The Nobel Prize in Physics 2024 : A Celebration of Machine Learning's Foundations in Physics

• J. MAHAMBA

The 2024 Nobel Prize in Physics has been awarded to the Professors John J. Hopfield and Geoffrey E. Hinton *"for foundational discoveries and inventions that enable machine learning with artificial neural networks,"* acknowledging the profound impact of these *energy-based* networks on the development of machine learning and the foundational role of physics in this process. This award reflects not only the transformative influence of neural networks on artificial intelligence but also highlights the power of interdisciplinary research in revolutionizing modern science. By integrating concepts from mathematics, physics, neuroscience, and computer science, Hopfield and Hinton laid the groundwork for the neural networks and deep learning models that led to the AI revolution we are currently experiencing.

Introduction

In this article, we aim to review the basics of *Hopfield Networks*, what they are, how they work and why they are said to be one of the building blocks of recent architecture of machine learning. We'll also explore the analogy with the *Ising model* from statistical physics, which will unveil the relationship between neural networks and physical systems. Limitations of Hopfield Networks, both theoretical and practical, will be considered in the end, which will also prepare the ground for introducing *Boltzmann Machines* and more advanced neural models in future articles.

1. Quick presentation of the laureates and their work

John J. Hopfield and Geoffrey E. Hinton's contributions to machine learning in the 1980s laid the foundation for the neural networks we utilize today. Hopfield's 1982 paper, *Neural networks and physical systems with emergent collective computational abilities* [4], introduced a new class of recurrent neural networks that could function as associative memories, enabling them to recall stored information from partial or noisy inputs. Three years later, Hinton's work on Boltzmann Machines expanded Hopfield's by introducing stochastic dynamics, which added a crucial element of randomness to neural networks, allowing them to escape *undesired stable states* during learning [1].

Their pioneering work brought machine learning into the domain of physics by framing neural networks as systems that *minimize an energy function* — a concept directly borrowed from thermodynamics and statistical mechanics. The connection between neural networks and physical systems such as *spin glasses* became a textbook case for demonstrating how simple components could combine to perform complex computations, an insight that the Nobel Committee recognized by awarding them the prize in physics.

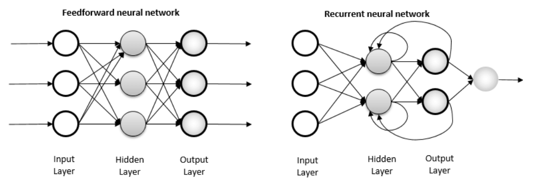
2. The Foundations of Hopfield Networks

Hopfield Networks are a type of *recurrent neural network* (RNN) designed to model *associative memory*, which refers to the ability to recall complete *patterns* from incomplete or noisy input, the same way you might recognize a song by hearing its first notes. Associative memory in Hopfield Networks functions analogously to how humans recognize a face even when some features are obscured. The key idea is that the network stores several *patterns* — the memories — and can retrieve them by "filling in" missing parts based on partial information.

2.1 – Architecture and Components

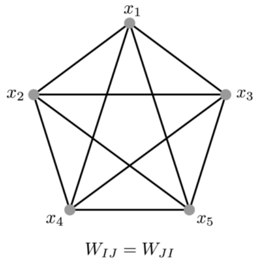
A recurrent neural network, contrasting with the classic feedforward neural networks, contains connections that cycle on themselves allowing information from previous states to influence future outputs. Unlike feedforward networks, which pass information unidirectionally from input to output, RNNs can process sequences and maintain a form of memory by feeding information from one time step into the next (*cf. Figure 1*). A Hopfield Network belongs to the class of RNN's but all the neurons in this kind of network are connected to each other and the connectivity relation is symmetrical.

FIGURE 1 – Feedforward Neural Network vs. Recurrent Neural Network. Source : ResearchGate.



But this type of networks is a bit more specific. While all the neurons in the network are connected to each other, it is important to note that only one neuron is updated at a time, receiving input from all the others (cf. Figure 2). In such a recurrent architecture that is fully connected, symmetrical and updates asynchronously, the network can attain equilibrium converging towards a stable state that represents a stored memory.

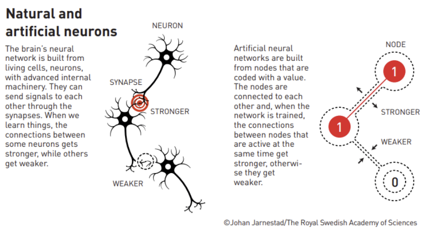
FIGURE 2 – Hopfield Network : there are actually two edges connecting each pair of neurons but the figure is simplified by symmetry. Source : Wikipedia.



A Hopfield Network consists of the following components :

- **Neurons** : These are the basic units of the network, and each neuron can take one of two states : $+1$ or -1, much like a binary switch. These states can be interpreted as "firing" or "not firing," in reference to biological neurons' activity in our brain and similar to the behavior of spins in physical systems — here "up" and "down" (*cf. Figure 3*).
- **Weights** : These are the strengths of the connections between pairs of neurons, determining how much the state of one neuron influences another. The weights are symmetric, meaning that the influence of neuron i on neuron j is the same as the influence of neuron j on neuron i ($w_{ij} = w_{ji}$).
- **Patterns** : These are the configurations of neuron states that the network is designed to store and recall. A pattern is essentially a memory, represented as a specific arrangement of $+1$ and -1 states across the network.

FIGURE 3 – Neural and Artificial Neurons. Source : The Nobel Prize



The goal of the network is to "remember" these patterns such that, when a partial or corrupted version of a pattern is given as an input to the network, it can reconstruct the full pattern through its internal dynamics.

2.2 – Learning and Storage of Patterns

To store a pattern in a Hopfield Network, the synaptic weights are adjusted such that the pattern becomes an *attractor* of the system. An attractor is a state towards which the network naturally evolves, like the center of a bowl *attracting* a marble. Hopfield proposed a rule to update the weights based on Hebbian learning, often summarized as *"neurons that fire together wire together."*

For p patterns $\{\xi^\mu\}$, where each pattern is a binary vector of length N (the number of neurons), the weights are determined by :

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$$

This learning rule ensures that the network "remembers" the patterns, meaning that they correspond to stable configurations with low energy, as we will now see.

2.3 – Energy Function and Network Dynamics

A central concept in Hopfield Networks is the *energy function*, which governs the dynamics of the network. The energy function is designed so that the network evolves naturally towards configurations with lower energy, analogous to how physical systems settle into low-energy states.

The energy of a given configuration of neurons is defined as :

$$E = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} s_i s_j - \sum_{i=1}^{N} \theta_i s_i$$

Here, s_i represents the state of neuron i ($+1$ or -1), and w_{ij} is the weight between neurons i and j. The term θ_i denotes an external bias (which can often be set to zero) which indicates the neuron's preferred state, being the one in which this neuron is in the closest pattern. The energy function behaves similarly to the *Hamiltonian* in physics, representing the total energy of a system. The network's objective is to minimize this energy function, and it does so through a series of neuron state updates.

At each update step, the network chooses a neuron to update and changes its state based on the weighted input from other neurons. The update rule is straightforward :

1. Calculate weighted sum of all the neurons linked to the neuron i :

$$h_i = \sum_{j=1}^{N} w_{ij} s_j$$

   This sum represents the net's agreement to the state of the neuron i : if the signs are the same, it means that the neuron i is in the state that minimizes the energy function, and if not, its state must be changed. Hence, we then

2. Update the state of neuron i based on the sign of this input :

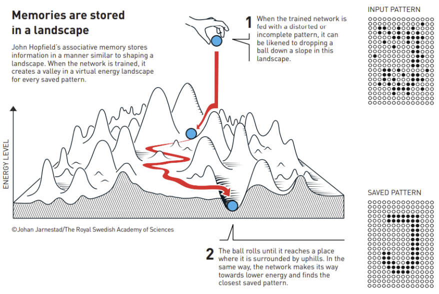
$$s_i = \mathrm{sign}(h_i)$$

This process is repeated iteratively until the network converges to a stable state, meaning no further updates will change the neuron states. This final state corresponds to a local minimum of the energy function, which is a stored pattern.

Indeed, the energy function is designed so that each neuron update reduces or maintains the overall energy of the system. Since the energy function is bounded from below, the network is guaranteed to eventually reach a state where no further updates can reduce the energy. This is called a local minimum.

In the context of memory recall, the network starts in a high-energy state when given a noisy or incomplete version of a pattern. As it updates the neurons, it "falls" into the nearest local minimum, which corresponds to the stored memory (*cf. Figure 4*). However, there can be issues where the network converges to *spurious states* — undesired configurations that are stable but do not correspond to any of the stored patterns. These spurious minima represent a limitation of the classical Hopfield model. These can appear when too many memories are stored, when two memories are too close to each other or when the energy landscape is relatively flat.

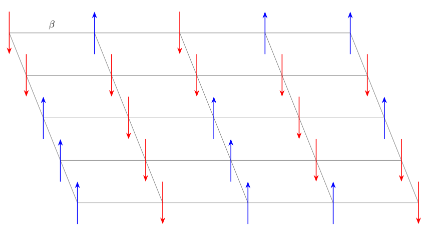
Figure 4 – Memory Storage and Retrieval in Hopfield Networks. Source : The Nobel Prize



3. Hopfield Networks and the Ising Model

Hopfield's analogy between neural networks and the *Ising model* of statistical physics is one of the most insightful aspects of his work. The Ising model describes magnetic materials by modeling interactions between spins that can be either up (+1) or down (−1) (*cf. Figure* 5) [6]. These spins, which can be understood as mini-magnets at the subatomic level, interact with their neighbors, and the system tends to settle into low-energy configurations — according to the *second law of thermodynamics* — where neighboring spins are aligned. The Hopfield Network functions in a similar way, where neurons (analogous to spins) interact through synaptic weights and settle into stable configurations that minimize the system's energy.

Figure 5 – Representation of the 2D Ising (nearest neighbor) model. Positive spins (up) are arbitrarily represented in blue, and negative spins (down) in red.



In both the Hopfield Network and the Ising model, the system can be represented as an energy landscape. Each configuration of neurons or spins corresponds to a point in this landscape, and the system evolves to lower its energy, moving towards the nearest minimum following a gradient-descent-like algorithm. In the case of Hopfield Networks, these minima correspond to stored patterns, while in the Ising model, they represent ordered magnetic states — wich are low entropy states in both cases [4].

This analogy deepens when we consider how both systems exhibit emergent behavior. In both cases, complex global patterns arise from simple, local interactions between units (neurons or spins). This connection to statistical mechanics provides a powerful framework for understanding how neural networks can perform complex tasks like memory recall and pattern recognition.

4. Significance of Hopfield Networks

Hopfield Networks have had a lasting impact on both cognitive science and machine learning. In cognitive science, they provided a formal model for associative memory, illustrating how the brain might retrieve memories based on partial input. This model captures the essence of how humans recognize faces, songs, or other complex stimuli even when presented with incomplete information.

Besides its applications in cognitive science, Hopfield networks were among the first generative neural network models in the machine learning industry and were already able to solve optimization problems through energy minimization. They showed how a collection of simple, connected elements could store and process complex information by evolving towards stable configurations that represent solutions.

However, Hopfield Networks have notable limitations :

- **Storage Capacity** : The number of patterns that can be stored in a Hopfield Network is limited to approximately $0.138N$, where N is the number of neurons [2].
- **Spurious States** : These are stable configurations that do not correspond to stored patterns, leading to incorrect memory recall.
- **Binary Neuron States** : The binary nature of neurons limits the flexibility of the model. Real-world systems often require continuous values for more nuanced control.

5. Extensions and Solutions to Hopfield's Limitations

An ever-increasing interest from the scientific community in the domain of neural networks has translated into numerous modifications or expansions of the classical model described by Hopfield. One such improvement is the Boltzmann Machine, developed by Geoffrey E. Hinton in 1985. With this machine, the neuron update rule is probabilistic rather than deterministic ; thus, it provides a stochastic dynamics to the framework — much more like what actually happens in the Ising model. This allows some neurons to jump out of a local minimum within the energy landscape and explore different directions before converging towards a final solution [1, 3].

More recent innovations include *continuous Hopfield Networks*, which extend the binary neuron states to continuous values, allowing for smoother optimization. Another advancement is *dense associative memory models*, which extend the storage capacity of Hopfield Networks exponentially by leveraging modern techniques in neural computation [8].

Looking ahead, *quantum Hopfield Networks* (or *qHops*) represent a promising new direction in this field. Following the laws of quantum mechanics, these models are able to perform associative memory tasks more effectively thanks to the properties of quantum *superposition* and *entanglement*. They have the potential to outrun some major issues of the classical Hopfield models. For example, qHops can, in theory, explore an exponentially larger state space leaving the network with superior storage capacity while being more impervious to spurious states which they could escape more easily thanks to *quantum tunneling* [7].

6. The Future of Hopfield's Legacy in Machine Learning

The work of John J. Hopfield continues to shape the landscape of machine learning today. Many modern neural network architectures, including deep learning models, can trace their conceptual roots back to energy-based models like Hopfield Networks. Innovations such as deep belief networks and autoencoders are direct descendants of this early kind of architecture [5].

As researchers explore the potential of quantum computing and information, quantum Hopfield Networks may represent the next frontier, offering the possibility of solving problems beyond the reach of classical systems. These models could provide new ways of performing associative memory tasks with enhanced flexibility, error correction and computational power.

7. Conclusion

The 2024 Nobel Prize in Physics highlights the enduring importance of Professors John J. Hopfield and Geoffrey E. Hinton's foundational work on neural networks. Hopfield's introduction of associative memory through energy minimization provided a powerful framework for understanding how networks of simple units could interact to produce complex and emergent behaviors. His work forged a bridge between statistical mechanics and machine learning, showing that insights from physics can profoundly shape the principles of intelligent systems.

While Hopfield Networks are foundational, their limitations have driven researchers to develop new models, such as Boltzmann Machines and continuous Hopfield Networks. Today, these ideas continue to influence the cutting edge research in AI, from classical deep learning to quantum neural networks. As we look forward, the legacy of Hopfield's work remains central to ongoing advancements, and it continues to inspire innovations that could fundamentally transform the future of artificial intelligence.

Références

[1] D. H. Ackley, G. E. Hinton et T. J. Sejnowski. « A learning algorithm for Boltzmann machines ». *Cognitive Science* **9**, nº 1 (1985), p. 147-169.

[2] D. J. Amit, H. Gutfreund et H. Sompolinsky. « Storing infinite numbers of patterns in a spin-glass model of neural networks ». *Physical Review Letters* **55**, nº 14 (1985), p. 1530-1533.

[3] G. E. Hinton et R. R. Salakhutdinov. « Reducing the dimensionality of data with neural networks ». *Science* **313**, nº 5786 (2006), p. 504-507.

[4] J. J. Hopfield. « Neural networks and physical systems with emergent collective computational abilities ». *Proceedings of the National Academy of Sciences* **79**, nº 8 (1982), p. 2554-2558.

[5] Y. LeCun, Y. Bengio et G. Hinton. « Deep learning ». *Nature* **521**, nº 7553 (2015), p. 436-444.

[6] J. A. Mahamba. *Mathematical Exploration of Phase Transitions in the Ising Model for Its Relevance in University-Level Statistical Physics Education*. Undergraduate Research Thesis. 2024.

[7] P. Rebentrost et al. « Quantum Hopfield neural network ». *Physical Review A* **98**, nº 4 (2018), p. 042308.

[8] M. Widrich et al. « Modern Hopfield Networks and Attention for Immune Repertoire Classification ». In : *Advances in Neural Information Processing Systems*. Vol. 33. 2020.

Jorys Mahamba

CentraleSupélec, 91190 Gif-sur-Yvette, France

jorys.mahamba@student-cs.fr

Jorys Mahamba is a master's student at CentraleSupélec with an interest in applying mathematics to the intersection of physics and computer science.