

# The ‘Life Machine’: A Quantum Metaphor for Living Matter

Mario Rasetti<sup>1,2</sup>

Received: 6 June 2016 / Accepted: 1 November 2016 / Published online: 21 November 2016  
© Springer Science+Business Media New York 2016

**Abstract** Aim of this paper is to provide a scheme for the construction of a conceptual, virtual *machine* (the term has here a significance analogous to that of the Turing machine, i.e., a formal device which manipulates and evolves ‘states’), able to perform all that *living matter* – as distinguished from *inert matter* – can do and inanimate matter cannot, in a setting consistent exclusively with the quantum laws. In other words, the objective is to create a theoretical construct, in the form of a conceptual framework representing and providing the operational tools of a “Life Machine”.

## 1 Philosophical Introduction

Through a famous series of lectures at the Dublin Institute for Advanced Studies in 1943, published a year later as *What is Life?* Schrödinger [1] had a major impact on the development of twentieth century biology, especially upon Crick and Watson [2] and other founders of molecular biology and on a generation of physicists and chemists who were later engaged to serve biology. Schrödinger was not breaking new ground, but he rather gathered together several strands of research and stated his questions in a terse and provocative manner. In the decades to follow, knowledge about the protein and nucleic acid basis of living systems continued to be obtained at an accelerating rate, with the sequencing of the human genome as a major landmark along this path of discovery. The *self-replicating* DNA has thus become a major metaphor for understanding all of life.

---

A fond tribute to the memory of David Finkelstein.

---

✉ Mario Rasetti  
rasetti@isi.it

<sup>1</sup> ISI Foundation, Via Alassio 11-C, 10126, Torino, Italy

<sup>2</sup> ISI Global Science Foundation, 33 W 42nd St, New York, NY 10036, USA

The world of living systems is roughly divided into two sets of entities: the *replicators*, basic objects crucial to control development, generating the fundamental level of action for natural selection, and the *interactors*, the set of molecules and structures coded by the replicators. Dawkins [3] goes as far as relegating organisms to the status of epiphenomenal gene-vehicles, “survival machines”. In time, a number of critical reactions emerged to what was perceived as an over-emphasis on nucleic acid replication; however the ever more rapid progress in gene sequencing, producing as it does fundamental insights into the relationship between genes and morphology, has added more and more important dimensions to our understanding of evolutionary phenomena in terms of this paradigm.

On the other hand, a large body of work mostly inspired by still another pillar of Schrödinger’s contribution to physics, thermodynamics, bearing on how organisms gain order from disorder through the thermodynamics of open systems far from equilibrium, influenced scientists involved in non-equilibrium thermodynamics, such as Prigogine [4], to try and understand how organisms produce their internal order while affecting their environment through not only their activities but the creation of disorder in it. The issue is the relation between the energy flow and the production of biological organization; internal order can be produced by gradients of matter and energy flow through living systems. The structures thus generated help not only to draw more energy through the system, lengthening its retention time, but also to dissipate degraded energy, or entropy, to the environment. In this perspective living systems can then be seen as an instance of a more general phenomenon of dissipative structures.

However, thermodynamics can deal only with the possibility that something occur spontaneously: whether self-organizing phenomena happen or not depends upon (casual) actual specific initial and boundary conditions as well as on the relationships among components. Seeing the cell as a thermodynamic dissipative structure cannot be considered as reducing the cell to physics, but rather as the wider physics of what Weaver [5] called ‘organized complexity’; the new physics of open systems and the dissipative structures that arise in them. The emergent, self-organizing spatio-temporal patterns observed in certain chemical reactions (Belousov-Zhabotinski [6, 7]) are also seen in biological systems; indeed, self-organizational phenomena pervade biology.

Important to such phenomena are the dynamics of non-linear interactions, as the responses in a living system can be much larger than the stimulus, and autocatalytic cycles: reaction sequences closed on themselves, in which a larger quantity of one or more original materials is made through specific processes. The catalysts in biological systems are coded in the genes of the DNA, so once more the place to start defining life is in viewing living systems as informed, autocatalytic cyclic entities that develop and evolve under the dual dictates of the second law of thermodynamics and of natural selection. Such an approach non-reductively connects the phenomenology of living systems with basic laws of physics and chemistry, namely of quantum mechanics. Others, such as Kauffman [8], intuit that an even richer physics might be needed to adequately capture the self-organizing phenomena observed in biology and speculate that a *fourth law of thermodynamics* may ultimately be required.

The tools of complexity sciences are increasingly deployed to develop better models of living systems. Yet, as Rosen [9] claims, complexity is not life itself but what he terms the *habitat of life* and we need to focus on functions and their interrelations as well. Living beings exhibit complex functional organization as well as the ability to become more adapted to their environment over generational time: these represent the true challenge to a physically-based representation largely grounded on reductionistic assumptions.

To implicitly provide an insight into the growing hierarchy of difficulty in the pathway our scheme proposes, as a tribute to David Finkelstein's deep and wide humanistic culture, we articulate it in five steps, that we name 'Circles' – inspired by the first five circles in Dante's *Inferno*. The Circles are focused: the first [in Dante's *Divine Comedy* [10], Limbo: a state of neglect or oblivion] on the rigorous statement of the terms and objectives of the problem to be approached; the subsequent three [Lust: a passionate desire, Gluttony: greed or excess, Avarice & Prodigality: having or giving on a lavish scale] on the coherent setting up of the tools needed by the different facets of the architecture; the Fifth and here last – we stop there: there is no Heresy yet! – [Wrath & Sullenness: sulkiness; petulance] on the assemblage of the parts in such a way as to cope with the objectives.

## 2 First Circle

Main aim of the First Circle is a quick discussion of the basic problem associated with *LIVING MATTER*. In this Circle we try to outline clearly the ground we need to explore: the physical reasons why living matter is so manifestly different from lifeless matter, in spite of the evidence that both obey the same set of physical laws. The origin and nature of life are unavoidably dependent on the writing and reading of hereditary records at single molecule level, as well as on the sharp distinction between the genetic description and the phenotypic construction processes.

Living matter is first of all distinguished from non-living matter by its collective behavior in the course of time. We know from the experiments of molecular biology that there is almost certainly no microscopic or short-time interaction (or reaction) within living cells that does not follow the laws of ordinary physics and chemistry, i.e., of quantum mechanics. But what is qualitatively exceptional about living matter is not – or not only – the complexity of its detailed dynamics, but the time evolution of those constraints which harness these motions to execute simple collective functions. It is this simplicity of function, that emerges out of an extremely complex detailed dynamics, which implies how, beginning from a common set of dynamical laws for the microscopic motion that cannot but be those of quantum mechanics, living matter evolves hierarchies of collective order whereas non-living matter evolves essentially a collective disorder.

Under such perspective, the two crucial questions for developmental biology emerge: how do we tell when there is communication in living systems? how does developmental control depend on the meaning of such communication? These questions are particularly relevant when one tries to understand the deep origin of life, which requires a profound comprehension of the processes whereby molecules can function symbolically, that is as records, codes and signals. We need to know how a molecule becomes a message; and we want to answer these other questions as well: what is the simplest set of physical conditions that would allow matter to branch into two pathways – the living and lifeless – but under a single set of microscopic dynamical laws? and how large a system one must consider before biological function can have a meaning?

Once more, what we want to know is how to distinguish communication between molecules from the normal physical interactions between molecules, which physical theory believes to account for all their states of motion. Such distinction needs to be made at the simplest possible level, since the answer to the basic question about the origin of life cannot of course come from highly evolved organisms, where communication processes are instead clearer and more distinct. We need to know how messages are originated at the most basic level.

The leading idea is here that a molecule does not become a message because of any particular shape or structure or physical behavior of the molecule itself: a molecule becomes a message only in the context of a larger system of physical constraints, which can be thought of as a “language”.

What we aim to understand is indeed a higher level than that of conventional quantum physics: namely not that of molecular structures but that of the structure of the language they mutually communicate with. And we need both the experimental evidence and a theoretical framework to figure out where the origin of living matter’s dynamical and hierarchical organization stems out of matter tout court; of how the hierarchy of constraints giving rise to a structured “language” can actually originate from the normal physical constraints that hold molecules together, and from the laws that govern their motion irrespective of whether they belong to a living organism or to an inert piece of matter. Single, isolated cells clearly exhibit developmental control in the growth of their structure: hence messages must unavoidably be generated by the interactions of the growing cell with its own structure.

Obviously the structures and functions that we call life are certainly not a spontaneous state of matter, but rather the consequence of over three billion years of hereditary evolution. Therefore the question of the origin of life should not be identified with a better and better description of the highly-evolved details of present cells: it should rather be related with the fundamental quantum mechanical description of the very primeval mechanism of hereditary evolution that has actually generated all living cells.

For this reason, one should keep clearly in mind the true meaning of hereditary processes. Most biology texts assert that the mystery of heredity is solved at the molecular level by the structure of DNA and the laws of chemistry. The current molecular biological interpretation of hereditary transmission begins with DNA, which replicates by a template process, and then – in a similar way – passes its hereditary information to RNA, which in turn is believed to code the synthesis of proteins. Proteins function primarily as enzyme catalysts controlling the rates of specific reactions, and as structural elements in the cell. The central dogma [11] of biology asserts that hereditary information passes from the nucleic acids to the proteins, and never the other way around: for this reason, it should be natural from the molecular biologist’s point of view to think that the most primitive, fundamental hereditary reactions at the origin of life occurred in template replica of nucleic acid molecules.

This, if from the one hand departs only slightly from the traditional concept of heredity, on the other fails to fulfill the logical and physical descriptions of hereditary processes. Indeed, the traditional idea of hereditary process involves the transmission from parent to offspring of certain traits from a set of alternatives; the traits actually transmitted depending on some description of the traits recorded or remembered from earlier times. Manifestly, the crucial logical point is here that the hereditary propagation of a trait, involving a code as it does, should imply a classification process that cannot simply be the outcome of the physical laws of motion on a set of initial conditions, nor a process of selection implementing one of several possible evolution patterns. These dynamical laws depend only on the immediate past and only through the notion of a physical system able to manipulate information can be associated with the concept of memory, description, code and classification.

### 3 Second Circle

Focus of the Second Circle is the intriguing possible interplay of QUANTUM MECHANICS & LIFE-LIKE BEHAVIOR. In this Circle we argue that quantum mechanics can

naturally (i.e., without further assumption or constraint) exhibit features and behavior that are consistent with what we call ‘living’ as well as ‘intelligent’. The main limitation we need to deal with in this context is that physical objects endowed with these properties must have a certain degree of complexity and be characterized by a specific articulated architecture. Following step by step the thought process of Albert [12] the latter can be summarized in this way: the (living/intelligent) object should consist of two interacting parts: one that we simply refer to as the *system*,  $\mathcal{S}$ , which is a (complex) quantum dynamical system, characterized by a complete set of observables, interfaced with the rest of the world; the other – possibly a subset of  $\mathcal{S}$  – that we call *automaton*,  $\mathcal{A}$ . An automaton is a ‘quantum machine’, namely a device, perhaps virtual – i.e., a collection of potential processes. A device in the sense of the Turing machine, namely a mathematical black box that obeys preset instructions, endowed with mechanisms for the input and output of information, with tools to measure physical observables of the system, with an inner ‘program’ consisting of the set of rules necessary to extrapolate the future behavior of all physical systems, including itself, given their initial conditions. These rules are or can be made consistent with and only with (i.e., no additional assumption or postulate is required) quantum mechanics.

The state of  $\mathcal{S}$  can be characterized, for any given observable, say  $X$ , by the value assumed by  $X$  in that state, say  $x$ . Until  $\mathcal{S}$  remains unperturbed  $x$  does not change in time. On the other hand,  $\mathcal{A}$  has the capacity, i.e., instructions in its program, to measure, register and store the value  $x$  of  $X$ , and once such operation has been completed, one can ‘interrogate’  $\mathcal{A}$  about the value of  $X$ . If it is a good automaton,  $\mathcal{A}$  should accurately answer that  $X$  has value  $x$ . Since also  $\mathcal{A}$  must be in a specific quantum state, interrogate it about something consists in fact in measuring some specific observable of  $\mathcal{A}$ , say  $\mathcal{X}_X$ , whose measurement gives the value established for  $X$  by  $\mathcal{S}$ . Here the requirement that the automaton is ‘good’ is the property that once the measurement of  $X$  is accomplished, the state of the composed object  $\mathcal{O} = \mathcal{S} \uplus \mathcal{A}$  be the tensor product of the state of  $\mathcal{S}$  in which  $X$  assumes value  $x$  with the state of  $\mathcal{A}$  in which  $\mathcal{X}_X$  assumes the same value  $x$ .

All of this implies in turn that at the end of the measurement  $\mathcal{O}$  must be in a state in which also the observable, say  $X'$ , which in  $\mathcal{O}$  corresponds to  $X$  in  $\mathcal{S}$ , assumes the same value  $x$ . Introducing in  $\mathcal{O}$  the uncertainty variable  $\mathcal{E}_X \doteq \mathcal{X}_X - X$ , saying that the automaton prediction about  $X$  is accurate means that in the state in which  $\mathcal{S}$  lives  $\mathcal{E}_X = 0$ .

Let now  $P$  be another observable of  $\mathcal{S}$ , canonically conjugate to  $X$  (i.e.,  $[X, P] \neq 0$ ). It is a mere consequence of the linearity of quantum mechanics that the state of  $\mathcal{O}$  in which  $P'$  has value, say,  $p$ , is the normalized superposition (a ‘Schrödinger’s cat’ state) of all states corresponding to all possible values, say  $x_j$  (for  $j$  in some appropriate index set  $\mathcal{J}$ ) assumed by  $X$ ; each of which, in turn, is the tensor product of the state in  $\mathcal{S}$  in which  $X$  assumes value  $x_j$ , times the state in  $\mathcal{A}$  in which  $\mathcal{X}_X$  assumes the same value  $x_j$ . Interrogated about  $X$  in such state of  $\mathcal{O}$ , the automaton is only able to predict with some probability  $\pi_j$  but accurately the value  $x_j$ . Notice how, even though  $\mathcal{X}_X$  and  $X$  are themselves not well defined in this state, saying that the prediction of the automaton in it is certainly accurate means that if  $\mathcal{X}_X$  and  $X$  were measured in this state, even though the result of these measures will be in itself unpredictable, one can expect with certainty that the result of the measurement of  $\mathcal{X}_X$  will be equal to that of the measurement of  $X$ .

We increase now progressively the complexity of the questions posed to  $\mathcal{A}$ . Ask it first to measure  $X$  – operation that sets it in the condition to predict with certainty  $X$  – and successively to predict the result of the measurement of  $P$ . For  $\mathcal{A}$  to be able to make accurate predictions about both  $X$  and  $P$ ,  $\mathcal{X}_X$  and the homologous  $\mathcal{P}_P$  should be compatible, which means that  $X$  and  $P$  should be compatible, i.e., commute, and this of course cannot

happen as  $X$  and  $P$  are incompatible by assumption. Simultaneous predictions of the quantum automaton relative to incompatible observables of an external system cannot be both accurate: the best  $\mathcal{A}$  can do is to predict the probability of a certain result.

As a further step along our complexity reach-out pathway, let us now assume  $\mathcal{O}$  to be prepared in the state in which  $P'$  has value  $p$ , and give instruction to  $\mathcal{A}$  to measure both  $X$  and  $P'$ , i.e., something not external but relative to itself, and to make then predictions about the results of successive measurements of both. The existence of  $P'$  is simply the consequence of the fundamental property of quantum mechanics that to every state of a quantum system are associated definite values of a complete set of observables. Thus, at the end of the operation, the state of  $\mathcal{O}$  – in which simultaneously the variable representing  $P'$ , and that,  $\mathcal{P}_{P'}$ , which the automaton predicts for  $P'$ , have the same value  $p$  – will be the superposition of several states: those in which  $\mathcal{S}$  is in the state in which  $X$  has value  $x_j$  and  $\mathcal{A}$  is in the corresponding states in which  $\mathcal{X}_X$  equals  $x_j$  and  $\mathcal{P}_{P'}$  is equal to  $p$ . In this state, even though  $X$  and  $P'$  are incompatible,  $\mathcal{E}_X$  and  $\mathcal{P}_{P'}$  are both zero. In other words, automaton  $\mathcal{A}$  within a composed state  $\mathcal{O}$  with system  $\mathcal{S}$  is able to measure something relative to itself.

A statement of this form will play a crucial role in the scheme at the hearth of our discussion. Two issues are here relevant. On the one hand,  $\mathcal{A}$  appears to be itself (just like a living molecule or a human brain) a composed system, for which  $\mathcal{X}_X$  and  $\mathcal{P}_{P'}$  may be thought of as observables of two physical subsystems fully separated (e.g., two fragments of the molecular DNA chain or two distinct memory registers) yet internal to the automaton itself. In the course of the measurement operation imagined, first  $\mathcal{X}_X$  correlates with  $X$ , then  $P'$  correlates to  $\mathcal{P}_{P'}$ , which is the observable of  $\mathcal{O}$  which belongs to the *memory register* for  $X$  of  $\mathcal{S}$ . On the other hand, the crucial fact should be noticed and kept in mind that  $\mathcal{X}_X$  and  $\mathcal{P}_{P'}$  may be treated as referring to distinct physical systems does not imply that these two systems could not be easily linked within a larger system, in such a way that they can function as storage system (code, memory) of a single automaton.

This latter point can be made more pregnant imagining metaphorically (but much more than a metaphor in the context of our construct)  $\mathcal{A}$  as a complex molecule able to exist in one of two states, inert or alive, or as a human brain, to which one might attribute discrete “mental states”. The fact that  $\mathcal{X}_X$  and  $\mathcal{P}_{P'}$  refer, say, to distinct physical systems, i.e., different nucleic amino-acids constituting a ‘living’ molecule or, analogously, distinct neurons within a ‘brain’, does not imply that such molecular complexes could not be mutually connected to form genetic material, or that such neurons could not be mutually linked to generate a larger system, in such a way as to operate as matter able to live or not to live, or as the memory of a same “mind”.

Correlations between subsystems within a unique collective global system may (most certainly do) require very complex processes when physically implemented: yet, once the link has taken place, even though the resulting automaton is in fact a composed system and the subsystems that constitute it may be expected to be able to perform complex measurements one on the other, nonetheless there is a well precise sense in which the information stored (‘memorized’) in different subsystems can be considered as proper to a single, larger automaton. When the latter measures an observable like  $P'$ , as we have said it is indeed measuring something about itself. The variable labeling the state of  $\mathcal{O}$  after such measurement, say  $P''$ , equals  $p$ , and this state is a very particular one: when in it,  $\mathcal{A}$  knows accurately and is able to provide the values of both  $X$  and  $P'$ . Should one measure  $P'$  and  $X$  (in this order:  $P'$  first then  $X$ ; the two operations do not commute) and then  $\mathcal{P}_{P'}$  and  $\mathcal{X}_X$  (once more in this order), one would find that  $\mathcal{E}_{P'}$  and  $\mathcal{E}_X$  are both zero. This means that  $\mathcal{A}$  can predict, with certainty and in advance, the results of both these measurements. This is equivalent to

stating that all that and only what “acquiring knowledge” may imply for  $\mathcal{A}$  is contained in the correlations characteristic of the state of  $\mathcal{O}$  after the measurement.

There ensues another intriguing feature. Imagine a second automaton, with the same structure as the first one. Distinguish the properties of the two automata writing for the prediction of automaton  $\mathcal{A}_\ell$ ,  $\ell = 1, 2$ , relative to observable  $Q$ ,  ${}^{(\ell)}Q_Q$  and for the error operator associated with such prediction by automaton  $\ell$ ,  ${}^{(\ell)}\mathcal{E}_Q$ . Moreover, as observable  $X$  pertains to  $\mathcal{S}$ , whereas  $P'$  does not refer only to  $\mathcal{S}$  but also to the prediction made by  $\mathcal{A}$ , introduce symbols  ${}^{(\ell)}P'$  to indicate the same thing indicated by  $P'$ , but for automaton  $\ell$ . In the state of  $\mathcal{O}$  considered, where  $P''$  has value  $p$ ,  ${}^{(1)}\mathcal{E}_X$  and  ${}^{(2)}\mathcal{E}_{(1)P'}$  are compatible and both zero, whereas  ${}^{(2)}\mathcal{E}_X$  and  ${}^{(2)}\mathcal{E}_{(1)P'}$  are not, nor will they ever be, because for the second automaton both variables  $X$  and  ${}^{(1)}P'$  belong to an external system. Thus, if one interrogated now the second automaton whether or not it is or it might ever be able to predict both  $X$  and  ${}^{(1)}P'$ , the answer would be “no”, even though the second automaton could be in a condition that allows it predict  ${}^{(1)}P'$  and make probabilistic statements about  $X$ .

There emerges, in this capacity of one of the two automata – depending on how the process is performed – to be able to predict simultaneously  $X$  and  ${}^{(1)}P'$ , an element of *subjectivity*, because such capacity depends not only on the structure of the automaton but also on which of the two automata it pertains to, that is to say on the automaton’s *identity*. This can be reformulated asserting that there are combinations of facts that can in principle be predicted by an automaton only a propos of itself.

Naturally one of the automata could try and communicate to the other which are its predictions for  $X$  and  ${}^{(1)}P'$ ; but the very act of communicating such predictions would make these uncertain or even false, because the interaction necessary for the process of communication, however weak, cannot but generate correlations between  $X$  and any variable other than  ${}^{(1)}\mathcal{X}_X$ , thus inducing unpredictable variations in the value of  ${}^{(1)}P'$ .

From all of this there emerges a hierarchical structure in the set of possible measurements that one can successively require automaton  $\mathcal{A}$  to perform. A certain set of measurements, like those related to  $X$ , produce knowledge concerning only the external system  $\mathcal{S}$ . A second set, instead, such as those of  $P'$ , produce knowledge concerning not only the external system, but also about the knowledge that the automaton has of such system. A third set, as those of  ${}^{(2)}P'$ , can be thought of as knowledge of the system  $\mathcal{S}$ , namely of the information that  $\mathcal{A}$  has acquired about  $\mathcal{S}$  as well as the knowledge that about it can acquire another observer. Such hierarchy can in principle be iteratively continued to more and more complex levels: quantum states can thus be envisioned such that when it lives in one of them  $\mathcal{A}$  has accurate knowledge of a large number of observables, all mutually incompatible.

In conclusion, the rules to combine measurements – if they must have predictive value – depend in general on the category in the hierarchy to which they belong. The rules to perform the measurements in the frame of the same set are controlled by the commutation rules, i.e., they are the same rules that apply to ordinary quantum measurements on external systems. However, the rules to combine measurements from different categories, are instead a radically different question: they depend, among other features, on the identity and can concern the ‘self’ of who is performing the measurement.

It is in this hierarchy of knowledge, due to which the correlations within a stratification of levels are the instrument for both acquisition and transfer of data (i.e., information coded in the states of the complex quantum system), the key to comprehend the mechanism of life. A mechanism whereby the critical combination of an appropriate level of complexity with the peculiar laws of quantum physics allows matter – up to that level necessarily inanimate – to know its own individuality and, in the frame of it, to measure its states and eventually

to transfer such information (even though within the stringent, rigid constraints imposed by quantum mechanics, by information theory [13, 14], and – don't forget – by the second law of thermodynamics) to the unorganized matter that surrounds it, in such a form as to allow the latter to self-organize to make a copy of itself. Or, quite analogously, that specific mechanism whereby a real neural network, made of neurons, axons and synapses, can generate organized knowledge.

#### 4 Third Circle

Objective of the Third Circle is the relationship between DATA & TOPOLOGY. The problem is of course that in a system candidate to give origin to life the amount of information to be dealt with is extremely large. In that case the system is faced with the task of tackling the hard process not only of handling information, but of doing so on very large data sets and with the highest possible efficiency. A number of basic questions emerge: can high dimensional, global structures (the hierarchies of information constructed in the scheme of the Second Circle) be inferred from low dimensional, local representations? and, in the affirmative, how? and how can the necessary reduction process be implemented in such a way as to preserve the maximal information about the global structure? The content of this Circle is mostly grounded in the work of M. Rasetti and E. Merelli [15]

Constructing a virtual, conceptual machine that mimics Life requires a novel pathway to face the challenges posed above – in particular, the issue of sustaining predictions about the dynamics of complex processes through the analysis, by the tools that the machine has, of very large data sets – paving the way for the creation of new high-level query languages that allow insignificant details to be suppressed and meaningful information to emerge as ‘mined out’ correlations. The corresponding theoretical framework will be described essentially as a non-linear topological field theory, similar yet alternative to conventional machine learning or other artificial intelligence data mining techniques, allowing for an efficient analysis and extraction of information from large sets of data.

An operational scheme of this sort must have its deep roots in the inference of ‘knowledge’ from globally rather than locally coded data features, its focus being essentially in the integration of the preeminent constructive elements of topological data analysis (*facts* as *forms*). The scheme is modeled as a topological field theory for the data space, which becomes in this way the logical space of forms, relying on the structural and syntactical features generated by the formal language whereby the transformation properties of the space of data are faithfully represented. The latter is a sort of *language of forms* recognized by an automaton naturally associated with the field theory.

Decisive outcome of this is a tool to extract directly from the space of observations (the collection of data) those relations that encode the emergent features of the complex systems represented by the data; *patterns* that data themselves describe as correlations among events at the global level, result of interactions among systemic components at the local level. Complex systems global properties are hard to represent and even harder to predict, however data provide the necessary information leading to the control of the characteristics of the system properties.

The three pillars the scheme rests on, which need to operate in mutual synergy, are: i) *Singular Homology Methods*: they furnish the necessary tools for the efficient (re-)construction of the (simplicial) topological structure of the space of data which encode patterns. It enables the system to perform Topological Data Analysis, homology driven and coherently consistent with the global topological, algebraic and combinatorial architectural features of the



space of data, when equipped with an appropriate ‘measure’; ii) *Topological Field Theory* : it is the construct, mimicking physical field theories (as connected to statistical field theories), to extract the necessary characteristic information about such patterns in a way that – in view of the field non-linearity and self-interaction – might generate as well, as feedback, the reorganization of the data set itself. Construction of the *Statistical/Topological Field Theory of Data Space*, is generated by the simplicial structure underlying data space, by an ‘action’ – depending on the topology of the space of data and on the nature of the data, as characterized by the properties of the processes through which they can be manipulated; a ‘gauge group’, which embodies these same two features: data space topology and process algebra structure and the corresponding fibre (block) bundle; iii) *Formal Language Theory*: this is the way to study the syntactical aspects of languages generated by the field theory through its algebraic structure, i.e., the inner configuration of its patterns, and to reason and understand how they behave. It allows us to map the semantics of the transformations implied by the non-linear field dynamics into an automated self-organized learning processes.

These three pillars are interlaced in such a way as to allow for the identification of structural patterns in large data sets, and efficiently perform there data mining. It is a sort of new *Pattern Discovery* method, based on extracting information from field correlations, that generates an automaton that is a *recognizer* of the data language.

**Step One: Topological Data Analysis** – Main pillar of our construction is the notion of *data space* [16], whose crucial feature is that it is neither a metric space nor a vector space – a property that is unfortunately still often assumed theoretical computer science – but is a topological space. This is at the root of most of the issues of the scheme proposed: whether the higher dimensional, global structures encoding relevant information can be efficiently inferred from lower dimensional, local representations; whether the reduction process performed (*filtration*; the progressive finer and finer simplicial complex interpolation of the data space) may be implemented in such a way as to preserve maximal information about the global structure of data space; whether the process can be carried over in a truly metric-free way [17]; whether from such global topological information *knowledge*, namely correlated information, can be extracted in the form of patterns in the data set.

The basic principles of this approach stem out of the seminal work of a number of authors, Carlsson, Edelsbrunner, Harer, Zamorodian, and others, see, e.g., [18–20]. The fundamental goal is to overcome the conventional method of converting the collection of points in data space  $\mathfrak{X}$  into a *network* – a graph  $\mathcal{G} \sim \mathfrak{X}$ , encompassing all relevant local topological features of  $\mathfrak{X}$ , whose edges are determined by the given notion of ‘proximity’, characterized by parameter  $\eta$  that fixes a coordinate-free measure for ‘data distance’. The key point is that, while  $\mathcal{G}$  captures pretty well *local* connectivity data, it ignores an abundance of higher order features, most of which have *global* nature and misses its rich and complex combinatorial structure. All these can instead be accurately perceived and captured by focusing on a different object than  $\mathcal{G}$ , say  $\mathcal{S}$  (also  $\mathcal{S} \sim \mathfrak{X}$ ).  $\mathcal{S}$  is a higher-dimensional, discrete object, of which  $\mathcal{G}$  is the 1-skeleton, generated by combinatorially completing the graph  $\mathcal{G}$  to a *simplicial complex*.  $\mathcal{S}$  is constructed from higher and higher dimensional simple pieces (simplices) identified combinatorially along their faces. It is this recursive and combinatorially exhaustive way of construction that makes the subtlest features of the data set, seen as a topological space  $\mathfrak{X} \sim \mathcal{S}$ , manifest and accessible.

In this representation  $\mathfrak{X}$  has an *hypergraph* structure whose hyperedges generate, for a given  $\eta$ , the set of relations induced by  $\eta$  as a measure of proximity. In other words, each hyperedge is a “many-body” *relational* simplex, i.e., a simplicial complex built by gluing

together lower-dimensional relational simplices that satisfy the  $\eta$  property. This makes  $\eta$  effectively metric independent. Dealing with the simplicial complex representation of  $\mathfrak{X}$  by the methods of algebraic topology [21], specifically the theory of persistent homology that explores it at various proximity levels by varying  $\eta$ , i.e., filtering relations by their robustness with respect to  $\eta$ , allows for the construction of a parameterized ensemble of inequivalent representations of  $\mathfrak{X}$ . The filtration process identifies those topological features which persist over a significant parameter range as signal, whereas those that are short-lived characterize noise.

Key ingredients of this form of analysis are the homology groups,  $H_i(\mathfrak{X})$ ,  $i = 0, 1, \dots$ , of  $\mathfrak{X}$  and in particular the associated *Betti numbers*  $b_i$ , the  $i$ -th Betti number,  $b_i = b_i(\mathfrak{X})$ , being the rank of  $H_i(\mathfrak{X})$  – a basic set of topological invariants of  $\mathfrak{X}$ . Intuitively, homology groups are functional algebraic tools, easy to deal with (as they are abelian), to pick up the qualitative features of a topological space represented by a simplicial complex. They are connected with the existence of *i-holes* (holes in  $i$  dimensions) in  $\mathfrak{X}$ . ‘Holes’ simply means  $i$ -dimensional cycles which don’t arise as boundaries of  $(i + 1)$ - or higher-dimensional objects. Indeed, the number of  $i$ -dimensional holes is  $b_i$ , the dimension of  $H_i(\mathfrak{X})$  because  $H_i(\mathfrak{X})$  is realized as the quotient vector space of the group of  $i$ -cycles with the group of  $i$ -boundaries. In the torsion-free case, knowing the  $b_i$ ’s is equivalent to knowing the full space homology and the  $b_i$  are sufficient to fully identify  $\mathfrak{X}$  as topological space.

Efficient algorithms are known for the computation of homology groups [22]. Indeed, for  $\mathcal{S}$  a simplicial complex of vertex-set  $\{v_0, \dots, v_N\}$ , a simplicial  $k$ -chain is a finite formal

sum  $\sum_{i=1}^N c_i \sigma_i$ , where each  $c_i$  is an integer and  $\sigma_i$  an oriented  $k$ -simplex  $\in \mathcal{S}$ . One can define

[23] on  $\mathcal{S}$  the group of  $k$ -chains  $C_k$  as the free abelian group which has a basis in one-to-one correspondence with the set of  $k$ -simplices in  $\mathcal{S}$ . The boundary operator  $\partial_k : C_k \rightarrow C_{k-1}$ ,

is the homomorphism defined by:  $\partial_k \sigma = \sum_{i=0}^k (-1)^i (v_0, \dots, \widehat{v}_i, \dots, v_k)$ , where the oriented

simplex  $(v_0, \dots, \widehat{v}_i, \dots, v_k)$  is the  $i$ -th face of  $\sigma$  obtained by deleting its  $i$ -th vertex.

In  $C_k$  elements of the subgroup  $Z_k = \ker(\partial_k)$  are referred to as *cycles*, whereas those of the subgroup  $B_k = \text{im}(\partial_{k+1})$  are called *boundaries*.

Direct computation shows that  $\partial^2 = 0$ , simply meaning that the boundary of anything has no boundary. The abelian groups  $(C_k, \partial_k)$  form a *chain complex* in which  $B_k$  is contained in  $Z_k$ .

The  $k$ -th homology group  $H_k$  of  $\mathcal{S}$  is defined to be the quotient abelian group  $H_k(\mathcal{S}) = Z_k/B_k$ . There follows that the homology group  $H_k(\mathcal{S})$  is nonzero exactly when there are  $k$ -cycles on  $\mathcal{S}$  which are not boundaries, meaning that there are  $k$ -dimensional holes in the complex.

Holes can be of different dimensions. The rank of the  $k$ -th homology group, and the number  $b_k = \text{rank}(H_k(\mathcal{S}))$ , the  $k$ -th Betti number of  $\mathcal{S}$  introduced before, is the number of  $k$ -dimensional holes in  $\mathcal{S}$ .

Persistent homology is generated recursively, starting with a specific complex  $\mathcal{S}_0$ , characterized by a given  $\eta = \eta_0$  and constructing from it the succession of chain complexes  $\mathcal{S}_\eta$  and chain maps for an increasing sequence of values of  $\eta$ , say  $\eta_0 \leq \eta \leq \eta_0 + \Lambda$ , for some  $\Lambda$ . The size of the  $\mathcal{S}_\eta$  grows monotonically with  $\eta$ , thus the chain maps generated by the filtration process can be naturally identified with a sequence of successive inclusions.

Most invariants in algebraic topology are difficult to compute efficiently, but homology is not: it is exceptional not only because – as we have seen – invariants arise as quotients

of finite-dimensional spaces but also because sometimes they can be derived from ‘physical’ models. In standard topology invariants were historically *constructed* out of geometric properties manifestly able to distinguish between objects of different shape. Other invariants were instead obtained in physics, and were in fact *discovered*, based, e.g., on topological quantum field theory technology [24]. These invariants provide information about properties that are purely topological but that one cannot even hint, on the basis of geometric representation.

It is in this perspective that can be grounded the idea of constructing a reliable ‘physical’ scenario for data space. ‘Physical’ refers here to a coherent formal framework in the abstract space of data, where no equation is available giving the information it encodes as outcome, but capable to describe through its topology the hidden correlation patterns that link data into information. One expects that a given amount of information coded in data might return the full topology of data space. A topological, nonlinear field theory can be designed over data space whereby global, topology-related pattern structures can indeed be reconstructed, providing a key to the information encoded.

All this bears of course on how patterns must be interpreted, as it deals rather with pattern discovery than pattern recognition. It is worth to remark that in logic there are approaches to the notion of pattern that, drawing on abstract algebra and on the theory of relations in formal languages – as opposed to others that deal with patterns via the theory of algorithms and effective constructive procedures – define a pattern as that kind of structural regularity, namely organization of configurations or regularity, that one identifies with the notion of *correlations* in (statistical) physics [25]. These logical paradigms will guide also our strategy.

Persistent homology groups can be computed based on the notion of filtered simplicial complex. In the process, the simplex generated at each given step in the recursive construction is associated with the order-number of the step, a time-like discrete parameter that orders the collection of complexes, giving rise to the representation of a *process*, endowed with an inherent discrete-time characteristic dynamics. The combinatorially different ways in which one may realize the sampling of (*inequivalent*) structures in this persistence construction process, varying the complex shape, gives rise to a ‘natural’ probability measure, constrained by and consistent with the data space invariants and transformation properties.

**Step Two: from Data Topology to Data Field** – Besides the customary filtrations due to Vietoris-Rips [26], whose  $k$ -simplices are the unordered  $(k+1)$ -tuples of points pairwise within distance  $\eta$ , and to Čech [27], where  $k$ -simplices are instead unordered  $(k+1)$ -tuples of points whose  $\frac{1}{2}\eta$ -ball neighborhoods intersect, another filtration enters here naturally into play, Morse filtration.

In the case of those simplicial complexes that are manifolds, Morse filtration is a filtration by excursion sets, in terms of what for differentiable manifolds would be curvature-like data. Here it is a non-smooth, discretized, intrinsic, metric-free setting, appropriate for that wild simplicial complex which is data space, thought of as the simplicial, combinatorial analog of the Hodge construction.

Even though it appears to deal with metric-dependent features, Morse filtration is purely topological, namely it is independent on both the Morse function and the pseudo-metric adopted through the choice of the (stable) proximity parameter. Moreover, Morse theory generates a set of inequalities for alternating sums of Betti numbers in terms of corresponding alternating sums of the numbers of critical points of the Morse function for each given index. The ensuing analogy with the Hodge scheme is far reaching: simplicial Morse

theory generates the analogues of intrinsic, discrete gradient vector field and gradient flow, associated to any given Morse function  $f_M$ .

The Morse complex built out of the critical points of (any) Morse function with support on the vertex set of  $\mathcal{S}$  has the same homology as the underlying space. In present case this implies that the induced Morse stratification [28] is essentially the same as the Harder-Narasimhan [29] stratification. Thus one can construct the PL analog of local ‘co-ordinates’ at the Morse critical points and give a viable representation of the normal bundle to the critical sets. The relation between Morse and homology theory is generated by the property that the number of critical points of index  $i$  of a given function  $f_M$  is equal to the number of  $i$  cells in the simplicial complex obtained ‘climbing’  $f_M$ , that manifestly bears on  $b_i$ . Morse homology is isomorphic to the singular homology; Morse and Betti numbers encode the same information, yet Morse numbers allow us to think of and underlying true ‘manifold’.

Gromov-Hausdorff (GH) topology [30, 31] appears to be the natural tool to construct a self-consistent measure over  $\mathcal{S}$ ; Gromov spaces of bounded geometries provide in fact the natural framework to address the measure-theoretic questions posed by simplicial geometry in higher dimensions. More specifically, it allows establishing sharp entropy estimates that characterize the distribution of combinatorially inequivalent simplicial configurations.

GH topology leads naturally to the construction of a statistical field theory of data, as its statistical features are fully determined by the *homotopy* types of the space [32]. Complexity and randomness of spaces of bounded geometry can be quite large in the case of large data sets, since the number of ‘coverings’ of a simplicial complex of bounded geometry grows exponentially with the volume. A sort of “thermodynamic limit” needs to be realized in the case of living matter over the growing filtrations of simplicial complexes, more and more random. In fact, a well defined statistical field theory requires dealing with the extension of the statistical notion of Gibbs field to the case where the substrate is not simply a graph but a simplicial complex, that requires to prove that the substrate underlying the Gibbs field might itself be in some way random. This can be done resorting to Gibbs ‘families’ [33], so that the ensemble of geometric systems behaves as a statistical mechanics object. It is intriguing that there ensues also the possibility of finding a critical behavior, as diversified phase structures may emerge, entailing a sort of phase transition when the system passes from an homotopy type to another. The deep connection between the simplicial complex structure of data space and the information that such space hides encoded at its deepest levels, resides in the property that data can be partitioned in a variety of equivalence classes classified by their homotopy type, all elements of each of which encode similar information. In our metaphor, in  $\mathfrak{X}$  information behaves as a sort of ‘order parameter’.

**The Topological Field Theory of Data (TDFT)** – A single mathematical object encompasses most of the information about the global topological structure of the data space: the Hilbert-Poincaré series  $\mathcal{P}(z)$  (in fact a polynomial in some indeterminate  $z$ ), generating function for the Betti numbers of the related simplicial complex.  $\mathcal{P}(z) = \sum_{i \geq 0} b_i z^i$ , can be generated through a field theory, as it is nothing but one of the functors of the theory itself for an appropriate choice of the field action.

The most appropriate metaphor to refer to, to describe the formal setup of TDFT – naturally keeping in mind not only the analogies but mostly the deep structural differences: continuous vs. discrete, tame vs. wild, finite vs. infinite gauge group – is Yang-Mills’ field theory (YMFT) [34]. In YMFT the variables are a connection field over a manifold  $M$  (in YMFT a Riemann surface), and the gauge group  $G$ , under which the Chern-Simons’ (CS)

action (i.e., the  $(2\kappa-1)$ -form defined in such a way that its exterior derivative equals the trace of the  $\kappa$ -th power of the curvature) is invariant, is  $SU(N)$ .

As Terry Tao [35] perfectly recounts it, one may think of a *gauge* simply as a global ‘coordinate system’ that varies depending on one’s location over the ambient space. A gauge transformation is nothing but a change of coordinates consistently performed at *each* such location, and a gauge theory is the model for a system whose dynamics is left unchanged if a gauge transformation is performed on it. A global coordinate system is an isomorphism between a set of geometric or combinatorial objects in a given class and a standard reference object in that same class. In a gauge-invariant framework all geometric quantities must be converted to the values they assume in the specific representation, and every geometric statement has to be invariant under coordinate changes. When this is achieved, the theory can be cast into coordinate-free form. Given the coordinate system and an isomorphism of the standard object, a new coordinate system is simply obtained by composing the global coordinate system and the standard object isomorphism, namely operating with the group of all transformations that leave the gauge invariant. Every coordinate system can be generated in this manner, and the space of coordinate systems is thus fully identified with the isomorphism group  $G$  of the standard object. This group is the *gauge group* for the class of objects considered. It is this very general definition that allows us to still think of ‘coordinates’, however, not as one does for vector spaces but simply as an intrinsic way to identify mutual relations between objects.

Returning to the YMFT analogy, the base-space for YMFT is a smooth manifold,  $M$ , over which the connection field is well defined and which allows for a consistent definition of the action, since the curvature – exterior derivative of the connection plus the wedge product of the connection by itself – is well defined everywhere. Field equations in this case are then nothing but a *variational* ‘machinery’ that takes a symmetry constraint, invariance with respect to  $G$ , as input and gives as output a field satisfying that constraint. To do calculus with the appropriate type of field one attaches to each point  $p$  of  $M$  a vector space – a *fiber*  $\mathfrak{f}$  over that point; such that the field at  $p$  is simply an element of  $\mathfrak{f}$ . The resulting collection of objects (manifold  $M$  and fiber  $\mathfrak{f}_p$  at every point  $p \in M$ ) is a *vector bundle*. In the presence of a gauge symmetry, every fiber must be a representation of the gauge group  $G$ , and the field structure is that of a  $G$ -bundle. Atiyah and Bott [36], via an infinite-dimensional Morse theory with the CS action functional playing the role of Morse function, and Harder and Narasimhan [29], via a purely combinatorial approach, have established a formula that expresses the Hilbert-Poincaré series as a functor of the YMFT, in a form that is reminiscent of the relation between grand-canonical and canonical partition functions in statistical mechanics.

In the case at hand, the vector bundles of the differential category, have a PL-category analogue, referred to as *block bundles* [37]. These allow us to reduce geometric and transformation problems characteristic of manifolds to homotopy theory for the groups and the complexes involved. This leads in natural way to reconstruct the  $G$ -bundle moduli space in a discretized setting. This can be done also for simplicial complexes that may possibly not be manifolds. Since the homotopy class of a map fully determines its homology class, the simplicial block-bundle construction furnishes as well all necessary tools to compute the Poincaré series. In view of this, in spite of its topological complexity, data space offers a natural simple choice for the action, or more precisely for the exponentiated action: the Heat Kernel  $\mathcal{K}$ . This because the trace of the Heat Kernel’s gives just the Poincaré series [38].  $\mathcal{K}$  can be constructed over the simplicial complex as the exponential of the intrinsic (metric-free) combinatorial Laplacian [39].

More specifically: for any finite oriented simplicial complex  $K$ , one in which all simplices, except for the vertices (empty simplex), are oriented, and any integer  $d \geq 0$ , the collection of  $d$ -chains of  $K$ ,  $\mathcal{C}_d$ , is a vector space over  $\mathbb{R}$  (these chains form a group; we refer to the set of chains of a given dimension as the chain group of that dimension). A basis for  $\mathcal{C}_d$  is the collection of elementary chains associated with the  $d$ -simplices of  $K$ , so  $\mathcal{C}_d$  has finite dimension  $D_d(K)$ . If the elements of  $\mathcal{C}_d$  are looked at as the coordinates relative to this basis of elementary chains, on these coordinate set one can define the standard inner product, and the basis of elementary chains is orthonormal. The  $d$ -th boundary operator is a linear transformation  $\partial_d : \mathcal{C}_d \rightarrow \mathcal{C}_{d-1}$ .

Each boundary operator  $\partial_d$  of  $K$  relative to the standard bases for  $\mathcal{C}_d$  and  $\mathcal{C}_{d-1}$  with given orderings has a matrix representation  $\mathbf{B}_d$ . The number of rows in  $\mathbf{B}_d$  is the number of  $(d-1)$ -simplices in  $K$ , and the number of columns is the number of  $d$ -simplices. Associated with the boundary operator  $\partial_d$  is its adjoint  $\partial^*$ ;  $\partial^* : \mathcal{C}_{d-1} \rightarrow \mathcal{C}_d$ .

The transpose of the matrix for the  $d$ -th boundary operator relative to the standard orthonormal basis of elementary chains with the given ordering,  $\mathbf{B}_d^t$ , is the matrix representation of the  $d$ -th adjoint boundary operator,  $\partial^*$  with respect to this same ordered basis. The  $d$ -th adjoint boundary operator of a finite oriented simplicial complex  $K$  is in fact the same as the  $d$ -th coboundary operator  $\delta_d : \mathcal{C}^{d-1}(K, \mathbb{R}) \rightarrow \mathcal{C}^d(K, \mathbb{R})$ , under the isomorphism  $\mathcal{C}^d(K, \mathbb{R}) = \text{Hom}(\mathcal{C}_d(K, \mathbb{R})) \simeq \mathcal{C}_d(K)$ .

For  $d \geq 0$  an integer, the  $d$ -th combinatorial Laplacian is the linear operator  $\Delta_d : \mathcal{C}_d \rightarrow \mathcal{C}_d$  given by  $\Delta_d = \partial_{d+1} \circ \partial_{d+1}^* + \partial_d^* \circ \partial_d$ . As for the group  $G$ , notice that the space of data has a deep, far reaching property: it is fully characterized only by its topological properties, neither metric nor geometric, thus – as the objects of the theory have no internal degrees of freedom, but are constrained by the manipulation processes they can be submitted to – there is only one natural symmetry it needs to satisfy: invariance under all those transformations of data that don't change its topology and are consistent with the constraints.

Summarizing, the large set of data is thought of as a discrete space, the *space of data*, naturally embedded into a unique *simplicial complex*  $\mathcal{S}$ .  $\mathcal{S}$  can be parametrized by a 'proximity' parameter  $\eta$ , and this converts the data set into a variable global topological object. The collection of all such topological complexes, one for each value of  $\eta$ , is dealt with by the theory of persistent homology, adapted to this parameterized family. One finally encodes the persistent homology of the data set in the form of a parameterized version of Betti numbers. This is a plausible basic scheme to model what living matter does when treating information.

The conventional way to convert a collection of points into a global object is to use the point cloud as vertex set of a combinatorial graph  $\mathfrak{G}$  whose edges are exclusively specified by a notion of proximity,  $\eta$ . Though a graph of this sort, a *network*, captures pretty well connectivity data,  $\mathfrak{G}$  ignores the wealth of higher order features beyond clustering. Such features can instead be accurately discerned by thinking of the graph as the scaffold of another, higher-dimensional, discrete, piecewise-linear object, generated by completion of the graph itself: just the *simplicial complex*  $\mathcal{S}$  – the space built from simple pieces of increasing dimension (simplices) identified combinatorially along their faces.

Notice that on a IT scale the large data sets dealt with by our 'Life Machine' could possibly exceed reasonable bounds, and even though homology is just linear algebra, excellent algorithms are necessary to compute it. Such algorithms are based on filtered simplicial complexes. The filtrations associated with Rips (or Čech) complexes, which provide the natural setting in which to implement persistence, naturally imply that also the Morse filtration can enter into play. The ensuing conceptual scheme is the following: Morse theory

provides a set of inequalities for alternating sums of Betti numbers in terms of corresponding alternating sums of the number of critical points of the Morse function for each given index. In this context a Morse complex behaves like a differential complex built out of the critical points of the discrete Morse function, having the same homology as the underlying piece-wise linear manifold.

As *block bundles* allow us to reduce geometric and transformation problems over manifolds to homotopy theory for the related complexes and groups and provide a way to reconstruct the entire moduli space of  $G$ -bundles in the PL setting, whereas the homotopy class of the maps involved fully determines the homology class, they provide the tool for realizing a simple and efficient method to compute the Poincaré polynomials. Moreover, standard methods allow us to invert the recursion relation for the Poincaré series of the open substack of bundles in a purely combinatorial way. Thus, based on the PL approach to vector bundle theory, the Harder-Narasimhan recursion for computing Betti numbers of moduli spaces can be realized in the envelope of a quite more general, recursively computable, set of symmetries. This leads to a canonical system in the related algebras. Such system is expected to comprise a surprising amount of information on the moduli spaces of representations, in particular the *quiver* representations of generalized path algebras. What is crucial here is that once one has a bundle of this sort, by a process of ‘reverse engineering’ a topological field theory over such manifold can be in principle selected, such that the Poincaré polynomial, generating function of the Betti numbers, which in the field theory is the value of a specific functor, reproduces just the parametrized Betti numbers derived from raw data. At this point one has a reliable self-consistent field theory, whose other functors ( $n$ -point correlation functions, etc.) provide the required information about the underlying *patterns*, directly re-conducible to those of the space of data.

There is yet another subtler and more articulated reason why Topology should be the natural tool to handle large, high-dimensional, sets of data, bearing on the very basis of computation theory. The pillar of computation logic is the Church-Turing Thesis [40] : ((Any well-defined procedure that can be grasped and performed by the human mind and ‘pencil/paper’, can be performed on a conventional digital computer with no bound on memory.))

Notice that this is not a theorem but a statement of belief concerning the universe we live in. This is why there are several intuition-based approaches to Church Turing thesis:

- The *Empirical Intuition* : no one has ever given a true counter-example to the thesis; i.e., a concrete example of process that humans can compute in a consistent and well defined way, yet cannot be programmed on a computer. Then the thesis is true.
- The *Mechanical Intuition* : the brain is a machine whose components obey physical laws. As such, in principle, a brain can be simulated on a digital computer, and any of its thoughts can be computed by a simulating computer. Then thesis is true.
- The *Quantum Intuition* : the brain is a machine, but not a classical one. It has quantum mechanical features, hence there are inherent barriers to its being simulated on a digital computer. Then the thesis is false. However, it remains true if we allow for quantum computers.
- The ‘*Beyond Turing*’ *Intuition* : the brain is inherently a quantum computing machine but able to compute uncomputable (or, better, non-Turing computable) functions. Then the thesis is false. A new tool is needed: the quantum Gandy machine. A Gandy machine is conjectured to be equivalent to a set (finite? countably infinite? uncountable?) of non-linearly interacting Turing machines: an innovative Topological Quantum Field Theory is necessary to make the thesis true.

Turing Machines can compute only the *recursive functions* introduced by Gödel as a class of *computable functions* defined by recursion. Indeed they are finite state machines that operate on an infinite one-dimensional tape (divided into squares) and do symbol manipulations on it. Between steps of the computation, the machine is in one of finitely many states. It reads one of the symbols on the tape, and changes its state depending on the state it is in and the symbol on the tape that it reads, and rewrites the symbol accordingly on the basis of a fixed set of rules. Then it moves to the symbol to the right or to the left. The flaw here is that – as Turing himself showed – for these machines the crucial question (*Halting Problem*): ( $\langle\langle$  Does a universal program  $\mathfrak{P}$  exist that can take any program  $\mathcal{P}$  and any input  $\mathcal{I}$  for  $\mathcal{P}$  and determine whether or not  $\mathcal{P}$  terminates/halts when run with input  $\mathcal{I}$ ? $\rangle\rangle$ ) has negative answer: such a universal program  $\mathfrak{P}$  cannot exist.

Gandy [41, 42] was able to define four criteria for incomputability, such that if even only one of these is violated, the computation is non-Turing:

- i : No description of the TM exists using hereditary finite sets for its parts (interactive computing).
- ii : The set theoretic rank required is unbounded (limitation of hierarchy).
- iii : The TM cannot be uniquely assembled from parts of bounded size (unboundedness).
- iv : Successive states of the TM along the computation cannot be reconstructed solely through local cause-effect relations. If each region of the machine has a neighbourhood, the region's next state is not due only to the state of its neighbourhood: non locality, both in space and time.

Topological Field Theory of Data is global (non-local), unbounded (discrete and finite base space, but infinite bundle), effectively interactive (non-linearly), and if quantized has clean hierarchical constraints.

The question of computability can be grounded – through the features of the gauge group – in Topological Data Field Theory, but the latter should be quantized. For this reason the role of ‘quantumness’ also in information manipulation is the central issue of the Fourth Circle.

## 5 Fourth Circle

The Fourth Circle deals with QUANTUM INFORMATION MANIPULATION & TOPOLOGY. We aim to show that quantum physics has indeed already modeled/implemented automaton  $\mathcal{A}$ , in a form [43] whereby it is able to tackle a number of problems in topology and that satisfies all the assumptions requested: the *spin network quantum automaton* [44, 45]. In the final (Fifth) ‘Circle’ we shall show how this very far reaching *machine* (in the sense of Turing) can actually be further generalized by the extension of the formal language it is able to recognize and use, so as to be able in principle to model living and/or intelligent matter.

It is by now generally accepted that each physical theory supports computation models whose power is determined (and also limited) only by the physical theory itself. Classical physics, quantum mechanics and topological quantum field theory are believed to support a multitude of different implementations of the Turing machine (or equivalent: circuits, automata, etc.) computational model. Even if none of the proposed solution goes to the Gandy machine level, this bears as well on hard questions such as undecidability or uncomputability.

Quantum Information Theory can efficiently approach, in the frame of spin network quantum automata, hard topological or formal language problems reducing their



computational weight – classically exponential or **NP** – to polynomial (both space and time) complexity. Topological quantum computation [46] is a scheme designed to comply with the behavior of partition and correlation functions of a non-abelian topological quantum field theory, with gauge group  $G = SU(2)$ . It provides therefore the scheme to approach topological data field theory, where however the gauge group is quite more complex. The connection of such bundle is a one-form  $A$  valued in the Lie algebra  $\mathfrak{g}$  of  $G$ . The theory is fully characterized by the gauge group  $G$  and the action is the non-linear Chern-Simons-Witten action [47], endowed with a coupling constant  $\kappa$ , referred to as the level of the theory.

Due to their invariance under gauge and diffeomorphic transformations, that freeze out local degrees of freedom, partition and correlation functions of such theory share a global, ‘topological’ character. It was a seminal result of Witten the discovery that they indeed encode deep topological information. The universal model of computation, referred to as the “Spin Network Quantum Automaton” capable of solving in the additive approximation a number of **#P** problems in topology and formal language theory in polynomial time, stems out of a discrete, finite version of such non-Abelian topological quantum field theory characterized by the Chern-Simons action [48]. The ‘Spin Network Quantum Automaton’ [49] can be thought of as an analog computer able to solve efficiently a wide variety of hard problems.

Among the various models of quantum computation, quantum finite states automata are universal. They are defined as 5-tuples  $\{\mathcal{H}; \mathcal{L}; \mathbf{U}; |s_i\rangle; |s_{acc}\rangle\}$ , where  $\mathcal{H}$  is a (finite dimensional) Hilbert space,  $\mathcal{L}$  is the language used to provide inputs to the automaton and  $\mathbf{U}$  is a set of transition rules which describe unitarily the evolution of the automaton.  $|s_i\rangle$  is the initial state of the automaton and  $|s_{acc}\rangle$  its final (accepted) state. Such model can be thought of as a Turing machine whose tape is constrained to move only in one direction. The states of the automaton coincide with the internal states of the machine, and  $\mathcal{L}$  is generated by the alphabet  $\mathfrak{A}$  used to write the symbols on the tape. The transition rules are a set of unitary operators, one for each ‘letter’ of  $\mathfrak{A}$ , to be applied whenever the automaton reads on the tape the corresponding symbol. In the ensuing context the transition rules are but unitary representations of words, namely finite sequences of symbols, in the alphabet. A word is accepted by the automaton with probability  $p$  if  $p$  is the quantum probability, i.e., the square absolute value for the evolution amplitude from the initial to the final state represented by that particular word in the unitary representation adopted.

The Spin-Network model of quantum computation exploits the whole tensor algebra associated with the (binary) coupling and recoupling theory of  $SU(2)$  quantum angular momenta. The model has been generalized, from *Simulator* to *Automaton* to embrace the tensor (co-)algebra associated with the quantum group  $[SU(2)]_q \sim U_q(\mathfrak{sl}(2))$ , necessary as gauge group in the associated topological quantum field theory.

The underlying idea is simple. Given  $n$  quantum angular momenta with assigned, fixed sum  $\mathbf{J}$ , the computational space, i.e., the finite dimensional Hilbert space in which computations are performed by means of unitary evolution operators, is naturally characterized as a graph  $\mathfrak{G}_n$  with finite number of vertices, corresponding to computational blocks, and a set of edges corresponding to the allowed elementary unitary evolutions (gates) connecting different blocks. A computation in the frame of this graph is naturally and straightforwardly interpreted as the evolution of a quantum finite state automaton.

The initial state of the automaton is a particular vector in the fiber bundle over  $\mathfrak{G}_n$ . The transition rules of the automaton – that generate the unitary processing of the input word – can then be easily recast into sequences of elementary unitary gates. A projective measurement on the final state of the automaton will provide the probability of acceptance

for the input word. Thus, on the spin-network graph a particular computation can be seen as a path (i.e., a sequence of edges) starting from the vertex corresponding to the initial state  $|s_i\rangle$  and ending into the vertex corresponding to the accepted final state  $|s_{acc}\rangle$  in  $\mathfrak{G}_n$ .

The algebraic content of the model bears on the different ways in which it is possible to decompose the tensor product of the spaces supporting irreducible representations of  $SU(2)$  ( $U_q(s\ell(2))$  for the spin network quantum automaton) angular momenta into a direct sum of irreps. Roughly speaking, each computational block of the spin-network represents a particular way of combining pairwise irreducible subspaces (one for each angular momentum  $\mathbf{J}_\ell$ ,

$\ell = 1, \dots, N$ ;  $\sum_{\ell=1}^N \mathbf{J}_\ell = \mathbf{J}$ ) of the Hilbert space associated with the given  $J$  (the eigenvalue

of  $\mathbf{J}^2$  being equal to  $J(J + 1)$ ). As for the elementary unitary operations to be associated with the edges of the spin-network graph, they can be realized in a combinatorial way by noticing that each computational block is actually a binary tree, whose leaves are labeled by the irreps of the incoming spins, while the root is labeled by the quantum number  $J$ . Any operation/transformation one can perform on such tree can be reduced to the application of a set of elementary moves [50] which are of one of only two possible types: the *twist* operation, that simply swaps two nearest Hilbert subspaces in the tensor product of the total Hilbert space, and the *rotation* operation, which changes the binary coupling structure of the concurrent Hilbert spaces in minimal way. The twist amounts to modifying the computational states by a phase factor, whereas the rotation is related to the unitary transformation implemented by an  $SU(2)$  (respectively  $U_q(2)$ )  $6j$ -symbol ( $6j_q$ ).

Notice that when one switches to  $U_q(2)$ , due to the breaking of symmetry between Hilbert spaces induced by the co-product, the basic element of the graph – the single three-valent elementary vertex – is turned into a topological object, a sphere with three holes referred to as *pants*, and the generation of the full graph by gluing basic elements becomes a sequence of *cobordism* operations. Moreover, with the  $q$ -deformed counterpart of the  $6j$ -symbol coming into play, the twist has a natural (unitary) generalization which accounts for the two basic operations associated with over/under crossings of braids and link diagrams. It is this feature that fully characterizes the topological structure of the Spin Network Quantum Automaton.

The fibered-graph  $\mathfrak{G}_n$  of the computational-space exhibits the same combinatorial properties of the related  $SU(2)$  Spin Network Quantum Simulator, because the combinatorial features of  $6j$  and  $6j_q$  coefficients are the same. However, on top of these, it benefits of the topological sensitivity – braids vs. permutations – induced by the deformation of the gauge group. The computational capacities derive indeed from the basic rules of quantum angular momenta composition, enriched by the braiding structure induced by the gauge group deformation. The architecture of the spin network quantum automaton can be summarized as the joint effect of the combinatorial setting associated with the simulator recoupling scheme, with superimposed an  $U_q(s\ell(2))$  fiber. It is in fact just the fiber bundle over the computational graph that encodes all possible computational Hilbert spaces as well as all possible unitary gates for any fixed number  $N = n + 1$  of incoming angular momenta. The cardinality of  $\mathfrak{G}_n$  is  $(2n)!/n!$ .

There is a one-to-one correspondence between the vertices of  $\mathfrak{G}_n$  and the computational Hilbert spaces of the automaton. For any given pair  $(n; J)$ , all the binary coupling schemes allowed involve the  $n + 1$  angular momentum quantum numbers  $J_1, \dots, J_{n+1}$  as well as  $k_1, \dots, k_{n-1}$  (corresponding to the  $n - 1$  intermediate angular momenta). The latter, together with the brackets defining the binary couplings, provide the alphabet in which quantum information is encoded (the rules and constraints of bracketing are instead part of the ‘syntax’ of the resulting coding language). The Hilbert spaces thus generated, each

$(2J + 1)$ -dimensional, are the simultaneous eigenspaces of the squares of the mutually commuting  $2(n + 1)$  hermitian, angular momentum operators giving fixed sum  $\mathbf{J}$ , of the intermediate angular momentum operators and of the operator  $J_z$  (projection of the total angular momentum  $\mathbf{J}$  along the quantization axis  $z$ ).

Hilbert spaces corresponding to different bracketing schemes, although isomorphic, are not identical since they actually correspond to partly different complete sets of physical observables. On the mathematical side this reflects the fact that the tensor product is not an associative operation. The quantum numbers  $J_\ell$  can be integers or half-(odd integers), while the range of the  $k_r$  is constrained by Clebsch-Gordan decompositions. As for unitary operations acting on the computational Hilbert spaces, the unitary transformations associated with recoupling coefficients ( $3nj$  symbols) of  $SU(2)$  can be split into elementary  $j$ -gates, namely Racah and phase transforms. A Racah transform is defined as  $\mathcal{R} : |\cdots((ab)_d)_f \cdots ; JM\rangle \rightarrow |\cdots(a(bc)_e)_f \cdots ; JM\rangle$ ; where  $a, b, c, \dots$  denote generic – that is to say, both incoming ( $J_\ell$ 's) and intermediate ( $k_j$ 's) – spin quantum numbers. The effect of a phase transform amounts instead to introducing a phase factor whenever two angular momentum labels are swapped  $\mathcal{P} : |\cdots(ab)_c \cdots ; JM\rangle$  to  $(-)^{a+b-c} |\cdots(ba)_c \cdots ; JM\rangle$ . These unitary operations are combinatorially encoded into the edge set of  $\mathfrak{G}_n$ .

The fundamental property of  $\mathfrak{G}_n$  originates in the compatibility conditions satisfied by the  $6j$  symbols: the Racah and the Biedenharn-Elliott [50] identities, and the orthogonality conditions. The latter ensure that any simple path in  $\mathfrak{G}_n$  with fixed endpoints can be freely deformed into any other (the topological genus of  $\mathfrak{G}_n$  is zero, contrary to what happens for data space), providing identical quantum transition amplitudes at the kinematical level. A program remains to be chosen, uniquely associating any particular input state to a well defined (set of) accepted output states. A program is a collection of single-step transition rules, namely a collection of elementary unitary operations (gates; sequences of Racah and/or phase transforms). Such prescriptions amount to selecting a family of directed paths in the fiber space structure, all starting from the same input state and ending in an admissible output state. A single path in this family is associated with a particular algorithm supported by the program. An actual computation – namely the choice of families of directed paths in the simulator's computational space – breaks the invariance with respect to intrinsic time translations that holds instead at the purely kinematical level.

Those programs that employ only gates which do not imply any transport along the fiber, share their structural features with (discretized) topological quantum field theories with gauge group coincident simply with the transformation group of  $\mathfrak{G}_n$ . The combinatorial structure is here prominent owing to the existence of a one-to-one correspondence between allowed elementary operations (Racah and phase transforms) and the edge set of  $\mathfrak{G}_n$ . When working in such purely discrete modes, the spin network complies with all Feynman's requirements for a universal simulator. In view of its topological features, the Spin Network Quantum Automaton, induced by replacing in the theory the gauge group  $SU(2)$  by its deformed (quantum) version,  $U_q(2)$ , recognizes the language of the braid group. The transformation from the Spin Network Quantum Simulator to the Spin Network Quantum Automaton reduces essentially to *cobordisms* and *pant* decompositions the purely topological operations of state transformation, instead of mere addition of quantum angular momenta and sum of the corresponding complex amplitudes (analogous to a Feynman path sum): the combinatorial structure is fully maintained.

Notice that the Biedenharn-Elliott identity, proper to the  $6j$  symbols (both conventional and 'quantum'), leads to pentagonal relations among the  $6j$  symbols. The Racah identity instead generates triangular relations, while the Orthonormality Condition (Catalan trees

with pairs of labels identified) gives rise to rectangular relations. This allows us to identify the global structure of the Spin Network Quantum Automaton as that of a Closed Symmetric Braided Monoidal Category [46].

Summarizing, the Spin Network Quantum Automaton is endowed with all the characteristics necessary to make it suitable as formal support of a ‘quantum life machine’ as it is pursued here: i) it recognizes the language of the Braid Group  $\mathfrak{B}_n$ ; ii) it can be thought of as a quantum recognizer in the sense of Wiesner and Crutchfield [51], a particular type of finite-state quantum automaton machine; iii) the scheme described holds efficiently in the *additive approximation*. The latter is defined in this way: a quantum machine of dimension  $\mathcal{O}(\text{poly}(n))$  is said to operate in the ‘additive approximation’ if in correspondence to the algorithm performed by the unitary  $U$  over an  $n$ -qubit pure state  $|\psi\rangle$ , which can in turn be prepared in  $\mathcal{O}(\text{poly}(n))$  time, it is possible to construct a statistical ensemble such that, sampling for a  $\mathcal{O}(\text{poly}(n))$  time two random variables  $X, Y$  one has  $E[X+iY] = \langle\psi|U|\psi\rangle$ .

## 6 Fifth Circle

The Fifth Circle tries and draws the conclusions of the entire pathway; LIFE – A FIELD THEORETICAL VIEW. The hierarchy of knowledge in which correlations among a stratification of levels are the effective instrument of acquisition and transfer of information/knowledge is embedded in data in turn encoded in the complex quantum states of the system. This is the key to comprehend that particular mechanism whereby the critical combination of an appropriate amount of complexity with the peculiar laws of quantum physics allows matter to know its own individuality and in the frame of it to measure the whole complex variety of its states, and to transfer eventually such information – within the constraints imposed by physical laws – to the unorganized matter that surrounds it, in such a form as to allow the latter to self-organize and make a copy of it. Can this be modeled in our field theoretical picture? And, in the affirmative, is it true that the range of functions of the brain, from simple memory up to self-consciousness can be fitted consistently in the same scheme?

Our final hypothesis is that living matter should be represented by a quantum automaton theory, generalized in such a way as to go beyond Turing’s scheme, endowed with a specific, very general fundamental symmetry. Out of such symmetry is expected to emerge the capacity of replicating themselves of the ‘objects’ of a world made of quantum entities. In other words, a quantum field theory, topological in its deep nature, with an action and a gauge symmetry capable of encompassing the enormous complexity of the task faced. Such generalized virtual machine is conceived, in a certain way, as something analogous to a von Neumann [52, 53] *constructor* and more specifically to its far reaching recent generalizations by D. Deutsch and C. Marletto [54, 55], but inherently quantum and topological field theory based. The associated symmetry, thought of in the inescapable PL representation, is what makes the system able to extract and propagate its own long-lived topological features, providing as it does the basic ingredient to equip the space of information topology with the fiber bundle structure necessary to support the appropriate field dynamics. In this sense such symmetry must be universal. A good, though most probably not exhaustive, inspiration comes from the work of Schlesinger [56]. In it universality consists in all features of the field being represented by ‘modular functors’. In this case, at least in so far as the state manifold can be thought of as a PL handlebody, the symmetry can be expected to be described by the group  $\mathcal{G}_{\mathcal{G}-\mathcal{T}}$  of Grothendieck-Teichmüller [57, 58], realized by a particular algebra

$\mathfrak{T}_p$ , the “algebra of periods” of Tate [59, 60] (defined over the rationals,  $\mathbb{Q}$  – sufficient for a complete information about ‘holes’ contained in the homology structure of the base space).

The main ingredients of the symmetry are here the algebra of periods and the ( $q$ -deformed, with  $q$  a root of unity) group  $\mathfrak{G}_{\mathcal{G}-\mathcal{T}}$ . Our conjecture stems from the fact that they act in a way, which is analogous to the relation between the arithmetic properties of the Galois group and its action on the geometry of fundamental groups. These two conceptual tools generate a sort of new number field; the ‘*natural numbers*’ of living matter. Notice that such symmetry over the space of quantum deformations of some finite-dimensional manifold, say  $\mathcal{M}$ , has a non-trivial realization linked with the action of  $\mathfrak{G}_{\mathcal{G}-\mathcal{T}}$  over the extended moduli space, that suggests a strong analogy of our theory with conformal quantum field theories.

A symmetry over the space of quantum deforms of finite-dimensional configuration manifold  $\mathcal{M}$ , in the form of a quotient action of the *motivic* Galois group has a non-trivial realization, linked with the action of  $\mathfrak{G}_{\mathcal{G}-\mathcal{T}}$  over the extended moduli space of certain generalized quantum field theories. The reason why these can be observed only when  $q$  is a root of unity is because then the center of  $U_q(\mathfrak{sl}(2))$  is much more extended and is strongly non-trivial with respect to the case of generic  $q$  (and even more so for  $q = 1$ ).

The depth of these rules has unimaginable reach. First, it implies that actually the entire supporting structure not only of living matter, but possibly of the complex organisms it generates and perhaps of their intelligence, is affected by quantization by deformation. Indeed in this picture even space and time themselves can no longer be thought of as an immutable stage over which events take place, because in turn they are directly involved in the process of quantization, which deforms the structure that every classical observer intuitively attributes to them.

We believe that it is the deformation by quantization of the finite-dimensional manifold  $\mathcal{M}$  – “ambient space” of living matter – which gives rise to the infinite-dimensional space necessary to realize the action of  $\mathfrak{G}_{\mathcal{G}-\mathcal{T}}$ . Also the algebra  $\mathfrak{T}_p$  itself is generated by this process of quantization by deformation. Thus, even if automaton  $\mathcal{A}$  – that is a quantum computer – is thought of as derived by a procedure of quantization by deformation of a classical model, the algebra of numbers physically natural for it is not ordinary arithmetic, but a wider structure, indeed just  $\mathfrak{T}_p$ . By this observation, we intend that while in a conventional quantum computer only the rationals can be realized in a natural way, it is the periods of  $\mathfrak{T}_p$  that emerge when an observer wants to probe non-local sub-systems, namely verify the long-time behavior of the ‘Life Machine’ for all possible types of input data, exploring essentially all its global topology. In other words, the ‘*numbers*’ of  $\mathfrak{T}_p$ , which do not have any classical analogue, emerge from an operation similar to what computer scientists refer to as “testing of the (here quantum) software”. In the language of formal logic they belong to a meta-level, as they refer to tests of the computer performed from outside, not to the system itself.

In a conceptual scheme of this sort not only there is no contradiction with the perspective proper to quantum information, but the latter is encompassed in a much wider and more general structure, because also testing a device is, under any point of view, a possible, permitted and necessary physical procedure.

One can finally thus concisely synthesize the perspectives derived from the proposed framework: i) all physical systems, including living ones, must be representable in real time by a quantum computer; the quantum automaton. Vice-versa quantum computers should be describable as physical systems which realize the quantization by deformation of some classical machine; ii) the right observables of the living physical world are those and only

those which can be determined by the observation of quantum computers by quantum computers.

It is this set of dynamical properties, together with the structural ones characterizing the automaton, that arguably provide the complete conceptual reference frame to answer the question posed at the beginning, giving a potential solution of principle to the problem of consistently describing living matter, i.e., all its properties and functions, only in terms of quantum physics and complex systems science.

These are the premises out of which our *Life Machine* must grow and evolve.

**Acknowledgments** Besides the legacy of the several unforgettable discussions with David, I wish to acknowledge the interminable exchanges of ideas of the past with Michael Conrad and Tullio Regge, both unfortunately no longer with us, and the lively fruitful debates with Emanuela Merelli and Chiara Marletto.

## References

1. Schrödinger, E.: *What is Life?* Cambridge University Press, Cambridge UK (1944)
2. Watson, J.D., Crick, F.H.C.: Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953)
3. Dawkins, R.: *The Selfish Gene*, 3rd edn. Oxford University Press, Oxford UK (2006)
4. Prigogine, I., Nicolis, G., Babloyants, A., *Thermodynamics of Evolution*, (part I.): *Physics Today* **25**, 23–28 (1972). & (part II): *Physics Today* **25**, 38–44 (December 1972)
5. Weaver, W.: Science and complexity. *American Scientist* **36**, 536–544 (1948)
6. Belousov, B.P.: Periodically acting reaction and its mechanism, Collection of abstracts on radiation medicine **147**, 145 – in Russian (1959)
7. Zhabotinsky, A.M.: Periodic processes of malonic acid oxidation in a liquid phase (a study of the kinetics of the Belousov reaction). *Biofizika* **9**, 306–311 (1964). – in Russian
8. Kauffman, S.: *Investigations*. Oxford University Press, Oxford UK (2000)
9. Rosen, R.: *Essays on Life Itself*. Columbia University Press, New York NY (1998)
10. Alighieri, D.: *The Divine Comedy* (English translation H.F. Cary), at <https://www.gutenberg.org/files/8800/8800-h/8800-h.htm/> ; last release (November 30, 2012)
11. Crick, F.H.C.: Central dogma of molecular biology. *Nature* **227**(5258), 561–563 (1970)
12. Albert, D.Z.: On quantum mechanical automata. *Phys. Lett.* **98**, 249–252 (1983)
13. Hartley, R.V.L.: Transmission of information. *Bell Syst. Tech. J.* **7**, 535–563 (1928)
14. Shannon, C.E.: *The Mathematical Theory of Communication*. University of Illinois Press 1998, Urbana IL (1998)
15. Rasetti, M., Merelli, E.: Topological Field Theory of Data: mining data beyond complex networks?. In: Contucci, P.L., et al. (eds.) *Advances in Disordered Systems, Random Processes and some applications*. Cambridge University Press 2016, Cambridge UK
16. Hopcroft, J., Kanna, R.: *Foundations of Data Science* (book available on web: <http://blogs.siam.org/the-future-of-computer-science/>) (2013)
17. Petri, G., Scolamiero, M., Donato, I., Vaccarino, F.: Topological Strata of Weighted Complex Networks. *PLoS ONE* **8**(6), e66506. doi:[10.1371/journal.pone.0066506](https://doi.org/10.1371/journal.pone.0066506)
18. Carlsson, G.: Topology and data. *Bull. AMS* **46**, 255–308 (2009)
19. Edelsbrunner, H., Harer, J.: *Computational Topology, an Introduction*, Amer. Math. Soc. 2010, Providence RI
20. Zomorodian, A.J.: *Topology of Computing*. Cambridge University Press, Cambridge UK (2009)
21. Hatcher, A.: *Algebraic Topology*. Cambridge University Press, Cambridge UK (2002)
22. Basu, S., Pollack, R., Roy, M.-F.: *Algorithms in Real Algebraic Geometry*. Springer-Verlag, New York (2006)
23. Hilton, P.J., Wylie, S.: *Homology Theory: an Introduction to Algebraic Topology*. Cambridge University Press, Cambridge UK (1967)
24. Witten, E.: *Quantum field theory and the Jones polynomial*, vol. 121 (1989)
25. Shalizi, C.R., Crutchfield, J.P.: Computational Mechanics: Pattern and Prediction, Structure and Simplicity. *J. Stat. Phys.* **104**, 817–879 (2001)
26. Vietoris, L.: Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen. *Math. Ann.* **97**, 454–472 (1927)

27. Čech, E.: Théorie générale de l'homologie dans un espace quelconque. *Fund. Math.* **19**, 149–183 (1932)
28. Harada, M., Wilkin, G.: Morse theory of the moment map for representations of quivers. *Geom. Dedicata* **150**, 307–353 (2011)
29. Harder, G., Narasimhan, M.S.: On the cohomology groups of moduli spaces of vector bundles on curves. *Math. Ann.* **212**, 215–248 (1974/75)
30. Gromov, M.: Structures métriques pour les variétés Riemanniennes, Conception Edn. Nathan, Paris FR (1981)
31. Fukaya, K.: Hausdorff convergence of riemannian manifolds and its applications. *Advanced Studies Pure Math* **18** (1990)
32. Wilkin, G.: *Homotopy groups of moduli spaces of stable quiver representations*, *Int. J. Math.*, in press, Preprint: [0901.4156](https://arxiv.org/abs/0901.4156) (2009)
33. Diaconis, P., Khare, K., Saloff-Coste, L.: Gibbs sampling, exponential families and orthogonal polynomials. *Stat. Sci.* **23**, 151–178 (2008)
34. Yang, C.N., Mills, R.: Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.* **96**(1), 191–195 (1954)
35. Tao, T.: (blog: <http://terrytao.wordpress.com/2008/09/27/what-is-a-gauge/>) (2008)
36. Atiyah, M.F., Bott, R.: The Yang-Mills equations over Riemann surfaces. *Philos. Trans. R. Soc. Lond. Ser. A* **308**(1505), 523–615 (1983)
37. Rourke, C.P., Sanderson, B.J.: Block Bundles: I, II, III. *Ann. Math.* **87**(1), 1–28 (1968). (2) 256–278, (3) 431–483
38. Knill, O.: *The Dirac operator of a graph*, Preprint: [1306.2166v1](https://arxiv.org/abs/1306.2166v1) (2013)
39. Hodge, W.V.D.: *The Theory and Applications of Harmonic Integrals*. Cambridge University Press, Cambridge UK (1941)
40. B., Copeland, J.: The Church-Turing Thesis. In: Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*. Summer 2015 Edn. on web: <http://plato.stanford.edu/archives/sum2015/entries/church-turing/>
41. Gandy, R.O.: Church's thesis and principles for mechanisms. In: Barwise, J., Keisler, H.J., Kunen, K. (eds.) *The Kleene symposium*, pp. 123–148. North-Holland, Amsterdam ND (1980)
42. Gandy, R.O.: On the impossibility of using analogue machines to calculate non-computable functions, unpublished (1993)
43. Hopcroft, J.E., Ullman, J.D.: *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, Reading MA (1979)
44. Marzuoli, A., Rasetti, M.: Spin network quantum simulator. *Phys. Lett. A* **306**, 79–87 (2002)
45. Garnerone, S., Marzuoli, A., Rasetti, M.: Quantum automata, braid group and link polynomials. *Quantum Inf. Comput.* **7**, 479–503 (2007)
46. Freedman, M.H., Kitaev, A., Larsen, M., Wang, Z.: Topological quantum computation. *Bull. Amer. Math. Soc.* **40**, 31–38 (2002)
47. Atiyah, M.F.: Topological quantum field theories. *Publ. Math. IHES* **68**, 175–186 (1989)
48. Chern, S.-S., Simons, J.: Characteristic forms and geometric invariants. *Ann. Math.* **99**, 48–69 (1974)
49. Marzuoli, A., Rasetti, M.: Computing spin networks. *Ann. Phys.* **318**, 345–407 (2005)
50. Biedenharn, L.C., Louck, J.D.: *The Racah-Wigner Algebra in Quantum Theory*. In: Rota, G.-C. (ed.) *Encyclopedia of Mathematics and its Applications*, vol. 9. Addison-Wesley Publ. Co., Reading MA (1981)
51. Wiesner, K., Crutchfield, J.P.: Computation in finitary stochastic and quantum processes. *Physica D: Nonlinear Phenomena* **237**, 1173–1195 (2008)
52. von Neumann, J.: The general and logical theory of automata, Hixon Symposium, September 20, 1948, Pasadena, California. In: Taub, A.H. (ed.) *John von Neumann Collected Works*, vol. V. Pergamon Press, Oxford UK (1963)
53. von Neumann, J.: *Theory of Self-reproducing Automata*. In: A.W. Burks (ed.) *University of Illinois Press, Urbana ILL* (1966)
54. Deutsch, D., Marletto, C.: Constructor theory of information. *Proc. R. Soc. A* **471**, 20140540 (2015)
55. Marletto, C.: Constructor theory of life. *J. R. Soc. Interface* **12**, 20141226 (2015)
56. Schlesinger, K.-G.: On the universality of string theory. *Found. Phys. Lett.* **15**, 523–536 (2002)
57. Schneps, L.: The Grothendieck-Teichmüller group GT: a survey. In: Schneps, L., Lochak, P. (eds.) *Geometric Galois Actions*, London Math. Soc. Lecture Note Ser., vol. 242, pp. 183–203. Cambridge University Press, Cambridge UK (1997)
58. Schlesinger, K.-G.: A quantum analogue of the Grothendieck-Teichmüller group. *J. Phys. A: Math. Gen.* **35**, 10189–10196 (2002)
59. Tate, J.: Relations between  $k_2$  and Galois cohomology. *Invent. Math.* **36**, 257–274 (1976)
60. Milne, J.S.: Periods of abelian varieties. *Compos. Math.* **140**, 1149–1175 (2004)