BELLMAN FUNCTION SITTING ON A TREE

NICOLA ARCOZZI, IRINA HOLMES, PAVEL MOZOLYAKO, AND ALEXANDER VOLBERG

Abstract. In this note we give a proof-by-formula of certain important embedding inequalities on a tree. We also consider the case of a bi-tree, where a different approach is explained.

0.1. Hardy operator on a tree. Let $I^0$ be a unit interval. Let us associate the dyadic lattice $D(I^0)$ and the uniform directed dyadic tree $T$ in a usual way. First we define the Hardy operator, the dual Hardy operator and the logarithmic potential: given a function $\varphi : T \to \mathbb{R}_+$ we let

$$(I\varphi)(\alpha) = \sum_{\beta \geq \alpha} \varphi(\beta), \quad \alpha \in T;$$

$$(I^*\varphi)(\alpha) = \sum_{\beta \leq \alpha} \varphi(\beta), \quad \beta \in T;$$

$$V^\varphi(\gamma) = (II^*\varphi)(\gamma), \quad \gamma \in T,$$

where $\leq$ is the natural order relation on $T$.

We always may think that the tree $T$ is finite (albeit very large). By the boundary $\partial T$ we understand the vertices that are not connected to smaller vertices.

Each dyadic interval $Q$ in $D(I^0)$ corresponds naturally to a vertex $\alpha_Q$. 

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

Key words and phrases. Carleson embedding on dyadic tree, bi-parameter Carleson embedding, Bellman function, capacity on dyadic tree and bi-tree.

Theorem 3.1 was obtained in the frameworks of the project 17-11-01064 by the Russian Science Foundation.

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

IH is partially supported by the NSF an NSF Postdoc under Award No.1606270.

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

AV is partially supported by the NSF grant DMS-160065.
Let $\mu$ be a measure on the tree $T$, so just the collection of non-negative numbers $\{\mu_P\}_{P \in T}$. Assuming $\mu$ to be a measure on $T$, we have

$$(I\mu)(\alpha_R) = \sum_{Q \supset R} \mu_Q, \quad Q, R \in \mathcal{D}(I^0);$$

$$(I^*\mu)(\alpha_Q) = \mu(Q) = \sum_{P \subset Q : \alpha_P \in \partial T} \mu_P, \quad Q \in \mathcal{D}(I^0);$$

$$V^\mu(\alpha_P) = (II^*\mu)(\alpha_P), \quad P \in \mathcal{D}(I^0),$$

the second equality is valid under the assumption of $\text{supp} \mu \subset \partial T$.

We will answer the question when $I : \ell^2(T) \to \ell^2(T, \mu)$. Passing to the adjoint operator we see that this is equivalent to the following inequality

$$(0.1) \sum_{Q \in T} \left( \sum_{P \leq Q} \varphi(P) \mu_P \right)^2 \lesssim \left( \sum_{R \in T} \varphi(R)^2 \mu_R \right).$$

**Theorem 0.1.** Operator $I$ is a bounded operator $I : \ell^2(T) \to \ell^2(T, \mu)$ if and only if

$$(0.2) \sum_{Q \in T; Q \leq R} \left( \sum_{P \leq Q} \mu_P \right)^2 \lesssim \left( \sum_{Q \leq R} \mu_Q \right) \forall R \in T.$$

This is proved in Theorem 1.3 below by the use of Bellman function.

1. **Bellman function on a tree**

**Theorem 1.1.** Let $dw$ be a positive measure on $I_0 := [0, 1]$. Let $\langle w \rangle_J$ denote $w(I)/|I|$. Let $\varphi$ be a measurable test function. Then if

$$(1.1) \frac{1}{|J|} \sum_{I \in \mathcal{D}(J)} \langle w \rangle_I^2 |I|^2 \leq \langle w \rangle_J \quad \forall J \in \mathcal{D}(I_0),$$

then

$$(1.2) \sum_{I \in \mathcal{D}(I_0)} \langle \varphi w \rangle_I^2 |I|^2 \lesssim \langle \varphi^2 w \rangle_{I_0} |I_0|,$$

This result can be obtained as a direct consequence of the weighted Carleson embedding theorem [4]:

**Theorem 1.2.** Let $\mathcal{D}$ be a dyadic lattice, $w$ be any weight, and $\{\alpha_I\}_{I \in \mathcal{D}}$ be a sequence of non-negative numbers. Then, if

$$(1.3) \frac{1}{|J|} \sum_{I \leq J} \alpha_I \langle w \rangle_I^2 \leq \langle w \rangle_J \quad \forall J \in \mathcal{D},$$

then

$$(1.4) \sum_{I \in \mathcal{D}} \alpha_I (\varphi \sqrt{w})_I^2 \lesssim \|\varphi\|^2_{L^2},$$

for all $\varphi \in L^2$. 
Clearly, the conclusion of (1.4) may be rewritten as
\[ \frac{1}{|I_0|} \sum_{I \subset I_0} \alpha_I \langle \varphi w \rangle_I^2 \lesssim \langle \varphi^2 w \rangle_{I_0}. \]
Letting \( \alpha_I = |I|^2 \) in (1.3), we obtain exactly Theorem 1.1.

We recall here that the proof of Theorem 1.2 in [4] was based upon the Bellman function
\[ B(F, f, A, v) := F - \frac{f^2}{v + A}, \]
and three main properties this function satisfies are:

1. \( B \) is defined on:
   \[ f^2 \leq Fv; A \leq v; \]
2. \( 0 \leq B \leq F \);
3. Main Inequality:
   \[ B(F, f, A, v) - \frac{1}{2} \left( B(F_-, f_-, A_-, v_-) + B(F_+, f_+, A_+, v_+) \right) \geq \frac{f^2}{v^2} m, \]
   for all points in the domain such that
   \[ F = \frac{F_- + F_+}{2}; f = \frac{f_- + f_+}{2}; v = \frac{v_- + v_+}{2}, \]
   and
   \[ A = m + \frac{A_- + A_+}{2}, \]
   for some \( m \geq 0. \)
In particular, we have that the function \( B \) is \emph{concave}.

1.1. \textbf{Carleson embedding theorem on a dyadic tree.} Now we wish to prove a version of Theorem 1.1 on a dyadic tree. Specifically, suppose we have a dyadic tree originating at some \( I_0 \in \mathcal{D} \). Define a measure \( \Lambda \) on the tree as follows: to each node \( I \in \mathcal{D}(I_0) \) we associate a non-negative number \( \lambda_I \geq 0 \). We may think of \( I \in \mathcal{D}(I_0) \) as an interval \emph{in the dyadic tree} by considering \( \{ K \in \mathcal{D}(I_0) : K \subset I \} \). Then we define
   \[ \Lambda(I) := \sum_{K \subset I} \lambda_K, \]
and the averaging operator
   \[ (\Lambda)_I := \frac{1}{|I|} \Lambda(I). \]
Given a function \( \varphi = \{ \varphi(I) \} \in \mathcal{D}(I_0) \) on the dyadic tree, we have
   \[ \int_I \varphi d\Lambda = \sum_{K \subset I} \varphi(K) \lambda_K, \]
and
\[(\varphi\Lambda)_I := \frac{1}{|I|} \int_I \varphi d\Lambda.\]

**Theorem 1.3** (Carleson embedding theorem for a dyadic tree). Let \(I_0 \in D\), the dyadic tree originating at \(I_0\) with notations as above, and \(\{\alpha_I\}_{I \subseteq I_0}\) be a sequence of non-negative numbers. Then, if

\[(1.8) \quad \frac{1}{|I|} \sum_{K \subseteq I} |K|^2 \alpha_K (\Lambda)_K \leq (\Lambda)_I, \quad \forall I \in D(I_0),\]

then

\[(1.9) \quad \frac{1}{|I_0|} \sum_{I \subseteq I_0} \alpha_I (\varphi\Lambda)_I \leq 4(\varphi^2)_I_0,\]

where

\[(\varphi\sqrt{\Lambda})_I := \frac{1}{|I|} \sum_{K \subseteq I} \varphi(K) \sqrt{\lambda_K} \quad \text{and} \quad (\varphi^2)_I_0 := \frac{1}{|I_0|} \sum_{I \subseteq I_0} \varphi(I)^2.\]

Note that the conclusion of (1.9) may be rewritten as

\[\sum_{I \subseteq I_0} \alpha_I (\varphi\Lambda)_I \leq 4(\varphi^2\Lambda)_I_0.\]

Letting \(\alpha_I = |I|^2\) in (1.8), we obtain:

**Corollary 1.4.** Let \(I_0 \in D\), the dyadic tree originating at \(I_0\) with notations as above. Then, if

\[(1.10) \quad \frac{1}{|I|} \sum_{K \subseteq I} |K|^2 (\Lambda)_K \leq (\Lambda)_I, \quad \forall I \in D(I_0),\]

then

\[\frac{1}{|I_0|} \sum_{I \subseteq I_0} |I|^2 (\varphi\Lambda)_I \leq 4(\varphi^2\Lambda)_I_0.\]

The proof of Theorem 1.3 is based also on the function \(B\) in (1.5), and on proving a more involved version of (1.6) – this will be Lemma 1.5. We recall here that if \(g\) is a concave, differentiable function on a convex domain \(S \subset \mathbb{R}^n\), then

\[g(x) - g(x^*) \leq \sum_{i=1}^n \frac{\partial g}{\partial x_i}(x^*) \cdot (x_i - x_i^*),\]

for all \(x, x^* \in S\). Denoting \(x := (F, f, A, v)\), for the function \(B\), this takes the particular form:

\[(1.10) \quad B(x) - B(x^*) \leq (F - F^*) - \frac{2f^*}{v^* + A^*}(f - f^*)\]
\[\quad + \frac{(f^*)^2}{(v^* + A^*)^2}(A - A^*) + \frac{(f^*)^2}{(v^* + A^*)^2}(v - v^*).\]
In particular:

\[(1.11) \quad B(F, f, A, v) - B(F, f, A - c, v) \geq c \frac{f^2}{(v + A)^2} \geq c \frac{f^2}{4v^2},\]

where the last inequality follows because \(0 \leq A \leq v\).

**Lemma 1.5.** The function \(B\) satisfies:

\[(1.12) \quad c \frac{f^2}{4v^2} \leq B(F, f, A, v) - \frac{1}{2} \left( B(F_-, f_-, A_-, v-) + B(F_+, f_+, A_+, v+) \right),\]

for all quadruplets in the domain of \(B\) such that

\[F = \tilde{F} + b^2; \quad f = \tilde{f} + ab; \quad A = \tilde{A} + c; \quad v = \tilde{v} + a^2,\]

and

\[\tilde{F} := \frac{F_- + F_+}{2}; \quad \tilde{f} := \frac{f_- + f_+}{2}; \quad \tilde{A} := \frac{A_- + A_+}{2}; \quad \tilde{v} := \frac{v_- + v_+}{2},\]

and \(a \geq 0, b \in \mathbb{R}, c \geq 0\) are some real numbers.

**Proof.** By (1.11):

\[c \frac{f^2}{4v^2} \leq \left( B(F, f, A, v) - B(\tilde{F}, \tilde{f}, A - c, \tilde{v}) \right) + \left( B(\tilde{F}, \tilde{f}, A - c, \tilde{v}) - B(F, f, A - c, v) \right).\]

We claim that the term in the second parenthesis is negative: apply (1.10) to obtain

\[(1.13) \quad B(\tilde{F}, \tilde{f}, A - c, \tilde{v}) - B(F, f, A - c, v) \leq -b^2 - \frac{2f}{v + A - c}(-ab) + \frac{f^2}{(v + A - c)^2}(-a^2)\]

\[- (b - \frac{af}{v + A - c})^2 \leq 0.\]

Then

\[c \frac{f^2}{4v^2} \leq \left( B(F, f, A, v) - B(\tilde{F}, \tilde{f}, A - c, \tilde{v}) \right) \leq B(F, f, A, v) - \frac{1}{2} \left( B(F_-, f_-, A_-, v-) + B(F_+, f_+, A_+, v+) \right),\]

where \((A_- + A_+)/2 = A - c\), and the last inequality follows by concavity of \(B\). This proves the lemma. \(\square\)

We are now ready to prove Theorem 1.3.
Proof of Theorem 1.3. For every \( I \in \mathcal{D}(I_0) \) define:

\[
\begin{align*}
v_I &:= (\Lambda)_I = \frac{1}{|I|}\lambda_I + \frac{1}{2}(v_{I_-} + v_{I_+}) \\
&= a_I^2 + \bar{v}_I, \text{ where } a_I := \sqrt{\frac{\lambda_I}{|I|}}; \\
F_I &:= (\varphi^2)_I = \frac{1}{|I|}\varphi(I)^2 + \frac{1}{2}(F_{I_-} + F_{I_+}) \\
&= b_I^2 + \bar{F}_I, \text{ where } b_I := \varphi(I)\sqrt{\frac{\lambda_I}{|I|}}; \\
f_I &:= (\varphi\sqrt{\Lambda})_I = \frac{\varphi(I)\sqrt{\lambda_I}}{|I|} + \frac{1}{2}(f_{I_-} + f_{I_+}) = a_I b_I + \bar{f}_I; \\
A_I &:= \frac{1}{|I|} \sum_{K \subset I} \alpha_K (\Lambda)_K = \frac{(\Lambda)_I^2}{|I|} + \frac{1}{2}(A_{I_-} + A_{I_+}) \\
&= c_I + \bar{A}_I, \text{ where } c_I := \frac{\alpha_I (\Lambda)_I^2}{|I|}.
\end{align*}
\]

Note then that

\[
\begin{align*}
\mathcal{B}(F_I, f_I, A_I, v_I) &= \mathcal{B}(b_I^2 + \bar{F}_I, a_I b_I + \bar{f}_I, c_I + \bar{A}_I, a_I^2 + \bar{v}_I),
\end{align*}
\]

so we may apply Lemma 1.5 and obtain

\[
\frac{1}{4}\alpha_I f_I^2 \leq |I|\mathcal{B}(x_I) - |I_-|\mathcal{B}(x_{I_-}) - |I_+|\mathcal{B}(x_{I_+}).
\]

Summing over \( I \in \mathcal{D}(I_0) \) and using the telescoping nature of the sum, we have

\[
\sum_{I \in I_0} \alpha_I f_I^2 \leq 4|I_0|\mathcal{B}(F_{I_0}, f_{I_0}, A_{I_0}, v_{I_0}) \leq 4|I_0|F_{I_0},
\]

which is exactly (1.9). \( \square \)

Remark 1.6. The big and essential difference with Theorem 1.1 now is that in the proof of Theorem 1.1 \( \{v_I\}_{I \in \mathcal{D}}, \{f_I\}_{I \in \mathcal{D}}, \{F_I\}_{I \in \mathcal{D}} \) are all martingales.

Now they are only supermartingales. In other words, we do not have the property (1.7) anymore.

Instead,

\[
\begin{align*}
F_I &\geq \frac{F_{I_-} + F_{I_+}}{2}; \\
f_I &\geq \frac{f_{I_-} + f_{I_+}}{2}; \\
v_I &\geq \frac{v_{I_-} + v_{I_+}}{2}.
\end{align*}
\]

This is an essential difference, because there is less cancellation, and, indeed, the fact that \( \{f_I\}_{I \in \mathcal{D}} \) is only a supermartingale can destruct the whole proof. As a small miracle the “good” supermartingale properties of \( \{v_I\}_{I \in \mathcal{D}}, \{F_I\}_{I \in \mathcal{D}} \) allow us to neutralize the “bad” supermartingale property of \( \{f_I\}_{I \in \mathcal{D}} \).
The exact calculation of this “good”–“bad” interplay was in (1.13). We wish to explain now why some supermartingales \( \{v_I\}_{I \in \mathcal{D}}, \{F_I\}_{I \in \mathcal{D}} \) are good and some, namely, \( \{f_I\}_{I \in \mathcal{D}} \) is bad.

The explanation is simple: the good ones are those that give positive partial derivative of \( \mathcal{B} \), the bad is the one that gives a negative partial derivative of \( \mathcal{B} \). In fact, 
\[
\frac{\partial \mathcal{B}}{\partial v} \geq 0, \quad \frac{\partial \mathcal{B}}{\partial F} = 1, \\
\frac{\partial \mathcal{B}}{\partial f} \leq 0.
\]

2. Maximal theorem on a tree

Now we are going to prove the result slightly more general than Corollary 1.4 from the previous section.

**Theorem 2.1.** Let \( I_0 \in \mathcal{D} \), the dyadic tree originating at \( I_0 \) with notations as above. Then, if 
\[
\frac{1}{|I|} \sum_{K \subset I} |K|^2 (\Lambda_K^2 \leq (\Lambda)_I, \quad \forall I \in \mathcal{D}(I_0),
\]
then 
\[
\frac{1}{|I_0|} \sum_{I \subset I_0} |I|^2 (\Lambda_I^2 \leq \sup_{K: I \subseteq K} \left( \frac{\varphi \Lambda_K}{\Lambda_K} \right)^2 \lesssim (\varphi^2 \Lambda)_{I_0}.
\]

The proof – for a change – is a stopping time proof and not a Bellman proof.

**Proof.** For every vertex \( H \) of the tree, let us introduce the set of vertices \( E_H \). Namely, let \( J \) be the first successor of \( H \) such that 
\[
\frac{(\varphi \Lambda)_J}{(\Lambda)_J} \geq 2 \frac{(\varphi \Lambda)_H}{(\Lambda)_H}.
\]

It may happen of course that \( J \) is not alone, and there are several first successors with this property. We call by \( E_H \) all vertices that are successors of all these \( J \)'s and also all such \( J \)'s.

Now we introduce another set of vertices associated with \( H \). Consider all successors of \( H \) which are not in \( E_H \). All of them plus \( H \) itself form the collection that is called \( O_H \). This set in never empty (it contains \( H \)) and can include all successors of \( H \).

Now we first assign \( H = I_0 \) and let \( \{J\} \) be the first successors of \( H \) with the property above. We call this family stopping vertices of first generation, and denote it by \( S_1 \). Then for any \( H \in S_1 \) we repeat the procedure thus having stopping vertices of the second generation: \( S_2 \).

For each \( j \) and each \( H \in S_j \), we have \( E_H \) and \( O_H \). Notice that all such \( O_H \) are disjoint. We call \( I_0 \) the stopping vertex of 0 generation, and let \( S = \bigcup_{j=0}^\infty S_j \).
Then
\[
\sum_{I \subset I_0} |I|^2(\Lambda)^2_I \sup_{K : I \subset K, |K| \leq 2} \left( \frac{(\varphi \Lambda)_K}{(\Lambda)_K} \right)^2
\]
\[
= \sum_{H \in S} \sum_{I \in O_H} |I|^2(\Lambda)^2_I \sup_{K : I \subset K, |K| \leq 2} \left( \frac{(\varphi \Lambda)_K}{(\Lambda)_K} \right)^2
\]
\[
\leq 4 \sum_{H \in S} \left( \frac{(\varphi \Lambda)_H}{(\Lambda)_H} \right)^2 \sum_{I \in O_H} |I|^2(\Lambda)^2_I
\]
\[
\leq 4 \sum_{H \in S} \left( \frac{(\varphi \Lambda)_H}{(\Lambda)_H} \right)^2 \Lambda(H).
\]
The last inequality uses the assumptions of the theorem. But notice that by definition of $E_H$ we easily get
\[
\Lambda(E_H) \leq \frac{1}{2} \Lambda(H) \Rightarrow \Lambda(H) \leq 2 \Lambda(O_H).
\]
Hence
\[
(2.1) \quad \sum_{I \subset I_0} |I|^2(\Lambda)^2_I \sup_{H : I \subset H, |H| \leq 2} \left( \frac{(\varphi \Lambda)_I}{(\Lambda)_I} \right)^2 \leq 8 \sum_{H \in S} \left( \frac{(\varphi \Lambda)_H}{(\Lambda)_H} \right)^2 \Lambda(O_H).
\]
Now, define $\beta_H$ for $H \subset I_0$ by $\beta_H := \Lambda(O_H)$ if $H \in S$, and $\beta_H := 0$ otherwise. Note that, by disjointness of $O_H$, we have
\[
\sum_{H \subset K} \beta_H \leq \Lambda(K), \quad \forall K \subset I_0.
\]
Therefore, if we let
\[
\alpha_H := \frac{\beta_H}{(\Lambda)_H}
\]
then the sequence $\alpha_H$ satisfies the requirements of the Carleson Embedding Theorem [1.3] for the dyadic tree. So we may rewrite the right hand side of (2.1) in terms of $\beta_H$ and apply Theorem [1.3]
\[
\frac{1}{|I_0|} \sum_{I \subset I_0} |I|^2(\Lambda)^2_I \sup_{H : I \subset H} \left( \frac{(\varphi \Lambda)_I}{(\Lambda)_I} \right)^2
\]
\[
\leq 8 \frac{1}{|I_0|} \sum_{H \subset I_0} \alpha_H (\varphi \Lambda)^2_H
\]
\[
\leq 32 (\varphi^2 \Lambda)_{I_0},
\]
completing the proof of Theorem 2.1.

\[\square\]

3. BI-TREE AND DYADIC RECTANGLES

Here we extend Theorem 1.1 to dyadic rectangles in $[0, 1)^2$. 

3.1. **Some additional notation.** We mostly follow previously established notation with some changes here and there. As before we identify the system of dyadic subintervals of a base interval $I^0$ with a rooted uniform directed dyadic tree $T$, and we also introduce the bi-tree $T^2 = T_x \times T_y$, where $T_x = T_y = T$. To elaborate, the vertex set of $T^2$ is the Cartesian product of the vertex sets of coordinate trees, and two points $\alpha = (\alpha_x, \alpha_y)$ and $\beta = (\beta_x, \beta_y)$ are connected by an edge when $\alpha_x = \alpha_y$ and $\beta_x$ is connected by an edge with $\beta_y$, or vice-versa. In what follows we do not use the edges of the graphs we are dealing with, so we identify the tree (bi-tree) and its vertex set. The direction for $T^2$ is induced by the coordinate trees as well. The associated representation is the system of dyadic subrectangles of a base rectangle $R^0 = I^0_x \times I^0_y$ (we might as well assume $R^0 = [0,1)^2$): each point $\alpha \in T^2$ corresponds to a unique rectangle $R_\alpha \subset R^0$, and each rectangle $R \subset R^0$ has its unique counterpart $\alpha_R \in T^2$. The order relation is given by inclusion: $\alpha \leq \beta$, if $R_\alpha \subset R_\beta$. In what follows we always assume that the trees and bi-trees in question are finite (of depth $K$ for some unimportant large number $K$), this allows us to disregard some unnecessary measure-theoretic details. We denote the boundary of (a finite) tree $T$ by $\partial T$ — a collection of dyadic intervals of rank $K = \text{depth}(T)$. Respectively, the distinguished boundary of a bi-tree $(\partial T)^2$ is a collection of rectangles $R = I_x \times I_y$ with $|I_x| = |I_y| = 2^{-K}$. Since the graphs we are dealing with are finite, it means that measures and functions are pretty much the same objects there — a collection of non-negative weights on the vertices. Nevertheless we always assume that the measure (say $\mu$) on the bi-tree is concentrated on its distinguished boundary, i.e. $\text{supp} \mu \subset (\partial T)^2$. Alternatively, one can think of $\mu$ as a measure on $[0,1)^2$ with piecewise-constant density with respect to $dx \, dy$,

$$d\mu(x,y) = \phi(x,y) \, dx \, dy,$$

where $\phi(t) \equiv \text{const}$ for $t = (t_x, t_y) \in r_{ij} = [i2^{-K}, (i+1)2^{-K}) \times [j2^{-K}, (j+1)2^{-K})$ (so that $\alpha_{r_{ij}} \in (\partial T)^2$). We use the same notation for objects on the bi-tree and its dyadic rectangle representation, the collection of dyadic subrectangles of $R^0$ is defined by $D(R^0)$ (again, we assume $R^0$ to be fixed).

3.1.1. **Potential-theoretic definitions.** As before, we define the Hardy operator, the dual and the logarithmic potential: given a function $\varphi : T^2 \to \mathbb{R}_+$ we let

$$\langle \mathbb{I} \varphi \rangle(\alpha) = \sum_{\beta \geq \alpha} \varphi(\beta), \quad \alpha \in T^2;$$

$$\langle \mathbb{I}^* \varphi \rangle(\alpha) = \sum_{\beta \leq \alpha} \varphi(\beta), \quad \beta \in T^2;$$

$$\mathbb{V} \varphi(\gamma) = \langle \mathbb{I}^* \varphi \rangle(\gamma), \quad \gamma \in T^2.$$

Every dyadic rectangle $Q$ corresponds to a vertex $\alpha_Q$, but we will often write just $Q$ instead of $\alpha_Q$. Assuming $\mu$ to be a measure and using the dyadic
rectangles, we have
\[(\mathbb{I}\mu)(R) = \sum Q \ni R \mu(Q), \quad Q, R \in \mathcal{D}(R^0);\]
\[(\mathbb{I}^*\mu)(Q) = \sum R \subset Q, R \in (\partial T)^2 \quad \mu(R) = \mu(Q), \quad Q, R \in \mathcal{D}(R^0);\]
\[\nabla^\mu(P) = \langle \mathbb{I}^*\mu \rangle(P), \quad P \in \mathcal{D}(R^0),\]
the second equality is due to our assumption of \(\text{supp } \mu \subset (\partial T)^2\) (so that \(\mu(R) = 0\), unless \(|R| = 2^{-2K}\)). As before we may consider the potential \(\nabla^\mu\) to be a piecewise-constant function on \(R^0\),
\[\nabla^\mu(t) = \nabla^\mu(P), \quad \text{where } P \text{ is a unique rectangle in } (\partial T)^2 \text{ containing } t.\]
A simple computation gives
\[\nabla^\mu(t) = \sum_{t \in Q \in \mathcal{D}(R^0)} \mu(Q) = \int_{R^0} k(t, s) d\mu(s),\]
where the kernel \(k(t, s)\) is
\[k(t, s) = \sharp\{Q \in \mathcal{D}(R^0) : t, s \in Q\}\]
the number of common ancestors (in \(T^2\)) of \(t\) and \(s\). In particular, if these points are "far" from each other in both coordinates (so that both \(|t_x - s_x|\) and \(|t_y - s_y|\) are much greater than \(2^{-K}\)), then this kernel looks like a product of two logarithms
\[k(t, s) \sim \log \frac{1}{|t_x - s_x|} \log \frac{1}{|t_y - s_y|},\]
hence the term "logarithmic potential".

**Definition.** We call a set \(E \subset R^0\) rectangular, if \(E = \bigcup_{j=1}^{N} R_j\) for some finite collection \(\{R_j\}\) of dyadic rectangles (any subset of \((\partial T)^2\) is rectangular, if \(T^2\) is finite).

Given such a set \(E \subset R^0\) we define a set \(\Omega_E\) of \(E\)-admissible functions
\[\Omega_E = \{f : T^2 \rightarrow \mathbb{R}_+ : \|f\| \geq 1 \text{ on } E\},\]
i.e. such functions that
\[\sum_{t \in Q} f(Q) \geq 1, \quad t \in E.\]
Next we define
\[\text{Cap } E = \inf_{f \in \Omega_E} \|f\|^2,\]
where
\[\|f\|^2 = \|f\|^2_{(T^2)} = \sum_{\mathcal{D}(R^0)} f^2(Q).\]
By general theory ([1], see [3] for details) there exists a minimizer $f_E$ and an equilibrium measure $\mu_E$, such that $f_E(Q) = \mu_E(Q) = (\mathbb{I} \ast \mu_E)(\alpha_Q))$, supp $\mu_E \subset E$, $\nabla^\mu E \equiv 1$ on supp $\mu_E$, and Cap $E = \mu_E(E)$ (here we use the finiteness of the bi-tree — all sets are clopen in natural topology).

Finally we define a mutual energy of two measures — given $\mu, \nu \geq 0$ we let $E[\mu, \nu] = \sum_{Q \in D(R^0)} \mu(Q) \nu(Q) = \int_{R^0} \nabla^\mu(t) d\nu(t) = \int_{R^0} \nabla^\nu(t) d\mu(t)$, we also let $E[\mu, \mu] = E[\mu]$. In particular, if $\mu = \mu_E$ is an equilibrium measure of some set $E$, then

$$\text{Cap } E = \mu_E(E) = |\mu_E| = \int_{R^0} \nabla^\mu E d\mu_E = E[\mu_E].$$

3.2. 2d version of Theorem 1.1.

**Theorem 3.1.** Let $\mu$ be a positive measure on $R^0 = [0,1)^2$. Let $\langle \mu \rangle_R$ denote $\frac{\mu(R)}{|R|}$. Let $\varphi$ be a measurable test function. Then, if

$$\sum_{Q \subset E, Q \in D(R^0)} \langle \mu \rangle_Q^2 |Q|^2 \leq \mu(E), \quad \forall \text{ rectangular } E,$$

then

$$\sum_{Q \in D(R^0)} \langle \varphi \mu \rangle_Q^2 |Q|^2 \lesssim \langle \varphi^2 \mu \rangle_{R^0} |R^0|.$$  

**Remark.** Like in 2d-version of Carleson measure theorem for the Hardy space we use here a modified version of the test condition (1.1) (we believe that it would be not enough to check (3.1) only for single rectangles).

3.3. One box condition and its corollary. In the next theorem it is essential to think that $\mu = \{\lambda_\beta\}$ is the measure on $\partial T^2$. Every dyadic rectangle $R$ corresponds to a node of $T^2$, and we will use this in the notations below.

**Theorem 3.2.** Let $\mu$ be a positive measure on $R^0 = [0,1)^2$. Let $\langle \mu \rangle_R$ denote $\frac{\mu(R)}{|R|}$. Let $\varphi$ be a measurable test function. Then, if

$$\sum_{Q \subset R, Q \in D(R^0)} \langle \mu \rangle_Q^2 |Q|^2 \leq \mu(R), \quad \forall \text{ rectangle } R,$$

then

$$\sum_{Q \in D(R^0)} \langle \varphi \mu \rangle_Q^2 |Q|^3 \lesssim \langle \varphi^2 \mu \rangle_{R^0} |R^0|.$$  

**Proof.** We consider exactly the same function $B(x), x = (F, f, A, v),

$$B(x) = F - \frac{f^2}{v + A}.$$
Given a rectangle \( R \) we consider
\[
F_R = \frac{1}{|R|} \sum_{\beta \leq R} \phi_{\beta}^2 \lambda_{\beta} = \frac{1}{|R|} \int_R \phi^2 \, d\mu, \\
f_R = \frac{1}{|R|} \sum_{\beta \leq R} \phi_{\beta} \lambda_{\beta} = \frac{1}{|R|} \int_R \phi \, d\mu,
\]
\[
v_R = \frac{1}{|R|} \sum_{\beta \leq R} \lambda_{\beta} = \frac{\mu(R)}{|R|}, \\
A_R = \frac{1}{|R|} \sum_{\beta \leq R} v_{\beta}^2 |R_{\beta}|^2, \quad x_R = (F_R, \ldots, v_R).
\]

Let \( R_+, R_- \) be right and left half-rectangles of \( R \), and \( R^t, R^b \) be top and bottom half-rectangles of \( R \) (so, e.g., if \( R = I \times J \), the \( R^t = I \times J_+ \)). Now let us estimate from below
\[
B(x_R) - \frac{1}{4} \left( B(x_{R_-}) + B(x_{R_+}) + B(x_{R^t}) + B(x_{R^b}) \right).
\]

As \( \mu \) is concentrated on the boundary, we see immediately, that
\[
F_R = \frac{1}{4} \left( F_{R_-} + F_{R_+} + F_{R^t} + F_{R^b} \right), \\
f_R = \frac{1}{4} \left( f_{R_-} + f_{R_+} + f_{R^t} + f_{R^b} \right).
\]

At the same time,
\[
A_R - \frac{1}{4} \left( A_{R_-} + A_{R_+} + A_{R^t} + A_{R^b} \right) \geq \frac{1}{|R|} \mu(R) \lambda_2, \\
v_R = \frac{1}{4} \left( v_{R_-} + v_{R_+} + v_{R^t} + v_{R^b} \right).
\]

The second equality here is just because \( v_R = \frac{1}{2} \left( v_{R_-} + v_{R_+} \right) \). And \( v_R = \frac{1}{2} \left( v_{R^t} + v_{R^b} \right) \). The first one because any \( \beta \)-term in \( A_R \) such that this term happens to be in two rectangles, e.g. in \( A_{R_-} \) and \( A_{R^t} \), will be cancelled in the difference. The terms that happen only in one rectangle (this is the case for \( R_{--} \) as an example) will be in coefficient \( \frac{1}{|R|} \) in \( A_R \), and only with coefficient \( \frac{1}{2 |R|} \) in \( \frac{1}{4} \left( A_{R_-} + A_{R_+} + A_{R^t} + A_{R^b} \right) \), so it gives a partial (positive) contribution to \( A_R - \frac{1}{4} \left( A_{R_-} + A_{R_+} + A_{R^t} + A_{R^b} \right) \). And of course, \( \frac{1}{|R|} \mu(R) \lambda_2 \) is in \( A_R \) and in none of \( A_{R_-}, A_{R_+}, A_{R^t}, A_{R^b} \), so it also the part of the contribution.

So we see that three variables \( F, f, v \) split in a “martingale” way, and for \( A_R \) we have the above “super-martingale” inequality.

Thus, considering
\[
x_R^* = (F_R; f_R; \frac{1}{4} (A_{R_-} + A_{R_+} + A_{R^t} + A_{R^b}); v_R)
\]

we can write
\[
B(x_R) - B(x_R^*) \geq \frac{\partial B}{\partial A}(x_R)(x_R - x_R^*);
\]
\[
B(x_R^*) \geq \frac{1}{4} \left( B(x_{R_-}) + B(x_{R_+}) + B(x_{R^t}) + B(x_{R^b}) \right).
\]
Here both inequalities are corollaries of the concavity of $B$, in the first one we used that all coordinates of $x_R, x^*_R$ coincide except the $A$-coordinate. Therefore, now we get

$$B(x_R) - \frac{1}{4} \left( B(x_{R_0}) + B(x_{R^+}) + B(x_{R^t}) + B(x_{R^b}) \right) \geq c \frac{\int_R \mu(R)^2}{v^2_R \ |R|} \left( \int_R \phi d\mu \right)^2 \ |R| \ .$$

Multiply this by $|R|^2$. We get a term of telescopic sum on the left:

$$|R|^2 B(x_R) - \left( |R^+_R|^2 B(x_{R^+_R}) + |R^t_R|^2 B(x_{R^t_R}) + |R^b_R|^2 B(x_{R^b_R}) \right) \geq c |R| \left( \int_R \phi d\mu \right)^2 \ |R| .$$

Notice that on the next step we pick up all terms $|R| \left( \int_R \phi d\mu \right)^2$ with $R := R_-, R^+, R^t, R^b$. Theorem is proved. 

3.4. **Proof of Theorem 3.1.** The proof we have now [3] is very non-direct and un-Bellman. It goes through (among other things) some capacitary estimates, though (at least on surface) neither the condition (3.1) nor the statement (3.2) mention anything about the capacity. The general scheme is as follows:

- First we notice that we only need to prove the only if part of the theorem, since (3.1) easily follows from (3.2) by testing on special functions;
- Then we use the recently obtained Stegenga-type characterization of measures satisfying (3.2) (basically, (3.2) holds for all $\phi$ if and only if $\mu$ is subcapacitary measure, i.e. $\mu(E) \leq \text{Cap } E$ for any rectangular $E$) — we are left to show that (3.1) implies subcapacitary property of $\mu$;
- We proceed by assuming that (3.2) is not true, i.e. (by reasoning above) $\mu$ can be very much non-subcapacitary — there exists a ”bad” rectangular set $F_0$ such that $\mu(F_0) \geq C \text{ Cap } F_0$ for some very large $C$;
- Using (3.1) we enlarge $F_0$ to obtain another set $F_1 \supset F_0$ that is at least ten times ”worse” than $F_0$;
- We iterate the step above until we exhaust the whole bi-tree (a propagation of ”badness” of sorts), so that at some point $j$ we have $F_j = R^0$ (we can repeat the enlarging step only so many times — the mass of $\mu$ is finite), which turns out to be in contradiction with (3.1).

**Remark.** Recently we come up with another proof (see [2]), which does not use capacity, still it is difficult although it is more direct.
3.4.1. Steps A and B. If we plug \( \varphi = \chi_E \) into (3.2), we obtain (3.1), and there goes the "if" part. From now on we assume (3.1) holds, and we aim to show (3.2).

Next we pass to the dual inequality, namely (3.2) is equivalent to

\[
\int_{R^0} ((\mathbb{I}f)(t))^2 d\mu(t) = \int_{R^0} \left( \sum_{t \in Q \in \mathcal{D}(R^0)} f(Q) \right)^2 d\mu(t)
\]

\[
\lesssim \|f\|^2 = \sum_{Q \in \mathcal{D}(R^0)} f^2(Q), \quad \forall f \in \ell^2(T^2).
\]

We claim that (3.5) follows from the subcapacitary condition

\[
\mu(E) \lesssim \text{Cap } E, \quad \forall \text{ rectangular } E.
\]

Indeed, the left-hand side of (3.5) is more or less \( \sum_{k \in \mathbb{Z}} 2^{2k} \mu(\mathbb{I}f \geq 2^k) \), hence (3.6) implies

\[
\int_{R^0} ((\mathbb{I}f)(t))^2 d\mu(t) \lesssim \sum_{k \in \mathbb{Z}} 2^{2k} \text{Cap}(\mathbb{I}f \geq 2^k).
\]

The last ingredient is the Strong Capacitary Inequality

\[
\sum_{k \in \mathbb{Z}} 2^{2k} \text{Cap}(\mathbb{I}f \geq 2^k) \lesssim \|f\|^2,
\]

which happens to be true on the bi-tree, see [3].

In what follows we will deduce the subcapacitary condition (3.6) from (3.1).

3.4.2. Step C. We argue by contradiction. First we note that for any a measure \( \sigma \) and a set \( E \), if

\[
\sigma(E) \geq C \text{Cap } E,
\]

then

\[
\mathcal{E}[\sigma|_E] \geq C \sigma(E) = C|\sigma|_E,
\]

here \( \sigma|_E \) is the restriction of \( \sigma \) on \( E \). Indeed, let \( \nu_E \) be the equilibrium measure for the set \( E \), then \( \mathcal{V}^{\nu_E} \geq 1 \) on \( E \), and the energy \( \mathcal{E}[\nu_E - \frac{1}{C}\sigma|_E] \) is positive,

\[
0 \leq \mathcal{E} \left[ \nu_E - \frac{1}{C}\sigma|_E \right] = \int \mathcal{V}^{\nu_E - \frac{1}{C}\sigma|_E} d\left( \nu_E - \frac{1}{C}\sigma|_E \right) = \left( \int \mathcal{V}^{\nu_E} d\nu_E - \frac{1}{C} \int \mathcal{V}^{\nu_E} d\sigma|_E \right) - \left( \frac{1}{C} \int \mathcal{V}^{\nu_E} d\sigma|_E - \frac{1}{C^2} \int \mathcal{V}^{\nu_E} d\sigma|_E \right) 
\]

\[
= \left( \text{Cap } E - \frac{1}{C} \int \mathcal{V}^{\nu_E} d\sigma|_E \right) - \left( \frac{1}{C} \int \mathcal{V}^{\nu_E} d\sigma|_E - \frac{1}{C^2} \mathcal{E}[\sigma|_E] \right) \leq \frac{1}{C}(C \text{Cap } E - |\sigma|_E) - \frac{1}{C^2}(C|\sigma|_E - \mathcal{E}[\sigma|_E]),
\]

and we are done.
Assume that there exists a set $F_0 \subset \mathbb{R}^0$ such that
\[ \mu(F_0) \geq C \text{Cap} F_0 \]
for some very large constant $C$ (to be specified later). Denote $\mu|_{F_0}$ by $\mu_0$.

The argument above yields
\[ (3.7) \quad \mathcal{E}[\mu_0] \geq C |\mu_0|. \]

Now we claim that (3.7) and (3.1) imply the existence of a set $F_1 \supset F_0$, such that
\[ (3.8) \quad \mathcal{E}[\mu_1] \geq 10C |\mu_1|, \]

here $\mu_1 = \mu|_{F_1}$.

**Remark.** The lack of subcapacitory property (3.6) was used only once — to produce (3.7). From now on we only use the domination of mass by energy condition.

3.4.3. Step D, part 1: making $\mu_0$ to be almost equilibrium. We start by introducing some additional notation. Given a (rectangular) set $E \subset \mathbb{R}^0$ and a measure $\nu$ we define the local energy of $\nu$ at $E$
\[ \mathcal{E}_E[\nu] := \sum_{Q \in \mathcal{D}(\mathbb{R}^0), Q \subset E} (\nu(Q))^2. \]

In particular we have $\mathcal{E}_{E_0}[\nu] = \mathcal{E}[\nu]$.

Now we have (3.7), that is
\[ \int \nabla^{\mu_0} d\mu_0 \geq \int C d\mu_0, \]
which means that $\nabla^{\mu_0} \geq \frac{C}{3}$ on a major part of supp $\mu_0$. For now we want to get rid of those points in supp $\mu_0$ where the potential is not large enough whilst conserving the total energy, we do so by the power of the following lemma

**Lemma 3.3.** Let $\nu$ be a non-negative measure on $\mathbb{R}^0$ (we remind that we always assume $\nu$ to have a piece-wise constant density on dyadic squares of $K$-th generation for some large number $K$), supp $\nu = E \subset \mathbb{R}^0$ and
\[ \mathcal{E}[\nu] = \int \nabla^{\nu} d\nu \geq C \nu(E) = C |\nu|. \]

Then there exists a set $E \subset E$ such that
\[ \nabla^{\tilde{\nu}} \geq \frac{C}{3}, \quad \text{on } \tilde{E}, \]
and
\[ \mathcal{E}[^{\tilde{\nu}}] \geq \frac{1}{6} \mathcal{E}[\nu]. \]

Here $\tilde{\nu} := \nu|_{E}$. 
Proof. First we assume that $C = 3$ (otherwise we just rescale).
Let $E_0 := \{ t \in E : \mathcal{V}^\nu \leq 1 \}$ and $\sigma_0 := \nu|_{E_0}$. Assume we have constructed
$\sigma_j$, $j = 0, \ldots, k - 1$, and the sets $E_j$. We then define $E_k$ to be
$$E_k = \{ \omega \in E \setminus \bigcup_{j=0}^{k-1} E_j : \mathcal{V}^\nu - \sum_{j=0}^{k-1} \sigma_j(\omega) \leq 1 \},$$
and we let $\sigma_k = \nu|_{E_k}$.
Since $T^2$ is finite, the procedure must stop at some (possibly very large)
number $N$, i.e. $E_j = \emptyset$ for $j > N$. We let $E_\infty = E \setminus \bigcup_{j=0}^{N} E_j$ (we do not
know yet if this set is non-empty), $\sigma_\infty = \nu|_{E_\infty}$. By construction we have
$$\forall^{\sigma_\infty}(\omega) \geq 1, \; \omega \in E_\infty.$$ Next we compute the energy of $\nu$,
$$\mathcal{E}[\nu] = \int \mathcal{V}^\nu d\nu = \sum_{N \geq j \geq 0} \sum_{N \geq k \geq 0} \int \mathcal{V}^{\sigma_j} d\sigma_k = 2 \sum_{N \geq k \geq 0} \sum_{N \geq j \geq k} \int \mathcal{V}^{\sigma_j} d\sigma_k =$$
$$2 \sum_{N \geq k \geq 0} \int \mathcal{V}^{\sum_{N \geq j \geq k} \sigma_j} d\sigma_k + 2 \int \mathcal{V}^{\sigma_\infty} d\sigma_\infty = 2 \sum_{N \geq k \geq 0} \int \mathcal{V}^{\nu - \sum_{N \geq j \geq k} \sigma_j} d\sigma_k +$$
$$2 \int \mathcal{V}^{\sigma_\infty} d\sigma_\infty \leq 2 \sum_{k=0}^{N} \int d\sigma_k + 2\mathcal{E}[\sigma_\infty] = 2|\nu| + 2\mathcal{E}[\sigma_\infty].$$
Since $\mathcal{E}[\nu] \geq 3|\nu|$ by assumption, we have
$$\mathcal{E}[\sigma_\infty] \geq \frac{1}{6} \mathcal{E}[\nu],$$
it remains to let $\tilde{\nu} := \sigma_\infty$, $\tilde{E} := E_\infty$, and we are done. \qed

We apply this lemma to $\mu_0$ and $F_0$ (we remind that $\mu_0 = \mu|_{F_0}$, so that
$\operatorname{supp} \mu_0 \subset F_0$) obtaining the set $F_0i \subset F_0$ and a measure $\mu_0i = \mu|_{F_0i}$
that satisfies $\forall^{\mu_0i} \geq C_1$ on $F_0i$ and $\mathcal{E}[\mu_0i] \geq \frac{1}{3} \mathcal{E}[\mu_0]$.

3.4.4. Step D, part 2: propagating 'badness' on a larger set. First we state
another lemma that allows us to estimate the total energy of an almost equi-
librium measure by its local energy at a certain level set (it is a byproduct of
Strong Capacitary Inequality on the bi-tree, see [3] for the proof and further
details)

Lemma 3.4. Let $\nu \geq 0$ be a measure on $R^0$ such that
$$\forall^\nu(t) \geq C_1, \; t \in \operatorname{supp} \nu.$$ Given $\varepsilon \in (0, 1)$ define
$$E_\varepsilon := \{ t \in R^0 : \forall^\nu(t) \geq \varepsilon C_1 \} .$$ There exists a constant $\varepsilon_0 > 0$, independent of $\mu$ and $C_1$, such that
$$\mathcal{E}_{E_{\varepsilon_0}}[\nu] \geq \frac{1}{2} \mathcal{E}[\nu].$$
Remark. The constant $\epsilon_0$ here actually depends on the dimension of the polytree $T^d$ (in this argument we are dealing with a bi-tree and $d = 2$), and possibly should look like $c^{-d}$ (with $c \sim 10$).

Define

\begin{equation}
F_{0ii} := \left\{ t \in R^0 : \forall \mu_{0i} (t) \geq \epsilon_0 \cdot \frac{C}{3} \right\},
\end{equation}

where $\epsilon_0$ is a constant from Lemma 3.4 and

\begin{equation}
F_1 := F_0 \bigcup F_{0ii},
\end{equation}

(the set $F_0$ may or may not be inside $F_{0ii}$, therefore we consider their union).

We aim to show (3.8).

Let $\mu_{0ii} = \mu|_{F_{0ii}}$. By (3.11) the mass of $\mu$ at $F_{0ii}$ dominates the local energy of $\mu$ on this set, which coincides with the local energy of $\mu_{0ii}$ there (since $\mu(Q) = \mu_{0ii}(Q)$ for $Q \subset F_{0ii}$). Clearly $E_{F_{0ii}}[\mu] \geq E_{F_{0ii}}[\mu_{0i}]$ ($\mu_{0i}$ is strictly smaller than $\mu$, being its restriction on $F_{0i}$). On the other hand, the set $F_{0ii}$ is chosen in such a way that the local energy of $\mu_{0i}$ dominates its total energy, $E_{F_{0ii}}[\mu_{0i}] \geq \frac{1}{2} \mathcal{E}[\mu_{0i}]$ (Lemma 3.4 with $\nu = \mu_{0i}$ and $E = F_{0ii}$).

We therefore have

$$
\mu(F_{0ii}) \geq \sum_{Q \subset F_{0ii}, Q \in \mathcal{D}(R^0)} (\mu)^2_Q |Q|^2 = \sum_{Q \subset F_{0ii}, Q \in \mathcal{D}(R^0)} \mu^2(Q) = E_{F_{0ii}}[\mu] \geq \frac{1}{2} \mathcal{E}[\mu_{0i}].
$$

By (3.11) the measure $\frac{3\mu_{0i}}{C \epsilon_0}$ is admissible for $F_{0ii}$ (i.e. $\frac{3\mu_{0i}}{C \epsilon_0} \geq 1$ on $F_{0ii}$), hence

$$
\text{Cap}_{F_{0ii}} \leq \frac{3^2}{C^2 \epsilon_0^2} \mathcal{E}[\mu_{0i}].
$$

Combining the estimates above we arrive at

$$
\mu(F_{0ii}) \geq \frac{1}{2} \mathcal{E}[\mu_{0i}] \geq \frac{C^2}{2 \cdot 9 \epsilon_0^2} \text{Cap}_{F_{0ii}}.
$$

One of the ways to define a capacity of a set is to pick up the largest mass $|\sigma|$ among all candidates $\sigma$ (supported on that set) such that $\mathcal{E}[\sigma] \leq |\sigma|$ (this is the argument from the beginning of subsection 3.4.2). Thus,

\begin{equation}
\mathcal{E}[\mu_{0i}] \geq \frac{C^2}{2 \cdot 9 \epsilon_0^2} |\mu_{0ii}|.
\end{equation}

It remains to see that (by Lemma 3.3) we have

$$
|\mu_{0ii}| = \mu(F_{0ii}) \geq \frac{1}{2} \mathcal{E}[\mu_{0i}] \geq \frac{1}{3} \frac{1}{2} \mathcal{E}[\mu_0] \geq \frac{C}{6} |\mu_0|.
$$
Plugging this estimate into (3.12) we obtain

\[
\mathcal{E}[\mu_1] \geq \mathcal{E}[\mu_{0\ii}] \geq \frac{C^2}{18\varepsilon_0} |\mu_{0\ii}| = \frac{C^2}{18\varepsilon_0} \left( \frac{1}{2}|\mu_{0\ii}| + \frac{1}{2}|\mu_{0\ii}| \right) \geq \frac{C^2}{18\varepsilon_0} \left( \frac{1}{2}|\mu_{0\ii}| + \frac{C}{12}|\mu_0| \right) \geq \frac{C^2}{36\varepsilon_0} (|\mu_{0\ii}| + |\mu_0|) = \frac{C^2}{36\varepsilon_0} (\mu(F_{0\ii}) + \mu(F_0)) \geq \frac{C^2}{36\varepsilon_0} \mu(F_1) \geq 10C|\mu_1|,
\]

if the constant C is large enough. We have (3.3).

3.4.5. The last step. Now we iterate the procedure — given a bad set \( F_k \) we can construct a larger and a worse set \( F_{k+1} \), until we exhaust the whole bi-tree (so \( F_{k_0} = R^0 \) for some \( k_0 \)). Clearly

\[
\mathcal{E}[\mu] = \mathcal{E}[\mu_{k_0}] \geq C \mu(F_{k_0}) = C \mu(R^0),
\]

and this is in a direct contradiction with (3.1) for \( E = F_{k_0} = R^0 \)

\[
\sum_{Q \in \mathcal{D}(R^0)} \mu(Q)^2 = \mathcal{E}[\mu] \leq \mu(R^0).
\]

References

[1] Adams, David R.; Hedberg, Lars Inge Function spaces and potential theory. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 314. Springer-Verlag, Berlin, 1996. xii+366 pp. ISBN: 3-540-57060-8

[2] Nicola Arcozzi, Irina Holmes, Pavel Mozolyako, Alexander Volberg, Bi-parameter embedding without capacity and measures with restriction energy condition, preprint, 2018, pp. 1–9.

[3] Nicola Arcozzi, Pavel Mozolyako, Karl-Mikael Perfekt, Giulia Sarfatti, Carleson measures for the Dirichlet space on the bidisc, preprint 2018.

[4] F. Nazarov, S. Treil, A. Volberg, The Bellman functions and two-weight inequalities for Haar multipliers, Journal of the AMS, Vol. 12, Number 4, 1999.