

## BOUNDEDNESS OF BILINEAR OPERATORS WITH NONSMOOTH SYMBOLS

JOHN E. GILBERT AND ANDREA R. NAHMOD\*

ABSTRACT. We announce the  $L^p$ -boundedness of general bilinear operators associated to a symbol or multiplier which need not be smooth. We establish a general result for multipliers that are allowed to have singularities along the edges of a cone as well as possibly at its vertex. It thus unifies earlier results of Coifman-Meyer for smooth multipliers and ones, such the Bilinear Hilbert transform of Lacey-Thiele, where the multiplier is not smooth.

### 1. Introduction and statement of the results

Let  $\mathcal{B} : \mathcal{S}(\mathbb{R}) \times \mathcal{S}(\mathbb{R}) \rightarrow \mathcal{S}'(\mathbb{R})$  be a continuous bilinear operator which commutes with simultaneous translations. Then there exists  $m$  in  $\mathcal{S}'(\mathbb{R} \times \mathbb{R})$ , the *symbol* or *multiplier*, such that

$$(1.1) \quad \mathcal{B}(f, g)(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m(\xi, \eta) \widehat{f}(\xi) \widehat{g}(\eta) e^{2\pi i x(\xi + \eta)} d\xi d\eta,$$

and  $\mathcal{B}$  commutes also with simultaneous dilations if  $m$  is homogeneous of degree 0. It is easy to see that  $f, g \rightarrow \mathcal{B}(f, g)$  is continuous as a mapping from  $\mathcal{S}(\mathbb{R}) \times \mathcal{S}(\mathbb{R})$  into  $L^2(\mathbb{R})$  when  $m$  is in  $L^\infty(\mathbb{R}^2)$ , and that  $\mathcal{B}(f, g)$  lies in the complex Hardy space  $H_{\mathbb{C}}^2(\mathbb{R})$  if in addition the support of  $m$  lies in the half-plane  $\xi + \eta \geq 0$ . The basic  $L^p$ -boundedness problem is to prescribe conditions on  $m = m(\xi, \eta)$  so that  $\mathcal{B}$  extends to a bounded operator from  $L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$  for  $p, q > 1$  and  $1/p + 1/q = 1/r$ .

In this note we report the  $L^p$ -boundedness result when  $m$  is not necessarily smooth, unifying previous results of Coifman-Meyer for smooth multipliers with ones for the non-smooth case, including the recent results of Lacey-Thiele for the Bilinear Hilbert transform. The first Main Theorem establishes a general result for multipliers that are allowed to have singularities along the edges of a cone as well as possibly at its vertex. Using a Whitney decomposition in

---

Received June 19, 2000.

1991 *Mathematics Subject Classification*. Primary 42B15, 42C15. Secondary 42B20, 42B25, 40A20

*Key words and phrases*. Bilinear operator, singular multiplier, wave-packet, paraproduct.

\*The second author was partially supported by NSF grant DMS 9971159 and acknowledges the support of MSRI, Berkeley (NSF grant DMS 9701755) and of IAS at Princeton (NSF grant DMS 9304580) during 1997-1998. She warmly thanks both Institutes for their hospitality while part of this research was being carried out.

the Fourier plane a general bilinear operator is represented as infinite discrete sums of time-frequency paraproducts obtained by associating wave-packets with tiles in phase-plane. Boundedness for the general bilinear operator then follows once the corresponding  $L^p$ -boundedness of time-frequency paraproducts is established. The latter result, the second Main Theorem, is proved using phase-plane analysis. The affine invariant structure of such operators in conjunction with the geometric properties of the associated phase-plane decompositions allow Littlewood-Paley techniques to be applied locally, ie. on trees. Boundedness of the full time-frequency paraproduct then follows using ‘almost orthogonality’ type arguments relying on estimates for tree-counting functions together with decay estimates. The results in this note represent research carried out over several years and completed in the summer of 1999. During that time period various aspects of this research and most of the ideas were presented by the authors in a number of lectures all around. Full details and proofs are contained in [8] [9].

**Main Theorem I.** *Let  $\Gamma$  be a closed one-sided cone with vertex at the origin and  $m = m(\xi, \eta)$  a function having derivatives of all orders inside  $\Gamma$  such that*

$$(1.2) \quad |D^\alpha m(\xi, \eta)| \leq \text{const.} \left( \frac{1}{\text{dist}((\xi, \eta), \partial\Gamma)} \right)^{|\alpha|}, \quad |\alpha| \geq 0.$$

*Then the bi-linear operator*

$$\mathcal{C}_\Gamma : f, g \longrightarrow \int_\Gamma m(\xi, \eta) \widehat{f}(\xi) \widehat{g}(\eta) e^{2\pi i x(\xi + \eta)} d\xi d\eta$$

*is bounded from  $L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$ ,  $1/p + 1/q = 1/r < 3/2$ , so long as no edge of  $\Gamma$  lies on the diagonal  $\xi + \eta = 0$  or on a coordinate axis. Furthermore, when  $\Gamma$  lies in the half-plane  $\xi + \eta > 0$  and  $r \geq 1$ , the operator  $\mathcal{C}_\Gamma$  has range in the complex Hardy space  $H_{\mathbb{C}}^r(\mathbb{R})$ .*

There is a corresponding Hardy space result when  $\Gamma$  lies in the half-plane  $\xi + \eta < 0$ . By changing variables  $\eta \longrightarrow -\eta$  we also obtain an equivalent result for sesqui-linear operators

$$\bar{\mathcal{C}}_\Gamma : f, g \longrightarrow \int_\Gamma m(\xi, \eta) \widehat{f}(\xi) \overline{\widehat{g}(\eta)} e^{2\pi i x(\xi - \eta)} d\xi d\eta.$$

**Remark.** In these results the multiplier  $m$  need only be smooth up to some sufficiently high order, but no attempt is made to quantify the necessary smoothness. If  $m$  is  $C^\infty$  everywhere in the plane except possibly at the origin its restriction to any cone  $\Gamma$  will satisfy (1.2) automatically provided

$$(1.3) \quad |D^\alpha m(\xi, \eta)| \leq \text{const.} \frac{1}{(|\xi| + |\eta|)^{|\alpha|}}, \quad |\alpha| \geq 0.$$

In particular, (1.3) will be satisfied whenever  $m$  is  $C^\infty$  and homogeneous of degree 0. For such multipliers the edges of the cone could be allowed to lie on one

or more of the coordinate axes. Thus, an easy corollary of Main Theorem I is the boundedness of the bilinear operators whose symbol is the degree zero homogeneous extension of a *piecewise- $C^\infty$* ( $\Sigma_1$ ) symbol, which is  $C^\infty$  in a neighborhood of  $(\xi, -\xi)$ . This result was conjectured in [7] and its existence suggested in [2].

The proof of Main Theorem I proceeds via special cases. For a given  $\theta$  let

$$\mathcal{C}_{P_\theta} : f, g \longrightarrow \int_{P_\theta} m(\xi, \eta) \widehat{f}(\xi) \widehat{g}(\eta) e^{2\pi i x(\xi+\eta)} d\xi d\eta$$

be the cone operator associated with the half-plane  $P_\theta = \{(\xi, \eta) : \xi \tan \theta - \eta > 0\}$  and  $\overline{\mathcal{C}}_{P_\theta}$  the corresponding sesqui-linear version.

**Theorem 1.4.** *Let  $m = m(\xi, \eta)$  be a function having derivatives of all orders in the half-plane  $P_\theta$  such that*

$$|D^\alpha m(\xi, \eta)| \leq \text{const.} \left( \frac{1}{\text{dist}((\xi, \eta), \partial P_\theta)} \right)^{|\alpha|}, \quad |\alpha| \geq 0.$$

*Then, if  $\partial P_\theta$  is not one of the coordinate axes,  $\mathcal{C}_{P_\theta}$  and  $\overline{\mathcal{C}}_{P_\theta}$  are bounded from  $L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$ ,  $1/p + 1/q = 1/r < 3/2$ , whenever  $\theta \neq -\pi/4$  and  $\pi/4$  respectively.*

Again the coordinate axes can be allowed if  $m$  satisfies (1.3) everywhere away from the origin in the plane. By taking  $m(\xi, \eta) \equiv 1$  we thus obtain all the Bilinear Hilbert transform results of Lacey-Thiele ([13], [14]).

**Remark.** Save for the restriction  $r > 2/3$ , theorem (1.4) also includes the well-known result of Coifman-Meyer establishing the boundedness of  $\mathcal{C}_{\mathbb{R}^2}(f, g)$  from  $L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$ ,  $r > 1/2$ , for any  $C^\infty$ -function  $m$  satisfying (1.3) (cf., [3, 4]). In fact, it is enough to write  $\mathcal{C}_{\mathbb{R}^2}$  as the sum  $\mathcal{C}_{P_\theta} + \mathcal{C}_{\mathbb{R}^2 \setminus P_\theta}$  for any allowed choice of  $\theta$ . It is interesting to note that a natural ‘miniaturization’ of the proof of Main Theorem I actually provides a proof of the  $L^p$ -boundedness of  $\mathcal{C}_{\mathbb{R}^2}$  for the full range of  $r$  ( cf.[10] [11] for other recent and independent proofs of the latter and more ). It also points to the reason for the failure to obtain the lower value of  $r$  in Main Theorem I. Indeed, in (1.3) the only singularity in the multiplier is at the origin - there is a preferred point in frequency, in other words - so that wave packets have only to contain translations in time and dilation. By contrast, in Main Theorem I there is no such preferred point because the singularities can lie on the full boundary of  $\Gamma$ . As a result wave packets now have to contain translation in frequency as well, *i.e.*, modulation. Even after including modulations, however, there is only one point in the proof, an application of the Hausdorff-Young inequality, at which it becomes essential to impose the condition  $r > 2/3$ . Save for this, the proof of Main Theorem I would be valid without restriction on  $r$ .

**Theorem 1.5.** *Let  $m = m(\xi, \eta)$  be a function having derivatives of all orders in the half-plane  $P_\theta$  such that*

$$|D^\alpha m(\xi, \eta)| \leq \text{const.} \left( \frac{1}{\text{dist}((\xi, \eta), \partial P_\theta)} \right)^{|\alpha|}, \quad |\alpha| \geq 0.$$

Then  $\mathcal{C}_{P_\theta}$  is bounded from  $L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$ ,  $1/p + 1/q = 1/r < 3/2$ , so long as  $0 < \theta \leq \pi/4$  while  $\overline{\mathcal{C}}_{P_\theta}$  is bounded if  $0 < \theta < \pi/4$ .

Granted (1.5), (1.4) follows easily and from it, Main Theorem I is readily established. Thus we concentrate on theorem (1.5). There are two fundamental ideas. The first is to represent  $\mathcal{C}_{P_\theta}$  in terms of a doubly-infinite sum of ‘discrete’ bilinear operators, and then secondly to establish  $L^p$ -boundedness for these discretizations.

**Time-frequency paraproducts.** Given positive numbers  $a_j$ , a positive rational  $\rho$ , and  $\mathcal{M}_\mu$ -test functions  $\phi^{(j)}$ , let

$$\phi_{k\ell n}^{(j)}(x) = \phi_Q^{(j)}(x) = s^{k/2} \phi_j(s^k x - a_j \ell) e^{2\pi i s^k x n}, \quad s = 2^\rho$$

be the corresponding wave packet associated with a tile  $Q \sim \{k, \ell, n\}$  in phase plane, incorporating translation in time, scaling, and modulation. By analogy with ‘standard’ paraproducts we form the sum

$$\mathcal{D}(f, g) = \sum_{k, \ell, n} s^{k/2} c_{k\ell n} \langle f, \phi_{k\ell n}^{(1)} \rangle \langle g, \phi_{k\ell n}^{(2)} \rangle \phi_{k\ell n}^{(3)}$$

over all tiles  $Q \sim \{k, \ell, n\}$  in phase plane, the coefficients  $c_{k\ell n}$  being in  $\ell^\infty$ . In ‘standard’ paraproducts there are no modulations and boundedness from  $\ell^\infty \times L^\infty(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^q(\mathbb{R})$  is well-known under the assumption that at least two of the ‘mother wave functions’ have vanishing moment (and more generally). Since modulation need not preserve vanishing moments, however, stronger conditions will have to be imposed to secure analogous  $L^p$ -boundedness results for  $\mathcal{D}(f, g)$ . Let  $w^{(j)}$  be finite intervals such that:

$$\text{supp } \widehat{\phi}^{(1)} \subseteq w^{(1)}, \quad \text{supp } \widehat{\phi}^{(2)} \subseteq w^{(2)}, \quad \text{supp } \widehat{\phi}^{(3)} \subseteq w^{(3)}$$

The substitute for vanishing moments is the requirement that the  $w^{(j)}$  have pairwise-disjoint closure.

**Definition 1.6.** Fix positive constants  $a_j$ , a positive rational  $\rho$ , and  $\mathcal{M}_\mu$ -test functions  $\phi^{(j)}$ . Then the bilinear operator

$$\mathcal{D} : f, g \longrightarrow \sum_{k, \ell, n} s^{k/2} c_{k\ell n} \langle f, \phi_{k\ell n}^{(1)} \rangle \langle g, \phi_{k\ell n}^{(2)} \rangle \phi_{k\ell n}^{(3)}, \quad s = 2^\rho$$

will be called a time-frequency paraproduct if the  $\phi^{(j)}$  have pairwise-disjoint Fourier support intervals  $w^{(j)}$ .

By a delicate phase-plane analysis in the spirit of C. Fefferman’s proof of Carleson’s theorem on the a.e. convergence of Fourier series of  $L^2$ -functions ([1], [5]) we have:

**Main Theorem II.** *Let  $\phi^{(j)}$  be  $\mathcal{M}_\mu(\mathbb{R})$ -test functions whose Fourier support intervals  $w^{(j)}$  have pairwise-disjoint closure. Then the time-frequency paraproduct*

$$\mathcal{D} : \{c_{k\ell n}\}, f, g \longrightarrow \sum_{k,\ell,n} s^{k/2} c_{k\ell n} \langle f, \phi_{k\ell n}^{(1)} \rangle \langle g, \phi_{k\ell n}^{(2)} \rangle \phi_{k\ell n}^{(3)}$$

where  $s = 2^\rho$ , is bounded from  $\ell^\infty \times L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$ , provided  $1/p + 1/q = 1/r < 3/2$ . Furthermore, the operator norm of  $\mathcal{D}$  satisfies the inequality

$$\|\mathcal{D}\|_{op} \leq \text{const. } P(\|\phi^{(1)}\|, \|\phi^{(2)}\|, \|\phi^{(3)}\|)$$

for some polynomial  $P$  depending only on  $a_j, \rho$  and the Fourier support intervals  $w^{(j)}$ .

Examples show that the restriction  $r > 2/3$  in Main Theorem II is sharp [12]. The boundedness results for the corresponding sesqui-linear version, follow from those for  $\mathcal{D}$ .

**Diagonalization of cone operators.** To ‘diagonalize’  $\mathcal{C}_{P_\theta}$  fix  $\theta \in (0, \pi/4]$  and recall that  $P_\theta$  is the half-plane  $\{(\xi, \eta) : \xi \tan \theta - \eta > 0\}$ . The basic idea is to generate a Whitney covering  $\{R_{kn}\}$  of  $P_\theta$  by translating and dilating a single square  $R$ . Then  $\mathcal{M}_\mu$ -test functions  $\psi^{(j)}$  arise as smooth bump functions associated with  $R$ . By taking Short Fourier transform expansions on each square  $R_{kn}$ , the operator  $\mathcal{C}_{P_\theta}$  can be represented as a doubly-infinite sum

$$\mathcal{C}_{P_\theta}(f, g) = \sum_{\lambda_1, \lambda_2 = -\infty}^{\infty} \mathcal{D}_{\lambda_1 \lambda_2}^{(\varphi)}(f, g)$$

of functions

$$\mathcal{D}_{\lambda_1 \lambda_2}^{(\varphi)}(f, g) = \sum_{k,\ell,n=-\infty}^{\infty} c_{kn}(\lambda_1, \lambda_2) s^{k/2} \langle f, \varphi_{k\ell n}^{(1)} \rangle \langle g, \varphi_{k\ell n}^{(2)} \rangle \varphi_{k\ell n}^{(3)}$$

in which  $\varphi^{(j)}(x) = \psi^{(j)}(x + a_j \lambda_j)$ , ( $j = 1, 2$ );  $\varphi^{(3)}(x) = \psi^{(3)}$  and the wave packets  $\varphi_{k\ell n}^{(j)}$  are defined by

$$\varphi_{k\ell n}^{(j)}(x) = s^{k/2} \varphi^{(j)}(s^k x - a \ell) e^{2\pi i s^k b_j n x}$$

for a fixed choice of positive (geometric) constants  $a_j, b_j$  and  $a$  independently of  $\lambda_1, \lambda_2$  ( eg.  $b_2$  controls how the constants behave as  $\theta \rightarrow 0$  ). The key requirements of the  $\mathcal{D}_{\lambda_1 \lambda_2}$  are readily apparent. For by the triangle inequality (taking  $r \geq 1$ , for example), Main Theorem II ensures that

$$\|\mathcal{C}_{P_\theta}(f, g)\|_r \leq C \left( \sum_{\lambda_1, \lambda_2} \left( \sup_{k,n} |c_{kn}(\lambda_1, \lambda_2)| \right) \|\mathcal{D}_{\lambda_1 \lambda_2}\|_{op} \right) \|f\|_p \|g\|_q.$$

Now (1.2) will guarantee that  $\sup_{k,n} |c_{kn}(\lambda_1, \lambda_2)|$  decays as fast as any polynomial in  $\lambda_1, \lambda_2$ , while Main Theorem II controls  $\|\mathcal{D}_{\lambda_1 \lambda_2}\|_{op}$ . In diagonalizing  $\mathcal{C}_{P_\theta}$ , therefore, it will be crucial to ensure that  $\|\mathcal{D}_{\lambda_1 \lambda_2}\|_{op}$  increases no faster than some *fixed* polynomial in  $\lambda_1, \lambda_2$ . It is here that translation in time plays a key role. Let  $\pi(a) : f(x) \longrightarrow a^{1/2} f(ax)$ ,  $a > 0$  denote the unitary action of dilation on  $L^2(\mathbb{R})$ . Dilation eliminates the  $b_j$  from the wave packets in  $\mathcal{D}^{(\varphi)}$  and we have that Main Theorem II yields:

**Theorem 1.7.** *The operator  $\mathcal{D}^{(\varphi)}$  above associated with wave packets*

$$\varphi_{k\ell n}^{(j)}(x) = s^{k/2} \varphi^{(j)}(s^k x - a_j \ell) e^{2\pi i s^k b_j n x}$$

*is bounded from  $\ell^\infty \times L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$ ,  $1/p + 1/q = 1/r < 3/2$  provided the Fourier support intervals  $w^{(j)}$  of the dilates  $\pi(1/b_j)\varphi^{(j)}$  have pairwise-disjoint closure. Furthermore, the operator norm of  $\mathcal{D}^{(\varphi)}$  satisfies the inequality*

$$\|\mathcal{D}^{(\varphi)}\|_{op} \leq \text{const. } P(\|\varphi^{(1)}\|, \|\varphi^{(2)}\|, \|\varphi^{(3)}\|)$$

*for some polynomial  $P$  depending only on  $a_j, b_j, \rho$  and the  $w^{(j)}$ .*

There are two crucial points to note.

- *The choice above forces the  $\varphi^{(j)}$  to have the same Fourier support interval as  $\psi^{(j)}$  for each  $j$ , independently of  $\lambda_1, \lambda_2$ . In turn this guarantees that the dilates  $\pi(1/b_j)\varphi^{(j)}$  too have Fourier support intervals independent of  $\lambda_1, \lambda_2$  for each  $j$ .*
- *The construction also ensures that the  $\varphi^{(j)}$  have disjoint Fourier support intervals which remain disjoint after dilation  $\varphi^{(j)} \longrightarrow \phi^{(j)} = \pi(1/b_j)\varphi^{(j)}$ , guaranteeing that theorem (1.7) above can be applied to each  $\mathcal{D}_{\lambda_1 \lambda_2}^{(\varphi)}$  to obtain boundedness.*

There is a corresponding representation of  $\overline{\mathcal{C}}_{P_\theta}$ . Some changes in the geometry are necessary due to the presence of the term  $\xi - \eta$  in  $\overline{\mathcal{C}}_{P_\theta}$ . Granted these, theorem (1.5) follows quickly and we are left to prove Main Theorem II.

## 2. Outline of the proof of Main Theorem II

The proof of Main Theorem II proceeds by reducing a general time-frequency paraproduct into ever more simple cases. Underlying a time-frequency paraproduct is an essential structural invariance in translation, modulation and dilation coming from the Schrödinger representation of the so-called Affine-Weyl-Heisenberg group (*cf.* [6]). By applying the same affine transformation in frequency to all the  $\phi^{(j)}$ , hence preserving disjointness of their Fourier support intervals, a general time-frequency paraproduct is represented as a finite sum of ones in which

- (i)  $s = 2^K$  for some  $K$  which we are free to specify, and
- (ii) the  $w^{(j)}$  all lie in some interval  $(\alpha, \alpha + \frac{1}{2})$ ,  $|\alpha| < \frac{1}{2}$ , which either contains the origin or is contained in  $(0, 1)$ .

Moreover the three  $w^{(j)}$  can be assumed to lie inside one of the basic intervals :

$$(0, 1), \quad M = 1; \quad \left( -\frac{2^{M-1} - 1}{2^M - 1}, \frac{2^{M-1} - 1}{2^M - 1} \right), \quad M > 1$$

which generate respective grids  $\mathcal{W}_M$  in  $\mathbb{R}$  via affine transformations  $\xi \longrightarrow 2^{Mk}\xi + n$ . The value of  $K$  is specified in terms of the separation of the  $w^{(j)}$ ; more precisely,  $s = 2^{MN}$  where  $N$  is chosen so large that in case  $M = 1$  there is at least one interval in  $\mathcal{W}_1$  of length  $1/2^N$  between adjacent  $w^{(j)}$  as well as one

between each end-point of  $[0, 1)$  and the nearest  $w^{(j)}$ , while in case  $M > 1$  there are corresponding intervals in  $\mathcal{W}_M$  of length  $\sim 1/2^{MN}$ . Hence in proving Main Theorem II it is enough to begin with time-frequency paraproduct

$$\mathcal{D}(f, g) = \sum_{k, \ell, n = -\infty}^{\infty} c_{k\ell n} 2^{MNk/2} \langle f, \phi_{k\ell n}^{(1)} \rangle \langle g, \phi_{k\ell n}^{(2)} \rangle \phi_{k\ell n}^{(3)}$$

where  $M$  is determined by which of the intervals above contains all the Fourier support intervals  $w^{(j)}$  and

$$\phi_{k\ell n}^{(j)}(x) = s^{k/2} \phi^{(j)}(s^k x - a_j \ell) e^{2\pi i s^k x n}, \quad s = 2^{MN}.$$

Such a time-frequency paraproduct will be said to be  $(M, N)$ -canonical form. The link of the Fourier support intervals with grid structures in frequency is crucial. We prove:

**Theorem 2.1.** *A time-frequency paraproduct*

$$\mathcal{D}(f, g) = \sum_{Q \in \mathbb{Q}_{M,N}^{(+)}} c_Q \frac{1}{\sqrt{|I_Q|}} \langle f, \phi_Q^{(1)} \rangle \langle g, \phi_Q^{(2)} \rangle \phi_Q^{(3)}$$

in  $(M, N)$ -canonical form is bounded as an operator from  $\ell^\infty \times L^p(\mathbb{R}) \times L^q(\mathbb{R})$  into  $L^r(\mathbb{R})$ , whenever  $1/p + 1/q = 1/r < 3/2$  and  $p, q > 1$ . Its operator norm satisfies the inequality

$$\|\mathcal{D}\|_{op} \leq \text{const. } P(\|\phi^{(1)}\|, \|\phi^{(2)}\|, \|\phi^{(3)}\|)$$

for some polynomial  $P$  depending only on  $a_j, \rho$  and the Fourier support intervals  $w^{(j)}$ .

The tiles  $Q \sim \{k, l, n\} \in \mathbb{Q}_{M,N}$  are defined via the affine transformations in frequency

$$\tau_Q : [0, 1) \longrightarrow w_Q, \quad M = 1; \quad \tau_Q : [-\alpha_M, \alpha_M) \longrightarrow w_Q, \quad M > 1;$$

ie.  $\tau_Q(\xi) = s^k(\xi + n)$ ,  $Q \sim \{k, l, n\}$ . The intervals  $w_Q^{(j)} = \tau_Q(w^{(j)})$  are then the Fourier support intervals of the wave packets  $\phi_Q^{(j)}$  and their geometric properties are fundamental to the restriction to time-frequency paraproducts in  $(M, N)$ -canonical form. By  $\mathbb{Q}_{M,N}^{(+)}$  we have denoted those  $Q \in \mathbb{Q}_{M,N}$  with  $n > 0$ . Hence  $\mathbb{Q}_{M,N} = \mathbb{Q}_{M,N}^{(-)} \cup \mathbb{Q}_{M,N}^{(0)} \cup \mathbb{Q}_{M,N}^{(+)}$ .

**Outline of the proof of Theorem 2.1.** The proof of Theorem (2.1) relies on a careful study of the phase plane associated with  $\mathcal{D}$ . Given  $\delta > 0$ ,  $\delta$  small, choose  $p, q > 1$  so that  $1/2 + 2\delta < 1/p + 1/q < 3/2 - 2\delta$ ,  $|1/p - 1/q| < 1/2 - 2\delta$ . The lower bound is needed to secure convergence of various geometric series occurring in the proof and is removed later using interpolation in exploiting the

symmetry and adjoint properties of the family of all  $\mathcal{D}$ 's. The upper bound is needed *solely* to prove the error estimate (2.2) below.

Set  $p_0 = \max\{p, p'\}$ ,  $q_0 = \max\{q, q'\}$  so that

$$\frac{1}{2} + 2\delta < 1/r_0 = 1/p_0 + 1/q_0 < \frac{3}{2} - 2\delta.$$

Now fix  $f \in L^p(\mathbb{R})$ ,  $g \in L^q(\mathbb{R})$  and  $\{c_Q\} \in l^\infty$ ; without loss of generality we assume  $\|\{c_Q\}\|_\infty = 1$ . The goal is to establish the weak type estimate

$$|\{x : |\mathcal{D}(f, g)(x)| \geq 2\gamma\}| \leq \text{const.} \left( \frac{\|f\|_p \|g\|_q}{\gamma} \right)^r, \quad \gamma > 0$$

with  $1/r = 1/p + 1/q$  as usual. The first step in the proof is reminiscent of the familiar Calderón-Zygmund decomposition. Fix a small  $\eta > 0$  to be specified later depending on the earlier choice of  $\delta$  and  $r_0$ . Set

$$E_{bad} = \{x : M_p(M(f)(x)) > s^{-1/\eta} \kappa_p\} \cup \{x : M_q(M(g)(x)) > s^{-1/\eta} \kappa_q\},$$

where

$$\kappa_p = \left( \frac{\|f\|_p^{1/q} \gamma^{1/p}}{\|g\|_q^{1/p}} \right)^r, \quad \kappa_q = \left( \frac{\|g\|_q^{1/p} \gamma^{1/q}}{\|f\|_p^{1/q}} \right)^r.$$

With these choices

$$|E_{bad}| \leq \text{const.} \left( \frac{\|f\|_p \|g\|_q}{\gamma} \right)^r.$$

As a function,  $\mathcal{D}(f, g) = \mathcal{D}_{bad}(f, g) + \mathcal{D}_{good}(f, g)$  decomposes into 'bad' and 'good' functions setting

$$\mathcal{D}_{bad}(f, g) = \sum_{I_Q \subseteq E_{bad}} \frac{1}{\sqrt{|I_Q|}} c_Q \langle f, \phi_Q^{(1)} \rangle \langle g, \phi_Q^{(2)} \rangle \phi_Q^{(3)},$$

The Hardy-Littlewood maximal function uniformly controls wave packet coefficients of  $f$  and  $g$ . Thus removal of all tiles with  $I_Q \subseteq E_{bad}$  ensures that the coefficients in

$$\mathcal{D}_{good}(f, g) = \sum_{I_Q \not\subseteq E_{bad}} \frac{1}{\sqrt{|I_Q|}} c_Q \langle f, \phi_Q^{(1)} \rangle \langle g, \phi_Q^{(2)} \rangle \phi_Q^{(3)}$$

satisfy uniform bounds. On the other hand, the  $\phi^{(i)}$  appearing in  $\mathcal{D}_{bad}(f, g)$  are 'concentrated' inside  $E_{bad}$ , so the bad function can be estimated sufficiently far away from  $E_{bad}$  using solely decay estimates on the  $\phi^{(i)}$  and Hausdorff-Young inequalities. Set  $E_1 = \bigcup_{I_Q \subseteq E_{bad}} s^2 I_Q$ .

**Theorem 2.2.** *The inequalities  $|E_1| \leq \text{const.} |E_{bad}|$  and*

$$\frac{1}{\gamma} \int_{\mathbb{R} \setminus E_1} |\mathcal{D}_{bad}(f, g)(x)| dx \leq \text{const.} |E_{bad}|$$

*hold uniformly in  $f, g$  and  $\gamma$  as well as the  $a_j$ .*

Clearly then

$$|\{x : |\mathcal{D}_{bad}(f, g)(x)| \geq \gamma\}| \leq \text{const.} \left( \frac{\|f\|_p \|g\|_q}{\gamma} \right)^r,$$

leaving only the proof of the corresponding estimate for  $\mathcal{D}_{good}(f, g)$ . This requires a very delicate decomposition of the ‘good’ function into the sum of functions associated with ‘trees’ of tiles defined using the partial order  $Q \leq Q' \iff I_Q \subseteq I_{Q'}, w_Q \supseteq w_{Q'}$  on  $\mathbb{Q}_{M,N}^{(+)}$ . A *tree*  $\mathbb{T}$  is set of tiles containing a tile  $\overline{Q}$  which is maximal in the sense that  $Q \in \mathbb{T} \iff Q \leq \overline{Q}$ . This maximal tile will be called the *tree-top* of  $\mathbb{T}$  and will often be denoted by  $I_{\mathbb{T}} \times w_{\mathbb{T}}$  to emphasize its dependence on  $\mathbb{T}$ . To each tree there corresponds a Carleson box or a tent in the usual upper half-plane and so there are intimate connections between trees and Tent spaces. The role of a tree, however, is to control in an efficient manner the oscillatory behaviour that an otherwise random group of tiles in phase-plane has. To illustrate this consider the *tree operator*

$$f, g \longrightarrow \mathcal{D}_{\mathbb{T}}(f, g) = \sum_{Q \in \mathbb{T}} c_Q \frac{1}{\sqrt{|I_Q|}} \langle f, \phi_Q^{(1)} \rangle \langle g, \phi_Q^{(2)} \rangle \phi_Q^{(3)}$$

obtained by summing only over tiles in  $\mathbb{T}$ , and suppose  $M = 1$ . For each tile  $Q \sim \{k, \ell, n\}$  in  $\mathbb{T}$  the tree structure ensures that  $n = [s^{-k} \lambda_{\mathbb{T}}]$ , where  $\lambda_{\mathbb{T}}$  is the left-hand endpoint of  $w_{\mathbb{T}}$ . After suitable conjugations by  $e^{2\pi i x \lambda_{\mathbb{T}}}$ , therefore,  $\mathcal{D}_{\mathbb{T}}$  can be rewritten in terms of modulated wave-packets all having roughly the *same oscillation* and hence  $\mathcal{M}_{\mu}$ -norm which is uniform in  $\mathbb{T}$ . To be precise their frequency satisfy the inequality  $0 \leq s^{-k} \lambda_{\mathbb{T}} - [s^{-k} \lambda_{\mathbb{T}}] < 1$ . A tree operator is thus a ‘standard’ paraproduct modulated by a single exponential  $e^{2\pi i x \lambda_{\mathbb{T}}}$ . Familiar techniques now produce  $L^2$ -norm estimates for  $\mathcal{D}_{\mathbb{T}}$  which are independent of  $\lambda_{\mathbb{T}}$  provided at least two of the modulated wave-packets  $\psi^{(i)}(x) = \phi^{(i)}(x) e^{2\pi i x (s^{-k} \lambda_{\mathbb{T}} - [s^{-k} \lambda_{\mathbb{T}}])}$  have vanishing moments. But  $s^{-k} \lambda_{\mathbb{T}} - [s^{-k} \lambda_{\mathbb{T}}] \notin w^{(i)} \implies$

$$\int_{-\infty}^{\infty} \psi^{(i)}(x) dx = \widehat{\phi}^{(i)}(s^{-k} \lambda_{\mathbb{T}} - [s^{-k} \lambda_{\mathbb{T}}]) = 0,$$

so we need to know that  $s^{-k} \lambda_{\mathbb{T}} - [s^{-k} \lambda_{\mathbb{T}}]$  fails to belong to at least two of the  $w^{(i)}$ . This, however, is exactly what disjointness of the Fourier support intervals guarantees. A corresponding argument applies in the case  $M \geq 2$ , setting  $\lambda_{\mathbb{T}} = \tau_{\overline{Q}}(0)$ . Hence we can also view this grouping of tiles into trees as

an ‘efficient localization’ in phase plane, for which the origin becomes once again a ‘distinguished’ point in frequency, in the sense that locally, *i.e.*, on each tree, Littlewood-Paley theory applies. The idea now is to choose families of trees. For each  $\nu \geq 0$ , we define nested families  $\{\mathbb{Q}_\nu\}$ ,

$$\emptyset \subseteq \dots \subseteq \mathbb{Q}_\nu \subseteq \mathbb{Q}_{\nu-1} \subseteq \dots \subseteq \mathbb{Q}_{-1} = \{Q \in \mathbb{Q}_{M,N}^{(+)} : I_Q \not\subseteq E_{bad}\}$$

of tiles recursively by choosing families  $\mathcal{F}_\nu = \bigcup_{i,j} \mathcal{F}_\nu^{(ij)}$  of trees so that

$$\mathbb{Q}_{\nu-1} \setminus \mathbb{Q}_\nu = \bigcup_{\mathbb{T} \in \mathcal{F}_\nu} \{Q : Q \in \mathbb{T}\}$$

We summarize the properties of  $\mathbb{Q}_\nu$  that follow immediately from the  $\nu$ -th stage construction. We list them for  $f$  but analogous ones hold for  $g$  with  $\phi_Q^{(1)}$  replaced by  $\phi_Q^{(2)}$ ,  $p_0$  by  $q_0$  and  $j \neq 2$  instead of  $j \neq 1$ . They are refinements of *a priori* estimates at the first stage, *ie.* in  $\mathbb{Q}_{-1}$

**Properties of  $\mathbb{Q}_\nu$ .** (i) For all  $Q$  in  $\mathbb{Q}_\nu$ ,

$$\frac{1}{\sqrt{|I_Q|}} |\langle f, \phi_Q^{(1)} \rangle| \leq \text{const}_\phi s^{-(1+\eta)(1+\nu)/p_0} s^{-1/\eta} \kappa_p$$

(ii) The inequality

$$\frac{1}{|I_\mathbb{T}|} \int_{-\infty}^{\infty} \left( \sum_{Q \in \mathbb{T}} \frac{1}{|I_Q|} |\langle f, \phi_Q^{(1)} \rangle|^2 \chi_{I_Q}(x) \right)^{1/2} dx \leq \text{const}_\phi \kappa_p s^{-(\nu+1)/p_0}$$

holds for all  $\Lambda^{(j)}$ -trees in  $\mathbb{Q}_\nu$ ,  $j \neq 1$ .

One remarkable consequence of this construction is that (ii) above remains valid for any interval  $J$  in  $I_\mathbb{T}$ , not just for  $I_\mathbb{T}$  itself, leading to a *Carleson measure type estimate*.

Then

$$\mathcal{D}_{good}(f, g) = \sum_{\nu=0}^{\infty} \left( \sum_{\mathbb{T} \in \mathcal{F}_\nu} \mathcal{D}_\mathbb{T}(f, g) \right), \quad \mathcal{F}_\nu = \bigcup_{i=1}^2 \left( \bigcup_{j=1}^3 \mathcal{F}_\nu^{(ij)} \right)$$

provides the desired decomposition. Note that there will be three different classes of trees, each specifying which two of the three wave-packets  $\phi_Q^{(i)}$ ,  $i = 1, 2, 3$ , have vanishing moments *uniformly* for tiles  $Q$  in that tree. All the difficulty comes in establishing  $L^2$ -estimates for  $\mathcal{D}_{good}$ .

Ideally, what one really wants is that each  $\mathcal{D}_\mathbb{T}(f, g)$  be an  $L^2$ -function and that pairs of such functions be ‘almost orthogonal’. So, armed with the Fourier support condition and the vanishing moment conditions available for each tree we prove:

(A) an  $L^2$ -norm estimate

$$\left( \frac{1}{\gamma^2} \int_{-\infty}^{\infty} |\mathcal{D}_{\mathbb{T}}(f, g)(x)|^2 dx \right)^{1/2} \leq \text{const. } s^{-\nu/r_0} |I_{\mathbb{T}}|^{1/2}$$

for each tree  $\mathbb{T}$  in  $\mathcal{F}_{\nu}$  and

(B) an  $L^{\sigma}$ -norm estimate for every  $1 \leq \sigma < \infty$

$$\left( \int_{-\infty}^{\infty} N_{\overline{\mathcal{F}}_{\nu}}(x)^{\sigma} dx \right)^{1/\sigma} \leq \text{const. } s^{(1+2\delta)\nu} \left( \frac{\|f\|_p \|g\|_q}{\gamma} \right)^{r/\sigma}$$

for the function  $N_{\overline{\mathcal{F}}_{\nu}} = N_{\overline{\mathcal{F}}_{\nu}}(x)$  counting the number of trees in  $\overline{\mathcal{F}}_{\nu}$  above  $x$  -where  $\overline{\mathcal{F}}_{\nu}$  is a suitable truncation of  $\mathcal{F}_{\nu}$ .

This counting function controls most aspects of the rest of the proof as it captures the interactions among trees. It enables us to sum ‘almost orthogonal’ tree *functions* in much the same spirit as almost orthogonal operators are summed in the Cotlar-Stein lemma. In the case of just one tree, for instance, it provides the  $L^2$ -bound

$$(\dagger) \quad \frac{1}{\gamma^2} \int_{-\infty}^{\infty} |\mathcal{D}_{\mathbb{T}}(f, g)(x)|^2 dx \leq \text{const. } s^{-2\delta\nu} \left( \frac{\|f\|_p \|g\|_q}{\gamma} \right)^r.$$

If this estimate for a single tree could be replaced by the sum over trees then the companion estimate

$$(\dagger\dagger) \quad |\{x : |\mathcal{D}_{good}(f, g)(x)| > \gamma\}| \leq \text{const.} \left( \frac{\|f\|_p \|g\|_q}{\gamma} \right)^r$$

to the one for the ‘bad’ function would follow immediately. Our approach has to be less direct, however, though it is the same in principle. We adopt the strategy Fefferman used at the corresponding point of his pointwise convergence proof ([5]):

- (a) ‘thin out’ the trees in  $\mathcal{F}_{\nu}$ , and seek families of new trees to be called *forests*;
- (b) decompose the ‘thinned’  $\mathcal{F}_{\nu}$  into  $O(\nu)$  forests whose trees still satisfy (A) and whose counting function satisfies the same  $L^{\sigma}$ -estimate (B) ;
- (c) ‘trim’ the new trees in each forest so that an estimate like  $(\dagger)$  holds now for the sum of trees in a forest ;
- (d) estimate the error terms created by this double pruning process.

Consequently, if we denote by  $\mathbb{S}^{trim}$  the new trees left after trimming then

$$\mathcal{D}_{good}(f, g) = \mathcal{D}_{dense}(f, g) + \mathcal{D}_{edge}(f, g) + \sum_{\nu=0}^{\infty} \left[ \sum_{n=1}^{O(\nu)} \left( \sum_{\mathbb{S} \in \mathcal{W}_n^{(\nu)}} \mathcal{D}_{\mathbb{S}^{trim}}(f, g) \right) \right]$$

where the error terms  $\mathcal{D}_{dense}(f, g)$  and  $\mathcal{D}_{edge}(f, g)$  are defined by summing respectively over tiles associated to the wave packets ‘concentrated’ in ‘leftover’ sets  $E_{dense}$  and  $E_{edge}$ . The counting function estimate ensures that once again there are estimates entirely analogous to (2.2) for these error terms after introducing exceptional sets defined from  $E_{dense}$  and  $E_{edge}$  in the same manner  $E_1$  was from  $E_{bad}$ . Hence the proof is reduced to establishing the following

**Theorem (Forest Estimate).** *The inequality*

$$\frac{1}{\gamma^2} \int_{-\infty}^{\infty} \left| \sum_{\mathbb{S} \in \mathcal{W}_n^{(\nu)}} \mathcal{D}_{\mathbb{S}^{trim}}(f, g)(x) \right|^2 dx \leq \text{const. } s^{-2\delta\nu} \left( \frac{\|f\|_p \|g\|_q}{\gamma} \right)^r$$

holds uniformly in  $f, g, \gamma$  and forest  $\mathcal{W}_n^{(\nu)}$ .

Combining all the previous estimates we finally deduce ( $\dagger\dagger$ ); thereby completing the proof of theorem (2.1) and hence of Main Theorem II.

### References

1. L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135–157.
2. R. R. Coifman, P. L. Lions, Y. Meyer and S. Semmes, *Compensated compactness and Hardy spaces*, J. Math. Pures Appl. **72** (1993), 247–286.
3. R. R. Coifman and Y. Meyer, *Au delà des opérateurs pseudo-différentiels*, Astérisque **57**, Société Mathématique de France (1978).
4. R. R. Coifman and Y. Meyer, *Fourier analysis of multilinear convolutions, Calderón’s theorem and analysis of Lipschitz curves*, Lecture Notes in Math. **779** (1979), 104–122.
5. C. Fefferman, *Pointwise convergence of Fourier series*, Ann. of Math. **98** (1973), 551–571.
6. G. B. Folland, *Annals of Mathematics Studies* **122** (1989), Princeton Univ. Press.
7. J. E. Gilbert, and A. R. Nahmod, *Hardy spaces and a Walsh model for Bilinear Cone Operators*, Trans. Amer. Math. Soc. **351** (1999), 3267–3300.
8. J. E. Gilbert, and A. R. Nahmod, *Bilinear Operators with Non-Smooth Symbol*, preprint (1999).
9. J. E. Gilbert and A. R. Nahmod,  *$L^p$ -Boundedness of Time-Frequency Paraproducts*, preprint (1999).
10. L. Grafakos, and R. H. Torres, *Multilinear Calderón-Zygmund Theory*, preprint (1999).
11. C. E. Kenig and E. M. Stein, *Multilinear estimates and fractional integration*, Math. Res. Lett. **6** (1999), 1–15.
12. M. Lacey, *The Bilinear Maximal function maps in  $L^p$  for  $2/3 < p \leq 1$* , Ann. of Math (2) **151** (2000), 35–57.
13. M. Lacey, and C. Thiele,  *$L^p$  estimates on the bilinear Hilbert transform,  $2 < p < \infty$* , Annals of Math. **146** (1997), 693–724.
14. M. Lacey, and C. Thiele, *On Calderón’s conjecture*, Annals of Math. **149** (1999), 475–496.

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF TEXAS AT AUSTIN, AUSTIN, TX 78712-1082.

*E-mail address:* gilbert@math.utexas.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, LEDERLE GRT, UNIVERSITY OF MASSACHUSETTS, AMHERST, MA 01003-4515.

*E-mail address:* nahmod@math.umass.edu