# Research Project: mutidimensional subshifts of finite type, open questions and interactions with computer science, physics and cognitive sciences

Silvère Gangloff

January 2, 2020

This research project is composed of three main axes of different natures. The first one [Section 1] lies at the interface between mathematics and computer science, and is concerned with a computational threshold phenomenon in the dynamics of multidimensional subshifts of finite type (SFT). I describe in this section the main open questions, conjectures and research directions that I would like to pursue, as well as extensions of this theme. The second one [Section 2] is related to mathematical physics, in particular about the methods used by physicists to compute the entropy of some multidimensional subshifts of finite type. The aim of this research is to convert these methods into rigorous ones. The third one [Section 3] lies at the interface between mathematics and cognitive sciences, and is based on a recent and still controversial theory of consciousness based on an aspect of conscious experience, namely integration of information. This third axis consists in an attempt to formalize this notion of integration in a way that can lead to novel ways to analyse dynamical systems, starting with multidimensional subshifts of finite type.

## 1 Characterizing a computational threshold phenomenon

The first axis of research is related to my doctoral thesis: *Computability of growth type invariants of multidimensional subshifts of finite type under dynamical constraints*, done under supervision of M. Sablik. In this section, I describe some problems that we have left open [Section 1.2], some conjectures of interest that are related to this theme on which I have given thoughts, some research directions and some extensions of this theme [Section 1.3].

### 1.1 Some definitions

Le us just recall that a bidimensional subshift of finite type (SFT) is a subset $X$ of some $\mathcal{A}^{\mathbb{Z}^2}$, where $\mathcal{A}$ is a finite set, defined by a finite set $\mathcal{F}$ of finite forbidden patterns:

$$X = \{x \in \mathcal{A}^{\mathbb{Z}^2} : \forall p \in \mathcal{F}, p \not\sqsubseteq x\}$$

The entropy of a bidimensional SFT $X$ is defined as:

$$h(X) = \inf_n \frac{\log_2(N_n(X))}{n^2}$$

Moreover, a bidimensional SFT is said $f$-block gluing, for a function $f : \mathbb{N} \to \mathbb{N}$, called gap function, when any pair of size $n$ square patterns appearing in an element of the SFT can be glued in a configuration in any relative positions, provided that the distance is greater than $f(n)$. A real number $x$ is said to be $\Pi_1$-computable when there is an algorithm which on input $n$ outputs a rational number $r_n$ such that for all $n$, $r_{n+1} \leq r_n$. This number is said to be computable when moreover $|x - r_n| \leq 2^{-n}$.

## 1.2 Open problems

In a first article with M. Sablik, we proved the following:

**Theorem 1** ([GS17a]). *A $f$-block gluing bidimensional subshift of finite type such that $f(n) = o(\log(n))$ has computable entropy. Moreover, the possible values of entropy for linear block gluing ($f(n) = O(n)$) ones are the $\Pi_1$-computable numbers.*

We have thus delimited a transition area for computability of entropy according to a quantification of the block gluing property. However, we were not able to make any statement on computability of entropy for bidimensional SFT for a gap function between logarithmic and linear, the main tools used in both statements of the theorem breaking in this area. Our conjecture is the following:

**Conjecture 1.** *A block gluing subshift of finite type whose gap function is between logarithmic and linear has computable entropy.*

### 1.2.1 Exploring the transition area for quantified block gluing property

In order to approach this problem, we aim at converting it into an intermediate difficulty one that would be at the same time tractable and non trivial. The most natural way to do it is to ask about the decidability of the language:

**Question 1.** *Does there exist such a SFT in the transition area whose language is undecidable ? In other words, is this possible to decide with an algorithm if a pattern appears in an element of the SFT or not ?*

We already know how to construct a bidimensional SFT with undecidable language which is linearly block gluing, which consists in implementing universal Turing machines over computing zones that consist in infinite triangles that can be distorted by shifting one row from the other in any of the two horizontal directions. This idea does not adapt directly to the transition area, so we need another one. One can lower again the difficulty of the problem and ask:

**Question 2.** *Does there exist a bidimensional subshift of finite type which is block gluing with gap function between logarithmic and linear ?*

The answer to this question seems to be positive, based on an on-going reflexion, for instance for a square root gap function. The idea is to generate infinite areas that grow from bottom to top with a speed that corresponds to the gap function. However, the process that generates this growth speed has some constraints to satisfy, which seem to be reducible to a condition on the direction of information propagation in this process, imposing that this propagation goes in only one horizontal direction. It seems that very few functions verify this condition. This thus leads to classification and enumeration questions:

**Question 3.** *What is the class of functions between logarithmic and linear which are the gap function of a bidimensional block gluing subshift of finite type ?*

**Question 4.** *Is this class finite or infinite?*

One possibility to explore this class is to use operators on subshifts of finite type, such as the one defined and used in [GS17a] that would act on the gap function of block gluing bidimensional SFT in order to construct new gap functions from existing ones.

### 1.2.2 Refining the knowledge of the threshold for decidable subshifts

In order to better understand this transition area, we worked with B. Hellouin on characterizing a threshold for a more flexible class of systems than subshifts of finite type: subsets of some $\mathcal{A}^{\mathbb{Z}^2}$ defined by forbidding patterns in a set $\mathcal{F}$ which is not finite but decidable, namely bidimensional decidable subshifts. We arrived at the following result:

**Theorem 2** ([GH18]). *If the series $\sum_k \frac{f(2^k)}{2^k}$ converges to a computable number, then the entropy of any decidable bidimensional subshift which is $f$-block gluing s computable. On the other hand, if the series $\sum_k \frac{f(2^k)}{2^k}$ converges to infinity, then the possible values of entropy for these subshifts are the $\Pi_1$-computable numbers.*

Although this characterization is very precise, a small gap remains: we don't know what happens when the series converges, but not to a computable number. Hence the following question:

**Question 5.** *Is the entropy of a bidimensional subshift which is $f$-block gluing is computable when $\sum_k \frac{f(2^k)}{2^k}$ converges ?*

One could also go further and get to better understand this threshold, by asking about a potential relation between the mode of convergence of the series $\sum_k \frac{f(2^k)}{2^k}$, for instance the speed of convergence, and the mode of convergence of entropy for $f$-block gluing decidable subshifts.

### 1.2.3 Characterizing values of entropies below the threshold

Another important problem is the characterization of values of entropy below this threshold. This is a very difficult problem, that can be declined into multiple variations:

**Question 6.** *What are the values of entropy for $f$-block gluing bidimensional decidable subshifts such that $\sum_k \frac{f(2^k)}{2^k}$ converges to a computable number ?*

**Question 7** (R. Pavlov, M. Schraudner)**.** *What are the values of entropy for $f$-block gluing bidimensional SFT such that $f$ is constant ?*

**Question 8.** *What are the values of entropy for $f$-block gluing bidimensional SFT such that $f(n) = o(\log(n))$ ?*

This question can be also formulated for other dynamical constraints, such as strong irreducibility (gluing patterns that are not necessarily square ones with distance greater than a constant) [HM10].

Some partial answer was given by R. Pavlov and M. Schraudner [PS15], realizing a subclass of computable numbers defined by a condition on the speed of computability as entropies of constant block gluing bidimensional SFT. Their construction is based on a natural operator that transforms any SFT into a constant block gluing one, which preserves entropy whenever the initial SFT, denoted $X$, satisfies a condition on the speed of convergence of $\log(N_n(X))/n^2$ to $h(X)$. They thus realize any number is the defined class as entropy of a SFT which satisfies this condition. This construction relies on the classical construction of R. Robinson [Rob71] of aperiodic tilings and embedding Turing machines computations, with the use of sturmian words in order to generate entropy, using the properties of their complexity function.

I would like to propose another approach of this question; the idea is basically to consider high dimensional subshifts of finite type, for instance three-dimensional ones, consisting in a stack of Robinson tilings with a direction of time, each of these tilings supporting Turing machines acting on only one bit of information. In the direction of time, the machine would act on this bit of information and the complexity generated by these bits together in order to correct the complexity function. Let us notice that the difficulty of this work is to control the number of *locally admissible patterns*, meaning those which do not contain any forbidden patterns, but do not necessarily appear in an element of the SFT. Considering higher dimensional subshifts of finite type would help to simplify the construction and see how one can act on this complexity function with Turing machines computations.

## 1.3 Other research directions

I describe here more research directions and reflexions that are broadly related to the research theme of the last section. An important part of my reflexion has been around the following conjecture:

**Conjecture 2** (B. Weiss). *Any bidimensional sofic subshift has a SFT cover which has the same entropy.*

Here a bidimensional sofic subshift is a subset of some $\mathcal{A}^{\mathbb{Z}^2}$ such that there exists an SFT that is mapped to this one by a local map. Such a SFT is called a cover of the sofic subshift. This conjecture is of interest for me since it would need to be answered to find new ways to construct subshifts of finite type with properties related to entropy.

My approach to the conjecture would be based on the idea that we can construct a sofic subshift without SFT cover having the same entropy, by coding a SFT whose elements consists in displays of "objects", such as curves or square areas, which are the support of a transfer of information aimed at synchronizing some bits of information, supported for instance by the extremities of the objects. The idea is that this transfer of information would need, in its definition, an orientation. Since the objects would not have specific orientation, the two possible orientations of the information transfer would be possible. The subshift obtained by forgetting the symbols representing this orientation would be a counterexample to the conjecture.

One would need then to understand the relation between the entropy of a subshift composed of 'objects' and the one enriched with random bits superimposed to specific parts of these objects, in order to prove the difference or the equality of these entropies. This question could be formulated apart from an approach to the conjecture.

Let us mention two other hard questions related to high dimensional SFT:

**Question 9.** *Does there exist a constant block gluing tridimensional SFT with uncomputable entropy ? With undecidable language ?*

**Question 10.** *Does there exist an aperiodic constant block gluing tridimensional SFT ?*

### 1.3.1 Other physical quantities: example of the spectral gap

Recently, another result of undecidability has been proved on another physical quantity defined on multidimensional subshifts of finite type (called lattice models in physics), namely the spectral gap.

In order to explain this notion, we need to recall first the notion of adjacency matrix: given a bidimensional SFT, its order $n$ adjacency matrix describes which pairs of length $n$ rows of symbols can be appear adjacent in an element of the SFT. The model is said to exhibit a spectral gap when the difference between the maximal eigenvalue of this matrix and the second largest stays above a constant for an infinity of numbers $n$. It has been proved the following:

**Theorem 3** ([CPGW]). *There is not algorithm that decides if a bidimensional SFT has a spectral gap or not.*

This theorem has been presented in a short version published in the journal *Nature* [CPGW']. The initial paper is very long and although the construction relies on Robinson tilings, using quantum Turing machines instead of simples ones, it seems to be independent from the developments around R. Robinson's construction made by M. Hochman and T. Meyerovitch [HM10] as well as the fixed-point constructions of B. Durand and A. Romashchenko [DR17]. Given that this type of constructions consists mainly in introducing new interacting gadgets in the initial construction of R. Robinson, it would be interesting to understand better this construction in detail, in order to extract new tools to deal with questions related to the computability of entropy.

On the other hand, one can transport the theme of effect of dynamical constraints on computability or decidability to the spectral gap problem:

**Question 11.** *Do there exist some dynamical constraints on bidimensional SFT that imply the decidability of the spectral gap ?*

Since many properties of lattice models are determined by the existence of a spectral gap, this problem is important for physicists [K17], and any constraint of this type could be interpreted as the localisation of the computable in the world of lattice models, having theoretical and practical implications.

### 1.3.2 A correspondance between dynamical restrictions and topological invariants

In another article with M. Sablik, we used the tools that we developed for the block gluing property in order to characterize the values of entropy dimension for minimal SFT. The entropy dimension, when it exists, is defined as follows, for a SFT $X$:

$$D_h(X) = \lim_n \frac{\log(\log(N_n(X)))}{\log(n)}$$

We proved, in particular, the following:

**Theorem 4** ([GS17b]). *The values of entropy dimension for minimal tridimensional subshifts of finite type are the $\Delta_2$-computable real numbers in* $[0, 2]$.

Here, a real number $x$ is $\Delta_2$-computable when there exists an algorithm which on input $n$ outputs a rational number $r_n$ such that the sequence $r$ converges to $x$.

Based on this result, there are two possible directions to follow:

1. One can introduce some quantification of the minimality property by the minimal distance that separates two occurrences of the same pattern in a configuration, and ask about the effect of this distance on the possible entropy dimensions of multidimensional SFT. This approach can be extended to other aspects of multidimensional SFT such as the discrete dynamical systems that can be simulated by a SFT under this constraint.

2. On can notice that the entropy of a minimal SFT is zero, and the only number that can be entropy dimension of a tridimensional SFT which is block gluing with a linear gap function, bounded from below by $n \mapsto Kn$ with $K$ arbitrarily large is 3. As a consequence, this draws an association between two topological invariants and two dynamical restrictions, since the possible values of any of these two invariants for tridimensional SFT under the corresponding dynamical restriction form a non trivial set, while this set is trivial when the invariants and restrictions are crossed. On can thus imagine drawing a wider correspondance between dynamical properties similar to the block gluing and the minimality, defined for instance by logical propositions on patterns, and topological invariants such as entropy and entropy dimension, by variation of the functions involved in the definition of the invariant. Based on an on-going reflexion with I. Torma, one can characterize a precise threshold for bidimensional SFT for the polynomial complexity [M11] and a notion of linear transitivity. Such a correspondance could possibly lead to a "threshold" on the function defining the topological invariant at which this threshold is transformed into a non-trivial transition area, enabling us to understand better this transition area for the entropy and the quantified block gluing property.

*All the questions in this part could benefit collaboration with specialists of discrete dynamical systems and computation models: P. Guillon (I2M), Guillaume Theyssier (I2M), A. Romashchenko (LIRMM), B. Hellouin (U. d'Orsay), Ilkka Torma (U. of Turku) as well as specialists of multidimensional dynamics such as M. Sablik (IMT), N. Aubrun (ENS Lyon), R. Pavlov (U. of Denver), M. Schraudner (U. of Chile).*

# 2 Exact computations of entropy

Multidimensional subshifts of finite type, as lattice models, have been considered by physicists for a long time, since the beginning of the XXth century, starting with the Ising model and its celebrated solution by L. Onsager. Here the term solution refers to the ability to compute formally (but not necessarily rigorously) some global quantities related to the model, such as entropy (free energy in physics). Although entropy computations have been considered as a pretext for the development of tools to compute more refined quantities such as correlation functions, much effort has been made to compute the entropy for few of these models. However, only a part of these works can be considered as fully rigorous, and the form of the litterature on the subject make these not accessible to mathematicians. My goal here is to provide some rigorous proofs by new means, or at least to complete the computation methods so that they appear completely rigorous. In this section, I would like to describe the work left to do in this direction [Section 2.1], another approach to the general problem of computing exactly the entropy of multidimensional SFT and some original questions that came out of my study of the square ice model [Section 2.2].

## 2.1 Rigorisation of exact computation methods

Besides for the dimers model, for which entropy has been computed rigorously by P.W.Kasteleyn [K61], and the square ice model, for which we have recently given full rigor to the computation of its entropy [G19a], there are two models I would like to consider: the eight-vertex, and hard hexagons models.

The eight-vertex model has been solved by R.Baxter [Baxter], and the computation method for its entropy can be decomposed, as well as for square ice, into a computation of a formula for the greatest eigenvalue of the adjacency matrices, and then an analysis part for the asymptotic behavior of this sequence of eigenvalues. After some work with P. Melotti, we concluded that this computation poses two types of problems: the first part of the computation is not directly accessible to mathematicians, and arguments are missing for the analysis part, although physicists admit that the eight-vertex model behaves as the square ice from this point of view. Hence the following research direction:

**Research direction 1.** *Review the literature and provide a rigorous version for the solution of the eight-vertex model. Adapt the tools developped for the analysis part of square ice entropy computation to the eight-vertex model.*

The case of the hard hexagons model is more difficult, since the computation of its entropy by R.Baxter [Baxter] is clearly non-rigorous. It relies on a functional equation verified by a path of diagonal adjacency matrices, for each of. the two diagonal directions. Due to special properties of these paths, these functional equations are transported to the greatest eigenvalues of these matrices. Then the value of the entropy is derived from a hypothesis of equality between the greatest eigenvalue of the two diagonal transfer matrices for the same size and parameter, the consequence of this hypothesis being that the entropy satisfies a system of functional equations that is sufficient to determine it as a function of the parameter. Hence the following direction:

**Research direction 2.** *Prove or disprove the equality of the greatest eigenvalues of the two diagonal adjacency matrices.*

On can also notice that there are important models for which no formula is conjectured for entropy: the hard core models, and non-attacking-kings subshift. There is also a whole literature in physics about solvable models produced out of abstract solutions of the Yang-Baxter equation and related to particuliar properties of algebraic structures, such as braid algebra. We don't know if these results are completely rigorous and if entropy is actually computed for some of these models. In any case, this would be important to consider this literature for the question of the computability of entropy of lattice models.

*The questions in this part could benefit collaboration with specialists of exactly solved models as K.Kozlowski, as well as specialists of percolation theory who are using results on exactly solved models, such as H. Duminil-Copin.*

### 2.2 Combinatorial approach to exact computation of entropy

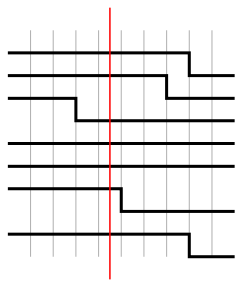
Here we propose some original approach and questions related to entropy computations for multidimensional SFT, that appeared during my study of square ice, motivated by developping combinatorial methods for computing the entropy. In my work with M. Sablik [GS17a], we used in the proof of the main theorem the exact computation of the entropy of a SFT obtained by some modifications on the square ice model. The proof is not trivial but uses only combinatorial arguments. One could try to extend this proof to other similar SFTs, by distorting the ones considered in [GS17a] in a way that the method remains valid, with some non-trivial adaptations. One could hope this way to attack the question:

**Question 12.** *Does there exist a combinatorial proof for the value of square ice entropy ?*

In [GS17a], we used a sequence $(T_r)_r$ of distortion operators on bidimensional subshifts of finite type, which act on the entropy by adding $\log_2(r+1)/r$. On could ask if it is possible to construct other operators that act on entropy in a given way, for instance:

**Question 13.** *Does there exist an operator on bidimensional SFT that transforms entropy into its square ?*

Also in [GS17a] , we used some subshifts of square ice defined by further imposing the vertical sections of any configuration to lie in a fixed unidimensional subshift of finite type on alphabet $\{0,1\}$, where a vertical section in a configuration is a bi-infinite word obtained by considering the restriction of the configuration to a column, and writing 1 on a position if it contains a segment of curve and 0 if not. One can see why it is called a section on the following figure:



One can ask about the relation between the restriction on the section, denoted $X$, and the bidimensional SFT obtained by this process, denoted $s(X)$. For instance :

**Question 14.** *Does there exist some unidimensional SFT $X$ whose entropy is positive such that the entropy of $s(X)$ is zero ? And conversely ? Is it possible to decide for which $X$ the entropy of $s(X)$ is positive ?*

## 3 Formalizing the phenomenon of integration

This third axis of research is the fruit of the work done during to my current post-doctoral year in the department of psychiatry of the University of Wisconsin-Madison. Some research is conducted here on a systematic approach of consciousness proposed by G. Tononi, called Integrated information theory, which is developed in an interdisciplinary environment involving philosophers, physicists, experimentalists, cognitive scientists, computer scientists. However, although the theory is formulated with mathematical notions, this theory seems to stay out of the scope of mathematicians, despite some recent attention [BZ12][PW18] on some earlier definition of the notion of neural complexity [TSE]. The reason seems to be mainly the ”language” barrier between broadly cognitive sciences and mathematics. My work here so far and the work left to be done consists in translating this theory into a coherent mathematical framework that would make sense for mathematicians, and extract mathematical problems of interest. I will attempt in this part to expose my understanding of the theory, the work to be done in this direction, and how interactions between cognitive theoretists and mathematicians could be interesting for both on this matter.

The theory is based on an axiomatisation of conscious experience, defining some *structural phenomena* (called axioms in the theory, which term is still subject of reflexion) that are verified by any conceivable experience [Section 3.1]. The most important one seems to be the phenomenon of *integration*, which corresponds to the idea that things in the experience "hold together", or that parts in the brain interact intensely, which is translated into the fact that a cut in the brain influence strongly its behavior. This phenomenon seems to have multiple interpretations in mathematical terms, that we expose informally in Section 3.2. We also relate in some interesting mathematical problems related to integration measures, and expose in Section 3.3 some formalization problem about an important direction of the theory, that searches to use integration to make sense of the fact that activation of grid-like neural networks in the brain seem to be directly related to consciousness.

## 3.1 Axiomatisation of conscious experience

The integrated information theory of G. Tononi relies on an axiomatisation of conscious experience. This is of matter however to understand that the axioms involved have a different nature from mathematical axioms, for the association of their formulation and their meaning is non trivial. It is still an interesting question to understand their nature, but I would give them the name of *structural phenomena*, since they are themselves part of the experience, although they are encountered in any possible experience. Moreover, their translation into mathematical terms is not direct and can even be multiple. These structural phenomena can be formulated shortly as follow (see for instance [AOT]):

1. **Intrinsic existence:** experience exists as such, independently from external observers.
2. **Composition:** any experience is composed of phenomenological distinctions: parts of experience can be distinguished from others, and experiences can be distinguished by the distinction of some of their parts.
3. **Information:** each experience is specific, and contains information about what it is not (alternative parts or alternative experiences).
4. **Integration:** any experience is unified: it can not be reduced to the concatenation of independent experiences.
5. **Exclusion:** an experience can be characterized by its phenomenal distinctions.

The aim of this axiomatisation is to direct the search of *neural correlates of consciousness*: any physical system that underlies consciousness (in the humain brain) has to account for these structural phenomena. For this purpose, the theory translates the axioms into physical postulates that rely on causality notions: it assumes first that a physical system is composed of interacting elements (composition) and the existence of a part of the system is translated in a non-trivial causal power over some other part of the system, including itself (intrinsic existence); phenomenal distinctions correspond to physical distinctions (the causal actions of parts of the systems over others) and these distinctions form together a causal structure : any part of the system *exists* as this causal structure (information and exclusion); a part of the system that underlies conscious experience must be integrated: any cut of this part into sub-parts has an effect on the causal structure, whose measure gives the level of integration.

## 3.2 Measures of integration

Let us focus here on the phenomenon of integration. The above description of the physical postulate corresponding to integration leads to the definition of a quantity of integration, denoted $\Phi$. Our purpose here is not to give a formal definition of this quantity. Let us notice however that it is very difficult to compute, even for simple systems (since it involves to enumerate and test all the candidate distinctions). This is an important issue, which motivates the search of other more tractable measures of integration, where the criterion for tractability would be the existence of sufficiently complex systems for which it is possible to compute this quantity, or even approximate with sufficient efficiency. Let us note here that the definition of $\Phi$ follows multiple other attempts of definitions by G. Tononi, such as the *neural complexity* also defined in terms of the effect of cuts, and the *functional clustering index* [EMRT], which is a ratio between internal exchange of information of a part of a system and the external exchange of information. Let us also mention quantities defined by other theoretists: the *perturbational complexity index* [PCI], which measure how much a system reacts stereotypically to perturbations, whose definition is mainly shaped to computation over data, and the *geometrical information integration* [GP], based on information geometry.

Theoretical tests to these proposed quantities, which are all based on notions of causality (for instance the type of reactions of a system to a perturbation shows how the elements of this system are causaly related) are often done on models that are in the scope of cognitive scientists, such as the ising model, or cellular automata such as Conway's game of life. My claim is that multidimensional subshifts of finite type such as the dimers model or the square ice model provide good benchmarks for selecting amongst these propositions, since on the contrary, their causal analysis is relatively simple. Hence the following research direction:

**Research direction 3.** *Test and compare the proposals of measures for integration and related causality notions over multidimensional subshifts of finite type. For the perturbational complexity index, find a pertinent definition for these systems, based on a definition of perturbation.*

Moreover, contrarily to a quantity such as entropy, the following questions are far from trivial for any of the above quantities:

**Research direction 4.** *Computing the maximal value over the class of multidimensional subshifts of finite type and finding some systems that attain this maximal value.*

This question is important since an answer would provide at the same time a confirmation of the pertinence of the measure and paradigmatic examples of integrated systems and principles of integration for further analysis of the class of multidimensional subshifts of finite type and dynamical systems in general. For this question, one approach could be to considered simplified version of the quantities: for instance in the definition of $\Phi$, one could consider only one possible cut, or a simple family of cuts, progressively augmenting it to recover the full complexity of $\Phi$.

Let us mention also that integrated information theory is intimately related to the notion of actual causation defined by the philosopher B. Weslake [Weslake]: an actual cause of an element to be in a particular state is a minimal set of elements (in some states) that is sufficient to determine the state of this element. L. Albantakis [[Albantakis] proposed a generalization of this notion of actual causation to causes of a set of elements $S_1$ (in itself, not as the collection of its elements) to another set of elements $S_2$, by measuring, with a function denoted $\alpha$, how cuts amongst the first set of elements influences its causal action on the second set. An actual cause of $S_2$ to be in a particular state is a maximum of this function $\alpha$. With L. Albantakis, we have shown that when the second set is reduced to one element, the actual cause coincides with the Weslake's notion. This gives some hope to characterize higher order causation in combinatorial terms:

**Research direction 5.** *Is it possible to provide a characterization of actual causes in combinatorial terms when $S_2$ is not reduced to one element ?*

An answer to this (simpler) question would lead to a better understanding of the notion of integration according to Integrated information theory.

### 3.3 Neural graphs having potential for integration

An important aim of Integrated information theory is to account for the shape of graphs of neural networks in the brain that seems to underly consciousness. G. Tononi distinguishes informally three types of graphs, based on the bidimensional grid vertex set, with different edge sets. The first one consists in the gathering of small groups of interconnected elements that are not themselves interconnected (i). The second one is the set of graphs obtained by attributing in.an irregular way to all the vertex some pattern of connectivity chosen amongst a prefixed set, that tells to which other vertices a vertex is connected with an edge (ii). The third type is the grid itself (iii).

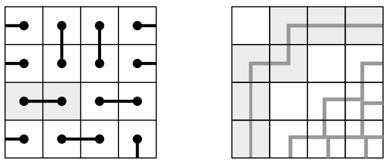
These three types of graphs are encountered in the brain, and amongst these, experiments show that conscious experience corresponds to activation of neural networks shaped as type (iii). One idea of the theory is that this can be accounted by the richness of set of integrated structures that can be generated by neural networks on this graph compared to other graphs, in particular of types (i) and (ii). Hence the following question:

**Research direction 6.** *State formally and prove the increase of richness of the set of integrated structures that can be generated by complex systems on the grid compared to types of graphs (i) and (ii).*

One can use subshifts of finite type on graphs to formalize this, and also enlarge the class of graphs considered, including for instance more regular graphs used to define subshifts of finite type as dynamical systems: cayley graphs of finitely generated groups, such as Baumslag-Solitar groups.

### 3.4 A notion of object

It appears when considering the notion of integration of Integrated information theory on simple multidimensional subshifts of finite type such as the dimers model or the square ice model, that the integrated structures appearing in them are numerous and do not correspond to the intuitive notion of what "exists" in an element of the subshift, where this notion would correspond to the dimers themselves in the dimers model and the curves in the square ice model (see the following figure where these are highlighted).



In order to define a notion of object that matches more this intuition, motivated at the same time by defining some notions that would be the base for a more tractable notion of integration (for instance the richness of the set of objects, or even types of objects (given that two dimers or two curves for instances are intuitively of the same type)) and a formalisation of the "gadgets" involved in the constructions of multidimensional SFT embedding Turing machines computations, I am aiming at defining a notion of object that uses actual causation. The idea is similar to the one of the functional clustering index, for which exchanges of information are quantified using information theory notions such as mutual information, but replacing mutual information by actual causation. More precisely, an object would be a pattern in an element of the subshift such that for all position in the pattern, the corresponding symbol has a cause mainly inside this set. One can prove, with a more formal definition that I won't detail here, that objects in the dimer model are exactly the dimers, and that the objects of the square ice model are exactly the curves.

Ideally, we would like to prove that this notion, or an adaptation of this it, is pertinent for more complex models such as Conway's game of life.

**Research direction 7.** *Adapt the notion of object so that it corresponds to the intuitive notion of objects in Conway's game of life.*

Since this cellular automaton is difficult to analyse in causal terms, one strategy would be to consider first some perturbations of the dimers and square ice model, replacing the dimers by trimers (or more complex shapes) for instance, as on the following figure,

or imposing constraints on the distance between the curves, progressively adapting the notion so that it covers all the models considered.

## References


[[Albantakis] L. Albantakis, E. Hoel, W. Marshall, G. Tnoni, What caused what? A quantitative account of actual causation using dynamical causal networks. *Entropy*, 2019, Vol. 21, Iss. 5.

[PCI] A.G. Casali et al., A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. *Science Translational Medicine*, 2013, Vol. 5, Iss. 198.

[GP] S. Amari and M. Oizumi and N. Tsuchiya, Unified framework for information integration based on information geometry. *Science Translational Medicine*, 2013, Vol. 5, Iss. 198.

[AOT] L.Albantakis and M.Oizumi and G.Tononi, From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. *PLoS Computational Biology*, 2014, Vol. 10, Iss. 5.

[Baxter] R.Baxter, Exxactly solved models in statistical mechanics *Academic Press*, 1982.

[B66] R. Berger Undecidability of the domino problem. *Memoirs of the American Mathematical Society*, 1966, Numb. 66.

[BZ12] J.Buzzi and L.Zambotti, Approximate maximizers of intricacy functionals. *Propability and related fields*, 2012, Vol. 153, Iss. 3-4.

[C97] M.Carvalho, Entropy dimension of dynamical systems. *Portualiae Mathematica*, 1997, Vol.54, Iss. 1.

[D15] H.Duminil-Copin and M. Gagnebin and M. Harel and I. Manolescu and V. Tassion, Discontinuity of the phase transition for the planar random-cluster and Potts models with $q > 4$. Arxiv 2016

[DR17] B.Durand and A.Romashchenko, On the expressive power of quasiperiodic tilings. Arxiv 2017

[EMRT] G.M.Edelman and A.R. McIntosh and D.P. Russel and G. Tononi, Functional clustering: identifying strongly interactive brain regions in neuroimaging data. *Neuroimage*, 1998, Vol. 7, Iss. 2.

[GHRS] S. Gangloff and A. Herrera and C. Rojas and M. Sablik, On the computability properties of topological entropy: a general approach. Arxiv 2019, Submitted

[GH18] S.Gangloff and B.Hellouin, Effect of quantified irreducibility on the computability of subshift entropy. *Discrete and continuous dynamical systems*, 2019, Vol. 39, Iss. 4.

[G19a] S.Gangloff, A proof that square ice entropy is $\frac{3}{2}\log_2(4/3)$. Arxiv, 2019. Submitted

[G19b] S.Gangloff, From coordinate to algebraic Bethe Ansatz for square ice. Arxiv, 2019. Submitted

[GT] S.Gangloff and A.Talon, Asymptotic growth rate of square grids dominating sets: a symbolic dynamics approach. Arxiv, 2019. Submitted

[GS17b] S. Gangloff and M. Sablik, A characterization of the entropy dimensions of minimal z3-sfts. arxiv, 2017. Submitted

[GS17c] S. Gangloff and M. Sablik. Simulation of minimal effective dynamical systems on the Cantor set by minimal tridimensional SFT. Arxiv, 2017. Submitted

[GS18] S. Gangloff and M. Sablik, Simulation of minimal effective dynamical systems on the Cantor set by minimal tridimensional SFT. Arxiv, 2018. Submitted

[GS17a] S. Gangloff and M. Sablik, Block gluing intensity of bidimensional sft: computability of the entropy and periodic points. *Journal d'Analyse math/'ematique*, to appear.

[HM10] M. Hochman and T. Meyerovitch, A characterization of the entropies of multidimensional shifts of finite type. *Annals of Mathematics*, Vol. 171, Iss. 3.

[K61] P.W.Kasteleyn, The statistics of dimers on a lattice : I. The number of dimer arrangements on a quadratic lattice. *Physica*, 1961, Vol. 27, Iss. 12.

[K15] K.K.Kozlowski, On condensation properties of Bethe roots associated with the XXZ chain. *Communications in Mathematical Physics*, 2018, Vol. 357.

[L67] E.H.Lieb Residual entropy of square ice *Physical review*, 1967, Vol. 162, Iss. 1.

[M11] T.Meyerovitch Growth-type invariants for Zd subshifts of finite type and classes arithmetical of real numbers *Inventiones Mathematicae*, 2011, Vol. 184, Iss. 3.

[PS15] R. Pavlov and M. Schraudner, Entropies realizable by block gluing shifts of finite type. *Journal d'Analyse Mathématique*, Vol. 126, Iss. 1.

[PW18] K.Petersen and B.Wilson Dynamical intricacy and average sample complexity. Arxiv 2018.

[Rob71] R. Robinson, Undecidability and nonperiodicity for tilings of the plane. *Inventiones Mathematicae*, 1971, Vol 12.

[R90] D.J. Rudolph, $\times 2 \times 3$ invariant measures and entropy. *Ergodic theory and dynamical systems*, 1990, Vol. 10, Iss. 2.

[Shannon] C. Shannon, A mathematical theory of communication. *Bell System Technical Journal*, 1948, Vol. 27, Iss. 3.

[TSE] G.Tononi and O.Sporns and M.Edelman, A measure for brain complexity: relating functional segregation and integration in the nervous system. *Processings of the National academy of science of United states of America*, !994, Vol. 91, Iss. 11.

[K17] V.Kreinovich Why some physicists are excited about the undecidability of the spectral gap problem and why should we. *Bulletin of EATCS*, 2017, Numb. 122.

[CPGW'] T. Cubitt and D.Perez-Garcia and M.M.Wolf, Undecidability of the spectral gap. *Nature*, 2015, Vol. 528.

[CPGW] T. Cubitt and D.Perez-Garcia and M.M.Wolf, Undecidability of the spectral gap. Arxiv 2015.

[Weslake] B. Weslake, A partial theory of actual causation. *Journal for the Philosophy of Science*, 2015.