A $T(b)$ THEOREM ON PRODUCT SPACES

YUMENG OU

ABSTRACT. The main result of this paper is a bi-parameter $T(b)$ theorem for the case that $b$ is a tensor product of two pseudo-accretive functions. In the proof, we also discuss the $L^2$ boundedness of different types of the $b$-adapted bi-parameter paraproducts.

1. INTRODUCTION

The study of the $T(1)/T(b)$ type theorems in the multi-parameter setting can be dated back to 1985, when Journé [14] proved the first multi-parameter $T(1)$ theorem by treating the singular integral operator as a vector-valued one-parameter operator. The result itself is very elegant except that some partial boundedness of the operator needs to be assumed. More recently, Pott and Villarroya in [20] prove a new bi-parameter $T(1)$ theorem with much weaker assumptions on the operator, where they formed different types of mixed conditions instead of assuming the partial boundedness. This is the point of view taken by Martikainen in [16], where he proved a representation theorem for bi-parameter singular integral operators which then implies a $T(1)$ result, and in his joint work with Hytönen [12], where they showed a bi-parameter $T(1)$ theorem in spaces of non-homogeneous type.

In this paper, for the first time, we prove a $T(b)$ theorem in product spaces, which is a natural extension of the work we have mentioned above.

1.1. Definition. A function $b \in L^\infty(\mathbb{R}^n \times \mathbb{R}^m)$ is called pseudo-accretive if there is a constant $C$ such that for any rectangle $R$ in $\mathbb{R}^n \times \mathbb{R}^m$ with sides parallel to axes, $\frac{1}{|R|} \int_R |b| > C$.

We will only discuss the case when $b = b_1 \otimes b_2$, where $b_1$ and $b_2$ are in $L^\infty(\mathbb{R}^n)$ and $L^\infty(\mathbb{R}^m)$, respectively. Then, the pseudo-accretivity and boundedness of $b$...
imply that there exists a constant $C$, such that for any cubes $K \subset \mathbb{R}^n, V \subset \mathbb{R}^m$, \[
frac{1}{|K|} \int_K b_1 > C \quad \text{and} \quad \nfrac{1}{|V|} \int_V b_2 > C,
\]
i.e. $b_1$ and $b_2$ are both pseudo-accretive in the classical sense. Although this seems to be too restrictive, it is actually quite natural. Note that $b = 1$ falls in this class. Moreover, in all of the papers mentioned above, some partial structures on the operator are required in order to treat those mixed problems risen in the bi-parameter setting. In other words, the singular integral operator itself we are looking at behaves like a tensor product in some sense. It is essential in our argument for $b$ to be a tensor product, otherwise, even defining $Tb$ would become a problem.

Just as in the situation for the bi-parameter $T(1)$ theorems, we still need to assume that besides $T, T^*$, the partial adjoints of $T$ also map $b$ to a $BMO$ function, an assumption shown by Journé [14] to be unnecessary for $T$ to be $L^2$ bounded. A more detailed discussion can be found in Section 6 of [14].

The main technique of the proof is to decompose $L^2$ functions into sums of martingale differences adapted to $b$, analyze each part of the sums, and show that they have good enough decay to be summed up. The advantage of analyzing martingale differences is that they are supported on dyadic rectangles, constant on each of their children, and have orthogonality. Martikainen followed a similar strategy in [16], using Haar functions. However, when we treat $b$ instead of 1, we have to create a bi-parameter $b$-adapted martingale difference decomposition, which makes the estimate of each part of the sum much less transparent. In the one-parameter setting, the idea of using such $b$-adapted martingale difference operators is well known and has been discussed by many authors in their proofs of different types of $Tb$ theorems, such as David, Journé and Semmes [5], Coifman, Jones and Semmes [3], Nazarov, Treil and Volberg [18], Hytönen and Martikainen [11]. But in the bi-parameter case, the $b$-adapted martingale difference has never been treated before.

The operator $T$ studied in this paper is initially defined as a continuous linear map from $bC_c^\infty(\mathbb{R}^n \times \mathbb{R}^m)$ to its dual. In order to justify the convergence of pairings of martingale differences, we also assume a priori that $T$ is bounded on $L^2(\mathbb{R}^n \times \mathbb{R}^m)$, although we will show that quantitatively the operator norm of $T$ is bounded by some constant depending only on the weak assumptions introduced in the following, but has nothing to do with the assumed $L^2 \to L^2$ norm. Note that although this a priori assumption is often unnecessary, it appears as a hypothesis in the proofs of some $T(1)$ theorems: many authors have added this assumption ([16, 12]), even in the one-parameter setting ([18, 9]). It is not a consequence of involving $b$, but results from the fact that one has an initially continuously defined operator which is treated dyadically. Thus, we are more interested in showing how those weak assumptions quantitatively control the
$L^2 \to L^2$ norm of $T$. However, in some specific examples that we will mention later, this a priori assumption can be removed.

The plan for the paper is the following. First, we introduce the assumptions on the operators as well as necessary preliminary on bi-parameter $b$-adapted martingale differences. Second, before stating and proving the $T(b)$ theorem, we discuss some types of bi-parameter $b$-adapted paraproducts, which will be used later. Next, we give an averaging formula in the same flavor as in [16], which enables us to use the concept of "goodness" of cubes in our estimate. Then, we will move on to the main body of this paper, prove the $T(b)$ theorem by a case by case estimate of terms in the averaging formula.

ACKNOWLEDGEMENT

The author would like to thank Jill Pipher for guiding her into this area, suggesting the topic and the numerous fruitful discussions. The author is also grateful to Michael Lacey and Brett Wick for useful discussions during her visit to Georgia Institute of Technology.

2. ASSUMPTIONS ON THE OPERATOR

Bi-parameter $b$-adapted martingale differences. As a preliminary, we begin with a quick introduction of the martingale difference decomposition adapted to our problem.

Let $\omega^n = (\omega^n_i)_{i \in \mathbb{Z}}$, where $\omega^n_i \in \{0, 1\}^n$. Let $D^n_0$ be the standard dyadic grid on $\mathbb{R}^n$. We define the shifted dyadic grid $D^n_\omega = \{I + \sum_{i:2^{-i} < \ell(I)} 2^{-i} \omega_i : I \in D^n_0\} = \{I + \omega^n : I \in D^n_0\}$, where $I + \omega^n := I + \sum_{i:2^{-i} < \ell(I)} 2^{-i} \omega^n_i$. There is a natural probability structure on $(\{0, 1\}^n)^\mathbb{Z}$, which gives us a random dyadic grid $D^n_\omega$ in $\mathbb{R}^n$. When there is no need to specify what is the $\omega^n$, most of the time, we just write $D^n$ for short. Interested readers can find more detailed discussion of random dyadic grids in [9] or [16].

Given a pseudo-accrative function $b = b_1 \otimes b_2$, and two fixed dyadic grids $D^n, D^m$ in $\mathbb{R}^n, \mathbb{R}^m$, respectively. For each $p \in \mathbb{Z}$, let $D^n_p$ be the collection of cubes of side length $2^{-p}$ in $D^n$, we have

$$E_{p}^{b_1} f = \sum_{I \in D^n_p} \frac{\int_I f b_1}{\int_I b_1} \chi_I, \quad E_{I}^{b_1} f = \chi_I E_{p}^{b_1} f.$$ 

Similarly, we have $E_q^{b_2}$ and $E_{J}^{b_2}$ defined for each $q \in \mathbb{Z}, J \in D^m$. Then their composition is a $b$-adapted double expectation operator:

$$E_{p,q}^{b} = E_{p}^{b_1} E_{q}^{b_2} = E_{q}^{b_2} E_{p}^{b_1}.$$
Let $\Delta^{b_1}_p = E^{b_1}_{p+1} - E^{b_1}_p$, $\Delta^{b_1}_I = \chi_I \Delta^{b_1}_p$ for each $I \in D^n_p$, and similarly for the other variable. The $b$-adapted double martingale difference is defined as

$$
\Delta^{b}_{p,q} = \Delta^{b_1}_p \Delta^{b_2}_q = \Delta^{b_2}_q \Delta^{b_1}_p.
$$

The following properties can be easily checked:

1. $\Delta^{b}_{I \times J} f$ is supported on the dyadic rectangle $I \times J$, and is a constant on each of its children;

2. $\int b_1 \Delta^{b}_{p,q} f \, dx_1 = \int b_2 \Delta^{b}_{p,q} f \, dx_2 = 0$;

3. $\Delta^{b}_{p,q} \Delta^{b}_{k,l} = 0$ unless $p = k, q = l$, and in this case it equals $\Delta^{b}_{p,q}$;

4. If $f \in L^2(\mathbb{R}^n \times \mathbb{R}^m)$, then $f = \sum_{p,q} \Delta^{b}_{p,q} f$ with convergence in $L^2$, and

$$
\|f\|_{L^2}^2 \lesssim \sum_{p,q} \|\Delta^{b}_{p,q} f\|_{L^2}^2 \lesssim \|f\|_{L^2}^2.
$$

Property (4) can be verified by iteration of the one-parameter martingale difference argument in [18].

Moreover, we observe that

$$
E^{b_1*}_{p,q} f = E^{b_1*}_p E^{b_2*}_q f = b \sum_{I \in D^n_p, J \in D^m_q} \frac{\int_{I \times J} f}{b} \chi_{I \times J}.
$$

and hence

$$
M_b \Delta^{b}_{p,q} = \Delta^{b_1*}_{p,q} M_b,
$$

where $M_b f = bf$ is the multiplication operator by $b$.

We now introduce the assumptions on $T$ that we will need throughout the argument. Fix two pseudo-accretive functions $b = b_1 \otimes b_2, b' = b'_1 \otimes b'_2$. For simplicity, denote $d = b_1 \otimes b'_2$ and $d' = b'_1 \otimes b_2$, then obviously $d, d'$ are also pseudo-accretive.

**Full Calderón-Zygmund structure.** If $f = f_1 \otimes f_2$ and $g = g_1 \otimes g_2$ with $f_1, g_1 \in C^\infty_0(\mathbb{R}^n), f_2, g_2 \in C^\infty_0(\mathbb{R}^m)$, $\text{spt} f_1 \cap \text{spt} g_1 = \emptyset$ and $\text{spt} f_2 \cap \text{spt} g_2 = \emptyset$, then we have the full kernel representation

$$
\langle M_b T M_b f, g \rangle = \int_{\mathbb{R}^{n+m}} \int_{\mathbb{R}^{n+m}} K(x, y) f(y) g(x) b(y) b'(x) \, dxdy.
$$

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

\[ |K(x, y)| \leq C \frac{1}{|x_1 - y_1|^n} \frac{1}{|x_2 - y_2|^m}. \]

(2) Hölder conditions

\[ |K(x, y) - K(x, (y_1, y_2')) - K(x, (y_1', y_2)) + K(x, y')| \leq C \frac{|y_1 - y_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{|y_2 - y_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \]

whenever \( |y_1 - y_1'| \leq |x_1 - y_1|/2 \) and \( |y_2 - y_2'| \leq |x_2 - y_2|/2 \).

\[ |K((x, x_2'), y) - K((x_1, x_2), y) + K(x', y)| \leq C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \]

whenever \( |x_1 - x_1'| \leq |x_1 - y_1|/2 \) and \( |x_2 - x_2'| \leq |x_2 - y_2|/2 \).

\[ |K(x, (y_1, y_2')) - K((x_1, x_2), y_1) + K(x, (y_1, y_2))| \leq C \frac{|x_1 - x_1'|^\delta}{|x_1 - y_1|^{n+\delta}} \frac{|y_2 - y_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \]

whenever \( |x_1 - x_1'| \leq |x_1 - y_1|/2 \) and \( |y_2 - y_2'| \leq |x_2 - y_2|/2 \).

(3) Mixed Hölder-size conditions

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

whenever \( |x_1 - x_1'| \leq |x_1 - y_1|/2 \),

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

whenever \( |y_1 - y_1'| \leq |x_1 - y_1|/2 \),

\[ |K((x_1, x_2'), y) - K((x_1, x_2), y)| \leq C \frac{1}{|x_1 - y_1|^n} \frac{|x_2 - x_2'|^\delta}{|x_2 - y_2|^{m+\delta}} \]

whenever \( |x_2 - x_2'| \leq |x_2 - y_2|/2 \),

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

whenever \( |y_2 - y_2'| \leq |x_2 - y_2|/2 \).
Partial Calderón-Zygmund structure. We also need some C-Z structure on \( \mathbb{R}^n \) and \( \mathbb{R}^m \) separately to deal with the case when \( f, g \) are only separated on one variable. If \( f = f_1 \otimes f_2, g = g_1 \otimes g_2 \) and \( \text{spt} f_1 \cap \text{spt} g_1 = \emptyset \), then we have the partial kernel representation

\[
\langle M_b T M_0 f, g \rangle = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K_{f_2,g_2}(x_1,y_1) f_1(y_1) g_1(x_1) b_1(y_1) b'_1(x_1) \, dx_1 dy_1.
\]

The partial kernel \( K_{f_2,g_2} \) defined on \( (\mathbb{R}^n \times \mathbb{R}^m) \setminus \{(x_1, y_1) \in \mathbb{R}^n \times \mathbb{R}^m : x_1 = y_1 \} \) is assumed to satisfy the following standard estimates:

1. **Size condition**

\[
|K_{f_2,g_2}| \leq C(f_2, g_2) \frac{1}{|x_1 - y_1|^n}.
\]

2. **Hölder conditions**

\[
|K_{f_2,g_2}(x_1, y_1) - K_{f_2,g_2}(x'_1, y_1)| \leq C(f_2, g_2) \frac{|x_1 - x'_1|^\delta}{|x_1 - y_1|^{n+\delta}}
\]

whenever \( |x_1 - x'_1| \leq |x_1 - y_1|/2 \),

\[
|K_{f_2,g_2}(x_1, y_1) - K_{f_2,g_2}(x_1, y'_1)| \leq C(f_2, g_2) \frac{|y_1 - y'_1|^\delta}{|x_1 - y_1|^{n+\delta}}
\]

whenever \( |y_1 - y'_1| \leq |x_1 - y_1|/2 \).

This assumption is in the same flavor of [16], and is important of defining \( T(b) \). In fact, we can weaken this by assuming the above only for the cases when

\[
(f_2, g_2) = (\chi_V, \chi_V), \quad (\chi_V, u_V b_2^{-1}), \quad \text{or} \quad (u_V b_2^{-1}, \chi_V),
\]

for any cube \( V \subset \mathbb{R}^m \), and \( u_V \) being a \( V \)-adapted function with zero-mean (i.e. \( \text{spt} u_V \subset V, |u_V| \leq 1 \) and \( \int u_V = 0 \)).

We also need to assume that there exists a universal constant \( C \), such that

\[
C(\chi_V, \chi_V) + C(\chi_V, u_V b_2^{-1}) + C(u_V b_2^{-1}, \chi_V) \leq C|V|.
\]

It is easily shown that both full and partial kernel representations also hold when \( f, g \) are finite linear combinations of characteristic functions, or even tensor products of compactly supported \( L^\infty \) functions, as long as for the required variable, they are still disjointly supported. To see this, when taking those functions, following from the standard condition on the kernels, both integrals are still convergent. We can use them to define the corresponding bilinear forms. After we finally show that \( T \) is bounded on \( L^2 \) (here we don’t even need the boundedness assumption on \( T \) a priori), use the density of \( C_0^\infty \) functions and Lebesgue dominated convergence theorem, we can show that the bilinear form has to be equal to the kernel representation, hence is well defined.
The partial C-Z structure assumption is natural. Recall how Journé defined his class of operators in [14]. Rephrasing in terms of our definition, Journé assumed that the partial kernel $K_{f_2,g_2}(x_1, y_1)$ is a bilinear form associated with a $L(L^2(\mathbb{R}^m), L^2(\mathbb{R}^m))$ valued standard C-Z kernel, which then implies the size and Hölder conditions (2.1), (2.2), (2.3). In the bi-parameter setting, the partial C-Z structure assumptions are required to both define $Tb$ and to handle the "mixed cases”. That arise because of the independent behavior in each variable. (See Section 6, 7, 9, 12 for discussions of different "mixed cases”). As far as we know, all the previous literature in this area needs some assumptions about the partial C-Z structure of the operator. For example, in Pott and Villa arroya’s most recent version of [20], they included such an assumption on the operator so that they can fully justify the definition of $T1$. Although it is formulated a little differently, but is in spirit the same as ours. Martikainen ([16]) also requires a similar assumption. (See Section 2 of [16]).

Note that in the case $f, g$ are separated in both variables, i.e. when we have the full kernel representation, the partial kernels are just

$$K_{f_2,g_2}(x_1, y_1) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^m} K(x, y)f_2(y_2)g_2(x_2)b_2(x_2)b_2'(y_2) \, dx_2 dy_2,$$

and both of the size and Hölder conditions follow easily.

We also assume that the symmetric partial kernel representation and corresponding conditions on kernel $K_{f_1,g_1}$ in the case $spt f_2 \cap spt g_2 = \emptyset$.

**Weak boundedness property.** We assume that there exists a constant $C$ such that, for any cube $K \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$,

$$|\langle M_b T M_b', \chi_K \otimes \chi_V \rangle| \leq C |K||V|.$$

**BMO conditions.** We assume $Tb, T^*b', T_1 d', T_1^* d \in BMO(\mathbb{R}^n \times \mathbb{R}^m)$, where $T_1$ is the partial adjoint of $T$ defined by

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

Here, by assuming that they are in $BMO(\mathbb{R}^n \times \mathbb{R}^m)$, equivalently, we mean that they are in $BMO_d(\mathbb{R}^n \times \mathbb{R}^m)$, the dyadic BMO space for any dyadic grid. It is proved by Pipher and Ward [19] that in the bi-parameter setting, the product BMO is the average of dyadic BMO. This result is then reproved and extended to multi-parameter by Treil [21] through a different method. We now run into a problem of defining $Tb$ (and similarly for the other three functions). In order to do this, we are going to show that $Tb$ lies in the dual of some properly selected subspace $A$ of $H^1_d(\mathbb{R}^n \times \mathbb{R}^m)$, i.e. the bilinear form $\langle g, Tb \rangle$ is well defined for any $g \in A$. 

A $T(b)$ THEOREM ON PRODUCT SPACES

7
Let \( A \) be the space consisting of all the functions

\[
\nu \sum_{\text{finite } I, J} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f
\]

where \( f \in C_{0}^{\infty}(\mathbb{R}^{n} \times \mathbb{R}^{m}), \ I \in \mathcal{D}^{m}, \ J \in \mathcal{D}^{m} \) and we are summing over a finite number of terms. It is easily seen that \( A \) is indeed a subspace of \( H_{d}^{1}(\mathbb{R}^{n} \times \mathbb{R}^{m}) \).

Hence, by linearity, it suffices to define \( \langle b^{I} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T b \rangle \).

Divide the bilinear form into four parts:

\[
\langle b^{I} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T (b_{X3I} \otimes \chi_{3J}) \rangle + \langle b^{I} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T (b_{X3I} \otimes \chi_{3J})^{c} \rangle
\]

\[
+ \langle b^{I} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T (b_{X3I}^{c} \otimes \chi_{3J}) \rangle + \langle b^{I} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T (b_{X3I}^{c} \otimes \chi_{3J})^{c} \rangle.
\]

Part one: \( \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f \) is a finite linear combination of characteristic functions. For each \( I_{i} \in \text{ch}(I), \ J_{j} \in \text{ch}(J) \),

\[
\langle b^{I} \chi_{I_{i}} \otimes \chi_{J_{j}}, \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T (b_{X3I} \otimes \chi_{3J}) \rangle = \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f |_{I_{i} \times J_{j}}, \langle b^{I} \chi_{I_{i}} \otimes \chi_{J_{j}}, T (b_{X3I} \otimes \chi_{3J}) \rangle,
\]

and

\[
\langle b^{I} \chi_{I_{i}} \otimes \chi_{J_{j}}, T (b_{X3I} \otimes \chi_{3J}) \rangle
\]

\[
= \langle b^{I} \chi_{I_{i}} \otimes \chi_{J_{j}}, T (b_{X3I} \otimes \chi_{3J}) \rangle + \langle b^{I} \chi_{I_{i}} \otimes \chi_{J_{j}}, T (b_{X3I} \otimes \chi_{3J}^{c}) \rangle
\]

\[
+ \langle b^{I} \chi_{I_{i}} \otimes \chi_{J_{j}}, T (b_{X3I}^{c} \otimes \chi_{3J}) \rangle + \langle b^{I} \chi_{I_{i}} \otimes \chi_{J_{j}}, T (b_{X3I}^{c} \otimes \chi_{3J}^{c}) \rangle.
\]

In the above, the first term makes sense due to the weak boundedness property. The second and third terms can be dealt with using partial kernel representation. Finally, the last term can be defined using full kernel representation.

Part two (and similarly for part three): Write

\[
\langle b^{I} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T (b_{X3I} \otimes \chi_{3J})^{c} \rangle = \sum_{i=1}^{2^{n}} \langle b^{I} \chi_{I_{i}} \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f, T (b_{X3I} \otimes \chi_{3J})^{c} \rangle
\]

\[
= \sum_{i=1}^{2^{n}} \langle b^{I} \chi_{I_{i}} (x_{1}) \otimes \Delta_{I}^{b_{1}} \Delta_{J}^{b_{2}} f |_{I_{i}} (x_{2}), T (b_{X3I} \otimes \chi_{3J})^{c} \rangle.
\]

Then for each term in the above, since the functions have good separation on one variable, we know that in the case that everything is compactly supported, it has
a partial kernel representation:
\[
\langle b'\chi_{I_i}(x_1) \otimes \Delta_{J_i}^{b_1} \Delta_{J_i}^{b_2} f(x_2), T(b\chi_{3I} \otimes \chi_{3J}) \rangle
\]
\[
= \int_{(3J)^e} \int_{I \times J} K_{X_{3I} \setminus X_{I_i}}(x_2, y_2) b_2(y_2) \Delta_{J_i}^{b_1} \Delta_{J_i}^{b_2} f_{|I_i}(x_2) b'_2(x_2) \, dx_2 dy_2
\]
\[
= \int_{(3J)^e} \int_{I \times J} \left[ K_{X_{3I} \setminus X_{I_i}}(x_2, y_2) - K_{X_{3J} \setminus X_{I_i}}(x_2, y_2) \right] b_2(y_2) \Delta_{J_i}^{b_1} \Delta_{J_i}^{b_2} f_{|I_i}(x_2) b'_2(x_2) \, dx_2 dy_2,
\]

While the integrand is not compactly supported, and the Hölder condition for partial kernels implies that the integral is convergent, it can be used to serve as the definition of the bilinear form on the left hand side.

Part four: In this part, the functions have good separations on both variables. As above, although we don’t have a full kernel representation for the bilinear form directly due to the fact that the integrand is not compactly supported, we can define it as follows:
\[
\int_{(3I)^e \times (3J)^e} \int_{I \times J} K(x, y) b(y)b'(x) \Delta_{J_i}^{b_1} \Delta_{J_i}^{b_2} f(x) \, dxdy
\]
and prove that the integral does converge. To see this last fact, we change \(K(x, y)\) to
\[
K(x, y) - K(c_1, x_2, y) - K(x_1, c_J, y) + K(c_1, c_J, y)
\]
by cancellation. Then the Hölder condition for the full kernel will imply the convergence of the integral.

Note that in parts two, three and four, we don’t give an arbitrary definition to those bilinear forms. A simple limiting argument shows that they are well defined. Consider part four for example. Let \(\phi\) be a cut-off function, such that \(\phi = 1\) on \(I \times J\), and \(\phi = 0\) outside \(3I \times 3J\). Denote dilation
\[
D_{k_1, k_2}(x) = \phi(x_1 k_1^{-1}, x_2 k_2^{-1}).
\]
Since \(\Delta_{J_i}^{b_1} \Delta_{J_i}^{b_2} f\) is a finite linear combination of characteristic functions, by the linearity of bilinear forms and full kernel representations, we have
\[
\langle b'\Delta_{J_i}^{b_1} \Delta_{J_i}^{b_2} f, T(bD_{k_1, k_2}\phi \chi_{3I} \otimes \chi_{3J}) \rangle
\]
\[
= \int_{(3I)^e \times (3J)^e} \int_{I \times J} K(x, y) D_{k_1, k_2}(y) b'_2(x) \Delta_{J_i}^{b_1} \Delta_{J_i}^{b_2} f(x) \, dxdy.
\]

Changing the kernel and using the Hölder condition for the full kernel as above, together with the boundedness of \(f\) and \(\phi\), we can show that the integrand is uniformly bounded by a constant multiple of
\[
\frac{1}{|x_1 - y_1|^{n+\delta}} \frac{1}{|x_2 - y_2|^{m+\delta}}.
\]
Then the Lebesgue dominated convergence theorem implies that
\[
\langle b' \Delta_I^{b'_1} \Delta_J f, T(b(3I)c \otimes \chi(3J)c) \rangle \\
= \lim_{k_1, k_2 \to \infty} \langle b' \Delta_I^{b'_1} \Delta_J f, T(bD_{k_1, k_2} \varphi(3I)c \otimes \chi(3J)c) \rangle \\
= \int_{(3I)c \times (3J)c} K(x, y) b(y) b'(x) \Delta_I^{b'_1} \Delta_J f(x) \, dx \, dy.
\]
And it’s easily seen that the above definition is independent of the choice of \( \varphi \).

Hence, \( Tb \) lies in the dual of \( A \). By saying that it belongs to \( BMO_d(\mathbb{R}^n \times \mathbb{R}^m) \), we mean that it is bounded on \( A \) and can be boundedly extended to a functional defined on the whole \( H_1^d(\mathbb{R}^n \times \mathbb{R}^m) \). And we can use the same technique above to give meanings to the other three objects similarly. Note that we can actually weaken this \( BMO \) assumption by only assuming that \( T(b) \) is a functional on \( A \), and similarly for the other three (but with differently chosen subspaces of \( H^1(\mathbb{R}^n \times \mathbb{R}^m) \)). We will see in the following that this is all we need.

**Diagonal BMO conditions.** There exists constant \( C \) such that, for any cube \( K \subset \mathbb{R}^n \), \( V \subset \mathbb{R}^m \), and any zero-mean functions \( a_K \), \( b_V \) which are \( K \), \( V \) adapted, respectively, the following hold:

- \( |\langle M_b T M_b(a_K b_1^{-1} \otimes \chi_V), \chi_K \otimes \chi_V \rangle| \leq C |K| |V| \)
- \( |\langle M_b T M_b(\chi_K \otimes \chi_V), a_K b_1^{-1} \otimes \chi_V \rangle| \leq C |K| |V| \)
- \( |\langle M_b T M_b(\chi_K \otimes b_V b_2^{-1}), \chi_K \otimes \chi_V \rangle| \leq C |K| |V| \)
- \( |\langle M_b T M_b(\chi_K \otimes \chi_V), \chi_K \otimes b_V b_2^{-1} \rangle| \leq C |K| |V| \)

3. **Bi-parameter \( b \)-adapted paraproducts**

In this section, we will discuss the boundedness of three different kinds of bi-parameter \( b \)-adapted paraproducts that will be used in the proof of our \( T(b) \) theorem.

**Partial paraproducts.** By partial paraproduct we mean a classical one-parameter \( b \)-adapted paraproduct with respect to one variable.

3.1. **Definition.** Let \( a \in BMO(\mathbb{R}^m) \). Then, for two fixed pseudo-accretive functions \( b_2, b'_2 \in L^\infty(\mathbb{R}^m) \), the operator \( \pi_{a, b_2}^{b'_2} \) is a partial paraproduct, acting on functions on \( \mathbb{R}^m \):

\[
\pi_{a, b_2}^{b'_2}(f) = \sum_{V \in \mathcal{D}^m} \langle f \rangle_{b'_2}^{b_2} M_{b_2} \Delta_{b'_2} a.
\]

Similarly, there is a symmetric partial paraproduct with respect to the other variable for fixed pseudo-accretive functions \( b_1, b'_1 \in L^\infty(\mathbb{R}^n) \), acting on functions on \( \mathbb{R}^n \).
3.2. Proposition. Partial paraproducts are bounded operators on $L^2$. Specifically,
\[
\|\mathcal{P}^{b_2}_{a} (f)\|_{L^2} \lesssim \|a\|_{BMO} \|f\|_{L^2},
\]
and a similar inequality holds for the symmetric one.

Proof. We only prove the first inequality. For any $f, g \in L^2$, \[
|\langle \mathcal{P}^{b_2}_{a} (f), g \rangle| = |\sum_{V \in \mathcal{D}^m} \langle f \rangle^2 \langle b_2 \rangle_{V} M_{b_2} \Delta_{b_2} a \cdot g | \\
= |\sum_{V} \langle f \rangle^2 \langle b_2 \rangle_{V} M_{b_2} \Delta_{b_2} a \cdot \Delta_{b_2} g | \\
\lesssim \sum_{V} \|\Delta_{b_2} a\|_{L^2} \|\Delta_{b_2} g\|_{L^2} \\
\leq \left( \sum_{V} \|\langle f \rangle^2 \Delta_{b_2} a\|_{L^2} \right)^{1/2} \left( \sum_{V} \|\Delta_{b_2} g\|_{L^2} \right)^{1/2} \\
\lesssim \left( \sum_{V} \|\langle f \rangle^2 \Delta_{b_2} a\|_{L^2} \right)^{1/2} \|g\|_{L^2}.
\]

where the fourth and fifth lines follow from Hölder inequality. Hence, it suffices to show that \[
\sum_{V} \|\langle f \rangle^2 \Delta_{b_2} a\|_{L^2} \lesssim \|a\|_{BMO} \|f\|_{L^2}.
\]
To see this, by the boundedness of $b_2$, \[
|\langle f \rangle^2 | \lesssim |V|^{-1} \int_{V} |f| = \langle |f| \rangle_{V}.
\]
Hence, it suffices to prove \[
\sum_{V} |\langle |f| \rangle_{V}| \|\Delta_{b_2} a\|_{L^2} \lesssim \|a\|_{BMO} \|f\|_{L^2}.
\]
Observing the above inequality, we see that by Carleson embedding theorem, all we need is to show that \{\|\Delta_{b_2} a\|_{L^2} \} is a Carleson sequence with constant \[
\lesssim \|a\|_{BMO}^2.
\]
Indeed, since \[
\|a\|_{BMO}^2 = \sup_{J} \frac{1}{|J|} \int_{J} |a - \langle a \rangle_{J}|^2 < \infty,
\]
for
any fixed dyadic $J$,

$$
\|a\|_{BMO}^2 \leq J \int_J |a - \langle a \rangle_J|^2 = \|\chi_J(a - \langle a \rangle_J)\|_2^2
$$

$$
\approx \sum_I \|\Delta_{I}^{b_2}(\chi_J(a - \langle a \rangle_J))\|_2^2
$$

$$
\geq \sum_{I \in D(J)} \|\Delta_{I}^{b_2}(\chi_J(a - \langle a \rangle_J))\|_2^2 = \sum_{I \in D(J)} \|\Delta_{I}^{b_2}(a - \langle a \rangle_J)\|_2^2
$$

$$
= \sum_{I \in D(J)} \|\Delta_{I}^{b_2} a\|_2^2.
$$

where the last equality is because $\Delta_{I}^{b_2}$ maps any constant function to 0. And this completes the proof. □

**Full paraproducts.** We now introduce a "real" bi-parameter $b$-adapted paraproduct, which is a natural generalization of the classical one-parameter one.

3.3. **Definition.** For $a \in BMO(\mathbb{R}^n \times \mathbb{R}^m)$, operator $\pi_{a}^{b_1,b_2}$ is called full paraproduct, defined as

$$
\pi_{a}^{b_1,b_2}(f) = \sum_{K \in D^n, V \in D^m} \langle f \rangle_{K \times V}^{b_1,b_2} M_b \Delta_{K}^{b_1} \Delta_{V}^{b_2} a.
$$

3.4. **Proposition.** Full paraproducts are bounded operators on $L^2(\mathbb{R}^n \times \mathbb{R}^m)$. Specifically,

$$
\|\pi_{a}^{b_1,b_2}(f)\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim \|a\|_{BMO(\mathbb{R}^n \times \mathbb{R}^m)} \|f\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)}.
$$

To prove this proposition, we need to first consider the space $H^1_b(\mathbb{R}^n \times \mathbb{R}^m)$, containing those functions $f$ such that $fb \in H^1(\mathbb{R}^n \times \mathbb{R}^m)$. It is easy to check that the dual space of $H^1_b(\mathbb{R}^n \times \mathbb{R}^m)$ is $BMO_b(\mathbb{R}^n \times \mathbb{R}^m)$, containing functions $f$ such that $f^{-1}b \in BMO(\mathbb{R}^n \times \mathbb{R}^m)$. It is well known that $H^1$ can be characterized using both martingale maximal function and square function with the norms being equivalent ([6]). Similarly, if we define a $b$-adapted maximal function

$$
f_b^*(x) = \sup_{p,q \in \mathbb{Z}} |E_p^{b_1} E_q^{b_2} f(x)| = \sup_{I \in D^n, J \in D^m} |E_I^{b_1} E_J^{b_2} f(x)|,
$$

then, we have the following fact

3.5. **Proposition.** A function $f$ belongs to $H^1_b(\mathbb{R}^n \times \mathbb{R}^m)$ if and only if $f_b^* \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$.

Now, define a $b$-adapted square function as

$$
S_b f(x) = (\sum_{p,q \in \mathbb{Z}} |\Delta_{p}^{b_1} \Delta_{q}^{b_2} f(x)|^2)^{1/2} = \left( \sum_{I \in D^n, J \in D^m} |\Delta_{I}^{b_1} \Delta_{J}^{b_2} f(x)|^2 \right)^{1/2}.
$$
and let the space $K_b^1(\mathbb{R}^n \times \mathbb{R}^m)$ consist of all the functions $f$ such that $S_b f \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$. We then have the following theorem.

3.6. Theorem. If $f \in K_b^1(\mathbb{R}^n \times \mathbb{R}^m)$, then $f \in H_b^1(\mathbb{R}^n \times \mathbb{R}^m)$. Moreover, for all $f \in K_b^1$, $\|f\|^*_L \leq \|S_b f\|_L$.

To prove Theorem 3.6, we use the idea of double martingale by Bernard and a technique involving atomic decomposition. See [1].

First, in our $b$-adapted case, the well known equivalence of $L^2$ norm between martingale maximal function and square function is still true. More specifically, we have

3.7. Proposition. If function $f \in L^2(\mathbb{R}^n \times \mathbb{R}^m)$, then both $f^*_b$ and $S_b f$ are in $L^2$, and their norms are equivalent to $\|f\|_L^2$.

Proof. Iteration of a well known one-parameter $L^2$ result (see [18]) gives

$$\|f\|^2_L \approx \sum_{p,q} \|\Delta^b_{p,q} f\|^2_{L^2}. $$

Hence,

$$\|S_b f\|^2_L = \int |S_b f|^2 = \sum_{p,q} \int |\Delta^b_{p,q} f|^2 = \sum_{p,q} \|\Delta^b_{p,q} f\|^2_{L^2} \approx \|f\|^2_L.$$

For martingale maximal function, $f \leq f^*_b$ a.e. gives $\|f\|_L \leq \|f^*_b\|_L$. On the other hand, by accretivity

$$f^*_b = \sup_{I,J} \left|E^b_{I_1} E^b_{J_1} f\right| \lesssim \sup_{I,J} \int_{I \times J} |f| \leq M^S f,$$

and the strong maximal function is bounded on $L^2$, it implies $\|f^*_b\|_L \lesssim \|f\|_L$. □

For simplicity, denote $f_{p,q} = E^b_{p} E^b_{q} f$, and for each pair $(p,q) \in \mathbb{Z} \times \mathbb{Z}$, let $\mathcal{F}_{p,q}$ be the $\sigma$-algebra generated by all the dyadic rectangles of size $2^{-p} \times 2^{-q}$.

3.8. Definition. The function $x \mapsto \tau(x) \subset \mathbb{Z} \times \mathbb{Z}$ is called a stopping time if $\{x : (p,q) \in \tau(x)\}$ is $\mathcal{F}_{p,q}$-measurable.

3.9. Definition. $a \in L^2(\mathbb{R}^n \times \mathbb{R}^m)$ is an atom of $K_b^1$ if there exists a stopping time $\tau$ such that

1. $\{|x : \tau(x) \neq \mathbb{Z} \times \mathbb{Z} \}| < \infty$;
2. Let $a_t = E_t^b a$, $t \in \mathbb{Z} \times \mathbb{Z}$, then $a_t(x) = 0$, $\forall t + 1 \in \tau(x)$;
3. $\|a^*_b\|_L \lesssim |\{x : \tau(x) \neq \mathbb{Z} \times \mathbb{Z}\}|^{1/2}$.

Note that if we call $F = \{x : \tau(x) \neq \mathbb{Z} \times \mathbb{Z}\}$, then from property (2) in the definition, both $a^*_b$ and $S_b a$ are supported on $F$. Also, such functions are called atoms because they have the following property.
3.10. **Proposition.** If $a$ is an atom, then $a \in C \cdot B$, where $B$ is the unit ball in $H^1_0$ or $K^1_b$, and $C$ is a universal constant independent of $a$.

**Proof.** Using the supports of $a^*_b$ and $S_b a$, Hölder inequality implies

$$\|a^*_b\|_{L^1} = \|a^*_b \chi_F\|_{L^1} \leq \|a^*_b\|_{L^2}|F|^{1/2} \lesssim 1.$$  

and

$$\|S_b a\|_{L^1} = \|S_b a \chi_F\|_{L^1} \leq \|S_b a\|_{L^2}|F|^{1/2} \approx \|a^*_b\|_{L^2}|F|^{1/2} \lesssim 1.$$ 

\[\square\]

We now state the theorem of atomic decomposition.

3.11. **Theorem.** Given $f \in K^1_b \cap L^2$, there exists a sequence of atoms $a^n$ and a sequence of scalars $\lambda_n$ such that

1. $f = \sum_n \lambda_n a^n$, a.e.
2. $\sum_n |\lambda_n| \lesssim \|f\|_{K^1_b}$.

Before stating the proof of Theorem 3.11, we show that this atomic decomposition result will imply Theorem 3.6.

**Proof.** (of Theorem 3.6) It suffices to show the result holds for $f \in L^2$. For any such function, atomic decomposition implies

$$f_t = \sum_n \lambda_n a^n_t, \text{ a.e. } \forall t \in \mathbb{Z} \times \mathbb{Z}.$$ 

Then,

$$f^*_b = \sup_t |f_t| = \sup_t \left| \sum_n \lambda_n a^n_t \right| \leq \sum_n |\lambda_n| \sup_t |a^n_t| = \sum_n |\lambda_n| (a^n)_b^*,$$

which implies

$$\|f^*_b\|_{L^1} \leq \sum_n |\lambda_n| \sup_n \|(a^n)_b^*\|_{L^1} \lesssim \sum_n |\lambda_n| \lesssim \|S_b f\|_{L^1}.$$ 

\[\square\]

We turn to the prove of Theorem 3.11.

**Proof.** (of Theorem 3.11) For any $n \in \mathbb{Z}$, let $F_n = \{x : S_b f(x) > 2^n\}$, and

$$\tau_n(x) = \{(p, q) : \forall t \leq (p, q), E_t(\chi_{F_n}) \leq 1/2\},$$

where $E_t$ is the classical expectation operator. It is easy to check that $\tau_n$ is a stopping time, and $\tau_n \subset \tau_{n+1}$.
For each $n$, define a new function $f^{\tau_n}(x) = \sum_{t \in \tau_n(x)} \Delta^b_{t-1} f(x)$, then

$$f^{\tau_{n+1}} - f^{\tau_n} = \sum_{t \in \tau_{n+1} \setminus \tau_n} \Delta^b_{t-1} f.$$ 

Using this, define

$$a^n = \frac{f^{\tau_{n+1}} - f^{\tau_n}}{2^n |\{x : \tau_n \neq \mathbb{Z} \times \mathbb{Z}\}|}, \quad \lambda_n = 2^n |\{x : \tau_n \neq \mathbb{Z} \times \mathbb{Z}\}|.$$ 

We claim that such $a^n$ and $\lambda_n$ satisfy all the properties required in the theorem.

To check property (2):

$$\sum_n |\lambda_n| = \sum_n 2^n |\{x : \tau_n \neq \mathbb{Z} \times \mathbb{Z}\}| = \sum_n 2^n |\{x : \sup_t |E_t(\chi_{F_n})| > 1/2\}|$$

$$\leq \sum_n 2^n \cdot 4 \int (\sup_t |E_t(\chi_{F_n})|)^2$$

$$= 4 \sum_n 2^n \|\chi_{F_n}^*\|^2_{L^2}$$

$$\lesssim 4 \sum_n 2^n \|\chi_{F_n}\|^2_{L^2}$$

$$= 4 \sum_n 2^n |\{x : S_b f(x) > 2^n\}| \lesssim \|S_b f\|_{L^1}.$$ 

In the above, the second line follows from Chebyshev Inequality, and the fourth line uses the $L^2$ boundedness of the classical martingale maximal function.

To check property (1): It suffices to check that

$$f = \sum_n (f^{\tau_{n+1}} - f^{\tau_n}) = \lim_{n \to \infty} f^{\tau_n} - \lim_{n \to -\infty} f^{\tau_n}, \text{ a.e.}$$

For the first limit, Chebyshev Inequality implies that

$$|F_n| \leq 2^{-n} \int S_b f = 2^{-n} \|S_b f\|_{L^1}.$$ 

So as $n \to \infty$, $|F_n| \to 0$ monotonically. Hence, $\|\sup_t |E_t(\chi_{F_n})|\|_{L^2}^2 \approx \|\chi_{F_n}\|_{L^2}^2 \to 0$.

By Fatou’s Lemma,

$$\int \lim inf_{n \to \infty} |(\chi_{F_n})^*|^2 \leq \lim inf_{n \to \infty} \int |(\chi_{F_n})^*|^2 = 0,$$

which implies $\lim_{n \to \infty} |E_t(\chi_{F_n})| = 0$ a.e. uniformly in $t$. So when $n$ is large enough, $\tau_n = \mathbb{Z} \times \mathbb{Z}$ a.e., i.e. $f^{\tau_n} = f$. 

For the second limit, if \( x \) is such that \( S_b f(x) = 0 \), then \( \Delta_b^1 f(x) = 0 \), \( \forall t \). Hence, \( f^{\tau_n}(x) = 0 \) \( \forall n \). Also, in this case, \( \forall q \) fixed, since \( f \in L^2(\mathbb{R}^n \times \mathbb{R}^m) \),

\[
\lim_{p \to -\infty} |E_p^{b_1} E_q^{b_2} f(x)| \lesssim \lim_{|t| \to \infty} \frac{\int_{|t|} |f|}{|I \times J|} \leq \lim_{|t| \to \infty} \frac{\|f\|_{L^2}}{|I \times J|^{1/2}} = 0.
\]

And similarly for \( E_p^{b_1} E_q^{b_2} f(x) \). So \( E_p^{b_1} \Delta_q^{b_2} f(x) = E_p^{b_1} \Delta_q^{b_2} f(x) = \lim_{p \to -\infty} E_p^{b_1} \Delta_q^{b_2} f(x) = 0 \), which means

\[
E_p^{b_1} E_q^{b_2} f(x) = E_p^{b_1} E_{q+1}^{b_2} f(x), \ \forall p, q.
\]

A similar limiting argument for the other variable implies \( E_p^{b_1} E_q^{b_2} f(x) = 0 \) \( \forall p, q \).

Hence,

\[
f(x) = \lim_{p,q \to \infty} f_{p,q}(x) = 0.
\]

Then the convergence is automatically true.

If \( x \) is such that \( S_b f(x) > 0 \), then for small enough \( n \), \( S_b f(x) > 2^n \), i.e. \( x \in F_n \subset X = \bigcup_n F_n = \{ x : S_b f(x) > 0 \} \). Also,

\[
\lim_{n \to \infty} f^{\tau_n}(x) = \lim_{n \to \infty} \sum_{t \in \tau_n} \Delta_{t-1}^b f(x) = \sum_{t \in \tau_n} \Delta_{t-1}^b f(x).
\]

We claim that all the terms appearing in the sum are 0, hence \( \lim_{n \to \infty} f^{\tau_n}(x) = 0 \).

For any \( t \in \tau_n(x) \), we have \( E_t(\chi_{F_n}) < 1/2 \), \( \forall n \). Let \( n \to -\infty \),

\[
E_t(\chi_X)(x) = \lim_{n \to -\infty} E_t(\chi_{F_n})(x) \leq 1/2.
\]

Say \( R = I \times J \) of size \( 2^{-t_1} \times 2^{-t_2} \) is the rectangle containing \( x \) of generation \( t \). Then \( R \) has nonempty intersection with \( X^c \) since otherwise \( E_t(\chi_X)(x) = 1 \). For any \( y \in R \cap X^c \), since \( S_b f(y) = 0 \), we have \( \Delta_{p,q}^b f(y) = 0 \), \( \forall p, q \).

However, since \( \Delta_{t-1}^b f(x) = \Delta_{t-1}^b f(y) \), it implies \( \Delta_{t-1}^b f(x) = 0 \), which proves the claim.

Then the only thing left to check is that all the \( a^n \) defined are indeed atoms.

To see this, firstly, \( a^n \in L^2 \). Indeed,

\[
\| f^{\tau_{n+1}} - f^{\tau_n} \|_{L^2}^2 \approx \int \sum_{t \in \tau_{n+1} \setminus \tau_n} |\Delta_{t-1}^b f|^2 \leq \| S_b f \|_{L^2}^2 \approx \| f \|_{L^2}^2.
\]

Secondly, just as how we argued for the second property above, we see that \(|\{ x : \tau_n(x) \neq \mathbb{Z} \times \mathbb{Z} \}| < \infty \).

Thirdly, if \( t + 1 \in \tau_n \), for any double integer \( s \) not satisfying \( s \leq t \), by a simple computation, we have

\[
E_t^{b_1} \Delta_{s-1}^{b_1} f = E_t^{b_1} (E_{s_1}^{b_1} E_{s_2}^{b_2} - E_{s_1-1}^{b_1} E_{s_2}^{b_2} - E_{s_1}^{b_1} E_{s_2-1}^{b_2} + E_{s_1}^{b_1} E_{s_2-1}^{b_2})f = 0.
\]
On the other hand, if \( s \leq t \), then \( s \in \tau_n \), hence,

\[
E^b_t \left( \sum_{s \leq t, s \in \tau_{n+1} \setminus \tau_n} \Delta^b_{s-1} f \right) = 0,
\]

which implies \( a^n_t = E^b_t(a^n) = 0 \).

Finally, to show \( \| (a^n)_b \|_{L^2}^2 \lesssim |\{ x : \tau_n(x) \neq \mathbb{Z} \times \mathbb{Z} \}|^{-1} \), it suffices to show

\[
\| S_b a^n \|_{L^2}^2 \lesssim |\{ x : \tau_n(x) \neq \mathbb{Z} \times \mathbb{Z} \}|^{-1},
\]

which is equivalent to

\[
\int \sum_{t \in \tau_{n+1} \setminus \tau_n} |\Delta^b_{t-1} f|^2 \lesssim 4^n |\{ x : \tau_n(x) \neq \mathbb{Z} \times \mathbb{Z} \}|.
\]

Write

\[
LHS = \int \sum_{t \in \tau_{n+1} \setminus \tau_n} |\Delta^b_{t-1} f|^2 \chi_{\{ \tau_n \leq 2^{n+1} \}} + \int \sum_{t \in \tau_{n+1} \setminus \tau_n} |\Delta^b_{t-1} f|^2 \chi_{\{ \tau_n > 2^{n+1} \}} = I + II.
\]

The first term can be dealt with trivially,

\[
I \leq 4^{n+1} \int_{\text{spt}(S_b a^n)} \chi_{\{ \tau_n \leq 2^{n+1} \}} \leq 4^{n+1} |\{ x : \tau_n(x) \neq \mathbb{Z} \times \mathbb{Z} \}|.
\]

For the second term, let \( \mathcal{D}_t \) denote all those dyadic rectangles of generation \( t \), then

\[
II = \sum_{t \in \tau_{n+1} \setminus \tau_n} \sum_{R \in \mathcal{D}_t} \int_R |\Delta^b_{t-1} f|^2 \chi_{\{ \tau_n > 2^{n+1} \}}
= \sum_{t \in \tau_{n+1} \setminus \tau_n} \sum_{R \in \mathcal{D}_t} |\Delta^b_{t-1} f|^2 |R| E_t(\chi_{F_{n+1}}) \chi_R
\leq \frac{1}{2} \sum_{t \in \tau_{n+1} \setminus \tau_n} \sum_{R \in \mathcal{D}_t} \int_R |\Delta^b_{t-1} f|^2
= \frac{1}{2} \int \sum_{t \in \tau_{n+1} \setminus \tau_n} |\Delta^b_{t-1} f|^2.
\]

In the above, the second lines follows from the fact that \( \Delta^b_{t-1} f \) is a constant on each \( R \), and the third line uses \( t \in \tau_{n+1} \). Combining \( I \) and \( II \) gives us

\[
\int \sum_{t \in \tau_{n+1} \setminus \tau_n} |\Delta^b_{t-1} f|^2 \lesssim 2 \cdot 4^{n+1} |\{ x : \tau_n(x) \neq \mathbb{Z} \times \mathbb{Z} \}|,
\]

which completes our proof for the theorem of atomic decomposition. \( \square \)
With the result of Theorem 3.6, we return to the full paraproducts, and give a proof of Proposition 3.4.

**Proof.** (of Proposition 3.4) For any \( f, g \in L^2(\mathbb{R}^n \times \mathbb{R}^m) \),

\[
|\langle \pi_{a,b}^b(f), g \rangle| = |\langle \sum_{K,V} \langle f \rangle_{K \times V} M_b \Delta_K^{b_1} \Delta_V^{b_2} a, g \rangle|
\]

\[
= |\langle a, M_b \sum_{K,V} \langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g \rangle|
\]

\[
\leq \|a\|_{BMO} \|M_b \sum_{K,V} \langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g \|_{H^1_b} = \|a\|_{BMO} \sum_{K,V} \langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g \|_{H^1_b}
\]

\[
\lesssim \|a\|_{BMO} \|S_b(\sum_{K,V} \langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g)\|_{L^1},
\]

where the last step in the above follows from Theorem 3.6. Hence, it suffices to show that

\[
\|S_b(\sum_{K,V} \langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g)\|_{L^1} \lesssim \|f\|_{L^2} \|g\|_{L^2}.
\]

To see this, notice that

\[
S_b^2(\sum_{K,V} \langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g) = \sum_{K,V} |\langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g|^2 \chi_K \otimes \chi_V
\]

\[
\leq |\sup_{K,V} \langle f \rangle_{K \times V} \chi_K \otimes \chi_V|^2 \sum_{K,V} |\Delta_K^{b_1} \Delta_V^{b_2} g|^2
\]

\[
\lesssim |M^S(f)|^2 S_b^2(g),
\]

where \( M^S(f) \) is the strong maximal function, which is bounded on \( L^2 \). Since \( S_b \) is also bounded on \( L^2 \), we have

\[
\|S_b(\sum_{K,V} \langle f \rangle_{K \times V} \Delta_K^{b_1} \Delta_V^{b_2} g)\|_{L^1} \lesssim \|M^S(f) S_b(g)\|_{L^1} \leq \|M^S(f)\|_{L^2} \|S_b(g)\|_{L^2} \lesssim \|f\|_{L^2} \|g\|_{L^2}.
\]

\( \square \)

**Mixed paraproducts.** Since we are working in the bi-parameter setting, there appears a new mixed type of \( b \)-adapted paraproducts which requires particular attention. Basically, it means we have an average on \( a \), and a difference on \( f \) with respect to one variable, and conversely with respect to the other.
3.12. **Definition.** For \( a \in BMO(\mathbb{R}^n \times \mathbb{R}^m) \), operator \( \tilde{\pi}^{b',b}_a \) is called a mixed paraproduct, defined as

\[
\tilde{\pi}^{b',b}_a(f) = \sum_{K \in D^n, V \in D^m} E^{b'_1}_K ((E^{b'_2}_V f) M_b \Delta^{b_1}_K \Delta^{b_2}_V a).
\]

3.13. **Proposition.** Mixed paraproducts are bounded operators on \( L^2(\mathbb{R}^n \times \mathbb{R}^m) \). Specifically,

\[
\| \tilde{\pi}^{b',b}_a(f) \|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim \|a\|_{BMO(\mathbb{R}^n \times \mathbb{R}^m)} \|f\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)}.
\]

Since we already have the \( b \)-adapted square function characterization of \( H^1_b \), this proposition can be proved in the same way as a similar result in [20].

**Proof.** For any \( f, g \in L^2(\mathbb{R}^n \times \mathbb{R}^m) \),

\[
|\langle \tilde{\pi}^{b',b}_a(f), g \rangle| = \sum_{K,V} \langle E^{b'_1}_K ((E^{b'_2}_V f) M_b \Delta^{b_1}_K \Delta^{b_2}_V a), g \rangle
\]

\[
= \sum_{K,V} \langle a, M_b (\Delta^{b_1}_K E^{b'_2}_V f) (\Delta^{b_2}_V E^{b'_1}_K g) \rangle
\]

\[
= |\langle a, \sum_{K,V} M_b (\Delta^{b_1}_K f)^{b'_2}_V \otimes (\Delta^{b_2}_V g)^{b'_1}_K \rangle|
\]

\[
\leq \|a\|_{BMO} \sum_{K,V} \| M_b (\Delta^{b_1}_K f)^{b'_2}_V \otimes (\Delta^{b_2}_V g)^{b'_1}_K \|_{H^1}
\]

\[
= \|a\|_{BMO} \sum_{K,V} \| (\Delta^{b_1}_K f)^{b'_2}_V \otimes (\Delta^{b_2}_V g)^{b'_1}_K \|_{H^1}
\]

\[
\lesssim \|a\|_{BMO} \| S_b (\sum_{K,V} (\Delta^{b_1}_K f)^{b'_2}_V \otimes (\Delta^{b_2}_V g)^{b'_1}_K) \|_{L^1}.
\]

We claim that

\[
\| S_b (\sum_{K,V} (\Delta^{b_1}_K f)^{b'_2}_V \otimes (\Delta^{b_2}_V g)^{b'_1}_K) \|_{L^1} \lesssim \|f\|_{L^2} \|g\|_{L^2}.
\]
To see this, note that

\[
S^2_b \left( \sum_{K,V} \langle \Delta_{K}^{b_1} f \rangle_{Y_i} \otimes \langle \Delta_{V}^{b_2} g \rangle_{K} \right)
\]

\[
= \sum_{K,V} |\Delta_{K}^{b_1}(f)^{b'_1}_{Y_i} \otimes \Delta_{V}^{b_2}(g)^{b'_2}_{K}|^2 \chi_K \otimes \chi_V
\]

\[
\leq \left( \sum_{K} \sup_{V} |\Delta_{K}^{b_1}(f)^{b'_1}_{Y_i}|^2 \chi_K \otimes \chi_V \right) \cdot \left( \sum_{V} \sup_{K} |\Delta_{V}^{b_2}(g)^{b'_2}_{K}|^2 \chi_K \otimes \chi_V \right)
\]

\[
\leq \left( \sum_{K} \sup_{V} |\Delta_{K}^{b_1}(f)^{b'_1}_{Y_i}|^2 \chi_K \otimes \chi_V \right) \cdot \left( \sup_{V} \sum_{K} |\Delta_{V}^{b_2}(g)^{b'_2}_{K}|^2 \chi_K \otimes \chi_V \right)
\]

\[
:= |S_b M^{b_2}(f)|^2 |M^{b_2} S_b(g)|^2,
\]

where the last two operators are just formally defined, but not the compositions of the square functions and maximal functions. Since pointwisely, \(|M^{b_1}(S_b g)| \leq |S_b(M^{b_1} g)|\), by symmetry, it suffices to prove that \(S_b M^{b_2} : L^2 \rightarrow L^2\). And this is true because

\[
\|S_b M^{b_2} f\|_{L^2} = \|(\sum_{K} \sup_{V} |\Delta_{K}^{b_1}(f)^{b'_1}_{Y_i}|^2 \chi_K \otimes \chi_V)^{1/2}\|_{L^2}
\]

\[
\lesssim \|(\sum_{K} |M_2(\Delta_{K}^{b_1} f)^{b'_1}_{K}|^2 \chi_K)^{1/2}\|_{L^2}
\]

\[
\lesssim \sum_{i=1}^{2^n} \left( \int_{\mathbb{R}^n} \left\| \left( \sum_{K} \chi_{K_i}(x) \otimes M_2(\Delta_{K}^{b_1} f)^{b'_1}_{K_i}(y) \right)^{1/2} \right\|^2_{L^2(\mathbb{R}^m)} dx \right)^{1/2}
\]

\[
\lesssim \sum_{i=1}^{2^n} \left( \int_{\mathbb{R}^n} \left\| \left( \sum_{K} |\Delta_{K}^{b_1} f|^2 \chi_{K_i} \right)^{1/2} \right\|^2_{L^2(\mathbb{R}^m)} dx \right)^{1/2}
\]

\[
\lesssim \left( \int_{\mathbb{R}^m} \left\| \left( \sum_{K} |\Delta_{K}^{b_1} f|^2 \right)^{1/2} \right\|^2_{L^2(\mathbb{R}^n)} dy \right)^{1/2}
\]

\[
= \left( \int_{\mathbb{R}^m} \|f_y\|^2_{L^2(\mathbb{R}^n)} dy \right)^{1/2} = \|f\|_{L^2}.
\]

In the above, \(M_2\) means the Hardy-Littlewood maximal function with respect to the second variable. In the fourth line, we used the Fefferman-Stein inequality. And in the sixth line, the operator \(S_b\) is the one-parameter \(b_1\)-adapted square function, defined as \(S_b f = (\sum_I |\Delta_{I}^{b_1} f|^2)^{1/2}\). It is straightforward to see that \(S_b\) is
an $L^2$ isometry up to some constant, which implies the seventh line in the above, where $f_y(x)$ denotes $f(x, y)$.

Hence, the $L^2$ boundedness of the mixed paraproduct is fully justified. □

4. MAIN THEOREM AND THE STRATEGY

We return to the main theorem of this paper. We will prove that, under the assumptions stated in Section 2, $T$ is bounded on $L^2(\mathbb{R}^n \times \mathbb{R}^m)$ with the operator norm depending only on the constants appearing in the above weak assumptions. By density and boundedness of $b, b'$, it suffices to show that for any $C_0^\infty$ functions $f, g$, there is a universal constant $C$ such that

$$|\langle M'_bTM_bf, g \rangle| \leq C \|f\|_{L^2} \|g\|_{L^2}.$$  

To prove this, recall that Martikainen [16] gave an averaging formula for the bilinear form $\langle Tf, g \rangle$ using a probabilistic concept called "goodness" of cubes. Here, if we decompose $f$ using the new defined $b$-adapted martingale difference instead, there is a natural generalization of the averaging formula as follows.

4.1. Proposition.

$$\langle M'_bTM_bf, g \rangle = \frac{1}{\pi_n \pi_m} E_{\omega^n} E_{\omega^m} \sum_{I_1, I_2 \in D^n, J_1, J_2 \in D^m} \chi_{\text{good}}(\text{small}(I_1, I_2)) \chi_{\text{good}}(\text{small}(J_1, J_2)) \langle M'_bTM_b\Delta_{I_1}^{b_1} \Delta_{J_1}^{b_1'} f, \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2'} g \rangle.$$  

To understand the above formula, recall that in [9], a cube $I \in D^n$ is called bad if there exists $\bar{I} \in D^n$ so that $\ell(\bar{I}) \geq 2^r \ell(I)$ and $d(I, \partial \bar{I}) \leq 2\ell(I)^\gamma \ell(\bar{I})^{1-\gamma}$. $\gamma_n = \delta/(2n + 2\delta)$, where $\delta > 0$ appears in the kernel estimates. And $\pi_n \pi_m := \mathbb{P}_{\omega^n}(I + \omega^n \text{ is good})$ is independent of $I \in D^n$. By lemma 2.3 in [9], the parameter $r$ can be chosen large enough such that $\pi_n \pi_m > 0$. Moreover, for a fixed $I \in D^n$ the position of $I + \omega^n$ depends on $\omega^n$ with $2^{-i} < \ell(I)$, while the goodness of $I + \omega^n$ depends on $\omega_i^n$ with $2^{-i} \geq \ell(I)$. Hence, they are independent. The proof of Proposition 4.1 is identical to the proof of Proposition 2.1 in [16], which we omit here.

Note that as in [9] and [16], we do not need to justify that the sum on the right hand side converges to the left hand side, which is the only place throughout the paper where we use the a priori $L^2 \rightarrow L^2$ boundedness of $T$. Indeed, by the convergence of expectation operators in $L^2$, the boundedness of $T$ will easily imply the convergences in the formula. However, when dealing with specific operators in practice, sometimes we can prove the convergence of the formula without assuming the boundedness assumption.
For example, if $T$ is canonically associated with a standard antisymmetric kernel $K(x, y)$, in the sense that
\[ K(x, y) = -K(y_1, x_2, x_1, y_2) = -K(x_1, y_2, y_1, x_2), \]
and $K$ satisfies all the size and Hölder conditions.

Then for any $f = f_1 \otimes f_2, g = g_1 \otimes g_2 \in C_0^\infty(\mathbb{R}^n \times \mathbb{R}^m),$
\[
\langle T f, g \rangle = \int \int \int \int K(x, y) f(y) g(x) \, dx \, dy
\]
is well defined. Hence, we automatically have the full and partial kernel representations. Also, by antisymmetry,
\[
\langle T(\chi_K \otimes \chi_V), \chi_K \otimes \chi_V \rangle = 0,
\]
which corresponds to the weak boundedness property for $b = b' = 1$. With these observations in mind, it is not hard to show that for any $f, g \in C_0^\infty$ and any fixed dyadic grid,
\[
\langle T f, g \rangle = \lim_{N \to \infty} \sum_{|p_i, q_i| < N} \langle T \Delta_{p_1} \Delta_{q_1} f, \Delta_{p_2} \Delta_{q_2} g \rangle.
\]
So the a priori boundedness of $T$ is not necessary any more.

With the averaging formula, it suffices to bound the sum on the right hand side uniformly for any fixed random grids, to do which, we will divide the sum into different parts according to the relative positions of the cubes, and discuss different cases one by one. By symmetry, except for one mixed case ($\ell(I_1) \leq \ell(J_1) > \ell(J_2)$), all the other cases are symmetric to ($\ell(I_1) \leq \ell(I_2), \ell(J_1) \leq \ell(J_2)$), which we will start with.

For the relative position of $I_1, I_2$, there are four different cases: separated (i.e. $\text{dist}(I_1, I_2) > \ell(I_1)^\gamma \ell(I_2)^{1-\gamma_0}$), inside (i.e. $I_1 \subset I_2$), equal, nearby (i.e. $\text{dist}(I_1, I_2) \leq \ell(I_1)^\gamma \ell(I_2)^{1-\gamma_0}$). Similarly, there are also four different cases for the second variable. Again using symmetry, it suffices to analyze the following ten cases:

- separated/separated, separated/inside, separated/equal, separated/nearby,
- inside/inside, inside/equal, inside/nearby,
- equal/equal, equal/nearby, nearby/nearby.

In preparation, we state two control lemmas here which will be repeatedly used when we deal with different cases in the following. For simplicity of notation,
write

\[
\sum_{I_1, I_2 \in K} = \sum_{I_1, I_2 \in \mathcal{D}^n(K)} \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \cdot \frac{|J_1|^{1/2} |J_2|^{1/2}}{|V|} \cdot \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \lesssim \|f\|_{L^2} \cdot \|g\|_{L^2}.
\]

where \( K \in \mathcal{D}^n \) and \( i_1, i_2 \in \mathbb{N} \).

**4.2. Lemma. (Full control lemma)** For fixed \( i_1, i_2, j_1, j_2 \in \mathbb{N} \) and any \( f \in L^2(\mathbb{R}^n \times \mathbb{R}^m) \), \( g \in L^2(\mathbb{R}^n \times \mathbb{R}^m) \),

\[
\sum_{K \in \mathcal{D}^n} \sum_{I_1, I_2 \in K} \sum_{J_1, J_2 \in V} |K|^{-1/2} |I_1|^{1/2} |I_2|^{1/2} |J_1|^{1/2} |J_2|^{1/2} \cdot \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \lesssim \|f\|_{L^2} \cdot \|g\|_{L^2}.
\]

**Proof.** It follows as a consequence of Hölder inequality.

\[
LHS \leq \left( \sum_{K, V} \sum_{I_1, J_1 \in K} \sum_{I_2, J_2 \in V} \frac{|I_1| |I_2|}{|K|^2} \cdot \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \right)^{1/2} \cdot \left( \sum_{K, V} \sum_{I_1, J_1 \in K} \sum_{I_2, J_2 \in V} \frac{|I_1| |I_2|}{|K|^2} \cdot \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \right)^{1/2}
\]

\[
= (2^{i_1 n} 2^{j_1 m}) \sum_{K, V} \sum_{I_1 \in K} \sum_{J_1 \in V} 2^{-i_1 n} 2^{-j_1 m} \cdot \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|}
\]

\[
= \left( \sum_{K, V} \sum_{I_1 \in K} \sum_{J_1 \in V} \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \right)^{1/2}
\]

In the last step above, we used the \( L^2 \) property of \( b \) adapted double martingale difference. \( \Box \)

**4.3. Lemma. (Partial control lemma)** For fixed \( i_1, i_2, j_1, j_2 \in \mathbb{N} \) and any \( f \in L^2(\mathbb{R}^n \times \mathbb{R}^m) \), \( g \in L^2(\mathbb{R}^n \times \mathbb{R}^m) \),

\[
\sum_{K \in \mathcal{D}^n} \sum_{I_1, I_2 \in K} \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \cdot \frac{|J_1|^{1/2} |J_2|^{1/2}}{|V|} \cdot \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \lesssim \|f\|_{L^2} \cdot \|g\|_{L^2},
\]

and

\[
\sum_{V \in \mathcal{D}^m} \sum_{J_1, J_2 \in V} \frac{|J_1|^{1/2} |J_2|^{1/2}}{|V|} \cdot \frac{|J_1|^{1/2} |J_2|^{1/2}}{|V|} \cdot \frac{\Delta t_1^b \Delta t_2^b f}{|L^2|} \cdot \frac{\Delta t_2^b \Delta t_2^b g}{|L^2|} \lesssim \|f\|_{L^2} \cdot \|g\|_{L^2}.
\]
These two inequalities are symmetric, and they can both be derived using a similar technique as for the above lemma. The only difference here is that we need to use the $L^2$ property of the $b$-adapted martingale difference of only one variable instead.

Before we move on to the main part of the proof of the theorem, i.e. the case by case estimate of summands in the averaging formula, let’s look at an example to see how our theory fits into some known results of boundedness of bi-parameter singular integral operators.

Consider operators associated with antisymmetric standard kernels. Journé, in [14], proved that if $K = LA$, the bicommutator of Calderón-Coifman type, where $L$ is any standard antisymmetric function, and

$$
\tilde{A}(x, y) = \frac{A(x_1, x_2) + A(y_1, y_2) - A(y_1, x_2) - A(x_1, y_2)}{(x_1 - y_1)(x_2 - y_2)},
$$

for some $A : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{C}$ such that $\partial_2^i A \in L^\infty$, then, the $L^2 \to L^2$ boundedness of the operator associated to $L$ implies $T1 \in BMO$, as well as the other $BMO$ conditions. It is also not hard to verify directly that $T$ satisfies the weak boundedness property and the four diagonal $BMO$ assumptions. (All of them are actually zero!) Hence, by our main theorem, $T$ is bounded on $L^2$ with operator norm controlled by the weak assumptions.

5. Separated/Separated: $\sigma_{\text{out/\text{out}}}$

Define $I_1 \cup I_2 = \bigcap_{K \in \mathcal{D}^n, K \supset I_1 \cup I_2} K$, i.e. the smallest $K$ such that $I_1 \cup I_2 \subset K$, and similarly for $J_1 \cup J_2$. Then since both of them are separated and $I_1, J_1$ are good, it is proved in [9] by Hytönen that $\ell(I_1) \ell(K)^{1-\gamma} \lesssim \text{dist}(I_1, I_2)$ and $\ell(J_1) \ell(V)^{1-\gamma} \lesssim \text{dist}(J_1, J_2)$.

Hence, we can write

$$
\sigma_{\text{out/\text{out}}} = \sum_{i_2=1}^{\infty} \sum_{i_1=1}^{\infty} \sum_{j_2=1}^{\infty} \sum_{j_1=1}^{\infty} \sum_{K, V} \sum_{I_1, I_2 \subset K} \sum_{J_1, J_2 \subset V} \langle M_b T M_b \Delta_{I_1}^b \Delta_{J_1}^b f, \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \rangle.
$$

The main goal of this section is to show that the following inequality holds.

5.1. Proposition.

$$
|\langle M_b T M_b \Delta_{I_1}^b \Delta_{J_1}^b f, \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \rangle| \\
\lesssim 2^{-i_1 \delta/2} 2^{-j_1 \delta/2} |I_1|^{1/2} |I_2|^{1/2} |J_1|^{1/2} |J_2|^{1/2} \frac{||\Delta_{I_1}^b \Delta_{J_1}^b f||_{L^2} ||\Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g||_{L^2}}{|K| |V|}.
$$
If this is true, then by the full control lemma we stated in the beginning, $\sigma_{\text{out}}$ can be bounded by $\|f\|_{L^2} \|g\|_{L^2}$.

**Proof.** Since the two functions are well separated on both variables, by the full kernel representation,

$$LHS = \left| \int_{I_1 \times J_1} \int_{I_2 \times J_2} K(x, y) \Delta_{i_1}^{b_1} \Delta_{j_1}^{b_2} f(y) b(y) \Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g(x) b'(x) \, dxdy \right|.$$

Using the cancellation properties of the martingale differences, we can replace $K(x, y)$ in the above by

$$K(x, y) - K(x, y_1, c_{I_1}) - K(x, c_{I_1}, y_2) + K(x, c_{I_1}, c_{I_2}).$$

Since $|y_1 - c_{I_1}| \leq \ell(I_1)/2 \leq \frac{1}{2} \ell(I_1)^\gamma \ell(I_2)^{1-\gamma} \leq \text{dist}(I_1, I_2)/2 \leq |x_1 - c_{I_1}|/2$, and similarly $|y_2 - c_{I_1}| \leq |x_2 - c_{I_1}|/2$, by the full Hölder condition,

$$|5.2| \lesssim \frac{|y_1 - c_{I_1}|^\delta}{|x_1 - c_{I_1}|^{n+\delta}} \frac{|y_2 - c_{I_1}|^\delta}{|x_2 - c_{I_1}|^{m+\delta}}$$

$$\lesssim \ell(I_1)^\delta \text{dist}(I_1, I_2)^{-n-\delta} \ell(J_1)^\delta \text{dist}(J_1, J_2)^{-m-\delta}$$

$$\lesssim \ell(I_1)^{\delta/2} \ell(K)^{-\gamma/2} |K|^{-1} \ell(J_1)^{\delta/2} \ell(V)^{-\gamma/2} |V|^{-1}$$

$$= 2^{-i_1 \delta/2} 2^{-j_1 \delta/2} |K|^{-1} |V|^{-1},$$

where for the third inequality we used $\ell(I_1)^\gamma \ell(K)^{1-\gamma} \lesssim \text{dist}(I_1, I_2)$ and $\ell(J_1)^\gamma \ell(V)^{1-\gamma} \lesssim \text{dist}(J_1, J_2)$. Then, by Hölder inequality and the boundedness of $b, b'$, this implies

$$LHS \lesssim 2^{-i_1 \delta/2} 2^{-j_1 \delta/2} |K|^{-1} |V|^{-1} \left( \int_{I_1 \times J_1} |\Delta_{i_1}^{b_1} \Delta_{j_1}^{b_2} f(y)| \, dy \right) \left( \int_{I_2 \times J_2} |\Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g(x)| \, dx \right)$$

$$\leq \text{RHS}.$$

\[ \square \]

6. **Separated/Inside: $\sigma_{\text{out}}/\text{in}$**

Since $J_1 \subseteq J_2$, $J_1$ is contained in some child of $J_2$, which we denote by $J_{2,1}$. Then $\Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g$ is constant with respect to $x_2$ on $J_{2,1}$, and we have

$$\langle M_b TM_a \Delta_{i_1}^{b_1} \Delta_{j_1}^{b_2} f, \Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g \rangle$$

$$= \langle M_b TM_a \Delta_{i_1}^{b_1} \Delta_{j_1}^{b_2} f, (\chi_{J_{2,1}} + \chi_{J_{2,1}^c}) \Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g \rangle$$

$$= \langle M_b TM_a \Delta_{i_1}^{b_1} \Delta_{j_1}^{b_2} f, \chi_{J_{2,1}} (\Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g - \langle \Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g \rangle_{J_{2,1}}) \rangle$$

$$+ \langle M_b TM_a \Delta_{i_1}^{b_1} \Delta_{j_1}^{b_2} f, \langle \Delta_{i_2}^{b_1} \Delta_{j_2}^{b_2} g \rangle_{J_{2,1}} (x_1) \otimes 1(x_2) \rangle$$

$$:= I + II.$$
where \( \langle f \rangle_{b_j} \) denotes the \( b_j \)-adapted average of \( f \) over \( J \) with respect to the second variable: \( (\int_J b_j)^{-1}(\int_J fb_j) \).

Write

\[
\sigma_{\text{out / in}} = \sum_{i_2=1}^\infty \sum_{i_1=i_2}^\infty \sum_{j_1=1}^\infty \sum_{J_2 \in D^m} \sum_{\text{dist}(I_1,J_2) > \ell(J_1)^\gamma m \ell(J_2)^{1-\gamma m}} \sum_{J_1 \subset J_2} I + II
\]

\[
:= \sigma_{\text{out / in}}' + \sigma_{\text{out / in}}''.
\]

**Part \( \sigma_{\text{out / in}}' \).** In order to bound \( \sigma_{\text{out / in}}' \) by \( \|f\|_{L^2} \|g\|_{L^2} \), by the full control lemma, it suffices to prove the following.

6.1. **Proposition.**

\[
|I| \lesssim \frac{|I_1|^{1/2}|I_2|^{1/2} |J_1|^{1/2}}{|K|} \frac{2^{-i_2 \delta/2} 2^{-j_1 \delta/2}}{|J_2|^{1/2}} \|\Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f\|_{L^2} \|\Delta_{I_2}^{b_1'} \Delta_{J_2}^{b_2'} g\|_{L^2}.
\]

**Proof.** Case 1: \( \ell(J_1) < 2^{-r} \ell(J_2) \).

The two functions in the pairing are separated in both variables, which enables us to use the full kernel representation:

\[
I = \int_{I_1 \times J_1} \int_{I_2 \times J_2} K(x,y) \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f(y) b(y)(\Delta_{I_2}^{b_1'} \Delta_{J_2}^{b_2'} g) dx dy.
\]

Since in this case, the size of \( J_1 \) is "significantly" small compared with \( J_2 \), by the goodness of \( J_1 \), \( \text{dist}(J_1, J_{2,1}) \geq 2 \ell(J_1)^{\gamma m} \ell(J_{2,1})^{1-\gamma m} \geq \ell(J_1)^{\gamma m} \ell(J_2)^{1-\gamma m} \), which implies good separation on both variables. Hence, using the cancellation property in \( y \) variable, we can change the kernel \( K(x,y) \) in the above to

\[
K(x,y) - K(x,y_1, c_{J_1}) - K(x, c_{I_1}, y_2) + K(x, c_{I_1}, c_{J_1}).
\]
By Hölder condition and a similar computation as in the Separated/Separated case,

$$|I| \lesssim \ell(I_1)^{\delta/2} \ell(K)^{-\delta/2} |K|^{-1} \ell(J_1)^{\delta} \left( \int_{I_1 \times J_1} |\Delta^{b_1}_{J_1} \Delta^{b_2}_{J_1} f| \, dy \right).$$

$$\left( \int_{I_2 \times J_{2,1}^c} \frac{1}{|x_2 - c_{J_1}|^{m+\delta}} |\Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g - \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \rangle_{J_{2,1}^c} | \, dx \right) \leq 2^{-i_1 \delta/2} |K|^{-1} \ell(J_1)^{\delta} \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \widehat{J}_{2,1}^c \rangle_{J_{2,1}^c} .$$

$$\left( \int_{I_2 \times J_{2,1}^c} \frac{1}{|x_2 - c_{J_1}|^{m+\delta}} \left( |\Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g| + \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \rangle_{J_{2,1}^c} \right) \, dx \right) \leq 2^{-i_1 \delta/2} \frac{|I_1|^{1/2}}{|K|} |J_2|^{1/2} \ell(J_1)^{\delta} \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \widehat{J}_{2,1}^c \rangle_{J_{2,1}^c} .$$

$$\left( \int_{I_2 \times J_{2,j}} \frac{1}{|x_2 - c_{J_1}|^{m+\delta}} \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \rangle_{J_{2,j}} \, dx + \sum_{j=2}^m \int_{I_2 \times J_{2,j}} \frac{1}{|x_2 - c_{J_1}|^{m+\delta}} \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \rangle_{J_{2,j}} \, dx \right) \leq 2^{-i_1 \delta/2} \frac{|I_1|^{1/2}}{|K|} |J_2|^{1/2} \ell(J_1)^{\delta} \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \widehat{J}_{2,1}^c \rangle_{J_{2,1}^c} \cdot$$

$$\left( \int_{I_2 \times J_{2,j}} \frac{1}{|x_2 - c_{J_1}|^{m+\delta}} \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \rangle_{J_{2,j}} \, dx \right) \leq 2^{-i_1 \delta/2} \frac{|I_1|^{1/2}}{|K|} |J_2|^{1/2} \ell(J_1)^{\delta/2} \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \widehat{J}_{2,1}^c \rangle_{J_{2,1}^c} .$$

$$= \text{LHS},$$

where in the third line, $J_{2,j}$ denotes all the children of $J_2$ except $J_{2,1}$, and we used the fact that $\Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g$ is constant with respect to $x_2$ on each child of $J_2$. And the fourth line follows from the estimate of those averages of $\Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g$.

**Case 2:** $2^{-r} \ell(J_2) \leq \ell(J_1) \leq \ell(J_2)$.

Let’s further split $I$ into two parts:

$$I' = \langle M_{g' T} M_b \Delta^{b_1}_{I_1} \Delta^{b_2}_{I_1} f, \chi_{3J_1 \cap J_{2,1}^c} (\Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g - \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \rangle_{J_{2,1}^c}) \rangle,$$

$$I'' = \langle M_{g' T} M_b \Delta^{b_1}_{I_1} \Delta^{b_2}_{I_1} f, \chi_{(3J_1)^c \cap J_{2,1}^c} (\Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g - \langle \Delta^{b_1}_{J_2} \Delta^{b_2}_{J_2} g \rangle_{J_{2,1}^c}) \rangle.$$
size of $J_1, J_2$ are comparable,

$$
|I''| \lesssim 2^{-ii\delta/2} \frac{|I_1|^{1/2} |J_2|^{1/2} |J_1|^{1/2}}{|K|} \left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \right\|_{L^2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2} \ell(J_1)^4 \int_{(3J_1)^c} \frac{1}{|x_2 - c_{J_1}|^{n+\delta}}
$$

$$
\lesssim 2^{-ii\delta/2} \frac{|I_1|^{1/2} |J_2|^{1/2} |J_1|^{1/2}}{|K|} \left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \right\|_{L^2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2}
$$

$$
\lesssim 2^{-ii\delta/2 - ji\delta/2} \frac{|I_1|^{1/2} |J_2|^{1/2} |J_1|^{1/2}}{|K|} \left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \right\|_{L^2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2}.
$$

Hence, the only thing left to deal with is $I'$. Since now the separation in the second variable is not good enough, we have to use the mixed Hölder-size condition instead. Again, in the full kernel representation, by cancellation property we can change the kernel to $K(x, y) - K(x, c_{J_1}, y_2)$, then

$$
|I'| \lesssim \int_{I_1 \times J_1} \int_{J_2 \times (3J_1 \cap J_2^c)} \frac{\ell(I_1)^4}{|x_1 - c_{J_1}|^{n+\delta}} \frac{\left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f(y) \right\|}{|x_2 - y_2|^m} \left( \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g(x) \right\| + \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g(x) \right\|_{J_2^c} \right) \, dx \, dy
$$

$$
\lesssim 2^{-ii\delta/2} |K|^{-1} \int_{I_1 \times J_1} \int_{J_2 \times (3J_1 \cap J_2^c)} \frac{\left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f(y) \right\|}{|x_2 - y_2|^m} \left( \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g(x) \right\| + \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g(x) \right\|_{J_2^c} \right) \, dx \, dy
$$

$$
\lesssim 2^{-ii\delta/2} |K|^{-1} \int_{I_1 \times J_1} \int_{J_2 \times (3J_1 \cap J_2^c)} \frac{\left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f(y) \right\|}{|x_2 - y_2|^m} \left( \sum_{j=2}^{2^n} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g_{J_2^c} \right\|_{J_2^c} \right) \, dx \, dy
$$

$$
\lesssim 2^{-ii\delta/2} |J_2|^{-1/2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2} \int_{I_1 \times J_1} \int_{3J_1 \cap J_2^c} \frac{\left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f(y) \right\|}{|x_2 - y_2|^m} \, dx \, dy
$$

$$
= 2^{-ii\delta/2} |J_2|^{-1/2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2} \sum_{i=1}^{2^n} \sum_{j=1}^{2^n} \int_{I_1 \times J_1} \int_{3J_1 \cap J_2^c} \frac{\left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f(y) \right\|_{J_1 \times J_1}}{|x_2 - y_2|^m} \, dx \, dy
$$

$$
\lesssim 2^{-ii\delta/2} |I_1|^{1/2} |J_2|^{-1/2} \left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \right\|_{L^2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2} \int_{J_1 \times J_1} \int_{3J_1 \cap J_1} \frac{1}{|x_2 - y_2|^m} \, dx \, dy
$$

$$
\lesssim 2^{-ii\delta/2} |I_1|^{1/2} |J_2|^{1/2} \left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \right\|_{L^2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2}
$$

$$
\lesssim 2^{-ii\delta/2 - ji\delta/2} |I_1|^{1/2} |J_2|^{1/2} \left\| \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \right\|_{L^2} \left\| \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \right\|_{L^2}.
$$

In the above, the fifth line is because $\Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f$ is a constant on each child of $I_1 \times J_1$, and the last line follows from the fact that the size of $J_1, J_2$ are comparable. This completes the proof of the proposition. \qed
Part $\sigma''_{out/in}$. For the part $\sigma''_{out/in}$ we are going to rewrite it into a form containing a partial $b$-adapted paraproduct. Rewrite

$$
\sigma''_{out/in} = \sum_{i_2=1}^{\infty} \sum_{i_1=i_2}^{\infty} \sum_{R=1}^{K} \sum_{dist(I_1,I_2) > t(I_1)^{\gamma_n} t(I_2)^{1-\gamma_n} J_1 \subseteq J_2} (i_1,i_2) \sum_{I_1 \cap I_2 = K} II,
$$

and first look at the innermost sum.

$$
\sum_{J_1 \subseteq J_2} \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_1} f, \langle \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g \rangle_{J_2 \subseteq J_1}^{b_2} (x_1) \otimes 1(x_2) \rangle
$$

$$
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_1} f, 1 \rangle_2 \langle \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g \rangle_{J_1}^{b_2} \rangle_1
$$

$$
= \sum_{J_1} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_1} f, 1 \rangle_2 \langle \sum_{J_2 \supseteq J_1} \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g \rangle_{J_1}^{b_2} \rangle_1
$$

$$
= \sum_{J_1} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_1} f, 1 \rangle_2 \langle \Delta_{I_2}^{b_2} g \rangle_{J_1}^{b_2} \rangle_1
$$

$$
= \sum_{V} \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{V}^{b_2} f, \langle \Delta_{I_2}^{b_2} g \rangle_{V}^{b_2} \otimes 1 \rangle
$$

$$
= \langle \Delta_{I_1}^{b_1} f, \sum_{V} M_b \Delta_{I_1}^{b_1} \Delta_{V}^{b_2} T^w (b_1' \langle \Delta_{I_2}^{b_2} g \rangle_{V}^{b_2} \otimes b_2) \rangle
$$

Notice that $\Delta_{I_1}^{b_1} f, \Delta_{I_2}^{b_2} g$ are constant with respect to $x_1$ on each child of $I_1, I_2$, respectively. If we decompose the above pairing into parts that are restricted on children of $I_1, I_2$, then

$$
\sum_{J_1 \subseteq J_2} II = \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \langle \chi_{I_1,t} \Delta_{I_1}^{b_1} f, \sum_{V} \langle \Delta_{I_2}^{b_2} g \rangle_{I_2,t,k}^{b_2} \rangle V M_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} T^w (\chi_{I_2,t,k} b_1' \otimes b_2) \rangle
$$

$$
= \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \langle \chi_{I_1,t} \Delta_{I_1}^{b_1} f, b_1 \otimes \chi_{h_{I_1,t,I_2,k}}^{b_2} (\Delta_{I_2}^{b_2} g) \rangle
$$

where $h_{I_1,t,I_2,k}(x_2) = \langle \Delta_{I_1}^{b_1} T^w (\chi_{I_2,k} b_1' \otimes b_2) \rangle_{I_1,t}$, and the following lemma guarantees that the partial paraproduct is well defined.

6.2. Lemma. $h_{I_1,t,I_2,k}$ is in $BMO(\mathbb{R}^m)$, and satisfies

$$
\|h_{I_1,t,I_2,k}\|_{BMO} \lesssim 2^{-i_1 \delta/2} |K|^{-1} |I_2|.
$$
We will assume the lemma to be true for the moment and prove it at the end of this section. The above pairing can be further rewritten as:

\[
\sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \langle \Delta_{I_1}^{b_1} f | I_{1,t}, \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_2}^{b_2} g) | I_{2,k} \rangle_2
\]

\[
= \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \langle \Delta_{I_1}^{b_1} f | I_{1,t}, \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_2}^{b_2} g) | I_{2,k} \rangle_2
\]

\[
= \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \left\langle \frac{\chi_{I_{2,k}}}{I_{2,k}} \otimes \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_1}^{b_1} f | I_{1,t}), \Delta_{I_2}^{b_2} g \right\rangle
\]

\[
= \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \left\langle \Delta_{I_1}^{b_2} \left( \frac{\chi_{I_{2,k}}}{I_{2,k}} \right) \otimes \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_1}^{b_1} f | I_{1,t}), \Delta_{I_2}^{b_2} g \right\rangle
\]

\[
= \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \left\langle b_1' \Delta_{I_2}^{b_2} \left( \frac{b_1^{-1} \chi_{I_{2,k}}}{I_{2,k}} \right) \otimes \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_1}^{b_2} f | I_{1,t}), \Delta_{I_2}^{b_2} g \right\rangle
\]

Then,

\[
|\sigma''_{\text{out/in}}| = \left| \sum_{i_2=1}^{2^n} \sum_{i_2=1}^{2^n} \sum_{K \in \text{dist}(I_1, I_2)} \sum_{I_{1,t}, I_{2,k} \in K} \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \langle b_1' \Delta_{I_2}^{b_2} \left( \frac{b_1^{-1} \chi_{I_{2,k}}}{I_{2,k}} \right) \otimes \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_1}^{b_1} f | I_{1,t}), \Delta_{I_2}^{b_2} g \rangle \right|
\]

\[
\lesssim \sum_{i_2=1}^{\infty} \sum_{i_2=1}^{\infty} \sum_{t=1}^{2^n} \sum_{k=1}^{2^m} \sum_{K \in \text{dist}(I_1, I_2)} \sum_{I_{1}, I_{2} \in K} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \langle b_1' \Delta_{I_2}^{b_2} \left( \frac{b_1^{-1} \chi_{I_{2,k}}}{I_{2,k}} \right) \otimes \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_1}^{b_1} f | I_{1,t}), \Delta_{I_2}^{b_2} g \rangle \|L^2\|g\|L^2\|
\]

We claim that for any \( t, k \),

\[
\text{(6.3)}
\]

\[
\| \sum_{K \in \text{dist}(I_1, I_2)} \sum_{i_2=1}^{\text{dist}(I_1, I_2)} \left( \int_{I_{1,t}} b_1 \, dx_1 \right) \langle b_1' \Delta_{I_2}^{b_2} \left( \frac{b_1^{-1} \chi_{I_{2,k}}}{I_{2,k}} \right) \otimes \pi_{h_{I_{1,t}, I_{2,k}}}^{b_1, b_2} (\Delta_{I_1}^{b_1} f | I_{1,t}), \Delta_{I_2}^{b_2} g \rangle \|L^2\|f\|L^2\|
\]

To see this, first observe that since \( b' \) is pseudo-accretive, for any \( L^2 \) function \( h \),

\[
\|h\|_{L^2} \approx \sup_{\|g\|_{L^2} \leq 1} \langle h, g \rangle = \sup_{\|g\|_{L^2} \leq 1} \int h g b'.
\]

And we have

\[
\langle \Delta_{I_2}^{b_2} h, g \rangle = \langle h, \Delta_{I_2}^{b_2} g \rangle.
\]
Hence by linearity, $LHS$ of (6.3) is comparable to

$$\sup_{\|g\|_{L^2} \leq 1} \sum_{K} \sum_{(i_1, i_2)} \left( \int_{I_1} b_1 \langle \Delta_{I_2}^{b_1} \chi_{I_2, k} \rangle \otimes \pi_{\delta_{I_1, i_2, k}}^{b_2, b_2^*} \langle \Delta_{I_1}^{b_1} f | I_1, t \rangle, g \rangle \right)^2$$

$$\lesssim \sup_{\|g\|_{L^2} \leq 1} \sum_{K} \sum_{(i_1, i_2)} |I_1| |I_2|^{1/2} |K|^{-1} \|\Delta_{I_1}^{b_1} f | I_1, t \|_{L^2(R^n)} \|\Delta_{I_2}^{b_2} g \|_{L^2}$$

Since $\|\Delta_{I_2}^{b_2} \chi_{I_2, k} \|_{L^2(R^n)} \lesssim \langle f | I_2 | |I_2|^{1/2} = |I_2|^{-1/2}$, and by Lemma 6.2 $\|h_{I_1, i_2, k}\|_{BMO} \lesssim 2^{-i_i/2} |K|^{-1/2} |I_2|$, the RHS of the above inequality

$$\lesssim 2^{-i_i/2} \sup_{\|g\|_{L^2} \leq 1} \sum_{K} \sum_{(i_1, i_2)} |I_1| |I_2|^{1/2} |K|^{-1} \|\Delta_{I_1}^{b_1} f | I_1, t \|_{L^2(R^n)} \|\Delta_{I_2}^{b_2} g \|_{L^2}$$

$$= 2^{-i_i/2} \sup_{\|g\|_{L^2} \leq 1} \sum_{K} \sum_{(i_1, i_2)} |I_1| |I_2|^{1/2} |K|^{-1} \left( \frac{1}{|I_1|} \int_{I_1, t} \int_{R^n} |\Delta_{I_1}^{b_1} f|^2 \right)^{1/2} \|\Delta_{I_2}^{b_2} g \|_{L^2}$$

$$\lesssim 2^{-i_i/2} \sup_{\|g\|_{L^2} \leq 1} \sum_{K} \sum_{(i_1, i_2)} |I_1|^{1/2} |I_2|^{1/2} |K|^{-1} \|\Delta_{I_1}^{b_1} f \|_{L^2} \|\Delta_{I_2}^{b_2} g \|_{L^2}$$

$$\lesssim 2^{-i_i/2} \|f\|_{L^2},$$

where the last step follows from the first partial control lemma we stated in the beginning.

Then, to complete this section, we give a proof of Lemma 6.2.

Proof. (of Lemma 6.2) It suffices to show that for any cube $V \subset R^n$, and any function $a$ satisfying $\text{spt} a \subset V$, $|a| \leq 1$, $\langle f, a \rangle = 0$, there holds

$$\langle h_{I_1, i_2, k}, a \rangle \lesssim 2^{-i_i/2} |K|^{-1/2} |I_2| |V|.$$

To see this,

$LHS = \langle (\Delta_{I_1}^{b_1} T^s (\chi_{I_2, k} b_1 \otimes b_2)), a \rangle$

$$= \langle \chi_{I_2, k} b_1 \otimes b_2, T(\Delta_{I_1}^{b_1} (\chi_{I_1, t}^s) \otimes a) \rangle$$

$$= \langle \chi_{I_2, k} b_1 \otimes b_2, T(\Delta_{I_1}^{b_1} (\chi_{I_1, t}^s) \otimes a) \rangle + \langle \chi_{I_2, k} b_1 \otimes \chi_{3V} b_2, T(\Delta_{I_1}^{b_1} (\chi_{I_1, t}^s) \otimes a) \rangle$$

$$:= (1) + (2).$$
For (2), since the two functions in the pairing have good separation on both variables, and \( \int a = \int \Delta_{I_1}^{b_1} (\frac{\chi_{I_1,t}}{|I_1|}) = 0 \), use full kernel representation and change the kernel to

\[
K(x, y) - K(x, y_1, c_V) - K(x, c_I, y_2) + K(x, c_I, c_V).
\]

Then, by Hölder condition,

\[
(2) \lesssim \ell(I_1)^{\delta} \ell(V)^{\delta} \int_{I_1 \times V} \int_{I_2 \times (3V)^c} \frac{1}{|x_1 - c_I|^{n+\delta}} \frac{1}{|x_2 - c_V|^{m+\delta}} |\Delta_{I_1}^{b_1} (\frac{\chi_{I_1,t}}{|I_1|})||a| \, dx \, dy
\]

\[
\lesssim 2^{-i_1 \delta / 2} |K|^{-1} \ell(V)^{\delta} |V||I_2| \int_{(3V)^c} \frac{1}{|x_2 - c_V|^{m+\delta}} \, dx_2
\]

\[
\lesssim 2^{-i_1 \delta / 2} |V| |K|^{-1} |I_2|.
\]

For (1), there is good separation on only one variable, so we need to use the partial kernel representation.

\[
(1) = \int_{I_1} \int_{I_2,h} K_{b_1^{-1}, a, \chi_{3V}} (x_1, y_1) \Delta_{I_1}^{b_1} (\frac{\chi_{I_1,t}}{|I_1|}) (y_1) b_1(x_1) \, dx_1 \, dy_1
\]

\[
= \int_{I_1} \int_{I_2,h} (K_{b_1^{-1}, a, \chi_{3V}} (x_1, y_1) - K_{b_1^{-1}, a, \chi_{3V}} (x_1, c_I)) \Delta_{I_1}^{b_1} (\frac{\chi_{I_1,t}}{|I_1|}) (y_1) b_1(x_1) \, dx_1 \, dy_1
\]

\[
\lesssim C(b_1^{-1}, a, \chi_{3V}) \left( \frac{\ell(I_1)}{\ell(K)} \right)^{\delta / 2} |K|^{-1} |I_2| \int_{I_1} |\Delta_{I_1}^{b_1} (\frac{\chi_{I_1,t}}{|I_1|})| \, dx_1
\]

\[
\lesssim 2^{-i_1 \delta / 2} |V| |K|^{-1} |I_2|.
\]

In the last step of the above, we used the partial C-Z assumption that \( C(b_1^{-1}, a, \chi_{3V}) \lesssim |V| \).

### 7. Separated / Equal: \( \sigma_{out} = \)

In this part,

\[
\sigma_{out} = \sum_{i_2=0}^{\infty} \sum_{i_1=1}^{\infty} \sum_{K \text{ dist}(I_1, I_2) < \ell(I_1)^{\gamma \eta} \ell(I_2)^{1-\gamma \eta}} \sum_{I_1 \cap I_2 = K} \sum_{V} \langle M_0 T M_0 \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle.
\]

By the full control lemma, it suffices to prove the following proposition.

#### 7.1. Proposition.

\[
|\langle M_0 T M_0 \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle| \lesssim 2^{-i_1 \delta / 2} \frac{|I_1|^{1/2} |I_2|^{1/2} \| \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f \|_2 \| \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \|_2}{|K| \| \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f \|_2 \| \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \|_2}.
\]
Proof.

\( |(M_b TM_b \Delta_{I_1}^b \Delta_{V}^b f, \Delta_{I_2}^b \Delta_{V}^b g)| \leq \sum_{V', V'' \in \text{ch}(V)} |(M_b TM_b (\chi_{V'} \Delta_{I_1}^b \Delta_{V}^b f), \chi_{V''} \Delta_{I_2}^b \Delta_{V}^b g)| \)

\[ + \sum_{V' \in \text{ch}(V)} |(M_b TM_b (\chi_{V'} \Delta_{I_1}^b \Delta_{V}^b f), \chi_{V'} \Delta_{I_2}^b \Delta_{V}^b g)| \]

\[ := (1) + (2). \]

For (2), the partial kernel representation gives

\[ (2) = \sum_{V' \in \text{ch}(V)} \left| \int_{I_1} \int_{I_2} (K_{x, V'} \chi_{V'}(x_1, y_1) - K_{x, V'} \chi_{V'}(x_1, c_1)) \right| \Delta_{I_1}^b \Delta_{I_2}^b f |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \Delta_{I_2}^b \Delta_{I_2}^b g |V'| \]

\[ \lesssim \sum_{V' \in \text{ch}(V)} C(\chi_{V'}, \chi_{V'}) 2^{-i\delta/2} |K|^{-1} |V'|^{-2} \left( \int_{I_1 \times V'} |\Delta_{I_1}^b \Delta_{V}^b f| \right) \left( \int_{I_2 \times V'} |\Delta_{I_2}^b \Delta_{V}^b g| \right) \]

\[ \lesssim 2^{-i\delta/2} |K|^{-1} |V'|^{-1} \| \Delta_{I_1}^b \Delta_{V}^b f \|_{L^2} \| \Delta_{I_2}^b \Delta_{V}^b g \|_{L^2} |I_1|^{1/2} |I_2|^{1/2} \sum_{V' \in \text{ch}(V)} |V'| \]

\[ = 2^{-i\delta/2} \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \| \Delta_{I_1}^b \Delta_{V}^b f \|_{L^2} \| \Delta_{I_2}^b \Delta_{V}^b g \|_{L^2}. \]

For (1), the full kernel representation and the mixed Hölder-size condition give

\[ (1) = \sum_{V', V'' \in \text{ch}(V)} | \int_{I_1 \times V'} \int_{I_2 \times V''} (K(x, y) - K(x, c_1, y_2)) \Delta_{I_1}^b \Delta_{V}^b f (y_1) \Delta_{I_2}^b \Delta_{V}^b g (x_1) b(y) b'(x) \, dx \, dy | \]

\[ \lesssim \sum_{V', V'' \in \text{ch}(V)} 2^{-i\delta/2} |K|^{-1} \left( \frac{1}{|V'|} \int_{I_1 \times V'} |\Delta_{I_1}^b \Delta_{V}^b f| \right) \left( \frac{1}{|V''|} \int_{I_2 \times V''} |\Delta_{I_2}^b \Delta_{V}^b g| \right) \left( \int_{V' \times V''} \frac{1}{|x_2 - y_2|^m} \right) \]

\[ \lesssim 2^{-i\delta/2} \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \| \Delta_{I_1}^b \Delta_{V}^b f \|_{L^2} \| \Delta_{I_2}^b \Delta_{V}^b g \|_{L^2} \sum_{V', V'' \in \text{ch}(V)} |V'|^{-1/2} |V''|^{-1/2} |V| \]

\[ \lesssim 2^{-i\delta/2} \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \| \Delta_{I_1}^b \Delta_{V}^b f \|_{L^2} \| \Delta_{I_2}^b \Delta_{V}^b g \|_{L^2}, \]

which completes the proof. \( \square \)
8. SEPARATED/NEARBY: $\sigma_{\text{out/near}}$

In this part, we still want to use the full control lemma to bound the pairing. Notice that since $J_1, J_2$ are near, from a simple lemma proved by Hytönen in [9], the cube $V = J_1 \vee J_2$ satisfies $\ell(V) \leq 2^r \ell(J_1)$, hence the size of $J_1, J_2$ and $V$ are comparable. Since

$$|\sigma_{\text{out/near}}| \leq \sum_{i_2=1}^{\infty} \sum_{i_1=i_2}^{\infty} \sum_{j_1=1}^{r} \sum_{(i_1, j_1)} \sum_{j_2=1}^{(j_1, j_2)} |\langle M_0 T M_0 \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f, \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g \rangle|$$

and $|J_1|^{1/2} |J_2|^{1/2} \approx C$, in order to bound $\sigma_{\text{out/near}}$, it suffices to show

$$|\langle M_0 T M_0 \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f, \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g \rangle| \lesssim 2^{-i_1 \delta/2} \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \|\Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f\|_{L^2} \|\Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g\|_{L^2}.$$ 

To see this, since now both variables are separated but only the first separation is good, by the full kernel representation and the mixed Hölder-size condition,

\[
LHS = \left| \int_{I_1 \times J_1} \int_{I_2 \times J_2} (K(x, y) - K(x, c_1, y_2)) \Delta_{I_1}^{b_1} f(y) b(y) \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g(x) b'(x) \, dx \, dy \right|
\]

$$\lesssim 2^{-i_1 \delta/2} |K|^{-1} \sum_{s, t=1}^{2m} \int_{I_1 \times J_1, s} \int_{I_2 \times J_2, t} |\langle \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \rangle_{J_1, s} \| \langle \Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g \rangle_{J_2, t} \| dx \, dy\]
$$

$$\lesssim 2^{-i_1 \delta/2} |K|^{-1} |I_1|^{1/2} |I_2|^{1/2} |J_1|^{-1/2} |J_2|^{-1/2} \|\Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f\|_{L^2} \|\Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g\|_{L^2} \sum_{s, t=1}^{2m} \int_{J_1, s \times J_2, t} \frac{1}{|x_2 - y_2|^m} \, dx_2 \, dy_2 \]
$$

$$\lesssim 2^{-i_1 \delta/2} \frac{|I_1|^{1/2} |I_2|^{1/2}}{|K|} \|\Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f\|_{L^2} \|\Delta_{I_2}^{b_2} \Delta_{J_2}^{b_2} g\|_{L^2},$$

where the last step follows from the fact that the size of $J_1, J_2$ and $V$ are comparable.

9. INSIDE/INSIDE: $\sigma_{\text{in/in}}$

This part is comparably difficult to deal with, and is also the first place where the assumed BMO conditions stated in the beginning come into play. We will also see that the boundedness of full paraproducts will play an important role in our estimates. To begin with, we first do the following decomposition. Let $I_1 \subset I_{2,1} \subset \text{ch}(I_2), J_1 \subset J_{2,1} \subset \text{ch}(J_2)$, then
\[
\langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle
\]
\[
= \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1}) \rangle
\]
\[
+ \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, (\Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g)_{I_2,1} (x_1) \otimes 1(x_2) \rangle
\]
\[
= \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} \times J_2, (\Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - (\Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g)_{I_2,1} + (\Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g)_{I_2,1}) \rangle
\]
\[
+ \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_2, 1(x_1, x_2) \rangle + \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, (\Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g)_{I_2,1} \times J_2, 1(x_1, x_2) \rangle \rangle
\]
\[
:= I + II + III + IV.
\]

**Part II, III.** These two parts are symmetric, so it suffices to estimate one of them, say part III. This can be similarly dealt with as the second part in section Separated/Inside, where we used partial paraproducts.

\[
\sum_{J_1 \subseteq J_2} III = \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, 1 \rangle_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1
\]
\[
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, 1 \rangle_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1
\]
\[
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1 \rangle_1
\]
\[
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1 \rangle_1
\]
\[
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1 \rangle_1
\]
\[
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1 \rangle_1
\]
\[
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1 \rangle_1
\]
\[
= \sum_{J_1 \subseteq J_2} \langle \langle M_b TM_b \Delta_{I_1}^{b_1} \Delta_{I_2}^{b_2} f, \chi_{I_2,1} (\langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} - \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2,1} \times J_1) \rangle_1 \rangle_1
\]
where \( s_{I_1,t,I_2,k}(x_2) = (\Delta_{I_1}^{b_1} T^*(\chi_{I_2,k} b_1' \otimes b_2' ))|_{I_1,t}, \ s_{I_1,t,I_2,k}'(x_2) = (\Delta_{I_1}^{b_1} T^*(\chi_{I_2,k} b_1' \otimes b_2' ))|_{I_1,t}. \)

Note that although formally, \( s_{I_1,t,I_2,k} \) is exactly the \( h_{I_1,t,I_2,k} \) we’ve encountered in section Separated/Inside, but here since the relative position of \( I_1, I_2 \) has changed, they are actually different functions. And we will prove later that although \( s_{I_1,t,I_2,k} \) is still in \( BMO(\mathbb{R}^m) \), the estimate of its norm is different from \( h_{I_1,t,I_2,k} \). More specifically,

9.1. Lemma.

\[
\|s_{I_1,t,I_2,k}\|_{BMO(\mathbb{R}^m)} \lesssim 2^{-i_1 \delta/2}, \quad \|s_{I_1,t,I_2,k}'\|_{BMO(\mathbb{R}^m)} \lesssim 2^{-i_1 \delta/2}
\]

Let’s assume this to be true right now. Then

\[
\sum_{J_1 \subseteq J_2} III = \sum_{k=2}^{2^n} \sum_{t=1}^{2^n} \left( \int_{I_1,t} b_1(\Delta_{I_1}^{b_1} f|_{I_1,t}, \pi_{s_{I_1,t,I_2,k}}^{(2)} (\Delta_{I_2}^{b_2} g|_{I_2,k}))_2 \right.
\]

\[
- \sum_{t=1}^{2^n} \left( \int_{I_1,t} b_1(\Delta_{I_1}^{b_1} f|_{I_1,t}, \pi_{s_{I_1,t,I_2,k}}^{(2)} (\Delta_{I_2}^{b_2} g|_{I_2,t}))_2 \right.
\]

\[
= \sum_{k=2}^{2^n} \sum_{t=1}^{2^n} \left( \int_{I_1,t} b_1(\Delta_{I_2}^{b_2} (\chi_{I_2,k} |_{I_2,k}) \otimes \pi_{s_{I_1,t,I_2,k}}^{(2)} (\Delta_{I_1}^{b_1} f|_{I_1,t}), g) \right.
\]

\[
- \sum_{t=1}^{2^n} \left( \int_{I_1,t} b_1(\Delta_{I_2}^{b_2} (\chi_{I_2,k} |_{I_2,k}) \otimes \pi_{s_{I_1,t,I_2,k}}^{(2)} (\Delta_{I_1}^{b_1} f|_{I_1,t}), g) \right.
\]

\[
:= (1) - (2).
\]

Note that part (1) is exactly the same as the pairing appeared in \( \sigma''_{out/in'} \) except that here the partial paraproduct is defined using a different \( BMO \) function. Hence, following exactly the same argument, for any \( t, k \), we have

\[
\|
\sum_{I_2} \sum_{I_1 \subseteq I_2} \left( \int_{I_1,t} b_1(\Delta_{I_2}^{b_2} (\chi_{I_2,k} |_{I_2,k}) \otimes \pi_{s_{I_1,t,I_2,k}}^{(2)} (\Delta_{I_1}^{b_1} f|_{I_1,t}))_2 \right.
\]

\[
\lesssim 2^{-i_1 \delta/2} \sup_{\|g\|_{L^2} \leq 1} \sum_{I_2} \sum_{I_1 \subseteq I_2} |I_1|^{1/2} \left( \int_{I_2} |\Delta_{I_1}^{b_1} f|_{L^2} \right) \left( \Delta_{I_2}^{b_2} g|_{L^2} \right)
\]

\[
\lesssim 2^{-i_1 \delta/2} \| f \|_{L^2},
\]

where again, in the last step, we used the first partial control lemma.
Similarly, although in part (2), the form of the pairing is a little bit different, however, when dealing with $\Delta_{I_2}^{b_1^*}(\frac{\chi_{I_2}}{|I_2|})$, we only need to bound it by

$$C\|\Delta_{I_2}^{b_1^*}(\frac{b_1^* - 1}{|I_2|})\|_{L^2(\mathbb{R}^n)} \lesssim |I_2|^{-1/2};$$

and since the norm of the BMO function has the same bound, so all the rest of the argument for part (1) still works here. i.e. This part satisfies the same estimate as part (1) does.

In conclusion,

$$|\sum_{I_1 \subseteq I_2} \sum_{J_1 \subseteq J_2} III| = |\sum_{i_1=1}^\infty \sum_{I_2} \sum_{J_1 \subseteq J_2} (i_1) III| = |\sum_{i_1=1}^\infty \sum_{I_2} (i_1) (1) - (2)|$$

$$\lesssim \sum_{i_1=1}^\infty 2^{-i_1\delta/2}\|f\|_{L^2}\|g\|_{L^2} \lesssim \|f\|_{L^2}\|g\|_{L^2}.$$

And we are only left to prove Lemma 9.1

Proof. (of Lemma 9.1) We only prove the inequality for $s_{I_1,t,I_2,k}$, since the other one follows from exactly the same argument. Let cube $V \subset \mathbb{R}^m$, $a$ is any function supported on $V$ such that $|a| \leq 1, \int a = 0$. It suffices to show $\langle s_{I_1,t,I_2,k}, a\rangle_2 \lesssim 2^{-i_1\delta/2}|V|$.

In the case $\ell(I_1) < 2^{-r}\ell(I_2)$, we have $\text{dist}(I_1, I_{2,1}) \geq \ell(I_1)^\gamma \ell(I_2)^{1-\gamma}$, i.e. the separation of $I_1$ and $I_2$ is good enough. Then following from the same reasoning in the proof of Lemma 6.2, and note that now $I_2 = K$, we have $\langle s_{I_1,t,I_2,k}, a\rangle_2 \lesssim 2^{-i_1\delta/2}|K||I_2||V| = 2^{-i_1\delta/2}|V|$.

Now let’s assume $2^{-r}\ell(I_2) \leq \ell(I_1) < \ell(I_2)$. Then the size of $I_1, I_2$ are comparable, i.e. $2^{-i_1} \approx C$, so it suffices to show $\langle s_{I_1,t,I_2,k}, a\rangle_2 \lesssim |V|$. Split

$$\langle s_{I_1,t,I_2,k}, a\rangle_2$$

$$= \langle \chi_{I_2}b_1^* \otimes b_2^*, T(\Delta_{I_1}^{b_1^*}(\frac{\chi_{I_1,t}}{|I_1,t|}) \otimes a) \rangle$$

$$= \langle \chi_{3I_1 \cap I_2}b_1^* \otimes \chi_{3V}b_2^*, T(\Delta_{I_1}^{b_1^*}(\frac{\chi_{I_1,t}}{|I_1,t|}) \otimes a) \rangle + \langle \chi_{3(I_1)^c \cap I_2}b_1^* \otimes \chi_{3V}b_2^*, T(\Delta_{I_1}^{b_1^*}(\frac{\chi_{I_1,t}}{|I_1,t|}) \otimes a) \rangle$$

$$+ \langle \chi_{3I_1 \cap I_2}b_1^* \otimes \chi_{3V^c}b_2^*, T(\Delta_{I_1}^{b_1^*}(\frac{\chi_{I_1,t}}{|I_1,t|}) \otimes a) \rangle + \langle \chi_{3(I_1)^c \cap I_2}b_1^* \otimes \chi_{3V^c}b_2^*, T(\Delta_{I_1}^{b_1^*}(\frac{\chi_{I_1,t}}{|I_1,t|}) \otimes a) \rangle$$

$$:=(1) + (2) + (3) + (4).$$
By the partial kernel representation and size condition for the partial kernel,

\[
(1) = \int_{I_1} \int_{3I_1 \cap I_{2,k}} K_{b^{-1}_2, a, \chi_{3V}}(x_1, y_1) \Delta_{I_1}^{b_1}(\frac{\chi_{I_1,t}}{|I_{1,t}|})b'_1(x_1) \, dx_1 dy_1
\]
\[
\lesssim C(b^{-1}_2 a, \chi_{3V})|I_1|^{-1} \int_{I_1} \int_{3I_1 \cap I_{2,k}} \frac{1}{|x_1 - y_1|^n} \, dx_1 dy_1 \lesssim |V|.
\]

By the partial kernel representation and Hölder condition for the partial kernel,

\[
(2) = \int_{I_1} \int_{(3I_1)^c \cap I_{2,k}} (K_{b^{-1}_2, a, \chi_{3V}}(x_1, y_1) - K_{b^{-1}_2, a, \chi_{3V}}(x_1, c_1)) \Delta_{I_1}^{b_1}(\frac{\chi_{I_1,t}}{|I_{1,t}|})b'_1(x_1) \, dx_1 dy_1
\]
\[
\lesssim C(b^{-1}_2 a, \chi_{3V})|I_1|^{-1} \int_{I_1} \int_{(3I_1)^c \cap I_{2,k}} \frac{\ell(I_1)^{\delta}}{|x_1 - c_1|^{n+\delta}} \lesssim |V|.
\]

By the full kernel representation and mixed Hölder-size condition,

\[
(3) = \int_{I_1 \times V} \int_{I_1 \cap I_{2,k} \times (3V)^c} (K(x, y) - K(x, y_1, c_1)) \Delta_{I_1}^{b_1}(\frac{\chi_{I_1,t}}{|I_{1,t}|})b'(x) \, dx dy
\]
\[
\lesssim |I_1|^{-1} \ell(V)^{\delta} \int_{I_1 \times V} \int_{I_1 \cap I_{2,k} \times (3V)^c} \frac{1}{|x_1 - y_1|^n} \frac{1}{|x_2 - c_1|^{m+\delta}} \, dx dy
\]
\[
\lesssim |I_1|^{-1} \ell(V)^{\delta} |V| |I_1| \ell(V)^{-\delta} = |V|.
\]

By the full kernel representation and Hölder condition,

\[
(4) = \int_{I_1 \times V} \int_{(3I_1)^c \cap I_{2,k} \times (3V)^c} (K(x, y) - K(x, c_1, y_2) - K(x, y_1, c_1) + K(x, c_1, c_1)) b'(x) \Delta_{I_1}^{b_1}(\frac{\chi_{I_1,t}}{|I_{1,t}|}) \, dx dy
\]
\[
\lesssim \ell(I_1)^{\delta} \ell(V)^{\delta} |I_1|^{-1} \int_{I_1 \times V} \int_{(3I_1)^c \cap I_{2,k} \times (3V)^c} \frac{1}{|x_1 - c_1|^{n+\delta}} \frac{1}{|x_2 - c_1|^{m+\delta}} \, dx dy
\]
\[
\lesssim |I_2|^{-1} \ell(V)^{\delta} |I_1|^{-1} |V| |I_2| \ell(V)^{-\delta} = |V|.
\]

Hence, the proof is complete. \(\square\)

Part I. In part I, since the functions in the pairing are separated on both variables, by an argument similar to what we did in the section Separated/Inside,

\[
|I| \lesssim \left( \frac{\ell(I_1)}{\ell(I_2)} \right)^{\delta/2} \left( \frac{\ell(J_1)}{\ell(J_2)} \right)^{\delta/2} \left( |I_1|^{1/2} |J_1|^{1/2} \right) \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f \|L^2\| \Delta_{I_2}^{b_1'} \Delta_{J_2}^{b_2'} g \|L^2\|,
\]
which combined with the full control lemma, will give the boundedness of part I. (Note that in order to prove the above inequality, we need to discuss four different cases depending on whether $\ell(I_1) < 2^{-r}\ell(I_2)$ and whether $\ell(J_1) < 2^{-r}\ell(J_2)$, and use size, Hölder, or mixed Hölder-size conditions accordingly in each case.)

**Part IV.** To deal with this part, we need to use the $b$-adapted full paraproducts and its $L^2 \to L^2$ boundedness. Write

$$\sum_{I_1 \subseteq I_2} \sum_{J_1 \subseteq J_2} \langle M_b T M_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f, \langle \Delta_{I_2}^{b_1} \Delta g \rangle_{I_1 \times J_1} \rangle$$

$$= \sum_{I_1 \subseteq I_2} \sum_{J_1 \subseteq J_2} \langle M_b T M_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f, \langle \sum_{I_2 \supseteq I_1} \Delta_{I_2}^{b_1} \Delta_{J_2}^{b_2} g \rangle_{I_1 \times J_1} \rangle$$

$$= \sum_{I_1 \subseteq I_2} \sum_{J_1 \subseteq J_2} \langle M_b T M_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f, \langle \Delta_{I_2}^{b_1} g \rangle_{I_1 \times J_1} \rangle$$

$$= \sum_{I_1} \sum_{J_1} \langle M_b T M_b \Delta_{I_1}^{b_1} \Delta_{J_1}^{b_2} f, \langle g \rangle_{I_1 \times J_1} \rangle$$

$$= \langle f, \pi_{T^* b'}(g) \rangle.$$

By assumption, $T^* b' \in BMO(\mathbb{R}^n \times \mathbb{R}^m)$, then use the $L^2$ boundedness of the full paraproduct, we have

$$\sum_{I_1 \subseteq I_2} \sum_{J_1 \subseteq J_2} IV \lesssim \|T^* b'\|_{BMO} \|f\|_{L^2} \|g\|_{L^2}.$$

**10. INSIDE/EQUAL AND INSIDE/NEARBY: $\sigma_{in/\sim}, \sigma_{in/\near}$**

The ways to estimate these two parts are similar, so we only explain the first one as an example. Let $I_1 \subseteq I_{2,1} \in \text{ch}(I_2)$, split

$$\sigma_{in/\sim} = \sum_{I_1 \subseteq I_2} \sum_{V} \langle M_b T M_b \Delta_{I_1}^{b_1} \Delta_{V}^{b_2} f, \Delta_{I_2}^{b_1} \Delta_{V}^{b_2} g \rangle$$

$$= \sum_{I_1 \subseteq I_2} \sum_{V} \langle M_b T M_b \Delta_{I_1}^{b_1} \Delta_{V}^{b_2} f, \chi_{I_{2,1}}(\Delta_{I_2}^{b_1} \Delta_{V}^{b_2} g - \langle \Delta_{I_2}^{b_1} \Delta_{V}^{b_2} g \rangle_{I_{2,1}}) \rangle$$

$$+ \sum_{I_1 \subseteq I_2} \sum_{V} \langle M_b T M_b \Delta_{I_1}^{b_1} \Delta_{V}^{b_2} f, 1(x_1) \otimes \langle \Delta_{I_2}^{b_1} \Delta_{V}^{b_2} g \rangle_{I_1}(x_2) \rangle$$

$$:= \sigma'_{in/\sim} + \sigma''_{in/\sim}.$$
To bound $\sigma_{in/=}^{x}$. In the case $\ell(I_1) < 2^{-r}\ell(I_2)$, it can be dealt with similarly as in the case Separated/Equal. In the case $2^{-r}\ell(I_2) \leq \ell(I_1) < \ell(I_2)$, we claim that

$$|\sigma_{in/=}^{x}| \lesssim \sum_{i_1=1}^{\infty} \sum_{I_2} \sum_{I_1 \subset I_2} \sum_{V} 2^{-i_1\delta/2} \frac{|I_1|^{1/2}}{|I_2|^{1/2}} \|\Delta_{I_1} b_1 \Delta_{V}^b f\|_{L^2} \|\Delta_{I_2} b_1 \Delta_{V}^b g\|_{L^2},$$

then the full control lemma implies the correct bound.

In order to prove the claim, further split

$$\langle M' T M_b \Delta_{I_1}^b \Delta_{V}^b f, \chi_{I_2,1} (\Delta_{I_2}^b \Delta_{V}^b g - \langle \Delta_{I_2}^b \Delta_{V}^b g \rangle_{I_2,1}) \rangle$$

$$= \sum_{V', V'' \in \text{ech}(V), V' \neq V''} \langle M' T M_b (\chi_{V'} \Delta_{I_1}^b \Delta_{V}^b f), \chi_{3 I_1 \cap I_2,1} \otimes \chi_{V''} (\Delta_{I_2}^b \Delta_{V}^b g - \langle \Delta_{I_2}^b \Delta_{V}^b g \rangle_{I_2,1}) \rangle$$

$$+ \sum_{V', V'' \in \text{ech}(V), V' \neq V''} \langle M' T M_b (\chi_{V'} \Delta_{I_1}^b \Delta_{V}^b f), \chi_{3 I_1 \cap I_2,1} \otimes \chi_{V''} (\Delta_{I_2}^b \Delta_{V}^b g - \langle \Delta_{I_2}^b \Delta_{V}^b g \rangle_{I_2,1}) \rangle$$

$$+ \sum_{V' \in \text{ech}(V)} \langle M' T M_b (\chi_{V'} \Delta_{I_1}^b \Delta_{V}^b f), \chi_{3 I_1 \cap I_2,1} \otimes \chi_{V'} (\Delta_{I_2}^b \Delta_{V}^b g - \langle \Delta_{I_2}^b \Delta_{V}^b g \rangle_{I_2,1}) \rangle$$

$$+ \sum_{V' \in \text{ech}(V)} \langle M' T M_b (\chi_{V'} \Delta_{I_1}^b \Delta_{V}^b f), \chi_{3 I_1 \cap I_2,1} \otimes \chi_{V'} (\Delta_{I_2}^b \Delta_{V}^b g - \langle \Delta_{I_2}^b \Delta_{V}^b g \rangle_{I_2,1}) \rangle$$

$$:= (1) + (2) + (3) + (4).$$

In part (1) and (2), both variables are separated, so we use the full kernel representation. And by the size condition and the mixed Hölder-size condition, respectively, they are bounded. In part (3) and (4), only the first variable is separated, so we need the partial kernel representation. By the size condition and Hölder condition for the partial kernel, respectively, they are bounded as well. We omit the details.
Now we deal with $\sigma''_{\text{in}}$, which needs the partial paraproduct argument, but is much easier than the cases we’ve seen before. As before, rewrite
\[
\sum_{I_1 \subseteq I_2} \langle M_M T_M b_1 \Delta_{I_2}^{b_1} b_2 f, \langle \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle_{I_2} \rangle
\]
\[
= \sum_{K} \langle M_M T_M b_1 \Delta_{K}^{b_1} b_2 f, 1(x_1) \otimes \langle \Delta_{K}^{b_2} g \rangle_{K} \rangle(x_2)
\]
\[
= \sum_{t=1}^{2^m} \sum_{k=1}^{2^m} \langle \chi_{V_t} \Delta_{V_t}^{b_2} f, \pi^{b_1}_{r_{V_t} V_k} (\Delta_{V_k}^{b_2} g|V_k) \otimes b_2 \rangle
\]
\[
= \sum_{t=1}^{2^m} \sum_{k=1}^{2^m} \left( \int_{V_t} b_2 \langle \pi^{b_1}_{r_{V_t} V_k} (\Delta_{V_t}^{b_2} f|V_t) \otimes \Delta_{V_t}^{b_2} (\frac{\chi_{V_k}}{|V_k|}) \rangle, g \right),
\]
where $r_{V_t} V_k (x_1) = (\Delta_{V_t}^{b_2} \chi_{V_k} (b_2))|V_t$ is a $BMO$ function whose norm satisfies the following lemma.

10.1. Lemma.
\[
\|r_{V_t} V_k\|_{BMO(\mathbb{R}^n)} \lesssim C.
\]

We postpone the proof, and assume this bound for the moment. Then
\[
|\sigma''_{\text{in}}| \lesssim \sum_{t=1}^{2^m} \sum_{k=1}^{2^m} \left\| \int_{V_t} b_2 \langle \pi^{b_1}_{r_{V_t} V_k} (\Delta_{V_t}^{b_2} f|V_t) \otimes \Delta_{V_t}^{b_2} (\frac{\chi_{V_k}}{|V_k|}) \rangle \| g \|_{L^2}.
\]

By a similar argument as in the previous two partial paraproducts, involving the estimate of the $BMO$ norm of $r_{V_t} V_k$ and the $L^2$ boundedness of the partial paraproduct, it is not hard to show that for any $t, k$,
\[
\| \int_{V_t} b_2 \langle \pi^{b_1}_{r_{V_t} V_k} (\Delta_{V_t}^{b_2} f|V_t) \otimes \Delta_{V_t}^{b_2} (\frac{\chi_{V_k}}{|V_k|}) \rangle \|_{L^2} \lesssim \| f \|_{L^2},
\]
which completes the estimate of part $\sigma''_{\text{in}}$.

Proof. (of Lemma 10.1) For any cube $K \subset \mathbb{R}^n$ and any function $a$ supported on $K$ such that $|a| \leq 1$, $\int a = 0$, we claim that $\langle r_{V_t} V_k, a \rangle_1 \lesssim |K|$.

To see this, write
\[
\langle r_{V_t} V_k, a \rangle_1 = \langle \chi_K b_1^{b_1} \otimes V_k b_2, T(a \otimes \Delta_{V_t}^{b_2} (\frac{\chi_{V_t}}{|V_t|})) \rangle
\]
\[
= \langle \chi_{(3K)} b_1^{b_1} \otimes V_k b_2, T(a \otimes \Delta_{V_t}^{b_2} (\frac{\chi_{V_t}}{|V_t|})) \rangle
\]
\[
= \langle \chi_{3K \setminus K} b_1^{b_1} \otimes V_k b_2, T(a \otimes \Delta_{V_t}^{b_2} (\frac{\chi_{V_t}}{|V_t|})) \rangle \]
\[
:= (1) + (2) + (3).
\]
For part (1), write
\[ (1) = \sum_{s=1}^{2m} \langle \chi_K b'_1 \otimes \chi_{V_k} b'_2, T(a \otimes \chi_{V_s} \Delta_V^{b_{2s}} \left( \frac{\chi_{V_s}}{|V_t|} \right)) \rangle. \]

If \( s \neq k \), use partial kernel representation and size condition for the partial kernel,
\[ \langle \chi_K b'_1 \otimes \chi_{V_k} b'_2, T(a \otimes \chi_{V_s} \Delta_V^{b_{2s}} \left( \frac{\chi_{V_s}}{|V_t|} \right)) \rangle \]
\[ = \int_{V_k} \int_{V_k} K_{b_1^{-1} a, \chi_K} (x_2, y_2) b'_2(x_2) \Delta_V^{b_{2s}} \left( \frac{\chi_{V_s}}{|V_t|} \right)(y_2) \, dx_2 \, dy_2 \]
\[ \lesssim C(b_1^{-1} a, \chi_K) |V|^{-1} \int_{V_k} \int_{V_k} \frac{1}{|x_2 - y_2|^m} \lesssim |K|. \]

If \( s = k \), by the first diagonal BMO condition,
\[ \langle \chi_K b'_1 \otimes \chi_{V_k} b'_2, T(a \otimes \chi_{V_s} \Delta_V^{b_{2s}} \left( \frac{\chi_{V_s}}{|V_t|} \right)) \rangle \]
\[ = (\Delta_V^{b_{2s}} \left( \frac{\chi_{V_k}}{|V_t|} \right)) \langle \chi_K b'_1 \otimes \chi_{V_k} b'_2, T(a \otimes \chi_{V_s} b_2) \rangle \]
\[ \lesssim |V|^{-1} |K| |V| = |K|. \]

For part (2) and (3), write
\[ (2) = \sum_{s=1}^{2m} \langle \chi_{(3K)^c} b'_1 \otimes \chi_{V_k} b'_2, T(a \otimes \chi_{V_s} \Delta_V^{b_{2s}} \left( \frac{\chi_{V_s}}{|V_t|} \right)) \rangle, \]
and similarly for (3).

If \( s \neq k \), since both variables are separated, we can use full kernel representation, and mixed Hölder-size condition for (2), size condition for (3). If \( s = k \), we use partial kernel representation, and Hölder condition for (2), size condition for (3). The details can be carried out similarly as for (1), and we omit them. \( \square \)

11. EQUAL/EQUAL, EQUAL/NEARBY AND NEARBY/NEARBY: \( \sigma =/\) =

We discuss these three cases together. When \( J_1, J_2 \) are near each other, the sizes of \( J_1, J_2, J_1 \cap J_2 \) are comparable, similarly for the other variable. So by the full control lemma, in either of these three cases, it suffices to show
\[ \langle M_{b} T M_{b} \Delta_{I_1}^{b_1} \Delta_{I_1}^{b_2} f, \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \rangle \lesssim \| \Delta_{I_1}^{b_1} \Delta_{I_1}^{b_2} f \|_{L^2} \| \Delta_{I_2}^{b_1} \Delta_{I_2}^{b_2} g \|_{L^2}. \]

We only prove the above for the case Equal/Equal, which is the most difficult one since there is no separation on either variable. Note that for Equal/Nearby, one can use partial kernel representation and size condition to prove it, and for Nearby/Nearby, the full kernel representation and size condition will do.
Write $I_1 = I_2 = K, J_1 = J_2 = V,$ and decompose the pairing into restrictions on each pair of their children,

$$\langle M_b TM_b \Delta_K^{b_1} \Delta_V^{b_2} f, \Delta_K^{b_1} \Delta_V^{b_2} g \rangle \leq \sum_{i,s=1}^{2^n} \sum_{j,t=1}^{2^m} \langle M_b TM_b (\chi_{K_i \times V_j} \Delta_K^{b_1} \Delta_V^{b_2} f), \chi_{K_i \times V_j} \Delta_K^{b_1} \Delta_V^{b_2} g \rangle.$$  

If $i \neq s, j \neq t,$ by the full kernel representation and size condition,

$$\langle M_b TM_b (\chi_{K_i \times V_j} \Delta_K^{b_1} \Delta_V^{b_2} f), \chi_{K_i \times V_j} \Delta_K^{b_1} \Delta_V^{b_2} g \rangle \leq \int_{K_i \times V_j} \int_{K_i \times V_j} \frac{1}{|x_1 - y_1|^n} \frac{1}{|x_2 - y_2|^m}.$$  

The case $i = s, j \neq t$ is symmetric to this one.

If $i = s, j = t,$ by the partial kernel representation and size condition for the partial kernel,

The case $i = s, j \neq t$ is symmetric to this one.

If $i = s, j = t,$ by the weak boundedness property,

$$\langle M_b TM_b \Delta_K^{b_1} \Delta_V^{b_2} f, \Delta_K^{b_1} \Delta_V^{b_2} g \rangle = \langle \Delta_K^{b_1} \Delta_V^{b_2} f, \Delta_K^{b_1} \Delta_V^{b_2} g \rangle.$$  

This completes this section, as well as all the cases when $\ell(I_1) \leq \ell(I_2), \ell(J_1) \leq \ell(J_2).$ Moreover, the cases when $\ell(I_1) > \ell(I_2), \ell(J_1) > \ell(J_2)$ can be dealt with symmetrically.

12. Mixed Cases

We now consider the mixed cases. It suffices to analyze the case when $\ell(I_1) \leq \ell(I_2), \ell(J_1) > \ell(J_2),$ and the only sub-case which is not symmetric to any of the above is the mixed Inside/Inside, which involves the boundedness of mixed...
paraproducts. By assumption, $I_1 \subset I_2, J_2 \subset J_1$. Suppose $I_1 \subset I_{2,1} \in \text{ch}(I_2)$ and $J_2 \subset J_{1,2} \subset \text{ch}(J_1)$. Split
\[
\langle M_{b^1} M_b \Delta_{j_1}^{b^1} \Delta_{j_2}^{b^2} f, \Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g \rangle
\]
\[
= \langle M_{b^1} M_b \Delta_{j_1}^{b^1} \Delta_{j_1}^{b^2}, f, 1(x) \otimes \langle \Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g \rangle_{l_{2,1},2}(x) \rangle
\]
\[
= \langle M_{b^1} M_b \Delta_{j_1}^{b^1} \Delta_{j_1}^{b^2} f, \chi_{l_{2,1}^1}, (\Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g - \langle \Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g \rangle_{l_{2,1},1}) \rangle
\]
\[
+ \langle M_{b^1} M_b \Delta_{j_1}^{b^1} \Delta_{j_1}^{b^2} f, \chi_{l_{2,1}^2}, (\Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g - \langle \Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g \rangle_{l_{2,1},1}) \rangle
\]
\[
+ \langle M_{b^1} M_b \Delta_{j_1}^{b^1} \Delta_{j_1}^{b^2} f, \chi_{l_{1,2}^1}, \delta_{l_{1,2}^1}, (\Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g - \langle \Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} g \rangle_{l_{2,1},1}) \rangle
\]
\[
+ \langle M_{b^1} M_b \Delta_{j_1}^{b^1} \Delta_{j_1}^{b^2} f, \chi_{l_{1,2}^2}, \delta_{l_{1,2}^2}, \Delta_{j_2}^{b^1} \Delta_{j_2}^{b^2} \rangle_{l_{2,1},1}(x)
\]
\[
:= I + II + III + IV.
\]

Part $I, II, III$ can be similarly estimated as the corresponding parts in the Inside/Inside case discussed above. Note that for part $II, III$, we need to use the partial adjoint operator $T_1$ to rewrite it into a form having partial paraproduct in it, and estimate some new one-parameter $BMO$ functions, which can be achieved by the same techniques we’ve seen before.

To estimate part $IV$, we need to apply the boundedness of mixed paraproducts.
\[
\sum_{I_1 \subseteq I_2} \sum_{J_2 \subseteq J_1} IV = \sum_{K,V} \langle M_{b^1} M_b \langle \Delta_{j_1}^{b^1} f \rangle_{V}, \langle \Delta_{j_2}^{b^2} g \rangle_{K} \rangle
\]
\[
= \sum_{K,V} \langle T(b_1 \langle \Delta_{j_1}^{b^1} f \rangle_{V} \otimes b_2), b_1 \otimes b_2 \langle \Delta_{j_2}^{b^2} g \rangle_{K} \rangle
\]
\[
= \sum_{K,V} \langle T_1(b_1 \otimes b_2), b_1 \langle \Delta_{j_1}^{b^1} f \rangle_{V} \otimes b_2 \langle \Delta_{j_2}^{b^2} g \rangle_{K} \rangle.
\]

Recall that by assumption, $d = b_1 \otimes b_2$, $d' = b_1' \otimes b_2$, so the above is
\[
\langle T_1(d'), \sum_{K,V} M_d \langle \Delta_{j_1}^{b_1} f \rangle_{V} \otimes \langle \Delta_{j_2}^{b_2} g \rangle_{K} \rangle
\]
\[
= \sum_{K,V} \langle T_1(d'), M_d \langle E_{V}^{d_2} \Delta_{K}^{d_1} f \rangle \langle E_{V}^{d_1} \Delta_{K}^{d_2} g \rangle \rangle
\]
\[
= \langle \sum_{K,V} E_{K}^{d_1} \langle \langle E_{V}^{d_2} f \rangle M_d \Delta_{K}^{d_1} \Delta_{K}^{d_2} T_1(d'), g \rangle \rangle
\]
\[
= \langle \hat{\pi}_{T_1}(d')(f), g \rangle
\]
\[
\lesssim \|T_1(d')\|_{BMO} \|f\|_{L^2} \|g\|_{L^2},
\]
and \( \|T_1(d')\|_{BMO} < \infty \) is one of our \( BMO \) assumptions. This completes the estimate of the mixed cases.

REFERENCES

[1] A. Bernard, *Espaces \( H^1 \) de martingales a deux indices. Dualité avec les martingales de type \( BMO \)*, Bull. Sc. Math. 103 (1979), 297–303.

[2] 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.

[3] R. R. Coifman, P. W. Jones and S. Semmes, *Two elementary proofs of the \( L^2 \) boundedness of Cauchy integrals on Lipschitz curves*, J. Amer. Math. Soc. 2 (1989), no. 3, 553–564.

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

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

[6] C. Fefferman and E. Stein, *\( H^p \) spaces of several variables*, Acta Math. 2 (1972), 137–193.

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

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

[9] T. Hytönen, *Representation of singular integrals by dyadic operators, and the \( A_2 \) theorem*, preprint (2011).

[10] 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).

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

[12] T. Hytönen and H. Martikainen, *Non-homogeneous \( T1 \) theorem for bi-parameter singular integrals*, preprint (2012).

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

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

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

[16] H. Martikainen, *Representation of bi-parameter singular integrals by dyadic operators*, preprint (2011).

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

[18] F. Nazarov, S. Treil, and A. Volberg, *The \( T \) \( b \)-theorem on non-homogeneous spaces*, Acta Math. 190 (2003), no. 2, 151–239.

[19] J. Pipher and L. A. Ward, *\( BMO \) from dyadic \( BMO \) on the bidisc*, J. Lond. Math. Soc. (2) 77 (2008), no. 2, 524–544.

[20] S. Pott and P. Villafruca, *A \( T(1) \) theorem on product spaces*, preprint (2013).

[21] S. Treil, *\( H^1 \) and dyadic \( H^1 \)*, preprint (2010).

DEPARTMENT OF MATHEMATICS, BROWN UNIVERSITY, PROVIDENCE, RI, USA

E-mail address: yumeng_ou@brown.edu