DOMINO TILINGS, HEIGHT FUNCTIONS AND CONWAY'S TILING GROUP

FLORESTAN BRUNCK

1. Introduction

Packing, covering and tiling problems are among the most basic combinatorial and geometrical problems, yet they were quickly understood to be quite difficult computationally. In the following we consider the square lattice $\mathbb{Z}^2$. A *tile* is a union of cells of this lattice. A tiling of a region R (obtained as a union of cells of $\mathbb{Z}^2$) by a set of tiles Σ is both a covering and a packing of R by translated copies of tiles of Σ, i.e. a covering of R by translated tiles of Σ whose interiors do not intersect. We will refer to the union of two adjacent cubic cells as a *domino*. Perhaps the most fundamental problems are the following two:

TILING PROBLEM:
INSTANCE: *A finite set Σ of tiles.*
QUESTION: *Does Σ tile the entire $\mathbb{Z}^2$ lattice?*

FINITE TILING PROBLEM:
INSTANCE: *A region R and a finite set Σ of tiles.*
QUESTION: *Does Σ tile R?*

The first problem was expected to be rather easy but was surprisingly shown to be undecidable. This owed to the belief at the time that aperiodic tilings did not exist. Indeed, if all tilings of $\mathbb{Z}^2$ were periodic, two simple semi-algorithms could be run in parallel to decide the tiling problem. First, a compactness argument shows that if one can cover larger and larger disks with tiles from Σ then one can tile the whole plane with tiles of Σ. By contrapositive, if Σ does not tile the whole plane, there must exist some disk D_n of radius n which cannot be covered by Σ and there is a semi-algorithm which attempts covering larger and larger disks until it fails to do so and stops to confirm that Σ cannot cover the whole plane. On the other hand, if all tilings are periodic then we need only exhaustively search for the period and try to tile bigger and bigger rectangles while trying to match left-right and top-bottom. If Σ tiles $\mathbb{Z}^2$, a periodic box will eventually be found. Running both semi-algorithms in parallel would decide the tiling problem were it not for the existence of aperiodic tilings. Against the expectations, Berger discovered the first aperiodic set of tiles [1] (with roughly 10 000 tiles, later reduced all the way to 2 by Penrose and eventually to 1 by Smith et al [7] earlier this year).

Theorem 1.1 (Berger)**.** *There exist aperiodic sets of tiles, namely, sets of tiles that tile the whole plane but only aperiodically.*

And the tiling problem was shown to be undecidable by Berger [1].

Theorem 1.2 (Berger)**.** *The tiling problem is undecidable.*

Very informally, the essence of the proof is a reduction to the halting problem where one shows how to simulate the computation of a Turing machine with a specifically designed set of tiles. Horizontal lines correspond to the tape of the Turing machine, marked with the current symbol being read and the current state transition being triggered. The tiles are designed so that the only way to obey the matching rules is to tile the horizontal row above to reflect the next state of the Turing machine and its tape. If the Turing machine terminates, the set of tiles cannot tile the whole plane and vice versa. Thus is one could decide the tiling problem, one could decide the halting problem.

The second problem, being finite, is obviously decidable by exhaustive search. However, it was shown to be in NP, and in fact NP-complete [4]. Thus we should not be surprised that even finite tiling problems can be *very* hard.

Theorem 1.3 (Lewis)**.** *The finite tiling problem is NP-complete.*

Interestingly though, the particular finite region to be tiled, as well as the tiles play a major role in the hardness of the problem, and a lot of subclasses of tilings have been given polynomial time algorithm already. For recent work in this fascinating topic, see for example [5]. We just mention in passing this theorem of Pak and Yang:

Theorem 1.4 (Pak, Yang)**.** *There exists a finite set $\mathcal{R}$ of at most 10^6 rectangular tiles, such that the tileability problem of simply connected regions with $\mathcal{R}$ is NP-complete.*

In addition, combinatorial colouring arguments starting emerging to give criteria for when a region would be guaranteed to be impossible to tile. Eventually, a number of these arguments coagulated and were subsumed by Conway's Tiling Group criteria [3, 8], which in the case of domino tilings for example gave rise to a necessary and sufficient criterion for tileability of a region by Thurston together with the discovery of *height functions*.

Before explaining what height functions are and where they come from, let us give a few important results first that arise from it.

Theorem 1.5 (Thurston)**.** *There exists a linear time (in the size of the region) to decide if a region admits a tiling by dominoes, and if so, output a tiling of it.*

Remark 1.6. This an improvement over the polynomial bound of $O(\sqrt{V}E)$ given by the best algorithms for perfect matchings in a graph $G = (V, E)$.

Perhaps the most remarkable and elegant perspective to come out of height functions is that the set of all domino tilings of a simply connected region forms a single distributive lattice whose partial order is given by a local operation called a flip [6].

Theorem 1.7. *The set of domino tilings of a fixed simply connected region forms a distributive lattice whose partial order coincides with the flip order.*

Which in particular gives connectivity under flips and easy formulas to compute flip distances.

Theorem 1.8. *Any given tiling by dominos of a simply connected region is connected by flips.*

2. What are height functions?

There are at least three different ways to define and prove properties about height functions:

(1) The physics way [2], which sees it as the gradient arising from the unique curl-free flow one can associate to a tiling by domino / a perfect matching

of the region involved. The construction may seem ad hoc from the outside but is elegant and quick. Although this is because the core property is now hidden in the proof that curl-free flows all arise from a gradient. Note that this uses ultimately Stokes' theorem which requires simple connectedness (and indeed all results fail if we allow the region R to have holes for examples. The main advantage is to make it easy to generalise to higher dimensions because this flow can be defined just as well, and even if the curl of a flow doesn't really exist in higher dimensions, we can treat directly with the flow itself.

(2) The computer science way [6], equally ad hoc, elementary and thorough but tedious and proving properties feels a little unecessary once one learns about (3). No direct hint of generalisation.

(3) The (geometric) group theoretic way [3, 8], which arguably motivates the definition/discovery the most. Quite a general method, with possible hints of generalisations.

The hope is to associate a unique function to a domino tiling of a fixed region so that we can fully encode tilings as height functions and only deal with the height functions themselves, which have very nice properties. In practice a height function is simply a function $h : V(R) \to \mathbb{Z}$ from the vertex set of the lattice region R to the integers.

(1) Curl-free Lattice Flows. The tiling is a described here first as a matching of the dual. Which we partition between odd and even vertices (even vertices are depicted in blue in Figure 1). The dual us oriented from even to odd vertices and a positive unit charge flows from a domino's even vertex to its matched odd vertex. Subtracting a reference flow where even vertices flow with a $-\frac{1}{2d}$ charge to their odd neighbours (where d is the dimension) makes this flow divergence free. This directly induces a curl-free flow in the original lattice by rotating all edges and their associated orientation 90 degrees (see Figure 1). Once again, curl-free flows on simply connected regions all arise from a gradient, which we define to be the height associated to that tiling. Schematically:

Tiling $\iff$ curl-free Lattice flow on $\mathbb{Z}^2$ $\iff$ gradient = height function

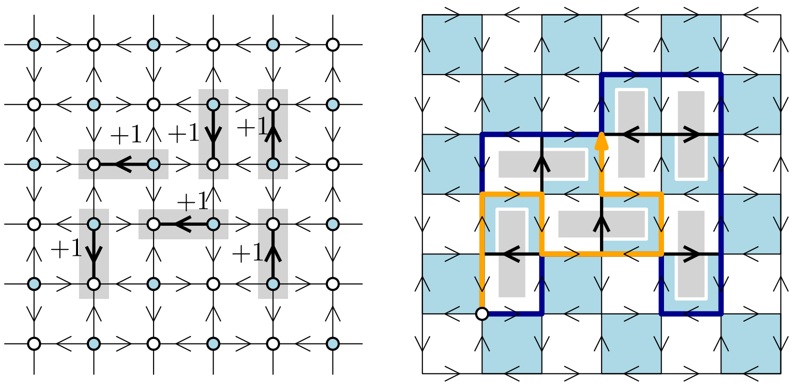


FIGURE 1: A divergence-free flow in the dual (on the left) corresponds to a curl-free flow in the square lattice (on the right).

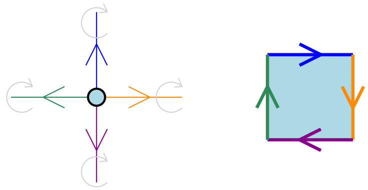


FIGURE 2: The correspondence is given by rotation each edge clockwise.

(2) Random Arrows on Checkerboards. This is essentialy the same as (3), but forgetting its origin and motivation. It amounts to defining the same orientation as in (1) and, starting from a designated reference point, following the arrows counting direct arrows as $+1$ and indirect as -1 to give a vertex its height. One easily shows that this is well defined as long as we only allow paths going around tiles (such as the orange one on Figure 1). We will refer to those as *tiling paths* in the next paragraph. It is then showed that if we restrict integer functions on the vertices to obey the same rules as the height functions arising from tilings (namely, increase either by 1 or by 3 following direct arrows), we get an expected 1-to-1 correspondence between tilings and this subset of height function.

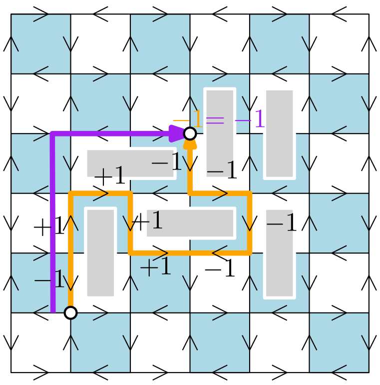


FIGURE 3: Any two *tiling paths* from and to the same points will yield the same height difference.

Once we equate tilings with height functions, it is quick to show that tilings form a distributive lattice, where meet and join correspond to min and max of the functions. And one can see that flips correspond to local min/local max.

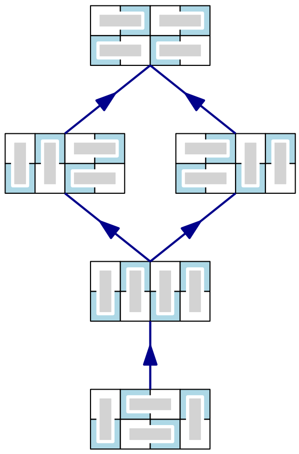


FIGURE 4: The lattice of domino tilings, going from the minimum tiling at the bottom to the maximum tiling at the top.

(3) Conway's Tiling Groups. The previous two definitions arguably hide *why* height functions exist at all and why we should expect this random colouring/orientation on the squares to work all. The price of smooth sailing is a lack of intuition and understanding as to how one might arrive at the definition in the first place. To remedy that, we want to show how it is in fact very natural to discover height functions as a lift of the original tiling to the *tiling group*, defined by Conway. And why this lifting idea is a very natural, assuming one is familiar with elementary rudiments of geometric group theory/algebraic topology. Deep down, we will see that it all boils down to a very simple application of one of the most beautiful connections between algebra and geometry: the correspondence between normal subgroups and covering spaces.

In group theory, we can define groups via *presentations*. Essentially, we describe a group by its generators (think $(0,1)$ and $(1,0)$ for the lattice $\mathbb{Z}^2$) and specify its structure by a list of the *relations* (also called relators) that it satisfies. For $\mathbb{Z}^2$, the only structure is really that we have squares, i.e. that ab and ba commute, so the only relation is $ab = ba$ or simply written as $aba^{-1}b^{-1}$ (meaning that this word is trivial in the group). The proper way of defining $\mathbb{Z}^2$ that way is to start with the free group on two generators $\mathbb{F}_2$, i.e. the group with two generators a and b and no structure/relation, so any word in a and b and their inverses (like $abbab^{-1}$) is valid and distinct. Then kill off (i.e. identify to the identity) all the words that correspond to squares. The proper way of doing that is to quotient by the normal subgroup we get by taking all the conjugates of squares (the so-called *normal closure*). Even without going into the apparent technicalities of why we need subgroups to be normal (more on that in a different post and how normality is a particular instance of the natural notion of *congruence*), this is what we want: conjugates of the square at the origin are all the other squares conjugated by the path leading up to them.

The next step is to start introducing geometry and talk about *Cayley graphs*. Given a group G with two generators a and b (the same works for any group of course), the Cayley graph $\Gamma(G)$ is the directed graph where vertices are group elements and we draw a directed edge between two group elements/vertices if they differ by a generator. So that for each vertex, we have four edges, two outgoing edges leading from w to wa and wb, and two incoming from wa^{-1} and wb^{-1}.

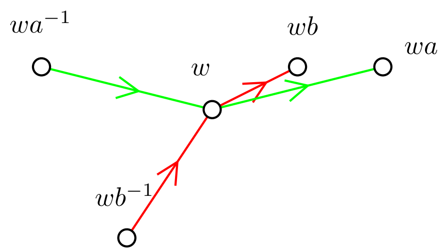


FIGURE 5: A vertex in the Cayley graph $\Gamma(G)$.

In a group defined by a presentation, a word in its generators corresponds to the identity if and only if it can be written as a product of conjugates of relators (that was our definition). Pictorially, in the Cayley graph $\Gamma(G)$, this corresponds to having the word under consideration trace a loop to and from the origin and tiling it with tiles corresponding to conjugates of the relators. So in $\mathbb{Z}^2$, we see that a word is trivial if it goes around the grid and come back to the origin. It is then tiles by conjugates of squares.

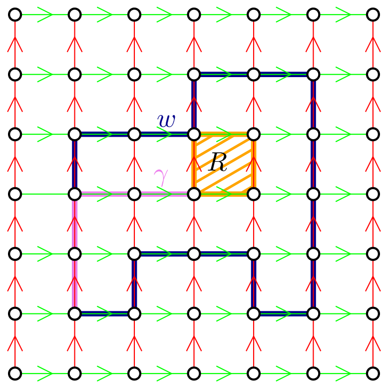


FIGURE 6: A word w is trivial in the Cayley graph $\Gamma(G)$ if it traces a loop to/from the origin (in dark blue) and can be tiled with conjugates of relators, such as the orange hatched square conjugated by the violet path γ, i.e. $\gamma(aba^{-1}b^{-1})\gamma^{-1}$.

Conway's simple but clever idea is then to define a group T with generators given by the underlying lattice of the tiling (the squares in our case) and such that its structure/relations correspond exactly to reading off words going around tiles in all their possible orientations. Because then deciding if a particular word (the word read going around a particular polygonal boundary in the lattice) is trivial in the tiling group is the same as deciding whether it can be tiled! This is the part that is usually proved by induction in definition (2): it is certainly true for a single tile, and then we can just cut up the boundary as the concatenation of two smaller loops to get the results. Note the small technicality that there are false positives if we allow the region to have cut vertices for example. But then each side of the cut vertex can be treated separately.

In the case of domino tilings, the generators say a and b are the standard $\mathbb{Z}^2$ generators $(1,0)$ and $(0,1)$. The contour on a domino then either reads the word $D_H = ab^2a^{-1}b^{-2}$ or $D_V = a^2ba^{-2}b^{-1}$ depending on whether they are laid horizontally or vertically. The tiling group (also called the domino group in this case) is then simply defined as the group D with presentation:

$$D = \langle a, b \mid D_H, D_V \rangle$$

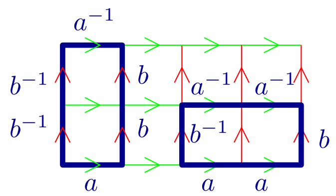


FIGURE 7: The relators D_V and D_H in the Domino group.

Now the magic is that D is a quotient of $\mathbb{Z}^2$, which means that it is a covering space of the planar lattice. So the hope is that we can *see* the Cayley graph of D directly hovering over $\mathbb{Z}^2$. We will see how to do just that but for now, we can already notice the core property that allows the first two definitions (1) and (2) to work: every path in the underlying lattice can be seen as a word in D and has a unique lift in D, and if we fix a region R (i.e. a word w which form a loop to the origin) and a tiling of it, then:

Proposition 2.1. *Any* tiling loop*, i.e. a closed path π which "follows the tiling" and never cuts a domino transversely, is trivial.*

This is really immediate: if we go around dominos, we're really following conjugates of relators and making sure our loop is filled with conjugates of relators, i.e. we're making sure that it is trivial.

Corollary 2.2. *(non necessarily closed) Tiling paths with the same start and end point have the same lift in D.*

This allows us to have a well defined lift of the entire tiling to the Cayley graph of D. Then the really nice thing in the case of dominoes or lozenge tilings is that the quotient of the projection from the tiling group to the lattice group is isomorphic to $\mathbb{Z}$, which allows one to associate a height to each lifter point instead. This gives rise to the promised height function. But in practice, what makes height functions work and give a criterion and algorithm to output a tiling if one exists and figure out if it does not is really the knowledge of the Cayley graph/tiling group.

Now onto building $\Gamma(G)$. We try to see what happens we move across the underlying lattice and figure out when we loop to the same vertex or end up in different places. Starting from a vertex w and walking along b we reach a new vertex. Multiplying by a^{-1} and generally following the boundary of the square in which w was the lower right corner, we end up with 5 different words/vertices by the time we're back directly under w. And we can see that this keeps going indefinitely, as we cycle around this square counter clockwise, all the words we obtain are different, so we can start by placing on a counterclockwise coil above this square (see Figure 8).

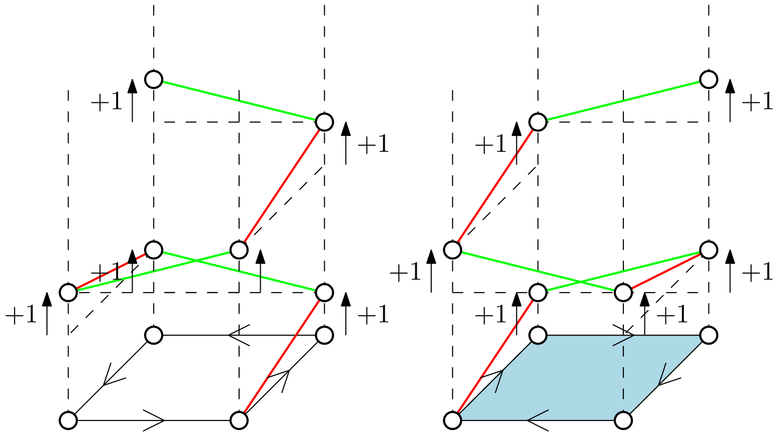


FIGURE 8: The Cayley graph $\Gamma(G)$ is constructed from joining clockwise and counterclockwise "coils" spiralling over each square of $\mathbb{Z}^2$.

This is true of all squares, so that we may think of placing a coil on top of every square. But the issue if then of course to figure out how these coils interact and match up on the boundaries of squares. If we explore the adjacent square in which w is the lower left corner, we realise that matching them up forces us to reverse the orientation of one of the coils, so that there is a clockwise coil on top of a square and a counter clockwise coil on top of all adjacent squares (see Figure 9).

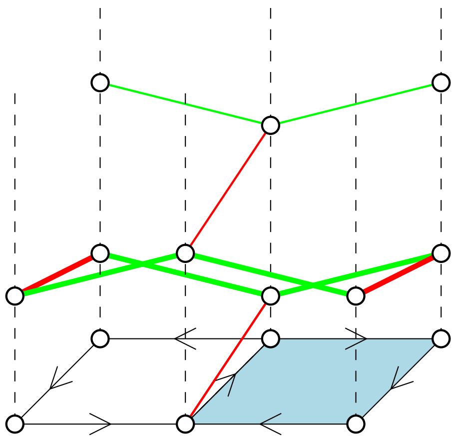


FIGURE 9: After joining coils, we see the relators of dominos appear (in bold on the figure).

Flips Seen in the Cayley Graph. So now, going back to (2) we can directly see the effect of a flip on the Cayley graph. As can be seen from the formulation of the meet-join distributive lattice (where the meet operation is taking the min and the join is taking the max), flip correspond to local minima or maxima within a square of 4 squares. In the Cayley graph, this looks like a bun made of 4 cells, the top 2 of which will be horizontal/vertical and the bottom 2 of which will be vertical/horizontal.

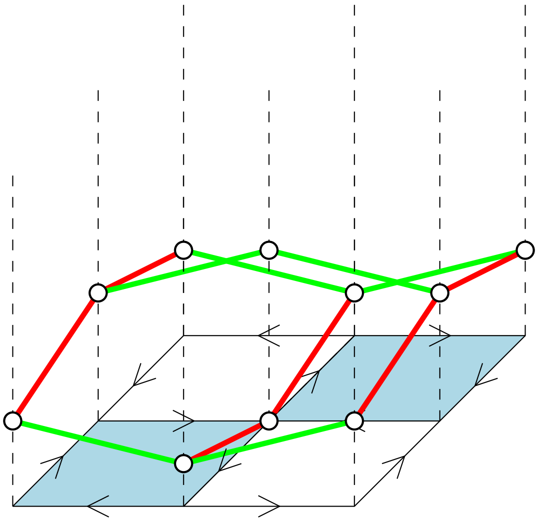
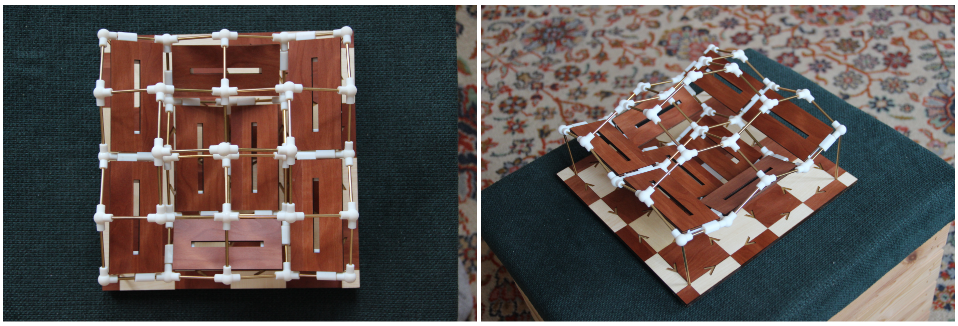


Figure 10: A flip corresponds to looking at either the top or the bottom of a two-by-two cell.

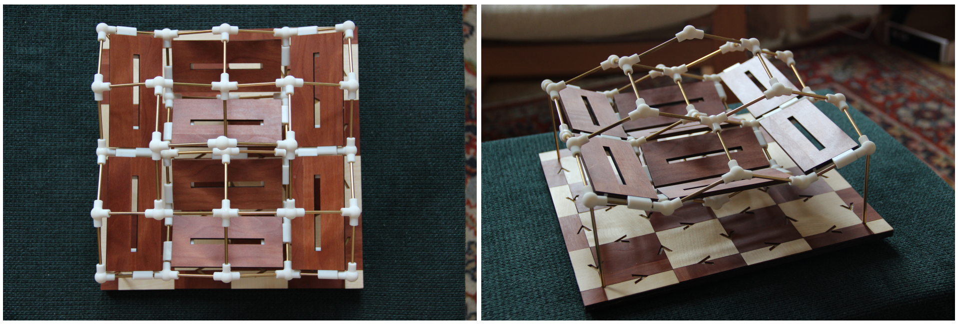
3. Thurston's algorithm

The idea here to output a single tiling of R (or to figure out that there's no such tiling at all) is to produce the one with the lowest lift. If there's a tiling at all, then there is a lowest tiling, and it is straightforward how to construct it. Being the lowest tiling, the highest vertices will be on the boundary, so we can start there (if it was not on the boundary, we could punch down cells, i.e. flip them to go lower). Due to the observation also that the boundary is necessarily a tiling path no matter the tiling, the heights of its vertices is independent of the tiling involved. So we start there. Given that our cells must go down from the boundary, the only possibility is to lay dominoes such that their centre is lined up against a local max on the boundary (see Figure 11). We then simply keep iterating this process, adjusting the boundary as we go. Either this process terminates and we reach the minimum tiling of our region, or we can no longer place a tile and conclude that our region cannot be tiled.



(A) : Seen from the top. (B) : Seen in perspective.

Figure 11: The minimum domino tiling for a four-by-four square.



(A) : Seen from the top. (B) : Seen in perspective.

FIGURE 12: A tiling obtained from flipping the top-centre two-by-two square in the minimum tiling.

REFERENCES

[1] R. Berger. The undecidability of the domino problem. *Amer. Math. Soc.*, 66, 1966. ↑1

[2] N. Chandgotia, S. Sheffield, and C. Wolfram. Large deviations for the 3d dimer model. *arXiv preprint arXiv:2304.08468*, 2023. ↑2

[3] J.H. Conway and J.C. Lagarias. Tiling with polyominoes and combinatorial group theory. *Journal of combinatorial theory, Series A*, 53(2):183–208, 1990. ↑2, ↑3

[4] H. R. Lewis. Complexity of solvable cases of the decision problem for the predicate calculus. In *19th Annual Symposium on Foundations of Computer Science (sfcs 1978)*, pages 35–47. IEEE, 1978. ↑2

[5] I. Pak and J. Yang. Tiling simply connected regions with rectangles. *Journal of Combinatorial Theory, Series A*, 120(7):1804–1816, 2013. ↑2

[6] E. Rémila. The lattice structure of the set of domino tilings of a polygon. *Theoretical computer science*, 322(2):409–422, 2004. ↑2, ↑3

[7] D. Smith, J.S. Myers, C.S. Kaplan, and C. Goodman-Strauss. An aperiodic monotile. *arXiv preprint arXiv:2303.10798*, 2023. ↑1

[8] P. Thurston, W. Conway's tiling groups. *The American Mathematical Monthly*, 97(8):757–773, 1990. ↑2, ↑3

INSTITUTE OF SCIENCE AND TECHNOLOGY AUSTRIA, AM CAMPUS 1, 3400, KLOSTERNEUBURG, LOWER AUSTRIA, AUSTRIA

Email address: florestan.brunck@mail.mcgill.ca