Commutators of multi-parameter flag singular integrals
and applications

Xuan Thinh Duong, Ji Li, Yumeng Ou, Jill Pipher and Brett D. Wick

Abstract: We introduce the iterated commutator for the Riesz transforms in the multi-parameter flag setting, and prove the upper bound of this commutator with respect to the symbol $b$ in the flag BMO space. Our methods require the techniques of semigroups, harmonic functions and multi-parameter flag Littlewood–Paley analysis. We also introduce the big commutator in this multi-parameter flag setting and prove the upper bound with symbol $b$ in the flag little-bmo space by establishing the “exponential–logarithmic” bridge between this flag little bmo space and the Muckenhoupt $A_p$ weights with flag structure. As an application, we establish the div-curl lemmas with respect to the appropriate Hardy spaces in the multi-parameter flag setting.

Keywords: multiparameter flag setting; flag commutator; Hardy and BMO space; div-curl lemma

Mathematics Subject Classification 2010: 42B30, 42B20, 42B35

1 Introduction and statement of main results

The Calderón–Zygmund theory of singular integrals has been central to the success and applicability of modern harmonic analysis in the last fifty years. This theory has had extensive applications to other fields of mathematics such as complex analysis, geometric measure theory and partial differential equations. In the setting of Euclidean spaces $\mathbb{R}^n$, a notable property of standard Calderón–Zygmund singular integrals, shared with the Hardy–Littlewood maximal operator, is that these operators commute with the classical one-parameter family of dilations on $\mathbb{R}^n$, $\delta \cdot x = (\delta x_1, \ldots, \delta x_n)$ for $\delta > 0$. See for example the monograph [39].

The product Calderón–Zygmund theory in harmonic analysis was introduced in the 70s, and studied extensively since then. The model case is a tensor product of classical singular integral operators; such operators arise in the context of questions about summation of multiple variable Fourier series. Early key work in this field includes that of Chang, C. Fefferman, R. Fefferman, Gundy, Journé, Stein [18, 13, 10, 11, 12, 2, 3, 4, 25, 38]. Included in these works are the identification of appropriate notions of product BMO and product Hardy space $H^p (\mathbb{R}^n \times \mathbb{R}^m)$.

More recently, the theory of (iterated) commutators has been developed in connection with the Chang–Fefferman BMO space, including paraproducts and multi-parameter div-curl lemmas; see, for example, [8, 15, 14, 21, 28, 29, 30]. In contrast with the classical Euclidean setting, the product Calderón–Zygmund singular integrals, and the strong maximal function operator, commute with the multi-parameter dilations on $\mathbb{R}^n$, $\delta \cdot x = (\delta_1 x_1, \ldots, \delta_n x_n)$ for $\delta = (\delta_1, \ldots, \delta_n) \in (0, \infty)^n$.

A new type of multi-parameter structure, which lies in between one-parameter and tensor product, was introduced by Muller, Ricci and Stein in [32] and [33], where they studied the $L^p$
boundedness of Marcinkiewicz multipliers $m(\mathcal{L}, iT)$ on Heisenberg group, where $\mathcal{L}$ is the sub-Laplacian and $T$ is the central invariant vector field, with $m$ being a multiplier of Marcinkiewicz-type. They showed that such Marcinkiewicz multipliers can be characterized by a convolution operator $f*K$ where $K$ is a so-called flag convolution kernel. This multi-parameter flag structure is not explicit, but only implicit in the sense that one can not formulate it in terms of an explicit dilation $\delta$ acting on $x$. Later, the notion of flag kernels (having singularities on appropriate flag varieties) and the properties of the corresponding singular integrals were then extended to the higher step case in Nagel, Ricci and Stein \cite{35} on Euclidean space and their applications on certain quadratic CR submanifolds of $\mathbb{C}^n$. Recently, Nagel, Ricci, Stein and Wainger \cite{36, 37} established the theory of singular integrals with flag kernels in a more general setting of homogeneous groups. They proved that, on a homogeneous group, singular integral operators with flag kernels are bounded on $L^p$, $1 < p < \infty$, and form an algebra. (See also \cite{16, 17} for related work.) Associated to this implicit multi-parameter flag structure, the Hardy space $H^1_F(\mathbb{R}^n \times \mathbb{R}^m)$ and BMO space $\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)$ were introduced by Han, Lu and Sawyer \cite{21, 22} through their creation of a flag type Littlewood–Paley theory. More recently, Han, Lee, and the second and fifth authors \cite{19} established a full characterization of $H^1_F(\mathbb{R}^n \times \mathbb{R}^m)$ via appropriate flag type non-tangential, radial maximal functions, Littlewood–Paley theory via Poisson integrals, the flag type Riesz transforms, as well as flag atomic decompositions.

In the multi-parameter setting, the dilation structure $\delta \cdot x = (\delta_1 x_1, \ldots, \delta_n x_n)$, for $\delta = (\delta_1, \ldots, \delta_n) \in (0, \infty)^n$, determines a geometry that is reflected by axes-parallel rectangles of arbitrary side-lengths. Indeed, the strong maximal function is defined as the supremum of averages over such rectangles, and the Chang–Fefferman product BMO space can also be characterized using such rectangles. When it comes to the flag setting, the lack of an explicit dilation structure makes its geometry much more obscure. However, from the study of properties of the flag singular integrals, such as the flag Riesz transforms that will be introduced below, one realizes that the flag geometry can be reflected by axes-parallel rectangles with certain restriction on the side-lengths. For example, the flag rectangles in $\mathbb{R}^n \times \mathbb{R}^m$ are the ones of the form $R = I \times J \subset \mathbb{R}^n \times \mathbb{R}^m$ with $\ell(I) \leq \ell(J)$. Compared to the multi-parameter setting, the restriction $\ell(I) \leq \ell(J)$ gives rise to new difficulties. For instance, a very useful trick in the study of problems in the multi-parameter setting is to take a sequence of rectangles $\{I \times J_i\}$ and let $J_i$ shrink to a point $y_0$ as $i \to \infty$. This can usually effectively reduce the problem to one-parameter. However, in the flag setting, such operation is not allowed any more. Other intrinsic difficulties of the flag setting can be better described from the analytic perspective, which will be discussed below.

A commutator of a classical Calderón–Zygmund singular integral with a BMO function is a bounded operator on $L^p$ with norm equivalent to the BMO norm of the symbol \cite{7}. Modern methods of proving the upper bound of these commutators in the multi-parameter product setting rely upon the existence of a wavelet basis for $L^2(\mathbb{R}^n)$, such as the Meyer wavelets or Haar wavelets, see for example \cite{27, 8}. It turns out that the behavior of the commutator is straightforward to analyze in terms of the wavelet basis. One method of proof shows that the commutator can be written as a linear combination of paraproducts and simple wavelet analogs of the Calderón–Zygmund operator in question. The other approach uses the wavelet basis to dominate the commutator by a composition of sparse operators. In the flag setting, we lack a suitable wavelet basis and this approach is not available. Essentially, the wavelet basis requires the construction of a suitable multi-resolution analysis, which we do not have in this flag setting. Hence, instead of the wavelet basis, we resort to using a method based on heat semi-groups and flag type Littlewood–Paley theory, exploiting the connection between the Reisz transforms and
Flag commutators

the Laplacian.

We now recall the flag Riesz transforms as studied in [19]. We use $R_j^{(1)}$ to denote the $j$-th Riesz transform on $\mathbb{R}^{n+m}$, $j = 1, 2, \ldots, n+m$, and we use $R_k^{(2)}$ to denote the $k$-th Riesz transform on $\mathbb{R}^{m}$, $k = 1, 2, \ldots, m$. Namely, we have that for $g^{(1)} \in L^2(\mathbb{R}^{n+m})$,

$$R_j^{(1)} g^{(1)}(x) = \text{p.v.} c_{n+m} \int_{\mathbb{R}^{n+m}} \frac{x_j - y_j}{|x - y|^{n+m+1}} g^{(1)}(y)dy, \quad x \in \mathbb{R}^{n+m};$$

and for $g^{(2)} \in L^2(\mathbb{R}^{m})$,

$$R_k^{(2)} g^{(2)}(z) = \text{p.v.} c_m \int_{\mathbb{R}^{m}} \frac{w_j - z_j}{|w - z|^{m+1}} g^{(2)}(w)dw, \quad z \in \mathbb{R}^{m}.$$

For $f \in L^2(\mathbb{R}^{n+m})$, we set

$$R_{j,k}(f) = R_j^{(1)} * R_k^{(2)} *_2 f,$$

that is, $R_{j,k}$ is the composition of $R_j^{(1)}$ and $R_k^{(2)}$. Note that the flag structure appears in $R_{j,k}$.

Given two functions $b, f \in L^2(\mathbb{R}^{n+m})$, we first recall the usual definition of commutator

$$[b, R_j^{(1)}](f)(x_1, x_2) := b(x_1, x_2) R_j^{(1)} f(x_1, x_2) - R_j^{(1)} (b \cdot f)(x_1, x_2). \quad (1.2)$$

The commutator can also act only on the second variable:

$$[b, R_k^{(2)}]_2(f)(x_1, x_2) := b(x_1, x_2) R_k^{(2)} *_2 f(x_1, x_2) - R_k^{(2)} *_2 (b \cdot f)(x_1, x_2). \quad (1.3)$$

Iterated commutators arise in the study of commutators of multi-parameter singular integral operators which are tensor products. In the flag setting, our iterated commutator takes the following form:

**Definition 1.1.** Given two functions $b, f \in L^2(\mathbb{R}^{n+m})$, the iterated commutator in the flag setting of $\mathbb{R}^n \times \mathbb{R}^m$ is

$$[[b, R_j^{(1)}], R_k^{(2)}]_2(f) := b(x_1, x_2) R_j^{(1)} * R_k^{(2)} *_2 f(x_1, x_2) - R_j^{(1)} *_2 (b \cdot R_k^{(2)} *_2 f)(x_1, x_2) - R_k^{(2)} *_2 (b \cdot R_j^{(1)} * f)(x_1, x_2) + R_k^{(2)} *_2 R_j^{(1)} * (b \cdot f)(x_1, x_2).$$

We point out that another possible definition via $[[b, R_k^{(2)}]_2, R_j^{(1)}](f)$ turns out to be equivalent; see Proposition 2.5 in Section 2.

We also introduce the big commutator in the flag setting as follows.

**Definition 1.2.** Given two functions $b, f \in L^2(\mathbb{R}^{n+m})$, the big commutator in the flag setting of $\mathbb{R}^n \times \mathbb{R}^m$ is

$$[b, R_{j,k}](f)(x) := b(x) R_{j,k}(f)(x) - R_{j,k}(b f)(x). \quad (1.4)$$

The main results, below, of this paper relate iterated and big commutator bounds to flag BMO spaces. As the definition of the space $\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)$ is very technical, we refer the reader to Section 2. Definition 2.4 for details.
Theorem 1.3. Suppose \( b \in \text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m) \) and \( 1 < p < \infty \). Then for every \( j = 1, \ldots, n+m, k = 1, \ldots, m, f \in L^p(\mathbb{R}^{n+m}), \)
\[
\|[[b, R^{(1)}_j], R^{(2)}_k]](f)\|_{L^p(\mathbb{R}^{n+m})} \lesssim \|b\|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|f\|_{L^p(\mathbb{R}^{n+m})}. \tag{1.5}
\]

Lacking methods related to analyticity ([14] for the Hilbert transform) or wavelets ([27, 28, 8]), we instead obtain this upper bound using the duality argument and the tools of semigroups, harmonic function extensions and techniques from multi-parameter analysis.

Next, we introduce the little flag BMO space. The flag structure has a geometry which is reflected by the axes-parallel rectangles \( R = I \times J \subset \mathbb{R}^{n+m} \) satisfying \( \ell(I) \leq \ell(J) \), the collection of which is referred to as flag rectangles, denoted by \( \mathcal{R}_F \). One can then define the little flag BMO space and the flag type Muckenhoupt weights \( A_{F,p} \) with respect to \( \mathcal{R}_F \).

Definition 1.4. A locally integrable function \( b \) is in little flag BMO space, denoted by \( \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m) \), if
\[
\|b\|_{\text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m)} := \sup_{R \in \mathcal{R}_F} \frac{1}{|R|} \int_R |b(x, y) - \langle b\rangle_R| \, dx \, dy < \infty, \tag{1.6}
\]

where \( \langle b\rangle_R = \frac{1}{|R|} \int_R b(x_1, x_2) \, dx_1 \, dx_2 \).

Theorem 1.5. Suppose \( T_F \) is a flag singular integral operator on \( \mathbb{R}^n \times \mathbb{R}^m \), \( b \in \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m) \) and \( 1 < p < \infty \). Then for \( f \in L^p(\mathbb{R}^{n+m}), \)
\[
\|[[b, T_F]](f)\|_{L^p(\mathbb{R}^{n+m})} \lesssim \|b\|_{\text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|f\|_{L^p(\mathbb{R}^{n+m})}. \tag{1.7}
\]

In the above, the flag singular integral \( T_F \) can be taken as the Riesz transform \( R_{j,k} \). The class of flag singular integral operators \( T_F \) naturally generalize the Riesz transforms \( R_{j,k} \) and are assumed to be associated to kernels having a standard flag structure. We refer the reader to Definition 1.4 in Section 4 for its precise definition. To obtain this upper bound, we study the little flag BMO space \( \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m) \) and find the connection with the John–Nirenberg BMO space on \( \mathbb{R}^{n+m} \) and on \( \mathbb{R}^m \). We also establish the bridge between functions in \( \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m) \) and weights in \( A_{F,p} \). These structures lead to the upper bound for \( \|[[b, R_{j,k}]](f)\).

As application, the commutator estimates obtained above imply certain versions of div-curl lemmas, which seem to be first of their kind in the flag setting. Roughly speaking, a div-curl lemma says that if vector fields \( E \) and \( B \) initially in \( L^2 \) have some cancellation (e.g. divergence or curl zero) then one can expect their dot product \( E \cdot B \) to belong to a better space of functions instead of just \( L^1 \) (as provided for by Cauchy-Schwarz). The cancellation conditions allow one to deduce some type of cancellation, e.g. \( \int E \cdot B = 0 \), suggesting that the function should belong to a suitable Hardy space since it is integrable and has mean zero. The algebraic structure of \( E \cdot B \) coupled with the duality between Hardy spaces and BMO spaces then points to the use of the commutator theorem to arrive at the membership of \( E \cdot B \) in the Hardy space; different commutator results suggest different div-curl lemmas that can be explored. In the classical one-parameter setting, the div-curl lemma says that given two vector fields, one with divergence zero and the other with curl zero, their dot product belongs to a Hardy space [6]. Later on, Lacey, Petermichl, and the fourth and the fifth authors proved multiple versions of div-curl lemmas in the multi-parameter setting [29], which are expected since the multi-parameter setting offers several different interpretations of the Hardy and BMO spaces. Thus, it is natural that our Theorems 1.3 and 1.5 lead to two versions of flag type div-curl lemmas.

First, consider vector fields on \( \mathbb{R}^n \times \mathbb{R}^m \) that take values in \( \mathcal{M}_{n+m,n+m} \) and are associated with the flag structure (see Section 5 for the precise definitions and details). We establish the div-curl
Flag commutators

Lemma in the flag setting with respect to the flag Hardy space below, which is a consequence of Theorem 1.3.

**Theorem 1.6.** Let $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$. Suppose that $E, B$ are vector fields on $\mathbb{R}^n \times \mathbb{R}^m$ taking the values in $\mathcal{M}_{n+m,m}$, associated with the flag structure. Moreover, suppose $E = E^{(1)} \ast_2 E^{(2)} \in L^p_{\nu}(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})$ and $B = B^{(1)} \ast_2 B^{(2)} \in L^q_{\nu}(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})$ satisfy that

$$\text{div}(x,y) E_j^{(1)}(x,y) = 0 \quad \text{and} \quad \text{curl}(x,y) B_j^{(1)}(x,y) = 0, \quad \forall k$$

and

$$\text{div}_y E_k^{(2)}(x,y) = 0 \quad \text{and} \quad \text{curl}_y B_k^{(2)}(x,y) = 0, \quad \forall x \in \mathbb{R}^n, \forall j.$$

Then $E \cdot B$ belongs to the flag Hardy space $H^1_{\nu}(\mathbb{R}^n \times \mathbb{R}^m)$ with

$$\|E \cdot B\|_{H^1_{\nu}(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})} \|B\|_{L^q(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})}. \quad (1.8)$$

We also prove another version of the div-curl lemma in the flag setting, which is with respect to the Hardy spaces on $\mathbb{R}^{n+m}$ and on $\mathbb{R}^m$, respectively. This version relies on the intermediate result in the proof of Theorem 1.3, namely, the structure of the flag little bmo space.

**Theorem 1.7.** Let $1 < p, q < \infty$ with $\frac{1}{p} + \frac{1}{q} = 1$. Suppose that $E, B$ are vector fields on $\mathbb{R}^n \times \mathbb{R}^m$ taking the values in $\mathbb{R}^{n+m}$. Moreover, suppose $E \in L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})$ and $B \in L^q(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})$ satisfy that

$$\text{div}(x,y) E(x,y) = 0 \quad \text{and} \quad \text{curl}(x,y) B(x,y) = 0$$

and

$$\text{div}_y E(x,y) = 0 \quad \text{and} \quad \text{curl}_y B(x,y) = 0, \quad \forall x \in \mathbb{R}^n.$$

Then we have

$$\|E \cdot B\|_{H^1_{\nu}(\mathbb{R}^{n+m})} \lesssim \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})} \|B\|_{L^q(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})}. \quad (1.9)$$

and

$$\int_{\mathbb{R}^m} \|E(\cdot,y) \cdot_2 B(\cdot,y)\|_{H^1_{\nu}(\mathbb{R}^n)} \, dy \lesssim \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})} \|B\|_{L^q(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})}, \quad (1.10)$$

where

$$E(x,y) \cdot_2 B(x,y) := \sum_{k=1}^m E_{n+k}(x,y) B_k(x,y).$$

It is known that the div-curl lemma in the classical setting has many applications in PDE and compensated compactness [6]. Similarly, we expect that the flag type div-curl lemmas described above would have interesting implications in these directions as well. For instance, following the ideas in [6], one can study weak convergence problems in the flag Hardy space. And it would be interesting to know whether one can use the flag type regularity (implied by our div-curl lemmas) of certain nonlinear quantities to obtain improved regularity results for certain nonlinear PDE.

This paper is organised as follows. In Section 2 we provide necessary preliminaries with respect to the flag structures. In Section 3 we study the flag iterated commutators as in Definition 1.1 and prove Theorem 1.3. In Section 4 we give a complete treatment of the flag little bmo spaces and flag type Muckenhoupt $A_p$ weights, toward the proof of Theorem 1.5. In the last section, we apply the boundedness of flag commutators from Theorems 1.3 and 1.5 to establish the flag div-curl results, Theorems 1.6 and 1.7.
2 Preliminaries in the flag setting

Recall the classical Poisson kernel on $\mathbb{R}^n$:

$$P(x) := \frac{c_n}{(1 + |x|^2)^{n/2}}.$$  

And we define

$$P_t(x) := \frac{1}{t^n} P(\frac{x}{t}).$$

For $f \in L^1(\mathbb{R}^n)$, let $F(x, t) := P_t * f(x)$. Then we have the following standard pointwise estimates for the Poisson integral, see in particular Stein ([39]).

**Proposition 2.1.** Suppose $f \in L^1(\mathbb{R}^n)$. Then

$$\sup_{(x,t)\in \mathbb{R}_+^{n+1}} t^{n+1} |\nabla_{x,t}^k F(x,t)| \leq C \|f\|_{L^1(\mathbb{R}^n)}.$$  

(2.1)

We now recall the flag Poisson kernel given by

$$P(x, y) = P^{(1)} \ast_{\mathbb{R}^m} P^{(2)}(x, y) = \int_{\mathbb{R}^m} P^{(1)}(x, y - z) P^{(2)}(z)dz$$

where

$$P^{(1)}(x, y) = \frac{c_{n+m}}{(1 + |x|^2 + |y|^2)^{(n+m+1)/2}}$$

and

$$P^{(2)}(z) = \frac{c_m}{(1 + |z|^2)^{(m+1)/2}}$$

are the classical Poisson kernels on $\mathbb{R}^{n+m}$ and $\mathbb{R}^m$, respectively. Then we have

$$P_{t_1, t_2}(x, y) = P^{(1)}_{t_1} \ast_{\mathbb{R}^m} P^{(2)}_{t_2}(x, y).$$

We define the Lusin area function with respect to $u = P_{t_1, t_2} * f$ as follows.

**Definition 2.2.** For $f \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$ and $u(x_1, x_2, t_1, t_2) = P_{t_1, t_2} * f(x_1, x_2)$, $S_F(u)$, the Lusin area integral of $u(x_1, x_2, t_1, t_2)$ is defined by

$$S_F(u)(x_1, x_2) \quad (2.2)$$

$$= \left\{ \int_{\mathbb{R}_+^{n+1}} \int_{\mathbb{R}_+^{m+1}} \chi_{t_1,s}(x_1 - w_1, x_2 - w_2) |t_1 \nabla^{(1)} t_2 \nabla^{(2)} u(w_1, w_2, t_1, t_2)|^2 \frac{dw_1 dt_1 dw_2 dt_2}{t_1^{n+m+1} t_2^{m+1}} \right\}^{\frac{1}{2}},$$

where $\nabla^{(1)} = (\partial_{t_1}, \partial_{w_1,1} \cdots \partial_{w_{1,m}}, \partial_{w_2,1} \cdots \partial_{w_{2,m}})$ is the standard gradient on $\mathbb{R}^{n+m+1}$, and $\nabla^{(2)} = (\partial_{t_2}, \partial_{w_{2,1}} \cdots \partial_{w_{2,m}})$ is the standard gradient on $\mathbb{R}^{m+1}$, and

$$\chi_{t_1, t_2}(x_1, x_2) := \chi^{(1)}_{t_1} \ast_{\mathbb{R}^m} \chi^{(2)}_{t_2}(x_1, x_2),$$

$$\chi^{(1)}_{t_1}(x_1, x_2) := t_1^{-(n+m)} \chi^{(1)}(\frac{x_1}{t_1}, \frac{x_2}{t_1}), \chi^{(2)}(z) := t_2^{-m} \chi^{(2)}(\frac{z}{t_2}), \chi^{(1)}(x, y) \text{ and } \chi^{(2)}(z) \text{ are the indicator function of the unit balls of } \mathbb{R}^{n+m} \text{ and } \mathbb{R}^m, \text{ respectively.}$$

**Definition 2.3.** The flag Hardy space $H^1_F(\mathbb{R}^n \times \mathbb{R}^m)$ is defined to be the collection of $f \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$ such that $S_F(u) \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$. The norm of $H^1_F(\mathbb{R}^n \times \mathbb{R}^m)$ is defined by

$$\|f\|_{H^1_F(\mathbb{R}^n \times \mathbb{R}^m)} = \|S_F(u)\|_{L^1(\mathbb{R}^n \times \mathbb{R}^m)}.$$  

(2.4)
We now recall the definition of the flag BMO space.

**Definition 2.4.** The flag BMO space $\text{BMO}_P(\mathbb{R}^n \times \mathbb{R}^m)$ is defined to be the collection of $b \in L^1_{\text{loc}}(\mathbb{R}^n \times \mathbb{R}^m)$ such that

$$
\|b\|_{\text{BMO}_P(\mathbb{R}^n \times \mathbb{R}^m)} := \sup_{\Omega} \left( \frac{1}{|\Omega|} \int_{T(\Omega)} |t_1 \nabla^{(1)} t_2 \nabla^{(2)} u(w_1, w_2, t_1, t_2)|^2 \frac{dw_1 dt_1 dw_2 dt_2}{t_1 t_2} \right)^{\frac{1}{2}} < \infty,
$$

where the supremum is taken over all open sets in $\mathbb{R}^n \times \mathbb{R}^m$ with finite measures, and $T(\Omega) = \cup_{R \subset \Omega} T(R)$ with the rectangle $R = I \times J$, $\ell(I) \leq \ell(J)$ and $T(R) = I \times (\frac{\ell(I)}{2}, \ell(I)) \times J \times (\frac{\ell(J)}{2}, \ell(J))$.

**Proposition 2.5.** Given two functions $b, f \in L^2(\mathbb{R}^{n+m})$, we have

$$
[[b, R^{(1)}_j], R^{(2)}_k]_2(f) = [[b, R^{(2)}_k], R^{(1)}_j](f).
$$

**Proof.** By definition, we see that

$$
[[b, R^{(1)}_j], R^{(2)}_k]_2(f)(x_1, x_2) = [b, R^{(1)}_j]_2 [R^{(2)}_k * 2 f](x_1, x_2) - [R^{(2)}_k * 2 f, R^{(1)}_j]_2 (b, R^{(1)}_j)(f)(x_1, x_2)
$$

$$
= b(x_1, x_2) R^{(1)}_j * R^{(2)}_k * 2 f(x_1, x_2) - R^{(2)}_k * 2 ((b, R^{(1)}_j)(f))(x_1, x_2)
$$

$$
- R^{(2)}_k * 2 (b \cdot R^{(1)}_j \cdot f - R^{(1)}_j \cdot (b \cdot f))(x_1, x_2)
$$

$$
= b(x_1, x_2) R^{(1)}_j * R^{(2)}_k * 2 f(x_1, x_2) - R^{(1)}_j * (b \cdot R^{(2)}_k * 2 f)(x_1, x_2)
$$

$$
- R^{(2)}_k * 2 (b \cdot R^{(1)}_j \cdot f)(x_1, x_2) + R^{(2)}_k * 2 R^{(1)}_j * (b \cdot f)(x_1, x_2).
$$

And we also have

$$
[[b, R^{(2)}_k], R^{(1)}_j](f)(x_1, x_2) = [b, R^{(2)}_k]_2 R^{(1)}_j * f(x_1, x_2) - R^{(1)}_j * ([b, R^{(2)}_k]_2(f))(x_1, x_2)
$$

$$
= b(x_1, x_2) R^{(2)}_k * 2 R^{(1)}_j * f(x_1, x_2) - R^{(2)}_k * 2 (b \cdot R^{(1)}_j * f)(x_1, x_2)
$$

$$
- R^{(1)}_j * (b \cdot R^{(2)}_k * 2 f - R^{(2)}_k * 2 (b \cdot f))(x_1, x_2)
$$

$$
= b(x_1, x_2) R^{(2)}_k * 2 R^{(1)}_j * f(x_1, x_2) - R^{(1)}_j * (b \cdot R^{(2)}_k * 2 f)(x_1, x_2)
$$

$$
- R^{(1)}_j * (b \cdot R^{(2)}_k * 2 f)(x_1, x_2) + R^{(1)}_j * R^{(2)}_k * 2 (b \cdot f)(x_1, x_2).
$$

It is direct to see that, by changing of variables,

$$
R^{(2)}_k * 2 R^{(1)}_j * f(x_1, x_2) = \int R^{(2)}_k(x_2 - z) R^{(1)}_j(x_1 - y_1, z - y_2) f(y_1, y_2) \, dz \, dy_1 \, dy_2
$$

$$
= \int R^{(2)}_k(\tilde{z} - y_2) R^{(1)}_j(x_1 - y_1, x_2 - \tilde{z}) f(y_1, y_2) \, d\tilde{z} \, dy_1 \, dy_2
$$

$$
= \int R^{(1)}_j(x_1 - y_1, x_2 - \tilde{z}) R^{(2)}_k(\tilde{z} - y_2) f(y_1, y_2) \, d\tilde{z} \, dy_1 \, dy_2
$$

$$
= R^{(1)}_j * R^{(2)}_k * 2 f(x_1, x_2),
$$

which implies that (2.6) holds. ☐
3 Upper bound of the iterated commutator $[[b, R_i^{(1)}], R_j^{(2)}]_2$

In this section, we prove Theorem 1.3, i.e., the upper bound of the iterated commutator $[[b, R_i^{(1)}], R_j^{(2)}]_2$. As we pointed out earlier, in the flag setting, there is lack of a suitable wavelet basis or Haar basis and hence the approaches in [27, 8] are not available. We establish a fundamental duality argument (Lemma 3.3) with respect to general flag type area integrals and flag Carleson measures, and then apply the technique of harmonic expansion to obtain the full versions of flag type Carleson measure inequalities (Proposition 3.5), which plays the role of “paraproducts”. Then, by considering the bilinear form associated with the iterated commutator $[[b, R_i^{(1)}], R_j^{(2)}]_2$ and by integration by part, we can decompose the bilinear form into a summation of different versions of “paraproducts”. Then the upper bound of the iterated commutator $[[b, R_i^{(1)}], R_j^{(2)}]_2$ follows from applying Proposition 3.5 to each “paraproducts”.

3.1 Extension via flag Poisson operator

For any $f \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$, we define the flag Poisson integral of $f$ by

$$F(x_1, x_2, t_1, t_2) := P_{t_1, t_2} * f(x_1, y_2),$$

where

$$P_{t_1, t_2}(x_1, x_2) = P_{t_1}^{(2)} * P_{t_2}^{(2)}(x_1, x_2).$$

Since $P(x_1, x_2) \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$, it easy to see that $F(x_1, x_2, t_1, t_2)$ is well-defined. Moreover, for any fixed $t_1$ and $t_2$, $P_{t_1, t_2} * f(x_1, x_2)$ is a bounded $C^\infty$ function and the function $F(x_1, x_2, t_1, t_2)$ is harmonic in $(x_1, x_2, t_1)$ and $(x_2, t_2)$, respectively. $F(x_1, x_2, t_1, t_2)$ is the flag harmonic extension of $f$ to $\mathbb{R}^{n+m+1}_+$. More precisely,

$$\begin{align}
\Delta_{\mathbb{R}^{n+m+1}} F(x_1, x_2, t_1, t_2) &= (\partial_{t_1}^2 + \Delta_{x_1, x_2}) F(x_1, x_2, t_1, t_2) = 0 \quad \text{in } \mathbb{R}^{n+m+1}_+; \\
\Delta_{\mathbb{R}^{n+m}} F(x_1, x_2, t_1, t_2) &= (\partial_{t_2}^2 + \Delta_{x_1, x_2}) F(x_1, x_2, t_1, t_2) = 0 \quad \text{in } \mathbb{R}^{n+m}_+; \\
\lim_{t_1 \to 0} \partial_{t_1} F(x_1, x_2, t_1, t_2) &= - (\Delta_{x_1, x_2})^{1/2} P_{t_2}^{(2)} * f(x_1, x_2) \quad \text{on } \mathbb{R}^{n+m}; \\
\lim_{t_2 \to 0} \partial_{t_2} F(x_1, x_2, t_1, t_2) &= - (\Delta_{x_2})^{1/2} P_{t_1}^{(2)} * f(x_1, x_2) \quad \text{on } \mathbb{R}^{n+m}; \\
\lim_{t_1 \to 0} F(x_1, x_2, t_1, t_2) &= P_{t_2}^{(2)} * f(x_1, x_2) \quad \text{on } \mathbb{R}^{n+m}; \\
\lim_{t_2 \to 0} F(x_1, x_2, t_1, t_2) &= P_{t_1}^{(2)} * f(x_1, x_2) \quad \text{on } \mathbb{R}^{n+m}; \\
\lim_{t_1 \to 0, t_2 \to 0} F(x_1, x_2, t_1, t_2) &= f(x_1, x_2) \quad \text{on } \mathbb{R}^{n+m}; \\
\lim_{|x_1, x_2, t_1| \to \infty} F(x_1, x_2, t_1, t_2) &= 0; \\
\lim_{|x_2, t_2| \to \infty} F(x_1, x_2, t_1, t_2) &= 0.
\end{align}$$

We then have the following lemma providing a connection between the boundary values $f$ and the flag harmonic extension $F$. This follows from the decay of the flag harmonic extensions of $f$ and repeated applications of integration by parts in the variables $t_1$ and $t_2$. 
Lemma 3.1. For $f \in L^1(\mathbb{R}^n \times \mathbb{R}^m)$, let $F$ be the same as in (3.1). Then we have
\[
\int_{\mathbb{R}^n \times \mathbb{R}^m} f(x_1, x_2) dx_1 dx_2 = \int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{m+1}_+} t_1 \partial^2_{t_1} t_2 \partial^2_{t_2} F(x_1, x_2, t_1, t_2) dx_1 dx_2 dt_1 dt_2.
\] (3.4)

Proof. We start from the right-hand side of (3.4). We write
\[
\int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{m+1}_+} t_1 \partial^2_{t_1} t_2 \partial^2_{t_2} F(x_1, x_2, t_1, t_2) dx_1 dx_2 dt_1 dt_2 \\
= \int_{\mathbb{R}^{m+1}_+} t_2 \partial^2_{t_2} P^{(2)}_{t_2} \ast \left( \int_{\mathbb{R}^{n+1}_+} t_1 \partial^2_{t_1} P^{(1)}_{t_1} \ast f(x_1, x_2) dx_1 dt_1 \right) dx_2 dt_2 \\
= \int_{\mathbb{R}^m} \left( \int_{\mathbb{R}^{n+1}_+} t_1 \partial^2_{t_1} P^{(1)}_{t_1} \ast f(x_1, x_2) dx_1 dt_1 \right) dx_2,
\]
where the last equality follows from decay of the flag harmonic extensions of $f$ and using the integration by part in the variables $t_2$. To continue, we write the right-hand side of the last equality above as
\[
\int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{m+1}_+} t_1 \partial^2_{t_1} P^{(1)}_{t_1} \ast f(x_1, x_2) dx_1 dt_1 = \int_{\mathbb{R}^{n+m}_+} f(x_1, x_2) dx_1 dx_2,
\]
which yields (3.4). Again, the last equality follows from decay of the flag harmonic extensions of $f$ and using the integration by part in the variables $t_1$.

3.2 Flag area functions and estimates

We also have a more general version of the area function.

Definition 3.2. For a function $G(x_1, x_2, t_1, t_2)$ defined on $\mathbb{R}^{n+1}_+ \times \mathbb{R}^{m+1}_+$, the general flag type Lusin area integral of $G$ is defined by
\[
S_{F,L}(G)(x_1, x_2) := \left\{ \int_{\mathbb{R}^{n+1}_+} \int_{\mathbb{R}^{m+1}_+} \chi_{t,s}(x_1 - w_1, x_2 - w_2) |G(w_1, w_2, t_1, t_2)|^2 \frac{dw_1 dt_1}{t_1^{n+m+1}} \frac{dw_2 dt_2}{t_2^{n+m+1}} \right\}^{1/2}.
\] (3.5)

Lemma 3.3. Suppose $F(x_1, x_2, t_1, t_2)$ and $G(x_1, x_2, t_1, t_2)$ are defined on $\mathbb{R}^{n+1}_+ \times \mathbb{R}^{m+1}_+$. Then the following estimate holds:
\[
\int_{\mathbb{R}^{n+1}_+} \int_{\mathbb{R}^{m+1}_+} F(x_1, x_2, t_1, t_2) G(x_1, x_2, t_1, t_2) dx_1 dx_2 dt_1 dt_2 \\
\leq C \sup_{\Omega \subset \mathbb{R}^n \times \mathbb{R}^m} \left( \frac{1}{|\Omega|} \int_{T(\Omega)} t_1 t_2 |F(y_1, y_2, t_1, t_2)|^2 dy_1 dy_2 dt_1 dt_2 \right)^{1/2} \\
\times \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \left( \int_{\mathbb{R}^{n+1}_+} \int_{\mathbb{R}^{m+1}_+} \chi_{t_1,t_2}(x_1 - y_1, x_2 - y_2) |G(y_1, y_2, t_1, t_2)|^2 \frac{dy_1 dy_2 dt_1 dt_2}{t_1^{n+m+1} t_2^{n+m+1}} \right)^{1/2} dx_1 dx_2.
\] (3.6)

Proof. Suppose both factors on the right-hand side above are finite, since otherwise there is nothing to prove. We also note that the second factor is actually $\|S_{F,G}\|_{L^1(\mathbb{R}^n \times \mathbb{R}^m)}$. 
We now let
\[ \Omega_k := \{(x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^m : S_{F,L}(G)(x_1, x_2) > 2^k \} \]
and define
\[ B_k := \{ R = I_1 \times I_2 : |(I_1 \times I_2) \cap \Omega_k| > \frac{1}{2} |I_1 \times I_2|, |(I_1 \times I_2) \cap \Omega_{k+1}| \leq \frac{1}{2} |I_1 \times I_2| \}, \]
where \( I_1 \) and \( I_2 \) are dyadic cubes in \( \mathbb{R}^n \) and \( \mathbb{R}^m \) with side-lengths \( \ell(I) \) and \( \ell(J) \) satisfying \( \ell(I) \leq \ell(J) \). Moreover, we define
\[ \Omega_k = \bigcup_{R \in B_k} R \]
and
\[ \tilde{\Omega}_k = \left\{ (x_1, x_2) \in \mathbb{R}^n \times \mathbb{R}^m : M_{flag}(\chi_{\Omega_k})(x_1, x_2) > \frac{1}{2} \right\}. \]
Next, we have
\[
\int_{\mathbb{R}^n} \int_{\mathbb{R}^m} F(x_1, x_2, t_1, t_2) G(x_1, x_2, t_1, t_2) \, dx_1 dx_2 dt_1 dt_2 \\
= \sum_k \sum_{R \in B_k} \int_{T(R)} \sqrt{t_1 t_2} F(x_1, x_2, t_1, t_2) \frac{G(x_1, x_2, t_1, t_2)}{\sqrt{t_1 t_2}} \, dx_1 dx_2 dt_1 dt_2 \\
\leq \sum_k \left( \sum_{R \in B_k} \int_{T(R)} t_1 t_2 |F(x_1, x_2, t_1, t_2)|^2 \, dx_1 dx_2 dt_1 dt_2 \right)^{1/2} \\
\times \left( \sum_{R \in B_k} \int_{T(R)} |G(x_1, x_2, t_1, t_2)|^2 \frac{dx_1 dx_2 dt_1 dt_2}{t_1 t_2} \right)^{1/2} \\
= \sum_k \left( \frac{1}{|\Omega_k|} \sum_{R \in B_k} \int_{T(R)} t_1 t_2 |F(x_1, x_2, t_1, t_2)|^2 \, dx_1 dx_2 dt_1 dt_2 \right)^{1/2} \\
\times \left( \frac{1}{|\Omega_k|} \sum_{R \in B_k} \int_{T(\Omega_k)} |G(x_1, x_2, t_1, t_2)|^2 \frac{dx_1 dx_2 dt_1 dt_2}{t_1 t_2} \right)^{1/2} \\
\leq \sum_k \left( \frac{1}{|\tilde{\Omega}_k|} \sum_{R \in B_k} \int_{T(\Omega_k)} t_1 t_2 |F(x_1, x_2, t_1, t_2)|^2 \, dx_1 dx_2 dt_1 dt_2 \right)^{1/2} \\
\times \left( \frac{1}{|\Omega_k|} \sum_{R \in B_k} \int_{T(\Omega_k)} |G(x_1, x_2, t_1, t_2)|^2 \frac{dx_1 dx_2 dt_1 dt_2}{t_1 t_2} \right)^{1/2} \\
\leq \sup_{\Omega \subseteq \mathbb{R}^n \times \mathbb{R}^m} \left( \frac{1}{|\Omega|} \int_{T(\Omega)} t_1 t_2 |F(x_1, x_2, t_1, t_2)|^2 \, dx_1 dx_2 dt_1 dt_2 \right)^{1/2} \\
\times \sum_k \left( \frac{1}{|\tilde{\Omega}_k|} \sum_{R \in B_k} \int_{T(\Omega_k)} |G(x_1, x_2, t_1, t_2)|^2 \frac{dx_1 dx_2 dt_1 dt_2}{t_1 t_2} \right)^{1/2}.
\]
As for the second factor in the last inequality above, note that
\[
2^{2k} |\tilde{\Omega}_k \setminus \Omega_k| \\
\geq \int_{\Omega_k \setminus \Omega_k} S_{F,L}(G)(x_1, x_2)^2 \, dx_1 dx_2
\]
Flag commutators

Thus, we have

\[
\int_{R_+^{n+1}} \int_{R_+^{m+1}} F(x_1, x_2, t_1, t_2) G(x_1, x_2, t_1, t_2) \, dx_1 dx_2 dt_1 dt_2 \\
\leq \sup_{\Omega \subset \mathbb{R}^n \times \mathbb{R}^m} \left( \frac{1}{|\Omega|} \int_{T(\Omega)} t_1 t_2 |F(x_1, x_2, t_1, t_2)|^2 \, dx_1 dx_2 dt_1 dt_2 \right)^{1/2} \\
\times \sum_k \left( \tilde{\Omega}_k |2^{2k} \tilde{\Omega}_k \setminus \tilde{\Omega}_k| \right)^{1/2} \\
\leq \sup_{\Omega \subset \mathbb{R}^n \times \mathbb{R}^m} \left( \frac{1}{|\Omega|} \int_{T(\Omega)} |t_1 t_2 F(x_1, x_2, t_1, t_2)|^2 \frac{dx_1 dx_2 dt_1 dt_2}{t_1 t_2} \right)^{1/2} \\
\times \sum_k |\Omega_k| 2^k \\
\leq \sup_{\Omega \subset \mathbb{R}^n \times \mathbb{R}^m} \left( \frac{1}{|\Omega|} \int_{T(\Omega)} |t_1 t_2 F(x_1, x_2, t_1, t_2)|^2 \frac{dx_1 dx_2 dt_1 dt_2}{t_1 t_2} \right)^{1/2} \\
\times \|S_{F,L}(G)\|_{L^1(\mathbb{R}^n \times \mathbb{R}^m)},
\]

which gives (3.6). This completes the proof of the Lemma 3.3.

From Lemma 3.3 above and the definition of $\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)$, we can obtain the following Corollary immediately.

**Corollary 3.4.** Suppose $G(x_1, x_2, t_1, t_2)$ is defined on $\mathbb{R}_+^{n+1} \times \mathbb{R}_+^{m+1}$, and $F(x_1, x_2, t_1, t_2) := P_{t_1,t_2} * f(x_1, x_2)$, where $f \in \text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)$. Then we have:

\[
\int_{R_+^{n+1}} \int_{R_+^{m+1}} |\nabla^{(1)} \nabla^{(2)} F(x_1, x_2, t_1, t_2)| \cdot |G(x_1, x_2, t_1, t_2)| \, dx_1 dx_2 dt_1 dt_2 \tag{3.7}
\leq C \|f\|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|S_{F,L}(G)\|_{L^1(\mathbb{R}^n \times \mathbb{R}^m)}.
\]

Moreover, based on Lemma 3.3 we can also establish the following estimates.

**Proposition 3.5.** Suppose $F(x_1, x_2, t_1, t_2) = P_{t_1,t_2} * f(x_1, x_2)$, $G(x_1, x_2, t_1, t_2) = P_{t_1,t_2} * g(x_1, x_2)$, and $B(x_1, x_2, t_1, t_2) = P_{t_1,t_2} * b(x_1, x_2)$. Then we have

1.

\[
\int_{R_+^{n+1} \times R_+^{m+1}} t_1 t_2 |\nabla^{(1)} \nabla^{(2)} B(x_1, x_2, t_1, t_2)| \cdot |\nabla_{x_1,x_2} \nabla_{x_2} \nabla^{(1)} \nabla^{(2)} G(x_1, x_2, t_1, t_2)| \\
\times |\nabla^{(1)} \nabla^{(2)} F(x_1, x_2, t_1, t_2)| \, dx_1 dx_2 dt_1 dt_2 \\
\leq C \|b\|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|(-\Delta_{x_1,x_2})^{1/2}(-\Delta_{x_2})^{1/2} g\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \|(-\Delta_{x_1,x_2})^{1/2}(-\Delta_{x_2})^{1/2} f\|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)}; \tag{3.8}
\]
2. \[
\int_{\mathbb{R}^n_+ \times \mathbb{R}^m_+} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| |\nabla x_1 x_2 \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2)| \leq C |b|_{\text{BMO}_p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_1 x_2)^{1/2} (-\Delta x_2)^{1/2} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_1 x_2)^{1/2} f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)}; \\
\]

3. \[
\int_{\mathbb{R}^n_+ \times \mathbb{R}^m_+} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| |\nabla x_1 x_2 \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2)| \leq C |b|_{\text{BMO}_p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_1 x_2)^{1/2} (-\Delta x_2)^{1/2} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_2)^{1/2} f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)}; \\
\]

4. \[
\int_{\mathbb{R}^n_+ \times \mathbb{R}^m_+} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| |\nabla x_1 x_2 \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2)| \leq C |b|_{\text{BMO}_p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_1 x_2)^{1/2} (-\Delta x_2)^{1/2} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)}; \\
\]

5. \[
\int_{\mathbb{R}^n_+ \times \mathbb{R}^m_+} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| |\nabla x_1 x_2 \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2)| \leq C |b|_{\text{BMO}_p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_1 x_2)^{1/2} (-\Delta x_2)^{1/2} f \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)}; \\
\]

6. \[
\int_{\mathbb{R}^n_+ \times \mathbb{R}^m_+} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| |\nabla x_1 x_2 \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2)| \leq C |b|_{\text{BMO}_p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_1 x_2)^{1/2} (-\Delta x_2)^{1/2} f \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)}; \\
\]

7. \[
\int_{\mathbb{R}^n_+ \times \mathbb{R}^m_+} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| |\nabla x_1 x_2 \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2)| \leq C |b|_{\text{BMO}_p(\mathbb{R}^n \times \mathbb{R}^m)} \| (\Delta x_1 x_2)^{1/2} (-\Delta x_2)^{1/2} f \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)}; \\
\]
8. \[
\int_{\mathbb{R}^{n+1}_{+} \times \mathbb{R}^{n+1}_{+}} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| \mid \nabla_{x_1, x_2} \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2) | \leq C \| \nabla \|_{BMO_{F}(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_1, x_2)^{\frac{1}{2}} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)} ;
\]

9. \[
\int_{\mathbb{R}^{n+1}_{+} \times \mathbb{R}^{n+1}_{+}} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| \mid \nabla_{x_2} \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2) | \leq C \| \nabla \|_{BMO_{F}(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_2)^{\frac{1}{2}} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)} ;
\]

10. \[
\int_{\mathbb{R}^{n+1}_{+} \times \mathbb{R}^{n+1}_{+}} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| \mid \nabla_{x_2} \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2) | \leq C \| \nabla \|_{BMO_{F}(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_2)^{\frac{1}{2}} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_2)^{\frac{1}{2}} f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)} ;
\]

11. \[
\int_{\mathbb{R}^{n+1}_{+} \times \mathbb{R}^{n+1}_{+}} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| \mid \nabla_{x_2} \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2) | \leq C \| \nabla \|_{BMO_{F}(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_2)^{\frac{1}{2}} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_2)^{\frac{1}{2}} f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)} ;
\]

12. \[
\int_{\mathbb{R}^{n+1}_{+} \times \mathbb{R}^{n+1}_{+}} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| \mid \nabla_{x_2} \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2) | \leq C \| \nabla \|_{BMO_{F}(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_2)^{\frac{1}{2}} g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)} ;
\]

13. \[
\int_{\mathbb{R}^{n+1}_{+} \times \mathbb{R}^{n+1}_{+}} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| \mid \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2) | \leq C \| \nabla \|_{BMO_{F}(\mathbb{R}^n \times \mathbb{R}^m)} \| g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_1, x_2)^{\frac{1}{2}} (-\Delta x_2)^{\frac{1}{2}} f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)} ;
\]

14. \[
\int_{\mathbb{R}^{n+1}_{+} \times \mathbb{R}^{n+1}_{+}} t_1 t_2 |\nabla (1) \nabla (2) B(x_1, x_2, t_1, t_2)| \mid \nabla (1) \nabla (2) G(x_1, x_2, t_1, t_2) | \leq C \| \nabla \|_{BMO_{F}(\mathbb{R}^n \times \mathbb{R}^m)} \| g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| (-\Delta x_1, x_2)^{\frac{1}{2}} f \|_{L^{p'}(\mathbb{R}^n \times \mathbb{R}^m)} ;
\]
To begin with, we first point out that for \( f \in C_0^\infty(\mathbb{R}^{n+m}) \), \( F(x_1,x_2,t_1,t_2) = P_{t_1,t_2} \ast f(x_1,x_2) \) are the Hardy-Littlewood maximal functions on \( \mathbb{R}^{n+m} \) and \( \mathbb{R}^m \), respectively.

Next, based on the estimate above and from the property of the Poisson semigroup, we have

\[
\sup_{(y_1,y_2,t_1,t_2) : \chi_{t_1,t_2}(x_1-y_1,x_2-y_2) \neq 0} |P_{t_1,t_2} \ast f(y_1,y_2)| \leq M_1 \left( M_2 \left( \left( -\Delta_{x_1,x_2} \right)^{\frac{1}{2}} f \right)(1,\cdot) \right)(x_1,x_2).
\]

Also, we have

\[
\sup_{(y_1,y_2,t_1,t_2) : \chi_{t_1,t_2}(x_1-y_1,x_2-y_2) \neq 0} |\nabla y_1 \nabla y_2 F(y_1,y_2,t_1,t_2)| \leq M_1 \left( M_2 \left( \left( -\nabla_{1,2} \nabla_{1,2} f \right)(1,\cdot) \right)(x_1,x_2) \right).
\]

Then, we first consider (3.3). Based on the estimates above and Corollary 3.4, we have

\[
\int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} |\nabla_{x_1,x_2} \nabla_{x_1,x_2} \nabla_{x_2} f(x_1,x_2,t_1,t_2)| dt_1 dt_2.
\]
\[ \leq C \| b \|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \int_{\mathbb{R}^n \times \mathbb{R}^m} S_f \left( t_1 t_2 \nabla_{x_1, x_2} \nabla^{(1)} G \right)(x_1, x_2) \times \left( M_1 \left( M_2 \left( \left( (\Delta_{x_1, x_2})^{1/2} (\Delta_{x_2})^{1/2} f \right) (1, \cdot) \right) (x_1, x_2) \right) + M_1 \left( M_2 \left( \left( \nabla_{1, 2} \nabla_{2} f \right) (1, \cdot) \right) (x_1, x_2) \right) \right) dx_1 dx_2 \leq C \| b \|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \int_{\mathbb{R}^n \times \mathbb{R}^m} S_f \left( \nabla_{x_1, x_2} \nabla_{x_2} (\Delta_{x_1, x_2})^{1/2} (\Delta_{x_2})^{1/2} G \right)(x_1, x_2) \times \left( M_1 \left( M_2 \left( \left( (\Delta_{x_1, x_2})^{1/2} (\Delta_{x_2})^{1/2} f \right) (1, \cdot) \right) (x_1, x_2) \right) + M_1 \left( M_2 \left( \left( \nabla_{1, 2} \nabla_{2} f \right) (1, \cdot) \right) (x_1, x_2) \right) \right) dx_1 dx_2 \leq C \| b \|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \left\| (\Delta_{x_1, x_2})^{1/2} (\Delta_{x_2})^{1/2} g \right\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \left\| (\Delta_{x_1, x_2})^{1/2} (\Delta_{x_2})^{1/2} f \right\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \right. \]

where in the second inequality the area function \( S_f \) is defined as in Definition 2.2, and the last inequality follows from Hölder’s inequality and boundedness of the maximal functions as well as the boundedness of the flag Riesz transforms. Hence we see that (3.8) holds.

By using similar estimate as above, we can obtain the estimates in (3.9), (3.24). We omit the details here since they are straightforward. \( \square \)

### 3.3 Upper bound for iterated commutators

**Theorem 3.6.** For every \( b \in \text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m) \), \( g \in C^\infty_c(\mathbb{R}^n \times \mathbb{R}^m) \) and for any \( i = 1, 2, \ldots, m+n, j = 1, \ldots, n \), there exists a positive constant \( C \) depending only on \( p, n \) and \( m \) such that

\[ \left\| \left[ b, R_i^{(1)} \right], R_j^{(2)} \right\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \leq C \| b \|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \| g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)}. \] (3.24)

**Proof.** Recall that

\[ \left[ b, R_i^{(1)} \right], R_j^{(2)} \right) \right) dx_1 dx_2 = \left( b(x_1, x_2) R_i^{(1)} * R_j^{(2)} g(x_1, x_2) \right)_2 \left( b(x_1, x_2) R_i^{(1)} * R_j^{(2)} g(x_1, x_2) \right) \]

\[ - R_j^{(2)} * \left( b \cdot R_i^{(1)} * g \right) \left( x_1, x_2 \right) + R_j^{(2)} \left. \left( b \cdot R_i^{(1)} * g \right) \right) \left( x_1, x_2 \right). \]

Hence, for every \( f \in C^\infty_c(\mathbb{R}^n \times \mathbb{R}^m) \), we have

\[ \langle f, \left[ b, R_i^{(1)} \right], R_j^{(2)} \right) \rangle = \langle f * b, R_i^{(1)} * R_j^{(2)} * g \rangle + \langle R_i^{(1)} * f, b \cdot R_j^{(2)} * g \rangle \]

\[ + \langle R_j^{(2)} * f, b \cdot R_i^{(1)} * g \rangle + \langle R_j^{(2)} * f, b \cdot R_i^{(1)} * g \rangle. \]

Denote by \( B, F, G \) the flag harmonic extension of the functions \( b, f, g \), respectively, as defined in (3.1). And for each fixed \( i, j \), denote by \( R_i^{(1)} * f \), \( R_j^{(2)} * f \) and \( R_i^{(1)} * R_j^{(2)} * f \) the flag harmonic extension of \( R_i^{(1)} * f \), \( R_j^{(2)} * f \) and \( R_i^{(1)} * R_j^{(2)} * f \).
Then we write
\[
\langle f, \left[ [b, R_i^{(1)}], R_j^{(2)} \right]_2 \rangle (g) = \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 \left( F \cdot B \cdot R_i^{(1)} \ast \widetilde{R_j^{(2)}} \ast_2 g + R_i^{(1)} \ast f \cdot B \cdot \widetilde{R_j^{(2)}} \ast_2 g \right. \\
+ R_j^{(2)} \ast_2 f \cdot B \cdot R_i^{(1)} \ast g + R_j^{(2)} \ast_2 R_i^{(1)} \ast f \cdot B \cdot G \right) dx_1 dx_2 dt_1 dt_2.
\]

We now claim that the right-hand side of (3.25) is bounded by
\[
C \| b \|_{BMO_{\mathcal{P}(\mathbb{R}^n \times \mathbb{R}^m)}} \| g \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \| f \|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)}.
\]

To see this, we compute the derivatives \( t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 \) for the integrand in the right-hand side of (3.25). Then we have the following terms:

\[
\begin{align*}
C_1 &= \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_2} \left( F \cdot R_i^{(1)} \ast \widetilde{R_j^{(2)}} \ast_2 g + t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_2} \left( R_i^{(1)} \ast f \cdot \widetilde{R_j^{(2)}} \ast_2 g \right. \\
&\quad + t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_2} \left( R_j^{(2)} \ast_2 f \cdot R_i^{(1)} \ast g \right) \\
&\quad + t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_2} \left( R_j^{(2)} \ast_2 R_i^{(1)} \ast f \cdot G \right) \right) dx_1 dx_2 dt_1 dt_2; \\
C_2 &= \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( F \cdot R_i^{(1)} \ast \widetilde{R_j^{(2)}} \ast_2 g \right) + t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( R_i^{(1)} \ast f \cdot \widetilde{R_j^{(2)}} \ast_2 g \right) \\
&\quad + t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( R_j^{(2)} \ast_2 f \cdot R_i^{(1)} \ast g \right) \\
&\quad + t_1 \partial_{t_1}^2 t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( R_j^{(2)} \ast_2 R_i^{(1)} \ast f \cdot G \right) dx_1 dx_2 dt_1 dt_2; \\
C_3 &= \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( F \cdot R_i^{(1)} \ast \widetilde{R_j^{(2)}} \ast_2 g \right) + t_1 \partial_{t_1} t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( R_i^{(1)} \ast f \cdot \widetilde{R_j^{(2)}} \ast_2 g \right) \\
&\quad + t_1 \partial_{t_1} t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( R_j^{(2)} \ast_2 f \cdot R_i^{(1)} \ast g \right) \\
&\quad + t_1 \partial_{t_1} t_2 \partial_{t_2}^2 B \cdot \partial_{t_1} \left( R_j^{(2)} \ast_2 R_i^{(1)} \ast f \cdot G \right) dx_1 dx_2 dt_1 dt_2; \\
C_4 &= \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \partial_{t_2} B \cdot \partial_{t_1} t_2 \partial_{t_2} \left( F \cdot R_i^{(1)} \ast \widetilde{R_j^{(2)}} \ast_2 g \right) \\
&\quad + t_1 \partial_{t_1} t_2 \partial_{t_2} B \cdot \partial_{t_1} t_2 \partial_{t_2} \left( R_i^{(1)} \ast f \cdot \widetilde{R_j^{(2)}} \ast_2 g \right) + t_1 \partial_{t_1} t_2 \partial_{t_2} B \cdot \partial_{t_1} t_2 \partial_{t_2} \left( R_j^{(2)} \ast_2 f \cdot R_i^{(1)} \ast g \right) \\
&\quad + t_1 \partial_{t_1} t_2 \partial_{t_2} B \cdot \partial_{t_1} t_2 \partial_{t_2} \left( R_j^{(2)} \ast_2 R_i^{(1)} \ast f \cdot G \right) dx_1 dx_2 dt_1 dt_2; \\
C_5 &= \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial_{t_2} \left( F \cdot R_i^{(1)} \ast \widetilde{R_j^{(2)}} \ast_2 g \right) + t_1 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial_{t_2} \left( R_i^{(1)} \ast f \cdot \widetilde{R_j^{(2)}} \ast_2 g \right)
\end{align*}
\]
Flag commutators

\[ + t_1 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( \widetilde{R}_j^{(2)} \star_2 f \cdot R_i^{(1)} \star g \right) \]

\[ + t_1 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( \widetilde{R}_j^{(2)} \star_2 R_i^{(1)} \star f \cdot G \right) \]

\[ C_6 = \int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{n+1}_+} t_1 t_2 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( F \cdot R_i^{(1)} \star R_j^{(2)} \star_2 g \right) + t_1 t_2 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( R_i^{(1)} \star f \cdot R_j^{(2)} \star_2 g \right) \]

\[ + t_1 t_2 \partial_{t_1} \partial^2_{t_2} B \cdot \partial_{t_1} \partial^2_{t_2} \left( R_j^{(2)} \star_2 f \cdot R_i^{(1)} \star g \right) + t_1 t_2 \partial_{t_1} \partial^2_{t_2} B \cdot \partial_{t_1} \partial^2_{t_2} \left( R_j^{(2)} \star_2 R_i^{(1)} \star f \cdot G \right) \]

\[ dx_1 dx_2 dt_1 dt_2; \]  \hspace{1cm} (3.31)

\[ C_7 = \int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{n+1}_+} t_1 t_2 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( F \cdot R_i^{(1)} \star R_j^{(2)} \star_2 g \right) + t_1 t_2 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( R_i^{(1)} \star f \cdot R_j^{(2)} \star_2 g \right) \]

\[ + t_1 t_2 \partial^2_{t_1} B \cdot \partial^2_{t_1} \left( R_j^{(2)} \star_2 f \cdot R_i^{(1)} \star g \right) + t_1 t_2 \partial^2_{t_1} B \cdot \partial^2_{t_1} \left( R_j^{(2)} \star_2 R_i^{(1)} \star f \cdot G \right) \]

\[ dx_1 dx_2 dt_1 dt_2; \]  \hspace{1cm} (3.32)

\[ C_8 = \int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{n+1}_+} t_1 t_2 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( F \cdot R_i^{(1)} \star \widetilde{R}_j^{(2)} \star_2 g \right) + t_1 t_2 \partial_{t_1} t_2 B \cdot \partial_{t_1} \partial^2_{t_2} \left( \widetilde{R}_i^{(1)} \star f \cdot \widetilde{R}_j^{(2)} \star_2 g \right) \]

\[ + t_1 t_2 \partial^2_{t_1} B \cdot \partial^2_{t_1} \left( \widetilde{R}_j^{(2)} \star_2 f \cdot \widetilde{R}_i^{(1)} \star g \right) + t_1 t_2 \partial^2_{t_1} B \cdot \partial^2_{t_1} \left( \widetilde{R}_j^{(2)} \star_2 \widetilde{R}_i^{(1)} \star f \cdot G \right) \]

\[ dx_1 dx_2 dt_1 dt_2; \]  \hspace{1cm} (3.33)

\[ C_9 = \int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{n+1}_+} t_1 t_2 B \cdot \partial^2_{t_1} \partial^2_{t_2} \left( F \cdot \widetilde{R}_i^{(1)} \star \widetilde{R}_j^{(2)} \star_2 g \right) + t_1 t_2 B \cdot \partial^2_{t_1} \partial^2_{t_2} \left( \widetilde{R}_i^{(1)} \star f \cdot \widetilde{R}_j^{(2)} \star_2 g \right) \]

\[ + t_1 t_2 B \cdot \partial^2_{t_1} \partial^2_{t_2} \left( \widetilde{R}_j^{(2)} \star_2 f \cdot \widetilde{R}_i^{(1)} \star g \right) + t_1 t_2 B \cdot \partial^2_{t_1} \partial^2_{t_2} \left( \widetilde{R}_j^{(2)} \star_2 \widetilde{R}_i^{(1)} \star f \cdot G \right) \]

\[ dx_1 dx_2 dt_1 dt_2. \]  \hspace{1cm} (3.34)

We first consider \( C_1 \). Note that \( \partial^2_{t_1} B = -\Delta x_2 B = -\nabla x_2 \cdot \nabla x_2 B \) and that \( \partial^2_{t_1} B = -\Delta x_1 x_2 B = -\nabla x_1 x_2 \cdot \nabla x_1 x_2 B \). So, integration by parts gives

\[ C_1 = \int_{\mathbb{R}^{n+1}_+ \times \mathbb{R}^{n+1}_+} t_1 t_2 \nabla x_1 x_2 \nabla x_2 B \cdot \nabla x_1 x_2 \nabla x_2 \left( F \cdot \widetilde{R}_i^{(1)} \star R_j^{(2)} \star_2 g \right) \]

\[ + t_1 t_2 \nabla x_1 x_2 \nabla x_2 B \cdot \nabla x_1 x_2 \nabla x_2 \left( \widetilde{R}_i^{(1)} \star f \cdot R_j^{(2)} \star_2 g \right) \]

\[ + t_1 t_2 \nabla x_1 x_2 \nabla x_2 B \cdot \nabla x_1 x_2 \nabla x_2 \left( R_j^{(2)} \star_2 f \cdot \widetilde{R}_i^{(1)} \star g \right) \]

\[ + t_1 t_2 \nabla x_1 x_2 \nabla x_2 B \cdot \nabla x_1 x_2 \nabla x_2 \left( R_j^{(2)} \star_2 \widetilde{R}_i^{(1)} \star f \cdot G \right) \]

\[ dx_1 dx_2 dt_1 dt_2 \]

\[ =: C_{1,1} + C_{1,2} + C_{1,3} + C_{1,4}. \]
For the first term, it is clear that

\[ C_{1,1} = \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} t_1 t_2 \nabla_x \nabla_y B \cdot \nabla_x \nabla_y F \cdot R^{(1)}_i \ast R^{(2)}_j \ast 2 \ g \ dx_1 dx_2 dt_1 dt_2 \]

\[ = \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} t_1 t_2 \nabla_x \nabla_y B \cdot \nabla_x \nabla_y F \cdot R^{(1)}_i \ast R^{(2)}_j \ast 2 \ g \ dx_1 dx_2 dt_1 dt_2 \]

\[ = \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} t_1 t_2 \nabla_x \nabla_y B \cdot \nabla_x \nabla_y F \cdot R^{(1)}_i \ast R^{(2)}_j \ast 2 \ g \ dx_1 dx_2 dt_1 dt_2 \]

\[ = \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} t_1 t_2 \nabla_x \nabla_y B \cdot F \cdot R^{(1)}_i \ast R^{(2)}_j \ast 2 \ g \ dx_1 dx_2 dt_1 dt_2 \]

\[ = C_{1,1,1} + C_{1,1,2} + C_{1,1,3} + C_{1,1,4}. \]

It is direct that \( C_{1,1,1} \) and \( C_{1,1,4} \) can be handled by using (3.9), and \( C_{1,1,2} \) and \( C_{1,1,3} \) can be handled by using (3.10), which gives that

\[ C_{1,1} \leq C \|b\|_{\text{BMO}^p(\mathbb{R}^n \times \mathbb{R}^m)} \|g\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \|f\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)}. \]

Symmetrically we obtain the estimate for \( C_{1,4} \) and using similar estimates we can handle \( C_{1,2} \) and \( C_{1,3} \). All these three terms are have the same upper as \( C_{1,1} \) above.

Next, for \( C_2 \), note that \( \partial^2 B = -\Delta_x \nabla_x \cdot \nabla_x \nabla_x B \). Thus, similar to the term \( C_1 \), by integration by parts, we have

\[ C_2 = -\int_{\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}} t_1 t_2 \nabla_x \nabla_y \partial_x \partial_y B \cdot \nabla_x \nabla_y \partial_x \partial_y \left( F \cdot R^{(1)}_i \ast R^{(2)}_j \ast 2 \ g \right) \]

\[ + t_1 t_2 \nabla_x \nabla_y \partial_x \partial_y B \cdot \nabla_x \nabla_y \partial_x \partial_y \left( R^{(1)}_i \ast f \ast R^{(2)}_j \ast 2 \ g \right) \]

\[ + t_1 t_2 \nabla_x \nabla_y \partial_x \partial_y B \cdot \nabla_x \nabla_y \partial_x \partial_y \left( R^{(2)}_j \ast 2 \ R^{(1)}_i \ast f \cdot G \right) \ dx_1 dx_2 dt_1 dt_2 \]

\[ =: C_{2,1} + C_{2,2} + C_{2,3} + C_{2,4}. \]

Again, the upper bounds from the four terms above can be obtained by applying Proposition 3.5 and they are all controlled by

\[ C \|b\|_{\text{BMO}^p(\mathbb{R}^n \times \mathbb{R}^m)} \|g\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \|f\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)}. \]

The term \( C_3 \) can be handled symmetrically to \( C_2 \) and we obtain the same upper bounds.

For the term \( C_4 \), by noting that \( |\partial_x \partial_y B(x_1, x_2, t_1, t_2)| \) is bounded by \( |\nabla^{(1)} \nabla^{(2)} B(x_1, x_2, t_1, t_2)| \), we obtain that \( C_4 \) is bounded by

\[ C \|b\|_{\text{BMO}^p(\mathbb{R}^n \times \mathbb{R}^m)} \|g\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)} \|f\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)}, \]

where we apply again the upper bounds in Proposition 3.5.

We now turn to the term \( C_9 \). We first point out the following equalities:

\[ \partial_x \nabla^{(1)} \ast R^{(2)}_j \ast 2 \ g(x_1, x_2) = -c \partial_x (x_1, x_2) \ast R^{(2)}_j \ast 2 \ g(x_1, x_2) \]
Then for the term $C_9$, we get

$$
\partial_t^2 \partial^2_t \left( F \cdot R_i^{(1)} \ast R_j^{(2)} \ast 2 g + R_i^{(1)} \ast f \cdot R_j^{(2)} \ast 2 g + R_j^{(2)} \ast 2 f \cdot R_i^{(1)} \ast g + R_j^{(2)} \ast 2 R_i^{(1)} \ast f \cdot G \right)
= 4 \partial_t \partial^2_t \left( F \cdot R_i^{(1)} \ast R_j^{(2)} \ast 2 g \right)
+ 2 \partial_t \partial^2_t \left( F \cdot \nabla_{x_1,x_2} R_i^{(1)} + f \cdot R_j^{(2)} \ast 2 g \right)
+ 2 \partial_t \partial^2_t \left( \nabla_{x_1,x_2} F \cdot R_i^{(1)} \ast g \right)
+ 2 \partial_t \partial^2_t \left( \nabla_{x_1,x_2} R_i^{(1)} \ast f \cdot G \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} R_i^{(1)} \ast f \cdot G \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} R_j^{(2)} \ast 2 f \cdot R_i^{(1)} \ast g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} R_j^{(2)} \ast 2 f \cdot R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} F \cdot R_i^{(1)} \ast g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} F \cdot R_j^{(2)} \ast 2 g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} R_i^{(1)} \ast f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} R_j^{(2)} \ast 2 f \cdot R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_i^{(1)} \ast R_j^{(2)} \ast 2 g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_j^{(2)} \ast 2 f \cdot R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} F \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} F \cdot \nabla_{x_1,x_2} R_j^{(2)} \ast 2 g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} R_i^{(1)} \ast f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} R_j^{(2)} \ast 2 f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_i^{(1)} \ast R_j^{(2)} \ast 2 g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_j^{(2)} \ast 2 f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} R_i^{(1)} \ast f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} R_j^{(2)} \ast 2 f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_i^{(1)} \ast R_j^{(2)} \ast 2 g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_j^{(2)} \ast 2 f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} R_i^{(1)} \ast f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_j^{(2)} \ast 2 f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_i^{(1)} \ast R_j^{(2)} \ast 2 g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_j^{(2)} \ast 2 f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right)
+ \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_i^{(1)} \ast R_j^{(2)} \ast 2 g \right)
- \partial_t \partial^2_t \left( \nabla_{x_1,x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} \nabla_{x_2} F \cdot R_j^{(2)} \ast 2 f \cdot \nabla_{x_1,x_2} R_i^{(1)} \ast g \right).
$$
Thus, we input the above 25 terms back into the right-hand side of $C_9$ and obtain the terms as follows:

\[
C_9 = \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 t_2 B \cdot \partial^2_{x_1} F \cdot R_i^{(1)} \bigg( F \cdot R_i^{(1)} \bigg) \ast_2 g + R_i^{(2)} \ast_2 f \cdot R_j^{(2)} \ast_2 g + R_j^{(2)} \ast_2 f \cdot R_i^{(2)} \ast_2 g + R_j^{(2)} \ast_2 R_i^{(1)} \ast f \cdot G \bigg) dx_1 dx_2 dt_1 dt_2
\]

\[
= 4 \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 t_2 \partial(x_{x_1,x_2}) \partial(x_{x_2}) B \cdot \partial_{x_1} \partial_{x_2} (FG) dx_1 dx_2 dt_1 dt_2
\]

\[
- 2 \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 t_2 \partial(x_{x_1,x_2}) \partial(x_{x_2}) B \cdot \partial_{x_2} \bigg( \nabla x_{x_1,x_2} R_i^{(1)} \ast f \cdot G \bigg) dx_1 dx_2 dt_1 dt_2
\]

\[
\ldots
\]

\[
= C_{9,1} + C_{9,2} + \ldots + C_{9,25},
\]

where we get all these terms from the equality $\partial^2_{x_1} \partial^2_{x_2} (\ldots)$ by integration by parts and taking all the gradients or partial derivatives with respect to $x_1, x_2$ to the function $B$. By applying Proposition 3.3 to all these terms, we obtain that they are all controlled by

\[
C' \|b\|_{BMO_e(\mathbb{R}^n \times \mathbb{R}^m)} \|g\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m) \|f\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m)}.
\]

Next we consider the term $C_5$, which can be consider as a cross term in between $C_1$ and $C_9$. To continue, we write

\[
\partial^2_{x_2} \bigg( F \cdot R_i^{(1)} \bigg) \ast_2 g + R_i^{(1)} \ast f \cdot R_i^{(2)} \ast_2 g + R_i^{(2)} \ast_2 f \cdot R_i^{(1)} \ast f \cdot G
\]

\[
= \partial^2_{x_2} \bigg( F \cdot R_j^{(2)} \ast_2 (R_i^{(1)} \ast g) + R_j^{(2)} \ast_2 f \cdot R_i^{(1)} \ast g
\]

\[
+ \partial^2_{x_2} \bigg( R_i^{(1)} \ast f \cdot R_j^{(2)} \ast_2 g + R_j^{(2)} \ast_2 (R_i^{(1)} \ast f) \cdot G
\]

\[
= E_1 + E_2.
\]

For the term $E_1$, we write

\[
E_1 = -2 \partial x_{x_2,j} \partial x_{x_2} \bigg( F \cdot R_i^{(1)} \ast g \bigg) + \nabla x_{x_2} \bigg( \nabla x_{x_2} R_j^{(2)} \ast_2 f \cdot R_i^{(1)} \ast g + F \cdot \nabla x_{x_2} R_j^{(2)} \ast_2 R_i^{(1)} \ast g
\]

\[
- \nabla x_{x_2} F \cdot R_j^{(2)} \ast_2 R_i^{(1)} \ast g - R_j^{(2)} \ast_2 f \cdot \nabla x_{x_2} R_i^{(1)} \ast g \bigg).
\]

For the term $E_2$, we write

\[
E_2 = -2 \partial x_{x_2,j} \partial x_{x_2} \bigg( R_i^{(1)} \ast f \cdot G \bigg) + \nabla x_{x_2} \bigg( \nabla x_{x_2} R_j^{(2)} \ast_2 R_i^{(1)} \ast f \cdot G + R_i^{(1)} \ast f \cdot \nabla x_{x_2} R_j^{(2)} \ast_2 g
\]

\[
- \nabla x_{x_2} R_i^{(1)} \ast f \cdot R_j^{(2)} \ast_2 g - R_j^{(2)} \ast_2 f \cdot \nabla x_{x_2} G \bigg).
\]
As a consequence, by substituting the above 10 terms in the right-hand side of the equalities $E_1$ and $E_2$ back in to the term $C_5$, we have that

$$C_5 = 2 \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \partial_{x_2} B \cdot \partial_{t_2} \partial_{t_2} \left( F \cdot R_i^{(1)} * g \right) dx_1 dx_2 dt_1 dt_2$$

$$- \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( \nabla_{x_2} R_j^{(2)} * 2 \cdot R_i^{(1)} * g \right) dx_1 dx_2 dt_1 dt_2$$

$$- \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( F \cdot \nabla_{x_2} R_j^{(2)} * 2 \cdot R_i^{(1)} * g \right) dx_1 dx_2 dt_1 dt_2$$

$$+ \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( \nabla_{x_2} F \cdot R_j^{(2)} * 2 \cdot R_i^{(1)} * g \right) dx_1 dx_2 dt_1 dt_2$$

$$+ \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( R_j^{(2)} * 2 \cdot \nabla_{x_2} R_i^{(1)} * g \right) dx_1 dx_2 dt_1 dt_2$$

$$+ 2 \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \partial_{x_2, \alpha} B \cdot \partial_{t_2} \partial_{t_2} \left( R_i^{(1)} * f \cdot G \right) dx_1 dx_2 dt_1 dt_2$$

$$- \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( \nabla_{x_2} R_j^{(2)} * 2 \cdot R_i^{(1)} * f \cdot G \right) dx_1 dx_2 dt_1 dt_2$$

$$- \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( R_i^{(1)} * f \cdot \nabla_{x_2} R_j^{(2)} * 2 g \right) dx_1 dx_2 dt_1 dt_2$$

$$+ \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( \nabla_{x_2} R_i^{(1)} * f \cdot R_j^{(2)} * 2 g \right) dx_1 dx_2 dt_1 dt_2$$

$$+ \int_{\mathbb{R}^{n+1} \times \mathbb{R}^{m+1}} t_1 \partial_{t_1} t_2 \nabla_{x_2} B \cdot \partial_{t_1} \left( R_j^{(2)} * 2 f \cdot \nabla_{x_2} G \right) dx_1 dx_2 dt_1 dt_2$$

$$=: C_{5,1} + \cdots + C_{5,10}.$$
Then by applying Proposition 3.5 to these terms, we obtain that they are all controlled by
\[ C \| b \|_{BMO(R^n \times R^m)} \| g \|_{L^p(R^n \times R^m)} \| f \|_{L^{p'}(R^n \times R^m)}. \]

The estimates for the term \( C_8 \) can be handled symmetrically, and we get the same upper bound for \( C_7 \) above.

4 Upper bound of the big commutator \([b, R_{j,k}]\)

We derive a general upper bound result for commutators of any flag singular integral. The proof is based on the \( A_{F,p} \) weighted estimate of flag singular integral operators and a Cauchy integral trick that goes back to the work of Coifman, Rochberg, and Weiss [7]. Roughly speaking, this technique allows one to bootstrap the weighted estimate for an arbitrary linear operator to that of its commutators of any order. This is the first time this idea is explored in the multi-parameter flag setting. In fact, although not needed for our upper bound proof, we demonstrate the bootstrapping result in the general higher order, two-weight setting.

4.1 \( A_p \) weight and little \( bmo \) in the flag setting

To begin with, we define the Muckenhoupt \( A_p \) weights in the flag setting, which consists of positive, locally integrable functions \( w \) satisfying

\[ [w]_{A_{F,p}} := \sup_{R \subset \mathbb{R}_x} \left( \frac{1}{|R|} \int_{R} w(x,y) \, dxdy \right) \left( \frac{1}{|R|} \int_{R} w(x,y)^{1-p'} \, dxdy \right)^{p-1} < \infty, \quad 1 < p < \infty, \]

where \( p' \) denotes the Hölder conjugate of \( p \). The following result of Wu [40] provides a way of approaching the \( A_{F,p} \) weights via the classical weights:

\[ A_{F,p} = A_p \cap A_p^{(2)}, \quad \forall 1 < p < \infty, \]

where \( A_p \) is the classical Muckenhoupt \( A_p \) class of weights on \( \mathbb{R}^{n+m} \), and \( A_p^{(2)} \) consists of weights \( w(x,y) \) such that \( w(x,\cdot) \in A_p \) with uniformly bounded characteristics for a.e. fixed \( x \in \mathbb{R}^n \).

We first show that a similar relation holds true for \( bmo_F \), which will be a useful tool for us in the study of this space.

**Lemma 4.1.** Let \( BMO(\mathbb{R}^{n+m}) \) denote the classical John-Nirenberg \( BMO \) space on \( \mathbb{R}^{n+m} \), and \( BMO^{(2)}(\mathbb{R}^m) \) be the space consisting of functions \( f(x,y) \) such that \( f(x,\cdot) \in BMO(\mathbb{R}^m) \) for a.e. fixed \( x \in \mathbb{R}^n \) with uniformly bounded norm. There holds

\[ bmo_F(\mathbb{R}^{n+m}) = BMO(\mathbb{R}^{n+m}) \cap BMO^{(2)}(\mathbb{R}^m) \]

with comparable norms.

**Proof.** The inclusion

\[ bmo_F(\mathbb{R}^{n+m}) \subset BMO(\mathbb{R}^{n+m}) \cap BMO^{(2)}(\mathbb{R}^m) \]

can be easily verified. Indeed, the inclusion \( bmo_F(\mathbb{R}^{n+m}) \subset BMO(\mathbb{R}^{n+m}) \) is obvious from the definition. Now fix \( x \in \mathbb{R}^n \). For any cube \( J \subset \mathbb{R}^m \), one can find a sequence of cubes \( I_k \subset \mathbb{R}^n \)
such that \( \ell(I_k) \leq \ell(J) \) and \( I_k \) shrinks to the point \( \{x\} \) as \( k \to \infty \). The containment thus follows from the Lebesgue differentiation theorem.

The other inclusion ("\( \supset \)") of the lemma follows from Proposition 4.2 below, which establishes the exp-log connection between \( A_{F,p} \) weights and \( \text{bmo}_F(\mathbb{R}^{n+m}) \), similarly as in the one-parameter and the product setting.

**Proposition 4.2.** Suppose \( w \) is a weight and \( 1 < p < \infty \). We have

(i) if \( w \in A_{F,p} \), then \( \log w \in \text{bmo}_F(\mathbb{R}^{n+m}) \);

(ii) if \( \log w \in \text{bmo}_F(\mathbb{R}^{n+m}) \), then \( w^\eta \in A_{F,p} \) for sufficiently small \( \eta > 0 \).

**Proof.** One observes directly from the definition that

\[ A_{F,p} \subset A_{F,q}, \quad \forall 1 < p \leq q < \infty, \]

and

\[ w \in A_{F,p} \iff w^{1-p'} \in A_{F,p'}, \quad \forall 1 < p < \infty. \]

Therefore, it suffices to prove the case \( p = 2 \).

We first prove (i). Suppose \( w \in A_{F,2} \) and let \( \varphi = \log w \). Then, for any \( R \in \mathcal{R}_F \) the \( A_{F,2} \) condition implies that

\[ \left( \frac{1}{|R|} \int_R e^{\varphi(x,y)-\langle \varphi \rangle_R} \, dxy \right) \left( \frac{1}{|R|} \int_R e^{\langle \varphi \rangle_R - \varphi(x,y)} \, dxdy \right) \leq [w]_{A_{F,2}} < \infty. \]

By Jensen’s inequality we have each of the factors above is at least 1 and at most \([w]_{A_{F,2}}\). Therefore, the inequality below holds:

\[ \frac{1}{|R|} \int_R e^{\varphi(x,y)-\langle \varphi \rangle_R} \, dxdy \leq 2[w]_{A_{F,2}}, \]

which, using the trivial estimate \( t \leq e^t \), implies that

\[ \frac{1}{|R|} \int_R |\varphi(x,y)-\langle \varphi \rangle_R| \, dxdy \leq 2[w]_{A_{F,2}}. \]

Hence, \( \varphi \in \text{bmo}_F(\mathbb{R}^{n+m}) \).

We now prove (ii). Let \( \varphi = \log w \in \text{bmo}_F(\mathbb{R}^{n+m}) \), it follows from (4.1) that \( \varphi \in \text{BMO}(\mathbb{R}^{n+m}) \) and \( \varphi \in \text{BMO}^{(2)}(\mathbb{R}^m) \). According to the classical exp-log connection between BMO and \( A_2 \), there hold for sufficiently small \( \eta > 0 \) that

\[ e^{\eta \varphi(\cdot, \cdot)} \in A_2(\mathbb{R}^{n+m}) \]

and

\[ e^{\eta \varphi(x, \cdot)} \in A_2(\mathbb{R}^m) \quad \text{uniformly in } x \in \mathbb{R}^n. \]

Hence, (4.2) implies that \( e^{\eta \varphi} \in A_{F,2} \) for sufficiently small \( \eta > 0 \), which completes the proof. \( \Box \)
4.2 Upper bound of the commutator

Given an operator $T$, define its $k$-th order commutator as

$$C_k^b(T) := [b_k, [b_{k-1}, \ldots, [b_1, T] \cdots]],$$

where each $b_j$ is a function on $\mathbb{R}^n \times \mathbb{R}^m$, $\forall 1 \leq j \leq k$.

**Theorem 4.3.** Let $\nu$ be a fixed weight on $\mathbb{R}^n \times \mathbb{R}^m$, $1 < p < \infty$, and $T$ be a linear operator satisfying

$$\|T\|_{L^p(\mu) \to L^p(\lambda)} \leq C_{n,m,p,T} ([\mu]_{A_{F,p}}, [\lambda]_{A_{F,p}}),$$

where $C_{n,m,p,T}(\cdot, \cdot)$ is an increasing function of both components, with $\mu, \lambda \in A_{F,p}$ and $\mu/\lambda = \nu^p$. For $k \geq 1$, let $b_j \in bmo_F(\mathbb{R}^n \times \mathbb{R}^m)$, $1 \leq j \leq k$, then there holds

$$\|C_k^b(T)\|_{L^p(\mu) \to L^p(\lambda)} \leq C_{n,m,p,k,T} ([\mu]_{A_{F,p}}, [\lambda]_{A_{F,p}}) \prod_{j=1}^{k} \|b_j\|_{bmo_F}.$$

Assuming Theorem 4.3 in order to derive (even unweighted) upper estimate for commutator of operator $T$, it suffices to know the corresponding weighted estimate for $T$ itself. When $T$ is a flag singular integral operator (which includes the flag Riesz transform $R_{j,k}$), such a result is obtained by Han, Lin and Wu in [20].

**Definition 4.4.** A flag singular integral $T_F : f \mapsto K \ast f$ is defined via a flag kernel $K$ on $\mathbb{R}^n \times \mathbb{R}^m$, which is a distribution on $\mathbb{R}^{n+m}$ that coincides with a $C^\infty$ function away from the coordinate subspace $\{(0,y)\} \subset \mathbb{R}^{n+m}$ and satisfies

(i) **(differential inequalities)** For each $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\beta = (\beta_1, \ldots, \beta_n)$

$$|\partial_x^\alpha \partial_y^\beta K(x,y)| \lesssim |x|^{-n-|\alpha|}(|x|+|y|)^{-m-|\beta|}$$

for all $(x,y) \in \mathbb{R}^{n+m}$ with $|x| \neq 0$;

(ii) **(cancellation conditions)**

(a) $$\left| \int_{\mathbb{R}^m} \partial_x^\alpha K(x,y) \psi_1(\delta y) dy \right| \leq C_\alpha |x|^{-n-|\alpha|}$$

for every multi-index $\alpha$ and for every normalized bump function $\psi_1$ on $\mathbb{R}^m$ and every $\delta > 0$;

(b) $$\left| \int_{\mathbb{R}^n} \partial_y^\beta K(x,y) \psi_2(\delta y) dy \right| \leq C_\beta |y|^{-m-|\beta|}$$

for every multi-index $\beta$ and for every normalized bump function $\psi_2$ on $\mathbb{R}^n$ and every $\delta > 0$;

(c) $$\left| \int_{\mathbb{R}^{n+m}} K(x,y) \psi_3(\delta_1 x, \delta_2 y) dx dy \right| \leq C$$

for every normalized bump function $\psi_3$ on $\mathbb{R}^{n+m}$ and every $\delta_1, \delta_2 > 0$. 
Theorem 4.5 (Remark 1.4 of [20]). Let $1 < p < \infty$ and $w \in A_{F,p}(\mathbb{R}^{n+m})$, there holds

$$\|T_f(f)\|_{L^p_0(\mathbb{R}^{n+m})} \leq C_p \|f\|_{L^p_0(\mathbb{R}^{n+m})}, \quad \forall f \in L^p_w(\mathbb{R}^{n+m}).$$

Applying Theorem 4.3 (with the choice $\mu = \lambda = w$) together with Theorem 1.5 one obtains immediately the following.

**Corollary 4.6.** Let $w \in A_{F,p}, 1 < p < \infty$ and $T$ be a flag singular integral operator as defined above. For any $k \geq 1$, $\vec{b} = (b_1, \cdots, b_k)$ where $b_j \in bmo_F(\mathbb{R}^n \times \mathbb{R}^m), j = 1, \ldots, k$, there holds

$$\|C^k_{\vec{b}}(T)\|_{L^p(w) \to L^p(w)} \leq C_{n,m,p,k,w,T} \prod_{j=1}^k |b_j|^{bmo_F}.$$

Obviously, the result above in the first order unweighted case is precisely the desired upper bound estimate in Theorem 1.5.

The core of the proof of Theorem 4.3 lies in a complex function representation of the commutators and the Cauchy integral formula. This method has been widely used to obtain upper estimates for linear and multilinear commutators in various settings, see [5, 7, 24, 1, 26] for examples. The main new challenge in our problem is the unique structure of the little flag BMO space and flag weights, which for instance doesn’t seem to fall into the category of spaces recently studied in [1].

**Proof of Theorem 4.3.** Observe that

$$C^k_{\vec{b}}(T) = \partial z_1 \cdots \partial z_k F(\vec{0}), \quad F(z) := e^{\sum_{j=1}^k b_j z_j} \left( e^{-\sum_{j=1}^k b_j z_j} \right),$$

which generalizes a classical formula representing higher order commutators. We remark that when all the symbol functions $b_j$ are the same, one can work instead with a simpler formula using single variable complex functions and their $k$-th order derivatives. According to the Cauchy integral formula on polydiscs,

$$C^k_{\vec{b}}(T) = \frac{1}{(2\pi i)^k} \oint \cdots \oint \frac{F(\vec{z}) \, dz_1 \cdots dz_k}{z_1^{\delta_1} \cdots z_k^{\delta_k}},$$

where each integral is over any closed path around the origin in the corresponding variable. For fixed $(\delta_1, \ldots, \delta_k)$ which will be determined later, there holds by Minkowski inequality that

$$\|C^k_{\vec{b}}(T)\|_{L^p(\mu) \to L^p(\lambda)} \leq \frac{1}{(2\pi i)^k} \oint \cdots \oint \|T\|_{L^p(\mu)} \|e^{\sum_{j=1}^k b_j z_j} \|_{L^p(\lambda)} \, \frac{|z_1| \cdots |z_k|}{\delta_1^{\delta_1} \cdots \delta_k^{\delta_k}},$$

$$\leq \frac{1}{(2\pi i)^k} \oint \cdots \oint C_{n,m,p,T} \left[ \|e^{\sum_{j=1}^k b_j z_j} \|_{A_{F,p}} \right] A_{F,p} \left[ e^{\sum_{j=1}^k b_j z_j} \right] A_{F,p} \, \frac{|z_1| \cdots |z_k|}{\delta_1^{\delta_1} \cdots \delta_k^{\delta_k}}.$$

where we have used the fact that $(e^{\sum_{j=1}^k b_j z_j} \mu, e^{\sum_{j=1}^k b_j z_j} \lambda)$ is a pair of weights satisfying

$$\frac{e^{\sum_{j=1}^k b_j z_j} \mu}{e^{\sum_{j=1}^k b_j z_j} \lambda} = \frac{\mu}{\lambda} \left( \frac{\mu}{\lambda} \right)^p.$$
Now we choose \( \{ \delta_j \} \) according to Lemma 4.7 below, which is the key ingredient of the proof concerning the relation between \( A_{F,p} \) weights and little flag BMO functions. Let

\[
\delta_1 := \frac{\epsilon_{n,m,p}}{\max \left( (\mu)_{A_{F,p}}, (\lambda)_{A_{F,p}} \right) \| b_1 \|_{\text{bmo}_F}}.
\]

where for any \( w \in A_{F,p} \)

\[
(w)_{A_{F,p}} := \max \left( [w]_{A_{F,p}}, [\sigma]_{A_{F,p}} \right).
\]

Here we have used the notation \( \sigma := w^{1-p'} \) to denote the dual weight of \( w \), and the relevant property of \( (w)_{A_{F,p}} \) to us is that

\[
(w)_{A_{F,p}} = \max([w]_{A_{F,p}}, [w]^{p'-1}_{A_{F,p}}) = [w]^{\max(1, p'-1)}_{A_{F,p}}.
\]

Recursively, for any \( j \geq 2 \), choose

\[
\delta_j := \frac{\epsilon_{n,m,p}}{\sup_{\{ z_i \}}: |z_i|=\delta_i, \ldots, |z_{j-1}|=\delta_{j-1} \max \left( (e^{P \Re(\sum_{i=1}^{j-1} b_i z_i)} \mu)_{A_{F,p}}, (e^{P \Re(\sum_{i=1}^{j-1} b_i z_i)} \lambda)_{A_{F,p}} \right) \| b_j \|_{\text{bmo}_F}}.
\]

Then applying Lemma 4.7 iteratively shows that

\[
[e^{P \Re(\sum_{j=1}^{k} b_j z_j)} \mu]_{A_{F,p}} \leq C_{n,m,p} [e^{P \Re(\sum_{j=1}^{k-1} b_j z_j)} \mu]_{A_{F,p}} \leq \cdots \leq C_{n,m,p}^k [\mu]_{A_{F,p}},
\]

and similarly

\[
[e^{P \Re(\sum_{j=1}^{k} b_j z_j)} \lambda]_{A_{F,p}} \leq C_{n,m,p}^k [\lambda]_{A_{F,p}},
\]

which in turn via the monotonicity of \( C_{n,m,p,T}(\cdot, \cdot) \) leads to

\[
C_{n,m,p,T} \left( [e^{P \Re(\sum_{j=1}^{k} b_j z_j)} \mu]_{A_{F,p}}, [e^{P \Re(\sum_{j=1}^{k} b_j z_j)} \lambda]_{A_{F,p}} \right) \leq C_{n,m,p,k,T}^d ([\mu]_{A_{F,p}}, [\lambda]_{A_{F,p}}).
\]

Therefore,

\[
\| C_b^k(T) \|_{L^p(\mu) \to L^p(\lambda)} \leq \frac{1}{\delta_1 \cdots \delta_k} C_{n,m,p,k,T}^d ([\mu]_{A_{F,p}}, [\lambda]_{A_{F,p}}) \prod_{j=1}^k \| b_j \|_{\text{bmo}_F}.
\]

The proof is thus complete. \( \square \)

**Lemma 4.7.** Let \( w \in A_{F,p}, 1 < p < \infty \), and \( b \in \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m) \). There are constants \( \epsilon_{n,m,p}, C_{n,m,p} > 0 \) such that

\[
[e^{P \Re(b z)} w]_{A_{F,p}} \leq C_{n,m,p} [w]_{A_{F,p}}
\]

whenever \( z \in \mathbb{C} \) satisfies

\[
|z| \leq \frac{\epsilon_{n,m,p}}{\| b \|_{\text{bmo}_F} (w)_{A_{F,p}}},
\]

where \( (w)_{A_{F,p}} \) is defined as in (4.3).
Proof. This estimate is a consequence of \([4.2]\), Lemma 4.1 and a one-parameter version proven by Hytönen in [24], which states that for any \(w \in A_p\), the classical Muckenhoupt \(A_p\) class on \(\mathbb{R}^d\), \(1 < p < \infty\), there exist \(c_{d,p}, C_{d,p} > 0\) such that
\[
[e^{\Re(bz)}w]_{A_p} \leq C_{d,p}[w]_{A_p}
\]
for all \(z \in \mathbb{C}\) with
\[
|z| \leq \frac{\epsilon_{n,p}}{\|b\|_{\text{BMO}(w)} A_p}.
\]

To see this, by (4.2) and Lemma 4.1, given \(w \in A_{pRF, p} \cap A_{p}^{(2)}\) and \(b \in \text{BMO}(\mathbb{R}^{n+m}) \cap \text{BMO}(\mathbb{R}^{m})\). Hence, taking \(\epsilon_{n,m,p} > 0\) sufficiently small, for all \(z \in \mathbb{C}\) satisfying
\[
|z| \leq \frac{\epsilon_{n,m,p}}{\|b\|_{\text{BMO}(w)} A_{pRF, p}},
\]
one has
\[
[e^{\Re(bz)}w]_{A_p} \leq C_{n,m,p}[w]_{A_p} \leq C_{n,m,p}[w]_{A_{pRF, p}},
\]
and
\[
[e^{\Re(b(x,\cdot)z)}w(x, \cdot)]_{A_p} \leq C_{m,p}[w(x, \cdot)]_{A_p} \leq C_{n,m,p}[w]_{A_{pRF, p}}, \quad \text{a.e. } x \in \mathbb{R}^n,
\]
by observing that
\[
\|b\|_{\text{BMO}_p} \gtrsim \max (\|b\|_{\text{BMO}(\mathbb{R}^{n+m})}, \sup_{x \in \mathbb{R}^n} \|b(x, \cdot)\|_{\text{BMO}(\mathbb{R}^{m})})
\]
and that
\[
(w)_{A_{pRF, p}} \gtrsim \max([w]_{A_p}, \sup_{x \in \mathbb{R}^n} [w(x, \cdot)]_{A_p}).
\]
The proof is thus complete. \(\square\)

5 Applications: div-curl lemmas in the flag setting

Let \(E^{(1)}\) be a vector field on \(\mathbb{R}^{n+m}\) taking the values in \(\mathbb{R}^{n+m}\), and let \(E^{(2)}\) be a vector field on \(\mathbb{R}^m\) taking the values in \(\mathbb{R}^m\). Now let \(\mathcal{M}_{n+m,m}\) denote the set of all \((n+m) \times m\) matrices. We now consider the following version of vector fields on \(\mathbb{R}^n \times \mathbb{R}^m\) taking the values in \(\mathcal{M}_{n+m,m}\), associated with the flag structure:

\[
E = E^{(1)} \ast_2 E^{(2)} := \begin{bmatrix}
E_1^{(1)} \ast_2 E_1^{(2)} & \ldots & E_1^{(1)} \ast_2 E_m^{(2)} \\
\vdots & \ddots & \vdots \\
\vdots & \ddots & \vdots \\
E_{n+m}^{(1)} \ast_2 E_1^{(2)} & \ldots & E_{n+m}^{(1)} \ast_2 E_m^{(2)}
\end{bmatrix},
\]

(5.1)

where
\[
E_j^{(1)} \ast_2 E_k^{(2)}(x, y) = \int_{\mathbb{R}^m} E_j^{(1)}(x, y - z) E_k^{(2)}(z) \, dz.
\]

Next we consider the following \(L^p\) space via projections. Suppose \(1 < p < \infty\). We define \(L^p_F(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})\) to be the set of vector fields \(E\) in \(L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})\) such that there
exist \( r_1, r_2 \in (1, \infty) \) with \( \frac{1}{r_1} + \frac{1}{r_2} = \frac{1}{p} + 1 \), \( E^{(1)} \in L^{r_1}(\mathbb{R}^{n+m}; \mathbb{R}^{n+m}) \), \( E^{(2)} \in L^{r_2}(\mathbb{R}^m; \mathbb{R}^m) \) and that \( E = E^{(1)} \ast_2 E^{(2)} \), moreover,
\[
\| E \|_{L^p_x(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m})} := \inf \| E^{(1)} \|_{L^{r_1}(\mathbb{R}^{n+m}; \mathbb{R}^{n+m})} \| E^{(2)} \|_{L^{r_2}(\mathbb{R}^m; \mathbb{R}^m)},
\]
where the infimum is taken over all possible \( r_1, r_2 \in (1, \infty) \), \( E^{(1)} \in L^{r_1}(\mathbb{R}^{n+m}; \mathbb{R}^{n+m}) \), \( E^{(2)} \in L^{r_2}(\mathbb{R}^m; \mathbb{R}^m) \).

Given two matrices \( A, B \in \mathcal{M}_{n+m,m} \), we define the “dot product” between \( A \) and \( B \) by
\[
A \cdot B = \sum_{j=1}^{n+m} \sum_{k=1}^{m} A_{j,k} B_{j,k}.
\]

We point out that this is the Hilbert-Schmidt inner product for two matrices and more generally this is referred to as the Schur product of two matrices.

We now prove Theorem 1.6.

**Proof of Theorem 1.6** Note that \( B \) is a vector field on \( \mathbb{R}^n \times \mathbb{R}^m \) taking the values in \( \mathcal{M}_{n+m,m} \), associated with the flag structure \((5.1)\). Then there exist certain vector field \( B^{(1)} \) on \( \mathbb{R}^{n+m} \) taking the values in \( \mathbb{R}^{n+m} \) and vector field \( B^{(2)} \) on \( \mathbb{R}^m \) taking the values in \( \mathbb{R}^m \) such that \( B = B^{(1)} \ast_2 B^{(2)} \) and that
\[
\| B \|_{L^p_x(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})} \approx \inf \| B^{(1)} \|_{L^{r_1}(\mathbb{R}^{n+m}; \mathbb{R}^{n+m})} \| B^{(2)} \|_{L^{r_2}(\mathbb{R}^m; \mathbb{R}^m)}
\]
with \( \frac{1}{pq_1} + \frac{1}{pq_2} = \frac{1}{p} + 1 \).

Thus, \( \text{curl}_{(x,y)} B^{(1)} = 0 \) implies that there exists \( \phi^{(1)} \in L^q(\mathbb{R}^{n+m}) \) such that
\[
B^{(1)} = (R_1^{(1)} \phi^{(1)}, \ldots, R_{n+m}^{(1)} \phi^{(1)}),
\]
with \( \| B^{(1)} \|_{L^{r_1}(\mathbb{R}^{n+m}; \mathbb{R}^{n+m})} \approx \| \phi^{(1)} \|_{L^{r_1}(\mathbb{R}^{n+m})} \). Again, \( \text{curl}_y B^{(2)} = 0 \) implies that there exists \( \phi^{(2)} \in L^{q_2}(\mathbb{R}^{n+m}) \) such that
\[
B^{(2)} = (R_1^{(2)} \phi^{(2)}, \ldots, R_m^{(2)} \phi^{(2)}),
\]
with \( \| B^{(2)} \|_{L^{r_2}(\mathbb{R}^m; \mathbb{R}^m)} \approx \| \phi^{(2)} \|_{L^{r_2}(\mathbb{R}^m)} \). As a consequence we get that the matrix \( B \) has elements
\[
B_{j,k} = R_{j,k} \ast \phi, \quad j = 1, \ldots, n + m, \quad k = 1, \ldots, m,
\]
where \( \phi = \phi^{(1)} \ast_2 \phi^{(2)} \) and \( \| B \|_{L^p_x(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})} \approx \| \phi \|_{L^q(\mathbb{R}^{n+m})} \).

Similarly, note that \( E \) is a vector field on \( \mathbb{R}^n \times \mathbb{R}^m \) taking the values in \( \mathcal{M}_{n+m,m} \), associated with the flag structure \((5.1)\). Then there exist certain vector field \( E^{(1)} \) on \( \mathbb{R}^{n+m} \) taking the values in \( \mathbb{R}^{n+m} \) and vector field \( E^{(2)} \) on \( \mathbb{R}^m \) taking the values in \( \mathbb{R}^m \) such that \( E = E^{(1)} \ast_2 E^{(2)} \) and that
\[
\| E \|_{L^p_x(\mathbb{R}^n \times \mathbb{R}^m; \mathcal{M}_{n+m,m})} \approx \inf \| E^{(1)} \|_{L^{p_1}(\mathbb{R}^{n+m}; \mathbb{R}^{n+m})} \| E^{(2)} \|_{L^{p_2}(\mathbb{R}^m; \mathbb{R}^m)}
\]
with \( \frac{1}{p_1} + \frac{1}{p_2} = \frac{1}{p} + 1 \).

Thus, the conditions \( \text{div}_{(x,y)} E^{(1)} = 0 \) and \( \text{div}_y E^{(2)} = 0 \) imply that
\[
\sum_{j=1}^{n+m} R_j^{(1)} \ast E_j^{(1)}(x,y) = 0 \quad \text{and} \quad \sum_{k=1}^{m} R_k^{(2)} \ast E_k^{(2)}(y) = 0.
\]
Hence, we get that

\[ \sum_{j=1}^{n+m} R_j^{(1)} * E_{j,k}(x,y) = 0 \quad \text{and} \quad \sum_{k=1}^{m} R_k^{(2)} *_2 E_{j,k}(x,y) = 0. \]

With these facts, we have that

\[ E(x,y) \cdot B(x,y) = \sum_{j=1}^{n+m} \sum_{k=1}^{m} E_{j,k}(x,y) B_{j,k}(x,y) = \sum_{j=1}^{n+m} \sum_{k=1}^{m} E_{j,k}(x,y) R_{j,k} * \phi(x,y) \]

\[ = \sum_{j=1}^{n+m} \sum_{k=1}^{m} \left\{ E_{j,k}(x,y) R_{j,k} * \phi(x,y) + R_j^{(1)} * E_{j,k}(x,y) R_k^{(2)} *_2 \phi(x,y) \right\} \]

\[ + R_k^{(2)} *_2 E_{j,k}(x,y) R_j^{(1)} * \phi(x,y) + R_{j,k} * E_{j,k}(x,y) \phi(x,y) \right\} \]

Now testing this equality over all functions in the flag BMO space, i.e., for every \( b \in \text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m) \), and then unravelling the expression with the Riesz transforms we see that

\[ \int_{\mathbb{R}^n \times \mathbb{R}^m} E(x,y) \cdot B(x,y) b(x,y) \, dx \, dy \]

\[ = \sum_{j=1}^{n+m} \sum_{k=1}^{m} \int_{\mathbb{R}^n \times \mathbb{R}^m} \left[ [b, R_j^{(1)}], R_k^{(2)} \right]_2 (E_{j,k})(x,y) \phi(x,y) \, dx \, dy. \]

Then based on Theorem 1.3 since \( b \in \text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m) \) we have that each of the above commutators is a bounded operator on \( L^p(\mathbb{R}^n \times \mathbb{R}^m) \) with norm controlled by the norm of \( b \), i.e., \( \|b\|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)}. \)

As a consequence, we get that

\[ \left| \int_{\mathbb{R}^n \times \mathbb{R}^m} E(x,y) \cdot B(x,y) b(x,y) \, dx \, dy \right| \]

\[ \lesssim \|b\|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; M_{n+m,m})} \|\phi\|_{L^q(\mathbb{R}^{n+m})} \]

\[ \lesssim \|b\|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; M_{n+m,m})} \|B\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; M_{n+m,m})}. \]

Then from the duality of \( H^1_F(\mathbb{R}^n \times \mathbb{R}^m) \) with \( \text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m) \), we obtain that

\[ \|E \cdot B\|_{H^1_F(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim \|b\|_{\text{BMO}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; M_{n+m,m})} \|B\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; M_{n+m,m})}. \]

This finishes the proof of Theorem 1.6.

**Proof of Theorem 1.7.** Suppose that \( E, B \) are vector fields on \( \mathbb{R}^n \times \mathbb{R}^m \) taking values in \( \mathbb{R}^{n+m} \). Moreover, suppose \( E \in L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m}) \) and \( B \in L^2(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m}) \) satisfy that

\[ \text{div}_{(x,y)} E(x,y) = 0 \quad \text{and} \quad \text{curl}_{(x,y)} B(x,y) = 0 \]

and

\[ \text{div}_x E(x,y) = 0 \quad \text{and} \quad \text{curl}_y B(x,y) = 0, \quad \forall x \in \mathbb{R}^n. \]
We now define the projection operator $\mathcal{P}$ as

$$\mathcal{P}E = \left( E_1 + R_1^{(1)} \left( \sum_{k=1}^{n+m} R_k^{(1)} E_k \right), \ldots, E_{n+m} + R_{n+m}^{(1)} \left( \sum_{k=1}^{n+m} R_k^{(1)} E_k \right) \right).$$

Then by definition, it is direct that

$$\text{div}_{x,y} \mathcal{P}E = 0$$

since

$$\sum_{j=1}^{n+m} R_j^{(1)} \left( E_j + R_j^{(1)} \left( \sum_{k=1}^{n+m} R_k^{(1)} E_k \right) \right) = 0. \quad (5.2)$$

Moreover, we also have $\mathcal{P} \circ \mathcal{P}E = \mathcal{P}E$. Next, we point out that applying $[b, \mathcal{P}]$ to the vector field $E$, we can get that the $j$th component is given by

$$\sum_{k=1}^{n+m} [b, R_j^{(1)} R_k^{(1)}](E_k).$$

Suppose now $b \in \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m)$. Then from Lemma 4.1 we know that

$$\text{bmo}_F(\mathbb{R}^{n+m}) = \text{BMO}(\mathbb{R}^{n+m}) \cap \text{BMO}^{(2)}(\mathbb{R}^m)$$

with comparable norms. Hence, we have that $b \in \text{BMO}(\mathbb{R}^{n+m})$ with

$$\|b\|_{\text{BMO}(\mathbb{R}^{n+m})} \lesssim \|b\|_{\text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m)}.$$
Then, again, by definition, we have that
\[ \text{div}_y \mathcal{P}^{(2)} E = 0 \]
since
\[ \sum_{j=1}^m R_j^{(2)} \left( E_{n+j} + R_j^{(2)} \left( \sum_{k=1}^m R_k^{(2)} E_{n+k} \right) \right) = 0. \] (5.3)

Now fix \( x \in \mathbb{R}^n \), by using the definition of \( \mathcal{P}^{(2)} \) and the fact (5.3), we get that for \( b \in \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m) \),
\[ \int_{\mathbb{R}^m} E(x, y) \cdot_2 B(x, y) b(x, y) \, dy = \int_{\mathbb{R}^m} \left[ b(x, \cdot), \mathcal{P}^{(2)} \right] E(x, y) \psi(x, y) \, dy. \]
Integrating the above equality over \( \mathbb{R}^n \), we have
\[
\left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} E(x, y) \cdot_2 B(x, y) b(x, y) \, dy \, dx \right|
\leq \int_{\mathbb{R}^n} \|b(x, \cdot)\|_{\text{BMO}(\mathbb{R}^n)} \|E(x, \cdot)\|_{L^p(\mathbb{R}^m)} \|B(x, \cdot)\|_{L^q(\mathbb{R}^m)} \, dx
\leq \|b\|_{\text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m)} \int_{\mathbb{R}^n} \|E(x, \cdot)\|_{L^p(\mathbb{R}^m)} \|B(x, \cdot)\|_{L^q(\mathbb{R}^m)} \, dx
\leq \|b\|_{\text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m)} \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})} \|B\|_{L^q(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})}.
\]
Here we use again Lemma 4.1 and Hölder’s inequality. Taking the supremum over all \( b \in \text{bmo}_F(\mathbb{R}^n \times \mathbb{R}^m) \) we obtain that
\[ \int_{\mathbb{R}^m} \|E(\cdot, y) \cdot_2 B(\cdot, y)\|_{H^1(\mathbb{R}^m)} \, dy \leq \|E\|_{L^p(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})} \|B\|_{L^q(\mathbb{R}^n \times \mathbb{R}^m; \mathbb{R}^{n+m})}. \]
This finishes the proof of Theorem 1.7.

**Acknowledgments:** X. T. Duong, J. Li and J. Pipher are supported by ARC DP 160100153. X. T. Duong and J. Li are also supported by Macquarie University Research Seeding Grant. B. D. Wick’s research is supported in part by National Science Foundation grant DMS #1560955. This paper started in July 2016 during J. Li’s visiting J. Pipher at Brown University. J. Li would like to thank the Department of Mathematics of Brown University for its hospitality. Part of this material is based upon work supported by the National Science Foundation under Grant No. DMS-1440140 while Y. Ou and J. Pipher were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2017 semester.

**References**

[1] A. Bényi, J.M. Martell, K. Moen, E. Stachura and R.H. Torres, Boundedness results for commutators with BMO functions via weighted estimates: A comprehensive approach, arXiv: 1710.08515.
[2] S-Y. A. Chang and R. Fefferman, Some recent developments in Fourier analysis and $H^p$ theory on product domains, Bull. A. M. S. **12** (1985), 1–43.

[3] S-Y. A. Chang and R. Fefferman, The Calderón-Zygmund decomposition on product domains, Amer. J. Math. **104** (1982), 455–468.

[4] S-Y. A. Chang and R. Fefferman, A continuous version of duality of $H^1$ with $BMO$ on the bidisc, Annals of math. **112** (1980), 179–201.

[5] D. Chung, M.C. Pereyra, and C. Pérez, Sharp bounds for general commutators on weighted Lebesgue spaces, Trans. Amer. Math. Soc. **364(3)** (2012), 1163–1177.

[6] R. Coifman, P.L. Lions, Y. Meyer and S. Semmes, Compensated compactness and Hardy spaces, *J. Math. Pures Appl.*, **72** (1993), 247–286.

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

[8] L. Dalenc and Y. Ou, Upper bound for multi-parameter iterated commutators, *Publ. Mat.* **60** (2016), 191–220.

[9] C. Fefferman and E. M. Stein, $H^p$ spaces of several variables, Acta Math. **129** (1972), 137–194.

[10] R. Fefferman, Multi-parameter Fourier analysis, Study **112**, Beijing Lectures in Harmonic Analysis, Edited by E. M. Stein, 47-130. Annals of Mathematics Studies Princeton University Press.

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

[12] R. Fefferman, Multiparameter Calderón-Zygmund theory, Harmonic analysis and partial differential equations (Chicago, IL, 1996), 207-221, Chicago Lectures in Math., Univ. Chicago Press, Chicago, IL, 1999.

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

[14] S.H. Fergusson and C. Sadosky, Characterizations of bounded mean oscillation on the polydisk in terms of Hankel operators and Carleson measures, *J. Anal. Math.* **81** (2000), 239–267.

[15] S. H. Ferguson and M. T. Lacey, A characterization of product BMO by commutators, *Acta Math.*, **189** (2002), 143–160.

[16] P. Glowacki, Composition and $L^2$-boundedness of flag kernels, Colloq. Math. **118** (2010), 581–585.

[17] P. Glowacki, Correction to “Composition and $L^2$-boundedness of flag kernels”, Colloq. Math. **120** (2010), 331.

[18] R. Gundy and E. M. Stein, $H^p$ theory for the polydisk, Proc. Nat. Acad. Sci., **76** (1979), 1979.

[19] Y.S. Han, M.Y. Lee, J. Li and B.D. Wick, Characterizations of flag Hardy space via Riesz transforms, maximal functions and Littlewood–Paley theory, arXiv:1611.05296.

[20] Y.S. Han, C.C. Lin and X.F. Wu, Boundedness of singular integrals with flag kernels on weighted flag Hardy spaces, submitted.

[21] Y.S. Han and G. Lu, Discrete Littlewood-Paley-Stein theory and multi-parameter Hardy spaces associated with flag singular integrals, arXiv:0801.1701.

[22] Y.S. Han, G. Lu and E. Sawyer, Flag Hardy spaces and Marcinkiewicz multipliers on the Heisenberg group: an expanded version, Anal. PDE, 7 (2014) 1465–1534.

[23] I. Homes, S. Petermichl and B.D. Wick, Weighted little bmo and two-weight inequalities for Journé commutators, arXiv:1701.06526.

[24] T.P. Hytönen, Holmes–Wick theorem on two-weight bounds for higher order commutators revisited, Arch. Math. 107 (2016) 389–395.
Flag commutators

[25] J. L. Journé, Calderón-Zygmund operators on product spaces, Rev. Mat. Iberoamericana 1 (1985), 55–92.
[26] I. Kunwar and Y. Ou, Two-weight inequalities for multilinear commutators, arXiv:1710.07392.
[27] M. Lacey, S. Petermichl, J. Pipher and B. D. Wick, Multiparameter Riesz commutators, Amer. J. Math., 131 (2009), 731–769.
[28] M. Lacey, S. Petermichl, J. Pipher and B. D. Wick, Iterated Riesz commutators: a simple proof of boundedness, Harmonic analysis and partial differential equations, Contemp. Math., 505 (2010), 171–178.
[29] M. Lacey, S. Petermichl, J. Pipher and B. D. Wick, Multi-parameter Div-Curl lemmas, Bull. Lond. Math. Soc., 44 (2012), no. 6, 1123–1131.
[30] M. Lacey and E. Terwilleger, Hankel operators in several complex variables and product BMO, Houston J. Math., 35 (2009), 159–183.
[31] E. Lenzmann and A. Schikorra, Sharp commutator estimates via harmonic extensions, arXiv:1609.08547.
[32] D. Müller, F. Ricci, and E. M. Stein, Marcinkiewicz multipliers and multi-parameter structure on Heisenberg(-type) groups, I, Invent. math. 119 (1995), 119–233.
[33] D. Müller, F. Ricci, and E. M. Stein, Marcinkiewicz multipliers and multi-parameter structure on Heisenberg(-type) groups, II, Math. Z. 221 (1996), 267–291.
[34] A. Nagel and E. M. Stein, On the product theory of singular integrals, Rev. Mat. Iberoamericana 20 (2004), 531–561.
[35] A. Nagel, F. Ricci, and E. M. Stein, Singular integrals with flag kernels and analysis on quadratic CR manifolds, J. Func. Anal. 181 (2001), 29–118.
[36] A. Nagel, F. Ricci, E. M. Stein, and S. Wainger, Singular integrals with flag kernels on homogeneous groups: I, Rev. Mat. Iberoamericana 28 (2012), 631–722.
[37] A. Nagel, F. Ricci, E. M. Stein, and S. Wainger, Algebras of singular integral operators with kernels controlled by multiple norms, arxiv:1511.05702.
[38] J. Pipher, Journé’s covering lemma and its extension to higher dimensions, Duke Mathematical Journal, 53 (1986), 683–690.
[39] E.M. Stein, Harmonic Analysis: Real-variable Methods, Orthogonality, and Oscillatory Integrals. Princeton, NJ: Princeton University Press, 1993
[40] X.F. Wu, Weighted norm inequalities for flag singular integrals on homogeneous groups, Taiwanese Journal of Mathematics, 18 (2014), 357–369.

X. Duong, Department of Mathematics, Macquarie University, NSW, 2109, Australia.
E-mail: xuan.duong@mq.edu.au

J. Li, Department of Mathematics, Macquarie University, NSW, 2109, Australia.
E-mail: ji.li@mq.edu.au

Y. Ou, Department of Mathematics, Baruch College, CUNY, New York, NY 10025, USA
E-mail: yumeng.ou@baruch.cuny.edu

J. Pipher, Department of Mathematics, Brown University, Providence, RI 02912, USA
E-mail: jpipher@math.brown.edu

B. Wick, Department of Mathematics, Washington University–St. Louis, St. Louis, MO 63130-4899 USA
E-mail: wick@math.wustl.edu