Product Space Singular Integrals with Mild Kernel Regularity

Emil Airta[1] · Henri Martikainen[1] · Emil Vuorinen[1]

Abstract

We develop product space theory of singular integrals with mild kernel regularity. We study these kernel regularity questions specifically in situations that are very tied to the $T1$ type arguments and the corresponding structural theory. In addition, our results are multilinear.

Keywords Singular integrals · Commutators · Weighted estimates · Kernel regularity · Multilinear analysis · Multi-parameter analysis

Mathematics Subject Classification 42B20

1 Introduction

The usual definition of a singular integral operator (SIO)

$$Tf(x) = \int_{\mathbb{R}^d} K(x, y) f(y) \, \mathrm{d}y$$

✉ Emil Vuorinen
  emil.vuorinen@helsinki.fi

  Emil Airta
  emil.airta@helsinki.fi

  Henri Martikainen
  henri.martikainen@helsinki.fi

[1] Department of Mathematics and Statistics, University of Helsinki, P.O.B. 68, 00014 Helsinki, Finland

involves a Hölder-continuous kernel K with a power-type continuity-modulus $t \mapsto t^\gamma$. However, many results continue to hold with significantly more general assumptions. Such kernel regularity considerations become non-trivial especially in connection with results that go beyond the classical Calderón–Zygmund theory—an example is the A_2 theorem of Hytönen [21] with Dini-continuous kernels by Lacey [24]. Estimates for SIOs with mild kernel regularity are, for instance, linked to the theory of rough singular integrals, see e.g. [22].

The fundamental question concerning the L^2 (or L^p) boundedness of an SIO T is usually best answered by so-called $T1$ theorems, where the action of the operator T on the constant function 1 is key. We study kernel regularity questions specifically in situations that are very tied to the $T1$ type arguments and the corresponding structural theory—a big part of the modern product space theory of SIOs relies on such analysis. The proofs of $T1$ theorems display a fundamental structural decomposition of SIOs into their cancellative parts and so-called paraproducts. It is this structure that is extremely important for obtaining further estimates beyond the initial scalar-valued L^p boundedness. Refined versions of $T1$ theorems provide exact identities in terms of model operators and are called representation theorems, see [20,21,32].

A concrete definition of kernel regularity is as follows. It concerns the required regularity of the continuity-moduli ω appearing in the various kernel estimates, such as,

$$|K(x, y) - K(x', y)| \leq \omega\left(\frac{|x - x'|}{|x - y|}\right)\frac{1}{|x - y|^d}, \quad |x - x'| \leq |x - y|/2.$$

Recently, Grau de la Herrán and Hytönen [17] proved that the modified Dini condition

$$\|\omega\|_{\mathrm{Dini}_\alpha} := \int_0^1 \omega(t)\left(1 + \log\frac{1}{t}\right)^\alpha \frac{dt}{t}$$

with $\alpha = \frac{1}{2}$ is sufficient to prove a $T1$ theorem even with an underlying measure μ that can be non-doubling. This matches the best known sufficient condition for the classical homogeneous $T1$ theorem [10]—such results are implicit in Figiel [16] and explicit in Deng et al. [11]. The exponent $\alpha = \frac{1}{2}$ has a fundamental, even sharp, feeling in all of the existing arguments.

In [17] a new type of representation theorem appears, where the key difference to the original representation theorems [20,21] is that the decomposition of the cancellative part is in terms of different operators that package multiple dyadic shifts into one and offer more efficient bounds when it comes to kernel regularity. Some of the ideas of the decomposition in [17] are rooted in the work of Figiel [15,16]. We simultaneously extend [17] both to the multilinear [12–14,27,33] and multi-parameter [23,30,32,35] settings. The proofs of the representation theorems appear to be now converging to their final and most elegant form, and the arguments are simultaneously efficient and sharp.

Linear bi-parameter SIOs, for example, have kernels with singularities on $x_1 = y_1$ or $x_2 = y_2$, where $x, y \in \mathbb{R}^d$ are written as $x = (x_1, x_2)$, $y = (y_1, y_2) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ for a fixed partition $d = d_1 + d_2$. For $x, y \in \mathbb{C} = \mathbb{R} \times \mathbb{R}$, compare e.g. the one-parameter

Beurling kernel $1/(x - y)^2$ with the bi-parameter kernel $1/[(x_1 - y_1)(x_2 - y_2)]$—the product of Hilbert kernels in both coordinate directions. In general, the product space analysis is quite different from one-parameter analysis and seems to resist many techniques—in part due to the failure of bi-parameter sparse domination methods, see [3] (see also [4] however), representation theorems are even more important in bi-parameter than in one-parameter. For example, the dyadic representation methods have proved very fruitful in connection with bi-parameter commutators and weighted analysis, see Holmes–Petermichl–Wick [19], Ou–Petermichl–Strouse [36] and [28]. See also [1,2].

We discuss various applications throughout. For example, we prove the following two-weight estimate for commutators. The result (1) extends [29] and the result (2) extends [19] and [28].

Theorem 1.1 *Suppose that $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ is the underlying bi-parameter space, $p \in (1, \infty)$, $\mu, \lambda \in A_p(\mathbb{R}^d)$ are bi-parameter weights and $\nu = \mu^{1/p}\lambda^{-1/p} \in A_2(\mathbb{R}^d)$ is the Bloom weight.*

(1) *If T_i, $i = 1, 2$, is a one-parameter ω_i-CZO on $\mathbb{R}^{d_i}$, where $\omega_i \in \mathrm{Dini}_{3/2}$, then*

$$\|[T_1, [T_2, b]]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}_{\mathrm{prod}}(\nu)}.$$

(2) *Suppose that T is a bi-parameter (ω_1, ω_2)-CZO. Then we have*

$$\|[b_m, \cdots [b_2, [b_1, T]] \cdots]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \prod_{j=1}^m \|b_j\|_{\mathrm{bmo}(\nu^{1/m})}$$

if one of the following conditions holds:

  (a) *T is paraproduct free and $\omega_i \in \mathrm{Dini}_{m/2+1}$;*
  (b) *$m = 1$ and $\omega_i \in \mathrm{Dini}_{3/2}$;*
  (c) *$\omega_i \in \mathrm{Dini}_{m+1}$.*

See the main text for all of the definitions and for additional results. These Bloom-style two-weight estimates have recently been one of the main lines of development concerning commutators, see e.g. [1,2,18,19,25,26,28,29] for a non-exhaustive list.

2 Basic Notation and Fundamental Estimates

Throughout this paper $A \lesssim B$ means that $A \leq CB$ with some constant C that we deem unimportant to track at that point. We write $A \sim B$ if $A \lesssim B \lesssim A$.

Dyadic Notation. Given a dyadic grid $\mathcal{D}$, $I \in \mathcal{D}$ and $k \in \mathbb{Z}$, $k \geq 0$, we use the following notation:

(1) $\ell(I)$ is the side length of I.
(2) $I^{(k)} \in \mathcal{D}$ is the kth parent of I, i.e., $I \subset I^{(k)}$ and $\ell(I^{(k)}) = 2^k\ell(I)$.
(3) $\mathrm{ch}(I)$ is the collection of the children of I, i.e., $\mathrm{ch}(I) = \{J \in \mathcal{D}: J^{(1)} = I\}$.
(4) $E_I f = \langle f \rangle_I 1_I$ is the averaging operator, where $\langle f \rangle_I = \fint_I f = \frac{1}{|I|} \int_I f$.

(5) $E_{I,k}f$ is defined via

$$E_{I,k}f = \sum_{\substack{J \in \mathcal{D} \\ J^{(k)}=I}} E_J f.$$

(6) $\Delta_I f$ is the martingale difference $\Delta_I f = \sum_{J \in \text{ch}(I)} E_J f - E_I f$.

(7) $\Delta_{I,k}f$ is the martingale difference block

$$\Delta_{I,k}f = \sum_{\substack{J \in \mathcal{D} \\ J^{(k)}=I}} \Delta_J f.$$

(8) $P_{I,k}f$ is the following sum of martingale difference blocks

$$P_{I,k}f = \sum_{j=0}^{k} \Delta_{I,j}f = \sum_{\substack{J \in \mathcal{D} \\ J \subset I \\ \ell(J) \geq 2^{-k}\ell(I)}} \Delta_J f.$$

A fundamental fact is that we have the square function estimate

$$\|S_{\mathcal{D}}f\|_{L^p} \sim \|f\|_{L^p}, \quad p \in (1, \infty), \quad S_{\mathcal{D}}f := \left(\sum_{I \in \mathcal{D}} |\Delta_I f|^2\right)^{1/2}. \tag{2.1}$$

See e.g. [7,8] for even weighted $\|S_{\mathcal{D}}f\|_{L^p(w)} \sim \|f\|_{L^p(w)}$, $w \in A_p$, square function estimates and their history. A weight w (i.e. a locally integrable a.e. positive function) belongs to the weight class $A_p(\mathbb{R}^d)$, $1 < p < \infty$, if

$$[w]_{A_p(\mathbb{R}^d)} := \sup_Q \frac{1}{|Q|} \int_Q w \left(\frac{1}{|Q|} \int_Q w^{1-p'}\right)^{p-1} < \infty,$$

where the supremum is taken over all cubes $Q \subset \mathbb{R}^d$.

Lemma 2.2 *Let $p \in (1, \infty)$. There holds that*

$$\left\|\left(\sum_{K \in \mathcal{D}} |P_{K,k}f|^2\right)^{1/2}\right\|_{L^p} \sim \sqrt{k+1}\|f\|_{L^p}, \quad k \in \{0, 1, 2, \ldots\}.$$

Proof If $f_i \in L^p$ then

$$\left\|\left(\sum_{i=0}^{\infty} \sum_{I \in \mathcal{D}} |\Delta_I f_i|^2\right)^{1/2}\right\|_{L^p} \sim \left\|\left(\sum_{i=0}^{\infty} |f_i|^2\right)^{1/2}\right\|_{L^p}. \tag{2.3}$$

This follows by extrapolating the corresponding weighted L^2 version of (2.3), which, in turn, simply follows from $\|S_{\mathcal{D}} f\|_{L^2(w)} \sim \|f\|_{L^2(w)}$, $w \in A_2$. Recall that the classical extrapolation theorem of Rubio de Francia says that if $\|h\|_{L^{p_0}(w)} \lesssim \|g\|_{L^{p_0}(w)}$ for some $p_0 \in (1, \infty)$ and all $w \in A_{p_0}$, then $\|h\|_{L^p(w)} \lesssim \|g\|_{L^p(w)}$ for all $p \in (1, \infty)$ and all $w \in A_p$.

Let $K \in \mathcal{D}$. We have that

$$\sum_{I \in \mathcal{D}} |\Delta_I P_{K,k} f|^2 = \sum_{j=0}^{k} |\Delta_{K,j} f|^2.$$

Thus, (2.3) gives that

$$\left\| \left(\sum_{K \in \mathcal{D}} |P_{K,k} f|^2 \right)^{1/2} \right\|_{L^p} \sim \left\| \left(\sum_{K \in \mathcal{D}} \sum_{j=0}^{k} |\Delta_{K,j} f|^2 \right)^{1/2} \right\|_{L^p}$$

$$= \left\| \left(\sum_{j=0}^{k} \sum_{I \in \mathcal{D}} |\Delta_I f|^2 \right)^{1/2} \right\|_{L^p} \sim \sqrt{k+1} \|f\|_{L^p}.$$

$\square$

We will also have use for the Fefferman–Stein inequality

$$\left\| \left(\sum_{k} |M f_k|^2 \right)^{1/2} \right\|_{L^p(w)} \lesssim \left\| \left(\sum_{k} |f_k|^2 \right)^{1/2} \right\|_{L^p(w)}, \qquad p \in (1, \infty), w \in A_p,$$

where M is the Hardy–Littlewood maximal function. Often, the lighter Stein's inequality

$$\left\| \left(\sum_{I \in \mathcal{D}} |E_I f_I|^2 \right)^{1/2} \right\|_{L^p(w)} \lesssim \left\| \left(\sum_{I \in \mathcal{D}} |f_I|^2 \right)^{1/2} \right\|_{L^p(w)}, \qquad p \in (1, \infty), w \in A_p,$$

is sufficient.

For an interval $J \subset \mathbb{R}$ we denote by J_l and J_r the left and right halves of J, respectively. We define $h_J^0 = |J|^{-1/2} 1_J$ and $h_J^1 = |J|^{-1/2}(1_{J_l} - 1_{J_r})$. Let now $I = I_1 \times \cdots \times I_d \subset \mathbb{R}^d$ be a cube, and define the Haar function h_I^η, $\eta = (\eta_1, \ldots, \eta_d) \in \{0, 1\}^d$, by setting

$$h_I^\eta = h_{I_1}^{\eta_1} \otimes \cdots \otimes h_{I_d}^{\eta_d}.$$

If $\eta \neq 0$ the Haar function is cancellative: $\int h_I^\eta = 0$. We exploit notation by suppressing the presence of η, and write h_I for some h_I^η, $\eta \neq 0$. Notice that for $I \in \mathcal{D}$ we have $\Delta_I f = \langle f, h_I \rangle h_I$ (where the finite η summation is suppressed), $\langle f, h_I \rangle := \int f h_I$.

Bi-parameter Variants A weight $w(x_1, x_2)$ (i.e. a locally integrable a.e. positive function) belongs to the bi-parameter weight class $A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$, $1 < p < \infty$, if

$$[w]_{A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})} := \sup_R \frac{1}{|R|} \int_R w \left(\frac{1}{|R|} \int_R w^{1-p'} \right)^{p-1} < \infty,$$

where the supremum is taken over $R = I^1 \times I^2$ and each $I^i \subset \mathbb{R}^{d_i}$ is a cube. Thus, this is the one-parameter definition but cubes are replaced by rectangles.

We have

$$[w]_{A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})} < \infty \text{ iff } \max \left(\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{d_1}} [w(x_1, \cdot)]_{A_p(\mathbb{R}^{d_2})}, \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}^{d_2}} [w(\cdot, x_2)]_{A_p(\mathbb{R}^{d_1})} \right) < \infty,$$

and that

$$\max \left(\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{d_1}} [w(x_1, \cdot)]_{A_p(\mathbb{R}^{d_2})}, \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}^{d_2}} [w(\cdot, x_2)]_{A_p(\mathbb{R}^{d_1})} \right) \le [w]_{A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})},$$

while the constant $[w]_{A_p}$ is dominated by the maximum to some power. For basic bi-parameter weighted theory see e.g. [19]. We say $w \in A_\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ if

$$[w]_{A_\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})} := \sup_R \frac{1}{|R|} \int_R w \exp \left(\frac{1}{|R|} \int_R \log(w^{-1}) \right) < \infty.$$

It is well-known that

$$A_\infty(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}) = \bigcup_{1 < p < \infty} A_p(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}).$$

We do not have any important use for the A_∞ constant. The $w \in A_\infty$ assumption can always be replaced with the explicit assumption $w \in A_s$ for some $s \in (1, \infty)$, and then estimating everything with a dependence on $[w]_{A_s}$.

We denote a general dyadic grid in $\mathbb{R}^{d_i}$ by $\mathcal{D}^i$. We denote cubes in $\mathcal{D}^i$ by I^i, J^i, K^i, etc. Thus, our dyadic rectangles take the forms $I^1 \times I^2$, $J^1 \times J^2$, $K^1 \times K^2$ etc.

If A is an operator acting on $\mathbb{R}^{d_1}$, we can always let it act on the product space $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ by setting $A^1 f(x) = A(f(\cdot, x_2))(x_1)$. Similarly, we use the notation $A^2 f(x) = A(f(x_1, \cdot))(x_2)$ if A is originally an operator acting on $\mathbb{R}^{d_2}$. Our basic bi-parameter dyadic operators – martingale differences and averaging operators—are obtained by simply chaining together relevant one-parameter operators. For instance, a bi-parameter martingale difference is $\Delta_R f = \Delta_{I^1}^1 \Delta_{I^2}^2 f$, $R = I^1 \times I^2$. Bi-parameter estimates, such as the square function bound

$$\left\| \left(\sum_{R \in \mathcal{D}^1 \times \mathcal{D}^2} |\Delta_R f|^2 \right)^{1/2} \right\|_{L^p(w)} = \left\| \left(\sum_{I^i \in \mathcal{D}^i} |\Delta_{I^1}^1 \Delta_{I^2}^2 f|^2 \right)^{1/2} \right\|_{L^p(w)} \sim \|f\|_{L^p(w)},$$

where $p \in (1, \infty)$ and w is a bi-parameter A_p weight, are easily obtained using vector-valued versions of the corresponding one-parameter estimates. The required vector-valued estimates, on the other hand, follow simply by extrapolating the obvious weighted $L^2(w)$ estimates.

We systematically collect maximal function and square function bounds now. First, some notation. When we integrate with respect to only one of the parameters we may e.g. write

$$\langle f, h_{I_1} \rangle_1 (x_2) := \int_{\mathbb{R}^{d_1}} f(x_1, x_2) h_{I_1}(x_1) \, dx_1.$$

If $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$ we define the dyadic bi-parameter maximal function

$$M_{\mathcal{D}} f := \sup_{R \in \mathcal{D}} 1_R \langle |f| \rangle_R.$$

Now define the square functions

$$S_{\mathcal{D}} f = \left(\sum_{R \in \mathcal{D}} |\Delta_R f|^2 \right)^{1/2}, \quad S_{\mathcal{D}^1}^1 f = \left(\sum_{I^1 \in \mathcal{D}^1} |\Delta_{I^1}^1 f|^2 \right)^{1/2}$$

and define $S_{\mathcal{D}^2}^2 f$ analogously. Define also

$$S_{\mathcal{D},M}^1 f = \left(\sum_{I^1 \in \mathcal{D}^1} \frac{1_{I^1}}{|I^1|} \otimes \left[M_{\mathcal{D}^2} \langle f, h_{I^1} \rangle_1 \right]^2 \right)^{1/2},$$

$$S_{\mathcal{D},M}^2 f = \left(\sum_{I^2 \in \mathcal{D}^2} \left[M_{\mathcal{D}^1} \langle f, h_{I^2} \rangle_2 \right]^2 \otimes \frac{1_{I^2}}{|I^2|} \right)^{1/2}.$$

Let $k = (k_1, k_2)$, where $k_i \in \{0, 1, 2, \ldots, \}$, and $K = K^1 \times K^2 \in \mathcal{D}$. We set

$$P_{K^1, k_1}^1 f = \sum_{\substack{I^1 \in \mathcal{D}^1 \\ I^1 \subset K^1 \\ \ell(I^1) \geq 2^{-k_1} \ell(K^1)}} \Delta_{I^1}^1 f$$

and define similarly P_{K^2, k_2}^2. Then, we define $P_{K,k} := P_{K^1, k_1}^1 P_{K^2, k_2}^2$.

Lemma 2.4 *For $p \in (1, \infty)$ and a bi-parameter weight $w \in A_p$ we have*

$$\| f \|_{L^p(w)} \sim \| S_{\mathcal{D}} f \|_{L^p(w)} \sim \| S_{\mathcal{D}^1}^1 f \|_{L^p(w)} \sim \| S_{\mathcal{D}^2}^2 f \|_{L^p(w)}.$$

For $k = (k_1, k_2)$, $k_i \in \{0, 1, \ldots, \}$, we have the estimates

$$\left\| \left(\sum_{K \in \mathcal{D}} |P_{K,k} f|^2 \right)^{1/2} \right\|_{L^p(w)} \lesssim \sqrt{k_1 + 1} \sqrt{k_2 + 1} \|f\|_{L^p(w)},$$

$$\left\| \left(\sum_{K^1 \in \mathcal{D}^1} |P^1_{K^1,k_1} f|^2 \right)^{1/2} \right\|_{L^p(w)} \lesssim \sqrt{k_1 + 1} \|f\|_{L^p(w)}$$

and the analogous estimate with $P^2_{K^2,k_2}$.

Moreover, for $p, s \in (1, \infty)$ we have the Fefferman–Stein inequality

$$\left\| \left(\sum_j |Mf_j|^s \right)^{1/s} \right\|_{L^p(w)} \lesssim \left\| \left(\sum_j |f_j|^s \right)^{1/s} \right\|_{L^p(w)}.$$

Here M can e.g. be $M^1_{\mathcal{D}^1}$ or $M_{\mathcal{D}}$. Finally, we have

$$\|S^1_{\mathcal{D},M} f\|_{L^p(w)} + \|S^2_{\mathcal{D},M} f\|_{L^p(w)} \lesssim \|f\|_{L^p(w)}.$$

3 Bi-parameter Singular Integrals

Bi-parameter SIOs We say that ω is a modulus of continuity if it is an increasing and subadditive function with $\omega(0) = 0$. A relevant quantity is the modified Dini condition

$$\|\omega\|_{\mathrm{Dini}_\alpha} := \int_0^1 \omega(t) \left(1 + \log \frac{1}{t} \right)^\alpha \frac{dt}{t}, \qquad \alpha \geq 0. \tag{3.1}$$

In practice, the quantity (3.1) arises as follows:

$$\sum_{k=1}^\infty \omega(2^{-k}) k^\alpha = \sum_{k=1}^\infty \frac{1}{\log 2} \int_{2^{-k}}^{2^{-k+1}} \omega(2^{-k}) k^\alpha \frac{dt}{t} \lesssim \int_0^1 \omega(t) \left(1 + \log \frac{1}{t} \right)^\alpha \frac{dt}{t}. \tag{3.2}$$

For many standard arguments $\alpha = 0$ is enough. For the $T1$ type arguments we will always need $\alpha = 1/2$. Some further applications can require a higher α.

Let $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and consider an n-linear operator T on $\mathbb{R}^d$. We define what it means for T to be an n-linear bi-parameter SIO. Let ω_i be a modulus of continuity on $\mathbb{R}^{d_i}$. Let $f_j = f^1_j \otimes f^2_j$, $j = 1, \ldots, n+1$.

First, we set up notation for the adjoints of T. We let T^{j*}, $j \in \{0, \ldots, n\}$, denote the full adjoints, i.e., $T^{0*} = T$ and otherwise

$$\langle T(f_1, \ldots, f_n), f_{n+1}\rangle = \langle T^{j*}(f_1, \ldots, f_{j-1}, f_{n+1}, f_{j+1}, \ldots, f_n), f_j\rangle.$$

A subscript 1 or 2 denotes a partial adjoint in the given parameter—for example, we define

$$\langle T(f_1, \ldots, f_n), f_{n+1}\rangle = \langle T_1^{j*}(f_1, \ldots, f_{j-1}, f_{n+1}^1 \otimes f_j^2, f_{j+1}, \ldots, f_n), f_j^1 \otimes f_{n+1}^2\rangle.$$

Finally, we can take partial adjoints with respect to different parameters in different slots also—in that case we denote the adjoint by $T_{1,2}^{j_1*, j_2*}$. It simply interchanges the functions $f_{j_1}^1$ and f_{n+1}^1 and the functions $f_{j_2}^2$ and f_{n+1}^2. Of course, we e.g. have $T_{1,2}^{j*, j*} = T^{j*}$ and $T_{1,2}^{0*, j*} = T_2^{j*}$, so everything can be obtained, if desired, with the most general notation $T_{1,2}^{j_1*, j_2*}$. In any case, there are $(n+1)^2$ adjoints (including T itself). Similarly, the dyadic model operators that we later define always have $(n+1)^2$ different forms.

Full Kernel Representation Here we assume that given $m \in \{1, 2\}$ there exists $j_1, j_2 \in \{1, \ldots, n+1\}$ so that spt $f_{j_1}^m \cap$ spt $f_{j_2}^m = \emptyset$. In this case we demand that

$$\langle T(f_1, \ldots, f_n), f_{n+1}\rangle = \int_{\mathbb{R}^{(n+1)d}} K(x_{n+1}, x_1, \ldots, x_n) \prod_{j=1}^{n+1} f_j(x_j) \, dx,$$

where

$$K : \mathbb{R}^{(n+1)d} \setminus \{(x_{n+1}, x_1, \ldots, x_n) \in \mathbb{R}^{(n+1)d} : x_1^1 = \cdots = x_{n+1}^1 \text{ or}$$
$$x_1^2 = \cdots = x_{n+1}^2\} \to \mathbb{C}$$

is a kernel satisfying a set of estimates which we specify next.

The kernel K is assumed to satisfy the size estimate

$$|K(x_{n+1}, x_1, \ldots, x_n)| \lesssim \prod_{m=1}^{2} \frac{1}{\left(\sum_{j=1}^{n} |x_{n+1}^m - x_j^m|\right)^{d_m n}}.$$

We also require the following continuity estimates—to which we continue to refer to as Hölder estimates despite the general continuity moduli. For example, we require that we have

$$|K(x_{n+1}, x_1, \ldots, x_n) - K(x_{n+1}, x_1, \ldots, x_{n-1}, (c^1, x_n^2))$$
$$- K((x_{n+1}^1, c^2), x_1, \ldots, x_n) + K((x_{n+1}^1, c^2), x_1, \ldots, x_{n-1}, (c^1, x_n^2))|$$

$$
\lesssim \omega_1 \left(\frac{|x_n^1 - c^1|}{\sum_{j=1}^n |x_{n+1}^1 - x_j^1|} \right) \frac{1}{\left(\sum_{j=1}^n |x_{n+1}^1 - x_j^1| \right)^{d_1 n}}
$$

$$
\times \, \omega_2 \left(\frac{|x_{n+1}^2 - c^2|}{\sum_{j=1}^n |x_{n+1}^2 - x_j^2|} \right) \frac{1}{\left(\sum_{j=1}^n |x_{n+1}^2 - x_j^2| \right)^{d_2 n}}
$$

whenever $|x_n^1 - c^1| \le 2^{-1} \max_{1 \le i \le n} |x_{n+1}^1 - x_i^1|$ and $|x_{n+1}^2 - c^2| \le 2^{-1} \max_{1 \le i \le n} |x_{n+1}^2 - x_i^2|$. Of course, we also require all the other natural symmetric estimates, where c^1 can be in any of the given $n + 1$ slots and similarly for c^2. There are, of course, $(n + 1)^2$ different estimates.

Finally, we require the following mixed Hölder and size estimates. For example, we ask that

$$
|K(x_{n+1}, x_1, \ldots, x_n) - K(x_{n+1}, x_1, \ldots, x_{n-1}, (c^1, x_n^2))|
$$

$$
\lesssim \omega_1 \left(\frac{|x_n^1 - c^1|}{\sum_{j=1}^n |x_{n+1}^1 - x_j^1|} \right) \frac{1}{\left(\sum_{j=1}^n |x_{n+1}^1 - x_j^1| \right)^{d_1 n}} \cdot \frac{1}{\left(\sum_{j=1}^n |x_{n+1}^2 - x_j^2| \right)^{d_2 n}}
$$

whenever $|x_n^1 - c^1| \le 2^{-1} \max_{1 \le i \le n} |x_{n+1}^1 - x_i^1|$. Again, we also require all the other natural symmetric estimates.

Partial Kernel Representations Suppose now only that there exists $j_1, j_2 \in \{1, \ldots, n+1\}$ so that spt $f_{j_1}^1 \cap$ spt $f_{j_2}^1 = \emptyset$. Then we assume that

$$
\langle T(f_1, \ldots, f_n), f_{n+1} \rangle = \int_{\mathbb{R}^{(n+1)d_1}} K_{(f_j^2)}(x_{n+1}^1, x_1^1, \ldots, x_n^1) \prod_{j=1}^{n+1} f_j^1(x_j^1) \, \mathrm{d}x^1,
$$

where $K_{(f_j^2)}$ is a one-parameter ω_1-Calderón–Zygmund kernel as e.g. in [17] but with a constant depending on the fixed functions $f_1^2, \ldots, f_{n+1}^2$. For example, this means that the size estimate takes the form

$$
|K_{(f_j^2)}(x_{n+1}^1, x_1^1, \ldots, x_n^1)| \le C(f_1^2, \ldots, f_{n+1}^2) \frac{1}{\left(\sum_{j=1}^n |x_{n+1}^1 - x_j^1| \right)^{d_1 n}}.
$$

The continuity estimates are analogous.

We assume the following $T1$ type control on the constant $C(f_1^2, \ldots, f_{n+1}^2)$. We have

$$
C(1_{I^2}, \ldots, 1_{I^2}) \lesssim |I^2| \tag{3.3}
$$

and

$$
C(a_{I^2}, 1_{I^2}, \ldots, 1_{I^2}) + C(1_{I^2}, a_{I^2}, 1_{I^2}, \ldots, 1_{I^2}) + \cdots + C(1_{I^2}, \ldots, 1_{I^2}, a_{I^2}) \lesssim |I^2|
$$

for all cubes $I^2 \subset \mathbb{R}^{d_2}$ and all functions a_{I^2} satisfying $a_{I^2} = 1_{I^2} a_{I^2}$, $|a_{I^2}| \leq 1$ and $\int a_{I^2} = 0$.

Analogous partial kernel representation on the second parameter is assumed when spt $f_{j_1}^2 \cap$ spt $f_{j_2}^2 = \emptyset$ for some j_1, j_2.

Definition 3.4 If T is an n-linear operator with full and partial kernel representations as defined above, we call T an n-linear bi-parameter (ω_1, ω_2)-SIO.

Bi-parameter CZOs We say that T satisfies the weak boundedness property if

$$|\langle T(1_R, \dots, 1_R), 1_R \rangle| \lesssim |R| \tag{3.5}$$

for all rectangles $R = I^1 \times I^2 \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

An SIO T satisfies the diagonal BMO assumption if the following holds. For all rectangles $R = I^1 \times I^2 \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and functions a_{I^i} with $a_{I^i} = 1_{I^i} a_{I^i}$, $|a_{I^i}| \leq 1$ and $\int a_{I^i} = 0$ we have

$$|\langle T(a_{I^1} \otimes 1_{I^2}, 1_R, \dots, 1_R), 1_R \rangle| + \cdots + |\langle T(1_R, \dots, 1_R), a_{I^1} \otimes 1_{I^2} \rangle| \lesssim |R| \tag{3.6}$$

and

$$|\langle T(1_{I^1} \otimes a_{I^2}, 1_R, \dots, 1_R), 1_R \rangle| + \cdots + |\langle T(1_R, \dots, 1_R), 1_{I^1} \otimes a_{I^2} \rangle| \lesssim |R|.$$

The product BMO space is originally by Chang and Fefferman [5,6], and it is the right bi-parameter BMO space for many considerations. An SIO T satisfies the product BMO assumption if it holds

$$S1 \in \mathrm{BMO}_{\mathrm{prod}}$$

for all the $(n+1)^2$ adjoints $S = T_{1,2}^{j_1*, j_2*}$. Here $S1 := S(1, \dots, 1)$. This can be interpreted in the sense that

$$\|S1\|_{\mathrm{BMO}_{\mathrm{prod}}} = \sup_{\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2} \sup_{\Omega} \left(\frac{1}{|\Omega|} \sum_{\substack{R = I^1 \times I^2 \in \mathcal{D} \\ R \subset \Omega}} |\langle S1, h_R \rangle|^2 \right)^{1/2} < \infty,$$

where $h_R = h_{I^1} \otimes h_{I^2}$, the supremum is over all dyadic grids $\mathcal{D}^i$ on $\mathbb{R}^{d_i}$ and open sets $\Omega \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with $0 < |\Omega| < \infty$, and the pairings $\langle S1, h_R \rangle$ can be defined, in a natural way, using the kernel representations.

Definition 3.7 An n-linear bi-parameter (ω_1, ω_2)-SIO T satisfying the weak boundedness property, the diagonal BMO assumption and the product BMO assumption is called an n-linear bi-parameter (ω_1, ω_2)-Calderón–Zygmund operator $((\omega_1, \omega_2)$-CZO).

Bi-parameter Model Operators For hybrid operators we will use suggestive notation, such as, $(S\pi)_i$ to denote a bi-parameter operator that behaves like an ordinary n-linear shift S_i on the first parameter and like an n-linear paraproduct π on the second—but this is just notation and our operators are not of tensor product form.

Shifts Let $i = (i_1, \ldots, i_{n+1})$, where $i_j = (i_j^1, i_j^2) \in \{0, 1, \ldots\}^2$. An n-linear bi-parameter shift S_i takes the form

$$\langle S_i(f_1, \ldots, f_n), f_{n+1} \rangle = \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(i_j)} = K}} a_{K,(R_j)} \prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{R_j} \rangle.$$

Here $K, R_1, \ldots, R_{n+1} \in \mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$, $R_j = I_j^1 \times I_j^2$, $R_j^{(i_j)} := (I_j^1)^{(i_j^1)} \times (I_j^2)^{(i_j^2)}$ and $\widetilde{h}_{R_j} = \widetilde{h}_{I_j^1} \otimes \widetilde{h}_{I_j^2}$. Here we assume that for $m \in \{1, 2\}$ there exist two indices $j_0, j_1 \in \{1, \ldots, n+1\}$, $j_0 \neq j_1$, so that $\widetilde{h}_{I_{j_0}^m} = h_{I_{j_0}^m}$, $\widetilde{h}_{I_{j_1}^m} = h_{I_{j_1}^m}$ and for the remaining indices $j \notin \{j_0, j_1\}$ we have $\widetilde{h}_{I_j^m} \in \{h_{I_j^m}^0, h_{I_j^m}\}$. Moreover, $a_{K,(R_j)} = a_{K,R_1,\ldots,R_{n+1}}$ is a scalar satisfying the normalization

$$|a_{K,(R_j)}| \leq \frac{\prod_{j=1}^{n+1} |R_j|^{1/2}}{|K|^n}. \tag{3.8}$$

We continue to define modified shifts—they are important for the weak kernel regularity. Let

$$A_{R_1,\ldots,R_{n+1}}^{j_1,j_2}(f_1, \ldots, f_{n+1}) = A_{R_1,\ldots,R_{n+1}}^{j_1,j_2} := \prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{R_j} \rangle, \tag{3.9}$$

where $\widetilde{h}_{R_j} = \widetilde{h}_{I_j^1} \otimes \widetilde{h}_{I_j^2}$, $\widetilde{h}_{I_{j_1}^1} = h_{I_{j_1}^1}$, $\widetilde{h}_{I_j^1} = h_{I_j^1}^0$, $j \neq j_1$, $\widetilde{h}_{I_{j_2}^2} = h_{I_{j_2}^2}$, $\widetilde{h}_{I_j^2} = h_{I_j^2}^0$, $j \neq j_2$. A modified n-linear bi-parameter shift Q_k, $k = (k_1, k_2)$, takes the form

$$\langle Q_k(f_1, \ldots, f_n), f_{n+1} \rangle = \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k)} = K}} a_{K,(R_j)} \Big[A_{R_1,\ldots,R_{n+1}}^{j_1,j_2} - A_{I_{j_1}^1 \times I_1^2, \ldots, I_{j_1}^1 \times I_{n+1}^2}^{j_1,j_2}$$

$$- A_{I_1^1 \times I_{j_2}^2, \ldots, I_{n+1}^1 \times I_{j_2}^2}^{j_1,j_2} + A_{I_{j_1}^1 \times I_{j_2}^2, \ldots, I_{j_1}^1 \times I_{j_2}^2}^{j_1,j_2} \Big]$$

for some j_1, j_2. Moreover, $a_{K,(R_j)} = a_{K,R_1,\ldots,R_{n+1}}$ is a scalar satisfying the usual normalization (3.8).

We now define the hybrid operators that behave like a modified shift in one of the parameters and like a standard shift in the other. A modified/standard n-linear

bi-parameter shift $(QS)_{k,i}$, $i = (i_1, \ldots, i_{n+1})$, $k, i_j \in \{0, 1, \ldots\}$, takes the form

$$\langle (QS)_{k,i}(f_1, \ldots, f_n), f_{n+1} \rangle$$

$$= \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k,i_j)} = K}} a_{K,(R_j)} \left[\prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{R_j} \rangle - \prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{I_{j_0}^1 \times I_j^2} \rangle \right]$$

for some j_0. Here we assume that $\widetilde{h}_{I_{j_0}^1} = h_{I_{j_0}^1}$, $\widetilde{h}_{I_j^1} = h_{I_j^1}^0$ for $j \neq j_0$, and that there exist two indices $j_1, j_2 \in \{1, \ldots, n+1\}$, $j_1 \neq j_2$, so that $\widetilde{h}_{I_{j_1}^2} = h_{I_{j_1}^2}$, $\widetilde{h}_{I_{j_2}^2} = h_{I_{j_2}^2}$ and for the remaining indices $j \notin \{j_1, j_2\}$ we have $\widetilde{h}_{I_j^2} \in \{h_{I_j^2}^0, h_{I_j^2}\}$. Moreover, $a_{K,(R_j)} = a_{K, R_1, \ldots, R_{n+1}}$ is a scalar satisfying the usual normalization (3.8). Of course, $(SQ)_{i,k}$ is defined symmetrically.

Partial Paraproducts Partial paraproducts are hybrids of π and S or π and Q.

Let $i = (i_1, \ldots, i_{n+1})$, where $i_j \in \{0, 1, \ldots\}$. An n-linear bi-parameter partial paraproduct $(S\pi)_i$ with the paraproduct component on $\mathbb{R}^{d_2}$ takes the form

$$\langle (S\pi)_i(f_1, \ldots, f_n), f_{n+1} \rangle = \sum_{K=K^1 \times K^2} \sum_{\substack{I_1^1, \ldots, I_{n+1}^1 \\ (I_j^1)^{(i_j)} = K^1}} a_{K,(I_j^1)} \prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{I_j^1} \otimes u_{j,K^2} \rangle,$$

$$\tag{3.10}$$

where the functions $\widetilde{h}_{I_j^1}$ and u_{j,K^2} satisfy the following. There are $j_0, j_1 \in \{1, \ldots, n+1\}$, $j_0 \neq j_1$, so that $\widetilde{h}_{I_{j_0}^1} = h_{I_{j_0}^1}$, $\widetilde{h}_{I_{j_1}^1} = h_{I_{j_1}^1}$ and for the remaining indices $j \notin \{j_0, j_1\}$ we have $\widetilde{h}_{I_j^1} \in \{h_{I_j^1}^0, h_{I_j^1}\}$. There is $j_2 \in \{1, \ldots, n+1\}$ so that $u_{j_2, K^2} = h_{K^2}$ and for the remaining indices $j \neq j_2$ we have $u_{j,K^2} = \frac{1_{K^2}}{|K^2|}$. Moreover, the coefficients are assumed to satisfy

$$\| (a_{K,(I_j^1)})_{K^2} \|_{\mathrm{BMO}} \leq \frac{\prod_{j=1}^{n+1} |I_j^1|^{1/2}}{|K^1|^n}.$$

Of course, $(\pi S)_i$ is defined symmetrically.

A modified n-linear partial paraproduct $(Q\pi)_k$ with the paraproduct component on $\mathbb{R}^{d_2}$ takes the form

$$\langle (Q\pi)_k(f_1, \ldots, f_n), f_{n+1} \rangle$$

$$= \sum_{K=K^1 \times K^2} \sum_{\substack{I_1^1, \ldots, I_{n+1}^1 \\ (I_j^1)^{(k)} = K^1}} a_{K,(I_j^1)} \left[\prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{I_j^1} \otimes u_{j,K^2} \rangle - \prod_{j=1}^{n+1} \langle f_j, \widetilde{h}_{I_{j_0}^1} \otimes u_{j,K^2} \rangle \right]$$

for some j_0—here $\widetilde{h}_{I^1_{j_0}} = h_{I^1_{j_0}}, \widetilde{h}_{I^1_j} = h^0_{I^1_j}$ for $j \neq j_0$ and u_{j,K^2} are like in (3.10). The constants satisfy the same normalization.

Full Paraproducts An n-linear bi-parameter full paraproduct Π takes the form

$$\langle \Pi(f_1, \ldots, f_n), f_{n+1} \rangle = \sum_{K=K^1 \times K^2} a_K \prod_{j=1}^{n+1} \langle f_j, u_{j,K^1} \otimes u_{j,K^2} \rangle,$$

where the functions u_{j,K^1} and u_{j,K^2} are like in (3.10). The coefficients are assumed to satisfy

$$\|(a_K)\|_{\text{BMO}_{\text{prod}}} = \sup_\Omega \left(\frac{1}{|\Omega|} \sum_{K \subset \Omega} |a_K|^2 \right)^{1/2} \leq 1,$$

where the supremum is over open sets $\Omega \subset \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with $0 < |\Omega| < \infty$.

Comparison to the Usual Model Operators The modified model operators can be written as suitable sums of the standard operators. This is practical when one is willing to lose $\frac{1}{2}$ of kernel regularity or if some estimates are too difficult to carry out for the more complicated modified operators. However, some regularity is always lost if this decomposition is used, so it is preferable to make do without it. To communicate the gist we only give the following formulation.

Lemma 3.11 *Let Q_k, $k = (k_1, k_2)$, be a modified n-linear bi-parameter shift. Then*

$$Q_k = C \sum_{u=1}^{c} \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} S^{u,i_1,i_2},$$

where each $S = S^{u,i_1,i_2}$ is a standard n-linear bi-parameter shift of complexity $i^m_{S,j}$, $j \in \{1, \ldots, n+1\}$, $m \in \{1, 2\}$, satisfying

$$i^m_{S,j} \leq k_m.$$

Similarly, a modified/standard shift can be represented using standard shifts and a modified partial paraproduct can be represented using standard partial paraproducts.

Proof For notational convenience we consider a shift Q_k of the particular form

$$\langle Q_k(f_1, \ldots, f_n), f_{n+1} \rangle = \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R^{(k)}_j = K}} a_{K,(R_j)} \left[A^{n+1,n+1}_{R_1, \ldots, R_{n+1}} - A^{n+1,n+1}_{I^1_{n+1} \times I^2_1, \ldots, I^1_{n+1} \times I^2_{n+1}} \right.$$

$$\left. - A^{n+1,n+1}_{I^1_1 \times I^2_{n+1}, \ldots, I^1_{n+1} \times I^2_{n+1}} + A^{n+1,n+1}_{I^1_{n+1} \times I^2_{n+1}, \ldots, I^1_{n+1} \times I^2_{n+1}} \right]. \tag{3.12}$$

There is no essential difference in the general case.

We define

$$b_{K,(R_j)} = |R_1|^{n/2} a_{K,(R_j)}$$

and

$$B_{R_1,\ldots,R_{n+1}}^{n+1,n+1} = \prod_{j=1}^{n} \langle f_j \rangle_{R_j} \langle f_{n+1}, h_{R_{n+1}} \rangle.$$

We can write the shift with these similarly as in (3.12) just by replacing a with b and A with B.

For the moment we define the following shorthand. For a cube I and integers $l, j_0 \in \{1, 2, \ldots\}$ we define

$$D_{I,l}(j, j_0) = \begin{cases} E_I, & \text{if } j \in \{1, \ldots, j_0 - 1\}, \\ P_{I,l-1}, & \text{if } j = j_0, \\ \text{id}, & \text{if } j \in \{j_0 + 1, j_0 + 2, \ldots\}, \end{cases} \tag{3.13}$$

where id denotes the identity operator.

Let $R_1, \ldots, R_{n+1}$ be as in the summation of Q_k. We use the above notation in both parameters, and we denote this, as usual, with superscripts $D_{I,l}^1(j, j_0)$ and $D_{I,l}^2(j, j_0)$. With some work (we omit the details) it can be shown that

$$B_{R_1,\ldots,R_{n+1}}^{n+1,n+1} = \sum_{m_1,m_2=1}^{n+1} \prod_{j=1}^{n} \langle D_{K^1,k_1}^1(j, m_1) D_{K^2,k_2}^2(j, m_2) f_j \rangle_{R_j} \langle f_{n+1}, h_{R_{n+1}} \rangle,$$

which gives that

$$\sum_K \sum_{\substack{R_1,\ldots,R_{n+1} \\ R_j^{(k)}=K}} B_{R_1,\ldots,R_{n+1}}^{n+1,n+1} =: \sum_{m_1,m_2=1}^{n+1} \Sigma_{m_1,m_2}^1.$$

Also, we have that

$$B_{I_{n+1}^1 \times I_1^2,\ldots,I_{n+1}^1 \times I_{n+1}^2}^{n+1,n+1} = \sum_{m_2=1}^{n+1} \prod_{j=1}^{n} \langle D_{K^2,k_2}^2(j, m_2) f_j \rangle_{I_{n+1}^1 \times I_j^2} \langle f_{n+1}, h_{R_{n+1}} \rangle$$

and

$$B_{I_1^1 \times I_{n+1}^2,\ldots,I_{n+1}^1 \times I_{n+1}^2}^{n+1,n+1} = \sum_{m_1=1}^{n+1} \prod_{j=1}^{n} \langle D_{K^1,k_1}^1(j, m_1) f_j \rangle_{I_j^1 \times I_{n+1}^2} \langle f_{n+1}, h_{R_{n+1}} \rangle,$$

which gives that

$$\sum_K \sum_{\substack{R_1,\ldots,R_{n+1} \\ R_j^{(k)}=K}} B^{n+1,n+1}_{I^1_{n+1}\times I^2_1,\ldots,I^1_{n+1}\times I^2_{n+1}} =: \sum_{m_2=1}^{n+1} \Sigma^2_{m_2}$$

and

$$\sum_K \sum_{\substack{R_1,\ldots,R_{n+1} \\ R_j^{(k)}=K}} B^{n+1,n+1}_{I^1_1\times I^2_{n+1},\ldots,I^1_{n+1}\times I^2_{n+1}} =: \sum_{m_1=1}^{n+1} \Sigma^3_{m_1}.$$

Finally, we write that

$$\sum_K \sum_{\substack{R_1,\ldots,R_{n+1} \\ R_j^{(k)}=K}} B^{n+1,n+1}_{I^1_{n+1}\times I^2_{n+1},\ldots,I^1_{n+1}\times I^2_{n+1}} =: \Sigma^4.$$

Using the above decompositions we have the identity

$$\langle Q_k(f_1,\ldots,f_n), f_{n+1}\rangle = \sum_{m_1,m_2=1}^{n} \Sigma^1_{m_1,m_2} + \sum_{m_2=1}^{n} (\Sigma^1_{n+1,m_2} - \Sigma^2_{m_2})$$
$$+ \sum_{m_1=1}^{n} (\Sigma^1_{m_1,n+1} - \Sigma^3_{m_1}) + (\Sigma^1_{n+1,n+1} - \Sigma^2_{n+1} - \Sigma^3_{n+1} + \Sigma^4).$$

The terms $\Sigma^1_{m_1,m_2}$ with $m_1, m_2 \in \{1,\ldots,n\}$ and the terms inside the parentheses will be written as sums of standard shifts.

First, we take one $\Sigma^1_{m_1,m_2}$ with $m_1, m_2 \in \{1,\ldots,n\}$. For convenience of notation we choose the case $m_1 = m_2 =: m$. Recall that

$$\Sigma^1_{m,m} = \sum_K \sum_{\substack{R_1,\ldots,R_{n+1} \\ R_j^{(k)}=K}} b_{K,(R_j)} \prod_{j=1}^{m-1} \langle f_j\rangle_K \langle P_{K,(k_1-1,k_2-1)} f_m\rangle_{R_m} \prod_{j=m+1}^{n} \langle f_j\rangle_{R_j} \langle f_{n+1}, h_{R_{n+1}}\rangle.$$

Expanding

$$\langle P_{K,(k_1-1,k_2-1)} f_m\rangle_{R_m} = \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \sum_{L^{(i_1,i_2)}=K} \langle f_m, h_L\rangle \langle h_L\rangle_{R_m}$$

there holds that

$$\Sigma_{m,m}^1 = \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \sum_{K} \sum_{\substack{L^{(i_1,i_2)}=K \\ R_j^{(k)}=K}} \sum_{R_{m+1},\ldots,R_{n+1}} \left(\sum_{\substack{R_1,\ldots,R_{m-1} \\ R_j^{(k)}=K}} \sum_{\substack{R_m \subset L \\ R_m^{(k)}=K}} \frac{b_{K,(R_j)}\langle h_L \rangle_{R_m}}{|K|^{(m-1)/2}|R_{n+1}|^{(n-m)/2}} \right)$$

$$\prod_{j=1}^{m-1} \langle f_j, h_K^0 \rangle \langle f_m, h_L \rangle \prod_{j=m+1}^{n} \langle f_j, h_{R_j}^0 \rangle \langle f_{n+1}, h_{R_{n+1}} \rangle.$$

Since

$$\left| \sum_{\substack{R_1,\ldots,R_{m-1} \\ R_j^{(k)}=K}} \sum_{\substack{R_m \subset L \\ R_m^{(k)}=K}} \frac{b_{K,(R_j)}\langle h_L \rangle_{R_m}}{|K|^{(m-1)/2}|R_{n+1}|^{(n-m)/2}} \right| \leq \frac{|K|^{(m-1)/2}|L|^{1/2}|R_{n+1}|^{(n-m+1)/2}}{|K|^n},$$

we see that

$$\Sigma_{m,m}^1 = \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \langle S_{(0,\ldots,0,(i_1,i_2),k,\ldots,k)}(f_1,\ldots,f_n), f_{n+1} \rangle,$$

where $S_{(0,\ldots,0,(i_1,i_2),k,\ldots,k)}$ is a standard n-linear bi-parameter shift. The case of general m_1, m_2 is analogous.

We turn to the terms $\Sigma_{n+1,m_2}^1 - \Sigma_{m_2}^2$. The terms $\Sigma_{m_1,n+1}^1 - \Sigma_{m_1}^3$ are symmetrical. Let $m_2 \in \{1,\ldots,n\}$. After expanding P_{K^2,k_2-1}^2 in the slot m_2 we have that $\Sigma_{n+1,m_2}^1 - \Sigma_{m_2}^2$ can be written as

$$\sum_{i_2=0}^{k_2-1} \sum_{K} \sum_{\substack{(L^2)^{(i_2)}=K^2 \\ R_j^{(k)}=K}} \sum_{R_1,\ldots,R_{n+1}} b_{K,(R_j)}\langle h_{L^2} \rangle_{I_{m_2}^2}$$

$$\left[\prod_{j=1}^{m_2-1} \langle f_j \rangle_K \left\langle f_{m_2}, \frac{1_{K^1}}{|K_1|} \otimes h_{L^2} \right\rangle \prod_{j=m_2+1}^{n} \langle f_j \rangle_{K^1 \times I_j^2} \right.$$

$$\left. - \prod_{j=1}^{m_2-1} \langle f_j \rangle_{I_{n+1}^1 \times K^2} \left\langle f_{m_2}, \frac{1_{I_{n+1}^1}}{|I_{n+1}^1|} \otimes h_{L^2} \right\rangle \prod_{j=m_2+1}^{n} \langle f_j \rangle_{I_{n+1}^1 \times I_j^2} \right] \langle f_{n+1}, h_{R_{n+1}} \rangle.$$

This splits the difference $\Sigma_{n+1,m_2}^1 - \Sigma_{m_2}^2$ as

$$\Sigma_{n+1,m_2}^1 - \Sigma_{m_2}^2 =: \sum_{i_2=0}^{k_2-1} \Sigma_{m_2,i_2}^{1,2}.$$

We fix one i_2 at this point.

Let $g_j^{m_2} := g_j = \langle f_j \rangle_{K^2}^2$ for $j \in \{1, \ldots, m_2 - 1\}$, $g_{m_2}^{m_2} := g_{m_2} = \langle f_{m_2}, h_{L^2} \rangle_2$ and $g_j^{m_2} := g_j = \langle f_j \rangle_{I_j^2}^2$ for $j \in \{m_2 + 1, \ldots, n\}$. Using this notation we have that the term inside the brackets is $\prod_{j=1}^n \langle g_j \rangle_{K^1} - \prod_{j=1}^n \langle g_j \rangle_{I_{n+1}^1}$. We write that

$$\prod_{j=1}^n \langle g_j \rangle_{K^1} - \prod_{j=1}^n \langle g_j \rangle_{I_{n+1}^1} = - \sum_{i_1=0}^{k_1-1} \left(\prod_{j=1}^n \langle g_j \rangle_{(I_{n+1}^1)^{(i_1)}} - \prod_{j=1}^n \langle g_j \rangle_{(I_{n+1}^1)^{(i_1+1)}} \right).$$

Then, we write $\prod_{j=1}^n \langle g_j \rangle_{(I_{n+1}^1)^{(i_1)}} - \prod_{j=1}^n \langle g_j \rangle_{(I_{n+1}^1)^{(i_1+1)}}$ as the sum

$$\sum_{m_1=1}^n \prod_{j=1}^{m_1-1} \langle g_j \rangle_{(I_{n+1}^1)^{(i_1+1)}} \langle \Delta_{(I_{n+1}^1)^{(i_1+1)}} g_{m_1} \rangle_{I_{n+1}^1} \prod_{j=m_1+1}^n \langle g_j \rangle_{(I_{n+1}^1)^{(i_1)}}.$$

Expanding

$$\langle \Delta_{(I_{n+1}^1)^{(i_1+1)}} g_{m_1} \rangle_{I_{n+1}^1} = \langle g_{m_1}, h_{(I_{n+1}^1)^{(i_1+1)}} \rangle \langle h_{(I_{n+1}^1)^{(i_1+1)}} \rangle_{I_{n+1}^1}$$

we get that $\prod_{j=1}^n \langle g_j \rangle_{K^1} - \prod_{j=1}^n \langle g_j \rangle_{I_{n+1}^1}$ equals

$$- \sum_{i_1=0}^{k_1-1} \sum_{m_1=1}^n \prod_{j=1}^{m_1-1} \langle g_j \rangle_{(I_{n+1}^1)^{(i_1+1)}} \langle g_{m_1}, h_{(I_{n+1}^1)^{(i_1+1)}} \rangle \langle h_{(I_{n+1}^1)^{(i_1+1)}} \rangle_{I_{n+1}^1} \prod_{j=m_1+1}^n \langle g_j \rangle_{(I_{n+1}^1)^{(i_1)}}.$$

This identity splits $\Sigma_{m_2,i_2}^{1,2}$ further as $\Sigma_{m_2,i_2}^{1,2} =: - \sum_{i_1=0}^{k_1-1} \sum_{m_1=1}^n \Sigma_{m_1,m_2,i_1,i_2}^{1,2}$.

We fix some m_1 and i_1 and consider the corresponding term. For convenience of notation we look at the case $m_1 = m_2 =: m$. There holds that

$$\Sigma_{m,m,i_1,i_2}^{1,2} = \sum_K \sum_{\substack{(L^2)^{(i_2)} = K^2}} \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k)} = K}} b_{K,(R_j)} \langle h_{(I_{n+1}^1)^{(i_1+1)} \times L^2} \rangle_{I_{n+1}^1 \times I_m^2}$$

$$\prod_{j=1}^{m-1} \langle f_j \rangle_{(I_{n+1}^1)^{(i_1+1)} \times K^2} \left\langle f_m, h_{(I_{n+1}^1)^{(i_1+1)} \times L^2} \right\rangle$$

$$\prod_{j=m+1}^n \langle f_j \rangle_{(I_{n+1}^1)^{(i_1)} \times I_j^2} \langle f_{n+1}, h_{R_{n+1}} \rangle.$$

This is seen as a standard shift once we reorganize the summation and verify the normalization. We take $(I_{n+1}^1)^{(i_1+1)}$ as the new "top cube" in the first parameter $((I_{n+1}^1)^{(i_1+1)}$ corresponds to $(L^1)^{(1)}$ in the summation below). There holds that

$\Sigma_{m,m,i_1,i_2}^{1,2}$ equals

$$
\sum_{K^1} \sum_{(L^1)^{(k_1-i_1)}=K^1} \sum_{(I_{n+1}^1)^{(i_1)}=L^1} \sum_{K^2} \sum_{(L^2)^{(i_2)}=K^2} \sum_{\substack{I_{m+1}^2,\dots,I_{n+1}^2 \\ (I_j^2)^{(k_2)}=K^2}} c_{K^1,L^1,I_{n+1}^1,K^2,L^2,I_{m+1}^2,\dots,I_{n+1}^2}
$$

$$
\prod_{j=1}^{m-1} \langle f_j \rangle_{(L^1)^{(1)} \times K^2} \Big\langle f_m, h_{(L^1)^{(1)} \times L^2} \Big\rangle \prod_{j=m+1}^{n} \langle f_j \rangle_{L^1 \times I_j^2} \langle f_{n+1}, h_{R_{n+1}} \rangle,
$$

where

$$
c_{K^1,L^1,I_{n+1}^1,K^2,L^2,I_{m+1}^2,\dots,I_{n+1}^2}
$$
$$
= \sum_{\substack{I_1^1,\dots,I_n^1 \\ (I_j^1)^{(k_1)}=K^1}} \sum_{\substack{I_1^2,\dots,I_{m-1}^2 \\ (I_j^2)^{(k_2)}=K^2}} \sum_{\substack{I_m^2 \subset L^2 \\ (I_m^2)^{(k_2)}=K^2}} b_{K,(R_j)} \langle h_{(L^1)^{(1)} \times L^2} \rangle_{I_{n+1}^1 \times I_m^2}.
$$

We have the estimate

$$
\left| c_{K^1,L^1,I_{n+1}^1,K^2,L^2,I_{m+1}^2,\dots,I_{n+1}^2} \right|
$$
$$
\leq \frac{|(L^1)^{(1)}|^{n/2} |I_{n+1}^1|^{1/2}}{|(L^1)^{(1)}|^n} \frac{|K^2|^{(m-1)/2} |L^2|^{1/2} |I_{n+1}^2|^{(n-m+1)/2}}{|K^2|^n}
$$
$$
\times |(L^1)^{(1)}|^{(n-1)/2} |K^2|^{(m-1)/2} |I^2|^{(n-m)/2}.
$$

Notice that the term in the first line in the right hand side is $2^{d_1(n-m)/2}$ times the right normalization of the shift, since in $\Sigma_{m,m,i_1,i_2}^{1,2}$ we have the cubes L^1 related to f_j with $j \in \{m+1,\dots,n\}$. Also, the term in the second line is almost cancelled out when one changes the averages in $\Sigma_{m,m,i_1,i_2}^{1,2}$ into pairings against non-cancellative Haar functions.

We conclude that for some $C \geq 1$ we have

$$
C^{-1} \Sigma_{m,m,i_1,i_2}^{1,2} = \langle S_{(0,\dots,0,(0,i_2),(1,k_2),\dots,(1,k_2),(i_1+1,k_2)}(f_1,\dots,f_n), f_{n+1} \rangle,
$$

where S is a standard n-linear bi-parameter shift of the given complexity. The case of general m_1, m_2 is analogous.

Finally, we look at the term $\Sigma^1_{n+1,n+1} - \Sigma^2_{n+1} - \Sigma^3_{n+1} + \Sigma^4$ which by definition is

$$
\sum_K \sum_{\substack{R_1,\ldots,R_{n+1} \\ R_j^{(k)}=K}} b_{K,(R_j)}
$$

$$
\left[\prod_{j=1}^n \langle f_j \rangle_K - \prod_{j=1}^n \langle f_j \rangle_{I^1_{n+1} \times K^2} - \prod_{j=1}^n \langle f_j \rangle_{K^1 \times I^2_{n+1}} \right. \tag{3.14}
$$

$$
\left. + \prod_{j=1}^n \langle f_j \rangle_{R_{n+1}} \right] \langle f_{n+1}, h_{R_{n+1}} \rangle.
$$

Consider the rectangles $K, R_1, \ldots, R_{n+1}$ as fixed for the moment. There holds that $\langle f_j \rangle_K - \prod_{j=1}^n \langle f_j \rangle_{I^1_{n+1} \times K^2}$ equals

$$
-\sum_{i_1=0}^{k_1-1} \sum_{m_1=1}^n \langle h_{(I^1_{n+1})^{(i_1+1)}} \rangle_{I^1_{n+1}} \tag{3.15}
$$

$$
\prod_{j=1}^{m_1-1} \langle f_j \rangle_{(I^1_{n+1})^{(i_1+1)} \times K^2} \left\langle f_{m_1}, h_{(I^1_{n+1})^{(i_1+1)}} \otimes \frac{1_{K^2}}{|K^2|} \right\rangle \prod_{j=m_1+1}^n \langle f_j \rangle_{(I^1_{n+1})^{(i_1)} \times K^2}.
$$

Similarly, we have that $-\prod_{j=1}^n \langle f_j \rangle_{K^1 \times I^2_{n+1}} + \prod_{j=1}^n \langle f_j \rangle_{R_{n+1}}$ equals

$$
\sum_{i_1=0}^{k_1-1} \sum_{m_1=1}^n \langle h_{(I^1_{n+1})^{(i_1+1)}} \rangle_{I^1_{n+1}}
$$

$$
\prod_{j=1}^{m_1-1} \langle f_j \rangle_{(I^1_{n+1})^{(i_1+1)} \times I^2_{n+1}} \left\langle f_{m_1}, h_{(I^1_{n+1})^{(i_1+1)}} \otimes \frac{1_{I^2_{n+1}}}{|I^2_{n+1}|} \right\rangle \prod_{j=m_1+1}^n \langle f_j \rangle_{(I^1_{n+1})^{(i_1)} \times I^2_{n+1}}. \tag{3.16}
$$

Let $g_j^{m_1,i_1} = \langle f_j \rangle^1_{(I^1_{n+1})^{(i_1+1)}}$ for $j \in \{1,\ldots,m_1-1\}$, $g_{m_1}^{m_1,i_1} = \langle f_{m_1}, h_{(I^1_{n+1})^{(i_1+1)}} \rangle_1$ and $g_j^{m_1,i_1} = \langle f_j \rangle^1_{(I^1_{n+1})^{(i_1)}}$ for $j \in \{m_1+1,\ldots,n\}$. The sum of (3.15) and (3.16) can similarly be split as

$$
\sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \sum_{m_1,m_2=1}^n \langle h_{R_{n+1}^{(i_1+1,i_2+1)}} \rangle_{R_{n+1}} \tag{3.17}
$$

$$
\prod_{j=1}^{m_2-1} \langle g_j^{m_1,i_1} \rangle_{(I^2_{n+1})^{(i_2+1)}} \langle g_{m_2}^{m_1,i_1}, h_{(I^2_{n+1})^{(i_2+1)}} \rangle \prod_{j=m_2+1}^n \langle g_j^{m_1,i_1} \rangle_{(I^2_{n+1})^{(i_2)}}.
$$

When one recalls the definition of the functions $g_j^{m_1, i_1}$ and writes this in terms of the functions f_j, one has that in the first parameter f_j is paired with $1_{(I_{n+1}^1)^{(i_1+1)}}/|(I_{n+1}^1)^{(i_1+1)}|$ for $j = 1, \ldots, m_1 - 1$, f_{m_1} with $h_{(I_{n+1}^1)^{(i_1+1)}}$ and f_j with $1_{(I_{n+1}^1)^{(i_1)}}/|(I_{n+1}^1)^{(i_1)}|$ for $j = m_1 + 1, \ldots, n$. Each f_j is paired similarly in the second parameter. In the case $m_1 = m_2 =: m$ the summand in (3.17) can be written as

$$\langle h_{R_{n+1}^{(i_1+1, i_2+1)}} \rangle_{R_{n+1}} \prod_{j=1}^{m-1} \langle f_j \rangle_{R_{n+1}^{(i_1+1, i_2+1)}} \langle f_{m_1}, h_{R_{n+1}^{(i_1+1, i_2+1)}} \rangle \prod_{j=m+1}^{n} \langle f_j \rangle_{R_{n+1}^{(i_1, i_2)}}. \quad (3.18)$$

The splitting in (3.17) gives us the identity

$$\Sigma_{n+1, n+1}^1 - \Sigma_{n+1}^2 - \Sigma_{n+1}^3 + \Sigma^4 =: \sum_{i_1=0}^{k_1-1} \sum_{i_2=0}^{k_2-1} \sum_{m_1, m_2=1}^{n} \Sigma_{m_1, m_2, i_1, i_2}^{1,2,3,4}.$$

We fix some i_1 and i_2 and consider the case $m_1 = m_2 =: m$. From (3.18) we see that

$$\Sigma_{m, m, i_1, i_2}^{1,2,3,4} = \sum_{K} \sum_{L^{(k_1-i_1, k_2-i_2)}=K} \sum_{R_{n+1}^{(i_1, i_2)}=L} c_{K, L, R_{n+1}}$$
$$\prod_{j=1}^{m-1} \langle f_j \rangle_{L^{(1,1)}} \langle f_{m_1}, h_{L^{(1,1)}} \rangle \prod_{j=m+1}^{n} \langle f_j \rangle_L \langle f_{n+1}, h_{R_{n+1}} \rangle,$$

where

$$c_{K, L, R_{n+1}} = \sum_{\substack{R_1, \ldots, R_n \\ R_j^{(k)}=K}} b_{K, (R_j)} \langle h_{L^{(1,1)}} \rangle_{R_{n+1}}.$$

The coefficient satisfies the estimate

$$|c_{K, L, R_{n+1}}| \leq \frac{|R_{n+1}|^{1/2}}{|L^{(1,1)}|^{1/2}} = \frac{|L^{(1,1)}|^{n/2} |R_{n+1}|^{1/2}}{|L^{(1,1)}|^n} |L^{(1,1)}|^{(n-1)/2}.$$

Thus, we see that $C^{-1} \Sigma_{m, m, i_1, i_2}^{1,2,3,4}$ is a standard n-linear bi-parameter shift. The complexity of the shift is $((0, 0), \ldots, (0, 0), (1, 1), \ldots, (1, 1), (i_1 + 1, i_2 + 1))$ with m zeros. The case of general m_1 and m_2 is analogous. □

Bi-parameter Representation Theorem We set

$$\sigma = (\sigma_1, \sigma_2) \in (\{0, 1\}^{d_1})^{\mathbb{Z}} \times (\{0, 1\}^{d_2})^{\mathbb{Z}}, \qquad \sigma_i = (\sigma_i^k)_{k \in \mathbb{Z}},$$

and denote the expectation over the product probability space by

$$\mathbb{E}_\sigma = \mathbb{E}_{\sigma_1} \mathbb{E}_{\sigma_2} = \mathbb{E}_{\sigma_2} \mathbb{E}_{\sigma_1} = \iint d\mathbb{P}_{\sigma_1} \, d\mathbb{P}_{\sigma_2}.$$

We also set $\mathcal{D}_0 = \mathcal{D}_0^1 \times \mathcal{D}_0^2$, where $\mathcal{D}_0^i$ is the standard dyadic grid of $\mathbb{R}^{d_i}$. We use the notation

$$I_i + \sigma_i := I_i + \sum_{k:\, 2^{-k} < \ell(I_i)} 2^{-k}\sigma_i^k, \qquad I_i \in \mathcal{D}_0^i.$$

Given $\sigma = (\sigma_1, \sigma_2)$ and $R = I_1 \times I_2 \in \mathcal{D}_0$ we set

$$R + \sigma = (I_1 + \sigma_1) \times (I_2 + \sigma_2) \qquad \text{and} \qquad \mathcal{D}_\sigma = \{R + \sigma:\ R \in \mathcal{D}_0\} = \mathcal{D}_{\sigma_1} \times \mathcal{D}_{\sigma_2}.$$

Theorem 3.19 *Suppose that T is an n-linear bi-parameter (ω_1, ω_2)-CZO, where $\omega_i \in$ Dini$_{1/2}$. Then we have*

$$\langle T(f_1, \ldots, f_n), f_{n+1} \rangle$$
$$= C\mathbb{E}_\sigma \sum_{k=(k_1,k_2)\in\mathbb{N}^2} \sum_{u=0}^{c_{d,n}} \omega_1(2^{-k_1})\omega_2(2^{-k_2})\langle V_{k,u,\sigma}(f_1, \ldots, f_n), f_{n+1} \rangle,$$

where

$$V_{k,u,\sigma} \in \{Q_k,\, S_{((k_1,k_2),\ldots,(k_1,k_2))},\, (QS)_{k_1,(k_2,\ldots,k_2)},\, (SQ)_{(k_1,\ldots,k_1),k_2},$$
$$(Q\pi)_{k_1},\, (\pi Q)_{k_2},\, (S\pi)_{(k_1,\ldots k_1)},\, (\pi S)_{(k_2,\ldots,k_2)},\, \Pi\}$$

defined in $\mathcal{D}_\sigma$, and if the operator does not depend on k_1 or k_2 then that particular $k_i = 0$.

Proof We decompose

$$\langle T(f_1, \ldots, f_n), f_{n+1} \rangle$$
$$= \mathbb{E}_\sigma \sum_{R_1,\ldots,R_{n+1}} \langle T(\Delta_{R_1} f_1, \ldots, \Delta_{R_n} f_n), \Delta_{R_{n+1}} f_{n+1} \rangle$$
$$= \sum_{j_1,j_2=1}^{n+1} \mathbb{E}_\sigma \sum_{\substack{R_1,\ldots,R_{n+1} \\ \ell(I_{i_1}^1) > \ell(I_{j_1}^1) \text{ for } i_1 \neq j_1 \\ \ell(I_{i_2}^2) > \ell(I_{j_2}^2) \text{ for } i_2 \neq j_2}} \langle T(\Delta_{R_1} f_1, \ldots, \Delta_{R_n} f_n), \Delta_{R_{n+1}} f_{n+1} \rangle$$
$$+ \mathbb{E}_\sigma \operatorname{Rem}_\sigma,$$

where $R_1, \ldots, R_{n+1} \in \mathcal{D}_\sigma = \mathcal{D}_{\sigma_1} \times \mathcal{D}_{\sigma_2}$ for some $\sigma = (\sigma_1, \sigma_2)$ and $R_j = I_j^1 \times I_j^2$.

The Main Terms For j_1, j_2 we let

$$\Sigma_{j_1, j_2, \sigma} = \sum_{\substack{R_1, \ldots, R_{n+1} \\ \ell(I^1_{i_1}) > \ell(I^1_{j_1}) \text{ for } i_1 \neq j_1 \\ \ell(I^2_{i_2}) > \ell(I^2_{j_2}) \text{ for } i_2 \neq j_2}} \langle T(\Delta_{R_1} f_1, \ldots, \Delta_{R_n} f_n), \Delta_{R_{n+1}} f_{n+1} \rangle.$$

These are symmetric and we choose to deal with $\Sigma_\sigma := \Sigma_{n, n+1, \sigma}$. After collapsing the relevant sums we have

$$\Sigma_\sigma = \sum_{\substack{R_1, \ldots, R_{n+1} \\ \ell(R_1) = \cdots = \ell(R_{n+1})}} \langle T(E_{R_1} f_1, \ldots, E_{R_{n-1}} f_{n-1}, \Delta^1_{I^1_n} E^2_{I^2_n} f_n), E^1_{I^1_{n+1}} \Delta^2_{I^2_{n+1}} f_{n+1} \rangle,$$

where $\ell(R_j) := (\ell(I^1_j), \ell(I^2_j))$ for $R_j = I^1_j \times I^2_j$.
   For $R = I^1 \times I^2$ we define

$$h_R = h_{I^1} \otimes h_{I^2}, \ h^0_R = h^0_{I^1} \otimes h^0_{I^2}, \ h^{1,0}_R = h_{I^1} \otimes h^0_{I^2} \text{ and } h^{0,1}_R = h^0_{I^1} \otimes h_{I^2}.$$

Using this notation we write

$$\langle T(E_{R_1} f_1, \ldots, E_{R_{n-1}} f_{n-1}, \Delta^1_{I^1_n} E^2_{I^2_n} f_n), E^1_{I^1_{n+1}} \Delta^2_{I^2_{n+1}} f_{n+1} \rangle$$
$$= \langle T(h^0_{R_1}, \ldots, h^0_{R_{n-1}}, h^{1,0}_{R_n}), h^{0,1}_{R_{n+1}} \rangle A^{n,n+1}_{R_1, \ldots, R_{n+1}} (f_1, \ldots, f_{n+1}),$$

where $A^{n,n+1}_{R_1, \ldots, R_{n+1}} (f_1, \ldots, f_{n+1}) = A^{n,n+1}_{R_1, \ldots, R_{n+1}}$ is defined in (3.9).
   We have

$$A^{n,n+1}_{R_1, \ldots, R_{n+1}} = A^{n,n+1}_{R_1, \ldots, R_{n+1}} - A^{n,n+1}_{I^1_n \times I^2_1, \ldots, I^1_n \times I^2_{n+1}} + A^{n,n+1}_{I^1_n \times I^2_1, \ldots, I^1_n \times I^2_{n+1}} \qquad (3.20)$$

and

$$A^{n,n+1}_{I^1_n \times I^2_1, \ldots, I^1_n \times I^2_{n+1}} = (A^{n,n+1}_{I^1_n \times I^2_1, \ldots, I^1_n \times I^2_{n+1}} - A^{n,n+1}_{I^1_n \times I^2_{n+1}, \ldots, I^1_n \times I^2_{n+1}}) + A^{n,n+1}_{I^1_n \times I^2_{n+1}, \ldots, I^1_n \times I^2_{n+1}}.$$
$$(3.21)$$

Then, we further have that the difference of the first two terms in the right hand side of (3.20) equals

$$\left[A^{n,n+1}_{R_1, \ldots, R_{n+1}} - A^{n,n+1}_{I^1_1 \times I^2_1, \ldots, I^1_n \times I^2_{n+1}} - A^{n,n+1}_{I^1_1 \times I^2_{n+1}, \ldots, I^1_{n+1} \times I^2_{n+1}} + A^{n,n+1}_{I^1_n \times I^2_{n+1}, \ldots, I^1_n \times I^2_{n+1}} \right]$$
$$+ \left\{ A^{n,n+1}_{I^1_1 \times I^2_{n+1}, \ldots, I^1_{n+1} \times I^2_{n+1}} - A^{n,n+1}_{I^1_n \times I^2_{n+1}, \ldots, I^1_n \times I^2_{n+1}} \right\}.$$
$$(3.22)$$

This gives us the decomposition

$$A^{n,n+1}_{R_1,\ldots,R_{n+1}} = [\,\cdot\,] + \{\,\cdot\,\} + (\,\cdot\,) + A^{n,n+1}_{I_n^1 \times I_{n+1}^2,\ldots,I_n^1 \times I_{n+1}^2}, \tag{3.23}$$

where inside the brackets we have the corresponding term as in (3.21) and (3.22).

The identity (3.23) splits Σ_σ into four terms $\Sigma_\sigma = \Sigma_\sigma^1 + \Sigma_\sigma^2 + \Sigma_\sigma^3 + \Sigma_\sigma^4$.

The Shift Case Σ_σ^1 We begin by looking at Σ_σ^1, that is, the term coming from $[\,\cdot\,]$ in (3.23). Let us further define the abbreviation

$$\varphi_{R_1,\ldots,R_{n+1}} := \langle T(h_{R_1}^0, \ldots, h_{R_{n-1}}^0, h_{R_n}^{1,0}), h_{R_{n+1}}^{0,1} \rangle$$

$$\times \left[A^{n,n+1}_{R_1,\ldots,R_{n+1}} - A^{n,n+1}_{I_n^1 \times I_1^2,\ldots,I_n^1 \times I_{n+1}^2} - A^{n,n+1}_{I_1^1 \times I_{n+1}^2,\ldots,I_{n+1}^1 \times I_{n+1}^2} + A^{n,n+1}_{I_n^1 \times I_{n+1}^2,\ldots,I_n^1 \times I_{n+1}^2} \right] \tag{3.24}$$

so that

$$\Sigma_\sigma^1 = \sum_{\substack{R_1,\ldots,R_{n+1} \\ \ell(R_1)=\cdots=\ell(R_{n+1})}} \varphi_{R_1,\ldots,R_{n+1}}.$$

If $R = I^1 \times I^2$ is a rectangle and $m = (m^1, m^2) \in \mathbb{Z}^{d_1} \times \mathbb{Z}^{d_2}$, then we define $I^i \dot{+} m^i := I^i + m^i \ell(I^i)$ and $R \dot{+} m := (I^1 \dot{+} m^1) \times (I^2 \dot{+} m^2)$. Notice that if $I_i^1 = I_j^1$ for all i, j or $I_i^2 = I_j^2$ for all i, j then $\varphi_{R_1,\ldots,R_{n+1}} = 0$. Thus, there holds that

$$\Sigma_\sigma^1 = \sum_{\substack{m_1,\ldots,m_{n+1}\in\mathbb{Z}^{d_1}\times\mathbb{Z}^{d_2} \\ (m_1^1,\ldots,m_{n+1}^1)\neq 0,\ m_n^1=0 \\ (m_1^2,\ldots,m_{n+1}^2)\neq 0,\ m_{n+1}^2=0}} \sum_R \varphi_{R\dot{+}m_1,\ldots,R\dot{+}m_{n+1}}$$

$$= \sum_{k_1,k_2=2}^{\infty} \sum_{\substack{m_1,\ldots,m_{n+1}\in\mathbb{Z}^{d_1}\times\mathbb{Z}^{d_2} \\ \max|m_j^1|\in(2^{k_1-3},2^{k_1-2}],\ m_n^1=0 \\ \max|m_j^2|\in(2^{k_2-3},2^{k_2-2}],\ m_{n+1}^2=0}} \sum_R \varphi_{R\dot{+}m_1,\ldots,R\dot{+}m_{n+1}}.$$

As in [17] we say that $I \in \mathcal{D}_{\sigma_i}$ is k-good for $k \geq 2$—and denote this by $I \in \mathcal{D}_{\sigma_i,\text{good}}(k)$—if $I \in \mathcal{D}_{\sigma_i}$ satisfies

$$d(I, \partial I^{(k)}) \geq \frac{\ell(I^{(k)})}{4} = 2^{k-2}\ell(I). \tag{3.25}$$

Notice that for all $I \in \mathcal{D}_0^i$ we have

$$\mathbb{P}(\{\sigma_i : I + \sigma_i \in \mathcal{D}_{\sigma_i,\text{good}}(k)\}) = 2^{-d_i}.$$

Next, we consider $\mathbb{E}_\sigma \Sigma_\sigma^1$ and add goodness to the rectangles R. Recall that $\mathbb{E}_\sigma = \mathbb{E}_{\sigma_1} \mathbb{E}_{\sigma_2}$. We write $\mathcal{D}_{\sigma,\text{good}}(k_1, k_2) := \mathcal{D}_{\sigma_1,\text{good}}(k_1) \times \mathcal{D}_{\sigma_2,\text{good}}(k_2)$. There holds that

$$\mathbb{E}_\sigma \sum_{R \in \mathcal{D}_\sigma} \varphi_{R \dotplus m_1, \ldots, R \dotplus m_{n+1}} = 2^d \mathbb{E}_\sigma \sum_{R \in \mathcal{D}_{\sigma,\text{good}}(k_1,k_2)} \varphi_{R \dotplus m_1, \ldots, R \dotplus m_{n+1}}.$$

Therefore, we have shown that

$$\mathbb{E}_\sigma \Sigma_\sigma^1 = 2^d C \sum_{k_1,k_2=2}^{\infty} \omega_1(2^{-k_1}) \omega_2(2^{-k_2}) \langle Q_{k_1,k_2}(f_1, \ldots, f_n), f_{n+1} \rangle, \quad (3.26)$$

where

$$\langle Q_{k_1,k_2}(f_1, \ldots, f_n), f_{n+1} \rangle$$
$$:= \frac{1}{C\omega_1(2^{-k_1})\omega_2(2^{-k_2})} \sum_{\substack{m_1,\ldots,m_{n+1} \in \mathbb{Z}^{d_1} \times \mathbb{Z}^{d_2} \\ \max|m_j^1| \in (2^{k_1-3}, 2^{k_1-2}], \, m_n^1=0 \\ \max|m_j^2| \in (2^{k_2-3}, 2^{k_2-2}], \, m_{n+1}^2=0}} \sum_{R \in \mathcal{D}_{\sigma,\text{good}}(k_1,k_2)} \varphi_{R \dotplus m_1, \ldots, R \dotplus m_{n+1}}$$

and C is a large enough constant.

Let $m_1, \ldots, m_{n+1}$ and $R = I^1 \times I^2$ be as in the definition of Q_{k_1,k_2}. The goodness of the rectangle R easily implies (we omit the details, see [17]) that $(R \dotplus m_j)^{(k_1,k_2)} = R^{(k_1,k_2)} =: K$ for all $j \in \{1, \ldots, n+1\}$. Recall the definition of $\varphi_{R \dotplus m_1, \ldots, R \dotplus m_{n+1}}$ from (3.24). Therefore, to conclude that Q_{k_1,k_2} is a modified bi-parameter n-linear shift it remains to prove the normalization

$$|\langle T(h_{R \dotplus m_1}^0, \ldots, h_{R \dotplus m_{n-1}}^0, h_{R \dotplus m_n}^{1,0}), h_{R \dotplus m_{n+1}}^{0,1} \rangle| \lesssim \omega_1(2^{-k_1}) \omega_2(2^{-k_2}) \frac{|R|^{(n+1)/2}}{|K|^n}.$$
$$(3.27)$$

Let us first assume that $k_1 \sim 1 \sim k_2$. Since $m_i^1 \neq 0$ and $m_j^2 \neq 0$ for some i and j we may use the full kernel representation of T to have that the left hand side of (3.27) is less than

$$\int_{\mathbb{R}^{(n+1)d}} |K(x_{n+1}, x_1, \ldots, x_n)| \prod_{j=1}^{n+1} h_{R \dotplus m_j}^0(x_j) \, dx.$$

Applying the size of the kernel K this is further dominated by

$$\int_{\mathbb{R}^{(n+1)d_1}} \frac{1}{\left(\sum_{j=1}^n |x_{n+1}^1 - x_j^1|\right)^{nd_1}} \prod_{j=1}^{n+1} h_{I^1 \dotplus m_j^1}^0 (x_j^1) \, dx^1$$

$$\times \int_{\mathbb{R}^{(n+1)d_2}} \frac{1}{\left(\sum_{j=1}^n |x_{n+1}^2 - x_j^2|\right)^{nd_2}} \prod_{j=1}^{n+1} h_{I^2 \dotplus m_j^2}^0 (x_j^2) \, dx^2 \lesssim \frac{1}{|I^1|^{(n-1)/2} |I^2|^{(n-1)/2}}.$$

Notice that this is the right estimate, since $\omega_i(2^{-k_i}) \sim 1$ and $|K| = |R^{(k_1,k_2)}| \sim |R| = |I^1||I^2|$.

Suppose then that k_1 and k_2 are large enough so that we can use the continuity assumption of the full kernel K. Using the zero integrals of h_{I^1} and h_{I^2} there holds that the left hand side of (3.27) equals

$$\left| \int_{\mathbb{R}^{(n+1)d}} \left(K(x_{n+1}, x_1, \ldots, x_n) - K(x_{n+1}, x_1, \ldots, x_{n-1}, (c_{I^1}, x_n^2)) \right. \right.$$

$$- K((x_{n+1}^1, c_{I^2}), x_1, \ldots, x_n) + K((x_{n+1}^1, c_{I^2}), x_1, \ldots, x_{n-1}, (c_{I^1}, x_n^2)) \Big)$$

$$\times \prod_{j=1}^{n-1} h_{R \dotplus m_j}^0 (x_j) h_{R \dotplus m_n}^{1,0} (x_n) h_{R \dotplus m_{n+1}}^{0,1} (x_{n+1}) \, dx \bigg|, \tag{3.28}$$

where c_{I^i} denotes the center of the corresponding cube. Here one can use the continuity assumption of K which leads to a product of two one-parameter integrals which can be easily estimated.

What remains is the case that for example $k_1 \sim 1$ and k_2 is large. This is done similarly as the above two cases using the mixed size and continuity assumption of K. This concludes the proof of (3.27) and we are done dealing with $\mathbb{E}_\sigma \Sigma_\sigma^1$.

The Partial Paraproduct Cases Σ_σ^2 and Σ_σ^3 Next, we look at the symmetric terms $\mathbb{E}_\sigma \Sigma_\sigma^2$ and $\mathbb{E}_\sigma \Sigma_\sigma^3$. We explicitly consider $\mathbb{E}_\sigma \Sigma_\sigma^2$ here. Recall that Σ_σ^2 equals

$$\sum_{\substack{R_1, \ldots, R_{n+1} \\ \ell(R_1) = \cdots = \ell(R_{n+1})}} \langle T(h_{R_1}^0, \ldots, h_{R_{n-1}}^0, h_{R_n}^{1,0}), h_{R_{n+1}}^{0,1} \rangle$$

$$\left\{ A_{I_1^1 \times I_{n+1}^2, \ldots, I_{n+1}^1 \times I_{n+1}^2}^{n,n+1} - A_{I_n^1 \times I_{n+1}^2, \ldots, I_n^1 \times I_{n+1}^2}^{n,n+1} \right\}.$$

Since the difference $A_{I_1^1 \times I_{n+1}^2, \ldots, I_{n+1}^1 \times I_{n+1}^2}^{n,n+1} - A_{I_n^1 \times I_{n+1}^2, \ldots, I_n^1 \times I_{n+1}^2}^{n,n+1}$ depends only on the cube I_{n+1}^2 in the second parameter we can further rewrite this as

$$\Sigma_\sigma^2 = \sum_{\substack{I_1^1,\dots,I_{n+1}^1,I^2 \\ \ell(I_1^1)=\cdots=\ell(I_{n+1}^1)}} \langle T(h_{I_1^1}^0 \otimes 1, \dots, h_{I_{n-1}^1}^0 \otimes 1, h_{I_n} \otimes 1), h_{I_{n+1}^1 \times I^2}^{0,1} \rangle$$

$$\times \Bigg\{ \prod_{j=1}^{n-1} \Big\langle f_j, h_{I_j^1}^0 \otimes \frac{1_{I^2}}{|I^2|} \Big\rangle \cdot \Big\langle f_n, h_{I_n} \otimes \frac{1_{I^2}}{|I^2|} \Big\rangle \langle f_{n+1}, h_{I_{n+1}^1 \times I^2}^{0,1} \rangle \tag{3.29}$$

$$- \prod_{j=1}^{n-1} \Big\langle f_j, h_{I_n^1}^0 \otimes \frac{1_{I^2}}{|I^2|} \Big\rangle \cdot \Big\langle f_n, h_{I_n} \otimes \frac{1_{I^2}}{|I^2|} \Big\rangle \langle f_{n+1}, h_{I_n^1 \times I^2}^{0,1} \rangle \Bigg\}.$$

Let us write the summand in (3.29) as $\varphi_{I_1^1,\dots,I_{n+1}^1,I^2}$. By proceeding in the same way as above with $\mathbb{E}_\sigma \Sigma_\sigma^1$ we have that

$$\mathbb{E}_\sigma \Sigma_\sigma^2 = 2^{d_1} C \mathbb{E}_\sigma \sum_{k=2}^\infty \omega_1(2^{-k}) \langle (Q\pi)_k(f_1, \dots, f_n), f_{n+1} \rangle, \tag{3.30}$$

where

$$\langle (Q\pi)_k(f_1, \dots, f_n), f_{n+1} \rangle$$

$$:= \frac{1}{C\omega_1(2^{-k})} \sum_{\substack{m \in \mathbb{Z}^{(n+1)d_1} \\ \max |m_j| \in (2^{k-3}, 2^{k-2}] \\ m_n = 0}} \sum_{\substack{I^1 \in \mathcal{D}_{\sigma_1,\text{good}}(k) \\ I^2 \in \mathcal{D}_{\sigma_2}}} \varphi_{I^1 \dotplus m_1, \dots, I^1 \dotplus m_{n+1}, I^2}.$$

The k-goodness of I^1 implies that here $(I^1 \dotplus m_j)^{(k)} = (I^1)^{(k)} =: K^1$ for all j. Therefore, to conclude that $(Q\pi)_k$ is a modified partial paraproduct with the paraproduct component in $\mathbb{R}^{d_2}$ it remains to show that if we fix $m_1, \dots, m_{n+1}$ and I^1 as in the above sum then

$$\|(\langle T(h_{I^1 \dotplus m_1}^0 \otimes 1, \dots, h_{I^1 \dotplus m_{n-1}}^0 \otimes 1, h_{I^1} \otimes 1), h_{(I^1 \dotplus m_{n+1}) \times I^2}^{0,1}\rangle)_{I^2 \in \mathcal{D}_{\sigma_2}}\|_{\text{BMO}}$$

$$\lesssim \omega_1(2^{-k}) \frac{|I^1|^{(n+1)/2}}{|K^1|^n}. \tag{3.31}$$

We verify the above BMO condition by taking a cube I^2 and a function a_{I^2} such that $a_{I^2} = a_{I^2} 1_{I^2}$, $|a_{I^2}| \leq 1$ and $\int a_{I^2} = 0$, and showing that

$$|\langle T(h_{I^1 \dotplus m_1}^0 \otimes 1, \dots, h_{I^1 \dotplus m_{n-1}}^0 \otimes 1, h_{I^1} \otimes 1), h_{(I^1 \dotplus m_{n+1})}^0 \otimes a_{I^2} \rangle|$$

$$\lesssim \omega_1(2^{-k}) \frac{|I^1|^{(n+1)/2}}{|K^1|^n} |I^2|. \tag{3.32}$$

For a suitably large constant C (so that we can use the continuity assumption of the kernel below) we split the pairing as

$$\langle T(h^0_{I^1 \dotplus m_1} \otimes 1_{(CI^2)^c}, h^0_{I^1 \dotplus m_2} \otimes 1, \dots, h^0_{I^1 \dotplus m_{n-1}} \otimes 1, h_{I^1} \otimes 1), h^0_{(I^1 \dotplus m_{n+1})} \otimes a_{I^2} \rangle$$
$$+ \langle T(h^0_{I^1 \dotplus m_1} \otimes 1_{CI^2}, h^0_{I^1 \dotplus m_2} \otimes 1, \dots, h^0_{I^1 \dotplus m_{n-1}} \otimes 1, h_{I^1} \otimes 1), h^0_{(I^1 \dotplus m_{n+1})} \otimes a_{I^2} \rangle. \tag{3.33}$$

Let us show that the first term in (3.33) is dominated by $\omega_1(2^{-k})|I^1|^{(n+1)/2}|I^2|/|K^1|^n$. We have two cases. The case that $k \sim 1$ is handled with the mixed size and continuity assumption of K. The case that k is large is handled with the continuity assumption of K. We show the details for the case $k \sim 1$. The other case is done similarly (see also the paragraph containing (3.28)).

We assume that $k \sim 1$. Since a_{I^2} has zero integral the pairing that we are estimating equals (by definition)

$$\int_{\mathbb{R}^{(n+1)d}} \left(K(x_{n+1}, x_1, \dots, x_n) - K((x^1_{n+1}, c_{I^2}), x_1, \dots, x_n) \right)$$
$$\times \prod_{j=1}^{n-1} h^0_{I^1 \dotplus m_j}(x^1_j) h_{I^1}(x^1_n) h^0_{(I^1 \dotplus m_{n+1})}(x^1_{n+1}) 1_{(CI^2)^c}(x^2_1) a_{I^2}(x^2_{n+1}) \, dx.$$

The mixed size and continuity property of K implies that the absolute value of the last integral is dominated by

$$\int_{\mathbb{R}^{(n+1)d_1}} \frac{1}{\left(\sum_{j=1}^n |x^1_{n+1} - x^1_j| \right)^{nd_1}} \prod_{j=1}^{n+1} h^0_{I^1 \dotplus m_j}(x^1_j) \, dx^1$$
$$\times \int_{\mathbb{R}^{(n+1)d_2}} \omega_2 \left(\frac{|x^2_{n+1} - c_{I^2}|}{\sum_{j=1}^n |c_{I^2} - x^2_j|} \right) \frac{1}{\left(\sum_{j=1}^n |c_{I^2} - x^2_j| \right)^{nd_2}} 1_{(CI^2)^c}(x^2_1) 1_{I^2}(x^2_{n+1}) \, dx^2.$$

The integral related to $\mathbb{R}^{d_1}$ is dominated by $|I^1|^{-(n-1)/2}$.

Consider the integral related to $\mathbb{R}^{d_2}$. By first estimating that

$$\omega_2 \left(\frac{|x^2_{n+1} - c_{I^2}|}{\sum_{j=1}^n |c_{I^2} - x^2_j|} \right) \leq \omega_2 \left(\frac{|x^2_{n+1} - c_{I^2}|}{|c_{I^2} - x^2_1|} \right)$$

with some work we see that the integral over $\mathbb{R}^{(n+1)d_2}$ is dominated by

$$\int_{I^2} \int_{(CI^2)^c} \omega_2 \left(\frac{|x^2_{n+1} - c_{I^2}|}{|c_{I^2} - x^2_1|} \right) \frac{1}{|c_{I^2} - x^2_1|^{d_2}} \, dx^2_1 \, dx^2_{n+1}$$
$$\lesssim |I^2| \int_{(CI^2)^c} \omega_2 \left(\frac{\ell(I^2)}{|c_{I^2} - x^2_1|} \right) \frac{1}{|c_{I^2} - x^2_1|^{d_2}} \, dx^2_1 \lesssim |I^2| \sum_{k=0}^{\infty} \omega_2(2^{-k}) \lesssim |I^2|.$$

In conclusion, we showed that the first term in (3.33) is dominated by $|I^1|^{-(n-1)/2}|I^2|$, which is the right estimate in the case $k \sim 1$.

We turn to consider the second term in (3.33). We again split it into two by writing $1 = 1_{(CI^2)^c} + 1_{CI^2}$ in the second slot. The part with $1_{(CI^2)^c}$ is estimated in the same way as above and then one continues with the part related to 1_{CI^2}. This is repeated until we are only left with the term

$$\langle T(h^0_{I^1 \dotplus m_1} \otimes 1_{CI^2}, \dots, h^0_{I^1 \dotplus m_{n-1}} \otimes 1_{CI^2}, h_{I^1} \otimes 1_{CI^2}), h^0_{(I^1 \dotplus m_{n+1})} \otimes a_{I^2}\rangle. \tag{3.34}$$

The estimate for this uses the partial kernel representations of T. Again, we have the two cases that either $k \sim 1$ or k is large. These are handled in the same way using either the size or the continuity of the partial kernels. We consider explicitly the case that k is large. Using the zero integral of h_{I^1} we have that the above pairing equals

$$\int_{\mathbb{R}^{(n+1)d_1}} \left(K_{1_{CI^2}, \dots, 1_{CI^2}, a_{I^2}}(x^1_{n+1}, x^1_1, \dots, x^1_n) - K_{1_{CI^2}, \dots, 1_{CI^2}, a_{I^2}}(x^1_{n+1}, x^1_1, \dots, c_{I^1}) \right)$$

$$\times \prod_{j=1}^{n-1} h^0_{I^1 \dotplus m_j}(x^1_j) h_{I^1}(x_n) h^0_{I^1 \dotplus m_{n+1}}(x^1_{n+1}) \, dx^1.$$

Taking absolute values and using the continuity of the partial kernel leads to

$$C(1_{CI^2}, \dots, 1_{CI^2}, a_{I^2}) \int_{\mathbb{R}^{(n+1)d_1}} \omega_1 \left(\frac{|x^1_n - c_{I^1}|}{\sum_{j=1}^n |x^1_{n+1} - x^1_j|} \right)$$

$$\times \frac{1}{\left(\sum_{j=1}^n |x^1_{n+1} - x^1_j| \right)^{nd_1}} \prod_{j=1}^{n+1} h^0_{I^1 \dotplus m_j}(x^1_j) \, dx^1.$$

By assumption there holds that $C(1_{CI^2}, \dots, 1_{CI^2}, a_{I^2}) \lesssim |I^2|$ and the integral is dominated by $\omega_1(2^{-k})|I^1|^{(n+1)/2}|K^1|^n$. This concludes the proof of (3.32) and also finishes our treatment of $\mathbb{E}_\sigma \Sigma^2_\sigma$.

The Full Paraproduct Σ^4_σ Recall that

$$\Sigma^4_\sigma = \sum_{\substack{R_1, \dots, R_{n+1} \\ \ell(R_1) = \dots = \ell(R_{n+1})}} \langle T(1_{R_1}, \dots, 1_{R_{n-1}}, h_{I^1_n} \otimes 1_{I^2_n}), 1_{I^1_{n+1}} \otimes h_{I^2_{n+1}} \rangle \prod_{j=1}^{n-1} \langle f_j \rangle_{I^1_n \times I^2_{n+1}}$$

$$\times \left\langle f_n, h_{I^1_n} \otimes \frac{1_{I^2_{n+1}}}{|I^2_{n+1}|} \right\rangle \left\langle f_{n+1}, \frac{1_{I^1_n}}{|I^1_n|} \otimes h_{I^2_{n+1}} \right\rangle,$$

which equals

$$\sum_{R = K^1 \times K^2} \langle T(1, \dots, 1, h_{K^1} \otimes 1), 1 \otimes h_{K^2} \rangle$$

$$\prod_{j=1}^{n-1} \langle f_j \rangle_R \Big\langle f_n, h_{K^1} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \Big\langle f_{n+1}, \frac{1_{K^1}}{|K^1|} \otimes h_{K^2} \Big\rangle.$$

This is directly a full paraproduct as

$$\langle T(1, \ldots, 1, h_{K^1} \otimes 1), 1 \otimes h_{K^2} \rangle = \langle T_1^{n*}(1, \ldots, 1), h_R \rangle,$$

and so we are done with this term. Therefore, we are done with the main terms, and no more full paraproducts will appear.

The Remainder Rem_σ To finish the proof of the bi-parameter representation theorem it remains to discuss the remainder term Rem_σ. Some of the weak boundedness type assumptions are used here—but there is nothing surprising on how they are used and we do not focus on that. We only explain the structural idea.

An $(n + 1)$-tuple $(I_1^i, \ldots, I_{n+1}^i)$ of cubes $I_j^i \in \mathcal{D}_{\sigma_i}$ belongs to $\mathcal{I}_{\sigma_i}$ if the following holds: if j is an index such that $\ell(I_j^i) \leq \ell(I_k^i)$ for all k, then there exists at least one index $k_0 \neq j$ so that $\ell(I_j^i) = \ell(I_{k_0}^i)$. The remainder term can be written as

$$\mathrm{Rem}_\sigma = \sum_{j_1=1}^{n+1} \sum_{\substack{I_1^1, \ldots, I_{n+1}^1 \\ \ell(I_i^1) > \ell(I_{j_1}^1) \text{ for } i \neq j_1}} \sum_{(I_1^2, \ldots, I_{n+1}^2) \in \mathcal{I}_{\sigma_2}} \langle T(\Delta_{R_1} f_1, \ldots, \Delta_{R_n} f_n), \Delta_{R_{n+1}} f_{n+1} \rangle$$

$$+ \sum_{j_2=1}^{n+1} \sum_{\substack{I_1^2, \ldots, I_{n+1}^2 \\ \ell(I_i^2) > \ell(I_{j_2}^2) \text{ for } i \neq j_2}} \sum_{(I_1^1, \ldots, I_{n+1}^1) \in \mathcal{I}_{\sigma_1}} \langle T(\Delta_{R_1} f_1, \ldots, \Delta_{R_n} f_n), \Delta_{R_{n+1}} f_{n+1} \rangle$$

$$+ \sum_{\substack{(I_1^1, \ldots, I_{n+1}^1) \in \mathcal{I}_{\sigma_1} \\ (I_1^2, \ldots, I_{n+1}^2) \in \mathcal{I}_{\sigma_2}}} \langle T(\Delta_{R_1} f_1, \ldots, \Delta_{R_n} f_n), \Delta_{R_{n+1}} f_{n+1} \rangle,$$

where as usual $R_i = I_i^1 \times I_i^2$. Let us write this as

$$\mathrm{Rem}_\sigma = \sum_{j_1=1}^{n+1} \mathrm{Rem}_{\sigma, j_1}^1 + \sum_{j_2=1}^{n+1} \mathrm{Rem}_{\sigma, j_2}^2 + \mathrm{Rem}_\sigma^3 .$$

First, we look at the terms $\mathrm{Rem}_{\sigma, j_1}^1$ and $\mathrm{Rem}_{\sigma, j_2}^2$ which are analogous. Consider for example $\mathrm{Rem}_{\sigma, n+1}^1$. We further divide $\mathcal{I}_{\sigma_2}$ into subcollections by specifying the slots where the smallest cubes are. For example, we consider here the part of the sum with the tuples $(I_1^2, \ldots, I_{n+1}^2)$ such that $\ell(I_i^2) > \ell(I_n^2) = \ell(I_{n+1}^2)$ for all $i = 1, \ldots, n - 1$. By collapsing the relevant sums of martingale differences the term we are dealing with

can be written as

$$\sum_{\substack{R_1,\ldots,R_{n+1} \\ \ell(R_i)=\ell(R_j)}} \langle T(E_{R_1}f_1,\ldots,E_{R_{n-1}}f_{n-1}, E^1_{I^1_n}\Delta^2_{I^2_n}f_n), \Delta_{R_{n+1}}f_{n+1}\rangle. \qquad (3.35)$$

In the first parameter there is only one martingale difference and in the second parameter there are two (in the general case at least two). Thus, the strategy is that we will write this in terms of model operators that have a modified shift or a paraproduct structure in the first parameter and a standard shift structure in the second parameter. We omit the details.

Finally, we consider Rem^3_σ. This is also divided into several cases by specifying the places of the smallest cubes in both parameters. For example, for notational convenience we take the part where $\ell(I^1_1) = \ell(I^1_{n+1}) < \ell(I^1_i)$ and $\ell(I^2_1) = \ell(I^2_{n+1}) < \ell(I^2_i)$ for all $i = 2, \ldots, n$. Notice that in general the places and the number of the smallest cubes do not need to be the same in both parameters. After collapsing the relevant sums of martingale differences the term we are looking at is

$$\mathbb{E}_\sigma \sum_{\substack{R_1,\ldots,R_{n+1} \\ \ell(R_i)=\ell(R_j)}} \langle T(\Delta_{R_1}f_1, E_{R_2}f_2,\ldots, E_{R_n}f_n), \Delta_{R_{n+1}}f_{n+1}\rangle. \qquad (3.36)$$

Here we have two (in the general case at least two) martingale differences in each parameter so this will be written in terms of standard bi-parameter n-linear shifts. We omit the details. This completes the proof. $\qquad\square$

Corollaries We indicate some corollaries—we start with the most basic unweighted boundedness on the Banach range of exponents.

Proposition 3.37 *Let $p_j \in (1,\infty)$, $j = 1,\ldots,n+1$, be such that $\sum_{j=1}^{n+1} 1/p_j = 1$. Suppose that Q_k is a modified n-linear bi-parameter shift. Then the estimate*

$$|\langle Q_k(f_1,\ldots,f_n), f_{n+1}\rangle| \lesssim \sqrt{k_1}\sqrt{k_2} \prod_{j=1}^{n+1} \|f_j\|_{L^{p_j}}$$

holds.

Suppose that $(QS)_{k,i}$ is a modified/standard shift (here $k \in \{1, 2, \ldots\}$ and $i = (i_1,\ldots,i_{n+1})$). Then the estimate

$$|\langle (QS)_{k,i}(f_1,\ldots,f_n), f_{n+1}\rangle| \lesssim \sqrt{k} \prod_{j=1}^{n+1} \|f_j\|_{L^{p_j}}$$

holds.

Proof We only prove the statement for the operator Q_k. This essentially contains the proof for $(QS)_{k,i}$.

We assume Q_k has the explicit form

$$\langle Q_k(f_1, \ldots, f_n), f_{n+1}\rangle = \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k)}=K}} a_{K,(R_j)} \left[\prod_{j=1}^n \langle f_j, h_{R_j}^0 \rangle - \prod_{j=1}^n \langle f_j, h_{I_{n+1}^1 \times I_j^2}^0 \rangle \right.$$

$$\left. - \prod_{j=1}^n \langle f_j, h_{I_j^1 \times I_{n+1}^2}^0 \rangle + \prod_{j=1}^n \langle f_j, h_{R_{n+1}}^0 \rangle \right] \langle f_{n+1}, h_{R_{n+1}} \rangle.$$

Using the notation (3.13) there holds that

$$\prod_{j=1}^n \langle f_j, h_{R_j}^0 \rangle = \sum_{m_1, m_2=1}^{n+1} \prod_{j=1}^n \langle D_{K^1, k_1}^1(j, m_1) D_{K^2, k_2}^2(j, m_2) f_j, h_{R_j}^0 \rangle.$$

We do the same decomposition with the other three terms inside the bracket $[\cdot]$. This splits $[\cdot]$ into a sum over $m_1, m_2 \in \{1, \ldots, n+1\}$. Then, we notice that all the terms in the sum with $m_1 = n+1$ or $m_2 = n+1$ cancel out. Thus, we get a splitting of $\langle Q_k(f_1, \ldots, f_n), f_{n+1}\rangle$ into a sum over $m_1, m_2 \in \{1, \ldots, n\}$. All the terms with different m_1 and m_2 are estimated separately.

In what follows—for notational convenience—we will focus on the case $m_1 = m_2 =: m \in \{1, \ldots, n\}$, and we define $D_{K^1, k_1}^1(j, m) D_{K^2, k_2}^2(j, m) =: D_{K, k}(j, m)$. The term in the splitting of $\langle Q_k(f_1, \ldots, f_n), f_{n+1}\rangle$ corresponding to $m = m_1 = m_2$ can be written as the sum

$$\sum_{i=1}^4 \langle U_i(f_1, \ldots, f_n), f_{n+1}\rangle,$$

where

$$\langle U_1(f_1, \ldots, f_n), f_{n+1}\rangle = \sum_K \sum_{\substack{R_1, \ldots, R_{n+1} \\ R_j^{(k)}=K}} a_{K,(R_j)} \prod_{j=1}^n \langle D_{K,k}(j, m) f_j, h_{R_j}^0 \rangle \langle f_{n+1}, h_{R_{n+1}} \rangle,$$

and U_2, U_3 and U_4 are defined similarly just by replacing $h_{R_j}^0$, $j \in \{1, \ldots, n\}$, by $h_{I_{n+1}^1 \times I_j^2}^0$, $h_{I_j^1 \times I_{n+1}^2}^0$ and $h_{R_{n+1}}^0$, respectively.

With some direct calculations it can be shown that for all $i \in \{1, \ldots, 4\}$ we have

$$|\langle U_i(f_1, \ldots, f_n), f_{n+1}\rangle| \leq \int \prod_{\substack{j=1 \\ j \neq m}}^n M f_j \left(\sum_K |M P_{K,(k_1-1, k_2-1)} f_m|^2 \right)^{1/2} S_{\mathcal{D}} f_{n+1}.$$

$$(3.38)$$

From here the estimate can be finished by Hölder's inequality, the Fefferman–Stein inequality and square function estimates, see Lemma 2.4.   □

Next, we look at the modified partial paraproducts. We will use the well known one-parameter H^1-BMO duality estimate

$$\sum_I |a_I b_I| \lesssim \|(a_I)\|_{\mathrm{BMO}} \left\| \left(\sum_I |b_I|^2 \frac{1_I}{|I|} \right)^{1/2} \right\|_{L^1}, \tag{3.39}$$

where the cubes I are in some dyadic grid.

Proposition 3.40 *Let $p_j \in (1, \infty)$, $j = 1, \ldots, n+1$, be such that $\sum_{j=1}^{n+1} 1/p_j = 1$. Suppose $(Q\pi)_k$ is a modified n-linear partial paraproduct. Then the estimate*

$$|\langle (Q\pi)_k(f_1, \ldots, f_n), f_{n+1}\rangle| \lesssim \sqrt{k} \prod_{j=1}^{n+1} \|f_j\|_{L^{p_j}}$$

holds.

Proof We assume that $\langle (Q\pi)_k(f_1, \ldots, f_n), f_{n+1}\rangle$ has the form

$$\sum_K \sum_{\substack{I_1^1, \ldots, I_{n+1}^1 \\ (I_j^1)^{(k)}=K^1}} a_{K,(I_j^1)} \left[\prod_{j=1}^n \left\langle f_j, h_{I_j^1}^0 \otimes \frac{1_{K^2}}{|K^2|} \right\rangle - \prod_{j=1}^n \left\langle f_j, h_{I_{n+1}^1}^0 \otimes \frac{1_{K^2}}{|K^2|} \right\rangle \right] \langle f_{n+1}, h_{I_{n+1}^1 \times K^2} \rangle.$$

We decompose

$$\prod_{j=1}^n \left\langle f_j, h_{I_j^1}^0 \otimes \frac{1_{K^2}}{|K^2|} \right\rangle = \sum_{m=1}^{n+1} \prod_{j=1}^n \left\langle D_{K^1, k_1}^1 (j, m) f_j, h_{I_j^1}^0 \otimes \frac{1_{K^2}}{|K^2|} \right\rangle$$

and similarly with the other term inside the bracket $[\,\cdot\,]$. Notice that the terms with $m = n + 1$ cancel out. Thus, we get a decomposition of $\langle (Q\pi)_k(f_1, \ldots, f_n), f_{n+1}\rangle$ into a sum over $m \in \{1, \ldots, n\}$. The terms with different m are estimated separately.

Fix one m. The term from the decomposition of $\langle (Q\pi)_k(f_1, \ldots, f_n), f_{n+1}\rangle$ related to m is

$$\sum_{i=1}^2 \langle U_1(f_1, \ldots, f_n), f_{n+1}\rangle,$$

where $\langle U_1(f_1, \ldots, f_n), f_{n+1}\rangle$ equals

$$\sum_K \sum_{\substack{I_1^1, \ldots, I_{n+1}^1 \\ (I_j^1)^{(k)}=K^1}} a_{K,(I_j^1)} \prod_{j=1}^n \left\langle D_{K^1, k_1}^1 (j, m) f_j, h_{I_j^1}^0 \otimes \frac{1_{K^2}}{|K^2|} \right\rangle \langle f_{n+1}, h_{I_{n+1}^1 \times K^2} \rangle \tag{3.41}$$

and $\langle U_2(f_1, \ldots, f_n), f_{n+1}\rangle$ is defined similarly just be replacing $h^0_{I^1_j}$, $j = 1, \ldots, n$, with $h^0_{I^1_{n+1}}$.

We consider U_1 first. From the one-parameter H^1-BMO duality estimate (3.39) we have that, with fixed K^1 and $I^1_1, \ldots, I^1_{n+1}$, the sum over K^2 of the absolute value of the summand in (3.41) is dominated by

$$\frac{|I^1_{n+1}|^{(n+1)/2}}{|K^1|^n} \int_{\mathbb{R}^{d_2}} \left(\sum_{K^2} \Big| \prod_{j=1}^{n} \Big\langle D^1_{K^1,k_1}(j,m) f_j, h^0_{I^1_j} \otimes \frac{1_{K^2}}{|K^2|} \Big\rangle \langle f_{n+1}, h_{I^1_{n+1} \times K^2} \rangle \Big|^2 \frac{1_{K^2}}{|K^2|} \right)^{1/2}$$

$$\leq \frac{|I^1_{n+1}|^{(n+1)/2}}{|K^1|^n} \int_{\mathbb{R}^{d_2}} \prod_{j=1}^{n} \langle M^2 D^1_{K^1,k_1}(j,m) f_j, h^0_{I^1_j} \rangle_1 \langle S^2_{\mathcal{D}_2} \Delta_{K^1,k_1} f_{n+1}, h^0_{I^1_{n+1}} \rangle_1.$$

The sum of this over K^1 and $I^1_1, \ldots, I^1_{n+1}$ such that $(I^1_j)^{(k)} = K^1$ is less than

$$\int_{\mathbb{R}^d} \prod_{\substack{j=1 \\ j \neq m}}^{n} M^1 M^2 f_j \left(\sum_{K^1} (M^1 M^2 P^1_{K^1,k_1-1} f_m)^2 \right)^{1/2} \left(\sum_{K^1} (S^2_{\mathcal{D}^2} \Delta_{K^1,k_1} f_{n+1})^2 \right)^{1/2}.$$

$$(3.42)$$

Notice that the square function related to f_{n+1} is just the bi-parameter square function $S_{\mathcal{D}}$. To finish the estimate it remains to use the Fefferman–Stein inequality and square function estimates, see Lemma 2.4.

The second term $|\langle U_2(f_1, \ldots, f_n), f_{n+1}\rangle|$ satisfies the same upper bound (3.42), and can therefore be estimated in the same way. The proof is concluded. □

The above, together with known estimates for standard operators, directly leads to Banach range boundedness of n-linear bi-parameter (ω_1, ω_2)-CZOs with $\omega_i \in$ Dini$_{1/2}$. We do not push this further in this paper. For state-of-the-art estimates with genuinely multilinear weights (in the full multilinear range) see [31]. There we recorded some of the estimates with Dini$_1$ using the above representation theorem and the decomposition of modified operators in terms of standard operators.

We are unable to perform the estimates of [31] with the regularity Dini$_{\frac{1}{2}}$. However, the linear case is special: the weighted estimates of linear modified model operators with a bound depending on the square root of the complexity are easy. Notice that in principle we have already done all the necessary work. For example, if we want to estimate $\|Q_k f\|_{L^p(w)}$, we study the unweighted pairings $\langle Q_k f, g \rangle$. Then, we proceed as in the linear case of Proposition 3.37. Depending on the form of the shift this leads us to terms corresponding to (3.38) such as

$$\int \left(\sum_K |M P_{K,(k_1-1,k_2-1)} f|^2 \right)^{1/2} S_{\mathcal{D}} g.$$

By Hölder's inequality this is less than

$$\left\| \left(\sum_K |M P_{K,(k_1-1,k_2-1)} f|^2 \right)^{1/2} \right\|_{L^p(w)} \|S_{\mathcal{D}} g\|_{L^{p'}(w^{1-p'})}$$
$$\lesssim \sqrt{k_1} \sqrt{k_2} \|f\|_{L^p(w)} \|g\|_{L^{p'}(w^{1-p'})}.$$

Proposition 3.43 *For every $p \in (1, \infty)$ and bi-parameter A_p weight w we have*

$$\|Q_k f\|_{L^p(w)} \lesssim \sqrt{k_1} \sqrt{k_2} \|f\|_{L^p(w)}.$$

For completeness, we record the corresponding result for CZOs. Again, for multilinear weighted estimates with the optimal weight classes see [31].

Corollary 3.44 *Let $p_j \in (1, \infty)$, $j = 1, \ldots, n+1$, be such that $\sum_{j=1}^{n+1} 1/p_j = 1$. Suppose that T is an n-linear bi-parameter (ω_1, ω_2)-CZO, where $\omega_i \in \mathrm{Dini}_{1/2}$. Then we have the Banach range estimate*

$$|\langle T(f_1, \ldots, f_n), f_{n+1} \rangle| \lesssim \prod_{j=1}^{n+1} \|f_j\|_{L^{p_j}}. \tag{3.45}$$

In the linear case $n = 1$ we have the weighted estimate

$$\|Tf\|_{L^p(w)} \lesssim \|f\|_{L^p(w)} \tag{3.46}$$

whenever $p \in (1, \infty)$ and $w \in A_p$ is a bi-parameter weight.

4 Commutator Estimates

The basic form of a commutator is $[b, T] \colon f \mapsto bTf - T(bf)$. We are interested in various iterated versions in the multi-parameter setting and with mild kernel regularity.

For a bi-parameter weight $w \in A_2(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ and a locally integrable function b we define the weighted product BMO norm

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod}}(w)} = \sup_{\mathcal{D}} \sup_{\Omega} \left(\frac{1}{w(\Omega)} \sum_{\substack{R \in \mathcal{D} \\ R \subset \Omega}} \frac{|\langle b, h_R \rangle|^2}{\langle w \rangle_R} \right)^{\frac{1}{2}}, \tag{4.1}$$

where the supremum is over all dyadic grids $\mathcal{D}^i$ on $\mathbb{R}^{d_i}$ and $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2$, and over all open sets $\Omega \subset \mathbb{R}^d := \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ for which $0 < w(\Omega) < \infty$. The following theorem, which is the two-weight Bloom version of [9], was proved in [29] with $\omega_i(t) = t^{\gamma_i}$.

Theorem 4.2 *Suppose that T_i is a one-parameter ω_i-CZO, where $\omega_i \in \text{Dini}_{3/2}$. Let $b\colon \mathbb{R}^d \to \mathbb{C}$, $p \in (1, \infty)$, $\mu, \lambda \in A_p(\mathbb{R}^d)$ be bi-parameter weights and $\nu = \mu^{1/p} \lambda^{-1/p} \in A_2(\mathbb{R}^d)$ be the associated bi-parameter Bloom weight. Then we have*

$$\|[T_1, [T_2, b]]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \|b\|_{\text{BMO}_{\text{prod}}(\nu)}.$$

Proof Let $\|b\|_{\text{BMO}_{\text{prod}}(\nu)} = 1$. We need to e.g. bound $\|[Q_{k_1}, [Q_{k_2}, b]]f\|_{L^p(\lambda)}$ for one-parameter modified shifts (which have a similar definition as in the bi-parameter case). It seems non-trivial to fully exploit the operators Q_k here and we content on splitting the operators to standard shifts and bounding

$$\sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \|[S_{k_1, j_1}, [S_{k_2, j_2}, b]]f\|_{L^p(\lambda)}$$

and other similar terms, where S_{k_i, j_i} is a linear one-parameter shift on $\mathbb{R}^{d_i}$ of complexity (k_i, j_i). Reaching Dini_1 would require replacing this step with a sharper estimate.

On page 11 of [29] it is recorded that

$$\|[S_{u_1, v_1}, [S_{u_2, v_2}, b]]f\|_{L^p(\lambda)} \lesssim (1 + \max(u_1, v_1))(1 + \max(u_2, v_2))\|f\|_{L^p(\mu)}.$$

Interestingly, this part of the argument can be improved: there actually holds that

$$\|[S_{u_1, v_1}, [S_{u_2, v_2}, b]]f\|_{L^p(\lambda)} \lesssim (1 + \max(u_1, v_1))^{1/2}(1 + \max(u_2, v_2))^{1/2}\|f\|_{L^p(\mu)}.$$
$$(4.3)$$

We will get back to this after completing the proof. Therefore, we have

$$\sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \|[S_{k_1, j_1}, [S_{k_2, j_2}, b]]f\|_{L^p(\lambda)} \lesssim (1 + k_1)^{3/2}(1 + k_2)^{3/2}\|f\|_{L^p(\mu)}.$$

Handling the other terms of the shift expansion of $[Q_{k_1}, [Q_{k_2}, b]]$ similarly, we get

$$\|[Q_{k_1}, [Q_{k_2}, b]]f\|_{L^p(\lambda)} \lesssim (1 + k_1)^{3/2}(1 + k_2)^{3/2}\|f\|_{L^p(\mu)}.$$

Controlling commutators like $[Q_{k_1}, [\pi, b]]$ similarly we get the claim.

We return to (4.3) now. Decompositions are very involved in the bi-commutator case, and we prefer to give the idea of the improvement (4.3) by studying the simpler one-parameter situation $[b, S_{i,j}]$, where

$$S_{i,j} = \sum_K \sum_{I^{(i)}=J^{(j)}=K} a_{IJK} \langle f, h_I \rangle h_J$$

is a one-parameter shift on $\mathbb{R}^d$ and $b \in \mathrm{BMO}(\nu)$;

$$\|b\|_{\mathrm{BMO}(\nu)} := \sup_{I \subset \mathbb{R}^d \text{ cube}} \frac{1}{\nu(I)} \int_I |b - \langle b \rangle_I| \sim \sup_{\mathcal{D}} \sup_{I_0 \in \mathcal{D}} \left(\frac{1}{\nu(I_0)} \sum_{\substack{I \in \mathcal{D} \\ I \subset I_0}} \frac{|\langle b, h_I \rangle|^2}{\langle \nu \rangle_I} \right)^{\frac{1}{2}} < \infty.$$

Here we only have use for the expression on the right-hand side, which is the analogue of the bi-parameter definition (4.1). However, it is customary to define things as on the left-hand side in this one-parameter situation. The equivalence follows from the weighted John–Nirenberg [34]

$$\sup_{I \subset \mathbb{R}^d \text{ cube}} \frac{1}{\nu(I)} \int_I |b - \langle b \rangle_I| \sim \sup_{I \subset \mathbb{R}^d \text{ cube}} \left(\frac{1}{\nu(I)} \int_I |b - \langle b \rangle_I|^2 \nu^{-1} \right)^{1/2}, \qquad \nu \in A_2.$$

Of course, one-parameter commutators $[b, T]$ can be handled even with Dini_0, but e.g. sparse domination proofs [25,26] are restricted to one-parameter, unlike these decompositions. To get started, we define the one-parameter paraproducts (with some implicit dyadic grid)

$$A_1(b, f) = \sum_I \Delta_I b \Delta_I f, \quad A_2(b, f) = \sum_I \Delta_I b E_I f \quad \text{and} \quad A_3(b, f) = \sum_I E_I b \Delta_I f.$$

By writing $b = \sum_I \Delta_I b$ and $f = \sum_J \Delta_J f$, and collapsing sums such as $1_I \sum_{J : I \subsetneq J} \Delta_J f = E_I f$, we formally have

$$bf = \sum_I \Delta_I b \Delta_I f + \sum_{I \subsetneq J} \Delta_I b \Delta_J f + \sum_{J \subsetneq I} \Delta_I b \Delta_J f = \sum_{k=1}^{3} A_k(b, f).$$

We now decompose the commutator as follows

$$[b, S_{i,j}] f = b S_{i,j} f - S_{i,j}(bf)$$
$$= \sum_{k=1}^{2} A_k(b, S_{i,j} f) - \sum_{k=1}^{2} S_{i,j}(A_k(b, f)) + [A_3(b, S_{i,j} f) - S_{i,j}(A_3(b, f))].$$

We have the well-known fact that $\|A_k(b, f)\|_{L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}(\nu)} \|f\|_{L^p(\mu)}$ for $k = 1, 2$—this can be seen by using the weighted H^1-BMO duality [37] (with $a_I = \langle b, h_I \rangle$)

$$\sum_I |a_I||b_I| \lesssim \|(a_I)\|_{\mathrm{BMO}(\nu)} \left\| \left(\sum_I |b_I|^2 \frac{1_I}{|I|} \right)^{1/2} \right\|_{L^1(\nu)}, \tag{4.4}$$

where

$$\|(a_I)\|_{\mathrm{BMO}(v)} = \sup_{I_0 \in \mathcal{D}} \left(\frac{1}{v(I_0)} \sum_{\substack{I \in \mathcal{D} \\ I \subset I_0}} \frac{|a_I|^2}{\langle v \rangle_I} \right)^{\frac{1}{2}}.$$

Combining this with the well-known estimate $\|S_{i,j}f\|_{L^p(w)} \lesssim \|f\|_{L^p(w)}$ for all $w \in A_p$ it follows that

$$\left\| \sum_{k=1}^{2} A_k(b, S_{i,j}f) - \sum_{k=1}^{2} S_{i,j}(A_k(b, f)) \right\|_{L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}(v)} \|f\|_{L^p(\mu)}.$$

The complexity dependence is coming from the remaining term

$$A_3(b, S_{i,j}f) - S_{i,j}(A_3(b, f)) = \sum_{K} \sum_{I^{(i)}=J^{(j)}=K} [\langle b \rangle_J - \langle b \rangle_I] a_{IJK} \langle f, h_I \rangle h_J.$$

There are many ways to bound this, but the following way based on the H^1-BMO duality—and executed in the particular way that we do below—gives the best dependence that we are aware of:

$$\|A_3(b, S_{i,j}f) - S_{i,j}(A_3(b, f))\|_{L^p(\lambda)} \lesssim (1 + \max(i, j))^{1/2} \|b\|_{\mathrm{BMO}(v)} \|f\|_{L^p(\mu)}.$$

We write

$$\langle b \rangle_J - \langle b \rangle_I = [\langle b \rangle_J - \langle b \rangle_K] - [\langle b \rangle_I - \langle b \rangle_K],$$

where we further write

$$\langle b \rangle_J - \langle b \rangle_K = \sum_{J \subsetneq L \subset K} \langle \Delta_L b \rangle_J = \sum_{J \subsetneq L \subset K} \langle b, h_L \rangle \langle h_L \rangle_J,$$

and similarly for $\langle b \rangle_I - \langle b \rangle_K$. We dualize and e.g. look at

$$\sum_{K} \sum_{I^{(i)}=J^{(j)}=K} \sum_{J \subsetneq L \subset K} |\langle b, h_L \rangle| |\langle h_L \rangle_J| |a_{IJK}| |\langle f, h_I \rangle| |\langle g, h_J \rangle|$$

$$= \sum_{K} \sum_{\substack{L \subset K \\ \ell(L) > 2^{-j}\ell(K)}} |\langle b, h_L \rangle| |L|^{-1/2} \sum_{\substack{I^{(i)}=J^{(j)}=K \\ J \subset L}} |a_{IJK}| |\langle f, h_I \rangle| |\langle g, h_J \rangle|$$

$$\lesssim \|b\|_{\mathrm{BMO}(v)} \sum_{K} \int \left(\sum_{\substack{L \subset K \\ \ell(L) > 2^{-j}\ell(K)}} \frac{1_L}{|L|^2} \left[\sum_{\substack{I^{(i)}=J^{(j)}=K \\ J \subset L}} |a_{IJK}| |\langle f, h_I \rangle| |\langle g, h_J \rangle| \right]^{2} \right)^{1/2} v,$$

where we used the weighted H^1-BMO duality. Here

$$\sum_{\substack{I^{(i)}=J^{(j)}=K \\ J\subset L}} |a_{IJK}||\langle f, h_I\rangle||\langle g, h_J\rangle| \leq \frac{1}{|K|} \int_K |\Delta_{K,i}f| \int_L |\Delta_{K,j}g|,$$

and we can bound

$$\sum_K \int \left(\sum_{\substack{L\subset K \\ \ell(L)>2^{-j}\ell(K)}} 1_L \langle|\Delta_{K,i}f|\rangle_K^2 \langle|\Delta_{K,j}g|\rangle_L^2\right)^{1/2} \nu$$

$$\leq j^{1/2} \sum_K \int (M\Delta_{K,i}f)(M\Delta_{K,j}g)\nu$$

$$\leq j^{1/2} \left\|\left(\sum_K |M\Delta_{K,i}f|^2\right)^{1/2}\right\|_{L^p(\mu)} \left\|\left(\sum_K |M\Delta_{K,j}g|^2\right)^{1/2}\right\|_{L^{p'}(\lambda^{1-p'})}$$

$$\lesssim j^{1/2} \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda^{1-p'})}.$$

We are done with the one-parameter case—the desired bi-parameter case can now be done completely similarly by tweaking the proof in [29] using the above idea. □

Remark 4.5 The previous way to use the H^1-BMO duality was to look at

$$\sum_K \sum_{\substack{L\subset K \\ \ell(L)=2^{-l}\ell(K)}} |\langle b, h_L\rangle||L|^{-1/2} \sum_{\substack{I^{(i)}=J^{(j)}=K \\ J\subset L}} |a_{IJK}||\langle f, h_I\rangle||\langle g, h_J\rangle|,$$

where $l = 0, \ldots, j-1$ is fixed, and to apply the H^1-BMO duality to the whole K, L summation. With l fixed this yields a uniform estimate, and there is also a curious 'extra' cancellation present—we can even bound

$$\sum_{\substack{I^{(i)}=J^{(j)}=K \\ J\subset L}} |a_{IJK}||\langle f, h_I\rangle||\langle g, h_J\rangle| \leq \frac{1}{|K|} \int_K |\Delta_{K,i}f| \int_L |g|,$$

that is, forget the $\Delta_{K,j}$ from g. Then it remains to sum over l which yields the dependence j instead of $j^{1/2}$. The way in our proof above is more efficient and we see that we utilize all of the cancellation as well.

Remark 4.6 An interesting question is can we have $\alpha = 1$ instead of $\alpha = 3/2$ by somehow more carefully exploiting the operators Q_k—this would appear to be the optimal result theoretically obtainable by the current methods.

We also note that it is certainly possible to handle higher order commutators, such as, $[T_1, [T_2, [b, T_3]]]$.

We will continue with more multi-parameter commutator estimates – the difference to the above is that now even the singular integrals are allowed to be multi-parameter.

For a weight w on $\mathbb{R}^d := \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ we say that a locally integrable function $b \colon \mathbb{R}^d \to \mathbb{C}$ belongs to the weighted little BMO space $\mathrm{bmo}(w)$ if

$$\|b\|_{\mathrm{bmo}(w)} := \sup_R \frac{1}{w(R)} \int_R |b - \langle b \rangle_R| < \infty,$$

where the supremum is over rectangles $R = I^1 \times I^2 \subset \mathbb{R}^d$. If $w = 1$ we denote the unweighted little BMO space by bmo. There holds that

$$\|b\|_{\mathrm{bmo}(w)} \sim \max \left(\operatorname*{ess\,sup}_{x_1 \in \mathbb{R}^{d_1}} \|b(x_1, \cdot)\|_{\mathrm{BMO}(w(x_1, \cdot))}, \operatorname*{ess\,sup}_{x_2 \in \mathbb{R}^{d_2}} \|b(\cdot, x_2)\|_{\mathrm{BMO}(w(\cdot, x_2))} \right),$$
(4.7)

see [19]. Here $\mathrm{BMO}(w(x_1, \cdot))$ and $\mathrm{BMO}(w(\cdot, x_2))$ are the one-parameter weighted BMO spaces. For example,

$$\|b(x_1, \cdot)\|_{\mathrm{BMO}(w(x_1, \cdot))} := \sup_{I^2} \frac{1}{w(x_1, \cdot)(I^2)} \int_{I^2} |b(x_1, y_2) - \langle b(x_1, \cdot) \rangle_{I^2}| \, dy_2,$$

where the supremum is over cubes $I^2 \subset \mathbb{R}^{d_2}$.

The following theorem was proved in [28] with $\omega_i(t) = t^{\gamma_i}$. The first order case $[b, T]$ appeared before in [19]. See also [29] for the optimality of the space $\mathrm{bmo}(\nu^{1/m})$ in the case $b_1 = \cdots = b_m = b$.

Theorem 4.8 *Let $p \in (1, \infty)$, $\mu, \lambda \in A_p$ be bi-parameter weights and $\nu := \mu^{1/p} \lambda^{-1/p}$. Suppose that T is a bi-parameter (ω_1, ω_2)-CZO and $m \in \mathbb{N}$. Then we have*

$$\|[b_m, \cdots [b_2, [b_1, T]] \cdots]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \prod_{i=1}^m \|b_i\|_{\mathrm{bmo}(\nu^{1/m})}$$

if one of the following conditions holds:

(1) *T is paraproduct free and $\omega_i \in \mathrm{Dini}_{m/2+1}$;*
(2) *$m = 1$ and $\omega_i \in \mathrm{Dini}_{3/2}$;*
(3) *$\omega_i \in \mathrm{Dini}_{m+1}$.*

Proof The proof is similar in spirit to that of Theorem 4.2. We use Lemma 3.11 and estimates for the commutators of the usual bi-parameter model operators. If we use the bounds from [28] directly, we e.g. immediately get

$$\|[b_m, \cdots [b_2, [b_1, Q_{k_1, k_2}]] \cdots]\|_{L^p(\mu) \to L^p(\lambda)}$$
$$\lesssim (1 + k_1)(1 + k_2)(1 + \max(k_1, k_2))^m \prod_{i=1}^m \|b_i\|_{\mathrm{bmo}(\nu^{1/m})}.$$
(4.9)

Similarly, we can read an estimate for all the other model operators from [28]. This gives us the result under the higher regularity assumption (3). Indeed, when using the estimate (4.9) in connection with the representation theorem one ends up with the series

$$\sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \omega_1(2^{-k_1})\omega_2(2^{-k_2})(1+k_1)(1+k_2)(1+\max(k_1,k_2))^m.$$

We split this into two according to whether $k_1 \leq k_2$ or $k_1 > k_2$ and, for example, there holds that

$$\sum_{k_1=0}^{\infty} \omega_1(2^{-k_1})(1+k_1) \sum_{k_2=k_1}^{\infty} \omega_2(2^{-k_2})(1+k_2)^{m+1} \lesssim \sum_{k_1=0}^{\infty} \omega_1(2^{-k_1})(1+k_1)\|\omega_2\|_{\text{Dini}_{m+1}}$$

$$\lesssim \|\omega_1\|_{\text{Dini}_1}\|\omega_2\|_{\text{Dini}_{m+1}}.$$

The first order case $m = 1$ with the desired regularity (assumption (2)) follows as the papers [1,2,19] dealing with commutators of the form $[T_1, [T_2, \ldots [b, T_k]]]$, where each T_k can be multi-parameter, include the proof of the first order case with the H^1-BMO duality strategy. And this strategy can be improved to give the additional square root save as in Theorem 4.2.

For $m \geq 2$ the new square root save becomes tricky. The paper [28] is not at all based on the H^1-BMO duality strategy on which this save is based on (see the proof of Theorem 4.2). We can improve the strategy of [28] for shifts. Thus, we are able to make the square root save for paraproduct free T (assumption (1)). By this we mean that (both partial and full) paraproducts in the dyadic representation of T vanish, which could also be stated in terms of (both partial and full) "$T1 = 0$" type conditions. The reader can think of convolution form SIOs.

We start considering $[b_2, [b_1, S_i]]$, where $i = (i_1, i_2)$, $i_j = (i_j^1, i_j^2)$ and S_i is a standard bi-parameter shift of complexity i. The reductions in pages 23 and 24 of [28] (Sect. 5.1) give that we only need to bound the key term

$$\langle U^{b_1,b_2}f, g \rangle := \sum_{K} \sum_{\substack{R_1,R_2 \\ R_j^{(i_j)}=K}} a_{K,R_1,R_2}[\langle b_1 \rangle_{R_2} - \langle b_1 \rangle_{R_1}][\langle b_2 \rangle_{R_2} - \langle b_2 \rangle_{R_1}]\langle f, h_{R_1} \rangle\langle g, h_{R_2} \rangle,$$

where as usual $K = K^1 \times K^2$ and $R_j = I_j^1 \times I_j^2$.

We write

$$\langle b_i \rangle_{R_2} - \langle b_i \rangle_{R_1} = [\langle b_i \rangle_{R_2} - \langle b_i \rangle_{K^1 \times I_2^2}] + [\langle b_i \rangle_{K^1 \times I_2^2} - \langle b_i \rangle_K]$$

$$+ [\langle b_i \rangle_K - \langle b_i \rangle_{K^1 \times I_1^2}] + [\langle b_i \rangle_{K^1 \times I_1^2} - \langle b_i \rangle_{R_1}].$$

This splits U^{b_1,b_2} into 16 different terms $U^{b_1,b_2}_{m_1,m_2}$, where $m_i \in \{1,\ldots,4\}$ tells which one of the above terms we have for b_i. These can be handled quite similarly, but there are some variations in the arguments. We will handle two representative ones.

We begin by looking at the term

$$\langle U_{3,4}^{b_1,b_2} f, g \rangle := \sum_K \sum_{\substack{R_1,R_2 \\ R_j^{(i_j)}=K}} a_{K,R_1,R_2} [\langle b_1 \rangle_{K^1 \times I_1^2} - \langle b_1 \rangle_K][\langle b_2 \rangle_{R_1}$$

$$- \langle b_2 \rangle_{K^1 \times I_1^2}] \langle f, h_{R_1} \rangle \langle g, h_{R_2} \rangle.$$

Write

$$\langle b_1 \rangle_{K^1 \times I_1^2} - \langle b_1 \rangle_K = \sum_{I_1^2 \subsetneq L^2 \subset K^2} \langle \Delta_{L^2} \langle b_1 \rangle_{K^1,1} \rangle_{I_1^2}$$

$$= \sum_{I_1^2 \subsetneq L^2 \subset K^2} \left\langle b_1, \frac{1_{K^1}}{|K^1|} \otimes h_{L^2} \right\rangle \langle h_{L^2} \rangle_{I_1^2} \qquad (4.10)$$

and

$$\langle b_2 \rangle_{R_1} - \langle b_2 \rangle_{K^1 \times I_1^2} = \sum_{I_1^1 \subsetneq L^1 \subset K^1} \langle \Delta_{L^1} \langle b_2 \rangle_{I_1^2,2} \rangle_{I_1^1} = \sum_{I_1^1 \subsetneq L^1 \subset K^1} \left\langle b_2, h_{L^1} \otimes \frac{1_{I_1^2}}{|I_1^2|} \right\rangle \langle h_{L^1} \rangle_{I_1^1}.$$

Writing $\langle b_1, \frac{1_{K^1}}{|K^1|} \otimes h_{L^2} \rangle = \int_{\mathbb{R}^{d_1}} \langle b_1, h_{L^2} \rangle_2 \frac{1_{K^1}}{|K^1|}$ and similarly for $\langle b_2, h_{L^1} \otimes \frac{1_{I_1^2}}{|I_1^2|} \rangle$ we arrive at

$$\int_{\mathbb{R}^d} \sum_K \sum_{\substack{L=L^1 \times L^2 \subset K \\ \ell(L^j) > 2^{-i_1^j} \ell(K^j)}} |\langle b_1, h_{L^2} \rangle_2||L^2|^{-1/2} |\langle b_2, h_{L^1} \rangle_1||L^1|^{-1/2}$$

$$\sum_{\substack{R_1^{(i_1)}=R_2^{(i_2)}=K \\ R_1 \subset L}} |a_{K,R_1,R_2} \langle f, h_{R_1} \rangle \langle g, h_{R_2} \rangle| \frac{1_{K^1}}{|K^1|} \frac{1_{I_1^2}}{|I_1^2|}.$$

The last line can be dominated by

$$|L^1| \langle M^2 \Delta_{K,i_1} f \rangle_{L^1,1} \langle |\Delta_{K,i_2} g| \rangle_K \frac{1_{K^1}}{|K^1|} 1_{L^2}.$$

We have now reached the term

$$\int_{\mathbb{R}^d} \sum_K \langle |\Delta_{K,i_2} g| \rangle_K \frac{1_{K^1}}{|K^1|} \sum_{\substack{L^2 \subset K^2 \\ \ell(L^2) > 2^{-i_1^2} \ell(K^2)}} |\langle b_1, h_{L^2} \rangle_2||L^2|^{-1/2} 1_{L^2}$$

$$\sum_{\substack{L^1 \subset K^1 \\ \ell(L^1) > 2^{-i_1^1} \ell(K^1)}} |\langle b_2, h_{L^1} \rangle_1| |L^1|^{1/2} \langle M^2 \Delta_{K, i_1} f \rangle_{L^1, 1}.$$

Recall that with fixed x_2 we have $b(\cdot, x_2) \in \mathrm{BMO}(\nu^{1/2}(\cdot, x_2))$, see (4.7). By weighted H^1-BMO duality we now have that

$$\sum_{\substack{L^1 \subset K^1 \\ \ell(L^1) > 2^{-i_1^1} \ell(K^1)}} |\langle b_2, h_{L^1} \rangle_1 (x_2)| |L^1|^{1/2} \langle M^2 \Delta_{K, i_1} f \rangle_{L^1, 1}(x_2)$$

$$\lesssim \|b_2\|_{\mathrm{bmo}(\nu^{1/2})} \int_{\mathbb{R}^{d_1}}$$

$$\left(\sum_{\substack{L^1 \subset K^1 \\ \ell(L^1) > 2^{-i_1^1} \ell(K^1)}} 1_{L^1}(y_1) (\langle M^2 \Delta_{K, i_1} f \rangle_{L^1, 1}(x_2))^2 \right)^{1/2} \nu^{1/2}(y_1, x_2) \, dy_1$$

$$\leq (i_1^1)^{1/2} \|b_2\|_{\mathrm{bmo}(\nu^{1/2})} |K^1| \langle M^1 M^2 \Delta_{K, i_1} f \cdot \nu^{1/2} \rangle_{K^1, 1}(x_2).$$

The term $(i_1^1)^{1/2} \|b_2\|_{\mathrm{bmo}(\nu^{1/2})}$ is fine and we do not drag it along in the following estimates. We are left with the task of bounding

$$\int_{\mathbb{R}^d} \sum_K \langle |\Delta_{K, i_2} g| \rangle_K 1_{K^1} \sum_{\substack{L^2 \subset K^2 \\ \ell(L^2) > 2^{-i_1^2} \ell(K^2)}} |\langle b_1, h_{L^2} \rangle_2| |L^2|^{-1/2} 1_{L^2}$$

$$M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot \nu^{1/2}).$$

We now put the $\int_{\mathbb{R}^{d_2}}$ inside and get the term

$$\int_{\mathbb{R}^{d_2}} 1_{L^2} M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot \nu^{1/2}) = |L^2| \langle M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot \nu^{1/2}) \rangle_{L^2, 2}.$$

Then, we are left with

$$\int_{\mathbb{R}^{d_1}} \sum_K \langle |\Delta_{K, i_2} g| \rangle_K 1_{K^1} \sum_{\substack{L^2 \subset K^2 \\ \ell(L^2) > 2^{-i_1^2} \ell(K^2)}} |\langle b_1, h_{L^2} \rangle_2| |L^2|^{1/2}$$

$$\langle M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot \nu^{1/2}) \rangle_{L^2, 2}.$$

By weighted H^1-BMO duality we have analogously as above that

$$\sum_{\substack{L^2 \subset K^2 \\ \ell(L^2) > 2^{-i_1^2} \ell(K^2)}} |\langle b_1, h_{L^2} \rangle_2| |L^2|^{1/2} \langle M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot v^{1/2}) \rangle_{L^2, 2}$$

$$\lesssim (i_1^2)^{1/2} \|b_1\|_{\mathrm{bmo}(v^{1/2})} \int_{\mathbb{R}^{d_2}} 1_{K^2} M^2 M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot v^{1/2}) v^{1/2}.$$

Forgetting the factor $(i_1^2)^{1/2} \|b_1\|_{\mathrm{bmo}(v^{1/2})}$, which is as desired, we are then left with

$$\int_{\mathbb{R}^d} \sum_K \langle |\Delta_{K, i_2} g| \rangle_K 1_K M^2 M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot v^{1/2}) v^{1/2}$$

$$\leq \int_{\mathbb{R}^d} \sum_K M^2 M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot v^{1/2}) \cdot M^1 M^2 \Delta_{K, i_2} g \cdot v^{1/2}.$$

Writing $v^{\frac{1}{2}} = \mu^{\frac{1}{2p}} \lambda^{\frac{1}{2p}} \cdot \lambda^{-\frac{1}{p}}$ we bound this with

$$\left\| \left(\sum_K [M^2 M^1 (M^1 M^2 \Delta_{K, i_1} f \cdot v^{1/2})]^2 \right)^{1/2} \right\|_{L^p(\mu^{1/2} \lambda^{1/2})}$$

multiplied by

$$\left\| \left(\sum_K [M^1 M^2 \Delta_{K, i_2} g]^2 \right)^{1/2} \right\|_{L^{p'}(\lambda^{1-p'})}.$$

It remains to use square function bounds together with the Fefferman–Stein inequality. For the more complicated term with the function f the key thing to notice is that first $\mu^{1/2} \lambda^{1/2} \in A_p$ and then that $v^{p/2} \mu^{1/2} \lambda^{1/2} = \mu$. We have controlled $\langle U_{3,4}^{b_1, b_2} f, g \rangle$.

The bound for $\langle U^{b_1, b_2} f, g \rangle$ follows by handling the other similar terms $U_{m_1, m_2}^{b_1, b_2}$. There is a slight variation in the argument needed, for example, in the following term

$$\langle U_{1,1}^{b_1, b_2} f, g \rangle := \sum_K \sum_{\substack{R_1, R_2 \\ R_j^{(i_j)} = K}} a_{K, R_1, R_2} [\langle b_1 \rangle_{R_2} - \langle b_1 \rangle_{K^1 \times I_2^2}] [\langle b_2 \rangle_{R_2}$$

$$- \langle b_2 \rangle_{K^1 \times I_2^2}] \langle f, h_{R_1} \rangle \langle g, h_{R_2} \rangle.$$

We expand the differences of averages as

$$[\langle b_1 \rangle_{R_2} - \langle b_1 \rangle_{K^1 \times I_2^2}][\langle b_2 \rangle_{R_2} - \langle b_2 \rangle_{K^1 \times I_2^2}]$$

$$= \sum_{I_2^1 \subsetneq U^1 \subset K^1} \sum_{I_2^1 \subsetneq V^1 \subset K^1} \left\langle b_1, h_{U^1} \otimes \frac{1_{I_2^2}}{|I_2^2|} \right\rangle \langle h_{U^1} \rangle_{I_2^1} \left\langle b_2, h_{V^1} \otimes \frac{1_{I_2^2}}{|I_2^2|} \right\rangle \langle h_{V^1} \rangle_{I_2^1}.$$

The key difference to the above term $U_{3,4}^{b_1,b_2}$ is that we need to further split this into two by comparing whether we have $V^1 \subset U^1$ or $U^1 \subsetneq V^1$. The related two terms are handled symmetrically. The absolute value of the one coming from "$V^1 \subset U^1$" can be written as

$$\int_{\mathbb{R}^{d_2}} \int_{\mathbb{R}^{d_2}} \sum_{K} \sum_{\substack{U^1 \subset K^1 \\ \ell(U^1) > 2^{-i_2^1} \ell(K^1)}} \\ \sum_{\substack{V^1 \subset U^1 \\ \ell(V^1) > 2^{-i_2^1} \ell(K^1)}} |\langle b_1, h_{U^1} \rangle_1(x_2)||U^1|^{-1/2}|\langle b_2, h_{V^1} \rangle_1(y_2)||V^1|^{-1/2} \\ \sum_{\substack{(I_1^1)^{(i_1^1)} = (I_2^1)^{(i_2^1)} = K^1 \\ I_2^1 \subset V^1}} \sum_{(I_1^2)^{(i_1^2)} = (I_2^2)^{(i_2^2)} = K^2} |a_{K,R_1,R_2} \langle f, h_{R_1} \rangle \langle g, h_{R_2} \rangle| \frac{1_{I_2^2}(x_2)}{|I_2^2|} \frac{1_{I_2^2}(y_2)}{|I_2^2|}.$$

The last line can be dominated by

$$\langle |\Delta_{K,i_1} f| \rangle_K |V^1| \sum_{(I_2^2)^{(i_2^2)} = K^2} \langle |\Delta_{K,i_2} g| \rangle_{V^1 \times I_2^2} \frac{1_{I_2^2}(x_2)}{|I_2^2|} 1_{I_2^2}(y_2).$$

Using the weighted H^1-BMO duality as above we have

$$\int_{\mathbb{R}^{d_2}} \sum_{\substack{V^1 \subset U^1 \\ \ell(V^1) > 2^{-i_2^1} \ell(K^1)}} |\langle b_2, h_{V^1} \rangle_1(y_2)||V^1|^{1/2} \langle |\Delta_{K,i_2} g| \rangle_{V^1 \times I_2^2} 1_{I_2^2}(y_2) \, dy_2$$

$$\leq (i_2^1)^{1/2} \|b_2\|_{\mathrm{bmo}(v^{1/2})} |U^1||I_2^2| \langle M^1 M^2 \Delta_{K,i_2} g \cdot v^{1/2} \rangle_{U^1 \times I_2^2}.$$

Forgetting the factor $(i_2^1)^{1/2} \|b_2\|_{\mathrm{bmo}(v^{1/2})}$ we have reached the term

$$\int_{\mathbb{R}^{d_2}} \sum_K \langle |\Delta_{K,i_1} f| \rangle_K \sum_{(I_2^2)^{(i_2^2)} = K^2} 1_{I_2^2} \sum_{\substack{U^1 \subset K^1 \\ \ell(U^1) > 2^{-i_2^1} \ell(K^1)}} |\langle b_1, h_{U^1} \rangle_1||U^1|^{1/2}$$

$$\langle M^1 M^2 \Delta_{K,i_2} g \cdot v^{1/2} \rangle_{U^1 \times I_2^2},$$

which—after using the H^1-BMO duality—produces $(i_2^1)^{1/2}\|b_1\|_{\mathrm{bmo}(\nu^{1/2})}$ multiplied by

$$\int_{\mathbb{R}^d} \sum_K \langle |\Delta_{K,i_1} f| \rangle_K M^1 M^2 (M^1 M^2 \Delta_{K,i_2} g \cdot \nu^{1/2}) \nu^{1/2} 1_K.$$

Similarly as with $U_{3,4}^{b_1,b_2}$, this term is under control. The term with $U^1 \subsetneq V^1$ is symmetric, and so we are also done with $U_{1,1}^{b_1,b_2}$.

This ends our treatment of U^{b_1,b_2}, since the above arguments showcased the only major difference between the various terms $U_{m_1,m_2}^{b_1,b_2}$. Thus, we are done with $[b_2, [b_1, S_i]]$. By Lemma 3.11 we conclude that

$$\|[b_2, [b_1, Q_{k_1,k_2}]]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim (1+k_1)(1+k_2)(1+\max(k_1,k_2)) \prod_{i=1}^2 \|b_i\|_{\mathrm{bmo}(\nu^{1/2})}.$$

By handling the higher order commutators similarly, we get the claim related to assumption (1). We omit these details. □

Remark 4.11 The new square root save from the H^1-BMO arguments reduces the required regularity from $m + 1$ to $m/2 + 1$. In these higher order commutators this is more significant than the save that could theoretically be obtained by not using Lemma 3.11. This could change the $+1$ to $+1/2$.

Theorem 4.2 involves only one-parameter CZOs in its estimate

$$\|[T_1, [T_2, b]]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}_{\mathrm{prod}}(\nu)},$$

while the basic estimate

$$\|[b, T]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \|b\|_{\mathrm{bmo}(\nu)}$$

of Theorem 4.8 involves a bi-parameter CZO T. A joint generalization—considered in the unweighted case in [36]—is an estimate for

$$\|[T_1, [T_2, \ldots [b, T_k]]]\|_{L^p(\mu) \to L^p(\lambda)},$$

where each T_i can be a completely general m-parameter CZO. Then the appearing BMO norm is some suitable combination of little BMO and product BMO. See [1,2] for a fully satisfactory Bloom type upper estimate in this generality – however, only for CZOs with the standard kernel regularity. The general case of [1,2] is hard to digest, but let us formulate a model theorem of this type with mild kernel regularity.

Theorem 4.12 *Let $\mathbb{R}^d = \prod_{i=1}^4 \mathbb{R}^{d_i}$ be a product space of four parameters and let $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2\}$, where $\mathcal{I}_1 = \{1, 2\}$ and $\mathcal{I}_2 = \{3, 4\}$, be a partition of the parameter space $\{1, 2, 3, 4\}$. Suppose that T_i is a bi-parameter $(\omega_{1,i}, \omega_{2,i})$-CZO on $\prod_{j \in \mathcal{I}_i} \mathbb{R}^{d_j}$,*

where $\omega_{j,i} \in \text{Dini}_{3/2}$. Let $b: \mathbb{R}^d \to \mathbb{C}$, $p \in (1, \infty)$, $\mu, \lambda \in A_p(\mathbb{R}^d)$ be 4-parameter weights and $\nu = \mu^{1/p}\lambda^{-1/p}$ be the associated Bloom weight. Then we have

$$\|[T_1, [T_2, b]]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \|b\|_{\text{bmo}^{\mathcal{I}}(\nu)}.$$

Here $\text{bmo}^{\mathcal{I}}(\nu)$ is the following weighted little product BMO *space:*

$$\|b\|_{\text{bmo}^{\mathcal{I}}(\nu)} = \sup_{\bar{u}} \|b\|_{\text{BMO}^{\bar{u}}_{\text{prod}}(\nu)},$$

where $\bar{u} = (u_i)_{i=1}^2$ is such that $u_i \in \mathcal{I}_i$ and $\text{BMO}^{\bar{u}}_{\text{prod}}(\nu)$ is the natural weighted bi-parameter product BMO *space on the parameters $\bar{u}$. For example,*

$$\|b\|_{\text{BMO}^{(1,3)}_{\text{prod}}(\nu)} := \sup_{x_2 \in \mathbb{R}^{d_2}, x_4 \in \mathbb{R}^{d_4}} \|b(\cdot, x_2, \cdot, x_4)\|_{\text{BMO}_{\text{prod}}(\nu(\cdot, x_2, \cdot, x_4))},$$

where the last weighted product BMO *norm is defined in (4.1).*

The proof is again a combination of Lemma 3.11 with the known estimates for the commutators of standard model operators [1,2]. However, there is again the additional square root save. There are no new significant challenges with this, which was not the case with Theorem 4.8 above, since these references are completely based on the H^1-BMO strategy. In this regard the situation is closer to that of Theorem 4.2.

References

1. Airta, E.: Two-weight commutator estimates: general multi-parameter framework. Publ. Mat. **64**(2), 681–729 (2020)
2. Airta, E., Li, K., Martikainen, H., Vuorinen, E.: Some new weighted estimates on product space. Indiana Univ. Math. J. (2019), to appear. https://www.iumj.indiana.edu/IUMJ/Preprints/8807.pdf
3. Barron, A., Conde-Alonso, J.M., Ou, Y., Rey, G.: Sparse domination and the strong maximal function. Adv. Math. **345**, 1–26 (2019)
4. Barron, A., Pipher, J.: Sparse domination for bi-parameter operators using square functions (2017), preprint. arXiv:1709.05009
5. Chang, S.-Y.A., Fefferman, R.: A continuous version of duality of H1 with BMO on the bidisc. Ann. Math. (2) **112**(1), 179–201 (1980)
6. Chang, S.-Y.A., Fefferman, R.: Some recent developments in Fourier analysis and Hp-theory on product domains. Bull. Am. Math. Soc. (N.S.) **12**(1), 1–43 (1985)

7. Chang, S.-Y.A., Wilson, J.M., Wolff, T.H.: Some weighted norm inequalities concerning the Schrödinger operators. Comment. Math. Helv. **60**(2), 217–246 (1985)

8. Cruz-Uribe, D., Martell, J.M., Pérez, C.: Sharp weighted estimates for classical operators. Adv. Math. **229**(1), 408–441 (2012)

9. Dalenc, L., Ou, Y.: Upper bound for multi-parameter iterated commutators. Publ. Mat. **60**(1), 191–220 (2016)

10. David, G., Journé, J.L.: A boundedness criterion for generalized Calderón-Zygmund operators. Ann. Math. (2) **120**(2), 3710397 (1984)

11. Deng, D., Yan, L., Yang, Q.: Blocking analysis and T (1) theorem. Sci. China Ser. A **41**(8), 801–808 (1998)

12. Di Plinio, F., Li, K., Martikainen, H., Vuorinen, E.: Banach-valued multilinear singular integrals with modulation invariance. Int. Math. Res. Not. (2020), to appear. https://doi.org/10.1093/imrn/rnaa234

13. Di Plinio, F., Li, K., Martikainen, H., Vuorinen, E.: Multilinear operator-valued Calderón-Zygmund theory. J. Funct. Anal. **279**(8), 108666 (2020)

14. Di Plinio, F., Li, K., Martikainen, H., Vuorinen, E.: Multilinear singular integrals on non-commutative Lp spaces. Math. Ann. **378**(3–4), 1371–1414 (2020)

15. Figiel, T.: On equivalence of some bases to the Haar system in spaces of vector-valued functions. Bull. Polish Acad. Sci. Math. 36 (1988), no. 3-4, 119–131 (1989)

16. Figiel, T.: Singular integral operators: a martingale approach. In: Geometry of Banach Spaces (Strobl, 1989), (1990), pp. 95–110

17. Grau de la Herrán, A., Hytönen, T.: Dyadic representation and boundedness of nonhomogeneous Calderón- Zygmund operators with mild kernel regularity. Mich. Math. J. **67**(4), 757–786 (2018)

18. Holmes, I., Lacey, M.T., Wick, B.D.: Commutators in the two-weight setting. Math. Ann. **367**(1–2), 51–80 (2017)

19. Holmes, I., Petermichl, S., Wick, B.D.: Weighted little bmo and two-weight inequalities for Journé commutators. Anal. PDE **11**(7), 1693–1740 (2018)

20. Hytönen, T.: Representation of singular integrals by dyadic operators, and the A2 theorem. Expo. Math. **35**, 166–205 (2011)

21. Hytönen, T.P.: The sharp weighted bound for general Calderón-Zygmund operators. Ann. Math. (2) **175**(3), 1473–1506 (2012)

22. Hytönen, T.P., Roncal, L., Tapiola, O.: Quantitative weighted estimates for rough homogeneous singular integrals. Isr. J. Math. **218**(1), 133–164 (2017)

23. Journé, J.-L.: Calderón-Zygmund operators on product spaces. Rev. Mat. Iberoam. **1**, 550–91 (1985)

24. Lacey, M.T.: An elementary proof of the A2 bound. Isr. J. Math. **217**(1), 181–195 (2017)

25. Lerner, A.K., Ombrosi, S., Rivera-Ríos, I.P.: On pointwise and weighted estimates for commutators of Calderón-Zygmund operators. Adv. Math. **319**, 153–181 (2017)

26. Lerner, A.K., Ombrosi, S., Rivera-Ríos, I.P.: Commutators of singular integrals revisited. Bull. Lond. Math. Soc. **51**(1), 107–119 (2019)

27. Li, K., Martikainen, H., Ou, Y., Vuorinen, E.: Bilinear representation theorem. Trans. Am. Math. Soc. **371**(6), 4193–4214 (2019)

28. Li, K., Martikainen, H., Vuorinen, E.: Bloom-type inequality for bi-parameter singular integrals: efficient proof and iterated commutators. Int. Math. Res. Not. IMRN (2019), rnz072. https://doi.org/10.1093/imrn/rnz072

29. Li, K., Martikainen, H., Vuorinen, E.: Bloom type upper bounds in the product BMO setting. J. Geom. Anal. **30**, 3181–3203 (2019)

30. Li, K., Martikainen, H., Vuorinen, E.: Bilinear Calderón-Zygmund theory on product spaces. J. Math. Pures Appl. **138**, 356–412 (2020)

31. Li, K., Martikainen, H., Vuorinen, E.: Genuinely multilinear weighted estimates for singular integrals in product spaces. Adv. Math. **393** (2021). https://doi.org/10.1016/j.aim.2021.108099

32. Martikainen, H.: Representation of bi-parameter singular integrals by dyadic operators. Adv. Math. **229**(3), 1734–1761 (2012)

33. Martikainen, H., Vuorinen, E.: Dyadic-probabilistic methods in bilinear analysis. Mem. Am. Math. Soc. **274** (2021), no. 1344. https://doi.org/10.1090/memo/1344

34. Muckenhoupt, B., Wheeden, R.L.: Weighted bounded mean oscillation and the Hilbert transform. Studia Math. 54 (1975/76), no. 3, 221–237

35. Ou, Y.: Multi-parameter singular integral operators and representation theorem. Rev. Mat. Iberoam. **33**(1), 325–350 (2017)

36. Ou, Y., Petermichl, S., Strouse, E.: Higher order Journé commutators and characterizations of multi-parameter BMO. Adv. Math. **291**, 24–58 (2016)
37. Wu, S.: A wavelet characterization for weighted Hardy spaces. Rev. Mat. Iberoam. **8**(3), 329–349 (1992)