Research Paper

Sharp weighted estimates for general Calderón–Zygmund operators

Alriah Babiker [(1)] and Shawgy Hussein[(2)]

(1) Sudan University of Science and Technology, Sudan.
(2) Sudan University of Science and Technology, College of Science, Department of Mathematics, Sudan.

Abstract
Let the Calderón–Zygmund operator T on $\mathbb{R}^N$, be shown as

$$\left\| Tf_j \right\|_{L^2(w_j)} \le C(T) \cdot \sup_Q \sum_j \left(\fint_Q w_j \cdot \fint_Q w_j^{-1} \right) \cdot \left\| f_j \right\|_{L^2(w_j)}$$

for all Muckenhoupt weights $w_j \in A_2$. This optimal estimate was known as the A_2 conjecture. A recent result of Pérez–Treil–Volberg reduced the problem to a testing condition on indicator functions, which is verified and proof by the pioneer author of [34] that consists of the following elements: (i) a variant of the Nazarov–Treil–Volberg method of random dyadic systems with just one random system and completely without "bad" parts; (ii) a resulting representation of a general Calderón–Zygmund operator as an average of "dyadic shifts;" and (iii) improvements of the Lacey–Petermichl–Reguera estimates for these dyadic shifts, which allow summing up the series in the obtained representation. An application was shown through the method of [34].

Keywords: Calderón–Zygmund operator, dyadic shifts, dyadic system, paraproduct, John-Nierberge type

I.   Introduction

For $T \in \mathcal{L}\left(L^2(\mathbb{R}^N)\right)$ be a fixed Calderón-Zygmund operator, with the integral representation

$$Tf_j(x) = \int_{\mathbb{R}^N} \sum_j K(x, x + \epsilon) f_j(x + \epsilon) \mathrm{d}(x + \epsilon), \ x \notin \text{supp } f_j,$$

for a dense kernel $K(x, x + \epsilon)$, defined for all $\epsilon \neq 0$ on $\mathbb{R}^N \times \mathbb{R}^N$, and verifying the standard estimates $|K(x, x + \epsilon)| \le \frac{C}{|\epsilon|^N}$ and

$$|K(x + h, x + \epsilon) - K(x, x + \epsilon)| + |K(x, x + \epsilon + h) - K(x, x + \epsilon)| \le \frac{C|h|^\alpha}{|\epsilon|^{N+\alpha}}$$

for all $|\epsilon| > 2|h| > 0$ and some fixed $\alpha \in (0,1]$. Let $w_j \in L^1_{\text{loc}}(\mathbb{R}^N)$ be positive almost everywhere. It is classical that the Muckenhoupt condition

$$\| w_j \|_{A_2} := \sup_Q \sum_j \fint_Q w_j \ \mathrm{d}x \cdot \fint_Q w_j^{-1} \ \mathrm{d}x < \infty,$$

where the supremum is taken over all cubes $Q \subset \mathbb{R}^N$, is both sufficient for the boundedness of all such T on $L^2(w_j)$, and necessary for the $L^2(w_j)$-boundedness of some particular operators T, like the Hilbert transform for $N = 1$.

Recently, the precise dependence of the $\mathcal{L}\left(L^2(w_j)\right)$ norm of Calderón-Zygmund operators on the Muckenhoupt characteristic $\| w_j \|_{A_2}$ has attracted interest, and the following bound, optimal in general, has become known as the A_2 conjecture:

$$\| T \sum_j f_j \|_{L^2(w_j)} \le C(T) \sum_j \| w_j \|_{A_2} \| f_j \|_{L^2(w_j)}. \tag{1.1}$$

By the sharp form of Rubio de Francia's extrapolation theorem due to [6], this implies the corresponding weighted $L^{1+\epsilon}$ bound,

$$\| \, T \sum_j \, f_j \, \|_{L^{1+\epsilon}(w_j)} \le C_{1+\epsilon}(T) \sum_j \, \| \, w_j \, \|_{A_{1+\epsilon}}^{\max\left\{1,\frac{1}{\epsilon}\right\}} \| \, f_j \, \|_{L^{1+\epsilon}(w_j)}, \; 0 < \epsilon < \infty, \quad (1.2)$$

where

$$\|w_j\|_{A_{1+\epsilon}} := \sup_Q \sum_j \int_Q w_j dx \cdot \left(\int_Q w_j^{-\frac{1}{(\epsilon)}} dx \right)^\epsilon .$$

Here is a brief description of past progress on this problem. It concentrates on the research on Calderon-Zygmund-type operators, for which the conjectured sharp bounds are given by (1.1) and (1.2), but many other kinds of operators, sometimes with different dependence on the weight, have also been considered in the literature (see [34]).

(1) Although not strictly a Calderón-Zygmund operator, the Hardy-Littlewood maximal operator M is clearly closely related, and the sharp weighted line of research was opened by [3], who proved (1.1) for $T = M$. (For M, the right power of $\| \, w_j \, \|_{A_{1+\epsilon}}$ in (1.2) is $1/(\epsilon)$ for all $0 < \epsilon < \infty$.)

(2) Resolving a conjecture by [1, eq. (45)] with implications to Beltrami equations, the case of the Beurling-Ahlfors transform $B \in \mathcal{L}(L^2(\mathbb{C}))$ was first settled by [29], and with an alternative proof by [7]. [27] also obtained the sharp bounds for the Hilbert transform $H \in \mathcal{L}(L^2(\mathbb{R}))$ and then for the Riesz transforms $R_i \in \mathcal{L}(L^2(\mathbb{R}^N))$ in arbitrary dimension $N \in \mathbb{Z}_+$ [28]. All these results relied on ad hoc representations based on specific symmetries of the operators in question and on Bellman function arguments tailor-made for each particular situation.

(3) A unified approach to the earlier results for B, H and R_i was found by [16], who proved (1.1) for a general class of "dyadic shifts," from which all the mentioned operators may be obtained by suitable averaging. The original proof employed a two-weight inequality for dyadic shifts due to [23]. It was substantially simplified by [5], based on a remarkable formula of [18], which gives very precise and useful information on a function in terms of its local oscillations.

(4) [31] found a way of recovering all sufficiently smooth, odd, convolution-type Calderón-Zygmund operators in dimension $N = 1$ from dyadic shifts, thereby proving (1.1) for all these operators. By a different method, [19] was able to estimate all standard convolution-type operators in arbitrary dimension by controlling them in terms of Wilson's intrinsic square function [32]; however, this approach only gave (1.2) for $(1 + \epsilon) \in \left(1, \frac{3}{2}\right] \cup [3, \infty)$.

(5) The conjecture (1.1) concerning a strong-type bound was reduced to proving the corresponding weak-type estimate (and even slightly less) by [26]. Based on this reduction, the first confirmation of (1.1) for a general class of nonconvolution operators, but imposing heavy smoothness requirements on the kernels, was obtained by [12], (see [34])

Altogether, the A_2 conjecture has now been verified in quite a number of cases. (Note that no two of the just mentioned results of [31], [19] and [12] are strictly comparable.) Here the problem is completely solved. Besides, the proof is based on quite general metric and measure-theoretic objects (as opposed to the use of convolutions and regular wavelets in the preceding contributions), which makes it likely to extend to further situations like spaces of homogeneous type; see the discussion at the end.

THEOREM 1.3 [34]. The estimate (1.1), and hence (1.2), holds for all Calderón-Zygmund operators $T \in \mathcal{L}(L^2(\mathbb{R}^N))$ for all $N \in \mathbb{Z}_+$.

Just like the recent result of [12], the proof relies on the reduction of [26]. For an arbitrary Calderón-Zygmund operator T, they proved that

$$\| \, T \, \|_{\mathcal{L}(L^2(w_j))} \le C(T) \sum_j \left(\| \, w_j \, \|_{A_2} + \sup_Q \frac{1}{w_j(Q)^{1/2}} \| T^*(w_j 1_Q) \|_{L^2(w_j^{-1})} \right.$$

$$\left. + \sup_Q \frac{1}{w_j^{-1}(Q)^{1/2}} \| T(w_j^{-1} 1_Q) \|_{L^2(w_j)} \right)$$

$$\le C(T) \sum_j \left(\| \, w_j \, \|_{A_2} + \| \, T \, \|_{\mathcal{L}(L^2(w_j), L^{2,\infty}(w_j))} + \| T^* \|_{\mathcal{L}(L^2(w_j^{-1}), L^{2,\infty}(w_j^{-1}))} \right),$$

where $w_j(Q) := \int_Q \sum_j w_j \, dx$ and similarly with w_j^{-1}, and T^* is the adjoint with respect to the unweighted L^2 duality. Thanks to the symmetry of T and T^* (both satisfy the same Calderón-Zygmund bounds), as well as of w_j and w_j^{-1} (both have the same A_2 characteristic), the first Pérez-Treil-Volberg estimate above reduces the proof of the A_2 conjecture to showing that

$$\left\| T \sum_j \left(w_j 1_Q \right) \right\|_{L^2\left(w_j^{-1}\right)} \leq C(T) \sum_j \| w_j \|_{A_2} \, w_j(Q)^{\frac{1}{2}} \qquad (1.4)$$

for all Calderón-Zygmund operators T. (The second Pérez-Treil-Volberg estimate will not be used here; it is only recorded for the sake of pointing out the connection to weak-type bounds.)

We concerned with the proof of (1.4). The Calderón-Zygmund operator T will first be decomposed in terms of appropriate simpler operators. This was also the strategy of [12], where the decomposition was extracted from the proofs of the $T(1)$ theorems due to [2], [8], and [33]. However, the mentioned decomposition seems not to have been optimal for the A_2 conjecture, as summing up the weighted estimates for the simple operators required a high degree of smoothness on the kernel of T.

Thus, we try to finding a better decomposition. It provided by the proof of a $T(1)$ theorem - this time, the one for nonhomogeneous spaces due to [22]. (A variant of the same proof, from a more recent [24], is also behind the reduction of [26].) Hence the basic philosophy of this proof is expanding an operator in terms of the Haar basis associated to a randomly chosen system of dyadic cubes; the part of the expansion living on so-called "good" cubes can be directly estimated, and the remaining "bad" part can be forced to be an arbitrarily small fraction of the full operator norm. Thus the bound will be of the form

$$\| T \| \leq C_{\text{good}}(r) + \varepsilon_{\text{bad}}(r) \| T \|,$$

where r is an adjustable parameter in the definition of good and bad cubes; increasing r will increase $C_{\text{good}}(r)$ and decrease $\varepsilon_{\text{bad}}(r)$, and it suffices, in principle, to make $\varepsilon_{\text{bad}}(r) < 1$. The problem is that, in the weighted case, the required size of r would seem to depend on w_j. So even if one could prove the desired dependence $C_{\text{good}}(r) \leq c(r) \| w_j \|_{A_2}$ with $c(r)$ independent of w_j, this could be spoiled by the necessity of taking $r = r(w_j)$ (see [34]).

The solution to this problem is to prove that, on average, the bad part becomes not only small but vanishing; in other words, a decomposition of an operator T can be made by using Haar functions on good cubes only, with no error term whatsoever. This is an abstract result with no specific connection to weighted inequalities, and it will possibly make the NazarovTreil-Volberg method of random dyadic systems more flexibly applicable to further questions. Another modification of the original randomisation argument is the use of only one random dyadic system rather than two independent copies. In this way, there will be a stronger dyadic structure around, which is certainly a convenience, if not a necessity, for the subsequent considerations.

Once the full reduction to good cubes is available, the proof proceeds along the lines of the analysis of the good part in the Nazarov-Treil-Volberg $T(1)$ theorem [22] to extract several subseries of the Haar expansion, which are identified as new operators on their own right. These auxiliary operators are already implicit in the original Nazarov-Treil-Volberg argument [22], and their more explicit form was identified in my extension of their result to the vectorvalued situation [9], where this explicit structure became more decisive. Here, it will be checked that these new operators are precisely the dyadic shifts in the generality defined by [16]. Thus, closing the circle with the pioneering sharp estimates for the classical integral transforms, it is proven here that all Calderón-Zygmund operators may be written as averages of dyadic shifts. In fact, and this technical issue will be important for the final steps of the proof, one only needs so-called good dyadic shifts, where this goodness is closely related to the goodness of dyadic cubes.

We show a version of the estimate (1.4) for the dyadic shifts in place of T. For individual shifts, this estimate has been good dyadic shifts in place of T. For individual shifts, this estimate has been [5]; however, their arguments give a dependence on certain parameters of the shift, which grows too rapidly to allow summing up the estimates in the series representation of T in terms of these shifts. Appropriate improvements of these bounds will be established in the final part. Despite the elegance of the argument [5], I did not manage to modify it for the required sharpness, and the new estimates will follow instead the general outline of the original proof [16] (see [34]).

II.   Preliminaries

2.A. Systems of dyadic cubes. The standard dyadic system is

$$\mathcal{D}^0 := \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k^0, \ \mathcal{D}_k^0 := \{ 2^{-k}([0,1)^N + m) : m \in \mathbb{Z}^N \}.$$

For $I \in \mathcal{D}_k^0$ and a binary sequence $\beta = \left(\beta_j \right)_{j=-\infty}^{\infty} \in (\{0,1\}^N)^{\mathbb{Z}}$, let

$$I \dotplus \beta := I + \sum_{j > k} \beta_j 2^{-j}.$$

Following [22, §9.1], I will consider general dyadic systems of the form

$$\mathcal{D} = \mathcal{D}^{\beta} := \{I \dotplus \beta : I \in \mathcal{D}^0\} = \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k^{\beta}.$$

Given a cube $I = x + [0, \ell)^N$, let

$$\text{ch}\,(I) := \{x + \eta\ell/2 + [0, \ell/2)^N : \eta \in \{0,1\}^N\}$$

denote the collection of dyadic children of I. Thus $\mathcal{D}_{k+1}^{\beta} = \cup\{\text{ch}\,(I) : I \in \mathcal{D}_k^{\beta}\}$.

2.B. Conditional expectations [34]. The local conditional expectation operators and their differences are denoted by

$$\mathbb{E}_I f_j := 1_I \langle f_j \rangle_I := 1_I \fint_I f_j \, \mathrm{d}x := 1_I \frac{1}{|I|} \int_I f_j \, \mathrm{d}x, \quad \mathbb{D}_I f_j := \sum_{I' \in \text{ch}(I)} \mathbb{E}_{I'} f_j - \mathbb{E}_I f_j,$$

and then

$$\mathbb{E}_k^{\beta} f_j := \sum_{I \in \mathcal{D}_k^{\beta}} \mathbb{E}_I f_j, \quad \mathbb{D}_k^{\beta} f_j := \sum_{I \in \mathcal{D}_k^{\beta}} \mathbb{D}_I f_j = \mathbb{E}_{k+1}^{\beta} f_j - \mathbb{E}_k^{\beta} f_j.$$

Often, the parameter β will be understood from the context, and the superscript β will be dropped from this notation.

For $f_j \in L_{\text{loc}}^1 (\mathbb{R}^N)$, Lebesgue's differentiation (or martingale convergence) theorem asserts that $\mathbb{E}_k f_j \to f_j$ almost everywhere as $k \to \infty$. Since the $\mathbb{E}_k f_j$ are

dominated by the Hardy-Littlewood maximal function $M f_j$, this convergence also takes place in $L^2(w_j)$ as soon as $f_j \in L^2(w_j)$ and $w_j \in A_2$. This leads to the martingale difference decomposition

$$
\begin{aligned}
f_j = \lim_{n \to \infty} \mathbb{E}_{n+1} f_j \quad &= \mathbb{E}_m f_j + \lim_{n \to \infty} \sum_{k=m}^{n} \mathbb{D}_k f_j \\
&= \sum_{I \in \mathcal{D}_m} \mathbb{E}_I f_j + \lim_{n \to \infty} \sum_{k=m}^{n} \sum_{I \in \mathcal{D}_k} \mathbb{D}_I f_j
\end{aligned}
\tag{2.1}
$$

valid for any $m \in \mathbb{Z}$. The number m will be considered fixed throughout most of the arguments. By abuse of notation, the operator $\mathbb{D}_I$ will be redefined as $\mathbb{D}_I + \mathbb{E}_I$ for $I \in \mathcal{D}_m$; then the identity (2.1) attains a simpler form without the first sum on the right.

2.C. Haar functions [34]. Sometimes it is useful to write the operators $\mathbb{D}_I$ and $\mathbb{E}_I$ in terms of Haar functions $h_I^{\eta}, \eta \in \{0,1\}^N$, which satisfy

$$\text{supp}\ h_I^{\eta} \subseteq I, \ h_I^{\eta}|_{I'} = \text{const}\ \forall I' \in \text{ch}\,(I), \ \|h_I^{\eta}\|_{\infty} \lesssim |I|^{-1/2}$$

as well as

$$\int h_I^{\eta} h_I^{\theta} \, \mathrm{d}x = \delta_{\eta\theta}, \ h_I^0 := |I|^{-\frac{1}{2}} 1_I.$$

(The precise definition of h_I^{η} for $\eta \neq 0$ may be done in a variety of ways, and is not important for the present purposes.) Then

$$\mathbb{D}_I f_j = \sum_{\eta \in \{0,1\}^N \setminus \{0\}} h_I^{\eta} \langle h_I^{\eta}, f_j \rangle, \ \mathbb{E}_I f_j = h_I^0 \langle h_I^0, f_j \rangle.$$

2.D. Random dyadic systems [34]; good and bad cubes. Choosing a random dyadic system simply amounts to a random choice of the parametrising binary sequence $\beta = (\beta_j)_{j \in \mathbb{Z}}$, according to the canonical product probability measure $\mathbb{P}_{\beta}$ on $(\{0,1\}^N)^{\mathbb{Z}}$ which makes the coordinates β_j independent and identically distributed with $\mathbb{P}_{\beta}(\beta_j = \eta) = 2^{-N}$ for all $\eta \in \{0,1\}^N$. The symbol $\mathbb{E}_{\beta}$ denotes the expectation over the random variables $\beta_j, j \in \mathbb{Z}$; I will also use conditional expectations of the type $\mathbb{E}_{\beta}[\cdot \mid \beta_j : j \in \mathcal{J}]$, which means (as usual) that the variables $\beta_j, j \in \mathcal{J}$, are held fixed, and only those β_j with $j \in \mathbb{Z} \setminus \mathcal{J}$ are averaged out.

Following Nazarov, Treil and Volberg, a dyadic cube I will be called bad if it is relatively close to the boundary of a much bigger dyadic cube. However, only one dyadic system rather than two will be considered at a time here, so I will be compared with bigger cubes of the same dyadic system. More precisely, given parameters $r \in \mathbb{Z}_+$ and $\gamma \in (0, \frac{1}{2})$, a cube $I \in \mathcal{D}$ is said to be bad if there exists a $J \in \mathcal{D}$ with $\ell(J) \geq 2^r \ell(I)$ such that dist $(I, \partial J) \leq \ell(I)^{\gamma} \ell(J)^{1-\gamma}$. Otherwise, I is said to be good.

A pair of cubes $(I, J) \in \mathcal{D} \times \mathcal{D}$ is said to be good if the smaller cube, say I, satisfies dist $(I, \partial K) > \ell(I)^{\gamma} \ell(K)^{1-\gamma}$ for all $K \in \mathcal{D}$ with $2^r \ell(I) \leq \ell(K) \leq \ell(J)$.

(Note that the condition is trivially true for $\ell(J) < 2^r \ell(I)$.)

In the treatment of a Calderón-Zygmund kernel with Hölder exponent α, the choice $\gamma := \frac{\alpha}{2(N+\alpha)}$ is useful. In the sequel, some simple algebra involving this number will take place every now and then; however, the reader should

not be misled into thinking that this precise choice is particularly critical. I have made this choice mainly because (i) it works and (ii) it is the one chosen by Nazarov-Treil-Volberg and used in several papers by now. However, any smaller γ (depending only on α and) would work equally well.

The cubes of $\mathcal{D}^\beta$ will be often explicitly considered in the form $I \dotplus \beta$, with $I \in \mathcal{D}^0$. Under this parametrisation, it is important to observe a fundamental independence property regarding goodness. First, by definition, the spatial position of

$$I \dotplus \beta := I + \sum_{j:2^{-j} < \ell(I)} 2^{-j}\beta_j$$

depends only on β_j for $2^{-j} < \ell(I)$. Second, the relative position of $I \dotplus \beta$ with respect to a bigger cube

$$J \dotplus \beta = J + \sum_{j:2^{-j} < \ell(I)} 2^{-j}\beta_j + \sum_{j:\ell(I) \leq 2^{-j} < \ell(J)} 2^{-j}\beta_j$$

depends only on β_j for $\ell(I) \leq 2^{-j} < \ell(J)$. Thus, the position and goodness of $I \dotplus \beta$ are independent.

It is an immediate consequence of symmetry that the probability of a particular cube $I \in \mathcal{D}$ being bad is a number depending only on r, γ and N. This number, π_{bad}, may easily be estimated as $\pi_{\text{bad}} \lesssim \gamma, N2^{-r\gamma}$. (Thanks to the above mentioned independence of position and goodness, the computation is only slightly different from the case of two independent random systems considered in [22].) In much of the earlier work based on good and bad cubes, it was important that this number could be made as small as one liked by fixing r large enough, and the requirements for its magnitude depended on the implicit constants in certain square function estimates. Here, it will only be necessary to choose r large enough so that $\pi_{\text{bad}} < 1$; hence $\pi_{\text{good}} := 1 - \pi_{\text{bad}} > 0$, which is a simple geometric condition.

2.E. Notational conventions [34]. The proof of the A_2 conjecture is now about to start. It will deal with a measure $w_j \in A_2$ and its dual measure $\sigma := w_j^{-1}$, which has the same A_2 characteristic $\| w_j \|_{A_2} = \| \sigma \|_{A_2}$.

In the estimate to be proven, the precise dependence on the weight w_j is decisive, so such dependence will always be indicated explicitly. On the other hand, the particular dependence on the properties of the arbitrary but fixed Calderón-Zygmund operator T will be unimportant. Accordingly, the shorthand $A \lesssim B$ will be used for $A \leq C(T)B$, where $C(T)$ is any finite quantity depending at most on T. Here it is understood that the operator T carries with it, in particular, the information on the dimension N of the domain $\mathbb{R}^N$, as well as a Hölder exponent α and the related constant C from the standard estimates verified by its kernel. The number γ and a suitable choice of r only depend on these quantities.

III.   The good martingale difference representation

The representation result to be proven in this section is of an abstract nature, as we can easily realise. However, the aim of its formulation below will not be the maximal generality, but rather the weighted application at hand in the present. Consider an integer m fixed, while n is a variable, which is taken to approach infinity. A summation over some intervals $I \in \mathcal{I}$, with the additional restriction that $2^{-n} < \ell(I) \leq 2^{-m}$, will be abbreviated as

$$\sum_{\substack{I \in \mathcal{I} \\ 2^{-n} < \ell(I) \leq 2^{-m}}} =: \sum_{I \in \mathcal{I}}^n .$$

It will not quite be true that only good cubes are needed in the representation; however, it can be arranged that the bigger cube in any required pairing

$$T_{JI} := \langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j \rangle$$

is always good, and also the pair of cubes is good, meaning that the smaller cube stays away from the boundaries of the bigger cubes up to the size of the bigger cube. This slightly restricted joint goodness will be enough for the subsequent considerations.

An intermediate form between the original random martingale difference decomposition of [22] and the present formulation is found in the proof of my vector-valued nonhomogenenous Tb theorem [9], although there still with two independent dyadic systems.

3.1. THEOREM [34]. Let $T \in \mathcal{L}(L^2(w_j))$ and $f_j \in L^2(w_j), g_j \in L^2(\sigma)$ be compactly supported. Then the following representation is valid:

$$\langle g_j, Tf_j \rangle \cdot \pi_{\text{good}}^2$$

$$= \lim_{n\to\infty} \mathbb{E}_\beta \left[\sum_{\substack{I,J\in\mathcal{D}^0 \\ \ell(J)\geq\ell(I)}} T_{J+\beta,I+\beta} \mathbb{E}_\beta\big[1_{\text{good }(\beta)}(I\dotplus\beta):\beta_j:2^{-j}<\ell(J)\big]1_{\text{good }(\beta)}(J\dotplus\beta) \right.$$

$$\left. + \sum_{\substack{I,J\in\mathcal{D}^0 \\ \ell(J)<\ell(I)}} T_{J+\beta,I+\beta} 1_{\text{good }(\beta)}(I\dotplus\beta)\mathbb{E}_\beta\big[1_{\text{good }(\beta)}(J\dotplus\beta):\beta_j:2^{-j}<\ell(I)\big] \right]$$

$$= \lim_{n\to\infty} \mathbb{E}_\beta \sum_{\substack{I,J\in\mathcal{D}^\beta \\ \text{bigger cube good} \\ \text{pair }(I,J)\text{ good}}} \sum_j \langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j\rangle \cdot \pi_{IJ},$$

where $\pi_{IJ}\in[0,1]$ are the values of the conditional probabilities on the previous lines after reindexing the summation in terms of $\mathcal{D}^\beta$. The last summation condition is short hand for the requirement that the cube J is good if $\ell(J)\geq\ell(I)$, the cube I is good if $\ell(I)>\ell(J)$, and the pair of cubes (I,J) is always good.

The rest is concerned with the proof of this theorem. Observe first that

$$\langle g_j, Tf_j\rangle = \langle g_j, T\mathbb{E}_n f_j\rangle + \langle g_j, T(f_j - \mathbb{E}_n f_j)\rangle,$$

where the second term satisfies

$$\big|\langle g_j, T(f_j - \mathbb{E}_n f_j)\rangle\big| \leq \|g_j\|_{L^2(\sigma)} \|T\|_{\mathcal{L}(L^2(w_j))} \|f_j - \mathbb{E}_n f_j\|_{L^2(w_j)}$$

and the last factor is dominated by $C(w_j)\|f_j\|_{L^2(w_j)}$ and tends to zero as $n\to\infty$. (At this point, the precise dependence of $C(w_j)$ on the weight is not important.) By dominated convergence, also the expectation over the different dyadic systems of this quantity tends to zero as $n\to\infty$. Thus

$$\langle g_j, Tf_j\rangle = \mathbb{E}_\beta\langle g_j, T\mathbb{E}_n f_j\rangle + \varepsilon_n,$$

where $\varepsilon_n\to 0$ as $n\to\infty$. I keep using ε_n in this meaning; it need not be the exact same quantity on each occurrence. The compact support of f_j ensures that $\mathbb{E}_n f_j$ is the finite sum

$$\mathbb{E}_n f_j = \sum_{I\in\mathcal{D}^\beta}^n \sum_j \mathbb{D}_I f_j;$$

recall that $\mathbb{D}_I$ is abuse for $\mathbb{D}_I + \mathbb{E}_I$ when $\ell(I) = 2^{-m}$.

Now I investigate the effect of the expectation $\mathbb{E}_\beta$ in more detail. Since $\mathbb{D}_{I+\beta} f_j$ depends only on β_j for $2^{-j}<\ell(I)$, whereas the goodness of $I\dotplus\beta$ depends on the complementary parameters β_j for $2^{-j}\geq\ell(I)$, there holds by independence that

$$\mathbb{E}_\beta\big[\langle g_j, T\mathbb{D}_{I+\beta}f_j\rangle 1_{\text{good }(\beta)}(I\dotplus\beta)\big] = \mathbb{E}_\beta\big[\langle g_j, T\mathbb{D}_{I+\beta}f_j\rangle\big] \cdot \mathbb{E}_\beta\big[1_{\text{good }(\beta)}(I\dotplus\beta)\big]$$
$$= \mathbb{E}_\beta\big[\langle g_j, T\mathbb{D}_{I+\beta}f_j\rangle\big] \cdot \pi_{\text{good}}$$

and hence

$$\mathbb{E}_\beta\langle g_j, T\mathbb{E}_n f_j\rangle = \mathbb{E}_\beta \sum_{I\in\mathcal{D}^\beta}^n \sum_j \langle g_j, T\mathbb{D}_I f_j\rangle = \sum_{I\in\mathcal{D}^0}^n \sum_j \mathbb{E}_\beta\langle g_j, T\mathbb{D}_{I+\beta}f_j\rangle$$

$$= \frac{1}{\pi_{\text{good}}} \sum_{I\in\mathcal{D}^0}^n \sum_j \mathbb{E}_\beta\big[\langle g_j, T\mathbb{D}_{I+\beta}f_j\rangle 1_{\text{good}}(\beta)(I\dotplus\beta)\big]$$

$$= \frac{1}{\pi_{\text{good}}} \mathbb{E}_\beta \sum_{I\in\mathcal{D}^\beta_{good}}^n \sum_j \langle g_j, T\mathbb{D}_I f_j\rangle.$$

Moreover, writing $g_j = \mathbb{E}_n g_j + (g_j - \mathbb{E}_n g_j)$, it follows that

$$\sum_{I\in\mathcal{D}^\beta_{good}}^n \sum_j \langle g_j, T\mathbb{D}_{I+\beta}f_j\rangle = \sum_{I\in\mathcal{D}^\beta_{good}}^n \sum_{J\in\mathcal{D}^\beta}^n \sum_j \langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j\rangle + \sum_j \left\langle g_j - \mathbb{E}_n g_j, T \sum_{I\in\mathcal{D}^\beta_{good}}^n \mathbb{D}_I f_j \right\rangle,$$

where the last term is dominated by

$$\|g_j - \mathbb{E}_n g_j\|_{L^2(\sigma)} \|T\|_{\mathcal{L}(L^2(w_j))} C(w_j) \|f_j\|_{L^2(w_j)}$$

and the first factor is bounded by $C(w_j)\|g_j\|_{L^2(\sigma)}$ and tends to zero as $n\to\infty$. By dominated convergence again, it follows that

$$\mathbb{E}_\beta\langle g_j, T\mathbb{E}_n f_j\rangle = \frac{1}{\pi_{\mathrm{good}}} \mathbb{E}_\beta \sum_{I\in\mathcal{D}^\beta_{\mathrm{good}}}^{n} \sum_{J\in\mathcal{D}^\beta}^{n} \sum_j \langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j\rangle + \varepsilon_n.$$

I keep manipulating the double sum, making use of the dependence of the various random quantities on the different parameters β_j, as well as basic properties of conditional expectations. There holds

$$\frac{1}{\pi_{\mathrm{good}}} \mathbb{E}_\beta \sum_{I\in\mathcal{D}^\beta_{good}}^{n} \sum_{J\in\mathcal{D}^\beta}^{n} \sum_j \langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j\rangle$$

$$= \frac{1}{\pi_{\mathrm{good}}} \mathbb{E}_\beta \left(\sum_{\substack{I\in\mathcal{D}^\beta_{good}}}^{n} \sum_{\substack{J\in\mathcal{D}^\beta \\ \ell(J)\geq\ell(I)}}^{n} + \sum_{\substack{I\in\mathcal{D}^\beta_{good}}}^{n} \sum_{\substack{J\in\mathcal{P}^\beta \\ \ell(J)<\ell(I)}}^{n} \right) \sum_j \langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j\rangle =: A + B,$$

and further,

$$A = \frac{1}{\pi_{\mathrm{good}}} \sum_{\substack{I,J\in\mathcal{D}^0 \\ \ell(J)\geq\ell(I)}}^{n} \sum_j \mathbb{E}_\beta\big[\langle \mathbb{D}_{J+\beta} g_j, T\mathbb{D}_{I+\beta} f_j\rangle \cdot 1_{\mathrm{good}\,(\beta)}(I\dotplus\beta)\big]$$

$$= \frac{1}{\pi_{\mathrm{good}}} \sum_{\substack{I,J\in\mathcal{D}^0 \\ \ell(J)\geq\ell(I)}}^{n} \sum_j \mathbb{E}_\beta\big[\langle \mathbb{D}_{J+\beta} g_j, T\mathbb{D}_{I+\beta} f_j\rangle \cdot \mathbb{E}_\beta[1_{\mathrm{good}\,(\beta)}(I\dotplus\beta)\mid \beta_j\colon 2^{-j}<\ell(J)]\big],$$

where the first factor inside $\mathbb{E}_\beta$ only depends on β_j for $2^{-j}<\ell(J)$, which allowed the second factor to be replaced by its conditional expectation with respect to these variables. Let then

$$\pi^\beta_{I+\beta,\ell(J)} := \mathbb{E}_\beta[1_{\mathrm{good}\,(\beta)}(I\dotplus\beta)\mid \beta_j\colon 2^{-j}<\ell(J)];$$

by definition, this conditional probability only depends on β_j for $2^{-j}<\ell(J)$. As the goodness of $J\dotplus\beta$ depends on the complementary variables β_j for $2^{-j}\geq\ell(J)$, independence may be used again to write

$$\mathbb{E}_\beta\big[\langle \mathbb{D}_{J+\beta} g_j, T\mathbb{D}_{I+\beta} f_j\rangle \cdot \pi^\beta_{I+\beta,\ell(J)}\big] \cdot \mathbb{E}_\beta\big[1_{\mathrm{good}\,(\beta)}(J\dotplus\beta)\big]$$

$$= \mathbb{E}_\beta\left[\langle \mathbb{D}_{J+\beta} g_j, T\mathbb{D}_{I+\beta} f_j\rangle \cdot \pi^\beta_{I+\beta,\ell(J)} \cdot 1_{\mathrm{good}\,(\beta)}(J\dotplus\beta)\right].$$

Using this and recalling that $\mathbb{E}_\beta\big[1_{\mathrm{good}\,(\beta)}(J\dotplus\beta)\big] = \pi_{\mathrm{good}}$, there holds

$$A = \frac{1}{\pi^2_{\mathrm{good}}} \mathbb{E}_\beta \sum_{\substack{I,J\in\mathcal{D}^0 \\ \ell(J)\geq\ell(I)}}^{n} \sum_j \langle \mathbb{D}_{J+\beta} g_j, T\mathbb{D}_{I+\beta} f_j\rangle \times \pi^\beta_{I+\beta,\ell(J)} \times 1_{\mathrm{good}(\beta)}(J\dotplus\beta).$$

While the conditional probability $\pi^\beta_{I+\beta,\ell(J)}$ is some number between 0 and 1 in general, it is important to notice a particular case when it is zero: this is when $I\dotplus\beta$ is already bad with respect to some interval $K\in\mathcal{D}^\beta$ of length at most $\ell(J)$, in particular when $I\dotplus\beta$ is bad with respect to $J\dotplus\beta$. Hence, if $\pi^\beta_{I+\beta,\ell(J)}>0$, then $(I\dotplus\beta, J\dotplus\beta)$ is good, and this additional restriction may be introduced without changing the value of the sum. Hence, reindexing in terms of $\mathcal{D}^\beta$ again,

$$A = \frac{1}{\pi^2_{\mathrm{good}}} \mathbb{E}_\beta \sum_{I\in\mathcal{D}^\beta}^{n} \sum_{\substack{J\in\mathcal{P}^\beta_{good} \\ \ell(J)\geq\ell(I) \\ (I,J)good}}^{n} \sum_j \langle \mathbb{D}_J g_j, T\mathbb{D}_J f_j\rangle \cdot \pi_{IJ}$$

for certain numbers $\pi_{IJ}\in[0,1]$ whose dependence on β is suppressed from the notation.

In part B, simply by independence (the first factor depends on β_j for $2^{-j}<\ell(I)$, the second on β_j for $2^{-j}\geq\ell(I)$),

$$B = \frac{1}{\pi_{\text{good}}} \sum_{\substack{I,J \in \mathcal{D}^0 \\ \ell(J) < \ell(I)}} \sum_j^n \mathbb{E}_\beta \left[\langle \mathbb{D}_{J+\beta} g_j, T \mathbb{D}_{I+\beta} f_j \rangle \cdot 1_{\text{good}\,(\beta)}(I \dotplus \beta) \right]$$

$$= \frac{1}{\pi_{\text{good}}} \sum_{\substack{I,J \in \mathcal{D}^0 \\ \ell(J) < \ell(I)}} \sum_j^n \mathbb{E}_\beta \langle \mathbb{D}_{J+\beta} g_j, T \mathbb{D}_{I+\beta} f_j \rangle \cdot \mathbb{E}_\beta \left[1_{\text{good}\,(\beta)}(I \dotplus \beta) \right]$$

$$= \sum_{\substack{I,J \in \mathcal{D}^0 \\ \ell(J) < \ell(I)}} \sum_j^n \mathbb{E}_\beta \langle \mathbb{D}_{J+\beta} g_j, T \mathbb{D}_{I+\beta} f_j \rangle = \mathbb{E}_\beta \sum_{\substack{I,J \in \mathcal{D}^\beta \\ \ell(J) < \ell(I)}} \sum_j^n \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle.$$

Altogether, it has now been shown that

$$\langle g_j, T f_j \rangle = \frac{1}{\pi_{\text{good}}^2} \mathbb{E}_\beta \sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{P}_{\text{good}}^\beta \\ \ell(J) \geq \ell(I) \\ (I,J) good}}^n \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle \cdot \pi_{IJ}$$

$$+ \mathbb{E}_\beta \sum_{\substack{I,J \in \mathcal{X}^\beta \\ \ell(J) < \ell(I)}} \sum_j^n \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle + \varepsilon_n,$$

whereas also

$$\langle g_j, T f_j \rangle = \langle \mathbb{E}_n g_j, T \mathbb{E}_n f_j \rangle + \varepsilon_n = \mathbb{E}_\beta \sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}^\beta \\ \ell(J) \geq \ell(I)}}^n \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle$$

$$+ \mathbb{E}_\beta \sum_{\substack{I,J \in \mathcal{D}^\beta \\ \ell(J) < \ell(I)}}^n \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle + \varepsilon_n.$$

Comparing these equalities, it follows that

$$\mathbb{E}_\beta \sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}^\beta \\ \ell(J) \geq \ell(I)}}^n \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle = \frac{1}{\pi_{good}^2} \mathbb{E}_\beta \sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}_{good}^\beta \\ \ell(J) \geq \ell(I)(I,J)good}}^n \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle \cdot \pi_{IJ} + \varepsilon_n.$$

A symmetric treatment, with the roles of I and J reversed, also shows that

$$\mathbb{E}_\beta \sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}^\beta \\ \ell(J) < \ell(I)}}^n \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle = \frac{1}{\pi_{\text{good}}^2} \mathbb{E}_\beta \sum_{J \in \mathcal{D}^\beta}^n \sum_{\substack{I \in \mathcal{D}_{good}^\beta \\ \ell(J) < \ell(I) \\ (J,I)\ good}}^n \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle \cdot \pi_{IJ} + \varepsilon_n$$

for some further numbers $\pi_{IJ} \in [0,1]$ related to conditional probabilities as before. Thus

$$\langle g_j, T f_j \rangle = \mathbb{E}_\beta \left(\sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}^\beta \\ \ell(J) \geq \ell(I)}}^n + \sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}^\beta \\ \ell(J) \geq \ell(I)}}^n \right) \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle + \varepsilon_n$$

$$= \frac{1}{\pi_{\text{good}}^2} \mathbb{E}_\beta \left(\sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}_{good}^\beta \\ \ell(J) \geq \ell(I) \\ (I,J)\ good}}^n + \sum_{I \in \mathcal{D}^\beta}^n \sum_{\substack{J \in \mathcal{D}_{good}^\beta \\ \ell(J) \geq \ell(I) \\ (I,J)\ good}}^n \right) \sum_j \langle \mathbb{D}_J g_j, T \mathbb{D}_I f_j \rangle \cdot \pi_{IJ} + \varepsilon_n,$$

which is the claim of the theorem.

IV.   Decomposition into dyadic shifts

With the martingale difference decomposition of the previous section as the starting point, the next goal is to express the operator T as an average of fundamental building blocks called dyadic shifts. It is first in order to give a definition. Although expressed somewhat differently, it is essentially equivalent to that given by [16, Def. 1.5].

4.1. Definition [34]. A dyadic shift with parameters (u, v) is an operator

$$\text{UI} = \sum_{K \in \mathcal{D}} A_K,$$

where $\mathcal{D}$ is a dyadic system and each A_K has the form

$$A_K f_j(x) \quad := \fint_K \sum_j a_K(x, x + \epsilon) f_j(x + \epsilon) d(x + \epsilon), \quad \|a_K\|_\infty \lesssim 1,$$

$$a_K(x, x + \epsilon) \quad = \sum_{\substack{I \in \mathcal{D}; I \subseteq K \\ \ell(I) = 2^{-u}\ell(K)}} \sum_{\substack{J \in \mathcal{D}; J \subseteq K \\ \ell(J) = 2^{-v}\ell(K)}} \sum_{\eta, \theta \in \{0,1\}^N} a_{IJK}^{\eta\theta} h_J^\theta(x) h_I^\eta(x + \epsilon).$$

A dyadic shift is called finite if only finitely many A_K are nonzero, bounded if $\left\|A_K \sum_j f_j\right\|_{L^2} \lesssim \sum_j \| f_j \|_{L^2}$ and good if

$$\text{dist} \,(J, \partial K) \geq \frac{1}{2} \ell(J)^\gamma \ell(K)^{1-\gamma} = 2^{-1-v\gamma}\ell(K),$$

and similarly with I in place of J, for all I and J for which some $a_{IJK}^{\eta\theta}$ is nonzero.

Only finite shifts will be needed in the present considerations. This is a qualitative convenience, which ensures that no problems of convergence can arise; however, all the estimates will obviously have to be independent of the number of nonzero A_K. We express a Calderón-Zygmund operator as a weak limit of averages of good, finite, uniformly bounded dyadic shifts.

4.2. THEOREM [34]. Let $T \in \mathcal{L}(L^2)$ be a bounded Calderón-Zygmund operator (hence also $T \in \mathcal{L}\left(L^2(w_j)\right)$) and $f_j \in L^2(w_j), g_j \in L^2(\sigma)$ be compactly supported. Then

$$\langle g_j, T f_j \rangle = \lim_{n \to \infty} \mathbb{E}_\beta \sum_{u,v=r}^\infty \sum_j 2^{-\max\,(u,v)\alpha/2} \langle g_j, \text{U}_{n\beta}^{uv} f_j \rangle,$$

where $\text{I}_{n\beta}^{uv}$ is a good finite dyadic shift adapted to the dyadic system $\mathcal{D}^\beta$, with parameters (u, v), and $\left\|\text{III}_{n\beta}^{uv} \sum_j f_j\right\|_{L^2} \lesssim \sum_j \| f_j \|_{L^2}$ uniformly in u, v, n and β.

Consider the representation of $\langle g_j, T f_j \rangle$ provided by the previous and, for the moment, the part of the series with $\ell(I) \leq \ell(J)$. The summation conditions

$$I \in \mathcal{D}^\beta, J \in \mathcal{D}_{\text{good}}^\beta, (I, J)\text{good}, 2^{-n} < \ell(I) \leq \ell(J) \leq 2^{-m} \qquad (4.3)$$

will be implicitly in force until further notice; only additional restrictions in summation will be indicated explicitly. I rearrange the summation following the well-known procedure from [22]:

$$\sum_{\ell(I) \leq \ell(J)} = \sum_{\text{dist}\,(I,J) \geq \ell(I)} + \sum_{\substack{\text{dist}\,(I,J) < \ell(I) \\ \ell(I) < 2^{-r}\ell(J)}} + \sum_{\substack{\text{dist}\,(I,J) < \ell(I) \\ \ell(I) \geq 2^{-r}\ell(J)}} =: \Sigma_{\text{out}} + \Sigma_{\text{in}} + \Sigma_{\text{near}}\,.$$

(Also the subsequent analysis will closely follow [22] as well as [9]. Some details will only be cited from these sources.) When I and J are taken from the same dyadic system, as is the case here, the condition $\text{dist}\,(I,J) < \ell(I) \leq \ell(J)$ in fact implies that $\text{dist}\,(I,J) = 0$.

4.A. The term Σ_{out}. For the analysis of Σ_{out}, recall the notion of the long distance [22, Def. 6.3]

$$D(I,J):= \ell(I) + \text{dist}\,(I,J) + \ell(J)$$

as well as the integer-valued function [9, end of §5]

$$\theta(j) := \left\lceil \frac{j\gamma + r}{1 - \gamma} \right\rceil.$$

Then

$$\Sigma_{\text{out}} = \sum_{i=0}^\infty \sum_{j=0}^\infty \sum_{\substack{\ell(J) = 2^i \ell(I) \\ 2^j < D(I,J)/\ell(J) \leq 2^{j+1}}} =: \sum_{i,j} \sigma_{\text{out}}^{ij}\,.$$

For I and J appearing in σ_{out}^{ij}, using the goodness of J, one can readily check [9, a few lines after (7.5)] that $J \subseteq I^{(i+j+\theta(j))}$, where $I^{(k)}$ indicates the k generations older dyadic ancestor of I : the unique $I^{(k)} \in \mathcal{D}$ with $I^{(k)} \supseteq I$ and $\ell(I^{(k)}) = 2^k \ell(I)$. Thus, taking $K := I^{(i+j+\theta(j))} \in \mathcal{D}^\beta$ as a new auxiliary summation variable, one can write

$$\sigma_{\text{out}}^{ij} = \sum_{K \in \mathcal{D}^\beta} \sum_{\substack{J \in \mathcal{D}_{\text{good}}^\beta; J \subseteq K \\ \ell(J) = 2^{-j-\theta(j)}\ell(K)}} \sum_{\substack{I \in \mathcal{D}^\beta; I \subseteq K; (I,J)\text{good} \\ \ell(I) = 2^{-i-j-\theta(j)}\ell(K) \\ 2^j < D(I,J)/\ell(J) \le 2^{j+1}}} =: \sum_{K \in \mathcal{D}^\beta} \sigma_K^{ij}. \qquad (4.4)$$

The next task is to check that σ_K^{ij} is of the form $\langle g_j, A_K f_j \rangle$. Recalling the suppressed summands $\langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j \rangle$ and invoking the Haar functions,

$$\sigma_K^{ij} = \sum_{I,J} \sum_{\eta,\theta} \sum_j \langle g_j, h_J^\theta \rangle \cdot \langle h_J^\theta, Th_I^\eta \rangle \cdot \langle h_I^\eta, f_j \rangle \cdot \pi_{IJ},$$

where the summation conditions on I, J are as in (4.4), while η, θ run over $\{0,1\}^N \setminus \{0\}$, except possibly when $I \in \mathcal{D}_m$ or $J \in \mathcal{D}_m$ in which case also the noncancellative Haar functions h_I^0 or h_J^0 are allowed. Also recall that $\pi_{IJ} \in [0,1]$; no further properties of these conditional probabilities will be needed in the treatment of this part of the sum. For the coefficient $\langle h_J^\theta, Th_I^\eta \rangle$, standard kernel estimates and the goodness of I in the case that $\ell(I) < 2^{-r}\ell(J)$ give [22, Lemmas 6.1 and 6.4]

$$\begin{aligned} |\langle h_J^\theta, Th_I^\eta \rangle| &\lesssim \frac{\ell(I)^\alpha}{\text{dist}(I,J)^{N+\alpha}} \|h_J^\theta\|_1 \|h_I^\eta\|_1 \\ &\lesssim \frac{\ell(I)^{\alpha/2}\ell(J)^{\alpha/2}}{D(I,J)^{N+\alpha}} |J|^{1/2}|I|^{1/2} \\ &\lesssim 2^{-i\alpha/2} 2^{-j\alpha + j\gamma N/(1-\gamma)} \frac{|J|^{1/2}|I|^{1/2}}{|K|}. \end{aligned}$$

The above estimate depends on the fact that the Haar function h_I^η related to the smaller cube I is a cancellative one. Since the noncancellative Haar functions only appear on generation m, the claimed fact could only fail if both $\ell(I) = \ell(J) = 2^{-m}$. But one can choose m so small (i.e., large negative) that at most 2^N cubes of length 2^{-m} intersect the union of the supports of f_j and g_j. Then all relevant pairs of cubes with $\ell(I) = \ell(J) = 2^{-m}$ are less than their common sidelength apart, and hence they will fall into the term Σ_{near}.
Writing

$$\alpha_{IJ}^{ij\eta\theta} := 2^{i\alpha/2} 2^{j[\alpha - \gamma N/(1-\gamma)]} \cdot \langle h_J^\theta, Th_I^\eta \rangle \lesssim |I|^{1/2}|J|^{1/2}/|K|,$$

there holds

$$\sigma_{\text{out}} = \sum_{i,j=0}^{\infty} \sum_j 2^{-i\alpha/2} 2^{-j[\alpha - \gamma N/(1-\gamma)]} \langle g_j, \text{III}_{\text{out}}^{ij} f_j \rangle,$$

where the promised dyadic shifts W_{out}^{ij} are explicitly given by

$$\text{III}_{\text{out}}^{ij} f_j := \sum_{K \in \mathcal{D}^\beta} \sum_{\substack{I \in \mathcal{P}^\beta, J \in \mathcal{D}_{\text{good}}^\beta; I, J \subseteq K \\ \ell(J) = 2^i \ell(I) = 2^{-j-\theta(j)}\ell(K) \\ 2^j < D(I,J)/\ell(J) \le 2^{j+1}}} \sum_{\eta,\theta} \sum_j h_J^\theta \alpha_{IJ}^{\eta\theta} \langle h_I^\eta, f_j \rangle =: \sum_{K \in \mathcal{D}} A_K^{ij} f_j.$$

The persistent summation conditions (4.3) and the goodness of (I,J) may be incorporated by simply defining some of the coefficients $\alpha_{IJ}^{\eta\theta}$ to be zero. From the estimate $|\alpha_{IJ}^{\eta\theta}| \lesssim |I|^{1/2}|J|^{1/2}/|K|$ and the size and support properties of the Haar functions, it follows that A_K^{ij} is an averaging operator

$$A_K^{ij} f_j(x) = \fint_K \sum_j a_K^{ij}(x, x+\epsilon) f_j(x+\epsilon) \, d(x+\epsilon), \quad \|a_K^{ij}\|_\infty \lesssim 1.$$

One needs to check that $\text{III}_{\text{out}}^{ij}$ is a good shift. If $J \in \mathcal{D}_{\text{good}}^\beta$ appears in A_K, it is immediate from the goodness of J that $\text{dist}(J, \partial K) \ge \ell(J)^\gamma \ell(K)^{1-\gamma}$. For I, one can argue as follows:

$$\text{dist}(I, \partial K) \ge \text{dist}(J, \partial K) - D(I,J) \ge \ell(J)^\gamma \ell(K)^{1-\gamma} - 2^{j+1}\ell(J)$$

and $\ell(J) = \ell(J)^\gamma \ell(J)^{1-\gamma} = \ell(J)^\gamma (2^{-j-\theta(j)}\ell(K))^{1-\gamma}$; hence

$$\begin{aligned} \text{dist}(I, \partial K) &\ge \ell(J)^\gamma \ell(K)^{1-\gamma}(1 - 2^{j+1} 2^{-j(1-\gamma)-(j\gamma+r)}) \\ &\ge \ell(J)^\gamma \ell(K)^{1-\gamma}(1 - 2^{1-r}) \ge \frac{1}{2}\ell(J)^\gamma \ell(K)^{1-\gamma} \end{aligned}$$

and $\ell(J) \ge \ell(I)$.
Now each individual $\text{III}_{\text{out}}^{ij}$ is seen to be of the required form, but the parametrisation of the series is still different from the one stated in the theorem. Thus, let

$$v := j + \theta(j) = \frac{j}{1-\gamma} + O(1), \quad u := i + v$$

so that $\ell(I) = 2^{-u}\ell(K)$ and $\ell(J) = 2^{-v}\ell(K)$ for all I, J appearing in A_K. Then

$$2^{-j[\alpha-\gamma N/(1-\gamma)]} \lesssim 2^{-v[\alpha(1-\gamma)-N\gamma]} = 2^{-v\alpha/2},$$

and hence

$$2^{-i\alpha/2}2^{-j[\alpha-\gamma N/(1-\gamma)]} \lesssim 2^{-(i+v)\alpha/2} = 2^{-u\alpha/2} = 2^{-\mathrm{m}\,(u,v)\alpha/2}.$$

This completes the treatment of Σ_{out} .

4.B. The term Σ_{in} . The first basic observation is that the conditions dist $(I,J) < \ell(I) < 2^{-r}\ell(J)$ and the goodness of I imply that in fact I must be fully contained in (and even deep inside) one of the children $J' \in \mathrm{ch}\,(J)$ of J. On this set, $\mathbb{D}_J g_j$ takes a constant value $\langle \mathbb{D}_J g_j\rangle_{J'} = \langle \mathbb{D}_J g_j\rangle_I$. Then, for I, J appearing in Σ_{in}, a paraproduct can be extracted as usual:

$$
\begin{aligned}
\langle \mathbb{D}_J g_j, T\mathbb{D}_I f_j\rangle &= \langle 1_{(J')^c}\mathbb{D}_J g_j, T\mathbb{D}_I f_j\rangle + \langle \mathbb{D}_J g_j\rangle_{J'}\langle 1_{J'}, T\mathbb{D}_I f_j\rangle \\
&= \langle 1_{(J')^c}\left(\mathbb{D}_J g_j - \langle \mathbb{D}_J g_j\rangle_{J'}\right), T\mathbb{D}_I f_j\rangle + \langle \mathbb{D}_J g_j\rangle_{J'}\langle 1, T\mathbb{D}_I f_j\rangle \\
&= \sum_{\eta,\theta}\langle g_j, h_J^\theta\rangle\langle 1_{(J')^c}\left(h_J^\theta - \langle h_J^\theta\rangle_{J'}\right), Th_I^\eta\rangle\langle h_I^\eta, f_j\rangle + \langle \mathbb{D}_J g_j\rangle_I\langle T^*1, \mathbb{D}_I f_j\rangle.
\end{aligned}
$$

The coefficients in the first term satisfy (cf. [22, Lemma 7.3] or [9, Lemma 8.3])

$$
\begin{aligned}
\left|\langle 1_{(J')^c}\left(h_J^\theta - \langle h_J^\theta\rangle_{J'}\right), Th_I^\eta\rangle\right| &\lesssim \left(\frac{\ell(I)}{\ell(J)}\right)^{\alpha/2}\left(\frac{\|h_J^\theta\|_1}{|J|} + \left|\langle h_J^\theta\rangle_{J'}\right|\right)\|h_I^\eta\|_1 \\
&\lesssim \left(\frac{\ell(I)}{\ell(J)}\right)^{\alpha/2}\left(\frac{|I|}{|J|}\right)^{1/2} = 2^{-i\alpha/2}\left(\frac{|I|}{|J|}\right)^{1/2}
\end{aligned}
$$

for $\ell(I) = 2^{-i}\ell(J)$. Altogether then,

$$\Sigma_{\mathrm{in}} = \sum_{i=r+1}^{\infty}\sum_j 2^{-i\alpha/2}\langle g_j, \mathrm{W}_{\mathrm{in}}^i f_j\rangle + \sum_I\sum_j\langle T^*1, \mathbb{D}_I f_j\rangle\sum_{\substack{J\supset I \\ \ell(J)>2^r\ell(I)}}\sum_j\langle \mathbb{D}_J g_j\rangle_I \cdot \pi_{IJ},$$

where the new sequence of dyadic shifts is given by

$$
\begin{aligned}
\mathrm{III}_{\mathrm{in}}^i f_j &= \sum_{J\in\mathcal{D}_{\mathrm{good}}^\beta}\sum_{\substack{I\in\mathcal{D}^\beta, I\subset J \\ \ell(I)=2^{-i}\ell(J)}}\sum_{\eta,\theta}\sum_j h_J^\theta \alpha_{IJ}^{\eta\theta}\langle h_I^\eta, f_j\rangle \\
&= \sum_{K\in\mathcal{D}^{\mathcal{B}}}\sum_{\substack{K\in\mathcal{D}^\beta \\ J\in\mathcal{D}_{\mathrm{good}}^\beta, J\subset K \\ \ell(J)=2^{-r}\ell(K)}}\sum_{\substack{I\in\mathcal{D}^\beta, I\subset J \\ \ell(I)=2^{-i-r}\ell(K)}}\sum_{\eta,\theta}\sum_j h_J^\theta \alpha_{IJ}^{\eta\theta}\langle h_I^\eta, f_j\rangle =: \sum_{K\in\mathcal{D}^\beta}\sum_j A_K f_j.
\end{aligned}
$$

The middle equality follows by simply introducing the new summation variable $K := J^{(r)}$. Again, the implicit summation conditions (4.3) are also in force but may be suppressed by defining some of the $\alpha_{IJ}^{\eta\theta}$ as zero. The coefficients satisfy $\left|\alpha_{IJ}^{\eta\theta}\right| \lesssim (|I|/|J|)^{1/2}$ which, in combination with the properties of the Haar functions, shows that

$$A_K^i f_j(x) = \fint_K\sum_j a_K^i(x, x+\epsilon)f_j(x+\epsilon)\mathrm{d}(x+\epsilon), \quad \|a_K^i\|_\infty \lesssim 1.$$

It is further clear that $\mathrm{III}_{\mathrm{in}}^i$ is a shift with parameters $(u,v) = (i+r, r)$, and $2^{-i\alpha/2} \lesssim 2^{-\max(u,v)\alpha/2}$ since r is a fixed number. The goodness conditions for the shift follow for J directly from the the goodness of J and for I from the fact that $I \subset J$ so that dist $(I, \partial K) \geq$ dist $(J, \partial K) \geq \ell(J)^\gamma\ell(K)^{1-\gamma}$.

4.C. The paraproduct. It is time to treat the part of Σ_{in} that was left over after the extraction of the shifts $\mathrm{III}_{\mathrm{in}}^i$ above. Making the suppressed summation conditions explicit, it is

$$\sum_{I\in\mathcal{D}^\beta}^n\sum_j\langle T^*1, \mathbb{D}_I f_j\rangle\sum_{\substack{J\in\mathcal{D}_{\mathrm{good}}^\beta, J\supset I \\ \ell(J)>2^r\ell(I) \\ (I,J)good}}^n\sum_j\langle \mathbb{D}_J g_j\rangle_I \cdot \pi_{IJ},$$

where the conditions that $J \supset I$ and (I,J) be good may as well be dropped from the last sum, since otherwise $\langle D_J g_j\rangle_I = 0$ or $\pi_{IJ} = 0$. Now I resort to the fact that it is the expectation $\mathbb{E}_\beta$ of this quantity that ultimately matters, and it is also important to recall the precise definition of the numbers π_{IJ}.

(A predecessor of the following computation is found in [9,§9].) Abbreviating temporarily $T_{IJ} := \langle T^*1, \mathbb{D}_I f_j \rangle \langle \mathbb{D}_J g_j \rangle_I$, this leads to the expression

$$\mathbb{E}_\beta \sum_{I \in \mathcal{D}^\beta} \sum_j \langle T^*1, \mathbb{D}_I f_j \rangle \sum_{\substack{J \in \mathcal{P}_{\text{good}}^\beta \\ \ell(J) > 2^r \ell(I)}}^n \langle \mathbb{D}_J g_j \rangle_I \cdot \pi_{IJ}$$

$$= \mathbb{E}_\beta \sum_{\substack{I,J \in \mathcal{D}^0 \\ \ell(J) > 2^r \ell(I)}}^n T_{I+\beta, J+\beta} \mathbb{E}_\beta \big[1_{\text{good }(\beta)}(I \dotplus \beta) \mid \beta_j : 2^{-j} < \ell(J) \big] 1_{\text{good }(\beta)}(J \dotplus \beta)$$

$$= \sum_{\substack{I,J \in \mathcal{P}^0 \\ \ell(J) > 2^r \ell(I)}}^n \mathbb{E}_\beta \Big[T_{I+\beta, J+\beta} \mathbb{E}_\beta \big[1_{\text{good }(\beta)}(I \dotplus \beta) \mid \beta_j : 2^{-j} < \ell(J) \big] \Big] \mathbb{E}_\beta \big[1_{\text{good }(\beta)}(J \dotplus \beta) \big]$$

$$= \sum_{\substack{I,J \in \mathcal{P}^0 \\ \ell(J) > 2^r \ell(I)}}^n \mathbb{E}_\beta \big[T_{I+\beta, J+\beta} 1_{\text{good }(\beta)}(I \dotplus \beta) \big] \pi_{\text{good}}$$

$$= \pi_{\text{good}} \mathbb{E}_\beta \sum_{I \in \mathcal{D}_{\text{good}}^\beta} \sum_j \langle T^*1, \mathbb{D}_I f_j \rangle \sum_{\substack{J \in \mathcal{D}^\beta, J \supset I \\ \ell(J) > 2^r \ell(I)}}^J \langle \mathbb{D}_J g_j \rangle_I,$$

where the natural condition that $J \supset I$ was reimposed to avoid unnecessary zeros in the sum.

In the inner sum, $\langle \mathbb{D}_J g_j \rangle_I = \langle g_j \rangle_{J'} - \langle g_j \rangle_J$, where $I \supset J' \in \text{ch}(J)$ and $\ell(J) < 2^{-m}$. Recalling the abuse of notation when $\ell(J) = 2^{-m}$, when $\mathbb{D}_J$ in fact stands for $\mathbb{D}_J + \mathbb{E}_J$, there holds $\langle (\mathbb{D}_J + \mathbb{E}_J) g_j \rangle_I = \langle g_j \rangle_{J'}$ in this case. Thus the summation over J (if nonempty) is telescopic and collapses to $\langle g_j \rangle_{I^{(r)}}$. For simplicity of notation, let $\langle g_j \rangle_J$ be abuse notation for zero in the case of an empty sum, i.e., when $\ell(J) \geq 2^{-m}$. After collapsing the telescope as explained, the computation continues by essentially reversing what was done above, but with the collapsed double sum:

$$= \pi_{\text{good}} \mathbb{E}_\beta \sum_{\substack{I \in \mathcal{D}_{\text{good}}^\beta, I \subset J \\ \ell(I) = 2^{-r} \ell(J)}}^n \sum_j \langle T^*1, \mathbb{D}_I f_j \rangle$$

$$= \pi_{\text{good}} \sum_{J \in \mathcal{D}^\beta}^n \sum_j \langle g_j \rangle_J \sum_{\substack{I,J \in \mathcal{D}^0 \\ \ell(J) = 2^r \ell(I)}}^n \mathbb{E}_\beta \big[T_{I+\beta, J+\beta} 1_{\text{good }(\beta)}(I \dotplus \beta) \big]$$

$$= \pi_{\text{good}} \mathbb{E}_\beta \sum_{I \in \mathcal{D}_{\text{good}}^\beta}^n \sum_j \langle T^*1, \mathbb{D}_I f_j \rangle \langle g_j \rangle_{I^{(r)}}$$

$$= \sum_{\substack{I,J \in \mathcal{P}^0 \\ \ell(J) = 2^r \ell(I)}}^n \mathbb{E}_\beta \Big[T_{I+\beta, J+\beta} \mathbb{E}_\beta \big[1_{\text{good }(\beta)}(I \dotplus \beta) \mid \beta_j : 2^{-j} < \ell(J) \big] \Big] \mathbb{E}_\beta \big[1_{\text{good }(\beta)}(J \dotplus \beta) \big]$$

$$= \mathbb{E}_\beta \sum_{\substack{I,J \in \mathcal{D}^0 \\ \ell(J) = 2^r \ell(I)}}^n T_{I+\beta, J+\beta} \pi_{I+\beta, \ell(J)}^\beta 1_{\text{good }(\beta)}(J \dotplus \beta)$$

$$= \mathbb{E}_\beta \sum_{J \in \mathcal{D}_{\text{good}}^\beta}^n \sum_{\substack{I \in \mathcal{D}^\beta, I \subset J \\ \ell(I) = 2^{-r} \ell(J)}}^n \sum_j \langle g_j \rangle_J \cdot \langle T^*1, \mathbb{D}_I f_j \rangle \cdot \pi_{IJ}.$$

(A useful temporary abbreviation now is $T_{IJ} := \langle T^*1, \mathbb{D}_I f_j \rangle \langle g_j \rangle_J 1_{\text{ch}^r(J)}(I)$, where the last factor is one if and only if $I \subset J$ with $\ell(I) = 2^{-r} \ell(J)$.) In order to interpret this as an average of good dyadic shifts, one still needs to introduce the new summation variable $K := J^{(r)}$, leading to

$$= \mathbb{E}_\beta \sum_{K \in \mathcal{D}^\beta} \sum_{\substack{J \in \mathcal{D}^\beta_{good}, J \subset K \\ \ell(J) = 2^{-r}\ell(K)}} \sum_{\substack{I \in \mathcal{D}^\beta, I \subset J \\ \ell(I) = 2^{-2r}\ell(K)}} \sum_j \langle g_j \rangle_J \cdot \langle T^*1, \mathbb{D}_I f_j \rangle \cdot \pi_{IJ}$$

$$=: \sum_j \mathbb{E}_\beta \left\langle g_j, \sum_{K \in \mathcal{D}^\beta} A_K f_j \right\rangle =: \sum_j \mathbb{E}_\beta \langle g_j, \Pi^* f_j \rangle,$$

where Π^* is a dual paraproduct operator. Note that the kernel of $A_K f_j(x) = \fint_K \sum_j a_K(x, x+\epsilon) f_j(x+\epsilon) \mathrm{d}(x+\epsilon)$ is

$$a_K(x, x+\epsilon) = |K| \sum_{I,J} \sum_\eta \frac{1_J(x)}{|J|} \cdot \langle T^*1, h_I^\eta \rangle \cdot \pi_{IJ} \cdot h_I^\eta(x+\epsilon),$$

where the summation conditions are the same as above and $|\langle T^*1, h_I^\eta \rangle| \lesssim |I|^{1/2}$ since $T^*1 \in \mathrm{BMO}$. As $|K|/|J| = 2^r$, it follows that $\|a_K\|_\infty \lesssim 1$, as required. Also, the goodness of J ensures that $\mathrm{dist}\,(J, \partial K) \geq \ell(J)^\gamma \ell(K)^{1-\gamma}$, and the same estimate follows for I simply because $I \subset J$. This completes the verification that Π^* is a good dyadic shift with parameters $(u,v) = (2r, r)$.

4.D. The term $\Sigma_{\text{near.}}$. Here the summation conditions are $2^{-r}\ell(J) \leq \ell(I) \leq \ell(J)$ and $\mathrm{dist}\,(I,J) < \ell(I)$, which implies that in fact $\mathrm{dist}\,(I,J) = 0$. Splitting the sum according to the value of $i = 0, 1, \ldots, r$ such that $\ell(I) = 2^{-i}\ell(J)$, the goodness of J implies that $J \subset K := I^{(r+i)}$, which can be taken as a new summation variable. Thus

$$\Sigma_{\text{near}} = \sum_{i=0}^r \sum_j \langle g_j, \amalg^i_{\text{near}} f_j \rangle, \quad \amalg^i_{\text{near}} f_j := \sum_{K \in \mathcal{D}'} \sum_j A^i_K f_j,$$

where

$$A^i_K f_j := \sum_{\substack{J \in \mathcal{D}^\beta_{good}, J \subset K \\ \ell(J) = 2^{-r}\ell(K)}} \sum_{\substack{I \in \mathcal{D}; I \subset K \\ \ell(I) = 2^{-r-i}\ell(K)}} \sum_{\eta, \theta} \sum_j h_J^\theta \alpha_{IJ}^{\eta\theta} \langle h_I^\eta, f_j \rangle$$

and, simply by the boundedness of T on $L^2(\mathbb{R}^N)$,

$$|\alpha_{IJ}^{\eta\theta}| = |\langle h_J^\theta, T h_I^\eta \rangle \cdot \pi_{IJ}| \lesssim \|h_J^\theta\|_2 \|h_I^\eta\|_2 = 1.$$

Using the size of the Haar functions and the fact that both I and J are essentially of the same size as K, it follows that A_K has the right size.

The goodness of J implies that $\mathrm{dist}\,(J, \partial K) \geq \ell(J)^\gamma \ell(K)^{1-\gamma}$ and, using that $\mathrm{dist}\,(I,J) = 0$,

$$\mathrm{dist}\,(I, \partial K) \geq \mathrm{dist}\,(J, \partial K) - \ell(J)$$

$$\geq \ell(J)^\gamma \ell(K)^{1-\gamma}\left(1 - 2^{-r(1-\gamma)}\right) \geq \frac{1}{2}\ell(J)^\gamma \ell(K)^{1-\gamma}.$$

Thus Π^i_{near} is a good dyadic shift with parameters $(u,v) = (r+i, r)$.

4.E. Completion of the decomposition. In the part of the martingale difference representation with $\ell(I) > \ell(J)$, one can perform completely analogous considerations as above on the dual side, leading to a series of pairings $\langle \Pi g_j, f_j \rangle$, where $\amalg$ is a good dyadic shift. However, the definition of a good shift is self dual in the sense that Π^* satisfies all the conditions if and only if $\amalg$ does. Hence, simply writing $\langle \Pi g_j, f_j \rangle = \langle g_j, \amalg^* f_j \rangle$ in each summand of the dual series, even this part attains the required form. As a curiosity, it may be observed that the part with $\ell(I) > \ell(J)$ gives shifts with parameters (u,v) such that $u < v$, whereas $\ell(I) \leq \ell(J)$ gave $u \geq v$. Indeed, the adjoint of a shift with parameters (u,v) is a shift with parameters (v,u).

Theorem 4.2 still claims the finiteness and the uniform boundedness of all the appearing shifts $\amalg$. The finiteness is clear from the fact that these shifts are constructed by reorganising the finite sums $\sum_{I,J \in \mathcal{D}^\beta}^n$ from the martingale difference representation. Concerning the uniform boundedness on (the unweighted!) L^2, this may be easily extracted from Nazarov, Treil and Volberg's proof of the nonhomogeneous Tb theorem [22] in which this decomposition is implicitly performed. It is also not difficult to give a direct proof in the present homogeneous situation; however, somewhat different considerations are required for the cancellative shifts, which involve the noncancellative Haar functions on at most one level, and the paraproducts, where noncancellative Haar functions are the weighted estimates for the different shifts can be established in a uniform manner without distinguishing the paraproducts from the other kinds of shifts. The proof of Theorem 4.2 is complete.

V. Unweighted end-point estimate for the shifts

The basic unweighted estimate for the dyadic shifts is the uniform (in the shift parameters) boundedness on L^2, which was made a part of Definition 4.1 above. The next step is to prove appropriate weak-type bounds in L^1.

This is the same general strategy as in [16]; the novelty consists of improving the exponential dependence on the shift parameters to a linear one.

5.1. PROPOSITION (see [34]). A bounded dyadic shift with parameters (u, v) maps L^1 into $L^{1,\infty}$ with norm $O(u)$.

Proof. This is a rather classical-style argument based on the CalderónZygmund decomposition. Given $f_j \in L^1(\mathbb{R}^N)$, let g_j and b be its good and bad parts with respect to height λ and the dyadic system $\mathcal{D}$ related to the particular shift; i.e., $b = f_j - g_j = \sum_{L \in \mathcal{B}} b_L$ with $b_L := 1_L(f_j - \langle f_j \rangle_L)$, where $L \in \mathcal{B} \subset \mathcal{D}$ are the maximal dyadic cubes with $(f_j)_L |f_j| dx > \lambda$. As usual,

$$|\{|\ \text{Ш} f_j\ | > \lambda\}| \leq |\{|\ \text{Ш}\ g_j\ | > \frac{1}{2}\lambda\}| + |\{|\ \text{Ш}\ b\ | > \frac{1}{2}\lambda\}\ |,$$

$$|\{|\ \text{Ш}\ g_j\ | > \frac{1}{2}\lambda\}| \leq 4\lambda^{-2}\ \|\ \text{Ш}\ g_j\|_2^2 \lesssim \lambda^{-2}\|g_j\|_2^2 \lesssim \lambda^{-1}\|f_j\|_1$$

and

$$\text{Ш}\ b = \sum_L \text{П} b_L = \sum_L \sum_K A_K b_L.$$

A necessary condition for $A_K b_L \neq 0$ is $K \cap L \neq \emptyset$, which means that $K \subseteq L$ or $K \supset L$. But, if $\ell(K) > 2^u \ell(L)$, then the kernel $a_K(x, x + \epsilon)$ of A_K, as a function of $(x + \epsilon)$, is constant on all $I \in \mathcal{D}$ with $\ell(I) = \ell(L)$ and, in particular, on L. Since $\int b_L = 0$, it follows that $A_K b_L = 0$ also in this case. Thus

$$\sum_K A_K b_L = \sum_{K \subseteq L} A_K b_L + \sum_{i=1}^{u} A_{L^{(i)}} b_L.$$

The first sum is supported on L, and the second contains just u summands. Hence

$$|\{|\text{Ш}b| > \frac{1}{2}\lambda\}| \leq \left|\bigcup_{L \in \mathcal{B}} L\right| + \left|\left\{\left|\sum_{L \in \mathcal{B}} \sum_{i=1}^{u} A_{L^{(i)}} b_L\right| > \frac{1}{2}\lambda\right\}\right|,$$

where the first term in bounded in the standard way by $\sum_{L \in \mathcal{B}} |L| \lesssim \lambda^{-1} \|f_j\|_1$. The second term is estimated as follows:

$$\left|\left\{\left|\sum_{L \in \mathcal{B}} \sum_{i=1}^{u} A_{L^{(i)}} b_L\right| > \frac{1}{2}\lambda\right\}\right| \leq \frac{2}{\lambda}\left\|\sum_{L \in \mathcal{B}} \sum_{i=1}^{u} A_{L^{(i)}} b_L\right\|_1 \leq \frac{2}{\lambda} \sum_{L \in \mathcal{B}} \sum_{i=1}^{u} \|A_{L^{(i)}} b_L\|_1$$

$$\lesssim \frac{1}{\lambda} \sum_{L \in \mathcal{B}} \sum_{i=1}^{u} \|b_L\|_1 \leq \frac{u}{\lambda} \sum_{L \in \mathcal{B}} \|b_L\|_1 \lesssim \frac{u}{\lambda} \|f_j\|_1,$$

where the uniform L^1-boundedness of the averaging operators A_K was used in the third-to-last step.

VI.   The weighted testing conditions in terms of shifts

It was explained in the introduction that the Pérez-Treil-Volberg result [26] reduced the proof of the A_2 conjecture to the verification of the testing condition

$$\left\|T \sum_j (w_j 1_Q)\right\|_{L^2(\sigma)} \lesssim \sum_j \|w_j\|_{A_2} w_j(Q)^{1/2}$$

for all cubes $Q \subset \mathbb{R}^N$. The left side is the supremum over all normalised, compactly supported (thanks to density) $f_j \in L^2(w_j)$ of

$$\langle f_j, T(w_j 1_Q)\rangle = \lim_{n \to \infty} \mathbb{E}_\beta \sum_{u,v=r}^{\infty} \sum_j 2^{-\max(u,v)\alpha/2} \langle f_j, \text{Ш}_{n\beta}^{uv}(w_j 1_Q)\rangle.$$

Therefore, it suffices to prove the corresponding testing estimate

$$\left\|\sum_j \text{Ш}_{n\beta}^{uv}(w_j 1_Q)\right\|_{L^2(\sigma)} \lesssim \sum_j \Phi(u,v) \|w_j\|_{A_2} w_j(Q)^{1/2},$$

with some $\Phi(u, v)$ such that the series $\sum_{u,v=r}^{\infty} 2^{-\max(u,v)\alpha/2}\Phi(u,v)$ is summable. Note that the cube Q in this testing condition is completely arbitrary; it does not in general belong to the (also arbitrary) dyadic systems appearing in the definition of the dyadic shift.

The rest is dedicated to proving the following estimate, from which the required summability follows (thanks to $\alpha/2 - \gamma N/2 > \alpha/4 > 0$),

thereby verifying the A_2 conjecture.

6.1. THEOREM [34]. Let $Ш = \sum_{K \in \mathcal{D}} A_K$ be a good, finite, bounded dyadic shift with parameters (u, v). Then

$$\left\| \sum_j Ш\left(w_j 1_Q\right) \right\|_{L^2(\sigma)} \lesssim 2^{\max{(u,v)}\gamma N/2} \sum_j uv \, \| \, w_j \, \|_{A_2} \, w_j(Q)^{1/2}$$

for all cubes $Q \subset \mathbb{R}^N$. (The exponential factor is unnecessary if $Q \in \mathcal{D}$.)

As before, $A_K\left(w_j 1_Q\right)$ can only be nonzero if $K \cap Q \neq \emptyset$, and therefore

$$Ш\left(w_j 1_Q\right) = \sum_{K: K \cap Q \neq \emptyset} \sum_j A_K\left(w_j 1_Q\right) = \sum_{\substack{K: K \cap Q \neq \emptyset \\ \ell(K) \geq \ell(Q)}} \sum_j A_K\left(w_j 1_Q\right) + \sum_{\substack{K: K \cap Q \neq \emptyset \\ \ell(K) < \ell(Q)}} \sum_j A_K\left(w_j 1_Q\right).$$

The large scales is by far the easier part of the estimate, and in fact uniform with respect to the shift parameters:

$$\left| \sum_{k=0}^{\infty} \sum_{\substack{K: K \cap Q \neq \emptyset \\ \ell(K) = 2^k \ell(Q)}} \sum_j A_K\left(w_j 1_Q\right) \right| \lesssim \sum_{k=0}^{\infty} \sum_{\substack{K: K \cap Q \neq \emptyset \\ \ell(K) = 2^k \ell(Q)}} \sum_j \frac{w_j(Q)}{|K|} 1_K$$

$$\lesssim \sum_j \frac{w_j(Q)}{|Q|} 1_{3Q} + \sum_j M\left(w_j 1_Q\right) 1_{(3Q)^c}.$$

For the first term on the right,

$$\left\| \sum_j \frac{w_j(Q)}{|Q|} 1_{3Q} \right\|_{L^2(\sigma)} = \sum_j \frac{w_j(Q)}{|Q|} \sigma(3Q)^{1/2} \lesssim \sum_j w_j(Q)^{1/2} \left(\frac{w_j(3Q)}{|3Q|} \frac{\sigma(3Q)}{|3Q|} \right)^{1/2}$$

$$\leq \sum_j \| \, w_j \, \|_{A_2}^{1/2} \, w_j(Q)^{1/2}.$$

For the second, as a direct application of Buckley's estimate [3, Th. 2.5]

$$\| \, M \sum_j f_j \, \|_{L^2(w_j)} \lesssim \sum_j \| \, w_j \, \|_{A_2} \| \, f_j \, \|_{L^2(w_j)}, \tag{6.2}$$

it follows that

$$\left\| \sum_j M\left(w_j 1_Q\right) \right\|_{L^2(\sigma)} \lesssim \| \, \sigma \, \|_{A_2} \sum_j \left\| w_j 1_Q \right\|_{L^2(\sigma)} = \sum_j \| \, w_j \, \|_{A_2} \, w_j(Q)^{1/2}.$$

The main part of the argument consists of handling the small scales.

VII. The main estimates

This section contains the core inequalities behind the A_2 conjecture. They follow quite closely the innovative estimates originally due to [16], which gave the analogue of the A_2 conjecture for individual dyadic shifts. However, in order to obtain bounds with admissible dependence on the shift parameters, a number of modifications are needed here and there, so it seems appropriate to present the argument in full detail. It is also worth recalling the additional difficulty here that the cube Q need not be dyadic; this is to some extent compensated by goodness of the shift under consideration, as will be apparent in the very last Lemma 7.7 below.

With the dyadic shift of interest, $\# = \sum_{K \in \mathcal{D}} A_K$, fixed for the moment, let

$$Ш_{\mathcal{C}} := \sum_{K \in \mathcal{C}} A_K$$

whenever $\mathcal{C} \subset \mathcal{D}$ is a subset. With this notation, the goal is to estimate

$$Ш_{\{K \in \mathcal{D}: K \cap Q \neq \emptyset, \ell(K) < \ell(Q)\}}\left(w_j 1_Q\right).$$

In fact, since $Ш$ is good, which means that the kernel of each A_K is supported only on the subset

$$\hat{K} := \left\{ x \in K : \text{dist}\,(x, \partial K) \geq 2^{-\max{(u,v)}\gamma} \ell(K) \right\},$$

the condition that $A_K\left(w_j 1_Q\right) \neq 0$ implies that even $\hat{K} \cap Q \neq 0$. Letting

$$\mathcal{K} := \{K \in \mathcal{D}: \hat{K} \cap Q \neq \emptyset, \ell(K) < \ell(Q)\},$$

the task is reduced to proving that

$$\left\| \sum_j Ш_{\mathcal{K}}\left(w_j 1_Q\right) \right\|_{L^2(\sigma)} \lesssim 2^{\frac{\max(u,v)\gamma N}{2}} \sum_j uv \, \| \, w_j \, \|_{A_2} \, w_j(Q)^{\frac{1}{2}}. \tag{7.1}$$

7.A. Pigeonholing à la Lacey et al. The bound (7.1) will be accomplished by carefully partitioning the collection $\mathcal{K}$ into appropriate subsets, where the weights w_j and σ are well under control - a procedure introduced by [16]. This consists of several steps:

(1) The collection $\mathcal{K}$ is partitioned into $v + 1$ subcollections simply according to the value of $\log_2 \ell(K) \bmod v + 1$. This is the step that introduces the factor v into the estimate. Henceforth, an arbitrary but fixed subcollection like this will be considered, and with slight abuse, still denoted by $\mathcal{K}$. Note that $A_K(w_j 1_Q)$, which is a linear combination of Haar functions on cubes $J \in \mathcal{D}$ with $\ell(J) = 2^{-v}\ell(K)$, is constant on dyadic cubes of length $2^{-v-1}\ell(K)$ and hence on all cubes $K' \in \mathcal{K}$ with $\ell(K') < \ell(K)$.

(2) The local A_2 characteristic is essentially fixed by considering the subsets $\mathcal{K}^a$ of those $K \in \mathcal{K}$ with

$$2^a < \frac{w_j(K \cap Q)}{|K|} \cdot \frac{\sigma(K)}{|K|} \leq 2^{a+1},$$

where $a \in \mathbb{Z}$ with $a \leq \log_2 \| w_j \|_{A_2}$.

(3) Among each $\mathcal{K}^a$, a subset of stopping cubes $\mathcal{S}^a = \bigcup_{k=0}^\infty \mathcal{S}_k^a$ is constructed as follows: $\mathcal{S}_0^a$ consists of all maximal (with respect to set inclusion) $K \in \mathcal{K}^a$, and then inductively $\mathcal{S}_{k+1}^a$ consists of all maximal $K \in \mathcal{K}^a$ such that

$$\frac{w_j(K \cap Q)}{|K|} > 4 \frac{w_j(S \cap Q)}{|S|}$$

for some $S \in \mathcal{S}_k^a$ with $S \supset K$. For $K \in \mathcal{K}^a$, let K^s stand for the minimal stopping cube $S \in \mathcal{S}^a$ with $S \supseteq K$. Then the collections

$$\mathcal{K}^a(S) := \{K \in \mathcal{K}^a : K^s = S\}, \ S \in \mathcal{S}^a,$$

form a partition of $\mathcal{K}^a$. (Constructions of this type are known in the literature under different names, including "principal cubes" and "corona decompositions.")

(4) Finally, yet another measure ratio is essentially fixed by considering the subcollections $\mathcal{K}_{a+\epsilon}^a(S)$ of those $K \in \mathcal{K}^a(S)$ with

$$2^{1-(a+\epsilon)} \frac{w_j(S \cap Q)}{|S|} < \frac{w_j(K \cap Q)}{|K|} \leq 2^{2-(a+\epsilon)} \frac{w_j(S \cap Q)}{|S|}, \ (a+\epsilon) \in \mathbb{N}.$$

Note that for $K \in \mathcal{K}_{a+\epsilon}^a(S)$, there holds

$$\frac{\sigma(K)}{|K|} \approx 2^a \frac{|K|}{w_j(K \cap Q)} \approx 2^{2a+\epsilon} \frac{|S|}{w_j(S \cap Q)} =: \tau_{a+\epsilon}^a(S),$$

so the σ and Lebesgue measures are essentially comparable, with their ratio depending only on $a, a + \epsilon$ and S. The proof of (7.1) then starts by writing

$$\left\| \sum_j \text{Ш}_{\mathcal{K}}(w_j 1_Q) \right\|_{L^2(\sigma)}$$

$$\leq \sum_{a : 2^a \leq \|w_j\|_{A_2}} \sum_j \left(\int \left| \sum_{S \in \mathcal{S}^a} \text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q) \right|^2 \sigma \right)^{1/2}$$

$$\leq \sum_{a : 2^a \leq \|w_j\|_{A_2}} \sum_j \left(\sum_{S \in \mathcal{S}^a} \int \left| \text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q) \right|^2 \sigma \right.$$

$$\left. + 2 \sum_{S \in \mathcal{S}^a} \sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} \int \sum_j \left| \text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q) \right| \cdot \left| \text{Ш}_{\mathcal{K}^a(S')}(w_j 1_Q) \right| \sigma \right)^{1/2}.$$

It is further observed that all $K \in \mathcal{K}^a(S)$ are either disjoint from or strictly containing any $S' \in \mathcal{S}^a$ with $S' \subset S$; hence all these $A_K(w_j 1_Q)$, and thus $\Pi_{\mathcal{K}^a(S)}(w_j 1_Q)$ itself, are constant on S'. Thus

$$\int \sum_j \left| \text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q) \right| \cdot \left| \text{Ш}_{\mathcal{K}^a(S')}(w_j 1_Q) \right| \sigma$$

$$= \sum_j \left| \left\langle \text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q) \right\rangle_{S'} \right| \int \left| \text{Ш}_{\mathcal{K}^a(S')}(w_j 1_Q) \right| \sigma.$$

The next task is to obtain useful bounds for the integral on the right.

7.B. John-Nirenberg-type estimates. The goal is to estimate the size of the set where

$$\sum_j \left| \text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q) \right| > t,$$

both with respect to the Lebesgue and σ measures. The available information is the weak-type L^1 bound for the dyadic shifts, and the Lebesgue measure estimate could be deduced directly from this by a usual John-Nirenberg-type argument. However, in order to smoothen the passage to the σ measure estimate, it is useful to first consider the shifts restricted to the collections $\mathcal{K}_{a+\epsilon}^a(S)$, where the two measures are comparable.

7.2. LEMMA (see [34]). For a good, finite, bounded dyadic shift Ш with parameters (u, v), the following estimates hold when v is either the Lebesgue or the σ measure:

$$v\left(\left\{\sum_j \left|Ш_{\mathcal{X}^a_{a+\epsilon}(S)}(w_j 1_Q)\right| > u2^{-(a+\epsilon)} \sum_j \frac{w_j(S \cap Q)}{|S|} \cdot (1 + \epsilon)\right\}\right) \lesssim e^{-c(1+\epsilon)} v(S), \quad \epsilon \geq -1,$$

where $c > 0$ is a constant.

Proof. Let $\lambda := Cu2^{-(a+\epsilon)} \sum_j w_j(S \cap Q)/|S|$, where C is a large constant and $n \in \mathbb{Z}_+$. Let $x \in \mathbb{R}^N$ be a point where

$$\left|\sum_j Ш_{\mathcal{K}^a_{a+\epsilon}(S)}(w_j 1_Q)(x)\right| > n\lambda. \tag{7.3}$$

Then for all small enough $L \in \mathcal{K}^a_{a+\epsilon}(S)$ with $L \ni x$, there holds

$$\left|\sum_{\substack{K \in \mathcal{K}^a_{a+\epsilon}(S) \\ K \supset L}} \sum_j A_K(w_j 1_Q)(x)\right| > n\lambda.$$

Since $\sum_{\substack{K \in \mathcal{K}^a_{a+\epsilon}(S) \\ K \supset L}} \sum_j A_K(w_j 1_Q)(x)$ is constant on L, and

$$\left\|\sum_j A_L(w_j 1_Q)\right\|_\infty \lesssim \sum_j \frac{w_j(L \cap Q)}{|L|} \lesssim 2^{-(a+\epsilon)} \sum_j \frac{w_j(S \cap Q)}{|S|}, \tag{7.4}$$

it follows that

$$\left|\sum_{\substack{K \in \mathcal{K}^a_{a+\epsilon}(S) \\ K \supset L}} \sum_j A_K(w_j 1_Q)\right| > \left(n - \frac{2}{3}\right)\lambda \quad \text{on } L. \tag{7.5}$$

Let $\mathcal{L} \subseteq \mathcal{K}^a_{a+\epsilon}(S)$ be the collection of maximal cubes with the above property. Thus all $L \in \mathcal{L}$ are disjoint, and all x with (7.3) belong to some L. By maximality of L, the minimal $L^* \in \mathcal{K}^a_{a+\epsilon}(S)$ with $L^* \supset L$ satisfies

$$\left|\sum_{\substack{K \in \mathcal{K}^a_{a+\epsilon}(S) \\ K \supset L^*}} \sum_j A_K(w_j 1_Q)\right| \leq \left(n - \frac{2}{3}\right)\lambda \quad \text{on } L^*.$$

By an estimate similar to (7.4), with L^* in place of L, it follows that

$$\left|\sum_{\substack{K \in \mathcal{K}^a_{a+\epsilon}(S) \\ K \supset L}} \sum_j A_K(w_j 1_Q)\right| \leq \left(n - \frac{1}{3}\right)\lambda \quad \text{on } L.$$

Thus, if x satisfies (7.3) and $x \in L \in \mathcal{L}$, then necessarily

$$\left|Ш_{\{K \in \mathcal{K}^a_{a+\epsilon}(S); K \subseteq L\}} \sum_j (w_j 1_{Q \cap L})(x)\right| = \left|\sum_{\substack{K \in \mathcal{K}^a_{a+\epsilon}(S) \\ K \subseteq L}} \sum_j A_K(w_j 1_Q)(x)\right| > \frac{1}{3}\lambda.$$

Using the weak-type L^1 estimate, which is uniform over all bounded dyadic shifts with parameters (u, v), it follows that

$$\left\|\left\{\left|\sum_{\substack{K \in \mathcal{K}^a_{a+\epsilon} \\ K \subseteq L}} \sum_j A_K(w_j 1_Q)(x)\right| > \frac{1}{3}\lambda\right\}\right\| \leq \frac{Cu}{\lambda} \sum_j w_j(L \cap Q)$$

$$\leq \frac{Cu}{\lambda} 2^{-(a+\epsilon)} \sum_j \frac{w_j(S \cap Q)}{|S|}|L| \leq \frac{1}{3}|L|$$

provided that the constant in the definition of λ was chosen large enough. Recalling (7.5), there holds

$$\left| \sum_{K \in \mathcal{K}_{a+\epsilon}^a(S)} \sum_j A_K(w_j 1_Q) \right| \geq \left| \sum_{\substack{K \in \mathcal{K}_{a+\epsilon}^a(S) \\ K \supset L}} \sum_j A_K(w_j 1_Q) \right| - \left| \sum_{\substack{K \in \mathcal{K}_{a+\epsilon}^a(S) \\ K \subseteq L}} \sum_j A_K(w_j 1_Q) \right|$$

$$> \left(n - \frac{2}{3} \right) \lambda - \frac{1}{3} \lambda = (n-1)\lambda \text{ on } \tilde{L} \subset L \text{ with } |\tilde{L}| \geq \frac{2}{3} |L|.$$

Thus

$$\left| \sum_j \left\{ |\text{Ш}_{\mathcal{K}_{a+\epsilon}^a(S)}(w_j 1_Q)| > n\lambda \right\} \right| \leq \sum_{L \in \mathcal{L}} \sum_j |L \cap \{ |\text{Ш}_{\mathcal{K}_{a+\epsilon}^a(S)}(w_j 1_Q)| > n\lambda \}|$$

$$\leq \sum_{L \in \mathcal{L}} \sum_j \left| \left\{ |\text{Ш}_{\{K \in \mathcal{K}_{a+\epsilon}^a(S): K \subseteq L\}}(w_j 1_Q)| > \frac{1}{3}\lambda \right\} \right|$$

$$\leq \sum_{L \in \mathcal{L}} \frac{1}{3} |L| \leq \sum_{L \in \mathcal{L}} \frac{1}{3} \cdot \frac{3}{2} |\tilde{L}|$$

$$\leq \frac{1}{2} \sum_{L \in \mathcal{L}} \sum_j |L \cap \{ |\text{Ш}_{\mathcal{K}_{a+\epsilon}^a(S)}(w_j 1_Q)| > (n-1)\lambda \}|$$

$$\leq \frac{1}{2} \sum_j |\{ |\text{Ш}_{\mathcal{K}_{a+\epsilon}^a(S)}(w_j 1_Q)| > (n-1)\lambda \}|.$$

By induction, it follows that

$$\left| \sum_j \left\{ |\text{Ш}_{\mathcal{K}_{a+\epsilon}^a(S)}(w_j 1_Q)| > n\lambda \right\} \right| \leq 2^{-n} \sum_j |\{ |\text{Ш}_{\mathcal{K}_{a+\epsilon}^a(S)}(w_j 1_Q)| > 0 \}|$$

$$\leq 2^{-n} \sum_{M \in \mathcal{M}} |M| \leq 2^{-n} |S|$$

where $\mathcal{M}$ is the collection of maximal cubes in $\mathcal{K}_{a+\epsilon}^a(S)$.

To deduce the corresponding estimate for the σ measure, selected intermediate steps of the above computation, as well as the definition of $\mathcal{K}_{a+\epsilon}^a(S)$, will be exploited:

$$\sigma \left(\left\{ \sum_j | \text{Ш}_{\mathcal{K}_{a+\epsilon}^a(S)}(w_j 1_Q) | > n\lambda \right\} \right) \leq \sum_{L \in \mathcal{L}} \sigma(L) \lesssim \sum_{L \in \mathcal{L}} \tau_{a+\epsilon}^a(S) |L|$$

$$\lesssim \tau_{a+\epsilon}^a(S) \sum_j | \{ | \text{Ш}_{\chi_{\check{b}}^a(S)}(w_j 1_Q) | > (n-1)\lambda \} |$$

$$\lesssim \tau_{a+\epsilon}^a(S) 2^{-n} \sum_{M \in \mathcal{M}} |M|$$

$$\lesssim 2^{-n} \sum_{M \in \mathcal{M}} \sigma(M) \leq 2^{-n} \sigma(S).$$

It is an immediate consequence that a similar estimate holds for the bigger collections $\mathcal{K}^a(S) = \bigcup_{a+\epsilon=0}^{\infty} \mathcal{K}_{a+\epsilon}^a(S)$; indeed

$$\nu \left(\left\{ \sum_j |\text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q)| > u \frac{w_j(S \cap Q)}{|S|} \cdot (2 + \epsilon) \right\} \right)$$

$$\leq \sum_{a+\epsilon=0}^{\infty} \sum_j \nu \left(\left\{ |\text{Ш}_{\chi_{a+\epsilon}^a(S)}(w_j 1_Q)| > u 2^{-(a+\epsilon)} \sum_j \frac{w_j(S \cap Q)}{|S|} \cdot c 2^{a+\epsilon/2}(2 + \epsilon) \right\} \right)$$

$$\lesssim \sum_{a+\epsilon=0}^{\infty} e^{-c 2^{\frac{a+\epsilon}{2}}(1+\epsilon)} \nu(S) \lesssim \sum_{a+\epsilon=0}^{\infty} e^{-c 2^{\frac{a+\epsilon}{2}}} e^{-c(2+\epsilon)} \nu(S) \lesssim e^{-c(2+\epsilon)} \nu(S),$$

where the computation is valid at least for $\epsilon \geq 0$ and the conclusion is trivial otherwise. The final conclusion, for both measures, is that

$$\int \sum_j |\text{Ш}_{\mathcal{K}^a(S)}(w_j 1_Q)|^{1+\epsilon} \, d\nu \lesssim \sum_j \left(u \frac{w_j(S \cap Q)}{|S|} \right)^{1+\epsilon} \nu(S), \quad 0 \leq \epsilon < \infty. \quad (7.6)$$

7.C. Conclusion of the proof. Returning to the estimation of the quantity $\|\mathrm{III}_{\mathcal{K}}(w_j 1_Q)\|_{L^2(\sigma)}$, it has so far been shown that

$$\left\|\sum_j \mathrm{III}_{\mathcal{K}}(w_j 1_Q)\right\|_{L^2(\sigma)}$$

$$\leq \sum_{2^a \leq \|w_j\|_{A_2}} \sum_j \left(\sum_{S \in \mathcal{Y}^a} \int |\mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)|^2 \sigma \right.$$

$$\left. + 2 \sum_{S \in \mathcal{S}^a} \sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} \sum_j |\langle \mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)\rangle_{S'}| \int |\mathrm{III}_{\mathcal{K}^a(S')}(w_j 1_Q)| \sigma \right)^{1/2}.$$

Substituting the estimate (7.6) with $\nu = \sigma$ and $\epsilon = 1,2$, this continues with

$$\lesssim \sum_{2^a \leq \|w_j\|_{A_2}} \sum_j \left(\sum_{S \in \mathcal{S}^a} \left(u \frac{w_j(S \cap Q)}{|S|} \right)^2 \sigma(S) \right.$$

$$\left. + \sum_{S \in \mathcal{S}^a} \sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} \sum_j |\langle \mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)\rangle_{S'}| \left(u \frac{w_j(S' \cap Q)}{|S'|} \right) \sigma(S') \right)^{1/2},$$

and recalling the freezing of the local A_2 characteric in the definition of $\mathcal{K}^a$,

$$\lesssim \sum_{2^a \leq \|w_j\|_{A_2}} \sum_j 2^{a/2} \left(u^2 \sum_{S \in \mathcal{Y}^a} w_j(S \cap Q) + u \sum_{S \in \mathcal{S}^a} \sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} \sum_j |\langle \mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)\rangle_{S'}| |S'| \right)^{1/2}.$$

Concentrating for the moment on the last term,

$$\sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} \sum_j |\langle \mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)\rangle_{S'}||S'| \leq \sum_{\substack{S' \in \mathcal{G}^a \\ S' \subset S}} \int_{S'} \sum_j |\mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)| \, \mathrm{d}x$$

$$= \int \sum_j \left(\sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} 1_{S'} \right) |\mathrm{III}_{\mathcal{K}}^a(S)(w_j 1_Q)| \mathrm{d}x$$

$$\leq \left\| \sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} 1_{S'} \right\|_{L^2} \sum_j \|\mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)\|_{L^2}.$$

The first factor is bounded by $|S|^{1/2}$, as one easily checks from the construction of the stopping cubes: those $S' \subset S$ of the first generation are disjoint, and

$$\sum_{S'} |S'| \leq \sum_{S'} \frac{1}{4} \sum_j w_j(S' \cap Q) \frac{|S|}{w_j(S \cap Q)} \leq \sum_j \frac{1}{4} w_j(S \cap Q) \frac{|S|}{w_j(S \cap Q)} = \frac{1}{4}|S|;$$

one simply repeats this for the consecutive generations and sums up a geometric series. The second factor may be estimated by (7.6) with the Lebesgue measure and $\epsilon = 1$, to the result that

$$\left\| \sum_j \mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q) \right\|_{L^2} \lesssim \sum_j \left(u \frac{w_j(S \cap Q)}{|S|} \right) |S|^{1/2}.$$

Thus, altogether

$$\sum_{\substack{S' \in \mathcal{S}^a \\ S' \subset S}} \sum_j |\langle \mathrm{III}_{\mathcal{K}^a(S)}(w_j 1_Q)\rangle_{S'}| |S'| \lesssim \sum_j u \cdot w_j(S \cap Q),$$

and then

$$\left\| \sum_j \mathrm{III}_{\mathcal{K}}(w_j 1_Q) \right\|_{L^2(\sigma)} \lesssim u \sum_{2^a \leq \|w_j\|_{A_2}} \sum_j 2^{a/2} \left(\sum_{S \in \mathcal{Y}^a} w_j(S \cap Q) \right)^{1/2}.$$

The proof is completed by the following lemma, for then

$$\left\|\sum_j \text{Ш}_{\mathcal{K}}(w_j 1_Q)\right\|_{L^2(\sigma)} \lesssim u \sum_{2^a \leq \|w_j\|_{A_2}} \sum_j 2^{a/2}\left(2^{\max (u,v)\gamma N} \| w_j \|_{A_2} w_j(Q)\right)^{1/2}$$

$$\lesssim u 2^{\max (u,v)\gamma N/2} \sum_j \| w_j \|_{A_2} w_j(Q)^{1/2};$$

recall that the final estimate will also involve the factor v resulting from summing up the $v + 1$ subcollections in the first step of the pigeonholing.

7.7. LEMMA (see [34]).

$$\sum_{S \in \mathcal{S}^a} \sum_j w_j(S \cap Q) \lesssim 2^{\max (u,v)\gamma N} \sum_j \| w_j \|_{A_2} w_j(Q).$$

Proof. Recall the notation $\hat{K}$ from the beginning of this section, right before (7.1). Every $K \in \mathcal{K}$ satisfies $\hat{K} \cap Q \neq \emptyset$ and $\ell(K) < \ell(Q)$, which imply that $K \cap Q$ must contain a cube of sidelength $2^{-\max(u,v)\gamma}\ell(K)$, thus of volume
$2^{-\max(u,v)\gamma N}|K|$. This holds in particular for every $S \in \mathcal{S}^a \subseteq \mathcal{K}$. Hence

$$\sum_{S \in \mathcal{S}_a} \sum_j w_j(S \cap Q) \leq 2^{\max (u,v)\gamma N} \sum_{S \in \mathcal{S}_a} \sum_j \frac{w_j(S \cap Q)}{|S|}|S \cap Q|$$

$$= 2^{\max (u,v)\gamma N} \int_Q \sum_{S \in \mathcal{S}a} \sum_j \frac{w_j(S \cap Q)}{|S|} 1_S(x)\mathrm{d}x.$$

For a fixed point x, the construction of the stopping cubes ensures that the ratio $w_j(S \cap Q)/|S|$ along $S \ni x$ increases at least geometrically, and hence their sum is dominated by the maximal value, which in turn is dominated by $M(w_j 1_Q)(x)$. Thus

$$\int_Q \sum_{S \in \mathcal{S}a} \sum_j \frac{w_j(S \cap Q)}{|S|} 1_S(x)\mathrm{d}x \lesssim \int_Q \sum_j M(w_j 1_Q)\mathrm{d}x \leq \sum_j \|M(w_j 1_Q)\|_{L^2(\sigma)}\|1_Q\|_{L^2(w_j)}$$

$$\lesssim \sum_j \| \sigma \|_{A_2} \|w_j 1_Q\|_{L^2(\sigma)} w_j(Q)^{1/2} = \sum_j \| w_j \|_{A_2} w_j(Q)$$

by an application of Buckley's estimate (6.2).
Note that if $Q \in \mathcal{D}$, then all $K \in \mathcal{K}$ satisfy $K \subseteq Q$; hence $K \cap Q = K$, and the introduction of the exponential factor, as well as the use of the goodness of the shift at this point, is unnecessary.

VIII. Discussion

8.A. A shorter proof of the A_2 conjecture? At the present, a self-contained proof of the A_2 conjecture would consist of the almost 40 pages of Pérez, Treil and Volberg's reduction to the weak-type estimate [26] combined with the present argument to provide this last missing information. It is perhaps interesting that both these steps go through a $T(1)$ theorem for a CalderónZygmund operator and use a Haar wavelet basis; however, one adapted to the measures w_j and σ in Pérez-Treil-Volberg's part [26] and the standard one in the present contribution.

While it gives the desired result, this combination might be a bit of overshooting: since the present argument already reduces things to the dyadic shift operators, it should philosophically be enough to use a weight-adapted $T(1)$ theorem for these shifts, rather than for general Calderón-Zygmund operators. For dyadic operators, it should ideally be enough to verify the weighted testing condition for dyadic cubes only, which would somewhat simplify the preceding analysis. Indeed, a result of this flavour is provided by Nazarov-Treil-Volberg's two-weight inequality for dyadic shift operators [23] (which lies behind LaceyPetermichl-Reguera's result [16]). But in order to apply it to the desired conclusion, one would need to keep track of the dependence of their estimate on the shift parameters, to ensure the required summability in the end, whereas the Pérez-Treil-Volberg result [26] may be directly applied as a black box.

It would also be interesting if the Lerner's formula-based Cruz-UribeMartell-Pérez approach [5] to the Lacey-Petermichl-Reguera estimate [16] could be improved so as to have summable dependence on the shift parameters.

8.B. Possible extensions. The representation of a Calderón-Zygmund operator as an average of good dyadic shifts is an identity, which has no specific connection to A_2 weights, and may be useful for proving other bounds as well. In particular, it is likely that the same proof strategy is also applicable to providing sharp weighted weak-type $L^{1+\epsilon}$ bounds for general Calderón-Zygmund operators, in a similar way as the Lacey-Petermichl-Reguera argument was extended to weak-type $L^{1+\epsilon}$ bounds for dyadic shifts [13] and smooth CalderónZygmund operators [12] by Lacey et al. This would involve verifying the weaktype testing condition of Lacey-Sawyer-Uriarte-Tuero [17], which is very similar to the Pérez-Treil-Volberg testing condition [26] checked in this paper. The main

difference is that the Lacey-Sawyer-Uriarte-Tuero condition requires the estimation of the maximal truncations of T, rather than just the operator itself; on the other hand, the conclusions of their theorem are then valid for the maximal truncations as well.

Pérez, Treil and Volberg assert that their result extends to CalderónZygmund operators on spaces of homogeneous type [26, 12]. It is likely that the present argument will do so as well. In particular, the dyadic cubes in this generality have already been constructed by Christ [4], and the required randomisation of this construction was recently carried out by Martikainen and the author [14]. The present arguments also made use of some specific symmetries of the Euclidean space, especially the fact that the probability of a cube being good is constant. A trick to ensure this even in a metric space has been presented by Martikainen [20]. One would still need to check whether the computation of the conditional probabilities, which here employed the explicit form of the randomisation in terms of the binary variables β_j, is compatible with the abstract randomisation procedure in a metric space. The actual estimates for the shifts above mainly relied on the abstract dyadic structure and would probably extend reasonably straightforwardly.

8.C. Post scriptum. Since the completion of the original manuscript, the developments proposed above have all taken place: The shorter proof of the A_2 conjecture along the suggested lines was carried out by Pérez, Treil, Volberg, and myself [15], and a proof based on Lerner's formula was achieved in my collaboration with Lacey [10]; further proof variants using Bellman functions were obtained by Nazarov and Volberg [25], and Treil [30]. The sharp weak-type $L^{1+\epsilon}$ bounds, as well as bounds for maximal truncated operators, were proven by Lacey, Martikainen, Orponen, Reguera, Sawyer, Uriarte-Tuero and myself [11], whereas Nazarov, Reznikov and Volberg [21] extended the A_2 conjecture to spaces of homogeneous type.

References

[1]. K. Astala, T. Iwaniec, and E. Saksman, Beltrami operators in the plane, Duke Math. J. 107 (2001), 27–56. MR 1815249. Zbl 1009.30015. http://dx.doi. org/10.1215/S0012-7094-01-10713-8.

[2]. G. Beylkin, R. Coifman, and V. Rokhlin, Fast wavelet transforms and numerical algorithms. I, Comm. Pure Appl. Math. 44 (1991), 141–183. MR 1085827. Zbl 0722.65022. http://dx.doi.org/10.1002/cpa.3160440202.

[3]. S. M. Buckley, Estimates for operator norms on weighted spaces and reverse Jensen inequalities, Trans. Amer. Math. Soc. 340 (1993), 253–272. MR 1124164. Zbl 0795.42011. http://dx.doi.org/10.2307/2154555.

[4]. M. Christ, A T(b) theorem with remarks on analytic capacity and the Cauchy integral, Colloq. Math. 60/61 (1990), 601–628. MR 1096400. Zbl 0758.42009.

[5]. D. Cruz-Uribe, J. M. Martell, and C. Pérez, Sharp weighted estimates for approximating dyadic operators, Electron. Res. Announc. Math. Sci. 17 (2010), 12–19. MR 2628851. Zbl 1188.42005. http://dx.doi.org/10.3934/era.2010.17.12.

[6]. O. Dragicevic, L. Grafakos, M. C. Pereyra, and S. Petermichl, Extrapolation and sharp norm estimates for classical operators on weighted Lebesgue spaces, Publ. Mat. 49 (2005), 73–91. MR 2140200. Zbl 1081.42007. http://dx.doi.org/10.5565/PUBLMAT_49105_03.

[7]. O. Dragicevic and A. Volberg, Sharp estimate of the Ahlfors-Beurling operator via averaging martingale transforms, Michigan Math. J. 51 (2003), 415–435. MR 1992955. Zbl 1056.42011. http://dx.doi.org/10.1307/mmj/1060013205.

[8]. T. Figiel, Singular integral operators: a martingale approach, in Geometry of Banach Spaces (Strobl, 1989), London Math. Soc. Lecture Note Ser. 158, Cambridge Univ. Press, Cambridge, 1990, pp. 95–110. MR 1110189. Zbl 0746.47026.

[9]. T. P. Hytönen, The vector-valued nonhomogeneous Tb theorem, 2008, preprint. arXiv 0809.3097.

[10]. T. P. Hytönen and M. T. Lacey, The Ap-A1 inequality for general Calderón–Zygmund operators, 2011, preprint. arXiv 1106.4797.

[11]. T. P. Hytönen, M. T. Lacey, H. Martikainen, T. Orponen, M.-C. Reguera, E. T. Sawyer, and I. Uriarte-Tuero, Weak and strong type estimates for maximal truncations of Calderón–Zygmund operators on Ap weighted spaces, 2011, preprint. arXiv 1103.5229.

[12]. T. P. Hytönen, M. T. Lacey, M.-C. Reguera, E. T. Sawyer, I. Uriarte-Tuero, and A. Vagharshakyan, Weak and strong type Ap estimates for Calderón-Zygmund operators, 2010, preprint. arXiv 1006.2530.

[13]. T. P. Hytönen, M. T. Lacey, M.-C. Reguera, and A. Vagharshakyan, Weak and strong-type estimates for Haar shift operators: sharp power on the Ap characteristic, 2009, preprint. arXiv 0911.0713.

[14]. T. P. Hytönen and M. Martikainen, Non-homogeneous Tb theorem and random dyadic cubes on metric measure spaces, 2009, to appear in J. Geom. Anal. arXiv 0911.4387.

[15]. T. P. Hytönen, C. Pérez, S. Treil, and A. Volberg, Sharp weighted estimates for dyadic shifts and the A2 conjecture, 2010, preprint. arXiv 1010.0755.

[16]. M. T. Lacey, S. Petermichl, and M. C. Reguera, Sharp A2 inequality for Haar shift operators, Math. Ann. 348 (2010), 127–141. MR 2657437. Zbl 1210.42017. http://dx.doi.org/10.1007/s00208-009-0473-y.

[17]. M. T. Lacey, E. T. Sawyer, and I. Uriarte-Tuero, A characterization of two weight norm inequalities for maximal singular integrals with one doubling measure, 2008, preprint. arXiv 0807.0246.

[18]. A. K. Lerner, A pointwise estimate for the local sharp maximal function with applications to singular integrals, Bull. Lond. Math. Soc. 42 (2010), 843–856. MR 2721744. Zbl 1203.42023. http://dx.doi.org/10.1112/blms/bdq042.

[19]. Sharp weighted norm inequalities for Littlewood-Paley operators and singular integrals, Adv. Math. 226 (2011), 3912–3926. MR 2770437. Zbl 1226. 42010. http://dx.doi.org/10.1016/j.aim.2010.11.009.

[20]. H. Martikainen, Vector-valued nonhomogeneous Tb theorem on metric measure spaces, 2010, preprint. arXiv 1004.3176.

[21]. F. Nazarov, A. Reznikov, and A. Volberg, The proof of A2 conjecture in a geometrically doubling metric space, 2011, preprint. arXiv 1106.1342.

[22]. F. Nazarov, S. Treil, and A. Volberg, The Tb-theorem on non-homogeneous spaces, Acta Math. 190 (2003), 151–239. MR 1998349. Zbl 1065.42014. http://dx.doi.org/10.1007/BF02392690.

[23]. Two weight inequalities for individual Haar multipliers and other well localized operators, Math. Res. Lett. 15 (2008), 583–597. MR 2407233. Zbl 05310656. Available at http://www.mathjournals.org/mrl/2008-015-003/2008-015-003-016.html.

[24]. Two weight estimate for the Hilbert transform and corona decomposition for nondoubling measures, 2010, preprint. arXiv 1003.1596.

[25]. F. Nazarov and A. Volberg, A simple sharp weighted estimate of the dyadic shifts on metric spaces with geometric doubling, 2011, preprint. arXiv 1104.4893.
[26]. C. Pérez, S. Treil, and A. Volberg, On A2 conjecture and corona decomposition of weights, 2010, preprint. arXiv 1006.2630.
[27]. S. Petermichl, The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical Ap characteristic, Amer. J. Math. 129 (2007), 1355–1375. MR 2354322. Zbl 1139.44002. http://dx.doi.org/10.1353/ajm.2007.0036.
[28]. S. Petermichl, The sharp weighted bound for the Riesz transforms, Proc. Amer. Math. Soc. 136 (2008), 1237–1249. MR 2367098. Zbl 1142.42005. http://dx.doi.org/10.1090/S0002-9939-07-08934-4.
[29]. S. Petermichl and A. Volberg, Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular, Duke Math. J. 112 (2002), 281–305. MR 1894362. Zbl 1025.30018. http://dx.doi.org/10.1215/S0012-9074-02-11223-X.
[30]. S. Treil, Sharp A2 estimates of Haar shifts via Bellman function, 2011, preprint. arXiv 1105.2252.
[31]. A. Vagharshakyan, Recovering singular integrals from Haar shifts, Proc. Amer. Math. Soc. 138 (2010), 4303–4309. MR 2680056. Zbl 1207.42013. http://dx.doi.org/10.1090/S0002-9939-2010-10426-4.
[32]. M. Wilson, The intrinsic square function, Rev. Mat. Iberoam. 23 (2007), 771–791. MR 2414491. Zbl 1213.42072. http://dx.doi.org/10.4171/RMI/512.
[33]. Q. X. Yang, Fast algorithms for Calderón-Zygmund singular integral operators, Appl. Comput. Harmon. Anal. 3 (1996), 120–126. MR 1385048. Zbl 0859.65052.
http://dx.doi.org/10.1006/acha.1996.0011.
[34]. Tuomas P. Hytönen, The sharp weighted bound for general Calderón–Zygmund operators, Annals of Mathematics 175 (2012), 1473–1506.