Séminaire BOURBAKI Juin 2006
58ème année, 2005-2006, n$^{\text{o}}$ 962

MODULATION INVARIANT AND MULTILINEAR SINGULAR INTEGRAL OPERATORS
[after Lacey and Thiele]

by **Michael CHRIST**

Contents

INTRODUCTION

The bilinear Hilbert transform is the operator

$$BH(f_1,f_2)(x)=\int_{\mathbb{R}} f_1(x-t)f_2(x+t)\,t^{-1}dt \tag{1}$$

where $x,t\in\mathbb{R}$ and $f_j\in L^{p_j}(\mathbb{R})$. If t^{-1} were an integrable function then this integral would become absolutely convergent, for almost every x for appropriate exponents p_j. The question of the finiteness of the conditional integral, and of inequalities in L^p norms, was an open problem from roughly the mid-1960's to the late 1990's, when Michael Lacey and Christoph Thiele showed in a series of breakthrough papers that BH is well-defined and bounded on appropriate L^p spaces. This operator is prototypical for a class of multilinear operators with modulation symmetry, and their work has been followed by significant further developments too numerous to cite in the space available.

In this expository article I discuss the background and origins of the problem, outline the main lines of the analysis, and indicate the connection with the almost everywhere

Supported in part by NSF grant DMS-040126.

convergence of Fourier integrals. This article is not intended as an exhaustive survey, but merely as an introduction to the main ideas of the original articles [19],[20],[21].

1. HISTORICAL BACKGROUND

1.1. Singular integrals

The most fundamental example of a Calderón-Zygmund singular integral operator is the Hilbert transform $Hf(x) = \pi^{-1} \int_{\mathbb{R}} f(x-t)\, t^{-1}\, dt$ for $x \in \mathbb{R}$. The integral fails to converge absolutely in general, and is defined as the limit as $\varepsilon \to 0$ of the integral over $|t| > \varepsilon$.

H plays a fundamental role in the theory of convergence of the Fourier transform, as well as in one-dimensional complex analysis. It satisfies $\widehat{Hf}(\xi) = i\,\mathrm{sgn}(\xi)\widehat{f}(\xi)$ for all $\xi \neq 0$, where $\mathrm{sgn}(\xi) = \pm 1$ according to whether $\xi > 0$ or < 0. Thus $P = \frac{1}{2}(I - iH)$, where I is the identity, is the projection operator onto positive frequencies: $\widehat{Pf}(\xi) = \widehat{f}(\xi)\chi_{\xi>0}$. The "partial sum" operators $\widehat{P_N f}(\xi) = \widehat{f}(\xi)\chi_{|\xi|\le N}$ can be synthesized out of P together with shifts of the Fourier variable, in such a way that uniform boundedness of P_N on L^p is equivalent to boundedness of H on L^p. This is the basis of the classical theorem of M. Riesz on L^p norm convergence of Fourier series.

Somewhat more general CZ operators can be expressed as Fourier multiplier operators

$$\widehat{Tf}(\xi) = m(\xi)\widehat{f}(\xi) \qquad \text{where } m(r\xi) \equiv m(\xi) \text{ for all } r > 0 \tag{2}$$

and $m \in C^\infty(\mathbb{R}^d \setminus \{0\})$. General Fourier multipliers $\widehat{Tf}(\xi) = m(\xi)\widehat{f}(\xi)$ with $m \in L^\infty$ preserve $L^p(\mathbb{R})$ only for $p = 2$; there is no characterization of L^p functions in terms of the absolute values of their Fourier coefficients for $p \neq 2$.

The most general Calderón-Zygmund operators in $\mathbb{R}^d$ lack convolution structure, taking the form $\int_{\mathbb{R}^d} K(x,y)f(y)\,dy$ where

$$|K(x,y)| \le C|x-y|^{-d} \quad \text{and} \quad |\nabla_{x,y}K| \le C|x-y|^{-d-1}; \tag{3}$$

again I slur over the issue of interpretation of this typically absolutely divergent integral. Roughly speaking, (3) says that the portions of f, g microlocalized in phase space near (x,ξ) and (x',ξ') respectively interact quite weakly unless $|\xi| + |\xi'| \le C|x-x'|^{-1}$. According to the uncertainty principle, any stronger restriction of this general type is meaningless. A basic theorem [2] states that *if* such an operator is bounded on L^2, then it is also bounded on L^p for all $p \in (1,\infty)$.

The basic symmetries of this theory are translation and dilation; if $K(x,y)$ is a Calderón-Zygmund kernel then so are $K(x-z, y-z)$ and $r^d K(rx, ry)$, *uniformly* for all $r > 0$ and $z \in \mathbb{R}^d$. The individual operators need not exhibit these symmetries, but the class as a whole does.

A third basic symmetry, with respect to *modulation*, is totally lacking in this theory. Multiplying $K(x,y)$ by $e^{i(ax+by)}$ for any nonzero $(a,b) \in \mathbb{R}^2$ destroys the bound on ∇K. This lack of symmetry is perhaps even more apparent in (2), in the convolution case

$K(x-y)$, where $\xi = 0$ plays a privileged role. Of course, such a modulation does not affect L^p estimates, but as we will see, the bilinear Hilbert transform can be regarded as an infinite sum of modulated Calderón-Zygmund operators with different modulating frequencies, in such a way that boundedness of the sum cannot easily be inferred by summing bounds for the individual summands.

1.2. Calderón's commutator

Calderón had an abiding interest in partial differential equations with nonsmooth coefficients and on nonsmooth domains. He had employed algebras of singular integral operators in studying PDE, for instance in his work on uniqueness in the Cauchy problem [3]. Thus he was naturally led to investigate compositions of operators such as the canonical example H, the operator M_A of multiplication by a function A having limited smoothness, and $\frac{d}{dx}$. He showed in 1965 [4] that the commutator $[H, M_A]$ is smoothing, in the sense that $\mathcal{C}_A = \frac{d}{dx} \circ [H, M_A]$ is bounded on $L^2(\mathbb{R}^1)$, whenever A is Lipschitz continuous, that is, whenever $a = dA/dx \in L^\infty$. Formally

$$\mathcal{C}_A f(x) = \int_{\mathbb{R}} f(y) \frac{A(x) - A(y)}{(x-y)^2}\, dy, \tag{4}$$

which satisfies the Calderón-Zygmund assumptions (3) when $a = \frac{dA}{dx}$ belongs to L^∞. These operators possess translation and dilation invariance as a family, even though individually they lack it.

Since the commutator operator is not translation-invariant, Plancherel's theorem can not be invoked directly to establish its L^2 boundedness. A key realization of Calderón was that it could profitably be regarded as a *bilinear* operator, and that the full force of Fourier analysis and complex variables methods should be brought to bear on a.

An intriguing alternative expression is obtained by writing $A(x) - A(y) = (x - y) \int_0^1 a(sx + (1-s)y)\, ds$ to obtain a decomposition $\mathcal{C}_A(f) = \int_0^1 \mathcal{C}_s(f, a)\, ds$ where

$$\mathcal{C}_s(f, a)(x) = \int_{\mathbb{R}} f(x-t) a(x+st) t^{-1}\, dt. \tag{5}$$

Thus bounds for $\mathcal{C}_s$ from $L^2 \times L^\infty$ to L^2 would imply corresponding bounds for the commutator operator. The special case $\mathcal{C}_1$ is traditionally called the *bilinear Hilbert transform*, but all the operators $\mathcal{C}_s$ for $s \neq 0, -1$ have essentially the same intrinsic qualities and stature. Calderón asked[(1)]. whether these operators do map $L^2 \times L^\infty$ to L^2. The problem became notorious, but was not resolved until the work of Lacey and Thiele [19],[20] in the late 1990s.

Thought of as linear operators acting on f, $\mathcal{C}_s$ have nonsmooth kernels $K(x, y) = (x-y)^{-1} a(xs + (1-s)y)$ which satisfy no gradient estimate. Viewed as bilinear operators, they are singular in the sense that $\mathcal{C}_s(f, a)(x)$ depends on $a(y_1) f(y_2)$ only for (y_1, y_2) in a one-dimensional subset of $\mathbb{R}^2$.

[(1)]The question is widely attributed to Calderón, though I know of no reference.

It is remarkable that these building blocks $\mathcal{C}_s$ not only retain translation and dilation symmetry, but gain new *modulation* symmetries: defining $\mathcal{M}_\eta f(x) = e^{ix\eta} f(x)$,

$$\boxed{\mathcal{C}_s(\mathcal{M}_{s\eta} f, \mathcal{M}_\eta a) \equiv \mathcal{M}_{(1+s)\eta} \mathcal{C}_s(f, a).} \tag{6}$$

These are partial symmetries; there is no relation for $\mathcal{C}_s(\mathcal{M}_\eta f, \mathcal{M}_{\tilde{\eta}} a)$ unless $s\eta = \tilde{\eta}$.

In terms of the Fourier transform the operator is written

$$\mathcal{C}_s(f, a)(x) = c \iint e^{ix(\xi_1+\xi_2)} \operatorname{sgn}(s\xi_2 - \xi_1) \widehat{f}(\xi_1) \widehat{a}(\xi_2) \, d\xi_1 \, d\xi_2 \tag{7}$$

for a certain constant c, and the modulation symmetry is reflected in the invariance of the Fourier multiplier $\operatorname{sgn}(s\xi_2 - \xi_1)$ under $\xi \mapsto \xi + (s\eta, \eta)$. This multiplier is nonsmooth along an entire line, rather than merely at the origin.

It is (perhaps) a general principle that more symmetric operators are more difficult to analyze; a featureless wall presents no cracks which can naturally be enlarged into gaps. A fundamental point to look for in the discussion below is how the symmetry is broken; see §5.

1.3. Carleson's maximal operator

Carleson [6] proved in 1966 that for any periodic function $f \in L^2$ of one real variable, the partial sums of the Fourier series converge to f almost everywhere. The essentially equivalent statement for the real line is that $(2\pi)^{-1} \int_{|\xi| \le N} \widehat{f}(\xi) e^{ix\xi} \, d\xi$ converges to $f(x)$ as $N \to \infty$, for almost every $x \in \mathbb{R}$. The main ingredient is an estimate for Carleson's maximal operator $\mathcal{C}^\star f(x) = \sup_{N < \infty} \big| \int_{|\xi| \le N} \widehat{f}(\xi) e^{ix\xi} \, d\xi \big|$, which is essentially the same as

$$\mathcal{C}^* f(x) = \sup_{N \in \mathbb{R}} \Big| \int_{\mathbb{R}} f(x-t) e^{iNt} t^{-1} \, dt \Big|. \tag{8}$$

Carleson proved that $\mathcal{C}^*$ maps L^2 to weak L^2, that is, $|\{x : \mathcal{C}^* f(x) > \lambda\}| \le C\lambda^{-2} \|f\|_{L^2}^2$ uniformly for all $\lambda > 0$ and $f \in L^2$. Almost everywhere convergence follows immediately from this inequality since it holds trivially for functions whose Fourier transforms have compact support.

It is equivalent to establish bounds for the linear operators $\int_{\mathbb{R}} f(x-t) e^{iN(x)t} t^{-1} \, dt$ which are *uniform* over all measurable real-valued selection functions N. Once again these operators enjoy forms of translation, dilation, and modulation invariance. For instance, $\mathcal{L}(\mathcal{M}_\eta f)(x) = \mathcal{M}_\eta \mathcal{L}' f(x)$, where $\mathcal{L}'$ is obtained from $\mathcal{L}$ by replacing the function $N(x)$ by $N(x) - \eta$.

Fefferman [15] later gave a second proof of Carleson's theorem. Lacey and Thiele used elements of both of these analyses to prove[(2)]

THEOREM 1.1. — *Let $p_1, p_2, q \in (1, \infty]$ satisfy $q^{-1} = p_1^{-1} + p_2^{-1}$, and assume that no more than one of these exponents is infinite. Then there exists $C < \infty$ such that $\|BH(f_1, f_2)\|_{L^q} \le C \|f_1\|_{L^{p_1}} \|f_2\|_{L^{p_2}}$ for all Schwartz class functions.*

[(2)] Their theorem actually applies for all $q > \frac{2}{3}$.

1.4. Two roads diverge

Calderón proved the bound he sought for the commutator operator without understanding the bilinear Hilbert transform, and went on to analyze [5] the Cauchy integral associated to Lipschitz curves with small Lipschitz constant by an extension of those ideas. Further developments have included a vast literature on elliptic boundary problems on Lipschitz domains, analytic capacity in one complex variable [25], the work of Coifman-Meyer-Mcintosh [12] on the Cauchy integral, and the $T(1)$ theorem of David and Journé [13]. A theory of multilinear Calderón-Zygmund singular operators was developed [9][10], which however does not include $\mathcal{C}_s$; it encompasses operators which have a Fourier representation like (7) with $\operatorname{sgn}(s\xi_2 - \xi_1)$ replaced by functions smooth away from $\xi = 0$ and satisfying $m(r\xi) \equiv m(\xi)$ for $r > 0$. These operators lack modulation invariance, and are less singular. Some of that theory provides essential building blocks for the analysis outlined here.

2. LOCALIZED FOURIER COEFFICIENTS

2.1. A frame with a preferred scale

Let $\psi : \mathbb{R}^1 \to \mathbb{C}$ be an infinitely differentiable function supported in $(0, 2)$ such that $\sum_{n\in\mathbb{Z}} \psi(t - n) \equiv 1$ for all $t \in \mathbb{R}$. Then the set of all functions $\{\psi_{k,n} = e^{ikt}\psi(t - n) : k, n \in \mathbb{Z}\}$ is a *frame* for $L^2(\mathbb{R}^1)$; for any $f \in L^2$,

$$f = c \sum_{k,n} \psi_{k,n} \langle f, \psi_{k,n} \rangle \tag{9}$$

for a certain constant c whose precise value is of no consequence for the type of inequality in question here. The inverse Fourier transform of ψ is a Schwartz function, and multiplying it by $(2\pi)^{-1/2}$ yields a function φ such that $\{\varphi_{k,n}(x) = e^{inx}\varphi(x - k)\}$ is likewise a frame for L^2. It is good intuition to think of $\varphi_{k,n}(x)$ as being essentially $c_0 e^{inx} e^{-|x-k|^2}$, although this is not quite correct because these functions lack compactly supported Fourier transforms.

One thinks of $\langle f, \varphi_{k,n} \rangle$ as being *localized Fourier coefficients*. Such a frame is quite different from celebrated *wavelet bases*. The lesser difference is that $\{\varphi_{k,n}\}$ is not an orthonormal system; there is some oversampling here. The significant difference is that whereas a wavelet basis treats all scales equally, this frame prefers one scale. An advantage of this frame, not shared by wavelet-type bases, is its invariance under modulation by integral frequencies.

The rank one operator $f \mapsto \langle f, \varphi_{k,n} \rangle \varphi_{k,n}$ heuristically represents the orthogonal projection of $L^2(\mathbb{R})$ onto the subspace consisting of all functions g such that g is supported in $I = [k, k + 1]$ and $\widehat{g}$ is supported in $\omega = [n, n + 1]$, although this is not exactly true. The entire *phase space* $\mathbb{R} \times \mathbb{R}$ is tiled by these sets $I \times \omega$, and corresponding to this geometric decomposition is the analytic decomposition of the identity operator as a sum of projections.

2.2. Scaled frames and tiles in phase space

Other frames can be constructed by scaling: For each integer r we form $\{2^{-r/2}\varphi_{k,n}(2^{-r}x) : k, n \in \mathbb{Z}\}$, which is likewise a frame. This corresponds to tiling $\mathbb{R}^2$ by rectangles $I \times \omega$ where I has length 2^r and ω has length 2^{-r}.

A *dyadic interval* is a closed bounded subinterval of $\mathbb{R}^1$ of the form $[k2^n, (k+1)2^n]$ for arbitrary integers k, n. The set of all dyadic intervals enjoys an often useful combinatorial property: If two such intervals do overlap, then one is contained in the other.

DEFINITION 2.1. — *A* tile *is a subset of the phase space* $\mathbb{R}^2$ *of the form* $I \times \omega$*, where* I, ω *are arbitrary dyadic intervals satisfying* $|I| \cdot |\omega| = 1$.

In contrast to dyadic intervals, no tile is properly contained in another. Tiles and dyadic intervals are said to be *nonoverlapping* if their interiors are disjoint.(3)

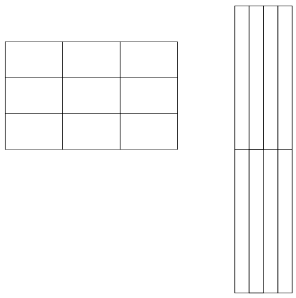

FIGURE 1. Two tilings of phase space. According to the uncertainty principle, no significant refinement of either tiling is meaningful.

2.3. Decomposition of the bilinear Hilbert transform by scales

The bilinear Hilbert transform has no preferred scale; with respect to the operators $D_\lambda f(x) = f(\lambda x)$, there is the dilation symmetry $BH(D_\lambda f, D_\lambda g) = D_\lambda(BH(f,g))$. Thus none of our frames is well adapted to the operator.

Let $\eta : \mathbb{R} \to \mathbb{C}$ be a smooth function supported(4) in $(8, 32)$ such that $\sum_{r\in\mathbb{Z}} \eta(2^r\xi) \equiv 1$ for all $\xi > 0$. BH is thus decomposed as $\sum_{r\in\mathbb{Z}} BH_r$ where

$$BH_r(f_1, f_2)(x) = c \int_{\mathbb{R}^2} e^{ix(\xi_1+\xi_2)} \operatorname{sgn}(\xi_1 - \xi_2)\eta(2^r(\xi_1 - \xi_2))\widehat{f_1}(\xi_1)\widehat{f_2}(\xi_2)\, d\xi_1\, d\xi_2, \tag{10}$$

plus a second, very similar, infinite sum obtained by replacing $\eta(s)$ by $\eta(-s)$.

There is an alternative expression

$$BH_r(f_1, f_2)(x) = \int_{\mathbb{R}} f_1(x+t)f_2(x-t)2^{-r}h(2^{-r}t)\, dt \tag{11}$$

(3) I will sometimes abuse language and notation by saying that intervals are disjoint when they merely do not overlap.

(4) The precise numbers $8, 32$ are of no significance; one could replace these by any A, B such that $B > 2A$ and $A > 0$ is sufficiently large.

for a certain Schwartz function h. An immediate consequence of Hölder's inequality is that BH_r maps $L^{p_1} \times L^{p_2}$ to L^q whenever p_1, p_2, q are ≥ 1 and satisfy $q^{-1} = p_1^{-1} + p_2^{-1}$. and the scaling symmetry ensures that the operator norm is independent of r. The sole issue is the summation over r.

BH_r clearly has a preferred spatial scale, 2^r, and retains the modulation invariance of BH. For each $r \in \mathbb{Z}$, BH_r will be analyzed in terms of the frame described above with the same parameter r. Thus one works simultaneously with infinitely many frames, one for each scale.

2.4. Phase space decomposition of the bilinear Hilbert transform

For each $r \in \mathbb{Z}$ denote by $\mathcal{P}_r$ the set of all phase space tiles $P = I \times \omega$ of dimensions $2^r \times 2^{-r}$. Consider the trilinear form $\mathcal{H}_r(f_1, f_2, f_3) = \langle f_3, BH_r(f_{,}f_2)\rangle$ for an arbitrary test function f_3. Then $\langle BH(f_1, f_2), f_3\rangle = \sum_{r=-\infty}^{\infty} \mathcal{H}_r(f_1, f_2, f_3)$.

Decompose each $f_k = \sum_{P^k \in \mathcal{P}_r} \langle f_k, \phi_{P^k}\rangle \phi_{P^k}$ to obtain

$$\mathcal{H}_r(f_1, f_2, f_3) = \sum_{P=(P^1,P^2,P^3)\in\mathcal{P}_r^3} \beta_P \prod_{k=1}^{3} \langle f_k, \phi_{P^k}\rangle$$

for certain coefficients β_P independent of $\{f_k\}$. Upper bounds for the "interaction amplitudes" β_P are required, and elementary estimates combined with the information that η is supported in $(8, 32)$ give:

Lemma 2.2. — *Let $P = (P^1, P^2, P^3)$ where each $P^k = I_k \times \omega_k$ is a phase space tile of dimensions $2^r \times 2^{-r}$. Then for any finite exponent N,*

$$|\beta_P| \leq C_N 2^{-r/2} \Big(\max_{k,l\in\{1,2,3\}} (1 + 2^{-r} \operatorname{distance}(I_k, I_l)) \Big)^{-N} \tag{12}$$

and

$$\beta_P \equiv 0 \tag{13}$$

unless the centerpoints c_k of the frequency space intervals ω_k satisfy $c_2 - c_1 \in [2\cdot 2^{-r}, 38\cdot 2^{-r}]$ and $c_3 - 2c_1 \in [10 \cdot 2^{-r}, 70 \cdot 2^{-r}]$.

This is a reflection of a fundamental characteristic of the bilinear Hilbert transform: $BH(e^{i\xi_1\cdot}, e^{i\xi_2\cdot})(x) = \pi i \operatorname{sgn}(\xi_1 - \xi_2) e^{i\xi_3 x}$ where $\boxed{\xi_3 = \xi_1 + \xi_2}$.

The numbers $2, 38, 10, 70$ are insignificant artifacts of certain nearly arbitrary choices. What is important is the following consequence, which will be a source of orthogonality in the analysis.

Fact 2.3. — *Let $P, Q \in \cup_r \mathcal{P}_r^3$ and suppose that β_P, β_Q are both nonzero. Suppose that for some index $k \in \{1, 2, 3\}$, $\omega_{P^k}, \omega_{Q^k}$ overlap, and that $|\omega_{P^k}| < |\omega_{Q^k}|$. Then for each $i \neq k$, ω_{P^i} is disjoint from ω_{Q^i}, and they are separated by a distance comparable to $|\omega_{Q^i}|$.*

2.5. Nuisance technicalities involving tiles

If I, J are dyadic intervals, then either one is contained in the other, or they do not overlap. This makes dyadic intervals well suited to stopping time arguments, in which one begins with such an interval, subjects it to a test, and if it fails, subdivides it into halves and subjects the two halves separately to (rescaled versions of) the same test, repeating indefinitely.

On the other hand, the set of all dyadic intervals has no reasonable translation invariance. Consider the interval $[0,1]$, which for any integer $N \geq 1$ is contained in the larger dyadic interval $[0, 2^N]$. These larger intervals have the unnatural feature that they extend only to the right of $[0,1]$, never to the left; their union is only half of the real axis. Thus analysis based on these intervals is likely to disregard interactions between the two halves of the real axis. This defect is essentially reprised at every dyadic point $j2^n$, $j, n \in \mathbb{Z}$.

This difficulty arises commonly and has been sidestepped in various ways by various authors; see [6], [16], [24]. It is also helpful to thin out the sum by partitioning the set of all dyadic intervals into finitely many subfamilies, so that for any two intervals I, J belonging to any common subfamily, if $I \subset J$ and $I \neq J$ then $|J| \geq 2^K |I|$ where K is a large constant. This leads to a decomposition into finitely many suboperators, all having the same structure.

In this exposé I will systematically slur over these technicalities, which are of no intrinsic interest. I do not pretend to give a full proof, only a conceptually accurate outline. Statements made below are correct, but under the proviso that these technicalities have been dealt with.

3. ALMOST-ORTHOGONALITY

3.1. Introduction

The space L^2 plays a special role in the classical singular integral operator theory, partly because methods relying on Hilbert space structure are available. In particular, Plancherel's theorem can be applied to easily establish L^2 estimates for translation-invariant operators; but it is not directly applicable to the commutator operator or its more degenerate relatives.

A rather flexible almost-orthogonality principle was introduced by Knapp and Stein [18], who were motivated by problems in the representation theory of semisimple groups to establish L^2 bounds for singular integral operators invariant with respect to certain (nilpotent) Lie group structures. They showed that if an abstract operator T is decomposed as a sum of bounded operators T_j such that $\|T_i T_j^*\|, \|T_i^* T_j\| \leq c_{i-j}$ for all $i, j \in \mathbb{Z}$, then $\|T\| \leq \sum_j c_j^{1/2}$. A trivial case is when the summands have pairwise orthogonal ranges, and likewise for their adjoints.

A satisfying explanation of the L^2 boundedness of Calderón-Zygmund operators was finally obtained by David and Journé, who showed that any such (bounded) operator can be decomposed as a sum of three parts, one of which has a natural almost-orthogonal decomposition in the sense of Knapp and Stein. The other two parts have a different structure, related to the concepts of Carleson measures and paraproducts [10]. It is remarkable that on this level the ultimate understanding of the fundamental L^2 estimate rests on the theory surrounding BMO, which is one limit of L^p as $p \to \infty$.

3.2. Orthogonality via phase space disjointness

If two functions f, g have disjoint supports, then of course $\langle f, g\rangle = 0$. The same goes if $\widehat{f}, \widehat{g}$ have disjoint supports. If P, Q are nonoverlapping tiles, then either ω_P and ω_Q are nonoverlapping, or I_P and I_Q are. In the former case, $\langle \phi_P, \phi_Q\rangle = 0$, but in the latter case the two supports can't be disjoint (both functions are real analytic). Nonetheless, since $\phi_P(x)$ decays rapidly as x moves away from I_P, $\langle \phi_P, \phi_Q\rangle$ is relatively small if I_P, I_Q are far apart. Therefore one hopes to retain some form of orthogonality. The following lemma(5) is analogous to Bessel's inequality for Fourier coefficients, but is slightly weakened by the necessity of a supplementary hypothesis(6) (14).

LEMMA 3.1. — *Let $\mathcal{S}$ be any set of pairwise nonoverlapping tiles. Let $f \in L^2$ and $\lambda > 0$. Suppose that for every $P \in \mathcal{S}$,*

$$\lambda |I_P|^{1/2} \le |\langle f, \phi_P\rangle| \le 2\lambda |I_P|^{1/2}. \tag{14}$$

Then $\boxed{\sum_{P\in\mathcal{S}} |\langle f, \phi_P\rangle|^2 \le C\|f\|_{L^2}^2.}$

The constant C is independent of $f, \lambda, \mathcal{S}$. An equivalent statement of the conclusion is $\sum_{P\in\mathcal{S}} |I_P| \le C\lambda^{-2}\|f\|_{L^2}^2$; $\sum_P |I_P|$ is a weighted count of the number of tiles satisfying (14). For a complete proof of this fundamental fact see §8.

A localized variant is often useful: Suppose that there is given an interval J such that $I_P \subset J$ for all $P \in \mathcal{S}$. Then under hypothesis (14),

$$\sum_{P\in\mathcal{S}} |\langle f, \phi_P\rangle|^2 \le C_M \int_{\mathbb{R}} |f(x)|^2 \left(1 + \frac{\text{distance}\,(x, J)}{|J|}\right)^{-M} dx \tag{15}$$

for any finite M. This is natural, since $|\phi_P(x)|$ decays rapidly as x moves away from J.

(5)The original proofs of Lacey and Thiele [19],[20] were more complicated than the one outlined here, in large part because they had only a weaker version of Lemma 3.1 at their disposal. Lemma 3.1, in a slightly more sophisticated form, is in their paper [21].
(6)Without a supplementary hypothesis of some kind, Lemma 3.1 is false.

4. TREES, TOWERS, AND MULTITREES

The set of all tiles is endowed with a partial ordering.

DEFINITION 4.1. — *$P \le Q$ if and only if $I_P \subset I_Q$ and $\omega_P \supset \omega_Q$. Also $P < Q$ means that $P \le Q$ and $P \ne Q$.*

DEFINITION 4.2. — *A* multitile *P is an ordered* 3*-tuple of tiles (P^1, P^2, P^3) such that $I_{P^i} = I_{P^j}$ for all indices i, j, and the centers of the associated frequency intervals ω_{P^j} satisfy the constraints listed following* (13)*.*

This common interval I_{P^k} is denoted by I_P. To any multitile are associated three functions ϕ_{P^k}, $k \in \{1, 2, 3\}$. For any single P, the associated three rectangles $I_P \times \omega_P^k$ all share the same dimensions.

4.1. Model operator

A simplified model for the sum that represents the bilinear Hilbert transform is

$$\mathcal{H}_{\text{model}}(f_1, f_2, f_3) = \sum_P |I_P|^{-1/2} \prod_{k=1}^{3} \left|\langle f_k, \phi_{P^k}\rangle\right| \tag{16}$$

where the sum ranges over all multitiles, or some large collection of multitiles. These terms are normalized so that $|I_P|^{-1/2} \prod_k \left|\langle f_k, \phi_{P^k}\rangle\right| \le C \prod_k \|f_k\|_{L^{p_k}}$ whenever $\sum_k p_k^{-1} = 1$, with C independent of P. The theorem asserts that this sum of infinitely many uniformly bounded operators is bounded.

The condition (12) that β_P is small whenever I_{P^k}, I_{P^i} are far apart has been simplified in this model to the condition that $\beta_P = 0$ whenever $I_{P^k} \ne I_{P^i}$. Only this model will be discussed further; but see [19] for an explanation of how the bilinear Hilbert transform can actually be realized as a limit of averages of such model operators.

One should regard each factor $\langle f_k, \phi_{P^k}\rangle$ as expressing one irreducible bit of information, and each term in the sum as being likewise irreducible. This minimality would be lost if a single frame for $L^2(\mathbb{R})$ were used.

4.2. Trees, towers, and multitrees

The analysis proceeds by decomposing the full sum (16) over all multitiles into sums over various subfamilies enjoying additional structure.

DEFINITION 4.3. — *A* tower *is a nonempty finite set T of tiles such that there exists a tile $\mathbf{top}(T) \in T$, called the top of T, such that every $P \in T$ satisfies $P \le \mathbf{top}(T)$, that is, $I_P \subset I_{\mathbf{top}(T)}$ and $\omega_P \supset \omega_{\mathbf{top}(T)}$.*

DEFINITION 4.4. — *A* tree *T is a finite set of tiles with an element $\mathbf{top}(t) \in T$, such that for any $P \ne Q \in T$: (i) $I_P \subset I_{\mathbf{top}(T)}$, (ii) P, Q do not overlap, (iii) if $|I_P| \ne |I_Q|$ then $\omega_P \cap \omega_Q = \emptyset$. Moreover, (iv) there exists $\xi_0 \in \mathbb{R}$ such that* distance $(\omega_P, \xi_0) \sim |\omega_P|$ *for all $P \in T$.*

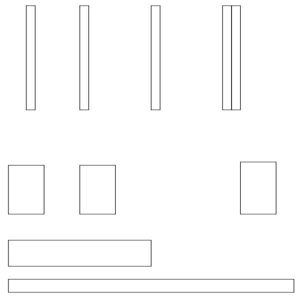

FIGURE 2. A tree. The x axis is horizontal; the ξ axis is vertical. Towers are similar, except that there exists some horizontal line which crosses every rectangle.

FIGURE 3. All intervals I_P associated to tiles $P \in T$, where T is either a tree or a tower. These intervals are contained in $\mathbb{R}$, but each is displaced vertically for purposes of illustration. $I_{\mathbf{top}(T)}$ is the topmost interval. The corresponding frequency space intervals ω_P have lengths inversely proportional to $|I_P|$. These form a nested set if T is a tower, whereas any two of different lengths are nonoverlapping if T is a tree.

DEFINITION 4.5. — *Let $k \in \{1,2,3\}$. A finite set T of multitiles is said to be a k-*tower *if $T^k = \{P^k : P \in T\}$ is a tower. A* multitree *is a set of multitiles which is a k-tower for some index k.*

A k-tower has a unique topmost element $\mathbf{top}(T)$, such that $\mathbf{top}(T)^k$ is the top of the tower T^k.

4.3. One origin of orthogonality

Plancherel's and Parseval's theorems and Bessel's inequality lie at the heart of many more sophisticated arguments in classical Fourier analysis. Here a fundamental role is played by inequalities based on orthogonality, reflecting the principle that ϕ_P, ϕ_Q are nearly orthogonal when the subsets P, Q of phase space do not overlap. We now discuss one property of the set of all multitiles, which leads to families of nonoverlapping tiles in the analysis of multilinear singular integrals.

Towers T are sets of tiles which totally lack useful almost-orthogonality, since $\omega_P \cap \omega_Q \neq \emptyset$ for any $P, Q \in T$. In the case where $0 \in \omega_{\mathbf{top}(T)}$, the rank one operators $f \mapsto \langle f, \phi_Q \rangle \phi_Q$ for $Q \in T$ is are closely analogous to averaging operators $f \mapsto$

$(|I_Q|^{-1}\int_{I_Q} f)\cdot\chi_{I_Q}$ (and more generally to frequency-modulated generalizations $f\mapsto e^{ix\xi_0}(|I_Q|^{-1}\int_{I_Q} e^{-ix\xi_0}f)\cdot\chi_{I_Q}(x)$ for $\xi_0\in\omega_{\mathbf{top}(T)}$). A tower could include a large family of Q such that $|I_Q|\to 0$ and I_Q approaches some point x_0. For almost all such x_0, $|I_Q|^{-1}\int_{I_Q} f\to f(x_0)$ by Lebesgue's differentiation theorem. Thus for a tower, no upper bound may be available for the coefficients $|I_Q|^{-1/2}\langle f,\phi_Q\rangle$ beyond their uniform boundedness.

However, in the context of multitiles, there is substantial compensation for this lack of orthogonality.

Fact 4.6. — *If a family T of multitiles is an i-tower for some $i\in\{1,2,3\}$ then* $\boxed{\{P^j : P\in T\} \textit{ is a tree}}$ *for any $j\neq i$ in $\{1,2,3\}$.*

This crucial consequence of Fact 2.3 partly explains the terminology "multitree".

Expressions $(\sum_{Q\in T}|\langle f,\phi_Q\rangle|^2)^{1/2}$, where T is a tree, are closely analogous to classical Littlewood-Paley expressions and are central to the analysis. An important part of this analogy is the next bound.

Lemma 4.7. — *Uniformly for all trees T and all functions f,*

$$\sum_{P\in T}|\langle f,\phi_P\rangle|^2\le C\min\left(\|f\|_{L^2}^2,\ |I_{\mathbf{top}(T)}|\ \|f\|_{L^\infty}^2\right). \tag{17}$$

These inequalities lend Fact 4.6 great significance. The second conclusion is a localized version of the first.

The basic principle underlying Lemma 4.7 is that since the tiles $P\in T$ form a nearly disjoint family of subsets of phase space, $\{\phi_P : P\in T\}$ is an almost-orthogonal family of functions. The lemma is not quite a consequence of Lemma 3.1 on almost-orthogonality via nonoverlapping tiles, but follows easily from a direct examination of the magnitudes of matrix coefficients $\langle\phi_P,\phi_Q\rangle$ for $P,Q\in T$.

4.4. Energy and mass of multitrees

The j-energy $\mathcal{E}_j(T)$ and j-mass $\mathcal{M}_j(T)$ of a multitree T are defined for $j\in\{1,2,3\}$ to be

$$\mathcal{E}_j(T)=|I_{\mathbf{top}(T)}|^{-1/2}(\sum_{P\in T}|\langle f_j,\phi_{P^j}\rangle|^2)^{1/2} \tag{18}$$

$$\mathcal{M}_j(T)=\sup_{P\in T}|I_P|^{-1/2}|\langle f_j,\phi_{P^j}\rangle|. \tag{19}$$

The individual terms are normalized so that $|I_P|^{-1/2}|\langle f_j,\phi_{P^j}\rangle|\le CMf_j(x)$ for $x\in I_P$, and if $\{P^j : P\in T\}$ is a tree then $\mathcal{E}_j(T)\le C\|f_j\|_{L^\infty}$ (by Lemma 4.7).

If i,j,k are the three elements of $\{1,2,3\}$ in any order then

$$|\mathcal{H}_T(f_1,f_2,f_3)|\le\mathcal{E}_j(T)\mathcal{E}_k(T)\mathcal{M}_i(T)|I_{\mathbf{top}(T)}|. \tag{20}$$

(20) is a direct consequence of definitions via Cauchy-Schwarz, and should be regarded as a manifesto of intent rather than as a genuine estimate. Proposition 7.2 below provides alternative upper bounds for $\mathcal{H}_T$, not subsumed in the mass-energy bound.

5. ORGANIZING THE TOTAL SUM INTO SUBSUMS

The stage has been set for a discussion, in this section and the next, of the heart of the proof: a sorting algorithm, a counting problem, and a counting estimate based on phase space orthogonality. To any finite set $\mathcal{S}$ of multitiles and any three functions $f_j : \mathbb{R}^1 \to \mathbb{C}$ is associated the operator expression

$$\mathcal{H}_{\mathcal{S}}(f_1, f_2, f_3) = \sum_{P \in \mathcal{S}} |I_P|^{-1/2} \prod_{j=1}^{3} |\langle f_j, \phi_{P^j} \rangle|. \tag{21}$$

We will partition the set of all multitiles into subsets having special structure, derive a reasonable bound for the contribution of each subset, and sum those bounds.

5.1. The sorting algorithm's output

In the proof we examine the sum $\mathcal{H}_{\mathcal{S}}$ of contributions of an arbitrary finite set of multitiles $\mathcal{S}$. The sorting procedure detailed below partitions the collection $\mathcal{S}$ into multitrees. It constructs families $\mathcal{F}_{n,i,j}$ of multitrees, indexed by $n \in \mathbb{Z}$ and $i, j \in \{1, 2, 3\}$, with the following main properties:

(i) $\mathcal{S}$ is the disjoint union, over all n and all ordered pairs (i, j), of all $T \in \mathcal{F}_{n,i,j}$.

(ii) Each $T \in \mathcal{F}_{n,i,j}$ is an i-tower.

(iii) For any $T \in \mathcal{F}_{n,i,j}$,

$$|\mathcal{H}_T(f_1, f_2, f_3)| \le C 2^{3n} |I_{\mathbf{top}(T)}|. \tag{22}$$

(iv) If $j = i$ then

$$|I_{\mathbf{top}(T)}| \le 2^{-2n} |\langle f_i, \phi_{\mathbf{top}(T)^i} \rangle|^2. \tag{23}$$

(v) If $j \ne i$ then

$$|I_{\mathbf{top}(T)}| \le 2^{-2n} \sum_{P \in T} |\langle f_j, \phi_{P^j} \rangle|^2. \tag{24}$$

(vi) Each $T \in \mathcal{F}_{n,i,j}$ enjoys certain maximality properties.

5.2. The sorting algorithm

The algorithm seeks out the enemy in the form of subsums corresponding to i-towers; such subsums are potentially unfavorable because of the lack of orthogonality discussed above. We proceed to construct families $\mathcal{F}_{n,i,j}$ of multitrees $T \subset \mathcal{S}$, for all $n \in \mathbb{Z}$ and all $(i,j) \in \{1,2,3\}^2$. The construction proceeds by descending induction on n. Begin with a very large positive n. Order the 9 ordered pairs of indices $(i,j) \in \{1,2,3\}$ arbitrarily.

Fix a small constant $c_0 > 0$. For the first pair (i,j), if $i = j$ then consider all i-towers $T \subset \mathcal{S}$ such that
(i) $|I_P|^{-1/2}|\langle f_i, \phi_{P^i}\rangle| \geq c_0 2^n$ for every $P \in T$
(ii) T is maximal with respect to set inclusion. That is, there exists no i-tower $T' \subset S$ properly containing T which satisfies (i).

If there exists a nonempty i-tower $T \subset \mathcal{S}$ satisfying (i), then there also exists one satisfying both criteria. Choose any one, T, put it into $\mathcal{F}_{n,i,j}$, and delete all tiles $P \in T$ from $\mathcal{S}$. Repeat the procedure with this reduced set $\mathcal{S}$ for the same pair (i,j) until no multitrees satisfying (i) remain. Then move on to the next pair (i,j). Observe how modulation symmetry is broken; the decomposition depends on the the localized Fourier coefficients $\langle f_i, \phi_{P^i}\rangle$ of f_i.

If $j \neq i$ then do the same, retaining (ii) but replacing (i) by
(i)* $\mathcal{E}_j(T) \geq 2^n$
and imposing a supplementary condition (iii)* whose role will not be visible in this exposé.[7]

Continue with a given index n until all 9 pairs (i,j) have been fully examined, and no multitrees satisfying the criteria remain. Then replace n by $n-1$, and repeat the selection again, with n replaced by $n-1$ in criteria (i) and (i)*. Continue by descending induction on n until only tiles satisfying $\langle f_k, \phi_{P^k}\rangle = 0$ for all $k \in \{1,2,3\}$ remain. Those contribute nothing to the operator, and may be discarded.

Given a finite set $\mathcal{S}$, there exists n_0 so large that (i), (i)* cannot possibly hold with factors of 2^{n_0}. The induction begins with such an n_0.

(23) and (24) follow directly from (i), (i)*. Moreover $\mathcal{M}_k(T) \leq c_0 2^{n+1}$ and $\mathcal{E}_k(T) \leq 2^{n+1}$ for all $k \in \{1,2,3\}$, because all tiles of T were available throughout the selection of $\mathcal{F}_{n,i,k}$, yet T was not selected. Therefore T did not satisfy the selection critera (i), (i)* at stage $n+1$. (22) follows by Cauchy-Schwarz.

This purely formal discussion applies to any expression $\sum_{P\in\mathcal{S}} |I_P|^{-1/2} \prod_{k=1}^3 |a_k(P)|$; the actual meaning of the coeffients $a_k = \langle f_k, \phi_{P^k}\rangle$ has not yet been exploited.

[7]Criterion (iii)*: If $i < j$, then a candidate i-tower T is not selected if there is some other candidate multitree T' such that ω_T^i lies strictly to the left of $\omega_{T'}^i$. (An interval $[a,b]$ is said to lie strictly to the left of $[c,d]$ if $b \leq c$.) The same goes if $i > j$, with "left" replaced by "right". The reader is urged to disregard this point for the present. Likewise the smallness of the constant c_0 in (i) plays a role which will not be apparent at the level of detail of this exposé.

6. COUNTING MULTITREES

The total sum is the sum of the contributions of all T, and we have a bound[8] of $2^{3n}|I_{\mathbf{top}}(t)|$ for each T, so it would suffice to have a suitable bound for

$$\boxed{\textit{the weighted count } \sum_T |I_{\mathbf{top}(T)}| \textit{ of the number of multitrees}}$$

in each family $\mathcal{F}_{n,i,j}$. Here an estimate with genuine content must finally be established.

Since an upper bound for $\sum_T |I_{\mathbf{top}(T)}|$ is required, it is advantageous to place as many tiles as possible into each multitree, consistent with the selection criteria (i), (i)*. This motivates the maximality criterion (ii).

6.1. A small reduction

For each $j \in \{1,2,3\}$ let $E_j \subset \mathbb{R}$ be an arbitrary measurable set with $|E_j| < \infty$. Let f_j be any function which is $\equiv 0$ on $\mathbb{R} \setminus E_j$, and $|f_j(x)| \leq 1$ for all $x \in E_j$. The theorem is a consequence of the following inequality via a simple interpolation argument.[9]

PROPOSITION 6.1. — *There exists $C < \infty$ such that for any finite set $\mathcal{S}$ of multitiles, sets E_j, and functions f_j,*

$$\sum_{P \in \mathcal{S}} |I_P|^{-1/2} \prod_{k=1}^{3} |\langle f_k, \phi_{P^k} \rangle| \leq C \prod_{k=1}^{3} |E_k|^{1/p_k} \tag{25}$$

for any exponents $p_k \in (1,\infty)$ satisfying $\sum_k p_k^{-1} = 1$.

6.2. Counting multitrees

Here is the crux of the entire analysis.

LEMMA 6.2 (Counting multitrees). — *For any $n \in \mathbb{Z}$ and $i,j \in \{1,2,3\}$, for any measurable sets E_1, E_2, E_3, the multitrees selected by the sorting algorithm satisfy*

$$\sum_{T \in \mathcal{F}_{n,i,j}} |I_{\mathbf{top}(T)}| \leq C 2^{-2n} |E_j|. \tag{26}$$

We'll discuss only the case $j = i$ in detail; see §7.1 for a brief discussion of the case $j \neq i$. The proof combines structural information built into the sorting algorithm with almost-orthogonality of the rank one operators $f \mapsto \langle f, \phi_{P^k} \rangle \phi_{P^k}$ associated to collections of nonoverlapping (scalar) tiles P. This structural information is:

FACT 6.3. — *The tiles $\{\mathbf{top}(T)^i : T \in \mathcal{F}_{n,i,i}\}$ do not overlap.*

---

[8] An important alternative bound is also available; see Proposition 7.2.

[9] Any function can be decomposed as $f = \sum_{m=-\infty}^{\infty} 2^m f_m$ where the sets f_m have pairwise disjoint supports and satisfy $\|f_m\|_{L^\infty} \leq 1$. Interpolation amounts to making this substitution for f_k for each $k \in \{1,2,3\}$ to produce an infinite sum of expressions of the type controlled by (25), and summing the resulting bounds.

Proof. — If not, there exist two distinct i-towers $T, T' \in \mathcal{F}_{n,i,i}$ satisfying $\mathbf{top}(T)^i < \mathbf{top}(T')^i$. If T' was chosen before T, we reach a contradiction because $T' \cup \{\mathbf{top}(T)\}$ is an i-tower which properly contains T', contradicting the maximality criterion (ii) for T'. If T was chosen before T' then (ii) is again contradicted, since $T \cup \{\mathbf{top}(T')\}$ is an i-tower (whose top is $\mathbf{top}(T')$), which properly contains T. □

By the sorting algorithm, $|I_{\mathbf{top}(T)}| \le 2^{-2n}|\langle f_i, \phi_{\mathbf{top}(T)^i}\rangle|^2 \le 4|I_{\mathbf{top}(T)}|$ for all $T \in \mathcal{F}_{n,i,i}$, uniformly in n, i and in $\{f_k\}, \{E_k\}$. Since these tiles $\mathbf{top}(T)^i$ are nonoverlapping and $|f_i| \le \chi_{E_i}$, the almost-orthogonality lemma gives

$$\sum_{T \in \mathcal{F}_{n,i,i}} |I_{\mathbf{top}(T)}| \le C 2^{-2n} \sum_{T \in \mathcal{F}_{n,i,i}} |\langle f_i, \phi_{\mathbf{top}(T)^i}\rangle|^2 \le C 2^{-2n} \|f_i\|_{L^2}^2 \le C 2^{-2n}|E_i|,$$

establishing Lemma 6.2. □

6.3. Summation with respect to n

By the bound of Lemma 4.7 in terms of $\|f\|_{L^\infty}$, the selection criteria (i), (i)* for parameters (n, i, j) cannot be satisfied unless n does not exceed a certain finite n_0, independent of the functions f_k and sets E_k. Therefore $\mathcal{F}_{n,i,j}$ is empty for all larger n. Thus

$$\sum_n \sum_{i,j} \sum_{T \in \mathcal{F}_{n,i,j}} \mathcal{H}_T(f_1, f_2, f_3) \le C \sum_{n \le n_0} 2^{3n} 2^{-2n} \sum_{k=1}^{3} |E_k| \le C \sum_{k=1}^{3} |E_k|. \tag{27}$$

In the case where all three sets E_j have comparable measures, this is the bound of Proposition 6.1. Although it remains to treat the general case, this suffices to exhibit essential elements of the analysis:[10] localized Fourier decomposition, the connection between phase space disjointness and orthogonality, sorting of tiles into towers, the relation between towers and trees, the key role of the weighted count of all resulting multitrees, and the role of orthogonality in establishing that count.

7. REFINEMENT

In this more technical section we briefly discuss two steps omitted above. The first is the counting of multitrees of type (n, i, j) for $j \ne i$. The second is the refinement of the above argument to replace the crude bound $\sum_{k=1}^{3} |E_k|$ by $\prod_{k=1}^{3} |E_k|^{1/p_k}$.

[10] It also fully proves a nontrivial result, an L^1 inequality for $BH(f_1, f_2)$ for arbitrary L^∞ functions f_j supported on sets of boundedly finite measures.

7.1. Counting multitrees of type (n,i,j) with $j \neq i$

The total "number" $\sum_{T\in\mathcal{F}_{n,i,j}} |I_{\mathbf{top}(T)}|$ of multitrees in $\mathcal{F}_{n,i,j}$ is likewise $\le C2^{-2n}|E_j|$. The proof relies on the following variant of the almost-orthogonality Lemma 3.1. A collection $\mathcal{F}$ of trees is said to be *strongly disjoint* if for any $T \neq T' \in \mathcal{F}$, any $P \in T$ and any $Q \in T'$, if $\omega_Q \supset \omega_P$ then not only must $I_Q \cap I_P = \emptyset$, but furthermore $I_Q \cap I_{\mathbf{top}(T)} = \emptyset$.

LEMMA 7.1. — *Let $f \in L^2$ and $\lambda > 0$ be arbitrary. Let $\mathcal{F}$ be a finite collection of strongly disjoint trees satisfying $\lambda \le \operatorname{energy}(T) \le 2\lambda$ for every $T \in \mathcal{F}$. Then*

$$\sum_{T\in\mathcal{F}} |I_{\mathbf{top}(T)}| \le C\lambda^{-2}\|f\|_{L^2}^2. \tag{28}$$

There is a localization in the same spirit as (15).

With this lemma in hand, the reasoning for the case $j \neq i$ is parallel to that for $j = i$, though a bit more complicated. It is almost but not quite true that $\{T^j : T \in \mathcal{F}_{n,i,j}\}$, where $T^j = \{P^j : P \in T\}$, is a strongly disjoint family of trees for any fixed n, i, j with $j \neq i$. See [19] for details.

7.2. Classical trilinear bound

So far, the analysis has relied entirely on L^2 estimates, but genuine L^p inequalities for linear operators do come into play. L^p estimates are available because expressions $\mathcal{H}_T$ associated to multitrees T are subsumed[11] by the theory of singular integral operators, as developed by Coifman and Meyer [9],[10]. If $f_k \in L^\infty$ then such a sum can be rewritten (with $\{1,2,3\} = \{i,j,k\}$) as $\langle \mathcal{T} f_i, f_j\rangle$ where $\mathcal{T}$ is a classical singular integral operator, associated to a kernel K which satisfies (3) with a constant C proportional to $\|f_k\|_{L^\infty}$.

Let E_j be measurable sets satisfying $|E_j| < \infty$, and let f_j be measurable functions supported on E_j satisfying $|f_j(x)| \le 1$ for almost every $x \in E_j$ and $f_j(x) = 0$ for $x \notin E_j$.

PROPOSITION 7.2. — *Let T be any multitree. For all 3-tuples of exponents $p_k \in (1,\infty]$ satisfying the scaling relation $\sum_k p_k^{-1} = 1$ with at most one exponent equal to ∞,*

$$\sum_{P\in T} |I_P|^{-1/2} \prod_{k=1}^{3} |\langle f_k, \phi_{P^k}\rangle| \le C \prod_{k=1}^{3} |E_k|^{1/p_k}. \tag{29}$$

The constant C depends on the exponents but not on the sets, functions, tower, or index l. This is of course a very particular case of Proposition 6.1.

As is often the case in this subject, a localized version is also available: $|E_k|$ can be replaced by $\int_{E_k} (1 + \frac{\operatorname{distance}(x, I_{\mathbf{top}(T)})}{|I_{\mathbf{top}(T)}|})^{-2}\,dx$.

[11] More exactly, these expressions are *modulated* singular integral operators; the associated kernels K are multiplied by factors $e^{i(x-y)\xi}$ for some arbitrary ξ.

7.3. An alternative bound

Consider for simplicity only the case $j = i$. If $\{i, l, m\} = \{1, 2, 3\}$ then using (29) in place of the mass-energy bound for each multitree leads to an alternative bound

$$\sum_{T \in \mathcal{F}_{n,i,i}} \mathcal{H}_T(f_1, f_2, f_3) \le C_\varepsilon 2^{-(1+\varepsilon)n} |E_m|^{1-2\varepsilon} |E_l|^\varepsilon |E_i|^\varepsilon \tag{30}$$

for arbitrarily small $\varepsilon > 0$.

To prove this write $\{1, 2, 3\} = \{i, l, m\}$. Application of the almost-orthogonality lemma as above shows that uniformly for all dyadic intervals J,

$$\sum_{T \in \mathcal{F}_{n,i,i} : I_{\mathbf{top}(T)} \subset J} |I_{\mathbf{top}(T)}| \le C 2^{-2n} \int_{E_i} \big(1 + \frac{\text{distance}\,(x, J)}{|J|}\big)^{-2} \, dx \le C 2^{-2n} |J|. \tag{31}$$

The *John-Nirenberg lemma* says that this “self-similar” inequality implies a stronger version of itself, to the effect that $\sum_{T \in \mathcal{F}_{n,i,i}} \chi_{I_{\mathbf{top}(T)}}$ is nearly a bounded function:

$$\int_J \big(\sum_{T \in \mathcal{F}_{n,i,i}} \chi_{I_{\mathbf{top}(T)}} \big)^r \le C_r 2^{-2nr} |J| \tag{32}$$

uniformly for all dyadic intervals J, for any finite exponent r.

Write $\{1, 2, 3\} = \{i, l, m\}$. Concerning the contributions of f_m, f_l, classical Calderón-Zygmund theory almost gives

$$\sum_{P \in T} |\langle f_m, \phi_{P^m} \rangle| \, |\langle f_l, \phi_{P^l} \rangle| \le C |E_l \cap I_{\mathbf{top}(T)}|^{1/p_l} |E_m \cap I_{\mathbf{top}(T)}|^{1/p_m} \tag{33}$$

for any exponents in $(1, \infty)$ satisfying $p_l^{-1} + p_m^{-1} = 1$; in particular, p_m can be taken to be arbitrarily close to 1. This inequality is merely almost true; $|E_j \cap I_{\mathbf{top}(T)}|$ must be replaced by $\int_{E_j} (1 + \frac{\text{distance}(x,I)}{|I|})^{-2} \, dx$, where $I = I_{\mathbf{top}(T)}$. This sort of routine complication is controlled satisfactorily by the Hardy-Littlewood maximal function.

Consider all intervals J which are maximal among the collection $\{I_{\mathbf{top}(T)} : T \in \mathcal{F}_{n,i,i}\}$ with respect to set inclusion. Combining (33) with (32) and Hölder’s inequality for each interval J, summing over J, and finally exploiting the inclusions $J \subset \{x : M(\chi_{E_i})(x) \ge c 2^{2n}\}$ leads to (30). Details are left to the experts.

(30) is unfavorable for n near $-\infty$, but the bound $2^n |E_i|$ is still favorable for all sufficiently negative n. To conclude the argument, modify the selection algorithm by replacing the thresholds 2^n in criteria (i), (i)* by $2^{n a_j}$ where $a_j > 0$ are parameters. Adjusting these appropriately, depending on the relative sizes of $|E_i|, |E_m|, |E_l|$, makes the minimum of our two bounds favorable in all cases, and permits summation of the infinite series over n.

8. PROOF OF THE ALMOST-ORTHOGONALITY LEMMA

I have emphasized the leading role played by L^2 arguments based on almost-orthogonality. The required almost-orthogonality lemma has a relatively simple proof. First, a preliminary fact: Define $h_P(x) = |I_P|^{-1}(1 + \frac{\text{distance}(x, I_P)}{|I_P|})^{-2}$. These functions belong to $L^1(\mathbb{R})$ uniformly in P. If $|I_Q| \le |I_P|$ then

$$|\langle \phi_P, \phi_Q \rangle| \le C|I_P|^{1/2}|I_Q|^{-1/2} \int_{I_Q} h_P.$$

$h_P(x)$ decays rapidly as x moves away from P, on a scale comparable to $|I_P|$.

Proof of Lemma 3.1. — Set $\beta_P = \langle f, \phi_P \rangle$ and $X^2 = \sum_{P \in \mathcal{F}} |\beta_P|^2$. Then

$$X^2 = \sum_P \langle f, \phi_P \rangle \langle \phi_P, f \rangle = \Big\langle \sum_P \beta_P \phi_P,\, f \Big\rangle \le \|f\|_{L^2} \|\sum_P \beta_P \phi_P\|_{L^2}.$$

Now

$$\|\sum_P \beta_P \phi_P\|_{L^2}^2 \le 2 \sum_P \sum_{Q: \omega_P \subset \omega_Q} |\beta_P \beta_Q \langle \phi_P, \phi_Q \rangle|$$

because $\langle \phi_P, \phi_Q \rangle = 0$ unless ω_P, ω_Q overlap, and dyadic intervals cannot overlap unless one contains the other. Thus

$$\begin{aligned} \|\sum_P \beta_P \phi_P\|_{L^2}^2 &\le 8\lambda^2 \sum_P |I_P|^{1/2} \sum_{Q: \omega_P \subset \omega_Q} |I_Q|^{1/2} |\langle \phi_P, \phi_Q \rangle| \\ &\le C\lambda^2 \sum_P |I_P| \sum_{Q: \omega_P \subset \omega_Q} \int_{I_Q} h_P \le C\lambda^2 \sum_P |I_P| \int_{\mathbb{R}} h_P \le C \sum_P |\beta_P|^2 \end{aligned}$$

since the collection of all intervals I_Q such that $Q \in \mathcal{F}$ and $\omega_P \subset \omega_Q$ is nonoverlapping by hypothesis. Thus $X^4 \le C\|f\|_{L^2}^2 X^2$. □

9. CARLESON'S MAXIMAL OPERATOR, ACCORDING TO LACEY AND THIELE

Here I give only a brief outline of a rather condensed treatment [21], in order to exhibit the parallel between the analyses of the bilinear Hilbert transform and Carleson's maximal operator.

Expression in terms of localized Fourier coefficients. For any dyadic interval ω, denote by $\omega^\flat$ and $\omega^\sharp$ the left and right halves of ω, respectively. Likewise to any tile $P = I \times \omega$ are associated the two semitiles $P^\flat = I \times \omega^\flat$ and $P^\sharp = I \times \omega^\sharp$.

Let functions ϕ_P be as above, with the single change that $\widehat{\phi_P}$ is supported in $\omega_P^\flat$. The linearized Carleson operator $\int_{\mathbb{R}} e^{iN(x)t} f(x-t)\, t^{-1} dt$ is modeled by sums

$$\mathcal{C}^{**}(f)(x) = \sum_P \langle f, \phi_P \rangle \phi_P(x) \chi(x, P) \tag{34}$$

where $\chi(x,P)=1$ if $N(x)$ lies in $\omega_P^\sharp$, and $=0$ otherwise. In fact, this operator can be realized as a limit of averages of such models. The goal is an inequality $|\langle \mathcal{C}^{**}(f),\chi_E\rangle| \le C\|f\|_{L^2}|E|^{1/2}$ for all measurable sets E.

Towers. A tower T is a set of tiles for which there exists some tile $\mathbf{top}(T)$ such that for every $P\in T$, $P\le \mathbf{top}(T)$; in this discussion $\mathbf{top}(T)$ is *not* required to be an element of T. $\sharp$-towers and $\flat$-towers are defined in the same way, with the requirements $P^\sharp \le \mathbf{top}(T)^\sharp$ and $P^\flat \le \mathbf{top}(T)^\flat$, respectively.

There are two fundamental sources of orthogonality. Firstly, if T is a $\flat$-tower, then for any $P,Q\in T$ satisfying $|I_P|\ne|I_Q|$, the intervals $\omega_P^\sharp,\omega_Q^\sharp$ do not overlap.(12) Therefore the sets $\{x\in E : N(x)\in\omega_P^\sharp\}$ and $\{x\in E : N(x)\in\omega_Q^\sharp\}$ are disjoint. Secondly, any $\sharp$-tower T is a $\flat$-tree: whenever $P,Q\in T$ satisfy $|I_P|\ne|I_Q|$, $\omega_P^\flat$ and $\omega_Q^\flat$ do not overlap. Therefore $\langle\phi_P,\phi_Q\rangle=0$.

The energy of a tower T is

$$\mathcal{E}(T)=\sup_{T'\subset T}|I_{\mathbf{top}(T')}|^{-1/2}\Big(\sum_{P\in T'}|\langle f,\phi_P\rangle|^2\Big)^{1/2}, \tag{35}$$

where the supremum is taken over all $\sharp$-towers $T'\subset T$. The mass of a tower is

$$\mathcal{M}(T)=\sup_{P\in T}\sup_{Q\ge P}|I_Q|^{-1}\int_{\{x\in E:N(x)\in\omega_Q\}}(1+|I_Q|^{-1}\,\mathrm{distance}\,(x,I_Q))^{-2}\,dx. \tag{36}$$

Classical mass-energy bound for a single tower. In the analysis of the bilinear Hilbert transform, the contribution of a single multitree was estimated by a trivial mass-energy bound. While there is a mass-energy bound here, it is not trivial. For any tower,

$$\sum_{P\in T}|\langle f,\phi_P\rangle|\cdot|\langle\phi_P,\chi_{N(x)\in\omega_P^\sharp}\chi_E\rangle|\le C\mathcal{E}(T)\mathcal{M}(T)|I_{\mathbf{top}(T)}|. \tag{37}$$

There is also a very simple alternative bound $\mathcal{M}(T)\le C<\infty$, uniformly for all sets E, because $\chi_E\in L^\infty$ with norm 1.

(37) is roughly on the level of Proposition 7.2, and is proved as follows. Any $\sharp$-tower is a $\flat$-tree, and the associated sum is essentially a (frequency-modulated generalization of a) truncated singular integral operator

$$\int_{|x-y|\ge\varepsilon(x)}K(x,y)f(y)\,dy \tag{38}$$

for some function $\varepsilon(x)$ and Calderón-Zygmund kernel K. Such truncations are basic objects in the classical theory. On the other hand, any $\flat$-tower enjoys the first form of orthogonality described above, and this leads quite easily to the upper bound (37) for its contribution.

Sorting algorithm. As in the analysis of BH, a selection algorithm partitions any finite set of tiles by selecting certain collections $\mathcal{F}_{n,\natural}$ of towers, for each integer n and

(12)This is correct under the usual proviso that technicalities concerning dyadic intervals have been appropriately dealt with.

index $\natural \in \{\flat, \sharp\}$. For $\mathcal{F}_{n,\flat}$ the main selection criterion is that $\mathcal{M}(T) \geq 2^{2n}$, while for $\mathcal{F}_{n,\sharp}$ it is that $\mathcal{E}(T) \geq 2^n$. In both cases T is required to be maximal, and for $\mathcal{F}_{n,\sharp}$ there is an analogue of criterion (iii)*.

Counting the towers. The key point is again an upper bound for the weighted number of trees in $\mathcal{F}_{n,\natural}$.

LEMMA 9.1. — *Uniformly for all functions f and all measurable sets $E \subset \mathbb{R}$ satisfying $\|f\|_{L^2} \leq 1$ and $|E| \leq 1$,*

$$\sum_{T \in \mathcal{F}_{n,\sharp} \cup \mathcal{F}_{n,\flat}} |I_{\mathbf{top}(T)}| \leq C 2^{-2n}. \tag{39}$$

For $\mathcal{F}_{n,\flat}$, (39) is essentially a bound for the Hardy-Littlewood maximal function of χ_E. The proof for $\mathcal{F}_{n,\sharp}$ is essentially the same as that of the counting bound for $\mathcal{F}_{n,i,j}$ with $i \neq j$ in the bilinear Hilbert transform analysis; the primary ingredient is the almost-orthogonality lemma for strongly disjoint trees, Lemma 7.1. The strong disjointness of the collection $\mathcal{F}_{n,\sharp}$ of trees is a consequence of the selection algorithm.

In this argument there is no *a priori* upper bound on n, but summation over all $n \in \mathbb{Z}$ yields the desired uniform upper bound anyway (the alternative bound $\mathcal{M}(T) \leq C < \infty$ is used for $n \geq 0$).

10. OPEN PROBLEMS

Higher-degree multilinear operators. The bilinear Hilbert transform can be generalized to

$$T(f_1, \cdots, f_m)(x) = \int_{\mathbb{R}} \prod_{k=1}^{m} f_k(x - \alpha_k t)\, t^{-1}\, dt \tag{40}$$

where the α_j are pairwise distinct and nonzero. The analysis outlined above fails to apply to this operator for $m \geq 3$, and it is an open problem whether this formal expression has any meaning for functions in appropriate spaces; scaling dictates that the natural estimate would be $\|T(f_1, f_2, f_3)\|_{L^q} \leq C \prod \|f_k\|_{L^{p_k}}$ where $q^{-1} = \sum_k p_k^{-1}$. In fact, for $m \geq 4$, when this operator is expanded in terms of localized Fourier coefficients, the resulting sum actually fails to converge absolutely. See [22] and the references cited there for some positive results in this direction.

I mention in passing that for nonsingular expressions $T(f_1, \ldots, f_m)(x) = \int_{-1}^{1} \prod_{k=1}^{m} f_k(x - \alpha_k t)\, dt$, with no singular factor t^{-1}, there are interesting questions (for $m \geq 3$) [7] which make contact with work of Bourgain [1] and of Katz-Tao [17] on a problem in additive number theory related to the Kakeya problem.

Scattering for one-dimensional Dirac operators. Associated to any potential $V : \mathbb{R} \to \mathbb{C}$ is the Dirac operator $D_V = \begin{pmatrix} -i\frac{d}{dx} & \bar{V} \\ V & i\frac{d}{dx} \end{pmatrix}$. If V has compact support then for each $\lambda \in \mathbb{R}$ there exist unique scattering coefficients $a(\lambda), b(\lambda)$ for which there exists a solution $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ of the generalized eigenfunction equation $D_V u = \lambda u$ taking the form $u(x) \equiv \begin{pmatrix} e^{i\lambda x} \\ 0 \end{pmatrix}$ as $x \to -\infty$ and $u(x) \equiv \begin{pmatrix} a(\lambda)e^{i\lambda x} \\ b(\lambda)e^{-i\lambda x} \end{pmatrix}$ as $x \to +\infty$.

For potentials $V \in L^1(\mathbb{R})$, it is elementary that there still exist solutions with these asymptotics, for every λ. For $V \in L^p$ for $1 < p < 2$, this continues to hold for Lebesgue-almost every λ. However, there are indications [14] that the natural class of potentials is $V \in L^2$.

For any x, $u(x)$ can be written in the form $u(x) = \begin{pmatrix} a(x,\lambda)e^{i\lambda x} \\ b(x,\lambda)e^{-i\lambda x} \end{pmatrix}$ where $b(\lambda) = \lim_{x\to+\infty} b(x,\lambda)$ and likewise for a. The mapping $V \to \begin{pmatrix} a(x,\lambda) \\ b(x,\lambda) \end{pmatrix}$ is fully nonlinear. A Taylor-type expansion about $V = 0$ gives the linearized expression

$$b(x,\lambda) = i\int_{-\infty}^{x} e^{2i\lambda y}V(y)\,dy \qquad \text{plus higher-order terms.} \tag{41}$$

The question is whether $\lim_{x\to+\infty} b(x,\lambda)$ exists for almost every λ, for every $V \in L^2(\mathbb{R})$. By writing $V = \widehat{f}$ we see that the linearization of this problem about $V = 0$ is simply a restatement of Carleson's theorem. Thus the almost everywhere existence of these scattering coefficients, for general $V \in L^2$, is a nonlinear extension of the problem of almost everywhere convergence of Fourier integrals.

For $1 < p < 2$, this was proved [8] using an expansion of the mapping $V \mapsto b(x,\lambda)$ as an infinite sum of multilinear expressions acting on V. However, Muscalu, Tao, and Thiele [23] have shown that even the first nonlinear expression in this series diverges for general $V \in L^2$, and the problem remains open despite an interesting positive result of those authors concerning a related model problem.

There is an almost identical problem for one-dimensional Schrödinger operators, with the added complication that a WKB-type phase correction must be incorporated [8].

Epilogue. There is a great more to be said, both about subsequent developments to which other authors have made important contributions, and about other types of multilinear operators which appear in contemporary analysis. Space-time inequalities prevent this author from discussing those matters here. I particularly regret having had to give short shrift to other authors' contributions in order to discuss the themes chosen in the space allotted.

REFERENCES

[1] J. BOURGAIN – *On the dimension of Kakeya sets and related maximal inequalities.* Geom. Funct. Anal. 9 (1999), no. 2, 256–282.

[2] A. P. CALDERÓN and A. ZYGMUND – *On the existence of certain singular integrals.* Acta Math. 88, (1952). 85–139.

[3] A. P. CALDERÓN – *Uniqueness in the Cauchy problem for partial differential equations.* Amer. J. Math. 80 1958 16–36.

[4] A. P. CALDERÓN – *Commutators of singular integral operators.* Proc. Nat. Acad. Sci. U.S.A. 53 1965 1092–1099.

[5] A. P. CALDERÓN – *Cauchy integrals on Lipschitz curves and related operators.* Proc. Nat. Acad. Sci. U.S.A. 74 (1977), no. 4, 1324–1327.

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

[7] M. CHRIST – *On certain elementary trilinear operators.* Math. Res. Lett. 8 (2001), no. 1-2, 43–56.

[8] M. CHRIST and A. KISELEV – *WKB asymptotic behavior of almost all generalized eigenfunctions for one-dimensional Schrödinger operators with slowly decaying potentials.* J. Funct. Anal. 179 (2001), no. 2, 426–447.

[9] R. R. COIFMAN and Y. MEYER, *Au delà des opérateurs pseudo-différentiels.* Astérisque, 57. Société Mathématique de France, Paris, 1978.

[10] R. R. COIFMAN and Y. MEYER, *Ondelettes et opérateurs. III. Opérateurs multilinéaires.* Actualités Mathématiques. Hermann, Paris, 1991.

[11] R. R. COIFMAN and Y. MEYER, *Le théorème de Calderón par les "méthodes de variable réelle".* Seminar on Harmonic Analysis, 1978–1979, pp. 49–55, Publ. Math. Orsay 79, 7, Univ. Paris XI, Orsay, 1979.

[12] R. R. COIFMAN, Y. MEYER, and A. MCINTOSH – *L'intégrale de Cauchy définit un opérateur borné sur L^2 pour les courbes lipschitziennes.* Ann. of Math. (2) 116 (1982), no. 2, 361–387.

[13] G. DAVID and J-L. JOURNÉ – *A boundedness criterion for generalized Calderón-Zygmund operators.* Ann. of Math. (2) 120 (1984), no. 2, 371–397.

[14] P. DEIFT and R. KILLIP, *On the absolutely continuous spectrum of one-dimensional Schr̈dinger operators with square summable potentials.* Comm. Math. Phys. 203 (1999), no. 2, 341–347.

[15] C. FEFFERMAN – *Pointwise convergence of Fourier series*, Ann. of Math. 98 (1973), 551–571.

[16] J. B. GARNETT and P. W. JONES – *BMO from dyadic BMO.* Pacific J. Math. 99 (1982), no. 2, 351–371.

[17] N. H. KATZ and T. TAO – *Bounds on arithmetic projections, and applications to the Kakeya conjecture.* Math. Res. Lett. 6 (1999), no. 5-6, 625–630.

[18] A. KNAPP and E. M. STEIN – *Intertwining operators for semisimple groups.* Ann. of Math. (2) 93 (1971), 489–578.

[19] M.T. LACEY and C.M. THIELE – *L^p bounds for the bilinear Hilbert transform, $p > 2$*, Ann. of Math. 146 (1997), 693–724.

[20] M.T. LACEY and C.M. THIELE – *On Calderón's Conjecture*, Ann. of Math. 149 (1999), 475–496.

[21] M.T. LACEY and C.M. THIELE – *A proof of boundedness of the Carleson operator*, Math. Res. Lett. 7 (2000), no. 4, 361–370.

[22] C. MUSCALU, T. TAO and C. THIELE – *Multi-linear operators given by singular multipliers.* J. Amer. Math. Soc. 15 (2002), no. 2, 469–496.

[23] C. MUSCALU, T. TAO and C. THIELE – *A counterexample to a multilinear endpoint question of Christ and Kiselev.* Math. Res. Lett. 10 (2003), no. 2-3, 237–246.

[24] F. NAZAROV, S. TREIL and A. VOLBERG – *The Tb-theorem on non-homogeneous spaces.* Acta Math. 190 (2003), no. 2, 151–239.

[25] H. PAJOT – *Capacité analytique et le probléme de Painlevé.* Séminaire Bourbaki. Vol. 2003/2004. Astérisque No. 299 (2005), Exp. No. 936, ix, 301–328.

Michael CHRIST

Department of Mathematics
University of California
Berkeley, CA 94720-3840 – U.S.A.
E-mail : mchrist@math.berkeley.edu