WHY DOES THE FOUNDATION OF ARTIFICIAL INTELLIGENCE DESERVE A NOBEL PRIZE IN PHYSICS?

SUMMARY: The 2024 Nobel Prize in Physics was awarded to John Hopfield (Princeton University) and Geoffrey Hinton (University of Toronto) “for foundational discoveries and inventions that enable machine learning with artificial neural networks” [1, 2]. Even though their fundamental works contributed significantly to the development of today’s artificial intelligence, the community questioned whether a recognition with a Physics Nobel Prize falls in the appropriate discipline. However, the development of the fundamental artificial neural network models introduced by Hopfield and Hinton was inspired by spin glasses and is thus based on principles from many-body physics.

By **STEFANIE CZISCHEK** <stefanie.czischek@uottawa.ca>

Department of Physics, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada

This year’s Nobel Prize in Physics caused puzzling debates in the community. The prize was awarded to John Hopfield (Princeton University) and Geoffrey Hinton (University of Toronto) “for foundational discoveries and inventions that enable machine learning with artificial neural networks” [1, 2]. While there is no doubt that a Nobel Prize is well deserved by the two researchers whose pioneering works laid the foundation for technologies that now shape our everyday life, many were confused about why it is Physics that lays claim on these fundamental developments. However, taking a closer look at the research behind the awards sheds some light into the discussion.

HOPFIELD’S CONTRIBUTION: HOPFIELD NETWORKS

Both Nobel laureates developed very fundamental network structures that revolutionized the field of machine learning with artificial neural networks [2]. John Hopfield’s major contribution was the introduction of Hopfield networks [3]. These consist of a set of neurons, which are binary units, interacting via symmetrically weighted, bidirectional all-to-all connections, as illustrated in Figure 1. The weights of these connections are *learned* when the network evolves to minimize a chosen energy function for a given input. The energy function E_{HN} defined over a Hopfield network is given by

$$E_{\mathrm{HN}} = -\frac{1}{2}\sum_{i,j} w_{i,j} s_i s_j,$$

where $w_{i,j} = w_{j,i}$ is the connecting weight between neurons i and j with $w_{i,i} = 0$. The state of neuron i is denoted by $s_i \in \{-1, 1\}$, where the index i runs over all N neurons in the network. By optimizing the connection weights $w_{i,j}$ for a given input state $\boldsymbol{s}$, Hopfield networks can recover patterns from noisy inputs, making them powerful candidates to deal with incomplete data [3]. The chosen setup of a fully connected network of binary units resembles a spin glass in condensed matter physics, specifically a Sherrington-Kirkpatrick model [4] and the energy function takes the familiar form of an Ising model, revealing the field and experience that inspired Hopfield's work.

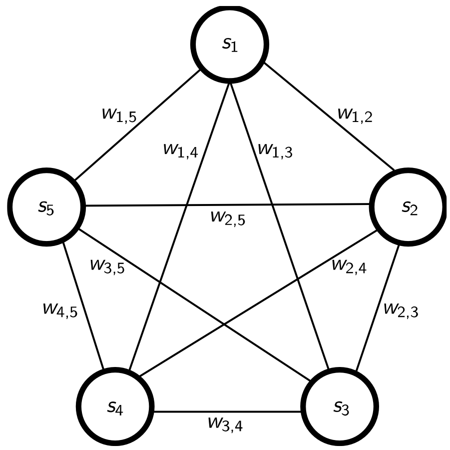

Figure 1. Hopfield network with $N = 5$ neurons s_i and connecting weights $w_{i,j} = w_{j,i}$.

HINTON'S CONTRIBUTION: RESTRICTED BOLTZMANN MACHINES

Similarly, Geoffrey Hinton introduced another fundamental network architecture, the restricted Boltzmann machine, a special case of the general Boltzmann machine [5], that finds applications in tasks like data generation [6] or classification [7]. Restricted Boltzmann machines consist of two layers of binary neurons, one interpretable *visible* layer and one *hidden* layer that characterizes the network's expressive power and enables efficient network training and data generation [6, 8]. While there are weighted bidirectional and symmetric all-to-all connections between neurons from different layers, no intralayer connections are allowed in the restricted Boltzmann machine, as illustrated in Figure 2. Based on this setup, an energy is defined for the overall network which is minimized by optimizing the connection weights during the network training process. This energy term E_{RBM} is defined similarly to the energy in Hopfield networks, but takes the restricted connectivity into account,

$$E_{\mathrm{RBM}} = -\sum_{i,j} v_i w_{i,j} h_j - \sum_i a_i v_i - \sum_j b_j h_j \, .$$

Here, $v_i \in \{0,1\}$ denotes the state of each visible neuron and, accordingly, $h_j \in \{0,1\}$ denotes the state of each hidden neuron, where the indices i and j run over all N_{v} visible and N_{h} hidden neurons, respectively. The parameters $w_{i,j}$ denote the connecting weights and a_i and b_j denote a bias factor for the visible and hidden neurons, respectively, which are also optimized during the network training process. Just like the Hopfield network, the restricted Boltzmann machine describes the Sherrington-Kirkpatrick model [4] as a spin glass in an external field. Furthermore, from the network energy E_{RBM}, a Boltzmann distribution $P(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{Z}\exp[-E_{\mathrm{RBM}}]$, with partition function Z as normalization constant, can be defined over all possible states of the neuron ensemble. By drawing samples of visible neuron configurations from the distribution encoded in the trained network, interpretable data can be generated. The specific network architecture of the restricted Boltzmann machine provides an efficient algorithm for sample generation. This ability led to novel achievements in tasks like image or text generation [6, 8].

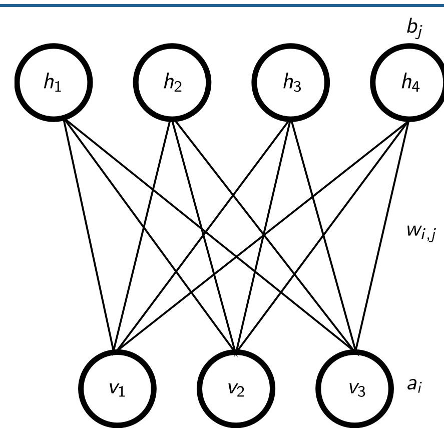

Figure 2. Restricted Boltzmann machine with $N_{\mathrm{v}} = 3$ visible and $N_{\mathrm{h}} = 4$ hidden neurons v_i and h_j, connecting weights $w_{i,j}$, and visible and hidden biases a_i and b_j, respectively.

Just like Hopfield networks, the introduction of the restricted Boltzmann machine as a fundamental artificial neural network architecture demonstrates the importance that physical models play in the field of machine learning and artificial intelligence. These developments show how the detailed understanding of spin glass models and the physical phenomena observed in those find applications far beyond many-body systems and directly affect commonly used technologies. The fact that physical principles have big impacts on computational algorithms clearly showcases how intertwined different research disciplines are and emphasizes the importance of collaborations to combine knowledge and approaches across specialized topics.

Nowadays, Hopfield networks and restricted Boltzmann machines are outdated and have been replaced by advanced, more powerful artificial neural network architectures [9]. Those include recurrent neural networks, or transformer models whose immense strength is demonstrated in recent groundbreaking artificial intelligences like ChatGPT [10]. However, all these impressive advances are built on the foundation of Hopfield networks and restricted Boltzmann machines. The introduction of those models in the 1980s pointed out that artificial neural networks are more powerful than experienced before and with this ended a decay period known as the *AI winter*. Together with the introduction of efficient learning algorithms for the restricted Boltzmann machine by Geoffrey Hinton and co-workers in the mid 2000s [11], they motivated the enhancements that led to today's state-of-the-art artificial neural network architectures.

While physics has significantly inspired the fundamental developments of today's artificial neural networks, at the same time artificial neural networks play an important role in the developments in state-of-the-art classical and quantum physics [12, 13]. Besides optimizing data evaluation, suggesting efficient experimental setups, or automatizing the tuning of experimental devices, artificial neural networks can be used for phase transition detection, to model quantum many-body systems, and more. Especially in the field of numerically simulating dynamics and ground states of quantum many-body systems, restricted Boltzmann machines have been the driving force that initiated the field [14, 15, 16, 17]. Early works used restricted Boltzmann machines as a general wavefunction ansatz and showed that such a general approach can accurately and efficiently model even higher-dimensional qubit systems.

CONCLUSIONS

Overall, there is a strong bidirectional connection between artificial neural networks and physics where one takes advantage of the other. While today's artificial neural networks play a significant role not only in our everyday life but also in various fields of classical and quantum physics, we would not have reached this stage without many-body physics that inspired the fundamental developments of artificial neural networks. With this, the 2024 Nobel Prize in Physics has surely been awarded in the right discipline. In the end, physics has not laid claim on those fundamental developments with this Nobel Prize, it has rather always been the understanding and phenomena of physical models that made those developments possible.

ACKNOWLEDGEMENTS

S. Czischek acknowledges discussions with B. Joos and J. Schulz.

REFERENCES

[1] Press release. NobelPrize.org., Nobel Prize Outreach AB 2024, 2024.

[2] Advanced information. NobelPrize.org, Nobel Prize Outreach AB 2024, 2024.

[3] J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc Natl Acad Sci USA **79(8)**, pp. 2554-2558, 1982.

[4] D. Sherrington and S. Kirkpatrick, "Solvable Model of a Spin-Glass," Phys Rev Lett **35**, p. 1792, 1975.

[5] G. Hinton, "Boltzmann machine," Scholarpedia **2(5)**, p. 1668, 2007.

[6] P. Smolensky, "Chapter 6: Information Processing in Dynamical Systems: Foundations of Harmony Theory," in Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations, MIT Press, 1986.

[7] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," Proceedings of the 25th international conference on Machine Learning - ICML '08, p. 536, 2008.

[8] G. Hinton, "Training products of experts by minimizing contrastive divergence," Neural Computation **14(8)**, pp. 1711-1800, 2002.

[9] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.

[10] OpenAI, ChatGPT, 2023.

[11] G. Hinton, "A Practical Guide to Training Restricted Boltzmann Machines," UTML TR **2010**, p. 003, 2010.

[12] A. Dawid, J. Arnold, B. Requena, A. Gresch, M. Plodzien, K. Donatella, K. Nicoli, P. Stornati, R. Koch, M. Büttner, R. Okula, G. Munoz-Gil, R. Vargas-Hernandez, A. Cervera-Lierta, J. Carrasquilla, V. Dunjko, M. Gabrie, P. Huembeli, E. van Nieuwenburg, F. Vicentini, L. Wang, S. Wetzel, G. Carleo, E. Greplova, R. Krems, F. Marquardt, M. Tomza, M. Lewenstein, A. Dauphin, "Modern applications of machine learning in quantum sciences," arXiv:2204.04198 [quant-ph], 2022.

[13] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto and L. Zdeborová, "Machine learning and the physical sciences," Rev Mod Phys **91**, p. 045002, 2019.

[14] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," Science **355(6325)**, pp. 602-606, 2017.

[15] G. Torlai and R. Melko, "Learning thermodynamics with Boltzmann machines," Phys Rev B **94**, p. 165134, 2016.

[16] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko and G. Carleo, "Neural-network quantum state tomography," Nature Physics **14**, pp. 447-450, 2018.

[17] R. Melko, G. Carleo, J. Carrasquilla and J. Cirac, "Restricted Boltzmann machines in quantum physics," Nature Physics **15**, pp. 887-892, 2019.