Nobel Prize 2024: Neural networks

– DOI: https://doi.org/10.1051/2024502

The year 2024 will be remembered as a landmark: for the first time, the Nobel Prizes in Physics and in Chemistry were awarded to neural networks used for modern machine learning.

The physics prize goes to John Hopfield and Geoffrey Hinton for the development of methods that are the founda-tion of todays powerful machine learning, the chemistry prize for the use of machine learning in order to predict the spatial structure of proteins. These two simultaneous awards confirm the importance of the development of these neural networks, and their formidable impact on all branches of scientific activity. This impact might be compared to what happened when the digital methods, and with them the numerical simulations, changed many aspects of science.

How did physics play a role in this breakthrough? It is interesting to step back and trace the historical development of the main conceptual steps.

Emergence

The first notion is the idea of emergence, which originates in the study of phase transitions. In a ferromagnetic system, the energy of neighboring spins is lower when they point parallel. In a simple model of Ising spins $s_i = \pm 1$, the energy is a sum of pair interactions between neighbouring spins $E = -\sum_{ij} J s_i s_j$. When such a system is thermalized at a temperature T, the spin configurations appear with a probability given by the Boltzmann weight $P(s) \propto e^{-E(s)/(kT)}$. The well-known ferromagnetic transition occurs at a critical temperature T_c. When $T < T_c$, the majority of the spins point in one given direction, giving rise to a macroscopic ferromagnetic state which spontaneously breaks the symmetry of reversal of all spins. There are two ferromagnetic states, with magnetizations $\pm M$. A ferromagnetic state is richer than it looks at first sight. At finite T, the spins are not all aligned; they evolve in time. But in a large system, at any time, the fraction of spins pointing up is the same, equal to $(1 + M)/2$. The state thus stores collectively some information; it is very robust to changes in the microscopic properties of the spins, their interactions, or even the disappearance of some of these spins.

Multiple states and spin glasses

A ferromagnet can provide collective encoding for one bit of information (the magnetization is $+M$ or $-M$). How can one go beyond? The insight came from another development of condensed matter physics that took place in the seventies, the study of disordered magnetic materials called spin glasses. In their simplest version [3], they differ from ferromagnets by the fact that two magnetic moment s_i and s_j now have an interaction energy $-J_{ij}s_is_j$ with a coupling constant J_{ij} that explicitely depends on the pair, and which can take positive or negative values. The coexistence of these disordered ferromagnetic and antiferromagnetic interactions gives rise to a low temperature 'spin glass' phase where the spins can order in many different states. This was gradually realized by careful studies [14] of the mean field model of Sherrington and Kirkpatrick [13], which culminated in its exact solution by Parisi [12, 11].

Neural networks: designing attractors

In his seminal paper of 1982 [8], John Hopfield proposed that these properties could be used in order to build completely new types of systems with distributed computation, and that this could provide a basis for how the interactions of neurons in the brain can lead to a collective processing of in-formation. As he phrases it in his introduction: "*In physical systems made from a large number of simple elements, interactions among large numbers of elementary components yield collective phenomena such as the stable magnetic orientations [4, 6]. Do analogous collective phenomena in a system of simple interacting neurons have useful "computational" correlates?*".

In order for the many states of spin glasses to be used in information processing, one needs a way to design the energy landscape, so that the various states can encode a desired information. This can be done by carefully choosing the coupling constants J_{ij} between pairs of spins (or artificial neu-rons) in the spin-glass. The "Hopfield model" designs a memory network by a reinforcement process inspired from Hebb's work [5]: given a certain number of patterns, which are neural configurations to be stored, it reinforces the interaction J_{ij} between neurons i and j whenever they are parallel in each stored patterns. The resulting network, initialized from a noisy version of a pattern, can relax spontaneously to the stored pattern in a zero temperature dynamics where the energy is decreased at each step (see Fig. 1). The exact computation of the storage capacity using spin-glass theory methods [1] gave a solid base to this whole approach, which prepared the ground for many subsequent developments.

Boltzmann machines

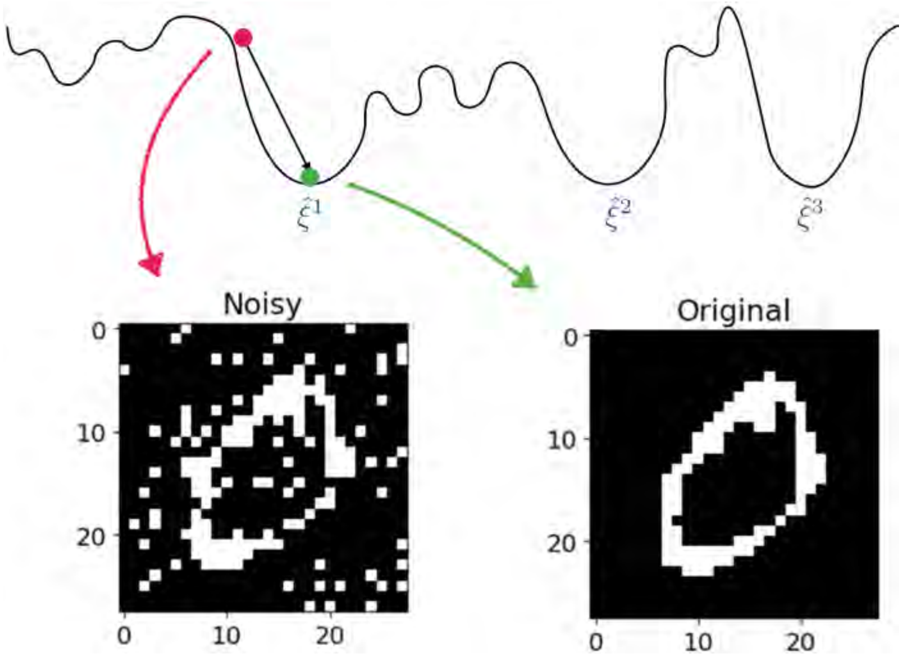
The model was soon generalized by Geoffrey Hinton and collaborators [4, 6] who introduced the "Boltzmann machine" where the dynamics is not only a decrease in energy, but leads to a Boltzmann equilibrium configuration at a finite temperature. They also introduced a difference between visible neurons, which encode the input from external world, and some "hidden neurons", which encode information that can be used for reasoning. For instance, in memorizing images of two different categories, landscapes and faces, there will be two clusters of states, each corresponding to a category, and ●●●

••• this can be encoded in the activity of the hidden neurons, opening the way to the use of Boltzmann machines for more complicated classification tasks.

The design algorithm again relied on modifying gradually the couplings J_{ij}, so that the correlations of neurons i and j get closer to the one in a given database. However this process is very slow and the practical use of Boltzmann machines was limited. Two major developments allowed to turn it into a more efficient tool: the use of restricted Boltzmann machines where the interactions are only between visible and hidden neurons, and the invention of a new learning algorithm, contrastive divergence [7], which gives a much faster way to design the coupling matrices.

Cornucopia

In a parallel development, the feedforward neural networks initiated by the work of Rosenblatt on the perceptron were developed notably for image analysis through the use of convolutional neural networks. These two lines of research on artificial neural networks, energy-based models and feedforward networks, had many interactions (for instance the pre-training of a multilayer feedforward network can be done using a sequence of restricted Boltzmann machines). The recent breakthroughs of machine learning in applications (including in many branches of physics!), based on multilayer networks and recognized in 2018 by the Turing Award to Bengio Hinton and LeCun[10], certainly have deep conceptual links to the statistical physics studies of emergent properties in systems of interacting artificial neurons initiated in the 80's. Hopfield's paper also played a major role in the birth of computational neurosciences, which has turned into a full-fledgeg field. The award this year of the Brain Prize to Larry Abbott, Terrence Sejnowski and Haim Sompolinsky, all three trained as physicists, attests of this impact. Hopfield realized that his model of interaction between artificial neurons, even if they are much simplified object, could serve as a new paradigm for brain func-tioning. Quoting him again: "... *each "neuron" has elementary properties, and the network has little structure. Nonetheless, collective computational properties spontaneously arose. Memories are retained as stable entities or Gestalts and can be correctly recalled from any reasonably sized subpart. [...] These properties [...] do not appear to be strongly dependent on precise details of the modeling. This robustness suggests that similar effects will be obtained even when more neurobiological details are added.*" This prediction has been vindicated by four decades of work in computational neuroscience. Phil Anderson once described spin glasses as a cornucopia, and this was well taken. Building on a few decades of work, statistical physics has helped to understand and control the collective behaviour of many simple elements, artificial neurons, in order to build very fancy information processing devices and open new routes towards understanding how the brain works. ■



▲ Illustration of associative memory in a Hopfield network: on top, the energy landscape. Here three patterns are stored (beware that the landscape is actually in large dimensions). Bottom: on the right, a handwritten digit, pattern number one. On the left, a noisy version of the pattern. The dynamics of the neurons is such that the neural network, initialized with noisy versions of the pattern, evolves towards the full memorized pattern. It can thus correct errors at the level of hardware (artificial neurons).

■ **Marc Mézard,**
Bocconi University, Milano

References

[1] D. J. Amit, H. Gutfreund and H. Sompolinsky, *Physical Review Letters* **55(14)**, 1530 (1985).

[2] D.J. Amit, *Behavioral and brain sciences* **18**, 617 (1995).

[3] S. F. Edwards and P. W. Anderson. Theory of spin glasses, *J. of Physics F* **5**, 965 (1975).

[4] Scott E. Fahlmann, Geoffrey Hinton, and Terrence J. Sejnowski. Massively parallel architectures for al: Netl, thistle, and boltzmann machines, In Proceedings of AAAI-83 Conference, 1 (1983)

[5] D. O. Hebb. *The organization of behavior*. Wiley and Sons, NY (1949).

[6] Geoffrey Hinton and Terrence J. Sejnowski. Optimal perceptual inference. In Proceedings ofthe IEEE Conference on Computer Vision and Pattern Recognition, 448, IEEE (1983).

[7] Geoffrey E Hinton, *Neural computation* **14(8)**, 1771 (2002).

[8] J.J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, **79(8)**, 2554 (1982).

[9] Mikail Khona and Ila R Fiete, *Nature Reviews Neuroscience* **23(12)**, 744 (2022).

[10] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton, *Nature* **521**(7553), 436 (2015).

[11] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. World Scientific Publishing Company (1987).

[12] Giorgio Parisi, *Journal of Physics A: Mathematical and General* **13(3)**, 1101 (1980).

[13] David Sherrington and Scott Kirkpatrick, *Physical Review letters* **35(26)**, 1792 (1975).

[14] David J Thouless, Philip W Anderson, and Robert G Palmer, *Philosophical Magazine* **35(3)**, 593 (977).