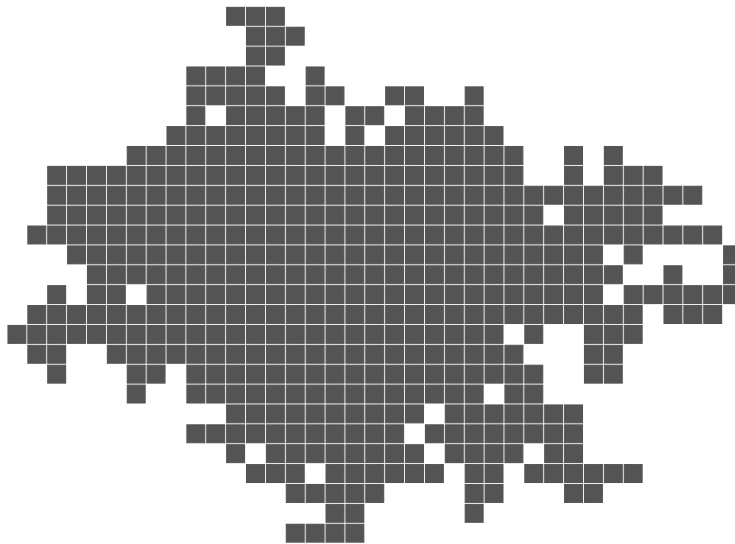# CIMAT

Centro de Investigación en Matemáticas

Doctoral Thesis

Topological, geometric and combinatorial properties of random polyominoes

Author:
Érika B. Roldán Roa

Supervisors:
Dr. Matthew Kahle
(OSU)
Dr. Víctor Pérez-Abreu C.
(CIMAT)

Declaration of Authorship

I, Érika B. ROLDÁN ROA, declare that this thesis titled, "Topological, geometric and combinatorial properties of random polyominoes" and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this Institute.

- Where I have consulted the published work of others, this is always clearly attributed.

- In particular, I indicate in any result that is not mine, or in joint work with my advisors and collaborators, the names of the authors and where it has been originally published (if I know it). The results that are not mine, or in joint work with my advisors and collaborators, are numerated with letters.

- All the images and figures contained in this work have been generated by me, except for Figure 1 and the content of Table 2.2.2.

- The algorithms and the code written for implementing them that were used for the figure constructions, computational experiments, and random simulations are my own work.

- Just in case God(s) exists, my parents and grandparents have always told me that I should acknowledge that She is the real author of everything created in this Universe.

Signed:

_____

Date:

_____

"Toys are not really as innocent as they look. Toys and games are precursors to serious ideas."
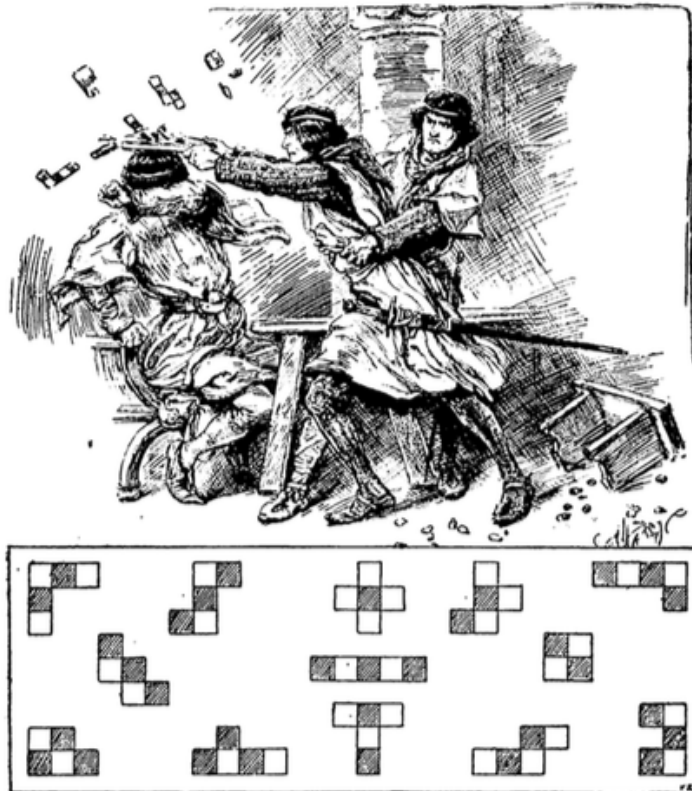
Charles Eames

FIGURE 1: Our story begins in 1907 with the 74th Canterbury puzzle [8]. This puzzle shows a chessboard broken into thirteen pieces (polyominoes). Twelve of them have five squares (pentominoes) and one of them has four squares (tetromino). The puzzle is solved by reconstructing the 8×8 chessboard using these thirteen pieces.

Acknowledgements

I would like to thank Víctor Pérez-Abreu (CIMAT) and Matthew Kahle (OSU) for being the best possible thesis advisors I could ask for. They have guided me and supported me with wisdom, patience, and kindness throughout my PhD studies.

Also, I am grateful to my Thesis Committee: Víctor Rivero, Antonio Rieser, Noé Torres, Hannah Alpert, Matthew Kahle and Víctor Pérez-Abreu, for their valuable comments and suggestions that enriched this thesis. English is not my native language, but I received a lot of help with the language edition (grammar, syntaxis, spelling, and some logic mistakes) of this thesis and other related documents. I would like to thank for these valuable comments and suggestions to: Benjamin Schweinhart, Celia Kahle, Kristina Raave, Hannah Alpert, Katherin Ritchey, Misty Adams, Nicholas Barrett, and my Thesis Committee. To CIMAT, and in particular to the Probability and Statistics Department, I thank all the academic and administrative support that they gave me during my PhD studies.

I was a visitor scholar during the 2017-2018 scholastic year at The Ohio State University; I thank to this institution, the hospitality and the academic richness that they provided me during this visit. In particular, I thank the TGDA interdepartmental research group at OSU. During this year I was supported by NSF DMS-1352386, CIMAT, CONACYT, and FORDECYT 265667 (Proyecto: Programa para un avance global e integrado de la matemática mexicana).

At OSU, in Columbus, and Providence, I spent a wonderful time with and received a lot of support from my friends, academic family, and research group pals: Andrew Newman, Benjamin Schweinhart, Fedor Manin, Gregory Malen, Hannah Alpert, Katherin Ritchey, Kyle Parsons, Jessica Zehel, and Matthew Kahle. I met a lot of wonderful people who are too numerous to mention here, but that were an important support when things were difficult and also for enjoying Columbus in a non-academic way. Finally, I could not thank enough all the kind support that I have received from Maritza Sirvent, James Carlson, and Dylan Carlson; they were my family in Columbus.

I thank ICERM (Institute for Computational and Experimental Research in Mathematics) for hosting me as a visitor scholar during the wonderful 2016's Autumn semester: Topology in Motion, and I am grateful to Matthew Kahle for encouraging me to attend this program. I am grateful to Ileana Streinu and Kavita Ramanan, who were my Mentors at ICERM, for the conversations, guidance and support that they gave me during this research visit. That semester at ICERM, I did part of the research work contained in this thesis, I meet and had inspiring math conversations and discussions with a lot of great mathematicians, and I made a lot of friends whose friendship (I am sure) will enrich the rest of my life.

Last, but not least, I would like to thank Isa, Isa's family (to her parents, Lupita, Juan Pablo, and Silvio) and my whole family (grandparents, ants, uncles, cousins and new generations, Nicholas Barrett and Edith Jimenez) for their love and support during these four years. And, of course, I thank Engracia Castro for keeping me sane enough to finish this thesis.

Contents

x

Dedicated to Isa: My Star... All my Stars.
To my Grandparents and to my dearest friend Nick.

Chapter 1

Introduction

Random shapes have served as important models for a large class of natural phenomena in a wide range of natural science disciplines, including meteorology, statistical mechanics, biology, chemistry, and astronomy.

We study two models of random shapes in this thesis — the Eden Cell Growth Model (EGM) and uniform and percolation distributed polyominoes (also known as lattice-based animals). These models have long been of interest in mathematical physics, probability, and statistical mechanics [11, 15].

Both structures have interesting topological, combinatorial, and geometrical properties. However, until now, tools from stochastic topology and topological data analysis have not been used to study these properties. Here, we present the first studies of EGM and polyominoes using these methods.

This thesis contains three original sets of contributions that are outlined in the next three sections, where we also describe polyominoes and the EGM, and where we state and provide context for our original results by surveying the existing mathematical literature.

1.1 Polyominoes with maximally many holes

In 1954, Solomon W. Golomb [13] defined an n-omino (which is a polyomino with area of n) as a rook-connected subset of n squares of the infinite checkerboard.

Even though polyominoes have been studied by the mathematical community for more than 60 years, it is still unknown how to find a formula for computing the exact number of polyominoes with a fixed number of squares.

Murray Eden was the first to propose and study the polyomino enumeration problem from an asymptotic combinatorics point of view [10]. In that work, Eden gives the first upper and lower bounds of the number of polyominoes with area n, represented as a_n

$$(3.14)^n \leq a_n \leq \left(\frac{27}{4}\right)^n. \tag{1.1}$$

Eden [10] also pointed out the similarity between polyominoes and Random Walks (RW), Self Avoiding Walks (SAW), and percolation processes on the regular square lattice on the plane. Because of these similarities, polyominoes have been used to study problems in crystallography, solid state, statistical mechanics, and

chemistry [15].

In 1965, David A. Klarner [20] improved the lower bound of the number of poly-ominoes of area n to $(3.20)^n \leq a_n$. In 1967, Klarner [19] proved that the sequence $\{a_n\}_{n=1}^{\infty}$ satisfies the following limit

$$\lim_{n \to \infty} (a_n)^{\frac{1}{n}} = \lambda. \tag{1.2}$$

In the proof of this limit he used the following important submultiplicativity in-equality

$$a_n \cdot a_m \leq a_{n+m}.$$

This inequality can be proved by a concatenation argument. In that paper, Klarner also improved the lower bound of the number of n-ominoes to $(3.73)^n \leq a_n$ for $n \gg 1$. After Klarner's papers the constant λ has been called the Klarner's constant.

The last lower bound improvement of the Klarner's constant, published in 2006 [3], is $3.980137 \leq \lambda$; the last improvement of the upper bound of Klarner's constant, published in 2015 [2], is $\lambda \leq 4.5685$. It is believed that $a_n \approx C\lambda^n n^{\theta}$ where $\theta = -1$ and $\lambda = 4.0625696$ but this has not been proved yet.

Although polyominoes with less than seven tiles do not have holes, we show in Chapter 3 that polyominoes with holes grow exponentially faster than polyominoes without holes. Thus, it is important to study polyominoes that have holes.

The first original contributions of this thesis, that we present in Chapter 2, are about the extremal combinatorial behavior of the maximum number of holes that a polyomino can have.

We prove a tight bound for the asymptotic behavior of the maximal number of holes of an n-omino as $n \to \infty$. We also give an exact formula for this quantity for an infinite sequence of natural numbers.

Denote the maximum number of holes that a polyomino can have by $f(n)$. In Theorem 1 (that we prove in Section 2.4) we show that if

$$n_k = \left(2^{2k+1} + 3 \times 2^{k+1} + 4\right)/3 \quad \text{and} \quad h_k = \left(2^{2k} - 1\right)/3,$$

then $f(n_k) = h_k$ for every $k \geq 1$.

In Section 2.5 we prove Theorem 2 that gives general bounds which hold for all n

$$\frac{1}{2}n - \sqrt{\frac{5}{2}n} + o(\sqrt{n}) \leq f(n) \leq \frac{1}{2}n - \sqrt{\frac{3}{2}n} + o(\sqrt{n}),$$

for large enough n.

1.2   Topology of random polyominoes

Percolation theory models on the square regular lattice are relevant in various areas of theoretical and experimental physics like material science [30], thermodynamics

[6], and statistical mechanics [4].

This gives relevance to the study of the topological features, such as the homology groups, of the finite clusters of the percolation model on the regular square lattice of the plane. These finite clusters correspond to polyominoes.

In Chapter 3 we present our second set of contributions of this thesis about the growth rate of the expectation of the number of holes in a polyomino with uniform and percolation distributions. We prove the existence of linear bounds for the expected number of holes of an n-omino with respect to both the uniform and percolation distributions. Furthermore, we exhibit particular constants for the upper and lower bounds in the uniform distribution case.

We prove in Theorem 3 that with the uniform distribution defined on the set of all n-ominoes, there exist constants C_1 and C_2 (not depending on n) such that

$$C_1 \cdot n \leq \mathbb{E}[\beta_1] \leq C_2 \cdot n, \tag{1.3}$$

for sufficiently large values of n. We also give exact values for C_1 and C_2.

For percolation distributions, we prove in Theorem 5 that for any $p \in (0,1)$, with the percolation distribution π_p defined on the set of polyominoes, there exist constants C_1 and C_2, not depending on n, but depending on p, such that

$$C_1 \cdot n \leq \mathbb{E}[\beta_1] \leq C_2 \cdot n.$$

As far as we know, this is the first time that these problems are studied for uniform and percolation distributed polyominoes.

In Appendix A, we describe Markov Chain Monte Carlo algorithms that we have implemented to sample random polyominoes.

1.3 The Eden Cell Growth Model

In 1961, Eden [10] introduced a two dimensional cell growth model based on polyominoes. He generated some random polyominoes under a cell growth process with n tiles for particular values of n with $n \leq 2^{15}$. In Figure 1.1 we show a cell colony with $10,000$ sites that we have simulated with the EGM algorithm.

The cell growth model can be described as follows [9, 10, 16]: it starts with a polyomino consisting of a single cell and it grows by adding one cell at a time uniformly and randomly with the restriction that the newly added cell needs to share at least one side with any other cell already present in the polyomino. Murray Eden has mentioned [10] that the first person to model the cell growth process from a mathematical point of view was Alan M.Turing [31] in 1952. However, Turing's cell growth model is not based on polyominoes; he used a one-dimensional structure to study the cell growth problem.

The EGM corresponds to a site First Passage Percolation model in the regular square lattice on the plane after applying the right time-change [1]. As far as we know, the evolution of the topology of the EGM (measured by its homology) has not
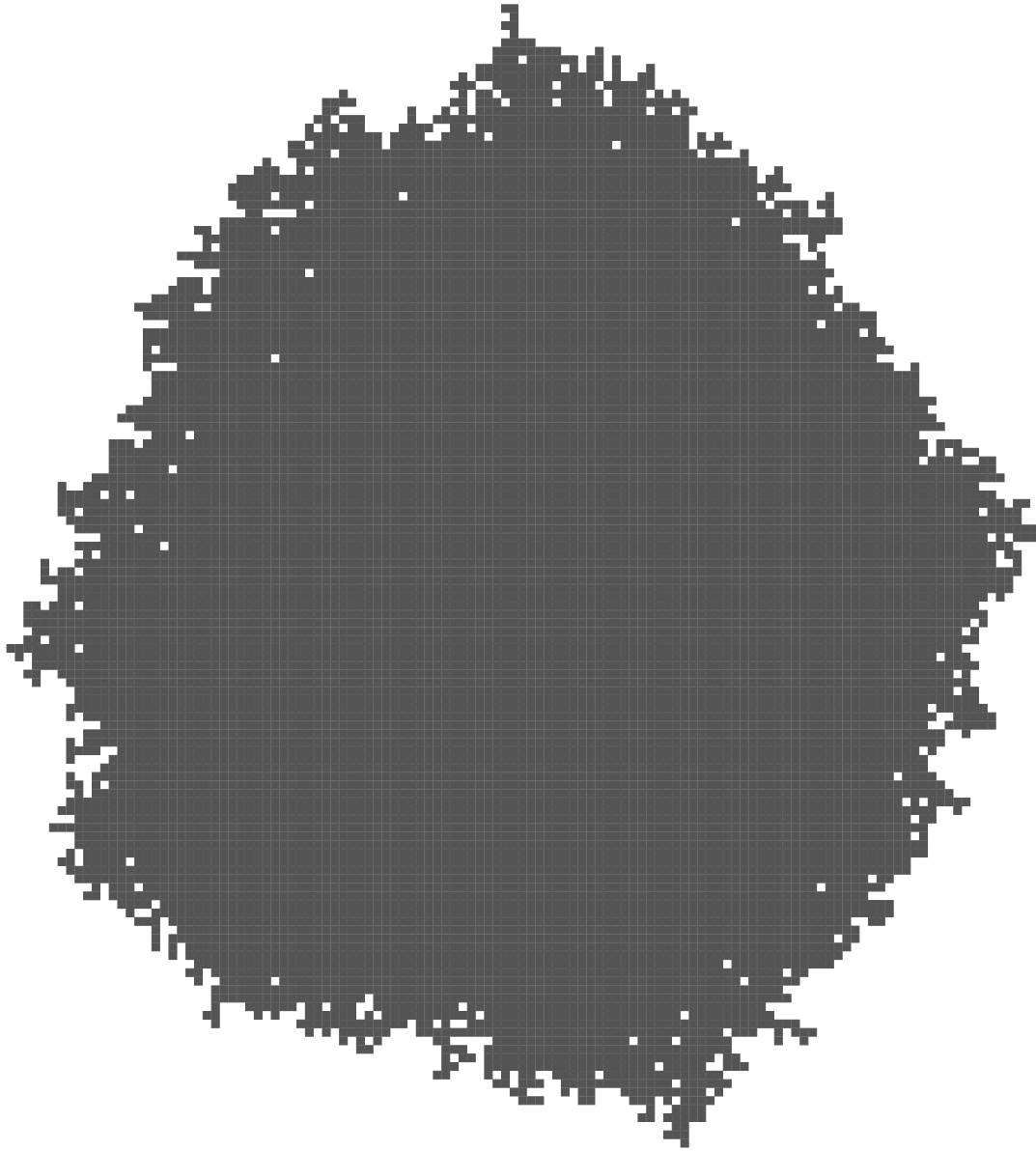
FIGURE 1.1: Eden Cell Growth Model with 10,000 tiles.

yet been studied and the same is true for any of the First Passage Percolation models.

This gives importance to our third set of contributions that we present in Chapter 4, which consists of importing new techniques from stochastic topology and topological data analysis to study the topological evolution of First Passage Percolation models. In particular, we are the first to study the persistent homology associated to the evolution of the homology of the EGM. Even more, we are the first to study the persistent homology associated to First Passage Percolation models.

In Theorem 7 we characterize the change in time of the rank of the first homology group on the stochastic process defined by the EGM. This allows us to design and implement a new algorithm that computes the persistent homology associated to this stochastic process at each time and that keeps track of geometric features of

the homology of the process like the area and location of the holes. Also, our algorithm keeps track of the persistent homology splitting tree—see [26] as a reference for splitting trees.

In Section 4.5, we present and analyze the results of the computational experiments that we have made with this algorithm. One of the conjectures that we have, based on these experiments, is about the asymptotic behavior of the number of holes. We present this conjecture in what follows.

Let β_t be the random variable that measures the rank of the first homology group of the EGM stochastic process at time t (the number of holes at time t), then for sufficiently large values of t

$$C_1 \sqrt[\alpha]{t} \leq \mathbb{E}[\beta_t] \leq C_2 \sqrt[\alpha]{t}, \tag{1.4}$$

for some constants $C_1, C_2 > 0$ and $\alpha \geq \frac{1}{2}$. We suspect that $\alpha = 0.5$, $C_1 > 1$, and $C_2 < 1.5$.

In Appendix B, we present more results of simulation experiments described in Section 4.5.6.

Chapter 2

Maximizing the number of holes in polyominoes

Denote the maximum number of holes that a polyomino with n tiles can have by $f(n)$. Our first contribution of the thesis is finding the exact value of $f(n)$ for an infinite sequence of values of n and giving tight bounds for the asymptotic behavior of $f(n)$ when n tends to infinity.

In particular, in Theorem 1 (that we prove in Section 2.4) we show that if

$$n_k = \left(2^{2k+1} + 3 \times 2^{k+1} + 4\right)/3 \quad \text{and} \quad h_k = \left(2^{2k} - 1\right)/3,$$

then $f(n_k) = h_k$ for every $k \geq 1$. For proving this result, we construct a sequence of polyominoes that has n_k tiles and h_k holes. To our surprise and delight this sequence of polyominoes has interesting geometric properties that we study in Section 2.6.

We also give general bounds which hold for all n. In Theorem 2 we prove that

$$\frac{1}{2}n - \sqrt{\frac{5}{2}n} + o(\sqrt{n}) \leq f(n) \leq \frac{1}{2}n - \sqrt{\frac{3}{2}n} + o(\sqrt{n}),$$

for large enough n.

Mathematicians have been studying properties of polyominoes for more than 60 years [15]. As far as we know, this is the first time that the asymptotic behavior of $f(n)$ has been studied. Also, exact values for $f(n)$ were only known for $1 \leq n \leq 28$ [27]. These values of f are contained in Table 2.2.2.

2.1 We have been playing with polyominoes since childhood

In this section we introduce the concept of polyominoes in an informal setting, and we explain why we are interested in studying polyominoes with holes.

Polyominoes can be made by gluing together, edge-to-edge, finitely many, non-overlapping unit squares on the plane. The simplest polyomino that can be constructed (and that begins to fascinate us already as mere toddlers) is the polyomino with only one square: the monomino.

But the most popular polyominoes are dominoes and tetrominoes. Usually, one can buy a set of double-six (28 tiles) dominoes in any department store. Tetrominoes are the main characters in the world-famous Tetris video game —see Figure 2.1.
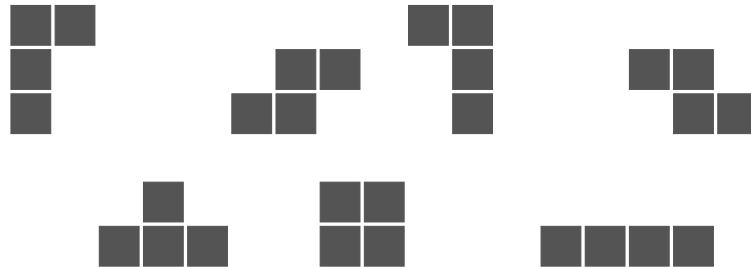
FIGURE 2.1: This figure shows the seven tetrominoes that comprise the video game Tetris. This game was originally released in the Soviet Union in 1984. Nintendo introduced it to the rest of the world in 1989.

With five squares we get the twelve pentominoes that are depicted in Figure 1. As the description of the figure explains, these twelve pentominoes, in addition to the square shaped tetromino, can tessellate an 8×8 chessboard.

There are 35 hexominoes and 108 heptominoes (under some equivalent reflection and rotation relationships that we will make precise in the next section). As far as I am aware, humanity has only computed the exact number of polyominoes with no more than 56 squares. Currently, there is no known formula for the number of polyominoes with any given number of squares. Even more, there is not an agreement on what it means for an enumerative combinatorial function to have a formula [24].

Any puzzle or problem that asks about covering a certain subset of the infinite checkerboard with a set of polyominoes is a tiling problem. Tiling problems involving polyominoes are well studied—see, for example, [14] or [28]. In tiling problems, one almost always restricts to simply-connected polyominoes, i.e., polyominoes without holes. But as we mentioned in Section 1.1, if $n > 6$ then n-ominoes with holes exist—see Figure 2.2.

FIGURE 2.2: A heptomino and an octomino with one hole.

Moreover, as the number of squares gets bigger and bigger, polyominoes with holes begin to outnumber polyominoes without holes. In Chapter 3, we prove that most polyominoes have holes as the number of polyominoes with holes grows exponentially faster than the number of polyominoes without holes. This is one of the main reasons why studying polyominoes with holes, in particular polyominoes with maximally many holes, matters.

2.2 Polyominoes, holes, and statement of main results

In this section we define polyominoes and holes in polyominoes. We also specify in detail what is known about $f(n)$.

2.2.1 Polyominoes

In 1954, S. W. Golomb [13] defined an n-omino (which is a polyomino with area of n) as a rook-connected subset of n squares of the infinite checkerboard. This definition implies that a polyomino is the union of a non-empty, finite subset of unitary square tiles with a connected interior of the plane's regular unitary square tiling. We also require polyominoes to be closed subsets of $\mathbb{R}^2$; thus, a polyomino contains the edges and vertices of its unitary squares.

Based on this definition of a polyomino, if we want to make sense of the set of polyominoes with a given number of tiles and its cardinality, we need first to establish when two n-ominoes will be considered equivalent. The three most common ways to define this equivalence relationship are known as: free, one-sided, and fixed polyominoes.

- Two free polyominoes are considered equivalent if they are congruent after performing any required translations, reflections, and/or rotations. For example, there is only one free domino, two free triominoes, five free tetrominoes, and twelve free pentominoes. Free pentominoes are depicted in Figure 1.

- Two one-sided polyominoes are considered equivalent if they are congruent after performing any required translations and/or rotations. For example, there is only one one-sided domino, two one-sided triominoes, seven one-sided tetrominoes, and 18 one-sided pentominoes. One-sided tetrominoes are depicted in Figure 2.1.

- Two fixed polyominoes are considered equivalent if they are congruent after translation. For example, there are two one-sided dominoes, six one-sided triominoes, 19 one-sided tetrominoes, and 63 one-sided pentominoes.

Counting polyominoes defined with one of these equivalent relationships gives bounds for counting polyominoes with the other two equivalent relationships. These bounds depend on the symmetry of the plane's regular square lattice:

$$\text{fixed} \leq 2\,(\text{one-sided}) \leq 8\,(\text{free}).$$

Because the number of holes in a polyomino is invariant under rotations, translations, and reflections, unless we explicitly mention it, the results that we prove in this chapter are true for polyominoes with any one of these three equivalent relationships. We will work with free polyominoes for the rest of the chapter. We denote the set of all free n-ominoes by $\mathcal{A}_n$. In Chapter 3 and Chapter 4, we specify other geometric properties of elements in $\mathcal{A}_n$ that depend on their positions in the plane. Although for now, we do not take into consideration the geometric position of the polyominoes on the plane.

2.2.2 Counting holes and the statement of main results

As we have mentioned before, our main interest in this chapter is in maximizing the number of holes that a polyomino can have. Figure 2.3 illustrates an 8-omino, a 20-omino, and a 60-omino with 1, 5, and 21 holes, respectively. We will show in Section 2.4 that these are polyominoes with maximally many holes.

To be precise about the topology, remember that we consider the tiles to be closed unit squares on the plane. Polyominoes are finite unions of these closed squares, so they are compact. The holes of a polyomino are defined to be the bounded connected components of its complement on the plane.

In Chapter 3, when given a polyomino, we will construct a corresponding CW cubical complex such that the number of holes would be equal to the rank of the first homological group of this cubical complex; for the purpose of this chapter, the definitions and notions of polyominoes and holes given above are enough.
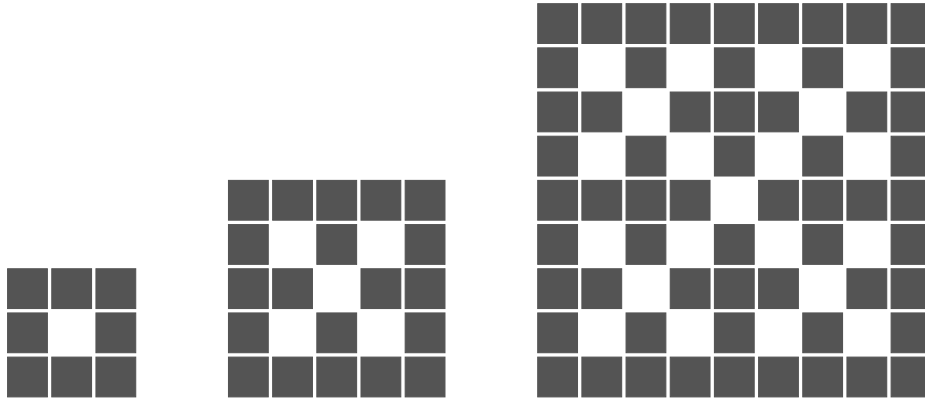
FIGURE 2.3: Polyominoes with maximally many holes.

Definition 1. *Given a polyomino* A, *we denote the number of holes in* A *by* $h(A)$. *For* $n \geq 1$ *we define*

$$f(n) = \max_{A \in \mathcal{A}_n} h(A). \tag{2.1}$$

Similarly, we define $g(m)$ to be the minimum number N such that there exists an N-omino with m holes. Observe that f and g are positive functions and we intuit that they are non-decreasing. We check below that g is a right inverse of f.

In Table 2.2.2, we present the computational results obtained by Tomás Oliveira e Silva [27] on enumerating free polyominoes according to area and number of holes up to $n = 28$. At the time of this writing, his computations were state of the art. As a corollary of his calculations, we know the value of $f(n)$ for $n \leq 28$ and of $g(m)$ for $m \leq 8$.

The function g is listed on the website of The On-Line Encyclopedia Of Integer Sequence as sequence A118797. This webpage lists the values of $g(m)$ for $m \leq 8$, and states that the best upper bound known for $g(m)$ is $g(m) \leq 3m + 5$.

We prove in this thesis a significant improvement for this upper bound as a corollary of Theorem 2. For large enough values of m

$$g(m) \leq 2m + C\sqrt{m} + c,$$

where C and c are positive fixed constants not depending on m.

Also in Theorem 2, we give the exact value of the functions $g(m)$ and $f(n)$ for an infinite sequence of natural numbers. We state these results in detail in what follows.

Here and throughout

$$n_k = \frac{1}{3}\left(2^{2k+1} + 3 \cdot 2^{k+1} + 4\right), \tag{2.2}$$

and

$$h_k = \frac{1}{3}\left(2^{2k} - 1\right). \tag{2.3}$$

We describe in Section 2.4 a construction of a sequence of polyominoes with n_k tiles and h_k holes. From this sequence we generate other sequences of polyominoes with $n_k - i$ tiles for $i \in \{1, 2, 3\}$ for which we find the values of $f(n_k - i)$.

Theorem 1. *For $k \geq 1$ we have $f(n_k) = h_k$. Moreover, $f(n_k - 1) = h_k$, and $f(n_k - 2) = h_k - 1$.*

As a consequence of this theorem, we get the exact values of $g(m)$ for infinite different values of m. This is proved using Theorem 1 and the properties of f and g that we study in Section 2.3.

Corollary 1. $g(h_k) = n_k - 1$ *for all $k \in \mathbb{N}$.*

We also establish the first and second order asymptotic behavior of f when we give tight bounds for $f(n)$ for large enough values of n.

Theorem 2. *Let $f(n)$ denote the maximum number of holes that a polyomino with n squares can have. Given $C_1 > \sqrt{5/2}$ and $C_2 < \sqrt{3/2}$, there exists an $n_0 = n_0(C_1, C_2)$ such that*

$$\frac{1}{2}n - C_1\sqrt{n} \leq f(n) \leq \frac{1}{2}n - C_2\sqrt{n},$$

for $n > n_0$.

Squares	No holes	1 hole	2 holes	3 holes	4 holes	5 holes	6 holes	7 holes	8 holes
1	1								
2	1								
3	2								
4	5								
5	12								
6	35								
7	107	1							
8	363	6							
9	1248	37							
10	4460	195							
11	16094	975	4						
12	58937	4622	41						
13	217117	21128	346						
14	805475	94109	2384	3					
15	3001127	410820	14560	69					
16	11230003	1766591	81863	798					
17	42161529	7506179	433172	7021	8				
18	158781106	31596240	2192752	51775	179				
19	599563893	131991619	10726252	339958	2509	1			
20	2269506062	547992183	51094203	2053872	25609	21			
21	8609442688	2263477612	238259280	11665593	214932	573			
22	32725637373	9309386178	1092053068	63192945	1578984	9140			
23	124621833354	38150082057	4934652298	329798278	10536260	105417	64		
24	475368834568	155859235424	22034767837	1670466031	65411645	982767	2131		
25	1816103345752	635067478628	97407519119	8256762365	383981499	7919375	38022	4	
26	6948228104703	2581737704039	426916828181	39992202198	2156114468	57387444	480713	329	
27	26618671505989	10474587325120	1857253575577	190431003084	11678888362	383757344	4872679	10332	
28	102102788362303	42422970467980	8027749130623	893726550231	61413242603	2410292366	42360239	188221	37

Table 2.2.2: Enumeration of n-ominoes by holes for $1 \leq n \leq 28$ [27].

Also, the sequence of polyominoes constructed in the proof of Theorem 1 allows us to show that the constant $\sqrt{3/2}$ involved in the upper bound of f given in Theorem 2 cannot be improved.

2.3 Preliminary results

In this section we prove all the preliminary results and define the geometric concepts that we need in order to prove Theorem 1 and Theorem 2.

We use a result proved by F. Harary and H. Harborth [17], stated in Theorem A below, that provides a formula for calculating the minimum possible edge perimeter that a polyomino with a given area can have. Analogous results to Theorem A are known for the triangular and the hexagonal regular lattices of the plane and for the cubical lattice in the space. Thus, all the preliminary results and theorems that we prove in this chapter can be derived for these other regular lattices and their respective polyforms.

We begin by analyzing the properties of the functions f and g by establishing how they relate to one another.

2.3.1 Properties of f and g

By definition, f is a step function with values in the natural numbers. In Figure 2.4, we can observe that $f(n)$ is monotonically increasing up to $n = 28$; and that $f(n+1) - f(n) \leq 1$ for n in this range. Intuitively, we expect that f is a monotonically increasing function and that $f(n+1) - f(n) \leq 1$ holds for all $n \in \mathbb{N}$.

We prove first that f is a non-decreasing function. It is always possible to attach a square tile to the outer perimeter of an n-omino and obtain an $(n+1)$-omino with at least the same number of holes. This implies that $f(n) \leq f(n+1)$ for evey $n \geq 1$, so f is non-decreasing. The following lemma tells us that f never increases in one step by more than one.

Lemma 1. *For every* $n \geq 1$
$$f(n+1) - f(n) \leq 1.$$

Proof. We will show that if A is any $(n+1)$-omino, then there exists an n-omino B such that $h(B) \geq h(A) - 1$.

Let A be an $(n+1)$-omino and k be the number of tiles in the bottom row of A and denote the leftmost tile in this row by l.

If $k = 1$, then l is only connected to one other tile. This allows us to delete l without disconnecting A or destroying any holes.

Now, suppose that the statement holds true whenever $k = 1, 2, \ldots, m$ (this is the induction hypothesis). Let $k = m + 1$ and denote by l_1, l_2, and l_3 the tile sites that share boundary with and are up and to the right of l—see Figure 2.5. Each of these three tile sites (l_1, l_2, and l_3) could either be occupied by tiles in A or not. All six possibilities for l_1, l_2 and l_3 are depicted in Figure 2.5. Any other combination not depicted in this figure would result in A having a disconnected interior and thus would not possible.
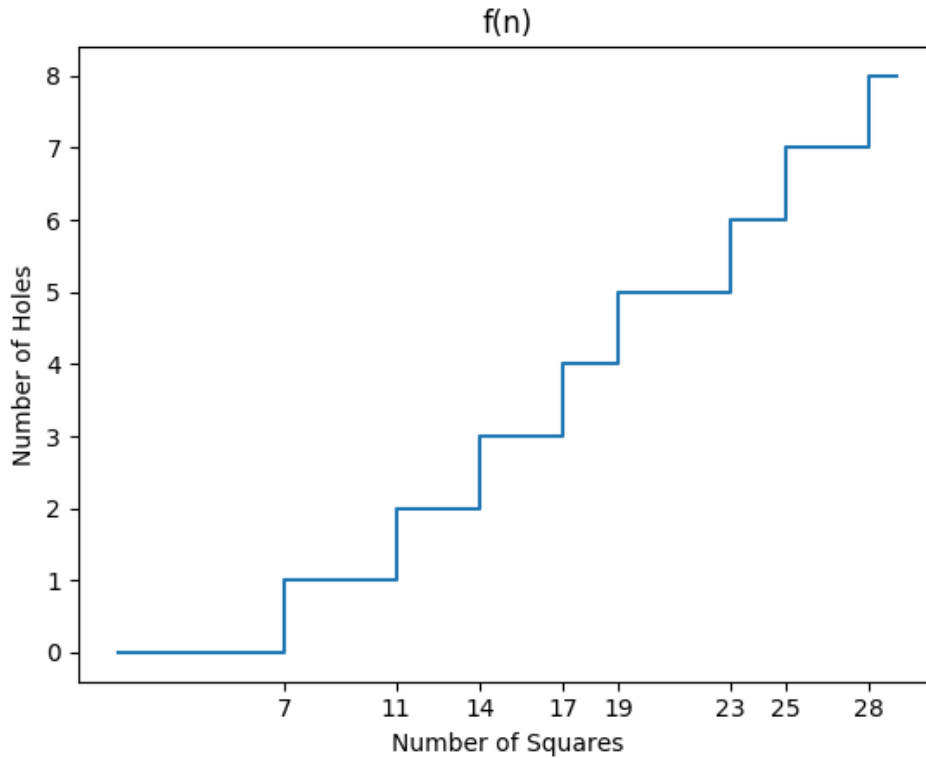
FIGURE 2.4: The step function $f(n)$ for $1 \leq n \leq 28$. These values of $f(n)$ were obtained from Table 2.2.2

If A is such that the local tile structure around l coincides with C_1, C_2, C_3, C_5, or C_6, then it is already possible to delete l from A to generate an n-omino B such that $h(B) \geq h(A) - 1$.

However, if C_4 is the local structure around l, then it is possible that deleting l disconnects A. In this case, we delete l and then add a new tile at the empty tile site l_2. This yields a new polyomino A' with the same number of tiles. If the addition of this new tile causes the coverage of a hole, then $h(A') = h(A) - 1$, and C_1 or C_3 must then be the new local structure around the leftmost tile of the bottom row of A'. This allows us to terminate the process by deleting the bottom, leftmost tile from A' without destroying more holes. If we have not destroyed any holes, then we have an $(n+1)$-omino A' with $h(A) = h(A')$ and with $k = m$ tiles in the bottom row. Hence, we can apply the induction hypothesis and the desired result follows. $\qquad \square$

In the next lemma we prove that g is the right inverse of f and that the values of g capture the times of f increasements.

Lemma 2. *For every $m \geq 1$, we have $f(g(m)) = m$. Also, $g(m) = n$ if and only if $f(n) = m$ and $f(n-1) = m-1$.*

Proof. By the definitions of f and g, we have that $f(g(m)) \geq m$. Supposing by way of contradiction that for some m we have $f(g(m)) \geq m+1$. By Lemma 1, $f(g(m) - 1) \geq m$. This implies that there exists a polyomino with $g(m) - 1$ tiles and at least m holes, but this contradicts the definition of g. We then conclude that $f(g(m)) = m$ for every m. It then follows that $g(m) = n$ if and only if $f(n) = m$ and $f(n-1) = m-1$. $\qquad \square$
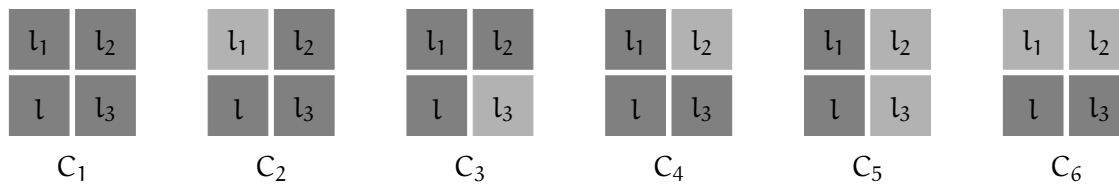
FIGURE 2.5: The tile l denotes the leftmost tile in the bottom row of a polyomino A. If the tile sites l_1, l_2, or l_3 are in A, they are colored dark gray; otherwise, they are light gray. C_1-C_6 are the six possible combinations for these tile sites. All other possibilities are rejected because they give a square structure with a non-connected interior but we are supposing that A is a polyomino.

As a consequence of this lemma, we get that g is also a monotonically increasing function. In Figure 2.6, we can observe that up to $m = 8$ the steps of g are of a length less than or equal to four and no less than two. This can also be observed in Figure 2.4 because the steps of f are located no more than four and at least two values apart. In the next lemma we partially prove that this property of g is always true.
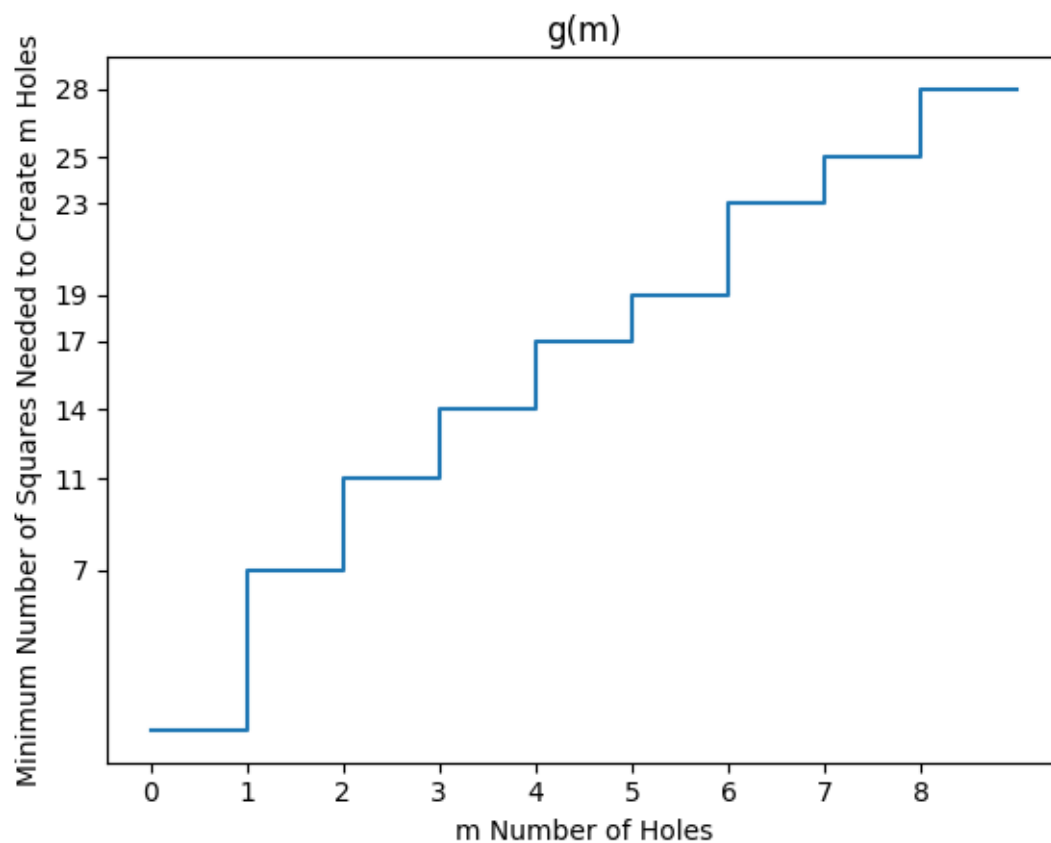

FIGURE 2.6: The step function $g(m)$ for $1 \leq m \leq 8$. These values of $g(m)$ were obtained from Table 2.2.2

Lemma 3. *For every* $m \geq 1$

$$1 \leq g(m+1) - g(m) \leq 4.$$

Proof. Assume that $g(m+1) = n$ and $g(m) = n$ for a natural number m. Then, from Lemma 2, we get $f(n) = m+1$, $f(n-1) = m$, $f(n) = m$, and $f(n-1) = m-1$ which is a contradiction. It follows that $g(m+1)$ cannot be equal to $g(m)$ for any natural number m. Because g is a monotonically increasing function, we conclude that for all $m \in \mathbb{N}$
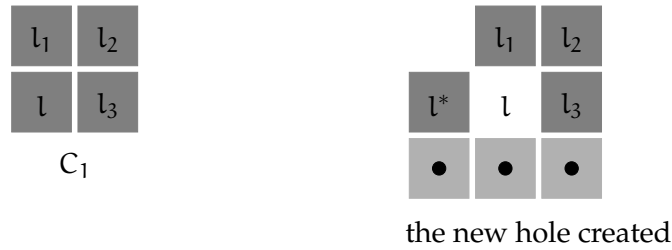
$$1 \leq g(m+1) - g(m).$$

FIGURE 2.7: Given a polyomino denote the leftmost tile in the bottom row by l. We are assuming that the local configuration around the tile l is C_1 — see Figure 2.5. The added tiles have a black dot in the middle and we have moved the tile that was in l to the l^* position.

FIGURE 2.8: Given a polyomino denote the leftmost tile in the bottom row by l. We are assuming that the local configuration around the tile l is C_2 — see Figure 2.5. In this case T can be part of the polyomino. If T is in the polyomino, then we generate one more hole as depicted in Option A). If T is not in the polyomino, then we generate one more hole as depicted in Option B). The added tiles have a black dot in the middle. In the case of Option B) we have moved the tile that was in l to the position l_1. This movement does not cover an already existing hole because T is not a tile in the polyomino.
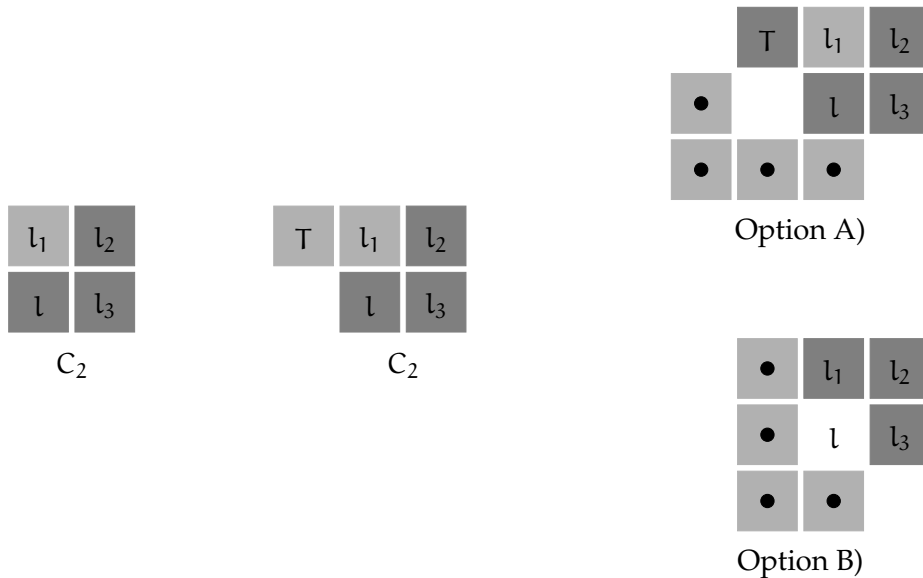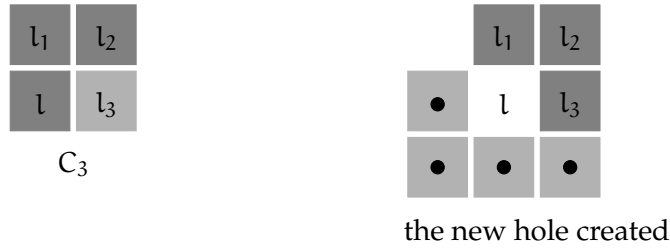
the new hole created

FIGURE 2.9: Given a polyomino denote the leftmost tile in the bottom row by l. We are assuming that the local configuration around the tile l is C_3—see Figure 2.5. The added tiles have a black dot in the middle and we have moved the tile that was in l to the position l_3. This movement does not cover a hole because l_3 is in the bottom row of the polyomino.

It is possible to add four tiles or less to any polyomino (for example, around the leftmost tile in the bottom row) and, if needed, relocate some tiles already present in the polyomino to create one more hole—see Figures 2.7, 2.8, 2.9, 2.10, 2.11 and 2.12. This implies that for all $m \in \mathbb{N}$

$$g(m+1) - g(m) \leq 4.$$

$\square$

We have not yet been able to prove that $2 \leq g(m+1) - g(m)$ for all $m \in \mathbb{N}$, but I strongly believe this is the case. This is stated in the next conjecture.

Conjecture 1. *For every* $m \geq 1$

$$2 \leq g(m+1) - g(m) \leq 4.$$

2.3.2 Perimeter

The edge perimeter $\mathrm{per}(A)$ of a polyomino $A \in \mathcal{A}_n$ is defined as the number of edges on the topological boundary of A. As an example, the 8-omino in Figure 2.3 has a perimeter equal to 16.

For all $n \in \mathbb{N}$ we denote $p_{\min}(n)$ and $p_{\max}(n)$ as the minimum and maximum edge perimeters possible for a polyomino with an area of n.

In 1976, F. Harary and H. Harborth [17] found an exact formula for the minimum perimeter possible in an n-omino.

Theorem A. *(F. Harary and H. Harborth, 1976)*

$$p_{\min}(n) = 2\lceil 2\sqrt{n} \rceil. \tag{2.4}$$

In particular, they exhibited a family of spiral-like polyominoes that achieve the minimum perimeter. Any square polyomino is a spiral-like polyomino—see Figure 2.13. In [21], S. Kurz counts the number of polyominoes (up to rotations, translations, and reflections) that have a perimeter equal to the minimum perimeter.

Let $A \in \mathcal{A}_n$. The number of edges that are on the boundary of two squares of A will be denoted by $b(A)$, in which case the edges are contained in the interior of the
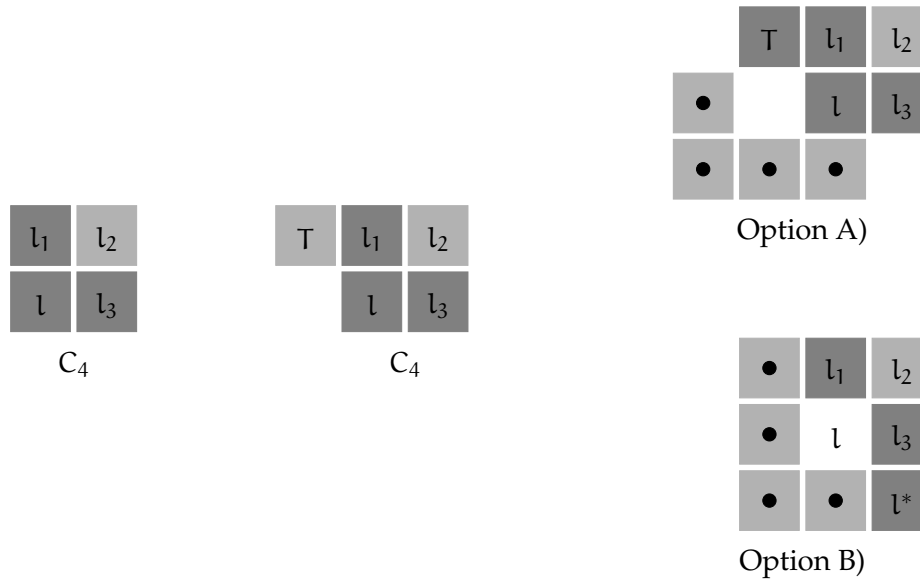
FIGURE 2.10: Given a polyomino denote the leftmost tile in the bottom row by l. We are assuming that the local configuration around the tile l is C_4— see Figure 2.5.In this case T can be in the polyomino. If T is in the polyomino, then we generate one more hole as depicted in Option A). If T is not in the polyomino, then we generate one more hole as depicted in Option B). The added tiles have a black dot in the middle. In the case of Option B) we have moved the tile that was in l to the position l^*. This movement does not disconnect the polyomino because the added tiles guarantee that the tiles l_1 and l_3 are still path connected by a path contained in the interior of the polyomino.

polyomino. Observe that all the edges of the squares of a polyomino either belong to the perimeter or to the interior of the polyomino. This means that

$$4n = \text{per}(A) + 2b(A). \tag{2.5}$$

For example, if A is any 7-omino depicted in Figure 2.14, then $b(A) = n - 1 = 6$ and $\text{per}(A) = 4n - 2(n - 1) = 16$.

Let $b_{\min}(n)$ be the minimum number of edges shared by two squares that an n-omino can have. It is possible to associate a **dual graph** with any polyomino by considering each square as a vertex and by connecting any two of these vertices if they share an edge.

Let A be an n-omino. Because any polyomino has to have a connected interior, then the associated dual graph of A is connected and there exists a spanning tree of this graph. If the dual graph has n vertices, then there are at least $n - 1$ edges in the spanning tree. This implies that there are at least $n - 1$ different common edges in A. That is, $b(A) \geq n - 1$, which is true for any n-omino. This then gives us that $b_{\min}(n) \geq n - 1$. Observing that the n-omino C with only one column has $b(C) = n - 1$, we conclude that $b_{\min}(n) = n - 1$.

As a consequence of this and the equality (2.5) for every $n \geq 1$

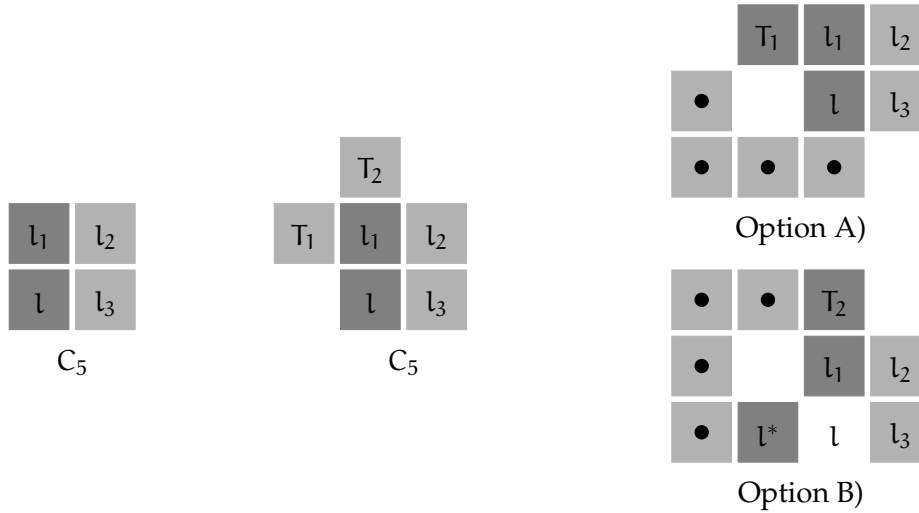$$p_{\max}(n) \leq 4n - 2b_{\min}(n) = 4n - 2(n - 1) = 2n + 2, \tag{2.6}$$

FIGURE 2.11: Given a polyomino denote the leftmost tile in the bottom row by l. We are assuming that the local configuration around the tile l is C_5— see Figure 2.5. In this case either T_1 or T_2 (or both) must be in the polyomino. If T_1 is in the polyomino, then we generate one more hole as depicted in Option A). If T_1 is not in the polyomino, then we generate one more hole as depicted in Option B). The added tiles have a black dot in the middle. In the case of Option B) we have moved the tile that was in l to the l^* position. Also, it might be the case that in Option B) less than four tiles are needed to be added to the polyomino if they were already in the polyomino.

and equality is only achieved by polyominoes with the number of common edges equal to $b_{min}(n)$.

We need to distinguish when an edge that is on the perimeter of a polyomino $A \in \mathcal{A}_n$ is an edge that forms part of a hole in the polyomino. To do this, we define such edges as being part of the hole perimeter. We represent by $p_h(A)$ the number of edges on the hole perimeter of A. We define the outer perimeter of A, denoted by $p_o(A)$, as the difference between the perimeter $per(A)$ and the hole perimeter

$$p_o(A) = per(A) - p_h(A).$$

If a polyomino A is simply connected, then $per(A) = p_o(A)$. In general, $per(A) = p_o(A) + p_h(A)$ by definition.

Polyominoes with holes might achieve the maximum perimeter. However, the next lemma checks the intuitive fact that the minimum perimeter cannot be achieved by polyominoes with holes.

Lemma 4. *If $A \in \mathcal{A}_n$ and A has at least one hole, then $p_{min}(n) < per(A)$.*

Proof. By equation (2.4), an n-omino with $k > 1$ holes, has an outer perimeter at least equal to $p_{min}(k + n)$. A polyomino with k holes has a hole perimeter greater or equal to $4k$. We have

$$per(A) = p_o(A) + p_h(A) \geq p_{min}(n + k) + 4k. \tag{2.7}$$

the new hole created
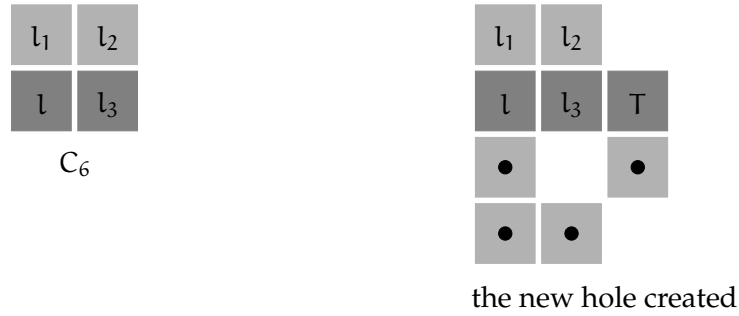
FIGURE 2.12: We generate one additional hole in a polyomino by adding to it three more tiles around its leftmost tile in the bottom row, denoted by l. We are assuming that the local configuration around the tile l is C_6— see Figure 2.5. The added tiles have a black dot in the middle. Observe that the tile T has to be in the polyomino to connect the tiles l and l_3 to the other tiles in the polyomino.

FIGURE 2.13: Polyominoes that achieve the minimum perimeter.

Because the function $h(x) = 2\lceil 2\sqrt{x} \rceil$ is a non-decreasing function, we can conclude from expression (2.7) that

$$\mathrm{per}(A) \geq p_{min}(n+k) + 4k \geq p_{min}(n) + 4k > p_{min}(n).$$

$\square$

We denote this minimum outer perimeter over all polyominoes with n tiles and m holes (whenever such polyominoes exist) by $p_{min}^{out}(n, m)$. Note that we always have

$$p_{min}(n+m) \leq p_{min}^{out}(n, m), \tag{2.8}$$

by definition.

Equality in (2.8) is attained by some polyominoes. For example, the equality holds if $n = 7$ and $m = 1$, and if $n = 14$ and $m = 3$ (for the last example see Figure 2.15).

FIGURE 2.14: Polyominoes that achieve the maximum perimeter.

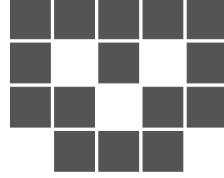FIGURE 2.15: Achieving the minimum outer perimeter and the maximum number of holes. This polyomino A has 14 tiles, 3 holes, and $p_o(A) = 18$. We know that $f(14) = 3$. From (2.4) we have $p_{min}(14 + 3) = 2\lceil 2\sqrt{17} \rceil = 18$. We conclude that $p_{min}(14 + 3) = p_{min}^{out}(14, 3)$.

In order to show a polyomino for which the equality is not attained, we need an upper bound for the maximum number of holes that polyominoes with the same number of tiles can have. There exists a polyomino for which $p_{min}(n + m) < p_{min}^{out}(n, m)$.

2.3.3 Main upper bound

We prove in the next lemma a general result which allows us to generate an upper bound of f from any lower bound of f. We will apply this lemma again in Sections 2.4 and 2.5.

Lemma 5. *Let n be any natural number. If $f(n)$ denotes the maximum number of holes that an element of $\mathcal{A}_n$ can have, then*

$$f(n) \le \frac{4n - 2b_{min}(n) - p_{min}(n + f(n))}{4}. \tag{2.9}$$

Proof. Let A be an element in $\mathcal{A}_n$ and let $h(A)$ denote the number of holes in A. Then

$$p_h(A) = 4n - 2b(A) - p_o(A) \le 4n - 2b_{min}(n) - p_o(A). \tag{2.10}$$

By (2.8) we have

$$p_o(A) \ge p_{min}^{out}(n, h(A)) \ge p_{min}(n + h(A)). \tag{2.11}$$

From inequalities (2.10) and (2.11) we get

$$p_h(A) \le 4n - 2b_{min}(n) - p_{min}(n + h(A)). \tag{2.12}$$

Then, if A is a polyomino that has $f(n)$ holes, we get

$$f(n) \le \frac{p_h(A)}{4} \le \frac{4n - 2b_{min}(n) - p_{min}(n + f(n))}{4},$$

which establishes inequality (2.9).

$\square$

As a corollary from Lemma 5, for any lower bound of f we get the following upper bound of f.

Corollary 2. *If* $lb_f(n) \leq f(n)$, *then*

$$f(n) \leq \frac{1}{2}n - \frac{1}{2}\left\lceil 2\sqrt{n + lb_f(n)} \right\rceil + \frac{1}{2}. \tag{2.13}$$

Proof. Let $lb_f(n) \leq f(n)$ for a natural number n, then $p_{min}(n + lb_f(n)) \leq p_{min}(n + f(n))$ by monotonicity. This inequality allows us to obtain (2.13) by substituting $p_{min}(n + f(n))$ with $p_{min}(n + lb_f(n))$ in (2.9). $\square$

2.4   Polyominoes that attain the maximum number of holes

In this section we construct a sequence $\{S_k\}_{k=1}^{\infty}$ of polyominoes with n_k squares such that S_k have h_k holes for $k \geq 1$. We show in Lemma 6 that these polyominoes have maximally many holes and that $h(S_k) = f(n_k)$ for all $k \in \mathbb{N}$. Finally, we prove Theorem 1.

In addition to having maximally many holes, the sequence $\{S_k\}_{k=1}^{\infty}$ has very interesting geometric properties that we analyze in Section 2.6.

2.4.1   Construction of the sequence

Remember that n_k and h_k were respectively defined in (2.2) and (2.3).

The first three elements, S_1, S_2, and S_3, of the sequence are shown in Figure 2.3.
To generate the rest of the sequence for $n \geq 2$, we use the following recursion process:

- First, place a rotation point in the center of the top right tile of S_{n-1}.

- Then, rotate S_{n-1} with respect to this point ninety degrees four times creating four, overlapping copies.

- Finally, remove the tile containing the rotation point.

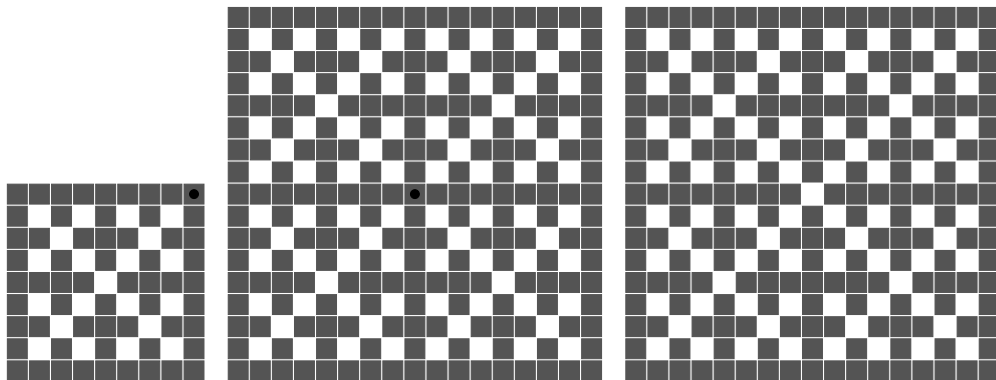In Figure 2.16 we construct S_4 from S_3 following this recursive algorithm.

FIGURE 2.16: Generating S_{n+1} from S_n (L to R). (1) The polyomino S_3, (2) four overlapping rotated copies of S_3, and (3) the polyomino S_4 made by removing the tile of rotation.

2.4.2 Number of holes in S_k and proof of Theorem 1

From this construction, we observe that for $k \geq 1$ we have the recursion $h(S_k) = 4h(S_{k-1}) + 1$. The factor of four is due to the four reflected copies of S_{k-1} generated in the process of constructing S_k. The one hole added is generated by the square removed after the rotation process. Then, because $h(S_1) = 1$, we get $h(S_k) = h_k$ for all $k \geq 1$.

Let s_k be the number of tiles in S_k for $k \geq 1$. The sequence s_k satisfies the recursion

$$s_{k+1} = 4s_k - 4(2^k + 1),$$

because the polyominoes S_k have side lengths of $2^k + 1$ tiles and, in the rotation process, $4(2^k + 1)$ tiles overlap. Additionally, the sequence n_k satisfies the relationship

$$n_{k+1} = 4n_k - 4(2^k + 1).$$

Then, because both n_k and s_k satisfy the same recursion relationship and are equal in the first element ($s_1 = 8$, $n_1 = 8$), we can conclude that n_k and s_k are the same sequences. We have proved the following lemma.

Lemma 6. *There exists a sequence of polyominoes $\{S_k\}_{k=1}^{\infty}$, such that S_k has n_k tiles and h_k holes.*

Now we can prove Theorem 1.

Proof of Theorem 1. First we show that $f(n_k) = h_k$. From Lemma 6 we know

$$h_k \leq f(n_k). \tag{2.14}$$

Substituting this lower bound in (2.13) we have

$$f(n_k) \leq \frac{1}{2}n_k - \frac{1}{2}\left\lceil 2\sqrt{n_k + h_k} \right\rceil + \frac{1}{2}. \tag{2.15}$$

From the easily verified identity

$$h_k + \frac{1}{2} = \frac{1}{2}n_k - \frac{1}{2}\left\lceil 2\sqrt{n_k + h_k} \right\rceil + \frac{1}{2},$$

and inequalities (2.14) and (2.15), we get

$$h_k \leq f(n_k) \leq h_k + \frac{1}{2}. \tag{2.16}$$

This implies that

$$f(n_k) = h_k, \tag{2.17}$$

because $f(n_k)$ and h_k are integers.

Now, we prove that $f(n_k - 1) = h_k$. By removing the upper leftmost square from each S_k, it is possible to generate a sequence of polyominoes $\{A_k\}_{k=1}^{\infty}$ with $n_k - 1$ tiles each A_k, such that $h(A_k) = h_k$. This implies that $h_k \leq f(n_k - 1)$. Then, because $f(n_k - 1) \leq f(n_k)$ and $f(n_k) = h_k$, we can conclude that

$$f(n_k - 1) = h_k.$$

Finally, we prove that $f(n_k - 2) = h_k - 1$. Because $f(n_k - 1) = h_k$ and f is a non-decreasing function, consequently $f(n_k - 2) \leq h_k$. If we assume by way of contradiction that $f(n_k - 2) = h_k$, then using (2.13) it must be that

$$h_k = f(n_k - 2) \leq \frac{1}{2}(n_k - 2) - \frac{1}{2}\left\lceil 2\sqrt{(n_k - 2) + h_k} \right\rceil + \frac{1}{2}. \tag{2.18}$$

Substituting (2.2) and (2.3) in this inequality leads to a contradiction. Then $f(n_k - 2) < h_k$. From this inequality and Lemma 1, we can conclude that

$$f(n_k - 2) = h_k - 1. \tag{2.19}$$

$\square$

2.5   General bounds and proof of Theorem 2

The main purpose of this section is to give a proof of Theorem 2. For this, we first get a lower bound for f for a particular sequence of natural numbers. Then, we extend this lower bound for all $f(n)$ for large enough $n \in \mathbb{N}$. Finally, we apply Corollary 2 to this general lower bound to obtain a general upper bound for $f(n)$.

2.5.1   General lower bound

We begin by describing how to construct the sequence of polyominoes $\{R_k\}_{k=1}^{\infty}$ of polyominoes with $m_k = 40k^2 + 20k$ tiles and $t_k = 20k^2$ holes.

The pattern S, depicted in Figure 2.17, covers a three-by-two rectangle of the regular square lattice. It has four occupied tiles and two empty tiles.

FIGURE 2.17: The pattern S consists of four occupied tiles and two empty tiles.

We want to construct polyominoes that contain as many non-overlapping copies of S as possible and in such a way that the two empty tiles in the pattern S are transformed into holes. Of course, some extra tiles will be added to guarantee that the resulting tile structure is indeed a polyomino. Our task, then, is to minimize the number of tiles that we need to add to have connectivity and to maximize the copies of S that can be concatenated. We solved this min-max problem by performing some algebra and calculus computations. We describe in detail the resulting construction in what follows.

We first place $10k^2$ copies of the pattern S into a rectangle with 6k tiles high and 10k tiles long. Then, we add a top row of 10k tiles and a leftmost column of 6k tiles. Finally, we attach 2k vertically aligned dominoes (for a total of 4k tiles) to the rightmost column. The polyomino R_2 is depicted in Figure 2.18, and the $40k^2$ tiles just described are colored with the lightest gray in this figure. The initial, repeated pattern S is bordered in black within R_2. From this construction we get the following
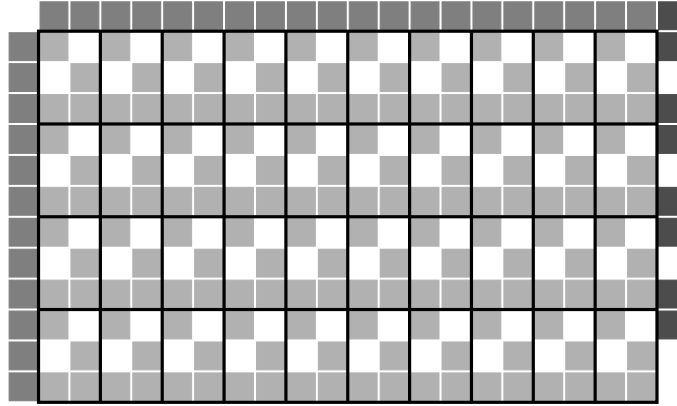
result.

FIGURE 2.18: The polyomino R_2 has 200 tiles and 80 holes.

Lemma 7. *The elements of the sequence of polyominoes* $\{R_k\}_{k=1}^\infty$ *have* $m_k = 40k^2 + 20k$ *tiles and* $t_k = 20k^2$ *holes.*

Proof. By construction, R_k has m_k tiles. It also has two holes per instance of the S pattern in R_k. There are $10k^2$ non-overlapping copies of S in R_k. This implies that there are $t_k = 20k^2$ holes in R_k. Thus, we have constructed a sequence of polyominoes $\{R_k\}$ such that each polyomino R_k has m_k tiles and t_k holes. $\qquad\square$

As a consequence of Lemma 7 we get the lower bound $t_k \leq f(m_k)$. By expressing t_k in terms of m_k, we get the following lemma.

Lemma 8. *For each* $k \geq 1$

$$\frac{1}{2}m_k - \sqrt{\frac{5}{2}\left(m_k + \frac{5}{2}\right)} + \frac{5}{2} \leq f(m_k). \tag{2.20}$$

We now extend the lower bound given in (2.20) to all $f(n)$ for sufficiently large values of n. We extend this lower bound by constructing an n-omino that gives a lower bound for $f(n)$ if $n > m_{22}$. This gives the same lower bound that we have found for $f(m_k)$ whenever $m_k < n < m_{k+1}$. First, we describe an algorithm for the construction of such an n-omino.

Let $n > m_{22}$ and k be the smallest number such that $m_k < n$. For constructing a polyomino P_n with n tiles by adding tiles to R_k, we use the following algorithm:

Step 0 Let $P = R_k$, $t = n - m_k$, and $c = k$. If at any part of the process $t = 1$, then go to Step 4. If $t = 0$, then make $P_n = P$, return P_n, and the process is completed. Observe that if $n = m_k$ then $P_n = R_k$.

Step 1 Add one vertically-oriented domino-patterned tile on the column next to the rightmost column of P at the highest possible position so that it creates a new hole with an area of 1. Decrease t by 2 ($t = t - 2$).

Step 2 Continue to add dominoes to this column in such a way that a new hole is created with each new domino placed. After adding each domino, decrease t by 2. If $c - 1$ dominoes were added to this column of P, go to the next step.

Step 3 Redefine P to be the polyomino resulting from Step 1 and Step 2, decrease c by 1 ($c = c - 1$), and go back to Step 1.

Step 4 If $t = 1$, then place this last tile in any location around the outer perimeter of P, set $P_n = P$, and set $t = t - 1$.
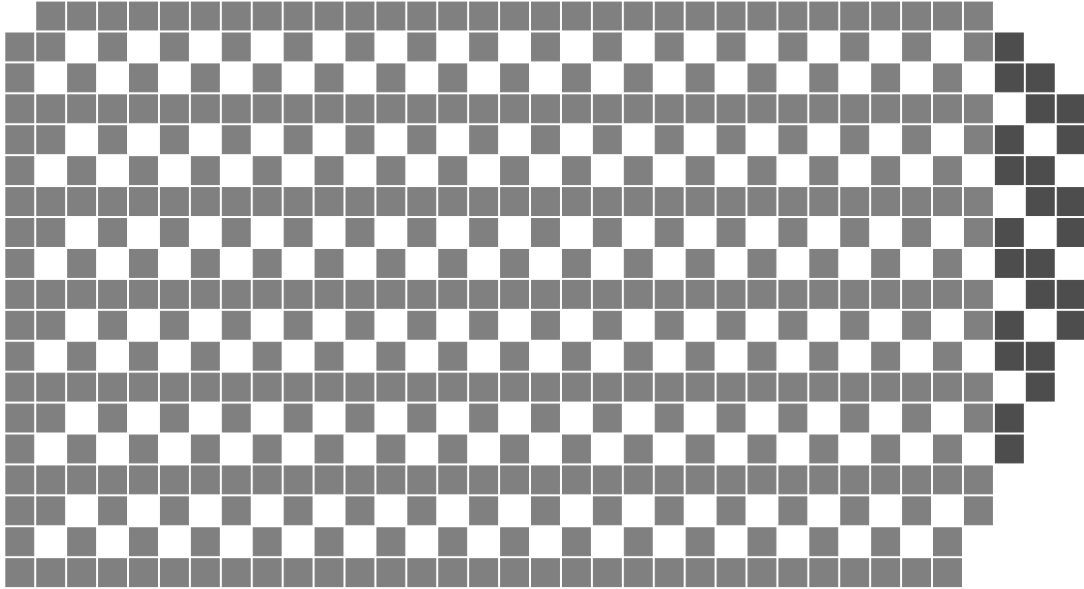
FIGURE 2.19: R_3 has 420 tiles and 180 holes. Adding tiles in a domino pattern on the right side yields P_{444}, which has 444 tiles and 192 holes.

When $k \geq 22$, this algorithm ensures that t will equal zero before c is zero. This property is extremely important for the construction of the polyomino P_n. It guarantees that we finish the construction process of P_n by placing all the $n - m_k$ tiles as dominoes on the left columns of R_k and that we get a new hole for each domino added. For instance, $444 < m_{22}$ and $460 < m_{22}$. Nevertheless, this algorithm works for constructing a polyomino with 444 tiles but does not work for constructing a polyomino with 460 tiles.

As with this algorithm we create one new hole for each added domino to R_k (Figure 2.19), we get

$$h(R_k) + \lfloor (n - m_k)/2 \rfloor \leq h(P_n).$$

Hence, by Lemma 8 and this last inequality of $h(P_n)$, we have proved the following lemma.

Lemma 9. *If $n > m_{22}$, then*

$$\frac{1}{2} n - \sqrt{\frac{5}{2} \left(n + \frac{5}{2} \right)} + 2 \leq f(n). \tag{2.21}$$

2.5.2 Proof of Theorem 2.

Proof. Substituting (2.21) as a lower bound of f in (2.13) and using properties of the ceiling function, we get

$$f(n) \leq \frac{1}{2}n - \frac{1}{2}\left\lceil 2\sqrt{n + \frac{n}{2} - C_1\sqrt{n}} \right\rceil + \frac{1}{2} \tag{2.22}$$

$$\leq \frac{1}{2}n - \sqrt{\frac{3n}{2} - C_1\sqrt{n} + 1} \tag{2.23}$$

$$\leq \frac{1}{2}n - C_2\sqrt{n}, \tag{2.24}$$

for $C_2 < \sqrt{3/2}$ and large enough n.

Inequalities (2.21) and (2.24) imply that for any positive numbers $C_1 > \sqrt{5/2}$ and $C_2 < \sqrt{3/2}$ there exists a natural number n_0 such that

$$\frac{1}{2}n - C_1\sqrt{n} \leq f(n) \leq \frac{1}{2}n - C_2\sqrt{n},$$

for all $n > n_0$. This completes the proof of Theorem 2.

$\square$

The constant $\sqrt{3/2}$ cannot be improved because the number of holes of our main sequence of polyominoes $\{S_k\}_{k=1}^{\infty}$ constructed in Section 2.4 attains this constant. This can be proved by substituting (2.2) and (2.3) in equality (2.17) to get

$$f(n_k) = \frac{1}{2}n_k - \sqrt{\frac{3}{2}}\sqrt{n_k + o(n_k)} + c. \tag{2.25}$$

We do not know if it is possible to find an exact formula for f, but the number of holes attained by the sequence of polyominoes $\{S_k\}_{n=1}^{\infty}$ shows that

$$f(n) = \frac{1}{2}n - \sqrt{\frac{3}{2}n + \frac{1}{4}} + \frac{1}{2},$$

infinitely often.

We end this chapter by analyzing some interesting geometric and asymptotic features of the main sequence of polyominoes $\{S_k\}$ that we constructed in Section 2.4.

2.6 Some geometric properties of the sequence S_k

The recursive construction of the main sequence of polyominoes $\{S_k\}$ suggests that the inner edge boundary is approaching a limiting fractal shape. During a conversation with Elliot Paquette (at The Ohio State University in 2018), he pointed out that one way to make sense of this idea is to consider the "inner boundary" of a polyomino S_n to be an immersed circle in $\mathbb{R}^2$. After appropriately rescaling and reparameterizing, these circles seem to converge to a space-filling curve in $[0,1]^2$ — see Figure 2.20.
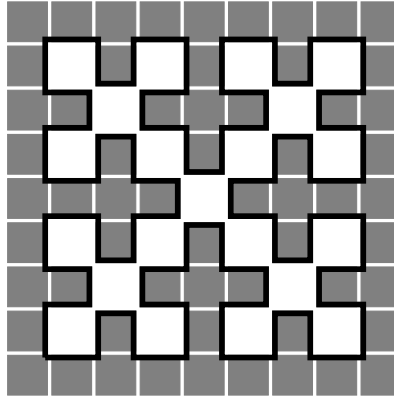
FIGURE 2.20: The inner boundary of S_3. It can be seen as an immersed circle.

Also, it is possible to derive an aperiodic tiling of the plane from our sequence. Define the planar dual of a polyomino S_k as the planar graph obtained by placing one vertex at the center of every tile and connecting two of these vertices if the two tiles containing them share an edge or if the two vertices are contained in tiles that have one of the patterns depicted in Figure 2.21.

FIGURE 2.21: For constructing the planar dual of a polyomino we place vertices in each square of the polyomino and we connect two vertices whenever they share an edge or whenever they share only one vertex. The other two tiles in the smallest square that contains them both are empty. This figure shows the two possibilities for this structure and the edge that has to be drawn.

With this construction we get one bounded face for every hole—see Figure 2.22. We can construct a limiting infinite polyomino by centering every S_k at the origin and taking the union of all of them. The planar dual of this infinite polyomino is an aperiodic tiling with three different prototiles: squares, pentagons, and hexagons.
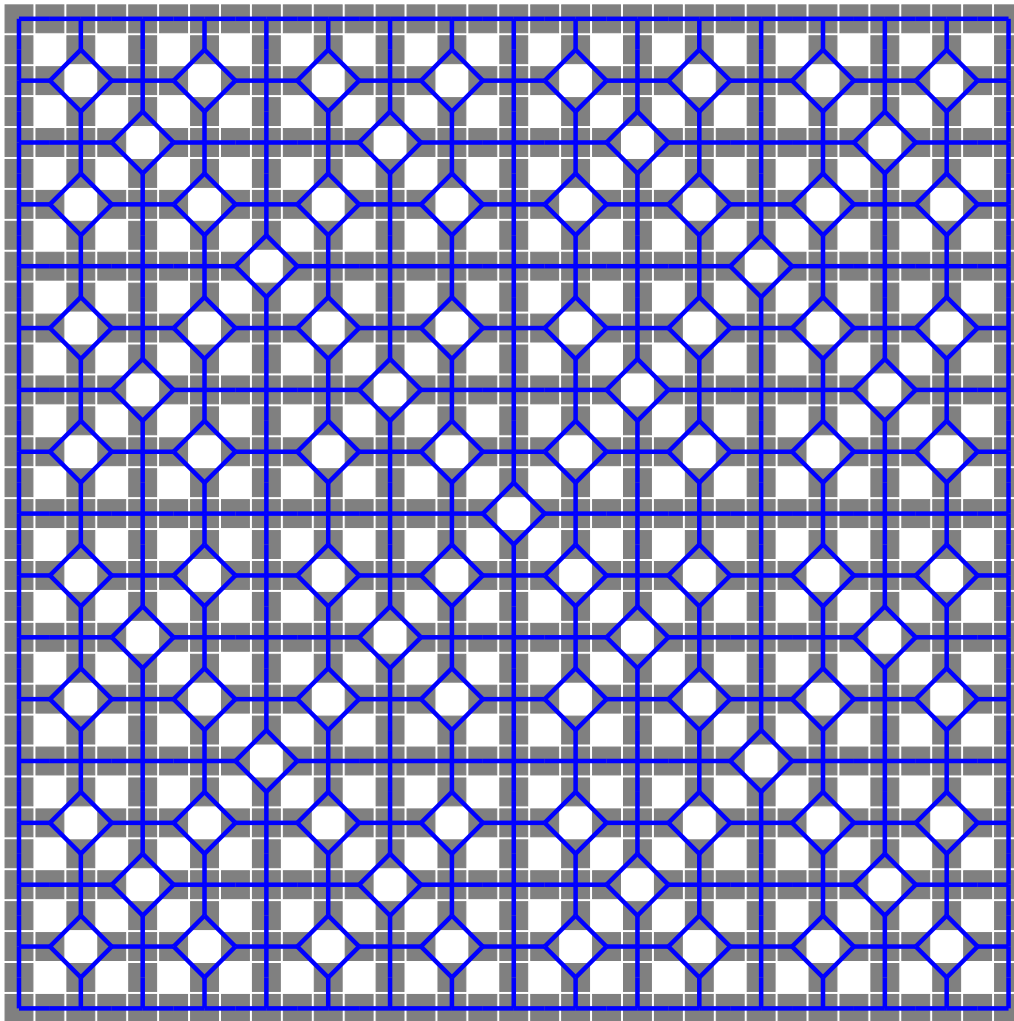
FIGURE 2.22: The polyomino S_5 with its planar dual superimposed
in blue.

For analyzing the next interesting geometric property of our main sequence we
need to construct *on top of a polyomino* a graph that captures the edge connectivity of
its tiles. In Section 2.3.2 we have already described this construction and we have
called this graph the dual graph of a polyomino. We recall here the definition for the
convenience of the reader.

Given a polyomino P we generate its dual graph by placing vertices in the center
of each one of the tiles of P and connecting any two vertices if their corresponding
tiles share an edge. This graph is commonly referred to in the literature as a site
animal and we will also use this construction in the following chapters.

Any element A_k of the sequence of polyominoes $\{A_k\}_{k=1}^\infty$ constructed in the proof
of Theorem 1, and that has been derived from our main sequence $\{S_k\}_{k=1}^\infty$ by deleting
one of its tile corners, has the property that the site animal associated with it is a tree.
This is a consequence of the fact that this sequence attains the upper bound for $f(n)$.

In this chapter we have studied the maximum number of holes that a polyomino
can have. In the next chapter we study how many holes a typical polyomino is

expected to have. In order to do so, we will give a more precise topological definition of the topological structure of a polyomino and of a hole in a polyomino.

Chapter 3

Asymptotic behavior of the homology of random polyominoes

Percolation theory models on regular lattices, such as the square regular lattice, are relevant in various areas of theoretical and experimental physics like material science [30], thermodynamics [6], and statistical mechanics [4]. These models have interesting critical phenomena at phase transitions. There are also several techniques for making simulations and for doing computational experiments to explore them.

This gives importance to the study of the topological features, such as homology groups, of the finite clusters of the percolation model on the regular square lattice of the plane. These finite clusters correspond to polyominoes as we have defined them in Chapter 2.

In this chapter we present the second set of contributions of this thesis. We establish the growth rate of the expected number of holes of polyominoes for these families of distributions that are related to the percolation model on the plane and for the uniform distribution. We prove that the expected number of holes grows linearly with respect to the number of tiles.

The precise statements of these results are stated in Theorem 3 and Theorem 5. We prove these theorems in Section 3.2 and Section 3.4, respectively. Also, we exhibit explicit constants for the upper and lower linear bounds in the uniform distribution case in Corollary 5.

As far as we know, this is the first time that these problems have been studied for percolation models, for the uniform distribution and for any other distributions on regular lattices.

3.1   (Algebraic) Topological definition of polyominoes and holes

In this section we give a precise algebraic topological definition of polyominoes and of the random variable that captures the number of holes in a polyomino. We are assuming that the reader knows how to construct, given a CW-complex, the sequence of homology vector spaces over the field $\mathbb{F}_2 = \{0, 1\}$ associated with the complex— see [18].

We are interested in defining this complex structure and the corresponding homology vector spaces because, intuitively, the ranks of these spaces capture the number of holes that a polyomino or a higher-dimensional polyform may have. For instance, on the plane, the rank of the first homology group corresponds exactly with the number of holes that a polyomino has. In what follows we provide the definitions needed for constructing a CW-complex for any given polyomino.

It is possible to construct an infinite geometric CW-Complex on $\mathbb{R}^2$ from the square lattice $\mathbb{Z}^2$ (this lattice can be seen as the square regular tessellation of the plane or the infinite checkerboard defined in Chapter 2). Then, given a polyomino, we can construct a subcomplex of this infinite cellular structure. To make these definitions more precise, we first give a cell-decomposition on $\mathbb{R}^2$:

- The set of 0-cells, denoted by X^0, consists of all the points in the integer lattice on the plane, i.e., $X^0 = \mathbb{Z}^2$.

- The set of all 1-cells denoted by X^1 is defined as the set of all closed segments of length one between elements of X^0 that are either parallel to the vector $\overrightarrow{(1,0)}$ or $\overrightarrow{(0,1)}$.

- The set of all 2-cells denoted by X^2 contain all closed squares with area 1 that have as their boundary four 1-cells.

We can define an infinite CW-structure on $\mathbb{R}^2$ consisting of the filtration $\{X^0, X^1, X^2\}$ and the infinite CW-complex on the plane consists of $\mathbb{R}^2$ endowed with this CW-structure, which we denote by $CW(\mathbb{Z}^2)$.

With this definition, for any given polyomino A we can construct a subcomplex of $CW(\mathbb{Z}^2)$. A is a (topological) subspace of $\mathbb{R}^2$ and A can be decomposed as the set union of elements in X^0, X^1, and X^2. This gives us a natural filtration of A by 0-cells, 1-cells, and 2-cells that we represent by A^0, A^1, and A^2, respectively.

Not all subcomplexes of $CW(\mathbb{Z}^2)$ correspond to a polyomino. A polyomino is a pure 2-dimensional finite subcomplex of $CW(\mathbb{Z}^2)$, with connected interior.

The site perimeter of a polyomino A, that we represent by $\mathrm{per}_{\mathrm{site}}(A)$, is the set of 2-cells in $CW(\mathbb{Z}^2)$ that share boundary with A. We will also use the notation $\mathrm{per}_{\mathrm{site}}(A)$ to refer to the cardinality of this set of cells. It will be clear from the context whether we are referring to the set of cells or its cardinality.

For each polyomino A we construct the sequence of homology vector spaces over the field $\mathbb{F}^2$ and we denote them by $\{H_k(A)\}_{k=0}^\infty$. The dimensions of these vector spaces have the topological information that will allow us to count the number of holes in A.

As usual, we represent by $\beta_i(A)$ the dimension of the finite i-th homology vector space:

$$\beta_i(A) := \dim(H_i(A)).$$

These ranks are commonly referred as Betti numbers. The value $\beta_0(A)$ gives the number of connected components in A that, in our case, is always equal to one because polyominoes are connected. As we have mentioned before, $\beta_1(A)$ coincides with the number of holes in A if the holes are defined as in Chapter 2. Finally,

$\beta_i(A) = 0$ for all $i > 1$. This is not always the case for polyforms or other finite cell structures defined in higher-dimension, regular lattices (tessellations).

We notice that if A and B are two polyominoes with m holes, then they are homotopy equivalent (and in fact homotopy equivalent to a bouquet of m circles) independently of the number of tiles that they have. Hence, their homology vector spaces are isomorphic, i.e, $H_k(A) \cong H_k(B)$ for all $k \in \mathbb{N}$. This, in our case, reduces to $\beta_1(A) = \beta_1(B)$. If $m \in \mathbb{N}$ and n is fixed, the enumeration of all $A \in \mathcal{A}_n$ for which $\beta_1(A) = m$ is an open problem.

From the values contained in Table 2.2.2, it is possible to compute the exact value of $\mathbb{E}[\beta_1]$ on each set $\mathcal{A}_n$ for all $1 \leq n \leq 28$. For instance, the expectation is zero for $n \leq 6$.

From the information that is known [27], it is not possible to compute the exact value of $\mathbb{E}[\beta_1]$ for polyominoes with more than 28 tiles. Though, from Theorem 2 we can derive that, independently of the probability distribution defined on $\mathcal{A}_n$, the expectation of β_1 grows at most linearly with respect to the number of tiles.

Lemma 10. *For any natural number n and any probability distribution defined on the set $\mathcal{A}_n$, we have*

$$\mathbb{E}[\beta_1] \leq \frac{1}{2}n. \tag{3.1}$$

Using the techniques that we present in Section 3.3, it is possible to improve this upper bound. Before giving a lower bound for $\mathbb{E}[\beta_1]$, we need to introduce new notation, new concepts, and new techniques.

3.2 Asymptotic bounds of homology for uniform random polyominoes

In this section we study the expected number of holes for random polyominoes. More precisely, we prove that the expectation of the random variable β_1, with respect to the uniform distribution defined on $\mathcal{A}_n$, grows linearly with respect to the number of tiles.

From Lemma 10 we know that this is true for the upper bound of $\mathbb{E}[\beta_1]$. But this is the extreme case as, intuitively, we expect that the majority of polyominoes do not have maximally many holes or that they are far from having maximally many holes.

For example, in Figure 3.1 we show a typical polyomino with 50 tiles; the maximum number of holes in a polyomino with 50 tiles is at least 14, but this polyomino only has one hole.

In the rest of this section we show that, for uniform distributed polyominoes, $\mathbb{E}[\beta_1]$ also has a lower bound that grows linearly with respect to the number of tiles in a polyomino. This result, the main conclusion of this section, is stated in Theorem 3. An essential part of the proof of this theorem relies on Theorem 4 and Theorem B. Theorem B was proved by N. Madras in [23]. We have used and adapted the techniques and ideas that he used to prove this result for the special case of the uniform distribution given in Theorem 4.
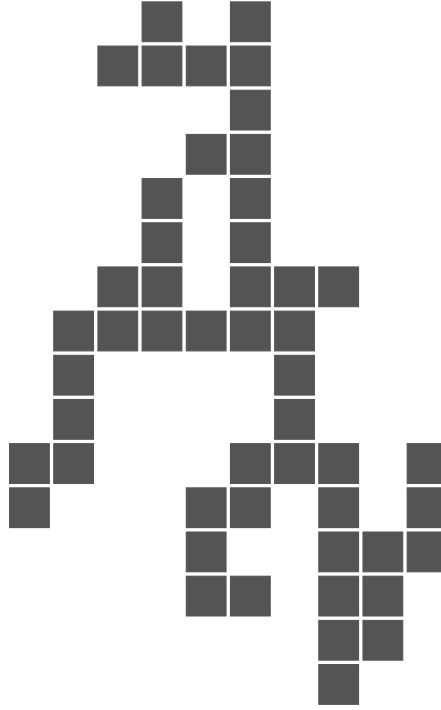
FIGURE 3.1: A *uniform random polyomino* with 50 tiles and only one hole. The maximum number of holes that a polyomino with 50 tiles can have is at least 14. This polyomino was sampled using a Markov chain Monte Carlo Metropolis algorithm that we study and define in detail in Chapter A.

As we will see in Section 3.3, Theorem B applies to more general structures than the square lattice and polyominoes with holes. It applies for regular patterns or subgraphs of regular lattices of $\mathbb{R}^n$ (in this same subsection we give precise definitions of regular patterns and regular lattices).

Although Theorem B is not stated for topological structures such as CW-complexes, we were able to use these techniques for proving our results using the useful graph representation of polyominoes by its dual graph also known in the literature as the site animal (remember from Section 2.6 that a dual graph of a polyomino (site animal) is constructed by placing vertices on the squares of a polyomino and connecting any two vertices if their corresponding squares share an edge).

These techniques could also be applied to more general topological structures. This is certainly the case for cubical complexes in higher dimensions and other polyforms in any regular lattice of $\mathbb{R}^n$. Thus, the results that we prove in this section are also true in these other general cases. And, it is also possible to find exact bounds for the uniform distribution and the percolation distributions defined on polyforms of a given number of tiles on these regular lattices.

The main result of this section is the following.

Theorem 3. *With the uniform distribution defined on $\mathcal{A}_n$, there exist constants C_1 and C_2 (not depending on n) such that*

$$C_1 \cdot n \leq \mathbb{E}[\beta_1] \leq C_2 \cdot n, \tag{3.2}$$

for sufficiently large values of n.

From Lemma 10 we can make for any natural number n the constant $C_2 = \frac{1}{2}$ in inequality (3.2) independently of the underlying distribution. Thus, we already have proved that the upper bound part of Theorem 3 holds for the uniform distribution. In what remains, we prove the lower bound part of this theorem.

The idea behind the proof of the lower bound part of Theorem 3 relies on the intuitive fact that the square configuration in a certain region of a random polyomino should be independent of the square configuration of another region of the same polyomino, provided that there is enough distance between the two non-overlapping regions.

More specifically, we will show that the pattern P depicted in Figure 3.2 occurs a fraction-number-of-the-tiles times in large polyominoes asymptotically almost surely.
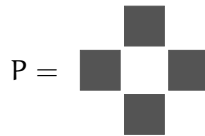

FIGURE 3.2: A pattern with one hole and four tiles.

The pattern P covers an area of five tiles. Observe that this pattern is not a polyomino because it does not have a connected interior. But, it captures the local structure that is necessary and sufficient to have one hole with area of one.

Theorem 4 (The Pattern Theorem on the uniform distribution case)**.** *Given a positive real number* m, *denote by* E^n_m *the set of all polyominoes with* n *tiles and at most* m *copies of* P, *i.e.*

$$E^n_m := \{a \in \mathcal{A}_n \mid a \text{ contains at most } m \text{ translations of } P\}.$$

Let $\lambda = \lim_{n\to\infty} |\mathcal{A}_n|^{\frac{1}{n}}$. *Then, there exists* $\epsilon > 0$ *such that*

$$\limsup_{n\to\infty} |E^n_{\epsilon n}|^{\frac{1}{n}} < \lambda. \tag{3.3}$$

Theorem 4 assures that there exists a fixed $\epsilon > 0$ such that the cardinality of the set of polyominoes that have at most ϵn holes with an area of one and n tiles grows exponentially slower than the total number of polyominoes with n tiles. We observe that the set of n-ominoes without holes is a subset of $E^n_{\epsilon n}$. As we have mentioned before, this theorem (and the proof that we give below) is an adaptation of Theorem B (and its proof) to the uniform distribution, the pattern P, and polyominoes seen as site animals.

We mentioned in Chapter 1 that Klarner's constant λ exists but that its exact value is unknown. We will also use in the proof of Theorem 4 that the sequence $|\mathcal{A}_n|$ is an increasing sequence.

3.2.1   Proof of Theorem 4

In what follows we construct the structures, establish useful notation, and give preliminary results that we use to prove Theorem 4. The two preliminary results are Lemma 11 and Lemma 12 that we state and prove bellow. Proving these lemmas in full detail allows us to compute exact values of C_1 for the inequality in Theorem 3.

We enclose P in the smallest tile configuration that will allow us to place the configuration at any location on a polyomino without disconnecting the polyomino. The configuration D that we show in Figure 3.3 has this property. The tile in black is part of the configuration. We use this black tile to have a canonical way to place the pattern D in a particular tile of a polyomino.

Let A be an n-omino. If we place D at any tile on A, replacing the local configuration of A with this placement, then A is transformed into a polyomino B with an m number of tiles. Clearly, $n - 1 \leq m \leq n + 7$. We define $\kappa = 7$. Then we have $n - \kappa \leq m \leq n + \kappa$. Thus, if we place s non-overlapping copies of D on A, we get a new polyomino with m number of tiles such that $n - s\kappa \leq m \leq n + s\kappa$.

$$D = \quad \blacksquare$$

FIGURE 3.3: Configuration D that contains P. If we place D on a tile t of a polyomino A by placing the black tile of D on t, then the structure of D assures that we get another polyomino. After placing D on A we could get a polyomino with a different number of tiles.

Let α be a positive constant such that: for any n and any n-omino A there exists a set of tiles in A with at least $\lfloor \alpha n \rfloor$ elements, such that if we place $\lfloor \alpha n \rfloor$ copies of the configuration D (one in each one of these $\lfloor \alpha n \rfloor$ tiles) then these copies of D placed in A are non-overlapping.
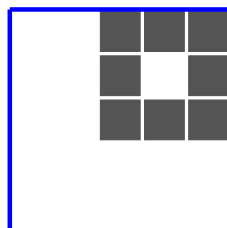
FIGURE 3.4: The region enclosed by the big square covers an area of five-by-five tiles. A sufficient restriction for placing another copy of D in such a way that it does not intersect this copy is that it has to be placed outside this enclosed region. This structure allow us to define $\alpha = 1/25$.

One possible value for α is $1/25$ because, if we place D at some location in a polyomino and we place another copy of D at any tile of the polyomino that is not included in the five-by-five square depicted in Figure 3.4, then the two copies of D

do not overlap.

Fix a number δ (we will choose a specific value for δ later in the proof) such that $0 < \delta < \alpha$. We know that there exists a set $T(A)$ of different tiles in A with cardinality $\lfloor \alpha n \rfloor$ such that if we place $\lfloor \alpha n \rfloor$ copies of the configuration D (one in each one of these $\lfloor \alpha n \rfloor$ tiles) then these copies of $D$ placed in A are non-overlapping. Now, from $T(A)$ we select an ordered tuple $W = (w_1, ..., w_{\lfloor \delta n \rfloor})$ with $\lfloor \delta n \rfloor$ tiles. Then, we construct $\lfloor \delta n \rfloor$ polyominoes as follows: let $A_0 = A$ and A_i be the polyomino obtained after placing the pattern D on the w_i tile in the polyomino A_{i-1} for $1 \leq i \leq \lfloor \delta n \rfloor$. Let $H = A_{\lfloor \delta n \rfloor}$. We observe that H has between $n - \lfloor \delta n \rfloor \kappa$ and $n + \lfloor \delta n \rfloor \kappa$ tiles. For an n-omino A we have constructed a triple (A, H, W) by choosing a particular $T(A)$ and an ordered $\lfloor \delta n \rfloor$-tuple $W \subset T(A)$. It is important to notice that the set of tiles W in A that we have selected are also tiles in H.

Let $\epsilon > 0$ (we will choose ϵ to be $\delta/2$ later in the proof) and let $\mathcal{H}$ be the collection of all triples (A, H, W), constructed as we described before but restricting to all $A \in E^n_{n\epsilon}$ (recall that if $A \in E^n_{n_\epsilon}$ then A has at most $\lfloor \epsilon n \rfloor$ translations of the pattern P). Denote by $\mathcal{H}_H$ the set of all polyominoes H such that there exists a triple $(A, H, W) \in \mathcal{H}$. Given two polyominoes $A \in E^n_{\epsilon n}$ and $H \in \mathcal{H}_H$ we say that A is associated to H if there exists a triple (A, H, W) in the set $\mathcal{H}$.

Clearly we can construct more than one triple (A, H, W) for any given polyomino $A \in E^n_{\epsilon n}$. In particular, once we fix A and the set $T(A)$, there are

$$\frac{\lfloor \alpha n \rfloor !}{(\lfloor \alpha n \rfloor - \lfloor \delta n \rfloor)!}, \tag{3.4}$$

different ways of choosing W from $T(A)$ (for some polyominoes there could also be more than one way to choose the set $T(A)$).

Lemma 11. *Let $E^n_{\epsilon n}$ be the set of all n-ominoes that have at most $\lfloor \epsilon n \rfloor$ copies of P. Then we have*

$$\mid \mathcal{H} \mid \geq \mid E^n_{\epsilon n} \mid \frac{\lfloor \alpha n \rfloor !}{(\lfloor \alpha n \rfloor - \lfloor \delta n \rfloor)!}. \tag{3.5}$$

Proof. For any element A in $E^n_{\epsilon n}$ we know that there exists at least one set $T(A)$ of tiles in A with $\lfloor \alpha n \rfloor$ different tiles such that if we place $\lfloor \alpha n \rfloor$ copies of the configuration D (one in each one of these $\lfloor \alpha n \rfloor$ tiles) then these copies of D placed in A are non-overlapping. Considering all possible ways of selecting the ordered sequence $W = (w_1, ..., w_{\lfloor \delta n \rfloor})$, from (3.4) we conclude (3.5). $\square$

Observe that Lemma 11 is giving us an upper bound for $\mid E^n_{\epsilon n} \mid$. This is convenient for proving inequality (3.3), but we still need more information on this upper bound before we are able to prove Theorem 4. Our next goal is to get an upper bound for $\mid \mathcal{H} \mid$ that will give us the upper bound that we want for $\mid E^n_{\epsilon n} \mid$.

For a fixed $H \in \mathcal{H}_H$ denote by $\mathcal{H}_{(H, \cdot)}$ the set of all possible ordered tuples of tiles $W = (w_1, ..., w_{\lfloor \delta n \rfloor})$ such that there exists an $A \in E^n_{\epsilon n}$ with $(A, H, W) \in \mathcal{H}$. Now, given a $W \in \mathcal{H}_{(H, \cdot)}$, we denote by $\mathcal{H}_{(\cdot, H, W)}$ the set of all polyominoes $A \in E^n_{\epsilon n}$ such that (A, H, W) is in $\mathcal{H}$. In what follows we find an upper bound for the cardinality of the sets $\mathcal{H}_{(H, \cdot)}$ and $\mathcal{H}_{(\cdot, H, W)}$ for any $H \in \mathcal{H}_H$ and all possibles tuples $W \in \mathcal{H}_{(H, \cdot)}$.

We represent by q the number of possible ways in which P and D could intersect without covering the hole at D or P. By construction, P is fully contained in D in a unique way (we are only allowed to place a configuration in a polyomino by translations). Also, there are 12 other possible intersections of P and D. The tile at the right-most column of P could intersect D in any tile of D located in its left-most column. Each one of the four tiles of P has an analogous way of intersecting D. Then $q = 13$.

Thus, for any n-omino A, if A has k translates of P, then we can be sure that A_1, which is obtained from A after placing D on a specific tile in A, has at most $k + q$ translates of P. From this fact and because $A \in E_{\epsilon n}^n$, using a recursive argument, if $H \in \mathcal{H}_H$, then H has at most $\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor$ translates of P. This implies that if $H \in \mathcal{H}_H$ and we want to find an upper bound for the cardinality of the set of possible $\lfloor \delta n \rfloor$-tuples of locations at which we have placed copies of D for getting H from an $A \in E_{\epsilon n}^n$, i.e. an upper bound for $| \mathcal{H}_{(H, \cdot)} |$, we have that

$$| \mathcal{H}_{(H, \cdot)} | \leq \frac{(\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor)!}{(\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor - \lfloor \delta n \rfloor)!}, \tag{3.6}$$

for any fixed $H \in \mathcal{H}$.

Let Z be the number of all square configurations that are subsets of the configuration obtained by adding the middle tile to D (the complete three-by-three square configuration). Clearly $Z = 2^9$. Then, for a polyomino $H \in \mathcal{H}_H$ and an element $W \in \mathcal{H}_{(H, \cdot)}$ there exist at most $Z^{\lfloor \delta n \rfloor}$ elements in $\mathcal{H}_{(\cdot, H, W)}$. We have all the ingredients that we need to give a useful upper bound for $| \mathcal{H} |$.

Lemma 12. *Let E_m^n be as in Theorem 4, and let $\mathcal{H}$ be the collection all triples of the form (A, H, W) with $A \in E_{\epsilon n}^n$. Then we have*

$$| \mathcal{H} | \leq (n + 2\kappa \lfloor \delta n \rfloor) | \mathcal{A}_{n + \kappa \lfloor \delta n \rfloor} | \frac{\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor!}{(\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor - \lfloor \delta n \rfloor)!} Z^{\lfloor \delta n \rfloor}. \tag{3.7}$$

Proof. (Lemma 12) For a fixed $H \in \mathcal{H}_H$ let S_H be the set of all triples of the form (A, H, W) such that $(A, H, W) \in \mathcal{H}$. From the previous discussion about Z and inequality (3.6) we get

$$
\begin{aligned}
| \mathcal{H} | &\leq \sum_{H \in \mathcal{H}_H} S_H \\
&\leq \sum_{H \in \mathcal{H}_H} \sum_{W \in \mathcal{H}_{(H, \cdot)}} | \mathcal{H}_{(\cdot, H, W)} | \\
&\leq \sum_{H \in \mathcal{H}_H} \frac{(\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor)!}{(\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor - \lfloor \delta n \rfloor)!} Z^{\lfloor \delta n \rfloor},
\end{aligned}
$$

which implies

$$| \mathcal{H} | \leq | \mathcal{H}_H | \frac{(\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor)!}{(\lfloor \epsilon n \rfloor + q \lfloor \delta n \rfloor - \lfloor \delta n \rfloor)!} Z^{\lfloor \delta n \rfloor}. \tag{3.8}$$

Observing that any $H \in \mathcal{H}_H$ must be an m-omino with $n - \lfloor \delta n \rfloor \kappa \leq m \leq n + \lfloor \delta n \rfloor \kappa$ and that the sequence $|\mathcal{A}_n|$ is increasing, we get

$$|\mathcal{H}_H| \leq \sum_{j=n-\kappa\lfloor\delta n\rfloor}^{j=n+\kappa\lfloor\delta n\rfloor} |\mathcal{A}_j| \leq (n + 2\kappa\lfloor\delta n\rfloor) |\mathcal{A}_{n+\kappa\lfloor\delta n\rfloor}|. \tag{3.9}$$

Finally, from (3.8) and (3.9) we conclude that (3.7) holds true. $\square$

We have all the ingredients that we need for the proof of Theorem 4 and computing a numeric value for ϵ in inequality (3.3).

Proof. (Theorem 4) Combining inequalities (3.5) and (3.7), we have

$$(n + 2\kappa\lfloor\delta n\rfloor) |\mathcal{A}_{n+\kappa\lfloor\delta n\rfloor}| \frac{(\lfloor\epsilon n\rfloor + q\lfloor\delta n\rfloor)!(\lfloor\alpha n\rfloor - \lfloor\delta n\rfloor)!}{(\lfloor\epsilon n\rfloor + q\lfloor\delta n\rfloor - \lfloor\delta n\rfloor)!\lfloor\alpha n\rfloor!} Z^{\lfloor\delta n\rfloor} \geq |E_{\epsilon n}^n|. \tag{3.10}$$

Then, taking n-th roots on both sides of inequality (3.10), using Stirling's formula, making $n \to \infty$, and substituting

$$\lambda^{1+\kappa\delta} = \lim_{n\to\infty}\left[|\mathcal{A}_{n(1+\kappa\delta)}|^{\frac{1}{n(1+\kappa\delta)}}\right]^{1+\kappa\delta},$$

we get

$$\lambda^{1+\kappa\delta}Z^{\delta}\frac{(\epsilon+q\delta)^{\epsilon+q\delta}(\alpha-\delta)^{\alpha-\delta}}{(\epsilon+q\delta-\delta)^{(\epsilon+q\delta-\delta)}\alpha^{\alpha}} \geq \limsup_{n\to\infty}|E_{\epsilon n}^n|^{\frac{1}{n}}. \tag{3.11}$$

Because $0 < \delta < \alpha$ there exists a $t \in (0,1)$, which we specify later, such that $\delta = t\alpha$. Also, we let $\epsilon = \delta/2$. Then, from inequality (3.11) we obtain

$$\lambda^{\kappa\delta}Z^{\delta}\left(\frac{(\frac{1}{2}+q)^{\frac{1}{2}+q}}{(\frac{1}{2}+q-1)^{\frac{1}{2}+q-1}}\right)^{\delta}(t^t(1-t)^{1-t})^{\alpha} \geq \frac{\limsup_{n\to\infty}|E_{\epsilon n}^n|^{\frac{1}{n}}}{\lambda}. \tag{3.12}$$

Now, we carefully choose the value of t to allow us conclude the proof. Let

$$Q = \lambda^{\kappa}Z\frac{(\frac{1}{2}+q)^{\frac{1}{2}+q}}{(q-\frac{1}{2})^{q-\frac{1}{2}}}. \tag{3.13}$$

Then we can rewrite inequality (3.12) as

$$(t^t(1-t)^{1-t}Q^t)^{\alpha} \geq \frac{\limsup_{n\to\infty}|E_{\epsilon n}^n|^{\frac{1}{n}}}{\lambda}. \tag{3.14}$$

Inequality (3.14) holds for any $t \in (0,1)$. By setting $t = 1/(1+Q)$ we get

$$\left(\frac{Q}{1+Q}\right)^{\alpha} \geq \frac{\limsup_{n\to\infty}|E_{\epsilon n}^n|^{\frac{1}{n}}}{\lambda}.$$

Finally, because $Q/(1+Q) < 1$ we conclude that

$$\lambda > \limsup_{n\to\infty}|E_{\epsilon n}^n|^{\frac{1}{n}}.$$

$\square$

Polyominoes with n tiles without holes form a subset of $E_{\epsilon n}^n$ (the set of all polyominoes that have at most ϵn holes with an area of one and n tiles). Therefore, we get the following corollary from Theorem 4.

Corollary 3. *The number of polyominoes with holes grows exponentially faster than the number of polyominoes without holes.*

Now, we can restate these results in terms of the probability of the event of a polyomino with n tiles having more than ϵn holes.

Corollary 4. *Let $H_{\epsilon n}$ be the event that an n-omino has more than ϵn holes. Then*

$$\lim_{n \to \infty} \mathbb{P}[H_{\epsilon n}] = 1. \tag{3.15}$$

Proof. For any natural number n we have $(E_{\epsilon n}^n)^C \subset H_{\epsilon n}$, which implies

$$\mathbb{P}[H_{\epsilon n}] \geq \mathbb{P}[(E_{\epsilon n}^n)^C] = 1 - \mathbb{P}[E_{\epsilon n}^n] = 1 - \frac{|E_{\epsilon n}^n|}{|\mathcal{A}_n|}. \tag{3.16}$$

From Theorem 4 we know that the cardinality of the set $E_{\epsilon n}^n$ grows exponentially slower than the cardinality of the set $\mathcal{A}_n$. Then, we conclude the proof by making $n \to \infty$ in (3.16). $\qquad\square$

3.2.2   Proof of Theorem 3

Now, we complete the proof for Theorem 3 by giving a lower bound for $\mathbb{E}[\beta_1]$.

Proof. (Theorem 3) Let $\mu = \limsup_{n \to \infty} |E_{\epsilon n}^n|^{\frac{1}{n}}$. We know from Theorem 4 that if $\psi := (\mu/\lambda)$ then $\psi < 1$. This implies that there exists a natural number n_0 such that for any $n \in \mathbb{N}$ if $n > n_0$ then

$$\mathbb{P}[(E_{\epsilon n}^n)] \leq \psi^n < \frac{1}{2}. \tag{3.17}$$

Using this bound and Corollary 4, we get for any $n > n_0$ that

$$
\begin{aligned}
\mathbb{E}[\beta_1] &= \sum_{k=0}^{\infty} k \cdot \mathbb{P}[\beta_1 = k] \\
&= \sum_{k=0}^{\lfloor \epsilon n \rfloor} k \cdot \mathbb{P}[\beta_1 = k] + \sum_{k=\lfloor \epsilon n \rfloor + 1}^{\infty} k \cdot \mathbb{P}[\beta_1 = k] \\
&\geq \epsilon n \sum_{k=\lfloor \epsilon n \rfloor + 1}^{\infty} \mathbb{P}[\beta_1 = k] \\
&\geq \epsilon n \mathbb{P}[\beta_1 \geq \lfloor \epsilon n \rfloor + 1] \\
&\geq \epsilon n \mathbb{P}[H_{\epsilon n}] \\
&\geq \epsilon n \left(\frac{1}{2}\right).
\end{aligned}
$$

Thus, $\mathbb{E}[\beta_1] \geq (1/2)\epsilon n$ for any $n > n_0$. $\qquad\square$

Finally, we compute an exact value for C_1 in Theorem 3 by finding the value of ϵ in Theorem 4. Notice that we already gave a value for C_2 in Lemma 10.

Corollary 5. *With the uniform distribution defined on $\mathcal{A}_n$, there exists a natural number n_0 such that for any $n > n_0$*

$$(2.65 \times 10^{-11}) \cdot n \leq \mathbb{E}[\beta_1] < \frac{1}{2} \cdot n. \tag{3.18}$$

The upper bound can be improved by applying the same techniques used to prove Theorem 4 for the pattern P' depicted in Figure 3.5. The reason is that if a polyomino A has $\epsilon'n$ non overlapping copies of P' in its configuration, then the maximum possible number of holes in A is reduced by at least $\epsilon'n$.
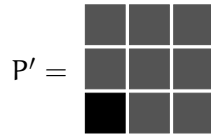
$$P' = $$

FIGURE 3.5: The pattern P'. If a polyomino has this configuration, then it will not reach the maximum number of holes.

For this pattern P', the proof is almost the same as the one given for the pattern P in Theorem 4. The reason is that the corresponding configuration for P' is D' (for P is the configuration D) and D and D' only differ by one tile. This changes only the value of κ (now $\kappa = 8$ instead of $\kappa = 7$) and the value of the constant q (now $q = 25$ instead of $q = 13$) is the value of the constant Q that is defined in (3.13). After computing the value of ϵ for this pattern P', we get the next improvement in the constants given in Corollary 5.

Corollary 6 (An improvement to Corollary 5). *With the uniform distribution defined on $\mathcal{A}_n$, there exists a natural number n_0 such that for any $n > n_0$*

$$(2.65 \times 10^{-11}) \cdot n \leq \mathbb{E}[\beta_1] < \left(\frac{1}{2} - (3.03 \times 10^{-12}) \right) \cdot n. \tag{3.19}$$

Improving these bounds in a significant way will require different techniques than those we have used for proving Theorem 4. But, by performing simulations of random polyominoes, it is possible to conjecture a better value for C_1 in the lower bound for $\mathbb{E}[\beta_1]$, to estimate a better value for C_2, or even conjecture a possible value of the limit

$$\lim_{n \to \infty} \frac{\mathbb{E}[\beta_1]}{n},$$

if the limit exists. In Appendix A we study a method for sampling random polyominoes with distributions converging to the uniform distribution and to the percolation distributions using Markov chain Monte Carlo Metropolis algorithms.

Theorem 3 holds for more general square configurations than P that fulfill certain regularity properties. Also, it is true for other underlying probability distributions defined on $\mathcal{A}_n$ and to polyforms defined in other geometric regular lattices different from $\mathbb{Z}^2$ in the plane and in other dimensions.

Before proving that $\mathbb{E}[\beta_1]$ also has upper and lower bounds that grow linearly with respect to the number of tiles in a polyomino for the percolation distributions defined on $\mathcal{A}_n$, we state in the next section Theorem B. As we mentioned before, this theorem is a generalization of Theorem 3.

3.3   A Pattern Theorem for Lattice Clusters

We begin by defining all the concepts that are involved in Theorem B. In the process, we use the special case of polyominoes in the regular square lattice as an example of the more general definitions.

Polyominoes are defined in the infinite square regular lattice on $\mathbb{Z}^2$. This lattice is an example of a more general family of euclidean space lattices with nice, regular, geometric properties on which the pattern theorem holds. A d-dimensional lattice is any periodic, locally-finite embedding in $\mathbb{R}^d$ of a connected infinite graph. By a locally-finite embedding we mean that any vertex has finite degree and that any bounded subset of $\mathbb{R}^d$ contains a finite number of vertices. With this definition, any locally-finite tessellation of $\mathbb{R}^d$ is a d-dimensional lattice.

If L is a d-dimensional lattice, we denote the set of its vertices by $S(L)$ and the set of its edges by $B(L)$. It is important to notice that both $S(L)$ and $B(L)$ are subsets of $\mathbb{R}^d$. We will refer to the elements of $S(L)$ as the sites of L and to the elements of $B(L)$ as bonds of L. Given any subset $G \subset L$, we denote by $S(G)$ and $B(G)$ the set of sites and the set of bonds of G respectively.

It is clear that in the d-dimensional square regular lattice (referred to for the rest of this section as $\mathbb{Z}^d$) it is possible to translate any subset of sites by a vector with integer entries and get another subset of sites in $\mathbb{Z}^d$. The same is true for any subset of bonds in $\mathbb{Z}^d$. This does not hold if the translation vector has a non-integer entry. Motivated by this translation characterization of the lattice $\mathbb{Z}^2$, we define for any d-dimensional lattice L the translation invariant set $S^*(L)$ of L as all the vectors in $\mathbb{R}^d$ that leave L invariant:

$$S^*(L) := \{u \in \mathbb{R}^d \mid L + u = L\}.$$

The set $S^*(L)$ is a group with 0 as the identity element (the vector with all entries equal to zero). Assuming that 0 is a site in L, we get that $S^*(L)$ is a subset of $S(L)$. In the case of $\mathbb{Z}^d$, these sets are the same; that is, $S^*(L) = S(L)$.

The next step is to define the objects that we want to study in L. In $\mathbb{Z}^2$ we have been studying polyominoes. For each $n \in \mathbb{N}$ we have defined a set $\mathcal{A}_n$ containing all the polyominoes of size n; in the set $\mathcal{A}_n$, we have only included polyominoes that have their lower, left-most column site at the origin. This guarantees that there is a unique element in $\mathcal{A}_n$ representing all polyominoes of the same shape that are equal under any integer translation in the plane.

In any lattice L, a family of clusters is a sequence $\{\mathcal{C}_n\}_{n=1}^{\infty}$ of subsets of finite subgraphs of L such that

- $\mathcal{C}_n \cap \mathcal{C}_m = \varnothing$ whenever $n \neq m$, and

- for every $n \in \mathbb{N}$, if $x, y \in \mathcal{C}_n$, then x is not possible to obtain from y under any $s \in S(L^*)$. In other words, as we have done with the set of polyominoes $\mathcal{A}_n$,

we only include in $\mathcal{C}_n$ one cluster representing all clusters that are equal under any translation contained in $S^*(L)$.

We represent by $\mathcal{C}_{<\infty}$ the set of all clusters

$$\mathcal{C}_{<\infty} := \bigcup_{n=1}^{\infty} \mathcal{C}_n;$$

and we refer to the elements of $\mathcal{C}_n$ as clusters of size n.

In the previous sections of this chapter we have been studying the asymptotic topological properties of $\mathcal{A}_{<\infty}$ with the uniform distribution defined on each A_n. The uniform distribution is related to the function $w : \mathcal{A}_{<\infty} \to (0, \infty)$ that assigns to each cluster $A \in \mathcal{A}_{<\infty}$ the value $w(A) = 1$. Intuitively, we can think of w as a function that is assigning a weight to each cluster contained in $\mathcal{A}_{<\infty}$.

Definition 2. *Let* L *be a lattice with clusters* $C_{<\infty}$ *and* $w : \mathcal{C}_{<\infty} \to (0, \infty)$*. For any given subset* $\mathcal{U}$ *of* $\mathcal{C}_{<\infty}$ *define*

$$W(\mathcal{U}) = \sum_{U \in \mathcal{U}} w(U). \tag{3.20}$$

We denote $W(\mathcal{C}_n)$ *as* W_n *for any natural number* n*. We say that* $w : \mathcal{C}_{<\infty} \to (0, \infty)$ *is a weight function on* $\mathcal{C}_{<\infty}$ *if it has the following two properties:*

1) *For each natural number* m *there exists a positive real number* α_m *such that*

$$\frac{1}{\alpha_m} w(A) \leq w(B) \leq \alpha_m w(A), \tag{3.21}$$

   whenever $A, B \in \mathcal{C}_n$ *such that they differ by at most* m *sites and bonds.*

2) *The limit*

$$\limsup_{n \to \infty} (W_n)^{\frac{1}{n}} \tag{3.22}$$

   exists and is finite.

In practice, one important characteristic to keep in mind when defining a weight function is that it should be invariant under translation. This means that, if we want to naturally extend it to all clusters in L, it should assign the same value to any two congruent clusters (under translations by elements of $S^*(L)$) in the lattice.

Given any weight function we can construct a probability distribution on each $\mathcal{C}_n$, and vice versa, by defining the probability of the singletons as

$$\mathbb{P}[\{A\}] = \frac{w(A)}{\sum_{B \in \mathcal{C}_n} w(B)} = \frac{w(A)}{W_n}, \tag{3.23}$$

for any $A \in \mathcal{C}_n$.

Finally, we give the properties that a pattern in a lattice L needs to fulfill in order to be used in The Pattern Theorem. Let $P_1 \neq \varnothing$ and P_2 be two finite, disjoint subsets of the lattice L. We think of P_1 and P_2 as in a fixed location with respect to each other. If $C \in \mathcal{C}_{<\infty}$ is a cluster, then we say that C contains P if C contains P_1 and does not contain P_2. In Figure 3.6 we show a possible pair P_1 and P_2 that generates the pattern

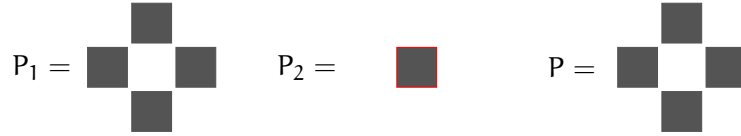P that we used in Theorem 4 and that we will use in Theorem 6.

FIGURE 3.6: P is a pattern with one hole and 4 tiles. In the ordered
pair $P = (P_1, P_2)$, the tile configuration P_1 contains the sites that we
want to be present in a cluster and P_2 contains the sites that we do
not want to be present in a cluster.

We refer to the ordered pair $P = (P_1, P_2)$ as the pattern P.

Definition 3. *Given a pattern* $P = (P_1, P_2)$, *we say that* P *is a proper pattern if it has the
following two properties:*

1 *For any* $n \in \mathbb{N}$ *there exists a cluster* $C \in \mathcal{C}_n$ *such that* P *is contained in* C.

2 *There exists a finite set* D *of sites and bonds such that, for every cluster* $C \in \mathcal{C}_{<\infty}$ *and
every site* s *in* C, *there exists another cluster* C' *and a translation vector* $u \in S^*(L)$
such that $s \in D + u$, C' *contains* $P + u$, *and* C' *has the same structure of* C *outside of*
$D + u$.

The pattern P, depicted in Figure 3.6, has these two properties. The configura-
tion D corresponding to this pattern is depicted in Figure 3.3. As we have seen,
this configuration plays a key role in the proof of Theorem 4 and for computing the
constants that we give in Corollary 5 and Corollary 6. For any given pattern P, the
goal is to find its respective configuration D in such a way that it has the minimum
possible sites and bonds but that it has the two properties mentioned above.

Define for all natural numbers n and m the set of clusters

$$E_m^n := \{a \in \mathcal{C}_n \mid a \text{ contains at most } m \text{ translates of P}\},$$

and remember that we are assuming that the limit

$$\limsup_{n \to \infty}(W_n)^{\frac{1}{n}} = \lambda,$$

exists (because w is a weight function). We have defined all the concepts that we
need to state The Pattern Theorem.

Theorem B (The Pattern Theorem, M. Madras 1999)**.** *Let* L *be a* $d-$*dimensional lattice
with a family of lattice clusters* $C_{<\infty}$ *and* P *a proper pattern in this lattice. If* w *is a weight
function on* $C_{<\infty}$, *then there exists an* $\epsilon > 0$ *such that*

$$\limsup_{n \to \infty}[W(E_m^n)]^{\frac{1}{n}} < \lambda. \tag{3.24}$$

For the proof of this theorem we refer the reader to [23]. We use Theorem B in
the proof of Theorem 5, that we state in the next section, as we have used Theorem
4 to prove Theorem 3. First, we give a precise definition of percolation distributions
over $\mathcal{A}_n$.

3.4 Homology of percolation distributed polyominoes

The percolation site model in the regular square lattice on the plane is described as follows. Choose a probability $p \in (0,1)$; then, include each one of the 2-cells of $CW(\mathbb{Z}^2)$ with probability p and exclude it with probability $1 - p$. In Figure 3.7, we show a simulation of the percolation model with $p = 1/2$.
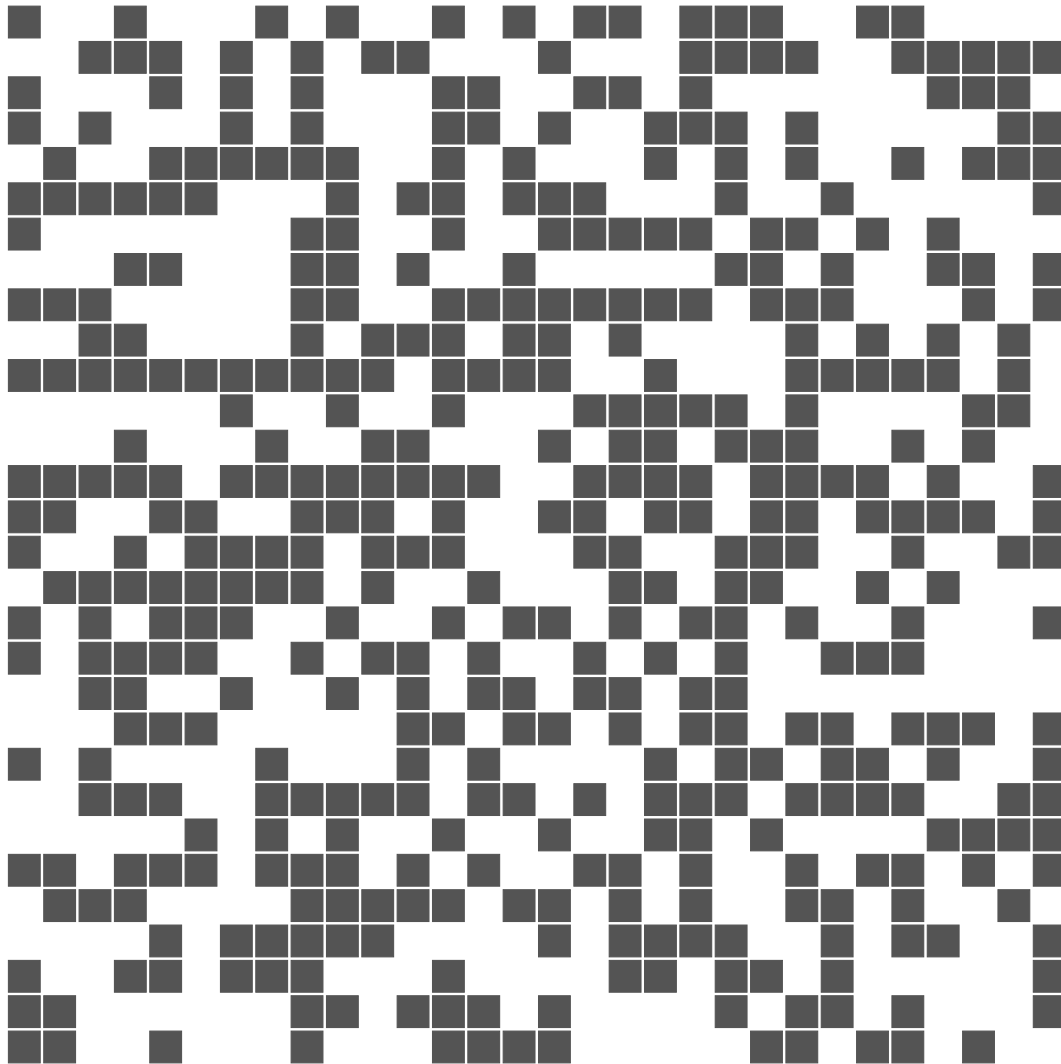

FIGURE 3.7: Percolation model over $CW(\mathbb{Z}^2)$ with $p = .5$. The components with connected interiors are polyominoes. In percolation theory they are refer more commonly as lattice animals or percolation clusters.

A relevant feature of this discrete percolation model is that every microscopic configuration can be separated unambiguously into clusters with connected interiors. These clusters are polyominoes. In the literature, these clusters are more commonly referred to as lattice animals. As we have mentioned before, (site) lattice animals are in one-to-one correspondence with polyominoes, but lattice animals do not capture the topology of polyominoes.

One of the most studied properties of this model is the probability threshold for the appearance of an infinite connected component. For every $p \in (0,1)$ the probability of the event that an infinite polyomino exists in a realization of this percolation model is either zero or one as a consequence of Kolmogorov's zero-one law. Also, a threshold for this event is known. There exists a probability $p_c \in (0,1)$ such that for any $p < p_c$ with a probability of one there does not exist an infinite polyomino in a realization of this percolation model and, if $p > p_c$ with probability one, there exists such an infinite polyomino (lattice animal).

Since the 1970s, different theoretical results and computational experiments of the percolation model on the regular square lattice suggested that there was a qualitative difference in the finite polyominoes in this percolation model above and below the threshold p_c. Some of these properties of the finite clusters (polyominoes) that they studied using computational experiments on this percolation model were density profile, radius, and perimeter [29]. In Figure 3.8, we show a polyomino sampled using a Metropolis–Hasting algorithm that converges to the probability distribution on $\mathcal{A}_{50}$ associated to the percolation model taking $p = p_c$. We analyze the properties of this algorithm in Appendix A.
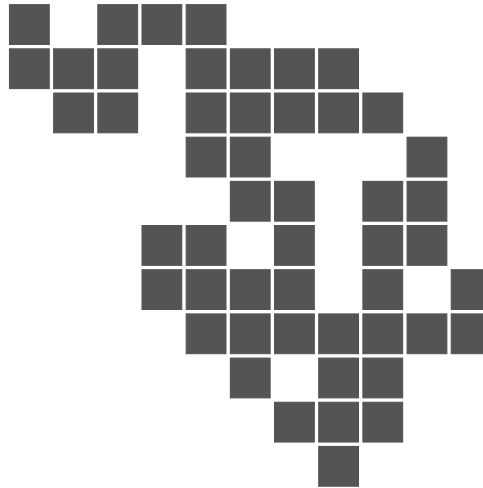

FIGURE 3.8: A polyomino that we sampled from the percolation distribution at criticality $p = p_c$ by implementing and running a Monte Carlo Markov chain Metropolis–Hasting algorithm. This polyomino has 50 tiles and four holes. It is showing a qualitative difference with respect to the polyomino that we have sampled with the uniform distribution that is shown in Figure 3.1. We give more details about this sampling method in Appendix A.

In what follows we give a precise definition of the probability distributions related to this percolation model.

Fix a natural number $n \in \mathbb{N}$ and a probability $p \in (1,0)$. We define the distribution π_p as the finite probability distribution on $\mathcal{A}_n$ that assigns to each element $A \in \mathcal{A}_n$, with site perimeter $t_A = per_{site}(A)$, the probability

$$\pi_p(A) = \frac{p^n(1-p)^{t_A}}{z}, \quad \text{with} \quad z = \sum_{B \in \mathcal{A}_n} p^n(1-p)^{t_B}. \tag{3.25}$$

This percolation distribution assigns different probabilities to polyominoes with different site perimeters. Thus, there does not exist a probability $p \in (0, 1)$ such that the distribution π_p is equal to the uniform distribution defined on $\mathcal{A}_n$—see Figure 3.9. However, the uniform distribution is related to the percolation model distributions because as $p \to 0$ the distribution π_p converges to the uniform distribution.

$$A = \blacksquare\blacksquare\blacksquare\blacksquare \qquad B = \begin{array}{c}\blacksquare\blacksquare\\\blacksquare\blacksquare\end{array}$$

FIGURE 3.9: The percolation distribution π_p assigns different probabilities to polyominoes with different site perimeters. Therefore, these percolation distributions cannot be equal to the uniform distribution. For example, if $p = 1/2$, then $z \cdot \pi_p(A) = 1/16384$ and $z \cdot \pi_p(B) = 1/4094$, with z defined as in (3.25).

We proved in Theorem 3 that with respect to the uniform distribution $\mathbb{E}[\beta_1]$ grows linearly with respect to the number of tiles in a polyomino. This result also holds for the percolation model distributions defined on $\mathcal{A}_n$. This is precisely stated in the next theorem.

Theorem 5. *Let $p \in (0, 1)$. With the percolation distribution π_p defined on $\mathcal{A}_n$, there exist constants C_1 and C_2, not depending on n but depending on p, such that*

$$C_1 \cdot n \leq \mathbb{E}[\beta_1] \leq C_2 \cdot n.$$

For proving this result we follow an analogous proof to the one that we used to prove Theorem 3. We want to apply The Pattern Theorem in its more general version given in Theorem B. Hence, for each distribution π_p, we need to define its related weight function $w_p : \mathcal{A}_{<\infty} \to (0, 1)$ satisfying properties (3.22) and (3.21) given in the previous subsection.

By substituting in (3.23) the probability that the distribution π_p assigns to a polyomino $A \in \mathcal{A}_{<\infty}$ with n tiles and site perimeter $t_A = per_{site}(A)$, we define the value of the function w_p, corresponding to π_p, as

$$w_p(A) := p^n (1-p)^{t_A}. \tag{3.26}$$

For any given subset $\mathcal{A} \subset \mathcal{A}_{<\infty}$ we represent

$$W_p(\mathcal{A}) = \sum_{A \in \mathcal{A}} w_p(A), \tag{3.27}$$

and we denote $W_p(\mathcal{A}_n)$ as $W_p(n)$ for any natural number n.

It is clear from its definition that the function w_p is invariant under translations by any vector with integer entries in the plane because the site perimeter of a polyomino does not change if we translate a polyomino on the plane. The next lemma [23] states that the function w_p has properties (3.22) and (3.21). That is, w_p is a weight function defined on $\mathcal{A}_{<\infty}$.

Lemma A. *Let* $p \in (1, 0)$ *and* $w_p : \mathcal{A}_{<\infty} \to (0, \infty)$ *the function defined as in (3.26). Then* w_p *is a weight function on* $\mathcal{A}_{<\infty}$*.*

From Lemma A and Theorem B, we get the following result.

Theorem 6 (The Pattern Theorem: for the percolation model distribution). *Let* $\lambda_p = \lim_{n\to\infty} W_p(n)^{\frac{1}{n}}$*. Then, there exists* $\epsilon > 0$ *such that*

$$\limsup_{n\to\infty} \mid W(E_{\epsilon n}^n) \mid^{\frac{1}{n}} < \lambda. \tag{3.28}$$

Theorem 5 can be proved using Theorem 6 by following an analogous argument to the one that we used to prove Theorem 3 from Theorem 4. Thus, as in the uniform distribution case, we have proved the existence of linear upper and lower bounds for the expected number of holes in polyominoes if the underlying probability distribution is π_p.

In the next chapter we define and study another probability distribution on the set of polyominoes such that, based on the computational experiments that we have performed, the expectation $\mathbb{E}[\beta_1]$ does not behave linearly with respect to the number of tiles in a polyomino. This distribution is related to a stochastic process defined over the sequence $\mathcal{A}_n$. We will use topological data analysis techniques to analyze the behavior of the variable β_1 as the number of tiles increases in this process.

Chapter 4

Stochastic growth process with polyominoes

In this chapter we study and perform computational experiments on the persistent homology and some geometric features related to the homology of a stochastic process defined on the set of all polyominoes. This process corresponds to a cell growth model called the Eden Growth Model (EGM). It can be described as follows [9, 10]: Start with a polyomino having only one tile and make it grow by adding one tile at a time uniformly at random to its site perimeter, with the restriction that the newly added tile needs to share at least one side with any other tile already present in the polyomino.

Eden mentioned in [10] that the first person to model the cell growth process from a mathematical point of view was A.M. Turing [31] in 1952. However, Turing's cell growth model is not based on polyominoes as, for the most part, he used a one-dimensional structure to study the cell growth problem.

Also, as highlighted in the recent survey paper [1], the EGM corresponds to a site First Passage Percolation model in the regular square lattice in the plane after applying the right time-change (where the time is measured by the number of added tiles). In a personal conversation with one of the authors of [1], A. Auffinger, at ICERM (Brown, University) in Fall 2016, he mentioned that the evolution of the topology of the EGM measured by its homology has not been studied yet and that the same was true for any of the First Passage Percolation models.

This emphasizes the relevance of our third set of contributions that we present in this chapter, which consists of importing from stochastic topology and topological data analysis new techniques to study the topological evolution of First Passage Percolation models. Also, in topological data analysis, the persistent homology has not been used to measure the evolution of the homology and geometric properties related to the homology of a stochastic process defined over polyforms in regular lattices as we do in this thesis.

In Section 4.3 we characterize—see Theorem 7—the possible change in the rank of the first homology group for the stochastic process defined by the EGM. This allowed us to design and implement a new algorithm that computes the persistent homology associated to this stochastic process at each time and that keeps track of geometric features of the homology of the process. We give a precise description of this algorithm in Section 4.4.

We present and analyze the results of the computational experiments that we have made with this algorithm in Section 4.5. One conjecture that we have, based on these experiments, is about the asymptotic behavior of the number of holes.

Conjecture 2. *Let β_1^t be the random variable that measures the rank of the first homology group of the EGM stochastic process at time t. Then for sufficiently large values of t,*

$$C_1 \sqrt[\alpha]{t} \leq \mathbb{E}[\beta_1^t] \leq C_2 \sqrt[\alpha]{t}, \tag{4.1}$$

for some constants $C_1, C_2 > 0$ and $\alpha \geq \frac{1}{2}$. We suspect that $\alpha = .5$, $C_1 > 1$, and $C_2 < 1.5$.

4.1   Preliminary definitions and notation

Remember that polyominoes are subsets with connected interiors of $\mathbb{R}^2$ that are formed by the closed tiles of the infinite checkerboard. We are considering these tiles to have an area of 1 and their vertices to be the elements of the set

$$\mathbb{Z}^2 := \{(x, y) \in \mathbb{R}^2 \mid n, m \in \mathbb{Z}\}.$$

Given two polyominoes, $A \in \mathcal{A}_n$ and $B \in \mathcal{A}_m$, we say that A is a subpolyomino of B if $A \subset B$ as a subset of $\mathbb{R}^2$, and we say that A is a related subpolyomino of B if there exists a polyomino $A' \in \mathcal{A}_n$ such that $A \sim A'$ and A' is a subpolyomino of B.

By the canonical closed unit square we refer to the polyomino given by the square of area 1 with vertices at points $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$. We will use the variable t for measuring the time of the EGM stochastic process. This time t is the same as the number of tiles that the EGM has at each time.

Polyomino B   Polyomino A

FIGURE 4.1: In both polyominoes the yellow square represents the canonical unit square. Polyomino A is a subpolyomino of B. If we place the canonical unit square of only one of the polyominoes on a different gray tile, then it is no longer true that polyomino A is a subpolyomino of B.

We denote the area of a polyomino A by $a(A)$. Then, $a(A) = n$ if A is an n-omino. We also want to measure the area covered by the holes of a polyomino. We denote by $a_h(A)$ the total area covered by all the holes in A and by $a_h^{max}(A)$ the maximum area over all holes in A. The number of holes with area $m \in \mathbb{N}$ in a polyomino will be denoted by $a_h^m(A)$.

4.2   (Algebraic) Topological definition of the EGM

With the definitions given in Section 3.1, we can define the stochastic process of the EGM from an algebraic topological point of view as follows.

Definition 4. *We define the EGM stochastic process as the sequence $\mathcal{G} = \{G_t\}$ of random variables such that each G_t takes values in the set of all polyominoes (on $CW(\mathbb{Z}^2)$) with $t + 1$ 2-cells, such that*

a) *G_0 is the subcomplex defined by the canonical closed unit square.*

b) *G_1 selects uniformly at random among all the polyominoes that have as subcomplex G_0 and have two 2-cells.*

c) *In general, for all $t > 1$, G_t selects uniformly at random among all the polyominoes that have as subcomplex G_t and that have t 2-cells.*

Every polyomino A can be regarded as a planar graph that we represent by $G(A)$ embedded in $\mathbb{R}^2$ with A^0 as the set of all vertices of $G(A)$ and A^1 as the set of all edges of $G(A)$. The faces of $G(A)$ contain the elements of A^2, the infinite face, and the faces formed by the holes contained in A that will exist whenever A is not simply connected. We denote by $v(A)$, $e(A)$, and $f(A)$ the cardinalities of these sets of vertices, edges, and faces of $G(A)$, respectively.

It is important to notice that this planar graph $G(A)$ associated to a polyomino A is different from the planar dual graph and the site animal (which is also a graph) defined in Section 2.6. Also, we are using a different meaning for the function f from the one that we used in Chapter 2. But, fear not, because they have a different domain.

4.3 Behavior of β_1 of the EGM

Euler's formula for planar graphs gives a relation between the vertices, edges and faces of a planar graph G embedded in $\mathbb{R}^2$. It states that the number of vertices, minus the number of edges, plus the number of faces of G (including the infinite face), equals two for any possible injective embedding of G in the plane.

Then, as a consequence of Euler's formula, for all polyominoes $A \in \mathcal{A}_n$ we have $v(A) - e(A) + f(A) = 2$, which implies $v(A) - e(A) + (n + \beta_1(A) + 1) = 2$. Finally, we get the equation

$$\beta_1(A) = e(A) - n - v(A) + 1. \tag{4.2}$$

Equation (4.2) and Lemma 4.3 are important ingredients for the algorithm that we designed and implemented for computing $\beta_1^t(G_t)$ in simulations of the EGM as defined in Definition 4 above. For stating in a more precise way what topological information we want to measure in the EGM simulations, we need first to establish the notation we are going to use.

For each simulation $\{G_t(\omega)\}$ of the EGM stochastic process, we can construct the sequence of random variables $\boldsymbol{\beta_0} = {}^{\beta_0^0, \beta_0^1, \beta_0^2 \cdots}$, and $\boldsymbol{\beta_1} = \beta_1^0, \beta_1^1, \beta_1^2 \cdots$, where $\beta_i^t = \dim[H_i(G_t(\omega))]$ for $0 \leq i \leq 1$, and $0 \leq t \leq \infty$. All other sequences corresponding to higher dimensional homology groups are trivial because our CW-complexes are embedded in $\mathbb{R}^2$. We remind the reader that for each polyomino A the sequence of homology vector spaces over the field $\mathbb{F}_2$ is denoted by $\{H_i(A)\}_{i=0}^{\infty}$. The dimensions of these vector spaces have the topological information that allows us to count the number of holes in A.

Because at each time G_t has only one component, $\mathbb{P}[\beta_0^t(G_t) = 1]$ is equal to 1 for all $t \in \mathbb{N}$. This means that the topological information of the EGM process is contained in the sequence $\boldsymbol{\beta_1} = \beta_1^0, \beta_1^1, \beta_1^2....$ Hence, whenever possible, for the rest of this chapter, we will suppress the superscript t from the random variables β_1^t and use the notation β_1.

We have observed in Chapter 2 that the set of polyominoes (or polycomplexes), with less than 7 tiles and at least one hole, is empty. Then, for all $0 \leq i \leq 6$

$$\mathbb{P}[\beta_1(G_i) = 0] = 1.$$

If the same constructions that we have made for $CW(\mathbb{Z}^2)$ are made for the d-dimensional cubical lattice $\mathbb{Z}^d$, then the sequences corresponding to all homology groups of dimensions $0 \leq k \leq d-1$ will keep track of all the interesting changes of the topology of the stochastic process (again, G_t having only one component at each t).

In the following theorem we give a characterization of the possible differences between two consecutive elements of $\boldsymbol{\beta_1}$.

Theorem 7. *Let $\boldsymbol{\beta_1}$ be obtained as an outcome of a simulation of the EGM stochastic process. Then, for all $t > 1$, there are only four possible values for β_1^t in terms of β_1^{t-1}: $\beta_1^t = \beta_1^{t-1}$, $\beta_1^t = \beta_1^{t-1} - 1$, $\beta_1^t = \beta_1^{t-1} + 1$, or $\beta_1^t = \beta_1^{t-1} + 2$.*

Before proving Theorem 7, we need to characterize the way in which a cell can be added to the polyomino from one step of the process to the next one. Then, we prove that in each one of these possible cases Theorem 7 holds. We need first to provide some definitions.

Given a cell a in a polyomino A we define the 1-neighborhood of a, that we denote by $N_1(a)$, as the subset of cells of the infinite complex $CW(\mathbb{Z}^2)$ that have a non-empty intersection with a. Observe that $N_1(a)$ has a and eight other cells different from a in it, see Figure 4.2.
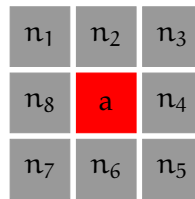
n_1	n_2	n_3
n_8	a	n_4
n_7	n_6	n_5

FIGURE 4.2: The nine cells that form the 1-neighborhood of a cell a with the corresponding notation.

Given a polyomino A and a 2-cell a not in A, it will be useful to consider the set $A^2 \cap N_1(a)$ that contains all the 2-cells that are part of A and, at the same time, are in the 1-neighborhood of a. Observe that a tile a that is not contained in A is on the site perimeter of A if and only if $A^2 \cap N_1(a) \neq \varnothing$. The proof of the next result involves a tedious counting of all possibilities of the set of tiles $A^2 \cap N_1(a)$.

Lemma 13. *Suppose that only one 2-cell, denoted by a, is added to a polyomino A in such a way that the new structure B is again a polyomino. In the eight graphic representations*

of the intersection $A^2 \cap N_1(A)$ *depicted in Figure 4.3, let the green cells (they also have a black dot inside) represent cells that might or might not be present in A, let the gray cells be present in A, let the red cell to be* a*, and let the white cells (the rest of the cells) be those not present in A. Then, ignoring the symmetries of the regular square tessellation of the plane, one and only one of the nine possible cases shown in Figure 4.3 holds for* $A^2 \cap N_1(a)$*.*

Proof. (Theorem 7)

Let $\boldsymbol{\beta_1} = (\beta_1^0, \beta_1^1, \beta_1^2...)$ be a sequence obtained as an outcome of a simulation $\{G_t(\omega)\}$ of the EGM stochastic process. Suppose that at time t one tile a on the site perimeter of $G_{t-1}(\omega)$ was added to the polyomino $G_{t-1}(\omega)$ (as described in Definition 4). Then, by Lemma 13, one and only one of the eight cases depicted in Figure 4.3 holds. The rest of the proof relies on Euler's formula, stated in Equation (4.2), as follows.

We start by analyzing in detail the difference between β_1^t and β_1^{t-1} for Case 4, and for the remaining cases, the proof can be completed in an analogous way.

- Case 4.1: In this case no vertices or edges or faces are added to the graph $G(G_{t-1}(\omega))$. This implies that $v(G_t(\omega)) = v(G_{t-1}(\omega))$, $e(G_t(\omega)) = e(G_{t-1}(\omega))$, and $f(G_t(\omega)) = f(G_{t-1}(\omega))$. From Equation (4.2), we have

$$\beta_1(G_{t-1}(\omega)) = e(G_{t-1}(\omega)) - t - v(G_{t-1}\omega) + 1, \tag{4.3}$$

and

$$\beta_1(G_t(\omega)) = e(G_t(\omega)) - (t+1) - v(G_t\omega) + 1. \tag{4.4}$$

Then, from equations (4.3) and (4.4), we get

$$\beta_1(G_t(\omega)) - \beta_1(G_{t-1}(\omega)) = -1. \tag{4.5}$$

- Case 1.1, 1.2, and 1.3: For these cases we get that

$$\beta_1(G_t(\omega)) - \beta_1(G_{t-1}(\omega)) = 0. \tag{4.6}$$

- Cases 2.1, 2.2, and 2.3: For these cases, after examining each one, we get that

$$\beta_1(G_t(\omega)) - \beta_1(G_{t-1}(\omega)) = 1. \tag{4.7}$$

- Case 3.1: In this case, we are adding three edges and one face. This implies

$$\beta_1(G_t(\omega)) - \beta_1(G_{t-1}(\omega)) = 2. \tag{4.8}$$

From equations (4.5) to (4.8), we conclude that $-1 \leq \beta_1^t - \beta_1^{t-1} \leq 2$, and this is true for all $t \in (\mathbb{N} \cap \{0\})$.

$\square$

4.4 Computational experiments: the algorithms

In this section we describe the algorithms that we have implemented to run our simulations of the EGM stochastic process and to keep track of the homology at each time in the process.

Case 1.1

Case 1.2

Case 1.3

Case 2.1

Case 2.2

Case 2.3

Case 3.1

Case 4.1
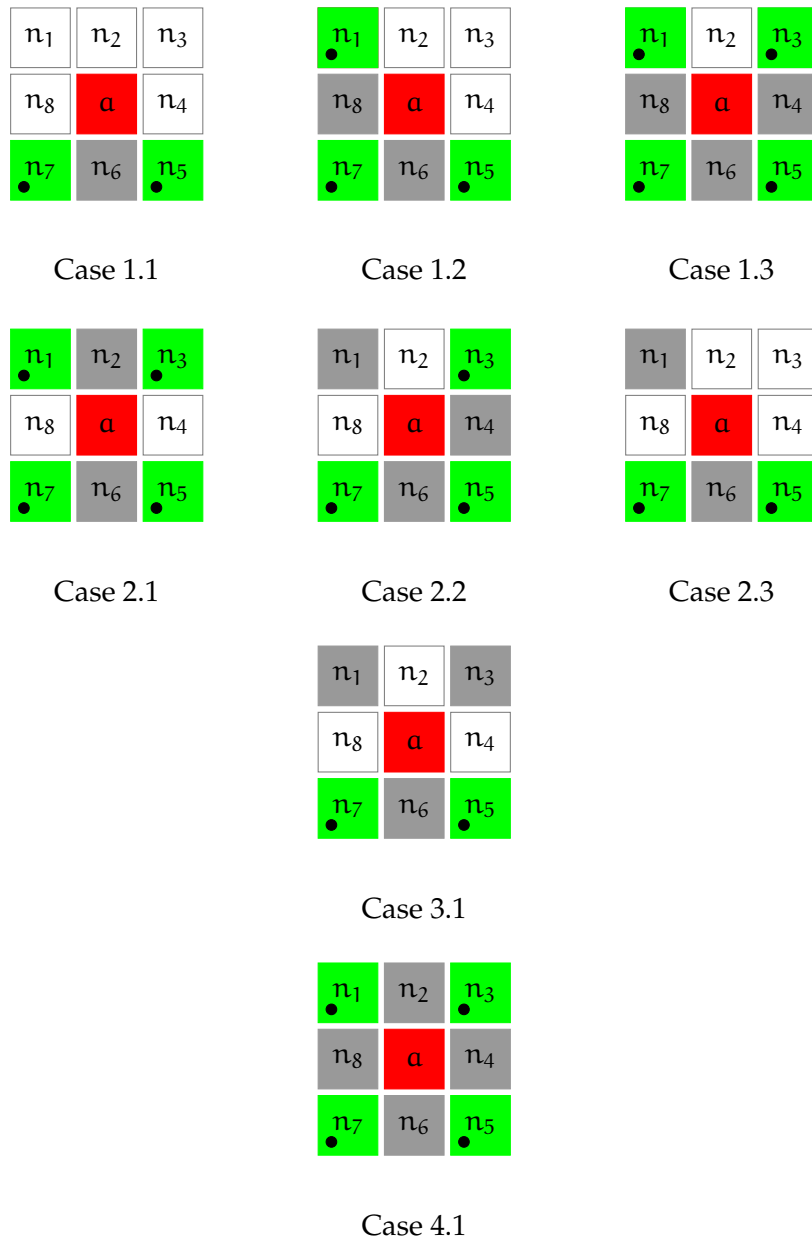
FIGURE 4.3: All possibilities for $A^2 \cap N_1(\mathfrak{a})$.

It is desirable to give, for any computational mathematical experiment, a detailed report describing the experiments in order to make it possible to reproduce them and also to state in detail the conditions and algorithms under which the computational experiments were done. Because of this, in what follows we describe in detail the algorithms that we have implemented.

We have different algorithms depending on which variables we are interested in measuring. This allows us to sample and simulate the EGM for very large numbers of cells. In particular, there is a difference between the algorithm that we use to keep track of the rank of the first homology group at each iteration and the algorithm that we have implemented for generating the persistent homology diagrams and barcodes.

We designed and implemented some of the algorithms in R [25], which is a language and environment for statistical computing. We have also conducted the statistical analysis of the results using R. We designed and implemented the algorithms related to the persistent homology of the EGM in Python.

4.4.1 The algorithm for measuring the rank of the first homology group

We have made simulations of the EDG stochastic process using the algorithm:

Step 0) Select a natural number N and set the counter $r = 0$.

Step 1) To generate G_0 we start with the canonical closed unit square. At this step $\beta_1^0 = 0$. Set $r = 1$.

Step 2) To generate G_1, add a new cell to the site perimeter of the canonical closed unit square, choosing uniformly at random one cell contained in the site perimeter of the canonical closed unit square. As we have discussed in Section 3.1, at this step $\beta_1^1 = 0$.

Step 3) In order to generate G_{k+1}, for $k \geq 1$, add a new cell a to the site perimeter of G_k, choosing a uniformly at random from the set of all cells contained in the site perimeter of G_k.

Step 4) Determine in which case of Figure 4.3 is the structure of $N_1(a) \cap G_k$.

Step 5) Based on the result obtained in Step 4, find the value of β_1^{k+1} by applying Theorem 7. Set $r = r + 1$.

Step 6) If $r < N$, return to Step 3). Otherwise, finish the process.

Observe that at each time t it is possible to keep track of the values of $v(G_t)$, $e(G_t)$ and $f(G_t)$, depending on the result of Step 4). In our algorithm, we have calculated the Betti numbers based on these quantities and Euler's Formula following the results of Theorem 7.

From Step 4) it is also possible to know the perimeter of G_t at each time, and this gives us the number of sites that are on the topological boundary of G_t. We are also able to keep track of the areas of the holes at each time during the process. Alas, this takes more computational time than restricting the algorithm to compute only the change in the homology at each time.

4.5   Obtained results of the computational experiments

4.5.1   On the Evolution and Asymptotic Behavior of the Homology of the Process

The next figure shows a log-log plot. The zlog axis shows the number of tiles and the xlog axis shows β_1^t. For each time t (that is equal to the number of tiles), we ran a simulation of the EGM up to time t and computed β_1^t for G_t. For each t, a new simulation was made. The vector of t was from $t = 5,000$ to $t = 10,000$, increasing by 100 between any two consecutive t. It is important to consider that only one simulation was plotted for each t. Then, on the log-log plot, a line was fit to the point cloud obtaining a slope of .56334 and an intercept of 2.08344.
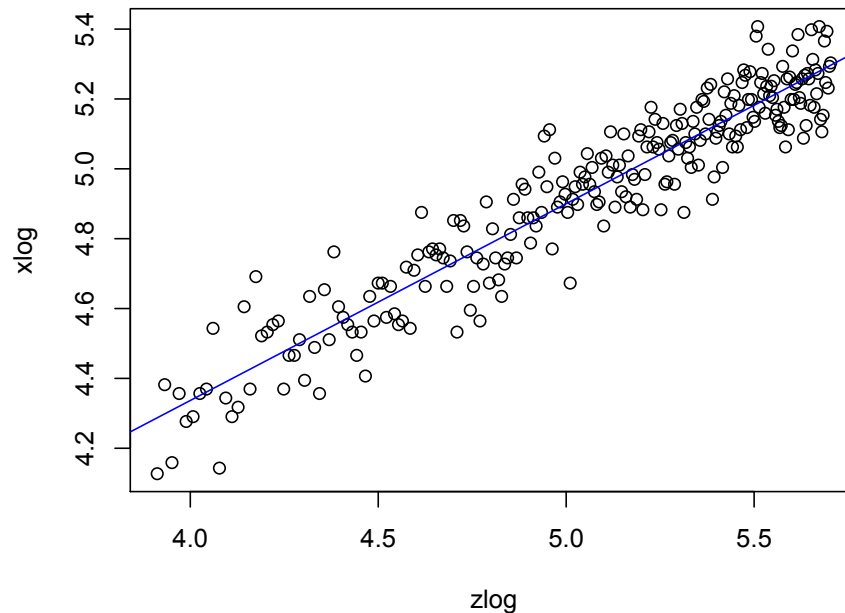
The adjustment was statistically significant. The statistical details of the adjustment are summarized in the next figure.

```
> summary(lm(xlog~zlog))

Call:
lm(formula = xlog ~ zlog)

Residuals:
      Min        1Q    Median        3Q       Max
-0.237333 -0.057748  0.004449  0.066413  0.256321

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.08344    0.06183   33.70   <2e-16 ***
zlog         0.56334    0.01216   46.32   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09282 on 249 degrees of freedom
Multiple R-squared:  0.896,     Adjusted R-squared:  0.8956
F-statistic:  2145 on 1 and 249 DF,  p-value: < 2.2e-16
```

In another experiment, we fixed t for $t = 10,000$ and we made 100 simulations of the Eden Growth Model up to time t. For each one of the simulations we calculated $\beta_1^t = \beta_1^{10,000}$. The next histogram was made using these results.

The mean was 109.17
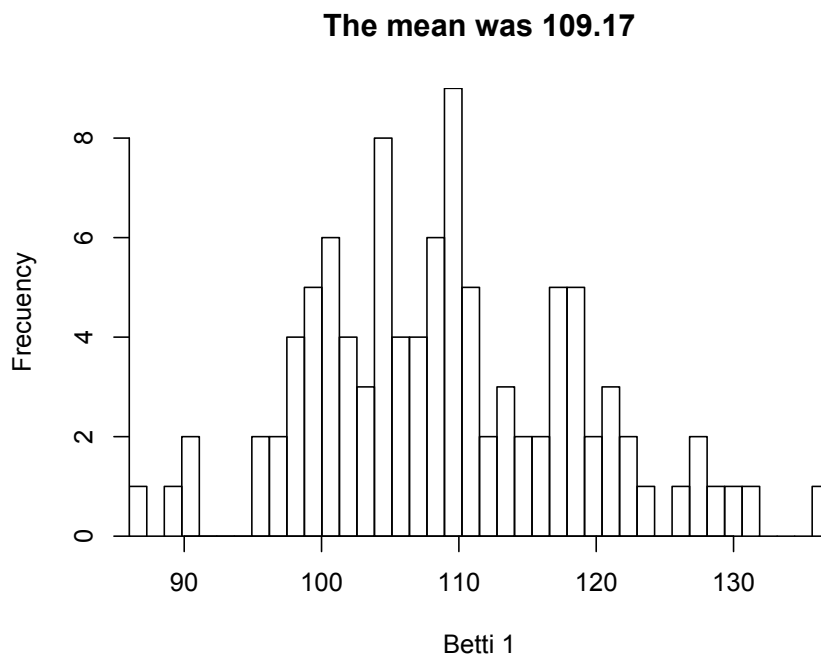
We observed a mean of 109.17, which is a little bit bigger than $\sqrt[2]{10,000}$—see Conjecture 2.

We then repeated the experiment, but this time we ran 1,000 simulations for $t = 10,000$. The next histogram shows the results of these simulations. The mean that we obtained this time was 109.7 which is almost the same as the mean we obtained from the experiment with only 100 repetitions of the simulation.

Histogram of x
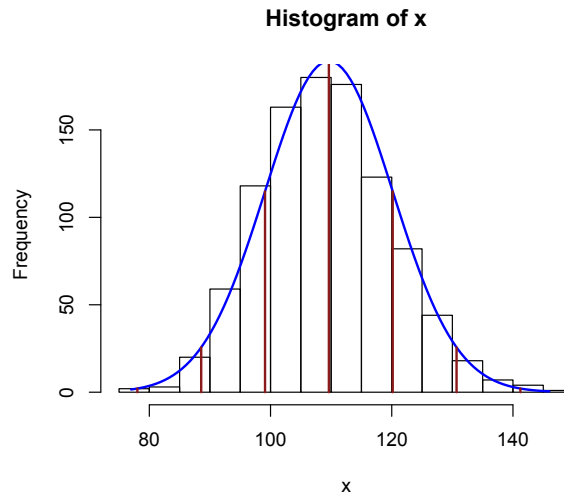
We also made a simulation for $t = 1,000,000$, and the number of holes at the end of this simulation came to $\beta_1 = 1227$. These results suggest that a Central Limit Theorem must hold for $\mathbb{E}[\beta_1]$.

4.5.2   On the change of β_1

We have proved in Theorem 7 that at each time a cell is added, we have one and only one of the following possibilities regarding changes in the homology of the process:

- $\beta_1^{t+1} = \beta_1^t$,

- $\beta_1^{t+1} = \beta_1^t - 1$,

- $\beta_1^{t+1} = \beta_1^t + 1$,

- $\beta_1^{t+1} = \beta_1^t + 2$.

As we have seen in Section 4.4.1, the algorithm that we have implemented can keep track of how the Betti numbers (the number of holes in the polyominos) are changing in time. This implies that, at the end of the process, we can know how many occurrences there have been of each one of these four options.

In the next figure, we show a simulation of the EGM for up to $1,000$ tiles that shows the change of β_1 at each time t for $1 \le t \le 1,000$.
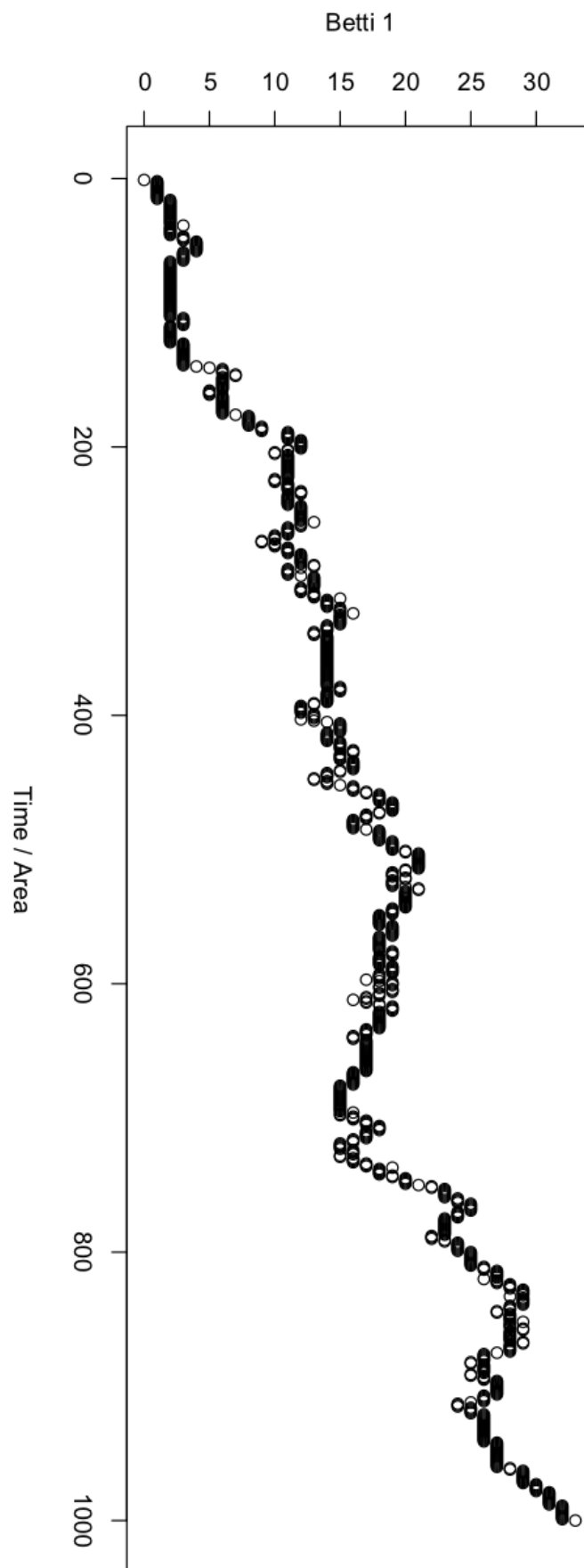
In Table 4.1, we summarize the results that we obtained in terms of the differences between β_1^t and β_1^{t-1} in simulations of the EGM for different t. The column labeled as i, for $i \in \{-1, 0, 1, 2\}$, has the number of times when $\beta_1^t - \beta_1^{t-1} = i$.

TABLE 4.1

t	-1	0	1	2	Total of Holes	Last time t, when $\beta_1^t = 0$
1000	70	832	95	1	27	
2000	187	1581	221	9	52	
3000	297	2361	331	9	52	
5000	530	3887	561	20	71	
5000	536	3883	554	25	68	
5000	508	3928	539	23	67	
6000	660	4639	666	33	72	
6000	633	4685	658	22	69	
6000	628	4690	656	24	76	
10,000	1151	7628	1162	58	127	
10,000	1153	7650	1129	67	110	
10,000	1158	7636	1143	62	109	
20,000	2374	15203	2311	111	159	19
20,000	2359	15204	2340	96	173	19
20,000	2360	15186	2369	84	177	33
1,000,000					1227	

We conjecture from these computational experiments that, as time goes to infinity, the probabilities of creating and destroying one hole are converging to a value near to 0.1. Also, we conjecture that the probability of creating zero holes is bigger than 0.75, and we expect it to be around 0.8. Finally, the probability of creating two holes is smaller that the rest of the probabilities, and we expect it to be less than 0.005. It is important to mention that, up to now, we do not know if these probabilities are stabilizing when $t \to \infty$.

4.5.3   For which t does $\beta_1^t = 0$ and how often does this happen?

We have observed in each of our simulations that there is a time t that varies between 20 and 90—see Table 4.1—at which β_1^t stops being zero (and never goes back to take the value of zero). The conjecture is:

Conjecture 3. *If we define* $h(t) := \mathbb{P}[\beta_1^t = 0]$, *then* $\lim_{t\to\infty} h(t) = 0$ *and the latter should converge to zero exponentially fast.*

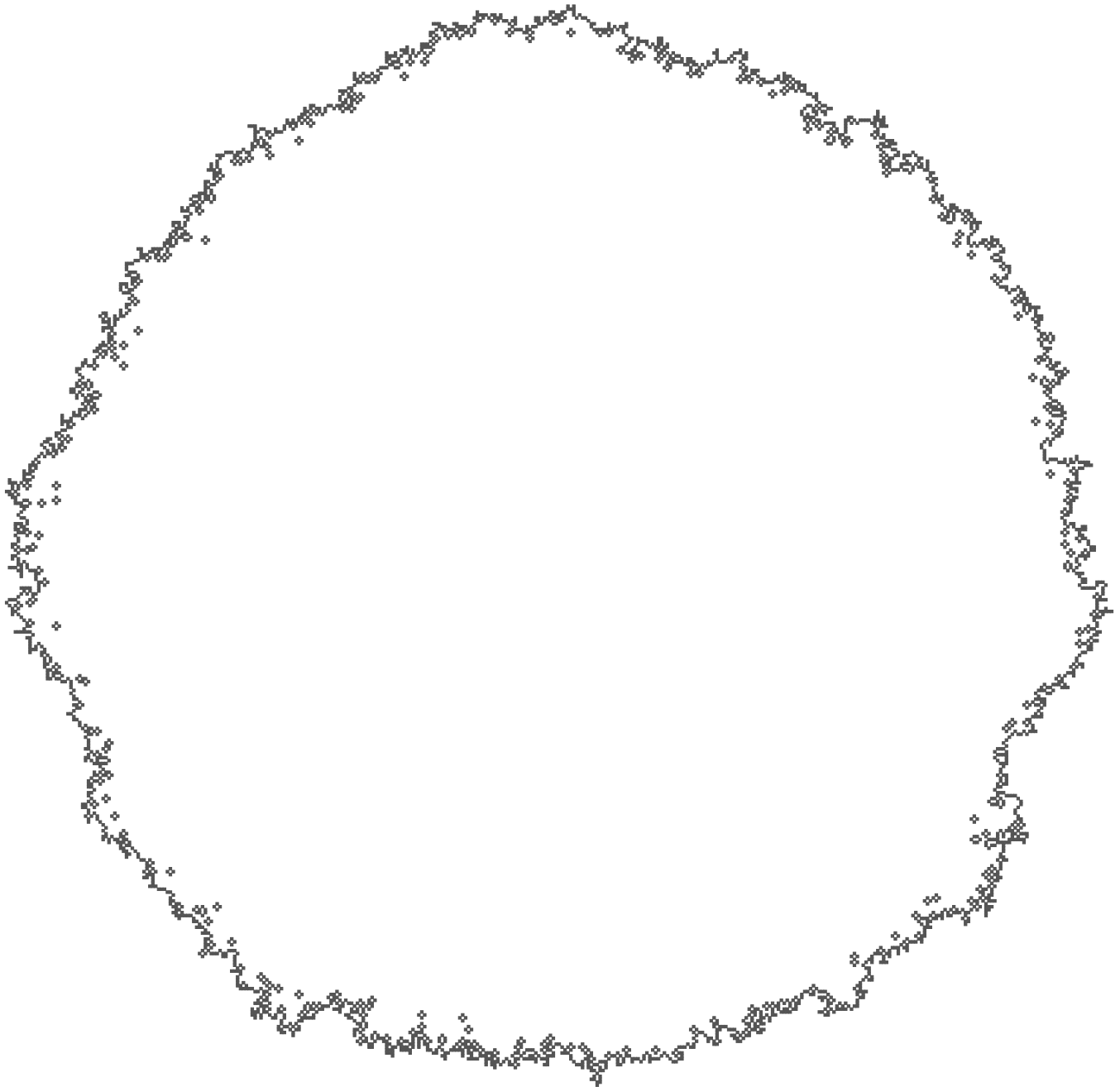### 4.5.4 On the perimeter and limiting shape

FIGURE 4.4: The perimeter of a simulation of the EGM with 100,000
cells.

With our algorithms we can keep track of the evolution of the perimeter of the EGM
in the simulations. Understanding more about the behavior of the perimeter is im-
portant in order to understand the evolution of the topology of the EGM. In the
long run, almost all the holes will be located in an epsilon neighborhood of the
perimeter—see Figure 4.4.

From these results we also conjecture that there exists a constant C such that the site perimeter tends to $C\sqrt{t}$ as t goes to infinity. Because of Theorem A, we know that $C > 4$. We could also explore, from a combinatorial point of view, the set of polyominoes with holes only *near the perimeter*. We then would like to answer the question: What is the maximum possible number of holes, located near the perimeter, that a polyomino can have?

4.5.5 About the Area of the Holes and the Persistent of the Holes

In the simulations we can keep track of the area of each hole, the time at which each hole was created and destroyed, and how the area of each hole changes in time. This allows us to:

a) Have a persistence diagram on the filtration given by the polyominoes that captures how the topology of the process is evolving in time.

b) Explore the average time that a hole, created at time t, will persist. This *persistent time* should be increasing as the process evolves in time because the probability of destroying a hole at time t is inversely proportional to the perimeter of G_t, and the perimeter is increasing with time (it will also depend on the perimeter of the hole). The goal is to find the precise rate at which the time of *a hole being alive* is increasing.

c) We have observed that most of the holes have area 1. But, the probability that big holes are created is increasing, at a very low rate, as the time increases.

d) The reason for not observing holes with big area very often is that they are very unstable in the sense that the probability that a tile in the site perimeter is placed on the perimeter of a hole is proportional to the area of the hole.

Because we are keeping track of the area of each hole we have geometric information about the homology of the EGM and this is giving us another dimension of measure of the persistent homology, not only the time at which the holes are created and covered. The birth and death times associated with a hole will be proportional to the size of a hole after rescaling.

Also, it is important to mention that the algorithm that we have designed and implemented keeps track of the persistent homology splitting tree [26].

### 4.5.6 On persistent homology: barcodes and persistence diagrams

Finally, we show some persistence diagrams and their associated barcodes. We have run the extended algorithm that we have implemented, that can keep track of each one of the holes created in the EGM process, for different times. Results for $t = 500$, $t = 5,000$, and $t = 50,000$ are shown below. In Appendix A.1 we present results for several other values of t.
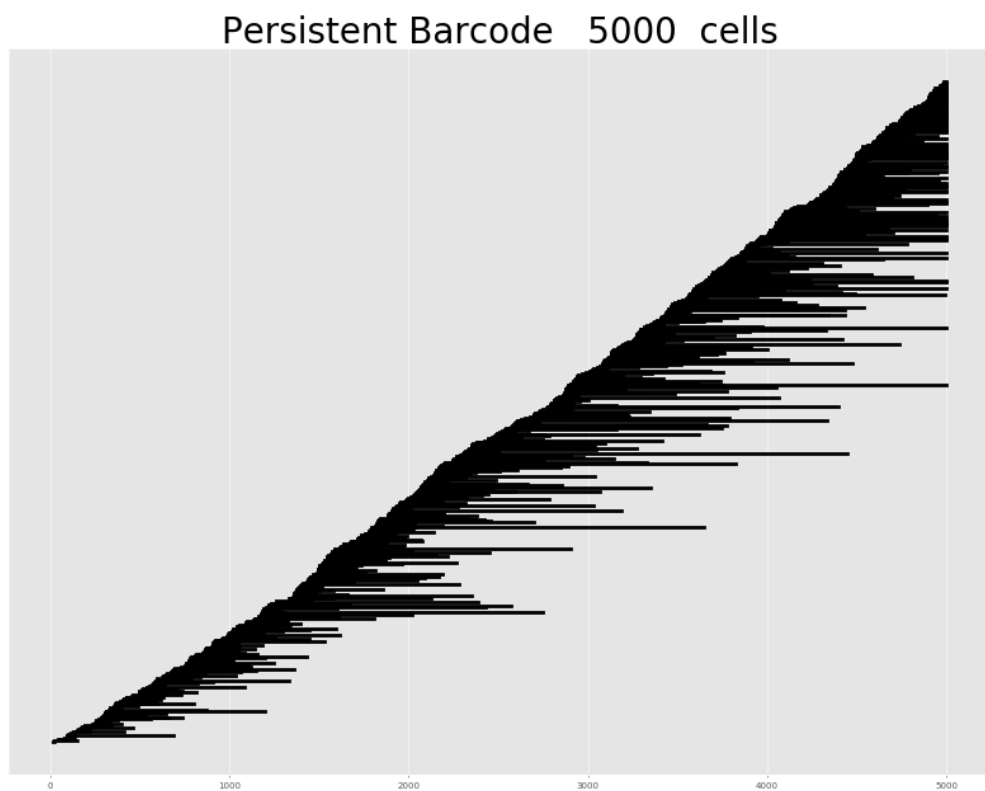
Below we show a simulation for $t = 500$. The green tile is the starting tile of the EGM process and the red tiles are the holes of the polyomino.

Persistent Diagram   500 cells

Persistent Barcode   500 cells
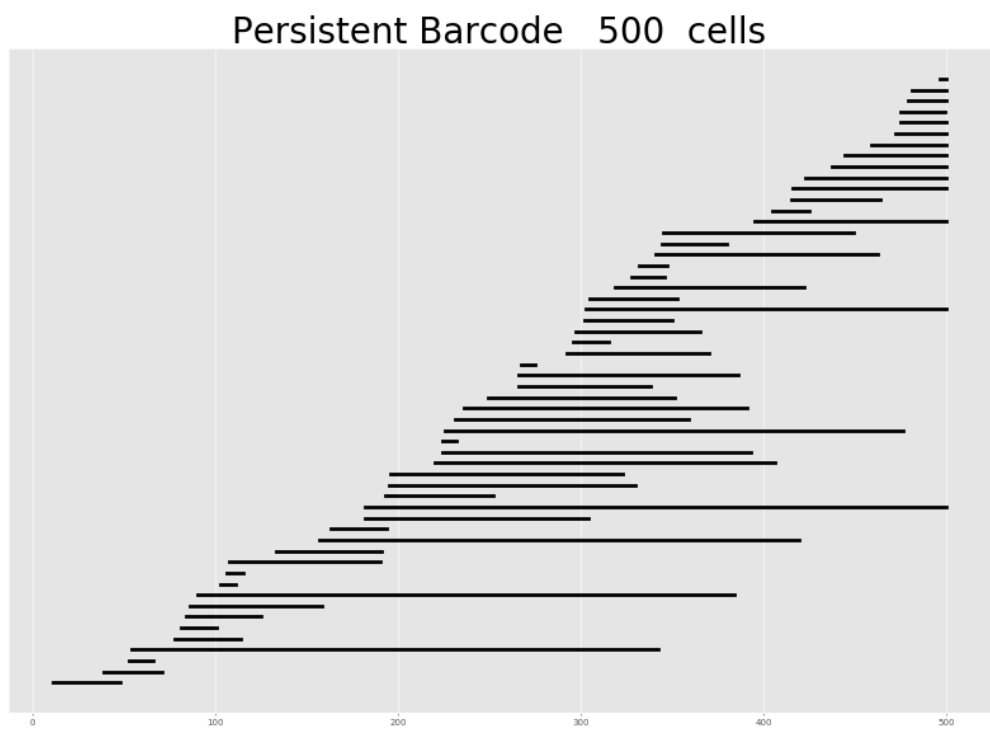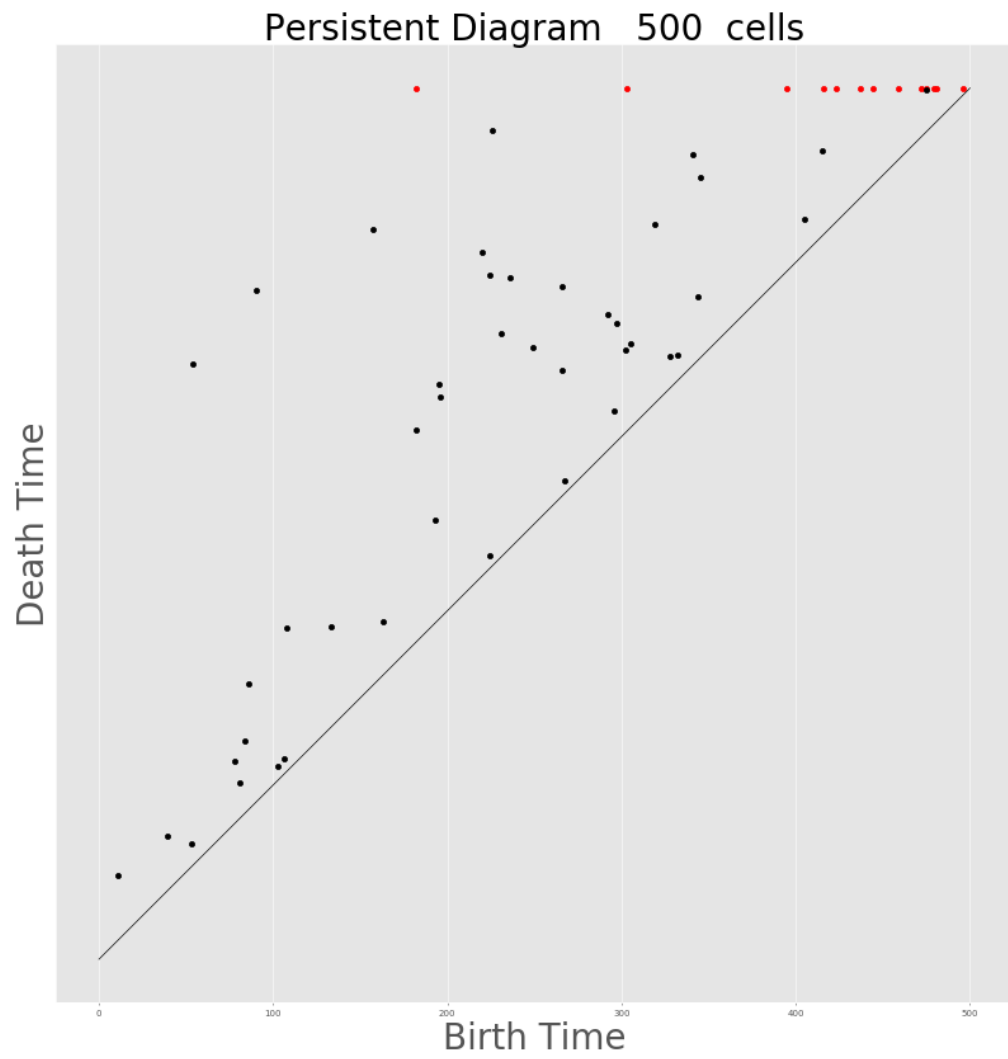
Below we show a simulation for t = 5,000. The green tile is the starting tile of the EGM process and the red tiles are the holes of the polyomino.

Persistent Diagram   5000 cells

Persistent Barcode   5000 cells
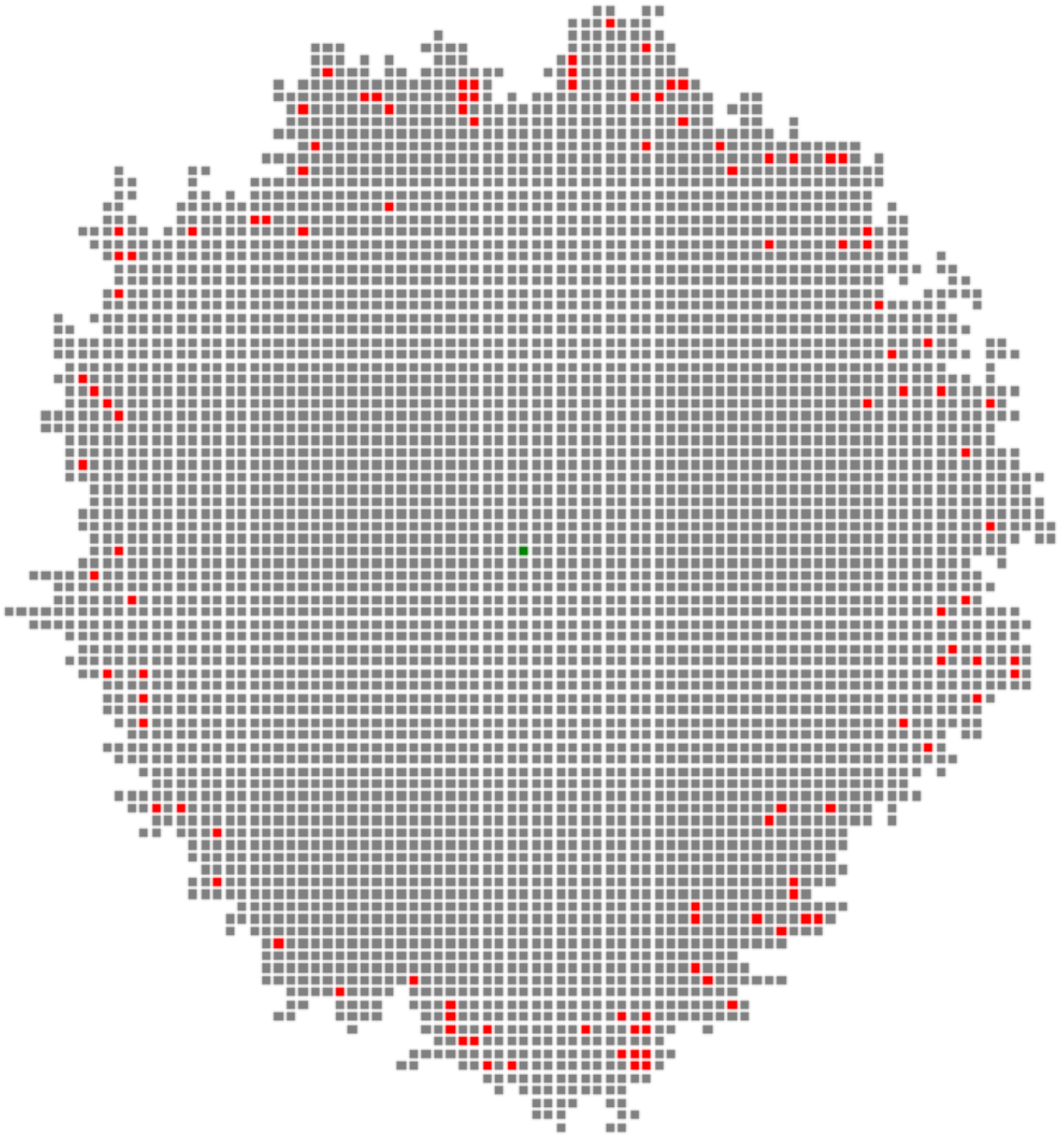
Below we show a simulation for $t = 50,000$. The green tile is the starting tile of the EGM process and the red tiles are the holes of the polyomino.

Persistent Diagram   50000   cells

Persistent Barcode   50000   cells
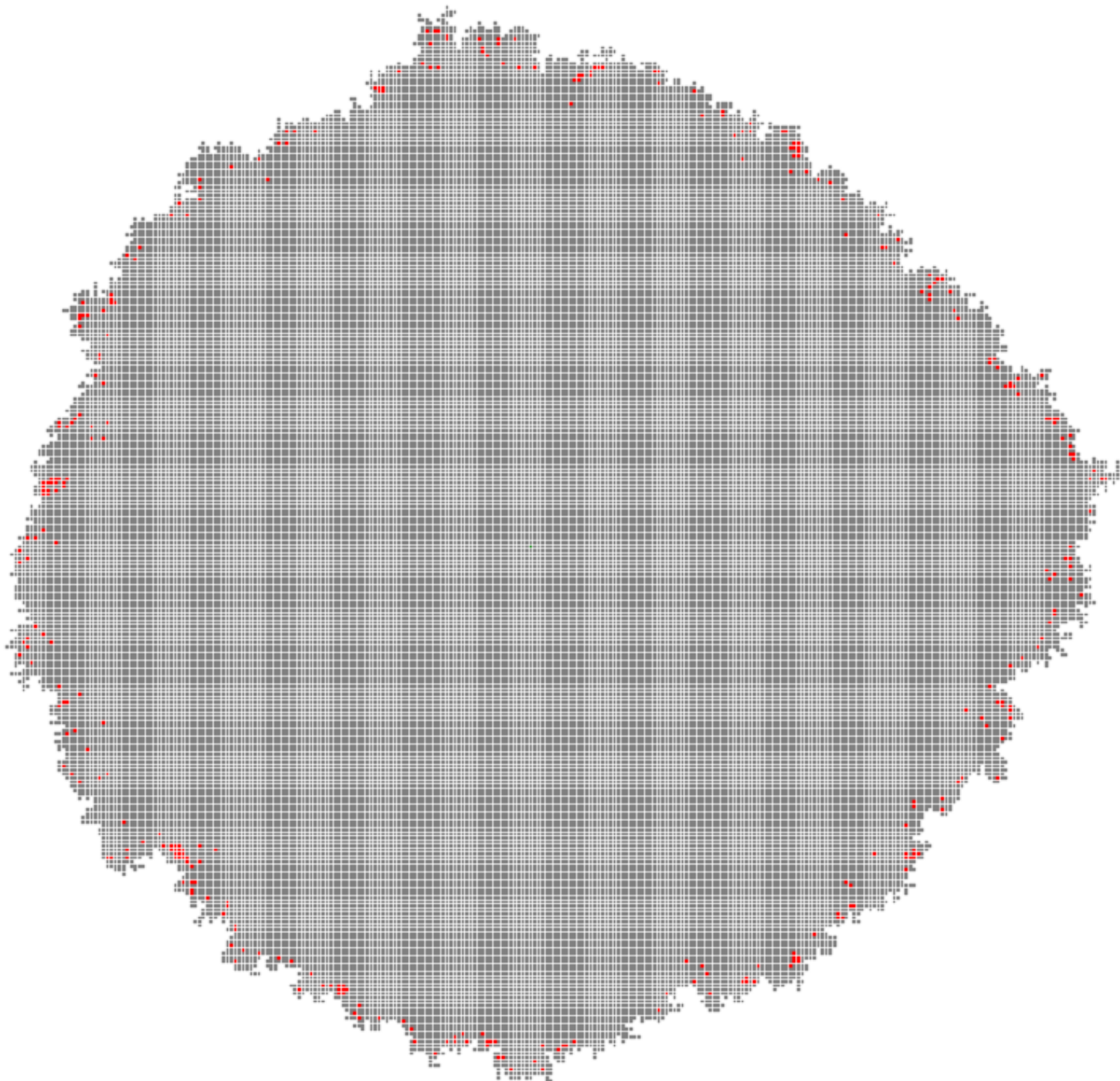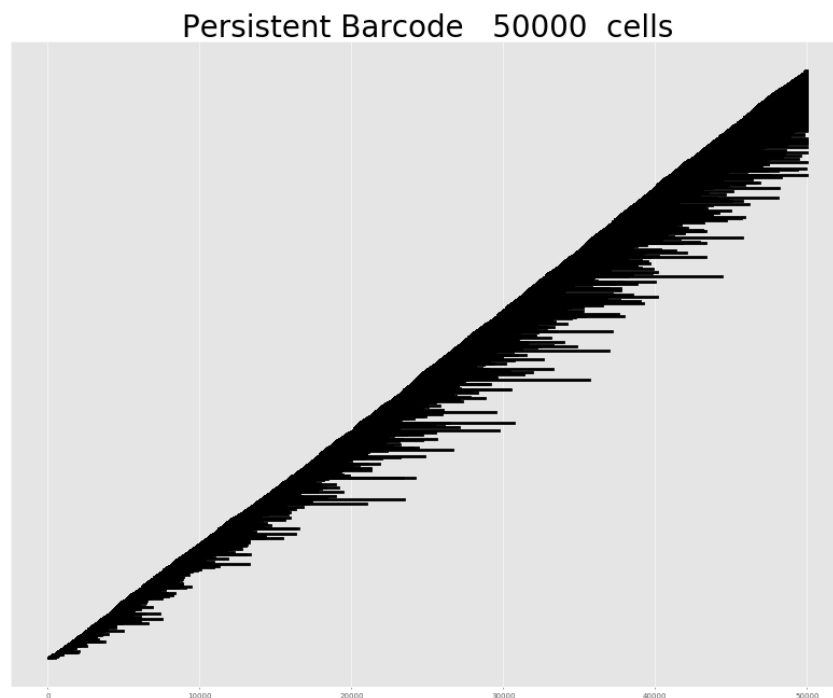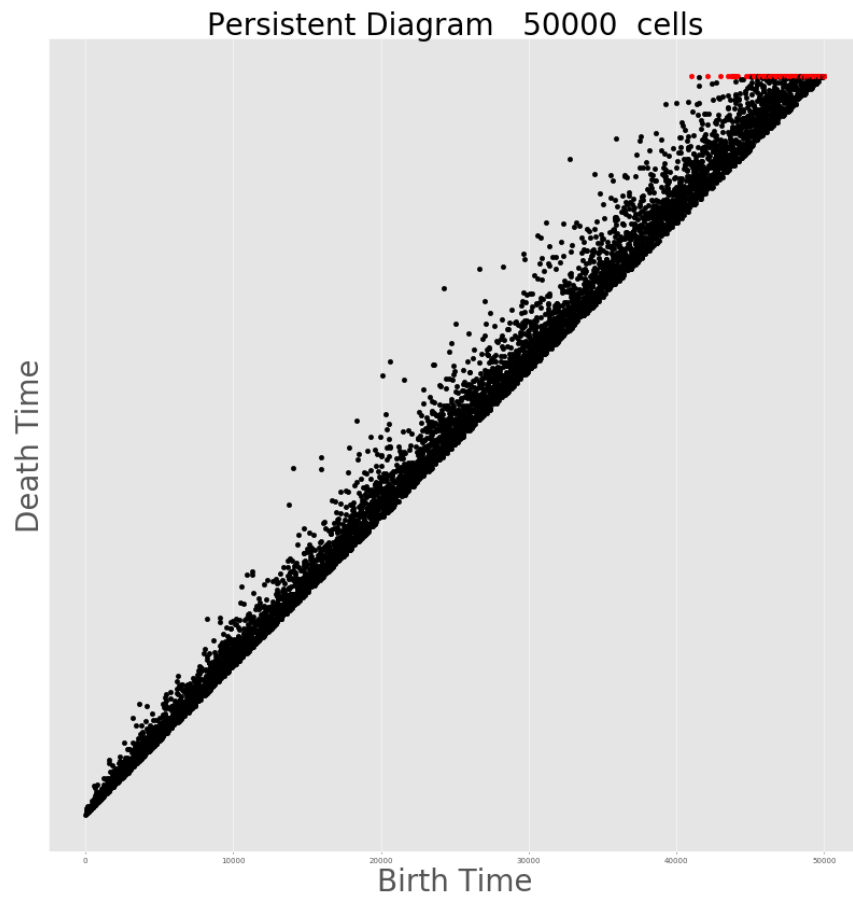
These experiments give experimental evidence for the Conjecture 2 to be true. This conjecture is stated in the introduction of this chapter. We show in Appendix B persistence diagrams and barcodes for simulations of the EGM for different times than $t = 500$ and $t = 50,000$.

Appendix A

Sampling random polyominoes

A.1 Simulation of random polyominoes

Markov Chain Monte Carlo Metropolis and Metropolis–Hasting algorithms are commonly used to sample random structures with a target distribution of very complex combinatorial and geometrical objects. For example, for sampling random states of disc configuration spaces, random trees, and random structures with crystals and quasi-crystal molecular arrangements [5, 7, 12].

In this appendix we briefly describe the Markov Chain Monte Carlo (MCMC) algorithms that we have implemented to sample random polyominoes. We omit the proofs of the results contained in this appendix.

A.1.1 Simulation of uniformly random polyominoes

For generating Figure 3.1 we have implemented a MCMC Metropolis algorithm with the uniform distribution as the target stationary distribution. First, we describe this MCMC Metropolis algorithm and then we state in Theorem 8 that its associated, homogeneous in time, Markov process converges to the uniform distribution. Select a natural number $n \in \mathbb{N}$. Then:

Step (0) Set $A = W$, where W is the n-Worm which is the unique polyomino that has its n tiles in the same row.

Step (1) Select one of the tiles contained in A with uniform probability. We represent this tile by x.

Step (2) Remove the tile x from A. We represent the resulting configuration, which has $n-1$ tiles, by $A \setminus \{x\}$. Now, select with uniform probability one of the tiles that are on the site perimeter of $A \setminus \{x\}$. Denote this tile by y.

Step (3) Place a tile on the site y selected in the previous step. Denote the obtained structure by $(A \setminus \{x\}) + \{y\}$. Then, if $(A \setminus \{x\}) + \{y\}$ is a polyomino make $A = (A \setminus \{x\}) + \{y\}$ and go to **Step (1)**. If $(A \setminus \{x\}) + \{y\}$ is not a polyomino, do not make changes in A (remove y and replace x where it was) and return to **Step (1)**.

Stop the algorithm after finishing **Step (3)** a desired number of times. We discuss in the next section how this stopping time should/could be selected.

The previous algorithm defines the transition probabilities of a homogeneous (in time) Markov chain with $\mathcal{A}_n$ as its state space. Denote these transition probabilities by $P(A, B)$ for all $A, B \in \mathcal{A}_n$. The value of $P(A, B)$ gives the probability of getting the

polyomino B from the polyomino A with the algorithm described above. Because the Markov chain is homogeneous in time, we have $P^t(A, B) = P(A, B)$ for all $t \in \mathbb{N}$ and all $A, B \in \mathcal{A}_n$.

This Markov chain has interesting properties that guarantee that it converges to the uniform distribution. As we did before in this thesis, we are denoting by π the uniform distribution defined over the set $\mathcal{A}_n$.

Theorem 8. *The Markov chain defined by the transition probabilities* $P(\cdot, \cdot)$ *is an aperiodic and irreducible Markov chain that fulfills, with respect to the uniform distribution* π, *the Detailed Balance Condition:*

$$\pi(A)P(A, B) = \pi(B)P(B, A),$$

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

Theorem 8 implies that the uniform distribution is the unique stationary distribution for this Markov chain. Hence, if we run the algorithm associated with this Markov chain for long enough time, then we will sample an n-omino from a distribution that is close to the uniform distribution.

But, how do we know how close we are from sampling uniform distributed random n-ominoes if we run the algorithm for a certain time? Or, how do we know for how long we need to run the algorithm if we want to be very close to the uniform distribution? Moreover, how do we measure the distance between the uniform distribution and the distribution that we are sampling from if we set a fixed time that we are running the algorithm? These questions can be stated in terms of the mixing time associated with a Markov chain that we define in the next section.

A.1.2   Mixing time

If $P(x, y)$ defines the probability of going from state x to state y in an homogeneous Markov chain process with state space χ, and we know that it converges in distribution to the stationary distribution π, then we would like to determine the right mixing time: How long should we run the Markov chain process to be *close enough* to π? In this scenario it is common to measure distances between two distributions by the total variation distance defined as

$$\| P_x^t - \pi \|_{TV} = \frac{1}{2} \sum_y | P^t(x, y) - \pi(y) | = \max_{A \subset \chi} | P^t(x, A) - \pi(A) |, \qquad (A.1)$$

where P_x^t is a column of probabilities that corresponds to the transition probabilities from x to all other states at time t.

Now, we can restate the problem of finding the mixing time as follows: Given P^t, π, x, and ϵ, how large should t be to guarantee that $\| P_x^t - \pi \|_{TV} < \epsilon$? We will denote this mixing time as $t_{mix}(\epsilon)$.

Commonly, the mixing time is selected to be $t_{mix}(1/2)$. We refer the reader to the monograph [7] by P. Diaconis for an explanation of why to select $1/2$ and the book [22] by Y. Peres and D. Levin for general reference on mixing times of Markov chains.

It is relevant to mention that it is not known if the mixing time associated with this Markov chain is polynomial with respect to the number of tiles in a polyomino. It could be the case that this mixing time is polynomial with respect to the number of polyominoes with a fixed area, which would imply that the mixing time grows exponentially fast (at the rate of the cardinality of $\mathcal{A}_n$). Another reason that makes the analysis of this Markov chain so difficult is that it is not transitive (because it is not regular). Nevertheless, as we have mentioned in Section 3.4, this algorithm has been used to sample random polyominoes (lattice animals) since the 1970s [29].

A.1.3 Simulation of polyominoes with percolation distributions

We have generated Figure 3.8 by implementing a Markov Chain Monte Carlo Metropolis–Hasting (MH) algorithm. This algorithm allows us to (asymptotically) sample polyominoes from the percolation distributions π_p defined on $\mathcal{A}_n$—see Section 3.4.

This algorithm is determined by a conditional density and a target density that in our case is π_p. The conditional density is given by the transition probabilities $\rho(\cdot, \cdot)$ defined by: $\rho(A_1, A_2) = 0$ if $P(A_1, A_2) = 0$, and

$$\rho(A_1, A_2) = \min\left\{\frac{\pi_p(A_2)}{\pi_p(A_1)}\frac{P_Y(A_2, A_1)}{P_Y(A_1, A_2)}, 1\right\} = \min\left\{(1-p)^{t_2-t_1}, 1\right\}$$

if $P_Y(A_1, A_2) \neq 0$, for all $A_1, A_2 \in \mathcal{A}^n$. The transition probabilities $P(\cdot, \cdot)$ were defined in the previous section. The values t_1 and t_2 represent the site perimeter of A_1 and A_2, respectively. The algorithm is the following:

Step (0) Set $A = W$, where W is the n-Worm which is the unique polyomino that has its n tiles in the same row.

Step (1) Select one of the tiles contained in A with uniform probability. We represent this tile by x.

Step (2) Remove the tile x from A. We represent the resulting configuration, which has $n - 1$ tiles, by $A \setminus \{x\}$. Now, select with uniform probability one of the tiles that are on the site perimeter of $A \setminus \{x\}$. Denote this tile by y.

Step (3) Place a tile on the site y selected in the previous step. Denote the obtained structure by $(A \setminus \{x\}) + \{y\}$. If $(A \setminus \{x\}) + \{y\}$ is not a polyomino, do not make changes in A (remove y and replace x where it was) and return to **Step (1)**. If $(A \setminus \{x\}) + \{y\}$ is a polyomino go to the next step.

Step (4) Define $B = (A \setminus \{x\}) + \{y\}$. Then, with probability $\rho(A, B)$, set $A = B$ and go to **Step (1)**; and, with probability $1 - \rho(A, B)$, do not make changes in A (remove y and replace x where it was) and return to **Step (1)**.

By the construction of the algorithm, its associated Markov chain is converging to the percolation distribution π_p because it fulfills the detail balance condition with respect to this distribution.

Also, this Markov chain has the same properties as the Markov chain that we have defined for the uniform distribution case. Therefore, we do not know the mixing time for this case either, even though we have used this algorithm to sample

polyominoes with the percolation distributions π_p. It remains as an open problem to establish the mixing time of this MCMC algorithm.

Appendix B

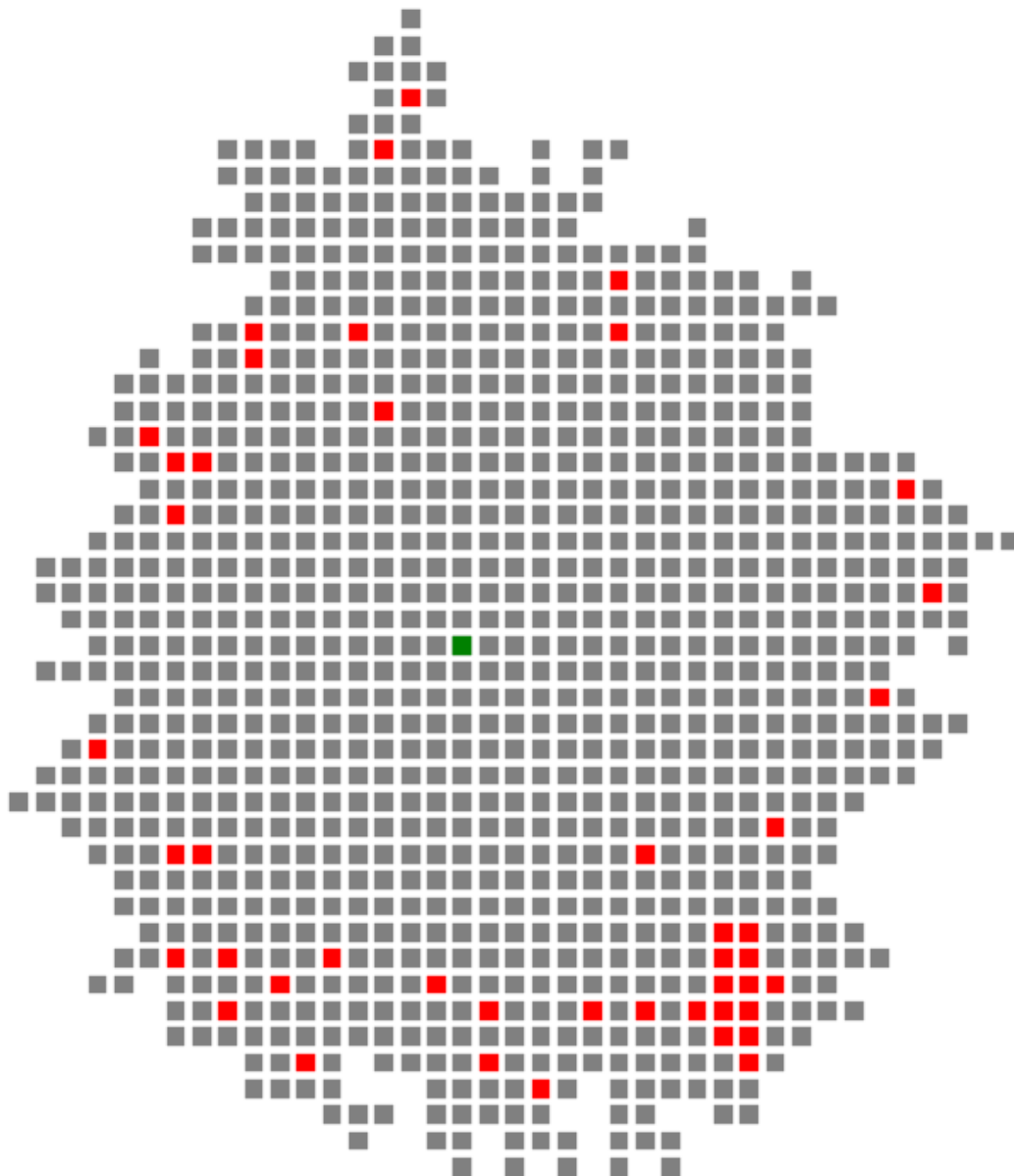# Persistence diagrams and barcodes for the EGM stochastic process

In this appendix we present results of the simulation experiments described in Section 4.5.6 for several additional values of t.
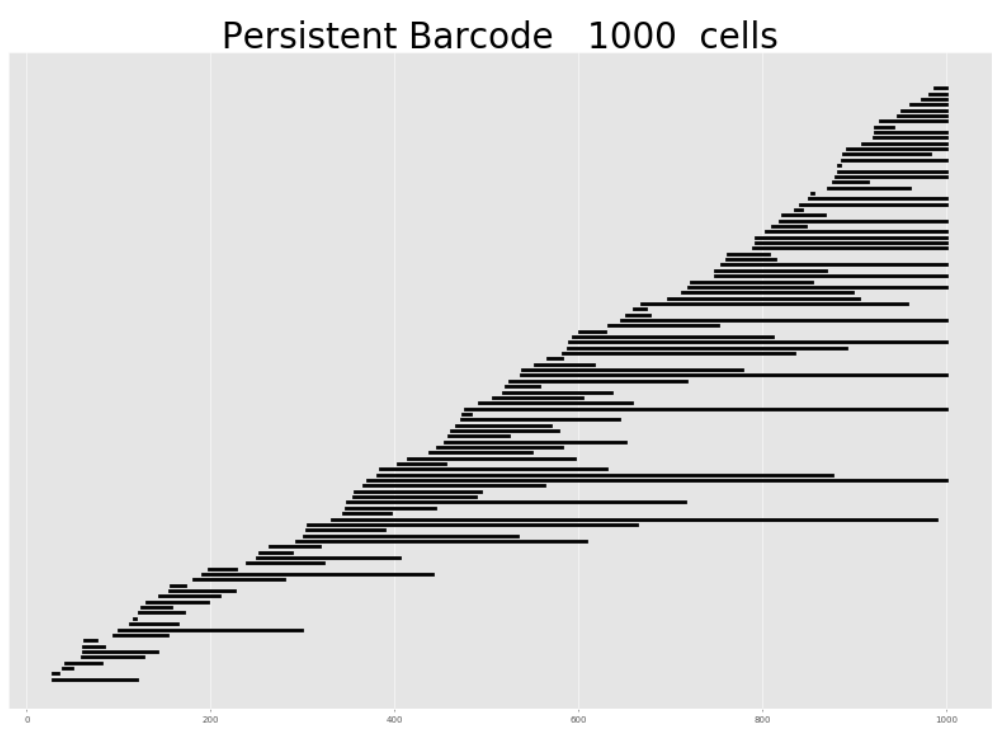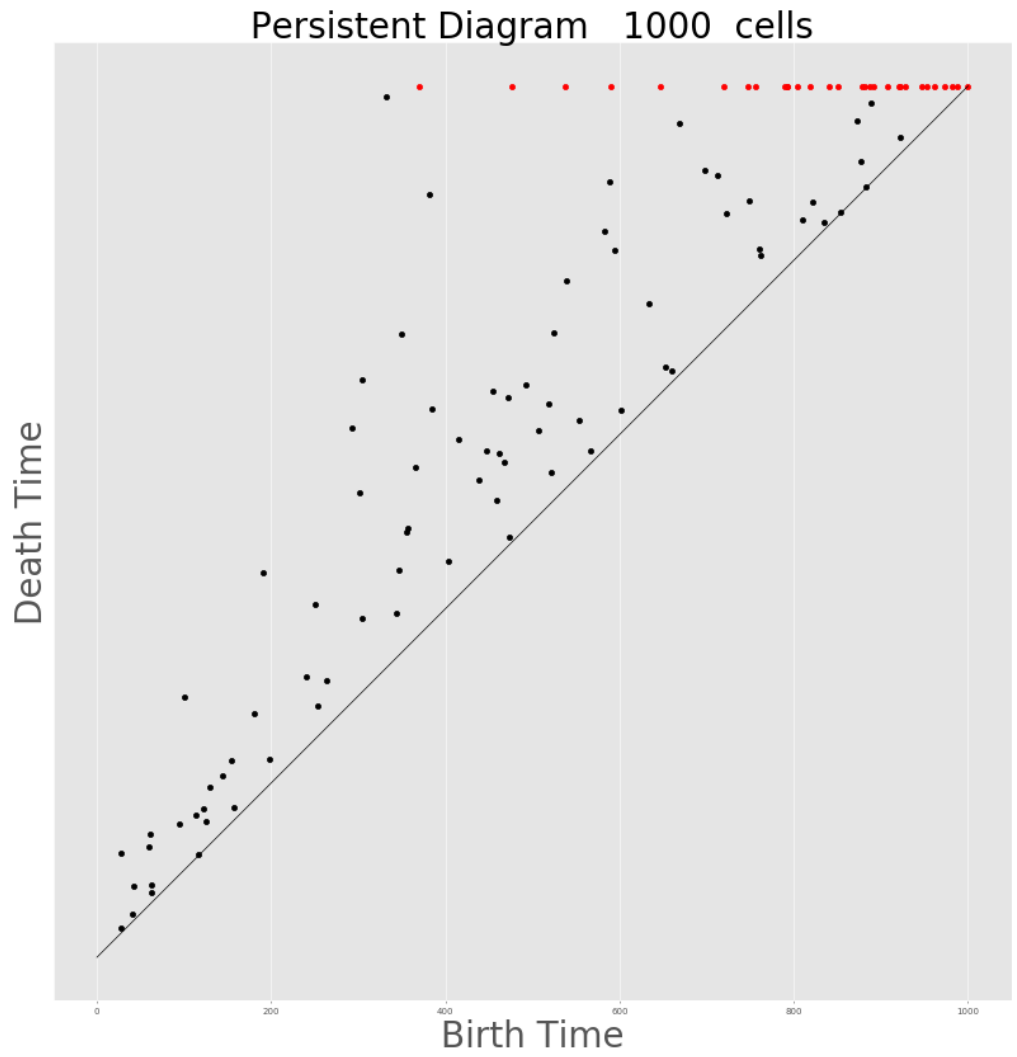
Below, we show a simulation of the EGM stochastic process for t = 100, its persistence diagrams, and barcodes. The green tile is the starting tile of the EGM process and the red tiles are the holes of the polyomino.
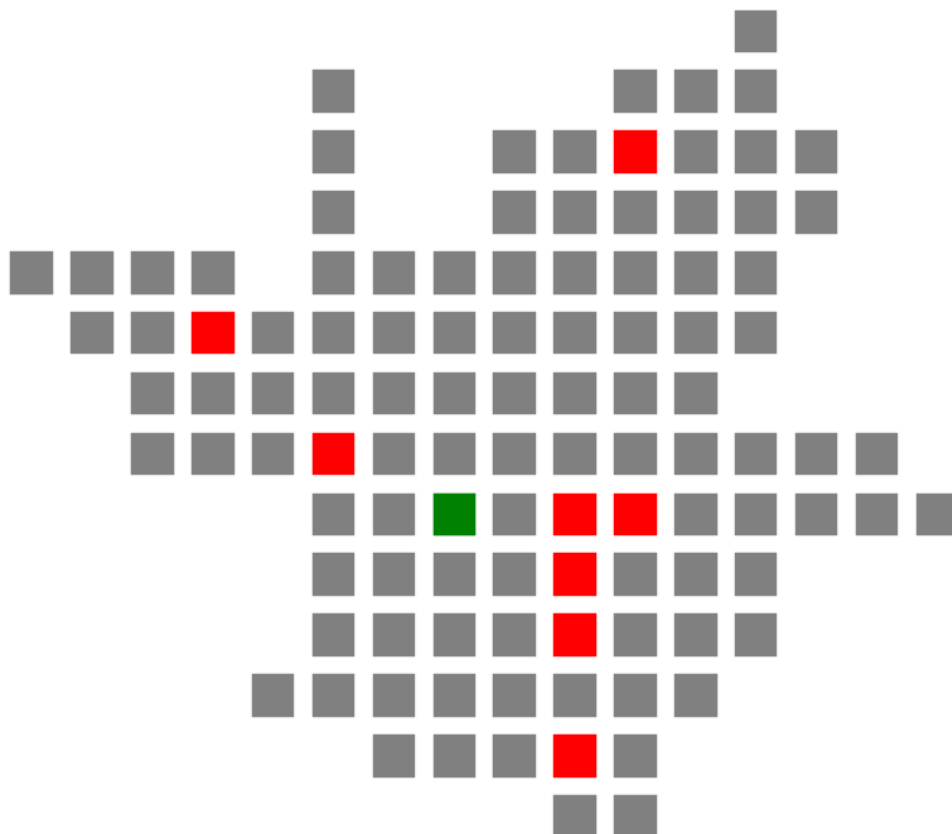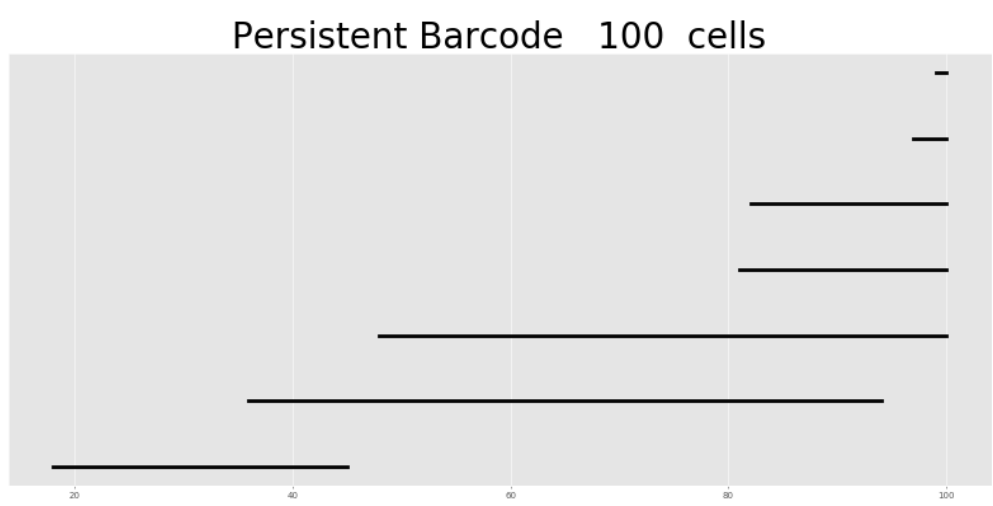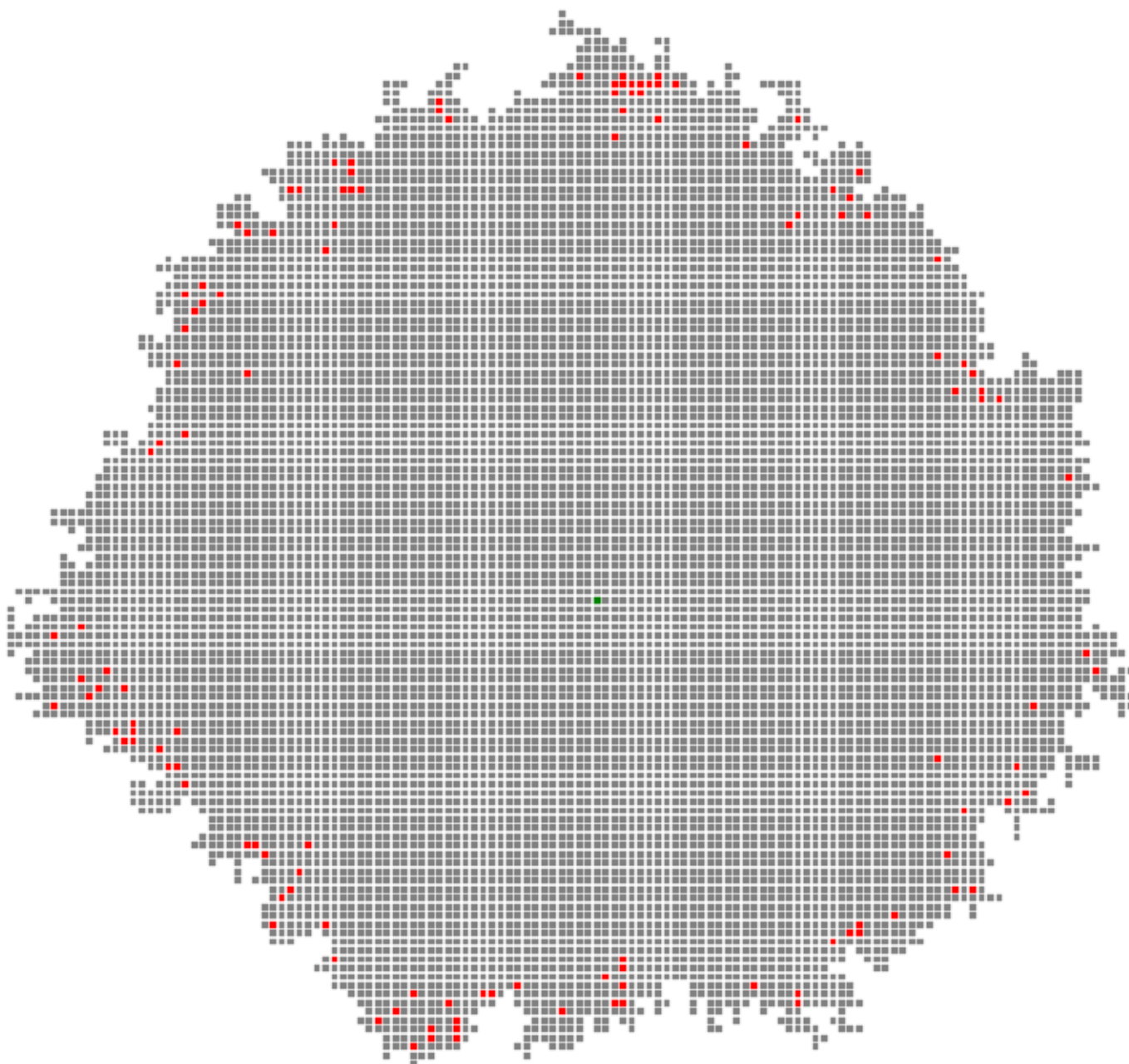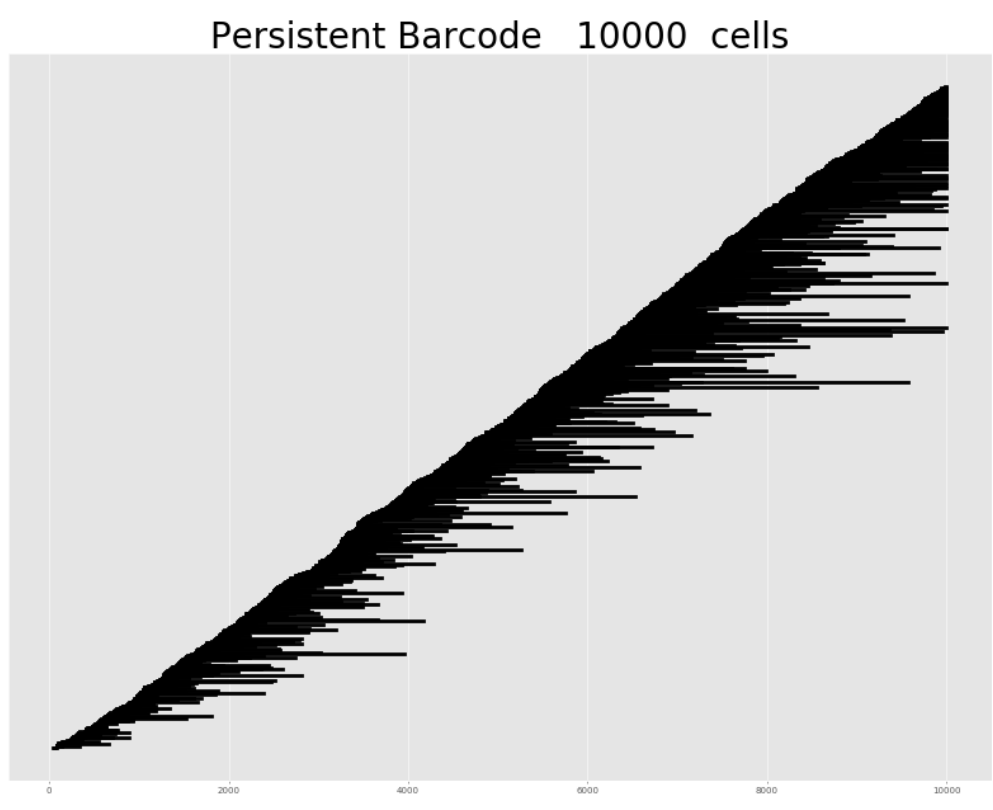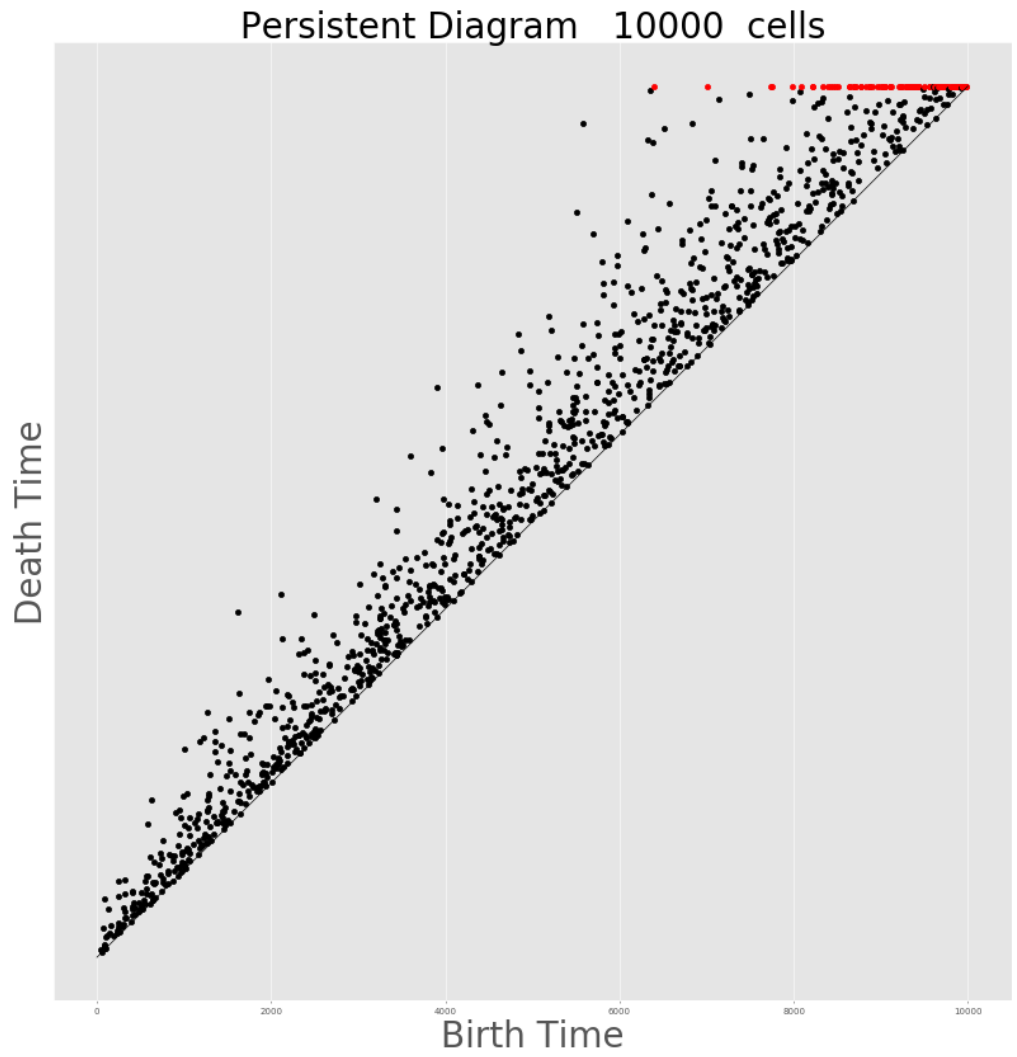
Below, we show a simulation of the EGM stochastic process for t = 1,000, its persistence diagrams, and barcodes. The green tile is the starting tile of the EGM process and the red tiles are the holes of the polyomino.

Persistent Diagram   1000 cells

Persistent Barcode   1000 cells

Here we show a simulation of the EGM stochastic process for t = 10,000, its persistence diagrams, and barcodes. The green tile is the starting tile of the EGM process and the red tiles are the holes of the polyomino.
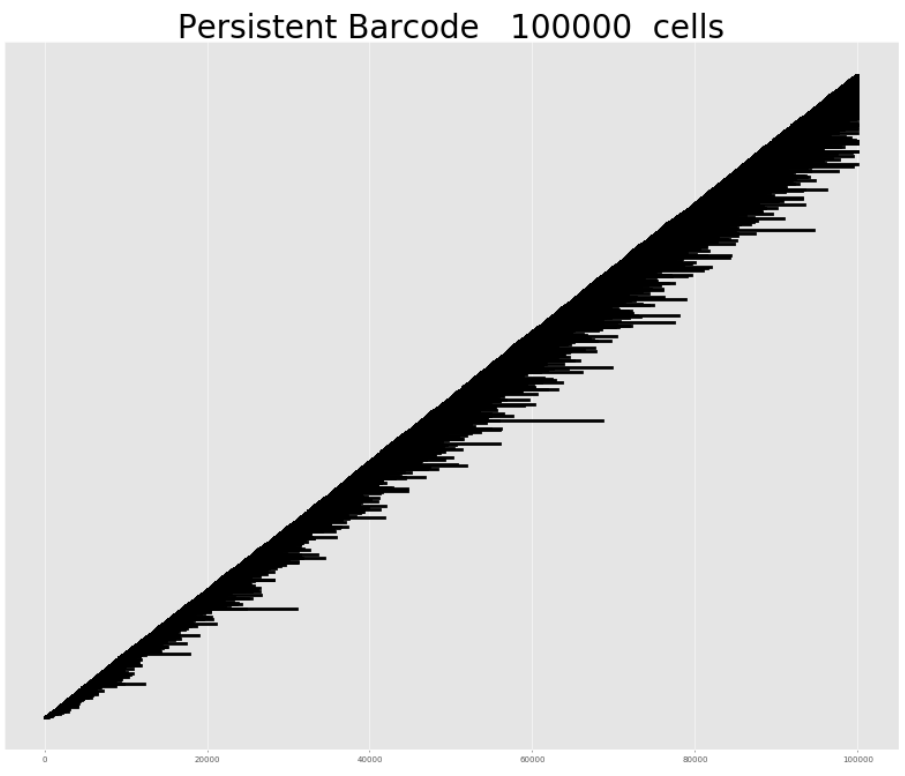
Persistent Diagram   10000 cells

Persistent Barcode   10000 cells

Lastly, we show a simulation of the EGM stochastic process for t = 100, 000, its persistence diagrams, and barcodes. The green tile is the starting tile of the EGM process and the red tiles are the holes of the polyomino.
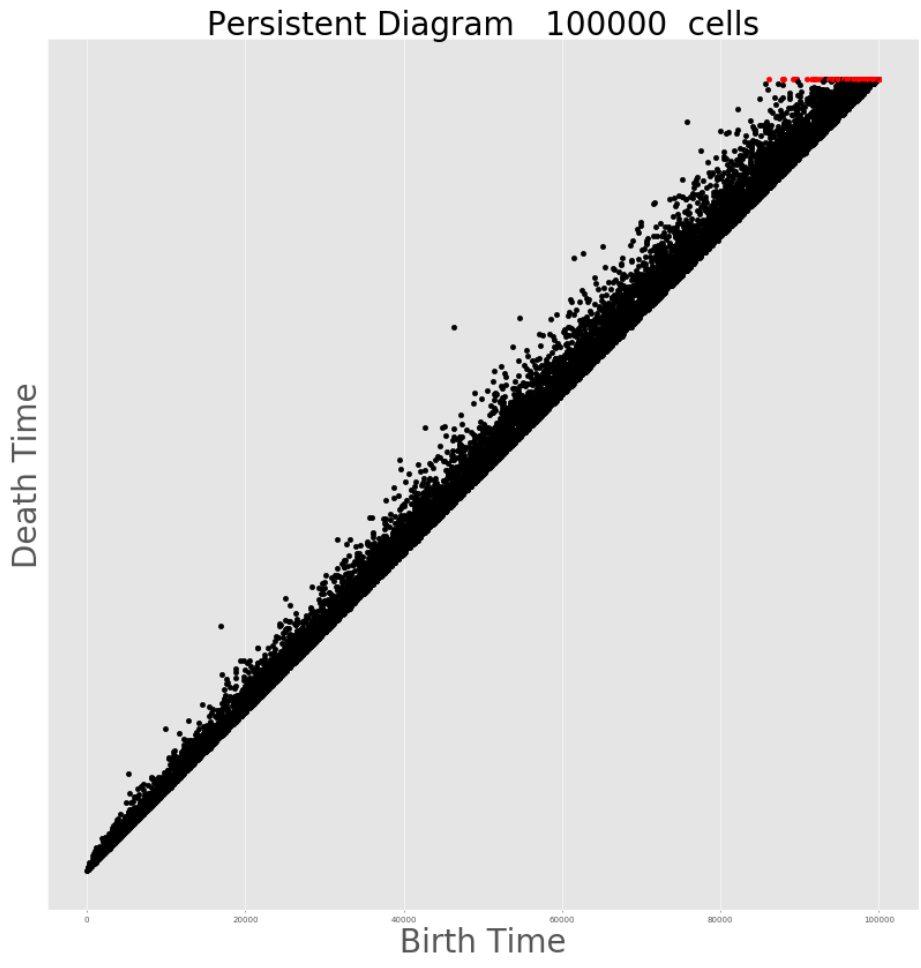
Persistent Diagram   100000 cells

Persistent Barcode   100000 cells
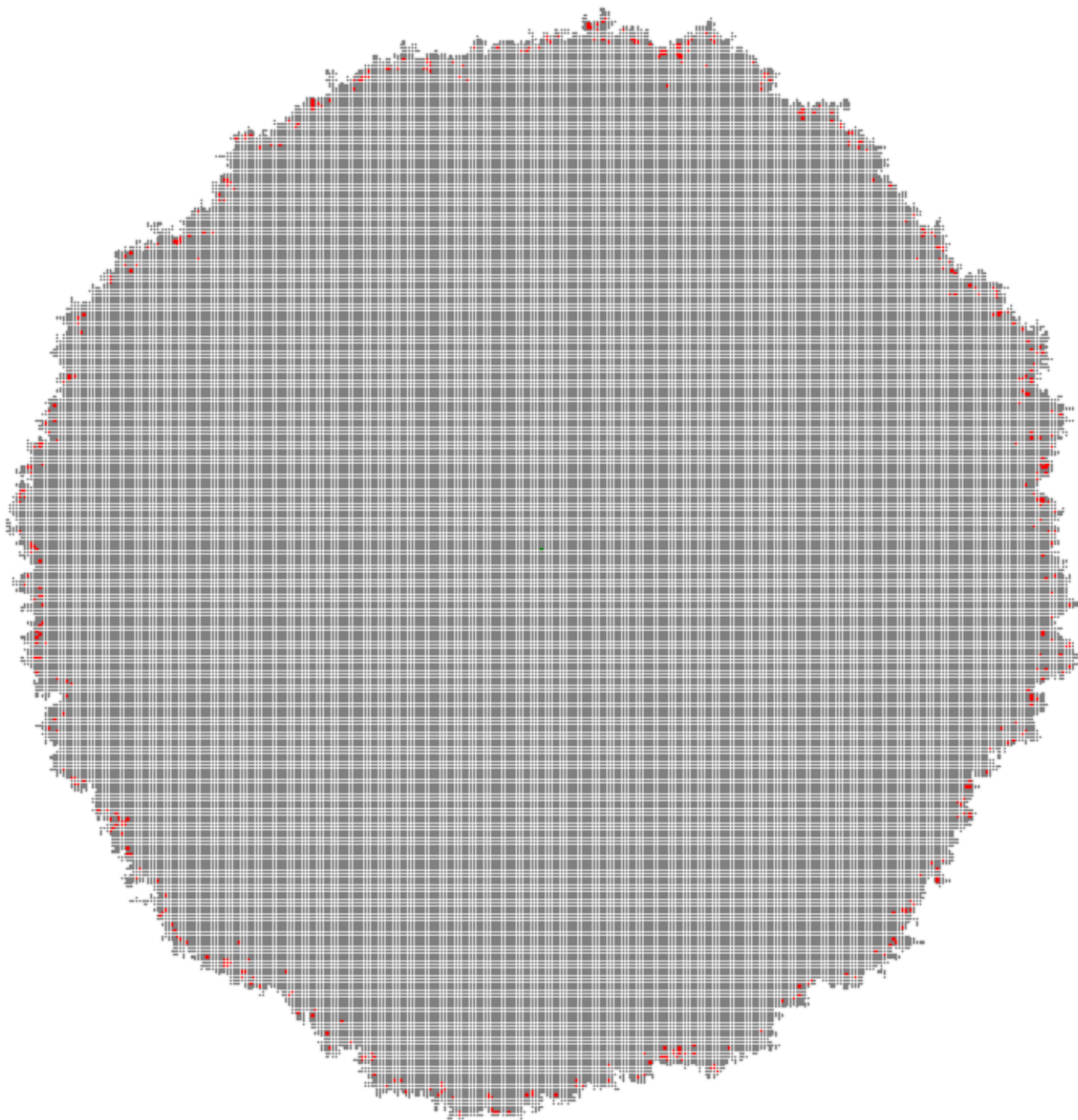
All this simulations provide supporting evidence for the Conjecture 2.

Bibliography

[1] Antonio Auffinger, Michael Damron, and Jack Hanson. *50 Years of First-Passage Percolation*. Vol. 68. American Mathematical Soc., 2017.

[2] Gill Barequet and Ronnie Barequet. "An Improved Upper Bound on the Growth Constant of Polyominoes". In: *Electronic Notes in Discrete Mathematics* 49 (2015), pp. 167–172.

[3] Gill Barequet et al. "Counting polyominoes on twisted cylinders". In: *Integers* 6 (2006), A22.

[4] Rodney Baxter. *Exactly solved models in statistical mechanics*. Elsevier, 2016.

[5] Gunnar Carlsson et al. "Computational topology for configuration spaces of hard disks". In: *Physical Review E* 85.1 (2012), p. 011303.

[6] Jacques Curély. "Thermodynamics of the 2D-Heisenberg classical square lattice Zero-field partition function". In: *Physica B: Condensed Matter* 245.3 (1998), pp. 263–276.

[7] Persi Diaconis. "The Markov chain Monte Carlo revolution". In: *Bulletin of the American Mathematical Society* 46.2 (2009), pp. 179–205.

[8] Henry Dudeney. *The Canterbury Puzzles*. Courier Corporation, 2002.

[9] Murray Eden. "A probabilistic model for morphogenesis". In: *Symposium on information theory in biology*. Pergamon Press, New York. 1958, pp. 359–370.

[10] Murray Eden. "A two-dimensional growth process". In: *Dynamics of fractal surfaces* 4 (1961), pp. 223–239.

[11] Murray Eden and Philippe Thevenaz. "History of a stochastic growth model". In: *Sixth International Workshop on Digital Image Processing and Computer Graphics*. International Society for Optics and Photonics. 1998, pp. 43–54.

[12] Deborah Franzblau. "Computation of ring statistics for network models of solids". In: *Physical Review B* 44.10 (1991), p. 4925.

[13] Solomon Golomb. "Checker boards and polyominoes". In: *American Mathematical Monthly* 61 (1954), pp. 675–682.

[14] Solomon W Golomb. *Polyominoes*. Second. Princeton University Press, Princeton, NJ, 1994.

[15] Anthony J Guttmann. *Polygons, polyominoes and polycubes*. Vol. 775. Springer, 2009.

[16] Frank Harary. "Unsolved problems in the enumeration of graphs". In: *Selected Papers of Alfréd Rényi* (1960), p. 63.

[17] Frank Harary and Heiko Harborth. "Extremal animals". In: *Journal of Combinatorics Information Syst. Sci.* 1.1 (1976), pp. 1–8.

[18] Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2002.

[19] David Klarner. "Cell growth problems". In: *Canadian Journal of Mathematics* 19.4 (1967), p. 851.

[20] David Klarner. "Some results concerning polyominoes". In: *Fibonacci Quarterly* (1965), pp. 9–20.

[21] Sascha Kurz. "Counting polyominoes with minimum perimeter". In: *Ars Combinatoria* 88 (2008), pp. 161–174.

[22] David Levin and Yuval Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.

[23] Neal Madras. "A pattern theorem for lattice clusters". In: *Annals of Combinatorics* 3.2-4 (1999), pp. 357–384.

[24] Igor Pak. "Complexity problems in enumerative combinatorics". In: *arXiv preprint arXiv:1803.06636* (2018).

[25] Austria. R Development Core Team. R Foundation for Statistical Computing Vienna. "R". In: *URL http://www.R-project.org* (2008).

[26] Benjamin Schweinhart. "Statistical Topology of Embedded Graphs". PhD thesis. Princeton University, 2015.

[27] Tomás Oliveira e Silva. *Animal enumerations on the* $\{4,4\}$ *Euclidian tiling*. http://sweet.ua.pt/tos/animals/a44.html. December 2015.

[28] Alexander Soifer. *Geometric etudes in combinatorial mathematics*. Springer, New York, 2010.

[29] Dietrich Stauffer. "Monte Carlo Study of Density Profile, Radius, and Perimeter for Percolation Clusters and Lattice Animals". In: *Physical Review Letters* 41.20 (1978), p. 1333.

[30] Jacob D Stevenson, Jörg Schmalian, and Peter G Wolynes. "The shapes of cooperatively rearranging regions in glass-forming liquids". In: *Nature Physics* 2.4 (2006), p. 268.

[31] Alan M Turing. "The chemical basis of morphogenesis". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 237.641 (1952), pp. 37–72.