BILINEAR REPRESENTATION THEOREM

KANGWEI LI, HENRI MARTIKAINEN, YUMENG OU, AND EMIL VUORINEN

ABSTRACT. We represent a general bilinear Calderón–Zygmund operator as a sum of simple dyadic operators. The appearing dyadic operators also admit a simple proof of a sparse bound. In particular, the representation implies a so called sparse $T1$ theorem for bilinear singular integrals.

1. INTRODUCTION

In this paper we show the exact dyadic structure behind *bilinear* Calderón–Zygmund operators by representing them using simple dyadic operators, namely some cancellative bilinear shifts and bilinear paraproducts. In the linear case Petermichl [14] first represented the Hilbert transform in this way, and later Hytönen [4] proved a representation theorem for all linear Calderón–Zygmund operators.

The representation theorems were originally motivated by the sharp weighted A_p theory, but certainly also have other value and intrinsic interest. For example, a representation theorem holds also in the bi-parameter setting as shown by one of us [10] (the multi-parameter extension of this is also by one of us [12]), and in this context the representation has proved to be very useful e.g. in connection with bi-parameter commutators, see [3] and [13].

Outside the multi-parameter context it is true that sparse domination results yield sharp weighted bounds, and that sparse domination can also be proved directly (without going through a representation). Such proofs usually start from the unweighted boundedness assumption, then conclude some weak type estimates, and then finally go about proving the sparse domination. However, we think that the idea of a so called sparse $T1$, as coined by Lacey–Mena [8], is extremely practical. This amounts to concluding a sparse bound directly from the $T1$ assumptions (by modifying the probabilistic $T1$ proof), and then noting that the sparse bound implies all the standard boundedness properties (even weak type). Such a combination gives everything in one blow.

We think that a very efficient way to go about things is to first prove a sharp form of a representation theorem working directly from the $T1$ assumptions. This is interesting on its own right, entails $T1$, gives an explicit equality containing the full dyadic structure of the operator, and can even be used to transfer sparse

---

2010 *Mathematics Subject Classification.* 42B20.

Key words and phrases. Dyadic analysis, Calderón–Zygmund operators, model operators, dyadic shifts, bilinear analysis, representation theorems, $T1$ theorems, weighted theory.

bounds, at least in the form sense, from the model dyadic operators to the singular integral. This strategy was employed in the linear setting by Culiuc, Di Plinio and one of us in [2], but of course they were able to cite the linear representation theorem with $T1$ assumptions from previous literature [5]. It is also to be noted that sparse bounds are remarkably simple to prove for dyadic model operators using the method of [2].

In this paper we, for the first time, prove a representation theorem in the bilinear setting, and we do it starting from the bilinear $T1$ assumptions. Moreover, we carry out the above strategy in the bilinear setting i.e. we prove sparse domination for our model operators and then transfer them back to the singular integral. In particular, we get a sparse bilinear $T1$ implying directly the boundedness of singular integrals from $L^p \times L^q$ to L^r for all $1 < p, q < \infty$ and $1/2 < r < \infty$ satisfying $1/p + 1/q = 1/r$, and even the boundedness from $L^1 \times L^1$ to $L^{1/2,\infty}$, just from the $T1$ assumptions. Of course, one can also recover known sharp weighted bounds (see e.g. [7]) from sparse domination. It is to be noted though that we prove sparse domination in the trilinear form sense, as such bounds are easy to transfer using the representation. A caveat regarding weighted bounds is that outside the Banach range the literature currently seems to lack an argument giving *sharp* weighted bounds from form type domination (but such bounds can be derived using pointwise sparse domination [1], [9]).

The proof of the representation entails finding a dyadic–probabilistic proof technique which produces only simple model operators. Some bilinear dyadic–probabilistic methods were studied by two of us in [11] in the non-homogeneous setting. However, there seems to be a plethora of possible ways to decompose things in the bilinear setting, and one has to be quite careful to really get only nice shifts and nice paraproducts (such that can easily be seen to obey sparse domination). We now move on to formulating some basic definitions and stating our theorems.

A function

$$K\colon (\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n) \setminus \Delta \to \mathbb{C}, \qquad \Delta := \{(x,y,z) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \colon\ x = y = z\},$$

is called a standard bilinear Calderón–Zygmund kernel if for some $\alpha \in (0,1]$ and $C_K < \infty$ it holds that

$$|K(x,y,z)| \le \frac{C_K}{(|x-y| + |x-z|)^{2n}},$$

$$|K(x,y,z) - K(x',y,z)| \le C_K \frac{|x-x'|^\alpha}{(|x-y| + |x-z|)^{2n+\alpha}}$$

whenever $|x-x'| \le \max(|x-y|, |x-z|)/2$,

$$|K(x,y,z) - K(x,y',z)| \le C_K \frac{|y-y'|^\alpha}{(|x-y| + |x-z|)^{2n+\alpha}}$$

whenever $|y-y'| \le \max(|x-y|,|x-z|)/2$, and

$$|K(x,y,z)-K(x,y,z')| \le C_K \frac{|z-z'|^\alpha}{(|x-y|+|x-z|)^{2n+\alpha}}$$

whenever $|z-z'| \le \max(|x-y|,|x-z|)/2$. The best constant C_K is denoted by $\|K\|_{\mathrm{CZ}_\alpha}$.

Given a standard bilinear Calderón–Zygmund kernel K we define

$$T_\varepsilon(f,g)(x) = \iint_{\max(|x-y|,|x-z|)>\varepsilon} K(x,y,z)f(y)g(z)\,dy\,dz.$$

The above is well-defined as an absolutely convergent integral if e.g. $f \in L^{p_1}$ and $g \in L^{p_2}$ for some $p_1, p_2 \in [1,\infty)$, since then

$$\iint_{\max(|x-y|,|x-z|)>\varepsilon} |K(x,y,z)f(y)g(z)|\,dy\,dz \lesssim \frac{1}{\varepsilon^{n(1/p_1+1/p_2)}}\|f\|_{L^{p_1}}\|g\|_{L^{p_2}}.$$

For us a bilinear Calderón–Zygmund operator is essentially the family of truncations $(T_\varepsilon)_{\varepsilon>0}$. In particular, this means that boundedness in some L^p spaces is understood in the sense that all T_ε are bounded uniformly in $\varepsilon > 0$.

We shall also define some smoother truncations. Suppose $\varphi \in \mathcal{A}$, where $\mathcal{A}$ consists of smooth functions $\varphi\colon [0,\infty) \to [0,1]$ satisfying that $\varphi = 0$ on $[0,1/2]$, $\varphi = 1$ on $[1,\infty)$ and $\|\varphi'\|_{L^\infty} \le 10$. Define the smoothly truncated singular integrals

$$T_\varepsilon^\varphi(f,g)(x) = \iint K_\varepsilon^\varphi(x,y,z)f(y)g(z)\,dy\,dz, \qquad \varepsilon > 0,$$

where

$$K_\varepsilon^\varphi(x,y,z) = K(x,y,z)\varphi\Big(\frac{|x-y|+|x-z|}{\varepsilon}\Big).$$

The point is that T_ε^φ, $\varepsilon > 0$, are operators with standard bilinear n-dimensional kernels (with the kernel bounds being independent of ε). Moreover, we have

$$|T_\varepsilon(f,g)(x) - T_\varepsilon^\varphi(f,g)(x)| \lesssim \mathcal{M}(f,g)(x) := \sup_{r>0} \big\langle |f| \big\rangle_{B(x,r)} \big\langle |g| \big\rangle_{B(x,r)},$$

where $\big\langle f \big\rangle_A := \frac{1}{|A|}\int_A f$. If $0 < \varepsilon_1 < \varepsilon_2$ we denote by $T^\varphi_{\varepsilon_1,\varepsilon_2}$ the operator

$$\begin{aligned} T^\varphi_{\varepsilon_1,\varepsilon_2}(f,g)(x) &= T^\varphi_{\varepsilon_1}(f,g)(x) - T^\varphi_{\varepsilon_2}(f,g)(x) \\ &= \iint K^\varphi_{\varepsilon_1,\varepsilon_2}(x,y,z)f(y)g(z)\,dy\,dz, \end{aligned}$$

where $K^\varphi_{\varepsilon_1,\varepsilon_2} = K^\varphi_{\varepsilon_1} - K^\varphi_{\varepsilon_2}$.

The notation T^{1*} and T^{2*} stand for the adjoints of a bilinear operator T, i.e.

$$\langle T(f,g), h\rangle = \langle T^{1*}(h,g), f\rangle = \langle T^{2*}(f,h), g\rangle.$$

We can now state our main theorem. For the exact definitions of the various objects and notions (random dyadic grids, bilinear cancellative shifts, bilinear paraproducts, weak boundedness, $T_\delta(1,1)$, sparse collections etc.) see the following two sections.

1.1. **Theorem.** *Let K be a bilinear Calderón–Zygmund kernel so that $\|K\|_{\mathrm{CZ}_\alpha} < \infty$, and let $(T_\varepsilon)_{\varepsilon>0}$ be the corresponding bilinear singular integral. Assume that*

$$\sup_{\delta>0}[\|T_\delta\|_{\mathrm{WBP}} + \|T_\delta(1,1)\|_{\mathrm{BMO}} + \|T_\delta^{1*}(1,1)\|_{\mathrm{BMO}} + \|T_\delta^{2*}(1,1)\|_{\mathrm{BMO}}] < \infty.$$

Let also $\varphi \in \mathcal{A}$. Then there is a constant $C = C(n,\alpha) < \infty$ so that for all $\varepsilon > 0$ and all compactly supported and bounded functions f, g and h it holds that

$$\begin{aligned}\langle T_\varepsilon^\varphi(f,g),h\rangle =& C(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0}\|T_\delta\|_{\mathrm{WBP}})E_\omega\sum_{k=0}^\infty\sum_{i=0}^k 2^{-\alpha k/2}\langle U^{i,k}_{\varepsilon,\varphi,\omega}(f,g),h\rangle \\ &+ C(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0}\|T_\delta(1,1)\|_{\mathrm{BMO}})E_\omega\langle \Pi_{\alpha_0(\varepsilon,\varphi,\omega)}(f,g),h\rangle \\ &+ C(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0}\|T_\delta^{1*}(1,1)\|_{\mathrm{BMO}})E_\omega\langle \Pi^{1*}_{\alpha_1(\varepsilon,\varphi,\omega)}(f,g),h\rangle \\ &+ C(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0}\|T_\delta^{2*}(1,1)\|_{\mathrm{BMO}})E_\omega\langle \Pi^{2*}_{\alpha_2(\varepsilon,\varphi,\omega)}(f,g),h\rangle,\end{aligned}$$

where each $U^{i,k}_{\varepsilon,\varphi,\omega}$ is a sum of cancellative bilinear shifts $S^{i,i,k}_{\varepsilon,\varphi,\omega}$, $S^{i,i+1,k}_{\varepsilon,\varphi,\omega}$ and adjoints of such operators, and Π_α stands for a bilinear paraproduct with α as in (3.1)*. For a fixed ω the operators above are defined using the dyadic lattice $\mathcal{D}_\omega$.*

The following corollary follows from the sparse domination of shifts and paraproducts (see Section 5), and the trivial sparse bound for $\mathcal{M}$.

1.2. **Corollary.** *There exist dyadic grids $\mathcal{D}_i$, $i = 1,\ldots,3^n$, with the following property. Let $\eta \in (0,1)$. For compactly supported and bounded functions f, g and h there is a dyadic grid $\mathcal{D}_i$ and an η-sparse collection $\mathcal{S} = \mathcal{S}(f,g,h,\eta) \subset \mathcal{D}_i$ so that the following holds.*

Let K be any standard bilinear Calderón–Zygmund kernel and $(T_\varepsilon)_{\varepsilon>0}$ be the corresponding bilinear singular integral. Then we have

$$\sup_{\varepsilon>0}|\langle T_\varepsilon(f,g),h\rangle| \le C_{T,K}\Lambda_{\mathcal{S}}(f,g,h),$$

where

$$\begin{aligned}C_{T,K} := C(\|K\|_{\mathrm{CZ}_\alpha} &+ \sup_{\varepsilon>0}\|T_\varepsilon(1,1)\|_{\mathrm{BMO}} + \sup_{\varepsilon>0}\|T_\varepsilon^{1*}(1,1)\|_{\mathrm{BMO}} \\ &+ \sup_{\varepsilon>0}\|T_\varepsilon^{2*}(1,1)\|_{\mathrm{BMO}} + \sup_{\varepsilon>0}\|T_\varepsilon\|_{\mathrm{WBP}})\end{aligned}$$

for some $C = C(n,\alpha,\eta) < \infty$ and

$$\Lambda_{\mathcal{S}}(f,g,h) := \sum_{Q\in\mathcal{S}}|Q|\langle |f|\rangle_Q\langle |g|\rangle_Q\langle |h|\rangle_Q.$$

Additional notation. We write $A \lesssim B$, if there is an absolute constant $C > 0$ (depending only on some fixed constants like n and α etc.) so that $A \le CB$. Moreover, $A \lesssim_\tau B$ means that the constant C can also depend on some relevant given parameter $\tau > 0$. We may also write $A \sim B$ if $B \lesssim A \lesssim B$.

We then define some notation related to cubes. If Q and R are two cubes we set:

- $\ell(Q)$ is the side-length of Q;
- If $a > 0$, we denote by aQ the cube that is concentric with Q and has sidelength $a\ell(Q)$;
- $d(Q, R) = \operatorname{dist}(Q, R)$ denotes the distance between the cubes Q and R;
- $\operatorname{ch}(Q)$ denotes the dyadic children of Q;
- If Q is in a dyadic grid, then $Q^{(k)}$ denotes the unique dyadic cube S in the same grid so that $Q \subset S$ and $\ell(S) = 2^k \ell(Q)$;
- If $\mathcal{D}$ is a dyadic grid, then $\mathcal{D}_k = \{Q \in \mathcal{D}\colon\ \ell(Q) = 2^{-k}\}$;

The notation $\langle f, g\rangle$ stands for the pairing $\int fg$.

The following maximal functions are also used:

$$\begin{aligned} M_{\mathcal{D}} f(x) &= \sup_{Q\in\mathcal{D}} 1_Q(x)\langle |f|\rangle_Q \qquad (\mathcal{D} \text{ is a dyadic grid}); \\ Mf(x) &= \sup_{r>0} \langle |f|\rangle_{B(x,r)}. \end{aligned}$$

Here $B(x, r) = \{y\colon\ |x - y| < r\}$. The bilinear variants are defined in the natural way, e.g.

$$\mathcal{M}(f, g)(x) = \sup_{r>0} \langle |f|\rangle_{B(x,r)} \langle |g|\rangle_{B(x,r)}.$$

Acknowledgements. K.L. is supported by Juan de la Cierva - Formación 2015 FJCI-2015-24547, the Basque Government through the BERC 2014-2017 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323. H.M. is supported by the Academy of Finland through the grants 294840 and 306901, and is a member of the Finnish Centre of Excellence in Analysis and Dynamics Research. Y.O. was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2017 semester when this work was carried out, and was supported by the National Science Foundation under Grant No. DMS-1440140. E.V. is a member of the Finnish Centre of Excellence in Analysis and Dynamics Research.

2. Basic definitions

2.1. **Random dyadic grids, martingales, Haar functions.**

Let $\omega = (\omega^i)_{i\in\mathbb{Z}}$, where $\omega^i \in \{0, 1\}^n$. Let $\mathcal{D}_0$ be the standard dyadic grid on $\mathbb{R}^n$. We define the new dyadic grid

$$\mathcal{D}_\omega = \Big\{ I + \sum_{i\colon 2^{-i}<\ell(I)} 2^{-i}\omega^i \,:\, I \in \mathcal{D}_0 \Big\} = \{I + \omega \,:\, I \in \mathcal{D}_0\},$$

where we simply have defined $I + \omega := I + \sum_{i\colon 2^{-i}<\ell(I)} 2^{-i}\omega^i$. There is a natural product probability measure $\mathbb{P}_\omega = \mathbb{P}$ on $(\{0, 1\}^n)^{\mathbb{Z}}$ – this gives us the notion of random dyadic grids $\omega \mapsto \mathcal{D}_\omega$.

A cube $I \in \mathcal{D} = \mathcal{D}_\omega$ is called bad if there exists such a cube $J \in \mathcal{D}$ that $\ell(J) \geq 2^r \ell(I)$ and

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

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

For $I \in \mathcal{D}$ and a locally integrable function f we define the martingale difference

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

We have the standard estimate

$$\Big\| \Big(\sum_{I \in \mathcal{D}} |\Delta_I f|^2 \Big)^{1/2} \Big\|_{L^p} \sim \|f\|_{L^p}, \qquad 1 < p < \infty.$$

Writing $I = I_1 \times \cdots \times I_n$ we can define the Haar function h_I^η, $\eta = (\eta_1, \ldots, \eta_n) \in \{0,1\}^n$, by setting

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

where $h_{I_i}^0 = |I_i|^{-1/2} 1_{I_i}$ and $h_{I_i}^1 = |I_i|^{-1/2}(1_{I_{i,l}} - 1_{I_{i,r}})$ for every $i = 1, \ldots, n$. Here $I_{i,l}$ and $I_{i,r}$ are the left and right halves of the interval I_i respectively. If $\eta \neq 0$ the Haar function is cancellative: $\int h_I^\eta = 0$. We have that

$$\Delta_I f = \sum_{\eta \in \{0,1\}^n \setminus \{0\}} \langle f, h_I^\eta \rangle h_I^\eta,$$

but for convenience we understand that the η summation is suppressed and simply write

$$\Delta_I f = \langle f, h_I \rangle h_I.$$

In this paper h_I always denotes a cancellative Haar function (i.e. $h_I = h_I^\eta$ for some $\eta \neq 0$). A non-cancellative Haar function is explicitly denoted by h_I^0.

2.2. **Testing conditions: BMO and WBP.** Let K be a standard bilinear Calderón-Zygmund kernel, and let $\{T_\varepsilon\}_{\varepsilon > 0}$ be the related family of truncated operators. We recall a usual interpretation of $T_\varepsilon(1,1)$ and what is means that it belongs BMO.

Fix some $\varepsilon > 0$. Let $R \subset \mathbb{R}^n$ be a closed cube and let ϕ be an L^∞ function supported in R such that $\int \phi = 0$. Let $C = C(\varepsilon) \geq 3$ be any large constant so that $2^{-1}(C-1)\ell(R) > \varepsilon$, whence $|x - y| > \varepsilon$ for all $x \in R$ and $y \notin CR$. We define

$$\begin{aligned} \langle T_\varepsilon(1,1), \phi \rangle &:= \langle T_\varepsilon(1_{CR}, 1_{CR}), \phi \rangle \\ &\quad + \iiint \big(K(x,y,z) - K(c_R, y, z)\big) 1_{(CR \times CR)^c}(y,z) \phi(x) \, dy \, dz \, dx. \end{aligned} \tag{2.1}$$

Applying the x-Hölder estimate of the kernel it is seen that the integral is absolutely convergent. It is straightforward to check that the right hand side of (2.1) is independent of the cube R and the constant C as long as ϕ is supported in R and $2^{-1}(C-1)\ell(R) > \varepsilon$, $C \geq 3$.

If $\varphi \in \mathcal{A}$ and ϕ is as above, we define

$$\begin{aligned} \langle T_\varepsilon^\varphi(1,1), \phi \rangle &:= \langle T_\varepsilon^\varphi(1_{CR}, 1_{CR}), \phi \rangle \\ &\quad + \iiint \big(K_\varepsilon^\varphi(x,y,z) - K_\varepsilon^\varphi(c_R,y,z)\big) 1_{(CR\times CR)^c}(y,z)\phi(x)\,dy\,dz\,dx \end{aligned} \tag{2.2}$$

for any closed cube R containing the support of ϕ and any $C \geq 3$, say.

2.3. **Definition.** Let $\varepsilon > 0$. Suppose K is a standard bilinear Calderón-Zygmund kernel, and let T_ε be the related truncated operator. We say that $T_\varepsilon(1,1)$ is in BMO, and write $T_\varepsilon(1,1) \in \mathrm{BMO}$, if there exists a constant C so that for all closed cubes R and all functions ϕ supported in R such that $\|\phi\|_{L^\infty} \leq 1$ and $\int \phi = 0$ there holds

$$\frac{|\langle T_\varepsilon(1,1), \phi \rangle|}{|R|} \leq C. \tag{2.4}$$

We denote the smallest constant C in (2.4) by $\|T_\varepsilon(1,1)\|_{\mathrm{BMO}}$.

If $\varphi \in \mathcal{A}$, the corresponding definition for the smoothly truncated operator T_ε^φ is obtained just by replacing T_ε by T_ε^φ.

In the representation theorem we will assume that $T_\varepsilon(1,1) \in \mathrm{BMO}$. The following simple lemma shows that the conditions $T_\varepsilon(1,1) \in \mathrm{BMO}$ and $T_\varepsilon^\varphi(1,1) \in \mathrm{BMO}$ are equivalent.

2.5. **Lemma.** *Suppose K is a standard bilinear Calderón-Zygmund kernel and let $\varepsilon > 0$ and $\varphi \in \mathcal{A}$. Then*

$$\|T_\varepsilon^\varphi(1,1)\|_{\mathrm{BMO}} \leq C\big(\|K\|_{\mathrm{CZ}_\alpha} + \|T_\varepsilon(1,1)\|_{\mathrm{BMO}}\big)$$

and

$$\|T_\varepsilon(1,1)\|_{\mathrm{BMO}} \leq C\big(\|K\|_{\mathrm{CZ}_\alpha} + \|T_\varepsilon^\varphi(1,1)\|_{\mathrm{BMO}}\big).$$

Proof. Fix a closed cube R and a function ϕ supported in R such that $\|\phi\|_{L^\infty} \leq 1$ and $\int \phi = 0$. Then, using the definitions (2.1) and (2.2), one sees that

$$\begin{aligned} \big|\langle T_\varepsilon(1,1), \phi \rangle - \langle T_\varepsilon^\varphi(1,1), \phi \rangle\big| &= \big|\langle T_\varepsilon(1_{CR}, 1_{CR}), \phi \rangle - \langle T_\varepsilon^\varphi(1_{CR}, 1_{CR}), \phi \rangle\big| \\ &\lesssim \langle \mathcal{M}(1_{CR}, 1_{CR}), |\phi| \rangle \leq \int |\phi| dx \leq |R|. \end{aligned}$$

The claim follows from this estimate. □

For the convenience of the reader we state the following lemma on the equivalence of some BMO type conditions – although $T_\varepsilon^\varphi(1,1)$ is not stricly speaking a function, the lemma nevertheless follows from John–Nirenberg by standard arguments. Therefore, the paraproducts we will encounter can be made to obey the normalisation in (3.1).

2.6. **Lemma.** *Suppose K is a standard bilinear Calderón-Zygmund kernel and let $\varepsilon > 0$ and $\varphi \in \mathcal{A}$. Suppose $\mathcal{D}$ is a dyadic lattice. Then*

$$\sup_{R \in \mathcal{D}} \frac{1}{|R|} \sum_{\substack{Q \in \mathcal{D} \\ Q \subset R}} \big|\big\langle T_\varepsilon^\varphi(1,1), h_Q \big\rangle\big|^2 \le C \|T_\varepsilon^\varphi(1,1)\|_{\mathrm{BMO}}^2$$

for some absolute constant C.

Next, we give the definition of weak boundedness property.

2.7. **Definition.** The weak boundedness property constant $\|T_\varepsilon\|_{\mathrm{WBP}}$ is the best constant C so that the inequality

$$|\big\langle T_\varepsilon(1_I, 1_I), 1_I \big\rangle| \le C|I|$$

holds for all cubes $I \subset \mathbb{R}^n$.

2.3. **Sparse collections.** A collection $\mathcal{S}$ of cubes is said to be *η-sparse* (or just sparse), $0 < \eta < 1$, if for any $Q \in \mathcal{S}$ there exists $E_Q \subset Q$ so that $|E_Q| > \eta |Q|$ and $\{E_Q : Q \in \mathcal{S}\}$ are pairwise disjoint. The definition does not require the cubes to be part of some fixed dyadic grid. Although, it can be convenient to know that in Corollary 1.2 the sparse family $\mathcal{S}$ can always be found inside one of the fixed dyadic grids $\mathcal{D}_i$, $\#i \lesssim 1$.

3. Bilinear shifts

In this section all cubes are part of some fixed dyadic grid $\mathcal{D}$. We will introduce certain cancellative shifts and paraproducts in this section. We will also show their boundedness $L^p \times L^q \to L^r$ in the simple case $1 < p, q, r < \infty$ satisfying $1/p + 1/q = 1/r$. The restriction $r > 1$ can be lifted after we have shown the sparse domination (see Section 5).

3.1. **Cancellative bilinear shifts.** Define for $i, j, k \ge 0$ the bilinear shift $(f, g) \mapsto S^{i,j,k}(f, g)$ by setting

$$S^{i,j,k}(f, g) = \sum_Q A_Q^{i,j,k}(f, g),$$

where

$$A_Q^{i,j,k}(f, g) = \sum_{\substack{I, J, K \subset Q \\ \ell(I) = 2^{-i}\ell(Q) \\ \ell(J) = 2^{-j}\ell(Q) \\ \ell(K) = 2^{-k}\ell(Q)}} \alpha_{I,J,K,Q} \langle f, \tilde{h}_I \rangle \langle g, \tilde{h}_J \rangle h_K$$

and

$$(\tilde{h}_I, \tilde{h}_J) \in \big\{(h_I, h_J), (h_I^0, h_J)(h_I, h_J^0)\big\}.$$

We also demand that

$$|\alpha_{I,J,K,Q}| \le \frac{|I|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2}.$$

Such a shift will be considered to be a *cancellative bilinear shift*. Also the duals of these operators will be used in the representation.

Let $1 < p, q, r < \infty$ be such that $1/p + 1/q = 1/r$. We show that

$$\|S^{i,j,k}(f,g)\|_{L^r} \lesssim \|f\|_{L^p}\|g\|_{L^q}$$

with the constant independent of the shift in question, and only depending on p, q, r. To do this, we may assume without loss of generality that for example $\tilde{h}_I = h_I$ for all I (a general shift can be split into two shifts where $\tilde{h}_I = h_I$ for all I in one of them and $\tilde{h}_J = h_J$ for all J in the other). Notice that we have the pointwise estimate

$$|A_Q^{i,j,k}(f,g)| \le \langle |f| \rangle_Q \langle |g| \rangle_Q 1_Q.$$

Define also

$$D_Q^i f = \sum_{\substack{I \subset Q \\ \ell(I) = 2^{-i}\ell(Q)}} \langle f, h_I \rangle h_I.$$

Since $A_Q^{i,j,k}(f,g) = A_Q^{i,j,k}(D_Q^i f, g)$, we have

$$|A_Q^{i,j,k}(f,g)| \le M_{\mathcal{D}} g \langle |D_Q^i f| \rangle_Q 1_Q.$$

Notice that

$$\Big\| \Big(\sum_Q |D_Q^i f|^2 \Big)^{1/2} \Big\|_{L^p} = \Big\| \Big(\sum_I |\Delta_I f|^2 \Big)^{1/2} \Big\|_{L^p} \sim \|f\|_{L^p}, \qquad 1 < p < \infty.$$

Let $1 < p, q, r < \infty$ be such that $1/p + 1/q = 1/r$. Using the above we see that

$$\|S^{i,j,k}(f,g)\|_{L^r} \sim \Big\| \Big(\sum_Q |D_Q^k(S^{i,j,k}(f,g))|^2 \Big)^{1/2} \Big\|_{L^r} = \Big\| \Big(\sum_Q |A_Q^{i,j,k}(f,g)|^2 \Big)^{1/2} \Big\|_{L^r}.$$

Now, we have

$$\begin{aligned}
\Big\| \Big(\sum_Q |A_Q^{i,j,k}(f,g))|^2 \Big)^{1/2} \Big\|_{L^r} &\le \Big\| M_{\mathcal{D}} g \Big(\sum_Q \langle |D_Q^i f| \rangle_Q^2 1_Q \Big)^{1/2} \Big\|_{L^r} \\
&\le \Big\| \Big(\sum_Q \langle |D_Q^i f| \rangle_Q^2 1_Q \Big)^{1/2} \Big\|_{L^p} \|M_{\mathcal{D}} g\|_{L^q} \\
&\lesssim \Big\| \Big(\sum_Q |D_Q^i f|^2 \Big)^{1/2} \Big\|_{L^p} \|g\|_{L^q} \lesssim \|f\|_{L^p}\|g\|_{L^q}.
\end{aligned}$$

3.2. **Bilinear paraproduct.** Let $\alpha = \{\alpha_K\}_{K \in \mathcal{D}}$ be sequence of complex numbers such that

$$\frac{1}{|K_0|} \sum_{K:\, K \subset K_0} |\alpha_K|^2 \le 1 \tag{3.1}$$

for all $K_0 \in \mathcal{D}$. We define the bilinear paraproduct

$$\Pi_\alpha(f,g) = \sum_K \alpha_K \langle f \rangle_K \langle g \rangle_K h_K.$$

To deal with this it is useful to recall the usual (linear) paraproduct

$$\pi_\alpha f = \sum_K \alpha_K \langle f \rangle_K h_K.$$

It is well known that $\pi_\alpha \colon L^r \to L^r$ boundedly for $1 < r < \infty$ because of the condition (3.1). An elegant way to do this directly in L^r is in [6]. It follows that $\Pi_\alpha \colon L^p \times L^q \to L^r$ boundedly for $1 < p, q, r < \infty$ satisfying $1/r = 1/p + 1/q$. Indeed, it holds that

$$\begin{aligned}
\|\Pi_\alpha(f,g)\|_{L^r} &\sim \Big\| \Big(\sum_K |\alpha_K|^2 |\langle f \rangle_K|^2 |\langle g \rangle_K|^2 \frac{1_K}{|K|} \Big)^{1/2} \Big\|_{L^r} \\
&\le \Big\| \Big(\sum_K |\alpha_K|^2 \langle \mathcal{M}_{\mathcal{D}}(f,g) \rangle_K^2 \frac{1_K}{|K|} \Big)^{1/2} \Big\|_{L^r} \\
&\sim \|\pi_\alpha(\mathcal{M}_{\mathcal{D}}(f,g))\|_{L^r} \lesssim \|\mathcal{M}_{\mathcal{D}}(f,g)\|_{L^r} \lesssim \|f\|_{L^p} \|g\|_{L^q}.
\end{aligned}$$

4. Proof of the bilinear representation theorem, Theorem 1.1

Consider an arbitrary $\varepsilon_1 > 0$ and let f, g and h be bounded functions with compact support. For the moment, let $\varepsilon_2 > \varepsilon_1$ be arbitrary, and write $T = T^{\varphi}_{\varepsilon_1,\varepsilon_2}$ and $K = K^{\varphi}_{\varepsilon_1,\varepsilon_2}$. This is an a priori bounded operator (for example in the $L^4 \times L^4 \to L^2$ sense), which makes the calculations below legit. We will decompose $\langle T(f,g), h \rangle$ first, and take the limit $\epsilon_2 \to \infty$ at the end.

Begin by decomposing $\langle T(f,g), h \rangle$ as

$$\begin{aligned}
\langle T(f,g), h \rangle &= E_\omega \sum_{K \in \mathcal{D}_\omega} \sum_{\substack{I \in \mathcal{D}_\omega \\ \ell(K) \le \ell(I)}} \sum_{\substack{J \in \mathcal{D}_\omega \\ \ell(K) \le \ell(J)}} \langle T(\Delta_I f, \Delta_J g), \Delta_K h \rangle \\
&+ E_\omega \sum_{I \in \mathcal{D}_\omega} \sum_{\substack{J \in \mathcal{D}_\omega \\ \ell(I) \le \ell(J)}} \sum_{\substack{K \in \mathcal{D}_\omega \\ \ell(I) < \ell(K)}} \langle T^{1*}(\Delta_K h, \Delta_J g), \Delta_I f \rangle \\
&+ E_\omega \sum_{J \in \mathcal{D}_\omega} \sum_{\substack{I \in \mathcal{D}_\omega \\ \ell(J) < \ell(I)}} \sum_{\substack{K \in \mathcal{D}_\omega \\ \ell(J) < \ell(K)}} \langle T^{2*}(\Delta_I f, \Delta_K h), \Delta_J g \rangle =: \Sigma^1 + \Sigma^2 + \Sigma^3.
\end{aligned}$$

We focus on the first sum Σ^1, and at this point write

$$\sum_{K \in \mathcal{D}_\omega} \sum_{\substack{I \in \mathcal{D}_\omega \\ \ell(K) \le \ell(I)}} \sum_{\substack{J \in \mathcal{D}_\omega \\ \ell(K) \le \ell(J)}} \langle T(\Delta_I f, \Delta_J g), \Delta_K h \rangle = \sum_{K \in \mathcal{D}_\omega} \langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K h \rangle,$$

where

$$E^\omega_{\ell(K)/2} f = \sum_{\substack{I \in \mathcal{D}_\omega \\ \ell(I) = \ell(K)/2}} 1_I \big\langle f \big\rangle_I.$$

The point of doing this is to gain the needed independence for the argument below (this seems to be a new simpler way to add goodness than in [5], and is straightforward to use also in this bilinear setting). Write now $\mathcal{D}_\omega = \mathcal{D}_0 + \omega$ to the end that

$$\sum_{K \in \mathcal{D}_\omega} \big\langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K h \big\rangle = \sum_{K \in \mathcal{D}_0} \big\langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_{K+\omega} h \big\rangle.$$

Next, we write

$$\begin{aligned}
\Sigma^1 &= E_\omega \sum_{K \in \mathcal{D}_0} \big\langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_{K+\omega} h \big\rangle \\
&= \frac{1}{\pi_{\text{good}}} \sum_{K \in \mathcal{D}_0} E_\omega[1_{\text{good}}(K+\omega)] E_\omega[\big\langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_{K+\omega} h \big\rangle] \\
&= \frac{1}{\pi_{\text{good}}} E_\omega \sum_{K \in \mathcal{D}_{\omega,\, \text{good}}} \big\langle T(E^\omega_{\ell(K)/2} f, E^\omega_{\ell(K)/2} g), \Delta_K h \big\rangle =: \frac{1}{\pi_{\text{good}}} E_\omega \Sigma^1(\omega),
\end{aligned}$$

where we used independence: $1_{\text{good}}(K+\omega)$ depends on ω_j for $2^{-j} \ge \ell(K)$ while $E^\omega_{\ell(K)/2} f$ depends on ω_j for $2^{-j} < \ell(K)/2 < \ell(K)$, same for $E^\omega_{\ell(K)/2} g$, and $\Delta_K h$ depends on ω_j for $2^{-j} < \ell(K)$.

Fix ω and let $\mathcal{D}_\omega = \mathcal{D}$. We will now start finding the shift structure in the sum $\Sigma^1(\omega)$ i.e.

$$\sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{I \in \mathcal{D} \\ \ell(K) \le \ell(I)}} \sum_{\substack{J \in \mathcal{D} \\ \ell(K) \le \ell(J)}} \big\langle T(\Delta_I f, \Delta_J g), \Delta_K h \big\rangle.$$

The double sum $\sum_{\substack{I \in \mathcal{D} \\ \ell(K) \le \ell(I)}} \sum_{\substack{J \in \mathcal{D} \\ \ell(K) \le \ell(J)}}$ can be organised as

$$\sum_{\substack{I \in \mathcal{D} \\ \ell(K) \le \ell(I)}} \sum_{\substack{J \in \mathcal{D} \\ \ell(I) \le \ell(J)}} + \sum_{\substack{J \in \mathcal{D} \\ \ell(K) \le \ell(J)}} \sum_{\substack{I \in \mathcal{D} \\ \ell(J) < \ell(I)}}.$$

This leads to the fact that

$$\begin{aligned}
&\sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{I \in \mathcal{D} \\ \ell(K) \le \ell(I)}} \sum_{\substack{J \in \mathcal{D} \\ \ell(K) \le \ell(J)}} \big\langle T(\Delta_I f, \Delta_J g), \Delta_K h \big\rangle \\
&\quad = \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{I \in \mathcal{D} \\ \ell(K) \le \ell(I)}} \big\langle T(\Delta_I f, E_{\ell(I)/2} g), \Delta_K h \big\rangle \\
&\quad + \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{J \in \mathcal{D} \\ \ell(K) \le \ell(J)}} \big\langle T(E_{\ell(J)} f, \Delta_J g), \Delta_K h \big\rangle =: \sigma^1 + \sigma^2.
\end{aligned}$$

We will now mostly focus on the part

$$\sigma^1 = \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{I \in \mathcal{D} \\ \ell(K) \le \ell(I)}} \big\langle T(\Delta_I f, E_{\ell(I)/2} g), \Delta_K h \big\rangle$$

$$= \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{I \in \mathcal{D} \\ \ell(K) \le \ell(I)}} \sum_{\substack{J \in \mathcal{D} \\ \ell(J) = \ell(I)/2}} \big\langle T(\Delta_I f, 1_J \langle g \rangle_J), \Delta_K h \big\rangle.$$

However, to get a simple paraproduct it is crucial to combine i.e. sum up the paraproduct parts from these two parts σ^1 and σ^2.

Step I: separated part. In this section we consider

$$\sigma^1_1 := \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{I, J \in \mathcal{D}: \\ \ell(K) \le \ell(I) = 2\ell(J) \\ \max(d(K,I), d(K,J)) > \ell(K)^\gamma \ell(J)^{1-\gamma}}} \big\langle T(\Delta_I f, 1_J \langle g \rangle_J), \Delta_K h \big\rangle$$

$$= \sum_{K \in \mathcal{D}_{\text{good}}} \sum_{\substack{I, J \in \mathcal{D}: \\ \ell(K) \le \ell(I) = 2\ell(J) \\ \max(d(K,I), d(K,J)) > \ell(K)^\gamma \ell(J)^{1-\gamma}}} \big\langle T(h_I, h_J^0), h_K \big\rangle \big\langle f, h_I \big\rangle \big\langle g, h_J^0 \big\rangle \big\langle h, h_K \big\rangle.$$

We need the existence of certain nice parents, the proof in the bilinear setting is essentially the same as in [5].

4.1. **Lemma.** *For I, J, K as in σ^1_1 there exists a cube $Q \in \mathcal{D}$ so that $I \cup J \cup K \subset Q$ and*

$$\max(d(K,I), d(K,J)) \gtrsim \ell(K)^\gamma \ell(Q)^{1-\gamma}.$$

Proof. Let $Q \in \mathcal{D}$ be the minimal parent of K for which both of the following two conditions hold:

- $\ell(Q) \ge 2^r \ell(K)$;
- $\max(d(K,I), d(K,J)) \le \ell(K)^\gamma \ell(Q)^{1-\gamma}$.

Since $\ell(Q) \ge 2^r \ell(K)$, the goodness of K gives that

$$\ell(K)^\gamma \ell(Q)^{1-\gamma} < d(K, Q^c).$$

If we would have that $I \subset Q^c$ or $J \subset Q^c$ we would get

$$\ell(K)^\gamma \ell(Q)^{1-\gamma} < \max(d(K,I), d(K,J)) \le \ell(K)^\gamma \ell(Q)^{1-\gamma},$$

which is a contradiction. Therefore, we have $I \cap Q \ne \emptyset$ and $J \cap Q \ne \emptyset$. Moreover, we have

$$\ell(K)^\gamma \ell(J)^{1-\gamma} < \max(d(K,I), d(K,J)) \le \ell(K)^\gamma \ell(Q)^{1-\gamma}$$

implying that $\ell(Q) > \ell(J)$, and so also $\ell(Q) \ge \ell(I)$. This implies $I \cup J \cup K \subset Q$.

It remains to note that the estimate $\max(d(K,I), d(K,J)) \gtrsim \ell(K)^\gamma \ell(Q)^{1-\gamma}$ is a trivial consequence of the minimality of Q. Indeed, there is something to check only if Q is minimal because $\ell(Q) \lesssim \ell(K)$. But then $\ell(Q) \lesssim \ell(J)$ and we get

$$\ell(K)^\gamma \ell(Q)^{1-\gamma} \lesssim \ell(K)^\gamma \ell(J)^{1-\gamma} < \max(d(K,I), d(K,J)).$$

$\square$

For I, J, K as in σ_1^1 we let $Q = I \vee J \vee K$ be the minimal cube $Q \in \mathcal{D}$ so that $I \cup J \cup K \subset Q$. We then know that

$$\max(d(K,I), d(K,J)) \gtrsim \ell(K)^\gamma \ell(Q)^{1-\gamma}. \tag{4.2}$$

Let us write

$$\sigma_1^1 = \sum_{k=0}^{\infty} \sum_{i=0}^{k} \sum_{Q \in \mathcal{D}} \sum_{\substack{I,J \in \mathcal{D},\, K \in \mathcal{D}_{\text{good}} \\ \max(d(K,I),d(K,J)) > \ell(K)^\gamma \ell(J)^{1-\gamma} \\ 2\ell(J) = \ell(I) = 2^{-i}\ell(Q),\, \ell(K) = 2^{-k}\ell(Q) \\ I \vee J \vee K = Q}} \big\langle T(h_I, h_J^0), h_K \big\rangle \big\langle f, h_I \big\rangle \big\langle g, h_J^0 \big\rangle \big\langle h, h_K \big\rangle.$$

Next, we define

$$\alpha_{I,J,K,Q} = \frac{\big\langle T(h_I, h_J^0), h_K \big\rangle}{C(\ell(K)/\ell(Q))^{\alpha/2}}$$

if $I, J \in \mathcal{D}$, $K \in \mathcal{D}_{\text{good}}$, $\max(d(K,I), d(K,J)) > \ell(K)^\gamma \ell(J)^{1-\gamma}$, $\ell(K) \le \ell(I) = 2\ell(J)$ and $I \vee J \vee K = Q$, and $\alpha_{I,J,K,Q} = 0$ otherwise. We can then write for fixed $k \ge 0$ and $i \le k$ that

$$\begin{aligned} &\sum_{Q \in \mathcal{D}} \sum_{\substack{I,J \in \mathcal{D},\, K \in \mathcal{D}_{\text{good}} \\ \max(d(K,I),d(K,J)) > \ell(K)^\gamma \ell(J)^{1-\gamma} \\ 2\ell(J) = \ell(I) = 2^{-i}\ell(Q),\, \ell(K) = 2^{-k}\ell(Q) \\ I \vee J \vee K = Q}} \big\langle T(h_I, h_J^0), h_K \big\rangle \big\langle f, h_I \big\rangle \big\langle g, h_J^0 \big\rangle h_K \\ &= C 2^{-\alpha k/2} \sum_{Q \in \mathcal{D}} \sum_{\substack{I,J,K \subset Q \\ \ell(I) = 2^{-i}\ell(Q) \\ \ell(J) = 2^{-i-1}\ell(Q) \\ \ell(K) = 2^{-k}\ell(Q)}} \alpha_{I,J,K,Q} \big\langle f, h_I \big\rangle \big\langle g, h_J^0 \big\rangle h_K =: C 2^{-\alpha k/2} S^{i,i+1,k}(f,g), \end{aligned}$$

which gives

$$\sigma_1^1 = C \sum_{k=0}^{\infty} \sum_{i=0}^{k} 2^{-k\alpha/2} \big\langle S^{i,i+1,k}(f,g), h \big\rangle.$$

It remains to verify that

$$|\alpha_{I,J,K,Q}| \le \frac{|I|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2}$$

for an appropriate choice of the constant C depending on the kernel estimates. We fix I, J, K, Q so that $\alpha_{I,J,K,Q} \ne 0$. Notice that $|x - c_K| \le \ell(K)/2$ (we are using the ℓ^∞ distance) for $x \in K$ while

$$\max(|x-y|, |x-z|) \ge \max(d(K,I), d(K,J)) > \ell(K)^\gamma \ell(J)^{1-\gamma} \ge 2^\gamma \frac{\ell(K)}{2} \ge 2^\gamma |x - c_K|$$

for $x \in K$, $y \in I$ and $z \in J$. Therefore, we have by the Hölder estimate in the x variable and the estimate (4.2) that

$$\begin{aligned}
|\langle T(h_I, h_J^0), h_K\rangle| &\lesssim \|h_I\|_{L^1}\|h_J^0\|_{L^1}\|h_K\|_{L^1}\frac{\ell(K)^\alpha}{\max(d(K,I),d(K,J))^{2n+\alpha}} \\
&\lesssim |I|^{1/2}|J|^{1/2}|K|^{1/2}\frac{\ell(K)^\alpha}{(\ell(K)^\gamma\ell(Q)^{1-\gamma})^{2n+\alpha}} \\
&= \frac{|I|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2}\Big(\frac{\ell(K)}{\ell(Q)}\Big)^{\alpha-\gamma(2n+\alpha)} \\
&= \frac{|I|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2}\Big(\frac{\ell(K)}{\ell(Q)}\Big)^{\alpha/2}.
\end{aligned}$$

This establishes the desired normalisation, and therefore we are done with σ_1^1.

Step II: diagonal. Here we look at the sum

$$\begin{aligned}
\sigma_2^1 :&= \sum_{K\in\mathcal{D}_{\text{good}}} \sum_{\substack{I,J\in\mathcal{D}:\\ \ell(K)\le\ell(I)=2\ell(J)\\ \max(d(K,I),d(K,J))\le\ell(K)^\gamma\ell(J)^{1-\gamma}\\ K\cap I=\emptyset \text{ or } K=I \text{ or } K\cap J=\emptyset}} \langle T(\Delta_I f, 1_J\langle g\rangle_J), \Delta_K h\rangle \\
&= \sum_{K\in\mathcal{D}_{\text{good}}} \sum_{\substack{I,J\in\mathcal{D}:\\ \ell(K)\le\ell(I)=2\ell(J)\le 2^r\ell(K)\\ \max(d(K,I),d(K,J))\le\ell(K)^\gamma\ell(J)^{1-\gamma}\\ K\cap I=\emptyset \text{ or } K=I \text{ or } K\cap J=\emptyset}} \langle T(h_I, h_J^0), h_K\rangle\langle f, h_I\rangle\langle g, h_J^0\rangle\langle h, h_K\rangle.
\end{aligned}$$

The goodness of the cube K was used to conclude that we cannot have $\ell(I) > 2^r\ell(K)$. Indeed, in the case $K\cap I=\emptyset$ this would imply $d(K,I) > \ell(K)^\gamma\ell(I)^{1-\gamma} \ge \ell(K)^\gamma\ell(J)^{1-\gamma}$ – a contradiction. In the case $K\cap J=\emptyset$ we would have (as $\ell(J)\ge 2^r\ell(K)$) that $d(K,J) > \ell(K)^\gamma\ell(J)^{1-\gamma}$ – a contradiction.

4.3. **Lemma.** *For I, J, K as in σ_2^1 there exists a cube $Q\in\mathcal{D}$ so that $I\cup J\cup K\subset Q$ and $\ell(Q)\le 2^r\ell(K)$.*

Proof. Define $Q = K^{(r)}$. Then $\ell(Q) = 2^r\ell(K) \ge \ell(I) > \ell(J)$. Therefore, it suffices to show that $I\cap Q\ne\emptyset$ and $J\cap Q\ne\emptyset$. But this is essentially the same argument as previously: If we would have that $I\subset Q^c$ or $J\subset Q^c$, we would get

$$\ell(K)^\gamma\ell(Q)^{1-\gamma} < d(K,Q^c) \le \max(d(K,I),d(K,J)) \le \ell(K)^\gamma\ell(J)^{1-\gamma},$$

which implies $\ell(J) > \ell(Q)$ – a contradiction. □

We can now write

$$\sigma_2^1 = \sum_{k=0}^{r}\sum_{i=0}^{k}\sum_{Q} \sum_{\substack{I,J\in\mathcal{D},\, K\in\mathcal{D}_{\text{good}}:\\ \max(d(K,I),d(K,J))\le\ell(K)^\gamma\ell(J)^{1-\gamma}\\ K\cap I=\emptyset \text{ or } K=I \text{ or } K\cap J=\emptyset\\ 2\ell(J)=\ell(I)=2^{-i}\ell(Q),\, \ell(K)=2^{-k}\ell(Q)\\ I\vee J\vee K=Q}} \langle T(h_I, h_J^0), h_K\rangle\langle f, h_I\rangle\langle g, h_J^0\rangle\langle h, h_K\rangle.$$

Notice that if $K \cap I = \emptyset$ then

$$
\begin{aligned}
|\langle T(h_I, h_J^0), h_K\rangle| &\lesssim |I|^{-1/2}|J|^{-1/2}|K|^{-1/2} \int_{10I\setminus I}\int_I \frac{dy\,dx}{|x-y|^n} \\
&\lesssim |I|^{1/2}|J|^{-1/2}|K|^{-1/2} \sim \frac{|I|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2}\Big(\frac{\ell(K)}{\ell(Q)}\Big)^{\alpha/2}.
\end{aligned}
$$

We get the same bound also if $K \cap J = \emptyset$ with an analogous calculation. So we only need to estimate in the case $K = I$ and $J \in \operatorname{ch}(K)$. Then we have

$$
|\langle T(h_K, h_J^0), h_K\rangle| \lesssim |K|^{-3/2} \sum_{K',K''\in \operatorname{ch}(K)} |\langle T(1_{K'}, 1_J), 1_{K''}\rangle|.
$$

If $K' \ne J$ or $K'' \ne J$ then $|\langle T(1_{K'}, 1_J), 1_{K''}\rangle| \lesssim |K|$ simply by the size estimate of the kernel. In the case $K' = K'' = J$ we have using the weak boundedness property that $|\langle T(1_J, 1_J), 1_J\rangle| \lesssim |K|$. So in the case $K = I$ and $J \in \operatorname{ch}(K)$ we also have

$$
|\langle T(h_K, h_J^0), h_K\rangle| \lesssim |K|^{-1/2} \sim \frac{|I|^{1/2}|J|^{1/2}|K|^{1/2}}{|Q|^2}\Big(\frac{\ell(K)}{\ell(Q)}\Big)^{\alpha/2}.
$$

The above lets us write

$$
\sigma_2^1 = C\sum_{k=0}^{r}\sum_{i=0}^{k} 2^{-k\alpha/2}\langle S^{i,i+1,k}(f,g), h\rangle
$$

for cancellative bilinear shifts $S^{i,i+1,k}$, where C depends on the kernel estimates and the weak boundedness property. We point out at this point that since $T = T^{\varphi}_{\varepsilon_1,\varepsilon_2} = T^{\varphi}_{\varepsilon_1} - T^{\varphi}_{\varepsilon_2}$, there holds

$$
\|T\|_{\mathrm{WBP}} \le C'\big(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0}\|T_\delta\|_{\mathrm{WBP}}).
$$

4.1. **Step III: error terms.** Here we start working with the sum

$$
\begin{aligned}
\sigma_3^1 &:= \sum_{\substack{I,J\in\mathcal{D},\, K\in\mathcal{D}_{\mathrm{good}} \\ \ell(I)=2\ell(J) \\ K\subset J\subset I}} \langle T(\Delta_I f, 1_J\langle g\rangle_J), \Delta_K h\rangle \\
&= \sum_{\substack{J\in\mathcal{D},\, K\in\mathcal{D}_{\mathrm{good}} \\ K\subset J}} \langle T(\Delta_{J^{(1)}} f, 1_J), \Delta_K h\rangle\langle g\rangle_J.
\end{aligned}
$$

We split

$$
\begin{aligned}
\langle T(\Delta_{J^{(1)}} f, 1_J), \Delta_K h\rangle &= \langle T(1_{J^c}(\Delta_{J^{(1)}} f - \langle \Delta_{J^{(1)}} f\rangle_J), 1_J), \Delta_K h\rangle \\
&\quad - \langle \Delta_{J^{(1)}} f\rangle_J \langle T(1, 1_{J^c}), \Delta_K h\rangle + \langle \Delta_{J^{(1)}} f\rangle_J\langle T(1,1), \Delta_K h\rangle.
\end{aligned}
$$

This gives us the decomposition $\sigma_3^1 = \sigma_{3,e}^1 + \sigma_{3,\pi}^1$, where the first two terms of the above decomposition are part of $\sigma_{3,e}^1$.

In this section we only deal with the error term $\sigma^1_{3,e}$. Notice that

$$\begin{aligned}\sigma^1_{3,e} = \sum_{\substack{J\in\mathcal{D},\, K\in\mathcal{D}_{\text{good}}\\ K\subset J}} &|J|^{-1/2}\big[\big\langle T(s_J, 1_J), h_K\big\rangle \\ &- \big\langle h_{J^{(1)}}\big\rangle_J \big\langle T(1, 1_{J^c}), h_K\big\rangle\big]\big\langle f, h_{J^{(1)}}\big\rangle \big\langle g, h^0_J\big\rangle \big\langle h, h_K\big\rangle,\end{aligned}$$

where $s_J := 1_{J^c}(h_{J^{(1)}} - \langle h_{J^{(1)}}\rangle_J)$ satisfies $|s_J| \lesssim |J|^{-1/2}$ and $\operatorname{spt} s_J \subset J^c$.

We will first bound $|\langle T(s_J, 1_J), h_K\rangle|$. In the case $\ell(J) \sim \ell(K)$ we are looking for the bound $|\langle T(s_J, 1_J), h_K\rangle| \lesssim 1$. This follows by writing

$$|\langle T(s_J, 1_J), h_K\rangle| \le |\langle T(1_{3J}s_J, 1_J), h_K\rangle| + |\langle T(1_{(3J)^c}s_J, 1_J), h_K\rangle|,$$

and using the size and Hölder estimate in the x-variable respectively. If $\ell(J) \ge 2^r\ell(K)$ we have $d(K, J^c) \ge \ell(K)^\gamma \ell(J)^{1-\gamma} \ge \ell(K)^{1/2}\ell(J)^{1/2}$. Therefore, Hölder estimate in the x-variable gives

$$\begin{aligned}|\langle T(s_J, 1_J), h_K\rangle| &\lesssim |K|^{-1/2}|J|^{-1/2}\ell(K)^\alpha \int_K \int_{J^c} \frac{dy}{|x-y|^{n+\alpha}}\, dx \\ &\lesssim |K|^{1/2}|J|^{-1/2}\Big(\frac{\ell(K)}{\ell(J)}\Big)^{\alpha/2}.\end{aligned}$$

Notice that this is ~ 1 if $\ell(J) \sim \ell(K)$, so the same estimate holds in both cases.

It is now also obvious, using almost exactly the same calculations as above, that

$$|\langle T(1, 1_{J^c}), h_K\rangle| \lesssim |K|^{1/2}\Big(\frac{\ell(K)}{\ell(J)}\Big)^{\alpha/2}.$$

But as $|\langle h_{J^{(1)}}\rangle_J| \lesssim |J|^{-1/2}$ we have the same bound as above. Therefore, we can write

$$\sigma^1_{3,e} = C\sum_{k=1}^\infty 2^{-\alpha k/2}\big\langle S^{0,1,k}(f,g), h\big\rangle$$

for some cancellative bilinear shifts and for some C depending on the kernel estimates.

4.2. **Part IV: paraproduct.** Here we combine

$$\sigma^1_{3,\pi} = \sum_{\substack{J\in\mathcal{D},\, K\in\mathcal{D}_{\text{good}}\\ K\subset J}} \big\langle T(1,1), \Delta_K h\big\rangle \big\langle \Delta_{J^{(1)}} f\big\rangle_J \big\langle g\big\rangle_J$$

with the relevant paraproduct type term coming from σ^2, namely

$$\sigma^2_{3,\pi} = \sum_{\substack{J\in\mathcal{D},\, K\in\mathcal{D}_{\text{good}}\\ K\subset J}} \big\langle T(1,1), \Delta_K h\big\rangle \big\langle f\big\rangle_{J^{(1)}} \big\langle \Delta_{J^{(1)}} g\big\rangle_J.$$

Notice the key cancellation

$$\big\langle \Delta_{J^{(1)}} f\big\rangle_J \big\langle g\big\rangle_J + \big\langle f\big\rangle_{J^{(1)}} \big\langle \Delta_{J^{(1)}} g\big\rangle_J = \big\langle f\big\rangle_J \big\langle g\big\rangle_J - \big\langle f\big\rangle_{J^{(1)}} \big\langle g\big\rangle_{J^{(1)}}.$$

Therefore, we get

$$\sigma^1_{3,\pi} + \sigma^2_{3,\pi} = \sum_{K \in \mathcal{D}_{\mathrm{good}}} \big\langle T(1,1), \Delta_K h \big\rangle \big\langle f \big\rangle_K \big\langle g \big\rangle_K .$$

Define

$$\alpha_K = \frac{\big\langle T(1,1), h_K \big\rangle}{C\big(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0} \|T_\delta(1,1)\|_{\mathrm{BMO}}\big)} \tag{4.4}$$

if K is good, where C is a large enough absolute constant, and otherwise set $\alpha_K = 0$. Recall that $T = T^\varphi_{\varepsilon_1,\varepsilon_2} = T^\varphi_{\varepsilon_1} - T^\varphi_{\varepsilon_2}$, whence in view of Lemma 2.5 and Lemma 2.6 the numbers α_K satisfy the correct normalisation (3.1). Hence we can write

$$\sigma^1_{3,\pi} + \sigma^2_{3,\pi} = C\big(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0} \|T_\delta(1,1)\|_{\mathrm{BMO}}\big) \langle \Pi_\alpha(f,g), h \rangle .$$

4.3. **Synthesis.** Let us collect the pieces of the above steps together. Recall that the operator T is actually $T^\varphi_{\varepsilon_1,\varepsilon_2}$. We have shown that

$$\begin{aligned}\Sigma^1(\omega) &= C(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0} \|T_\delta\|_{\mathrm{WBP}}) \sum_{k=0}^{\infty} \sum_{i=0}^{k} 2^{-\alpha k/2} \big\langle U^{i,k}_{\varepsilon_1,\varepsilon_2,\varphi,\omega}(f,g), h \big\rangle \\ &\quad + C(\|K\|_{\mathrm{CZ}_\alpha} + \sup_{\delta>0} \|T_\delta(1,1)\|_{\mathrm{BMO}}) \big\langle \Pi_{\alpha_0(\varepsilon_1,\varepsilon_2,\varphi,\omega)}(f,g), h \big\rangle,\end{aligned}$$

where each $U^{i,k}_{\varepsilon_1,\varepsilon_2,\varphi,\omega}$ is a sum of cancellative shifts $S^{i,i,k}_{\varepsilon_1,\varepsilon_2,\varphi,\omega}$ and $S^{i,i+1,k}_{\varepsilon_1,\varepsilon_2,\varphi,\omega}$, and where $\Pi_{\alpha_0(\varepsilon_1,\varepsilon_2,\varphi,\omega)}$ is the paraproduct related to the sequence defined around Equation (4.4). Collecting together the symmetric parts we get the result of Theorem 1.1 except we have the dependence on ϵ_2 on both sides. However, it is clear that $\big\langle T^\varphi_{\varepsilon_1,\varepsilon_2}(f,g), h \big\rangle = \big\langle T^\varphi_{\varepsilon_1}(f,g), h \big\rangle$ if ϵ_2 is large enough (depending on the supports of f, g and h.) Thus, it is enough to do some limiting argument $\epsilon_2 \to \infty$ on the right hand side also.

The operators $U^{i,k}_{\varepsilon_1,\varepsilon_2,\varphi,\omega}$ depend on ε_1, ε_2 and φ because the coefficients of the shifts are defined using the operator $T^\varphi_{\varepsilon_1,\varepsilon_2}$. Let $U^{i,k}_{\varepsilon_1,\varphi,\omega}$ be the corresponding operator, but where the coefficients of the shifts are defined with the operator $T^\varphi_{\varepsilon_1}$ instead. Do the similar thing with the paraproducts. Dominated convergence theorem shows that it is enough to show that

$$\big\langle U^{i,k}_{\varepsilon_1,\varepsilon_2,\varphi,\omega}(f,g), h \big\rangle \to \big\langle U^{i,k}_{\varepsilon_1,\varphi,\omega}(f,g), h \big\rangle,$$

when $\epsilon_2 \to \infty$, and similarly for the paraproducts. The convergence of the above pairings is simply based on the fact that the coefficients of the shifts defined with $T^\varphi_{\varepsilon_1,\varepsilon_2}$ approach to the ones defined with $T^\varphi_{\varepsilon_1}$. Let us quickly show the argument for the paraproduct, the same reasoning applies for the cancellative shifts.

It is enough to show that

$$\lim_{\epsilon_2 \to \infty} \Big| \sum_{K \in \mathcal{D}_{\mathrm{good}}} \big\langle T^\varphi_{\varepsilon_2}(1,1), h_K \big\rangle \big\langle f \big\rangle_K \big\langle g \big\rangle_K \big\langle h, h_K \big\rangle \Big| = 0.$$

Fix $M > 0$. Notice that using $\sup_{\delta>0} \|T^{\varphi}_{\delta}(1,1)\|_{\mathrm{BMO}} < \infty$ and the boundness of the paraproduct there holds for every $\epsilon_2 > 0$ that

$$\begin{aligned}\Big| \sum_{K:\, \ell(K)<1/M \text{or } \ell(K)>M} &\big\langle T^{\varphi}_{\varepsilon_2}(1,1), h_K\big\rangle \big\langle f\big\rangle_K \big\langle g\big\rangle_K \big\langle h, h_K\big\rangle \Big| \\ &\lesssim \|f\|_{L^4}\|g\|_{L^4}\Big(\sum_{K:\, \ell(K)<1/M \text{or } \ell(K)>M} \|\Delta_K h\|_{L^2}^2\Big)^{1/2} = c(M),\end{aligned}$$

where $c(M) \to 0$ when $M \to \infty$. This gives that

$$\begin{aligned}\Big| \sum_K &\big\langle T^{\varphi}_{\varepsilon_2}(1,1), h_K\big\rangle \big\langle f\big\rangle_K \big\langle g\big\rangle_K \big\langle h, h_K\big\rangle \Big| \\ &\le c(M) + \Big| \sum_{K:\, 1/M \le \ell(K) \le M} \big\langle T^{\varphi}_{\varepsilon_2}(1,1), h_K\big\rangle \big\langle f\big\rangle_K \big\langle g\big\rangle_K \big\langle h, h_K\big\rangle \Big|.\end{aligned}$$

The latter sum is finite as h has compact support. Since $\big\langle T^{\varphi}_{\varepsilon_2}(1,1), h_K\big\rangle \to 0$ when $\epsilon_2 \to \infty$, we have that

$$\lim_{\epsilon_2 \to \infty} \Big| \sum_{K \in \mathcal{D}_{\mathrm{good}}} \big\langle T^{\varphi}_{\varepsilon_2}(1,1), h_K\big\rangle \big\langle f\big\rangle_K \big\langle g\big\rangle_K \big\langle h, h_K\big\rangle \Big| \le c(M).$$

The claim follows by letting $M \to \infty$.

We are done with the proof of Theorem 1.1.

5. Sparse form domination for shifts

Let us first introduce a general framework of trilinear forms. Let $\mathcal{D}$ be a fixed dyadic grid on $\mathbb{R}^n$ and i, j, k be nonnegative integers. Define the trilinear form

$$\begin{aligned}\mathbb{S}^{\rho}(f_1, f_2, f_3) &:= \sum_{Q \in \mathcal{D}} S_Q(f_1, f_2, f_3) \\ &:= \sum_{Q \in \mathcal{D}} \iiint_{Q\times Q\times Q} K_Q(x_1, x_2, x_3) \prod_{j=1}^{3} f_j(x_j)\, dx_1\, dx_2\, dx_3,\end{aligned}$$

where $\rho \ge 0$. Assume it satisfies the following:

A. The kernels $K_Q : Q \times Q \times Q \to \mathbb{C}$ satisfy $\|K_Q\|_{L^\infty} \le |Q|^{-2}$.

B. There exist exponents $p, q, r \in (1, \infty)$ such that $1/p + 1/q = 1/r$ and a constant $\mathcal{B}$ so that for every subcollection $\mathcal{Q} \subset \mathcal{D}$ of dyadic cubes the truncated form

$$\mathbb{S}^{\rho}_{\mathcal{Q}}(f_1, f_2, f_3) := \sum_{Q \in \mathcal{Q}} S_Q(f_1, f_2, f_3).$$

satisfies

$$|\mathbb{S}^{\rho}_{\mathcal{Q}}(f_1, f_2, f_3)| \le \mathcal{B}\|f_1\|_{L^p}\|f_2\|_{L^q}\|f_3\|_{L^{r'}}.$$

C. K_Q is constant on sets of the form $Q_1 \times Q_2 \times Q_3$, where $Q_i^{(\rho+1)} = Q$.

It can easily be seen that trilinear forms associated to both cancellative bilinear shifts and paraproducts fall into the above class of forms. Corollary 1.2 follows from Thereom 1.1 by using two results from this section, namely Proposition 5.1 and Corollary 5.8.

We state the next proposition for only dyadic grids without quadrants – these are dyadic grids where every sequence of cubes I_k with $I_k \subsetneq I_{k+1}$ satisfy $\mathbb{R}^n = \bigcup_k I_k$. Since almost every dyadic grid has this property, this generality is already enough for us to conclude everything we need. Of course, the proposition would hold in every grid but since this is not needed, we prefer this technical simplification.

5.1. **Proposition.** *Let $\eta \in (0,1)$, $\mathcal{D}$ be a dyadic grid without quadrants and f_1, f_2, f_3 be compactly supported and bounded functions. Then there exists an η-sparse collection $\mathcal{S} = \mathcal{S}(f_1, f_2, f_3, \eta) \subset \mathcal{D}$, so that for all $\mathbb{S}^\rho$ defined in $\mathcal{D}$ there holds*

$$\mathbb{S}^\rho(f_1, f_2, f_3) \lesssim_\eta (\mathcal{B}+\rho) \sum_{Q \in \mathcal{S}} |Q| \prod_{j=1}^{3} \langle |f_j| \rangle_Q =: (\mathcal{B}+\rho)\Lambda_{\mathcal{S}}(f_1, f_2, f_3). \tag{5.2}$$

Proof. Let $Q_0 \in \mathcal{D}$ be so that it contains the supports of all of the three functions f_j. Define $\mathcal{E}$ to be the collection of maximal cubes $Q \in \mathcal{D}$, $Q \subset Q_0$, such that

$$\max\left(\frac{\langle |f_1| \rangle_Q}{\langle |f_1| \rangle_{Q_0}}, \frac{\langle |f_2| \rangle_Q}{\langle |f_2| \rangle_{Q_0}}, \frac{\langle |f_3| \rangle_Q}{\langle |f_3| \rangle_{Q_0}} \right) > C_0.$$

For $C_0 = C_0(\eta)$ large enough there holds

$$\sum_{Q \in \mathcal{E}} |Q| \le (1-\eta)|Q_0|.$$

The cube Q_0 is the first cube to be included in $\mathcal{S}$, and $E_{Q_0} := Q_0 \setminus \bigcup_{Q \in \mathcal{E}} Q$.

Let $\mathcal{G} = \mathcal{G}(Q_0) := \{Q \in \mathcal{D} \colon Q \subset Q_0 \text{ and } Q \not\subset Q' \text{ for every } Q' \in \mathcal{E}\}$, and for $Q \in \mathcal{D}$ write $\mathcal{D}(Q) = \{R \in \mathcal{D} \colon R \subset Q\}$. Then we have the decomposition

$$\begin{aligned} \mathbb{S}^\rho(f_1, f_2, f_3) &= \sum_{\substack{Q \in \mathcal{D} \\ Q \supsetneq Q_0}} S_Q(f_1, f_2, f_3) + \mathbb{S}^\rho_{\mathcal{G}}(f_1, f_2, f_3) \\ &\quad + \sum_{Q \in \mathcal{E}} \mathbb{S}^\rho_{\mathcal{D}(Q)}(f_1 1_Q, f_2 1_Q, f_3 1_Q), \end{aligned} \tag{5.3}$$

where we applied the fact that the functions are supported in Q_0. The size property $\|K_Q\|_{L^\infty} \le |Q|^{-2}$ of the kernels implies that

$$\Big| \sum_{\substack{Q \in \mathcal{D} \\ Q \supsetneq Q_0}} S_Q(f_1, f_2, f_3) \Big| \le \sum_{\substack{Q \in \mathcal{D} \\ Q \supsetneq Q_0}} \frac{\|f_1\|_{L^1} \|f_2\|_{L^1} \|f_3\|_{L^1}}{|Q|^2} \sim |Q_0| \prod_j \langle |f_j| \rangle_{Q_0}.$$

We will prove the estimate

$$\mathbb{S}^\rho_{\mathcal{G}}(f_1, f_2, f_3) \lesssim_\eta (\mathcal{B} + \rho)|Q_0| \prod_j \langle |f_j| \rangle_{Q_0}. \tag{5.4}$$

From (5.3) and (5.4) it is then seen that the collection $\mathcal{S}$ can be obtained by iterating this process, in the second step beginning with $\mathbb{S}^\rho_{\mathcal{D}(Q)}(f_1 1_Q, f_2 1_Q, f_3 1_Q)$ for some $Q \in \mathcal{E}$. Hence, to conclude the proof, it remains to show (5.4).

We prove (5.4) by performing a Calderón-Zygmund decomposition to f_j with respect to the collection $\mathcal{E}$, obtaining for each $j = 1, 2, 3$ that

$$f_j = g_j + b_j := g_j + \sum_{Q \in \mathcal{E}} b_{j,Q}, \quad b_{j,Q} := \Big(f_j - \langle f_j \rangle_Q\Big) 1_Q.$$

For every $Q \in \mathcal{E}$ there hold the standard properties

$$\|g_j\|_{L^\infty} \lesssim_\eta \langle |f_j| \rangle_{Q_0}, \quad \int_Q b_{j,Q} = 0, \quad \|b_{j,Q}\|_{L^1} \lesssim_\eta |Q| \langle |f_j| \rangle_{Q_0}.$$

Decompose the left hand side of (5.4) into eight parts:

$$\mathbb{S}^\rho_{\mathcal{G}}(g_1, g_2, g_3),\ \mathbb{S}^\rho_{\mathcal{G}}(b_1, b_2, b_3),\ \mathbb{S}^\rho_{\mathcal{G}}(g_1, g_2, b_3),\ \mathbb{S}^\rho_{\mathcal{G}}(g_1, b_2, g_3),\ \cdots$$

The part with three good functions can be directly estimated via the boundedness of $\mathbb{S}^\rho_{\mathcal{G}}$ and the estimates $\|g_j\|_{L^\infty} \lesssim_\eta \langle |f_j| \rangle_{Q_0}$:

$$|\mathbb{S}^\rho_{\mathcal{G}}(g_1, g_2, g_3)| \le \mathcal{B} \|g_1\|_{L^p} \|g_2\|_{L^q} \|g_3\|_{L^{r'}} \lesssim_\eta \mathcal{B}|Q_0| \prod_j \langle |f_j| \rangle_{Q_0}.$$

In all the other parts, there is at least one bad function involved. All of these terms vanish by assumption C if $\rho = 0$, so assume now that $\rho \ge 1$. By symmetry we consider a term of the form $\mathbb{S}^\rho_{\mathcal{G}}(b_1, h_2, h_3)$, where h_j can either be g_j or b_j. We further decompose $\mathcal{G}$ into ρ subcollections each of which, denoted by $\mathcal{G}'$, satisfies that $\ell(I_1) \ge 2^\rho \ell(I_2)$ whenever $I_1, I_2 \in \mathcal{G}'$, $I_1 \supsetneq I_2$. It suffices to show that

$$|\mathbb{S}^\rho_{\mathcal{G}'}(b_1, h_2, h_3)| \lesssim_\eta |Q_0| \prod_j \langle |f_j| \rangle_{Q_0}. \tag{5.5}$$

Because of the assumption C, the defining property of $\mathcal{G}'$ and the fact that $\int b_{1,Q} = 0$ for every $Q \in \mathcal{E}$, we have that for every $Q \in \mathcal{E}$ there exists at most one $R \in \mathcal{G}'$ such that $Q \subsetneq R$ and $S_R(b_{1,Q}, h_2, h_3) \neq 0$. If such a cube R exists we denote it by $R(Q)$. Therefore,

$$\begin{aligned} |\mathbb{S}^\rho_{\mathcal{G}'}(b_1, h_2, h_3)| &\le \sum_{R \in \mathcal{G}'} \sum_{\substack{Q \in \mathcal{E} \\ R(Q) = R}} |S_R(b_{1,Q}, h_2, h_3)| \\ &\le \sum_{R \in \mathcal{G}'} \sum_{\substack{Q \in \mathcal{E} \\ R(Q) = R}} \frac{\|b_{1,Q}\|_{L^1} \|h_2 1_R\|_{L^1} \|h_3 1_R\|_{L^1}}{|R|^2}, \end{aligned} \tag{5.6}$$

where the size estimate $\|K_R\|_{L^\infty} \le |R|^{-2}$ was applied.

Let $j = 2, 3$ and fix some $R \in \mathcal{G}'$ for the moment. We will prove $\|h_j 1_R\|_{L^1} \lesssim_\eta |R| \langle |f_j| \rangle_{Q_0}$. The L^∞ property of g_j implies that $\|g_j 1_R\|_{L^1} \lesssim_\eta |R| \langle |f_j| \rangle_{Q_0}$. The estimates $\|b_{j,Q}\|_{L^1} \lesssim_\eta |Q| \langle |f_j| \rangle_{Q_0}$ give

$$\|b_j 1_R\|_{L^1} = \sum_{Q \in \mathcal{E} \colon Q \subset R} \|b_{j,Q}\|_{L^1} \lesssim_\eta \sum_{Q \in \mathcal{E} \colon Q \subset R} |Q| \langle |f_j| \rangle_{Q_0} \le |R| \langle |f_j| \rangle_{Q_0}.$$

Now we proceed from (5.6) as

$$\begin{aligned} |\mathbb{S}^\rho_{\mathcal{G}'}(b_1, h_2, h_3)| &\lesssim_\eta \sum_{R \in \mathcal{G}'} \sum_{\substack{Q \in \mathcal{E} \\ R(Q) = R}} \|b_{1,Q}\|_{L^1} \langle |f_2| \rangle_{Q_0} \langle |f_3| \rangle_{Q_0} \\ &\lesssim_\eta |Q_0| \langle |f_1| \rangle_{Q_0} \langle |f_2| \rangle_{Q_0} \langle |f_3| \rangle_{Q_0}. \end{aligned}$$

This completes the proof of (5.5), and hence the proof of the proposition. □

For clarity we give the proof of the following lemma – it is a simple argument that can be extracted from the proof of Lemma 4.7 in Lacey–Mena [8].

5.7. **Lemma.** *Let $0 < \eta_1, \eta_2 < \infty$. Suppose $\mathcal{D}$ is a dyadic grid and $f_1, f_2, f_3 \in L^1$. Then there is an η_2-sparse family $\mathcal{U} = \mathcal{U}(f_1, f_2, f_3, \eta_2) \subset \mathcal{D}$ so that for all η_1-sparse $\mathcal{S} \subset \mathcal{D}$ there holds that*

$$\Lambda_{\mathcal{S}}(f_1, f_2, f_3) \lesssim_{\eta_1, \eta_2} \Lambda_{\mathcal{U}}(f_1, f_2, f_3).$$

Proof. We first construct the family $\mathcal{U}$. Let $C = C(\eta_2) \ge 8^n$ be a large enough constant depending on η_2. For each $k \in \mathbb{Z}$ define

$$\mathcal{U}_k = \Big\{ \text{maximal cubes } Q \in \mathcal{D} \text{ so that } \prod_j \langle |f_j| \rangle_Q > C^k \Big\}.$$

Notice that if $Q \in \mathcal{U}_k$ then

$$C^k < \prod_j \langle |f_j| \rangle_Q \le 8^n \prod_j \langle |f_j| \rangle_{Q^{(1)}} \le 8^n C^k \le C^{k+1}.$$

This means that a given $Q \in \mathcal{D}$ can belong to at most one of the collections $\mathcal{U}_k$. Define

$$\mathcal{U} = \bigcup_{k \in \mathbb{Z}} \mathcal{U}_k.$$

Let us show that this is an η_2-sparse collection. Let $Q \in \mathcal{U}$ and fix k so that $Q \in \mathcal{U}_k$. Notice first that

$$\Big| \bigcup_{\substack{R \in \mathcal{U} \\ R \subsetneq Q}} R \Big| = \Big| \bigcup_{\substack{R \in \mathcal{U}_{k+1} \\ R \subset Q}} R \Big| = \sum_{\substack{R \in \mathcal{U}_{k+1} \\ R \subset Q}} |R|.$$

If $R \in \mathcal{U}_{k+1}$ is such that $R \subset Q$, then

$$\prod_j \langle |f_j| \rangle_R > C^{k+1} \ge \frac{C}{8^n} \prod_j \langle |f_j| \rangle_Q,$$

and so

$$\max_j \frac{\langle |f_j| \rangle_R}{\langle |f_j| \rangle_Q} > \frac{C^{1/3}}{2^n}.$$

This implies that

$$\Big| \bigcup_{\substack{R \in \mathcal{U} \\ R \subsetneq Q}} R \Big| \le \frac{3 \cdot 2^n}{C^{1/3}} |Q| \le (1-\eta_2)|Q|$$

provided $C = C(\eta_2)$ is large enough. It is now clear that the sets

$$E_Q := Q \setminus \bigcup_{\substack{R \in \mathcal{U} \\ R \subsetneq Q}} R, \qquad Q \in \mathcal{U},$$

are disjoint and satisfy $|E_Q| \ge \eta_2 |Q|$, which proves that $\mathcal{U}$ is η_2-sparse.

Consider an arbitrary $\mathcal{S} \subset \mathcal{D}$, which is η_1-sparse. If $Q \in \mathcal{S}$ satisfies $\prod_j \langle |f_j| \rangle_Q \ne 0$, then there is a cube $R \in \mathcal{U}$ so that $Q \subset R$. Let $\pi_{\mathcal{U}} Q$ denote the minimal $R \in \mathcal{U}$ so that $Q \subset R$. Suppose $\pi_{\mathcal{U}} Q \in \mathcal{U}_k$. Then we cannot have $\prod_j \langle |f_j| \rangle_Q > C^{k+1}$ (as otherwise $\pi_{\mathcal{U}} Q$ would not be minimal), and so

$$\prod_j \langle |f_j| \rangle_Q \le C^{k+1} \le C \prod_j \langle |f_j| \rangle_{\pi_{\mathcal{U}} Q} \lesssim_{\eta_2} \prod_j \langle |f_j| \rangle_{\pi_{\mathcal{U}} Q}.$$

Finally, we get

$$\begin{aligned}
\Lambda_{\mathcal{S}}(f_1, f_2, f_3) &= \sum_{R \in \mathcal{U}} \sum_{\substack{Q \in \mathcal{S} \\ \pi_{\mathcal{U}} Q = R}} |Q| \prod_j \langle |f_j| \rangle_Q \\
&\lesssim_{\eta_2} \sum_{R \in \mathcal{U}} \prod_j \langle |f_j| \rangle_R \sum_{\substack{Q \in \mathcal{S} \\ Q^a = R}} |Q| \\
&\lesssim_{\eta_1} \sum_{R \in \mathcal{U}} |R| \prod_j \langle |f_j| \rangle_R = \Lambda_{\mathcal{U}}(f_1, f_2, f_3).
\end{aligned}$$

□

5.8. **Corollary.** *There exists dyadic grids $\mathcal{D}_i$, $i = 1, \ldots, 3^n$, with the following property. Let $\eta_1, \eta_2 \in (0,1)$. Suppose $f_1, f_2, f_3 \in L^1$. Then for some i there exists an η_2-sparse collection $\mathcal{U} = \mathcal{U}(f_1, f_2, f_3, \eta_2) \subset \mathcal{D}_i$, so that for all η_1-sparse collections of cubes $\mathcal{S}$ we have*

$$\Lambda_{\mathcal{S}}(f_1, f_2, f_3) \lesssim_{\eta_1, \eta_2} \Lambda_{\mathcal{U}}(f_1, f_2, f_3). \tag{5.9}$$

Proof. We can let $(\mathcal{D}_i)_i$ be any collection of 3^n dyadic grids with the property that for any cube $P \subset \mathbb{R}^n$ there exists $R \in \bigcup_i \mathcal{D}_i$ so that $P \subset R$ and $\ell(R) \le 6\ell(P)$. Then it is easy to find a $6^{-n}\eta_1$-sparse collections $\mathcal{S}_i \subset \mathcal{D}_i$ (depending on $\mathcal{S}$) so that

$$\Lambda_{\mathcal{S}}(f_1, f_2, f_3) \lesssim_{\eta_1} \sum_i \Lambda_{\mathcal{S}_i}(f_1, f_2, f_3).$$

Let $\mathcal{U}_i = \mathcal{U}_i(f_1, f_2, f_3, \eta_2) \subset \mathcal{D}_i$ be the universal sparse collections given by Lemma 5.7. Then we have that

$$\Lambda_{\mathcal{S}}(f_1, f_2, f_3) \lesssim_{\eta_1} \sum_i \Lambda_{\mathcal{S}_i}(f_1, f_2, f_3) \lesssim_{\eta_1, \eta_2} \sum_i \Lambda_{\mathcal{U}_i}(f_1, f_2, f_3) \lesssim \Lambda_{\mathcal{U}_{i_0}}(f_1, f_2, f_3)$$

for some i_0. We are done. □

REFERENCES

[1] J. M. Conde–Alonso, G. Rey, A pointwise estimate for positive dyadic shifts and some applications, Math. Ann. 365 (2016) 1111–1135.

[2] A. Culiuc, F. Di Plinio, Y. Ou, Uniform sparse domination of singular integrals via dyadic shifts, Math. Res. Lett., to appear, arXiv:1610.01958, 2016.

[3] I. Holmes, S. Petermichl, B. Wick, Weighted little bmo and two-weight inequalities for Journé commutators, Anal. & PDE, to appear, arXiv:1701.06526, 2017.

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

[5] T. Hytönen, Representation of singular integrals by dyadic operators, and the A_2 theorem, Expo. Math., 35 (2017), no. 2, 166–205.

[6] T. Hänninen, T. Hytönen, Operator-valued dyadic shifts and the $T(1)$ theorem, Monatsh. Math. 180 (2016) 213–253.

[7] K. Li, K. Moen, W. Sun, The sharp weighted bound for multilinear maximal functions and Calderón–Zygmund operators, J. Fourier Anal. Appl. 20 (2014) 751–765.

[8] M. Lacey, D. Mena, The sparse $T1$ theorem, Houston J. Math., 43 (2017), no. 1, 111–127.

[9] A. Lerner, F. Nazarov, Intuitive dyadic calculus: the basics, preprint, arXiv:1508.05639, 2015.

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

[11] H. Martikainen, E. Vuorinen, Dyadic–probabilistic methods in bilinear analysis, preprint, arXiv:1609.01706, 2016.

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

[13] Y. Ou, S. Petermichl, E. Strouse, Higher order Journé commutators and characterizations of multi-parameter BMO, Adv. Math. 291 (2016) 24–58.

[14] S. Petermichl, Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol. C. R. Acad. Sci. Paris Sér. I Math. 330 (2000), no. 6, 455–460.

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

E-mail address: kli@bcamath.org

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

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

(Y.O.) DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 77 MASSACHUSETTS AVENUE, CAMBRIDGE, MA 02139, USA

E-mail address: yumengou@mit.edu

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

E-mail address: emil.vuorinen@helsinki.fi