

# BILINEAR CALDERÓN–ZYGmund THEORY ON PRODUCT SPACES

KANGWEI LI, HENRI MARTIKAINEN, AND EMIL VUORINEN

ABSTRACT. We develop a wide general theory of bilinear bi-parameter singular integrals  $T$ . This includes general Calderón–Zygmund type principles in the bilinear bi-parameter setting: easier bounds, like estimates in the Banach range, imply boundedness in the full bilinear range  $L^p \times L^q \rightarrow L^r$ ,  $1/p + 1/q = 1/r$ ,  $1 < p, q \leq \infty$ ,  $1/2 < r < \infty$ , weighted estimates, mixed-norm estimates, and so on. We build this Calderón–Zygmund theory using the very useful perspective of dyadic representation theorems that hold under testing conditions. New weighted estimates are developed and used effectively throughout. We also develop commutator decompositions and use them to show estimates in the full range for commutators and iterated commutators, like  $[b_1, T]_1$  and  $[b_2, [b_1, T]_1]_2$ , where  $b_1$  and  $b_2$  are little BMO functions. Our commutator method contains several novelties – one is that the new method can be used to simplify and improve linear commutator proofs, even in the two-weight Bloom setting. We also quickly show commutator lower bounds by using and developing the recent median method.

RÉSUMÉ. Nous développons une vaste théorie générale des intégrales singulières bilinéaires à deux paramètres  $T$ . Ceci inclut les principes généraux du type Calderón–Zygmund dans le cadre bilinéaire à deux paramètres : des limites plus faciles, comme les estimations dans la gamme de Banach, impliquent la bornitude de la gamme bilinéaire complète  $L^p \times L^q \rightarrow L^r$ ,  $1/p + 1/q = 1/r$ ,  $1 < p, q \leq \infty$ ,  $1/2 < r < \infty$ , les estimations pondérées, les estimations de normes mixtes, etc. Nous construisons cette théorie de Calderón–Zygmund en utilisant la perspective très utile des théorèmes de la représentation dyadique qui sont valables sous les conditions de test. Des nouvelles estimations pondérées sont développées et utilisées efficacement partout. Nous développons également des décompositions de commutateurs et les utilisons pour montrer des estimations dans toute la gamme des commutateurs et des commutateurs itérés, comme  $[b_1, T]_1$  et  $[b_2, [b_1, T]_1]_2$ , où  $b_1$  et  $b_2$  sont fonctions de BMO petites. Notre méthode de commutateur contient plusieurs nouveautés – une est que la nouvelle méthode peut être utilisé pour simplifier et améliorer les épreuves de commutateur linéaire, même dans le cadre Bloom à deux poids. Nous montrons aussi rapidement que les limites inférieures du commutateur en utilisant et en développant la récente méthode médiane.

## 1. INTRODUCTION

### 1.1. Estimates for singular integrals in various settings.

*Linear and multilinear theory.* A singular integral operator (SIO) refers to a linear operator  $T$  for which

$$\langle T f_1, f_2 \rangle := \int_{\mathbb{R}^n} T f_1(x) f_2(x) dx = \iint_{\mathbb{R}^n \times \mathbb{R}^n} K(x, y) f_1(y) f_2(x) dy dx$$

---

2010 *Mathematics Subject Classification.* 42B20.

*Key words and phrases.* Calderón–Zygmund operators, bi-parameter analysis, bilinear analysis, representation theorems, weighted estimates, commutators.

for some suitable kernel  $K$  and for all nice functions  $f_1, f_2$  that are disjointly supported. A fundamental problem is to prove that  $T$  is bounded in  $L^p$ ,  $p \in (1, \infty)$ , under the assumption that  $T$  is bounded in some  $L^{p_0}$ . A key feature of the classical theory is that the following argument achieves this: prove weak  $(1, 1)$  via Calderón–Zygmund decomposition, interpolate and dualise. See any standard textbook such as Stein [59].

We discuss the bilinear theory of SIOs. A heuristic model of a bilinear SIO  $T$  in  $\mathbb{R}^n$  is  $T(f_1, f_2)(x) := \tilde{T}(f_1 \otimes f_2)(x, x)$ , where  $x \in \mathbb{R}^n$ ,  $f_i: \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $(f_1 \otimes f_2)(x_1, x_2) = f_1(x_1)f_2(x_2)$  and  $\tilde{T}$  is a linear SIO in  $\mathbb{R}^{2n}$ . More precisely, a bilinear SIO  $T$  has a kernel  $K$  satisfying estimates that are obtained from the above heuristic via the linear estimates, and if  $\text{spt } f_i \cap \text{spt } f_j = \emptyset$  for some  $i, j$  then

$$\langle T(f_1, f_2), f_3 \rangle = \iiint_{\mathbb{R}^{3n}} K(x, y, z) f_1(y) f_2(z) f_3(x) dy dz dx.$$

A fundamental aspect brought to play by the bilinearity is that now the natural range for boundedness is  $L^p \times L^q \rightarrow L^r$  for  $p, q \in (1, \infty]$  and  $r \in (1/2, \infty)$  satisfying the scaling  $1/p + 1/q = 1/r$ . If  $p, q, r \in (1, \infty)$  we talk about the Banach range, and otherwise about the quasi-Banach range, where we may have  $r < 1$ . There are a lot of ways, including bilinear interpolation, sparse domination and good lambda type arguments, to go from the boundedness  $L^{p_0} \times L^{q_0} \rightarrow L^{r_0}$ , where  $p_0, q_0, r_0$  are fixed, to the whole range. Some of these methods go through the endpoint estimate  $L^1 \times L^1 \rightarrow L^{1/2, \infty}$ , which again follows via the Calderón–Zygmund decomposition. See e.g. Grafakos–Torres [24]. More modern aspects are cited and used e.g. in [47].

*Multi-parameter theory.* We deal with bi-parameter theory. An example of a bi-parameter SIO in the product space  $\mathbb{R}^n \times \mathbb{R}^m$  is  $T_n \otimes T_m$ , where  $T_n$  and  $T_m$  are linear SIOs in  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively, and  $(T_n \otimes T_m)(f_1 \otimes f_2)(x) = T_n f_1(x_1) T_m f_2(x_2)$  for  $x = (x_1, x_2) \in \mathbb{R}^{n+m}$  and two functions  $f_1, f_2$  defined in  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively. Noticing that  $T_n \otimes T_m = T_n^1 T_m^2$ , where e.g.  $T_n^1 f(x) = T_n(f(\cdot, x_2))(x_1)$ , it follows by Fubini that  $\|T_n \otimes T_m\|_{L^p \rightarrow L^p} \leq \|T_n\|_{L^p \rightarrow L^p} \|T_m\|_{L^p \rightarrow L^p}$ . Theory of non-tensor form bi-parameter SIOs is interesting. The classical references Fefferman–Stein [19] and Journé [37] deal with convolution form and general kernels, respectively. The kernel structure of a non-tensor form SIO can be deduced from the tensor product case. We require two kinds of kernel representations depending whether we have separation in both  $\mathbb{R}^n$  and  $\mathbb{R}^m$  (a full kernel representation), or just in  $\mathbb{R}^n$  or  $\mathbb{R}^m$  (a partial kernel representation). This viewpoint on the kernels is recent, see [50] by one of us and Pott-Villarroya [57]. Grau de la Herrán [25] showed that the operator-valued formulations of Journé [37] are equivalent to the more tangible full and partial kernel assumptions of [50].

The one-parameter machinery for achieving  $L^p$  boundedness from boundedness in a single  $L^{p_0}$  via weak  $(1, 1)$  does not work as this endpoint is false. An alternative route via the boundedness  $L^\infty \rightarrow \text{BMO}_{\text{prod}}$ , where  $\text{BMO}_{\text{prod}}$  is the product BMO of Chang and Fefferman [3, 4], exists. That this can be interpolated via the implication  $T^*: H^1 \rightarrow L^1$  requires Chang–Fefferman [5], which heavily involves product Hardy space machinery. The estimate  $L^\infty \rightarrow \text{BMO}_{\text{prod}}$  is also very important for the  $T1$  theory. This is proved via a difficult argument involving Journé’s covering theorem [36]. For a modern proof see e.g. Hytönen–Martikainen [35], which works even in the non-doubling situation. For dyadic viewpoints on product BMO and  $H^1$  see Pipher–Ward [56] and Treil [60].

An alternative way to prove  $T: L^p \rightarrow L^p$ ,  $p \in (1, \infty)$ , is to show that  $T1$  assumptions imply  $L^p$  boundedness (and not just  $L^2$ ). This avoids the  $H^1$  theory and follows e.g. from the representation theorem [50]. If a bi-parameter SIO  $T$  and its partial adjoint  $T_1$  (defined via  $\langle T_1(f_1 \otimes f_2), g_1 \otimes g_2 \rangle = \langle T(g_1 \otimes f_2), f_1 \otimes g_2 \rangle$ ) are bounded in some  $L^{p_0}$ , possibly with a different  $p_0$ , the estimate  $L^\infty \rightarrow \text{BMO}_{\text{prod}}$  implies the  $T1$  assumptions for  $T, T^*, T_1, T_1^*$ , and so  $T$  is bounded in all  $L^p$ . The partial adjoint has to be mentioned as it need not be bounded even if  $T$  is. We will show that at least this viewpoint has a useful analog in the bilinear bi-parameter setting. Of course, the original main point of  $T1$  theorems also holds: they show the boundedness just from testing conditions.

*Multilinear multi-parameter theory.* Again, we focus on the bilinear bi-parameter setting. A model of a bilinear bi-parameter singular integral in  $\mathbb{R}^n \times \mathbb{R}^m$  is

$$(T_n \otimes T_m)(f_1 \otimes f_2, g_1 \otimes g_2)(x) := T_n(f_1, g_1)(x_1)T_m(f_2, g_2)(x_2),$$

where  $f_1, g_1: \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $f_2, g_2: \mathbb{R}^m \rightarrow \mathbb{C}$ ,  $x = (x_1, x_2) \in \mathbb{R}^{n+m}$  and  $T_n, T_m$  are bilinear SIOs defined in  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively. Unlike in the linear bi-parameter case, the theory of the tensor products is already non-trivial – we will return to this point. As in the linear bi-parameter case, the natural kernel structure of a general bilinear bi-parameter SIO can be deduced from the tensor product case. These are the operators that we study in this paper. Difficult estimates for some concrete bilinear bi-parameter SIOs have had interesting applications – we get to these later. A systematic abstract theory, with all the underlying general principles discovered, has been missing. We claim that developing such theory has big advantages.

The most fundamental question is the boundedness of a bilinear bi-parameter SIO  $T$  in the full natural range  $L^p \times L^q \rightarrow L^r$ , where again  $p, q \in (1, \infty]$  and  $r \in (1/2, \infty)$  satisfy  $1/p + 1/q = 1/r$ . An easy to explain corollary of our theory is that if  $T$  is a bilinear bi-parameter SIO (which refers just to possessing the kernel structure),

$$(1.1) \quad \|T(f_1, f_2)\|_{L^{r_0}} \lesssim \|f_1\|_{L^{p_0}} \|f_2\|_{L^{q_0}}, \quad 1/p_0 + 1/q_0 = 1/r_0 < 1,$$

and the same estimate holds for the partial adjoints of  $T$  (maybe with different exponents  $p_0, q_0, r_0$ ), then we have in the full range that  $\|T(f_1, f_2)\|_{L^r} \lesssim \|f_1\|_{L^p} \|f_2\|_{L^q}$ . If  $T(1, 1) = 0$ , and the same holds for the adjoints and the partial adjoints, we call  $T$  free of full paraproducts (this terminology will become clear later). In this case

$$(1.2) \quad \|T(f_1 \otimes f_2, g_1 \otimes g_2)\|_{L^{r_0}} \lesssim \|f_1\|_{L^{p_0}} \|f_2\|_{L^{p_0}} \|g_1\|_{L^{q_0}} \|g_2\|_{L^{q_0}}$$

for some  $p_0, q_0, r_0$  with  $1/p_0 + 1/q_0 = 1/r_0 \leq 1$  gives the boundedness of  $T$  everywhere.

In general, we get the boundedness of SIOs in the full range under boundedness and cancellation conditions, which consist of product BMO type conditions such as  $T(1, 1) \in \text{BMO}_{\text{prod}}$  (this follows from (1.1) but also from weaker assumptions) and some mild conditions for which the tensor boundedness (1.2) is sufficient. *The story we have been trying to tell is that while this type of theory can be used to prove the boundedness via the  $T1$  assumptions (which is as fundamental here as it is in the classical theory), it is also a machine yielding general Calderón–Zygmund type principles, e.g. that the boundedness in the full range follows from the boundedness in a fixed tuple  $1/p_0 + 1/q_0 = 1/r_0 < 1$ , in the bilinear bi-parameter generality.*

Bilinear bi-parameter multipliers (that are special SIOs, see the Appendix) were studied by Muscalu–Pipher–Tao–Thiele [51], and they applied their theory to Leibniz rule type estimates relevant in non-linear PDE. The difficulty was to establish the boundedness in the quasi-Banach range, as the boundedness in the Banach range was easy – see

however the above point in cursive. Other references for multilinear multi-parameter multipliers are Chen–Lu [7] (e.g. a weighted bound and some restricted smoothness results), Grafakos–He–Nguyen–Yan [23] (restricted smoothness) and Di Plinio–Ou [13] (e.g. mixed-norm estimates). We also mention Lacey–Metcalfe [41].

In the free of full paraproducts case we also get bilinear weighted estimates involving bi-parameter Muckenhoupt  $A_p$  classes, and we can in fact use this as the technology to get the full range via extrapolation. The proof of the weighted estimates requires novel methods. We also prove mixed-norm estimates, where  $L^p(\mathbb{R}^{n+m})$  is replaced by  $L^{p_1}(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))$ . Such estimates are topical (e.g. the multipliers are free of full paraproducts). For a bilinear bi-parameter SIO free of full paraproducts the simple condition (1.2) (or some weaker conditions) yields **all** these difficult estimates.

In Journé [38] tensor products  $U_n \otimes U_m$  of “multilinear singular integral forms” are studied – see Christ–Journé [8] for the definition of these forms that they call  $\delta$ - $n$  SIF. It is stated that a tensor product of general bilinear operators, both bounded  $L^\infty \times L^2 \rightarrow L^2$ , need not be bounded  $L^\infty \times L^2 \rightarrow L^2$ . It was proved that this is true, however, for their singular forms, and this was further applied to certain Cauchy type operators improving a previous application of the bi-parameter  $T1$  theorem [38]. We mention that we get that  $T_n \otimes T_m$  – a tensor product of bounded bilinear SIOs – is weighted bounded in the full range, even without paraproduct free assumptions – see Remark 6.18.

*Commutators and methodology.* Our second contribution deals with commutator estimates and methods to prove them. Commutator estimates have remained at the heart of modern harmonic analysis since the result of Coifman–Rochberg–Weiss [11] showing that

$$\|b\|_{\text{BMO}} \lesssim \|[b, T]\|_{L^p \rightarrow L^p} \lesssim \|b\|_{\text{BMO}}, \text{ where } [b, T]f := bTf - T(bf),$$

for a class of non-degenerate CZOs  $T$ . Both the upper and lower bounds of this estimate are non-trivial, and allow deep extensions – notably to the multi-parameter setting, see e.g. Ferguson–Sadosky [21] and Ferguson–Lacey [20]. We significantly contribute to the methodology by developing commutator decompositions and estimates. Then we prove upper and lower bound commutator estimates in the bilinear bi-parameter framework.

A technical backbone for all of the above is obtained by first developing further the probabilistic–dyadic methods, which were initially pioneered by Nazarov–Treil–Volberg [52] in their study of non-doubling singular integrals. These methods have previously also been significantly developed by many leading analysts to prove state-of-the-art weighted results, and also to prove results in the border of harmonic analysis and geometric measure theory (see Tolsa’s book [61]). Our modern approach is key to our results, but also offers many new avenues for future research in the bilinear bi-parameter setup – such avenues e.g. include Bloom type two-weight estimates and non-doubling analysis.

We now move on to give a more detailed account of the field and our theory.

**1.2. Dyadic representation theorems.** A representation theorem aims to represent SIOs by using some natural dyadic model operators (DMOs). The aim is to reduce questions to these simpler DMOs. Such theorems have reshaped the modern thinking of singular integrals, and the development of these methods is the reason why many problems in the bi-parameter scene are more attainable today.

Recently, representation theorems have been shown to hold for both the bi-parameter and bilinear SIOs, but not in the simultaneous presence of both these difficulties. Moreover, the theory of the DMOs themselves and the usage of the representation is much

more complicated in the bilinear bi-parameter framework. The bi-parameter representation theorem is by one of us [50] and the bilinear case is by Li–Martikainen–Ou–Vuorinen [47]. The representations provide precise structural information, which is often of key importance. That is the case also in this paper.  $T1$  theorems are byproducts.

In the linear bi-parameter context the representation theorem [50] has proved to be very useful e.g. in connection with bi-parameter commutators and weighted analysis, see e.g. Holmes–Petermichl–Wick [30] and Ou–Petermichl–Strouse [54]. In the bilinear bi-parameter theory quasi–Banach estimates provide a key additional challenge not present in the linear situation, and similarly weighted estimates are harder to obtain.

The representation theorems were originally motivated by the sharp weighted  $A_p$  theory. Petermichl [55] showed a representation for the Hilbert transform, and Hytönen [31] proved a representation theorem for all bounded linear one-parameter SIOs. While the rise of sparse domination methods has somewhat reduced the need of one-parameter representation theorems, this is not the case in multi-parameter situations, where a fully satisfying theory of sparse domination is missing (although see Barron–Pipher [1]).

There are three types of DMOs that are relevant for us: bilinear bi-parameter shifts, partial paraproducts and full paraproducts. We show, starting from  $T1$  type assumptions, that our SIOs have a representation as an average of a rapidly decaying sum of DMOs. While the formulation is naturally of similar nature as in the previous representation theorems, the underlying dyadic structure is extremely involved in this situation, and the proof requires a very delicate decomposition. The representation is a very important technical component of the paper, but we regard e.g. the subsequent weighted and commutator estimates more novel.

The following terminology is convenient. A bilinear bi-parameter SIO should have the suitable kernel representations – see Definition 3.1. A SIO satisfying the  $T1$  type assumptions – see Definition 3.5 – is a Calderón–Zygmund operator (CZO).

**1.3. Theorem.** *Suppose  $T$  is a bilinear bi-parameter CZO. Then we have*

$$\langle T(f_1, f_2), f_3 \rangle = C_T \mathbb{E}_\omega \sum_{\substack{k=(k_1, k_2, k_3) \in \mathbb{Z}_+^3 \\ v=(v_1, v_2, v_3) \in \mathbb{Z}_+^3}} \alpha_{k,v} \langle U_\omega^{k,v}(f_1, f_2), f_3 \rangle,$$

where  $\omega = (\omega_1, \omega_2)$  is associated to random dyadic grids  $\mathcal{D}_\omega = \mathcal{D}_{\omega_1}^n \times \mathcal{D}_{\omega_2}^m$ ,  $C_T \lesssim 1$ , the numbers  $\alpha_{k,v} > 0$  decay exponentially in complexity  $(k, v)$ , and  $U_\omega^{k,v}$  denotes some bilinear bi-parameter dyadic model operator of complexity  $(k, v)$  defined in the lattice  $\mathcal{D}_\omega$ .

The free of full paraproducts assumption that appears in some statements below implies precisely that  $T$  has a representation with shifts and partial paraproducts only. This requires that  $T(1, 1) = 0$  in the sense that  $\langle T(1, 1), h_I \otimes h_J \rangle = 0$  (where  $h_I$  is a cancellative Haar function on a cube  $I$ ) and the same for the adjoints and partial adjoints.

Multiplicators are paraproduct free CZOs (even the partial paraproducts vanish). An example of an operator that is free of full paraproducts but not necessarily of partial paraproducts is  $T_n \otimes T_m$ , where  $T_n$  is free of paraproducts. Regarding terminology it seems that continuous analogs of what we call shifts are often called paraproducts. For us a paraproduct is an operator involving BMO philosophies. The shifts are the simplest DMOs as their boundedness is based on the size of the individual coefficients only.

**1.3. Boundedness properties of  $T$ .** The following theorem states our estimates for CZOs.

**1.4. Theorem.** *Suppose  $T$  is a bilinear bi-parameter CZO. Then we have*

$$\|T(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}$$

for all  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfying  $1/p + 1/q = 1/r$ .

Suppose further that  $T$  is free of full paraproducts. Then we have the weighted estimate

$$\|T(f_1, f_2)\|_{L^r(v_3)} \leq C([w_1]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}, [w_2]_{A_q(\mathbb{R}^n \times \mathbb{R}^m)}) \|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}$$

for all  $1 < p, q < \infty$  and  $1/2 < r < \infty$  satisfying  $1/p + 1/q = 1/r$ , and for all bi-parameter weights  $w_1 \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ ,  $w_2 \in A_q(\mathbb{R}^n \times \mathbb{R}^m)$  with  $v_3 := w_1^{r/p} w_2^{r/q}$ . In the unweighted case we also have the mixed-norm estimates

$$\|T(f_1, f_2)\|_{L^{r_1}(\mathbb{R}^n; L^{r_2}(\mathbb{R}^m))} \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))} \|f_2\|_{L^{q_1}(\mathbb{R}^n; L^{q_2}(\mathbb{R}^m))}$$

for all  $1 < p_i, q_i \leq \infty$  and  $1/2 < r_i < \infty$  with  $1/p_i + 1/q_i = 1/r_i$ , except that if  $r_2 < 1$  we have to assume  $\infty \notin \{p_1, q_1\}$ .

**1.5. Remark.** In Remark 6.18 we show that if  $T = T_n \otimes T_m$  is a tensor product of bounded bilinear SIOs, the weighted bounds hold even without the paraproduct free assumption. In the new note [46] we improve these weighted estimates further in the tensor product case and also remove the restriction in the mixed-norm estimates using fancier extrapolation that allows to reach the exponent  $\infty$ .

The weighted estimates are first proved for the shifts and partial paraproducts, and then moved to full paraproduct free CZOs using the representation and the multilinear extrapolation by Grafakos–Martell (and Duoandikoetxea) [15, 22]. Extrapolation is needed as quasi-Banach estimates cannot be moved to  $T$  directly via the representation due to the presence of the averaging. The mixed-norm estimates follow from the weighted estimates and operator-valued analysis.

Weighted estimates for singular integrals in the linear bi-parameter setting are already quite difficult: see Fefferman–Stein [19] and Fefferman [16], [17]. Recently, Holmes–Petermichl–Wick [30] showed weighted bounds for bi-parameter singular integrals using the representation theorem [50]. We use this approach also here, but the nature of the bilinear weights poses a problem with the usage of duality (notice e.g. that if  $w_1, w_2 \in A_4$  then  $v_3 := w_1^{1/2} w_2^{1/2} \in A_4$  while we need to work in  $L^2(v_3)$ ). Therefore, the proof of the weighted boundedness of DMOs cannot proceed as in the linear case. This is a problem, as many bi-parameter proofs, especially those related to full paraproducts, are based on  $H^1$ -BMO type duality arguments. That is, the bi-parameter nature of the problem would favour duality based proofs, but the bilinear nature does not.

We manage to bound the shifts and partial paraproducts with duality free proofs with a careful usage of the  $A_\infty$  condition (so we only exploit the fact that  $v_3 \in A_\infty$ ). For example, we use the  $A_\infty$  extrapolation by Cruz-Uribe–Martell–Pérez [9] and known lower square function bounds valid for  $A_\infty$  weights. The weighted proof of the partial paraproducts also exploits sparse bounds of the one-parameter bilinear paraproducts.

Regarding the first part of the theorem, we still need to show unweighted bounds in the full range for (averages of) full paraproducts. Here we use a different set of tools than above: weak type arguments and interpolation. The general method is from the paper [51], but especially in the commutator setting, where we also use these weak type methods, the setup is much more complicated.

#### 1.4. Commutator estimates.

*Upper bounds.* We define the commutators

$$[b, T]_1(f_1, f_2) = bT(f_1, f_2) - T(bf_1, f_2) \text{ and } [b, T]_2(f_1, f_2) = bT(f_1, f_2) - T(f_1, bf_2).$$

A function  $b$  is in little BMO,  $\text{bmo}(\mathbb{R}^{n+m})$ , if  $b(\cdot, x_2)$  and  $b(x_1, \cdot)$  are uniformly in BMO. By studying DMOs we prove the following upper bound for commutators.

**1.6. Theorem.** *Let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ , and let  $b \in \text{bmo}(\mathbb{R}^{n+m})$ . Suppose  $T$  is a bilinear bi-parameter CZO. Then we have*

$$\|[b, T]_1(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|b\|_{\text{bmo}(\mathbb{R}^{n+m})} \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})},$$

and similarly for  $[b, T]_2$ . Analogous results hold for iterated commutators like  $[b_2, [b_1, T]_1]_2$ .

We discuss the proof of Theorem 1.6 later. The theory of commutator estimates is extremely vast and important. By bounding commutators of shifts Ou, Petermichl and Strouse [54] proved that  $[b, T]$  is  $L^2$  bounded, when  $T$  is a bi-parameter SIO as in [50] and free of paraproducts. This is the important base case for more complicated multi-parameter commutator estimates – i.e., the proof method using the DMOs lends itself to more complicated estimates. For example, suppose that  $T_1$  and  $T_2$  are paraproduct free linear bi-parameter singular integrals satisfying the assumptions of the representation theorem [50] in  $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$  and  $\mathbb{R}^{n_3} \times \mathbb{R}^{n_4}$  respectively. Then according to [54] we have

$$\begin{aligned} \|[T_1, [b, T_2]]f\|_{L^2(\prod_{i=1}^4 \mathbb{R}^{n_i})} &\lesssim \max \left( \sup_{x_2, x_4} \|b(\cdot, x_2, \cdot, x_4)\|_{\text{BMO}_{\text{prod}}}, \sup_{x_2, x_3} \|b(\cdot, x_2, x_3, \cdot)\|_{\text{BMO}_{\text{prod}}}, \right. \\ &\quad \left. \sup_{x_1, x_4} \|b(x_1, \cdot, \cdot, x_4)\|_{\text{BMO}_{\text{prod}}}, \sup_{x_1, x_3} \|b(x_1, \cdot, x_3, \cdot)\|_{\text{BMO}_{\text{prod}}} \right) \|f\|_{L^2(\prod_{i=1}^4 \mathbb{R}^{n_i})}. \end{aligned}$$

Related to this see also Dalenc–Ou [12]. The dyadic methods are also of crucial importance in many other deeper instances, like in the Bloom type estimates discussed below.

The paper [54] was eventually generalised to concern all bi-parameter singular integrals satisfying  $T1$  conditions by Holmes–Petermichl–Wick [30]. Importantly, they also prove *Bloom type two-weight bounds* for commutators of the particular form  $[b, T]$ . With a Bloom type inequality we understand the following. Given some operator  $A^b$ , the definition of which depends naturally on some function  $b$ , we seek for a two-weight estimate

$$\|A^b\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}(\nu)},$$

where  $p \in (1, \infty)$ ,  $\mu, \lambda \in A_p$ ,  $\nu := \mu^{1/p} \lambda^{-1/p}$ , and  $\text{BMO}(\nu)$  is some suitable weighted BMO space. Usually  $A^b$  is some commutator like  $[b, T]$  in [30].

Recently, in a sequel of the current paper [48] we gave an efficient proof of [30] exploiting our modified commutator decompositions from this paper (see below), and generalised the result of [30] to iterated commutators of the form  $[b_k, \dots [b_2, [b_1, T]] \dots]$ . This kind of iteration in the Bloom setting is not obvious with the appearance of weighted BMO spaces like  $\text{bmo}(\nu^{1/k})$  – see Lerner–Ombrosi–Rivera–Ríos [45] for the one-parameter setting and sparse domination. Related previous one parameter commutator results include Holmes–Lacey–Wick [28, 29] and Lerner–Ombrosi–Rivera–Ríos [44]. In the very recent paper [49] we also use the new commutator decomposition strategy of the current paper to generalise Dalenc–Ou [12] to the Bloom setting:

$$\|[T_n^1, [b, T_m^2]]\|_{L^p(\mu) \rightarrow L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \|b\|_{\text{BMO}_{\text{prod}}(\nu)}$$

for a bounded singular integral  $T_n$  in  $\mathbb{R}^n$  and a bounded singular integral  $T_m$  in  $\mathbb{R}^m$ . That is, we can even handle product BMO type commutators in the Bloom setting.

The linear bi-parameter results discussed so far rely on the bi-parameter representation theorem [50] and its multi-parameter generalisation by Y. Ou [53], and on increasingly sophisticated ways to bound the appearing model operators and their commutators. Similarly, we rely on our new bilinear bi-parameter representation theorem. However, in Theorem 1.6 we do not consider Bloom type theory – such bilinear bi-parameter Bloom theory is left as a very interesting question. In the one-parameter setting multilinear Bloom type results appear in Kunwar–Ou [39] (where sparse domination is used).

*Proof of Theorem 1.6 and commutator decompositions.* The main challenge in going from [54] to [30], apart from the Bloom setting, appeared to be that the various paraproducts include non-cancellative Haar functions and have a more complicated structure than the shifts. In [30] everything was reduced to a so called remainder term, which entailed always expanding  $bf$  in the bi-parameter sense. However, this remainder term has a particularly nice structure only when there are no non-cancellative Haar functions.

Our guideline is to expand  $bf$  using bi-parameter martingales in  $\langle bf, h_I \otimes h_J \rangle$ , using one-parameter martingales in  $\langle bf, h_I^0 \otimes h_J \rangle$  (or  $\langle bf, h_I \otimes h_J^0 \rangle$ ), and not to expand at all in  $\langle bf, h_I^0 \otimes h_J^0 \rangle$  (here  $h_I^0$  is a non-cancellative Haar function). Moreover, when a non-cancellative Haar function appears, a suitable average of  $b$  is added and subtracted. Working like this the proof of the Banach range boundedness of bilinear commutators is surprisingly short. For the quasi-Banach range we use weak type estimates as already mentioned, and this is quite delicate. Iterated commutators greatly benefit from the new clear decomposition.

*The Cauchy trick and its limitations.* The Cauchy integral trick by Coifman–Rochberg–Weiss [11] is a trick, where one e.g. defines  $F(z) = e^{bz}T(e^{-bz}f_1, f_2)$ , notices that formally  $F'(0) = [b, T]_1(f_1, f_2)$ , uses the Cauchy integral formula and then hopes to employ *weighted* bounds of  $T$  to get bounds for  $[b, T]_1$ . Weighted bounds for  $T$  are always needed as some weights of the form  $e^{\operatorname{Re}(bz)}$  appear in this argument. This trick is tied to the BMO properties of  $b$  by requiring exponential integrability as given by the classical inequality of John–Nirenberg. It still works in the bi-parameter setting if  $b$  is in little BMO, however.

The simplest reason why we do not use this trick is that we currently get weighted bounds for  $T$  only in the free of full paraproducts case, while we prove the commutator result for all  $T$ . Also, our proof of the weighted bound for the partial paraproducts utilises one-parameter sparse domination, and therefore does not work in three or higher parameters. Thus, we think that another proof for the boundedness of the partial paraproducts in the full range is of interest. While we do not show such proofs explicitly, we perform weak type arguments in this more complicated commutator setting from which they can easily be deduced.

The most important reason for developing and using the dyadic commutator estimates is that they can be used in many instances where the Cauchy trick fails. The trick does not work for the product BMO type commutators discussed in the linear situation above, and we are also currently investigating certain natural bilinear analogs of product BMO type commutators. Moreover, in the Bloom setting one has to directly prove the first order case no matter what, and in the most general little bmo type iterations  $b$  is not in

bmo (recall that the Cauchy trick relies on the classical BMO property of  $b$ ). So in the Bloom setting the Cauchy trick is not really useful even in the little BMO context.

*Lower bounds.* Lower bounds for commutators of particular singular integrals (like the Riesz transform) complement the theory of the upper bounds discussed above. For the history and interesting new results and methods see the recent paper by Hytönen [34].

In the multi-parameter setting the deepest lower bounds concern the product BMO type commutators. The Hilbert case  $[H_1, [b, H_2]]$  is considered in Ferguson–Lacey [20] – see also Lacey–Petermichl–Pipher–Wick [42, 43] for the Riesz setting. Ou–Petermichl–Strouse [54] also contains general bounds of such type. Notice that e.g. in [54] the proved lower bounds for Riesz transforms go through some well-chosen bi-parameter CZOs of not tensor product type – this is yet another example of the use of general theory.

Little BMO type lower bounds are not as difficult as the product BMO ones. In this paper we are dealing with bilinear little BMO type commutators. In the linear bi-parameter case at least [30] contains a little BMO type lower bound, even in the Bloom setting. Guo–Lian–Wu [27] proved lower bounds in the bilinear one-parameter setting – even for exponents in the quasi-Banach range. Previous bilinear results like Chaffee [2] worked only in the Banach range. However, we do not try to adapt any of these methods to the bilinear bi-parameter setting, since a bilinear bi-parameter version of the wonderful median method of Hytönen [34] and Lerner–Ombrosi–Rivera-Ríos [45] can be developed. This allows us to state the lower bounds for quite general non-degenerate kernels using very weak off-diagonal type testing conditions – see Section 10 for the statements.

Sections 2, 3 and 4 are devoted to definitions and setting the stage. In Section 5 we prove the representation theorem. Sections 6 and 7 contain the estimates for the DMOs and the CZOs, respectively. Sections 8 and 9 give the commutator upper bounds.

## 2. BASIC DEFINITIONS AND RESULTS

**2.1. Basic notation.** We denote  $A \lesssim B$  if  $A \leq CB$  for some constant  $C$  that can depend on the dimension of the underlying spaces, on integration exponents, and on various other constants appearing in the assumptions. We denote  $A \sim B$  if  $B \lesssim A \lesssim B$ .

We work in the bi-parameter setting in the product space  $\mathbb{R}^{n+m}$ . In such a context  $x = (x_1, x_2)$  with  $x_1 \in \mathbb{R}^n$  and  $x_2 \in \mathbb{R}^m$ . We often take integral pairings with respect to one of the two variables only: If  $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  and  $h: \mathbb{R}^n \rightarrow \mathbb{C}$ , then  $\langle f, h \rangle_1: \mathbb{R}^m \rightarrow \mathbb{C}$  is defined by  $\langle f, h \rangle_1(x_2) = \int_{\mathbb{R}^n} f(y_1, x_2)h(y_1) dy_1$ .

**2.2. Dyadic notation, Haar functions and martingale differences.** We denote a dyadic grid in  $\mathbb{R}^n$  by  $\mathcal{D}^n$  and a dyadic grid in  $\mathbb{R}^m$  by  $\mathcal{D}^m$ . The dyadic rectangles are denoted by  $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ . If  $I \in \mathcal{D}^n$ , then  $I^{(k)}$  denotes the unique dyadic cube  $S \in \mathcal{D}^n$  so that  $I \subset S$  and  $\ell(S) = 2^k \ell(I)$ . Here  $\ell(I)$  stands for side length. Also,  $\text{ch}(I)$  denotes the dyadic children of  $I$  – this means that  $I' \in \text{ch}(I)$  if  $(I')^{(1)} = I$ . The measure of a cube  $I$  is simply denoted by  $|I|$  no matter what dimension we are in.

When  $I \in \mathcal{D}^n$  we denote by  $h_I$  a cancellative  $L^2$  normalised Haar function. This means the following. Writing  $I = I_1 \times \dots \times I_n$  we can define the Haar function  $h_I^\eta$ ,  $\eta = (\eta_1, \dots, \eta_n) \in \{0, 1\}^n$ , by setting  $h_I^\eta = h_{I_1}^{\eta_1} \otimes \dots \otimes h_{I_n}^{\eta_n}$ , where  $h_{I_i}^0 = |I_i|^{-1/2} 1_{I_i}$  and  $h_{I_i}^1 = |I_i|^{-1/2} (1_{I_{i,l}} - 1_{I_{i,r}})$  for every  $i = 1, \dots, n$ . Here  $I_{i,l}$  and  $I_{i,r}$  are the left and right halves of the interval  $I_i$  respectively. If  $\eta \in \{0, 1\}^n \setminus \{0\}$  the Haar function is cancellative:  $\int h_I^\eta = 0$ . We usually suppress the presence of  $\eta$  and simply write  $h_I$  for some  $h_I^\eta$ ,  $\eta \in$

$\{0, 1\}^n \setminus \{0\}$ . Then  $h_I h_J$  can stand for  $h_I^{\eta_1} h_J^{\eta_2}$ , but we always treat such a product as a non-cancellative function.

For  $I \in \mathcal{D}^n$  and  $f \in L_{\text{loc}}^1(\mathbb{R}^n)$  we define the martingale difference

$$\Delta_I f = \sum_{I' \in \text{ch}(I)} [\langle f \rangle_{I'} - \langle f \rangle_I] 1_{I'}.$$

Here  $\langle f \rangle_I = \frac{1}{|I|} \int_I f$ . We also write  $E_I f = \langle f \rangle_I 1_I$ . Now, we have  $\Delta_I f = \sum_{\eta \neq 0} \langle f, h_I^\eta \rangle h_I^\eta$ , or suppressing the  $\eta$  summation,  $\Delta_I f = \langle f, h_I \rangle h_I$ , where  $\langle f, h_I \rangle = \int f h_I$ . A martingale block is defined by

$$\Delta_{K,i} f = \sum_{\substack{I \in \mathcal{D}^n \\ I^{(i)} = K}} \Delta_I f, \quad K \in \mathcal{D}^n.$$

Next, we define bi-parameter martingale differences. Let  $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{C}$  be locally integrable. Let  $I \in \mathcal{D}^n$  and  $J \in \mathcal{D}^m$ . We define the martingale difference

$$\Delta_I^1 f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}, \Delta_I^1 f(x) := \Delta_I(f(\cdot, x_2))(x_1).$$

Define  $\Delta_J^2 f$  analogously, and also define  $E_I^1$  and  $E_J^2$  similarly. We set

$$\Delta_{I \times J} f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}, \Delta_{I \times J} f(x) = \Delta_I^1(\Delta_J^2 f)(x) = \Delta_J^2(\Delta_I^1 f)(x).$$

Notice that  $\Delta_I^1 f = h_I \otimes \langle f, h_I \rangle_1$ ,  $\Delta_J^2 f = \langle f, h_J \rangle_2 \otimes h_J$  and  $\Delta_{I \times J} f = \langle f, h_I \otimes h_J \rangle h_I \otimes h_J$  (suppressing the finite  $\eta$  summations). Martingale blocks are defined in the natural way

$$\Delta_{K \times V}^{i,j} f = \sum_{I: I^{(i)} = K} \sum_{J: J^{(j)} = V} \Delta_{I \times J} f = \Delta_{K,i}^1(\Delta_{V,j}^2 f) = \Delta_{V,j}^2(\Delta_{K,i}^1 f).$$

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

$$[w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)} := \sup_R \langle w \rangle_R \langle w^{1-p'} \rangle_R^{p-1} < \infty,$$

where the supremum is taken over  $R = I \times J$ , where  $I \subset \mathbb{R}^n$  and  $J \subset \mathbb{R}^m$  are cubes with sides parallel to the axes (we simply call such  $R$  rectangles). We have

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

and that  $\max \left( \text{ess sup}_{x_1 \in \mathbb{R}^n} [w(x_1, \cdot)]_{A_p(\mathbb{R}^m)}, \text{ess sup}_{x_2 \in \mathbb{R}^m} [w(\cdot, x_2)]_{A_p(\mathbb{R}^n)} \right) \leq [w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}$ , while the constant  $[w]_{A_p}$  is dominated by the maximum to some power. Of course,  $A_p(\mathbb{R}^n)$  is defined similarly as  $A_p(\mathbb{R}^n \times \mathbb{R}^m)$  – just take the supremum over cubes  $Q$ . For the basic theory of bi-parameter weights consult e.g. [30].

Also, recall that  $w \in A_\infty(\mathbb{R}^n)$  if

$$[w]_{A_\infty(\mathbb{R}^n)} = \sup_Q \left( \frac{1}{|Q|} \int_Q w \right) \exp \left( \frac{1}{|Q|} \int_Q \log w^{-1} \right) < \infty,$$

where the supremum is taken over all the cubes  $Q \subset \mathbb{R}^n$ . We will only use the fact that  $A_p \subset A_\infty$ , and that certain key estimates hold for the larger class of  $A_\infty$  weights.

We record the following standard weighted square function estimates.

**2.1. Lemma.** For  $p \in (1, \infty)$  and  $w \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$  we have

$$\begin{aligned} \|f\|_{L^p(w)} &\sim_{[w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}} \left\| \left( \sum_{\substack{I \in \mathcal{D}^n \\ J \in \mathcal{D}^m}} |\Delta_{I \times J} f|^2 \right)^{1/2} \right\|_{L^p(w)} \\ &\sim_{[w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}} \left\| \left( \sum_{I \in \mathcal{D}^n} |\Delta_I^1 f|^2 \right)^{1/2} \right\|_{L^p(w)} \sim_{[w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}} \left\| \left( \sum_{J \in \mathcal{D}^m} |\Delta_J^2 f|^2 \right)^{1/2} \right\|_{L^p(w)}. \end{aligned}$$

**2.4. Maximal functions.** Given  $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  and  $g: \mathbb{R}^n \rightarrow \mathbb{C}$  we denote the dyadic maximal functions by

$$M_{\mathcal{D}^n} g := \sup_{I \in \mathcal{D}^n} \frac{1}{|I|} \int_I |g| \quad \text{and} \quad M_{\mathcal{D}} f := \sup_{R \in \mathcal{D}} \frac{1}{|R|} \int_R |f|.$$

The non-dyadic variants are simply denoted by  $M$ , as it is clear what is meant from the context. We also set  $M_{\mathcal{D}^n}^1 f(x_1, x_2) = M_{\mathcal{D}^n}(f(\cdot, x_2))(x_1)$ . The operator  $M_{\mathcal{D}^m}^2$  is defined similarly. For various maximal functions  $M$  we define  $M_s$  by setting  $M_s f = (M|f|^s)^{1/s}$ .

Standard weighted estimates involving maximal functions are recorded below.

**2.2. Lemma.** For  $p, s \in (1, \infty)$  and  $w \in A_p$  we have the Fefferman–Stein inequality

$$\left\| \left( \sum_j |M f_j|^s \right)^{1/s} \right\|_{L^p(w)} \leq C([w]_{A_p}) \left\| \left( \sum_j |f_j|^s \right)^{1/s} \right\|_{L^p(w)}.$$

We also have

$$\|\varphi_{\mathcal{D}^n}^1 f\|_{L^p(w)} \sim_{[w]_{A_p}} \left\| \left( \sum_{I \in \mathcal{D}^n} \frac{1}{|I|} \otimes [M \langle f, h_I \rangle_1]^2 \right)^{1/2} \right\|_{L^p(w)} \lesssim_{[w]_{A_p}} \|f\|_{L^p(w)},$$

where  $\varphi_{\mathcal{D}^n}^1 f := \sum_{I \in \mathcal{D}^n} h_I \otimes M \langle f, h_I \rangle_1$ . The function  $\varphi_{\mathcal{D}^m}^2 f$  is defined in the symmetric way and satisfies the same estimates.

**2.5. BMO spaces.** We say that a locally integrable function  $b: \mathbb{R}^n \rightarrow \mathbb{C}$  belongs to the dyadic BMO space  $\text{BMO}(\mathcal{D}^n)$  if

$$\|b\|_{\text{BMO}(\mathcal{D}^n)} := \sup_{I \in \mathcal{D}^n} \frac{1}{|I|} \int_I |b - \langle b \rangle_I| < \infty.$$

The ordinary space  $\text{BMO}(\mathbb{R}^n)$  is defined by taking the supremum over all cubes.

We say that a locally integrable function  $b: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  belongs to the dyadic little BMO space  $\text{bmo}(\mathcal{D})$ , where  $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ , if

$$\|b\|_{\text{bmo}(\mathcal{D})} := \sup_{R \in \mathcal{D}} \frac{1}{|R|} \int_R |b - \langle b \rangle_R| < \infty.$$

The non-dyadic space  $\text{bmo}(\mathbb{R}^{n+m})$  is defined in the natural way – take the supremum over all rectangles. We have

$$\|b\|_{\text{bmo}(\mathcal{D})} \sim \max \left( \text{ess sup}_{x_1 \in \mathbb{R}^n} \|b(x_1, \cdot)\|_{\text{BMO}(\mathcal{D}^m)}, \text{ess sup}_{x_2 \in \mathbb{R}^m} \|b(\cdot, x_2)\|_{\text{BMO}(\mathcal{D}^n)} \right).$$

We also have the John–Nirenberg property

$$\|b\|_{\text{bmo}(\mathbb{R}^{n+m})} \sim \sup_{R \subset \mathbb{R}^{n+m}} \left( \frac{1}{|R|} \int_R |b - \langle b \rangle_R|^p \right)^{1/p}, \quad 1 < p < \infty.$$

Moreover, we need to know that  $\text{bmo}(\mathbb{R}^{n+m}) \subset \text{BMO}_{\text{prod}}(\mathbb{R}^{n+m})$ . The reader can consult e.g. [30, 54]. A short proof of a weighted version of this inclusion is included in [48].

Finally, we have the product BMO space. Set

$$\|b\|_{\text{BMO}_{\text{prod}}(\mathcal{D})} := \sup_{\Omega} \left( \frac{1}{|\Omega|} \sum_{\substack{I \times J \in \mathcal{D} \\ I \times J \subset \Omega}} |\langle b, h_I \otimes h_J \rangle|^2 \right)^{1/2},$$

where the supremum is taken over those sets  $\Omega \subset \mathbb{R}^{n+m}$  such that  $|\Omega| < \infty$  and such that for every  $x \in \Omega$  there exist  $I \times J \in \mathcal{D}$  so that  $x \in I \times J \subset \Omega$ . The non-dyadic product BMO space  $\text{BMO}_{\text{prod}}(\mathbb{R}^{n+m})$  can be defined using the norm defined by the supremum over all dyadic grids of the above dyadic norms.

**2.6. Commutators.** We set

$$[b, T]_1(f_1, f_2) = bT(f_1, f_2) - T(bf_1, f_2) \quad \text{and} \quad [b, T]_2(f_1, f_2) = bT(f_1, f_2) - T(f_1, bf_2).$$

These are understood in a situation, where we e.g. know that  $T: L^3(\mathbb{R}^{n+m}) \times L^3(\mathbb{R}^{n+m}) \rightarrow L^{3/2}(\mathbb{R}^{n+m})$ , and  $b$  is locally in  $L^3$ . Then we study the case that  $f_1$  and  $f_2$  are, say, bounded and compactly supported, so that e.g.  $bf_2 \in L^3(\mathbb{R}^{n+m})$  and  $bT(f_1, f_2) \in L^1_{\text{loc}}(\mathbb{R}^{n+m})$ .

**2.7. Duality.** For a bilinear operator  $T$  acting on functions defined on the product space  $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$  we define the following duals

$$\begin{aligned} & \langle T(f_1 \otimes f_2, g_1 \otimes g_2), h_1 \otimes h_2 \rangle \\ &= \langle T^{1*}(h_1 \otimes h_2, g_1 \otimes g_2), f_1 \otimes f_2 \rangle = \langle T^{2*}(f_1 \otimes f_2, h_1 \otimes h_2), g_1 \otimes g_2 \rangle \\ &= \langle T_1^{1*}(h_1 \otimes f_2, g_1 \otimes g_2), f_1 \otimes h_2 \rangle = \langle T_1^{2*}(f_1 \otimes f_2, h_1 \otimes g_2), g_1 \otimes h_2 \rangle \\ &= \langle T_2^{1*}(f_1 \otimes h_2, g_1 \otimes g_2), h_1 \otimes f_2 \rangle = \langle T_2^{2*}(f_1 \otimes f_2, g_1 \otimes h_2), h_1 \otimes g_2 \rangle \\ &= \langle T_{1,2}^{1*,2*}(h_1 \otimes f_2, g_1 \otimes h_2), f_1 \otimes g_2 \rangle = \langle T_{2,1}^{1*,2*}(f_1 \otimes h_2, h_1 \otimes g_2), g_1 \otimes f_2 \rangle. \end{aligned}$$

One should understand that one cannot claim that the boundedness of  $T$  would e.g. imply the boundedness of  $T_1^{1*}$  – partial duals like this need not be bounded even if  $T$  is. This is a standard caveat of bi-parameter analysis.

### 3. BILINEAR BI-PARAMETER SIOS AND CZOS

We begin by formulating what it means for  $T$  to be a bilinear bi-parameter singular integral operator (SIO). Let  $\mathcal{F}_n$  consist of all the finite linear combinations of indicators of cubes  $I \subset \mathbb{R}^n$ , define  $\mathcal{F}_m$  analogously, and let  $\mathcal{F}$  consist of all the finite linear combinations of indicators of rectangles  $R \subset \mathbb{R}^{n+m}$ . Assume that we have bilinear operators  $T, T^{1*}, T^{2*}, T_1^{1*}, T_1^{2*}, T_2^{1*}, T_2^{2*}, T_{1,2}^{1*,2*}$  and  $T_{2,1}^{1*,2*}$ , each acting on tuples  $(f_1, f_2) \in \mathcal{F} \times \mathcal{F}$  and mapping them into locally integrable functions, and assume that we have the duality identities of Section 2.7 for all  $f_1, g_1, h_1 \in \mathcal{F}_n$  and  $f_2, g_2, h_2 \in \mathcal{F}_m$ .

**3.1. Kernel representations and SIOs.** Let  $f_i = f_i^1 \otimes f_i^2$ , where  $f_i^1 \in \mathcal{F}_n$  and  $f_i^2 \in \mathcal{F}_m$ .

3.1.1. *Full kernel representation.* Here we assume that  $\text{spt } f_i^1 \cap \text{spt } f_j^1 = \emptyset$  for some  $i, j$ , and  $\text{spt } f_{i'}^2 \cap \text{spt } f_{j'}^2 = \emptyset$  for some  $i', j'$ . In this case we demand that

$$\langle T(f_1, f_2), f_3 \rangle = \iint_{\mathbb{R}^{n+m}} \iint_{\mathbb{R}^{n+m}} \iint_{\mathbb{R}^{n+m}} K(x, y, z) f_1(y) f_2(z) f_3(x) dx dy dz,$$

where  $K: (\mathbb{R}^{n+m})^3 \setminus \{(x, y, z) \in (\mathbb{R}^{n+m})^3: x_1 = y_1 = z_1 \text{ or } x_2 = y_2 = z_2\} \rightarrow \mathbb{C}$  is a kernel satisfying a set of estimates which we specify next. Note that this implies kernel representations also for  $T^{1*}, T^{2*}, T_1^{1*}, T_1^{2*}, T_2^{1*}, T_2^{2*}, T_{1,2}^{1*,2*}$  and  $T_{2,1}^{1*,2*}$ . We denote their kernels by  $K^{1*}, K^{2*}$  and so forth. They all have an obvious formula using  $K$ .

Let  $\alpha \in (0, 1]$ . The kernel  $K$  is assumed to satisfy the size estimate

$$|K(x, y, z)| \leq C \frac{1}{(|x_1 - y_1| + |x_1 - z_1|)^{2n}} \frac{1}{(|x_2 - y_2| + |x_2 - z_2|)^{2m}},$$

the Hölder estimate

$$\begin{aligned} & |K(x, y, z) - K((x_1, x'_2), y, z) - K((x'_1, x_2), y, z) + K(x', y, z)| \\ & \leq C \frac{|x_1 - x'_1|^\alpha}{(|x_1 - y_1| + |x_1 - z_1|)^{2n+\alpha}} \frac{|x_2 - x'_2|^\alpha}{(|x_2 - y_2| + |x_2 - z_2|)^{2m+\alpha}} \end{aligned}$$

whenever  $|x_1 - x'_1| \leq \max(|x_1 - y_1|, |x_1 - z_1|)/2$  and  $|x_2 - x'_2| \leq \max(|x_2 - y_2|, |x_2 - z_2|)/2$ , and the mixed Hölder and size estimate

$$|K(x, y, z) - K((x'_1, x_2), y, z)| \leq C \frac{|x_1 - x'_1|^\alpha}{(|x_1 - y_1| + |x_1 - z_1|)^{2n+\alpha}} \frac{1}{(|x_2 - y_2| + |x_2 - z_2|)^{2m}}$$

whenever  $|x_1 - x'_1| \leq \max(|x_1 - y_1|, |x_1 - z_1|)/2$ . These estimates are also assumed from the kernels  $K^{1*}, K^{2*}$ , and so forth.

3.1.2. *Partial kernel representations.* Suppose that  $\text{spt } f_i^1 \cap \text{spt } f_j^1 = \emptyset$  for some  $i, j$ . Then we assume that

$$\langle T(f_1, f_2), f_3 \rangle = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} K_{f_1^2, f_2^2, f_3^2}(x_1, y_1, z_1) f_1^1(y_1) f_2^1(z_1) f_3^1(x_1) dx_1 dy_1 dz_1,$$

where  $K_{f_1^2, f_2^2, f_3^2}: \mathbb{R}^{3n} \setminus \{(x_1, y_1, z_1) \in \mathbb{R}^{3n}: x_1 = y_1 = z_1\} \rightarrow \mathbb{C}$  is a kernel satisfying the estimates of a standard bilinear Calderón-Zygmund kernel in  $\mathbb{R}^n$ , but with a constant depending on the functions  $f_1^2, f_2^2, f_3^2$ . This means that we have the size condition

$$|K_{f_1^2, f_2^2, f_3^2}(x_1, y_1, z_1)| \leq C(f_1^2, f_2^2, f_3^2) \frac{1}{(|x_1 - y_1| + |x_1 - z_1|)^{2n}}$$

and the Hölder estimate

$$|K_{f_1^2, f_2^2, f_3^2}(x_1, y_1, z_1) - K_{f_1^2, f_2^2, f_3^2}(x'_1, y_1, z_1)| \leq C(f_1^2, f_2^2, f_3^2) \frac{|x_1 - x'_1|^\alpha}{(|x_1 - y_1| + |x_1 - z_1|)^{2n+\alpha}}$$

whenever  $|x_1 - x'_1| \leq \max(|x_1 - y_1|, |x_1 - z_1|)/2$ . The analogous Hölder estimates in the  $y_1$  and  $z_1$  slots are also assumed.

We assume that

$$C(1_J, 1_J, 1_J) + C(a_J, 1_J, 1_J) + C(1_J, a_J, 1_J) + C(1_J, 1_J, a_J) \leq C|J|$$

for all cubes  $J \subset \mathbb{R}^m$  and all functions  $a_J \in \mathcal{F}_m$  satisfying  $a_J = 1_J a_J$ ,  $|a_J| \leq 1$  and  $\int a_J = 0$ . Analogous partial kernel representation is assumed when  $\text{spt } f_{i'}^2 \cap \text{spt } f_{j'}^2 = \emptyset$  for some  $i', j'$ .

**3.1. Definition.** If  $T$  is a bilinear operator with full and partial kernel structure as defined in Section 3.1.1 and Section 3.1.2, respectively, we call  $T$  a bilinear bi-parameter SIO.

**3.2. Reformulation of a bilinear bi-parameter SIO.** We present a vector-valued reformulation of our SIOs, which is needed in the proof of Proposition 3.6.

Suppose  $T$  is a bilinear bi-parameter SIO. We want to define for  $x_1, z_1, y_1 \in \mathbb{R}^n$  (such that we don't have  $x_1 = y_1 = z_1$ ) a bilinear SIO  $U_1(x_1, y_1, z_1)$  in  $\mathbb{R}^m$  so that we have the properties (1) and (2) below.

(1) If  $f_1, f_2, f_3: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  are functions in  $\mathcal{F}$  so that for some  $i, j$  we have  $\text{spt}_{\mathbb{R}^n} f_i \cap \text{spt}_{\mathbb{R}^n} f_j = \emptyset$ , then we have

$$\langle T(f_1, f_2), f_3 \rangle = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \langle U_1(x_1, y_1, z_1)(f_1(y_1, \cdot), f_2(z_1, \cdot)), f_3(x_1, \cdot) \rangle dx_1 dy_1 dz_1.$$

Here  $\text{spt}_{\mathbb{R}^n} f$  is the closure of those  $x_1$  for which there exists  $x_2$  so that  $f(x_1, x_2) \neq 0$ .

(2) For a bilinear SIO  $U$  in  $\mathbb{R}^m$  denote

$$\begin{aligned} \|U\| := & \|U\|_{\text{CZ}_\alpha} + \|U(1, 1)\|_{\text{BMO}(\mathbb{R}^m)} + \|U^{1*}(1, 1)\|_{\text{BMO}(\mathbb{R}^m)} \\ & + \|U^{2*}(1, 1)\|_{\text{BMO}(\mathbb{R}^m)} + \sup\{|\langle U(1_V, 1_V), 1_V \rangle| : V \subset \mathbb{R}^m \text{ cube}\}, \end{aligned}$$

where  $\|U\|_{\text{CZ}_\alpha}$  denotes the best constant in the bilinear size and  $\alpha$ -Hölder estimates (meaning standard bilinear kernel estimates like in Section 3.1.2). We demand the natural size and Hölder estimates in this norm, for instance that

$$\|U_1(x_1, y_1, z_1) - U_1(x'_1, y_1, z_1)\| \lesssim \frac{|x_1 - x'_1|^\alpha}{(|x_1 - y_1| + |x_1 - z_1|)^{2n+\alpha}}$$

whenever  $|x_1 - x'_1| \leq \max(|x_1 - y_1|, |x_1 - z_1|)/2$ . Of course, we also want to define  $U_2(x_2, y_2, z_2)$  (a bilinear SIO on  $\mathbb{R}^n$ ) analogously.

In the bi-parameter setting Journé [37] originally used an operator-valued formulation in this spirit. The proof that in the linear bi-parameter setting assumptions like in [50] with partial and full kernels imply operator-valued type assumptions like in [37] is by A. Grau de la Herrán [25]. This argument can be adapted with ease to our bilinear bi-parameter setting to give the desired  $U_1$  satisfying (1) and (2) above. We omit the details.

The conditions in (2) are only useful in conjunction with a quantitative bilinear  $T1$  theorem. A suitable version can at least be found in [47]. Let  $U$  be a bilinear SIO in  $\mathbb{R}^m$  with  $\|U\|_{\text{CZ}_\alpha} < \infty$ . Then we have

$$\|U\|_{L^p(\mathbb{R}^m) \times L^q(\mathbb{R}^m) \rightarrow L^r(\mathbb{R}^m)} \lesssim \|U\|$$

for all  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfying  $1/p + 1/q = 1/r$ . Moreover, we have  $\|U\|_{L^\infty(\mathbb{R}^m) \times L^\infty(\mathbb{R}^m) \rightarrow \text{BMO}(\mathbb{R}^m)} \lesssim \|U\|$  and  $\|U\|_{L^1(\mathbb{R}^m) \times L^1(\mathbb{R}^m) \rightarrow L^{1/2, \infty}(\mathbb{R}^m)} \lesssim \|U\|$ .

### 3.3. Boundedness and cancellation assumptions and CZOs.

*Weak boundedness assumptions.* We say that  $T$  satisfies the weak boundedness property if  $|\langle T(1_I \otimes 1_J, 1_I \otimes 1_J), 1_I \otimes 1_J \rangle| \leq C|I||J|$  for all cubes  $I \subset \mathbb{R}^n$  and  $J \subset \mathbb{R}^m$ .

*Diagonal BMO assumption.* SIO  $T$  satisfies the diagonal BMO assumption if the following holds. If  $I \subset \mathbb{R}^n$  and  $J \subset \mathbb{R}^m$  are cubes and  $a_I \in \mathcal{F}_n$  and  $a_J \in \mathcal{F}_m$  are functions satisfying  $a_I = 1_I a_I$ ,  $|a_I| \leq 1$  and  $\int a_I = 0$  and  $a_J = 1_J a_J$ ,  $|a_J| \leq 1$  and  $\int a_J = 0$ , then we have

$$\begin{aligned} & |\langle T(a_I \otimes 1_J, 1_I \otimes 1_J), 1_I \otimes 1_J \rangle| + |\langle T(1_I \otimes 1_J, a_I \otimes 1_J), 1_I \otimes 1_J \rangle| \\ & + |\langle T(1_I \otimes 1_J, 1_I \otimes 1_J), a_I \otimes 1_J \rangle| + |\langle T(1_I \otimes a_J, 1_I \otimes 1_J), 1_I \otimes 1_J \rangle| \\ & + |\langle T(1_I \otimes 1_J, 1_I \otimes a_J), 1_I \otimes 1_J \rangle| + |\langle T(1_I \otimes 1_J, 1_I \otimes 1_J), 1_I \otimes a_J \rangle| \leq C|I||J|. \end{aligned}$$

*Product BMO assumption.* SIO  $T$  satisfies the product BMO assumption if it holds  $S(1, 1) \in \text{BMO}_{\text{prod}}(\mathbb{R}^{n+m})$  for all the nine choices  $S \in \{T, T^{1*}, T^{2*}, T_1^{1*}, T_1^{2*}, T_2^{1*}, T_2^{2*}, T_{1,2}^{1*,2*}, T_{2,1}^{1*,2*}\}$ .

**3.2. Remark.** Recall that  $T(1, 1) \in \text{BMO}_{\text{prod}}(\mathbb{R}^{n+m})$  means that  $\|T(1, 1)\|_{\text{BMO}(\mathcal{D})} \lesssim 1$  for all  $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ . To make sense of this, we need only to define the pairings  $\langle T(1, 1), h_I \otimes h_J \rangle$ . This can be done as follows. It is by definition the sum of the following nine terms:

$$(3.3) \quad \begin{aligned} & \langle T(1_{(3I)^c} \otimes 1_{(3J)^c}, 1 \otimes 1), h_I \otimes h_J \rangle, \quad \langle T(1_{(3I)^c} \otimes 1_{3J}, 1 \otimes 1_{(3J)^c}), h_I \otimes h_J \rangle, \\ & \langle T(1_{3I} \otimes 1_{(3J)^c}, 1_{(3I)^c} \otimes 1), h_I \otimes h_J \rangle, \quad \langle T(1_{3I} \otimes 1_{3J}, 1_{(3I)^c} \otimes 1_{(3J)^c}), h_I \otimes h_J \rangle, \end{aligned}$$

$$(3.4) \quad \begin{aligned} & \langle T(1_{(3I)^c} \otimes 1_{3J}, 1 \otimes 1_{3J}), h_I \otimes h_J \rangle, \quad \langle T(1_{3I} \otimes 1_{3J}, 1_{(3I)^c} \otimes 1_{3J}), h_I \otimes h_J \rangle, \\ & \langle T(1_{3I} \otimes 1_{(3J)^c}, 1_{3I} \otimes 1), h_I \otimes h_J \rangle, \quad \langle T(1_{3I} \otimes 1_{3J}, 1_{3I} \otimes 1_{(3J)^c}), h_I \otimes h_J \rangle, \end{aligned}$$

and  $\langle T(1_{3I} \otimes 1_{3J}, 1_{3I} \otimes 1_{3J}), h_I \otimes h_J \rangle$ . However, only the last term is defined from the get go. The terms in (3.3) can be defined in a natural way using the full kernel of  $T$  and the terms in (3.4) via the partial kernels.

**3.5. Definition.** A bilinear bi-parameter SIO  $T$  satisfying the weak boundedness property, the diagonal BMO assumption and the product BMO assumption is called a bilinear bi-parameter Calderón–Zygmund operator (CZO).

The following proposition can be shown quite similarly to the linear case – for a modern proof in the linear setting see both [35] and [25]. The version below requires the known John–Nirenberg for dyadic product BMO. We omit the details.

**3.6. Proposition.** *If  $T$  is a bilinear bi-parameter SIO and for some  $r > 1$  we have*

$$\int_{\Omega} |T(1_{\Omega}, 1_{\Omega})|^r \lesssim |\Omega|$$

*for all finite unions of rectangles  $\Omega$ , then  $T(1, 1) \in \text{BMO}_{\text{prod}}(\mathbb{R}^{n+m})$ .*

This is key for the statements made in the Introduction, where boundedness in one tuple (for  $T$  and all its adjoints) implies boundedness in the full range.

#### 4. DMOs: SHIFTS, PARTIAL PARAPRODUCTS AND FULL PARAPRODUCTS

We define the dyadic model operators (DMOs) that appear in the bilinear bi-parameter representation theorem. Let  $f_1, f_2: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  be two given functions. The operators are here defined in some fixed grids  $\mathcal{D}^n$  and  $\mathcal{D}^m$ .

**4.1. Bilinear bi-parameter shifts.** For triples of integers  $k = (k_1, k_2, k_3)$ ,  $k_1, k_2, k_3 \geq 0$ , and  $v = (v_1, v_2, v_3)$ ,  $v_1, v_2, v_3 \geq 0$ , and cubes  $K \in \mathcal{D}^n$  and  $V \in \mathcal{D}^m$ , define

$$A_{K,k}^{V,v}(f_1, f_2) = \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_1^{(k_1)} = I_2^{(k_2)} = I_3^{(k_3)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}^m \\ J_1^{(v_1)} = J_2^{(v_2)} = J_3^{(v_3)} = V}} a_{K,V,(I_i),(J_j)} \\ \times \langle f_1, h_{I_1} \otimes h_{J_1} \rangle \langle f_2, h_{I_2} \otimes h_{J_2} \rangle h_{I_3}^0 \otimes h_{J_3}^0.$$

We also demand that the scalars  $a_{K,V,(I_i),(J_j)}$  satisfy the estimate

$$|a_{K,V,(I_i),(J_j)}| \leq \frac{|I_1|^{1/2} |I_2|^{1/2} |I_3|^{1/2} |J_1|^{1/2} |J_2|^{1/2} |J_3|^{1/2}}{|K|^2 |V|^2}.$$

A shift of complexity  $(k, v)$  of a particular form (the non-cancellative Haar functions are in certain positions) is

$$S_k^v(f_1, f_2) = \sum_{K \in \mathcal{D}^n} \sum_{V \in \mathcal{D}^m} A_{K,k}^{V,v}(f_1, f_2).$$

An operator of the above form, but having the non-cancellative Haar functions  $h_I^0$  and  $h_J^0$  in some of the other slots, is also a shift. So there are shifts of nine different types, and we could e.g. also have for all  $K, V$  that

$$A_{K,k}^{V,v}(f_1, f_2) = \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_1^{(k_1)} = I_2^{(k_2)} = I_3^{(k_3)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}^m \\ J_1^{(v_1)} = J_2^{(v_2)} = J_3^{(v_3)} = V}} a_{K,V,(I_i),(J_j)} \\ \times \langle f_1, h_{I_1}^0 \otimes h_{J_1} \rangle \langle f_2, h_{I_2} \otimes h_{J_2} \rangle h_{I_3} \otimes h_{J_3}^0.$$

**4.2. Bilinear bi-parameter partial paraproducts.** First, we need bilinear one-parameter paraproducts. Let  $b: \mathbb{R}^m \rightarrow \mathbb{C}$  be a function and define

$$A_b(g_1, g_2) := \sum_{V \in \mathcal{D}^m} \langle b, h_V \rangle \langle g_1 \rangle_V \langle g_2 \rangle_V h_V,$$

where  $g_i: \mathbb{R}^m \rightarrow \mathbb{C}$ . An operator  $\pi_b$  is called a dyadic bilinear paraproduct in  $\mathbb{R}^m$  if it is of the form  $A_b$ ,  $A_b^{1*}$  or  $A_b^{2*}$ .

Let  $k = (k_1, k_2, k_3)$ ,  $k_1, k_2, k_3 \geq 0$ . For each  $K, I_1, I_2, I_3 \in \mathcal{D}^n$  we are given a function  $b_{K,I_1,I_2,I_3}: \mathbb{R}^m \rightarrow \mathbb{C}$  such that  $\|b_{K,I_1,I_2,I_3}\|_{\text{BMO}(\mathbb{R}^m)} \leq \frac{|I_1|^{1/2} |I_2|^{1/2} |I_3|^{1/2}}{|K|^2}$ . A partial paraproduct of complexity  $k$  of a particular form is

$$P_k(f_1, f_2) = \sum_{K \in \mathcal{D}^n} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_1^{(k_1)} = I_2^{(k_2)} = I_3^{(k_3)} = K}} h_{I_3}^0 \otimes \pi_{b_{K,I_1,I_2,I_3}}(\langle f_1, h_{I_1} \rangle_1, \langle f_2, h_{I_2} \rangle_1),$$

where  $\pi_{b_{K,I_1,I_2,I_3}}$  denotes a bilinear paraproduct in  $\mathbb{R}^m$ , and is of the same form for all  $K, I_1, I_2, I_3$ . Again, an operator of the above form, but having the non-cancellative Haar function  $h_I^0$  in some other slot, is also a partial paraproduct. Therefore, we have nine different possibilities as usual (the bilinear paraproducts can be of one of the three different types, and the non-cancellative Haar function in  $\mathbb{R}^n$  can appear in one of the three slots).

We also have partial paraproducts with shift structure in  $\mathbb{R}^m$  and paraproducts in  $\mathbb{R}^n$ .

**4.3. Bilinear bi-parameter full paraproducts.** Let  $b: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  be a function with  $\|b\|_{\text{BMO}_{\text{prod}}(\mathbb{R}^{n+m})} \leq 1$ . A full paraproduct  $\Pi_b$  of a particular form is

$$\Pi_b(f_1, f_2) = \Pi(f_1, f_2) = \sum_{\substack{K \in \mathcal{D}^n \\ V \in \mathcal{D}^m}} \lambda_{K,V}^b \langle f_1 \rangle_{K \times V} \langle f_2 \rangle_{K \times V} h_K \otimes h_V,$$

where the function  $b$  determines the coefficients  $\lambda_{K,V}^b$  via the formula  $\lambda_{K,V}^b := \langle b, h_K \times h_V \rangle$ . Again, an operator of the above form, but having the cancellative Haar functions  $h_K$  or  $h_V$  in some other slots, is also a full paraproduct. There are nine different cases as the Haar functions present in the coefficients  $\lambda_{K,V}^b$  are not allowed to move, i.e. we always have  $\lambda_{K,V}^b := \langle b, h_K \times h_V \rangle$ . For example,  $\Pi_b$  could also be of the form

$$\Pi_b(f_1, f_2) = \sum_{\substack{K \in \mathcal{D}^n \\ V \in \mathcal{D}^m}} \lambda_{K,V}^b \langle f_1 \rangle_{K \times V} \left\langle f_2, \frac{1_K}{|K|} \otimes h_V \right\rangle h_K \otimes \frac{1_V}{|V|}.$$

## 5. THE REPRESENTATION THEOREM

We need the following notation regarding random dyadic grids. Let  $\mathcal{D}_0^n$  and  $\mathcal{D}_0^m$  denote the standard dyadic grids on  $\mathbb{R}^n$  and  $\mathbb{R}^m$  respectively. For  $\omega_1 = (\omega_1^i)_{i \in \mathbb{Z}} \in (\{0, 1\}^n)^{\mathbb{Z}}$ ,  $\omega_2 = (\omega_2^i)_{i \in \mathbb{Z}} \in (\{0, 1\}^m)^{\mathbb{Z}}$ ,  $I \in \mathcal{D}_0^n$  and  $J \in \mathcal{D}_0^m$  denote

$$I + \omega_1 := I + \sum_{i: 2^{-i} < \ell(I)} 2^{-i} \omega_1^i \quad \text{and} \quad J + \omega_2 := J + \sum_{i: 2^{-i} < \ell(J)} 2^{-i} \omega_2^i.$$

Then we define the random lattices

$$\mathcal{D}_{\omega_1}^n = \{I + \omega_1 : I \in \mathcal{D}_0^n\} \quad \text{and} \quad \mathcal{D}_{\omega_2}^m = \{J + \omega_2 : J \in \mathcal{D}_0^m\}.$$

There is a natural probability product measure  $\mathbb{P}_{\omega_1}$  in  $(\{0, 1\}^n)^{\mathbb{Z}}$  and  $\mathbb{P}_{\omega_2}$  in  $(\{0, 1\}^m)^{\mathbb{Z}}$ . We set  $\omega = (\omega_1, \omega_2) \in (\{0, 1\}^n)^{\mathbb{Z}} \times (\{0, 1\}^m)^{\mathbb{Z}}$ , and denote the expectation over the product probability space by  $\mathbb{E}_{\omega} = \mathbb{E}_{\omega_1, \omega_2} = \mathbb{E}_{\omega_1} \mathbb{E}_{\omega_2} = \mathbb{E}_{\omega_2} \mathbb{E}_{\omega_1} = \iint d\mathbb{P}_{\omega_1} d\mathbb{P}_{\omega_2}$ .

We also set  $\mathcal{D}_0 = \mathcal{D}_0^n \times \mathcal{D}_0^m$ . Given  $\omega = (\omega_1, \omega_2)$  and  $R = I \times J \in \mathcal{D}_0$  we may set

$$R + \omega = (I + \omega_1) \times (J + \omega_2) \quad \text{and} \quad \mathcal{D}_{\omega} = \{R + \omega : R \in \mathcal{D}_0\}.$$

**5.1. Theorem.** *Suppose  $T$  is a bilinear bi-parameter CZO. Then  $T$  can be extended to act on bounded and compactly supported functions  $f_i: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  (or e.g.  $L^3(\mathbb{R}^{n+m})$  functions), and we have*

$$\langle T(f_1, f_2), f_3 \rangle = C_T \mathbb{E}_{\omega} \sum_{\substack{k=(k_1, k_2, k_3) \in \mathbb{Z}_+^3 \\ v=(v_1, v_2, v_3) \in \mathbb{Z}_+^3}} \alpha_{k,v} \sum_u \langle U_{k,u, \mathcal{D}_{\omega}}^v(f_1, f_2), f_3 \rangle,$$

where  $C_T \lesssim 1$ ,  $\alpha_{k,v} = 2^{-\alpha \max k_i / 2} 2^{-\alpha \max v_j / 2}$ , the summation over  $u$  is finite, and  $U_{k,u, \mathcal{D}_{\omega}}^v$  is always either a shift of complexity  $(k, v)$ , a partial paraproduct of complexity  $k$  or  $v$  (this requires  $k = 0$  or  $v = 0$ ) or a full paraproduct (this requires  $k = v = 0$ ) associated with some product BMO function  $b$  satisfying  $\|b\|_{\text{BMO}_{\text{prod}}} \leq 1$ .

*Proof.* We show the proof under the additional assumption that  $T$  is a priori bounded, say from  $L^3 \times L^3 \rightarrow L^{3/2}$ . At the end we comment why this is enough.

We show that given  $N$  we have for all bounded  $f_i$  with  $\text{spt } f_i \subset B(0, c_0 2^N)$  that

$$\langle T(f_1, f_2), f_3 \rangle = \lim_{M \rightarrow \infty} C_T \mathbb{E}_\omega \sum_{\substack{k=(k_1, k_2, k_3) \in \mathbb{Z}_+^3 \\ v=(v_1, v_2, v_3) \in \mathbb{Z}_+^3}} \alpha_{k,v} \sum_u \langle U_{k,u, \mathcal{D}_\omega}^{v, N, M}(f_1, f_2), f_3 \rangle,$$

where  $U_{k,u, \mathcal{D}_\omega}^{v, N, M}$  are dyadic model operators so that all the appearing cubes  $I_1, I_2, I_3$  and  $J_1, J_2, J_3$  satisfy  $2^{-M} \leq \ell(I_i), \ell(J_j) \leq 2^N$ , and we allow non-cancellative Haar functions in all slots on the top level  $2^N$  except for the Haar functions related to paraproducts  $\pi$  (as part of a partial paraproduct) or  $\Pi$  (a full paraproduct). A simple limiting argument, which we omit, gives the desired conclusion.

To get started fix  $N$  and three bounded functions  $f_1, f_2$  and  $f_3$  with  $\text{spt } f_i \subset B(0, c_0 2^N)$ . Define

$$E_{2^{-M_1}, 2^{-M_2}}^\omega f_1 = \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_1) = 2^{-M_1}}} \sum_{\substack{J_1 \in \mathcal{D}_{\omega_2}^m \\ \ell(J_1) = 2^{-M_2}}} \langle f_1 \rangle_{I_1 \times J_1} 1_{I_1 \times J_1},$$

and also write  $E_{2^{-M}}^\omega f_1 = E_{2^{-M}, 2^{-M}}^\omega f_1$ . Using the boundedness of  $T$  we have

$$\langle T(f_1, f_2), f_3 \rangle = \lim_{M \rightarrow \infty} \langle T(E_{2^{-M}}^\omega f_1, E_{2^{-M}}^\omega f_2), E_{2^{-M}}^\omega f_3 \rangle.$$

By dominated convergence we further have that

$$\langle T(f_1, f_2), f_3 \rangle = \lim_{M \rightarrow \infty} \mathbb{E}_\omega \langle T(E_{2^{-M}}^\omega f_1, E_{2^{-M}}^\omega f_2), E_{2^{-M}}^\omega f_3 \rangle.$$

We fix  $M$  for now.

Define  $\Delta_I^{\leq N} = \Delta_I$  if  $\ell(I) < 2^N$ ,  $\Delta_I^{\leq N} = \Delta_I + E_I$  if  $\ell(I) = 2^N$  and  $\Delta_I^{\leq N} = 0$  if  $\ell(I) > 2^N$ . As  $N$  is fixed for the whole argument, we will be abusing notation and writing  $\Delta_I^1 = \Delta_I^{1, \leq N}$ ,  $\Delta_J^2 = \Delta_J^{2, \leq N}$  and  $\Delta_{I \times J} = \Delta_I^{1, \leq N} \Delta_J^{2, \leq N} = \Delta_I^1 \Delta_J^2$ . Similarly,  $h_I$  can stand for  $h_I$  or  $h_I^0$  if  $\ell(I) = 2^N$ . It is only inside this proof that this can happen.

Now, with this notation we have with any  $M_1, M_2$  such that  $2^{-M_i} < 2^N$  that

$$(5.2) \quad E_{2^{-M_1}, 2^{-M_2}}^\omega f_1 = \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n, J_1 \in \mathcal{D}_{\omega_2}^m \\ 2^{-M_1} < \ell(I_1) \leq 2^N \\ 2^{-M_2} < \ell(J_1) \leq 2^N}} \Delta_{I_1 \times J_1} f_1.$$

Using this with  $M_1 = M_2 = M$  we have for all large  $M$  that

$$\begin{aligned} & \langle T(E_{2^{-M}}^\omega f_1, E_{2^{-M}}^\omega f_2), E_{2^{-M}}^\omega f_3 \rangle \\ &= \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ 2^{-M} < \ell(I_i) \leq 2^N}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}_{\omega_2}^m \\ 2^{-M} < \ell(J_i) \leq 2^N}} \langle T(\Delta_{I_1 \times J_1} f_1, \Delta_{I_2 \times J_2} f_2), \Delta_{I_3 \times J_3} f_3 \rangle. \end{aligned}$$

Notice that the sums are finite as the functions have compact support. This makes all the reorganisations of sums and interchanging  $\mathbb{E}_\omega$  and sums readily true. If the reader is already wondering about the reason for the exact support condition  $\text{spt } f_i \subset B(0, c_0 2^N)$ , this is related to the fact that we want to ensure that the smallest cubes are strictly smaller than  $2^N$  in the so called separated sums below. This will become clear at the end.

We split

$$\begin{aligned}
 \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ 2^{-M} < \ell(I_i) \leq 2^N}} &= \sum_{\substack{I_3 \in \mathcal{D}_{\omega_1}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_3) \leq \ell(I_1) \leq 2^N}} \sum_{\substack{I_2 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_3) \leq \ell(I_2) \leq 2^N}} \\
 &+ \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n \\ 2^{-M} < \ell(I_1) < 2^N}} \sum_{\substack{I_2 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_1) \leq \ell(I_2) \leq 2^N}} \sum_{\substack{I_3 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_1) < \ell(I_3) \leq 2^N}} \\
 &+ \sum_{\substack{I_2 \in \mathcal{D}_{\omega_1}^n \\ 2^{-M} < \ell(I_2) < 2^N}} \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_2) < \ell(I_1) \leq 2^N}} \sum_{\substack{I_3 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_2) < \ell(I_3) \leq 2^N}}.
 \end{aligned}$$

We split the  $J_1, J_2, J_3$  sum similarly. This gives us a splitting

$$\langle T(E_{2^{-M}}^\omega f_1, E_{2^{-M}}^\omega f_2), E_{2^{-M}}^\omega f_3 \rangle = \sum_{i=1}^9 \Sigma^i(\omega),$$

where the enumeration is so that we first multiplied the first  $I$  summation with all the three possible  $J$  summations to get  $\Sigma^1, \Sigma^2, \Sigma^3$ , and so on. We consider  $N$  and  $M$  fixed at this point of the argument, and therefore just write  $\Sigma^i$  instead of  $\Sigma_{N,M}^i$ . For example,  $\Sigma^1(\omega)$  equals

$$\begin{aligned}
 \sum_{\substack{I_3 \in \mathcal{D}_{\omega_1}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_3) \leq \ell(I_1) \leq 2^N}} \sum_{\substack{I_2 \in \mathcal{D}_{\omega_1}^n \\ \ell(I_3) \leq \ell(I_2) \leq 2^N}} \\
 \sum_{\substack{J_3 \in \mathcal{D}_{\omega_2}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \sum_{\substack{J_1 \in \mathcal{D}_{\omega_2}^m \\ \ell(J_3) \leq \ell(J_1) \leq 2^N}} \sum_{\substack{J_2 \in \mathcal{D}_{\omega_2}^m \\ \ell(J_3) \leq \ell(J_2) \leq 2^N}} \langle T(\Delta_{I_1 \times J_1} f_1, \Delta_{I_2 \times J_2} f_2), \Delta_{I_3 \times J_3} f_3 \rangle.
 \end{aligned}$$

These are the nine main symmetries, which are handled relatively similarly. We choose to deal with  $\Sigma^1(\omega)$ .

*Adding goodness by a collapse trick.* A cube  $I \in \mathcal{D}_{\omega_1}^n$  is called bad if there exists such a cube  $P \in \mathcal{D}_{\omega_1}^n$  that  $\ell(P) \geq 2^r \ell(I)$  and

$$d(I, \partial P) \leq \ell(I)^{\gamma_n} \ell(P)^{1-\gamma_n}.$$

Here  $\gamma_n = \alpha / (2[2n + \alpha])$ , where  $\alpha > 0$  appears in the kernel estimates. Otherwise a cube is called good. We note that  $\pi_{\text{good},n} := \mathbb{P}_{\omega_1}(I + \omega_1 \text{ is good})$  is independent of the choice of  $I \in \mathcal{D}_0^n$ . The appearing parameter  $r$  is a large enough fixed constant so that  $\pi_{\text{good},n} > 0$  (and also  $\pi_{\text{good},m} > 0$ ). Moreover, for a fixed  $I \in \mathcal{D}_0^n$  the set  $I + \omega_1$  depends on  $\omega_1^i$  with  $2^{-i} < \ell(I)$ , while the goodness of  $I + \omega_1$  depends on  $\omega_1^i$  with  $2^{-i} \geq \ell(I)$ . These notions are independent by the product probability structure.

The way of adding goodness like below is from Li–Martikainen–Ou–Vuorinen [47] – it is a significantly cleaner way to add goodness to the smallest cubes than the one originally used by Hytönen [31]. In fact, in this generality the original way would probably cause a tedious mess.

We want the smallest cubes (i.e.  $I_3$  and  $J_3$ ) to be good. This can be achieved with the averaged sum  $\mathbb{E}_\omega \Sigma^1(\omega)$ . The following argument requires us to collapse certain sums at

this point to gain enough independence. Therefore, we write

$$\Sigma^1(\omega) = \sum_{\substack{I_3 \in \mathcal{D}_{\omega_1}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{\substack{J_3 \in \mathcal{D}_{\omega_2}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \langle T(E_{\ell(I_3)/2, \ell(J_3)/2}^\omega f_1, E_{\ell(I_3)/2, \ell(J_3)/2}^\omega f_2), \Delta_{I_3 \times J_3} f_3 \rangle,$$

where we used (5.2) with  $2^{-M_1} = \ell(I_3)/2$  and  $2^{-M_2} = \ell(J_3)/2$ .

Now the whole pairing depends only on  $\omega_1^i$  and  $\omega_2^j$  for  $2^{-i} < \ell(I_3)$  and  $2^{-j} < \ell(J_3)$ . However, the goodness of  $I_3$  depends on  $\omega_1^i$  for  $2^{-i} \geq \ell(I_3)$  (and similarly for  $J_3$ ). Let

$$\Sigma_{\text{good}}^1(\omega) = \sum_{\substack{I_3 \in \mathcal{D}_{\omega_1, \text{good}}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{\substack{J_3 \in \mathcal{D}_{\omega_2, \text{good}}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \langle T(E_{\ell(I_3)/2, \ell(J_3)/2}^\omega f_1, E_{\ell(I_3)/2, \ell(J_3)/2}^\omega f_2), \Delta_{I_3 \times J_3} f_3 \rangle.$$

By independence we get  $\mathbb{E}_\omega \Sigma^1(\omega) = C_1 \mathbb{E}_\omega \Sigma_{\text{good}}^1(\omega)$  for  $C_1 = (\pi_{\text{good}, n} \pi_{\text{good}, m})^{-1} < \infty$ .

We now fix an arbitrary  $\omega = (\omega_1, \omega_2)$  for the rest of the argument, and study  $\Sigma^1 := \Sigma_{\text{good}}^1(\omega)$ . We will also be writing  $\mathcal{D}^n$  instead of  $\mathcal{D}_{\omega_1}^n$ , and so on. Next, we re-expand the functions  $E_{\ell(I_3)/2, \ell(J_3)/2} f_1$  and  $E_{\ell(I_3)/2, \ell(J_3)/2} f_2$  to see that  $\Sigma^1$  equals

$$\sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{\substack{I_1 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_1) \leq 2^N}} \sum_{\substack{I_2 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_2) \leq 2^N}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \sum_{\substack{J_1 \in \mathcal{D}^m \\ \ell(J_3) \leq \ell(J_1) \leq 2^N}} \sum_{\substack{J_2 \in \mathcal{D}^m \\ \ell(J_3) \leq \ell(J_2) \leq 2^N}} \langle T(\Delta_{I_1 \times J_1} f_1, \Delta_{I_2 \times J_2} f_2), \Delta_{I_3 \times J_3} f_3 \rangle.$$

*Bilinear collapse argument.* The double sum  $\sum_{\ell(I_3) \leq \ell(I_1) \leq 2^N} \sum_{\ell(I_3) \leq \ell(I_2) \leq 2^N}$  equals

$$\sum_{\substack{I_1 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_1) \leq 2^N}} \sum_{\substack{I_2 \in \mathcal{D}^n \\ \ell(I_1) \leq \ell(I_2) \leq 2^N}} + \sum_{\substack{I_2 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_2) < 2^N}} \sum_{\substack{I_1 \in \mathcal{D}^n \\ \ell(I_2) < \ell(I_1) \leq 2^N}}.$$

A similar decomposition is performed on the  $J_1, J_2$  summations. This decomposes  $\Sigma^1 = \sum_{i=1}^4 \sigma_1^i$ , where the enumeration is again so that we first multiplied the first  $I$  summation with the two possible  $J$  summations to get  $\sigma_1^1, \sigma_1^2$ . We can now collapse sums in all of the terms  $\sigma_1^i$ , which is useful. The downside is that we have the main symmetries  $\Sigma^1, \dots, \Sigma^9$ , but splitting them further like here causes the following phenomena: whenever a paraproduct like term appears in some  $\sigma_k^i$ , it has to be combined (i.e. summed up) with the analogous terms from certain  $\sigma_k^j$  in order to get a simple paraproduct.

We now collapse the innermost  $I$  and  $J$  summations in all of the terms  $\sigma_1^i$ . This gives

$$\sigma_1^1 = \sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{\substack{I_1, I_2 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_1) = 2\ell(I_2) \leq 2^N}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \sum_{\substack{J_1, J_2 \in \mathcal{D}^m \\ \ell(J_3) \leq \ell(J_1) = 2\ell(J_2) \leq 2^N}} \langle T(\Delta_{I_1 \times J_1} f_1, E_{I_2 \times J_2} f_2), \Delta_{I_3 \times J_3} f_3 \rangle;$$

$$\sigma_1^2 = \sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{I_1, I_2 \in \mathcal{D}^n} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \sum_{\substack{J_1, J_2 \in \mathcal{D}^m \\ \ell(J_3) \leq \ell(J_1) = \ell(J_2) < 2^N}} \langle T(E_{J_1}^2 \Delta_{I_1}^1 f_1, E_{J_2}^1 \Delta_{J_2}^2 f_2), \Delta_{I_3 \times J_3} f_3 \rangle;$$

$$\sigma_1^3 = \sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{I_1, I_2 \in \mathcal{D}^n} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \sum_{\substack{J_1, J_2 \in \mathcal{D}^m \\ \ell(J_3) \leq \ell(J_1) = 2\ell(J_2) \leq 2^N}} \langle T(E_{I_1}^1 \Delta_{J_1}^2 f_1, E_{J_2}^2 \Delta_{I_2}^1 f_2), \Delta_{I_3 \times J_3} f_3 \rangle;$$

$$\sigma_1^4 = \sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) \leq 2^N}} \sum_{I_1, I_2 \in \mathcal{D}^n} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) \leq 2^N}} \sum_{\substack{J_1, J_2 \in \mathcal{D}^m \\ \ell(J_3) \leq \ell(J_1) = \ell(J_2) < 2^N}} \langle T(E_{I_1 \times J_1} f_1, \Delta_{I_2 \times J_2} f_2), \Delta_{I_3 \times J_3} f_3 \rangle.$$

*Bilinear bi-parameter organisation of sums.* We start dealing with  $\sigma_1^1$ , but due to the need to combine paraproduct like terms, this systemically requires the need to combine the terms coming from  $\sigma_1^1$  with some of the analogous terms coming from  $\sigma_1^2$ ,  $\sigma_1^3$  and  $\sigma_1^4$ . Therefore, we in essence deal with all of them simultaneously.

With a fixed  $I_3$  we split

$$\begin{aligned} \sum_{\substack{I_1, I_2 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_1) = 2\ell(I_2) \leq 2^N}} &= \sum_{\substack{I_1, I_2 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_1) = 2\ell(I_2) \leq 2^N \\ \max(d(I_1, I_3), d(I_2, I_3)) > \ell(I_3)^{\gamma_n} \ell(I_2)^{1-\gamma_n}}} &+ \sum_{\substack{I_1, I_2 \in \mathcal{D}^n \\ \ell(I_3) \leq \ell(I_1) = 2\ell(I_2) \leq 2^N \\ \max(d(I_1, I_3), d(I_2, I_3)) \leq \ell(I_3)^{\gamma_n} \ell(I_2)^{1-\gamma_n} \\ I_1 \cap I_3 = \emptyset \text{ or } I_1 = I_3 \text{ or } I_2 \cap I_3 = \emptyset}} \\ &+ \sum_{\substack{I_1, I_2 \in \mathcal{D}^n \\ \ell(I_1) = 2\ell(I_2) \leq 2^N \\ I_3 \subset I_2 \subset I_1}}. \end{aligned}$$

These parts are titled separated, diagonal and nested, respectively. A similar decomposition of  $\sum_{\substack{J_1, J_2 \in \mathcal{D}^m \\ \ell(J_3) \leq \ell(J_1) = 2\ell(J_2) \leq 2^N}}$  is done with a fixed  $J_3$ . This splits  $\sigma_1^1$  into nine parts, which are titled like separated/separated, separated/diagonal, etc.

We begin with the most difficult case – i.e. Nested/Nested. The remaining cases are easier, except for the fact that finding a common parent (which is required for the shift structure) is easiest (as it is obvious) in the nested case. We only briefly comment on the other cases after handling Nested/Nested.

*Nested/Nested.* We have by definition that

$$\sigma_{1,\text{nes,nes}}^1 = \sum_{\substack{I_2 \in \mathcal{D}^n: \ell(I_2) < 2^N \\ I_3 \in \mathcal{D}_{\text{good}}^n: \ell(I_3) > 2^{-M} \\ I_3 \subset I_2}} \sum_{\substack{J_2 \in \mathcal{D}^m: \ell(J_2) < 2^N \\ J_3 \in \mathcal{D}_{\text{good}}^m: \ell(J_3) > 2^{-M} \\ J_3 \subset J_2}} \langle T(\Delta_{I_2^{(1)} \times J_2^{(1)}} f_1, E_{I_2 \times J_2} f_2), \Delta_{I_3 \times J_3} f_3 \rangle.$$

We write

$$\begin{aligned} \langle T(\Delta_{I_2^{(1)} \times J_2^{(1)}} f_1, E_{I_2 \times J_2} f_2), \Delta_{I_3 \times J_3} f_3 \rangle &= \langle f_1, h_{I_2^{(1)}} \otimes h_{J_2^{(1)}} \rangle \langle f_2 \rangle_{I_2 \times J_2} \langle f_3, h_{I_3} \otimes h_{J_3} \rangle \\ &\quad \times \langle T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1_{I_2 \times J_2}), h_{I_3} \otimes h_{J_3} \rangle. \end{aligned}$$

First, we split

$$\begin{aligned} T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1_{I_2 \times J_2}) &= T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1_{I_2^c} \otimes 1_{J_2^c}) - T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1 \otimes 1_{J_2^c}) \\ &\quad - T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1_{I_2^c} \otimes 1) + T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1). \end{aligned}$$

Using the identity

$$h_{I_2^{(1)}} = \langle h_{I_2^{(1)}} \rangle_{I_2} + s_{I_2}, \quad s_{I_2} := 1_{I_2^c} [h_{I_2^{(1)}} - \langle h_{I_2^{(1)}} \rangle_{I_2}],$$

we further write

$$-T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1 \otimes 1_{J_2^c}) = -T(s_{I_2} \otimes h_{J_2^{(1)}}, 1 \otimes 1_{J_2^c}) - \langle h_{I_2^{(1)}} \rangle_{I_2} T(1 \otimes h_{J_2^{(1)}}, 1 \otimes 1_{J_2^c})$$

and

$$-T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1_{I_2^c} \otimes 1) = -T(h_{I_2^{(1)}} \otimes s_{J_2}, 1_{I_2^c} \otimes 1) - \langle h_{J_2^{(1)}} \rangle_{J_2} T(h_{I_2^{(1)}} \otimes 1, 1_{I_2^c} \otimes 1)$$

and

$$\begin{aligned} T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1) &= T(s_{I_2} \otimes s_{J_2}, 1) + \langle h_{I_2^{(1)}} \rangle_{I_2} T(1 \otimes s_{J_2}, 1) \\ &\quad + \langle h_{J_2^{(1)}} \rangle_{J_2} T(s_{I_2} \otimes 1, 1) + \langle h_{I_2^{(1)}} \rangle_{I_2} \langle h_{J_2^{(1)}} \rangle_{J_2} T(1, 1). \end{aligned}$$

We have split  $\sigma_{1,\text{nes,nes}}^1$  into four essentially different parts  $\sigma_{1,\text{nes,nes}}^1 = \sum_{u=1}^4 \sigma_{1,\text{nes,nes},u}^1$ , where we agree that  $\sigma_{1,\text{nes,nes},1}^1$  corresponds to

$$\begin{aligned} T(h_{I_2^{(1)}} \otimes h_{J_2^{(1)}}, 1_{I_2^c} \otimes 1_{J_2^c}) - T(s_{I_2} \otimes h_{J_2^{(1)}}, 1 \otimes 1_{J_2^c}) \\ - T(h_{I_2^{(1)}} \otimes s_{J_2}, 1_{I_2^c} \otimes 1) + T(s_{I_2} \otimes s_{J_2}, 1), \end{aligned}$$

$\sigma_{1,\text{nes,nes},2}^1$  corresponds to

$$\langle h_{I_2^{(1)}} \rangle_{I_2} [-T(1 \otimes h_{J_2^{(1)}}, 1 \otimes 1_{J_2^c}) + T(1 \otimes s_{J_2}, 1)],$$

$\sigma_{1,\text{nes,nes},3}^1$  corresponds to

$$\langle h_{J_2^{(1)}} \rangle_{J_2} [-T(h_{I_2^{(1)}} \otimes 1, 1_{I_2^c} \otimes 1) + T(s_{I_2} \otimes 1, 1)],$$

and  $\sigma_{1,\text{nes,nes},4}^1$  corresponds to  $\langle h_{I_2^{(1)}} \rangle_{I_2} \langle h_{J_2^{(1)}} \rangle_{J_2} T(1, 1)$ .

We start with the terms  $\sigma_{1,\text{nes,nes},2}^1$ ,  $\sigma_{1,\text{nes,nes},3}^1$  and  $\sigma_{1,\text{nes,nes},4}^1$ . The first two are handled similarly (they both produce partial paraproducts), and we only deal with  $\sigma_{1,\text{nes,nes},3}^1$ . The term  $\sigma_{1,\text{nes,nes},4}^1$  produces a full paraproduct.

The term  $\sigma_{1,\text{nes,nes},3}^1$ . We are looking at the term

$$\begin{aligned} \sigma_{1,\text{nes,nes},3}^1 &= \sum_{\substack{I_2 \in \mathcal{D}^n: \ell(I_2) < 2^N \\ I_3 \in \mathcal{D}_{\text{good}}^n: \ell(I_3) > 2^{-M} \\ I_3 \subset I_2}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) < 2^N}} \langle f_3, h_{I_3} \otimes h_{J_3} \rangle \langle -T(h_{I_2^{(1)}} \otimes 1, 1_{I_2^c} \otimes 1) \\ &+ T(s_{I_2} \otimes 1, 1), h_{I_3} \otimes h_{J_3} \rangle \sum_{\substack{k \geq 0 \\ 2^k \leq 2^{N-1}/\ell(J_3)}} \langle h_{J_3^{(k+1)}} \rangle_{J_3^{(k)}} \langle \langle f_1, h_{I_2^{(1)}} \rangle_1, h_{J_3^{(k+1)}} \rangle \langle f_2 \rangle_{I_2 \times J_3^{(k)}}. \end{aligned}$$

Notice that

$$\begin{aligned} \langle h_{J_3^{(k+1)}} \rangle_{J_3^{(k)}} \langle \langle f_1, h_{I_2^{(1)}} \rangle_1, h_{J_3^{(k+1)}} \rangle &= \langle \Delta_{J_3^{(k+1)}} \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k)}} \\ &= \langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k)}} - \langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k+1)}} \end{aligned}$$

if  $2^k < 2^{N-1}/\ell(J_3)$ , and it equals  $\langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k)}}$  if  $2^k = 2^{N-1}/\ell(J_3)$ .

The sum does not collapse. To achieve this, we need to sum  $\sigma_{1,\text{nes,nes},3}^1$  with the term

$$\begin{aligned} \sigma_{1,\text{nes,nes},3}^2 &= \sum_{\substack{I_2 \in \mathcal{D}^n: \ell(I_2) < 2^N \\ I_3 \in \mathcal{D}_{\text{good}}^n: \ell(I_3) > 2^{-M} \\ I_3 \subset I_2}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) < 2^{N-1}}} \langle f_3, h_{I_3} \otimes h_{J_3} \rangle \langle -T(h_{I_2^{(1)}} \otimes 1, 1_{I_2^c} \otimes 1) \\ &+ T(s_{I_2} \otimes 1, 1), h_{I_3} \otimes h_{J_3} \rangle \sum_{\substack{k \geq 0 \\ 2^k < 2^{N-1}/\ell(J_3)}} \langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k+1)}} [\langle f_2 \rangle_{I_2 \times J_3^{(k)}} - \langle f_2 \rangle_{I_2 \times J_3^{(k+1)}}]. \end{aligned}$$

This formula for  $\sigma_{1,\text{nes,nes},3}^2$  can be seen by calculating similarly as we did with  $\sigma_{1,\text{nes,nes},3}^1$ . Noting the key cancellation the result is that  $\sigma_{1,\text{nes,nes},3}^1 + \sigma_{1,\text{nes,nes},3}^2$  equals

$$\begin{aligned} &\sum_{\substack{I_2 \in \mathcal{D}^n \\ I_3 \in \mathcal{D}_{\text{good}}^n \\ \ell(I_2) < 2^N \\ \ell(I_3) > 2^{-M} \\ I_3 \subset I_2}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) < 2^N}} \langle f_3, h_{I_3} \otimes h_{J_3} \rangle \langle -T(h_{I_2^{(1)}} \otimes 1, 1_{I_2^c} \otimes 1) + T(s_{I_2} \otimes 1, 1), h_{I_3} \otimes h_{J_3} \rangle \\ &\times \sum_{\substack{k \geq 0 \\ 2^k \leq 2^{N-1}/\ell(J_3)}} [\langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k)}} \langle f_2 \rangle_{I_2 \times J_3^{(k)}} - \langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k+1)}} \langle f_2 \rangle_{I_2 \times J_3^{(k+1)}}], \end{aligned}$$

where we understand that  $[\langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k)}} \langle f_2 \rangle_{I_2 \times J_3^{(k)}} - \langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k+1)}} \langle f_2 \rangle_{I_2 \times J_3^{(k+1)}}]$  is replaced with  $\langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3^{(k)}} \langle f_2 \rangle_{I_2 \times J_3^{(k)}}$  if  $2^k = 2^{N-1}/\ell(J_3)$ . Now, this last sum collapses to  $\langle \langle f_1, h_{I_2^{(1)}} \rangle_1 \rangle_{J_3} \langle f_2 \rangle_{I_2 \times J_3}$ . Define

$$b_{I_2, I_3} = C^{-1} \left( \frac{\ell(I_2)}{\ell(I_3)} \right)^{\alpha/2} |I_2|^{-1/2} \langle -T(h_{I_2^{(1)}} \otimes 1, 1_{I_2^c} \otimes 1) + T(s_{I_2} \otimes 1, 1), h_{I_3} \rangle_1$$

whenever  $I_2 \in \mathcal{D}^n$ ,  $I_3 \in \mathcal{D}_{\text{good}}^n$  and  $I_3 \subset I_2$ , and otherwise set  $b_{I_2, I_3} = 0$ . We have that  $\sigma_{1, \text{nes}, \text{nes}, 3}^1 + \sigma_{1, \text{nes}, \text{nes}, 3}^2$  equals

$$C \sum_{\substack{I_2, I_3 \in \mathcal{D}^n \\ 2^{-M} < \ell(I_2), \ell(I_3) < 2^N \\ I_3 \subset I_2}} \left( \frac{\ell(I_3)}{\ell(I_2)} \right)^{\alpha/2} \langle h_{I_3} \otimes \pi_{\mathcal{D}_{\text{good}}^m, b_{I_2, I_3}}^{N, M} (\langle f_1, h_{I_2^{(1)}} \rangle_1, \langle f_2, h_{I_2^0} \rangle_1), f_3 \rangle,$$

where

$$\pi_{\mathcal{D}_{\text{good}}^m, b_{I_2, I_3}}^{N, M} (g_1, g_2) = \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) < 2^N}} \langle b_{I_2, I_3}, h_{J_3} \rangle_{J_3} \langle g_2 \rangle_{J_3} h_{J_3}.$$

This is seen to be a sum of (truncated) partial paraproducts with indices  $0, 1, k, k \geq 1$ . This is because of the following lemma, which gives the required BMO norm bound for the functions  $b_{I_2, I_3}$ . We give the proof relatively carefully as a model.

**5.3. Lemma.** *Let  $I_2 \in \mathcal{D}^n$ ,  $I_3 \in \mathcal{D}_{\text{good}}^n$  and  $I_3 \subset I_2$ . Then we have*

$$\begin{aligned} |\langle T(h_{I_2^{(1)}} \otimes 1, 1_{I_2^c} \otimes 1), h_{I_3} \otimes a_J \rangle| + |\langle T(s_{I_2} \otimes 1, 1), h_{I_3} \otimes a_J \rangle| \\ \lesssim \left( \frac{\ell(I_3)}{\ell(I_2)} \right)^{\alpha/2} \left( \frac{|I_3|}{|I_2|} \right)^{1/2} |J| \end{aligned}$$

for all cubes  $J \subset \mathbb{R}^m$  and all functions  $a_J$  such that  $\text{spt } a_J \subset J$ ,  $|a_J| \leq 1$  and  $\int a_J = 0$ . Notice that here  $h_{I_3}$  really stands for a cancellative Haar function (as is the case above where  $\ell(I_3) < 2^N$ ), but whether  $h_{I_2^{(1)}}$  is cancellative or not does not matter.

*Proof.* We only explicitly deal with  $|\langle T(s_{I_2} \otimes 1, 1), h_{I_3} \otimes a_J \rangle|$ , and estimate

$$\begin{aligned} |\langle T(s_{I_2} \otimes 1, 1), h_{I_3} \otimes a_J \rangle| &\leq |\langle T(s_{I_2} \otimes 1_{3J}, 1 \otimes 1_{3J}), h_{I_3} \otimes a_J \rangle| \\ &\quad + |\langle T(s_{I_2} \otimes 1_{3J}, 1 \otimes 1_{(3J)^c}), h_{I_3} \otimes a_J \rangle| \\ &\quad + |\langle T(s_{I_2} \otimes 1_{(3J)^c}, 1), h_{I_3} \otimes a_J \rangle| = A_1 + A_2 + A_3. \end{aligned}$$

We first consider the case  $\ell(I_3) < 2^{-r} \ell(I_2)$  ( $r$  is related to goodness). We have

$$\begin{aligned} A_1 = \left| \int_{\mathbb{R}^n} \int_{I_2^c} \int_{I_3} [K_{1_{3J}, 1_{3J}, a_J}(x_1, y_1, z_1) \right. \\ \left. - K_{1_{3J}, 1_{3J}, a_J}(c_{I_3}, y_1, z_1)] s_{I_2}(y_1) h_{I_3}(x_1) dx_1 dy_1 dz_1 \right|, \end{aligned}$$

which gives

$$\begin{aligned} A_1 &\lesssim C(1_{3J}, 1_{3J}, a_J) \ell(I_3)^\alpha \int_{\mathbb{R}^n} \int_{I_2^c} \int_{I_3} \frac{|s_{I_2}(y_1) h_{I_3}(x_1)| dx_1 dy_1 dz_1}{(|c_{I_3} - y_1| + |c_{I_3} - z_1|)^{2n+\alpha}} \\ &\lesssim |J| |I_2|^{-1/2} |I_3|^{1/2} \ell(I_3)^\alpha \int_{I_2^c} \frac{dy_1}{|c_{I_3} - y_1|^{n+\alpha}} \lesssim \left( \frac{\ell(I_3)}{\ell(I_2)} \right)^{\alpha/2} \left( \frac{|I_3|}{|I_2|} \right)^{1/2} |J|. \end{aligned}$$

The last estimate used the fact that  $d(I_3, I_2^c) > \ell(I_3)^\gamma \ell(I_2)^{1-\gamma} \geq \ell(I_3)^{1/2} \ell(I_2)^{1/2}$ , which follows from the goodness of the cube  $I_3$  and the fact that  $\ell(I_3) < 2^{-r} \ell(I_2)$ . The estimate for  $A_2$  is very similar – just use the full kernel representation and a Hölder estimate of

the full kernel instead. The  $\mathbb{R}^n$  integrals are handled using exactly the same argument as above, while the  $|J|$  is obtained from the estimate

$$\ell(J)^\alpha \int_{(3J)^c} \int_{3J} \int_J \frac{dx_2 dy_2 dz_2}{(|c_J - y_2| + |c_J - z_2|)^{2m+\alpha}} \lesssim |J| \ell(J)^\alpha \int_{(3J)^c} \frac{dz_2}{|c_J - z_2|^{m+\alpha}} \lesssim |J|.$$

As we just disregarded the  $3J$  in the  $y_2$  integral above, we see that  $A_3$  follows by essentially the same estimate. We are done with the case  $\ell(I_3) < 2^{-r} \ell(I_2)$ .

Next, we deal with the remaining case where  $\ell(I_3) \sim \ell(I_2)$ . Here we use the above splitting to  $A_i$ ,  $i = 1, 2, 3$ , but also further split each of these to  $A_i^k$ ,  $k = 1, 2$ , where  $k = 1$  means that we have replaced  $s_{I_2}$  with  $1_{3I_3} s_{I_2}$  and  $k = 2$  means that we have replaced  $s_{I_2}$  with  $1_{(3I_3)^c} s_{I_2}$ . The required estimate for  $A_1^1$  follows using the partial kernel representation, the size estimate for the kernel, and the estimate  $\int_{3I_3 \setminus I_3} \int_{I_3} \frac{dx_1 dy_1}{|x_1 - y_1|^n} \lesssim |I_3|$ . The estimate for  $A_1^2$  follows using the partial kernel representation, the Hölder estimate for the kernel, and standard estimates. All the remaining estimates follow using similar estimates, we just point out which kernel estimate is used. In  $A_2^1$  use the full kernel and a mixed Hölder and size estimate, and in  $A_2^2$  use the full kernel and a Hölder estimate. The remaining terms  $A_3^1$  and  $A_3^2$  are handled as  $A_2^1$  and  $A_2^2$  respectively.  $\square$

*The term  $\sigma_{1,\text{nes},\text{nes},4}^1$ .* Here we need to sum up all the terms  $\sigma_{1,\text{nes},\text{nes},4}^i$ ,  $i = 1, \dots, 4$ . We first sum up  $\sigma_{1,\text{nes},\text{nes},4}^1$  and  $\sigma_{1,\text{nes},\text{nes},4}^2$ . Using calculations like above we see that

$$\begin{aligned} \sigma_{1,\text{nes},\text{nes},4}^1 + \sigma_{1,\text{nes},\text{nes},4}^2 &= \sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) < 2^N}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) < 2^N}} \langle f_3, h_{I_3} \otimes h_{J_3} \rangle \langle T(1, 1), h_{I_3} \otimes h_{J_3} \rangle \\ &\quad \sum_{\substack{k \geq 0 \\ 2^k \leq 2^{N-1}/\ell(I_3)}} \langle h_{I_3^{(k+1)}} \rangle_{I_3^{(k)}} \langle \langle f_1, h_{I_3^{(k+1)}} \rangle_1 \rangle_{J_3} \langle f_2 \rangle_{I_3^{(k)} \times J_3}. \end{aligned}$$

Notice that here

$$\begin{aligned} \langle h_{I_3^{(k+1)}} \rangle_{I_3^{(k)}} \langle \langle f_1, h_{I_3^{(k+1)}} \rangle_1 \rangle_{J_3} &= \langle h_{I_3^{(k+1)}} \otimes \langle f_1, h_{I_3^{(k+1)}} \rangle_1 \rangle_{I_3^{(k)} \times J_3} \\ &= \langle \Delta_{I_3^{(k+1)}}^1 f_1 \rangle_{I_3^{(k)} \times J_3} = \langle f_1 \rangle_{I_3^{(k)} \times J_3} - \langle f_1 \rangle_{I_3^{(k+1)} \times J_3} \end{aligned}$$

if  $2^k < 2^{N-1}/\ell(I_3)$ , and it equals  $\langle f_1 \rangle_{I_3^{(k)} \times J_3}$  if  $2^k = 2^{N-1}/\ell(I_3)$ . In exactly the same way we see that

$$\begin{aligned} \sigma_{1,\text{nes},\text{nes},4}^3 + \sigma_{1,\text{nes},\text{nes},4}^4 &= \sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) < 2^{N-1}}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) < 2^N}} \langle f_3, h_{I_3} \otimes h_{J_3} \rangle \langle T(1, 1), h_{I_3} \otimes h_{J_3} \rangle \\ &\quad \sum_{\substack{k \geq 0 \\ 2^k < 2^{N-1}/\ell(I_3)}} \langle f_1 \rangle_{I_3^{(k+1)} \times J_3} [\langle f_2 \rangle_{I_3^{(k)} \times J_3} - \langle f_2 \rangle_{I_3^{(k+1)} \times J_3}]. \end{aligned}$$

Noticing the cancellation we have

$$\sum_{i=1}^4 \sigma_{1,\text{nes},\text{nes},4}^i = \sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n \\ 2^{-M} < \ell(I_3) < 2^N}} \sum_{\substack{J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(J_3) < 2^N}} \langle f_3, h_{I_3} \otimes h_{J_3} \rangle \langle T(1, 1), h_{I_3} \otimes h_{J_3} \rangle$$

$$\sum_{\substack{k \geq 0 \\ 2^k \leq 2^{N-1}/\ell(I_3)}} [\langle f_1 \rangle_{I_3^{(k)} \times J_3} \langle f_2 \rangle_{I_3^{(k)} \times J_3} - \langle f_1 \rangle_{I_3^{(k+1)} \times J_3} \langle f_2 \rangle_{I_3^{(k+1)} \times J_3}],$$

where we understand that  $[\langle f_1 \rangle_{I_3^{(k)} \times J_3} \langle f_2 \rangle_{I_3^{(k)} \times J_3} - \langle f_1 \rangle_{I_3^{(k+1)} \times J_3} \langle f_2 \rangle_{I_3^{(k+1)} \times J_3}]$  is replaced with  $\langle f_1 \rangle_{I_3^{(k)} \times J_3} \langle f_2 \rangle_{I_3^{(k)} \times J_3}$  if  $2^k = 2^{N-1}/\ell(I_3)$ . This last sum collapses to  $\langle f_1 \rangle_{I_3 \times J_3} \langle f_2 \rangle_{I_3 \times J_3}$ . Therefore, we have shown that

$$\sum_{i=1}^4 \sigma_{1,\text{nes},\text{nes},4}^i = \langle \Pi_{\mathcal{D}_{\text{good}}^n, \mathcal{D}_{\text{good}}^m, T(1,1)}^{N,M}(f_1, f_2), f_3 \rangle,$$

where  $\Pi_{\mathcal{D}_{\text{good}}^n, \mathcal{D}_{\text{good}}^m, T(1,1)}^{N,M}(f_1, f_2)$  equals

$$\sum_{\substack{I_3 \in \mathcal{D}_{\text{good}}^n, J_3 \in \mathcal{D}_{\text{good}}^m \\ 2^{-M} < \ell(I_3), \ell(J_3) < 2^N}} \langle T(1,1), h_{I_3} \otimes h_{J_3} \rangle \langle f_1 \rangle_{I_3 \times J_3} \langle f_2 \rangle_{I_3 \times J_3} h_{I_3} \otimes h_{J_3}.$$

It remains to consider the term  $\sigma_{1,\text{nes},\text{nes},1}^1$ .

The term  $\sigma_{1,\text{nes},\text{nes},1}^1$ . This term clearly produces a sum of shifts of the type  $(0, 1, k)$ ,  $(0, 1, v)$ ,  $k, v \geq 1$ , provided that we check the estimate

$$|I_2|^{-1/2} |J_2|^{-1/2} |\langle T(s_{I_2} \otimes s_{J_2}, 1), h_{I_3} \otimes h_{J_3} \rangle| \lesssim \left( \frac{\ell(I_3)}{\ell(I_2)} \right)^{\alpha/2} \frac{|I_3|^{1/2}}{|I_2|} \left( \frac{\ell(J_3)}{\ell(J_2)} \right)^{\alpha/2} \frac{|J_3|^{1/2}}{|J_2|},$$

and the three other completely analogous estimates. This estimate is checked using similar estimates as in Lemma 5.3 (but only the full kernel representation is needed). Therefore, we are done with this term, and so with the whole Nested/Nested part.

*Remaining parts.* As mentioned, the other parts are essentially easier than the one we just handled, namely Nested/Nested. We now briefly indicate how they are handled. First, we cover the lemmata which give the existence of common ancestors (the existence of the common parent was obvious above because of the nestedness). The following lemma gives the existence of the common dyadic parent in the separated case.

**5.4. Lemma.** *Let  $I_1, I_2 \in \mathcal{D}^n$  and  $I_3 \in \mathcal{D}_{\text{good}}^n$  be such that  $\ell(I_3) \leq \ell(I_1) = 2\ell(I_2)$  and  $\max(d(I_1, I_3), d(I_2, I_3)) > \ell(I_3)^{\gamma_n} \ell(I_2)^{1-\gamma_n}$ . Then there exists a cube  $K \in \mathcal{D}^n$  so that  $I_1 \cup I_2 \cup I_3 \subset K$  and*

$$\max(d(I_3, I_1), d(I_3, I_2)) \gtrsim \ell(I_3)^{\gamma_n} \ell(K)^{1-\gamma_n}.$$

The existence of the common dyadic parent in the diagonal case is given next.

**5.5. Lemma.** *Let  $I_1, I_2 \in \mathcal{D}^n$  and  $I_3 \in \mathcal{D}_{\text{good}}^n$  be such that we have  $\ell(I_3) \leq \ell(I_1) = 2\ell(I_2)$ ,  $\max(d(I_1, I_3), d(I_2, I_3)) \leq \ell(I_3)^{\gamma_n} \ell(I_2)^{1-\gamma_n}$  and either  $I_1 \cap I_3 = \emptyset$  or  $I_1 = I_3$  or  $I_2 \cap I_3 = \emptyset$ . Then there exists a cube  $K \in \mathcal{D}^n$  so that  $I_1 \cup I_2 \cup I_3 \subset K$  and  $\ell(K) \leq 2^r \ell(I_3)$ .*

The results are by Hytönen (see [31] and [32]) in the linear case, and the above formulations are proved in Li–Martikainen–Ou–Vuorinen [47]. Denote the smallest common dyadic ancestor by  $I_1 \vee I_2 \vee I_3$  (which exists by the above lemmata). Using these results we arrange the desired shift structure by reorganising the summations to the cases where  $I_1 \vee I_2 \vee I_3 = K$  for a given  $K \in \mathcal{D}^n$  (and similarly for the summations over  $\mathcal{D}^m$ ).

As the Nested/Nested is the only place, where a full paraproduct appears, we only get shifts or partial paraproducts from the remaining terms. The collapse of the various partial paraproducts require the same kind of arguments as above.

Having taken care of the structural considerations, things boil down to proving the desired bounds required by the shifts and partial paraproducts. Now, we explain the role of the assumption  $\text{spt } f_i \subset B(0, c_0 2^N)$ . This ensures that e.g. all the appearing cubes  $I_i$  satisfy  $d(I_i, I_{i'}) \leq 2c_0 2^N$ . So if the smallest cube  $I_3$  satisfies  $\ell(I_3) = 2^N$ , then

$$\max(d(I_1, I_3), d(I_2, I_3)) \leq 2c_0 2^N < \ell(I_3)^{\gamma_n} \ell(I_2)^{1-\gamma_n}$$

if  $c_0$  is chosen appropriately. This means that  $\ell(I_3) < 2^N$  in the separated sum, which is key for obtaining the desired estimates using Hölder bounds of the kernels. On the other hand, in the diagonal sum it does not matter whether the Haar functions are cancellative or not (we demonstrate this below). And, in the various nested cases we sum as above.

Therefore, in the majority of the cases we use the kernel estimates like we did above, and combine these with the quantitative bounds for  $I_1 \vee I_2 \vee I_3$  given by the above two lemmata. However, we did not explain the role of the diagonal BMO assumptions nor the weak boundedness assumptions yet. The need arises in the cases Diagonal/Nested (and Nested/Diagonal) and Diagonal/Diagonal respectively. To conclude, we give an idea how the diagonal BMO assumptions are used in the Diagonal/Nested case.

*Diagonal/Nested and the diagonal BMO assumptions.* If we extract the partial paraproduct here as previously, we are e.g. left with the task of showing the BMO bound

$$|\langle T(h_{I_1} \otimes 1, 1_{I_2} \otimes 1), h_{I_3} \otimes a_J \rangle| \lesssim |J|,$$

whenever  $I_1, I_2$  and  $I_3$  are like in the diagonal summation,  $J \subset \mathbb{R}^m$  is a cube, and  $a_J$  is a function such that  $\text{spt } a_J \subset J$ ,  $|a_J| \leq 1$  and  $\int a_J = 0$ . If  $I_1 \cap I_3 = \emptyset$  or  $I_2 \cap I_3 = \emptyset$ , this is not difficult. So we only consider the remaining case where  $I_1 = I_3$  and  $I_2 \in \text{ch}(I_1)$ . Next, we use the familiar splitting  $1 = 1_{3J} + 1_{(3J)^c}$ , and also split  $h_{I_1} = |I_1|^{-1/2} \sum_{I' \in \text{ch}(I_1)} \epsilon_{I'} 1_{I'}$ , where  $\epsilon_{I'} = \pm 1$ . So we are left with terms like

$$|I_1|^{-1} \sum_{I', I'' \in \text{ch}(I_1)} |\langle T(1_{I'} \otimes 1_{(3J)^c}, 1_{I_2} \otimes 1), 1_{I''} \otimes a_J \rangle|$$

and

$$|I_1|^{-1} \sum_{I', I'' \in \text{ch}(I_1)} |\langle T(1_{I'} \otimes 1_{3J}, 1_{I_2} \otimes 1_{3J}), 1_{I''} \otimes a_J \rangle|.$$

In the first sum if  $I' \neq I_2$  or  $I'' \neq I_2$ , we use the full kernel representation. If  $I' = I'' = I_2$  we use the partial kernel representation with  $K_{1_{I_2}, 1_{I_2}, 1_{I_2}}$ . In the second sum if  $I' \neq I_2$  or  $I'' \neq I_2$  we use the partial kernel representation with  $K_{1_{3J}, 1_{3J}, 1_{a_J}}$ . Finally, in the second sum the case  $I' = I'' = I_2$  requires us to bound  $|I_1|^{-1} |\langle T(1_{I_2} \otimes 1_{3J}, 1_{I_2} \otimes 1_{3J}), 1_{I_2} \otimes a_J \rangle|$ . But this is dominated by  $|J|$  using one of the diagonal BMO assumptions. This ends our treatment of  $\Sigma^1$ , and thus the proof of the a priori bounded case.

**$T$  is not a priori bounded.** It is possible to first prove a representation theorem in a certain finite set up, where no a priori boundedness is needed. Reductions of this type in the linear one-parameter situation appear in [26] and [33]. We omit the technical details in our setting as they are similar. A corollary of such a special representation is the boundedness of  $T$ , say from  $L^3 \times L^3 \rightarrow L^{3/2}$ . After this, we can run the above argument.

□

## 6. BOUNDEDNESS RESULTS FOR DMOS

6.1. **Weighted bounds for shifts and partial paraproducts.** It is well-known that

$$(6.1) \quad \|f\|_{L^2(v)}^2 \lesssim_{[v]_{A_\infty(\mathbb{R}^n)}} \int_{\mathbb{R}^n} \sum_{I \in \mathcal{D}^n} |\Delta_I f|^2 v$$

for  $f: \mathbb{R}^n \rightarrow \mathbb{C}$  and  $v \in A_\infty(\mathbb{R}^n)$ . The fact that this holds for  $A_\infty$  weights is important for us (due to the nature of bilinear weights). This can be proved using some good lambda estimates as in [6]. See also e.g. Theorem 3.4 in [14] for an explicit proof of this in a more general setting. The following  $A_\infty$  extrapolation result is [9, Theorem 2.1].

6.2. **Lemma.** *Let  $(f, g)$  be a pair of positive functions defined on  $\mathbb{R}^m$ . Suppose that there exists some  $0 < p_0 < \infty$  such that for every  $w \in A_\infty(\mathbb{R}^m)$  we have*

$$\int_{\mathbb{R}^m} f^{p_0} w \leq C([w]_\infty, p_0) \int_{\mathbb{R}^m} g^{p_0} w.$$

Then for all  $0 < p < \infty$  and  $w \in A_\infty(\mathbb{R}^m)$  we have

$$\int_{\mathbb{R}^m} f^p w \leq C([w]_\infty, p) \int_{\mathbb{R}^m} g^p w.$$

Therefore, by extrapolation (6.1) improves to

$$\int_{\mathbb{R}^n} \left( \sum_j |f_j|^2 \right)^{p/2} v \lesssim_{[v]_{A_\infty(\mathbb{R}^n)}} \int_{\mathbb{R}^n} \left( \sum_j \sum_{I \in \mathcal{D}^n} |\Delta_I f_j|^2 \right)^{p/2} v, \quad p \in (0, \infty).$$

A direct corollary is that

$$(6.3) \quad \|f\|_{L^p(v)}^p \lesssim \iint_{\mathbb{R}^{n+m}} \left( \sum_{I \in \mathcal{D}^n} |\Delta_I^1 f|^2 \right)^{p/2} v$$

and

$$(6.4) \quad \|f\|_{L^p(v)}^p \lesssim \iint_{\mathbb{R}^{n+m}} \left( \sum_{\substack{I \in \mathcal{D}^n \\ J \in \mathcal{D}^m}} |\Delta_{I \times J} f|^2 \right)^{p/2} v$$

for  $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$ ,  $p \in (0, \infty)$  and weights  $v$  satisfying  $\text{ess sup}_{x_2 \in \mathbb{R}^m} [v(\cdot, x_2)]_{A_\infty(\mathbb{R}^n)} < \infty$  and  $\text{ess sup}_{x_1 \in \mathbb{R}^n} [v(x_1, \cdot)]_{A_\infty(\mathbb{R}^m)} < \infty$  (and the implicit constants depend on these norms). We will apply this with  $v = v_3 = w_1^{r/p} w_2^{r/q} \in A_{2r}(\mathbb{R}^n \times \mathbb{R}^m)$ , where  $p, q \in (1, \infty)$ ,  $1/p + 1/q = 1/r$  and  $w_1 \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ ,  $w_2 \in A_q(\mathbb{R}^n \times \mathbb{R}^m)$ .

*Shifts.* We prove the weighted estimates for shifts.

6.5. **Proposition.** *Let  $1 < p, q < \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ ,  $w_1 \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$  and  $w_2 \in A_q(\mathbb{R}^n \times \mathbb{R}^m)$  be bi-parameter weights, and set  $v_3 := w_1^{r/p} w_2^{r/q}$ . Suppose that  $(S_\omega)_\omega$  is a collection of bilinear bi-parameter shifts of the same type, where  $S_\omega$  is defined in the grid  $\mathcal{D}_\omega$ . Then we have*

$$\|\mathbb{E}_\omega S_\omega(f_1, f_2)\|_{L^r(v_3)} \leq C([w_1]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}, [w_2]_{A_q(\mathbb{R}^n \times \mathbb{R}^m)}) \|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}.$$

6.6. *Remark.* Here and below the non-averaged cases, where there is only one operator in a fixed grid, are true and easier to prove.

*Proof of Proposition 6.5.* We allow the implicit constants to depend on  $[w_1]_{A_p(\mathbb{R} \times \mathbb{R}^m)}$  and  $[w_2]_{A_p(\mathbb{R} \times \mathbb{R}^m)}$ . We first prove the statement in a fixed grid directly. Let  $S = S_{\mathcal{D}}$  for some fixed  $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ .

The easiest case is when we have  $h_{I_3}^0 \otimes h_{J_3}^0$  (i.e. all the non-cancellative functions are outside) – then we don't need the  $A_\infty$  square function estimates at all, and we can work directly. So we omit writing more about this case.

Consider the case  $S(f_1, f_2) = \sum_{K \times V \in \mathcal{D}} A_{K,V}(f_1, f_2)$ , where  $A_{K,V}(f_1, f_2)$  has the form

$$\sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_i^{(k_i)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}^m \\ J_i^{(v_i)} = V}} a_{K,V,(I_i),(J_i)} \langle f_1, h_{I_1}^0 \otimes h_{J_1} \rangle \langle f_2, h_{I_2} \otimes h_{J_2} \rangle h_{I_3} \otimes h_{J_3}^0.$$

Using the lower square function estimate (6.3) we get that

$$\|S(f_1, f_2)\|_{L^r(v_3)} \lesssim \left\| \left( \sum_{K \in \mathcal{D}^n} \left| \sum_{V \in \mathcal{D}^m} A_{K,V}(f_1, f_2) \right|^2 \right)^{1/2} \right\|_{L^r(v_3)}.$$

Using the cancellative Haar functions and the normalisation of the coefficients  $a_{K,V,(I_i),(J_i)}$  we get  $|A_{K,V}(f_1, f_2)| \leq M_{\mathcal{D}}(\Delta_{V,v_1}^2 f_1) M_{\mathcal{D}}(\Delta_{K \times V}^{k_2, v_2} f_2)$ . Combining these gives that

$$\begin{aligned} \|S(f_1, f_2)\|_{L^2(v_3)} &\lesssim \left\| \left( \sum_{K \in \mathcal{D}^n} \left| \sum_{V \in \mathcal{D}^m} M_{\mathcal{D}}(\Delta_{V,v_1}^2 f_1) M_{\mathcal{D}}(\Delta_{K \times V}^{k_2, v_2} f_2) \right|^2 \right)^{1/2} \right\|_{L^r(v_3)} \\ &\leq \left\| \left( \sum_{V \in \mathcal{D}^m} [M_{\mathcal{D}} \Delta_{V,v_1}^2 f_1]^2 \right)^{1/2} \right\|_{L^p(w_1)} \left\| \left( \sum_{K \times V \in \mathcal{D}} [M_{\mathcal{D}} \Delta_{K \times V}^{k_2, v_2} f_2]^2 \right)^{1/2} \right\|_{L^q(w_2)}, \end{aligned}$$

which is clearly dominated by  $\|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}$ .

There is no essential difference with any of the other cases. If we have  $h_{I_3} \otimes h_{J_3}$  just use (6.4) in the beginning instead. We are done proving the non-averaged case. For the averaged case of the proposition, notice that it follows from the non-averaged case if e.g.  $p = q = 4$  and  $r = 2$  (the point being that  $r > 1$ ). Bilinear extrapolation [15, 22] then gives the averaged version for all exponents.  $\square$

*Partial paraproducts.* We need the following lemma.

**6.7. Lemma.** *Let  $v \in A_\infty(\mathbb{R}^m)$  and  $b \in \text{BMO}(\mathbb{R}^m)$ . Then for a bilinear paraproduct  $\pi_b$  in  $\mathbb{R}^m$  we have*

$$(6.8) \quad \int_{\mathbb{R}^m} |\pi_b(g_1, g_2)| v \lesssim [v]_{A_\infty(\mathbb{R}^m)} \|b\|_{\text{BMO}} \int_{\mathbb{R}^m} (M_{\mathcal{D}^m} g_1 M_{\mathcal{D}^m} g_2) v.$$

*Proof.* Using sparse domination in the form sense for  $\pi_b$ , as stated e.g. in [47], we have

$$\begin{aligned} \int_{\mathbb{R}^m} |\pi_b(g_1, g_2)| v &= \left\langle \pi_b(g_1, g_2), \text{sgn}(\pi_b(g_1, g_2)) v \right\rangle \\ &\lesssim \|b\|_{\text{BMO}} \sum_{Q \in \mathcal{S}} \langle |g_1| \rangle_Q \langle |g_2| \rangle_Q v(Q) \\ &\leq \|b\|_{\text{BMO}} \sum_{Q \in \mathcal{S}} \left( \langle (M_{\mathcal{D}^m} g_1)^{1/2} (M_{\mathcal{D}^m} g_2)^{1/2} \rangle_Q^v \right)^2 v(Q) \\ &\lesssim \|b\|_{\text{BMO}} [v]_{A_\infty(\mathbb{R}^m)} \|(M_{\mathcal{D}^m} g_1)^{1/2} (M_{\mathcal{D}^m} g_2)^{1/2}\|_{L^2(v)}^2, \end{aligned}$$

where in the last step we have used the Carleson embedding theorem.  $\square$

Now we use the  $A_\infty$  extrapolation, Lemma 6.2, in a more crucial way than above. Notice that via Lemma 6.2 we can improve Lemma 6.7 as follows. By extrapolating the estimate (6.8) we get that if  $v \in A_\infty(\mathbb{R}^m)$  and  $b_k \in \text{BMO}(\mathbb{R}^m)$  with  $\|b_k\|_{\text{BMO}(\mathbb{R}^m)} \leq 1$ ,  $k \in \mathbb{Z}$ , then for all  $p \in (0, \infty)$  we have

$$(6.9) \quad \int_{\mathbb{R}^m} \left| \sum_{k \in \mathbb{Z}} \pi_{b_k}(g_{1,k}, g_{2,k}) \right|^p v \leq C([v]_{A_\infty}, p) \int_{\mathbb{R}^m} \left( \sum_{k \in \mathbb{Z}} M_{\mathcal{D}^m} g_{1,k} M_{\mathcal{D}^m} g_{2,k} \right)^p v.$$

Similarly, one concludes from (6.9) that if  $\|b_{j,k}\|_{\text{BMO}(\mathbb{R}^m)} \leq 1$  and  $v \in A_\infty(\mathbb{R}^m)$ , then

$$(6.10) \quad \begin{aligned} & \int_{\mathbb{R}^m} \left( \sum_{j \in \mathbb{Z}} \left| \sum_{k \in \mathbb{Z}} \pi_{b_{j,k}}(g_{1,j,k}, g_{2,j,k}) \right|^p \right)^{q/p} v \\ & \leq C([v]_{A_\infty}, p, q) \int_{\mathbb{R}^m} \left( \sum_{j \in \mathbb{Z}} \left| \sum_{k \in \mathbb{Z}} M_{\mathcal{D}^m} g_{1,j,k} M_{\mathcal{D}^m} g_{2,j,k} \right|^p \right)^{q/p} v \end{aligned}$$

for every  $p, q \in (0, \infty)$ .

We are ready to prove the weighted bound for the partial paraproducts.

**6.11. Proposition.** *Let  $1 < p, q < \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ ,  $w_1 \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$  and  $w_2 \in A_q(\mathbb{R}^n \times \mathbb{R}^m)$  be bi-parameter weights, and set  $v_3 := w_1^{r/p} w_2^{r/q}$ . Suppose that  $(P_\omega)_\omega$  is a collection of partial paraproducts of the same type, where  $P_\omega$  is defined in the grid  $\mathcal{D}_\omega$ . Then*

$$\|\mathbb{E}_\omega P_\omega(f_1, f_2)\|_{L^r(v_3)} \leq C([w_1]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}, [w_2]_{A_q(\mathbb{R}^n \times \mathbb{R}^m)}) \|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}.$$

*Proof.* Similarly as in Proposition 6.5, by bilinear extrapolation it is enough to prove the non-averaged version. We first consider the case that  $P$  has the form

$$P(f_1, f_2) := \sum_{K \in \mathcal{D}^n} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_i^{(k_i)} = K}} h_{I_3}^0 \otimes \pi_{b_{K, (I_i)}}(\langle f_1, h_{I_1} \rangle_1, \langle f_2, h_{I_2} \rangle_1),$$

where  $\pi_{b_{K, (I_i)}}$  is a bilinear paraproduct with

$$\|b_{K, (I_i)}\|_{\text{BMO}(\mathbb{R}^m)} \leq \frac{|I_1|^{1/2} |I_2|^{1/2} |I_3|^{1/2}}{|K|^2} =: a_{K, (I_i)}.$$

In this case, notice that we may apply (6.9) to conclude that

$$(6.12) \quad \begin{aligned} & \|P(f_1, f_2)\|_{L^r(v_3)} \\ & \lesssim \left\| \sum_{K \in \mathcal{D}^n} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_i^{(k_i)} = K}} a_{K, (I_i)} h_{I_3}^0 \otimes M_{\mathcal{D}^m} \langle f_1, h_{I_1} \rangle_1 M_{\mathcal{D}^m} \langle f_2, h_{I_2} \rangle_1 \right\|_{L^r(v_3)} \\ & = \left\| \sum_{K \in \mathcal{D}^n} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_i^{(k_i)} = K}} a_{K, (I_i)} h_{I_3}^0 \otimes \langle \varphi_{\mathcal{D}^n}^1(f_1), h_{I_1} \rangle_1 \langle \varphi_{\mathcal{D}^n}^1(f_2), h_{I_2} \rangle_1 \right\|_{L^r(v_3)}, \end{aligned}$$

where we recall that  $\varphi_{\mathcal{D}^n}^1(f) = \sum_{I \in \mathcal{D}^n} h_I \otimes M_{\mathcal{D}^m} \langle f_1, h_I \rangle_1$ .

For a fixed  $x_2 \in \mathbb{R}^m$  the last function in (6.12) as a function of  $x_1$  is a one-parameter bilinear shift acting on  $\varphi_{\mathcal{D}^n}^1(f_1)(\cdot, x_2)$  and  $\varphi_{\mathcal{D}^n}^1(f_2)(\cdot, x_2)$ . Therefore, the weighted boundedness of one-parameter bilinear shifts implies that  $\|P(f_1, f_2)\|_{L^r(v_3)}$  is dominated by

$$(6.13) \quad \left( \int_{\mathbb{R}^m} \|\varphi_{\mathcal{D}^n}^1(f_1)(\cdot, x_2)\|_{L^p(w_1(\cdot, x_2))}^r \|\varphi_{\mathcal{D}^n}^1(f_2)(\cdot, x_2)\|_{L^q(w_2(\cdot, x_2))}^r dx_2 \right)^{1/r} \\ \lesssim \|\varphi_{\mathcal{D}^n}^1(f_1)\|_{L^p(w_1)} \|\varphi_{\mathcal{D}^n}^1(f_2)\|_{L^q(w_2)} \lesssim \|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}.$$

By symmetry it remains to consider the case when  $P$  has the form

$$P(f_1, f_2) := \sum_{K \in \mathcal{D}^n} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_i^{(k_i)} = K}} h_{I_3} \otimes \pi_{b_{K, (I_i)}} (\langle f_1, h_{I_1}^0 \rangle_1, \langle f_2, h_{I_2} \rangle_1).$$

Applying the lower square function estimate (6.3) and the estimate (6.10), we have similarly as in (6.12) that

$$(6.14) \quad \|P(f_1, f_2)\|_{L^r(v_3)} \\ \lesssim \left\| \left( \sum_{K \in \mathcal{D}^n} \left| \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_i^{(k_i)} = K}} h_{I_3} \otimes \pi_{b_{K, (I_i)}} (\langle f_1, h_{I_1}^0 \rangle_1, \langle f_2, h_{I_2} \rangle_1) \right|^2 \right)^{1/2} \right\|_{L^r(v_3)} \\ \lesssim \left\| \left( \sum_{K \in \mathcal{D}^n} \left| \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_i^{(k_i)} = K}} a_{K, (I_i)} h_{I_3} \otimes \langle M_{\mathcal{D}} f_1, h_{I_1}^0 \rangle_1 \langle \varphi_{\mathcal{D}^n}^1(f_2), h_{I_2} \rangle_1 \right|^2 \right)^{1/2} \right\|_{L^r(v_3)},$$

where we also estimated that  $M_{\mathcal{D}^m} \langle f_1, h_{I_1}^0 \rangle_1 \leq \langle M_{\mathcal{D}} f_1, h_{I_1}^0 \rangle_1$ . With a fixed  $x_2 \in \mathbb{R}^m$ , consider the function in the last line in (6.14) as a function of  $x_1 \in \mathbb{R}^m$ . It is the square function of a one-parameter bilinear shift. Therefore, similarly as in (6.12) and (6.13), we have that the last term in (6.14) is dominated by

$$\|M_{\mathcal{D}} f_1\|_{L^p(w_1)} \|\varphi_{\mathcal{D}^n}^1(f_2)\|_{L^q(w_2)} \lesssim \|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}.$$

We are done.  $\square$

## 6.2. Unweighted boundedness of full paraproducts.

**6.15. Proposition.** *Let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ . Suppose that  $(\Pi_\omega)_\omega$  is a collection of bilinear bi-parameter full paraproducts of the same type, where  $\Pi_\omega$  is defined in the grid  $\mathcal{D}_\omega$ . Then we have*

$$\|\mathbb{E}_\omega \Pi_\omega(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}.$$

*Proof.* The proof of the Banach range boundedness  $p, q, r \in (1, \infty)$  is a standard argument using duality. Suppose now that  $1/p + 1/q = 1/r$  with  $1 < p, q < \infty$  and  $\frac{1}{2} < r < 1$ , and  $E \subset \mathbb{R}^{n+m}$  with  $0 < |E| < \infty$ . We prove that for some  $E' \subset E$  with  $|E'| \geq |E|/2$  we have

$$|\langle \mathbb{E}_\omega \Pi_\omega(f_1, f_2), f_3 \rangle| \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})} |E|^{1/r'}, \quad |f_3| \leq 1_{E'}.$$

We omit the details of this core part of the proof, as later on we have to perform the argument in the commutator setting, which is quite a bit more demanding, and we give the full details there. However, we take this opportunity to explain the completely standard interpolation argument, which can always be used to end the proof after such a weak type estimate. This general scheme is used e.g. in [51].

At this point we know that

$$(6.16) \quad \|\mathbb{E}_\omega \Pi_\omega(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}, \quad \text{if } r > 1,$$

$$(6.17) \quad \|\mathbb{E}_\omega \Pi_\omega(f_1, f_2)\|_{L^{r,\infty}(\mathbb{R}^{n+m})} \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}, \quad \text{if } r < 1.$$

Notice that for  $r = 1$  we may easily get that if  $0 < |E_i| < \infty$ ,  $i = 1, 2, 3$ ,  $|f_1| \leq 1_{E_1}$  and  $|f_2| \leq 1_{E_2}$  then there exists  $E'_3 \subset E_3$  so that  $|E'_3| \geq |E_3|/2$ , and so that for all  $|f_3| \leq 1_{E'_3}$  we have

$$|\langle \mathbb{E}_\omega \Pi_\omega(f_1, f_2), f_3 \rangle| \lesssim |E_1|^{1/p} |E_2|^{1/p'}.$$

This follows by taking convex combinations of the estimates (6.16), (6.17). Then use e.g. Theorem 3.8 (or rather its proof) in Thiele's book [58] to update all of our estimates that are either weak type (if  $r < 1$ ) or restricted weak type (if  $r = 1$ ) into strong type bounds. The cases  $p = \infty$  or  $q = \infty$  follow by duality.  $\square$

6.18. *Remark.* If  $b = b_1 \otimes b_2$ , where  $\|b_1\|_{\text{BMO}(\mathbb{R}^n)}, \|b_2\|_{\text{BMO}(\mathbb{R}^m)} \leq 1$ , we get the weighted estimate for full paraproducts. Denoting  $b_K^1 = \langle b_1, h_K \rangle$ ,  $b_V^2 = \langle b_2, h_V \rangle$  we e.g. have

$$\begin{aligned} & \left\| \sum_{K,V} b_K^1 b_V^2 \langle f_1 \rangle_{K \times V} \langle f_2 \rangle_{K \times V} h_K \otimes h_V \right\|_{L^r(v_3)} \\ & \lesssim \left\| \left( \sum_K |b_K^1|^2 \frac{1_K}{|K|} \otimes \left| \sum_V b_V^2 \langle f_1 \rangle_{K \times V} \langle f_2 \rangle_{K \times V} h_V \right|^2 \right)^{1/2} \right\|_{L^r(v_3)} \\ & \lesssim \left\| \left( \sum_K |b_K^1|^2 \frac{1_K}{|K|} \otimes [M \langle f_1 \rangle_{K,1} M \langle f_2 \rangle_{K,1}]^2 \right)^{1/2} \right\|_{L^r(v_3)} \\ & \lesssim \left( \mathbb{E} \left\| \sum_K \epsilon_K b_K^1 h_K \otimes \langle M^2 f_1 \rangle_{K,1} \langle M^2 f_2 \rangle_{K,1} \right\|_{L^r(v_3)}^r \right)^{1/r} \lesssim \|M^1 M^2 f_1 M^1 M^2 f_2\|_{L^r(v_3)}, \end{aligned}$$

from which the claim follows. We used (6.10) and Kahane–Khinchine. Other forms of full paraproducts are similar but even easier.

## 7. WEIGHTED, QUASI-BANACH AND MIXED-NORM ESTIMATES FOR CZOS

Recall that a bilinear bi-parameter SIO  $T$  is free of full paraproducts if the following holds:  $\langle S(1, 1), h_I \otimes h_J \rangle = 0$  for all  $S \in \{T, T^{*1}, T^{*2}, T_1^{1*}, T_1^{2*}, T_2^{1*}, T_2^{2*}, T_{1,2}^{1*,2*}, T_{2,1}^{1*,2*}\}$  and all cubes  $I \subset \mathbb{R}^n$ ,  $J \subset \mathbb{R}^m$ . We begin with the improved bounds that hold in this case.

**7.1. Theorem.** *Let  $T$  be a bilinear bi-parameter CZO that is free of full paraproducts. Then we have*

$$(7.2) \quad \|T(f_1, f_2)\|_{L^r(v_3)} \lesssim_{[w_1]_{A_p}, [w_2]_{A_q}} \|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}$$

for all  $1 < p, q < \infty$  and  $1/2 < r < \infty$  satisfying  $1/p + 1/q = 1/r$ , and for all bi-parameter weights  $w_1 \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ ,  $w_2 \in A_q(\mathbb{R}^n \times \mathbb{R}^m)$  with  $v_3 := w_1^{r/p} w_2^{r/q}$ .

We also have

$$(7.3) \quad \|T(f_1, f_2)\|_{L^{r_1}(v_{3,1}; L^{r_2}(v_{3,2}))} \lesssim_{[w_{i,j}]} \|f_1\|_{L^{p_1}(w_{1,1}; L^{p_2}(w_{2,1}))} \|f_2\|_{L^{q_1}(w_{2,1}; L^{q_2}(w_{2,2}))}$$

for all  $1 < p_i, q_i < \infty$  and  $1/2 < r_i < \infty$  with  $1/p_i + 1/q_i = 1/r_i$ , for all  $w_{1,1} \in A_{p_1}(\mathbb{R}^n)$ ,  $w_{2,1} \in A_{q_1}(\mathbb{R}^m)$ ,  $w_{1,2} \in A_{p_2}(\mathbb{R}^m)$ ,  $w_{2,2} \in A_{q_2}(\mathbb{R}^m)$  with  $v_{3,1} := w_{1,1}^{r_1/p_1} w_{2,1}^{r_1/q_1}$  and  $v_{3,2} := w_{1,2}^{r_2/p_2} w_{2,2}^{r_2/q_2}$ .

In the unweighted case we have

$$\|T(f_1, f_2)\|_{L^{r_1}(\mathbb{R}^n; L^{r_2}(\mathbb{R}^m))} \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))} \|f_2\|_{L^{q_1}(\mathbb{R}^n; L^{q_2}(\mathbb{R}^m))}$$

for all  $1 < p_i, q_i \leq \infty$  and  $1/2 < r_i < \infty$  with  $1/p_i + 1/q_i = 1/r_i$ , except that if  $r_2 < 1$  we have to assume  $\infty \notin \{p_1, q_1\}$ .

**7.4. Remark.** Let  $1 < u_1, u_2 < \infty$  and  $1/2 < u_3 < \infty$  satisfy  $1/u_1 + 1/u_2 = 1/u_3$ . Recall that weighted bounds imply vector-valued bounds, so we have in the situation of the previous theorem that for instance

$$\left\| \left( \sum_j |T(f_1^j, f_2^j)|^{u_3} \right)^{1/u_3} \right\|_{L^r(v_3)} \lesssim \left\| \left( \sum_j |f_1^j|^{u_1} \right)^{1/u_1} \right\|_{L^p(w_1)} \left\| \left( \sum_j |f_2^j|^{u_2} \right)^{1/u_2} \right\|_{L^q(w_2)},$$

whenever  $1 < p, q < \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ , and  $w_1 \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ ,  $w_2 \in A_q(\mathbb{R}^n \times \mathbb{R}^m)$  with  $v_3 := w_1^{r/p} w_2^{r/q}$ .

*Proof of Theorem 7.1.* The implicit constants are allowed to depend on  $A_p$  characteristics below. Write the pointwise identity

$$(7.5) \quad T(f_1, f_2) = C_T \sum_{\substack{k=(k_1, k_2, k_3) \in \mathbb{Z}_+^3 \\ v=(v_1, v_2, v_3) \in \mathbb{Z}_+^3}} \alpha_{k,v} \sum_u \mathbb{E}_\omega U_{k,u, \mathcal{D}_\omega}^v(f_1, f_2),$$

where the DMOs are shifts or partial paraproducts. The results for averages of model operators in Propositions 6.5 and 6.11 imply (7.2) by using  $\|\sum_i g_i\|_{L^r(v_3)}^r \leq \sum_i \|g_i\|_{L^r(v_3)}^r$ , if  $r < 1$ , or by using the normal triangle inequality otherwise.

We now simply use extrapolation again to get mixed-norm estimates. Suppose  $1/p_2 + 1/q_2 = 1/r_2$ , where  $1 < p_2, q_2 < \infty$ ,  $1/2 < r_2 < \infty$ . Fix some weights  $w_{1,2} \in A_{p_2}(\mathbb{R}^m)$ ,  $w_{2,2} \in A_{q_2}(\mathbb{R}^m)$  and set  $v_{3,2} := w_{1,2}^{r_2/p_2} w_{2,2}^{r_2/q_2}$ . Define

$$H_{r_2, v_{3,2}}(x_1) = H(x_1) := \left( \int_{\mathbb{R}^m} |T(f_1, f_2)(x_1, x_2)|^{r_2} v_{3,2}(x_2) dx_2 \right)^{1/r_2},$$

$$F_{1, p_2, w_{1,2}}(x_1) = F_1(x_1) := \left( \int_{\mathbb{R}^m} |f_1(x_1, x_2)|^{p_2} w_{1,2}(x_2) dx_2 \right)^{1/p_2},$$

and

$$F_{2, q_2, w_{2,2}}(x_1) = F_2(x_1) := \left( \int_{\mathbb{R}^m} |f_2(x_1, x_2)|^{q_2} w_{2,2}(x_2) dx_2 \right)^{1/q_2}.$$

Suppose  $w_{1,1} \in A_{p_2}(\mathbb{R}^n)$ ,  $w_{2,1} \in A_{q_2}(\mathbb{R}^n)$  and define  $v_{3,1} := w_{1,1}^{r_2/p_2} w_{2,1}^{r_2/q_2}$ . Let  $w_1 := w_{1,1} \otimes w_{1,2} \in A_{p_2}(\mathbb{R}^n \times \mathbb{R}^m)$ ,  $w_2 := w_{2,1} \otimes w_{2,2} \in A_{q_2}(\mathbb{R}^n \times \mathbb{R}^m)$  and define the related weight  $v_3 := w_1^{r_2/p_2} w_2^{r_2/q_2}$ .

Applying (7.2) with  $p = p_2$ ,  $q = q_2$ ,  $r = r_2$  and with the weights  $w_1, w_2$  and  $v_3$  we have that  $\|H\|_{L^{r_2}(v_{3,1})} \lesssim \|F_1\|_{L^{p_2}(w_{1,1})} \|F_2\|_{L^{q_2}(w_{2,1})}$ . Because this holds for all weights  $w_{1,1} \in A_{p_2}(\mathbb{R}^n)$  and  $w_{2,1} \in A_{q_2}(\mathbb{R}^n)$  with  $v_{3,1} = w_{1,1}^{r_2/p_2} w_{2,1}^{r_2/q_2}$ , then bilinear extrapolation (see e.g. [15]) implies that

$$(7.6) \quad \|H\|_{L^{r_1}(v_{3,1})} \lesssim \|F_1\|_{L^{p_1}(w_{1,1})} \|F_2\|_{L^{q_1}(w_{2,1})}$$

holds for all exponents  $1 < p_1, q_1 < \infty$ ,  $1/2 < r_1 < \infty$ , and all weights  $w_{1,1} \in A_{p_1}(\mathbb{R}^n)$ ,  $w_{2,1} \in A_{q_1}(\mathbb{R}^n)$  and  $v_{3,1} := w_{1,1}^{r_1/p_1} w_{2,1}^{r_1/q_1}$ . But (7.6) is exactly what is claimed in (7.3).

We now discuss the unweighted mixed-norm estimates, where we can allow some of the exponents  $p_1, q_1, p_2, q_2$  to be  $\infty$ . We fix exponents  $p_2, q_2 \in (1, \infty]$  such that  $1/p_2 + 1/q_2 = 1/r_2 \in (0, 1]$ .

Let  $U_{k,u,\mathcal{D}_\omega}^v$  be a model operator from the representation, and write  $U_{k,u,\mathcal{D}_\omega}^v(f_1, f_2)$  in the form

$$\sum_{\substack{K \in \mathcal{D}_{\omega_1}^n \\ V \in \mathcal{D}_{\omega_2}^m}} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ I_i^{(k_i)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}_{\omega_2}^m \\ J_i^{(v_i)} = V}} a_{K,V,(I_i),(J_j)} \langle f_1, \tilde{h}_{I_1} \otimes \tilde{h}_{J_1} \rangle \langle f_2, \tilde{h}_{I_2} \otimes \tilde{h}_{J_2} \rangle \tilde{h}_{I_3} \otimes \tilde{h}_{J_3},$$

where  $\tilde{h}_I \in \{h_I, h_I^0\}$  and  $\tilde{h}_J \in \{h_J, h_J^0\}$ . Think of  $U_{k,u,\mathcal{D}_\omega}^v$  as a dyadic model operator in  $\mathbb{R}^n$  with operator coefficients in the sense that

$$\langle U_{k,u,\mathcal{D}_\omega}^v(f_1, f_2), f_3 \rangle = \sum_{K \in \mathcal{D}_{\omega_1}^n} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ I_i^{(k_i)} = K}} \langle b_{K,(I_i)}(\langle f_1, \tilde{h}_{I_1} \rangle_1, \langle f_2, \tilde{h}_{I_2} \rangle_1), \langle f_3, \tilde{h}_{I_3} \rangle_1 \rangle_{\mathbb{R}^m},$$

where  $\langle \cdot, \cdot \rangle_{\mathbb{R}^m}$  refers to the integral pairing in  $\mathbb{R}^m$  and  $b_{K,(I_i)}$  is the one-parameter model operator determined by the family of coefficients  $\{a_{K,V,(I_i),(J_j)}\}_{V,(J_j)}$  and the Haar functions  $\tilde{h}_{J_j}$ . The normalization of  $b_{K,(I_i)}$  depends on  $K$  and  $(I_i)$ .

Using this viewpoint one can adapt the bilinear sparse domination argument from Section 5 of [47] to see that if  $f_i \in L_c^\infty(\mathbb{R}^{n+m})$  then there exists a sparse collection  $\mathcal{S} = \mathcal{S}(f_1, f_2, f_3, p_2, q_2)$  of cubes in  $\mathbb{R}^n$  so that

$$(7.7) \quad |\langle U_{k,u,\mathcal{D}_\omega}^v(f_1, f_2), f_3 \rangle| \lesssim (1 + \max k_i) \sum_{Q \in \mathcal{S}} \langle \|f_1\|_{p_2} \rangle_Q \langle \|f_2\|_{q_2} \rangle_Q \langle \|f_3\|_{r'_2} \rangle_Q |Q|.$$

This uses the fact that if  $P$  is a one-parameter or a bi-parameter model operator, then

$$\|P(f_1, f_2)\|_{L^{r_2}} \lesssim \|f_1\|_{L^{p_2}} \|f_2\|_{L^{q_2}}.$$

The case  $p_2, q_2 < \infty$  has been proved above (or in [47]) and for example the case  $q_2 = \infty$  is interpreted via the dual estimate

$$(7.8) \quad \|P^{2*}(f_1, f_2)\|_{L^1} \lesssim \|f_1\|_{L^{p_2}} \|f_2\|_{L^{p'_2}}.$$

The same collection  $\mathcal{S}$  in (7.7) can be made to work for all model operators  $U_{k,u,\mathcal{D}_\omega}^v$ , see Lacey–Mena [40] or [47]. Related to this notice that  $\mathcal{S}$  is not claimed to be a subset of  $\mathcal{D}_{\omega_1}^n$ .

We infer using the representation theorem that for  $f_i \in L_c^\infty(\mathbb{R}^{n+m})$  there exists a sparse collection  $\mathcal{S} = \mathcal{S}(f_1, f_2, f_3, p_2, q_2)$  of cubes in  $\mathbb{R}^n$  so that

$$(7.9) \quad |\langle T(f_1, f_2), f_3 \rangle| \lesssim \sum_{Q \in \mathcal{S}} \langle \|f_1\|_{p_2} \rangle_Q \langle \|f_2\|_{q_2} \rangle_Q \langle \|f_3\|_{r'_2} \rangle_Q |Q|.$$

We can extend (7.9) to hold also when, for example,  $f_2 \in L^\infty(\mathbb{R}^{n+m})$  and  $f_1, f_3 \in L_c^\infty(\mathbb{R}^{n+m})$ . To see this, let  $R_1 \subset R_2 \subset \dots$  be a sequence of rectangles so that  $\bigcup R_k = \mathbb{R}^{n+m}$ . Using that  $T^{2*}(f_1, f_3) \in L^1$  we have

$$\langle T(f_1, f_2), f_3 \rangle := \langle T^{2*}(f_1, f_3), f_2 \rangle = \lim_{k \rightarrow \infty} \langle T^{2*}(f_1, f_3), 1_{R_k} f_2 \rangle = \lim_{k \rightarrow \infty} \langle T(f_1, 1_{R_k} f_2), f_3 \rangle.$$

Let  $k$  be so large that  $2|\langle T(f_1, 1_{R_k} f_2), f_3 \rangle| \geq |\langle T(f_1, f_2), f_3 \rangle|$ . Use the sparse estimate to  $|\langle T(f_1, 1_{R_k} f_2), f_3 \rangle|$  and then in the right hand side estimate up by omitting the indicator.

Let  $M(g_1, g_2) := \sup_Q \langle |g_1| \rangle_Q \langle |g_2| \rangle_Q 1_Q$  be the bilinear maximal function in  $\mathbb{R}^n$ . The sparse estimate implies similarly as in Lemma 6.7 that if  $f_1, f_2 \in L_c^\infty(\mathbb{R}^{n+m})$  then

$$\|T(f_1, f_2)\|_{L^1(\rho; L^{r_2}(\mathbb{R}^m))} \lesssim \|M(\|f_1\|_{L^{p_2}(\mathbb{R}^m)}, \|f_2\|_{L^{q_2}(\mathbb{R}^m)})\|_{L^1(\rho)}$$

holds for all  $\rho \in A_\infty(\mathbb{R}^n)$ . Now  $A_\infty$  extrapolation implies that if  $f_1, f_2 \in L_c^\infty(\mathbb{R}^{n+m})$ ,  $p_1, q_1 \in (1, \infty]$  and  $1/r_1 := 1/p_1 + 1/q_1 > 0$ , then

$$(7.10) \quad \begin{aligned} \|T(f_1, f_2)\|_{L^{r_1}(\mathbb{R}^n; L^{r_2}(\mathbb{R}^m))} &\lesssim \|M(\|f_1\|_{L^{p_2}(\mathbb{R}^m)}, \|f_2\|_{L^{q_2}(\mathbb{R}^m)})\|_{L^{r_1}(\mathbb{R}^n)} \\ &\lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))} \|f_2\|_{L^{q_1}(\mathbb{R}^n; L^{q_2}(\mathbb{R}^m))}. \end{aligned}$$

We first treat the cases where  $p_1, p_2 < \infty$ . If also  $q_1, q_2 < \infty$ , there is nothing to do by density. Since in the sparse estimate we were able to have that  $f_1, f_3 \in L_c^\infty(\mathbb{R}^{n+m})$  and  $f_2 \in L^\infty(\mathbb{R}^{n+m})$ , we can allow  $f_1 \in L_c^\infty(\mathbb{R}^{n+m})$  and  $f_2 \in L_c^\infty(\mathbb{R}^n; L^\infty(\mathbb{R}^m))$  in (7.10) if  $p_1, p_2, q_1 < \infty$  and  $q_2 = \infty$  or  $f_1 \in L_c^\infty(\mathbb{R}^{n+m})$  and  $f_2 \in L^\infty(\mathbb{R}^{n+m})$  if  $p_1, p_2 < \infty$  and  $q_1 = q_2 = \infty$ . We can then again extend by density. The case  $p_1, p_2, q_2 < \infty$  and  $q_1 = \infty$  is symmetric to the case below, where  $p_1 = \infty$  and  $p_2, q_1, q_2 < \infty$ .

It only remains to consider the case  $p_1 = \infty$  or  $p_2 = \infty$ . But the case  $p_2 = \infty$  is symmetric with the case  $q_2 = \infty$ , so we only deal with the case  $p_1 = \infty$  and  $p_2 < \infty$ . In particular, we already know at this point that everything works if  $p_2 = \infty$  and  $p_1, q_1, q_2 < \infty$ . This is used below when  $r_2 = 1$ .

So let finally  $p_1 = \infty$ ,  $q_1 = r_1 \in (1, \infty)$  and  $p_2 < \infty$ . Assume first that  $r_2 > 1$ . If  $f_1 \in L^\infty(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))$  and  $f_2 \in L^{q_1}(\mathbb{R}^n; L^{q_2}(\mathbb{R}^m))$ , then we define  $T(f_1, f_2) \in L^{q_1}(\mathbb{R}^n; L^{r_2}(\mathbb{R}^m))$  by setting for  $f_3 \in L^{q_1'}(\mathbb{R}^n; L^{r_2'}(\mathbb{R}^m))$  that

$$\langle T(f_1, f_2), f_3 \rangle := \langle T^{1*}(f_3, f_2), f_1 \rangle.$$

Notice that from above we know that  $T^{1*}$  maps  $L^{q_1'}(\mathbb{R}^n; L^{r_2'}(\mathbb{R}^m)) \times L^{q_1}(\mathbb{R}^n; L^{q_2}(\mathbb{R}^m))$  into  $L^1(\mathbb{R}^n; L^{p_2'}(\mathbb{R}^m))$ . Assume then that  $r_2 = 1$ , which implies that  $q_2 < \infty$ . Let  $f_1 \in L^\infty(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))$  and  $f_2 \in L_c^\infty(\mathbb{R}^{n+m})$ . By thinking of  $f_2$  as a function in  $L^s(\mathbb{R}^{n+m})$  for some sufficiently large  $s \in (1, \infty)$ , we know by the previous case of  $r_2 > 1$  that  $T(f_1, f_2)$  is well-defined as a function. If  $f_3 \in L_c^\infty(\mathbb{R}^{n+m})$ , then

$$\begin{aligned} |\langle T(f_1, f_2), f_3 \rangle| &= |\langle T^{1*}(f_3, f_2), f_1 \rangle| \\ &\lesssim \|f_3\|_{L^{q_1'}(\mathbb{R}^n; L^\infty(\mathbb{R}^m))} \|f_2\|_{L^{q_1}(\mathbb{R}^n; L^{q_2}(\mathbb{R}^m))} \|f_1\|_{L^\infty(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))}, \end{aligned}$$

which implies that  $T(f_1, f_2) \in L^{q_1}(\mathbb{R}^n; L^1(\mathbb{R}^m))$  with the right estimate. We are done.  $\square$

We also have the following unweighted estimates for all bilinear bi-parameter CZOs.

**7.11. Theorem.** *Let  $T$  be a bilinear bi-parameter CZO. Then we have*

$$\|T(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}$$

for all  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfying  $1/p + 1/q = 1/r$ . Moreover, we have the mixed-norm estimates

$$\|T(f_1, f_2)\|_{L^{r_1}(\mathbb{R}^n; L^{r_2}(\mathbb{R}^m))} \lesssim \|f_1\|_{L^{p_1}(\mathbb{R}^n; L^{p_2}(\mathbb{R}^m))} \|f_2\|_{L^{q_1}(\mathbb{R}^n; L^{q_2}(\mathbb{R}^m))}$$

for all  $1 < p_i, q_i \leq \infty$  with  $1/p_i + 1/q_i = 1/r_i$ ,  $i = 1, 2$ ,  $1/2 < r_1 < \infty$ ,  $1 \leq r_2 < \infty$ .

*Proof.* Write again the pointwise identity (7.5), which this time contains all model operators. Then argue as before but also use Proposition 6.15. This gives the desired non-mixed estimate in the full range, which can be lifted to the claimed mixed-norm estimate by exactly the same sparse domination argument as above.  $\square$

## 8. DECOMPOSITIONS AND UPPER BOUNDS FOR FIRST ORDER COMMUTATORS

We begin by considering first order commutators.

**8.1. Theorem.** *Let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ , and let  $b \in \text{bmo}(\mathbb{R}^{n+m})$ . Suppose  $T$  is a bilinear bi-parameter CZO. Then we have*

$$\|[b, T]_1(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|b\|_{\text{bmo}(\mathbb{R}^{n+m})} \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})},$$

and similarly for  $[b, T]_2$ .

*Proof.* Using Theorem 5.1 write the pointwise identity

$$[b, T]_1(f_1, f_2) = C_T \sum_{\substack{k=(k_1, k_2, k_3) \in \mathbb{Z}_+^3 \\ v=(v_1, v_2, v_3) \in \mathbb{Z}_+^3}} \alpha_{k,v} \sum_u \mathbb{E}_\omega [b, U_{k,u, \mathcal{D}_\omega}^v]_1(f_1, f_2).$$

Averages of commutators of model operators map in the full range with a bound polynomial in complexity – this is proved carefully below (see Propositions 8.14, 8.17 and 8.21). Using this we get the claim.  $\square$

The corresponding results for iterated commutators are proved in Section 9.

**8.1. Martingale difference expansions of products.** The idea is that a product  $bf$  paired with Haar functions is expanded in the bi-parameter fashion only if both of the Haar functions are cancellative. In a mixed situation we expand only in  $\mathbb{R}^n$  or  $\mathbb{R}^m$ , and in the remaining fully non-cancellative situation we do not expand at all. When pairing with a non-cancellative Haar function we add and subtract a suitable average of  $b$ .

Let  $\mathcal{D}^n$  and  $\mathcal{D}^m$  be some fixed dyadic grids in  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively, and write  $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ . In what follows we sum over  $I \in \mathcal{D}^n$  and  $J \in \mathcal{D}^m$ .

*Paraproduct operators.* Let us first define certain paraproduct operators:

$$\begin{aligned} A_1(b, f) &= \sum_{I, J} \Delta_{I \times J} b \Delta_{I \times J} f, & A_2(b, f) &= \sum_{I, J} \Delta_{I \times J} b E_I^1 \Delta_J^2 f, \\ A_3(b, f) &= \sum_{I, J} \Delta_{I \times J} b \Delta_I^1 E_J^2 f, & A_4(b, f) &= \sum_{I, J} \Delta_{I \times J} b \langle f \rangle_{I \times J}, \end{aligned}$$

and

$$\begin{aligned} A_5(b, f) &= \sum_{I, J} E_I^1 \Delta_J^2 b \Delta_{I \times J} f, & A_6(b, f) &= \sum_{I, J} E_I^1 \Delta_J^2 b \Delta_I^1 E_J^2 f, \\ A_7(b, f) &= \sum_{I, J} \Delta_I^1 E_J^2 b \Delta_{I \times J} f, & A_8(b, f) &= \sum_{I, J} \Delta_I^1 E_J^2 b E_I^1 \Delta_J^2 f. \end{aligned}$$

The operators are grouped into two collections, since they are handled differently (using product BMO or little BMO estimates, respectively). Also recall that  $\text{bmo} \subset \text{BMO}_{\text{prod}}$ . When the underlying grid needs to be written, we write  $A_{i, \mathcal{D}}(b, f)$ . When desired, these operators can be written with Haar functions using the standard formulas. Recall that

then one has to be slightly careful when a term like  $h_I h_I$  or  $h_J h_J$  appears (as they do e.g. when expanding  $A_1$  using Haar functions). This really can be of the form  $h_I^{\epsilon_1} h_I^{\epsilon_2}$  for possibly different  $\epsilon_1, \epsilon_2$ . However, the only property we will use is that  $|h_I h_I| = 1_I/|I|$ , i.e. we always treat such products as non-cancellative objects.

We also define

$$a_1^1(b, f) = \sum_I \Delta_I^1 b \Delta_I^1 f \quad \text{and} \quad a_2^1(b, f) = \sum_I \Delta_I^1 b E_I^1 f.$$

The operators  $a_1^2(b, f)$  and  $a_2^2(b, f)$  are defined analogously. Again, we can also e.g. write  $a_{1, \mathcal{D}^n}^1(b, f)$  to emphasise the underlying dyadic grid.

**8.2. Lemma.** *Let  $H_b$  be  $A_i(b, \cdot)$ ,  $i = 1, \dots, 8$ , or  $a_j^1(b, \cdot)$ ,  $a_j^2(b, \cdot)$ ,  $j = 1, 2$ . Let also  $b \in \text{bmo}(\mathbb{R}^{n+m})$ ,  $p \in (1, \infty)$  and  $w \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ . Then we have*

$$\|H_b f\|_{L^p(w)} \lesssim C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|b\|_{\text{bmo}(\mathbb{R}^{n+m})} \|f\|_{L^p(w)}.$$

*Proof.* The proofs follow by completely standard arguments using  $H^1$ -BMO type duality estimates. See e.g. [30] (the operators  $a_j^1(b, \cdot)$ ,  $a_j^2(b, \cdot)$  do not appear in [30], and the operators  $A_i(b, \cdot)$  appear in a bit different Haar function forms).  $\square$

Let now  $f \in L^p(\mathbb{R}^{n+m})$  for some  $p \in (1, \infty)$ , and  $b \in \text{bmo}(\mathbb{R}^{n+m})$ . We know that  $b \in L_{\text{loc}}^p(\mathbb{R}^{n+m})$  by the John–Nirenberg valid for little BMO. For  $I_0 \in \mathcal{D}^n$  and  $J_0 \in \mathcal{D}^m$  we will now introduce our expansions of  $\langle bf, h_{I_0} \otimes h_{J_0} \rangle$ ,  $\langle bf, h_{I_0} \otimes \frac{1_{J_0}}{|J_0|} \rangle$  and  $\langle bf \rangle_{I_0 \times J_0}$ .

We first consider  $\langle bf, h_{I_0} \otimes h_{J_0} \rangle$ . There holds

$$1_{I_0 \times J_0} b = \sum_{\substack{I_1 \times J_1 \in \mathcal{D} \\ I_1 \times J_1 \subset I_0 \times J_0}} \Delta_{I_1 \times J_1} b + \sum_{\substack{J_1 \in \mathcal{D}^m \\ J_1 \subset J_0}} E_{I_0}^1 \Delta_{J_1}^2 b + \sum_{\substack{I_1 \in \mathcal{D}^n \\ I_1 \subset I_0}} \Delta_{I_1}^1 E_{J_0}^2 b + E_{I_0 \times J_0} b.$$

Let us denote these terms by  $I_j$ ,  $j = 1, 2, 3, 4$ , in the respective order. We have the corresponding decomposition of  $f$ , whose terms we denote by  $II_i$ ,  $i = 1, 2, 3, 4$ . Calculating carefully the pairings  $\langle I_j II_i, h_{I_0} \otimes h_{J_0} \rangle$  we see that

$$(8.3) \quad \langle bf, h_{I_0} \otimes h_{J_0} \rangle = \sum_{i=1}^8 \langle A_i(b, f), h_{I_0} \otimes h_{J_0} \rangle + \langle b \rangle_{I_0 \times J_0} \langle f, h_{I_0} \otimes h_{J_0} \rangle.$$

We now consider  $\langle bf, h_{I_0} \otimes \frac{1_{J_0}}{|J_0|} \rangle$ . This time we write  $1_{I_0} b = \sum_{\substack{I_1 \in \mathcal{D}^n \\ I_1 \subset I_0}} \Delta_{I_1}^1 b + E_{I_0}^1 b$ , and similarly for  $f$ . Calculating  $\langle bf, h_{I_0} \rangle_1$  we see that

$$(8.4) \quad \begin{aligned} \left\langle bf, h_{I_0} \otimes \frac{1_{J_0}}{|J_0|} \right\rangle &= \sum_{i=1}^2 \left\langle a_i^1(b, f), h_{I_0} \otimes \frac{1_{J_0}}{|J_0|} \right\rangle \\ &\quad + \langle (\langle b \rangle_{I_0, 1} - \langle b \rangle_{I_0 \times J_0}) \langle f, h_{I_0} \rangle_1 \rangle_{J_0} + \langle b \rangle_{I_0 \times J_0} \left\langle f, h_{I_0} \otimes \frac{1_{J_0}}{|J_0|} \right\rangle. \end{aligned}$$

When we have  $\langle bf \rangle_{I_0 \times J_0}$  we do not expand at all:

$$(8.5) \quad \langle bf \rangle_{I_0 \times J_0} = \langle (b - \langle b \rangle_{I_0 \times J_0}) f \rangle_{I_0 \times J_0} + \langle b \rangle_{I_0 \times J_0} \langle f \rangle_{I_0 \times J_0}.$$

All of our commutators are simply decomposed using (8.3), (8.4) (and its symmetric form) and (8.5) whenever the relevant pairings/averages appear. This splits the commutator into several parts. Only the parts which come from the last terms in (8.3), (8.4) and

(8.5) need to be combined with another part of the commutator of the same form. All the other parts are handled separately.

**8.2. Adapted maximal functions.** For  $b \in \text{BMO}(\mathbb{R}^n)$  and  $f: \mathbb{R}^n \rightarrow \mathbb{C}$  define

$$M_b f = \sup_I \frac{1_I}{|I|} \int_I |b - \langle b \rangle_I| |f|.$$

In the situation  $b \in \text{bmo}(\mathbb{R}^{n+m})$  and  $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  define  $M_b$  analogously but take the supremum over all rectangles. The dyadic variants could also be defined, and denoted by  $M_{\mathcal{D}^n, b}$  and  $M_{\mathcal{D}, b}$ . For a little BMO function  $b \in \text{bmo}(\mathbb{R}^{n+m})$  define

$$\varphi_{\mathcal{D}^m, b}^2(f) = \sum_{J \in \mathcal{D}^m} M_{\langle b \rangle_{J,2}} \langle f, h_J \rangle_2 \otimes h_J,$$

and similarly define  $\varphi_{\mathcal{D}^n, b}^1(f)$ . For our later usage it is important to not to use the dyadic variant  $M_{\mathcal{D}^n, \langle b \rangle_{J,2}}$ , as it would induce an unwanted dependence on  $\mathcal{D}^n$  (which has relevance in some randomisation considerations).

**8.6. Lemma.** *Suppose  $\|b_i\|_{\text{BMO}(\mathbb{R}^n)} \leq 1$ ,  $1 < u, p < \infty$  and  $w \in A_p(\mathbb{R}^n)$ . Then we have*

$$(8.7) \quad \left\| \left( \sum_i [M_{b_i} f_i]^u \right)^{1/u} \right\|_{L^p(w)} \lesssim C([w]_{A_p(\mathbb{R}^n)}) \left\| \left( \sum_i |f_i|^u \right)^{1/u} \right\|_{L^p(w)}.$$

*The same bound holds with  $\|b_i\|_{\text{bmo}(\mathbb{R}^n \times \mathbb{R}^m)} \leq 1$  and  $w \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ . For a function  $b$  with  $\|b\|_{\text{bmo}(\mathbb{R}^n \times \mathbb{R}^m)} \leq 1$  we also have*

$$\|\varphi_{\mathcal{D}^m, b}^2(f)\|_{L^p(w)} \leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|f\|_{L^p(w)}, \quad 1 < p < \infty, w \in A_p(\mathbb{R}^n \times \mathbb{R}^m).$$

*Proof.* We begin by proving the bound  $\|M_b f\|_{L^p(w)} \leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|f\|_{L^p(w)}$  – the proof is the same in the one-parameter case. Fix  $w \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$  and choose  $s = s([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \in (1, p)$  so that  $[w]_{A_{p/s}(\mathbb{R}^n \times \mathbb{R}^m)} \leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)})$ . This can be done using the reverse Hölder inequality – the well-known bi-parameter version is stated and proved e.g. in Proposition 2.2. of [30]. Using Hölder's inequality and the John–Nirenberg for little bmo we get that  $M_b f \leq C(s) M_s f = C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) M_s f$ . Now, using that  $M: L^q(v) \rightarrow L^q(v)$  for all  $q \in (1, \infty)$  and  $v \in A_q(\mathbb{R}^n \times \mathbb{R}^m)$  we have

$$\|M_b f\|_{L^p(w)} \leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|M|f|^s\|_{L^{p/s}(w)}^{1/s} \leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|f\|_{L^p(w)}.$$

The bi-parameter version of (8.7) (and (8.7) itself) now follow by extrapolation.

Next, using the estimate (8.7) and the fact that  $\|\langle b \rangle_{J,2}\|_{\text{BMO}(\mathbb{R}^n)} \lesssim 1$  we get

$$\begin{aligned} \|\varphi_{\mathcal{D}^m, b}^2(f)\|_{L^p(w)} &\leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \left\| \left( \sum_J (M_{\langle b \rangle_{J,2}} \langle f, h_J \rangle_2)^2 \otimes \frac{1_J}{|J|} \right)^{1/2} \right\|_{L^p(w)} \\ &\leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \left\| \left( \sum_J |\langle f, h_J \rangle_2|^2 \otimes \frac{1_J}{|J|} \right)^{1/2} \right\|_{L^p(w)} \\ &\leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|f\|_{L^p(w)}. \end{aligned}$$

□

We collect a few trivial, but useful, lemmas below.

**8.8. Lemma.** For  $I \in \mathcal{D}^n$  and  $J \in \mathcal{D}^m$  we have

$$|\langle (\langle b \rangle_{J,2} - \langle b \rangle_{I \times J}) \langle f, h_J \rangle_2 \rangle_I| \leq \left\langle \varphi_{\mathcal{D}^m, b}^2(f), \frac{1_I}{|I|} \otimes h_J \right\rangle$$

and

$$|\langle (b - \langle b \rangle_{I \times J}) f \rangle_{I \times J}| \lesssim \langle M_b f \rangle_{I \times J}.$$

*Proof.* There holds

$$|\langle (\langle b \rangle_{J,2} - \langle b \rangle_{I \times J}) \langle f, h_J \rangle_2 \rangle_I| \leq \langle M_{\langle b \rangle_{J,2}} \langle f, h_J \rangle_2 \rangle_I = \left\langle \varphi_{\mathcal{D}^m, b}^2(f), \frac{1_I}{|I|} \otimes h_J \right\rangle,$$

where the last inequality follows from orthogonality. The second claimed inequality is even more immediate.  $\square$

The validity of the following lemma is obvious.

**8.9. Lemma.** Suppose  $I^{(i)} = Q^{(q)} = K$  and  $J^{(j)} = R^{(r)} = V$ . If  $\|b\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$  then we have  $|\langle b \rangle_{Q \times R} - \langle b \rangle_{I \times J}| \lesssim \max(i, j, q, r)$ .

**8.3. Banach range boundedness of commutators of DMOs.** For the Banach range theory of commutators we only need the fact that all the DMOs from Section 4 are of the following general type. Fix dyadic grids  $\mathcal{D}^n$  and  $\mathcal{D}^m$ . Let  $U = U_k^v$ ,  $0 \leq k_i \in \mathbb{Z}$  and  $0 \leq v_i \in \mathbb{Z}$ ,  $i = 1, 2, 3$ , be a bilinear bi-parameter operator such that

$$\begin{aligned} \langle U(f_1, f_2), f_3 \rangle &= \sum_{\substack{K \in \mathcal{D}^n \\ V \in \mathcal{D}^m}} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_1^{(k_1)} = I_2^{(k_2)} = I_3^{(k_3)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}^m \\ J_1^{(v_1)} = J_2^{(v_2)} = J_3^{(v_3)} = V}} a_{K, V, (I_i), (J_j)} \\ &\quad \times \langle f_1, \tilde{h}_{I_1} \otimes \tilde{h}_{J_1} \rangle \langle f_2, \tilde{h}_{I_2} \otimes \tilde{h}_{J_2} \rangle \langle f_3, \tilde{h}_{I_3} \otimes \tilde{h}_{J_3} \rangle, \end{aligned}$$

where  $a_{K, V, (I_i), (J_j)}$  are constants and for all  $i = 1, 2, 3$  we have  $\tilde{h}_{I_i} = h_{I_i}$  for all  $I_i \in \mathcal{D}^n$  or  $\tilde{h}_{I_i} = h_{I_i}^0$  for all  $I_i \in \mathcal{D}^n$ , and similarly with the functions  $\tilde{h}_{J_j}$ . We assume that for all  $p, q, r \in (1, \infty)$  with  $1/p + 1/q = 1/r$  we have

$$\begin{aligned} (8.10) \quad &\sum_{\substack{K \in \mathcal{D}^n \\ V \in \mathcal{D}^m}} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_1^{(k_1)} = I_2^{(k_2)} = I_3^{(k_3)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}^m \\ J_1^{(v_1)} = J_2^{(v_2)} = J_3^{(v_3)} = V}} |a_{K, V, (I_i), (J_j)}| \\ &\quad \times \langle f_1, \tilde{h}_{I_1} \otimes \tilde{h}_{J_1} \rangle \langle f_2, \tilde{h}_{I_2} \otimes \tilde{h}_{J_2} \rangle \langle f_3, \tilde{h}_{I_3} \otimes \tilde{h}_{J_3} \rangle \\ &\lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})} \|f_3\|_{L^{r'}(\mathbb{R}^{n+m})}. \end{aligned}$$

We do not assume anything else about the constants  $a_{K, V, (I_i), (J_j)}$ . In particular,  $U$  can be a bilinear bi-parameter shift, a partial paraproduct or a full paraproduct.

The proof of the following Banach range boundedness result is now surprisingly easy.

**8.11. Proposition.** Let  $p, q, r \in (1, \infty)$ ,  $1/p + 1/q = 1/r$ ,  $0 \leq k_i \in \mathbb{Z}$  and  $0 \leq v_i \in \mathbb{Z}$ ,  $i = 1, 2, 3$ . Let  $U = U_k^v$  be a general bilinear bi-parameter DMO satisfying (8.10). In particular,  $U$  can be a bilinear bi-parameter shift, a partial paraproduct or a full paraproduct. Then for  $b$  such that  $\|b\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$  we have

$$\| [b, U]_1(f_1, f_2) \|_{L^r(\mathbb{R}^{n+m})} \lesssim (1 + \max(k_i, v_i)) \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}.$$

*Proof.* We separately treat the different possible combinations of cancellative and non-cancellative Haar functions. The proof depends only on what Haar functions we have paired with  $f_1$  and  $f_3$  in  $\langle U(f_1, f_2), f_3 \rangle$ . All the model operators fall into one of the following cases:

- (1) We have  $\langle f_1, h_{I_1} \otimes h_{J_1} \rangle \langle f_3, h_{I_3} \otimes h_{J_3} \rangle$ .
- (2) We have  $\langle f_1, h_{I_1}^0 \otimes h_{J_1} \rangle \langle f_3, h_{I_3} \otimes h_{J_3} \rangle$ , or one of the three other symmetric cases.
- (3) We have  $\langle f_1, h_{I_1}^0 \otimes h_{J_1} \rangle \langle f_3, h_{I_3} \otimes h_{J_3}^0 \rangle$ , or the symmetric case.
- (4) We have  $\langle f_1, h_{I_1}^0 \otimes h_{J_1}^0 \rangle \langle f_3, h_{I_3} \otimes h_{J_3} \rangle$ , or the symmetric case.
- (5) We have  $\langle f_1, h_{I_1}^0 \otimes h_{J_1} \rangle \langle f_3, h_{I_3}^0 \otimes h_{J_3} \rangle$ , or the symmetric case.
- (6) We have  $\langle f_1, h_{I_1}^0 \otimes h_{J_1}^0 \rangle \langle f_3, h_{I_3}^0 \otimes h_{J_3} \rangle$ , or one of the other three symmetric cases.
- (7) We have  $\langle f_1, h_{I_1}^0 \otimes h_{J_1} \rangle \langle f_3, h_{I_3}^0 \otimes h_{J_3}^0 \rangle$ .

**Case 1.** Expand using (8.3). Using Lemma 8.9, the boundedness property (8.10) and the boundedness of the operators  $A_i(b, \cdot)$ ,  $i = 1, \dots, 8$ , we have that

$$|\langle [b, U]_1(f_1, f_2), f_3 \rangle| \lesssim \max(k_i, v_i) \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})} \|f_3\|_{L^{r'}(\mathbb{R}^{n+m})}.$$

**Case 2.** This time we expand using (8.3) and (8.4). Then we use Lemma 8.8, Lemma 8.9, the boundedness property (8.10), the boundedness of the operators  $A_i(b, \cdot)$ ,  $i = 1, \dots, 8$ , the boundedness of the operators  $a_i^1(b, \cdot)$ ,  $a_i^2(b, \cdot)$ ,  $i = 1, 2$ , and Lemma 8.6. This gives us the desired bound. For example, one calculates like

$$\begin{aligned} & \sum_{\substack{K \in \mathcal{D}^n \\ V \in \mathcal{D}^m}} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_1^{(k_1)} = I_2^{(k_2)} = I_3^{(k_3)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}^m \\ J_1^{(v_1)} = J_2^{(v_2)} = J_3^{(v_3)} = V}} |a_{K, V, (I_i), (J_j)}| \\ & \times |\langle \langle b \rangle_{J_1, 2} - \langle b \rangle_{I_1 \times J_1} \rangle \langle f_1, h_{J_1} \rangle_2, h_{I_1}^0 \rangle| |\langle f_2, \tilde{h}_{I_2} \otimes \tilde{h}_{J_2} \rangle| |\langle f_3, h_{I_3} \otimes h_{J_3} \rangle| \\ & \lesssim \sum_{\substack{K \in \mathcal{D}^n \\ V \in \mathcal{D}^m}} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}^n \\ I_1^{(k_1)} = I_2^{(k_2)} = I_3^{(k_3)} = K}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}^m \\ J_1^{(v_1)} = J_2^{(v_2)} = J_3^{(v_3)} = V}} |a_{K, V, (I_i), (J_j)}| \\ & \times \langle \varphi_{\mathcal{D}^m, b}^2(f_1), h_{I_1}^0 \otimes h_{J_1} \rangle |\langle f_2, \tilde{h}_{I_2} \otimes \tilde{h}_{J_2} \rangle| |\langle f_3, h_{I_3} \otimes h_{J_3} \rangle| \\ & \lesssim \|\varphi_{\mathcal{D}^m, b}^2(f_1)\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})} \|f_3\|_{L^{r'}(\mathbb{R}^{n+m})} \\ & \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})} \|f_3\|_{L^{r'}(\mathbb{R}^{n+m})}. \end{aligned}$$

**Cases 3.-7.** We keep operating as above, and so these cases are very similar.  $\square$

#### 8.4. Quasi-Banach estimates for commutators of averages of DMOs.

*Duality lemma.* We present a technical modification of the  $H^1$ -BMO type duality estimate.

For a sequence of scalars  $\{a_R\}_{R \in \mathcal{D}_\omega}$  denote

$$\|\{a_R\}_{R \in \mathcal{D}_\omega}\|_{\text{BMO}_{\text{prod}}(\mathcal{D}_\omega)} := \sup_{\Omega} \left( \frac{1}{|\Omega|} \sum_{\substack{R \in \mathcal{D}_\omega \\ R \subset \Omega}} |a_R|^2 \right)^{1/2},$$

where the supremum is taken over those sets  $\Omega \subset \mathbb{R}^{n+m}$  such that  $|\Omega| < \infty$  and such that for every  $x \in \Omega$  there exist  $K \times V \in \mathcal{D}_\omega$  so that  $x \in K \times V \subset \Omega$ .

**8.12. Lemma.** *Let  $\omega = (\omega_1, \omega_2)$  be a random parameter and  $F \subset \mathbb{R}^{n+m}$ . Suppose  $\mathcal{C} \subset \mathcal{D}_0$  is a collection of rectangles such that  $|R \cap F| \geq \frac{99}{100}|R|$  for all  $R \in \mathcal{C}$ . Let  $\{a_{R+\omega}\}_{R \in \mathcal{C}}$  and  $\{b_R\}_{R \in \mathcal{C}}$  be two collections of scalars. Then*

$$\sum_{R \in \mathcal{C}} |a_{R+\omega} b_R| \lesssim \|\{a_{R+\omega}\}_{R \in \mathcal{C}}\|_{\text{BMO}_{\text{prod}}(\mathcal{D}_\omega)} \iint_F \left( \sum_{R \in \mathcal{C}} |b_R|^2 \frac{1_R}{|R|} \right)^{1/2}.$$

*Proof.* Write  $S = \left( \sum_{R \in \mathcal{C}} |b_R|^2 \frac{1_R}{|R|} \right)^{1/2}$ . We may suppose that  $\|1_F S\|_{L^1} < \infty$ . For  $u \in \mathbb{Z}$  let  $\Omega_u = \{1_F S > 2^{-u}\}$  and  $\tilde{\Omega}_u = \{M1_{\Omega_u} > c\}$ , where  $c = c(n, m) \in (0, 1)$  is small enough. Define the collections  $\hat{\mathcal{C}}_u = \{R \in \mathcal{C} : |R \cap \Omega_u| \geq |R|/100\}$  and write  $\mathcal{C}_u = \hat{\mathcal{C}}_u \setminus \hat{\mathcal{C}}_{u-1}$ .

Let  $R \in \mathcal{C}$  be such that  $b_R \neq 0$ . Then for all  $x = (x_1, x_2) \in R \cap F$  there holds that  $0 < |b_R| |R|^{-1/2} \leq 1_F(x) S(x)$ . Since  $|R \cap F| \geq \frac{99}{100}|R|$ , this implies that  $R \in \hat{\mathcal{C}}_u$  for all large  $u \in \mathbb{Z}$ . On the other hand, since  $|\Omega_u| \rightarrow 0$ , as  $u \rightarrow -\infty$ , we have  $R \notin \hat{\mathcal{C}}_u$  for all small  $u \in \mathbb{Z}$ . Therefore, (because  $\Omega_{u-1} \subset \Omega_u$  for all  $u$ ) it holds that

$$\sum_{R \in \mathcal{C}} |a_{R+\omega} b_R| = \sum_{u \in \mathbb{Z}} \sum_{R \in \mathcal{C}_u} |a_{R+\omega} b_R|.$$

For fixed  $u$  estimate

$$\sum_{R \in \mathcal{C}_u} |a_{R+\omega} b_R| \leq \left( \sum_{R \in \mathcal{C}_u} |a_{R+\omega}|^2 \right)^{1/2} \left( \sum_{R \in \mathcal{C}_u} |b_R|^2 \right)^{1/2}.$$

Suppose  $R \in \mathcal{C}_u$ . Because  $|R \cap \Omega_u| \geq \frac{1}{100}|R|$ , there holds that  $R + \omega \subset 3R \subset \tilde{\Omega}_u$  if  $c = c(n, m)$  is fixed small enough. Therefore, since  $|\tilde{\Omega}_u| \lesssim |\Omega_u|$ , we get that

$$\left( \sum_{R \in \mathcal{C}_u} |a_{R+\omega}|^2 \right)^{1/2} \lesssim \|\{a_{R+\omega}\}_{R \in \mathcal{C}}\|_{\text{BMO}_{\text{prod}}(\mathcal{D}_\omega)} |\Omega_u|^{1/2}.$$

Let again  $R \in \mathcal{C}_u$ . Every  $R \in \mathcal{C}$  satisfies by assumption that  $|R \cap F| \geq \frac{99}{100}|R|$ , and because  $R \notin \hat{\mathcal{C}}_{u-1}$ , there holds that  $|R \cap \Omega_{u-1}^c| \geq \frac{99}{100}|R|$ . Thus, we have  $|R \cap F \cap \Omega_{u-1}^c| \geq \frac{98}{100}|R|$ , and this gives (noting also that  $R \subset \tilde{\Omega}_u$  for every  $R \in \mathcal{C}_u$ ) that

$$\left( \sum_{R \in \mathcal{C}_u} |b_R|^2 \right)^{1/2} \lesssim \left( \iint_{\tilde{\Omega}_u \setminus \Omega_{u-1}} 1_F \sum_{R \in \mathcal{C}_u} |b_R|^2 \frac{1_R}{|R|} \right)^{1/2} \leq \left( \iint_{\tilde{\Omega}_u \setminus \Omega_{u-1}} 1_F S^2 \right)^{1/2} \lesssim 2^{-u} |\Omega_u|^{1/2}.$$

The claim follows by summing over  $u$ .  $\square$

**8.13. Remark.** Set

$$\|\{a_V\}_{V \in \mathcal{D}_{\omega_2}^m}\|_{\text{BMO}(\mathcal{D}_{\omega_2}^m)} := \sup_{V_0 \in \mathcal{D}_{\omega_2}^m} \left( \frac{1}{|V_0|} \sum_{\substack{V \in \mathcal{D}_{\omega_2}^m \\ V \subset V_0}} |a_V|^2 \right)^{1/2}.$$

For weak type estimates of partial paraproducts we use the following special case of Lemma 8.12. Let  $F \subset \mathbb{R}^{n+m}$  and  $K_0 \in \mathcal{D}_0^n$ . Suppose  $\mathcal{C} \subset \mathcal{D}_0^m$  is a collection of cubes such that  $|(K_0 \times V) \cap F| \geq \frac{99}{100}|K_0 \times V|$  for every  $V \in \mathcal{C}$ . Let  $\omega_2$  be a random parameter, and let  $\{a_{V+\omega_2}\}_{V \in \mathcal{C}}$  be a collection of scalars. Then, for all scalars  $\{b_V\}_{V \in \mathcal{C}}$  we have

$$\sum_{V \in \mathcal{C}} |a_{V+\omega_2} b_V| \lesssim \|\{a_{V+\omega_2}\}_{V \in \mathcal{C}}\|_{\text{BMO}(\mathcal{D}_{\omega_2}^m)} \iint_F \frac{1_{K_0}}{|K_0|} \otimes \left( \sum_{V \in \mathcal{C}} |b_V|^2 \frac{1_V}{|V|} \right)^{1/2}.$$

This follows from Lemma 8.12. To see this, let  $\tilde{\mathcal{C}} = \{K \times V \in \mathcal{D}_0: K = K_0, V \in \mathcal{C}\}$ . For  $K \times V \in \mathcal{C}$  define  $\tilde{a}_{K \times (V + \omega_2)} = a_{V + \omega_2}$ . Then we have

$$\|\{\tilde{a}_{K \times (V + \omega_2)}\}_{K \times V \in \tilde{\mathcal{C}}}\|_{\text{BMO}_{\text{prod}}(\mathcal{D}_0^n \times \mathcal{D}_{\omega_2}^m)} = \frac{1}{|K_0|^{1/2}} \|\{a_{V + \omega_2}\}_{V \in \mathcal{C}}\|_{\text{BMO}(\mathcal{D}_{\omega_2}^m)}.$$

We move on to proving the quasi-Banach estimates.

*Shifts.*

**8.14. Proposition.** *Let  $\|b\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$ , and let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ . Suppose that  $(S_{k,\omega}^v)_\omega$  is a collection of bilinear bi-parameter shifts of the same type and of complexity  $(k, v)$ , where  $S_{k,\omega}^v$  is defined in the grid  $\mathcal{D}_\omega$ . Then we have*

$$\|\mathbb{E}_\omega[b, S_{k,\omega}^v]_1(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim (1 + \max(k_i, v_i)) \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}.$$

*Proof.* We omit this shift case, and only show the corresponding proof for partial and full paraproducts below. However, we demonstrate the iterated case for shifts later.  $\square$

*Partial paraproducts.* Below it is convenient to write the partial paraproducts in the following way. Let  $\omega = (\omega_1, \omega_2) \in (\{0, 1\}^n)^\mathbb{Z} \times (\{0, 1\}^m)^\mathbb{Z}$  and  $k = (k_1, k_2, k_3)$ ,  $k_1, k_2, k_3 \geq 0$ . For each  $K, I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n$  with  $I_i^{(k_i)} = K$  and  $V \in \mathcal{D}_{\omega_2}^m$  we are given a constant  $a_{K,V,(I_i)}^\omega$  such that for all  $K, I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n$  with  $I_i^{(k_i)} = K$  we have

$$\|\{a_{K,V,(I_i)}^\omega\}_{V \in \mathcal{D}_{\omega_2}^m}\|_{\text{BMO}(\mathcal{D}_{\omega_2}^m)} \leq \frac{|I_1|^{1/2} |I_2|^{1/2} |I_3|^{1/2}}{|K|^2}.$$

We consider a partial paraproduct  $P_{k,\omega}$  of complexity  $k$  and of a particular form

$$(8.15) \quad \begin{aligned} \langle P_{k,\omega}(f_1, f_2), f_3 \rangle &= \sum_{K \in \mathcal{D}_{\omega_1}^n} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ I_i^{(k_i)} = K}} \sum_{V \in \mathcal{D}_{\omega_2}^m} a_{K,V,(I_i)}^\omega \langle f_1, h_{I_1}^0 \otimes h_V \rangle \\ &\quad \times \left\langle f_2, h_{I_2} \otimes \frac{1_V}{|V|} \right\rangle \left\langle f_3, h_{I_3} \otimes \frac{1_V}{|V|} \right\rangle. \end{aligned}$$

Before we estimate the commutators of partial paraproducts we define two auxiliary operators and prove their boundedness. Given  $\omega_2$  set for  $g: \mathbb{R}^m \rightarrow \mathbb{C}$  and  $f: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  that

$$\tilde{S}_{\omega_2} g = \left( \sum_{V \in \mathcal{D}_0^m} |\langle g, h_{V + \omega_2} \rangle|^2 \frac{1_V}{|V|} \right)^{1/2}, \quad \tilde{S}_{\omega_2}^2 f = \left( \sum_{V \in \mathcal{D}_0^m} |\langle f, h_{V + \omega_2} \rangle|^2 \otimes \frac{1_V}{|V|} \right)^{1/2}.$$

Recall that  $\mathbb{E}_\omega = \mathbb{E}_{\omega_1} \mathbb{E}_{\omega_2}$ . Let  $\Phi_1$  and  $\Phi_2^l$ ,  $0 \leq l \in \mathbb{Z}$ , be the operators

$$\Phi_1(f) = \mathbb{E}_{\omega_2} M^1 \tilde{S}_{\omega_2}^2 (\varphi_{\omega_2, b}^2 f), \quad \Phi_2^l(f) = \left( \sum_{K \in \mathcal{D}_0^n} \mathbb{E}_{\omega_1} (M^1 \Delta_{K + \omega_1, l}^1 \varphi_{\omega_1}^1 f)^2 \right)^{1/2},$$

where  $\varphi_{\omega_2, b}^2 f_1 := \varphi_{\mathcal{D}_{\omega_2}^m, b}^2 f_1$  is the function from Section 8.2 and  $\varphi_{\omega_1}^1 f := \varphi_{\mathcal{D}_{\omega_1}^n}^1 f$  was introduced in Lemma 2.2.

**8.16. Lemma.** *We have for all  $l \in \mathbb{Z}$ ,  $l \geq 0$ , that*

$$\|\Phi_1(f)\|_{L^s(\mathbb{R}^{n+m})} + \|\Phi_2^l(f)\|_{L^s(\mathbb{R}^{n+m})} \lesssim \|f\|_{L^s(\mathbb{R}^{n+m})}, \quad s \in (1, \infty),$$

where the bound is independent of  $l$ .

*Proof.* We use weights and extrapolation. This is useful with  $\Phi_2^l$  in order to reduce to  $L^2$  estimates where we can take the expectation out. With  $\Phi_1$  we could do without weights by estimating directly in  $L^s$ . Take  $w \in A_2(\mathbb{R}^n \times \mathbb{R}^m)$ .

Notice that for all  $V \in \mathcal{D}_0^m$  we have  $|\langle f, h_{V+\omega_2} \rangle_2| \otimes \frac{1_V}{|V|^{1/2}} \lesssim M^2(\langle f, h_{V+\omega_2} \rangle_2 \otimes h_{V+\omega_2})$ . This implies

$$\|\tilde{S}_{\omega_2}^2 f\|_{L^2(w)} \lesssim \left\| \left( \sum_{V \in \mathcal{D}_{\omega_2}^m} [M^2(\langle f, h_V \rangle_2 \otimes h_V)]^2 \right)^{1/2} \right\|_{L^2(w)} \leq C([w]_{A_2}) \|f\|_{L^2(w)},$$

where we used weighted maximal function and weighted square function estimates. Lemma 8.6 says that  $\|\varphi_{\omega_2, b}^2(f)\|_{L^2(w)} \leq C([w]_{A_2}) \|f\|_{L^2(w)}$ , and obviously  $M^1$  satisfies the same bound. The  $L^2(w)$  result for  $\Phi_1$  follows, and we can extrapolate.

Next, we have

$$\|\Phi_2^l(f)\|_{L^2(w)}^2 = \mathbb{E}_{\omega_1} \left\| \left( \sum_{K \in \mathcal{D}_{\omega_1}^n} (M^1 \Delta_{K, l}^1 \varphi_{\omega_1}^1 f)^2 \right)^{1/2} \right\|_{L^2(w)}^2 \leq C([w]_{A_2}) \|f\|_{L^2(w)}^2$$

using weighted maximal function and weighted square function estimates and Lemma 2.2. We can extrapolate to finish.  $\square$

We are ready to treat the first order commutator of partial paraproducts.

**8.17. Proposition.** *Let  $\|b\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$ , and let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ . Suppose that  $(P_{k, \omega})_\omega$  is a collection of partial paraproducts of the same type and of fixed complexity  $k = (k_1, k_2, k_3)$ , where  $P_{k, \omega}$  is defined in the grid  $\mathcal{D}_\omega$ . Then we have*

$$\|\mathbb{E}_\omega [b, P_{k, \omega}]_1(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim (1 + \max k_i) \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}.$$

*Proof.* The case  $r > 1$  is clear by Proposition 8.11. Our remaining task is to prove a weak type estimate when  $r < 1$ , which combined with the Banach range boundedness implies Proposition 8.17 via interpolation. This general scheme was already explained in the proof of Proposition 6.15.

All the different forms of partial paraproducts are treated in a similar way. Here we assume that the operators are of the form (8.15), and therefore the commutator is split using the identity (8.4). This leads to several terms, of which the two coming from the last term in (8.4) need to be combined – this part of the commutator produces the dependence on the complexity via Lemma 8.9. All the other parts of the commutator are estimated separately. Here we focus on the part of the commutator involving  $\langle (\langle b \rangle_{V, 2} - \langle b \rangle_{I_1 \times V}) \langle f_1, h_V \rangle_2, h_{I_1}^0 \rangle$ . All the other parts are estimated relatively similarly – see also the proof of Proposition 8.21 (the full paraproduct case), where a different type of term is handled.

Let  $\Phi_1$  and  $\Phi_2^{k_2}$  be the auxiliary operators from Lemma 8.16, and recall the operator  $\varphi_{\omega_2, b}^2$  involved in the definition of  $\Phi_1$ . By Lemma 8.8 we have that

$$(8.18) \quad \left| \langle (\langle b \rangle_{V, 2} - \langle b \rangle_{I_1 \times V}) \langle f_1, h_V \rangle_2, h_{I_1}^0 \rangle \right| \leq \langle \varphi_{\omega_2, b}^2 f_1, h_{I_1}^0 \otimes h_V \rangle.$$

Since  $\Phi_1$  and  $\Phi_2^{k_2}$  are bounded, it is enough to show that for all  $f_1 \in L^p(\mathbb{R}^{n+m})$ ,  $f_2 \in L^q(\mathbb{R}^{n+m})$  with  $\|\Phi_1(f_1)\Phi_2^{k_2}(f_2)\|_{L^r(\mathbb{R}^{n+m})} = 1$  and  $E \subset \mathbb{R}^{n+m}$  with  $0 < |E| < \infty$  there

exists  $E' \subset E$  with  $|E'| \geq \frac{99}{100}|E|$  so that

$$(8.19) \quad \sum_{\substack{K \in \mathcal{D}_0^n \\ V \in \mathcal{D}_0^m}} \Lambda_{K,V}(f_1, f_2, f_3) \lesssim |E|^{1/r'}$$

holds for all  $f_3$  such that  $|f_3| \leq 1_{E'}$ . Here  $\Lambda_{K,V}$  is defined to act on three functions by

$$\begin{aligned} \Lambda_{K,V}(f_1, f_2, f_3) = \mathbb{E}_\omega \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ I_i^{(k_i)} = K + \omega_1}} & \left| a_{K+\omega_1, V+\omega_2, (I_i)}^\omega \langle \varphi_{\omega_2, b}^2 f_1, h_{I_1}^0 \otimes h_{V+\omega_2} \rangle \right. \\ & \left. \times \left\langle f_2, h_{I_2} \otimes \frac{1_{V+\omega_2}}{|V|} \right\rangle \left\langle f_3, h_{I_3} \otimes \frac{1_{V+\omega_2}}{|V|} \right\rangle \right|. \end{aligned}$$

We turn to prove (8.19). Define the sets  $\Omega_u = \{\Phi_1(f_1)\Phi_2^{k_2}(f_2) > C2^{-u}|E|^{-1/r'}\}$ ,  $u \geq 0$ . For a small enough  $c = c(n, m) \in (0, 1)$  define the enlargement by  $\tilde{\Omega}_u = \{M1_{\Omega_u} > c\}$ . Define  $E' = E \setminus \tilde{\Omega}_0$ . By choosing the constant  $C$  in the definition of the sets  $\Omega_u$  to be large, we have  $|E'| \geq \frac{99}{100}|E|$ . Let  $\hat{\mathcal{R}}_u$  be the collection of rectangles  $\hat{\mathcal{R}}_u = \{R \in \mathcal{D}_0 : |R \cap \Omega_u| \geq |R|/100\}$ , and write  $\mathcal{R}_u = \hat{\mathcal{R}}_u \setminus \hat{\mathcal{R}}_{u-1}$  when  $u \geq 1$ .

Now, we fix an arbitrary function  $f_3$  such that  $|f_3| \leq 1_{E'}$  and consider (8.19) with  $f_3$ . Let  $K \times V \in \mathcal{D}_0$ . First, we show that if  $\Lambda_{K,V}(f_1, f_2, f_3) \neq 0$  then  $K \times V \in \hat{\mathcal{R}}_u$  for some  $u$ . We have for all  $\omega = (\omega_1, \omega_2)$  and every  $(x_1, x_2) \in K \times V$  that

$$\begin{aligned} \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n \\ I_1^{(k_1)} = K + \omega_1}} \frac{|I_1|^{1/2}}{|K||V|^{1/2}} |\langle \varphi_{\omega_2, b}^2 f_1, h_{I_1}^0 \otimes h_{V+\omega_2} \rangle| & \leq \sum_{\substack{I_1 \in \mathcal{D}_{\omega_1}^n \\ I_1^{(k_1)} = K + \omega_1}} \frac{|I_1|^{1/2}}{|K|} \langle \tilde{S}_{\omega_2}^2(\varphi_{\omega_2, b}^2 f_1), h_{I_1}^0 \rangle_1(x_2) \\ & \lesssim M^1 \tilde{S}_{\omega_2}^2(\varphi_{\omega_2, b}^2 f_1)(x_1, x_2) \end{aligned}$$

and

$$\begin{aligned} \sum_{\substack{I_2 \in \mathcal{D}_{\omega_1}^n \\ I_2^{(k_2)} = K + \omega_1}} \frac{|I_2|^{1/2}}{|K|} \left| \left\langle f_2, h_{I_2} \otimes \frac{1_{V+\omega_2}}{|V|} \right\rangle \right| & \lesssim \sum_{\substack{I_2 \in \mathcal{D}_{\omega_1}^n \\ I_2^{(k_2)} = K + \omega_1}} \frac{|I_2|^{1/2}}{|K|} M \langle f_2, h_{I_2} \rangle_1(x_2) \\ & = \sum_{\substack{I_2 \in \mathcal{D}_{\omega_1}^n \\ I_2^{(k_2)} = K + \omega_1}} \frac{|I_2|^{1/2}}{|K|} \langle \varphi_{\omega_1}^1 f_2, h_{I_2} \rangle_1(x_2) \lesssim M^1 (\Delta_{K+\omega_1, k_2}^1 \varphi_{\omega_1}^1 f_2)(x_1, x_2). \end{aligned}$$

Therefore, if  $\Lambda_{K,V}(f_1, f_2, f_3) \neq 0$ , then for every  $x \in K \times V$  there holds that

$$\begin{aligned} 0 < \mathbb{E}_\omega \sum_{\substack{I_1, I_2 \in \mathcal{D}_{\omega_1}^n \\ I_i^{(k_i)} = K + \omega_1}} \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|^2|V|^{1/2}} & \left| \langle \varphi_{\omega_2, b}^2 f_1, h_{I_1}^0 \otimes h_{V+\omega_2} \rangle \left\langle f_2, h_{I_2} \otimes \frac{1_{V+\omega_2}}{|V|} \right\rangle \right| \\ & \lesssim \Phi_1(f_1)(x) \Phi_2^{k_2}(f_2)(x). \end{aligned}$$

The inequality " $<$ " holds since the integrand is positive for  $\omega$  in a set of positive measure. From this it follows that  $K \times V \subset \Omega_u$  if  $u$  is large enough, and so in particular  $K \times V \in \hat{\mathcal{R}}_u$ .

If  $K \times V \in \widehat{\mathcal{R}}_u$ , then for all  $\omega$  there holds that  $(K \times V) + \omega \subset (3K) \times (3V) \subset \widetilde{\Omega}_u$ . The constant  $c = c(n, m)$  in the definition of  $\widetilde{\Omega}_u$  is chosen so that this inclusion holds. If  $K \times V \in \widehat{\mathcal{R}}_0$ , then  $(K \times V) + \omega \subset \widetilde{\Omega}_0 \subset (E')^c$  for all  $\omega$ , which combined with the fact that  $|f_3| \leq 1_{E'}$  implies that  $\Lambda_{K,V}(f_1, f_2, f_3) = 0$ .

With the above observations we have that

$$\sum_{\substack{K \in \mathcal{D}_0^n \\ V \in \mathcal{D}_0^m}} \Lambda_{K,V}(f_1, f_2, f_3) = \sum_{u=1}^{\infty} \sum_{K \times V \in \mathcal{R}_u} \Lambda_{K,V}(f_1, f_2, f_3).$$

We fix  $u$  and estimate the corresponding term.

First, notice that if  $K \times V \in \mathcal{R}_u$ , then  $\Lambda_{K,V}(f_1, f_2, f_3) = \Lambda_{K,V}(f_1, f_2, 1_{\widetilde{\Omega}_u} f_3)$ . For  $K \in \mathcal{D}_0^n$  define  $\mathcal{C}_{K,u} = \{V \in \mathcal{D}_0^m : K \times V \in \mathcal{R}_u\}$ , which allows us to write  $\sum_{K \times V \in \mathcal{R}_u} = \sum_{K \in \mathcal{D}_0^n} \sum_{V \in \mathcal{C}_{K,u}}$ . If  $K \in \mathcal{D}_0^n$ , then every  $V \in \mathcal{C}_{K,u}$  satisfies  $|(K \times V) \cap \Omega_{u-1}^c| \geq \frac{99}{100}|K \times V|$ . Therefore, Remark 8.13 gives that

$$(8.20) \quad \sum_{K \in \mathcal{D}_0^n} \sum_{V \in \mathcal{C}_{K,u}} \Lambda_{K,V}(f_1, f_2, 1_{\widetilde{\Omega}_u} f_3) \lesssim \sum_{K \in \mathcal{D}_0^n} \mathbb{E}_{\omega} \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ I_i^{(k_i)} = K + \omega_1}} \left[ \frac{|I_1|^{1/2} |I_2|^{1/2} |I_3|^{1/2}}{|K|^2} \right. \\ \left. \times \iint_{\widetilde{\Omega}_u \setminus \Omega_{u-1}} \frac{1_K}{|K|} \otimes \widetilde{S}_{\omega_2} \langle \varphi_{\omega_2, b}^2 f_1, h_{I_1}^0 \rangle_1 M \langle f_2, h_{I_2} \rangle_1 M \langle 1_{\widetilde{\Omega}_u} f_3, h_{I_3} \rangle_1 \right],$$

where we used the estimate  $|\langle f_2, h_{I_2} \otimes \frac{1_V + \omega_2}{|V|} \rangle| \lesssim M \langle f_2, h_{I_2} \rangle_1(x_2)$ ,  $x_2 \in V$ , and the same estimate with  $1_{\widetilde{\Omega}_u} f_3$ . We were able to insert the restriction  $\widetilde{\Omega}_u$  to the integration area since  $K \times V \subset \widetilde{\Omega}_u$  for every  $K \times V \in \mathcal{R}_u$ .

Notice that  $\widetilde{S}_{\omega_2} \langle \varphi_{\omega_2, b}^2 f_1, h_{I_1}^0 \rangle_1(x_2) \leq \langle \widetilde{S}_{\omega_2}^2(\varphi_{\omega_2, b}^2 f_1), h_{I_1}^0 \rangle_1(x_2)$  for all  $x_2$ , and recall that  $M \langle f_2, h_{I_2} \rangle_1 = \langle \varphi_{\omega_1}^1 f_2, h_{I_2} \rangle_1$ . Thus, the inner sum over the cubes  $I_i$  in the right hand side of (8.20) is dominated by

$$\iint_{\widetilde{\Omega}_u \setminus \Omega_{u-1}} M^1 \widetilde{S}_{\omega_2}^2(\varphi_{\omega_2, b}^2 f_1) M^1(\Delta_{K+\omega_1, k_2}^1 \varphi_{\omega_1}^1 f_2) M^1(\Delta_{K+\omega_1, k_3}^1 \varphi_{\omega_1}^1(1_{\widetilde{\Omega}_u} f_3)).$$

Taking expectation  $\mathbb{E}_{\omega} = \mathbb{E}_{\omega_1} \mathbb{E}_{\omega_2}$ , using Hölder's inequality with respect to  $\omega_1$  and summing over  $K \in \mathcal{D}_0^n$  shows that

$$\sum_{K \in \mathcal{D}_0^n} \sum_{V \in \mathcal{C}_{K,u}} \Lambda_{K,V}(f_1, f_2, 1_{\widetilde{\Omega}_u} f_3) \lesssim \iint_{\widetilde{\Omega}_u \setminus \Omega_{u-1}} \Phi_1(f_1) \Phi_2^{k_2}(f_2) \Phi_2^{k_3}(1_{\widetilde{\Omega}_u} f_3).$$

By definition we have  $\Phi_1(f_1)(x) \Phi_2^{k_2}(f_2)(x) \lesssim 2^{-u} |E|^{-1/r}$  for all  $x \in \Omega_{u-1}^c$ . Also, just by using the  $L^2$ -boundedness of  $\Phi_2^{k_3}$  and the fact that  $\|f_3\|_{L^\infty} \leq 1$  there holds that

$$\iint_{\widetilde{\Omega}_u \setminus \Omega_{u-1}} \Phi_2^{k_3}(1_{\widetilde{\Omega}_u} f_3) \lesssim |\widetilde{\Omega}_u| \lesssim |\Omega_u| \lesssim 2^{ur} |E|.$$

These combined give that  $\sum_{K \in \mathcal{D}_0^n} \sum_{V \in \mathcal{C}_{K,u}} \Lambda_{K,V}(f_1, f_2, f_3) \lesssim 2^{-u(1-r)} |E|^{1/r'}$ , which can be summed over  $u$  since  $r < 1$ . This finishes the proof of (8.19), and therefore we have

proved the desired weak type estimate for the part of the commutator involving the terms  $\langle (\langle b \rangle_{V,2} - \langle b \rangle_{I_1 \times V}) \langle f_1, h_V \rangle_2, h_{I_1}^0 \rangle$ . The corresponding estimate for the other parts gives the weak type estimate for  $\mathbb{E}_\omega [b, P_{k,\omega}]_1$ , when  $P_{k,\omega}$  is of the form (8.15). This ends our treatment of Proposition 8.17.  $\square$

*Full paraproducts.*

**8.21. Proposition.** *Let  $\|b\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$ , and let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ . Suppose that  $(\Pi_\omega)_\omega$  is a collection of full paraproducts of the same type, where  $\Pi_\omega$  is defined in the grid  $\mathcal{D}_\omega$ . Then we have*

$$\|\mathbb{E}_\omega [b, \Pi_\omega]_1(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})}.$$

*Proof.* As in the partial paraproduct case we are done after showing the weak type estimate for  $r < 1$ . The different forms of full paraproducts are handled separately with analogous arguments. Let us consider here the case where every  $\Pi_\omega$  is of the form

$$(8.22) \quad \langle \Pi_\omega(f_1, f_2), f_3 \rangle = \sum_{K \times V \in \mathcal{D}_\omega} a_{K,V}^\omega \langle f_1 \rangle_{K \times V} \left\langle f_2, \frac{1_K}{|K|} \otimes h_V \right\rangle \left\langle f_3, h_K \otimes \frac{1_V}{|V|} \right\rangle,$$

where  $\|\{a_{K,V}^\omega\}_{K \times V \in \mathcal{D}_\omega}\|_{\text{BMO}_{\text{prod}}(\mathcal{D}_\omega)} \leq 1$ .

The commutators  $[b, \Pi_\omega]$  are again split with the identities from Section 8.1, this time using (8.4) and (8.5). The resulting terms are handled separately with similar arguments. Here we consider the term

$$(8.23) \quad \mathbb{E}_\omega \sum_{K \times V \in \mathcal{D}_\omega} a_{K,V}^\omega \langle f_1 \rangle_{K \times V} \left\langle f_2, \frac{1_K}{|K|} \otimes h_V \right\rangle \left\langle a_{i,\omega_1}^1(b, f_3), h_K \otimes \frac{1_V}{|V|} \right\rangle$$

for some  $i \in \{1, 2\}$ .

This time, let  $\Phi_1$  and  $\Phi_2$  be the auxiliary operators

$$\Phi_1(f) = \mathbb{E}_{\omega_2} \left( \sum_{V \in \mathcal{D}_0^m} (M \langle f, h_{V+\omega_2} \rangle_2)^2 \otimes \frac{1_V}{|V|} \right)^{1/2}$$

and

$$\Phi_2(f) = \mathbb{E}_{\omega_1} \left( \sum_{K \in \mathcal{D}_0^n} \frac{1_K}{|K|} \otimes (M \langle a_{i,\omega_1}^1(b, f), h_{K+\omega_1} \rangle_1)^2 \right)^{1/2}.$$

Similarly as in Lemma 8.16 these are bounded in  $L^s$  for every  $s \in (1, \infty)$ , which uses the fact that  $a_{i,\omega_1}^1(b, \cdot)$  is bounded by Lemma 8.2.

Let now  $f_1 \in L^p(\mathbb{R}^{n+m})$  and  $f_2 \in L^q(\mathbb{R}^{n+m})$  be such that  $\|M f_1 \Phi_1(f_2)\|_{L^r} = 1$  and let  $E \subset \mathbb{R}^{n+m}$  with  $0 < |E| < \infty$ . We show that there exists a set  $E' \subset E$  with  $|E'| \geq \frac{99}{100}|E|$  so that

$$(8.24) \quad \sum_{K \times V \in \mathcal{D}_0} \Lambda_{K,V}(f_1, f_2, f_3) \lesssim |E|^{1/r'}$$

holds for all  $f_3$  such that  $|f_3| \leq 1_{E'}$ , where  $\Lambda_{K,V}$  acts on a triple of functions by

$$\begin{aligned} \Lambda_{K,V}(f_1, f_2, f_3) &= \mathbb{E}_\omega \left| a_{K+\omega_1, V+\omega_2}^\omega \langle f_1 \rangle_{(K \times V) + \omega} \left\langle f_2, \frac{1_{K+\omega_1}}{|K|} \otimes h_{V+\omega_2} \right\rangle \right. \\ &\quad \left. \times \left\langle a_{i,\omega_1}^1(b, f_3), h_{K+\omega_1} \otimes \frac{1_{V+\omega_2}}{|V|} \right\rangle \right|. \end{aligned}$$

Since  $\|Mg_1\Phi_1(g_2)\|_{L^r} \lesssim \|g_1\|_{L^p}\|g_2\|_{L^q}$  for all  $g_1$  and  $g_2$ , this gives the estimate that we want for the term (8.23). Together with the corresponding estimates for all the other parts of  $\mathbb{E}_\omega[b, \Pi_\omega]$ , this proves the weak type estimate we wanted to show.

We turn to prove (8.24). For  $u \geq 0$  let  $\Omega_u = \{Mf_1\Phi_1(f_2) > C2^{-u}|E|^{-1/r}\}$  and  $\tilde{\Omega}_u = \{M1_{\Omega_u} > c\}$ , where  $c = c(n, m) \in (0, 1)$  is a small constant. Set  $E' = E \setminus \tilde{\Omega}_0$ . By choosing the constant  $C$  to be large enough, we have that  $|E'| \geq \frac{99}{100}|E|$ . Then, define  $\hat{\mathcal{R}}_u = \{R \in \mathcal{D}_0: |R \cap \Omega_u| \geq \frac{1}{100}|R|\}$  for  $u \geq 0$  and  $\mathcal{R}_u = \hat{\mathcal{R}}_u \setminus \hat{\mathcal{R}}_{u-1}$  for  $u \geq 1$ .

Suppose  $K \times V \in \mathcal{D}_0$  is such that  $\Lambda_{K,V}(f_1, f_2, f_3) \neq 0$ . Then, we see that

$$0 < \mathbb{E}_\omega \frac{1}{|V|^{1/2}} \left| \langle f_1 \rangle_{(K \times V) + \omega} \left\langle f_2, \frac{1_{K+\omega_1}}{|K|} \otimes h_{V+\omega_2} \right\rangle \right| \lesssim Mf_1(x)\Phi_1(f_2)(x)$$

for all  $x \in K \times V$ . The first “ $<$ ” holds since the integrand is positive for  $\omega$  in a set of positive measure. Thus,  $K \times V \subset \Omega_u$  for large enough  $u$ , so  $K \times V \in \hat{\mathcal{R}}_u$ .

If  $R \in \hat{\mathcal{R}}_u$ , then  $R + \omega \subset 3R \subset \tilde{\Omega}_u$  for all  $\omega$ , which is based the fact that  $c = c(n, m)$  in the definition of  $\tilde{\Omega}_u$  is small enough. Notice that

$$(8.25) \quad \left\langle a_{i,\omega_1}^1(b, f_3), h_{K+\omega_1} \otimes \frac{1_{V+\omega_2}}{|V|} \right\rangle = \left\langle a_{i,\omega_1}^1(b, 1_{(K \times V) + \omega} f_3), h_{K+\omega_1} \otimes \frac{1_{V+\omega_2}}{|V|} \right\rangle.$$

Thus, if  $K \times V \in \hat{\mathcal{R}}_0$ , then  $\Lambda_{K,V}(f_1, f_2, f_3) = 0$  since  $(K \times V) + \omega \subset (E')^c$  for all  $\omega$ , and  $|f_3| \leq 1_{E'}$ .

Now, we have that

$$\sum_{K \times V \in \mathcal{D}_0} \Lambda_{K,V}(f_1, f_2, f_3) = \sum_{u=1}^{\infty} \sum_{K \times V \in \mathcal{R}_u} \Lambda_{K,V}(f_1, f_2, 1_{\tilde{\Omega}_u} f_3),$$

where it was legitimate to replace  $f_3$  with  $1_{\tilde{\Omega}_u} f_3$  because of (8.25). We fix one  $u$  and estimate the related term.

If  $K \times V \in \mathcal{R}_u$ , then by definition  $|(K \times V) \cap \Omega_{u-1}^c| \geq \frac{99}{100}|K \times V|$ . Therefore, using Lemma 8.12 and then the estimate  $\left| \left\langle f_2, \frac{1_{K+\omega_1}}{|K|} \otimes h_{V+\omega_2} \right\rangle \right| \lesssim M \langle f_2, h_{V+\omega_2} \rangle_2(x_1)$ ,  $x_1 \in K$ , and a corresponding estimate related to  $f_3$ , we have that

$$\sum_{K \times V \in \mathcal{R}_u} \Lambda_{K,V}(f_1, f_2, 1_{\tilde{\Omega}_u} f_3) \lesssim \iint_{\tilde{\Omega}_u \setminus \Omega_{u-1}} Mf_1\Phi_1(f_2)\Phi_2(1_{\tilde{\Omega}_u} f_3).$$

The restriction to  $\tilde{\Omega}_u$  in the integration came from the fact that every  $R \in \mathcal{R}_u$  satisfies  $R \subset \tilde{\Omega}_u$ . Since  $Mf_1(x)\Phi_1(f_2)(x) \lesssim 2^{-u}|E|^{-1/r}$  for  $x \in \Omega_{u-1}^c$ , the operator  $\Phi_2$  is  $L^2$  bounded and  $\|f_3\|_{L^\infty} \leq 1$ , there holds that

$$\iint_{\tilde{\Omega}_u \setminus \Omega_{u-1}} Mf_1\Phi_1(f_2)\Phi_2(1_{\tilde{\Omega}_u} f_3) \lesssim 2^{-u}|E|^{-1/r}|\tilde{\Omega}_u| \lesssim 2^{-u(1-r)}|E|^{1/r'}.$$

This can be summed over  $u$  since  $r < 1$ , which finishes the proof of (8.24).  $\square$

## 9. UPPER BOUNDS FOR ITERATED COMMUTATORS

We give the details only for shifts in this section (in the first order case we gave the details for partial and full paraproducts). First, we record the following lemma – the

proof is straightforward with our by now familiar method. We omit the details but see Lemma 5.1 of [48]. Parts of this lemma also appear in [30].

**9.1. Lemma.** *Let  $\|b_1\|_{\text{bmo}(\mathbb{R}^{n+m})} = \|b_2\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$ ,  $1 < p < \infty$  and  $w \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ . Then for  $i = 1, \dots, 8$  and  $j = 1, 2$  we have*

$$\|[b_2, A_i(b_1, f)]\|_{L^p(w)} + \|[b_2, a_j^1(b_1, f)]\|_{L^p(w)} \leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|f\|_{L^p(w)}.$$

Next, we set up some additional notation and record the boundedness of certain deterministic square functions. Let  $i, j \in \mathbb{Z}$ ,  $i, j \geq 0$ . Suppose that we have a family of operators  $U = \{U_\omega\}_\omega$  such that for all  $\omega$  there holds

$$\|U_\omega f\|_{L^2(w)} \leq C([w]_{A_2(\mathbb{R}^n \times \mathbb{R}^m)}) \|f\|_{L^2(w)}, \quad f \in L^2(w),$$

for all  $w \in A_2(\mathbb{R}^n \times \mathbb{R}^m)$ . Define

$$S_U^{i,j} f = \left( \sum_{K \times V \in \mathcal{D}_0} \mathbb{E}_\omega (M \Delta_{(K+\omega_1) \times (V+\omega_2)}^{i,j} U_\omega f)^2 \right)^{1/2}.$$

Similarly, given  $U = \{U_{\omega_1}\}_{\omega_1}$  or  $U = \{U_{\omega_2}\}_{\omega_2}$  we set

$$S_{i,U}^1 = \left( \sum_{K \in \mathcal{D}_0^n} \mathbb{E}_{\omega_1} (M \Delta_{K+\omega_1, i}^1 U_{\omega_1} f)^2 \right)^{1/2}, \quad S_{j,U}^2 = \left( \sum_{V \in \mathcal{D}_0^m} \mathbb{E}_{\omega_2} (M \Delta_{V+\omega_2, j}^2 U_{\omega_2} f)^2 \right)^{1/2}.$$

We write  $S^{i,j}$ ,  $S_i^1$  and  $S_j^2$  if there is no  $U$  present. The proof of the next lemma follows from extrapolation similarly as in Lemma 8.16.

**9.2. Lemma.** *For all  $p \in (1, \infty)$  and  $w \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$  there holds*

$$\|S_U^{i,j} f\|_{L^p(w)} + \|S_{i,U}^1 f\|_{L^p(w)} + \|S_{j,U}^2 f\|_{L^p(w)} \leq C([w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}) \|f\|_{L^p(w)}.$$

We are ready to handle the iterated commutators of shifts.

**9.3. Proposition.** *Let  $\|b_1\|_{\text{bmo}(\mathbb{R}^{n+m})} = \|b_2\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$ , and let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ . Suppose that  $(S_\omega)_\omega = (S_{k,\omega}^v)_\omega$  is a collection bilinear bi-parameter shifts of the same type and of complexity  $(k, v)$ , each defined in the grid  $\mathcal{D}_\omega$ . Then we have*

$$\|\mathbb{E}_\omega [b_2, [b_1, S_\omega]_1]_2(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim (1 + \max(k_i, v_i))^2 \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})},$$

and similarly for  $\mathbb{E}_\omega [b_2, [b_1, S_\omega]_1]_1$ .

*Proof.* We focus on the harder quasi-Banach range proof, which contains the decomposition of the iterated commutator that is used also in the Banach range case. Some comments about the Banach range case are made along the way. We only consider the iterated commutator of the form  $[b_2, [b_1, S_\omega]_1]_2$ , the other being analogous.

The shifts we consider here are of the form

$$\langle S_\omega(f_1, f_2), f_3 \rangle = \sum_{\substack{K \in \mathcal{D}_0^n \\ V \in \mathcal{D}_0^m}} A_{K \times V}^\omega(f_1, f_2, f_3),$$

where

$$\begin{aligned} A_{K \times V}^\omega(f_1, f_2, f_3) &= \sum_{\substack{I_1, I_2, I_3 \in \mathcal{D}_{\omega_1}^n \\ I_i^{(k_i)} = K + \omega_1}} \sum_{\substack{J_1, J_2, J_3 \in \mathcal{D}_{\omega_2}^m \\ J_i^{(v_i)} = V + \omega_2}} a_{K + \omega_1, V + \omega_2, (I_i), (J_j)}^\omega \\ &\quad \times \langle f_1, h_{I_1}^0 \otimes h_{J_1} \rangle \langle f_2, h_{I_2} \otimes h_{J_2} \rangle \langle f_3, h_{I_3} \otimes h_{J_3}^0 \rangle. \end{aligned}$$

Shifts of other form are handled analogously.

The first step of the decomposition is to expand  $\langle [b_1, S_\omega]_1(f_1, f_2), f_3 \rangle$  using our usual rules. This leads to multiple terms, and their contributions to the second order commutator  $[b_2, [b_1, S_\omega]_1]_2$  are considered separately. One of the terms is

$$(9.4) \quad \sum_{\substack{K \in \mathcal{D}_0^n \\ V \in \mathcal{D}_0^m}} A_{K \times V}^\omega(f_1, f_2, a_{i, \omega_1}^1(b_1, f_3)).$$

We consider its contribution first, and briefly comment the others at the end.

We need to prove the following key statement. Let  $r < 1$ , and suppose  $f_1 \in L^p(\mathbb{R}^{n+m})$  and  $f_2 \in L^q(\mathbb{R}^{n+m})$  satisfy

$$\left\| S_{v_1}^2 f_1 S^{k_2, v_2} f_2 + \sum_{j=1}^8 S_{v_1}^2 f_1 S_{A_j, b_2}^{k_2, v_2} f_2 \right\|_{L^r} = 1,$$

where  $A_{j, b_2}$  denotes the family  $\{A_{j, \omega}(b_2, \cdot)\}_\omega$ , and the square functions are defined as above. Given  $E \subset \mathbb{R}^{n+m}$  with  $0 < |E| < \infty$  there exists a subset  $E' \subset E$  with  $|E'| \geq \frac{99}{100}|E|$  so that for all functions  $f_3$  satisfying  $|f_3| \leq 1_{E'}$  there holds

$$(9.5) \quad \left| \mathbb{E}_\omega \sum_{\substack{K \in \mathcal{D}_0^n \\ V \in \mathcal{D}_0^m}} [A_{K \times V}^\omega(f_1, f_2, a_{i, \omega_1}^1(b_1, b_2 f_3)) - A_{K \times V}^\omega(f_1, b_2 f_2, a_{i, \omega_1}^1(b_1, f_3))] \right| \\ \lesssim (1 + \max(k_i, v_i)) |E|^{1/r'}.$$

In view of Lemma 9.2 this is enough.

We define

$$\Omega_u = \left\{ S_{v_1}^2 f_1 S^{k_2, v_2} f_2 + \sum_{j=1}^8 S_{v_1}^2 f_1 S_{A_j, b_2}^{k_2, v_2} f_2 > C_0 2^{-u} |E|^{-1/r} \right\}, \quad u \geq 0,$$

and  $\tilde{\Omega}_u = \{M 1_{\Omega_u} > c_1\}$ , where  $c_1 > 0$  is a small enough dimensional constant. Then we choose  $C_0 = C_0(c_1)$  so large that the set  $E' := E \setminus \tilde{\Omega}_0$  satisfies  $|E'| \geq \frac{99}{100}|E|$ . Define the collections  $\hat{\mathcal{R}}_u = \{R \in \mathcal{D}_0 : |R \cap \Omega_u| \geq |R|/2\}$ , and set  $\mathcal{R}_u = \hat{\mathcal{R}}_u \setminus \hat{\mathcal{R}}_{u-1}$  for  $u \geq 1$ . Fix now some function  $f_3$  such that  $|f_3| \leq 1_{E'}$ . We abbreviate

$$\Lambda_{K \times V}^\omega(f_1, f_2, f_3) = A_{K \times V}^\omega(f_1, f_2, a_{i, \omega_1}^1(b_1, b_2 f_3)) - A_{K \times V}^\omega(f_1, b_2 f_2, a_{i, \omega_1}^1(b_1, f_3)).$$

Notice the localisation property

$$\Lambda_{K \times V}^\omega(f_1, f_2, f_3) = \Lambda_{K \times V}^\omega(f_1, f_2, 1_{(K + \omega_1) \times (V + \omega_2)} f_3).$$

With some thought, similarly as in the proofs of Proposition 8.17 and Proposition 8.21, we see that we may write

$$\mathbb{E}_\omega \sum_{\substack{K \in \mathcal{D}_0^n \\ V \in \mathcal{D}_0^m}} \Lambda_{K \times V}^\omega(f_1, f_2, f_3) = \sum_{u=1}^{\infty} \sum_{K \times V \in \mathcal{R}_u} \mathbb{E}_\omega \Lambda_{K \times V}^\omega(f_1, f_2, 1_{\tilde{\Omega}_u} f_3).$$

We now fix  $u$ , and our goal is to prove

$$\sum_{K \times V \in \mathcal{R}_u} \mathbb{E}_\omega |\Lambda_{K \times V}^\omega(f_1, f_2, 1_{\tilde{\Omega}_u} f_3)| \lesssim (1 + \max(k_i, v_i)) 2^{-u(1-r)} |E|^{1/r'}.$$

By adding and subtracting  $A_{K \times V}^\omega(f_1, f_2, b_2 a_{i, \omega_1}^1(b_1, 1_{\tilde{\Omega}_u} f_3))$ , we see that

$$(9.6) \quad \begin{aligned} \Lambda_{K \times V}^\omega(f_1, f_2, 1_{\tilde{\Omega}_u} f_3) &= -A_{K \times V}^\omega(f_1, f_2, [b_2, a_{i, \omega_1}^1(b_1, \cdot)](1_{\tilde{\Omega}_u} f_3)) \\ &\quad + A_{K \times V}^\omega(f_1, f_2, b_2 a_{i, \omega_1}^1(b_1, 1_{\tilde{\Omega}_u} f_3)) \\ &\quad - A_{K \times V}^\omega(f_1, b_2 f_2, a_{i, \omega_1}^1(b_1, 1_{\tilde{\Omega}_u} f_3)). \end{aligned}$$

We now consider the first term of (9.6). The corresponding term in the Banach range proof can be handled immediately by just using Lemma 9.1. Here we need to work a bit harder. Denote the family  $\{[b_2, a_{i, \omega_1}^1(b_1, \cdot)]\}_{\omega_1}$  by  $[b_2, a_{i, b_1}^1]$ . We see that

$$\begin{aligned} &\sum_{K \times V \in \mathcal{R}_u} \mathbb{E}_\omega |A_{K \times V}^\omega(f_1, f_2, [b_2, a_{i, \omega_1}^1(b_1, \cdot)](1_{\tilde{\Omega}_u} f_3))| \\ &\lesssim \int_{\tilde{\Omega}_u \setminus \Omega_{u-1}} S_{v_1}^2 f_1 S^{k_2, v_2} f_2 S_{k_1, [b_2, a_{i, b_1}^1]}^1(1_{\tilde{\Omega}_u} f_3). \end{aligned}$$

From here the estimate can be concluded in the familiar way, using that  $S_{v_1}^2 f_1 S^{k_2, v_2} f_2 \lesssim 2^{-u} |E|^{-1/r}$  in the complement of  $\Omega_{u-1}$  and that the square function  $S_{k_1, [b_2, a_{i, b_1}^1]}^1$  is bounded.

The boundedness of this square function follows from Lemmas 9.1 and 9.2.

We now consider the last difference in (9.6). In the Banach range this can simply be handled by using the boundedness of  $[b_2, S_\omega]_2$  and  $a_{i, \omega_1}^1(b_1, \cdot)$ . Here we need to split this further by expanding the products  $b_2 a_{i, \omega_1}^1(b_1, 1_{\tilde{\Omega}_u} f_3)$  and  $b_2 f_2$  using our usual expansions rules determined by the form  $A_{K \times V}^\omega$ .

This produces several terms, one of them being  $A_{K \times V}^\omega(f_1, A_{j, \omega}(b_2, f_2), a_{i, \omega_1}^1(b_1, 1_{\tilde{\Omega}_u} f_3))$ ,  $j = 1, \dots, 8$ . Its contribution can be handled by estimating as

$$\begin{aligned} &\sum_{K \times V \in \mathcal{R}_u} \mathbb{E}_\omega |A_{K \times V}^\omega(f_1, A_{j, \omega}(b_2, f_2), a_{i, \omega_1}^1(b_1, 1_{\tilde{\Omega}_u} f_3))| \\ &\lesssim \int_{\tilde{\Omega}_u \setminus \Omega_{u-1}} S_{v_1}^2 f_1 S_{A_j, b_2}^{k_2, v_2} f_2 S_{k_3, a_{i, b_1}^1}^1(1_{\tilde{\Omega}_u} f_3), \end{aligned}$$

and then concluding as usual. The other produced terms from the expansion are handled similarly. We have proved (9.5), and this finishes the treatment of the contribution of (9.4) to  $[b_2, [b_1, S_\omega]_1]_2$ .

We explain what happens if instead of (9.4) we consider another term coming from the decomposition of  $\langle [b_1, S_\omega]_1(f_1, f_2), f_3 \rangle$ . The term that we handled is the only one which uses the add and subtract trick from above, which lead to (9.6). The other ones are handled by directly expanding the products  $b_2 f_2$  and  $b_2 f_3$  in the usual way. Some of

the terms are particularly easy in the Banach range case, since they do not even require expansions and essentially reduce to the boundedness of  $[b_2, S_\omega]_2$ . We omit these details.  $\square$

We end this section by stating the corresponding result for CZOs.

**9.7. Theorem.** *Let  $\|b_1\|_{\text{bmo}(\mathbb{R}^{n+m})} = \|b_2\|_{\text{bmo}(\mathbb{R}^{n+m})} = 1$ , and let  $1 < p, q \leq \infty$  and  $1/2 < r < \infty$  satisfy  $1/p + 1/q = 1/r$ . Let  $T$  be a bilinear bi-parameter CZO. Then we have*

$$\|[b_2, [b_1, T]_1]_2(f_1, f_2)\|_{L^r(\mathbb{R}^{n+m})} \lesssim \|f_1\|_{L^p(\mathbb{R}^{n+m})} \|f_2\|_{L^q(\mathbb{R}^{n+m})},$$

and similarly for  $[b_2, [b_1, T]_1]_1$ .

## 10. LOWER BOUNDS FOR COMMUTATORS

We quickly prove lower bounds by developing bilinear bi-parameter analogs of some methods present in the very recent paper [34] by Hytönen. Let  $K$  be a kernel satisfying the standard estimates of a full kernel (as in Section 3.1.1). We also assume that  $K$  is uniformly non-degenerate. In our setup this means that for all  $y \in \mathbb{R}^{n+m}$  and  $r_1, r_2 > 0$  there exists  $x \in \mathbb{R}^{n+m}$  such that  $|x_1 - y_1| > r_1, |x_2 - y_2| > r_2$  and

$$(10.1) \quad |K(x, y, y)| \gtrsim \frac{1}{r_1^{2n} r_2^{2m}}.$$

**10.2. Remark.** Regarding the assumed Hölder conditions of the kernel  $K$ , similarly as in [34], a weaker modulus of continuity could be enough. Also, this bilinear non-degeneracy condition could perhaps be relaxed somewhat. We did not pursue these avenues.

**10.3. Example.** Define the kernels of the bilinear one-parameter Riesz transforms  $R_i^n$ ,  $i = 1, \dots, 2n$ , by setting for  $(x, y, z) \in \mathbb{R}^{3n}$  with  $x \neq y$  or  $x \neq z$  that

$$K_{R_i^n}(x, y, z) = \frac{x_i - y_i}{(|x - y|^2 + |x - z|^2)^{(2n+1)/2}} \quad \text{or} \quad K_{R_i^n}(x, y, z) = \frac{x_i - z_i}{(|x - y|^2 + |x - z|^2)^{(2n+1)/2}}$$

depending whether  $i = 1, \dots, n$  or  $i = n + 1, \dots, 2n$ , respectively. Let  $R_{i,j} = R_{i,j}^{n,m} := R_i^n \otimes R_j^m$  denote the corresponding bilinear bi-parameter Riesz transform on  $\mathbb{R}^{n+m}$ , and let  $K_{i,j}$  denote its full kernel. Obviously  $K_{i,j}$  satisfies (10.1).

Notice that (10.1) implies the following: given a rectangle  $R = I \times J$  and  $C_0 > 0$ , there exists a rectangle  $\tilde{R} = \tilde{I} \times \tilde{J}$  such that  $\ell(I) = \ell(\tilde{I}), \ell(J) = \ell(\tilde{J}), d(I, \tilde{I}) \geq C_0 \ell(I), d(J, \tilde{J}) \geq C_0 \ell(J)$  and such that for some  $\sigma \in \mathbb{C}$  with  $|\sigma| = 1$  we have for all  $x \in \tilde{R}$  and  $y, z \in R$  that

$$\text{Re } \sigma K(x, y, z) \gtrsim_{C_0} \frac{1}{|R|^2}.$$

To see this, let, for a big enough constant  $A = A(C_0)$ , the centre  $c_{\tilde{R}}$  of  $\tilde{R}$  be the point  $x$  given by (10.1) applied to  $y = c_R, r_1 = A\ell(I)$  and  $r_2 = A\ell(J)$ . Choose  $\sigma$  so that  $\sigma K(c_{\tilde{R}}, c_R, c_R) = |K(c_{\tilde{R}}, c_R, c_R)|$ , and use mixed Hölder and size estimates repeatedly.

Let  $k \geq 1$  and  $b \in L_{\text{loc}}^k(\mathbb{R}^{n+m})$ . Let  $\gamma_1, \gamma_2 \in \{0, 1, \dots, k\}$  be such that  $\gamma_1 + \gamma_2 = k, r > 0$  and  $C_0 > 0$ . We define  $\Gamma = \Gamma(K, b, k, r, \gamma_1, \gamma_2, C_0)$  by setting

$$\Gamma = \sup \frac{1}{|R|^{1/r}} \left\| x \mapsto 1_{\tilde{R}}(x) \iint_{A \times A} (b(x) - b(y))^{\gamma_1} (b(x) - b(z))^{\gamma_2} K(x, y, z) \, dy \, dz \right\|_{L^{r, \infty}(\mathbb{R}^{n+m})},$$

where the supremum is taken over all rectangles  $R, \tilde{R}$  with  $\ell(I) = \ell(\tilde{I})$ ,  $\ell(J) = \ell(\tilde{J})$ ,  $d(I, \tilde{I}) \geq C_0 \ell(I)$  and  $d(J, \tilde{J}) \geq C_0 \ell(J)$ , and all subsets  $A \subset R$ .

10.4. *Remark.* It is clear that if  $T$  is a bilinear bi-parameter singular integral with a full kernel  $K$ , then the function inside the  $L^{r,\infty}$  norm in  $\Gamma$  is equal to

$$1_{\tilde{R}}(x)[b, \dots [b, [b, T]_{i_1}]_{i_2} \dots]_{i_k}(1_A, 1_A)(x),$$

where  $i_j \in \{1, 2\}$  with exactly  $\gamma_1$  of them being 1 and  $\gamma_2$  of them being 2. Therefore, if  $1/p + 1/q = 1/r$  for some  $p, q \in (1, \infty]$  and this iterated commutator maps  $L^p \times L^q \rightarrow L^{r,\infty}$ , then  $\Gamma$  is dominated by this norm. However, the constant  $\Gamma$  is significantly weaker and depends only on the kernel  $K$  and some off-diagonal assumptions.

10.5. **Proposition.** *Suppose  $K$  is a uniformly non-degenerate bilinear bi-parameter full kernel,  $k \geq 1$  and  $b \in L^k_{\text{loc}}(\mathbb{R}^{n+m})$  is real-valued. Let  $\gamma_1, \gamma_2 \in \{0, 1, \dots, k\}$  be such that  $\gamma_1 + \gamma_2 = k$ ,  $r > 0$  and  $C_0 > 0$ . Then for  $\Gamma = \Gamma(K, b, k, r, \gamma_1, \gamma_2, C_0)$  we have  $\|b\|_{\text{bmo}(\mathbb{R}^{n+m})} \lesssim \Gamma^{1/k}$ .*

*Proof.* Let  $R = I \times J$  be a fixed rectangle. As we saw above we can find  $\tilde{R} = \tilde{I} \times \tilde{J}$  such that  $\ell(I) = \ell(\tilde{I})$ ,  $\ell(J) = \ell(\tilde{J})$ ,  $d(I, \tilde{I}) \geq C_0 \ell(I)$ ,  $d(J, \tilde{J}) \geq C_0 \ell(J)$  and such that for some  $\sigma \in \mathbb{C}$  with  $|\sigma| = 1$  we have for all  $x \in \tilde{R}$  and  $y, z \in R$  that  $\text{Re } \sigma K(x, y, z) \gtrsim_{C_0} |R|^{-2}$ . For  $t \in \mathbb{R}$  let  $t_+ = \max(t, 0)$ . For an arbitrary  $\alpha \in \mathbb{R}$  and  $x \in \tilde{R} \cap \{b \geq \alpha\}$  we have

$$\begin{aligned} \left( \frac{1}{|R|} \int_R (\alpha - b)_+ \right)^k &\leq \frac{1}{|\tilde{R}|^2} \iint_{[R \cap \{b \leq \alpha\}]^2} (b(x) - b(y))^{\gamma_1} (b(x) - b(z))^{\gamma_2} dy dz \\ &\lesssim \text{Re } \sigma \iint_{[R \cap \{b \leq \alpha\}]^2} (b(x) - b(y))^{\gamma_1} (b(x) - b(z))^{\gamma_2} K(x, y, z) dy dz. \end{aligned}$$

Letting  $\alpha$  be a median of  $b$  in  $\tilde{R}$  we have that

$$\begin{aligned} |R|^{1/r} \left( \frac{1}{|R|} \int_R (\alpha - b)_+ \right)^k &\lesssim |\tilde{R} \cap \{b \geq \alpha\}|^{1/r} \left( \frac{1}{|R|} \int_R (\alpha - b)_+ \right)^k \\ &\lesssim \left\| x \mapsto 1_{\tilde{R}}(x) \iint_{[R \cap \{b \leq \alpha\}]^2} (b(x) - b(y))^{\gamma_1} (b(x) - b(z))^{\gamma_2} K(x, y, z) dy dz \right\|_{L^{r,\infty}(\mathbb{R}^{n+m})}. \end{aligned}$$

Using the definition of  $\Gamma$  we get  $\int_R (\alpha - b)_+ \lesssim \Gamma^{1/k} |R|$ . Combining with the symmetric estimate  $\int_R (b - \alpha)_+ \lesssim \Gamma^{1/k} |R|$  we are done.  $\square$

## APPENDIX A. BILINEAR BI-PARAMETER MULTIPLIERS

We show that the multipliers  $T_m$  studied e.g. in [51] are paraproduct free (even the partial paraproducts vanish) bilinear bi-parameter CZOs. Recall that in Grafakos–Torres [24] it was shown that the bilinear one-parameter multipliers of Coifman–Meyer [10] are bilinear one-parameter singular integrals satisfying  $T1$  type assumptions. So what we do is in this spirit but in the bi-parameter setting.

Let  $m \in L^\infty(\mathbb{R}^{n+m} \times \mathbb{R}^{n+m})$  be smooth outside the set  $\Delta := \{(\xi, \eta) \in \mathbb{R}^{n+m} \times \mathbb{R}^{n+m} : |\xi_1| + |\eta_1| = 0 \text{ or } |\xi_2| + |\eta_2| = 0\}$ , and assume that it satisfies the estimate

$$|\partial_{\xi_1}^{\alpha_1} \partial_{\xi_2}^{\alpha_2} \partial_{\eta_1}^{\beta_1} \partial_{\eta_2}^{\beta_2} m(\xi, \eta)| \lesssim (|\xi_1| + |\eta_1|)^{-|\alpha_1| - |\beta_1|} (|\xi_2| + |\eta_2|)^{-|\alpha_2| - |\beta_2|}.$$

For Schwartz functions  $f_1, f_2: \mathbb{R}^{n+m} \rightarrow \mathbb{C}$  and  $x \in \mathbb{R}^{n+m}$  define

$$T_m(f_1, f_2)(x) = \iint_{\mathbb{R}^{n+m}} \iint_{\mathbb{R}^{n+m}} m(\xi, \eta) \hat{f}_1(\xi) \hat{f}_2(\eta) e^{2\pi i x \cdot (\xi + \eta)} d\xi d\eta,$$

where  $\widehat{f}_1(\xi) = \iint_{\mathbb{R}^{n+m}} f_1(x) e^{-2\pi i x \cdot \xi} dx$ .

Fix a smooth function  $p_1: \mathbb{R}^{2n} \rightarrow [0, 1]$  so that  $p_1(\xi_1, \eta_1) = 1$  when  $|\xi_1| + |\eta_1| \leq 1$  and  $p_1(\xi_1, \eta_1) = 0$  when  $|\xi_1| + |\eta_1| \geq 2$ . Set  $\delta_1(\xi_1, \eta_1) = p_1(\xi_1, \eta_1) - p_1(2\xi_1, 2\eta_1)$ . Define  $\delta_2$  similarly in  $\mathbb{R}^{2m}$ , and set

$$m_{j_1, j_2}(\xi, \eta) = m(\xi, \eta) \delta_1(2^{-j_1} \xi_1, 2^{-j_1} \eta_1) \delta_2(2^{-j_2} \xi_2, 2^{-j_2} \eta_2), \quad j_1, j_2 \in \mathbb{Z}.$$

For  $J \in \mathbb{N}$  define

$$m_J(\xi, \eta) = \sum_{|j_1|, |j_2| \leq J} m_{j_1, j_2}(\xi, \eta).$$

We show that  $T_{m_J}$  is a paraproduct free (in the strong sense that also the partial paraproducts vanish) bilinear bi-parameter singular integral in our class uniformly on  $J$ .

**Full kernel representation.** Set  $K_J(x, y) = \sum_{|j_1|, |j_2| \leq J} K_{j_1, j_2}(x, y)$ , where

$$K_{j_1, j_2}(x, y) = \iint_{\mathbb{R}^{n+m}} \iint_{\mathbb{R}^{n+m}} m_{j_1, j_2}(\xi, \eta) e^{2\pi i(x \cdot \xi + y \cdot \eta)} d\xi d\eta.$$

Notice that  $K_J$  is a Schwartz function as  $m_J$  is, and that for all  $x \in \mathbb{R}^{n+m}$  we have

$$T_{m_J}(f_1, f_2)(x) = K_J * (f_1 \otimes f_2)(x, x) = \iint_{\mathbb{R}^{n+m}} \iint_{\mathbb{R}^{n+m}} K_J(x - y, x - z) f_1(y) f_2(z) dz dy.$$

We claim that  $(x, y, z) \mapsto K_J(x - y, x - z)$  satisfies the estimates required from a full kernel uniformly on  $J$ . Using standard arguments involving integration by parts as in Stein's book [59] pp. 245–246 (which is the linear one-parameter case), we see that

$$\sum_{j_1, j_2 \in \mathbb{Z}} |\partial_{x_1}^{\alpha_1} \partial_{x_2}^{\alpha_2} \partial_{y_1}^{\beta_1} \partial_{y_2}^{\beta_2} K_{j_1, j_2}(x, y)| \lesssim (|x_1| + |y_1|)^{-2n - |\alpha_1| - |\beta_1|} (|x_2| + |y_2|)^{-2m - |\alpha_2| - |\beta_2|}.$$

These derivative bounds imply the desired full kernel estimates for  $K_J$ , uniformly on  $J$ .

**Partial kernel representation.** Let  $f_i = f_i^1 \otimes f_i^2$ , where  $f_i^1: \mathbb{R}^n \rightarrow \mathbb{C}$  and  $f_i^2: \mathbb{R}^m \rightarrow \mathbb{C}$ ,  $i = 1, 2, 3$ . We write

$$\langle T_{m_J}(f_1, f_2), f_3 \rangle = \langle T_{m_{J, f_1^2, f_2^2, f_3^2}}(f_1^1, f_2^1), f_3^1 \rangle,$$

where

$$m_{J, f_1^2, f_2^2, f_3^2}(\xi_1, \eta_1) = \iiint_{\mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m} m_J(\xi, \eta) \widehat{f_1^2}(\xi_2) \widehat{f_2^2}(\eta_2) \widehat{f_3^2}(x_2) e^{2\pi i x_2 \cdot (\xi_2 + \eta_2)} d\xi_2 d\eta_2 dx_2$$

and

$$T_{m_{J, f_1^2, f_2^2, f_3^2}}(f_1^1, f_2^1)(x_1) = \iint_{\mathbb{R}^n \times \mathbb{R}^n} m_{J, f_1^2, f_2^2, f_3^2}(\xi_1, \eta_1) \widehat{f_1^1}(\xi_1) \widehat{f_2^1}(\eta_1) e^{2\pi i x_1 \cdot (\xi_1 + \eta_1)} d\xi_1 d\eta_1.$$

Notice that

$$\partial_{\xi_1}^{\alpha_1} \partial_{\eta_1}^{\beta_1} m_{J, f_1^2, f_2^2, f_3^2}(\xi_1, \eta_1) = \langle T_{\partial_{\xi_1}^{\alpha_1} \partial_{\eta_1}^{\beta_1} m_{J, (\cdot, \cdot, \cdot)}}(f_1^2, f_2^2), f_3^2 \rangle,$$

where

$$T_{\partial_{\xi_1}^{\alpha_1} \partial_{\eta_1}^{\beta_1} m_{J, (\cdot, \cdot, \cdot)}}(f_1^2, f_2^2)(x_2) = \iint_{\mathbb{R}^m \times \mathbb{R}^m} \partial_{\xi_1}^{\alpha_1} \partial_{\eta_1}^{\beta_1} m_J(\xi, \eta) \widehat{f_1^2}(\xi_2) \widehat{f_2^2}(\eta_2) e^{2\pi i x_2 \cdot (\xi_2 + \eta_2)} d\xi_2 d\eta_2.$$

Since we have

$$|\partial_{\xi_2}^{\alpha_2} \partial_{\eta_2}^{\beta_2} \partial_{\xi_1}^{\alpha_1} \partial_{\eta_1}^{\beta_1} m_J(\xi_1, \xi_2, \eta_1, \eta_2)| \lesssim (|\xi_1| + |\eta_1|)^{-|\alpha_1| - |\beta_1|} (|\xi_2| + |\eta_2|)^{-|\alpha_2| - |\beta_2|},$$

Coifman–Meyer [10] and Grafakos–Torres [24] tell us that

$$|\langle T_{\partial_{\xi_1}^{\alpha_1} \partial_{\eta_1}^{\beta_1} m_{J, (\xi_1, \cdot, \eta_1, \cdot)}}(f_1^2, f_2^2), f_3^2 \rangle| \lesssim (|\xi_1| + |\eta_1|)^{-|\alpha_1| - |\beta_1|} \|f_1^2\|_{L^4(\mathbb{R}^m)} \|f_2^2\|_{L^4(\mathbb{R}^m)} \|f_3^2\|_{L^2(\mathbb{R}^m)}.$$

But this means that the same upper bound holds for  $|\partial_{\xi_1}^{\alpha_1} \partial_{\eta_1}^{\beta_1} m_{J, f_1^2, f_2^2, f_3^2}(\xi_1, \eta_1)|$ . Another application of [10], [24] tells us that for  $x_1 \in \mathbb{R}^n$  we have

$$T_{m_{J, f_1^2, f_2^2, f_3^2}}(f_1^1, f_2^1)(x_1) = \iint_{\mathbb{R}^n \times \mathbb{R}^n} K_{J, f_1^2, f_2^2, f_3^2}(x_1 - y_1, x_1 - z_1) f_1^1(y_1) f_2^1(z_1) dy_1 dz_1,$$

where  $(x_1, y_1, z_1) \mapsto K_{J, f_1^2, f_2^2, f_3^2}(x_1 - y_1, x_1 - z_1)$  is a bilinear Calderón–Zygmund kernel with a constant dominated by  $\|f_1^2\|_{L^4(\mathbb{R}^m)} \|f_2^2\|_{L^4(\mathbb{R}^m)} \|f_3^2\|_{L^2(\mathbb{R}^m)}$ . In particular, we have the desired partial kernel representation in  $\mathbb{R}^n$  with the correct bounds. Of course, we can similarly also obtain the partial kernel representation in  $\mathbb{R}^m$ .

We record here that the above application of [10], [24] actually gives the bound

$$(A.1) \quad \begin{aligned} |\langle T_{m_J}(f_1, f_2), f_3 \rangle| &= |\langle T_{m_{J, f_1^2, f_2^2, f_3^2}}(f_1^1, f_2^1), f_3^1 \rangle| \\ &\lesssim \|f_1^1\|_{L^4(\mathbb{R}^n)} \|f_1^2\|_{L^4(\mathbb{R}^m)} \|f_2^1\|_{L^4(\mathbb{R}^n)} \|f_2^2\|_{L^4(\mathbb{R}^m)} \|f_3^1\|_{L^2(\mathbb{R}^n)} \|f_3^2\|_{L^2(\mathbb{R}^m)}, \end{aligned}$$

i.e. boundedness for functions of the tensor product form.

**Boundedness and cancellation assumptions.** Notice that (A.1) immediately implies the weak boundedness and diagonal BMO assumptions.

We claimed that  $T_{m_J}$  is a paraproduct free bilinear bi-parameter singular integral. It remains to check that for all bounded functions  $f_i: \mathbb{R}^n \rightarrow \mathbb{C}$ ,  $g_i: \mathbb{R}^m \rightarrow \mathbb{C}$ ,  $i = 1, 2$ , all

$$S_J \in \{T_{m_J}, T_{m_J}^{1*}, T_{m_J}^{2*}, (T_{m_J})_1^{1*}, (T_{m_J})_1^{2*}, (T_{m_J})_2^{1*}, (T_{m_J})_2^{2*}, (T_{m_J})_{1,2}^{1*,2*}, (T_{m_J})_{2,1}^{1*,2*}\}$$

and all rectangles  $R \subset \mathbb{R}^{n+m}$  there holds

$$\langle S_J(1 \otimes g_1, 1 \otimes g_2), h_R \rangle = \langle S_J(f_1 \otimes 1, f_2 \otimes 1), h_R \rangle = 0.$$

By symmetry it is enough to show  $\langle T_{m_J}(1 \otimes g_1, 1 \otimes g_2), h_R \rangle = 0$ . We simply have that (the integral converges as  $K_J$  is a Schwartz function)

$$T_{m_J}(1 \otimes g_1, 1 \otimes g_2)(x) = \iint_{\mathbb{R}^{n+m}} \iint_{\mathbb{R}^{n+m}} K_J(y_1, x_2 - y_2, z_1, x_2 - z_2) g_1(y_2) g_2(z_2) dy dz$$

is a constant function on  $x_1$ , and therefore the claim follows using  $\int_{\mathbb{R}^n} h_R(x_1, x_2) dx_1 = 0$ .

**Conclusion.** With some more care it is possible to prove that  $T_m$  itself is a paraproduct free bilinear bi-parameter CZO, and not only that

$$\langle T_m(f_1, f_2), f_3 \rangle = \lim_{J \rightarrow \infty} \langle T_{m_J}(f_1, f_2), f_3 \rangle,$$

where  $T_{m_J}$  are such uniformly on  $J$ .

**Funding:** The work of K.L. was supported by the Basque Government BERC 2018-2021 program; and Spanish Ministry of Economy and Competitiveness MINECO through Juan de la Cierva - Formación 2015 [FJCI-2015-24547], project [MTM2017-82160-C2-1-P] and BCAM Severo Ochoa excellence accreditation [SEV-2013-0323].

The work of H.M. was supported by the Academy of Finland [294840, 306901]; and the University of Helsinki three-year research grant [75160010].

The work of E.V. was supported by the Academy of Finland [306901, 307333].

## ACKNOWLEDGEMENTS

We thank Francesco Di Plinio and Yumeng Ou for useful conversations when the last two named authors visited MIT in December 2017. We thank MIT for the hospitality.

## REFERENCES

- [1] A. Barron, J. Pipher, Sparse domination for bi-parameter operators using square functions, preprint, arXiv:1709.05009, 2017.
- [2] L. Chaffee, Characterizations of bounded mean oscillation through commutators of bilinear singular integral operators, *Proc. Roy. Soc. Edinburgh Sect. A* 146 (2016) 1159–1166.
- [3] S.-Y. A. Chang, R. Fefferman, A continuous version of duality of  $H^1$  with BMO on the Bidisc, *Ann. of Math.* 112 (1980) 179–201.
- [4] S.-Y. A. Chang, R. Fefferman, Some recent developments in Fourier analysis and  $H^p$  theory on product domains, *Bull. Amer. Math. Soc.* 12 (1985) 1–43.
- [5] S.-Y. A. Chang, R. Fefferman, The Calderón–Zygmund Decomposition on Product Domains, *Amer. J. Math.* 104 (1982) 455–468.
- [6] S.-Y. A. Chang, J. M. Wilson, T. H. Wolff, Some weighted norm inequalities concerning the Schrödinger operators, *Comment. Math. Helv.* 60 (1985) 217–246.
- [7] J. Chen, G. Lu, Hörmander type theorems for multi-linear and multi-parameter Fourier multiplier operators with limited smoothness, *Nonlinear Anal.* 101 (2014) 98–112.
- [8] M. Christ, J.-L. Journé, Polynomial growth estimates for multilinear singular integral operators, *Acta Math.* 159 (1987) 51–80.
- [9] D. Cruz-Uribe, J.M. Martell, C. Pérez, Extrapolation results for  $A_\infty$  weights and applications, *J. Funct. Anal.* 213 (2004) 412–439.
- [10] R. Coifman, Y. Meyer, Au delà des opérateurs pseudo-différentiels, *Astérisque* 57 (1978) 1–185.
- [11] R. Coifman, R. Rochberg, G. Weiss, Factorization theorems for Hardy spaces in several variables, *Ann. of Math.* (2) 103 (1976) 611–635.
- [12] L. Dalenc, Y. Ou, Upper bound for multi-parameter iterated commutators *Publ. Mat.* 60 (2016) 191–220.
- [13] F. Di Plinio, Y. Ou, Banach-valued multilinear singular integrals, *Indiana Univ. Math. J.* 67 (2018) 1711–1763.
- [14] K. Domelevo, P. Ivanisvili, S. Petermichl, S. Treil, A. Volberg, On the failure of lower square function estimates in the non-homogeneous weighted setting, *Math. Ann.* (2018) <https://doi.org/10.1007/s00208-018-1787-4>.
- [15] J. Duoandikoetxea, Extrapolation of weights revisited: New proofs and sharp bounds, *J. Funct. Anal.* 260 (2011) 1886–1901.
- [16] R. Fefferman, Harmonic analysis on product spaces, *Ann. of Math.* 126 (1987) 109–130.
- [17] R. Fefferman,  $A^p$  weights and singular integrals, *Amer. J. Math.* 110 (1988) 975–987.
- [18] C. Fefferman, E. Stein, Some maximal inequalities, *Amer. J. Math.* 93 (1971) 107–115.
- [19] R. Fefferman, E. Stein, Singular integrals on product spaces, *Adv. Math.* 45 (1982) 117–143.
- [20] S. H. Ferguson, M. T. Lacey, A characterization of product BMO by commutators, *Acta Math.* 189 (2002) 143–160.
- [21] S. Ferguson, C. Sadosky, Characterizations of bounded mean oscillation on the polydisk in terms of Hankel operators and Carleson measures, *J. Anal. Math.* 81 (2000) 239–267.
- [22] L. Grafakos, J.M. Martell, Extrapolation of weighted norm inequalities for multivariable operators and applications, *J. Geom. Anal.* 14 (2004) 19–46.
- [23] L. Grafakos, D. He, H. V. Nguyen, L. Yan, Multilinear multiplier theorems and applications, *J. Fourier Anal. Appl.*, (2018) <https://doi.org/10.1007/s00041-018-9606-6>.
- [24] L. Grafakos, R. Torres, Multilinear Calderón–Zygmund theory, *Adv. Math.* 165 (2002) 124–164.
- [25] A. Grau de la Herrán, Comparison of  $T1$  conditions for multi-parameter operators, *Proc. Amer. Math. Soc.* 144 (2016) 2437–2443.
- [26] A. Grau de la Herrán, T. Hytönen, Dyadic representation and boundedness of non-homogeneous Calderón–Zygmund operators with mild kernel regularity, *Michigan Math. J.* 67 (2018) 757–786.
- [27] W. Guo, J. Lian, H. Wu, The unified theory for the necessity of bounded commutators and applications, preprint, arXiv:1709.08279, 2017.

- [28] I. Holmes, M. Lacey, B. Wick, Commutators in the two-weight setting, *Math. Ann.* 367 (2017) 51–80.
- [29] I. Holmes, M. Lacey, B. Wick, Bloom’s inequality: commutators in a two-weight setting, *Arch. Math. (Basel)* 106 (2016) 53–63.
- [30] I. Holmes, S. Petermichl, B. Wick, Weighted little bmo and two-weight inequalities for Journé commutators, *Anal. PDE* 11 (2018) 1693–1740.
- [31] T. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, *Ann. of Math.* 175 (2012) 1473–1506.
- [32] T. Hytönen, Representation of singular integrals by dyadic operators, and the  $A_2$  theorem, *Expo. Math.* 35 (2017) 166–205.
- [33] T. Hytönen, The two-weight inequality for the Hilbert transform with general measures, *Proc. Lond. Math. Soc.* 117 (2018) 483–526.
- [34] T. Hytönen, The  $L^p$ -to- $L^q$  boundedness of commutators with applications to the Jacobian operator, preprint, arXiv:1804.11167, 2018.
- [35] T. Hytönen, H. Martikainen, Non-homogeneous  $T1$  theorem for bi-parameter singular integrals, *Adv. Math.* 261 (2014) 220–273.
- [36] J.-L. Journé, A covering lemma for product spaces, *Proc. Amer. Math. Soc.* 96 (1986) 593–598.
- [37] J.-L. Journé, Calderón-Zygmund operators on product spaces, *Rev. Mat. Iberoam.* 1 (1985) 55–91.
- [38] J.-L. Journé, Two problems of Calderón-Zygmund theory on product-spaces, *Ann. Inst. Fourier* 38 (1988) 111–132.
- [39] I. Kunwar, Y. Ou, Two-weight inequalities for multilinear commutators, *New York J. Math.* 24 (2018) 980–1003.
- [40] M. T. Lacey, D. Mena, The sparse  $T1$  theorem, *Houston J. Math.* 43 (2017) 111–127.
- [41] M. Lacey, J. Metcalfe, Paraproducts in one and several parameters, *Forum Math.* 19 (2) (2007) 325–351.
- [42] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Multiparameter Riesz commutators, *Amer. J. Math.* 131 (2009) 731–769.
- [43] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Multi-parameter Div-Curl Lemmas, *Bull. London Math. Soc.* 10 (2012) 1123–1131.
- [44] A. Lerner, S. Ombrosi, I. Rivera-Ríos, On pointwise and weighted estimates for commutators of Calderón-Zygmund operators, *Adv. Math.* 319 (2017) 153–181.
- [45] A. Lerner, S. Ombrosi, I. Rivera-Ríos, Commutators of singular integrals revisited, *Bull. London Math. Soc.* 51 (2019) 107–119.
- [46] K. Li, J.M. Martell, H. Martikainen, S. Ombrosi, E. Vuorinen, End-point estimates, extrapolation for multilinear Muckenhoupt classes, and applications, preprint arxiv:1902.04951, 2019.
- [47] K. Li, H. Martikainen, Y. Ou, E. Vuorinen, Bilinear representation theorem, *Trans. Amer. Math. Soc.* 371 (2019) 4193–4214.
- [48] K. Li, H. Martikainen, E. Vuorinen, 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>.
- [49] K. Li, H. Martikainen, E. Vuorinen, Bloom type upper bounds in the product BMO setting, *J. Geom. Anal.* (2019) <https://doi.org/10.1007/s12220-019-00194-3>.
- [50] H. Martikainen, Representation of bi-parameter singular integrals by dyadic operators, *Adv. Math.* 229 (2012) 1734–1761.
- [51] C. Muscalu, J. Pipher, T. Tao, C. Thiele, Bi-parameter paraproducts, *Acta Math.* 193 (2) (2004) 269–296.
- [52] F. Nazarov, S. Treil, A. Volberg, The  $Tb$ -theorem on non-homogeneous spaces, *Acta Math.* 190 (2003) 151–239.
- [53] Y. Ou, Multi-parameter singular integral operators and representation theorem, *Rev. Mat. Iberoam.* 33 (2017) 325–350.
- [54] Y. Ou, S. Petermichl, E. Strouse, Higher order Journé commutators and characterizations of multi-parameter BMO, *Adv. Math.* 291 (2016) 24–58.
- [55] S. Petermichl, Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol, *C. R. Acad. Sci. Paris Sér. I Math.* 330 (2000) 455–460.
- [56] J. Pipher, L. A. Ward, BMO from dyadic BMO on the bidisc, *J. London Math. Soc.* 77 (2) (2008) 524–544.
- [57] S. Pott, P. Villarroya, A  $T(1)$  theorem on product spaces, preprint, arXiv:1105.2516, 2011.

- [58] C. Thiele, Wave packet analysis, CBMS Regional Conference Series in Mathematics 105. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006.
- [59] E. M. Stein, Harmonic Analysis: Real-variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press, Princeton, New Jersey, 1993.
- [60] S. Treil,  $H^1$  and dyadic  $H^1$ , in: Linear and complex analysis, in: Amer. Math. Soc. Transl. Ser. 2, vol. 226, Amer. Math. Soc., Providence, RI, 2009, pp. 179–193.
- [61] X. Tolsa, Analytic capacity, the Cauchy transform, and non-homogeneous Calderón–Zygmund theory, Progress in Mathematics, Vol. 307, Birkhäuser Verlag, Basel, 2014.

(K.L.) BCAM (BASQUE CENTER FOR APPLIED MATHEMATICS), ALAMEDA DE MAZARREDO 14, 48009 BILBAO, SPAIN

*E-mail address:* kangwei.nku@gmail.com

(H.M.) DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF HELSINKI, P.O.B. 68, FI-00014 UNIVERSITY OF HELSINKI, FINLAND

*E-mail address:* henri.martikainen@helsinki.fi

(E.V.) DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF HELSINKI, P.O.B. 68, FI-00014 UNIVERSITY OF HELSINKI, FINLAND

*E-mail address:* j.e.vuorin@gmail.com