Ark. Mat., **46** (2008), 315–336
DOI: 10.1007/s11512-008-0080-2
© 2008 by Institut Mittag-Leffler. All rights reserved

A Wiener–Wintner theorem for the Hilbert transform

Michael Lacey and Erin Terwilleger

Abstract. We prove the following extension of the Wiener–Wintner theorem and the Carleson theorem on pointwise convergence of Fourier series: For all measure-preserving flows (X, μ, T_t) and $f \in L^p(X, \mu)$, there is a set $X_f \subset X$ of probability one, so that for all $x \in X_f$,

$$\lim_{s\downarrow 0} \int_{s<|t|<1/s} e^{i\theta t} f(\mathrm{T}_t\, x)\, \frac{dt}{t} \quad \text{exists for all } \theta.$$

The proof is by way of establishing an appropriate oscillation inequality which is itself an extension of Carleson's theorem.

1. The main theorem

We are concerned with quantitative inequalities related to the pointwise convergence of singular integrals that are uniform with respect to modulation. To state our results, define *dilation* and *modulation operators* by

$$\operatorname{Dil}_s^{(p)} f(x) \stackrel{\text{def}}{=} s^{-1/p} f\left(\frac{x}{s}\right), \qquad 0<s,p<\infty, \tag{1.1}$$

$$\operatorname{Dil}_s^{(\infty)} f(x) \stackrel{\text{def}}{=} f\left(\frac{x}{s}\right), \qquad 0<s<\infty,$$

$$\operatorname{Mod}_\xi f(x) \stackrel{\text{def}}{=} e^{ix\xi} f(x), \qquad \xi\in\mathbb{R}. \tag{1.2}$$

Let K be a distribution. The most important example will be $K_H(y) \stackrel{\text{def}}{=} \zeta(y)/y$, where ζ is a smooth, symmetric, compactly supported function. This is a distribution associated with a truncation of the Hilbert transform kernel.

Our principal concern is the convergence of terms $(\operatorname{Dil}_s^{(1)} K)*f(x)$ in a pointwise sense, and in one that is, in addition, uniform over all modulations. To do

The first author is a Guggenheim fellow, and his research was supported in part by the National Science Foundation.

this, we use the following definition.

$$\operatorname{Osc}_n(K;f)^2 \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \sup_{k_j \le l < l' < k_{j+1}} |[\operatorname{Dil}^{(1)}_{2^{l/n}} K - \operatorname{Dil}^{(1)}_{2^{l'/n}} K] * f|^2. \tag{1.3}$$

This definition depends upon a choice of an increasing sequence of integers $k_j \in \mathbb{Z}$, a dependence that we suppress as relevant constants are independent of the choice of $\{k_j\}_{j=1}^{\infty}$. It also depends upon a choice of positive integer n, which we have incorporated into the notation. This only permits dilations of the form $2^{l/n}$ for integers l.

Theorem 1.4. *Fix a smooth, symmetric, compactly supported function ζ. For integers $n>0$ and $1<p<\infty$ there is a constant $C_{n,p,\zeta}$ so that we have the inequality*

$$\Big\| \sup_N \operatorname{Osc}_n(K_H; \operatorname{Mod}_N f) \Big\|_p \le C_{n,p,\zeta} \|f\|_p. \tag{1.5}$$

The inequality holds for all choices of increasing sequences $\{k_j\}_{j=1}^{\infty}$ satisfying $k_{j+1} \ge k_j + n$.

Our primary interest in this theorem is the corollary below, which is a Hilbert transform counterpart to the well known Wiener–Wintner theorem for ergodic averages. Deriving the corollary below is a standard part of the literature, with the roots of the argument going back to Calderón [6]. The use of an oscillation inequality to establish convergence was introduced by Bourgain [5]. See also the papers of Campbell–Jones–Reinhold–Wierdl [7], and Jones–Kaufmann–Rosenblatt–Wierdl [13]. A proof of the corollary below is in the next section.

Corollary 1.6. *For all measure-preserving flows $\{\mathrm{T}_t : t \in \mathbb{R}\}$ on a probability space (X,μ) and functions $f \in L^p(\mu)$, there is a set $X_f \subset X$ of probability one, so that for all $x \in X_f$ we have that*

$$\lim_{s \downarrow 0} \int_{s<|t|<1/s} e^{i\theta t} f(\mathrm{T}_t\, x)\, \frac{dt}{t} \quad \textit{exists for all } \theta.$$

This is a common extension of two classical theorems: Carleson's theorem [9] on Fourier series with Hunt's extension [12], and the Wiener–Wintner theorem [23] on ergodic averages.

Carleson's theorem. *We have the inequality*

$$\Big\| \sup_N \Big| \int_{-\infty}^{\infty} \operatorname{Mod}_N f(x-y)\, \frac{dy}{y} \Big| \Big\|_p \lesssim \|f\|_p, \quad 1<p<\infty.$$

Wiener–Wintner theorem. *For all measure-preserving flows* $\{\mathrm{T}_t : t\in\mathbb{R}\}$ *on a probability space* (X,μ) *and functions* $f\in L^p(X,\mu)$*, there is a set* $X_f\subset X$ *of probability one, so that for all* $x\in X_f$ *we have that*

$$\lim_{s\to\infty}\frac{1}{s}\int_{-s}^{s} e^{i\theta t} f(\mathrm{T}_t\, x)\,dt \quad \textit{exists for all } \theta,$$

$$\lim_{s\to 0}\frac{1}{s}\int_{-s}^{s} e^{i\theta t} f(\mathrm{T}_t\, x)\,dt \quad \textit{exists for all } \theta.$$

The Wiener–Wintner theorem can been seen as an extension of the Birkhoff ergodic theorem. The Carleson theorem is a deep result from the 1960s, and since then several proofs have been offered. An extensive survey and bibliography on this subject can be found in [14].

The possibility of extending the Wiener–Wintner theorem to the setting of the Hilbert transform was first raised in the paper of Campbell and Petersen [8]. The specific result proved there was essentially Carleson's theorem on the integers, with a transference to measure-preserving systems. Part of this was contained in a prior work of Mátë [17], a work that was overlooked until much later.

Assani [1] and [2] proved our Corollary 1.6 on a class of dynamical systems he termed *Wiener–Wintner systems*. The definition of these systems depends upon particular properties not enjoyed by all dynamical systems.

Our tool to prove convergence in the Hilbert transform setting is the oscillation inequality (1.5), an idea first employed in ergodic theory in the pioneering work of Bourgain on the ergodic theorem along arithmetic sequences [5]. The use of oscillation has subsequently been systematically studied in e.g. [7] and [13] and in references therein.

The main goal of this paper is a proof of Theorem 1.4. Clearly, we follow the lines of a proof of Carleson's theorem. In particular we employ the Lacey–Thiele approach [16] and refine one part of it to deduce our main theorem. We will also appeal to the 'restricted weak type argument' of C. Muscalu, T. Tao and C. Thiele [18], and L. Grafakos, T. Tao and E. Terwilleger [11].

Acknowledgements. The authors have benefited from conversations with Jim Campbell, Anthony Quas and Mate Wierdl. Part of this research was completed at the Schrödinger Institute, Vienna, Austria. For one of us (ML), discussions with Karl Petersen about this question formed our introduction to Carleson's theorem, for which we have been indebted to him ever since. Finally, we thank the referee for a detailed and helpful report.

2. Deduction of Corollary 1.6

This is a known argument. Let us begin with the following simple observation. Suppose that $\{a_n\}_{n=1}^{\infty}$ is a numerical sequence, so that for all increasing sequences of integers k_j,

$$\sum_{j=1}^{\infty} \sup_{k_j<n<n'\leq k_{j+1}} |a_n - a_{n'}|^2 < \infty. \tag{2.1}$$

It follows that $\lim_{n\to\infty} a_n$ exists. Indeed, to find a contradiction assume that $\lim_{n\to\infty} a_n$ does not exist. Then, there is an $\varepsilon>0$ so that for all integers K, $\limsup_{n'>n>K} |a_n-a_{n'}|>\varepsilon$. And therefore, we can build a sequence of integers $\{k_j\}_{j=1}^{\infty}$ to contradict (2.1).

We will use an extension of this. Suppose that $\{a_n(\theta):n\in\mathbb{N}$ and $\theta\in\mathbb{R}\}$ is a collection of numerical sequences indexed by $\theta\in\mathbb{R}$. Suppose that for all increasing sequences of integers k_j,

$$\sup_{\theta} \sum_{j=1}^{\infty} \sup_{k_j<n<n'\leq k_{j+1}} |a_n - a_{n'}|^2 < \infty. \tag{2.2}$$

It follows that $\lim_{n\to\infty} a_n(\theta)$ exists for all $\theta\in\mathbb{R}$.

Let us begin with the analytical inequality of Theorem 1.4. Using the observation of Calderón [6], one sees that the oscillation inequality continues to hold on the measure-preserving flow (X,μ,T_t).

Fix an integer $n\in\mathbb{N}$, and a smooth, symmetric, compactly supported function ζ. Fix an $f\in L^p(X)$, $1<p<\infty$. We see that there is a set $X_{f,\zeta,n}\subset X$ of probability one, so that for all $x\in X_f$ we have that

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} e^{i\theta t} f(\mathrm{T}_t\, x)\zeta(2^{-k/n}t)\,\frac{dt}{t} \quad \text{exists for all } \theta. \tag{2.3}$$

Now, the countable union of null sets is null, thus we can take a single set $X_{f,\zeta}$ of probability one so that (2.3) holds for all $n\in\mathbb{N}$.

We need to replace $\zeta(t)t^{-1}$ by $\mathbf{1}_{|t|\leq 1}t^{-1}$. To do so, consider a sequence of smooth, symmetric, compactly supported functions ζ_k satisfying

$$\mathbf{1}_{[-1+2^{-k},1-2^{-k}]}(t) \leq \zeta_k(t) \leq \mathbf{1}_{[-1,1]}.$$

Thus, these functions converge pointwise to $\zeta_\infty \stackrel{\text{def}}{=} \mathbf{1}_{[-1,1]}$. For each k, there is a single null set X_{f,ζ_k} of probability one so that (2.3) holds for all $n\in\mathbb{N}$ and all $\zeta=\zeta_k$. Therefore, there is a single null set X_f so that (2.3) holds for all $n\in\mathbb{N}$. We can further assume that for $x \in X_f$ that we have

$$\mathrm{M}\, f(x) \stackrel{\text{def}}{=} \sup_t \frac{1}{2t} \int_{-t}^{t} |f(\mathrm{T}_s\, x)|\, ds < \infty.$$

To conclude, observe that for all $0<a<\infty$,

$$\left| \int e^{i\theta t} f(\mathrm{T}_t\, x) \left[\zeta_k\left(\frac{t}{a}\right) - \zeta_\infty\left(\frac{t}{a}\right) \right] \frac{dt}{t} \right| \lesssim 2^{-k}\, \mathrm{M}\, f(x).$$

Thus, we see that the corollary holds.

3. Deduction of Theorem 1.4

There are two more technical estimates that we prove. Specifically, let ψ be some Schwartz function which satisfies

(3.1) $$0 \le \hat{\psi}(\xi) \le C_0,$$

(3.2) $$\hat{\psi} \text{ is supported in } \left[-2, -\tfrac{1}{2}\right],$$

(3.3) $$|\psi(y)| \le C_1 \min(|y|^{-\nu}, |y|^{\nu}).$$

Here, ν will be a large constant whose exact value we need not specify. And we will not have complete freedom in precisely which Schwartz function ψ we can take here. It should arise in a particular way described in the proof of Proposition 3.4, and will be non-zero! The purpose of this section is to describe how a particular result for *any choice of non-zero* ψ as above will lead to a proof of our main theorem.

Consider the distribution

$$\Psi \stackrel{\text{def}}{=} \sum_{v=1}^{\infty} \mathrm{Dil}^{(1)}_{2^{-v}}\, \psi.$$

We will prove the following two propositions in the next section.

Proposition 3.4. *With the assumptions* (3.2)–(3.3), *the inequality* (1.5) *holds with* $n=1$ *and the distribution* K_H *replaced by* Ψ.

Proposition 3.5. *We have the inequality*

(3.6) $$\left\| \sup_N \left[\sum_{j=-\infty}^{\infty} |(\mathrm{Dil}^{(1)}_{2^j}\, \psi) * \mathrm{Mod}_N\, f|^2 \right]^{1/2} \right\|_p \lesssim \|f\|_p, \quad 1<p<\infty.$$

Note that for fixed modulation, (3.6) is a Littlewood–Paley inequality, making the inequality above a ‘Carlesonized Littlewood–Paley’ inequality. Inequalities like this have been proved by Prestini and Sjölin [20]. They also follow from the method of Lacey and Thiele.

Both propositions follow from our Proposition 4.9 of the next section, which is phrased in a language conducive to the methods of Lacey and Thiele [16]. These methods have been applied in a number of variants of Carleson's theorem, see e.g. Pramanik and Terwilleger [19] and Grafakos, Tao, and Terwilleger [11].

We turn to the deduction of Theorem 1.4. Observe that the two previous propositions immediately prove that when we consider dilations which are powers of $2^{1/n}$ we have

$$\Big\|\sup_N \operatorname{Osc}_n(\Psi;\operatorname{Mod}_N f)\Big\|_p \lesssim n\|f\|_p, \quad n\in\mathbb{N},\ 1<p<\infty.$$

Thus we need not concern ourselves with these features of Theorem 1.4 and Corollary 1.6.

For a distribution K, set

$$\|K\|_{*,p}=\sup_{\|f\|_p=1}\Big\|\sup_N \operatorname{Osc}_1(K;\operatorname{Mod}_N f)\Big\|_p.$$

Note that since our definition incorporates differences, this is a seminorm on distributions K. That is, it obeys the triangle inequality (which we use), but can be zero for non-zero distributions. In particular, for a Dirac point mass δ we have $\|\delta\|_{*,p}=0$, and similarly for the distribution K with $\widehat{K}=\mathbf{1}_{[0,\infty)}$.

Our task is to show that $\|K_H\|_{*,p}<\infty$, where $K_H(y)=y^{-1}\zeta(y)$ for some smooth, symmetric, compactly supported Schwartz function. Our Proposition 3.4 is, with this notation, the assertion that $\|\Psi\|_{*,p}<\infty$. The same inequality will hold for a kernel which can be obtained as a convex combination of dilations of ψ and Ψ. Thus, set

$$\Psi_0 \overset{\text{def}}{=} \int_0^1 \operatorname{Dil}^{(1)}_{2^s}\Psi\,\frac{ds}{s}.$$

In this integral, we are careful to integrate against the measure ds/s, which is the Haar measure for the positive reals under multiplication, the underlying group for the dilation operators. In particular, it follows that Ψ_0 is a distribution whose Fourier transform is a non-zero constant on $(-\infty,-1)$ and is 0 on $\left(-\frac{1}{2},\infty\right)$. Thus by Proposition 3.4, we clearly have $\|\Psi_0\|_{*,p}<\infty$.

Now we will show that $\|D_0\|_{*,p}<\infty$ for the distribution

$$D_0(y)=y^{-1}\zeta(y)-c(\Psi_0(y)-\overline{\Psi_0(y)}),$$

where we choose the complex constant c so that $\lim_{\xi\to\infty} \widehat{D}_0(\xi)=0$. In fact, it is a well-known elementary fact that for $c=i\pi$,

$$\int_{-\infty}^{\infty} \zeta(y)e^{i\xi y}\,\frac{dy}{y} = c+O(|\xi|^{-1}). \tag{3.7}$$

We will decompose the distribution D_0 into a sum which can be treated with Proposition 3.5. Then using that $\|\Psi_0\|_{*,p}<\infty$ and $\|D_0\|_{*,p}<\infty$, we obtain the desired inequality for $K(y)=y^{-1}\zeta(y)$.

Choose χ to be a smooth function supported on $\frac{1}{2}\le|\xi|\le 2$ so that

$$\sum_{k=-\infty}^{\infty} \operatorname{Dil}^{\infty}_{2^{-k}} \chi = \mathbf{1}_{\mathbb{R}\setminus\{0\}},$$

and set $\hat{\Delta}_k=\widehat{D}_0 \operatorname{Dil}^{\infty}_{2^{-k}} \chi$. The following lemma finishes the proof of Theorem 1.4.

Lemma 3.8. *We have*

$$\|\Delta_k\|_{*,p} \lesssim 2^{-|k|}, \quad k\in\mathbb{Z}.$$

Proof. We will verify that

$$\|\hat{\Delta}_k\|_{\infty} \lesssim 2^{-|k|}, \qquad k\in\mathbb{Z}, \tag{3.9}$$

$$\hat{\Delta}_k \text{ is supported on } 2^{-k-1}\le|\xi|\le 2^{-k+1}, \tag{3.10}$$

$$|\Delta_k(y)| \lesssim 2^{-k-|k|}(1+2^{-k}|y|)^{-\nu}, \qquad k\in\mathbb{Z},\ y\in\mathbb{R}, \tag{3.11}$$

with implied constants independent of $k\in\mathbb{Z}$ and ν being the large, unspecified constant that appears in (3.3). With decay in $|k|$ in both (3.9) and (3.11), the lemma then follows from a trivial change of scale and from Proposition 3.5.

Let us recall the trivial estimate which follows from the symmetry of ζ,

$$|\widehat{K}_H(\xi)| = \left|\int_{-\infty}^{\infty} \zeta(y)\frac{e^{i\xi y}}{y}\,dy\right| \lesssim |\xi|. \tag{3.12}$$

In addition we have the estimate below, applied for $|\xi|\le 1$,

$$\left|\frac{d^w}{d\xi^w}\widehat{K}_H(\xi)\right| = \left|\int_{-\infty}^{\infty} \zeta(y)y^{w-1}e^{i\xi y}\,dy\right| \lesssim \begin{cases} |\xi|, & w \text{ even},\\ 1, & w \text{ odd}.\end{cases} \tag{3.13}$$

Whereas for $|\xi|\ge 1$, we have

$$\left|\frac{d^w}{d\xi^w}\widehat{K}_H(\xi)\right| \lesssim |\xi|^{-\nu}, \quad |\xi|>1,\ 0<w\le\nu. \tag{3.14}$$

That is, we have very rapid decay in a large number of derivatives.

Now, (3.10) is true by definition of Δ_k. To see (3.9) for $k\geq 2$, note that this is only determined by the Fourier transform of K_H since $\widehat{\Psi}_0$ and $\widehat{\widehat{\Psi}}_0$ are zero. The result easily follows by the inequality (3.12) and property (3.10). For $k\leq 2$, the inequality follows from the construction of D_0, and in particular the property in (3.7).

We turn to the last condition, (3.11). It is well known that decay of order ν in spatial variables is implied by differentiability of a function in frequency variables. Observe that

$$(y^\nu \Delta_k(y))\hat{}(\xi) = i^{-\nu}\,\frac{d^\nu}{d\xi^\nu}\hat{\Delta}_k(\xi) = i^{-\nu}\,\frac{d^\nu}{d\xi^\nu}\,\mathrm{Dil}^{(\infty)}_{2^{-k}}\,\chi(\xi)\widehat{K}_H(\xi).$$

Hence,

$$|(y^\nu \Delta_k(y))\hat{}(\xi)| \leq \sum_{w=0}^{\nu} 2^{kw}\left|\frac{d^{\nu-w}}{d\xi^{\nu-w}}\sup_{2^{-k-1}\leq|\xi|\leq 2^{-k+1}}\widehat{K}_H(\xi)\right|.$$

For $k\geq 1$, this sum is dominated by the last two terms. To control them, use (3.13), supplying the estimate $\lesssim 2^{(\nu-1)k}$. This is better by a factor of 2^{-k} than the trivial estimate, so that Fourier inversion proves (3.11) in this case.

The case of $k\leq 0$ is easier, due to the rapid decay in (3.14). □

4. Decomposition and main proposition

We state the definitions needed for the main proposition and conclude this section with the argument of how this proposition proves the results of the previous section, namely Propositions 3.4 and 3.5.

In addition to the modulation and dilation operators in (1.1) and (1.2), we need *translation operators*

$$\mathrm{Tran}_y\, f(x) \stackrel{\mathrm{def}}{=} f(x-y), \quad y\in\mathbb{R}. \tag{4.1}$$

We set $\mathcal{D}$ to be the dyadic grid and say that $I\times\omega\in\mathcal{D}\times\mathcal{D}$ is a *tile* if $|\omega||I|=1$. Let $\mathcal{T}$ denote the set of all tiles.

We think of ω as a frequency interval and I as a spatial interval; our definition of a tile is a reflection of the uncertainty principle for the Fourier transform. We will plot frequency intervals in the vertical direction. Each dyadic interval ω is a union of two dyadic intervals of half the length of ω. We call them ω_+ and ω_- and view ω_+ as above ω_-.

We take a fixed Schwartz function φ with frequency support in the interval $[-1/\nu, 1/\nu]$. For a tile $s = I_s \times \omega_s$, define

$$\varphi_s \stackrel{\text{def}}{=} \operatorname{Mod}_{c(\omega_{s,-})} \operatorname{Tran}_{c(I_s)} \operatorname{Dil}^2_{|I_s|} \varphi. \tag{4.2}$$

Here, $c(J)$ is the center of the interval J, and $\omega_{s,-}$ is the lower half of the interval ω_s. Thus, this function is localized to be supported in the time-frequency plane close to the rectangle $I_s \times \omega_{s,-}$.

There are companion functions which depend on different choices of certain measurable functions. These functions should be thought of as those choices of modulation and indices that will achieve, up to a constant multiple, the supremums in the oscillation function. To linearize the modulation, let

$$N\colon \mathbb{R} \longrightarrow \mathbb{R} \text{ be a measurable function (a modulation parameter).}$$

We define another function related to the rectangle $I_s \times \omega_{s,+}$ which tells us when the linearized modulation parameter is at a certain frequency. Let

$$\phi_s(x) \stackrel{\text{def}}{=} \mathbf{1}_{\omega_{s,+}}(N(x))\varphi_s(x). \tag{4.3}$$

Now define a tile variant of the oscillation operator by

$$\text{Tile-osc}(f) \stackrel{\text{def}}{=} \left[\sum_{j=1}^{\infty} \sup_{k_j \le l < l' < k_{j+1}} \Bigg| \sum_{\substack{s \in \mathcal{T} \\ 2^l \le |I_s| \le 2^{l'}}} \langle f, \varphi_s \rangle \phi_s \Bigg|^2 \right]^{1/2}. \tag{4.4}$$

Here, an increasing sequence of integers $\{k_j\}_{j=1}^{\infty}$ are specified in advance. We make the definition for clarity's sake, as we will not explicitly work with it. Rather we prefer to fully linearize this maximal operator. This requires the additional choices of functions

$$\alpha_j\colon \mathbb{R} \longrightarrow \mathbb{R}, \quad \sum_{j=1}^{\infty} |\alpha_j(x)|^2 \le 1 \text{ for all } x, \tag{4.5}$$

$$\ell_{j-}, \ell_{j+}\colon \mathbb{R} \longrightarrow \mathbb{Z}, \quad k_j \le \ell_{j-} < \ell_{j+} < k_{j+1}. \tag{4.6}$$

And we set

$$F_{s,j} \stackrel{\text{def}}{=} \{x : 2^{\ell_{j-}(x)} \le |I_s| < 2^{\ell_{j+}(x)}\}, \tag{4.7}$$

$$f_{s,j}(x) \stackrel{\text{def}}{=} \mathbf{1}_{F_{s,j}}(x)\alpha_j(x)\phi_s(x). \tag{4.8}$$

The sequences of functions $\ell_{j\pm}$ are selecting the level at which the maximal difference occurs. The α_j are chosen to realize the ℓ^2- norm in the definition of oscillation. We make all of these choices in order to linearize the oscillation operator.

Our main proposition is the following result.

Proposition 4.9. *For all choices of $N(x)$ and increasing sequences of integers $\{k_j\}_{j=1}^{\infty}$, the operator* Tile-osc *extends to a bounded sublinear operator on L^p, $1<p<\infty$. In particular, for sets $G,H\subset\mathbb{R}$ of finite measure, there exists a set $H'\subset H$ such that $|H'|\geq\frac{1}{2}|H|$ and*

$$\sum_{j=1}^{\infty}\sum_{s\in\mathcal{T}}|\langle \mathbf{1}_G,\varphi_s\rangle\langle \mathbf{1}_{H'},f_{s,j}\rangle|\lesssim \min(|G|,|H|)\left(1+\left|\log\frac{|G|}{|H|}\right|\right). \tag{4.10}$$

Note that the inequality above implies that

$$|\langle \text{Tile-osc}(\mathbf{1}_G),\mathbf{1}_{H'}\rangle|\lesssim |G|^{1/p}|H|^{1-1/p},\quad 1<p<\infty.$$

That is, we have the restricted weak-type inequality for all $1<p<\infty$.([1]) Then we use that a function is in $L^{p,\infty}$ if for every measurable set H, there is a subset H' of H such that $|H'|\geq\frac{1}{2}|H|$ and the integral of the function over the set H' is at most a constant times $|H|^{1-1/p}$. Hence, by Marcinkiewicz interpolation [21] and the restricted weak-type reduction of Stein and Weiss [22], we can obtain the estimate

$$\|\text{Tile-osc}(f)\|_p\lesssim\|f\|_p,\quad 1<p<\infty. \tag{4.11}$$

The deduction of Proposition 3.4 and Proposition 3.5

For $\xi\in\mathbb{R}$ and $l\in\mathbb{Z}$, consider the operators

$$\mathrm{A}_{\xi,l}\,f\stackrel{\text{def}}{=}\sum_{|I_s|=2^l}\mathbf{1}_{\xi\in\omega_{s,+}}\langle f,\varphi_s\rangle\varphi_s.$$

The tile oscillation operator is built up from these operators. Observe that these operators enjoy the properties

$$\mathrm{A}_{\xi,l}\,\mathrm{Trans}_{n2^l}=\mathrm{Trans}_{n2^l}\,\mathrm{A}_{\xi,l},\qquad n\in\mathbb{Z}, \tag{4.12}$$

$$\mathrm{A}_{\xi,l}\,\mathrm{Dil}^{(2)}_{2^{-l'}}=\mathrm{Dil}^{(2)}_{2^{-l'}}\,\mathrm{A}_{\xi2^{-l'},l+l'},\qquad l'\in\mathbb{Z}, \tag{4.13}$$

$$\mathrm{A}_{\xi,l}\,\mathrm{Mod}_{-\theta}=\mathrm{Mod}_{-\theta}\,\mathrm{A}_{\xi+\theta,l},\qquad \theta\in\mathbb{R}. \tag{4.14}$$

Notice that these conditions tell us that the operators $\mathrm{A}_{\xi,l}$ have a near translation invariance, a certain modulation invariance, and are related to each other through dilations. In addition, these operators are bounded on L^2 uniformly in ξ and l, a fact well represented in the literature.

([1]) In fact, the estimate (4.10) gives a favorable upper bound on the behavior of the constant with respect to p, namely that they are no more than $\max(p,p/(p-1))$, see [11].

We will now define

$$\mathrm{B}_{\xi,l}\, f \stackrel{\mathrm{def}}{=} \lim_{\substack{K\to\infty\\ L\to\infty}} \frac{1}{4KL} \int_{-K}^{K}\int_{-L}^{L} \mathrm{Mod}_{-\theta}\, \mathrm{Trans}_{-y}\, \mathrm{A}_{(\xi+\theta),l}(\mathrm{Trans}_y\, \mathrm{Mod}_\theta\, f)\, dy\, d\theta.$$

By periodicity of the integrand in y and θ, for all Schwartz functions f, the averages in the right-hand side converge pointwise to $\mathrm{B}_{\xi,l}\, f(x)$ as $K, L\to\infty$.

Let us make some observations about the operators $\mathrm{B}_{\xi,l}$. First, (4.12) and periodicity of the integrand in y imply that $\mathrm{B}_{\xi,l}$ commutes with translations. Second, it is a bounded, positive, semidefinite operator, as is easy to see. Hence, it is given by convolution. Indeed, (4.14) implies that

$$\mathrm{B}_{\xi,l}\, f = \mathrm{Mod}_\xi\, \beta_l * \mathrm{Mod}_{-\xi}\, f$$

for a function β_l that we turn to next. The equality (4.13) implies that $\beta_l = \mathrm{Dil}^{(1)}_{2^l}\, \beta$, where β is given such that β_0 is a smooth Schwartz function satisfying the conditions (3.1)–(3.3), a routine exercise to verify.

Assuming Proposition 4.9, it follows that we can conclude Propositions 3.4 and 3.5 for non-zero functions $\psi = \beta_0$. Our proof is complete.

5. Main lemmas

To prove Proposition 4.9, we must first obtain the set $H' \subset H$. Consider the set

$$\Omega = \left\{ x : \mathrm{M}\, \mathbf{1}_G(x) > 6 \min\left(1, \frac{|G|}{|H|}\right) \right\},$$

where M is the Hardy–Littlewood maximal function. Setting $H' = H \setminus \Omega$ and using that M is of weak type (1,1) with constant at most 3, we see that $|H'| \geq \frac{1}{2}|H|$. To prove (4.10), we split the sum over $s \in \mathcal{T}$ into the sum over s such that $I_s \subset \Omega$ and the sum over s such that $I_s \not\subset \Omega$. The former sum can be taken care of by an argument of M. Lacey and C. Thiele [15, Section 4] done in the bilinear setting. It also appears, modified to the linear case as is our situation here, in L. Grafakos, T. Tao, and E. Terwilleger [11, Section 3]. The reader can also see the same argument in [14, (7.8)]. Thus we restrict our attention to the tiles s where $I_s \not\subset \Omega$.

We begin with some concepts needed to phrase the proof. There is a natural partial order on tiles. We say that $s < s'$ if $\omega_s \supset \omega_{s'}$ and $I_s \subset I_{s'}$. Note that the time variable of s is localized to that of s', and the frequency variable of s is similarly localized, up to the variability allowed by the uncertainty principle. Note that two tiles are incomparable with respect to the '$<$' partial order if and only if the tiles, as rectangles in the time-frequency plane, do not intersect. A *maximal tile* will be one that is maximal with respect to this partial order.

Let $\mathcal{S}$ denote an arbitrary set of tiles. We call a set of tiles $\mathbf{T}\subset\mathcal{S}$ a *tree* if there is a tile $I_{\mathbf{T}}\times\omega_{\mathbf{T}}$, called the *top of the tree,* such that for all $s\in\mathbf{T}$, $s<I_{\mathbf{T}}\times\omega_{\mathbf{T}}$. We note that the top is not uniquely defined. An important point is that a tree top specifies a location in the time variable for the tiles in the tree, namely inside $I_{\mathbf{T}}$, and localizes the frequency variables, identifying $\omega_{\mathbf{T}}$ as a nominal origin.

We say that *the count of* $\mathcal{S}$ *is at most* A if $\mathcal{S}=\bigcup_{\mathbf{T}\subset\mathcal{S}}\mathbf{T}$, where each $\mathbf{T}\subset\mathcal{T}$ is a tree which is maximal with respect to inclusion and

$$\operatorname{Count}(\mathcal{S})\stackrel{\text{def}}{=}\sum_{\mathbf{T}\subset\mathcal{S}}|I_{\mathbf{T}}|\le A.$$

Fix $\chi(x)=(1+|x|)^{-\nu}$, where ν is, as before, a large constant whose exact value is unimportant to us. Define

$$\chi_I:=\operatorname{Trans}_{c(I)}\operatorname{Dil}^{(1)}_{|I|}\chi, \tag{5.1}$$

$$\operatorname{dense}(s):=\sup_{s<s'}\int_{N^{-1}(\omega_{s'})\cap H}\chi_{I_{s'}}\,dx, \tag{5.2}$$

$$\operatorname{dense}(\mathcal{S}):=\sup_{s\in\mathcal{S}}\operatorname{dense}(s),\qquad \mathcal{S}\subset\mathcal{T}.$$

The first and most natural definition of a 'density' of a tile, would be $|I_s|^{-1}|N^{-1}(\omega_{s,+})\cap I_s|$. However φ is supported on the whole real line, although it does decay faster than the inverse of any polynomial. We refer to this as a 'Schwartz tails problem'. The definition of density as $\int_{N^{-1}(\omega_s)}\chi_{I_s}\,dx$, as it turns out, is still not adequate. That we should take the supremum over $s<s'$ only becomes evident in the proof of the 'tree lemma' below.

The following 'density lemma' is well known in the literature. See [16, Proposition 3.1] or the survey [14, Lemma 3.6].

Lemma 5.3. *Any subset* $\mathcal{S}\subset\mathcal{T}$ *is a union of* $\mathcal{S}_{\text{heavy}}$ *and* $\mathcal{S}_{\text{light}}$ *for which*

$$\operatorname{dense}(\mathcal{S}_{\text{light}})<\tfrac{1}{2}\operatorname{dense}(\mathcal{S}),$$

and the collection $\mathcal{S}_{\text{heavy}}$ *satisfies*

$$\operatorname{Count}(\mathcal{S}_{\text{heavy}})\lesssim\operatorname{dense}(\mathcal{S})^{-1}|H|. \tag{5.4}$$

What is significant is that this relatively simple lemma admits a non-trivial variant intimately linked to the tree structure and orthogonality. We should refine the notion of a tree. Recall that $\omega_{s,+}$ is the top half and $\omega_{s,-}$ the bottom half of the frequency interval ω_s associated to a tile s. Call a tree $\mathbf{T}$ with top $I_{\mathbf{T}}\times\omega_{\mathbf{T}}$

a *+tree* if for each $s \in \mathbf{T}$, aside from the top, $I_{\mathbf{T}} \times \omega_{\mathbf{T}} \cap I_s \times \omega_{s,+}$ is not empty and a *−tree* if for each $s \in \mathbf{T}$, aside from the top, $I_{\mathbf{T}} \times \omega_{\mathbf{T}} \cap I_s \times \omega_{s,-}$ is not empty. Any tree is a union of a +tree and a −tree . If $\mathbf{T}$ is a +tree , observe that the rectangles $\{I_s \times \omega_{s,-} : s \in \mathbf{T}\}$ are disjoint. We see that

$$\sum_{s \in \mathbf{T}} |\langle f, \varphi_s \rangle|^2 \lesssim \|f\|_2^2.$$

This motivates the definition

$$\operatorname{size}(\mathcal{S}) := \sup\left\{ \left[\frac{1}{|I_{\mathbf{T}}|} \sum_{s \in \mathbf{T}} |\langle f, \varphi_s \rangle|^2 \right]^{1/2} : \mathbf{T} \subset \mathcal{S} \text{ is a +tree} \right\}. \tag{5.5}$$

The following 'size lemma' has also appeared in the literature. See [16, Proposition 3.2] or the survey [14, Lemma 3.9].

Lemma 5.6. *Assume that $f = \mathbf{1}_G$. Any subset $\mathcal{S} \subset \mathcal{T}$ is a union of $\mathcal{S}_{\mathrm{big}}$ and $\mathcal{S}_{\mathrm{small}}$ for which*

$$\operatorname{size}(\mathcal{S}_{\mathrm{small}}) < \tfrac{1}{2} \operatorname{size}(\mathcal{S}),$$

and the collection $\mathcal{S}_{\mathrm{big}}$ satisfies

$$\operatorname{Count}(\mathcal{S}_{\mathrm{big}}) \lesssim \operatorname{size}(\mathcal{S})^{-2} |G|. \tag{5.7}$$

Concerning the quantity size, we need an additional piece of information about it. Recall that M is the Hardy–Littlewood maximal function.

Lemma 5.8. *Suppose that $\mathcal{S}$ is the set of tiles with*

$$I_s \not\subset \Omega, \quad s \in \mathcal{S}.$$

Then it is the case that $\operatorname{size}(\mathcal{S}) \lesssim \min(1, |G|/|H|)$.

This fact, a delicate consequence of the Calderón–Zygmund decomposition, was proved by Muscalu, Tao, and Thiele [18, Lemma 7.8] in the multilinear case and will not be proved in this paper. See also [14, Lemma 7.10] for an alternative proof, and in [11, Lemma 1] for the adaptation to higher dimensions.

For a set of tiles $\mathcal{S}$, set

$$\operatorname{Sum}(\mathcal{S}) \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \sum_{s \in \mathcal{S}} |\langle \mathbf{1}_G, \varphi_s \rangle \langle \mathbf{1}_H, f_{s,j} \rangle|.$$

Our final lemma relates trees, density and size. It is the 'tree lemma'.

Lemma 5.9. *For any tree* $\mathbf{T}$

$$\mathrm{Sum}(\mathbf{T}) \lesssim \mathrm{size}(\mathbf{T})\,\mathrm{dense}(\mathbf{T})|I_{\mathbf{T}}|. \tag{5.10}$$

Of course for any set of tiles $\mathcal{S}$, we would then have

$$\mathrm{Sum}(\mathcal{S}) \lesssim \sum_{\mathbf{T}\subset\mathcal{S}} \mathrm{size}(\mathbf{T})\,\mathrm{dense}(\mathbf{T})|I_{\mathbf{T}}|.$$

Proof of Proposition 4.9. Recall that we are considering the tiles $s\in\mathcal{S}$ such that $I_s\not\subset\Omega$. This style of proof is well represented in the literature. See Muscalu, Tao, and Thiele [18] and Grafakos, Tao, and Terwilleger [11] for two examples. We include details for the sake of completeness.

To fix ideas, let us first consider the case of $|G|\simeq|H|$. We inductively apply Lemmas 5.6 and 5.3 so that the estimates on the counts of the collections we get in (5.7) and (5.4) are approximately the same. Thus, we can decompose the set of tiles $\mathcal{S}$ into pairwise disjoint subcollections of tiles $\mathcal{S}_n$, $n\geq 0$, such that

$$\mathrm{size}(\mathcal{S}_n) \lesssim 2^{-n}, \quad \mathrm{dense}(\mathcal{S}_n) \lesssim 2^{-2n} \quad \text{and} \quad \mathrm{Count}(\mathcal{S}_n) \lesssim 2^{2n}|G|.$$

Thus, applying the tree lemma, we see that

$$\mathrm{Sum}(\mathcal{S}_n) \lesssim 2^{-n}|G|, \quad n\geq 0.$$

This is summable in n to our desired estimate. This case is complete.

The next case is that of $|G|>3|H|$. In this case, the best estimate on the count that we can hope to get from the size lemma starts off much larger than that from the density lemma. Therefore, it is more efficient to initially just apply the density lemma.

For $0\leq n\leq\log(|G|/|H|)$, we apply the density lemma to get subcollections of tiles $\mathcal{S}_n$ such that

$$\mathrm{size}(\mathcal{S}_n) \lesssim 1, \quad \mathrm{dense}(\mathcal{S}_n) \lesssim 2^{-2n} \quad \text{and} \quad \mathrm{Count}(\mathcal{S}_n) \lesssim 2^{2n}|H|.$$

Clearly, we have

$$\mathrm{Sum}(\mathcal{S}_n) \lesssim |H|, \quad 0\leq n\leq \log\frac{|G|}{|H|}.$$

And this is summed over n to get the claimed bound of $|H|\log(|G|/|H|)$.

For the remaining collection of tiles, we proceed as in the first case. Namely, for $n\geq\log(|G|/|H|)$, apply the size lemma or density lemma, as needed, to get subcollections of tiles $\mathcal{S}_n$ satisfying

$$\mathrm{size}(\mathcal{S}_n) \lesssim 2^{-n}, \quad \mathrm{dense}(\mathcal{S}_n) \lesssim 2^{-2n} \quad \text{and} \quad \mathrm{Count}(\mathcal{S}_n) \lesssim 2^{2n}|G|.$$

Then we have

$$\operatorname{Sum}(\mathcal{S}_n) \lesssim 2^{-n}|G|, \quad n \geq \log \frac{|G|}{|H|}.$$

Again this is summed over n to get the claimed bound of $|H| \lesssim |H| \log(|G|/|H|)$.

The last case, that of $3|G|<|H|$, is the most intricate of the three cases and is the one where the refined size estimate is needed.(2) Applying Lemma 5.8, we obtain the estimate $\operatorname{size}(\mathcal{S}) \lesssim |G|/|H|$.

Initially applying the density lemma to $\mathcal{S}$ will yield a better bound on the count than from the size lemma. Thus, for $0 \leq n \leq \log(|H|/|G|)$, we apply the density lemma to get subcollections $\mathcal{S}_n \subset \mathcal{S}$ such that

$$\operatorname{size}(\mathcal{S}_n) \lesssim \frac{|G|}{|H|}, \quad \operatorname{dense}(\mathcal{S}_n) \lesssim 2^{-n} \quad \text{and} \quad \operatorname{Count}(\mathcal{S}_n) \lesssim 2^n |H|.$$

Clearly we have

$$\operatorname{Sum}(\mathcal{S}_n) \lesssim |G|,$$

and summing over $0 \leq n \leq \log(|H|/|G|)$ yields the bound $|G| \log(|H|/|G|)$.

For $n \geq \log(|H|/|G|)$, we apply the size lemma or the density lemma, as needed, to decompose the remaining tiles into subcollections $\mathcal{S}_n$ with

$$\operatorname{size}(\mathcal{S}_n) \lesssim \sqrt{2^{-n} \frac{|G|}{|H|}}, \quad \operatorname{dense}(\mathcal{S}_n) \lesssim 2^{-n} \quad \text{and} \quad \operatorname{Count}(\mathcal{S}_n) \lesssim 2^n |H|.$$

For these collections, we have

$$\operatorname{Sum}(\mathcal{S}_n) \lesssim \sqrt{2^{-n}|G||H|},$$

which summed over $n \geq \log(|H|/|G|)$ will give us the bound of $\lesssim |G|$. □

6. Proof of Lemma 5.9

The tree lemma, with its adaptation to the setting of oscillation, is the primary new step in this paper. This lemma appears in [16] and [11] with the functions ϕ_s rather than the functions $f_{s,j}$. These new functions have the oscillation built into them and dealing with them brings about new difficulties.

We begin with some remarks about oscillation operators and a particular form of the same that we shall use at a critical point of this proof. Let ζ be a smooth

(2) This case is the one that is needed to deduce the oscillation inequalities for $1<p<2$.

function with Fourier transform supported in $[-1-\varepsilon, 1+\varepsilon]$ for a fixed, small, positive ε and equal to 1 on $[-1,1]$. Set

$$\operatorname{Osc}(f)^2 \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} \sup_{2^{k_j} \le |I| \le |I'| \le 2^{k_{j+1}}} \bigl|\operatorname{Dil}^{(1)}_{|I|} \zeta * f - \operatorname{Dil}^{(1)}_{|I'|} \zeta * f\bigr|^2.$$

It is known that this is bounded on L^2, and in this situation we will give an elementary proof of this fact below.

We shall have recourse to not only this bound, but a particular refinement. Let $\mathcal{J}$ be a partition of $\mathbb{R}$ into dyadic intervals. To each $J \in \mathcal{J}$, associate a subset $E(J) \subset J$ with $|E(J)| \le \delta |J|$, where $0<\delta<1$ is fixed. Consider

$$\operatorname{Osc}_\delta(f)^2 \stackrel{\text{def}}{=} \sum_{J \in \mathcal{J}} \mathbf{1}_{E(J)} \sum_{j=1}^{\infty} \sup_{\substack{J \subset I \subset I' \\ 2^{k_j} \le |I| \le |I'| \le 2^{k_{j+1}}}} |\mathbf{1}_I \langle \zeta_I, f\rangle - \mathbf{1}_{I'} \langle \zeta_{I'}, f\rangle|^2 \tag{6.1}$$

We estimate the norm of this operator.

Lemma 6.2. *We have the estimate*

$$\|\operatorname{Osc}_\delta(f)\|_2 \lesssim \sqrt{\delta}\, \|f\|_2 \tag{6.3}$$

for all $f \in L^2$.

Proof. Let us begin with a proof that $\|\operatorname{Osc}\|_{2\to 2} \lesssim 1$. That is, we do not have the additional information about the partition $\mathcal{J}$, and sets $E(J)$ for $J \in \mathcal{J}$. For a sequence of increasing integers k_j and function $f \in L^2$, set

$$\hat{f}_j = \mathbf{1}_{\{\xi : 2^{-k_{j+1}-1} \le |\xi| \le 2^{-k_j+1}\}} \hat{f}$$

Then, we certainly have $\sum_{j=1}^{\infty} \|f_j\|_2^2 \le 3\|f\|_2^2$. Moreover, due to our assumption about the function ζ,

$$\sup_{2^{k_j} \le |I| \le |I'| \le 2^{k_{j+1}}} |\operatorname{Dil}^{(1)}_{|I|} \zeta * f| \le \mathrm{M} f_{j-1} + \mathrm{M} f_j + \mathrm{M} f_{j+1},$$

where M is the usual maximal function. Thus, by the boundedness of the maximal function on L^2 we have

$$\|\operatorname{Osc}(f)\|_2^2 \le 3 \sum_{j=1}^{\infty} \|\mathrm{M} f_j\|_2^2 \lesssim \|f\|_2^2.$$

It is hardly surprising that the proof above appeals to the boundedness of the maximal function, since the estimate on the oscillation operator implies that for the maximal function. Likewise, our lemma implies a bound for a certain variant of the maximal function. As it turns out, we need this variant in the course of the proof.

Define

$$\mathrm{M}_\delta f(x) \stackrel{\text{def}}{=} \sum_{J\in\mathcal{J}} \mathbf{1}_{E(J)}(x) \sup_{J\subset I} \langle |f|, \chi_I\rangle,$$

where χ_I is defined as in (5.1). Then the estimate we claim is $\|\mathrm{M}_\delta\|_2 \lesssim \sqrt{\delta}$. Indeed, for any point $x\in E(J)$, we have the inequality

$$\mathrm{M}_\delta f(x) \lesssim \inf_{y\in J} \mathrm{M} f(y),$$

where M is the usual maximal function. Therefore, we can estimate

$$\|\mathrm{M}_\delta f\|_2^2 = \sum_{J\in\mathcal{T}} \int_{E(J)} \mathrm{M}_\delta f(x)^2\,dx \lesssim \sum_{J\in\mathcal{T}} |E(J)| \inf_{y\in J} \mathrm{M} f(y)^2 \le \delta \int_{-\infty}^{\infty} \mathrm{M} f(x)^2\,dx.$$

This proves our claim.

To conclude the proof, we can estimate

$$\int_{\bigcup_{J\in\mathcal{T}} E(J)} \mathrm{Osc}_\delta(f)(x)^2\,dx \lesssim \sum_{j=1}^{\infty} \|\mathrm{M}_\delta f_j\|_2^2 \lesssim \delta \sum_{j=1}^{\infty} \|f_j\|_2^2 \lesssim \delta \|f\|_2^2.$$

Our proof is complete. □

We begin the main line of the argument. Let $\delta=\mathrm{dense}(\mathbf{T})$, and $\sigma=\mathrm{size}(\mathbf{T})$. By a modification of the functions $\alpha_j(x)$ by a choice of signs, we can assume the identity

$$\sum_{j=1}^{\infty} \sum_{s\in\mathbf{T}} |\langle \mathbf{1}_G, \varphi_s\rangle \langle f_{s,j}, \mathbf{1}_{H'}\rangle| = \int_{H'} \sum_{j=1}^{\infty} \sum_{s\in\mathbf{T}} \langle \mathbf{1}_G, \varphi_s\rangle f_{s,j}(x)\,dx.$$

As we have no particular control on the set H', we will need the following partition of the real line induced by the tree $\mathbf{T}$. Let $\mathcal{J}$ be the partition of $\mathbb{R}$ consisting of the maximal dyadic intervals J such that $3J$ does not contain any I_s for $s\in\mathbf{T}$. It is helpful to observe that for such J, if $|J|\le|I_{\mathbf{T}}|$, then $J\subset 3I_{\mathbf{T}}$, and if $|J|\ge|I_{\mathbf{T}}|$, then $\mathrm{dist}(J, I_{\mathbf{T}})\gtrsim|J|$. The integral above is at most the sum of the two terms

$$\sum_{j=1}^{\infty} \sum_{J\in\mathcal{J}} \sum_{\substack{s\in\mathbf{T} \\ |I_s|\le 2|J|}} |\langle \mathbf{1}_G, \varphi_s\rangle| \int_{J\cap H'} |f_{s,j}(x)|\,dx, \tag{6.4}$$

$$\sum_{j=1}^{\infty} \sum_{J\in\mathcal{J}} \int_{J\cap H'} \Bigg| \sum_{\substack{s\in\mathbf{T} \\ |I_s|>2|J|}} \langle \mathbf{1}_G, \varphi_s\rangle f_{s,j}(x) \Bigg|\,dx. \tag{6.5}$$

Notice that for the second sum to be non-zero, we must have $J\subset 3I_{\mathbf{T}}$.

The first term (6.4) is controlled by an appeal to the 'Schwartz tails'. Fix an integer $n\geq -1$, and only consider those $s\in\mathbf{T}$ for which $|I_s|=2^{-n}|J|$. Recalling that $f_{s,j}(x)=\mathbf{1}_{F_{s,j}}(x)\alpha_j(x)\mathbf{1}_{\omega_{s,+}}(N(x))\varphi_s(x)$, we see that

$$\sum_{j=1}^{\infty}\sum_{\substack{s\in\mathbf{T}\\|I_s|=2^{-n}|J|}}|\langle\mathbf{1}_G,\varphi_s\rangle|\int_{J\cap H'}|f_{s,j}(x)|\,dx\lesssim\sum_{\substack{s\in\mathbf{T}\\|I_s|=2^{-n}|J|}}\sigma\delta(|I_s|^{-1}\mathrm{dist}(I_s,J))^{-10}|I_s|$$
$$\lesssim\sigma\delta 2^{-n}\min(|J|,|J|(|I_{\mathbf{T}}|^{-1}\mathrm{dist}(J,I_{\mathbf{T}}))^{-5}).$$

Observe that for each s above, only one value of j contributes to the left-hand sum. In addition, we have used the fact that there are only a bounded number of tiles s for which $|I_s|^{-1}\mathrm{dist}(I_s,J)$ is essentially constant. In addition, for the case $|J|\leq|I_{\mathbf{T}}|$, we used that the distance from I_s to J is at least $\gtrsim|J|$. In the case $|J|>|I_{\mathbf{T}}|$, use $|I_s|^{-1}\mathrm{dist}(I_s,J)\geq|I_{\mathbf{T}}|^{-1}\mathrm{dist}(J,I_{\mathbf{T}})$. The estimate above can then be summed over $n\geq -1$ and $J\in\mathcal{J}$ to bound (6.4) by $\lesssim\sigma\delta|I_{\mathbf{T}}|$, as required.

Now we turn to the control of (6.5). The integral in this quantity is supported in the set

$$E(J)=J\cap\bigcup_{\substack{s\in\mathbf{T}\\|I_s|>2|J|}}(N^{-1}(\omega_{s,+})\cap H'). \tag{6.6}$$

Then the critical observation is that $|E(J)|\lesssim\delta|J|$. To see this, let J' be the next larger dyadic interval that contains J. Then $3J'$ must contain some $I_{s'}$ for $s'\in\mathbf{T}$. Hence there exists a tile s' with $I_{s'}\subset I_{s'}\subset I_{\mathbf{T}}$ such that $|I_{s'}|=2|J|$ or $|I_{s'}|=4|J|$, and $\omega_{\mathbf{T}}\subset\omega_{s'}\subset\omega_{s'}$. Then, $s'<s'$, and by the definition of density,

$$\frac{|J\cap H'\cap N^{-1}(\omega_{s'})|}{|I_{s'}|}\leq\int_{H'\cap N^{-1}(\omega_{s'})}\chi_{I_{s'}}\,dx\leq\delta.$$

But, for each s as in (6.6), we have $\omega_s\subset\omega_{s'}$, so that $E(J)\subset N^{-1}(\omega_{s'})$. Our claim follows.

Suppose that $\mathbf{T}$ is a $-$tree. This means that the tiles $\{I_s\times\omega_{s,+}:s\in\mathbf{T}\}$ are disjoint and thus the functions $f_{s,j}$ are disjointly supported. In particular, the oscillation that arises from such functions is trivially bounded by their ℓ^∞-norm. Then the bound for (6.5) is no more than

$$\sum_{j=1}^{\infty}|E(J)|\Bigg\|\sum_{\substack{s\in\mathbf{T}\\|I_s|>2|J|}}|\langle\mathbf{1}_G,\varphi_s\rangle f_{s,j}|\Bigg\|_\infty\lesssim\delta\sigma|J|.$$

This is summed over $J\subset 3I_{\mathbf{T}}$ to get the desired bound.

Suppose that $\mathbf{T}$ is a +tree. This is the interesting case. At this point, we will appeal to the norm bound for oscillation, (6.3), applied to the function

$$\Gamma \stackrel{\text{def}}{=} \operatorname{Mod}_{-c(\omega_{\mathbf{T}})} \sum_{s\in\mathbf{T}} \langle \mathbf{1}_G, \varphi_s\rangle \varphi_s.$$

This is an assumption that can be assumed by an appropriate modulation of the fixed L^2-function f. In the definition of Γ, it is useful to us that we only use the 'smooth' functions φ_s in the definition of this function. Note that $\|\Gamma\|_2 \lesssim \sigma\sqrt{|I_{\mathbf{T}}|}$, which is a consequence of the definition of size and the (near) orthogonality of the functions φ_s in the case of the +tree.

The purpose of these next remarks is to relate the sums over a +tree to oscillation. Recall that the oscillation is defined relative to a sequence of integers k_j. For each J, consider $x\in J$ and integers ℓ such that $\max(2|J|, 2^{\ell_{j-}(x)}) < 2^\ell < 2^{\ell_{j+}(x)}$. We have

$$\sum_{\substack{s\in\mathbf{T}\\ |I_s|=2^\ell}} \langle \mathbf{1}_G, \varphi_s\rangle f_{s,j}(x) = \sum_{\substack{s\in\mathbf{T}\\ |I_s|=2^\ell}} \langle \mathbf{1}_G, \varphi_s\rangle \varphi_s(x)\alpha_j(x).$$

This is because all of the intervals $\omega_{s,+}$ are nested and must contain $\omega_{\mathbf{T}}$, and if $N(x)\in\omega_{s,+}$, then it must also be in every other $\omega_{s',+}$ that is the same size or larger. What is significant here is that on the right we have a particular scale of (a modulation of) the sum that defines Γ.

Furthermore, consider the functions

$$\Gamma_{j,J}(x) \stackrel{\text{def}}{=} \operatorname{Mod}_{-c(\omega_{\mathbf{T}})} \sum_{\substack{s\in\mathbf{T}\\ \max(2|J|, 2^{\ell_{j-}(x)}) \le |I_s| \le 2^{\ell_{j+}(x)}}} \langle \mathbf{1}_G, \varphi_s\rangle \varphi_s.$$

In particular, we can choose ζ as in the definition of our oscillation operator (6.1) so that

$$\operatorname{Dil}^{(1)}_{2^\ell} \zeta * \Gamma = \operatorname{Mod}_{-c(\omega_{\mathbf{T}})} \sum_{\substack{s\in\mathbf{T}\\ |I_s|\ge 2^\ell}} \langle \mathbf{1}_G, \varphi_s\rangle \varphi_s.$$

Therefore, we have

$$\Gamma_{j,J} = \big[\operatorname{Dil}^{(1)}_{\max(2|J|, 2^{\ell_{j-}(x)})} \zeta - \operatorname{Dil}^{(1)}_{2^{\ell_{j+}(x)+1}} \zeta\big] * \Gamma.$$

We conclude that for $x\in E(J)$,

$$\sum_{j=1}^\infty \Bigg| \sum_{\substack{s\in\mathbf{T}\\ 2|J|<|I_s|}} \langle \mathbf{1}_G, \varphi_s\rangle f_{s,j}(x)\Bigg| \le \bigg(\sum_{j=1}^\infty |\Gamma_{j,J}(x)|^2\bigg)^{1/2} \bigg(\sum_{j=1}^\infty |\alpha_j(x)|^2\bigg)^{1/2} \lesssim \operatorname{Osc}_\delta \Gamma(x),$$

where we are using the oscillation operator defined in (6.1). We are able to use this operator here since $2|J|<|I_s|$ and $3J$ does not contain any I_s, which implies that $J\subset 3I_s$.

The conclusion of this proof is now at hand. By Lemma 6.2 we have

$$\begin{aligned}\sum_{\substack{J\in\mathcal{J}\\ |J|\le 3|I_{\mathbf{T}}|}}\int_{E(J)}\sum_{j=1}^{\infty}\Bigg|\sum_{\substack{s\in\mathbf{T}\\ 2|J|<|I_s|}}\langle\mathbf{1}_G,\varphi_s\rangle f_{s,j}(x)\Bigg|\,dx &\lesssim \int_{\bigcup_{|J|\le 3|I_{\mathbf{T}}|}E(J)}|\mathrm{Osc}_\delta\,\Gamma(x)|\,dx\\ &\lesssim \Bigg|\bigcup_{|J|\le 3|I_{\mathbf{T}}|}E(J)\Bigg|^{1/2}\|\mathrm{Osc}_\delta\,\Gamma\|_2\\ &\lesssim \delta\sqrt{|I_{\mathbf{T}}|}\|\Gamma\|_2\\ &\lesssim \sigma\delta|I_{\mathbf{T}}|.\end{aligned}$$

7. Concluding remarks

Let us pose a conjecture concerning the kernel $J_H(y)\stackrel{\text{def}}{=}\mathbf{1}_{[-1,1]}y^{-1}$, that is the Hilbert transform kernel with a sharp cut off.

Conjecture 7.1. We have the inequality valid for all $n\ge 1$,

$$\|\mathrm{Osc}_n(J_H;f)\|_p\lesssim\|f\|_p,\quad 1<p<\infty.$$

In fact, the implied constant can be taken independent of n.

The proof as currently presented does not permit the deduction of this. Given the central role the Fourier transform plays in our proof, the technical difficulty we come to has a succinct description in terms of $\hat{J}_H$. Namely, the variation of $\hat{J}_H$ is infinite. But as the variation is only logarithmically infinite, one suspects that a proof of the conjecture above would have to revisit the proof of Carleson's theorem, with this example in mind.

The following is a corollary of Theorem 1.4.

Corollary 7.2. *For any measure-preserving system (X,μ,T) and $f\in L^p(X,\mu)$ for $1<p<\infty$, there is a set X_f of probability one for which for all $x\in X_f$,*

$$\lim_{N\to\infty}\sum_{0<|k|<N}\frac{e^{i\theta k}}{k}f(\mathrm{T}^k x)\quad\textit{exists for all }\theta.$$

The proof would begin by transferring the oscillation inequality (1.5), valid on $\mathbb{R}$, to the integers $\mathbb{Z}$. This kind of transference can be done directly; it is also possible that the necessary result follows from known transference results such as Auscher and Carro [3]. Details are left to the reader.

Likewise, the method of proof that we employ throughout the paper could be adapted to shed light on more general singular integrals, as well as the original Wiener–Wintner theorem. Indeed, an oscillation result could be proved for the latter theorem. We do not however pursue these lines here.

The Wiener–Wintner theorem has a deep extension to the return time theorem of Bourgain [4], see also the appendix to [5]. This theorem, which we do not recall in detail here, has certain extensions and variants that are currently only approachable via the phase plane methods of the type used in this paper. The return time is however a more sophisticated result, and the phase plane methods required are correspondingly more difficult. These issues will be explored in forthcoming papers of C. Demeter, M. Lacey, T. Tao, and C. Thiele. *Note added during revision of this paper*: The paper mentioned here is [10].

References

1. Assani, I., Caractérisation spectrale des systèmes dynamiques du type Wiener–Wintner, *C. R. Acad. Sci. Paris Sér. I Math.* **332** (2001), 321–324.
2. Assani, I., *Wiener Wintner Ergodic Theorems*, World Scientific, River Edge, NJ, 2003.
3. Auscher, P. and Carro, M. J., On relations between operators on $\mathbf{R}^N$, $\mathbf{T}^N$ and $\mathbf{Z}^N$, *Studia Math.* **101** (1992), 165–182.
4. Bourgain, J., Temps de retour pour les systèmes dynamiques, *C. R. Acad. Sci. Paris Sér. I Math.* **306** (1988), 483–485.
5. Bourgain, J., Pointwise ergodic theorems for arithmetic sets, *Inst. Hautes Études Sci. Publ. Math.* **69** (1989), 5–45.
6. Calderón, A., Ergodic theory and translation-invariant operators, *Proc. Nat. Acad. Sci. U.S.A.* **59** (1968), 349–353.
7. Campbell, J., Jones, R. L., Reinhold, K. and Wierdl, M., Oscillation and variation for the Hilbert transform, *Duke Math. J.* **105** (2000), 59–83.
8. Campbell, J. and Petersen, K., The spectral measure and Hilbert transform of a measure-preserving transformation, *Trans. Amer. Math. Soc.* **313** (1989), 121–129.
9. Carleson, L., On convergence and growth of partial sums of Fourier series, *Acta Math.* **116** (1966), 135–157.
10. Demeter, C., Lacey, M., Tao, T. and Thiele, C., Breaking the duality in the return times theorem, *Duke Math. J.* **143**(2) (2008), 281–355.
11. Grafakos, L., Tao, T. and Terwilleger, E., L^p bounds for a maximal dyadic sum operator, *Math. Z.* **246** (2004), 321–337.
12. Hunt, R. A., On the convergence of Fourier series, in *Orthogonal Expansions and their Continuous Analogues* (*Proc. Conf., Edwardsville, IL, 1967*), pp. 235–255, Southern Illinois Univ. Press, Carbondale, IL, 1968.

13. Jones, R. L., Kaufman, R., Rosenblatt, J. M. and Wierdl, M., Oscillation in ergodic theory, *Ergodic Theory Dynam. Systems* **18** (1998), 889–935.
14. Lacey, M., Carleson's theorem: proof, complements, variations, *Publ. Mat.* **48** (2004), 251–307.
15. Lacey, M. and Thiele, C., On Calderón's conjecture, *Ann. of Math.* **149** (1999), 475–496.
16. Lacey, M. and Thiele, C., A proof of boundedness of the Carleson operator, *Math. Res. Lett.* **7** (2000), 361–370.
17. Máté, A., Convergence of Fourier series of square integrable functions, *Mat. Lapok* **18** (1967), 195–242.
18. Muscalu, C., Tao, T. and Thiele, C., Multi-linear operators given by singular multipliers, *J. Amer. Math. Soc.* **15** (2002), 469–496.
19. Pramanik, M. and Terwilleger, E., A weak L^2 estimate for a maximal dyadic sum operator on $\mathbb{R}^n$, *Illinois J. Math.* **47** (2003), 775–813.
20. Prestini, E. and Sjölin, P., A Littlewood–Paley inequality for the Carleson operator, *J. Fourier Anal. Appl.* **6** (2000), 457–466.
21. Stein, E. M., *Singular Integrals and Differentiability Properties of Functions*, Princeton Mathematical Series **30**, Princeton University Press, Princeton, NJ, 1970.
22. Stein, E. M. and Weiss, G., An extension of a theorem of Marcinkiewicz and some of its applications, *J. Math. Mech.* **8** (1959), 263–284.
23. Wiener, N. and Wintner, A., On the ergodic dynamics of almost periodic systems, *Amer. J. Math.* **63** (1941), 794–824.

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

Erin Terwilleger
Department of Mathematics, U-3009
University of Connecticut
Storrs, CT 06269
U.S.A.
terwilleger@math.uconn.edu

Received 22 December 2005
in revised form 13 October 2007
published online June 5, 2008