Rev. Mat. Iberoam. 38 (2022), no. 7, 2069–2116
DOI 10.4171/RMI/1378

Carleson embedding on the tri-tree and on the tri-disc

Pavel Mozolyako, Georgios Psaromiligkos, Alexander Volberg and Pavel Zorin-Kranich

Abstract. We prove a multi-parameter dyadic embedding theorem for the Hardy operator on the multi-tree. We also show that for a large class of Dirichlet spaces in the bi-disc and the tri-disc, this proves the embedding theorem of those Dirichlet spaces of holomorphic functions on the bi- and tri-disc. We completely describe the Carleson measures for such embeddings. The result below generalizes the embedding result of Arcozzi et al. from the bi-tree to the tri-tree and from the Carleson–Chang condition to the Carleson box condition. One of our embedding descriptions is similar to the Carleson–Chang–Fefferman condition, and involves dyadic open sets. On the other hand, the unusual feature is that the embedding on the bi-tree and the tri-tree turns out to be equivalent to the one box Carleson condition. This is in striking difference to works of Chang–Fefferman and the well-known Carleson quilt counterexample. Finally, we explain the obstacle that prevents us from proving our results on poly-discs of dimension four and higher.

To Antonio Córdoba, who was an inspiration for multi-parameter research of ours, including this paper.

1. Introduction and the main result

The present article treats a two weight problem on multi-parameter paraproduct operators. Singular bi-parameter and multi-parameter operators enjoyed and continue to enjoy much attention, see [6, 12–14, 19, 20, 31]. They are notoriously difficult. Two weight problems for singular integrals were studied in a series of papers by Nazarov, Treil, and Volberg for dyadic singular operators, and in a series of papers by Lacey, Shen, Sawyer, and Uriarte-Tuero for the Hilbert transform, see [22, 23, 29, 30], and the references therein. Another example is a recent paper by Iosevich, Krause, Sawyer, Taylor, and Uriarte-Tuero [18] on the two weight problem for the spherical maximal operator motivated by Falconer's distance set problem.

Classically, an estimate of paraproduct tri-linear forms [16] is based on the $T1$ theorem of David and Journé. The theory of Carleson measures (or classical BMO theory)

2020 Mathematics Subject Classification: Primary 42B99; Secondary 47A99.
Keywords: Carleson embedding on dyadic tree, multi-parameter Carleson embedding.

is involved. It is well known [9, 10, 19, 20] that in the multi-parameter setting all these results and concepts of Carleson measure, BMO, John–Nirenberg inequality, Calderón–Zygmund decomposition are much more delicate. The paper [27] develops a completely new approach to prove natural tri-linear bi-parameter estimates on bi-parameter paraproducts, especially outside of the Banach range. In [27], Journé's lemma [19] was used, but the approach did not generalize to multi-parameter paraproduct forms. This issue was resolved in [28], where a simplified method was used to address the multi-parameter paraproducts.

We consider here bi-parameter and tri-parameter paraproducts and reveal the obstacle to treat the dimension 4 objects. Our paraproducts are only dyadic ones, and we estimate them only in L^2. But we consider a two weight problem. One weight is arbitrary and the other one is dictated by the problem from complex analysis in the poly-disc (our original motivation). This other weight has the product structure because of this original motivation. We are able to give the necessary and sufficient condition for the two weight boundedness of such multi-parameter paraproducts in the two and three parameter cases (and of course in the one parameter case).

Three remarks are in order: a) the general two weight problem even for two parameter paraproducts seems to not having a simple necessary and sufficient criterion at all (unlike a one parameter case of dyadic paraproducts, whose solution is basically due to Eric Sawyer); so it is a "miracle" that the full solution exists when one measure is arbitrary, and another one has a product structure; b) this solution continues to amaze us because it seemingly goes against a famous Carleson counterexample in the theory of Chang–Fefferman product BMO; c) it is also amazing that the problem about holomorphic functions in the poly-disc can be reduced to dyadic problems having nothing to do with complex analysis; the information – in many cases – is not getting lost.

1.1. Background. A geometric problem and two weight estimates

To wet the appetite, consider first the following very simple geometric problem. We are given a collection of non-negative numbers $\{\alpha_I\}_{I\in\mathcal{D}(I_0)}$ enumerated by the family $\mathcal{D}$ of dyadic subintervals of unit interval $I_0 = [0, 1]$. We wish to find an assignment $I \to E_I$, $I \in \mathcal{D}$, of measurable sets in such a way that

(1) the sets E_I are pairwise disjoint;

(2) $m(E_I) = \alpha_I$.

There is an obvious necessary condition:

$$\forall J \in \mathcal{D}(I_0), \quad \sum_{I\in\mathcal{D}(J)} \alpha_I \le m(J)\,. \tag{1.1}$$

A simple and very well-known elementary construction shows that (1.1) is not only necessary but also sufficient. Moreover, such a condition (called the Carleson packing condition with constant $C = 1$) is necessary and sufficient if I_0 is a unit cube in $\mathbb{R}^d$ rather than a unit interval, and when $\mathcal{D}$ means the collection of all dyadic sub-cubes of the unit cube. The Lebesgue measure m can be replaced here by any finite Borel measure without point masses.

Now let us make the problem harder. We just mentioned that replacing dyadic intervals dyadic cubes $\mathcal{D}$ represents a simple problem. But what if we augment the collection of sets? It is very natural and useful, see e.g. [6], to consider the collection of dyadic rectangles $\mathcal{D}^k = \mathcal{D} \times \cdots \times \mathcal{D}$ (k times), $k \geq 2$. It is much harder to prove that the condition

$$\forall \mathcal{S} \subset \mathcal{D}^2, \quad \sum_{I \times J \in \mathcal{S}} \alpha_{I \times J} \leq \mu\big(\bigcup_{I \times J \in \mathcal{S}} I \times J\big) \tag{1.2}$$

for a finite Borel measure without point masses, is sufficient for the existence of the assignment $I \times J \to E_{I \times J}$, for all dyadic rectangles $I \times J$ of measurable sets, in such a way that

(1) the sets $E_{I \times J}$ are pairwise disjoint;

(2) $\mu(E_{I \times J}) = \alpha_{I \times J}$.

Obviously, (1.2) is necessary for the existence of such a measurable assignment. However, several non-trivial proofs exist. The methods range from geometric ones (see Barron–Pipher [6]), to convex analysis/functional analysis (see Hänninen [17], using a result of Dor [11]). Here $k = 2$, but this is not essential, the same result holds for dyadic rectangles in all dimensions.

Moreover, Hänninen [17] proved that dyadic rectangles can be replaced by arbitrary collections of Borel sets.

Definition 1.1. (Carleson coefficients in the generality of a collection of Borel sets). Let μ be a locally finite Borel measure on $\mathbb{R}^d$. Let $\mathcal{S}$ be a countable collection of Borel sets. A family $\{\alpha_S\}_{S \in \mathcal{S}}$ of non-negative reals is Carleson (with constant $C = 1$) if we have

$$\sum_{S \in \mathcal{S}, S \subset \Omega} \alpha_S \leq \mu(\Omega)$$

for every union Ω of sets of the collection $\mathcal{S}$.

Hänninen proved that the disjoint measurable assignment exists if and only if the sequence $\{\alpha_S\}_{s \in \mathcal{S}}$ satisfies this general Carleson packing condition, that can be also written as

$$\forall \mathcal{S}' \subset \mathcal{S}, \quad \sum_{S \in \mathcal{S}'} \alpha_S \leq \mu\big(\bigcup_{S \in \mathcal{S}'} S\big).$$

Now we would like to indicate the connection of the above mentioned "combinatorial" problems to two-weight embedding theorems that have another (equivalent) disguise: two weight paraproduct estimates.

We start again with the simple 1-dimensional dyadic case. We fix the canonical bijection between intervals of $\mathcal{D}(I_0)$ and a dyadic tree T, whose vertices we will still call I, and where $I_0 = [0, 1]$ is the root of T. We fix a measure μ on I_0; it will be one of our two weights. The second weight lives on T and it is just a sequence of non-negative numbers enumerated by vertices (= dyadic intervals): $w := \{w_I\}_{I \in T}$.

The two-weighted problem is to find necessary and sufficient conditions on (w, μ) to have

$$\sum_{I \in T} w_I \cdot \Big(\int_I f \, d\mu\Big)^2 \leq C \int f^2 \, d\mu. \tag{1.3}$$

Let us show what this has in common with the previously considered geometric question. There is an obvious necessary condition for (1.3) to hold: just plug $f = \mathbf{1}_J$, $J \in \mathcal{D}$, to obtain

$$\forall J \in \mathcal{D}, \quad \sum_{I \in \mathcal{D}(J)} w_I \cdot \mu(I)^2 \le C\,\mu(J)\,. \tag{1.4}$$

We can now use the assignment mentioned above for $\alpha_I := w_I \mu(I)^2/C$. We will get a disjoint family $\{E_I\}_{I\in\mathcal{D}}$. As a next step, one can use that the dyadic maximal function with respect to any μ is bounded in $L^2(I_0, \mu)$. This will finish the proof that (1.4) is also sufficient for the embedding (1.3). The fact that (1.4) is necessary and sufficient for the embedding (1.3) is called the Carleson–Sawyer theorem. Carleson proved it in the 60's, and used it in his interpolation and famous corona results. Sawyer's generalization appeared in the 80's. Both results are fundamental in the dyadic approach to the theory of Calderón–Zygmund operators.

Two (or multi) parameter paraproducts require a solution of a much more involved two-weight problem. We fix a measure μ on $[0, 1]^2$; it will be one of our two weights. The second weight lives on T^2 and it is just a sequence of non-negative numbers enumerated by vertices (= dyadic rectangles): $w := \{w_{I\times J}\}_{I,J\in T}$.

The two-weighted problem is to find necessary and sufficient conditions on (w, μ) to have

$$\sum_{I,J\in T} w_{I\times J} \cdot \Big(\int_{I\times J} f\, d\mu\Big)^2 \le C \int_{[0,1]^2} f^2\, d\mu. \tag{1.5}$$

A bi-tree T^2 is a rooted graph with vertices being dyadic rectangles, and the root being $I_0 \times I_0 = [0, 1]^2$. It is a much more complicated graph than the simple T, in particular, it has cycles. However, again there are simple necessary conditions for (1.5). We get one by plugging $f = \mathbf{1}_{I_1\times J_1}$, $I_1, J_1 \in \mathcal{D}$. But Carleson gave an example of weight w on T^2 such that even with $\mu = m_2$, the Lebesgue measure on the plane, this necessary condition is *not* sufficient. But there is a stronger necessary condition. Choose now $f = \mathbf{1}_{\cup_{k=1}^\infty I_k\times J_k}$. In other words, choose a subset $S' \subset \mathcal{D}(I_0) \times \mathcal{D}(I_0)$, consider $\Omega = \cup_{R'\in S'} R'$, and choose $f = \mathbf{1}_\Omega$ to plug into (1.5). Then we immediately and trivially get the following necessary condition for the embedding (1.5) (called the Carleson–Chang packing condition):

$$\forall S' \subset \mathcal{D}(I_0) \times \mathcal{D}(I_0), \quad \sum_{R\subset \cup_{R'\in S'} R'} w_R \cdot (\mu(R))^2 \le C\,\mu\big(\textstyle\bigcup_{R'\in S'} R'\big)\,. \tag{1.6}$$

Again, the assignment of disjoint E_R, $R \in \mathcal{D}(I_0) \times \mathcal{D}(I_0)$, is the first step, but the second step breaks down: the strong maximal (even dyadic strong maximal) operator with respect to μ is rarely bounded in $L^2(\mu)$. But maybe one does not need maximal operators to prove the embedding as above?

This is what we know about (1.5) and its analogs for the tri-tree and higher multi-trees.

(1) A. S.-Y. Chang proved that if $\mu = m_2$ (or $\mu = m_d$) then the necessary condition (1.6) is sufficient, and this holds for any w on T^2 (and correspondingly T^d).

(2) For any μ such that the strong dyadic maximal function is bounded in $L^2(\mu)$, (1.6) is sufficient, and this holds for any w on T^2 (and correspondingly T^d if we consider the measure μ on $[0,1]^d$).

(3) Moreover, if (1.6) is sufficient for the embedding (1.5) with arbitrary w (maybe with a different constant), then μ is such that the strong dyadic maximal function is bounded in $L^2(\mu)$. This holds in any dimension d.

(4) There exists w such that (1.6) does *not* hold, but the following simplified version does hold:

$$\forall I_1 \times J_1 \in \mathcal{D}(I_0) \times \mathcal{D}(I_0), \qquad \sum_{R \subset I_1 \times J_1} w_R \cdot (\mu(R))^2 \le C\, \mu(I_1 \times J_1)\,.$$

(5) Such an example exists even with $\mu = m_2$ (Carleson [7], Tao [32]).

(6) There exists (w, μ) such that (1.6) does hold, but the following more complicated (but obviously necessary for the embedding (1.5), just plug $f = \mathbf{1}_F$ into (1.5)) condition does *not* hold:

$$\forall F \subset [0,1]^2,\ \forall S' \subset \mathcal{D}(I_0) \times \mathcal{D}(I_0), \qquad \sum_{R \subset \cup_{R' \in S'} R'} w_R \cdot (\mu(R \cap F))^2 \le C\, \mu(F)\,.$$

(7) The latter example has w having only values 1 and 0, and moreover the support of w is a connected subgraph of T^2.

(8) In general, the necessary and sufficient conditions for (1.5) are unknown.

(9) The case $w \equiv 1$ is interesting and has interesting applications to complex analysis.

(10) Whatever is μ, for the case $w \equiv 1$ for T^2 and T^3, we can give simple necessary and sufficient conditions for the embedding (1.5) to hold

(11) We conjecture that the same answer holds for T^d, $d \ge 4$, but we cannot prove this.

(12) Our answer for the case $w \equiv 1$ for T^2 and T^3 is counterintuitive. At the first glance, it seems to contradict Carleson's example. Of course it does not. The answer is that the embedding (1.5) holds if an only if (we give it for $d = 2$, the same answer with obvious changes holds for $d = 3$, and this is the main result of the current article):

$$\forall I_1 \times J_1 \in \mathcal{D}(I_0) \times \mathcal{D}(I_0), \qquad \sum_{R \subset I_1 \times J_1} (\mu(R))^2 \le C_0\, \mu(I_1 \times J_1)\,. \tag{1.7}$$

Of course, the constant C in (1.5) can be calculated by C_0 in (1.7), but it is a non-linear relationship.

1.2. Background. Embedding from $L^2(m_2)$ to $\ell^2(T^2, \{\beta_R^2\})$, where $\{R\}$ are dyadic rectangles

Everywhere below, the angular brackets $\langle \cdot \rangle_S$ mean the average over the set S. The measure, if not indicated otherwise, is the Lebesgue measure.

Lennart Carleson showed in [7] that the natural generalization, 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 [8] 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

$$\Pi_b\varphi := \sum_R \langle\varphi\rangle_R \,(b, h_R)\, h_R\,.$$

The paraproduct Π_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

$$\sum_R \langle\varphi\rangle_R^2\, \beta_R^2 \le C \int \varphi^2\, dm_2, \tag{1.8}$$

where $\beta_R := (b, h_R)$ are the Haar coefficients of the function b. In analogy to the one-parameter Carleson embedding, one could ask whether (1.8) is equivalent to the “box” condition

$$\sum_{R\subseteq R_0} \beta_R^2 \le C' m_2(R_0) \tag{1.9}$$

for every dyadic rectangle $R_0 \subseteq [0,1]^2$. A counterexample showing that (1.9) does not imply (1.8) was constructed by Carleson [7, 32].

It was observed by Chang [8] (in a continuous setting) that (1.8) is equivalent to the *bi-parameter Carleson* (*or Carleson–Chang*) *condition*

$$\sum_{R\subset\Omega} \beta_R^2 \le C' m_2(\Omega)\,,$$

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 [9] to characterize the dual of the Hardy space on the bi-disc $H^1(\mathbb{D}^2)$. The same embedding result holds in dimension $d > 2$, from $L^2(m_d)$ to $\ell^2(T^d, \{\beta_R^2\})$.

1.3. Terminology and notation

We begin with order-theoretic conventions.

Definition 1.2. A *finite tree* T is a finite partially ordered set such that, for every $\omega \in T$, the set $\{\alpha \in T : \alpha \ge \omega\}$ is totally ordered (we allow trees to have several maximal elements).

An *d-tree* T^d is a Cartesian product of d (possibly different) finite trees with the product order.

A subset $\mathcal{U}$ (respectively, $\mathcal{D}$) of a partially ordered set T is called an *up-set* (respectively, *down-set*) if, for every $\alpha \in \mathcal{U}$ and $\beta \in T$ with $\alpha \le \beta$ (respectively, $\beta \le \alpha$), we also have $\beta \in \mathcal{U}$ (respectively, $\beta \in \mathcal{D}$).

The *Hardy operator* on an d-tree T^d is defined by

$$\mathbf{I}\phi(\gamma) := \sum_{\gamma' \geq \gamma} \phi(\gamma') \quad \text{for any } \phi : T^d \to \mathbb{R}.$$

In the one-parameter case $d = 1$, we denote it by I, and in the two-parameter case $d = 2$, by $\mathbb{I}$. The adjoint $\mathbf{I}^*$ of the Hardy operator $\mathbf{I}$ is given by the formula

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

Definition 1.3. Let μ and w be positive functions on T^d. The *box constant* is the smallest number $[w,\mu]_{\mathrm{Box}}$ such that

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

The *Carleson constant* is the smallest number $[w,\mu]_{\mathrm{C}}$ such that

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

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

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

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

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

holds for all functions ψ on T^d.

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 multi-tree or the weights w, μ.

1.4. Main result

The inequalities

$$[w,\mu]_{\mathrm{Box}} \leq [w,\mu]_{\mathrm{C}} \leq [w,\mu]_{\mathrm{HC}} \leq [w,\mu]_{\mathrm{CE}} \tag{1.11}$$

are obvious. The converse inequalities for 1-trees were proved in [29]. For 2-trees, in the case $w \equiv 1$, the converse inequality

$$[1,\mu]_{\mathrm{CE}} \lesssim [1,\mu]_{\mathrm{C}}$$

was proved in [3]. In [4], it was proved that, more generally,

$$[w, \mu]_{\mathrm{CE}} \lesssim [w, \mu]_{\mathrm{Box}}$$

for weights w of tensor product form on 2-trees. In this article, we extend this result to 3-trees.

Theorem 1.4. *Let $\mu: T^3 \to [0, \infty)$. Let $w: T^3 \to [0, \infty)$ be of tensor product form. Then the converses of the inequalities in* (1.11) *also hold*:

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

Theorem 1.4 will follow from conditional results on d-trees, namely Theorem 6.3 and Theorem 7.3.

1.5. The methods

The methods of proving this main result of ours are mostly by potential theory and some combinatorics. But this potential theory is very far from the classical one. It is a potential theory on graphs with cycles, in particular, there will be no maximum principle for the potentials considered below. This is the main difficulty and the main attraction of what follows.

2. Holomorphic function spaces in the poly-disc

Another way to interpret the Hardy inequality (or more precisely, its weighted version, see below) is to consider its connection to certain problems in the theory of Hilbert spaces of analytic functions on the (poly-)disc. It was actually this connections that motivated the study of this inequality in [5] and [3].

We start with some additional notation. Given an integer $d \geq 1$ and $s = (s_1, \ldots, s_d) \in \mathbb{R}^d$, we consider a Hilbert space $\mathcal{H}_s(\mathbb{D}^d)$ of analytic functions on the poly-disc $\mathbb{D}^d$ with the norm

$$\|f\|^2_{\mathcal{H}_s(\mathbb{D}^d)} := \sum_{n_1,\ldots,n_d \geq 0} |\hat{f}(n_1, \ldots, n_d)|^2 \, (n_1 + 1)^{s_1} \cdots (n_d + 1)^{s_d},$$

where

$$f(z) = \sum_{n_1,\ldots,n_d \geq 0} \hat{f}(n_1, \ldots, n_d) \, z_1^{n_1} \cdots z_d^{n_d}, \quad z = (z_1, \ldots, z_d) \in \mathbb{D}^d.$$

Observe that, clearly,

$$\mathcal{H}_{\vec{s}}(\mathbb{D}^d) = \bigotimes_{j=1}^{d} \mathcal{H}_{s_j}(\mathbb{D}).$$

In particular, the choice $s = (0, \ldots, 0)$ gives a classical Hardy space on the poly-disc; on the other hand, $s = (1, \ldots, 1)$ corresponds to the Dirichlet space.

2.1. Embedding (Carleson) measures on the poly-disc

A measure ν on $\mathbb{D}^d$ is called a Carleson measure for $\mathcal{H}_s$ if there is a constant C_ν such that

$$\int_{\mathbb{D}^d} |f(z)|^2\, d\nu(z) \le C_\nu \|f\|^2_{\mathcal{H}_s(\mathbb{D}^d)},$$

or, in other words, the embedding Id: $\mathcal{H}_s(\mathbb{D}^d) \to L^2(\mathbb{D}^d, d\nu)$ is bounded.

For brevity, we concentrate below on the case $d = 2$, indicating the changes necessary for other d. Consider first the case of $s = (1, 1)$.

Given a holomorphic function

$$f(z_1, z_2) = \sum_{m,n\ge 0} a_{mn} z_1^m z_2^n$$

on $\mathbb{D}^2$, we let

$$\|f\|^2_{\mathcal{D}(\mathbb{D}^2)} = \sum_{m,n\ge 0} |a_{mn}|^2 (m+1)(n+1).$$

This norm can also be written as follows:

$$\begin{aligned}\|f\|^2_{\mathcal{D}(\mathbb{D}^2)} &= \frac{1}{\pi^2}\int_{\mathbb{D}^2} |\partial_{z_1,z_2} f(z_1,z_2)|^2\, dz_1\, dz_2 + \frac{1}{2\pi^2}\int_{\mathbb{T}}\int_{\mathbb{D}} |\partial_{z_1} f(z_1, e^{it})|^2\, dz_1\, dt \\ &\quad + \frac{1}{2\pi^2}\int_{\mathbb{D}}\int_{\mathbb{T}} |\partial_{z_2} f(e^{it}, e^{is})|^2\, ds\, dz_2 + \frac{1}{4\pi^2}\int_{\mathbb{T}}\int_{\mathbb{T}} |f(e^{is}, e^{it})|^2\, ds\, dt \\ &= \|f\|_*^2 + \text{ other terms,}\end{aligned}$$

where $\|f\|_*$ is a semi-norm which is invariant under biholomorphisms of the bidisc. In what follows, however, we use an equivalent norm, arising from the representation $\mathcal{D}(\mathbb{D}^2) = \mathcal{D}(\mathbb{D}) \otimes \mathcal{D}(\mathbb{D})$ (this particular choice will be justified in a few lines). For $f \in \mathrm{Hol}(\mathbb{D})$, let

$$\|f\|^2_{\mathcal{D}} := \frac{1}{\pi}\int_{\mathbb{D}} |f'|^2(z)\, dz + C_0 |f(0)|^2, \tag{2.1}$$

where $C_0 > 0$ is a constant to be chosen shortly. It is a classical fact that the Dirichlet space on the unit disc is a reproducing kernel Hilbert space ([5]), and, consequently, $\mathcal{D}(\mathbb{D}^2)$ is one as well. The reproducing kernel K_z, $z \in \mathbb{D}^2$ (generated by $\|\cdot\|_{\mathcal{D}}$), is

$$K_z(w) = \Big(C_1 + \log\frac{1}{1-\bar z_1 w_1}\Big)\Big(C_1 + \log\frac{1}{1-\bar z_2 w_2}\Big), \quad z, w \in \mathbb{D}^2$$

(so it is a product of reproducing kernels for $\mathcal{D}(\mathbb{D})$ in the respective variables), and $C_1 > 0$ is a constant depending on C_0.

The definition of norm in (2.1) implies that K_z enjoys the following important property:

$$\Re K_z(w) \sim |K_z(w)|, \quad z, w \in \mathbb{D}^2,$$

if we take C_1 (respectively, C_0) to be large enough.

Let $\mu, w\colon T^d \to \mathbb{R}_+$. We define a weighted Hardy operator to be

$$\mathbf{I}_w f(\alpha) := \sum_{\beta \geq \alpha} f(\beta)\, w(\beta).$$

We call (μ, w) a trace pair for the weighted Hardy inequality if

$$\int_{T^d} (\mathbf{I}_w f)^2\, d\mu \lesssim \int_{T^d} f^2\, dw \tag{2.2}$$

for any $f\colon T^d \to \mathbb{R}_+$, i.e., the operator $\mathbf{I}_w\colon L^2(T^d, dw) \to L^2(T^d,\, d\mu)$ is bounded. The dual version is

$$\int_{T^d} (\mathbf{I}^*(\varphi\mu))^2\, dw \leq \int_{T^d} \varphi^2\, d\mu \tag{2.3}$$

for any $\varphi\colon T^d \to \mathbb{R}_+$, where

$$\mathbf{I}^*\varphi(\beta) := \sum_{\alpha \leq \beta} \varphi(\alpha).$$

It turns out that trace pairs for the weighted Hardy inequality and Carleson measures for $\mathcal{H}_s$ are closely related. Below we give a brief overview of this relationship. We gloss over most of the technical parts of this short exposition, for more details see [5] and Section 2 of [3], where it was presented for $d = 1$, $s = s_1 \in (0, 1]$ and $d = 2$, $s = 1$, respectively.

We start by assuming that $s \in (0, 1]^d$ (so that $\mathcal{H}_s(\mathbb{D}^d)$ is a *weighted Dirichlet space on the poly-disc*), and that $\operatorname{supp} \nu \subset r\mathbb{D}^d$ for some $r < 1$ (the latter is just a convenience assumption that allows us to make the corresponding graphs to be finite; no estimate below will depend on r, or on the depth of the graph).

It is well known that $\mathcal{H}_{s_j}(\mathbb{D})$, $1 \leq j \leq d$, is a *reproducing kernel Hilbert space* (RKHS) with kernel K_{s_j} satisfying (possibly after a suitable change of norm)

$$\begin{aligned} |K_{s_j}|(z_j, \zeta_j) &\asymp |1 - z_j \bar{\zeta}_j|^{s_j - 1}, \quad 0 < s_j < 1,\\ |K_{s_j}|(z_j, \zeta_j) &\asymp \log |1 - z_j \bar{\zeta}_j|^{-1}, \quad s_j = 1. \end{aligned}$$

Moreover, it is not hard to verify that

$$\Re K_s \asymp |K_s|, 0 < s \leq 1\,.$$

However, the case $s = 0$ is a special case, as

(2.4) the Poisson kernel is not equivalent to the absolute value of the Cauchy kernel.

It follows immediately that $\mathcal{H}_{\vec{s}}(\mathbb{D}^d)$ is a reproducing kernel Hilbert space as well, and

$$K_{\vec{s}}(z, \zeta) = \prod_{j=1}^{d} K_{s_j}(z_j, \zeta_j), \quad z, \zeta \in \mathbb{D}^d.$$

Going back to the Carleson embedding, we see that Id: $\mathcal{H}_{\vec{s}}(\mathbb{D}^d) \to L^2(\mathbb{D}^d, d\nu)$ is bounded if and only if its adjoint Θ is bounded as well. Let us compute its action on a function $g \in L^2(\mathbb{D}^d, d\nu)$:

$$(\Theta g)(z) = \langle \Theta g, K_{\vec{s}}(z,\cdot)\rangle_{\mathcal{H}_{\vec{s}}(\mathbb{D})} = \langle g, K_{\vec{s}}(z,\cdot)\rangle_{L^2(\mathbb{D}^d,d\nu)} = \int_{\mathbb{D}^d} g(\zeta)\, \overline{K_{\vec{s}}(z,\zeta)}\, d\nu(\zeta).$$

Hence, for Θ to be bounded, it must satisfy

$$\|g\|^2_{L^2(\mathbb{D}^d,d\nu)} \gtrsim \|\Theta g\|_{\mathcal{H}_{\vec{s}}(\mathbb{D}^d)} = \langle g, \Theta g\rangle_{L^2(\mathbb{D}^d,d\nu)} = \int_{\mathbb{D}^{2d}} g(z)\, \overline{g(\zeta)}\, K_{\vec{s}}(z,\zeta)\, d\nu(z)\, d\nu(\zeta). \tag{2.5}$$

If inequality (2.5) holds, then trivially the following holds:

$$\|g\|^2_{L^2(\mathbb{D}^d,d\nu)} \gtrsim \int_{\mathbb{D}^{2d}} g(z)\, g(\zeta)\, K_{\vec{s}}(z,\zeta)\, d\nu(z)\, d\nu(\zeta), \quad g \ge 0\,. \tag{2.6}$$

If we would know that the real part of the coordinate reproducing kernel is comparable to its absolute value, we would deduce that Θ is bounded if and only if

$$\int_{\mathbb{D}^{2d}} g(z)\, g(\zeta)\, |K_{\vec{s}}(z,\zeta)|\, d\nu(z)\, d\nu(\zeta) \lesssim \|g\|^2_{L^2(\mathbb{D}^d,d\nu)} \tag{2.7}$$

for any positive g on $\mathbb{D}^d$.

In fact, (2.5) implies (2.6), and we can take the real part of both sides of (2.6), putting the real part on the kernel. Now if we would know that

$$\Re K_{\vec{s}}(z,\zeta) = \Re \prod_{j=1}^{d} K_{s_j}(z_j,\zeta_j) \asymp \Big|\prod_{j=1}^{d} K_{s_j}(z_j,\zeta_j)\Big| = |K_{\vec{s}}(z,\zeta)|, \quad z,\zeta \in \mathbb{D}^d, \tag{2.8}$$

we would deduce (2.5)$\Rightarrow$(2.7). The only thing we need for this implication is the above pointwise equivalence (2.8). On the other hand, the implication (2.7)$\Rightarrow$(2.5) obviously always holds.

We conclude that, in the presence of the pointwise equivalence (2.8), we have that (2.5)$\equiv$(2.7).

However, the equivalence (2.8) – ultimately important for us to prove the equivalence of dyadic and analytic embeddings (see below) – has limitations. First of all, it is false even for the 1D case $d = 1$ if $s = 0$, see (2.4). That makes the case $s = 0$ quite special. It is well known that for the 1D case, the embedding measures for the Poisson and Cauchy kernels on $L^2(\mathbb{T})$ are the same. This is rather simple, but should be considered as a "miracle". Already in the 2D situation, the fact that the embedding measures for the Poisson $P_{z_1} P_{z_2}$ and the Cauchy $K_{\vec{0}}(z,\zeta) = (1 - z_1\bar{\zeta}_1)^{-1}(1 - z_2\bar{\zeta}_2)^{-1}$ kernels on $L^2(\mathbb{T}^2)$ are the same is a subtle fact that will be considered in [25] separately.

Another interesting distinction of the case $s = 0$ is again about (2.4). The reader will see that for $s > 0$, we will characterize the embedding in terms of a simple box (rectangular) test. As it is well known from the works of Chang, Fefferman and Carleson [7, 8, 13, 32], such a characterization is not possible for the Poisson embedding of $L^2(\mathbb{T}^d)$ if $d \geq 2$. We would wish to attribute this phenomena to the fact that the Poisson kernel has a special shape.

Let $q(\alpha)$ be the Whitney rectangle described at the beginning of Section 2.4 below. Let

$$P(\alpha, \beta) := \sup_{z \in q(\alpha), \zeta \in q(\beta)} P(z, \zeta)\,.$$

See the definition of $w_{\vec{0}}$ below in (2.13).

In the language of the tree T^d, the fact that the Poisson kernel has a special shape means that the following inequality is false in general:

$$\mathbf{I}_{w_{\vec{0}}}\mathbf{1}(\alpha \vee \beta) \lesssim P(\alpha, \beta),$$

This finishes the discussion of $\vec{s} = \vec{0}$, which corresponds to the Hardy space in the poly-disc. Now let $\vec{s} = (s_j)$ and $0 < s_j \leq 1$.

2.2. Unweighted Dirichlet space in the poly-disc

We first consider the case when all $s_j = 1$. For brevity, we assume $d = 2$. For the unweighted Dirichlet space, this is not a restriction of generality, as we will see soon. The reproducing kernel is $K_{\vec{1}}(z, \zeta) = \log(1 - z_1\bar{\zeta}_1)\log(1 - z_2\bar{\zeta}_2) = K_1(z_1, \zeta_1)K_1(z_2, \zeta_2)$. The first idea is to see that our inequality (2.5) (equivalent to the embedding)

$$\int_{\mathbb{D}^2} g(z)\,\overline{g(\zeta)}\,K_{\vec{1}}(z, \zeta)\,d\nu(z)\,d\nu(\zeta) \leq A\,\|g\|^2_{L^2(\mathbb{D}^2, d\nu)} \tag{2.9}$$

implies that for every $C \geq 0$ we have

$$\int_{\mathbb{D}^2} g(z)\,\overline{g(\zeta)}(C + K_1(z_1, \zeta_1))\,(C + K_1(z_2, \zeta_2))\,d\nu(z)\,d\nu(\zeta) \leq B(C)\|g\|^2_{L^2(\mathbb{D}^2, d\nu)}. \tag{2.10}$$

To deduce this inequality from (2.9), one should open the brackets and consider four terms in the left-hand side. The term with $K_1(z_1, \zeta_1))K_1(z_2, \zeta_2)$ is $\lesssim \|g\|^2_{L^2(\mathbb{D}^2, d\nu)}$ by (2.9). The term with $C^2 \int_{\mathbb{D}^2} g(z)\,\overline{g(\zeta)}\,d\nu(z)\,d\nu(\zeta)$ is obviously $\lesssim \|g\|^2_{L^2(\mathbb{D}^2, d\nu)}$ by the Hölder inequality. Consider one of mixed terms (they are treated symmetrically):

$$C \int_{\mathbb{D}^2} g(z)\,\overline{g(\zeta)}\,K_1(z_1, \zeta_1)\,d\nu(z)\,d\nu(\zeta) =: C \cdot I,$$

skip C, and, using the disintegration theorem and the pushing forward of ν to the first coordinate (we call that push forward ν_1), we write I as follows:

$$I = \int_{\mathbb{D}} G(z_1)\,\overline{G(\zeta_1)}\,K_1(z_1, \zeta_1))\,d\nu_1(z_1)\,d\nu_1(\zeta_1),$$

where

$$G(w) := \int g(w, u)\,d\nu_w(u)$$

and $d\nu_w(u)$ are slicing measures: $\nu(E) = \int \nu_w(E)d\nu_1(w)$.

The push forward measure ν_1 on $\mathbb{D}$ is obviously a Carleson measure for 1D Dirichlet space if ν is a Carleson measure for the Dirichlet space in 2D. Therefore,

$$\begin{aligned}&\int_{\mathbb{D}} G(z_1)\,\overline{G(\zeta_1)}\,K_1(z_1,\zeta_1))\,d\nu_1(z_1)\,d\nu_1(\zeta_1) \le B\int_{\mathbb{D}} |G(z_1)|^2\,d\nu_1(z_1)\\ &\quad\le B\int_{\mathbb{D}}\Big(\int_{\mathbb{D}}|g(z_1,z_2)|\,d\nu_{z_1}(z_2)\Big)^2 d\nu_1(z_1) \le B'\int_{\mathbb{D}^2}|g(z_1,z_2)|^2\,d\nu_{z_1}(z_2)d\nu_1(z_1)\\ &\quad\le B'\int_{\mathbb{D}^2}|g(z_1,z_2)|^2\,d\nu(z)\,.\end{aligned}$$

We deduced (2.10) from (2.9) by the use of the disintegration theorem and slicing measures. Notice that the nature of the kernel did not play any role. We could have done this with any dimension d and any kernel $K_{\vec{s}}$ instead of $K_{\vec{1}}$.

The fact that we worked with precisely $K_{\vec{1}}$ is crucial. In fact, the values of $1-z\bar{\zeta}$, for $z,\zeta\in\mathbb{D}$, are obviously in the right half-plane. Hence, as $\Im K_1$ is the argument of $\log\frac{1}{1-z\bar{\zeta}}$, we have

$$|\Im K_1(z,\zeta)|\le \pi/2. \tag{2.11}$$

Hence, by adding a sufficiently large constant $C>0$ to $K_1(z,\zeta)$ we achieve a) $|\Re(C+K_1)|\gg|\Im(C+K_1)|$, and b) $|\Re(C^d+K_{\vec{1}}(z,\zeta))|\ge c\Re(\Pi_{j=1}^d(C+K_1)(z_j,\zeta_j)))$ for any dimension d; it is enough to choose $C=C(d)$ a large positive number. The latter inequality implies that

$$\Re\Pi_{j=1}^d(C+K_1(z_j,\zeta_j))\asymp|\Pi_{j=1}^d(C+K_1(z_j,\zeta_j))|\,.$$

Therefore, for $\vec{s}=\vec{1}$, by modifying the kernel we can achieve (2.8) without changing the class of Carleson measures. This is shown by (2.10). This means that without changing the set of embedding measures, we can equivalently replace the inequality (2.5) by (2.7). This reasoning works for $\vec{s}=\vec{1}$ and any dimension d.

2.3. Weighted Dirichlet space in the poly-disc

Now let $\vec{s}=(s_j)_{j=1}^d$, $0<s_j\le 1$, but $\vec{s}\ne\vec{1}$. We are unable to repeat the trick that was successful in the previous section. In fact, for $K_s=(1-z\bar{\zeta})^{s-1}$, with $0<s<1$, (2.11) does not hold, the imaginary part will not be bounded, and so the previous reasoning with adding a large constant to each kernel of each variable does not work.

However, to reduce the analytic embedding (2.5) to the dyadic embedding on multi-trees, we seem to really need to show that (2.5) implies (2.7) (the converse implication being always trivial).

Here we have only partial results, namely for the case when

$$1-\varepsilon(d)\le s_j\le 1$$

for $\varepsilon(d)$ sufficiently close to 0.

We just notice that $1 - z\bar{\zeta}$ lies in the right half-plane if $z, \zeta \in \mathbb{D}$, and so $(1 - z\bar{\zeta})^\varepsilon$ lies in the cone $C_\varepsilon = \{u + iv, u \geq 0, |v| \leq u \cdot \tan \pi\varepsilon\}$. Therefore, for every $s_j \in (1 - \varepsilon, 1)$,

$$|\Im K_{s_j}(z_j, \zeta_j)| \leq \tan \pi\varepsilon \cdot \Re K_{s_j}(z_j, \zeta_j) .$$

This implies that if ε is sufficiently small (depending on the dimension d), then (2.8) holds, which, as we have already explained, gives us the equivalence of (2.5) and (2.7).

From (2.7), we will now proceed to conclude that the dyadic embedding holds. Then we will explain why the dyadic embedding implies (2.7), thus closing the circular argument.

2.4. From the embedding of analytic functions in the poly-disc to the dyadic multi-parameter embedding

Consider a fixed dyadic lattice $\mathcal{D}$ on $\mathbb{T}$. By this, we mean the following. For any dyadic arc I on $\mathbb{T}$, the symbol $Q(I)$ denotes a Carleson box:

$$Q(I) := \{z = re^{i\theta} \in \mathbb{D} : \theta \in I, 1 - |I| \leq r \leq 1\} .$$

By $q(I)$ we understand its top half:

$$q(I) := \{z = re^{i\theta} \in \mathbb{D} : \theta \in I, 1 - |I| \leq r \leq 1 - |I|/2\} .$$

Sets $q(I)$ form a classical Whitney decomposition of $\mathbb{D}$ into dyadic Carleson half-boxes.

This Whitney decomposition corresponds to a chosen dyadic lattice. Clearly there is a one-to-one correspondence between these half boxes and the vertices of a dyadic tree T just because vertices of T and dyadic intervals of $\mathcal{D}$ are in one-to-one correspondence. So each half box has an address α, which is a vertex of T, so we can write

$$q(\alpha) = q(I) .$$

We can choose a fixed dyadic lattice for each coordinate tori $\mathbb{T}$. Consequently, the Whitney decomposition of $\mathbb{D}^d$ generated by Cartesian products of the respective coordinate decompositions can be encoded by vertices of T^d, i.e., each (multi-)half box q corresponds to a point $\alpha_q \in T^d$, and vice-versa, each $\alpha \in T^d$ has a unique counterpart $q(\alpha)$, where now α is a multi-index corresponding to the vertex of T^d.

The reader should keep it in mind when we will consider boxes constructed by random choices of dyadic lattices $\omega := (\mathcal{D}_1, \dots, \mathcal{D}_d)$. Notice that the collections $\omega := (\mathcal{D}_1, \dots, \mathcal{D}_d)$ of dyadic lattices form a natural measure space provided with a probability measure: $(\Omega, \mathbb{P})$. For future purposes, notice that given a point z in the poly-disc $\mathbb{D}^d$, and a random multi-lattice ω, we will denote the address of the box that contains z by the symbol $\alpha^\omega(z)$ (any fixed z is contained in an open box almost surely, and, thus, the address is uniquely defined by z and ω). The box should be called $q(\alpha^\omega(z))$; often we will skip ω.

As a result, we can define a family $\Lambda = \Lambda^\omega : \mathrm{Meas}^+(\mathbb{D}^d) \to \mathrm{Meas}^+(T^d)$ of canonical maps given by

$$\Lambda\nu(\alpha) = \nu(q(\alpha)). \tag{2.12}$$

Similarly, given a function $g \in L^2(\mathbb{D}^d, d\nu)$, we write

$$\Lambda g(\alpha) := \frac{1}{\nu(q(\alpha))} \int_{q(\alpha)} g(z)\, d\nu(z).$$

A vertex $\alpha \in T^d$ corresponds to a rectangle $R(\alpha) = I_1 \times \cdots \times I_d$. Let $\vec{s} = (s_1, \ldots, s_d)$, $0 < s_i \le 1$, $i = 1, \ldots, d$. Put

$$w_{\vec{s}}(\alpha) = |I_1|^{s_1 - 1} \cdots |I_d|^{s_d - 1}. \tag{2.13}$$

This is the weight on T^d that is associated to the embedding theorem on T^d. This theorem corresponds to the embedding theorem of the class $\mathcal{H}_s$ on $\mathbb{D}^d$.

Define a random kernel as follows. Fix $\omega \in \Omega$ and $(z, \zeta) \in \mathbb{D}^{2d}$. In the dyadic multi-lattice ω, find α^ω and β^ω such that $z \in q(\alpha^\omega)$ and $\zeta \in q(\beta^\omega)$. Up to measure zero of ω, z and ζ lie in corresponding open boxes, hence the boxes are uniquely defined, and so α^ω and β^ω are well-defined. Then consider

$$k^\omega(z, \zeta) := (\mathbf{I}_{w_{\vec{s}}} \mathbf{1})(\alpha^\omega(z) \vee \beta^\omega(\zeta)).$$

where $\alpha \vee \beta$ is the least common ancestor of α and β in the geometry of T^d. In particular, for $\vec{s} = \vec{1}$, the multi-tree kernel $\mathbf{I}_{w_{\vec{1}}} \mathbf{1}(\alpha \vee \beta)$ is the number of ancestors that are common for α and β. If $\vec{s} \ne \vec{1}$, the kernel counts the weighted number of ancestors.

An elementary computation gives that, independently of ω, the following inequality holds if $s_i \ne 0, i = 1, \ldots, d$:

$$k^\omega_{\vec{s}}(z, \zeta) \lesssim |K_{\vec{s}}|(z, \zeta). \tag{2.14}$$

The implied constant depends only on d and $s_i \ne 0, i = 1, \ldots, d$.

Remark 2.1. If some s_i vanish, we have "a phase transition" in the kernel, and (2.14) stops to be true in general. This explains the special role of Hardy spaces on the poly-disc. If the reader thinks that the Chang–Fefferman theory gives the embedding theorem for the Hardy space $H^2(\mathbb{D}^d)$ (the case $s_i = 0, i = 1, \ldots, d$), we should upset the reader by saying that this is not so. The Chang–Fefferman theory gives the characterization of embedding measures in the d-harmonic space $h^2(\mathbb{D}^d)$. As, obviously, the Hardy space of holomorphic functions in the poly-disc is such that $H^2(\mathbb{D}^d) \subset h^2(\mathbb{D}^d)$, the Chang–Fefferman theory gives the sufficient condition for a measure to be an embedding measure for the Hardy class, but whether it is a necessary condition (we believe it is) is not known outside the classical case $d = 1$. If the influential paper [15] were correct, then its proof could be modified to give this necessity, but unfortunately the note [34] indicated a counterexample to the reasoning (but not to the result) of [15].

The inverse inequality to (2.14) is generally not true due to the difference between hyperbolic geometry on the unit disc and that of a dyadic tree. However, one can verify that if one considers the family of dyadic lattices $\Omega\ \omega = (\mathcal{D}_1, \ldots, \mathcal{D}_d)$ on $\mathbb{T}^d$ with a natural probability measure on this family, then the following holds:

$$\forall (z, \zeta) \in \mathbb{D}^{2d},\ \exists\, \Omega(z, \zeta) \subset \Omega: \tag{2.15}$$
$$\text{a) } \mathbb{P}(\Omega(z, \zeta)) \ge c_d > 0, \quad \text{and} \quad \text{b) } \forall \omega \in \Omega(z, \zeta),\ |K_{\vec{s}}|(z, \zeta) \le C_d\, k^\omega_{\vec{s}}(z, \zeta).$$

where c_d and C_d depend only on dimension d.

Now (2.14) and (2.15) give us

$$K_{\vec{s}}(z,\zeta) \asymp \mathbb{E}_\omega k^\omega_{\vec{s}}(z,\zeta)\,. \tag{2.16}$$

By Tonelli's theorem, we have

$$\sum_{\alpha\in T^d}\sum_{\beta\in T^d} \Lambda g(\alpha)\,\Lambda g(\beta)\,\mathbf{I}_{w_{\vec{s}}}\mathbf{1}(\alpha\vee\beta)\,\Lambda\nu(\alpha)\,\Lambda\nu(\beta) = \int_{T^d} (\mathbf{I}^*(\Lambda g\Lambda\nu))^2\,dw_{\vec{s}}, \tag{2.17}$$

For fixed ω, and for $z\in\mathbb{D}^d, \zeta\in\mathbb{D}^d$, we use that $k\vec{s}^\omega(z,\zeta)$ is constant on each pair of boxes from the multi-lattice ω detected by the pair (z,ζ) to write

$$\begin{aligned}\int_{\mathbb{D}^d}\int_{\mathbb{D}^d} & g(z)\,g(\zeta)\,k^\omega_{\vec{s}}\,d\nu(z)\,d\nu(\zeta)\\ &= \sum_{q(\alpha^\omega)}\sum_{q(\beta^\omega)} \Lambda g(q(\alpha^\omega))\,\Lambda g(q(\alpha^\omega))\,\mathbf{I}_{w_{\vec{s}}}\,\mathbf{1}(\alpha^\omega\vee\beta^\omega)\,\Lambda\nu(q(\alpha^\omega))\,\Lambda\nu((q(\beta^\omega))\\ &\lesssim \|\Lambda g\|^2_{L^2(\Lambda\nu)} \le \|g\|^2_{L^2(\nu)}\,,\end{aligned} \tag{2.18}$$

where the constants of equivalence depend only on the dimension. Here we used (2.17) and the boundedness of the operator with kernel $\mathbf{I}_{w_{\vec{s}}}\mathbf{1}(\alpha\vee\beta)$ on the graph T^d.

Now let us hit (2.18) by expectation in ω and use (2.16). Therefore (2.7) follows from (2.3) for $\mu=\Lambda\nu$ and $w=w_{\vec{s}}$.

Assume now that (2.7) holds. Fix a measure μ on T^d. Fix any ω. Let ν be any measure on $\mathbb{D}^d$ such that $\Lambda\nu=\mu$. Then

$$\int_{\mathbb{D}^d}\int_{\mathbb{D}^d} g(z)\,g(\zeta)\,k^\omega_{\vec{s}}\,d\nu(z)\,d\nu(\zeta) \lesssim \|g\|^2_{L^2(\nu)}$$

just because of (2.14). Apply this inequality to the special non-negative g that assumes constant values on each given box $q(\alpha^\omega)$. We can choose those constants arbitrarily with the only condition that $\|g\|^2_{L^2(\nu)} = \|g\|^2_{L^2(\Lambda\nu)} < \infty$. Then we get (2.3) for $\mu=\Lambda\nu$ and $w=w_{\vec{s}}$.

2.5. Verifying (2.15)

It is enough to verify it for $d=1$, because then we can use the product structure of the kernel $|K_{\vec{s}}|$ and the independence of the lattices $\mathcal{D}_1,\dots,\mathcal{D}_d$. Put $D(z,\zeta) := |z-\zeta| + 1-|z| + 1-|\zeta|$ (it is a sort of distance). Then

$$K_s(z,\zeta) = |1-z\bar{\zeta}|^{s-1} \approx D(z,\zeta)^{s-1}\,.$$

Now we define the analogous dyadic distance that depends on a dyadic lattice, say lattice L. We define $D^L(z,\zeta)$ as the smallest length of the dyadic arcs from L that are larger than $\max(1-|z|, 1-|\zeta|)$ and contain the shorter arc that has end-points z and ζ.

Then right-hand side of (2.15) (for $d=1$) is $\approx (D^L(z,\zeta))^{s-1}$ and D^L is always $\ge cD$, where c is an absolute constant. Of course we have

$$(\mathrm{dist}^L(z,\zeta))^{s-1} \approx \mathbf{I}_{w_s}\mathbf{1}(\alpha\vee\beta)\,.$$

To prove (2.15) (for $d = 1$), it is enough to prove that

$$\mathrm{dist}^L(z,\zeta) \le C\, D(z,\zeta)$$

for a set of dyadic lattices of a fixed probability. Let the full family of dyadic lattices be just the rotation of one fixed lattice provided with the natural probability measure $d\theta/2\pi$.

Let I be a dyadic arc of length $2\pi \cdot 2^{-m}$. Given z and ζ, let us calculate the probability of being a bad dyadic lattice, where bad means that the inequality displayed above is false with constant $C = 8$. Each dyadic lattice has two end-points of the first division, four end-points of the second division, etc.

Then the probability for the first division points to be inside I is $\frac{2}{2\pi/|I|} = 2 \cdot 2^{-m}$ (as we have two such points). The probability for the second division points to be inside I is $4 \cdot 2^{-m}$. We continue until we find the $(m-4)$-th division points, for which the probability such a point is in I is almost $2^{m-4} \cdot 2^{-m}$. These are all bad scenarios. Their probability is at most $1/8$.

Hence, the probability none of these points are in I (which we can call a "good" event) is at least $7/8$. But if none of these division points are inside I, we have

$$D_L(z,\zeta) \le 10\, D(z,\zeta)\,.$$

Inequality (2.15) is proved.

Let us formulate the reduction from the d-disc to the d-tree by a theorem.

Theorem 2.2. *Let $\vec{s} = (s_1,\dots,s_d)$, $s_i \in (0,1]$, $i = 1,\dots,d$, where all s_i are sufficiently close to* 1: $1 - s_i \le \varepsilon_d$, *for a certain positive absolute constant ε_d and $i = 1,\dots,d$. Let ν be a measure in $\mathbb{D}^d$. Then the embedding operator* id: $\mathcal{H}_{\vec{s}}(\mathbb{D}^d) \to L^2(\mathbb{D}^d,\nu)$ *is bounded if and only if for any dyadic multi-lattice ω on $\mathbb{T}^d$, the measure $\mu = \Lambda\nu$ on T^d obtained by formula* (2.12) *and the weight $w_{\vec{s}}$ on T^d from* (2.13) *give us the embedding pair on T^d in the sense that*

$$\sum_{\alpha\in T^d} (\mathbf{I}^*\psi\mu)^2(\alpha)\, w_{\vec{s}}(\alpha) \le C \int_{T^d} \psi^2\, d\mu\,.$$

Notice that the weight $w_{\vec{s}}(\alpha)$ here has a tensor product form:

$$\alpha = (\alpha_1,\dots,\alpha_d) \ \Rightarrow\ w_{\vec{s}}(\alpha) = u_{s_1}(\alpha_1)\cdots u_{s_d}(\alpha_d)\,.$$

Remark 2.3. Notice that for $\vec{s} = \vec{1}$ (Dirichlet space case), integration in (2.17) with respect to $dw_{\vec{s}}$ means just summation over all vertices of T^d. For other $\vec{s}$, a natural weight appears (it weights the vertices), and the summation has to be with respect to this weight. In our situation of the scale $\mathcal{H}_{\vec{s}}$ (of various spaces of analytic functions in the poly-disc described at the beginning of this section), the weight that appears is always the product of weights in each coordinate. *This emphasizes why we especially care about the results with product weights.*

To summarize, the problem of characterizing Carleson measures for the weighted Dirichlet space $\mathcal{H}_{\vec{s}}$ can be often moved to a discrete medium (for $\vec{s} = \vec{1}$ can be *always* done, for any dimension d), and after that this problem interpreted (without any loss of information) as the problem of characterizing a trace pair $(\mu, w_{\vec{s}})$. For instance, we will see that (2.3) is equivalent to a single box condition (since $w_{\vec{s}}$ has a product structure)

$$\sum_{\beta \leq \alpha} (\mathbf{I}^* \Lambda \nu)^2(\beta)\, w_{\vec{s}}(\beta) \lesssim \mathbf{I}^* \Lambda \nu(\beta)$$

for any $\beta \in T^d$. On the poly-disc, this condition transforms to

$$\sum_{R \subset Q} \nu^2(T(R))\, w_{\vec{s}}(R) \lesssim \nu(T(Q)), \quad \text{for any } Q,$$

where Q and R are dyadic rectangles on the (poly-)torus $\mathbb{T}^d$, and $T(Q)$ is the usual tent area above Q. One can also check that this condition is necessary by testing Carleson embedding on appropriate functions.

The argument above fails, for a number of reasons, if even one of the parameters s_j becomes zero. However, for the classical Hardy space on the poly-disc one can still make a connection between the Carleson embedding and the Hardy inequality, but now using the direct embedding (2.2) instead of the dual (2.3), and the roles of μ and w are reversed. It is done in Section 3.

3. The end-point case $s = 0$

We repeat ourselves: the equivalence (2.8) – ultimately important for us to prove equivalence of dyadic and analytic embeddings – has limitations. First of all (2.8), is false even for the case $d = 1$ if $s = 0$, see (2.4). That makes the case $s = 0$ quite special. It is well known that for the $d = 1$ case, the embedding measures for the Poisson and the Cauchy kernels on $L^2(\mathbb{T})$ are the same. This is rather classical, but should be considered as a "miracle", exactly because (2.8) fails. Already in the 2D situation, the fact that the embedding measures for the Poisson $P_{z_1} P_{z_2}$ and the Cauchy $K_{\vec{0}}(z, \zeta) = (1 - z_1 \bar{\zeta}_1)^{-1}(1 - z_2 \bar{\zeta}_2)^{-1}$ kernels on $L^2(\mathbb{T}^2)$ are the same is a subtle fact that will be considered in [25] separately. It is based on Ferguson–Lacey's characterization of symbols of "little" Hankel operators [15, 21].

Another interesting distinction of the case $s = 0$ is about (2.4). The reader will see that for $s > 0$ we characterize the embedding in terms of a simple box (rectangular) test. As it is well known from the works of Chang, Fefferman and Carleson [7, 8, 13, 32], such characterization is not possible for the Poisson embedding of $L^2(\mathbb{T}^d)$ if $d \geq 2$. We would wish to attribute this phenomena to the fact that the Poisson kernel has a special shape. In our language, this means that unlike (2.14) above, that holds for $s \neq 0$, the same type of inequality for the Poisson kernel,

$$\mathbf{I}_{w_{\vec{0}}} \mathbf{1}(\alpha \vee \beta) \lesssim P(\alpha, \beta),$$

is false, where P is a multi-parameter Poisson kernel, $P(\alpha, \beta) := \sup_{z \in q(\alpha), \zeta \in q(\beta)} P(z, \zeta)$.

For $s = 0$, the space $\mathcal{H}_s(\mathbb{D}^d) =: H^2(\mathbb{D}^d)$ is the Hardy space on the poly-disc. The embedding Id: $H^2(\mathbb{D}^d) \to L^2(\mathbb{D}^d, d\nu)$ can be still equivalently described as inequality (2.5), but cannot be described any longer as inequality (2.7). The reason is that the reproducing kernel $K_0(z,\zeta) = (1 - z\bar{\zeta})^{-1}$ does not satisfy anymore the property that its real part is equivalent to its absolute value.

Still, we want to deduce the embedding theorem Id: $H^2(\mathbb{D}^d) \to L^2(\mathbb{D}^d, d\nu)$ from the dyadic statement of the type (2.3). Notice that the embedding of the Hardy space of analytic functions in the poly-disc follows from the Poisson embedding. Also notice that for dimension $d = 1$, these two embeddings are equivalent, in the sense that the classes of embedding measures in the disc are the same.

This is absolutely not obvious for $d > 1$. So below we consider only the embedding of $L^2(\mathbb{T}^d)$ by the means of the multi-Poisson kernel. We do not touch upon the question of equivalence of this Poisson embedding of $L^2(\mathbb{T}^d)$ and the (Poisson) embedding of $H^2(\mathbb{T}^d)$. The relation between the two embeddings (that of $L^2(\mathbb{T}^d)$ and that of $H^2(\mathbb{T}^d)$) for $d > 1$ will be addressed in [25]. It is a really subtle question, that requires the extension of [15]. To our utmost consternation, this question has not been addressed in the literature.

To this end, we stop to consider the adjoint operator to the embedding Id: $H^2(\mathbb{D}^d) \to L^2(\mathbb{D}^d, d\nu)$. Instead we consider this embedding directly, namely, if P^k denotes the Poisson extension in the k-th variable, we write down our embedding as the following inequality:

$$\int_{\mathbb{D}^d} [P^1 \dots P^d f]^2 \, d\nu \le \int_{\mathbb{T}^d} |f|^2 \, dm_d \,, \tag{3.1}$$

where $\mathbb{T}^d$ is the torus and m_d its Lebesgue measure. We emphasize again that this should hold for any $f \in L^2(\mathbb{T}^d, m_d)$. Let $\{q(\alpha)\}_{\alpha \in T^d}$ be the Whitney decomposition of $\mathbb{D}^d$ generated by Cartesian products of the respective coordinate decompositions. By [8], we know that the inequality (3.1) is equivalent to the Carleson–Chang condition:

$$\sum_{\alpha:\, q(\alpha) \cap \mathrm{Tent}(\Omega) \neq \emptyset} \nu(q(\alpha)) \le C m_d(\Omega) \quad \forall \text{ open } \Omega \subset \mathbb{T}^d \,. \tag{3.2}$$

So we wish to deduce the implication (3.2) $\Rightarrow$ (3.1) by using only the dyadic multi-tree statement that we will formulate now.

Let $f\colon T^d \to [0,\infty)$ and let m_d be the Lebesgue measure on $\partial T^d := (\partial T)^d$ given by $m_d(\omega) = 2^{-Nd}$. Now let Ω be an arbitrary union of elementary cubes ω's of size 2^{-N}. Call such sets *dyadic open sets*. For any $\alpha \in T^d$, we denote by R_α the dyadic d-subrectangle of the unit cube that corresponds to α. Let $\nu\colon T^d \to [0,\infty)$ be such that

$$\sum_{\alpha:\, R_\alpha \subset \Omega} (m_d(R_\alpha))^2 \nu(\alpha) \le C m_d(\Omega) \quad \forall \text{ dyadic open } \Omega \subset \partial T^d \,. \tag{3.3}$$

We consider the inequality on the multi-tree T^d:

$$\int_{T^d} \left(\mathbf{I}^*(f\,dm_d)\right)^2 d\nu \le C_1 \int_{\partial T^d} f^2 \, dm_d \,. \tag{3.4}$$

Suppose we know that (3.3) $\Rightarrow$ (3.4) (with different constants, but without dependence on N). We want to use this implication as the only tool to prove implication (3.2) $\Rightarrow$ (3.1).

This requires some work even for the case $d = 1$. Below is the way to do this reduction for $d = 1, 2$. General d follows the same steps.

For an interval I of $\mathbb{R}$, Q_I denotes a Carleson box, and T_I denotes its upper half. Similarly, for a rectangle $R = I \times J$ in $\mathbb{R}^2$, we have $Q_R := Q_I \times Q_J$ and $T_R := T_I \times T_J$. If I runs over a certain dyadic lattice of intervals, then the T_I tile the upper half-plane. Similarly, if R runs over dyadic system of rectangles, the T_R tile $\mathbb{C}^2_+$. In the next two subsections, I_0 always denotes $[-1, 1]$, and we let Q_0 be always $Q_{[-1,1]}$.

3.1. The one dimensional case

Let Pf mean the Poisson extension of f. We first consider the 1D case. Let ν be a measure that lies in the upper half plane, and that satisfies the following box Carleson condition:

$$\nu(Q_I) \le C_1\, |I|, \quad \forall I\,. \tag{3.5}$$

Let f be a nonnegative test function on the real line with support in $[-1/2, 1/2]$, We want to give a new proof of the Carleson embedding

$$\int_{Q_0} [Pf]^2\, d\nu \le C_2 \int f^2 dx, \tag{3.6}$$

where C_2 depends only on C_1.

As we have Harnack's inequality for Pf, we always may assume that ν is a doubling measure in the Poincaré metric of $\mathbb{C}_+$.

We wish to prove the implication (3.5) $\Rightarrow$ (3.6) by allowing ourselves to use only the implication (3.7) $\Rightarrow$ (3.8), where given a dyadic lattice $\mathcal{D}$, we have

$$\nu(Q_I) \le C_1\, |I|, \quad \forall I \in \mathcal{D}, \tag{3.7}$$

$$\sum_{J\in\mathcal{D}} \langle f\rangle_J^2\, \nu(T_J) \le C \int_{\mathbb{R}} f^2 dx\,. \tag{3.8}$$

Here are several notations: as always, for a given I, λI means the interval with the same center, but with length $\lambda|I|$. If I is an interval of a dyadic lattice $\mathcal{D}$, then I^j is its ancestor such that $|I^j| = 2^j|I|$. We denote by $c_I = x_I + iy_I$ the center of T_I, and by P_I, the Poisson kernel with pole at c_I. As $P_I f$ is bounded by an absolute constant times the convex combination of the averages $\langle f\rangle_{2^k I}$, $k = 0, \dots, \log(1/|I|)$, and the average $\langle f\rangle_{I_0}$, we can choose k_I that gives the maximum to $\langle f\rangle_{2^k I}$, $k = 0, \dots, \log(1/|I|)$, and then

$$Pf(c_I) = P_I f \le A_1\langle f\rangle_{2^{k_I}\cdot I} + A_2\langle f\rangle_{I_0}\,.$$

Our goal is to give a new way to prove (3.6). Traditionally it is deduced from (3.5) by a interpolation argument. We wish to deduce it using only the dyadic L^2 estimate. The second term is trivial to estimate.

To estimate the first term, we will do the following. We consider the probabilistic space of dyadic lattices built as follows. Divide $\mathbb{R}$ into equal intervals of size 2^{-N}, where N is very large. We do it to have $[-1/2, 1/2]$ tiled. Now we can toss the coin and choose which pair is united to one dyadic interval of size 2^{-N+1}. These are the fathers. Toss the coin again to choose who are grandfathers. Now for a given interval of size 2^{-N} we have already 4 different grandfathers, each with probability $1/4$. We continue this tossing for a total number of $N+4$ tossings. For any interval of size 2^{-N} inside $[-1/2, 1/2]$, the most senior ancestor will contain $Q_0 = [-1/2, 1/2]$ with probability $15/16$. We call the collection of such dyadic lattices Ω (it is a finite family of lattices). All dyadic lattices in Ω have equal probability, and we just renormalize the probability to have $\mathbb{P}(\Omega) = 1$.

The thus obtained random dyadic lattices will be called $\mathcal{D}(\omega)$, their probability space will be called $(\Omega, \mathbb{P})$. Now fix $\omega \in \Omega$ (meaning fix one of those lattices), and consider some small $I \in \mathcal{D}(\omega)$ of size 2^{-N}. We consider c_I and find k_I as above. Consider $2^{k_I} I$. It may not be dyadic, but it has the same center x_I as the dyadic I, so consider I^{k_I} and I^{k_I+10} and check whether I^{k_I} is inside $\frac{3}{4} I^{k_I+10}$. Suppose it is. Then obviously, as $x_I \in I^{k_I}$, we will have that

$$2^{k_I} I \subset I^{k_I+10}.$$

It is very easy to see that

$$\mathbb{P}\big\{I^{k_I} \text{ is inside } \tfrac{3}{4} I^{k_I+10}\big\} \geq 1/2.$$

Thus

$$\mathbb{P}\{2^{k_I} I \subset I^{k_I+10}\} \geq 1/2.$$

If the event $2^{k_I} I \subset I^{k_I+10}$ happens, then we call T_I good, we color it red, and we color I^{k_I+10} also red; but we take the measure ν on T_I, color it blue and move this blue mass to $T_{I^{k_I+10}}$. No measure then is left in T_I. All measure movements are "up". It never happens that the measure is moved into a square Q_J, $J \in \mathcal{D}(\omega)$, from outside of Q_J. Therefore, the new measure satisfies the same Carleson condition (3.5) for all boxes Q_J, where J is in this $\mathcal{D}(\omega)$.

Otherwise, we call T_I bad, and we color it white. We do nothing else.

Then we look at intervals of size 2^{-N+1} and repeat all that. We do this for every $\mathcal{D}(\omega)$. Obviously the same T_I can be good for some ω and bad for others. We established above that the probability to be good is at least $1/2$.

It may happen that a certain T_J has blue mass (moved from below) and original mass. If we need to move mass from T_J we color blue and move only original mass, the "new" mass; the blue mass, which came from below, rests unmoved.

When we finish the procedure we have a new measure, and we color it all blue (many parts of it are already colored blue), and we call it $\nu_b(\omega)$ (it is random, and it also depends on f). But it is dyadic Carleson like (3.5) for all boxes Q_I, $I \in \mathcal{D}(\omega)$.

After this procedure, it may very well happen that for a given $\mathcal{D}(\omega)$ and $J \in \mathcal{D}(\omega)$, T_J is colored white, but J is colored red and T_J contains blue mass particles.

For every ω we also have subdomains R (colored red) and W (colored white) of Q_0, $W = W(\omega)$ consisting of bad T_I, $I \in \mathcal{D}(\omega)$, and $R = R(\omega)$, consisting of good T_I, $I \in \mathcal{D}(\omega)$. Now

$$\begin{aligned}\int_R [Pf]^2\, d\nu &\lesssim \sum_{I \in \mathcal{D}(\omega), T_I \text{ good}} \langle f\rangle^2_{I^{k_I+10}}\, \nu_b(T_{I^{k_I+10}}) + \langle f\rangle^2_{I_0} |\nu| \\ &\leq \sum_{J \in \mathcal{D}(\omega), J \text{ red}} \langle f\rangle^2_J\, \nu_b(T_J) \leq C \int_{\mathbb{R}} f^2\, dx,\end{aligned}$$

This is because we always preserve the dyadic box Carleson (3.5) property for $\nu_b(\omega)$ in the corresponding $\mathcal{D}(\omega)$. On the other hand, let us denote by $\mathcal{F}$ the union of all I's in all dyadic lattices $\mathcal{D}(\omega), \omega \in \Omega$, such that $2^{-N} \leq |I| \leq 2^4$. Then

$$\int_\Omega \int_{R(\omega)} [Pf]^2 \, d\nu \, d\mathbb{P}(\omega) \geq \frac{1}{2} \int_{Q_0} [Pf]^2 \, d\nu,$$

because each T_I, for $I \in \mathcal{F}$, will be red at least half of the time (meaning that $\mathbb{P}\{T_I \text{ is red}\} \geq 1/2$).

3.2. The multi-dimensional case

Now the measure ν is in Q_0^n. We will consider for brevity only the case $n = 2$. The measure ν satisfies the Chang–Carleson condition. For any open set $G \subset Q_0$, consider its tent: $T_G = (\{z, w) \in \mathbb{C}_+ : R(z, w) \subset G\}$, where

$$R(z, w) := [\Re z - \Im z, \Re z + \Im z] \times [\Re w - \Im w, \Re w + \Im w] \,.$$

The Chang–Carleson condition is

$$\nu(T_G) \leq C_1 \, |G| \,, \tag{3.9}$$

where $|G|$ denotes the plane Lebesgue measure of G.

As we have Harnack's inequality, we always may assume that ν is a doubling measure in the natural metric of $\mathbb{C}_+^2$.

This allows us to notice the following. Consider any system of dyadic rectangles. Choose any finite family of dyadic rectangles $R = I \times J$ of this system, we call their union O "a dyadic open set". It has a dyadic tent T_O^d. Now, by definition, it is the union of all T_Q for all dyadic Q (of the same system) such that $Q \subset O$.

The doubling property above (which we assume without loss of generality because of Harnack's principle) allows us to conclude that if ν has property (3.9) it also has the following dyadic Chang–Carleson property:

$$\nu(T_O^d) \leq C \, |O| \,. \tag{3.10}$$

Now let $\mathcal{P} = P^1 P^2$ be the bi-Poisson extension. Fix a test function $f \geq 0$ supported in $[-1/2, 1/2]^2$. Consider two dyadic lattices of one variable as before $\mathcal{D}(\omega_x)$, $\mathcal{D}(\omega_y)$, and consider the system of dyadic rectangles $R = I \times J$, $I \in \mathcal{D}(\omega_x)$, $J \in \mathcal{D}(\omega_y)$. Call this system $\mathcal{D}(\omega)$, $\omega := (\omega_x, \omega_y)$. Let $c_R = (c_I, c_J)$, where c_I is the center of T_I and c_J is the center of T_J.

Let P_I^1 be the Poisson kernel with pole at c_I, and let P_J^2 be the Poisson kernel with pole at c_J. The bi-Poisson extension $P_I^1 P_J^2$ is bounded by an absolute constant times the convex combination of the averages $\langle f \rangle_{2^k I \times 2^m J}$, $k = 0, \ldots, \log(1/|I|)$, $m = 0, \ldots, \log(1/|J|)$, and the average $\langle f \rangle_{Q_0}$. We can choose k_I, m_J that gives the maximum to $\langle f \rangle_{2^k I \times 2^m J}$, $k = 0, \ldots, \log(1/|I|)$, $m = 0, \ldots, \log(1/|J|)$, and then we have

$$\mathcal{P} f(c_I, c_J) = P_I^1 P_J^2 f \leq A_1 \langle f \rangle_{2^{k_I} I \times 2^{m_J} J} + A_2 \langle f \rangle_{Q_0} \,.$$

Again we can ensure that

$$\mathbb{P}\{2^{k_I} I \subset I^{k_I+10}, 2^{m_J} I \subset J^{m_J+10}\} \geq 1/4\,. \tag{3.11}$$

Then we just repeat the coloring scheme from subsection 3.1. This time we color the 4D rectangles T_R, $R \in \mathcal{D}(\omega)$, white if T_R is bad, namely, if the event in (3.11) does not happen, and we color it red and call it good if that event does happen. From red T_R we scoop all the measure ν, color its particles blue and move to $T_{\hat{R}}$ for the ancestor $\hat{R} := I^{k_I+10} \times J^{m_J+10}$ of $R = I \times J$.

Again we will have that the random blue measure $\nu(\omega)$ satisfies (3.10) as the original measure ν does. Then we repeat the calculation of subsection 3.1. We should prove the embedding

$$\int\int_{Q_0\times Q_0} [P^1 P^2 f]^2\, d\nu \leq C \int\int_{I_0\times I_0} f^2\, dm_2\,.$$

We just repeat the averaging over the probability calculation of subsection 3.1.

4. Surrogate maximum principle

From now on, our paper is devoted only to the multi-tree case (the dyadic n-rectangles case). We will need to overcome a major difficulty: the potential theory on multi-trees does not allow a maximum principle.

Let μ be a positive function on an d-tree T^d. Its *energy* is defined as

$$\mathcal{E}[\mu] := \int w(\mathbf{I}^*\mu)^2.$$

We view the weight $w\colon T^d \to [0,\infty)$ as fixed, and keep it implicit in the notation. The energy can be written in terms of the *potential*

$$\mathbf{V}^\mu := \mathbf{I}(w\,\mathbf{I}^*\mu),$$

as $\mathcal{E}[\mu] = \int_{T^d} \mathbf{V}^\mu\, \mathrm{d}\mu$. Consider the truncated potential and the partial energy:

$$\mathbf{V}^\mu_\delta := \mathbf{I}(\mathbf{1}_{\mathbf{V}^\mu\leq\delta}\, w\,\mathbf{I}^*\mu) \quad \text{and} \quad \mathcal{E}_\delta[\mu] := \int_{T^d} \mathbf{V}^\mu_\delta\, \mathrm{d}\mu = \int_{\{\mathbf{V}^\mu\leq\delta\}} w(\mathbf{I}^*\mu)^2.$$

On a 1-tree, we have the maximum principle

$$\mathbf{V}^\mu_\delta \leq \delta. \tag{4.1}$$

It follows that, for any positive function ρ on T, we have

$$\begin{aligned}\int_T \mathbf{V}^\mu_\delta\, \mathrm{d}\rho &= \int_{\{\mathbf{V}^\mu\leq\delta\}} wI^*\mu I^*\rho \leq \min(\delta|\rho|, \mathcal{E}[\mu]^{1/2}\mathcal{E}[\rho]^{1/2}) \\ &\leq (\delta|\rho|)^\kappa\, (\mathcal{E}[\mu]\,\mathcal{E}[\rho])^{(1-\kappa)/2}\end{aligned} \tag{4.2}$$

for every $\kappa \in (0, 1]$, where

$$|\rho| := \int_T \rho.$$

A similar estimate on 2-trees, with a specific κ, was obtained in [4]. In this section, we give a streamlined proof of such an estimate on 2-trees and extend it to 3-trees.

We do not know how to deal with d-trees with $d \geq 4$.

If $T^d = T_1 \times \cdots \times T_d$ is an d-tree, then we denote by $I_1, \ldots, I_n$ the Hardy operators acting in the respective coordinates, so that $\mathbf{I} = I_1 \cdots I_n$. We use a similar index convention for the operators $\Delta_1, \ldots, \Delta_n$.

4.1. 1-trees

Lemma 4.1. *Let T be a tree, and let $f, g\colon T \to [0, \infty)$ be any functions. Then*

$$(If)(Ig) \leq I(If \cdot g + f \cdot Ig).$$

Proof. We have

$$\begin{aligned}
If(\alpha)Ig(\alpha) &\leq If(\alpha)Ig(\alpha) + I(fg)(\alpha) \\
&= \sum_{\alpha' \geq \alpha, \alpha'' \geq \alpha} f(\alpha')g(\alpha'') + \sum_{\alpha' \geq \alpha} f(\alpha')g(\alpha') \\
&= \sum_{\alpha' \geq \alpha'' \geq \alpha} f(\alpha')g(\alpha'') + \sum_{\alpha'' \geq \alpha' \geq \alpha} f(\alpha')g(\alpha'') \\
&= \sum_{\alpha'' \geq \alpha} If(\alpha'')g(\alpha'') + \sum_{\alpha' \geq \alpha} f(\alpha')Ig(\alpha') \\
&= I(If \cdot g)(\alpha) + I(f \cdot Ig)(\alpha).
\end{aligned}$$

∎

Definition 4.2. 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 β:

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

A function $g\colon T \to \mathbb{R}$ is called *superadditive* if for every $\beta \in T$ we have

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

The difference operator is defined by

$$\Delta g(\beta) := g(\beta) - \sum_{\beta' \in \operatorname{ch}(\beta)} g(\beta').$$

Lemma 4.3 (Partial summation). *Let T be a finite tree. For any functions $f, g\colon T \to \mathbb{R}$, we have*

$$\sum_{\alpha \in T} f(\alpha)\, g(\alpha) = \sum_{\alpha' \in T} \Delta f(\alpha')\, Ig(\alpha'). \tag{4.3}$$

Proof. By induction on the size of the tree, one can show

$$f(\alpha) = \sum_{\alpha' \le \alpha} \Delta f(\alpha').$$

It follows that

$$\sum_{\alpha} f(\alpha) g(\alpha) = \sum_{\alpha,\alpha' : \alpha' \le \alpha} \Delta f(\alpha') g(\alpha) = \sum_{\alpha'} \Delta f(\alpha') \sum_{\alpha : \alpha' \le \alpha} g(\alpha) = \sum_{\alpha' \in T} \Delta f(\alpha') I g(\alpha').$$

■

Lemma 4.4. *Let T be a tree, and let $f, g \colon T \to \mathbb{R}$. Then*

$$I^*(fg) = I^*(\Delta f \cdot Ig) - f(Ig - g).$$

Proof. For $\beta \in T$, write $\downarrow\beta := \{\alpha \in T \mid \alpha \le \beta\}$. This is again a sub-tree, on which we can apply the partial summation identity (4.3). Hence,

$$I^*(fg)(\beta) = \int_{\downarrow\beta} fg = \int_{\downarrow\beta} \Delta f \cdot I(g \mathbf{1}_{\downarrow\beta})$$

For each $\alpha \in \downarrow\beta$, we have

$$I(g\mathbf{1}_{\downarrow\beta})(\alpha) = \sum_{\gamma : \alpha \le \gamma \le \beta} g(\gamma) = \sum_{\gamma : \alpha \le \gamma} g(\gamma) - \sum_{\gamma : \beta \le \gamma} g(\gamma) + g(\beta) = Ig(\alpha) - Ig(\beta) + g(\beta).$$

Therefore,

$$\begin{aligned} I^*(fg)(\beta) &= \int_{\downarrow\beta} \Delta f \cdot (Ig - Ig(\beta) + g(\beta)) = \int_{\downarrow\beta} \Delta f \cdot Ig - (Ig(\beta) - g(\beta)) \int_{\downarrow\beta} \Delta f \\ &= I^*(\Delta f \cdot Ig)(\beta) - (Ig(\beta) - g(\beta)) f(\beta). \end{aligned}$$

■

Corollary 4.5 (cf. Lemma 2.2 in [4]). *Let T be a tree, and let $f, g \colon T \to [0, \infty)$. Then*

$$I^*(fg) \le I^*(\Delta f \cdot Ig).$$

4.2. 2-trees

In this section we prove a version of (4.2) on 2-trees that refines Lemma 4.1 in [4]. Recall that $\mathbb{I} = I_1 I_2$.

Lemma 4.6. *Let T^2 be a bi-tree, and let $f, g \colon T^2 \to [0, \infty)$. Then*

$$(\mathbb{I} f)(\mathbb{I} g) \le \mathbb{I}(\mathbb{I} f \cdot g + I_1 f \cdot I_2 g + I_2 f \cdot I_1 g + f \cdot \mathbb{I} g).$$

Proof. The linear operators I_1 and I_2 commute and $\mathbb{I} = I_1 I_2$. To each of I_1 and I_2 we can apply Lemma 4.1. Hence,

$$(\mathbb{I} f)(\mathbb{I} g) = (I_1 I_2 f)(I_1 I_2 g) \le I_1\big((I_1 I_2 f)(I_2 g) + (I_2 f)(I_1 I_2 g)\big).$$

By Lemma 4.1, the sum in the bracket is

$$\begin{aligned} &= (I_2 I_1 f)(I_2 g) + (I_2 f)(I_2 I_1 g) \\ &\le I_2\big((I_2 I_1 f)(g) + (I_1 f)(I_2 g)\big) + I_2\big((I_2 f)(I_1 g) + (f)(I_2 I_1 g)\big). \end{aligned}$$

Hence,

$$\begin{aligned} (\mathbb{I} f)(\mathbb{I} g) &\le I_1\big(I_2\big((I_2 I_1 f)(g) + (I_1 f)(I_2 g)\big) + I_2\big((I_2 f)(I_1 g) + (f)(I_2 I_1 g)\big)\big) \\ &= \mathbb{I}(\mathbb{I} f \cdot g + I_1 f \cdot I_2 g + I_2 f \cdot I_1 g + f \cdot \mathbb{I} g). \end{aligned}$$ ■

The following result will not be used in our current treatment of bi-trees. We include it to illustrate the relation of Lemma 4.6 with the argument in [4].

Corollary 4.7 (cf. Theorem 3.1 in [4]). *Let $0 < \delta \le \lambda/4$. Let $f : T^2 \to [0, \infty)$ with* $\operatorname{supp} f \subseteq \{\mathbb{I} f \le \delta\}$. *Then*

$$(\mathbb{I} f)\, \mathbf{1}_{\mathbb{I} f \ge \lambda} \le 4\lambda^{-1} \mathbb{I}\, (I_1 f \cdot I_2 f).$$

Proof. Substituting $f = g$, Lemma 4.6 implies that

$$(\mathbb{I} f)^2 \le 2\mathbb{I}\, (I_1 f \cdot I_2 f + f \cdot \mathbb{I} f).$$

Using the support condition, this implies

$$\begin{aligned} (\mathbb{I} f)\, \mathbf{1}_{\mathbb{I} f \ge \lambda} &\le \lambda^{-1}\, (\mathbb{I} f)^2\, \mathbf{1}_{\mathbb{I} f \ge \lambda} \le \lambda^{-1} 2\mathbb{I}\, (I_1 f \cdot I_2 f + \delta f)\, \mathbf{1}_{\mathbb{I} f \ge \lambda} \\ &\le 2\lambda^{-1}\, \mathbb{I}\, (I_1 f \cdot I_2 f) + 2\delta \lambda^{-1}\, \mathbb{I} f\, \mathbf{1}_{\mathbb{I} f \ge \lambda}. \end{aligned}$$

Since $2\delta \lambda^{-1} \le 1/2$, this implies

$$(\mathbb{I} f) \mathbf{1}_{\mathbb{I} f \ge \lambda} \le 4\lambda^{-1} \mathbb{I}\, (I_1 f \cdot I_2 f)$$ ■

4.2.1. Energy bound.

Lemma 4.8. *Let T^2 be a 2-tree, and let $f : T^2 \to [0, \infty)$ be a function that is superadditive in each parameter separately. Let $w : T^2 \to [0, \infty)$ be of tensor product form. Suppose that* $\operatorname{supp} f \subseteq \{\mathbb{I}(wf) \le \delta\}$. *Then*

$$\int_{T^2} wf \cdot I_1(w_1 f) \cdot I_2(w_2 f) \cdot \mathbb{I}(wf) \le \delta^2 \int_{T^2} wf^2.$$

Proof. By the hypothesis, the left-hand side of the conclusion is

$$\begin{aligned} &\le \delta \int_{T^2} wf \cdot I_1(w_1 f) \cdot I_2(w_2 f) = \delta \int_{T^2} w_1 f \cdot I_1(wf) \cdot I_2(w_2 f) \\ &= \delta \int_{T^2} wf \cdot I_1^*(w_1 f \cdot I_2(w_2 f)) = \delta \int_{T^2} wf \cdot I_1^*(f \cdot I_2(wf)). \end{aligned} \tag{4.4}$$

By Corollary 4.5, we have

$$I_1^*(f \cdot I_2(wf)) \le I_1^*(\Delta_1 f \cdot I_1 I_2(wf)).$$

Since $\{\mathbb{I}(wf) \le \delta\}$ is an up-set, $\Delta_1 f$ is supported on this set. Since f is superadditive, $\Delta_1 f \ge 0$. Hence,

$$I_1^*(f \cdot I_2(wf)) \le I_1^*(\Delta_1 f \cdot \mathbb{I}(wf)) \le I_1^*(\Delta_1 f \cdot \delta) = \delta f. \tag{4.5}$$

Inserting (4.5) into (4.4), we obtain the claim. ∎

Lemma 4.9. *Let T^2 be a 2-tree, and let $f: T^2 \to [0,\infty)$ be a function that is superadditive in each parameter separately. Let $w: T^2 \to [0,\infty)$ be of tensor product form. Suppose that* supp $f \subseteq \{\mathbb{I}(wf) \le \delta\}$. *Then*

$$\int_{T^2} w(I_1 w_1 f)^2 (I_2 w_2 f)^2 \le 4\delta^2 \int_{T^2} w f^2.$$

Proof. By Lemma 4.1 and the commutativity of operations in different coordinates,

$$\begin{aligned}
\int_{T^2} w(I_1 w_1 f)^2 (I_2 w_2 f)^2 &\le 4 \int_{T^2} w I_1(w_1 f \cdot I_1(w_1 f)) \cdot I_2(w_2 f \cdot I_2(w_2 f)) \\
&= 4 \int_{T^2} I_1(w_1 f \cdot I_1(wf)) \cdot I_2(w_2 f \cdot I_2(wf)) \\
&= 4 \int_{T^2} I_2^*(w_1 f \cdot I_1(wf)) \cdot I_1^*(w_2 f \cdot I_2(wf)) \\
&= 4 \int_{T^2} w I_2^*(f \cdot I_1(wf)) \cdot I_1^*(f \cdot I_2(wf)).
\end{aligned}$$

Using (4.5), we obtain the claim. ∎

The next results improve Lemma 4.1 in [4].

Lemma 4.10 (Small energy majorization on bi-tree). *Let T^2 be a 2-tree, and let $f: T^2 \to [0,\infty)$ be a function that is superadditive in each parameter separately. Let $w: T^2 \to [0,\infty)$ be of tensor product form. Suppose that* supp $f \subseteq \{\mathbb{I}(wf) \le \delta\}$. *Let $\lambda \ge 4\delta$. Then there exists $\varphi: T^2 \to [0,\infty)$ such that*

$$\text{a) } \mathbb{I}w\varphi \ge \mathbb{I}wf, \text{ where } \mathbb{I}wf \in [\lambda, 2\lambda], \quad \text{b) } \int_{T^2} w\varphi^2 \le C\frac{\delta^2}{\lambda^2}\int_{T^2} wf^2,$$

where C is an absolute constant.

Proof. Since $2\delta\lambda^{-1} \le 1/2$, we have

$$(\mathbb{I} f)\mathbf{1}_{\mathbb{I} f \ge \lambda} \le 4\lambda^{-1}\mathbb{I}(I_1 f \cdot I_2 f)$$

And thus

$$(\mathbb{I} f)\mathbf{1}_{\lambda \le \mathbb{I} f \le 2\lambda} \le 4\lambda^{-1}\mathbb{I}(I_1 f \cdot I_2 f)\mathbf{1}_{\lambda \le \mathbb{I} f \le 2\lambda} \le 4\lambda^{-1}\mathbb{I}(I_1 f \cdot I_2 f \cdot \mathbf{1}_{\mathbb{I} f \le 2\lambda}).$$

Put

$$\varphi := 4\lambda^{-1}(I_1 f \cdot I_2 f \cdot \mathbf{1}_{\mathbb{I} f \le 2\lambda})$$

Then φ does already satisfy condition a) of the statement of the lemma. Now apply Lemma 4.9 to see that condition b) of the statement of the lemma is satisfied as well. ∎

4.2.2. The lack of maximum principle and the capacity of bad sets. In [4] we proved the analogous small energy majorization statement on the bi-tree T^2 but with δ/λ in the right-hand side of b).

Let us see why we care. Let μ be a measure on ∂T^2 and let its potential $\mathbb{V}^\mu \le 1$ on $\operatorname{supp}\mu$. In the "usual" potential theory, the maximum principle would imply that the potential $\mathbb{V}^\mu \le 1$ everywhere (or at least that $\mathbb{V}^\mu \le C$ with absolute constant C, see [1]).

This is not true for potential theory on multi-trees. The reader can find the counter-examples in [4].

The natural question arises: given $\lambda \gg 1$, what is the size of the set $\{\mathbb{V}^\mu \ge \lambda\}$? Let us introduce the usual notion of capacity on T^2. Given a set E, we consider all φ such that $\mathbb{I}\varphi \ge 1$ on E and

$$\operatorname{cap}(E) := \inf \int_{T^2} \varphi^2,$$

where the infimum is taken over such φ. So one would like to estimate the capacity $\operatorname{cap}(\{\mathbb{V}^\mu > \lambda\})$ of the bad set in terms of λ, if $\mathbb{V}^\mu \le 1$ on $\operatorname{supp}\mu$.

Theorem 4.11. *Let us be on T^2, and suppose $\mathbb{V}^\mu \le 1$ on $\operatorname{supp}\mu$. Then*

$$cap(\{\mathbb{V}^\mu > \lambda\}) \le \frac{C\,\mathcal{E}[\mu]}{\lambda^4}$$

for $\lambda \ge 1$, where C is an absolute constant.

Proof. Consider $f = \mathbb{I}^*\mu$, $\delta = 1$. If $f(\alpha) \ne 0$, then there is $\beta \le \alpha$ such that $\beta \in \operatorname{supp}\mu$. But then, by assumption, $\mathbb{I} f(\beta) = \mathbb{I}\mathbb{I}^*\mu(\beta) = \mathbb{V}^\mu(\beta) \le 1$. By the monotonicity of $\mathbb{I}$, we have that $\mathbb{I} f(\alpha) \le 1$. Hence

$$\operatorname{supp} f \subset \{\mathbb{I} f \le \delta = 1\},$$

and we are in the assumptions of the small energy majorization Lemma 4.10 on the bi-tree. We apply it with data $(f, \delta = 1, \lambda := 2^m\lambda)$ to get functions φ_m, $m = 0, 1, \ldots$, such that

$$\mathbb{I}\varphi_m \ge \mathbb{I} f = \mathbb{V}^\mu, \quad \text{where } \mathbb{V}^\mu \in [2^m\lambda, 2^{m+1}\lambda],$$

which means that

$$2^{-m}\lambda^{-1}\mathbb{I}\varphi_m \ge 1, \quad \text{where } \mathbb{V}^\mu \in [2^m\lambda, 2^{m+1}\lambda],$$

On the other hand, putting $\varphi := \sum_m 2^{-m}\lambda^{-1}\varphi_m$, we get firstly

$$\mathbb{I}\varphi \ge 1, \quad \text{where } \mathbb{V}^\mu \in [\lambda, \infty),$$

and secondly,

$$\int \varphi^2 \le \Big(\lambda^{-1}\sum_m 2^{-m}\big(\int_{T^2}\varphi_m^2\big)^{1/2}\Big)^2$$
$$\le C\,\Big(\lambda^{-1}\sum_m \lambda^{-1}2^{-2m}\big(\int_{T^2} f^2\big)^{1/2}\Big)^2 \le C'\,\lambda^{-4}\int_{T^2} f^2.$$

As $f = \mathbb{I}^*\mu$, we have $\int_{T^2} f^2 = \int_{T^2}\mathbb{I}^*\mu\mathbb{I}^*\mu = \int_{T^2}\mathbb{I}\mathbb{I}^*\mu\,d\mu = \int_{T^2}\mathbb{V}^\mu\,d\mu = \mathcal{E}[\mu]$, which proves the theorem. ∎

Remark 4.12. We do not know how precise is the rate λ^{-4} in Theorem 4.11. We do not even know whether the sharp rate should be polynomial or exponential. What we do know (see [4]) is that for any large λ there exists a measure μ, such that $\mathbb{V}^{\mu} \le 1$ on $\operatorname{supp}\mu$ but such that, with an absolute positive constant c, the following holds:

$$\operatorname{cap}(\{\mathbb{V}^{\mu} > \lambda\}) \ge c e^{-2\lambda}\,. \tag{4.6}$$

4.2.3. Continuation of energy estimates.

Lemma 4.13. *Let μ and ρ be positive measures on T^2, and let $\delta > 0$. Let $w\colon T^2 \to [0,\infty)$ be of tensor product form. Then*

$$\Big(\int \mathbb{V}_{\delta}^{\mu}\,\mathrm{d}\rho\Big)^4 \le 28\cdot\delta^2\,\mathcal{E}_{\delta}[\mu]\,\mathcal{E}[\rho]\,|\rho|^2.$$

Proof. Let $f := \mathbf{1}_{\mathbb{V}^{\mu}\le\delta}\mathbb{I}^*\mu$. Then

$$\begin{aligned}\int \mathbb{V}_{\delta}^{\mu}\,\mathrm{d}\rho &= \int \mathbb{I}(wf)\,\mathrm{d}\rho \le |\rho|^{1/2}\Big(\int (\mathbb{I}(wf))^2\,\mathrm{d}\rho\Big)^{1/2}\\ &\overset{\text{Lemma 4.6}}{\le} |\rho|^{1/2}\Big(2\int \mathbb{I}(I_1(wf)\cdot I_2(wf) + (wf)\cdot\mathbb{I}(wf))\,\mathrm{d}\rho\Big)^{1/2}\\ &= 2^{1/2}|\rho|^{1/2}\Big(\int w(I_1(w_1 f)\cdot I_2(w_2 f) + f\cdot\mathbb{I}(wf))\mathbb{I}^*\rho\Big)^{1/2}\\ &\le 2^{1/2}|\rho|^{1/2}\,\mathcal{E}[\rho]^{1/4}\Big(\int w(I_1(w_1 f)\cdot I_2(w_2 f) + f\cdot\mathbb{I}(wf))^2\Big)^{1/4};\end{aligned}$$

and expanding the square and using Lemma 4.8 and Lemma 4.9,

$$\begin{aligned}&\le 2^{1/2}|\rho|^{1/2}\,\mathcal{E}[\rho]^{1/4}\Big(7\delta^2\int wf^2\Big)^{1/4}\\ &= 28^{1/4}|\rho|^{1/2}\,\mathcal{E}[\rho]^{1/4}\,\delta^{1/2}\,\mathcal{E}_{\delta}[\mu]^{1/4}.\end{aligned}$$ ∎

4.3. 3-trees

Similarly to Lemma 4.6, we obtain the following result for 3-trees.

Lemma 4.14. *Let T^3 be a 3-tree, and let $f, g\colon T^3 \to [0,\infty)$. Then*

$$(\mathbf{I}f)(\mathbf{I}g) \le \mathbf{I}\Big(\sum_{A\subseteq\{1,2,3\}} I_A f\cdot I_{A^c} g\Big),$$

where $I_A = \prod_{i\in A} I_i$.

Corollary 4.15. *Let $0 < \delta \le \lambda/4$. Let $f\colon T^3 \to [0,\infty)$ with $\operatorname{supp} f \subseteq \{\mathbf{I}f \le \delta\}$. Then*

$$(\mathbf{I}f)\,\mathbf{1}_{\lambda\le\mathbf{I}f\le 2\lambda} \le 4\lambda^{-1}\,\mathbf{I}\Big(\sum_{i\in\{1,2,3\}} I_i f\cdot I_{(i)} f\cdot\mathbf{1}_{\mathbf{I}f\le 2\lambda}\Big),$$

where $I_{(i)} = \prod_{j\ne i} I_j$.

Proof. Substituting $f = g$, Lemma 4.14 implies that

$$(\mathbf{I}f)^2 \le \mathbf{I}\Big(2 \sum_{i \in \{1,2,3\}} I_i f \cdot I_{(i)} f + 2f \cdot \mathbf{I}f\Big).$$

Using the support condition, this implies

$$\begin{aligned}(\mathbf{I}f)\mathbf{1}_{\lambda \le \mathbf{I}f \le 2\lambda} &\le \lambda^{-1}(\mathbf{I}f)^2\, \mathbf{1}_{\lambda \le \mathbf{I}f \le 2\lambda} \le \lambda^{-1}\,\mathbf{I}\Big(2 \sum_{i \in \{1,2,3\}} I_i f \cdot I_{(i)} f + 2\delta f\Big) \\ &\le \lambda^{-1}\,\mathbf{I}\Big(2 \sum_{i \in \{1,2,3\}} I_i f \cdot I_{(i)} f\Big) + 2\delta\,\lambda^{-1}\,\mathbf{I}f.\end{aligned}$$

Since $2\delta\lambda^{-1} \le 1/2$, this implies

$$\begin{aligned}(\mathbf{I}f)\, \mathbf{1}_{\lambda \le \mathbf{I}f \le 2\lambda} &\le 2\lambda^{-1}\,\mathbf{I}\Big(2 \sum_{i \in \{1,2,3\}} I_i f \cdot I_{(i)} f\Big)\mathbf{1}_{\lambda \le \mathbf{I}f \le 2\lambda} \\ &\le 2\lambda^{-1}\,\mathbf{I}\Big(2 \sum_{i \in \{1,2,3\}} I_i f \cdot I_{(i)} f \cdot \mathbf{1}_{\mathbf{I}f \le 2\lambda}\Big).\end{aligned}$$

■

4.3.1. Energy bound.

Lemma 4.16. *Let $f\colon T^3 \to [0,\infty)$ be superadditive. Let $w\colon T^3 \to [0,\infty)$ be a tensor product. Suppose that* $\operatorname{supp} f \subseteq \{\mathbf{I}(wf) \le \delta\}$. *Then*

$$\int w(I_1(w_1 f) \cdot I_2 I_3(w_2 w_3 f))^2\, \mathbf{1}_{\mathbf{I}(wf) \le \lambda} \le 2\delta\,\lambda \int w\, f^2.$$

Proof. By Lemma 4.1, we have

$$\begin{aligned}\int & w(I_1(w_1 f) \cdot I_2 I_3(w_2 w_3 f))^2\, \mathbf{1}_{\mathbf{I}(wf) \le \lambda} \\ &\le 2 \int w I_1(w_1 f \cdot I_1(w_1 f)) \cdot (I_2 I_3(w_2 w_3 f))^2\, \mathbf{1}_{\mathbf{I}(wf) \le \lambda} \\ &= 2 \int I_1(w_1 f \cdot I_1(wf)) \cdot (I_2 I_3(w_2 w_3 f)) \cdot (I_2 I_3(wf))\, \mathbf{1}_{\mathbf{I}(wf) \le \lambda} \\ &= 2 \int w_1 f \cdot I_1(wf) \cdot I_1^*\big((I_2 I_3(w_2 w_3 f)) \cdot (I_2 I_3(wf))\, \mathbf{1}_{\mathbf{I}(wf) \le \lambda}\big).\end{aligned} \tag{4.7}$$

By Corollary 4.5, we have

$$\begin{aligned}I_1^*\big(&(I_2 I_3(w_2 w_3 f)) \cdot (I_2 I_3(wf))\mathbf{1}_{\mathbf{I}(wf) \le \lambda}\big) \\ &\le I_1^*\big(\Delta_1(\mathbf{1}_{\mathbf{I}(wf) \le \lambda} \cdot I_2 I_3(w_2 w_3 f)) \cdot I_1(I_2 I_3(wf))\big).\end{aligned}$$

Since $\{\mathbf{I}(wf) \le \lambda\}$ is an up-set and f is superadditive in the first coordinate, we have $\Delta_1(\mathbf{1}_{\mathbf{I}(wf) \le \lambda} \cdot I_2 I_3(w_2 w_3 f)) \ge 0$, and $I_1(I_2 I_3 wf) = \mathbf{I}wf \le \lambda$ on the support of the former function. Hence,

$$\begin{aligned}I_1^*\big((I_2 I_3(w_2 w_3 f)) \cdot (I_2 I_3(wf))\mathbf{1}_{\mathbf{I}(wf) \le \lambda}\big) &\le I_1^*\big(\Delta_1(\mathbf{1}_{\mathbf{I}(wf) \le \lambda} \cdot I_2 I_3(w_2 w_3 f)) \cdot \lambda\big) \\ &= \lambda\, \mathbf{1}_{\mathbf{I}(wf) \le \lambda} \cdot I_2 I_3(w_2 w_3 f).\end{aligned}$$

Using this bound, we obtain

$$\begin{aligned}(4.7) \le 2\lambda \int w_1 f \cdot I_1(wf) \cdot I_2 I_3(w_2 w_3 f) &= 2\lambda \int f \cdot I_1(wf) \cdot I_2 I_3(wf)\\ &= 2\lambda \int wf \cdot I_1^*(f \cdot I_2 I_3(wf)).\end{aligned}$$

As in (4.5), we see that

$$I_1^*(f \cdot I_2 I_3(wf)) \le \delta f.$$

This implies the conclusion of the lemma. ■

Compare the next result with Lemma 4.10.

Lemma 4.17 (Small energy majorization on the tri-tree). *Let T^3 be a 3-tree, and let $f: T^3 \to [0,\infty)$ be a function that is superadditive in each parameter separately. Let $w: T^3 \to [0,\infty)$ be of tensor product form. Suppose that* $\operatorname{supp} f \subseteq \{\mathbf{I}(wf) \le \delta\}$. *Let $\lambda \ge 4\delta$. Then there exists $\varphi: T^3 \to [0,\infty)$ such that*

$$\text{a) } \mathbf{I}(w\varphi) \ge \mathbf{I}(wf), \text{ where } \mathbf{I}(wf) \in [\lambda, 2\lambda], \quad \text{b) } \int_{T^3} w\varphi^2 \le C\frac{\delta}{\lambda}\int_{T^3} wf^2,$$

where C is an absolute constant.

Proof. Since $2\delta\lambda^{-1} \le 1/2$, we have

$$\begin{aligned}(\mathbf{I}wf)\,\mathbf{1}_{\lambda \le \mathbf{I}f \le 2\lambda} &\le 2\lambda^{-1}\mathbf{I}\Big(2\sum_{i\in\{1,2,3\}} I_i w_i f \cdot I_{(i)} w_{(i)} f\Big)\mathbf{1}_{\lambda \le \mathbf{I}f \le 2\lambda}\\ &\le 2\lambda^{-1}\mathbf{I}\Big(2\sum_{i\in\{1,2,3\}} I_i w_i f \cdot I_{(i)} w_{(i)} f \cdot \mathbf{1}_{\mathbf{I}f \le 2\lambda}\Big).\end{aligned}$$

Put

$$\varphi := 2\lambda^{-1}\Big(2\sum_{i\in\{1,2,3\}} I_i w_i f \cdot I_{(i)} w_{(i)} f \cdot \mathbf{1}_{\mathbf{I}f \le 2\lambda}\Big).$$

Then we have just seen that a) is satisfied. To prove b), just apply Lemma 4.16. ■

4.3.2. The lack of maximum principle and the capacity of bad sets. The reader can compare this subsection with Subsection 4.2.2.

Let μ be a measure on ∂T^3 and let its potential $\mathbf{V}^\mu \le 1$ on $\operatorname{supp}\mu$. As we already mentioned, in the "usual" potential theory, the maximum principle would imply that potential $\mathbb{V}^\mu \le 1$ everywhere (or at least that $\mathbf{V}^\mu \le C$ with absolute constant C, see [1]).

As we also already mentioned, see Subsection 4.2.2, this is not true for potential theory on multi-trees.

A natural question arises: given $\lambda \gg 1$, what is the size of the set $\{\mathbf{V}^\mu \ge \lambda\}$? Let us introduce the usual notion of capacity on T^3. Given a set E, we consider all φ such that $\mathbf{I}\varphi \ge 1$ on E and

$$\operatorname{cap}(E) := \inf \int_{T^3} \varphi^2,$$

where the infimum is taken over such φ. So one would like to estimate the capacity of the bad set $\operatorname{cap}(\{\mathbf{V}^\mu > \lambda\})$ in terms of λ, if $\mathbf{V}^\mu \le 1$ on $\operatorname{supp}\mu$.

Theorem 4.18. *Let us be on T^3, and let $\mathbf{V}^\mu \le 1$ on $\operatorname{supp}\mu$. Then*

$$cap(\{\mathbb{V}^\mu > \lambda\}) \le \frac{C\,\mathcal{E}[\mu]}{\lambda^3}$$

for $\lambda \ge 1$, where C is an absolute constant.

Proof. Consider $f = \mathbf{I}^*\mu, \delta = 1$. If $f(\alpha) \ne 0$, then there is $\beta \le \alpha$ such that $\beta \in \operatorname{supp}\mu$. But then, by assumption,

$$\mathbf{I}f(\beta) = \mathbf{II}^*(\beta) = \mathbf{V}^\mu(\beta) \le 1.$$

By the monotonicity of $\mathbf{I}$, we have that $\mathbf{I}f(\alpha) \le 1$. Hence

$$\operatorname{supp} f \subset \{\mathbf{I}f \le \delta = 1\},$$

and we are in the assumptions of the small energy majorization Lemma 4.17 on the tri-tree. We apply it with data $(f, \delta = 1, \lambda := 2^m\lambda)$ to finish the proof in exactly the same manner as this has been done in Theorem 4.11. ∎

Remark 4.19. We do not know how precise is the rate λ^{-3} in Theorem 4.18. The lower bound (4.6) still applies, since the Cartesian product of a 2-tree and a singleton is a 3-tree, but there is a lot of room between the upper and the lower bound.

4.3.3. Continuation of energy estimates. The next result is a version of (4.2) for 3-trees. The proof closely follows that of Lemma 4.1 in [4].

Lemma 4.20. *Let μ and ρ be positive measures on T^3, and let $\delta > 0$. Let $w : T^3 \to [0,\infty)$ be of tensor product form. Then*

$$\Big(\int \mathbf{V}_\delta^\mu \,\mathrm{d}\rho\Big)^3 \lesssim \delta\, \mathcal{E}_\delta[\mu]\, \mathcal{E}[\rho] |\rho|. \tag{4.8}$$

Proof. Without loss of generality, $\mathcal{E}_\delta[\mu] \ne 0$ and $\rho \not\equiv 0$. Let $\lambda > 0$, to be chosen later.

Let $f := \mathbf{I}^*\mu \cdot \mathbf{1}_{\mathbf{V}^\mu \le \delta}(\alpha)$. This function is superadditive. Also, $\mathbf{I}(wf) = \mathbf{V}_\delta^\mu \le \mathbf{V}^\mu \le \delta$ on $\operatorname{supp} f$, and $\mathcal{E}_\delta[\mu] = \int wf^2$.

For $m = 0, 1, \dots$ let

$$\phi_m := 4(2^m\lambda)^{-1}\Big(\sum_{i\in\{1,2,3\}} I_i(w_i f)\cdot I_{(i)}(w_{(i)} f)\cdot \mathbf{1}_{\mathbf{I}(wf)\le 2^{m+1}\lambda}\Big).$$

Then, by Corollary 4.15 with wf in place of f, we have

$$\mathbf{I}(wf)\cdot \mathbf{1}_{2^m\lambda < \mathbf{I}(wf) \le 2^{m+1}\lambda} \le \mathbf{I}(w\phi_m),$$

and, by Lemma 4.16, we have

$$\int w\phi_m^2 \lesssim \frac{\delta}{2^m\lambda}\int wf^2.$$

Hence,

$$\begin{aligned}\int \mathbf{V}_\delta^\mu \, \mathrm{d}\rho &= \int_{\{\mathbf{V}_\delta^\mu \le \lambda\}} \mathbf{V}_\delta^\mu \, \mathrm{d}\rho + \sum_{m=0}^\infty \int_{\{2^m\lambda < \mathbf{V}_\delta^\mu \le 2^{m+1}\lambda\}} \mathbf{V}_\delta^\mu \, \mathrm{d}\rho \\ &\le \lambda|\rho| + \sum_{m=0}^\infty \int \mathbf{I}(w\phi_m) \, \mathrm{d}\rho = \lambda|\rho| + \sum_{m=0}^\infty \int w\,\phi_m \mathbf{I}^* \, \mathrm{d}\rho \\ &\le \lambda|\rho| + \sum_{m=0}^\infty \Big(\int w\phi_m^2\Big)^{1/2} \mathcal{E}[\rho]^{1/2} \le \lambda|\rho| + \sum_{m=0}^\infty C(\delta/(2^m\lambda))^{1/2} \mathcal{E}_\delta[\mu]^{1/2}\mathcal{E}[\rho]^{1/2} \\ &\le \lambda|\rho| + C(\delta/\lambda)^{1/2}\,\mathcal{E}_\delta[\mu]^{1/2}\,\mathcal{E}[\rho]^{1/2}.\end{aligned}$$

Substituting $\lambda = (\delta\,\mathcal{E}_\delta[\mu]\,\mathcal{E}[\rho])^{1/3}\,|\rho|^{-2/3}$, we obtain (4.8). ∎

Corollary 4.21. *Let μ and ρ be positive measures on T^3, and let $\delta > 0$. Then*

$$\int \mathbf{V}_\delta^\mu \, \mathrm{d}\rho \le C_{(4.8)}^{1/2}\,\delta^{1/2}\,\mathcal{E}[\mu]^{1/6}\,|\mu|^{1/6}\,\mathcal{E}[\rho]^{1/3}\,|\rho|^{1/3}.$$

Proof. By Lemma 4.20 and Theorem 4.23, we have

$$\Big(\int \mathbf{V}_\delta^\mu \, \mathrm{d}\rho\Big)^3 \le C_{(4.8)}\,\delta\,\mathcal{E}_\delta[\mu]\,\mathcal{E}[\rho]\,|\rho| \le C_{(4.8)}\,\delta\,\big(C_{(4.8)}\,\delta\,\mathcal{E}[\mu]\,|\mu|\big)^{1/2}\,\mathcal{E}[\rho]\,|\rho|$$ ∎

4.4. *d*-trees

We say that a weight w satisfies the *surrogate maximum principle* if, for some $\kappa > 0$, $C < \infty$, and every positive functions $\mu, \rho \colon T^d \to [0,\infty)$ and $\delta > 0$, we have

$$\int \mathbf{V}_\delta^\mu \, \mathrm{d}\rho \le C\,(\delta|\rho|)^\kappa\,(\mathcal{E}_\delta[\mu]\,\mathcal{E}[\rho])^{(1-\kappa)/2}. \tag{4.9}$$

When $d \in \{1,2,3\}$, every weight w of tensor product form satisfies the surrogate maximum principle with $\kappa = 1/d$ and C independent of w. For $d = 1$, this follows from the maximum principle (4.1). For $d = 2$, this holds by Lemma 4.13, and for $d = 3$, by Lemma 4.20. This leads us to the following conjecture.

Conjecture 4.22 (Surrogate maximum principle). *Let $w : T^d \to [0,\infty)$ be of tensor product form. Then w satisfies the surrogate maximum principle with $\kappa = 1/n$ and $C = C(n)$ independent of w.*

In what follows, we will work conditionally on the surrogate maximum principle. All implicit constants are allowed to depend on κ, C in (4.9), but not otherwise on w. In particular, our results hold unconditionally for w of tensor product form if $n \in \{1,2,3\}$.

Taking $\rho = \mu$ in (4.9), we obtain Lemma 4.23 below.

Lemma 4.23. *Let $w \colon T^d \to [0,\infty)$ be such that the surrogate maximum principle (4.9) holds. Let μ be a positive measure on T^d, and let $\delta > 0$. Then*

$$\int \mathbf{V}_\delta^\mu \, \mathrm{d}\mu \le C_{(4.9)}^{2/(1+\kappa)}\,(\delta|\mu|)^{2\kappa/(1+\kappa)}\,\mathcal{E}[\mu]^{(1-\kappa)/(1+\kappa)}. \tag{4.10}$$

When using Lemma 4.23, we can also denote $2\kappa/(1+\kappa)$ by the letter κ again, which proves (4.9) for $d = 1, 2, 3$.

Conjecture 4.24. *For all positive integers d,*

$$\int \mathbf{V}_\delta^\mu \, \mathrm{d}\mu \le C_d \, (\delta|\mu|)^{2/(d+1)} \, \mathcal{E}[\mu]^{(d-1)/(d+1)}.$$

Notice that for $d = 1$ we have the best possible estimate, it is linear in δ. For $d = 2$, we proved above the estimate with $\delta^{2/3}$. To our big surprise, we managed to improve this result: in [24] we have proved the estimate $\le C_\tau(\delta|\mu|)^{1-\tau}\mathcal{E}[\mu]^\tau$ for any $\tau > 0$. And we can prove that the estimate with $\tau = 0$ is false for $d = 2$, see [26].

5. Carleson condition implies hereditary Carleson condition

For an arbitrary set $E \subseteq T^d$, let

$$\mathcal{E}_E[\mu] := \int_E w(\mathbf{I}^*\mu)^2.$$

Lemma 5.1. *Let $w\colon T^d \to [0,\infty)$ be such that the surrogate maximum principle (4.9) holds. Let $\nu\colon T^d \to [0,\infty)$ and define*

$$E := \left\{\mathbf{V}^\nu > (2C_{(4.10)})^{-1/\kappa} \frac{\mathcal{E}[\nu]}{|\nu|}\right\} \subseteq T^d.$$

Then

$$\mathcal{E}_E[\nu] := \sum_{\alpha\in E} w(\alpha)(\mathbf{I}^*\nu(\alpha))^2 \ge \frac{1}{2}\mathcal{E}[\nu].$$

Proof. Put $\delta := (2C_{(4.10)})^{-1/\kappa}\,\mathcal{E}[\nu]/|\nu|$. By Lemma 4.23, we have

$$\mathcal{E}_E[\nu] = \mathcal{E}[\nu] - \mathcal{E}_\delta[\nu] \ge \mathcal{E}[\nu] - C_{(4.10)}\,(\delta|\nu|)^\kappa\,\mathcal{E}[\nu]^{1-\kappa} = \mathcal{E}[\nu]/2,$$

and the claim follows. ∎

Theorem 5.2. *Let $w\colon T^d \to [0,\infty)$ be such that the surrogate maximum principle (4.9) holds. Then, for every $\mu\colon T^d \to [0,\infty)$, we have*

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

Proof. Without loss of generality, $[w,\mu]_{\mathrm{Car}} = 1$. Let

$$A := [w,\mu]_{\mathrm{HC}} = \sup_{E\subseteq T^d,\mu(E)\neq 0} \frac{\mathcal{E}[\mu\mathbf{1}_E]}{\mu(E)}. \tag{5.1}$$

Since T^d is finite, the constant A is finite, and there exists a maximizer E for (5.1). Let $\nu := \mu\mathbf{1}_E$ and set

$$\mathcal{D} := \{\mathbf{V}^\nu > cA\}$$

with a small constant c. Then, by Lemma 5.1, we have

$$\mathcal{E}_{\mathcal{D}}[\nu] \geq \frac{1}{2}\,\mathcal{E}[\nu]\,.$$

Hence, $0 < \mathcal{E}[\nu] \leq 2\mathcal{E}_{\mathcal{D}}[\nu] \leq 2\mathcal{E}_{\mathcal{D}}[\mu] \leq 2\mu(\mathcal{D})$. In particular, $\mu(\mathcal{D}) \neq 0$.

By definition, we have $\mathbf{V}^{\nu} > cA$ on $\mathcal{D}$, and therefore

$$cA\mu(\mathcal{D}) \leq \int_{\mathcal{D}} \mathbf{V}^{\nu}\,\mathrm{d}\mu \leq \mathcal{E}[\nu]^{1/2}\,\mathcal{E}[\mu\mathbf{1}_{\mathcal{D}}]^{1/2} \leq (2\mu(\mathcal{D}))^{1/2}(A\mu(\mathcal{D}))^{1/2}.$$

It follows that $A \lesssim 1$. ■

6. Hereditary Carleson condition implies Carleson embedding

Theorem 6.1. *Let $w\colon T^d \to [0,\infty)$ be such that the surrogate maximum principle (4.9) holds. Let μ and ρ be positive measures on T^d with*

$$[w,\mu]_{\mathrm{HC}} \leq 1 \quad \textit{and} \quad [w,\rho]_{\mathrm{HC}} \leq 1. \tag{6.1}$$

Then, for some $\kappa' > 0$, we have

$$\int \mathbf{V}^{\mu}\,\mathrm{d}\rho \lesssim |\mu|^{1/2-\kappa'}|\rho|^{1/2+\kappa'}.$$

Remark 6.2. This improves upon the estimate

$$\int \mathbf{V}^{\mu}\,\mathrm{d}\rho \leq \mathcal{E}[\mu]^{1/2}\,\mathcal{E}[\rho]^{1/2} \lesssim |\mu|^{1/2}\,|\rho|^{1/2}$$

that is immediate by Cauchy–Schwarz and the Carleson condition.

Proof. Let $\delta > 0$, to be chosen later. By (4.9) and (6.1), we obtain

$$\int \mathbf{V}^{\mu}_{\delta}\,\mathrm{d}\rho \lesssim \delta^{\kappa}|\mu|^{(1-\kappa)/2}|\rho|^{(1+\kappa)/2}.$$

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

$$\int (\mathbf{V}^{\mu} - \mathbf{V}^{\mu}_{\delta})\,\mathrm{d}\rho = \int_E w\mathbf{I}^*\mu\,\mathbf{I}^*\rho \leq \mathcal{E}_E[\mu]^{1/2}\,\mathcal{E}_E[\rho]^{1/2} \leq \mu(E)^{1/2}\,\mathcal{E}[\rho]^{1/2}.$$

Note that

$$\delta\mu(E) \leq \int_E \mathbf{V}^{\mu}\,\mathrm{d}\mu \leq \mathcal{E}[\mu]^{1/2}\,\mathcal{E}[\mu\mathbf{1}_E]^{1/2} \leq \mathcal{E}[\mu]^{1/2}\,\mu(E)^{1/2}$$

by definition (1.10) of the hereditary Carleson constant. Hence,

$$\mu(E)^{1/2} \leq \delta^{-1}\mathcal{E}[\mu]^{1/2},$$

and it follows that

$$\int (\mathbf{V}^\mu - \mathbf{V}^\mu_\delta)\,\mathrm{d}\rho \le \delta^{-1}\mathcal{E}[\rho]^{1/2}\mathcal{E}[\mu]^{1/2}.$$

Hence,

$$\int \mathbf{V}^\mu\,\mathrm{d}\rho \le C\delta^\kappa |\mu|^{(1-\kappa)/2}|\rho|^{(1+\kappa)/2} + \delta^{-1}|\rho|^{1/2}|\mu|^{1/2}.$$

Optimizing in δ, we obtain

$$\int \mathbf{V}^\mu\,\mathrm{d}\rho \lesssim |\mu|^{\frac{1/2}{1+\kappa}}|\rho|^{\frac{1/2+\kappa}{1+\kappa}}.$$ ∎

Exactly as in Theorem 6.3 of [4], we can now prove the following result.

Theorem 6.3. *Let $w\colon T^d \to [0,\infty)$ be such that the surrogate maximum principle (4.9) holds. Then, for every $\mu\colon T^d \to [0,\infty)$, we have*

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

Alternatively, we can argue as follows. Let $\mu, \tilde{\mu}$ be measures on T^d with $[w,\mu]_{\mathrm{HC}} \le 1$ and $[w,\tilde{\mu}]_{\mathrm{HC}} \le 1$. By Theorem 6.1 and the definition of the hereditary Carleson constant (1.10), the positive bilinear map

$$(\psi,\tilde{\psi}) \mapsto \int w(\mathbf{I}^*\psi\mu)(\mathbf{I}^*\tilde{\psi}\tilde{\mu}) \tag{6.2}$$

is bounded on $L^{p,1}(\mu)\times L^{p',1}(\tilde{\mu})$ and on $L^{p',1}(\mu)\times L^{p,1}(\tilde{\mu})$, where $1/p = 1/2-\kappa'$. By restricted type interpolation, it follows that the map (6.2) is also bounded on $L^2(\mu)\times L^2(\tilde{\mu})$. Theorem 6.3 arises in the case $\mu = \tilde{\mu}$, $\psi = \tilde{\psi}$.

7. Box condition implies hereditary Carleson condition

7.1. Main estimate

Define

$$\mathbf{V}^\nu_P(\omega) := \sum_{Q:\,\omega\le Q\le P} w(Q)\,\mathbf{I}^*\nu(Q), \tag{7.1}$$

$$\mathbf{V}^\mu_{\varepsilon',\mathrm{good}}(\omega) := \sum_{P\ge\omega:\,\mathbf{V}_P(\omega)>\varepsilon'} (w\mathbf{I}^*\mu)(P).$$

Lemma 7.1. *Let $n \ge 2$ and $\mu\colon T^d \to [0,\infty)$. Let $w\colon T^d \to [0,\infty)$ be such that the surrogate maximum principle (4.9) holds. Assume that $\mathcal{E}[\mu] \le |\mu|$ and that*

$$\mathbf{V}^\mu \ge 1/3 \quad \text{on } \operatorname{supp}\mu. \tag{7.2}$$

Then, if ε' is small enough, we have

$$\int \mathbf{V}^\mu_{\varepsilon',\mathrm{good}}\,\mathrm{d}\mu \gtrsim |\mu|.$$

Proof of Lemma 7.1. It suffices to show that, for some ε' and ε_{n-1}, we have

$$\mu\{\omega \in T^d \mathbf{V}^{\mu}_{\varepsilon',\mathrm{good}}(\omega) \geq \varepsilon_{n-1}\} \geq |\mu|/2.$$

Let $\varepsilon > 0$, to be chosen later, and define

$$\varepsilon_1 := \varepsilon, \quad \varepsilon_2 := \varepsilon\varepsilon_1^{1/\kappa}, \quad \varepsilon_3 := \varepsilon\varepsilon_2^{1/\kappa}, \ldots$$

By Lemma 4.23, we have

$$\int \mathbf{V}^{\mu}_{\varepsilon_j}\, \mathrm{d}\mu \lesssim \varepsilon_j^{\kappa}\, |\mu|^{\kappa}\, \mathcal{E}[\mu]^{1-\kappa} \lesssim \varepsilon_j^{\kappa} \int \mathrm{d}\mu$$

for some $\kappa > 0$. By Chebyshov's inequality, it follows that

$$\mathbf{V}^{\mu}_{\varepsilon_j}(\omega) \leq (\varepsilon_j/\varepsilon)^{\kappa}/10 \tag{7.3}$$

for a proportion $\geq (1 - C\varepsilon^{\kappa})$ of ω's. So we only consider ω's for which (7.3) holds for all $j = 1, \ldots, n-1$. Similarly, we may restrict to those ω's for which $\mathbf{V}^{\mu}(\omega) \lesssim 1$.

Let

$$\varepsilon' := \varepsilon \cdot \varepsilon_1 \cdots \varepsilon_{n-1}.$$

For a fixed ω, let

$$\mathcal{U} := \{Q \geq \omega \mid \mathbf{V}_Q(\omega) > \varepsilon'\} \tag{7.4}$$

and

$$\mathcal{W}_j := \{Q \geq \omega \mid \mathbf{V}^{\mu}(Q) \leq \varepsilon_j\}, \quad 1 \leq j \leq n-1.$$

For $p \in T^d$, write

$$\uparrow p := \{\alpha \in T^d \mid \alpha \geq p\}.$$

For $p \in \uparrow\omega$, let

$$\downarrow p := \{\alpha \in T^d \mid \omega \leq \alpha \leq p\}.$$

If $\mathcal{U} \not\subseteq \mathcal{W}_{n-1}$, then this means that there exists $p \notin \mathcal{W}_{n-1}$ with $\uparrow p \subseteq \mathcal{U}$. Hence,

$$\mathbf{V}^{\mu}_{\varepsilon',\mathrm{good}}(\omega) \geq \sum_{p' \in \uparrow p} w\,\mu(p') = \mathbf{V}^{\mu}(p) \geq \varepsilon_{n-1}.$$

Assume now that $\mathcal{U} \subseteq \mathcal{W}_{n-1}$. In this case, we will cover $\uparrow\omega \setminus \mathcal{W}_1$ by boundedly many sets of the form $\downarrow q$ with $q \in \uparrow\omega \setminus \mathcal{U}$. This will lead to a contradiction with (7.2), since, by (7.3) and (7.4), the integral of

$$f := w\,\mathbf{I}^*\mu$$

is small on $\mathcal{W}_1$ and on each such set $\downarrow q$.

For a set of coordinates $J \subseteq \{1, \ldots, n\}$ and a point $p \in T^d$, let

$$\uparrow_J p := \{q \in T^d \mid q_j \geq p_j \text{ for } j \in J,\ q_j = p_j \text{ for } j \notin J\}.$$

Given $J \subseteq \{1, \dots, n\}$ with $J \neq \emptyset$ and $p \in T^d$, we define a set $\mathcal{Q}_J(p) \subset T^d$ as follows. If $|J| = 1$, then $\mathcal{Q}_J(p)$ consists of the (unique) maximal element of $\uparrow_J p \setminus \mathcal{U}$, if the latter set is nonempty, and is empty otherwise. If $|J| \geq 2$, then $\mathcal{Q}_J(p)$ is a maximal set of maximal elements of $\uparrow_J p \setminus \mathcal{W}_{n-|J|+1}$ such that the sets $\uparrow_J q \setminus \mathcal{W}_{n-|J|+2}$ are pairwise disjoint for $q \in \mathcal{Q}_J(p)$.

Then, recursively, let $\mathcal{R}_\emptyset(p) := \{p\}$,

$$\mathcal{R}_J(p) := \bigcup_{J' \subset J} \bigcup_{p' \in \mathcal{Q}_J(p)} \mathcal{R}_{J'}(p'),$$

where the first union runs over all subsets of J with cardinality $|J'| = |J| - 1$.

We claim that, for every $p \in \uparrow\omega$ and every $J \subseteq \{1, \dots, n\}$ with $J \neq \emptyset$, we have

$$\bigcup_{p' \in \mathcal{R}_J(p)} \downarrow p' \supseteq \uparrow_J p \setminus \mathcal{W}_{n-|J|+1}, \tag{7.5}$$

where we set $\mathcal{W}_n := \mathcal{U}$ to simplify notation. We prove (7.5) by induction on $|J|$. For $|J| = 1$, the claim (7.5) obviously holds. Let now J with $|J| \geq 2$ be given, and suppose that (7.5) is known for all proper subsets of J. Let

$$\mathcal{D} := \bigcup_{p' \in \mathcal{R}_J(p)} \downarrow p', \quad \mathcal{P} := \uparrow_J p \setminus \mathcal{W}_{n-|J|+1}.$$

By the inductive hypothesis,

$$\mathcal{D} \supseteq \uparrow_{J'} p' \setminus \mathcal{W}_{n-|J|+2} \tag{7.6}$$

for every $p' \in Q_J(p)$ and every $J' \subsetneq J$. Suppose that

$$\mathcal{D} \not\supseteq \mathcal{P}. \tag{7.7}$$

Choose a maximal $q \in \mathcal{P} \setminus \mathcal{D}$. Since $\mathcal{D}$ is a down-set, q is also a maximal element of $\mathcal{P}$. We claim that

$$(\uparrow_J q \cap \uparrow_J p') \setminus \mathcal{W}_{n-|J|+2} = \emptyset \quad \text{for all } p' \in \mathcal{Q}_J(p). \tag{7.8}$$

Indeed, suppose for a contradiction that there exists $q' \in (\uparrow_J q \cap \uparrow_J p') \setminus \mathcal{W}_{n-|J|+2}$, and let q' be minimal with this property. Since $\mathcal{W}_{n-|J|+2}$ is an up-set, q' is also a minimal element of $\uparrow_J q \cap \uparrow_J p'$. Since $q, p' \in \uparrow_J p$, q' is in fact the coordinate-wise maximum of q, p'. Since q and p' are distinct maximal elements of $\mathcal{P}$, in fact q' coincides with p' in at least one coordinate, so $q' \in \uparrow_{J'} p'$ for some $J' \subsetneq J$. Now, (7.6) implies that $q' \in \mathcal{D}$, and, since $\mathcal{D}$ is a down-set and $q' \geq q$, also $q \in \mathcal{D}$, a contradiction.

Therefore, (7.8) holds. But this contradicts the maximality of $\mathcal{Q}_J(p)$. So the assumption (7.7) is false, and we obtain (7.5).

Let $p \geq \omega$. For $2 \leq |J| \leq n$, we have

$$1 \gtrsim \mathbf{V}^\mu(\omega) \geq \mathbf{V}^\mu(p) \geq \sum_{q \in \mathcal{Q}_J(p)} \int_{\uparrow_J q \setminus \mathcal{W}_{n-|J|+2}} f \geq \sum_{q \in \mathcal{Q}_J(p)} (\mathbf{I} f(q) - \mathbf{I}(f \mathbf{1}_{\mathcal{W}_{n-|J|+2}})(\omega))$$

and, by definition of $q \in \mathcal{W}_{n-|J|+1}$ and by (7.3), this is

$$\geq \sum_{q \in \mathcal{Q}_J(p)} (\varepsilon_{n-|J|+1} - (\varepsilon_{n-|J|+2}/\varepsilon)^{1/2}/10) \gtrsim |\mathcal{Q}_J(p)| \varepsilon_{n-|J|+1}.$$

It follows that

$$\varepsilon_1 \cdots \varepsilon_{n-1} |\mathcal{R}_{\{1,\dots,n\}}(\omega)| \lesssim 1.$$

Hence, by (7.5),

$$\mathbf{V}^\mu(\omega) - \mathbf{V}^\mu_\varepsilon(\omega) = \int_{\uparrow\omega \setminus \mathcal{W}_1} f \leq \sum_{p' \in \mathcal{R}_{\{1,\dots,n\}}(\omega)} \int_{\downarrow p'} f = \sum_{p' \in \mathcal{R}_{\{1,\dots,n\}}(\omega)} \mathbf{V}^\mu_{p'}(\omega)$$
$$\leq \varepsilon' |\mathcal{R}_{\{1,\dots,n\}}(\omega)| \lesssim \frac{\varepsilon'}{\varepsilon_1 \cdots \varepsilon_{n-1}} = \varepsilon.$$

Therefore, by (7.3),

$$1/3 \leq \mathbf{V}^\mu(\omega) = (\mathbf{V}^\mu(\omega) - \mathbf{V}^\mu_\varepsilon(\omega)) + \mathbf{V}^\mu_\varepsilon(\omega) \leq C\varepsilon + 1/10.$$

This inequality is false if ε is sufficiently small, contradicting the assumption $\mathcal{U} \subseteq \mathcal{W}_{n-1}$. ■

7.2. Box condition implies hereditary Carleson condition

We refer to Lemma 3.1 in [2] or Lemma 7.1 in [4] for the following lemma.

Lemma 7.2 (Balancing lemma). *Let $\nu \colon T^d \to [0,\infty)$ with*

$$\mathcal{E}[\nu] = \int \mathbf{V}^\nu \, \mathrm{d}\nu \geq A|\nu|.$$

Then there exists a down-set $\tilde{E} \subset T^d$ such that for the measure $\tilde{\nu} := \nu \mathbf{1}_{\tilde{E}}$ we have

$$\mathbf{V}^{\tilde{\nu}} \geq \frac{A}{3} \quad \text{on } \tilde{E} \quad \text{and} \quad \mathcal{E}[\tilde{\nu}] \geq \frac{1}{3} \mathcal{E}[\nu].$$

The next result contains the last missing inequality in Theorem 1.4.

Theorem 7.3. *Let $n \geq 2$. Let $w \colon T^d \to [0,\infty)$ be such that the surrogate maximum principle (4.9) holds. Then, for every $\nu \colon T^d \to [0,\infty)$, we have*

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

Proof. By scaling, we may assume $[w,\nu]_{\mathrm{Box}} = 1$ without loss of generality. Let $A := [w,\nu]_{\mathrm{HC}}$. Let $E \subset T^2$ be a subset such that $\mu = \nu \mathbf{1}_E \neq 0$ and $\mathcal{E}[\mu] = A|\mu|$ (such a subset exists because we assume that T^d is finite). By Lemma 7.2, there exists a further subset $\tilde{E} \subset T^2$ such that $\tilde{\mu} := \mu \mathbf{1}_{\tilde{E}}$ satisfies

$$\mathbf{V}^{\tilde{\mu}} \geq \frac{A}{3} \quad \text{on } \tilde{E}$$

and $\tilde{\mu} \neq 0$. Thus, replacing μ by $\tilde{\mu}$, we may assume $\mathbf{V}^\mu \geq A/3$ on $\operatorname{supp} \mu$.

By Lemma 7.1 applied with μ/A in place of μ, for sufficiently small $\varepsilon, \theta > 0$, we have

$$\int \mathbf{V}^{\mu}_{\varepsilon A,\mathrm{good}}\,\mathrm{d}\mu \geq 2\theta \mathcal{E}[\mu].$$

We claim that, with these values of ε and θ, we have

$$\mathcal{E}[\mu] \leq \frac{\theta}{1-\theta} \sum_{\alpha\,:\,\theta\varepsilon A \mathbf{I}^*\mu(\alpha) \leq \mathcal{E}_\alpha[\mu]} w(\alpha)(\mathbf{I}^*\mu(\alpha))^2. \tag{7.9}$$

Indeed, suppose that α is such that

$$\theta\,\varepsilon\,A\,\mathbf{I}^*\mu(\alpha) > \mathcal{E}_\alpha[\mu] = \sum_{\omega\leq\alpha} \mu(\omega)\,\mathbf{V}^{\mu}_{\alpha}(\omega), \quad \mathbf{V}^{\mu}_{\alpha}(\omega) = \sum_{\beta:\omega\leq\beta\leq\alpha} w(\beta)(\mathbf{I}^*\mu)(\beta),$$

where the latter definition is from (7.1). Then we have

$$\begin{aligned}\sum_{\omega\leq\alpha:\mathbf{V}^{\mu}_{\alpha}(\omega)\leq\varepsilon A} \mu(\omega) &= \mathbf{I}^*\mu(\alpha) - \sum_{\omega\leq\alpha:\mathbf{V}^{\mu}_{\alpha}(\omega)>\varepsilon A} \mu(\omega) \geq \mathbf{I}^*\mu(\alpha) - \frac{1}{\varepsilon A}\sum_{\omega\leq\alpha} \mathbf{V}^{\mu}_{\alpha}(\omega)\mu(\omega)\\ &\geq (1-\theta)\,\mathbf{I}^*\mu(\alpha).\end{aligned}$$

It follows that

$$\begin{aligned}\sum_{\alpha\,:\,\theta\varepsilon A\mathbf{I}^*\mu(\alpha)>\mathcal{E}_\alpha[\mu]} w(\alpha)(\mathbf{I}^*\mu(\alpha))^2 &\leq \sum_{\alpha} w(\alpha)\,\mathbf{I}^*\mu(\alpha)\,\frac{1}{1-\theta} \sum_{\omega\leq\alpha\,:\,\mathbf{V}^{\mu}_{\alpha}(\omega)\leq\varepsilon A} \mu(\omega)\\ &= \frac{1}{1-\theta}\sum_{\omega}\mu(\omega) \sum_{\alpha\geq\omega\,:\,\mathbf{V}^{\mu}_{\alpha}(\omega)\leq\varepsilon A} w(\alpha)\,\mathbf{I}^*\mu(\alpha)\\ &= \frac{1}{1-\theta}\sum_{\omega}\mu(\omega)\,(\mathbf{V}^{\mu}-\mathbf{V}^{\mu}_{\mathrm{good},\varepsilon A})(\omega) \leq \frac{1-2\theta}{1-\theta}\,\mathcal{E}[\mu].\end{aligned}$$

This implies the claim (7.9).

By Lemma 4.23 again, and since $\mathbf{V}^{\mu} \geq A/4$ on $\operatorname{supp}\mu$, we also have

$$\mathcal{E}_{c'A}[\mu] \lesssim (c'A)^{\kappa}\,|\mu|^{\kappa}\,\mathcal{E}[\mu]^{1-\kappa} \lesssim (c')^{\kappa}\,\mathcal{E}[\mu]. \tag{7.10}$$

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

$$\mathcal{E}[\mu] \lesssim \sum_{\alpha\in\mathcal{R}} w(\alpha)(\mathbf{I}^*\mu(\alpha))^2, \quad \mathcal{R} := \{\alpha\in T^d \mid \theta\varepsilon A\mathbf{I}^*\mu(\alpha)\leq\mathcal{E}_\alpha[\mu], \mathbf{V}^{\mu}(\alpha)\geq c'A\}.$$

For each $\alpha\in\mathcal{R}$, we have

$$\theta\,\varepsilon\,A\,\mathbf{I}^*\mu(\alpha) \leq \mathcal{E}_\alpha[\mu] \leq \mathcal{E}_\alpha[\nu] \leq [w,\nu]_{\mathrm{Box}}\,\mathbf{I}^*\nu(\alpha) = \mathbf{I}^*\sigma(\alpha),$$

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

$$A^2\,\mathcal{E}[\mu] \lesssim \mathcal{E}[\sigma]. \tag{7.11}$$

On the other hand, using the definition of A, the fact that $\mathbf{V}^{\mu} \gtrsim A$ on $\operatorname{supp}\sigma$, and the Cauchy–Schwarz inequality, we obtain

$$\mathcal{E}[\sigma] \leq A|\sigma| \lesssim \int \mathbf{V}^{\mu}\,\mathrm{d}\sigma \leq \mathcal{E}[\mu]^{1/2}\,\mathcal{E}[\sigma]^{1/2}. \tag{7.12}$$

From (7.12), we obtain $\mathcal{E}[\sigma] \lesssim \mathcal{E}[\mu]$, and inserting this into (7.11) gives $A \lesssim 1$. ∎

8. What we cannot prove

The main problem with pushing the results to d-trees, $d \geq 4$, lies with Lemma 4.10 and Lemma 4.17. Let us start with majorization on a simple dyadic tree. All trees below are big but finite. Let f and g be two non-negative functions on a simple dyadic tree T. As always, $If(v)$ means summing $f(u)$ "up" from v to root o.

Here is the analog of Lemma 4.10 and Lemma 4.17. The big difference of the lemma below is that it involves two functions: f and g. This is not the case for Lemma 4.10 and Lemma 4.17, that involve one function.

Lemma 8.1. *Let* $\operatorname{supp} f \subset \{Ig \leq \delta\}$. *Let g be a superadditive function. There exists $\varphi\colon T \to \mathbb{R}_+$ such that*

$$\text{a) } I\varphi(\omega) \geq If(\omega), \quad \forall \omega \in \partial T\colon Ig(\omega) \in [\lambda, 2\lambda], \tag{8.1}$$

$$\text{b) } \int_T \varphi^2 \leq C\frac{\delta}{\lambda}\int_T f^2. \tag{8.2}$$

Proof. Put

$$\varphi = \lambda^{-1} If \cdot g \cdot \mathbf{1}_{Ig \leq 4\lambda},$$

and see [4]. ∎

Now let us see what happens on the bi-tree T^2. As before, $\mathbb{I} f(v)$ means summing $f(u)$ "up" over all ancestors of v from v to root o. Notice that now a vertex may have two parents.

Conjecture 8.2. *Let* $\operatorname{supp} f \subset \{\mathbb{I}g \leq \delta\}$. *Let g be a function superadditive in its both variables separately. There exists $\varphi\colon T^2 \to \mathbb{R}_+$ such that*

$$\text{a) } \mathbb{I}\varphi(\omega) \geq \mathbb{I} f(\omega), \quad \forall \omega \in \partial T^2\colon \mathbb{I}g(\omega) \in [\lambda, 2\lambda]; \quad \text{b) } \int_{T^2} \varphi^2 \leq C\Big(\frac{\delta}{\lambda}\Big)^\tau \int_{T^2} f^2$$

for some positive τ.

By analogy with the previous section, one may think that given f, g on T^2 such that

$$\operatorname{supp} f \subset \{\mathbb{I}g \leq \delta\}$$

and having g (super)additive on T^2, one constructs φ as in Lemma 8.1 by the formula

$$\varphi = \lambda^{-1}\, \mathbb{I} f \cdot g \cdot \mathbf{1}_{\mathbb{I}g \leq 4\lambda}.$$

However, this is false.

What is true is the following: let $\operatorname{supp} f \subset \{\mathbb{I}g \leq \delta\}$ and let $\lambda \geq 10\delta$ and

$$\varphi := \lambda^{-1}(I_1 f \cdot I_2 g + I_1 g \cdot I_2 f + g \cdot \mathbb{I} f).$$

Then

$$\mathbb{I}(\mathbf{1}_{\mathbb{I}g \leq 2\lambda} \cdot \varphi) \geq \mathbb{I} f, \quad \text{where } \mathbb{I}g \in [\lambda, 2\lambda]. \tag{8.3}$$

So a) from the previous lemma can be generalized to the bi-tree with the following formula for φ:

$$\varphi = \lambda^{-1}(I_1 f \cdot I_2 g + I_1 g \cdot I_2 f + g \cdot \mathbb{I} f) \cdot \mathbf{1}_{\mathbb{I} g \le 2\lambda}. \tag{8.4}$$

The main difficulty in generalizing Lemma 8.1 to bi-trees is that we cannot prove b) of this lemma for the case of the bi-tree. This is because we have not a good estimate of $\int_{T^2} (\mathbb{I} f)^2 g^2$ via $\int_{T^2} f^2$ for g that is *separately superadditive* in both variables.

Notice that this hurdle is removed if $f = g$ because then

$$\mathbb{I}(\lambda^{-1} g \mathbb{I} f) = \mathbb{I}(\lambda^{-1} f \mathbb{I} f) \le \frac{\delta}{\lambda} \mathbb{I} f \le \frac{1}{10} \mathbb{I} f,$$

and we have another φ for majorization: $\tilde{\varphi} := c\lambda^{-1}(2I_1 f \cdot I_2 f)$, where $c = 10/9$. In fact, from (8.3) it now follows that

$$\mathbb{I}(\mathbf{1}_{\mathbb{I} f \le 2\lambda} \cdot \tilde{\varphi}) \ge \mathbb{I} f, \quad \text{where } \mathbb{I} f \in [\lambda, 2\lambda].$$

The analog of inequality b) of Lemma 8.1 $\equiv$ (8.2) on the bi-tree now follows from Lemma 4.9.

For the tri-tree we do not have the analog of Lemma 8.1 with two functions f, g, as we do not have it even on the bi-tree.

But similarly to (8.4), we can put

$$\begin{aligned} \varphi = \lambda^{-1}(&I_1 f \cdot \mathbb{I}_{23} g + I_2 f \cdot \mathbb{I}_{13} g + I_3 f \cdot \mathbb{I}_{12} g \\ &+ I_1 g f \cdot \mathbb{I}_{23} f + I_2 g \cdot \mathbb{I}_{13} f + I_3 g \cdot \mathbb{I}_{12} f + g \mathbf{I} f). \end{aligned} \tag{8.5}$$

Again this function φ satisfies

$$\mathbf{I}(\mathbf{1}_{\mathbf{I} \le 2\lambda} \cdot \varphi) \ge \mathbf{I} f, \quad \text{where } \mathbf{I} g \in [\lambda, 2\lambda], \tag{8.6}$$

which is the analog of a) of Lemma 8.1 (and the analog of (8.3)). However, we cannot prove the analog of b) of Lemma 8.1 for this function.

The main difficulty in generalizing Lemma 8.1 to tri-trees is that we cannot prove b) of this lemma on tri-tree. This misfortune happens by the same reason it happens on the bi-tree: we have not a good estimate of $\int_{T^3} (\mathbf{I} f)^2 g^2$ via $\int_{T^3} f^2$ for g that is *separately superadditive* in both variables.

Notice that this hurdle is removed if $f = g$ because then

$$\mathbf{I}(\lambda^{-1} g \mathbf{I} f) = \mathbf{I}(\lambda^{-1} f \mathbf{I} f) \le \frac{\delta}{\lambda} \mathbf{I} f \le \frac{1}{10} \mathbf{I} f,$$

and in place of φ from (8.5), we have another φ for majorization:

$$\tilde{\varphi} := c\lambda^{-1}(2I_1 f \cdot \mathbb{I}_{23} f + 2I_2 f \cdot \mathbb{I}_{13} f + 2I_3 f \cdot \mathbb{I}_{12} f),$$

where $c = 10/9$. In fact, from (8.6) it now follows that

$$\mathbb{I}(\mathbf{1}_{\mathbf{I} f \le 2\lambda} \cdot \tilde{\varphi}) \ge \mathbf{I} f, \quad \text{where } \mathbf{I} f \in [\lambda, 2\lambda].$$

The analog of inequality b) of Lemma 8.1 $\equiv$ (8.2) on the tri-tree now follows from Lemma 4.16.

8.1. What goes wrong on 4-tree

The reader has the right to ask: you do not know how to estimate $\int_{T^2}(\mathbb{I}f)^2g^2$ via $\int_{T^2}f^2$, or $\int_{T^3}(\mathbf{I}f)^2g^2$ via $\int_{T^2}f^2$, but you know how to remove this hurdle in the case $f = g$? Perhaps one could also remove this hurdle for $f = g$ on the d-tree, $d \geq 4$.

Unfortunately, we can see now that the trick does not work for $d \geq 4$. Let us notice that by the analogy with (8.4), (8.5), we can construct φ for the 4-tree:

$$\varphi = \lambda^{-1}(I_1 f \cdot \mathbb{I}_{234}g + I_2 f \cdot \mathbb{I}_{134}g + I_3 f \cdot \mathbb{I}_{124}g + I_4 f \cdot \mathbb{I}_{123}g + I_1 g \cdot \mathbb{I}_{234}f + I_2 g \cdot \mathbb{I}_{134}f + I_3 g \cdot \mathbb{I}_{124}f + I_4 g \cdot \mathbb{I}_{123}f + \mathbb{I}_{12}g \cdot \mathbb{I}_{34}f + \mathbb{I}_{23}g \cdot \mathbb{I}_{14}f + \mathbb{I}_{34}g \cdot \mathbb{I}_{12}f + \mathbb{I}_{12}f \cdot \mathbb{I}_{34}g + \mathbb{I}_{23}f \cdot \mathbb{I}_{14}g + \mathbb{I}_{34}f \cdot \mathbb{I}_{12}g + g\mathbf{I}f).$$

Here $\mathbf{I}$ means summation in all four variables, the Hardy operator on T^4. Let us consider what happens for the case $g = f$. We again can absorb the last term $g\mathbf{I}f = f\mathbf{I}f \leq \delta f$ into the left-hand side because $\operatorname{supp} f \subset \{\mathbf{I}f \leq \delta\}$.

But to prove the analog of b) of Lemma 8.1, we would need to know how to estimate, e.g.,

$$\int_{T^4}(\mathbb{I}_{12}f \cdot \mathbb{I}_{34}f)^2 \leq C\int_{T^4} f^2\,.$$

We do not know how to achieve such an estimate.

To feel this difficulty better, let us prove Lemma 8.1, where the main point is the following "weighted" estimate:

$$\operatorname{supp} f \subset \{Ig \leq \delta\} \Rightarrow \int_T (If)^2 g^2 \leq C\delta\|Ig\|_\infty \int_T f^2 \quad \text{for superadditive } g\,. \tag{8.7}$$

8.1.1. The proof of Lemma 8.1 and the explanation where the proof breaks down on the bi-tree. We just repeat the proof from [4], but we emphasize why the proof does not work for very similar estimate of $\int_{T^2}(\mathbb{I}f)^2g^2$. We are in the assumptions of Lemma 8.1. That is, we are given two functions f, g on the tree T, and

1) $\operatorname{supp} f \subset \{Ig \leq \delta\}$, 2) g is a superadditive function.

We need to see why the key estimate (8.7) works on T and will not work on T^2 if one replaces I by $\mathbb{I}$ and T by T^2 everywhere.

We start with a lemma that holds regardless of operators and measures.

Lemma 8.3. *Let K be an integral operator with a positive kernel and let f and g be positive functions. Then*

$$\int (Kf)^2 g \leq \Big(\sup_{\operatorname{supp} g} KK^*g\Big)\int f^2.$$

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

$$\int (Kf)^2 g = \int fK^*(Kf\cdot g) \leq \|f\|_2\,\|K^*(Kf\cdot g)\|_2.$$

We call the operator and its kernel by the same letter K. By the hypothesis,

$$Kh(x) = \int K(x, y)h(y),$$

with a positive kernel K. Hence

$$\begin{aligned}
\|K^*(Kf \cdot g)\|_2^2 &= \int K^*(Kf \cdot g)\, K^*(Kf \cdot g) \\
&= \int K(x, y)((If)(x)g(x))\, K(x', y)\,((Kf)(x')g(x'))\, \mathrm{d}(x, x', y) \\
&\le \int \frac{1}{2}(Kf(x)^2 + Kf(x')^2)K(x, y)(g(x))K(x', y)(g(x'))\, \mathrm{d}(x, x', y) \\
&= \frac{1}{2}\int K^*((Kf)^2 \cdot g)\, K^*(g) + \int K^*(g)\, K^*((Kf)^2 \cdot g) \\
&= \int (KK^*g) \cdot (Kf)^2 \cdot g \le \Big(\sup_{\operatorname{supp} g} KK^*g\Big) \int (Kf)^2 \cdot g.
\end{aligned}$$

Substituting the second displayed estimate into the first we obtain

$$\int (Kf)^2 g \le \|f\|_2 \Big(\sup_{\operatorname{supp} g} KK^*g\Big)\Big(\int (Kf)^2 \cdot g\Big)^{1/2}.$$

The conclusion follows. ■

In the preceding lemma, the operator K could have been either I on T or $\mathbb{I}$ on T^2; this did not matter. But in the next lemma, it matters whether we are on T or T^2.

Lemma 8.4. *Let T be a finite tree, and let $g, h\colon T \to [0, \infty)$. Assume that g is superadditive and $\lambda = \|Ih\|_{L^\infty(\operatorname{supp} g)}$. Then for every $\beta \in T$, we have*

$$I(gh)(\beta) = \sum_{\alpha \le \beta} g(\alpha)h(\alpha) \le \lambda\, g(\beta).$$

Proof. Without loss of generality, we may consider the case when β is the unique maximal element of T and $T = \operatorname{supp} g$. 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 the superadditivity of g, we have

$$\begin{aligned}
\sum_{\alpha \le \beta} g(\alpha)h(\alpha) &= g(\beta)h(\beta) + \sum_{\beta' \in \operatorname{ch}(\beta)} \sum_{\alpha \le \beta'} g(\alpha)h(\alpha) \\
&\le g(\beta)h(\beta) + \sum_{\beta' \in \operatorname{ch}(\beta)} g(\beta') \sup_{\alpha \le \beta'} \sum_{\alpha \le \alpha' \le \beta'} h(\alpha') \\
&\le g(\beta)h(\beta) + \sum_{\beta' \in \operatorname{ch}(\beta)} g(\beta') \sup_{\alpha < \beta} \sum_{\alpha \le \alpha' < \beta} h(\alpha') \\
&\le^{\text{key}} g(\beta)h(\beta) + g(\beta) \sup_{\alpha < \beta} \sum_{\alpha \le \alpha' < \beta} h(\alpha') = g(\beta) \sup_{\alpha \le \beta} \sum_{\alpha \le \alpha' \le \beta} h(\alpha').
\end{aligned}$$ ■

Remark 8.5. It seems that this claim fails to be true on T^2. At least, the reasoning fails. In Conjecture 8.2 we had to assume that g is superadditive in its both variables separately. This assumption is indispensable for us, because in our applications of such a lemma on T^2, the function g on T^2 always comes from some function (measure) f additive on T^3 in each of its three variables. The function g is always defined by a simple rule $g = I_i f \cdot \mathbf{1}_{\mathbf{I} f \le t}$, $i = 1$ or 2 or 3. But such a function g is automatically separately superadditive in each of its two variables.

But if g is separately superadditive in its both variables, then the key estimate in the above lemma does not work. In fact, instead of having $\sum_{\beta' \in \mathrm{ch}(\beta)} g(\beta') \le g(\beta)$, we will have to write

$$\sum_{\beta' \in \mathrm{ch}(\beta)} g(\beta') \le 2g(\beta)\,.$$

This seemingly innocuous change leads to accumulation of constants in the above proof. The above proof breaks down if it cannot keep constant 1 at every stage of the induction.

Now we present the proof of Lemma 8.1 by means of Lemma 8.3 and Lemma 8.4. Let $\varphi = 2\lambda^{-1} I f \cdot g \cdot \mathbf{1}_{Ig \le 4\lambda}$. Let ω be such that $Ig(\omega) \ge \lambda$. Then $f(\omega) = 0$ and $f(\gamma) = 0$ for all ancestors of ω up to the first γ' such that $Ig(\gamma') \le \delta$. Hence, on such ω,

$$\sum_{\gamma \ge \omega} If \cdot g \cdot \mathbf{1}_{Ig \le 4\lambda} = \sum_{\gamma \ge \omega} If \cdot g = If(\omega)(\lambda - \delta) \ge \frac{\lambda}{2}\, If(\omega).$$

We checked (8.1) of Lemma 8.1.

To check (8.2), we first apply Lemma 8.3 with

$$K := I \circ \mathbf{1}_{Ig \le \delta},$$

which a composition of multiplication operator and I. Then

$$\int_T \varphi^2 = \frac{4}{\lambda^2} \int_T (If)^2 (g \mathbf{1}_{Ig \le 4\lambda})^2 \le \frac{4}{\lambda^2} \sup_{\mathrm{supp}\, g} KK^*(g^2 \mathbf{1}_{Ig \le 4\lambda}) \int_T f^2\,.$$

To understand $\sup_{\mathrm{supp}\, g} KK^*(g^2 \mathbf{1}_{Ig \le 4\lambda})$, we use Lemma 8.4. By this lemma, for any node α,

$$K^*(g^2 \mathbf{1}_{Ig \le 4\lambda})(\alpha) \le I^*(g^2 \mathbf{1}_{Ig \le 4\lambda})(\alpha) \le 4\lambda\, g(\alpha)\,.$$

Now we are left to estimate $Kg = I(\mathbf{1}_{Ig \le \delta} g)$. But just by the definition of I, we have

$$I(\mathbf{1}_{Ig \le \delta} g) \le \delta. \tag{8.8}$$

So

$$\sup_{\mathrm{supp}\, g} KK^*(g^2 \mathbf{1}_{Ig \le 4\lambda}) \le 4\delta\lambda\,,$$

and we get

$$\int_T \varphi^2 \le \frac{16\delta}{\lambda} \int f^2\,.$$

Remark 8.6. We already observed one obstacle to prove Conjecture 8.2. We did this in Remark 8.5. Now let us observe, that even if we could manage to overcome this first difficulty, we still have another very serious one: the analog of inequality (8.8) is blatantly false on T^2. The following inequality is generically false on the bi-tree:

$$\mathbb{I}(\mathbf{1}_{\mathbb{I}g\le\delta}\, g) \le \delta.$$

Remark 8.7. Unfortunately, Conjecture 8.2 on T^2 turned out to be false. The counter-example is built in [26].

Acknowledgements. We are grateful to Sergei Treil for indicating a faulty reasoning in Section 2.4; we corrected this reasoning.

Funding. PM is supported by the Russian Science Foundation grant 20-61-46016. The results of Sections 3 and 4 were obtained in the framework of the Russian Science Foundation grant 20-61-46016. 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).

References

[1] Adams, R. and Hedberg, L.: *Function spaces and potential theory*. Grundlehren der mathematischen Wissenschaften 314, Springer, Berlin, 1996.

[2] Arcozzi, N., Holmes, I., Mozolyako, P. and Volberg, A.: Bi-parameter embedding and measures with restriction energy condition. *Math. Ann.* **377** (2020), no. 1-2, 643–674.

[3] Arcozzi, N., Mozolyako, P., Perfekt, K.-M. and Sarfatti, G.: Bi-parameter potential theory and Carleson measures for the Dirichlet space on the bidisc. Preprint 2018, arXiv: 1811.04990.

[4] Arcozzi, N., Mozolyako, P., Psaromiligkos, G., Volberg, A. and Zorin-Kranich, P.: Bi-parameter Carleson embeddings with product weights. Preprint 2020, arXiv: 1906.11150.

[5] Arcozzi, N., Rochberg, R., Sawyer, E. and Wick, B. D.: Potential theory on trees, graphs and Ahlfors-regular metric spaces. *Potential Anal.* **41** (2014), no. 2, 317–366.

[6] Barron, A. and Pipher, J.: Sparse domination for bi-parameter operators using square functions. Preprint 2017, arXiv: 1709.05009.

[7] Carleson, L.: *A counter example for measures bounded on H^p for the bi-disc*. Mittag-Leffler report no. 7, 1974.

[8] Chang, S.-Y. A.: Carleson measure on the bi-disc. *Ann. of Math. (2)* **109** (1979), no. 3, 613–620.

[9] Chang, S.-Y. A. and Fefferman, R.: A continuous version of duality of H^1 with BMO on the bidisc. *Ann. of Math. (2)* **112** (1980), no. 1, 179–201.

[10] Chang, S.-Y. A. and Fefferman, R.: Some recent developments in Fourier analysis and H^p-theory on product domains. *Bull. Amer. Math. Soc. (N.S.)* **12** (1985), no. 1, 1–43.

[11] Dor, L. E.: On projections in L^1. *Ann. of Math. (2)* **102** (1975), no. 3, 463–474.

[12] Fefferman, R.: Calderón–Zygmund theory for product domains: H^p spaces. *Proc. Nat. Acad. Sci. U.S.A.* **83** (1986), no. 4, 840–843.

[13] Fefferman, R.: Harmonic analysis on product spaces. *Ann. of Math. (2)* **126** (1987), 109–130.

[14] Fefferman, R.: Some recent developments in Fourier analysis and H^p theory on product domains. II. In *Function spaces and applications (Lund, 1986)*, pp. 44–51. Lecture Notes in Math. 1302, Springer, Berlin, 1988.

[15] Ferguson, S. and Lacey, M. T.: A characterization of product BMO by commutators. *Acta Math.* **189** (2002), no. 2, 143–160.

[16] Grafakos, L. and Torres, R.: Multilinear Calderón–Zygmund theory. *Adv. Math.* **165** (2002), no. 1, 124–164.

[17] Hänninen, T.: Equivalence of sparse and Carleson coefficients for general sets. *Ark. Mat.* **56** (2018), no. 2, 333–339.

[18] Iosevich, A., Krause, B., Sawyer, E., Taylor, K. and Uriarte-Tuero. I.: Maximal operators: scales, curvature and the fractal dimension. *Anal. Math.* **45** (2019), no. 1, 63–86.

[19] Journé, J.-L.: Calderón–Zygmund operators on product spaces. *Rev. Mat. Iberoamericana* **1** (1985), no. 3, 55–91.

[20] Journé, J.-L.: Two problems of Calderón–Zygmund theory on product-spaces. *Ann. Inst. Fourier (Grenoble)* **38** (1988), no. 1, 111–132.

[21] Lacey, M. T.: Lectures on Nehari's theorem on the polydisk. In *Topics in harmonic analysis and ergodic theory*, pp. 185–213. Contemp. Math. 444, American Mathematical Society, Providence, RI, 2007.

[22] Lacey, M. T.: Two-weight inequality for the Hilbert transform: a real variable characterization II. *Duke Math. J.* **163** (2014), no. 15, 2821–2840.

[23] Lacey, M. T., Sawyer, E. T., Shen, C.-Y. and Uriarte-Tuero, I.: Two-weight inequality for the Hilbert transform: a real variable characterization I. *Duke Math. J.* **163** (2014), no. 15, 2795–2820.

[24] Mozolyako, P., Psaromiligkos, G., Volberg, A. and Zorin-Kranich, P.: Improved surrogate biparameter maximum principle. Preprint 2021, arXiv: 2101.01094v2.

[25] Mozolyako, P., Treil, S. and Volberg, A.: Ferguson–Lacey's theorem for vector valued symbols and embedding measures for Hardy space in the poly-disc. Preprint, 2020.

[26] Mozolyako, P. and Volberg, A.: Essential differences of potential theories on a tree and on a bi-tree. *C. R. Math. Acad. Sci. Paris* **360** (2022), 1039–1048.

[27] Muscalu, C., Pipher, J., Tao, T. and Thiele, C.: Bi-parameter paraproducts. *Acta Math.* **193** (2004), no. 2, 269–296.

[28] Muscalu, C., Pipher, J., Tao, T. and Thiele, C.: Multi-parameter paraproducts. *Rev. Mat. Iberoamericana* **22** (2006), no. 3, 963–976.

[29] Nazarov, F., Treil, S. and Volberg, A.: The Bellman functions and two-weight inequalities for Haar multipliers. *J. Amer. Math. Soc.* **12** (1999), no. 4, 909–928.

[30] Nazarov, F., Treil, S. and Volberg, A.: Two weight inequalities for individual Haar multipliers and other well localized operators. *Math. Res. Lett.* **15** (2008), no. 3, 583–597.

[31] Pipher, J.: Journé's covering lemma and its extension to higher dimensions. *Duke Math. J.* **53** (1986), no. 3, 683–690.

[32] Tao, T.: Dyadic product H^1, BMO, and Carleson's counterexample. Unpublished note (1999), available at http://www.math.ucla.edu/ tao/preprints/Expository/product.dvi.

[33] Verbitsky, I. E.: Imbedding and multiplier theorems for discrete Littlewood–Paley spaces. *Pacific J. Math.* **176** (1996), no. 2, 529–556.

[34] Volberg, A.: Non-symmetry for Fourier transform of positive functions of two variables. Preprint, 2019, 2 pages.

Received June 3, 2021. Published online December 23, 2022.

Pavel Mozolyako
Department of Mathematics and Computer Science, Saint-Petersburg University, 29 Line 14th (Vasilyevsky Island), 199178 Saint-Petersburg, Russia;
pmzlcroak@gmail.com

Georgios Psaromiligkos
Department of Mathematics, Michigan Sate University, East Lansing, MI 48823, USA;
psaromil@math.msu.edu

Alexander Volberg
Department of Mathematics, Michigan Sate University, East Lansing, MI 48823, USA;
volberg@math.msu.edu

Pavel Zorin-Kranich
Mathematical Institute, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany;
pzorin@uni-bonn.de