Random Dyadic Tilings of the Unit Square

Svante Janson, Dana Randall and Joel Spencer

April 26, 2002

Abstract

A "dyadic rectangle" is a set of the form $R = [a2^{-s}, (a+1)2^{-s}] \times [b2^{-t}, (b+1)2^{-t}]$, where s and t are non-negative integers. A dyadic tiling is a tiling of the unit square with dyadic rectangles. In this paper we study n-tilings, which consist of 2^n nonoverlapping dyadic rectangles, each of area 2^{-n}, whose union is the unit square. We discuss some of the underlying combinatorial structures, provide some efficient methods for uniformly sampling from the set of n-tilings, and study some limiting properties of random tilings.

1 Introduction

We shall examine tilings of the unit square of a special type. By a *dyadic rectangle* we mean a set of the form

$$R = [a2^{-s}, (a+1)2^{-s}] \times [b2^{-t}, (b+1)2^{-t}]$$

where s, t are nonnegative integers and a, b are integers with $0 \leq a < 2^s$ and $0 \leq b < 2^t$. An n-tiling of the unit square is a set of 2^n dyadic rectangles, each of area 2^{-n}, whose union is the unit square $[0,1] \times [0,1]$. (Overlap at the edges does not concern us.) Figure 1 gives examples of such n-tilings. We shall often just speak of tilings when the value n is understood. Let $\mathcal{T}_n$ be the set of all n-tilings.

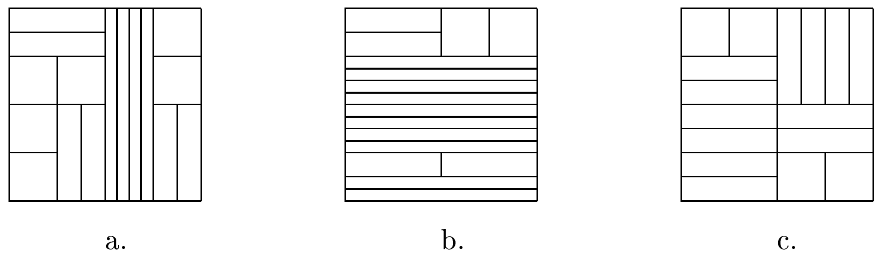

Figure 1: Examples of dyadic tilings

Definition. A tiling has a *vertical cut* if the line $x = \frac{1}{2}$ cuts through none of its rectangles. It has a *horizontal cut* if the line $y = \frac{1}{2}$ cuts through none of its rectangles.

We note that Figure 1a has a vertical cut, Figure 1b has a horizontal cut, and Figure 1c has both a vertical and a horizontal cut. We emphasize that cuts, as opposed to the struts of Section 6.2, must cut the square precisely in half. Cuts play a critical role in the analysis of tilings due to the following result.

Theorem 1.1. *Every tiling has either a vertical cut or a horizontal cut. (It may have both.)*

Proof. If $x = \frac{1}{2}$ cuts through a rectangle then that rectangle must be of the form $R = [0,1] \times I$. If also $y = \frac{1}{2}$ cuts through a rectangle then that rectangle must be of the form $S = J \times [0,1]$. But then S, T overlap in $I \times J$. $\square$

Let A_n denote the number of n-tilings. The square itself provides the unique 0-tiling so that $A_0 = 1$. $A_1 = 2$ since the square may be split into left and right halves or top and bottom halves. Some effort yields $A_2 = 7$, and by the following recursion [10] we obtain $A_3 = 82$, $A_4 = 11047$, $A_5 = 198860242$, $A_6 = 64197955389505447$, $\ldots$. (For convenience we let $A_{-1} = 0$.)

Theorem 1.2. *For $n \geq 1$,*

$$A_n = 2A_{n-1}^2 - A_{n-2}^4. \tag{1.1}$$

Proof [10]. Consider n-tilings with a vertical cut. These consist of n-tilings of $[0, \frac{1}{2}] \times [0,1]$ and $[\frac{1}{2}, 1] \times [0,1]$. Dilating $x \to 2x$, n-tilings of $[0, \frac{1}{2}] \times [0,1]$ are equivalent to $(n-1)$-tilings of the unit square. Dilating $x \to 2x - 1$, n-tilings of $[\frac{1}{2}, 1] \times [0,1]$ are equivalent to $(n-1)$-tilings of the unit square. Hence there are A_{n-1}^2 such tilings. Similarly there are A_{n-1}^2 n-tilings with a horizontal cut. This gives, by Theorem 1.1, all n-tilings but we have overcounted by those n-tilings with both horizontal and vertical cuts. Such tilings consist of n-tilings of each of the four subsquares $[0, \frac{1}{2}] \times [0, \frac{1}{2}]$, $[\frac{1}{2}, 1] \times [0, \frac{1}{2}]$, $[0, \frac{1}{2}] \times [\frac{1}{2}, 1]$, $[\frac{1}{2}, 1] \times [\frac{1}{2}, 1]$. Dilating $(x, y) \to (2x, 2y)$, n-tilings of $[0, \frac{1}{2}] \times [0, \frac{1}{2}]$ are equivalent to $(n-2)$-tilings of the unit square, and similarly for the other three subsquares. Hence there are A_{n-2}^4 n-tilings with both horizontal and vertical cuts. $\square$

The recursion of Theorem 1.2 is believed not to admit a closed solution. The asymptotics of A_n have been carefully studied in [10]; here we note only that

$$A_n \sim \phi^{-1} \omega^{2^n} \tag{1.2}$$

where $\omega = 1.84454757\ldots$ (ω does not appear to have a nice form) and ϕ is the golden ratio $\phi = (1 + \sqrt{5})/2 = 1.6180\ldots$.

Dyadic rectangles were used previously to analyze the packing of random axis-parallel rectangles of arbitrary size [7]. The more constrained dyadic tilings proved to have a fascinating structure, which motivated our current efforts. Much remains to be studied however, and several open problems are given below.

We give some deterministic results on the set of all n-tilings in Section 2. In Section 3 we define two representations as labeled binary trees, which will play an important role later on.

By a *random tiling* we mean a uniformly sampled tiling in $\mathcal{T}_n$, for some given n; in other words, each tiling is chosen with the same probability $1/A_n$. In Section 4, we present a method to randomly sample tilings with this uniform distribution. The method is both practically useful if one wants to generate random tilings, and theoretically useful in some of our later proofs. In Section 5 we discuss an alternative method to generate random tilings by running a Markov process; we show that two natural Markov processes are rapidly mixing. In Section 6 we study the asymptotic behavior as $n \to \infty$ of some properties of random tilings.

Problem 1.3. More generally, one may naturally define dyadic boxes in d dimensions and from that, n-tilings. For $d \geq 3$, however, Theorem 1.1 fails, as n-tilings do not necessarily have any cuts. The structure of the family of n-tilings and the asymptotics of the number $A_{n,d}$ of such tilings remain completely open. In particular, we do not know whether, for $d \geq 3$ fixed, the 2^n-th root of $A_{n,d}$ approaches infinity.

2 The lattice of tilings

There is a useful height function which lends insight into the set of dyadic tilings. Define the *height* $h(t)$ of a dyadic $2^{-k} \times 2^{-l}$ rectangle t with area 2^{-n} to be k (equivalently, $n - l$), and define the total height $H(T)$ of a tiling T to be the sum of the heights of all rectangles in it. Note that since the height of a single rectangle of area 2^{-n} is one of the numbers $\{0, \dots, n\}$,

$$0 \leq H(T) \leq n2^n, \qquad \text{for all } \ T \in \mathcal{T}_n. \tag{2.1}$$

If T is a tiling and $p \in [0,1]^2$, we let $T(p)$ be the dyadic rectangle in T containing p. (If there are two or more such rectangles in T, which happens only if p lies on their boundaries, we choose for definiteness the one containing points north-east of p. This is not important, and we could avoid this complication completely by considering only irrational $p \in [0,1]^2$.) A tiling $T \in \mathcal{T}_n$ then is completely described by its *height function* $h(T) : [0,1]^2 \to \{0, 1, \dots, n\}$ defined by $h(T)(p) := h(T(p))$. Note that

$$H(T) = 2^n \int_{[0,1]^2} h(T)(p)\, dp. \tag{2.2}$$

The height function allows us to define a partial order on the set of n-tilings. Given two tilings $T_1, T_2 \in \mathcal{T}_n$, we say $T_1 \preceq T_2$ if $h(T_1) \leq h(T_2)$, i.e., if $h(T_1(p)) \leq h(T_2(p))$ for all $p \in [0,1]^2$. With these definitions, we find the following.

Theorem 2.1. *The partial order on $\mathcal{T}_n$ defines a distributive lattice.*

Proof. First, there are unique highest and lowest elements in $\mathcal{T}_n$. Namely, the highest tiling is the all vertical tiling, consisting of only $2^{-n} \times 1$ rectangles (which has height function constant n), and the lowest is the all horizontal tiling, consisting of 1×2^{-n} rectangles (which has height function constant 0).

Let T_1 and T_2 be any two tilings in $\mathcal{T}_n$. We define the join

$$T_1 \vee T_2 = \{\max(T_1(p), T_2(p)) : p \in [0,1]^2\},$$

where $\max(T_1(p), T_2(p))$ is the tile with larger height. Note that this is well-defined since every irrational point $p \in [0,1]^2$ lies on exactly one dyadic tile of each height. Similarly, define the meet

$$T_1 \wedge T_2 = \{\min(T_1(p), T_2(p)) : p \in [0,1]^2\},$$

where $\min(T_1(p), T_2(p))$ is the tile with smaller height.

We need to argue that the meet and join always yield valid tilings. Consider first $T_1 \vee T_2$. Clearly every point in $[0,1]^2$ is covered at least once. Suppose that there exists a point p_1 which is covered by (the interior of) two tiles t and t' in $T_1 \vee T_2$. Clearly t and t' have different heights, and further they must come from different tilings. We can assume without loss of generality that $t \in T_1$, $t' \in T_2$ and that $h(t) > h(t')$. Recalling that $t' \in \{\max(T_1(p), T_2(p))\}$, there must exist an irrational point $p_2 \in t'$ such that $T_2(p_2) = t'$ and $h(t') \geq h(t'')$, where $t'' = T_1(p_2)$. Let $t = I \times J$, $t' = I' \times J'$ and $t'' = I'' \times J''$. Since $p_1 \in t \cap t'$ and $h(t) > h(t')$, and all intervals are dyadic, we have $I \subset I'$ and $J \supset J'$. Since $p_2 \in t' \cap t''$ and $h(t') \geq h(t'')$, we have $I' \subseteq I''$ and $J' \supseteq J''$. Consequently, $t \neq t''$ but $t \cap t'' = I \times J'' \neq \emptyset$, which contradicts the fact that t and t'' are different tiles in T_1. Hence tiles in $T_1 \vee T_2$ can never intersect. An analogous argument shows that $T_1 \wedge T_2$ is a proper tiling.

Note that the height function satisfies

$$h(T_1 \vee T_2) = \max(h(T_1), h(T_2)), \tag{2.3}$$

$$h(T_1 \wedge T_2) = \min(h(T_1), h(T_2)). \tag{2.4}$$

It follows that the height function defines an order-preserving bijection of $\mathcal{T}_n$ onto a sublattice of the distributive lattice of all functions $[0,1]^2 \to \{0, 1, \ldots, n\}$. Therefore $\mathcal{T}_n$ also forms a distributive lattice. ☐

Let $\widetilde{T}_k$ denote the special tiling with $2^{-k} \times 2^{k-n}$ rectangles, $k = 0, \ldots, n$; thus $\widetilde{T}_k$ has height function constant k. As noted above, $\widetilde{T}_0$ is the lowest tiling and $\widetilde{T}_n$ is the highest. Other of these special tilings also have useful properties.

Theorem 2.2. *An n-tiling T has a horizontal cut if and only if $T \preceq \widetilde{T}_{n-1}$. It has a vertical cut if and only if $T \succeq \widetilde{T}_1$.*

Proof. T has a horizontal cut if and only if it contains no $2^{-n} \times 1$ rectangle, i.e., if and only if $h(T)(p) \leq n-1$ for every $p \in [0,1]^2$. The second part is similar. ☐

Theorem 2.3. *Suppose that T_1, T_2 are n-tilings with $n \geq 2$. If $T_1 \preceq T_2$, T_1 has a horizontal cut, and T_2 has a vertical cut, then there exists a tiling T_3 with both vertical and horizontal cuts such that $T_1 \preceq T_3 \preceq T_2$.*

Proof. Define $T_3 = T_1 \vee \widetilde{T}_1$. By Theorem 2.2, $\widetilde{T}_1 \preceq T_2$, so $T_1 \preceq T_3 \preceq T_2$. By Theorem 2.2 again, $T_1 \preceq \widetilde{T}_{n-1}$, and trivially $\widetilde{T}_1 \preceq \widetilde{T}_{n-1}$, so $\widetilde{T}_1 \preceq T_3 \preceq \widetilde{T}_{n-1}$, which by a final application of Theorem 2.2 completes the proof. ☐

If T is a tiling and R is a dyadic rectangle such that R is a union of tiles in T, we can obtain new tilings by rotating the part of the tiling inside R by a multiple of $90°$. (By "rotating" a non-square region, we really mean a rotation followed by appropriate dilations in the coordinate directions to make the result fit in the same region again.) Rotations will be important later. Here we are only concerned with the simplest non-trivial case.

We make $\mathcal{T}_n$ into a directed graph by defining an edge $T_1 \to T_2$ if T_1 and T_2 can be obtained from each other by rotating a dyadic rectangle of area 2^{-n+1}, with $T_1 \prec T_2$.

In particular, $T_1 \to T_2$ if there is a dyadic rectangle R of area 2^{-n+1} such that T_1 and T_2 coincide outside R, T_1 contains the two horizontal halves of R, while T_2 contains the two vertical halves of R. It follows that if $T_1 \to T_2$, then $H(T_2) = H(T_1) + 2$.

The following theorem yields a natural connection between this graph structure and the partial order on $\mathcal{T}_n$. Namely, we include an edge $T_1 \to T_2$ if and only if T_2 is a minimal successor of T_1. It further follows from the theorem that $T_1 \to T_2$ if and only if $T_1 \preceq T_2$ and $H(T_2) = H(T_1) + 2$, thus yielding yet another characterization.

Theorem 2.4. *Let $T_1, T_2 \in \mathcal{T}_n$. Then $T_1 \preceq T_2$ if and only if there exists an oriented path from T_1 to T_2 in the directed graph $\mathcal{T}_n$. Every such path has length $\frac{1}{2}H(T_2) - \frac{1}{2}H(T_1)$.*

Proof. The existence of such a path whenever $T_1 \preceq T_2$ is clear for $n \le 1$. For larger n we use induction in n. If T_1 has a vertical cut, then T_2 also has (by Theorem 2.2), and the existence of an oriented path follows from the induction hypothesis by considering the left and right halves of the tilings separately. If both T_1 and T_2 have horizontal cuts, we consider similarly the upper and lower halves separately. By Theorem 1.1, the only remaining case is when T_1 has a horizontal cut and T_2 a vertical cut. In this case we use Theorem 2.3 and find a tiling T_3 such that, by the previous cases, there exist paths from T_1 to T_3 and from T_3 to T_2; we combine these paths into one. This completes the proof of existence of an oriented path from T_1 to T_2 whenever $T_1 \preceq T_2$. The converse is immediate.

The final assertion follows because $H(T_2) = H(T_1) + 2$ when $T_1 \to T_2$. □

Corollary 2.5. *The total height $H(T)$ is twice the common length of the oriented paths from the lowest tiling $\widetilde{T}_0$ to T.* □

Theorem 2.6. *Ignoring orientations, $\mathcal{T}_n$ is a connected graph. The distance $d(T_1, T_2)$ between two tilings T_1 and T_2 in this graph equals*

$$2^{n-1} \int_{[0,1]^2} |h(T_1)(p) - h(T_2)(p)| \, dp = 2^{n-1} \|h(T_1) - h(T_2)\|_{L^1([0,1]^2)}.$$

Proof. Combining the oriented paths given by Theorem 2.4 from T_1 and T_2 to $T_1 \vee T_2$, and reversing the latter, we obtain a path from T_1 to T_2 whose length, by (2.2) and (2.3), is

$$\begin{aligned}
\tfrac{1}{2}H(T_1 \vee T_2) &- \tfrac{1}{2}H(T_1) + \tfrac{1}{2}H(T_1 \vee T_2) - \tfrac{1}{2}H(T_2) \\
&= 2^{n-1} \int_{[0,1]^2} \big(2h(T_1 \vee T_2)(p) - h(T_1)(p) - h(T_2)(p)\big) \, dp \\
&= 2^{n-1} \int_{[0,1]^2} |h(T_1)(p) - h(T_2)(p)| \, dp.
\end{aligned}$$

Conversely, if the distance $d(T_1, T_2) = 1$, then $T_1 \to T_2$ or $T_2 \to T_1$, and $\|h(T_1) - h(T_2)\|_{L^1([0,1]^2)} = 2^{-n}|H(T_1) - H(T_2)| = 2^{1-n}$. By the triangle inequality for the L^1 metric, we have in general $\|h(T_1) - h(T_2)\|_{L^1([0,1]^2)} \le 2^{1-n} d(T_1, T_2)$. □

Corollary 2.7. *The diameter of the undirected graph $\mathcal{T}_n$ is $n2^{n-1}$. This distance is attained by the lowest and highest tilings $\widetilde{T}_0$ and $\widetilde{T}_n$.* □

It is easily seen that the distance between any other pair of tilings is strictly smaller.

Problem 2.8. Study further properties of $\mathcal{T}_n$ as a lattice or graph. For example, what is the distribution of the vertex degrees?

3 Tree representations

The existence of cuts promotes a useful representation of dyadic tilings in terms of labeled binary trees. Trees are natural in this context because of the hierarchical relationship of the cuts. The first cut divides the unit square into two halves, each of which can be interpreted as a dyadic tiling in $\mathcal{T}_{n-1}$ through dilation. These in turn have cuts, and so forth. The labels on a tree capture whether corresponding cuts are horizontal or vertical.

We will use two different versions of this idea: one (HV-trees) where the labels specify the *absolute* orientations of the cuts and one (AD-trees) where the labels specify the *relative* orientations.

Recall that a binary tree either is empty or consists of a root and two (binary) subtrees attached to the root. We find it convenient to say that the root of a binary tree has height 1. Thus a complete binary tree of height n has $2^n - 1$ nodes: 2^{k-1} with height k, for $k = 1, \ldots, n$. We also say that the nodes with height k lie on level k. The empty tree has height 0.

3.1 HV-trees

A complete binary tree of height n whose $2^n - 1$ nodes are labeled H or V defines an n-tiling by the following procedure:

1. If the tree is empty ($n = 0$) then Exit.
2. If the root is labeled H, make a horizontal cut. If the root is labeled V, make a vertical cut.
3. Continue recursively with the two halves separately, using the left subtree of the root for one half (for definiteness, the lower or left half, say) and the right subtree for the other half.

Conversely, Theorem 1.1 implies that every n-tiling is produced in this way by some labeled complete binary tree. However, the tree is in general not unique, since the unit square (or a subrectangle) may have both a vertical and a horizontal cut; indeed, there are 2^{2^n-1} complete binary trees of height n whose nodes are labeled H or V, which is far greater than the number of tilings in $\mathcal{T}_n$.

In order to obtain a unique representation by labeled binary trees, we decide to use the label V whenever we have a choice of proceeding with either H or V. We make the following definition.

Definition. A complete binary tree whose nodes are labeled H or V is an *HV-tree* if there is no node labeled H that has two children labeled V.

Let $\mathcal{T}_n^{HV}$ be the set of HV-trees of height n.

Lemma 3.1. *An HV-tree with root labeled H defines a tiling without vertical cut.*

Proof. This is trivial for height 0 or 1. For larger trees we use induction. The root has, by the definition of HV-trees, at least one subtree whose root is labeled H, so by induction the lower or upper half of the tiling does not have a vertical cut. □

Theorem 3.2. *The construction above yields a bijection between $\mathcal{T}_n^{HV}$ and $\mathcal{T}_n$.*

Proof. Given a tiling, we create a corresponding HV-tree as follows. If there is a vertical cut in the tiling, then label the root V and continue recursively with the left and right halves. If there is no vertical cut, there is by Theorem 1.1 a horizontal cut; we then label the root H and continue recursively with the lower and upper halves.

Note that if the root is labeled H, there is no vertical cut and thus at least one of the two halves produced by the first cut has no vertical cut; consequently, the two children of the root cannot both be labeled V. The same applies to all later stages of the construction, which shows that we have constructed an HV-tree.

Clearly, the tree defines the given tiling. Moreover, it follows from Lemma 3.1 that any two HV-trees defining the same tiling have to have the same root label, and by recursion they have to be identical. Hence each tiling corresponds to a unique HV-tree. □

3.2 AD-trees

In the second representation, we use complete binary trees with the labels A (agree) or D (disagree) to indicate whether the cut is parallel or orthogonal to its parent (i.e., the preceding cut). We arbitrarily define the (absent) parent of the first cut to have vertical orientation. We formalize the construction of a tiling given a complete binary tree labeled with A and D as follows.

1. Initialize by defining the parent cut to be the left edge of the unit square.
2. If the tree is empty ($n = 0$) then Exit.
3. If the root is labeled A, make a cut parallel to the parent cut. If the root is labeled D, make a cut orthogonal to the parent cut.
4. Continue recursively (from Step 2) with the two halves separately, using the two subtrees of the root and in both cases setting the parent cut equal to the cut just made. More precisely, if the root is labeled A, use the left subtree for the half nearest the parent cut, and if the root is labeled D, use the left subtree for the left half, viewed from the parent cut.

The specification in Step 4 of the order of the subtrees is chosen such that changing a single label in the tree corresponds to rotating the corresponding subtiling $\pm 90°$ (clockwise if changing an A to a D and counterclockwise if changing a D to an A). This labeling is illustrated in Figure 3.2.

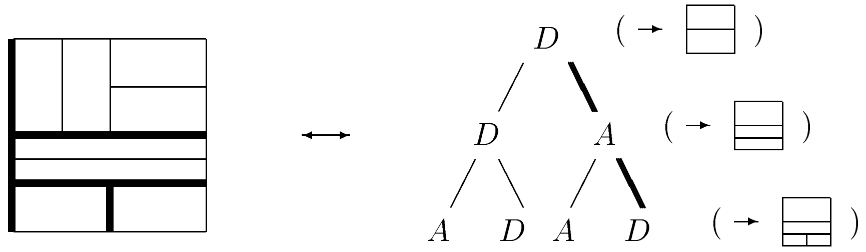

Figure 2: The correspondence between dyadic tilings and AD-trees.

Just as for HV-trees, every tiling is defined by some labeled complete binary tree, but the correspondence between labeled trees and tilings is not bijective. More precisely, whenever a node is labeled A and its two children are labeled D, we would get the same tiling if we were to label all three nodes D, followed by appropriate relabelings at the descendants of these nodes. We resolve this by only allowing the labeling where all three such nodes are labeled D. We say that any tree with an invalid labeling (i.e., a node labeled A which has two children labeled D) has a *badly labeled subtree*. This motivates the following definition.

Definition. A complete binary tree whose nodes are labeled A or D is an *AD-tree* if there is no node labeled A that has two children labeled D.

Let $\mathcal{T}_n^{AD}$ be the set of AD-trees of height n.

Theorem 3.3. *The construction above yields a bijection between $\mathcal{T}_n^{AD}$ and $\mathcal{T}_n$.*

Proof. This is very similar to the proof of 3.2. Given a tiling, we can create a unique AD-tree recursively, beginning at the root, by always choosing D when we have a choice. We omit the details. □

4 Recursive algorithms for sampling

The recursive formula for the number of tilings of each size suggests a natural method for sampling tilings, or equivalently HV-trees or AD-trees, uniformly. Starting with the unit square, we calculate the probability that there is a horizontal or vertical cut, and then recursively determine probabilities of each cut in the two halves that are formed. Analogously, we can use these probabilities to determine the label of the root of the tree. Each subtree is then labeled so as to avoid introducing badly labeled subtrees. We formalize how to do this in what follows. It turns out to be convenient to use the tree representations.

4.1 Probabilities at the root

We start by introducing some notation which will be useful for determining the relevant probabilities. Let

$$p_n = \mathbf{P}(\text{a random tiling in } \mathcal{T}_n \text{ has a vertical cut}) = \frac{A_{n-1}^2}{A_n}, \qquad n \geq 0. \tag{4.1}$$

By symmetry, the probability of a horizontal cut is the same. We have $p_0 = 0$, $p_1 = 1/2$, $p_2 = 4/7, \ldots$. Theorem 1.2 yields, by dividing by A_{n-1}^2, the recursion

$$p_n = \frac{1}{2 - p_{n-1}^2}, \qquad n \geq 1. \tag{4.2}$$

It follows easily from (4.2) that p_n increases to the smallest positive root of $x = 1/(2 - x^2)$, i.e.,

$$p_n \to \phi^{-1} = \phi - 1 = (\sqrt{5} - 1)/2, \qquad \text{as } n \to \infty.$$

Moreover, using $\frac{d}{dx}(2 - x^2)^{-1} = 2x(2 - x^2)^{-2} \leq 2\phi^{-3}$ for $0 \leq x \leq \phi^{-1}$, it follows by the mean value theorem and induction that

$$p_n = \phi - 1 + O((\phi^3/2)^{-n}). \tag{4.3}$$

4.2 The recursive construction

We will construct a random HV-tree in $\mathcal{T}_n^{HV}$ with a uniform distribution recursively, beginning by choosing the label of the root. Since there is an obvious bijection between HV-trees and AD-trees, given by relabeling $H \leftrightarrow A$ and $V \leftrightarrow D$, the same method can be used to construct uniformly distributed random AD-trees. Any of the bijections in Section 3 then yields a uniform random n-tiling. (An HV-tree and the AD-tree obtained by relabeling correspond to different tilings, so the tilings produced by a particular simulation will depend on whether we use HV-trees and AD-trees, but both constructions yield the same uniform distribution. Similarly, note that the algorithm below is highly asymmetric with respect to H and V, although we know that the resulting distribution of tilings is invariant under rotation.)

If the root is labeled V, its two subtrees can be any trees in $\mathcal{T}_{n-1}^{HV}$, and we may continue recursively. If the root is labeled H, however, and $n \geq 2$, we have the constraint that its two children must not both be labeled V; this introduces a dependency between the subtrees that makes a straight-forward recursion impossible. In order to overcome this difficulty we look ahead, which we formalize as follows.

Definition. The *type* of a node in a HV-tree is one of the four symbols V, H_{HH}, H_{HV}, H_{VH}, chosen according to the following rules:

1. If the node is labeled V, its type is V.
2. If the node is labeled H and it is not a leaf, its type is H_{xy}, where x and y are the labels of its children.
3. If the node is labeled H and it is a leaf, its type is H_{HH}.

Note that the type of a node determines the label, but not conversely; however, the labeling of the whole tree determines the types, and conversely.

We define the type type(T) of a non-empty HV-tree T to be the type of its root, and define type$(\emptyset) = H_{HH}$.

Consequently, an HV-tree $T \in \mathcal{T}_n^{HV}$, $n \geq 1$, is described by its type and two trees $T_1, T_2 \in \mathcal{T}_{n-1}^{HV}$, with the constraints that

$$\begin{aligned} \text{type}(T) = H_{HH} &\implies \text{type}(T_1), \text{type}(T_2) \in \{H_{HH}, H_{HV}, H_{VH}\}; \\ \text{type}(T) = H_{HV} &\implies \text{type}(T_1) \in \{H_{HH}, H_{HV}, H_{VH}\},\ \text{type}(T_2) = V; \\ \text{type}(T) = H_{VH} &\implies \text{type}(T_1) = V,\ \text{type}(T_2) \in \{H_{HH}, H_{HV}, H_{VH}\}. \end{aligned} \tag{4.4}$$

Apart from these constraints, T_1 and T_2 may be any HV-trees in $\mathcal{T}_{n-1}^{HV}$. This allows, in principle, a recursive generation of all trees in $\mathcal{T}_n^{HV}$. (This quickly becomes practically impossible, since $|\mathcal{T}_n^{HV}|$ grows very quickly. We will turn this recursive generation into a practical procedure for random generation of HV-trees.)

For trees of type V, there are no constraints on T_1 and T_2, so the number of such trees is A_{n-1}^2; hence the number of trees of other types is $A_n - A_{n-1}^2$. It follows from the rules above that the number of trees in $\mathcal{T}_n^{HV}$ $(n \geq 1)$ of the four types are

$$\begin{aligned} V &: \quad A_{n-1}^2 = p_n A_n \\ H_{HH} &: \quad (A_{n-1} - A_{n-2}^2)^2 = p_n (1 - p_{n-1})^2 A_n \\ H_{HV} &: \quad A_{n-2}^2 (A_{n-1} - A_{n-2}^2) = p_n p_{n-1} (1 - p_{n-1}) A_n \\ H_{VH} &: \quad A_{n-2}^2 (A_{n-1} - A_{n-2}^2) = p_n p_{n-1} (1 - p_{n-1}) A_n \end{aligned} \tag{4.5}$$

(Together, these numbers add up, as they should, to $2A_{n-1}^2 - A_{n-2}^4 = A_n$.

4.3 Recursive generation of random tilings

Let $\tau^{(n)}$ denote a random type $\tau \in \{V, H_{HH}, H_{HV}, H_{VH}\}$ with the distribution given by $\mathbf{P}(\tau^{(n)} = V) = p_n$, $\mathbf{P}(\tau^{(n)} = H_{HH}) = p_n(1 - p_{n-1})^2$, $\mathbf{P}(\tau^{(n)} = H_{HV}) = \mathbf{P}(\tau^{(n)} = H_{VH}) = p_n p_{n-1}(1 - p_{n-1})$. Let further $\tau_H^{(n)}$ denote $\tau^{(n)}$ conditioned on $\tau^{(n)} \neq V$; thus $\mathbf{P}(\tau_H^{(n)} = H_{HH}) = (1 - p_{n-1})/(1 + p_{n-1})$, $\mathbf{P}(\tau_H^{(n)} = H_{HV}) = \mathbf{P}(\tau_H^{(n)} = H_{VH}) = p_{n-1}/(1 + p_{n-1})$.

It follows from (4.5) that $\tau^{(n)}$ has the same distribution as the type of a randomly chosen tree in $\mathcal{T}_n^{HV}$, and thus $\tau_H^{(n)}$ has the same distribution as the type of a random tree in $\mathcal{T}_n^{HV}$ with the root labeled H. The discussion in the preceding section now shows that the following algorithm generates a uniformly distributed random element of $\mathcal{T}_n^{HV}$, for any given $n \geq 1$.

Algorithm 4.1.

1. *Select randomly a type for the root with the distribution* $\tau^{(n)}$.

2. *Recursively assign types to all other nodes such that if a node of height* k, $1 \leq k < n$, *is assigned a type* τ, *then its left and right children get types* τ_1 *and* τ_2 *selected as follows.*

   $\tau = V$**:** *Choose* τ_1 *and* τ_2*, independently, both with the distribution of* $\tau^{(n-k)}$.

   $\tau = H_{HH}$**:** *Choose* τ_1 *and* τ_2*, independently, both with the distribution of* $\tau_H^{(n-k)}$.

$\tau = H_{HV}$: *Choose* τ_1 *with the distribution of* $\tau_H^{(n-k)}$ *and let* $\tau_2 = V$.

$\tau = H_{VH}$: *Let* $\tau_1 = V$ *and choose* τ_2 *with the distribution of* $\tau_H^{(n-k)}$.

3. *All vertices with type* V *are labeled* V*; the others are labeled* H.

Next observe that since $p_n \to \phi^{-1}$ as $n \to \infty$, it follows that $\tau^{(n)} \xrightarrow{d} \tau^{(\infty)}$ and $\tau_H^{(n)} \xrightarrow{d} \tau_H^{(\infty)}$, with (using $\phi^2 = \phi + 1$ repeatedly)

$$\begin{aligned}
\mathbf{P}(\tau^{(\infty)} = V) &= \phi^{-1} = \phi - 1, & \mathbf{P}(\tau_H^{(\infty)} = V) &= 0,\\
\mathbf{P}(\tau^{(\infty)} = H_{HH}) &= \phi^{-5} = 5\phi - 8, & \mathbf{P}(\tau_H^{(\infty)} = H_{HH}) &= \phi^{-3} = 2\phi - 3,\\
\mathbf{P}(\tau^{(\infty)} = H_{HV}) &= \phi^{-4} = 5 - 3\phi, & \mathbf{P}(\tau_H^{(\infty)} = H_{HV}) &= \phi^{-2} = 2 - \phi,\\
\mathbf{P}(\tau^{(\infty)} = H_{VH}) &= \phi^{-4} = 5 - 3\phi, & \mathbf{P}(\tau_H^{(\infty)} = H_{VH}) &= \phi^{-2} = 2 - \phi.
\end{aligned}$$

We can use these asymptotic distributions to construct an asymptotic version of Algorithm 4.1.

Algorithm 4.2. *This is the same as Algorithm 4.1, but using the distributions* $\tau^{(\infty)}$ *and* $\tau_H^{(\infty)}$ *in Steps 1 and 2.*

Consequently, for fixed N and $n \to \infty$, the labels of the top N levels of a (uniform) random tree in $\mathcal{T}_n^{HV}$ converges (in distribution) to the outcome of Algorithm 4.2 (with $n = N$).

Remark 4.3. In Algorithms 4.1 and 4.2, we may process the nodes in any order such that a node is visited before its children. One natural choice, easily expressed as a recursive algorithm, is to travel depth-first, beginning with the left child, its left child, and so on.

Another choice is breadth-first, where we assign the types level by level, in order of the heights of the nodes. This version is useful for the arguments in Section 6. Moreover, it means that Algorithm 4.2 not only generates a random HV-tree of any given height; we can also regard Algorithm 4.2 as generating a random *infinite* HV-tree. The remarks above show that this random infinite tree is the limit in distribution of a (uniform) random tree in $\mathcal{T}_n^{HV}$, as $n \to \infty$, in the sense that the distribution of labels on any fixed finite part converges.

Remark 4.4. We have shown that Algorithm 4.1 generates uniformly distributed random HV-trees and thus uniformly distributed random tilings in $\mathcal{T}_n$. Evidently, one can also produce random tilings in $\mathcal{T}_n$ by the following simpler algorithm: *Make a vertical or horizontal cut, with probability* $1/2$ *each, and continue recursively in each half for* $n-1$ *iterations (with all choices independent).* This method, however, does *not* give a uniformly distributed tiling when $n \geq 2$. For example, the probability of obtaining the all horizontal tiling $\widetilde{T}_0$ is $2^{-(2^n-1)} \ll A_n^{-1}$. Moreover, a branching process argument similar to the one in Section 6.2, shows that for the random tiling generated by this procedure, $\mathbf{P}$(there is a vertical cut) $\to 1$ as $n \to \infty$, in contrast to (4.1). This simpler method is equivalent to choosing a random labeling of the complete binary tree with H and V (or A and D) uniformly among all 2^{2^n-1} possibilities without any restrictions, and constructing the corresponding tiling as in Section 3.

5 Dynamic sampling algorithms

An alternative method for sampling AD-trees and dyadic tilings is by simulating suitable Markov chains. A simple Markov chain on the state space $\mathcal{T}_n$ starts from any tiling T_0. At each step it chooses a random dyadic rectangle R of any size larger than 2^{-n} and checks whether R has a nontrivial intersection with any of the tiles in the current tiling. If not, then it rotates the part of the tiling that falls within R by $90°$ in one of the two directions; otherwise it does nothing. When R has area 2^{-n+1}, this rotation changes exactly two tiles and is very similar to a Markov chain previously studied in the context of domino tilings of the chessboard [12, 11, 13].

It will be useful to interpret tilings in terms of AD-trees. We first define a second, very simple Markov chain on AD-trees and show this is rapidly mixing. Of course this immediately defines a Markov chain on the set of dyadic tilings, but this chain is less natural in this context. Hence, we conclude this section by comparing the mixing rates of the two Markov chains on tilings to show that they both define efficient sampling algorithms.

5.1 A Markov chain on AD-trees

The Markov chain on AD-trees $\mathcal{T}_n^{AD}$ successively changes the labeling at a single node of the tree, while avoiding badly labeled subtrees. If $x \in \mathcal{T}_n^{AD}$ is a labeled tree, then let $x(v)$ be the label which x assigns to vertex v.

The Markov chain $\widetilde{\mathcal{M}}_n$ starts at a fixed starting point x_0, say the tree such that $x_0(v) = D$ for all nodes v. Given that $\widetilde{\mathcal{M}}_n$ is at state x_t at time t, it moves to state x_{t+1} as follows: Let V_n be the set of the $2^n - 1$ vertices of the tree. Pick $(v, b) \in V_n \times \{A, D\}$ uniformly. For all $w \neq v$, $x_{t+1}(w) = x_t(w)$. Next, set $x_{t+1}(v) = b$ if it leads to a valid configuration (i.e., if it does not create a badly labelled tree). If changing the label at v creates a badly labeled subtree, then we reject the move and remain at the current configuration, so $x_{t+1}(w) = x_t(w)$ for all w. The transition probabilities $\widetilde{P}(\cdot,\cdot)$ of $\widetilde{\mathcal{M}}_n$ are

$$\widetilde{P}_n(x,y) = \begin{cases} \frac{1}{2|V_n|} & \text{if there is a unique node } v \text{ such that} \\ & \quad x(v) \neq y(v); \\ 1 - \sum_{y' \neq x} \widetilde{P}_n(x,y') & \text{if } x = y; \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.6 shows that the state space is connected. We prove a stronger lemma which will be useful later. Given any two configurations $x, y \in \mathcal{T}_n^{AD}$, let us define the distance $\Phi(x,y)$ to be the Hamming distance between them. That is, $\Phi(x,y)$ is the number of vertices which are assigned different labels.

Lemma 5.1. *Let $x, y \in \mathcal{T}_n^{AD}$ be any two configurations. Then there is a sequence of states $z_0, z_1, \ldots, z_d$ such that $z_0 = x$, $z_d = y$, $d = \Phi(x,y)$ and for all $0 \leq i < d$, $\Phi(z_i, z_{i+1}) = 1$.*

Proof. We consider two cases. First, suppose that there exists at least one vertex that is labeled D in x and A in y. Let w be any such vertex which is as far from the root as possible. Define z_1 by changing the label of x at w, i.e., $z_1(w) = A$ and $z_1(v) = x(v)$, for all $v \neq w$. To see that z_1 is a valid labeling, notice that if the children of w in z_1 are both labeled D, then they must also be labeled D in x, and hence, by our choice of w, they are labeled D in y. This contradicts our assumption that y is a valid AD-tree. Consequently, $\Phi(x, z_1) = 1$ and $\Phi(z_1, y) = d - 1$, and the result follows by induction on d.

If all of the vertices which have different labels in x and y are labeled A in x, we interchange the roles of x and y in the argument. □

Corollary 5.2. *The Markov chain $\widetilde{\mathcal{M}}_n$ is ergodic and converges to the uniform distribution on $\mathcal{T}_n^{AD}$.* □

5.2 Bounding the mixing rate of $\widetilde{\mathcal{M}}_n$

To bound the mixing rate, or convergence time, of our Markov chain, we will use a simple path coupling argument.

Starting in any given initial state x, we measure the deviation of the distribution $P^t(x, \cdot)$ at time t from the uniform distribution π by the *variation distance*:

$$\Delta_x(t) = \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)|.$$

The *mixing time* of the Markov chain is defined as

$$\tau(\epsilon) = \max_x \min\{t : \ \Delta_x(t') \leq \epsilon \text{ for all } t' \geq t\}.$$

If $\tau(\epsilon)$ is polylogarithmic in the size of Ω, for fixed ϵ, then we say that the Markov chain is *rapidly mixing*. Recall that our state space has size which is doubly exponential in n (see equation (1.2)), so this means that $\tau(\epsilon)$ is exponential in n. All of our algorithms must have mixing time which is at least 2^n, the time it takes to write down a single configuration.

One strategy for bounding $\tau(\epsilon)$ is to construct a *coupling* for the Markov chain, i.e., a stochastic process $(x_t, y_t)_{t=0}^{\infty}$ on $\Omega \times \Omega$ such that each of the processes x_t and y_t is a faithful copy of $\mathcal{M}$ (given initial states $x_0 = x$ and $y_0 = y$), and if $x_t = y_t$, then $x_{t+1} = y_{t+1}$.

The expected time taken for the processes to meet provides a good bound on the mixing time of $\mathcal{M}$. To state this formally, for initial states x, y set

$$T_{x,y} = \min\{t : x_t = y_t \mid x_0 = x,\ y_0 = y\},$$

and define the *coupling time* to be $T = \max_{x,y} \mathbf{E}\, T_{x,y}$. The following result relates the mixing time to the coupling time (see, e.g., [1]):

Theorem 5.3. $\tau(\epsilon) \leq Te\lceil \ln \epsilon^{-1} \rceil$. □

The method of path coupling simplifies our goal by letting us bound the mixing rate of a Markov chain by considering only a small subset of $\Omega \times \Omega$. (See [6, 9].) We use the following theorem, obtained by combining Theorems 2.1 and 2.2 in Dyer and Greenhill [9].

Theorem 5.4. *Let Φ be an integer valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \ldots, B\}$. Let U be a subset of $\Omega \times \Omega$ such that for all $(x, y) \in \Omega \times \Omega$ there exists a path $x = z_0, z_1, \ldots, z_r = y$ between x and y such that $(z_i, z_{i+1}) \in U$ for $0 \leq i < r$ and*

$$\sum_{i=0}^{r-1} \Phi(z_i, z_{i+1}) = \Phi(x, y).$$

Let $\mathcal{M}$ be a Markov chain on Ω with transition matrix P. Consider any random function $f : \Omega \to \Omega$ such that $\mathbf{P}[f(x) = y] = P(x, y)$ for all $x, y \in \Omega$, and define a coupling of the Markov chain by $(x_t, y_t) \to (x_{t+1}, y_{t+1}) = (f_t(x_t), f_t(y_t))$, where $(f_t)_{t=0}^{\infty}$ are independent copies of f. Let $\Delta\Phi(x_t, y_t) = \Phi(x_{t+1}, y_{t+1}) - \Phi(x_t, y_t)$. Suppose that, conditioned on any pair of states x_t and y_t,

(i). $\mathbf{E}(\Delta\Phi(x_t, y_t)) \leq 0$ *when* $(x_t, y_t) \in U$,

(ii). $\mathbf{P}[\Phi(x_{t+1}, y_{t+1}) \neq \Phi(x_t, y_t)] \geq \alpha$ *when* $x_t \neq y_t$, *for some constant* $\alpha > 0$.

Then the mixing time of $\mathcal{M}$ satisfies

$$\tau(\epsilon) \leq \left\lceil \frac{eB^2}{\alpha} \right\rceil \lceil \ln \epsilon^{-1} \rceil. \qquad \square$$

The random function f thus updates all states of the Markov chain simultaneously. This is known as a *complete coupling.*

To apply this machinery to analyze the Markov chain $\widetilde{\mathcal{M}}_n$, we first need to define the random function f that defines the coupling. We do this by choosing $(v, b) \in V_n \times \{D, A\}$ uniformly and independently and then for any configuration x defining $f(x)$ by changing $x(v)$ to b, if possible. This defines a simultaneous update of all states, and thus a complete coupling with the correct marginal distributions.

Let U be the pairs of configurations (x, y) such that the Hamming distance $\Phi(x, y) = 1$. The following lemma establishes that the expected distance between such a pair is never increasing.

Lemma 5.5. *Let $x_t, y_t \in \mathcal{T}_n^{AD}$ be two configurations such that $\Phi(x_t, y_t) = 1$. Then the expected change in distance $\mathbf{E}[\Delta\Phi(x_t, y_t)] \leq 0$ after one step of the coupled Markov chain.*

Proof. Let $x_t, y_t \in \mathcal{T}_n^{AD}$ such that $\Phi(x_t, y_t) = 1$ and let w be the vertex where they are labeled differently. Assume without loss of generality that w is labeled D in x_t. Suppose we choose $(v, b) \in V_n \times \{A, D\}$ for our next move. We consider the set of moves that can change the distance between the configurations. Let $p(w)$ be the parent of w, $s(w)$ be the sibling of w, and $l(w)$ and $r(w)$ be the left and right children of w.

1. If $v = w$, then $x_{t+1} = y_{t+1}$ for any choice of b.

2. If $v = p(w)$, then there is exactly one way that the distance can increase. Namely, $x_t(p(w)) = D$, $x_t(s(w)) = D$ and $b = A$. (This move would increase the distance between configurations because the label on $x_{t+1}(p(w)) = x_t(p(w))$ but $y_{t+1}(p(w)) \neq y_t(p(w))$.)

3. If $v = s(w)$, then again there is exactly one way for the distance to increase. This can only occur if $x_t(p(w)) = A$, $x_t(s(w)) = A$ and $b = D$.
4. If $v = l(w)$ is the left child of w, then we can increase the distance between configurations only if $x_t(l(w)) = A$, $x_t(r(w)) = D$ and $b = D$, where $r(w)$ is the right child of w.
5. Likewise, if $v = r(w)$, the distance can increase only if $x_t(l(w)) = D$, $x_t(r(w)) = A$ and $b = D$.

Initially it looks as though there are many more possibilities for increasing the distance than decreasing it. However, we are quite fortunate in that not all of these potentially bad events can be present simultaneously. In particular, the bad events described in the second and third cases cannot occur simultaneously, nor can the last two cases. Hence, there are at most two choices of (v, b) which will increase the distance to 2 and exactly two choices of (v, b) which will decrease the distance to 0. All other moves are neutral. Summarizing this discussion, we find that

$$\mathbf{E}[\Delta\Phi(x_t, y_t)] \leq 0,$$

since all of these moves are equally likely. □

This lemma provides the crucial ingredient towards our path coupling argument for bounding the mixing rate of $\widetilde{\mathcal{M}}_n$. Finally, notice that for any two tilings at distance 1 there are always two possible moves which decrease the distance. Hence we can take $\alpha = (2^n - 1)^{-1}$.

Referring to Theorem 5.4 and Lemma 5.1, we find

Corollary 5.6. *The mixing time of the Markov chain $\widetilde{\mathcal{M}}_n$ on labeled trees $\mathcal{T}_n^{AD}$ satisfies $\tau(\epsilon) \leq 2^{3n} e \lceil \ln \epsilon^{-1} \rceil$.* □

5.3 A natural Markov chain on tilings

The Markov chain $\widetilde{\mathcal{M}}_n$ defined on AD-trees can be reinterpreted in terms of tilings; at each step one of $2^n - 1$ possible dyadic rectangles is identified, and if certain conditions are met, the subtiling can be rotated. Our definition of AD-trees restricts both the set of possible rectangles as well as the direction of rotation. For example, if the tiling has both horizontal and vertical cuts (of the unit square), then we may allow the top or bottom half to be rotated in this fashion, but we would not allow the left or right halves to be rotated, nor the whole square.

We rectify this asymmetry by introducing a new Markov chain $\mathcal{M}_n$, referred to at the beginning of this section. Two tilings are connected by a single step of the Markov chain if one can be transformed into the other by rotating the part of the tiling contained in some dyadic rectangle in the square.

Let b_n be the number of dyadic rectangles (regions) in a unit square with area at least 2^{-n+1}. Since there are $(k+1)2^k$ dyadic rectangles with area exactly 2^{-k} (since there are $k+1$ choices for the shape and each shape can appear in exactly 2^k positions), we find

$$b_n = \sum_{k=0}^{n-1} (k+1)2^k = (n-1)2^n + 1.$$

This gives the transition probabilities $P_n(\cdot,\cdot)$ of our Markov chain $\mathcal{M}_n$:

$$P_n(T_1,T_2) = \begin{cases} \frac{1}{4b_n} & \text{if } T_1 \text{ and } T_2 \text{ differ by rotating the subtiling} \\ & \quad \text{in a dyadic subrectangle by } \pm 90^\circ \text{ or } 180^\circ; \\ 1 - \sum_{T' \neq T_1} P_n(T_1,T') & \text{if } T_1 = T_2; \\ 0 & \text{otherwise.} \end{cases}$$

It is not hard to see that this Markov chain connects the state space; starting with any tiling, we can always rotate subtilings starting with large dyadic rectangles and continuing with smaller ones until all the cuts are vertical. (Theorem 2.6 shows a stronger statement.) In addition, all transitions (besides self-loops) have equal probability, so detailed balance implies that the Markov chain converges to the uniform distribution. Summarizing this, we find:

Theorem 5.7. *The Markov chain $\mathcal{M}_n$ is ergodic and converges to the uniform distribution over dyadic tilings $\mathcal{T}_n$.* □

5.4 Showing that $\mathcal{M}_n$ is rapidly mixing

We conclude by showing that $\mathcal{M}_n$ is rapidly mixing by comparing its mixing rate to that of $\widetilde{\mathcal{M}}_n$, using the comparison method of Diaconis and Saloff-Coste [8]. Here we describe a special case of the comparison theorem which is sufficient for our application.

Let P and $\widetilde{P}$ be transition matrices of two reversible Markov chains on state space Ω which have the same stationary distribution π. We would like to express the mixing rate of a Markov chain (P,π,Ω) (for example, $\mathcal{M}_n$, the Markov chain on tilings) in terms of the mixing rate of $(\widetilde{P},\pi,\Omega)$ (for example, $\widetilde{\mathcal{M}}_n$, the rapidly mixing Markov chain on AD-trees).

To apply the comparison method it is necessary to map each transition of $\widetilde{P}$ to a path described by some number of transitions in P. In our application this is trivial since $\widetilde{P}(x,y) \neq 0$ implies that $P(x,y) \neq 0$ for every $x,y \in \Omega$. Hence we can use the identity map and all of our paths have length 1. Using the formulation of the comparison method as given in [13, Proposition 4] (slightly modified here), we have the following theorem.

Theorem 5.8. *Let (P,π,Ω) and $(\widetilde{P},\pi,\Omega)$ be two reversible Markov chains such that $\widetilde{P}(x,y) \neq 0$ implies $P(x,y) \neq 0$ for all $x,y \in \Omega$. Let $\pi_* = \min_{x\in\Omega} \pi(x)$. Then, for $0 < \epsilon < 1/2$,*

$$\tau(\epsilon) \leq 4\ln(1/(\epsilon\pi_*))A \max\Big(\frac{\widetilde{\tau}(\epsilon)}{\ln(1/(2\epsilon))}, 1\Big), \tag{5.1}$$

where

$$A = \max_{x\neq y, \widetilde{P}(x,y)>0} \frac{\widetilde{P}(x,y)}{P(x,y)}.$$ □

To apply this to our Markov chains $\mathcal{M}_n$ and $\widetilde{\mathcal{M}}_n$, consider any pair of distinct states $x, y \in \Omega$ such that $\widetilde{P}(x,y) > 0$. We find

$$\begin{aligned}\frac{\widetilde{P}(x,y)}{P(x,y)} &\le \frac{(2|V_n|)^{-1}}{(4b_n)^{-1}} \\ &= \frac{4\big((n-1)\,2^n+1\big)}{2(2^n-1)} \\ &\le 2n.\end{aligned}$$

In addition, the stationary probability π is uniform over dyadic tilings, so $\pi_*^{-1} \le 2^{2^n}$. Applying these bounds, Theorem 5.8 gives

$$\tau(\epsilon) \le c(\epsilon) n\, 2^n\, \widetilde{\tau}(\epsilon),$$

for some constant $c(\epsilon)$. Hence, by Corollary 5.6,

Corollary 5.9. *The Markov chain $\mathcal{M}_n$ on dyadic tilings $\mathcal{T}_n$ is rapidly mixing.* □

Problem 5.10. It is also natural to consider the analogous Markov chain where we only rotate rectangles of area 2^{-n+1}. This is the same as a random walk on $\mathcal{T}_n$ regarded as an undirected graph as in Section 2. Is this Markov chain also rapidly mixing?

6 Random dyadic tilings

We now turn our attention to limiting properties of random tilings such as the expected height of a tiling and the likelihood of long, thin rectangles. As in Section 4.2, we shall make use of the partition of tilings into types according to whether there are vertical or horizontal cuts. We also use the recursive random construction from Section 4.3.

6.1 Total height

Recall the definition of the total height of a tiling from Section 2. We will here study the normalized height function defined by

$$\tilde{H}(T) = 2^{-n}H(T) - n/2, \qquad T \in \mathcal{T}_n. \tag{6.1}$$

This, combined with equation (2.1), tells us that $-n/2 \le \tilde{H}(T) \le n/2$. We let H_n and $\tilde{H}_n$ denote the random variables $H(T)$ and $\tilde{H}(T)$ obtained by choosing a random tiling $T \in \mathcal{T}_n$.

By symmetry (a 90° rotation transforms $\tilde{H}(T)$ to $-\tilde{H}(T)$), $\tilde{H}_n$ is a symmetric random variable. In particular, $\mathbf{E}\,\tilde{H}_n = 0$.

Theorem 6.1. *There exists a symmetric random variable $\tilde{H}_\infty$ such that*

(i). *As $n \to \infty$, $\tilde{H}_n \xrightarrow{d} \tilde{H}_\infty$, with convergence of all moments.*

(ii). *For any real t,*

$$\mathbf{E}\exp(t\tilde{H}_n) \le \exp(\tfrac{1}{4}\phi^4 t^2), \qquad 1 \le n \le \infty.$$

(iii). *For any $a \geq 0$,*

$$\mathbf{P}(\tilde{H}_n \geq a) \leq \exp(-\phi^{-4}a^2), \qquad 1 \leq n \leq \infty.$$

(iv). *The moment generating function $\psi(z) = \mathbf{E}\, e^{z\tilde{H}_\infty}$ is an entire function satisfying the functional equation*

$$\psi(z) = (\sqrt{5}-1)\cosh(z/2)\big(\psi(z/2)\big)^2 - (\sqrt{5}-2)\big(\psi(z/4)\big)^4. \tag{6.2}$$

(v). $\mathbf{Var}\, \tilde{H}_\infty = \mathbf{E}\, \tilde{H}_\infty^2 = (6\phi-2)/11 = (3\sqrt{5}+1)/11 = 0.7007458\cdots$.

Remark 6.2. Higher moments of $\tilde{H}_\infty$ may be found recursively by differentiation of (6.2). In particular, $\mathbf{E}\, \tilde{H}_\infty^4 = (71230+7902\sqrt{5})/80465 = 1.10482\cdots$. (All odd moments vanish by symmetry.)

Remark 6.3. This theorem justifies the definition of the normalization $\tilde{H}_n$, which at first might appear odd. There are 2^n rectangles, each with height in $\{0,\dots,n\}$, so that if the heights were independent the variance would be at most $n^2 2^n$. Theorem 6.1 shows, in contrast, that $\mathbf{Var}\, H_n = 2^{2n}\,\mathbf{Var}\, \tilde{H}_n \sim c2^{2n}$, with $c = \mathbf{Var}\, \tilde{H}_\infty = (3\sqrt{5}+1)/11$, which indicates a very high correlation. Roughly speaking, if an early cut creates long thin rectangles then all of its subrectangles will tend also to be long and thin.

Proof. We use the bijection with HV-trees, and define $H(T)$ for an HV-tree $T \in \mathcal{T}_n^{HV}$ to be the total height of the corresponding tiling. It is easily seen (by induction) that each of the 2^{n-1} paths in the tree from the root to a leaf corresponds to two congruent tiles in the tiling, whose height equals the number of labels V in the path. Let $v_k(T)$ denote the number of nodes at level k in the tree T that are labeled V. Since each node at level k lies on the path to 2^{n-k} leaves, we obtain by summing over all paths

$$H(T) = 2\sum_{x \text{ leaf}} (\text{number of } V \text{ on the path to } x) = 2\sum_{k=1}^{n} 2^{n-k} v_k(T) \tag{6.3}$$

and thus

$$\tilde{H}(T) = \sum_{k=1}^{n} (2^{1-k} v_k(T) - \tfrac{1}{2}). \tag{6.4}$$

We label the HV-tree T by types as in Section 4.2 and define the *k-type* $\mathrm{type}_k(T)$ to be the subtree of nodes up to height k, labeled with their types in T (for $k = 1,\dots,n$). Thus $\mathrm{type}_1(T)$ is the root and its type, or equivalently just the type of the root, which we already have called the type of the tree, i.e., $\mathrm{type}_1(T) = \mathrm{type}(T)$.

For each n we define a martingale $X_0^{(n)}, X_1^{(n)}, \dots, X_n^{(n)} = \tilde{H}_n$ by setting $X_0^{(n)} = \mathbf{E}\, \tilde{H}_n = 0$ and $X_k^{(n)} = \mathbf{E}(\tilde{H}_n \mid \mathrm{type}_k)$ for $k \geq 1$. In other words, $X_k^{(n)}(T)$ is defined as the average of $\tilde{H}_n(T')$ over all HV-trees T' having the same k-type as T.

Let us first consider $X_1^{(n)} = \mathbf{E}(\tilde{H}_n \mid \mathrm{type})$. If $T \in \mathcal{T}_n^{HV}$ is of type V, and the two subtrees of the root are $T_1, T_2 \in \mathcal{T}_{n-1}^{HV}$, then $H(T) = H(T_1) + H(T_2) + 2^n$ by (6.3), or directly by considering the corresponding tilings. Hence

$$\tilde{H}(T) = \tfrac{1}{2}\big(\tilde{H}(T_1) + \tilde{H}(T_2)\big) + \tfrac{1}{2}, \qquad \text{if } \ \mathrm{type}(T) = V. \tag{6.5}$$

Similarly, if T has type H_{HH}, H_{HV} or H_{VH}, then $H(T) = H(T_1) + H(T_2)$ and

$$\tilde{H}(T) = \tfrac{1}{2}\big(\tilde{H}(T_1) + \tilde{H}(T_2)\big) - \tfrac{1}{2}, \qquad \text{if} \quad \text{type}(T) \neq V. \tag{6.6}$$

As discussed in Section 4.2, for trees with $\text{type}(T) = V$, T_1 and T_2 may be any trees in $\mathcal{T}_{n-1}^{HV}$, and thus

$$\mathbf{E}(\tilde{H}_n \mid \text{type} = V) = \tfrac{1}{2}\,\mathbf{E}\,\tilde{H}_{n-1} + \tfrac{1}{2}\,\mathbf{E}\,\tilde{H}_{n-1} + \tfrac{1}{2} = \tfrac{1}{2}, \qquad n \geq 1. \tag{6.7}$$

Combining this with

$$\mathbf{E}(\tilde{H}_n \mid \text{type} = V)\,\mathbf{P}(\text{type} = V) + \mathbf{E}(\tilde{H}_n \mid \text{type} \neq V)\,\mathbf{P}(\text{type} \neq V) = \mathbf{E}\,\tilde{H}_n = 0$$

and $\mathbf{P}(\text{type} = V) = p_n$ by (4.1), we find

$$\mathbf{E}(\tilde{H}_n \mid \text{type} \neq V) = -\frac{p_n}{2(1 - p_n)}. \tag{6.8}$$

Similarly, it follows from (6.6) and the constraints (4.4), using (6.7) and (6.8), that

$$\begin{aligned} \mathbf{E}(\tilde{H}_n \mid \text{type} = H_{HH}) &= \mathbf{E}(\tilde{H}_{n-1} \mid \text{type} \neq V) - \tfrac{1}{2} \\ &= -\frac{p_{n-1}}{2(1 - p_{n-1})} - \frac{1}{2} = -\frac{1}{2(1 - p_{n-1})}. \end{aligned} \tag{6.9}$$

and

$$\begin{aligned} \mathbf{E}(\tilde{H}_n \mid \text{type} = H_{HV}) &= \mathbf{E}(\tilde{H}_n \mid \text{type} = H_{VH}) \\ &= \tfrac{1}{2}\,\mathbf{E}(\tilde{H}_{n-1} \mid \text{type} = V) + \tfrac{1}{2}\,\mathbf{E}(\tilde{H}_{n-1} \mid \text{type} \neq V) - \tfrac{1}{2} \\ &= -\frac{1}{4(1 - p_{n-1})}. \end{aligned} \tag{6.10}$$

Consequently, $X_1^{(n)} - X_0^{(n)} = X_1^{(n)}$ is a random variable taking the three values in (6.7), (6.9) and (6.10) with probabilities $\mathbf{P}(\tau^{(n)} = V) = p_n$, $\mathbf{P}(\tau^{(n)} = H_{HH}) = p_n(1 - p_{n-1})^2$ and $\mathbf{P}(\tau^{(n)} = H_{HV}) + \mathbf{P}(\tau^{(n)} = H_{VH}) = 2p_n p_{n-1}(1 - p_{n-1})$, respectively, where as in Section 4.3 $\tau^{(n)}$ is the type of a random HV-tree in $\mathcal{T}_n^{HV}$ (see (4.5)). In particular,

$$|X_1^{(n)}| \leq \frac{1}{2(1 - p_{n-1})} \leq \frac{\phi^2}{2}. \tag{6.11}$$

For future use we define $Y^{(n)} = X_1^{(n)}$, and let further $Y_H^{(n)}$ and $Y_V^{(n)}$ denote the random variables obtained by conditioning $Y^{(n)}$ on type $\neq V$ and type $= V$, respectively. (Thus, by (6.7), $Y_V^{(n)} = \frac{1}{2}$ really is non-random.) Moreover, define $\bar{Y}^{(n)} = Y^{(n)} - \mathbf{E}\,Y^{(n)} = Y^{(n)}$, $\bar{Y}_H^{(n)} = Y_H^{(n)} - \mathbf{E}\,Y_H^{(n)}$ and $\bar{Y}_V^{(n)} = Y_V^{(n)} - \mathbf{E}\,Y_V^{(n)} = 0$. By (6.11), a similar calculation for $\bar{Y}_H^{(n)}$ and trivially for $\bar{Y}_V^{(n)}$, we have the common bound

$$|\bar{Y}^{(n)}|,\ |\bar{Y}_H^{(n)}|,\ |\bar{Y}_V^{(n)}| \leq \frac{\phi^2}{2}. \tag{6.12}$$

We proceed to studying the martingale differences $X_{k+1}^{(n)} - X_k^{(n)}$ for higher k. Let $T \in \mathcal{T}_n^{HV}$ and $1 \le k < n$. There are 2^{k-1} nodes on level k in T, and each of them carries two subtrees of height $n-k$. Denoting these 2^k subtrees by $T_1, \dots, T_{2^k} \in \mathcal{T}_{n-k}^{HV}$, we find from (6.4)

$$\tilde{H}(T) = \sum_{i=1}^{k} (2^{1-i} v_i(T) - \tfrac{1}{2}) + 2^{-k} \sum_{j=1}^{2^k} \tilde{H}(T_j). \tag{6.13}$$

Now consider all trees with a given k-type τ. The k-type determines $v_1, \dots, v_k$, and it specifies some of the labels on level $k+1$, i.e., some of the root labels of the trees T_j (the ones attached to a node on level k labeled H); say that τ specifies m_H labels H and m_V labels V on level $k+1$, leaving $2^k - m_H - m_V$ unspecified. By the recursive construction in Section 4, the trees T_j can be any trees with the right root labels, and (6.13) yields

$$\begin{aligned}\mathbf{E}\bigl(\tilde{H}(T) \mid \operatorname{type}_k(T) = \tau\bigr) = \sum_{i=1}^{k} (2^{1-i} v_i(T) - \tfrac{1}{2}) + 2^{-k} \bigl(m_H \,\mathbf{E}(\tilde{H}_{n-k} \mid \operatorname{type} \neq V) \\ + m_V \,\mathbf{E}(\tilde{H}_{n-k} \mid \operatorname{type} = V)\bigr), \end{aligned} \tag{6.14}$$

where the conditional expectations on the right hand side are given by (6.7), (6.8).

Now suppose that we extend the k-type τ to a $(k+1)$-type τ' by specifying also the types at level $k+1$. In (6.13), this means that we now have specified the types of $T_1, \dots, T_{2^k}$. If τ' specifies $\operatorname{type}(T_j) = \tau_j$, we thus obtain from (6.13) and (6.14), since the trees T_j otherwise are arbitrary trees in $\mathcal{T}_{n-k}^{HV}$,

$$\begin{aligned}&\mathbf{E}(\tilde{H}(T) \mid \operatorname{type}_{k+1}(T) = \tau') - \mathbf{E}(\tilde{H}(T) \mid \operatorname{type}_k(T) = \tau) \\ &= 2^{-k} \Bigl(\sum_{j=1}^{2^k} \mathbf{E}(\tilde{H}_{n-k} \mid \operatorname{type} = \tau_j) - m_H \,\mathbf{E}(\tilde{H}_{n-k} \mid \operatorname{type} \neq V) - m_V \,\mathbf{E}(\tilde{H}_{n-k} \mid \operatorname{type} = V) \Bigr).\end{aligned}$$

If we use the recursive construction in Section 4.3, then the types τ_j are assigned independently, given τ, and it follows that conditioned on $\operatorname{type}_k(T) = \tau$, we have

$$X_{k+1}^{(n)}(T) - X_k^{(n)}(T) = 2^{-k} \sum_{j=1}^{2^k} \bar{Y}_j, \tag{6.15}$$

where $\bar{Y}_1, \dots, \bar{Y}_{2^k}$ are independent random variables, and each $\bar{Y}_j$ is distributed as one of $\bar{Y}^{(n-k)}$, $\bar{Y}_H^{(n-k)}$ and $\bar{Y}_V^{(n-k)}$. Since every $\bar{Y}_j$ has mean 0 and, by (6.12), variance at most $c = \phi^4/4$, we obtain

$$\mathbf{E}\bigl((X_{k+1}^{(n)} - X_k^{(n)})^2 \mid \operatorname{type}_k = \tau\bigr) = 2^{-2k} \sum_{j=1}^{2^k} \mathbf{E}(\bar{Y}_j)^2 \le c 2^{-k}$$

and thus

$$\mathbf{E}(X_{k+1}^{(n)} - X_k^{(n)})^2 \le c 2^{-k}, \qquad 0 \le k < n. \tag{6.16}$$

Since martingale differences are orthogonal, and $\tilde{H}_n = X_n^{(n)}$, this yields

$$\mathbf{E}(\tilde{H}_n - X_k^{(n)})^2 = \sum_{i=k}^{n-1} \mathbf{E}(X_{i+1}^{(n)} - X_i^{(n)})^2 \le 2c2^{-k}, \qquad 0 \le k < n. \tag{6.17}$$

Next let $k \ge 0$ be fixed, and let $n \to \infty$. Since then $p_n \to \phi^{-1}$, the conditional expectations in (6.7)–(6.10) converge to some limits, and the right hand side in (6.14) converges to some number $X_k(\tau)$. Moreover, the k-type of a random tiling $T \in \mathcal{T}_n^{HV}$ is given by the first k levels of Algorithm 4.1, and thus its distribution converges to the distribution of the random k-type τ_k^∞ generated by Algorithm 4.2, see Remark 4.3. Hence, defining $X_k = X_k(\tau_k^\infty)$, we find that $X_k^{(n)} \xrightarrow{d} X_k$ as $n \to \infty$, for every fixed $k \ge 0$.

Moreover, we may generate all τ_k^∞, $k = 0, 1, \dots$ together as the k-types of the random infinite HV-tree constructed in Remark 4.3, and then $X_0, X_1, \dots$ becomes a martingale, as is easily seen by construction or by taking the limit of the martingales $\{X_k^{(n)}\}$. Moreover, by letting $n \to \infty$ in (6.16), we see that

$$\mathbf{E}(X_{k+1} - X_k)^2 \le c2^{-k}, \qquad k \ge 0,$$

and hence the martingale $\{X_k\}$ is L^2-bounded, whence it converges in L^2; thus there exists a random variable $\tilde{H}_\infty$ such that $\mathbf{E}(\tilde{H}_\infty - X_k)^2 \to 0$ as $k \to \infty$. This, the convergence $X_k^{(n)} \xrightarrow{d} X_k$ for every fixed k, and the uniform bound (6.17), where the bound $2c2^{-k}$ tends to 0 as $k \to \infty$, implies that $\tilde{H}_n \xrightarrow{d} \tilde{H}_\infty$ by a standard 3ϵ argument [5, Theorem 4.2].

This proves the first claim in (i). Convergence of all moments follows from this when we have shown the uniform bound (ii).

For (ii), we return to the representation (6.15) for given n, k and τ. Using again $\mathbf{E}\,\bar{Y}_j = 0$ and $|\bar{Y}_j| \le \phi^2/2$, it follows as in e.g. [2, proof of Theorem A.1.16] that

$$\mathbf{E}\exp(t\bar{Y}_j) \le \cosh(\tfrac{1}{2}\phi^2 t) \le \exp(\tfrac{1}{8}\phi^4 t^2), \qquad j = 1, \dots, 2^k,$$

and thus

$$\begin{aligned}\mathbf{E}\big(\exp(t(X_{k+1}^{(n)} - X_k^{(n)})) \mid \mathrm{type}_k = \tau\big) \\ &= \mathbf{E}\exp\Big(t2^{-k}\sum_{j=1}^{2^k}\bar{Y}_j\Big) = \prod_{j=1}^{2^k}\mathbf{E}\exp\big(t2^{-k}\bar{Y}_j\big) \\ &\le \exp\big(2^k\tfrac{1}{8}\phi^4(2^{-k}t)^2\big) = \exp(2^{-k-3}\phi^4 t^2).\end{aligned}$$

Consequently, since type_k determines $X_k^{(n)}$,

$$\begin{aligned}\mathbf{E}\exp\big(tX_{k+1}^{(n)}\big) &= \mathbf{E}\big(\mathbf{E}\big(\exp(t(X_{k+1}^{(n)} - X_k^{(n)})) \mid \mathrm{type}_k\big)\exp(tX_k^{(n)})\big) \\ &\le \exp(2^{-k-3}\phi^4 t^2)\exp\big(tX_k^{(n)}\big),\end{aligned}$$

and thus, recalling $\tilde{H}_n = X_n^{(n)}$ and $X_0^{(n)} = 0$,

$$\mathbf{E}\exp(t\tilde{H}_n) \le \prod_{k=0}^{n-1} \exp(2^{-k-3}\phi^4 t^2) \le \exp(2^{-2}\phi^4 t^2).$$

This is (ii) for finite n. The estimate for $n = \infty$ follows by taking the limit as $n \to \infty$ (or by the same argument).

(iii) follows from (ii) by a standard application of Markov's inequality. (See, e.g., [2, Appendix A].)

For (iv), we first observe that (ii) implies that $\mathbf{E}\, e^{z\tilde{H}_\infty}$ exists for every complex z and defines an entire function, and further that $\mathbf{E}\, e^{z\tilde{H}_n} \to \mathbf{E}\, e^{z\tilde{H}_\infty}$ as $n \to \infty$.

In order to show the functional equation (6.2), we return to the argument used to show (1.1). Consider a random tiling in $\mathcal{T}_n^{HV}$ and let $\mathcal{V}$ and $\mathcal{H}$ denote the events that there is a vertical or horizontal cut, respectively, and let $\mathbf{1}_\mathcal{V}$ and $\mathbf{1}_\mathcal{H}$ denote the corresponding indicator functions. Then, since there is at least one cut by Theorem 1.1,

$$\begin{aligned}\mathbf{E}\, e^{z\tilde{H}_n} &= \mathbf{E}(e^{z\tilde{H}_n}\mathbf{1}_\mathcal{V}) + \mathbf{E}(e^{z\tilde{H}_n}\mathbf{1}_\mathcal{H}) - \mathbf{E}(e^{z\tilde{H}_n}\mathbf{1}_{\mathcal{V}\cap\mathcal{H}}) \\ &= \mathbf{E}(e^{z\tilde{H}_n} \mid \mathcal{V})\,\mathbf{P}(\mathcal{V}) + \mathbf{E}(e^{z\tilde{H}_n} \mid \mathcal{H})\,\mathbf{P}(\mathcal{H}) - \mathbf{E}(e^{z\tilde{H}_n} \mid \mathcal{V}\cap\mathcal{H})\,\mathbf{P}(\mathcal{V}\cap\mathcal{H}). \quad (6.18)\end{aligned}$$

Moreover, using (6.5),

$$\mathbf{E}(e^{z\tilde{H}_n} \mid \mathcal{V}) = e^{z/2}\,\mathbf{E}(e^{\frac{z}{2}\tilde{H}_{n-1}})\,\mathbf{E}(e^{\frac{z}{2}\tilde{H}_{n-1}})$$

and similarly, or by symmetry,

$$\mathbf{E}(e^{z\tilde{H}_n} \mid \mathcal{H}) = e^{-z/2}\,\mathbf{E}(e^{\frac{z}{2}\tilde{H}_{n-1}})\,\mathbf{E}(e^{\frac{z}{2}\tilde{H}_{n-1}}).$$

Furthermore, if $T \in \mathcal{T}_n^{HV}$ is a tiling with both vertical and horizontal cuts, it is composed of four (arbitrary) tilings $T_1, T_2, T_3, T_4 \in \mathcal{T}_{n-2}^{HV}$, with $\tilde{H}(T) = \frac{1}{4}\sum_1^4 \tilde{H}(T_i)$, which leads to

$$\mathbf{E}(e^{z\tilde{H}_n} \mid \mathcal{V}\cap\mathcal{H}) = \left(\mathbf{E}(e^{\frac{z}{4}\tilde{H}_{n-2}})\right)^4.$$

Since $\mathbf{P}(\mathcal{V}) = \mathbf{P}(\mathcal{H}) = p_n \to \phi - 1 = (\sqrt{5}-1)/2$ and thus $\mathbf{P}(\mathcal{V}\cap\mathcal{H}) = 2p_n - 1 \to 2\phi - 3 = \sqrt{5} - 2$, (6.2) now follows by letting $n \to \infty$ in (6.18).

Finally, differentiating (6.2) twice at $z = 0$ yields, with $\sigma^2 = \mathbf{Var}\,\tilde{H}_\infty$ and using $\mathbf{E}\,\tilde{H}_\infty = 0$,

$$\sigma^2 = 2(\phi-1)\tfrac{1}{4}(1+2\sigma^2) - (2\phi-3)\tfrac{4}{16}\sigma^2,$$

which yields (v) after elementary calculations. □

6.2 Spanning rectangles

We call a subrectangle of the unit square is a *strut* if it spans the unit square vertically. (Hence a dyadic rectangle of area 2^{-n} is a strut if its height as defined in Section 2 is n.) We will study the distribution of $S_n(T)$, the number of struts in a random tiling T in $\mathcal{T}_n$.

We begin by observing that T has a horizontal cut if and only if there are no struts, i.e., if $S_n(T) = 0$. Hence, by (4.1),

$$\mathbf{P}(S_n = 0) = p_n \to \phi - 1.$$

To proceed, we again use HV-trees. As remarked in the proof of Theorem 6.1, a path from the root to a leaf in an HV-tree defines two congruent tiles in the corresponding tiling, and these tiles are struts if and only if all nodes on the path are labeled V. Thus S_n equals two times the number of such paths in a random HV-tree.

Theorem 6.4. *$S_n/(\sqrt{5}-1)^n \xrightarrow{d} Z$ as $n \to \infty$, for some random variable Z such that:*

(i). $\mathbf{P}(Z = 0) = \lim_{n\to\infty} \mathbf{P}(S_n = 0) = \phi - 1$.

(ii). $\mathbf{E}\, Z = \beta$ *and* $\mathbf{Var}\, Z = 2\phi\beta^2$, *where* $\beta = \prod_{n=1}^{\infty}(p_n\phi) = 0.702845\cdots$.

(iii). *Besides the point mass at* 0, *Z has an absolutely continuous distribution on* $(0,\infty)$, *with a continuous and strictly positive density.*

Proof. For an HV-tree $T \in \mathcal{T}_n^{HV}$ and $1 \le k \le n$, let $X_k^{(n)}(T)$ be two times the number of paths from the root of T to a node of height k such that the k nodes on the path all are labeled V. Thus $X_n^{(n)} = S_n$. Further, let $X_0^{(n)} = 1$.

It follows from the recursive construction in Section 4.3 that $X_{k+1}^{(n)}$ has the distribution of a sum of $X_k^{(n)}$ independent variables $Y_j^{(n-k)}$ (for $1 \le j \le X_k^{(n)}$), where $Y_j^{(m)}$ has the distribution

$$\begin{aligned} \mathbf{P}(Y_j^{(m)} = 2) &= p_m, \\ \mathbf{P}(Y_j^{(m)} = 0) &= 1 - p_m. \end{aligned} \tag{6.19}$$

In other words, the random sequence $X_0^{(n)}, \dots, X_n^{(n)}$ is an inhomogeneous branching process where the kth generation has the offspring distribution given by $Y_j^{(n-k)}$ in (6.19). It follows that

$$\mathbf{E}(X_{k+1}^{(n)} \mid X_0^{(n)}, \dots, X_k^{(n)}) = (\mathbf{E}\, Y_1^{(n-k)}) X_k^{(n)} = 2p_{n-k} X_k^{(n)};$$

hence, defining $m_k^{(n)} = \mathbf{E}\, X_k^{(n)}$ and $W_k^{(n)} = X_k^{(n)}/m_k^{(n)}$, we see that $W_0^{(n)}, \dots, W_n^{(n)}$ is a martingale and that

$$m_k^{(n)} = \mathbf{E}\, X_k^{(n)} = \prod_{i=0}^{k-1} 2p_{n-i} = \prod_{i=n-k+1}^{n} 2p_i, \qquad 0 \le k \le n. \tag{6.20}$$

Moreover, again by the branching process,

$$\mathbf{E}\big((X_{k+1}^{(n)} - 2p_{n-k}X_k^{(n)})^2 \mid X_k^{(n)}\big) = \mathbf{Var}(Y_1^{(n-k)}) X_k^{(n)} = 4p_{n-k}(1-p_{n-k}) X_k^{(n)}$$

and thus

$$\mathbf{E}(X_{k+1}^{(n)} - 2p_{n-k}X_k^{(n)})^2 = 4p_{n-k}(1-p_{n-k}) m_k^{(n)}.$$

and

$$\mathbf{E}(W_{k+1}^{(n)} - W_k^{(n)})^2 = 4p_{n-k}(1-p_{n-k})m_k^{(n)}/(m_{k+1}^{(n)})^2 = \frac{1-p_{n-k}}{p_{n-k}}(m_k^{(n)})^{-1}.$$

Since $p_1 = 1/2$ and $p_i \geq p_2 = 4/7$ for $i \geq 2$, (6.20) implies $m_k^{(n)} \geq (8/7)^{k-1}$, and thus, for $0 \leq k \leq n$, since martingale differences are orthogonal and $S_n = X_n^{(n)}$,

$$\begin{aligned} \mathbf{E}(S_n/m_n^{(n)} - W_k^{(n)})^2 &= \sum_{i=k}^{n-1} \mathbf{E}(W_{i+1}^{(n)} - W_i^{(n)})^2 \leq \sum_{i=k}^{n-1} (m_i^{(n)})^{-1} \\ &\leq \sum_{i=k}^{n-1} (7/8)^{i-1} \leq 8(7/8)^{k-1}. \end{aligned} \tag{6.21}$$

The variable $Y_j^{(n-k)}$ defined by (6.19) evidently converges in distribution, as $n \to \infty$ with fixed k, to a limit variable Y_j with

$$\begin{aligned} \mathbf{P}(Y_j = 2) &= \phi - 1, \\ \mathbf{P}(Y_j = 0) &= 2 - \phi. \end{aligned} \tag{6.22}$$

Consider the standard (Galton–Watson) branching process $X_0, X_1, \dots$ with $X_0 = 1$ and offspring distribution given by (6.22), and the corresponding expectations $m_k = (\mathbf{E}\, Y_1)^k = (\sqrt{5}-1)^k$ and martingale $W_k = X_k/m_k$. As is well-known [4, Section 1.6] (and easy to prove), this martingale converges, and thus $W_k \to W$ as $k \to \infty$ for some W.

Moreover, it is obvious that for every fixed $k \geq 0$, as $n \to \infty$, we have $X_k^{(n)} \xrightarrow{d} X_k$, $m_k^{(n)} \to m_k$ and thus $W_k^{(n)} \xrightarrow{d} W_k$. Together with the uniform bound (6.21), this implies $S_n/m_n^{(n)} \xrightarrow{d} W$, using again [5, Theorem 4.2]. Furthermore, as $n \to \infty$,

$$\frac{m_n^{(n)}}{(\sqrt{5}-1)^n} = \prod_{i=1}^{n} \frac{2p_i}{2\phi - 2} \to \prod_{i=1}^{\infty} \frac{p_i}{\phi - 1},$$

where the infinite product converges because of (4.3). Denoting this product by β, we thus have $S_n/(\sqrt{5}-1)^n \xrightarrow{d} \beta W$. The proof is completed by using well-known properties of the limit W, see e.g. [4, Th. I.6.2, Cor. I.12.1] and [3, Sec. 3.6]. ☐

We might also consider horizontal struts, which are the tiles with height 0. By symmetry, the same results hold for the number $S_n^{(0)}$ of them. Note that, by Theorem 1.1, a tiling in $\mathcal{T}_n$ (with $n \geq 1$) cannot contain both horizontal and vertical struts, so at least one of the numbers S_n and $S_n^{(0)}$ is always 0.

Problem 6.5. We leave for further study the analysis of the number $S_n^{(h)}$ of rectangles of a given height h for $0 < h < n$. For h fixed, we expect asymptotic distributions similar in nature to those of S_n. The situation is less clear when $h \sim cn$ with $0 < c < 1$. In particular, what is the limiting distribution of the number $S_{2n}^{(n)}$ of squares?

Problem 6.6. Let $h_{\min}(T)$ denote the minimum height of all tiles in the tiling T. Then

$$\mathbf{P}(h_{\min}(T) = 0) = \mathbf{P}(\text{there is a horizontal strut}) = 1 - p_n \to 2 - \phi.$$

Does $h_{\min}$ have an asymptotic distribution (as seems likely)? What is it? In other words, does $\mathbf{P}(h_{\min}(T) = h)$ have a limit as $n \to \infty$ for every fixed $h \geq 1$, and what is the limit?

Acknowledgement

The main results of this paper were discovered during an excursion to the Zaniemyśl Forest during the Random Graphs conference in Poznań, Poland, August 1999.

References

[1] Aldous, D. Random walks on finite groups and rapidly mixing Markov chains. *Séminaire de Probabilités* XVII, 1981/82, Springer Lecture Notes in Mathematics 986, pp. 243–297.

[2] Alon, N. and Spencer, J. *The Probabilistic Method.* 2nd Edition, Wiley, New York, 2000.

[3] Asmussen, S. and Hering, H. *Branching Processes.* Birkhäuser, Basel, 1983.

[4] Athreya, K.B. and Ney, P.E. *Branching processes.* Springer, Berlin, 1972.

[5] Billingsley, P. *Convergence of Probability Measures.* Wiley, New York, 1968.

[6] Bubley, R. and Dyer, M. Path coupling: A technique for proving rapid mixing in Markov chains. *Proc. 38th Annual IEEE Symp. on Foundations of Computer Science* 223–231, 1997.

[7] Coffman, E.G., Lueker, G.S., Spencer, J. and Winkler, P.M. Packing random rectangles. *Probability Theory and Related Fields* **120**:585–599, 2001. SJ

[8] Diaconis, P. and Saloff-Coste, L. Comparison theorems for reversible Markov chains. *Ann. Appl. Probability* **3**:696–730, 1993.

[9] Dyer, M. and Greenhill, C. A more rapidly mixing Markov chain for graph colorings. *Random Structures and Algorithms* **13**:285–317, 1998.

[10] Lagarias, J., Spencer, J. and Vinson, J. Dyadic Equipartitions of the Unit Square. *Discrete Mathematics* (to appear).

[11] Luby, M., Randall, D. and Sinclair, A. Markov chain algorithms for planar lattice structures. *Proc. 36th IEEE Symposium on Foundations of Computer Science (FOCS '95)* 150–159, 1995.

[12] Propp, J. and Wilson, D. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* **9**:223–252, 1996.

[13] Randall, D. and Tetali, P. Analyzing Glauber dynamics by comparison of Markov chains. *J. Mathematical Physics* **41**:1598–1615, 2000.

Svante Janson, Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
E-mail address: `svante@math.uu.se`
Web address: `http://www.math.uu.se/~svante/`

Dana Randall, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0160, USA
E-mail address: `randall@math.gatech.edu`
Web address: `http://www.math.gatech.edu/~randall/`

Joel Spencer, Courant Institute, 251 Mercer St., New York, NY 10012, USA
E-mail address: `spencer@cs.nyu.edu`
Web address: `http://www.cs.nyu.edu/cs/faculty/spencer/`