BI-PARAMETER CARLESON EMBEDDINGS
WITH PRODUCT WEIGHTS

NICOLA ARCOZZI, PAVEL MOZOLYAKO, GEORGIOS PSAROMILIGKOS,
ALEXANDER VOLBERG, AND PAVEL ZORIN-KRANICH

Abstract. We show that a (weighted) Carleson embedding from the bi-torus to
the bi-disc is equivalent to a simple “box” condition, for product weights on the bi-
disc and arbitrary weights on the bi-torus. This gives a new simple necessary and
sufficient condition for the embedding of the whole scale of weighted Dirichlet
spaces of holomorphic functions on the bi-disc. This scale includes the usual
Dirichlet space on the bi-disc. Our result is in contrast to the classical situation
on the bi-disc considered by Chang and Fefferman, when a counterexample due
to Carleson shows that the “box” condition does not suffice for the embedding to
hold. Our result can be viewed as a new and unexpected combinatorial property
of all positive finite planar measures.

1. Introduction

1.1. Background. Lennart Carleson showed in [Car74] that the natural generaliza-
tion, using a “box” condition, from the one parameter case (disc) to the bi-parameter
case (bi-disc) of his embedding theorem does not work. Sun-Yang A. Chang in
[Cha79] found the necessary and sufficient condition for the validity of the Carleson
embedding for bi-harmonic extensions into the bi-disc.

The discrete versions of these results can be motivated by considering a bi-parameter
dyadic paraproduct. For a dyadic rectangle \( R = I \times J \subseteq [0,1]^2 \) denote by \( h_R(x,y) = h_I(x)h_J(y) \) an associated \( L^2 \) normalized Haar function. The simplest example of a
bi-parameter dyadic paraproduct is the operator

\[
\Pi_b \varphi := \sum_R \langle \varphi \rangle_R \beta_R h_R.
\]

The paraproduct \( \Pi_b \) is a bounded operator on \( L^2 \) with respect to the Lebesgue
measure \( m \) on \([0,1]^2\) if and only if we have

\[
\sum_R \langle \varphi \rangle^2_R \beta^2_R \leq C \int \varphi^2 dm,
\]

where \( \beta_R := (b, h_R) \) are Haar coefficients of the function \( b \). In analogy to the one-
parameter Carleson embedding one could ask whether (1.1) is equivalent to the “box”
condition

\[
\sum_{R \subseteq R_0} \beta^2_R \leq C' m(R_0)
\]

for every dyadic rectangle \( R_0 \subseteq [0,1]^2 \). A counterexample showing that (1.2) does
not imply (1.1) was constructed by Carleson [Car74; Tao].

It was observed by Chang [Cha79] (in a continuous setting) that (1.1) is equivalent
to the bi-parameter Carleson (or Carleson–Chang) condition

\[
\sum_{R \subseteq \Omega} \beta^2_R \leq C' m(\Omega),
\]

2010 Mathematics Subject Classification. 42B20, 42B35, 47A30.

NA is partially supported by the grants INDAM-GNAMPA 2017 “Operatori e disuguaglianze
integrali in spazi con simmetrie” and PRIN 2018 “Varietà reali e complesse: geometria, topologia e
analisi armonica”.

PM is supported by the Russian Science Foundation grant 17-11-01064.

AV is partially supported by the NSF grant DMS-160065 and DMS 1900268 and by Alexander
von Humboldt foundation.

PZ was partially supported by the Hausdorff Center for Mathematics (DFG EXC 2047).
where the constant $C'$ is uniform for all subsets $\Omega \subseteq [0,1]^2$ that are finite unions of dyadic rectangles. This necessary and sufficient condition was later used by Chang and Fefferman [CF80] to characterize the dual of the Hardy space on the bi-disc $H^1(\mathbb{D}^2)$. In Appendix A we recall the proof of the equivalence between (1.1) and (1.3) in the more general setting of bi-trees.

The present article treats a certain two weight problem about bi-parameter para-product operators. Singular bi-parameter operators enjoyed and continue to enjoy much attention, see [Fef81; Fef86; Fef88], [Pip86], [Jou88], [BP17]. They are notoriously difficult. Two weight problems for singular integrals were studied in a series of papers by Nazarov, Treil, and Volberg on dyadic singular operators and in a series of papers by Lacey, Shen, Sawyer, and Uriarte-Tuero on the Hilbert transform, see [NTV07], [LSSU14], and the references therein. Another example is a very recent paper by Iosevich, Krause, Sawyer, Taylor, and Uriarte-Tuero [Ios+19] on the two weight problem for the spherical maximal operator motivated by Falconer’s distance set problem.

1.2. Summary. In the works mentioned in Section 1.1 the “underlying” measure was the Lebesgue measure on bi-torus, and the “embedding” measure on the bi-disc was a priori arbitrary. In this article we switch the constraints on two measures involved: the “underlying” measure on the bi-torus is arbitrary, while the “embedding” measure on the bi-disc has a product structure. This problem appeared in trying to understand the embedding of Dirichlet space on bi-disc [AMPS18], this connection will be discussed in more detail in Section 1.5. For uniform “embedding” measures several necessary and sufficient conditions of Carleson–Chang type for a Carleson embedding of Dirichlet space on the bi-disc were found in [AMPS18; AHMV18b].

In this article we show the unexpected fact that for “embedding” measures with product structure a condition of the “box” type (1.2) turns out to characterize the Carleson embedding (1.1). This is a striking difference to the case considered by Chang and Fefferman, where Carleson’s counterexample explicitly forbids for such an effect to happen. We also construct several counterexamples in the absence of product structure.

1.3. Terminology and notation. We begin with order-theoretic conventions.

**Definition 1.4.** A finite tree $T$ is a finite partially ordered set such that for every $\omega \in T$ the set $\{\alpha \in T : \alpha \geq \omega\}$ is totally ordered (we allow trees to have several maximal elements).

A bitree $T^2$ is a cartesian product of 2 finite trees with the product order.

A subset $U$ (resp. $D$) of a partially ordered set $T$ is called an up-set (resp. down-set) if for every $\alpha \in U$ and $\beta \in T$ with $\alpha \leq \beta$ (resp. $\beta \leq \alpha$) we also have $\beta \in U$ (resp. $\beta \in D$).

The Hardy operator on the bitree $T^2$ is defined by

$$I\phi(\gamma) := \sum_{\gamma' \geq \gamma} \phi(\gamma')$$

for any $\phi : T^2 \to \mathbb{R}$. (1.5)

In the one-parameter case $T^1 = T$ we denote it by $I$. The adjoint $I^*$ of the Hardy operator $I$ is given by the formula

$$I^*\psi(\gamma) = \sum_{\gamma' \leq \gamma} \psi(\gamma').$$

(1.6)

**Definition 1.7.** Let $\mu, w$ be positive functions on $T^2$. The box constant is the smallest number $[w, \mu]_{Box}$ such that

$$E_\beta[\mu] := \sum_{\alpha \leq \beta} w(\alpha)(I^* \mu(\alpha))^2 \leq [w, \mu]_{Box} \sum_{\alpha \leq \beta} \mu(\alpha), \quad \forall \beta \in T^2.$$

(1.8)

The Carleson constant is the smallest number $[w, \mu]_C$ such that

$$\sum_{\alpha \in D} w(\alpha)(I^* \mu(\alpha))^2 \leq [w, \mu]_C \mu(D), \quad \forall D \subset T^2 \text{ down-set}.$$

(1.9)
The hereditary Carleson constant (or restricted energy condition constant or REC constant) is the smallest constant \([w, \mu]_{HC}\) such that

\[
\mathcal{E}[\mu 1_E] = \sum_{\alpha \in T^2} w(\alpha)(\mathbb{I}^*(\mu 1_E)(\alpha))^2 \leq [w, \mu]_{HC} \mu(E), \quad \forall E \subset T^2.
\] (1.10)

The Carleson embedding constant is the smallest constant \([w, \mu]_{CE}\) such that the adjoint embedding

\[
\sum_{\alpha \in T^2} w(\alpha)[\mathbb{I}^*(\psi)(\alpha)]^2 \leq [w, \mu]_{CE} \sum_{\omega \in T^2} |\psi(\omega)|^2 \mu(\omega)
\] (1.11)

holds for all functions \(\psi\) on \(T^2\).

In order to make a connection to the objects in Section 1.1 consider the bi-tree \(T^2\) that consists of the dyadic rectangles in \([0, 1]^2\) with side lengths between 1 and \(2^{-N}\), where \(N\) is a large finite integer. Let

\[
\mu(\omega) = \begin{cases} m(\omega) & \text{if } \omega \text{ is a } (2^{-N} \times 2^{-N})\text{-square}, \\ 0 & \text{otherwise}. \end{cases}
\] (1.12)

Suppose also \(\beta_R^2 = m(R)^2 w_R\). Then (1.3) is equivalent to (1.9) with constant independent of \(N\), while (1.1) is equivalent to (1.11) with constant independent of \(N\). The box condition (1.2) is equivalent to (1.8), again with constant independent of \(N\). All these just mentioned equivalences hold for Lebesgue measure \(\mu\) as in (1.12). In the present article we will describe all embedding measures \(\mu\) such that (1.11) is valid if \(w\) has product structure.

For positive numbers \(A, B\) we write \(A \lesssim B\) if \(A \leq CB\) with an absolute constant \(C\), that in particular does not depend on the tree or bi-tree or the weights \(w, \mu\).

1.4. Main result. The inequalities

\[
[w, \mu]_{Box} \leq [w, \mu]_{C} \leq [w, \mu]_{HC} \leq [w, \mu]_{CE}
\] (1.13)

are obvious. It turns out that for product measures \(w\) there are also converse inequalities.

**Theorem 1.14.** Let \(\mu, w\) be positive measures on \(T^2\). Assume that \(w\) is of the product form

\[
w(\alpha) = w(\alpha_x, \alpha_y) = w_x(\alpha_x)w_y(\alpha_y).
\] (1.15)

Then the reverses of the inequalities in (1.13) also hold:

\[
[w, \mu]_{CE} \lesssim [w, \mu]_{HC} \lesssim [w, \mu]_{C} \lesssim [w, \mu]_{Box}.
\]

**Remark 1.16.** In Carleson’s counterexample \(\mu = m\), but the coefficients \(\beta_R\) are not of the form \(m(R)^2 w_R\) with \(w_R\) having product structure. The leading particular case of Theorem 1.14 is \(w \equiv 1\), which corresponds to the description of all embedding measures for the unweighted Dirichlet space on the bi-disc. Already in this case most combinatorial features of the proof for general weights appear.

The index 2 in the notation \(T^2\) stands for the fact that the tree has two parameters. In particular we allow \(T^2 = T \times T'\), where \(T\) and \(T'\) are different simple trees. When the second tree \(T'\) consists of one point, Theorem 1.14 recovers the two weight Carleson embedding in [NTV99]. Alternative proofs of that result can be found in [AHMV18a] and in Appendix A. In the case \(w \equiv 1\) equivalence of the Carleson (1.9) and the Carleson embedding (1.11) conditions on (dyadic) bi-trees was proved in [AMPS18], and in addition an equivalent capacitary condition was found in that article. An alternative proof of this equivalence in the case \(w \equiv 1\) was given in [AHMV18b].

But neither [AMPS18] nor [AHMV18b] proved the equivalence of embedding and box condition (1.8) even for \(w \equiv 1\). This is partially because in view of Carleson’s counterexample it was hard to imagine the equivalence of box condition and embedding.
We would also like to mention a result of Sawyer [Saw85], who obtained a complete description of the two-weight embedding for a special (hooked) weight \( w \). We will discuss Sawyer’s result in more detail in Section 7.4.

The main contribution of the present article is the unexpected equivalence of the box condition (1.8) with the other conditions listed above, see Theorem 7.7. The extension of the previous results to product weights \( w \) is also new. This extension follows the argument in [AHMV18b], which we streamline considerably in Theorems 5.6 and 6.3.

The second part of this article contains counterexamples with general \( w \) to all equivalences proved in the case of product \( w \). One such example is of course due to Carleson [Car74], but the weight \( w = \{ w_R \} \) is not very simple in that example. We will give examples in which for each dyadic rectangle \( R \) the weight \( w_R \) is either 0 or 1. This corresponds to embeddings on subgraphs of bi-trees.

The question of embedding to weighted \( k \)-trees, \( k \geq 3 \), their subgraphs, and more general graphs with cycles remains wide open. The difference between bi-parameter singular integrals and \( k \)-parameter, \( k \geq 3 \), is not new for multi-parameter theory. See, for example, [Jou88], where J.-L. Journé constructed a 3-parameter singular Calderón–Zygmund operator \( S \) mapping (as it should) \( L^2(\mathbb{R}^3) \) to itself, but not mapping \( L^\infty \) to product (Chang–Fefferman) \( BMO \) even though \( S \) acts in the expected way on the rectangular atoms introduced in [Fef86]. We are grateful to Jill Pipher who indicated to us the paper [Jou88]. The counterexample in [Jou88] is related to the counterexample of Carleson [Car74] showing that rectangular Carleson measures are not, on the bi-disc, a good substitute for classical Carleson measures.

1.5. Applications to complex analysis on the bi-disc. As already mentioned, the results of Carleson, Chang, and Fefferman correspond to the situation where the measure \( \mu \) is fixed to be the Lebesgue measure, and the weight \( w \) is arbitrary. Here we impose a condition (product structure) on the weight (for example we can consider \( w \equiv 1 \)), and vary the measure. One reason to consider such a problem stems from the problems about Carleson measures for a certain scale of Hilbert spaces of analytic functions on the bi-disc. Namely, given \( s = (s_1, s_2) \in \mathbb{R}^2 \) consider the space \( \mathcal{H}_s \) of analytic functions \( f \) on the unit bi-disc for which the norm

\[
\sum_{n_1, n_2 \geq 0} (n_1 + 1)^{s_1} (n_2 + 1)^{s_2} |\hat{f}(n_1, n_2)|^2 =: \|f\|_s^2
\]

is finite. One then can ask for which measures \( \mu \) embedding \( \mathcal{H}_s \to L^2(\mathbb{D}^2, d\mu) \) is bounded. The Hardy space on the bi-disc (i.e., \( \mathcal{H}_{(0,0)} \)) corresponds to the Carleson–Chang–Fefferman case. It turns out that the special form of the second case, arbitrary \( \mu \) and \( w \equiv 1 \), describes the embedding of the Dirichlet space \( \mathcal{H}_{(1,1)} \) on the bi-disc (see [AMPS18]). Moreover, adjusting the pair \( (\mu, w) \), one can get a boundedness criterion for the embedding of any space \( \mathcal{H}_s \) on the Hardy–Dirichlet scale on the bi-disc.

1.6. Combinatorial property of all planar positive measures. The equivalence

\[
[1, \mu]_{\text{Box}} \asymp [1, \mu]_{\text{CE}}
\]

still amazes us. More explicitly, the essence of this equivalence is that for any positive measure \( \mu \) on a square \( Q_0 \) the property that

\[
\sum_{R \subseteq R_0} \mu(R)^2 \leq \mu(R_0)
\]

for all dyadic rectangles \( R_0 \)

implies

\[
\sum_{R \subseteq \bigcup_i R_i} \mu(R)^2 \leq \mu(\bigcup_i R_i)
\]

for all unions dyadic rectangles \( R_i \).

This contradicts—at first glance—the examples of Carleson type [Car74; Tao], and goes against the general feeling that huge overlap among the dyadic rectangle must prevent this equivalence from happening. But it happens, see below.
2. Small energy majorization on an ordinary tree

Definition 2.1. Given a finite tree $T$, the set of children of a vertex $\beta \in T$ consists of the maximal elements of $T$ that are strictly smaller than $\beta$:

$$\text{ch} \beta := \max \{ \beta' \in T \mid \beta' < \beta \}$$

A function $g : T \to [0, \infty)$ is called superadditive if for every $\beta \in T$ we have

$$g(\beta) \geq \sum_{\beta' \in \text{ch}(\beta)} g(\beta').$$

Lemma 2.2. Let $T'$ be a finite tree and $g, h : T' \to [0, \infty)$. Assume that $g$ is superadditive and $Ih \leq \lambda$ on $\text{supp} g$. Then for every $\beta \in T'$ we have

$$\sum_{\alpha \leq \beta} g(\alpha)h(\alpha) \leq \lambda g(\beta).$$

Proof. Without loss of generality we may consider the case when $\beta$ is the unique maximal element of $T'$ and $T' = \text{supp} g$. We induct on the depth of the tree. Let $T'$ be given and suppose that the claim is known for all its branches. Then by the inductive hypothesis and superadditivity we have

$$\sum_{\alpha \leq \beta} g(\alpha)h(\alpha) = g(\beta)h(\beta) + \sum_{\beta' \in \text{ch}(\beta)} g(\beta') \sum_{\alpha \leq \beta'} h(\alpha')$$

$$\leq g(\beta)h(\beta) + \sum_{\beta' \in \text{ch}(\beta)} g(\beta') \sup_{\alpha \leq \beta'} \sum_{\alpha \leq \beta' \leq \beta} h(\alpha')$$

$$\leq g(\beta)h(\beta) + \sum_{\beta' \in \text{ch}(\beta)} g(\beta') \sup_{\alpha \leq \beta} \sum_{\alpha \leq \alpha' < \beta} h(\alpha')$$

$$= g(\beta) \sup_{\alpha \leq \beta} \sum_{\alpha \leq \alpha' < \beta} h(\alpha').$$

Lemma 2.3. Let $I$ be an integral operator with a positive kernel and $f, g$ positive functions. Then

$$\int (If)^2 g \leq \left( \sup_{\text{supp} g} II^* g \right) \int f^2.$$

Proof. Without loss of generality $f$ is positive. By duality we have

$$\int (If)^2 g = \int fI^*(If \cdot g) \leq \|f\|_2 \|I^*(If \cdot g)\|_2.$$

By the hypothesis $Ih(x) = \int K(x, y)h(y)$ with a positive kernel $K$. Hence

$$\|I^*(If \cdot g)\|_2^2 = \int I^*(If \cdot g)I^*(If \cdot g)$$

$$= \int K(x, y)((If)(x)g(x))K(x', y)((If)(x')g(x')) \, d(x, x', y)$$

$$\leq \int \frac{1}{2}(If(x)^2 + If(x')^2)K(x, y)(g(x))(g(x')) \, d(x, x', y)$$

$$= \frac{1}{2} \int \left( II^* g \right) (If)^2 \cdot g + \int I^*(g)I^*((If)^2 \cdot g)$$

$$= \int (II^* g) \cdot (If)^2 \cdot g$$

$$\leq \left( \sup_{\text{supp} g} II^* g \right) \int (If)^2 \cdot g.$$
The conclusion follows by rearranging the terms. □

Lemma 2.4. Let \( g, f, w : T \to [0, \infty) \) be positive functions and \( \lambda, \delta > 0 \). Assume that \( g \) is superadditive and \( I(wg) \leq \delta \) on \( \text{supp} \, f \). Then there exists a positive function \( \phi : T \to [0, \infty) \) such that

\[
I(w\phi) \gtrsim I(wf) \quad \text{on} \quad \{ \lambda/2 < I(wg) \leq 2\lambda \},
\]

and

\[
\int w\phi^2 \lesssim \delta \int wf^2.
\]

Proof. Without loss of generality we may assume \( \lambda \geq 4\delta \). Define

\[
\phi(\alpha) := \frac{1}{\lambda} 1_{\delta < I(wg)(\alpha) \leq 2\lambda} I(wf)(\alpha)g(\alpha)
\]

We prove first (2.5). Let \( \omega \in T \) be such that \( \lambda/2 < I(wg)(\omega) \leq 2\lambda \). Then for every \( \alpha \in T \) with \( \alpha \geq \omega \) and \( I(wg)(\alpha) \geq \delta \) we have

\[
I(wf)(\alpha) = I(wf)(\omega).
\]

It follows that

\[
I(w\phi)(\omega) = \frac{1}{\lambda} \sum_{\alpha \geq \omega, \delta < I(wg)(\alpha) \leq 2\lambda} I(wf)(\alpha)(wg)(\alpha)
\]

\[
= I(wf)(\omega) \frac{1}{\lambda} \sum_{\alpha \geq \omega, \delta < I(wg)(\alpha)} (wg)(\alpha)
\]

\[
= I(wf)(\omega) \frac{1}{\lambda} (I(wf)(\omega) - I(wg)(\alpha_{\text{min}))),
\]

where \( \alpha_{\text{min}} \) is the smallest \( \alpha \) outside of the summation range if it exists (otherwise that term is omitted). But then \( I(wg)(\omega) \geq \lambda/2 \) and \( I(wg)(\alpha_{\text{min}}) \leq \delta \), and (2.5) follows.

Next we will prove the energy estimate (2.6). Let \( \mathcal{U} := \{ I(wg) \leq \delta \} \), so that \( \mathcal{U} \) is an up-set and \( f \) is supported on \( \mathcal{U} \). By Lemma 2.3 with the operator \( I(\sqrt{w}1_{\mathcal{U}}) \) and functions \( f\sqrt{w} \) and \( g^2w1_{I(wg) \leq 2\lambda} \) we can estimate

\[
\int w\phi^2 \leq \frac{1}{\lambda^2} \sum_{\alpha \geq \omega, I(wg)(\alpha) \leq 2\lambda} I(wf)(\alpha)^2g(\alpha)^2w(\alpha)
\]

\[
\leq \frac{1}{\lambda^2} \left( \int wf^2 \right) \sup_{\mathcal{U}} I(w1_{\mathcal{U}} I^*(g^2w1_{I(wg) \leq 2\lambda})).
\]

By Lemma 2.2 with the superadditive function \( g1_{I(wg) \leq 2\lambda} \) and the function \( h = wg \) we can estimate

\[
I^*(g^2w1_{I(wg) \leq 2\lambda}) \leq 2\lambda g.
\]

Moreover, since \( \mathcal{U} \) is an up-set on a simple tree we have

\[
I(w1_{\mathcal{U}}g) \leq \sup_{\mathcal{U}} I(wg) \leq \delta.
\]

Combining the last three displays we obtain the energy estimate (2.6). □

3. SMALL ENERGY MAJORIZATION ON A BITREE

Theorem 3.1. Let \( m : T^2 \to [0, \infty) \) be a positive function and \( \lambda, \delta > 0 \). Let \( w \) be a positive measure on \( T^2 \) that is of the product form (1.15). Assume that \( m \) is superadditive and that \( \mathbb{I}(wm) \leq \delta \) on \( \text{supp} \, m \). Then there exists a positive function \( \phi \) on \( T^2 \) such that

\[
\mathbb{I}(w\phi) \gtrsim \mathbb{I}(wm) \quad \text{on} \quad \{ \lambda < \mathbb{I}(wm) \leq 2\lambda \},
\]

and

\[
\int w\phi^2 \lesssim \frac{\delta}{\lambda} \int wm^2.
\]
Proof of Theorem 3.1. We may assume \( \lambda \gg \delta \). For a fixed \( \alpha_y \in T_y \) we will apply Lemma 2.4 data \((g^{\alpha_y}, f^{\alpha_y}, w_x, \lambda)\) that we will now construct. Let

\[
g^{\alpha_y}(\beta_x) := \sum_{\alpha_y \geq \alpha_y} m(\beta_x \times \alpha_y')w_y(\alpha_y'), \quad \beta_x \in T_x.
\]

Superadditivity of \( m \) implies that \( g^{\alpha_y} \) is superadditive on \( T_x \). Also

\[
I_x(w_xg^{\alpha_y})(\beta_x) = \sum_{\beta_x \geq \beta_x, \alpha_y \geq \alpha_y} w_xm(\beta_x \times \alpha_y')w_y(\alpha_y') = \|w_xw_ym\|(\beta_x \times \alpha_y),
\]
where \( I_x \) denotes the Hardy operator on the tree \( T_x \). Let

\[
f^{\alpha_y}(\beta_x) := m(\beta_x \times \alpha_y).
\]

From (3.4) we see that

\[
I_x(w_xg^{\alpha_y})(\omega_x) \geq I_x(w_xf^{\alpha_y})(\omega_x) \quad \text{if} \quad \lambda/2 < I_x(w_xg^{\alpha_y})(\omega_x) \leq 2\lambda,
\]

\[
\int_{T_x} w_x(\phi^{\alpha_y})^2 \leq \frac{\delta}{\lambda} \int_{T_x} w_x(f^{\alpha_y})^2.
\]

It is nice to have the following formulas in mind (the second one follows from (2.7))

\[
g^{\alpha_y}(\beta_x) = I_y[w_yf^{\alpha_y}(\beta_x)]
\]

\[
\phi^{\alpha_y}(\beta_x) = \frac{1}{\lambda} \mathbf{1}_{\delta < I(w_g)(\alpha) \leq 2\lambda} I_x[w_xf^{\alpha_y}](\beta_x)g^{\alpha_y}(\beta_x)
\]

\[
= \frac{1}{\lambda} \mathbf{1}_{\delta < I(w_g)(\alpha) \leq 2\lambda} I_x[w_xf^{\alpha_y}](\beta_x)I_y[w_yf^{\alpha_y}(\beta_x)]
\]

Now put

\[
\phi(\beta_x \times \alpha_y) := \phi^{\alpha_y}(\beta_x).
\]

Summing (3.6) over all \( \alpha_y \in T_y \) we obtain

\[
\int_{T_x} w_xw_y\phi^2 = \sum_{\alpha_y} w_y(\alpha_y) \int_{T_x} w_x(\phi^{\alpha_y})^2 \\
\leq \frac{\delta}{\lambda} \sum_{\alpha_y} w_y(\alpha_y) \int_{T_x} w_x(f^{\alpha_y})^2 \\
= \frac{\delta}{\lambda} \int_T w_xw_ym^2.
\]

This shows (3.3).

Let now \( \omega = \omega_x \times \omega_y \in T^2 \) with \( \lambda < \|w_xw_ym\|(\omega) \leq 2\lambda \). Summing (3.5) in \( \alpha_y \) we obtain

\[
\|w_xw_y\phi\|(\omega) \geq \sum_{\lambda/2 < \|w_xw_ym\|(\omega_x \times \alpha_y) \leq 2\lambda} I_x(w_x\phi^{\alpha_y})(\omega_x)w_y(\alpha_y)
\]

\[
\geq \sum_{\lambda/2 < \|w_xw_ym\|(\omega_x \times \alpha_y)} I_x(w_xf^{\alpha_y})(\omega_x)w_y(\alpha_y)
\]

\[
= \|w_xw_ym\|(\omega_x \times \omega_y) - \|w_xw_ym\|(\omega_x \times \alpha_y, \min),
\]

where \( \alpha_y, \min \) is the smallest index outside the summation range if it exists (otherwise this term is omitted). But then \( \|w_xw_ym\|(\omega_x \times \omega_y) > \lambda \) and \( \|w_xw_ym\|(\omega_x \times \alpha_y, \min) \leq \lambda/2 \). This shows (3.2). \( \square \)
4. Main lemma

Define
\[ \mathbb{V}^\mu := \mathbb{I}(w1^\mu), \]
\[ E_\delta := \{ \mathbb{V}^\mu \leq \delta \} \subseteq T^2, \]
\[ \mathbb{V}^\mu_\delta := \mathbb{I}(w1_{E_\delta}1^\mu), \]
\[ \mathcal{E}_\delta[\mu] := \int_{T^2} \mathbb{V}^\mu_\delta \, d\mu = \int_{E_\delta} w(1^\mu)^2, \]
\[ \mathcal{E}[\mu] := \int_{T^2} \mathbb{V}^\mu \, d\mu = \sum_{\alpha \in T^2} w(\alpha)((1^\mu)(\alpha))^2. \]

**Lemma 4.1.** Let \( \mu, \rho, w \) be positive measures on \( T^2 \) and \( \delta > 0 \). Assume that \( w \) is of the product form \( (1.15) \). Then
\[ \left( \int \mathbb{V}^\mu_\delta \, d\rho \right)^3 \lesssim \delta \mathcal{E}_\delta[\mu] \mathcal{E}[\rho]|\rho|. \] (4.2)

**Remark 4.3.** On an ordinary tree Lemma 4.1 is trivial because in this case \( \mathbb{V}^\mu_\delta \leq \delta \), so (4.2) can be recovered from the pair of estimates
\[ \int \mathbb{V}^\mu_\delta \, d\rho \leq \delta |\rho|, \quad \int \mathbb{V}^\mu_\delta \, d\rho \leq \mathcal{E}_\delta[\mu]^{1/2} \mathcal{E}[\rho]^{1/2}. \]

On a bitree the potential \( \mathbb{V}^\mu_\delta \) can be considerably bigger than \( \delta \).

**Proof.** Without loss of generality \( \mathcal{E}_\delta[\mu] \neq 0 \) and \( \rho \neq 0 \). Let \( \lambda > 0 \) be chosen later.

Let \( m(\alpha) := \mathbb{I}(\mu)(\alpha)1_{E_\delta}(\alpha) \). Since \( E_\delta \) is an up-set, this function is superadditive.

Also, \( \mathbb{I}(wm) = \mathbb{V}^\mu_\delta \leq \mathbb{V}^\mu \leq \delta \) on \( \text{supp} \, m \), and \( \mathcal{E}_\delta[\mu] = \int w \mu^2 \). By Theorem 3.1 with data \((g, f, 2^n \lambda, 3)\) we obtain functions \( \phi_n \) such that \( \mathbb{I}(w \phi_n) \geq \mathbb{I}(wm) \) on the set \( \{2^n \lambda < \mathbb{I}(wm) \leq 2^{n+1} \lambda \} \) and \( \int w \phi_n^2 \leq \delta/(2^n \lambda) \cdot \int w \mu^2 \). Then
\[ \int \mathbb{V}^\mu_\delta \, d\rho = \int \{ \mathbb{V}^\mu_\delta \leq \lambda \} \mathbb{V}^\mu_\delta \, d\rho + \sum_{n=0}^{\infty} \int \{2^n \lambda < \mathbb{V}^\mu_\delta \leq 2^{n+1} \lambda \} \mathbb{V}^\mu_\delta \, d\rho \]
\[ \leq \lambda |\rho| + \sum_{n=0}^{\infty} \int \{2^n \lambda < \mathbb{V}^\mu_\delta \leq 2^{n+1} \lambda \} \mathbb{I}(w \phi_n) \, d\rho \]
\[ \leq \lambda |\rho| + \sum_{n=0}^{\infty} (\int w \phi_n^2)^{1/2} \mathcal{E}[\rho]^{1/2} \]
\[ \leq \lambda |\rho| + \sum_{n=0}^{\infty} C(\delta/(2^n \lambda))^{1/2} \mathcal{E}_\delta[\mu]^{1/2} \mathcal{E}[\rho]^{1/2}. \]
\[ \leq \lambda |\rho| + C(\delta/\lambda)^{1/2} \mathcal{E}_\delta[\mu]^{1/2} \mathcal{E}[\rho]^{1/2}. \]

Substituting \( \lambda = (\delta \mathcal{E}_\delta[\mu] \mathcal{E}[\rho])^{1/3} |\rho|^{-2/3} \) we obtain (4.2).

**Remark 4.4.** By the Cauchy–Schwarz inequality we have the trivial estimate
\[ \int \mathbb{V}^\mu_\delta \, d\rho \leq \mathcal{E}_\delta[\mu]^{1/2} \mathcal{E}[\rho]^{1/2}. \]

Lemma 4.1 improves upon this estimate if \( \delta |\rho| \lesssim \mathcal{E}_\delta[\mu]^{1/2} \mathcal{E}[\rho]^{1/2} \), which is equivalent to \( \lambda/\delta \gtrsim 1 \) in the proof.

5. Carleson implies hereditary Carleson

Taking \( \rho = \mu \) in Lemma 4.1 we obtain the following result.

**Corollary 5.1.** Let \( \mu, w \) be positive measures on \( T^2 \) and \( \delta > 0 \). Assume that \( w \) is of the product form \((1.15)\). Then
\[ (\mathcal{E}_\delta[\mu])^2 \leq (4.2) \delta \mathcal{E}[\mu] |\mu|. \] (5.2)
Corollary 5.3. Let \( \nu, w \) be positive measures on \( T^2 \) and

\[
E := \left\{ \mathbb{V}^\nu > \frac{1}{4C_{(4.2)}} \frac{\mathcal{E}[\nu]}{|\nu|} \right\} \subseteq T^2.
\]  

(5.4)

Assume that \( w \) is of the product form (1.15). Then

\[
\mathcal{E}_E[\nu] := \sum_{\alpha \in E} w(\alpha)(\mathbb{I}^n \nu(\alpha))^2 \geq \frac{1}{2} \mathcal{E}[\nu].
\]  

(5.5)

Proof. By Corollary 5.1 we have

\[
E \mathbb{E}[\nu] = E \mathbb{E}[\nu] \geq (C_{(4.2)} \frac{1}{4C_{(4.2)}} \frac{\mathcal{E}[\nu]}{|\nu|})^{1/2} = \mathcal{E}[\nu]/2,
\]

and the claim follows.

\( \square \)

Theorem 5.6. Let \( \mu, w \) be positive measures on \( T^2 \). Assume that \( w \) is of the product form (1.15). Then

\[
|w, \mu|_{HC} \lesssim |w, \mu|_{C}.
\]

Theorem 5.6 is also contained in Theorem 7.7, but we give a separate short proof.

Proof. Without loss of generality \( \mu \not\equiv 0 \) and \( |w, \mu|_{C} = 1 \). Let

\[
A := |w, \mu|_{HC} = \sup_{E \subseteq T^2, \mu(E) \neq 0} \frac{\mathcal{E}[\mu 1_E]}{\mu(E)}
\]  

(5.7)

be the hereditary Carleson constant. Since \( T^2 \) is finite, the constant \( A \) is finite, and there exists a maximizer \( E \) for (5.7). Let \( \nu := \mu 1_E \) and

\[
D := \left\{ \mathbb{V}^\nu > \frac{1}{4C_{(4.2)}} A \right\}.
\]  

(5.8)

Since \( A = \frac{\mathcal{E}[\nu]}{|\nu|} \), by Corollary 5.3, the trivial inequality \( \nu \leq \mu \), and the Carleson condition (1.9) we have

\[
\frac{1}{2} \mathcal{E}[\nu] \leq \mathcal{E}_D[\nu] \leq \mathcal{E}_D[\mu] \leq \mu(D).
\]  

(5.9)

In particular, \( \mu(D) \neq 0 \). On the other hand, by (5.8), the Cauchy–Schwarz inequality, (5.9), and the definition of \( A \) we have

\[
\frac{A}{4C_{(4.2)}} \mu(D) \leq \int_D \mathbb{V}^\nu d\mu \leq \mathcal{E}[\nu]^{1/2} \mathcal{E}[\mu 1_D]^{1/2} \leq 2^{1/2} A^{1/2} \mu(D).
\]  

(5.10)

It follows that \( A \leq 2^{5/2} \frac{C^2}{C_{(4.2)}} \).

\( \square \)

It is better to run this argument with \((4/5)^2\) in place of \(1/4\) in (5.4). Then it gives \(5^5/2^8 \leq 13\) in place of \(2^5 = 32\).

6. HEREDITARY CARLESON IMPLIES CARLESON EMBEDDING

Theorem 6.1. Let \( \mu, \rho, w \) be positive measures on \( T^2 \). Assume that \( w \) is of the product form (1.15) and

\[
|w, \mu|_{HC} \leq 1, \quad |w, \rho|_{HC} \leq 1.
\]

Then

\[
\int \mathbb{V}^\mu d\rho \leq 2^2 C_{(4.2)}^{1/4} |\mu|^{3/8} |\rho|^{5/8}.
\]

This improves upon the estimate

\[
\int \mathbb{V}^\mu d\rho \leq \mathcal{E}[\mu]^{1/2} \mathcal{E}[\rho]^{1/2} \lesssim |\mu|^{1/2} |\rho|^{1/2}
\]

that is immediate by Cauchy–Schwarz and the Carleson condition.
Proof. Let $\kappa > 0$ be chosen later. By Lemma 4.1 and the Carleson condition we obtain

$$\left( \int \mathbb{V}_\kappa \mu \, d\rho \right)^3 \leq C(4.2)\kappa |\mu||\rho|^2.$$  

Consider the down-set $E := \{ \mathbb{V}^\mu > \kappa \} \subset T^2$. By the Cauchy–Schwarz inequality and the Carleson condition we have

$$\int (\mathbb{V}^\mu - \mathbb{V}_\kappa^\mu) \, d\rho = \int_E \mathbb{V}^\mu \, \mu \, \rho \leq \mathcal{E}_E[\mu]^{1/2} \mathcal{E}_E[\rho]^{1/2} \leq \mu(E)^{1/2} |\rho|^{1/2}.$$  

Note that

$$\kappa \mu(E) \leq \int_E \mathbb{V}^\mu \, d\mu \leq \mathcal{E}[\mu]^{1/2} \mathcal{E}_1[\mu]^{1/2} \leq |\mu|^{1/2} \mu(E)^{1/2}$$

by the Carleson condition and (1.10), so $\mu(E)^{1/2} \leq \kappa^{-1}|\mu|^{1/2}$. Hence

$$\int \mathbb{V}^\mu \, d\rho \leq (C(4.2)\kappa |\mu||\rho|^2)^{1/3} + \kappa^{-1}|\rho|^{1/2} |\mu|^{1/2}.$$  

Choosing $\kappa = C(4.2)^{-1/4} |\mu|^{1/8} |\rho|^{-1/8}$ gives the claim.  

Remark 6.2. A careful reading of the proof of Theorem 6.1 gives the estimate

$$\int \mathbb{V}^\mu \, d\rho \lesssim \mathcal{E}[\mu]^{3/8} \mathcal{E}[\rho]^{3/8} |\rho|^{1/4} |w, \mu|_{HC}^{1/8}.$$  

Theorem 6.3. Let $\mu, w : T^2 \to [0, \infty)$. Assume that $w$ is of the product form (1.15). Then

$$[w, \mu]_{CE} \lesssim [w, \mu]_{HC}.$$  

The argument below is similar to the proof of [AH96, Theorem 7.1.1].  

Proof. Without loss of generality $[w, \mu]_{HC} = 1$. Let $\psi : T^2 \to [0, \infty)$ and consider

$$\psi \mu = \int_{t=0}^\infty \mu_t \, dt, \quad \mu_t := \mu 1_{\{\psi > t\}}.$$  

Then $[w, \mu_t]_{HC} \leq [w, \mu]_{HC} = 1$ for every $0 < t < \infty$. Expanding the square we obtain

$$\int w(\mathbb{V}^\mu(\psi \mu))^2 \sim 2 \int_{0 < s < t < \infty} \int_{T^2} w(\mathbb{V}^\mu_\mu s)(\mathbb{V}^\mu_\mu t) \, ds \, dt$$

by Theorem 6.1 \leq 2 \int_{0 < s < t < \infty} |\mu|^{3/8} |\mu_t|^{5/8} \, ds \, dt

$$= 2 \int_{r=0}^1 \int_{t=0}^\infty t |\mu_t|^{3/8} |\mu_t|^{5/8} \, dt \, dr$$

$$= 2 \int_{r=0}^1 r^{-3/4} \int_{t=0}^\infty (r^2 t |\mu_t|)^{3/8} (t |\mu_t|)^{5/8} \, dt \, dr$$

by Hölder \leq 2 \int_{r=0}^1 r^{-3/4} \left( \int_{t=0}^\infty r^2 t |\mu_t| \, dt \right)^{3/8} \left( \int_{t=0}^\infty t |\mu_t| \, dt \right)^{5/8} \, dr

$$= \left( \int_{r=0}^1 r^{-3/4} \, dr \right) \left( 2 \int_{t=0}^\infty |\mu_t| \, dt \right)$$

$$\leq 2 \int_{t=0}^\infty |\mu_t| \, dt$$

$$= \int \psi^2 \, d\mu.$$  

7. Box implies (hereditary) Carleson

Clearly the box constant (1.8) is smaller than the hereditary Carleson constant (1.10). In this section we show a converse inequality for product weights $w$. This counterintuitive result represents a certain combinatorial property of all planar measures.
7.1. Balancing lemma. We include the proof of the next result for completeness.

**Lemma 7.1** ([AHMV18b, Lemma 3.1]). Let \( \nu : T^2 \to [0, \infty) \) be a non-negative function with
\[
\mathcal{E}[\nu] = \int \mathcal{V}^\nu \, d\nu \geq A|\nu|.
\]
Then there exists a down-set \( \check{E} \subset T^2 \) such that for the measure \( \check{\nu} := \nu \mathbf{1}_{\check{E}} \) we have
\[
\mathcal{V}^{\check{\nu}} \geq \frac{A}{3} \quad \text{on } \check{E},
\]
and
\[
\mathcal{E}[\check{\nu}] \geq \frac{1}{3} \mathcal{E}[\nu].
\]

**Proof.** Replacing \( \nu \) by \( 3\nu/A \) we may assume \( A = 3 \). Let \( E_0 := T^2 \) and \( \nu_0 := \nu \mathbf{1}_{E_0} \). We then define inductively
\[
E_{k+1} := E_k \setminus \{ \mathcal{V}^{\nu_k} \leq 1 \}, \quad \nu_{k+1} := \nu \mathbf{1}_{E_{k+1}}.
\]
The sequence \( (E_k) \) consists of down-sets in \( T^2 \) and is decreasing, and since \( T^2 \) is finite it must stabilize, that is, \( E_{N+1} = E_N \) for some \( N \). By construction we have
\[
\mathcal{V}^{\nu_N} \geq 1 \quad \text{on } E_N.
\]
Let \( \sigma_k := \nu_k - \nu_{k+1} = \nu \mathbf{1}_{E_k \setminus E_{k+1}} \). Then
\[
\mathcal{E}[\nu] = \int \mathcal{V}^{\nu_0} \, d\nu_0 = \int \mathcal{V}^{\nu_1} \, d\nu_1 + \int \mathcal{V}^{\nu_0} \, d\sigma_0 + \int \mathcal{V}^\nu \, d\nu_1 + 2 \int \mathcal{V}^{\nu_0} \, d\sigma_0 \leq \mathcal{E}[\nu_1] + 2|\sigma_0| \leq \cdots \leq \mathcal{E}[\nu_N] + 2|\sigma_0| + \cdots + 2|\sigma_{N-1}|
\]
\[
\leq \mathcal{E}[\nu_N] + 2|\nu|.
\]
Since \( \mathcal{E}[\nu] \geq 3|\nu| \) by assumption, we obtain
\[
\mathcal{E}[\nu_N] \geq \frac{1}{3} \mathcal{E}[\nu].
\]
This gives the conclusion with \( \check{E} := E_N \). \( \square \)

7.2. Stopping time argument. Let
\[
\eta^{\mu}_{\text{good}, \epsilon}(\omega) := \sum_{\alpha \geq \omega: \sum_{\beta \leq \alpha} w(\beta)\mathbf{1}^* \mu(\beta) > \epsilon} w(\alpha)\mathbf{1}^* \mu(\alpha).
\]

**Lemma 7.2.** Let \( \mu, w \) be positive measures on \( T^2 \) and \( \epsilon > 0 \). Then for every \( \omega \in T^2 \) at least one of the following conditions holds.

1. \( \forall^{\mu}_{\text{good}, \epsilon}(\omega) > \epsilon \), or
2. \( \forall^{\mu}_{4\epsilon}(\omega) \geq \forall^{\mu}(\omega)/2 \).

**Proof.** Fix \( \omega \in T^2 \) with
\[
\forall^{\mu}_{\text{good}, \epsilon}(\omega) \leq \epsilon. \quad (7.3)
\]
The set
\[
\mathcal{U} := \{ \alpha \geq \omega: \sum_{\omega \leq \beta \leq \alpha} w(\beta)\mathbf{1}^* \mu(\beta) > \epsilon \} \quad (7.4)
\]
is an up-set. If \( \mathcal{U} = \emptyset \), then in fact \( \forall^{\mu}(\omega) \leq \epsilon \), so \( \forall^{\mu}_{4\epsilon}(\omega) = \forall^{\mu}(\omega) \). Suppose henceforth \( \mathcal{U} \neq \emptyset \). Let
\[
\mathcal{W} := \{ \alpha \geq \omega: \forall^{\mu}(\alpha) \leq 4\epsilon \}. \quad (7.5)
\]
Without loss of generality we may assume \( \omega \notin \mathcal{W} \), since otherwise \( \forall^{\mu}_{4\epsilon}(\omega) = \forall^{\mu}(\omega) \).
For \( \alpha \in \mathcal{U} \) we have
\[
\forall^{\mu}(\alpha) = \sum_{\alpha' \geq \alpha} w(\alpha')\mathbf{1}^* \mu(\alpha') \leq \sum_{\alpha' \in \mathcal{U}} w(\alpha')\mathbf{1}^* \mu(\alpha') = \forall^{\mu}_{\text{good}, \epsilon}(\omega) \leq \epsilon \leq 4\epsilon,
\]
so \( \mathcal{U} \subseteq \mathcal{W} \).
Let $\alpha_{0,x}$ be minimal such that there exists $\alpha_{0,y}$ with $\alpha_0 = \alpha_{0,x} \times \alpha_{0,y} \in \mathcal{W}$, and let $\alpha_{0,y}$ be minimal with this property. By assumption $\alpha_0 \neq \omega$. For $j \geq 0$ construct inductively

$$\alpha_{j+1,x} := \min\{\tau_x : \tau_x \times \alpha_{j,y} \in \mathcal{U}\}, \quad \alpha_{j+1,y} := \min\{\tau_y : \alpha_{j+1,x} \times \tau_y \in \mathcal{W}\}. $$

Notice $\alpha_{j+1,x} \geq \alpha_{j,x}$ and $\alpha_{j+1,y} \leq \alpha_{j,y}$. We stop at $J$ for which either the minimum in the definition of $\alpha_{j+1,x}$ is taken over the empty set or $\alpha_{J,y} = \omega_y$.

We claim that for every $j$, except possibly $j = J$ if $J > 0$, we have

$$\sum_{\alpha_{j,x} \leq \tau_x < \alpha_{j+1,x}} \sum_{\alpha_{j,y} \leq \tau_y} (w^\mu)(\tau_x \times \tau_y) \geq 2\epsilon. \tag{7.6}$$

Here we interpret the restriction by $\alpha_{j+1,x}$ as nonexistent if $j = J$.

Consider first the case that $\alpha_{j,y} > \omega_y$ (see Figure 1). Let $\tilde{\alpha}_{j,y} \in T_y$ be the maximal element with $\alpha_{j,y} > \tilde{\alpha}_{j,y} \geq \omega_y$. Let $\tilde{\alpha}_{j,x} \in T_x$ be the maximal element with $\tilde{\alpha}_{j,x} \geq \omega_x$ and $\tilde{\alpha}_{j,x} \times \tilde{\alpha}_{j,y} \notin \mathcal{U}$. Then

$$\sum_{\alpha_{j,x} \leq \tau_x < \alpha_{j+1,x}} \sum_{\alpha_{j,y} \leq \tau_y} (w^\mu)(\tau_x \times \tau_y) \geq \mathcal{V}^\mu(\alpha_{j,x} \times \tilde{\alpha}_{j,y}) - \sum_{\omega_x \leq \tau_x \leq \tilde{\alpha}_{j,x}} (w^\mu)(\tau_x \times \tilde{\alpha}_{j,y}) - \mathcal{V}_{\mathrm{good},\epsilon}(\omega) \geq 4\epsilon - \epsilon - \epsilon,$$

where in the last inequality we used the definition (7.5) of $\mathcal{W}$ to bound the first term, the definition (7.4) of $\mathcal{U}$ to bound the middle term, and the hypothesis (7.3) to bound the last term, see Figure 1.

Consider now the case that $\alpha_{j,y} = \omega_y$. Then necessarily $j = J$ and as we do not prove the relation for $j = J$, $J > 0$ we may assume $J = 0$ and so $j = 0$. By assumption there is no restriction by $\alpha_{1,x}$ as $j = J$. Since $\omega \notin \mathcal{W}$, we have $\alpha_{0,x} > \omega_x$. Let $\tilde{\alpha}_x \in T_x$ be the maximal element with $\alpha_{0,x} > \tilde{\alpha}_x \geq \omega_x$. Let $\tilde{\alpha}_y \in T_y$ be the maximal element with $\tilde{\alpha}_y \geq \omega_y$. Then by construction $\tilde{\alpha}_x \times \tilde{\alpha}_y \notin \mathcal{W}$, hence also $\tilde{\alpha}_x \times \tilde{\alpha}_y \notin \mathcal{U}$. It follows that

$$\sum_{\tau_x \geq \alpha_{0,x}} \sum_{\tau_y \geq \alpha_{0,y}} (w^\mu)(\tau_x \times \tau_y) \geq \mathcal{V}^\mu(\tilde{\alpha}_x \times \alpha_{0,y}) - \sum_{\omega_y \leq \tau_y \leq \tilde{\alpha}_y} (w^\mu)(\tilde{\alpha}_x \times \tau_y) \geq 4\epsilon - \epsilon = 3\epsilon,$$

where in the last inequality we again used the definition (7.5) of $\mathcal{W}$ to bound the first term and the definition (7.4) of $\mathcal{U}$ to bound the second term. This finishes the proof of (7.6).

Since the summation sets in (7.6) are disjoint and contained in $\mathcal{W}$ we obtain

$$\mathcal{V}_{4\epsilon}(\omega) = \sum_{\alpha \in \mathcal{W}} w(\alpha)\mu(\alpha) \geq 2\epsilon \max(J,1) \geq \epsilon(J + 1). $$

---

**Figure 1.** Proof of the lower bound (7.6) for the hatched area: \( \square \geq \mathcal{U} - \mathcal{W} - \square \).
On the other hand,
\[ \mathbb{V}^\mu(\omega) - \mathbb{V}_{4\epsilon}^\mu(\omega) = \sum_{\alpha < \omega: \alpha \notin \mathcal{W}} w(\alpha) \mu(\alpha) \leq \sum_{j=0}^{J} \sum_{\alpha < \omega: \alpha \leq \alpha_j} w(\alpha) \mu(\alpha) \leq \epsilon(J+1), \]

since each inner sum is over a rectangular region and the maximum of that region is not in \( \mathcal{U} \). The last two displays show that the second alternative in the statement of the lemma holds.

\[ \square \]

7.3. **Box condition implies hereditary Carleson.** The next result contains the last missing inequality in Theorem 1.14.

**Theorem 7.7.** Let \( \nu, w : T^2 \to [0, \infty) \) be positive functions. Assume that \( w \) is of the product form (1.15). Then

\[ [w, \nu]_{\text{HC}} \lesssim [w, \nu]_{\Box}. \]

**Proof.** By scaling we may assume \([w, \nu]_{\Box} = 1\) without loss of generality. Let \( A := [w, \nu]_{\text{HC}} \). Let \( E \subset T^2 \) be a subset such that \( \mu = \nu 1_E \neq 0 \) and \( \mathcal{E}[\mu] \geq \frac{2}{3} A |\mu| \).

By Lemma 7.1 there exists a further subset \( \tilde{E} \subset T^2 \) such that \( \tilde{\mu} := \nu 1_{\tilde{E}} \) satisfies

\[ \forall \tilde{\mu} \geq \frac{A}{4} \text{ on } \tilde{E} \]

and \( \tilde{\mu} \neq 0 \). Thus, replacing \( \mu \) by \( \tilde{\mu} \), we may assume \( \forall \mu \geq A/4 \) on \( \text{supp } \mu \).

By Corollary 5.1 we know

\[ \int \mathbb{V}_{4\epsilon}^\mu d\mu \lesssim (\epsilon \mathcal{E}[\mu]|\mu|)^{1/2} \leq (\epsilon A)^{1/2} |\mu|. \]

Taking \( \epsilon = cA \) for a small constant \( c \) we can ensure that \( \mu \{ \forall \mu \leq A/10 \} \geq |\mu|/2 \).

By Lemma 7.2 we have \( \mathbb{V}_{\text{good, } \epsilon}^\mu > \epsilon \) on that set, so

\[ \int \mathbb{V}_{\text{good, } \epsilon}^\mu d\mu \gtrsim \epsilon |\mu| \geq c \mathcal{E}[\mu]. \quad (7.8) \]

We claim that for a sufficiently small absolute constant \( 0 < \theta \) we have

\[ \mathcal{E}[\mu] \lesssim \sum_{\alpha: \theta \mathbb{V}_\alpha^\mu(\alpha) \leq \mathcal{E}[\mu]} w(\alpha)(\mathbb{V}_\alpha^\mu(\alpha))^2. \quad (7.9) \]

Indeed, suppose that \( \alpha \) is such that

\[ \theta |\mathbb{V}_\alpha^\mu(\alpha)| > \mathcal{E}[\mu] = \sum_{\omega \leq \alpha} |\mathbb{V}_\omega^\mu(\omega)|, \quad \mathbb{V}_\alpha^\mu(\alpha) = \sum_{\beta: \omega \leq \beta \leq \alpha} w(\beta)(\mathbb{V}_\beta^\mu(\beta)). \]

Then by the Tchebyshov inequality we have

\[ \sum_{\omega \leq \alpha: \mathbb{V}_\omega^\mu(\omega) \leq \epsilon} \mu(\omega) = \mathbb{V}_\alpha^\mu(\alpha) - \sum_{\omega \leq \alpha: \mathbb{V}_\omega^\mu(\omega) > \epsilon} \mu(\omega) \geq (1-\theta) \mathbb{V}_\alpha^\mu(\alpha). \]

It follows that

\[ \sum_{\alpha: \theta \mathbb{V}_\alpha^\mu(\alpha) > \mathcal{E}[\mu]} w(\alpha)(\mathbb{V}_\alpha^\mu(\alpha))^2 \leq \sum_{\alpha} w(\alpha) \mathbb{V}_\alpha^\mu(\alpha) \frac{1}{1-\theta} \sum_{\omega \leq \alpha: \mathbb{V}_\omega^\mu(\omega) \leq \epsilon} \mu(\omega) \]

\[ = \frac{1}{1-\theta} \sum_{\omega} \mu(\omega) \sum_{\alpha: \mathbb{V}_\alpha^\mu(\alpha) \leq \epsilon} w(\alpha) \mathbb{V}_\alpha^\mu(\alpha) = \frac{1}{1-\theta} \sum_{\omega} \mu(\omega)(\mathbb{V}_\omega^\mu - \mathbb{V}_{\text{good, } \epsilon}^\mu)(\omega). \]

Therefore

\[ \sum_{\alpha: \theta \mathbb{V}_\alpha^\mu(\alpha) \leq \mathcal{E}[\mu]} w(\alpha)(\mathbb{V}_\alpha^\mu(\alpha))^2 \geq \int \mathbb{V}_\omega^\mu d\mu - \frac{1}{1-\theta} \int (\mathbb{V}_\omega^\mu - \mathbb{V}_{\text{good, } \epsilon}^\mu) d\mu \]

\[ = -\frac{\theta}{1-\theta} \int \mathbb{V}_\omega^\mu d\mu + \frac{1}{1-\theta} \int \mathbb{V}_{\text{good, } \epsilon}^\mu d\mu. \]

By (7.8) the latter term is \( \gtrsim \mathcal{E}[\mu] \), and it dominates the first term if \( \theta \) is sufficiently small. This finishes the proof of the claim (7.9).
By Corollary 5.1 again and since $\forall \mu \geq A/4$ on $\text{supp} \mu$ we also have

$$E_{cA}[\mu] \lesssim (c')^{1/2} E[\mu].$$

(7.10)

Taking $c'$ sufficiently small and combining (7.10) with (7.9) we obtain

$$E[\mu] \lesssim \sum_{\alpha \in R} w(\alpha)(\Pi^* \mu(\alpha))^2, \quad R = \{\alpha: \theta \epsilon \Pi^* \mu(\alpha) \leq E[\mu], \forall^\mu(\alpha) \geq c'A\}.$$

For each $\alpha \in R$ we have

$$\theta \epsilon \Pi^* \mu(\alpha) \leq E_\alpha[\mu] \leq [w, \nu]_{Box} \Pi^* \nu(\alpha) = \Pi^* \sigma(\alpha),$$

where $\sigma = \nu 1_F$, $F = \{\beta \in T^2: \exists \alpha \in R, \alpha \geq \beta\}$. It follows that

$$\epsilon^2 E[\mu] \lesssim E[\sigma].$$

(7.11)

On the other hand, using definition of $A$, the fact that $\forall^\mu \geq A$ on $\text{supp} \sigma$, and the Cauchy–Schwarz inequality we obtain

$$E[\sigma] \leq A|\sigma| \lesssim \int \forall^\mu \ d\sigma \leq E[\mu]^{1/2} E[\sigma]^{1/2}.$$

(7.12)

From (7.12) we obtain $E[\sigma] \lesssim E[\mu]$, and inserting this into (7.11) gives $\epsilon \lesssim 1$. By the choice of $\epsilon$ this in turn implies $A \lesssim 1$. \qed

7.4. A comparison with a result of E. Sawyer. The setting in Sawyer’s paper [Saw85] can be reduced to the following situation. Assume that the weight $w$ has a very special structure: it is supported on ancestors of one small square. Namely, there exists a point $\omega_0 \in \partial T^2$ such that

$$\text{supp} w \subset \{\alpha \in T^2 | \alpha \geq \omega_0\}.$$

(7.13)

If such $w$ is identically 1 on ancestors of one small square $\omega_0$, then it is a product weight, and we know the answer: condition (7.15c) below is necessary and sufficient for embedding. But for general weight supported on ancestors of one small square $\omega_0$, a careful reading of [Saw85] provides the following result.

Theorem 7.14 (cf. [Saw85, Theorem 1(A)]). Suppose that the weight $w$ satisfies the support condition (7.13) for some small square $\omega_0 \in T^2$. Then $[w, \mu]_{CE}$ is finite if and only for some $A < \infty$ and every $\beta \in T^2$ with $\beta \geq \omega_0$ the following conditions hold:

$$\Pi^* \mu(\beta) \Pi w(\beta) \leq A^2,$$

(7.15a)

$$\sum_{\alpha \geq \beta \geq \omega_0} \mu(\alpha)(\Pi w(\alpha))^2 \leq A^2 \sum_{\alpha \geq \beta} w(\alpha),$$

(7.15b)

$$\sum_{\omega_0 \leq \alpha \leq \beta} w(\alpha)(\Pi^* \mu(\alpha))^2 \leq A^2 \sum_{\alpha \geq \beta} \mu(\alpha).$$

(7.15c)

No two of these conditions suffice to ensure $[w, \mu]_{CE} < \infty$.

Remark 7.16. The last condition (7.15c) is just the box condition (1.8). In other words, if we restrict the weight to be supported only on the hooked rectangles, but drop the requirement that it has a product structure, we see that the single box test (1.8) is getting replaced by three single box tests for the pair $(\mu, w)$.

To summarize: Sawyer considers the weight concentrated on the ancestors of a single point, and we consider the weight with product structure restrictions. It would be interesting to find a characterization of the Carleson embedding (1.11) for general $w$, which would also cover any subgraph of the bitree. One can even consider such a problem on an arbitrary directed graph without directed cycles, but even for the bitree the problem seems to be very difficult.
8. General setting for constructing counterexamples

Let $T^2 = T^2_N$ be a finite (but very deep) dyadic bi-tree, that is, the set of all dyadic rectangles in the square $Q_0 = [0,1]^2$ with side lengths at least $2^{-N}$ ordered by inclusion. We denote the set of minimal elements of this bi-tree, that is, the small squares of size $2^{-N} \times 2^{-N}$, by $\partial T^2$, and elements of this set will be denoted by $\omega$. We denote sets of $\omega$’s by $\mathcal{E} \subset \partial T^2$ and identify them with their union, so we will write $Q \subset E$ if $Q$ is covered by the elements of $E$.

In all examples the measure $\mu$ will be supported on the square $Q_0$. As in (1.12) we identify it with a function on $T^2$ by setting $\tilde{\mu}(\omega) := \mu(\omega)$ for $\omega \in \partial T^2$ and $\tilde{\mu}(Q) := 0$ for $Q \notin \partial T^2$. Then $\mathbb{I}^* \tilde{\mu}(Q) = \mu(Q)$. With this convention the box condition (1.8) for the measure $\mu$ and weight $w = \{w_Q\}$ becomes

$$\sum_{Q \in T^2, Q \subset R} \mu^2(Q)w_Q \leq C \mu(R), \quad \text{for any } R \in T^2. \tag{8.1}$$

The Carleson condition (1.9) becomes

$$\sum_{Q \in T^2, Q \subset E} \mu^2(Q)w_Q \leq C \mu(E), \quad \text{for any } E \subset (\partial T)^2, \tag{8.2}$$

the hereditary Carleson (or Restricted Energy) condition (1.10) becomes

$$\sum_{Q \in \mathcal{D}} \mu^2(Q \cap E)w_Q \leq C \mu(E), \quad \text{for any } E \subset (\partial T)^2, \tag{8.3}$$

and the Carleson embedding (1.11) becomes

$$\sum_{Q \in \mathcal{D}} \left( \int_Q \varphi \, d\mu \right)^2 w_Q \leq C \int_{Q_0} \varphi^2 \, d\mu \quad \text{for any } \varphi \in L^2(Q_0, d\mu). \tag{8.4}$$

The implications

$$(8.1) \iff (8.2) \iff (8.3) \iff (8.4)$$

hold for arbitrary measures $\mu$ and weights $w$. For product weights $w$ the converse implications hold by Theorem 1.14, so all these conditions are in fact equivalent. We will show that the converse implications do not hold in general. We will do this by constructing $N$-coarse measures $\mu$ and weights $w$ on finite bi-trees $T^2$ of depth $N$ such that the discrepancies between box, Carleson, REC, and embedding constants grow with $N$.

9. Box condition does not imply Carleson condition

In [Car74] Carleson constructed families $\mathcal{R}$ of dyadic sub-rectangles of $Q = [0,1]^2$ as having the following two properties:

$$\forall R_0 \in T^2, \quad \sum_{R \subset R_0, R \in \mathcal{R}} m_2(R) \leq C_0 m_2(R_0), \tag{9.1}$$

but

$$\sum_{R \in \mathcal{R}} m_2(R) > C_1 m_2(\bigcup_{R \in \mathcal{R}} R), \tag{9.2}$$

with arbitrarily large ratios $C_1/C_0$, where $m_2$ is the planar Lebesgue measure. Choosing $\mu = m_2$ and

$$w_R := \begin{cases} \frac{1}{m_2(R)}, & R \in \mathcal{R}, \\ 0, & \text{otherwise} \end{cases}$$

we can identify the left-hand sides of (9.1) and (9.2) with the left-hand sides of (8.1) and (8.2), respectively. Hence the box condition (8.1) holds with constant $C_0$, while the Carleson condition (8.2) can only hold with constant $\geq C_1$.

The weight $w$ is rather wild here. But there is also a counterexample with $w_R \in \{0,1\}$ for all $R$, see [HPV19].
10. Carleson condition does not imply REC

Our aim here is to show that for general \( w, \mu \) the Carleson condition (8.2) is no longer sufficient for the embedding (8.4) or even the Restricted Energy Condition (8.3). Namely we prove the following statement.

**Proposition 10.1.** For any \( \delta > 0 \) there exists a number \( N \in \mathbb{N} \), a weight \( w : T_N^2 \to \{0, 1\} \), and a measure \( \mu \) on \( \partial T^2 \) such that \( \mu \) satisfies the Carleson condition (8.2) with the constant \( C_\mu = \delta \):

\[
\sum_{Q \subseteq E} \mu^2(Q)w_Q \leq \delta \mu(E), \quad \text{for any } E \subseteq (\partial T)^2, \quad (10.2)
\]

but there exists a set \( F \subseteq Q_0 \) such that

\[
\sum_{Q \in \mathcal{D}} \mu^2(Q \cap F)w_Q > \mu(F), \quad (10.3)
\]

hence the constant in (8.3) is at least 1.

We intend to give two examples of this kind. The first example is quite simple and is inspired by the counterexample for \( L^2 \)-boundedness of the biparameter maximal function (as it should be by Proposition A.2). In this example the weight \( w \) is supported on a very small subset of the bi-tree, which differs greatly from the original graph. The second example is somewhat more involved, but the weight \( w \) is supported on a much larger portion of the bi-tree; in fact it has the monotonicity property \( w_R \geq w_Q \) for \( R \supseteq Q \). Nevertheless there are not enough rectangles in the support of \( w \) to have the Carleson-REC equivalence.

10.1. **Notation.** In Sections 10 and 11 we consider dyadic bi-trees \( T^2 = T_N^2 \) of large but finite depth \( N \). We will construct examples in which the ratio between the constants goes to \( \infty \) as \( N \to \infty \).

We denote by \( \omega_0 := [0, 2^{-N}]^2 \) the left lower corner of the unit square. Given a dyadic rectangle \( R = [a, b] \times [c, d] \) let

\[
R^+ := [(a + b)/2, b] \times [c, d],
\]

\[
R^{++} := [a, b] \times ([c + d]/2, d],
\]

\[
R^{++} := [(a + b)/2, b] \times ([c + d]/2, d],
\]

be its right half, upper half, and upper right quadrant, respectively.

10.2. **A simple example.** Let \( Q_i = [0, 2^{-i+1}] \times [0, 2^{-N+i}] \) for \( j = i, \ldots, N \). Let measure \( \mu \) have mass 1 on \( \omega_0 \) and each of \( Q_i^{++} \), and mass 0 everywhere else. Let

\[
w_R := \begin{cases} 1 & \text{if } R \in \{\omega_0, Q_1, Q_2, \ldots, Q_N\}, \\ 0 & \text{else}. \end{cases}
\]

So we have \( N + 1 \) weights \( w_R \) equal to 1. For the set \( E = \omega_0 \) we have

\[
\mathcal{E}[\mu|E] = \mu(\omega_0)^2 + \sum_{i=1}^N \mu(\omega_0 \cap Q_i)^2 = (N + 1) = (N + 1)\mu(E).
\]

So the REC constant (8.3) is \( \geq N + 1 \).

Denoting \( Q_0 := \omega_0 \), for an arbitrary \( E \subseteq \partial T^2 \) we have

\[
\mathcal{E}_E[\mu] = \sum_{R \subseteq E, w_R \neq 0} \mu(R)^2 = \sum_{j:Q_j \subseteq E} \mu(Q_j)^2.
\]

Then since \( Q_i^{++} \cap Q_j = \emptyset \) unless \( i \in \{0, j\} \), we have

\[
\mathcal{E}_E[\mu] \leq \sum_{j:Q_j \subseteq E} 2^2 \leq 4\mu(E)
\]

So the Carleson condition (8.2) holds with constant 4.
10.3. The lack of maximal principle matters. In this section we construct a more complicated example in which the Carleson condition holds, but the restricted energy condition fails. The weight \( w \) still has values either 0 or 1, but the support \( \mathcal{R} \) of \( w \) is an up-set, that is, it contains every ancestor of every rectangle in \( \mathcal{R} \).

The example is based on the fact that potentials on bi-tree may not satisfy the maximal principle. So we start with constructing an \( N \)-coarse \( \mu \) such that we have
\[
\forall \mu \lesssim 1 \quad \text{on supp } \mu,
\] (10.4)
but
\[
\max \forall \mu \geq \forall \mu(\omega_0) \gtrsim \log N.
\] (10.5)

We define a collection of rectangles
\[
Q_j := [0, 2^{-2j}] \times [0, 2^{-2-j}N], \quad j = 1, \ldots, M \approx \log N.
\] (10.6)

Now we put
\[
\mathcal{R} := \{ R : Q_j \subset R \text{ for some } j = 1 \ldots M \}
\]
\[
w_Q := 1_{\mathcal{R}}(Q)
\]
\[
\mu(\omega) := \frac{1}{N} \sum_{j=1}^M \frac{1}{|Q_j^{++}|} \mathbf{1}_{Q_j^{++}}(\omega).
\] (10.7)

Here \( |Q| \) denotes the total amount of points \( \omega \in (\partial T)^2 \cap Q \), i.e. the amount of the smallest possible rectangles (of size \( 2^{-N} \times 2^{-N} \)) in \( Q \).

Observe that on \( Q_j \) the measure is basically a uniform distribution of the mass \( \frac{1}{N} \) over the upper right quarter \( Q_j^{++} \) of the rectangle \( Q_j \) (and these quadrants are disjoint).

To prove (10.4) we fix \( \omega \in Q_j^{++} \) and split
\[
\forall \mu(\omega) = \forall \mu_{Q_j^{++}}(\omega) + \mu(Q_j^+) + \mu(Q_j^+) + \forall \mu(\omega),
\]
where the first term sums up \( \mu(Q) \) for \( Q \) between \( \omega \) and \( Q_j^{++} \). It is easy to see that \( \forall \mu_{Q_j^{++}}(\omega) \lesssim \frac{1}{N} \) (the left-hand side is a double geometric sum). Trivially \( \mu(Q_j^+) + \mu(Q_j^+) \lesssim \frac{1}{N} \). The non-trivial part is the estimate
\[
\forall \mu(Q_j) \lesssim 1.
\] (10.8)

For each dyadic rectangle \( R \supseteq \omega_0 \) and each \( j' \) we have
\[
either \ Q_{j'} \subseteq R, \ or \ Q_{j'}^{++} \cap R = \emptyset.
\] (10.9)

Moreover, since the sides of rectangles \( Q_j \) are nested, the set \( \{ j' : Q_{j'} \subseteq R \} \) is an interval that contains \( j \). For an interval of integers \( [m, m+k] \) let
\[
C^{[m,m+k]} := \{ R \supseteq \omega_0 : \{ j' : Q_{j'} \subseteq R \} = [m, m+k] \}.
\]

Since each rectangle in \( C^{[m,m+k]} \) contains \([0, 2^{-2m}] \times [0, 2^{-2-m-k}N] \), we have
\[
\#C^{[m,m+k]} \leq (2^m + 1)(2^{-m-k}N + 1) \lesssim 2^{-k}N.
\] (10.10)

It follows that
\[
\forall \mu(Q_j) = \sum_{[m,m+k] \ni j} (\#C^{[m,m+k]})(k+1) \frac{1}{N} \lesssim \sum_{k \geq 0} (k+1)2^{-k}N \frac{1}{N} \lesssim 1.
\] (10.11)

This shows (10.8), and hence (10.4) is also proved.

Now we will estimate \( \forall \mu(\omega_0) \) from below. To this end we need a more careful lower bound on \( \#C^{[m,m+k]} \). The set \( C^{(j)} \) contains all rectangles \( R \) that contain \( Q_j \) and are contained in \([0, 2^{-2j-1}] \times [0, 2^{-2-j-1}N-1] \), so
\[
\#C^{(j)} \gtrsim 2^{j-1} \cdot 2^{-j-1}N \gtrsim N.
\] (10.12)
Hence
\[ V^\mu(\omega_0) \geq \sum_{j=1}^{M} \left( \#C^{(j)} \right) \frac{1}{N} \gtrsim M. \]  
(10.13)

This shows (10.5) as \( M \asymp \log N \).

**Remark 10.14.** In this example \( V^\mu \leq 1 \) on \( \text{supp} \mu \), and \( \text{cap}(\omega_0) = \frac{1}{N^2} \). Denote \( \lambda := \log N \). Then
\[ \text{cap}\{ \omega : V^\mu \geq \lambda \} \geq e^{-c\lambda}, \quad c > 0. \]  
(10.15)

Here capacity is the bi-tree capacity defined e.g. in [AMPS18]. So the example above essentially uses the non-existence of maximal principle for bi-tree potential. It would be interesting to prove the opposite inequality, namely, that while the maximal principle fails for the bi-tree potential, the set where it fails has small capacity:
\[ \text{cap}\{ \omega : V^\mu \geq \lambda \} \leq e^{-c\lambda}, \quad c > 0. \]  
(10.16)

Notice that Lemma 4.1 (see also [AMPS18, Theorem 1.6], [AHMV18b, Lemma 2.6]) gives us a much weaker estimate
\[ \text{cap}\{ \omega : V^\mu \geq \lambda \} \leq C\lambda^3. \]  
(10.17)

Now we construct the second example of \( \nu \) and \( w \) such that the Carleson condition holds, but the REC (restricted energy condition) fails. The weight \( w \) is chosen as in (10.7), so this time it is the indicator function of an up-set. With the measure \( \mu \) that we have just constructed we put
\[ \nu := \mu + \nu|_{\omega_0}, \]
where \( \nu|_{\omega_0} \) is the uniformly distributed over \( \omega_0 \) measure of total mass \( \frac{1}{N} \).

### 10.4. REC constant is large

Let us first give a lower bound for the REC constant. Consider \( F = \omega_0 \). Then by (10.12) we have
\[ \mathcal{E}[\nu|F] \geq \sum_{j=1}^{M} \left( \#C^{(j)} \right) \nu(\omega_0)^2 \gtrsim MN \cdot \nu(\omega_0)^2. \]

This shows that \( [w, \nu]_{HC} \gtrsim \nu(\omega_0) \cdot NM = M \).

### 10.5. Carleson constant is small

Next we will verify that the Carleson condition (8.2) holds with a small constant. We may remove from the sum on the left-hand side of (8.2) all rectangles \( Q \notin \mathcal{R} \). Then we can replace \( E \) by the union of remaining \( Q \)'s without changing the left-hand side and decreasing the right-hand side. Hence we may reduce to the case when \( E \) is a union of members of \( \mathcal{R} \). By (10.9) it follows that for each \( j \) we have either \( Q_j \subseteq E \) or \( Q_j^{++} \cap E = \emptyset \). Let \( \mathcal{J} := \{ j : Q_j \subseteq E \} \). Then we obtain
\[ LHS(8.2) \leq \sum_{[m,m+k] \subseteq \mathcal{J}} \sum_{Q \in C^{[m,m+k]}} ((k+1)/N + 1/N)^2. \]  
(10.18)

Using (10.10) this implies
\[ LHS(8.2) \lesssim \sum_{[m,m+k] \subseteq \mathcal{J}} 2^{-k}N(k+2)^2(1/N)^2 \lesssim (\#J)/N \leq \mu(E) \leq \nu(E), \]
so that \( [w, \nu]_{Car} \lesssim 1 \).
Remark 10.19. Let \( w = \sum_{j=1}^{M} w_j \) be the sum of \( M \) weights each of which is of the tensor product form (1.15). By subadditivity of the Hereditary Carleson constant in the first argument, Theorem 1.14, and monotonicity of the Carleson constant in the first argument, we have

\[
[w, \nu]_{HC} \leq \sum_{j=1}^{M} [w_j, \nu]_{HC} \lesssim \sum_{j=1}^{M} [w_j, \nu]_{C} \leq \sum_{j=1}^{M} [w, \nu]_{C} = M[w, \nu]_{C}.
\] (10.20)

A modification of the above example shows that the constant \( M \) in this inequality is optimal. Namely, let

\[
\mathcal{R}_j := \{ R : Q_j \subset R \}, \quad j = 1 \ldots M,
\]

\[
w_Q := \sum_{j=1}^{M} \mathbf{1}_{\mathcal{R}_j}(Q).
\] (10.21)

With this new \( w \) the above proof of the bound \( [w, \nu]_{Car} \lesssim 1 \) remains valid. Indeed, the only change is that additional factors \((k + 1)\) appear in (10.11) and (10.18). The proof of the lower bound \( [w, \nu]_{HC} \gtrsim M \) remains unchanged.

The example in Section 10.2 also shows that the growth rate \( M \) in (10.20) cannot be improved.

11. REC does not imply embedding

In this section we emulate the previous construction, we start with \( \{Q_j\} \) and measure \( \mu \) but instead of adding \( \omega_0 \) we will add a more sophisticated piece of measure.

We define \( Q_j, \mu, \mathcal{R}, w \) as in the previous section. We continue with denoting

\[
Q_{0,j} := Q_j, \quad \mu_0 := \mu \text{ from the previous section}.
\]

Next we continue with defining a sequence of collections \( Q_k, k = 0, \ldots, K \approx \log M \) of dyadic rectangles as follows

\[
Q_k := \left\{ Q_{k,j} := \bigcap_{i=j}^{j+2^k-1} Q_{0,i}, \; j = 1, \ldots, M - 2^k \right\}, \; k = 1, \ldots, K.
\] (11.1)
In other words, \( Q_k \) consists of the intersections of \( 2^k \) consecutive elements of the basic collection \( Q_0 \). The total amount of rectangles in \( Q_k \) is denoted by \( M_k = M - 2^k + 1 \).

For \( k = 1, \ldots, K \) let
\[
\mu_k(\omega) := \frac{2^{-2k}}{N} \sum_{j=1}^{M_k} \frac{1}{|Q_{k,j}^+|} 1_{Q_{k,j}^+}(\omega), \quad \omega \in (\partial T)^2,
\]
and define
\[
\mu := \mu_0 + \sum_{k=1}^{K} \mu_k.
\]

11.1. **Embedding constant is large.** By duality the inequality (1.11) is equivalent to the Carleson embedding inequality
\[
\int (\mathbb{I}(fw))^2 \, d\mu \leq [w, \mu]_{CE} \sum f^2 \cdot w.
\] (11.2)

We test the inequality (11.2) with the function
\[
f(R) := \mu_0(R) = \mathbb{I}^* \mu_0(R).
\]
Using (10.4) we obtain
\[
\sum f^2 \cdot w = \int \forall \mu_0 \, d\mu_0 \lesssim \|\mu_0\| = \frac{M}{N}.
\] (11.3)

On the other hand, by definition (11.1) and replacing \( M \) by \( 2^k \) in (10.13) we obtain
\[
\forall \mu_0(Q_{k,j}) \gtrsim 2^k N \cdot \frac{1}{N} = 2^k.
\] (11.4)

It follows that
\[
\int (\mathbb{I}(fw))^2 \, d\mu = \int (\forall \mu_0)^2 \, d\mu = \sum_{k=1}^{K} \int (\forall \mu_0)^2 \, d\mu_k \gtrsim \sum_{k=1}^{K} 2^k \|\mu_k\| \sim \frac{M}{N} \log M.
\] (11.5)

Substituting (11.3) and (11.5) in (11.2) we obtain \([w, \mu]_{CE} \gtrsim \log M\).

11.2. **REC constant is small.** We claim that \([w, \mu]_{HC} \lesssim 1\). This means that for any collection \( A \) of dyadic rectangles, setting \( A := \bigcup_{R \in A} R \), we have
\[
\mathcal{E}[\mu|A] \lesssim \mu(A).
\] (11.6)

To show (11.6) let \( \nu_k := \mu_k|A, k = 0, \ldots, K \). Then
\[
\mathcal{E}[\mu|A] = \sum_{n,k} \int \forall \nu_n \nu_k \leq 2 \sum_{n \geq k} \int \forall \nu_n \nu_k \leq 2 \sum_{n \geq k} \int \forall \nu_n \nu_k.
\]
Since \( \text{supp} \nu_k \subseteq \text{supp} \mu_k \) it suffices to show
\[
\sum_{n \geq k} \forall \nu_n \lesssim 1 \quad \text{on} \quad \text{supp} \mu_k.
\] (11.7)

The claim (11.7) has the advantage that it does not depend on \( A \) any more.

For every \( R \in \mathcal{R} \) we have
\[
\mu_n(R) = 2^{-2n} \# \{Q_{n,j} \subseteq R\} \leq 2^{-2n}(\# \{Q_{0,j} \subseteq R\} + 2^n)
\leq 2^{-n}(\# \{Q_{0,j} \subseteq R\} + 1) \leq 2 \cdot 2^{-n} \mu_0(R).
\]

It follows that
\[
\forall \mu_0(Q_{k,j}) \lesssim 2^{-n} \forall \mu_0(Q_{k,j}) \leq 2^{-n} \sum_{i=j}^{j+2^k-1} \forall \mu_0(Q_{0,i}) \lesssim 2^{k-n},
\]
where the last inequality follows from (10.4). This implies (11.7) and therefore (11.6).
Remark 11.8. We can repeat the change of weight in the manner we did in Remark 10.19 in order to have

\[ w = \sum_{i=1}^{M} \tau_i \otimes \eta_i. \]

The discrepancy between REC constant and embedding constant is at least of the order \( \log M \).

Remark 11.9. Observe also that in the examples of Sections 10 and 11 the weight \( w \) was supported only on the predecessors of one fixed point on the boundary of the bi-tree. This means that the weighted bi-tree is essentially a weighted subset of \( \mathbb{Z}^2 \), which puts us in the context of [Saw85]. As we mentioned before, see Theorem 7.14, Sawyer showed that in this case the embedding is equivalent to three single box conditions, in particular any combination of two of them does not imply the embedding anymore. Here we supplemented his results by showing that even the dual multiple box test is not enough to get any of the relations in Theorem 1.14. In other words, the analog of Chang–Fefferman type condition (we called it Carleson condition) and even stronger REC (= hereditary Carleson condition) condition is not enough to ensure embedding even in the simplest case when weight \( w \) was supported only on the predecessors of one fixed point on the boundary of the bi-tree.

**Appendix A. Maximal function versus embedding**

The definition of the Hardy operator (1.5) and its adjoint (1.6) makes sense on an arbitrary finite partially ordered set \( P \). Given a weight \( \mu \) on \( P \) we define the corresponding maximal operator by

\[ \mathcal{M}_\mu \psi (\omega) := \sup_{\alpha \geq \omega} |\psi(\alpha)| \mu(\alpha), \quad (\psi)_\mu (\alpha) := \frac{I^*(\mu)(\alpha)}{I^*(\psi)(\alpha)}, \quad (A.1) \]

with the convention \( 0/0 = 0 \). This definition recovers the usual dyadic maximal operator on the tree and the bi-parameter maximal operator on the bi-tree. We can also define Carleson (1.9) and Carleson embedding (1.11) constants for pairs of weights on an arbitrary partially ordered set. A classical argument [Ste70, p. 236] gives the following relation between these constants.

**Proposition A.2.** Let \( \mu \) be a measure on a partially ordered set \( P \). Then

\[ \sup_{w:[w,\mu] \leq 1} \|w,\mu\|_{CE} = \|\mathcal{M}_\mu\|_{L^2(\mu)} \rightarrow L^2(\mu). \quad (A.3) \]

**Example A.4.** If \( P = T \) is a usual tree, then the maximal function \( \mathcal{M}_\mu \) is essentially the martingale maximal function, and it is well-known that it is bounded on \( L^p(\mu) \) with norm at most \( p' \). In particular the right-hand side of (A.3) equals 4. This is the sharp constant in the Carleson embedding theorem on the tree [NTV99].

**Example A.5.** If \( P = T^2 = T_x \times T_y \) is a bi-tree and \( \mu = \mu_x \times \mu_y \) is a product measure, then the two-parameter maximal operator (A.1) can be majorized by the composition of two one-parameter maximal operators

\[ \mathcal{M}_\mu \psi \leq \mathcal{M}_{x,\mu_x} \mathcal{M}_{y,\mu_y} \psi, \]

which are also defined by (A.1) but on the trees \( T_x \) and \( T_y \). Using \( L^2 \) bounds for the one-parameter maximal operators we see that the right-hand side of (A.3) is bounded by 16. Hence for product measures \( \mu \) and arbitrary weights \( w \) Proposition A.2 gives the implication (1.3) \( \implies \) (1.1).
Proof of Proposition A.2. We begin with the inequality $\leq$ in (A.3). Let $\psi : P \to [0, \infty)$ be a non-negative function. Then
\[
\sum_\alpha w(\alpha) \|\psi(\mu)(\alpha)\|^2 = \sum_\alpha w(\alpha) \|\psi(\mu)(\alpha)\|^2 \int_0^{(\psi)_\mu(\alpha)} 2s \, ds \\
= \int_0^\infty 2s \sum_{\alpha: (\psi)_\mu(\alpha) > s} w(\alpha) \|\psi(\mu)(\alpha)\|^2 \, ds \\
\leq \int_0^\infty 2s \sum_{\alpha: M_\mu \psi(\alpha) > s} w(\alpha) \|\psi(\mu)(\alpha)\|^2 \, ds \\
\leq [w, \mu]_C \int_0^\infty 2s \mu(\alpha > s) \, ds \\
= [w, \mu]_C \|M_\mu \psi\|^2_{L^2(\mu)} \\
\leq [w, \mu]_C \|M_\mu \psi\|^2_{L^2(\mu)} \|\psi\|^2_{L^2(\mu)}.
\]
Here we have used that the superlevel sets $\{\alpha : M_\mu \psi(\alpha) > s\}$ are down-sets. Taking the supremum over $\psi$ and $w$ we obtain the inequality $\leq$ in (A.3).

Now we will show the inequality $\geq$ in (A.3). Let $\psi : P \to [0, \infty)$ be a non-negative function such that $\|\psi\|_{L^2(\mu)} > 0$. For each $\alpha \in P$ with $(\psi)_\mu(\alpha) \neq 0$ let
\[
A'(\alpha) := \{\omega \leq \alpha : M_\mu \psi(\omega) = (\psi)_\mu(\alpha)\},
\]
and let $A'(\alpha) := \emptyset$ otherwise. Enumerate $P = \{\alpha_1, \alpha_2, \ldots\}$ and set
\[
A(\alpha_j) := A'(\alpha_j) \setminus \bigcup_{j' < j} A'(\alpha_{j'}). \]

Then
\[
\sum_{\beta \in P} (M_\mu \psi)^2(\beta) \mu(\beta) = \sum_{\alpha} \sum_{\beta \in A(\alpha)} (\psi)_\mu(\alpha)^2 \mu(\beta) \\
= \sum_{\alpha} w(\alpha) (\|\psi(\mu)(\alpha)\|^2,
\]
where
\[
w(\alpha) := (\|\psi(\mu)(\alpha)\|^2)^{-1} \sum_{\beta \in A(\alpha)} \mu(\beta)
\]
with the convention $w(\alpha) = 0$ if $A(\alpha) = \emptyset$ (where we might be dividing by zero in the above formula). Since the sets $A(\alpha)$ are disjoint and consist of elements smaller than $\alpha$, for every down-set $D \subseteq P$ we have
\[
\sum_{\alpha \in D} w(\alpha) (\|\psi(\mu)(\alpha)\|^2 \leq \sum_{\alpha \in D} \sum_{\beta \in A(\alpha)} \mu(\beta) \leq \sum_{\alpha \in D} \mu(\alpha),
\]
so $[w, \mu]_C \leq 1$. On the other hand, by the above calculation
\[
[w, \mu]_C \geq \frac{\|M_\mu \psi\|^2_{L^2(\mu)}}{\|\psi\|^2_{L^2(\mu)}}.
\]
By first taking the supremum over $w$ with $[w, \mu]_C \leq 1$ and then over $\psi$ we obtain the inequality $\geq$ in (A.3).

\[
\Box
\]

A.1. A sparse proof. In this section we give an alternative proof of a special case of the inequality $\leq$ in Proposition A.2 using an argument going back to [Ver96] and using [Dor75].

Alternative proof of $\leq$ in Proposition A.2. Let $X$ be a standard Borel space with an atom-free measure $\mu$. Suppose that $P$ is a finite collection of Borel subsets of $X$ ordered by inclusion, which we denote by $\leq$, $\geq$. Assume that each $Q \in P$ is the disjoint union of minimal elements of $P$ contained in it (as is the case in the dyadic bi-tree of finite depth). Let $\tilde{\mu}(\omega) := \mu(\omega)$ for $\omega$ minimal and 0 otherwise. We prove
the inequality for \( \mu \) as we identify \( \mu \) with \( \tilde{\mu} \). Let \( w \) be a sequence with \([w, \tilde{\mu}]_{CE} = 1\). This implies for all \( \Omega \) which are union of elements of \( \mathcal{P} \)
\[
\sum_{Q \subseteq \Omega} w(Q) \mu(Q)^2 \leq \mu(\Omega). \tag{A.6}
\]
as \( \tilde{\mu}(\Omega) = \mu(\Omega) \) for such \( \Omega \).

By [Hän18, Theorem 1.3] (see also [Bar19] for a geometric proof) this implies the existence of pairwise disjoint measurable subsets \( E_Q \subset Q, Q \in \mathcal{P}, \) such that
\[
w(Q) \cdot \mu(Q)^2 \leq \mu(E_Q). \tag{A.7}
\]
Let an arbitrary positive function \( \psi \) on \( \mathcal{P} \). We define a new function \( \tilde{\psi} \) on \( \bigcup \omega \in \mathcal{P} \) by
\[
\tilde{\psi}(x) := \psi(\omega(x)) \text{ where } \omega(x) \text{ is the unique minimal element of } \mathcal{P} \text{ containing } x.
\]
We can easily deduce that \( \mathcal{M}_{\tilde{\mu}} \psi(\omega(x)) = \mathcal{M}'_{\mu} \tilde{\psi}(x) \) for \( \mathcal{M}'_{\mu} \tilde{\psi}(x) := \sup_{Q \ni x} \frac{1}{\mu(Q)} \int_Q \tilde{\psi} \, d\mu \).

From this, we also get \( \|\mathcal{M}_{\tilde{\mu}} \psi\|_{L^2(\mathcal{P}, \tilde{\mu})} = \|\mathcal{M}'_{\mu} \tilde{\psi}\|_{L^2(\bigcup \omega, \mu)} \). Finally, recall that \( \|\tilde{\mu}(Q) = \mu(Q) \). Then,
\[
\sum_{Q \in \mathcal{P}} (\tilde{\mu}(\tilde{\mu})(Q))^2 w(Q) = \sum_{Q \in \mathcal{P}} (\psi)^2 w(Q) \leq \sum_{Q \in \mathcal{P}} (\psi)^2 \cdot \mu(E_Q)
\]
\[
\leq \sum_{Q \in \mathcal{P}} (\inf_{\omega \subseteq Q} \mathcal{M}_{\tilde{\mu}} \psi(\omega))^2 \cdot \mu(E_Q)
\]
\[
\leq \int_{E_Q} \left( \mathcal{M}'_{\mu} \tilde{\psi} \right)^2 \, d\mu
\]
\[
\leq \int_{\bigcup \omega} \left( \mathcal{M}'_{\mu} \tilde{\psi} \right)^2 \, d\mu
\]
\[
= \int_{\mathcal{P}} \left( \mathcal{M}_{\tilde{\mu}} \psi \right)^2 \, d\tilde{\mu}
\]
\[
\leq \|\mathcal{M}_{\tilde{\mu}}\|^2_{L^2(\tilde{\mu}) \to L^2(\tilde{\mu})} \|\psi\|^2_{L^2(\tilde{\mu})}.
\]
Taking the supremum over \( \psi \) we obtain \([w, \tilde{\mu}]_{CE} \leq \|\mathcal{M}_{\tilde{\mu}}\|^2_{L^2(\tilde{\mu}) \to L^2(\tilde{\mu})} \). \( \square \)

**References**

[AH96] D. R. Adams and L. I. Hedberg. *Function spaces and potential theory*. Vol. 314. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1996. pp. xii+366. MR: 1414441 (cit. on p. 10).

[AHMV18a] N. Arcozzi, I. Holmes, P. Mozolyako, and A. Volberg. “Bellman function sitting on a tree”. Preprint. Sept. 2018. arXiv: 1809.03397 [math.CA] (cit. on p. 3).

[AHMV18b] N. Arcozzi, I. Holmes, P. Mozolyako, and A. Volberg. “Bi-parameter embedding and measures with restriction energy condition”. Preprint. Nov. 2018. arXiv: 1811.00978v6 [math.CA] (cit. on pp. 2–4, 11, 18).

[AMP18] N. Arcozzi, P. Mozolyako, K.-M. Perfekt, and G. Sarfatti. “Bi-parameter Potential theory and Carleson measures for the Dirichlet space on the bidisc”. Preprint. Nov. 2018. arXiv: 1811.04990 [math.CV] (cit. on pp. 2–4, 18).

[Bar19] A. Barron. “Sparse bounds in harmonic analysis and semiperiodic estimates”. PhD thesis. Brown University, 2019 (cit. on p. 23).

[BP17] A. Barron and J. Pipher. “Sparse domination for bi-parameter operators using square functions”. Preprint. Mar. 2017. arXiv: 1709.05009 [math.CA] (cit. on p. 2).

[Car74] L. Carleson. *A counterexample for measures bounded on \( H^p \) for the bi-disc*. Tech. rep. 7. Mittag–Leffler, 1974 (cit. on pp. 1, 4, 15).
[CF80] S.-Y. A. Chang and R. Fefferman. “A continuous version of duality of $H^{sp}1$ with BMO on the bidisc”. In: Ann. of Math. (2) 112.1 (1980), pp. 179–201. MR: 584078 (cit. on p. 2).

[Cha79] S.-Y. A. Chang. “Carleson measure on the bi-disc”. In: Ann. of Math. (2) 109.3 (1979), pp. 613–620. MR: 534766 (cit. on p. 1).

[Dor75] L. E. Dor. “On projections in $L_1$”. In: Ann. of Math. (2) 102.3 (1975), pp. 463–474. MR: 0420244 (cit. on p. 22).

[Fef81] R. Fefferman. “Strong differentiation with respect to measures”. In: Amer. J. Math. 103.1 (1981), pp. 33–40. MR: 601461 (cit. on p. 2).

[Fef86] R. Fefferman. “Calderón-Zygmund theory for product domains: $H^p$ spaces”. In: Proc. Nat. Acad. Sci. U.S.A. 83.4 (1986), pp. 840–843. MR: 826217 (cit. on pp. 2, 4).

[Fef88] R. Fefferman. “Some recent developments in Fourier analysis and $H^p$ theory on product domains. II”. In: Function spaces and applications (Lund, 1986). Vol. 1302. Lecture Notes in Math. Springer, Berlin, 1988, pp. 44–51. MR: 942256 (cit. on p. 2).

[Hän18] T. S. Hänninen. “Equivalence of sparse and Carleson coefficients for general sets”. In: Ark. Mat. 56.2 (2018), pp. 333–339. arXiv: 1709.10457 [math.CA]. MR: 3893778 (cit. on p. 23).

[HPV19] I. Holmes, G. Psaromiligkos, and A. Volberg. “A comparison of box and Carleson conditions on bi-trees”. Preprint. Mar. 2019. arXiv: 1903.02478 [math.CA] (cit. on p. 19).

[Ios+19] A. Iosevich, B. Krause, E. Sawyer, K. Taylor, and I. Uriarte-Tuero. “Maximal operators: scales, curvature and the fractal dimension”. In: Anal. Math. 45.1 (2019), pp. 63–86. MR: 3916130 (cit. on p. 2).

[Jou88] J.-L. Journé. “Two problems of Calderón-Zygmund theory on product-spaces”. In: Ann. Inst. Fourier (Grenoble) 38.1 (1988), pp. 111–132. MR: 949001 (cit. on pp. 2, 4).

[LSSU14] M. T. Lacey, E. T. Sawyer, C.-Y. Shen, and I. Uriarte-Tuero. “Two-weight inequality for the Hilbert transform: a real variable characterization, I”. In: Duke Math. J. 163.15 (2014), pp. 2795–2820. arXiv: 1201.4319 [math.CA]. MR: 3285857 (cit. on p. 2).

[NTV07] F. Nazarov, S. Treil, and A. Volberg. “Two weight inequalities for individual Haar multipliers and other well localized operators”. In: Math. Res. Lett. 15.3 (2007), pp. 583–597. arXiv: math/0702758. MR: 2407233 (cit. on p. 2).

[NTV99] F. Nazarov, S. Treil, and A. Volberg, “The Bellman functions and two-weight inequalities for Haar multipliers”. In: J. Amer. Math. Soc. 12.4 (1999), pp. 909–928. MR: 1685781 (cit. on pp. 3, 21).

[Pip86] J. Pipher. “Journé’s covering lemma and its extension to higher dimensions”. In: Duke Math. J. 53.3 (1986), pp. 683–690. MR: 860666 (cit. on p. 2).

[Ste70] E. M. Stein. Singular integrals and differentiability properties of functions. Princeton Mathematical Series, No. 30. Princeton University Press, Princeton, N.J., 1970, pp. xiv+290. MR: 0290095 (cit. on p. 21).

[Tao] T. Tao. “Dyadic product $H^p$, BMO, and Carleson’s counterexample”. Short story (cit. on pp. 1, 4).

[Ver96] I. E. Verbitsky. “Imbedding and multiplier theorems for discrete Littlewood-Paley spaces”. In: Pacific J. Math. 176.2 (1996), pp. 529–556. MR: 1435004 (cit. on p. 22).