] is a Carleson embedding of the wave packet transform, see (1.4). Di Plinio and Ou in [8] later recovered L^p estimates for the bilinear Hilbert transform in the full range of [16] by proving a localized Carleson embedding for the same wave packet transform; here localized is in the sense of (1.9). We also mention the work by Uraltsev [26] in reproving L^p estimates for the variational Carleson operator, first obtained in [23], which relies on Carleson embeddings for a modified wave packet transform in addition to the wave packet transform (1.4). Note the Carleson embedding results stated in this paragraph are done with respect to functions f in Lebesgue L^p and embeddings into a "non-weighted" outer measure space.

The purpose of this paper is to explore the outer measure framework of time-frequency analysis in the context of weighted inequalities. Specifically, we seek to understand generalized Carleson embeddings

$$\big\|F(f)\big\|_{\mathcal{L}^p(X,\sigma^w,\mathcal{S}^w)} \leq C(p,w)\|f\|_{L^p(\mathbb{R},w)}$$

where w is a weight on $\mathbb{R}$ and $(X,\sigma^w,\mathcal{S}^w)$ is an outer measure space dependent on w. The motivation is in part due to the recent progress in identifying weighted L^p estimates in harmonic analysis. Using a weighted time-frequency analysis based on the model sum (1.1), the first author with Lacey [9, 10] obtained novel weighted estimates for the variational Carleson operator and Walsh counterpart. Part of the analysis in the series focused on inequalities in traditional time-frequency analysis analog to a Carleson embedding (1.3) for weighted L^p functions. We are interested in understanding the embeddings in the outer measure framework.

It is worth mentioning there are suitable alternatives for obtaining weighted L^p estimates in time-frequency analysis which have been recently explored. The application of sparse domination techniques for instance has seen success with the highlight being the remarkable find by Culiuc, Di Plinio, and Ou [6] in determining weighted estimates for the bilinear Hilbert transform, the first of its kind[1]. We point out more recent work with weighted estimates for the bilinear Hilbert transform and similar operators by Cruz-Uribe and Martell [5], and Benea and Muscalu [1, 2]. Note that sparse domination has also been used to obtain weighted norm inequalities for the variational Carleson operator [7] which are an improvement of [9]. It is questionable however if sparse domination can be used to establish weighted norm estimates for operators with less symmetry such as the truncated bilinear Hilbert transform [11] or the biest operator [21, 22] whose weighted results are unknown.

1.1. **Continuous Wave Packet Transform and Main Result.** The space we work over is upper 3-space $X = \mathbb{R}\times\mathbb{R}\times\mathbb{R}_+$ whose coordinates are viewed as parameterizations of symmetries on the class of modulation invariant integral operators. The primary outer L^p embedding map of interest in this work is the *wavelet projection operator* of a function $f:\mathbb{R}\to\mathbb{C}$ into upper 3-space

$$P(f)(y,\eta,t) := f * \phi_{\eta,t}(y) = \int_{\mathbb{R}} f(x) e^{i\eta(y-x)} \frac{1}{t}\phi\Big(\frac{y-x}{t}\Big)\,dx, \qquad (y,\eta,t)\in X \tag{1.4}$$

where $\phi_{\eta,t}(y)$ is a modulated wave function of a Schwartz function ϕ on $\mathbb{R}$ with compact frequency support. We also refer to the formulation $P(f)(y,\eta,t) = \langle f, \phi_{y,\eta,t}\rangle$ where

$$\phi_{y,\eta,t}(x) = e^{-i\eta(y-x)}\frac{1}{t}\overline{\phi\Big(\frac{y-x}{t}\Big)} \tag{1.5}$$

[1] Xiaochun Li [20] has some unpublished results about weighted estimates for the bilinear Hilbert transform.

is the *wave packet* of ϕ at $(y,\eta,t)\in X$, here as usual $\langle f,g\rangle=\int f\overline{g}$. In this regard, $P(f)$ is also referred to as the *continuous wave packet transform* of f.

The wave packet transform (1.4) serves as a projection of modulation invariant operators in upper 3-space. As shown in [12], the wave packet representation of the bilinear Hilbert transform in X is a linear combination of integrals whose integrand is a pointwise product of wave packet transforms. To recover L^p estimates for the bilinear Hilbert transform, the key result in [12, Theorem 5.1] is that the wave packet transform is a generalized Carleson embedding from $L^p(\mathbb{R})$ to some outer L^p space on X for $2<p<\infty$,

$$\|P(f)\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\leq C_{p,\phi}\,\|f\|_{L^p(\mathbb{R})}. \tag{1.6}$$

It was also stated in [12] and later shown in [26] that $P(f)$ is one of two embedding maps arising from the decomposition (1.2) in connection with the Carleson and variational Carleson operators.

The main result of this paper is an extension of (1.6) to $\mathbb{A}_p$ weighted spaces. Given $1<p<\infty$, recall a weight $w:\mathbb{R}\to[0,\infty]$ belongs to the class of $\mathbb{A}_p$ weights if

$$[w]_{\mathbb{A}_p}:=\sup_{\text{interval } I\subset\mathbb{R}}\left(\frac{1}{|I|}\int_I w(x)\,dx\right)\left(\frac{1}{|I|}\int_I w(x)^{-\frac{1}{p-1}}\,dx\right)^{p-1}<\infty. \tag{1.7}$$

Theorem 1. *Fix a Schwartz function ϕ on $\mathbb{R}$ whose Fourier transform is supported in a small neighborhood $(-\delta,\delta)$. Let $2<q<\infty$ and $w\in\mathbb{A}_{q/2}$. Then*

$$\|P(f)\|_{\mathcal{L}^q(X,\sigma^w,\mathcal{S}^w)}\lesssim\|f\|_{L^q(\mathbb{R},w)} \tag{1.8}$$

for all $f\in L^q(\mathbb{R},w)$ where the implicit constant depends on ϕ, q, and $[w]_{\mathbb{A}_{q/2}}$.

The details concerning the outer L^p space in (1.8) are postponed to Section 2. In the scenario $w=1$ is associated with Lebesgue measure, the theorem immediately implies the strong embedding result (1.6) from [12, Theorem 5.1] as $w=1$ is an $\mathbb{A}_p$ weight for all $p>1$.

We envision (1.8) can be used akin to the Lebesgue version (1.6) to establish weighted estimates in time-frequency analysis. The weighted results previously mentioned which use the outer measure framework are based on embeddings which send Lebesgue L^p functions f into non-weighted outer measure spaces. Developing a weighted outer measure framework in time-frequency analysis[2] could lead to natural self-contained proofs and potentially be a tool to examine the open problems previously mentioned. We stress that using the weighted outer measure framework in this work to obtain weighted L^p estimates for modulation invariant operators is beyond the scope of the paper.

Another question not being addressed in this work is the prospect of a localized version of (1.8). While the Carleson embedding of $P(f)$ (1.6) is not bounded for $1\leq p\leq 2$, Di Plinio and Ou [8, Theorem 1] showed there is a localized extension for $1<p<2$ in the sense

$$\|P(f)1_{X\setminus E_f}\|_{\mathcal{L}^q(X,\sigma,\mathcal{S})}\leq C_{p,q}\|f\|_{L^p(\mathbb{R})},\qquad p'<q\leq\infty \tag{1.9}$$

where E_f is an exceptional set dependent on large L^p averages of f. Localized embeddings of form (1.9) are key ingredients in recent papers concerning L^p estimates for modulation invariant operators, cf. [26, 8, 6, 7] but it is open whether a localized version of (1.8) holds; this question is left for further study.

[2] We point out work by Thiele, Treil, and Volberg [25] which uses weighted outer measure spaces in the context of martingale multipliers.

1.2. **Structure of Paper.** In Section 2, we setup the outer measure space over upper 3-space X and the corresponding outer L^p space which is the setting for Theorem 1. The section concludes with relevant properties for general outer L^p spaces which are needed in the paper. Section 3 discusses L^2 restriction estimates for the wave packet transform in upper 3-space. These estimates are key to the proof of Theorem 1 which is pushed to Section 4.

1.3. **Notation.** Given a finite interval I with center c_I, denote aI as the interval with center c_I and length $|aI| = a|I|$. For a fixed finite interval I, let

$$\chi_I(x) = \Big[1 + \Big(\frac{|x - c_I|}{|I|}\Big)^2\Big]^{-1}. \tag{1.10}$$

Let $\mathscr{S}(\mathbb{R})$ denote the space of Schwartz functions on $\mathbb{R}$. We write the Fourier transform of $f \in \mathscr{S}(\mathbb{R})$ as

$$\widehat{f}(\xi) = \int_{\mathbb{R}} e^{-i\xi x} f(x)\, dx.$$

Given a weight $w : \mathbb{R} \to [0, \infty]$, let $w(E) = \int_E w(x)\, dx$ for all Lebesgue measurable sets E on $\mathbb{R}$. When $E = (a, b)$ is an interval on $\mathbb{R}$, we write $w(a, b) = w((a, b))$ for convenience. We denote weighted L^p spaces on $\mathbb{R}$ as $L^p(w) = L^p(\mathbb{R}, w)$ and the norm as $\|f\|_{L^p(w)} = \|f\|_{L^p(\mathbb{R},w)}$ with similar convention for weak L^p spaces. Finally, for a dyadic grid $\mathcal{D}$, we write the dyadic (weighted) L^p maximal functions over $\mathbb{R}$ as

$$M_{p,w}(f)(x) = \sup_{\text{dyadic } Q \ni x} \left(\frac{1}{|Q|} \int_Q |f(x)|^p\, w(x) dx \right)^{1/p}$$

where $M = M_{1,1}$ is the standard dyadic maximal function and $M_p = M_{p,1}$ is the dyadic L^p maximal function.

2. Outer L^p Spaces

This section sets up the outer L^p space over upper 3-space in Theorem 1. This setup is built upon the outer L^p definitions and concepts formulated in [12]. For convenience, we record useful properties of outer L^p spaces in Section 2.2.

2.1. **Outer L^p Spaces over X.** We work with the outer L^p space associated with outer measure space $(X, \sigma, \mathcal{S})$ where $X = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+$ is upper 3 space, σ is a pre-measure on X with respect to a distinguished collection of Borel sets $\mathbb{E}$, and $\mathcal{S}$ is a *size*, i.e., a quasi sub-additive averaging map over each collection $E \in \mathbb{E}$. In keeping with the language developed in time-frequency analysis, the first coordinate of X represents *time*, the second coordinate represents *frequency*, and the third coordinate represents *scale*.

2.1.1. *Outer Measure Spaces and 3D Tents.* The distinguished collection of Borel sets in X which our outer measure space is built over is the collection of *3D tents* (or tents for short) in upper 3-space. Fix a triplet $\Theta = (C_1, C_2, b)$ such that $\min(C_1, C_2) > b > 0$ where b is a sufficiently small parameter to be used later. For each $(x, \xi, s) \in X$, define the 3D tent

$$T_\Theta(x, \xi, s) := \left\{ (y, \eta, t) \in X \ : \ t < s, |y - x| < s - t, -\frac{C_1}{t} < \eta - \xi < \frac{C_2}{t} \right\}.$$

A tent $T_\Theta(x, \xi, s)$ is asymmetric in frequency unless $C_1 = C_2$. An image of the center component in a generic 3D tent $T_\Theta(x, \xi, s)$ with $C_1 \neq C_2$ is shown in Figure 1.

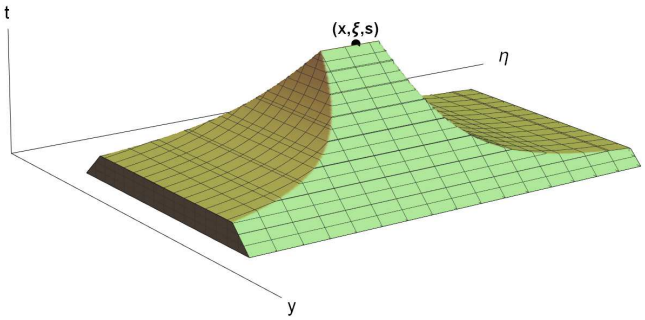

FIGURE 1. The center component of a generic 3D tent $T_\Theta(x,\xi,s)$ with $C_1 \neq C_2$. Note that the image is only a portion of the whole tent $T_\Theta(x,\xi,s)$ as tents have unbounded support in frequency.

We further subdivide a 3D tent $T_\Theta(x,\xi,s)$ into *core* and *lacunary* components. The overlapping or core of the tent $T_\Theta(x,\xi,s)$, denoted by $T^b_\Theta(x,\xi,s)$, is defined as

$$T^b_\Theta(x,\xi,s) := \big\{(y,\eta,t) \in T_\Theta(x,\xi,s) \ : \ |\eta-\xi| \le bt^{-1}\big\}.$$

The lacunary part of the tent $T_\Theta(x,\xi,s)$, denoted by $T^\ell_\Theta(x,\xi,s)$ is the asymmetric shell which is disjoint from the core,

$$T^\ell_\Theta(x,\xi,s) := T_\Theta(x,\xi,s) \setminus T^b_\Theta(x,\xi,s).$$

Figure 2 shows two-dimensional projections of a tent which helps distinguish the separation between the core and lacunary parts. The choice in $b < \min(C_1,C_2)$ is to ensure the shells of T_Θ are nontrivial. In the context of Theorem 1, we set $\delta = 2^{-8}b$ for the frequency support of the kernel ϕ in the wave packet transform.

We now consider a pre-measure over the distinguished collection of 3D tents. Fixing Θ, let $\mathbb{E}$ be the collection of all tents T_Θ in X. For a fixed weight function $w : \mathbb{R} \to [0,\infty]$, define the pre-measure $\sigma^w : \mathbb{E} \to [0,\infty)$ where

$$\sigma^w\big(T_\Theta(x,\xi,s)\big) := w(x-s,x+s) = \int_{\mathbb{R}} 1_{|u-x|<s} w(u)\,du.$$

To extend σ^w to an outer measure μ^w on X, define for an arbitrary set $E \subset X$,

$$\mu^w(E) := \inf\Big\{ \sum_{T_\Theta \in \mathbb{E}'} \sigma^w(T_\Theta) \ : \ E \subset \bigcup_{T_\Theta \in \mathbb{E}'} T_\Theta \Big\}$$

where the infimum is over all countable sub-collections $\mathbb{E}'$ of $\mathbb{E}$ which covers E. It is straightforward to check that $\mu^w(T_\Theta) = \sigma^w(T_\Theta)$ for all tents $T_\Theta \in \mathbb{E}$.

It remains to define a non-negative averaging operator called a *size* on the space $\mathcal{B}(X)$ of Borel-measurable functions over X. A size is a map $\mathcal{S} : \mathcal{B}(X) \to [0,\infty]^{\mathbb{E}}$ such that the following properties hold for all $F,G \in \mathcal{B}(X)$ and all $T_\Theta \in \mathbb{E}$.

(1) [Monotone] If $|F| \le |G|$, then $\mathcal{S}(F)(T_\Theta) \le \mathcal{S}(G)(T_\Theta)$.
(2) [Scaling] If $\lambda \in \mathbb{C}$, then $\mathcal{S}(\lambda F)(T_\Theta) = |\lambda|\,\mathcal{S}(F)(T_\Theta)$.
(3) [Quasi Triangle] There exists constant $C = C(\mathcal{S}) \ge 1$ such that

$$\mathcal{S}(F+G)(T_\Theta) \le C\Big[\mathcal{S}(F)(T_\Theta) + \mathcal{S}(G)(T_\Theta)\Big]. \tag{2.1}$$

The infimum of all such C is the *quasi-triangle constant of size* $\mathcal{S}$.

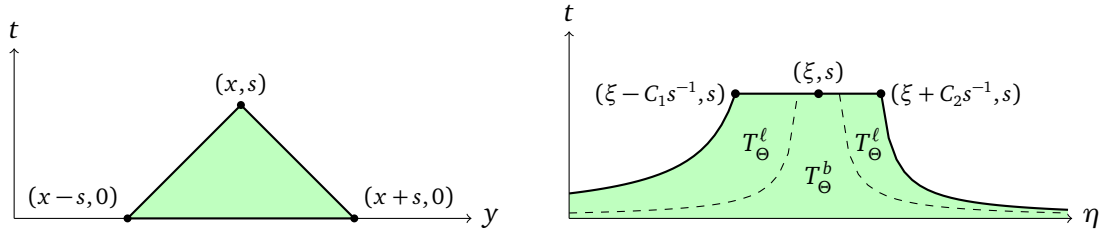

FIGURE 2. Projections of a tent $T_\Theta(x,\xi,s)$ with $C_1 \neq C_2$. The left image is the time-scale projection and the right image is the frequency-scale projection with the partition between the core T^b_Θ and lacunary T^ℓ_Θ regions.

To construct the size in Theorem 1, we first denote S_{T_Θ} as the continuous square function operator of a Borel function $F \in \mathcal{B}(X)$ restricted to the lacunary part of a fixed tent T_Θ,

$$S_{T_\Theta}(F)(u) = \left(\int_{T^\ell_\Theta(x,\xi,s)} |F(y,\eta,t)|^2 1_{|y-u|<t} \, dy\, d\eta \frac{dt}{t} \right)^{1/2}.$$

The size $\mathcal{S}^w$ in Theorem 1 is then a superposition of an L^∞ norm over the core of a tent and an $L^2(w)$ average norm for the square function S_{T_Θ}. Formally,

$$\mathcal{S}^w(F)(T_\Theta(x,\xi,s)) := \frac{1}{\sqrt{w(x-s,x+s)}} \left\| S_{T_\Theta}(F) \right\|_{L^2(w)} + \sup_{(y,\eta,t)\in T^b_\Theta(x,\xi,s)} |F(y,\eta,t)|$$

$$= \left(\frac{1}{w(x-s,x+s)} \int_{T^\ell_\Theta(x,\xi,s)} |F(y,\eta,t)|^2 w(y-t,y+t) \, dy\, d\eta \frac{dt}{t} \right)^{1/2} + \sup_{(y,\eta,t)\in T^b_\Theta(x,\xi,s)} |F(y,\eta,t)|.$$

It is straightforward to check $\mathcal{S}^w$ is a size on X with quasi-triangle constant 1. The triplet $(X,\sigma^w,\mathcal{S}^w)$ as defined above is therefore the outer measure space for this paper. Rather than denoting the space with μ^w, we use the pre-measure σ^w for it is implicitly in terms of the collection of tents $\mathbb{E}$. As we are working with a fixed $\Theta = (C_1, C_2, b)$, we drop the Θ notation out of convenience and write a 3D tent as $T(x,\xi,s) = T_\Theta(x,\xi,s)$. Be aware that the implicit constant in Theorem 1 is also in terms of Θ.

2.1.2. *Outer L^p Spaces.* We formulate the outer integrable spaces with respect to the outer measure space $(X,\sigma^w,\mathcal{S}^w)$. Given $\lambda > 0$ and $F \in \mathcal{B}(X)$, define the *super level measure* associated with μ^w by

$$\mu^w\big(\mathcal{S}^w(F) > \lambda\big) := \inf\Big\{ \mu^w(E) \,:\, E \subset X \text{ Borel} \quad \text{s.t.} \quad \sup_{T\in\mathbb{E}} \mathcal{S}^w(F1_{X\setminus E})(T) \leq \lambda \Big\}.$$

For each $0 < p < \infty$ and $F \in \mathcal{B}(X)$, consider the *outer L^p maps*

$$\|F\|_{\mathcal{L}^p(X,\sigma^w,\mathcal{S}^w)} := \left(\int_0^\infty p\lambda^{p-1} \mu^w\big(\mathcal{S}^w(F) > \lambda\big) \, d\lambda \right)^{1/p}$$

$$\|F\|_{\mathcal{L}^{p,\infty}(X,\sigma^w,\mathcal{S}^w)} := \sup_{\lambda>0} \Big(\lambda^p \mu^w\big(\mathcal{S}^w(F) > \lambda\big) \Big)^{1/p}$$

$$\|F\|_{\mathcal{L}^{\infty,\infty}(X,\sigma^w,\mathcal{S}^w)} = \|F\|_{\mathcal{L}^\infty(X,\sigma^w,\mathcal{S}^w)} := \sup_{T\in\mathbb{E}} \mathcal{S}^w(F)(T)$$

and set $\mathcal{L}^p(X,\sigma^w,\mathcal{S}^w)$, $\mathcal{L}^{p,\infty}(X,\sigma^w,\mathcal{S}^w)$ as the set of Borel functions whose corresponding outer L^p map is finite. As in classical L^p theory, $\mathcal{L}^p(X,\sigma^w,\mathcal{S}^w)$ is contained in $\mathcal{L}^{p,\infty}(X,\sigma^w,\mathcal{S}^w)$.

2.2. **Properties of outer L^p spaces.** We record useful properties concerning outer L^p spaces and their weak versions. The properties hold for general outer measure spaces so we use an abstract outer measure space $(X,\sigma,\mathcal{S})$ (where X is a metric space) and outer L^p space $\mathcal{L}^p(X,\sigma,\mathcal{S})$.

The following proposition from [12, Proposition 3.1] shows $\|\cdot\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}$ is a quasi seminorm.

Proposition 1. *Let $(X,\sigma,\mathcal{S})$ be an outer measure space. Consider $F,G\in\mathcal{B}(X)$ and $0<p\le\infty$.*

(1) *If $|F|\le|G|$ then $\|F\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\le\|G\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}$*
(2) *If $\lambda\in\mathbb{C}$, $\|\lambda F\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}=\lambda\|F\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}$*
(3) *Let C be the quasi-triangle inequality of size $\mathcal{S}$. Then*

$$\|F+G\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\le C_p\Big(\|F\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}+\|G\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\Big) \tag{2.2}$$

$$\text{where } C_p=\begin{cases}2^{1/p}C & 0<p<1\\ 2C & 1\le p<\infty\\ C & p=\infty\end{cases}.$$

The above properties also hold for the weak space $\mathcal{L}^{p,\infty}(X,\sigma,\mathcal{S})$.

Recall size $\mathcal{S}^w$ has quasi-triangle constant of 1. As such, both $\mathcal{L}^p(X,\sigma^w,\mathcal{S}^w)$ and $\mathcal{L}^{p,\infty}(X,\sigma^w,\mathcal{S}^w)$ for $p>1$ have a quasi-triangle constant of 2.

Note the quasi-triangle inequality (2.2) can be generalized to a summation of n functions F_j by

$$\Big\|\sum_{j=1}^n F_j\Big\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\le\sum_{j=1}^n C_p^j\|F_j\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}$$

with a similar inequality if the summation is with respect to $\mathcal{L}^{p,\infty}(X,\sigma,\mathcal{S})$. Assuming the sequence $\|F_j\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}$ has sufficient decay when $j\to\infty$, this can be extended to an infinite series. One application is the following domination property presented by Uraltsev [26, Corollary 2.1].

Proposition 2 (Dominated Convergence)**.** *Fix an outer measure space $(X,\sigma,\mathcal{S})$ and $0<p\le\infty$. Consider Borel functions F, $F_j\in\mathcal{B}(X)$ satisfying the following properties.*

(1) *$|F|\le\limsup_{j\to\infty}|F_j|$ pointwise on X.*
(2) *There exists $C_p'>C_p\ge1$ where C_p is a quasi-triangle constant for $\mathcal{L}^p(X,\sigma,\mathcal{S})$ such that*

$$\sup_{j\ge1}(C_p')^j\|F_{j+1}-F_j\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\lesssim\|F_1\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}. \tag{2.3}$$

Then $\|F\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\lesssim_{C_p,C_p'}\|F_1\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}$. Moreover, if $\|F_1\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\lesssim C$ and the upper estimate in (2.3) is replaced by C, then $\|F\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\lesssim C$. A similar result holds in the context of weak outer L^p spaces.

We finish by recording an outer L^p version of classical Marcinkiewicz interpolation as shown in [12, Proposition 3.5].

Proposition 3 (Outer Marcinkiewicz interpolation)**.** *Let $(X,\sigma,\mathcal{S})$ be an outer measure space and (Y,ν) be a measure space. Fix $1\le p_1<p_2\le\infty$. Let $F:L^{p_j}(Y,\nu)\to\mathcal{B}(X)$ be an operator for $j=1,2$ such that for any $f,g\in L^{p_1}(Y,\nu)+L^{p_2}(Y,\nu)$ and $\lambda\ge0$ we have*

(1) *$|F(\lambda f)|=|\lambda F(f)|$.*
(2) *$|F(f+g)|\le C(|F(f)|+|F(g)|)$ for some constant $C>0$.*
(3) *$\|F(f)\|_{\mathcal{L}^{p_j,\infty}(X,\sigma,\mathcal{S})}\le A_j\|f\|_{L^{p_j}(Y,\nu)}$ for $j=1,2$.*

Suppose $p\in(p_1,p_2)$. Then there exists a constant $C=C(p_1,p_2,p)>0$ such that

$$\|F(f)\|_{\mathcal{L}^p(X,\sigma,\mathcal{S})}\le CA_1^{\theta_1}A_2^{\theta_2}\|f\|_{L^p(Y,\nu)}$$

where $0<\theta_1,\theta_2<1$ satisfy $\frac{1}{p}=\frac{\theta_1}{p_1}+\frac{\theta_2}{p_2}$.

3. L^2 RESTRICTION ESTIMATES FOR THE WAVELET PROJECTION OPERATOR

In this section we collect and prove some local L^2 estimates for the wavelet projection operator

$$P(f)(y,\eta,t) = \langle f, \phi_{y,\eta,t}\rangle, \qquad (y,\eta,t) \in X := \mathbb{R}\times\mathbb{R}\times\mathbb{R}_+,$$

and recall that $\phi_{y,\eta,t}$ is the L^1 normalized wave packet of ϕ at $(y,\eta,t)\in X$, defined in (1.5). For the convenience of the reader, we recall this definition below,

$$\phi_{y,\eta,t}(x) = e^{-i\eta(y-x)}\frac{1}{t}\overline{\phi\Big(\frac{y-x}{t}\Big)},$$

here ϕ is often referred to as the mother wavelet in the literature. Our methods actually work for more general setups, where the mother wavelet ϕ may depend on (y,η,t), as long as it satisfies uniform decay estimates of Schwartz type and uniform frequency support conditions. For the simplicity of the presentation, the proof is only presented in our simpler setup where the wave packets all share the same mother wavelet ϕ.

We look for geometric conditions on $Y\subset X$ such that there is a nontrivial improvement of the basic estimate

$$\Big(\int_Y |\langle f,\phi_{y,\eta,t}\rangle|^2\,dyd\eta dt\Big)^{1/2} \lesssim \sup_{(y,\eta,t)\in Y}|\langle f,\phi_{y,\eta,t}\rangle|\sqrt{|Y|}.$$

Here $|Y|$ is the 3D Lebesgue measure of Y. Note that if Y is the lacunary region of a tent then

$$\Big(\int_Y |\langle f,\phi_{y,\eta,t}\rangle|^2\,dyd\eta dt\Big)^{1/2} \lesssim \|f\|_2$$

thanks to Calderón Zymgund theory. Thus, we expect that the tent structure and $\|f\|_2$ will play an important role in the estimate and in the assumed geometric structure of Y. In fact, if we normalize $\|f\|_2=1$ then a first step is to obtain some geometric condition on Y such that there is an improvement of the following nature

$$\Big(\int_Y |\langle f,\phi_{y,\eta,t}\rangle|^2 dyd\eta dt\Big)^{1/2} \lesssim_\epsilon \Big(1+\sup_{(y,\eta,t)\in Y}|\langle f,\phi_{y,\eta,t}\rangle|\sqrt{|Y|}\Big)^{1-\epsilon},$$

for some $\epsilon\in(0,1)$. This will be the main theme of this section.

3.1. Discrete restriction estimates.

We first consider a discretized variant of the above local L^2 norm, namely geometric conditions on a set of points $E\subset X$ such that there is some $\epsilon\in(0,1)$ such that

$$\Big(\sum_{(y,\eta,t)\in E} t|\langle f,\phi_{y,\eta,t}\rangle|^2\Big)^{1/2} \lesssim \|f\|_2^\epsilon\left(\|f\|_2+\sup_{(y,\eta,t)\in E}|\langle f,\phi_{y,\eta,t}\rangle|\Big(\sum_{(y,\eta,t)\in E} t\Big)^{1/2}\right)^{1-\epsilon}. \tag{3.1}$$

A preliminary result we need is the following standard estimate regarding inner products of wave packets; see [24] for proof in phase plane analog.

Lemma 1. *Let $\phi\in\mathscr{S}(\mathbb{R})$. Consider points (y,η,t), $(y',\eta',t')\in X$ with $t'\geq t$. Then for any integer $N\geq 1$,*

$$\big|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle\big| \lesssim_{\phi,N} (t')^{-1}\Big[1+\Big(\frac{|y-y'|}{t'}\Big)^2\Big]^{-N}.$$

We now define a notion of well-separation for a discrete collection of points in X which is an extension of the analog of well-separation in phase plane analysis.

Definition 1. *A collection of points $E \subset X$ is* **well-separated** *if there exists constants $\alpha, \beta > 0$ such that for all $(y,\eta,t),(y',\eta',t') \in E$, either*

$$|y-y'| > \alpha \max(t,t') \qquad or \qquad |\eta-\eta'| > \beta \max\big(t^{-1},(t')^{-1}\big). \tag{3.2}$$

Over a discrete collection of well-separated points in X, it is possible to obtain a version of estimate 3.1 in the following form.

Lemma 2. *Fix $\phi \in \mathscr{S}(\mathbb{R})$ such that supp $\widehat{\phi} \subset (-\delta,\delta)$. Consider a countable collection of well separated points $E = \{(y_k,\eta_k,t_k)\} \subset X$ with separation constants $\alpha > 0$ and $\beta \geq 4\delta$ as in (3.2). Suppose $s \in (0,1)$. Then for any $f \in L^2(\mathbb{R})$,*

$$\Big(\sum_k t_k |\langle f, \phi_{y_k,\eta_k,t_k}\rangle|^2\Big)^{1/2} \lesssim_{\phi,\alpha,s} \|f\|_2 + \Big[\sup_k \big|\langle f, \phi_{y_k,\eta_k,t_k}\rangle\big| \Big(\sum_k t_k\Big)^{1/2}\Big]^s \|f\|_2^{1-s}. \tag{3.3}$$

The proof of Lemma 2 follows a format similar to corresponding proofs in the phase plane setting, cf. [17, 23, 9], and in the continuum setting [12]. We first prove the $s = 1/3$ case by appealing to standard phase plane analysis on wave packets. The general $0 < s < 1$ case follows from a logarithmic argument.

Proof. Assume the sum of t_k is finite else nothing to show. By scaling, assume without loss of generality that $\|f\|_2 = 1$. For each k, let $\phi_k = \phi_{y_k,\eta_k,t_k}$ and denote $D = \sup_k |\langle f,\phi_k\rangle|$.

Case $s = 1/3$. Let $A = \sum_k t_k |\langle f,\phi_k\rangle|^2 > 0$. Apply Hölder's inequality to get

$$A^2 \leq \Big\| \sum_k t_k \langle f,\phi_k\rangle\, \phi_k \Big\|_2^2 = \sum_{k,j} t_k t_j \,\langle f,\phi_k\rangle\, \langle \phi_k,\phi_j\rangle\, \langle \phi_j, f\rangle\,.$$

Split the right-hand double summation into a "diagonal" term $\{8^{-1}t_k \leq t_j \leq 8t_k\}$ and an "off-diagonal" term $\{8t_j < t_k\} \cup \{8t_k < t_j\}$. By symmetry, the summation is bounded by

$$\sum_{k,j\,:\,8^{-1}t_k \leq t_j \leq 8t_k} t_k t_j \,\langle f,\phi_k\rangle\, \langle \phi_k,\phi_j\rangle\, \langle \phi_j, f\rangle \tag{3.4}$$

$$+ 2 \sum_{k,j\,:\,8t_j < t_k} t_k t_j \,\langle f,\phi_k\rangle\, \langle \phi_k,\phi_j\rangle\, \langle \phi_j, f\rangle\,. \tag{3.5}$$

To estimate the diagonal summand (3.4), use symmetry to bound the smaller of the f-inner products with the large one to obtain

$$\sum_{k,j\,:\,8^{-1}t_k \leq t_j \leq 8t_k} t_k t_j |\,\langle f,\phi_k\rangle\, \langle \phi_k,\phi_j\rangle\, \langle \phi_j, f\rangle\,| \leq 2 \sum_k t_k |\langle f,\phi_k\rangle|^2 \Big(\sum_{j\,:\,8^{-1}t_k \leq t_j \leq 8t_k} t_j |\langle \phi_k,\phi_j\rangle| \Big).$$

We now show the inner summation of wave packets is bounded uniformly over all k, j. For fixed k, assume without loss of generality that $\langle \phi_k,\phi_j\rangle \neq 0$ for all j terms. Application of Lemma 1 and equivalence of scales $t_k \sim t_j$ yield the inner product estimate

$$\sum_{j\,:\,8^{-1}t_k \leq t_j \leq 8t_k} t_j |\langle \phi_k,\phi_j\rangle| \lesssim_\phi \sum_{j\,:\,8^{-1}t_k \leq t_j \leq 8t_k} \Big[1 + \Big|\frac{y_j - y_k}{t_k}\Big|^2\Big]^{-1} \lesssim_\phi \sum_{m=0}^{\infty} (1+m^2)^{-1} \#E_k(m)$$

where

$$E_k(m) := \Big\{(y,\eta,t) \in E \,:\, 8^{-1}t_k \leq t \leq 8t_k,\ \langle \phi_k, \phi_{y,\eta,t}\rangle \neq 0,\ \frac{m}{4}t_k \leq |y_k - y| < \frac{m+1}{4}t_k\Big\}.$$

It remains to show $\#E_k(m)$ is uniformly bounded over all m and k. The nonzero inner product condition between ϕ_k and any $\phi_{y,\eta,t} \in E_k(m)$ implies they have overlapping frequency support with $|\eta_k - \eta| \leq 9\delta t_k^{-1}$. Thus, all $(y,\eta,t) \in E_k(m)$ must lie over a rectangular region with bounded

area dependent on δ. Moreover, well separation of E implies for any (y,η,t), $(y',\eta',t') \in E_k(m)$ that either $|y-y'| > 2^{-3}\alpha t_k$ or $|\eta-\eta'| > 2^{-1}\delta {t_k}^{-1}$. Therefore, the order of $E_k(m) \lesssim \alpha^{-1}$ is uniformly bounded for all k,m and so the diagonal term (3.4) has the estimate

$$\sum_{k,j\,:\,8^{-1}t_k \le t_j \le 8t_k} t_k t_j \langle f,\phi_k\rangle \langle \phi_k,\phi_j\rangle \langle \phi_j, f\rangle \lesssim_{\phi,\alpha} A. \tag{3.6}$$

Turing our attention to the off-diagonal summand (3.5), application of the Cauchy Schwarz inequality yields the bound

$$\sum_{k,j\,:\,8t_j < t_k} t_k t_j \langle f,\phi_k\rangle \langle \phi_k,\phi_j\rangle \langle \phi_j, f\rangle \le A^{1/2} H^{1/2}$$

where

$$H := \sum_k \Big[\sum_{j\,:\,8t_j<t_k} (t_k)^{1/2} t_j |\langle \phi_k,\phi_j\rangle \langle \phi_j, f\rangle| \Big]^2 \le D^2 \sum_k t_k \Big[\sum_{j\,:\,8t_j<t_k} t_j |\langle \phi_k,\phi_j\rangle| \Big]^2.$$

Fix k and without loss of generality assume $\langle \phi_k,\phi_j\rangle \neq 0$ for all j terms in inner summation above. By consideration of frequency support, $|\eta_k - \eta_j| < 2\delta t_j^{-1} < \beta t_j^{-1}$ and so well separation dictates $|y_k - y_j| > \alpha t_k$. By Lemma 1,

$$t_j \big|\langle \phi_k,\phi_j\rangle\big| \lesssim_\phi t_k^{-1}\Big[1+\Big(\frac{y_k-y_j}{t_k}\Big)^2\Big]^{-1} t_j \lesssim_\phi (t_k)^{-1} \int_{[y_j-\frac{\alpha}{4}t_j, y_j+\frac{\alpha}{4}t_j]} \Big[1+\Big(\frac{y_k-x}{t_k}\Big)^2\Big]^{-1} dx.$$

We claim the intervals $[y_j-\frac{\alpha}{4}t_j, y_j+\frac{\alpha}{4}t_j]$ above are all disjoint from each other. Consider another point $(y_\ell,\eta_\ell,k_\ell) \in E$ such that $8t_\ell < t_k$ and $\langle \phi_k,\phi_\ell\rangle \neq 0$. As ϕ_k and ϕ_ℓ have overlapping frequency support (just as ϕ_k and ϕ_j do),

$$|\eta_j - \eta_\ell| \le 2\delta(t_j^{-1}+t_\ell^{-1}) \le \beta \max(t_j^{-1}, t_\ell^{-1}).$$

Well separation between (y_j,η_j,t_j) and $(y_\ell,\eta_\ell,t_\ell)$ then requires $|y_j-y_\ell| \ge \alpha \max(t_j,t_\ell)$ which prevents the intervals from intersecting nontrivially. Hence all such intervals are disjoint and by summing over all such j,

$$\sum_k t_k \Big[\sum_{j\,:\,8t_j<t_k} t_j |\langle \phi_k,\phi_j\rangle| \Big]^2 \lesssim_\phi \sum_k t_k \Big(\frac{1}{t_k} \int_{\mathbb{R}} \Big[1+\Big(\frac{y_k-x}{t_k}\Big)^2\Big]^{-1} dx \Big)^2 \lesssim C_\phi \sum_k t_k.$$

This gives us the off-diagonal inequality

$$\sum_{k,j\,:\,8t_j<t_k} t_k t_j \langle f,\phi_k\rangle \langle \phi_k,\phi_j\rangle \langle \phi_j, f\rangle \lesssim_\phi A^{1/2} D \Big(\sum_k t_k\Big)^{1/2}. \tag{3.7}$$

Applying the diagonal estimate (3.6) and off diagonal estimate (3.7) yields the bound

$$A^2 \lesssim_\phi A + D\Big(\sum_k t_k\Big)^{1/2} A^{1/2}$$

and the desired inequality for case $s = 1/3$ follows by rearrangement.

Case $s \in (0,1)$. Assume $D = \sup_k |\langle f,\phi_k\rangle|$ is finite. Subdivide the collection $E = \bigcup_{j\ge 0} A_j$ where

$$A_j = \big\{k : 2^{-(j+1)}D < |\langle f,\phi_k\rangle| \le 2^{-j}D\big\}.$$

Let $A_{\ge j} = \bigcup_{i\ge j} A_i$ be the union of indices corresponding to the points in all levels below A_j. For fixed j, the sub-collection corresponding to $A_{\ge j}$ is still well separated. Apply the $s = 1/3$ estimate

to $A_{\geq j}$ to get

$$\Big(\sum_{k\in A_{\geq j}} t_k|\langle f,\phi_k\rangle|^2\Big)^{1/2} \lesssim_{\phi,\alpha} 1+\Big[\big(2^{-j}D\big)\Big(\sum_{k\in A_{\geq j}} t_k\Big)^{1/2}\Big]^{1/3}.$$

Note the right-hand side above decays to 1 for large j. By taking $j\geq \max(0,\log_2[D(\sum_k t_k)^{1/2}])$,

$$\sum_{k\in A_{\geq j}} t_k|\langle f,\phi_k\rangle|^2 \lesssim_{\phi,\alpha} 1. \tag{3.8}$$

A similar bound exists over each individual level A_j. Indeed, the points associated with each A_j are themselves well separated so application of the $s=1/3$ case shows

$$\Big(\sum_{k\in A_j} t_k|\langle f,\phi_k\rangle|^2\Big)^{1/2} \lesssim_{\phi,\alpha} 1+\Big[\big(2^{-j}D\big)\Big(\sum_{k\in A_j} t_k\Big)^{1/2}\Big]^{1/3} \sim_{\phi,\alpha} 1+\Big[\Big(\sum_{k\in A_j} t_k|\langle f,\phi_k\rangle|^2\Big)^{1/2}\Big]^{1/3}$$

and by rearrangement,

$$\sum_{k\in A_j} t_k|\langle f,\phi_k\rangle|^2 \lesssim_{\phi,\alpha} 1. \tag{3.9}$$

Fix the smallest positive integer $j\geq \max(0,\log_2[D(\sum_k t_k)^{1/2}])$ and apply (3.8)-(3.9) to obtain the logarithmic estimate

$$\begin{aligned}\sum_k t_k|\langle f,\phi_k\rangle|^2 &= \sum_{k\in A_{\geq j}} t_k|\langle f,\phi_k\rangle|^2+\sum_{i=0}^{j-1}\sum_{k\in A_i} t_k|\langle f,\phi_k\rangle|^2\\ &\lesssim_{\phi,\alpha} 1+\max\Big(0,\log_2[D(\sum_k t_k)^{1/2}]\Big).\end{aligned} \tag{3.10}$$

The passage to arbitrary $s\in(0,1)$ uses the standard logarithmic properties $s\log(x)=\log(x^s)$ and $\max(0,\log x)\leq x$ for all $x>0$,

$$\Big(\sum_k t_k|\langle f,\phi_k\rangle|^2\Big)^{1/2} \lesssim_{\phi,\alpha,s} \left(2s+\Big[D\Big(\sum_k t_k\Big)^{1/2}\Big]^{2s}\right)^{1/2} \lesssim_{\phi,s} 1+\Big[D\Big(\sum_k t_k\Big)^{1/2}\Big]^s. \qquad \square$$

3.2. **Continuous restriction estimates.** In this section, we describe a set of geometric conditions on a set $Y\subset X$ such that there is some nontrivial estimate for

$$\int_Y |\langle f,\phi_{y,\eta,t}\rangle|^2\,dy\,d\eta\,dt.$$

As indicated at the beginning of this section, these conditions will involve some union of tents, or more precisely union of lacunary parts of tents.

We now define a notion of well separation with regards to a collection of 3D tents; assume all tents have the same parameterization $\Theta=(C_1,C_2,b)$. Below by *partial tents* we mean subsets of tents.

Definition 2. *Consider a collection of partial tents $E=\{T_k^*\}$ where $T_k^*\subset T_k=T(x_k,\xi_k,s_k)$ and the T_k^* are pairwise disjoint from each other. The collection E is **well-separated** if there exists constants $\alpha\geq 1$, $\beta>0$, and $B>1$ such that if $(y,\eta,t)\in T_k^*$ and $(y',\eta',t')\in\bigcup_j T_j^*$ satisfy $Bt'<t$ then either*

$$|\eta-\eta'|>\beta(t')^{-1} \qquad or \qquad |x_k-y'|>\alpha(s_k-t'). \tag{3.11}$$

The well separation mandates points sufficiently small in scale with respect to a partial tent T_k^* must either be sufficiently far in frequency from another point in T_k^* or in another tent altogether. Note that when T_k^* is a singleton set with coordinates comparable to the tent T_k and $\alpha > 1$, this reduces to well separation of points. It would be interesting to see an appropriate well separation definition for an arbitrary $Y \subset X$ but this is beyond the scope of the paper.

Similar to the well separation of points, the projection operator $P(f)$ over well separated partial tents is almost L^2.

Lemma 3. *Fix $\phi \in \mathscr{S}(\mathbb{R})$ such that supp $\widehat{\phi} \subset (-\delta, \delta)$. Consider a collection of well separated partial tents $E = \{T_k^* = T^*(x_k, \xi_k, s_k)\}$ with separation constants $\alpha \geq 1$, $\beta \geq 4\delta$, and $B > 1$ as in (3.11). Suppose $s \in (0,1)$. Then for any $f \in L^2(\mathbb{R})$,*

$$\left(\int_{\bigcup_k T_k^*} |\langle f, \phi_{y,\eta,t} \rangle|^2 dy d\eta dt \right)^{1/2} \lesssim \|f\|_2 + \left[\sup_{(y,\eta,t) \in \bigcup_k T_k^*} \left| \langle f, \phi_{y,\eta,t} \rangle \right| \left(\sum_k s_k \right)^{1/2} \right]^s \|f\|_2^{1-s} \tag{3.12}$$

where the implicit constant depends on ϕ, Θ, B, β, and s.

The proof uses the same structure as in Lemma 2 where one proves the $s = 1/3$ case and then generalizes to $0 < s < 1$. In order to show the general $0 < s < 1$ inequality, we need the following simpler estimate.

Lemma 4. *Fix $\phi \in \mathscr{S}(\mathbb{R})$ such that supp $\widehat{\phi} \subset (-\delta, \delta)$. Consider a subset $Y \subset \bigcup_k T_k^*$ with nonzero (three-dimensional) Lebesgue measure where the collection of partial tents T_k^* is well-separated as in Lemma 3. Then*

$$\left(\int_Y |\langle f, \phi_{y,\eta,t} \rangle|^2 dy d\eta dt \right)^{1/2} \lesssim \|f\|_2 + \left[\sup_{(y,\eta,t) \in Y} \left| \langle f, \phi_{y,\eta,t} \rangle \right| \sqrt{|Y|} \right]^{1/3} \|f\|_2^{2/3}.$$

Proof of Lemma 4. By scaling invariant we may assume that $\|f\|_2 = 1$. Let $A = \int_Y |\langle f, \phi_{y,\eta,t} \rangle|^2 dy d\eta dt$. By Cauchy Schwarz,

$$A^2 \lesssim \left\| \int_Y \langle f, \phi_{y,\eta,t} \rangle \, \phi_{y,\eta,t} dy d\eta dt \right\|_2^2$$

$$= \int_{Y \times Y} \langle f, \phi_{y,\eta,t} \rangle \, \langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle \, \langle \phi_{y',\eta',t'}, f \rangle \, dy \, d\eta \, dt \, dy' \, d\eta' \, dt'.$$

Split the integral into three parts: the diagonal $B^{-1}t' \leq t \leq Bt'$, the upper half $t > Bt'$, and the lower half $t' > Bt$. We may estimate the diagonal part by

$$\leq 2 \int_Y |\langle f, \phi_{y,\eta,t} \rangle|^2 \left[\int_{(y',\eta') \in \mathbb{R}^2, B^{-1}t' \leq t \leq Bt'} |\langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle| dy' d\eta' dt' \right] dy d\eta dt.$$

Claim 1. *We have the following uniform estimate for all $(y, \eta, t) \in Y$.*

$$\int_{B^{-1}t \leq t' \leq Bt} \int_{\mathbb{R}^2} |\langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle| \, dy' d\eta' \, dt' \lesssim 1.$$

Using Claim 1, it is clear that the diagonal part is $O(A)$. To see Claim 1 is true, we first note that by taking supremum of t' over $B^{-1}t \leq t' \leq Bt$ and integrating dt' over that region, we obtain

$$\int_{B^{-1}t \leq t' \leq Bt} \int_{\mathbb{R}^2} |\langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle| \, dy' d\eta' \, dt' \lesssim_B \sup_{B^{-1}t \leq t' \leq Bt} \int_{\mathbb{R}^2} t \, |\langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle| \, dy' d\eta'$$

To integrate frequency η', recall that nonzero inner poduct of wave packets $\phi_{y,\eta,t}$ and $\phi_{y',\eta',t'}$ implies they have overlapping support in frequency. Restricing to such wave packets, it follows

that η' has to be contained in an interval of length comparable to t^{-1}. Thus, the last display is bounded from above by

$$\lesssim_{B,\beta} \sup_{B^{-1}t\le t'\le Bt}\sup_{\eta'\in\mathbb{R}}\Big(\int_{\mathbb{R}}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|\,dy'\Big)$$

$$\lesssim_{\phi,B,\beta} \sup_{B^{-1}t\le t'\le Bt}\sup_{\eta'\in\mathbb{R}}\Big(t^{-1}\int_{\mathbb{R}}\Big[1+\Big|\frac{y-y'}{t}\Big|^2\Big]^{-1}dy'\Big)\lesssim_{\phi,B,\beta} 1$$

where Lemma 1 was used in the passage to the second line. This establishes Claim 1 and the estimate on the diagonal terms.

For the off-diagonal terms, we show the details for the upper half term where $t > Bt'$ noting the lower half term may be treated similarly. Denote $D = \sup_{(y,\eta,t)\in Y}|\langle f,\phi_{y,\eta,t}\rangle|$ for notation convenience. By Cauchy-Schwarz,

$$\int_Y \langle f,\phi_{y,\eta,t}\rangle\int_{(y',\eta',t')\in Y:Bt'<t}\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle\,\langle\phi_{y',\eta',t'},f\rangle\,dy'd\eta'dt'dyd\eta dt$$

$$\le A^{1/2}\left(\int_Y\Big(\int_{(y',\eta',t')\in Y:Bt'<t}\big|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle\,\langle\phi_{y',\eta',t'},f\rangle\big|\,dy'd\eta'dt'\Big)^2dyd\eta dt\right)^{1/2}$$

$$\le A^{1/2}D\left(\int_Y\Big(\int_{(y',\eta',t')\in Y:Bt'<t}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|dy'd\eta'dt'\Big)^2dyd\eta dt\right)^{1/2}.$$

It remains to show

$$\sup_{(y,\eta,t)\in Y}\int_{(y',\eta',t')\in Y:Bt'<t}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|dy'd\eta'dt'\lesssim_\phi 1.$$

Fix $(y,\eta,t)\in Y$. As $Y\subset\bigcup_j T_j^*$, it suffices to show the following claim.

Claim 2. *We have the following uniform estimate for all k and $(y,\eta,t)\in T_k^*$,*

$$\sum_j\int_{T_j^*:Bt'\le t}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|\,dy'd\eta'dt'\lesssim_\phi\Big[1+\Big(\frac{s_k-|x_k-y|}{t}\Big)\Big]^{-1}.$$

Fix $(y,\eta,t)\in T_k^*$ and consider $(y',\eta',t')\in T_j^*$ such that $Bt'\le t$. We may assume without loss of generality that $\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle\neq 0$ which means their Fourier transforms have overlapping support. As $Bt'<t$,

$$|\eta-\eta'|\le 2\delta(t')^{-1}\le\beta(t')^{-1}$$

so we are integrating $d\eta'$ with respect to an interval of length $2\beta(t')^{-1}$. By the well-separation criteria, we can also restrict the integral dy' to the region $|y'-x_k|>s_k-t$.

It remains to consider the integral with respect to dt'. For fixed position y', we claim that any points of the form $(y',\eta'',t'')\in\bigcup_j T_j^*$ with $Bt''<t$ and $\langle\phi_{y,\eta,t},\phi_{y',\eta'',t''}\rangle\neq 0$ must be within a factor of B of each other in scale. Indeed, recall $(y',\eta',t')\in T_j^*$ and consider $(y'',\eta'',t'')\in T_m^*$ where $\langle\phi_{y,\eta,t},\phi_{y'',\eta'',t''}\rangle\neq 0$ and $Bt''\le t$. If we were to assume $Bt''\le t'$ then it follows that

$$|\eta'-\eta''|\le 4\delta(t'')^{-1}\le\beta(t')^{-1}$$

and so well-separation dictates

$$|y''-x_j|>s_j-t''>s_j-t'\ge|y'-x_j|.$$

This shows $y''\neq y'$ so setting $y''=y'$ implies that $t''\sim_B t'$.

Given $y' \in \mathbb{R}$, let $E(y')$ be the collection of scales t'' for which there exists η'' and T_j^* such that $(y', \eta'', t'') \in T_j^*$ with $Bt'' < t$ and $\langle \phi_{y,\eta,t}, \phi_{y',\eta'',t''} \rangle \neq 0$. For each $y' \in \mathbb{R}$ with $E(y') \neq \emptyset$, there exists an interval $I(y') = [\alpha(y'), B\alpha(y')]$ which contains $E(y')$. Here, one can check

$$\alpha(y) := B^{-1} \sup_{t'' \in E(y')} t'' < \infty.$$

Applying Lemma 1 the well separation observations above with yields

$$\sum_j \int_{T_j^* : Bt' \leq t} |\langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle| \, dy' d\eta' dt'$$

$$\lesssim \int_{|y'-x_k|>s_k-t} \int_{I(y')} \int_{[\eta-\beta/t', \eta+\beta/t']} |\langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle| \, d\eta' dt' dy'$$

$$\lesssim_B \int_{|y'-x_k|>s_k-t} \left(\sup_{t' \in I(y')} \int_{[\eta-\beta/t', \eta+\beta/t']} t' |\langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle| \, d\eta' \right) dy'$$

$$\lesssim_{\phi,B,\beta} \int_{|y'-x_k|>s_k-t} t^{-1} \Big[1 + \Big(\frac{|y-y'|}{t} \Big)^2 \Big]^{-2} dy' \lesssim \Big[1 + \Big(\frac{s_k - |x_k - y|}{t} \Big) \Big]^{-1}$$

as desired, verifying Claim 2. Collecting the diagonal and off-diagonal estimates, we have

$$A^2 \lesssim A + A^{1/2} D |Y|^{1/2}$$

and from here the desired result follows. □

Proof of Lemma 3. Assume $D = \sup_{(y,\eta,t) \in \bigcup_k T_k^*} |\langle f, \phi_{y,\eta,t} \rangle|$ and the sum of s_k are finite. By scaling, we may assume without loss of generality that $\|f\|_2 = 1$.

Case $s = 1/3$. Let $A := \sum_k \int_{T_k^*} |\langle f, \phi_{y,\eta,t} \rangle|^2 \, dy d\eta dt$. By Cauchy-Schwarz,

$$A^2 \lesssim \sum_{k,j} \int_{T_k^* \times T_j^*} \langle f, \phi_{y,\eta,t} \rangle \, \langle \phi_{y,\eta,t}, \phi_{y',\eta',t'} \rangle \, \langle \phi_{y',\eta',t'}, f \rangle \, dy d\eta dt \, dy' d\eta' dt'.$$

Split the integral the diagonal part $B^{-1}t' \leq t \leq Bt'$, the upper half $t > Bt'$, and the lower half $t' > Bt$. It remains to show the diagonal part is bounded above by $O(A)$ and the off diagonal parts are bounded above by

$$O\left(\Big[\sup_{(y,\eta,t) \in \cup_k T_k^*} |\langle f, \phi_{y,\eta,t} \rangle| \Big] \Big(\sum_k s_k \Big)^{1/2} A^{1/2} \right).$$

This would imply

$$A^2 \lesssim_\phi A + \Big[\sup_{(y,\eta,t) \in \cup_k T_k^*} |\langle f, \phi_{y,\eta,t} \rangle| \Big] \Big(\sum_k s_k \Big)^{1/2} A^{1/2}$$

and the $s = 1/3$ case follows by rearrangement.

To estimate the diagonal part, we have by symmetry

$$\begin{aligned}
&2\sum_{k,j=1}\int_{T_k^*\times T_j^*:B^{-1}t\le t'\le Bt}|\langle f,\phi_{y,\eta,t}\rangle|^2\,|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|\,dyd\eta dt\,dy'd\eta'dt'\\
&\le 2A\sup_{k,(y,\eta,t)\in T_k^*}\left(\sum_j\int_{T_j^*:B^{-1}t\le t'\le Bt}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|\,dy'd\eta'\,dt'\right)\\
&\le 2A\sup_{k,(y,\eta,t)\in T_k^*}\left(\int_{B^{-1}t\le t'\le Bt}\int_{\mathbb{R}^2}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|\,dy'd\eta'\,dt'\right)
\end{aligned}$$

where last part is due the disjointness of T_j^*. Using Claim 1, the last display is clearly $O(A)$.

We now focus on the off-diagonal terms. We will prove the desired estimate for the upper half and the lower half will follow by similar construction. Taking note that the partial tents T_k^* are pairwise disjoint, apply Cauchy-Schwarz to obtain estimates

$$\sum_{k,j}\int_{T_k^*\times T_j^*}\langle f,\phi_{y,\eta,t}\rangle\,\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle\,\langle\phi_{y',\eta',t'},f\rangle\,dyd\eta dt\,dy'd\eta'dt'\le A^{1/2}\Big(\sum_k H_k\Big)^{1/2}$$

where

$$\begin{aligned}
H_k&:=\int_{T_k^*}\left(\sum_j\int_{T_j^*:Bt'\le t}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle\,\langle\phi_{y',\eta',t'},f\rangle|\,dy'd\eta'dt'\right)^2dyd\eta dt\\
&\le\Big[\sup_{(y,\eta,t)\in\cup_k T_k^*}|\langle f,\phi_{y,\eta,t}\rangle|\Big]^2\int_{T_k^*}\left(\sum_j\int_{T_j^*:Bt'\le t}|\langle\phi_{y,\eta,t},\phi_{y',\eta',t'}\rangle|\,dy'd\eta'dt'\right)^2dyd\eta dt
\end{aligned}$$

For simplicity in notation, let D be the supremum above. Using Claim 2 above, we have

$$H_k\lesssim_{\phi,B,\beta}D^2\int_{T_k^*}\Big[1+\Big(\frac{s_k-|x_k-y|}{t}\Big)\Big]^{-2}dyd\eta dt$$

$$\lesssim_{\phi,B,\beta}D^2\int_0^{s_k}\int_{x_k-s_k}^{x_k+s_k}\int_{\xi-C_1t^{-1}}^{\xi+C_2t^{-1}}\Big[1+\Big(\frac{s_k-|x_k-y|}{t}\Big)\Big]^{-2}d\eta dydt\lesssim_{\phi,\Theta}D^2s_k$$

Summing over all k gives

$$\Big(\sum_k H_k\Big)^{1/2}\lesssim_\phi\Big[\sup_{(y,\eta,t)\in\cup_k T_k^*}|\langle f,\phi_{y,\eta,t}\rangle|\Big]\Big(\sum_k s_k\Big)^{1/2}$$

and the desired off-diagonal estimate now follows.

Case $s\in(0,1)$. Generalization to $s\in(0,1)$ is essentially the same argument as in Lemma 2 with appropriate modifications. Recall $D=\sup_{(y,\eta,t)\in\bigcup_k T_k^*}|\langle f,\phi_{y,\eta,t}\rangle|<\infty$. Subdivide the collection E of partial tents into

$$A_j=\Big\{(y,\eta,t)\in\bigcup_k T_k^*\,:\,2^{-(j+1)}D<|\langle f,\phi_{y,\eta,t}\rangle|\le 2^{-j}D\Big\}.$$

Denote $A_{\ge j}=\bigcup_{i\ge j}A_i$, and for convenience let $A_{\ge j,k}=T_k^*\cap A_{\ge j}$. For fixed j, the sub-collection $\{A_{\ge j,k}\}$ of partial tents is well-separated so by applying the $s=1/3$ estimate,

$$\left(\sum_k\int_{A_{\ge j,k}}|\langle f,\phi_{y,\eta,t}\rangle|^2\,dyd\eta dt\right)^{1/2}\lesssim_{\phi,B,\beta}1+\left[\Big(2^{-j}D\Big)\Big(\sum_k s_k\Big)^{1/2}\right]^{1/3}.$$

Note the upper estimate above decays such that for $j \geq \max(0, \log_2[D(\sum_k s_k)^{1/2}])$,

$$\int_{A_{\geq j}} |\langle f, \phi_{y,\eta,t}\rangle|^2 \, dy d\eta dt \lesssim 1. \tag{3.13}$$

To estimate each individual level A_j, use Lemma 4 to obtain

$$2^{-(j+1)} D|A_j|^{1/2} \leq \Big(\int_{A_j} |\langle f, \phi_{y,\eta,t}\rangle|^2 \, dy d\eta dt \Big)^{1/2} \lesssim_{\phi,B,\beta} 1 + \Big[\Big(2^{-j} D \Big) |A_j|^{1/2} \Big]^{1/3}.$$

From this estimate, it follows immediately that $2^{-j} D|A_j|^{1/2} \lesssim 1$, and consequently,

$$\Big(\int_{A_j} |\langle f, \phi_{y,\eta,t}\rangle|^2 \, dy d\eta dt \Big)^{1/2} \lesssim 1. \tag{3.14}$$

Fix the smallest integer $j \geq \max(0, \log_2[D(\sum_k s_k)^{1/2}])$. By (3.13) and (3.14), we obtain the following logarithmic estimate spanning all of $\bigcup_k T_k^* = \bigcup A_j$.

$$\begin{aligned} \sum_k \int_{T_k^*} |\langle f, \phi_{y,\eta,t}\rangle|^2 \, dy d\eta dt &= \int_{A_{\geq j}} |\langle f, \phi_{y,\eta,t}\rangle|^2 \, dy d\eta dt + \sum_{k=1}^{j-1} \int_{A_k} |\langle f, \phi_{y,\eta,t}\rangle|^2 \, dy d\eta dt \\ &\lesssim_{\phi,B,\beta} 1 + \max\Big(0, \log_2[D(\sum_k s_k)^{1/2}]\Big). \end{aligned}$$

The desired result for $s \in (0,1)$ is then obtained by using the same logarithmic argument as in the conclusion of Lemma 2. □

4. Proof of Theorem 1

We start with reductions which reduces the proof of Theorem 1 and make some useful observations. The outline of the proof will be then stated in Section 4.2 followed by the specifics of the proofs in Sections 4.3, 4.4.

4.1. **Preliminary reductions and observations.**

4.1.1. *Reduction - Discrete Parameter Tents.* Following [12], we pass our tent paramterizations to a discrete subset of X and prove Theorem 1 in the discrete parameter setting. The chosen discrete subset is chosen such that tents in $\mathbb{E}$ are *centrally contained* within the discrete parameter tents. Note that a point $(x', \xi', s') \in T(x, \xi, s)$ is centrally contained if the following inequalities hold:

$$2^{-3} s \leq s' \leq 2^{-2} s, \qquad |x' - x| \leq 2^{-4} s, \qquad |\xi' - \xi| \leq 2^{-8} b s^{-1}. \tag{4.1}$$

Consider the subset X_Δ of points $(x, \xi, s) \in X$ such that there exists integers k, m, n satisfying

$$x = 2^{k-4} n, \quad \xi = 2^{-k-8} bm, \quad s = 2^k.$$

Let $\mathbb{E}_\Delta$ be the collection of all tents $T(x, \xi, s)$ with $(x, \xi, s) \in X_\Delta$. We recall the following relation [12, Lemma 5.2] regarding tents in $\mathbb{E}$ and $\mathbb{E}_\Delta$.

Lemma 5. *For any $(x', \xi', s') \in X$ there exists (x, ξ_-, s), $(x, \xi_+, s) \in X_\Delta$, such that (x', ξ', s') is centrally contained in the tents $T(x, \xi_- s)$, $T(x, \xi_+, s)$ as specified in (4.1) and satisfy*

$$T(x', \xi', s') \subset T(x, \xi_-, s) \cup T(x, \xi_+, s),$$

$$T(x', \xi', s') \cap T^b(x, \xi_-, s) \cap T^b(x, \xi_+, s) \subset T^b(x', \xi', s').$$

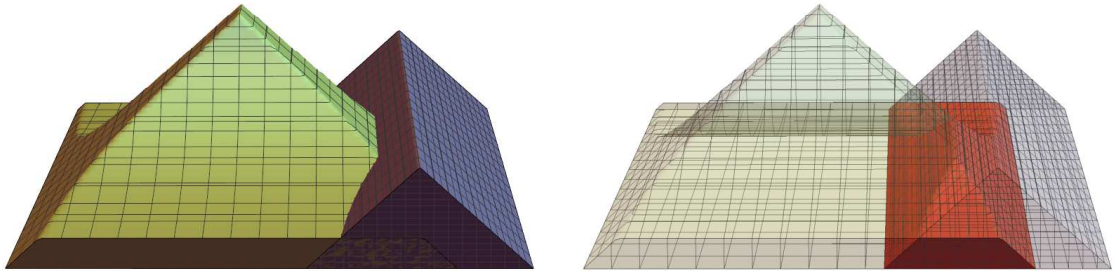

FIGURE 3. The intersection of a 3D tent (in green) and a triangular strip (in blue) is another 3D tent (in red).

Let $\mathcal{S}^w_\Delta$, σ^w_Δ, and μ^w_Δ be S^w, σ^w, and μ^w respectively restricted to the generating sub-collection $\mathbb{E}_\Delta$. Given $T \in \mathbb{E}$, there exist $T^+, T^- \in \mathbb{E}_\Delta$ satisfying $T \subset T^+ \cup T^-$ and

$$\sigma^w(T) \le \sigma^w_\Delta(T^+) + \sigma^w_\Delta(T^-) \le C\sigma^w(T)$$

where right-most inequality uses the doubling property of w. This implies the outer measures μ^w and μ^w_Δ are equivalent. Furthermore, given $F \in \mathcal{B}(X)$ and tent $T \in \mathbb{E}$, application of $T^+, T^- \in \mathbb{E}_\Delta$ as above shows

$$\mathcal{S}^w(F)(T) \le C\Big(\mathcal{S}^w_\Delta(F)(T^+) + \mathcal{S}^w_\Delta(F)(T^-)\Big)$$

where C is dependent on the doubling constant of w.

We therefore have an equivalence of the outer L^p spaces and weak outer L^p spaces

$$\mathcal{L}^p(X, \sigma^w, \mathcal{S}^w) \sim \mathcal{L}^p(X, \sigma^w_\Delta, \mathcal{S}^w_\Delta), \qquad \mathcal{L}^{p,\infty}(X, \sigma^w, \mathcal{S}^w) \sim \mathcal{L}^{p,\infty}(X, \sigma^w_\Delta, \mathcal{S}^w_\Delta)$$

for $1 \le p \le \infty$. Thus, to prove Theorem 1, it suffices to establish the following theorem with respect to the discrete parameter setting.

Theorem 2. *Let $\phi \in \mathscr{S}(\mathbb{R})$ such that supp $\widehat{\phi} \subset (-2^{-8}b, 2^{-8}b)$. Given a locally integrable function f on $\mathbb{R}$, let $P(f)$ be the wave packet transform (1.4). Suppose $2 < q < \infty$ and $w \in \mathbb{A}_{q/2}$. Then*

$$\|P(f)\|_{\mathcal{L}^q(X, \sigma^w_\Delta, \mathcal{S}^w_\Delta)} \lesssim_{\Theta, \phi, q, [w]_{\mathbb{A}_{q/2}}} \|f\|_{L^q(w)}.$$

4.1.2. *Useful Observations.*

Remark 1 (Tent Containment)**.** *The proof of Theorem 2 requires us to pass from a collection of selected tents to another while maintaining well separation in the style of Section 3. To that end, we use the following geometric observation regarding tents.*

Given $x' \in \mathbb{R}$ and $s' > 0$, denote the triangular strip

$$E_{x',s'} = \{(y, \eta, t) \in X \,:\, 0 < t < s', |y - x'| < s' - t\}$$

Consider a tent $T(x, \xi, s) \in \mathbb{E}$ such that $T(x, \xi, s) \cap E_{x',s'}$ is nonempty. Then the intersection itself is a 3D tent $T(x'', \xi, s'') \in \mathbb{E}$ where $(x''-s'', x''+s'') = (x-s, x+s) \cap (x'-s', x'+s')$. Figure 3 illustrates this intersection. As the frequency parameters play no role in this observation, the containment extends to the core and lacunary partial tents:

$$T^b(x, \xi, s) \cap E_{x',s'} = T^b(x'', \xi, s'') \qquad \text{and} \qquad T^\ell(x, \xi, s) \cap E_{x',s'} = T^\ell(x'', \xi, s'').$$

Remark 2 ($\mathbb{A}_p$ weights are not L^1 integrable)**.** *We briefly point out the standard fact that $\mathbb{A}_p$ weights are not integrable on $\mathbb{R}^n$. Otherwise, reverse Hölder's inequality implies for all cubes Q*

$$\int_Q \big[w(x)\big]^r \, dx \lesssim_r |Q|^{1-r}$$

where $r > 1$ is dependent on w. Application of monotone convergence theorem shows $w = 0$ a.e. which isn't an $\mathbb{A}_p$ weight. We state this remark as we haven't found another reference mentioning it.

Remark 3 (Shifted Dyadic Intervals)**.** *While the discrete parameter tents are not necessarily over dyadic intervals, we will implement another pass to work with dyadic grids on $\mathbb{R}$. The dyadic intervals we are interested are*

$$\mathcal{D}_k = \left\{ \left[2^{-j}(m + (-1)^j k/3),\, 2^{-j}(m + 1 + (-1)^j k/3) \right) : j, m \in \mathbb{Z} \right\} \qquad k \in \{0, 1, 2\}$$

where $\mathcal{D}_0$ is the standard dyadic grid, $\mathcal{D}_1$ is the 1/3 shifted grid, and $\mathcal{D}_2$ is the 2/3 shifted grid. We recall the three grids lemma (see [4]) where a finite interval $I \subset \mathbb{R}$ is comparable to a dyadic interval $J \in \mathcal{D}_j$ for some $j \in \{0, 1, 2\}$ such that $I \subset J$ and $3|I| \leq |J| \leq 6|I|$.

4.2. **Structure of Proof.** Fix $2 < q < \infty$ and weight $w \in \mathbb{A}_{q/2}$. To prove Theorem 2, it suffices to establish the following weak outer L^p estimates.

$$\|P(f)\|_{\mathcal{L}^\infty(X, \sigma_\Delta^w, \mathcal{S}_\Delta^w)} \lesssim_{\phi, q, [w]_{\mathbb{A}_{q/2}}} \|f\|_{L^\infty(w)} \tag{4.2}$$

$$\|P(f)\|_{\mathcal{L}^{q,\infty}(X, \sigma_\Delta^w, \mathcal{S}_\Delta^w)} \lesssim_{\phi, q, [w]_{\mathbb{A}_{q/2}}} \|f\|_{L^q(w)} \tag{4.3}$$

If true, reverse Hölder's inequality says (4.3) holds with q replaced by some $q - \epsilon$ and then invoke outer Marcinkiewicz interpolation (see Proposition 3) to pass to the strong estimate at q itself. Moreover, it suffices to show (4.3) holds for a Schwartz function f on $\mathbb{R}$. Indeed, given $f \in L^q(w)$ pick a sequence of Schwartz functions f_k converging to f in $L^q(w)$ norm satisfying

- $\|f_1\|_{L^q(w)} \leq C\|f\|_{L^q(w)}$ and
- $\|f_{k+1} - f_k\|_{L^q(w)} \leq C2^{-10k}\|f\|_{L^q(w)}$.

It is clear $P(f) = \lim_k P(f_k)$ pointwise and if we assume (4.3) holds for Schwartz functions,

$$\|P(f_{k+1}) - P(f_k)\|_{\mathcal{L}^{q,\infty}(X, \sigma_\Delta^w, \mathcal{S}_\Delta^w)} \lesssim \|f_{k+1} - f_k\|_{L^q(w)} \lesssim C2^{-10k}\|f\|_{L^q(w)}.$$

The reduction follows by application of Proposition 2 to the sequence $P(f_k)$.

We prove the outer L^∞ estimate (4.2) in Section 4.3 by appealing to Littlewood-Paley square function estimates to control each tent. Section 4.4 handles the more complicated weak outer L^q estimate (4.3) by good-lambda type arguments restricted to tents with large size. We remain in the discrete parameter setting for the rest of the paper, unless otherwise stated. As such, we drop the Δ notation and denote $\mathbb{E} = \mathbb{E}_\Delta$, $\mathcal{S} = \mathcal{S}_\Delta^w$, and $\mu = \mu_\Delta^w$.

4.3. **The outer L^∞ embedding (4.2).** Consider $f \in L^\infty(w)$. The goal is to show

$$\mathcal{S}\big(P(f)\big)(T(x, \xi, s)) \leq C\|f\|_{L^\infty(w)}$$

with implicit constant C independent of $(x, \xi, s) \in X_\Delta$ and f. As $\mathcal{S}$ is the sum of an L^∞ size and an $L^2(w)$ size, we prove this inequality for each size separately. The estimate for the L^∞ portion is clear from the definition of the wave packet transform.

$$\sup_{(y,\eta,t) \in T^b(x,\xi,s)} |\langle f, \phi_{y,\eta,t} \rangle| \leq \|f\|_\infty \|\phi\|_1 \leq C\|f\|_{L^\infty(w)}.$$

It remains to control the $L^2(w)$ portion of size $\mathcal{S}$,

$$\int_{T^\ell(x,\xi,s)} |P(f)(y, \eta, t)|^2 w(y - t, y + t) dy\, d\eta \frac{dt}{t} \leq Cw(x - s, x + s)\|f\|_\infty^2. \tag{4.4}$$

Fix $(x,\xi,s)\in X_\Delta$ and consider the decomposition $f=f_1+f_2$ where $f_1:=f1_{(x-2s,x+2s)}$. For each (y,η,t) such that $y\in(x-s,x+s)$ and $t<s$,

$$|P(f_2)(y,\eta,t)|\le\int_{[-s,s]^c}\big|f_2(y-z)\frac{1}{t}\phi\Big(\frac{z}{t}\Big)\big|dz\le C\frac{t}{s}\|f\|_\infty.$$

As $(y,\eta,t)\in T^\ell(x,\xi,s)$, we have the containment $(y-t,y+t)\subset(x-s,x+s)$. Therefore

$$\big\|S_T\big(P(f_2)\big)\big\|^2_{L^2(w)}=\int_{T^\ell(x,\xi,s)}|P(f_2)(y,\eta,t)|^2w(y-t,y+t)dydy\frac{dt}{t}$$

$$\le C\int_0^s\int_{x-s}^{x+s}\int_{\xi-C_1t^{-1}}^{\xi+C_2t^{-1}}\Big(\frac{t}{s}\|f\|_\infty\Big)^2w(y-t,y+t)\,d\eta dy\frac{dt}{t}\le Cw(x-s,x+s)\|f\|^2_\infty$$

which establishes the desired estimate on f_2.

For the f_1 piece, it suffices to show the square function estimate

$$\big\|S_T\big(P(h)\big)\big\|_{L^q(w)}\le C_{\phi,q,[w]_{\mathbb{A}_{q/2}}}\|h\|_{L^q(w)},\tag{4.5}$$

for any $h\in L^q(w)$. As $S_T(P(f_1))$ is compactly supported, we can pass to (4.5) by applying Hölder's inequality to the left side of (4.4). From there, appeal to the compact support of f_1 and doubling property of w to control $\|f_1\|_{L^q(w)}$ by the desired result. By symmetry, we only need to establish (4.5) where T^ℓ is replaced by $T^\ell\cap\{\eta>\xi\}$. Consider a change in variables $\eta=\xi+\gamma/t$, and an absolute constant $C'>\max(C_1,C_2)$. It remains to show

$$\Big\|\Big(\int_0^\infty\int_{\mathbb{R}}\int_b^{C'}\big|P(h)(y,\xi+\frac{\gamma}{t},t)\big|^2 1_{|y-u|<t}d\gamma dy\frac{dt}{t^2}\Big)^{1/2}\Big\|_{L^q_u(w)}\le C\,\|h\|_{L^q(w)}.$$

Let $g(u)=h(u)e^{-i\xi u}$ and $M_\gamma\phi(z)=e^{i\gamma z}\phi(z)$. Since $b\le\gamma\le C'$ and $\widehat{\phi}$ is supported in $(-2^{-8}b,2^{-8}b)$, the frequency support of $M_\gamma\phi$ is bounded away from 0 and ∞ and $M_\gamma\phi$ satisfies the usual decay estimates (where the implicit constant can be chosen uniformly over $b\le\gamma\le C'$). Observe that

$$P(h)\big(y,\xi+\frac{\gamma}{t},t\big)=e^{i\xi y}\Big(g*(M_\gamma\phi)_t(y)\Big)$$

where $(M_\gamma\phi)_t(z)=t^{-1}M_\gamma\phi(z/t)$. Uniformly over $b\le\gamma\le C'$, we have

$$\Big\|\Big(\int_0^\infty\int_{\mathbb{R}}\big|P(h)(y,\xi+\frac{\gamma}{t},t)\big|^2 1_{|y-u|<t}dy\frac{dt}{t^2}\Big)^{1/2}\Big\|_{L^q_u(w)}=$$

$$=\Big\|\Big(\int_0^\infty\int_{\mathbb{R}}\big|g*(M_\gamma\phi)_t(y)\big|^2 1_{|y-u|<t}dy\frac{dt}{t^2}\Big)^{1/2}\Big\|_{L^q_u(w)}\le C\|g\|_{L^q(w)}=C\|h\|_{L^q(w)}$$

where the second-to-line inequality is a consequence of $L^q(w)$ boundedness for continuous square function estimates with $w\in\mathbb{A}_q\supset\mathbb{A}_{q/2}$; see [18, 19] for details. This establishes (4.5) and concludes the proof of the outer L^∞ embedding (4.2).

4.4. **The weak outer L^q embedding (4.3).** As mentioned, we may assume that f is a Schwartz function on $\mathbb{R}$. Given $\lambda>0$ we need to find a countable collection of points $Q\subset X_\Delta$ such that

$$\sum_{(x,\xi,s)\in Q}w(x-s,x+s)\lesssim\lambda^{-q}\|f\|^q_{L^q(w)},$$

and for every $T\in\mathbb{E}_\Delta$ we have, with $E:=\bigcup_{(x,\xi,s)\in Q}T(x,\xi,s)$,

$$\mathcal{S}\big(P(f)1_{X\setminus E}\big)(T)\le\lambda.\tag{4.6}$$

We first reduce to the case when $\widehat{f}$ is compactly supported. Indeed, we may select frequencies $\xi_0 = 0 < \xi_1 < \xi_1 < \dots$ such that with $\widehat{f_k} = \widehat{f} 1_{\xi_{k-1} \leq |\xi| \leq \xi_k}$, $k \geq 1$, satisfy

$$\|f_k\|_{L^q(w)} \leq C 2^{-10k} \|f\|_{L^q(w)}.$$

Using the special case to each f_k with $\lambda_k = 2^{-k}\lambda$ we obtain collections Q_k, and clearly

$$\sum_{k \geq 1} \sum_{(x,\xi,s) \in Q_k} w(x-s, x+s) \leq C\lambda^{-q} \sum_k 2^{qk} \|f_k\|_{L^q(w)}^q \leq C\lambda^{-q} \|f\|_{L^q(w)}^q.$$

On the other hand using subadditivity of the size, we may estimate

$$\mathcal{S}\big(P(f)1_{X \setminus E}\big)(T) \leq \sum_{k \geq 1} \mathcal{S}\big(P(f_k)1_{X \setminus E}\big)(T) \leq \sum_{k \geq 1} 2^{-k}\lambda \leq \lambda.$$

Thus from now on we may assume that f is Schwartz such that $\widehat{f}$ is compactly supported.

4.4.1. *Treatment for the L^∞ part of the size.* We first isolate tents $T \in \mathbb{E}_\Delta$ which contain all $(y, \eta, t) \in X$ such that $|P(f)(y, \eta, t)| > \lambda$. Note that by Cauchy-Schwarz,

$$|P(f)(y, \eta, t)| = \Big| \int_{\mathbb{R}} f(y-z) e^{i\eta z} \frac{1}{t} \phi\big(\frac{z}{t}\big) dz \Big| \leq t^{-1/2} \|f\|_2 \|\phi\|_2 = O(t^{-1/2}), \tag{4.7}$$

so if $|P(f)(y, \eta, t)| > \lambda$ then there is an a priori bound $t < O(\lambda^{-2})$. We remark this estimate is not essential; the upper bound is only needed to ensure that the L^∞ selection algorithm described below will terminate and it will never be used quantitatively.

***L^∞* Selection Algorithm**. Suppose there is some $(y, \eta, t) \in X$ such that $|P(f)(y, \eta, t)| > \lambda$. By Lemma 5, we can associate with (y, η, t) a point $(x, \xi, s) \in X_\Delta$ such that (y, η, t) is centrally contained in $T(x, \xi, s)$. In particular s has to be bounded a priori by (4.7) and is always a power of 2. We may therefore choose $(y_1, \eta_1, t_1) \in X$ and $(x_1, \xi_1, s_1) \in X_\Delta$ so that

- $|P(f)(y_1, \eta_1, t_1)| > \lambda$,
- $(y_1, \eta_1, t_1) \in T(x_1, \xi_1, s_1)$ centrally, and
- s_1 is maximal.

We iterate this process. Assume that we have already selected $(y_k, \eta_k, t_k) \in X$ and $(x_k, \xi_k, s_k) \in X_\Delta$ for $1 \leq k < n$. Suppose there is a point $(y, \eta, t) \in X$ outside the union of selected tents $T(x_k, \xi_k, s_k)$ satisfying $|P(f)(y, \eta, t)| > \lambda$. We now choose $(y_n, \eta_n, t_n) \in X$ and $(x_n, \xi_n, s_n) \in X_\Delta$ such that

- $(y_n, \eta_n, t_n) \notin \bigcup_{k=1}^{n-1} T(x_k, \xi_k, s_k)$,
- $|P(f)(y_n, \eta_n, t_n)| > \lambda$,
- $(y_n, \eta_n, t_n) \in T(x_n, \xi_n, s_n)$ centrally, and
- s_n is maximal.

Figure 4 gives a visual representation of the selection. For each $k \geq 1$, denote $T_k = T(x_k, \xi_k, s_k)$ and $I_k = (x_k - s_k, x_k + s_k)$ for notation conveinence.

Our goal is to show that

$$\sum_{k=1}^{n} w(I_k) \leq C\lambda^{-q} \|f\|_{L^q(w)}^q \tag{4.8}$$

for all n. Assuming (4.8) holds, we justify the algorithm successfully terminates. If the algorithm finishes after n steps, then all points (y, η, t) outside the union of tents T_k for $1 \leq k \leq n$ satisfy $|P(f)(y, \eta, t)| \leq \lambda$. Suppose the algorithm doesn't terminate after a finite number of steps. We observe the sequence of selected heights s_k must decay to 0 in this scenario. This is due to the selection process being independent of the chosen weight w meaning (4.8) would hold for the

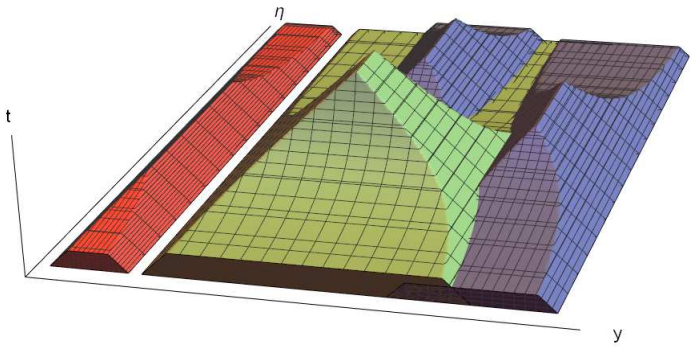

FIGURE 4. Example of tents from the L^∞ selection algorithm. Here, the green tent (with largest scale) is selected first, followed by the two blue tents and finally the red tent (with smallest scale).

Lebesgue case $w = 1$. Therefore, if we consider some (y, η, t) outside the union $\bigcup_{k\geq 1} T_k$ then $t > s_j$ for some j. As T_j is the tallest tent at the j-step with respect to containing a centralized point whose wavelet projection is greater than λ, we conclude $|P(f)(y, \eta, t)| \leq \lambda$. Thus, up to the proof of (4.8), all points (y, η, t) outside the union of the selected tents in the algorithm must satisfy $|P(f)(y, \eta, t)| \leq \lambda$.

It remains to prove weighted estimate (4.8). We first make the following observation regarding the well separation of the selected points (y_k, η_k, t_k).

Claim 3. *The collection of points $\{(y_k, \eta_k, t_k)\}_{k\geq 1}$ is well separated in the sense of (3.2) with separation constants $\alpha = 2^{-2}$ and $\beta = 2^{-6}b$; here, b is the parameter in $\Theta = (C_1, C_2, b)$.*

To verify the claim, consider (y_j, η_j, t_j), (y_k, η_k, t_k) such that $j < k$. By the selection algorithm, $(y_j, \eta_j, t_j) \in T_j$ was selected prior to $(y_k, \eta_k, t_k) \in T_k$. Suppose to the contrary that

$$|\eta_j - \eta_k| \leq 2^{-6} b \max(t_j^{-1}, t_k^{-1}) \quad \text{and} \quad |y_j - y_k| \leq 2^{-2} \max(t_j, t_k).$$

Central containment and $s_j \geq s_k$ implies $|y_j - y_k| \leq 2^{-4} s_j$ and $|\eta_j - \eta_k| \leq 2^{-3} b s_k^{-1}$. Thus,

$$|y_k - x_j| \leq |y_k - y_j| + |y_j - x_j| \leq 2^{-3} s_j \leq s_j - t_k$$

and

$$\eta_k - \xi_j = (\eta_k - \eta_j) + (\eta_j - \xi_j) \leq t_k^{-1} 2^{-4} b \leq t_k^{-1} C_2$$

with similar work gives $\eta_k - \xi_j \geq -t_k^{-1} C_1$. As $t_k < s_k \leq s_j$, this means $(y_k, \eta_k, t_k) \in T_j$ which contradicts the assumption the point was selected after T_j was removed. Thus, Claim 3 is true.

By the three grids trick (see Remark 3), each I_k is contained in a dyadic interval J_k of comparable length where J_k is either in the standard dyadic grid $\mathcal{D}_0$, the 1/3 shifted grid $\mathcal{D}_1$, or the 2/3 shifted grid $\mathcal{D}_2$. It suffices to prove

$$\sum_{k=1}^{n} w(J_k) \leq C \lambda^{-q} \|f\|^q_{L^q(w)} \tag{4.9}$$

where J_k is an element of $\mathcal{D}_0$, $\mathcal{D}_1$, or $\mathcal{D}_2$ such that $I_k \subset J_k$ and $3|I_k| \leq |J_k| \leq 6|I_k|$. Without loss of generality, we may assume all J_k belong to the same dyadic grid, which for convenience we assume to be the standard grid $\mathcal{D}_0$.

For each $m \geq 0$ let K_m be the set of k such that $2^m\lambda \leq |P(f)(y_k, \eta_k, t_k)| \leq 2^{m+1}\lambda$. It suffices to show that for any m

$$\sum_{k \in K_m} w(J_k) \leq C(2^m\lambda)^{-q}\|f\|^q_{L^q(w)}.$$

We show this for $m = 0$ as the general case can be obtained by simply letting $\lambda' = 2^m\lambda$ and repeating the argument. Fixing $m = 0$, assume without loss of generality that $K = K_0$ so all k are automatically inside K_0.

For convenience let

$$N = \sum_k 1_{J_k} \qquad \text{and} \qquad N_I = \sum_{k: J_k \subset I} 1_{J_k}$$

be the tent counting function and counting function restricted to interval I respectively. The following lemma is a localization inequality for the $L^1(w)$ norm of N_I.

Lemma 6. *Fix a dyadic interval I, $\alpha > 0$, and integer $n \geq 0$. Let $\chi_I(x)$ be as defined in (1.10). Then*

$$\lambda\|N_I\|^{1/q}_{L^1(w)} \lesssim_{n,\alpha} \|N_I\|^{\alpha}_{\infty}\|f\,\chi_I^n\|_{L^q(w)}.$$

Proof. Fix interval I, without loss of generality we may assume $N = N_I$. We first consider the simpler proof for $n = 0$. For convenience, denote $P(f)(y_k, \eta_k, t_k) = \langle f, \phi_k\rangle$ where

$$\phi_k(x) \equiv \phi_{y_k,\eta_k,t_k}(x) = \frac{1}{t_k}e^{-i\eta_k(y_k - x)}\overline{\phi\Big(\frac{y_k - x}{t_k}\Big)}.$$

As $\lambda \leq |P(f)(y_k, \eta_k, t_k)| \leq 2\lambda$, we obtain

$$\lambda^2 \sum_k w(J_k) \lesssim \sum_{k \in K_m} w(J_k)|P(f)(y_k, \eta_k, t_k)|^2 = \|S(f)\|^2_{L^2(w)}$$

where

$$S(f)(x) := \Big(\sum_k |\langle f, \phi_k\rangle|^2 1_{J_k}(x)\Big)^{1/2}$$

is the square function summed over the selected points.

We now implement a sharp maximal inequality argument. Note first that $supp\big(S(f)\big) \subset supp(N)$. Using Hölder's inequality and the sharp maximal inequality we have

$$\big\|S(f)\big\|_{L^2(w)} \leq w\big(supp(N)\big)^{\frac{1}{2}-\frac{1}{q}}\big\|S(f)\big\|_{L^q(w)} \lesssim \Big(\sum_k w(J_k)\Big)^{\frac{1}{2}-\frac{1}{q}}\big\|(S(f))^{\#}\big\|_{L^q(w)}$$

where $(S(f))^{\#}$ is the dyadic sharp maximal function, taken with respect to the grid that contains all intervals J_k. Consequently,

$$\lambda\Big[\sum_k w(J_k)\Big]^{1/q} \lesssim \big\|(S(f))^{\#}\big\|_{L^q(w)}. \tag{4.10}$$

We will now establish the pointwise estimate

$$(S(f))^{\#}(x) \lesssim_{\phi,s} M_2(f)(x) + \lambda^s[M(N)(x)]^{s/2}[M_2(f)(x)]^{1-s} \tag{4.11}$$

where M is the usual dyadic maximal function and $s \in (0, 1)$. For each dyadic interval J it suffices to show there is some constant $\alpha_J > 0$ such that

$$\left(\frac{1}{|J|}\int_J |S(f)(x) - \alpha_J|^2 dx\right)^{1/2} \lesssim \inf_{x \in J} M_2(f)(x) + \lambda^s \inf_{x \in J}[M(N)(x)]^{s/2} \inf_{x \in J}[M_2(f)(x)]^{1-s}$$

Setting c_J as the center of J, let $\alpha_J = S_{out}(f)(c_J)$ where

$$S_{out}(f)(x) := \Big(\sum_{k: J \subset J_k} |\langle f, \phi_k \rangle|^2 1_{J_k}(x) \Big)^{1/2}.$$

For convenience, let $S_{in}(f)(x) = (S(f)(x)^2 - S_{out}(f)(x)^2)^{1/2}$ and consider $f_J = f\,\chi_J^N$ for some large constant N. By the square function estimate (from Lebesgue theory) we have

$$\int_J |S(f)(x) - \alpha_J|^2 dx \le \int_J |S_{in}(f)(x)|^2 dx = \int_J |\widetilde{S}_{in}(f_J)(x)|^2 dx$$

where $\widetilde{S}_{in}$ uses the mollified wave functions $\chi_J^{-N}\phi_k$. Note the mollified wave function has the same frequency support as ϕ_k and has sufficient decay while localization over $(y_k - t_k, y_k + t_k)$ in the sense Lemma 1 is applicable for appropriate exponents. Recall from Claim 3 that the selected points (y_k, η_k, t_k) are well-separated with separation constants dependent solely on Θ. We can therefore apply Lemma 2 along with the observation $t_k \le |J_k|$ at each k to get

$$\begin{aligned}\Big(\int_J |S(f)(x) - \alpha_J|^2 dx \Big)^{1/2} &\lesssim \|f_J\|_2 + \Big[\sup_k \big| P(f)(y_k, \eta_k, t_k) \big| \Big(\sum_{k: J_k \subset J} |J_k| \Big)^{1/2} \Big]^s \|f_J\|_2^{1-s} \\ &\lesssim |J|^{1/2} \inf_{x \in J} M_2(f)(x) + \lambda^s |J|^{1/2} \inf_{x \in J} [M(N)(x)]^{s/2} \inf_{x \in J} [M_2(f)(x)]^{1-s}\end{aligned}$$

as desired.

Now, using (4.11) and the fact that $w \in \mathbb{A}_{q/2}$,

$$\begin{aligned}\big\| (S(f))^{\#} \big\|_{L^q(w)} &\lesssim \|f\|_{L^q(w)} + \lambda^s \|N\|_{L^{q/2}(w)}^{s/2} \|f\|_{L^q(w)}^{1-s} \\ &\lesssim \|f\|_{L^q(w)} + \lambda^s \|N\|_\infty^{(\frac{s}{q})(\frac{q}{2}-1)} \|N\|_{L^1(w)}^{s/q} \|f\|_{L^q(w)}^{1-s}.\end{aligned}$$

As $\|N\|_{L^1(w)} = \sum_k w(J_k)$, combine (4.10) with the above sharp maximal function bounds to get

$$\lambda \|N\|_{L^1(w)}^{1/q} \lesssim \|N\|_\infty^{\frac{s}{1-s}(\frac{1}{2}-\frac{1}{q})} \|f\|_{L^q(w)}.$$

By selecting $0 < s < 1$ suitably we obtain the desired conclusion for any $\alpha > 0$, but for $n = 0$. For $n > 0$, apply the argument above for the mollified wave functions $\widetilde{\phi}_k(x) = \chi_I(x)^{-n}\phi_k(x)$, and note that $\langle f, \phi_k \rangle = \langle f\,\chi_I^n, \widetilde{\phi}_k \rangle$ for all k. As previously stated, the mollified wave function $\widetilde{\phi}_k$ has the same frequency support as ϕ_k and sufficient decay while still localized over $(y_k - t_k, y_k + t_k)$ so the same analysis as before applies. □

With the $L^1(w)$ norm estimate on N_I, we deduce a similar estimate for $L^{r,\infty}(w)$ with $0 < r < 1$.

Corollary 1. *Fix dyadic interval I. For any $r \in (0,1)$ and $n > 0$ it holds that*

$$\|N_I\|_{L^{r,\infty}(w)} \lesssim (\lambda^{-1} \|f\,\chi_I^n\|_{L^q(w)})^{q/r}.$$

Proof. Arguing as before, without loss of generality we may assume that $N = N_I$ and $n = 0$. Given any $t > 0$, we have

$$w(N > t) = \sum_{k \ge 0} w(2^k t < N \le 2^{k+1} t).$$

For each k, let A_k be the collection of k such that $J_k \cap \{N \le 2^{k+1}t\} \neq \emptyset$ and denote N_k as the counting function N restricted to A_k. Then for every $x \in \{N \le 2^{k+1}t\}$ we have $N(x) = N_k(x)$, therefore

$$\{2^k t < N \le 2^{k+1} t\} \subset \{N_k > 2^k t\}.$$

Furthermore, we also have

$$\|N_k\|_\infty \leq 2^{k+1}t.$$

Indeed, N_k is locally constant, and any interval on which N_k is constant must be part of some J_k that in turn intersects $\{N \leq 2^{k+1}t\}$, and clearly $N_k \leq N$.

Thus, applying Lemma 6 with $\alpha > 0$ we obtain

$$\lambda \|N_k\|_{L^1(w)}^{1/q} \lesssim \|N_k\|_\infty^\alpha \|f\|_{L^q(w)} \lesssim 2^{(k+1)\alpha} t^\alpha \|f\|_{L^q(w)}$$

therefore

$$\lambda^q w(N_k > 2^{k+1}t) \leq 2^{-(k+1)} t^{-1} \lambda^q \|N_k\|_{L^1(w)} \lesssim 2^{-(k+1)(1-\alpha q)} t^{-(1-\alpha q)} \|f\|_{L^q(w)}^q$$

Summing over $k \geq 0$, it follows that

$$\lambda^q w(N > t) \lesssim t^{-(1-\alpha q)} \|f\|_{L^q(w)}^q$$

provided that $0 < \alpha < 1/q$. Letting $r = 1 - \alpha q$ we obtain

$$t w(N > t)^{1/r} \lesssim \lambda^{-q/r} \|f\|_{L^q(w)}^{q/r}.$$

To get $n > 0$, apply the argument above for mollified wave functions $\widetilde{\phi}_k = \chi_I^{-n} \phi_k$. □

Next, we establish a good λ inequality with respect to N. Below, let $M_{q,w} f$ be the weighted L^q-maximal function.

Lemma 7. *Consider $L > 0$ and $r \in (0,1)$. There is some $c > 0$ such that for any $t > 0$,*

$$w(N > t) \leq \frac{1}{L} w(N > t/4) + w(M_{q,w} f > c\lambda t^{r/q}).$$

Proof. Let $\mathbf{I}$ be the collection of all maximal dyadic intervals that are subsets of $\{N > t/4\}$. Suppose that $I \in \mathbf{I}$ and $I \cap \{M_{q,w} f \leq c\lambda t^{q/r}\} \neq \emptyset$. Then for every $x \in \{N > t\} \cap I$ we have $N(x) - N_I(x) \leq t/4$, so in particular $x \in \{N_I > t/4\}$. It follows that, using Corollary 1,

$$w(\{N > t\} \cap I) \leq w(N_I > t/4) \lesssim t^{-r} \lambda^{-q} \|f \chi_I^n\|_{L^q(w)}^q$$

$$\lesssim t^{-r} \lambda^{-q} w(I) \inf_{x \in I} M_{q,w}(f)(x)^q \lesssim t^{-r} \lambda^{-q} w(I) (c\lambda t^{r/q})^q \lesssim c^q w(I).$$

Consequently, by choosing $c > 0$ sufficiently small (independent of I) we obtain

$$w(\{N > t\} \cap I) \leq \frac{1}{L} w(I).$$

Summing over I we obtain the desired claim. □

We are now ready to show (4.9). Integrating over $t > 0$ in the good λ estimate provided by Lemma 7, it follows (from the standard argument) that

$$\|N\|_{L^1(w)} \lesssim \int_0^\infty w(M_{q,w}(f) > c\lambda t^{r/q}) dt = (c\lambda)^{-q/r} \int_0^\infty w(M_{q,w}(f) > s) s^{\frac{q}{r}-1} ds$$

$$\lesssim \lambda^{-q/r} \|M_{q,w}(f)\|_{L^{q/r}(w)}^{q/r} \lesssim \lambda^{-q/r} \|f\|_{L^{q/r}(w)}^{q/r}.$$

Using the fact that the $\mathbb{A}_q$ condition is an open condition, we actually have $w \in \mathbb{A}_{qr/2}$ for some $r \in (0,1)$. Thus, by replacing q with qr,

$$\|N\|_{L^1(w)} \lesssim \lambda^{-q} \|f\|_{L^q(w)}^q.$$

This completes the proof of (4.9) which in turn implies the desired estimate (4.8).

We finish by denoting Q_0 as the collection of (x_k,ξ_k,s_k) selected and

$$E_0=\bigcup_{(x,\xi,s)\in Q_0} T(x,\xi,s).$$

We have $|P(f)(y,\eta,t)|\le\lambda$ for all $(y,\eta,t)\in E_0^c$, while the weighted outer measure of E_0 is $O(\lambda^{-q}\|f\|_{L^q(w)})$. We free the notations (y_k,η_k,t_k), (x_k,ξ_k,s_k), T_k, I_k, $k\ge 1$.

4.4.2. *Treatment for the L^2 part of the size.* We now select tents over which the $L^2(w)$ portion of size $\mathcal{S}$ is large with respect to λ. We split a tent $T(x,\xi,s)$ into its upper half

$$T_+(x,\xi,s)=T(x,\xi,s)\cap\{(y,\eta,t)\in X\,:\,\eta\ge\xi\}$$

and its lower half $T_-(x,\xi,s)=T(x,\xi,s)\backslash T_+(x,\xi,s)$. We define $T_\pm^b$ and $T_\pm^\ell$ similarly.

This section focuses on finding a countable collection of points $Q_+\subset X_\Delta$ such that

$$\sum_{(x,\xi,s)\in Q_+} w(x-s,x+s)\le C\lambda^{-q}\|f\|_{L^q(w)}^q$$

and for every tent $T(x,\xi,s)\in\mathbb{E}_\Delta$,

$$\frac{1}{w(x-s,x+s)}\int_{T_+^\ell(x,\xi,s)\backslash E_+}|P(f)(y,\eta,t)|^2w(y-t,y+t)dyd\eta\frac{dt}{t}\le\lambda^2$$

where

$$E_+=E_0\cup\bigcup_{(x.\xi,s)\in Q_+}T(x,\xi,s).$$

We remark that the work to generate Q_+ and E_+ can be applied symmetrically to find a countable collection of points $Q_-\subset X_\Delta$ such that

$$\sum_{(x,\xi,s)\in Q_-} w(x-s,x+s)\le C\lambda^{-q}\|f\|_{L^q(w)}^q$$

and if

$$E_-=E_0\cup\bigcup_{(x,\xi,s)\in Q_-}T(x,\xi,s),$$

then

$$\frac{1}{w(x-s,x+s)}\int_{T_-^\ell(x,\xi,s)}|P(f)(y,\eta,t)1_{X\backslash E_-}(y,\eta,t)|^2w(y-t,y+t)dyd\eta\frac{dt}{t}\le\lambda^2$$

for all tents $T(x,\xi,s)\in\mathbb{E}_\Delta$. By setting $Q=Q_0\cup Q_+\cup Q_-$ and $E:=\bigcup_{(x,\xi,s)\in Q}T(x,\xi,s)$, the proof of the q-endpoint estimate (4.3) will then be completed.

L^2 Selection Algorithm. Let C be larger than the doubling constant of w. Suppose there exists $(x,\xi,s)\in X_\Delta$ such that

$$\frac{1}{w(x-s,x+s)}\int_{T_+^\ell(x,\xi,s)\backslash E_0}|P(f)(y,\eta,t)|^2w(y-t,y+t)dyd\eta\frac{dt}{t}\ge C^{-1}\lambda^2. \tag{4.12}$$

Applying (4.5), it follows that $w(x-s,x+s)$ is bounded from above a priori. By substituting $f=f\chi^N_{(x-s,x+s)}\chi^{-N}_{(x-s,x+s)}$ in (4.12) where $N>q$, similar work in conjunction with Remark 2 shows s itself is bounded a priori. Indeed, as $w\notin L^1$ is a doubling weight, it forces $|x|/s$ to be sufficiently large whenever s itself is large. Consequently, $\|f\chi^N_{(x-s,x+s)}\|_\infty$ is sufficiently small. As

$$\|\chi^N_{(x-s,x+s)}\|_{L^1(w)}\lesssim w(x-s,x+s),$$

we conclude the value of s for points (x,ξ,s) satisfying (4.12) must be bounded.

Let $\widetilde{s}$ be the least upper bound on s. As such, ξ is a discrete parameter since it is a multiple of $2^{-8}b(\widetilde{s})^{-1}$. Since $\widehat{f}$ is compactly supported, there is an upper bound on ξ meaning there is a maximal possible value $\xi = \xi_{\max}$ to consider in (4.5). Select $(x_1, \xi_1, s_1) \in X_\Delta$ satisfying (4.12) such that $\xi_1 = \xi_{\max}$ and s_1 is maximal with respect to the restriction $\xi_1 = \xi_{\max}$. Let $T_1 = T(x_1, \xi_1, s_1)$ and $I_1 = (x_1 - s_1, x_1 + s_1)$ for convenience of notation. By the maximality of s_1 and the doubling property of w, note

$$\frac{1}{w(I_1)} \int_{T_+^\ell(x_1,\xi_1,s_1)\setminus E_0} |P(f)(y,\eta,t)|^2 w(y-t, y+t) dy d\eta \frac{dt}{t} \leq \lambda^2.$$

We now iterate the argument. Assume that we have selected $(x_k, \xi_k, s_k) \in X_\Delta$ for $1 \leq k \leq n-1$ and set $E_n = E_0 \cup \bigcup_{k=1}^{n-1} T_k$. Suppose there is some $(x, \xi, s) \in X_\Delta$ such that

$$\frac{1}{w(x-s, x+s)} \int_{T_+^\ell(x,\xi,s)\setminus E_n} |P(f)(y,\eta,t)|^2 w(y-t, y+t) dy d\eta \frac{dt}{t} \geq C^{-1}\lambda^2$$

We now select such a point (x_n, ξ_n, s_n) such that ξ_n is a (possibly new) maximal $\xi_{\max}$ and s_n is maximized with respect to $\xi_{\max}$. Denote $T_n = T(x_n, \xi_n, s_n)$ and $I_n = (x_n - s_n, x_n + s_n)$. Again, by the maximality of s_n,

$$\frac{1}{w(I_n)} \int_{T_+^\ell(x_n,\xi_n,s_n)\setminus E_n} |P(f)(y,\eta,t)|^2 w(y-t, y+t) dy d\eta \frac{dt}{t} \leq \lambda^2.$$

Our goal is to show that

$$\sum_{k=1}^{n} w(I_k) \leq C\lambda^{-q} \|f\|_{L^q(w)}^q. \tag{4.13}$$

where C is independent of n. Assuming (4.13) is valid, we now justify the termination of the L^2 selection algorithm. If the algorithm terminates after selecting n tents, then

$$\frac{1}{w(x-s, x+s)} \int_{T_+^\ell(x,\xi,s)\setminus E_{n+1}} |P(f)(y,\eta,t)|^2 w(y-t, y+t) dy d\eta \frac{dt}{t} \leq C^{-1}\lambda^2.$$

holds for all $(x, \xi, s) \in \mathbb{E}_\Delta$ and we set Q_+ as the collection of (x_k, ξ_k, s_k) for $1 \leq k \leq n$. In the scenario the algorithm does not terminate after a finite number of steps, set $E_{(1)} = E_0 \cup \bigcup_{k\geq 1} T_k$ and $Q_{+,(1)}$ as the collection of selected (x_k, ξ_k, s_k) for $k \geq 1$. Note the selected ξ_k form a non-increasing sequence of elements in the lattice $\mathbb{Z}2^{-8}b(\widetilde{s})^{-1}$. In the case ξ_k tends to negative infinity, suppose there is some $(x, \xi, s) \in X_\Delta$ satisfying

$$\frac{1}{w(x-s, x+s)} \int_{T_+^\ell(x,\xi,s)\setminus E_{(1)}} |P(f)(y,\eta,t)|^2 w(y-t, y+t) dy d\eta \frac{dt}{t} \geq C^{-1}\lambda^2. \tag{4.14}$$

As $\xi_k \to -\infty$, then $\xi_j < \xi$ for some j which contradicts the selection of T_j. Thus, the converse inequality to (4.14) must hold for all $(x, \xi, s) \in X_\Delta$ and we set $Q_+ = Q_{+,(1)}$.

Now suppose ξ_k does not tend to negative infinity and instead stabilizes at some finite $\xi_{(1)}$. We restart the algorithm and redefine the selected tents T_k by $T_{(1),k} = T(x_{(1),k}, \xi_{(1),k}, s_{(1),k})$ and intervals I_k by $I_{(1),k}$. Observe in this scenario that the tail of the sequence $s_{(1),k}$ decays to 0. Indeed, if $s_{(1),k}$ converges to some nonzero $s_{(1)}$ then the tail of $Q_{+,(1)}$ is of the form $(x_{(1),k}, \xi_{(1)}, s_{(1)})$ where $x_{(1),k}$ is an element in the lattice $\mathbb{Z}2^{-4}s_{(1)}$. The corresponding intervals $I_{(1),k}$ therefore eventually slide towards infinity or negative infinity. By recycling the argument for the a priori bound on s, such a sequence cannot occur.

Consider $(x,\xi,s)\in X_\Delta$ where (4.14) holds. As before, there is a maximal frequency $\xi_{\max}$ to consider for such a point. Given the previous selection of tents, we know $\xi_{\max}\le\xi_{(1)}$. If $\xi_{\max}=\xi_{(1)}$ and $(x,\xi_{\max},s)$ satisfies (4.14) then $s>s_{(1),k}$ for some k by the previous paragraph which contradicts the selection of $T_{(1),k}$. Therefore, it follows that $\xi_{\max}<\xi_{(1)}$. Choose a point $(x_{(2),1}\xi_{(2),1},s_{(2),1})$ satisfying (4.14) such that $s_{(2),1}$ is maximized under the condition $\xi_{(2),1}=\xi_{\max}$. Iterate the selection algorithm as before to obtain a sequence of tents and intervals

$$T_{(2),k}=T(x_{(2),k},\xi_{(2),k},s_{(2),k}),\qquad\text{and}\qquad I_{(2),k}=(x_{(2),k}-s_{(2),k},x_{(2),k}+s_{(2),k}).$$

The proof of (4.13), to be shown, will naturally extend here to give

$$\sum_{k=1}^{\infty}w(I_{(1),k})+\sum_{k\ge1}w(I_{(2),k})\le C\lambda^{-q}\|f\|^q_{L^q(w)}.$$

Continue the process as shown above. If we eventually have a sequence of frequencies $\xi_{(m),k}$ (m is fixed) which either terminates after finitely many k or $\xi_{(m),k}\to-\infty$ then there are no more $(x,\xi,s)\in X_\Delta$ for which the (now updated) version of (4.14) holds. At worst, we eventually obtain a double sequence of tents $T_{(m),k}$ where the double summation of $w(I_{(m),k})$ is $O(\lambda^{-q}\|f\|^q_{L^q(w)})$. In addition, the sequence of stabilizing points $\xi_{(k)}$ in this case is strictly decreasing in a discrete lattice and tending to negative infinity. We finish by setting Q_+ as the collection of $(x_{(j),k},\xi_{(j),k},s_{(j),k})$ where $j,k\in\mathbb{N}$ and E_+ as the union of tents $T(x_{(j),k},\xi_{(j),k},s_{(j),k})$. We can therefore conclude that there are no more points $(x,\xi,s)\in X_\Delta$ such that

$$\frac{1}{w(x-s,x+s)}\int_{T^\ell_+(x,\xi,s)\setminus E_+}|P(f)(y,\eta,t)|^2w(y-t,y+t)\,dy\,d\eta\frac{dt}{t}\ge C^{-1}\lambda^2$$

and so the L^2 algorithm terminates.

It remains to prove the weighted estimate (4.13). Note the proof follows similar steps as in the L^∞ treatment in Section 4.4.1. For convenience of notation let

$$T^*_k=T^\ell_+(x_k,\xi_k,s_k)\setminus E_k\qquad\text{for } k\ge1$$

with $E_1=E_0$. We first note the well separation of the partial tents T^*_k.

Claim 4. *The collection of partial tents $\{T^*_k\}$ is well-separated in the sense of (3.11) with separation constants $\alpha=1$, $\beta=2^{-6}b$, and $B=2^8\max(C_1,C_2)b^{-1}$ in terms of $\Theta=(C_1,C_2,b)$.*

To verify the claim, consider $(y,\eta,t)\in T^*_j$ and $(y',\eta',t')\in T^*_k$ such that $Bt'<t$. Assuming $|\eta-\eta'|\le2^{-6}bt'^{-1}$, it follows from (y',η',t') being in the upper lacunary part of T_k that $\xi_j>\xi_k$.

$$\xi_j-\xi_k=(\eta-\eta')-(\eta-\xi_j)+(\eta'-\xi_k)$$

$$\ge-2^{-6}bt'^{-1}-C_2t^{-1}+bt'^{-1}\ge(1-2^{-6}-2^{-8})bt'^{-1}>0$$

This means tent T_j was selected prior to T_k so $(y',\eta',t')\notin T_j$ by the selection process. Furthermore, observe that $t'<t<s_j$ and

$$\eta'-\xi_j=(\eta'-\eta)+(\eta-\xi_j)\le\left(2^{-7}b+C_2B^{-1}\right)(t')^{-1}\le C_2(t')^{-1}$$

with similar work showing $\eta'-\xi_j\ge-C_1(t')^{-1}$. As $(y',\eta',t')\notin T_j$, we then require

$$|y'-x_j|>s_j-t'>s_j-t$$

which verifies Claim 4.

As in the L^∞ selection argument, application of the three grids trick (see Remark 3) means it suffices to show

$$\sum_k w(J_k) \leq C\lambda^{-q}\|f\|^q_{L^q(w)} \tag{4.15}$$

where all J_k are dyadic intervals in the standard grid $\mathcal{D}_0$, 1/3 shifted grid $\mathcal{D}_1$, or 2/3 shifted grid $\mathcal{D}_2$ such that $I_k \subset J_k$ and $3|I_k| \leq |J_k| \leq 6|I_k|$. We may assume without loss of generality that all J_k belong to the standard dyadic grid.

As before, let $N(x)$ be the counting function over the selected intervals J_k and $N_I(x)$ be the counting function N restricted to J_k contained in interval I. We need the following analogue of Lemma 6, and the rest of the proof for the L^2 portion (with respect to upper half of tents) is exactly the same as the L^∞ portion in Section 4.4.1.

Lemma 8. *Fix a dyadic interval I, $\alpha > 0$, and integer $n \geq 0$. Let $\chi_I(x)$ be as defined in (1.10). Then*

$$\lambda\|N_I\|^{1/q}_{L^1(w)} \lesssim_{n,\alpha} \|N_I\|^\alpha_\infty \|f\,\chi_I^n\|_{L^q(w)}.$$

Proof. As before, we may assume without loss of generality that $N = N_I$ (freeing up the notation of interval I) and set $n = 0$.

Let S be the following square function

$$S(F)(u) := \left(\sum_{k:J_k}\int_{T_k^*} |F(y,\eta,t)|^2 1_{|y-u|<t}\,dy\,d\eta\frac{dt}{t}\right)^{1/2}, \quad u \in \mathbb{R}.$$

Appealing to the doubling property of w and the selection criteria for the tents T_k,

$$\lambda^2\sum_k w(J_k) \lesssim \sum_k \int_{T_k^*} |P(f)(y,\eta,t)|^2 w(y-t,y+t)\,dy\,d\eta\frac{dt}{t} = \|S(P(f))\|^2_{L^2(w)}.$$

We note that if $(y,\eta,t) \in T(x_k,\xi_k,s_k)$ and $|y-u| < t$ then $u \in (y-t,y+t) \subset (x_k - s_k, x_k + s_k)$. Therefore $supp(SF) \subset \bigcup_{k:J_k} J_k$. Consequently, by an application of Hölder's inequality,

$$\lambda\|N\|^{1/q}_{L^1(w)} \lesssim \|S(P(f))\|_{L^q(w)}.$$

It remains to control the $L^q(w)$ norm of this square function. Our main idea here is to cover $(y-t,y+t)$ using a (shifted) dyadic interval I comparable to $(y-t,y+t)$ via the three grids trick and pass to (shifted) square functions over these grids. More precisely, the interval I will belong to either the classical grid $\mathcal{D}_0$ or one of the shifted grids $\mathcal{D}_1$, $\mathcal{D}_2$ mentioned prior. We may bound

$$S(F)(u) \leq \sum_{j=0}^2 S_j(F)(u)$$

where $S_j(F)$ is a square function over grid $\mathcal{D}_j$. Namely, for each grid $\mathcal{D}_j$ we may define for some $C > 1$ (sufficiently large absolute constant)

$$S_j(F)(u) = \left(\sum_{k:J_k}\int_{T_k^*} |F(y,\eta,t)|^2 \sum_{I\in\mathcal{D}_j : t/C \leq |I| \leq Ct} 1_I(y)1_I(u)\,dy\,d\eta\frac{dt}{t}\right)^{1/2}.$$

Fix j. Standard estimates show

$$\|S_j(P(f))\|_{L^q(w)} \lesssim \|\big(S_j(P(f))\big)^\sharp\|_{L^q(w)}$$

where $(\cdot)^\sharp$ is the dyadic sharp maximal function, with intervals from the grid $\mathcal{D}_j$. Now, for each dyadic $J \in \mathcal{D}_j$, we then let

$$\alpha(u) = \Big(\sum_{k:I_k} \int_{T_k^*} |P(f)(y,\eta,t)|^2 \sum_{I\in D_j : C^{-1}t \le |I| \le Ct\,, J\subset I} 1_I(y) 1_I(u)\, dy d\eta \frac{dt}{t} \Big)^{1/2}.$$

Note that $\alpha(u)$ is constant over $u \in J$, which we now refer to as α_J (despite its dependence on j). For each $u \in J$,

$$\begin{aligned} |S_j\big(P(f)\big)(u) - \alpha_J| &\le \Big(\sum_{k:J_k} \int_{T_k^*} |P(f)(y,\eta,t)|^2 \sum_{I\in D_j : C^{-1}t \le |I| \le Ct\,, I\subset J} 1_I(y) 1_I(u)\, dy d\eta \frac{dt}{t} \Big)^{1/2} \\ &\lesssim \Big(\sum_{k:I_k} \int_{T_k^*} |P(f)(y,\eta,t)|^2 1_{|y-u|=O(t)} 1_{t=O(|J|)}\, dy d\eta \frac{dt}{t} \Big)^{1/2} \end{aligned}$$

thus using Lebesgue theory we have

$$\frac{1}{|J|}\int_J |S_j\big(P(f)\big)(u) - \alpha_J| \lesssim \Big(\frac{1}{|J|} \sum_{k:J_k} \int_{T_k^*} |P(f)(y,\eta,t)|^2 1_{(y-t,y+t)\subset CJ}\, dy d\eta dt \Big)^{1/2}$$

where $C > 0$ is sufficiently large. Recall $(y-t, y+t) \subset CJ$ is equivalent to $0 < t \le C|J|/2$ and $|y - c_J| < C|J|/2 - t$. By Remark 1, $T_k^* \cap \{(y-t, y+t) \subset CJ\}$ is a subset of tent T_k' with top interval $I_k' = I_k \cap CJ$. Note the collection of partial tents $T_k^* \cap \{(y-t,y+t)\subset CJ\}$ in T_k' is still well separated with same separation constants as Claim 4.

Let $f_J = f\chi_{CJ}^N$ for some large constant N. Using Lemma 3 with $|P(f)(y,\eta,t)| = |\langle f_J, \chi_{CJ}^{-N}\phi_{y,\eta,t}\rangle|$ and $|P(f)(y,\eta,t)| \le \lambda$ for all points in $T_k^* \cap \{(y-t,y+t)\subset CJ\}$, we have

$$\begin{aligned} \int_J \big|S_j(P(f))(u) - \alpha_J\big|\, du &\lesssim \|f_J\|_2 + \big[\lambda \|N\|_{L^1(CJ)}^{1/2}\big]^s \|f_J\|_2^{1-s} \\ &\lesssim |J|^{1/2} \inf_{x\in J} M_2(f)(x) + |J|^{1/2}\lambda^s \inf_{x\in J}\big[M(N)(x)\big]^{s/2} \inf_{x\in J}\big[M_2(f)(x)\big]^{1-s} \end{aligned}$$

and so

$$\big(S_j(P(f))\big)^\sharp(x) \lesssim M_2(f)(x) + \lambda^s \big[M(N)(x)\big]^{s/2}\big[M_2(f)(x)\big]^{1-s}. \tag{4.16}$$

Combining (4.16) and the fact $w \in \mathbb{A}_{q/2}$,

$$\begin{aligned} \|S_j(P(f))\|_{L^q(w)} &\lesssim \|f\|_{L^q(w)} + \lambda^s \|N\|_{L^{q/2}(w)}^{s/2} \|f\|_{L^q(w)}^{1-s} \\ &\lesssim \|f\|_{L^q(w)} + \lambda^s \|N\|_\infty^{(\frac{s}{q})(\frac{q}{2}-1)} \|N\|_{L^1(w)}^{s/q} \|f\|_{L^q(w)}^{1-s} \end{aligned}$$

Summing over $j = 0, 1, 2$ we obtain

$$\lambda \|N\|_{L^1(w)}^{1/q} \lesssim \|f\|_{L^q(w)} + \lambda^s \|N\|_\infty^{(\frac{s}{q})(\frac{q}{2}-1)} \|N\|_{L^1(w)}^{s/q} \|f\|_{L^q(w)}^{1-s}$$

and we get by rearrangement

$$\lambda \|N\|_{L^1(w)}^{1/q} \lesssim \|N\|_\infty^{\frac{s}{1-s}(\frac{1}{2}-\frac{1}{q})} \|f\|_{L^q(w)}.$$

We obtain the desired result for the $n = 0$ case by choosing $s \in (0,1)$ sufficiently small. For the case $n \ge 1$, consider the mollified wave packet $\widetilde{\phi}_{y,\eta,t} = \chi_I^{-n}\phi_{y,\eta,t}$ and apply work above to $P(f)(y,\eta,t) = \langle f\chi_I^n, \widetilde{\phi}_{y,\eta,t}\rangle$. □

REFERENCES

[1] C. Benea and C. Muscalu. Sparse domination via the helicoidal method. arXiv preprint arXiv:1707.05484, 2017.
[2] C. Benea and C. Muscalu. The helicoidal method. In *Operator theory: themes and variations*, volume 20 of *Theta Ser. Adv. Math.*, pages 45–96. Theta, Bucharest, 2018.
[3] L. Carleson. On convergence and growth of partial sums of Fourier series. *Acta Math.*, 116:135–157, 1966.
[4] M. Christ. Weak type $(1,1)$ bounds for rough operators. *Ann. of Math. (2)*, 128(1):19–42, 1988.
[5] D. Cruz-Uribe and J. M. Martell. Limited range multilinear extrapolation with applications to the bilinear Hilbert transform, 2018.
[6] A. Culiuc, F. Di Plinio, and Y. Ou. Domination of multilinear singular integrals by positive sparse forms. *J. Lond. Math. Soc. (2)*, 98(2):369–392, 2018.
[7] F. Di Plinio, Y. Q. Do, and G. N. Uraltsev. Positive sparse domination of variational Carleson operators. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 18(4):1443–1458, 2018.
[8] F. Di Plinio and Y. Ou. A modulation invariant Carleson embedding theorem outside local L^2. *J. Anal. Math.*, 135(2):675–711, 2018.
[9] Y. Do and M. Lacey. Weighted bounds for variational Fourier series. *Studia Math.*, 211(2):153–190, 2012.
[10] Y. Do and M. Lacey. Weighted bounds for variational Walsh-Fourier series. *J. Fourier Anal. Appl.*, 18(6):1318–1339, 2012.
[11] Y. Do, R. Oberlin, and E. A. Palsson. Variational bounds for a dyadic model of the bilinear Hilbert transform. *Illinois J. Math.*, 57(1):105–119, 2013.
[12] Y. Do and C. Thiele. L^p theory for outer measures and two themes of Lennart Carleson united. *Bull. Amer. Math. Soc. (N.S.)*, 52(2):249–296, 2015.
[13] C. Fefferman. Pointwise convergence of Fourier series. *Ann. of Math. (2)*, 98:551–571, 1973.
[14] R. A. Hunt. On the convergence of Fourier series. In *Orthogonal Expansions and their Continuous Analogues (Proc. Conf., Edwardsville, Ill., 1967)*, pages 235–255. Southern Illinois Univ. Press, Carbondale, Ill., 1968.
[15] M. Lacey and C. Thiele. L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$. *Ann. of Math. (2)*, 146(3):693–724, 1997.
[16] M. Lacey and C. Thiele. On Calderón's conjecture. *Ann. of Math. (2)*, 149(2):475–496, 1999.
[17] M. Lacey and C. Thiele. A proof of boundedness of the Carleson operator. *Math. Res. Lett.*, 7(4):361–370, 2000.
[18] A. K. Lerner. Sharp weighted norm inequalities for Littlewood-Paley operators and singular integrals. *Adv. Math.*, 226(5):3912–3926, 2011.
[19] A. K. Lerner. On sharp aperture-weighted estimates for square functions. *J. Fourier Anal. Appl.*, 20(4):784–800, 2014.
[20] X. Li. personal communication.
[21] C. Muscalu, T. Tao, and C. Thiele. L^p estimates for the biest. I. The Walsh case. *Math. Ann.*, 329(3):401–426, 2004.
[22] C. Muscalu, T. Tao, and C. Thiele. L^p estimates for the biest. II. The Fourier case. *Math. Ann.*, 329(3):427–461, 2004.
[23] R. Oberlin, A. Seeger, T. Tao, C. Thiele, and J. Wright. A variation norm Carleson theorem. *J. Eur. Math. Soc. (JEMS)*, 14(2):421–464, 2012.
[24] C. Thiele. *Wave packet analysis*, volume 105 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006.
[25] C. Thiele, S. Treil, and A. Volberg. Weighted martingale multipliers in the non-homogeneous setting and outer measure spaces. *Adv. Math.*, 285:1155–1188, 2015.
[26] G. Uraltsev. Variational Carleson embeddings into the upper 3-space. preprint ArXiv:1610.07657, 2016.

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF VIRGINIA, CHARLOTTESVILLE, VA 22904-4137
Email address: yendo@virginia.edu
Email address: mel3jp@virginia.edu