ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume **23**, Number 3, 2020, 311–329

Probabilistic modeling of the self-assembly of the 1-dimensional DNA structures

Alex AMARIOAREI , Gefry BARAD , Eugen CZEIZLER* , Andrei PAUN , and Romica TRANDAFIR

National Institute of Research and Development for Biological Sciences, Department of Bioinfonnatics, Bucharest, Romania
*Computational Biomodeling Laboratory and Department of Computer Science, Abo Akademi University, Turku, Finland
E-mail: gefry.barad@incdsb.ro; gbarad@gmail.com

Abstract. In a recent paper, using one of the algorithmic assembly formalisms of DNA nanotechnology, we proved that one tile can self-assemble length n structures and $n \times n$ squares, which are basic shapes in the study of DNA origami. This new result within a classic Tile Assembly Model (TAM) would not have been possible without the following programming topics: how can we simulate one-dimensional staged self-assembly using the signal-passing TAM, and how can we program staged self-assembly using the available software? We provide probabilistic approaches for investigating the assembly of tile-based one-dimensional structures. We obtain a probabilistic proof of Han's hook length formula in Enumerative Combinatorics. We identify algebraic and combinatorial structures underlying these algorithmic and information theory results.

Key-words: natural computing; self-assembly; signal and staged TAM; probabilistic models; hook length formulas; operads and Hopf Algebras

1. Introduction

Wang tiles (1961) [1] were initially used to study aperiodic covering of the plane: square tiles with colored edges inserted in the integer lattice only when adjacent tiles have the same color. Inspired by Wang tiles (this set tiles the plane), the abstract Tile Assembly Model was developed in [2] and has been used to study the Kolmogorov complexity of DNA structures i.e. in simple cases, finding the smallest number of tiles required to deterministically self-assemble a given shape. We informally review some definitions [3], [4] of the staged self-assembly [7] and the hierarchical TAM or 2HAM [5]. Finally our approach produces pre-experimental insight required by 'in silico' programming and simulations, as well as the proofs of two mathematical results.

We analyze the assembly and dis-assembly probabilities for one-dimensional tile structures, for a given assembly path. We examine whether there exist properly defined probabilistic models which would ensure that the final assembly graph has a specific probability.

Self-assembly is a process in which a disordered system of pre-existing components forms an organized structure as a consequence of local interactions among the components themselves. Several capsomeres form a virus capsid. We can say that the virus is the output of a biological computation. In 1994 Adleman solved a particular instance of the Hamiltonian Path Problem using DNA tiles [6], which was a keystone in DNA computing.

Several structural properties of the DNA (DeoxyriboNucleic Acid) molecule, for example the pairing principle called the Watson-Crick complementarity or the double stranded geometric shape have been theoretically speculated in the new field of DNA computing (officially born in 1994). The abstract models of computations have been inspired by tiles made out of several DNA helices.

1.1. Models of tile self-assemby: the abstract TAM, hierarchical and the signal-passing self-assembly models

A tile assembly system (TAS) associated with an abstract Tile assembly model (aTAM) is a triple (T, σ, t). T is a finite set of tile types, σ is a finite seed assembly and t is the temperature. A tile type is a 1×1 square with edges $\{N, S, W, E\} = e$. There is a function $g : e \to A$ (a finite set of symbols) which specifies the glue of each edge. An assembly sequence is a sequence (α_i) in which $\alpha_0 = \sigma$ and each α_{i+1} is obtained from α_i by a t-stable addition of a single tile. The system has a glue strength function $h : A \times A \to \mathbb{R}_+$. An addition is t-stable if the sum of the glue values $h(u, v)$ involved in the addition is $\geq t$, where u and v are the glues of 2 adjacent edges, one from α_i and the second one from the newly added tile. The result of an assembly sequence is $res(\overline{\alpha}) = \cup dom\,(\alpha_i)$. $A[T]$ is the set of all producible assemblies of T.

The hierarchical TAM (HTAM or 2HAM) is a generalization of the aTAM: 2 assemblies consisting of one or more tiles can attach to each other as long as t-stability is preserved along the entire newly formed structure. There is a discrete time evolution parameter (u). At time (u), one or more t-stable supertiles are created. Then, a supertile is inserted into the plane domain of the terminal assembly. There is a sequence of sets of supertile additions and a massive parallel creation of supertiles.

1.2. The staged self-assembly model

The staged self-assembly of Wang tiles was formalized in [7]. It is used in biotechnology and encompass the researcher's ability to perform actions as part of a computation. So it is a kind of 3-dimensional programming. In this model, the assembly process is allowed to take place in different, separated test-tubes (or bins) in parallel. At time (u) the content of two or more test tubes are mixed in a separate bin. There is also a filtration procedure: a number of products created in a bin are selected for a new stage. After several stages (HTAM in each test-tube) the

terminal product is assembled in a specific bin. The input of the model is the classical input of a hierarchical tile assembly model: a temperature, a set of tile types placed in a marked initial test tubes, a glue strength function, and a directed acyclic assembly graph G which specify the order of the stages, the initial assemblies of each bin and the final products to be used for the next stages. The nodes of the assembly graph G are the bins. The oriented edges show that some final products from a bin are used as initial products of a reaction taken place in a different bin. Any planar shape can be assembled using a staged-self assembly with 16 tile types.

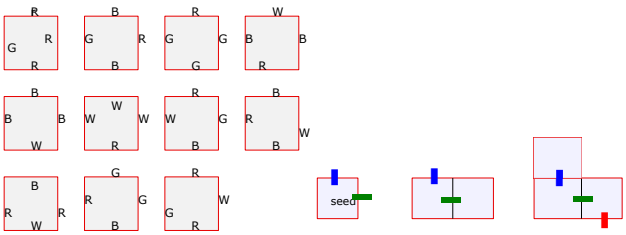

Fig. 1. 11 four-coloured Wang tiles described in Jeandel and Rao (2015), which can tile the plane aperiodically; a self assembly of several tiles

The output of a self-assembly software. We can simulate the 2-dimensional self-assembly of square-shaped tiles using the following software tools: BioNetGen a software package for rule-based modeling of biochemical systems; Xgrow (2003), ISU TAS (2009)- Tile Assembly Model simulators written in C for a Windows environment. A natural problem is to simulate staged self-assembly using a classical one pot self-assembly.

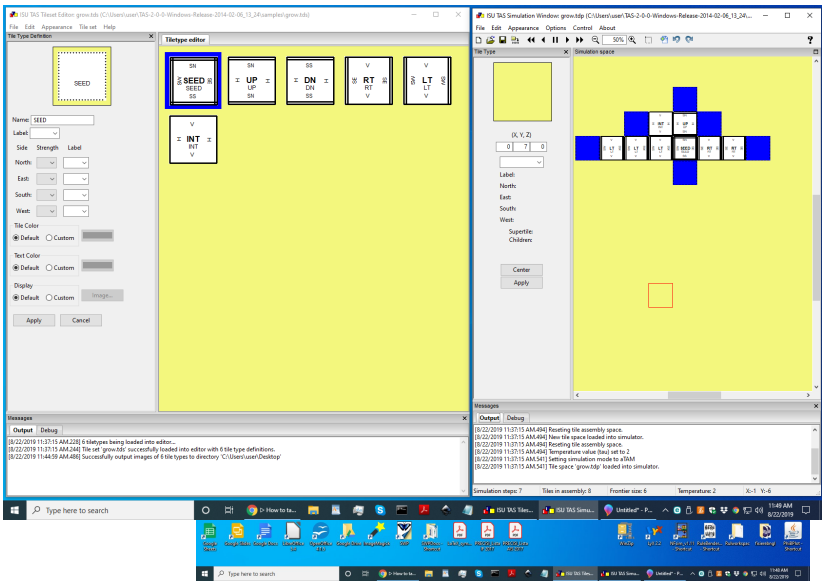

Fig. 2. Screenshot of an ISU-TAS session: 6 given tiles; simulation assembly after 7 steps; possible places to add a new tile - in blue; the next simulation step will choose one of these blue squares with equal probabilities if the temperature is 1

In the next screenshot, there are 4 tiles which are able to generate only one final 1-dimensional structure. Using the probabilistic models from the next sections, we will be able to precisely answer the following question: what is the probability of a given planar binary tree to be the assembly template for a 1-dim structure as above, after a huge number of simulations. The ISU-TAS software visualize the discrete computational steps: the way tiles form super-tiles converging to the final structure.

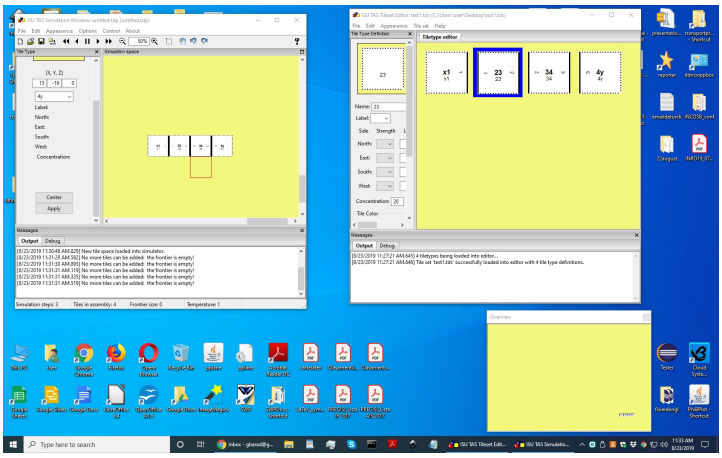

Fig. 3. Compared to the last model which evolves from a seeded, marked tile, the HTAM does not have an initial seed as a starting point. The model is deterministic: we know there is only one final structure, a length n bar and multiple ways to form it, using smaller super-tiles.

2. The formation of a linear structure

We consider the following 1-dimensional hierarchical TAM; the map g specifies the glues.

There are n types of square tiles. There are $n-1$ glues. The tile denoted $(i, i+1)$ has $g(W) = i-1$ and $g(E) = i$, i from 2 to $n-1$. There are 2 additional tile types: the left terminal one given by $g(E) = 1$ and the right terminal one give by $g(W) = n-1$. Other edges do not have glues (they have the empty glue) so the system is 1-dimensional. The glue strength function is $h(i, i) = 1$ and the temperature is 1. After $n-1$ binding events the terminal assembly is a bar of length n. Random Parallel (HTAM) bindings in at most $n-1$ units of time are necessary to build the length n-bar. We are interested is the path, the mixing graph which describe the way the bar of length n is formed. For example, the figure below describes a planar binary tree traced by n tiles.

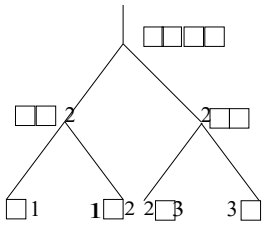

Fig. 4. a planar binary tree as a mixing graph (evolution scenario) of the 1-dim HAM; identical glues labeled by numbers allow binding; there are 2 stages.

The configuration space is given by the set of planar binary trees with n leaves, counted by the Catalan number $\frac{1}{n+1}C_{2n}^{n}$. A planar binary tree is an acyclic graph with $2n-1$ vertices (n univalent and $n-1$ trivalent vertices) , and a marked univalent vertex called the root. The graph is embedded in the plane – it has a specific planar representation. The bin (reactor) can have N tiles of each type and N bars of length n as the final product, each having a possible different planar binary tree as the mixing graph. A natural probabilistic problem is to ask about the distribution of the paths above when N goes to infinity.

The following result was proved in [8]:

Theorem (*Theorem 3 in [8]*). *Given a planar binary tree of dimension n, there is a STAM using only one tile type and having $O(n)$ glues and signal-passing capabilities, which assemble a length n bar using the given binary tree as the mixing graph.*

This theorem is important for the following reason. STAM is the abbreviation for *signal-passing tile assembly model* and it was defined by Padilla et. al. [9]. It is an self-assembly abstract model (it could be hierarchical or seeded) where after a binding between two connected components, some glues from these components can change its state: they can become active, or in state off, or some glues can change its strength. Based on some signals activated at the binding non-local effects can appear. Every tile has information channels, like electrical wires which allow the transmision of these commands to some glues according to the predefined rules.

The theorem states that we can control the path of formation of a length n bar, if we enhance the tiles with a specific system of channels which control the states of the glues. For a population of N tiles, we can design a percent $p(t)$ of tiles which evolve towards the final structure using a predefined path t. In this way, we can model using self-assembly any possible probabilities of formation paths, at least in the 1-dimensional case: knowing that we will have $\frac{N}{n}$ final products, length n bars, we can enhance from the beginning the tiles with a signal-passing system which will realize these desired concentrations. So, the next probabilistic models can be simulated using algorithmic self-assembly.

For example, the tile which self-assemble a length 4 bar, using the tree of Figure 4 as the formation graph has the following description in the STAM formalism.

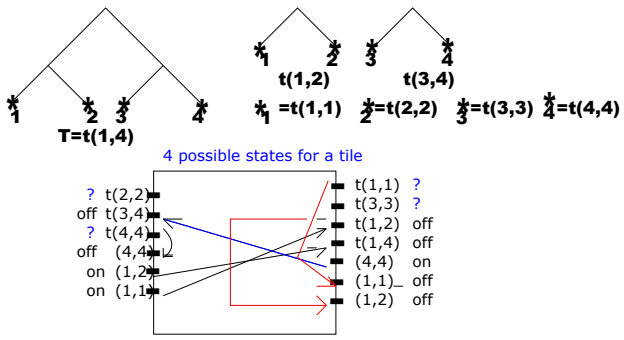

Fig. 5. a planar binary tree as a mixing graph (evolution scenario) of the 1-dim HAM; identical glues labeled by numbers allow binding; there are 2 stages.

Initially, a signal-passing tile can have one of the 4 possible states: only one glue of type t(x,x) is on and the others are off; the initial state of the glues of type (a,b) is fixed; active identical glues of type (a,b) can bind; active complementary tree glues can bind pairs of complementary trees: t(1,1) and t(2,2); t(3,3) and t(4,4); t(1,2) and t(3,4)

3. Probabilistic models and parameters of programming. Combinatorial aspects

3.1. A top-down approach

We consider a bar of length n, formed using n tiles indexed $1, 2, \ldots, n$. We can imagine a terminal self-assembly product to begin to disassemble as a result of temperature variations. The input: the parameters $(d_{i,j})$, where $0 < i < i + j \leq n$ which satisfy $d_{i,j} > 0$ and $\sum_{\alpha=1}^{m} d_{\alpha,m+1-\alpha} = 1$, $\forall\, 1 \leq m \leq n-1$.

$(d_{i,j})$ is the probability of a bar of length $i + j$ to be separated in two bars of lengths i and j (the glue strength between the tiles i and $i + 1$ is below t). We also denote $(d_{i,j})$ by $d(i, j)$ if there is a risk of confusion.

In HTAM, after $n - 1$ separation-events, we will have n elementary tiles. We also have as observable the planar binary tree T which describes the parallel process of disassembly (i.e. the creation of connected components). We try to define a numerical function defined on the set of planar binary trees (notation: P_T or $P(T)$), as the probability of formation of T.

If $T = T_1 \vee T_2$, then $P(T) = d\left(|T_1|, |T_2|\right) P\left(T_1\right) P\left(T_2\right)$ because of the parallel nature of HTAM. $|T|$:= the number of its univalent vertices=the number of tiles.

We also have $\sum_{T \in PBT(n)} P(T) = 1$, where $PBT(n)$ = the set of planar binary trees with n

leaves (n univalent vertices). We define $S_m = \sum_{T\in PBT(m)} P(T)$

$$\sum_{T\in PBT(n)} P(T) = \sum_{k=1}^{n-1} \sum_{\substack{T_1\in PBT(k)\\ T=T_1\vee T_2}} d(k,n-k)P\left(T_1\right)P\left(T_2\right) =$$

$$= \sum_{k=1}^{n-1} \sum_{T_1\in PBT(k)} d(k,n-k)P\left(T_1\right)S_{n-k} = \sum_{k=1}^{n-1} d(k,n-k)S_kS_{n-k}$$

An induction argument implies S_m=1 for any m, $1 \leq m \leq n$. We try to choose $d(i,j)$ symmetric in some sense: $d(i,j) = d(j,i)$ such that if the trees T_1 and T_2 are isomorphic as (spatial) graphs, then $P\left(T_1\right) = P\left(T_2\right)$. A possible choice is given by n numbers, $a_1, a_2, \ldots, a_n$ such that

$$d(i,j) = \frac{a_ia_j}{\sum_{k=1}^{i+j-1} a_ka_{i+j-k}}$$

.

The setting above allow us to give a probabilistic proof of an important result of Han [10] on Enumerative Combinatorics. We define $PBT = \bigcup_{n\geq 1} PBT(n)$. The following equalities are called hook length formulas for planar binary trees [10].

$$\sum_{T\in PBT} x^{|T|} \prod_{v\in V_T} \frac{1}{h_v 2^{h_v-1}} = e^x$$

$$\sum_{T\in PBT} X^{|T|} \prod_{v\in V_T} \left(1+\frac{1}{h_v}\right) = \sum_{n\geq 0} (n+1)^{n-1}\frac{(2X)^n}{n!}$$

V_T = the set of internal trivalent vertices of the planar binary tree T, including its root.

h_v = the number of internal trivalent vertices (# leaves-1) of the sub-tree of T whose root is v.

For example, the tree from Figure 4 has the following hook length numbers: 3,1,1. Han provides a general procedure to find these types of formulas. Given $\rho : \mathbb{N} \to \mathbb{R}$, called a hook-function, define the series $f(X)$. PBT is the set of all planar binary trees.

$$\sum_{T\in PBT} X^{|T|} \prod_{v\in V_T} \rho\left(h_v\right) = 1 + f_1X + f_2X^2 + \ldots + f_nX^n + \ldots$$

Knowing $\rho(X) = \sum_k \rho(k)X^k$, we have $f(X)$. We also have the following theorem of Han ([10] Formula 6.3 of Sect. 6): $\rho(n) = \frac{f_n}{\sum_{k=0}^{n-1} f_kf_{n-k-1}} = \frac{[X^n]f(X)}{[X^{n-1}]f^2(X)}$,

where $[X^m]\, f(X)$ is the coefficient of X^m of the series f. $f_0 = 1$. This general inversion formula was generalized for other classes of trees and for Feynman rules of algebraic quantum field theory in [11].

We provide the following probabilistic proof of Han's formula using the framework of this section. First of all, we define a new sequence of numbers. $g_{k+1} = f_k$ for any $0 \leq k$. The series f is given.

We consider a HTAM which disassemble bars, with probabilities $d(i,j) = \dfrac{g_i g_j}{\sum\limits_{\substack{\alpha+\beta=i+j \\ \alpha,\beta\geq 1}} g_\alpha g_\beta}$.

We define $\rho(n) = \dfrac{g_{n+1}}{\sum\limits_{\substack{\alpha+\beta=i+j=n+1 \\ \alpha,\beta\geq 1}} g_\alpha g_\beta} = d(i,j) \cdot \dfrac{g_{n+1}}{g_i g_j} \Rightarrow d(i,j) = \rho(n) \cdot \frac{g_i g_j}{g_{n+1}}$.

Using the recursive definition of the probability of a tree $T = T_1 \vee T_2$,
$P(T) = d\left(|T_1|, |T_2|\right) P\left(T_1\right) P\left(T_2\right) \Rightarrow P(T) = \prod\limits_{v \in V_T} d_v(i,j) = \prod\limits_{v \in V_T} \rho\left(h_v\right) \cdot \dfrac{g_{(|T_v^1|)} g_{(|T_v^2|)}}{g_{(|T_v|)}}$,
where $T_v = T_v^1 \vee T_v^2$ is the sub-tree of T whose root is v. The product of ρ's is exactly the quantity which enters in the Han's formula. The product of the fractions $\dfrac{g_a g_b}{g_c}$ is equal to $\dfrac{1}{g_{n+1}} = \dfrac{1}{f_n}$ for the following reason: for any vertex v which is not the root, the product $g_a g_b$ appears as numerator. But separately, g_a and g_b appear at denominators when we consider the fractions associated with the vertices v_1 and v_2. g_c appears at numerator, coloring the edge from a vertex towards its left or a right sub-tree.

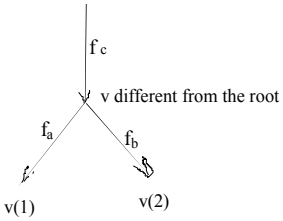

Fig. 6. To compute the probability of a tree using the product of $d(i,j)$, we color the vertices using ρ; the edges of the tree are colored by $g_{|t|}$

The sum of the probabilities $\sum\limits_{T \in PBT(n)} P(T) = \sum\limits_{T \in PBT(n)} \dfrac{1}{f_n} \prod\limits_{v \in V_T} \rho\left(h_v\right) = 1\ ,\ \Rightarrow f_n = \sum\limits_{T \in PBT(n)} \cdot \prod\limits_{v \in V_T} \rho\left(h_v\right) = g_{n+1}$. So ρ, as defined above using the series f and used as a hook length function, generates the series f.

$$\rho(m) = \frac{f(m)}{f(0)f(m-1) + f(1)f(m-2) + \ldots + f(m-1)f(0)}$$

3.2. Relations among probabilities

Given the numbers $(P_T)_{T \in PBT(n)}$, is there a probability model P_d defined by $(d_{i,j})_{1 \leq i < i+j \leq n}$ such that $P_T = P(T)$?

The numbers $(d_{i,j})$ define a probability distribution on $PBT(m)$ for any $1 \leq m \leq n$. We denote $T = T_1 \vee T_2 = (T_1, T_2)$ the decomposition of a $T \in PBT$ into its left and right subtrees. $P\left(T_1 \vee T_2\right) = d(i,j) P\left(T_1\right) P\left(T_2\right)$ implies the following relations.

$|T|$=the number of the leaves (tiles). The letters $i, j, m, n, ..$ below denote trees of different dimensions

$$P((i,r))P((j,s)) = P((j,r))P((i,s)) \qquad \text{if } |i| = |j| \text{ and } |r| = |s|$$

$$P((m,(i,r)))P((n,(j,s))) = P((m,(i,s)))P((n,(j,r)) \qquad \text{if } |m| = |n|$$

$$P((i,r),m)P((j,s),n) = P((i,s),m)P((j,r),n).$$

These relations are not sufficient to determine $d(i,j)$. A relation is given by $P(A)P(B) = P(C)P(D)$, where $A, B, C, D \in PBT(n)$. We say that a relation is of level k if the trees A, B, C and D contain k pairs of parentheses, as above. If two relations of degrees k and l are written symbolically as $P(a)P(b) = P(c)P(d)$ and $P(x)P(y) = P(z)P(t)$, then the relation of degree $k + l + 1$ is written:

$$P((a,x))P((b.y)) = P((c,z))P((d,t)).$$

This exponential number of relations $P(A)P(B) = P(C)P(D)$ can be described recursively as above: for a given relation, there is a parenthesization of variables which are named by trees of lower dimension, such that the left and the right trees of this parenthesization are lower dimension relations.

Proposition 1. *Given the numbers $(P_T)_{T \in PBT(n)}$ with $\sum P_T = 1$ satisfying all the recursive relations above, up to level n, there are unique numbers $(d_{i,j})$ such that in the HTAM probabilistic model P($(d_{i,j})$), the probability of any trees is P_T.*

Proof. We use induction over n. For any k, let a_{n-k} be a fixed tree of dimension $n - k$. Define a probability distribution on $PBT(k)$, for any k, $1 \leq k \leq n$ given by

$$\frac{P(t_1)}{P(t_2)} = \frac{P(t_1 \vee a_{n-k})}{P(t_2 \vee a_{n-k})}, \qquad \forall\, t_i \in PBT(k).$$

We can find $P(t_i)$ knowing their sum $= 1$ and their relative proportions.
The relations are satisfied for the trees in $PBT(k)$, for any k, $1 \leq k \leq n-1$, so these numbers P_T are probabilities built using $d_{i,j}$, for $i + j \leq n - 1$. If $i + j = n$, define $d_{i,j} = \dfrac{P(a \vee b)}{P(a)P(b)}$ if $|a| = i$ and $|b| = j$. The way lower degree probabilities were defined and the relations imply $(d_{i,j})$ are well-defined and $\sum_{i=1}^{m-1} d_{i,m-i} = 1$.

Remark. Given a tree $T \in PBT(n)$, $P(T) = \prod_{k=2}^{n} \prod_{p=1}^{\alpha_k} d(i_p, j_p)^{\beta_p}$, if $i_p + j_p = k$.
So, for every k there are α_k sub-trees of T of dimension k.

If $\alpha_k \leq 1$, $\forall\, k$, then $\forall\, \varepsilon \in (0,1)$, there are parameters $d_{i,j}$ such that $P(T) > 1 - \varepsilon$. We can increase the $P(T)$ as close as we want to 1. So, we can simulate a staged self-disassembly using T.

If there is k such that $\alpha_k > 1$, then $P(T) < \dfrac{1}{2}$ for any choice of the parameters. (A consequence of the Hölder mean inequality.)

For example, in the tree which disassemble a bar of length 8 in 4 stages: $8- > (4,4)- > ((2,2),(1,3))- > (((1,1),(1,1)),(1,(1,2)))- > 8\ tiles$

$$P(T) = d(4,4)d(1,2)d(2,2)d(1,3) < \frac{1}{2}.$$

In this case we cannot assign T as a privileged path for disassembly. Specific chemical models could require to maximize/minimize $\sum P(T_i) \mid T_i \in A \subset PBT(n)$. In this case some numbers have to vanish; other $d_{i,j}$ have to satisfy a non-linear system of equations by taking derivatives of the P(T)'s. Compactness of the parameter space implies there is always a solution.

The trees $T \in PBT(n)$ for which the following property is true: for any $\epsilon > 0$ we can find a top-down $(d_{i,j})$ such that its probability $P(T) > 1 - \epsilon$ are characterized by the following fact: any internal vertex v is connected with a leaf– an univalent vertex.

In Biosciences particular types of the top-down $(d_{i,j})$ probabilistic models appeared long time ago. The evolutionary relationships among a collection of organisms are displayed in the form of a tree whose leaves are labeled by the organisms. Internal vertices of the tree represent common ancestors. Trees are built using statistical and combinatorial techniques and DNA sequences. Phylogenetic trees are used in epidemiology, ecology and linguistics. Useful models include Yule (or branching Markovian) models where d(i,j) are proportional to $(ij)^z$, to $c(i)c(j)$ where $c(i)$ are Catalan numbers or to C_{i+j}^{i} [12], [13].

4. The bottom-up approach

In this case we are in a classic HTAM framework.

The parameters are n; $\{p(k; l_1, l_2, \ldots, l_m)\}_{\substack{k=1\ldots m-1 \\ m=1\ldots n}}$ positive numbers. The m-uples of strictly positive natural numbers have to satisfy $l_1 + l_2 + \ldots + l_m = n$.

We require $\sum_{k=1}^{m-1} p(k; l_1, l_2, \ldots, l_m) = 1$, for any $l_1, l_2, \ldots, l_m$. Given the n tiles $(i, i+1)$, at each moment we have a set of linearly ordered bars of length $l_1, l_2, \ldots, l_m$. With probability $p(k, l_1, l_2, \ldots, l_m)$ we join the bars of length l_k and l_{k+1}. After $n-1$ bindings we have a bar of length n and the observable: the mixing planar binary tree $T \in PBT(n)$.

We have the following reversibility result, consistent with the visual intuition:

Proposition 2. *Given any previous top-down model defined by $(d_{i,j})_n$, there is a bottom-up model $(P\{p(k,l_i)\})_n$ such that $\forall T \in PBT(n)$, the probabilities of T computed in both probabilistic models are equal.*

4.1. Aspects of the bottom-up model $(P\{p(k,l_i)\})$

We consider the following oriented graph G_n. The vertices are indexed over the interval partitions of n: $(l_1, l_2, \ldots, l_m)$, where $l_1 + l_2 + \ldots + l_m = n$. $l_i \in \mathbb{N}^*$.
An oriented edge indexed $(k; l_1, l_2, \ldots, l_m)$, where $1 \leq k \leq m-1$ will join $(l_1, l_2, \ldots, l_m)$ to $(l_1, \ldots, l_{k-1}, l_k + l_{k+1}, \ldots, l_m)$.

The graph G_n is the poset of the subsets of $\{1, 2, \ldots, n-1\}$ i.e. the boolean lattice, the 1-skeleton of the $n-1$ -dim cube via the following bijection between vertices. $(l_1, l_2, \ldots, l_m) \rightarrow$the characteristic function of $(l_1, l_1 + l_2, \ldots, l_1 + l_2 + \ldots l_{m-1}) \subset \{1, 2, \ldots, n-1\} \rightarrow$ the coordinates of a vertex of the cube $\{0, 1\}^{n-1}$.

We assign the numbers $p\,(k; l_1, l_2, \ldots, l_m)$ to the corresponding edges from G_n. We define the probability (or the weight) of an oriented path of length $n-1$, which join the vertex (1,1...1) to (n) to be the product of the coefficients $p(k; l_i)$ which color the edges of the path.

LEMMA 4..1. *The sum of the weights associated to all* $n-1-$*paths as above is* 1.

Proof. We partition the vertices of G_n into n levels.
$L_{n+1-m} = \{(l_1, l_2, \ldots, l_m) \mid l_1 + l_2 + \ldots + l_m = n\}$.

We define the probability (or the weight) of an oriented path d of length k to be W_d:=the product of the numbers $p\left(k_i; l_1^i, l_2^i, \ldots, l_{\alpha_i}^i\right)$ from its edges $S_m = \sum\limits_d W_d$ where d is an oriented path of length m which start from $(1, 1, \ldots, 1)$.

We will prove by induction that $S_m = S_{m+1}, \forall\, m \in \{1, 2, \ldots, n-2\}$.

In particular $S_{n-1} = S_1 = \sum\limits_{k=1}^{n-1} p(k; 1, 1, \ldots, 1) = 1$.

$$S_{m+1} = \sum_{|d|=m+1} W_d = \sum_{v \in L_{m+1}} \sum_{d:(1,1,\ldots,1)\rightarrow v} W_d =$$

$$= \sum_{v \in L_m} \sum_{d:(1,1,\ldots,1)\rightarrow v} W_d \cdot W_e = \sum_{|d|=m} W_d \left(\sum_e W_e \right) = \sum_{|d|=m} W_d = S_m,$$

where e is the edge between the vertex v and the level L_{m+1}. The sum from parentheses is $1 : \sum\limits_{k=1}^{m-1} p\,(k; l_1, l_2, \ldots, l_m) = 1$. □

Any oriented path d of length $n-1$ in G_n specifies a possible binding path for the HTAM: the order and the position of the $n-1$ bindings. In particular, it defines a planar binary tree T_d for the mixing of the initial n tiles. It defines a function Ψ, called the Loday-Ronco map $\Psi : \{n-1\text{-paths in } G_n\} \rightarrow PBT(n)$ [14].

There are $(n-1)!$ oriented paths of length n-1 in G_n and there are in bijection with the permutations of $\{1, 2, \ldots, n-1\}$. This bijection allow us to define the probability of a permutation and $\Psi : S_{n-1} \rightarrow PBT(n)$ [14].

Any path of length $n-1$ in G_n defines a path of length $n-1$ in the $(n-1)-dim$ cube which join the vertices $(0, 0, \ldots, 0)$ and $(1, 1, \ldots, 1)$. This path in cube is the 1-skeleton of a simplex in $\mathbb{R}^{n-1}$ defined as $\left\{(X_1, X_2, \ldots, X_{n-1}) \mid 0 \leq X_{\sigma(1)} \leq X_{\sigma(2)} \leq \ldots \leq X_{\sigma(n-1)} \leq 1\right\}$ for a specific permutation $\sigma \in S_{n-1}$ [15].

The map $\Psi_n : S_{n-1} \rightarrow PBT(n)$ can be defined recursively $\Psi_n(\sigma(1), \sigma(2), \ldots, \sigma(n-1) = \Psi_{k-1}\left(std(\sigma(1), \sigma(2), \ldots, \sigma(k-1)\right) \vee \Psi_{n-k}\left[(std(\sigma(k+1), \ldots, \sigma(n-1)\right]$

std=the standardization map associates a permutation to any word of distinct letters from a totally ordered set. Example: $std(9, 6, 2) = (3, 2, 1) \qquad std(acb) = 132$

In HTAM a permutation of $\{1, 2, \ldots, n-1\}$, $\sigma(1), \sigma(2), \ldots, \sigma(n-1)$ specifies the space between (super)tiles which will be joined at the moment k.

Example: □□□□ → □□□□ → □□□□ → □□□□ gives the permutation (312).

DEFINITION 4..2. *Let $T \in PBT(n)$. The probability of the tree T in the bottom-up $p(k; l_i)$ model is*
$P(T)= \sum_{\substack{\sigma \in S_{n-1} \\ \Psi(\sigma)=T}} P(\sigma) = \sum_{\substack{d \text{ is n-1-path in } G_n \\ \Psi(d)=T}} P(d).$

Knuth's Theorem [16] counts the number of the terms from the definition above: for any planar binary tree T, there are $\frac{n!}{\prod h_v}$ paths in G_n which realize T. It is the number of permutations which can decorate the trivalent vertices of T increasingly, with respect to the partial order given by the paths from the root of T to its leaves. h_v is the hook-length of a vertex of T.

Remark. Because of the Definition 4.2, $\forall\, \varepsilon \in (0, 1)$, given $T \in PBT(n)$ there is a set of parameters such that $P(T) > 1 - \epsilon$. We can choose a particular path d which realize T and we increase to $1 - \epsilon/n$ all the specific parameters $p(k_j; l_t)$ which color the edges of d. So we can model a staged self-assembly with mixing graph T using this probabilistic model.

Proposition 3. *Given any positive numbers (p_d), indexed over the paths of length n-1 from G_n (or n-1-permutations), $\sum_d p_d = 1$, there is at most one HTAM probabilistic model $\{P(k, l_i)\}$ such that its probabilities satisfy $p(d) = p_d$.*

Proof. We determine recursively the numbers $p\,(k; l_1, l_2, \ldots, l_m)$ for the edges between the levels L_m and L_{m+1}, for $m = 1, 2, \ldots, n-1$. We define for any path d of length k, $W_d = \prod_{i=1}^{k} W_{e_i}$.

For the given edge e decorated by $p(k; 1, 1, \ldots, 1)$, $p(k; 1, 1, \ldots, 1) = \sum_{e \in d} p_d$

$e \in d$ means the n-1 -length path d contains e. It begins with e.

For the edge $e_2 = p(k; 1, 1, \ldots, 2, \ldots, 1)$ between L_2 and L_3 we consider the unique edge e_1 between L_1 and L_2 with its head = the tip of $p(k; \ldots 2, \ldots, 1) = e_2$. Then $W_{e_2} = \frac{\sum_{e_1, e_2 \in d} p_d}{W_{e_1}}$.
Summation is over all paths d which contain the edges e_1 and e_2.

In general for an edge e_k between L_k and L_{k+1}, we consider a particular path $e_1, e_2, \ldots, e_{k-1}$ which join $(1, 1, \ldots, 1)$ to the tip of e_k. Using the same types of summation as in Lemma 4.1 from Section 4.1, we get $W_{e_k} = \frac{\sum_{e_1, e_2 \ldots e_k \in d} p_d}{W_{e_1} W_{e_2} \ldots W_{e_{k-1}}}$.
Summation is over all paths d which contain the edges e_1 ,e_2...e_k.

So, there is at most one system $p(k; l_i)$ which generate (p_d). We have a solution if the formula for W_{e_k} above does not depend on the path $e_1, e_2 \ldots e_{k-1}$.
$W_{e_{k-1}} = \frac{\sum_{e_1, e_2 \ldots e_{k-1} \in d} p_d}{W_{e_1} W_{e_2} \ldots W_{e_{k-2}}} \Rightarrow W_{e_1} W_{e_2} \ldots W_{e_{k-1}} = \sum_{e_1, e_2 \ldots e_{k-1} \in d} p_d \Rightarrow W_{e_k} = \frac{\sum_{e_1, e_2 \ldots e_k \in d} p_d}{\sum_{e_1, e_2 \ldots e_{k-1} \in d} p_d}$

A consequence of these formulas is the following:

Proposition 4. *The numbers $(p_d)_{d \in G_n}$ are the probabilities of a bottom-up probability system $\{p(K, l)\}$ if and only if the following conditions are satisfied: their sum is 1 and*

$\forall\, v \in L_k$, $\forall\, e$ *edge with* $tip(e) = v$; $\forall$ 2 *paths* a *and* b *of length* $k-1$ *between* $(1,1,\ldots,1)$ *and* v,

$$\frac{\sum\limits_{d \ni ae} p_d}{\sum\limits_{d \ni a} p_d} = \frac{\sum\limits_{d \ni be} p_d}{\sum\limits_{d \ni b} p_d}.$$

We say that the numbers (p_d) forms a solvable system. In this case, W_{e_k} does not depend on the path $e_1, \ldots, e_{k-1}$.

Corollary a) If the numbers (p_d) satisfies the following conditions:

$\forall\, v \in L_k$, $\forall\, e, f$ edges with $tip(e) = tip(f) = v$ and

$\forall$ 2 oriented paths of length $k-1$ between $(1,1,\ldots,1)$ and v, called a and b

$$\left(\sum_{d \ni ae} p_d\right) \cdot \left(\sum_{d \ni bf} p_d\right) = \left(\sum_{d \ni be} p_d\right) \left(\sum_{d \ni af} p_d\right),$$

then the system (p_d) is solvable.

Proof. We sum over f and we obtain the general conditions from Proposition 4.

Corollary b) *The numbers* (p_d) *forms a solvable system if for any vertex* v *and any 4 paths* a, b, c, d *which satisfy:* $head(a) = head(b) = tip(c) = tip(d) = v$ $|a| = |b| = n - 1 - |c| = n - 1 - |d|$ *we have the equality* $p_{ac}p_{bd} = p_{ad}p_{bc}$

Proof. Let e be the first edge from d. Let f be first edge from c.

We sum over the paths c with fixed f and over the paths d with fixed e and we apply Corollary a).

The next corollary was announced in Section 4 (Proposition 2).

Corollary c) We consider a top-down probability system $p\,(d_{i,j})_n$, which generate the probabilities $(P_T)_{T \in PBT(n)}$. To each path of length $n-1$, $d \in G_n$, we associate $p_d = \dfrac{P_{\Psi(d)}}{|A_d|}$, Ψ is the Loday-Ronco map, $A_d = \{\text{path } e \mid \Psi(d) = \Psi(e)\}$

.

If $\Psi(d) =$ the tree T, then $|A_d| = \dfrac{n!}{\prod\limits_{v \in V_T} h_v}$ (Knuth's formula).

Then it is not hard to see that the system (p_d) satisfies the conditions from Corollary b), so it is solvable. In particular, there is a bottom-up system $\{p\,(K, l_i)\}$ which generate the same probabilities for trees as the given (d_{ij})-disassembly model.

Corollary c) is a consequence of the Knuth's formula –which equalize the denominators from the relations from Corollary b)–and of the following relation between the probabilities generated by $\{d_{ij}\}$. By definition the tree $T[a, b, c, \ldots]$ is obtained from T by glueing the roots of $a, b, c, \ldots$ to the leaves of T.

We consider T_1 and $T_2 \in PBT(m)$. We consider the trees t_i^1 and $t_i^2 \in PBT(l_i)$, $i = 1, 2, \ldots, m$. We consider the four trees $T_i^j = T_i \left[t_1^j, t_2^j, \ldots, t_n^j\right]$, $i, j \in \{1, 2\}$.
Then $p\left(T_1^1\right) p\left(T_2^2\right) = p\left(T_1^2\right) p\left(T_2^1\right)$

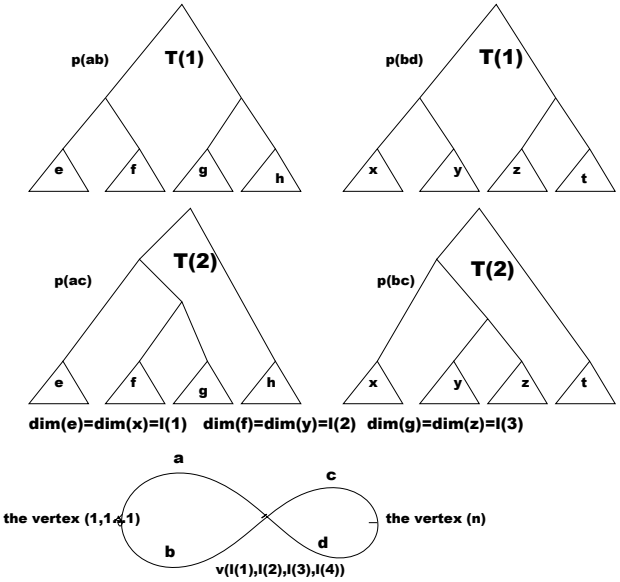

Fig. 7. the four trees involved in the relation p(ad)p(bc)=p(ac)p(bd) of Corollary b)

The probabilities p(e),p(f)...p(z),p(t) appear on the right-hand and in the left hand sides. The contribution of the internal vertices: d(i,j) depend on the structures of the trees T_1 and T_2, and on the numbers l_i.
The products of d(i,j) for the internal vertices of T_1 in p(ad) and p(bd) are the same. The hook lengths of the internal vertices of T_p in T_i^j depend on T_p and on the numbers l_i – not on the specific shapes of the bottom trees e,f,g,h...x,y,z,t.

Proposition 4 and Corollaries a) and b) are valid not only for the Boolean lattice, but for any finite directed graph with one source, one sink, and whose vertices are partitioned by levels.

For n=4, a top-down model is given by d(1,2)+d(2,1)=1 , d(1,3)+d(2,2)+d(3,1)=1
The bottom up (classical HTAM) $P(p(k; l_j); n)$ model which provide the same probabilities for trees is given by: p(1; 1,1,1,1)=0.5* d(2,2)+ d(3,1)d(2,1);
p(3; 1,1,1,1)=0.5* d(2,2)+ d(1,3)d(1,2);
p(2; 1,1,1,1)=d(3,1)d(1,2)+ d(1,3)d(2,1);
p(1; 1,2,1)=d(3,1)d(1,2)/(d(3,1)d(1,2)+ d(1,3)d(2,1)) and p(2; 1,2,1)=1-p(1; 1,2,1)
p(1; 2,1,1)=d(3,1)d(2,1)/(0.5* d(2,2)+ d(3,1)d(2,1)) and p(2; 2,1,1)=1-p(1; 2,1,1)
p(1; 1,1,2)=0.5*d(2,2)/(0.5* d(2,2)+ d(1,3)d(1,2)) and p(2; 1,1,2)=1-p(1; 1,1,2)

The results are relevant for the 1-dim HTAM model. We can define a Markov chain which model the assembly and the dis-assembly of bars of at most n tiles: the Boolean lattice G_n has its directed edges weighted by $p(k; l_i)$. The inverse directions are weighted by $d_{i,j}$ (the bottom-up

model is Corollary c) counterpart of the $d_{i,j}$ model). With probabilities p and $1-p$ respectively, the system assembly bars – it evolves from level L_k to L_{k+1} – or dis-assembly bars. These are semi-local rules, similar to certain parameters of forward and backward rates from the kinetic model kTAM [22].

*Example.*We define a Markov chain which model the assembly and the disassembly of bars of 4 tiles: with probabilities 1-p and p a set of 2 bars can evolve to a bar of length 4, or to disassemble. With probabilities 1-q and q , one bar of length 2 can disassemble or can be involved in a binding with a tile. We simulated this model in Mathematica, which provided several computable properties of this discrete time Markov chain on the Boolean lattice. Implementation and results: Parameters: d12 := 0.4; d21 := 1 - d12; d13 := 0.25; d22 := 0.35; d31 := 1 - d22 - d13; p := 0.25; q := 0.75

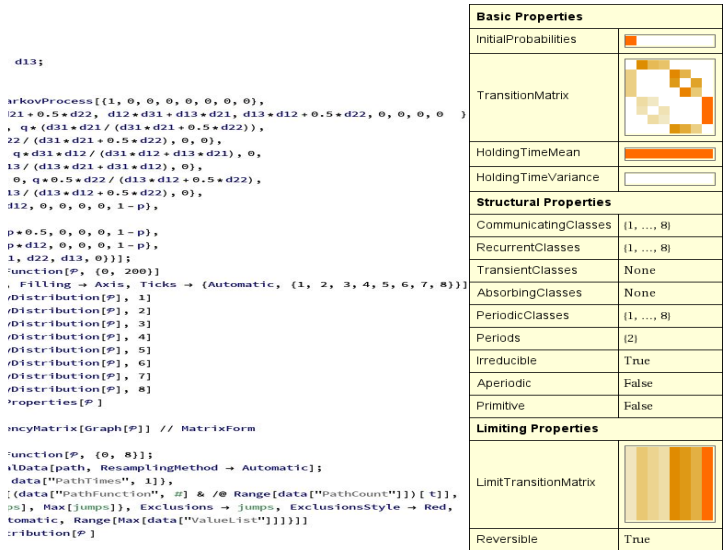

Fig. 8. Automatic $Mathematica$ computation of basic Markov chain properties

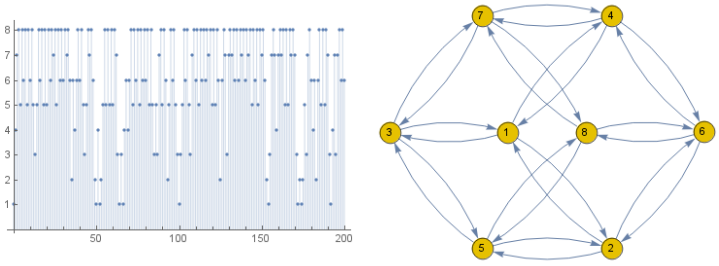

Fig. 9. Random walk of 200 steps in G(4); The graph:=1-skeleton of the 3-dim. cube.
Using the above parameters, the most probable final structures are 4-tile bars

4.2. Observations and Generalizations

Multiplicative relations as above $p\left(T_1^1\right) p\left(T_2^2\right) = p\left(T_1^2\right) p\left(T_2^1\right)$ are not necessary for the proof of Proposition 1 from Section 3.2. They are needed for the proof of Proposition 2 of Section 4, using Corollary b). They can be formalized using the algebraic structures of Operads and Combinatorial Hopf Algebras.

4.3. Definitions of the algebraic structures

DEFINITION 4..3. *A non-Σ operad O is a collection of sets $O(n)$, n $\geq$ 1 such that:*
There is a composition law f: $O(m) \otimes O(n_1) \otimes ... \otimes O(n_m) \longrightarrow O(n_1 + ... + n_m)$
There is a unit e $\in O(1)$. *f(g; e,e,e...e)=g for any g* $\in O(k)$. *The composition law f is associative:*
$f[f(g; g_1, g_2, ...g_n); r_1^1, r_2^1, ...r_{x_1}^1, r_1^2, r_2^2, ...r_{x_2}^2, ...r_1^n, r_2^n, ...r_{x_n}^n] =$
$= f(g; f(g_1; r_1^1, r_2^1, ...r_{x_1}^1), f(g_2; r_1^2, r_2^2, ...r_{x_2}^2)...f(g_n; r_1^n, r_2^n, ...r_{x_n}^n))$

In the case of the binary trees, the associativity of the composition law (the glueing of some trees to the univalent vertices of a bigger tree), shows that it does not matter the order of this operation if we apply it two times: $Graft(Graft(T; t_j); a_k)$=$Graft(T; Graft(t_j; a_k))$. Operads and Hopf Algebras connected with Algorithms are met in [18], [20], [19].

We consider the sets $(O_n)_{n\geq 1}$. O_n is the set of oriented paths of length n-1 from G_n.
We consider the following maps: $\beta^m_{l_1,l_2,...,l_m} : O_m \times O_{l_1} \times O_{l_2} \times \ldots \times O_{l_m} \rightarrow O_{l_1+l_2+...+l_m=n}$.
If $(d, d_1, d_2, \ldots, d_m) \in Domain(\beta)$, the map associates the following mixing path to n tiles. We join the first l_1 tiles using d_1, creating a bar of length l_1.
Next we create a bar of length l_2, using d_2 for the tiles $l_1 + 1 \ldots l_1 + l_2$ and so on.
Then we join the bars of lengths $l_1, l_2, \ldots, l_m$ using the path prescribed by d. In this way we get a path in O_n, called $\beta^m_{l_1,\ldots,l_m}(d, d_1, \ldots, d_m)$
The set of maps (β) forms the algebraic structure of Operad. The dissection of trees is associative. Every map (β) is injective, but not surjective – we did not shuffle or entangle the paths.

The top-down model $P\left(d_{i,j}\right)_n$ allow us to define probabilities on $PBT(m)$ for any $m \leq n$. This is not the case for the $P(\{p(k, l_i\})$ model which is restricted to the trees of a given dimension n. For any path d, the level L_k separates d into one path to the vertex (n), and k smaller paths d_{l_j} from the vertex (1,1...1) which show how these k bars were formed using elementary tiles.

Beginning with any distributions on O_1, O_2, O_3 we can define distributions (probabilities for each path) on all O_n, $n > 3$ using the formula $P(d) = \frac{1}{n} \sum_{k=1}^{n} \frac{l_1! l_2! \ldots l_k!}{n!} P_{d_k} P_{d_{l_1}} P_{d_{l_2}} \ldots P_{d_{l_k}}$ We conjecture good properties of these recursively defined distributions concerning solvability and the ability to assign the same probabilities for paths representing the same planar binary trees.

These probabilities built for trees are based on clusters of n tiles and a kind of static, scale-free reasoning. There are also probabilistic models (Prof. Trandafir) where $\{p(k, l_i\}$ depend on the space s_i between the bars of lengths l_i and l_{i+1}. In this case it is useful to assume that the probabilitiles are inverse-proportional to s_i, and the space parameters s_i are updated according to the level of the boolean lattice. It is also possible to assume random generation of parameters for each simulation.

The probabilities mentioned in Section 1.2, for specific ISU-TAS simulations (caption of Fig.2 and the next paragraph) are equal, $\{p(k,l_i\} = 1/m$, the weight of each maximal chain is $1/n!$, so the probability of each tree T is $\frac{1}{\prod h_v}$ according to the Knuth's formula. The sum of these probabilities, over all planar binary trees of a given dimension n is 1. This result appears as Theorem 6.1 in [10] where it was proved using the Knuth hook-length formula.

We consider the set H of formal power series $g(x) = 1+f_1(x-1)+f_2(x-1)^2+\ldots+f_n(x-1)^n+\ldots$ which are local diffeomorphisms in a neighborhood of 1 and g(1)=1. We consider the following operator $P(g)(z) = 1/(g(z)^{-1}/z)^{-1}$, where b(z)=$f(z)^{-1}$ is the compositional inverse: b(f(z))=f(b(z))=z. Then P(P(g))=g.

This involution can be generalized to a set M with 2 distributive group structures $\circ$ and $\bullet$, one of them being abelian (instead of multiplication and series composition). (x $\bullet$ y)$\circ$z = (x$\circ$z)$\bullet$(y$\circ$z) The unit of the abelian $\bullet$ is $1_\bullet$. The inverse of a is $a^{-1,\bullet}$.The unit of $\circ$ is $1_\circ$.
The inverse of a with respect to $\circ$ is $a^{-1,\circ}$. Let B be the element $(1_\circ)^{-1,\bullet}$. It is the function $1/z$ if we work in H. Then $P(g) = B \circ (B \bullet g^{-1,\circ})^{-1,\circ}$ is an involution. It is probably the minimal involution built using every symbol above. Its search was motivated by an analogue of Han inversion formula for the general weights $p(k;l_1,...l_m)$.

5. Conclusions and future work

The parallelism of HTAM in the top-down and in the bottom-up models is reflected in different ways at the level of dissecting trees. The first model has a recursive left-right structure, speculated in hook-length formulae. The second model use the levels of the boolean lattice. This self-assembly parallelism was connected to Operads and Hopf Algebras used in Quantum Field Theory by Manin, Marcolli and Yeats. The bottom-up model can be used to simulate with probability close to 1 any one dimensional staged assembly with any given mixing graph. An open problem is to describe the assembly pathways of the following two-dimensional assembly system: given two planar binary trees T_1 and T_2, we can consider a tile which will self-assembly a length n bar using T_1 for the East-West direction, and using T_2 for the North-South direction. The result will be a single tile which self-assembly a $n \times n$ square. In this case, even if there is only one final product, there are multiple ways to arrive to it: not easy to describe, count and geometrically visualize. There are rich mathematical structures involved in the simulation of one-dimensional staged self-assembly by a one-pot self-assembly. Staged self-assembly is three-dimensional and self-assembly algorithmic formalism provides the insight of several mathematical results. Propositions 2 and 4 can be generalized to other directed graphs of interest.

Acknowledgements. This research was funded by the Romanian National Authority for Scientific Research project MoDASyS, co-funded by the European Fund of Regional Developement grant number POC P-37-257.

References

[1] WANG, H. *Proving theorems by pattern recognitionII.* Bell system technical journal **40**.1 (1961): 1-41.

[2] WINFREE, E. *Simulations of computing by self-assembly.* Proc. of the Fourth DIMACS Meeting on DNA Based Computers. 1998.

[3] ROTHEMUND, PWK, and E. Winfree. *The program-size complexity of self-assembled squares* Proceedings of the thirty-second annual ACM symposium on Theory of computing. 2000.

[4] PATITZ, M. J. *An introduction to tile-based self-assembly and a survey of recent results.* Natural Computing, 2014, **13.2**: 195-224.

[5] AGGARWAL G., et al. *Complexities for generalized models of self-assembly* SIAM Journal on Computing, 2005, **34.6**: 1493-1515.

[6] ADLEMAN, Leonard M. *Molecular computation of solutions to combinatorial problems* Science, 1994, 1021-1024.

[7] DEMAINE, Erik D., et al. *Staged self-assembly: nanomanufacture of arbitrary shapes with O (1) glues.* Natural Computing, 2008, **7.3**: 347-370.

[8] BARAD, G., et al. *Simulation of one dimensional staged DNA tile assembly by the signal-passing hierarchical TAM.* Procedia Computer Science, 2019, **159**: 1918-1927.

[9] PADILLA, J. E.; LIU, W.; SEEMAN, N. C. *Hierarchical self assembly of patterns from the Robinson tilings: DNA tile design in an enhanced tile assembly model*. Natural Computing, 2012, **11.2**: 323-338.

[10] HAN, G. N. *Discovering hook length formulas by an expansion technique*, The Electronic Journal of Combinatorics, **15**(1), pp. 133-174.

[11] JONES, B. R. & YEATS, K. (2015). *Tree hook length formulae, Feynman rules and B-series*, Annales de lInstitut Henri Poincar D, **2**(4), pp.413-430.

[12] BLUM, M. GB; FRANOIS, O. *Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance.* Systematic Biology, 2006, **55.4**: 685-691.

[13] HAGEN, O.; STADLER, T. *TreeSimGM: Simulating phylogenetic trees under general BellmanHarris models with lineagespecific shifts of speciation and extinction in R*. Methods in ecology and evolution, 2018, **9.3**: 754-760.

[14] LODAY, J. L. & RONCO, M. O. (1998). *Hopf Algebra of the Planar Binary Trees*, Advances in Mathematics, **139**(2), pp.293-309.

[15] FOISSY, L. & UNTERBERGER, J., *Ordered Forests, Permutations, and Iterated Integrals*, IMRN 2013 **(4)**, pp.846-885.

[16] HIVERT, F., NOVELLI, J. C. & THIBON, J. Y. *Trees, functional equations, and combinatorial Hopf algebras* European Journal of Combinatorics, **29**(7), pp.1682-1695.

[17] Wolfram Research, Inc., Mathematica, Version 11.2, Champaign, IL (2018); (www.wolfram.com) (2018)

[18] KRAHMER, U. *From computer algorithms to quantum field theory: an introduction to operads*, Snapshots of modern mathematics from Oberwolfach **17**/2015 , pp.1-10

[19] MANIN, Y. I. (2014). *Complexity vs energy: theory of computation and theoretical physics* Journal of Physics: Conference Series, Vol. **532**, No. 1, p. 012018. IOP Publishing.

[20] TANASA, A. *Combinatorial Hopf algebras in noncommutative quantum field theory*, Romanian Journal of Physics, **55**(9-10), pp.1142-1155.

[21] DUNN, K. E.,et. al. *Guiding the folding pathway of DNA origami* Nature, **525**(7567)

[22] MOHAMMED, A.; CZEIZLER, E.; CZEIZLER, E. *Computational modelling of the kinetic Tile Assembly Model using a rule-based approach.* Theoretical Computer Science, 2017, **701** 203-215.

[23] AMARIOAREI, A., et al. *One Dimensional DNA Tiles Self Assembly Model Simulation* International Journal of Unconventional Computing **13** (2018).

[24] MARDANLOU, V., et al. *A coarse-grained model captures the temporal evolution of DNA nanotube length distributions* .Natural Computing, 2018, **17**.1: 183-199.

[25] APOSTOLOPOULOU, I.; MARCULESCU, D. *Tractable learning and inference for large-scale probabilistic Boolean networks.* IEEE transactions on neural networks and learning systems, 2019, **30**.9: 2720-2734.