The Hopfield model and its role in the development of synthetic biology.

Andrea Loettgers

Abstract—Neural network models make extensive use of concepts coming from physics and engineering. How do scientists justify the use of these concepts in the representation of biological systems? How is evidence for or against the use of these concepts produced in the application and manipulation of the models? It will be shown in this article that neural network models are evaluated differently depending on the scientific context and its modeling practice. In the case of the Hopfield model, the different modeling practices related to theoretical physics and neurobiology played a central role for how the model was received and used in the different scientific communities. In theoretical physics, where the Hopfield model has its roots, mathematical modeling is much more common and established than in neurobiology which is strongly experiment driven. These differences in modeling practice contributed to the development of the new field of synthetic biology which introduced a third type of model which combines mathematical modeling and experimenting on biological systems and by doing so mediates between the different modeling practices.

I. INTRODUCTION

A main characteristic of neural network models, such as the Hopfield model [3], is the application of concepts from physics and engineering in the representation of brain structures and functions. The integration of concepts from these diverse fields locates these models at the interface of neurobiology, physics, and engineering. The first part of the paper focuses on the dependency of the evaluation of neural network models on specific scientific contexts, discussing the case of the Hopfield model. As will be shown, the evaluation of this model in theoretical physics and neurobiology was dependent on the specific modeling and scientific practices of the respective fields. Theoretical physics, especially statistical physics and solid state physics had developed a specific modeling strategy in the context of critical phenomena triggered by the success of the Ising model in studying the mechanism of phase transitions. This strategy consists in focusing on the identification of the basic components and interactions which give rise to specific phenomena such as phase transitions. The strategy leaves aside details of the system even if they are known. Neurobiology, especially in the 1980's when the Hopfield model had been introduced, was very much experiment driven. Neurobiologists were "in touch" with the immense complexity and wealth of details of biological systems on a daily basis in their experiments. Being "in touch" with the biological system was a feature which they missed in the Hopfield model. The strategy which Hopfield had used in reducing the complexity of the biological neural

Andrea Loettgers is with the Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA (phone: 626-395-5969; fax: 626-449-0756; email: andreal@hss.caltech.edu).

network were alien to neurobiologists. The epistemic culture in which the Hopfield model had been developed and the epistemic culture in which this system and its function that it should represented were located were separated.

The second part of the paper will discuss how synthetic biology, with its specific modeling approach of reconstructing biological networks [6], can be understood as an attempt to mediate between the two different epistemic cultures. Synthetic biologists use biological components, genes and proteins, for the engineering of genetic networks. The engineering of genetic networks is based on genuine engineering concepts as for example feedback loops. By implementing the engineered network into cells, synthetic biologists test the performance and behavior of their models under the constraints of the biological systems. Similar to the approach taken by Hopfield, synthetic biologists try to identify those components of a genetic network that give rise to a specific function, such as the circadian rhythm. The materiality of the models and the testing of the network in the cell produce evidence for or against this specific approach of abstracting the complexity of biological systems and for or against the use of concepts from engineering and physics. This evidence goes beyond the evidence that is gained from mathematical models. Our example will be the *Repressilator*, a genetic network producing oscillations on the level of gene expressions that was introduced by Michael Elowitz and Stanislas Leibler [2]. It will be seen that the results gained by synthetic models are not conclusive. There remains an uncertainty regarding the approach and its concepts. But the models provide tools that allows scientists to ask more precise questions and that force them to get "in touch" with the biological system.

II. THE HOPFIELD MODEL AND THE INTERPRETATION OF BIOLOGICAL SYSTEMS AS COMPLEX COMPUTATIONAL SYSTEMS

Following the construction process of the Hopfield model one can make the argument that one of the main motivations in the design of the model was the attempt to concretize the view of the brain as a complex computational system. This was done by giving it a foundation in physics, which means finding the physical laws that determine the mechanism underlying and governing specific brain functions such as autoassociative memory. Hopfield wrote in an article entitled: *Physics, Biological Computation and Complementarity* [5] which is based on a talk he gave at the Niels Bohr centenary symposium in 1985:

> The laws of elementary particle physics and cosmology and the history of the universe serve

merely to determine the nature of a planetary environment. The dynamical equations of quantum mechanics and quantum electrodynamics (and their classical equivalents when appropriate) are the essential elemental laws of physics which lead to biology. Some physicists make claims that 'we shall never understand life until we understand the origins of the elementary particles'. But the real mysteries of biology lie in the way in which these dynamical laws of physics, and the substrate of electrons, photons and nuclei on which they operate, produce the complex set of counter-intuitive phenomena labeled with the term biology [5] p.52.

By drawing an analogy to disordered magnetic systems Hopfield constructed a distributed model of auto-associative memory which he introduced 1982 in a paper entitled: *Neural Networks and physical system with emergent collective computational abilities* [3]. An example of the kind of problems that can be investigated with the Hopfield model is the problem of character recognition. Hopfield's model provides a mechanism that allows just this kind of retrieval of information from incomplete data. In the model, the human brain's neurons are binary variables which form an interconnected network. The network may be storing a certain number of patterns. In the retrieval phase, one starts with some initial configuration and the network dynamics evolves towards the stored pattern which is closest to the initial configuration. The model belongs to the class of attractor neural networks. The basic structure of the Hopfield model is described in the following paragraph.

III. THE HOPFIELD MODEL

In the model, the human brain's neurons N are binary variables $\sigma_i(t)$, where $\sigma_i(t) = +1(0)$ characterizes an active (passive) neuron. The model neurons are connected by exchange couplings J_{ij}. The exchange couplings in the case of the Hopfield model are used for storing the memory items. Hopfield used the Hebb rule which states: a simultaneous activation of two connected neurons results in a strengthening of the synaptic coupling between the two neurons (Hebb 1949). This rule is formalized in the Hopfield model as follows

$$J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu, \tag{1}$$

where the ξ_i^μ are variables that describe a pattern, i.e. a given configuration of active and inactive neurons. The number of patterns stored in the network is given by p and in each pattern the number of neurons is equal to the total number of neurons in the network, N. In the Hopfield model, the postsynaptic potential $h_i(t)$ is produced by the activating signals arriving from all the other neurons to which the neuron i is connected. Mathematically, it is given by:

$$h_i(t) = \sum_{j=1}^{N} J_{ij}\sigma_j(t). \tag{2}$$

Hopfield introduced the following dynamics into his model:

$$\sigma_i(t) = \begin{cases} +1, & \text{if } h_i > 0, \\ 0, & \text{if } h_i < 0 \,. \end{cases} \tag{3}$$

In the course of this dynamics, a neuron σ_i is randomly chosen and updated according to the update rule. Under the right conditions (symmetric couplings, no self couplings, and a limited pattern number), this dynamics will drive the system from an incomplete pattern into the associated complete pattern. This property is due to the energy function which is given by:

$$E = -\frac{1}{2} \sum_{i \neq j} J_{ij}\sigma_i\sigma_j, \tag{4}$$

and which is, as Hopfield could show, a Lyapunov function which decreases in the course of the dynamics. Starting from an incomplete input pattern the dynamics will lead into the nearest-by energy minimum, the one that corresponds to the stored pattern.

When the model was first introduced in the begining of the 1980's physicists and neurobiologists reacted very differently. Neurobiologists rejected the Hopfield model whereas physicists on the other hand found the model very appealing. How can one explain these different reactions? I am going to argue that physicists and neurobiologists used different standards in evaluating the model which are rooted in their specific modeling and scientific practices.

A. Different standards for model evaluation

Taking into account the complexity of biological neural networks the Hopfield model is of a surprisingly simple structure. John Hopfield is a theoretical physicists by training. In the construction of the model he related to a specific strategy common in statistical physics and solid state physics of identifying the critical components and interactions in a complex system. This strategy consists in simplifying and abstracting the system under investigation in such a way that details are left out even if they are known in order to find the essential properties which explain the phenomena under investigation. This strategy is to a good part the outcome of the success story of the Ising model. The Ising model was introduced by Ernst Ising in 1924 and models the property of ferromagnetism as a cooperative phenomena. In the standard Ising model, the magnetic moments S_i are fixed on a rectangular lattice and can take on values $S_i = \pm 1$, corresponding to the different spin directions. Only spins located on nearest neighbor lattice points interact in the model. The interaction is cwdetermined by a single constant J, representing the interaction energy, and is a function of S_i and S_j. It is negative if S_i and S_j have the same sign and it is positive if they have opposite signs. The total interaction energy E is given by:

$$E = -\sum_{<i,j>} J S_i S_j, \tag{5}$$

where $< i, j >$ runs over all nearest neighbors and is at its minimum at maximal order of the system i.e. when all spins point to the same direction. This means that the simple Ising model has two energy minima corresponding to the cases where all spins are up or all spins are down.

The system represented by the Ising model shows what physicists call critical behavior around a critical temperature T_c, the so-called Curie temperature, where the system undergoes a phase transition from the paramagnetic, the disordered state, into the ferromagnetic, the ordered state. This behavior of magnetic systems below the critical temperature is an example of collective behavior.

The Ising model is part of the tradition in statistical physics of investigating the properties of complex systems such as crystals, gases, liquids, and magnetic systems. In this tradition, models are based on the assumption that the *interaction* between the elements are responsible for the observed phenomena such as ferromagnetism. For most of the collective behavior of such systems, the properties of the individual elements are not important. As has been already discussed by others such as R.I.G. Hughes [7], the Ising model has been used extensively in the study of critical phenomena. In this context the fact that the model shows only little correspondence to real magnetic systems was not regarded as a disadvantage. The Ising model functions as a *representation* for some class of other objects or processes in the world that share similar features [4]. The unspecific and abstract structure of the Ising model made it possible for physicists to use it for representing various systems, such as gases, alloys, and liquids, which share the property that they show critical behavior of a certain kind.

Thus, a simple model allows one to study and explain such a very basic and important properties as phase transitions and furthermore it allows one to do so in such a general fashion that the Ising model could function as a representation for other systems. This fact had an important impact on the modeling practice in this part of physics. It showed:

- Physical phenomena, such as ferromagnetism, are the result of collective phenomena. This means they result from the interactions between the components of the system and not from the specific properties of the components themselves.
- The details of the system are not important for the production of the observed phenomena.

Hopfield operated in this modeling tradition when he constructed his model by drawing an analogy to the Sherrington-Kirkpatrick (SK) spin glass model [8], [9]. The SK spin glass model is a model of a disordered magnetic system and has a very similar structure as the Ising model. The main difference spertains to the interaction. Sherrington and Kirkpatrick use competing exchange couplings for modeling the disorder in the system.

B. *The Serrington-Kirkptrick spin glass model*

In SK spin glass model, the disorder arises the fact that both ferromagnetic and antiferromagnetic couplings are present. A classical example for a spin glass is an alloy of gold and iron ($AuFe$). Here Gold, Au, a non-magnetic substance is weakly diluted with iron, Fe. Such a system does not have a regular lattice structure such as one finds in conventional crystals. The Fe atoms are randomly distributed in an Au matrix. This means that also their magnetic moments are randomly distributed. The couplings are a function of the distance between the magnetic moments $J_{ij} = J(R_i - R_j)$, with R_i and R_j denoting the positions of the spins. The values of the couplings are randomly distributed and can be positive, i.e. ferromagnetic, or negative, i.e. antiferromagnetic. The concurrent existence of ferromagnetic and antiferromagnetic couplings leads to a 'conflict' for some spins of the system, when they cannot 'satisfy' at the same time both ferromagnetic and antiferromagnetic couplings. Ferromagnetic couplings tend to align the spins, antiferromagnetic couplings make them point in opposite directions. Marc Mezard, Giorgio Parisi, and Miguel Angel Virasoro explain this phenomena by the following illustrative example:

> In a classical tragedy the scenario may be the following: there is a fight between two groups and the various characters on the scene have to choose sides. In addition they all have strong personal feelings, positive or negative, towards each other (it is a tragedy!). Some of them are friends and some of them are enemies. For simplicity we will assume that all feelings are reciprocal; otherwise the system may never reach equilibrium [...]. Let us consider three characters (A, B, and C); if A and B, B and C, A and C do like each other, there is no problem: they will all choose the same side. In a similar way, if A and B are friends and C is an enemy of both, then A and B can be on one side and C will be on the other. Frustration follows, instead, if A, B and C hate each other because two personal enemies must then fight on the same side. [10]

The competing interaction between the ferromagnetic and antiferromagnetic couplings is at the origin of the observed complex behavior in spin glasses. But as the simple example shows, this form of competition could also be the origin of complex behavior in other systems, where instead of a competing interaction one has, for example, competing goals.

The complex behavior of spin glasses appears at low temperatures after the system undergoes something like a phase transition. The systems exhibit a 'freezing transition' to a state with a new kind of 'order' in which the spins are aligned in random directions [1]. At high temperatures the system is in a paramagnetic phase. Cooling down of the substance leads to a freezing of the spins, pointing in a random direction. This transition is not an ordinary phase transition because it does not lead into an ordered phase.

Scattering with neutrons for example shows, that the material in the spin glass phases does not have a periodical order.

In 1975 David Sherrington and Scott Kirkpatrick introduced their model for spin glasses in an article entitled 'Solvable Model of a Spin-Glass' [8]. This seminal paper was followed in 1978 by a longer article 'Infinite-ranged models of spin-glasses' [9] in which they undertook a more detailed analytical and numerical examination of the properties of their model.

The basic structure of the model is relatively simple: a system of interacting binary variables where the binary variables represent the spins, S_i, or equivalently the magnetic moments. The S_i can take either value $+1$, which corresponds to spin direction 'up', or -1, which corresponds to spin direction 'down'. The coupling between two spins S_i and S_j is described by parameters J_{ij} which enter into an energy function of the form:

$$E = -\sum_{i,j} J_{ij} S_i S_j. \qquad (6)$$

The disorder is modelled by choosing the J_{ij} Gauss distributed,

$$P(J_{ij}) = \frac{1}{\sqrt{2\Pi J^2}} \exp\left[-\frac{(J_{ij} - J_0)^2}{2J^2} \right]. \qquad (7)$$

$P(J_{ij})dJ_{ij}$ is the probability that a given coupling takes a value between J_{ij} and $J_{ij} + dJ_{ij}$. Positive values of J_{ij} correspond to ferromagnetic and negative values to antiferromagnetic couplings.

A comparison between the SK and the Hopfield models shows that the two models agree in their basic structure. The Hopfield model is an inverse spin glass model in the sense that in the case of the Hopfield model the energy minima are given (the stored pattern) where in the case of the SK model the energy minima are the unknown.

In abstracting the complexity of biological neural networks Hopfield followed the modeling tradition exemplified by the Ising model. Neurobiologists soon brought forward a very basic critique: why should in biological neural networks the same mechanism be at work as in disordered magnetic systems? What do disordered magnetic systems and biological neural networks have in common? The fact that the Hopfield model was able to mimic auto-associative memory by using the competitive interactions taken from spin glass model was not enough. Experimental verification would have required measurements of the activity of single neurons in different parts of the brain which was not possible. The Hopfield model also contradicted the complexity of the biological systems which neurobiologists observed day by day in their experiments.

In contrast to neurobiologists physicists were familiar with the modeling tradition in which Hopfield developed his model. They recognized in the Hopfield model a success because it extended this modeling tradition beyond the boundaries of physics into the field of biology. By doing so it showed a possible way of relating biological phenomena to physical phenomena by some general mechanism, such as

the competing interactions observed in disordered magnetic systems.

Physicists evaluated the Hopfield model using different standards, based in the specific modeling tradition developed in the context of critical phenomena. Neurobiologists on the other hand evaluated the mode on the basis of their insight into the complexity of biological systems which they had gained in their experiments. The Hopfield model obviously contradicted this complexity but agreed with the abstraction strategy used by physicists working on magnetic systems.

Physicists and neurobiologists worked in different epistemic cultures and they missed insights into each others approaches. Physicists were not as familiar with the experimental approached taken by neurobiologists. On the other hand mathematical models were not the preferred research tools of neurobiologists.

The emerging field of synthetic biology can be understood as an attempt of reducing the gap between the epistemic cultures of physicists, engineering and modeling biological neural networks, and neurobiologists, investigating biological neural networks mainly by performing experiments.

Many synthetic biologists have a background in physics or engineering. A central point in the research practice of this new field consists in constructing physical models of biological systems based on concepts coming from physics and engineering. By using biological components and testing their performance in biological systems, synthetic biologists try to create evidence which goes beyond the mathematical model.

It is an interesting question to ask what developments led to the emergence of this approach. A comprehensive answer can not be given at this point. Further detailed examination of developments in the 1990's related to the modeling of neural networks and biological systems in general is required. Only one point will be discussed in the following, the development of specific programs or centers for computational neuroscience. We will see that the foundation of these centers were important for physicists and engineers to enter biology on a detour. They allowed them to develop the approach of engineering genetic networks for investigating the underlying mechanisms of biological functions such as DNA replication, chemotaxis, protein synthesis, and circadian clock.

C. Computational Neuroscience and the development of the synthetic approach

Models, such as the Hopfield model, with their success of being able to model biological phenomena using concepts from physics and engineering turned neurobiology and biology into very attractive fields for physicists and engineers. This development was supported by specific programs as for example the *Center for Computation and Neural Systems*, founded in 1989 with the strong support of John Hopfield at the California Institute of Technology. Its aim was to introduce physicists and engineers to neuroscience, and enable its students to use theories, methods, and techniques from physics and engineering to approach problems from neuroscience. The Center describes itself as:

[...] an interdisciplinary Ph.D. program to study problems arising at the interface between neuroscience, electrical engineering, computer science, and physics. The unifying theme of today's program is the relationship between the physical structure of a computational system (molecular, neuronal or electronic hardware), the dynamics of its operation, and the computational problems that it can efficiently solve.

Similar centers and programs were subsequently founded at many other universities. These centers and the students graduating from their programs contributed to the development of synthetic biology by starting to get the concepts from physics and engineering in touch with biology, by starting to experiment with biological systems. Standardization of biological methods and techniques especially in genetics and modelcular biology made it possible for physicists and engineers to start their own experimental work in this field.

One can argue that actually working with biological systems and performing experiments on biological systems had the effect that physicists experienced the missing touch with the natural system represented by mathematical models as the Hopfield model. As a consequence they looked for new ways of verifying the approach of using concepts from physics and engineering in the investigation of biological systems. One way is practiced in synthetic biology. It consists in engineering "simple biological system," for example the *Repressilator*, and in finding the sufficient components and interactions of a genetic network that give rise to oscillations.

IV. THE *Repressilator*

The *Repressilator* was introduced in 2000 by Michael Elowitz and Stanislav Leibler [2]. The *Repressilator* is a genetic network consisting of three transcriptional repressors arranged in a feedback loop producing oscillations on the level of gene expressions. The *Repressilator* had been constructed in the bacteria *E. coli* and the oscillations had been observed using fluorescent reporter gene.

The basic structure of the *Repressilator* is a DNA ring containing three genes: *tetR*, *lacI*, and λ *cI*, which encode the proteins TetR, LacI, and λ cI. In the Repressilator the proteins selectively bind to another's promoter and inhibit the expression of the gene and respectively the production of the protein. This arrangement of a negative feedback loop leads to temporal oscillations in the protein concentrations. These oscillations in protein production can be made visible when one of the protein products in the feedback loop is linked to a gene for making a green fluorescent protein. If the circuit is introduced into the bacteria *E. coli* the oscillation in the protein production can be observed by the blinking of the bacteria.

The specific design, components and interactions had been based on a mathematical model. The model consists of six coupled differential equations of the following form:

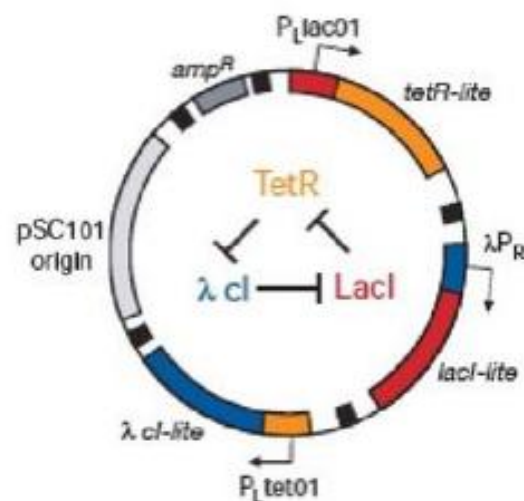$$\frac{dm_i}{dt} = -m_i + \frac{\alpha}{(1+p_i^n)} + \alpha_0, \qquad (8)$$

Fig. 1. The Repressilator network [2].

$$\frac{dp_i}{dt} = -\beta(p_i - m_i), \qquad (9)$$

with p_i denoting the three repressor concentrations, and their corresponding mRNA concentrations, m_i (where i is *lacI*, *tetR*, or *cl*. Furthermore α_0 is the number of protein copies per cell produced from a given promoter type in the presence of saturating amounts of repressor (describing the "leakiness"), and $\alpha + \alpha_0$ in its absence. β denotes the ratio of the protein decay rate to the mRNA decay rate; and n is the Hill coefficient describing the cooperativity. The analysis of the mathematical model showed that:

> [...] oscillations are favoured by strong promoters coupled to efficient ribosome binding sites, tight transcriptional repression (low 'leakiness'), cooperative repression characteristics, and comparable protein and mRNA decay rates.

Elowitz and Leibler chose the components and design of their artificial network in a way that these properties are fulfilled. By doing so they designed a network which did not allowed them to study natural network designs directly. But it allowed them to study possible network designs which may give insight as to what the sufficient components and interactions of genetic networks are and which may be used in the engineering of alternative network designs.

In order to study the oscillation of the *Repressilator* it was implemented into *E. coli* cells and the timecourse of the fluorescence of one *E. coli* was monitored under the microscope as it grew into small colonies. The oscillation did not turn out, as was hoped, to be regular. Elowitz and Leibler concluded that noise may be responsible for the observed irregular behavior and that further studies will be necessary to distinguish between intrinsic noise due to stochastic effects in gene expression and fluctuations in the amount of other cellular components.

V. CONCLUSION

The result that the engineered network did not show the behavior that was predicted by the mathematical model is an

important one. It shows that synthetic biology may have the potential to lead beyond mathematical modeling. The materiality of the synthetic model could be of crucial importance in providing evidence for or against a mathematical model and for or against the use of specific concepts from physics and engineering for describing structures of genetic networks. The approach taken by synthetic biology crucially depends on the assumption of the modular organization of biological systems. Making the assumption of a modular organization of biological systems allows for the decomposition of biological systems in functional modules. Each of these functional modules is assigned to a specific function, such as the circadian clock. The modules are assumed to be autonomous which allows one to model specific functions and to test them by introducing them into cells where the model should not interfere with other functions in the cell. If the assumption of the modular organization of biological systems is true, synthetic models allow a very controlled way of designing and testing possible design principles. The emphasis is put on possible. This means that engineering genetic networks will not lead to a conclusive answer about the underlying mechanism of specific function. But by combing mathematical modeling and experimenting on biological systems synthetic biology will mediate between the two approaches.

REFERENCES

[1] Binder, K., Staufer, D. 1979. "Monte Carlo Studies of Systems with Disorder." In *Monte Carlo Methods in Statistical Physics* (K. Binder, Ed.), Berlin-Heidelberg-New York: Springer.

[2] Elowitz, Michael, Leibler, Stanislas. 2000 "A synthetic oscillator network of transcriptional regulators", *Nature* **40**: 335.

[3] Hopfield, J. J. 1982. "Neural Networks and physical system with emergent collective computational abilities." *Proc. Natl. Acad. Sci. USA* **79**: 2554-2558.

[4] Morgan, Mary. 2003. "Experiments without Material Intervention." In *The Philosophy of Scientific Experimentation* (Hans Radder, ed.), Pittsburgh University Press.

[5] Hopfield, John J: 'Physics, Biological Computation, and Complementarity', In *The Lesson of Quantum Theory*, Proceedings of the Niels Bohr Centenary Symposium, (de Boer, D., Dahl, E, Ulfbeck, O Ed.), North-Holland, 1986, p. 295.

[6] Sprinzak, David, Elowitz, Michael. 2005 "Reconstruction of genetic circuitsA", *Nature* **438**: 442-448.

[7] Hughes, R.I.G.. 1999 "The Ising model, computer simulation, and universal physics." In *Models as Mediators*, edited by Mary Morgan and Margaret Morrison, 97-146. Cambridge University Press.

[8] Kirkpatrick, S., Sherrington, D. 1975. "Solvable Model of a Spin-Glass." *Physical Review Letters* **35**, 26 : 1792-1796.

[9] Kirkpatrick, S., Sherrington, D. 1978. "Infinite-ranged models of spin-glasses." *Physical Review B* **17** : 4384.

[10] Mezzard, Marc, Parisi, Giorgio, Virasoro, Miguel Angel. 1987. *Spin Glass Theory and Beyond* Singapore, New Jersey, Hong Kong: World Scientific.