REPRESENTATION OF BI-PARAMETER SINGULAR INTEGRALS BY DYADIC OPERATORS

HENRI MARTIKAINEN

ABSTRACT. We prove a dyadic representation theorem for bi-parameter singular integrals. That is, we represent certain bi-parameter operators as rapidly decaying averages of what we call bi-parameter shifts. A new version of the product space $T1$ theorem is established as a consequence.

1. INTRODUCTION

We study certain bi-parameter singular integrals T acting on some class of functions with product domain $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$. Our aim is to prove a representation theorem for them as an average of bi-parameter shifts S:

$$\langle Tf, g\rangle = C_T \mathbb{E}_{w_n}\mathbb{E}_{w_m} \sum_{\substack{(i_1,i_2)\in\mathbb{Z}^2_+ \\ (j_1,j_2)\in\mathbb{Z}^2_+}} 2^{-\max(i_1,i_2)\delta/2} 2^{-\max(j_1,j_2)\delta/2} \langle S^{i_1 i_2 j_1 j_2}_{\mathcal{D}_n\mathcal{D}_m} f, g\rangle.$$

Here the average is taken over all the dyadic grids $\mathcal{D}_n$ in $\mathbb{R}^n$ (parametrized by the random parameter w_n) and all the dyadic grids $\mathcal{D}_m$ in $\mathbb{R}^m$ (parametrized by the random parameter w_m). An exact formulation of everything is given after the introduction. Such a representation theorem exists for ordinary Calderón–Zygmund operators, and this was proven by Hytönen [7] in connection with the proof of the A_2 conjecture for general singular integrals.

In the one-parameter case such general representation theorems have already been utilized several times after [7]. The simplified proof of the A_2 conjecture by Hytönen, Pérez, Treil and Volberg [10] offered among other things a bit easier formulation of the representation theorem. In [8] the author together with Hytönen, Lacey, Orponen, Reguera, Sawyer and Uriarte–Tuero used the representation theorem to study sharp weak and strong type weighted bounds for maximal truncations $T_\#$. Modifying the metric randomization by Hytönen and the author [9] these representation theorems were lifted to the generality of metric spaces by Nazarov, Reznikov and Volberg [14]. Several other applications in the weighted context also already exist.

2010 *Mathematics Subject Classification.* 42B20.

Key words and phrases. Haar shift, bi-parameter singular integral, $T1$ theorem, non-homogeneous analysis.

The author is supported by the Academy of Finland through the project "L^p methods in harmonic analysis".

The reason why the representation theorem is so useful in the one-parameter case is that it can be used to reduce problems considering a general singular integral T into purely dyadic problems considering shifts only. Because of this, there is no particular reason why the applications should be limited to weighted questions. This just happens to be the case, since the representation theorem was originally developed for this purpose and is still very new a result. This is motivation enough for us to develop the analogous theory in the bi-parameter case. It would, of course, be interesting to study sharp weighted theory in the bi-parameter setting. Our theorem might be useful for this, however, it is a very difficult problem.

Regarding multi-parameter singular integrals, and multi-parameter harmonic analysis in general, there is a very large existing theory. After the classical $T1$ and Tb type theory by David and Journé [2] and David, Journé and Semmes [3], the first $T1$ type theorem for product spaces was proved by Journé [11]. Regarding other classical theory, we only mention the work of Chang and Fefferman [1], Fefferman [4] and Fefferman and Stein [5]. These three concern singular integrals and various spaces, like the BMO, on the product setting. There is a wide body of more recent developments of which we here only mention the papers by Ferguson and Lacey [6], Lacey and Metcalfe [12] and Muscalu, Pipher, Tao and Thiele [13]. These have to do with various multi-parameter paraproducts and characterizations for some product spaces. Some bi-parameter paraproducts appear also in our proof, and the product BMO space is thus important for us.

The classical multi-parameter singular integral theory of Journé [11] involves formulations written in the language of vector-valued Calderón–Zygmund theory. Very recently Pott and Villarroya [16] formulated and proved a new type of $T1$ theorem for product spaces. There such vector-valued formulations are replaced by several new mixed type conditions. Here we define our bi-parameter operators inspired by [16]. The conditions we use are not exactly the same. We, for example, do not work with smooth testing conditions. Establishing the correct shift structure is our primary task. However, we do get, as a by product, a pretty nice form of the product space $T1$ theorem.

In this paper we bring the superbly useful machinery of non-homogeneous analysis pioneered by Nazarov, Treil and Volberg (see for example [15]) to the context of bi-parameter theory. The use of non-homogeneous analysis gives additional decay for certain matrix elements involved in the expansion of $\langle Tf, g\rangle$. Just like in Hytönen's proof of the representation theorem for one-parameter singular integrals, the proof is a $T1$ style proof with ingredients from non-homogeneous analysis. In our case, we have to deal with the much added complexity of the bi-parameter situation. Indeed, there are more cases than in the one-parameter setting, and many of these are interesting mixed type phenomena. The non-homogeneous analysis makes this splitting into cases nicely transparent getting rid of rare geometric complications.

Acknowledgements. The author wishes to thank Tuomas Hytönen for suggesting this topic and for useful discussions.

2. Definitions, strategy and the main result

Structural assumptions. Let us formulate the Calderón–Zygmund structure of our operators. The basic assumption is that if $f = f_1 \otimes f_2$ (meaning $f(x) = f_1(x_1)f_2(x_2)$ for $x = (x_1, x_2)$) and $g = g_1 \otimes g_2$ with $f_1, g_1 \colon \mathbb{R}^n \to \mathbb{C}$, $f_2, g_2 \colon \mathbb{R}^m \to \mathbb{C}$, $\operatorname{spt} f_1 \cap \operatorname{spt} g_1 = \emptyset$ and $\operatorname{spt} f_2 \cap \operatorname{spt} g_2 = \emptyset$, then we have the kernel representation

$$\langle Tf, g\rangle = \int_{\mathbb{R}^{n+m}} \int_{\mathbb{R}^{n+m}} K(x,y) f(y) g(x)\, dx\, dy.$$

The kernel $K \colon (\mathbb{R}^{n+m} \times \mathbb{R}^{n+m}) \setminus \{(x,y) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} : x_1 = y_1 \text{ or } x_2 = y_2\} \to \mathbb{C}$ is assumed to satisfy the size condition

$$|K(x,y)| \le C \frac{1}{|x_1 - y_1|^n} \frac{1}{|x_2 - y_2|^m}$$

and the Hölder conditions

$$\begin{aligned} |K(x,y) - K(x,(y_1,y_2')) - K(x,(y_1',y_2)) + K(x,y')| \\ \le C \frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{|y_2 - y_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \end{aligned}$$

whenever $|y_1 - y_1'| \le |x_1 - y_1|/2$ and $|y_2 - y_2'| \le |x_2 - y_2|/2$,

$$\begin{aligned} |K(x,y) - K((x_1,x_2'),y) - K((x_1',x_2),y) + K(x',y)| \\ \le C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \end{aligned}$$

whenever $|x_1 - x_1'| \le |x_1 - y_1|/2$ and $|x_2 - x_2'| \le |x_2 - y_2|/2$,

$$\begin{aligned} |K(x,y) - K((x_1,x_2'),y) - K(x,(y_1',y_2)) + K((x_1,x_2'),(y_1',y_2))| \\ \le C \frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \end{aligned}$$

whenever $|y_1 - y_1'| \le |x_1 - y_1|/2$ and $|x_2 - x_2'| \le |x_2 - y_2|/2$, and

$$\begin{aligned} |K(x,y) - K(x,(y_1,y_2')) - K((x_1',x_2),y) + K((x_1',x_2),(y_1,y_2'))| \\ \le C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{|y_2 - y_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \end{aligned}$$

whenever $|x_1 - x_1'| \le |x_1 - y_1|/2$ and $|y_2 - y_2'| \le |x_2 - y_2|/2$.

Furthermore, we assume the mixed Hölder and size conditions

$$|K(x,y) - K((x_1',x_2),y)| \le C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{1}{|x_2 - y_2|^m}$$

whenever $|x_1 - x_1'| \le |x_1 - y_1|/2$,

$$|K(x,y) - K(x,(y_1', y_2))| \le C \frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{1}{|x_2 - y_2|^m}$$

whenever $|y_1 - y_1'| \le |x_1 - y_1|/2$,

$$|K(x,y) - K((x_1, x_2'), y)| \le C \frac{1}{|x_1 - y_1|^n} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{m+\delta}}$$

whenever $|x_2 - x_2'| \le |x_2 - y_2|/2$, and

$$|K(x,y) - K(x,(y_1, y_2'))| \le C \frac{1}{|x_1 - y_1|^n} \frac{|y_2 - y_2'|^\delta}{|x_2 - y_2|^{m+\delta}}$$

whenever $|y_2 - y_2'| \le |x_2 - y_2|/2$. We use, for minor convenience, ℓ^∞ metrics on $\mathbb{R}^n$ and $\mathbb{R}^m$.

We also need some Calderón–Zygmund structure on $\mathbb{R}^n$ and $\mathbb{R}^m$ separately. If $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ with $\operatorname{spt} f_1 \cap \operatorname{spt} g_1 = \emptyset$, then we assume the kernel representation

$$\langle Tf, g\rangle = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K_{f_2,g_2}(x_1,y_1) f_1(y_1) g_1(x_1)\, dx_1\, dy_1.$$

The kernel $K_{f_2,g_2} \colon (\mathbb{R}^n \times \mathbb{R}^n) \setminus \{(x_1,y_1) \in \mathbb{R}^n \times \mathbb{R}^n :\ x_1 = y_1\}$ is assumed to satisfy the size condition

$$|K_{f_2,g_2}(x_1,y_1)| \le C(f_2,g_2) \frac{1}{|x_1 - y_1|^n}$$

and the Hölder conditions

$$|K_{f_2,g_2}(x_1,y_1) - K_{f_2,g_2}(x_1',y_1)| \le C(f_2,g_2) \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n+\delta}}$$

whenever $|x_1 - x_1'| \le |x_1 - y_1|/2$, and

$$|K_{f_2,g_2}(x_1,y_1) - K_{f_2,g_2}(x_1,y_1')| \le C(f_2,g_2) \frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n+\delta}}$$

whenever $|y_1 - y_1'| \le |x_1 - y_1|/2$. Let $|A|$ denote the Lebesgue measure of a set A and χ_A be the characteristic function of A. We need the above representations and some control for $C(f_2,g_2)$ only in the diagonal in the following sense. For every cube $V \subset \mathbb{R}^m$ we assume that there holds $C(\chi_V, \chi_V) + C(\chi_V, u_V) + C(u_V, \chi_V) \le C|V|$, whenever u_V is such a function that $\operatorname{spt} u_V \subset V$, $|u_V| \le 1$ and $\int u_V = 0$. Functions u_V are called V-adapted with zero-mean (so V-adapted means just the first two conditions on the support and size). We also assume the analogous representation and properties with a kernel K_{f_1,g_1} in the case $\operatorname{spt} f_2 \cap \operatorname{spt} g_2 = \emptyset$.

Boundedness and cancellation assumptions. Define the partial adjoint T_1 of T by setting

$$\langle T_1(f_1 \otimes f_2), g_1 \otimes g_2\rangle = \langle T(g_1 \otimes f_2), f_1 \otimes g_2\rangle.$$

We assume that $T1, T^*1, T_1(1)$ and $T_1^*(1)$ belong to the product BMO on $\mathbb{R}^n \times \mathbb{R}^m$. We recall the definition of this space later in this section.

We assume that $|\langle T(\chi_K \otimes \chi_V), \chi_K \otimes \chi_V\rangle| \le C|K||V|$ for every cube $K \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$. This is the weak boundedness property for T.

We also assume the following diagonal BMO conditions: for every cube $K \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ and for every zero-mean functions a_K and b_V which are K and V adapted respectively (one has $\operatorname{spt} a_K \subset K$, $|a_K| \le 1$ and $\int a_K = 0$, and similarly for b_V):

(i) $|\langle T(a_K \otimes \chi_V), \chi_K \otimes \chi_V\rangle| \le C|K||V|$,
(ii) $|\langle T(\chi_K \otimes \chi_V), a_K \otimes \chi_V\rangle| \le C|K||V|$,
(iii) $|\langle T(\chi_K \otimes b_V), \chi_K \otimes \chi_V\rangle| \le C|K||V|$,
(iv) $|\langle T(\chi_K \otimes \chi_V), \chi_K \otimes b_V\rangle| \le C|K||V|$.

Haar functions. Let h_I be a L^2 normalized Haar function related to $I \in \mathcal{D}_n$, where $\mathcal{D}_n$ is a dyadic grid on $\mathbb{R}^n$. With this we mean that h_I, $I = I_1 \times \cdots \times I_n$, is one of the 2^n functions h_I^η, $\eta = (\eta_1, \ldots, \eta_n) \in \{0,1\}^n$, defined by

$$h_I^\eta = h_{I_1}^{\eta_1} \otimes \cdots \otimes h_{I_n}^{\eta_n},$$

where $h_{I_i}^0 = |I_i|^{-1/2}\chi_{I_i}$ and $h_{I_i}^1 = |I_i|^{-1/2}(\chi_{I_{i,l}} - \chi_{I_{i,r}})$ for every $i = 1, \ldots, n$. Here $I_{i,l}$ and $I_{i,r}$ are the left and right halves of the interval I_i respectively. If $\eta \ne 0$ the Haar function is cancellative: $\int h_I = 0$. All the cancellative Haar functions form an orthonormal basis of $L^2(\mathbb{R}^n)$. If $a \in L^2(\mathbb{R}^n)$ we may thus write $a = \sum_{I \in \mathcal{D}_n} \sum_{\eta \in \{0,1\}^n \setminus \{0\}} \langle a, h_I^\eta\rangle h_I^\eta$. However, we suppress the finite η summation and just write $a = \sum_I \langle a, h_I\rangle h_I$. Given a dyadic grid $\mathcal{D}_m$ on $\mathbb{R}^m$ and a cube $J \in \mathcal{D}_m$, we denote a L^2 normalized Haar function on J by u_J.

Product BMO on $\mathbb{R}^n \times \mathbb{R}^m$. Let us be given a dyadic grid $\mathcal{D}_n$ in $\mathbb{R}^n$ and a dyadic grid $\mathcal{D}_m$ in $\mathbb{R}^m$. We define the square function

$$S_{\mathcal{D}_n\mathcal{D}_m} f = \Big[\sum_{K \in \mathcal{D}_n} \sum_{V \in \mathcal{D}_m} |\langle f, h_K \otimes u_V\rangle|^2 \frac{\chi_K \otimes \chi_V}{|K||V|}\Big]^{1/2}.$$

Then the product Hardy space $H^1_{\mathcal{D}_n\mathcal{D}_m}(\mathbb{R}^n \times \mathbb{R}^m)$ consists of the locally integrable functions f with $\|f\|_{H^1_{\mathcal{D}_n\mathcal{D}_m}(\mathbb{R}^n \times \mathbb{R}^m)} = \|S_{\mathcal{D}_n\mathcal{D}_m} f\|_1 < \infty$. The dual of this space is the product BMO space $\mathrm{BMO}_{\mathcal{D}_n\mathcal{D}_m}(\mathbb{R}^n \times \mathbb{R}^m)$.

For us, the condition that $b \in \{T1, T^*1, T_1(1), T_1^*(1)\}$ is in the product BMO is defined to mean that $\|b\|_{\mathrm{BMO}_{\mathcal{D}_n\mathcal{D}_m}(\mathbb{R}^n \times \mathbb{R}^m)} \le C$ with every dyadic grid $\mathcal{D}_n$ in $\mathbb{R}^n$ and every dyadic grid $\mathcal{D}_m$ in $\mathbb{R}^m$.

Bi-parameter shifts. A bi-parameter shift on $\mathbb{R}^n \times \mathbb{R}^m$ is tied to a dyadic grid $\mathcal{D}_n$ on $\mathbb{R}^n$, a dyadic grid $\mathcal{D}_m$ on $\mathbb{R}^m$ and non-negative integers i_1, i_2, j_1, j_2. Such an operator is denoted by $S^{i_1 i_2 j_1 j_2}_{\mathcal{D}_n \mathcal{D}_m}$ and is of the form

$$S^{i_1 i_2 j_1 j_2}_{\mathcal{D}_n \mathcal{D}_m} f = \sum_{K \in \mathcal{D}_n} \sum_{V \in \mathcal{D}_m} A^{i_1 i_2 j_1 j_2}_{KV} f,$$

where

$$A^{i_1 i_2 j_1 j_2}_{KV} f = \sum_{\substack{I_1, I_2 \subset K \\ \ell(I_1) = 2^{-i_1}\ell(K) \\ \ell(I_2) = 2^{-i_2}\ell(K)}} \sum_{\substack{J_1, J_2 \subset V \\ \ell(J_1) = 2^{-j_1}\ell(V) \\ \ell(J_2) = 2^{-j_2}\ell(V)}} a_{I_1 I_2 K J_1 J_2 V} \langle f, h_{I_1} \otimes u_{J_1} \rangle h_{I_2} \otimes u_{J_2}$$

with

$$|a_{I_1 I_2 K J_1 J_2 V}| \le \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|} \frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|}.$$

Here, of course, $I_1, I_2 \in \mathcal{D}_n$ and $J_1, J_2 \in \mathcal{D}_m$, and $\ell(I)$ denotes the side length of a cube I. It is also required that all the subshifts

$$S^{i_1 i_2 j_1 j_2}_{\mathcal{A}\mathcal{B}} = \sum_{K \in \mathcal{A}} \sum_{V \in \mathcal{B}} A^{i_1 i_2 j_1 j_2}_{KV} f, \qquad \mathcal{A} \subset \mathcal{D}_n,\ \mathcal{B} \subset \mathcal{D}_m,$$

map $L^2(\mathbb{R}^n \times \mathbb{R}^m) \to L^2(\mathbb{R}^n \times \mathbb{R}^m)$ with norm at most one. If all of the Haar functions $h_{I_1}, h_{I_2}, u_{J_1}, u_{J_2}$ appearing are cancellative, the shift is called cancellative. Otherwise, it is called non-cancellative. The last requirement concerning the L^2 boundedness of all of the subshifts follows from the other conditions for cancellative shifts.

In practice, it is useful to observe that a bi-parameter shift S of type (i_1, i_2, j_1, j_2) related to some dyadic grids is simply of the form

$$\begin{aligned} Sf(x) = \sum_{K,V} A_{KV} f(x) &= \sum_{K,V} \frac{1}{|K \times V|} \int_{K \times V} K_{A_{KV}}(x,y) f(y)\,dy \\ &= \int_{\mathbb{R}^{n+m}} K_S(x,y) f(y)\,dy, \end{aligned}$$

where first of all spt $K_{A_{KV}} \subset (K \times V) \times (K \times V)$ and $|K_{A_{KV}}(x,y)| \le 1$. Moreover, $K_{A_{KV}}$ is constant with respect to x on dyadic rectangles $I \times J \subset K \times V$ for which $\ell(I) < 2^{-i_2}\ell(K)$ and $\ell(J) < 2^{-j_2}\ell(V)$, and $K_{A_{KV}}$ is constant with respect to y on dyadic rectangles $I \times J \subset K \times V$ for which $\ell(I) < 2^{-i_1}\ell(K)$ and $\ell(J) < 2^{-j_1}\ell(V)$. Note also that clearly

$$|K_S(x,y)| \le C \frac{1}{|x_1 - y_1|^n} \frac{1}{|x_2 - y_2|^m}.$$

2.1. **Random dyadic grids and the basic averaging formula.** Let $w_n = (w_n^i)_{i\in\mathbb{Z}}$ and $w_m = (w_m^j)_{j\in\mathbb{Z}}$, where $w_n^i \in \{0,1\}^n$ and $w_m^j \in \{0,1\}^m$. Let $\mathcal{D}_n^0$ and $\mathcal{D}_m^0$ be the standard dyadic grids on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. In $\mathbb{R}^n$ we define the new dyadic grid $\mathcal{D}_n = \{I + \sum_{i:\, 2^{-i}<\ell(I)} 2^{-i}w_n^i : \, I \in \mathcal{D}_n^0\} = \{I + w_n : \, I \in \mathcal{D}_n^0\}$, where we simply have defined $I + w_n := I + \sum_{i:\, 2^{-i}<\ell(I)} 2^{-i}w_n^i$. The dyadic grid $\mathcal{D}_m$ in $\mathbb{R}^m$ is similarly defined. There is a natural product probability structure on $(\{0,1\}^n)^{\mathbb{Z}}$ and $(\{0,1\}^m)^{\mathbb{Z}}$. So we have independent random dyadic grids $\mathcal{D}_n$ and $\mathcal{D}_m$ in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Even if $n = m$ we need two independent grids.

A cube $I \in \mathcal{D}_n$ is called bad if there exists $\tilde{I} \in \mathcal{D}_n$ so that $\ell(\tilde{I}) \geq 2^r\ell(I)$ and $d(I, \partial\tilde{I}) \leq 2\ell(I)^{\gamma_n}\ell(\tilde{I})^{1-\gamma_n}$. Here $\gamma_n = \delta/(2n+2\delta)$, where $\delta > 0$ appears in the kernel estimates. One notes that $\pi_{\text{good}}^n := \mathbb{P}_{w_n}(I + w_n \text{ is good})$ is independent of $I \in \mathcal{D}_n^0$. The parameter r is a fixed constant so that $\pi_{\text{good}}^n, \pi_{\text{good}}^m > 0$. Furthermore, it is important to note that for a fixed $I \in \mathcal{D}_n^0$ the set $I + w_n$ depends on w_n^i with $2^{-i} < \ell(I)$, while the goodness of $I + w_n$ depends on w_n^i with $2^{-i} \geq \ell(I)$. In particular, these notions are independent. Analogous definitions and remarks related to $\mathcal{D}_m$ hold.

We prove the basic averaging formula of Hytönen [7] but in the bi-parameter setting. This is the only part of the proof where probabilistic arguments are needed, and here independence plays a big role, even more so in the bi-parameter setting. We note that the functions f and g in this paper are always taken from some particularly nice dense subset of functions.

2.1. **Proposition.** *There holds*

$$\langle Tf, g\rangle = C\mathbb{E} \sum_{I_1,I_2\in\mathcal{D}_n} \sum_{J_1,J_2\in\mathcal{D}_m} \chi_{\textit{good}}(\textit{smaller}(I_1,I_2))\chi_{\textit{good}}(\textit{smaller}(J_1,J_2)) \langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes u_{J_2}\rangle \langle f, h_{I_1}\otimes u_{J_1}\rangle\langle g, h_{I_2}\otimes u_{J_2}\rangle,$$

where $\mathbb{E} = \mathbb{E}_{w_n}\mathbb{E}_{w_m}$ *and* $C = 1/(\pi_{\textit{good}}^n\pi_{\textit{good}}^m)$.

2.2. *Remark.* Here all the appearing Haar functions are, of course, cancellative and we recall that the finite summations over the 2^n-1 or 2^m-1 different cancellative Haar functions per cube are simply suppressed from the notation.

Proof. Define $\langle f, h_I\rangle_1(y) = \int f(x,y)h_I(x)\,dx$, $y \in \mathbb{R}^m$. We may write

$$f = \sum_{I_1\in\mathcal{D}_n} h_{I_1}\otimes\langle f, h_{I_1}\rangle_1 = \sum_{I_1\in\mathcal{D}_n^0} h_{I_1+w_n}\otimes\langle f, h_{I_1+w_n}\rangle_1$$

so that by independence

$$\begin{aligned}\langle Tf, g\rangle &= E_{w_n}\sum_{I_1\in\mathcal{D}_n^0}\langle T(h_{I_1+w_n}\otimes\langle f, h_{I_1+w_n}\rangle_1), g\rangle \\ &= \frac{1}{\pi_{\text{good}}^n}E_{w_n}\sum_{I_1\in\mathcal{D}_n^0}\chi_{\text{good}}(I_1+w_n)\langle T(h_{I_1+w_n}\otimes\langle f, h_{I_1+w_n}\rangle_1), g\rangle.\end{aligned}$$

After expanding g similarly as f above, one sees that this equals

$$\frac{1}{\pi^n_{\text{good}}} E_{w_n} \sum_{I_1, I_2 \in \mathcal{D}_n^0} \chi_{\text{good}}(I_1 + w_n) \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle$$

$$= \frac{1}{\pi^n_{\text{good}}} E_{w_n} \sum_{\substack{I_1, I_2 \in \mathcal{D}_n^0 \\ \ell(I_1) \le \ell(I_2)}} \chi_{\text{good}}(I_1 + w_n) \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle$$

$$+ E_{w_n} \sum_{\substack{I_1, I_2 \in \mathcal{D}_n^0 \\ \ell(I_1) > \ell(I_2)}} \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle.$$

Here we again used independence in the latter summation. Comparing to the trivial representation

$$\langle Tf, g\rangle = E_{w_n} \sum_{I_1, I_2 \in \mathcal{D}_n^0} \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle$$

we conclude that

$$\pi^n_{\text{good}} E_{w_n} \sum_{\substack{I_1, I_2 \in \mathcal{D}_n^0 \\ \ell(I_1) \le \ell(I_2)}} \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle$$

$$= E_{w_n} \sum_{\substack{I_1, I_2 \in \mathcal{D}_n^0 \\ \ell(I_1) \le \ell(I_2)}} \chi_{\text{good}}(I_1 + w_n) \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle.$$

First expanding g and proceeding like above one gets the symmetric formula

$$\pi^n_{\text{good}} E_{w_n} \sum_{\substack{I_1, I_2 \in \mathcal{D}_n^0 \\ \ell(I_2) < \ell(I_1)}} \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle$$

$$= E_{w_n} \sum_{\substack{I_1, I_2 \in \mathcal{D}_n^0 \\ \ell(I_2) < \ell(I_1)}} \chi_{\text{good}}(I_2 + w_n) \langle T(h_{I_1 + w_n} \otimes \langle f, h_{I_1+w_n}\rangle_1), h_{I_2+w_n} \otimes \langle g, h_{I_2+w_n}\rangle_1\rangle.$$

Splitting the trivial representation in to these two parts allows us to conclude that

$$\langle Tf, g\rangle = \frac{1}{\pi^n_{\text{good}}} E_{w_n} \sum_{I_1, I_2 \in \mathcal{D}_n} \chi_{\text{good}}(\text{smaller}(I_1, I_2)) \langle T(h_{I_1} \otimes \langle f, h_{I_1}\rangle_1), h_{I_2} \otimes \langle g, h_{I_2}\rangle_1\rangle.$$

We now expand on $\mathbb{R}^m$. One may write

$$\langle f, h_{I_1}\rangle_1 = \sum_{J_1 \in \mathcal{D}_m} \langle f, h_{I_1} \otimes u_{J_1}\rangle u_{J_1}$$

so that

$$h_{I_1} \otimes \langle f, h_{I_1}\rangle_1 = \sum_{J_1 \in \mathcal{D}_m} \langle f, h_{I_1} \otimes u_{J_1}\rangle h_{I_1} \otimes u_{J_1}.$$

We may then follow the recipe from above: insert this to the above formula for $\langle Tf, g\rangle$, add goodness to J_1 by independence, expand $h_{I_2} \otimes \langle g, h_{I_2}\rangle_1$, split the summation to $\ell(J_1) \le \ell(J_2)$ and $\ell(J_1) > \ell(J_2)$, remove the goodness from J_1 in the latter summation by independence and, finally, compare to the appropriate trivial identity. One also does the symmetric thing, where one first expands $h_{I_2} \otimes \langle g, h_{I_2}\rangle_1$ and adds the goodness to J_2. Combining these gives the claim of the proposition. □

2.3. *Remark.* One may also use full expansions like $f = \sum_{I_1 \in \mathcal{D}_n} \sum_{J_1 \in \mathcal{D}_m} \langle f, h_{I_1} \otimes u_{J_1}\rangle h_{I_1} \otimes u_{J_1}$ in the beginning of the proof. Following the usual trickery this leads to the formula

$$\langle Tf, g\rangle = \frac{1}{\pi^n_{\text{good}}} \mathbb{E} \sum_{I_1, I_2 \in \mathcal{D}_n} \sum_{J_1, J_2 \in \mathcal{D}_m} \chi_{\text{good}}(\text{smaller}(I_1, I_2)) \\ \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes u_{J_2}\rangle \langle f, h_{I_1} \otimes u_{J_1}\rangle \langle g, h_{I_2} \otimes u_{J_2}\rangle.$$

Here it may at first seem that there is no longer enough independence to add the goodness to J_1. However, one may simply write the summation as

$$\sum_{I_1, I_2 \in \mathcal{D}_n} \sum_{J_1 \in \mathcal{D}_m} \chi_{\text{good}}(\text{smaller}(I_1, I_2)) \langle T(h_{I_1} \otimes u_{J_1}), g_{I_2}\rangle \langle f, h_{I_1} \otimes u_{J_1}\rangle,$$

where one realizes that

$$g_{I_2} = \sum_{J_2 \in \mathcal{D}_m} \langle g, h_{I_2} \otimes u_{J_2}\rangle h_{I_2} \otimes u_{J_2} = h_{I_2} \otimes \langle g, h_{I_2}\rangle_1$$

does not depend on w_m. Then one may add the goodness to J_1 using independence and repeat the basic recipe to get the proposition.

Strategy and formulation of the main theorem. We fix the random variables w_n and w_m which fixes the dyadic grids $\mathcal{D}_n$ and $\mathcal{D}_m$ respectively. Then we study the summation

$$\sum_{\substack{\ell(I_1) \le \ell(I_2) \\ I_1 \text{ good}}} \sum_{\substack{\ell(J_1) \le \ell(J_2) \\ J_1 \text{ good}}} \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes u_{J_2}\rangle \langle f, h_{I_1} \otimes u_{J_1}\rangle \langle g, h_{I_2} \otimes u_{J_2}\rangle.$$

We more often than not suppress from the notation the important fact that I_1 and J_1 are good. Then we perform the splitting

$$\sum_{\ell(I_1) \le \ell(I_2)} = \sum_{\substack{\ell(I_1) \le \ell(I_2) \\ d(I_1, I_2) > \ell(I_1)^{\gamma_n} \ell(I_2)^{1-\gamma_n}}} + \sum_{I_1 \subsetneq I_2} + \sum_{I_1 = I_2} + \sum_{\substack{\ell(I_1) \le \ell(I_2) \\ d(I_1, I_2) \le \ell(I_1)^{\gamma_n} \ell(I_2)^{1-\gamma_n} \\ I_1 \cap I_2 = \emptyset}},$$

and similarly for the summation over the grid $\mathcal{D}_m$. Here $d(A, B)$ denotes the distance of the sets A and B (recall that we use the ℓ^∞ metric). The first sum is the separated sum, then we have the inside sum, the equal sum and the nearby sum. The summation over both the grids is split in to various types which also includes several mixed types. The list is: separated/separated, separated/inside, separated/equal, separated/nearby, inside/inside, inside/equal, inside/nearby, equal/equal, equal/nearby, nearby/nearby and some symmetric mixed sums. It seems reasonable to deal with these separately.

Note that actually the mixed sums where $\ell(I_1) \le \ell(I_2)$ and $\ell(J_1) > \ell(J_2)$ or $\ell(I_1) > \ell(I_2)$ and $\ell(J_1) \le \ell(J_2)$ are not completely symmetrical to this case. However, the relevant difference is only in the full paraproduct that appears in the corresponding inside/inside part. There one gets a bit different paraproducts, which are related to the assumptions that $T_1(1)$ and $T_1^*(1)$ belong to the product BMO of $\mathbb{R}^n \times \mathbb{R}^m$. We comment more on this on Remark 7.2.

The goal is to represent all of these different parts as a sum of shifts with a good decay factor in front. Combining all these cases together leads to our main theorem:

2.4. **Theorem.** *For a bi-parameter singular integral operator T as defined above, there holds for some bi-parameter shifts $S^{i_1 i_2 j_1 j_2}_{\mathcal{D}_n \mathcal{D}_m}$ that*

$$\langle Tf, g\rangle = C_T \mathbb{E}_{w_n}\mathbb{E}_{w_m} \sum_{\substack{(i_1,i_2)\in\mathbb{Z}^2_+ \\ (j_1,j_2)\in\mathbb{Z}^2_+}} 2^{-\max(i_1,i_2)\delta/2} 2^{-\max(j_1,j_2)\delta/2} \langle S^{i_1 i_2 j_1 j_2}_{\mathcal{D}_n \mathcal{D}_m} f, g\rangle,$$

where non-cancellative shifts may only appear if $(i_1, i_2) = (0, 0)$ or $(j_1, j_2) = (0, 0)$.

2.5. **Corollary.** *A bi-parameter singular integral T as defined above is L^2 bounded.*

We note that all of the appearing non-cancellative shifts will have a certain paraproduct structure, and this structure is explicit in the proof. For example in [8], where the one-parameter representation theorem is applied, it is important to know the explicit structure of the non-cancellative shifts.

The rest of the paper is dedicated to the piece by piece proof of this theorem. We use $X \lesssim Y$ to mean $X \le CY$ for some constant C and $X \sim Y$ to mean $Y \lesssim X \lesssim Y$. Of course, we cannot absorb just any constants, but only ones that depend on the dimensions or the various constants from the assumptions concerning T.

3. SEPARATED/SEPARATED

Let $I_1 \vee I_2 = \bigcap_{K\in\mathcal{D}_n,\, K\supset I_1\cup I_2} K$ and $J_1 \vee J_2 = \bigcap_{V\in\mathcal{D}_m,\, V\supset J_1\cup J_2} V$. The separation conditions together with goodness imply $\ell(I_1)^{\gamma_n}\ell(I_1\vee I_2)^{1-\gamma_n} \lesssim d(I_1, I_2)$ and $\ell(J_1)^{\gamma_m}\ell(J_1\vee J_2)^{1-\gamma_m} \lesssim d(J_1, J_2)$.

Let us write

$$\sum_{\substack{\ell(I_1)\le \ell(I_2) \\ d(I_1,I_2)>\ell(I_1)^{\gamma_n}\ell(I_2)^{1-\gamma_n}}} \sum_{\substack{\ell(J_1)\le \ell(J_2) \\ d(J_1,J_2)>\ell(J_1)^{\gamma_m}\ell(J_2)^{1-\gamma_m}}} = \sum_{\substack{i_2\ge 1\\ j_2\ge 1}} \sum_{\substack{i_1\ge i_2\\ j_1\ge j_2}} \sum_{\substack{K\in\mathcal{D}_n\\ V\in\mathcal{D}_m}} \sum_{\substack{d(I_1,I_2)>\ell(I_1)^{\gamma_n}\ell(I_2)^{1-\gamma_n}\\ I_1\vee I_2=K \\ \ell(I_1)=2^{-i_1}\ell(K),\, \ell(I_2)=2^{-i_2}\ell(K)}} \sum_{\substack{d(J_1,J_2)>\ell(J_1)^{\gamma_m}\ell(J_2)^{1-\gamma_m}\\ J_1\vee J_2=V \\ \ell(J_1)=2^{-j_1}\ell(V),\, \ell(J_2)=2^{-j_2}\ell(V)}}.$$

3.1. **Lemma.** *For I_1, I_2, J_1, J_2 in the above summation, we have the estimate*

$$\begin{aligned}|\langle T(h_{I_1}\otimes u_{J_1}), &h_{I_2}\otimes u_{J_2}\rangle| \\ &\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}\Big(\frac{\ell(J_1)}{\ell(V)}\Big)^{\delta/2} \\ &= 2^{-i_1\delta/2}\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}2^{-j_1\delta/2}\frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|}.\end{aligned}$$

Proof. Given a cube I we denote by c_I its center. We may write

$$\begin{aligned}&\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes u_{J_2}\rangle \\ &= \int_{I_1\times J_1}\int_{I_2\times J_2} K(x,y)h_{I_1}(y_1)u_{J_1}(y_2)h_{I_2}(x_1)u_{J_2}(x_2)\,dxdy,\end{aligned}$$

where we may, using cancellation, replace $K(x,y)$ by

$$K(x,y)-K(x,(y_1,c_{J_1}))-K(x,(c_{I_1},y_2))+K(x,(c_{I_1},c_{J_1})).$$

Since $|y_1-c_{I_1}|\le \ell(I_1)/2\le \frac{1}{2}\ell(I_1)^{\gamma_n}\ell(I_2)^{1-\gamma_n}\le d(I_1,I_2)/2\le |x_1-c_{I_1}|/2$ and similarly $|y_2-c_{J_1}|\le |x_2-c_{J_1}|/2$, we have

$$\begin{aligned}|K(x,y)-&K(x,(y_1,c_{J_1}))-K(x,(c_{I_1},y_2))+K(x,(c_{I_1},c_{J_1}))| \\ &\lesssim \frac{|y_1-c_{I_1}|^\delta}{|x_1-c_{I_1}|^{n+\delta}}\frac{|y_2-c_{J_1}|^\delta}{|x_2-c_{J_1}|^{m+\delta}} \\ &\lesssim \ell(I_1)^\delta d(I_1,I_2)^{-n-\delta}\ell(J_1)^\delta d(J_1,J_2)^{-m-\delta} \\ &\lesssim \ell(I_1)^\delta[\ell(I_1)^{\gamma_n}\ell(K)^{1-\gamma_n}]^{-n-\delta}\ell(J_1)^\delta[\ell(J_1)^{\gamma_m}\ell(V)^{1-\gamma_m}]^{-m-\delta} \\ &= \ell(I_1)^{\delta/2}\ell(K)^{-\delta/2}|K|^{-1}\ell(J_1)^{\delta/2}\ell(V)^{-\delta/2}|V|^{-1}.\end{aligned}$$

Here we used $\ell(I_1)^{\gamma_n}\ell(K)^{1-\gamma_n}\lesssim d(I_1,I_2)$ and $\gamma_n n+\gamma_n\delta=\delta/2$ (and the analogous estimates involving J_1, J_2, V and m). Recalling the L^2 normalization of the Haar functions and the fact that $\ell(I_1)/\ell(K)=2^{-i_1}$ and $\ell(J_1)/\ell(V)=2^{-j_1}$ completes the proof. □

We write

$$\begin{aligned}\langle T(h_{I_1}\otimes u_{J_1}),&h_{I_2}\otimes u_{J_2}\rangle\langle f, h_{I_1}\otimes u_{J_1}\rangle\langle g, h_{I_2}\otimes u_{J_2}\rangle \\ &= C2^{-i_1\delta/2}2^{-j_1\delta/2}\frac{\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes u_{J_2}\rangle}{C2^{-i_1\delta/2}2^{-j_1\delta/2}}\langle\langle f, h_{I_1}\otimes u_{J_1}\rangle h_{I_2}\otimes u_{J_2}, g\rangle.\end{aligned}$$

Define

$$a_{I_1 I_2 K J_1 J_2 V} = \frac{\langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes u_{J_2} \rangle}{C 2^{-i_1 \delta/2} 2^{-j_1 \delta/2}}$$

if all the various goodness and separation conditions appearing in the summations are satisfied, and otherwise set $a_{I_1 I_2 K J_1 J_2 V} = 0$. This enables us to write

$$\sum_{\substack{\ell(I_1) \le \ell(I_2) \\ d(I_1, I_2) > \ell(I_1)^{\gamma_n} \ell(I_2)^{1-\gamma_n}}} \sum_{\substack{\ell(J_1) \le \ell(J_2) \\ d(J_1, J_2) > \ell(J_1)^{\gamma_m} \ell(J_2)^{1-\gamma_m}}} \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes u_{J_2} \rangle \\ \langle f, h_{I_1} \otimes u_{J_1} \rangle \langle g, h_{I_2} \otimes u_{J_2} \rangle$$

in the form

$$C \sum_{\substack{i_2 \ge 1 \\ j_2 \ge 1}} \sum_{\substack{i_1 \ge i_2 \\ j_1 \ge j_2}} 2^{-i_1 \delta/2} 2^{-j_1 \delta/2} \sum_{K, V} \langle A_{KV}^{i_1 i_2 j_1 j_2} f, g \rangle,$$

where

$$A_{KV}^{i_1 i_2 j_1 j_2} f = \sum_{\substack{I_1, I_2 \subset K \\ \ell(I_1) = 2^{-i_1} \ell(K) \\ \ell(I_2) = 2^{-i_2} \ell(K)}} \sum_{\substack{J_1, J_2 \subset V \\ \ell(J_1) = 2^{-j_1} \ell(V) \\ \ell(J_2) = 2^{-j_2} \ell(V)}} a_{I_1 I_2 K J_1 J_2 V} \langle f, h_{I_1} \otimes u_{J_1} \rangle h_{I_2} \otimes u_{J_2}$$

with

$$|a_{I_1 I_2 K J_1 J_2 V}| \le \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \frac{|J_1|^{1/2} |J_2|^{1/2}}{|V|}.$$

The corresponding bi-parameter shift with indices i_1, i_2, j_1, j_2 is by definition

$$S^{i_1 i_2 j_1 j_2} f = \sum_{K,V} A_{KV}^{i_1 i_2 j_1 j_2} f.$$

4. Separated/inside

As $J_1 \subsetneq J_2$, there is a child $J_{2,1}$ of J_2 such that $J_1 \subset J_{2,1}$. We decompose

$$\langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes u_{J_2} \rangle = \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes s_{J_1 J2} \rangle \\ + \langle u_{J_2} \rangle_{J_1} \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes 1 \rangle,$$

where $s_{J_1 J_2} = \chi_{J_{2,1}^c} [u_{J_2} - \langle u_{J_2} \rangle_{J_{2,1}}]$. The relevant properties of $s_{J_1 J_2}$ are $|s_{J_1 J_2}| \le 2 |J_2|^{-1/2}$ and $\operatorname{spt} s_{J_1 J_2} \subset J_{2,1}^c$.

We write

$$\sum_{\substack{\ell(I_1) \le \ell(I_2) \\ d(I_1, I_2) > \ell(I_1)^{\gamma_n} \ell(I_2)^{1-\gamma_n}}} \sum_{J_1 \subsetneq J_2} \\ = \sum_{i_2 \ge 1} \sum_{i_1 \ge i_2} \sum_{j_1 \ge 1} \sum_{K \in \mathcal{D}_n} \sum_{J_2 \in \mathcal{D}_m} \sum_{\substack{d(I_1, I_2) > \ell(I_1)^{\gamma_n} \ell(I_2)^{1-\gamma_n} \\ I_1 \vee I_2 = K \\ \ell(I_1) = 2^{-i_1} \ell(K), \ell(I_2) = 2^{-i_2} \ell(K)}} \sum_{\substack{J_1 \subset J_2 \\ \ell(J_1) = 2^{-j_1} \ell(J_2)}}.$$

4.1. **Lemma.** *For I_1, I_2, J_1, J_2 in the above summation, we have the estimate*

$$\begin{aligned}|\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes s_{J_1J2}\rangle| \\ &\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\frac{|J_1|^{1/2}}{|J_2|^{1/2}}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}\Big(\frac{\ell(J_1)}{\ell(J_2)}\Big)^{\delta/2} \\ &= 2^{-i_1\delta/2}\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}2^{-j_1\delta/2}\frac{|J_1|^{1/2}}{|J_2|^{1/2}}.\end{aligned}$$

Proof. There is good separation by the goodness of J_1 if $\ell(J_1) < 2^{-r}\ell(J_2)$. Indeed, in this case there holds $d(J_1, J_{2,1}^c) \ge 2\ell(J_1)^{\gamma_m}\ell(J_{2,1})^{1-\gamma_m} \ge \ell(J_1)^{\gamma_m}\ell(J_2)^{1-\gamma_m}$. Then we may write

$$\begin{aligned}\langle T(h_{I_1}\otimes u_{J_1}),&h_{I_2}\otimes s_{J_1J2}\rangle \\ &= \int_{I_1\times J_1}\int_{I_2\times J_{2,1}^c} K(x,y)h_{I_1}(y_1)u_{J_1}(y_2)h_{I_2}(x_1)s_{J_1J_2}(x_2)\,dx\,dy,\end{aligned}$$

and replace $K(x,y)$ by $K(x,y) - K(x,(y_1,c_{J_1})) - K(x,(c_{I_1},y_2)) + K(x,(c_{I_1},c_{J_1}))$ using the cancellation of u_{J_1} and h_{I_1}. We may utilize the kernel estimates to get

$$\begin{aligned}|K(x,y) - K(x,(y_1,c_{J_1}))-&K(x,(c_{I_1},y_2)) + K(x,(c_{I_1},c_{J_1}))| \\ &\lesssim \ell(I_1)^{\delta/2}\ell(K)^{-\delta/2}|K|^{-1}\ell(J_1)^{\delta}\frac{1}{|x_2-c_{J_1}|^{m+\delta}}.\end{aligned}$$

This yields

$$\begin{aligned}|\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes s_{J_1J2}\rangle| \\ &\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}\frac{|J_1|^{1/2}}{|J_2|^{1/2}}\ell(J_1)^{\delta}\int_{J_{2,1}^c}\frac{dx_2}{|x_2-c_{J_1}|^{m+\delta}},\end{aligned}$$

where

$$\begin{aligned}\int_{J_{2,1}^c}\frac{dx_2}{|x_2-c_{J_1}|^{m+\delta}} &\lesssim \int_{\mathbb{R}^m\setminus B(c_{J_1},d(J_1,J_{2,1}^c))}\frac{dx_2}{|x_2-c_{J_1}|^{m+\delta}} \\ &\lesssim d(J_1,J_{2,1}^c)^{-\delta} \lesssim \ell(J_1)^{-\delta/2}\ell(J_2)^{-\delta/2}.\end{aligned}$$

Therefore, we have

$$\begin{aligned}|\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes s_{J_1J2}\rangle| \\ &\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}\frac{|J_1|^{1/2}}{|J_2|^{1/2}}\Big(\frac{\ell(J_1)}{\ell(J_2)}\Big)^{\delta/2}.\end{aligned}$$

We still need to deal with the case $2^{-r}\ell(J_2) \le \ell(J_1)(\le \ell(J_2))$. This time we split

$$\begin{aligned}\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes s_{J_1J2}\rangle &= \langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes(\chi_{3J_1}s_{J_1J2})\rangle \\ &\quad + \langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes(\chi_{(3J_1)^c}s_{J_1J2})\rangle.\end{aligned}$$

We have that $\langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes (\chi_{3J_1} s_{J_1 J2})\rangle$ equals

$$\int_{I_1 \times J_1} \int_{I_2 \times (3J_1 \setminus J_{2,1})} [K(x,y) - K(x,(c_{I_1}, y_2))] h_{I_1}(y_1) u_{J_1}(y_2) h_{I_2}(x_1) s_{J_1 J_2}(x_2)\, dx\, dy$$

so we can estimate using the mixed Hölder and size estimate that

$$\begin{aligned}
&|\langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes (\chi_{3J_1} s_{J_1 J2})\rangle| \\
&\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} |J_1|^{-1/2}|J_2|^{-1/2} \int_{J_1}\int_{3J_1 \setminus J_1} \frac{1}{|x_2 - y_2|^m}\, dx_2\, dy_2 \\
&\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} \frac{|J_1|^{1/2}}{|J_2|^{1/2}} \\
&\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} \frac{|J_1|^{1/2}}{|J_2|^{1/2}} \Big(\frac{\ell(J_1)}{\ell(J_2)}\Big)^{\delta/2}.
\end{aligned}$$

In the term $\langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes (\chi_{(3J_1)^c} s_{J_1 J2})\rangle$ we have good separation everywhere, so the Hölder estimate for K yields

$$\begin{aligned}
|\langle T(h_{I_1} &\otimes u_{J_1}), h_{I_2} \otimes (\chi_{(3J_1)^c} s_{J_1 J2})\rangle| \\
&\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} \frac{|J_1|^{1/2}}{|J_2|^{1/2}} \ell(J_1)^{\delta} \int_{(3J_1)^c} \frac{dx_2}{|x_2 - c_{J_1}|^{m+\delta}} \\
&\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} \frac{|J_1|^{1/2}}{|J_2|^{1/2}} \\
&\lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} \frac{|J_1|^{1/2}}{|J_2|^{1/2}} \Big(\frac{\ell(J_1)}{\ell(J_2)}\Big)^{\delta/2}.
\end{aligned}$$

□

The above lemma enables us to write

$$\sum_{\substack{\ell(I_1) \le \ell(I_2) \\ d(I_1, I_2) > \ell(I_1)^{\gamma_n} \ell(I_2)^{1-\gamma_n}}} \sum_{J_1 \subsetneq J_2} \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes s_{J_1 J2}\rangle \langle f, h_{I_1} \otimes u_{J_1}\rangle \langle g, h_{I_2} \otimes u_{J_2}\rangle$$

in the form

$$C \sum_{i_2 \ge 1} \sum_{i_1 \ge i_2} \sum_{j_1 \ge 1} 2^{-i_1 \delta/2} 2^{-j_1 \delta/2} \langle S^{i_1 i_2 j_1 0} f, g\rangle.$$

Next, we deal with the series with the term $\langle u_{J_2}\rangle_{J_1} \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes 1\rangle$. This will yield shifts of the type $(i_1, i_2, 0, 0)$ which are non-cancellative (their $\mathbb{R}^m$ parts are paraproducts in a certain sense). As these shifts will be non-cancellative, we will also have to worry about their L^2 boundedness properties.

Write

$$
\begin{aligned}
&\sum_{J_1 \subsetneq J_2} \langle u_{J_2}\rangle_{J_1} \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes 1\rangle \langle f, h_{I_1} \otimes u_{J_1}\rangle \langle g, h_{I_2} \otimes u_{J_2}\rangle \\
&= \sum_{J_1} \Big\langle \sum_{J_2} \langle g, h_{I_2} \otimes u_{J_2}\rangle u_{J_2} \Big\rangle_{J_1} \langle T(h_{I_1} \otimes u_{J_1}), h_{I_2} \otimes 1\rangle \langle f, h_{I_1} \otimes u_{J_1}\rangle \\
&= \sum_{V} \langle \langle g, h_{I_2}\rangle_1 \rangle_V \langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes 1\rangle \langle f, h_{I_1} \otimes u_V\rangle.
\end{aligned}
$$

The summands can further be written in the form

$$
|V|^{-1/2} \langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes 1\rangle \langle \langle f, h_{I_1} \otimes u_V\rangle h_{I_2} \otimes u_V^0, g\rangle,
$$

where $u_V^0 = |V|^{-1/2}\chi_V$. Written in this way it is evident that we will have the required shift structure of the type $(i_1, i_2, 0, 0)$.

4.2. **Lemma.** *The correct normalization*

$$
|\langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes 1\rangle| \lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|} \Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} |V|^{1/2}
$$

holds.

Proof. Let us first split

$$
\langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes 1\rangle = \langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes \chi_{3V}\rangle + \langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes \chi_{(3V)^c}\rangle.
$$

We have

$$
\begin{aligned}
|\langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes \chi_{3V}\rangle| \le |V|^{-1/2} \sum_{V' \in \mathrm{ch}(V)} \Big[&|\langle T(h_{I_1} \otimes \chi_{V'}), h_{I_2} \otimes \chi_{3V\setminus V'}\rangle| \\
&+ |\langle T(h_{I_1} \otimes \chi_{V'}), h_{I_2} \otimes \chi_{V'}\rangle| \Big]
\end{aligned}
$$

For the first time, we use the kernel representations in $\mathbb{R}^n$ to write $\langle T(h_{I_1} \otimes \chi_{V'}), h_{I_2} \otimes \chi_{V'}\rangle$ in the form

$$
\int_{I_1} \int_{I_2} [K_{\chi_{V'},\chi_{V'}}(x_1, y_1) - K_{\chi_{V'},\chi_{V'}}(x_1, c_{I_1})] h_{I_1}(y_1) h_{I_2}(x_1)\, dx_1\, dy_1.
$$

This gives that

$$
\begin{aligned}
|\langle T(h_{I_1} \otimes \chi_{V'}), h_{I_2} \otimes \chi_{V'}\rangle| &\le C(\chi_{V'}, \chi_{V'}) \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|} \Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} \\
&\lesssim |V| \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|} \Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}.
\end{aligned}
$$

Notice that by the mixed Hölder and size estimates for K we have the same bound also for the term $|\langle T(h_{I_1} \otimes \chi_{V'}), h_{I_2} \otimes \chi_{3V\setminus V'}\rangle|$, and so there holds

$$
|\langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes \chi_{3V}\rangle| \lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|} \Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2} |V|^{1/2}.
$$

The term $\langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes \chi_{(3V)^c}\rangle$ is in control by the full kernel representation and the Hölder estimate for K. □

These are non-cancellative shifts so we must separately demonstrate the L^2 boundedness. For this, we prefer to write things in a different way:

$$\begin{aligned}
&\sum_V \langle\langle g, h_{I_2}\rangle_1\rangle_V \langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes 1\rangle\langle f, h_{I_1} \otimes u_V\rangle \\
&= \sum_V \langle\langle g, h_{I_2}\rangle_1\rangle_V \langle\langle T^*(h_{I_2} \otimes 1), h_{I_1}\rangle_1, u_V\rangle\langle\langle f, h_{I_1}\rangle_1, u_V\rangle \\
&= C2^{-i_1\delta/2}\Big\langle \langle f, h_{I_1}\rangle_1, \sum_V \langle\langle g, h_{I_2}\rangle_1\rangle_V \langle b_{I_1I_2}, u_V\rangle u_V\Big\rangle \\
&= C2^{-i_1\delta/2}\langle\langle f, h_{I_1}\rangle_1, \Pi_{b_{I_1I_2}}(\langle g, h_{I_2}\rangle_1)\rangle \\
&= C2^{-i_1\delta/2}\langle \Pi^*_{b_{I_1I_2}}(\langle f, h_{I_1}\rangle_1), \langle g, h_{I_2}\rangle_1\rangle \\
&= C2^{-i_1\delta/2}\langle h_{I_2} \otimes \Pi^*_{b_{I_1I_2}}(\langle f, h_{I_1}\rangle_1), g\rangle,
\end{aligned}$$

where $b_{I_1I_2} = \langle T^*(h_{I_2} \otimes 1), h_{I_1}\rangle_1/(C2^{-i_1\delta/2})$ and $\Pi_{b_{I_1I_2}}$ is the related paraproduct on $\mathbb{R}^m$ defined by the general formula

$$\Pi_b a = \sum_V \langle a\rangle_V \langle b, u_V\rangle u_V.$$

4.3. **Lemma.** *We have* $b_{I_1I_2} \in \mathrm{BMO}(\mathbb{R}^m)$ *with the bound*

$$\|b_{I_1I_2}\|_{\mathrm{BMO}(\mathbb{R}^m)} \le c\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}.$$

Proof. Let V be any cube in $\mathbb{R}^m$ and a be any function in $\mathbb{R}^m$ such that $\operatorname{spt} a \subset V$, $|a| \le 1$ and $\int a = 0$. It suffices to show that

$$|\langle T(h_{I_1} \otimes a), h_{I_2} \otimes 1\rangle| \lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}|V|.$$

This is done by splitting $1 = \chi_{3V} + \chi_{(3V)^c}$ and using kernel estimates in a similar fashion as before. □

4.4. *Remark.* The strengthening of Lemma 4.2 to the related BMO estimate of Lemma 4.3 requires one to have the control $C(u_V, \chi_V) \le C|V|$ for V-adapted functions u_V with zero-mean. It is precisely for these type of BMO reasons that merely the assumption $C(\chi_V, \chi_V) \le C|V|$ does not seem to be enough for the results of this paper.

Let us abbreviate

$$\sum_{\substack{I_1,\, I_2 \subset K \\ \ell(I_1)=2^{-i_1}\ell(K),\, \ell(I_2)=2^{-i_2}\ell(K)}} = \sum_{I_1,\, I_2 \subset K}^{(i_1,i_2)}.$$

We are ready to show the boundedness of our non-cancellative shifts of type $(i_1, i_2, 0, 0)$.

4.5. **Proposition.** *There holds*

$$\Big\| \sum_K \sum_{I_1,\, I_2 \subset K}^{(i_1,i_2)} h_{I_2} \otimes \Pi^*_{b_{I_1 I_2}}(\langle f, h_{I_1}\rangle_1) \Big\|_2 \le \|f\|_2.$$

Proof. There holds by orthogonality that

$$\begin{aligned} \Big\| \sum_K \sum_{I_1,\, I_2 \subset K}^{(i_1,i_2)} & h_{I_2} \otimes \Pi^*_{b_{I_1 I_2}}(\langle f, h_{I_1}\rangle_1) \Big\|_2^2 \\ &= \sum_K \sum_{I_2 \subset K}^{(i_2)} \Big\| \sum_{I_1 \subset K}^{(i_1)} \Pi^*_{b_{I_1 I_2}}(\langle f, h_{I_1}\rangle_1) \Big\|_2^2 \\ &\le \sum_K \sum_{I_2 \subset K}^{(i_2)} \Big(\sum_{I_1 \subset K}^{(i_1)} \|\Pi^*_{b_{I_1 I_2}}(\langle f, h_{I_1}\rangle_1)\|_2 \Big)^2. \end{aligned}$$

Let $p_K^{i_1}$ be the orthogonal projection from $L^2(\mathbb{R}^n)$ to $\operatorname{span}\{h_{I_1} : I_1 \subset K,\, \ell(I_1) = 2^{-i_1}\ell(K)\}$. Write also $f_y(x) = f(x,y)$. There holds by the boundedness of paraproducts defined by BMO functions and the previous lemma that

$$\begin{aligned} \|\Pi^*_{b_{I_1 I_2}}(\langle f, h_{I_1}\rangle_1\|_2 &\le \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|} \|\langle f, h_{I_1}\rangle_1\|_2 \\ &\le \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|} \Big(\int_{\mathbb{R}^m} \int_{I_1} |p_K^{i_1} f_y(x)|^2 \,dx\,dy \Big)^{1/2}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &\Big\| \sum_K \sum_{I_1,\, I_2 \subset K}^{(i_1,i_2)} h_{I_2} \otimes \Pi^*_{b_{I_1 I_2}}(\langle f, h_{I_1}\rangle_1) \Big\|_2^2 \\ &\le \sum_K \frac{1}{|K|} \Big(\sum_{I_1 \subset K}^{(i_1)} |I_1|^{1/2} \Big(\int_{\mathbb{R}^m} \int_{I_1} |p_K^{i_1} f_y(x)|^2 \,dx\,dy \Big)^{1/2} \Big)^2 \\ &\le \sum_K \frac{1}{|K|} \Big(\sum_{I_1 \subset K}^{(i_1)} |I_1| \Big) \Big(\sum_{I_1 \subset K}^{(i_1)} \int_{\mathbb{R}^m} \int_{I_1} |p_K^{i_1} f_y(x)|^2 \,dx\,dy \Big) \\ &\le \sum_K \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} |p_K^{i_1} f_y(x)|^2 \,dx\,dy \\ &= \int_{\mathbb{R}^m} \|f_y\|_2^2 \,dy = \|f\|_2^2, \end{aligned}$$

where we again utilized orthogonality. □

We end this this section by concluding that

$$\sum_{\substack{\ell(I_1)\le\ell(I_2)\\ d(I_1,I_2)>\ell(I_1)^{\gamma_n}\ell(I_2)^{1-\gamma_n}}} \sum_{J_1\subsetneq J_2} \langle u_{J_2}\rangle_{J_1}\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes 1\rangle\langle f, h_{I_1}\otimes u_{J_1}\rangle\langle g, h_{I_2}\otimes u_{J_2}\rangle$$
$$= C\sum_{i_2\ge 1}\sum_{i_1\ge i_2} 2^{-i_1\delta/2}\langle S^{i_1 i_2 00} f, g\rangle.$$

5. Separated/equal

There holds that

$$|\langle T(h_{I_1}\otimes u_V), h_{I_2}\otimes u_V\rangle| \lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}.$$

Indeed, to see this, first estimate

$$|\langle T(h_{I_1}\otimes u_V), h_{I_2}\otimes u_V\rangle| \le |V|^{-1}\Big[\sum_{\substack{V',V''\in \mathrm{ch}(V)\\ V'\ne V''}} |\langle T(h_{I_1}\otimes \chi_{V'}), h_{I_2}\otimes \chi_{V''}\rangle|$$
$$+ \sum_{V'\in\mathrm{ch}(V)} |\langle T(h_{I_1}\otimes \chi_{V'}), h_{I_2}\otimes \chi_{V'}\rangle|\Big].$$

We have by the kernel representation in $\mathbb{R}^n$ that

$$|\langle T(h_{I_1}\otimes \chi_{V'}), h_{I_2}\otimes \chi_{V'}\rangle| \le C(\chi_{V'},\chi_{V'})\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}$$
$$\lesssim |V|\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}.$$

For $V'\ne V''$ the estimate

$$|\langle T(h_{I_1}\otimes \chi_{V'}), h_{I_2}\otimes \chi_{V''}\rangle| \lesssim |V|\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}$$

follows from the full kernel representation using the mixed Hölder and size estimate of K.

We may thus immediately write that

$$\sum_{\substack{\ell(I_1)\le\ell(I_2)\\ d(I_1,I_2)>\ell(I_1)^{\gamma_n}\ell(I_2)^{1-\gamma_n}}} \sum_{V}\langle T(h_{I_1}\otimes u_V), h_{I_2}\otimes u_V\rangle\langle f, h_{I_1}\otimes u_V\rangle\langle g, h_{I_2}\otimes u_V\rangle$$
$$= C\sum_{i_2\ge 1}\sum_{i_1\ge i_2} 2^{-i_1\delta/2}\langle S^{i_1 i_2 00} f, g\rangle,$$

where in this case $S^{i_1 i_2 00}$ are cancellative shifts.

6. SEPARATED/NEARBY

For the J_1 and J_2 in the nearby summation it is evident that $V = J_1 \vee J_2$ satisfies $\ell(V) \leq 2^r \ell(J_1)$. Thus, we may write

$$\sum_{\substack{\ell(I_1)\leq\ell(I_2) \\ d(I_1,I_2)>\ell(I_1)^{\gamma_n}\ell(I_2)^{1-\gamma_n}}} \sum_{\substack{\ell(J_1)\leq\ell(J_2) \\ d(J_1,J_2)\leq\ell(J_1)^{\gamma_m}\ell(J_2)^{1-\gamma_m} \\ J_1\cap J_2=\emptyset}}$$

$$= \sum_{i_2\geq 1}\sum_{i_1\geq i_2}\sum_{j_1=1}^{r}\sum_{j_2=1}^{j_1}\sum_K\sum_V \sum_{\substack{d(I_1,I_2)>\ell(I_1)^{\gamma_n}\ell(I_2)^{1-\gamma_n} \\ I_1\vee I_2=K \\ \ell(I_1)=2^{-i_1}\ell(K),\, \ell(I_2)=2^{-i_2}\ell(K)}} \sum_{\substack{d(J_1,J_2)\leq\ell(J_1)^{\gamma_m}\ell(J_2)^{1-\gamma_m},\, J_1\cap J_2=\emptyset \\ J_1\vee J_2=V \\ \ell(J_1)=2^{-j_1}\ell(V),\, \ell(J_2)=2^{-j_2}\ell(V)}}.$$

It is easy to get the required estimate

$$|\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes u_{J_2}\rangle| \lesssim \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\Big(\frac{\ell(I_1)}{\ell(K)}\Big)^{\delta/2}$$

by using the full kernel representation and the mixed Hölder and size estimate of K. Therefore, we are able to realize this part in the form

$$C\sum_{i_2\geq 1}\sum_{i_1\geq i_2}\sum_{j_1=1}^{r}\sum_{j_2=1}^{j_1} 2^{-i_1\delta/2}2^{-j_1\delta/2}\langle S^{i_1i_2j_1j_2}f, g\rangle.$$

7. INSIDE/INSIDE

We decompose

$$\begin{aligned}\langle T(h_{I_1}\otimes u_{J_1}), h_{I_2}\otimes u_{J_2}\rangle &= \langle T(h_{I_1}\otimes u_{J_1}), s_{I_2I_2}\otimes s_{J_1J_2}\rangle \\ &\quad + \langle u_{J_2}\rangle_{J_1}\langle T(h_{I_1}\otimes u_{J_1}), s_{I_2I_2}\otimes 1\rangle \\ &\quad + \langle h_{I_2}\rangle_{I_1}\langle T(h_{I_1}\otimes u_{J_1}), 1\otimes s_{J_1J_2}\rangle \\ &\quad + \langle h_{I_2}\rangle_{I_1}\langle u_{J_2}\rangle_{J_1}\langle T(h_{I_1}\otimes u_{J_1}), 1\rangle,\end{aligned}$$

where $s_{I_1I_2} = \chi_{I_{2,1}^c}(h_{I_2} - \langle h_{I_2}\rangle_{I_{2,1}})$ and $s_{J_1J_2} = \chi_{J_{2,1}^c}[u_{J_2} - \langle u_{J_2}\rangle_{J_{2,1}}]$. The relevant properties are spt $s_{I_1I_2} \subset I_{2,1}^c$, spt $s_{J_1J_2} \subset J_{2,1}^c$, $|s_{I_1I_2}| \leq 2|I_2|^{-1/2}$ and $|s_{J_1J_2}| \leq 2|J_2|^{-1/2}$.

7.1. **Lemma.** *There holds*

$$|\langle T(h_{I_1}\otimes u_{J_1}), s_{I_2I_2}\otimes s_{J_1J_2}\rangle| \lesssim \frac{|I_1|^{1/2}}{|I_2|^{1/2}}\Big(\frac{\ell(I_1)}{\ell(I_2)}\Big)^{\delta/2}\frac{|J_1|^{1/2}}{|J_2|^{1/2}}\Big(\frac{\ell(J_1)}{\ell(J_2)}\Big)^{\delta/2}.$$

Proof. In the case that $\ell(I_1) < 2^{-r}\ell(I_2)$ and $\ell(J_1) < 2^{-r}\ell(J_2)$ one may use the Hölder estimate of K. In the case $2^{-r}\ell(I_2) \leq \ell(I_1)(\leq \ell(I_2))$ and $2^{-r}\ell(J_2) \leq \ell(J_1)(\leq$

$\ell(J_2))$ one splits

$$\begin{aligned}\langle T(h_{I_1}\otimes u_{J_1}), s_{I_2I_2}\otimes s_{J_1J_2}\rangle &= \langle T(h_{I_1}\otimes u_{J_1}), (\chi_{3I_1}s_{I_2I_2})\otimes(\chi_{3J_1}s_{J_1J_2})\rangle \\ &+ \langle T(h_{I_1}\otimes u_{J_1}), (\chi_{3I_1}s_{I_2I_2})\otimes(\chi_{(3J_1)^c}s_{J_1J_2})\rangle \\ &+ \langle T(h_{I_1}\otimes u_{J_1}), (\chi_{(3I_1)^c}s_{I_2I_2})\otimes(\chi_{3J_1}s_{J_1J_2})\rangle \\ &+ \langle T(h_{I_1}\otimes u_{J_1}), (\chi_{(3I_1)^c}s_{I_2I_2})\otimes(\chi_{(3J_1)^c}s_{J_1J_2})\rangle.\end{aligned}$$

The first term is controlled by the size estimate of the full kernel:

$$\begin{aligned}&|\langle T(h_{I_1}\otimes u_{J_1}), (\chi_{3I_1}s_{I_2I_2})\otimes(\chi_{3J_1}s_{J_1J_2})\rangle| \\ &\le |I_1|^{-1/2}|I_2|^{-1/2}\int_{I_1}\int_{3I_1\setminus I_1}\frac{dx_1\,dy_1}{|x_1-y_1|^n}\cdot|J_1|^{-1/2}|J_2|^{-1/2}\int_{J_1}\int_{3J_1\setminus J_1}\frac{dx_2\,dy_2}{|x_2-y_2|^m} \\ &\lesssim \frac{|I_1|^{1/2}}{|I_2|^{1/2}}\frac{|J_1|^{1/2}}{|J_2|^{1/2}} \lesssim \frac{|I_1|^{1/2}}{|I_2|^{1/2}}\Big(\frac{\ell(I_1)}{\ell(I_2)}\Big)^{\delta/2}\frac{|J_1|^{1/2}}{|J_2|^{1/2}}\Big(\frac{\ell(J_1)}{\ell(J_2)}\Big)^{\delta/2}.\end{aligned}$$

The two terms after that are controlled using the mixed size and Hölder estimates of K. The last term is controlled using the Hölder estimate of K. The mixed cases where $2^{-r}\ell(I_2)\le\ell(I_1)(\le\ell(I_2))$ and $\ell(J_1)<2^{-r}\ell(J_2)$ or $\ell(I_1)<2^{-r}\ell(I_2)$ and $2^{-r}\ell(J_2)\le\ell(J_1)(\le\ell(J_2))$ are handled similarly. □

The above lemma shows that

$$\sum_{I_1\subsetneq I_2}\sum_{J_1\subsetneq J_2}\langle T(h_{I_1}\otimes u_{J_1}), s_{I_2I_2}\otimes s_{J_1J_2}\rangle\langle f, h_{I_1}\otimes u_{J_1}\rangle\langle g, h_{I_2}\otimes u_{J_2}\rangle$$

can be realized in the form

$$C\sum_{i_1=1}^{\infty}\sum_{j_1=1}^{\infty}2^{-i_1\delta/2}2^{-j_1\delta/2}\langle S^{i_10j_10}f,g\rangle.$$

The part

$$\sum_{I_1\subsetneq I_2}\sum_{J_1\subsetneq J_2}\langle u_{J_2}\rangle_{J_1}\langle T(h_{I_1}\otimes u_{J_1}), s_{I_2I_2}\otimes 1\rangle\langle f, h_{I_1}\otimes u_{J_1}\rangle\langle g, h_{I_2}\otimes u_{J_2}\rangle$$

can be written in the form

$$C\sum_{i_1=1}^{\infty}2^{-i_1\delta/2}\langle S^{i_1000}f,g\rangle,$$

where

$$S^{i_1000}f=\sum_K\sum_{\substack{I_1\subset K\\ \ell(I_1)=2^{-i_1}\ell(K)}}h_K\otimes\Pi^*_{b_{I_1K}}(\langle f,h_{I_1}\rangle_1)$$

and $b_{I_1K}=\langle T^*(s_{I_1K}\otimes 1),h_{I_1}\rangle_1/C2^{-i_1\delta/2}$. Since one can check $\|b_{I_1K}\|_{\mathrm{BMO}(\mathbb{R}^m)}\le c|I_1|^{1/2}/|K|^{1/2}$, it is similarly as has already been done in the separated/inside case seen that $\|S^{i_1000}f\|_2\le\|f\|_2$. The proof of the BMO estimate is similar to the proof of the previous lemma.

Completely analogously one can write

$$\sum_{I_1 \subsetneq I_2} \sum_{J_1 \subsetneq J_2} \langle h_{I_2} \rangle_{I_1} \langle T(h_{I_1} \otimes u_{J_1}), 1 \otimes s_{J_1 J_2} \rangle \langle f, h_{I_1} \otimes u_{J_1} \rangle \langle g, h_{I_2} \otimes u_{J_2} \rangle$$

in the form

$$C \sum_{j_1=1}^{\infty} 2^{-j_1 \delta/2} \langle S^{00 j_1 0} f, g \rangle,$$

where S^{00j_10} is a non-cancellative L^2 bounded shift.

The last part

$$\sum_{I_1 \subsetneq I_2} \sum_{J_1 \subsetneq J_2} \langle h_{I_2} \rangle_{I_1} \langle u_{J_2} \rangle_{J_1} \langle T(h_{I_1} \otimes u_{J_1}), 1 \rangle \langle f, h_{I_1} \otimes u_{J_1} \rangle \langle g, h_{I_2} \otimes u_{J_2} \rangle$$

collapses to

$$\sum_{K,V} \langle g \rangle_{K \times V} \langle T^* 1, h_K \otimes u_V \rangle \langle f, h_K \otimes u_V \rangle = C \langle \Pi^*_{T^*1/C} f, g \rangle,$$

where

$$\Pi_b f = \sum_{K,V} \langle f \rangle_{K \times V} \langle b, h_K \otimes u_V \rangle h_K \otimes u_V$$

is a bounded shift of the type $(0,0,0,0)$ for b in the product BMO of $\mathbb{R}^n \times \mathbb{R}^m$. So here we can set $S^{0000} = \Pi^*_{T^*1/C}$. Note that the correct normalization for this shift would follow just from the various kernel estimates and the weak boundedness property.

7.2. *Remark.* In the proof of this representation theorem there are paraproducts of essentially three different types. We have seen two types already: the full paraproduct

$$\Pi_b f = \sum_{K,V} \langle f \rangle_{K \times V} \langle b, h_K \otimes u_V \rangle h_K \otimes u_V$$

and some half paraproducts, like

$$f \mapsto \sum_{K} \sum_{\substack{I_1 \subset K \\ \ell(I_1) = 2^{-i_1} \ell(K)}} h_K \otimes \Pi^*_{b_{I_1 K}} (\langle f, h_{I_1} \rangle_1),$$

which have a paraproduct part only in the $\mathbb{R}^n$ or $\mathbb{R}^m$ variable. The third type of paraproduct does not surface in our current sum, where $\ell(I_1) \le \ell(I_2)$ and $\ell(J_1) \le \ell(J_2)$. However, for example in the mixed case, where $\ell(I_1) \le \ell(I_2)$ and $\ell(J_1) > \ell(J_2)$, one has in the corresponding inside/inside part the mixed full paraproduct

$$\begin{aligned} f \mapsto & \sum_{K,V} |K \times V|^{-1} \langle T_1(1), h_K \otimes u_V \rangle \langle f, h_K \otimes \chi_V \rangle \chi_K \otimes u_V \\ &= \sum_{K,V} \langle T_1(1), h_K \otimes u_V \rangle \langle f, h_K \otimes u_V^2 \rangle h_K^2 \otimes u_V, \end{aligned}$$

which is L^2 bounded as $T_1(1)$ belongs to the product BMO of $\mathbb{R}^n \times \mathbb{R}^m$ by assumption. It is rather straightforward and well-known that both of these full paraproducts are bounded on L^2 if they are defined by functions in the dyadic product BMO. This can be proven by duality – see for example [16].

8. Inside/equal

One splits

$$\langle T(h_{I_1} \otimes u_V), h_{I_2} \otimes u_V\rangle = \langle T(h_{I_1} \otimes u_V), s_{I_1I_2} \otimes u_V\rangle + \langle h_{I_2}\rangle_{I_1}\langle T(h_{I_1} \otimes u_V), 1 \otimes u_V\rangle,$$

where $s_{I_1I_2} = \chi_{I_{2,1}^c}(h_{I_2} - \langle h_{I_2}\rangle_{I_{2,1}})$ satisfies spt $s_{I_1I_2} \subset I_{2,1}^c$ and $|s_{I_1I_2}| \le 2|I_2|^{-1/2}$.

One may write

$$\sum_{I_1 \subsetneq I_2} \sum_V \langle T(h_{I_1} \otimes u_V), s_{I_1I_2} \otimes u_V\rangle\langle f, h_{I_1} \otimes u_V\rangle\langle g, h_{I_2} \otimes u_V\rangle$$

in the form

$$C\sum_{i_1=1}^{\infty} 2^{-i_1\delta/2}\langle S^{i_1000}f, g\rangle$$

with cancellative shifts. For this one needs that

$$|\langle T(h_{I_1} \otimes u_V), s_{I_1I_2} \otimes u_V\rangle| \lesssim \frac{|I_1|^{1/2}}{|I_2|^{1/2}}\Big(\frac{\ell(I_1)}{\ell(I_2)}\Big)^{\delta/2}.$$

Estimate

$$|\langle T(h_{I_1} \otimes u_V), s_{I_1I_2} \otimes u_V\rangle| \le |V|^{-1}\Big[\sum_{\substack{V',V'' \in \mathrm{ch}(V)\\ V' \ne V''}} |\langle T(h_{I_1} \otimes \chi_{V'}), s_{I_1I_2} \otimes \chi_{V''}\rangle| + \sum_{V' \in \mathrm{ch}(V)} |\langle T(h_{I_1} \otimes \chi_{V'}), s_{I_1I_2} \otimes \chi_{V'}\rangle|\Big].$$

In the case $V' \ne V''$ use the full kernel representation. In the diagonal case use the kernel representation in $\mathbb{R}^n$. If $\ell(I_1) < 2^{-r}\ell(I_2)$, use the mixed size and Hölder estimate of K (in the case $V' \ne V''$) or the Hölder estimate for the kernel $K_{\chi_{V'},\chi_{V'}}$ (in the case $V' = V''$). In the case $2^{-r}\ell(I_2) \le \ell(I_1)$ split $s_{I_1I_2} = \chi_{3I_1}s_{I_1I_2} + \chi_{(3I_1)^c}s_{I_1I_2}$. For $V' \ne V''$ use the size estimate of K for the first term and the mixed size and Hölder estimate of K for the second term. In the case $V' = V''$ use the size estimate of $K_{\chi_{V'},\chi_{V'}}$ for the first term, and the Hölder estimate of $K_{\chi_{V'},\chi_{V'}}$ for the second term.

One writes

$$\sum_{I_1 \subsetneq I_2} \sum_V \langle h_{I_2}\rangle_{I_1}\langle T(h_{I_1} \otimes u_V), 1 \otimes u_V\rangle\langle f, h_{I_1} \otimes u_V\rangle\langle g, h_{I_2} \otimes u_V\rangle$$

in the form

$$C\langle S^{0000}f, g\rangle,$$

where in this case

$$S^{0000} f = \sum_V \Pi^*_{b_V}(\langle f, u_V\rangle_2) \otimes u_V$$

and $b_V = \langle T^*(1 \otimes u_V), u_V\rangle_2/C$. This is indeed a non-cancellative shift of the type $(0,0,0,0)$.

8.1. **Lemma.** *There holds* $\|b_V\|_{\mathrm{BMO}(\mathbb{R}^n)} \le c$.

Proof. Fix a cube $K \subset \mathbb{R}^n$ and a function a so that $\operatorname{spt} a \subset K$, $|a| \le 1$ and $\int a = 0$. We need to show that $|\langle T(a \otimes u_V), 1 \otimes u_V\rangle| \lesssim |K|$. We begin with the split

$$\langle T(a \otimes u_V), 1 \otimes u_V\rangle = \langle T(a \otimes u_V), \chi_K \otimes u_V\rangle + \langle T(a \otimes u_V), \chi_{3K\setminus K} \otimes u_V\rangle + \langle T(a \otimes u_V), \chi_{(3K)^c} \otimes u_V\rangle.$$

There holds

$$\langle T(a \otimes u_V), \chi_{3K\setminus K} \otimes u_V\rangle \le |V|^{-1}\Big[\sum_{\substack{V',V''\in \mathrm{ch}(V)\\ V'\ne V''}} |\langle T(a \otimes \chi_{V'}), \chi_{3K\setminus K} \otimes \chi_{V''}\rangle| + \sum_{V'\in \mathrm{ch}(V)} |\langle T(a \otimes \chi_{V'}), \chi_{3K\setminus K} \otimes \chi_{V'}\rangle|\Big],$$

where

$$|\langle T(a \otimes \chi_{V'}), \chi_{3K\setminus K} \otimes \chi_{V''}\rangle| \le \int_K \int_{3K\setminus K} \frac{1}{|x_1-y_1|^n}\,dx_1\,dy_1 \cdot \int_{V'}\int_{V''} \frac{1}{|x_2-y_2|^m}\,dx_2\,dy_2 \lesssim |K||V|$$

and

$$|\langle T(a \otimes \chi_{V'}), \chi_{3K\setminus K} \otimes \chi_{V'}\rangle| \le C(\chi_{V'}, \chi_{V'}) \int_K \int_{3K\setminus K} \frac{1}{|x_1-y_1|^n}\,dx_1\,dy_1 \lesssim |K||V|.$$

Furthermore, we have

$$|\langle T(a \otimes u_V), \chi_{(3K)^c} \otimes u_V\rangle| \le |V|^{-1}\Big[\sum_{\substack{V',V''\in \mathrm{ch}(V)\\ V'\ne V''}} |\langle T(a \otimes \chi_{V'}), \chi_{(3K)^c} \otimes \chi_{V''}\rangle| + \sum_{V'\in \mathrm{ch}(V)} |\langle T(a \otimes \chi_{V'}), \chi_{(3K)^c} \otimes \chi_{V'}\rangle|\Big],$$

where

$$|\langle T(a \otimes \chi_{V'}), \chi_{(3K)^c} \otimes \chi_{V''}\rangle| \lesssim |K| \cdot \ell(K)^\delta \int_{(3K)^c} \frac{dx_1}{|x_1-c_K|^{n+\delta}} \cdot \int_{V'}\int_{V''} \frac{1}{|x_2-y_2|^m}\,dx_2\,dy_2 \lesssim |K||V|$$

and

$$|\langle T(a \otimes \chi_{V'}), \chi_{(3K)^c} \otimes \chi_{V'} \rangle| \le C(\chi_{V'}, \chi_{V'}) \int_K \int_{(3K)^c} \frac{\ell(K)^\delta}{|x_1 - c_K|^{n+\delta}}\, dx_1\, dy_1 \lesssim |K||V|.$$

For the first term we again begin with the estimate

$$|\langle T(a \otimes u_V), \chi_K \otimes u_V \rangle| \lesssim |V|^{-1} \sum_{V', V'' \in \operatorname{ch}(V)} |\langle T(a \otimes \chi_{V'}), \chi_K \otimes \chi_{V''} \rangle|.$$

Let us consider the case $V' \ne V''$. In this case we have

$$\begin{aligned} |\langle T(a \otimes \chi_{V'}), \chi_K \otimes \chi_{V''} \rangle| &= \Big| \int_{V'} \int_{V''} K_{a,\chi_K}(x_2, y_2)\, dx_2\, dy_2 \Big| \\ &\le C(a, \chi_K) \int_{V'} \int_{V''} \frac{1}{|x_2 - y_2|^m}\, dx_2\, dy_2 \lesssim |K||V|. \end{aligned}$$

Thus, we are only left with the need for the estimate $|\langle T(a \otimes \chi_{V'}), \chi_K \otimes \chi_{V'} \rangle| \lesssim |K||V|$ – but this is one of the diagonal BMO assumptions.

□

Because of this lemma, one can show, similarly but with a bit less effort than in Proposition 4.5, that S^{0000} is L^2 bounded.

9. INSIDE/NEARBY

This goes very much so in the same vein as the inside/equal case. In fact, this is easier since the nearby cubes do not intersect by definition. From the series with the matrix element $\langle T(h_{I_1} \otimes u_{J_1}), s_{I_1 I_2} \otimes u_{J_2} \rangle$ we get

$$C \sum_{i_1=1}^{\infty} \sum_{j_1=1}^{r} \sum_{j_2=1}^{j_1} 2^{-i_1\delta/2} 2^{-j_1\delta/2} \langle S^{i_1 0 j_1 j_2} f, g \rangle.$$

From the series with the matrix element $\langle h_{I_2} \rangle_{I_1} \langle T(h_{I_1} \otimes u_{J_1}), 1 \otimes u_{J_2} \rangle$ we get

$$C \sum_{j_1=1}^{r} \sum_{j_2=1}^{j_1} 2^{-j_1\delta/2} \langle S^{00 j_1 j_2} f, g \rangle$$

with bounded non-cancellative shifts.

10. EQUAL/EQUAL

This part can be realized in the form $C\langle S^{0000} f, g \rangle$ for a cancellative shift, since one can just estimate $|\langle T(h_K \otimes u_V), h_K \otimes u_V \rangle| \lesssim 1$. This estimate is an easy consequence of the weak boundedness property and the size estimates of our kernels.

11. EQUAL/NEARBY

This part is clearly of the form

$$C\sum_{j_1=1}^{r}\sum_{j_2=1}^{j_1}2^{-j_1\delta/2}\langle S^{00j_1j_2}f,g\rangle,$$

where the shifts are cancellative. Here one can again just use the estimate $|\langle T(h_K\otimes u_{J_1}),h_K\otimes u_{J_2}\rangle|\lesssim 1$, which follows just from the size estimates of our kernels.

12. NEARBY/NEARBY

This part is of the form

$$C\sum_{i_1=1}^{r}\sum_{i_2=1}^{i_1}\sum_{j_1=1}^{r}\sum_{j_2=1}^{j_1}2^{-i_1\delta/2}2^{-j_1\delta/2}\langle S^{i_1i_2j_1j_2}f,g\rangle$$

once again because of the easy estimate $|\langle T(h_{I_1}\otimes u_{J_1}),h_{I_2}\otimes u_{J_2}\rangle|\lesssim 1$. This follows from the size estimate for the full kernel.

REFERENCES

[1] S. Y. A. Chang and R. Fefferman, *Some recent developments in Fourier analysis and H^p theory on product domains*, Bull. Amer. Math. Soc. **12** (1985), 1–43.

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

[3] G. David, J.-L. Journé, and S. Semmes, *Opérateurs de Calderón-Zygmund, fonctions para-accretives et interpolation*, Rev. Mat. Iberoamericana **1** (1985), no. 4, 1–56.

[4] R. Fefferman, *Harmonic analysis on product spaces*, Ann. Math. **7** (1987), 109–130.

[5] R. Fefferman and E. Stein, *Singular integrals on product spaces*, Adv. in Math. **45** (1982), no. 2, 117–143.

[6] S. Ferguson and M. Lacey, *A characterization of product BMO by commutators*, Acta Math. **189** (2002), 143–160.

[7] T. Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. Math., to appear; preprint (2010).

[8] T. Hytönen, M. Lacey, H. Martikainen, T. Orponen, M. C. Reguera, E. Sawyer, and I. Uriarte–Tuero, *Weak and strong type estimates for maximal truncations of Calderón-Zygmund operators on A_p weighted spaces*, J. Anal. Math., to appear; preprint (2011).

[9] T. Hytönen and H. Martikainen, *Non-homogeneous Tb theorem and random dyadic cubes on metric measure spaces*, J. Geom. Anal., to appear; preprint (2009).

[10] T. Hytönen, C. Pérez, S. Treil, and A. Volberg, *Sharp weighted estimates for dyadic shifts and the A_2 conjecture*, preprint (2010).

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

[12] M. Lacey and J. Metcalfe, *Paraproducts in one and several parameters*, Forum Math. **19** (2007), 325–351.

[13] C. Muscalu, J. Pipher, T. Tao, and C. Thiele, *Bi-parameter paraproducts*, Acta Math. **193** (2004), 269–296

[14] F. Nazarov, A. Reznikov, and A. Volberg, *The proof of A_2 conjecture in a geometrically doubling metric space*, preprint (2011).

[15] F. Nazarov, S. Treil, and A. Volberg, *The Tb-theorem on non-homogeneous spaces*, Acta Math. **190** (2003), no. 2, 151–239.
[16] S. Pott and P. Villarroya, *A $T(1)$ theorem on product spaces*, preprint (2011).

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF HELSINKI, P.O.B. 68, FI-00014 HELSINKI, FINLAND
E-mail address: henri.martikainen@helsinki.fi