2D Digital Balls
Daniel Vainsencher
2D Digital Balls
Research Thesis

In Partial Fulfillment of The
Requirements for the Degree of
Master of Science in Computer Science

Daniel Vainsencher

Submitted to the Senate of
the Technion- Israel Institute of Technology

Cheshvan, 5770 Haifa
October 2009
The Research Thesis Was Done Under The Supervision of Alfred M. Bruckstein in the Department of Computer Science.

The Generous Financial Help Of the Miriam and Aaron Gutwirth Scholarship is Gratefully Acknowledged.
Contents

1 Abstract 1

2 Introduction 2

3 Brief Overview of Results 5

4 Pareto Optimality on the Eight Connected Grid 8
   4.1 Optimality conditions 8
   4.2 Some suboptimal shapes and notation 9
   4.3 The optimal shapes on the 8 connected grid are rectangles 13
   4.4 The optimal rectangles on the 8 connected grid have almost equal sides 16

5 Pareto Optimality on the Four Connected Grid 18
   5.1 Describing optimal simple shapes in $X_4$ 21
   5.2 Spines of optimal shapes are short 24
   5.3 Relation to the sequence of integers closest to $\frac{z^2}{8}$ 27

6 Pareto Optimality on the Hexagonal Grid 28
   6.1 Definitions 28
   6.2 Optimal shapes in $X_6$ are simple 31
   6.3 The optimal simple shapes 33

7 Context and Motivations 38

8 Concluding Remarks 41
## List of Figures

1. Three shapes with optimal boundary for their size. $A$ and $B$ show non-uniqueness of optimal shapes, $C$ can be enlarged without affecting the boundary size. ........................................ 3
2. Feasible $(a(S), p(S))$ pairs with maximal and minimal boundaries, delimiting $F$ in $X_4$. The squares have maximal perimeters, the circles have minimal perimeters, the full circles also have maximal area for their perimeter. For example a single tile has 4 neighbors and is extremal in both ways. Thus every shape $S$ with area $a(S)$ must have perimeter $p(S)$ between a circle and a square. ................................................................. 4
3. A prefix of the Wang & Wang spiral. ...................................... 5
4. Shapes generating all optimal shapes for squares, when neighbors share a side or corner. ...................................................... 6
5. Shapes generating all optimal shapes for squares, when neighbors share a side. ................................................................. 6
6. Shapes generating all optimal shapes for hexagons. ...................... 6
7. Any two shapes can be translated to share 3 neighbors in $X_8$. .... 9
8. The neighbor directions in $X_8$, induced by an arbitrary tile $x_{orig}$ and its neighbors $x_1$, $x_2$ and so forth. .......................... 10
9. A union of separated shapes is not optimal. .............................. 11
10. A shape and two coordinate systems. ........................................... 12
11. In this example $A$ is as before, $B$ is its bounding rectangle (black background). $a = a_1; b = a_2; \ldots$ similarly $na = n_1 \ldots$. The white circles are in $N(A)$, the grey ones are in $N(B) \setminus N(A)$. Here $N_{3,5}$ consists of the two circles between $nd$ and $na$. The direction associated with each element of $N(B)$ is shown as an arrow. ............................... 15
12. Diagram of a separation lemma argument, if $p(n)$ is not finite. Each line divides the plane into a left half, which we call $C$, and a right half called $D$. The vertical line is created when $n = nc = n_5$, then each of $C, D$ contains one of $nb, nd = n_3, n_7$. The diagonal line corresponds to the case in which $n \in N(B)$ is not one of $\{n_1\}$, but in $N_{3,5}$ placing each of the delimiters of $N_{3,5}$, which are $nb, nc = n_3, n_5$ in one of $C, D$. .................................................... 16
13. Any two shapes can be translated to share 2 neighbors in $X_4$. 19
14. A simple shape and its dimensions. ........................................... 20
15. Construction of the four elements $\{n_1\}$, on three distinct $a_i$. 21
16. If the vertical and horizontal rays meet, then $n6$ in the diagram (which refers to $n_6$) is separated from $n4$ ($n_4$). Both are neighbors of $A$, which allows a separation argument against the optimality of $A$. ................................................................. 21
17. A degenerate shape $A$, and a variation $A'$ which clarifies that $A$ is non optimal ................................................................. 22
18. Expanding a simple shape to one side .................................... 23
19. Spine types and their areas .................................................. 24
20 The optimal spines ................................................. 27
21 Three optimal degenerate shapes .................................. 27
22 One coordinate system for the hexagonal grid .................. 29
23 A is the gray shape, then $\psi_{i,i+1}(A)$ for $i = 3$ is the hexagon with coordinates (3,0) .................................................. 30
24 Two shapes placed so they share 2 neighbors. .................... 31
25 The shape A is given in dark gray, the hexagons not in A but in the minimal simple shape B containing A are lighter colored, and $N(B)$ are lightest. Here each arrow in direction $d_i$ starts at $n_i$, to find $a_i$ look in the reverse direction. Note that $a_i = a_j$ is possible, but $n_i = n_j$ is not. .................................................. 32
26 Here we see $d(n)$ for each n. Note that $n - k \cdot d(n)$ for $k \in \mathbb{Z}$ passes through some element of $B$, therefore through exactly two elements of $N(B)$. .................................................. 33
27 This is a simple shape in $X_6$ that can be described as having dimensions $d = 6, e = 5$ and ear sizes $j = 2; k = 3$. ............... 34
28 Naphtalene is a simple Polycyclic Aromatic Hydrocarbon (multiple carbon hexagons terminated by hydrogen atoms). Larger PAH molecules (and in the limit, graphene sheets) correspond to connected shapes in the hexagon tiling (when omitting the hydrogens from the diagram). .................................................. 40
29 Left shape has 13 neighbors, right shape has 14 neighbors, but both have 22 edges and 8 internal vertices. Then the neighbor based perimeter is not equivalent to perimeter as defined by edges or internal vertices. .................................................. 40

List of Tables

1 For each spine that may be optimal, we find $i$ such that the perimeter of that spine after $m$ expansions is $4(m + 1) + i$. Then we calculate the expanded areas for those spines, and find some spines that result in shapes of smaller areas than those resulting from a different spine with the same $i$ therefore the same perimeters. .................................................. 26
1 Abstract

In the plane, the way to enclose the most area with a given perimeter and to use the shortest perimeter to enclose a given area, is always to use a circle. If we replace the plane by a regular tiling of it, and construct polyforms i.e. shapes as sets of tiles, things become more complicated. We need to redefine the area and perimeter measures, and study the consequences carefully. A spiral construction often provides, for every integer number of tiles (area), a shape that is most compact in terms of the perimeter or boundary measure; however it may not exhibit all optimal shapes. We characterize all shapes that have both shortest boundaries and maximal areas for three common planar discrete spaces.
2 Introduction

In the continuous plane $\mathbb{R}^2$, disks are special in that they have maximal area for a given length of boundary. They also have minimal length of boundary, for any given area. These two claims are equivalent in $\mathbb{R}^2$ because disks can be scaled arbitrarily to have any boundary length or any area. The disks, equivalent up to scaling, are also the only minimal shapes. These facts may be summarized by a theorem called the isoperimetric inequality:

**Theorem 2.1.** Let $\gamma$ be a simple rectifiable closed curve in $\mathbb{R}^2$, then by the Jordan theorem, it encloses a finite area $A$. If we denote the length of the curve $L$, then $A \leq \frac{L^2}{4\pi}$, with equality achieved only by circles.

Digital geometry is concerned, among other things, with regular tilings of the plane, for example by squares or hexagons and the shapes that arise by considering sets of tiles, also called polyforms. For example, digital straight lines (DSLs) \[3\] may be defined as subsets of $\mathbb{Z}^2$ that fulfill a linear inequality: \[
\{(i, j) \in \mathbb{Z}^2 | j \leq ai + b\},
\] and their interesting properties arise from the interaction between the reals $a, b$ and the integer grid.

In digital geometry, we can scale any shape up by any integer number $n$, by replacing each tile with a “meta” tile of “side size” $n$, and this results in a valid shape. However, this kind of scaling is usually impossible for non integer ratios. Even worse, this form of scaling does not generally preserve geometric properties that we may care about. For example, scaling the digital straight line $\{(i, j) | j \leq i\}$ by a factor of 2 on $\mathbb{Z}^2$ does not result in a digital straight line.

Therefore, it should not come as a surprise that the above discussed equivalence between the two interpretations of the isoperimetric inequality in $\mathbb{R}^2$ does not extend to discrete cases. The discrete case reveals its secrets via a delicate analysis of the interaction between suitably defined concepts of boundary size (in two dimensions, this is the perimeter) and area for each regular tiling of the plane. Before we present one existing approach to such problems and describe our own results, we shall define our aims and terms more precisely on a particular example.

We identify the tiling of the plane by squares of unit length sides with $\mathbb{Z}^2$. It is then natural to define the area of a shape as the number of tiles in it. When it comes to defining its perimeter, however, there is more than one natural way to do so. Considering length of boundary literally, we might count the number of sides that squares inside the shape share with squares outside it. Another natural definition of boundary size is to count the number of squares that are neighbors of the shape, where neighbors may be defined as sharing a side, or as sharing a side or corner with squares in the shape. Unfortunately these three definitions are not equivalent, so we must select one of them before we can explore the relation between the area of a shape, and the size of its boundary.

For example, let us choose to define the boundary size as the number of squares sharing a side with the shape. Now we can consider the set of geometrically distinct shapes in Figure 1, and note the following: $A$ and $B$ are obviously different shapes with the same area and the same neighborhood size.
Two questions naturally arise.

1. Is there a unique shape of maximal area for a given boundary size, or a unique shape of minimal boundary size for a given area? The answer is no, because as will become clear later, A and B are optimal in both senses.

2. Is every shape of maximal boundary for its area also of maximal area for its boundary size? Again the answer is no, because it is easy to add a square to shape C without increasing its boundary size, despite the fact (which is harder to see) that C has optimal boundary size for its area.

Therefore, given the above definitions, the two optimality conditions are clearly not equivalent. The relations between shapes, areas and (at least) this definition of boundary size are more complicated than those in the continuous case.

From a general perspective, we have a set $O$ of objects of interest, i.e. the shapes or subsets of a tiling, and two ways to measure the size of an object, i.e. the area and boundary size (or perimeter). Hence, to each shape $S$ we associate its area $a(S) \in \{0\} \cup \mathbb{N}$ and its perimeter $p(S) \in \{0\} \cup \mathbb{N}$. Then, to learn about the interactions between these two measures for a given tiling and suitable definitions of $a(S), p(S)$, we may look at the set of feasible points $F = \{(a(S), p(S)) | S \in O\} \subset (\{0\} \cup \mathbb{N})^2$.

For example, in the space we considered above, each square of a shape has at most 4 neighbors, therefore $p(S) \leq 4 \cdot a(S)$, equality being achieved by shapes in which every two tiles are far enough to not share neighbors. This reflects a fundamental difference between this space and $\mathbb{R}^2$, in which there exist shapes with finite area but infinite boundary length, such as the Koch snowflake or other fractals [8].

Let us return to the isoperimetric issues. Since the shapes in a tiling also correspond to shapes in $\mathbb{R}^2$, we might reasonably expect that there exists a constant $\gamma$ such that $\sqrt{a(S)} \cdot \gamma \leq p(S)$, for reasonable definitions of perimeter. Then a diagram of $F$ is presented in Figure 2.

In this diagram, the lower boundary, representing the price in perimeter to be paid for increasing the area, raises various questions of interest. The set of feasible pairs $(a(S), p(S))$ along this boundary corresponds to extremal shapes highlighting the discrete isoperimetric inequality. Note that it is possible that a feasible pair is satisfied by more than one shape, then we shall be interested in

![Figure 1: Three shapes with optimal boundary for their size. A and B show non-uniqueness of optimal shapes, C can be enlarged without affecting the boundary size.](image-url)
Figure 2: Feasible \((a(S), p(S))\) pairs with maximal and minimal boundaries, delimiting \(F\) in \(X_4\). The squares have maximal perimeters, the circles have minimal perimeters, the full circles also have maximal area for their perimeter. For example a single tile has 4 neighbors and is extremal in both ways. Thus every shape \(S\) with area \(a(S)\) must have perimeter \(p(S)\) between a circle and a square.

mathematically characterizing those shapes. We shall deal in this paper with the set of shapes that, like disks in \(\mathbb{R}^2\), are optimal both in having the maximal area for their perimeter and in having the minimal perimeter for their area. These shapes are said to form the so called Pareto optimal frontier of the diagram of \(F\) in Figure 2. Similar diagrams and the concept of Pareto Optimality [9] are indeed used in economics and in engineering in situations where conflicting criteria need to be jointly optimized.

A common way of generating shapes on the boundary of \(F\) defining the isoperimetric inequality corresponding to a particular tiling, is a spiral construction. It starts with one tile, and sequentially adds tiles in a circular order growing the shape layer by layer. Harary and Harborth [6] use spirals to explore the extremal “animals” (simply connected subsets of tilings of the plane by regular triangles, hexagons and squares), which minimize the length of the perimeter for a number of tiles. Wang and Wang [13] provided a “spiral like” construction for the space \(\mathbb{Z}^n\) for any \(n\), generating an infinite sequence \(T^{(n)} = (x_i)_{i=1}^{\infty}\) of
tiles $x_i \in \mathbb{Z}^n$, such that any finite prefix of this sequence corresponds to a shape that has maximal compactness in the sense of having minimal boundary size for its number of $\mathbb{Z}^n$ tiles. The initial part of this construction for $\mathbb{Z}^2$ is given in Figure 3.

![Figure 3: A prefix of the Wang & Wang spiral.](image)

As discussed above, “spiral constructions” are known to yield isoperimetric inequalities and thus the lower boundary of the $F$ diagram for the corresponding discrete spaces. Although very pleasing, these constructions leave several interesting geometric questions open, and we shall address some of them herein.

Let us look at the set of shapes in Figure 3 again. Since $p(S)$ (the perimeter) increases more slowly than $a(S)$, and both are in $\mathbb{Z}_+$, not every shape in the sequence can have a larger perimeter than its predecessor. Therefore, not all shapes in this sequence can have maximal area for their perimeter. Which of those shapes are optimal in this sense as well? Does this sequence of shapes exhibit all shapes with smallest perimeter for their area? The answer to this second question is no. We have already seen two shapes $A$ and $B$ that will turn out to be Pareto optimal and have the same area, and clearly only one of them can appear in Wang and Wang’s sequence, or any similar spiral construction.

We will present our results in the sequel, provide proofs in sections 4-6, and discuss both related analyses and some possible applications in Section 7. Some conclusions about this style of analysis compared to others are proposed in Section 8.

3 Brief Overview of Results

In this work we shall completely characterize the set of Pareto optimal shapes for three planar discrete spaces. In the case considered above ($\mathbb{Z}^2$ with the perimeter defined as the number of neighboring tiles sharing a side with the shape) our characterization will also exhibit Pareto optimal shapes that are not generated by the spiral construction. In the other two cases we shall prove the opposite: that a spiral construction does generate all the Pareto optimal shapes.

**Definition 3.1.** The expansion of a shape is the union of the shape with its neighbors. The expansions of a shape are the shapes that result by iterating this process.

In the tiling of $\mathbb{R}^2$ by squares, we can define the perimeter of a shape as the number of tiles outside of it that share with it a **side or a corner**. This we call the eight neighbor boundary, and its association with the tiling we call it the eight connected grid, or sometimes $X_8$. 


Theorem 3.2. A shape in the 8 connected grid is Pareto optimal if and only if it is the empty shape, or one of the shapes in Figure 4 or generated by iterated expansions of such a shape.

Figure 4: Shapes generating all optimal shapes for squares, when neighbors share a side or corner.

In the tiling of $\mathbb{R}^2$ by squares, we can define the perimeter of a shape as the number of tiles outside of it that share with it a side. This we call the four neighbor boundary, and its association with the tiling we call it the four connected grid, or sometimes $X_4$.

Theorem 3.3. A shape in $X_4$ is Pareto optimal if and only if it is the empty shape, or one of those shown in Figure 5, or it is generated by iterated expansions of such a shape.

Figure 5: Shapes generating all optimal shapes for squares, when neighbors share a side.

Theorem 3.4. Consider the tiling of $\mathbb{R}^2$ by hexagons, defining the perimeter of a shape as the number of tiles outside of it that share with it a side. Then a shape is Pareto optimal if and only if it is the empty shape, or one of those shown in Figure 6, or it is generated by iterated expansions of such a shape.

Figure 6: Shapes generating all optimal shapes for hexagons.

Brunvoll et al [4] mention that the shapes in Figure 6 arising from similar but distinct requirements, in the interesting context of organic chemistry, and we expand on this in Section 7.
We recall that digital straight lines were defined as half planes governed by real valued coefficients. It is not hard to see that each of the initial shapes above is generated by the intersection of several DSLs. We shall see below in the proofs of these results that all the optimal shapes can be described so, and the coefficients of the DSLs are very particular ones.

Another way to describe the Pareto frontier is to consider the sequences of \((a(S), p(S))\) pairs of Pareto optimal shapes. The sequences for neighborhood sizes are boring in that they include almost all natural numbers - after a short finite prefix, the 8 neighborhood case allows exactly all even perimeters, and the others allow all perimeters. As we might expect, the sequences for area grow more or less quadratically in their perimeter for all of these spaces, and there are some interesting connections. The connection between these shapes and the sequences were previously known, except apparently for Theorem 3.6.

**Theorem 3.5.** For the square tiling with the 8-neighbor boundary, for any \(n \in \mathbb{N}\), the area of a non-empty optimal shape with the \(n\)th smallest feasible perimeter is the maximal product of two integers whose sum is \(n\).

The characterization as the maximal product of two integers whose sum is \(n\) is given by Sloane in his wonderful Encyclopedia of Integer Sequences [11] for the sequence identified there as A002620. The first few values of this sequence are as follows: 0, 0, 1, 2, 4, 6, 9, 12, 16, 20, 25, 30, 36, 42, 49, 56, 64, 72. Note that the theorem above uses \(n \geq 1\), so 0 appears only once.

**Theorem 3.6.** For the square tiling with the 4-neighbor boundary, the sequence of areas of non-empty optimal shapes in order of increasing boundary size is given by \(B(0, m) \leq B(1, m) \leq B(2, m) \leq B(3, m) \leq B(0, m + 1) \leq B(1, m + 1) \leq \ldots\) for \(m \in \{-1, 0\} \cup \mathbb{N}\), where \(B(i, m)\) is defined as follows:

- For \(i = 0\), \(B(i, m) = 2m^2 + 6m + 5\).
- For \(i \in \{1, 2, 3\}\), \(B(i, m) = 2m^2 + (6 + i) m + 4 + 2i\).

This sequence of areas begins with 1, 2, 3, 5, 6, 8, 10, 13, 15, 18, 21, as seen in Figure 2. Consider the sequence \(\left\lfloor \frac{n^2}{8} + \frac{1}{2} \right\rfloor\), which begins 0, 0, 1, 1, 2, 3, 5, 6, 8, 10... and is numbered A001971 in the Sloane encyclopedia. The agreement between these sequences (starting from the fourth element of the latter) is not limited to the finite prefix we show here, it is described and proved in in subsection 5.3.

A better, geometric proof of this fact remains a challenge for the future.

**Theorem 3.7.** For the hexagonal tiling, the sequence of areas of non-empty optimal shapes in order of increasing boundary size is given by one occurrence of \(A(1, 0)\) followed by \(A(2, j) \leq A(3, j) \leq A(4, j) \leq A(5, j) \leq A(1, j + 1) \leq A(6, j) \leq A(2, j + 1) \leq \ldots\) for \(j \in \{0\} \cup \mathbb{N}\), where \(A(i, j)\) is defined so:

- For \(i = 1\), \(A(i, j) = 3j^2 + 3j + 1\).
- For \(i \in \{2, 3, 4, 5\}\), \(A(i, j) = 3j^2 + (3 + i) j + i\).
- For \(i = 6\), \(A(i, j) = 3j^2 + 10j + 8\).

This sequence begins with 1, 2, 3, 4, 5, 7, 8, 10, 12, 14, 16, 19, 21, 24, 27, 30 and is already known as sequence A001399. One of the characterizations given for
it is: the numbers of the tiles along a spiral at which “folds” occur, which is equivalent to the areas of the shapes occurring during that construction that have no extra corners, so that they can be defined naturally by 6 DSLs.

Let us next proceed to prove the above results. We shall analyze the 8 connected grid first, and it will serve as a template for the more complex spaces.

4 Pareto Optimality on the Eight Connected Grid

4.1 Optimality conditions

We consider \( \mathbb{Z}^2 = \{(i, j) | i, j \in \mathbb{Z}\} \) as the vertices of a graph, which has an edge between \((i, j)\) and \((m, n)\) iff \( \max\{|i-m|, |j-n|\} = 1\). Thus every vertex has 8 neighbors. We define \( d_8(x, y) \) to be the distance between \( x, y \in \mathbb{Z}^2 \) induced by this graph, which is simply the length of the minimal path between \( x \) and \( y \), and equal to \( \max\{|i-m|, |j-n|\} \). We note that \( d_8 \) is a natural measure of distance since it fulfills the requirements of a metric, including the triangle inequality. We will analyze shapes in the space consisting of the points of \( \mathbb{Z}^2 \) with the distance \( d_8 \), denoted \( X_8 = (\mathbb{Z}^2, d_8) \), and called the 8 connected grid.

We will construct other spaces by varying the set of points and the distance function. In the following discussion we assume \( X = X_8 \), though we will later see that many of the definitions and claims can be generalized to other spaces.

**Definition 4.1.** A shape \( A \) in \( X \) is a finite subset of \( X \). The neighborhood or boundary of \( A \) is \( N(A) = \{x \in X | d(x, A) = 1\} \).

Note that we do not restrict our analysis to connected shapes, a common assumption.

**Definition 4.2.** We call a shape \( A \) area optimal if it has the maximal area for its boundary size, or formally, if \( \forall B (|N(A)| \geq |N(B)|) \Rightarrow (|A| \geq |B|) \).

**Definition 4.3.** We call a shape \( A \) boundary optimal if it has a minimal boundary area for its area, or formally, if \( \forall B (|A| \leq |B|) \Rightarrow (|N(A)| \leq |N(B)|) \).

We shall call a shape optimal if it is Pareto optimal, meaning that it is both area optimal and boundary optimal. Our goal will be to characterize the set of optimal shapes.

The isometries of \( X \) are those mappings \( X \rightarrow X \) that preserve distances so \( f \) is an isometry if \( \forall x, y \in X d(x, y) = d(f(x), f(y)) \). Note that isometries of \( X \) are bijections, and also preserve the sizes of neighborhoods, therefore they preserve also optimality. The isometries of \( X_8 \) include all finite compositions of the mappings \((x, y) \mapsto (-x, y) \) (reflection), \((x, y) \mapsto (y, -x) \) (rotation by \( \frac{\pi}{2} \)) and \((x, y) \mapsto (x+1, y) \) (translation).

Note that these definitions are general, and will make sense even if we replaced \( d_8 \) by a different metric, induced by a different neighbor relation, and possibly with a different set of isometries. We shall indeed analyze two other spaces using many of the same definitions and techniques, after we’ve found the optimal shapes of \( X_8 \).
4.2 Some suboptimal shapes and notation

Given any shape $A$, we can find the rightmost column on which it has elements, and denote $a_{r,b} = (x,y)$ the element of that column which is lowest (in $A$’s the rightmost column, take the bottom element). Given a shape $B$, we can similarly find $b_{l,b}$ (in the leftmost column, the bottom element). Then there exists a translation $T$, such that $T(b_{l,b}) = (x+2,y)$. By the construction of $T$ and the definitions of $a_{r,b}$ and $b_{l,b}$, we find that $d(A,T(B)) = 2$, and that $|N(A) \cap N(T(B))| \geq |N(a_{r,b}) \cap N(b_{l,b})| = 3$. An example of this construction is given in Figure 7.

![Figure 7: Any two shapes can be translated to share 3 neighbors in $X_8$.](image)

We state a slightly weaker conclusion formally:

**Lemma 4.4.** For any two shapes $A, B$ in $X$, there exists a translation $T$ such that $A \cap T(B) = A \cap N(T(B)) = A \cap B = \emptyset$ so that $|A \cup T(B)| = |A| + |T(B)| = |A| + |B|$, and also $|N(A \cup T(B))| \leq |N(A)| + |N(B)| - 2$.

The reasoning given above actually supports $|N(A \cup T(B))| \leq |N(A)| + |N(B)|$, but the lemma as given holds across other spaces we shall discuss. We shall use this lemma to diagnose non-optimal shapes.

**Lemma 4.5.** Let $A$ be the union of non empty sets $B, C$ such that $B \cap C = B \cap N(C) = N(B) \cap C = \emptyset$ (so each part overlaps neither the other part nor its neighbors), and $|N(B) \cap N(C)| \leq 1$, so the overlap between their neighborhoods is small. Then $A$ is not boundary optimal.

**Proof.** Because $B \cap C = \emptyset$, $|A| = |B| + |C|$. Next we show that under the conditions described, $|N(A)| \geq |N(C)| + |N(B)| - 1$.

$$N(A) = \{x | d(x,A) = 1\} = \{x | d(x,C) \leq 1\} = \{x | d(x,C) = 1 \land d(x,B) \geq 1\} \cup \{x | d(x,C) = 1 \land d(x,B) \leq 1\} = (N(C) \setminus B) \cup (N(B) \setminus C) = N(C) \cup N(B)$$

Then $|N(A)| = |N(C) \cup N(B)| = |N(C)| + |N(B)| - |N(C) \cap N(B)|$ by the inclusion exclusion principle, but we are given that $|N(C) \cap N(B)| \leq 1$, finishing this part. Lemma 4.4 tells us that there exists $T$ such that $|B \cup T(C)| = |A|$, and $|N(C) \cup N(T(B))| \leq |N(C)| + |N(B)| - 2$. Then $B \cup T(C)$ has the same area as $B \cup C$, but strictly less neighbors, then $C \cup B$ is not optimal. \(\square\)
We now take some arbitrary $x_{\text{orig}} \in X$ as our origin. $x_{\text{orig}}$ has eight neighbors, we name them $x_0, \ldots, x_7$, moving counter clockwise from the neighbor to the left and below $x$. We represent directions in this space as $d_i = x_i - x_{\text{orig}}$ (Figure 8), and say that $i$ is the direction index of $d_i$. Denoting $A = \{1, 3, 5, 7\}$, $\{d_i\}_{i \in A}$ correspond to directions parallel to the axes. It is natural to treat the direction indices cyclically, so that the direction following $d_7$ is $d_0$. One way to state this is to consider direction indices as belonging to the finite group $\mathbb{Z}/8\mathbb{Z}$ in which $7 + 1 = 0$. Then $d_{i+4} = -d_i$ is the direction exactly opposed to $d_i$.

**Definition 4.6.** Let $i$ be a neighbor direction index (so that $d(x, x + d_i) = 1$), then the ray from $x_0$ in direction $i$ is the set $R_{x_0, i} = \{x_0 + pd_i | p \in \{0\} \cup \mathbb{N}\}$.

**Remark 4.7.** Clearly, $R_{x_0, i}$ is a connected subgraph of $X$. Note that connectedness in this paper always refers to the graph theoretic concept, not a topological one.

**Definition 4.8.** Let $i \neq j$. We define the cut along $R_{x_0, i}$ and $R_{x_0, j}$ to be the set $L_{x_0, i, j} = \{x_0\} \cup R_{x_0, i} \cup R_{x_0, j}$.

Again, $L_{x_0, i, j}$ is always a connected shape. This does not, however, guarantee that $X \setminus L_{x_0, i, j}$ is not also connected, for the same reason that the two diagonals of a chess board are each connected in this way, despite crossing. This is a peculiarity of the $\mathbb{Z}^2$ grid which can be solved by considering the complementary metric for background images [12]. For our purposes it is enough to note that a cut $L_{x_0, i, j}$ in $X_8$ partitions $X \setminus L_{x_0, i, j}$ into two infinite sets that do not share a side.

**Definition 4.9.** Let $L_{x_0, i, j}$ be a cut in $X_8$. We can write $X \setminus L_{x_0, i, j} = A \cup B$ where $A, B$ are disjoint, infinite, and do not share a side. Then we shall say that $A, B$ are separated by $L_{x_0, i, j}$.

**Lemma 4.10.** (The separation lemma) Let $A$ be a shape, and let $B, C$ be the two parts of $X \setminus L_{x_0, i, j}$ so that $N(A) \cap B$ and $N(A) \cap C$ are not empty, and $|N(A) \cap L_{x_0, i, j}| \leq 1$. Then $A$ is not boundary optimal.

This formalizes a fact that seems almost obvious from inspection of Figure 9.
Proof. First we will show that $A \cap L_{x_0,i,j} = \emptyset$, so that $A$ is the disjoint union of subsets of $B, C$. Assume that $a \in A \cap L_{x_0,i,j}$, then $a$ has at least two neighbors along $L_{x_0,i,j}$ (because $i, j$ are neighborhood directions). Each of them either is, or is not in $A$. If it is not, then we have found an element of $N(A) \cap L_{x_0,i,j}$. If it is in $A$ the same argument can be applied again. This will only occur a finite number of times because $A$ is finite, then eventually we will find a neighbor of $A$ in each of the two directions along $L_{x_0,i,j}$. So the presence of $a$ entails $|N(A) \cap L_{x_0,i,j}| \geq 2$, which contradicts a given assumption.

Let $n \in X \setminus L_{x_0,i,j}$, wlog we assume that $n \in B$. Now let us assume also that $n \in N(A \cap C)$, and show this leads to a contradiction. We have seen that in $X_8$, elements of $B$ and of $C$ cannot share a side (then in particular, this is true of $n$ and each element of $A \cap C$). Then because $n \in N(A \cap C)$, they must share a corner, we assume wlog that $a_C$ is the tile in $A \cap C$ sharing this corner with $n$. Then $n, a_C$ are on a diagonal crossing $L_{x_0,i,j}$. Then $a_C$ shares sides with two elements of $L_{x_0,i,j}$, in contradiction to given assumptions. This shows that $A \cap C$ does not have a neighbor across $L_{x_0,i,j}$. Clearly the same holds for $A \cap B$.

From this we conclude three facts:

1. Any element of $N(A) \cap B$ is a neighbor of $A \cap B$, and similarly any element of $N(A) \cap C$ is a neighbor of $A \cap C$. In particular, neither of $A \cap C$ and $A \cap B$ is empty.

2. Any neighbors shared by $A \cap B$ and $A \cap C$ must be in $L_{x_0,i,j}$.

3. Neither of $A \cap B$, $A \cap C$ can overlap a neighbor of the other.

Then we have shown that $A$ is a disjoint union of subsets of $B, C$, neither subset is empty, each of $A \cap B, A \cap C$ does not overlap the others neighborhood, and the only possible common neighbor of both is in $L_{x_0,i,j}$. Then we can apply Lemma 4.5 to conclude that $A$ is not optimal.

Remark 4.11. Note that most of this proof depended only on Lemma 4.5 and definitions. The only point at which we considered features particular to $X_8$ was when noting that a subset of $A$ on one side cannot have a neighbor from the other.

Lemma 4.10 will turn out to be very easily usable in proving that optimal shapes have specific forms. Here is a sketch of its use. Assume that we have
found $L_{x_0,i,j}$ such that $x_0$ is the only neighbor of a shape $A$ along $L_{x_0,i,j}$. Assume also that each of the connected components of $X \setminus L_{x_0,i,j}$ contains an element of $N(A)$, then by the lemma $A$ is not optimal.

At the beginning of this section, we expended some effort to specify particular elements in shapes. During the rest of this paper, we will find ourselves doing this more than once, so it is useful to take a moment to introduce some concise and general notation.

Let $i \in \mathbb{A}$ then $d_i$ (called the main direction) is parallel to the axes. Let $j \notin \{i, i+4\}$ (so $d_j$ is neither $d_i$ nor its inverse). Any such choice induces a coordinate system on $X$, because for each $x$ there exists unique $p, q \in \mathbb{Z}$ such that $x = x_{\text{orig}} + pd_i +qd_j$. We cannot have both directions be diagonal, because then coloring the plane as a chess board, we would find that we can represent using integer coefficients only those tiles sharing the color of the origin.

An aside: in a finite dimensional vector space, for a large enough set of elements to be a basis it is sufficient for them to be linearly independent. Here we require another condition because $\mathbb{Z}^2$ is not a vector space, following the fact that $\mathbb{Z}$ is not a field. In algebraic terms, $\mathbb{Z}$ is a commutative ring with a unit, and $\mathbb{Z}^2$ is a $\mathbb{Z}$ module, and we require our two elements to form a basis in that sense. A necessary and sufficient condition for two vectors to form a basis for $\mathbb{Z}^2$ is that the triangle they define not contain any vertices of $\mathbb{Z}^2$ other than the origin and vector ends - this follows from Pick’s theorem for triangles [7].

Figure 10 shows two examples of coordinate systems induced by such choices of directions. One is rather standard, and the other is not. In the second, if we set $a = x_{\text{orig}}$, then $b = x_{\text{orig}} + 1d_i + 0d_j$, $c = x_{\text{orig}} - 1d_i + 1d_j$, $d = x_{\text{orig}} - 2d_i + 2d_j$, $e = x_{\text{orig}} - 1d_i + 2d_j$.

**Figure 10:** A shape and two coordinate systems.

**Definition 4.12.** Given $i, j$ as described above, we define the function $\phi_{i,j} : X \to \mathbb{Z}^2$ as $x \mapsto (p, q)$ where $x = x_{\text{orig}} + pd_i +qd_j$. The function $\phi_{i,j}$ is called the coordinate mapping with directions $i, j$.

So $\phi_{i,j}$ assigns to every element in $X$ its coordinates in the coordinate system defined by $x_{\text{orig}}, i, j$. We will always assume $x_{\text{orig}}$ to be constant throughout, and only $i, j$ vary, so $\phi_{i,j}$ is well identified.
Note that in our example, for \( j = 4 \), tiles \( c, e \) have the same first coordinate. So the second direction \( j \), which can be diagonal or not, allows us to divide the plane into equivalence classes, each of which is a line in the direction \( d_j \). The direction \( i \) induces an order on these lines.

**Definition 4.13.** Let \( \phi_{i,j} \) be a coordinate mapping and \( A \) a shape. Partition \( A \) into layers \( A_k = \{ x \in A | \phi_{i,j}(x) \in \{ k \} \times \mathbb{Z} \} \) so that all elements of a layer have the same coordinate in the direction \( i \), and vary in the \( j \) coordinate. Let \( k_{\text{max}} = \max_{A_k \neq \emptyset} k \) so that \( A_{k_{\text{max}}} \) is the layer with maximal \( k \) that is not empty. Then \( \psi_{i,j} : \mathcal{P}(X) \rightarrow X \) is defined so that \( \psi_{i,j}(A) \) is the element of \( A_{k_{\text{max}}} \) with maximal \( j \) coordinate.

In Figure 10, \( \psi_{3,5}(B) = e \) and \( \psi_{3,4}(B) = b \). This notation allows us to easily identify some extremal elements in a shape, for example in the beginning of this section, we could have written simply that \( a_{r,b} = \psi_{3,1}(A) \) and \( b_{l,b} = \psi_{7,1}(B) \).

### 4.3 The optimal shapes on the 8 connected grid are rectangles

Starting from the next subsection, we will consider the optimality of shapes from a limited set of rather simple shapes.

**Definition 4.14.** We call \( A \) a rectangle if \( A = \{ (a, b) | a \in [c, d] \land b \in [e, f] \} \). In this case, we say the dimensions of \( A \) are \( j = d - c + 1, k = f - e + 1 \).

We note that the dimensions of a rectangle define it up to translation. Optimality does not depend on location, but only on dimensions, so we will consider only those. Similarly optimality is invariant to rotation by a quarter turn, so we can assume that \( j \leq k \). Our task of characterizing the optimal shapes will clearly be simpler when we restrict ourselves to the rectangles, but to justify doing so, we must first prove they are sufficient.

**Lemma 4.15.** Any optimal shape in \( X_8 \) is a rectangle.

The proof of this claim is one of three very similar proofs we will see. The following ideas occur in each of them (Figures 11 and 12 illustrate them):

1. We assume that there exists an optimal non-simple shape \( A \). To show this leads to a contradiction, we consider the minimal simple shape \( B \) s.t. \( A \subset B \). Then \( A \neq B \Rightarrow |A| < |B| \).
2. Then if we show a mapping \( m : N(B) \rightarrow N(A) \) that is one to one, then \( B \) has no more neighbors than \( A \) does, then \( A \) is not area optimal.
3. We construct this mapping by assigning to each element \( n \) in \( N(B) \) a direction into \( B \). We will later choose the direction of each ray more precisely, so that it points towards \( A \). We then define its end point \( m(n) \) to be the first element of \( N(A) \) encountered by a ray from \( n \) along this direction.
4. In order to make the mapping be well defined and one to one, use the following geometric facts to choose the assigned directions:

(a) Rays from adjacent elements of \(N(B)\) which are in parallel directions cannot meet. But for all the rays to point inward, we must allow changes of directions, and cannot assume all rays will be parallel.

(b) \(B\) is minimal, therefore some extremal elements of \(A\) are also extremal in \(B\). So \(B\) shares some neighbors with \(A\). If \(n \in N(A) \cap N(B)\), then a ray from \(n\) also ends in \(n\), never leaves it and is in no danger of collision with other rays. This allows to change the assigned directions wherever \(A\) is adjacent to \(N(B)\).

(c) For a moment suppose that a ray divides \(N(B)\) into two parts, but never meets an element from \(N(A)\), then our mapping might not be well defined, because the ray does not end. But our ray is into \(B\), so it intersects \(N(B)\) at another point (call it \(q\)) dividing it into two parts. Because the direction associated with \(q\) must also be into \(B\), it cannot be the same as that associated with \(p\), therefore somewhere along each of the two parts of \(N(B)\), which are paths from \(p\) to \(q\), the direction must have changed. Direction changes are only associated with points at which \(N(A)\) and \(N(B)\) coincide, then each component of \(X\setminus L\) contains an element of \(N(A)\), therefore an element of \(A\), and we can apply the separation lemma to contradict the assumed optimality of \(A\).

(d) If two rays meet from different directions and their points \(n_a, n_b \in N(B)\) are not adjacent, then as we argued above, on each of the two paths between them along \(N(B)\), the directions changed, allowing the use of the separation lemma.

5. We conclude that if \(A\) were optimal, then the mapping \(m\) would be well defined and one to one, proving that \(A\) is dominated by the minimal simple shape containing it \(B\). Then optimal, non-simple shapes do not exist.

Now we will follow this plan for \(X_8\).

**Proof.** Let \(A\) be some optimal shape, then let \(B\) be the minimal rectangle such that \(A \subset B\). It is clear that \(|B| \geq |A|\), we assume the inequality is proper, because otherwise \(A = B\) and we are done. Then it remains to show that \(|N(A)| \geq |N(B)|\). We will do this by showing a mapping \(m : N(B) \to N(A)\), and then proving that it is one to one.

Let \(i \in k\), then \(d_i\) is parallel to one of the axes (for example, it may be the negative vertical direction), then we denote \(a_i = \psi_{i, i+2}(A)\) (in this case, the rightmost element in the bottom row of \(A\)). We denote also \(n_i = a_i + d_i\), then note that \(n_i \in N(A) \cap N(B)\). \(N(B)\) has its sides along the rectangle \(B\), and \(\{n_i\}_{i \text{ odd}}\) are each on one of these sides, and not on its corners. Thus \(\{n_i\}_{i \text{ odd}}\) partition \(N(B)\) into 4 connected components, each of which is denoted \(N_{i,i+2}\) if it is bounded by \(n_i, n_{i+2}\). An example of this construction is provided in
Figure 11. \( N_{i,i+2} \) is non empty because it contains at least a corner of \( N(B) \), so that \( \{n_i\} \cup N_{i,i+2} \cup \{n_{i+2}\} \) is shaped like an “L” (as in Figure 12). To each \( n \in N(B) \) we associate a direction \( d(n) \) so: if \( n = n_i \), then \( d(n) = d_i \), if \( n \in N_{i,i+2} \), then \( d(n) = d_{i+1} \).

Note that for every \( n \in N(B) \), \( n - d(n) \) is a member of \( B \) that neighbors. Our mapping \( m \) will simply go further along a ray in the same direction until the first neighbor of \( A \) is reached. Formally we will define \( p(n) = \min \{q \in \{0\} \cup \mathbb{N} : \{n - q d(n) \in N(A)\} \}, \) allowing us to define the mapping \( m : N(B) \to N(A) \) as \( m(n) = n - p(n) \cdot d(n) \). One might reasonably worry about the soundness of these definitions if there exists no \( q \in \{0\} \cup \mathbb{N} \) such that \( n - q d(n) \in N(A) \). We will assume this is the case, show this leads to a contradiction, and conclude such a \( q \) must exist.

Now we consider \( L = L_{n,d(n),-d(n)} \), and note that it separates \( X \) into two connected components \( C, D \) (as in Figure 12) so that their disjoint union is \( X \setminus L \), and so that \( N(A) \subset C \cup D \). In particular, \(|N(A) \cap L_{n} |= 0 \leq 1 \), then if we show that each of \( C, D \) has an element of \( N(A) \), we can conclude that \( A \) is non optimal by the separation lemma, in contradiction to our assumption.

If \( n = n_i \) for some odd \( i \), then \( L \) separates the rectangle \( N(B) \) into two components each containing a whole side, and in particular one of \( n_{i+2} \). If \( n \notin \{n_i\} \), odd then we write \( d_i = d(n) \), we then note that \( L \) separates \( N_{i,i+2} \) into two components such that each is continued by one of \( n_i, n_{i+2} \). In either case, each of \( C, D \) contains at least one element from \( \{n_i\} \), odd .

Then a finite \( q \) exists, and the definitions of \( p(n) \) and \( m(n) \) are sound.

Now it remains to show that \( m(n) \) is one to one. To do this, we assume that there exist distinct elements \( n, n' \in N(B) \) that have \( m = m(n) = m(n') \), and show this leads to a contradiction. We note that by the definition of \( m \), the only element in \( L_{m,d(n),d(n')} \), that can belong to either \( A \) or \( N(A) \) is \( m \in N(A) \). Again we will use the separation lemma to contradict the optimality of \( A \), and
Figure 12: Diagram of a separation lemma argument, if \( p(n) \) is not finite. Each line divides the plane into a left half, which we call \( C \), and a right half called \( D \). The vertical line is created when \( n = nc = n_5 \), then each of \( C, D \) contains one of \( nb, nd = n_3, n_7 \). The diagonal line corresponds to the case in which \( n \in N(B) \) is not one of \( \{n_i\} \), but in \( N_{3.5} \) placing each of the delimiters of \( N_{3.5} \), which are \( nb, nc = n_3, n_5 \) in one of \( C, D \).

It is necessary to show each component contains an element from \( N(A) \). We note that \( d(n) \neq d(n') \), because otherwise their lines are parallel and they do not have a common \( m \).

If we can write \( n \in N_{i,i+2} \) and \( n' \in N_{j,j+2} \) where \( i \neq j \) are odd, then \( n_i \) and \( n_{i+2} \) are separated by \( L_{m,d(n),d(n')} \).

Otherwise, at least one of the rays starts at one of the special elements \( n_i \), and wlog we assume \( n = n_i \), then we note that \( m(n_i) = n_i \), which is in \( N(A) \cap N(B) \). Now we consider in turn all the possibilities for the direction \( j \) corresponding to \( n' \). If \( j \) is \( i \pm 1 \), then the ray corresponding to it must arrive at \( n_i \) from outside the rectangle \( B \cup N(B) \), which is inconsistent with the construction. Similarly, \( j = i \pm 2 \) would require \( n' = n_j \) to be at a corner of \( N(B) \) which is again inconsistent with the construction, because then \( n_j - d_j \notin A \). For other \( j \), \( L_{m,d_j,d_i} \) admits a separation argument, using \( n_{i-1} \) and \( n_{i+1} \).

4.4 The optimal rectangles on the 8 connected grid have almost equal sides

In this section, knowing that all optimal shapes are rectangles, we will pause to relate our notation to well known facts about rectangles, then find a subset of rectangles that includes all optimal ones, and then prove that this subset is exactly the set of optimal shapes.

**Lemma 4.16.** A rectangle of dimensions \( j, k \), has area \( |A| = j \cdot k \), and boundary size \( |A'| = 2j + 2k + 4 = 2 \cdot (j + k + 2) \).

**Claim 4.17.** Let \( A \) be an optimal rectangle, then \( |j - k| < 2 \).
Proof. Because rotation by \( \frac{\pi}{2} \) is an isometry, it does not affect optimality. Thus, we may assume wlog that \( j \geq k \). Now we assume that \( j = k + 1 + m \), where \( m \in \mathbb{N} \), and show \( A \) is not optimal.

Now we consider \( B \), a rectangle with dimensions \( j - 1, k + 1 \), then \( |B'| = 2 \cdot (j - 1 + k + 1) = 2 \cdot (j + k + 2) = |A'| \).

On the other hand, \(|A| = k (k + 1 + m) = k^2 + k + km < k^2 + km + k + m = (k + 1) (k + m) = |B|\), then \( A \) is not area optimal. \( \square \)

Claim 4.18. All rectangles with \( |j - k| < 2 \) are optimal.

We already know that \( |j - k| < 2 \) is a necessary conditions for optimality, now we want to show it is sufficient as well. In other words, that for no shape \( A \) fulfilling the condition, there exists a shape \( B \) strictly improving on it. To do this it is sufficient to compare \( A \) only to other shapes \( B \) that also fulfill the necessary condition - but we need to explain why.

Lemma 4.19. Let \( O \) be the set of optimal shapes in \( X \), and let \( S \supseteq O \). If for every two shapes \( A, B \in S \), \( B \) does not improve on \( A \), then \( S = O \).

Proof. Assume for a moment that the shape \( B \) improves on \( A \) (we will write this here as \( B > A \)). Then by definition, either \( B \) is optimal or \( (\exists B_1) B_1 > B \). Let \( (B_i)_{i \in A} \) be a maximal sequence so \( A \) is a prefix of \( \mathbb{N} \) and \( (\forall i \in A) B_i \succ B_{i+1} \).

If the sequence is finite, then its maximality implies that its last element is optimal, therefore fulfills the necessary condition. So if \( A \) can be improved by \( B \), it can also be improved by some optimal shape \( B_{m+} \), which must fulfill our necessary condition. Now all that remains is to show that infinite sequences of improvement are not possible.

Assume by contradiction we have an infinite sequence of improvements, so \( A = \mathbb{N} \). The neighborhood size, being a natural number, can only be decreased a finite number of times, then there exists \( N \in \mathbb{N} \) such that if \( i > N \) then \( N (B_i) = N (B_{i+1}) = \mathcal{N} \) and \(|B_i| < |B_{i+1}|\). In other words, from some point, our sequence of shapes must have strictly increasing areas. Now consider \( B_i \) as a sequence of subsets of \( \mathbb{R}^2 \), normalizing the tiling so that the area of a single tile is 1. Then the area of \( B_i \) in \( \mathbb{R}^2 \) is simply the number of tiles in \( B_i \), therefore it tends to infinity as \( i \) does. Next we will use the number of neighbors \( \mathcal{N} \) to bound from above the length of the curve that is the boundary of \( B_i \) in \( \mathbb{R}^2 \), by noting that the boundary of \( B_i \) is a subset of the boundaries of its \( \mathcal{N} \) neighbors. Our tiles have four sides, each of length one, then the \( \mathcal{N} \) neighbors of \( B_i \) are contained by a curve of length at most \( 4 \cdot \mathcal{N} \). Then so does \( B_i \). Now the length of the boundary of \( B_i \) in \( \mathbb{R}^2 \) is bounded independently of \( i \), then the isoperimetric inequality for \( \mathbb{R}^2 \) tells us that the area of \( B_i \) is also uniformly bounded, contradicting the infinite growth we saw above. \( \square \)

Note that this argument also holds for different spaces and tilings of \( \mathbb{R}^2 \), as long as their tiles have bounded number of sides and length of sides. Now we return to prove the claim.
Proof. Assume wlog that \( j > k \). Then any shape with \( |j - k| < 2 \) has one of the forms \((k, k)\) or \((k, k + 1)\). We define the partial order \( \prec \) among these pairs using the total order \((0, 0) \prec (0, 1) \prec (1, 1) \prec (1, 2) \prec \cdots \prec (k - 1, k) \prec (k, k) \prec (k, k + 1) \prec \ldots \). Now we note that the areas agree with this order: \((k - 1) \cdot k < k \cdot k < k \cdot (k + 1)\), and the neighborhoods sizes also agree with it: \(2 \cdot (k + k - 1 + 2) \leq 2 \cdot (k + k + 2) \leq 2 \cdot (k + k + 1 + 2)\). Then for any two shapes \( A, B \) fulfilling the necessary conditions, \( A \prec B \Rightarrow N (A) < N (B) \wedge |A| < |B|\), thus neither strictly improves on the other. 

Thus we can conclude:

Lemma 4.20. The set of optimal shapes in the eight-connected grid is the set of rectangles whose height and width differ by at most 1.

This directly results in Theorem 3.5. Theorem 3.2 also results, though it generates the rectangles using expansions. Expansions are used for consistency with the results of the following sections, where they arise quite naturally.

Thus we have characterized all the optimal shapes in the 8 grid, and found them to consist of a simple family of rectangles. The basic steps we took in this process - proof of a lemma giving us the separation corollary, proof that optimal shapes must be simple rectangles and the elimination of non-optimal rectangles, will be repeated in the next few sections for two other spaces, though the simple shapes there will not be rectangles. For each of the 4-connected grid and the hexagonal grid, we will need new tactics to identify the non-optimal simple shapes, but the structure of the proof and many of the tools remain the same.

5 Pareto Optimality on the Four Connected Grid

In this section, we will consider two squares to be adjacent only if they share an edge. Formally, let \( x = (i, j) ; y = (m, n) \), so \( x, y \in \mathbb{Z}^2 \), then in our graph \( G = (\mathbb{Z}^2, E) \) there will be an edge \((x, y) \in E\) iff \(|i - m| + |j - n| = 1\). This induces the distance \( d_4 ((a, b), (c, d)) = |i - m| + |j - n| \), and in this section we consider \( X_4 = (\mathbb{Z}^2, d_4)\).

The definitions of the previous section carry over quite cleanly. Any finite set \( A \subset \mathbb{Z}^2 \) is a shape in \( X \), its neighborhood is the set of squares at distance exactly one, and formally \( N (A) = \{ x \in \mathbb{Z}^2 | d_4 (A, x) = 1 \} \). The new distance changes the shapes of neighborhoods, naturally affecting which shapes are optimal. As we noted, isometries must preserve neighborhoods, which again allows only translations, reflections, rotations by \( \frac{\pi}{2} \) and their compositions.

The possible directions \( \{d_i\}_{i=0}^7 \) are the same as in the previous section, though now \( x + d_i \) is a neighbor of \( x \) iff \( d_i \) is parallel to one of the axes, or equivalently, if \( i \in \mathbb{A} \). Similarly, the definitions for \( \phi_{i,j}, \psi_{i,j} \), rays \( R_{x_0, i} \) and cuts \( L_{x_0, i,j} \) are all unchanged. Note that in \( X_4 \), the neighborhood direction indices \( i, j \) allowed in rays and cuts are just those in \( \mathbb{A} \). Then it is easy to see that in \( X_4 \), any cut \( L_{x_0, i,j} \) separates \( X \setminus L_{x_0, i,j} \) into two connected components in \( X_4 \).
In the previous section we proved from general arguments the separation corollary, relying on Lemma 4.4 we proved in $X_8$. This lemma holds also in $X_4$, though with a slightly different proof, exemplified in Figure 13.

**Proof.** Let $A, B$ be shapes in $X_4$, and let $a = \psi_{3,4}(A)$ (among the diagonals from bottom left to top right, take the lowest passing through $A$, then $a$ is the rightmost element of $A$ on it) and $b = \psi_{7,0}(B)$ (by symmetry, the leftmost element on the topmost such diagonal of $B$). Then there exists a translation $T$ such that $T(b) = a + d_2$ (see an example in Figure 13). By the construction of $T$ and the definitions of $a, b$, we find that $d(A, T(B)) = 2$, and that $N(A) \cap N(T(B)) \supset N(a_{r,b}) \cap N(b_{l,b}) = \{a + d_1, a + d_3\}$, then $|N(A \cup T(B))| \leq |N(A)| + |N(B)| - 2$.

![Figure 13: Any two shapes can be translated to share 2 neighbors in $X_4$.](image)

From this proof of Lemma 4.4 for $X_4$, Corollary 4.5 for $X_4$ follows. Next we will prove also Corollary 4.10 for $X_4$.

**Proof.** See the proof for $X_8$, what remains to be proved here is that given $L_{x_0,i,j}$ such that $A \subset X \setminus L_{x_0,i,j}$, and $X \setminus L_{x_0,i,j} = B \cup C$, and $n \in B$, it is impossible that $n \in N(A \cap C)$. But here this is clear because $B, C$ are not connected.

In $X_8$, we considered rectangles which are shapes defined as the points fulfilling some inequalities on their coordinates. In $X_4$ we define the set of simple shapes in a similar way.

**Definition 5.1.** We call a shape $A$ simple in $X_4$ if it can be written in the form $A = \{(a, b) | a + b \in [j_1, j_2] \cap a - b \in [k_1, k_2]\}$. We will say its dimensions are $j = j_2 - j_1$ and $k = k_2 - k_1$.

See Figure 14. Recalling that isometries and in particular rotation by $\frac{\pi}{2}$ do not affect optimality, we may assume wlog that $j \leq k$.

**Lemma 5.2.** Optimal shapes in $X_4$ are simple.

**Remark 5.3.** The following proof is very similar to that for $X_8$. In particular, we will again show a one to one mapping into the neighbors of an arbitrary optimal shape from the neighbors of a simple shape with at least as much are. The construction of this mapping is changed only in its details due to the different geometry.
Proof. Let $A$ be an optimal shape, we assume by contradiction that it is not simple, then it is a strict subset of $B$, the minimal simple shape such that $A \subseteq B$. Then it will suffice to show a one to one mapping $m : N(B) \to N(A)$, because then $B$ has no more neighbors and greater area than $A$, therefore $A$ is not optimal.

Let $i$ be even, then $d_i$ is a diagonal direction, then we denote $a_i = \psi_{i,i+2}(A)$. In this case we denote also $n_i = a_i + d_{i+1}$, then note that $n_i \in N(A) \cap N(B)$. \{n_i\}_i\text{even} are each on at least one of the sides of $N(B)$ (possibly more - $n_i$ can be in a corner). Thus \{n_i\}_i\text{odd} partition $N(B)$ into (at most) 4 connected components, each of which is denoted $N_{i,i+2}$ if it is bounded by $n_i, n_{i+2}$. To each $n \in N(B)$ we associate a direction $d(n)$ so: if $n \in \{n_i\} \cup N_{i,i+2}$, then $d(n) = d_{i+1}$. An example is given in Figure 15.

Now we define $p : N(B) \to \mathbb{Z}$ as $p(n) = \min_{q \in \mathbb{N} \cup \{0\}} \{n - q \cdot d(n) \in N(A)\}$ so that $p(n)$ is the number of steps necessary to reach any element of $N(A)$ from a particular $n \in N(B)$, in the direction $-d(n)$. For $n \in \{n_i\}_i\text{even}$, this is always zero, because $n \in N(B) \cap N(A)$. The same holds for the element $n \in N_{i,i+2}$ that is nearest to $n_{i+2}$ because it is adjacent the same element of $A$. For other $n \in N_{i,i+2}$, $p(n)$ is finite, otherwise a separation argument applied to $n_i, n_{i+2}$ tells us $A$ is not optimal.

Now we can define the mapping $m : N(B) \to N(A)$ as $m(n) = n - p(n) \cdot d(n)$. Then it remains only to show it is one to one. Assume by contradiction that $n_a \neq n_b$, and $m = m(n_a) = m(n_b)$, then consider $L = L_{m,d(n_a),d(n_b)}$. $d(n_a) = d(n_b)$ implies a contradiction because then the rays from $m$ are in the same direction, therefore cross $N(B)$ at the same point, and $n_a = n_b$. We also know that $n_a, n_b \notin \{n_i\}_i\text{even}$ because if either is there, then $m \in N(B)$, and then again $n_a = n_b$ because the rays from $m$ cannot pass through another

![Figure 14: A simple shape and its dimensions.](image-url)
Figure 15: Construction of the four elements \( \{n_i\} \), on three distinct \( a_i \).

element of \( N(B) \). The case that remains is that \( n_a \in N_{i,i+2} \), and \( n_b \in N_{j,j+2} \), for \( i,j \) even and distinct. Then we can apply a separation argument with \( L \) separating \( n_i \) from \( n_j \), as in Figure 16.

Figure 16: If the vertical and horizontal rays meet, then \( n_6 \) in the diagram (which refers to \( n_6 \)) is separated from \( n_4 \) (\( n_4 \)). Both are neighbors of \( A \), which allows a separation argument against the optimality of \( A \).

5.1 Describing optimal simple shapes in \( X_4 \)

It now remains to identify which of the simple shapes are optimal. These come in more varieties than rectangles, thus it will take a few steps, and it will be...
useful to introduce a few tools for the description of such shapes.

**Definition 5.4.** A simple shape $A$ with $j = 0$ is called a degenerate shape.

Note that degenerate shapes behave differently from other simple shapes (for example, in the degenerate case $k$ cannot have odd values, because the Manhattan distance between two tiles on a diagonal is always even).

**Lemma 5.5.** The only optimal degenerate shapes have an area of 0, 1 or 2.

*Proof.* If $A$ is degenerate, then $j = 0$. We assume by contradiction that $k \geq 4$ (see Figure 17) and $A$ is optimal. But the shape $B$, created by placing all the tiles in the same column has exactly as many neighbors (two horizontal neighbors per tile, and two additional vertical neighbors), and the same area, but is not simple, therefore is not optimal. Then $A$ cannot be optimal. The shapes with $k \leq 3$ have areas as described, as can be seen in Figure 21. \(\square\)

We have seen that only a small and finite set of degenerate shapes is of interest to our discussion. Henceforth assume any mention of simple shapes refers to those that are not degenerate, if these are not mentioned explicitly.

**Figure 17:** A degenerate shape $A$, and a variation $A'$ which clarifies that $A$ is non optimal

**Lemma 5.6.** Every simple shape has $j + k + 4 \ 4$-Neighbors.

*Proof.* By induction on $j$ and $k$. This is true for the shape of two neighboring tiles (i.e $j = k = 1$), and 6 neighbors.

In the induction step, we assume validity for $j, k$ and prove it for $j + 1$ (the same reasoning applies to expansion in $k$). Adding 1 to $j$ causes one diagonal side (having $r$ tiles) to expand to some direction (an expansion up is illustrated in Figure 18).

As a result, $r$ neighbors in that direction become new tiles, and $r$ vacant beyond those in the same direction become neighbors, not modifying the neighborhood size yet. However, the new tile that is last in the direction of advancement is exposed to a new neighbor from the side. Having been diagonal to an extreme tile in the shape, it was not a neighbor before (i.e. increasing $j$ or $k$ adds one neighbor), thus a shape of dimensions $(j+1), k$ has $j + 1 + k + 4$ neighbors, completing the induction step. \(\square\)

Next we will describe each simple shape as a spine expanded by an iterative expansion process.
Lemma 5.7. Let $A$ be a simple shape of dimensions $j, k$. Then its expansion (as in Definition 3.1) is a simple shape of dimensions $j + 2, k + 2$.

Thus, the number of tiles of the shape increases by $j + k + 4$, and the number of neighbors grows by 4. See Figure 19 for an example.

Lemma 5.8. Let $A$ be a simple shape with dimensions $j, k$. After $s$ expansion steps, its neighborhood grows by $4s$ and its area grows by $E(j, k, s) = s(2 + j + k + 2s)$.

Proof. The neighborhood grows linearly, being equal to $j + k + 4$. $E(j, k, s)$ is defined as the number of tiles added to a simple shape of dimensions $j, k$ by a sequence of $s$ expansions, therefore:

$$E(j, k, s) = \sum_{i=0}^{s-1} ((j + 2i) + (k + 2i) + 4) =$$

$$s \cdot (4 + j + k) + 4 \cdot \sum_{i=0}^{s-1} i =$$

$$s \cdot (4 + j + k) + 4 \cdot \frac{s \cdot (s - 1)}{2} =$$

$$s (2 + j + k + 2s)$$

Definition 5.9. A simple shape such that $j \in \{1, 2\}$ is called a spine.

Lemma 5.10. A simple shape $A$ can be described as a spine, expanded some finite number (possibly zero) of times. This description is unique.
We shall later show that there are only 4 kinds of spines. Thus, since we know the area added by each expansion step, we will be able to calculate the areas of all simple shapes.

Proof. If \( j \) of \( A \) is even, we say that \( A_s \) has dimensions \( 2, k - j + 2 \), otherwise \( 1, k - j + 1 \). Either way, \( A_s \) is a spine and expanding it \( s = \lceil \frac{j}{2} \rceil - 1 \) times yields exactly \( A \). Then the area of every simple shape is the sum of the area of its spine \( A_s \) and the area added in the expansions.

We note that starting from any other spine will result in the wrong shape - a different initial width results in wrong parity of the final width, and the same spine width but different different length results in a wrong difference between length and width. Therefore this description is unique.

**Lemma 5.11.** Let \( A_s \) be a spine of dimensions \( j, k \), then its area is given by (See Figure 19):

1. If \( j = 1 \), the area is \( k + 1 \)
2. If \( j = 2 \), then we have the following options:
   
   (a) If \( k \) is odd, then \( |A_s| = \frac{3(k+1)}{2} \).
   
   (b) \( k \) is even, of type 1, then \( |A_s| = \frac{3k}{2} + 1 \).
   
   (c) \( k \) is even, of type 2, then \( |A_s| = \frac{3k}{2} + 2 \).

Proof. For \( j = 1 \), there are \( k \) tiles at distances 0 to \( k - 1 \) from one line, and one more. For \( j = 2 \), there are \( \lfloor \frac{k+1}{2} \rfloor \) triplets of tiles. Note that there are two ways of getting from an odd \( k \) to an even one, depending on which boundary is moved, resulting in different area increases.

**Figure 19: Spine types and their areas**

5.2 Spines of optimal shapes are short

**Lemma 5.12.** Let \( A_s \) be a spine with dimensions \( j, k \) of an optimal shape \( A \), then \( j + 4 > k \).
Later we will show that this result, while necessary in our analysis, is not tight.

Proof. We assume by contradiction that $A$ is an optimal shape with spine $j+4 \leq k$, extended $s$ times. Then we take the same skeleton with $b$ shortened by 4, and expanding it $s + 1$ times we get $B$, such that $|N(A)| = |N(B)|$.

We will now show that $|B| > |A|$, contradicting the optimality of $A$. $|A|$ is the sum of spine size and $E(a, b, s)$.

First we note that $E(j, k - 4, s + 1) - (E(j, k, s)) = j + k$, then if $j = 1$, the area of the skeletons is $k - 4 + 1$ and $k + 1$, then

$$k - 4 + j \geq j + 4 - 4 + j > 0 \Rightarrow (k - 4 + 1) + (j + k) > k + 1.$$

If $j = 2$, in all the variations, subtracting 4 from $k$ reduces the skeleton area by precisely 6, but the expansions more than offset that because $j + k \geq 2j + 4 > 6$.

Corollary 5.13. The dimensions of spines of optimal shapes are a subset of: 

\{(1, 1), (1, 2), (1, 3), (1, 4), (2, 2), (2, 3), (2, 4), (2, 5)\}

Recalling Lemma 5.11, we note that spines of dimensions \{(2, 2), (2, 4)\} mentioned above come in two types. As we saw there, type 2 spines have strictly more area than those of type 1, with the same neighborhood size. Therefore only type 2 spines can result in optimal shapes. In this context, each set of spine dimensions results in a certain optimal spine area and neighborhood size.

This allows us to restrict our attention to a set of shapes small enough to apply elimination.

Lemma 5.14. Let $A$ be a non degenerate optimal shape with dimensions $j, k$, so that $|N(A)| = 4(m + 1) + i$, with $i \in \{0, 1, 2, 3\}$. Then: $|A| = 2m^2 + (1 + i)m + \max\{1, i\}$

Proof. Let $a, b$ be the dimensions of $A$’s spine, then remembering each expansion increases the neighborhood size by 4, we see that $4(m + 1) + i = j + k + 4 = a + b + 4(s + 1)$. One conclusion is that $a + b \equiv i \mod 4$, and another is that $s = \frac{4m + i - a - b}{4}$. Hence, denoting $|A_s|$ the area of the skeleton of dimensions $a, b$, the total area for such a shape is exactly $|A| = |A_s| + E(a, b, \frac{4m + i - a - b}{4})$.

Table 1 describes for each $i$ the possible spines for optimal shapes with $|N(A)| = 4m + i$, the shape’s area for each spine, and the spines resulting in shapes that are sub-optimal for that neighborhood size. The spines in Table 1 that are not marked as suboptimal fulfill the formula claimed.

We have shown one necessary condition for optimality of degenerate shapes and now another for non-degenerate shapes. As we have done for $X_k$, we now show a partial order $\prec$ on all shapes fulfilling the necessary conditions that agrees with area and with perimeter, thus proving the necessary conditions are also sufficient.
Table 1: For each spine that may be optimal, we find \( i \) such that the perimeter of that spine after \( m \) expansions is \( 4(m + 1) + i \). Then we calculate the expanded areas for those spines, and find some spines that result in shapes of smaller areas than those resulting from a different spine with the same \( i \) therefore the same perimeters.

A degenerate shape \( A \) fulfilling our condition has has area \( |A| \in \{0, 1, 2\} \) and perimeter \( N(A) \in \{0, 4, 6\} \) respectively, then the comparison \( \prec \) of two degenerate shapes is obvious. Next we consider its definition for two non degenerate shapes allowed by Lemma 5.14.

First we note that for any specific number of expansion \( m \), \( |A| \) and \( N(A) \) are strictly monotonous in \( i \). Furthermore \( |A| \) is strictly monotonous in \( m \) because 2 \((m + 1)^2 + 2(m + 1) + 1 = 2m^2 + 4m + 2 + 2m + 2 + 1 = 2m^2 + 6m + 5 > 2m^2 + 5m + 3 \). It is clear that \( N(A) \) is also monotonous in the number of expansions. Then we can choose \( \prec \) for non degenerate shapes to agree with the lexicographic ordering comparing first \( m \) and then \( i \), and thus have it agree with \( |A| \) and \( N(A) \).

Now all that remains is to consider the comparison of a degenerate shape to a non degenerate shape. The first (in terms of \( \prec \)) non degenerate shape we consider is that with spine \((1, 1)\) and no expansions (call it \( N_1 \)), and it has exactly the same area and perimeter as the last degenerate shape (call it \( D_2 \)), so it is sufficient to consider them equivalent in \( \prec \) as well, and this defines the result of comparing any degenerate shape \( A \) to any non degenerate shape \( B \): \( A \preceq D_2 = N_1 \preceq B \).

Thus, all shapes fulfilling our necessary conditions are optimal. We now state this result formally.

**Lemma 5.15.** The non-degenerate optimal shapes are those simple shapes whose spine \((a, b)\) is one of the following: \{\((1, 1)\), \((1, 2)\), \((2, 2)\), \((2, 3)\), \((2, 4)\)\}. These spines appear in Figure 20.
Lemma 5.16. The degenerate simple shapes with areas 0, 1, 2 are all optimal.

All these appear in Figure 21.

Corollary 5.17. Let $A_s$ be a spine with dimension $a, b$ of an optimal shape $A$, then $a+3 > b$.

This is a tighter version of Lemma 5.12, and can now be verified by inspection of the list of optimal spines.

Now to prove theorem 3.3, consider the union of degenerate optimal shapes and non-degenerate optimal shapes, noting that the non empty optimal degenerate shapes, when expanded once, are equal to some of the non-degenerate optimal shapes. Theorem 3.6 results from applying the formulas for area developed in Lemma 5.14, and the order induced by $\prec$.

5.3 Relation to the sequence of integers closest to $\frac{n^2}{8}$

Theorem 5.18. Let $(A_n, p_n)$ be the $n$th area-perimeter pair exhibited by the optimal shapes in $X_4$ (including the empty shape), when ordered with increasing area, for $n \geq 3$. Then $A_n$ is the integer closest to $\frac{(n+1)^2}{8}$ (rounding $n + \frac{1}{2}$ to $n + 1$).

Proof. The sequence of areas for non-empty shapes is given for $m \in \mathbb{N} \cup \{-1, 0\}$ and $i \in \{0, 1, 2, 3\}$ so:
1. For \( i = 0 \), \( B(0, m) = 2m^2 + 6m + 5 \).

2. For \( i \in \{1, 2, 3\} \), \( B(i, m) = 2m^2 + (6 + i)m + 4 + 2i \).

The proof is by induction on \( m \) for each of the four values of \( i \). Thus the base case should consider each \( i \). \( n = 3 \) requires us to consider the third feasible pair \((A_3, p_3)\), which is \((2, 6)\). This corresponds to a spine of dimensions \((1, 1)\), and parameters \( i = 2; m = -1 \), so that \( B(2, -1) = 2 \) is the desired area. Then all that remains is to see that this agrees with \( \left(\frac{n+1}{8}\right)^2 = \frac{2}{4} = 2 \). Similar verification for the other cases may be done by substitution as follows:

1. \( A_4 = 3 \) has spine \((1, 2)\), and parameters \( i = 3; m = -1 \).

2. \( A_5 = 5 \) has spine \((2, 2)\), and parameters \( i = 0; m = 0 \).

3. \( A_6 = 6 \) has spine \((2, 3)\), and parameters \( i = 1; m = 0 \).

Induction step. We assume the claim holds for \( i, m \), and prove it for \( i, m+1 \). In particular, if \( i, m \) corresponds to \( A_n \), then \( i, m+1 \) corresponds to \( A_{n+4} \), and so should the areas. Note that \( i, m \) corresponds to \( n \) iff \( n = 1 + i + 4(m + 1) \).

Now we note that \( \frac{n^2}{8} - \frac{(n-4)^2}{8} = \frac{8n-16}{8} = n - 2 \). Since the difference between the two is an integer, they are both rounded the same way. Then we need to prove that \( B(i, m) - B(0, m-1) = n - 2 = 4(m + 1) + i \). For \( i \in \{1, 2, 3\} \)

\[
2m^2 + (6 + i)m + 4 + 2i - \left(2(m - 1)^2 + (6 + i)(m - 1) + 4 + 2i\right) = \\
2m^2 + (6 + i)m - \left(2(m - 1)^2 + m(6 + i) - 1(6 + i)\right) = \\
2m^2 - 2(m - 1)^2 + (6 + i) = \\
4m + 4 + i
\]

Which is what we needed for \( i \neq 0 \). And the formula for \( B(0, m) \) differs from that for \( B(i, m) \) only by a constant, which is cancelled by the subtraction, so we can see the same holds there.

**Remark** 5.19. Note that \( A_1 \), which is the empty shape and has area 0, also corresponds to \( \frac{(n+1)^2}{8} = \frac{1}{2} \) which we round to 1, therefore the claim does not hold for \( n = 1 \). However \( A_2 \), having area 1, gives \( \frac{(2+1)^2}{8} = 1 + \frac{1}{8} \) and the closest integer to it is 1, as needed. This case, however, does not conveniently fall under the proof above.

### 6 Pareto Optimality on the Hexagonal Grid

#### 6.1 Definitions

We will now consider the tiling of \( \mathbb{R}^2 \) by hexagons. We will denote \( H \) the set of hexagons, and consider them as vertices in a graph. Two vertices in the graph
will be considered adjacent iff the corresponding hexagons share a side. We denote $d_{H} : H \times H \to \mathbb{R}^+$ the distance induced by this graph, so $d_{H}(h_a, h_b)$ is the length of the shortest path from hexagon $h_a$ to $h_b$, and consider the space $X_6 = (H, d_{H})$.

The definitions of shape, neighborhood, optimality given for $X_8$ and $X_4$ did not depend on the space $\mathbb{Z}^2$ beyond the metrics that were induced. Then we can apply the same definitions here as well, so in $X_6$ a shape $A$ is a finite subset of $H$, its neighborhood is $H \supset N(A) = \{ h \in H | d_{H}(A, h) = 1 \}$, and the optimality of a shape depends on the number of hexagons in it and on $|N(A)|$ as usual.

Two aspects of $X_6$ that are different to $X_4$, $X_8$ are the set of isometries and the set of available directions. Choosing $x_{\text{orig}} \in H$, we see it has 6 neighbors, and denoting them $x_0, \ldots, x_5$ in counterclockwise direction and starting from the lower neighbor on the left, we define the 6 available directions $d_i = x_i - x_{\text{orig}}$ for $i \in \{0, \ldots, 5\}$. Isometries preserve distances, therefore cannot change neighborhood relations, then in particular rotation by $\pi \frac{3}{2}$ is not allowed, because it does not map hexagons to hexagons. In fact, the allowed isometries are generated by rotation around any hexagon by $\pi \frac{1}{2}$, by reflection around $d_1$ and by translation in direction $d_1$.

We have defined directions in $X_6$, all of which are neighbor directions, and this is sufficient for us to apply here the definitions of rays $R_{x_0,i}$ and the cuts from $x_0$ along $i$ and $j$ denoted $L_{x_0,i,j}$ given for $X_8$.

For reasons that will become clear later, it is useful to use multiple coordinate systems on the hexagon tiling. To identify a particular one, we choose some particular hexagon $a \in H$, and choose $b, c \in N(\{h_0\})$ so that $b, c$ are neighbors also, and the cycle $a \to b \to c \to a$ is arranged counter clockwise. Then we note that $b - a$ specifies one direction $d_b$, and $c - a$ specifies another we will denote $d_c$, and for any hexagon $h$, there exists a unique pair $(p, q) \in \mathbb{Z}^2$ such that $h = a + pd_b + qd_c$, so that $h$ can be reached from $a$ by taking $p$ steps in direction $d_b$, and $q$ steps in direction $d_c$. Such a coordinate system can be seen in Figure 22.

![Figure 22: One coordinate system for the hexagonal grid](image-url)
The mapping $\phi : H \rightarrow \mathbb{Z}^2$ defined as $h_0 + pd_b + qd_c \mapsto (p, q)$ is one to one and onto, and allows us to induce a neighbor relation on $\mathbb{Z}^2$. In this neighbor relationship, clearly $(0, 1)$ and $(1, 0)$ are neighbors of $(0, 0)$, but to be consistent with $X_6$ as we defined it, $(1, 0)$ must also be a neighbor of $(0, 1)$. Then with this definition of neighborhood relation consistent with that of $H$, we can construct a distance function $d_6$ on $\mathbb{Z}_2$ such that $X = \langle H, d_H \rangle$ is isometric to $\langle \mathbb{Z}_2, d_6 \rangle$.

Note also that in this construction we were free to choose the direction $d_b$ as we wish, and each choice produces a different coordinate system over all hexagons.

Then in particular we can define $\psi_{i,i+1}(A)$ in a fashion consistent to that in previous chapters, with one example in Figure 23.

**Definition 6.1.** Divide $A$ into layers $A_k = \{x \in A | \phi(x) \in \{k\} \times \mathbb{Z}\}$ so that all elements of a layer have the same coordinate in the direction $i$, and vary in the $i+1$ coordinate. Let $k_{\text{max}} = \max_{A_k \neq \emptyset} k$ so that $A_{k_{\text{max}}}$ is the layer with maximal $k$ that is not empty. Then $\psi_{i,i+1}$ is the element of $A_{k_{\text{max}}}$ with maximal $i+1$ coordinate.

![Figure 23](image)

Figure 23: $A$ is the gray shape, then $\psi_{i,i+1}(A)$ for $i = 3$ is the hexagon with coordinates $(3, 0)$

In order to justify using the separation corollary, we will prove Lemma 4.4 for $X_6$ as well. Figure 24 illustrates the proof.

**Proof.** Let $A, B \subset H$ be two non-empty shapes, we need to show a translation $T$ such that $A$ and $T(B)$ are disjoint, and have at least two shared neighbors. Using the same coordinates we have used in the example, we note that $a = \psi_{3,4}(A)$ is the topmost element on the rightmost column of $A$. We recall that in $X_6$ the direction index opposite to $i$ is $i+3$ modulo 6, for example $3+3 = 6 = 0$. Then from symmetry it is clear that $b = \psi_{0,1}(B)$ is the bottom element on the leftmost column of $B$. Then we can choose $T$ so that $T(b) = a + d_2 + d_3$, or in words, after the translation, the bottom element of the left most column of $B$ is two columns to the right and at the same height as the top element in the rightmost column of $A$. Then clearly they are disjoint, and it is easy to see that
Figure 24: Two shapes placed so they share 2 neighbors.

$a + d_2$ and $a + d_3$ are neighbors of $T(b)$ and of $a$, and therefore joint neighbors of $T(B)$ and of $A$.

**Definition 6.2.** A subset of $H$ of form

$$\{ h \mid h = \phi_{i,i+1}(p,q) \land \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \in \{a,\ldots,d\} \times \{b,\ldots,e\} \times \{c,\ldots,f\} \}$$

where $a, b, c, d, e, f \in \mathbb{Z}$ and $a \leq d \land b \leq e \land c \leq f$, is called a simple shape.

As an example, the leftmost shape in Figure 24 (which we called $A$) is a simple shape, with $a = -3; d = -2; b = -2; e = 0; c = -3; f = -1$.

**Remark 6.3.** Note that simple shapes, while defined above in terms of one particular mapping to $(\mathbb{Z}^2, d_6)$, in fact specify the bounds in terms of the directions in $\mathbb{Z}^2$ that correspond real directions in $(H, d_H)$. Thus the definition is in fact a geometric one in $X_6$, and tells us that simple shapes are the intersections of 6 half planes, one along each of $\{d_j\}_{j=0}^5$, and is thus independent of the direction $i$ that appears in the formal definition.

### 6.2 Optimal shapes in $X_6$ are simple

**Lemma 6.4.** All optimal subsets of $H$ are simple shapes.

**Proof.** Let $A$ be any finite subset of $H$, and let $B$ be the minimal simple shape such that $A \subset B$. Clearly $|A| \leq |B|$, then it is sufficient to show that $N(B) \leq N(A)$. This is because then if $|A| < |B|$, then $A$ is not optimal, and if $|A| = |B|$, because $A \subset B$, we conclude $A = B$, and then $A$ is simple. First we dispatch a special case in order to avoid it in the general proof. For $|B| = 1$, it is clear that $|N(A)| \in \{0,6\} \leq 6 = |N(B)|$, so we can now prove this for $|B| > 1$.

In order to show that $N(B) \leq N(A)$, we will construct a mapping $m : N(B) \rightarrow N(A)$, and show it is one to one.

31
Let \( d_i \) be one of the six directions, then we consider the coordinate system associated with \( h_0, h_i, h_{i+1} \). Then for every \( a \in A \), \( \phi_{i,i+1}(a) = (p,q) \) where \( a = h_0 + pd_i + gd_{i+1} \). Let \( a_i = \psi_{i,i+1}(A) \), then as we can see for an example in Figure 25 by the choice of \( a_i \), \( n_i = a_i + d_i \in N(A) \cap N(B) \), and \( a_i + kd_i \notin B \cup N(B) \) for any \( k \geq 2 \).

![Figure 25: The shape A is given in dark gray, the hexagons not in A but in the minimal simple shape B containing A are lighter colored, and N(B) are lightest. Here each arrow in direction \( d_i \) starts at \( n_i \), to find \( a_i \) look in the reverse direction. Note that \( a_i = a_j \) is possible, but \( n_i = n_j \) is not.](image)

This way we define \( n_i \) for \( i = 0...5 \). We associate with \( n_i \) the direction \( d_i \), and write \( d(n_i) = d_i \). Note that \( N(B) \) is a simple cycle in \( H \), and \( \{n_i\} \) partition \( N(B) \setminus \{n_i\} \) into at most 6 components (less if two or more of \( \{n_i\} \) are adjacent to one another). We denote \( N_{i,i+1} \) to be the component delimited between \( n_i, n_{i+1} \). If \( n \in N_{i,i+1} \) then we associate with it the direction \( d_i \), and write \( d(n) = d_i \). Now \( d \) maps to each element \( n \in N(B) \) a direction \( d(n) \) (see Figure 26). Note that for the general \( n \) also, it is true that \( n + kd(n) \notin B \cup N(B) \) for any \( k \geq 1 \).

Let \( n \in N(B) \), and let \( d_j = d(n) \). Assume by contradiction \( n - pd_j \notin N(A) \) for every non-negative integer \( p \). This is clearly false for any of our special \( n_i \), so we assume that \( n \in N_{j,j+1} \). Now consider the cut \( L_{n,j,j+1} \), which separates \( X_6 \) into two connected components, we’ll call them \( T, S \) one of which includes \( n_j \) and the other of which includes \( n_{j+1} \). In Figure 25, for \( n = (2,-1) \), the cut would be parallel to the arrow through \((1,-2)\), separating it from \((2,1)\).

If \( L_{n,j,j+3} \) included elements of \( A \), then it would have to include elements of \( N(A) \), and we’re assuming it does not. Then because \( n_j, n_{j+1} \in N(A) \), each of \( T, S \) includes at least one element of \( A \), the neighbor of \( n_j \) or of \( n_{j+1} \) as appropriate. But then by the separation corollary, \( A \) is not optimal. Then our assumption is false, therefore there exists \( p \in N \cup \{0\} \) such that \( n - pd_j \in N(A) \). We denote the smallest such number \( p(n) \). Now we can define our mapping \( m : N(B) \to N(A) \) as \( m(n) = n - p(n) d_j \in N(A) \).

Now it remains to show that \( m \) is one to one.
Let \( s, t \in N(B) \), and let \( m(s) = m(t) \), then we need to prove that \( s = t \). If \( d(t) = d(s) \), then \( s = t \), because we’ve noted that \( n + k \cdot d(n) \notin B \cup N(B) \) for any \( n \) and \( k \geq 1 \). Now we assume that \( d(t) \neq d(s) \), and wlog that \( d(t) = d_i; d(s) = d_j \) and \( j < i \). If \( t, s \) are neighbors and \( m(t) = m(s) \), then by the assignment of directions, \( t \) is one of the special \( n_i \). But then \( s - d_i \in A \) and is also a neighbor of \( t \), then \( t \in N(A) \Rightarrow t = m(t) \Rightarrow t = s \). If \( t, s \) are not neighbors, then \( L_{m(t) = m(s), d_i, d_j} \) allows a separation argument, with at least one element of \( N(B) \) that separated \( t \) from \( s \) in each direction ending up in each component of \( X \). Then \( A \) is not optimal, which is a contradiction.

6.3 The optimal simple shapes

Having proved that the optimal shapes are all simple, now we turn to finding which of the simple shapes actually are optimal. We will use two tactics to make this easier. The first is to make the most of the characterization of simple shapes in terms of \((\mathbb{Z}^2, d_6)\). The second is to assume as much as we can without loss of generality in order to reduce the number of cases we need to deal with.

We recall that a simple shape has the form

\[
\begin{cases}
    h \in H \mid h = \phi_{i+1} (p, q) \land \left( \begin{array}{ll}
        1 & 0 \\
        0 & 1 \\
        1 & 1
    \end{array} \right) \in \{a, \ldots, d\} \times \{b, \ldots, e\} \times \{c, \ldots, f\}
\end{cases}
\]

where \( a, b, c, d, e, f \in \mathbb{Z} \) and \( a \leq d \land b \leq e \land c \leq f \). Because \( \phi_{i+1} \) is an isomorphism, we may instead consider the shape

\[
\begin{cases}
    (p, q) \in \mathbb{Z}^2 \mid \left( \begin{array}{ll}
        1 & 0 \\
        0 & 1 \\
        1 & 1
    \end{array} \right) \left( \begin{array}{l}
        p \\
        q
    \end{array} \right) \in \{a, \ldots, d\} \times \{b, \ldots, e\} \times \{c, \ldots, f\}
\end{cases}
\]

Then \( a, d, b, e \) describe bounds parallel to the axes, while \( c, f \) describe upper and lower bounds on \( p + q \). Because translations do not affect optimality, we may
assume wlog that \( a, b = 0 \). Then a general simple shape can be described by just four numbers in \( \mathbb{N} \cup \{0\} \), which are \( d, e, c, f \). Geometrically, it is a rectangle of size \( d \times e \) except that its bottom left and top right corners might be cropped. The effect of the \( c, f \) bounds is to remove an isosceles and right triangle from those two corners. Then in the description of the shape, we may specify instead the number of diagonal rows that have been removed from each corner, which we denote \( j \) and \( k \). Then we say that this shape has dimensions \( d, f \) and ear sizes \( j, k \). Figure 27 provides an example of this notation.

![Figure 27](image)

Figure 27: This is a simple shape in \( X_6 \) that can be described as having dimensions \( d = 6, e = 5 \) and ear sizes \( j = 2; k = 3 \).

**Lemma 6.5.** A simple shape \( A \) in \( X_6 \) of dimensions \( a, b \) and ear sizes \( j, k \) has area \( |A| = a \cdot b - \frac{j(j+1)}{2} - \frac{k(k+1)}{2} \). It has perimeter \( |N(A)| = 2(a + b + 1) - j - k \).

**Proof.** The formula for area is obvious.

We recall that this shape is in \( (\mathbb{Z}^2, d_4) \), in which two tiles are neighbors if they either neighbors according to \( d_4 \), or on a diagonal with constant sum of coordinates. Then a rectangle of dimensions has \( 2(a + b + 1) \) neighbors. Now we apply an inductive argument in \( j \) (or \( k \)). From such a triangle, removing one tile from the top right (or bottom left) corner removes the two neighbors that corner had, and adds as a neighbor the tile that was removed from the shape, in total removing one neighbor. In general, the \( n \)th diagonal removed includes \( n \) tiles, which have \( n \) neighbors to their right (left) and one additional neighbor above (below) the top (bottom) tile in the diagonal. Therefore we remove \( n + 1 \) neighbors, and add \( n \) that were tiles, and in total remove 1 neighbor, for the \( n \)th diagonal removed.

**Lemma 6.6.** The lengths of the sides of a simple shape with dimensions \( a, b \) and ears \( j, k \) are \( a - j, j + 1, b - j, a - k, k + 1, b - k \), counter clockwise, starting from the top left side.

**Proof.** The inductive argument is similar to that for the previous lemma. If \( j = k = 0 \), then the side sizes are \( a, 1, b, a, 1, b \). Removing a diagonal increases by one that side, and reduces by one each of the adjacent sides.
Lemma 6.7. Let $m, n$ be the lengths of two sides of an optimal simple shape of $X_6$, that are not adjacent. Then $|m - n| < 2$.

Proof. Assume that the two sides are parallel, then there is a direction $i$ such that the two sides are mapped in $\mathbb{Z}^2$ to the ears, then $m = j + 1; n = k + 1$. We assume that $|m - n|$ is odd to complete the proof will conclude the shape is not optimal. It follows that $j - k \geq 2$, assume wlog that $j \geq k + 2$. Now we compare this shape $A$ with dimensions $a, b$ and ears $j, k$ to the simple shape $B$ with dimensions $a, b$ and ears $j - 1, k + 1$. The perimeters are equal: $2(a + b + 1) - j - k = 2(a + b + 1) - (j - 1) - (k + 1)$. The sizes are not:

$$a \cdot b - \frac{(j - 1) \cdot j}{2} - \frac{(k + 1)(k + 2)}{2} - \left(a \cdot b - \frac{j(j + 1)}{2} - \frac{k(k + 1)}{2}\right) = \frac{1}{2}(-j(j - 1) - k + 1) - \frac{1}{2}(2j - 2(k + 1)) = \frac{k + 2 - k - 1}{2} = 1$$

Then $A$ is not optimal, because $B$ has larger area for the same perimeter.

Now assume that the two sides had a single side separating them. Then there exists a direction $i$ such that the side separating them is an ear. Then we assume wlog that $m = a - j, n = b - j$. We assume again that $|m - n|$ is odd, and wlog that $m \geq 2 + n$, then $a - j \geq 2 + b - j \Rightarrow a \geq 2 + b$. A similar comparison argument with the shape with the same ears $j, k$ but side sizes $a - 1; b + 1$ shows the original shape is not optimal.

Corollary 6.8. For any isometry of an optimal simple shape $A$ into $(\mathbb{Z}^2, \mathcal{d}_6)$, $|j - k| < 2$, $|a - b| < 2$, $|j + 1 - (a - k)| < 2$ and $|(b - j) - (k + 1)| < 2$.

Proof. $|j - k| < 2$ because they are opposite. $|j + 1 - (a - k)| < 2$ because the sides are $a - j, j + 1, b - j, a - k, k + 1, b - k$, and there they are separated only by the side of length $b - j$. If $|(b - j) - (k + 1)| < 2$ because they are separated only by $a - k$. $|a - b| < 2$ because $a - j, b - j$ are separated only by $j + 1$.

We have found a set of geometric constraints on optimal shapes, and expressed them in terms of the parameters describing the shape that is induced by a particular isometry of the hexagon plane into $\mathbb{Z}^2$. We will explore the set of parameters that fit these constraints, to find which geometric shapes they represent. In order to avoid duplicated work, we wish to eliminate degrees of freedom that come from having different possible isometries. Thus we will make some symmetry breaking assumptions.

For whatever pair of opposed sides we choose to be the ears, we can choose wlog that $a \leq b$ and $j \leq k$. This allows us to assume that $a - k \leq a - j, b - k \leq b - j$. Of the three pairs of parallel sides, we choose the one containing the
smallest side to be the pair of ears, so that \( j + 1 \leq a - k \), and \( j + 1 \) is the length of the shortest side of the shape.

To better express the symmetry eliminating assumptions, we change some variables by writing \( b = a + d, k = j + \delta, a - k = j + 1 + \Delta \Rightarrow a = 2j + \delta + \Delta + 1 \).

Then the symmetry breaking assumptions are \( d, \delta \geq 0 \) and \( \Delta \geq 0 \), and the geometric results translate into \( d, \delta, \Delta \leq 1 \) and also:

\[
\begin{align*}
2 &> |(b - j) - (k + 1)| = \\
&> |(a + d - j) - (j + \delta + 1)| = \\
&> |a + d - j - \delta - 1| = \\
&> |a + d - 2j - \delta - 1| = \\
&> |(2j + \delta + \Delta + 1) + d - 2j - \delta - 1| = \\
&> |\Delta + d| = \\
&> \Delta + d
\end{align*}
\]

Then because \( \Delta, d \geq 0 \), we conclude that at least one of \( \Delta, d \) is zero.

Using this notation, we see that optimal shapes have the form \((a, b, j, k) = (2j + \delta + \Delta + 1, 2j + \delta + \Delta + 1 + d, j, j + \delta)\) allowing us to describe such a shape using the parameters \( j, \delta, d, \Delta \). Here \( j + 1 \in \mathbb{N} \) is the length of the shortest side, and \( \delta, d, \Delta \) play a role similar to that of the skeleton in the previous section. Due to the symmetry eliminating inequalities, and writing \( p_i = (\delta, d, \Delta) \), only \( p_1 = (0, 0, 0), p_2 = (0, 1, 0), p_3 = (1, 0, 0), p_4 = (0, 0, 1), p_5 = (1, 0, 1), p_6 = (1, 1, 0) \) may be optimal. We note that the shape \( j, p_i \), when expanded once (all its neighbors are added to it), results in the shape \( j + 1, p_i \).

Now we substitute \((a, b, j, k) = (2j + \delta + \Delta + 1, 2j + \delta + \Delta + 1 + d, j, j + \delta)\) into the formulae we already know.

The neighborhood size is:

\[
2(a + b + 1) - j - k = \\
2(2j + \delta + \Delta + 1 + 2j + \delta + \Delta + 1 + d + 1) - j - j - \delta = \\
2(4j + 2\Delta + 4\delta + 2d + 3) - 2j - \delta = \\
6j + 4\Delta + 4\delta + 2d + 6 - \delta = \\
6j + 4\Delta + 3\delta + 2d + 6
\]

The area of such a shape is \( a \cdot b = (2j + \delta + \Delta + 1) \cdot \left(\frac{i(i+1)}{2} + \frac{k(k+1)}{2}\right) \) where \( a \cdot b = (2j + \delta + \Delta + 1) \).
(2j + δ + Δ + 1 + d) and because δ ∈ {0, 1}

\[
\begin{align*}
\frac{j(j + 1)}{2} + \frac{k(k + 1)}{2} &= \\
\frac{j(j + 1)}{2} + \frac{(j + δ)(j + δ + 1)}{2} &= \\
\frac{j^2 + j}{2} + \frac{j^2 + 2jδ + δ^2 + j + δ}{2} &= \\
\frac{j^2 + j + δj + δ}{2} &= \\
\frac{j^2 + j(1 + δ) + δ}{2} &= \\
\end{align*}
\]

**Lemma 6.9.** Let \( A(i, j) \) be the area of the simple shape of form \((a, b, j, k) = (2j + δ + Δ + 1, 2j + δ + Δ + 1 + d, j, j + δ)\), where \( j + 1 \) is the shortest side and the other parameters \((d, δ, Δ)\) equal to \( p_i \) which is the \( i \)th element of the tuple \((\{0,0,0\}, \{0,1,0\}, \{0,0,1\}, \{1,0,0\}, \{0,1,1\}, \{1,1,0\})\). Then for every \( j \in \mathbb{N} \cup \{0\}\): \( A(1, j) \leq A(2, j) \leq A(3, j) \leq A(4, j) \leq A(5, j) \leq A(1, j + 1) \leq A(6, j) \leq A(2, j + 1) \)

**Proof.** This requires nothing but tedious substitution. As we saw above, the area is \( a \cdot b - \left( \frac{j(j + 1)}{2} + \frac{k(k + 1)}{2} \right) \). In \( A(2, j + 1) \),

\[
a \cdot b = \\
(2(j + 1) + δ + Δ + 1) \cdot (2(j + 1) + δ + Δ + 1 + d) = \\
(2(j + 1) + 1 + 1 + 1) \cdot (2(j + 1) + 1 + 1 + 1 + 0) = \\
(2j + 4)^2
\]

Then

\[
A(2, j + 1) = \\
(2j + 4)^2 - (j + 1)^2 - (j + 1)(1 + 0) - 0 = \\
(2j + 4)^2 - (j + 1)^2 - (j + 1) = \\
4j^2 + 16j + 16 - j^2 - 2j - 1 - j - 1 = 3j^2 + 13j + 14
\]

Similarly we find:

\[
A(6, j) = 3j^2 + 10j + 8 \\
A(1, j + 1) = 3j^2 + 9j + 7 \\
A(5, j) = 3j^2 + 8j + 5 \\
A(4, j) = 3j^2 + 7j + 4 \\
A(3, j) = 3j^2 + 6j + 3 \\
A(2, j) = 3j^2 + 5j + 2 \\
A(1, j) = 3j^2 + 3j + 1
\]

And indeed these fulfill the order required. \( \Box \)
Lemma 6.10. For shapes having particular $i, j$ parameters as in the previous lemma, we define $N(i, j)$ to be the number of its neighbors. Then for every $j \in \mathbb{N} \cup \{0\}$, $N(1, j) \leq N(2, j) \leq N(3, j) \leq N(4, j) \leq N(5, j) \leq N(1, j + 1) \leq N(6, j) \leq N(2, j + 1)$.

Proof. Again this is a matter of simple substitution. As we saw above, the area of a shape with parameters $p_i$ and shortest side length $j + 1$ is $N(i, j) = 6j + 4\Delta + 3\delta + 2d + 6$.

\[
\begin{align*}
N(1, j) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 + 6 = 6j + 6 \\
N(2, j) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 1 + 6 = 6j + 8 \\
N(3, j) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 1 + 2 \cdot 0 + 6 = 6j + 9 \\
N(4, j) &= 6 \cdot j + 4 \cdot 1 + 3 \cdot 0 + 2 \cdot 0 + 6 = 6j + 10 \\
N(5, j) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 1 + 2 \cdot 1 + 6 = 6j + 11 \\
N(1, j + 1) &= 6 \cdot (j + 1) + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 + 6 = 6j + 12 \\
N(6, j) &= 6 \cdot j + 4 \cdot 1 + 3 \cdot 1 + 2 \cdot 1 + 6 = 6j + 13 \\
N(2, j + 1) &= 6 \cdot (j + 1) + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 1 + 6 = 6j + 14
\end{align*}
\]

And these fulfill the required order. □

Lemma 6.11. Let $W$ be the set of simple shapes with smallest side size $j + 1$ for $j \in \mathbb{N} \cup \{0\}$ and other parameters equal to $p_i$, where $i \in \{1, \ldots, 6\}$. Then every shape in $W$ is an optimal shape in $X_6$.

Proof. We have shown that all the optimal shapes of $X_6$ are in $W$. Then if any shape in $W$ is not Pareto optimal, so that it can be improved, then in particular it can be improved by some other optimal set, therefore by one of the shapes in $W$. This relies on the general argument against the existance of infinite chains of improvement, made in Lemma 4.19.

We have shown a linear ordering of the elements of $W$, such that every element has strictly larger area than its predecessors, and strictly larger perimeter. Then none of the shapes in $W$ is improved upon in one sense without compromise in the other, by any other shape in $W$. Combined with the argument of the previous paragraph, we find every shape in $W$ to be Pareto optimal. □

Note that this lemma completes the proof of Theorem 3.4, and then Theorem 3.7 results directly from applying the calculations in the proof of Lemma 6.9.

7 Context and Motivations

This work was originally motivated by a specific application of the isoperimetric results, but also fits into a long standing mathematical exploration of digital spaces. We will briefly present the motivation for such problems next.

Diverse phenomena spread in space, like fire in a forest, a contamination or a contagious disease. The dynamic cleaning problem [1] is an abstraction
of the challenge to contain and mitigate such phenomena on a simple discrete domain, the regular grid. In this model, a connected subset of the grid (the contaminated area) expands to incorporate all its neighbors every few time steps. The cleaning is carried out by a set of agents which move along the edge of the contaminated area, removing contamination on their way. The number of agents and their assumed capabilities for movement, computation, and coordination vary, and their efficiency in mitigating the contamination is explored both analytically and by simulation. An upper bound on this efficiency can be derived by abstracting away the positions of the contaminated tiles and the cleaning robots, and using only their numbers to estimate their progress. For the robots we assume each robot cleans one tile at each time step - this is the ideal performance, which realistically could be hindered for example by overcrowding of robots in a small area. The complementary question about the contamination is how slowly can it possibly spread? This is equivalent to asking, what is the smallest number of neighbors that a connected subset of the grid of cardinality $n$ can have? This is clearly a discrete isoperimetric problem.

A second possible area of application is in domain partitioning for parallel computation. Given a computation defined over a planar square grid, such that the result at each cell requires information at each adjacent cell (for example, in the sense of sharing an edge), how do we partition the cells among a set of computing nodes, so as to minimize communications between nodes? We assume that the number of cells is very large (so we ignore the effect of the perimeter of the domain). Then partitioning the domain using shapes with maximal area and minimal perimeter is desirable, if it is possible. It is clear by inspection that for many of the optimal shapes we consider, a one shape tiling of the plane is possible.

An analysis somewhat similar to ours was applied to understanding the class of “polyomino weak full set achievement games” by Sieben [10]. Here one player wishes to expand a connected set of controlled tiles by taking over one free neighbor at each of his turns, while his adversary tries to bound his progress by capturing $q$ arbitrary uncontrolled tiles at each of his turns. One way to prove that the first player can win is to show that it is impossible for the second player to ever capture as many tiles as the controlled set has neighbors. If we abstract over the order of moves, and assume a worst case strategy for player 1, this too becomes an isoperimetric question over connected shapes. Note that this application, like the previous one, can also be transferred to the hexagonal grid by using the results and techniques presented in the current work.

Isoperimetric results and achievement games [5] are aspects of a general study of subsets of discrete planar grids and various specializations. We have mentioned before the study of “animals”, a name given to simply connected subsets, which often focuses on various integer parameters of “animals”, like their areas and various perimeter measures. The study of ranges achieved for those parameters thus naturally includes the isoperimetric question, among others. The hexagonal case arises in the study of planar molecules with formulas $C_nH_s$, which consist of hexagonal rings of carbon, whose otherwise free bonds are taken by the hydrogen atoms. The structure of Naphtalene, which is a sim-
A simple example of this structure, may be seen in Figure 28. In addition to the two

![Naphtalene molecule](image)

Figure 28: Naphtalene is a simple Polycyclic Aromatic Hydrocarbon (multiple carbon hexagons terminated by hydrogen atoms). Larger PAH molecules (and in the limit, graphene sheets) correspond to connected shapes in the hexagon tiling (when omitting the hydrogens from the diagram).

integer numbers $n$ and $s$ of carbon and hydrogen atoms respectively, we may consider also the number of hexagons and the length of the perimeter, or the number of hexagons and the number of internal vertices (carbon atoms that link only to other carbons). The so-called circular animals [4] are defined as those maximizing the number of hexagons relative to the perimeter length or to the number of internal vertices. The circular shapes are also exactly the shapes optimal in the sense considered by this work, despite the non-equivalence of the measures of perimeter, which is demonstrated in Figure 29.

![Circular shapes](image)

Figure 29: Left shape has 13 neighbors, right shape has 14 neighbors, but both have 22 edges and 8 internal vertices. Then the neighbor based perimeter is not equivalent to perimeter as defined by edges or internal vertices.

A natural question is generalization - can such results be found for different spaces, perhaps in higher than two dimensions? we have considered our discrete spaces as graphs at times, but relied heavily on extra geometric properties, such as the existence of directions. The isoperimetric inequality and spiral construction questions have been explored for more general types of graphs, such as graph products, as reviewed by Bezrukov [2]. In the language of graphs, the edge isoperimetric problem (EIP) and the vertex isoperimetric problem (VIP) correspond to counting the sides a shape shared with its complement, and to counting the shapes neighbors respectively. We show by a simple argument that
results exactly analogous to our own do not exist even for graphs corresponding to $\mathbb{Z}^3$.

Balls are known to be optimal (they appear in the spiral constructions) and the boundaries of balls grow quadratically with the radius. Then the number of possible perimeters between those of successive balls is proportional to the radius (as opposed to being constant as in the planar case we considered). Is it possible then, that all the optimal shapes are generated only by iterated expansions of a finite set of simple shapes? if expansions of optimal sets preserve their ordering by area and perimeter, then we would have a constant number of optimal shapes between successive balls. But the eventually the largest gap between perimeters of two successive optimal shapes ($A$ and $B$) must be larger than 6, allowing the introduction of a shape between them by adding an isolated point to $A$. Because this shape is dominated by neither $A$ or $B$, there must be an optimal shape between $A$ and $B$ that does dominate it, in contradiction to $A$ and $B$ being successive optimal shapes. Then any characterizations of isoperimetric shapes for higher dimensional spaces must be made in different terms.

8 Concluding Remarks

This paper gives a simple geometric characterization, for three common digital spaces ($\mathbb{Z}^2$ with 4- and 8-neighborhoods, and the hexagonal tiling) the planar polyforms (shapes) that are optimal in having both the smallest perimeter for their area and the largest area for their perimeter. As a byproduct, the results generated interesting integer sequences, one of which was previously known only for a surprisingly different reason. The lack of regularity in some of the results also illustrates the difficulties in dealing with discretizations of natural geometric concepts like the basic isoperimetric inequality. We note that many variations in the definition of perimeter are possible and will almost certainly lead to different and possibly very interesting results.

While previous explorations of discretized versions of the discrete isoperimetric problems concentrated on inequalities and on constructing some shapes that are boundary optimal, we focus on geometric properties common to all Pareto optimal shapes. For example, instead of assuming connectedness properties of shapes, we prove the optimal shapes are such (and identify the unique exception). While complete characterizations of the simple form we present here are not possible for higher dimensions, the approach of exploring geometric commonalities might still be fruitful. Are isoperimetric discrete shapes in higher dimension also connected (except perhaps for a finite number of cases)? Can they also be described as intersections of half spaces? if so, how many half spaces are required? It seems to us there is more geometric richness to isoperimetric shapes than spiral constructions illustrate.
References


משהו של מספותו את הוראות האפסטיפולייםCLUSIONES לתרבויות רישומיות ומדעים אקדמיים.

ובטוחות זו טענה שלושה מתרבויות המודרניות, ע"יリンクות של המימוש לזרועות וא
משויקות חיזב, את הוראות ההשאלה всемי תיירות. מסיבות של הזזות בין הזרועות הממדיות
تسمות, גם תכנית את השאלות עם או בין הזרועות איננה קיימת. בהellung של מנהיגות
בוצק ריצי צהות שמעימה שיכונים ע缣ום." המנהיגים של מנהיגות פעילות המשמשים
כניזי הוראה של צוות זהה או פשוט, ובכל מקום ששל ViewData, מחברים את הזרועות המודרניטים
המילים, קבעו את הוראות האפסטיפוליים ממשקים, כ"י קבעו סופי הקטנה של הזרועה
החלונות הזרועות ביחס לאורח פַמְמָי

איור 2: זירות הזרועות עבורה שהותディים עד משות

איור 3: זירות הזרועות עבורה שהותדיים פיים משות

איור 4: זירות הזרועות עבורה ריצי חיות משותי

נושה של יל מזרחי הזדמנת למלג מטרונים ריפויים שלובים בשל מificacion
החברות כלכ"ל שרטטו מספים עם הזרועות, בחרו את מתכונת הזרועות פיות
מסתובב, גם רוקדים של פרפר מבדוח אפסטיפוליים שלישית קבוצת מתכונת יを与え
 ויותר פ שויה, לקבוצת מתכונת אייםを与え, וכמה בוחנים מונגים מחכים: אם מזיכים
במעבדה או עם מי מזרחי נוספים, ע"י הלקט ה 3
לציבור האורח, עד זיור לחתופי או
ברוחב 요청 של צוות הזרועת

יש לייצר שלמה- acompaña ריבים, סמיכויות נבוכו, מרכזים שונים של העבורה
של צוות בכל גודל אפריז. מלכתחילה אפסטים של עבורה אפסטיפוליים בצולם, אל מתכונת

ii
A \leq \frac{L_2}{\pi}:

\begin{align*}
A & \leq \frac{L_2}{\pi} \\
& = L_2 - \frac{1}{\pi}.
\end{align*}

\begin{align*}
A & \leq \frac{L_2}{\pi} \\
& = \frac{L_2}{\pi} - 1.
\end{align*}

\begin{align*}
A & \leq \frac{L_2}{\pi} \\
& = \frac{L_2}{\pi} - 1.
\end{align*}
החברה מעשה ברוחו יד פורפיו אלפרד בורקנשטיין מעשה יד."}

אף מודע لكل זה והقصير ועל המיים הכספים וה_APB הסים והשיטפון.
כדורים דיגיטליים בשני ממדים

היבר על מחקרה

לשם مليי חלקי שלHDR לשחרור לקבלי התוכן
מגיסטר למדעי מידע מחשב

דניאל ווינשנקר

הונג לסנט הטכניו - מכון טכנולוגי לישראל
השון והש"ע חיפה נובמבר 2009
כדורים דו-גליליים בשני ממדים
dינהל ווינשנקר