KUNGL.
VETENSKAPS-
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

8 OCTOBER 2024

Scientific Background to the Nobel Prize in Physics 2024

"FOR FOUNDATIONAL DISCOVERIES AND INVENTIONS THAT ENABLE MACHINE LEARNING WITH ARTIFICIAL NEURAL NETWORKS"

The Nobel Committee for Physics

THE ROYAL SWEDISH ACADEMY OF SCIENCES has as its aim to promote the sciences and strengthen their influence in society.

Nobel Prize® and the Nobel Prize® medal design mark
are registrated trademarks of the Nobel Foundation

KUNGL.
VETENSKAPS-
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

The Royal Swedish Academy of Sciences has decided to award
the Nobel Prize in Physics 2024 jointly to

John J. Hopfield and **Geoffrey E. Hinton**

*"for foundational discoveries and inventions that enable machine learning
with artificial neural networks"*

Introduction

With its roots in the 1940s, machine learning based on artificial neural networks (ANNs) has developed over the past three decades into a versatile and powerful tool, with both everyday and advanced scientific applications. With ANNs the boundaries of physics are extended to host phenomena of life as well as computation.

Inspired by biological neurons in the brain, ANNs are large collections of "neurons", or nodes, connected by "synapses", or weighted couplings, which are trained to perform certain tasks rather than asked to execute a predetermined set of instructions. Their basic structure has close similarities with spin models in statistical physics applied to magnetism or alloy theory. This year's Nobel Prize in Physics recognizes research exploiting this connection to make breakthrough methodological advances in the field of ANN.

Historical background

The first electronic-based computers appeared in the 1940s, and were invented for military and scientific purposes. They were intended to carry out computations that were cumbersome and time-consuming for humans. In the 1950s, the opposite need emerged, namely to get computers to do what humans and other mammals are good at – pattern recognition.

This artificial intelligence-oriented objective was first approached by mathematicians and computer scientists, who developed programs based on logical rules. This approach was pursued until the 1980s, but the computational resources that were required for the exact classifications, for example, of images became prohibitive.

In parallel, efforts had been initiated to find out how biological systems solve the pattern recognition problem. As early as 1943, Warren McCulloch and Walter Pitts [1], a neuroscientist and a logician, respectively, had proposed a model for how the neurons in the brain cooperate. In their model, a neuron formed a weighted sum of binary incoming signals from other neurons, which determined a binary outgoing signal. Their work became a launch pad for later research into both biological and artificial neural networks.

Another influential early contribution came from the psychologist Donald Hebb [2]. In 1949, Hebb proposed a mechanism for learning and memories, where the simultaneous and repeated activation of two neurons leads to an increased strength of the synapse between them.

In the ANN area, two architectures for systems of interconnected nodes were explored, "recurrent" and "feedforward" networks, where the former allows for feedback interactions (Figures 1 and 2). A feedforward network has input and output layers and may also contain additional layers of hidden nodes sandwiched in-between.

In 1957, Frank Rosenblatt proposed a feedforward network for image interpretation, which was also implemented in computer hardware [3]. It had three layers of nodes, with adjustable weights only between the middle and output layers. Those weights were determined in a systematic fashion.

Rosenblatt's system attracted considerable attention, but it had limitations when it came to non-linear problems. A simple example is the "one or the other but not both" (XOR) problem. These limitations were pointed out in an influential book by Marvin Minsky and Seymour Papert in 1969 [4], which led to a hiatus funding-wise for ANN research.

A parallel development took inspiration from magnetic systems, which were to create models for recurrent neural networks and investigate their collective properties [5-10].

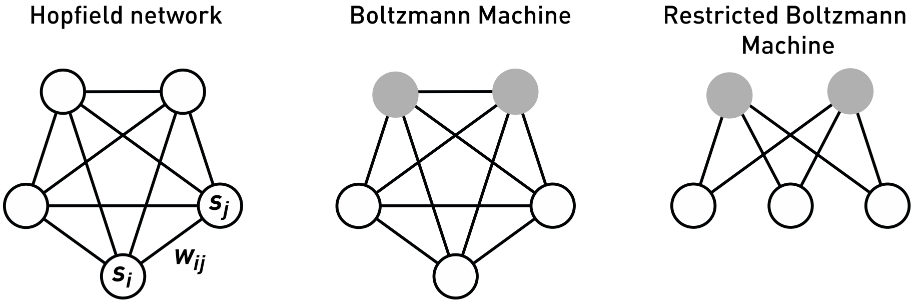

Figure 1. *Recurrent networks of N binary nodes s_i (0 or 1), with connection weights w_{ij}. (Left) The Hopfield model. (Centre) Boltzmann machine. The nodes are divided into two groups, visible (open circles) and hidden (grey) nodes. The network is trained to approximate the probability distribution of a given set of visible patterns. Once trained, the network can be used to generate new instances from the learned distribution. (Right) Restricted Boltzmann Machine (RBM). Same as the Boltzmann machine, but without any couplings within the visible layer or between hidden nodes. This variant can be used for layer-by-layer pre-training of deep networks.*

The 1980s

The 1980s saw major breakthroughs in the areas of both recurrent and feedforward neural networks, which led to a rapid expansion of the ANN field.

John Hopfield, a theoretical physicist, is a towering figure in biological physics. His seminal work in the 1970s examined electron transfer between biomolecules [11] and error correction in biochemical reactions (kinetic proofreading) [12].

In 1982, Hopfield published a dynamical model for an associative memory based on a simple recurrent neural network [13]. Collective phenomena frequently occur in physical systems, such as domains in magnetic systems and vortices in fluid flow. Hopfield asked whether emergent collective phenomena in large collections of neurons could give rise to “computational” abilities.

Noting that collective properties in many physical systems are robust to changes in model details, he addressed this question using a neural network with N binary nodes s_i (0 or 1). The dynamics were asynchronous with threshold updates of individual nodes at random times. The new value of a node s_i was determined by a weighted sum over all other nodes,

$$h_i = \sum_{j \neq i} w_{ij} s_j,$$

and was set to s_i=1 if h_i>0, and s_i=0 otherwise (with the threshold set to zero). The couplings w_{ij} were assumed symmetric and to reflect pairwise correlations between the nodes in stored memories, which is referred to as the Hebb rule. The symmetry of the weights guarantees stable dynamics. Stationary states were identified as memories, distributed over the *N* nodes in a non-local storage. Furthermore, the network was assigned an energy *E* given by

$$E = -\sum_{i<j} w_{ij} s_i s_j,$$

which is a monotonically decreasing function under the dynamics of the network. Notable is that the connection between the world of physics, as defined in the 1980s, and ANNs was obvious already from these two equations. The first equation can be used to represent the Weiss molecular field (after the French physicist Pierre Weiss) that describes how atomic magnetic moments align in a solid, and the latter is often used to evaluate the energy of a magnetic configuration, e.g. a ferromagnet. Hopfield was naturally well aware of how these equations were used to describe magnetic materials.

Metaphorically, the dynamics drive the system with *N* nodes to the valleys of an *N*-dimensional energy landscape, in which the stationary states are located. The stationary states represent memories learned by the Hebb rule. Initially, the number of memories that could be stored in Hopfield's dynamical model was limited. Methods to alleviate this problem were developed in later work [14].

Hopfield used his model as an associative memory or as a method for error correction or pattern completion. A system initialized with an incorrect pattern, perhaps a misspelled word, is attracted to the nearest local energy minimum in his model, whereby a correction occurs. The model gained additional traction when it became clear that basic properties, such as the storage capacity, could be understood analytically, by using methods from spin glass theory [15,16].

A legitimate question at the time was whether the properties of this model are an artifact of its crude binary structure. Hopfield answered this question by creating an analog version of the model [17], with continuous-time dynamics given by the equations of motion for an electronic circuit. His analysis of the analog model demonstrated that the binary nodes could be replaced by analog ones without losing the emergent collective properties of the original model. The stationary states of the analog model corresponded to mean-field solutions of the binary system at an effective adjustable temperature, and approached the stationary states of the binary model at low temperature.

The close correspondence between the analog and binary models was subsequently used by Hopfield and David Tank [18,19] to develop a method for solving difficult discrete optimization problems based on the continuous-time dynamics of the analog model. Here, the optimization problem to be solved, including constraints, is encoded in the interaction parameters (weights) of the network. They chose to use the dynamics of the analog model in order to have a "softer" energy landscape and thereby facilitate the search. The above-mentioned effective temperature of the analog system was gradually decreased, as in global optimization with simulated annealing [20]. Optimization occurs through integration of the equations of motion of an electronic circuit, during which the nodes evolve without instructions from a central unit. This approach constitutes a pioneering example of using a dynamical system to seek solutions to difficult discrete optimization problems [21]. A more recent example is quantum annealing [22].

By creating and exploring the above physics-based dynamical models – not only the milestone associative memory model but also those that followed – Hopfield made a foundational contribution to our understanding of the computational abilities of neural networks.

In 1983–1985 Geoffrey Hinton, together with Terrence Sejnowski and other coworkers, developed a stochastic extension of Hopfield's model from 1982, called the Boltzmann machine [23,24]. Here, each state **s**=(s_1,…,s_N) of the network is assigned a probability given by the Boltzmann distribution

$$P(\boldsymbol{s}) \propto e^{-E/T} \qquad\qquad E = -\sum_{i<j} w_{ij} s_i s_j - \sum_i \theta_i s_i$$

where T is a fictive temperature and θ_i is a bias, or local field.

The Boltzmann machine is a generative model. Unlike the Hopfield model, it focuses on statistical distributions of patterns rather than individual patterns. It contains visible nodes that correspond to the patterns to be learned as well as additional hidden nodes, where the latter are included to enable modelling of more general probability distributions.

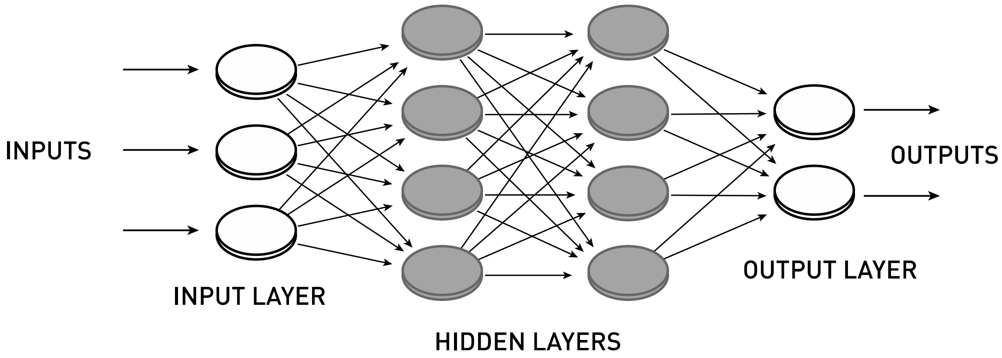

Figure 2. *Feedward network with two layers of hidden nodes between the input and output layers.*

The weight and bias parameters of the network, which define the energy *E*, are determined so that the statistical distribution of visible patterns generated by the model deviates minimally from the statistical distribution of a given set of training patterns. Hinton and his colleagues developed a formally elegant gradient-based learning algorithm for the parameter determination [24]; however, each step of the algorithm involves time-consuming equilibrium simulations for two different ensembles.

While theoretically interesting, in practice, the Boltzmann machine was initially of limited use. However, a slimmed-down version of it with fewer weights, called the restricted Boltzmann machine, developed into a versatile tool (see next section).

Both the Hopfield model and the Boltzmann machine are recurrent neural networks. The 1980s also saw important progress on feedforward networks. A key advance was the demonstration by David Rumelhart, Hinton and Ronald Williams in 1986 of how architectures with one or more hidden layers could be trained for classification using an algorithm known as backpropagation [25]. Here, the objective is to minimize the mean square deviation, *D*, between output from the network and training data, by gradient descent. This requires computing the partial derivatives of *D* with respect to all weights in the network. Rumelhart, Hinton and Williams reinvented a scheme for this, which had previously been applied to related problems by others [26,27]. Additionally, and more importantly, they demonstrated that networks with a hidden layer could be trained by this method to perform tasks known to be unsolvable without such a layer. Furthermore, they elucidated the function of hidden nodes.

Toward deep learning

The methodological breakthroughs in the 1980s were soon followed by successful applications, including pattern recognition in images, languages and clinical data. An important method was multilayered convolutional neural networks (CNN) trained by backpropagation, as advanced by Yann LeCun and Yoshua Bengio [28,29]. The CNN architecture had its roots in the neocognitron method created by Kunihiko Fukushima [30], who in turn was inspired by work of David Hubel and Torsten Wiesel, Nobel Prize Laureates in Physiology or Medicine in 1981. The CNN approach developed by LeCun and coworkers became used by several American banks for classifying handwritten digits on checks from the mid-1990s. Another successful example from this period is the long short-term memory method created by Sepp Hochreiter and Jürgen Schmidhuber [31]. This is a recurrent network for processing sequential data, as in speech and language, and can be mapped to a multilayered network by unfolding in time.

While certain multilayered architectures led to successful applications in the 1990s, it remained a challenge to train deep multilayered networks with many connections between consecutive layers. To many researchers in the field, training dense multilayered networks seemed out of reach. The situation changed in the 2000s. A leading figure in this breakthrough was Hinton, and an important tool was the restricted Boltzmann machine (RBM).

An RBM network has weights only between visible and hidden nodes, and no weights connect two nodes of the same type. For RBMs, Hinton created an efficient approximate learning algorithm [32], called contrastive divergence, which was much faster than that for the full Boltzmann machine [24]. With Simon Osindero and Yee-Whye Teh, he then developed a pre-training procedure for multilayer networks, in which the layers are trained one by one using an RBM [33]. An early application of this approach was an autoencoder network for dimensional reduction [34,35]. After pre-training, it became possible to perform a global parameter fine-tuning using the backpropagation algorithm. The pre-training with RBMs picked up structures in data, such as corners in images, without using labelled training data. Having found these structures, labelling those by backpropagation turned out to be a relatively simple task.

By linking layers pre-trained in this way, Hinton was able to successfully implement examples of deep and dense networks, a milestone toward what is now known as deep learning. Later on, it became possible to replace RBM-based pre-training by other methods to achieve the same performance of deep and dense ANNs.

ANNs as powerful tools in physics and other scientific disciplines

Much of the above discussion is focused on how physics has been a driving force underlying inventions and development of ANNs. Conversely, ANNs are increasingly playing an important role as a powerful tool for modelling and analysis in almost all of physics.

In some applications, ANNs are employed as a function approximator [36]; i.e. the ANNs are used to provide a "copycat" for the physics model in question. This can significantly reduce the computational resources required, thereby allowing larger systems to be probed at higher resolution. Significant advances have been achieved in this way, e.g. for quantum-mechanical many-body problems [37-39]. Here, deep learning architectures are trained to reproduce energies of phases of materials, as well as the shape and strength of interatomic forces, with an accuracy comparable to *ab initio* quantum-mechanical models. With these ANN trained atomic models, considerably faster determination of phase stabilities and the dynamics of new materials can be made. Examples showing the success of these methods involve the prediction of new photovoltaic materials.

With these models, it is also possible to study phase transitions [40] as well as the thermodynamical properties of water [41]. Similarly, the development of ANN representations has made it possible to reach higher resolutions in explicit physics-based climate models [42,43] without resorting to additional computing power.

During the 1990s, ANNs became a standard data analysis tool within particle physics experiments of ever-increasing complexity. Highly sought-after fundamental particles, such as the Higgs boson, only exist for a fraction of a second after being created in high-energy collisions (e.g. ~10^{-22} s for the Higgs boson). Their presence needs to be inferred from tracking information and energy deposits in large electronic detectors. Often the anticipated detector signature is very rare and could be mimicked by more common background processes. To identify particle decays and increase the efficiency of analyses, ANNs were trained to pick out specific patterns in the large volumes of detector data being generated at a high rate.

ANNs improved the sensitivity of searches for the Higgs boson at the CERN Large Electron-Position (LEP) collider during the 1990s [44], and were used in the analysis of data that led to its discovery at the CERN Large Hadron Collider in 2012 [45]. ANNs were also used in studies of the top quark at Fermilab [46].

In astrophysics and astronomy, ANNs have also become a standard data analysis tool. A recent example is an ANN-driven analysis of data from the IceCube neutrino detector at the South Pole, which resulted in a neutrino image of the Milky Way [47]. Exoplanet transits have been identified by the Kepler Mission using ANNs [48]. The Event Horizon Telescope image of the black hole at the centre of the Milky Way used ANNs for data processing [49].

So far, the most spectacular scientific breakthrough using deep learning ANN methods is the AlphaFold tool for prediction of three-dimensional protein structures, given their amino acid sequences [50]. In modelling of industrial physics and chemistry applications, ANNs also play an increasingly important role.

ANNs in everyday life

The list of applications used in everyday life that are based on ANNs is long. These networks are behind almost everything we do with computers, such as image recognition, language generation, and more.

Decision support within health care is also a well-established application for ANNs. For example, a recent prospective randomized study of mammographic screening images showed a clear benefit of using machine learning in improving detection of breast cancer [51]. Another recent example is motion correction for magnetic resonance imaging (MRI) scans [52].

Concluding remarks

The pioneering methods and concepts developed by Hopfield and Hinton have been instrumental in shaping the field of ANNs. In addition, Hinton played a leading role in the efforts to extend the methods to deep and dense ANNs.

With their breakthroughs, that stand on the foundations of physical science, they have showed a completely new way for us to use computers to aid and to guide us to tackle many of the challenges our society face. Simply put, thanks to their work Humanity now has a new item in its toolbox, which we can choose to use for good purposes. Machine learning based on ANNs is currently revolutionizing science, engineering and daily life. The field is already on its way to enable breakthroughs toward building a sustainable society, e.g. by helping to identify new functional materials. How deep learning by ANNs will be used in the future depends on how we humans choose to use these incredibly potent tools, already present in many aspects of our lives.

References

1. W.S. McCulloch and W. Pitts, *Bull. Math. Biophys.* **5**, 115 (1943).
2. D.O. Hebb, *The organization of behavior* (Wiley & Sons, New York, 1949).
3. F. Rosenblatt, *Principles of neurodynamics:Perceptrons and theory of brain mechanisms* (Spartan Book, Washigton D.C., 1962).
4. M.L. Minsky and S.A. Papert, *Perceptrons: An introduction to computational geometry* (MIT Press, Cambridge, 1969).
5. B.G. Cragg and H.N.V. Temperley, *Brain* **78**, 304 (1955).
6. E.R. Caianiello, *J. Theor. Biol.* **2**, 204 (1961).
7. K. Nakano, *IEEE Trans., Syst., Man, Cybern*. **SMC-2**, 380 (1972).
8. S.-I. Amari, *IEEE Trans. Comput.* **C-21**, 1197 (1972).
9. W.A. Little, *Math. Biosci.* **19**, 101 (1974).
10. W.A. Little and G.L. Shaw, *Math. Biosci.* 39, 281 (1978).
11. J.J. Hopfield, *Proc. Natl. Acad. Sci USA* **71**, 3640 (1974).
12. J.J. Hopfield, *Proc. Natl. Acad. Sci USA* **71**, 4135 (1974).
13. J.J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
14. D. Krotov and J.J. Hopfield. In *Advances in Neural Information Processing Systems* **29**, 1172 (2016).
15. D. J. Amit, H. Gutfreund and H. Sompolinsky, *Phys. Rev. A* **32**, 1007 (1985).
16. M. Mézard, G. Parisi and M. Virasoro, *Spin glass theory and beyond: An introduction to the replica method and its applications* (World Scientific, Singapore, 1987).
17. J.J. Hopfield, *Proc. Natl. Acad. Sci. USA* **81**, 3088 (1984).
18. J.J. Hopfield and D.W. Tank, *Biol. Cybern.* **52**, 141 (1985).
19. J.J. Hopfield and D.W. Tank, *Science* **233**, 625 (1986).
20. S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, *Science* **220**, 671 (1983).
21. N. Mohseni, P. McMahon and T. Byrnes, *Nat. Phys. Rev.* **4**, 363 (2022).
22. T. Kadowaki and H. Nishimori, *Phys. Rev. E* **58**, 5355 (1998).

23. S.E. Fahlman, G.E. Hinton and T.J. Sejnowski. In *Proceedings of the AAAI-83 conference*, pp. 109-113 (1983).

24. D.H. Ackley, G.E. Hinton and T.J. Sejnowski, *Cogn. Sci.* **9**, 147 (1985).

25. D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Nature* **323**, 533 (1986).

26. P.J. Werbos. In *System Modeling and Optimization*, pp. 762-770 (1982).

27. S. Linnainmaa, Master's thesis (in Finnish), Univ. Helsinki (1970); published in *BIT* **16**, 146 (1976).

28. Y. LeCun, B.Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard and L.D. Jackel, *Neural Comput.* **1**, 541 (1989).

29. Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, *Proc. IEEE* **86**, 2278 (1998).

30. K. Fukushima, *Biol. Cybern.* **36**, 193 (1980).

31. S. Hochreiter and J. Schmidhuber, *Neural Comput.* **9**, 1735 (1997).

32. G.E. Hinton, *Neural Comput.* **14**, 1771 (2002).

33. G.E. Hinton, S. Osindero and Y.-W. The, *Neural Comput.* **18**, 1527 (2006).

34. Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle. In *Advances in Neural Information Processing Systems* **19**, 153 (2006).

35. G.E. Hinton and R. Salakhutdinov, *Science* **313**, 504 (2006).

36. K. Hornik, *Neural Netw.* **4**, 251 (1991).

37. J. Behler and M. Parrinello, *Phys. Rev. Lett.* **98**, 146401 (2007).

38. G. Carleo and M. Troyer, *Science* **355**, 602 (2017).

39. P.M. Piaggi, J. Weis, A.Z. Panagiotopoulos, P.G. Debenedetti and R. Car, *Proc. Natl. Acad. Sci. USA* **119**, e2207294119 (2022).

40. R. Jinnouchi, J. Lahnsteiner, F. Karsai, G. Kresse and M. Bokdam, *Phys. Rev. Lett.* **122**, 225701 (2019).

41. P.M. de Hijes, C. Dellago, R. Jinnouchi, B. Schmiedmayer and G. Kresse, *J. Chem. Phys.* **160**, 114107 (2024).

42. S. Rasp, M.S. Pritchard and P. Gentine, *Proc. Natl. Acad. Sci USA* **115**, 9684 (2018).

43. C. Wong, *Nature* **628**, 710 (2024).

44. ALEPH Collaborations, *Phys. Lett B* **447**, 336 (1999).

45. ATLAS Collaboration, *Phys. Lett. B* **716,** 1 (2012).

46. D0 Collaboration, *Phys. Rev. Lett.* **103**, 092001 (2009).

47. IceCube Collaboration, *Science* **380**, 1338 (2023).

48. K.A. Pearson, L. Palafox and C.A. Griffith, *Mon. Not. R. Astron. Soc.* **474**, 478 (2017).

49. EHT Collaboration, *ApJL* **930**, L15 (2022).

50. J. Jumper *et al.*, *Nature* **596**, 583 (2021).

51. K. Lång *et al.*, *Lancet Oncol.* **24**, 936 (2023).

52. V. Spieker *et al.*, *IEEE Trans. Med. Imaging* **43**, 846 (2024).