SOME NEW WEIGHTED ESTIMATES ON PRODUCT SPACES

EMIL AIRTA, KANGWEI LI, HENRI MARTIKAINEN, AND EMIL VUORINEN

ABSTRACT. We complete our theory of weighted $L^p(w_1) \times L^q(w_2) \to L^r(w_1^{r/p} w_2^{r/q})$ estimates for bilinear bi-parameter Calderón–Zygmund operators under the assumption that $w_1 \in A_p$ and $w_2 \in A_q$ are bi-parameter weights. This is done by lifting a previous restriction on the class of singular integrals by extending a classical result of Muckenhoupt and Wheeden regarding weighted BMO spaces to the product BMO setting. We use this extension of the Muckenhoupt-Wheeden result also to generalise some two-weight commutator estimates from bi-parameter to multi-parameter. This gives a fully satisfactory Bloom type upper estimate for $[T_1, [T_2, \dots [b, T_k]]]$, where each T_i can be a completely general multi-parameter Calderón–Zygmund operator.

1. INTRODUCTION

Singular integral operators (SIOs) are operators of the form

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

They include many important linear transformations that arise in the analysis connecting geometric measure theory, partial differential equations, harmonic analysis and functional analysis. Classical one-parameter kernels are singular when $x = y$. Product space theory (multi-parameter theory), on the other hand, is concerned with kernels whose singularities are more spread out. To get an idea, for $x, y \in \mathbb{C} = \mathbb{R} \times \mathbb{R}$, compare the one-parameter Beurling kernel $1/(x-y)^2$ with the bi-parameter kernel $1/[(x_1 - y_1)(x_2 - y_2)]$ – the product of Hilbert kernels in both coordinate directions.

Multi-parameter SIOs arise naturally in applications involving a product type estimate. A simple example is given by the multiplier operators. A multiplier $m \colon \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ satisfying $|\partial^\alpha m(\xi)| \lesssim |\xi|^{-|\alpha|}$ for all multi-indices $\alpha = (\alpha_1, \alpha_2)$ and $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2 \setminus \{0\}$ gives rise to a convolution form one-parameter SIO T_m with $\widehat{T_m f}(\xi) = m(\xi)\widehat{f}(\xi)$. However, if m only satisfies *the less demanding* estimate $|\partial_{\xi_1}^{\alpha_1} \partial_{\xi_2}^{\alpha_2} m(\xi)| \lesssim |\xi_1|^{-|\alpha_1|} |\xi_2|^{-|\alpha|_2}$, we get a bi-parameter SIO T_m. For the classical linear multi-parameter theory and some of its original applications see e.g. Fefferman–Stein [16] and Journé [22].

On the other hand, a heuristic model of an n-linear SIO T in $\mathbb{R}^d$ is obtained by setting

$$T(f_1, \dots, f_n)(x) = U(f_1 \otimes \dots \otimes f_n)(x, \dots, x), \qquad x \in \mathbb{R}^d, \, f_i \colon \mathbb{R}^d \to \mathbb{C},$$

where U is a linear singular integral operator in $\mathbb{R}^{nd}$. See e.g. Grafakos–Torres [12] for the basic theory. Multilinear SIOs arise naturally from applications to partial differential equations, complex function theory and ergodic theory, among others. For instance, L^p estimates for the homogeneous fractional derivative $D^\alpha f = \mathcal{F}^{-1}(|\xi|^\alpha \widehat{f}(\xi))$ of a product of two or more functions, often referred to as *fractional Leibniz rules*, are widely employed in the study of dispersive equations. This started from the work of Kato and Ponce [23], and such estimates descend from the multilinear Hörmander-Mihlin multiplier theorem of Coifman-Meyer [4]. A variety of formulations may be found e.g. in Grafakos–Oh [11].

Finally, multilinear multi-parameter estimates arise naturally in applications whenever a multilinear phenomena, like the fractional Leibniz rules, are combined with product type estimates, such as those that arise when we want to take different *partial* fractional derivatives. We refer to our recent work [28] for a thorough background on the subject, its significance and for new developments. In the product setting this multilinear theory of SIOs is historically significantly more limited than in the one-parameter setting. For example, in the one-parameter – linear or multilinear – setting, the following Calderón–Zygmund type principle is standard: if an SIO is bounded with some exponents, it is bounded with all eligible exponents. In the linear bi-parameter setting such principles follow from [22] or [29], but they are already more involved. In [28] we finally were able to develop such general principles in the bilinear bi-parameter setting: simpler estimates in the Banach range ($r > 1$), imply boundedness in the full bilinear range $L^p \times L^q \to L^r$, $1/p + 1/q = 1/r$, $1 < p, q \le \infty$, $1/2 < r < \infty$, weighted estimates, mixed-norm estimates, and so on.

However, our weighted estimates in [28] (see Section 2 for the definition of A_p weights) still had the restriction that we needed the cancellation $T(1,1) = 0$, and the same for the adjoints and partial adjoints. It is easy to come up with singular integrals, where this cancellation does not hold: already a tensor product of two bilinear one-parameter SIOs does not in general satisfy it. In this paper we remove this final restriction, which leads to a complete and satisfactory theory for general bilinear bi-parameter SIOs.

The assumptions on objects like $T(1,1)$ have to do with $T1$ type arguments. Already in the linear one-parameter setting the question of the boundedness of T with a generic kernel is often best answered by so-called $T1$ theorems, where the action of the operator T on the constant function 1 plays a critical role, and it is only the convolution kernels $K(x,y) = k(x-y)$, which are conveniently studied via formulae like $\widehat{Tf}(\xi) = \widehat{k}(\xi)\widehat{f}(\xi)$. While the assumption $T1 = 0$ is more general than T being of convolution form, it is morally related.

The reason why we care about weighted estimates is that, beyond their significant intrinsic interest, they are of fundamental use in proving other (unweighted) estimates, like obtaining the full bilinear range of exponents from a single tuple (p_0, q_0, r_0) and proving vector-valued and mixed-norm estimates. This is due to the very powerful bilinear extrapolation results – see e.g. [8, 10, 24, 25, 31]. In the product setting this viewpoint is particularly useful as many of the classical one-parameter tools are crudely missing. However, extrapolation is already very convenient in the one-parameter multilinear theory due to, for example, the complicated nature of multilinear interpolation.

For example, we obtained a range of mixed-norm $L^{p_1}L^{p_2}$ estimates in [28] via weighted estimates, and, thus, they required the same restriction on the class of SIOs. See e.g. Di Plinio–Ou [7] for some previous mixed-norm estimates. With the understanding that a

Calderón–Zygmund operator (CZO) is an SIO satisfying natural $T1$ type assumptions, our improvement of [28] now reads:

1.1. **Theorem.** *Let T be a bilinear bi-parameter CZO as defined in* [28]. *Then we have the weighted estimate*

$$\|T(f,g)\|_{L^r(w)} \leq C([w_1]_{A_p}, [w_2]_{A_q})\|f\|_{L^p(w_1)}\|g\|_{L^q(w_2)}$$

for all $1 < p, q < \infty$ and $1/2 < r < \infty$ with $1/r = 1/p + 1/q$, 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 $w = w_1^{r/p} w_2^{r/q}$. In the unweighted case we also have the mixed-norm estimates

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

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 $\infty \notin \{p_1, q_1\}$.

Bilinear weights pose a problem with duality: notice e.g. that if $w_1, w_2 \in A_4$ then $w := w_1^{1/2} w_2^{1/2} \in A_4$, while we need to work in $L^2(w)$. This is a relevant problem and often makes bilinear bi-parameter weighted estimates different and harder than in the linear case. For the linear estimates see Fefferman–Stein [16] and Fefferman [13, 14], and the much more recent Holmes–Petermichl–Wick [21] that is rooted on the modern dyadic-probabilistic methods [29]. In particular, already some linear paraproduct estimates depend on suitable H^1-BMO type duality arguments, and, for this reason, we could not previously handle weighted estimates for certain model operators. This led to the restriction on the class of CZOs. We now remove this restriction by developing some new theory for the product BMO space of Chang and Fefferman [2, 3]. Theorem 3.2 below is an extension of a classical result of Muckenhoupt and Wheeden [30] to the product BMO setting – it says that certain weighted product BMO spaces are actually the same as the unweighted product BMO space. This gives a useful way to construct objects in some genuinely weighted product BMO spaces by starting with an object in the unweighted product BMO. We prove this result of independent interest in its full generality, and apply its special case to prove our bilinear bi-parameter weighted estimates.

Finally, we present another application of Theorem 3.2. This concerns the recently hot topic of two-weight estimates for commutators of SIOs. The classical result of Coifman–Rochberg–Weiss [5] showed that

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

for a class of non-degenerate one-parameter SIOs T. Commutator estimates then e.g. yield factorizations for Hardy functions, imply div-curl lemmas relevant in compensated compactness, and are connected to the Jacobian problem $Ju = f$ in L^p (see Hytönen [19]). The two-weight problem concerns estimates from $L^p(\mu)$ to $L^p(\lambda)$ for two different weights μ, λ and has attracted significant interest after the recent work by Holmes–Lacey–Wick [21]. For the product space versions of these two-weight estimates see the recent works [1, 21, 26, 27]. In this paper we remove the final restriction on the most general estimate thus far, [1, Theorem 1.5], and prove a two-weight upper bound on $[T_1, [T_2, \ldots [b, T_k]]]$, where each T_i can be a completely general multi-parameter CZO as in [32]. Previously, in certain situations it was explicitly required in [1] that T_i can be at most bi-parameter, or else it needs to satisfy some extra cancellation.

1.2. **Theorem.** *Let $\mathbb{R}^d = \prod_{i=1}^m \mathbb{R}^{d_i}$ be a product space of m-parameters. Let $k \le m$ be given and $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_k\}$ be a partition of $\{1, \ldots, m\}$. Let $k_i = \#\mathcal{I}_i$, and for each each $i = 1, \ldots, k$ let T_i be a multi-parameter CZO as in [32], which is defined in the product space $\prod_{j \in \mathcal{I}_i} \mathbb{R}^{d_j}$ and has k_i parameters. Let $b\colon \mathbb{R}^d \to \mathbb{C}$, $p \in (1, \infty)$, $\mu, \lambda \in A_p(\mathbb{R}^d)$ be m-parameter weights and $\nu = \mu^{1/p}\lambda^{-1/p}$ be the associated Bloom weight. Then we have*

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

where in this formula every T_i is extended to act on the whole product space $\mathbb{R}^d$, and $\mathrm{bmo}^{\mathcal{I}}(\nu)$ is a suitable weighted little product BMO space as in [1].

We manage this extension in Section 5 by using the full power of Theorem 3.2. For the various definitions see Section 2 and Section 5.

1.1. **In-depth discussion.** In [1, Theorem 1.5] two-weight commutator estimates were shown for

$$[T_1, [T_2, \ldots [b, T_k]]],$$

where each T_i is a multi-parameter CZO as in Ou [32] (a multi-parameter dyadic representation theorem generalising [29]). However, it was required that either T_i satisfies some cancellation (is paraproduct free) or is of at most two parameters. In Theorem 1.2 we remove the restrictions completing the upper bound theory of multi-parameter two-weight commutator bounds. We stress that this is philosophically different to what we do in Theorem 1.1 – there we aim for complete *bi-parameter* theory in the strictly different bilinear context by removing some previous cancellation restrictions present in the bilinear bi-parameter theory [28]. In [1] no restrictions appear on the bi-parameter case to begin with – restrictions are needed only if some of the operators T_i are m-parameter, $m \ge 3$.

The passage from bi-parameter to multi-parameter is known to be very non-trivial in some occasions. In [1] difficulties arose as certain so-called partial paraproduct terms (model operators appearing in the dyadic representation given by [32]) seemed to require a sparse domination type treatment already used in [26], and this could only be done if the paraproduct component of the partial paraproduct had only one parameter. In tri-parameter theory (and general multi-parameter theory) it was then necessary to assume that such partial paraproducts vanish (that is, the SIOs have some extra cancellation), as otherwise partial paraproducts with multi-parameter paraproduct components arise.

All of this is somehow connected to the very innocent result of [28, Lemma 6.7]. This particular one-parameter lemma is proved using sparse domination and can be extrapolated to get useful vector-valued estimates, which we need in various places in [28, 26]. We now know how to prove some strong enough analogs of [28, Lemma 6.7] without sparse domination and using Theorem 3.2 instead, and this is extendable to multi-parameter. See Lemma 5.5 below. This allows us to remove the restrictions present in [1]. The reader might still reasonably wonder would it now be easy enough to state a multi-parameter version of Theorem 1.1 as well. The short answer is that this might still require more work. This is because the linear two-weight theory only requires some special versions of [28, Lemma 6.7], where averages are outside as in Lemma 5.5, but the weighted estimates for bilinear bi-parameter partial paraproducts [28, Proposition 6.11] use all of the symmetries of [28, Lemma 6.7].

2. NOTATIONS AND PRELIMINARIES

Basic notation. Throughout this paper $A \lesssim B$ means that $A \leq CB$ with some constant C that we deem unimportant to track at that point. In particular, we often do not track the dependence on the weight constants. We write $A \sim B$ if $A \lesssim B \lesssim A$.

Before Section 5 we work in the bi-parameter setting in the product space $\mathbb{R}^{n+m} = \mathbb{R}^n \times \mathbb{R}^m$. We write $x = (x_1, x_2)$ with $x_1 \in \mathbb{R}^n$ and $x_2 \in \mathbb{R}^m$ and for $f : \mathbb{R}^{n+m} \to \mathbb{C}$ and $h : \mathbb{R}^n \to \mathbb{C}$, we define

$$\langle f, h \rangle_1(x_2) := \int_{\mathbb{R}^n} f(x_1, x_2) h(x_1) \, \mathrm{d}x_1.$$

In Section 5, which is the only place in the paper where we do multi-parameter theory (as opposed to bi-parameter), some additional notation will be introduced.

Haar functions. We denote a dyadic grid in $\mathbb{R}^n$ by $\mathcal{D}^n$ and a dyadic grid in $\mathbb{R}^m$ by $\mathcal{D}^m$. We often write $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ for the related dyadic rectangles.

For an interval $I \subset \mathbb{R}$ we denote by I_l and I_r the left and right halves of I, respectively. We define $h_I^0 = |I|^{-1/2} 1_I$ and $h_I^1 = |I|^{-1/2}(1_{I_l} - 1_{I_r})$. Let now $Q = I_1 \times \cdots \times I_n \in \mathcal{D}^n$, and define the Haar function h_Q^η, $\eta = (\eta_1, \ldots, \eta_n) \in \{0, 1\}^n$, by setting

$$h_Q^\eta = h_{I_1}^{\eta_1} \otimes \cdots \otimes h_{I_n}^{\eta_n}.$$

If $\eta \neq 0$ the Haar function is cancellative: $\int h_Q^\eta = 0$. We exploit notation by suppressing the presence of η, and write h_Q for some h_Q^η, $\eta \neq 0$. If $R = I \times J \in \mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$, we define $h_R = h_I \otimes h_J$.

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

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

where the supremum is taken over $R = I \times J$, 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

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

and that $\max \big(\operatorname{ess\,sup}_{x_1 \in \mathbb{R}^n} [w(x_1, \cdot)]_{A_p(\mathbb{R}^m)}, \operatorname{ess\,sup}_{x_2 \in \mathbb{R}^m} [w(\cdot, x_2)]_{A_p(\mathbb{R}^n)} \big) \leq [w]_{A_p(\mathbb{R}^n \times \mathbb{R}^m)}$, while the constant $[w]_{A_p}$ is dominated by the maximum to some power. We say $w \in A_\infty(\mathbb{R}^n \times \mathbb{R}^m)$ if

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

It is well-known (see e.g. [9, Section 7]) that

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

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

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. A modern reference for the basic theory of bi-parameter weights is e.g. [21].

Square functions and maximal functions. 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

$$M_{\mathcal{D}^n} g := \sup_{I \in \mathcal{D}^n} 1_I \langle |g| \rangle_I \quad \text{and} \quad M_{\mathcal{D}} f := \sup_{R \in \mathcal{D}} 1_R \langle |f| \rangle_R,$$

where $\langle f \rangle_A = |A|^{-1} \int_A f$. We can e.g. write $\langle f \rangle_I^1$ if $I \subset \mathbb{R}^n$ and we average only over the first parameter. 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. The weighted maximal function is defined by

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

where $\langle |f| \rangle_R^w = w(R)^{-1} \int_R |f| w$. We require the following very nice result of Fefferman [15]. For a modern reference see [26, Proposition B.1]. Notice that in the bi-parameter setting this result is very non-trivial as w is not of tensor form.

2.1. Lemma. *Let $w \in A_\infty(\mathbb{R}^n \times \mathbb{R}^m)$. Then for all $1 < p \le \infty$ we have*

$$\|M_{\mathcal{D}}^w f\|_{L^p(w)} \lesssim \|f\|_{L^p(w)}.$$

Now define the square functions

$$S_{\mathcal{D}} f = \left(\sum_{R \in \mathcal{D}} |\langle f, h_R \rangle|^2 \frac{1_R}{|R|} \right)^{1/2}, \quad S_{\mathcal{D}^n}^1 f = \left(\sum_{I \in \mathcal{D}^n} \frac{1_I}{|I|} \otimes |\langle f, h_I \rangle_1|^2 \right)^{1/2}$$

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

$$S_{\mathcal{D},M}^1 f = \left(\sum_{I \in \mathcal{D}^n} \frac{1_I}{|I|} \otimes \left[M_{\mathcal{D}^m} \langle f, h_I \rangle_1 \right]^2 \right)^{1/2}, \quad S_{\mathcal{D},M}^2 f = \left(\sum_{J \in \mathcal{D}^m} \left[M_{\mathcal{D}^n} \langle f, h_J \rangle_2 \right]^2 \otimes \frac{1_J}{|J|} \right)^{1/2}.$$

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

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

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

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

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

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

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

See [26, Lemma 2.1] for an indication on how to prove this standard result. Of key importance to us is the following lower square function estimate valid for A_∞ weights.

2.3. **Lemma.** *There holds*

$$\|f\|_{L^p(w)} \lesssim \|S^1_{\mathcal{D}^n} f\|_{L^p(w)}$$

and

$$\|f\|_{L^p(w)} \lesssim \|S_{\mathcal{D}} f\|_{L^p(w)}$$

for all $p \in (0, \infty)$ and bi-parameter weights $w \in A_\infty$.

See [28, Section 6] for an explanation of this well-known inequality. This is important to us as the weight $w = w_1^{r/p} w_2^{r/q}$ in Theorem 1.1 is at least A_∞ – it is in fact A_{2r}.

Singular integrals. A kernel $K \colon \mathbb{R}^d \times \mathbb{R}^d \setminus \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d \colon x = y\} \to \mathbb{C}$ is a standard Calderón-Zygmund kernel on $\mathbb{R}^d$ if we have

$$|K(x, y)| \le \frac{C}{|x - y|^d}$$

and, for some $\alpha \in (0, 1]$,

$$|K(x, y) - K(x', y)| + |K(y, x) - K(y, x')| \le C \frac{|x - x'|^\alpha}{|x - y|^{d + \alpha}}$$

whenever $|x - x'| \le |x - y|/2$.

A singular integral operator (SIO) is a linear operator T (initially defined, for example, on bounded and compactly supported functions) so that there is a standard kernel K for which

$$\langle Tf, g \rangle = \iint_{\mathbb{R}^d \times \mathbb{R}^d} K(x, y) f(y) g(x) \, \mathrm{d}y \, \mathrm{d}x$$

whenever the functions f and g are nice and have disjoint supports. A Calderón–Zygmund operator (CZO) is an SIO T, which satisfies the $T1$ condition

$$\int_I |T1_I| + \int_I |T^* 1_I| \lesssim |I|$$

for all cubes $I \subset \mathbb{R}^d$. Here T^* is the linear adjoint of T. A $T1$ theorem says that an SIO is a CZO if and only if it is bounded from $L^p \to L^p$ for all (equivalently for some) $p \in (1, \infty)$. We know a lot about the structure of a CZO T. Indeed, we can represent T with certain dyadic model operators (DMOs) – see [17, 18]. The DMOs take the concrete form of so-called dyadic shifts and paraproducts. Moreover, the $T1$ theorem is a consequence of the representation theorem. We will indicate how a representation theorem looks soon.

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 $T1$ conditions are more complicated than in one-parameter – for the complete definitions see [29, Section 2]. As in the one-parameter case, we can represent a bi-parameter CZO – a bi-parameter SIO satisfying the bi-parameter $T1$ conditions – using bi-parameter DMOs. We will only

need these model operators in this paper, and they are defined when they are needed. Everything in Theorem 1.1 and Theorem 1.2 is reduced, via representation theorems, to estimates of model operators.

The analogous multi-parameter theory is presented in [32]. An inherent complication of these bi-parameter and multi-parameter representations is the presence not only of "pure" paraproducts and cancellative shifts, but also of their hybrid combinations – partial paraproducts – that are completely new compared to the one-parameter case.

For Theorem 1.1 we still need to discuss the bilinear theory. A heuristic model of a bilinear one-parameter SIO T in $\mathbb{R}^d$ is $T(f_1, f_2)(x) := U(f_1 \otimes f_2)(x, x)$, where $x \in \mathbb{R}^d$, $f_i \colon \mathbb{R}^d \to \mathbb{C}$, $(f_1 \otimes f_2)(x_1, x_2) = f_1(x_1)f_2(x_2)$ and U is a linear SIO in $\mathbb{R}^{2d}$. In detail, a bilinear SIO T has a kernel K satisfying estimates that are obtained from the above heuristic via the linear estimates, and if $\operatorname{spt} f_i \cap \operatorname{spt} f_j = \emptyset$ for some i, j then

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

Finally, 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 $f_1, g_1 \colon \mathbb{R}^n \to \mathbb{C}$, $f_2, g_2 \colon \mathbb{R}^m \to \mathbb{C}$, $x = (x_1, x_2) \in \mathbb{R}^{n+m}$ and T_n, T_m are bilinear SIOs defined in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. See [28, Section 3] for the rather long definition in the full generality. We reiterate that we will only meet the DMOs in this paper. This is because by [28, Theorem 1.3] we have that if T is a bilinear bi-parameter CZO, then

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

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 $\mathcal{D}_\omega$. Crucially, we have shown the desired weighted estimates in [28] for all but one type of model operators – the remaining model operators are defined carefully in Section 4 and the weighted estimate is proved for them. This shows Theorem 1.1.

3. WEIGHTED BMO SPACES

Let $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ be a lattice of dyadic rectangles, $A = (a_R)_{R\in\mathcal{D}}$ be a sequence of scalars and $\Omega \subset \mathbb{R}^{n+m}$. We define

$$S_A(x) = \left(\sum_{R\in\mathcal{D}} |a_R|^2 \frac{1_R(x)}{|R|} \right)^{1/2} \quad \text{and} \quad S_{A,\Omega}(x) = \left(\sum_{\substack{R\in\mathcal{D} \\ R\subset\Omega}} |a_R|^2 \frac{1_R(x)}{|R|} \right)^{1/2}.$$

Let $p \in (0, \infty)$. Define

$$\|A\|_{\mathrm{BMO}_{\mathrm{prod}}(p)} = \sup_\Omega \frac{1}{|\Omega|^{1/p}} \|S_{A,\Omega}\|_{L^p},$$

where Ω is open and $0 < |\Omega| < \infty$. There are many possibilities how to define a weighted version. The following is not the 'correct' definition for many things. Nonetheless, it will

be of key use to us. Thus, let $w \in A_\infty$ and set

$$\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(p)} = \sup_\Omega \frac{1}{w(\Omega)^{1/p}} \|S_{A,\Omega}\|_{L^p(w)}.$$

We set $\|A\|_{\mathrm{BMO}_{\mathrm{prod}}} = \|A\|_{\mathrm{BMO}_{\mathrm{prod}}(2)}$ and $\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}} = \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(2)}$. The weight turns out not to play a role here – that is, we have $\|A\|_{\mathrm{BMO}_{\mathrm{prod}}} = \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}$ for all bi-parameter weights $w \in A_\infty$. To prove this we need the bi-parameter John-Nirenberg. The unweighted version $\|A\|_{\mathrm{BMO}_{\mathrm{prod}}} \sim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}(p)}$ is well-known. However, we need to know it in the following form, which is a priori stronger. The proof is similar, though, but requires the very non-trivial Lemma 2.1.

3.1. **Proposition.** *For all bi-parameter weights $w \in A_\infty$ we have*

$$\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}} \sim \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(p)}, \qquad 0 < p < \infty.$$

Proof. By Hölder's inequality we only need to prove $\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(q)} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(p)}$ for $p < q$ and $q > 2$. We may assume $a_R \neq 0$ for only finitely many $R \in \mathcal{D}$. We fix Ω and, for a large enough $N > 0$, denote $E := \{S_{A,\Omega} > N\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(p)}\}$. We now have

$$w(E) \leq (N\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(p)})^{-p}\|S_{A,\Omega}\|_{L^p(w)}^p \leq N^{-p}w(\Omega).$$

Split $\mathcal{D} = \mathcal{R}_1 \cup \mathcal{R}_2$, where

$$\mathcal{R}_1 := \{R \colon w(E \cap R) > w(R)/2\}, \mathcal{R}_2 := \{R \colon w(E \cap R) \leq w(R)/2\}.$$

Notice that clearly for all $R \in \mathcal{R}_1$ we have

$$R \subset \{M_{\mathcal{D}}^w(1_E) > 1/2\} =: \tilde{E}.$$

Since $w \in A_\infty$, by Lemma 2.1 we have

$$w(\tilde{E}) \lesssim \|M_{\mathcal{D}}^w(1_E)\|_{L^2(w)}^2 \lesssim w(E).$$

We now fix N so that we always have $w(\tilde{E}) \leq w(\Omega)/2^q$, and then notice that

$$\left\|\left(\sum_{\substack{R \in \mathcal{R}_1 \\ R \subset \Omega}} |a_R|^2 \frac{1_R}{|R|}\right)^{\frac{1}{2}}\right\|_{L^q(w)} \leq \|S_{A,\tilde{E}}\|_{L^q(w)}$$

$$\leq \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(q)} w(\tilde{E})^{\frac{1}{q}} \leq \frac{1}{2}\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(q)} w(\Omega)^{\frac{1}{q}}.$$

This is absorbable, so we now move on to consider the sum, where $R \in \mathcal{R}_2$. As $q > 2$ we may calculate

$$\left\|\left(\sum_{\substack{R \in \mathcal{R}_2 \\ R \subset \Omega}} |a_R|^2 \frac{1_R}{|R|}\right)^{\frac{1}{2}}\right\|_{L^q(w)}^2 = \sup_{\|g\|_{L^{(q/2)'}(w)}=1} \left|\sum_{\substack{R \in \mathcal{R}_2 \\ R \subset \Omega}} |a_R|^2 \langle gw \rangle_R\right|$$

$$\leq 2 \sup_{\|g\|_{L^{(q/2)'}(w)}=1} \|1_{E^c} S_{A,\Omega}^2 M_{\mathcal{D}}^w g\|_{L^1(w)}$$

$$\leq 2 \sup_{\|g\|_{L^{(q/2)'}(w)}=1} \|1_{E^c} S_{A,\Omega}\|_{L^q(w)}^2 \|M_{\mathcal{D}}^w g\|_{L^{(q/2)'}(w)}$$

$$\lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(p)}^2 w(\Omega)^{2/q},$$

where we used Lemma 2.1 in the last step. The proof is done as we have shown that

$$\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(q)} \leq \frac{1}{2}\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(q)} + C\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(p)}.$$

$\square$

3.2. **Theorem.** *For all bi-parameter weights $w \in A_\infty$ we have*

$$\|A\|_{\mathrm{BMO}_{\mathrm{prod}}} \sim \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}.$$

Proof. Fix $w \in A_\infty$. Then there exists $s > 2$ so that $w \in A_s$. We first prove $\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}}$. Define the linear bi-parameter paraproduct

$$\Pi f = \Pi_A f = \sum_{R \in \mathcal{D}} a_R \langle f \rangle_R h_R.$$

It is well-known (see e.g. [21]) that

$$\|\Pi f\|_{L^s(w)} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}} \|f\|_{L^s(w)}.$$

Then by Lemma 2.2 we have

$$\left\| \left(\sum_{R \in \mathcal{D}} |a_R|^2 |\langle f \rangle_R|^2 \frac{1_R}{|R|} \right)^{\frac{1}{2}} \right\|_{L^s(w)} = \|S_{\mathcal{D}}(\Pi f)\|_{L^s(w)} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}} \|f\|_{L^s(w)}.$$

Testing with $f = 1_\Omega$ we get

$$\|S_{A,\Omega}\|_{L^s(w)} \leq \left\| \left(\sum_{R \in \mathcal{D}} |a_R|^2 |\langle 1_\Omega \rangle_R|^2 \frac{1_R}{|R|} \right)^{\frac{1}{2}} \right\|_{L^s(w)} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}} w(\Omega)^{\frac{1}{s}}.$$

This means that $\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}(s)} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}}$. By Proposition 3.1 we conclude that $\|A\|_{\mathrm{BMO}_{\mathrm{prod},w}} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}}$.

It remains to prove $\|A\|_{\mathrm{BMO}_{\mathrm{prod}}} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}$. For $0 \leq f \in L^s(w)$ and $0 \leq g \in L^{(s/2)'}(w)$, we have

$$\begin{aligned}
\sum_{R \in \mathcal{D}} |a_R|^2 \langle w \rangle_R \langle f \rangle_R^2 \langle g \rangle_R^w &= \int_0^\infty \sum_{\substack{R \in \mathcal{D} \\ \langle f \rangle_R^2 \langle g \rangle_R^w > t}} |a_R|^2 \langle w \rangle_R \, \mathrm{d}t \\
&\leq \int_0^\infty \sum_{\substack{R \in \mathcal{D} \\ R \subset \{(M_{\mathcal{D}}f)^2 M_{\mathcal{D}}^w g > t\}}} |a_R|^2 \langle w \rangle_R \, \mathrm{d}t \\
&\leq \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}^2 \int_0^\infty w(\{(M_{\mathcal{D}}f)^2 M_{\mathcal{D}}^w g > t\}) \, \mathrm{d}t \\
&= \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}^2 \|(M_{\mathcal{D}}f)^2 M_{\mathcal{D}}^w g\|_{L^1(w)} \\
&\leq \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}^2 \|M_{\mathcal{D}}f\|_{L^s(w)}^2 \|M_{\mathcal{D}}^w g\|_{L^{(s/2)'}(w)} \\
&\lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}^2 \|f\|_{L^s(w)}^2 \|g\|_{L^{(s/2)'}(w)},
\end{aligned}$$

where we have used Lemma 2.1 in the last step. Testing the above inequality with $f = w^{-\frac{1}{s}}1_\Omega$, $g = w^{-\frac{1}{(s/2)'}}1_\Omega$ we get

$$\sum_{\substack{R\in\mathcal{D}\\R\subset\Omega}} |a_R|^2 \langle w^{-\frac{1}{s}}\rangle_R^2 \langle w^{\frac{2}{s}}\rangle_R \lesssim \|A\|^2_{\mathrm{BMO}_{\mathrm{prod},w}}|\Omega|.$$

We conclude the proof by noticing that $1 \leq \langle w^{-\frac{1}{s}}\rangle_R^2 \langle w^{\frac{2}{s}}\rangle_R$. $\qquad\square$

3.3. *Remark.* In the one-parameter case, the equivalence between BMO and BMO_w, where $w \in A_\infty$, is due to Muckenhoupt and Wheeden [30].

Finally, we define the actual weighted product BMO by setting

$$\|A\|_{\mathrm{BMO}_{\mathrm{prod}}(w)} = \sup_\Omega \frac{1}{w(\Omega)^{1/2}}\left\|\left(\sum_{\substack{R\in\mathcal{D}\\R\subset\Omega}}|a_R|^2\frac{1_R(x)}{w(R)}\right)^{1/2}\right\|_{L^2(\,\mathrm{d}x)} = \sup_\Omega\left(\frac{1}{w(\Omega)}\sum_{\substack{R\in\mathcal{D}\\R\subset\Omega}}\frac{|a_R|^2}{\langle w\rangle_R}\right)^{\frac{1}{2}}.$$

The previous theorem is of independent interest, but also yields the following key lemma.

3.4. **Lemma.** *If $A \in \mathrm{BMO}_{\mathrm{prod}}$ define $A_w = (a_R\langle w\rangle_R)_{R\in\mathcal{D}}$ for $w \in A_\infty$. Then we have*

$$\|A_w\|_{\mathrm{BMO}_{\mathrm{prod}}(w)} \sim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}}.$$

Proof. Notice that

$$\|A_w\|_{\mathrm{BMO}_{\mathrm{prod}}(w)} = \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}} \sim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}}.$$

Here the first equality is obvious and the second estimate is Theorem 3.2. $\qquad\square$

3.5. **Corollary.** *For sequences of scalars $A = (a_R)$ and $B = (b_R)$ we have*

$$\sum_{R\in\mathcal{D}} |a_R|\langle w\rangle_R|b_R| \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}}\|S_B\|_{L^1(w)}$$

whenever $w \in A_\infty$.

Proof. Follows from the known, see e.g. [21, Proposition 4.1], weighted H^1-BMO duality

$$\text{(3.6)} \qquad \sum_{R\in\mathcal{D}} |a_R||b_R| \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}(w)}\|S_B\|_{L^1(w)}$$

and Lemma 3.4. $\qquad\square$

4. PROOF OF THEOREM 1.1

A bilinear bi-parameter full paraproduct on a grid $\mathcal{D} = \mathcal{D}^n \times \mathcal{D}^m$ has the form

$$\text{(4.1)} \qquad \Pi_A(f_1, f_2) = \Pi(f_1, f_2) = \sum_{R=I\times J\in\mathcal{D}} a_R\left\langle f_1, h_I \otimes \frac{1_J}{|J|}\right\rangle\langle f_2\rangle_R \frac{1_I}{|I|} \otimes h_J,$$

where $\|A\|_{\mathrm{BMO}_{\mathrm{prod}}} \leq 1$. What is important is that there are actually nine different types of full paraproducts – the full paraproduct above corresponds to the tuples $\left(h_I, \frac{1_I}{|I|}, \frac{1_I}{|I|}\right)$ and $\left(\frac{1_J}{|J|}, \frac{1_J}{|J|}, h_J\right)$, but the h_I can be in any of the three slots and so can the h_J.

It follows from [28] that to prove Theorem 1.1 it suffices to prove the following weighted estimate for the full paraproducts.

4.2. **Proposition.** *Let $1 < p, q < \infty$ and $1/2 < r < \infty$ satisfy $1/p + 1/q = 1/r$, $w_1 \in A_p$ and $w_2 \in A_q$ be bi-parameter weights, and set $w := w_1^{r/p} w_2^{r/q}$. Then we have*

$$\|\Pi(f_1, f_2)\|_{L^r(w)} \lesssim \|f_1\|_{L^p(w_1)} \|f_2\|_{L^q(w_2)}.$$

Proof. **Case 1.** Suppose that there is a full average over $R = I \times J$ at least in f_1 or f_2. In such cases the bilinear paraproduct estimate decouples reducing to linear estimates. For example, suppose Π has the form (4.1). Then using the weighted lower square function estimate, Lemma 2.3, and the basic Lemma 2.2 we have

$$\|\Pi(f_1, f_2)\|_{L^r(w)} \lesssim \left\| \left(\sum_J \left| \sum_I |a_R| \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \langle f_2 \rangle_{I \times J} \frac{1_I}{|I|} \right|^2 \frac{1_J}{|J|} \right)^{\frac{1}{2}} \right\|_{L^r(w)}$$

$$\lesssim \left\| M_{\mathcal{D}} f_2 \left(\sum_J \left(\sum_I |a_R| \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \left| \frac{1_I}{|I|} \right)^2 \frac{1_J}{|J|} \right)^{\frac{1}{2}} \right\|_{L^r(w)}$$

$$\leq \|M_{\mathcal{D}} f_2\|_{L^q(w_2)} \left\| \left(\sum_J \left(\sum_I |a_R| \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \left| \frac{1_I}{|I|} \right)^2 \frac{1_J}{|J|} \right)^{\frac{1}{2}} \right\|_{L^p(w_1)}$$

$$\lesssim \|f_2\|_{L^q(w_2)} \|S^2_{\mathcal{D}^m} h\|_{L^p(w_1)},$$

where

$$h = \sum_{R = I \times J} |a_R| \left| \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \right| \frac{1_I}{|I|} \otimes h_J.$$

This is just a standard linear bi-parameter paraproduct, and thus satisfies the weighted estimate $\|h\|_{L^p(w_1)} \lesssim \|f_1\|_{L^p(w_1)}$ (see e.g. [21]). Thus, we are done by Lemma 2.2.

Case 2. Out of the remaining cases we choose the symmetry

$$\Pi(f_1, f_2) = \sum_{R = I \times J} a_R \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \left\langle f_2, \frac{1_I}{|I|} \otimes h_J \right\rangle \frac{1_R}{|R|}.$$

Equipped with our current tools we can prove the desired estimate directly for any $p_0, q_0 \in (1, \infty)$ and $r_0 \in [1, \infty)$ satisfying $1/r_0 = 1/p_0 + 1/q_0$. By bilinear extrapolation [8, 10] it is enough to prove the estimate with only one fixed tuple, so this is certainly enough to get the claimed full range. For example, in the case $r_0 = 1$ we get

$$\|\Pi(f_1, f_2)\|_{L^1(w)} \leq \sum_{R = I \times J} |a_R| \langle w \rangle_R \left| \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \right| \left| \left\langle f_2, \frac{1_I}{|I|} \otimes h_J \right\rangle \right|$$

$$\lesssim \left\| \left(\sum_{R = I \times J} \left| \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \right|^2 \left| \left\langle f_2, \frac{1_I}{|I|} \otimes h_J \right\rangle \right|^2 \frac{1_R}{|R|} \right)^{1/2} \right\|_{L^1(w)}$$

$$\leq \|S^1_{\mathcal{D}, M} f_1\|_{L^{p_0}(w_1)} \|S^2_{\mathcal{D}, M} f_2\|_{L^{q_0}(w_2)} \lesssim \|f_1\|_{L^{p_0}(w_1)} \|f_2\|_{L^{q_0}(w_2)},$$

where we have used Corollary 3.5 in the second estimate and Lemma 2.2 in the last step. $\square$

4.3. *Remark.* The advantage of the case $r_0 = 1$ in Case 2 above is that then $w \in A_2$, so that proving the required estimate $\|A\|_{\mathrm{BMO}_{\mathrm{prod}, w}} \lesssim \|A\|_{\mathrm{BMO}_{\mathrm{prod}}}$ does not require the John-Nirenberg inequality and thus not even Lemma 2.1. However, we still note that the case

$r_0 > 1$ could be done with a similar calculation, but it requires bounding $|\langle \Pi(f_1, f_2), f_3 w \rangle|$ for $f_3 \in L^{r'_0}(w)$ with

$$\sum_{R=I \times J} |a_R| \langle w \rangle_R \left| \left\langle f_1, h_I \otimes \frac{1_J}{|J|} \right\rangle \right| \left| \left\langle f_2, \frac{1_I}{|I|} \otimes h_J \right\rangle \right| \langle |f_3| \rangle_R^w,$$

using the fuller strength of Corollary 3.5 and also Lemma 2.1.

5. PROOF OF THEOREM 1.2

We now move on to Theorem 1.2 – this directly forces us to deal with m-parameters, $m \geq 3$. Indeed, recall that [1, Theorem 1.5] has no restrictions as long as the appearing singular integral operators are at most bi-parameter. Notice that it is obvious how m-parameter weights in the product space $\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_m}$ are defined – simply work with rectangles $R = I_1 \times \cdots \times I_m$. Notice also that if we set $\mathcal{D} = \prod_{i=1}^{m} \mathcal{D}^i$, where $\mathcal{D}^i$ is a dyadic grid on $\mathbb{R}^{d_i}$, and if we have a sequence $A = (a_R)_{R \in \mathcal{D}}$, the various product BMO norms of A have a completely analogous definition in m-parameters.

We begin by noticing that a multi-parameter version of Theorem 3.2 is true – the proof is exactly the same.

5.1. Theorem. *For all m-parameter weights $w \in A_\infty$ on $\mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_m}$ we have*

$$\|A\|_{\mathrm{BMO}_{\mathrm{prod}}} \sim \|A\|_{\mathrm{BMO}_{\mathrm{prod},w}}.$$

To communicate the main idea regarding Theorem 1.2, including the definition of the space $\mathrm{bmo}^{\mathcal{I}}(\nu)$ appearing in the theorem, we now restrict the discussion to the tri-parameter space $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}$. Let $\mathcal{D}^i$ be a dyadic grid in $\mathbb{R}^{d_i}$ and set $\mathcal{D} = \mathcal{D}^1 \times \mathcal{D}^2 \times \mathcal{D}^3$. We denote cubes in $\mathcal{D}^i$ by I_i, J_i, K_i, etc. It is obvious how to define, just like in the bi-parameter setting, the square functions $S^1_{\mathcal{D}^1}$, $S^{1,2}_{\mathcal{D}^1 \times \mathcal{D}^2}$, $S_{\mathcal{D}}$, and so on. We now consider $b \colon \mathbb{R}^d \to \mathbb{C}$ and a tri-parameter Bloom weight $\nu = \mu^{1/p} \lambda^{-1/p}$, where $p \in (1, \infty)$ and $\mu, \lambda \in A_p(\mathbb{R}^d)$ are tri-parameter A_p weights. In Section 3 we had a sequence $A = (a_R)$, but if we would have been talking about a function b there, we would have simply considered the sequence $(\langle b, h_R \rangle)$. For the commuting function b we prefer to understand various BMO spaces directly via the dualised forms (3.6). This is because it makes various relationships between the BMO spaces much more transparent. Here we follow [1, Section 2], where the facts used in the following explanation are also proved.

Theorem 1.2 is not dyadic – thus, we need the following inequalities to hold uniformly for all dyadic grids $\mathcal{D}^i$. If for $i \in \{1, 2, 3\}$ we have

$$|\langle b, f \rangle| \leq C \|S^i_{\mathcal{D}^i} f\|_{L^1(\nu)},$$

we denote the optimal constant C by $\|b\|_{\mathrm{BMO}^i(\nu)}$. Moreover, e.g. the norm $\|b\|_{\mathrm{BMO}^{1,2}_{\mathrm{prod}}(\nu)} = \|b\|_{\mathrm{BMO}^{1,2}(\nu)}$ has the obvious dual definition as the best constant in the inequality

$$|\langle b, f \rangle| \leq \|b\|_{\mathrm{BMO}^{1,2}_{\mathrm{prod}}(\nu)} \|S^{1,2}_{\mathcal{D}^1 \times \mathcal{D}^2} f\|_{L^1(\nu)},$$

or could be defined as in Section 3. Similarly, it is obvious how to e.g. define $\|b\|_{\mathrm{BMO}^{1,3}_{\mathrm{prod}}(\nu)} = \|b\|_{\mathrm{BMO}^{1,3}(\nu)}$ and $\|b\|_{\mathrm{BMO}^{1,2,3}_{\mathrm{prod}}(\nu)} = \|b\|_{\mathrm{BMO}^{1,2,3}(\nu)}$.

We can always estimate up by adding more parameters to square functions so that e.g.

$$\|S_{\mathcal{D}^i}^i f\|_{L^1(\nu)} \lesssim \|S_{\mathcal{D}^i \times \mathcal{D}^j}^{i,j} f\|_{L^1(\nu)}.$$

This explains the convenience of using the dual formulations here directly, as from such estimates we can immediately see that e.g. $\mathrm{BMO}^1(\nu) \subset \mathrm{BMO}^{1,2}_{\mathrm{prod}}(\nu)$. Another thing is that, for instance, the $\mathrm{BMO}^1(\nu)$ condition implies that

(5.2) $$|\langle b(\cdot, x_2, x_3), g\rangle| \leq \|b\|_{\mathrm{BMO}^1(\nu)} \|S_{\mathcal{D}^1} g\|_{L^1(\nu(\cdot, x_2, x_3))}$$

uniformly on (almost every) x_2, x_3, which is sometimes useful.

Given a partition $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_k\}$, $k \leq 3$, of the parameter space $\{1, 2, 3\}$, we define

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

where $\bar{u} = (u_i)_{i=1}^k$ is such that $u_i \in \mathcal{I}_i$. We illustrate this with examples. If $k = 3$ we look at $[T_1, [T_2, [T_3, b]]]$, where each T_i is a one-parameter CZO on $\mathbb{R}^{d_i}$, and the related condition is $\|b\|_{\mathrm{bmo}^{\{\{1\}, \{2\}, \{3\}\}}(\nu)} = \|b\|_{\mathrm{BMO}^{1,2,3}(\nu)} = \|b\|_{\mathrm{BMO}^{1,2,3}_{\mathrm{prod}}(\nu)}$. If $k = 2$ we may e.g have $\mathcal{I} = \{\{1, 2\}, \{3\}\}$ and look at $[T_1, [T_2, b]]$, where T_1 is a bi-parameter CZO on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and T_2 is a one-parameter CZO on $\mathbb{R}^{d_3}$, and the related condition is $\|b\|_{\mathrm{bmo}^{\{\{1,2\}, \{3\}\}}(\nu)} = \sup\{\|b\|_{\mathrm{BMO}^{1,3}(\nu)}, \|b\|_{\mathrm{BMO}^{2,3}(\nu)}\}$. Both the cases $k = 3$ and $k = 2$ are already covered by [1, Theorem 1.5], since all the CZOs are at most bi-parameter. The remaining case $k = 1$ deals with the commutator $[b, T]$, where T is a tri-parameter CZO on $\mathbb{R}^d$, and the related condition is $\|b\|_{\mathrm{bmo}^{\{1,2,3\}}(\nu)} = \sup_i \|b\|_{\mathrm{BMO}^i(\nu)}$.

The top-level strategy is always to first use the dyadic representation theorems [17, 18, 29, 32] and reduce to commutators of dyadic model operators in some fixed dyadic grids $\mathcal{D}^i$. The difficulty levels of the different cases $k \in \{1, 2, 3\}$ are not strictly comparable. If $k = 3$ we have the triple commutator $[T_1, [T_2, [T_3, b]]]$, where the commutator structure – and the related decomposition of the commutator – is most complicated, but the operators themselves are one-parameter and the model operator decomposition [18] of each of them is the simplest. The remaining difficulty in [1, Theorem 1.5] was **not** that it was not known how to efficiently decompose arbitrary commutators $[T_1, [T_2, \ldots [b, T_k]]]$ or to use the complicated BMO spaces $\mathrm{bmo}^{\mathcal{I}}(\nu)$. The problem was that if one T_i is of k_i-parameters, $k_i \geq 3$, then the model operator decomposition [32] produces a very complicated model operator – a multi-parameter partial paraproduct with a k-parameter, $k \geq 2$, paraproduct component.

This is the problem we address in Theorem 1.2. In the proof we deal with the $k = 1$ case in the tri-parameter situation. Moreover, we do not go through all the cases $[b, U]$, where U is a tri-parameter model operator [32]. Rather, we only focus on the new problem that $U = P$ for a tri-parameter partial paraproduct with a bi-parameter paraproduct component. It is then possible to take a completely general commutator $[T_1, [T_2, \ldots [b, T_k]]]$, follow the decomposition strategy of [1] and handle the main new difficulty similarly as we do here. See also Section 1.1 to understand the role of Lemma 5.5 and, thus, how everything depends on Theorem 3.2. One more point: the main Lemma 5.5 is applied to the **unweighted** product BMO coefficients appearing in the dyadic model operators, and it is not related to the commuting function b. If we would e.g. deal with a 4-parameter partial paraproduct with a tri-parameter paraproduct component, we would need the full generality of Theorem 5.1 stated above.

Due to the above discussion, we will now only control $[b, P]$, where P is a tri-parameter partial paraproduct with a bi-parameter paraproduct component. We will now fix a 'little BMO' function $b \in \bigcap_{i=1}^3 \mathrm{BMO}^i(\nu)$ with the normalisation

$$\sup_i \|b\|_{\mathrm{BMO}^i(\nu)} \leq 1.$$

Further, we will fix a tri-parameter partial paraproduct

$$Pf = \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} a_{(K_j), I_1, J_1} \Big\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \Big\rangle h_{J_1} \otimes \frac{1_{K_2}}{|K_2|} \otimes h_{K_3},$$

where for all K_1, I_1, J_1 like above we have

$$\big\| (a_{(K_j), I_1, J_1})_{K_2, K_3} \big\|_{\mathrm{BMO}_{\mathrm{prod}}^{2,3}} \leq \frac{|I_1|^{1/2} |J_1|^{1/2}}{|K_1|}.$$

We will show the tri-parameter Bloom estimate

(5.3) $$\|[b, P]f\|_{L^p(\lambda)} \lesssim \|f\|_{L^p(\mu)},$$

where the implicit constant depends on the norms $[\mu]_{A_p}$ and $[\lambda]_{A_p}$, and we have polynomial dependency on the complexity, but this is not emphasised.

To show this, we will need some particular paraproducts. For $i \in \{1, 2, 3\}$ we define

$$A_1^i(b, f) = \sum_{I_i \in \mathcal{D}^i} \Delta_{I_i}^i b \Delta_{I_i}^i f, \quad A_2^i(b, f) = \sum_{I_i \in \mathcal{D}^i} \Delta_{I_i}^i b E_{I_i}^i f \quad \text{and} \quad A_3^i(b, f) = \sum_{I_i \in \mathcal{D}^i} E_{I_i}^i b \Delta_{I_i}^i f.$$

In one-parameter $\Delta_I g = \langle g, h_I \rangle h_I$ and $E_I g = 1_I \langle g \rangle_I$ are the usual martingale difference and averaging operators, and then e.g. $\Delta_{I_1}^1 f(x) = (\Delta_{I_1} f(\cdot, x_2, x_3))(x_1)$. For $i_1, i_2 \in \{1, 2, 3\}$ and $j_1, j_2 \in \{1, 2, 3\}$ define formally

$$A_{j_1, j_2}^{i_1, i_2}(b, f) = A_{j_1}^{i_1} A_{j_2}^{i_2}(b, f)$$

so that e.g.

$$A_{1,2}^{1,3}(b, f) = \sum_{I_3 \in \mathcal{D}^3} A_1^1(\Delta_{I_3}^3 b, E_{I_3}^3 f) = \sum_{\substack{I_1 \in \mathcal{D}^1 \\ I_3 \in \mathcal{D}^3}} \Delta_{I_1}^1 \Delta_{I_3}^3 b \Delta_{I_1}^1 E_{I_3}^3 f.$$

What we need now is that according to [1, Lemma 3.1] we have

$$\|A_{j_1, j_2}^{i_1, i_2}(b, f)\|_{L^p(\lambda)} \lesssim \|f\|_{L^p(\mu)}$$

as long as $(j_1, j_2) \neq (3, 3)$. Indeed, for $j_1 \neq 3$ and $j_2 \neq 3$ we have

$$\|A_{j_1, j_2}^{i_1, i_2}(b, f)\|_{L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}_{\mathrm{prod}}^{i_1, i_2}(\nu)} \|f\|_{L^p(\mu)} \lesssim \|f\|_{L^p(\mu)},$$

and if e.g. $j_1 = 3$ and $j_2 \neq 3$ we have

$$\|A_{3, j_2}^{i_1, i_2}(b, f)\|_{L^p(\lambda)} \lesssim \|b\|_{\mathrm{BMO}^{i_2}(\nu)} \|f\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)}.$$

We now decompose the commutator $[b, P]$ using the paraproducts. In $[b, P]f = bPf - P(bf)$ we decompose

$$bPf = \sum_{j_1, j_2 = 1}^3 A_{j_1, j_2}^{1,3}(b, Pf) \qquad \text{and} \qquad bf = \sum_{j_1, j_2 = 1}^3 A_{j_1, j_2}^{1,2}(b, f).$$

Because of the above paraproduct estimates and the weighted boundedness of P (see e.g. [21, Proposition 7.6]), to control $\|[b, P]f\|_{L^p(\lambda)}$ we only need to control the $L^p(\lambda)$ norm of

$$A_{3,3}^{1,3}(b, Pf) - P(A_{3,3}^{1,2}(b, f))$$

$$= \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} a_{(K_j), I_1, J_1} \Big\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \Big\rangle h_{J_1} \otimes \langle b \rangle_{J_1 \times K_3}^{1,3} \frac{1_{K_2}}{|K_2|} \otimes h_{K_3}$$

$$- \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} a_{(K_j), I_1, J_1} \Big\langle \langle b \rangle_{I_1 \times K_2}^{1,2} \langle f, h_{I_1} \otimes h_{K_2} \rangle_{1,2} \Big\rangle_{K_3} h_{J_1} \otimes \frac{1_{K_2}}{|K_2|} \otimes h_{K_3}.$$

In the first term we write $\langle b \rangle_{J_1 \times K_3}^{1,3} = [\langle b \rangle_{J_1 \times K_3}^{1,3} - \langle b \rangle_{J_1 \times K_2 \times K_3}] + \langle b \rangle_{J_1 \times K_2 \times K_3}$ and in the second term we write $\langle b \rangle_{I_1 \times K_2}^{1,2} = [\langle b \rangle_{I_1 \times K_2}^{1,2} - \langle b \rangle_{I_1 \times K_2 \times K_3}] + \langle b \rangle_{I_1 \times K_2 \times K_3}$. We then combine the last two terms and further add and subtract $\langle b \rangle_{K_1 \times K_2 \times K_3}$. We begin by dealing with one of the resulting terms

$$E_1 := \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} \gamma_{(K_j), I_1, J_1} \Big\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \Big\rangle h_{J_1} \otimes \frac{1_{K_2}}{|K_2|} \otimes h_{K_3},$$

where $\gamma_{(K_j), I_1, J_1} = a_{(K_j), I_1, J_1} [\langle b \rangle_{I_1 \times K_2 \times K_3} - \langle b \rangle_{K_1 \times K_2 \times K_3}]$. We write

$$\langle b \rangle_{I_1 \times K_2 \times K_3} - \langle b \rangle_{K_1 \times K_2 \times K_3} = \sum_{l=1}^{i_1} \iint_{\mathbb{R}^{d_2} \times \mathbb{R}^{d_3}} \langle b, h_{I_1^{(l)}} \rangle_1 \langle h_{I_1^{(l)}} \rangle_{I_1} \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|}.$$

Allowing a polynomial dependency on the complexity, we can fix l and study the dualised form

$$\iint_{\mathbb{R}^{d_2} \times \mathbb{R}^{d_3}} \sum_{\substack{K_1 \\ L_1^{(i_1-l)} = K_1}} \sum_{\substack{I_1^{(l)} = L_1 \\ J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} |a_{(K_j), I_1, J_1}| |L_1|^{-1/2} |\langle b, h_{L_1} \rangle_1| \Big| \Big\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \Big\rangle \Big|$$

$$\Big| \Big\langle g, h_{J_1} \otimes \frac{1_{K_2}}{|K_2|} \otimes h_{K_3} \Big\rangle \Big| \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|}.$$

Using (5.2) we reduce to

$$\text{(5.4)} \quad \int_{\mathbb{R}^d} \Bigg(\sum_{\substack{K_1 \\ L_1^{(i_1-l)} = K_1}} \frac{1_{L_1}}{|L_1|^2} \otimes \Bigg[\sum_{\substack{I_1^{(l)} = L_1 \\ J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} |a_{(K_j), I_1, J_1}| \Big| \Big\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \Big\rangle \Big|$$

$$\Big| \Big\langle g, h_{J_1} \otimes \frac{1_{K_2}}{|K_2|} \otimes h_{K_3} \Big\rangle \Big| \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \Bigg]^2 \Bigg)^{1/2} \nu.$$

A certain vector-valued inequality will now be derived using A_∞ extrapolation, and the base case estimate for the extrapolation will rely on Corollary 3.5.

5.5. Lemma. *Suppose $0 < p, q < \infty$ and $w \in A_\infty(\mathbb{R}^{d_2} \times \mathbb{R}^{d_3})$ is a bi-parameter A_∞ weight. Suppose for all $k, j \in \mathbb{N}$ we are given a sequence $(a_{K_2,K_3}^{k,j})_{K_2,K_3}$ satisfying*

$$\big\|(a_{K_2,K_3}^{k,j})_{K_2,K_3}\big\|_{\mathrm{BMO}_{\mathrm{prod}}^{2,3}} \leq 1.$$

Then for all locally integrable $f_{j,k}, g_{j,k} \colon \mathbb{R}^{d_2} \times \mathbb{R}^{d_3} \to \mathbb{C}$ we have

$$\iint \left(\sum_j \left[\sum_k \sum_{K_2,K_3} |a_{K_2,K_3}^{k,j}| \Big| \Big\langle f_{j,k}, h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \Big\rangle \Big| \Big| \Big\langle g_{j,k}, \frac{1_{K_2}}{|K_2|} \otimes h_{K_3} \Big\rangle \Big| \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \right]^p \right)^{q/p} w$$

$$\lesssim \iint \left(\sum_j \left[\sum_k \left(\sum_{K_2,K_3} \langle |\langle f_{j,k}, h_{K_2} \rangle_2|\rangle_{K_3}^2 \langle |\langle g_{j,k}, h_{K_3} \rangle_3|\rangle_{K_2}^2 \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \right)^{1/2} \right]^p \right)^{q/p} w.$$

Proof. By repeated use of A_∞ extrapolation [6, Theorem 2.1], it is enough to prove that

$$\iint \left(\sum_{K_2,K_3} |a_{K_2,K_3}| \Big| \Big\langle f, h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \Big\rangle \Big| \Big| \Big\langle g, \frac{1_{K_2}}{|K_2|} \otimes h_{K_3} \Big\rangle \Big| \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \right) w$$

$$\lesssim \iint \left(\sum_{K_2,K_3} \langle |\langle f, h_{K_2} \rangle_2|\rangle_{K_3}^2 \langle |\langle g, h_{K_3} \rangle_3|\rangle_{K_2}^2 \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \right)^{1/2} w.$$

We note that the used extrapolation results are stated for the so-called Muckenhoupt bases and directly cover multi-parameter situations. The remaining estimate now follows from Corollary 3.5 almost immediately. $\qquad\square$

Applying the $p = 2$ and $q = 1$ case of the previous lemma with a fixed x_1 and $\nu(x_1, \cdot) \in A_\infty(\mathbb{R}^{d_2} \times \mathbb{R}^{d_3})$ to (5.4) we reduce to

$$\int_{\mathbb{R}^d} \left(\sum_{\substack{K_1 \\ L_1^{(i_1-l)}=K_1}} \frac{1_{L_1}}{|L_1|^2 |K_1|^2} \otimes \left[\sum_{\substack{I_1^{(l)}=L_1 \\ J_1^{(j_1)}=K_1}} |I_1|^{1/2} |J_1|^{1/2} \right. \right.$$

$$\left. \left. \left(\sum_{K_2,K_3} \langle |\langle f, h_{I_1} \otimes h_{K_2} \rangle_{1,2}|\rangle_{K_3}^2 \langle |\langle g, h_{J_1} \otimes h_{K_3} \rangle_{1,3}|\rangle_{K_2}^2 \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \right)^{1/2} \right]^2 \right)^{1/2} \nu.$$

We now estimate

$$\left(\sum_{K_2,K_3} \langle |\langle f, h_{I_1} \otimes h_{K_2} \rangle_{1,2}|\rangle_{K_3}^2 \langle |\langle g, h_{J_1} \otimes h_{K_3} \rangle_{1,3}|\rangle_{K_2}^2 \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \right)^{1/2}$$

$$\leq |I_1|^{1/2} |J_1|^{1/2} \left(\sum_{K_2,K_3} \langle |\langle f, h_{K_2} \rangle_2|\rangle_{I_1 \times K_3}^2 \langle |\langle \Delta_{K_1,j_1}^1 g, h_{K_3} \rangle_3|\rangle_{J_1 \times K_2}^2 \frac{1_{K_2}}{|K_2|} \frac{1_{K_3}}{|K_3|} \right)^{1/2}.$$

Here $\Delta_{K_1,j_1}^1 g = \sum_{J_1^{(j_1)}=K_1} \Delta_{J_1}^1 g$ is a martingale block, and we essentially threw away the cancellation in the h_{I_1}, which we can do as the nature of our argument produces one

extra cancellation. Using this we reduce to estimating

$$\int_{\mathbb{R}^d} M_{\mathcal{D}^1}^1 \left(\sum_{K_2} \frac{1_{K_2}}{|K_2|} \otimes \left[M_{\mathcal{D}^1 \times \mathcal{D}^3} \langle f, h_{K_2} \rangle_2 \right]^2 \right)^{1/2}$$

$$\left(\sum_{K_1} \left[M_{\mathcal{D}^1}^1 \left(\sum_{K_3} \left[M_{\mathcal{D}^1 \times \mathcal{D}^2} \langle \Delta_{K_1, j_1}^1 g, h_{K_3} \rangle_3 \right]^2 \otimes \frac{1_{K_3}}{|K_3|} \right)^{1/2} \right]^2 \right)^{1/2} \nu.$$

Using Hölder's inequality and the standard estimates of Lemma 2.2, together with some vector-valued improvements of Lemma 2.2, which can be obtained by extrapolating the estimates of the same lemma, we derive the desired upper bound $\|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\lambda^{1-p'})}$. We have shown that

$$|E_1| \lesssim \|f\|_{L^p(\mu)}.$$

The error term with $\langle b \rangle_{J_1 \times K_2 \times K_3} - \langle b \rangle_{K_1 \times K_2 \times K_3}$ is handled similarly.

To complete the proof, we now deal with the error term

$$E_2 = \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} a_{(K_j), I_1, J_1} \left\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \right\rangle$$

$$h_{J_1} \otimes \left[\langle b \rangle_{J_1 \times K_3}^{1,3} - \langle b \rangle_{J_1 \times K_2 \times K_3} \right] \frac{1_{K_2}}{|K_2|} \otimes h_{K_3},$$

which is in some sense simpler than the error term E_1 and does not require the new techniques of this paper. This is essentially because in the following expansion of the function b we do not get a sum over the cubes in the shift parameter as above, but rather over $J_2 \in \mathcal{D}^2$. So we expand

$$\left(\langle b \rangle_{J_1 \times K_3}^{1,3} - \langle b \rangle_{J_1 \times K_2 \times K_3} \right) 1_{K_2} = \sum_{J_2 \subset K_2} \left\langle b, \frac{1_{J_1}}{|J_1|} \otimes h_{J_2} \otimes \frac{1_{K_3}}{|K_3|} \right\rangle h_{J_2}$$

and dualise, and use this to reduce to

$$\iint_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_3}} \sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} \sum_{J_2 \subset K_2} |a_{(K_j), I_1, J_1}| \frac{|\langle b, h_{J_2} \rangle_2|}{|K_2|}$$

$$\left| \left\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \right\rangle \right| \left| \langle g, h_{J_1} \otimes h_{J_2} \otimes h_{K_3} \rangle \right| \frac{1_{J_1}}{|J_1|} \frac{1_{K_3}}{|K_3|}.$$

Using an estimate like (5.2) we get

$$\sum_{\substack{K_1 \\ I_1^{(i_1)} = J_1^{(j_1)} = K_1}} \sum_{K_2, K_3} |a_{(K_j), I_1, J_1}| \left| \left\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|} \right\rangle \right|$$

(5.6)

$$\left\langle \left(S_{\mathcal{D}^2} \langle g, h_{J_1} \otimes h_{K_3} \rangle_{1,3} \right) \langle \nu \rangle_{J_1 \times K_3}^{1,3} \right\rangle_{K_2}.$$

We now define the auxiliary operator

$$Ug = \sum_{V_1, V_3} h_{V_1} \otimes \left(S_{\mathcal{D}^2} \langle g, h_{V_1} \otimes h_{V_3} \rangle_{1,3} \right) \langle \nu \rangle_{V_1 \times V_3}^{1,3} \otimes h_{V_3},$$

and notice that the term in (5.6) equals

$$\sum_{\substack{K_1 \\ I_1^{(i_1)}=J_1^{(j_1)}=K_1}} \sum_{K_2,K_3} |a_{(K_j),I_1,J_1}| \left|\left\langle f, h_{I_1} \otimes h_{K_2} \otimes \frac{1_{K_3}}{|K_3|}\right\rangle\right| \left|\left\langle Ug, h_{J_1} \otimes \frac{1_{K_2}}{|K_2|} \otimes h_{K_3}\right\rangle\right|.$$

That is, using the original partial paraproduct P, we can see this as the pairing $\langle Pf, Ug\rangle$ with absolute values in (which is of no significance). Thus, using the known weighted boundedness of P, we can simply dominate things by $\|f\|_{L^p(\mu)}\|Ug\|_{L^{p'}(\mu^{1-p'})}$.

It remains to show that $\|Ug\|_{L^{p'}(\mu^{1-p'})} \lesssim \|g\|_{L^{p'}(\lambda^{1-p'})}$. We will show this now. Using again variants of the standard estimates of Lemma 2.2 we get

$$\|Ug\|_{L^{p'}(\mu^{1-p'})} \sim \left\|\left(\sum_{V_1,V_3} \frac{1_{V_1}}{|V_1|} \otimes \left[(S_{\mathcal{D}^2}\langle g, h_{V_1} \otimes h_{V_3}\rangle_{1,3})\langle\nu\rangle_{V_1\times V_3}^{1,3}\right]^2 \otimes \frac{1_{V_3}}{|V_3|}\right)^{1/2}\right\|_{L^{p'}(\mu^{1-p'})}$$

$$\lesssim \left\|\left(\sum_{V_1,V_3} \frac{1_{V_1}}{|V_1|} \otimes \left(S_{\mathcal{D}^2}\langle g, h_{V_1} \otimes h_{V_3}\rangle_{1,3}\right)^2 \otimes \frac{1_{V_3}}{|V_3|}\right)^{1/2}\nu\right\|_{L^{p'}(\mu^{1-p'})}$$

$$= \left\|\left(\sum_{V_1,V_3} \frac{1_{V_1}}{|V_1|} \otimes \left(S_{\mathcal{D}^2}\langle g, h_{V_1} \otimes h_{V_3}\rangle_{1,3}\right)^2 \otimes \frac{1_{V_3}}{|V_3|}\right)^{1/2}\right\|_{L^{p'}(\lambda^{1-p'})}$$

$$\lesssim \|g\|_{L^{p'}(\lambda^{1-p'})}.$$

We have shown that $|E_2| \lesssim \|f\|_{L^p(\mu)}$. The error term with $\langle b\rangle_{I_1\times K_2}^{1,2} - \langle b\rangle_{I_1\times K_2\times K_3}$ is handled similarly. This ends our proof of the tri-parameter Bloom estimate (5.3).

REFERENCES

[1] E. Airta, Two-weight commutator estimates: general multi-parameter framework, Publ. Mat., to appear, arXiv:1906.10983, 2019.

[2] 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.

[3] 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.

[4] R. Coifman, Y. Meyer, Au delà des opérateurs pseudo-différentiels, Astérisque 57 (1978) 1–185.

[5] R. Coifman, R. Rochberg, G. Weiss, Factorization theorems for Hardy spaces in several variables, Ann. of Math. (2) 103 (1976) 611–635.

[6] D. Cruz-Uribe, J.M. Martell, C. Pérez, Extrapolation from A_∞ weights and applications, J. Funct. Anal. 213 (2004) 412–439.

[7] F. Di Plinio, Y. Ou, Banach-valued multilinear singular integrals, Indiana Univ. Math. J. 67 (2018) 1711–1763.

[8] J. Duoandikoetxea, Extrapolation of weights revisited: New proofs and sharp bounds, J. Funct. Anal. 260 (2011) 1886–1901.

[9] J. Duoandikoetxea, F. Martín-Reyes, S. Ombrosi, On the A_∞ conditions for general bases. Math. Z. 282 (2016) 955–972.

[10] L. Grafakos, J.M. Martell, Extrapolation of weighted norm inequalities for multivariable operators and applications, J. Geom. Anal. 14 (2004) 19–46.

[11] L. Grafakos, S. Oh, The Kato-Ponce inequality, Comm. Partial Differential Equations 39 (2014) 1128–1157.

[12] L. Grafakos, R. Torres, Multilinear Calderón–Zygmund theory, Adv. Math. 165 (2002) 124–164.

[13] R. Fefferman, Harmonic analysis on product spaces, Ann. of Math. 126 (1987) 109–130.

[14] R. Fefferman, A^p weights and singular integrals, Amer. J. Math. 110 (1988) 975–987.

[15] R. Fefferman, Strong differentiation with respect to measures. Amer. J. Math. 103 (1981) 33–40.

[16] R. Fefferman, E. Stein, Singular integrals on product spaces, Adv. Math. 45 (1982) 117–143.

[17] T. Hytönen, The sharp weighted bound for general Calderón-Zygmund operators, Ann. of Math. 175 (2012) 1473–1506.

[18] T. Hytönen, Representation of singular integrals by dyadic operators, and the A_2 theorem, Expo. Math. 35 (2017)166–205.

[19] T. Hytönen, The L^p-to-L^q boundedness of commutators with applications to the Jacobian operator, preprint, arXiv:1804.11167, 2018.

[20] I. Holmes, M. Lacey, B. Wick, Commutators in the two-weight setting, Math. Ann. 367 (2017) 51–80.

[21] I. Holmes, S. Petermichl, B. Wick, Weighted little bmo and two-weight inequalities for Journé commutators. Anal. PDE 11 (2018) 1693–1740.

[22] J.-L. Journé, Calderón-Zygmund operators on product spaces, Rev. Mat. Iberoam. 1 (1985) 55–91.

[23] T. Kato, G. Ponce, Commutator estimates and the Euler and Navier-Stokes equations, Comm. Pure Appl. Math. 41 (1988) 891–907.

[24] K. Li, J. M. Martell, S. Ombrosi, Extrapolation for multilinear Muckenhoupt classes and applications to the bilinear Hilbert transform, preprint, arXiv:1802.03338, 2018.

[25] 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.

[26] 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`.

[27] 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.

[28] K. Li, H. Martikainen, E. Vuorinen, Bilinear Calderón-Zygmund theory on product spaces, J. Math. Pures Appl. (2019), `https://doi.org/10.1016/j.matpur.2019.10.007`.

[29] H. Martikainen, Representation of bi-parameter singular integrals by dyadic operators, Adv. Math. 229 (2012) 1734–1761.

[30] B. Muckenhoupt and R. L. Wheeden, Weighted bounded mean oscillation and the Hilbert transform, Studia Math. 54 (1975/76) 221–237.

[31] B. Nieraeth, Quantitative estimates and extrapolation for multilinear weight classes, Math. Ann. 375 (2019) 453–507.

[32] Y. Ou, Multi-parameter singular integral operators and representation theorem, Rev. Mat. Iberoam. 33 (2017) 325–350.

(E.A.) DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF HELSINKI, P.O.B. 68, FI-00014 UNIVERSITY OF HELSINKI, FINLAND
E-mail address: `emil.airta@helsinki.fi`

(K.L.) CENTER FOR APPLIED MATHEMATICS, TIANJIN UNIVERSITY, WEIJIN ROAD 92, 300072 TIANJIN, CHINA
E-mail address: `kli@tju.edu.cn`

(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`