HARDY SPACES AND A WALSH MODEL FOR BILINEAR CONE OPERATORS

John E. Gilbert and Andrea R. Nahmod

April 11, 1997

In memory of J.-A. Chao

Abstract. The study of bilinear operators associated to a class of non-smooth symbols can be reduced to the study of certain special bilinear cone operators to which a time frequency analysis using smooth wave-packets is performed. In this paper we prove that when smooth wave-packets are replaced by Walsh wave-packets the corresponding discrete Walsh model for the cone operators is not only L^p-bounded, as Thiele has shown in his thesis for the Walsh model corresponding to the bilinear Hilbert transform, but actually improves regularity as it maps into a Hardy space. The same result is expected to hold for the special bilinear cone operators.

1. Introduction.

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

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

If m is homogeneous of degree 0, such a $\mathcal{B}$ commutes also with simultaneous dilations. The basic L^p-boundedness problem is to prescribe conditions on m so that $\mathcal{B}$ extends to a bounded operator from $L^p(\mathbb{R}) \times L^q(\mathbb{R})$ into $L^r(\mathbb{R})$; in the dilation-invariant case the exponents p, q and r must follow the same pattern as in Hölder's inequality, *i.e.*, $1/p + 1/q = 1/r$.

A necessary condition for L^p-boundedness is that m be *bounded*, but it is not sufficient. On the other hand, *smoothness* of m is sufficient but not necessary. In particular, if m is sufficiently smooth, then $\mathcal{B}$ maps $L^p \times L^{p'}$ into L^1 ([8]); furthermore, for a *smooth* symbol the cancellation condition

$$m(\xi,-\xi) = 0 \qquad (\xi \neq 0) \tag{1.2}$$

1991 *Mathematics Subject Classification*. Primary 42B15, 42B30. Secondary 42B25.

improves the boundedness to $L^p \times L^{p'} \to H^1$ ([7]). Such an improvement in regularity comes into play in compensated compactness, because it allows use of weak* compactness arguments to ensure convergence of approximating solutions to partial differential equations. Notice that (1.2) is equivalent to the vanishing moment condition

$$\int_{-\infty}^{\infty} \mathcal{B}(f,g)(x)\,dx = 0 \tag{1.3}$$

whenever the integral is well-defined since

$$\int_{-\infty}^{\infty} \mathcal{B}(f,g)(x)\,dx = \int_{-\infty}^{\infty} m(\xi,-\xi)\,\hat{f}(\xi)\,\hat{g}(-\xi)\,d\,\xi\,.$$

Perhaps the most fundamental example having *non-smooth* symbol arises from the Hilbert transform $\mathcal{H}$ in the form of the pointwise product

$$f, g \longrightarrow \tfrac{1}{4}(f + i\mathcal{H}f)(x)(g + i\mathcal{H}g)(x) = \int_0^{\infty}\int_0^{\infty} \hat{f}(\xi)\hat{g}(\eta)\,e^{2\pi i x(\xi+\eta)}\,d\xi d\eta \tag{1.4}$$

of the projections of f, g on complex Hardy spaces. Virtually by construction it maps $L^p(\mathbb{R}) \times L^q(\mathbb{R})$ into complex Hardy space $H^r_{\mathbb{C}}(\mathbb{R})$, $1/p + 1/q = 1/r$. The imaginary part

$$f, g \longrightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\mathrm{sgn}(\xi) + \mathrm{sgn}(\eta))\hat{f}(\xi)\hat{g}(\eta)\,e^{2\pi i x(\xi+\eta)}\,d\xi d\eta \tag{1.5}$$

of (1.4) is the prototype $\mathcal{T} = \mathcal{H}$ of the Coifman-Rochberg-Weiss commutator $\mathcal{B}(f,g) = \mathcal{T}(f)g - f\mathcal{T}^*(g)$ occurring in the 'div-curl' lemma. This prototype satisfies (1.2) and maps $L^p(\mathbb{R}) \times L^q(\mathbb{R})$ into the real Hardy space $H^r(\mathbb{R})$, while any linear combination of $\mathcal{H} \otimes I$ and $I \otimes \mathcal{H}$ always maps $L^p(\mathbb{R}) \times L^q(\mathbb{R})$ into $L^r(\mathbb{R})$. What makes the L^p-boundedness so easy to establish for all of these examples is that they are sums of operators 'respecting' a tensor product structure in the sense that each such $\mathcal{B}$ is the composition

$$L^p(\mathbb{R}) \otimes_\gamma L^q(\mathbb{R}) \quad \overset{\mathcal{K}}{\longrightarrow} \quad L^p(\mathbb{R}) \otimes_\gamma L^q(\mathbb{R}) \quad \overset{\mathcal{M}}{\longrightarrow} \quad L^r(\mathbb{R}) \tag{1.6}$$

of a singular integral operator $\mathcal{K}$ and pointwise multiplication $\mathcal{M} : f \otimes g \to f(x)g(x)$. The decomposition (1.6) occurs whenever m is constant on each quadrant of $\mathbb{R} \times \mathbb{R}$, but if the symbol is, say, a rotation of one of

$$m(\xi,\eta) = -i(\mathrm{sgn}(\xi) + \mathrm{sgn}(\eta)), \qquad m(\xi,\eta) = -i\ \mathrm{sgn}(\xi)$$

through some angle other than a multiple of $\pi/2$, this simple argument fails and L^p-boundedness is no longer clear.

To understand the general non-smooth case, therefore, we need to look at operators whose symbol 'cuts across' quadrants. Such operators arise naturally not only in connection with compensated compactness phenomena but also with Calderón's first commutator

$$\mathcal{C}^{(1)}_A(f)(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{A(x) - A(y)}{(x-y)^2} f(y)\,dy$$

where A has a derivative $A' = a \in L^\infty(\mathbb{R})$. Indeed, it is well known that $\mathcal{C}_A^{(1)}$ can be realized as a weighted average

$$\langle \mathcal{C}_A^{(1)}(f),\, g\,\rangle = \int_0^{\pi/2} \langle B_{-\theta}(f,\, g),\, a\,\rangle \,\frac{1}{(\sin\theta+\cos\theta)^2}\, d\theta$$

of a family

$$B_\theta(f,\, g) = -i \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \operatorname{sgn}(\xi\cos\theta + \eta\sin\theta)\, \hat{f}(\xi)\,\hat{g}(\eta)\, e^{2\pi i x(\xi+\eta)}\, d\xi d\eta$$

of dilation-invariant bilinear operators whose symbols lack both smoothness and cancellation ([9]). Each B_θ can also be written as a principal value singular integral

$$B_\theta(f,g)(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(x-t\cos\theta)\, g(x-t\sin\theta)\,\frac{dt}{t}. \tag{1.7}$$

The simple cases $\mathcal{H}\otimes I$ and $I\otimes\mathcal{H}$ correspond to $\theta = 0,\, \pi/2,\, \pi, 3\pi/2$; all of these map $L^p \times L^{p'} \to L^1$. By contrast, $B_{\pi/4}$ reduces to the Hilbert transform $\mathcal{H}(fg)$ of the pointwise product of f, g and so is bounded from $L^2\times L^2$ into weak-L^1, not into L^1. The Bilinear Hilbert transform is the case $\theta = -\pi/4$; recently, Lacey and Thiele established its $L^p\times L^q \to L^r$-boundedness for $2 < p, q < \infty$ using a very delicate time-frequency analysis in the spirit of Fefferman's proof of the almost everywhere convergence of Fourier series of L^2-functions ([13], [19]).

The previous discussion suggests that the most inclusive criterion for L^p-boundedness should deal simultaneously with smooth and non-smooth symbols. For simplicity we shall concentrate here on the dilation-invariant case. By a *piecewise* C^k-function on the unit circle Σ_1 in $\mathbb{R}^2$, we shall mean a function having bounded, continuous derivatives up to order k on the complement of a finite set in Σ_1 and one-sided derivatives at the end-points.

(1.8) Conjecture. *Let $m_0 = m_0(\xi,\eta)$ be a piecewise C^k-function on Σ_1 which is C^k in a neighborhood of the points $(\xi,-\xi)$ and let $\mathcal{B}$ be the bilinear operator whose symbol is the degree zero homogeneous extension of m_0. Then $\mathcal{B} : L^p(\mathbb{R})\times L^{p'}(\mathbb{R}) \to L^1(\mathbb{R})$; in addition,*

(i) *if $m_0(\xi,-\xi) = 0$, then $\mathcal{B} : L^p(\mathbb{R})\times L^{p'}(\mathbb{R}) \to H^1(\mathbb{R})$, while*
(ii) *if $m_0(\xi,\eta) = 0$ when $\xi+\eta \le 0$, then $\mathcal{B} : L^p(\mathbb{R})\times L^{p'}(\mathbb{R}) \to H^1_{\mathbb{C}}(\mathbb{R})$.*

The idea underlying the conjecture is to rule out discontinuities of the symbol along the diagonal $\xi+\eta = 0$ and so avoid the operator $\mathcal{H}(fg)$ which does not map into L^1, while at the same time allowing all examples known to us of (dilation-invariant) bilinear operators with smooth or non-smooth symbol that map at least into L^1. By writing m_0 as a sum of functions supported on closed intervals in Σ_1 it is enough to consider bilinear *Cone operators*

$$\mathcal{C}_{\Gamma_\alpha}(f,g)(x) = \int\int_{\Gamma_\alpha} m(\xi,\eta)\hat{f}(\xi)\;\hat{g}(\eta)\; e^{2\pi i x\cdot(\xi+\eta)}\, d\xi d\eta \tag{1.9}$$

whose symbol m is homogeneous of degree 0 and has bounded continuous derivatives up to order k inside a cone Γ_α of aperture α having vertex at the origin; the hypotheses of the conjecture require that the diagonal $\xi+\eta=0$ not be an edge of the support cone. For example, set

$$A_{\theta,\alpha}(f,g)=\tfrac{1}{2}i\,\{B_{\theta+\alpha}(f,g)-B_\theta(f,g)\}.$$

Its symbol is supported on a *double cone* of aperture α inside of which it is constant. Neither cone intersects the diagonal $\xi+\eta=0$ so long as $-3\pi/4<\theta<\theta+\alpha<\pi/4$; in this case $A_{\theta,\alpha}$ can then be written as the sum of two cone operators corresponding to a cone lying in the half-plane $\xi+\eta>0$ and the reflection of this cone in the origin. In general we could allow the symbol in (1.9) to be non-homogeneous so long as the derivatives up to order k are still bounded.

The conjecture arises quite naturally from time-frequency analysis and Fourier plane geometry of the cone operators. Suppose first that $\alpha\le\pi/4$ and that one edge of Γ_α lies on the positive ξ-axis; the other edge is assumed not to lie on the diagonal $\xi+\eta=0$. We say then that Γ_α is in *special position*; the corresponding cone operators satisfy the hypotheses of (1.8)(ii). If f and g are now replaced by their wave packet expansions the geometry of such Γ_α eliminates all wave packets except those having vanishing moment and frequency in a fixed half-line. In such cases we expect $\mathcal{C}_{\Gamma_\alpha}$ to map $L^p(\mathbb{R})\times L^{p'}(\mathbb{R})$ into *complex* Hardy space $H^1_{\mathbb{C}}(\mathbb{R})$. Corresponding results for Γ_α in general position as well as for $\alpha>\pi/4$ would then follow via symmetry and simple examples such as (1.4), establishing the conjecture in full knowing (1.8)(ii) only for $\mathcal{C}_{\Gamma_\alpha}$ with Γ_α in special position. As illustration, note that the previous $A_{\theta,\alpha}$ satisfy the cancellation condition (1.2), and so by (1.8)(i) should map $L^p(\mathbb{R})\times L^{p'}(\mathbb{R})$ into $H^1(\mathbb{R})$. But this also would follow from (1.8)(ii) because one of the cone operators in the decomposition of $A_{\theta,\alpha}$ would then map into $H^1_{\mathbb{C}}(\mathbb{R})$, while, by symmetry, the second maps into $\overline{H^1_{\mathbb{C}}(\mathbb{R})}$. Boundedness for all the B_θ, $\theta\ne\pi/4$, too would follow from (1.8)(ii).

In this paper we shall establish strong supporting evidence for conjecture (1.8)(ii) on which the full conjecture rests. When the functions f and g in (1.9) are replaced by smooth wave packet expansions then $\mathcal{C}_{\Gamma_\alpha}(f,g)$ becomes an ℓ^1-sum of *discrete* bilinear operators which can be thought of as infinite sums of *modulated para-products*. When Γ_α is in special position its geometry ensures that the wave packets occurring in these infinite sums have frequency in one direction and so they belong to $H^1_{\mathbb{C}}$. The problem thus becomes one of establishing the $L^p(\mathbb{R})\times L^{p'}(\mathbb{R})\to H^1_{\mathbb{C}}(\mathbb{R})$-boundedness of such discrete bilinear operators. The time-frequency analysis of them, however, is complicated by the fact that smoothness of the wave-packets forces overlap, thereby eliminating sharp 'cut-offs'. By contrast, Walsh wavelet packets are not smooth, but the time-frequency analysis associated with them is simpler. Conjecture (1.8) can thus be tested on the discrete bilinear operator obtained by substituting Walsh wavelet packets for smooth wave packets in the infinite sums of modulated para-products.

Let $\mathrm{w}_{j,\ell,n}=2^{-j/2}W_n(2^{-j}x-\ell)$ be the usual Walsh wavelet packets derived from the Walsh functions W_n. The Walsh model for the Bilinear Hilbert Transform is the family of bilinear operators

$$H_w: f,g\longrightarrow\sum_{j,\ell,n}2^{-j/2}\big\langle f,\mathrm{w}_{j,\ell,4n+\epsilon_1}\big\rangle\,\big\langle g,\mathrm{w}_{j,\ell,4n+\epsilon_2}\big\rangle\,\mathrm{w}_{j,\ell,4n+\epsilon_3}\tag{1.10}$$

associated with all fixed triples $(\epsilon_1, \epsilon_2, \epsilon_3)$ having integer entries ϵ_k such that

$$0 \le \epsilon_k \le 3, \qquad j \ne k \Longrightarrow \epsilon_j \ne \epsilon_k. \tag{1.11}$$

Summation in (1.10) is taken over all $j, \ell \in \mathbb{Z}$ and all $n \ge 0$ but the same triple is used throughout the sum and the notation H_w will be used irrespective of the choice of triple. Estimates for Walsh functions show that the series converges unconditionally in $L^r(\mathbb{R})$ for all $g \in L^q(\mathbb{R})$, $q < r$, so long as f is a step function, but convergence in general is much more delicate. This family was first introduced by Thiele in his thesis - the so-called 'Quartile Operator' ([22]).

A natural structure has been built into the definition. By interchanging ϵ_1 and ϵ_2 we see that the Walsh model family is symmetric in the sense that f, g can be reversed in (1.10), while interchanging ϵ_3 with either of ϵ_1, ϵ_2 shows that the family is also self-adjoint. On the other hand, when dilation is defined by $\delta_r : f(x) \to r^{-1/2} f(r^{-1}x)$, it is clear that each individual operator commutes with simultaneous dyadic dilation since

$$H_w(\delta_{2^k} f, \delta_{2^k} g) = 2^{-k/2} \delta_{2^k}(H_w(f, g)), \qquad (k \in \mathbb{Z}). \tag{1.12}$$

Thus any $L^p \times L^q \to L^r$ boundedness result for the Walsh model family must again follow the same pattern as in Hölder's inequality, *i.e.*, $1/p + 1/q = 1/r$. The symmetry and self-adjoint structure together with a bilinear Marcinkiewicz Interpolation theorem will enable us to establish such strong type results from weak type estimates solely on the restricted range $1 < p < 2 \le q < \infty$.

In his thesis Thiele showed, among other results, that H_w maps $L^2(\mathbb{R}) \times L^2(\mathbb{R})$ into $L^1(\mathbb{R})$. By analogy with conjecture (1.8), therefore, the question is whether cancellation improves regularity. In the context of Walsh functions the appropriate Hardy space is dyadic H^1 and, in view of (1.3), the corresponding cancellation condition is the requirement that each 'exposed' function $\mathrm{w}_{j,\ell,4n+\epsilon_3}$ have vanishing moment. Since only the W_n, $n > 0$, have vanishing moment, the most natural family of Walsh model operators having 'cancellation' are thus the operators

$$D_w : f, g \longrightarrow \sum_{n>0} \sum_{j,\ell \in \mathbb{Z}} 2^{-j/2} \big\langle f, \mathrm{w}_{j,\ell,4n+\epsilon_1} \big\rangle \big\langle g, \mathrm{w}_{j,\ell,4n+\epsilon_2} \big\rangle \, \mathrm{w}_{j,\ell,4n+\epsilon_3} \tag{1.13}$$

obtained by restricting summation in (1.10) to $n > 0$ for each fixed choice of triple $(\epsilon_1, \epsilon_2, \epsilon_3)$ satisfying (1.11). This family will again be symmetric as well as self-adjoint, and each individual operator will be simultaneously invariant under dyadic dilation. The principal result proved in this paper is the following analogue of the smooth symbol result.

Main Theorem. *Every bilinear operator D_w is bounded from $L^2(\mathbb{R}) \times L^2(\mathbb{R})$ into the dyadic Hardy space $H^1_d(\mathbb{R})$.*

The difference between H_w and D_w is an operator

$$(H_w - D_w) : f, g \longrightarrow \sum_{j,\ell \in \mathbb{Z}} 2^{-j/2} \big\langle f, \mathrm{w}_{j,\ell,\epsilon_1} \big\rangle \big\langle g, \mathrm{w}_{j,\ell,\epsilon_2} \big\rangle \, \mathrm{w}_{j,\ell,\epsilon_3} \tag{1.5}$$

which will still increase regularity if the 'exposed' wavelet packet $\mathrm{w}_{j,\ell,\epsilon_3}$ has vanishing moment, *i.e.*, if $\epsilon_3 \ne 0$. For then one of the adjoints of $(H_w - D_w)$ will be a para-product since at least one of $\mathrm{w}_{j,\ell,\epsilon_1}, \mathrm{w}_{j,\ell,\epsilon_2}$ also will have vanishing moment for all ℓ, j (*cf.* [18] page 375). In such a case H_w itself will increase regularity. More precisely,

Corollary. *A Walsh model operator H_w associated with a triple $(\epsilon_1, \epsilon_2, \epsilon_3)$ is bounded from $L^2(\mathbb{R}) \times L^2(\mathbb{R})$ into the dyadic Hardy space $H^1_d(\mathbb{R})$ whenever $\epsilon_3 \neq 0$.*

The proof of the Main theorem uses characterizations of $H^1_d(\mathbb{R})$ by 4-adic martingale conjugate transforms just as the Hilbert transform characterizes $H^1(\mathbb{R})$. The basic idea is to replace the phase plane constructed with dyadic intervals by one constructed with 4-adic intervals and to replace a Walsh wavelet packet by its 4-tuple of conjugate transforms analogous to a wave packet being in complex H^1. In this way we obtain an $\mathbb{R}^4$-valued version $\mathcal{D}$ of D_w. We regard both D_w and $\mathcal{D}$ as Walsh models for cone operators - the former for the case of a double cone, the latter for one defined in (1.9) with Γ_α in special position. In complete analogy with the discussion of (1.8)(ii) for complex H^1-spaces we prove an $\mathbb{R}^4$-valued version of Thiele's result by showing that $\mathcal{D}$ maps $L^2(\mathbb{R}, \mathbb{R}^4) \times L^2(\mathbb{R}, \mathbb{R}^4)$ into $L^1(\mathbb{R}, \mathbb{R}^4)$. With this the proof of the Main Theorem is complete. One might expect to use a maximal function or square function argument to establish H^1-boundedness but this seems to add a non-linear veneer to an already involved proof. The non-linearity could perhaps be avoided in the case of square function arguments by using vector-valued ideas, but even here the use of integral transform characterizations would appear to be simpler since they reduce the proof to an L^1-result for $\mathbb{R}^4$-valued functions.

Our proof actually establishes a stronger result, namely the boundedness of

$$D_w : L^p_d(\mathbb{R}) \times L^q_d(\mathbb{R}) \longrightarrow H^r_d(\mathbb{R}) \qquad \left(\frac{1}{p} + \frac{1}{q} = \frac{1}{r}\right)$$

for all $1 < p,\, q < \infty$ so long as $r > r_0$ where r_0, $r_0 < 1$, is an as yet undetermined critical index. The appearance of this critical index can be traced to the conjugate transform characterization of $H^r_d(\mathbb{R})$, and not to the intrinsic structure of D_w.

This paper is organized as follows. In section 2 we introduce some basic facts about the Walsh wave-packets and martingales on $\mathbb{R}$. Section 3 contains the Chao-Janson characterization of $H^1_d(\mathbb{R})$ by 4-adic martingale conjugate transforms. In section 4 we introduce the scalar and vector-valued 4-adic Walsh models, study their relationship and explain how the strong-type results for the vector-valued 4-adic Walsh model yield the same results for the scalar-valued one; the Riesz transforms of the Walsh wavelet packets will be determined also. In section 5 we study the geometry of the 4-adic (or *even-scaled*) phase plane. Section 6 contains a formulation of Janson's real bilinear interpolation theorem. We also explain there how this, together with symmetry and duality, is exploited to obtain the strong-type results $L^p \times L^q \to L^r$ with $1 < p, q < \infty$ and $1/p + 1/q = 1/r < 3/2$ solely from the weak-type estimates $L^p \times L^q \to L^{r\infty}$ with $1 < p < 2 \leq q < \infty$ and $1/p + 1/q = 1/r$. Finally in section 7 we establish these weak type estimates for the vector-valued 4-adic Walsh model. The proof of these estimates is analogous to that given by Thiele in his thesis for the quartile operator H_w, both of which in turn were inspired by Fefferman's proof of Carleson's theorem ([13]).

Finally, Thiele's results together with our proof suggests that the *full result* should be that the bilinear operator D_w is bounded as an operator

$$D_w : H^p_d(\mathbb{R}) \times H^q_d(\mathbb{R}) \longrightarrow H^r_d(\mathbb{R})$$

for all p and q such that $0 < 1/p + 1/q = 1/r < 3/2$. By duality there will also be results for dyadic BMO and Lipschitz spaces. Using the symmetry and self-adjoint property of the

D_w family our proof establishes boundedness for all $1 < p, q < \infty$ and $r > \max\{\, r_0,\, 2/3\,\}$, knowing simply the result for $1 < p < 2 \leq q < \infty$. Computing r_0 is an algebraic problem; we omit any discussion of it here. On the other hand, extension of the possible values of p and q would come about if the Carleson measure estimate (7.8) can be established for all $H^p_d(\mathbb{R})$, $2/3 < p \leq 2$, rather than for $L^p(\mathbb{R})$, $1 < p \leq 2$. Again we omit any discussion of it here.

The authors wish to thank R. Coifman, M. Cwikel, J. Lakey, and M. Taibleson for enlightening conversations or correspondence during work on this paper.

2. Wavelet packets, martingales.

The Walsh functions $W_0, W_1, W_2, \ldots$ are defined recursively on $[0, 1)$ by

$$W_0 \equiv 1, \qquad W_{2n+1}(x) = \begin{cases} W_n(2x), & 0 \leq x < \frac{1}{2}, \\ -W_n(2x-1), & \frac{1}{2} \leq x < 1, \end{cases} \tag{2.1)(i}$$

and

$$W_{2n}(x) = \begin{cases} W_n(2x), & 0 \leq x < \frac{1}{2}, \\ W_n(2x-1), & \frac{1}{2} \leq x < 1. \end{cases} \tag{2.1)(ii}$$

The W_m will always be extended by zero outside the interval $[0, 1)$. An alternative definition in terms of the Rademacher functions

$$r_0(x) = \operatorname{sgn}\, \sin 2\pi x, \qquad r_n(x) = r_0(2^n x) \quad (n \geq 1) \tag{2.2}$$

and the dyadic representation of integers will be useful later: if $m = 2^{s_1} + 2^{s_2} + \cdots + 2^{s_k}$ with $s_1 > s_2 > \cdots > s_k$, then

$$W_m(x) = \begin{cases} \chi_{[0,1)}(x), & m = 0 \\ r_{s_1}(x)\, r_{s_2}(x) \ldots r_{s_k}(x)\, \chi_{[0,1)}(x), & m \geq 1. \end{cases} \tag{2.4}$$

where $\chi_{[0,1)}$ is the characteristic function of the interval $[0, 1)$. This last representation makes clear the fact that $|W_m(x)| = 1$ everywhere on its support.

There are deep connections with analysis on both the unit interval and the whole real line. It is well-known, for instance, that $\{W_m\}_{m \geq 0}$ is an orthonormal basis of $L^2[0, 1)$ with respect to the natural inner product on $L^2[0, 1)$; the *Walsh series* of a function f in $L^2[0, 1)$ is then defined by

$$W[f] = \sum_{m \geq 0} \langle f, W_m \rangle\, W_m.$$

Relations with martingales can be made through Walsh series: the partial sums

$$f_n(x) = W_{2^n}[f] = \sum_{0 \leq m \leq 2^n} \langle f, W_m \rangle\, W_m(x)$$

of order 2^n have the structure of a dyadic martingale; in fact, the representation by martingale differences

$$f \sim \sum_n \Delta_n f = \sum_n \Big(\sum_{2^n \le m \le 2^{n+1}} \langle f, W_m \rangle \, W_m \Big) \tag{2.5}$$

is the earliest and simplest form of a Littlewood-Paley decomposition ([14]).

Connections with analysis on the real line can be made through the theory of wavelets and wavelet packets. As W_1 is just the Haar mother wavelet, its dyadic translates and dilates $2^{-j/2} W_1(2^{-j}x - k)$ provide an orthonormal wavelet basis for $L^2(\mathbb{R})$. While this achieves excellent 'time' localization, an improvement in 'frequency' localization can be obtained by allowing translates and dilates of all Walsh functions ([10],[15]). The Walsh *wavelet packets* are thus defined by

$$\mathrm{w}_{j,\ell,m}(x) = 2^{-j/2} W_m(2^{-j}x - \ell) = (\delta_{2^k} \circ \tau_\ell) W_m \tag{2.6}$$

with respect to dyadic dilation and integer translation $\tau(\ell) : f(x) \to f(x - \ell)$. For each fixed resolution j the family$\{\mathrm{w}_{j,\ell,m} : \ell \in \mathbb{Z},\ m \ge 0\}$ is an orthonormal basis for $L^2(\mathbb{R})$ with respect to the inner product

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x) \overline{g(x)} \, dx.$$

Other orthonormal bases can be constructed using the covering criterion of Coifman-Meyer *et al.* ([11] page 466). Walsh wavelet packets are related to dyadic martingales on the real line in the same way that Walsh functions are related to dyadic martingales on $[0, 1)$. Let

$$\mathcal{E}_k : f \longrightarrow \sum_{I \in \mathcal{B}_k} \frac{1}{|I|} \langle f, \chi_I \rangle \, \chi_I(x)$$

be the conditional expectation with respect to the σ-field $\mathcal{B}_k$ generated by dyadic intervals of length 2^{-k} (*cf.* [21] page 188). The $\mathcal{E}_k$ commute with dyadic dilation and with integer translation in the sense that

$$\begin{aligned} \mathcal{E}_k(\delta_{2^j}(f)) &= \delta_{2^j}(\mathcal{E}_{j+k} f) \qquad (j, k \in \mathbb{Z}), \\ \mathcal{E}_k(\tau_\ell(f)) &= \tau_\ell(\mathcal{E}_k f) \qquad (k \ge 0,\ \ell \in \mathbb{Z}). \end{aligned} \tag{2.7}$$

Dyadic Hardy spaces can be introduced using the $\mathcal{E}_k$.

(2.8) Definition. *The dyadic Hardy space $H^p_d(\mathbb{R})$ consists of all dyadic martingales $f = \{f_k\}$ whose dyadic maximal function $f^*(x) = \sup_k |f_k(x)|$ is L^p-integrable.*

The norm on $H^p_d(\mathbb{R})$ is defined by

$$\|f\|_{H^p_d} = \Big(\int_{-\infty}^{\infty} |f^*(x)|^p \, dx \Big)^{1/p}.$$

Meyer has developed an alternative approach to dyadic Hardy spaces using the Haar wavelet $\mathrm{w}_{j,\ell,1}$ but the Walsh wavelet packets are much too overcomplete to provide the same martingale basis as Haar wavelets. Nonetheless, the $\mathrm{w}_{j,\ell,n}$ play an analogous role though we have to distinguish carefully between the case $m = 0$ and $m > 0$ for precisely the reasons that led to the distinction between H_w and D_w. So suppose first that $m > 0$. Then in view of (2.5) and (2.7),

$$\mathcal{E}_k(\mathrm{w}_{j,\ell,m}) = \begin{cases} \mathrm{w}_{j,\ell,m}, & k+j \geq s+1 \\ 0, & k+j < s+1, \end{cases}$$

whenever $2^s \leq m < 2^{s+1}$; consequently,

$$(\mathcal{E}_{k+1} - \mathcal{E}_k)\mathrm{w}_{j,\ell,m} = \begin{cases} \mathrm{w}_{j,\ell,m}, & s = k+j, \\ 0, & s \neq k+j \end{cases} \tag{2.9}$$

for the same range of m. In particular,

$$\mathrm{w}^*_{j,\ell,m}(x) = 2^{-j/2}\chi_{[0,1)}(2^{-j}x - \ell) \quad (m > 0),$$

so $\mathrm{w}_{j,\ell,m}$ belongs to $H^p_d(\mathbb{R})$, $0 < p < \infty$, for all $m > 0$. On the other hand,

$$W^*_0(x) \approx \frac{1}{1+x}\,\chi_{[0,\infty)}(x)$$

so none of the translates and dilates $\mathrm{w}_{j,\ell,0}$ of W_0 belong to $H^p_d(\mathbb{R})$, $p \leq 1$.

3. Characterizations of dyadic H^p.

Characterizations of dyadic Hardy spaces introduced by J.-A. Chao in the finite measure space case will be central to the proof of the Main theorem ([3], [4]). They depend on the fact that there is a one-to-one correspondence between dyadic martingales and 4-adic martingales on $\mathbb{R}$. Indeed, if $f = \{f_k\}_k$ is a dyadic martingale on $\mathbb{R}$ relative to the σ-fields $\mathcal{B}_k$, then $\phi = \{f_{2k}\}_k$ is a 4-adic martingale relative to the $\mathcal{B}_{2k}$; conversely, if $\phi = \{\phi_k\}_k$ is a 4-adic martingale relative to the $\mathcal{B}_{2k}$, then there is a dyadic martingale $f = \{f_k\}_k$ such that $f_{2k} = \phi_k$. Furthermore, a dyadic martingale $f = \{f_k\}_k$belongs to $H^p_d(\mathbb{R})$ if and only if its 4-adic maximal function

$$Mf(x) = \sup_k |f_{2k}(x)|, \tag{3.1}$$

is L^p-integrable. Consequently,

$$\text{const.}\,\|f\|_{H^p} \;\leq\; \left(\int_{-\infty}^{\infty} Mf(x)^p\,dx\right)^{1/p} \;\leq\; \text{const.}\,\|f\|_{H^p}. \tag{3.2}$$

Chao's proof of this result for a regular martingale on a finite measure space carries over to $\mathbb{R}$ because the σ-fields $\mathcal{B}_k$ also are regular ([2]). For consistency of notation throughout this section, $f = \{f_k\}$ will always denote a dyadic martingale, while $\phi = \{\phi_k\}, \psi = \{\psi_k\}$ will denote 4-adic martingales.

The freedom to apply 4-adic martingale theory allowed Chao to exploit characterizations of martingale H^1 by 'conjugate transforms' obtained earlier by Janson in a setting where initially Janson's results did not apply ([16]). It is this same freedom that prompts the introduction of the 4-adic Walsh model operator D in the next section. Again because of the regularity of the $\mathcal{B}_k$ all constructions and results for regular martingales on a finite measure spaces go over immediately to $\mathbb{R}$. Let $\phi = \{\phi_k\}_k$ be 4-adic martingale on $\mathbb{R}$ relative to the $\mathcal{B}_{2k}$ and let A be a real 4×4 matrix which maps the subspace

$$\mathcal{V} = \{v = (v_0, v_1, v_2, v_3) \in \mathbb{R}^4 : v_0 + v_1 + v_2 + v_3 = 0\,\}$$

of $\mathbb{R}^4$ into itself under matrix multiplication $A : v \to vA$; in other words, the matrix $A = [A_{ij}]$ has the property

$$v_0 + v_1 + v_2 + v_3 = 0 \implies \sum_j \Big(\sum_i v_i A_{ij} \Big) = 0. \tag{3.3}$$

This will always be the case, for instance, when the rows of A have average zero, *i.e.*, when $\sum_{j=0}^{3} A_{ij} = 0$. Now let $\phi = \sum_k \Delta_k \phi$ be the representation of ϕ in terms of its martingale differences $\Delta_k \phi = \phi_{k+1} - \phi_k$. To each 4-adic interval I, $|I| = 4^{-k}$, there corresponds a vector $v \in \mathcal{V}$ such that

$$(\Delta_k \phi)\big|_I = v_0 \chi_{I_0} + v_1 \chi_{I_1} + v_2 \chi_{I_2} + v_3 \chi_{I_3} = \sum_{j=0}^{3} \langle\, v, e_j \rangle\, \chi_{I_j} \qquad (v \in \mathcal{V})$$

where $I = \cup_j I_j$ is the partition of I into its 4 equal subintervals I_j, $|I_j| = 4^{-(k+1)}$. But then

$$\psi_k\big|_I = \sum_{j=0}^{3} \langle\, A(v), e_j \rangle\, \chi_{I_j} \qquad (I \in \mathcal{B}_{2k})$$

define martingale differences of a 4-adic martingale $\psi = \{\psi_k\}_k$ since $A : \mathcal{V} \to \mathcal{V}$. Thus, $\phi \to \psi = \sum_k \psi_k$ associates to each matrix A satisfying (3.3) a singular integral transform $\phi \to \psi = \mathcal{A}(\phi)$ on 4-adic martingales; we shall say that A is the *symbol* of $\mathcal{A}$. By construction each $\mathcal{A}$ commutes with 4-adic dilation in the sense that

$$(\delta_{4^k})(\mathcal{A} f) = \mathcal{A}(\delta_{4^k} f), \qquad (k \in \mathbb{Z}) \tag{3.4)(i}$$

while it commutes with pointwise multiplication at the martingale difference level meaning that

$$\mathcal{A}(\nu \Delta_k \phi) = \nu \mathcal{A}(\Delta_k \phi), \qquad (k \in \mathbb{Z}) \tag{3.4)(ii}$$

whenever ν is $\mathcal{B}_{2k}$-measurable. Although the $\mathcal{A}$ do not commute with integer translation, they do have the property:

$$\mathcal{A}(\tau_\ell(\Delta_k \phi)) = \tau_\ell(\mathcal{A}(\Delta_k \phi)), \qquad (k \geq 0, \ell \in \mathbb{Z}). \tag{3.4)(iii}$$

When the matrix A is diagonal such transforms reduce to a special case of the martingale transforms introduced by Burkholder (*cf.* [20] pps. 95–103).

To estimate the norm of $\mathcal{A}(f)$ we use square function ideas. On a 4-adic interval I,

$$|\psi_k(x)|^2 = \sum_{j=0}^{3} |\langle A(v), e_j\rangle|^2 \chi_{I_j}(x) \le \sup_j \Big(\sum_i |A_{ij}|^2\Big) \|v\|^2 \chi_I(x),$$

so

$$S(\mathcal{A}\phi)(x) = \Big(\sum_k |(\Delta_k \phi)(x)|^2\Big)^{1/2} \le \text{const.}\Big(\sum_k \mathcal{E}_{2k-2}(|\Delta_k\phi|^2)(x)\Big)^{1/2}.$$

But for regular martingales this last square function, like the first square function, is equivalent to the 4-adic maximal function in L^p-norm. Interpreting this for dyadic martingales, we deduce that the inequality

$$\int_{-\infty}^{\infty} |\mathcal{A}(f)(x)|^p \, dx \ \le \ \text{const.} \int_{-\infty}^{\infty} |f^*(x)|^p \, dx$$

holds for all p, $0 < p < \infty$, proving the next result.

(3.5) Theorem. *Let A be a real 4×4 matrix satisfying* (3.3)*. Then the associated integral transform $\mathcal{A}$ having A as symbol is bounded on $H^p_d(\mathbb{R})$, $0 < p < \infty$.*

The adjoint operator $\mathcal{A}^*$ of $\mathcal{A}$ also can be described in terms of A when, say, both the rows and columns of A have average zero. For then the transpose matrix $A^t = [A_{ji}]$ will be the symbol of an integral transform which is bounded on all H^p_d-spaces and which on $L^2(\mathbb{R})$ is easily seen to be the adjoint of $\mathcal{A}$ on $L^2(\mathbb{R})$. What we need is a vectorial version. Let $A^{(m)} = [A^{(m)}_{ij}], 1 \le m \le n$, be a family of 4×4 real matrices each of which satisfies (3.3), and let $A = [A_{ij}]$ be the 4×4 matrix whose entries

$$A_{ij} = (A^{(1)}_{ij}, A^{(2)}_{ij}, \ldots, A^{(n)}_{ij}) \tag{3.6}$$

belong to $\mathbb{R}^n$. Then the integral transform $\mathcal{A}$ having A as symbol is again well-defined and the same proof as before shows that $\mathcal{A}$ is bounded from $H^p_d(\mathbb{R})$ into $L^p(\mathbb{R}, \mathbb{R}^n)$, $0 < p < \infty$, since all the necessary martingale results remain valid for Hilbert space-valued martingales. If, in addition, every column in each $A^{(m)}$ has average zero, then again the adjoint of $\mathcal{A}$ is the integral transform whose symbol is the transpose of A.

Janson and Chao introduced simple algebraic conditions on the matrices $A^{(m)}$ to provide striking characterizations of $H^1_d(\mathbb{R})$ by the corresponding integral transforms in the same sense that a function in $L^1(\mathbb{R}^n)$ belongs to $H^1(\mathbb{R}^n)$ if and only if its Riesz transforms are $L^1(\mathbb{R}^n)$-integrable ([3], [5], [6], [16]).

(3.7) Theorem (Janson). *Let $A^{(m)}$, $1 \le m \le n$, be real 4×4 matrices having no common eigenvector in $\mathcal{V}$ and let $A^{(0)}$ be the 4×4 identity matrix. Then there exists r_0, $r_0 < 1$, such that the inequality*

$$\|a\|^p \ \le \ \frac{1}{4} \sum_{j=0}^{3} \Big(\sum_{m=0}^{n} |a_m + \langle\, A^{(m)}(v),\, e_j \,\rangle|^p \Big)$$

holds for each $p > r_0$ *uniformly in* $a = (a_0, \ldots, a_n) \in \mathbb{R}^{n+1}$ *and* $v \in \mathcal{V}$.

The algebraic condition on the $A^{(m)}$ ensures that $\mathcal{A}\phi$ satisfies a sub-martingale inequality. Let $\mathcal{A}_m$ be the integral transform having $A^{(m)}$ as its symbol and let $\Psi = \{\Psi_k\}$ be the $\mathbb{R}^{n+1}$-valued 4-adic martingale $\Psi = (\phi, \mathcal{A}_1\phi, \ldots, \mathcal{A}_n\phi)$ so that

$$\Psi_{k+1} = \Psi_k + \Delta_k\Psi = \Psi_k + \left(\sum_{j=0}^{3} \langle\, A^{(0)}(v),\, e_j\,\rangle \chi_{I_j}, \ldots, \sum_{j=0}^{3} \langle\, A^{(n)}(v),\, e_j\,\rangle \chi_{I_j}\right).$$

Then with $a = \Psi_k(x)$ in (3.7) we see that $\{\Psi_k\}$ satisfies the submartingale inequality

$$\|\Psi_k(x)\|^p \le \mathcal{E}_{2k}(\|\Psi_{k+1}\|^p)(x)$$

provided $p > r_0$. Hence by the martingale maximal theorem,

$$\begin{aligned}\int_{-\infty}^{\infty} \sup_k |\phi_k(x)|\, dx &\le \text{const.} \int_{-\infty}^{\infty} (\sup_k \|\Psi_k(x)\|^p)^{1/p}\, dx \\ &\le \text{const.} \sup_k \int_{-\infty}^{\infty} \|\Psi_{k+1}(x)\|\, dx\end{aligned}$$

for each p, $r_0 < p < 1$. Consequently, the inequality

$$\int_{-\infty}^{\infty} \sup_k |\phi_k(x)|\, dx \;\le\; \text{const.}\Big\{ \|\phi\|_{L^1} + \|\mathcal{A}_1(\phi)\|_{L^1} + \ldots \|\mathcal{A}_n(\phi)\|_{L^1} \Big\}$$

will hold whenever the algebraic condition of (3.7) is satisfied. Since the reverse inequality is always valid, this establishes the characterization of $H^1_d(\mathbb{R})$ on which the Main Theorem is based.

(3.8) Theorem (Chao-Janson). *Let* $A^{(m)}$, $1 \le m \le n$, *be real* 4×4 *matrices having no common eigenvector in* $\mathcal{V}$. *Then a function* f *in* $L^1(\mathbb{R})$ *belongs to* $H^1_d(\mathbb{R})$ *if and only if* $\mathcal{A}_1 f, \ldots, \mathcal{A}_n f$ *are* L^1*-integrable.*

The same proof shows also that a dyadic martingale $f = \{f_k\}$ will belong to $H^p_d(\mathbb{R})$, $p > r_0$, if and only if f and all the $\mathcal{A}_m f$, $1 \le m \le n$, are L^p-integrable.

We shall apply the Chao-Janson characterization using the matrices

(3.9)(i) $$R_1 = \frac{1}{2}\begin{bmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix}, \qquad R_2 = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

and

(3.9)(ii) $$R_3 = \frac{1}{2}\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

The associated integral transforms $\mathcal{R}_1, \mathcal{R}_2$ and $\mathcal{R}_3$ are bounded on $H^p_d(\mathbb{R})$ as are their adjoints. In particular,

$$f \longrightarrow F = \frac{1}{\sqrt{3}}(f, \mathcal{R}_1 f, \mathcal{R}_2 f, \mathcal{R}_3 f) \tag{3.10}$$

will be bounded from $H^p_d(\mathbb{R})$ into $L^p(\mathbb{R}, \mathbb{R}^4)$, $0 < p < \infty$. The factor $1/\sqrt{3}$ is a normalizing constant introduced to ensure that $\mathcal{R} : L^2(\mathbb{R}) \to L^2(\mathbb{R}, \mathbb{R}^4)$ is a partial isometry, *i.e.*,

$$\int_{-\infty}^{\infty} \|\mathcal{R}(f)(x)\|^2\,dx = \int_{-\infty}^{\infty} |f(x)|^2\,dx \qquad (f \in L^2(\mathbb{R})). \tag{3.11}$$

Indeed, the symbol of $\mathcal{R}^*\mathcal{R}$ is the matrix

$$\tfrac{1}{3}(I + R_1^t R_1 + R_2^t R_2 + R_3^t R_3) = I - \tfrac{1}{6}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

and this last matrix satisfies (3.3) because it annihilates all v in $\mathcal{V}$, so the corresponding singular integral transform annihilates all martingale differences. Consequently, $\mathcal{R}^*\mathcal{R} = I$, establishing (3.11).

It is easy to see that these R_i satisfy the algebraic condition in (3.8).

(3.12) Lemma. *The matrices R_1, R_2 and R_3 in* (3.9) *have no non-zero common eigenvector in $\mathcal{V}$.*

Proof. It is obviously enough to show that R_1 and R_2 have no non-zero common eigenvector in $\mathcal{V}$. Now the kernel of R_1 is generated by the vectors $[1, 1, 0, 0]$ and $[0, 0, 1, 1]$, so the only vectors in $\mathcal{V}$ that belong to this kernel are scalar multiples of $[1, 1, -1, -1]$. As R_2 maps $[1, 1, -1, -1]$ into the vector $[1, -1, 1, -1]$, however, no non-zero vector in the kernel of R_1 can be an eigenvector for R_2.

On the other hand, $R_2 + iR_1 = \overline{(R_2 - iR_1)}$. Consequently, if R_1 and R_2 have a non-zero common eigenvector in $\mathcal{V}$, say $R_1 v = \lambda v$ and $R_2 v = \mu v$, then v must also be an eigenvector for $\overline{(R_2 - iR_1)}$. But iR_1 and R_2 are both self-adjoint so all their eigenvalues are real. Thus $2\lambda v = 0$, which means that either $v = 0$ or else v is in the kernel of R_1, neither of which can be true if v is a non-zero eigenvector for R_2. $\square$

In view of (3.8) and (3.11), therefore, the $\mathcal{R}_i$ have properties analogous to classical Riesz transforms, so it seems reasonable to call them *4-adic Riesz transforms*.

4. The 4-adic Walsh model.

To exploit the characterizations of $H^1_d(\mathbb{R})$ given in the previous section the D_w operators have to be replaced by the 4-adic version and thence by a vector-valued version incorporating all the 4-adic Riesz transforms. Set

$$\omega_{j,\ell,n}(x) = 2^{-j} W_n(4^{-j}x - \ell) = (\delta_{4^j} \circ \tau_\ell) W_n. \tag{4.1}$$

Replacing the $\mathrm{w}_{j,\ell,n}$ in (1.13) with the $\omega_{j,\ell,n}$ we obtain the 4-adic version

$$D : f, g \longrightarrow \sum_{n>0} \sum_{j,\ell\in\mathbb{Z}} 2^{-j} \big\langle f, \omega_{j,\ell,4n+\epsilon_1} \big\rangle \, \big\langle g, \omega_{j,\ell,4n+\epsilon_2} \big\rangle \, \omega_{j,\ell,4n+\epsilon_3} \tag{4.2}$$

of D_w. Again the family of all such D will be symmetric and self-adjoint. There is very close relationship between D_w and D. By separating the sum over the resolution j into even and odd j we see that

$$D_w(f,g) = D(f,g) + \sqrt{2}(\delta_2 \circ D)(\delta_{1/2}(f), \delta_{1/2}(g)). \tag{4.3}$$

Replacing each wavelet packet $\omega_{j,\ell,m}$ in D by the $\mathbb{R}^4$-valued function $\mathcal{R}(\omega_{j,\ell,m})$ we finally obtain the Walsh model operator

$$\mathcal{D}(F,G) = \sum_{n>0} \sum_{j,\ell\in\mathbb{Z}} 2^{-j} \big\langle F, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_1}) \big\rangle \, \big\langle G, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_2}) \big\rangle \, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_3}) \tag{4.4}$$

defined on $\mathbb{R}^4$-valued functions F and G. Once again the family of all such $\mathcal{D}$ will be symmetric and self-adjoint. This family is the one we shall study in later sections.

Unconditional convergence of each of the series above in L^r-norm, at least for simple functions is easily seen.

(4.5) Theorem. *Let I be a dyadic interval in $\mathbb{R}$ and $\chi_{_I}$ its characteristic function. Then the inequalities*

$$\|D_w(\chi_{_I}, g)\|_{L^r} \le \sum_{n>0} \sum_{j,\ell\in\mathbb{Z}} 2^{-j/2} |\langle \chi_{_I}, \mathrm{w}_{j,\ell,4n+\epsilon_1}\rangle \, \langle g, \mathrm{w}_{j,\ell,4n+\epsilon_2}\rangle| \, \|\mathrm{w}_{j,\ell,4n+\epsilon_3}\|_{L^r}$$

$$\le \text{const.}\, |I|^{1/q'-1/r} \, \|g\|_{L^q} \qquad \left(\frac{1}{q} + \frac{1}{q'} = 1\right)$$

hold uniformly for all g in $L^q(\mathbb{R})$ and each $r > q\,,\ q > 1$.

Proof. Suppose $|I| = 2^{-k}$. Then by dilation invariance

$$\chi_{_I} = 2^{k/2}\, \delta_{2^k}(\chi_{_J}), \qquad D(\chi_{_I}, g) = (\delta_{2^{-k}} \circ D)(\chi_{_J}, \delta_{2^k} g)$$

for some dyadic interval J with $|J| = 1$. On the other hand,

$$\|D(\chi_{_J}, g)\|_{L^r} \le \sum_{j,\ell,n} 2^{-j/2} \, |\langle \, \mathrm{w}_{j,\ell,4n+\epsilon_1},\ \delta_{2^k} g\rangle| \, |\langle \chi_{_J},\ \mathrm{w}_{j,\ell,4n+\epsilon_2}\rangle| \, \|\mathrm{w}_{j,\ell,4n+\epsilon_3}\|_{L^r}.$$

We estimate each of the terms in this last infinite sum. Fix m with $2^s \le m < 2^{s+1}$. In view of (2.11),

$$|\langle \chi_{_J}, \mathrm{w}_{j,\ell,m}\rangle| = \begin{cases} 0, & j \le s, \ell \in \mathbb{Z}, \\ 0, & j > s, \ell \ne 0, \\ 2^{-j/2}, & j > s, \ell = 0, \end{cases}$$

while

$$\left|\left\langle\, \mathrm{w}_{j,\ell,m},\, \delta_{2^k} g\right\rangle\right| \le \|\delta_{2^k} g\|_{L^q}\, \|\mathrm{w}_{j,\ell,m}\|_{L^{q'}} = 2^{(j-k)(1/2-1/q)}\, \|g\|_{L^q}$$

by Hölder's inequality since

$$\|\mathrm{w}_{j,\ell,m}\|_{L^t} = 2^{j(1/t-1/2)} \qquad (0 < t < \infty).$$

Hence,

$$\|D(\chi_{_J},g)\|_{L^r} \le \text{const. } 2^{-k(1/2-1/q)}\, \|g\|_{L^q}$$
$$\times \sum_{s=0}^{\infty}\Big(\sum_{2^s\le m<2^{s+1}}\ \sum_{j>s} 2^{-j(1+1/q-1/r)}\Big) \le \text{const.}\, 2^{-k(1/2-1/q)}\, \|g\|_{L^q}$$

provided $r > q$. Thus

$$\|(\delta_{2^k}\circ D)(\chi_{_J}, g)\|_{L^r} \le \text{const. } 2^{-k(1-1/r-1/q)}\, \|g\|_{L^q} = \text{const. } |I|^{1/q'-1/r}\, \|g\|_{L^q},$$

completing the proof. $\square$

By applying this result to a finite number of dyadic intervals we obtain

(4.6) Corollary. *Let $f = \sum_I a_I \chi_{_I}$ be a finite linear combination of characteristic functions of dyadic intervals in $\mathbb{R}$. Then the inequalities*

$$\|D_w(f,g)\|_{L^r} \le \sum_{n>0}\sum_{j,\ell\in\mathbb{Z}} 2^{-j/2}\, |\langle f, \mathrm{w}_{j,\ell,4n+\epsilon_1}\rangle\, \langle g, \mathrm{w}_{j,\ell,4n+\epsilon_2}\rangle|\, \|\mathrm{w}_{j,\ell,4n+\epsilon_3}\|_{L^r}$$
$$\le \text{const. } \|g\|_{L^q}$$

hold for all g in $L^q(\mathbb{R})$ and each $r > q$ with constant depending on f; in particular, (1.13) *converges unconditionally in $L^r(\mathbb{R})$, $r > q$, for all $g \in L^q(\mathbb{R})$ whenever $f = \sum_I a_I \chi_{_I}$.*

Thus $D_w(f,g)$ is well-defined for all $g \in L^q(\mathbb{R})$ and all f in a dense subspace of $L^p(\mathbb{R})$. The analogous result will be true for both $D(f,g)$ and $\mathcal{D}(F,G)$. Since $\mathcal{R}(D(f,g))$ and

$$\sum_{n>0}\sum_{j,\ell\in\mathbb{Z}} 2^{-j}\left\langle f, \omega_{j,\ell,4n+\epsilon_1}\right\rangle \left\langle g, \omega_{j,\ell,4n+\epsilon_2}\right\rangle \mathcal{R}(\omega_{j,\ell,4n+\epsilon_3})$$

will also then be well-defined as functions in L^r, property (3.11) ensures that the identity

$$\mathcal{R}(D(f,g)) = (D(f,g),\, \mathcal{R}_1(D(f,g)), \ldots, \mathcal{R}_3(D(f,g))) = \mathcal{D}(\mathcal{R}f, \mathcal{R}g) \tag{4.7}$$

holds for all $g \in L^q(\mathbb{R})$ and $f = \sum_I a_I \chi_{_I}$. To prove that $D(f,g)$, and hence $D_w(f,g)$, is in $H^1_d(\mathbb{R})$, therefore, it is enough to prove

(4.8) Theorem. *The operator $\mathcal{D}$ is bounded from $L^2(\mathbb{R},\mathbb{R}^4)\times L^2(\mathbb{R},\mathbb{R}^4)$ into $L^1(\mathbb{R},\mathbb{R}^4)$ for all triples $(\epsilon_1,\epsilon_2,\epsilon_3)$ having integer entries satisfying* (1.11).

In fact more is true. We will actually prove that $\mathcal{D}$ is bounded from $L^p(\mathbb{R},\mathbb{R}^4)\times L^q(\mathbb{R},\mathbb{R}^4)$ into $L^r(\mathbb{R},\mathbb{R}^4)$ for all $p,q>1$ such that $1/p+1/q=1/r<3/2$.

The proof of (4.8) will run parallel to the corresponding scalar-valued one given by Thiele in his thesis once all the analogous 4-adic results and properties have been established. For the convenience of the reader we give most of the details, especially as those details have some novelty. To that end we first study the action of $\mathcal{R}$ on the 'even-scaled' Walsh wave packets $\omega_{j,\ell,m}$. Everything depends on the 4-adic decomposition of m. So fix m and choose σ so that $4^\sigma\le m<4^{\sigma+1}$; this requires m to lie in one of three frequency intervals

$$w_1=[4^\sigma,\ 2.4^\sigma),\qquad w_2=[2.4^\sigma,\ 3.4^\sigma),\qquad w_3=[3.4^\sigma,\ 4^{\sigma+1}) \tag{4.9}$$

each having the same length 4^σ.

(4.10) Theorem. *Let $\mathcal{R}_i(\omega_{j,\ell,m})$ be the 4-adic Riesz transforms of the Walsh wavelet packets $\omega_{j,\ell,m}$, $m\ge 1$, where the $\mathcal{R}_i$ are defined in section 3. Then*

$$\mathcal{R}_1(\omega_{j,\ell,m})=\begin{cases}0, & m\in w_1,\\ \omega_{j,\ell,m+4^\sigma}, & m\in w_2,\\ -\omega_{j,\ell,m-4^\sigma}, & m\in w_3,\end{cases}\qquad \mathcal{R}_2(\omega_{j,\ell,m})=\begin{cases}\omega_{j,\ell,m+4^\sigma}, & m\in w_1,\\ \omega_{j,\ell,m-4^\sigma}, & m\in w_2,\\ 0, & m\in w_3,\end{cases}$$

while

$$\mathcal{R}_3(\omega_{j,\ell,m})=\begin{cases}\omega_{j,\ell,m+2.4^\sigma}, & m\in w_1,\\ 0, & m\in w_2,\\ \omega_{j,\ell,m-2.4^\sigma}, & m\in w_3\end{cases}$$

where the frequency intervals w_1,w_2 and w_3 are defined in (4.9).

Proof. In view of (3.4)(i), (iii) and (2.9),

$$\mathcal{R}_i(\omega_{j,\ell,m})=(\delta_{4^j}\circ\tau_\ell)\mathcal{R}_i(W_m),$$

so it is enough to prove the theorem in the case $j=\ell=0$. But by (2.9) again

$$(\mathcal{E}_{2\sigma+2}-\mathcal{E}_{2\sigma})(W_m)=W_m.$$

On each time interval I of length $4^{-\sigma}$, therefore, the 4 values of W_m determine a vector $v\in\mathcal{V}$. To compute $\mathcal{R}_i(W_m)$ we need to find this vector; it will depend on which of the frequency intervals w_i the integer m belongs. Now

$$m=4^\sigma+m_1,\quad (m\in w_1);\qquad m=2.4^\sigma+m_2,\quad (m\in w_2);$$

$$m=2.4^\sigma+4^\sigma+m_3,\quad (m\in w_3)$$

with $0 \le m_i < 4^\sigma$ for all three i. Thus

$$W_m(x) = \begin{cases} r_0(4^\sigma x).W_{m_1}(x), & m \in w_1, \\ r_0(2.4^\sigma x).W_{m_2}(x), & m \in w_2, \\ r_0(4^\sigma x) r_1(4^\sigma x).W_{m_3}(x), & m \in w_3, \end{cases}$$

and so $\mathcal{R}(W_m) = W_{m_1}\,\mathcal{R}(r_0)$ when $m \in w_1$ since W_{m_1} will be $\mathcal{B}_{2\sigma}$-measurable; similar comments apply when $m \in w_2$ and $m \in w_3$. Consequently, there is a natural mapping

$$W_m \longrightarrow v = \begin{cases} [1, 1, -1, -1], & m \in w_1, \\ [1, -1, 1, -1], & m \in w_2, \\ [1, -1, -1, 1], & m \in w_3 \end{cases}$$

from W_m into $\mathcal{V}$. Furthermore, the particular 4-vector associated to W_m is independent of the time interval I, $|I| = 4^{-\sigma}$. But simple matrix multiplication shows that

$$[1, 1, -1, -1]R_i = \begin{cases} 0, & i = 1, \\ [1, -1, 1, -1], & i = 2, \\ [-1, 1, 1, -1], & i = 3, \end{cases}$$

$$[1, -1, 1, -1]R_i = \begin{cases} [1, -1, -1, 1], & i = 1, \\ [1, 1, -1, -1], & i = 2, \\ 0, & i = 3, \end{cases}$$

$$[1, -1, -1, 1]R_i = \begin{cases} -[1, -1, 1, -1], & i = 1, \\ 0, & i = 2, \\ [1, 1, -1, -1,], & i = 3, \end{cases}$$

Hence, by (3.4)(ii)

$$\mathcal{R}_1(W_m) = \begin{cases} 0, & m \in w_1, \\ W_{m+4^\sigma}, & m \in w_2, \\ -W_{m-4^\sigma}, & m \in w_3, \end{cases} \qquad \mathcal{R}_2(W_m) = \begin{cases} W_{m+4^\sigma}, & m \in w_1, \\ W_{m-4^\sigma}, & m \in w_2, \\ 0, & m \in w_3, \end{cases}$$

while

$$\mathcal{R}_3(W_m) = \begin{cases} W_{m+2.4^\sigma}, & m \in w_1, \\ 0, & m \in w_2, \\ W_{m-2.4^\sigma}, & m \in w_3. \end{cases}$$

This completes the proof of (4.10). $\square$

(4.11) Corollary. *Each function* $\mathcal{R}(\omega_{j,\ell,m})$ *has support on the interval* $[4^j\ell, 4^j(\ell+1))$, *and on this support* $\|\mathcal{R}(\omega_{j,\ell,m})(x)\| = 2^{-j}$. *In addition* $\|\mathcal{R}(\omega_{j,\ell,m})\|_{L^2} = 1$.

5. Geometry of even-scaled phase space.

In this section we develop the geometric properties of phase plane needed later. The details differ somewhat from those given by Thiele because 4-adic constructions are used and the frequency $n = 0$ has been excluded. Recall that an even-scaled dyadic interval or 4-adic interval is an interval of the form $[4^j\ell, 4^j(\ell+1))$, $j, \ell \in \mathbb{Z}$. The set $\mathcal{J}$ of all such intervals form a grid in the sense that

$$\text{(i)} \quad I \cap I' \in \{\emptyset, I, I'\}, \qquad \text{(ii)} \quad I \subset I',\ I \neq I' \implies 4|I| \leq |I'|.$$

for all $I, I' \in \mathcal{J}$; frequent use of this property will be made without comment. The even-scaled Walsh phase plane will be the open upper half-plane $\mathbb{R}^2_+$ where the first coordinate represents the time or spatial variable and the second coordinate represents the frequency variable. A *tile* P is a rectangle $I_P \times \omega_P$ in $\mathbb{R}^2_+$ such that $I_P,\ w_P \in \mathcal{J}$ and $|I_P|\,|w_P| = 1$. The set of all these tiles will be denoted by $\mathbb{P}$. To each tile

$$P = [4^j\ell, 4^j(\ell+1)) \times [4^{-j}m, 4^{-j}(m+1))$$

we associate the Walsh wave packet

$$\omega_P(x) = \omega_{j,\ell,m}(x) = 4^{-j/2} W_m(4^{-j}x - \ell),$$

thus providing a one-to-one correspondence $P \leftrightarrow \omega_P$ between $\mathbb{P}$ and the family of all Walsh wavelet packets introduced in (4.1). Note that $\omega_P(x)$ is supported on the time interval I_P of the P and that its absolute value is constant on this interval; corollary (4.11) says that the same is true of $\mathcal{R}(\omega_P)$.

(5.1) Remark. *Because the frequency* $n = 0$ *has been excluded, the distance from a tile* P *to the boundary* $\mathbb{R}$ *of* $\mathbb{R}^2_+$ *is at least* $1/|I_P|$.

A *quartile* Q is a 4-adic rectangle $I_Q \times w_Q$ in $\mathbb{R}^2_+$ of area 4 where both I_Q and w_Q are 4-adic intervals. The set of all these quartiles will be denoted by $\mathbb{Q}$. Each quartile $Q \in \mathbb{Q}$ consists of four *frequency sibling tiles* a_Q, b_Q, c_Q and d_Q, listed alphabetically with increasing frequency, all of which have the same time interval as Q. On the other hand, Q will be said to be *decomposable* if it can also be written as the union of four *time sibling tiles* $\alpha_Q, \beta_Q, \gamma_Q$ and δ_Q, listed alphabetically with increasing time, all of which have the same frequency interval as Q. Notice that every quartile can be written as the union of four frequency sibling tiles but by (5.1) there are quartiles which cannot be written as the union of four time sibling tiles. Such quartiles will be said to be *indecomposable*. To the frequency siblings correspond Walsh wavelet packets $\omega_{a_Q}, \omega_{b_Q}, \omega_{c_Q}$ and ω_{d_Q} respectively, while $\omega_{\alpha_Q}, \omega_{\beta_Q}, \omega_{\gamma_Q}$ and ω_{δ_Q} correspond to the respective time siblings.

There is a very delicate interplay between the geometry of 4-adic phase plane and boundedness estimates for the Walsh model operator $\mathcal{D}$. This is based on the *phase plane realization* of $\mathcal{D}$. For when $\epsilon_1 = 0$, $\epsilon_2 = 1$, and $\epsilon_3 = 2$, the corresponding operator in (4.4) can be written as

$$\mathcal{D}(F, G) = \sum_{Q \in \mathbb{Q}} \frac{1}{\sqrt{|I_Q|}} \langle F, \mathcal{R}(\omega_{a_Q}) \rangle \langle G, \mathcal{R}(\omega_{b_Q}) \rangle \mathcal{R}(\omega_{c_Q}).$$

The family of all $\mathcal{D}$-operators is obtained by varying the choice and order of frequency siblings used. The remainder of this section develops the interplay between the geometry of 4-adic phase plane and the structure of $\mathcal{D}$.

(5.2) Lemma. *Let Q and Q' be two different quartiles in $\mathbb{Q}$. Then at least three of the intersections $a_Q \cap a_{Q'}$, $b_Q \cap b_{Q'}$, $c_Q \cap c_{Q'}$, $d_Q \cap d_{Q'}$ are empty.*

Proof. First note that if $w_Q = [4^j m, 4^j(m+1))$ then each frequency sibling tile of Q has frequency interval $f_i = [4^j m + (i-1)4^{j-1}, 4^j m + i4^{j-1})$, $i = 1, \dots, 4$ respectively. If $Q \cap Q' = \emptyset$ we are done. So suppose $Q \cap Q' \neq \emptyset$. Without any loss of generality we can assume $w_{Q'} \subset w_Q$ since $Q \neq Q'$. But $w_{Q'} \subseteq f_i$ for some $1 \leq i \leq 4$ since $w_{Q'}$ is itself a 4-adic interval. Then all frequency sibling tiles in Q' have frequency intervals contained in f_{i_0}. Thus all of the frequency intervals of the four frequency sibling tiles of Q' have empty intersection with f_i for all $i \neq i_0$. $\square$

There is a natural *partial* order on $\mathbb{P}$ in which

$$P \leq P' \qquad \Longleftrightarrow \qquad I_P \subseteq I_{P'}, \quad w_P \supseteq w_{P'}.$$

Observe that two tiles are *comparable*, meaning that $P \leq P'$ or $P' \leq P$, if and only if $P \cap P' \neq \emptyset$.

If $\mathcal{P}$ is a set of tiles we denote by $\mathcal{P}^{\min}$ the set of minimal tiles in $\mathcal{P}$ and by $\mathcal{P}^{\max}$ the set of maximal tiles in $\mathcal{P}$.

(5.3) Remark. *The four frequency sibling tiles of an indecomposable quartile are always minimal tiles.*

(5.4) Lemma. *Let $\mathcal{P}$ be a finite set of pairwise disjoint tiles. Let $\widetilde{\mathcal{P}}$ be the set of all tiles $\tilde{P} \in \mathbb{P}$ such that $\tilde{P} \subset \bigcup_{P \in \mathcal{P}} P$. Then either $\mathcal{P} = (\widetilde{\mathcal{P}})^{\min}$ or $\mathcal{P}$ contains four frequency sibling tiles.*

Proof. The tiles in $(\widetilde{\mathcal{P}})^{\min}$ are minimal and therefore pairwise disjoint. They are all covered by tiles in $\mathcal{P}$. Therefore either $\mathcal{P} = (\widetilde{\mathcal{P}})^{\min}$ or there exists $P \in \mathcal{P}$ such that $P \notin (\widetilde{\mathcal{P}})^{\min}$. We pick such a P with *maximal time interval* I_P and we also pick $\tilde{P} \in \widetilde{\mathcal{P}}$ such that $\tilde{P} \leq P$, $\tilde{P} \neq P$. Then $\tilde{P}$ is less or equal than any of the three frequency sibling tiles to P. Since $\tilde{P}$ is covered by the tiles in $\mathcal{P}$ then for each of the three frequency sibling tiles of P there exists a $P' \in \mathcal{P}$ intersecting nontrivially that sibling, and such that $P' \cap P = \emptyset$. Therefore, each of the three frequency sibling tiles to P is less or equal than P'. But then, $P' \in \widetilde{\mathcal{P}}$ and P' is *not minimal*, moreover $|I_{P'}| \leq |I_P|$. Then P' must equal a frequency sibling tile to P and therefore $\mathcal{P}$ contains four frequency siblings. $\square$

(5.5) Lemma. *Let $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ be as in* (5.4)*. Then,*

$$\bigcup_{P \in \mathcal{P}} P = \bigcup_{\tilde{P} \in (\widetilde{\mathcal{P}})^{\min}} \tilde{P}$$

Proof. If $\mathcal{P} = (\widetilde{\mathcal{P}})^{\min}$ we are done. If $\mathcal{P} \neq (\widetilde{\mathcal{P}})^{\min}$, lemma (5.4) ensures that $\mathcal{P}$ contains four frequency sibling tiles none of which belong to $(\widetilde{\mathcal{P}})^{\min}$. So a quartile that is the union of these four frequency sibling tiles is decomposable by (5.3) and we can then replace the four frequency sibling tiles by the four time sibling tiles over the *same* quartile. It suffices now to establish the lemma for the new set of tiles. Doing this successively, the procedure must terminate because changing from frequency sibling tiles to time sibling tiles gives tiles with larger frequency intervals. Since there is an upper bound on the size of the frequency intervals of tiles in $\widetilde{\mathcal{P}}$ (the original $\mathcal{P}$ was a finite set), this procedure terminates when $\mathcal{P}$ becomes $(\widetilde{\mathcal{P}})^{\min}$. $\square$

(5.6) Lemma. *Let* $\mathcal{P}$ *and* $\widetilde{\mathcal{P}}$ *be as in* (5.4)*. Then*

$$\bigcup_{P\in\mathcal{P}} P = \bigcup_{\tilde{P}\in(\widetilde{\mathcal{P}})^{\max}} \tilde{P}$$

Proof. This is a consequence of the dual versions of (5.4) and (5.5). That is, exchange $\leq$ by $\geq$, the words maximal and minimal as well as "frequency sibling tiles" and "time sibling tiles." For the proof of the dual version of (5.5) note that any quartile that is the union of four time sibling tiles is always decomposable (so we can replace the four time sibling tiles by the four frequency sibling tiles over the same quartile). $\square$

(5.7) Corollary. *Let* $\mathcal{P}$ *and* $\widetilde{\mathcal{P}}$ *be as in* (5.4)*. Then for each* $\tilde{P}\in\widetilde{\mathcal{P}}$ *there is a set* $\mathcal{P}'$ *of pairwise disjoint tiles that contains* $\tilde{P}$ *and satisfies*

$$\bigcup_{P'\in\mathcal{P}'} P' = \bigcup_{P\in\mathcal{P}} P$$

Proof. By (5.5) we can assume $\mathcal{P} = (\widetilde{\mathcal{P}})^{\min}$ and also that each $P\in(\widetilde{\mathcal{P}})^{\min}$ intersects $\tilde{P}$ nontrivially. (Just remove from $(\widetilde{\mathcal{P}})^{\min}$ all those which don't. Call that set A.) We get then that $\tilde{P}$ is maximal in the set $\mathcal{B}$ of all tiles in $\bigcup_{P\in(\widetilde{\mathcal{P}})^{\min}\setminus A} P$. By (5.6) we are done once we choose $\mathcal{P}' = \mathcal{B}^{\max}\cup A$. $\square$

We define for quartiles the same partial order as for tiles. Then we have:

(5.8) Definition. *A set* $\mathcal{Q}$ *of quartiles is called "convex", if for all ordered triples* $Q\leq Q'\leq Q''$ *of quartiles we have that if* $Q, Q''\in\mathcal{Q}$ *then* $Q'\in\mathcal{Q}$*.*

(5.9) Lemma. *The union of a finite convex set* $\mathcal{Q}$ *of quartiles can be decomposed into a disjoint union of tiles.*

Proof. We use induction on the number of quartiles. Clearly the lemma is true for the empty set. So given a nonempty finite convex set $\mathcal{Q}$ of quartiles we can pick a *minimal* quartile Q in $\mathcal{Q}$. Let a_Q be one of the four frequency sibling tiles that constitute Q. Let Q' be the quartile which is the union of a_Q and its three time siblings (these tiles are at the same scale as the a_Q). The choice of Q' is uniquely determined because the union of the time interval of a_Q and its three time sibling tiles must be a 4-adic interval on $\mathbb{R}$ of size four times the size of the time interval of a_Q; and there exists a unique one with these conditions since the 4-adic intervals on $\mathbb{R}$ form a grid. If there exists $Q''\in\mathcal{Q}$ besides

Q which intersects a_Q nontrivially then its frequency interval must be contained in or be equal to the frequency interval of a_Q (Q is minimal) and thus also be contained or be equal to the frequency interval of Q'. So $Q \leq Q' \leq Q''$ which implies $Q' \in \mathcal{Q}$ by convexity. This means a_Q is either contained in or disjoint from (in the case that such a Q'' doesn't exist) the union of the quartiles in $\mathcal{Q} \setminus \{Q\}$. This is true for all four frequency sibling tiles that constitute Q, and so the desired result follows from the induction hypothesis (which says that $\mathcal{Q} \setminus \{Q\}$ can be decomposed into a disjoint union of tiles). □

(5.10) Lemma. *Let Q be a quartile. Then each of the wave packets associated to each of the four frequency sibling tiles of Q can be written as a linear combination of wave packets associated to each of the time sibling tiles, and vice versa. Namely,*

$$\omega_{j+1,\ell,4n} = \frac{1}{2}\big(\omega_{j,4\ell,n} + \omega_{j,4\ell+1,n} + \omega_{j,4\ell+2,n} + \omega_{j,4\ell+3,n}\big),$$

$$\omega_{j+1,\ell,4n+1} = \frac{1}{2}\big(\omega_{j,4\ell,n} + \omega_{j,4\ell+1,n} - \omega_{j,4\ell+2,n} - \omega_{j,4\ell+3,n}\big),$$

$$\omega_{j+1,\ell,4n+2} = \frac{1}{2}\big(\omega_{j,4\ell,n} - \omega_{j,4\ell+1,n} + \omega_{j,4\ell+2,n} - \omega_{j,4\ell+3,n}\big),$$

$$\omega_{j+1,\ell,4n+3} = \frac{1}{2}\big(\omega_{j,4\ell,n} - \omega_{j,4\ell+1,n} - \omega_{j,4\ell+2,n} + \omega_{j,4\ell+3,n}\big).$$

Proof. From the inductive definition of the Walsh functions it is easy to see that

$$\mathrm{w}_{j,\ell,2n} = \frac{1}{\sqrt{2}}\big(\mathrm{w}_{j-1,2\ell,n} + \mathrm{w}_{j-1,2\ell+1,n}\big),$$

$$\mathrm{w}_{j,\ell,2n+1} = \frac{1}{\sqrt{2}}\big(\mathrm{w}_{j-1,2\ell,n} - \mathrm{w}_{j-1,2\ell+1,n}\big).$$

By applying these twice we get the corresponding relationships between wave packets in the 4-adic case:

$$\begin{aligned}\omega_{j+1,\ell,4n} &= \frac{1}{\sqrt{2}}\big(\mathrm{w}_{2j+1,2\ell,2n} + \mathrm{w}_{2j+1,2\ell+1,2n}\big),\\ &= \frac{1}{2}\big(\mathrm{w}_{2j,4\ell,n} + \mathrm{w}_{2j,4\ell+1,n} + \mathrm{w}_{2j,4\ell+2,n} + \mathrm{w}_{2j,4\ell+3,n}\big),\\ &= \frac{1}{2}\big(\omega_{j,4\ell,n} + \omega_{j,4\ell+1,n} + \omega_{j,4\ell+2,n} + \omega_{j,4\ell+3,n}\big).\end{aligned}$$

The other identities can be established in exactly the same way. Notice that the coefficient matrix

$$B = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

is an orthogonal matrix. So we can change back and forth, increasing or decreasing the scale as we wish. □

(5.11) Lemma. *Let $\mathcal{P}$ and $\mathcal{P}'$ be two finite sets of pairwise disjoint tiles, which cover the same area in the phase plane. Then the two sets of wave packets $\{\omega_P\}_{P\in\mathcal{P}}$ and $\{\omega_{P'}\}_{P'\in\mathcal{P}'}$ span the same vector subspace of $L^2(\mathbb{R})$.*

Proof. As in (5.5) we replace each set of four frequency sibling tiles by four time sibling tiles without changing the spanned vector space in $L^2(\mathbb{R})$ or in $L^2(\mathbb{R},\mathbb{R}^4)$ by (5.10), until both sets become $(\widetilde{\mathcal{P}})^{\min}$. If the four frequency sibling tiles form an indecomposable quartile they already belong to $(\widetilde{\mathcal{P}})^{\min}$. □

(5.12) Corollary. *Let $\mathcal{P}$ and $\mathcal{P}'$ be two finite sets of pairwise disjoint tiles, which cover the same area in the phase plane. Then the two sets $\{\mathcal{R}(\omega_P)\}_{P\in\mathcal{P}}$ and $\{\mathcal{R}(\omega_{P'})\}_{P'\in\mathcal{P}'}$ span the same vector subspace of $L^2(\mathbb{R},\mathbb{R}^4)$.*

Proof. Let F be a linear combination of $\{\mathcal{R}(\omega_P)\}_{P\in\mathcal{P}}$. Since $\mathcal{R}$ is linear, we can express F as the image under $\mathcal{R}$ of a linear combination of the wave packets $\{\omega_P\}_{P\in\mathcal{P}}$. By (5.11) and the linearity of $\mathcal{R}$, F can be expressed as linear combination of $\{\mathcal{R}(\omega_{P'})\}_{P'\in\mathcal{P}'}$. □

(5.13) Corollary. *Let $S\subseteq\mathbb{R}^2_+$ be a disjoint union of tiles. Then there is a unique vector space in $L^2(\mathbb{R})$ associated to S, and a unique vector space in $L^2(\mathbb{R},\mathbb{R}^4)$ associated to S. Namely, the vector space spanned by $\{\omega_P\}_{P\in S}$ and by $\{\mathcal{R}(\omega_P)\}_{P\in S}$ respectively.*

We denote by Π_S both the orthogonal projection onto this subspace of $L^2(\mathbb{R})$ and onto this subspace of $L^2(\mathbb{R},\mathbb{R}^4)$. We say that the set S *defines a projection*.

(5.14) Corollary. *Let $\mathcal{P}$ be a finite set of pairwise disjoint tiles and let $\mathcal{A}$ denote their union. Then the wave packet ω_P of any tile $P\subset\mathcal{A}$ is contained in the vector space spanned by the wave packets of the tiles in $\mathcal{P}$. And moreover, $\mathcal{R}(\omega_P)$ is contained in the vector space spanned by the functions $\mathcal{R}$ of the wave packets of the tiles in $\mathcal{P}$.*

Proof. That ω_P is contained in the vector space spanned by the wave packets of the tiles in $\mathcal{P}$ is a consequence of Corollaries (5.7) and (5.13). The corresponding result for $\mathcal{R}(\omega_P)$ relies also on the linearity of $\mathcal{R}$. □

(5.15) Lemma. *If two tiles P, P' are disjoint, then the corresponding wave packets ω_P and $\omega_{P'}$ are orthogonal in $L^2(\mathbb{R})$.*

Proof. If $I_P\cap I_{P'}=\emptyset$ then $\langle\omega_P,\omega_{P'}\rangle=0$. If $I_P\cap I_{P'}\neq\emptyset$ then we must have that $w_P\cap w_{P'}=\emptyset$ since P and P' are disjoint. Without any loss of generality we can assume $I_P, I_{P'}\subseteq[0,1)$, since if $|I_P|=4^j$ and $|I_{P'}|=4^{j'}$ with $j\geq j'$, by a rescaling of the time axis we have that $\langle\omega_P,\omega_{P'}\rangle=2^{j-j'}\langle\omega_{\tilde P},\omega_{\tilde P'}\rangle$ where $\tilde P=[0,1)\times[m,m+1)$, $|I_{\tilde P'}|=4^{-(j-j')}$ and $w_{\tilde P}\cap w_{\tilde P'}=\emptyset$.

Let $\mathcal{P}$ be the set of tiles of the form $[0,1)\times w$ with $|w|=1$ which intersect P and let $\mathcal{P}'$ be the set of tiles of the same form which intersect P'. These two sets are disjoint. Moreover, the wave packets corresponding to tiles of the form $[0,1)\times w$ are Walsh functions restricted to $[0,1)$. Hence by (5.14) the two wave packets in question are contained in orthogonal subspaces of $L^2(\mathbb{R})$ since Walsh functions are orthogonal on $[0,1)$. □

The corresponding result for the $\mathcal{R}(\omega_P)$ follows at once from

$$\langle\mathcal{R}(\omega_P),\mathcal{R}(\omega_{P'})\rangle=\langle\omega_P,\omega_{P'}\rangle$$

since $\mathcal{R}^*\mathcal{R}=I$ on $L^2(\mathbb{R})$.

(5.16) Corollary. *If two tiles P, P' are disjoint, then $\mathcal{R}(\omega_P)$ and $\mathcal{R}(\omega_{P'})$ are orthogonal in $L^2(\mathbb{R}, \mathbb{R}^4)$.*

6. Bilinear Marcinkiewicz interpolation.

In the final analysis establishing boundedness for such bilinear operators as those in the previous sections always relies on linear or bilinear real interpolation between weak type estimates. In this section we collect together the requisite interpolation results and then show how the main theorem follows from them once weak type estimates have been established.

Because of the need to allow Lebesgue L^p-spaces and Lorentz L^{pq}-spaces with $p < 1$ it will be important to formulate all the interpolation results in the context of *quasi-Banach* spaces. It will also be important to allow Hilbert space-valued functions. Thus we shall be considering Lebesgue spaces $L^p(\mathbb{R}, \mathcal{H})$-spaces and Lorentz spaces $L^{pq}(\mathbb{R}, \mathcal{H})$-spaces with $0 < p < \infty$ and $\mathcal{H}$ a real or complex Hilbert space. Recall the following real interpolation between such spaces which is well-known in its scalar-valued form, though less-well known in general ([1] page 121).

(6.1) Theorem. *Let $\mathcal{H}$ be a real or complex Hilbert space. Then the equality*

$$(L^{p_0 q_0}(\mathbb{R}, \mathcal{H}),\ L^{p_1 q_1}(\mathbb{R}, \mathcal{H}))_{\theta, q} = L^{pq}(\mathbb{R}, \mathcal{H}) \qquad \left(\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}\right)$$

holds up to equivalence of norms for all p_i, q_i with $0 < p_i \leq \infty$ and $1 \leq q_i \leq \infty$.

As Michael Cwikel has pointed out to us, however, the vector-valued version of (6.1) for any Banach space A can be deduced from the scalar-valued version: let X_0, X_1 be Banach spaces of measurable functions on the same measure space Ω and let $X_i(A)$ be the space of A-valued strongly measurable functions $f : \Omega \to A$ for which $\| \|f(.)\|_A \|_{X_i}$ is finite. Then it is straightforward to show that the K-functionals associated with the respective compatible pairs $\mathbf{B} = (X_0(A), X_1(A))$ and $\mathbf{X} = (X_0, X_1)$ satisfy the identity

$$K(t, f : \mathbf{B}) = K(t, \|f(.)\|_A : \mathbf{X}) \qquad (t > 0)$$

for an A-valued function f. Interpolation between spaces of A-valued functions can thus be read off from the corresponding scalar-valued results.

Bilinear interpolation results paralleling Young's inequality for convolution have been known for Banach spaces from the early days of interpolation theory, but a result paralleling Hölder's inequality for pointwise multiplication precisely, valid for both p-Banach spaces and Banach spaces, is needed. This was given by Janson though his result will be formulated here exactly as it is required ([17]).

Fix p_0, p_1 and p_2 with $1 < p_0 < p_1 \leq p_2 < \infty$ and let $\mathcal{B} = \mathcal{B}(f, g)$ be a bilinear operator which is defined on some space $\mathcal{F}(\mathbb{R}, \mathcal{H})$ of functions which is dense in every $L^p(\mathbb{R}, \mathcal{H})$, $0 < p < \infty$. Suppose further that $\mathcal{B}$ extends to a bounded bilinear operator

$$\mathcal{B} : \begin{cases} L^{p_0}(\mathbb{R}, \mathcal{H}) \times L^{p_2}(\mathbb{R}, \mathcal{H}) \longrightarrow L^{r_0, \infty}(\mathbb{R}, \mathcal{H}), \\ L^{p_2}(\mathbb{R}, \mathcal{H}) \times L^{p_0}(\mathbb{R}, \mathcal{H}) \longrightarrow L^{r_0, \infty}(\mathbb{R}, \mathcal{H}), \end{cases} \qquad \left(\frac{1}{r_0} = \frac{1}{p_0} + \frac{1}{p_2}\right) \tag{6.2}$$

as well as

$$\mathcal{B}:\begin{cases} L^{p_1}(\mathbb{R},\mathcal{H})\times L^{p_2}(\mathbb{R},\mathcal{H})\longrightarrow L^{r_1,\infty}(\mathbb{R},\mathcal{H}),\\ L^{p_2}(\mathbb{R},\mathcal{H})\times L^{p_1}(\mathbb{R},\mathcal{H})\longrightarrow L^{r_1,\infty}(\mathbb{R},\mathcal{H}),\end{cases}\qquad \left(\frac{1}{r_1}=\frac{1}{p_1}+\frac{1}{p_2}\right). \tag{6.3}$$

The result of Janson replaces these weak type estimates with strong type estimates as in the classical Marcinkiewicz theorem by setting up an abstract interpolation structure.

(6.4) Theorem. *If $\mathcal{B}$ is bilinear operator which is defined on all functions in $\mathcal{F}(\mathbb{R},\mathcal{H})$, then it extends to a bilinear operator*

$$\mathcal{B}: L^p(\mathbb{R},\mathcal{H})\times L^q(\mathbb{R},\mathcal{H})\longrightarrow L^r(\mathbb{R},\mathcal{H})\qquad \left(\frac{1}{r}=\frac{1}{p}+\frac{1}{q}\right)$$

that is bounded for all

$$p_0<p,q<p_2,\qquad \frac{1}{r_1}<\frac{1}{p}+\frac{1}{q}<\frac{1}{r_0}$$

once (6.2) *and* (6.3) *are known to be bounded.*

The proof is simply an application of Janson's result in a particular case. Because the form of the conditions on p, q and r are the same as in Hölder's theorem, *i.e.* corresponding to pointwise multiplication, we shall call (6.4) the *Bilinear* Marcinkiewicz Interpolation theorem for multiplication type operators.

Proof of (6.4). Fix compatible pairs

$$\mathbf{X}_1=(L^{p_0}(\mathbb{R},\mathcal{H}),\, L^{p_2}(\mathbb{R},\mathcal{H}))=\mathbf{X}_2,\qquad \mathbf{Y}=(L^{r_0,\infty}(\mathbb{R},\mathcal{H}), L^{p_2/2,\infty}(\mathbb{R},\mathcal{H})).$$

of complete quasi-normed Abelian groups and let

$$\mathbf{X}_{i,\theta_i,q_i}=((L^{p_0}(\mathbb{R},\mathcal{H}),\, L^{p_2}(\mathbb{R},\mathcal{H}))_{\theta_i,q_i},\qquad \mathbf{Y}_{\theta,q}=(L^{r_0,\infty}(\mathbb{R},\mathcal{H}),\, L^{p_2/2,\infty}(\mathbb{R},\mathcal{H}))_{\theta,q}.$$

be the corresponding real interpolation spaces; these can be identified using (6.1). Now set $\theta=-1+\phi+\psi$ for each pair $(\phi,\psi)\in[0,1]\times[0,1]$. Janson's result relates the boundedness of $\mathcal{B}$ with the geometry of the set Φ of pairs (ϕ,ψ) for which

$$\mathcal{B}:\mathbf{X}_{1,\phi,q_1}\times\mathbf{X}_{2,\psi,q_2}\longrightarrow\mathbf{Y}_{\theta,\infty}$$

is bounded for some choice of q_1, q_2 varying possibly with (ϕ,ψ). Property (6.2) ensures that Φ contains the points $(1,0)$ and $(0,1)$, while (6.3) ensures that Φ contains the points $(\theta_1,1)$ and $(1,\theta_1)$ when θ_1 is chosen so that

$$L^{p_1}(\mathbb{R},\mathcal{H})=(L^{p_0}(\mathbb{R},\mathcal{H}),\, L^{p_2}(\mathbb{R},\mathcal{H}))_{\theta_1,p_1}.$$

Consequently,

$$\mathcal{B}: L^p(\mathbb{R},\mathcal{H})\times L^q(\mathbb{R},\mathcal{H})\longrightarrow L^r(\mathbb{R},\mathcal{H})\qquad \left(\frac{1}{r}=\frac{1}{p}+\frac{1}{q}\right)$$

will be bounded for any pair (ϕ, ψ) in the interior of the convex set having these four points in Ω as extreme points ([17], theorem 2). Such points correspond precisely to the restrictions on p, q given in (6.4), thus completing the proof of (6.4). $\square$

Although (6.1) can fail when the same Hilbert space or Banach space is not used throughout (*cf.* [12]), it is worth noting that the proof above remains valid for arbitrary Banach spaces A_0, A_1 and A_2 in the sense that a bilinear operator $\mathcal{B}$ will be bounded as a mapping

$$\mathcal{B} : L^p(\mathbb{R}, A_0) \times L^q(\mathbb{R}, A_1) \longrightarrow L^r(\mathbb{R}, A_2) \qquad \left(\frac{1}{r} = \frac{1}{p} + \frac{1}{q}\right) \tag{6.5}$$

for the same range as in (6.4) once

$$\mathcal{B} : \begin{cases} L^{p_0}(\mathbb{R}, A_0) \times L^{p_2}(\mathbb{R}, A_1) \longrightarrow L^{r_0,\infty}(\mathbb{R}, A_2), \\ L^{p_2}(\mathbb{R}, A_0) \times L^{p_0}(\mathbb{R}, A_1) \longrightarrow L^{r_0,\infty}(\mathbb{R}, A_2), \end{cases} \qquad \left(\frac{1}{r_0} = \frac{1}{p_0} + \frac{1}{p_2}\right) \tag{6.6}$$

as well as

$$\mathcal{B} : \begin{cases} L^{p_1}(\mathbb{R}, A_0) \times L^{p_2}(\mathbb{R}, A_1) \longrightarrow L^{r_1,\infty}(\mathbb{R}, A_2), \\ L^{p_2}(\mathbb{R}, A_0) \times L^{p_1}(\mathbb{R}, A_1) \longrightarrow L^{r_1,\infty}(\mathbb{R}, A_2), \end{cases} \qquad \left(\frac{1}{r_1} = \frac{1}{p_1} + \frac{1}{p_2}\right) \tag{6.7}$$

are known to bounded.

The next result reduces the proof of (4.8), and hence that of the Main Theorem, to establishing weak type estimates for $\mathcal{D}$, *i.e.*, to establishing the boundedness of the mapping $\mathcal{D} : L^p(\mathbb{R}) \times L^q(\mathbb{R}) \to L^{r\infty}(\mathbb{R})$.

(6.8) Theorem. *Suppose $\mathcal{D}$ is bounded from $L^p(\mathbb{R}) \times L^q(\mathbb{R})$ into $L^{r\infty}(\mathbb{R})$ for all choices of triples $(\epsilon_1, \epsilon_2, \epsilon_3)$ having integer entries satisfying* (1.11) *whenever* $1 < p < 2 \leq q < \infty$. *Then each such $\mathcal{D}$ extends to a bounded operator*

$$\mathcal{D} : L^p(\mathbb{R}) \times L^q(\mathbb{R}) \longrightarrow L^r(\mathbb{R})$$

for all $1 < p,\, q < \infty$ *for which* $1/p + 1/q = 1/r < 3/2$.

Proof. Choose $1 < p_0 < 2 \leq p_2 < \infty$. Then by symmetry, $\mathcal{D}$ satisfies (6.2) with $\mathcal{H} = \mathbb{R}^4$. Similarly, $\mathcal{D}$ will satisfy (6.3) also for any fixed choice of $p_1,\ p_0 < p_1 < 2$. Thus by the bilinear Marcinkiewicz theorem above, $\mathcal{D}$ will extend to a bounded operator

$$\mathcal{D} : L^p(\mathbb{R}) \times L^q(\mathbb{R}) \longrightarrow L^r(\mathbb{R})$$

for all $1 < p, q < \infty$ for which

$$\frac{1}{p_1} + \frac{1}{p_2} < \frac{1}{p} + \frac{1}{q} < \frac{1}{p_0} + \frac{1}{p_2}$$

By varying the choice of p_0, p_1 and p_2 we thus obtain (6.8) for all $1 < p,\, q < \infty$ with $1/2 < 1/p + 1/q < 3/2$.

To extend the range, we exploit the self-adjoint property of the family of $\mathcal{D}$ operators. Fix $q_0 > 4$. We prove first that $\mathcal{D}$ maps $L^{q_0} \times L^{q_0}$ into $L^{q_0/2}$ (this corresponds to the case proved separately by Thiele using tri-linear estimates). Let $L^{p_0}(\mathbb{R}, \mathcal{H})$, $1 < p_0 < 2$, be the dual space of $L^{q_0/2}$ and let $L^{q_1}(\mathbb{R}, \mathcal{H})$ be the dual space of $L^{q_0}(\mathbb{R}, \mathcal{H})$. Then for each H in $L^{p_0}(\mathbb{R}, \mathcal{H})$ and F, G in $L^{q_0}(\mathbb{R}, \mathcal{H})$,

$$\langle \mathcal{D}(F,G), H \rangle = \sum_{j,\ell,n} 2^{-j} \big\langle F, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_1})\big\rangle \big\langle G, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_2})\big\rangle \big\langle H, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_3})\big\rangle$$

$$= \Big\langle G, \Big\{ \sum_{j,\ell,n} 2^{-j} \big\langle H, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_3})\big\rangle \big\langle F, \mathcal{R}(\omega_{j,\ell,4n+\epsilon_1})\big\rangle \mathcal{R}(\omega_{j,\ell,4n+\epsilon_2}) \Big\} \Big\rangle.$$

But

$$\frac{1}{q_1} = \frac{1}{q_0} + \frac{1}{p_0} = \frac{1}{q_0} + \frac{q_0 - 2}{q_0} = 1 - \frac{1}{q_0} > \frac{1}{2}.$$

So by our earlier strong type results,

$$|\langle \mathcal{D}(F,\, G),\, H \rangle| \;\leq\; \text{const.}\; \|F\|_{L^{q_0}} \|G\|_{L^{q_0}} \|H\|_{L^{p_0}},$$

showing indeed that $\mathcal{D}$ maps $L^{q_0} \times L^{q_0}$ into $L^{q_0/2}$. Consequently, $\mathcal{D}$ is bounded as a bilinear operator

$$\mathcal{D} : \begin{cases} L^{p_0}(\mathbb{R}, \mathcal{H}) \times L^{q_0}(\mathbb{R}, \mathcal{H}) \longrightarrow L^{r_0,\infty}(\mathbb{R}, \mathcal{H}), \\ L^{q_0}(\mathbb{R}, \mathcal{H}) \times L^{p_0}(\mathbb{R}, \mathcal{H}) \longrightarrow L^{r_0,\infty}(\mathbb{R}, \mathcal{H}), \end{cases} \qquad \left(\frac{1}{r_0} = \frac{1}{p_0} + \frac{1}{q_0} \right)$$

as well as

$$\mathcal{D} : L^{q_0}(\mathbb{R}, \mathcal{H}) \times L^{q_0}(\mathbb{R}, \mathcal{H}) \longrightarrow L^{r_1,\infty}(\mathbb{R}, \mathcal{H}), \quad \left(\frac{1}{r_1} = \frac{2}{q_0} \right).$$

We can thus apply the bilinear Marcinkiewicz theorem again, establishing boundedness in (6.8) for all p, q such that

$$\frac{1}{p_0} + \frac{1}{q_0} > \frac{1}{r} = \frac{1}{p} + \frac{1}{q} > \frac{2}{q_0}.$$

The general result follows by letting $q_0 \to \infty$. □

7. Weak type estimates for $\mathcal{D}$.

In this section we will complete the proofs of the boundedness results for the various Walsh model operators by showing that the 4-adic version

$$\mathcal{D} : F, G \longrightarrow \sum_{Q \in \mathbb{Q}} \frac{1}{\sqrt{|I_Q|}} \langle\, F,\, \mathcal{R}(\omega_{a_Q})\,\rangle \,\langle\, G,\, \mathcal{R}(\omega_{b_Q})\,\rangle\, \mathcal{R}(\omega_{c_Q})$$

satisfies the weak type estimate

$$|\{x : \|\mathcal{D}(F,G)(x)\| > \lambda\,\}| \leq \text{const.} \left(\frac{\|F\|_{L^p} \|G\|_{L^q}}{\lambda} \right)^r \qquad (\lambda > 0) \tag{7.1}$$

provided
$$1 < p < 2 \le q < \infty, \qquad \frac{1}{p} + \frac{1}{q} = \frac{1}{r}.$$
To be able to exploit the structure of the family of such operators, however, it is important to observe the proof of (7.1) does not depend on which three of the four frequency siblings $\{a_Q, b_Q, c_Q, d_Q\}$ are used to define $\mathcal{D}$, nor on the order in which they appear, so long as the choice and order are fixed for all Q.

In proving (7.1), the lack of a natural total order in the geometry of $\mathcal{D}$ forces us to use L^2-techniques. We shall partition $\mathbb{R}$ into a 'good' set $\mathbb{R} \setminus E$ where L^2-estimates can be established and a 'bad' set E whose measure can be controlled. To that end fix F in $L^p(\mathbb{R})$, $1 < p < 2$, so that $G \to \mathcal{D}(F, G)$ becomes a linear operator; we shall further assume that F is a linear combination of characteristic functions of 4-adic intervals, though this requirement is not always necessary.

(7.2) Main Lemma. *For each $\kappa > 0$ there exists a set $E \subset \mathbb{R}$ such that the estimates*

(i)
$$|E| \le \left(\frac{\|F\|_{L^p}}{\kappa}\right)^p,$$

and

(ii)
$$\left(\int_{\mathbb{R}\setminus E} \|\mathcal{D}(F,G)(x)\|^2\, dx\right)^{1/2} \le \text{const.}\left(\frac{\kappa}{\|F\|_{L^p}}\right)^{\alpha} \|F\|_{L^p}\, \|G\|_{L^q}$$

hold uniformly for all $G \in L^q(\mathbb{R}, \mathbb{R}^4)$ with $2 \le q < \infty$ and $\alpha = p(1/r - 1/2)$.

Given $\lambda > 0$, the weak estimate (7.1) follows easily from (7.2) by choosing
$$\kappa = \left(\|F\|^p_{L^p} \frac{\lambda^q}{\|G\|^q_{L^q}}\right)^{1/p+q}.$$

For then
$$|E| \le \left(\frac{\|F\|_{L^p}}{\kappa}\right)^p \le \text{const.}\left(\frac{\|F\|_{L^p}\|G\|_{L^q}}{\lambda}\right)^r,$$
while
$$\int_{\mathbb{R}\setminus E} \|\mathcal{D}(F,G)(x)\|^2\, dx \le \text{const.}(\kappa^{\alpha}\|F\|^{1-\alpha}_{L^p}\|G\|_{L^q})^2$$
$$\le \text{const.}\ \lambda^2 \left(\frac{\|F\|_{L^p}\|G\|_{L^q}}{\lambda}\right)^r.$$

In this case
$$|\{x : \|\mathcal{D}(F,G)(x)\| > \lambda\}| \ \le\ \frac{1}{\lambda^2}\int_{\mathbb{R}\setminus E} \|\mathcal{D}(F,G)(x)\|^2\, dx + |E|$$
$$\le\ \text{const.}\left(\frac{\|F\|_{L^p}\|G\|_{L^q}}{\lambda}\right)^r,$$

establishing (7.1). □

We are thus reduced to proving the Main Lemma (7.2). It in turn relies on two key estimates:

- Proposition (7.4) which will control the measure of the set E, and
- Proposition (7.13) which will yield the L^2 estimate on $\mathbb{R} \setminus E$.

Their proof will run parallel to those given by Thiele in the scalar-valued setting. Given a family $\mathcal{P}$ of pairwise disjoint tiles in $\mathbb{P}$, set

$$N_{\mathcal{P}}(x) = \text{card}\{P \in \mathcal{P} : x \in I_P\}.$$

We call this the *counting function* for the set $\mathcal{P}$. Now let $\mathbb{P}_k$ be a set of pairwise disjoint tiles in $\mathbb{P}$ such that

$$\|\Pi_P(F)\|_{L^\infty} = \|\langle F, \mathcal{R}(\omega_P)\rangle \mathcal{R}(\omega_P)\|_{L^\infty} = \frac{|\langle F, \mathcal{R}(\omega_P)\rangle|}{\sqrt{|I_P|}} \geq 2^k, \qquad (P \in \mathbb{P}_k).$$

By Parseval's inequality and (5.16),

$$\int_{-\infty}^{\infty} N_{\mathbb{P}_k}(x)\, dx = \sum_{P \in \mathbb{P}_k} |I_P| \leq \text{const.} \left(\frac{\|F\|_{L^2}}{2^k} \right)^2. \tag{7.3}$$

We call such an estimate a *Carleson measure estimate*. Of course, both $\mathbb{P}_k$ and its Carleson measure estimate depend on the fixed choice of F.

(7.4) Proposition. *For each $F \in L^p(\mathbb{R}, \mathbb{R}^4)$, $1 < p < 2$, and $k \in \mathbb{Z}$ the counting function $N_{\mathbb{P}_k}$ satisfies the inequality*

$$\|N_{\mathbb{P}_k}^{1/p'+\varepsilon}\|_{L^p} \ \leq \ \text{const.} \frac{\|F\|_{L^p}}{2^k}$$

uniformly in F for each $\varepsilon > 0$ with constant depending on p and ε.

Proof. For each $x \in \mathbb{R}$, let $B(x) = \{\Pi_P(F)(x)\}_{P \in \mathbb{P}_k}$. Then the $L^p(\mathbb{R}, l^s)$-norm of this function $B = B(x)$ is given by

$$\|B\|_{L^p(l^s)} = \left(\int_{-\infty}^{\infty} \left(\sum_{P \in \mathbb{P}_k} \|\Pi_P(F)(x)\|^s \right)^{p/s} dx \right)^{1/p}.$$

Thus again by Parseval's inequality and (5.16)

$$\|B\|_{L^2(l^2)} = \sum_{P \in \mathbb{P}_k} \|\Pi_P(F)\|_{L^2}^2 \leq \|F\|_{L^2},$$

while

$$\|B\|_{L^{1+\varepsilon}(l^\infty)}^{1+\varepsilon} = \int_{-\infty}^{\infty} \left(\sup_{P \in \mathbb{P}_k} \|\Pi_P(F)(x)\| \right)^{1+\varepsilon} dx$$

$$\leq \int_{-\infty}^{\infty} (M(F)(x))^{1+\varepsilon}\, dx \leq \|F\|_{L^{1+\varepsilon}}^{1+\varepsilon}$$

for some small $\varepsilon > 0$ since

$$\|\Pi_P(F)(x)\| \le \frac{1}{|I_P|}\int_{I_P} \|F(y)\|\,dy \le M(F)(x).$$

But then by complex interpolation the inequality

$$\|B\|_{L^p(l^{p'+\tilde{\varepsilon}})} \;\le\; \text{const.}\,\|F\|_{L^p} \qquad (1<p<2) \tag{7.5}$$

holds for sufficiently small $\tilde{\varepsilon} > 0$ ([1]). On the other hand,

$$N_{\mathbb{P}_k}(x) \;=\; \sum_{P\in\mathbb{P}_k} \chi_P(x) \;\le\; \sum_{P\in\mathbb{P}_k} \big(2^{-k}\|\Pi_P(F)(x)\|\big)^{p'+\tilde{\varepsilon}}\,.$$

Hence, in view of (7.5),

$$\Big(\int_{-\infty}^{\infty} \big(N_{\mathbb{P}_k}(x)\big)^{p/p'+\tilde{\varepsilon}}\,dx\Big)^{1/p} \le \frac{1}{2^k}\Big(\int_{-\infty}^{\infty} \big(\|B(x)\|_{\ell^{p'+\tilde{\varepsilon}}}\big)^p dx\Big)^{1/p}$$

$$= \frac{1}{2^k}\,\|B\|_{L^p(\ell^{p'+\varepsilon})} \le \text{const.}\,\frac{\|F\|_{L^p}}{2^k},$$

completing the proof. □

When the projection $\Pi_P(F) = \langle F,\, \mathcal{R}(\omega_P)\,\rangle\, \mathcal{R}(\omega_P)$ onto the subspace spanned by the wavelet packet associated with a single tile is replaced by the projection

$$\Pi_Q(F) = \sum_{z=a}^{d} \langle F, \mathcal{R}(\omega_{z_Q})\rangle \mathcal{R}(\omega_{z_Q})$$

onto the subspace spanned by the wavelet packets associated with the four frequency sibling tiles $\{a_Q, b_Q, c_Q, d_Q\}$ in a given quartile Q, there is a companion Carleson measure estimate. To isolate the L^∞-norm of each of the terms in this sum, however, we set

$$\|\Pi_Q(F)\|_\infty = \max_{z=a,\dots,d} \|\Pi_{z_Q}(F)\|_{L^\infty} = \max_{z=a,\dots,d} \|\langle F, \mathcal{R}(\omega_{z_Q})\rangle \mathcal{R}(\omega_{z_Q})\|_{L^\infty} \tag{7.6}$$

where the maximum is taken over the four frequency siblings in Q.

(7.7) Corollary. *The Carleson measure estimates of* (7.4) *remain valid when the sets* $\mathbb{P}_k$ *of pairwise disjoint tiles are replaced by any set* $\mathbb{Q}_k$ *of pairwise disjoint quartiles such that* $\|\Pi_Q(F)\|_\infty \;\ge\; 2^k$ *for all* $Q \in \mathbb{Q}_k$.

Proof. If Q is a quartile in $\mathbb{Q}_k$ such that $\|\Pi_Q F\|_{L^\infty} \ge 2^k$, then one of the frequency sibling tiles in Q, say a_Q, satisfies

$$\|\Pi_{a_Q} F\|_{L^\infty} \ge 2^k.$$

For each quartile $Q \in \mathbb{Q}_k$ we pick one such tile and call this set of tiles $\mathfrak{P}_k$. The corollary follows applying (7.4) to this set $\mathfrak{P}_k$. □

Just as in Fefferman's proof of the almost everywhere convergence of Fourier series of L^2-functions, the proof of the Main Lemma proceeds by decomposing the operator $\mathcal{D}$ into sums over 'trees' which in our case will be collections of quartiles. Recall the partial order on the set $\mathbb{Q}$ of quartiles in which

$$Q \leq Q' \quad \Longleftrightarrow \quad I_Q \subseteq I_{Q'}, \quad w_{Q'} \subseteq w_Q .$$

We define the *density* of the quartile Q with respect to F by

$$\delta(Q, F) = \sup_{Q' \geq Q} \|\Pi_{Q'}(F)\|_\infty$$

and set

$$\mathbb{Q}_k(f) = \{ Q \in \mathbb{Q} : 2^k \leq \delta(Q, F) < 2^{k+1} \} \qquad (k \in \mathbb{Z}).$$

These are the quartiles having *homogeneous density*. They provide the initial decomposition

$$\mathcal{D}(F, G) = \sum_{k \in \mathbb{Z}} \Bigg(\sum_{Q \in \mathbb{Q}_k} \frac{1}{\sqrt{|I_Q|}} \langle F, \mathcal{R}(\omega_{a_Q}) \rangle \langle G, \mathcal{R}(\omega_{b_Q}) \rangle \mathcal{R}(\omega_{c_Q}) \Bigg)$$

of $\mathcal{D}(F, G)$. Notice also that each $\mathbb{Q}_k$ is finite since F is a finite linear combination of characteristic functions, so it makes sense to speak of maximal quartiles in $\mathbb{Q}_k$ and to count how many maximal quartiles in $\mathbb{Q}_k$ are greater than a given quartile. To quantify these ideas let $\mathbb{Q}_k^{\max}$ be the set of all such maximal quartiles in $\mathbb{Q}_k$.

(7.8) Definition. *For each $i \geq 0$ the set*

$$\mathbb{Q}_{k,i} = \{ Q \in \mathbb{Q}_k : 2^i \leq \operatorname{card}\{ Q' \in \mathbb{Q}_k^{\max} : Q' \geq Q \} < 2^{i+1} \}$$

of quartiles in $\mathbb{Q}_k$ will be called a forest. Conversely, for each maximal quartile $Q^{\max}$ in $\mathbb{Q}_{k,i}$, the set

$$\mathbb{T}_{Q^{\max}} = \{ Q \in \mathbb{Q}_{k,i} : Q \leq Q^{\max} \}$$

of quartiles in $\mathbb{Q}_{k,i}$ will be called a tree having tree top $Q^{\max}$.

With this decomposition, $\mathcal{D}(F, G)$ now becomes a sum

$$\mathcal{D}(F, G) = \sum_{k \in \mathbb{Z}} \Bigg(\sum_{i=0}^{\infty} \mathcal{F}_{\mathbb{Q}_{k,i}}(G) \Bigg) \tag{7.9}$$

of *Forest operators*

$$\mathcal{F}_{\mathbb{Q}_{k,i}}(G) = \sum_{Q \in \mathbb{Q}_{k,i}} \frac{1}{\sqrt{|I_Q|}} \langle F, \mathcal{R}(\omega_{a_Q}) \rangle \langle G, \mathcal{R}(\omega_{b_Q}) \rangle \mathcal{R}(\omega_{c_Q}).$$

The union of trees gives back a forest, but we can be much more precise,

(7.10) Lemma (Fefferman-Thiele). *Any two trees in $\mathbb{Q}_{k,i}$ are disjoint.*

Proof. Suppose not, then there exists $Q \in \mathbb{Q}_{k,i}$ such that $Q \leq Q^{\max}$ and $Q \leq {Q'}^{\max}$ where both $Q^{\max}$ and ${Q'}^{\max}$ are maximal quartiles in $\mathbb{Q}_{k,i}$. But $Q^{\max}$ and ${Q'}^{\max}$ must be less than at least 2^i maximal quartiles in $\mathbb{Q}_k$ since they both belong to $\mathbb{Q}_{k,i}$. Furthermore, these two collections of maximal quartiles must be disjoint because by maximality the frequency intervals of $Q^{\max}$ and ${Q'}^{\max}$ are disjoint. Hence Q is less than at least 2^{i+1} maximal quartiles in $\mathbb{Q}_k$, contradicting the fact that Q belongs to $\mathbb{Q}_{k,i}$. $\square$

Consequently, each Forest operator in (7.9) is itself a sum

$$\mathcal{F}_{\mathbb{Q}_{k,i}} = \sum_{Q^{\max} \in \mathbb{Q}_{k,i}} \mathcal{T}_{Q^{\max}}$$

of *Tree operators*

$$\mathcal{T}_{Q^{\max}}(G) = \sum_{Q \in \mathbb{T}_{Q^{\max}}} \frac{1}{\sqrt{|I_Q|}} \langle\, F,\, \mathcal{R}(\omega_{a_Q})\,\rangle\, \langle\, G,\, \mathcal{R}(\omega_{b_Q})\,\rangle\, \mathcal{R}(\omega_{c_Q}).$$

These are the most basic operators. They are *local* in the sense that $\mathcal{T}_{Q^{\max}}(G)$ depends only on the restriction of G to the maximal time interval $I_{Q^{\max}}$; on the other hand, the crucial point of (7.10) is that the Tree operators making up a forest operator are *orthogonal*. Both properties are key to determining the L^2-operator norm of a forest operator once the following two results are in hand.

(7.11) Lemma. *The set of all quartiles in a tree is a finite convex set of quartiles.*

Proof. By construction any tree can contain only finitely many quartiles because F was assumed to be a finite linear combination of step functions. To check convexity, let $\mathbb{T}_{Q^{\max}}$ be a tree in $\mathbb{Q}_{k,i}$ having $Q^{\max}$ for its top and suppose $Q \leq Q' \leq Q''$ holds for some pair of quartiles Q, Q'' in $\mathbb{T}_{Q^{\max}}$. Then $Q' \leq Q^{\max}$ and

$$2^k \leq \delta(Q, F) \leq \delta(Q', F) \leq \delta(Q'', F) < 2^{k+1},$$

so $Q' \in \mathbb{Q}_k$. On the other hand, since $Q \leq Q' \leq Q''$,

$$\{\tilde{Q} \in \mathbb{Q}_k^{\max} \,:\, \tilde{Q} \geq Q''\} \subseteq \{\tilde{Q} \in \mathbb{Q}_k^{\max} \,:\, \tilde{Q} \geq Q'\} \subseteq \{\tilde{Q} \in \mathbb{Q}_k^{\max} \,:\, \tilde{Q} \geq Q\}.$$

Hence, $Q' \in \mathbb{Q}_{k,i}$ and $Q' \in \mathbb{T}_{Q^{\max}}$ since $Q' \leq Q^{\max}$. $\square$

(7.12) Lemma. *Let $\mathbb{T}_{Q^{\max}}$ be a tree in $\mathbb{Q}_{k,i}$ having the quartile $Q^{\max}$ as its top, and let $\Pi_{\mathbb{T}_{Q^{\max}}}$ be the orthogonal projection onto the subspace of $L^2(\mathbb{R}, \mathbb{R}^4)$ defined by the union $A_{\mathbb{T}_{Q^{\max}}} = \cup_{Q \in \mathbb{T}_{Q^{\max}}} Q$ of the quartiles in $\mathbb{T}_{Q^{\max}}$. Then the inequality*

$$\|\Pi_{\mathbb{T}_{Q^{\max}}}(F)(x)\| \leq \text{const.}\, 2^k$$

holds for every $x \in I_{Q^{\max}}$.

Proof. By Lemma (5.9) and (7.11) a tree can be decomposed into the union of disjoint tiles. So if $\widetilde{\mathcal{P}} = \{P \in \mathbb{P} : P \subseteq A_{\mathbb{T}_{Q^{\max}}}\}$, by (5.5) and (5.13) the orthogonal projection onto the subspace of $L^2(\mathbb{R}, \mathbb{R}^4)$ associated with $A_{\mathbb{T}_{Q^{\max}}}$ may be realized as

$$\Pi_{\mathbb{T}_{Q^{\max}}}(F)(x) = \sum_{P \in (\widetilde{\mathcal{P}})^{\min}} \langle F, \mathcal{R}(\omega_P)\rangle \mathcal{R}(\omega_P)(x).$$

Fix $x \in I_{Q^{\max}}$ and let Q be the minimal quartile in $\mathbb{T}_{Q^{\max}}$ containing x. If Q is decomposable, let z_Q, $z = \alpha, \ldots, \gamma$, be its four time sibling tiles; without loss of generality we can assume $x \in I_{\alpha_Q}$. If $\alpha_Q \in (\widetilde{\mathcal{P}})^{\min}$, lemma (5.15) ensures that

$$\Pi_Q(F)(x) = \Pi_{\alpha_Q}(F)(x) = \Pi_{\mathbb{T}_{Q^{\max}}}(F)(x)$$

from which the inequality

$$\|\Pi_{\mathbb{T}_{Q^{\max}}}(F)(x)\| = \|\Pi_Q(F)(x)\| \leq \text{const.}\, 2^k$$

follows because of the construction of the tree.

If, however, $\alpha_Q \notin (\widetilde{\mathcal{P}})^{\min}$, then there must exist four time sibling tiles P_j, $0 \leq j \leq 3$, such that $\alpha_Q \subseteq \cup_{j=0}^3 P_j$, while

$$P_j \subseteq A_{\mathbb{T}_{Q^{\max}}}, \quad P_j \leq \alpha_Q, \qquad (0 \leq j \leq 3).$$

But, $\widetilde{Q} = \cup_{j=0}^3 P_j$ is a quartile such that $I_{\widetilde{Q}} = I_{\alpha_Q}$, and

$$\widetilde{Q} \leq Q \leq Q^{\max}, \qquad \widetilde{Q} \subseteq A_{\mathbb{T}_{Q^{\max}}},$$

so the convexity of the tree ensures that $\widetilde{Q} \in \mathbb{T}_{Q^{\max}}$, contradicting the minimality of Q among the quartiles in $\mathbb{T}_{Q^{\max}}$ whose time interval contains x. Thus $\alpha_Q \in (\widetilde{\mathcal{P}})^{\min}$.

On the other hand, if Q is indecomposable then, the four frequency sibling tiles that form Q belong to $(\widetilde{\mathcal{P}})^{\min}$. Moreover, because of the convexity of the tree, Q is a minimal quartile *in the tree*. If $x \in I_Q$, therefore, $\Pi_{\mathbb{T}_{Q^{\max}}}(F)(x) = \Pi_Q(F)(x)$ from which the the inequality for $\Pi_{\mathbb{T}_{Q^{\max}}}(F)$ follows as before. □

We can now prove the second of the two propositions needed in the proof of the Main Lemma.

(7.13) Proposition. *The inequality*

$$\|\mathcal{F}_{\mathbb{Q}_{k,i}}(G)\|_{L^2} \;\leq\; \text{const.}\, 2^k \, \|G\|_{L^2}$$

holds uniformly in k *and* i *for all* $G \in L^2(\mathbb{R}, \mathbb{R}^4)$.

Proof. It is enough to show that

$$\|\mathcal{T}_{Q^{\max}}(G)\|_{L^2} \;\leq\; \text{const.}\, 2^k \, \|G\|_{L^2} \tag{7.14}$$

holds for all $Q^{\max} \in \mathbb{Q}_{k,i}$ uniformly in k and i. For then

$$\|\mathcal{F}_{\mathbb{Q}_{k,i}}(G)\|_{L^2} = \| \sum_{Q^{\max}\in\mathbb{Q}_{k,i}} \mathcal{T}_{Q^{\max}}(G)\|_{L^2} \leq \text{ const.}\, 2^k \ \|G\|_{L^2}$$

because of the orthogonality properties of tree operators.

We split the tree $\mathbb{T}_{Q^{\max}}$ into two subsets and then sum separately over the two sets. Fix ξ in the frequency interval of $Q^{\max}$ and set

$$B_{Q^{\max}} = \{ Q \in \mathbb{T}_{Q^{\max}} : \xi \in b_Q \}, \qquad B'_{Q^{\max}} = \mathbb{T}_{Q^{\max}} \setminus B_{Q^{\max}}.$$

To determine the L^2-operator norm of

$$\mathcal{T}_{B_{Q^{\max}}}(G) = \sum_{Q\in B_{Q^{\max}}} \frac{1}{\sqrt{|I_Q|}} \langle\, F, \mathcal{R}(\omega_{a_Q})\,\rangle\, \langle\, G, \mathcal{R}(\omega_{b_Q})\,\rangle \mathcal{R}(\omega_{c_Q}).$$

note first that if Q, Q' belong to $B_{Q^{\max}}$ then $c_Q \cap c_{Q'} = \emptyset$. This follows from Lemma (5.2) because

$$Q, Q' \in B_{Q^{\max}}, \quad Q \cap Q' \neq \emptyset, \qquad \Longrightarrow \quad b_Q \cap b_{Q'} \neq \emptyset$$

Thus the exposed wavelet packets $\{\, \mathcal{R}(\omega_{c_Q}) \,:\, Q \in B_{Q^{\max}} \,\}$ form an orthogonal family (*cf.* (5.16)). Also we can replace F with the function $\Pi_{\mathbb{T}_{Q^{\max}}}(F)$ since the subspace of $L^2(\mathbb{R}, \mathbb{R}^4)$ associated with $A_{\mathbb{T}_{Q^{\max}}}$ contains all the $\{ \mathcal{R}(\omega_{a_Q}) : Q \in \mathbb{T}_{Q^{\max}} \}$. By Parseval's theorem, therefore,

$$\|\mathcal{T}_{B_{Q^{\max}}}(G)\|_{L^2} \ = \ \Big(\sum_{Q\in B_{Q^{\max}}} \frac{1}{|I_Q|}\, |\langle\, \Pi_{\mathbb{T}_{Q^{\max}}}(F),\, \mathcal{R}(\omega_{a_Q})\,\rangle|^2\, |\langle\, G,\, \mathcal{R}(\omega_{b_Q})\,\rangle|^2 \Big)^{1/2}.$$

The whole point of all the earlier combinatorics and operator decompositions is that we are left with an L^∞-function $\Pi_{\mathbb{T}_{Q^{\max}}}(F)$ such that

$$\|\Pi_{\mathbb{T}_{Q^{\max}}}(F)(x)\| \leq \text{const.}\, 2^k.$$

Now view $\mathbb{T}_b$ as a measure space with measure μ defined on each quartile Q by

$$\mu(\{Q\}) = |\langle\, \Pi_{\mathbb{T}_{Q^{\max}}}(F),\, \mathcal{R}(\omega_{a_Q})\rangle|^2.$$

Then

$$\|\mathcal{T}_{B_{Q^{\max}}}(G)\|_{L^2} \ = \Big(\int_0^\infty \mu(\{ Q : |\langle G, \mathcal{R}(\omega_{b_Q})\rangle|^2 \geq \lambda\, |I_Q| \})\, d\lambda \Big)^{1/2}.$$

But if Q is a quartile such that $|\langle\, G,\, \mathcal{R}(\omega_{b_Q})\rangle|^2 \geq \lambda\, |I_Q|$ then

$$I_Q \subseteq \{ x :\, M(G)^2(x) \geq \lambda \}$$

where M is the maximal function. Since the tiles a_Q are pairwise disjoint, therefore,

$$\mu(\{Q : |\langle G, \mathcal{R}(\omega_{b_Q})\rangle|^2 \geq \lambda\, |I_Q|\}) \;\leq\; \int_{\{x\,:\,M(G)^2(x)>\lambda\}} \|\Pi_{\mathbb{T}_{Q\max}}(F)(x)\|^2\, dx$$

$$\leq \text{const.}\, 2^{2k}\, |\{x :\; M(G)^2(x) \geq \lambda\,\}|.$$

Consequently,

$$\|\mathcal{T}_{B_{Q\max}}(G)\|_{L^2} \;\leq \text{const.}\, 2^k \Big(\int_0^\infty |\{x :\; M(G)^2(x) \geq \lambda\,\}|\, d\lambda\Big)^{1/2}$$

$$\leq\; \text{const.}\;\, 2^k \|M(G)\|_{L^2} \;\leq\; \text{const.}\;\, 2^k \|G\|_{L^2}.$$

In the second case, set

$$\mathcal{T}_{B'_{Q\max}}(G) = \sum_{Q\in B'_{Q\max}} \frac{1}{\sqrt{|I_Q|}} \langle\, F, \mathcal{R}(\omega_{a_Q})\,\rangle\, \langle\, G, \mathcal{R}(\omega_{b_Q})\,\rangle \mathcal{R}(\omega_{c_Q}).$$

But if Q, Q' belong to $B'_{Q\max}$, then $b_Q \cap b'_Q = \emptyset$ for the same reason as before and so the functions $\{\mathcal{R}(\omega_{b_Q}) :\; Q \in B'_{Q\max}\,\}$ form an orthonormal family. Thus the *adjoint* of $\mathcal{T}_{B'_{Q\max}}$ has the same structure as $\mathcal{T}_{B_{Q\max}}$ and its L^2-operator norm will satisfy the same estimate. On combining operator norm estimates we obtain (7.14), completing the proof. □

All that remains now is the proof of the Main Lemma (7.2).

Proof of Main Lemma. Given $\kappa > 0$ choose k_0 so that $2^{k_0} \leq \kappa < 2^{k_0+1}$; select also any $m > p' + \varepsilon$ where ε is the same as the one appearing in (7.4). For each fixed F we have shown already that $\mathcal{D}(F, G)$ admits a decomposition

$$\mathcal{D}(F, G)(x) = \sum_{k\in\mathbb{Z}} \sum_{i=0}^{\infty} \mathcal{F}_{\mathbb{Q}_{k,i}}(G),$$

in terms of forest operators $\mathcal{F}_{\mathbb{Q}_{k,i}}$ each of whose L^2-operator norm is essentially 2^k. We shall use the Carleson measure estimate to aid in summing over the forest operators.

Set

$$A = \{\, \mathbb{Q}_{k,i} :\; k < k_0,\; i < m(k_0 - k)\,\}$$

and let B be the set of all other forests. Then

$$\mathcal{D}(F, G)(x) = \sum_{\mathbb{Q}_{k,i}\in A} \mathcal{F}_{\mathbb{Q}_{k,i}}(G) + \sum_{\mathbb{Q}_{k,i}\in B} \mathcal{F}_{\mathbb{Q}_{k,i}}(G).$$

Now by construction,

$$\text{supp}(\mathcal{F}_{\mathbb{Q}_{k,i}}(G)) \;\subseteq\; \bigcup\{I_Q : Q \in \mathbb{Q}_{k,i}\,\} \;\subseteq\; \{x : N_{\mathbb{Q}_k}(x) \geq 2^i\,\}$$

where $N_{\mathbb{Q}_k}(x)$ counts the number of maximal quartiles in $\mathbb{Q}_k$ containing x. Thus for $\mathbb{Q}_{k,i} \in B$ the support of $\mathcal{F}_{\mathbb{Q}_{k,i}}(G)$ lies in the set

$$E_k = \begin{cases} \{x : N_{\mathbb{Q}_k}(x) \geq 1\}, & k > k_0, \\ \{x : N_{\mathbb{Q}_k}(x) \geq 2^{m(k_0-k)}\}, & k \leq k_0. \end{cases}$$

Now set $E = \cup_k E_k$. By Corollary (7.7),

$$|E_k| \leq \text{const.}\left(\frac{\|F\|_{L^p}}{2^k}\right)^p \qquad (k > k_0),$$

while

$$|E_k| \leq \text{const.}\left(\frac{\|F\|_{L^p}}{2^k 2^{m(k_0-k)/(p'+\varepsilon)}}\right)^p \qquad (k \leq k_0).$$

Thus the choice of k_0 and m ensure that

$$\begin{aligned} |E| \ &\leq \ \sum_k |E_k| = \sum_{k \leq k_0} |E_k| + \sum_{k > k_0} |E_k| \\ &\leq \text{const.}\left(\frac{\|F\|_{L^p}}{2^{k_0}}\right)^p \leq \text{const.}\left(\frac{\|F\|_{L^p}}{\kappa}\right)^p, \end{aligned}$$

establishing the first of the estimates in the Main lemma.

To establish the second, choose $G \in L^2 \cap L^q$ and observe first that

$$\left(\int_{\mathbb{R}\setminus E} \|\mathcal{D}(F,G)(x)\|^2 \, dx\right)^{1/2} \leq \sum_{\mathbb{Q}_{k,i} \in A} \|\mathcal{F}_{\mathbb{Q}_{k,i}}(G)\|_{L^2},$$

since on $\mathbb{R} \setminus E$ we find only the time intervals of quartiles in $\mathbb{Q}_{k,i} \in A$. But

$$\sum_{\mathbb{Q}_{k,i} \in A} \mathcal{F}_{\mathbb{Q}_{k,i}}(G) = \sum_{k \leq k_0} \sum_{i=0}^{m(k_0-k)} \mathcal{F}_{\mathbb{Q}_{k,i}}(G),$$

so by (7.13)

$$\left(\int_{\mathbb{R}\setminus E} \|\mathcal{D}(F,G)(x)\|^2 \, dx\right)^{1/2} \ \leq \ \text{const.}\, 2^{k_0} \|G\|_{L^2} \ \leq \ \text{const.}\, \kappa \, \|G\|_{L^2}.$$

On the other hand, the support of $\mathcal{F}_{\mathbb{Q}_{k,i}}(G)$ lies in the set $\{x : N_{\mathbb{Q}_k}(x) \geq 1\}$ whenever $\mathbb{Q}_{k,i}$ belongs to A, so

$$\mathcal{F}_{\mathbb{Q}_{k,i}}(G) = \mathcal{F}_{\mathbb{Q}_{k,i}}(G)\, \chi_{\{x : N_{\mathbb{Q}_k}(x) \geq 1\}} = \mathcal{F}_{\mathbb{Q}_{k,i}}(G \chi_{\{x : N_{\mathbb{Q}_k}(x) \geq 1\}}).$$

Thus, the function G in the right hand side of the last inequality can be replaced by $G\chi_{\{x:N_{\mathbb{Q}_k}(x)\geq 1\}}$. But, by (7.7) and Hölder's inequality,

$$\|G\,\chi_{\{x:N_{\mathbb{Q}_k}(x)\geq 1\}}\|_{L^2} \leq \|G\|_{L^q}\,|\{x : N_{\mathbb{Q}_k}(x) \geq 1\}|^{\frac{1}{2}-\frac{1}{q}}$$

$$\leq \text{const.}\,\|G\|_{L^q}\left(\frac{\|F\|_{L^p}}{2^k}\right)^{p\,(\frac{1}{2}-\frac{1}{q})} \leq \text{const.}\,\|G\|_{L^q}\left(\frac{\|F\|_{L^p}}{\kappa}\right)^{p\,(\frac{1}{2}-\frac{1}{q})}.$$

Hence

$$\left(\int_{\mathbb{R}\setminus E}\|\mathcal{D}(F,G)(x)\|^2\,dx\right)^{1/2} \leq \text{const.}\,\kappa\,\|G\|_{L^q}\left(\frac{\|F\|_{L^p}}{\kappa}\right)^{p\,(\frac{1}{2}-\frac{1}{q})}$$

from which the second estimate in the Main Lemma follows defining α so that

$$1-\alpha = p\left(\frac{1}{2}-\frac{1}{q}\right) = p\left(\frac{1}{2}-\frac{1}{r}+\frac{1}{p}\right).$$

This completes the proof of the Main Lemma, of (4.8) and of the Main Theorem as well. □

References

1. J. Bergh and J. Löfström, *Interpolation Spaces, An Introduction*, Springer-Verlag, Berlin-Heidelberg-New York, 1976.
2. J.-A. Chao, *Lusin area functions on local fields*, Pacific J. Math. **59** (1972), 383–390.
3. J.-A. Chao, *H^p-spaces of conjugate systems on local fields*, Studia Math. **49** (1974), 267–287.
4. J.-A. Chao, *Conjugate characterizations of H^1-spaces dyadic martingales*, Math. Ann. **240** (1979), 63–67.
5. J.-A. Chao and S. Janson, *A note on H^1 q-martingales*, Pacific J. Math. **97** (1981), 383–390.
6. J.-A. Chao and M.H. Taibleson, *A subregularity inequality of conjugate systems on local fields*, Studia Math. **46** (1973), 249–257.
7. R. R. Coifman, P. L. Lions, Y. Meyer and S. Semmes, *Compensated compactness and Hardy spaces*, Journal Pure Appl. **72** (1993), 247–286.
8. R. R. Coifman and Y. Meyer, *Au delà des opérateurs pseudo-differentiels*, Astérisque, Société Mathématique de France **57** (1978).
9. R. R. Coifman and Y. Meyer, *On commutators of singular integrals and bilinear singular integrals*, Trans. Amer. Math. Soc. **212** (1975), 315–331.
10. R. R. Coifman and Y. Meyer, *Nouvelles bases orthonormées de $L^2(\mathbb{R})$ ayant la structure du système de Walsh*, preprint, Yale University, New Haven, CT (1989).
11. R. R. Coifman, Y. Meyer and V. Wickerhauser, *Size properties of wavelet-packets*, Wavelets and Their Applications (M. B. Ruskai *et al.*, eds.), Jones and Bartlett, Boston, 1992, pp. 453–470.
12. M. Cwikel, *On* $(L_p(A_0), L_p(A_1))_{\theta q}$, Proc. Amer. Math. Soc. **44** (1974), 286–292.
13. C. Fefferman, *Pointwise convergence of Fourier series*, Annals of Math. **98** (1973), 551–571.
14. R.F. Gundy, *Martingale theory and pointwise convergence of certain orthogonal series*, Trans. Amer. Math. Soc. **124** (1966), 228–248.
15. E. Hernandez and G. Weiss, *A first course on wavelets*, CRC Press, Boca Raton, FL, 1996.
16. S. Janson, *Characterizations of H^1 by singular integral operators on martingales and $\mathbb{R}^n$*, Math. Scand. **41** (1977), 140–152.
17. S. Janson, *On interpolation of multi-linear operators*, Lecture Notes in Math., vol. 1302, Springer-Verlag, Berlin-Heidelberg-New York, 1988, pp. 290–302.
18. J.-P. Kahane and P. Lemarié-Rieusset, *Fourier Series and Wavelets*, Gordon and Breach, Amsterdam, 1995.

19. M. Lacey, and C. Thiele, *L^p estimates on the bilinear Hilbert transform,* $2 < p < \infty$, Annals of Math. **146** (1997), 683–724.
20. E.M. Stein, *Topics in harmonic analysis related to the Littlewood-Paley theory*, Ann. of Math. Studies No. 63, Princeton University Press, Princeton, NJ, 1970.
21. E.M. Stein, *Harmonic Analysis, Real variable methods, orthogonality, and oscillatory integrals*, Princeton University Press, Princeton, NJ, 1993.
22. C.M. Thiele, *Time-Frequency analysis in the Discrete Phase Plane*, Thesis, Yale University, 1995, (available at `http://www.math.yale.edu/pub/wavelets/papers/`).