HW1 Molecular Programming

MPRI 2.11.1

16.12.2021 - Due on Thu. 06/01 before 08:45

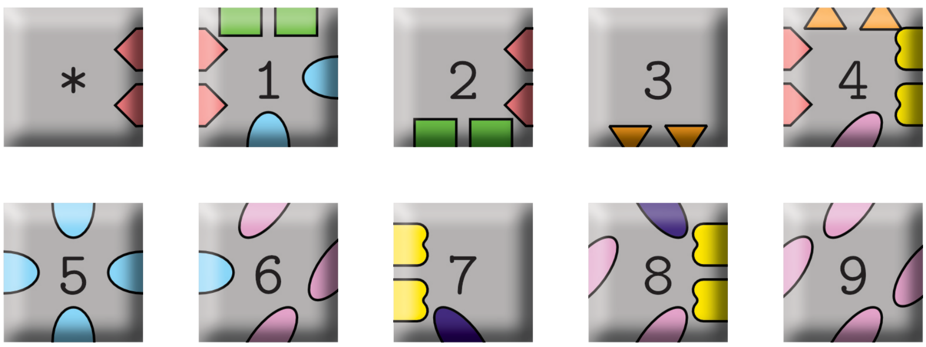
You are asked to complete the exercise marked with a [★] and to send me your solutions to:

nicolas.schabanel@ens-lyon.fr

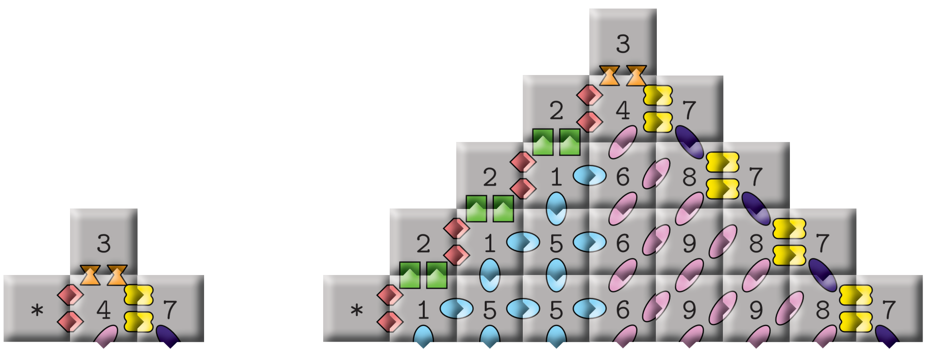
as a PDF file named **HW1-Lastname.pdf** on **Thu. 06/01 before 08:45**.

■ **Exercise 1 (Algorithmic Self-Assembly).** Recall that the self-assembly process consists in, given a finite tileset (with infinitely many tiles of each type), starting from the seed tile (marked with a ★), glueing tiles with matching colors to the current aggregate so that each new tile is attached by at least *two* links to the aggregate (either on the same border or on two borders). Recall that a shape is *final* if no tiles can be attached to it anymore.

► **Question 1.1)** *What is the exact family of final shapes self-assembled by the following tileset? (No proof nor justification is asked.) Indicate the local order of assembly by drawing arrows over the tiles of a generic final shape. Which are the two competing tiles that decide the size of the resulting final shape?*

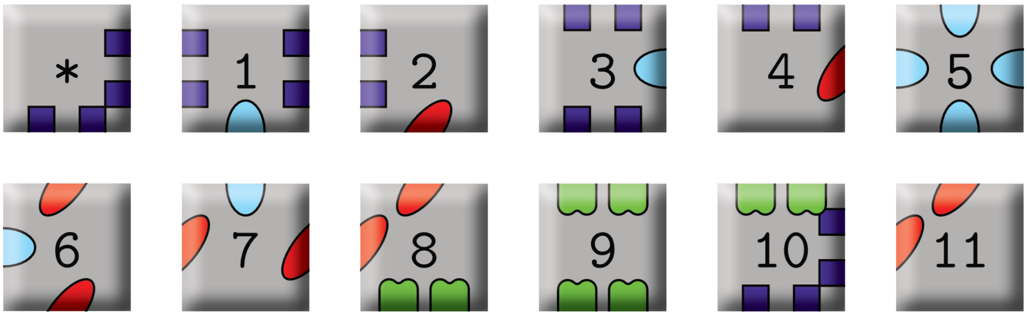


Answer. ▷ The final shapes are exactly the pyramides of height n and basis $2n-1$ with $n \geqslant 2$. The two competing tiles are 1 and 4; it's when 4 arrives that the construction stops. Here is a generic shape with the local order of assembly:



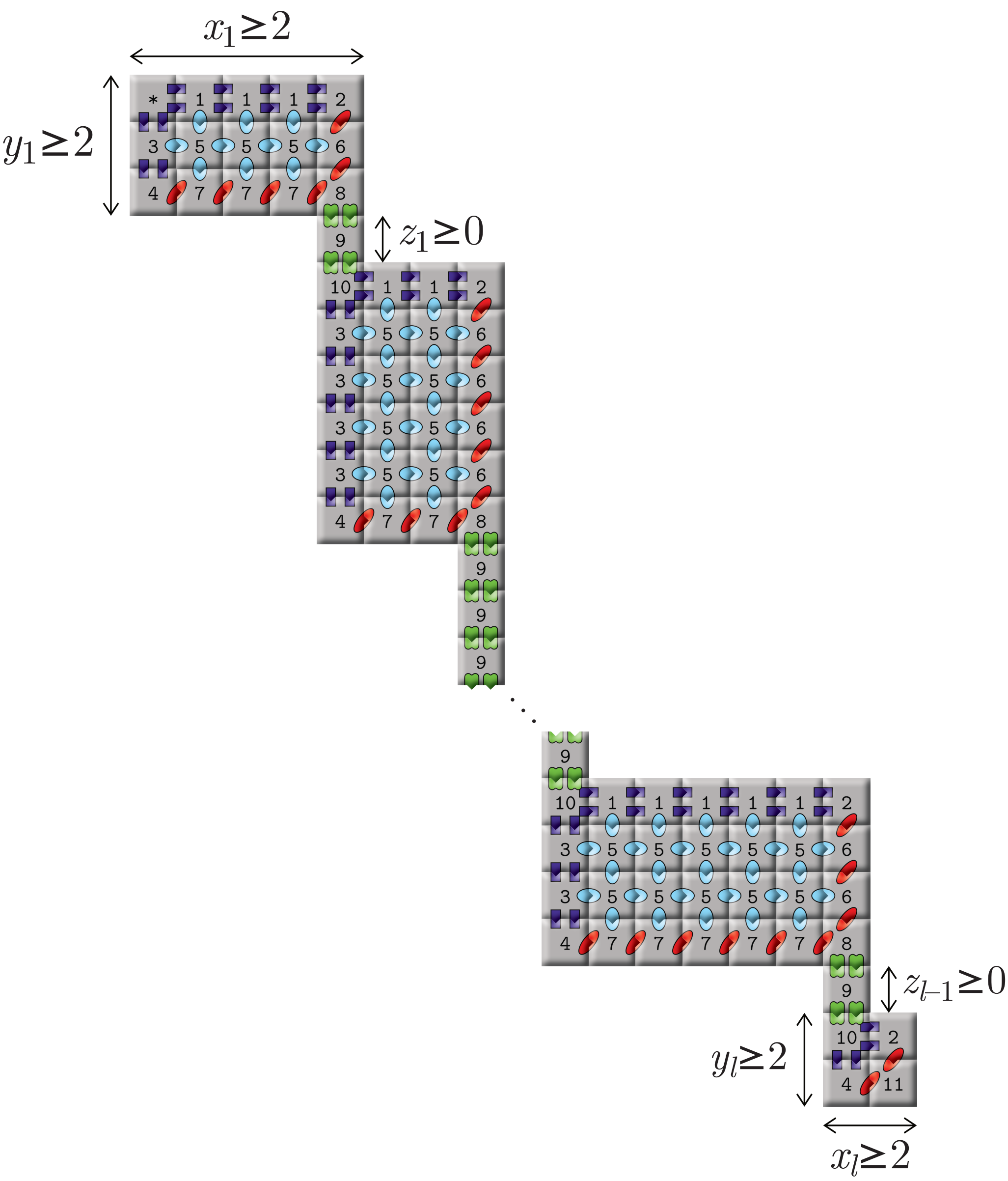
◁

[★] Exercise 2. What is the family of shapes built by this tileset at temperature $T^\circ = 2$? *(no justification asked)*

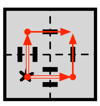


Indicate the assembly order with arrows[1] **on a generic production. Is this a well-ordered tileset?**

Answer. ▷ The tiles set assembles the following family of shapes in the order indicated by the arrows: $\ell \geqslant 1$ rectangles $x_i \times y_i$ with $x_1, y_1, \ldots, x_\ell, y_\ell \geqslant 2$ glued together by a bar $1 \times z_i$ with $z_1, \ldots, z_{\ell-1} \geqslant 0$ gluing the southeast to the northwest corner of each rectangle.



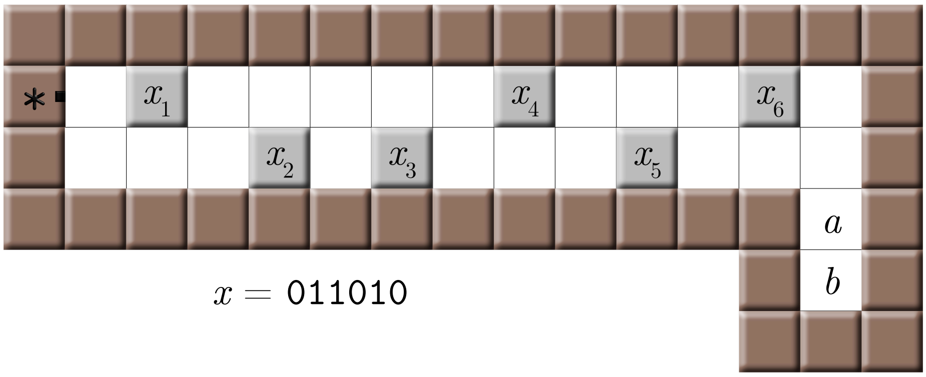
---

[1] Draw an arrow $(i,j) \rightarrow (i',j')$ iff tile (i,j) is attachedbefore tile (i',j'): for instance, 

◁

■ **Exercise 3 (Bit-cave).** We want to design a tileset that computes at temperature $T^\circ = 1$ the parity of a number x written in binary $x_1 \ldots x_n$: parity$(x) = (x_1 + \cdots + x_n) \mod 2$. Remember that *glue mismatch* are allowed: a tile can attach to the agregate as soon as there is at least on side of the tile that agrees with a side of the current agregate (even if the other sides do not agree).

Consider the "cave" seed (in brown below) where the i-th bit of x is encoded in the $(2i+1)$-th column by a (grey) tile placed either on the top row if $x_i = 0$ or in the bottom row if $x_i = 1$.

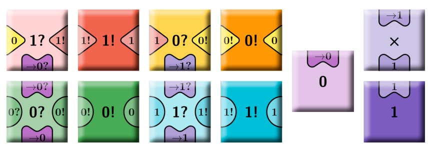


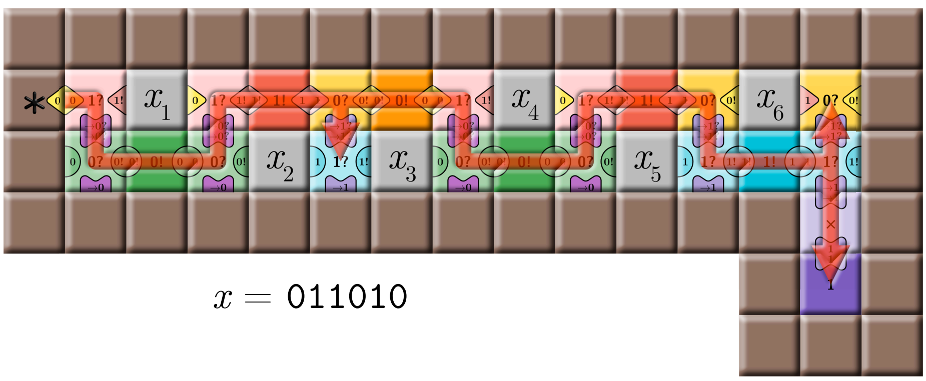
▶ **Question 3.1)** *Give a (absolute constant size) tileset that at temperature $T^\circ = 1$, in presence of the seed above:*

- *places a tile at position a and none at position b if and only if* parity$(x) = 0$*; and*
- *places a tile at position b if and only if* parity$(x) = 1$.

Give a generic assembly (run your tileset on the example above) indicating the order of assembly. How many tile types do you have? No mathematical justification asked however a simple explanation of the role of each tile type is necessary.

Answer. ▷ The figure below gives a solution with 11 tiles. Be aware that we must be extra careful when assigning glues at $T^\circ = 1$ as everything that can stick will stick. The principle of this tileset is to send two signals: one that tries the parity+1 going in the top row and one that tries the parity+0 going in the bottom row. If the bottom signal passes iff the next bit is 0, the top signal passes iff the next bit is 1, so that the parity is correctly computed. The last step is accomplished by trying to send a (purple) signal downwards to output the answer.





$\triangleleft$

■ **Exercise 4.** Assume a random Poisson model where the random time X between two consecutive appearances of a tile of a given type τ at a given empty location follows an exponential law: $p(x) = c \cdot e^{-cx}$ where $c > 1$ is the concentration of the tiles of type τ. We want to prove the following theorem:

Theorem 1 (Adleman et al, 2001). *Consider an ordered tile system $\mathcal{T}$ that assembles deterministically a single shape S. Let $\prec$ be the partial order of the assembly, i.e. such that $(i,j) \prec (k,l)$ if the tile at position (i,j) is attached before the tile at (k,l) by $\mathcal{T}$. With very high probably, the assembly time of a shape S by $\mathcal{T}$ is:*

$$O(\gamma \times \mathsf{rank}(S))$$

where γ only depends on the concentrations and $\mathsf{rank}(S)$ is the highest rank in the shape S (i.e. the length of the longest path in $\prec$).

▶ **Question 4.1)** *Let X be an exponential random variable such that $p(X = x) = ce^{-cx}$ for all real $x \geqslant 0$, for some $c > 0$. Show that X is memoryless, i.e. for all $u, t \geqslant 0$,*

$$p(X = t + u | X \geqslant u) = p(X = t)$$

Answer. $\triangleright$ $p(X = t + u|X \geqslant u) = \frac{p(X=t+u \wedge X \geqslant u)}{\Pr\{X \geqslant u\}} = \frac{p(X=t+u)}{\Pr\{X \geqslant u\}} = \frac{ce^{-c(t+u)}}{\int_u^\infty ce^{-cx}dx} = \frac{ce^{-c(t+u)}}{e^{-cu}} = ce^{-ct}$. $\triangleleft$

Let T be the assembly time of the shape S, i.e. the time at which the last tile of shape S is attached. Let $X_{i,j}$ be the independant exponential random variable for the time between two consecutive appearances of the tile to be attached at position (i,j) in S. We denote by $w(P)$ the random variable for the weight of a $\prec$-path P, defined as: $w(P) = \sum_{(i,j)\in P} X_{i,j}$.

▶ **Question 4.2)** *Show that:*

$$T = \max_{\prec\text{-path } P} w(P)$$

$\triangleright$ *Hint. Proceed by recurrence on the rank of the tiles and show that for all tile (i,j), its assembly time is the random variable $T_{ij} = \max_{\prec\text{-path } P \text{ from } (0,0) \text{ to } (i,j)} w(P)$.*

Answer. $\triangleright$ We proceed by recurrence and show that for all tile t of rank r attaching at position (i,j), the time at which it attaches is

$$T_{i,j} = \max_{\prec\text{-path } P \text{ from } (0,0) \text{ to } (i,j)} w(P) \quad (*)$$

Indeed, for the seed $T_{0,0} = 0 = \max_{\prec\text{-path } P \text{ from } (0,0) \text{ to } (0,0)} w(P)$. Assume now $(*)$ for all tiles with rank $< r$. Consider a tile t of rank r attaching at position (i,j). As $\mathcal{T}$ is ordered, t will attached only if its predecessors are already present, i.e. when it shows up for the first time after the attachement of the last of its predecessors. As the exponential variable for the time between two consecutive appearance of tile t is independent and memoryless, it is identically distributed as $X_{i,j}$. It follows that:

$$\begin{aligned} T_{i,j} &= X_{i,j} + \max_{(k,l)\prec(i,j)} T_{k,l} \\ &= X_{i,j} + \max_{(k,l)\prec(i,j)} \left(\max_{\prec\text{-path } P_{k,l} \text{ from } (0,0) \text{ to } (k,l)} w(P_{k,l}) \right) \\ &= \max_{\prec\text{-path } P \text{ from } (0,0) \text{ to } (i,j)} w(P) \end{aligned}$$

as the rank of any $(k,l) \prec (i,j)$ is $< r$. The result follows. $\triangleleft$

▶ **Question 4.3)** *Let $X_1, \ldots, X_\ell$ be ℓ independent exponential variables s.t. $p(X_i = x) = c_i e^{-c_i x}$ with $c_i > 1$. Show that there is γ which depends only of $\min_i c_i$ such that: for all $n \geqslant \ell$,*

$$\Pr\{X_1 + \cdots + \ell \geqslant \gamma \cdot n\} \leqslant 1/4^\ell \cdot e^{-\gamma(n-\ell)}$$

▷ *Hint. Note that $\mathbb{E}[e^{X_i}] < \infty$ and apply Markov inequality to $Z = e^{X_1 + \cdots + X_\ell}$.*

Answer. ▷ First, $\mathbb{E}[e^{X_i}] = \int_0^\infty c_i e^{-c_i x} e^x dx = \int_0^\infty c_i e^{(1-c_i)x} dx = \frac{c_i}{c_i - 1} < \infty$ as $c_i > 1$. Now,

$$\begin{aligned}
\Pr\{X_1 + \cdots + X_\ell \geqslant \gamma \cdot n\} &= \Pr\{e^{(X_1 + \cdots + X_\ell)} \geqslant e^{\gamma \cdot n}\} \\
&\underset{\text{(Markov)}}{\leqslant} \frac{\mathbb{E}[e^{(X_1 + \cdots + X_\ell)}]}{e^{\gamma n}} \underset{\text{(Independence of the } X_i\text{s)}}{=} \frac{\prod_i \mathbb{E}[e_i^X]}{e^{\gamma n}} \\
&\leqslant \left(\frac{c_{\min}}{e^\gamma (c_{\min} - 1)}\right)^\ell e^{-\gamma(n-\ell)} \leqslant 1/4^\ell \cdot e^{-\gamma(n-\ell)}
\end{aligned}$$

for any $\gamma \geqslant \ln(\frac{4c_{\min}}{c_{\min}-1})$. ◁

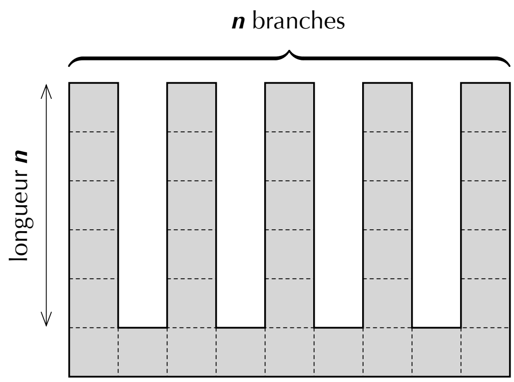
▶ **Question 4.4)** *Conclude.*

Answer. ▷ Let $n = \mathrm{rank}(S)$. Consider a $\prec$-path P, P has length at most n. By question 4.3, we have: $\Pr\{w(P) = \sum_{(i,j)\in P} X_{i,j} \geqslant \gamma \cdot n\} \leqslant 1/4^{|P|} e^{-\gamma(n-|P|)}$. Thus, by the union bound, and as there are at most 3^ℓ $\prec$-paths of length ℓ:

$$\begin{aligned}
\Pr\{T \geqslant \gamma \cdot n\} = \Pr\{\exists P, w(P) \geqslant \gamma \cdot n\} &\leqslant \sum_{\prec\text{-path } P} \Pr\{w(P) \geqslant \gamma \cdot n\} \\
&\leqslant \sum_{\ell=0}^{n} 3^\ell / 4^\ell e^{-\gamma(n-\ell)} = e^{-\gamma n} \sum_{\ell=0}^{n} (3e^\gamma/4)^\ell \\
&= e^{-\gamma n} \cdot \frac{(3e^\gamma/4)^{n+1} - 1}{3e^\gamma/4 - 1} = O((3/4)^n)
\end{aligned}$$

It follows that T is at most $O(\mathrm{rank}(S))$ with very high probability. ◁

■ **Exercise 5.** Propose a staged assembly scheme at temperature $T^\circ = 1$ of the shape family E of candelabrums with n branches of length n.



Describe the tiles, glues, their number, the number of stages and the number of different bechers needed. Give an illustration of the stages to build a generic production.

Answer. ▷ We proceed by cutting recursively the candelabrum (Fig. **??**) into pieces ot size 2^i, chosing two glues among Red, Green, Blue in a cyclic manner for the junctions so that the intermediate pieces do not stick together. We obtain 7 types of branches:

the straight leftmost branche "I", and the 6 L-shaped branches corresponding to the 6 possible pairs of distinct glues picked among Rouge, Vert, Bleu.

We then cut again each branch recursively into pieces of size 2^i using three glues Orange, Cyan, Rose cycling at the junctions, so that the branches do not tick to eachtoher when put together to assemble each branch. We then add the missing tile of the "L" using a Maroon glue.

The assembly proceeds at $T^\circ = 1$ in $O(\log n)$ steps using $O(\log n)$ bechers: $O(\log n)$ bechers for storing each length for each combination of colors; and $O(\log n)$ for the depth-first travel of the decomposition tree.

◁

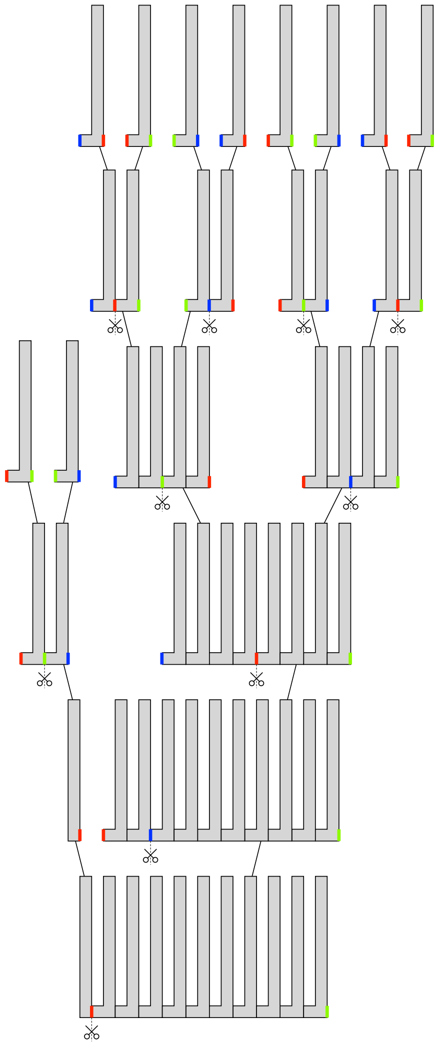


Figure 1: Cut and assembly of the candelabrum.

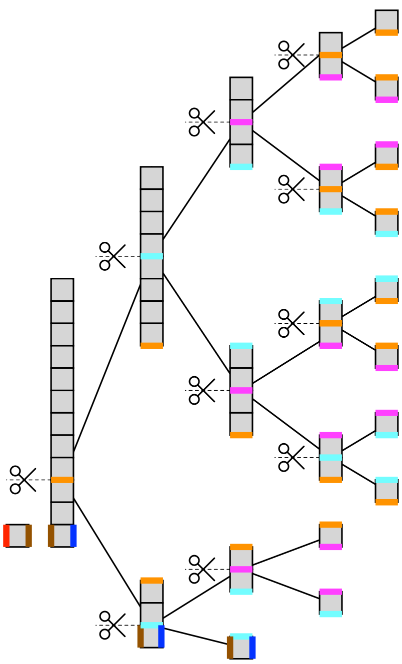


Figure 2: Cut and assembly of one of the seven possible branches.