MULTIPARAMETER RIESZ COMMUTATORS

By Michael T. Lacey, Stefanie Petermichl,
Jill C. Pipher, and Brett D. Wick

Abstract. It is shown that product BMO of S.-Y. A. Chang and R. Fefferman, defined on the space $\mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_t}$, can be characterized by the multiparameter commutators of Riesz transforms. This extends a classical one-parameter result of R. Coifman, R. Rochberg, and G. Weiss, and at the same time extends the work of M. Lacey and S. Ferguson and M. Lacey and E. Terwilleger, on multiparameter commutators with Hilbert transforms.

1. Introduction. In one parameter, a classical result of Coifman, Rochberg and Weiss [8], in turn an extension of the result of Nehari [26], shows that a function in the Hardy space H^1 on the ball can be weakly factored as a sum of products of functions in H^2 on the ball. Recently, Ferguson and Lacey [12] and Lacey and Terwilleger [19] proved the corresponding weak factorization for H^1 of the polydisc. In this paper, we prove the real variable generalization of these two sets of results.

Let $\mathrm{M}_b\varphi \stackrel{\text{def}}{=} b \cdot \varphi$ be the operator of pointwise multiplication by a function b. For Schwartz functions f on $\mathbb{R}^d$, let $\mathrm{R}_j f$ denote the jth Riesz transform of f, for $1 \le j \le d$. From time to time, we will use the notation R_0 for the identity operator.

We are concerned with product spaces $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_t}$ for vectors $\vec{d} = (d_1, \ldots, d_t) \in \mathbb{N}^t$. For Schwartz functions b,f on $\mathbb{R}^{\vec{d}}$, and for a vector $\vec{j} = (j_1, \ldots, j_t)$ with $1 \le j_s \le d_s$ for $s = 1, \ldots, t$ we consider the family of commutators

$$\mathrm{C}_{\vec{j}}(b,f) \stackrel{\text{def}}{=} [\cdots[[\mathrm{M}_b, \mathrm{R}_{1,j_1}], \mathrm{R}_{2,j_2}], \cdots], \mathrm{R}_{t,j_t}]f \tag{1.1}$$

where $\mathrm{R}_{s,j}$ denotes the jth Riesz transform acting on $\mathbb{R}^{d_s}$.

Manuscript received April 28, 2007.

Research of the first author supported in part by a National Science Foundation Grant. The author is a Guggenheim Fellow; research of the second author supported in part by a National Science Foundation Grant; research of the third author supported in part by a National Science Foundation Grant; research of the fourth author supported in part by a National Science Foundation RTG Grant to Vanderbilt University.

American Journal of Mathematics 131 (2009), 731–769.

MAIN THEOREM. *We have the estimates below, valid for* $1 < p < \infty$.

$$\sup_{\vec{j}} \left\| C_{\vec{j}}(b,\cdot) \right\|_{p\to p} \simeq \|b\|_{\mathrm{BMO}}\,. \tag{1.2}$$

By BMO*, we mean Chang–Fefferman* BMO.

To establish this result, we find it necessary to prove an extended version of this Theorem, proving the equivalence of norms not only for the Riesz transforms, but also a class of singular integral operators whose symbols are supported on cones. In this, and other ways, our methods shed new light on issues related to commutators even in the one parameter case.

It is well known that the result above has an equivalent formulation in terms of *weak factorization* of Hardy space; indeed, this equivalence is important to the proof of the Theorem. For $\vec{j}$ a vector with $1 \le j_s \le d_s$, and $s = 1, \dots, t$, let $\Pi_{\vec{j}}$ be the bilinear operator defined by the following equation

$$\langle C_{\vec{j}}(b,f), g\rangle \stackrel{\mathrm{def}}{=} \langle b, \Pi_{\vec{j}}(f,g)\rangle.$$

One can express $\Pi_{\vec{j}}$ as a linear combination of products of iterates of Riesz transforms, R_{s,j_s}, applied to the f and g. It follows immediately by duality from the Main Theorem that for sequences $f_k^{\vec{j}}, g_k^{\vec{j}} \in L^2(\mathbb{R}^{\vec{d}})$ with $\sum_{\vec{j}} \sum_{k=1}^{\infty} \left\|f_k^{\vec{j}}\right\|_2 \left\|g_k^{\vec{j}}\right\|_2 < \infty$ we have

$$\sum_{\vec{j}} \sum_{k=1}^{\infty} \Pi_{\vec{j}}(f_k^{\vec{j}}, g_k^{\vec{j}}) \in H^1(\mathbb{R}^{\vec{d}}).$$

With this observation, we define

$$L^2(\mathbb{R}^{\vec{d}}) \widehat{\odot} L^2(\mathbb{R}^{\vec{d}}) \stackrel{\mathrm{def}}{=} \left\{ f \in L^1(\mathbb{R}^{\vec{d}}) \colon f = \sum_{\vec{j}} \sum_{k=1}^{\infty} \Pi_{\vec{j}}(f_k^{\vec{j}}, g_k^{\vec{j}}) \right\}. \tag{1.3}$$

This is the projective product given by

$$\|f\|_{L^2(\mathbb{R}^{\vec{d}})\widehat{\odot}L^2(\mathbb{R}^{\vec{d}})} \stackrel{\mathrm{def}}{=} \inf \left\{ \sum_{\vec{j}} \sum_{k} \left\|f_k^{\vec{j}}\right\|_2 \left\|g_k^{\vec{j}}\right\|_2 \right\}$$

where the infimum is taken over all decompositions of f as in (1.3). This definition has an obvious extension to $L^p \widehat{\odot} L^{p'}$, for $1 \le p, p' < \infty$. We have the following corollary.

WEAK FACTORIZATION THEOREM. *For any* $1 < p < \infty$*, let* $p' = p/(p-1)$ *be the conjugate index. We have* $H^1(\mathbb{R}^{\vec{d}}) = L^p(\mathbb{R}^{\vec{d}}) \widehat{\odot} L^{p'}(\mathbb{R}^{\vec{d}})$*. Namely, for any* $f \in H^1(\mathbb{R}^{\vec{d}})$

there exist sequences $f_k^{\vec{\jmath}} \in L^p(\mathbb{R}^{\vec{d}})$ *and* $g_k^{\vec{\jmath}} \in L^{p'}(\mathbb{R}^{\vec{d}})$ *such that*

$$f = \sum_{\vec{\jmath}} \sum_{k=1}^{\infty} \Pi_{\vec{\jmath}}(f_k^{\vec{\jmath}}, g_k^{\vec{\jmath}}),$$

$$\|f\|_{H^1} \simeq \sum_{\vec{\jmath}} \sum_{k} \left\|f_k^{\vec{\jmath}}\right\|_p \left\|g_k^{\vec{\jmath}}\right\|_{p'}.$$

The result of Coifman, Rochberg and Weiss [8] has found a number of further applications. The original paper includes a weak factorization result for certain Bergman spaces, and there is the striking application to the theory of compensated compactness [7]. We anticipate that some of these applications persist into the higher parameter setting of this paper, but we do not purse these points in this paper.

The proofs given here are rather different from that of Coifman, Rochberg and Weiss [8]. Their proof of the upper bound on commutator norms relies upon a sharp function inequality, a method of proof that is quite powerful in the one parameter setting. This method admits only a weak extension to the higher parameter setting; instead, our proof of the upper bound, namely Theorem 5.3, follows from the decomposition of the commutators into a sum of simpler terms. These terms are paraproducts, composed on either side by Calderón–Zygmund operators. This method has been used in different settings, such as Petermichl [27], and Lacey [18]. Our formalization of this method in this paper could lead to further applications of this method.

The paraproducts that arise are of multiparameter form. The specific result needed is Theorem 4.4 below. This result is due to Journé [13]; more recent discussions of paraproducts are in [17, 24, 25].

For the lower bound, namely Theorem 6.1, we use the strategy of Ferguson–Lacey [12] and Lacey–Terwilleger [19]. One inducts on parameters, using a bootstrapping argument, and the Journé Lemma [14]. However, to implement this strategy, we have to prove a second version of the Main Theorem, one in which the Riesz transforms are replaced by the a family of Calderón–Zygmund operators whose symbols are adapted to cones. These kernels are described in the text preceding (5.8), and Theorem 5.10 is the extended version of our Main Theorem.

To start the induction, in the case of Riesz transforms, we can of course use the Coifman, Rochberg, and Weiss result. But for the cones, we appeal to the results of Uchiyama [29] and Song-Ying Li [21] which are deep extensions of the work of Coifman, Rochberg and Weiss.

In §2–4 we recall different aspects of the multiparameter theory in forms required for this investigation. §5 introduces the cone operators, and this establishes the upper bound on commutator norms. The initial stages of the lower bound on commutator norms is proved in §6. The more refined bootstrapping argument occupies §7.

Acknowledgments. The authors benefited from the Banff Research Station, through the Research in Teams program, and a very pleasant stay at the Université de Bordeaux. The authors also enjoyed a stay at Texas A&M University.

2. Wavelets in several dimensions. This discussion is initially restricted to a one parameter setting. We will use dilation and translation operators on $\mathbb{R}^d$

$$\mathrm{Tr}_y f(x) \stackrel{\text{def}}{=} f(x-y), \quad y \in \mathbb{R}^d, \tag{2.1}$$

$$\mathrm{Dil}_a^{(p)} f(x) \stackrel{\text{def}}{=} a^{-d/p} f(x/a), \quad a > 0,\ 0 < p \le \infty. \tag{2.2}$$

These will also be applied to sets, in an obvious fashion, in the case of $p = \infty$.

By the *(d dimensional) dyadic grid* in $\mathbb{R}^d$ we mean the collection of cubes

$$\mathcal{D}_d \stackrel{\text{def}}{=} \left\{ j2^k + [0,2^k)^d \colon\ j \in \mathbb{Z}^d,\ k \in \mathbb{Z} \right\}.$$

Wavelets arise from a *scaling function* or a *father wavelet* W. The principle requirement is that the functions $\{\mathrm{Tr}_{c(I)}\mathrm{Dil}_I^{(2)} w \colon\ I \in \mathcal{D}_1\}$ form an orthonormal basis for $L^2(\mathbb{R})$. Except for the fact that it is not smooth, $h = -\mathbf{1}_{(0,1/2)} + \mathbf{1}_{(1/2,1)}$ is a scaling function, with father wavelet $\mathbf{1}_{(0,1)}$. This generates the Haar basis for $L^2(\mathbb{R})$.

For $\varepsilon \in \{0,1\}$, set $w^0 = w$ and $w^1 = W$, the superscript ^{0} denoting that "the function has mean 0," while a superscript ^{1} denotes that "the function is an L^2 normalized indicator function." In one dimension, for an interval I, set

$$w_I^\varepsilon \stackrel{\text{def}}{=} \mathrm{Tr}_{c(I)}\mathrm{Dil}_{|I|}^{(2)} w^\varepsilon.$$

The father wavelet is of some convenience to us, as we have the useful facts, valid on the interval J. (Technically, these results are only true for *multiresolution analysis (MRA)* wavelets; both the Haar and Meyer wavelets are MRA wavelets.)

$$\sum_{I \supsetneq J} \langle f, w_I \rangle w_I = \langle f, w_J^1 \rangle w_J^1. \tag{2.3}$$

We will use the *Meyer wavelet* in later sections of the paper. This wavelet, found by Y. Meyer [22, 23], arises from a Schwartz wavelet w, with $\widehat{w}$ supported on $1/3 \le |\xi| \le 8/3$. Indeed, $\widehat{w}$ is identically equal to 1 on the intervals $1 \le |\xi| \le 2$. One of the reasons this is such a useful wavelet for us is the fact below which is exploited several times:

$$8\,|I| < |I'| \text{ implies } \widehat{w_I \cdot w_{I'}} \text{ is supported on } (4\,|I|)^{-1} < |\xi| < 3\,|I|^{-1}. \tag{2.4}$$

Let $\mathrm{Sig}_d \stackrel{\text{def}}{=} \{0,1\}^d - \{\vec{1}\}$, which we refer to as *signatures*. In d dimensions, for a cube Q with side $|I|$, i.e., $Q = I_1 \times \cdots \times I_d$, and a choice of $\varepsilon \in \mathrm{Sig}_d$, set

$$w_Q^{\varepsilon}(x_1, \ldots, x_d) \stackrel{\text{def}}{=} \prod_{j=1}^{d} w_{I_j}^{\varepsilon_j}(x_j).$$

It is then the case that the collection of functions

$$\mathrm{Wavelet}_{\mathcal{D}_d} \stackrel{\text{def}}{=} \{w_Q^{\varepsilon} \colon\ Q \in \mathcal{D}_d,\ \varepsilon \in \mathrm{Sig}_d\}$$

form a wavelet basis for $L^p(\mathbb{R}^d)$ for any choice of d dimensional dyadic grid $\mathcal{D}_d$. Here, we are using the notation $\vec{1} = (1, \ldots, 1)$. While we exclude the superscript $\vec{1}$ here, it plays a role in the theory of paraproducts.

We will use these bases in the tensor product setting. Thus, for a vector $\vec{d} = (d_1, \ldots, d_t)$, and $1 \leq s \leq t$, let $\mathcal{D}_{d_s}$ be a choice of d_s dimensional dyadic grid, and let

$$\mathcal{D}_{\vec{d}} = \otimes_{s=1}^{t} \mathcal{D}_{d_s}.$$

Also, let $\mathrm{Sig}_{\vec{d}} \stackrel{\text{def}}{=} \{\vec{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_t) : \varepsilon_s \in \mathrm{Sig}_{d_s}\}$. Note that each ε_s is a vector, and so $\vec{\varepsilon}$ is a "vector of vectors". For a rectangle $R = Q_1 \times \cdots \times Q_t$, being a product of cubes of possibly different dimensions, and a choice of vectors $\vec{\varepsilon} \in \mathrm{Sig}_{\vec{d}}$ set

$$w_R^{\vec{\varepsilon}}(x_1, \ldots, x_t) = \prod_{s=1}^{t} w_{Q_s}^{\varepsilon_s}(x_s).$$

These are the appropriate functions and bases to analyze multiparameter paraproducts and commutators.

Let

$$\mathrm{Wavelet}_{\mathcal{D}_{\vec{d}}} \stackrel{\text{def}}{=} \left\{ w_R^{\vec{\varepsilon}} \colon\ R \in \mathcal{D}_{\vec{d}},\ \vec{\varepsilon} \in \mathrm{Sig}_{\vec{d}} \right\}.$$

This is a basis in $L^p(\mathbb{R}^{\vec{d}})$, where we will use the notation

$$\mathbb{R}^{\vec{d}} \stackrel{\text{def}}{=} \mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_t}$$

to emphasize that we are in a tensor product setting.

3. Chang–Fefferman BMO. We describe the elements of product Hardy space theory, as developed by S.-Y. Chang and R. Fefferman [5, 6, 9–11] as well as Journé [13, 14]. By this, we mean the Hardy spaces associated with domains like $\otimes_{s=1}^{t} \mathbb{R}^{d_s}$.

Remark. The (real) Hardy space $H^1(\mathbb{R}^d)$ typically denotes the class of functions with the norm

$$\sum_{j=0}^{d} \|\mathrm{R}_j f\|_1 ,$$

where R_j denotes the jth Riesz transform. Here and below we adopt the convention that R_0, the 0th Riesz transform, is the identity. This space is invariant under the one parameter family of isotropic dilations, while $H^1(\mathbb{R}^{\vec{d}})$ is invariant under dilations of each coordinate separately. That is, it is invariant under a t parameter family of dilations, hence the terminology "multiparameter" theory.

As before, the space $H^1(\mathbb{R}^{\vec{d}})$ has a variety of equivalent norms, in terms of square functions, maximal functions and Riesz transforms. For our discussion of paraproducts, it is appropriate to make some definitions of translation and dilation operators which extend the definitions in (2.1)–(2.2). (Indeed, here we are adopting broader notation than we really need, in anticipation of a discussion of multiparameter paraproducts.) Define

$$\mathrm{Tr}_y f(x) \stackrel{\text{def}}{=} f(x-y), \qquad y \in \mathbb{R}^{\vec{d}}, \tag{3.1}$$

$$\mathrm{Dil}^p_{a_1,\dots,a_t} f(x_1,\dots,x_t) \stackrel{\text{def}}{=} \prod_{s=1}^{t} a_s^{-d_s/p} f(x_1/a_1,\dots,x_t/a_t), \tag{3.2}$$

$$a_1,\dots,a_t > 0,$$

$$\mathrm{Dil}^p_R \stackrel{\text{def}}{=} \mathrm{Tr}_{c(R)} \mathrm{Dil}^p_{|Q_1|,\dots,|Q_d|}. \tag{3.3}$$

In the last definition $R = Q_1 \times \cdots \times Q_t$ is a rectangle, each Q_s is a cube and the dilation incorporates the locations and scales associated with R. $c(R)$ is the center of R.

For a nonnegative smooth bump function φ^1 in $\mathbb{R}^{\vec{d}}$ with $\int \varphi^1 \, dx = 1$, define the (strong) maximal function by

$$\mathrm{M}f(x) \stackrel{\text{def}}{=} \sup_{R \in \mathcal{D}_{\vec{d}}} \mathrm{Dil}^2_R \varphi^1(x) \langle f, \mathrm{Dil}^2_R \varphi^1 \rangle.$$

For $s = 1,\dots,t$, choose radial bump functions φ^0_s on $\mathbb{R}^{d_s}$ with $\int_{\mathbb{R}^{d_s}} \varphi^0_s \, dx_s = 0$ and

$$\sup_{\xi} \int_0^\infty \left| \widehat{\varphi^0_s}(t\xi) \right|^2 \frac{dt}{t} < \infty.$$

Then, fix φ^0 so that

$$\varphi^0(x_1,\dots,x_t) \stackrel{\text{def}}{=} \prod_{s=1}^{t} \varphi_s^0(x_s).$$

As an analog of the Littlewood–Paley square function, set

$$\mathrm{S}f(x) \stackrel{\text{def}}{=} \left[\sum_{R\in\mathcal{D}_{\vec{d}}} [\mathrm{Dil}_R^2\varphi^0(x)]^2 \left|\langle f, \mathrm{Dil}_R^2\varphi^0\rangle\right|^2\right]^{1/2}.$$

THEOREM 3.4. (Equivalent forms of H^1 norm) *All of the norms below are equivalent, and can be used as a definition of* $H^1(\mathbb{R}^{\vec{d}})$.

$$\|\mathrm{M}f\|_1 \simeq \|f\|_1 + \|\mathrm{S}f\|_1 \simeq \sum_{\vec{0}\le\vec{j}\le\vec{d}} \left\|\prod_{s=1}^{t} \mathrm{R}_{s,j_s} f\right\|_1.$$

R_{s,j_s} *is the Riesz transform computed in the* j_s*th direction of the* s*th variable, and the* 0*th Riesz transform is the identity operator.*

3.1. BMO($\mathbb{R}^{\vec{d}}$). The dual of the real Hardy space is $H^1(\mathbb{R}^{\vec{d}})^* = \mathrm{BMO}(\mathbb{R}^{\vec{d}})$, the t-fold product BMO space. It is a Theorem of S.-Y. Chang and R. Fefferman [6] that this space has a characterization in terms of a product Carleson measure.

Define

$$\|b\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} \stackrel{\text{def}}{=} \sup_{U\subset\mathbb{R}^{\vec{d}}} \left[|U|^{-1} \sum_{R\subset U} \sum_{\vec{\varepsilon}\in\mathrm{Sig}_{\vec{d}}} \left|\langle b, w_R^{\vec{\varepsilon}}\rangle\right|^2\right]^{1/2}. \tag{3.5}$$

Here the supremum is taken over all open subsets $U \subset \mathbb{R}^{\vec{d}}$ with finite measure, and we use a wavelet basis $w_R^{\vec{\varepsilon}}$.

THEOREM 3.6. (Chang–Fefferman BMO) *We have the equivalence of norms*

$$\|f\|_{(H^1(\mathbb{R}^{\vec{d}}))^*} \approx \|f\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})}.$$

That is, $\mathrm{BMO}(\mathbb{R}^{\vec{d}})$ *is the dual to* $H^1(\mathbb{R}^{\vec{d}})$.

3.2. Journé's Lemma. The explicit definition of BMO in (3.5) is quite difficult to work with. In the first place, it is not an intrinsic definition, in that one needs some notion of wavelet to define it. Secondly, the supremum is over a very broad class of objects: All open subsets of $\mathbb{R}^{\vec{d}}$ of finite measure. There are simpler

definitions (also unfortunately not intrinsic) that in particular circumstances are sufficient.

Say that a collection of rectangles $\mathcal{U} \subset \mathcal{D}_{\vec{d}}$ *has* $t-1$ *parameters* if and only if there is a choice of coordinate s so that for all $R, R' \in \mathcal{U}$ we have $Q_s = Q'_s$, that is the sth coordinate of the rectangles are all one fixed d_s dimensional cube.

We then define

$$\|f\|_{\mathrm{BMO}_{-1}(\mathbb{R}^{\vec{d}})} \stackrel{\text{def}}{=} \sup_{\substack{\mathcal{U} \text{ has } t-1 \\ \text{parameters}}} \left[|\mathrm{sh}(\mathcal{U})|^{-1} \sum_{R \in \mathcal{U}} \sum_{\vec{\varepsilon} \in \mathrm{Sig}_{\vec{d}}} \left| \langle f, w_R^{\vec{\varepsilon}} \rangle \right|^2 \right]^{1/2} . \tag{3.7}$$

A collection of rectangles has a *shadow* given by $\mathrm{sh}(\mathcal{U}) \stackrel{\text{def}}{=} \bigcup \{R\colon\ R \in \mathcal{U}\}$. We use the "$-1$" subscript to indicate that we have "lost one parameter" in the definition. (In the two parameter case, our definition of BMO_{-1} is actually a slightly larger space than the more familiar rectangular BMO space.) Motivation for this definition comes from our use of induction on parameters in the proof of the lower bound for the commutators. See §6.1.

L. Carleson [4] produced examples of functions which acted as linear functionals on $\mathrm{H}^1(\mathbb{R}^{\vec{d}})$ with norm one, yet had arbitrarily small BMO_{-1} norm. This example is recounted at the beginning of R. Fefferman's article [9].

Journé's Lemma permits us, with certain restrictions, to dominate the BMO norm by the BMO_{-1} norm. We need a version of this statement with an additional refinement, see (3.10) that first appeared in [12], and is important to our "bootstrapping" argument in §7.

LEMMA 3.8. (Journé's Lemma) *Let $\mathcal{U}$ be a collection of rectangles whose shadow has finite measure. For any $\eta > 0$, we can construct $V \supset \mathrm{sh}(\mathcal{U})$ and a function* Emb: $\mathcal{U} \longrightarrow [1, \infty)$ *so that*

$$\mathrm{Emb}(R) \cdot R \subset V, \qquad R \in \mathcal{U}, \tag{3.9}$$

$$|V| < (1+\eta)\,|\mathrm{sh}(\mathcal{U})|\,, \tag{3.10}$$

$$\left\| \sum_{R \in \mathcal{U}} \sum_{\vec{\varepsilon} \in \mathrm{Sig}_{\vec{d}}} \mathrm{Emb}(R)^{-C} \langle f, w_R^{\vec{\varepsilon}} \rangle w_R^{\vec{\varepsilon}} \right\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} \leq K_\eta \, \|f\|_{\mathrm{BMO}_{-1}(\mathbb{R}^{\vec{d}})} . \tag{3.11}$$

The constant K_η depending only on η and $\vec{d}$, and the constant C, appearing in the last display, upon the vector $\vec{d}$.

Notice that the power on the embeddedness term in (3.11) is allowed to be quite big, a function of the parameters $\vec{d}$ that we do not specify. Also, concerning the conclusions, if we were to take $\mathrm{Emb}(R) \equiv 1$, then certainly the first conclusion (3.9) would be true. But, the last conclusion would be false for the Carleson examples in particular. This choice is obviously not permitted in general.

The formulations of Journé's Lemma given here are not the typical ones found in Journé's original Lemma, or J. Pipher's extension to the three dimensional case [28]. These papers give the more geometric formulation of these Lemmas, and J. Pipher's article implicitly contains the geometric formulation needed to prove the Lemma above (provided one is satisfied with the estimate $|V| \lesssim |\mathrm{sh}(\mathcal{U})|$). See Pipher [28]. Lemma 3.8, as formulated above, was found in Lacey and Terwilleger [19]; the two dimensional variant (which is much easier) appeared in Lacey and Ferguson [12]. The paper of Cabrelli, Lacey, Molter and Pipher [3] surveys some issues related to Journé's Lemma. See in particular Sections 2 and 4. We refer the reader to these references for more information on this subject.

4. Paraproducts. The paraproducts that arise are of a somewhat general nature, and so we make some definitions which will permit a reasonably general definition of a paraproduct.

Let $\chi(x) = (1 + |x|^2)^{-1}$. Let $\chi_Q^{(2)} = \mathrm{Dil}_Q^{(2)} \chi$. Say that φ is *adapted to* Q iff

$$|\mathrm{D}^m \varphi(x)| \lesssim |Q|^{-m/d} \, [\chi_Q^{(2)}(x)]^N, \qquad x \in \mathbb{R}^d. \tag{4.1}$$

This inequality should hold for all derivatives D^m, where $m \le d+1$, where d is the ambient dimension. The inequality should hold for all integers N. The implied constant can depend upon these parameters. Say that φ has a zero iff $\int \varphi \, dx = 0$.

We extend these definitions to functions φ on $\mathbb{R}^{\vec{d}}$. Say that φ is *adapted to* $R = \prod Q_s$ if and only if

$$\varphi(x_1, \dots, x_t) = \prod_{s=1}^{t} \varphi_s(x_s), \qquad \text{where } \varphi_s \text{ is adapted to } Q_s. \tag{4.2}$$

Say that φ *has zeros in the sth coordinate* if and only if

$$\int_{\mathbb{R}^{d_s}} \varphi(x_1, \dots, x_s, \dots x_t) \, dx_s = 0, \qquad \text{for all } x_1, \dots x_{s-1}, x_{s+1} \dots x_t. \tag{4.3}$$

The main Theorem on paraproducts that we will need concerns bilinear operators formed in this way. For $j = 1, 2, 3$ let $\{\varphi_{j,R} \colon R \in \mathcal{D}_{\vec{d}}\}$ be three families of functions adapted to the dyadic rectangles in $\mathcal{D}_{\vec{d}}$. Then define

$$\mathrm{B}(f_1, f_2) \stackrel{\mathrm{def}}{=} \sum_{R \in \mathcal{D}_{\vec{d}}} \frac{\langle f_1, \varphi_{1,R} \rangle}{|R|^{1/2}} \langle f_2, \varphi_{2,R} \rangle \, \varphi_{3,R}.$$

The following result is due to Journé [13, 15]. Also see [17, 24, 25].

THEOREM 4.4. *Assume that the family* $\{\varphi_{1,R}\}$ *has zeros in all coordinates. For every other coordinate* s*, assume that there is a choice of* $j = 2, 3$ *for which the*

the family $\{\varphi_{j,R}\}$ *has zeros in the sth coordinate. Then the operator* B *enjoys the property*

$$\mathrm{B}\colon \mathrm{BMO} \times L^p \longrightarrow L^p, \qquad 1 < p < \infty.$$

We will refer to the function φ_1 as the *symbol* of the paraproduct. This function plays the same role for paraproducts as does the symbol of the commutator. Particularly relevant for us is the following reformulation of this theorem: If B_1 and B_2 are bounded paraproducts, then the tensor product $\mathrm{B}_1 \otimes \mathrm{B}_2$ is a bounded paraproduct for symbols on the corresponding product BMO space.

In many applications of this result, the functions $\varphi_{1,R}$, acting on the symbol of the paraproduct, will be product wavelets.

A more particular form of the upper bound on commutators plays a role in both the upper and lower bounds for our Main Theorem. We state this variant of Theorem 4.4 for our use below. In particular, the estimate (4.11) is used in the lower bound. It holds when the symbol and the function the paraproduct is applied to have "separated wavelet support" in the sense of (4.10).

For a subset of coordinates $J \subset \{1, \dots, t\}$ set

$$\mathrm{F}_{\vec{l},J} \stackrel{\text{def}}{=} \sum_{\vec{\varepsilon} \in \mathrm{Sig}_{\vec{d}}} \sum_{\substack{\vec{k} \in \mathbb{Z}^t \\ k_s = l_s,\ s \in J \\ k_s \geq l_s, s \notin J}} \Delta \mathrm{F}_{\vec{k}}, \tag{4.5}$$

$$\Delta \mathrm{F}_{\vec{k}} \stackrel{\text{def}}{=} \sum_{\vec{\varepsilon} \in \mathrm{Sig}_{\vec{d}}} \sum_{\substack{R \in \mathcal{D}_{\vec{d}} \\ |Q_s| = 2^{k_s}}} w_R^{\vec{\varepsilon}} \otimes w_R^{\vec{\varepsilon}}. \tag{4.6}$$

For those coordinates $s \in J$, we take the wavelet projection onto that scale, while for those coordinates $s \notin J$, we sum larger scales. That means that we lose the zero in the coordinates not in J.

Write $R' \lesssim_J R$ if and only if $|Q'_s| \leq |Q_s|$ for $s \notin J$ and $|Q'_s| = |Q_s|$ for $s \in J$.

THEOREM 4.7. *Let* T *be a product Calderón–Zygmund kernel as in Theorem 5.2. For all* $J \subset \{1, \dots, t\}$, *and* $\vec{k} \in \mathbb{Z}^t$ *with*

$$3 \leq k_s \leq 8,\ s \notin J, \qquad -8 \leq k_s \leq 8,\ s \in J. \tag{4.8}$$

We have

$$\left\| \sum_{\vec{l} \in \mathbb{Z}^t} (\Delta \mathrm{F}_{\vec{l}}\, b) \cdot \mathrm{T} \mathrm{F}_{\vec{l}+\vec{k},J} \varphi \right\|_2 \lesssim \|b\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} \|\varphi\|_2 . \tag{4.9}$$

Moreover, suppose we have the following separation condition: Fix an integer

$A > 0$. Suppose that

$$\text{(4.10)}\quad \text{if for } \vec{\varepsilon} \text{ and } \vec{\varepsilon}', \ \langle b, w_{R'}^{\vec{\varepsilon}'}\rangle \neq 0, \ \langle \varphi, w_R^{\vec{\varepsilon}}\rangle \neq 0 \text{ with } R' \lesssim_J R, \text{ then } AR \cap R' = \emptyset.$$

We then have the estimate

$$\text{(4.11)}\qquad \left\| \sum_{\vec{l}\in\mathbb{Z}^t} (\Delta \mathrm{F}_{\vec{l}} b) \cdot \mathrm{TF}_{\vec{l}+\vec{k},J} \varphi \right\|_2 \lesssim A^{-100t} \|b\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} \|\varphi\|_2 .$$

Implied constants are independent of the choice of $\vec{k}$.

The operator in (4.9), though it fits into the category of paraproducts, it does not fit the precise definition we have given of a paraproduct above, and so we will postpone the proof of this Theorem until the end of this section.

Shift operators. There are different types of "shifts" on wavelets that also enter into our considerations. These are shifts of signature, scale and location. We discuss each of these.

Define a "signature shift" operator by a map ϵ: $\mathrm{Sig}_{\vec{d}} \times \mathcal{D}_{\vec{d}} \longrightarrow \mathrm{Sig}_{\vec{d}}$, where $\epsilon(\cdot, R)$ is one to one for each rectangle R. Then the operator is defined first on wavelets by

$$\text{Ш}_{\mathrm{Sig},\epsilon}(w_R^{\vec{\varepsilon}}) = w_R^{\epsilon(\vec{\varepsilon},R)}$$

and then extended linearly. The boundedness properties of these operators are straightforward.

PROPOSITION 4.12. *We have the estimate*

$$\left\|\text{Ш}_{\mathrm{Sig},\epsilon}\right\|_{p\to p} \lesssim C_p, \qquad 1 < p < \infty.$$

The proof follows immediately from the Littlewood–Paley inequalities. We omit the details.

Define a "scale shift" operator by a one to one map σ_ρ: $\mathcal{D}_{\vec{d}} \longrightarrow \mathcal{D}_{\vec{d}}$ that sends each dyadic rectangle R into a unique $\sigma(R) \subset R$, so that the ratio $\rho = |\sigma(R)| \,/\, |R|$ is independent of R. The parameter of this operator is ρ. Define a corresponding linear operator $\text{Ш}_{\mathrm{scale},\rho}$ by

$$\text{Ш}_{\mathrm{scale},\rho}(w_R^{\vec{\varepsilon}}) \stackrel{\text{def}}{=} \sqrt{\rho} \cdot w_{\sigma(R)}^{\vec{\varepsilon}}$$

and the operator is then uniquely defined by linearity. Our observation is that this shift is a uniformly bounded operator on product BMO.

THEOREM 4.13. *The operators* $\text{Ш}_{\text{scale},\rho}$ *map* $\text{BMO}(\mathbb{R}^{\vec{d}})$ *to itself. Moreover for all* $\kappa > 0$ *we have the estimate*

$$\|\text{Ш}_{\text{scale},\rho}\|_{\text{BMO}\to\text{BMO}} \lesssim \rho^{-\kappa}.$$

Proof. Given $f \in \text{BMO}$, and open set $U \subset \mathbb{R}^{\vec{d}}$, consider the set

$$V \stackrel{\text{def}}{=} \{\text{M}\mathbf{1}_U > c\rho\}$$

where M is the strong t parameter maximal function appropriate to this setting, namely

$$\text{M}f \stackrel{\text{def}}{=} \sup_{R\in\mathcal{D}_{\vec{d}}} \frac{\mathbf{1}_{2R}}{|2R|}\int_{2R} |f(y)|\ dy.$$

Observe that for appropriate c if $\sigma(R) \subset U$ then we have $R \subset V$.

We can estimate

$$\begin{aligned}\sum_{\vec{\varepsilon}\in\text{Sig}_{\vec{d}}}\sum_{R\subset U}\left|\langle \text{Ш}_{\text{scale},\rho}f, w_R^{\vec{\varepsilon}}\rangle\right|^2 &= \rho\sum_{\vec{\varepsilon}\in\text{Sig}_{\vec{d}}}\sum_{R\subset U}\left|\langle f, w_{\sigma^{-1}(R)}^{\vec{\varepsilon}}\rangle\right|^2\\ &\le \rho\sum_{\vec{\varepsilon}\in\text{Sig}_{\vec{d}}}\sum_{R\subset V}\left|\langle f, w_R^{\vec{\varepsilon}}\rangle\right|^2\\ &\le \rho\,\|b\|_{\text{BMO}}^2\,|V|\,.\end{aligned}$$

It remains to estimate $|V|$ in terms of $|U|$.

Using the L^p mapping properties of the maximal function, we can estimate

$$|V| \lesssim \rho^{-p}\,|U|\,.$$

Taking $p = 1 + \kappa$ proves our theorem. □

Remark. When the number of parameters $t = 1$, the operators $\text{Ш}_{\text{scale},\rho}$ are in fact uniformly bounded on BMO as follows from the weak L^1 bound for the maximal function. For $t > 1$, there is a logarithmic estimate:

$$\|\text{Ш}_{\text{scale},\rho}\|_{\text{BMO}\to\text{BMO}} \lesssim (1 + \log 1/\rho)^t.$$

The strong maximal function we are using satisfies $\|\text{M}\|_{p\to p} \lesssim (p-1)^{-t}$, aside from dimensional considerations from the individual components of $\vec{d}$. Using this estimate, and taking $p - 1 \simeq |\log\rho|^{-1}$, the estimate above follows.

We define "location shift" operators. Let $\lambda_n\colon \mathcal{D}_{\vec{d}} \longrightarrow \mathcal{D}_{\vec{d}}$ be a one to one map such that for all rectangles $R \in \mathcal{D}_{\vec{d}}$, the image rectangle $\lambda_n(R)$ has the same dimensions in each coordinate, namely

$$|Q_s| = |\lambda_n(Q)_s|\,, \qquad 1 \le s \le t.$$

Moreover, $\lambda(R) \subset nR$. The shift operator is then defined on wavelets by

$$\text{Ш}_{\text{loc},n} w_R^{\vec{\varepsilon}} = w_{\lambda_n(R)}^{\vec{\varepsilon}}, \tag{4.14}$$

and is then extended linearly. The parameter of this operator is said to be n.

The estimate we need concerns the L^p norms of this operator.

PROPOSITION 4.15. *We have the estimates below, valid for all integers n.*

$$\|\text{Ш}_{\text{loc},n}\|_{p\to p} \lesssim n^{|\vec{d}|}$$

where $\left|\vec{d}\right| = d_1 + d_2 + \cdots + d_t$ *depends only on* $\vec{d}$.

Proof. Since λ_n is one to one, it is clear that Ш_{loc} is bounded with norm one on L^2. For $p \neq 2$ we use the Littlewood–Paley inequalities, together with the obvious fact that

$$\mathbf{1}_{\lambda_n(R)} \lesssim n^{|\vec{d}|} \mathrm{M}\mathbf{1}_R$$

for all rectangles R. Then, using the Fefferman–Stein Maximal inequality, we have

$$\begin{aligned}
\|\text{Ш}_{\text{loc},n} f\|_p &\lesssim \left\| \left[\sum_{\vec{\varepsilon}\in \mathrm{Sig}_{\vec{d}}} \sum_{R\in\mathcal{D}^{\vec{d}}} \frac{\left|\langle f, w_R^{\vec{\varepsilon}}\rangle\right|^2}{|R|} \mathbf{1}_{\lambda_n(R)} \right]^{1/2} \right\|_p \\
&\lesssim n^{|\vec{d}|} \left\| \left[\sum_{\vec{\varepsilon}\in \mathrm{Sig}_{\vec{d}}} \sum_{R\in\mathcal{D}^{\vec{d}}} \frac{\left|\langle f, w_R^{\vec{\varepsilon}}\rangle\right|^2}{|R|} (\mathrm{M}\mathbf{1}_R)^2 \right]^{1/2} \right\|_p \\
&\lesssim n^{|\vec{d}|} \left\| \left[\sum_{\vec{\varepsilon}\in \mathrm{Sig}_{\vec{d}}} \sum_{R\in\mathcal{D}^{\vec{d}}} \frac{\left|\langle f, w_R^{\vec{\varepsilon}}\rangle\right|^2}{|R|} \mathbf{1}_R \right]^{1/2} \right\|_p \\
&\lesssim n^{|\vec{d}|} \|f\|_p\,.
\end{aligned}$$

□

Generalized paraproducts. Experience shows that paraproducts arise in a variety of ways. They do in this paper, and in this section, we adopt a notation to formalize the different ways that the paraproducts arise.

Given an operator P acting on $L^p(\mathbb{R}^{\vec{d}})$, we set

$$\|\mathrm{P}\|_{\mathrm{Para}} = \inf \left\{ \sum_{\epsilon} \sum_{\rho} \sum_{n} n^{|\vec{d}|} \rho^{-1/|\vec{d}|} \, |c(\epsilon, \rho, n)| \right\} \tag{4.16}$$

where the infimum is taken over all representations

$$\mathrm{P}f = \sum_{\epsilon} \sum_{\rho} \sum_{n} c(\epsilon, \rho, n) \cdot \mathrm{B}_{\epsilon,\rho,n}(\text{Ш}_{\mathrm{scale},\rho} b, \text{Ш}_{\mathrm{Sig},\epsilon} \text{Ш}_{\mathrm{loc},n} f).$$

In this display, the operators $\mathrm{B}_{\epsilon,\rho,n}$ are paraproducts as in Theorem 4.4, with norm at most one. The operators $\text{Ш}_{\mathrm{scale},\rho}$ are scale shift operators, with parameter ρ; the $\text{Ш}_{\mathrm{Sig},\epsilon}$ are signature shift operators; and $\text{Ш}_{\mathrm{loc},n}$ are location shift operators of parameter n.

We may combine the different results of this section into the estimate

$$\|\mathrm{P}\|_{p\to p} \lesssim \|\mathrm{P}\|_{\mathrm{Para}}, \qquad 1 < p < \infty. \tag{4.17}$$

Examples of how to use this norm are in the next proof.

Proof of Theorem 4.7. We will assume that the Calderón–Zygmund operator T is the identity. It is straightforward to supply the necessary additional details to accommodate the general case.

The "father wavelet" W permits us to rewrite the operator in (4.5). For a subset of coordinates $J \subset \{1, \dots, t\}$ we set

$$W_{R,J}(x_1, \dots, x_t) \stackrel{\mathrm{def}}{=} \prod_{s\in J} w_{Q_s}^{\varepsilon_s}(x_s) \cdot \prod_{s\notin J} W_{Q_s}(x_s).$$

Thus, in the coordinates in J we take the Meyer wavelet, and for those coordinates not in J we take a father wavelet. In particular, $\mathrm{F}_{\vec{l},J}$ as defined in (4.5) is

$$\mathrm{F}_{\vec{l},J} = \sum_{\vec{\varepsilon}\in \mathrm{Sig}_{\vec{d}}} \sum_{\substack{R\in\mathcal{D}^{\vec{d}} \\ |R_s|=2^{l_s}}} W_{R,J} \otimes W_{R,J}.$$

Let $\vec{k} \in \mathbb{Z}^{\vec{t}}$ be as in (4.8). Let R, R' be dyadic rectangles with

$$|Q_s| = 2^{k_s d_s} \left|Q_s'\right|, \quad 1 \le s \le t, \qquad A \simeq \mathrm{M}\mathbf{1}_R(c(R')). \tag{4.18}$$

The function

$$\zeta_{R,R',J} \stackrel{\text{def}}{=} A^{-N}\sqrt{|R|}w_R \cdot W_{R',J}$$

is adapted to R, in the sense of (4.2). Here N is a fixed large constant depending upon $\vec{d}$.

The assumption (4.8) plays an essential role in describing the zeros of the function $\zeta_{R,R',J}$. $W_{R',J}$ has zeros for $s \in J$, but certainly does not have zeros for $s \notin J$. The properties of the Meyer wavelet, and in particular (2.4), along with the assumption on $\vec{k}$ then imply that $\zeta_{R,R',J}$ has zeros for $s \notin J$.

Now, consider a map $\pi_A\colon \mathcal{D}_{\vec{d}} \longrightarrow \mathcal{D}_{\vec{d}}$ such that the pairs $R, \pi(R)$ satisfy (4.18). Set $\mu(\pi) = A$ where A is as in (4.18). The operator

$$\mathrm{B}_\pi(b,\varphi) \stackrel{\text{def}}{=} \sum_{R\in\mathcal{D}_{\vec{d}}} \frac{\langle b, w_R\rangle}{\sqrt{|R|}} \langle \varphi, W_{\pi(R),J}\rangle \, \zeta_{R,\pi(R),J}$$

is a paraproduct, composed with a change of location operator. Note that the function that falls on b has zeros in all coordinates; the function that falls on φ has zeros for $s \in J$, and $\zeta_{R,\pi(R)}$ has zeros for $s \notin J$. It is then clear that

$$\|\mathrm{B}_\pi\|_{\mathrm{Para}} \lesssim \mu(\pi)^N. \tag{4.19}$$

Now, a moments thought reveals that we can write, for appropriate choices of π_υ,

$$\sum_{\vec{l}\in\mathbb{Z}^t} (\Delta \mathrm{F}_{\vec{l}} b) \cdot \mathrm{F}_{\vec{l}+\vec{k},J}\varphi = \sum_{\upsilon=1}^{\infty} \mathrm{B}_{\pi_\upsilon}(b,\varphi). \tag{4.20}$$

Moreover, for all $0 < A < 1$, the number of π_υ occurring in the sum above with $\mu(\pi_\upsilon) \simeq A$ is at most A^{-C} where C depends upon $\vec{d}$. But then from (4.19), it is clear that (4.9) holds.

The second conclusion of the Lemma, (4.11), is quite important to the proof of our lower bounds on commutator norms. (Estimates of this type are also important to detailed information about norm bounds for paraproducts. See [17, §4.3].) But with the assumption (4.10), note that we can again have the equality (4.20), but with this additional property: For all υ, we have $\mu(\pi_\upsilon) \lesssim A$. It is then clear that (4.11) holds. □

5. The upper bound. Let K be a standard Calderón–Zygmund convolution kernel on $\mathbb{R}^d \times \mathbb{R}^d$. This means that the kernel is a distribution that satisfies the estimates below for $x \neq y$

$$\begin{aligned} \left|\nabla^j K(y)\right| &\leq N\,|y|^{-d-j}, \quad j = 0, 1, 2, \ldots, d+1. \\ \left\|\widehat{K}\right\|_{L^\infty(\mathbb{R}^{\vec{d}})} &\leq N. \end{aligned} \tag{5.1}$$

The first estimate combines the standard size and smoothness estimate. (Our proof requires a large number of derivatives on the kernel, due to an argument in §5.1.) The last, and critical, assumption is equivalent to assuming that the operator defined on Schwartz functions by

$$\mathrm{T}_K f(x) \stackrel{\text{def}}{=} \int K(x-y) f(y)\, dy$$

extends to a bounded operator on $L^2(\mathbb{R}^d)$. The least constant N satisfying the inequalities (5.1) and $\|\mathrm{T}_K\|_{2\to 2} \le N$ is some times referred to as the *Calderón–Zygmund norm of K*.

Now let $K_1, \dots, K_t$ be a collection of Calderón–Zygmund kernels, with K_s defined on $\mathbb{R}^{d_s} \times \mathbb{R}^{d_s}$. It is not obvious that the corresponding tensor product operator

$$T_{K_1} \otimes \cdots \otimes T_{K_t}$$

is a bounded operator on $L^p(\mathbb{R}^{\vec{d}})$. This is a consequence of the multiparameter Calderón–Zygmund theory. This is a basic fact for us, so we state it here.

PRODUCT CALDERÓN–ZYGMUND KERNELS 5.2. *Let $K_1, \dots, K_t$ be a collection of Calderón–Zygmund convolution kernels, with K_s defined on $\mathbb{R}^{d_s} \times \mathbb{R}^{d_s}$. Then*

$$\mathrm{T}_{K_1} \otimes \cdots \otimes \mathrm{T}_{K_t}$$

extends to a bounded linear operator from $L^p(\mathbb{R}^{\vec{d}})$ to itself for all $1 < p < \infty$.

It is also not at all clear that the multiparameter commutators are bounded operators, even in the case of the Riesz transforms. Thus, this is one of the principal results of this paper.

THEOREM 5.3. *We have the estimates below, valid for $1 < p < \infty$.*

$$\|[\mathrm{T}_{K_1}, \cdots [\mathrm{T}_{K_t}, \mathrm{M}_b] \cdots]\|_{p\to p} \lesssim \|b\|_{\mathrm{BMO}}\,. \tag{5.4}$$

By BMO*, we mean Chang–Fefferman* BMO*. The implied constant depends upon the vector $\vec{d}$, and the Calderón–Zygmund norm of the T_{K_s}.*

There are two corollaries of this that we will use below. For a symbol b on $\mathbb{R}^{\vec{d}}$, define

$$\|b\|_{\mathrm{Riesz},p} \stackrel{\text{def}}{=} \sup \|[\mathrm{R}_{j_1}, \cdots [\mathrm{R}_{j_t}, \mathrm{M}_b] \cdots]\|_{p\to p}\,, \qquad 1 < p < \infty. \tag{5.5}$$

where the supremum is formed over all choices of Riesz transforms R_{j_s} for $1 \le j_s \le d_s$. (In the case that $d_s = 1$, the Riesz transforms reduce to the Hilbert

transform.) The Riesz transforms of course fall under the purview of the Theorem above, so we see that $\|b\|_{\text{Riesz},p} \lesssim \|b\|_{\text{BMO}}$. This is half of our Main Theorem, and the other half is the reverse inequality.

We will have need of another class of singular integral operators besides the Riesz transforms, with the Fourier transform of these kernels—the symbol of the kernel—being well adapted to a *cone* in $\mathbb{R}^d$.

Suppose that the dimension $d \geq 2$. A *cone* $C \subset \mathbb{R}^d$ is specified by the data (ξ_C, Q) where $\xi_C \in \mathbb{R}^d$ is a unit vector referred to as the *direction of the cone* and $Q \subset \mathbb{R}^{d-1}$ is a cube centered at the origin. The cone consists of all vectors θ given in orthogonal coordinates $(\theta_\xi \xi, \theta_\perp)$ with $\theta_\xi = \theta \cdot \xi$, and $\theta_\perp \in \theta_\xi Q$. For $0 < \lambda$ by λC we mean the cone with data $(\xi_C, \lambda Q)$. By the *aperture of* C we mean $|Q|$.

The Fourier restriction operator specified by C should be bounded on all L^p spaces, Namely the operator defined by

$$\widehat{\mathrm{P}_C f} \stackrel{\text{def}}{=} \mathbf{1}_C \widehat{f} \tag{5.6}$$

should admit a uniform bound on all $L^p(\mathbb{R}^d)$ spaces. By taking the boundary of the cone to be a cube this is certainly the case: Compositions of Fourier projections onto half spaces yields P_C, so it will not be given by composition with respect to a (one parameter) Calderón–Zygmund kernel as in (5.1).

For a cone in $C \subset \mathbb{R}^d$, we fix a Calderón–Zygmund kernel K_C which satisfies the size and smoothness assumptions above, and in addition,

$$\mathbf{1}_C \leq \widehat{K_C} \leq \mathbf{1}_{(1+\kappa)C}. \tag{5.7}$$

Here, we introduce a small parameter κ which will depend upon dimension $\vec{d}$. Moreover, we choose the cone operator to make a sufficiently smooth transition from 0 to 1 that the operator T_C with symbol given by K_C defines a Calderón–Zygmund operator, bounded on all L^p, $1 < p < \infty$.

There is however an essential point to observe: That the kernel K_C satisfies the Calderón–Zygmund estimates (5.1), but with constants that tend to infinity as the aperture of the cone tends to infinity. In the limit, the kernels K_C tend to a projection of a one dimensional Calderón–Zygmund kernel. (The operators admit uniform L^p bound in the aperture, but we need to apply a Theorem of Song-Ying Li [21] which only applies if the kernels are Calderón–Zygmund on $\mathbb{R}^d$.)

But, with the aperture fixed, in each dimension, we can choose these kernels to be rotations of one another, so that they admit uniform bounds in $L^p(\mathbb{R}^d)$. We will refer to the operator T_C given by convolution with K_C as a *Cone transform.*

As a matter of convention, in the case of $d = 1$, there are two cones, $\mathbb{R}_\pm$. The Cone transforms are the corresponding projections onto the positive and negative frequency axes. These are of course linear combinations of the identity and the Hilbert transform, which coincide with the Riesz transforms.

We now define a third norm on a symbol b on $\mathbb{R}^{\vec{d}}$

$$\|b\|_{\mathrm{Cone},p} \stackrel{\mathrm{def}}{=} \sup \|[\mathrm{T}_{C_1}, \cdots [\mathrm{T}_{C_t}, \mathrm{M}_b] \cdots]\|_{p\to p}\,, \qquad 1 < p < \infty. \tag{5.8}$$

where the supremum is formed over all choices of Cone transforms T_{C_s} with $C_s \subset \mathbb{R}^{d_s}$ in which the aperture of the cone is fixed. Later in the proof, we will specify an aperture. It follows that we also have $\|b\|_{\mathrm{Cone},p} \lesssim \|b\|_{\mathrm{BMO}}$. This is an important observation for us, so let us formalize it in the following Corollary of Theorem 5.3, which includes half of our Main Theorem.

COROLLARY 5.9. *We have the inequalities*

$$\|b\|_{\mathrm{Riesz},p}\,,\ \|b\|_{\mathrm{Cone},p} \lesssim \|b\|_{\mathrm{BMO}}\,, \qquad 1 < p < \infty.$$

For the inequality concerning Cone operators, the implied constant depends upon the aperture of the cones.

Remark. In the one dimensional case, a "cone" is just a projection onto the positive axis, say, and most of the considerations of this section are not needed. For the sake of exposition, in this section we will assume that all the coordinates of $\vec{d} = (d_1, \ldots, d_t)$ are at least two. The case when some coordinates are one is technically easier, but more difficult in terms of accommodating the general argument into the notation.

Let us formalize the extended version of our Main Theorem.

EXTENSION OF MAIN THEOREM 5.10. *For all $t \geq 1$ and choices of $\vec{d}$ we have*

$$\|b\|_{\mathrm{Riesz},p} \simeq \|b\|_{\mathrm{Cone},p} \simeq \|b\|_{\mathrm{BMO}}\,, \qquad 1 < p < \infty.$$

The implied constants depend upon the vector $\vec{d}$ and the aperture of the cone.

We find it necessary to prove the equivalence between the BMO and Cone norms in order to deduce the equivalence with the Riesz norm.

5.1. A one parameter result. A commutator is a special form of a paraproduct. Our approach to Theorem 5.3 is obtain a decomposition of a one parameter commutator into a sum of paraproducts. The tensor product of the elements of our decomposition are themselves bounded operators, so we can then pass to the multiparameter statement of the Theorem.

Remark. The multiparameter setting is related to the tensor products of dilation groups. An essential difficulty is that the tensor product of bounded operators need not be bounded. See [16]. And so it will be incumbent upon us to describe

sufficient conditions on the operators for the tensor products to be bounded, and reduce the commutators above to these settings.

A result of this type, expressing a commutator as a sum of paraproducts, is known to experts, and has been used in [1], and may well have been formulated in this way before.

PROPOSITION 5.11. *For any Calderón–Zygmund kernel satisfying (5.1), and symbol b we can write the commutator* $[T_K, M_b]$ *as an absolutely convergent sum of paraproducts composed with signature, scale and location shifts. That is, using the notation in (4.16),*

$$\|[T_K, M_b]\|_{\mathrm{Para}} \lesssim 1.$$

Proof of Theorem 5.3. The Proposition above shows that a commutator is the absolutely convergent sum of bounded paraproducts. The result of Journé, Theorem 4.4, is that the tensor product of bounded paraproducts is bounded. As the commutators in our Theorem act on a tensor product space, we see that the commutators can be written as an absolutely convergent sum of tensor products of bounded paraproducts. Hence, the Theorem follows. □

Proof. A basic fact here is that if ϕ is adapted to a cube Q, then so is $T_K\phi$. Clearly, $T_K\phi$ has a zero. This in particular shows that for a paraproduct operator B, we have

$$\|T_K \circ B\|_{\mathrm{Para}} + \|B \circ T_K\|_{\mathrm{Para}} \lesssim \|B\|_{\mathrm{Para}}.$$

As we are working with convolution operators, we could use a classical Littlewood–Paley decomposition method to prove this result. We have however already introduced wavelets (which are essential later in this paper) so we prefer that method here.

We recall that the Meyer wavelet w in one dimension has Fourier transform identically equal to zero on a neighborhood of the origin. It follows from the rapid decrease of the wavelet that we then have

$$\int_{\mathbb{R}} x^k w(x)\, dx = 0, \qquad k > 0. \tag{5.12}$$

That is, the wavelet is orthogonal to all polynomials in x. This property extends to the multidimensional Meyer wavelet.

Let

$$F_j \stackrel{\mathrm{def}}{=} \sum_{\varepsilon \in \mathrm{Sig}_d} \sum_{|Q| \geq 2^{jd}} w_Q^\varepsilon \otimes w_Q^\varepsilon \tag{5.13}$$

be the Father wavelet projection. And let

$$\Delta \mathrm{F}_j \stackrel{\text{def}}{=} \mathrm{F}_j - \mathrm{F}_{j+1} = \sum_{\varepsilon \in \mathrm{Sig}_d} \sum_{|Q|=2^{jd}} w_Q^{\varepsilon} \otimes w_Q^{\varepsilon}$$

be the projection onto the wavelets of scale 2^j.

The property (2.4) is relevant to us. In particular, it follows from this that we have the Fourier transform of the product $\Delta_j \mathrm{F}_j b \cdot \mathrm{F}_{j+3} f$ is localized to $2^{-j-3} \leq |\xi| \leq 2^{-j+3}$.

We expand the commutator in these wavelet projections. Thus,

$$[\mathrm{T}_K, \mathrm{M}_b] f = \sum_{j,j'} [\mathrm{T}_K, \mathrm{M}_{\Delta \mathrm{F}_j b}] \Delta \mathrm{F}_{j'} f.$$

The principal term arises from $j+3 < j'$, where we do not have any cancellation in the commutator, and we write

$$\begin{aligned} \sum_{j+3<j'} [\mathrm{T}_K, \mathrm{M}_{\Delta \mathrm{F}_j b}] \Delta \mathrm{F}_{j'} f &= \mathrm{T}_K \circ \mathrm{B}_1(b,f) - \mathrm{B}_2(b,f), \\ \mathrm{B}_1(b,\phi) &\stackrel{\text{def}}{=} \sum_j \Delta \mathrm{F}_j \, b \cdot \mathrm{F}_{j+3} \phi, \\ \mathrm{B}_2(b,\phi) &\stackrel{\text{def}}{=} \sum_j \Delta \mathrm{F}_j \, b \cdot \mathrm{T}_K \circ \mathrm{F}_{j+3} \phi. \end{aligned}$$

It is important that the product $\Delta \mathrm{F}_j \, b \cdot \mathrm{F}_{j+3} f$ have no Fourier support in a neighborhood of the origin that has diameter proportional to 2^{-j}. Certainly, B_1 is a paraproduct. It follows that $\mathrm{T}_K \circ \mathrm{B}_1$ is as well. Upon inspection, one sees that B_2 is a paraproduct. It is also straightforward to verify that $\|\mathrm{T}_K \circ \mathrm{B}_1\|_{\mathrm{Para}} + \|\mathrm{B}_2\|_{\mathrm{Para}} \lesssim 1$.

In the remaining cases we expect terms which are substantially smaller. The principal point is this estimate. For $\varepsilon, \varepsilon' \in \mathrm{Sig}_d$,

$$\begin{aligned} \left| [\mathrm{T}_K, \mathrm{M}_{w_Q^{\varepsilon}}] w_{Q'}^{\varepsilon'}(x) \right| \lesssim & \left[\frac{|Q|}{|Q'|} \right]^{1+1/2d} \left(1 + \frac{\operatorname{dist}(Q, Q')}{|Q|^{1/d}} \right)^{-N} \\ & \times |Q|^{-1/2} \, [\chi_{Q'}^{(2)}(x)]^N, \qquad 2^{3d} |Q| \geq |Q'| \,. \end{aligned} \tag{5.14}$$

Here, $\chi_{Q'}^{(2)}$ is as in (4.1). In the language of the section on paraproducts, this shows that a large constant times this function is adapted to the cube Q'. The power $-N$ on the term involving distance holds for all large N; a power of $N > d$ is required; The power $1 + 1/2d$ on the ratio $|Q| / |Q'|$ follows from the number of derivatives we have on the kernel in (5.1); some power larger than one is required.

With the inequality (5.14), it is easy to verify that

$$\|[\mathrm{M}_b, \mathrm{T}_K] \cdot - \mathrm{T}_K \mathrm{B}_1(b, \cdot) + \mathrm{B}_2(b, \cdot)\|_{\mathrm{Para}} \lesssim 1.$$

The proof of (5.14) is taken in two steps. We have

$$\begin{aligned} \left|[\mathrm{T}_K, \mathrm{M}_{w_Q^\varepsilon}] w_{Q'}^{\varepsilon'}(x)\right| &\lesssim \left(1 + \frac{\operatorname{dist}(Q, Q')}{|Q|^{1/d}}\right)^{-N} \\ &\quad \times |Q|^{-1/2} [\chi_{Q'}^{(2)}(x)]^N, \qquad 2^{3d} |Q| \geq |Q'|. \end{aligned}$$

That is, we do not have the term involving $|Q| / |Q'|$ appearing on the right. This estimate is easy to obtain, and we omit the details.

The second estimate is

$$\left|[\mathrm{T}_K, \mathrm{M}_{w_Q^\varepsilon}] w_{Q'}^{\varepsilon'}(x)\right| \lesssim \left[\frac{|Q|}{|Q'|}\right]^{1+1/d} |Q|^{-1/2} [\chi_{Q'}^{(2)}(x)]^N, \quad 2^{3d} |Q| \geq |Q'|. \tag{5.15}$$

That is, we have a slightly larger power of $|Q| / |Q'|$ than is claimed in (5.14). Taking a geometric mean of these two estimates will prove (5.14).

To see (5.15), let us assume that $|Q| = 1$, which we can do as a dilation of K has the same Calderón–Zygmund norm as K. Observe that the commutator above is

$$\int \{w_Q^\varepsilon(x) - w_Q^\varepsilon(y)\} K(x - y) \cdot w_{Q'}^{\varepsilon'}(y)\, dy.$$

Write the leading term in the integral as

$$\{w_Q^\varepsilon(x) - w_Q^\varepsilon(y)\} K(x - y) = T(x - y) + E(x, y),$$

where $T(x - y)$ is the dth degree Taylor polynomial of the left-hand side, and the error term $E(x, y)$ satisfies $|E(x, y)| \lesssim |x - y|^{d+1}$. That we have this estimate follows from our assumptions (5.1) on the kernel K. The wavelet $w_{Q'}^{\varepsilon'}$, by choice of wavelet, is orthogonal to the Taylor polynomial, see (5.12). Thus, as claimed,

$$\begin{aligned} \left|\int \{w_Q^\varepsilon(x) - w_Q^\varepsilon(y)\} K(x - y) \cdot w_{Q'}^{\varepsilon'}(y)\, dy\right| &\leq \left|\int E(x, y) w_{Q'}^{\varepsilon'}(y)\, dy\right| \\ &\lesssim \int |x - y|^{d+1} \left|w_{Q'}^{\varepsilon'}(y)\right| dy \\ &\lesssim \left[\frac{|Q|}{|Q'|}\right]^{1+1/d} \chi_{Q'}^{(2)}(x). \end{aligned}$$

□

An estimate for Riesz transforms. For our use at the end of the proof of the lower bound on operator norms on Riesz commutators, we will need a more quantitative estimate on upper bounds of such commutators. This estimate is most convenient to state here.

PROPOSITION 5.16. *For all integers $a \geq 1$, consider the operator*

$$\mathrm{U}_a(f,g) \stackrel{\text{def}}{=} \sum_{|Q'|=2^a|Q|} \sum_{\varepsilon',\varepsilon\in \mathrm{Sig}_d} \langle f, w_{Q'}^{\varepsilon'}\rangle \langle g, w_Q^{\varepsilon}\rangle \, [\mathrm{M}_{w_{Q'}^{\varepsilon'}}, \mathrm{R}_s] w_Q^{\varepsilon}$$

where R_s is the sth Riesz transform on $\mathbb{R}^d$. We have the estimate

$$\|\mathrm{U}_a\|_{\mathrm{Para}} \lesssim 2^{-Ma}. \tag{5.17}$$

This estimate holds for all $a, M > 1$, and all Riesz transforms, with implied constant only being a function of M, and the dimension d.

The proof is a simple variant on the previous proof. Clearly the role of the signatures is unimportant, and we will ignore the sum on the signatures in the argument below. Note that U_a is a paraproduct, with zeros falling on f, and zeros falling on g. Now, Q and Q' have different scales, which means that w_Q and $w_{Q'}$ are not adapted to cubes of the same scale.

This was exactly the problem addressed with the inequality (5.14) above. However, the Riesz transform has an infinitely smooth kernel. Therefore, a stronger form of (5.14) holds. Namely, for all $m > 1$, we have

$$\begin{aligned}\left|[[\mathrm{M}_{w_{Q'}^{\epsilon}}, \mathrm{R}_s] w_Q^{\varepsilon}(x)\right| \lesssim{}& \left[\frac{|Q|}{|Q'|}\right]^m \left(1 + \frac{\operatorname{dist}(Q,Q')}{|Q'|^{1/d}}\right)^{-N} \\ &\times |Q'|^{-1/2} [\chi_Q^{(2)}(x)]^N, \qquad |Q'| = 2^a |Q|\,.\end{aligned}$$

The implied constant depends only upon m, through the growth of the constants in the relevant estimates of the Riesz transform kernels.

The proof of the estimate proceeds just as the proof of (5.14), so we omit the details. The derivation of the proposition from this last estimate is routine.

6. The lower bound. We turn to the converse to Corollary 5.9, namely the Theorem below, which includes half of our Main Theorem.

THEOREM 6.1. *We have the inequalities below, valid for all choices of $\vec{d}$:*

$$\|b\|_{\mathrm{BMO}} \lesssim \|b\|_{\mathrm{Riesz},p}\,, \ \|b\|_{\mathrm{Cone},p}\,, \qquad 1 < p < \infty,$$

where the two norms are defined in (5.5) and (5.8). For the inequality involving the cone norm, the implied constant depends upon the aperture of the cone.

Remark. It is enough to prove this inequality with the L^2 operator norm on the right-hand side. If a commutator is bounded from L^p to itself, then it is also bounded on the conjugate space $L^{p'}$, and so by interpolation bounded on L^2. That is, we have the inequality $\|b\|_{\text{Riesz},2} \lesssim \|b\|_{\text{Riesz},p}$, valid for all $1 < p < \infty$. The same inequality holds for the Cone norm.

For the rest of this paper, we will denote $\|b\|_{\text{Riesz},2}$ by $\|b\|_{\text{Riesz}}$ and similarly for the Cone norms.

We use induction on parameters, namely the number of coordinates in $\vec{d}$. The base case is $t = 1$. Coifman, Rochberg and Weiss [8] proved that $\|b\|_{\text{BMO}} \lesssim \|b\|_{\text{Riesz}}$. This is a well-known result, with a concise proof. We find it necessary to prove the same inequality for the cones as an aid to proving the result about Riesz transforms. Indeed, it was this part of the proof that motivated the definition of the cone norm.

In the case $t = 1$, we indeed have the inequality $\|b\|_{\text{BMO}} \lesssim \|b\|_{\text{Cone}}$. This is a consequence of a deep line of investigation begun by Uchiyama [29], in which both directions of the Coifman, Rochberg, and Weiss result were extended to more general Calderón–Zygmund operators. In particular, a result of Song-Ying Li gives us as a Corollary to his Theorem, this essential result, which completes our discussion of the base case $t = 1$ in our induction on parameters.

THEOREM 6.2. (Li [21]) *In the case of $t = 1$, for all $d \geq 1$ and symbols b on $\mathbb{R}^d$ we have*

$$\|b\|_{\text{BMO}} \lesssim \|b\|_{\text{Cone}}\,.$$

In the inductive stage of the proof, we use the induction hypothesis to derive a lower bound on the commutator norms in terms of the BMO_{-1} norm. In so doing, it is very useful to use the equivalent Weak Factorization Theorem.

We then "bootstrap" from this weaker inequality to the full inequality. Namely, we can work with a symbol b with BMO norm one, but with BMO_{-1} norm small. (That is, the function b is of the type found in Carleson's examples [4].) With b fixed, we select an appropriate commutator which will admit a lower bound on its operator norm. We select a test function which will show that the commutator has a large operator norm. Verification of this fact will depend critically on the assumption that the symbol has small BMO_{-1} norm, and the Journé Lemma.

6.1. The initial BMO_{-1} lower bound. We assume that $t \geq 2$ and use the induction hypothesis to establish a lower bound on the Riesz and Cone norms of a symbol. This norm is in terms of our BMO norm with $t - 1$ parameters.

LEMMA 6.3. *For $t \geq 2$, assume Theorem 6.1 in the case of $t - 1$ parameters. Then we have the estimate*

$$\|b\|_{\text{BMO}_{-1}} \lesssim \|b\|_{\text{Cone}}\,,\ \|b\|_{\text{Riesz}}\,, \tag{6.4}$$

where it is to be emphasized that the BMO_{-1} *norm on the left is the* BMO *norm of* $t-1$ *parameters.*

Proof. We only give the proof of $\|b\|_{\mathrm{BMO}_{-1}} \lesssim \|b\|_{\mathrm{Riesz}}$ explicitly. This proof uses an equivalent form of the induction hypothesis, namely a weak factorization result on H^1 in $t-1$ parameters. The same weak factorization result holds for Cone transforms. See Li [21] for the one parameter formulation of this result.

Using the notation of (1.3), it is a straightforward exercise in duality to demonstrate that

$$\sup_{\vec{j}} \left\| \mathrm{C}_{\vec{j}}(b,\cdot) \right\|_{2\to 2} \approx \|b\|_{(L^2(\mathbb{R}^{\vec{d}})\widehat{\odot}L^2(\mathbb{R}^{\vec{d}}))^*}. \tag{6.5}$$

Therefore to show (6.4), it is sufficient to demonstrate that the following inequality holds,

$$\|b\|_{(L^2\widehat{\odot}L^2)^*} \gtrsim \|b\|_{\mathrm{BMO}_{-1}}, \tag{6.6}$$

and this will be established by relying upon the truth of the Theorem in $t-1$ parameters.

Given a smooth symbol $b(x_1,\dots x_t) = b(x_1,x')$ of t parameters, we assume that $\|b\|_{\mathrm{BMO}_{-1}} = 1$. As the symbol is smooth, the supremum in the norm is achieved by a collection of rectangles $\mathcal{U}$ of $\mathcal{D}_{\vec{d}}$ of $t-1$ parameters. We can assume that the rectangles in $\mathcal{U}$ agree in the first coordinate, to a cube $Q \subset \mathbb{R}^{d_1}$. As there are free dilations in each coordinate, we can assume that $|Q| = 1$ and $|\mathrm{sh}(\mathcal{U})| \approx 1$. Then define

$$\psi = \sum_{R\in\mathcal{U}} \sum_{\vec{\varepsilon}\in \mathrm{Sig}_{\vec{d}}} \langle b, w_R^{\vec{\varepsilon}}\rangle w_R^{\vec{\varepsilon}}.$$

One notes that $\langle b,\psi\rangle = 1$. To prove the claim, it is then enough to demonstrate that $\|\psi\|_{L^2(\mathbb{R}^{\vec{d}})\widehat{\odot}L^2(\mathbb{R}^{\vec{d}})} \lesssim 1$. Next observe that $\psi(x) = \psi_1(x_1)\psi'(x')$ and $\psi_1 \in H^1(\mathbb{R}^{d_1})$ with

$$\|\psi_1\|_{H^1(\mathbb{R}^{d_1})} = 1.$$

For the function ψ_1, we use the one parameter weak factorization of $H^1(\mathbb{R}^{d_1})$ of Coifman, Rochberg and Weiss [8]: There exists functions $f_n^j, g_n^j \in L^2(\mathbb{R}^{d_1})$, $n\in\mathbb{N}$, $1 \le j_1 \le d_1$, such that

$$\psi_1 = \sum_{n=1}^{\infty} \sum_{j_1=1}^{d_1} \Pi_{1,j_1}(f_n^{j_1}, g_n^{j_1}),$$

where $\Pi_{1,j_1}(p,q) := R_{1,j_1}(p)q + pR_{1,j_1}(q)$. One next sees that $\psi' \in H^1(\otimes_{l=2}^t \mathbb{R}^{d_l})$ with norm controlled by a constant. This follows from the choice of $\mathcal{U}$ and

the square function characterization of the space $H^1(\otimes_{l=1}^{t-1}\mathbb{R}^{d_l})$. By the induction hypothesis in $t-1$ parameters, in particular that $H^1(\otimes_{l=2}^{t}\mathbb{R}^{d_l}) = L^2(\otimes_{s=2}^{t}\mathbb{R}^{d_s})\widehat{\odot}L^2(\otimes_{s=2}^{t}\mathbb{R}^{d_s})$, we have $f_m^{\vec{j}}, g_m^{\vec{j}} \in L^2(\otimes_{s=2}^{t}\mathbb{R}^{n_s})$ with $m \in \mathbb{N}$ and $\vec{j}$ a vector with $1 \le j_s \le d_s$ for $s = 2, \dots, t$ such that

$$\psi' = \sum_{m=1}^{\infty}\sum_{\vec{j}} \Pi_{\vec{j}}(f_m^{\vec{j}}, g_m^{\vec{j}}), \qquad \sum_{m=1}^{\infty}\sum_{\vec{j}} \left\|f_m^{\vec{j}}\right\|_2 \left\|g_m^{\vec{j}}\right\|_2 \lesssim 1.$$

This immediately implies (6.4) since $\psi = \psi_1\psi'$, and we have a weak factorization of ψ with $\|\psi\|_{L^2(\mathbb{R}^{\vec{d}})\widehat{\odot}L^2(\mathbb{R}^{\vec{d}})} \lesssim 1$. □

7. The bootstrapping argument. In this section, we assume that $\|b\|_{\mathrm{BMO}_{-1}} < \delta_{-1}$, is very small, for a constant $0 < \delta_{-1} < 1$ to be chosen. Under this additional assumption, we conclude the proof of Theorem 6.1.

This proof is intricate, and indeed at this stage we find it essential to first prove the result for cones, namely we first prove $\|b\|_{\mathrm{BMO}} \lesssim \|b\|_{\mathrm{Cone}}$. Elements of this proof in this case are essential to address the Riesz norm case.

7.1. The lower bound on the cone norm. This case follows the lines of the argument of Lacey and Terwilleger [20]. (The current argument is however somewhat simpler.) We make a remark about the cone norms with different apertures. Given a cone C with data (ξ, Q) where ξ is a unit vector in $\mathbb{R}^d$ and Q is a cube, consider a second cone C' with data (ξ', C'). We can map one cone into the other with an orthogonal rotation and a dilation in $d-1$ variables. Thus, the corresponding Calderón–Zygmund kernels K_C and $K_{C'}$ can be mapped one into the other by way of these same transformations.

A rotation preserves the Calderón–Zygmund norm of the kernel, but the dilation does not since it is not uniform in all coordinates. Nevertheless, this observation shows that the Cone norms associated to distinct apertures are comparable. Thus, to prove our result, it suffices to demonstrate the existence of *some* aperture for which the Theorem is true. This we will do by taking a somewhat large aperture, that approximates a half space.

For a choice of symbol b with $\|b\|_{\mathrm{BMO}} = 1$ and $\|b\|_{\mathrm{BMO}_{-1}} < \delta_{-1}$, there is an associated open set U for which we achieve the supremum in the BMO norm. After an appropriate dilation, we can assume $\frac{1}{2} < |\mathrm{sh}(\mathcal{U})| \le 1$. Let $\mathcal{U} = \{R \in \mathcal{D}_{\vec{d}}\colon\ R \subset U\}$.

For a collection of rectangles $\mathcal{T}$ define the wavelet projection onto $\mathcal{T}$ as

$$\mathrm{P}_{\mathcal{T}} b \stackrel{\mathrm{def}}{=} \sum_{R\in\mathcal{T}}\sum_{\vec{\varepsilon}\in\mathrm{Sig}_{\vec{d}}} \langle b, w_R^{\vec{\varepsilon}}\rangle w_R^{\vec{\varepsilon}}.$$

Define $\beta = \mathrm{P}_{\mathcal{U}} b$. We use this function to build a test function to demonstrate a lower bound on the Cone norm.

The purpose of the next steps is to select the cones we will use. This issue involves some subtleties motivated by subsequent steps in the proof. Of particular importance is that the selection of the cones be dependent only on dimension, as well as satisfy some particular estimates. It turns out to be a useful device to select two distinct cones: One will be used for the selection of the test function for the commutator, and the other for the cones we use to define the commutator.

Given a cone C with data (ξ, Q), let H_C be the convolution operator with symbol $\mathbf{1}_{(0,\infty)}(\xi \cdot \theta)$, which is to say that H_C is the Fourier projection onto a half space associated with C.

LEMMA 7.1. *Given $\vec{d}$ and $\kappa > 0$ we can select cones*

$$D_s \subset C_s \subset \mathbb{R}^{d_s}, \qquad 1 \le s \le t.$$

These cones have data (ξ_s, Q_s) and (ξ_s, Q_s') respectively. They are, up to a rotation, only a function of $\vec{d}$ and $\kappa > 0$, and they satisfy these properties. Defining

$$\gamma \stackrel{\text{def}}{=} \mathrm{T}_{D_1} \cdots \mathrm{T}_{D_t} \beta \tag{7.2}$$

we have

$$\|\gamma\|_2 \ge 4^{-t}. \tag{7.3}$$

$$\|\mathrm{H}_{D_1} \cdots \mathrm{H}_{D_t} \beta - \gamma\|_4 \le \kappa. \tag{7.4}$$

$$\left\| (\mathrm{H}_{C_1} \cdots \mathrm{H}_{C_t} - \mathrm{P}_{C_1} \cdots \mathrm{P}_{C_t}) \, |\gamma|^2 \right\|_2 \le \kappa. \tag{7.5}$$

Notice that (7.4) estimates an L^4 norm; and that (7.5) concerns the function $|\gamma|^2$, and we are estimating the difference between the projections onto the half spaces defined by the cones, and the projection onto the larger cones.

Proof. We begin with the selection of the cones D_s, which is a randomized procedure. Fix a small constant $0 < \eta < \frac{1}{10}$. In each dimension $\mathbb{R}^{d_s}$, fix an aperture Q_s so that the cone D_s with this aperture satisfies

$$\mathbb{P}(D_s \cap S^{d_s - 1} \mid S^{d_s - 1}) \ge \tfrac{1}{2} - \eta.$$

Here, S^{d_s-1} denotes the sphere in $\mathbb{R}^{d_s}$ endowed with the canonical normalized surface measure. The notation above is the standard way to denote conditional probability.

Now, let D_s' denote a random rotation of the cone D_s. Taking expectations of L^2 norms below, we have access to the Plancherel identity to see that

$$\mathbb{E} \left\| \mathrm{P}_{D_1'} \cdots \mathrm{P}_{D_t'} \beta \right\|_2^2 = c_{\vec{d}} \, \mathbb{E} \int_{D_1' \otimes \cdots \otimes D_t'} \left| \widehat{\beta}(\vec{\xi}) \right|^2 \, d\vec{\xi} \ge (\tfrac{1}{2} - \eta)^t.$$

But also, we must have

$$\mathbb{E}\left\|(\mathrm{H}_{D_1'}\cdots\mathrm{H}_{D_t'}-\mathrm{P}_{D_1'}\cdots\mathrm{P}_{D_t'})\beta\right\|_2^2\le\eta^t.$$

View these statements about the L^2 norm of nonnegative random variables. As concerns the first inequality, note that

$$\sup_{D_1',\dots,D_t'}\left\|\mathrm{P}_{D_1'}\cdots\mathrm{P}_{D_t'}\beta\right\|_2^2\ \le\ 1.$$

Hence, we see that

$$\mathbb{P}(\left\|\mathrm{P}_{D_1'}\cdots\mathrm{P}_{D_t'}\beta\right\|_2^2\ge(\tfrac14)^t)\ \ge\ (\tfrac12-\eta)^t,$$
$$\mathbb{P}(\left\|(\mathrm{H}_{D_1'}\cdots\mathrm{H}_{D_t'}-\mathrm{P}_{D_1'}\cdots\mathrm{P}_{D_t'})\beta\right\|_2^2\ge\eta^{t/2})\ \le\ \eta^{t/2}.$$

Therefore, for η sufficiently small, we can select cones $D_1',\dots,D_t'$ so that

$$\left\|\mathrm{P}_{D_1'}\cdots\mathrm{P}_{D_t'}\beta\right\|_2^2\ \ge\ (\tfrac14)^t$$
$$\left\|(\mathrm{H}_{D_1'}\cdots\mathrm{H}_{D_t'}-\mathrm{P}_{D_1'}\cdots\mathrm{P}_{D_t'})\beta\right\|_2^2\ \le\ \eta^{t/2}.$$

On the other hand, we automatically have

$$\left\|(\mathrm{H}_{D_1'}\cdots\mathrm{H}_{D_t'}-\mathrm{T}_{D_1'}\cdots\mathrm{T}_{D_t'})\beta\right\|_8\le C\,\|\beta\|_8\le K$$

for an absolute constant K. Keeping in mind the fact that the symbol of a cone operator T_C is identically one on the cone C, we see that we have proved (7.3) and (7.4).

We use proof by contradiction to find the cones C_s. Fix the cones D_s as above, and let us suppose that (7.5) fails for some $\kappa>0$. Then, we can find a sequence of cones

$$C_s^k\subsetneq C_s^{k+1},\qquad 1\le s\le t,\ k\ge1$$

with data $(\xi_k,Q_{s,k}')$, where the apertures $Q_{s,k}'$ increase to all of $\mathbb{R}^{d_s-1}$.

We can also find functions β_k satisfying

$$\|\beta_k\|_2=1,\qquad \|\beta_k\|_8\le K_8,$$
$$\left\|(\mathrm{H}_{C_1^k}\cdots\mathrm{H}_{C_t^k}-\mathrm{P}_{C_1^k}\cdots\mathrm{P}_{C_t^k})\,|\gamma_k|^2\right\|_2\ge\kappa.$$

where γ_k is defined as in (7.2). The constant K_8 depends only on $\vec{d}$, and the John Nirenberg inequality for BMO.

In particular, as we can assume an upper bound on the L^8 norm of the β_k, the sequence of functions $\{\beta_k\}$ are precompact in the L^2 topology. Letting β_∞ be a limit point of the sequence of functions, and defining γ_∞ as in (7.2), we see that for all large k,

$$\left\| (\mathrm{H}_{C_1^k} \cdots \mathrm{H}_{C_t^k} - \mathrm{P}_{C_1^k} \cdots \mathrm{P}_{C_t^k}) \, |\gamma_\infty|^2 \right\|_2 \geq \kappa.$$

But this is an absurdity, as in the limit, the symbol of this difference is supported on a subspace of codimension t. Therefore, (7.5) holds. □

The cones we form the commutator of are the C_s of the previous Lemma. We will test the commutator against the function $\overline{\gamma}$, where γ is as in (7.2).

By Journé's Lemma, in particular Lemma 3.8, there will exist an open set V which satisfies the conditions of that Lemma. Set

$$\mathcal{V} \stackrel{\text{def}}{=} \{R : R \subset V, \ R \not\subset \operatorname{sh}(\mathcal{U})\}. \tag{7.6}$$

Finally, let $\mathcal{W} \stackrel{\text{def}}{=} \mathcal{D}_{\vec{d}} - \mathcal{U} - \mathcal{V}$.

The function γ enjoys these properties of Lemma 7.1, as well as the ones below, which will conclude the proof.

$$\|[\mathrm{T}_{C_1}, \cdots [\mathrm{T}_{C_t}, \mathrm{M}_{P_{\mathcal{U}} b}] \cdots] \overline{\gamma}\|_2 \gtrsim 1, \tag{7.7}$$

$$\|[\mathrm{T}_{C_1}, \cdots [\mathrm{T}_{C_t}, \mathrm{M}_{P_{\mathcal{V}} b}] \cdots] \overline{\gamma}\|_2 \lesssim \delta_J^{1/4}, \tag{7.8}$$

$$\|[\mathrm{T}_{C_1}, \cdots [\mathrm{T}_{C_t}, \mathrm{M}_{P_{\mathcal{W}} b}] \cdots] \overline{\gamma}\|_2 \leq K_J \delta_{-1}. \tag{7.9}$$

Here $0 < \delta_J < 1$ is the constant associated with Journé's Lemma (called η in Lemma 3.8), that is to be specified. K_J is a function of δ_J. These estimates will lead to an absolute lower bound and prove Theorem 6.1. Namely, the implied constants in each of the inequalities depend only on $\vec{d}$, while δ_J and δ_{-1} are free to choose. Certainly we can choose δ_J first, and then with K_J specified in (7.9), select δ_{-1} to prove our Theorem.

Estimate (7.8) is straightforward. It is easy to see that

$$\|[\mathrm{T}_{C_1}, \cdots [T_{C_t}, \mathrm{M}_{P_{\mathcal{V}} b}] \cdots] \overline{\gamma}\|_2 \lesssim \|\mathrm{P}_{\mathcal{V}} b\|_4 \, \|\gamma\|_4 \lesssim \|\mathrm{P}_{\mathcal{V}} b\|_4 ,$$

where the implied constant depends upon the the L^4 norms of the Cone transforms.

But, by Journé's Lemma 3.8 and construction, we have that

$$\|\mathrm{P}_{\mathcal{V}} b\|_2 \leq \delta_J^{1/2}, \qquad \|\mathrm{P}_{\mathcal{V}} b\|_{\mathrm{BMO}} \leq 1,$$

which implies that

$$\|\mathrm{P}_{\mathcal{V}} b\|_4 \leq \delta_J^{1/4}.$$

These together give (7.8).

We turn to the verification of (7.7) and provide a lower bound for the L^2 norm below

$$\|[\mathrm{T}_{C_1}, \cdots [\mathrm{T}_{C_t}, \mathrm{M}_\beta] \cdots] \overline{\gamma}\|_2 .$$

Recall that $\beta = \mathrm{P}_{\mathcal{U}} b$ and the definition of γ in (7.2). The commutator is a linear combination of terms $\mathrm{T}[\beta \cdot \mathrm{T}' \overline{\gamma}]$, where T and T′ are either the identity, or a product in T_{C_s}, $1 \leq s \leq t$. (Each T_{C_s} must occur in either T or T′.) In each case that T′ is a nontrivial product, we have $\mathrm{T}' \overline{\gamma} = 0$. It then remains to consider the only term not of this type, namely

$$\mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\beta \cdot \overline{\gamma}].$$

Write $\beta = \gamma + \beta' + \beta''$, where the smaller cones from Lemma 7.1 enter again below.

$$\begin{aligned} \beta' &= (\mathrm{H}_{C_1} \cdots \mathrm{H}_{C_t} - \mathrm{T}_{D_1} \cdots \mathrm{T}_{D_t})\beta, \\ \beta'' &= (\mathrm{I} - \mathrm{H}_{C_1} \cdots \mathrm{H}_{C_t})\beta. \end{aligned} \tag{7.10}$$

Note that (7.3) provides information about β'. We need to consider $\mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\beta \cdot \overline{\gamma}]$, which is now divided into three terms. They are

$$\begin{aligned} \mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\beta \cdot \overline{\gamma}] = {} & \mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\beta'' \cdot \overline{\gamma}] + \mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\beta' \cdot \overline{\gamma}] \\ & + \mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\gamma \cdot \overline{\gamma}]. \end{aligned}$$

Now, $\overline{\gamma}$ and β'' are supported on the same product of halfspaces, which are complementary to the cones, thus

$$\mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\beta'' \cdot \overline{\gamma}] = 0. \tag{7.11}$$

For β' we do not attempt to find any cancellation, just relying on the favorable estimate from (7.4).

$$\left\|\mathrm{T}_{C_1} \cdots \mathrm{T}_{C_t}[\beta' \cdot \overline{\gamma}]\right\|_2 \leq \|\beta'\|_4 \cdot \|\gamma\|_4 \lesssim \kappa. \tag{7.12}$$

The last term holds the essence of this component of the argument. By (7.5),

$$
\begin{aligned}
(7.13)\qquad \|\mathrm{T}_{C_1}\cdots\mathrm{T}_{C_t}[\gamma\cdot\overline{\gamma}]\|_2-\kappa &\geq \|\mathrm{H}_{C_1}\cdots\mathrm{H}_{C_t}[\overline{\gamma}\cdot\gamma]\|_2\\
&\gtrsim \|\overline{\gamma}\cdot\gamma\|_2\\
&= \|\gamma\|_4^2\\
&\gtrsim \left\|\left[\sum_{\vec{\varepsilon}\in\mathrm{Sig}_{\vec{d}}}\sum_{R\in\mathcal{U}}\frac{\left|\langle\gamma,w_R^{\vec{\varepsilon}}\rangle\right|^2}{|R|}\mathbf{1}_R\right]^{1/2}\right\|_4^2\\
&\gtrsim \left\|\left[\sum_{\vec{\varepsilon}\in\mathrm{Sig}_{\vec{d}}}\sum_{R\in\mathcal{U}}\frac{\left|\langle\gamma,w_R^{\vec{\varepsilon}}\rangle\right|^2}{|R|}\mathbf{1}_R\right]^{1/2}\right\|_2^2\\
&\gtrsim 1.
\end{aligned}
$$

The second line follows as the Fourier transform of $\overline{\gamma}\cdot\gamma$ is symmetric with respect to the half planes determined by the cones; the third line is obvious; the fourth line uses the Littlewood–Paley inequalities; and the fifth line uses the fact that the rectangles in $\mathcal{U}$ are in a set of measure at most one. This completes the proof of (7.7).

The remainder of the paper is devoted to proving (7.9) which is taken up in the next section.

The Proof of (7.9). Theorem 4.7 will let us conclude estimate (7.9), and the proof of Theorem 6.1.

We have observed that the commutator we are considering simplifies considerably when applied to the function $\overline{\gamma}$. Letting $\mathrm{T}=\mathrm{T}_{C_1}\cdots\mathrm{T}_{C_t}$, and $\mathrm{T}'=\mathrm{T}_{D_1}\cdots\mathrm{T}_{D_t}$ the estimate we are to prove is

$$
\|\mathrm{T}(\mathrm{P}_{\mathcal{W}}b\cdot\overline{\gamma})\|_2=\left\|\mathrm{T}(\mathrm{P}_{\mathcal{W}}b\cdot\overline{\mathrm{T}'\beta})\right\|_2\lesssim K_J\delta_{-1}.
$$

It is critical to observe that the outermost T imposes a cancellation condition similar to the one defining paraproducts. For $R\in\mathcal{U}$ and $R'\in\mathcal{W}$, and choices of signatures $\vec{\varepsilon},\vec{\varepsilon}'\in\mathrm{Sig}_{\vec{d}}$, we have

$$
\mathrm{T}(w_{R'}^{\vec{\varepsilon}'}\cdot\overline{\mathrm{T}'w_R^{\vec{\varepsilon}}})=0\qquad\text{if for any } 1\leq s\leq t,\ \left|Q_s'\right|>64\left|Q_s\right|.
$$

Recall that we defined $\gamma=\mathrm{T}'\beta$. T and T′ are a convolution operators, so the

Fourier support of $\mathrm{T}w_R^{\vec{\varepsilon}}$ is contained in the Fourier support of $w_R^{\vec{\varepsilon}}$. Therefore, this property follows immediately from the properties of the Meyer wavelet.

Using this observation, we see the estimate to be proved is

$$\left\| \sum_{\vec{\varepsilon},\vec{\varepsilon}\in \mathrm{Sig}_{\vec{d}}} \sum_{(R,R')\in\mathcal{A}} \overline{\langle b, w_R^{\vec{\varepsilon}}\rangle} \langle b, w_{R'}^{\vec{\varepsilon}'}\rangle \, \mathrm{T}(w_{R'}^{\vec{\varepsilon}} \cdot \overline{\mathrm{T}' w_R^{\vec{\varepsilon}}}) \right\|_2 \lesssim \delta_{-1}, \tag{7.14}$$

$$\mathcal{A} \stackrel{\text{def}}{=} \{(R,R')\colon\ R\subset U,\ R'\not\subset V, |Q_s'| \le 64\,|Q_s|\,, 1\le s\le t\}.$$

Notice that in the L^2 norm, we are free to remove the operator T, as it is a bounded operator on $L^2(\mathbb{R}^{\vec{d}})$.

It is essential to observe that this last sum can be written as a sum of paraproducts, as in Theorem 4.7. The purpose of these next definitions is to decompose the collection $\mathcal{A}$ into appropriate parts to which Theorem 4.7 applies. For an integer $n \ge 1$, take

$$\mathcal{U}_n \stackrel{\text{def}}{=} \{R\in\mathcal{U}\colon\ 2^{n-1}\le \mathrm{Emb}(R)\le 2^n\},$$

$$\mathcal{A}_n \stackrel{\text{def}}{=} \{(R,R')\colon\ R\in\mathcal{U}_n,\ R'\not\subset V, |Q_s'| \le 64\,|Q_s|\,, 1\le s\le t\}.$$

Here, Emb is the function supplied to us by Journé's Lemma, Lemma 3.8. Hence, as a consequence we have

$$\|\mathrm{P}_{\mathcal{U}_n} b\|_{\mathrm{BMO}} \lesssim 2^{Cn}\delta_{-1},$$

where C is a large constant depending only on $\vec{d}$. Observe that for $(R,R')\in\mathcal{A}_n$ we necessarily have $2^{n-1}R\cap R' = \emptyset$. In particular, the assumption (4.10) will hold with $A\simeq 2^n$. From this separation, and the rapid decay of the Meyer wavelet, we will gain an arbitrarily large power of 2^{-n}. Thus the presence of the term 2^{Cn} in this last estimate turns out not to be a concern for us.

Our estimate below is a consequence of (4.11), after a further decomposition of the sum to account for the role of the location of the zeros, controlled by the set J in (4.11).

$$\left\| \sum_{\vec{\varepsilon},\vec{\varepsilon}'\in \mathrm{Sig}_{\vec{d}}} \sum_{(R,R')\in\mathcal{A}_n} \overline{\langle b, w_R^{\vec{\varepsilon}}\rangle} \langle b, w_{R'}^{\vec{\varepsilon}'}\rangle \, (w_{R'}^{\vec{\varepsilon}} \cdot \overline{\mathrm{T}' w_R^{\vec{\varepsilon}}}) \right\|_2 \lesssim 2^{-n}\delta_{-1}, \quad n\ge 1. \tag{7.15}$$

Summation in $n\ge 1$ will then prove (7.14).

Specifically, let $J\subset\{1,\dots,t\}$, and let integer $\vec{k}\in\mathbb{Z}^t$ satisfy

$$k_s = 8,\ s\notin J, \qquad -8\le k_s\le 8,\ s\in J.$$

Let $\mathcal{A}_{n,J,\vec{k}}$ be a subset of $\mathcal{A}_n$ given by

$$\mathcal{A}_{n,J,\vec{k}} \stackrel{\text{def}}{=} \{(R,R') \in \mathcal{A}_n \colon \ |Q'_s| \le 2^{-8}\,|Q_s|\,,\ s \notin J, \quad |Q'_s| = 2^{k_s}\,|Q_s|\,,\ s \in J\}.$$

For this collection, the estimate

$$\left\| \sum_{\vec{\varepsilon},\vec{\varepsilon}' \in \operatorname{Sig}_{\vec{d}}} \sum_{(R,R') \in \mathcal{A}_{n,J,\vec{k}}} \overline{\langle b, w_R^{\vec{\varepsilon}} \rangle}\, \langle b, w_{R'}^{\vec{\varepsilon}'} \rangle\, (w_{R'}^{\vec{\varepsilon}} \cdot \overline{\mathrm{T}' w_R^{\vec{\varepsilon}}}) \right\|_2 \lesssim 2^{-n} \delta_{-1}$$

is then a consequence of (4.11). This estimate is summed over $J \subset \{1,\dots,n\}$ and $\vec{k}$ to prove (7.15).

7.2. The lower bound on the Riesz transforms.

Properties of Riesz transforms. We need some special properties of Riesz transforms. Variants are in the paper of Coifman, Rochberg and Weiss [8].

PROPOSITION 7.16. *For each s, let T_s be a polynomial in the Riesz transforms on $\mathbb{R}^{d_s}$. Then, we have the inequality*

$$\|[\cdots[\mathrm{M}_b, \mathrm{T}_1], \cdots], \mathrm{T}_t]\|_{2\to 2} \lesssim \sup_{\vec{j}} \left\|\mathrm{C}_{\vec{j}}(b,\cdot)\right\|_{2\to 2}.$$

The implied constant depends upon the choice of polynomials T_s.

This in fact follows from the elementary identity

$$[\mathrm{M}_b, \mathrm{R}_j \mathrm{R}_k] = [\mathrm{M}_b, \mathrm{R}_j]\mathrm{R}_k + \mathrm{R}_j[\mathrm{M}_b, \mathrm{R}_k].$$

An operator T which is a polynomial in Riesz transforms is a convolution operator, with radial symbol. Below, we will only describe the symbols that we are interested in.

The selection of the operators which are polynomials in Riesz transforms is hardly obvious, and we identify their properties in the following Lemma.

LEMMA 7.17. *Given any $0 < \eta < 1$ and any cone C in $\mathbb{R}^d$, there is an operator U_C, a linear combination of the identity and a polynomial in Riesz transforms on $\mathbb{R}^d$, with symbols υ_C such that*

$$\begin{cases} |\upsilon_C(\xi) - 1| < \eta & \xi \in C \\ |\upsilon_C(\xi)| < \eta & \xi \in -C. \end{cases} \tag{7.18}$$

Finally, we have the estimate

$$\|\mathrm{U}_C\|_p \lesssim C_p, \qquad 1 < p < \infty. \tag{7.19}$$

The constant C_p is independent of the choice of the cone C and dimension d.

Proof. This depends upon particular properties of spherical harmonics and zonal polynomials. We were aided by [2] in our search for this proof.

It suffices to prove the following. For a choice of cone C in $\mathbb{R}^d$, with direction ξ_C, and $0 < \eta < 1$, we can choose operator U with symbol v which is odd with respect to ξ_C, $\|v\|_\infty \le 2$, and

$$\begin{cases} |v(\xi) + 1| < \eta & \xi \in -C, \\ |v(\xi) - 1| < \eta & \xi \in C. \end{cases} \tag{7.20}$$

Finally, v restricted to the unit sphere is a polynomial in ξ_j for $(\xi_1, \xi_2, \dots, \xi_d) \in S^{d-1}$. We will see that the degree is at most $\lesssim \eta^{-1} \log 1/\eta$.

Then, U is in fact a polynomial in Riesz transforms. Since v is odd, the method of rotations applies to provide us with an estimate of $\|\mathrm{U}\|_p \le C_p$, where C_p is absolute for $1 < p < \infty$. To get an operator with symbol as in our Lemma, we add the identity operator to U.

We obtain the symbol v by employing the Poisson kernel in the ball in $\mathbb{R}^d$, and as well an expansion of this kernel into zonal harmonics. Let us recall the properties we need. The Poisson kernel in $\mathbb{R}^d$ is

$$P(x, \zeta) = \frac{1 - |x|^2}{|x - \zeta|^d}, \qquad |x| < 1,\ |\zeta| = 1.$$

A homogeneous harmonic polynomial p of degree m on $\mathbb{R}^d$ has the reproducing formula [2, p. 97]

$$p(x) = |x|^m \int_{S^{d-1}} p(\zeta) Z_m(x, \zeta)\, \sigma(d\zeta).$$

Here, σ denotes normalized Lebesgue measure on the unit sphere S^{d-1} in $\mathbb{R}^d$. The polynomial $Z_m(x, \zeta)$ is a *zonal polynomial of degree m*. It follows that the Poisson kernel admits an expansion in terms of these polynomials

$$P(x, \xi) = \sum_{m=0}^{\infty} Z_m(x, \zeta).$$

This series is absolutely convergent sum in the interior of the unit ball, thanks to the elementary estimate

$$|Z_m(x,\zeta)| \lesssim m^{d-2}\,|x|^m\,. \tag{7.21}$$

It is a basic property of the zonal polynomials that they are only a function of $|x|$ and $x\cdot\zeta$. Indeed, they can be expanded as

$$Z_m(x,\zeta) = \sum_{k=0}^{[m/2]} c_{k,m}(x\cdot\zeta)^{m-2k}\,|x|^{2k}\,.$$

Here, $c_{k,m}$ are known real coefficients. In particular, for a function $\widetilde{v}$ on the unit sphere that is odd, the new function

$$\int \widetilde{v}(\zeta) Z_m(x,\zeta)\,d\sigma(\zeta)$$

is also odd, for $|x|$ held fixed.

To construct our operator U, via its symbol v, recall that the operator U is associated to a cone C with direction ξ_C. On the cone C and the opposite cone $-C$ we require rather precise information about the symbol v. Outside of these cones we only require an absolute upper bound on v. Hence, we have some freedom in taking an initial approximate to the symbol v. In what follows, we concentrate on defining the symbol on the sphere S^{d-1}.

Take as an initial approximate $\widetilde{\widetilde{v}}(\xi) = \operatorname{sign}(\xi\cdot\xi_C)$. For a small constant c, consider the function

$$\widetilde{v}(\theta) \stackrel{\text{def}}{=} \int \widetilde{\widetilde{v}}(\zeta) P((1-c\eta)\theta,\zeta)\,d\sigma(\zeta).$$

This function will be nonnegative, odd, bounded in absolute value by 1, and satisfy (7.20). It is not however a polynomial in spherical harmonics.

But each function

$$v_m(\theta) \stackrel{\text{def}}{=} \int \widetilde{\widetilde{v}}(\zeta) Z_m((1-c\eta)\theta,\zeta)\,d\sigma(\zeta)$$

is also odd, as we have already noted. By (7.21), we have the estimate

$$\Bigl\| \sum_{m=m_0}^{\infty} v_m \Bigr\|_{L^\infty(S^{d-1})} \le \eta/4, \qquad m_0 = C(\log 1/\eta)/\eta$$

where C depends upon the dimension d. Therefore, the function υ we need can be taken to be

$$\sum_{m=0}^{m_0} \upsilon_m.$$

Our proof is complete. □

The selection of a test function. We continue to assume that the symbol b satisfies $\|b\|_{\mathrm{BMO}} = 1$ while $\|b\|_{\mathrm{BMO}_{-1}} < \delta_{-1}$. We follow many of the initial stages of the proof of the lower bound on the Cone norm. We choose cones $D_s \subset C_s$, and cone operators $\mathrm{T}_{D_s}, \mathrm{T}_{C_s}$, for $1 \le s \le t$ just as in Lemma 7.1.

We continue to use the notations $\mathcal{U}$, $\mathcal{V}$ and $\mathcal{W}$, thus $|\mathrm{sh}(\mathcal{U})| \simeq 1$ and

$$\sum_{R \in \mathcal{U}} |\langle b, w_R \rangle|^2 \simeq |\mathrm{sh}(\mathcal{U})| .$$

$\mathcal{V}$ and $\mathcal{W}$ are defined as in (7.6), and $\beta = \mathrm{P}_{\mathcal{U}} b$. As before, we set $\gamma = \mathrm{T}_{D_1} \cdots \mathrm{T}_{D_t} \beta$.

For $0 < \eta < 1$ to be chosen, apply Lemma 7.17, obtaining operators T_s which are a linear combinations of the identity and polynomials in Riesz transforms on $\mathbb{R}^{d_s}$ which approximate the projection operator P_{C_s} in the sense of that Lemma. Let us see that we have the estimate

$$\|[\mathrm{T}_1, \cdots [\mathrm{T}_t, \mathrm{M}_\beta] \cdots] \overline{\gamma}\|_2 \gtrsim 1. \tag{7.22}$$

The commutator is a linear combination of 2^t terms of the form

$$\mathrm{T}[\beta \mathrm{T}' \overline{\gamma}]$$

where T and T′ are either the identity or a composition of the operators T_s. If T′ is not the identity, it follows that the symbol of T′ is at most η on the Fourier support of $\overline{\gamma}$. Therefore, we can estimate

$$\|\mathrm{T}[\beta \mathrm{T}' \overline{\gamma}]\|_2 \lesssim \|\beta \mathrm{T}' \overline{\gamma}\|_2 \lesssim \|\beta\|_4 \, \|\mathrm{T}' \overline{\gamma}\|_4 \lesssim \eta^{1/3}.$$

This last estimate follows from $\|\mathrm{T}' \overline{\gamma}\|_2 \lesssim \eta$, while $\|\mathrm{T}' \overline{\gamma}\|_8 \lesssim 1$. This point is imperative, and follows from the uniform L^p bounds we obtain from Lemma 7.17.

This leaves the term $\mathrm{T}_1 \cdots \mathrm{T}_t \beta \overline{\gamma}$. But, for a sufficiently small choice of η, we are free to use the same argument as in (7.13). This proves (7.22).

It then follows from Proposition 7.16 that for some choice of Riesz transforms R_{j_s} on $\mathbb{R}^{d_s}$ we have

$$\|\mathrm{C}(\beta, \overline{\gamma})\|_2 \gtrsim 1, \quad \text{where} \tag{7.23}$$

$$\mathrm{C}(f, g) \stackrel{\mathrm{def}}{=} [\mathrm{R}_{j_1}, \cdots [\mathrm{R}_{j_t}, \mathrm{M}_f] \cdots] g.$$

Now, it also follows that

$$\|\mathrm{C}(\mathrm{P}_{\mathcal{V}} b, \overline{\gamma})\|_2 \lesssim \delta_J^{1/4}. \tag{7.24}$$

Indeed, we can appeal to the same argument as used to prove (7.8).

Finally, we claim that

$$\|\mathrm{C}(\mathrm{P}_{\mathcal{W}} b, \overline{\gamma})\|_2 \lesssim \delta_{-1}. \tag{7.25}$$

This estimate requires the same argument as for (7.9), plus an additional estimate; the details are below. The three inequalities (7.23), (7.24) and (7.25) are then combined in in the same manner as in the proof of the lower bound on the Cone norm to complete the proof.

Proof of (7.25). The different quantitative estimates we have for paraproducts are brought to bear on this estimate. First, we expand the expression $\mathrm{C}(\mathrm{P}_{\mathcal{W}} b, \gamma)$ into the sum of commutators on different pairs of wavelets. This sum is further written as $D_1 + D_2$, where we define D_1 explicitly here.

$$D_1 \stackrel{\text{def}}{=} \sum_{\vec{\varepsilon},\vec{\epsilon}\in \mathrm{Sig}_{\vec{d}}} \sum_{(R,R')\in\mathcal{A}} \overline{\langle \gamma, w_R^{\vec{\varepsilon}} \rangle} \langle b, w_{R'}^{\vec{\epsilon}} \rangle \, \mathrm{C}(w_{R'}^{\vec{\epsilon}}, \overline{w_R^{\vec{\varepsilon}}}) \tag{7.26}$$

$$\mathcal{A} \stackrel{\text{def}}{=} \{(R,R') \colon R \in \mathcal{U},\ R' \not\subset V, |Q_s'| \le 64\,|Q_s|\,, 1 \le s \le t\}.$$

This is the part of the commutator that most closely resembles the part arising from the commutator arising from the Cone operators.

It is essential to observe that this last sum can be written as a finite sum of compositions of Riesz transforms and the "paraproducts" in Theorem 4.7, and in particular the more technical estimate (4.11), applied to the functions $\mathrm{P}_{\mathcal{U}} b$ and $\mathrm{P}_{\mathcal{W}} b$. We also comment that the Riesz transforms applied to the wavelet element $w_R^{\vec{\varepsilon}}$ do not substantially change the localization properties of the wavelet, and thus the Riesz transforms do not spoil the estimates that appear in Theorem 4.7. This sum varies of choices of $\vec{k}$ with $\left\|\vec{k}\right\|_\infty \le 8$, and arbitrary $J \subset \{1, \dots, t\}$. (The subset J consists of those coordinates s for which $|Q_s| = 2^{k_s}\,|Q_s'|$.)

We will need to decompose the collection $\mathcal{A}$ into appropriate parts to which this estimate applies. That is the purpose of this definition. For an integer $n \ge 1$, take

$$\gamma_n \stackrel{\text{def}}{=} \sum_{\vec{\varepsilon}\in \mathrm{Sig}_{\vec{d}}} \sum_{\substack{R\subset U \\ 2^{n-1} \le \mathrm{Emb}(R) \le 2^n}} \langle \gamma, w_R^{\vec{\varepsilon}} \rangle w_R^{\vec{\varepsilon}}$$

We claim that

$$\|\mathrm{C}(\mathrm{P}_{\mathcal{W}} b, \overline{\gamma_n})\|_2 \lesssim 2^{-n} \delta_{-1}. \tag{7.27}$$

It follows from Lemma 3.8 that we have the estimate

$$\|\gamma_n\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} \lesssim 2^{2tn}\delta_{-1}, \tag{7.28}$$

indeed, this is the point of this definition. From other parts of the expansion of the Riesz commutator we need to find some decay in n.

Nevertheless, from this estimate and the upper bound on Riesz commutator norms, we have the estimate

$$\|\mathrm{C}(\mathrm{P}_{\mathcal{W}}b, \overline{\gamma_n})\|_2 \lesssim \|b\|_{\mathrm{BMO}(\mathbb{R}^{\vec{d}})} \|\gamma_n\|_2 \lesssim 2^{2tn}\delta_{-1}.$$

We use this estimate for $n < 20$, say.

For $n \geq 20$, $R \in \mathcal{U}$ with $2^{n-1} \leq \mathrm{Emb}(R) \leq 2^n$, and rectangle R' with $(R, R') \in \mathcal{A}$, it follows that we must have $2^{n-9}R \cap R' = \emptyset$. That is, (4.10) is satisfied with the value of A in that display being $A \simeq 2^n$ for $n \geq 20$. Thus, we conclude that

$$\|\mathrm{C}(\mathrm{P}_{\mathcal{W}}b, \overline{\gamma_n})\|_2 \lesssim 2^{-50n}\delta_{-1}, \qquad n \geq 20.$$

This completes our proof of (7.27), and the proof estimate (7.9) .

It remains to estimate the term D_2. The principal tool here is the estimate for Riesz commutators given in Lemma 5.16, and in order to apply this lemma, as well as take advantage of our remaining freedom to select the δ_{-1} norm, we need a sophisticated decomposition of the sum that controls D_2. That is the point of these next definitions.

Let m be an integer. For a non empty subset $J \subset \{1, \dots, t\}$, and choices of integers $\vec{a} = (a_j)_{j \in J}$ with $a_j \geq 7$, we define

$$D(m, J, \vec{a}) \stackrel{\text{def}}{=} \sum_{\vec{\varepsilon}, \vec{\varepsilon}' \in \mathrm{Sig}_{\vec{d}}} \sum_{(R,R') \in \mathcal{A}(m,J,\vec{a})} \overline{\langle \gamma, w_R^{\vec{\varepsilon}} \rangle} \, \langle b, w_{R'}^{\vec{\varepsilon}'} \rangle \, \mathrm{C}(w_{R'}^{\vec{\varepsilon}'}, \overline{w_R^{\vec{\varepsilon}}}) \tag{7.29}$$

$$\begin{aligned} \mathcal{A}(m, J, \vec{a}) \stackrel{\text{def}}{=} \{(R, R') &: R \in \mathcal{U},\ R' \not\subset V;\ |Q_s'| \leq 64\,|Q_s|\,, s \notin J; \\ &|Q_s'| = 2^a\,|Q_s|\,,\ s \in J;\ 2^m \leq \mathrm{Emb}(R) \leq 2^{m+1}\}. \end{aligned}$$

With this definitions, we will have

$$D_2 = \sum_{m, J, \vec{a}} D(m, J, \vec{a}).$$

The estimate below holds.

$$\|D(m, J, \vec{a})\|_2 \lesssim 2^{-m - \sum_s a_s}\delta_{-1}. \tag{7.30}$$

This is summable over the parameters $m, J, \vec{a}$, and so completes the proof of the estimate for D_2.

Recall that we have the essential consequence of Journé's Lemma.

$$\left\| \sum_{\substack{R\in\mathcal{U}\\ 2^m\le \mathrm{Emb}(R)\le 2^{m+1}}} \sum_{\varepsilon\in \mathrm{Sig}_{\vec{d}}} \langle \gamma, w_R^{\varepsilon}\rangle w_R^{\varepsilon} \right\|_{\mathrm{BMO}} \lesssim 2^{C_{\vec{d}}\, m}\delta_{-1}. \tag{7.31}$$

The first subcase occurs when we have

$$\max_{s\in J} a_s \le \frac{m}{4}. \tag{7.32}$$

It follows from the definition of embeddedness that $Q_{s'} \cap 2^{m/4} Q_s = \emptyset$ for all $(R, R') \in \mathcal{A}(m, J, \vec{a})$. Therefore, the function

$$2^{-Nm}\sqrt{|Q_{s'}|}\,\mathrm{C}\,(w_{R'}^{\vec{\varepsilon}'}, \overline{w_R^{\vec{\varepsilon}}}) \tag{7.33}$$

is adapted to Q_s, where $N \ge 1$ can be taken arbitrarily. This sum can be understood as a paraproduct, with symbol given by $\mathrm{P}_{\mathcal{W}} b$, applied to γ. Zeros fall on γ for those coordinates $s \in J$. Using (7.31) and the estimate (7.32), we see that (7.30) holds in this case.

The second case is when (7.32) fails in any coordinate, say $s_0 \in J$. In this instance, we see that the sum we are considering in that coordinate is of the type considered in Lemma 5.16. That is, the sum is an operator whose paraproduct norm, as defined in (4.16) is at most $2^{-Na_{s_0}}$, for arbitrarily large N. In all other coordinates, the sum is an operator with paraproduct norm at most a constant. The tensor product of paraproducts is a bounded operator, therefore in this case, we have

$$\|D(m, J, \vec{a})\|_2 \lesssim 2^{-N\sum a_s} 2^{C_{\vec{d}} m}\delta_{-1} \lesssim 2^{-2\sum a_s}\delta_{-1}.$$

Here of course, we again rely upon (7.31). This completes the proof of (7.30). This in turn completes the proof of the lower bound on the norm of multiparameter Riesz commutators.

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332
E-mail: lacey@math.gatech.edu

Department of Mathematics, University of Texas at Austin, Austin, TX 78712
E-mail: stefanie@math.utexas.edu

Department of Mathematics, Brown University, Providence, RI 02912
E-mail: jpipher@math.brown.edu

Department of Mathematics, Vanderbilt University, Nashville, TN 37240
E-mail: brett.d.wick@vanderbilt.edu

REFERENCES

[1] P. Auscher and M. E. Taylor, Paradifferential operators and commutator estimates, *Comm. Partial Differential Equations* **20** (1995), no. 9–10, 1743–1775.

[2] S. Axler, P. Bourdon, and W. Ramey, *Harmonic Function Theory*, 2nd ed., *Grad. Texts in Math.*, vol. 137, Springer-Verlag, New York, 2001.

[3] C. Cabrelli, M. Lacey, U. Molter, and J. C. Pipher, Variations on the theme of Journe's Lemma, *Houston J. Math.* **32** (2006), no. 3, 833–861 (electronic).

[4] L. Carleson, A counterexample for measures bounded on H^p spaces for the bidisk, *Mittag-Leffler Rep.*, no. 7, Inst. Mittag-Leffler (1974).

[5] S.-Y. A. Chang and R. Fefferman, Some recent developments in Fourier analysis and H^p-theory on product domains, *Bull. Amer. Math. Soc. (N.S.)* **12** (1985), no. 1, 1–43.

[6] ———, A continuous version of duality of H^1 with BMO on the bidisc, *Ann. of Math. (2)* **112** (1980), no. 1, 179–201.

[7] R. Coifman, P.-L. Lions, Y. Meyer, and S. Semmes, Compensated compactness and Hardy spaces, *J. Math. Pures Appl. (9)* **72** (1993), no. 3, 247–286 (English, with English and French summaries).

[8] R. R. Coifman, R. Rochberg, and G. Weiss, Factorization theorems for Hardy spaces in several variables, *Ann. of Math. (2)* **103** (1976), no. 3, 611–635.

[9] R. Fefferman, Bounded mean oscillation on the polydisk, *Ann. of Math. (2)* **110** (1979), no. 2, 395–406.

[10] ———, A note on Carleson measures in product spaces, *Proc. Amer. Math. Soc.* **93** (1985), no. 3, 509–511.

[11] ———, Harmonic analysis on product spaces, *Ann. of Math. (2)* **126** (1987), no. 1, 109–130.

[12] S. H. Ferguson and M. T. Lacey, A characterization of product BMO by commutators, *Acta Math.* **189** (2002), no. 2, 143–160.

[13] J.-L. Journé, Calderón-Zygmund operators on product spaces, *Rev. Mat. Iberoamericana* **1** (1985), no. 3, 55–91.

[14] ———, A covering lemma for product spaces, *Proc. Amer. Math. Soc.* **96** (1986), no. 4, 593–598.

[15] ———, Two problems of Calderón-Zygmund theory on product-spaces, *Ann. Inst. Fourier (Grenoble)* **38** (1988), no. 1, 111–132.

[16] B. C. Krikeles, Tensor products of multilinear operators, *J. Math. Anal. Appl.* **116** (1986), no. 1, 230–244.

[17] M. T. Lacey and J. Metcalfe, Paraproducts in one and several parameters, *Forum Math.* **19** (2007), no. 2, 325–351.

[18] M. T. Lacey, Commutators with Riesz potentials in one and several parameters, *Hokkaido Math. J.* **36** (2007), no. 1, 175–191.

[19] M. T. Lacey and E. Terwilleger, Hankel operators in several complex variables and product BMO, *Houston J. Math.* (to appear).

[20] M. T. Lacey, E. Terwilleger, and B. D. Wick, Remarks on product VMO, *Proc. Amer. Math. Soc.* **134** (2006), no. 2, 465–474 (electronic).

[21] S.-Y. Li, Characterization of the boundedness for a family of commutators on L^p, *Colloq. Math.* **70** (1996), no. 1, 59–71.

[22] Y. Meyer, Wavelets and operators, *Analysis at Urbana*, vol. 1 (Urbana, IL, 1986–1987), 1989, pp. 256–365.

[23] ———, *Ondelettes et opérateurs. I*, *Actualités Mathématiques. [Current Mathematical Topics]*, Hermann, Paris, 1990.

[24] C. Mucalu, J. Pipher, T. Tao, and C. Thiele, Bi-parameter paraproducts, *Acta Math.* **193** (2004), no. 2, 269–296.

[25] ———, Multi-parameter paraproducts, *Rev. Mat. Iberoamicana* **22** (2006), no. 3, 963–976.

[26] Z. Nehari, On bounded bilinear forms, *Ann. of Math. (2)* **65** (1957), 153–162.

[27] S. Petermichl, Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol, *C. R. Acad. Sci. Paris Sér. I Math.* **330** (2000), no. 6, 455–460 (English, with English and French summaries).

[28] J. Pipher, Journé's covering lemma and its extension to higher dimensions, *Duke Math. J.* **53** (1986), no. 3, 683–690.

[29] A. Uchiyama, On the compactness of operators of Hankel type, *Tôhoku Math. J. (2)* **30** (1978), no. 1, 163–171.