Isoperimetrically Pareto-optimal Shapes on the Hexagonal Grid

Daniel Vainsencher and Alfred M. Bruckstein
{danielv,freddy}@cs.technion.ac.il
Computer Science Department
The Technion
32000 Haifa, Israel

January 20, 2008

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 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. In this paper we characterize all shapes that have both shortest boundaries and maximal areas for one particular boundary measure on the hexagon tiling. We show this set of Pareto optimal shapes is the same as that induced by a different boundary measure that was studied in the context of theoretical chemistry.

1 Introduction

A particularly nice tiling of the plane $\mathbb{R}^2$, the one favored by bees, is the tiling by regular hexagons. In the tiling of the plane by hexagons every finite set of tiles has a finite number of neighbors. What are the sets of hexagons that are maximal in the sense that adding another hexagon to them must increase the number of neighbors? Ideally, we would like the answer to this question to be a simple and geometric one, as occurs in the following continuous and more familiar variation. The shape that encloses a maximal area in $\mathbb{R}^2$ with a closed curve of any given length is a circular one. This family of shapes is also the set of minimal length closed curves enclosing a given area. This set of facts is often quantified by the isoperimetric inequality:

Theorem 1.1. *Let γ 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 $A \leq \frac{L^2}{4\pi}$.*

Life on the hexagonal grid is quite different from that in $\mathbb{R}^2$. While the length of the boundary curve is "the one" obvious measure for boundary size of a shape in the $\mathbb{R}^2$ plane, in the grid context the number of neighbors is just as natural. We will discuss some connections between the two possible boundary measures in Section 3.

There is another way in which life on the grid is different. We said above that in the plane the set of circles is the answer to two distinct questions, depending on whether the boundary size or the area is held constant. In $\mathbb{R}^2$ an argument by scaling can be used to show the equivalence of these questions in the plane. For sets of hexagons, they are not. One way to explore the optimal tradeoffs between the area and boundary size measures is to consider the Pareto optimal frontier, which is the set of objects such that increasing the value in one geometric measure requires strictly decreasing the value of another. For details about Pareto optimality and applications to game theory, see [4]). In our case, the isoperimetrically Pareto optimal shapes are those sets of hexagons that have both maximal area for their given boundary size, and minimal boundary size for their given area, and our goal is to give a characterization of all such shapes. The name frontier comes from the correspondence of such sets to boundaries between the feasible pairs of values of the geometric measures, and those that are not achievable, in diagrams exemplified by Figure 1.

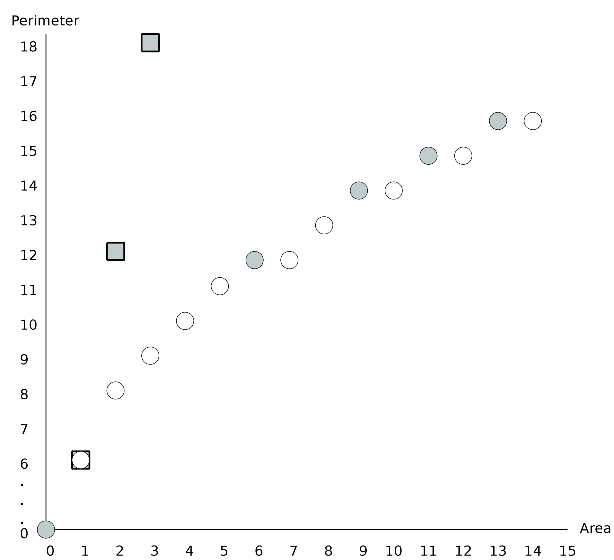

Figure 1: Some feasible measure pairs for subsets of the hexagon tiling. The squares have maximal perimeters for a given area, the circles have minimal perimeters, the white circles also have maximal area for their perimeter, and thus make up the Pareto optimal frontier. Every shape with a given area must have its perimeter between the circle and the square on that column, or on one of them.

Our main result is a geometric characterization of the isoperimetrically Pareto optimal shapes, using the concept of an expansion of a shape, which is defined as its union with its neighbors.

Theorem 1.2. *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 2, or it is generated by iterated expansions of such a shape.*

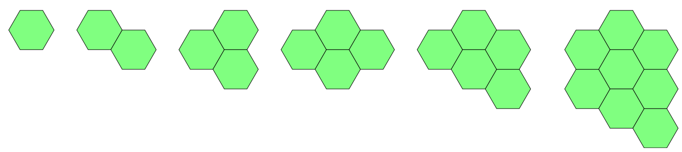

Figure 2: Shapes generating all optimal shapes for hexagons.

As Figure 1 hints, and can easily be seen from our explorations in Section2, for every neighborhood size larger than eight, there exists a Pareto optimal shape with that boundary. This provides a geometric answer to the question in the first paragraph: the shapes of maximal area are the Pareto optimal shapes. Another consequence of the fact that the perimeters of Pareto optimal shapes are consequtive is that to describe the set of optimal feasible measures we need only to provide the sequence of areas for the Pareto optimal frontier. That is done in the following result.

Theorem 1.3. *For the hexagonal tiling, the sequence of areas of non-empty optimal shapes in order of increasing boundary size is given by one occurence of $A(1,0)$ followed by $A(2,j) \leq A(3,j) \leq A(4,j) \leq A(5,j) \leq A(1,j+1) \leq A(6,j) \leq A(2,j+1) \leq \dots$ for $j \in \mathbb{N} \cup \{0\}$, where $A(i,j)$ is defined as follows:*
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 identified in Sloane's wonderful Encyclopedia of Integer Sequences [5] as sequence A001399. One of the characterizations given for it there is the following: the numbers of the tiles along a spiral on the hexagonal grid at which "folds" occur. Note that this spiral construction is well known, and we will say more about it when considering related work in Section 3.

2 Characterization of Pareto optimal shapes

2.1 Definitions and Basic Facts

We consider the tiling of $\mathbb{R}^2$ by hexagons, denoting the set of hexagons H. We will say that two hexagons are adjacent or neighbors if they share a side. This neighborhood relation can be represented as a graph in which hexagons are vertices, and an edge exists between two hexagons if they are adjacent.

Definition 2.1. A shape A in H is a finite subset of H. The neighborhood or boundary $N(A)$ of A consists of the vertices adjacent to it. Then $N(A) = \{h \in H | d(h, A) = 1\}$.

Definition 2.2. We call a shape A area optimal if it has the maximal area for its boundary size, or formally, if for any B such that $|N(A)| \geq |N(B)|$, it is also true that $|A| \geq |B|$.

Definition 2.3. We call a shape A boundary optimal if it has a minimal boundary area for its area, or formally, if for any B such that $|A| \leq |B|$, it is also true that $|N(A)| \leq |N(B)|$.

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

Because our exploration is a geometric one, we do not wish to distinguish between shapes differing only in location or orientation. In particular, we wish to ensure that such transformations do not affect the optimality of shapes, so we are free to perform them at our convenience. In fact, any translation, rotation by $\frac{\pi}{3}$ or reflection preserves neighborhood, and thus does not affect the optimality of a shape. The proof of the following lemma will use this fact about translations.

Our first lemma will show that shapes constructed in a certain way cannot be optimal. In the form given by its corollary, it will be easy to apply to quite a general class of shapes. But first we consider a simple construction. Let us take one hexagon h_0 and two distinct half-lines from it, and denote the infinite set of hexagons in it L. Then $H \backslash L$ has two connected components, call them A and B, so that $H = L \dot{\cup} A \dot{\cup} B$ where $\dot{\cup}$ denotes disjoint union, and we say that L separates A and B.

Lemma 2.4. *Let $S = C \cup D$, where $C \subset A$ and $D \subset B$ are non empty, and $|N(S) \cap L| \leq 1$. Then S is not optimal.*

Proof. C and D have at most one neighbor in common. To see this, note that the separation of A, B by L means that $d(A, B) = 2$, and in particular $d(C, B) = 2$. Then by the definition of neighborhoods, $N(C) \subset L \cup A$. Similarly, $N(D) \subset L \cup B$. Then $N(D) \cap N(C) \subset L$, and combining this with the fact that $|N(S) \cap L| \leq 1$, we conclude that $|N(D) \cap N(C)| \leq 1$.

But there exists a different way to combine C and D so that they have more common neighbors, without reducing their total area. Let c be the top hexagon in the rightmost column of C. Let d be the bottom hexagon in the leftmost column of D. Now we define $S' = C \cup T(D)$, where T is a translation such that $T(d)$ is exactly two columns to the right of c. Then it is easy to see that c and $T(d)$ now share exactly two neighbors in a column free of elements of S', and that C and $T(D)$ thus also share exactly two neighbors. This transformation is illustrated in Figure 3.

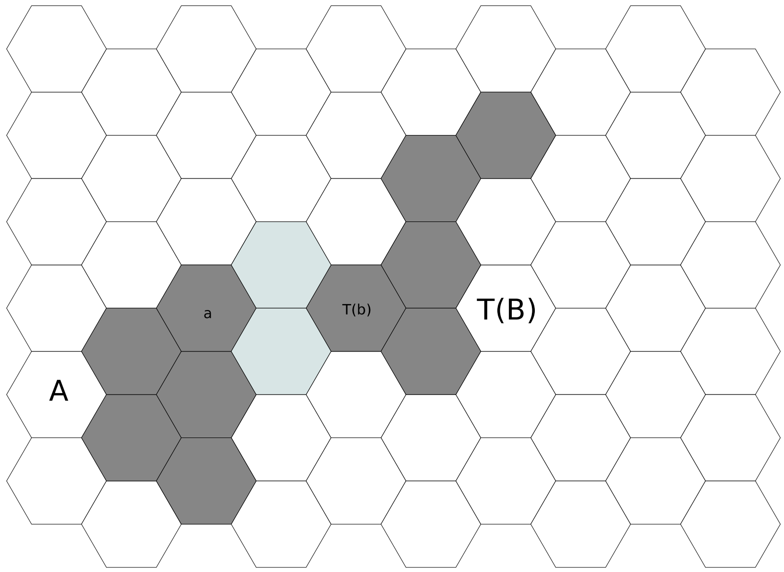

Figure 3: Two shapes A and B placed so they share 2 neighbors of the hexagons a and b.

Note that $C \cap N(D) = D \cap N(C) = \emptyset$ because they are separated by L, and the same applies to C and $T(D)$ which are separated by a column. Now simple set arithmetic shows that $N(S') < N(S)$. Similarly, $C \cap T(D) = \emptyset$ implies that $|S'| = |S|$, then S is not boundary optimal. $\square$

The next corollary makes this fact more easily applicable.

Corollary 2.5. *If L contains at most one neighbor of S, and each of A and B contains a neighbor of S, then S is not optimal.*

Proof. Initially, we will show that L does not overlap S, because if it did, then S would also have at least two neighbors in L. We note that any element in L has at least two neighbors in L. If $L \cap S = \{s\}$, this completes the argument. If S has at least two elements in L, then it divides L into two or three connected components, each of which must contribute one neighbor of S.

Let n_1 be the promised element from $N(S) \cap A$, and let $C = S \cap A$. Then to use the lemma, we need only to show that C is not empty, and apply symmetry to find a non empty subset of S in B. But we are given that there exists $s_1 \in S$ such that $d(n_1, s_1) = 1$, then $s_1 \notin B \Rightarrow s_1 \in L \cup A$ but as we have shown, S does not overlap L, then $s_1 \in A \Rightarrow s_1 \in C$. $\square$

To continue the analysis of shapes in H, we introduce a family of coordinate mappings on it. Let $a \in H$, and let b be a neighbor of h. Now we take c to be the neighbor of a and of b, so that the sequence a, b, c, a advances counter clockwise.

Then every hexagon $t \in H$ can be reached by starting at a and moving p steps in the direction of b and q steps in the direction of c (still relative to a). Thus every choice of a, b induces a mapping $H \to \mathbb{Z}^2$. In fact, what we are choosing is a starting point h (previously called a), and the direction in which to find b. Proceeding counterclockwise, we denote the six different directions d_i where $i \in \{0, 1, \ldots, 5\}$ are called the direction indexes, and naturally are treated cyclically, so that $5 + 1 = 0$ where direction indexes are concerned. We define $\phi_{h,i} : H \to \mathbb{Z}^2$ to be the mapping above defined by h and $b = h + d_i$. Continuing with this notation, $\phi_{h,i}(t) = (p, q)$ where $t = h + p \cdot d_i + q \cdot d_{i+1}$.

Such a coordinate system can be seen in Figure 4.

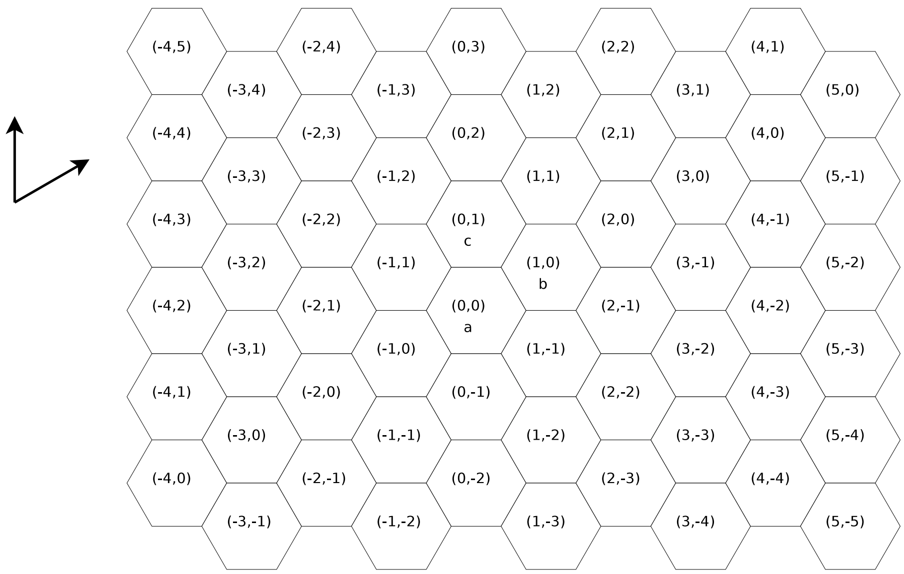

Figure 4: One coordinate system for the hexagonal grid

We will use such coordinate mappings ϕ in two ways. In the first, we note $\phi_{h,i}$ is a bijection onto $\mathbb{Z}^2$ and thus has an inverse function $\phi_{h,i}^{-1}$, allowing us to induce a neighborhood relation on $\mathbb{Z}^2$. Let $p, q \in \mathbb{Z}^2$, then we define them to be neighbors iff the hexagons $\phi_{h,i}^{-1}(p)$ and $\phi_{h,i}^{-1}(q)$ are neighbors. The neighbors of each point in Z^2 are the closest points along either axis, and an additional two along one diagonal. We will restrict our analysis to shapes whose images under coordinates mappings have a nice representation in $\mathbb{Z}^2$, after proving that all optimal shapes belong to this class. The second use of the ϕ coordinate mappings is in defining a concept useful to this proof.

Definition 2.6. Assume h, i are as above. Partition the shape A into layers $A_k = \{x \in A | \phi_{h,i}(x) \in \{k\} \times \mathbb{Z}\}$ so that all elements of a layer have the same coordinate in the direction i, and vary in the $i + 1$th coordinate. Let $k_{\max} = \max\{k | A_k \neq \emptyset\}$ so that $A_{k_{\max}}$ is the layer with maximal k that is not empty.

Then we define $\psi_i : \mathcal{P}(H) \rightarrow H$ so that $\psi_i(A)$ is the element of $A_{k_{\max}}$ with maximal value in the $i+1$th coordinate.

Thus for every direction i, $\psi_i(A)$ allows us to choose a unique element of A that is extremal in two directions. For example, in the proof of lemma 2.4, we could define $c = \psi_i(C)$ and $d = \psi_{i+3}(D)$ and obtain a transformation T' that is different from T, but allowing the same kind of proof, for any direction index i.

Definition 2.7. A subset of H of form

$$\left\{ \phi_{h,i}^{-1}(p,q) \,|\, p,q \in \mathbb{Z} \wedge \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \in \{a,\ldots,d\} \times \{b,\ldots,e\} \times \{c,\ldots,f\} \right\}$$

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

As an example, the leftmost shape in Figure 3 (designated A) is a simple shape, which in the coordinate system of Figure 4 has parameters $a = -3; d = -2; b = -2; e = 0; c = -3; f = -1$.

Remark 2.8. Note that simple shapes, while defined above in terms of one particular mapping to $\mathbb{Z}^2$, in fact specify the bounds in terms of those directions in $\mathbb{Z}^2$ that also correspond to ones in the hexagon tiling. Thus a simple shape is the intersections of 6 half planes, one along each of $\{d_j\}_{j=0}^{5}$. Then the definition is in fact a geometric one, and independent of the direction i that appears in it. In particular the set $\{d-a, e-b, f-c\}$ corresponding to any particular shape will be independent of i, since each element in it is the size of the shape in a different direction.

As another use of directions, we introduce a notation for a concept we have already used.

Definition 2.9. Let h_0 be a hexagon, and let d_i, d_j be two distinct directions. Then we denote $L_{h_0,i,j} = h_0 \cup \{h_0 + d_i \cdot k | k \in \mathbb{Z}\} \cup \{h_0 + d_j \cdot k | k \in \mathbb{Z}\}$ the union of h_0 with the half-lines in the given directions.

2.2 Pareto optimal shapes are simple

Lemma 2.10. *All optimal non empty shapes in H are simple shapes.*

Proof. Let A be any finite subset of H, and let B be the minimal simple shape such that $A \subset B$, then it is sufficient to show that $N(B) \leq N(A)$. This is because then in the case that $|A| < |B|$, we know that A is not optimal. Clearly $|A| \leq |B|$, then in the remaining case that $|A| = |B|$, because $A \subset B$, we conclude $A = B$, and then A is simple.

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. For each neighbor of B, we will first assign it a direction. Then we will show that moving from the neighbor along this direction eventually we find a neighbor of A, which will define our mapping.

Lastly, we will show that if two neighbors of B are thus mapped to the same neighbor of A, then A cannot be optimal, using the corollary of separation.

Let d_i be one of the six directions, then we consider the coordinate mapping $\phi_{h,i}$. Let $a_i = \psi_i(A)$, then as we can see for example in Figure 5 that by the choice of a_i, it must result that $n_i = a_i + d_i \in N(A) \cap N(B)$, then near our selected points a_i, in the selected directions, the first neighbor of A along $-d_i$ starting from n_i is n_i itself.

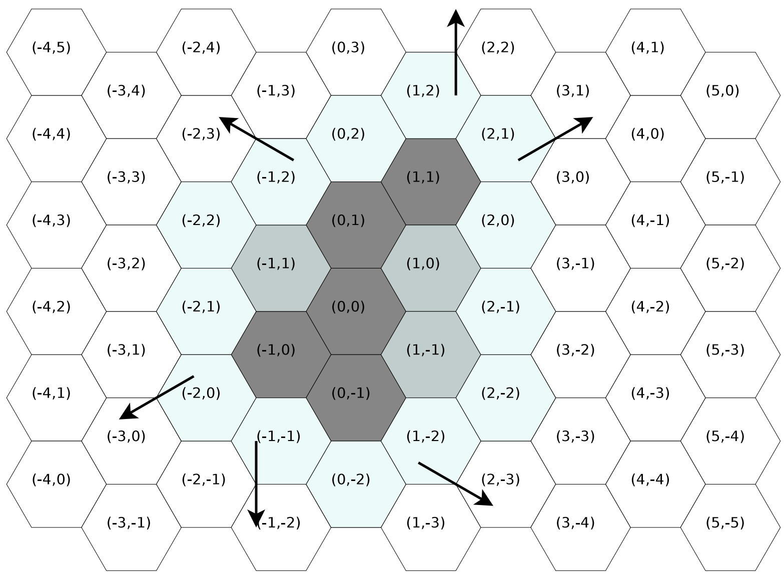

Figure 5: The shape A is given in dark gray, the hexagons not in A but in the minimal simple shape B containing A are lighter colored, and $N(B)$ are lightest. Here each arrow in direction d_i starts at n_i, to find a_i look in the reverse direction. Note that $a_i = a_j$ is possible, but $n_i = n_j$ is not.

This way we define n_i for $i = 0 \ldots 5$. Note that $N(B)$, as a subgraph of the graph we discussed on H, is a simple cycle. Then $\{n_i\}_{i=0}^5$ partition $N(B) \setminus \{n_i\}_{i=0}^5$ into at most 6 (possibly empty) connected components. Empty components will occur if two or more of $\{n_i\}_{i=0}^5$ are adjacent to one another. We denote $N_{i,i+1}$ to be the component bounded between n_i and n_{i+1}.

Now we define $d : N(B) \to \{d_i\}_{i=0}^5$ so that it maps to each element $n \in N(B)$ a direction $d(n)$ (see Figure 6). For each direction i, we have $d(n_i) = d_i$, and if $n \in N_{i,i+1}$ then $d(n) = d_i$. Note that because d_i is a direction in which a_i is extremal, and B is minimal such that $A \subset B$, a half line from n in direction d_i points away from B so that $n + kd(n) \notin B \cup N(B)$ for any $k \geq 1$. Because A is contained in B, we also know that $n + kd(n) \notin A \cup N(A)$.

Let $n \in N(B)$, and let $d_j = d(n)$, we will now show that there exists a non negative p such that $n - pd_j \in N(A)$. For $n = n_j$, this holds for $p = 0$, then

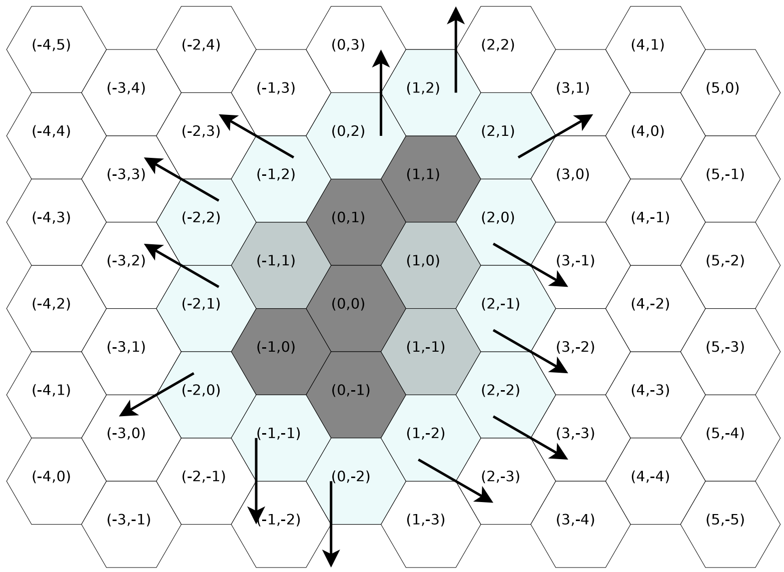

Figure 6: Here we see $d(n)$ for each n. Note that $n - k \cdot d(n)$ for $k \in \mathbb{Z}$ passes through some element of B, therefore through exactly two elements of $N(B)$.

we prove this only for $n \in N_{j,j+1}$. Assume by contradiction $n - pd_j \notin N(A)$ for every non negative integer p. Then the set $L_{n,j,j+3}$ is disjoint to $N(A)$, and has one of n_j, n_{j+1} on each side of it, then by the corollary of separation, A is not optimal, then our assumption is false. In Figure 5, for $n = (2, -1)$, the cut would be parallel to the arrow through $(1, -2)$, separating it from $(2, 1)$.

Then we conclude that for each $n \in N(B)$ there exists $p \in \mathbb{N} \cup \{0\}$ 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 are along the half line from $m(s)$ in direction $d(s)$, and on its intersection with $N(B)$, therefore they are equal, because we've noted that $n + kd(n) \notin B \cup N(B)$ for any n and $k \geq 1$, so that this intersection has only one hexagon. Now we assume that $d(t) \neq d(s)$, and wlog that $d(t) = d_i; d(s) = d_j$ and $j < i$. If neither of t, s is in $\{n_i\}_{i=0}^{5}$, $L_{m(t)=m(s),d_i,d_j}$ allows a separation argument, with at least one element of $\{n_i\}_{i=0}^{5}$ that separated t from s in each direction ending up in each component of X. Then A is not optimal, which is a contradiction. Otherwise, by the assignment of directions, t is one of the special n_i. Then $m(s) = m(t) = t$, then s is along the half line from t in direction d_j, and because $-d_j$ is assigned to be into B, we must have $d_j = d_i - 1$, and then by the construction of B, $s = t$. $\square$

2.3 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 their coordinate mappings. 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 simple shapes in H were defined in terms of linear constraints on its image in $\mathbb{Z}^2$ under a coordinate mapping $\phi_{h,i}$, having the form

$$\left\{ (p,q) \in \mathbb{Z}^2 | \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \in \{a,\ldots,d\} \times \{b,\ldots,e\} \times \{c,\ldots,f\} \right\}$$

In the following, we will refer to such subsets of $\mathbb{Z}^2$ also as simple shapes, and consider their properties. We note that a,d,b,e describe bounds parallel to the axes, while c,f describe upper and lower bounds on the sum of the coordinates. 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 each of 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 are $j = c-1$ and $k = a+b-f$. We will say that this shape has dimensions d,f and ear sizes j,k. Having discarded the location of a shape, we will now use a,b instead of d,f to denote its dimensions. Figure 7 provides an example of this notation.

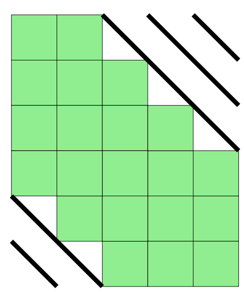

Figure 7: This is a simple shape that can be described as having dimensions $a=6, b=5$ and ear sizes $j=2; k=3$.

Lemma 2.11. *A simple shape A in $\mathbb{Z}^2$ 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 in this context we consider two tiles to be neighbors if they share a side or if they share a corner and are 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 rectangle, 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 reducing the number of neighbors by one. In general, the nth 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 $n+1$ neigbors, and add n that were tiles, and in total reduce the number of neighbors by one, for each diagonal removed. ☐

Lemma 2.12. *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. A similar inductive argument applies. 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 2.13. *Let m, n be the lengths of two sides of an optimal simple shape, that are not adjacent. Then $|m-n| < 2$.*

Proof. We assume throughout without loss of generality that $m \geq n$. First we consider the case in which the two sides are parallel, then there is a coordinate mapping such that the two sides are mapped in $\mathbb{Z}^2$ to the ears, then $m = j+1; n = k+1$. Assume by contradiction that $|m-n| \geq 2$, then $|j-k| \geq 2$, and we conclude 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:

$$\begin{aligned} &|B| - |A| = \\ a \cdot b - \frac{(j-1)\cdot j}{2} - \frac{(k+1)\,(k+2)}{2} - \left(a \cdot b - \frac{j\,(j+1)}{2} - \frac{k\,(k+1)}{2}\right) &= \\ \frac{1}{2}\left(-\left(j-1\right)\cdot j) - (k+1)\,(k+2) + j\,(j+1) + k\,(k+1)\right) &= \\ \frac{1}{2}\left(2j - 2(k+1)\right) &= \\ j-k-1 &\geq \\ k+2-k-1 &= 1 \end{aligned}$$

Then A was not optimal, because B had larger area for the same perimeter, contradicting our assumption.

Now we consider the case in which the two sides have a single side separating them. Then there exists a coordinate mapping such that the side separating them is an ear. We can choose it so that $m = a - j$, $n = b - j$. We assume by contradiction that $|m - n| \geq 2$, then $m \geq 2 + n$, and 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 this is again not optimal. $\square$

Corollary 2.14. *For any i, the image of an optimal simple shape A under $\phi_{h,i}$ fulfills the following relations, $|j - k| < 2$, $|a - b| < 2$, $|j + 1 - (a - k)| < 2$ 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 $b - j$. $|(b - j) - (k + 1)| < 2$ because they are separated only by $a - k$. $|a - b| < 2$ because $a - j, b - j$ are separated only by $j + 1$. $\square$

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:

$$\begin{aligned}
2 &> \\
|(b - j) - (k + 1)| &= \\
|(a + d - j) - (j + \delta + 1)| &= \\
|a + d - j - j - \delta - 1| &= \\
|a + d - 2j - \delta - 1| &= \\
|(2j + \delta + \Delta + 1) + d - 2j - \delta - 1| &= \\
|\Delta + d| &= \\
\Delta + d &
\end{aligned}$$

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

Using this notation, we see that the optimal shapes must 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, δ, d, Δ. Here $j + 1$ is the length of the shortest side, so any $j \in \mathbb{N} \cup \{0\}$ can occur. Considering the symmetry eliminating inequalities, and writing $p_i = (\delta, d, \Delta)$, we allow only $p_1 = (0,0,0)$, $p_2 = (0,1,0)$, $p_3 = (1,0,0)$, $p_4 = (0,0,1)$, $p_5 = (1,0,1)$, $p_6 = (1,1,0)$. 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:

$$\begin{aligned}
2(a+b+1)-j-k &= \\
2(2j+\delta+\Delta+1+2j+\delta+\Delta+1+d+1)-j-j-\delta &= \\
2(4j+2\Delta+2\delta+d+3)-2j-\delta &= \\
6j+4\Delta+4\delta+2d+6-\delta &= \\
6j+4\Delta+3\delta+2d+6 &
\end{aligned}$$

The area of such a shape is $a \cdot b - (\frac{j(j+1)}{2} + \frac{k(k+1)}{2})$. Now we will partially simplify these parts. By substitution $a \cdot b = (2j+\delta+\Delta+1)\cdot(2j+\delta+\Delta+1+d)$. We now use the fact that $\delta \in \{0,1\}$ to compute:

$$\begin{aligned}
\frac{j(j+1)}{2} + \frac{k(k+1)}{2} &= \\
\frac{j(j+1)}{2} + \frac{(j+\delta)(j+\delta+1)}{2} &= \\
\frac{j^2+j}{2} + \frac{j^2+2j\delta+\delta^2+j+\delta}{2} &= \\
\jmath^2+j+\delta j+\delta &= \\
\jmath^2+j(1+\delta)+\delta &
\end{aligned}$$

Lemma 2.15. *Let $A(i,j)$ be the area of the simple shape of form $(a,b,j,k) = (2j+\delta+\Delta+1, 2j+\delta+\Delta+1+d, j, j+\delta)$, where $j+1$ is the shortest side and the other parameters (d,δ,Δ) equal to p_i which is the ith 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 $j \in \mathbb{N} \cup \{0\}$: $A(1,j) \le A(2,j) \le A(3,j) \le A(4,j) \le A(5,j) \le A(1,j+1) \le A(6,j) \le A(2,j+1)$*

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

$$\begin{aligned}
a \cdot b &= \\
(2(j+1)+\delta+\Delta+1)\cdot(2(j+1)+\delta+\Delta+1+d) &= \\
(2(j+1)+1+1+1)\cdot(2(j+1)+1+1+1+0) &= \\
(2j+4)^2 &
\end{aligned}$$

Then

$$\begin{aligned} A\left(2, j+1\right) &= \\ \left(2j+4\right)^{2}-\left(j+1\right)^{2}-\left(j+1\right)\left(1+0\right)-0 &= \\ \left(2j+4\right)^{2}-\left(j+1\right)^{2}-\left(j+1\right) &= \\ 4j^{2}+16j+16-j^{2}-2j-1-j-1 &= 3j^{2}+13j+14 \end{aligned}$$

Similarly we find:

$$\begin{aligned} A\left(6, j\right) &= 3j^{2}+10j+8 \\ A\left(1, j+1\right) &= 3j^{2}+9j+7 \\ A\left(5, j\right) &= 3j^{2}+8j+5 \\ A\left(4, j\right) &= 3j^{2}+7j+4 \\ A\left(3, j\right) &= 3j^{2}+6j+3 \\ A\left(2, j\right) &= 3j^{2}+5j+2 \\ A\left(1, j\right) &= 3j^{2}+3j+1 \end{aligned}$$

And indeed these fulfill the order required. □

Lemma 2.16. *For shapes having particular i, j parameters as in the previous lemma, we define $N\left(i, j\right)$ to be the number of its neighbors. Then for every $j \in \mathbb{N} \cup \{0\}$, $N\left(1, j\right) \leq N\left(2, j\right) \leq N\left(3, j\right) \leq N\left(4, j\right) \leq N\left(5, j\right) \leq N\left(1, j+1\right) \leq N\left(6, j\right) \leq N\left(2, j+1\right)$*

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 $N\left(i, j\right) = 6j + 4\Delta + 3\delta + 2d + 6$.

$$\begin{aligned} N\left(1, j\right) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 + 6 = 6j+6 \\ N\left(2, j\right) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 1 + 6 = 6j+8 \\ N\left(3, j\right) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 1 + 2 \cdot 0 + 6 = 6j+9 \\ N\left(4, j\right) &= 6 \cdot j + 4 \cdot 1 + 3 \cdot 0 + 2 \cdot 0 + 6 = 6j+10 \\ N\left(5, j\right) &= 6 \cdot j + 4 \cdot 0 + 3 \cdot 1 + 2 \cdot 1 + 6 = 6j+11 \\ N\left(1, j+1\right) &= 6 \cdot \left(j+1\right) + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 0 + 6 = 6j+12 \\ N\left(6, j\right) &= 6 \cdot j + 4 \cdot 1 + 3 \cdot 1 + 2 \cdot 0 + 6 = 6j+13 \\ N\left(2, j+1\right) &= 6 \cdot \left(j+1\right) + 4 \cdot 0 + 3 \cdot 0 + 2 \cdot 1 + 6 = 6j+14 \end{aligned}$$

And these fulfill the required order. □

Lemma 2.17. *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.*

Proof. First note that a non optimal shape can be improved by some other shape. This shape is either optimal, or it can be in turn also improved. Continuing in this form as long as possible, we have two resulting scenarios: either a finite chain of improvement, ending in an optimal shape which clearly also improves on the initial shape, or an infinite chain of non-optimal shapes, each strictly improving on the previous ones.

Improvement means either reducing the boundary size, which can only be done a finite number of times, or increasing the area. The isoperimetric inequality for $\mathbb{R}^2$ then tells us that as the area grows unboundedly, so must the length of the curve enclosing it. But this curve is a subset of the union of the curves enclosing the neighbors of a shape. Since there is a bounded number of neighbors, and each has just 6 sides of fixed length, we cannot have infinite growth in area, therefore infinite improvement is impossible. Then any non optimal shape is in particular improved by some optimal shape.

We have shown that all the optimal shapes are in W. Then by the argument above, any shape in W that is not Pareto optimal, can be improved by one of the shapes in W. Next we show this does not occur.

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. $\square$

Note that this lemma completes the proof of Theorem 1.2, and then Theorem 1.3 results directly from applying the calculations in the proof of Lemma 2.15.

3 Related Work and Concluding Remarks

As we mentioned before, we can also consider a different boundary measure: the number of edges that the shape shares with hexagons outside it, or equivalently, the length of the minimal curve enclosing the shape. This edge count boundary measure was explored through its boundary optimal shapes by Harary and Harborth [3] for hexagons, rectangles and triangles, which are there given the name "extremal animals". The spiral construction is used in this study and others, because it provides for every value of the area, a boundary optimal shape. Thus a spiral generates many boundary optimal shapes, though certainly not all of them. In fact, it also generates many of the Pareto optimal shapes, though not always all of them [1].

Brunvoll, Cyvin and Cyvin, among others, explore the set of extremal animals, and in [2] describe the set of hexagonal shapes that are also area optimal, there called circular animals. Part of the interest in the shapes on the hexagonal grid comes from the field of mathematical chemistry, due to their correspondence to C_nH_s molecules, also known as benzenoids, in which the carbon atoms are organized in hexagons, and the hydrogen atoms appear along the otherwise free bonds along the edges of this planar molecule. The number of atoms of each kind are uniquely determined by the number of hexagons and the edge boundary size by the formula $(n; s) = \left(2 \cdot a + \frac{e}{2} + 1; \frac{e}{2} + 3\right)$ where a is the area or number of hexagons in the shape, and e is the number of edges in its boundary.

The fact that the edge count boundary measure is not equivalent to the boundary measure we considered before can be seen using the example in Figure 8. Thus, it is somewhat surprising that the sets of shapes that are Pareto optimal for the two measures are in fact equal, and thus appear quite often in the literature about hexagonal extremal animals. It is not clear however, whether the sets of boundary optimal shapes for the two measures are the same.

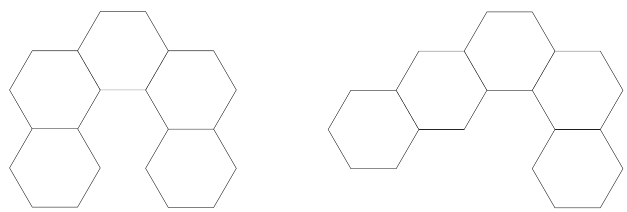

Figure 8: These shapes have 22 edge segments each. The shape on the left has only 13 neighbors, while the right has 14, because one of its neighbors is shared by two of its hexagons. It is clear that this proximity does not affect the count of edges.

A similar exploration, applied to two different boundary measures on the square tiling of the plane, in addition to the hexagonal case discussed here, is described in detail in a companion Technical Report [6].

References

[1] Yaniv Altshuler, Vladimir Yanovsky, Daniel Vainsencher, Israel A. Wagner, and Alfred M. Bruckstein. On minimal perimeter polyminoes. *Lecture Notes in Computer Science*, 4245:17, 2006.

[2] Jon Brunvoll, Bjørg N. Cyvin, and Sven J. Cyvin. More about extremal animals. *Journal of Mathematical Chemistry*, 1993.

[3] Frank Harary and Heiko Harborth. Extremal animals. *Journal of Combinatorics, Information and System Sciences*, 1(1):1–8, 1976.

[4] Martin J Osborne and Ariel Rubinstein. *A course in game theory.* MIT Press, Cambridge, Mass., 1994.

[5] N. J. A. Sloane. The on-line encyclopedia of integer sequences. Published electronically at `www.research.att.com/~njas/sequences/`, 1996-2006.

[6] Daniel Vainsencher and Alfred M. Bruckstein. Isoperimetrically optimal polyforms. CIS Technical Report 06, CS Department, Technion, 2007.