universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Existence and Uniqueness of the Measure of Maximal Entropy of the Domino Tiling Process on $\mathbb{Z}^2$

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag.rer.nat.)

Verfasser:	Philipp Deutsch
Matrikelnummer:	0348326
Studienkennzahl lt. Studienblatt:	A 405
Studienrichtung lt. Studienblatt:	Mathematik
Betreuer:	Univ.-Prof. Dr. Klaus Schmidt

Wien, im März 2010

Acknowledgments

First and foremost I want to thank my thesis advisor, Prof. Klaus Schmidt, for his encouragement to wander off the beaten track of mathematics and his patience in guiding me through the "jungle" of mathematical research.

I also wish to express my gratitude towards my parents who made it possible for me to pursue my interests and encouraged me through all those years.

Lastly I would like to thank all of those who supported me during my studies and the work on this thesis.

Philipp Deutsch

Contents

CHAPTER 1

Introduction

Since the beginning of the twentieth century the notion of randomness and information has been studied extensively in a mathematical context. In his paper from 1948, *A Mathematical Theory of Communication* [**29**], Claude Shannon used the term *entropy*, which was a concept known from statistical mechanics, and founded the field of mathematics now called information theory.

In physics, where it was first introduced, entropy is a term that is often loosely described as a measure of randomness, disorder, or uncertainty. Since Shannon's entropy takes the same form mathematically (both being logarithms of a probability measure), it made sense to use the same name.

Most famously entropy appears in the Second Law of Thermodynamics, which states that the entropy of an isolated system increases until it reaches a maximum value at equilibrium. This idea was later generalized in the *Principle of Maximum Entropy*, which postulates that when dealing with uncertainty, the probability distribution that best represents the current state of knowledge is the one maximizing entropy, given any known information. The investigation of entropy in various contexts grew rapidly, and the notion is now applied in fields reaching from physics to information theory to game theory and economics [**12**], [**11**].

Similar to entropy, domino tilings (or dimer models) were studied by physicists well before mathematicians. Imagine a substance made up of molecules which in turn consist of two atoms each. A simple model of this substance would predict that the molecules should arrange themselves into some sort of regular lattice, fitting together as tightly as possible.

There are, of course, many ways these molecules can be stacked, and each distribution has different entropy, which in turn dictates the thermodynamic properties of this

substance. Physicists are thus interested in the distribution of such dominoes that gives the highest possible entropy.

For mathematicians there are also a number of reasons why dimer models warrant deeper study. As it turns out, domino tiling models are extremely versatile, having analogies and equivalents in as diverse fields as spanning trees, random walks and electrical networks. They are fundamental in algebraic graph theory and combinatorial geometry. Furthermore, even though spanning trees have been used already in the nineteenth century, there is still a large number of interesting problems left to solve.

CHAPTER 2

Dynamical Systems and Entropy

In this chapter we will lay the foundation for the following sections and, ultimately, for the main result of this thesis, Theorem 5.6. Most importantly, we will make the notion of entropy precise and prove some of its properties.

As was mentioned in the Introduction, the term entropy arises in different contexts in mathematics and the natural sciences. We will be concerned with the measure-theoretic and topological entropy as it was first introduced by Shannon in 1948 [**29**] and subsequently developed by A. Kolmogorov in 1958 [**14**].

There is a quickly increasing number of books on the subjects of dynamical systems, ergodic theory and entropy. While we will mostly follow Walters' *An Introduction to Ergodic Theory* [**35**], additional information can be found in Ott [**24**] and Katok and Hasselblatt [**14**]. While Silva [**32**] provides a good, albeit basic, introduction to the subject, Keller [**15**] provides an advanced insight into the connections with statistical mechanics. For the measure-theoretic background we refer to Taylor [**34**], Doob [**8**] and Sheldon [**27**].

1. Ergodic Theory

The word "ergodic" origins from the greek words ἔργον (work) and ὁδός (path). It was introduced by Boltzmann, who studied actions on energy surfaces described by a Hamiltonian as they arise in statistical mechanics. Today, the field of ergodic theory, or measurable dynamics, studies the long-term behavior of abstract dynamical systems. That is, we start out with a set X and a transformation $T : X \to X$ acting on this set. In ergodic theory, X and T are given measurable structure. Sometimes we will also assume a topological space, usually a metric space.

Throughout this chapter we will be concerned with measure spaces:

DEFINITION. A triple $(X, \mathcal{A}, \mu)$ is called a *measure space*, where X is a set, $\mathcal{A}$ is a σ-algebra of subsets of X and μ is a measure on X. It is called a *probability space* if $\mu(X) = 1$. We will throughout denote an arbitrary σ-algebra with $\mathcal{A}$ and the Borel-σ-algebra with $\mathscr{B}$.

The transformation acting on a measure space should be compatible with this measure, in other words:

DEFINITION. A transformation T is called *measurable* if it is one-to-one and onto. In this case T^{-1} is also a transformation. A transformation T is called *measurable* if $T^{-1}A \in \mathcal{A}$ for all $A \in \mathcal{A}$. A transformation T is *measure-preserving* if it is measurable and

$$\mu(T^{-1}(A)) = \mu(A)$$

for every set $A \in \mathcal{A}$. We then say μ is *invariant* under T. The set of all T-invariant measures on X is denoted by $\mathcal{M}(X, T)$.

Assume that there is a set $A \subset X$ such that $x \in A$ iff $T(x) \in A$. Then $T|_A$ and $T|_{A^c}$ would both be independent dynamical systems and could be studied separately. We can think of systems which don't have such sets as indecomposable systems.

DEFINITION. A measure-preserving transformation T is called *ergodic* if for every set with the property $T^{-1}(A) = A$ it follows that either $\mu(A) = 0$ or $\mu(A^c) = 0$.

Another important concept is the notion of recurrence. Formally we have

DEFINITION. A measure-preserving transformation T is *recurrent*, if for every measurable set A with $\mu(A) > 0$ there exists a subset $N \subset A$, $\mu(N) = 0$ such that for every $x \in A \backslash N$ there exists an integer $n \in \mathbb{N}$ with $T^n(x) \in A$.

Intuitively this means that T is recurrent if for every set A with positive measure almost every point of A eventually returns to A.

The notions of ergodicity and recurrence alone are quite strong. Indeed we have the

THEOREM 2.1. *Let $(X, \mathcal{A}, \mu)$ be a σ-finite measure space, $T : X \to X$ be a measure preserving transformation and let A, B be arbitrary sets of positive measure. Then the following are equivalent:*

(1) T is recurrent and ergodic.

(2) $\mu\left(X \setminus \bigcup_{n=1}^{\infty} T^{-n}(A)\ \right) = 0$.

(3) For almost every $x \in X$ there exists an $n \in \mathbb{N}$ such that $T^n(x) \in A$.

(4) There exists an $n \in \mathbb{N}$ such that $T^{-n}(A) \cap B \neq \emptyset$.

(5) There exists an $n \in \mathbb{N}$ such that $\mu(T^{-n}(A) \cap B) = 0$.

(Without proof)

For a proof of this theorem we refer to the references cited at the beginning of this chapter.

Another important concept that we will need is *mixing*:

DEFINITION. Let $(X, \mathcal{A}, \mu, T)$ be a dynamical system. If for all $A, B \in \mathcal{A}$ we have

$$\lim_{n\to\infty} \mu(A \cap T^{-n}B) = \mu(A)\mu(B)$$

the system is called *strong mixing*.

Intuitively the name makes sense. Loosely we can say (interpreting T as the passage of time), that if we wait long enough the proportion of A (under iterations of T) in B approaches the size of A (i.e. the proportion of A in X).

REMARK. As the name suggests, there also exists *weak mixing* and also *topological mixing*. We will not be concerned with these here. For further information see for example Walters [**35**].

This definition can be extended to higher orders of mixing. For example, a *strong 3-mixing system* is a system for which

$$\lim_{m,n\to\infty} \mu(A \cap T^{-m}B \cap T^{-(m+n)}) = \mu(A)\mu(B)\mu C$$

holds for all $A, B, C \in \mathcal{A}$.

2. Entropy

The definition of entropy of a measure-preserving transformation T of a probability space $(X, \mathcal{A}, \mu)$ is done in three stages: the entropy of a finite sub-σ-algebra of $\mathcal{A}$, the entropy of the transformation relative to a finite sub-σ-algebra, and finally the entropy

of T. Some of the definitions involve logarithms – It is customary to use the natural logarithm, since this way the entropy ties in more neatly with some concepts in statistical mechanics.

DEFINITION. A *partition* of $(X, \mathcal{A}, \mu)$ is a disjoint collection of elements of $\mathcal{A}$ whose union is X. We will work only with finite partitions which we will denote by greek letters, for example $\alpha = \{A_1, \ldots, A_k\}$. If $\mathcal{C}$ is a finite sub-σ-algebra of $\mathcal{A}$, then the non-empty sets form a partition of X, denoted by $\xi(\mathcal{C})$.

For two finite partitions of $(X, \mathcal{A}, \mu)$, $\alpha = \{A_1, \ldots, A_k\}$ and $\beta = \{B_1, \ldots, B_l\}$ define their *join* of *refinement* as

$$\alpha \vee \beta := \{A_i \cap B_j : 1 \leq i \leq k, 1 \leq j \leq l\}.$$

It is again a partition of X.

Think of a partition α of a probability space as a list of the possible outcomes of an experiment, where the probability of each outcome A_i is $\mu(A_i)$. As mentioned in the Introduction, we will try to associate with that experiment a number $H_\mu(\alpha)$ that measures the uncertainty about the outcome of the experiment or equivalently the amount of information received when performing the experiment - the entropy.

What properties should this entropy function H_μ have? It should be 0 iff $\mu(A_i) = 1$ for some A_i, since this means that this A_i must occur and there is no uncertainty about the experiment. We would also like H_μ to be continuous and symmetric. When performing two experiments α and β the total information gained should equal the information gained from α plus the information gained from performing β, knowing the outcome of α. Finally the entropy should have its maximum when all the outcomes are equally likely, for then the uncertainty is largest.

It turns out that there is one function that satisfies all those conditions [**17**]:

DEFINITION. Let $\mathcal{C}$ be a finite sub-algebra of $\mathcal{A}$ with $\xi(\mathcal{C}) = \{C_1, \ldots, C_k\}$. The measure-theoretic entropy of $\mathcal{C}$ (or $\xi(\mathcal{C})$) is given by the real number

$$H_\mu(\mathcal{C}) = H_\mu(\xi(\mathcal{C})) = -\sum_{i=1}^{k} \mu(C_i) \log \mu(C_i).$$

We are now ready for the second step in defining a measure-preserving transformation's entropy.

DEFINITION. Let $T : X \to X$ be a measure-preserving transformation acting on a probability space $(X, \mathcal{A}, \mu)$, and let $\mathcal{C}$ be a finite sub-σ-algebra of $\mathcal{A}$. Then

$$h_\mu(T, \xi(\mathcal{C})) = h_\mu(T, \mathcal{C}) = \lim_{n \to \infty} \frac{1}{n} H_\mu \left(\bigvee_{i=0}^{n-1} T^{-i} \mathcal{C} \right)$$

is called the *measure-theoretic entropy of T with respect to $\mathcal{C}$*.

REMARK. The existence of this limit is of course not trivial. It can however be proved quite easily using some results from analysis, as demonstrated in [**35**].

It is easiest to interpret T as the passage of time. If, for example, the application of T means the passage of one day, then $\bigvee_{i=0}^{n-1} T^{-i}\ \mathcal{C}$ can be interpreted as performing the experiment represented by $\mathcal{C}$ on n consecutive days and combining the results. The entropy $h_\mu(T, \mathcal{C})$ is then the average information gained per day when the experiment is performed daily, forever.

DEFINITION. Let $(X, \mathcal{A}, \mu)$ be a probability space and let $T : X \to X$ be a transformation thereon. Then $h_\mu(T) = \sup h_\mu(T, \mathcal{C})$, where the supremum is taken over all finite sub-algebras of $\mathcal{A}$, is called the *measure-theoretic entropy of T*. Equivalently $h_\mu(T) = \sup h_\mu(T, \alpha)$, where the supremum is taken over all finite partitions of $(X, \mathcal{A}, \mu)$.

This value depends on μ. We will now develop another concept, called the topological entropy, which does not depend on the measure of the probability space. As the name implies this assumes a topological structure for X, i.e. we can work with open sets. In the following paragraphs we will be concerned with open covers of X, which we will denote by greek letters.

DEFINITION. Let α and β bet open covers of X. Their *join* is defined as

$$\alpha \vee \beta := \{A \cap B : A \in \alpha, B \in \beta\}.$$

The number $N(\alpha)$ denotes the number of sets in a finite subcover of α with smallest cardinality.

As before we can now define entropy in three steps:

DEFINITION. The *topological entropy of an open cover* α is the number

$$H(\alpha) := \log N(\alpha).$$

If $T : X \to X$ is a continuous map then the *topological entropy of* T *relative to* α is given by

$$h(T, \alpha) := \lim_{n\to\infty} \frac{1}{n} H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right).$$

Finally, the *topological entropy of* T is

$$h(T) := \sup h(T, \alpha),$$

where the supremum is taken over all open covers of X.

The treatment of these two notions of entropy begs the question whether they are linked in some way. As it turns out there is a beautiful and important connection between the measure-theoretic and the topological entropy, the variational principle: It says that the supremum of $h_\mu(T)$ taken over all T-invariant measures, equals the topological entropy of T. We will prove this in next section.

3. The Variational Principle

The inequality $sup\{h_\mu(T) : \mu \in \mathcal{M}(X, T)\} \leq h(T)$ has been known to be true since the 1960s. The equality was then proved in 1970, but the more elegant proof given here was found in 1976 by M. Misiurewicz [**35**], [**23**].

Before proving the principle we need some further definitions and lemmas.

LEMMA 2.2. *The function* $\phi : [0, \infty) \to \mathbb{R}$ *defined by*

$$\phi(x) = \begin{cases} 0, & x = 0 \\ x \log x, & x \neq 0 \end{cases}$$

is strictly convex. Further we have that $\phi\left(\sum_{i=1}^k a_i x_i\right) \leq \sum_{i=1}^k a_i \phi x_i$*, if* $x_i \in [0, \infty), a_i \geq 0, \sum_{i=1}^k a_i = 1$*, and equality holds iff all the* x_i *corresponding to non-zero* a_i *are equal.*

(Without proof)

Let $\mathcal{A}, \mathcal{C}$ be finite sub-σ-algebras of $\mathscr{B}$ and

$$\alpha(\mathcal{A}) = \{A_1, \ldots, A_k\}, \qquad \gamma(\mathcal{C}) = \{C_1, \ldots, C_p\}.$$

DEFINITION. The *entropy of $\mathcal{A}$ given $\mathcal{C}$ is the number*

$$\begin{aligned} H(\alpha(\mathcal{A})|\gamma(\mathcal{C})) = H(\mathcal{A}|\mathcal{C}) &= -\sum_{j=1}^{p} \mu(C_j) \sum_{i=1}^{k} \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \\ &= -\sum_{i,j} \mu(A_i \cap C_j) \log \frac{\mu(A_i \cap C_j)}{\mu(C_j)} \end{aligned}$$

There is another, equivalent, definition of the topological entropy given by Bowen, which makes use of spanning sets.

DEFINITION. For $n \geq 1$ and $\varepsilon > 0$ a finite $S \subset X$ is called a (n, ε)*-separated set* if for any two distinct points $x, y \in S$ there is an $i \in [0, n-1]$ such that $d(T^i x, T^i y) \geq \varepsilon$. The maximal cardinality of a (n, ε)-separated set is denoted by $s_n(n, \varepsilon) \in \mathbb{N}$, and we will write $s_n(\varepsilon, X, T)$ for the largest cardinality of any (n, ε)-separated set with respect to T.

DEFINITION. The topological entropy can be defined through spanning sets according to

$$h(T) = \lim_{\varepsilon \to 0} s(\varepsilon, X, T).$$

This definition is equivalent to the one given in the previous section.

THEOREM 2.3 (Variational Principle). *Let $T : X \to X$ be a continuous map on a compact metric space X equipped with the Borel-σ-algebra $\mathscr{B} = \mathscr{B}(X)$. Then*

(i) *For every $\mu \in \mathcal{M}(X, T)$ $h_\mu(T) \leq h(T)$ and*
(ii) $h(T) = \sup\{h_\mu(T) : \mu \in \mathcal{M}(X, T)\}$.

PROOF.

(i) Let $\alpha = \{A_1, \ldots, A_k\}$ be a finite partition of $(X, \mathcal{A})$ and let $\varepsilon > 0$ so that $\varepsilon < 1/(k \log k)$. Since μ is a Borel-measure it is regular and thus there exist compact sets $B_j \subset A_j$, $1 \leq j \leq k$ such that $\mu(A_j \backslash B_j) < \varepsilon$. Let β be the

partition $\beta = \{B_0, B_1, \ldots, B_k\}$ where $B_0 := X \setminus \bigcup_{j=1}^{k} B_j$. Then $\mu(B_0) < k\varepsilon$ and we have

$$
\begin{aligned}
H_\mu(\alpha|\beta) &= -\sum_{i=0}^{k}\sum_{j=0}^{k} \mu(B_i)\phi\left(\frac{\mu(B_i \cap A_j)}{\mu(B_i)}\right) \\
&= -\mu(B_0)\sum_{j=1}^{k} \phi\left(\frac{\mu(B_0 \cap A_j)}{\mu(B_0)}\right) \\
&\leq \mu(B_0)\log k \\
&< k\varepsilon \log k < 1.
\end{aligned} \tag{2.1}
$$

The second line follows because for all $i \neq 0$ we have that $\mu(B_i \cap A_j)/\mu(B_i) = 0$ or 1. The third line is a consequence of the fact that the function attains its maximum value, $\log k$ iff all k outcomes are equally likely.

The key is that for each $i \neq 0$, $B_0 \cup B_i = X \setminus \bigcup_{j \neq i} B_j$ is an open set and therefore $\tilde{\beta} = \{B_0 \cup B_1, \ldots, B_0 \cup B_k\}$ is an open cover of X. If $n \geq 1$ we have

$$
H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) \leq \log N\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) = \log\left(N\left(\bigvee_{i=0}^{n-1} T^{-i}\tilde{\beta}\right)\cdot 2^n\right) \tag{2.2}
$$

where $N\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right)$ denotes the number of non-empty sets in the partition. Continuing, we get

$$
h_\mu(T, \beta) \leq h(T, \beta) + \log 2 \tag{2.3}
$$

and

$$
h_\mu(T, \alpha) \leq h_\mu(T, \beta) + H_\mu(\alpha|\ \beta) \leq h(T) + \log 2 + 1. \tag{2.4}
$$

This gives a first upper bound $h_\mu(T) \leq h(T) + \log 2 + 1$. It can be shown that for every $n \geq 1$ we have $h_\mu(T^n) = nh_\mu(T)$ (this is known as Abramov's Theorem, see for example [**26**]), and using this we can complete the proof:

$$
\begin{aligned}
h_\mu(T) &= \lim_{n\to\infty} \frac{1}{n} h_\mu(T^n) \\
&\leq \lim_{n\to\infty} \frac{1}{n} h(T^n) + \lim_{n\to\infty} \frac{1}{n}(\log 2 + 1) = h(T).
\end{aligned} \tag{2.5}
$$

■

(ii) We will show that there exists a T-invariant probability measure μ such that $h_\mu(T) \geq s(\varepsilon, X, T)$. First, let $\varepsilon > 0$. Let then S_n be an (n, ε)-separated set of cardinality $s_n(\varepsilon, X)$.

We can now use the atomic measures defined uniformly on the points of those sets (which are not necessarily invariant)

$$\sigma_n := \frac{1}{s(n,\varepsilon)} \sum_{x \in S_n} \delta_x$$

to define a new measure

$$\mu_{n_j} := \frac{1}{n} \sum_{i=0}^{n-1} \sigma_n \circ T^{-i}.$$

It is easy to show that, by replacing $\{n\}$ by a subsequence if necessary, $\mu_n \to \mu \in \mathcal{M}(X, T)$ (see for example [**35**], Thm. 6.9). For notational simplicity we will rename this sequence again $\{n\}$.

Choose now a partition $\alpha = \{A_1, \ldots, A_k\}$ such that $\operatorname{diam}(A_i) < \varepsilon$ and $\mu(\partial A_i) = 0$ for all $i = 1, \ldots, k$. No member of the refinement $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ can contain more than one member of S_n. Thus, $s_n(n, \varepsilon)$ members of $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ each have σ_n-measure $1/s_n(\varepsilon, X)$, the others have σ_n-measure zero and we have

$$H_{\sigma_n}\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) = \log s_n(\varepsilon, X). \tag{2.6}$$

Fix $q, n \in \mathbb{N}$ with $1 < q < n$ and define $a(j) := [(n-j)/g]$ for $0 \leq j \leq q-1$, where $[x]$ denotes the integer part of x. Note that with this notation

$$\{0, 1, \ldots, n-1\} = \{j + rq + i : 0 \leq r \leq a(j) - 1, 0 \leq i \leq q-1\} \cup R, \tag{2.7}$$

where

$$R = \{0, 1, \ldots, j-1, j + a(j)q + j + a(j)q + 1, \ldots, n-1\}.$$

We can therefore write

$$\bigvee_{i=0}^{n-1} T^{-i}\alpha = \bigvee_{r=0}^{a(j)-1} T^{-(rq+j)} \bigvee_{i=0}^{q-1} T^{-i}\alpha \vee \bigvee_{l \in R} T^{-l}\alpha, \tag{2.8}$$

and R has cardinality at most $2q$. Continuing from equation 2.6 we have

$$\log s_n(\varepsilon, X) = H_{\sigma_n}\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \tag{2.9}$$
$$\leq \sum_{r=0}^{a(j)-1} H_{\sigma_n}\left(T^{-(rq+j)}\bigvee_{i=0}^{q-1} T^{-i}\alpha\right) + \sum_{k\in R} H_{\sigma_n}\left(T^{-k}\alpha\right)$$
$$\sum_{r=0}^{a(j)-1} H_{\sigma_n\circ T^{-(rq+j)}}\left(\bigvee_{i=0}^{q-1} T^{-i}\alpha\right) + 2q\log k$$

Summing over all j this inequality becomes

$$q\log s_n(\varepsilon, X) \leq \sum_{p=0}^{n-1} H_{\sigma_n\circ T^{-p}}\left(\bigvee_{i=0}^{q-1} T^{-i}\alpha\right) + 2q^2\log k, \tag{2.10}$$

and dividing by n we get

$$\frac{q}{n}\log s_n(\varepsilon, X) \leq H_{\mu_n}\left(\bigvee_{i=0}^{q-1} T^{-i}\alpha\right) + \frac{2q^2}{n}\log k. \tag{2.11}$$

REMARK. This follows from the property that for $\mu_i \in \mathcal{M}(X,T)$, $p_i \geq 0$, $\sum p_i = 1$ we have $\sum p_i H_{\mu_i}(\alpha) \leq H_{\sum p_i\mu_i}(\alpha)$ for any finite partition α.

Since the A_i have a boundary of μ-measure zero, so do the members of $\bigvee_{i=0}^{q-1} T^{-i}\alpha$ and thus we have $\lim_{j\to\infty}\mu_{n_j}(B) = \mu(B)$ for every $B \in \bigvee_{i=0}^{q-1} T^{-i}\alpha$ and therefore $H_{\mu_{n_j}}\left(\bigvee_{i=0}^{q-1} T^{-i}\alpha\right) \to H_\mu\left(\bigvee_{i=0}^{q-1} T^{-i}\alpha\right)$. Replacing $\{n\}$ by $\{n_j\}$ in equation 2.11 and letting j and q got to ∞ we get $s(\varepsilon, X, T) \leq h_\mu(T,\alpha) \leq h_\mu(T)$. □

CHAPTER 3

Connections between domino tilings and other fields

Burton and Pemantle [**7**] studied domino tilings of $\mathbb{Z}^d$ using a connection between spanning trees and domino tilings. In particular, they calculated the entropy of domino tilings of $\mathbb{Z}^2$ using this relationship. Domino tilings have interesting connections not only to spanning trees, but also to random walks, electrical networks, percolation and many other fields [**20**]. In each of these there are questions that are much easier to analyse after translating the problem into an equivalent one in another field [**16**]. For example, Burton and Pemantle showed that the uniform measure of spanning trees of the $n \times n$ square grid converges as $n \to \infty$ to the unique translation-invariant measure of maximal entropy by considering the dual problem for spanning trees. It follows that the domino tiling process on $\mathbb{Z}^2$ has a unique translation-invariant measure of maximal entropy. We don't know how to prove this directly for the perfect matching itself.

1. Basic definitions

A *graph* is a pair $G = (V, E)$ of sets of *vertices* (or *nodes* or *points*) V and *egdes* (or *lines*) E such that $E \subseteq [V]^2$ i.e. each edge connects two vertices with each other or one vertex with itself, in which case it is called a *self-edge*. If there is no risk of ambiguity we will refer to the graph simply as G. We will write V or $V(G)$ or V_G for the set of vertices and E or $E(G)$ or E_G for the set of edges of G.

A graph is called *planar*, if it can be drawn in the plane $\mathbb{R}^2$ without crossing any edges. Such a drawing partitions the plane into a finite number of regions, called *faces*. The set of all faces of G is $F = F(G) = F_G$. If there is a route along the edges connecting every vertex with every other, the graph is *connected*. A graph is *finite* if it has a finite number of vertices and edges, otherwise it is *infinite*.

A vertex v is *incident* with an edge e if $v \in e$, and e is then an edge *at* v. If x and y are two vertices incident with a common edge e, e *joins* x and y and we can

write $e = xy = yx$. We then say x and y are *ends* of e and they are called *adjacent* or *neighbours* or *connected*, written $x \backsim y$. Two edges are *adjacent* if they share a common end and two faces are *adjacent* if they share a common edge. Pairwise non-adjacent vertices or edges are called *independent*. The *degree* $d(v)$ or d_v of a vertex v is the number of edges incident to v, where self-edges are counted twice.

If the vertices can be divided into two disjoint sets U and W, so that every edge connects a vertex in U with one in W, we say the graph G is *bipartite*. If $|U| = |W|$ then G is a *balanced* bipartite graph. If the edges of a graph are replaced by directed edges, or *arrows*, the graph is called *directed*.

A *subgraph* G' of G is a subset of $G = (V, E)$ that is still a graph, written $G' \subseteq G$. A *forest* on G is a subgraph that does not contain a cycle and a *tree* is a connected forest. A *spanning tree, denoted* T (or *spanning forest*, denoted F) is a tree (or forest) that contains all the vertices of G. Given a subgraph H of G we write T_H (or F_H) for the set of edges of T (or F) contained in H. A planar *essential* spanning forest is a spanning forest of a planar graph G in which every component touches the outer boundary or is infinite, if G itself is infinite. A *directed tree* is tree where one vertex v has been designated the *root* and all the edges are directed away from v.

A set $M \subseteq E$ of independent edges in a graph is called a *matching*, it is *perfect* if it matches all vertices of the graph. A perfect matching is also called a *domino tiling*.

2. Dominoes and trees

Let G be a finite connected planar graph. We will now construct another graph, called the dual graph G^* of G, as follows: For every face in G draw a vertex in G^*. This includes one vertex for the unbounded outer face of G which we will call f^*. Vertices in the dual graph are connected iff the corresponding faces in G are adjacent. The set V_{G^*} is thus identified with F_G and E_{G^*} is identified with E_G. Fig. 3.1 shows a graph G (a) and its dual graph G^* (b). The vertex f^* is shown in extended form, i.e. spread-out instead of as a single dot.

Another graph can be obtained by putting these two together. Drawing the graphs G and G^* on top of each other and adding another vertex wherever lines cross we get the left hand side of fig. 3.2, where the added vertices are drawn in gray. This new graph is

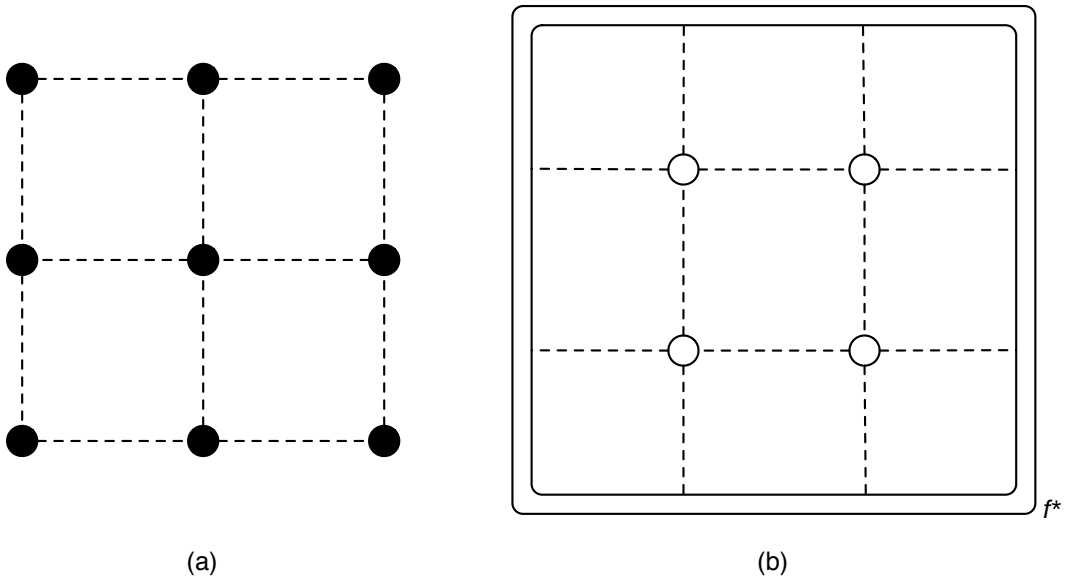

FIGURE 3.1. A graph G and its dual graph G^*

called the induced graph $\tilde{G}$ of G. To avoid confusion, we will say that $\tilde{G}$ has nodes and links whereas G has vertices and edges. There is a node in $\tilde{G}$ for every edge, vertex and face of G. If e is an edge of G we will write $\tilde{e}$ for the corresponding edge-node in $\tilde{G}$ and similarly for the faces and vertices of G.

Finally, let v be a vertex of G that is adjacent to f^*. We can now construct the induced subgraph, or domino graph, $\tilde{G}(v)$ of $\tilde{G}$, which will play an important role later on. It is obtained by deleting the vertices v and f^* from $\tilde{G}$, along with all incident edges, as seen on the right hand side of fig. 3.2.

We want to construct a domino tiling on this new graph $\tilde{G}(v)$. Since every edge-node in $\tilde{G}(v)$ is connected to a node corresponding to a vertex or a face in G, $\tilde{G}(v)$ is bipartite. To see that it is also balanced we need Euler's formula.

THEOREM 3.1 (Eulers formula). *Let G be a finite connected planar graph with vertex set V, edge set E and face set F. Then*

$$|V| - |E| + |F| = 2.$$

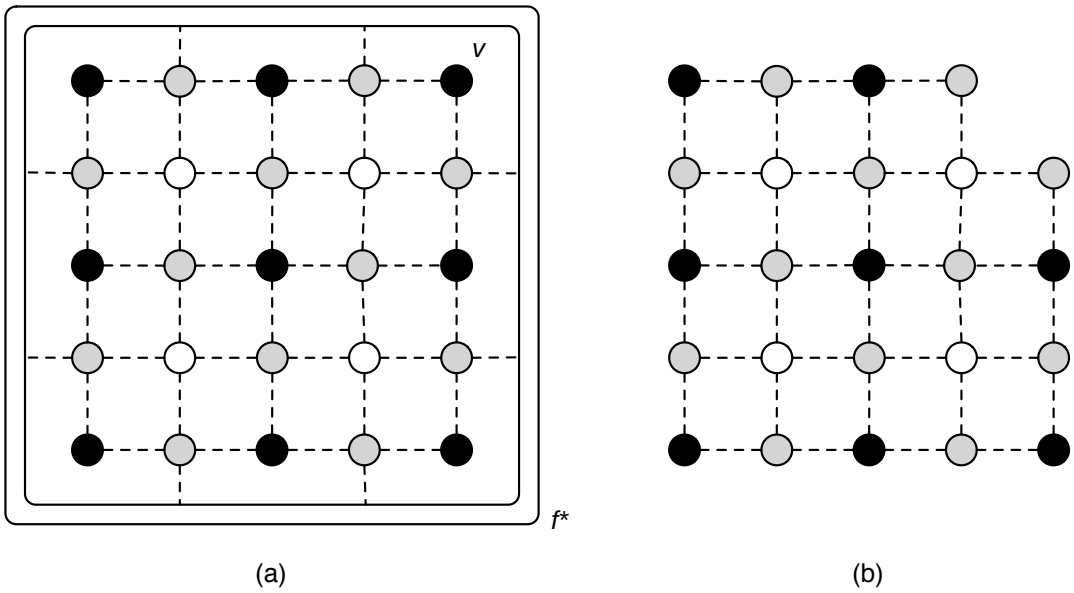

FIGURE 3.2. Construction of $\tilde{G}$ and $\tilde{G}(v)$

PROOF. Begin by choosing a spanning tree T for G and consider the subgraph T^* in the dual graph G^*, where an edge $e^* \in E_{G^*}$ is in T^* iff the corresponding edge in G is not in T. Since T does not contain a cycle, all the faces in G (and therefore all the vertices in G^*) are connected by T^*. But T^* cannot contain any cycles either, because otherwise it would separate some vertices of G from others, which is impossible since T is spanning. Therefore T^* is itself a spanning tree on G^*.

For every tree the number of vertices is one larger than the number of edges. To see this, pick one vertex as the root and direct all the edges of T away from it. There is a bijection between the non-root vertices and the edges by pairing each edge with the vertex it points to. We therefore have $|V| = |E_T| + 1$ and by the same reasoning, $|F| = |E_{T^*}| + 1$. Since an edge is in T^* iff the corresponding edge is not in T we have $|E_T| + |E_{T^*}| = |E|$ so that $|V| + |F| = |E_T| + 1 + |E_{T^*}| + 1 = |E| + 2$. □

REMARK. This relationship between the number of vertices, faces and edges of a planar graph was first mentioned by Euler in 1750, although without proof. The remarkable proof given here, which gets by without induction, was given by Aigner and Ziegler [**1**].

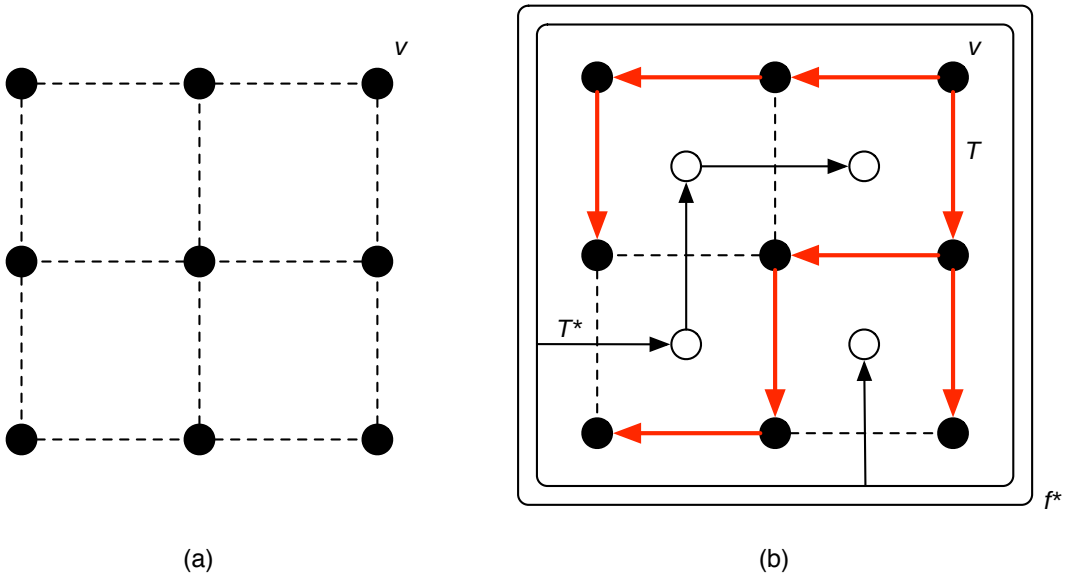

FIGURE 3.3. A planar graph G (a) and dual directed spanning trees on G and G^* (b)

Consider now a finite connected planar graph G and the domino graph $\tilde{G}(v)$. Applying Euler's formula to $\tilde{G}(v)$ we see that $(|V|-1)+(|F|-1)=|E|$, so it is balanced. As mentioned before, it is bipartite: This means we can expect to find perfect matchings on $\tilde{G}(v)$.

There is in fact a one-to-one and onto correspondence between the directed spanning trees rooted at v on G and the perfect matchings on $\tilde{G}(v)$:

THEOREM 3.2. *Let G be a finite connected planar graph and f^* be the vertex of the dual graph G^* that corresponds to the unbounded outer face of G. If v is incident with f^*, then there is a bijection between spanning trees of G rooted at v and perfect matchings of $\tilde{G}(v)$.*

PROOF. Let v be incident with f^* and let let T be a directed spanning tree on G rooted at v. We will show how to find the unique corresponding perfect matching on $\tilde{G}(v)$. As we have seen in the proof of theorem 3.1, T determines a spanning tree T^* on G^*, which we will consider as a directed spanning tree with root f^*. This situation is shown in fig. 3.3 (b).

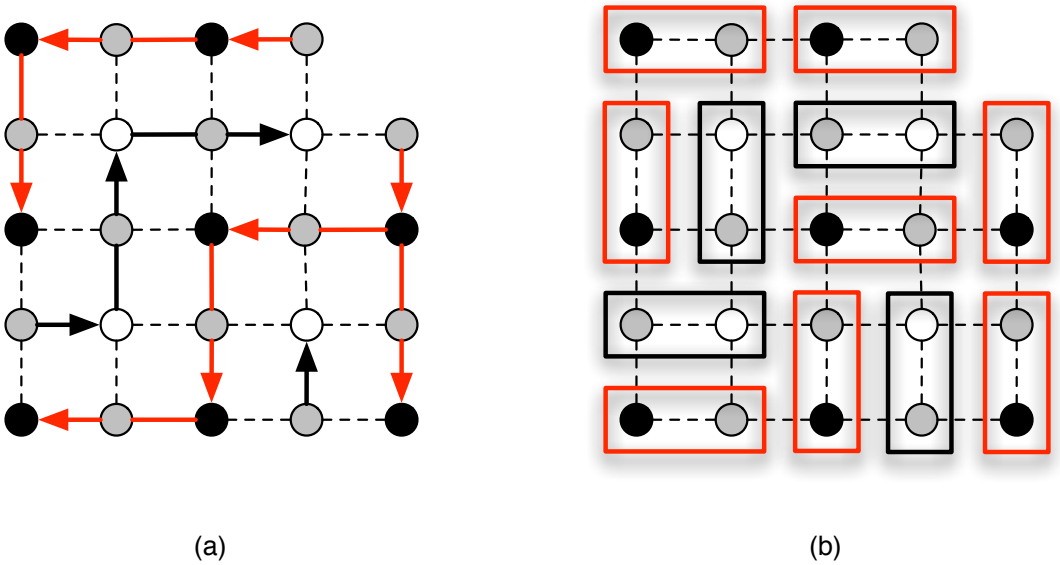

FIGURE 3.4. Obtaining a domino tiling from a directed spanning tree

Consider now the same tree as a subgraph of $\tilde{G}(v)$ (see fig. 3.4 (a)). Every edge of the tree T coincides with exactly one edge-node (depicted gray) and points at exactly one vertex-node (black) of $\tilde{G}(v)$. Similarly, every edge of the tree T^* corresponds to one edge-node and points at one face-node (white). Match every edge-node with the node that the corresponding edge points towards, as shown on the right of fig. 3.4.

Since every edge-node of $\tilde{G}(v)$ can be identified with one edge in T *or* in T^* this gives a unique way of pairing the nodes. As we have shown that $\tilde{G}(v)$ is balanced and bipartite we can be sure to obtain a domino tiling this way.

Conversely, given a domino tiling D on $\tilde{G}(v)$ we can construct a spanning tree $\tilde{T}$ on G simply by putting an edge e in $\tilde{T}$ iff $\tilde{e}$ is paired with a vertex-node for every $e \in E_G$. To see that this is in fact a spanning tree, observe that $\tilde{G}(v)$ has $|V_G| - 1$ vertex-nodes. Therefore $\tilde{T}$ has as many edges, as it should have. It remains to show that it is acyclic.

Suppose $\tilde{T}$ contained a cycle C, say of length n, which divides the plane into two regions, one inside and the other one outside of C.

CLAIM. *Each of those regions contains an odd number of nodes of* $\tilde{G}(v)$.

Suppose at least one of the regions contains an even number of nodes. Replace the other region with a single face (if both are even choose any one). The number of nodes in the resulting graph is given by $|V| + |E| + |F|$, which, by Euler's formula, must be even. The cycle itself contains $2n$ nodes (n vertices and n edges) and the modified region 1 node. Therefore the unmodified region must have an odd number of elements as well.

♢

As C separates $\tilde{G}(v)$ into two unconnected parts, D must be a domino tiling on both regions independent of the other, which is impossible since each region contains an odd number of nodes. This shows that $\tilde{T}$ is in fact a spanning tree on G and by construction $\tilde{T} = T$. □

REMARK. This bijection was first discovered by Temperley in 1974 for $m \times n$ rectangular grids. It was later generalized by Propp and, independently, by Burton and Pemantle to arbitrary (unweighted) planar graphs. The extension to the directed weighted case was found by Kenyon, Propp and Wilson [**16**].

3. Trees and random walks

A path $\mathscr{P}$ in a graph G is a sequence of vertices such that there is an edge in G connecting every pair of successive vertices in $\mathscr{P}$. For any vertex v define the *simple random walk* on G starting at v intuitively as follows: Consider a particle moving randomly on the graph, starting at v. At each time $t \in \mathbb{N}$ it chooses uniformly among the edges of the vertex it is currently at.

This defines a random path $(SRW_v^G(0), SRW_v^G(1), \dots)$ on G with the random function $SRW_v^G : \mathbb{N} \to V_G$ (Which we will denote only SRW if there is no danger of confusion.) Another important concept that brings us one step closer to trees is the *loop erasure* of a path:

DEFINITION (Loop erasure). Let $\mathscr{P}$ be any finite path $\langle v_0, v_1, \dots, v_l \rangle$ in G. Define the *loop erasure* $\mathrm{LE}(\mathscr{P})$ inductively: The first vertex u_0 of $\mathrm{LE}(\mathscr{P})$ is the first vertex v_0 of $\mathscr{P}$. Suppose now that we know u_j and let k be the last index such that $u_j = v_k$. Then put $u_{j+1} := v_{k+1}$ if $k < l$, otherwise let $\mathrm{LE}(\mathscr{P})$ remain $\langle u_0, \dots, u_j \rangle$.

REMARK. A better name for this procedure would be "cycle erasure", because usually a *loop* denotes an edge that connects a vertex to itself. Unfortunately, the name is standard.

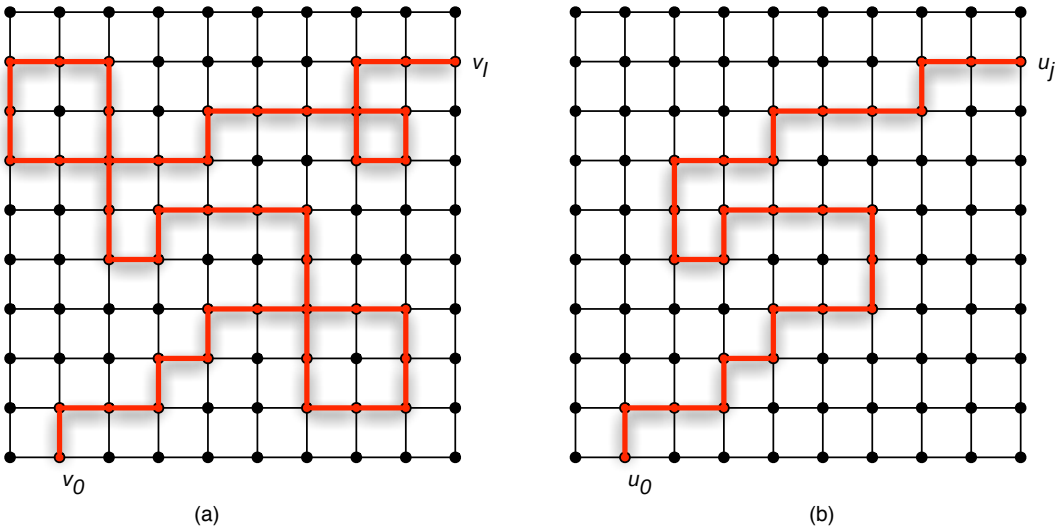

FIGURE 3.5. A path $\mathscr{P}$ (a) and LE($\mathscr{P}$) (b)

Intuitively it is clear how LE($\mathscr{P}$) is constructed: Move along the path $\mathscr{P}$ and delete any cycles in order in which they appear. An example for a path and the corresponding loop erasure is shown in figure 3.5.

Using this we can generate a random spanning tree: First, pick any vertex r to be the root and create a growing sequence T_i ($i \geq 0$) of trees inductively: Let $T_0 := r$. Suppose now that we know T_i. If it spans G, we're done. Otherwise, pick any vertex $v_{i+1} \notin T_i$ and start a simple random walk from there, stopping the first time it hits T_i. Create T_{i+1} by adding the loop erasure of this random walk to T_i. The set of edges of the last tree in this growing sequence, where the root r is simply forgotten (i.e. the last tree is considered as an undirected tree), is the output of *Wilsons's algorithm*. We have

THEOREM 3.3 (Wilson (1996)). *Let G be a finite graph. Wilson's algorithm gives a spanning tree on G with uniform distribution, no matter what vertices are chosen for the loop-erased random walks.*

For a complete proof see the original article by Wilson [**36**]. We will only give a brief

SKETCH OF PROOF. After using Wilson's algorithm, we get a spanning tree T and a number of erased cycles $C_1, \ldots, C_n$. The probability of SRW making exactly those transitions is

$$\left(\prod_{v \neq r} \frac{1}{d(v)}\right) \left(\prod_{i=1}^{n} \prod_{u \in C_i} \frac{1}{d(u)}\right). \tag{3.1}$$

This follows from the Markov property of SRW and the fact that in every step all the neighbors are equally likely transitions. We thus get a product of two terms, one involving only vertices in the tree and the other in the cycles. Therefore, the tree is independent of the cycles and all the trees have the same probability. □

4. Random walks and electrical networks

The connection between electrical networks and random walks (and thus spanning trees) is by itself a beautiful and rich field of research. It is also not new: Kirchhoff studied a similar relationship as early as 1847 [**18**]. An excellent treatment of all the main concepts, which we will follow closely in this section, can be found in Doyle and Snell's book *Random walks and electric networks* [**9**], which is also available online.

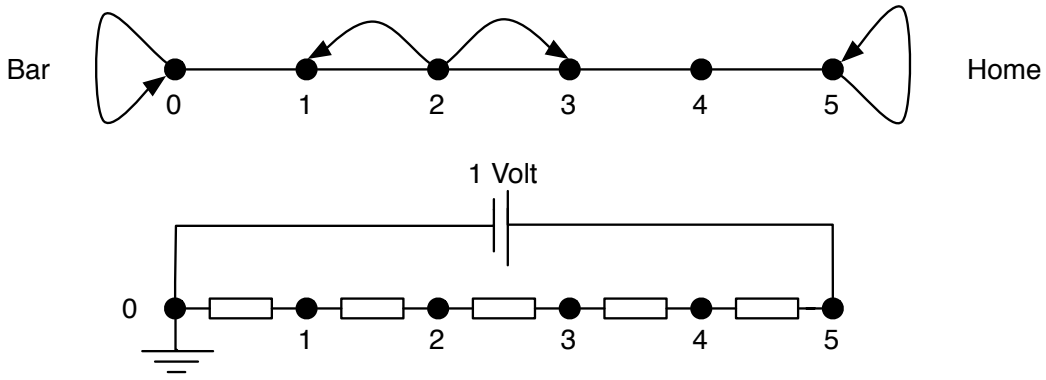

FIGURE 3.6. A random walk on a finite graph and the corresponding electrical network

Let's start with a random walk on a finite graph, see Fig. 3.6 top. Imagine a drunk walking along a street with 5 blocks. He starts at block x and chooses one direction

randomly with probability 1/2. He walks in this direction until he reaches another block, at which point he again chooses his direction randomly, as before. He does this until he reaches block 5, his home, or block 0, the bar. If he reaches either of those, he stays there.

The basic problem here is to determine the probability $p(x)$ that the man, starting at x, will reach home before reaching the bar. Obviously $p(x)$ has the three properties:

(1) $p(0) = 0$
(2) $p(5) = 1$
(3) $p(x) = \frac{1}{2}p(x-1) + \frac{1}{2}p(x+1)$ for $x = 1, 2, 3, 4$.

In this example it is easy to verify that $p(x) = x/5$.

The corresponding electrical network is shown in Fig.3.6, bottom. To construct it, replace each edge in the graph with a 1 ohm resistor and connect the ends with a unit voltage. We can now measure the voltage $v(x)$ at each vertex. Since $x = 0$ is grounded we have $v(0) = 0$ and $v(5) = 1$ so properties (1) and (2) from before are satisfied. We will now show that point (3) is met as well. We will be using some basic principles from electrotechnics which can be found in any introductory text to this field, for example Beuth [**4**].

Kirchhoff's first rule states that at each vertex x the current flowing into x is the same as the current flowing out. If x and y are connected by a resistance R, the current i_{xy} flowing from x to y is

$$i_{xy} := \frac{v(x) - v(y)}{R}.$$

This is Ohm's Law. We therefore have

$$0 = i_{(x-1)x} + i_{(x+1)x} = \frac{v(x-1) - v(x)}{R} + \frac{v(x+1) - v(x)}{R}. \tag{3.2}$$

Multiplying by R and solving for $v(x)$ gives

$$v(x) = \frac{v(x+1) + v(x-1)}{2}, \tag{3.3}$$

which is property (3). So in this case we can easily see that $p(x)$ and $v(x)$ are the same. This is also true in much more general graphs than a couple of points connected on a straight line [**9**], but for us it suffices to show this in two dimensions.

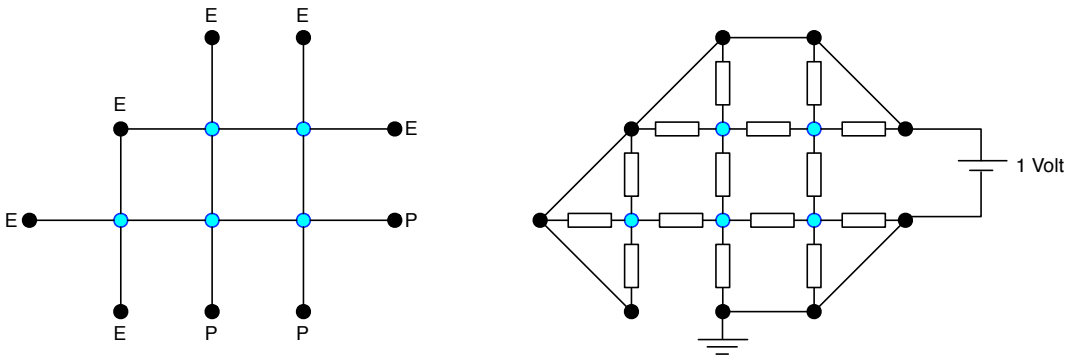

FIGURE 3.7. A random walk on a finite graph and the corresponding electrical network in two dimensions

As we will use this in the following chapters, we will now consider the case of a finite subgraph $G \subseteq \mathbb{Z}^2$ with the usual connections between neighbors, see figure 3.7, left. Consider a thief who starts at one of the blue vertices and chooses one of the four possible directions with equal probability until he either escapes at a point marked E or gets caught by a policeman at a point marked P. We want to find the probability $p(x)$ that the thief, starting at any point x, escapes before reaching a policeman.

As in one dimension we will consider the related electric circuit shown on the right of figure 3.7 obtained as follows: All the vertices marked P are connected and grounded, while all the vertices marked E are connected and kept at one volt by a battery. All the edges are equipped with a resistance R. Now we are looking for the voltage $v(x)$ at the blue vertices.

Again we would like to show that those two functions are identical. To do this, we split the graph G into two finite sets of vertices $G = D \cup B$ which satisfy:

(1) $D \cap B = \emptyset$.
(2) Every point in D has 4 neighbors in G.
(3) Every point in B has at least one neighbor in D.

We then call D the *interior* (depicted in blue in figure 3.7) and B the *boundary* of G. A function $f : G \to \mathbb{R}$ is called *harmonic* if

$$f(x,y) = \Delta f := \frac{f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1)}{4}, \qquad \forall (x,y) \in D.$$

Notice that this averaging property applies only to points in the interior of G.

The function $p(x)$ is harmonic: Since the probability of the thief moving in any one direction is $1/4$ this follows immediately by writing $p(x)$ as the sum of the four possible first steps. On the other hand, the fact that $v(x)$ is harmonic can be shown again using Kirchhoff's Laws. The current going in and out of $x = (a, b)$ adds to:

$$\begin{aligned} &\frac{v(a+1,b)-v(a,b)}{R}+\frac{v(a-1,b)-v(a,b)}{R} \\ &+\frac{v(a,b+1)-v(a,b)}{R}+\frac{v(a,b-1)-v(a,b)}{R}=0, \end{aligned} \tag{3.4}$$

from which we obtain the definition of harmonicity by multiplying by R and solving for $v(a, b)$. Therefore $p(x)$ and $v(x)$ are harmonic with identical boundary values. Our desired result now follows using the Uniqueness Principle which we will prove by way of the

THEOREM 3.4 (Maximum Principle). *A harmonic function f(x) defined on G takes on its maximum value M and its minimum value m at the boundary B.*

PROOF. Assume $f(P) = M$ for an interior point P. By the very definition of a harmonic function $f(P)$ is the average of the values at the surrounding points, therefore all these must equal M also. Working our way to the boundary we eventually reach a point $Q \in B$ for which $f(Q) = M$. The Maximum value is thus always attained at the boundary. The argument can remain unchanged for proving the assertion about the minimum value. □

THEOREM 3.5 (Uniqueness Prinicple). *If $f(x)$ and $g(x)$ are harmonic functions on G so that $f(x) = g(x)$ on B, then $f(x) = g(x)$ $\forall x \in G$.*

PROOF. Let $h(x) := f(x) - g(x)$ be the difference between the two harmonic functions. Then we have

$$\begin{aligned} \Delta h &= \frac{h(a+1,b) + h(a-1,b) + h(a,b+1) + h(a,b-1)}{4} = \\ &= \frac{f(a+1,b) + f(a-1,b) + f(a,b+1) + f(a,b-1)}{4} - \\ &= \frac{g(a+1,b) + g(a-1,b) + g(a,b+1) + g(a,b-1)}{4} \\ &= \Delta f - \Delta g = f(x) - g(x) = h(x). \end{aligned} \tag{3.5}$$

Thus the difference between two harmonic functions is itself an harmonic function. But $h(x) = 0$ for $x \in B$ and by the Maximum Principle the maximum and minimum values of h are 0, so that $h(x) = 0$ for every $x \in G$ and thus $f(x) = g(x)$ for every $x \in G$. □

We have shown that the symmetric random walk on subgraphs of $\mathbb{Z}^2$ is intimately connected with electrical networks. A result we will use later on is Raleygh's Monotonicity Law. In preparation for this we need to go a little bit deeper into the theory of electrical networks.

Setting up a voltage v between two points a and b means establishing a voltage $v_a := v(a) = v$ at a and $v_b = 0$. A current $i_a = \sum_x i_{ax}$ will enter the circuit from the battery. The total amount of current flowing depends on the total resistance of the circuit. We define the *effective resistance* R_{eff} between a and b by $R_{eff} = v_a/i_a$. Considering this from the viewpoint of a random walk we can interpret this quantity as an escape probability.

DEFINITION. Let R_{xy} be the resistance between two points x and y. When a current i_{xy} flows through a resistor, the *energy dissipation* is the quantity $i_{xy}^2 R_{xy}$, the product of the current and the voltage $v_{xy} = i_{xy}R_{xy}$. The *total energy dissipation* in the whole circuit is

$$E := \frac{1}{2} \sum_{x,y} i_{xy}^2 R_{xy}.$$

DEFINITION. A *flow* $\mathbf{j}$ from a to b is a collection of numbers j_{xy} for the edges xy which satisfy:

(1) $j_{xy} = -j_{yx}$

(2) $\sum_y j_{xy} = 0$ if $x \neq a, b$

(3) $j_{xy} = 0$ if $x \not\sim y$

If $v_a = 1$, the resulting flow from a to b is called a *unit current flow.*

THEOREM 3.6 (Conservation of Energy). *Let $\boldsymbol{j}$ be a flow from a to b and $w : E(G) \to \mathbb{R}$. Then*

$$(w_a - w_b)\, j_a = \frac{1}{2} \sum_{x,y} (w_x - w_y)\, j_{xy}.$$

PROOF. This can be shown directly by calculating the right-hand side:

$$\begin{aligned} \sum_{x,y} (w_x - w_y)\, j_{xy} &= \sum_x \left(w_x \sum_y j_{xy} \right) - \sum_y \left(w_y \sum_x j_{xy} \right) \\ &= w_a \sum_y j_{ay} + w_b \sum_y j_{by} - w_y \sum_x j_{xa} - w_b \sum_x j_x b \\ &= w_a j_a + w_b j_b - w_a\,(-j_a) - w_b\,(-j_b) \\ &= 2\,(w_a - w_b)\, j_a. \end{aligned} \tag{3.6}$$

□

THEOREM 3.7 (Thomson's Principle). *The unit flow $\boldsymbol{i}$ from a to b resulting from Kirchhff's Laws minimizes the energy dissipation $\frac{1}{2}\sum_{x,y} j_{xy}^2 R_{xy}$ among all unit flows $\boldsymbol{j}$ from a to b*

PROOF. Let $\mathbf{j}$ be an arbitrary unit flow from a to b and set $d_{xy} = j_{xy} - i_{xy}$, which is also a flow from a to b with $d_a = \sum_x d_{ax} = 1 - 1 = 0$. Calculating the energy dissipation we get:

$$\begin{aligned} \sum_{x,y} j_{xy}^2 R_{xy} &= \sum_{x,y} (i_{xy} + d_{xy})^2 R_{xy} \\ &= \sum_{x,y} i_{xy}^2 R_{xy} + 2 \sum_{x,y} i_{xy} R_{xy} d_{xy} + \sum_{x,y} d_{xy}^2 R_{xy} \\ &= \sum_{x,y} i_{xy}^2 R_{xy} + 2 \sum_{x,y} (v_x - v_y)\, d_{xy} + \sum_{x,y} d_{xy}^2 R_{xy}. \end{aligned} \tag{3.7}$$

But after the previous result the middle term equals $4\,(v_a - v_b)\,d_a = 0$, hence

$$\sum_{x,y} j_{xy}^2 R_{xy} = \sum_{x,y} i_{xy}^2 R_{xy} + \sum_{x,y} d_{xy}^2 R_{xy} \geq \sum_{x,y} i_{xy}^2 R_{xy}. \tag{3.8}$$

□

We are finally ready to prove

THEOREM 3.8 (Rayleigh's Monotonicity Law). *If the resistances of a circuit are increased, the effective resistance R_{eff} between any two points can only increase. If they are decreased, it can only decrease.*

REMARK. This Law seems trivial. Indeed, James Clerk Maxwell called it "self-evident" in his monumental *Treatise on Electricity and Magnetism* ([**22**], p. 354). Nonetheless, since it will play quite an important part in the arguments to follow, it is instructive to check why this must be so.

PROOF OF THEOREM 3.8. Pick any two points a and b and let $\mathbf{i}$ be the current flow between them with resistors R_{xy} and let $\mathbf{j}$ the current for another set of resistors $\bar{R}_{xy}$ with $\bar{R}_{xy} \geq R_{xy}$. We get

$$\bar{R}_{eff} = \frac{1}{2}\sum_{x,y} j_{xy}^2 \bar{R}_{xy} \geq \frac{1}{2}\sum_{x,y} j_{xy}^2 R_{xy} \geq \frac{1}{2}\sum_{x,y} i_{xy}^2 R_{xy} = R_{eff}, \tag{3.9}$$

where we used Thompson's Principle. The proof for the case where the resistances are decreasing is similar. □

As we've mentioned before, Rayleigh's Monotonicity Law seems trivial at first glance. But this is only true as long as we are thinking about electricity and current flows and their analogy to the flow of water: Obviously reducing the number of pipes in which water can flow will increase the amount of water in the other pipes. Thinking about this situation in terms of a random walk however makes it much less straight-forward.

Consider again the thief who tries to evade policemen. Rayleigh's Monotonicity Law states that the escape probability increases when another possible route is added. But why should this be so? Of course, this adds another possibility of escaping. On the other hand, this adds another possibility of returning to the starting point, too!

This is but one example of a situation where considering a difficult problem from a completely different point of view renders it almost trivial. Rayleigh's Monotonicity Law can be proved directly probabilistically (for example using Markov Chains, as in Doyle and Snell [**9**]), but the proof is rather tedious.

CHAPTER 4

Measures of Spanning Trees

For any finite graph G there is only a finite number of spanning trees on G. Therefore there exists a uniform measure on the spanning trees on G, which we'll denote by μ_G. In this chapter we will extend this notion to infinite graphs using an appropriate limit. Uniform measures concentrated on spanning trees or forests on infinite graphs were first studied by Pemantle in 1991 [**25**] using methods developed by Broder [**5**] and Aldous [**2**]. In 2001, Benjamini, Lyons, Peres and Schramm (hereafter referred to as BLPS) published an extensive paper [**3**] which redeveloped the theory using an algorithm for generating spanning trees due to Wilson [**36**].

For every connected graph there exists at least one spanning tree. For most graphs, the number of possible trees increases rapidly with the size of the graph and it is by no means a trivial problem to choose one at random. However, since generating trees at random according to the uniform measure is of interest not only to mathematicians but also to computer scientists, more and more efficient algorithms have been developed over the years. The first ones used Kirchhoff's Matrix-Tree Theorem:

THEOREM 4.1 (Matrix-Tree Theorem). *Let G be a finite connected graph. Define the negative Laplacian $L(G)$ as the matrix*

$$L(G)_{v,w} = \begin{cases} d(v), & \text{if } v = w \\ -1, & \text{if } v \backsim w \\ 0, & \text{else.} \end{cases}$$

Then the determinant of the submatrix of $L(G)$ obtained by deleting any row r and any column s from $L(G)$ gives the number of spanning trees of G.

(Without proof)

But this is hard to calculate since a large number of determinants has to be found, and it doesn't lend itself well to theoretical study. Later algorithms used connections between spanning trees and random walks as discussed in section 3 of the previous chapter, notably developed by Broder [**5**], Aldous [**2**] and Wilson [**36**]. They were not only faster but much easier to analyze for probabilists. Following BLPS [**3**], we will study random trees on infinite graphs using Wilson's algorithm (see Theorem 3.3 on page 20).

1. Infinite graphs

We will now take some time to discuss differences between finite and infinite graphs. However, we will only concern ourselves with graphs that have countably many vertices and finitely many edges at each vertex, such as the nearest-neighbor graph on $\mathbb{Z}^2$.

The terminology is the same as for finite graphs, except for a few new concept that arise because of the infinity of the vertex set. We say a graph is *locally finite* if all the vertices have finite degrees. An infinite graph (E, V) of the form

$$V = \{x_0, x_1, x_2, \dots\}, \qquad E = \{x_0x_1, x_1x_2, \dots\}$$

is called a *ray*. A *double ray* is an infinite graph of the form

$$V = \{\dots, x_{-1}, x_0, x_1, \dots\}, \qquad E = \{\dots, x_{-1}x_0, x_0x_1, \dots\},$$

where $x_i \neq x_j$ for every $i \neq j$. Finite (double) rays are called paths. The subrays of rays are called *tails*. Since rays are infinite, they have infinitely many tails, but any two of them only differ by finitely many initial vertices.

A concept of infinite graph theory that will play an important part later on is the notion of an end. An *end* of a graph G is an equivalence class of rays, where two rays are said to be equivalent if, for every finite subset of vertices $S \subseteq V(G)$, both have a tail in the same component of $G \backslash S$.

For a tree, ends are pretty straightforward: Two rays are equivalent if they share infinitely many vertices. One has to be careful though: Even locally finite trees can have uncountably many ends, the standard example being the binary tree T_2, in which every vertex set has exactly two upper neighbors, depicted in figure 4.1. It can be interpreted

as the set of finite 0-1 sequences with the empty sequence as the root. The ends of T_2 then correspond to infinite 0-1 sequences.

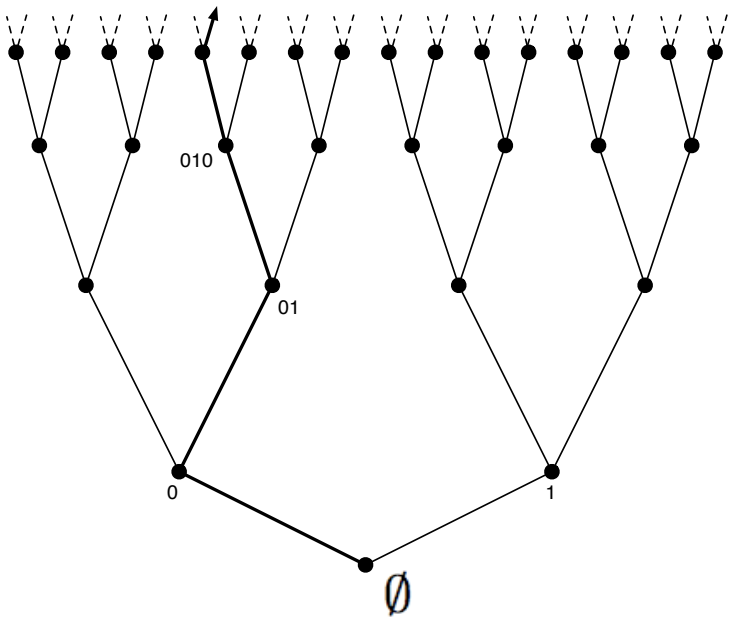

FIGURE 4.1. The binary tree T_2

An end of a graph can be interpreted as points at infinity, towards which its rays converge.

2. Measures on infinite graphs

We will use Wilson's algorithm to construct a measure on an infinite graph G. Let V be the vertex set of G and let $V_1 \subset V_s \subset \dots$ be finite connected subsets of V so that $\cup_{n=1}^{\infty} V_n = V$. Let $G_n = (V_n, E_n)$, where an edge (of G) is in G_n if both its endpoints are in V_n. We say G_n is *spanned* by V_n and $\langle G_n \rangle$ an *exhaustion* of G. Denote by μ_{G_n} the uniform measure on spanning trees of G_n.

It was proved by Pemantle [**25**] that if an infinite connected graph G is exhausted by a sequence G_n, the weak limit of $\langle \mu_{G_n} \rangle$ exists. This limit measure is now called the *free (uniform) spanning forest* on G, denoted FSF. Pemantle implicitly showed the existence of another limit measure, now called *wired (uniform) spanning forest*, short WSF. For

an overview over these concepts see Lyons (1998) [**20**]. BLPS (2001) [**3**], which we will follow in this section, gives an exhaustive treatment.

Let μ_n^F be the uniform spanning tree probability measure on G_n. For any finite set B of edges, there exists a natural number N such that for all $n \geq N$ we have $B \subseteq E_n$. It is obvious, that

$$\mu_n^F(B \subseteq T) \geq \mu_{n+1}^F(B \subseteq T), \tag{4.1}$$

which follows directly from Rayleigh's Monotonicity Law (Theorem 3.8). From this we see that the limit $\mu^F(B \subseteq T) := \lim_{n\to\infty} \mu_n^F(B \subseteq T)$ exists. μ^F is called the *free uniform spanning forest* measure on G and is independent of the exhaustion $\langle G_n \rangle$.

There is another method of considering limits of spanning trees. When we constructed μ^F we considered only G_n (and spanning trees thereon) but ignored the complement. There is however the possibility that a spanning tree of G connects the vertices on the boundary of G_n by a path *outside* of G_n. Therefore we look at the graphs G_n^W obtained from G_n by identifying all the boundary vertices and μ_n^W, the uniform spanning tree measure on G_n^W.

Using the same argument as before we can see that this defines again a limit measure μ^W, which again does not depend on the particular exhaustion. It is called the *wired uniform spanning forest*, or WSF. The term "wired" comes from the idea of "wiring" the boundary vertices of G_n together.

In fact, the WSF is easier to analyze and thus better understood than the FSF. We will however not dwell too much on the differences, but in fact often drop the terms "wired" and "free" altogether, since for the graph $G = \mathbb{Z}^2$ the two are identical.

This is true for a quite large number of graphs: As our next Theorem proves this holds for every *recurrent* graph, i.e. a graph on which the symmetric random walk is recurrent. For a good introduction to recurrence, transience and the related topics consult Sheldon [**27**].

The fact that $\mathbb{Z}^2$ is recurrent might be surprising given the fact that SRW on $\mathbb{Z}^d$ is transient for all $d \geq 3$. This was first proven by Pólya in 1921 and today the probabilities that a random walk on a d-dimensional lattice returns to the origin are called *Pólya's*

random walk constants. For an insight into these topics and a proof that $\mathbb{Z}^2$ is in fact recurrent, see Rudnick and Gaspari's treatise on random walks [**28**].

THEOREM 4.2. *Let G be an infinite recurrent graph. The random spanning tree T_G obtained by Wilson's method (with arbitrary root and ordering of the vertices) has the same distribution as the WSF and FSF.*

PROOF. Let $\langle G_n \rangle$ be an exhaustion of G by finite graphs. We would like to show that for any finite event $B \in 2^E$ (that is, the measurable space of all subsets of E with the Borel σ-field),

$$|\mathbb{P}[T_G \in B] - \mu_n^W[B]| \to 0 \text{ as } n \to \infty, \tag{4.2}$$

and similar fo μ_n^F. Let K_0 denote the set of vertices connected by the edges in B and let K be the union of K_0 and the set of vertices that precede some vertex in K_0 in the ordering given by Wilson's method. We call $\partial_V G_n$ the *vertex boundary* of G_n, i.e. the set of vertices in the complement of G_n that are adjacent to some vertex in G_n.

For a random walk $\langle X(n) \rangle$, we will use τ_v to denote the *hitting time*:

$$\tau_v := \inf\{n > 0 : X_n = v\}.$$

We then have

$$| \mathbb{P} [T_G \in B] - \mu_n^W[B]| \leq \sum_{v \in K} \mathbb{P}_v[\tau_{\partial_V G_n} < \tau_r], \tag{4.3}$$

but since SRW on $\mathbb{Z}^2$ is recurrent, the right hand side tends to 0 as n goes to infinity. The same argument applies to μ^F as well. □

REMARK. The converse is not true: There are many transient lattices for which the free and wired spanning forest measures coincide, for example $\mathbb{Z}^d$. For a precise discussion of when this is the case see BLPS [**3**].

At this point there is one question begging to be answered: Why are the limiting measures called wired/free spanning *forests*, instead of wired/free spanning *trees*? It turns out that in general the uniform measure of spanning trees on a graph does not necessarily converge to a measure concentrated on single trees. For example, in $\mathbb{Z}^d$ for

$d \geq 5$, the uniform spanning trees on G_n do converge to a limiting distribution, but this consists of an essential spanning forest with infinitely many components [**25**].

This seems surprising at first glance, but consider that it is by no means easy to distinguish a spanning tree from a forest on $\mathbb{Z}^d$. In fact, it is impossible to tell an essential spanning forest from a spanning tree with only a finite amount of information.

The question when WSF or FSF on an infinite graph actually consists of a single component (i.e. when it is a tree) has been discussed for a while. As mentioned above, this was settled by Pemantle for $\mathbb{Z}^d$, but to get the full picture we again refer to BLPS [**3**].

For $\mathbb{Z}^2$ the question can be answered very quickly: Since we know that the WSF and FSF coincide on $\mathbb{Z}^2$ and can be generated by Wilsons's method, the uniform measure on the integer lattice can be defined without a limiting measure since $\mathbb{Z}^2$ is recurrent and thus SRW hits every point. It is the clear from the construction that the resulting subgraph must be connected.

Much harder to investigate is the question of the *shape* of the components of these spanning trees. As we will see in the next section there are quite many interesting results, which will also be useful in the next chapter.

3. The shape of components of the uniform measure

We now turn to the most important question of this chapter: What can be said about the shape of elements of $\mu_{\mathbb{Z}^2}$? First we will need some additional definitions to describe the shapes of subgraphs of $\mathbb{Z}^2$.

Definition. A subset of $\mathbb{Z}^2$ has *density* α if for each sequence of rectangles $B_1 \subseteq B_2 \subseteq \ldots$ with $\bigcup_{n\geq 1} B_n = \mathbb{Z}^2$ the limit

$$\lim_{n\to\infty} \frac{|S \cap B_n|}{|B_n|}$$

exists and is equal to α. Otherwise S is called *rough*.

We also introduce the following notation: Let Y be a finite set with at least three elements. A *partition* of Y is a collection $P = \{P_1, P_2, P_3\}$ of three non-empty pairwise disjoint subsets of Y whose union is Y. Two partitions P and Q of Y are said *compatible*

if there is an ordering of each such that $Q_2 \cup Q_3 \subseteq P_1$, and a collection $\mathbf{P}$ of partitions is *compatible* if each pair $P, Q \in \mathbf{P}$ is compatible.

LEMMA 4.3. *If $\boldsymbol{P}$ is a compatible collection of partitions of Y, then $|\mathbf{P}| \leq |Y| - 2$.*

PROOF. We will show this by induction by $n = |Y|$. If $|Y| = 3$, then Y has only one partition and the lemma is valid. Suppose now that it holds for $|Y| < n$. Choose any partition $P \in \mathbf{P}$. Then $|P_i| = n_i > 0$ and $n_1 + n_2 + n_3 = n$. Divide $\mathbf{P} \setminus P$ into three classes $\mathbf{P_i}$ with $Q \in \mathbf{P_i}$ if

$$Q_2 \cup Q_3 \subseteq P_i,$$

possibly after reordering the indices of Q.

Fix $i = 1$ and define $Y_1 = P_1 \cup \{v\}$ for $v \notin Y$, and for each $Q \in \mathbf{P_1}$ define a partition $\tilde{Q}$ of Y as $\tilde{Q} = \{\tilde{Q_1}, Q_2, Q_3\}$ with $\tilde{Q_1} = (Q_1 \cap P_1) \cup \{v\}$. We have now constructed a compatible collection of partitions of Y_1:

$$\tilde{\mathbf{P_1}} = \{\tilde{Q} : Q \in \mathbf{P_1}\},$$

and for which $|Y_1| = n_1 + 1 < n_1 + n_2 + n_3 = n$.

By induction we then get

$$|\mathbf{P}| = \sum_i |\mathbf{P_i} + 1 \leq \sum_i ((n_i + 1) - 2) + 1 = n - 2. \tag{4.4}$$

□

A vertex x of a subgraph S of $\mathbb{Z}^2$ is called a *separator* if removing it leaves more than one infinite component, and a *branchpoint* if it leaves more than two infinite components. Burton and Keane (1989) [**6**] showed that the following important property of stationary measures:

THEOREM 4.4. *If μ is translation-invariant then μ-almost surely all trees have one or two topological ends.*

PROOF. Since we assume that μ is ergodic it follows that for μ-almost all trees all large enough rectangles B contain at least $\epsilon|B|$ branching points. We then only need to

show that for any tree and any rectangle B, the number of branching points of T in B is less than the number of points on the boundary of B. This will suffice, since

$$\epsilon|B| \not< |\partial B|$$

for large enough B and positive ϵ.

Let therefore T be a tree of $\mathbb{Z}^2$ and set

$$Y = T \cap \partial B.$$

If $x \in B$ is a branching point for T, then removal of it defines a partition

$$P = \{P_1, P_2, P_3\}$$

of Y so that $P_i \neq \emptyset$ for $1 \leq i \leq 3$. If $\tilde{x}$ is another branching point in B with partition $Q = \{Q_1, Q_2, Q_3\}$, then the indices for P and Q can be chosen as to satisfy $Q_1 \cup Q_2 \subseteq P_1$.

The previous lemma concerning compatible partitions implies that the number of branching points belonging to T is at most $|Y| - 2$ and summing over all the trees completes the proof. □

Consider now the uniform measure $\mu_{\mathbb{Z}^2}$. It is clear from the construction of the measure via random walks that it is concentrated on connected graphs, i.e. single trees, since in $\mathbb{Z}^2$ SRW hits every point. But this measure has also another important property:

THEOREM 4.5. *Under the uniform measure $\mu_{\mathbb{Z}^2}$ each tree has only one topological end almost surely.*

REMARK. In other words, the measure consists only of trees with the property that when any vertex is removed, the tree is split up into exactly one finite and one infinite part.

PROOF. We will proof this by contradiction. Since the set of branchpoints has density zero the tree on $\mathbb{Z}^2$ has at most two topological ends almost surely. Since the measure is translation-invariant, the number of topological ends must be almost surely constant. Assume that there are two topological ends almost surely.

The spanning tree then can be thought of as a doubly infinite line to which a finite tree has been added at every vertex. This line corresponds to the set of all separators

and has density, say, D. We now say that for any two vertices v_1 and v_2, the vertex v_3 separates v_1 and v_2 if the path in T from v_1 to v_2 passes through v_3. If three vertices lie on the infinite line of T then one of them has to separate the other two. We therefore have:

$$\sum_i \mathbb{P}[v_i \text{ separates the other two}] \geq \mathbb{P}[v_1, v_2 \text{ and } v_3 \text{ are separators}]. \tag{4.5}$$

We now move the vertices v_i apart so that the distances $|v_i - v_j|$ all go towards infinity. The measure μ is 3-mixing, so the right hand side goes to D^3. To reach the contradiction we have to show that the left-hand side goes to zero.

Let the pairwise distances be larger than L for some $L > 0$. Choose B_n large enough to contain the vertices v_i which we'll call v, w and x for simplicity sake, and fix n. We now start a random walk from v and denote by γ the segment until the first time it hits w. This segment determines whether x separates v and w:

If γ does not hit x then x doesn't separate v and w. Otherwise let γ_1 be the segment up until x and γ_2 be the second part, up until it hits w. Since the construction is the same, the path connecting v and w in T is the same as the one connecting them in $T(\gamma)$, thus:

$$LE(\gamma) = LE(\gamma_2 * \gamma_1))LE(LE(\gamma_2) * \gamma_1). \tag{4.6}$$

In this expression x appears only once: On the right-hand side, where the two segments join. The vertex x seperates v and w if and only if it does not get erased when we take the loop-erasure of $LE(\gamma_2) * \gamma_1$.This is only the case when γ_1 intersects with $LE(\gamma_2)$, so x is a separator iff γ_1 and $LE(\gamma_2)$ are disjoint except at x. It remains to show that the probability of this being the case goes to zero as $n \to \infty$ and $L \to \infty$.

Let now $M \in \mathbb{N}$ and let $\gamma \wedge M$ denote the first segment of the path γ up to $\gamma(M)$. The probability that $LE(\gamma_2) \wedge M$ and $\gamma_1 \wedge M$ are disjoint is an upper bound for the probability that $LE(\gamma_2)$ and γ_1 are disjoint. We are thus finished if we can show that

$$\inf_M \lim_L \lim_n \mathbb{P}[LE(\gamma_2) \wedge M \cap \gamma_1 \wedge M \neq \{x\}] = 0. \tag{4.7}$$

But γ_1 and γ_2 are independent random walks, and combined with the fact that the distribution is the same no matter in which direction we follow the SRW, we can rewrite this as

$$\inf_M \mathbb{P}[LERW_x \ \wedge M \cap SRW_x \wedge M \neq \{x\}] = 0, \tag{4.8}$$

which finishes the proof. □

CHAPTER 5

The measure of maximum entropy

We are now ready to tackle our main theorem: It states that the uniform measure of essential spanning trees is the unique translation-invariant measure on the set of essential spanning trees on $\mathbb{Z}^2$.

Let $\mathcal{E}_{\mathbb{Z}^2}$ be the set of probability measures concentrated on spanning forests on $\mathbb{Z}^2$ that are translation-invariant and that maximize entropy. Then we can write our main theorem very succinctly:

THEOREM 5.1. $\mathcal{E}_{\mathbb{Z}^2} = \{\mu_{\mathbb{Z}^2}\}$.

It will take some preliminaries before we can actually go about proving this. After some historical background on the search for this proof we will devote a section to the basic definitions and the results we will need before finally tackling the theorem and its implications.

1. Historical overview

The entropy of the domino process on $\mathbb{Z}^2$ was first calculated by Kasteleyn in 1961 ([**13**]) as the exponential growth rate of the number of tilings of a large rectangle. However, since the boundary conditions play a crucial role in this calculation, this does not show that this is in fact the largest entropy possible. In the concluding remarks to his paper *The Statistics of Dimers on a Lattice* [**13**] he writes:

> The effect of boundary conditions is, however, not entirely trivial and will be discussed in more detail in a subsequent paper.

This turned out to be something of an understatement. Hardly any progress was made until 1993, when Burton and Pemantle published a monumental paper [**7**], using results provided by Pemantle in 1991 [**25**], where the assertion was proved by linking the domino process to uniform spanning trees, as we have done in chapter 3. Unfortunately,

the proof, albeit quite short and elegant, turned out to be wrong. The error was pointed out by Lyons [**21**] and a year later Sheffield [**30**] found a valid proof for the more general family of graphs.

2. Preliminary considerations

A natural strategy for proving claims like Theorem 5.1 is to show that every translation-invariant measure of maximum entropy has a so called Gibbs-property and that this property in turn characterizes the measure. Burton and Pemantle used this strategy but (erroneously) implied that every maximizing measure satisfies the following property:

DEFINITION (Strong Gibbs Property). Let μ be a measure on $\mathbb{Z}^d$ and $H \subseteq \mathbb{Z}^d$ be a finite induced subgraph of $\mathbb{Z}^d$. We write $a \backsim_O b$ for two vertices on the border of H a and b if there is a path between them that consists of edges *outside* of H. This obviously depends on $F_{\mathbb{Z}^d \setminus H}$. Identify now the vertices equivalent under $\backsim_O$ and call the resulting graph $\tilde{H}$.

Construct now a measure $\tilde{\mu}$ as follows: First pick a spanning tree from $F_{\mathbb{Z}^d \setminus H}$ according to μ and from F_H uniformly from the set of all spanning trees on $\tilde{H}$ (this is no problem because H and $\tilde{H}$ share the same edges.)

We say μ satisfies the *Strong Gibbs Property* if $\mu = \tilde{\mu}$.

REMARK. This construction of the graph $\tilde{H}$ could make problems if an edge of the tree was removed when identifying vertices under $\backsim_O$. This cannot happen, since the tree can't contain loops per definition.

For any finite graph, every entropy maximizing measure has the Strong Gibbs Property. For infinite graphs this is not necessarily the case: Consider for example the case $\mathbb{Z}^d$ with $d \geq 4$. As we have seen in the preceeding chapter, F contains $\mu_{\mathbb{Z}^d}$-a.s. infinitely many trees, each of which has only one topological end. Therefore, given $F_{\mathbb{Z}^d \setminus H}$, all elements of F_H that join distinct infinite trees of $F_{\mathbb{Z}^d \setminus H}$ have probability zero.

Following Sheffield [**30**], we claim however that every $\mu \in \mathcal{E}_{\mathbb{Z}^2}$ does satisfy a slightly different property and will use this fact to prove Theorem 5.1:

DEFINITION (Weak Gibbs Property). Let $H \subseteq \mathbb{Z}^d$ as before. Let a and b be two vertices on the edge of H, as before. We now write $a \backsim_I b$ if a and b are connected by a path *inside* of H. For this, no knowledge of $F_{\mathbb{Z}^d \setminus H}$ is necessary. We then construct a probability measure $\tilde{\mu}$ as before:

Under $\tilde{\mu}$, all the spanning forests on H that give the same relationship occur with equal probability. Then we say μ has the *Weak Gibbs Property*, if $\mu = \tilde{\mu}$.

To see that every measure that maximizes entropy must have the Weak Gibbs Property, assume that there exists a $\mu \in \mathcal{E}_{\mathbb{Z}^2}$ which doesn't. This means that in a finite subgraph H, conditioned on $\backsim_I$, μ is not uniform. If we now take a random collection S of nonintersecting translates of H and resample $F_{\tilde{H}}$ independently for each $\tilde{H} \in S$ according to the conditional measure, we can construct another measure $\tilde{\mu}$ from μ that has higher entropy.

3. Proof of the main theorem

We will prove Theorem 5.1 in two steps: First we will show that when μ is translation-invariant, has the Weak Gibbs Property and μ-a.s. contains only one tree with one topological end, then $\mu = \mu_{\mathbb{Z}^2}$. We will then demonstrate that otherwise $h_\mu \leq h_{\mu_{\mathbb{Z}^2}}$.

The following lemma will be needed in the proof:

LEMMA 5.2. *Let (Ω, μ) be a probability space and $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$ be binary random variables so that for every $\omega \in \Omega$ with $|\{i | X_i \neq Y_i\}| \leq K$. Then*

$$|h_\mu(X) - h_\mu(Y)| < K \log(n),$$

where $h_\mu(X)$ denotes the measure-theoretic entropy of the partition given by X.

REMARK. In other words, we can estimate the change in entropy of two random variables for small changes in the output of the process.

PROOF. Let $Z_i := \mathbb{1}_{X_i \neq Y_i}$. Then $h_\mu(X) \leq h_\mu(X, Z) = h_\mu(Y, Z) \leq h_\mu(Y) + h_\mu(Z)$. Here the entropy $h_\mu(X, Z)$ is the measure-theoretic entropy of the random variable $(X_1, X_2, \ldots, X_n, Z_1, Z_2, \ldots, Z_n)$. This and an analogue calculation for $h_\mu(Y)$ gives

$|h_\mu(X) - h_\mu(Y)| \leq h_\mu(Z)$. But from the definition on page 6 we get $h_\mu(Z) \leq \log(n^K)$, since Z consists of n elements at most K of which are non-zero. □

We will now prove the first part of Theorem 5.1.

LEMMA 5.3. *If μ has the Weak Gibbs Property and μ-almost surely all trees in $\mathbb{Z}^2$ have only one topological end, then $\mu = \mu_{\mathbb{Z}^2}$.*

PROOF. Fix any finite induced subgraph $B \subseteq \mathbb{Z}^2$. We will show that μ and μ_G induce the same law on B. For this, we construct another larger finite set $C \subseteq \mathbb{Z}^2$ that contains B. Denote by C_f the set of vertices of C that are starting points for infinite paths in F that do not enter C after the first point and by $\tilde{C}$ the union of C_f and all vertices that lie on finite components of $F \setminus C_f$. The set $\tilde{C} \subseteq V(\mathbb{Z}^2)$ is then the set of vertices v for which every infinite path in F that contains v also includes an element of C.

Finally we construct another, even larger vertex set D which contains not only all the vertices of C but also all its neighbors. We will now use the Weak Gibbs Property using this set. It says that if we condition on $F_{\mathbb{Z}^2 \setminus D}$ and the relationship $\backsim_I$ using D, then all the possibilities of F_D that extend $F_{\mathbb{Z}^2 \setminus D}$ to an essential spanning forest on $\mathbb{Z}^2$ and preserve the relationship $\backsim_I$ are equally likely.

We can also condition on the event $\tilde{C} \subseteq D$ and on a particular choice of $\tilde{C}$ and C_f to see that all the spanning forests of $\tilde{C}$ rooted at C_f (that is, all the spanning trees on the graph obtained from $\tilde{C}$ by identifying all the vertices in C_f) have the same possibility of appearing as the restriction of F to $\tilde{C}$.

But the set D can be chosen to be as large as we want, therefore it contains $\tilde{C}$ with probability arbitrarily close to 1. Therefore we see that, more generally, conditioned on $\tilde{C}$ and C_f, the spanning forests on $\tilde{C}$ rooted at C_f are uniformly distributed. But C too can be picked arbitrarily large. It follows that μ is the uniform measure on $\mathbb{Z}^2$. □

For the second part, let B_n be an exhaustion of $\mathbb{Z}^2$.

LEMMA 5.4. *Let $\tilde{B_n}$ have the same vertex set as B_n, albeit with arbitrary boundary conditions and let $\tilde{\mu_n}$ give equal probability to each spanning forest on $\tilde{B_n}$ in which every component touches the boundary. Then $h_{\tilde{\mu}} \leq h_{\mu_{\mathbb{Z}^2}}$.*

PROOF. We will first construct another measure $\tilde{\nu_n}$ concentrated on the spanning trees of $\mathbb{Z}^2$: Partition $\mathbb{Z}^2$ with translates of B_n and sample from independent copies of $\tilde{\mu_n}$ on every translate. Add edges as necessary to make the resulting spanning forest into a spanning tree in every translate (this takes at most $O(n)$ additional edges) and finally connect each of the trees by a path so that the result is a spanning tree on the whole $\mathbb{Z}^2$.

This results in a measure concentrated on spanning trees on $\mathbb{Z}^2$ with two ends, and it can be made translation-invariant by averaging over all $\mathbb{Z}^2$-shifts in B_n. Using Lemma 5.2 we have for every n:

$$\frac{1}{|B_n|} h_{\tilde{\mu_n}}(\tilde{B_n}) \leq h_{\tilde{\nu_n}} + O(n \log n) \leq h_{\text{top}} + O(n \log n). \tag{5.1}$$

But since (following a similar argument)

$$\lim_{n \to \infty} \frac{1}{|B_n|} h_{\mu_{\mathbb{Z}^2}}(B_n) = h_{\text{top}}, \tag{5.2}$$

we see that $\tilde{\mu}$ has a smaller entropy than $\mu_{\mathbb{Z}^2}$. □

4. Numerical calculation

We will now derive a formula for calculating the maximum entropy of the domino tiling process of $\mathbb{Z}^2$. To do this we will first calculate the maximum entropy of the spanning forest process on the integer lattice and translate this resut using the connections established in chapter 3.

DEFINITION. Define the adjacency function

$$R(x) = \begin{cases} 1, & \text{if x} \backsim 0 \\ 0, & \text{otherwise} \end{cases} \qquad \forall x \in \mathbb{Z}^2$$

and the incidence matrix

$$M(x,y) = \begin{cases} 1, & \text{if x} \backsim \text{y} \\ 0, & \text{otherwise,} \end{cases} \qquad \forall x, y \in \mathbb{Z}^2$$

so that $M(x,y) = R(y-x)$. For a subgraph $B \subseteq \mathbb{Z}^2$ let M_B denote the incidence Matrix of the induced subgraph. Let T^2 be the 2-dimensional Torus $\mathbb{R}^2/\mathbb{Z}^2$. An element

$\alpha = (\alpha_1,\ \alpha_2) \in T^2$ can then be written as pair of numbers in (0,1]. For $\alpha \in T^2$ define $Q : T^2 \to \mathbb{C}$ by the finite sum

$$Q(\alpha) := \frac{1}{4} \sum_{x \in \mathbb{Z}^2} \mathrm{e}^{2\pi i \alpha \cdot x} R(x).$$

THEOREM 5.5 (Calculation of measure of maximal entropy). *The topological entropy of the uniform spanning forest process is given by*

$$h_{ST} = \int_{T^2} \log\left(4\left(1 - Q\left(\alpha\right)\right)\right) \mathrm{d}\alpha$$

PROOF. We'll use the Matrix-Tree Theorem (Theorem 4.1) to compute N_{B_n}.

CLAIM. *Given $\alpha = (\alpha_1/n, \alpha_2/n) \in T^2$, $4Q(\alpha)$ is an eigenvalue of the matrix M_n with eigenvector $v(\alpha) \otimes \xi^\alpha$.*

This can easily be checked directly:

$$\begin{aligned} &\sum_{y \in \mathbb{Z}^2} M_n(x,y) v(\alpha) \exp(2\pi i \alpha \cdot y) \\ &\qquad = \sum_{y \in \mathbb{Z}^2} R(y-x) \exp(2\pi i \alpha \cdot (y-x)) v(\alpha) \exp(2\pi i \alpha \cdot x) \\ &\qquad = 4Q(\alpha) v(\alpha) \exp(2\pi i \alpha \cdot x). \end{aligned} \tag{5.3}$$

♢

We can now use those eigenvalues of M_n to calculate the number of spanning trees:

$$\begin{aligned} N_{B_n} &= \frac{1}{(2n+1)^2} \prod_{\alpha} (4 - 4Q(\alpha)) \\ &= \frac{1}{(2n+1)^2} \left(\prod_{\alpha \neq 0} 4 \prod_{\lambda} \left(1 - \lambda\ (\alpha)\right) \right) \left(\prod_{\lambda \neq 1} \left(1 - \lambda\left(0\right)\right) \right) \\ &= \frac{1}{(2n+1)^2} \left(\prod_{\alpha \neq 0} 1 - Q(\alpha) \right) \left(\prod_{\lambda \neq 1} \left(1 - \lambda\left(0\right)\right) \right). \end{aligned} \tag{5.4}$$

Putting this into the definition of the topological entropy and, we get

$$
\begin{aligned}
h_{ST} &= \lim_{n\to\infty} \frac{1}{(2n+1)^2}\log(N_{B_n}) \\
&= \lim_{n\to\infty} \sum_{\alpha\neq 0} \log(4(1-Q(\alpha)))\frac{1}{(2n+1)^2} \\
&= \int_{T^2} \log(4(1-Q(\alpha)))\mathrm{d}\alpha,
\end{aligned} \tag{5.5}
$$

where the last equalitiy follows by approximating the integral by Riemann sums. The second term in the number of spanning trees can be neglected since they are logarithmically insignificant. In this 2-dimensional case it is clear that the integral is finite and by bounded convergence the sum converges to the integral. For higher dimensions a similar formula holds, as discussed in Burton and Pemantle [**7**], a result that we will also go over in the last section of this chapter. □

As we have seen in chapter 3, spanning trees and domino tilings are intimately connected. In fact, we have shown that there is a bijection between domino tilings and directed spanning trees, which we'll call Φ. Writing Ψ for the map that takes a directed pair of spanning trees (T, T^*) and leaves just the undirected T we have the following correspondence:

$$\text{Domino Tilings} \overset{\Phi}{\leftrightarrow} \text{Directed STs} \overset{\Psi}{\rightarrow} \text{STs}.$$

Generally, the map Ψ is not one-to-one, but if T is a one-ended tree, so is T^*, so there is only one way to orient the edges. Let now G be a $\mathbb{Z}^2$-periodic planar graph. We then have a well defined map $\Phi^{-1} \circ \Psi^{-1}$ from one-ended spanning trees of $\mathbb{Z}^2$ to domino tilings of $\tilde{\mathbb{Z}^2} \simeq \mathbb{Z}^2$. We have seen in chapter 4 that the uniform spanning tree measure $\mu_{\mathbb{Z}^2}$ on $\mathbb{Z}^2$ is supported on the set of one-ended trees, which means this map gives us a transported measure $\nu_{\mathbb{Z}^2}$ on domino tilings of $\mathbb{Z}^2$.

THEOREM 5.6. *This measure $\nu_{\tilde{\mathbb{Z}^2}}$ is the unique measure of maximal entropy among all shift invariant measures on domino tilings. It's entropy per vertex is $h_{\mu_{\mathbb{Z}^2}}/4$.*

PROOF. From the discussion leading up to Theorem 5.6 it is clear that $\nu_{\tilde{\mathbb{Z}^2}}$ is well defined and shift invariant. It remains only to show it's property concerning entropy.

Let $\tilde{\mu}$ be any $\mathbb{Z}^2$-invariant probability measure on domino tilings of $\tilde{G}$. We can transport this to a measure μ on essential spanning trees on $\mathbb{Z}^2$ by $\mu(B) = \tilde{\mu}(\Phi^{-1}(\Psi^{-1}(B)))$. which has the same entropy.

But $h_{\tilde{\mu}} \leq h_{\mu} \leq h_{\mu_{\mathbb{Z}^2}} \leq h_{\nu_{\mathbb{Z}^2}}$ per fundamental domain with equality iff $\mu = \mu_{\mathbb{Z}^2}$. But since the uniform measure is concentrated on one-ended spanning trees, $\nu_{\mathbb{Z}^2}$ is uniquely determined, which ensures that it is in fact the unique measure of maximum entropy on domino tilings.

We derived the entropy formula for the spanning tree process on $\mathbb{Z}^2$ per vertex. To get from there to the domino process formula on $\mathbb{Z}^2$ we have to take into account the way the domino graph is constructed (see chapter 3). For every vertex in the original graph $G = \mathbb{Z}^2$ we get four vertices in $\tilde{G} = \mathbb{Z}^2$. To convert our entropy value for spanning trees to the entropy value for dominoes all that is left is dividing by 4. □

Equipped with this result, we can finally calculate the numeric value of the entropy of the domino tiling process of $\mathbb{Z}^2$. Starting with the definition of $Q(\alpha)$ we obtain:

$$\begin{aligned} 4Q(\alpha) &= \sum_{x \in \mathbb{Z}^2} \mathrm{e}^{2\pi i \alpha \cdot x} R(x) = \sum_{x \backsim 0} \mathrm{e}^{2\pi i(\alpha_1 x_1 + \alpha_2 x_2)} \\ &= 2(\cos(2\pi\alpha_1) + \cos(2\pi\alpha_2)). \end{aligned} \tag{5.6}$$

Using this we can calculate the entropy of the spanning tree process using the formula in theorem 5.1:

$$H_{ST} = \int_0^1 \int_0^1 \log(4 - 2(\cos(2\pi\alpha_1) + \cos(2\pi\alpha_2)))\mathrm{d}\alpha_1 \mathrm{d}\alpha_2 \approx 1.16624. \tag{5.7}$$

This entropy has caused a lot of discussion since it was first calculated by Burton and Pemantle in 1993 [**7**]. It turns out that this value is the same as the entropy of a completely different system calculated by Lind, Schmidt and Ward in 1990 [**19**]. This fact remained a mystery until Solomyak showed in 1997 [**33**] that this is no coincidence but that the Matrix-Tree Theorem is responsible.

There is also a connection between this entropy and Catalan's constant G. We have:

$$H_{ST} = 4G/\pi,$$

where $G = \sum_{k=0}^{\infty}(-1)^k/(2k+1)^2 \approx 0.91596$ is Catalan's constant. For more information on this connection see Lyons [**21**], Shrock and Wu [**31**] and the references therein.

It follows then from theorem 5.6 that the entropy of the domino tiling process (on $\mathbb{Z}^2$) is a quarter of this value:

$$H_{DT} = G/\pi = \frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^2} \approx 0.29156.$$

As was already mentioned, this was first calculated by Kastelyn in 1961. He could not, however, prove that it is in fact the largest entropy possible. It took almost half a century to find the proof given here and until now we know of no possibility of showing this result without taking the "detour" through the field of spanning trees.

5. Higher dimensions

Up until now we almost exclusively concentrated on the two-dimensional case. This is historically the first situation studied in this context as well as one of the few cases in which a closed formula for the entropy value is known. However, the notion that the limit of the uniform spanning tree measure maximizes entropy can be, and has been made in a much more general setting.

The first problem arises when considering the limit of the uniform spanning tree measure. As we discussed in chapter 4, the limit can be taken with free (FSF) or wired (WSF) boundary conditions. In the case of $\mathbb{Z}^2$ they coincide and give a single tree. In fact, they coincide for $\mathbb{Z}^d$ for all $d \geq 2$ but only for $d \leq 4$ this limiting measure is itself a tree with only one topological end. For higher dimensions it can be shown that it is concentrated on spanning forests with infinitely many components, all of which have either one or two topological ends [**25**].

But also in this much more complicated case it can be shown that the uniform spanning tree measure converges to a translation-invariant measure (on essential spanning forests) for which the specific entropy is maximal [**30**]. Burton and Pemantle [**7**] found a formula for this limit measure:

$$H = \int_{T^d} \log(D^k\chi(Q(\alpha))(1))\mathrm{d}\alpha,$$

where $T^d = \mathbb{R}^d/\mathbb{Z}^d$ is the d-dimensional torus, D is the maximal degree of the vertices, the matrix $Q(\alpha)$ is the (finite) sum $D^{-1}\sum_{x\in\mathbb{Z}^d}\exp(2\pi i\alpha x)R(x)$, and $\chi(Q(\alpha))$ is the characteristic polynomial of $Q(\alpha)$. $R(x)$ is, as in our case, the adjacency function.

For higher dimensions no entropy values are known in simple terms of other known constants or functions (as is the case with $d = 2$ and the connection with Catalan's constant). There is also the problem that it is no trivial task to evaluate the integral in higher dimensions accurately. For more information on this interesting investigation we refer to Shrock and Wu (2000) [**31**] and Felker and Lyons (2003) [**10**].

Appendix

Abstract

Consider the nearest neighbor graph for the integer lattice $\mathbb{Z}^2$ and a random domino tiling, or perfect matching, thereon. It is shown that there is a measure on such domino tilings that maximizes the measure-theoretic entropy and is unique with this property and it's numerical value is calculated.

Results are discussed and proved concerning dynamical systems, ergodic theory, spanning trees, electrical networks, graph theoretic stochastic processes and measure theory.

Betrachte den Graph auf dem Gitter $\mathbb{Z}^2$, wobei die nächsten Nachbarn verbunden sind, und eine Domino Abdeckung darauf. In dieser Arbeit wird gezeigt dass dafür ein eindeutig bestimmtes Maß maximaler maßtheoretischer Entropie existiert und dessen numerischer Wert wird berechnet.

Ergebnisse die dabei diskutiert und bewiesen werden reichen aus den Gebieten dynamischer Systeme, Ergodentheorie, spanning Trees, elektrische Netzwerke und graphtheoretischer stochastischer Prozesse zu Maßtheorie.

Curriculum vitæ

CONTACT Kettenbrückengasse 23/2/4 *Phone:* (+43) 650 822 26 25
1050 Vienna *Email:* philipp.g.deutsch@gmail.com

PERSONAL DATA
- Born on January 8th, 1985 in Vienna
- Nationality: Austrian

EDUCATION

University of Vienna, Vienna

Mathematics (Diplom)
- August 2008 - February 2009: Exchange semester at UPMC, Paris VI, France
- January 2008: Completed undergraduate studies
- March 2005: Beginning of studies

Philosophy
- November 2004: Dropout
- March 2004: Beginning of studies

University of Technology, Vienna

Physics (Bachelor studies)
- February 2004: Beginning of studies

Piaristengymnasium, Vienna VIII
- June 2003: Final examination with honours
- School year 2002/2003: Vice student body president
- Sept. - Dec. 2000: Exchange semester at Frensham Heights, Surrey, UK

WORK EXPERIENCE
- Since Sept. 2009: Trainee at Erste Group Bank AG
- June 2007 - July 2008: Freelance worker ar Austrian Research Centers (ARC)

Bibliography

[1] M. Aigner and G. Ziegler. *Proofs from The Book*. Springer-Verlag Berlin Heidelberg New York, 3rd edition, 2004.

[2] D. J. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM Journal of Discrete Mathematics*, 3:450–465, 1990.

[3] I. Benjamini, R. Lyons, Y. Peres, and O. Schramm. Uniform spanning forests. *The Annals of Probability*, 29(1):1–65, 2001.

[4] K. Beuth and O. Beuth. *Elementare Elektronik*. Vogel Buchverlag, 2003.

[5] A. Broder. Generating random spanning trees. *30th Annual Symposium of Foundations of Computer Science*, pages 442–447, 1989.

[6] R. Burton and M. Keane. Density and uniqueness in percolation. *Communications in mathematical physics*, 121:501–505, Jan 1989.

[7] R. Burton and R. Pemantle. Local characteristics, entropy and limit theorems for spanning trees and domino tilings via transfer-inpedances. *The Annals of Probability*, pages 1329–1371, 1993.

[8] J. Doob. *Measure Theory*. Springer-Verlag Berlin Heidelberg New York, 1994.

[9] P. G. Doyle and J. L. Snell. *Random Walks and Electric Networks*. Mathematical Association of America, Washington, DC, 1984.

[10] J. L. Felker and R. Lyons. High-precision entropy values for spanning trees in lattices. *Journal of Physics A: Mathematical and General*, 36(31), Apr 2003.

[11] P. Grünwald and A. P. Dawid. Game theory, maximum entropy, minimum discrepancy and robust baysian decision theory. *Ann. Statist.*, 32(4):1367–1433, Aug 2004.

[12] P. Harremoes and F. Topsøe. Maximum entropy fundamentals. *Entropy*, 3:191–226, Oct 2001.

[13] P. W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27:1209–1225, 1961.

[14] A. Katok and B. Hasselblatt. *Introduction to the modern Theory of Dynamical Systems*. Cambridge University Press, 1995.

[15] G. Keller. *Equilibrium States in Ergodic Theory*. Cambridge University Press, 1998.

[16] R. Kenyon, J. Propp, and D. B. Wilson. Trees and matchings. *The Electronic Journal of Combinatorics*, 7, 2000.

[17] A. Khinchine. *Mathematical Foundations of Information Theory*. Dover Publications, 1957.

[18] G. Kirchhoff. Über die auflösung der gleichungen, auf welche man bei der untersuchung der linearen vertheilung galvanischer ströme geführt wird. *Annalen der Physik und Chemie*, 72(12):497–508, Jul 1847.

[19] D. Lind, K. Schmidt, and T. Ward. Mahler measure and entropy for commuting automorphisms of compact groups. *Inventiones Mathematicae*, 101:593–629, 1990.

[20] R. Lyons. A bird's-eye view of uniform spanning trees and forests. *Microsureys in Discrete Probability*, pages 135–162, 1998.

[21] R. Lyons. Asymptotic enumeration of spanning trees. *Combinator. Probab. Comp.*, 14(4):491–522, 2005.

[22] J. C. Maxwell. *A treatise on electricity and magnetism*, volume 1. Oxford: Clarendon Press, 1873.

[23] M. Misiurewicz. A short proof of the variational principle for a $\mathbb{Z}_+^n$ action on a compact space. *Astérisque*, 40:147–187, 1976.

[24] E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, 1993.

[25] R. Pemantle. Choosing a spanning tree for the integer lattice uniformly. *The Annals of Probability*, 19:1559–1574, 1991.

[26] M. Pollicott and M. Yuri. *Dynamical Systems and Ergodic Theory*. Cambridge University Press, 1998.

[27] S. M. Ross. *Introduction to Probability Models*. Academic Press, 9th edition, 2007.

[28] J. Rudnick and G. Gaspari. *Elements of the Random Walk: An Introduction for Advanced Students and Researchers*. Cambridge University Press, 2004.

[29] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

[30] S. Sheffield. Uniqueness of maximal entropy measure on essential spanning forests. *The Annals of Probability*, 34(3):857–864, 2006.

[31] R. Shrock and F. Wu. Spanning trees on graphs and lattices in d dimensions. *Journal of Physics A: Mathematical and General*, 33:3881–3902, 2000.

[32] C. Silva. *Invitation to Ergodic Theory*. American Mathematical Society, 2008.

[33] R. Solomyak. On coincidence of entropies for two classes of dynamical systems. *Ergodic Theory and Dynamical Systems*, 18:731–738, 1998.

[34] S. Talyor. *Introduction to Measure and Integration*. Cambridge University Press, 1966.

[35] P. Walters. *An Introduction to Ergodic Theory*. Springer-Verlag Berlin Heidelberg New York, 1982.

[36] D. B. Wilson. Generating random spanning trees more quickly than the cover time. *ACM Symposium on Theory of Computing*, pages 296–303, 1996.

SACHERSCHLIESSUNG

der

Fachbereichsbibliothek Mathematik, Statistik, Informatik der Universität Wien

Klassifikation

Aufstellungsort: ☐ HoA ☐ HoD

MSC2010: - -

CCS98: - -

ZDM88: - -

BK: - -

Schlagwortketten nach RSWK

Kontrollvermerk der Fachbibliothek:

_____ _____

Datum Unterschrift