BLOOM TYPE INEQUALITY FOR BI-PARAMETER SINGULAR INTEGRALS: EFFICIENT PROOF AND ITERATED COMMUTATORS

KANGWEI LI, HENRI MARTIKAINEN, AND EMIL VUORINEN

ABSTRACT. Utilising some recent ideas from our bilinear bi-parameter theory, we give an efficient proof of a two-weight Bloom type inequality for iterated commutators of linear bi-parameter singular integrals. We prove that if T is a bi-parameter singular integral satisfying the assumptions of the bi-parameter representation theorem, then

$$\|[b_k, \cdots [b_2, [b_1, T]] \cdots]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \prod_{i=1}^{k} \|b_i\|_{\mathrm{bmo}(\nu^{\theta_i})},$$

where $p \in (1, \infty)$, $\theta_i \in [0, 1]$, $\sum_{i=1}^{k} \theta_i = 1$, $\mu, \lambda \in A_p$, $\nu := \mu^{1/p} \lambda^{-1/p}$. Here A_p stands for the bi-parameter weights in $\mathbb{R}^n \times \mathbb{R}^m$ and $\mathrm{bmo}(\nu)$ is a suitable weighted little BMO space. We also simplify the proof of the known first order case.

1. INTRODUCTION

We recently developed in [23] a lot of theory for general bilinear bi-parameter singular integrals using modern dyadic analysis – in particular, we proved various bilinear bi-parameter commutator estimates. This lead us to discover an improved general principle for approaching bi-parameter commutator estimates of dyadic model operators. In this paper we use our method to give an efficient proof of Bloom type inequalities for iterated commutators of bi-parameter singular integrals. Our objective is to offer a proof with a very transparent structure. The iterated result is new in the bi-parameter setting, and its proof benefits greatly from this structure. Our proof of the first order case is short.

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

$$\|A^b\|_{L^p(\mu) \to L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}(\nu)},$$

where $p \in (1, \infty)$, $\mu, \lambda \in A_p$, $\nu := \mu^{1/p} \lambda^{-1/p}$, and $\mathrm{BMO}(\nu)$ is some suitable weighted BMO space. Usually A^b is some commutator, like $[b, T]f := bTf - T(bf)$, where T is a singular integral operator. Bloom [2] achieved such an inequality for $T = H$ – the Hilbert transform. Holmes–Lacey–Wick [12, 13] gave a modern proof and generalised Bloom's result to the case of a general (one-parameter) Calderón–Zygmund operator. The iterated case is by Holmes–Wick [15] (see also Hytönen [16] for a proof via the Cauchy integral trick). An improved iterated case is by Lerner–Ombrosi–Rivera-Ríos [21]: in [15, 16] there is some single $b \in \mathrm{BMO} \cap \mathrm{BMO}(\nu)$, while in [21] the iteration is taken using $b \in \mathrm{BMO}(\nu^{1/k}) \supset \mathrm{BMO} \cap \mathrm{BMO}(\nu)$ (see also the related paper [11] by García–Cuerva, Harboure, Segovia and Torrea). In [21] it is said that it seems that their bound cannot

2010 *Mathematics Subject Classification.* 42B20.
Key words and phrases. Representation theorems, iterated commutators, Bloom's inequality.

be obtained by a simple inductive argument. Some multilinear (one-parameter) Bloom type inequalities are considered by Kunwar–Ou [18]. Commutator estimates are in general very important and widely studied – for some other very recent references see e.g. Hytönen [17] and Lerner–Ombrosi–Rivera-Ríos [20].

A model of a bi-parameter singular integral operator in $\mathbb{R}^n \times \mathbb{R}^m$ is $T_1 \otimes T_2$, where T_1 and T_2 are usual singular integrals in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The general definition of a bi-parameter singular integral T requires that $\langle Tf_1, f_2\rangle$, $f_i = f_i^1 \otimes f_i^2$, can be written using different kernel representations depending on whether

(1) $\operatorname{spt} f_1^1 \cap \operatorname{spt} f_2^1 = \emptyset$ and $\operatorname{spt} f_1^2 \cap \operatorname{spt} f_2^2 = \emptyset$,
(2) $\operatorname{spt} f_1^1 \cap \operatorname{spt} f_2^1 = \emptyset$ or
(3) $\operatorname{spt} f_1^2 \cap \operatorname{spt} f_2^2 = \emptyset$.

In the first case we have a so-called full kernel representation, while in cases 2 and 3 a partial kernel representations holds in $\mathbb{R}^n$ or $\mathbb{R}^m$, respectively. The bi-parameter representation theorem [24] by one of us has enabled the development of deep commutator estimates also in the bi-parameter setting. The representation holds under natural $T1$ conditions involving the product BMO space of Chang and Fefferman [3] and some weak testing conditions. It allows to reduce the commutator estimates of singular integrals to those of model operators U, where U is a so-called bi-parameter shift, partial paraproduct or full paraproduct. We will only need these model operators in this paper and they are recalled in Section 4. As the somewhat lengthy kernel estimates and testing conditions of T are not explicitly needed here, we refer to [24] for the remaining details.

Using the dyadic representation theorem Ou, Petermichl and Strouse proved in [25] that $[b, T]\colon L^2(\mathbb{R}^{n+m}) \to L^2(\mathbb{R}^{n+m})$, when T is a paraproduct free bi-parameter singular integral and b is a little BMO function. This is the important base case for more complicated multi-parameter commutator estimates involving product BMO and iterated commutators of the form $[T_1, [b, T_2]]$ – see again [25] and Dalenc–Ou [7]. For the earlier deep commutator lower bounds in the Hilbert and Riesz settings see Ferguson–Lacey [10] and Lacey–Petermichl–Pipher–Wick [19]. The paper [25] was eventually generalised to concern all bi-parameter singular integrals satisfying $T1$ conditions by Holmes–Petermichl–Wick [14]. In fact, [14] proves much more: Bloom's inequality in the bi-parameter setting. The multi-parameter commutator scene is again very active, see also e.g. Duong–Li–Ou–Pipher–Wick [8], which is a very recent paper concerning commutators of multi-parameter flag singular integrals.

In [23] we explain that the presence of non-cancellative Haar functions h_I^0 in many of the bi-parameter model operators seem to have caused a lot of technical troubles in previous bi-parameter commutator estimates. 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 $\langle bf, h_I^0 \otimes h_J^0\rangle$. Moreover, when a non-cancellative Haar function appears a suitable average of b is added and subtracted. See Section 3 for the general details and e.g. (4.7) for an example of the resulting simple decomposition. In [14] everything was always reduced to a so called remainder term, which essentially entails expanding bf in the bi-parameter sense in all of the above situations. However, this remainder term has a particularly nice structure only when there are no non-cancellative Haar functions (the shift case) – otherwise it can lead to some difficult tail terms.

In this paper we want to use the above decomposition idea from [23] and showcase how it simplifies things in the linear bi-parameter setting. The Bloom setting is demanding, but the proof framework adapts nicely even to this generality. Our treatment of first order commutators is very different in many ways compared to [14] – that is, the simplifications in the decomposition itself, which are described above, are not the only difference – we also estimate differently. We exploit the known one-weight boundedness of the model operators even more: most terms arising from our new decomposition can be estimated directly by combining the weighted boundedness of the model operators and some Bloom type estimates of appropriate auxiliary operators, such as,

$$f \mapsto \sup_R \frac{1_R}{|R|} \int_R |b - \langle b \rangle_R||f|,$$

where $R = I \times J$ is a dyadic rectangle. The bounds for this maximal function presented in Proposition 4.13 rely on an interesting estimate of Fefferman [9] concerning the maximal function $f \mapsto \sup_R 1_R \langle |f| \rangle_R^\lambda$, $\langle f \rangle_R^\lambda := \lambda(R)^{-1} \int_R f \, \mathrm{d}\lambda$, defined using an A_∞-weight λ (see also Appendix B). Modern sparse domination methods are also useful in some parts of the proof – we use such estimates from [20] and [22, 23]. For example, a certain special term U^b associated to a model operator U and the commuting function b needs to be estimated directly. The estimate (4.11) that follows from sparse domination techniques is very effective for this.

We also can, for the first time, prove a Bloom type inequality for iterated commutators of bi-parameter singular integrals. Our main theorem is:

1.1. **Theorem.** *Let T be a bi-parameter singular integral satisfying the $T1$ type assumptions of the dyadic representation theorem* [24]. *Let also $p \in (1, \infty)$, $\mu, \lambda \in A_p$ and $\nu := \mu^{1/p}\lambda^{-1/p}$. Then we have*

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

and, more generally,

$$\|[b_k, \cdots [b_2, [b_1, T]] \cdots]\|_{L^p(\mu) \to L^p(\lambda)} \lesssim_{[\mu]_{A_p}, [\lambda]_{A_p}} \prod_{i=1}^k \|b_i\|_{\mathrm{bmo}(\nu^{\theta_i})},$$

where $\theta_i \in [0, 1]$ and $\sum_{i=1}^k \theta_i = 1$. Here A_p stands for the bi-parameter weights in $\mathbb{R}^n \times \mathbb{R}^m$ and $\mathrm{bmo}(\nu)$ is a suitable weighted little BMO space.

1.2. *Remark.* In the proof of the iterated commutator estimate, we will only prove the second order case, since the proof structure is such that it is clear how to continue the iteration.

Notice also that choosing $b_1 = \cdots = b_k = b$ and $\theta_1 = \cdots = \theta_k = 1/k$ we get a bi-parameter analog of [21], while choosing $\theta_1 = 1$ (and the rest zero) we get analogs of [15, 16]. However, the first is the better choice as $\mathrm{bmo}(\nu^{1/k}) \supset \mathrm{bmo} \cap \mathrm{bmo}(\nu)$. Indeed, similarly as in the one-parameter case [21], this is seen by using that $\langle \nu \rangle_R^\theta \lesssim_{[\nu]_{A_2}} \langle \nu^\theta \rangle_R$ for all $\theta \in (0, 1)$ and rectangles R (this estimate follows from Theorem 2.1 in [5] by iteration).

We also mention that some experts may find Appendix A interesting: it proves that little BMO is contained in the product BMO – even in the weighted situation – using only relatively elementary tools. We confess that we were only previously aware of a proof of this in the unweighted situation, and that proof depended on the deep commutator result of Ferguson–Lacey [10]. This weighted result is mentioned in [14] without proof.

2. Definitions and preliminaries

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 $f\colon \mathbb{R}^{n+m} \to \mathbb{C}$ and $h\colon \mathbb{R}^n \to \mathbb{C}$, then $\langle f, h\rangle_1 \colon \mathbb{R}^m \to \mathbb{C}$ is defined by

$$\langle f, h\rangle_1(x_2) = \int_{\mathbb{R}^n} f(y_1, x_2)h(y_1)\,\mathrm{d}y_1.$$

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 $\mathcal{D}^m$. 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, $\mathrm{ch}(I)$ denotes the dyadic children of I, i.e., $I' \in \mathrm{ch}(I)$ if $I' \in \mathcal{D}^n$, $I' \subset I$ and $\ell(I') = \ell(I)/2$. We sometimes write $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$.

When $I \in \mathcal{D}^n$ we denote by h_I a cancellative L^2 normalised Haar function. This means the following. Writing $I = I_1 \times \cdots \times I_n$ we can define the Haar function h_I^η, $\eta = (\eta_1, \ldots, \eta_n) \in \{0,1\}^n$, by setting

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

where $h_{I_i}^0 = |I_i|^{-1/2} 1_{I_i}$ and $h_{I_i}^1 = |I_i|^{-1/2}(1_{I_{i,l}} - 1_{I_{i,r}})$ for every $i = 1, \ldots, n$. Here $I_{i,l}$ and $I_{i,r}$ are the left and right halves of the interval I_i respectively. The reader should carefully notice that h_I^0 is the non-cancellative Haar function for us and that in some other papers a different convention is used. If $\eta \in \{0,1\}^n \setminus \{0\}$ the Haar function is cancellative: $\int h_I^\eta = 0$. We usually suppress the presence of η and simply write h_I for some h_I^η, $\eta \in \{0,1\}^n \setminus \{0\}$. Then $h_I h_I$ can stand for $h_I^{\eta_1} h_I^{\eta_2}$, but we always treat such a product as a non-cancellative function (which it is in the worst case scenario $\eta_1 = \eta_2$).

For $I \in \mathcal{D}^n$ and a locally integrable function $f\colon \mathbb{R}^n \to \mathbb{C}$, we define the martingale difference

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

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 η 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

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

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

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

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

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

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 η summations). Martingale blocks are defined in the natural way

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

2.3. **Weights.** A weight $w(x_1, x_2)$ (i.e. a locally integrable a.e. positive function) belongs to bi-parameter $A_p(\mathbb{R}^n \times \mathbb{R}^m)$, $1 < p < \infty$, 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). Here $1/p + 1/p' = 1$, i.e., p' is the dual exponent of p. We have

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

and that $\max\big(\operatorname{ess\,sup}_{x_1\in\mathbb{R}^n} [w(x_1,\cdot)]_{A_p(\mathbb{R}^m)}, \operatorname{ess\,sup}_{x_2\in\mathbb{R}^m} [w(\cdot,x_2)]_{A_p(\mathbb{R}^n)}\big) \le [w]_{A_p(\mathbb{R}^n\times\mathbb{R}^m)}$, 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. [14].

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

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

where the supremum is taken over all the cubes $Q \subset \mathbb{R}^n$. We will use that $A_p \subset A_\infty$, and also some estimates that are valid for A_∞ weights.

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

$$M_{\mathcal{D}^n} g(x) := \sup_{I\in\mathcal{D}^n} \frac{1_I(x)}{|I|}\int_I |g(y)|\,\mathrm{d}y$$

and

$$M_{\mathcal{D}^n,\mathcal{D}^m} f(x_1,x_2) := \sup_{R\in\mathcal{D}^n\times\mathcal{D}^m} \frac{1_R(x_1,x_2)}{|R|}\iint_R |f(y_1,y_2)|\,\mathrm{d}y_1\,\mathrm{d}y_2.$$

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. We record the following standard estimates, which are used repeatedly below.

2.1. **Lemma.** *For $p \in (1,\infty)$ and $w \in A_p(\mathbb{R}^n \times \mathbb{R}^m)$ the weighted square function estimates*

$$\|f\|_{L^p(w)} \sim_{[w]_{A_p(\mathbb{R}^n\times\mathbb{R}^m)}} \Big\|\Big(\sum_{\substack{I\in\mathcal{D}^n\\ J\in\mathcal{D}^m}} |\Delta_{I\times J} f|^2\Big)^{1/2}\Big\|_{L^p(w)}$$

$$\sim_{[w]_{A_p(\mathbb{R}^n\times\mathbb{R}^m)}} \Big\|\Big(\sum_{I\in\mathcal{D}^n} |\Delta_I^1 f|^2\Big)^{1/2}\Big\|_{L^p(w)} \sim_{[w]_{A_p(\mathbb{R}^n\times\mathbb{R}^m)}} \Big\|\Big(\sum_{J\in\mathcal{D}^m} |\Delta_J^2 f|^2\Big)^{1/2}\Big\|_{L^p(w)}$$

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

$$\Big\|\Big(\sum_j |Mf_j|^s\Big)^{1/s}\Big\|_{L^p(w)} \le C([w]_{A_p})\Big\|\Big(\sum_j |f_j|^s\Big)^{1/s}\Big\|_{L^p(w)}.$$

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

$$\|\varphi^1_{\mathcal{D}^n,\mathcal{D}^m} f\|_{L^p(w)} \sim_{[w]_{A_p}} \Big\|\Big(\sum_{I\in\mathcal{D}^n} \frac{1_I}{|I|}\otimes [M_{\mathcal{D}^m}\langle f, h_I\rangle_1]^2\Big)^{1/2}\Big\|_{L^p(w)} \lesssim_{[w]_{A_p}} \|f\|_{L^p(w)},$$

where

$$\varphi^1_{\mathcal{D}^n,\mathcal{D}^m} f := \sum_{I\in\mathcal{D}^n} h_I \otimes M_{\mathcal{D}^m}\langle f, h_I\rangle_1.$$

The function $\varphi^2_{\mathcal{D}^n,\mathcal{D}^m} f$ is defined in the symmetric way and satisfies the same estimates.

One easy way to show such estimates is to reduce to $p = 2$ via standard extrapolation. When $p = 2$ it is especially easy to use one-parameter results iteratively. See e.g. [4, 6] for one-parameter square function results and their history.

If an average is with respect to a different measure than the Lebesgue measure we can e.g. write $\langle f\rangle^\lambda_R := \frac{1}{\lambda(R)}\int_R f\,\mathrm{d}\lambda$, and similarly we can write $M_{\mathcal{D}^n,\mathcal{D}^m,\lambda} f = \sup_R 1_R\langle |f|\rangle^\lambda_R$.

2.5. **BMO spaces.** Given $w \in A_2(\mathbb{R}^n)$ we say that a locally integrable function $b\colon \mathbb{R}^n \to \mathbb{C}$ belongs to the weighted dyadic BMO space $\mathrm{BMO}_{\mathcal{D}^n}(w)$ if

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

The space $\mathrm{BMO}(w)$ can be defined using the norm defined by the supremum over all dyadic grids of the above dyadic norms.

Given $w \in A_2(\mathbb{R}^n\times\mathbb{R}^m)$ we say that a locally integrable function $b\colon \mathbb{R}^{n+m} \to \mathbb{C}$ belongs to the weighted dyadic little BMO space $\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)$ if

$$\|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)} := \sup_{R\in\mathcal{D}^n\times\mathcal{D}^m} \frac{1}{w(R)}\int_R |b - \langle b\rangle_R| < \infty.$$

Again, the space $\mathrm{bmo}(w)$ is defined via the supremum of the dyadic norms. We have

$$\|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)} \sim \max\big(\operatorname*{ess\,sup}_{x_1\in\mathbb{R}^n} \|b(x_1,\cdot)\|_{\mathrm{BMO}_{\mathcal{D}^m}(w(x_1,\cdot))}, \operatorname*{ess\,sup}_{x_2\in\mathbb{R}^m} \|b(\cdot,x_2)\|_{\mathrm{BMO}_{\mathcal{D}^n}(w(\cdot,x_2))}\big).$$

Moreover, we have the two-weight John–Nirenberg property

$$\|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(\nu)} \sim_{[\mu]_{A_p},[\lambda]_{A_p}} \sup_{R\in\mathcal{D}^n\times\mathcal{D}^m} \Big(\frac{1}{\mu(R)}\int_R |b - \langle b\rangle_R|^p\lambda\Big)^{1/p}, \tag{2.2}$$

if $p \in (1,\infty)$, $\mu, \lambda \in A_p$ and $\nu := \mu^{1/p}\lambda^{-1/p}$. Notice that here $\nu \in A_2$. For these see [14].

Finally, we have the product BMO space. Given $w \in A_2(\mathbb{R}^n \times \mathbb{R}^m)$ set

$$\|b\|_{\mathrm{BMO}^{\mathcal{D}^n,\mathcal{D}^m}_{\mathrm{prod}}(w)} := \sup_{\Omega} \Big(\frac{1}{w(\Omega)} \sum_{\substack{I\in\mathcal{D}^n, J\in\mathcal{D}^m\\ I\times J\subset\Omega}} |\langle b, h_I\otimes h_J\rangle|^2\langle w\rangle^{-1}_{I\times J}\Big)^{1/2},$$

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 $I \in \mathcal{D}^n$, $J \in \mathcal{D}^m$ so that $x \in I \times J \subset \Omega$. The non-dyadic product BMO space can be defined using the norm defined by the supremum over all dyadic grids of the above dyadic norms.

It is stated in [14] (without proof or reference) that $\mathrm{bmo}(w) \subset \mathrm{BMO}_{\mathrm{prod}}(w)$, $w \in A_2$. This embedding $\|b\|_{\mathrm{BMO}^{\mathcal{D}^n,\mathcal{D}^m}_{\mathrm{prod}}(w)} \lesssim_{[w]_{A_2}} \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)}$ is used in the main proof only

via the fact that it implies that (3.3) also holds for $\mathrm{bmo}(w)$ functions. We give a proof of this result in Appendix A.

2.6. **Commutators.** We briefly discuss one way to understand how the commutators are defined, and how all the pairings and expansions appearing in our proof can be seen to be well defined. For example, we discuss the second order case. Let $b_i \in \mathrm{bmo}(\nu^{\theta_i})$, $i = 1, 2$, be given, where $\nu = \mu^{1/p}\lambda^{-1/p}$, $\mu, \lambda \in A_p$, $p \in (1, \infty)$. Define

$$\mathcal{F} = \mathcal{F}(b_1, b_2) = \bigcup_{k=1}^{\infty} \{f \colon \mathbb{R}^{n+m} \to \mathbb{R} \colon \text{ spt } f \subset B(0,k) \cap \{|b_1|, |b_2| \le k\} \text{ and } |f| \le k\}.$$

For $f_1, f_2 \in \mathcal{F}$ the pairing $\langle [b_2, [b_1, T]] f_1, f_2 \rangle$ is well defined (if T is e.g. a singular integral satisfying the assumptions of the representation theorem) and $\mathcal{F}$ is dense in $L^p(\mu)$ and $L^{p'}(\lambda^{1-p'})$. Moreover, for some k we can replace b_i by $b_{i,k} = \max(\min(b_i, k), -k)$. Notice that $\|b_{i,k}\|_{\mathrm{bmo}(\nu^{\theta_i})} \lesssim \|b_i\|_{\mathrm{bmo}(\nu^{\theta_i})}$. This can be seen by using identities like $\max(c, d) = (c + d + |c - d|)/2$, and showing that $h \in \mathrm{bmo}(\nu)$ implies $|h| \in \mathrm{bmo}(\nu)$.

These considerations imply that below we may assume that the little BMO functions b_1, b_2 are bounded and f_1, f_2 are bounded and compactly supported, which makes everything legitimate.

3. Martingale difference expansions of products

We recall from [23] our modified strategy of expanding commutators. 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. Our protocol also entails the following: 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. We define certain standard paraproduct operators:

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

and

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

The operators are grouped into two collections, since they are handled differently (using product BMO or little BMO estimates, respectively).

We also define

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

The operators $a_1^2(b,f)$ and $a_2^2(b,f)$ are defined analogously.

3.1. **Lemma.** *Let π_b be $A_i(b,\cdot)$, $i=1,\dots,8$, or $a_j^1(b,\cdot)$, $a_j^2(b,\cdot)$, $j=1,2$. Suppose $b\in \mathrm{bmo}(\nu)$, where $\nu=\mu^{\frac{1}{p}}\lambda^{-\frac{1}{p}}$, $\mu,\lambda\in A_p$ and $p\in(1,\infty)$. Then*

$$\|\pi_b\|_{L^p(\mu)\to L^p(\lambda)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b\|_{\mathrm{bmo}(\nu)}.$$

3.2. *Remark.* Notice that $\nu=(\lambda^{1-p'})^{1/p'}(\mu^{1-p'})^{-1/p'}$, where $\lambda^{1-p'},\mu^{1-p'}\in A_{p'}$, so that the natural dual statement concerning $\|\pi_b\|_{L^{p'}(\lambda^{1-p'})\to L^{p'}(\mu^{1-p'})}$ follows.

Proof of Lemma 3.1. The operators $A_i(b,\cdot)$ (but in a somewhat different form) are already discussed in [14]. To aid the reader we note that the proofs essentially write themselves if one knows certain weighted H^1-BMO type duality estimates. For $i=1,\dots,4$ we use

$$\sum_{I,J}|\langle b,h_I\otimes h_J\rangle||A_{IJ}| \lesssim_{[\nu]_{A_2}} \|b\|_{\mathrm{BMO}_{\mathrm{prod}}^{\mathcal{D}^n,\mathcal{D}^m}(\nu)}\Big\|\Big(\sum_{I,J}|A_{IJ}|^2\frac{1_{I\times J}}{|I\times J|}\Big)^{1/2}\Big\|_{L^1(\nu)} \tag{3.3}$$

and the fact that $\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(\nu)\subset \mathrm{BMO}_{\mathrm{prod}}^{\mathcal{D}^n,\mathcal{D}^m}(\nu)$. A proof of (3.3) is recorded in [14] (but even this weighted version was well-known according to them). For $i=5,\dots,8$ we may use the one parameter analog of the estimate (3.3) in various ways – e.g. through the fact that it implies (as $[\langle\nu\rangle_{I,1}]_{A_2}\le[\nu]_{A_2}$ and $\|\langle b\rangle_{I,1}\|_{\mathrm{BMO}(\langle\nu\rangle_{I,1})}\le\|b\|_{\mathrm{bmo}(\nu)}$) that

$$\sum_J\Big|\Big\langle b,\frac{1_I}{|I|}\otimes h_J\Big\rangle\Big||A_{IJ}| \lesssim_{[\nu]_{A_2}} \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(\nu)}\Big\|\Big(\sum_J|A_{IJ}|^2\frac{1_J}{|J|}\Big)^{1/2}\Big\|_{L^1(\langle\nu\rangle_{I,1})}.$$

Of course, the operators $a_j^1(b,\cdot)$, $a_j^2(b,\cdot)$, $j=1,2$, can also be handled with the one parameter analog of (3.3). □

For $I_0\in\mathcal{D}^n$ and $J_0\in\mathcal{D}^m$ we will now introduce our expansions of $\langle bf,h_{I_0}\otimes h_{J_0}\rangle$, $\big\langle bf,h_{I_0}\otimes\frac{1_{J_0}}{|J_0|}\big\rangle$ and $\langle bf\rangle_{I_0\times J_0}$.

Expansion of $\langle bf,h_{I_0}\times h_{J_0}\rangle$. There holds

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

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 $\langle I_jII_i,h_{I_0}\otimes h_{J_0}\rangle$ we see that

$$\langle bf,h_{I_0}\otimes h_{J_0}\rangle=\sum_{i=1}^8\langle A_i(b,f),h_{I_0}\otimes h_{J_0}\rangle+\langle b\rangle_{I_0\times J_0}\langle f,h_{I_0}\otimes h_{J_0}\rangle. \tag{3.4}$$

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

$$\begin{aligned}\Big\langle bf,h_{I_0}\otimes\frac{1_{J_0}}{|J_0|}\Big\rangle&=\sum_{i=1}^2\Big\langle a_i^1(b,f),h_{I_0}\otimes\frac{1_{J_0}}{|J_0|}\Big\rangle\\&\quad+\big\langle(\langle b\rangle_{I_0,1}-\langle b\rangle_{I_0\times J_0})\langle f,h_{I_0}\rangle_1\big\rangle_{J_0}+\langle b\rangle_{I_0\times J_0}\Big\langle f,h_{I_0}\otimes\frac{1_{J_0}}{|J_0|}\Big\rangle.\end{aligned} \tag{3.5}$$

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

$$\langle bf\rangle_{I_0\times J_0} = \langle (b-\langle b\rangle_{I_0\times J_0})f\rangle_{I_0\times J_0} + \langle b\rangle_{I_0\times J_0}\langle f\rangle_{I_0\times J_0}. \tag{3.6}$$

All of our commutators are simply decomposed using (3.4), (3.5) (and its symmetric form) and (3.6) whenever the relevant pairings/averages appear.

4. First order commutator

Let $U = U^{k,v} = U^{k,v}_{\mathcal{D}^n,\mathcal{D}^m}$, $k=(k_i)$, $v=(v_i)$, $0\leq k_i\in\mathbb{Z}$ and $0\leq v_i\in\mathbb{Z}$, $i=1,2$, be a dyadic bi-parameter operator (defined using fixed dyadic grids $\mathcal{D}^n$ and $\mathcal{D}^m$) such that

$$\langle Uf_1,f_2\rangle = \sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\ \sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=I_2^{(k_2)}=K}}\ \sum_{\substack{J_1,J_2\in\mathcal{D}^m\\ J_1^{(v_1)}=J_2^{(v_2)}=V}} a_{K,V,(I_i),(J_j)}\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle,$$

where $a_{K,V,(I_i),(J_j)}$ are scalars and for all $i=1,2$ we have $\widetilde{h}_{I_i}=h_{I_i}$ (a cancellative Haar function) for all $I_i\in\mathcal{D}^n$ or $\widetilde{h}_{I_i}=1_{I_i}/|I_i|$ for all $I_i\in\mathcal{D}^n$, and similarly with the functions $\widetilde{h}_{J_j}$. To prove a Bloom type inequality for $[b,T]$, where T is a bi-parameter singular integral, it is enough to prove a Bloom type inequality for $[b,U]$, where U can be a so called bi-parameter shift, partial paraproduct or a full paraproduct (we will recall what these mean later). This is because of the dyadic bi-parameter representation theorem [24] – one only has to be maintain a polynomial dependence of k_1,k_2,v_1,v_2.

The basic structure is the following.

(1) The shift case: We have

$$\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle = \langle f_1,h_{I_1}\otimes h_{J_1}\rangle\langle f_2,h_{I_2}\otimes h_{J_2}\rangle.$$

(2) The partial paraproduct case: We have $k_1=k_2=0$ and

$$\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle = \Big\langle f_1,\frac{1_K}{|K|}\otimes h_{J_1}\Big\rangle\langle f_2,h_K\otimes h_{J_2}\rangle$$

or the symmetric case, or we have $v_1=v_2=0$ and

$$\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle = \Big\langle f_1,h_{I_1}\otimes\frac{1_V}{|V|}\Big\rangle\langle f_2,h_{I_2}\otimes h_V\rangle$$

or the symmetric case.

(3) The full paraproduct case: We have $k_1=k_2=v_1=v_2=0$ and

$$\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle = \langle f_1\rangle_{K\times V}\langle f_2,h_K\otimes h_V\rangle$$

or the symmetric case, or we have $k_1=k_2=v_1=v_2=0$ and

$$\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle = \Big\langle f_1,h_K\otimes\frac{1_V}{|V|}\Big\rangle\Big\langle f_2,\frac{1_K}{|K|}\otimes h_V\Big\rangle$$

or the symmetric case.

Most terms arising from our decomposition of $[b,U]$ can in fact be handled using the fact that all the model operators satisfy for all $1<p<\infty$ and $w\in A_p(\mathbb{R}^n\times\mathbb{R}^m)$ that

$$\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_i^{(k_i)}=K}}\sum_{\substack{J_1,J_2\in\mathcal{D}^m\\ J_j^{(v_j)}=V}}\left|a_{K,V,(I_i),(J_i)}\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle\right| \lesssim C([w]_{A_p})\|f_1\|_{L^p(w)}\|f_2\|_{L^{p'}(w^{1-p'})}. \tag{4.1}$$

Given some suitable BMO function b let us also define U^b via

$$\langle U^b f_1,f_2\rangle=\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_i^{(k_i)}=K}}\sum_{\substack{J_1,J_2\in\mathcal{D}^m\\ J_j^{(v_j)}=V}}a_{K,V,(I_i),(J_j)}[\langle b\rangle_{I_2\times J_2}-\langle b\rangle_{I_1\times J_1}] \times\langle f_1,\widetilde{h}_{I_1}\otimes\widetilde{h}_{J_1}\rangle\langle f_2,\widetilde{h}_{I_2}\otimes\widetilde{h}_{J_2}\rangle.$$

In the unweighted (or one weight case) the boundedness of U^b can be reduced to (4.1) via the simple observation that

$$|\langle b\rangle_{I_2\times J_2}-\langle b\rangle_{I_1\times J_1}|\lesssim\|b\|_{\mathrm{bmo}(\mathbb{R}^n\times\mathbb{R}^m)}\max(k_i,v_i).$$

However, if we want to prove a Bloom type inequality for U^b, and this is key for the Bloom type inequality for $[b,U]$, we have to run a harder adaptation of the proof of (4.1). This requires recalling more carefully what the assumptions about the coefficients $a_{K,V,\dots}$ are in each case. Notice also that $U^b=0$ when $k=v=0$ i.e. U^b does not arise in the full paraproduct case. Moreover, the Bloom type inequality for U^b is much harder when U is a partial paraproduct compared to the case that U is a shift (we use sparse bounds of bilinear paraproducts to handle the partial paraproduct case).

Despite having to deal with U^b separately, it is extremely convenient to blackbox (4.1). Such a weighted bound for all model operators was first recorded in [14]. The proof is essentially the same with or without weights (in the weighted case one just uses weighted versions of square function and maximal function bounds at the end). We note that a reader who is not familiar with the fundamental basic bound (4.1) can essentially read the proof from the current paper also. Indeed, for full paraproducts one can consult Lemma 3.1, and for the other model operators the bounds proved for U^b are harder, and in fact an easier version of those arguments can also be used to get (4.1).

4.1. **The shift case.** We show that if $U=U^{k,v}$ is a shift then

$$|\langle[b,U]f_1,f_2\rangle|\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}\|b\|_{\mathrm{bmo}(\nu)}(1+\max(k_i,v_i))\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.$$

Using our general decomposition philosophy from Section 3 we see that

$$\langle[b,U]f_1,f_2\rangle=\sum_{i=1}^{8}\langle Uf_1,A_i(b,f_2)\rangle-\sum_{i=1}^{8}\langle U(A_i(b,f_1)),f_2\rangle+\langle U^b f_1,f_2\rangle. \tag{4.2}$$

The first term is easy using Lemma 3.1 and (4.1) as

$$\begin{aligned}|\langle Uf_1,A_i(b,f_2)\rangle|&\le\|Uf_1\|_{L^p(\mu)}\|A_i(b,f_2)\|_{L^{p'}(\mu^{1-p'})}\\ &\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}\|b\|_{\mathrm{bmo}(\nu)}\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})},\end{aligned}$$

and the second one is handled similarly.

To handle U^b we begin by splitting

$$\begin{aligned}\langle b\rangle_{I_2\times J_2}-\langle b\rangle_{I_1\times J_1} &= [\langle b\rangle_{I_2\times J_2}-\langle b\rangle_{K\times J_2}]+[\langle b\rangle_{K\times J_2}-\langle b\rangle_{K\times V}]\\ &\quad+[\langle b\rangle_{K\times V}-\langle b\rangle_{K\times J_1}]+[\langle b\rangle_{K\times J_1}-\langle b\rangle_{I_1\times J_1}].\end{aligned}\tag{4.3}$$

The resulting four terms are essentially symmetric, so we only deal with the first one. There holds that

$$|\langle b\rangle_{I_2\times J_2}-\langle b\rangle_{K\times J_2}|\lesssim \|b\|_{\mathrm{bmo}(\nu)}\sum_{\substack{L\in\mathcal{D}^n\\ I_2\subsetneq L\subset K}}\frac{\nu(L\times J_2)}{|L\times J_2|}.\tag{4.4}$$

Using this we see that it is enough to fix one $l\in\{1,\dots,k_2\}$ and estimate the term

$$\begin{aligned}&\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{L\in\mathcal{D}^n\\ L^{(k_2-l)}=K}}\sum_{\substack{J_2\in\mathcal{D}^m\\ J_2^{(v_2)}=V}}\\ &\frac{\nu(L\times J_2)}{|L\times J_2|}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=K\\ I_2^{(l)}=L}}\sum_{\substack{J_1\in\mathcal{D}^m\\ J_1^{(v_1)}=V}}\big|a_{K,V,(I_i),(J_i)}\langle f_1,h_{I_1}\otimes h_{J_1}\rangle\langle f_2,h_{I_2}\otimes h_{J_2}\rangle\big|.\end{aligned}\tag{4.5}$$

Now we use the fact that

$$|a_{K,V,(I_i),(J_i)}|\le\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\frac{|J_1|^{1/2}|J_2|^{1/2}}{|V|},$$

which implies that

$$\begin{aligned}&\frac{\nu(L\times J_2)}{|L\times J_2|}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=K\\ I_2^{(l)}=L}}\sum_{\substack{J_1\in\mathcal{D}^m\\ J_1^{(v_1)}=V}}\big|a_{K,V,(I_i),(J_i)}\langle f_1,h_{I_1}\otimes h_{J_1}\rangle\langle f_2,h_{I_2}\otimes h_{J_2}\rangle\big|\\ &\le\nu(L\times J_2)\langle|\Delta^{k_1,v_1}_{K\times V}f_1|\rangle_{K\times V}\langle|\Delta^{k_2,v_2}_{K\times V}f_2|\rangle_{L\times J_2}\\ &\le\iint 1_{L\times J_2}M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_1,v_1}_{K\times V}f_1)M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_2,v_2}_{K\times V}f_2)\nu.\end{aligned}$$

Using this we see that (4.5) can be dominated by

$$\begin{aligned}&\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\iint M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_1,v_1}_{K\times V}f_1)M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_2,v_2}_{K\times V}f_2)\nu\\ &\le\Big\|\Big(\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}(M_{\mathcal{D}^n,\mathcal{D}^m}\Delta^{k_1,v_1}_{K\times V}f_1)^2\Big)^{1/2}\Big\|_{L^p(\mu)}\Big\|\Big(\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}(M_{\mathcal{D}^n,\mathcal{D}^m}\Delta^{k_2,v_2}_{K\times V}f_2)^2\Big)^{1/2}\Big\|_{L^{p'}(\lambda^{1-p'})}\\ &\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})},\end{aligned}$$

where in the second step we used that $\nu=\mu^{1/p}\lambda^{-1/p}$. We are done with the shifts.

4.2. **The partial paraproduct case.** We now deal with the partial paraproducts, and we choose the symmetry

$$\langle Uf_1, f_2\rangle = \sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}} \sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=I_2^{(k_2)}=K}} a_{K,V,(I_i)}\Big\langle f_1, h_{I_1}\otimes\frac{1_V}{|V|}\Big\rangle\langle f_2, h_{I_2}\otimes h_V\rangle. \tag{4.6}$$

We will show that

$$|\langle [b,U]f_1, f_2\rangle| \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b\|_{\mathrm{bmo}(\nu)}(1+\max(k_1,k_2))\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.$$

Using our general decomposition philosophy from Section 3 we see that

$$\begin{aligned}\langle [b,U]f_1, f_2\rangle &= \sum_{i=1}^{8}\langle Uf_1, A_i(b,f_2)\rangle - \sum_{i=1}^{2}\langle U(a_i^1(b,f_1)), f_2\rangle + \langle U^b f_1, f_2\rangle\\ &- \sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}} \sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=I_2^{(k_2)}=K}} a_{K,V,(I_i)}\big\langle(\langle b\rangle_{I_1,1}-\langle b\rangle_{I_1\times V})\langle f_1, h_{I_1}\rangle_1\big\rangle_V\langle f_2, h_{I_2}\otimes h_V\rangle.\end{aligned} \tag{4.7}$$

The first two terms are handled precisely as in the shift case. The last term is directly under control using (4.1) and the following lemma.

4.8. **Lemma.** *Let $p\in(1,\infty)$ and $\mu,\lambda\in A_p(\mathbb{R}^n\times\mathbb{R}^m)$. Assume that $b\in\mathrm{bmo}(\nu)$, where $\nu=\mu^{1/p}\lambda^{-1/p}$. Let $I\in\mathcal{D}^n$ and $J\in\mathcal{D}^m$. Then*

$$\big|\big\langle(\langle b\rangle_{I,1}-\langle b\rangle_{I\times J})\langle f, h_I\rangle_1\big\rangle_J\big| \lesssim_{[\nu]_{A_2}} \|b\|_{\mathrm{bmo}(\nu)}\Big\langle\varphi^{\nu,1}_{\mathcal{D}^n,\mathcal{D}^m}f, h_I\otimes\frac{1_J}{|J|}\Big\rangle,$$

where

$$\varphi^{\nu,1}_{\mathcal{D}^n,\mathcal{D}^m}f = \sum_{I\in\mathcal{D}^n} h_I\otimes M_{\mathcal{D}^m}(\langle f, h_I\rangle_1)\langle\nu\rangle_{I,1}.$$

Moreover, we have

$$\|\varphi^{\nu,1}_{\mathcal{D}^n,\mathcal{D}^m}\|_{L^p(\mu)\to L^p(\lambda)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} 1.$$

Proof. We will use the one parameter estimate

$$\frac{1}{|J|}\int_J |b_0-\langle b_0\rangle_J||g| \lesssim [w]_{A_\infty}\|b_0\|_{\mathrm{BMO}(w)}\frac{1}{|J|}\int_J M_{\mathcal{D}^m}(g)w.$$

Let us prove this. Using Lemma 5.1 in [20] we find a sparse family $\mathcal{S}=\mathcal{S}(J,b)$ such that

$$|b_0-\langle b_0\rangle_J|1_J \le 2^{n+2}\sum_{\substack{Q\in\mathcal{S}\\ Q\subset J}}\big\langle|b_0-\langle b_0\rangle_Q|\big\rangle_Q 1_Q. \tag{4.9}$$

Therefore, we have

$$\begin{aligned}\int_J |b_0-\langle b_0\rangle_J||g| &\lesssim \sum_{\substack{Q\in\mathcal{S}\\ Q\subset J}}\big\langle|b_0-\langle b_0\rangle_Q|\big\rangle_Q\int_Q|g|\\ &\le \|b_0\|_{\mathrm{BMO}(w)}\sum_{\substack{Q\in\mathcal{S}\\ Q\subset J}}\big\langle|g|\big\rangle_Q w(Q)\end{aligned}$$

$$\leq \|b_0\|_{\mathrm{BMO}(w)} \sum_{\substack{Q\in\mathcal{S}\\ Q\subset J}} [\langle M_{\mathcal{D}^m}(g)^{\frac{1}{2}}\rangle_Q^w]^2 w(Q)$$

$$\lesssim [w]_{A_\infty}\|b_0\|_{\mathrm{BMO}(w)} \int_J M_{\mathcal{D}^m}(g)w,$$

where in the last step we have used the Carleson embedding theorem (notice that Lebesgue sparse implies w-Carleson).

Now, we have

$$\begin{aligned}\big|\big\langle(\langle b\rangle_{I,1} - \langle b\rangle_{I\times J})\langle f,h_I\rangle_1\big\rangle_J\big| &\lesssim [\langle\nu\rangle_{I,1}]_{A_\infty}\|\langle b\rangle_{I,1}\|_{\mathrm{BMO}(\langle\nu\rangle_{I,1})}\big\langle M_{\mathcal{D}^m}(\langle f,h_I\rangle_1)\langle\nu\rangle_{I,1}\big\rangle_J\\ &= [\langle\nu\rangle_{I,1}]_{A_\infty}\|\langle b\rangle_{I,1}\|_{\mathrm{BMO}(\langle\nu\rangle_{I,1})}\Big\langle \varphi^{\nu,1}_{\mathcal{D}^n,\mathcal{D}^m} f, h_I\otimes\frac{1_J}{|J|}\Big\rangle,\end{aligned}$$

and then recall that $[\langle\nu\rangle_{I,1}]_{A_2}\leq[\nu]_{A_2}$ and $\|\langle b\rangle_{I,1}\|_{\mathrm{BMO}(\langle\nu\rangle_{I,1})}\leq\|b\|_{\mathrm{bmo}(\nu)}$.

Next, we have

$$\begin{aligned}\|\varphi^{\nu,1}_{\mathcal{D}^n,\mathcal{D}^m} f\|_{L^p(\lambda)} &\sim_{[\lambda]_{A_p}} \Big\|\Big(\sum_{I\in\mathcal{D}^n}\frac{1_I}{|I|}\otimes[M_{\mathcal{D}^m}(\langle f,h_I\rangle_1)\langle\nu\rangle_{I,1}]^2\Big)^{1/2}\Big\|_{L^p(\lambda)}\\ &\lesssim_{[\lambda]_{A_p}} \Big\|\Big(\sum_{I\in\mathcal{D}^n}\frac{1_I}{|I|}\otimes[M_{\mathcal{D}^m}\langle f,h_I\rangle_1]^2\Big)^{1/2}\nu\Big\|_{L^p(\lambda)}\\ &= \Big\|\Big(\sum_{I\in\mathcal{D}^n}\frac{1_I}{|I|}\otimes[M_{\mathcal{D}^m}\langle f,h_I\rangle_1]^2\Big)^{1/2}\Big\|_{L^p(\mu)} \lesssim_{[\mu]_{A_p}} \|f\|_{L^p(\mu)}.\end{aligned}$$

□

We now take care of the remaining U^b term. This key term also arises in [14] where it is actually omitted by saying that it goes similarly as a certain other term (which does not arise at all in our decomposition). To handle this term we find it necessary to use somewhat sophisticated tools via bilinear sparse domination.

Similarly as in the shift case it is enough to fix $l\in\{1,\dots,k_1\}$ and estimate the term

$$\begin{aligned}&\sum_{K\in\mathcal{D}^n}\sum_{\substack{L\in\mathcal{D}^n\\ L^{(k_1-l)}=K}}\int_{\mathbb{R}^n}\frac{1_L(x_1)}{|L|}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(l)}=L\\ I_2^{(k_2)}=K}}\\ &\qquad\sum_{V\in\mathcal{D}^m}\int_{\mathbb{R}^m}\frac{1_V(x_2)}{|V|}\Big|a_{K,V,(I_i)}\Big\langle f_1,h_{I_1}\otimes\frac{1_V}{|V|}\Big\rangle\langle f_2,h_{I_2}\otimes h_V\rangle\Big|\nu(x_1,x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1.\end{aligned}\tag{4.10}$$

From the sparse domination of bilinear paraproducts (see e.g. [22]) we can deduce (see Lemma 6.7 in [23]) that

$$\begin{aligned}&\int_{\mathbb{R}^m}\Big(\sum_{V\in\mathcal{D}^m}\big|a_{K,V,(I_i)}\big\langle\langle f_1,h_{I_1}\rangle_1\big\rangle_V\langle\langle f_2,h_{I_2}\rangle_1,h_V\rangle\big|\frac{1_V}{|V|}\Big)\rho\\ &\qquad\lesssim_{[\rho]_{A_\infty(\mathbb{R}^m)}}\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\int_{\mathbb{R}^m}M_{\mathcal{D}^m}(\langle f_1,h_{I_1}\rangle_1)M_{\mathcal{D}^m}(\langle f_2,h_{I_2}\rangle_1)\rho.\end{aligned}\tag{4.11}$$

This requires knowing that we have $\sup_{V_0\in\mathcal{D}^m}\Big(\frac{1}{|V_0|}\sum_{\substack{V\in\mathcal{D}^m\\ V\subset V_0}}|a_{K,V,(I_i)}|^2\Big)^{1/2}\leq\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}$. Recall the function $\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}$ from Lemma 2.1 and then notice the identity $M_{\mathcal{D}^m}(\langle f_1,h_{I_1}\rangle_1)=\langle\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_1,h_{I_1}\rangle_1$. We now see that (4.10) can be dominated in the $\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}$ sense by

$$\begin{aligned}
&\sum_{K\in\mathcal{D}^n}\sum_{\substack{L\in\mathcal{D}^n\\ L^{(k_1-l)}=K}}\iint\frac{1_L}{|L|}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(l)}=L\\ I_2^{(k_2)}=K}}\frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\langle\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_1,h_{I_1}\rangle_1\langle\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_2,h_{I_2}\rangle_1\nu\\
&\qquad\leq\sum_{K\in\mathcal{D}^n}\sum_{\substack{L\in\mathcal{D}^n\\ L^{(k_1-l)}=K}}\iint 1_L\langle|\Delta^1_{K,k_1}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_1|\rangle_{L,1}\langle|\Delta^1_{K,k_2}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_2|\rangle_{K,1}\nu\\
&\qquad\leq\sum_{K\in\mathcal{D}^n}\iint M^1_{\mathcal{D}^n}\Delta^1_{K,k_1}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_1\cdot M^1_{\mathcal{D}^n}\Delta^1_{K,k_2}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_2\cdot\nu\\
&\qquad\leq\Big\|\Big(\sum_{K\in\mathcal{D}^n}[M^1_{\mathcal{D}^n}\Delta^1_{K,k_1}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_1]^2\Big)^{1/2}\Big\|_{L^p(\mu)}\\
&\qquad\qquad\times\Big\|\Big(\sum_{K\in\mathcal{D}^n}[M^1_{\mathcal{D}^n}\Delta^1_{K,k_2}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m}f_2]^2\Big)^{1/2}\Big\|_{L^{p'}(\lambda^{1-p'})}\\
&\qquad\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.
\end{aligned}$$

4.3. **The full paraproduct case.** Depending on the form of U (we have two genuinely different symmetries here), we get different terms in the expansion (following Section 3) of $\langle[b,U]f_1,f_2\rangle$. However, after minor thought (recall also that $U^b=0$) the reader will understand that the only type of term that we have not seen before is

$$\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}a_{K,V}\big\langle(b-\langle b\rangle_{K\times V})f_1\big\rangle_{K\times V}\langle f_2,h_K\otimes h_V\rangle. \tag{4.12}$$

To handle this via (4.1) we introduce the following maximal function:

$$M^b_{\mathcal{D}^n,\mathcal{D}^m}(f)=\sup_R\frac{1_R}{|R|}\int_R|b-\langle b\rangle_R||f|,$$

where $R=I\times J\in\mathcal{D}^n\times\mathcal{D}^m$.

4.13. **Proposition.** *Let $p\in(1,\infty)$ and $b\in\mathrm{bmo}(\nu)$, where $\mu,\lambda\in A_p$ and $\nu=\mu^{1/p}\lambda^{-1/p}$. Then we have*

$$\|M^b_{\mathcal{D}^n,\mathcal{D}^m}\|_{L^p(\mu)\to L^p(\lambda)}\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}\|b\|_{\mathrm{bmo}(\nu)}.$$

Proof. There exists some $1<q<p$ such that $\mu,\lambda\in A_q$. Then Hölder's inequality implies that $\mu_0=\mu^{\frac{q}{p}}\lambda^{1-\frac{q}{p}}\in A_q$. Since $\mu_0^{\frac{1}{q}}\lambda^{-\frac{1}{q}}=\nu$, by the two-weight John-Nirenberg for little BMO (2.2) we have for all x and $R\ni x$ that

$$\begin{aligned}
\frac{1}{|R|}\int_R|b-\langle b\rangle_R||f|&\leq\Big(\frac{1}{|R|}\int_R|b-\langle b\rangle_R|^{q'}\mu_0^{1-q'}\Big)^{\frac{1}{q'}}\Big(\frac{1}{|R|}\int_R|f|^q\mu_0\Big)^{\frac{1}{q}}\\
&\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}\|b\|_{\mathrm{bmo}(\nu)}\frac{\lambda^{1-q'}(R)^{\frac{1}{q'}}}{|R|}\Big(\int_R|f|^q\mu_0\Big)^{\frac{1}{q}}
\end{aligned}$$

$$\lesssim_{[\lambda]_{A_p}} \|b\|_{\mathrm{bmo}(\nu)} \Big(\frac{1}{\lambda(R)}\int_R |f|^q \mu_0\Big)^{\frac{1}{q}}$$
$$\leq \|b\|_{\mathrm{bmo}(\nu)} M_{\mathcal{D}^n,\mathcal{D}^m,\lambda}(|f|^q\mu_0\lambda^{-1})(x)^{\frac{1}{q}}.$$

The claim now follows from the boundedness property $M_{\mathcal{D}^n,\mathcal{D}^m,\lambda}\colon L^{p/q}(\lambda) \to L^{p/q}(\lambda)$ and the observation that $(\mu_0\lambda^{-1})^{\frac{p}{q}}\lambda = \mu$. The first mentioned fact is non-trivial as λ is not of product form – but it has been proved by R. Fefferman in [9] using the A_∞ property of λ. For clarity we give a proof in our dyadic setting in Appendix B. □

Notice that

$$\big|\big\langle (b - \langle b\rangle_{K\times V})f_1\big\rangle_{K\times V}\big| \lesssim \langle M^b_{\mathcal{D}^n,\mathcal{D}^m} f_1\rangle_{K\times V}$$

so that (4.1) gives that the absolute value of (4.12) can be dominated with

$$C([\lambda]_{A_p})\|M^b_{\mathcal{D}^n,\mathcal{D}^m} f_1\|_{L^p(\lambda)}\|f_2\|_{L^{p'}(\lambda^{1-p'})} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b\|_{\mathrm{bmo}(\nu)}\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.$$

Here the last estimate used Lemma 4.13. We are done with the full paraproducts.

5. Iterated commutators

To study the Bloom type inequality for iterated commutators, we also need to consider the commutators of general paraproduct operators that appear in Section 3.

5.1. **Lemma.** *Let π_b be $A_i(b,\cdot)$, $i = 1,\cdots,8$ or $a^1_j(b,\cdot)$, $a^2_j(b,\cdot)$, $j = 1,2$. Suppose $b_1 \in \mathrm{bmo}(\nu^{\theta_1})$ and $b_2 \in \mathrm{bmo}(\nu^{\theta_2})$, where $\nu = \mu^{1/p}\lambda^{-1/p}$, $0 \leq \theta_1,\theta_2 \leq 1$, $\theta_1+\theta_2 = 1$ and $\mu,\lambda \in A_p$. Then*

$$\|[b_2,\pi_{b_1}]\|_{L^p(\mu)\to L^p(\lambda)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}.$$

Proof. With our existing tools there is no essential difference in the proof for different operators, and we e.g. choose $\pi_{b_1} = A_5(b_1,\cdot)$. We have

$$\langle A_5(b_1,f_1),f_2\rangle = \sum_{I,J}\Big\langle b_1, \frac{1_I}{|I|}\otimes h_J\Big\rangle\langle f_1, h_I\otimes h_J\rangle\langle f_2, h_I\otimes h_J h_J\rangle.$$

Using our decomposition philosophy (treating $h_J h_J$ as non-cancellative) we get

$$\begin{aligned}\langle [b_2,A_5(b_1,f_1)],f_2\rangle &= \sum_{i=1}^{2}\langle A_5(b_1,f_1),a^1_i(b_2,f_2)\rangle - \sum_{i=1}^{8}\langle A_5(b_1,A_i(b_2,f_1)),f_2\rangle \\ &+ \sum_{I,J}\Big\langle b_1,\frac{1_I}{|I|}\otimes h_J\Big\rangle\langle f_1,h_I\otimes h_J\rangle\big\langle(\langle b_2\rangle_{I,1} - \langle b_2\rangle_{I\times J})\langle f_2,h_I\rangle_1, h_J h_J\big\rangle.\end{aligned} \tag{5.2}$$

The first and second term are similar – we only deal with the second one. We begin with the only reasonable step:

$$|\langle A_5(b_1,A_i(b_2,f_1)),f_2\rangle| \leq \|A_5(b_1,A_i(b_2,f_1))\|_{L^p(\lambda)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.$$

We want to use the Bloom inequality for $A_5(b_1,\cdot)$ with $b_1 \in \mathrm{bmo}(\nu^{\theta_1})$. Thus, we write

$$\nu^{\theta_1} = (\mu^{\theta_1}\lambda^{1-\theta_1})^{1/p}\lambda^{-1/p}, \text{ where } \mu^{\theta_1}\lambda^{1-\theta_1},\lambda \in A_p,$$

and get

$$\|A_5(b_1,A_i(b_2,f_1))\|_{L^p(\lambda)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|A_i(b_2,f_1)\|_{L^p(\mu^{\theta_1}\lambda^{1-\theta_1})}.$$

Then we write

$$\nu^{\theta_2} = \mu^{1/p}(\mu^{\theta_1}\lambda^{1-\theta_1})^{-1/p}, \text{ where } \mu, \mu^{\theta_1}\lambda^{1-\theta_1} \in A_p,$$

and similarly get

$$\|A_i(b_2, f_1)\|_{L^p(\mu^{\theta_1}\lambda^{1-\theta_1})} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}\|f_1\|_{L^p(\mu)}.$$

For the third term in (5.2) we begin with Lemma 4.8, which gives us that

$$|\big\langle(\langle b_2\rangle_{I,1} - \langle b_2\rangle_{I\times J})\langle f_2, h_I\rangle_1, h_J h_J\big\rangle| \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}\Big\langle \varphi^{\nu^{\theta_2},1}_{\mathcal{D}^n,\mathcal{D}^m} f_2, h_I \otimes \frac{1_J}{|J|}\Big\rangle.$$

Then writing

$$\nu^{\theta_1} = \mu^{1/p}(\mu^{1-\theta_1}\lambda^{\theta_1})^{-1/p}, \text{ where } \mu, \mu^{1-\theta_1}\lambda^{\theta_1} \in A_p,$$

we get (again using the known Bloom for $A_5(b_1,\cdot)$) that the absolute value of the third term in (5.2) can be dominated in the $\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}}$ sense by

$$\|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|f_1\|_{L^p(\mu)}\|\varphi^{\nu^{\theta_2},1}_{\mathcal{D}^n,\mathcal{D}^m} f_2\|_{L^{p'}((\mu^{1-\theta_1}\lambda^{\theta_1})^{1-p'})}.$$

Then using Lemma 4.8 together with the identity

$$\nu^{\theta_2} = (\lambda^{1-p'})^{1/p'}((\mu^{1-\theta_1}\lambda^{\theta_1})^{1-p'})^{-1/p'}, \text{ where } \lambda^{1-p'}, (\mu^{1-\theta_1}\lambda^{\theta_1})^{1-p'} \in A_{p'},$$

we get

$$\|\varphi^{\nu^{\theta_2},1}_{\mathcal{D}^n,\mathcal{D}^m} f_2\|_{L^{p'}((\mu^{1-\theta_1}\lambda^{\theta_1})^{1-p'})} \lesssim \|f_2\|_{L^{p'}(\lambda^{1-p'})}.$$

We are done. □

5.1. **The shift case.** We show that if $U = U^{k,v}$ is a shift then

$$\begin{aligned}&|\langle[b_2,[b_1,U]]f_1, f_2\rangle| \\ &\quad\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}(1+\max(k_i,v_i))^2\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.\end{aligned}$$

We recall Equation (4.2) with $b = b_1$. In the iterated commutator $\langle[b_2,[b_1,U]]f_1, f_2\rangle$ the first term of (4.2) leads to the need to study

$$\langle Uf_1, A_i(b_1, b_2 f_2)\rangle - \langle U(b_2 f_1), A_i(b_1, f_2)\rangle,$$

which can be written (by adding and subtracting the obvious term) in the form

$$-\langle Uf_1, [b_2, A_i(b_1,\cdot)]f_2\rangle + \langle[b_2,U]f_1, A_i(b_1,f_2)\rangle.$$

We have using (4.1) and Lemma 5.1 that

$$\begin{aligned}|\langle Uf_1, [b_2, A_i(b_1,\cdot)]f_2\rangle| &\le \|Uf_1\|_{L^p(\mu)}\|[b_2, A_i(b_1,\cdot)]f_2\|_{L^{p'}(\mu^{1-p'})} \\ &\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.\end{aligned}$$

On the other hand, using the known Bloom type inequality for the first order commutator $[b_2, U]$ and also for $A_i(b_1,\cdot)$, we get arguing analogously as in the proof of Lemma 5.1 that

$$\begin{aligned}&|\langle[b_2,U]f_1, A_i(b_1,f_2)\rangle| \\ &\quad\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}(1+\max(k_i,v_i))\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.\end{aligned}$$

We have thus handled the contribution of the first term of (4.2) to $\langle[b_2,[b_1,U]]f_1, f_2\rangle$. The contribution of the second term of (4.2) to $\langle[b_2,[b_1,U]]f_1, f_2\rangle$ is handled in the same way.

Therefore, we are only left with bounding the contribution of the third term of (4.2) to $\langle[b_2,[b_1,U]]f_1, f_2\rangle$, i.e. bounding $\langle[b_2, U^{b_1}]f_1, f_2\rangle$. Expanding this as in (4.2), we are left

with some terms that can be handled using the already known Bloom type inequality for U^{b_1} and the Bloom type inequality for $A_i(b_2,\cdot)$, and also with the new term

$$\langle U^{b_1,b_2}f_1,f_2\rangle := \sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_i^{(k_i)}=K}}\sum_{\substack{J_1,J_2\in\mathcal{D}^m\\ J_j^{(v_j)}=V}} a_{K,V,(I_i),(J_i)}[\langle b_1\rangle_{I_2\times J_2}-\langle b_1\rangle_{I_1\times J_1}]$$
$$[\langle b_2\rangle_{I_2\times J_2}-\langle b_2\rangle_{I_1\times J_1}]\langle f_1,h_{I_1}\otimes h_{J_1}\rangle\langle f_2,h_{I_2}\otimes h_{J_2}\rangle.$$

To finish the shift case, we now bound U^{b_1,b_2}. We use (4.3) with $b=b_1$ and $b=b_2$. When we multiply these together, we get multiple different terms – we pick two representative ones

$$[\langle b_1\rangle_{K\times V}-\langle b_1\rangle_{K\times J_1}][\langle b_2\rangle_{K\times J_1}-\langle b_2\rangle_{I_1\times J_1}] \tag{5.3}$$

and

$$[\langle b_1\rangle_{I_2\times J_2}-\langle b_1\rangle_{K\times J_2}][\langle b_2\rangle_{K\times J_1}-\langle b_2\rangle_{I_1\times J_1}]. \tag{5.4}$$

The point is that in the first case we only have I_1,J_1 appearing in both terms (and no I_2,J_2), and the other one is a mixed case. Nevertheless, they can in fact be handled with completely analogous estimates. Therefore, we only deal with (5.3).

We use analogous estimates to (4.4), which leads to the need to bound

$$\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{H\in\mathcal{D}^m\\ H^{(v_1-h)}=V}}\frac{\nu^{\theta_1}(K\times H)}{|K\times H|}\sum_{\substack{L\in\mathcal{D}^n\\ L^{(k_1-l)}=K}}\sum_{\substack{J_1\in\mathcal{D}^m\\ J_1^{(h)}=H}}$$
$$\frac{\nu^{\theta_2}(L\times J_1)}{|L\times J_1|}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(l)}=L\\ I_2^{(k_2)}=K}}\sum_{\substack{J_2\in\mathcal{D}^m\\ J_2^{(v_2)}=V}}\big|a_{K,V,(I_i),(J_i)}\langle f_1,h_{I_1}\otimes h_{J_1}\rangle\langle f_2,h_{I_2}\otimes h_{J_2}\rangle\big|, \tag{5.5}$$

where $l\in\{1,\dots,k_1\}$ and $h\in\{1,\dots,v_1\}$. The second line of (5.5) is dominated by

$$\iint 1_{L\times J_1}M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_1,v_1}_{K\times V}f_1)\langle|\Delta^{k_2,v_2}_{K\times V}f_2|\rangle_{K\times V}\nu^{\theta_2}.$$

Therefore, continuing in the same way (5.5) can be dominated with

$$\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\iint M_{\mathcal{D}^n,\mathcal{D}^m}(M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_1,v_1}_{K\times V}f_1)\nu^{\theta_2})M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_2,v_2}_{K\times V}f_2)\nu^{\theta_1}.$$

Writing $\nu^{\theta_1}=(\mu^{\theta_1}\lambda^{\theta_2})^{1/p}\lambda^{-1/p}$ we can dominate this with

$$\Big\|\Big(\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}[M_{\mathcal{D}^n,\mathcal{D}^m}(M_{\mathcal{D}^n,\mathcal{D}^m}(\Delta^{k_1,v_1}_{K\times V}f_1)\nu^{\theta_2})]^2\Big)^{1/2}\Big\|_{L^p(\mu^{\theta_1}\lambda^{\theta_2})}$$
$$\times\Big\|\Big(\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}[M_{\mathcal{D}^n,\mathcal{D}^m}\Delta^{k_2,v_2}_{K\times V}f_2]^2\Big)^{1/2}\Big\|_{L^{p'}(\lambda^{1-p'})}.$$

We can conclude the case (5.3) by using Fefferman–Stein, square function estimates and also noting that $\nu^{\theta_2 p}\mu^{\theta_1}\lambda^{\theta_2}=\mu$. We are done with the shift case.

5.2. **The partial paraproduct case.** We show that if $U = U^k$ is a partial paraproduct of the form (4.6), then we have

$$|\langle [b_2,[b_1,U]]f_1, f_2\rangle| \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}(1+\max(k_1,k_2))^2\|f_1\|_{L^p(\mu)}\|f_2\|_{L^{p'}(\lambda^{1-p'})}.$$

Recall (4.7) with $b = b_1$. The contributions of the first two terms of (4.7) to $\langle [b_2,[b_1,U]]f_1, f_2\rangle$ are handled using the same general argument that we used with shifts.

We now bound the contribution of the third term of (4.7) to $\langle [b_2,[b_1,U]]f_1, f_2\rangle$, i.e. we bound $\langle [b_2,U^{b_1}]f_1, f_2\rangle$. Expanding this as in (4.7), the first two terms can be handled using the already known Bloom type inequality for U^{b_1} and the Bloom type inequality for $A_i(b_2,\cdot)$ and $a_i^1(b_2,\cdot)$, while the last term can be handled using Lemma 4.8 and the Bloom for U^{b_1}. Therefore, we are again facing the need to handle U^{b_1,b_2}.

Recall that $U = U^k$ is of the form (4.6). When written out, U^{b_1,b_2} includes terms of the form $(\langle b_1\rangle_{I_2\times V} - \langle b_1\rangle_{I_1\times V})(\langle b_2\rangle_{I_2\times V} - \langle b_2\rangle_{I_1\times V})$. As before, we split

$$\langle b_1\rangle_{I_2\times V} - \langle b_1\rangle_{I_1\times V} = [\langle b_1\rangle_{I_2\times V} - \langle b_1\rangle_{K\times V}] + [\langle b_1\rangle_{K\times V} - \langle b_1\rangle_{I_1\times V}],$$

and similarly with the function b_2. These multiplied together divides U^{b_1,b_2} into four parts, which are handled in the same way. To complement the case we handled with shifts (where we chose (5.3) instead of (5.4)), we choose here the part coming from the terms

$$(\langle b_1\rangle_{K\times V} - \langle b_1\rangle_{I_1\times V})(\langle b_2\rangle_{I_2\times V} - \langle b_2\rangle_{K\times V}).$$

We apply the estimate (4.4), and see that it suffices to bound the term

$$\sum_{K\in\mathcal{D}^n}\sum_{\substack{L_1,L_2\in\mathcal{D}^n\\ L_i^{(k_i-l_i)}=K}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_i^{(l_i)}=L_i}}\sum_{V\in\mathcal{D}^m}\frac{\nu^{\theta_1}(L_1\times V)}{|L_1\times V|}\frac{\nu^{\theta_2}(L_2\times V)}{|L_2\times V|} \times\Big|a_{K,V,(I_i)}\Big\langle f_1, h_{I_1}\otimes\frac{1_V}{|V|}\Big\rangle\langle f_2, h_{I_2}\otimes h_V\rangle\Big|, \tag{5.6}$$

where $l_i\in\{1,\dots,k_i\}$, $i\in\{1,2\}$.

If $\rho_1,\rho_2\in A_2(\mathbb{R}^m)$, then

$$\langle\rho_1\rangle_J\langle\rho_2\rangle_J \le [\rho_1]_{A_2}[\rho_2]_{A_2}\langle\rho_1\rho_2\rangle_J.$$

Indeed, by Hölder's inequality there holds for any cube $J\subset\mathbb{R}^m$ that

$$1 = \frac{1}{|J|}\int_J(\rho_1\rho_2)^{\frac13}\rho_1^{-\frac13}\rho_2^{-\frac13} \le \langle\rho_1\rho_2\rangle_J^{\frac13}\langle\rho_1^{-1}\rangle_J^{\frac13}\langle\rho_2^{-1}\rangle_J^{\frac13},$$

which combined with $\langle\rho_i^{-1}\rangle_J \le [\rho_i]_{A_2}\langle\rho_i\rangle_J^{-1}$ gives the claim. If $L_1,L_2\in\mathcal{D}^n$ and $V\in\mathcal{D}^m$, this shows that

$$\begin{aligned}\frac{\nu^{\theta_1}(L_1\times V)}{|L_1\times V|}\frac{\nu^{\theta_2}(L_2\times V)}{|L_2\times V|} &\le [\langle\nu^{\theta_1}\rangle_{L_1,1}]_{A_2}[\langle\nu^{\theta_2}\rangle_{L_2,1}]_{A_2}\big\langle\langle\nu^{\theta_1}\rangle_{L_1,1}\langle\nu^{\theta_2}\rangle_{L_2,1}\big\rangle_V\\ &\le [\nu]_{A_2}\big\langle\langle\nu^{\theta_1}\rangle_{L_1,1}\langle\nu^{\theta_2}\rangle_{L_2,1}\big\rangle_V.\end{aligned} \tag{5.7}$$

We turn to (5.6). If K, L_1, L_2, I_1 and I_2 are as in (5.6), then applying (5.7) one sees that the inner sum over $V \in \mathcal{D}^m$ is less than $[\nu]_{A_2}$ multiplied by

$$\int_{\mathbb{R}^m} \sum_{V\in\mathcal{D}^m} \big|a_{K,V,(I_i)}\big\langle\langle f_1, h_{I_1}\rangle_1\big\rangle_V \langle\langle f_2, h_{I_2}\rangle_1, h_V\rangle\big| \frac{1_V}{|V|}\langle\nu^{\theta_1}\rangle_{L_1,1}\langle\nu^{\theta_2}\rangle_{L_2,1}$$

$$\lesssim_{[\langle\nu^{\theta_1}\rangle_{L_1,1}\langle\nu^{\theta_2}\rangle_{L_2,1}]_{A_\infty}} \frac{|I_1|^{1/2}|I_2|^{1/2}}{|K|}\int_{\mathbb{R}^m} M_{\mathcal{D}^m}(\langle f_1, h_{I_1}\rangle_1) M_{\mathcal{D}^m}(\langle f_2, h_{I_2}\rangle_1)\langle\nu^{\theta_1}\rangle_{L_1,1}\langle\nu^{\theta_2}\rangle_{L_2,1},$$

where we used the application of sparse domination as in (4.11). Notice that Theorem 2.1 in [5] implies that $\langle\nu\rangle^{\theta_1}_{L_1,1} \lesssim_{[\nu]_{A_2}} \langle\nu^{\theta_1}\rangle_{L_1,1}$ (while the other direction is trivial by Hölder's inequality). This implies that $[\langle\nu^{\theta_1}\rangle_{L_1,1}\langle\nu^{\theta_2}\rangle_{L_2,1}]_{A_2} \leq C([\nu]_{A_2})$.

Recall the identity $M_{\mathcal{D}^m}(\langle f_i, h_{I_i}\rangle_1) = \langle\varphi^1_{\mathcal{D}^n,\mathcal{D}^m} f_i, h_{I_i}\rangle_1$. We sum the last estimate over K, L_1, L_2, I_1 and I_2, and move the summations inside the integral over $\mathbb{R}^m$. Then, for a fixed $x_2 \in \mathbb{R}^m$, we can use one parameter estimates in the same spirit that we used in connection with (5.5). This shows that (5.6) is dominated by $C([\mu]_{A_p}, [\lambda]_{A_p})$ multiplied by

$$\iint_{\mathbb{R}^{n+m}} \sum_{K\in\mathcal{D}^n} M^1_{\mathcal{D}^n}(\Delta^1_{K,k_1}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m} f_1) M^1_{\mathcal{D}^n}(M^1_{\mathcal{D}^n}(\Delta^1_{K,k_2}\varphi^1_{\mathcal{D}^n,\mathcal{D}^m} f_2)\nu^{\theta_2})\nu^{\theta_1}.$$

From here the estimate can be concluded by familiar steps. This ends our study of U^{b_1,b_2}.

To finish our treatment of partial paraproducts, we need to bound the contribution of the last term of (4.7) to $\langle[b_2,[b_1,U]]f_1, f_2\rangle$. Expanding using our usual rules leads us to the following sum of terms

$$\begin{aligned}
&\sum_{i=1}^{8}\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=I_2^{(k_2)}=K}} a_{K,V,(I_i)}\big\langle(\langle b_1\rangle_{I_1,1} - \big\langle b_1\rangle_{I_1\times V})\langle f_1, h_{I_1}\rangle_1\big\rangle_V \langle A_i(b_2,f_2), h_{I_2}\otimes h_V\rangle\\
&-\sum_{i=1}^{2}\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=I_2^{(k_2)}=K}} a_{K,V,(I_i)}\big\langle(\langle b_1\rangle_{I_1,1} - \big\langle b_1\rangle_{I_1\times V})\langle a^1_i(b_2,f_1), h_{I_1}\rangle_1\big\rangle_V \langle f_2, h_{I_2}\otimes h_V\rangle\\
&+\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=I_2^{(k_2)}=K}} a_{K,V,(I_i)}\big\langle(\langle b_1\rangle_{I_1,1} - \big\langle b_1\rangle_{I_1\times V})(\big\langle b_2\rangle_{I_1\times V} - \langle b_2\rangle_{I_1,1})\langle f_1, h_{I_1}\rangle_1\big\rangle_V\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\langle f_2, h_{I_2}\otimes h_V\rangle\\
&+\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}}\sum_{\substack{I_1,I_2\in\mathcal{D}^n\\ I_1^{(k_1)}=I_2^{(k_2)}=K}} a_{K,V,(I_i)}[\big\langle b_2\rangle_{I_2\times V} - \big\langle b_2\rangle_{I_1\times V}]\big\langle(\langle b_1\rangle_{I_1,1} - \big\langle b_1\rangle_{I_1\times V})\langle f_1, h_{I_1}\rangle_1\big\rangle_V\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\times\langle f_2, h_{I_2}\otimes h_V\rangle\\
&= I + II + III + IV.
\end{aligned}$$

The estimates for I and II follow quite directly from (4.1) and lemmas 3.1 and 4.8. The term IV can be handled using Lemma 4.8 and the Bloom type inequality of U^{b_2}. The term III requires a more careful treatment, which we will now proceed to give.

We will use the following one parameter estimate

$$\frac{1}{|J|}\int_J |b_1-\langle b_1\rangle_J||b_2-\langle b_2\rangle_J||g| \lesssim [w_1]_{A_\infty}[w_2]_{A_\infty}\|b_1\|_{\mathrm{BMO}(w_1)}\|b_2\|_{\mathrm{BMO}(w_2)} \times \frac{1}{|J|}\int_J [M_{\mathcal{D}^m}(M_{\mathcal{D}^m}(g)w_2)w_1 + M_{\mathcal{D}^m}(M_{\mathcal{D}^m}(g)w_1)w_2].$$

We can prove this by using (4.9) again:

$$\begin{aligned}
&\int_J |b_1-\langle b_1\rangle_J||b_2-\langle b_2\rangle_J||g| \\
&\lesssim \sum_{\substack{P\in\mathcal{S}_1\\ P\subset J}}\sum_{\substack{Q\in\mathcal{S}_2\\ Q\subset J}} \big\langle |b_1-\langle b_1\rangle_P|\big\rangle_P \big\langle |b_2-\langle b_2\rangle_Q|\big\rangle_Q \int_{P\cap Q}|g| \\
&\leq \|b_1\|_{\mathrm{BMO}(w_1)}\|b_2\|_{\mathrm{BMO}(w_2)} \sum_{\substack{P\in\mathcal{S}_1\\ P\subset J}}\sum_{\substack{Q\in\mathcal{S}_2\\ Q\subset P}} \langle w_1\rangle_P\langle w_2\rangle_Q \int_Q |g| \\
&+ \|b_1\|_{\mathrm{BMO}(w_1)}\|b_2\|_{\mathrm{BMO}(w_2)} \sum_{\substack{Q\in\mathcal{S}_2\\ Q\subset J}}\sum_{\substack{P\in\mathcal{S}_1\\ P\subset Q}} \langle w_1\rangle_P\langle w_2\rangle_Q \int_P |g| \\
&\lesssim [w_2]_{A_\infty}\|b_1\|_{\mathrm{BMO}(w_1)}\|b_2\|_{\mathrm{BMO}(w_2)} \sum_{\substack{P\in\mathcal{S}_1\\ P\subset J}} \langle w_1\rangle_P \int_P M_{\mathcal{D}^m}(g)w_2 \\
&+ [w_1]_{A_\infty}\|b_1\|_{\mathrm{BMO}(w_1)}\|b_2\|_{\mathrm{BMO}(w_2)} \sum_{\substack{Q\in\mathcal{S}_2\\ Q\subset J}} \langle w_2\rangle_Q \int_Q M_{\mathcal{D}^m}(g)w_1.
\end{aligned}$$

The desired one parameter estimate follows from this as previously. Define now

$$\varphi^{\nu_1,\nu_2,1}_{\mathcal{D}^n,\mathcal{D}^m} f_1 = \sum_{I\in\mathcal{D}^n} h_I \otimes M_{\mathcal{D}^m}(M_{\mathcal{D}^m}(\langle f_1,h_I\rangle_1)\langle\nu_1\rangle_{I,1})\langle\nu_2\rangle_{I,1},$$

and notice that we get

$$\begin{aligned}
&|\big\langle(\langle b_1\rangle_{I_1,1}-\big\langle b_1\rangle_{I_1\times V})(\big\langle b_2\rangle_{I_1\times V}-\langle b_2\rangle_{I_1,1})\langle f_1,h_{I_1}\rangle_1\big\rangle_V| \\
&\qquad \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}\Big\langle \varphi^{\nu^{\theta_1},\nu^{\theta_2},1}_{\mathcal{D}^n,\mathcal{D}^m} f_1 + \varphi^{\nu^{\theta_2},\nu^{\theta_1},1}_{\mathcal{D}^n,\mathcal{D}^m} f_1, h_{I_1}\otimes\frac{1_V}{|V|}\Big\rangle.
\end{aligned}$$

It is not hard to show (similarly as in Lemma 4.8) that

$$\big\|\varphi^{\nu^{\theta_1},\nu^{\theta_2},1}_{\mathcal{D}^n,\mathcal{D}^m} f_1\big\|_{L^p(\lambda)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|f_1\|_{L^p(\mu)}. \tag{5.8}$$

This, together with (4.1), ends our treatment of the term III. We are done with the partial paraproducts.

5.3. **The full paraproduct case.** The only term that arises here, which cannot be handled using exactly the same arguments that we have seen above with shifts and partial paraproducts, is

$$\sum_{\substack{K\in\mathcal{D}^n\\ V\in\mathcal{D}^m}} a_{K,V}\big\langle (b_1-\langle b_1\rangle_{K\times V})(b_2-\langle b_2\rangle_{K\times V})f_1\big\rangle_{K\times V}\langle f_2, h_K\otimes h_V\rangle.$$

The natural maximal function is now

$$M^{b_1,b_2}_{\mathcal{D}^n,\mathcal{D}^m} f = \sup_{R\in\mathcal{D}^n\times\mathcal{D}^m} \frac{1_R}{|R|}\int_R |b_1-\langle b_1\rangle_R||b_2-\langle b_2\rangle_R||f|.$$

We have

$$\|M^{b_1,b_2}_{\mathcal{D}^n,\mathcal{D}^m}\|_{L^p(\mu)\to L^p(\lambda)} \lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b_2\|_{\mathrm{bmo}(\nu^{\theta_2})}$$

as in Proposition 4.13. Indeed, using the same notation, and noticing that $\mu_0^{1-q'}, \lambda^{1-q'} \in A_{q'} \subset A_{q'/\theta}$ for any $\theta\in[0,1]$, we then have, for $\theta_1\in(0,1)$ and $x\in R$, that

$$\begin{aligned}
&\frac{1}{|R|}\int_R |b_1-\langle b_1\rangle_R||b_2-\langle b_2\rangle_R||f| \\
&\le \Big(\frac{1}{|R|}\int_R |f|^q\mu_0\Big)^{\frac{1}{q}}\Big(\frac{1}{|R|}\int_R |b_1-\langle b_1\rangle_R|^{\frac{q'}{\theta_1}}\mu_0^{1-q'}\Big)^{\frac{\theta_1}{q'}}\Big(\frac{1}{|R|}\int_R |b_2-\langle b_2\rangle_R|^{\frac{q'}{\theta_2}}\mu_0^{1-q'}\Big)^{\frac{\theta_2}{q'}} \\
&\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \Big(\frac{1}{|R|}\int_R |f|^q\mu_0\Big)^{\frac{1}{q}}\Big(\frac{\lambda^{1-q'}(R)}{|R|}\Big)^{\frac{\theta_1}{q'}}\|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\Big(\frac{\lambda^{1-q'}(R)}{|R|}\Big)^{\frac{\theta_2}{q'}}\|b\|_{\mathrm{bmo}(\nu^{\theta_2})} \\
&\lesssim \|b_1\|_{\mathrm{bmo}(\nu^{\theta_1})}\|b\|_{\mathrm{bmo}(\nu^{\theta_2})}\Big(\frac{1}{\lambda(R)}\int_R |f|^q\mu_0\Big)^{\frac{1}{q}}.
\end{aligned}$$

After this we can conclude as previously. We still comment on the case $\theta_1=0$. Let $s=q'/t'$, where $t=t([\lambda]_{A_p})\in(q,\frac{p+q}{2q'}+1)$ (these restrictions imply that $s>1$ and $p/st>1$) will be chosen later to be close enough to q. Set $\mu_1^{1-t'}=\mu_0^{1-q'}$. For $x\in R$ we have

$$\begin{aligned}
&\frac{1}{|R|}\int_R |b_1-\langle b_1\rangle_R||b_2-\langle b_2\rangle_R||f| \\
&\le \Big(\frac{1}{|R|}\int_R |b_1-\langle b_1\rangle_R|^{s'}\Big)^{\frac{1}{s'}}\Big(\frac{1}{|R|}\int_R |b_2-\langle b_2\rangle_R|^{st'}\mu_1^{1-t'}\Big)^{\frac{1}{st'}}\Big(\frac{1}{|R|}\int_R |f|^{st}\mu_1\Big)^{\frac{1}{st}} \\
&\lesssim_{[\mu]_{A_p},[\lambda]_{A_p}} \|b_1\|_{\mathrm{bmo}}\|b_2\|_{\mathrm{bmo}(\nu)}\Big(\frac{\lambda^{1-q'}(R)}{|R|}\Big)^{\frac{1}{q'}}\Big(\frac{1}{|R|}\int_R |f|^{st}\mu_1\Big)^{\frac{1}{st}} \\
&\le \|b_1\|_{\mathrm{bmo}}\|b_2\|_{\mathrm{bmo}(\nu)}[M_{\mathcal{D}^n,\mathcal{D}^m,\lambda}(M_{\mathcal{D}^n,\mathcal{D}^m}(f^{st}\mu_1)^{\frac{q}{st}}\lambda^{-1})(x)]^{\frac{1}{q}}.
\end{aligned}$$

To finish, we prove that $\lambda^{-\frac{p}{q}+1}\in A_{\frac{p}{st}}$. We first prove $\lambda^{-\frac{p}{q}+1}\in A_{\frac{p}{q}}$. We can take q close enough to p such that $(p/q)'>p$. Using Hölder's inequality with the exponent u, where $1/u:=(p-1)(p/q-1)<(p-1)(p'-1)=1$, we get

$$\frac{1}{|R|}\int_R \lambda^{-\frac{p}{q}+1}\Big(\frac{1}{|R|}\int_R\lambda\Big)^{\frac{p}{q}-1} \le \Big(\frac{1}{|R|}\int_R \lambda^{-\frac{1}{p-1}}\Big)^{(p-1)(\frac{p}{q}-1)}\Big(\frac{1}{|R|}\int_R\lambda\Big)^{\frac{p}{q}-1} \le [\lambda]_{A_p}^{\frac{p}{q}-1}.$$

Finally, we choose t very close to q – notice that if $t\to q$ then $s\to 1$ and so $p/st\to p/q$. Using the open property of $A_{p/q}$ weights we conclude that $\lambda^{-\frac{p}{q}+1}\in A_{\frac{p}{st}}$. Using this we can end the proof.

Appendix A. Embedding $\mathrm{bmo}(w)\subset \mathrm{BMO}_{\mathrm{prod}}(w)$

We give a proof of the embedding – we thank Prof. T. Hytönen for giving us an outline of the proof.

A.1. **Proposition.** *We have* $\|b\|_{\mathrm{BMO}_{\mathrm{prod}}^{\mathcal{D}^n,\mathcal{D}^m}(w)} \lesssim_{[w]_{A_2}} \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)}$ *if* $w \in A_2(\mathbb{R}^n \times \mathbb{R}^m)$.

Proof. First, using the ℓ^2-valued Kahane–Khintchine inequality notice that

$$\Big(\sum_{\substack{I\in\mathcal{D}^n\\ J\in\mathcal{D}^m}} |\Delta_{I\times J} f(x)|^2\Big)^{1/2} = \Big(\mathbb{E}\sum_I \Big|\sum_J \epsilon_J \Delta_{I\times J} f(x)\Big|^2\Big)^{1/2}$$
$$\sim \mathbb{E}\Big(\sum_I \Big|\Delta_I\Big(\sum_J \epsilon_J \Delta_J^2 f(\cdot, x_2)\Big)(x_1)\Big|^2\Big)^{1/2}.$$

Taking $L^1(w)$ norm, and using the known (see [6]) lower bound

$$\|g\|_{L^p(\rho)} \lesssim_{[\rho]_{A_\infty}} \Big\|\Big(\sum_I |\Delta_I g|^2\Big)^{1/2}\Big\|_{L^p(\rho)}, \qquad p \in (0,\infty),$$

we get

$$\Big\|\Big(\sum_{\substack{I\in\mathcal{D}^n\\ J\in\mathcal{D}^m}} |\Delta_{I\times J} f|^2\Big)^{1/2}\Big\|_{L^1(w)} \gtrsim_{[w]_{A_2}} \iint_{\mathbb{R}^{n+m}} \mathbb{E}'\Big|\sum_J \epsilon'_J \Delta_J^2 f(x_1,x_2)\Big| w(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$
$$\sim \Big\|\Big(\sum_{J\in\mathcal{D}^m} |\Delta_J^2 f|^2\Big)^{1/2}\Big\|_{L^1(w)}.$$

Next, using the weighted one parameter $H^1 - \mathrm{BMO}$ duality result (i.e. the one parameter analog of (3.3)), see Wu [26], we get

$$\begin{aligned}\Big|\iint_{\mathbb{R}^{n+m}} b(x) f(x)\,\mathrm{d}x\Big| &\lesssim_{[w]_{A_2}} \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)} \Big\|\Big(\sum_{J\in\mathcal{D}^m} |\Delta_J^2 f|^2\Big)^{1/2}\Big\|_{L^1(w)}\\ &\lesssim_{[w]_{A_2}} \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)} \Big\|\Big(\sum_{\substack{I\in\mathcal{D}^n\\ J\in\mathcal{D}^m}} |\Delta_{I\times J} f|^2\Big)^{1/2}\Big\|_{L^1(w)}.\end{aligned} \tag{A.2}$$

Borrowing a calculation from [14] we can conclude the proof. Indeed, for all Ω we have

$$\Big(\sum_{\substack{I\in\mathcal{D}^n, J\in\mathcal{D}^m\\ I\times J\subset\Omega}} |\langle b, h_I\otimes h_J\rangle|^2 \langle w\rangle_{I\times J}^{-1}\Big)^{1/2}$$
$$= \sup\Big\{\Big|\sum_{\substack{I\in\mathcal{D}^n, J\in\mathcal{D}^m\\ I\times J\subset\Omega}} \langle b, h_I\otimes h_J\rangle a_{I,J}\Big| : \sum_{\substack{I\in\mathcal{D}^n, J\in\mathcal{D}^m\\ I\times J\subset\Omega}} |a_{I,J}|^2 \langle w\rangle_{I\times J} = 1\Big\}.$$

Given such $a = (a_{I,J})$ define $f_a = \sum_{\substack{I,J\\ I\times J\subset\Omega}} a_{I,J} h_I \otimes h_J$. Then we have using (A.2) and Hölder's inequality that

$$\Big|\sum_{\substack{I\in\mathcal{D}^n, J\in\mathcal{D}^m\\ I\times J\subset\Omega}} \langle b, h_I\otimes h_J\rangle a_{I,J}\Big| = |\langle b, f_a\rangle|$$
$$\lesssim_{[w]_{A_2}} \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)} \Big\|\Big(\sum_{\substack{I\in\mathcal{D}^n, J\in\mathcal{D}^m\\ I\times J\subset\Omega}} |a_{I,J}|^2 \frac{1_I\otimes 1_J}{|I||J|}\Big)^{1/2}\Big\|_{L^1(w)}$$

$$\leq \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)} w(\Omega)^{1/2} \Big(\sum_{\substack{I\in\mathcal{D}^n, J\in\mathcal{D}^m \\ I\times J\subset\Omega}} |a_{I,J}|^2 \langle w\rangle_{I\times J} \Big)^{1/2} = \|b\|_{\mathrm{bmo}_{\mathcal{D}^n,\mathcal{D}^m}(w)} w(\Omega)^{1/2}.$$

We are done. □

A.3. *Remark.* Notice that actually the proof works just by assuming that w is uniformly in $A_\infty(\mathbb{R}^n)$ and $A_\infty(\mathbb{R}^m)$, and that we also do not need the full strength of the little BMO assumption: we can e.g. use $\operatorname{ess\,sup}_{x_1\in\mathbb{R}^n} \|b(x_1,\cdot)\|_{\mathrm{BMO}_{\mathcal{D}^m}(w(x_1,\cdot))}$ instead of the little BMO norm.

Appendix B. Boundedness of the strong maximal function

We give a proof of the following variant of a result of Fefferman [9]. The proof is quite clear in the dyadic setting. We note that we get a polynomial dependence on $[\lambda]_{A_p}$, while in Barron–Pipher [1] there seemed to be some exponential dependence.

B.1. **Proposition.** *Let $p\in(1,\infty)$ and $\lambda\in A_p(\mathbb{R}^n\times\mathbb{R}^m)$. Then for $s\in(1,\infty)$ we have*

$$\|M_{\mathcal{D}^n,\mathcal{D}^m,\lambda} f\|_{L^s(\lambda)} \lesssim [\lambda]_{A_p}^{1+1/s} \|f\|_{L^s(\lambda)}.$$

Proof. Write $\mathcal{D}=\mathcal{D}^n\times\mathcal{D}^m$. By interpolation it is enough to prove $\|M_{\mathcal{D},\lambda} f\|_{L^{s,\infty}(\lambda)} \lesssim [\lambda]_{A_p}^{1+1/s}\|f\|_{L^s(\lambda)}$. Fix f and $\alpha>0$, and set $\Omega=\Omega(\alpha)=\{M_{\mathcal{D},\lambda} f>\alpha\}$. Write $\Omega=\bigcup_{j=1}^\infty R_j$ for some rectangles $R_j\in\mathcal{D}$ with $\langle |f|\rangle^\lambda_{R_j}>\alpha$. It suffices to fix N and prove

$$\alpha\lambda\Big(\bigcup_{j=1}^N R_j\Big)^{1/s} \lesssim [\lambda]_{A_p}^{1+1/s}\|f\|_{L^s(\lambda)}. \tag{B.2}$$

Write $R_j=I_j\times J_j$, and reindex the cubes so that $\ell(J_{j+1})\leq \ell(J_j)$, $j=1,\dots,N-1$. We use Cordoba–Fefferman algorithm. Let $s_1=1$, and suppose $s_1<s_2<\dots<s_{l-1}<N$ have been chosen. Then s_l is defined, if it exists, to be the smallest integer $j\in(s_{l-1},N]$ so that

$$\Big|R_j\cap\bigcup_{i=1}^{l-1} R_{s_i}\Big| < \frac{|R_j|}{2}.$$

Write $\mathcal{J}=\{1,\dots,N\}$, $\mathcal{J}_s=\{s_i\}$ and $\mathcal{J}_s^c=\mathcal{J}\setminus\mathcal{J}_s$. Notice that for all $j_0\in\mathcal{J}$ we have

$$R_{j_0}\cap\bigcup_{\substack{j\in\mathcal{J}_s\\ j<j_0}} R_j = \Big[I_{j_0}\cap\bigcup_{\substack{j\in\mathcal{J}_s\\ j<j_0\\ J_{j_0}\subset J_j}} I_j\Big]\times J_{j_0}. \tag{B.3}$$

For $x_2\in\mathbb{R}^m$ we set $I_j(x_2)=I_j$ if $x_2\in J_j$, and $I_j(x_2)=\emptyset$ otherwise. Let $j_0\in\mathcal{J}_s^c$ be arbitrary. Then we have

$$\Big|R_{j_0}\cap\bigcup_{\substack{j\in\mathcal{J}_s\\ j<j_0}} R_j\Big| \geq \frac{|R_{j_0}|}{2}.$$

Using (B.3) we see that

$$\Big|I_{j_0}(x_2)\cap\bigcup_{j\in\mathcal{J}_s} I_j(x_2)\Big| \geq \Big|I_{j_0}(x_2)\cap\bigcup_{\substack{j\in\mathcal{J}_s\\ j<j_0}} I_j(x_2)\Big| \geq \frac{|I_{j_0}(x_2)|}{2}.$$

Using that for all cubes $I \subset \mathbb{R}^n$ and all subsets $E \subset I$ we have

$$\frac{\lambda(\cdot, x_2)(E)}{\lambda(\cdot, x_2)(I)} \geq [\lambda]_{A_p}^{-1} \Big(\frac{|E|}{|I|}\Big)^p \tag{B.4}$$

we conclude that

$$\lambda(\cdot, x_2)\Big(I_{j_0}(x_2) \cap \bigcup_{j \in \mathcal{J}_s} I_j(x_2)\Big) \geq c_1 [\lambda]_{A_p}^{-1} \lambda(\cdot, x_2)(I_{j_0}(x_2)), \qquad c_1 := 2^{-p}.$$

Since $j_0 \in \mathcal{J}_s^c$ was arbitrary we get for all $x_2 \in \mathbb{R}^m$ that

$$\bigcup_{j \in \mathcal{J}_s^c} I_j(x_2) \subset \Big\{ M_{\mathcal{D}^n, \lambda(\cdot, x_2)} \big(1_{\bigcup_{j \in J_s} I_j(x_2)}\big) \geq c_1 [\lambda]_{A_p}^{-1} \Big\}.$$

Using that $M_{\mathcal{D}^n, \lambda(\cdot, x_2)} \colon L^1(\lambda(\cdot, x_2)) \to L^{1,\infty}(\lambda(\cdot, x_2))$ (even with constant 1) we get

$$\lambda(\cdot, x_2)\Big(\bigcup_{j \in \mathcal{J}} I_j(x_2)\Big) \lesssim [\lambda]_{A_p} \lambda(\cdot, x_2)\Big(\bigcup_{j \in \mathcal{J}_s} I_j(x_2)\Big),$$

which, after integrating over $x_2 \in \mathbb{R}^m$, gives our first key inequality

$$\lambda\Big(\bigcup_{j \in \mathcal{J}} R_j\Big) \lesssim [\lambda]_{A_p} \lambda\Big(\bigcup_{j \in \mathcal{J}_s} R_j\Big). \tag{B.5}$$

Let now $j_0 \in \mathcal{J}_s$. Then by construction and using (B.3) we get for all x_2 that

$$\Big| I_{j_0}(x_2) \cap \bigcup_{\substack{j \in \mathcal{J}_s \\ j < j_0}} I_j(x_2) \Big| \leq \frac{|I_{j_0}(x_2)|}{2}.$$

Applying (B.4) to $E_{j_0}(x_2) := I_{j_0}(x_2) \setminus \bigcup_{\substack{j \in \mathcal{J}_s \\ j < j_0}} I_j(x_2)$ we have

$$\lambda(\cdot, x_2)(E_j(x_2)) \geq c_1 [\lambda]_{A_p}^{-1} \lambda(\cdot, x_2)(I_j(x_2)), \qquad j \in \mathcal{J}_s,\ x_2 \in \mathbb{R}^m.$$

Dualising against g with $\|g\|_{L^s(\lambda(\cdot, x_2))} \leq 1$, using the above sparseness property and using that $M_{\mathcal{D}^n, \lambda(\cdot, x_2)} \colon L^s(\lambda(\cdot, x_2)) \to L^s(\lambda(\cdot, x_2))$ (with a norm independent of λ) we get

$$\Big\| \sum_{j \in \mathcal{J}_s} 1_{I_j(x_2)} \Big\|_{L^{s'}(\lambda(\cdot, x_2))}^{s'} \lesssim [\lambda]_{A_p}^{s'} \lambda(\cdot, x_2)\Big(\bigcup_{j \in \mathcal{J}_s} I_j(x_2)\Big).$$

Integrating over $x_2 \in \mathbb{R}^m$ we get our second key inequality

$$\Big\| \sum_{j \in \mathcal{J}_s} 1_{R_j} \Big\|_{L^{s'}(\lambda)} \lesssim [\lambda]_{A_p} \lambda\Big(\bigcup_{j \in \mathcal{J}_s} R_j\Big)^{1/s'}. \tag{B.6}$$

Recalling that $\langle |f| \rangle_{R_j}^{\lambda} > \alpha$ and using (B.5) and (B.6) we get our claim (B.2):

$$\begin{aligned} \alpha \lambda\Big(\bigcup_{j=1}^N R_j\Big)^{1/s} &\lesssim [\lambda]_{A_p}^{1/s} \alpha \frac{\lambda\big(\bigcup_{j \in \mathcal{J}_s} R_j\big)}{\lambda\big(\bigcup_{j \in \mathcal{J}_s} R_j\big)^{1/s'}} \leq [\lambda]_{A_p}^{1/s} \frac{\sum_{j \in \mathcal{J}_s} \int_{R_j} |f| \lambda}{\lambda\big(\bigcup_{j \in \mathcal{J}_s} R_j\big)^{1/s'}} \\ &\leq [\lambda]_{A_p}^{1/s} \frac{\big\| \sum_{j \in \mathcal{J}_s} 1_{R_j} \big\|_{L^{s'}(\lambda)}}{\lambda\big(\bigcup_{j \in \mathcal{J}_s} R_j\big)^{1/s'}} \|f\|_{L^s(\lambda)} \lesssim [\lambda]_{A_p}^{1+1/s} \|f\|_{L^s(\lambda)}. \end{aligned}$$

□

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], BCAM Severo Ochoa excellence accreditation [SEV-2013-0323] and project [MTM2017-82160-C2-1-P] funded by (AEI/FEDER, UE) and acronym "HAQMEC".

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 Prof. T. Hytönen for his generous help with Appendix A. We also thank the referees for suggestions that helped to clarify the exposition. Part of this research was conducted when K. Li was visiting University of Helsinki – the hospitality of which is acknowledged. H.M. and E.V. are members of the Finnish Centre of Excellence in Analysis and Dynamics Research.

REFERENCES

[1] A. Barron, J. Pipher, Sparse domination for bi-parameter operators using square functions, preprint, arXiv:1709.05009, 2017.
[2] S. Bloom, A commutator theorem and weighted BMO, Trans. Amer. Math. Soc. 292 (1985) 103–122.
[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] D. Cruz–Uribe, J. M Martell, C. Pérez, Sharp weighted estimates for classical operators, Adv. Math. 229 (2012) 408–441.
[5] D. Cruz–Uribe, C.J. Neugebauer, The structure of the reverse Hölder classes, Trans. Amer. Math. Soc. 347 (1995) 2941–2960.
[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] L. Dalenc, Y. Ou, Upper bound for multi-parameter iterated commutators Publ. Mat. 60 (2016) 191–220.
[8] X.T. Duong, J. Li, Y. Ou, J. Pipher, B. Wick, Commutators of multi-parameter flag singular integrals and applications, Anal. PDE 12 (2019) 1325–1355.
[9] R. Fefferman, Strong differentiation with respect to measures, Amer. J. Math. 103 (1981) 33-40.
[10] S. H. Ferguson, M. T. Lacey, A characterization of product BMO by commutators, Acta Math. 189 (2002) 143–160.
[11] J. García–Cuerva, E. Harboure, C. Segovia, J. L. Torrea, Weighted norm inequalities for commutators of strongly singular integrals, Indiana Univ. Math. J. 40 (1991) 1397–1420.
[12] I. Holmes, M. Lacey, B. Wick, Commutators in the two-weight setting, Math. Ann. 367 (2017) 51–80.
[13] I. Holmes, M. Lacey, B. Wick, Bloom's inequality: commutators in a two-weight setting, Arch. Math. (Basel) 106 (2016) 53–63.
[14] I. Holmes, S. Petermichl, B. Wick, Weighted little bmo and two-weight inequalities for Journé commutators, Anal. PDE 11 (2018) 1693–1740.
[15] I. Holmes, B. Wick, Two weight inequalities for iterated commutators with Calderón-Zygmund operators, J. Operator Theory 79 (2018) 33–54.
[16] T. Hytönen, The Holmes-Wick theorem on two-weight bounds for higher order commutators revisited, Arch. Math. (Basel) 107 (2016) 389–395.
[17] T. Hytönen, The L^p-to-L^q boundedness of commutators with applications to the Jacobian operator, preprint, arXiv:1804.11167, 2018.
[18] I. Kunwar, Y. Ou, Two-weight inequalities for multilinear commutators, New York J. Math. 24 (2018) 980–1003.
[19] M. Lacey, S. Petermichl, J. Pipher, B. Wick, Multiparameter Riesz commutators, Amer. J. Math. 131 (2009) 731–769.
[20] A. Lerner, S. Ombrosi, I. Rivera-Ríos, On pointwise and weighted estimates for commutators of Calderón-Zygmund operators, Adv. Math. 319 (2017) 153–181.

[21] A. Lerner, S. Ombrosi, I. Rivera-Ríos, Commutators of singular integrals revisited, preprint arxiv:1709.04724, 2017.
[22] K. Li, H. Martikainen, Y. Ou, E. Vuorinen, Bilinear representation theorem, Trans. Amer. Math. Soc. 371 (2019) 4193–4214.
[23] K. Li, H. Martikainen, E. Vuorinen, Bilinear Calderón–Zygmund theory on product spaces, arXiv:1712.08135, 2018.
[24] H. Martikainen, Representation of bi-parameter singular integrals by dyadic operators, Adv. Math. 229 (2012) 1734–1761.
[25] Y. Ou, S. Petermichl, E. Strouse, Higher order Journé commutators and characterizations of multi-parameter BMO, Adv. Math. 291 (2016) 24–58.
[26] S. Wu, A wavelet characterization for weighted Hardy spaces, Rev. Mat. Iberoam. 8 (1992) 329–349.

(K.L.) BCAM (Basque Center for Applied Mathematics), Alameda de Mazarredo 14, 48009 Bilbao, Spain
E-mail address: kli@bcamath.org

(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: emil.vuorinen@helsinki.fi