

2024年诺贝尔物理学奖与学科交叉*

白若然 张丽茹 黄海平[†]

(中山大学物理学院 广州 510275)

2024-12-05收到

[†] email: huanghp7@mail.sysu.edu.cn

DOI: 10.7693/wl20250103

CSTR: 32040.14.wl20250103

The Nobel Prize in Physics 2024 and related interdisciplinary studies

BAI Ruo-Ran ZHANG Li-Ru HUANG Hai-Ping[†]

(School of Physics, Sun Yat-sen University, Guangzhou 510275, China)

摘要 2024年, 霍普菲尔德和辛顿因对神经网络和人工智能(AI)的基础性贡献而被授予诺贝尔物理学奖, 引发了科技界的广泛热议。AI是否属于物理学以及物理学如何为AI做出贡献成为争论的焦点。文章回顾了物理学如何促进AI早期发展的历史, 特别强调了两个重要的科学分支, 它们起源于通过应用物理学思想来研究神经网络的实践。之后还讨论了理解和进一步改进AI以及思考“智能的本质”的未来方向。

关键词 诺贝尔物理学奖, 霍普菲尔德网络, 玻尔兹曼机, 理论神经科学, 统计物理学, 人工智能

Abstract In 2024, Hopfield and Hinton were awarded the Nobel Prize in Physics for their fundamental contributions to neural-network based artificial intelligence (AI), which triggered much debate on whether AI belongs to physics and how physics contributes to AI. Here, we review the history of how physics contributed to the early development of AI, and in particular we highlight two important scientific branches that were initiated by applying physical concepts to study neural networks. We also discuss future directions for understanding and improving AI, and for resolving the nature of intelligence.

Keywords Nobel Prize in Physics, Hopfield networks, Boltzmann machine, theoretical neuroscience, statistical physics, artificial intelligence

1 引言

2024年的诺贝尔物理学奖授予了对神经网络基础理论发展做出卓越贡献的科学家霍普菲尔德和辛顿。对于这个结果, 许多人有这样的疑问: 人工智能似乎是一个计算机科学里的分支, 它跟物理是如何扯上关系的? 霍普菲尔德原是一名物

理学家, 后跨界到计算神经科学和计算机科学; 而辛顿则是一名地道的计算机科学家。在人们的一贯印象中, 研究大脑应该是医生或生物学家的事情, 而研究人工智能则是计算机科学家的任务。2024年诺贝尔物理学奖颁发给计算神经科学鼻祖霍普菲尔德和深度学习之父辛顿具有非凡的历史意义。本文将从学科交叉角度分析物理学与人工智能、脑科学学科交叉融合的历史与未来。

* 国家自然科学基金优秀青年基金(基金号: 12122515)资助项目

2 怎样的工作能获得诺贝尔奖的青睞？

在2024年诺贝尔物理学奖颁发之后，一名瑞士科学家 Andrew Akbashev 在社交媒体上写了他对于诺奖级工作的理解，这个分析非常有意思。其中第一点是无用性。这强调了获奖者一开始就不是奔着有用性，而是为了满足自己的好奇心才给世界带来了新知识和新见解，这当然就是科学的定义。第二点是敢冒天下之大不韪。这其实是在强调他们做了大多数人害怕做的课题。比如霍普菲尔德，在40多年前，他就在思考人类的大脑为什么会想起一件事情，以及该怎么用物理语言去描述它。基于当时的知识，思考这样的事情无疑是耗费心力的，一方面因为大脑的结构远非清楚；另一方面物理学概念能否描述如此复杂的过程尚属疑问。而辛顿亦是如此。20世纪80年代时，人工智能刚经历过寒冬，而辛顿坚信可以把神经网络变聪明并学习任意数据分布。他最终从物理学家那里汲取了灵感，发展了玻尔兹曼机算法。第三点是诺贝尔奖的工作能够孵化一个大的研究方向。霍普菲尔德的工作创建了理论神经科学学科，在21世纪初各国开展脑科学项目的背景下意义非凡，辛顿则开创了深度学习的范式，为人类逐步走向通用人工智能奠定了基石。最后一点是诺贝尔奖并不单纯奖励研究论文的引用数，而是奖励一个与人类科技文明息息相关的领域的思想引领者。从这一点上讲，两位科学家当之无愧。

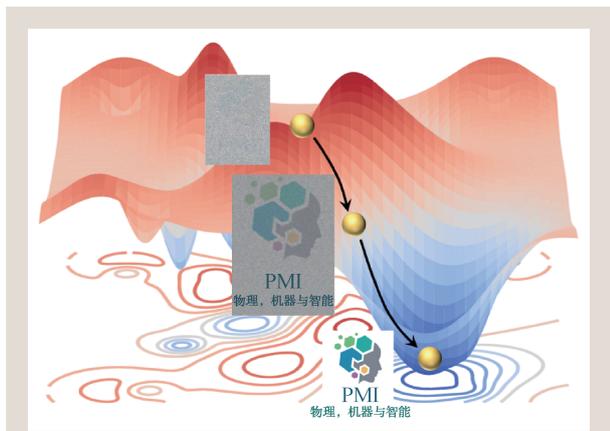


图1 联想记忆网络的能量景观和记忆提取动力学

3 霍普菲尔德与理论神经科学

霍普菲尔德曾经回忆他如何对大脑感兴趣的故事。当时他在美国西海岸从事研究工作，因为周围的同事对交叉学科抱有浓厚的兴趣，所以霍普菲尔德时常沉浸在这样跨学科的讨论中。他逐步认识到：“How mind emerges from brain is to me the deepest question posed by our humanity”^[1]。于是，他开启了对大脑记忆功能的理解。对于物理学家来说，理解就意味着能用数学去描述具体的过程。在经历过多次失败的尝试之后，霍普菲尔德忽然意识到物理学中能量的概念是一个关键的突破。

我们想象一个碗，在这个碗内侧放一个小球，小球就会滚下去。如果有摩擦力的话，小球最后就会停留在碗底。一个小球是你能够感受到的三维空间的物体，如果大脑只有三个细胞，那这个大脑的状态就可以变成三维空间中的小球，每个神经元的状态分别对应小球的三维坐标。但是我们的大脑有几百亿个神经元，每个神经元都有一个随时间变化的状态。所以人类没办法直观地在三维空间想象这个动力学过程，但是大脑的状态可以被拓展为高维空间的广义粒子。我们的记忆可以被映射为小球滚动的过程，这就是霍普菲尔德提出的联想记忆模型^[2]。你可以想象大脑中会存在一个势能，当大脑的状态连续改变的时候，就相当于一个广义粒子在高维空间的势能面上运动。你的记忆就是势能低点所对应的状态。有时候你看到了某个东西，它会突然勾起你美好或悲伤的记忆，这个回想的过程从物理学上讲就是粒子在受到梯度力的作用下往能量最低点运动的轨迹(图1)。这样的映射极具物理思维，因为一个在重力势场中运动的粒子对于每一个学过物理的学生来说是非常了解的事情。

霍普菲尔德对大脑如何实现记忆的思考离不开当时物理学的发展，尤其是自旋玻璃理论。他特别提到很多想法曾经受到凝聚态物理大师菲利普·安德森的影响。而同时代的帕里西是2021年

诺贝尔物理学奖得主，他有一个观点是：复杂系统的一个重要特征是很多平衡态或非平衡态共存，而生物这个复杂系统则是通过在不同态之间变换来适应多变且不确定的世界。帕里西在早期的研究中实现了从无序系统中提取隐藏的有序，并且开创了复本对称破缺的研究范式^[3]。这个范式在20世纪80年代中期就已经延伸到人工神经网络理论研究，并且早期的研究恰好集中于分析霍普菲尔德发明的联想记忆网络，这正是当时大量物理学家进入一个新兴交叉领域的根本原因。在那个时代，物理学的思想已经孕育了多个后来繁荣的研究方向，如计算神经科学和基于无监督学习的预训练方案。

霍普菲尔德的模型并不是大脑记忆机制的精确描述，但它开创了一个供众多学科的研究人员思考大脑复杂功能的思维起点。例如，2014年因发现大脑导航细胞而获得诺贝尔生理学医学奖的莫泽教授强调，“在研究神经元如何在记忆和导航中协同工作时，霍普菲尔德模型对神经科学家很有用。该模型将记忆描述为一个表面的低点，帮助研究人员可视化某些动力学状态，比如焦虑等”。因此，霍普菲尔德模型在认知科学、神经生物学同样有较广泛的影响。

霍普菲尔德非常关注大脑的研究，但他坚信，如果不能用数学语言来描述大脑，人们就无法真正理解它。这一点充分体现了他物理学的思维方式，而非生物学的思维模式。生物学研究往往是借助一些先进的技术工具对大脑皮层或海马体神经活动的记录和分析来研究记忆。当跨学科研究时，是否需要掌握所有相关学科的知识呢？他认为，如果完全依赖于生物学的细节，反而束缚了物理学的思想，因此他并未过多考虑生物细节。他的交叉学科研究风格值得我们借鉴。在如今的学术环境中，许多学生提到跨学科研究时，往往认为必须拼命地研读其他学科的书籍，这显然是十分困难的。然而，霍普菲尔德的做法却不同，他认为物理学训练为其打开了其他学科的大门，并且帮助他在跨学科领域做出了卓越贡献^[1]。因此，在跨学科的研究中，我们要保持物理的思维方式，而非考虑尽量多的细节。复杂系统可以在

很多层面上有其自身的规律，越往顶层越普适，越往底层越存在个体依赖性^[4]。

计算神经科学的发展历程可以追溯到1983年春在加州理工学院举办的“HopFest”讨论会^[5]。这个因霍普菲尔德网络的提出而由科学家发起的邀请制会议，充分表明了当时霍普菲尔德论文引发的研究热情，之后会议进一步扩大和公开，聚集了来自各领域的专家，探讨了神经网络和计算智能的可能性。HopFest讨论会被认为是首次正式的计算神经科学专业会议，奠定了该领域的基础，吸引了跨学科的广泛关注。这场讨论会并没有明确地将目标局限于人工神经网络，组织者的初衷是会议能够吸引生物学家参与，特别是关注大脑的生物学机制，并讨论人工神经网络与人类大脑智能之间的关系。然而，随着会议的召开，组织者的观点开始分化。部分学者认为，探索大脑的复杂性过于缓慢且不切实际，反而认为人工神经网络本身的发展和优化可以成为实现智能的真正路径。因此，在最初的几届会议中，参会者关于是否应该继续关注大脑的生物学细节产生了激烈的辩论。其中一部分人选择在人工智能领域研究智能，这个方向的努力由辛顿引领，演化出NeurIPS (Neural Information Processing Systems)会议，该会议逐渐成为全球顶尖的学术盛会，专注于神经信息处理和人工智能。截至2024年，NeurIPS会议已举办到第38届，汇聚了全球科学家和工程师，推动了计算神经科学、人工智能及其他相关学科的交叉融合。

与此同时，另一个重要的组织OCNS (Organization for Computational Neurosciences)也在这一时期形成，由(计算)神经生物学家主导，该会议迅速发展为计算神经科学领域的核心会议之一。截至2024年，OCNS会议已经举办到第33届，成为该领域专注于计算神经科学和脑科学实验的重要学术交流平台。

从HopFest讨论会开始，计算神经科学从初期的封闭学术探讨到形成多个重要的国际会议，不断推动着脑科学、人工智能和计算神经科学的跨学科发展。这一历程体现了学科交叉合作的重要性，为理解和探索智能系统、神经网络以及人

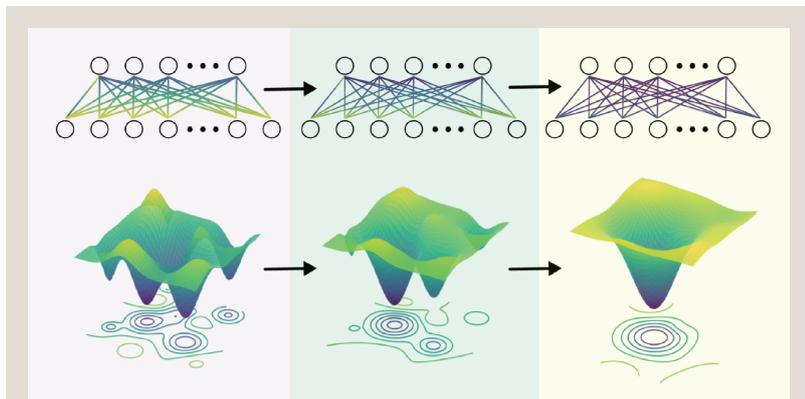


图2 玻尔兹曼机训练及其玻尔兹曼分布演化(低维近似)

类大脑的奥秘奠定了坚实的基础。如今人工智能的辉煌离不开40多年前一群为了好奇心而研究的科学家们的热情。值得一提的是，当年与辛顿合作提出玻尔兹曼机的特伦斯·谢诺夫斯基长期担任NeurIPS会议的主席。

4 辛顿与人工智能

辛顿在1985年发表了一篇重要论文，这篇文章由他与戴维·阿克利和特伦斯·谢诺夫斯基两位同事合作完成，文中提出了玻尔兹曼机(Boltzmann Machine)的概念^[6]，成为后来人工智能研究的基石之一。当时，他们讨论的焦点在于如何构建一个可以自主学习的神经网络，而不是霍普菲尔德网络那些确定的权重连接。辛顿等人将物理学中的概念引入神经网络，尝试通过加入“温度”变量(或者采样)来调节网络的连接强度，从而使网络不仅能记忆，还可以进行更复杂的学习。在40多年前提出的这一思路，尽管今天看来似乎较为简单，但在当时是极为创新的，而这完全是基于多学科科学家在一次小型研讨会上碰撞出的思想^[7]。辛顿构思的玻尔兹曼机实际上基于物理学的Ising模型，将神经元的激活状态视为物理系统的自旋状态，而不断变化的自旋状态促使神经连接不断更新，直至模型的玻尔兹曼分布趋于实际数据分布(图2)。这就是后来大名鼎鼎的对比散度(contrastive divergence, CD)算法的源头^[8]。因此辛顿的早期研究创造了一个可以动态调整的模式，以适应多变的真实数据分布。这某种程度上

摒弃了物理学家猜测模型的习惯，将科学研究在几十年之后带入数据驱动的时代(如当今大火的AI for Science, 简称AI4S)^[9]。

辛顿在2006年发表了另一篇重要的工作，巧妙地使用CD算法预训练深度神经网络，即对深度学习网络进行逐层训练，然后将权重进行微调^[10]。这一创新为深度学习的发展开辟了新路径，也直接促成了2012年在ImageNet图像识别竞赛中

深度学习算法的大获全胜，震惊了计算机界。这一成果标志着深度学习的崛起，也引发了人工智能领域的重大变革，吸引了众多科技巨头的关注和投资。辛顿等人提出的训练范式，即逐层预训练+微调，为深度学习的早期成功奠定了基础。例如，当前的聊天式机器人等大规模预训练模型也使用了预训练和微调的框架；另外近期不少无监督学习也是采用这种范式(虽然训练是端到端的)。因此，辛顿等人的早期研究成果为后续AI层出不穷的发展奠定了基础。总结一下，霍普菲尔德和辛顿从物理学思想出发，然后创造了另一个理解和应用智能的研究领域，如今服务于所有科学，当然也包括物理学本身，详情可见评述文章[11]。

最后，需要强调的是，与AI相关的研究不仅需要应用研究中利用和提升“预测”能力，还需要在理论研究中增强我们对黑盒子的“理解”。在现代社会，大家可能更关注模型的预测能力，但仍然需要铭记物理学家弗里曼·戴森所说的一句话：“严格理论赋予一个课题以智力的深度和精确；在你能证明一个严格理论之前，你不可能全面理解你所关注的概念的意义”(见其著名演讲《鸟与青蛙》)。因此真正理解AI的工作原理依然至关重要^[4, 12, 13]。只有对AI进行全面理解，才能推动这一场科技革命更好地服务于社会发展。

5 未来学科交叉趋势

2024年的诺贝尔化学奖颁发给设计AlphaFold的三名科学家，是AI4S的典型代表。AI4S让人

们看到深度学习的威力，比如 AlphaFold，彻底革新了从氨基酸序列到三维蛋白质结构的预测精度。在高维空间做统计预测显然是深度网络的强项。而在 AI 取得应用辉煌的同时，人们就不得不去面对 AI 本身的机制，即 Science for AI (S4AI 或者 Science of AI 更精确些)，也就是 2024 年的诺贝尔物理学奖。Science for AI 的关键词是“理解”，询问“为什么”是理论家的一个本能反应；理解了 AI 之后，我们就可以重新设计 AI，使它达到一个更高的层次。提升 AI，AI 发展科学，研究 AI 再度提升 AI，所以在 21 世纪，我们更希望 AI 和 Science 形成这样的一个闭环(图 3)。

我们接下来看看霍普菲尔德对物理学的定义^[1]。他回忆，因为父母都是物理学家，所以物理学对他来说不是特别陌生的东西。他小时候喜欢拆东西，拆完之后还喜欢研究这些东西到底是怎么组成在一起实现功能的。大家都以为拆东西是工程师的事情，但是在霍普菲尔德看来，拆完之后能够对其行为进行定量的解释，这就是物理学。所以从这个意义上讲，只要有对世界上万物的“想去拆解它，想看看它是怎么运行的”这种想法，这本身就是一种物理思维。所以物理学更像是一种世界观，即我们周围的世界可以通过个人努力、科学家的聪明才智和充足的资源以合理的定量方式被理解(并且这种理解带有很强的预言能力)。在 AI 不断取得进步的今天，我们应该重新去定义物理学，物理学已经超出了传统的研究物理对象(what)的范畴，而是如何运用(how)物理思维探索世界万物的学科(从 what 到 how 的跃迁)。从这个观点来看，辛顿和霍普菲尔德的获奖将鼓励更多理论家运用物理学的想象力进一步破解智能的本质。

从 2024 年诺贝尔奖的特点来看，我们正在进入一个全新时代：学科边界正在模糊，而同时重大创新正走向大道至简的原则。理查德·费曼说过，自然界是不分科的，而学科细分是人类为了方便研究而引入的。如今基于简单规则设计的 AI 正在走向通用人工智能；或者，在 AI 眼中，万物皆为高维向量，而所谓的“思考”只不过是一场矩阵、向量计算罢了。复杂系统领域有一条铁一

般的规律：任何复杂现象的背后都有简单的规律；而简单规则的多次迭代将会涌现复杂的结构和功能^[14]。

2023 年初，我们迎来了一个全新的“怪物”——ChatGPT (聊天式机器人)^[15]。这类模型的参数量达到了惊人的一万亿，虽然与人类大脑的连接数相比少了数百倍，但其工作方式却截然不同。ChatGPT 模型通过对海量数据的压缩和总结来生成知识，但这种知识并不是传统意义上的公式化理解和推导，而是基于数据模式的预测。然而，这种基于数据压缩的“知识”与传统的逻辑推理和理解之间存在矛盾。传统物理学依赖于方程和逻辑推导，而现代的大规模计算模型则是基于数据驱动的预测。这导致了一个重要的问题：计算是否等同于理解？当我们讲授经典物理学方程时，可以通过逻辑和推理让学生明白方程的物理含义。然而，对于像 ChatGPT 这样基于数据的预测模型来说，知识变成了编程和调参，模型的内部机理往往不透明。这是当今两大研究范式的冲突。

如今的大模型参数达到了万亿级别，而人类大脑的连接数是五百万亿左右。随着训练参数的增加，AI 展现出了越来越多的能力。80 亿参数的



图 3 AI 和 Science 形成闭环，AI4S 强调预测，而 S4AI 强调在机制层面的理解可以进一步提升 AI

模型仅能进行基本的语言理解和代数运算，而达到五千亿参数后，模型涌现出理解语言、回答问题、翻译、推理等多种能力，这种规模效应确实让人们惊叹和不解。我们有理由相信，AI将可以实现对已知概念范围内知识的完全泛化。但对于开创性研究，AI目前还做不到。因为在训练AI时总是要提供足够的算力和高质量的数据库，但是开创性研究是从无到有，想他人之所未想，并且有些更是反直觉的，而目前ChatGPT的计算原理是通过统计关联，对输入信息加权来实现直接计算，对于思维的小概率事件AI则无法评估其影响。而且人类从0到1的开创性研究往往取决于人的品味、审美、生活经历，但AI既没有生活经历，也没有社会体验，反而成为了AI的短板。所以AI的增强只能会对传统教育和渐进式科研产生巨大冲击。

总结来说，我们在20世纪基本解决了能源问题，未来可能需要应对的是如何闭环利用能源，

参考文献

- [1] Hopfield J J. Annual Review of Condensed Matter Physics, 2013, 5:1
 [2] Hopfield J J. Proc. Natl. Acad. Sci. USA, 1982, 79(8):2554
 [3] 黄海平. 科学, 2022, 74:40
 [4] Huang H. Front. Comput. Neurosci., 2024, 18:1388166
 [5] Bower J M. (NIPS) NeurIPS and Neuroscience: A Personal Historical Perspective, 2022
 [6] Ackley D H, Hinton G E, Sejnowski T J. Cognitive Science, 1985, 9:147
 [7] Stone J V. The Artificial Intelligence Papers: Original Research Pa-

pers With Tutorial Commentaries. Sebteel Press, 2024
 [8] Hinton G E. Neural Computation, 2002, 14 (8): 1771
 [9] Wang H C *et al.* Nature, 2023, 620:47
 [10] Hinton G E, Salakhutdinov R R. Science, 2006, 313(5786):504
 [11] Carleo G *et al.* Rev. Mod. Phys., 2019, 91:045002
 [12] 黄海平. 神经网络的统计力学(英文版). 北京: 高等教育出版社, 2021
 [13] 黄海平. 现代物理知识, 2024, 36 (6):49
 [14] Goldenfeld N, Kadanoff L P. Science, 1999, 284:87
 [15] Bubeck S *et al.* 2023, arXiv:2303.12712

同时打开AI的能力边界。发展高维计算的统计物理在这一过程中将扮演关键角色，同时也将引发人类对于“思考”和“认知”的完全理解。这可能需要新的数学和计算机架构。在这一目标还远未达到的今天，人们也应该警惕目前盛行的生成式算法可能会在某种程度上扭曲人类原有的文明，甚至把新生代的人类大脑塑造成数字大脑(digital brain)而不是生物大脑(biological brain)。因为大脑接触到的信息一旦发生改变，大脑就会调整自己的突触，最后人人都可能变成数字计算机。所以在仰慕AI能力的同时，我们应该加强AI的治理和安全管控，高度重视AI背后运行规则的理论研究。

2024年的诺贝尔物理学奖将开启一个思考“智能的本质是什么”的时代！物理学，尤其是统计物理将变得越来越重要，而且在探索智能的道路上将与统计学、计算机科学、认知科学和脑科学高度交叉融合。

读者和编者

《物理》有奖征集封面素材

为充分体现物理科学的独特之美，本刊编辑部欢迎广大读者和作者踊跃投寄与物理学相关的封面素材。要求图片清晰，色泽饱满，富有较强的视觉冲击力和很好的物理科学内涵。

一经选用，均有稿酬并赠阅该年度《物理》杂志。

请将封面素材以附件形式发至：physics@iphy.ac.cn；联系电话：010-82649029。

《物理》编辑部