Sharp weighted estimates for dyadic shifts and the A_2 conjecture

By *Tuomas Hytönen* at Helsinki, *Carlos Pérez* at Sevilla, *Sergei Treil* at Providence and *Alexander Volberg* at East Lansing

---

Abstract. We give a self-contained proof of the A_2 conjecture, which claims that the norm of *any* Calderón–Zygmund operator is bounded by the first degree of the A_2 norm of the weight. The original proof of this result by the first author relied on a subtle and rather difficult reduction to a testing condition by the last three authors. Here we replace this reduction by a new weighted norm bound for dyadic shifts – linear in the A_2 norm of the weight and quadratic in the complexity of the shift –, which is based on a new quantitative two-weight inequality for the shifts. These sharp one- and two-weight bounds for dyadic shifts are the main new results of this paper. They are obtained by rethinking the corresponding previous results of Lacey–Petermichl–Reguera and Nazarov–Treil–Volberg. To complete the proof of the A_2 conjecture, we also provide a simple variant of the representation, already in the original proof, of an arbitrary Calderón–Zygmund operator as an average of random dyadic shifts and random dyadic paraproducts. This method of the representation amounts to the refinement of the techniques from non-homogeneous Harmonic Analysis.

1. Introduction

A *Calderón–Zygmund operator* in $\mathbb{R}^d$ is an integral operator, bounded in L^2 and with kernel K satisfying the following growth and smoothness conditions:

(i) $|K(x, y)| \leq \dfrac{C_{\text{cz}}}{|x - y|^d}$ for all $x, y \in \mathbb{R}^d$, $x \neq y$.

(ii) There exists $\alpha > 0$ such that

$$|K(x, y) - K(x', y)| + |K(y, x) - K(y, x')| \leq C_{\text{cz}} \frac{|x - x'|^\alpha}{|x - y|^{d+\alpha}}$$

for all $x, x', y \in \mathbb{R}^d$ such that $|x - x'| < |x - y|/2$.

It is well known (see [12]) that a Calderón–Zygmund operator is bounded in the weighted space $L^2(w)$ if (and for many Calderón–Zygmund operators only if) the weight w satisfies the famous Muckenhoupt A_2 condition

$$(1.1) \qquad \sup_Q \Big(|Q|^{-1} \int_Q w\,dx \Big) \Big(|Q|^{-1} \int_Q w^{-1} dx \Big) =: [w]_{A_2} < \infty.$$

The quantity $[w]_{A_2}$ is called the *Muckenhoupt norm* of the weight w (although it is definitely not a norm).

It has been an old problem to describe how the norm of a Calderón–Zygmund operator in the weighted space $L^2(w)$ depends on the Muckenhoupt norm $[w]_{A_2}$ of w. A conjecture was that for a fixed Calderón–Zygmund operator T its norm is bounded by $C \cdot [w]_{A_2}$, where the constant C depends on the operator T (but not on the weight w). Simple counterexamples demonstrate that for the classical operators like Hilbert transform or Riesz transform, a better estimate than $C \cdot [w]_{A_2}$ is not possible.

This linear (in $[w]_{A_2}$) estimate of the norm has become known as the A_2 *conjecture*.

For the maximal function, the estimate $C \cdot [w]_{A_2}$ was proved by S. Buckley [3]: he also proved that this estimate is optimal for the maximal function. The first result for a singular "integral" operator was due to J. Wittwer [42], who proved the A_2 conjecture for the Haar multipliers. The same result for Beurling–Ahlfors transform (convolution with $\pi^{-1} z^{-2}$ in $\mathbb{C}$) was obtained first by Petermichl–Volberg [34] by using the combination of Bellman function technique and the heat extension, and later by Dragicevic–Volberg [7] via the representation of the Beurling–Ahlfors transform as an average of Haar multipliers over all dyadic lattices.

This result was used in [34] to answer positively an important question in the theory of quasiconformal maps, see [1], about whether a weakly quasiregular map is quasiregular (or equivalently whether there is a self-improvement of a solution of the Beltrami equation in the case of critical exponent).

Then S. Petermichl [32] proved the A_2 conjecture for the Hilbert transform, again using the representation of the Hilbert transform as an average of copies of a simple dyadic operator (the so-called dyadic, or Haar, shift of complexity 1 from [31]). Further cases of the A_2 conjecture were settled in [2, 33].

We should mention here an earlier paper by R. Fefferman and J. Pipher [9], where a linear estimate in terms of stronger A_1 norm of the weight w was obtained for the Hilbert transform. This result found its application in geometric questions pertinent to multi-parameter Harmonic Analysis, in particular for singular operators on the Heisenberg group. The result in [32] is a considerable strengthening of Fefferman–Pipher's theorem. For other types of results on A_1, see [18].

A recent paper [16] by M. Lacey, S. Petermichl and M. Reguera established the A_2 conjecture for general dyadic shifts. Another proof of the linear bound for dyadic shifts was obtained in Cruz-Uribe–Martell–Pérez [4, 5] in a very beautiful and concise approach based on a remarkable "formula" by Lerner [17]. Thus, the conjecture was proved for all operators which can be represented by taking for each dyadic grid a sum of finitely many dyadic shifts of uniformly bounded complexity (see definition below) and taking the average over all grids.

In particular, as it was shown by A. Vagharshakyan [39], any convolution Calderón–Zygmund operator on the real line $\mathbb{R}$ with sufficiently smooth kernel can be obtained by averaging copies of just one Haar shift, so the A_2 conjecture holds for such operators.

Note that estimates of the norms of the dyadic shifts obtained in [16] and in [4, 5] grew exponentially in the complexity of the shift, so it was only possible to estimate the Calderón–Zygmund operators obtained by averaging of finitely many such shifts.

Using linear estimates for the dyadic shifts and a special decomposition (in the form proposed by Yang [43]) of a Calderón–Zygmund operator Hytönen–Lacey–Reguera–Sawyer–Vagharshakyan–Uriarte-Tuero in [15] proved the A_2 conjecture for all Calderón–Zygmund operators with sufficiently smooth kernels (the smoothness was dependent on the dimension in [15]). However, the problem for general Calderón–Zygmund operators required (as we shall see) some probabilistic ideas rooted in non-homogeneous Harmonic Analysis [20, 26] (see also [6, 10, 13, 21, 24, 25, 37, 40]).

For general Calderón–Zygmund operators, the last three authors [30] reduced the A_2 conjecture to a weak type estimate by establishing the inequality

$$\|T\|_{L^2(w)\to L^2(w)} \le C \left([w]_{A_2} + \|T\|_{L^2(w)\to L^{2,\infty}(w)} + \|T'\|_{L^2(w^{-1})\to L^{2,\infty}(w^{-1})}\right).$$

In [30] it is also shown that the A_2 conjecture is equivalent to getting the linear in $[w]_{A_2}$ estimate on simplest test functions (this is a $T(1)$ theorem in the presence of weight). Using this result of Pérez–Treil–Volberg and the technique developed in [16], the first author [14] was able to prove the A_2 conjecture for general Calderón–Zygmund operators, i.e., the following theorem:

Theorem 1.1 ([14]). *Let T be a Calderón–Zygmund operator and w be an A_2 weight. Then*

$$\|Tf\|_{L^2(w)} \le C \cdot [w]_{A_2} \|f\|_{L^2(w)},$$

where the constant C depends only on the dimension d, the parameters C_{cz}, α of the Calderón–Zygmund operator and its norm in the non-weighted L^2.

A crucial new element in [14] was a clever averaging trick, allowing one to get rid of the so called *bad* cubes and thus represent an arbitrary Calderón–Zygmund operator as a weighted average of (infinitely many) dyadic shifts. This averaging trick was a development of the bootstrapping argument used by Nazarov–Treil–Volberg [26], where they exploited the fact that the bad part of a function can be made arbitrarily small. Using the original Nazarov–Treil–Volberg averaging trick would add an extra factor depending on $[w]_{A_2}$ to the estimate, so a new idea was necessary. A new observation in [14] was that as soon as the probability of a "bad" cube is less than 1, it is possible to completely ignore the bad cubes (at least in the situation where they cause troubles).

The paper [14], which itself is neither short or very simple, relies of a rather technically involved preprint [30]. Thus the necessity of a simpler, direct proof, not using the reduction to the weak type estimates seems pretty evident.

Such a direct proof of Theorem 1.1 is presented in this paper; moreover, we obtain new results on the dyadic shifts into which the Calderón–Zygmund operator T is decomposed. Indeed, the reduction of the A_2 conjecture to a testing condition, which in [30] was made on the level of the Calderón–Zygmund operator T, is here performed on the more elementary level of the dyadic shifts in the representation of T. The possibility of such a simplification in the proof of the A_2 conjecture was suggested in [14, Section 8.A], and here we carry out this program in detail.

The main components of the proof are as follows:

(i) An averaging trick, which is a version of the one from [14] (unlike [14] we do not need *good* shifts here, and this simplifies the matter). This trick allows us not to worry about "bad" cubes and represent a general Calderón–Zygmund operator as a weighted average of dyadic shifts with the weights decaying exponentially in the complexity of the shifts.

(ii) Sharp estimates, with all the constants written down, in the two weight $T(1)$ theorem from [28] in the setting of dyadic shifts (Theorem 3.4). Note that while most of the necessary estimates were done in [28], a formal application of the result from [28] would give an exponential (in complexity) growth of the norm.

To get the polynomial (in complexity) growth, one needs some non-trivial modifications. For the convenience of the reader we present the complete proof, not only the modifications: only describing modifications and referring the reader to the proof in [28] would make the paper unreadable.

(iii) A modification of the proof from [16], which gives polynomial in complexity, instead of exponential, as in [16], bound for the weighted norm of the dyadic shift (Theorem 5.1). The main difference compared to [16] is a better (linear in complexity instead of exponential) estimate of the (non-weighted) weak L^1 norm of a dyadic shift, which was obtained in [14].

The rest of the proof essentially follows the construction from [16], keeping track of constants, and clarifying parts of the proof that were presented there in a sketchy way. We note that a variant of such a modification of [16] already appeared in [14], where it was used to verify the required testing conditions for T, but not an explicit norm bound for the shifts themselves.

Aside from the new self-contained proof of Theorem 1.1, the above-mentioned Theorems 3.4 and 5.1, giving sharp quantitative two-weight and one-weight bounds for dyadic shifts, are the main new results of this paper. In this paper, the two-weight bounds serve mainly as a tool for sharp one-weight bounds, but they have been systematically studied in [22, 27, 29, 35, 36].

2. Dyadic lattices and martingale difference decompositions. Random dyadic lattices

2.1. Random dyadic lattices.
The standard dyadic system in $\mathbb{R}^d$ is

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

For $I \in \mathcal{D}_k^0$ and a binary sequence $\omega = (\omega_j)_{j=-\infty}^\infty \in (\{0,1\}^d)^{\mathbb{Z}}$, let

$$I \dotplus \omega := I + \sum_{j<k} \omega_j 2^j.$$

Following Nazarov–Treil–Volberg [26, Section 9.1], consider general dyadic systems of the form

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

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

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

denote the collection of dyadic children of I. Thus $\mathscr{D}_{k-1}^\omega = \bigcup \{ \mathrm{ch}(I) : I \in \mathscr{D}_k^\omega \}$. Note that, in line with [26] but contrary to [14], we use the "geometric" indexing of cubes, where larger k refers to larger cubes, rather than the "probabilistic" indexing, where larger k would refer to finer sigma-algebras.

Consider the standard probability measure on $\{0, 1\}^d$, which assigns equal probability 2^{-d} to every point. Define the measure $\mathbf{P}$ on $(\{0, 1\}^d)^{\mathbb{Z}}$ as the corresponding product measure.

2.2. Martingale difference decompositions and Haar functions. For a cube I in $\mathbb{R}^d$ let

$$\mathbb{E}_I f := \left(\fint_I f \, dx \right) \mathbf{1}_I := \left(|I|^{-1} \int_I f \, dx \right) \mathbf{1}_I, \qquad \Delta_I := -\mathbb{E}_I + \sum_{J \in \mathrm{ch}(I)} \mathbb{E}_J.$$

It is well known that for an arbitrary dyadic lattice $\mathscr{D}$ every function $f \in L^2(\mathbb{R}^d)$ admits the orthogonal decomposition

$$f = \sum_{I \in \mathscr{D}} \Delta_I f.$$

We also need the weighted martingale difference decomposition. Let μ be a Radon measure on $\mathbb{R}^d$. Define the weighted expectation and martingale differences as

$$\mathbb{E}_I^\mu f := \left((\mu(I))^{-1} \int_I f \, d\mu \right) \mathbf{1}_I, \qquad \Delta_I^\mu := -\mathbb{E}_I^\mu + \sum_{J \in \mathrm{ch}(I)} \mathbb{E}_J^\mu;$$

for the definiteness we set $\mathbb{E}_I^\mu f = 0$ if $\mu(I) = 0$.

For an arbitrary dyadic lattice $\mathscr{D}$ and $k \in \mathbb{Z}$, any function $f \in L^2(\mu)$ admits an orthogonal decomposition

$$(2.1) \qquad f = \sum_{I \in \mathscr{D}: \ell(I) = 2^k} \mathbb{E}_I^\mu f + \sum_{I \in \mathscr{D}: \ell(I) \le 2^k} \Delta_I^\mu f.$$

Given a cube Q in $\mathbb{R}^d$, any function in the martingale difference space $\Delta_Q L^2$ is called a Haar function (corresponding to Q) and is usually denoted by h_Q. Note that h_Q denotes a *generic* Haar function, not any particular one.

A *generalized* Haar function h_Q is a linear combination of a Haar function and $\mathbf{1}_Q$. In other words, a generalized Haar function h_Q is constant on the children of Q, but unlike the regular Haar function it is not orthogonal to constants.

Similarly a function $h \in \Delta_Q^\mu L^2(\mu)$ is called a weighted Haar function and is denoted as h_Q^μ.

3. Dyadic shifts. A sharp two weight estimate

Definition 3.1. An unweighted dyadic paraproduct is an operator Π of the form

$$\Pi f = \sum_{Q \in \mathscr{D}} (\mathbb{E}_Q f) h_Q,$$

where h_Q are some (non-weighted) Haar functions.

Definition 3.2.   Let $m, n \in \mathbb{N}$. An elementary dyadic shift with parameters m, n is an operator given by

$$\mathbb{S}f := \sum_{Q \in \mathcal{D}} \sum_{\substack{Q',Q'' \in \mathcal{D}, Q', Q'' \subset Q, \\ \ell(Q')=2^{-m}\ell(Q), \ell(Q'')=2^{-n}\ell(Q)}} |Q|^{-1} (f, h_{Q'}^{Q''}) h_{Q''}^{Q'},$$

where $h_{Q'}^{Q''}$ and $h_{Q''}^{Q'}$ are (non-weighted) Haar functions for the cubes Q' and Q'' respectively, subject to normalization

(3.1) $$\left\| h_{Q'}^{Q''} \right\|_\infty \cdot \left\| h_{Q''}^{Q'} \right\|_\infty \leq 1.$$

Notice that this implies, in particular, that

(3.2)
$$\mathbb{S}f(x) = \sum_{Q \in \mathcal{D}} |Q|^{-1} \int_Q a_Q(x, y) f(y) dy, \qquad \operatorname{supp} a_Q \subset Q \times Q, \ \|a_Q\|_\infty \leq 1,$$

where

(3.3) $$a_Q(x, y) = \sum_{\substack{Q',Q'' \in \mathcal{D}, Q', Q'' \subset Q, \\ \ell(Q')=2^{-m}\ell(Q), \ell(Q'')=2^{-n}\ell(Q)}} h_{Q''}^{Q'}(x) h_{Q'}^{Q''}(y).$$

The number $\max(m, n)$ is called the *complexity* of the dyadic shift.

Definition 3.3.   If in the above definition we allow some (or all) $h_{Q'}$, $h_{Q''}$ to be *generalized* Haar functions, we get what we will call an *elementary generalized dyadic shift*.

A dyadic shift with parameters m and n is a sum of at most $(2^d)^2$ elementary dyadic shifts (with parameters m and n). If we allow some (or all) of the elementary dyadic shifts to be *generalized* ones, we get the *generalized* dyadic shift.

Remark.   The paraproduct Π is an elementary generalized dyadic shift with parameters 0, 1, provided that $\|h_Q\|_\infty \leq 1$ for all cubes Q.

Remark.   The main difference between dyadic shifts and generalized ones is that a dyadic shift is always a bounded operator in L^2 (assuming the normalization (3.1)), while for the boundedness of a generalized dyadic shift some additional conditions are required.

We always think that our dyadic shifts $\mathbb{S}$ are *finite* dyadic shifts meaning that only finitely many Q's are involved in its definition above. All estimates will be independent of this finite number.

In the present section we consider a two weight $T(1)$ theorem for dyadic shifts. We fix two measures μ, ν on $\mathbb{R}^d$. Finite dyadic shifts are integral operators with kernel

$$A(x, y) = \sum_{Q \in \mathcal{D}} a_Q(x, y),$$

the sum being well defined as it is finite. We define now

$$\mathbb{S}_\mu f(x) := \int A(x, y) f(y) \, d\mu(y),$$

and its adjoint $\mathbb{S}_v^*$,

$$\mathbb{S}_v^* g(y) = \int \overline{A}(x,y) g(x) \, dv(x).$$

We need the notation

$$[\mu, v]_{A_2} := \sup_I \langle \mu \rangle_I \langle v \rangle_I,$$

where $\langle \sigma \rangle_I := |I|^{-1} \sigma(I)$.

The following theorem is the first new main result of this paper. It is essentially a quantified version of [28, Theorem 2.3].

Theorem 3.4. *Let $\mathbb{S}$ be an elementary generalized dyadic shift with parameters m and n. Let us suppose that there exists a constant B such that for any $Q \in \mathcal{D}$ we have*

$$(3.4) \qquad \int_Q |\mathbb{S}_\mu \mathbf{1}_Q|^2 dv \le B\mu(Q), \qquad \int_Q |\mathbb{S}_v^* \mathbf{1}_Q|^2 d\mu \le Bv(Q).$$

Then

$$(3.5) \qquad \|\mathbb{S}_\mu f\|_v \le C \Big(2^{d/2}(r+1)\big(B^{1/2} + [\mu,v]_{A_2}^{1/2}\big) + r^2 [\mu,v]_{A_2}^{1/2} \Big) \|f\|_\mu,$$

where $r = \max(m,n)$, and C is an absolute constant.

The idea of the proof of this theorem is quite simple. The operator $\mathbb{S}^\mu$ is represented essentially as the sum of *weighted paraproducts*, which are estimated using condition (3.4) and the operator with finitely many diagonals, which is estimated by $C[\mu,v]_{A_2}^{1/2}$.

Take two test functions f, g. Using martingale difference decomposition (2.1) we can decompose

$$f = \sum_{Q \in \mathcal{D}: \ell(I)=2^k} \mathbb{E}_Q^\mu f + \sum_{Q \in \mathcal{D}: \ell(I)<2^k} \Delta_Q^\mu f,$$

$$g = \sum_{Q \in \mathcal{D}: \ell(I)=2^k} \mathbb{E}_Q^v g + \sum_{Q \in \mathcal{D}: \ell(I)<2^k} \Delta_Q^v g.$$

We want to estimate the bilinear form $\langle \mathbb{S}^\mu f, g \rangle_v$. We will first concentrate on the nontrivial case $f = \sum_{Q \in \mathcal{D}} \Delta_Q^\mu f$, $g = \sum_{Q \in \mathcal{D}} \Delta_Q^v g$; adding the terms $\sum_{Q \in \mathcal{D}: \ell(I)=2^k} \mathbb{E}_Q^\mu f$ and $\sum_{Q \in \mathcal{D}: \ell(I)=2^k} \mathbb{E}_Q^v g$ will be easy.[1]

3.1. Weighted paraproducts. Fix an integer r. Then the paraproduct $\Pi^\mu = \Pi_{\mathbb{S}}^\mu$, acting (formally) from $L^2(\mu)$ to $L^2(v)$ is defined as

$$\Pi^\mu := \sum_{Q \in \mathcal{D}} \mathbb{E}_Q^\mu f \sum_{\substack{R \in \mathcal{D}, R \subset Q, \\ \ell(R)=2^{-r}\ell(Q)}} \Delta_R^v \mathbb{S}_\mu \mathbf{1}_Q.$$

---

[1] In fact, we will only apply this theorem in the situation when martingale difference decompositions not involving $\mathbb{E}_Q^\mu$ and $\mathbb{E}_Q^v$ are possible.

The paraproduct $\Pi^\nu = \Pi^\nu_{\mathbb{S}*}$, acting (formally) from $L^2(\nu)$ to $L^2(\mu)$, is defined similarly,

$$\Pi^\nu := \sum_{Q\in\mathscr{D}} \mathbb{E}^\nu_Q f \sum_{\substack{R\in\mathscr{D},R\subset Q,\\ \ell(R)=2^{-r}\ell(Q)}} \Delta^\mu_R \mathbb{S}^*_\nu 1_Q.$$

Notice that if $r \geq n$, then for any $f \in L^1_{\mathrm{loc}}(\mu)$ such that $f|_Q \equiv 1$, and for any $R \in \mathscr{D}$ such that $R \subset Q$ and $\ell(R) \leq 2^{-r}\ell(Q)$, we have

(3.6) $\Delta^\nu_R \mathbb{S}_\mu f = \Delta^\nu_R \mathbb{S}_\mu 1_Q.$

Indeed, in the decomposition

$$\langle \mathbb{S}_\mu(1_Q - f), h^\nu_R \rangle_\nu = \sum_{I\in\mathscr{D}} \sum_{\substack{I',I''\in\mathscr{D},I',I''\subset I,\\ \ell(I')=2^{-m}\ell(I),\ell(I'')=2^{-n}\ell(I)}} \langle 1_Q - f, h_{I'} \rangle_\mu \langle h_{I''}, h^\nu_R \rangle_\nu$$

only the terms with $I' \not\subset Q$ and $I'' \subset R$ can give a non-zero contribution. But the inclusions $I'' \subset R \subset Q$ together with size conditions on I'' and R imply that

$$\ell(I) = 2^n \ell(I'') \leq 2^r \ell(I'') \leq 2^r \ell(R) \leq \ell(Q),$$

so $I \subset Q$ (because $I \cap Q \supset I'' \neq \varnothing$, so the inclusion of the dyadic cubes is determined by their sizes). But the inclusion $I \subset Q$ implies $I' \subset Q$, so the conditions $I' \not\subset Q$ and $I'' \subset R$ are incompatible.

The equality (3.6) means that for $r \geq n$ we can replace 1_Q by 1, bringing our definition of the paraproduct more in line with the classical one.

Lemma 3.5. *Let $Q, R \in \mathscr{D}$, and let $r \geq n$. Then for the paraproduct $\Pi^\mu = \Pi^\mu_{\mathbb{S}*}$ defined above the following assertions hold:*

(i) *If $\ell(R) \geq 2^{-r}\ell(Q)$ then $\langle \Pi^\mu h^\mu_Q, h^\nu_R \rangle_\nu = 0$ for all weighted Haar functions h^μ_Q and h^ν_R.*

(ii) *If $R \not\subset Q$, then $\langle \Pi^\mu h^\mu_Q, h^\nu_R \rangle_\nu = 0$ for all weighted Haar functions h^μ_Q and h^ν_R.*

(iii) *If $\ell(R) < 2^{-r}\ell(Q)$, then for all weighted Haar functions h^μ_Q and h^ν_R*

$$\langle \Pi^\mu h^\mu_Q, h^\nu_R \rangle_\nu = \langle \mathbb{S}_\mu h^\mu_Q, h^\nu_R \rangle_\nu;$$

in particular, if $R \not\subset Q$, then both sides of the equality are 0.

Proof. Let us use Q' and R' for the summation indices in the paraproduct, i.e., let us write

$$\Pi^\mu h^\mu_Q := \sum_{Q'\in\mathscr{D}} \mathbb{E}^\mu_{Q'} h^\mu_Q \sum_{\substack{R'\in\mathscr{D},R'\subset Q',\\ \ell(R')=2^{-r}\ell(Q')}} \Delta^\nu_{R'} \mathbb{S}_\mu 1_{Q'}.$$

Since h^ν_R is orthogonal to ranges of all projections $\Delta^\nu_{R'}$ except Δ^ν_R we can write

(3.7) $\langle \Pi^\mu h^\mu_Q, h^\nu_R \rangle_\nu = \langle (E^\mu_{Q'} h^\mu_Q) \Delta^\nu_R \mathbb{S}_\mu 1_{Q'}, h^\nu_R \rangle_\nu = a \langle \mathbb{S}_\mu 1_{Q'}, h^\nu_R \rangle_\nu,$

where Q' is the ancestor of R of order r (i.e. the cube $Q' \supset R$ such that $\ell(Q') = 2^r \ell(R)$) and a is the value of $E^\mu_{Q'} h^\mu_Q$ on Q', $E^\mu_{Q'} h^\mu_Q = a 1_{Q'}$.

It is easy to see that $E_{Q'}^\mu h_Q^\mu \not\equiv 0$ (equivalently $a \ne 0$) only if $Q' \subsetneqq Q$. Therefore, see (3.7),

$$\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu \ne 0$$

only if $Q' \subsetneqq Q$ and statements (i) and (ii) of the lemma follow immediately.

Indeed, if $\ell(R) \ge 2^{-r}\ell(Q)$ and $\ell(Q') = 2^r\ell(R)$, the inclusion $Q' \subsetneqq Q$ is impossible, so

$$\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu = 0,$$

and the statement (i) is proved.

If $R \not\subset Q$, then the inclusion $Q' \subsetneqq Q$ (which, as it was discussed above, is necessary for $\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu \ne 0$) implies that $R \not\subset Q'$. This means that Q' is not an ancestor of R, however (3.7) again shows that for Q' to be an ancestor of R is necessary for $\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu \ne 0$.

Let us prove statement (iii). Let $\ell(R) < 2^{-r}\ell(Q)$. If $R \not\subset Q$ then by the statement (ii) of the lemma $\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu = 0$. On the other hand if M is the ancestor of order r of R, then $Q \cap M = \varnothing$, thus by (3.6)

$$\langle \mathbb{S}_\mu h_Q^\mu, h_R^\nu \rangle_\nu = \langle \mathbb{S}_\mu 0 \cdot \mathbf{1}_M, h_R^\nu \rangle_\nu = 0.$$

So, we only need to consider the case $R \subset Q$.

Let Q_1 be the "child" of Q containing R (i.e. $R \subset Q_1 \subset Q$, $\ell(Q_1) = \ell(Q)/2$), and let b be the value of h_Q^μ on Q_1. Then, since $\ell(R) \le 2^{-r}\ell(Q_1)$, (3.6) implies that

$$\langle \mathbb{S}_\mu h_Q^\mu, h_R^\nu \rangle_\nu = b\langle \mathbb{S}_\mu \mathbf{1}_{Q_1}, h_R^\nu \rangle_\nu.$$

On the other hand we have shown before, see (3.7), that

$$\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu = \langle (E_{Q'}^\mu h_Q^\mu)\Delta_R^\nu \mathbb{S}_\mu \mathbf{1}_{Q'}, h_R^\nu \rangle_\nu,$$

where $Q' \in \mathscr{D}$ is the ancestor of order r of R, meaning that $R \subset Q'$, $\ell(Q') = 2^r\ell(R)$. Therefore $Q' \subset Q_1$ and so $E_{Q'}^\mu h_Q^\mu = b\mathbf{1}_{Q'}$. We also know, see (3.6), that because $Q' \subset Q_1$ we have equality $\Delta_R^\nu \mathbb{S}_\mu \mathbf{1}_{Q'} = \Delta_R^\nu \mathbb{S}_\mu \mathbf{1}_{Q_1}$. Thus we can continue:

$$\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu = b\langle \Delta_R^\nu \mathbb{S}_\mu \mathbf{1}_{Q'}, h_R^\nu \rangle_\nu = b\langle \Delta_R^\nu \mathbb{S}_\mu \mathbf{1}_{Q_1}, h_R^\nu \rangle_\nu = b\langle \mathbb{S}_\mu \mathbf{1}_{Q_1}, h_R^\nu \rangle_\nu.$$

Therefore $\langle \Pi^\mu h_Q^\mu, h_R^\nu \rangle_\nu = \langle \mathbb{S}_\mu h_Q^\mu, h_R^\nu \rangle_\nu$, and the lemma is proved. □

3.2. Boundedness of the weighted paraproduct. We will need the following well-known theorem.

Let $f_R := \frac{1}{\mu(R)}\int_R f \, d\mu$ be the average of the function f with respect to the measure μ.

Theorem 3.6 (Dyadic Carleson embedding theorem). *If the numbers $a_Q \ge 0$, $Q \in \mathscr{D}$, satisfy the Carleson measure condition*

$$(3.8) \qquad \sum_{Q \subset R} a_Q \le \mu(R),$$

then for any $f \in L^2(\mu)$

$$\sum_{R \in \mathscr{D}} a_R |f_R|^2 \le 4 \cdot \|f\|_{L^2(\mu)}^2.$$

This theorem is very well known, cf. [8]. Usual proofs are based on a stopping time argument and the dyadic maximal inequality; the constant 4 appears as 2^2, where 2 is the norm of the dyadic maximal operator on $L^2(\mu)$. For an alternative proof using the Bellman function method, see [20]. It was also proved in [23] that the constant 4 is optimal. We should mention that in [19,23] this theorem was proved for $\mathbb{R}^1$, but the same proof works for general martingale setup. A proof for $\mathbb{R}^2$ was presented in [20], and the same proof works for $\mathbb{R}^d$.

Let us now show that the paraproduct $\Pi = \Pi_{\mathbb{S}}^{\mu}$ is bounded. Ranges of the projections Δ_R^{ν} are mutually orthogonal, so to prove the boundedness of the paraproduct $\Pi_{\mathbb{S}}^{\mu}$ it is sufficient to show that the numbers

$$a_Q := \sum_{\substack{R \in \mathscr{D}, R \subset Q, \\ \ell(R) = 2^{-r}\ell(Q)}} \|\Delta_R^{\nu} \mathbb{S}_{\mu} \mathbf{1}_R\|_{L^2(\nu)}^2$$

satisfy the Carleson measure condition (3.8) from Theorem 3.6. Let us prove this.

Consider a cube $\tilde{Q}$. We want to show that

$$\sum_{Q \subset \tilde{Q}} \sum_{\substack{R \in \mathscr{D}, R \subset Q, \\ \ell(R) = 2^{-r}\ell(Q)}} \|\Delta_R^{\nu} \mathbb{S}_{\mu} \mathbf{1}_Q\|_{L^2(\nu)}^2 \leq B\mu(\tilde{Q}).$$

By (3.6) we can replace $\mathbf{1}_Q$ by $\mathbf{1}_{\tilde{Q}}$, so the desired estimates become

$$\sum_{\substack{R \in \mathscr{D}, R \subset \tilde{Q}, \\ \ell(R) \leq 2^{-r}\ell(\tilde{Q})}} \|\Delta_R^{\nu} \mathbb{S}_{\mu} \mathbf{1}_{\tilde{Q}}\|_{L^2(\nu)}^2 \leq \sum_{R \subset \tilde{Q}} \|\Delta_R^{\nu} \mathbb{S}_{\mu} \mathbf{1}_{\tilde{Q}}\|_{L^2(\nu)}^2 \leq \|\mathbf{1}_{\tilde{Q}} \mathbb{S}_{\mu} \mathbf{1}_{\tilde{Q}}\|_{L^2(\nu)}^2.$$

By the assumption of Theorem 3.4, see (3.4),

$$\|\mathbf{1}_{\tilde{Q}} \mathbb{S}_{\mu} \mathbf{1}_{\tilde{Q}}\|_{L^2(\nu)}^2 := \int_{\tilde{Q}} |\mathbb{S}_{\mu} \mathbf{1}_{\tilde{Q}}|^2 d\nu \leq B\mu(\tilde{Q})$$

and so the sequence a_Q, $Q \in \mathscr{D}$ satisfies the condition (3.8). Thus the norm of the paraproduct Π^{μ} is bounded by $CB^{1/2}$ (we can pick $C = 2$ here) and similarly for Π^{ν}. $\square$

3.3. Boundedness of $\mathbb{S}$: essential part. Let $f \in L^2(\mu)$, $g \in L^2(\nu)$, $\|f\|_{\mu}, \|g\|_{\nu} \leq 1$. We want to estimate $|\langle \mathbb{S}_{\mu} f, g \rangle_{\nu}|$.

Consider first f and g of form

$$f = \sum_{Q \in \mathscr{D}} \Delta_Q^{\mu} f, \qquad g = \sum_{R \in \mathscr{D}} \Delta_R^{\nu} g, \qquad \|f\|_{\mu} \leq 1, \|g\|_{\nu} \leq 1.$$

Then by Lemma 3.5

$$(3.9) \quad \langle \mathbb{S}_{\mu} f, g \rangle_{\nu} = \langle \Pi_{\mathbb{S}_{\mu}}^{\mu} f, g \rangle_{\nu} + \langle f, \Pi_{\mathbb{S}_{\nu}^*}^{\nu} g \rangle_{\nu} + \sum_{\substack{Q, R \in \mathscr{D}, \\ 2^{-r} \leq \ell(R)/\ell(Q) \leq 2^r}} \langle \mathbb{S}_{\mu} \Delta_Q^{\mu} f, \Delta_R^{\nu} g \rangle_{\nu}.$$

We know that the paraproducts $\Pi_{\mathbb{S}_{\mu}}^{\mu}$ and $\Pi_{\mathbb{S}_{\nu}^*}^{\nu}$ are bounded, so the first two terms can be estimated together by $4B^{1/2}$. Thus it remains to estimate the last sum.

It is enough to estimate the operator S

$$\langle Sf, g \rangle_v := \sum_{\substack{Q, R \in \mathcal{D}, \\ 2^{-r} \ell(Q) \le \ell(R) \le \ell(Q)}} \langle \mathbb{S}_\mu \Delta_Q^\mu f, \Delta_R^v g \rangle_v$$

because the sum over $2^{-r} \ell(R) \le \ell(Q) < \ell(R)$ is estimated similarly. The operator S can be split as $S = \sum_{k=0}^r S_k$, where

$$\langle S_k f, g \rangle_v := \sum_{\substack{Q, R \in \mathcal{D}, \\ \ell(R) = 2^{-k} \ell(Q)}} \langle \mathbb{S}_\mu \Delta_Q^\mu f, \Delta_R^v g \rangle_v.$$

Each S_k can be in turn decomposed as $S_k = \sum_{j \in \mathbb{Z}} S_{k,j}$, where

$$\langle S_{k,j} f, g \rangle_v := \sum_{\substack{Q, R \in \mathcal{D}, \\ \ell(Q) = 2^j, \ell(R) = 2^{j-k}}} \langle \mathbb{S}_\mu \Delta_Q^\mu f, \Delta_R^v g \rangle_v.$$

For a fixed k the ranges $\operatorname{Ran} S_{k,j}$, $j \in \mathbb{Z}$ are mutually orthogonal in $L^2(v)$, and the dual ranges $\operatorname{Ran} S_{k,j}^*$, $j \in \mathbb{Z}$ are mutually orthogonal in $L^2(\mu)$. Therefore $\|S_k\| \le \max_{j \in \mathbb{Z}} \|S_{k,j}\|$, so we only need to uniformly estimate individual operators $S_{k,j}$.

So, if

$$f_j = \sum_{Q \in \mathcal{D}: \ell(Q) = 2^j} \Delta_Q^\mu f, \qquad g_{j-k} = \sum_{R \in \mathcal{D}: \ell(R) = 2^{j-k}} \Delta_Q^v g,$$

it is sufficient to estimate $\langle S_{k,j} f_j, g_{j-k} \rangle_v = \langle \mathbb{S}_\mu f_j, g_{j-k} \rangle_v$.

We can decompose the operator $S_{k,j}$ into *interior* and *outer* parts

$$\langle S_{k,j} f, g \rangle_v = \sum_{\substack{Q, R \in \mathcal{D}: R \subset Q, \\ \ell(Q) = 2^j, \ell(R) = 2^{j-k}}} \langle \mathbb{S}_\mu \Delta_Q^\mu f, \Delta_R^v g \rangle_v + \sum_{\substack{Q, R \in \mathcal{D}: R \cap Q = \varnothing, \\ \ell(Q) = 2^j, \ell(R) = 2^{j-k}}} \langle \mathbb{S}_\mu \Delta_Q^\mu f, \Delta_R^v g \rangle_v$$

$$=: \langle S_{k,j}^{\mathrm{int}} f, g \rangle_v + \langle S_{k,j}^{\mathrm{out}} f, g \rangle_v.$$

Let us estimate $S_{k,j}^{\mathrm{out}}$. For cubes $Q, R \in \mathcal{D}$, $R \cap Q = \varnothing$, $\ell(Q) = 2^j$, $\ell(R) = 2^{j-k}$ and the corresponding weighted Haar functions h_Q^μ and h_R^v we can write

$$(3.10) \quad \langle S_{k,j}^{\mathrm{out}} h_Q^\mu, h_R^v \rangle_v = \langle \mathbb{S}_\mu h_Q^\mu, h_R^v \rangle_v$$

$$= \sum_{M \in \mathcal{D}} |M|^{-1} \int_{M \times M} a_M(x, y) h_Q^\mu(y) \overline{h_R^v}(x) d\mu(y) dv(x),$$

where the kernels a_M are from (3.2).

If $\ell(M) \le \ell(Q) = 2^j$, then the cube M cannot contain both Q and R (because $R \cap Q = \varnothing$), so the corresponding integral in (3.10) is 0. On the other hand, if $\ell(M) > 2^r \ell(Q)$, $r = \max(m, n)$ being the complexity of the dyadic shift $\mathbb{S}$, then for any x the function $a_M(x, \cdot)$ is constant on Q, so the corresponding integral in (3.10) is again 0.

So in (3.10) we only need to count M, $2^j < \ell(M) \leq 2^{j+r}$, and therefore we can write

$$\left|\langle S_{k,j}^{\mathrm{out}} h_Q^\mu, h_R^\nu \rangle_\nu \right| = \left| \sum_{s=j+1}^{j+r} \int_{\mathbb{R}^d \times \mathbb{R}^d} A_s(x,y) h_Q^\mu(y) \overline{h_R^\nu}(x) d\mu(y) d\nu(x) \right|$$

$$\leq \sum_{s=j+1}^{j+r} \int_{\mathbb{R}^d \times \mathbb{R}^d} |A_s(x,y)| \cdot |h_Q^\mu(y)| \cdot |h_R^\nu(x)| d\mu(y) d\nu(x),$$

where $A_s(x,y) := \sum_{M \in \mathcal{D}:\ell(M)=2^s} |M|^{-1} a_M(x,y)$.

Adding extra non-negative terms (with $R \subset Q$) we can estimate

$$|\langle S_{k,j}^{\mathrm{out}} f, g \rangle_\nu|$$

$$\leq \sum_{s=j+1}^{j+r} \sum_{\substack{Q,R \in \mathcal{D}:R \cap Q = \varnothing, \\ \ell(Q)=2^j, \ell(R)=2^{j-k}}} \int_{\mathbb{R}^d \times \mathbb{R}^d} |A_s(x,y)| \cdot |\Delta_Q^\mu f(y)| \cdot |\Delta_R^\nu g(x)| d\mu(y) d\nu(x)$$

$$\leq \sum_{s=j+1}^{j+r} \int_{\mathbb{R}^d \times \mathbb{R}^d} |A_s(x,y)| \cdot |f_j(y)| \cdot |g_{j-k}(x)| d\mu(y) d\nu(x)$$

But each integral operator with kernel $|A_s|$ is the direct sum of the operators with kernels $|M|^{-1}|a_M|$, $M \in \mathcal{D}$, $\ell(M) = 2^s$ (recall that a_M is supported on $M \times M$).

Since $\|a_M\|_\infty \leq 1$ we can estimate the Hilbert–Schmidt norm

$$\int_{M \times M} |M|^{-2} |a_M(x,y)|^2 d\mu(y) d\nu(x) \leq [\mu,\nu]_{A_2}.$$

So the norm of each operator with kernel $|M|^{-1}|a_M(x,y)|$ is at most $[\mu,\nu]_{A_2}^{1/2}$. Therefore the norm of each operator with kernel $|A_s(x,y)|$ is estimated by $[\mu,\nu]_{A_2}^{1/2}$, and summing in s we get

$$(3.11) \qquad \|S_{k,j}^{\mathrm{out}}\|_{L^2(\mu) \to L^2(\nu)} \leq r[\mu,\nu]_{A_2}^{1/2}.$$

To estimate the norm of $S_{k,j}^{\mathrm{int}}$ we need the following simple lemma.

Lemma 3.7. *In the assumptions of Theorem* 3.4

$$\|\mathbf{1}_Q \mathbb{S}_\mu h_Q^\mu\|_\nu^2 \leq 2^d (B + 4[\mu,\nu]_{A_2}) \|h_Q^\mu\|_\mu^2$$

for any μ-Haar function h_Q^μ.

Proof. Let Q_k, $k = 1, 2, \ldots, 2^d$ be the dyadic children of Q. A μ-Haar function h_Q^μ can be represented as

$$(3.12) \qquad h_Q^\mu = \sum_{k=1}^{2^d} \alpha_k \mathbf{1}_{Q_k}, \qquad \sum_{k=1}^{2^d} \alpha_k \mu(Q_k) = 0$$

and

$$(3.13) \qquad \|h_Q^\mu\|_\mu^2 = \sum_{k=1}^{2^d} |\alpha_k|^2 \mu(Q_k).$$

By assumption (3.4) of Theorem 3.4

$$(3.14) \qquad \|\mathbf{1}_{Q_k} \mathbb{S}_\mu \mathbf{1}_{Q_k}\|_\nu^2 \le B\mu(Q_k).$$

Let us estimate $\|\mathbf{1}_{Q \setminus Q_k} \mathbb{S}_\mu \mathbf{1}_{Q_k}\|_\nu$. We know that

$$\mathbb{S}_\mu \mathbf{1}_{Q_k}(x) = \sum_{M \in \mathscr{D}} |M|^{-1} \int_{Q_k} a_M(x,y) \mathbf{1}_{Q_k}(y) d\mu(y).$$

Since the functions a_M are supported on $M \times M$, only the terms with $M \supset Q$ can give a non-zero contribution for $x \notin Q_k$. Therefore, summing the geometric series we get that

$$|\mathbb{S}_\mu \mathbf{1}_{Q_k}(x)| \le 2\mu(Q_k)|Q|^{-1} \quad \forall x \notin Q_k.$$

Then

$$\|\mathbf{1}_{Q \setminus Q_k} \mathbb{S}_\mu \mathbf{1}_{Q_k}\|_\nu^2 \le 4\mu(Q_k)^2 |Q|^{-2} \nu(Q),$$

and combining this estimate with (3.14) we get

$$\begin{aligned}
\|\mathbf{1}_Q \mathbb{S}_\mu \mathbf{1}_{Q_k}\|_\nu^2 &\le B\mu(Q_k) + 4\mu(Q_k)^2 |Q|^{-2} \nu(Q) \\
&\le B\mu(Q_k) + 4\mu(Q_k)\mu(Q)|Q|^{-2} \nu(Q) \\
&\le (B + 4[\mu,\nu]_{A_2})\mu(Q_k).
\end{aligned}$$

Therefore, we can get, recalling (3.12) and (3.13),

$$\begin{aligned}
\|\mathbf{1}_Q \mathbb{S}_\mu h_Q^\mu\|_\nu &\le \sum_{k=1}^{2^d} |\alpha_k| \|\mathbf{1}_Q \mathbb{S}_\mu \mathbf{1}_{Q_k}\|_\nu \\
&\le (B + 4[\mu,\nu]_{A_2})^{1/2} \sum_{k=1}^{2^d} |\alpha_k| \mu(Q_k)^{1/2} \\
&\le (B + 4[\mu,\nu]_{A_2})^{1/2} 2^{d/2} \left(\sum_{k=1}^{2^d} |\alpha_k|^2 \mu(Q_k) \right)^{1/2} \\
&= 2^{d/2}(B + 4[\mu,\nu]_{A_2})^{1/2} \|h_Q^\mu\|_\mu. \qquad \square
\end{aligned}$$

Using the above Lemma 3.7, we can easily estimate $S_{k,j}^{\text{int}}$. Namely,

$$\begin{aligned}
\|S_{k,j}^{\text{int}} f_j\|_\nu^2 &= \sum_{Q \in \mathscr{D}:\ell(Q)=2^j} \left\| \sum_{R \subset Q:\ell(R)=2^{j-k}} \Delta_R^\nu \mathbb{S}_\mu \Delta_Q^\mu f \right\|_\nu^2 \\
&\le \sum_{Q \in \mathscr{D}:\ell(Q)=2^j} \|\mathbf{1}_Q \mathbb{S}_\mu \Delta_Q^\mu f\|_\nu^2 \\
&\le 2^d (B + 4[\mu,\nu]_{A_2}) \sum_{Q \in \mathscr{D}:\ell(Q)=2^j} \|\Delta_Q^\mu f\|_\mu^2 \\
&= 2^d (B + 4[\mu,\nu]_{A_2}) \|f_j\|_\mu^2.
\end{aligned}$$

Combining this with the estimate (3.11) of $\|S_{k,j}^{\text{out}}\|$, we get that

$$\|S_{k,j}\|_{L^2(\mu)\to L^2(\nu)} \leq 2^{d/2}(B + 4[\mu,\nu]_{A_2})^{1/2} + r[\mu,\nu]_{A_2}^{1/2}.$$

Since the operator S_k is the orthogonal sum of $S_{k,j}$, we get the same estimate for $\|S_k\|$. To get the estimate for $\|S\|$, $S = \sum_{k=0}^{r} S_k$, we just multiply the above estimate by $r+1$.

Adding in (3.9) all the estimates together we get that for f and g of form

$$f = \sum_{Q\in\mathscr{D}} \Delta_Q^\mu f, \qquad g = \sum_{R\in\mathscr{D}} \Delta_R^\nu g, \qquad \|f\|_\mu \leq 1, \ \|g\|_\nu \leq 1,$$

we have

$$(3.15) \quad |\langle \mathbb{S}_\mu f, g\rangle_\nu| \leq 4B^{1/2} + 2\cdot(r+1)\big[2^{d/2}(B + 4[\mu,\nu]_{A_2})^{1/2} + r[\mu,\nu]_{A_2}^{1/2}\big];$$

the first term here comes from the paraproducts, and the extra factor 2 in the second term is to take into account the sum over $\ell(Q) < \ell(R)$ in (3.9).

3.4. Boundedness of $\mathbb{S}$: some little details.
We are almost done with the proof of Theorem 3.4, modulo a little detail: for arbitrary measures μ functions $f \in L^2(\mu)$ do not admit martingale difference decomposition $f = \sum_{Q\in\mathscr{D}} \Delta_Q^\mu f$.

Each compact subset of $\mathbb{R}^d$ is contained in at most 2^d cubes of the same size as the size of this compact subset, so let Q_k, $k = 1, 2, \ldots, 2^d$ be the dyadic cubes of some size 2^N containing supports of f and g. The correct decomposition is given by (2.1) which reads as

$$(3.16) \qquad f = \sum_{Q\in\mathscr{D}:\ell(Q)=2^k} \mathbb{E}_Q^\mu f + \sum_{Q\in\mathscr{D}:\ell(Q)\leq 2^k} \Delta_Q^\mu f$$

(here k is an arbitrary but fixed integer), and similarly for $g \in L^2(\nu)$,

$$(3.17) \qquad g = \sum_{Q\in\mathscr{D}:\ell(Q)=2^k} \mathbb{E}_Q^\nu g + \sum_{Q\in\mathscr{D}:\ell(Q)\leq 2^k} \Delta_Q^\nu g.$$

So we need to estimate some extra terms. Of course, in the situation when we apply the theorem ($d\mu = w\,dx$, $d\nu = w^{-1}dx$, w satisfies the A_2 condition) f and g can be represented via martingale difference decomposition, although some explanation will still be needed.

Fortunately, there is a very simple way to estimate the extra terms. Let us say that dyadic cubes $Q, R \in \mathscr{D}$ are *relatives* if they have a common ancestor, i.e. a cube $M \in \mathscr{D}$ such that $Q, R \subset M$. The importance of the notion of relatives stems from the trivial observation that if the cubes Q and R are not relatives, then $\mathbb{S}_\mu \mathbf{1}_Q \equiv 0$ on R.

It is sufficient to prove the estimate on a dense set of compactly supported functions. For compactly supported functions f and g only finitely many terms $\mathbb{E}_Q^\mu f$ and $\mathbb{E}_Q^\nu g$ in the decompositions (3.16) and (3.17) are non-zero. Let us slit the collection of corresponding cubes into equivalence classes of relatives, and for each equivalence class find a common ancestor (it is always possible because of finiteness).

Denote by $\mathscr{A}$ the set of these common ancestors. Then we can write instead of (3.16) and (3.17)

$$(3.18) \qquad f = \sum_{Q\in\mathscr{A}} \mathbb{E}_Q^\mu f + \sum_{Q\in\mathscr{A}} \sum_{R\in\mathscr{D}:R\subset Q} \Delta_R^\mu f =: f_e + f_d,$$

$$(3.19) \qquad g = \sum_{Q\in\mathscr{A}} \mathbb{E}_Q^\nu g + \sum_{Q\in\mathscr{A}} \sum_{R\in\mathscr{D}:R\subset Q} \Delta_R^\nu g =: g_e + g_d,$$

the indices "e" and "d" here mean *expectation* and *difference*. Let us decompose

$$\langle \mathbb{S}_\mu f, g \rangle_\nu = \langle \mathbb{S}_\mu (f_e + f_d), g_e + g_d \rangle_\nu$$
$$= \langle \mathbb{S}_\mu f_e, g \rangle_\nu + \langle \mathbb{S}_\mu f_d, g_e \rangle_\nu + \langle \mathbb{S}_\mu f_d, g_d \rangle_\nu.$$

The last term is estimated by (3.15) (note that $\|f\|_\mu^2 = \|f_e\|_\mu^2 + \|f_d\|_\mu^2$ and similarly for $\|g\|_\nu^2$), so we just need to estimate the first two terms.

Any two cubes $Q, Q' \in \mathcal{A}$, $Q \neq Q'$ are not relatives, so as we already mentioned $\mathbb{S}_\mu \mathbf{1}_Q \equiv 0$ on any $Q' \in \mathcal{A}$, $Q' \neq Q$. Therefore

$$|\langle \mathbb{S}_\mu \mathbb{E}_Q^\mu f, g \rangle_\nu| = |\langle \mathbb{S}_\mu \mathbb{E}_Q^\mu f, g \mathbf{1}_Q \rangle_\nu| \leq \|\mathbf{1}_Q \mathbb{S}_\mu \mathbb{E}_Q^\mu f\|_\nu \|g \mathbf{1}_Q\|_\nu$$
$$\leq B^{1/2} \|\mathbb{E}_Q^\mu f\|_\mu \|g \mathbf{1}_Q\|_\nu$$

(we use assumption (3.4) of Theorem 3.4 for the last inequality). Summing over all $Q \in \mathcal{A}$ and applying the Cauchy–Schwarz inequality we get

$$|\langle \mathbb{S}_\mu f_e, g \rangle_\nu| = \sum_{Q \in \mathcal{A}} |\langle \mathbb{S}_\mu \mathbb{E}_Q^\mu f, g \rangle_\nu| \leq B^{1/2} \sum_{Q \in \mathcal{A}} \|\mathbb{E}_Q^\mu f\|_\mu \|g \mathbf{1}_Q\|_\nu$$
$$\leq B^{1/2} \|f_e\|_\mu \|g\|_\nu \leq B^{1/2} \|f\|_\mu \|g\|_\nu.$$

Similarly

$$|\langle \mathbb{S}_\mu f_d, g_e \rangle_\nu| = |\langle f_d, \mathbb{S}_\nu^* g_e \rangle_\nu| \leq B^{1/2} \|f_d\|_\mu \|g_e\|_\nu \leq B^{1/2} \|f\|_\mu \|g\|_\nu,$$

so in general case we just need to add $2B^{1/2}$ to the right side of (3.15).

4. Dyadic shifts and random lattices

In this section we use a probabilistic approach to decompose an arbitrary Calderón–Zygmund operator as an average of simple blocks, namely, the dyadic shifts investigated above. More precisely, we prove the following result, which is a variant of [14, Theorem 4.2]. The decomposition here is easier than in [14], and there is a reason for that: the shifts in [14] needed to have an extra geometric property pertinent to being applied in conjunction with [30]. Here we do not need that as we are not basing our reasoning on a weighted $T1$ theorem of [30]. The idea of such decomposition goes back to methods of non-homogeneous Harmonic Analysis exploited in [26] or [41] for example.

Theorem 4.1. *Let T be a Calderón–Zygmund operator in $\mathbb{R}^d$ with parameter α. Then T can be represented as*

$$T = C \int_\Omega \sum_{m,n \in \mathbb{Z}_+} 2^{-(m+n)\alpha/2} \, \mathbb{S}_{m,n}^\omega \, d\mathbf{P}(\omega),$$

where $\mathbb{S}_{m,n}^\omega$ is a dyadic shift with parameters m, n in the lattice $\mathcal{D}_\omega$; the shifts with parameters $0, 1$ and $1, 0$ can be generalized shifts, and all other shifts are the regular ones.

The constant C depends only on the dimension d and the parameters of the Calderón–Zygmund operator T (the norm $\|T\|_{L^2 \to L^2}$, the smoothness α, and the constant C_{cz} in the Calderón–Zygmund estimates).

4.1. Getting rid of bad cubes.   Let $\mathcal{D}_\omega$, $\omega \in \Omega$ be the translated dyadic lattice in $\mathbb{R}^d$ as defined in Section 2.1 and let $\mathbf{P}$ be the canonical probability measure on Ω (also defined in Section 2.1).

Fix $r_0 \in \mathbb{N}$. Let $\gamma = \frac{\alpha}{2(d+\alpha)}$, where α is the Calderón–Zygmund parameter of the operator T.

Definition.   A cube $Q \in \mathcal{D}_\omega$ is called *bad* if there exists a bigger cube $R \in \mathcal{D}_\omega$ such that $\ell(Q) < 2^{-r_0}\ell(R)$ and
$$\operatorname{dist}(Q, R) < \ell(Q)^\gamma \ell(R)^{1-\gamma}.$$

Let us introduce some probabilistic notation we will use in this section. Let $\mathbf{E} = \mathbf{E}_\Omega$ denote the expectation with respect to the probability measure $\mathbf{P}$,
$$\mathbf{E}_\Omega F = \mathbf{E}_\Omega F(\omega) = \int_\Omega F(\omega)\, d\mathbf{P}(\omega);$$

slightly abusing the notation we will often write $\mathbf{E}_\Omega F(\omega)$ to emphasize that F is a random variable (depending on ω).

For $k \in \mathbb{Z}$ let $\mathfrak{A}_k$ be the sigma-algebra generated by the random variables ω_j, $j < k$, and let $\mathbf{E}_{\mathfrak{A}_k}$ be the corresponding conditional expectation. Because of the product structure of Ω, the conditional expectation $\mathbf{E}_{\mathfrak{A}_k}$ is easier to understand: it is just the integration with respect to a part of variables ω_j.

Namely, for $k \in \mathbb{Z}$ one can split $\omega = ({}^k\omega, \omega^k)$, where ${}^k\omega := (\omega_j)_{j<k}$, $\omega^k := (\omega_j)_{j\geq k}$, so Ω is represented as a product $\Omega = {}^k\Omega \times \Omega^k$. Note that the sets ${}^k\Omega$ and Ω^k are probability spaces with respect to the standard product measures. We will use the same letter $\mathbf{P}$ for these measures (probabilities), hoping that this will not lead to confusion.

Denote by $\Omega^k[{}^k\omega]$ the "slice" of Ω,
$$\Omega^k[{}^k\omega] = \{({}^k\omega, \omega^k) : \omega^k \in \Omega^k\}.$$

Then for almost all ${}^k\omega$, assuming that $\omega = ({}^k\omega, \omega^k)$ we have
$$(\mathbf{E}_{\mathfrak{A}_k} F)(\omega) = \mathbf{E}_{\Omega^k[{}^k\omega]} F := \int_{\Omega^k} F({}^k\omega, \tilde{\omega}^k)\, d\mathbf{P}(\tilde{\omega}^k),$$

so the conditional expectation $\mathbf{E}_{\mathfrak{A}_k}$ is just the integration over slices.

Finally, given a cube $Q \in \mathcal{D}_\omega$, $\ell(Q) = 2^k$, denote by $\Omega[Q]$ the slice $\Omega[Q] := \Omega^k[{}^k\omega]$ for the particular choice of the parameters ${}^k\omega = (\omega_j)_{j<k}$ determining the position of Q (and of all cubes of size 2^k). The notation $\mathbf{E}_{\Omega[Q]}$ then should be clear, and one also can define the conditional probability
$$\mathbf{P}\{\text{event}|Q\} := \mathbf{E}_{\Omega[Q]}\mathbf{1}_{\text{event}}.$$

Lemma 4.2. $\pi_{\text{bad}} = \pi_{\text{bad}}(r_0, \gamma, d) := \mathbf{P}\{Q \text{ is bad}|Q\} \leq C(d)\, 2^{-cr_0}$.

In words: given a cube Q, the probability that it is bad is a constant depending only on r_0, γ and d, and can be estimated as stated.

Proof.   The proof is an easy exercise for the reader.   □

From now on let us fix a sufficiently large r_0 such that $\pi_{\text{bad}} < 1$, so the probability of being good satisfies $\pi_{\text{good}} = 1 - \pi_{\text{bad}} > 0$.

Lemma 4.3. *Let T be a bounded operator in $L^2 = L^2(\mathbb{R}^d, dx)$. Then for all $f, g \in C_0^\infty$ one has*

$$\langle Tf, g \rangle = \pi_{\text{good}}^{-1} \int_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J), \\ I \text{ is good}}} \langle T\Delta_I f, \Delta_J g \rangle \, d\mathbf{P}(\omega)$$

$$+ \pi_{\text{good}}^{-1} \int_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) > \ell(J), \\ J \text{ is good}}} \langle T\Delta_I f, \Delta_J g \rangle \, d\mathbf{P}(\omega).$$

Proof. It is more convenient to use probabilistic notation in the proof. Let

$$f_{\text{good}, \omega} := \sum_{\substack{I \in \mathscr{D}_\omega, \\ I \text{ is good}}} \Delta_I f.$$

Then for any $f, g \in L^2$,

$$\mathbf{E}_\Omega \langle f_{\text{good}, \omega}, g \rangle = \mathbf{E}_\Omega \sum_{\substack{I \in \mathscr{D}_\omega, \\ I \text{ is good}}} \langle \Delta_I f, \Delta_I g \rangle$$

$$= \sum_{k \in \mathbb{Z}} \mathbf{E}_\Omega \mathbf{E}_{\mathfrak{A}_k} \sum_{\substack{I \in \mathscr{D}_\omega : \ell(I) = 2^k, \\ I \text{ is good}}} \langle \Delta_I f, \Delta_I g \rangle$$

$$= \sum_{k \in \mathbb{Z}} \mathbf{E}_\Omega \mathbf{E}_{\mathfrak{A}_k} \sum_{I \in \mathscr{D}_\omega : \ell(I) = 2^k} \langle \Delta_I f, \Delta_I g \rangle \mathbf{1}_{\{I \text{ is good}\}}(\omega).$$

To compute the conditional expectation let us notice that the position of the cubes $I \in \mathscr{D}_\omega$, $\ell(I) = 2^k$ depends only on the random variables ω_j, $j < k$. On the other hand, the event that a cube $I \in \mathscr{D}_\omega$, $\ell(I) = 2^k$ is good depends only on the variables ω_j, $j \geq k$, and for fixed variables ω_j, $j < k$ the corresponding conditional probability of this event is π_{good}, so we can write for the conditional expectation

$$(4.1) \qquad\qquad \mathbf{E}_{\mathfrak{A}_k} \mathbf{1}_{\{I \text{ is good}\}}(\omega) = \pi_{\text{good}}.$$

Therefore

$$\mathbf{E}_{\mathfrak{A}_k} \sum_{I \in \mathscr{D}_\omega : \ell(I) = 2^{-k}} \langle \Delta_I f, \Delta_I g \rangle \mathbf{1}_{\{I \text{ is good}\}}(\omega) = \pi_{\text{good}} \sum_{I \in \mathscr{D}_\omega : \ell(I) = 2^{-k}} \langle \Delta_I f, \Delta_I g \rangle,$$

which gives us

$$(4.2) \qquad\qquad \mathbf{E}_\Omega \langle f_{\text{good}, \omega}, g \rangle = \pi_{\text{good}} \langle f, g \rangle.$$

Applying this identity to $\langle Tf_{\text{good},\omega}, g\rangle = \langle f_{\text{good},\omega}, T^*g\rangle$ (with T^*g instead of g) we get

$$
\begin{aligned}
(4.3) \quad \pi_{\text{good}} & \langle Tf, g\rangle \\
&= \mathbf{E}_\Omega \langle Tf_{\text{good},\omega}, g\rangle \\
&= \mathbf{E}_\Omega \sum_{\substack{I,J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J), \\ I \text{ is good}}} \langle T\Delta_I f, \Delta_J g\rangle \\
&\qquad\qquad + \sum_{k \in \mathbb{Z}} \mathbf{E}_\Omega \mathbf{E}_{\mathfrak{A}_k} \sum_{\substack{I,J \in \mathscr{D}_\omega, \\ \ell(I)=2^k, \ell(I)>\ell(J)}} \langle T\Delta_I f, \Delta_J g\rangle \mathbf{1}_{\{I \text{ is good}\}} \\
&= \mathbf{E}_\Omega \sum_{\substack{I,J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J), \\ I \text{ is good}}} \langle T\Delta_I f, \Delta_J g\rangle + \pi_{\text{good}} \mathbf{E}_\Omega \sum_{\substack{I,J \in \mathscr{D}_\omega, \\ \ell(I)>\ell(J)}} \langle T\Delta_I f, \Delta_J g\rangle;
\end{aligned}
$$

here again in the last equality we used (4.1) and the fact that for $2^k = \ell(I) \geq \ell(J)$ the position of I and J depends on the variables ω_j, $j < k$, while the property of I depends on the variables ω_j, $j \geq k$ and is not influenced by the position of J.

Remark 4.4. To justify the interchange of the summation and expectation $\mathbf{E}_\Omega$ in (4.3) we first observe that for smooth f

$$
\|\Delta_I f\|_\infty \leq \begin{cases} C(d)\|\nabla f\|_\infty \ell(I) & \ell(I) < 1, \\ \|f\|_\infty |I|^{-1} & \ell(I) \geq 1. \end{cases}
$$

So, if we denote

$$
f_\omega^k := \sum_{I \in \mathscr{D}_\omega : \ell(I)=2^k} \Delta_I f, \qquad f_{\text{good},\omega}^k := \sum_{\substack{I \in \mathscr{D}_\omega : \ell(I)=2^k, \\ I \text{ is good}}} \Delta_I f,
$$

then, integrating the previous estimates, we have for $f \in C_0^\infty$

$$
\|f_\omega^k\|_{L^2}, \ \|f_{\text{good},\omega}^k\|_{L^2} \leq C(f) \min\{2^k, 2^{-kd}\},
$$

so

$$
\sum_{k \in \mathbb{Z}} \|f_\omega^k\|_{L^2} \leq C(f), \qquad \sum_{k \in \mathbb{Z}} \|f_{\text{good},\omega}^k\|_{L^2} \leq C(f).
$$

Then for $f, g \in C_0^\infty$

$$
\sum_{j,k \in \mathbb{Z}} |\langle Tf_{\text{good},\omega}^k, g_\omega^j\rangle| \leq \|T\| C(f) C(g),
$$

which justifies the first interchange of summation and integration in (4.3). The same estimate holds if we replace $f_{\text{good},\omega}^k$ by f_ω^k, and this justifies the second interchange.

Note also that the sum f_ω^k has at most $C(f,k)$ non-zero terms $\Delta_I f$ (where $C(f,k) < \infty$ does not depend on ω), so for fixed k and j we can interchange summation over I, $\ell(I) = 2^k$ and integration without any problems.

Let us continue with the proof of Lemma 4.3. Since for all $\omega \in \Omega$

$$\langle Tf, g \rangle = \sum_{I, J \in \mathscr{D}_\omega} \langle T\Delta_I f, \Delta_J g \rangle,$$

averaging over all ω we get

$$(4.4) \qquad \langle Tf, g \rangle = \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J)}} \langle T\Delta_I f, \Delta_J g \rangle + \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) > \ell(J)}} \langle T\Delta_I f, \Delta_J g \rangle.$$

Multiplying this identity by π_{good} and comparing with (4.3) we get that

$$(4.5) \qquad \pi_{\text{good}} \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J)}} \langle T\Delta_I f, \Delta_J g \rangle = \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J), \\ I \text{ is good}}} \langle T\Delta_I f, \Delta_J g \rangle.$$

Remark. Note that the above identity cannot be obtained by directly applying the above trick with the conditional expectation to the right side. If $2^s = \ell(I) < \ell(J) = 2^k$, then the position of I and J is defined by the variables ω_j, $j < k$, and the property of I being good depends on ω_j, $j \geq s$. Thus the conditional probability of I being good depends on the mutual position of I and J and so there is no splitting we used proving (4.2), (4.3).

We can repeat the reasoning leading to (4.4) without any changes to the splitting into $\ell(I) < \ell(J)$ and $\ell(I) \geq \ell(J)$ to get

$$\pi_{\text{good}} \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) < \ell(J)}} \langle T\Delta_I f, \Delta_J g \rangle = \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) < \ell(J), \\ I \text{ is good}}} \langle T\Delta_I f, \Delta_J g \rangle.$$

From the symmetry between I and J we can conclude that

$$(4.6) \qquad \pi_{\text{good}} \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) > \ell(J)}} \langle T\Delta_I f, \Delta_J g \rangle = \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) > \ell(J), \\ J \text{ is good}}} \langle T\Delta_I f, \Delta_J g \rangle.$$

Substituting (4.3) and (4.4) into (4.6) we get

$$\langle Tf, g \rangle = \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J)}} \langle T\Delta_I f, \Delta_J g \rangle + \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) > \ell(J)}} \langle T\Delta_I f, \Delta_J g \rangle$$

$$= \pi_{\text{good}}^{-1} \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \\ \ell(I) \leq \ell(J), \\ I \text{ is good}}} \langle T\Delta_I f, \Delta_J g \rangle + \pi_{\text{good}}^{-1} \mathbf{E}_\Omega \sum_{\substack{I, J \in \mathscr{D}_\omega, \ell(I) > \ell(J) \, J \text{ is good}}} \langle T\Delta_I f, \Delta_J g \rangle. \quad \square$$

4.2. Subtracting paraproducts. For a Calderón–Zygmund operator T in $L^2(\mathbb{R}^d)$ and a dyadic lattice $\mathscr{D}_\omega$, define the dyadic paraproduct Π_T^ω,

$$\Pi_T^\omega f := \sum_{Q \in \mathscr{D}_\omega} (\mathbb{E}_Q f) \Delta_Q T\mathbf{1}.$$

Here $\Delta_Q T\mathbf{1}$ is defined by duality,

$$\langle \Delta_Q T\mathbf{1}, g\rangle := \langle \mathbf{1}, T^* \Delta_Q g\rangle \quad \forall g \in L^2;$$

the right side here is well defined, as one can easily show that $T^* \Delta_Q g \in L^1$. (This is a pretty standard place in the theory of Calderón–Zygmund operators.)

Define operators $\tilde{T}_\omega$

$$\tilde{T}_\omega := T - \Pi_T^\omega - (\Pi_{T^*}^\omega)^*.$$

Remark 4.5. The matrix of the paraproduct Π_T^ω has a very special "triangular" form. Namely, a block $\Delta_R \Pi_T^\omega \Delta_Q$, $Q, R \in \mathcal{D}_\omega$ can be non-zero only if $R \subsetneqq Q$. Notice also, that if $\ell(Q) = 2^k$, then the block $\Delta_R \Pi_T^\omega \Delta_Q$ does not depend on the variables ω_j, $j \geq k$.

From the above observation it is easy to see that if $Q, R \in \mathcal{D}_\omega$, $\max\{\ell(Q), \ell(R)\} = 2^k$, then the block $\Delta_R \tilde{T}_\omega \Delta_Q$ does not depend on variables ω_j, $j \geq k$, and that

$$\Delta_R \tilde{T}_\omega \Delta_Q = \Delta_R T \Delta_Q$$

if $Q \cap R = \varnothing$ or $Q = R$.

The paraproducts were introduced in Calderón–Zygmund theory in the proofs of $T(1)$ and $T(b)$ theorems. The main idea is that one can estimate the operators $\tilde{T}_\omega$ by estimating the absolute values of the entries of its matrix in the Haar basis, but one cannot, in general, do the same with paraproducts (and so with a general Calderón–Zygmund operator T). The paraproducts however can be easily estimated by the Carleson embedding theorem, using the condition $T\mathbf{1} \in \text{BMO}$ ($Tb \in \text{BMO}$).

Definition. Let $D(Q, R)$ be the so-called *long distance* between the cubes Q and R, see [26],

$$D(Q, R) := \text{dist}(Q, R) + \ell(Q) + \ell(R).$$

Lemma 4.6. *Let T be a Calderón–Zygmund operator (with parameter α), and let $Q, R \in \mathcal{D}_\omega$, $\ell(Q) \leq \ell(R)$. Let h_Q and h_R be Haar functions, $\|h_Q\| = \|h_R\| = 1$. If Q is a good cube, then*

$$|\langle \tilde{T}_\omega h_Q, h_R\rangle|, \; |\langle \tilde{T}_\omega h_R, h_Q\rangle| \leq C \frac{\ell(Q)^{\alpha/2}\ell(R)^{\alpha/2}}{D(Q, R)^{d+\alpha}}|Q|^{1/2}|R|^{1/2},$$

where $C = C(r_0, d, \alpha, C_{\text{cz}}) < \infty$.

The proof is pretty standard, see for example [26].

Lemma 4.7. *Let $C = C(r_0, d, \alpha, C_{\text{cz}}) \neq 0$ be the constant from the above Lemma* 4.6, *and let $|a_{Q,R}| \leq 1$. Then for any dyadic lattice $\mathcal{D}_\omega$ and for any $m, n \in \mathbb{Z}_+$, $m \geq n$ the operator*

$$C^{-1} \sum_{M \in \mathcal{D}_\omega} \sum_{\substack{Q, R \in \mathcal{D}_\omega : Q, R \subset M, \\ \ell(Q) = 2^{-m}\ell(M), \\ \ell(R) = 2^{-n}\ell(M), \\ Q \text{ is good}}} a_{Q,R} 2^{(m+n)\alpha/2} \cdot \frac{D(Q, R)^{d+\alpha}}{\ell(M)^{d+\alpha}} \Delta_R \tilde{T}_\omega \Delta_Q$$

is a dyadic shift with parameters m, n, and the same holds if we replace $\Delta_R \tilde{T}_\omega \Delta_Q$ by $\Delta_Q \tilde{T}_\omega \Delta_R$.

Proof. We will need the notion of the *standard Haar basis* here. For an interval $I \subset \mathbb{R}$ let $h_I^0 := |I|^{-1/2} \mathbf{1}_I$, and let h_I^1 be the standard L^2-normalized Haar function,

$$h_I^1 := |I|^{-1/2}(\mathbf{1}_{I_+} - \mathbf{1}_{I_-}),$$

where I_+ and I_- are the right and the left halves of I respectively.

For a cube $Q = I_1 \times I_2 \times \cdots \times I_d \in \mathbb{R}^d$ and an index j, $0 \leq j < 2^d$, let

$$h_Q^j(x) := \prod_{k=1}^d h_{I_k}^{j_k}(x_k), \quad x = (x_1, x_2, \ldots, x_d),$$

where $j_k \in \{0, 1\}$ are the coefficients in the binary decomposition $j = \sum_{k=1}^d j_k 2^{k-1}$ of j.

The system h_Q^j, $j = 1, \ldots, 2^d - 1$ form an orthonormal basis in $\Delta_Q L^2$, which we will call the *standard Haar basis*.

Note that $h_Q^0 = |Q|^{-1/2} \mathbf{1}_Q$. The block $\Delta_R \tilde{T}_\omega \Delta_Q$ can be represented as

$$\Delta_R \tilde{T}_\omega \Delta_Q = \sum_{j,k=1}^{2^d-1} c_{j,k}(Q, R) \langle \cdot, h_Q^k \rangle h_R^j,$$

where $c_{j,k}(Q, R) = \langle \tilde{T}_\omega h_Q^k, h_R^j \rangle$.

Since $\|h_Q^j\|_\infty = |Q|^{-1/2}$ we can estimate, using Lemma 4.6,

$$(4.7) \qquad |c_{j,k}(Q, R)| \cdot \|h_Q^k\|_\infty \cdot \|h_R^j\|_\infty \leq C \frac{\ell(Q)^{\alpha/2} \ell(R)^{\alpha/2}}{D(Q, R)^{d+\alpha}},$$

where $C = C(r_0, d, \alpha, C_{cz})$ is the constant from Lemma 4.6.

Clearly for fixed j, k and the constant C from Lemma 4.6 we can write

$$C^{-1} \sum_{M \in \mathcal{D}_\omega} \sum_{\substack{Q,R \in \mathcal{D}_\omega : Q,R \subset M, \\ \ell(Q) = 2^{-m}\ell(M), \\ \ell(R) = 2^{-n}\ell(M), \\ Q \text{ is good}}} a_{Q,R} 2^{(m+n)\alpha/2} \cdot \frac{D(Q, R)^{d+\alpha}}{\ell(M)^{d+\alpha}} c_{j,k}(Q, R) \langle \cdot, h_Q^k \rangle h_R^j$$

$$= \sum_{M \in \mathcal{D}_\omega} \sum_{\substack{Q,R \in \mathcal{D}_\omega : Q,R \subset M, \\ \ell(Q) = 2^{-m}\ell(M), \\ \ell(R) = 2^{-n}\ell(M), \\ Q \text{ is good}}} \langle \cdot, h_Q \rangle h_R,$$

where h_Q and h_R are multiples of h_Q^k and h_R^j. This sum has the structure of an elementary dyadic shift, and to prove the lemma we only need to estimate $\|h_Q\|_\infty \|h_R\|_\infty$.

Using (4.7) we get for fixed cubes Q and R

$$\|h_Q\|_\infty \|h_R\|_\infty \leq \frac{\ell(Q)^{\alpha/2} \ell(R)^{\alpha/2}}{D(Q, R)^{d+\alpha}} 2^{(m+n)\alpha/2} \cdot \frac{D(Q, R)^{d+\alpha}}{\ell(M)^{d+\alpha}}$$

$$= \frac{1}{\ell(M)^d} \cdot \frac{\ell(Q)^{\alpha/2} \ell(R)^{\alpha/2}}{\ell(M)^\alpha} 2^{(m+n)\alpha/2} = \frac{1}{\ell(M)^d},$$

because $\ell(Q)/\ell(M) = 2^{-m}$, $\ell(R)/\ell(M) = 2^{-n}$.

So, the above sum is indeed an elementary dyadic shift with parameters m, n. Summing over all j, k we get the conclusion of the lemma.   □

4.3. Proof of Theorem 4.1. As we explained before, see Lemma 4.3, we can represent T as the average

$$T = \pi_{\text{good}}^{-1}\mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(Q)\leq\ell(R), \\ Q \text{ is good}}} \Delta_R T\Delta_Q + \pi_{\text{good}}^{-1}\mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(R)<\ell(Q), \\ R \text{ is good}}} \Delta_R T\Delta_Q;$$

here and below in this section the averages $\mathbf{E}_\Omega$ are understood in the weak sense, as equalities of the bilinear forms for $f, g \in C_0^\infty$. As it was explained before in the proof of Lemma 4.3, see Remark 4.4 there, in this case we can freely interchange the summation and expectation (integration) $\mathbf{E}_\Omega$.

Recalling the decomposition

$$T = \tilde{T}_\omega + \Pi_T^\omega + (\Pi_{T*}^\omega)^*,$$

and using the fact that for $Q, R \in \mathcal{D}_\omega$

$$\Delta_R \Pi_T^\omega \Delta_Q = 0, \qquad \Delta_Q (\Pi_{T*}^\omega)^* \Delta_R = 0$$

if $\ell(Q) \leq \ell(R)$, we can write

(4.8)
$$T = \pi_{\text{good}}^{-1}\mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(Q)\leq\ell(R), \\ Q \text{ is good}}} \Delta_R \tilde{T}_\omega \Delta_Q + \pi_{\text{good}}^{-1}\mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(R)<\ell(Q), \\ R \text{ is good}}} \Delta_R \tilde{T}_\omega \Delta_Q$$

$$+ \pi_{\text{good}}^{-1}\mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(Q)\leq\ell(R), \\ Q \text{ is good}}} \Delta_R (\Pi_{T*}^\omega)^* \Delta_Q + \pi_{\text{good}}^{-1}\mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(R)<\ell(Q), \\ R \text{ is good}}} \Delta_R \Pi_T^\omega \Delta_Q.$$

Lemma 4.8. *For the paraproducts Π_T^ω one has*

$$\mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(R)<\ell(Q), \\ R \text{ is good}}} \Delta_R \Pi_T^\omega \Delta_Q = \mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(R)\leq\ell(Q), \\ R \text{ is good}}} \Delta_R \Pi_T^\omega \Delta_Q = \pi_{\text{good}}\mathbf{E}_\Omega \Pi_T^\omega.$$

Proof. It is not hard to see from the definition of the paraproduct that for $f \in L^2$

$$\sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(R)<\ell(Q), \\ R \text{ is good}}} \Delta_R \Pi_T^\omega \Delta_Q f = \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(R)\leq\ell(Q), \\ R \text{ is good}}} \Delta_R \Pi_T^\omega \Delta_Q f = \sum_{\substack{R\in\mathcal{D}_\omega, \\ R \text{ is good}}} (\Delta_R T\mathbf{1})\mathbb{E}_R f.$$

Applying $\mathbf{E}_\Omega$ we get that

$$\mathbf{E}_\Omega \sum_{\substack{R \in \mathcal{D}_\omega, \\ R \text{ is good}}} (\Delta_R T\mathbf{1})\mathbb{E}_R f = \sum_{k \in \mathbb{Z}} \mathbf{E}_\Omega \mathbf{E}_{\mathfrak{A}_k} \sum_{\substack{R \in \mathcal{D}_\omega, \\ \ell(R)=2^k}} (\Delta_R T\mathbf{1})(\mathbb{E}_R f)\mathbf{1}_{R \text{ is good}}(\omega)$$

$$= \pi_{\text{good}} \sum_{k \in \mathbb{Z}} \mathbf{E}_\Omega \sum_{\substack{R \in \mathcal{D}_\omega, \\ \ell(R)=2^k}} (\Delta_R T\mathbf{1})\mathbb{E}_R f = \pi_{\text{good}} \mathbf{E}_\Omega \Pi_T^\omega f;$$

here we again used the fact that, by (4.1), $\mathbf{E}_{\mathfrak{A}_k} \mathbf{1}_{R \text{ is good}}(\omega) = \pi_{\text{good}}$ for $R \in \mathcal{D}_\omega, \ell(R) = 2^k$.
□

By Lemma 4.8 the second line in (4.8) is $\mathbf{E}_\Omega(\Pi_T^\omega + (\Pi_{T*}^\omega)^*)$. We know that the para-products Π_T^ω and $(\Pi_{T*}^\omega)^*$ are (up to a constant factor $C = C(\alpha, d, C_{\text{cz}}, \|T\|)$) generalized dyadic shifts with parameters 0, 1 and 1, 0 respectively.

So to prove the theorem we need to represent the first line in (4.8) as the average of dyadic shifts. Let us represent the first term. For $m, n \in \mathbb{Z}_+, m \geq n$, define the dyadic shifts $\mathbb{S}_{m,n}^\omega$ as

$$\mathbb{S}_{m,n}^\omega = \sum_{M \in \mathcal{D}_\omega} \sum_{\substack{Q,R \in \mathcal{D}_\omega:Q,R\subset M, \\ \ell(Q)=2^{-m}\ell(M), \\ \ell(R)=2^{-n}\ell(M), \\ Q \text{ is good}}} \pi(Q|R) \cdot \rho_{Q,R}^{-1} \cdot 2^{(m+n)\alpha/2} \cdot \frac{D(Q,R)^{d+\alpha}}{\ell(M)^{d+\alpha}} \Delta_R \tilde{T}_\omega \Delta_Q,$$

where

$$\pi(Q|R) = \mathbf{P}\{Q \text{ is good}|R\} = \mathbf{E}_{\Omega[R]}\mathbf{1}_{Q \text{ is good}}$$

(note that $\ell(Q) \leq \ell(R)$). The weights $\rho_{Q,R}, Q, R \in \mathcal{D}_\omega$, are defined by

$$(4.9) \qquad \rho_{Q,R} := \mathbf{E}_{\Omega[R]} \sum_{M \in \mathcal{D}_\omega:Q,R\subset M} \frac{D(Q,R)^{d+\alpha}}{\ell(M)^{d+\alpha}} \cdot \mathbf{1}_{Q \text{ is good}}(\omega);$$

note that in the above expression we assume (can assume) that the variables $\omega_j, j < k$, determining the position of R (and so of Q) are fixed.

Remark 4.9. In general, $\rho_{Q,R}$ can be zero. However, it is not hard to see that $\rho_{Q,R} > 0$ if $\pi(Q|R) > 0$, so the dyadic shifts $\mathbb{S}_{m,n}^\omega$ are well defined.

Averaging we get

$$\mathbf{E}_\Omega \sum_{m,n\in\mathbb{Z}:m\geq n} 2^{-(m+n)\alpha/2}\mathbb{S}_{m,n}^\omega$$

$$= \mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(Q)\leq\ell(R), \\ \pi(Q|R)\neq0}} \sum_{\substack{M\in\mathcal{D}_\omega, \\ Q,R\subset M}} \pi(Q|R) \cdot \rho_{Q,R}^{-1} \frac{D(Q,R)^{d+\alpha}}{\ell(M)^{d+\alpha}} \cdot \mathbf{1}_{Q \text{ is good}}(\omega)\Delta_R \tilde{T}_\omega \Delta_Q$$

$$= \mathbf{E}_\Omega \sum_{\substack{Q,R\in\mathcal{D}_\omega, \\ \ell(Q)\leq\ell(R), \\ \pi(Q|R)\neq0}} \mathbf{E}_{\Omega[R]}\pi(Q|R) \cdot \rho_{Q,R}^{-1} \cdot \Delta_R \tilde{T}_\omega \Delta_Q \sum_{\substack{M\in\mathcal{D}_\omega, \\ Q,R\subset M}} \frac{D(Q,R)^{d+\alpha}}{\ell(M)^{d+\alpha}} \cdot \mathbf{1}_{Q \text{ is good}}(\omega)$$

and recalling the definition of $\rho_{Q,R}$ we conclude

$$\mathbf{E}_{\Omega} \sum_{m,n\in\mathbb{Z}:m\geq n} 2^{-(m+n)\alpha/2}\mathbb{S}^{\omega}_{m,n} = \mathbf{E}_{\Omega} \sum_{\substack{Q,R\in\mathscr{D}_{\omega}, \\ \ell(Q)\leq\ell(R)}} \pi(Q|R)\Delta_R\tilde{T}_{\omega}\Delta_Q.$$

On the other hand

$$\mathbf{E}_{\Omega} \sum_{\substack{Q,R\in\mathscr{D}_{\omega}, \\ \ell(Q)\leq\ell(R), \\ Q \text{ is good}}} \Delta_R\tilde{T}_{\omega}\Delta_Q = \sum_{k\in\mathbb{Z}} \mathbf{E}_{\Omega}\mathbf{E}_{\mathfrak{A}_k} \sum_{\substack{Q,R\in\mathscr{D}_{\omega}, \\ \ell(Q)\leq\ell(R)=2^k}} \mathbf{1}_{Q \text{ is good}}(\omega) \cdot \Delta_R\tilde{T}_{\omega}\Delta_Q$$

$$= \sum_{k\in\mathbb{Z}} \mathbf{E}_{\Omega} \sum_{\substack{Q,R\in\mathscr{D}_{\omega}, \\ \ell(Q)\leq\ell(R)=2^k}} (\mathbf{E}_{\Omega[R]}\mathbf{1}_{Q \text{ is good}})\Delta_R\tilde{T}_{\omega}\Delta_Q$$

$$= \mathbf{E}_{\Omega} \sum_{\substack{Q,R\in\mathscr{D}_{\omega}, \\ \ell(Q)\leq\ell(R)}} \pi(Q|R)\Delta_R\tilde{T}_{\omega}\Delta_Q,$$

so

$$\mathbf{E}_{\Omega} \sum_{m,n\in\mathbb{Z}:m\geq n} 2^{-(m+n)\alpha/2}\mathbb{S}^{\omega}_{m,n} = \mathbf{E}_{\Omega} \sum_{\substack{Q,R\in\mathscr{D}_{\omega}, \\ \ell(Q)\leq\ell(R), \\ Q \text{ is good}}} \Delta_R\tilde{T}_{\omega}\Delta_Q.$$

It now remains to show that $\mathbb{S}^{\omega}_{m,n}$ are (up to a constant factor) the dyadic shifts. The operators $\mathbb{S}^{\omega}_{m,n}$ have the appropriate structure, so we only need to prove the estimates, i.e., prove that the weights $\rho_{Q,R}$ are uniformly bounded away from 0. The necessary estimate follows from Lemma 4.10 below.

So, we have decomposed the first term in (4.8) as the average of dyadic shifts. The decomposition of the second term is carried out similarly, so Theorem 4.1 is proved (modulo Lemma 4.10). □

Lemma 4.10. *Let* $Q, R \in \mathscr{D}_{\omega}$, $\ell(Q) \leq \ell(R)$. *Then*

(i) $\pi(Q|R) > 0$ *if and only if* Q *is "good up to the level of* R*", meaning that*

$$(4.10) \quad \text{dist}(Q, Q') \geq \ell(Q)^{\gamma}\ell(Q')^{1-\gamma} \quad \forall Q' \in \mathscr{D}_{\omega} : 2^{r_0}\ell(Q) < \ell(Q') \leq \ell(R);$$

note that the cubes Q' *do not depend on the variables* ω_j, $j \geq k$ *where* $2^k = \ell(R)$.

(ii) *There exists a constant* $c = c(d, r_0, \gamma)$ *such that*

$$\rho_{Q,R} \geq c(d, r_0) \quad \forall Q, R \in \mathscr{D}_{\omega} : \pi(Q|R) \neq 0.$$

Proof. We want to estimate conditional probability end expectation with R and Q fixed. That means the lattice up to the level of R is fixed, so nothing changes if we replace R by a cube in the same level. So, without loss of generality we can assume that $Q \subset R$.

Let us first consider a special case. Let $\ell(R) = \ell(Q)2^s$, where

$$(4.11) \qquad\qquad s \geq 2/\gamma + r_0 \cdot (1-\gamma)/\gamma,$$

and let

$$\text{dist}(Q, \partial R) \geq \frac{1}{4}\ell(R).$$

Then the estimate (4.11) implies that

$$\ell(Q)^\gamma [2^{r_0}\ell(R)]^{1-\gamma} = 2^{-s\gamma} 2^{r_0(1-\gamma)}\ell(R) \leq \frac{1}{4}\ell(R),$$

meaning that for any cube $M \in \mathcal{D}_\omega$, $\ell(R) \leq \ell(M) \leq 2^{r_0}\ell(R)$ (assuming that the lattice $\mathcal{D}_\omega$ is fixed up to the level of R)

$$(4.12) \qquad \ell(Q)^\gamma \ell(M)^{1-\gamma} \leq \frac{1}{4}\ell(R) \leq \text{dist}(Q, \partial R) \leq \text{dist}(Q, \partial M).$$

On the other hand, if $\ell(M) > 2^{r_0}\ell(R)$ and the pair R, M is good, meaning that

$$\text{dist}(R, \partial M) \geq \ell(R)^\gamma \ell(M)^{1-\gamma},$$

then

$$(4.13) \qquad \text{dist}(Q, \partial M) \geq \ell(Q)^\gamma \ell(M)^{1-\gamma},$$

so the pair Q, M is also good.

Therefore, if the cube R is good, then Q is good as well: as we just discussed, the inequality (4.13) holds if $\ell(M) > 2^{r_0}\ell(R)$, and it holds for $\ell(R) \leq \ell(M) \leq 2^{r_0}\ell(R)$ by (4.12). And the assumption (4.10) covers the remaining cases.

So, in our special case $\pi(Q|R) \geq \pi_{\text{good}}$.

The general case can be easily reduced to this special situation. Namely, if $Q \subsetneq R$, then with probability at least 2^{-d} the parent $\tilde{R}$ of R satisfies

$$\text{dist}(Q, \partial\tilde{R}) \geq \frac{1}{4}\ell(\tilde{R});$$

one can easily see that for $d = 1$, and considering the coordinates independently, one gets the conclusion.

Applying this procedure $s_0 - 1$ times, where s_0 is the smallest integer satisfying (4.11), we arrive (with probability at least $2^{-(s_0-1)d}$) at the special situation we just discussed. Therefore for $Q \subsetneq R$ (equivalently $\ell(Q) < \ell(R)$) statement (i) is proved with the estimate

$$(4.14) \qquad \pi(Q|R) \geq 2^{-(s_0-1)d}\pi_{\text{good}} =: \pi_0.$$

Finally, if $Q = R$, we arrive with probability 1 at the previous situation, so the statement (i) is now completely proved with estimate (4.14).

The statement (ii) is now easy. First note, that if $\tau \in \mathbb{Z}$ is such that $2^\tau > D(Q, R)$, then

$$(4.15) \qquad \mathbf{P}\{\exists M \in \mathcal{D}_\omega : \ell(M) = 2^\tau, \ Q, R \subset M \mid R\} \geq 1 - d \cdot 2D(Q, R)/2^\tau.$$

Indeed, in one dimension the probability that such M does not exist can be estimated above by $2D(Q, R)/2^\tau$, so to get the estimate of non-existence in $\mathbb{R}^d$ we can just multiply it by d. The extra factor 2 appears in the one-dimensional case because M cannot be moved continuously, but only in multiples of $\ell(R)$.

Define

$$\tau_0 := \lfloor \log_2(d D(Q, R)/\pi_0) \rfloor + 3,$$

so

$$d \cdot 2D(Q, R)/2^{\tau_0} \leq \pi_0/2.$$

Comparing the estimates (4.14) and (4.15) of probabilities, we can get that for fixed Q and R the probability that Q is good and that $Q, R \subset M$ for some $M \in \mathcal{D}_\omega, \ell(M) = 2^{\tau_0}$, is at least $\pi_0/2$.

On the other hand, the definition of τ_0 implies that $\ell(M) = 2^{\tau_0} \leq 8 \cdot d \cdot D(Q, R)/\pi_0$, so

$$D(Q, R)/\ell(M) \geq \pi_0/8.$$

Therefore, the contribution to the sum (4.9) defining $\rho_{Q,R}$ of the term with such M alone is at least

$$(\pi_0/8)^{d+\alpha} \pi_0/2.$$

That proves (ii) and so the lemma.   □

5. Sharp weighted estimate of dyadic shifts

Recall that for a dyadic shift $\mathbb{S}$ with parameters m and n its complexity is $r := \max(m, n)$. In this section we assume that a dyadic lattice $\mathcal{D}$ is fixed. Let $\mathbb{S}$ be an elementary (possibly generalized) dyadic shift

$$(5.1) \qquad \mathbb{S}f(x) = \sum_{Q \in \mathcal{D}} \int_Q a_Q(x, y) f(y) dy,$$

where a_Q are supported on $Q \times Q$, $\|a_Q\|_\infty \leq |Q|^{-1}$ (in this section we will incorporate $|Q|^{-1}$ into a_Q). Let $\mathcal{A} \subset \mathcal{D}$ be a collection of dyadic cubes. Define the restricted dyadic shift $\mathbb{S}_{\mathcal{A}}$ by taking the sum in (5.1) only over $Q \in \mathcal{A}$.

As it was shown by Theorem 4.1 that a Calderón–Zygmund operator T is a weighted average of dyadic shifts with exponentially (in complexity of shifts) decaying weights, to prove Theorem 1.1 it is sufficient to get an estimate of the norm of dyadic shifts which is polynomial in complexity. The following theorem, indeed, achieves a norm bound which is *quadratic* in complexity. This is the second new main result of this paper and represents a substantial quantitative improvement over earlier sharp weighted bounds for dyadic shifts [5, 16], which were exponential in complexity. Note that the paper [14], while using dyadic shifts as auxiliary operators in the original proof of Theorem 1.1, circumvented the question of actually estimating their norm. This is achieved in [14] by going through the test conditions of rather involved paper [30].

Theorem 5.1. *Let $\mathbb{S}$ be an elementary (possibly generalized) dyadic shift of complexity r in $\mathbb{R}^d$, such that all restricted shifts $\mathbb{S}_{\mathcal{A}}$ are uniformly bounded in L^2,*

$$(5.2) \qquad \sup_{\mathcal{A} \subset \mathcal{D}} \|\mathbb{S}_{\mathcal{A}}\|_{L^2 \to L^2} =: B_2 = B_{\mathbb{S}} < \infty.$$

Then for any A_2 weight w

$$(5.3) \qquad \|\mathbb{S}f\|_{L^2(w)} \le C 2^{3d/2}(r+1)^2(B_2^2+1)[w]_{A_2}\|f\|_{L^2(w)}, \quad \forall f \in L^2(w),$$

where C is an absolute constant.

Note that for dyadic shifts we are considering (that is non-generalized dyadic shifts and paraproducts), the assumption about uniform boundedness of $\mathbb{S}_{\mathcal{A}}$ is satisfied automatically. Namely, any non-generalized dyadic shift is a contraction in L^2, so (5.2) holds with $B = 1$. It is also easy to see that for the paraproducts $\|\mathbb{S}_{\mathcal{A}}\|_{L^2 \to L^2} \le \|\mathbb{S}\|_{L^2 \to L^2}$.

The estimate (5.3) with C depending exponentially on r was proved (for non-generalized dyadic shifts) in [16]. However, careful analysis of proofs there allows (after some modifications) to obtain polynomial estimates.

Compared to [16], the main new ingredients here are

- the sharp two weight estimate of Haar shifts, see the above Theorem 3.4, which is essentially the main result of [28] (with the additional assumptions about "size" of the operator), with the dependence of the estimates on all parameters spelled out;

- Proposition 5.1 of [14], reproduced as Theorem 5.2 below, which gives a *linear* in complexity of $\mathbb{S}$ estimate of the unweighted weak L^1 norm of $\mathbb{S}$; the corresponding estimate in [16] was exponential in complexity.

Replacing f in (5.3) by fw^{-1} and noticing that $\|fw^{-1}\|_{L^2(w)} = \|f\|_{L^2(w^{-1})}$ we can rewrite it as

$$(5.4) \quad \|\mathbb{S}(fw^{-1})\|_{L^2(w)}$$
$$\le C 2^{3d/2}(r+1)^2(B_2^2+1)[w]_{A_2}\|f\|_{L^2(w^{-1})}, \quad \forall f \in L^2(w^{-1}),$$

so we are in the settings of Theorem 3.4 with $d\mu = w^{-1}dx$, $d\nu = w\,dx$. By Theorem 3.4, to prove estimate (5.4) it is sufficient to show that

$$\int_Q |\mathbb{S}(\mathbf{1}_Q w^{-1})|^2 w\,dx \le B[w]_{A_2}^2 w^{-1}(Q), \quad \forall f \in L^2(w^{-1}),$$

$$(5.5) \qquad \int_Q |\mathbb{S}(\mathbf{1}_Q w)|^2 w^{-1}dx \le B[w]_{A_2}^2 w(Q), \quad \forall f \in L^2(w),$$

where

$$B^{1/2} = C 2^d (r+1)(B_2^2+1)$$

with an absolute constant C.

Since $[w^{-1}]_{A_2} = [w]_{A_2}$, one can get one estimate from the other by replacing w by w^{-1}. Thus, to prove Theorem 5.1 and so the main result (Theorem 1.1) we only need to prove one of the above estimates, for example (5.5).

The rest of the section is devoted to proving (5.5).

5.1. Weak type estimates for dyadic shifts. Let $\|\mathbb{S}\|_2$ be a shorthand for $\|\mathbb{S}\|_{L^2 \to L^2}$. We say that a shift $\mathbb{S}$ has *scales separated by r levels*, if all cubes Q with $a_Q \not\equiv 0$ in (5.1) satisfy $\log_2 \ell(Q) \equiv j \mod r$ for some fixed $j \in \{0, 1, \dots, r-1\}$.

The following result reproduces [14, Proposition 5.1] with an additional observation concerning shifts which have their scales separated. This seemingly technical variant allows us to obtain the asserted quadratic, rather than cubic, dependence on complexity in Theorem 5.1.

Theorem 5.2. *Let $\mathbb{S}$ be a generalized elementary dyadic shift with parameters m, n. Then $\mathbb{S}$ has weak type* 1-1 *with the estimate*

$$(5.6) \qquad \|\mathbb{S}\|_{L^1 \to L^{1,\infty}} \leq C(d, m, \|\mathbb{S}\|_2) = 2^{d+2}\|\mathbb{S}\|_2^2 + 1 + 4m,$$

meaning that for all $f \in L^1$ and for all $\lambda > 0$

$$\left|\{x : |\mathbb{S}f(x)| > \lambda\}\right| \leq \frac{C(d, m, \|\mathbb{S}\|_2)}{\lambda}\|f\|_1.$$

If $\mathbb{S}$ has scales separated by $r \geq m$ levels, then we have the improved estimate

$$\|\mathbb{S}\|_{L^1 \to L^{1,\infty}} \leq C(d, 1, \|\mathbb{S}\|_2) = 2^{d+2}\|S\|_2^2 + 5.$$

Proof. Our shift $\mathbb{S}$ can be written (see (3.2)) as

$$\mathbb{S}f(x) = \sum_{Q \in \mathcal{D}} \int a_Q(x, y) f(y) dy,$$

where a_Q is supported on $Q \times Q$ and $\|a_Q\|_\infty \leq |Q|^{-1}$ (we incorporated the factor $|Q|^{-1}$ from (3.2) into a_Q here). It follows from the representation (3.3) of a_Q that for fixed x the function $a_Q(x, \cdot)$ is constant on cubes $Q' \in \mathcal{D}, \ell(Q') < 2^{-m}\ell(Q)$.

To estimate its weak norm we use the standard Calderón–Zygmund decomposition at height $\lambda > 0$ with respect to the dyadic lattice $\mathcal{D}$. Namely, as it is well known, see for example [11, p. 286], given $f \in L^1$ there exists a decomposition $f = g + b, b = \sum_{Q \in \mathcal{Q}} b_Q$, where $\mathcal{Q} \subset \mathcal{D}$ is a collection of disjoint dyadic cubes, such that

 (i) $\|g\|_1 \leq \|f\|_1, \|g\|_\infty \leq 2^d \lambda$;

 (ii)   each function b_Q is supported on a cube Q and

$$\|b_Q\|_1 \leq 2 \cdot \|\mathbf{1}_Q f\|_1, \qquad \int_{\mathbb{R}^d} b_Q \, dx = 0;$$

(iii) $\displaystyle\sum_{Q \in \mathcal{Q}} |Q| \leq \lambda^{-1}\|f\|_1.$

The property (i) of the Calderón–Zygmund decomposition implies that

$$(5.7) \qquad\qquad\qquad \|f\|_2^2 \leq 2^d \lambda \|f\|_1.$$

As usual, we can estimate

$$\left|\{x : |\mathbb{S}f(x)| > \lambda\}\right| \leq \left|\{x : |\mathbb{S}g(x)| > \lambda/2\}\right| + \left|\{x : |\mathbb{S}b(x)| > \lambda/2\}\right|$$

(one of the two terms should be at least half of the sum). The measure of the first set is estimated using the boundedness of $\mathbb{S}$ in L^2:

$$\left|\{x : |\mathbb{S}g(x)| > \lambda/2\}\right| \leq \|\mathbb{S}\|_2^2 \|g\|_2^2 \frac{4}{\lambda^2} \leq \|\mathbb{S}\|_2^2 \frac{2^{d+2}}{\lambda}\|f\|_1,$$

where $\|\mathbb{S}\|_2$ is the shorthand for $\|\mathbb{S}\|_{L^2 \to L^2}$; we used (5.7) to get the second inequality.

To estimate $|\{x : |\mathbb{S}b(x)| > \lambda/2\}|$ we fix a $Q \in \mathcal{Q}$ and write a pointwise inequality:

$$|\mathbb{S}b_Q(x)| \leq \sum_{R \in \mathcal{D}: Q \subsetneq R} \left| \int_R a_R(x, y) b_Q(y) dy \right| + \left| \sum_{R \in \mathcal{D}: R \subset Q} \int_R a_R(x, y) b_Q(y) dy \right|.$$

Therefore, summing in $Q \in \mathcal{Q}$, we get

$$|\mathbb{S}b(x)| \leq \sum_{Q \in \mathcal{Q}} \sum_{R \in \mathcal{D}: Q \subsetneq R} \left| \int_R a_R(x, y) b_Q(y) dy \right| + \sum_{Q \in \mathcal{Q}} \left| \sum_{R \in \mathcal{D}: R \subset Q} \int_R a_R(x, y) b_Q(y) dy \right|$$
$$=: A(x) + B(x).$$

Hence, using again the fact that one of the two terms should be at least a half of the sum, we can estimate

$$\left| \{x : |\mathbb{S}b(x)| > \lambda/2\} \right| \leq \left| \{x : A(x) > \lambda/2\} \right| + \left| \{x : B(x) > 0\} \right|.$$

The second set is obviously inside $\bigcup_{Q \in \mathcal{Q}} Q$: indeed the function $B(x)$ vanishes outside this set because $a_R(x, y) = 0$ for all $x \notin R$, and $R \subset Q$. So, using the property (iii) of the Calderón–Zygmund decomposition, we can estimate the measure of the second set as

$$\left| \{x : B(x) > 0\} \right| \leq \sum_{Q \in \mathcal{Q}} |Q| \leq \frac{1}{\lambda} \|f\|_1.$$

To estimate the first measure we want to show that $\|A\|_1 \leq C \|f\|_1$, then clearly

(5.8) $$\left| \{x : A(x) > \lambda/2\} \right| \leq \frac{2}{\lambda} \|A\|_1 \leq \frac{2C}{\lambda} \|f\|_1.$$

We will estimate the norm of each term in A separately. Let us fix $Q \in \mathcal{Q}$ and let us consider

$$A_Q(x) := \sum_{R \in \mathcal{D}, Q \subsetneq R} \left| \int_R a_R(x, y) b_Q(y) dy \right|.$$

Since the function b_Q is orthogonal to constants, and the function $a_R(x, \cdot)$ is constant on cubes $Q \in \mathcal{D}$, $\ell(Q) < 2^{-m} \ell(R)$, we can see that the only cubes R which may contribute to A_Q are the ancestors of Q of orders $1, \ldots, m$. So, in general, there are at most m non-zero terms in A_Q; if $\mathbb{S}$ has scales separated by $r \geq m$ levels, there is at most one.

Recalling that for an integral operator T with kernel K

$$\|T\|_{L^1 \to L^1} = \operatorname*{ess\,sup}_y \|K(\cdot, y)\|_1,$$

we can see that the integral operator with kernel a_R is a contraction in L^1. Since at most m such operators contribute to A_Q,

$$\|A_Q\|_1 \leq m \|b_Q\|_1 \leq 2m \|\mathbf{1}_Q f\|_1;$$

the last inequality here holds because of property (ii) of the Calderón–Zygmund decomposition.

Summing over all $Q \in \mathcal{Q}$ we get

$$\|A\|_1 \le 2m \sum_{Q \in \mathcal{Q}} \|\mathbf{1}_Q f\|_1 \le 2m \|f\|_1.$$

So, see (5.8),

$$\left|\{x : A(x) > \lambda/2\}\right| \le \frac{2}{\lambda} \|A\|_1 \le \frac{4m}{\lambda} \|f\|_1.$$

If $\mathbb{S}$ has scales separated by $r \ge m$ levels, we can take 1 in place of m in the last few estimates.

$\square$

Using this improved weak type estimate one can get the desired estimate (5.5) by following the proof in [16] and keeping track of the constants. However, there are several other places in [16], where the curse of exponentiality appears. So for the convenience of the reader, we are doing all necessary estimates below. Note that an analogous modification of [16] was already carried out in [14]; here we present yet another argument in the spirit of [16] but with modifications pertinent to eliminating the curse of exponentiality.

5.2. First slicings. Let us fix $Q_0 \in \mathcal{D}$, and let us prove estimate (5.5) for $Q = Q_0$. Recall that $\mathbb{S}$ is an integral operator with kernel $\sum_{Q \in \mathcal{D}} a_Q(x, y)$, where a_Q as in the previous section ($|Q|^{-1}$ is incorporated in a_Q).

Define

$$f_Q(x) := \int_{Q_0} a_Q(x, y) w(y) dy,$$

so

$$\mathbb{S}(\mathbf{1}_{Q_0} w) = \sum_{Q \in \mathcal{D}: Q \cap Q_0 \neq \varnothing} f_Q =: f.$$

We can split f into "inner" and "outer" parts,

$$f = \sum_{Q \in \mathcal{D}: Q \subset Q_0} f_Q + \sum_{Q \in \mathcal{D}: Q_0 \subsetneqq Q} f_Q =: f_{\mathrm{i}} + f_{\mathrm{o}}.$$

The "outer" part f_{o} is easy to estimate. Since $\|a_Q(x, \cdot)\|_\infty \le |Q|^{-1}$, we can write for $Q_0 \subsetneqq Q$

$$|f_Q(x)| \le w(Q_0) |Q|^{-1}$$

and summing over all Q, $Q_0 \subsetneqq Q$,

$$|f_{\mathrm{o}}(x)| \le |Q_0|^{-1} w(Q_0) \sum_{Q \in \mathcal{D}: Q_0 \subsetneqq Q} |Q|^{-1} |Q_0|$$

$$= |Q_0|^{-1} w(Q_0) \sum_{k=1}^{\infty} 2^{-kd} \le |Q_0|^{-1} w(Q_0).$$

Therefore,

$$\int_{Q_0} |f_{\mathrm{o}}|^2 w^{-1} \le |Q_0|^{-2} w(Q_0)^2 w^{-1}(Q_0) \le [w]_{A_2} w(Q_0),$$

so $\|\mathbf{1}_{Q_0} f_{\mathrm{o}}\|_{L^2(w^{-1})} \le [w]_{A_2}^{1/2} w(Q_0)^{1/2}$, and it only remains to estimate $\|f_{\mathrm{i}}\|_{L^2(w^{-1})}$.

Now we perform the first splitting. Let r be the complexity of the shift $\mathbb{S}$. Let us split the lattice $\mathcal{D}$ into $r + 1$ lattices $\mathcal{D}_r^j$, $j = 0, 1, \ldots, r$, where each lattice $\mathcal{D}_r^j$ consists of the cubes $Q \in \mathcal{D}$ of size $2^{j-(r+1)\tau}$, $\tau \in \mathbb{Z}$.

If we can show that uniformly in j

$$(5.9) \qquad \int_{Q_0} \left| \sum_{Q \in \mathcal{D}_r^j} f_Q \right|^2 w^{-1} \le C \, 2^{2d} (B_2^2 + 1)^2 [w]_{A_2}^2 w(Q_0),$$

where C is an absolute constant, then we are done. Indeed, (5.9) gives us an estimate of the norm of the sum in the left-hand side. Taking the sum over all $j = 1, 2, \ldots, r$ we only multiply the estimate of the norm by $r + 1$. So, to get from the estimate (5.9) to the desired estimate (5.5), we just need to multiply the right side of (5.9) by $(r + 1)^2$.

The main reason for this splitting of $\mathcal{D}$ is that it simplifies the structure meaning that for $Q \in \mathcal{D}_r^j$ the function f_Q is constant on the children of Q in the lattice $\mathcal{D}_r^j$. Also note that the shift

$$\mathbb{S}^j f(x) := \sum_{Q \in \mathcal{D}_r^j} \int_Q a_Q(x, y) f(y) dy$$

has scales separated by $r + 1 > m$ levels, and $\mathbf{1}_{\mathcal{D}_r^j}(Q) \cdot f_Q = \mathbb{S}^j (\mathbf{1}_{Q_0} w)$.

Let us fix j, and let us from now on consider the lattice $\mathcal{D}_r := \mathcal{D}_r^j$. Since j is not important in what follows, we will skip it and use the notation $\mathcal{D}_r$, freeing the symbol j for use in a different context. We also denote $\mathbb{S}^j$ simply by $\mathbb{S}$, bearing in mind the separation of scales which allows the use of the sharper estimate in the weak-type bound of Theorem 5.2.

Now we split the lattice $\mathcal{D}_r$ into the collections $\mathcal{Q}_k$, $k \in \mathbb{Z}_+$, $k < \log_2([w]_{A_2})$, where each $\mathcal{Q}_k$ is the set of all cubes $Q \in \mathcal{D}_r$ such that

$$(5.10) \qquad 2^k \le \frac{w(Q)}{|Q|} \cdot \frac{w^{-1}(Q)}{|Q|} < 2^{k+1}.$$

We want to show that

$$(5.11) \qquad \int_{Q_0} \left| \sum_{Q \in \mathcal{Q}_k : Q \subset Q_0} f_Q \right|^2 w^{-1} \le C_1 2^k [w]_{A_2} w(Q_0),$$

where $C_1 = C 2^{2d}(B_2^2 + 1)^2$ is the constant in the right side of (5.9). Then, using the triangle inequality and summing the geometric progression we get

$$\left\| \mathbf{1}_{Q_0} \sum_{Q \in \mathcal{D}_r} f_Q \right\|_{L^2(w^{-1})} \le C_1^{1/2} [w]_{A_2}^{1/2} \sum_{k \in \mathbb{Z}_+ : k < \log_2([w]_{A_2})} 2^{k/2} w(Q_0)$$

$$< 4 C_1^{1/2} [w]_{A_2} w(Q_0),$$

so (5.11) implies that (5.9) holds with $C = 16C_1$.

So, we reduced the main result to the estimate (5.11) with $C_1 = C 2^{2d}(B_2^2 + 1)^2$. Note that if we prove (5.11) for $Q_0 \in \mathcal{Q}_k$, then we are done, because for general Q_0 we can add up the estimate for maximal subcubes of Q_0 belonging to $\mathcal{Q}_k$.

5.3. Stopping moments and Corona decomposition. Let us suppose that the weight w and the lattices $\mathcal{D}_r$ and $\mathcal{Q} = \mathcal{Q}_k \subset \mathcal{D}_r$ described above are fixed.

Given a cube $Q_0 \in \mathcal{Q} = \mathcal{Q}_k$ let us construct the generations

$$\mathcal{G}_\tau^* = \mathcal{G}_\tau^*(Q_0) = \mathcal{G}_\tau^*(Q_0, w, \mathcal{Q}), \quad \tau \in \mathbb{Z}_+$$

of stopping cubes as follows. Define the initial generation $\mathcal{G}_0^*$ to be the cube Q_0.

For all cubes $Q \in \mathcal{G}_\tau^*$ we consider maximal cubes $Q' \in \mathcal{Q}$, $Q' \subset Q$ such that

$$\frac{w(Q')}{|Q'|} > 4\frac{w(Q)}{|Q|};$$

the collection of all such cubes Q' is the next generation $\mathcal{G}_{\tau+1}^*$ of the stopping cubes.

Let $\mathcal{G}^* = \mathcal{G}^*(Q_0) := \bigcup_{\tau \geq 0} \mathcal{G}_\tau^*$ be the collection of all stopping cubes.

Note that if we start constructing stopping moments from a cube $Q \in \mathcal{G}^*$, the stopping moments $\mathcal{G}^*(Q)$ will agree with $\mathcal{G}^*$, meaning that

$$\mathcal{G}^*(Q) = \{Q' \in \mathcal{G}^* : Q' \subset Q\}.$$

Let us introduce the last piece of notation. For a cube $Q \in \mathcal{G}^*$ let us define

$$\mathcal{Q}(Q) := \{Q' \in \mathcal{Q} : Q' \subset Q\},$$

and let

$$\mathcal{P}(Q) := \mathcal{Q}(Q) \setminus \bigcup_{Q' \in \mathcal{G}^*: Q' \subsetneq Q} \mathcal{Q}(Q').$$

The above definitions make sense for arbitrary $Q \in \mathcal{Q}$, but we will use it only for $Q \in \mathcal{G}^*$, so we included this assumption in the definition. Note that for $Q_0 \in \mathcal{Q}$ the set $\mathcal{Q}(Q_0)$ admits the following disjoint decomposition:

$$(5.12) \qquad \mathcal{Q}(Q_0) = \bigcup_{Q \in \mathcal{Q}^*(Q_0)} \mathcal{P}(Q).$$

Properties of stopping moments. It follows from the construction of $\mathcal{G}^*$ that if $R \in \mathcal{G}^*$ and Q is a maximal cube in $\mathcal{G}^*$ such that $Q \subsetneq R$, then

$$(5.13) \qquad \frac{w(Q)}{|Q|} > 4\frac{w(R)}{|R|}.$$

The estimate (5.13) implies

$$(5.14) \qquad |Q| \leq \frac{|R|}{4} \cdot \frac{w(Q)}{w(R)},$$

and summing over all such maximal $Q \in \mathcal{G}^*$, $Q \subsetneq R$ (assume that $R \in \mathcal{G}_\tau^*$) we get

$$(5.15) \qquad \left| \bigcup_{Q \in \mathcal{G}^*: Q \subsetneq R} Q \right| = \sum_{Q \in \mathcal{G}_{\tau+1}^*: Q \subsetneq R} |Q| \leq \frac{|R|}{4w(R)} \sum_{Q \in \mathcal{G}_{\tau+1}^*: Q \subsetneq R} w(Q) \leq \frac{1}{4}|R|,$$

for all $R \in \mathcal{G}^*$.

Repeating this estimate for each Q and summing over the generations we get

$$\sum_{Q \in \mathscr{G}^*: Q \subsetneq R} |Q| \le |R| \sum_{n=1}^{\infty} 4^{-n} = \frac{1}{3}|R|.$$

Adding $|R|$ to this sum we get the following *Carleson property* of the stopping moments $\mathscr{G}^*$:

$$(5.16) \qquad \sum_{Q \in \mathscr{G}^*: Q \subset R} |Q| \le \frac{4}{3}|R|.$$

It is easy to see that this estimate holds for all $R \in \mathscr{D}$, not just for $R \in \mathscr{G}^*$: one just needs to consider maximal cubes $R' \in \mathscr{G}^*$, $R' \subset R$ and apply (5.16) to each of these cubes.

Iterating (5.15) and summing over all generations we get

$$(5.17) \qquad \left\| \sum_{Q \in \mathscr{G}^*, Q \subset R} \mathbf{1}_Q \right\|_2 \le |R|^{1/2} \sum_{k=0}^{\infty} 2^{-k} = 2|R|^{1/2}.$$

We need the following simple lemma.

Lemma 5.3. *For any $R \in \mathscr{D}$*

$$(5.18) \qquad \sum_{Q \in \mathscr{G}^*, Q \subset R} w(Q) \le C[w]_{A_2} w(R),$$

where C is an absolute constant.

Proof. The Carleson embedding theorem (see Theorem 3.6) applied to $\mathbf{1}_R$ together with the Carleson property (5.16) imply that

$$\sum_{Q \in \mathscr{G}^*, Q \subset R} \left(\fint_Q w^{1/2} \right)^2 |Q| \le C \|\mathbf{1}_R w^{1/2}\|_2^2 = C w(R)$$

(the best constant is $C = 4 \cdot 4/3$). But

$$\left(\fint w^{1/2} \right)^{-1} \le \fint w^{-1/2} \le \left(\fint w^{-1} \right)^{1/2} \quad \text{by Cauchy–Schwarz}$$

$$\le [w]_{A_2}^{1/2} \left(\fint_Q w \right)^{-1/2} \quad \text{because } \left(\fint_Q w \right)\left(\fint_Q w^{-1} \right) \le [w]_{A_2},$$

so

$$\fint_Q w \le [w]_{A_2} \left(\fint_Q w^{1/2} \right)^2$$

and the lemma is proved (with $C = 16/3$). $\qquad\square$

This proof was (essentially) presented in [38]. In [16] a different proof was presented, using a clever iteration argument and giving the better constant $C = 16/9$.

5.4. John–Nirenberg type estimates. Given a collection $\mathcal{A}$ of cubes, $\mathcal{A} \subset \mathcal{D}_r$, define the function $f_{\mathcal{A}}$ by

$$f_{\mathcal{A}} := \sum_{Q \in \mathcal{A}} f_Q.$$

For the cube $Q_0 \in \mathscr{G}^*$ consider the function $f_{\mathcal{Q}(Q_0)}$. By (5.12) the function $f_{\mathcal{Q}(Q_0)}$ can be decomposed as

$$(5.19) \qquad f_{\mathcal{Q}(Q_0)} = \sum_{R \in \mathscr{G}^*} f_{\mathcal{P}(R)},$$

where $\mathscr{G}^* := \mathscr{G}^*(Q_0)$ is the collection of stopping cubes.

The main reason for introducing this decomposition is that, as we will show below, the functions $f_{\mathcal{P}(R)}$ behave in many respects as BMO functions: they have exponentially decaying distribution functions, so, in particular all L^p norms for $p < \infty$ are equivalent.

In the proof of these facts the weak L^1 estimate of dyadic shifts (Theorem 5.2) is used.

The first lemma, which is [16, Lemma 3.15], is a simple observation, that for the John–Nirenberg estimates of the distribution function it is sufficient to have weak type estimates.

Recall that $\mathcal{D}_r$ is a 2^r-adic lattice, i.e., the children Q' of Q satisfy $\ell(Q') = 2^{-r}\ell(Q)$.

Definition 5.4. Let ϕ_Q, $Q \in \mathcal{D}_r$ be a collection of functions such that ϕ_Q is supported on Q and is constant on children (in $\mathcal{D}_r$) of Q. For $R_0 \in \mathcal{D}_r$ let $\phi_{R_0}^*$ be a maximal function

$$\phi_{R_0}^*(x) := \sup_{Q \in \mathcal{D}_r : Q \ni x} \left| \sum_{R \in \mathcal{D}_r : Q \subsetneq R \subset R_0} \phi_R(x) \right|.$$

Lemma 5.5. *Let ϕ_Q, $Q \in \mathcal{D}_r$ be a collection of functions such that*

(i) *ϕ_Q is supported on Q and constant on the children (in $\mathcal{D}_r$) of Q;*

(ii) *$\|\phi_Q\|_\infty \le 1$;*

(iii) *there exists $\delta \in (0,1)$ such that for all cubes $R \in \mathcal{D}_r$*

$$\left| \{ x \in R : \phi_R^*(x) > 1 \} \right| \le \delta |R|.$$

Then for all $R \in \mathcal{D}_r$ and for all $t \ge 0$

$$\left| \{ x \in R : \phi_R^*(x) > t \} \right| \le \delta^{(t-1)/2} |R|.$$

Proof. Let us prove the conclusion of the lemma for a fixed cube $R = R_0 \in \mathcal{D}_r$. Let $\mathcal{B}_1$ be the collection of all maximal cubes $Q \in \mathcal{D}_r$, $Q \subset R_0$ such that

$$(5.20) \qquad \left| \sum_{R \in \mathcal{D}_r : Q \subsetneq R \subset R_0} \phi_R(x) \right| > 1, \quad x \in Q;$$

note that the functions ϕ_R (and so the sum) are constant on the cube Q.

Define the set B_1,

$$B_1 := \bigcup_{Q \in \mathcal{B}_1} Q.$$

It follows from the construction that $\phi_{R_0}^* \le 1$ outside of B_1, and that for any $Q \in \mathcal{B}_1$ the sum in (5.20) is at most 2. Note also that by the assumption (iii) we have that $|B_1| \le \delta |R_0|$.

For each cube $\tilde{R} \in \mathcal{B}_1$ we repeat the above construction (with $\tilde{R}$ instead of R_0); we will get a collection of stopping cubes $\mathcal{B}_2$ and the set $B_2 = \bigcup_{Q \in \mathcal{B}_2} Q$, $B_2 \subset B_1$, $|B_2| \leq \delta^2 |R_0|$. It is easy to see that $\phi^*_{R_0} \leq 2 + 1 = 3$ outside of B_2 and that for any cube $Q \in \mathcal{B}_2$

$$\left| \sum_{R \in \mathcal{D}_r : Q \subsetneqq R \subset R_0} \phi_R(x) \right| \leq 4, \quad x \in Q$$

(sums outside of $\tilde{R} \in \mathcal{B}_1$ contribute at most 2, and the sums starting at $\tilde{R} \in \mathcal{B}_1$ contribute at most 1 outside of B_2 and at most 2 on $Q \in \mathcal{B}_2$.)

Repeating this procedure we get the collections $\mathcal{B}_n$ of "stopping cubes" and the decreasing sequence of sets $B_n = \bigcup_{Q \in \mathcal{B}_n} Q$, such that

$$|B_n| \leq \delta^n,$$
(5.21) $$\phi^*_{R_0} \leq 2n - 1 \quad \text{outside of } B_n,$$

and

$$\left| \sum_{R \in \mathcal{D}_r : Q \subsetneqq R \subset R_0} \phi_R(x) \right| \leq 2n \quad \forall Q \in \mathcal{B}_n, \, \forall x \in Q.$$

The last inequality is only needed for the inductive construction.

Given $t > 1$ let n be the largest integer such that $2n - 1 \leq t$,

$$n = \lfloor (t + 1)/2 \rfloor.$$

By (5.21)

$$\phi^*_{R_0} \leq 2n - 1 \leq t \quad \forall x \notin B_n,$$

so

$$\left| \left\{ x \in R_0 : \phi^*_{R_0}(x) > t \right\} \right| \leq |B_n| \leq \delta^n \leq \delta^{(t-1)/2}.$$

This completes the proof for $t > 1$, but for $0 \leq t \leq 1$ the conclusion is trivial. $\quad \square$

As it was shown above in Theorem 5.2, the weak L^1 norm of a dyadic shift $\mathbb{S}$ of complexity r, with scales separated by $r + 1$ levels, can be estimated by $C = 2^{d+2} \|\mathbb{S}\|_2^2 + 5$, so the weak L^1 norm of our dyadic shift $\mathbb{S}$ and all its subshifts $\mathbb{S}_{\mathcal{A}}$, $\mathcal{A} \subset \mathcal{D}_r$, can be estimated by

(5.22) $$B_1 = 2^{d+2} B_2^2 + 5,$$

where

$$B_2 = \sup_{\mathcal{A} \subset \mathcal{D}} \|\mathbb{S}_{\mathcal{A}}\|_{L^2 \to L^2}.$$

Now we need the following lemma, which is essentially [16, Lemma 4.7] with all constants written down; in fact, certain modifications in the argument are needed to avoid introducing exponential dependence on r, which was (implicitly) the case in [16]. Such a modification (with linear dependence on r) was first obtained in [14, Lemma 7.2]; here we even achieve an estimate uniform with respect to r by taking into account the separation of scales of our shift, and the resulting improvement in the estimate of Theorem 5.2.

Let $\mathscr{P} \subset \mathscr{D}_r$ be a collection of cubes. Define the maximal function $f^*_{\mathscr{P}}$ (compare with Definition 5.4) by

$$(5.23) \qquad f^*_{\mathscr{P}}(x) := \sup_{Q \in \mathscr{D}_r : Q \ni x} \left| \sum_{R \in \mathscr{P} : Q \subsetneq R} f_R(x) \right|.$$

For the function $f_{\mathscr{P}(R_0)}$, $R_0 \in \mathscr{G}^*$ defined above in the beginning of Section 5.4 we have $|f_{\mathscr{P}(R_0)}| \le f^*_{\mathscr{P}(R_0)}$, so we will use $f^*_{\mathscr{P}(R_0)}$ to estimate the distribution function of $|f_{\mathscr{P}(R_0)}|$.

Note that for $R \in \mathscr{P}$ we cannot guarantee that its children in $\mathscr{D}_r$ are in $\mathscr{P}$. So while in the above definition the sums are taken over all $R \in \mathscr{P}$, we need to take supremum over $Q \in \mathscr{D}_r$.

Lemma 5.6. *Let B_1 be given by (5.22). Then for any $R \in \mathscr{G}^*$ we have*

$$(5.24) \qquad \left| \left\{ x \in R : f^*_{\mathscr{P}(R)}(x) > 16t \frac{w(R)}{|R|} \right\} \right| \le 2\sqrt{2} \cdot 2^{-t/2B_1} |R|,$$

$$(5.25) \qquad w^{-1} \left(\left\{ x \in R : f^*_{\mathscr{P}(R)}(x) > 20t \frac{w(R)}{|R|} \right\} \right) \le 24 \cdot 2^{-t/2B_1} w^{-1}(R).$$

Proof. Now it is time to perform the last splitting. Namely, let us split the set $\mathscr{P}(R)$ into the sets $\mathscr{P}_\alpha(R)$, $\alpha \in \mathbb{Z}_+$, where the collection $\mathscr{P}_\alpha = \mathscr{P}_\alpha(R)$ consists of all cubes $Q \in \mathscr{P}(R)$ for which

$$(5.26) \qquad 4^{-\alpha} \frac{w(R)}{|R|} < \frac{w(Q)}{|Q|} \le 4^{-\alpha+1} \frac{w(R)}{|R|}.$$

Note that by the construction of stopping moments

$$\frac{w(Q)}{|Q|} \le 4 \frac{w(R)}{|R|},$$

so we do not need $\alpha < 0$. We can estimate

$$f^*_{\mathscr{P}(R)} \le \sum_{\alpha \in \mathbb{Z}_+} f^*_{\mathscr{P}_\alpha(R)}.$$

Now let us estimate the level sets of $f^*_{\mathscr{P}_\alpha(R)}$ using the above Lemma 5.5. For $Q \in \mathscr{P}_\alpha(R)$

$$|f_Q(x)| \le \frac{w(Q)}{|Q|} \le 4^{-\alpha+1} \frac{w(R)}{|R|} \le \frac{w(R)}{|R|} 2^{-2\alpha+3} B_1$$

(recall that $B_1 \ge 1$). To this end, let

$$\phi_Q := \mathbf{1}_{\mathscr{P}_\alpha(R)}(Q) \cdot \frac{2^{2\alpha-3}|R|}{B_1 w(R)} \cdot f_Q,$$

so that ϕ_Q satisfies the first two assumptions of Lemma 5.5.

Recall the notation $\phi^*_{R_1}$ from Definition 5.4. We want to use the weak type estimate for shifts to estimate the size of the set

$$\{ x \in R_1 : \phi^*_{R_1} > 1 \}.$$

Observe that this set is the union of the maximal cubes $M \in \mathcal{D}_r$ such that

$$\left| \sum_{M : M \subsetneq Q \subset R_1} \phi_Q(x) \right| > 1$$

for $x \in M$. Let $\mathcal{M}$ stand for the collection of these maximal cubes, and let

$$\mathcal{N} := \left\{ Q \in \mathcal{D}_r : Q \subset R_1; \nexists M \in \mathcal{M}, Q \subset M \right\}.$$

Then

$$\left\{ x \in R_1 : \phi_{R_1}^* > 1 \right\} = \left\{ x \in R_1 : \left| \sum_{Q \in \mathcal{N}} \phi_Q \right| > 1 \right\},$$

where

$$\sum_{Q \in \mathcal{N}} \phi_Q = \frac{2^{2\alpha-3}|R|}{B_1 w(R)} \sum_{Q \in \mathcal{P}_\alpha(R) \cap \mathcal{N}} f_Q = \frac{2^{2\alpha-3}|R|}{B_1 w(R)} \mathbb{S}_{\mathcal{P}_\alpha(R) \cap \mathcal{N}}(\mathbf{1}_{R_1} w);$$

hence, by Theorem 5.2 and $\|\mathbf{1}_{R_1} w\|_1 = w(R_1)$,

$$\left| \left\{ x \in R_1 : \phi_{R_1}^* > 1 \right\} \right| \le B_1 \frac{2^{2\alpha-3}|R|}{B_1 w(R)} w(R_1).$$

If $R_1 \in \mathcal{P}_\alpha(R)$, then the right side is directly dominated by $2^{2\alpha-3} \cdot 4^{-\alpha+1}|R_1| = \frac{1}{2}|R_1|$. For an arbitrary $R_1 \in \mathcal{D}_r$, observe that $\phi_{R_1}^* = \sum_P \phi_P^*$, where the summation ranges over the maximal $P \in \mathcal{P}_\alpha(R)$ with $P \subset R_1$. Since $\operatorname{supp} \phi_P^* \subset P$, and these cubes are disjoint, it follows that

$$\left| \left\{ x \in R_1 : \phi_{R_1}^* > 1 \right\} \right| = \sum_P \left| \left\{ x \in P : \phi_P^* > 1 \right\} \right| \le \sum_P \frac{1}{2}|P| \le \frac{1}{2}|R_1|.$$

Observing that

$$\phi_R^* = \frac{2^{2\alpha-3}|R|}{B_1 w(R)} \cdot f_{\mathcal{P}_\alpha(R)}^*,$$

Lemma 5.5 implies that

$$\left| \left\{ x \in R : f_{\mathcal{P}_\alpha(R)}^* > t \frac{w(R)}{|R|} 2^{-2\alpha+3} t \right\} \right| = \left| \left\{ x \in R : \phi_R^* > t \right\} \right| \le 2^{-(t-1)/2}|R|.$$

Rescaling t we can rewrite the inequality as

$$(5.27) \qquad \left| \left\{ x \in R : f_{\mathcal{P}_\alpha(R)}^*(x) > 16t \frac{w(R)}{|R|} \right\} \right| \le \sqrt{2} \cdot 2^{-t 4^\alpha / B_1}|R| \quad \forall t > 0.$$

Denote the set above as $E_\alpha(t)$,

$$E_\alpha(t) := \left\{ x \in R : f_{\mathcal{P}_\alpha(R)}^*(x) > 16t \frac{w(R)}{|R|} \right\}.$$

We want to estimate the set where

$$\sum_{\alpha=0}^{\infty} f_{\mathcal{P}_\alpha(R)}^*(x) > T.$$

If this happens for $x \in R$, then either $f^*_{\mathscr{P}_0(R)}(x) > T/2$, or

$$\sum_{\alpha=1}^{\infty} f^*_{\mathscr{P}_\alpha(R)}(x) > T/2.$$

The latter inequality implies that either $f^*_{\mathscr{P}_0(R)}(x) > T/4$ or

$$\sum_{\alpha=1}^{\infty} f^*_{\mathscr{P}_\alpha(R)}(x) > T/4,$$

and so on. Repeating this reasoning with $T = 16w(R)t/|R|$, we can see that

$$\left\{ x \in R : f^*_{\mathscr{P}(R)}(x) > 16t \frac{w(R)}{|R|} \right\} \subset \bigcup_{\alpha \geq 0} E_\alpha(2^{-\alpha-1}t).$$

So using (5.27) we get

$$|R|^{-1} \left| \left\{ x \in R : f^*_{\mathscr{P}(R)}(x) > 16t \frac{w(R)}{|R|} \right\} \right| \leq \sqrt{2} \sum_{\alpha=0}^{\infty} 2^{-t \cdot 2^{\alpha-1}/B_1}$$

$$\leq \sqrt{2} \sum_{\alpha=0}^{\infty} 2^{-t/2B_1 - \alpha} \quad \text{if } t \geq 2B_1$$

$$\leq 2\sqrt{2} \cdot 2^{-t/2B_1},$$

which proves (5.24). We have proved (5.24) for $t \geq 2B_1$, but for $t < 2B_1$ this estimate is trivial, because the right side is greater than $|R|$. Thus, (5.24) holds for all $t > 0$.

To prove (5.25), let us first recall that all our cubes are in $\mathcal{Q} = \mathcal{Q}_k$, so (5.10) holds for all of them. If, in addition $Q \in \mathscr{P}_\alpha(R)$, then (5.26) (the definition of $\mathscr{P}_\alpha(R)$) is satisfied, and combining these two estimates we get

$$(5.28) \qquad 2^k 4^{\alpha-1} \frac{|R|}{w(R)} \leq \frac{w^{-1}(Q)}{|Q|} \leq 2^{k+1} 4^\alpha \frac{|R|}{w(R)} \quad \forall Q \in \mathscr{P}_\alpha(R).$$

So $w^{-1}(Q)$ can be estimated via $|Q|$. Thus we will use the known estimates of the Lebesgue measure of level sets to get the estimates of the w^{-1} measure.

Let us consider the set where

$$f^*_{\mathscr{P}_\alpha(R)}(x) > 20t \frac{w(R)}{|R|}.$$

This set is a disjoint union of cubes $Q' \in \mathcal{D}_r$, which are the first (maximal) cubes Q for which the sum in (5.23) defining $f^*_{\mathscr{P}_\alpha(R)}$ exceeds $20t \cdot w(R)/|R|$. Unfortunately the cubes Q' are not necessarily in $\mathscr{P}_\alpha(R)$, so we cannot use (5.28) for them. But their parents are in $\mathscr{P}_\alpha(R)$ (because the summation is over $\mathscr{P}_\alpha(R)$).

So, let $\mathscr{E}_\alpha(t)$ be the collection of such parents, and let

$$\tilde{E}_\alpha(t) := \bigcup_{Q \in \mathscr{E}_\alpha(t)} Q.$$

Note that to get $\tilde{E}_\alpha(t)$ it is sufficient to take the union of the maximal cubes $Q \in \mathcal{E}_\alpha(t)$, so the set $\tilde{E}_\alpha(t)$ is a disjoint union of cubes $Q \subset \mathcal{P}_\alpha(R)$. Since for $Q \in \mathcal{P}_\alpha(R)$

$$|f_Q(x)| \le \frac{w(Q)}{|Q|} \le 4^{-\alpha+1} \frac{w(R)}{|R|},$$

we can conclude that for all $Q \in \mathcal{E}_\alpha(t)$ and all $t \ge 4^{-\alpha}$

$$\left| \sum_{R' \in \mathcal{P}_\alpha(R) : Q \subsetneq R'} f_{R'}(x) \right| \ge 20t \frac{w(R)}{|R|} - 4 \cdot 4^{-\alpha} \frac{w(R)}{|R|} \ge 16t \frac{w(R)}{|R|} \quad \forall x \in Q$$

(because the corresponding sum for one of the children Q' of Q exceeds $20t \cdot w(R)/|R|$ on Q', and the difference between the two sums is f_Q; we also use that the sum in the left hand side is constant on Q).

So $f^*_{\mathcal{P}_\alpha(R)}(x) > 16t \cdot w(R)/|R|$ on Q, and we conclude that for $t \ge 4^{-\alpha}$ the inclusion $\tilde{E}_\alpha(t) \subset E_\alpha(t)$ holds. Using the estimate (5.27) for $|E_\alpha(t)|$ (and replacing $\sqrt{2}$ by 2 there) we get that for $t \ge 4^{-\alpha}$

$$(5.29) \qquad |\tilde{E}_\alpha(t)| \le 2 \cdot 2^{-t4^\alpha/B_1} |R|.$$

Note that for $t < 4^{-\alpha}$ the above estimate is trivial, so it holds for all $t > 0$.

Since by (5.28) for all $Q \in \mathcal{P}_\alpha(R)$

$$w^{-1}(Q) \le 2^{k+1} 4^\alpha \frac{|R|}{w(R)} |Q|$$

summing over maximal cubes in $\mathcal{E}_\alpha(t)$ we get

$$\begin{aligned} w^{-1}(\tilde{E}_\alpha(t)) &\le 2^{k+1} 4^\alpha \frac{|R|}{w(R)} |\tilde{E}_\alpha(t)| \\ &\le 2^{k+1} 4^\alpha \frac{|R|}{w(R)} 2 \cdot 2^{-t4^\alpha/B_1} |R| \quad \text{by (5.29)} \\ (5.30) \qquad &\le 4^\alpha \cdot 2^2 \cdot 2^{-t4^\alpha/B_1} w^{-1}(R) \quad \text{by (5.10).} \end{aligned}$$

Now we want to estimate $w^{-1}(\tilde{E}(t))$, where

$$\tilde{E}(t) := \left\{ x \in R : f^*_{\mathcal{P}}(x) > 20t \frac{w(R)}{|R|} \right\}.$$

Let $T := 20t \cdot w(R)/R$. If for $x \in R$

$$\sum_{\alpha=0}^{\infty} f^*_{\mathcal{P}_\alpha(R)}(x) > T,$$

then either $f^*_{\mathcal{P}_0(R)}(x) > T/2$ (in which case $x \in \tilde{E}_0(t/2)$) or

$$\sum_{\alpha=1}^{\infty} f^*_{\mathcal{P}_\alpha(R)}(x) > T/2.$$

If the latter inequality holds, then either $f^*_{\mathscr{P}_1(R)}(x) > T/4$, so $x \in \tilde{E}_0(t/4)$, or

$$\sum_{\alpha=2}^{\infty} f^*_{\mathscr{P}_\alpha(R)}(x) > T/4.$$

Repeating this reasoning we get that

$$\tilde{E}(t) \subset \bigcup_{\alpha \geq 0} \tilde{E}_\alpha(t 2^{-\alpha-1}).$$

So we conclude

$$w^{-1}(\tilde{E}(t)) \leq \sum_{\alpha=0}^{\infty} w^{-1}(\tilde{E}_\alpha(t 2^{-\alpha-1}))$$

$$\leq 4w^{-1}(R) \sum_{\alpha=0}^{\infty} 4^\alpha 2^{-t 2^{\alpha-1}/B_1} \quad \text{by (5.30)}$$

$$\leq 4w^{-1}(R) \cdot 6 \cdot 2^{-t/2B_1} \quad \text{if } t \geq 2B_1.$$

To prove the last inequality we need to estimate for $t \geq 2B_1$ the sum

$$\sum_{\alpha=0}^{\infty} 2^{2\alpha - t 2^\alpha/2B_1}.$$

Since $2^\alpha \geq 3\alpha + 2$ for $\alpha \geq 4$, we can estimate for $\alpha \geq 4$ and $t \geq 2B_1$

$$2\alpha - t 2^\alpha/2B_1 \leq 2\alpha - t \cdot (3\alpha + 2)/2B_1$$
$$= (2\alpha - 2\alpha t/2B_1) - \alpha t/2B_1 - 2t/2B_1$$
$$\leq 0 - \alpha - t/2B_1,$$

so

$$\sum_{\alpha=4}^{\infty} 2^{2\alpha - t 2^\alpha/2B_1} \leq 2^{-t/2B_1} \sum_{\alpha=4}^{\infty} 2^{-\alpha} < \frac{1}{2} \cdot 2^{-t/2B_1}.$$

For $\alpha = 0, 1, 2, 3$ we can estimate

$$2^{2\alpha - t 2^\alpha/2B_1} \leq c_\alpha 2^{-t/2B_1}, \quad \text{where } c_0 = 1, \ c_1 = c_2 = 2, \ c_3 = \frac{1}{2},$$

so adding everything we get that

$$w^{-1}(\tilde{E}(t)) \leq 24 \cdot 2^{-t/2B} w^{-1}(R).$$

We proved that estimate for $t \geq 2B_1$, but for $t < 2B_1$ the estimate is trivial because the right side is bigger than $w^{-1}(R)$. So the estimate holds for all $t > 0$. □

5.5. Conclusion of the proof.

Lemma 5.7. *For any* $R \in \mathcal{G}^*$

$$(5.31) \qquad \|f_{\mathcal{P}(R)}\|_{L^2} \leq C_1 B_1 \frac{w(R)}{|R|} |R|^{1/2},$$

$$(5.32) \qquad \|f_{\mathcal{P}(R)}\|_{L^2(w^{-1})} \leq C_1 B_1 \frac{w(R)}{|R|} \sqrt{w^{-1}(R)},$$

where C_1 and C_2 are absolute constants and B_1 is given by (5.22).

This lemma is proved by using the distributional inequalities from Lemma 5.6 and computing the norms using distribution functions. That will give the desired estimates for the norms of the maximal function $f^*_{\mathcal{P}(R)}$, and since $|f_{\mathcal{P}(R)}| \leq f^*_{\mathcal{P}(R)}$, we get the conclusion of the lemma. We leave the details as a trivial exercise for the reader.

Recall that to prove the main result we need to prove estimate (5.11) for all cubes $Q_0 \in \mathcal{Q} = \mathcal{Q}_k$. For a cube $Q \in \mathcal{Q}$, let $\mathcal{Q}(Q) := \{Q' \in \mathcal{Q} : Q' \subset Q\}$. We want to estimate $\|f_{\mathcal{Q}(Q_0)}\|_{L^2(w^{-1})}$, $Q_0 \in \mathcal{Q}$, where

$$f_{\mathcal{Q}(Q_0)} := \sum_{Q \in \mathcal{Q}(Q_0)} f_Q.$$

Since (see (5.19))

$$f_{\mathcal{Q}(Q_0)} = \sum_{Q \in \mathcal{G}^*(Q_0)} f_{\mathcal{P}(Q)},$$

we can write

$$\|f_{\mathcal{Q}(Q_0)}\|^2_{L^2(w^{-1})} \leq \sum_{R \in \mathcal{G}^*(Q_0)} \|f_{\mathcal{P}(R)}\|^2_{L^2(w^{-1})} + 2 \sum_{R,Q \in \mathcal{G}^*(Q_0):Q \subsetneqq R} |\langle f_{\mathcal{P}(R)}, f_{\mathcal{P}(Q)} \rangle_{w^{-1}}|$$

$$= S_1 + S_2.$$

The first sum is easy to estimate. By (5.32)

$$\|f_{\mathcal{P}(R)}\|^2_{L^2(w^{-1})} \leq [C_1 B_1]^2 \frac{w(R)^2}{|R|^2} w^{-1}(R),$$

$$\leq [C_1 B_1]^2 2^{k+1} w(R) \quad \text{because } R \in \mathcal{Q} = \mathcal{Q}_k.$$

Summing over all $R \in \mathcal{G}^* = \mathcal{G}^*(Q_0)$ we get, using (5.18),

$$S_1 \leq 2[C_1 B_1]^2 2^k \sum_{R \in \mathcal{G}^*(G_0)} w(R) \leq C B_1^2 2^k [w]_{A_2} w(Q_0),$$

where C is an absolute constant.

Let us now estimate S_2. Let $Q, R \in \mathcal{G}^*$, $Q \subsetneqq R$. Then $f_{\mathcal{P}(R)}(x)$ is constant on Q; let us use the symbol $f_{\mathcal{P}(R)}(Q)$ to denote this constant. We then can estimate

$$|\langle f_{\mathcal{P}(R)}, f_{\mathcal{P}(Q)} \rangle_{w^{-1}}| \leq |f_{\mathcal{P}(R)}(Q)| \cdot (w^{-1}(Q))^{1/2} \|f_{\mathcal{P}(Q)}\|_{L^2(w^{-1})} \quad \text{by Cauchy–Schwarz}$$

$$\leq C_1 B_1 |f_{\mathcal{P}(R)}(Q)| \frac{w^{-1}(Q) w(Q)}{|Q|} \quad \text{by (5.32)}$$

$$(5.33) \qquad \leq C_1 B_1 |f_{\mathcal{P}(R)}(Q)| 2^{k+1} \cdot |Q| \quad \text{because } Q \in \mathcal{Q}_k.$$

Using this estimate we can write

$$
\begin{aligned}
S_2(R) &:= \sum_{Q \in \mathscr{G}^* : Q \subsetneqq R} |\langle f_{\mathscr{P}(R)}, f_{\mathscr{P}(Q)} \rangle_{w^{-1}}| \\
&\leq 2^{k+1} C_1 B_1 \sum_{Q \in \mathscr{G}^* : Q \subsetneqq R} |f_{\mathscr{P}(R)}(Q)| \cdot |Q| \quad \text{by (5.33)} \\
&= 2^{k+1} C_1 B_1 \int_R |f_{\mathscr{P}(R)}| \sum_{Q \in \mathscr{G}^* : Q \subsetneqq R} \mathbf{1}_Q \, dx \\
&\leq 2^{k+1} C_1 B_1 \| f_{\mathscr{P}(R)} \|_2 \cdot \Big\| \sum_{Q \in \mathscr{G}^* : Q \subsetneqq R} \mathbf{1}_Q \Big\|_2 \quad \text{by Cauchy–Schwarz} \\
&\leq 2^{k+2} [C_1 B_1]^2 w(R) \quad \text{by (5.31) and (5.17).}
\end{aligned}
$$

Therefore, using (5.18), we have

$$
S_2 \leq 2^{k+1} [C_1 B_1]^2 \leq 2^{k+1} [C_1 B_1]^2 \sum_{R \in \mathscr{G}^*(Q_0)} w(R) \leq C(B_1)^2 2^k [w]_{A_2} w(Q_0)
$$

with some absolute constant C. But that is exactly the estimate (5.11), so Theorem 5.1 is proved. $\qquad \square$

References

[1] *K. Astala*, *T. Iwaniec* and *E. Saksman*, Betrami operators in the plane, Duke Math J. **107** (2001), 27–56.

[2] *O. Beznosova*, Linear bound for the dyadic paraproduct on weighted Lebesgue space $L^2(w)$, J. Funct. Analysis **255** (2008), 994–1007.

[3] *S. M. Buckley*, Estimates for operator norms on weighted spaces and reverse Jensen inequalities, Trans. Amer. Math. Soc. **340** (1993), 53–272.

[4] *D. Cruz-Uribe*, *J. Martell* and *C. Perez*, Sharp weighted estimates for classical operators, Adv. Math. **229** (2012), 408–441.

[5] *D. Cruz-Uribe*, *J. Martell* and *C. Perez*, Sharp weighted estimates for approximating dyadic operators, Electron. Res. Announc. Math. Sci. **17** (2010), 12–19.

[6] *G. David*, Analytic capacity, Calderón–Zygmund operators, and rectifiability, Publ. Mat. **43** (1999), 3–25.

[7] *O. Dragičević* and *A. Volberg*, Sharp estimate of the Ahlfors-Beurling operator via averaging martingale transforms, Michigan Math. J. **51** (2003), 415–435.

[8] *P. L. Duren*, Theory of H^p spaces, Pure Appl. Math. **38**, Academic Press, New York 1970.

[9] *R. Fefferman* and *J. Pipher*, Multiparameter operators and sharp weighted inequalities, Amer. J. Math. **119** (1997), 337–369.

[10] *T. Figiel*, Singular integral operators: a martingale approach, in: Geometry of Banach spaces (Strobl 1989), London Math. Soc. Lecture Note Ser. **158**, Cambridge University Press, Cambridge (1990), 95–110.

[11] *L. Grafakos*, Classical Fourier analysis, 2nd ed., Grad. Texts in Math. **249**, Springer, New York 2008.

[12] *R. Hunt*, *B. Muckenhoupt* and *R. Wheeden*, Weighted norm inequalities for the conjugate function and the Hilbert transform, Trans. Amer. Math. Soc. **176** (1973), 227–251.

[13] *T. Hytönen*, The vector-valued nonhomogeneous Tb theorem, preprint 2008, `http://arxiv.org/abs/0809.3097`.

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

[15] *T. Hytönen*, *M. Lacey*, *M. C. Reguera*, *E. Sawyer*, *I. Uriarte-Tuero* and *A. Vagharshakyan*, Weak and Strong type A_p Estimates for Calderón–Zygmund Operators, preprint 2010, `http://arxiv.org/abs/1006.2530`.

[16] *M. Lacey*, *S. Petermichl* and *M. Reguera*, Sharp A_2 inequality for Haar shift operators, Math. Ann. **348** (2010), 127–141.

[17] *A. Lerner*, A pointwise estimate for the local sharp maximal function with applications to singular integrals, Bull. London Math. Soc. **42** (2010), 843–856.

[18] *A. Lerner*, *S. Ombrosi* and *C. Pérez*, A1 bounds for Calderón–Zygmund operators related to a problem of Muckenhoupt and Wheeden, Math. Res. Lett. **16** (2009), 149–156.

[19] *F. L. Nazarov* and *S. R. Treil*, The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis, Algebra i Analiz **8** (1996), 32–162.

[20] *F. Nazarov*, *S. Treil* and *A. Volberg*, Cauchy integral and Calderón–Zygmund operators on nonhomogeneous spaces, Int. Math. Res. Not. (1997), no. 5, 103–726.

[21] *F. Nazarov*, *S. Treil* and *A. Volberg*, Weak type estimates and Cotlar inequalities for Calderón–Zygmund operators on nonhomogeneous spaces, Int. Math. Res. Not. (1998), no. 9, 463–487.

[22] *F. Nazarov*, *S. Treil* and *A. Volberg*, The Bellman functions and two-weight inequalities for Haar multipliers, J. Amer. Math. Soc. **12** (1999), 909–928.

[23] *F. Nazarov*, *S. Treil* and *A. Volberg*, Bellman function in stochastic control and harmonic analysis, in: Systems, approximation, singular integral operators, and related topics (Bordeaux 2000), Oper. Theory Adv. Appl. **129**, Birkhäuser, Basel (2001), 393–423

[24] *F. Nazarov*, *S. Treil* and *A. Volberg*, Accretive system Tb theorems on nonhomogeneous spaces, Duke Math. J. **113** (2002), 259–312.

[25] *F. Nazarov*, *S. Treil* and *A. Volberg*, Nonhomogeneous Tb theorem which proves Vitushkin's conjecture, preprint no. 519, CRM, Barcelona (2002), 1–84.

[26] *F. Nazarov*, *S. Treil* and *A. Volberg*, The Tb theorem on non-homogeneous spaces, Acta Math. **190** (2003), 151–239.

[27] *F. Nazarov*, *S. Treil* and *A. Volberg*, Two weight $T1$ theorem for the Hilbert transform: the case of doubling measures, preprint 2004.

[28] *F. Nazarov*, *S. Treil* and *A. Volberg*, Two weight inequalities for individual Haar multipliers and other well localized operators, Math. Res. Lett. **15** (2008), 583–597.

[29] *F. Nazarov*, *S. Treil* and *A. Volberg*, Two weight estimate for the Hilbert transform and corona decomposition for non-doubling measures, preprint 2010, `http://arxiv.org/abs/1003.1596`.

[30] *C. Pérez*, *S. Treil* and *A. Volberg*, On A_2 conjecture and corona decomposition of weights, preprint 2010, `http://arxiv.org/abs/1006.2630`.

[31] *S. Petermichl*, Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol, C. R. Acad. Sci. Paris, Sér. I Math. **330** (2000), 455–460.

[32] *S. Petermichl*, The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical A_p characteristic, Amer. J. Math. **129** (2007), 1355–1375.

[33] *S. Petermichl*, The sharp weighted bound for the Riesz transforms, Proc. Amer. Math. Soc. **136** (2008), 1237–1249.

[34] *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.

[35] *E. Sawyer*, A characterization of a two-weight norm inequality for maximal operators, Studia Math. **75** (1982), 1–11.

[36] *E. Sawyer*, Two weight norm inequalities for certain maximal and integral operators, in: Harmonic analysis, Lecture Notes in Math. **908**, Springer (1982), 102–127.

[37] *X. Tolsa*, L^2 boundedness for the Cauchy linear operator for continuous measures, Duke Math. J. **98** (1999), 269–304.

[38] *S. Treil* and *A. Volberg*, Wavelets and the angle between past and future, J. Funct. Anal. **143** (1997), 269–308.

[39] *A. Vagharshakyan*, Recovering singular integral kernels from Haar shifts, Proc. Amer. Math. Soc. **138** (2010), 4303–4309.

[40] *A. Volberg*, Matrix A_p weights via S-function, J. Amer. Math. Soc. **10** (1997), 445–466.

[41] *A. Volberg*, Calderón–Zygmund capacities and operators on nonhomogeneous spaces, CBMS Reg. Conf. Ser. Math. **100**, American Mathematical Society, Providence 2003.

[42] *J. Wittwer*, A sharp estimate on the norm of the martingale transform, Math. Res. Lett. **7** (2000), 1–12.

[43] *Q. X. Yang*, Fast algorithms for Calderón–Zygmund singular integral operators, Appl. Comput. Harmon. Anal. **3** (1996), 120–126.

Tuomas Hytönen, Department of Mathematics and Statistics, P.O.B. 68, FI-00014 University of Helsinki, Finland
e-mail: tuomas.hytonen@helsinki.fi

Carlos Pérez, Department of Mathematics, Universidad de Sevilla, Sevilla, Spain
e-mail: carlosperez@us.es

Sergei Treil, Department of Mathematics, Brown University, Box 1917, Providence, RI 02912, USA
e-mail: treil@math.brown.edu

Alexander Volberg, Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA
e-mail: volberg@math.msu.edu