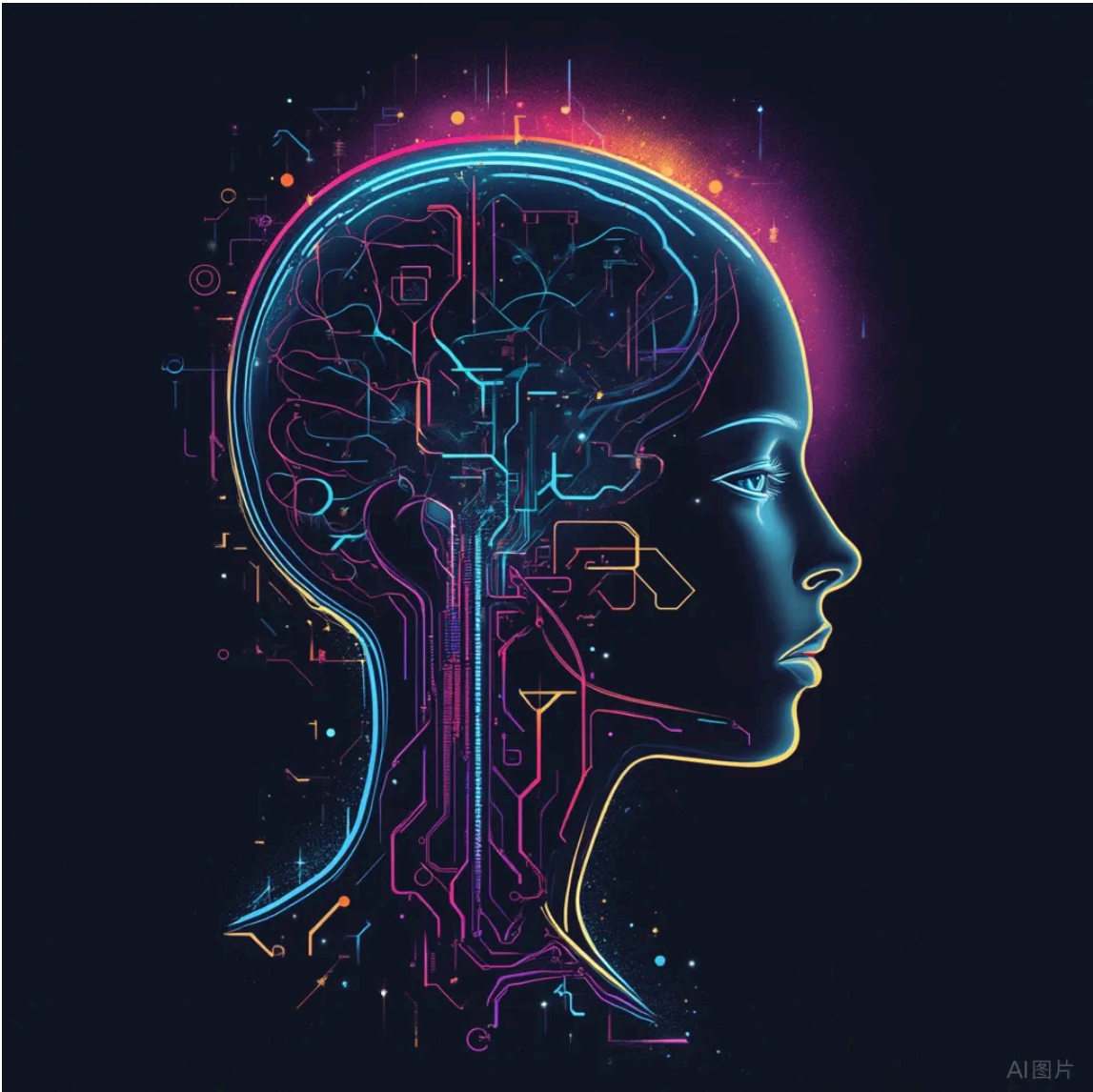


Nature综述：大规模神经形态计算

原创 Kudithipudi等 集智俱乐部 2025年05月03日 21:48 上海



摘要 神经形态计算（Neuromorphic Computing）指脑启发的硬件与算法设计方法，研究者们借鉴神经科学中的生物智能原理来设计高效的计算系统，尤其适用于对体积、重量和功耗有严格要求的应用场景。当前，该研究领域正处于发展的关键阶段，因此明确未来大规模神经形态计算系统的发展方向至关重要。本文探讨了构建规模可扩展的神经形态计算架构的方法，并总结了其中的关键特征。此外，我们分析了可以从规模扩展中获益的潜在应用场景，以及需要解决的主要挑战。进一步地，我们审视了支持该领域持续发展的完整技术生态系统，并探讨了规模扩展所带来的新机遇。我们的研究汇总了多个计算子领域的观点，为神经形态计算研究人员和从业者提供指导，以推动该领域的进一步发展。

研究领域：神经形态计算，脑启发，生物智能

Dhiresha Kudithipudi, Catherine Schuman, Craig M. Vineyard等 | 作者

Shenky20 | 译者

[nature](#) > [review articles](#) > [article](#)

Review | Published: 22 January 2025

Neuromorphic computing at scale

[Dhiresha Kudithipudi](#) , [Catherine Schuman](#), [Craig M. Vineyard](#), [Tej Pandit](#), [Cory Merkel](#), [Rajkumar Kubendran](#), [James B. Aimone](#), [Garrick Orchard](#), [Christian Mayr](#), [Ryad Benosman](#), [Joe Hays](#), [Cliff Young](#), [Chiara Bartolozzi](#), [Amitava Majumdar](#), [Suma George Cardwell](#), [Melika Payvand](#), [Sonia Buckley](#), [Shruti Kulkarni](#), [Hector A. Gonzalez](#), [Gert Cauwenberghs](#), [Chetan Singh Thakur](#), [Anand Subramoney](#) & [Steve Furber](#)

[Nature](#) **637**, 801–812 (2025) | [Cite this article](#)12k Accesses | 109 Altmetric | [Metrics](#)

Abstract

Neuromorphic computing is a brain-inspired approach to hardware and algorithm design that efficiently realizes artificial neural networks. Neuromorphic designers apply the principles of biointelligence discovered by neuroscientists to design efficient computational systems, often for applications with size, weight and power constraints. With this research field at a critical juncture, it is crucial to chart the course for the development of future large-scale neuromorphic systems. We describe approaches for creating scalable neuromorphic architectures and identify key features. We discuss potential applications that can benefit from scaling and the main challenges that need to be addressed. Furthermore, we examine a comprehensive ecosystem necessary to sustain growth and the new opportunities that lie ahead when scaling neuromorphic systems. Our work distils ideas from several computing sub-fields, providing guidance to researchers and practitioners of neuromorphic computing who aim to push the frontier forward.

论文题目: Neuromorphic computing at scale

论文地址: <https://www.nature.com/articles/s41586-024-08253-8>

随着神经网络在越来越多的应用领域产生广泛影响，人类大脑仍然是研究者们进一步模拟复杂计算能力并实现突破的重要灵感来源。然而，要实现受大脑启发的机器学习需要在计算平台的设计与构建方式上进行根本性的变革。在这一方向上，最具前景的研究之一是神经形态计算（Neuromorphic Computing），即受大脑启发的硬件与算法设计方法，能够高效实现人工神经网络[1]。神经形态计算领域研究者借鉴神经科学提出的生物智能（Biointelligence）原理来设计高效的计算系统，尤其适用于对体积、重量和功耗有严格限制的应用场景。

根据近年来神经形态计算原型（prototype）系统的发展速度进行推测，该领域在未来人工智能（artificial intelligence）应用中具有巨大的潜力。预计到2026年，神经形态计算芯片的市场规模将达到5.566亿美元[2]。一些神经形态计算芯片正快速进入早期商业市场，并已展现出在不同规模的计算任务中实现极低功耗和低延迟性能的潜力[3,4]。这一领域迅速发展的原因之一在于其系统的多功能性。例如，传统计算的进步通常专注于特定类型的计算架构——超算领域聚焦于百亿亿次（Exascale）计算，而嵌入式系统则强调小型化，这两种架构通常不会相互借鉴，而神经形态计算则能够通过统一的计算技术在这两类计算架构中同时带来变革，具备颠覆性的潜力。当前该领域面临的关键问题在于神经形态计算是否已经准备好实现真正的计算突破，类似于“AlexNet时刻”（见框1）。此外，评估某种方法的成熟度远比计算单一的性能指标更加复杂，规模化是衡量该领域进展的重要维度之一。如今神经形态计算正处于关键发展阶段，我们的目标是识别并解决关键需求，从而推动变革性影响的产生。

与传统深度学习加速器相比，神经形态计算系统（Neuromorphic Computing Systems）具备以下几个显著的计算优势：

1. 计算与存储紧密耦合，避免了计算单元与存储设备之间高成本的数据传输。
2. 基于脉冲或事件的稀疏分布式信息编码，能够携带时间信息。
3. 动态且局部的学习机制，无需通过反向传播在多个层级上传播误差，从而降低能耗。
4. 通过学习实现稳定的感知，并能够对感官信号进行预测。
5. 跨多个时间尺度进行动态计算，支持实时学习与处理。

然而，这些特性并非“万能解”。以上优化也可以通过完全自上而下的工程方法来实现，但我们认为，借鉴进化过程中形成的解决方案可能会带来更快速的发展路径。神经形态计算正是这一方向上的探索之一，其独特优势在于紧密联系神经科学与生物学，而后者已经通过自组织、动态重连、三维生长、模块化、高效信号编码、稀疏计算、基于事件的计算等机制，成功解决了类似的挑战。因此，神经科学与生物学原理有望为大规模神经形态计算系统的设计提供灵感。

在图1中，我们展示了神经形态计算系统的发展时间线，涵盖了该领域至今所取得的关键技术进展。这些系统的底层架构标志着计算在复杂性、通用性和异构性方面的重要里程碑。然而，在计算堆栈的每个层级仍然存在诸多挑战，必须加以解决，才能使大规模神经形态计算机真正走向实用化并被广泛应用。

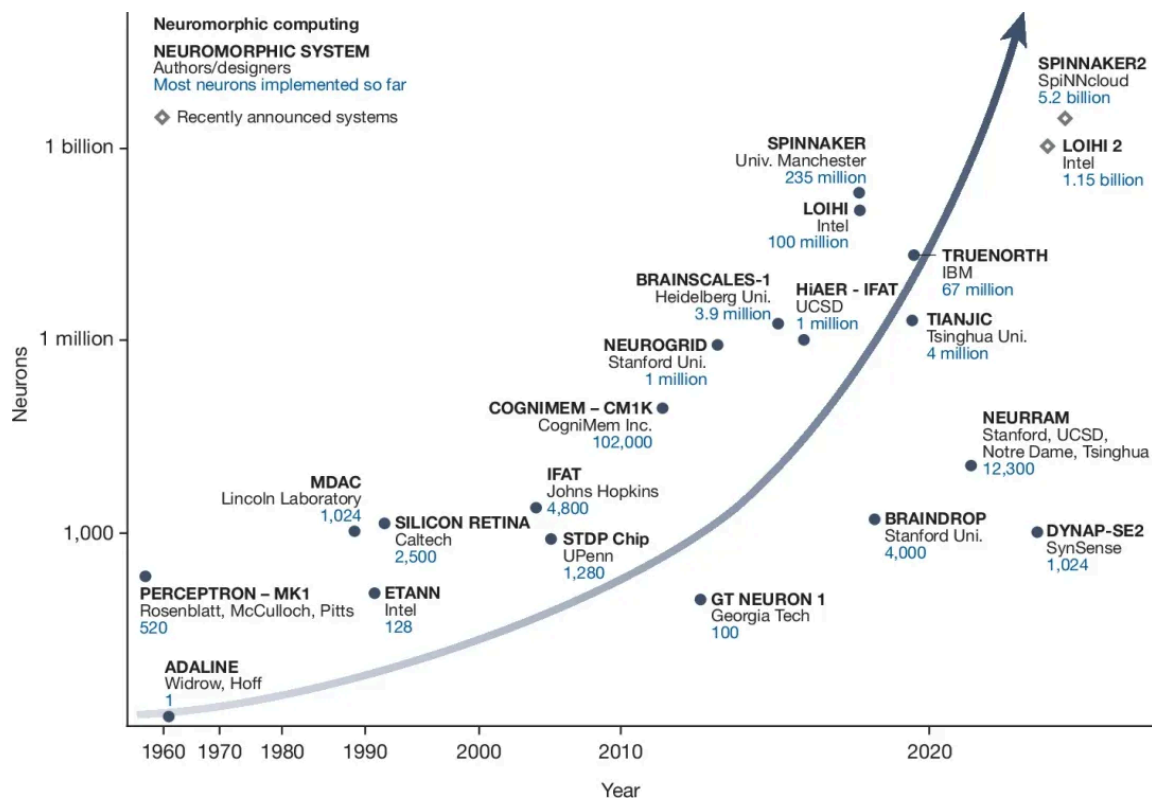


图1：神经形态计算系统的发展进程

图1展示了神经元数量（y轴）、芯片与计算板的规模随时间（x轴）增长的趋势。尽管这仅代表规模扩展的一个维度，但在这一过程中，计算架构和通信模型也在不断演进以支持系统的扩展。随着神经形态计算的发展，对规模化能力的提升需要综合考虑架构设计、部署策略和工具链等多个方面。为简洁起见，图中仅展示了部分公开记录的神经形态加速器[3,37,39,40,41,49,102,109,110,111,112,113,114,115,116,117,118,119,120,121]。目前，部分系统（如DeepSouth）采用现成芯片（例如现场可编程门阵列（FPGA, Field-Programmable Gate Array））以实现规模化扩展。而其他架构（如IBM NorthPole122）则探索通过类神经功能加速神经网络计算，其中神经元并非核心计算单元，因此这些架构的进展未纳入图中。神经形态计算领域正快速发展，预计2026年将出现拥有超过100亿个神经元的神经形态计算系统。

在本文中，我们探讨了可扩展且实用的神经形态计算基础设施[5]的本质、需求、重要性及面临的挑战。具体而言，我们分析了规模化计算的关键特征，并讨论如何使大规模神经形态计算基础设施更广泛地服务于不同领域的研究人员和从业者。此外，我们深入研究了神经形态计算的计算堆栈及工具链，识别了其中的关键挑战与潜在机遇。我们期待这些观点能够激发新的思考和跨学科合作，从而加速大规模神经形态计算系统的发展。

框1：神经形态计算的“AlexNet时刻”

在探索大规模神经形态计算的潜力时，我们可以从深度学习革命的演进历程中获得启发。两者都建立在不同层次的神经网络理解之上，并经历了长期的技术瓶颈期。类似于卷积神经网络（CNN, Convolutional Neural Network）在手写数字识别[140]方面的成熟，神经形态计算正处于即将迎来“AlexNet时刻”的阶段

——一个能够突破现有局限、充分展现其潜力的关键转折点[141]。本文总结了当前亟待解决的问题，这些问题的突破将加速神经形态计算的发展，使其影响力可与深度学习相媲美。

神经形态计算的进展可能取决于专用硬件，正如AlexNet（一种经典的卷积神经网络，由 Alex Krizhevsky、Ilya Sutskever 和 Geoffrey Hinton 在 2012 年 ImageNet 图像分类竞赛中提出，将图像分类的正确率提升到前所未有的高度）的成功依赖于通用图形处理单元（GPU）的性能提升。AlexNet的出现引发了深度学习模型的规模化浪潮，带来了更强的计算设备[142]、更大规模的神经网络[143]以及显著提升的芯片性能[144]。类比而言，为了释放其潜力，神经形态计算需要确定其关键的硬件需求。虽然当前大规模深度学习系统依赖数千颗加速芯片，但神经形态计算的突破可能来自一个小规模硬件配置——毕竟，AlexNet最初仅使用了两块GPU。如果神经形态计算能够在小规模系统上实现突破，则将激发更大规模系统的研发，进而重复深度学习的成功路径。因此，在小规模神经形态计算中涌现的“AlexNet时刻”，有望为未来大规模神经形态计算的落地应用奠定基础。

神经形态计算的规模化发展

我们将大规模神经形态计算定义为一个包括算法、硬件、架构和基础设施的计算系统，其具备在规模、速度和能耗等方面满足复杂现实任务需求的能力。这一目标可以通过在数据中心中利用多个大规模系统进行虚拟化访问，通过边缘设备构成的大型网络实现分布式智能，或两者相结合。要推动神经形态计算的规模化发展，必须超越实验室中的概念验证，实现面向现实任务的规模化部署。一旦规模化落地，神经形态计算可能会被数百万用户所采用，无论他们是否意识到其存在。因此规模化落地能够标志神经形态计算从传统研究范式向实际应用的重大转变。高性能计算（HPC）领域中近期神经形态应用的增长可以提供初步示范。大多数科学计算的高性能计算系统是基于计算资源，可以根据不同科学任务的需求进行资源扩展。尽管神经形态计算在本质上不同于冯·诺依曼（Von Neumann）架构，但其通用计算架构或许非常适合HPC场景中的按需计算资源：同一套神经形态系统既可以用于大规模科学计算模拟，也可以支持小型化的边缘计算和分布式智能应用。

随着神经形态计算系统的发展，规模化能力的提升需要综合考虑开发、部署和工具链等多个方面。目前，虽然大规模神经形态计算系统的研究正以极快的速度推进，但直到最近才开始产生较大影响。尽管未来的具体发展难以预测，但我们预计其带来的变革性影响不会亚于人工智能革命。

长期以来，微电子产业一直以“规模”作为创新进展的核心度量标准。例如：摩尔定律（Moore's Law）描述了硬件密度的提升趋势；超级计算机的计算能力以每秒浮点运算次数（FLOPS, Floating-Point Operations Per Second）来衡量。这一统一度量方式，为从资源受限的微控

制器到满足HPC需求的服务器级处理器提供了性能评估依据。因此，以类似的扩展度量体系衡量神经形态计算的规模化发展是合理的。这种方法既能量化计算性能（计算操作数量），也能直观地类比神经形态系统的神经元规模，例如将其与不同动物大脑的神经元数量进行对应[6]。然而，尽管这些度量方式能够反映计算能力的提升，但它们低估了神经形态计算的独特特性。规模化发展带来了制造、测试和可靠性等方面的挑战，尤其是在非受控环境下的性能表现、基础设施适配和用户可用性等方面。

目前，该领域的科研与初创投资正在空前增长，同时器件与架构也在不断趋于成熟[2]。神经科学正在加深对身体机制的理解[7]，进而启发神经形态计算工程师，推动感知计算系统的创新，并拓展至新的应用领域[8]。已有多个领域的概念验证案例展示了神经形态计算的潜力，例如科学计算，人工视觉[9,10]，机器人技术[11]，生物信号处理[12]，太空计算，计算神经科学[13,14]。这些案例（见图2）证明了神经形态计算的可行性，但与真正的规模化发展之间仍然存在较大鸿沟。推动神经形态计算规模化的关键在于明确规模化系统所需的核心特性，以及确定最有利于其广泛应用的条件。在本文中，我们将重点探讨对神经形态计算的研究、工程设计和应用起到关键作用的各个方面。

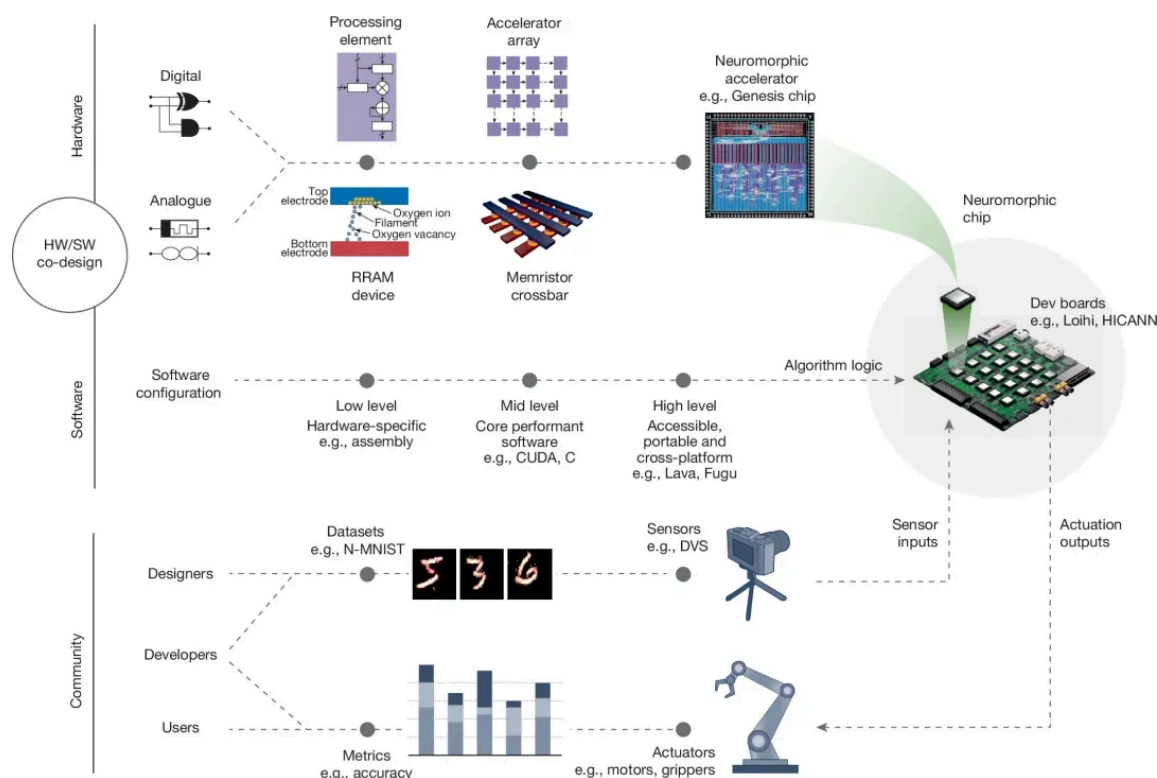


图2：神经形态计算生态系统

本文所描述的神经形态计算生态系统由以下几个层次构成：

1. 硬件与软件技术：硬件包括神经形态芯片与计算板[123,124,125]，软件涵盖则神经形态计算框架与工具链⁶⁶ [66,67,72,77,126,127]。
2. 硬件/软件协同设计 (Co-Design)：通过软硬件联合优化，提高系统的计算效率和灵活性。
3. 工作流 (Workflow)：形成完整的神经形态计算生态[4,128,129]，支持不同任务的适配。

4. 原型系统及应用领域：基于嵌入式传感与执行系统（Sensors[55] & Actuators[130,132]）的神经形态计算板（Prototype Boards）在具体领域[9,10,14,128,130,131]的应用。
5. 外部工具及框架集成：通过一系列外部工具[133,134,135,136,137]，与神经形态计算框架进行集成。
6. 社区反馈与研发推动：研究社区的反馈机制起到研发加速器的作用。

在数据中心中利用多个大规模神经形态系统进行远程访问，通过边缘设备构成的大规模智能网络，实现分布式计算，或者结合数据中心计算与边缘智能，从而形成完整生态，为神经形态计算的广泛应用提供了基础架构支持。

神经形态计算的关键特性

如何确定神经形态计算系统相比于深度学习加速器或传统冯·诺依曼（Von Neumann）处理器，具有更高效率和更强扩展性的特性，这仍然是一个挑战。因此，我们识别出一系列对于实现神经形态计算优势的关键特性（见图3）。需要强调的是，这些特性并不会取代神经形态计算的核心特性[2]，而是为了实现规模化部署需要进一步优化。

1. 分布式与层次化 (Distributed and Hierarchical)

神经形态计算系统应当具备分布式计算能力，并采用层次化结构，类似于大脑视觉皮层（Visual Cortex）等区域的层级组织方式[15,16]。这种架构能够解耦复杂输入[17,18]，在不同层次处理不同尺度的信息[18]，增强大规模模型的控制能力从而提高模型的可解释性，通过信息解耦（Disentangling）简化数据并减少计算冗余，避免线性或非层次化网络[19]的低效计算，不需要特殊机制[20]或监督。因此，支持分布式与层次化计算是神经形态计算实现规模化的关键因素之一。

1. 稀疏性 (Sparsity)

人脑在活动与连接模式上表现出高度的稀疏性，这是其可扩展性的关键因素。研究表明，大脑的突触在发育过程中经历稀疏连接、密度增强、大规模剪枝和稳定稀疏性维持的过程[21]。其中稀疏性可以降低表示复杂度，仅使用必要的神经元和连接，提升计算、存储和能耗效率，而不损失精度。稀疏化可以分为结构性稀疏（Structural Sparsity），如神经元、突触权重、注意力头的剪枝，以及瞬时稀疏（Ephemeral Sparsity），如激活值、梯度、误差的稀疏化[22]。尽管神经形态计算天然支持事件驱动（Event-Driven）计算[23,24,25,26]，但仍需进一步探索稀疏机制以实现更大规模计算和泛化能力。此外，稀疏性还可能通过采用非传统的空间布局[27]来加快模型扩展速度。

1. 神经元可扩展性 (Neuronal Scalability)

神经元可拓展性指能够支持大规模神经元的单芯片或多芯片系统，使深度脉冲神经网络（Spiking Neural Network, SNN）或基于脉冲频率的算法能够在各类机器学习应用中解决复杂现实问题[23,28,29]。这种可扩展性还使我们能够实时模拟完整的人脑，从而推动认知计算和神经科学研究的发展。目前规模化计算可以通过堆叠多个神经形态芯片来实现（见图1），支持数亿个神经元的计算能力，极大地拓展了应用范围，例如求解非确定性多项式时间（NP）完全问题、运行大规模神经模拟或执行复杂图算法[30,31]。

要推动神经形态计算的更广泛应用，必须引入可集成不同计算资源和神经形态元件[32,33]的特性，并支持外部工具和传感器的融合，这种能力被称为异构集成（Heterogeneous Integration）[34,35]，即能够在单个神经形态系统内整合多种设备技术，进行多传感器数据融合，并支持大规模混合人工神经网络（ANN）和脉冲神经网络框架的部署。

1. 异步通信（Asynchronous communication）

许多神经形态芯片采用基于事件的异步通信协议来支持地址事件表示（AER, Address Event Representation）架构下的数据收发。地址事件（Address Events）作为芯片的输入，代表接收输入事件或脉冲的神经元[36]。这些协议已有多个变体，并能够在不同芯片架构中实现，如第一代芯片（如硅视网膜感知系统）、皮层认知处理系统[3,37,38]、复杂片上网络（NoC, Network on Chip）架构[32,39,40,41]等。这些进展使得异步通信技术更加成熟，并促进了大规模系统的集成。

1. 动态可重构性（Dynamic reconfigurability）

大脑本质上是一个动态系统。研究表明，执行认知任务需要大脑区域之间的动态网络重组，并形成复杂且瞬时的通信模式[42]。为了实现类似功能，多个神经形态系统需要支持不同形式的动态可重构性[23,43,44,45,46]。这种能力可以作用于基础计算单元（如突触、神经元、轴突、树突）或整体系统。在某些芯片中，突触连接的可重构性和一对多通信能力可以用于线性滤波，包括边缘检测和平滑运算。

1. 冗余性与相关性（Redundancy and correlation）

神经元执行多种任务的能力部分归因于神经冗余性。神经元之间的相关性影响着数据在神经元群体中的编码方式和群体神经计算的解码方法[47]。研究表明，冗余的活动模式可能对神经计算有利[48]，使计算在神经动态不稳定的情况下仍能保持稳定并有效滤除噪声。尽管已有多个神经形态系统[3,32,39,49]支持类似机制（如跳零（Skip-Zero）与零权重（Zero-Weight）方法），但仍然有扩展空间。

1. 传感器与计算端的接口（Sensor and compute interfaces）

除核心计算系统外，应用的端到端开发还需要将外部传感器和执行器高效集成至神经形态计算工具链。目前，大多数商用传感器采用非脉冲式协议，需要内核级接口将数据转换为脉冲编码格式。优化的神经形态驱动可以极大提升低延迟任务（如音频和视觉处理）的性能[50,51]。通用神经信息格式（Neural Information Format）的采用[52,53,54]能够使神经形态驱动更加通用化，避免硬件依赖。神经形态传感器（如事件驱动相机[55]）可进一步减少响应时间，适用于高速运动捕捉任务。医疗设备（如硅电子耳蜗[56]、电子皮肤假肢[57]）则可提升感知的精度与灵敏度。

1. 资源感知 (Resource awareness)

为了提升效率与多功能性，加速器的设计必须具备资源感知能力。资源感知指系统在运行过程中追踪自身的能耗、计算与存储资源，并根据任务需求动态调整资源分配的能力。自适应（Self-Aware）架构和持续学习（Continual Learning）[58]加速器提供了现这一特性的可借鉴思路。

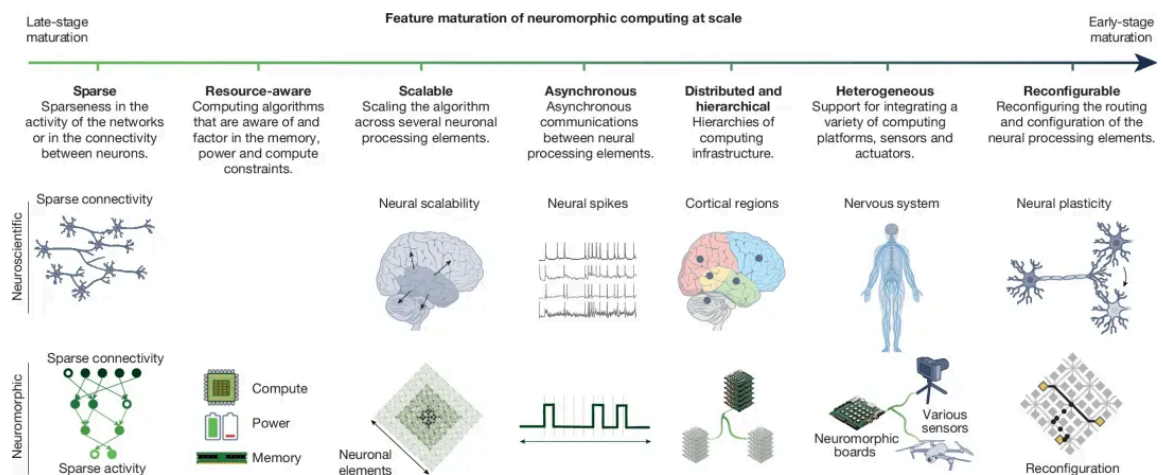


图3：大规模神经形态计算系统的关键特性及其成熟度演进时间线

这些特性增强了神经形态计算系统的核心特征，并在多个方面直接借鉴了大脑及神经处理机制[138,139]。不同特性的成熟度发展具有不同的时间尺度：部分特性将在短期内达到较高的成熟水平，另一些特性可能需要更复杂的技术手段才能实现相同水平的成熟度。

每种关键特性都可以从以下几个方面进一步探讨其影响，例如功耗（Power），性能（Performance），可扩展性（Scalability），自适应能力（Adaptability）和集成的多样性（Versatility of Integration）。此外我们还注意到，一些本质上属于神经形态计算的特性同样适用于大规模传统计算系统的设计。例如极端并行计算（Extreme Parallelism）是大规模超级计算机架构的基础，并不局限于神经形态计算。而在模型扩展性方面，调整扩展曲线的斜率可以训练比当前更大规模的模型，为突破现有计算能力限制带来一种可能性，或者在相同算力下使用较小模型但进行更长时间的训练、使用更大规模的数据集。这使得未来趋势可能是训练大量小型模型，每个模型针对特定任务的数据集优化，从而避免了单一大模型在小众任务上的性能下降，

尤其是在长尾任务和小样本应用中表现更优。因此，尽管某些计算设计概念并非神经形态计算独有，在统一计算模型中的协同作用才是其真正的价值所在。

挑战与机遇

在总结一些关键挑战与机遇时，我们需要注意的是目前对大脑的理解仍然有限，而神经形态计算则有望为加深对大脑认知提供机会。

尽管神经形态计算已取得了重大进展，并展现出在多个社会应用领域构建可持续且稳健技术的颠覆性潜力，但在其广泛落地并产生实际影响之前仍需克服诸多挑战。这些挑战涉及神经形态硬件设计者、用户及算法开发者，涵盖从软硬件协同设计（Co-Design）（见框2）到技术采纳与商业化等多个层面。

从开发者的角度来看，神经形态计算面临多个方面的挑战：

1. 硬件易用性：需要提供更高级的编程抽象（类似于从汇编语言向面向对象编程的过渡），使开发者无需深入理解底层硬件即可使用。由于神经形态硬件的多样性，模型在不同平台上的复用性较差，从而使得算法与特定硬件紧密绑定。计算原语（Computational Primitives）和硬件约束在不同平台间差异显著，因此需要制定通用的硬件与软件标准以减少跨平台算法的适配成本。
2. 跨平台兼容性：硬件抽象层（HAL, Hardware Abstraction Layer）的引入对于神经形态计算平台至关重要，这有助于构建通用编译方案（Compilation Scheme），使任意脉冲神经网络（SNN, Spiking Neural Network）模型能够移植到不同的硬件架构上。目前的神经形态计算生态系统仍然缺少一些关键组件（详见图4）。在现有的脉冲神经网络框架中，互操作性仍然有限，与传统人工神经网络框架的高互联性存在显著差距。
3. 开发体验优化：缺乏推理优化器、拖拽式编辑器和跨平台兼容性工具，影响开发的便捷性，并增加了新用户的学习成本。对于资深开发者和大型企业而言，缺乏集成的云端扩展、部署和生命周期管理资源不利于大规模开发和应用。综上，领域迫切需要跨框架与跨硬件的交叉编译（Cross-Compilation）工具以减少重复开发，并促进生态系统之间的迁移。近年来在弥合这些差距方面已经取得重大进展，部分关键技术空白正逐步被填补，例如AI视觉工具包、模拟器/分析工具和自动超参数搜索框架已集成到现有工具链中[40,50,52,53,59,60,61,62,63,64,65,66,67,68,69,70,71,72]，对动态架构和低级计算原语模拟的支持已可与主流AI工具相媲美。但虽然低级神经元计算原语的交换格式已被广泛采用，高层神经元拓扑结构的交换格式在不同框架间仍然缺乏统一标准。主流AI/ML工具链曾面临类似的挑战，最终通过ONNX格式的联合开发和广泛采用得以缓解[73,74]。因此，神经形态计算领域同样需要推动类似标准的建立或采用，以规范脉冲神经网络模型的描述与交换。
4. 提高工具与平台的可访问性：提供全面的文档支持和健全的社区反馈机制，将有助于吸引更多新用户，促进神经形态计算的发展。这一进展将推动集成系统组件的开发，形成可复用的生态系统（例如ROS、Linux），而不是专注于特定硬件任务的孤立功能模块。

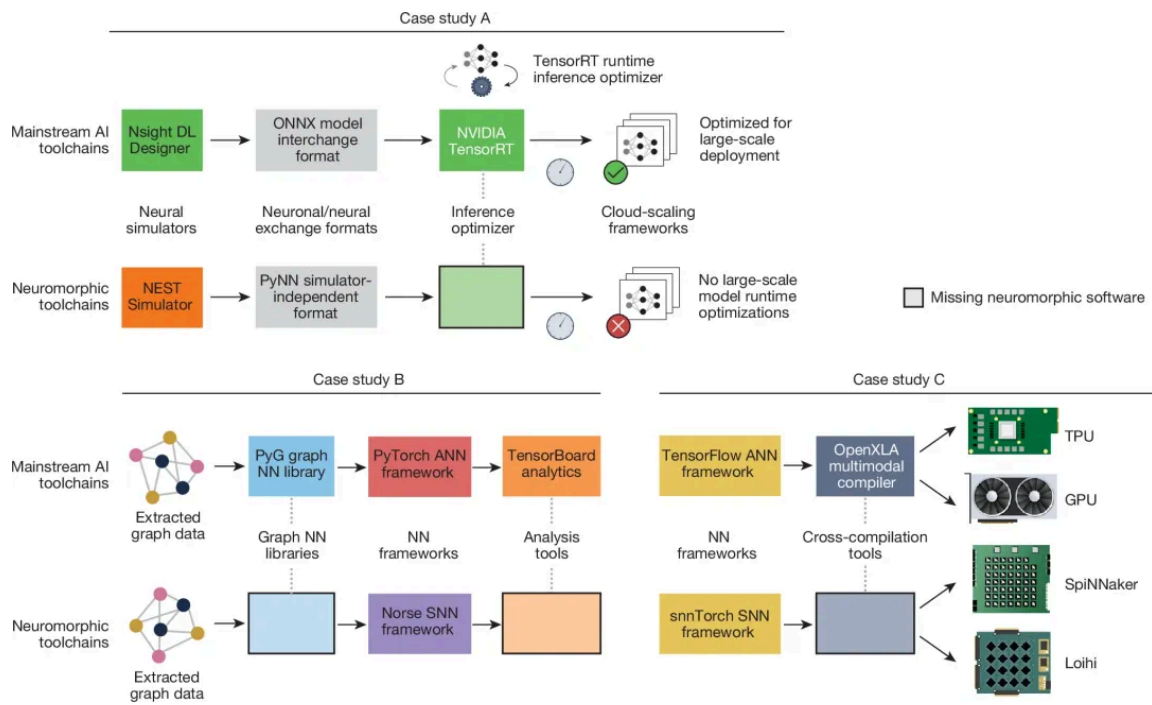


图4：神经形态计算软件生态系统中的缺口对比（与AI/ML案例研究对照）

主流人工智能/机器学习生态系统（图中上半部分，每个案例研究对应一列）已经建立了完善的开发管道（Pipeline），并且在框架、编译器和硬件之间实现了广泛的互操作性。而相比之下，神经形态计算生态系统（图中下半部分，每个案例研究对应一列）仍然存在较多关键缺失组件，需要进一步完善。识别并填补这些空白对于构建更全面、协作性更强的神经形态计算生态系统至关重要。我们提供了三个案例研究进行对比分析：案例A，大规模部署与推理优化；案例B，图神经网络（Graph Neural Networks, GNN）的开发与分析；案例C，AI模型到通用或特定硬件的编译。图中心标注了AI/ML生态系统和神经形态计算系统中共有的软件组件，而神经形态计算系统缺失的软件组件则以空白块表示。当前神经形态计算生态系统尚未完善的部分，需要进一步的工具和框架开发，以实现与人工智能/机器学习领域相类似的互操作性和生态成熟度。

科学研究社区偏好工具的灵活性，而终端用户和产品设计师更注重工具的效率。因此，提高神经形态计算系统采用率的一种策略是借鉴早期AI工具链（如Torch和Theano[75,76]）的发展模式，即在早期开发阶段优先考虑灵活性而非效率，从而吸引更多广泛的科学研究社区参与开发。这种社区驱动的发展模式促使更高效的后端（Back-End）计算框架的诞生，例如PyTorch[77]。

与此同时，应用开发者面临的一大障碍在于神经形态计算系统需要一种根本上不同的编程范式，即结合神经计算原语（Primitives）来编写“程序”。这对没有神经形态计算或计算神经科学经验的研究人员和工程师构成了进入门槛。因此，需要建立坚实的理论基础来指导如何使用神经计算模块（Building Blocks）来实现计算。例如，常微分方程（ODE, Ordinary Differential Equation）求解器可用于建模神经动力学。目前已有一些初步工作采用了数学建模方法来研究这一问题，例如神经工程框架（NEF, Neural Engineering Framework）[78]和动态神经场（Dynamic Neural Fields）[79]。

从设计者的角度来看，神经形态计算系统的开发需要贯穿整个计算栈（Stack）的各个层级：器件、电子电路、计算架构、算法和应用的发展应在神经形态系统中无缝集成，高度同步各个层级的进展以确保系统设计的整体性。值得注意的是，为了使这些系统能够扩展至复杂任务，系统组件必须具有模块化（Modularity）特性。这引出了一个核心问题：神经形态计算系统当前使用的神经元和突触是否是正确的抽象级别？目前领域的主流实现基本都基于神经元-突触模型，但这一抽象是否适用于未来的规模化发展仍有待讨论。此外，神经形态算法通常强调生物可解释性（Biological Plausibility），但这未必是实现最佳任务性能、可扩展性和高能效的最优选择。部分原因在于，神经形态计算的基础架构与生物神经系统存在根本性差异，并且生物系统的复杂性远超当前简化的生物计算模型，许多层级的复杂性未能在现有模型中体现。挑战在于如何开发合适的算法抽象层级，使我们能够在清晰而聚焦的框架下讨论算法的效率与可扩展性，同时忽略对不重要生物学细节的计算。要实现神经形态算法的规模化还需要特别关注异步和分布式算法，但这一方向的研究主要由不同于深度学习与神经形态计算主流社区的研究群体推动，尚未形成紧密的协作。确定关键接口需求（如高速传感器集成、多种传感器兼容性）同样是系统设计阶段必须解决的重要挑战。最后，如何持续吸引神经科学家参与跨学科合作，特别是在神经形态研究人员主导部分神经科学实验的情况下，也将是未来需要重点解决的问题之一。

从用户的角度来看，重大挑战则在于如何确定适合的应用领域以充分发挥神经形态计算的优势。一个有用的类比是量子计算（Quantum Computing）的发展。量子计算之所以能够获得数十亿美元的投资，部分原因在于其存在形式化的复杂度优势，例如Shor 算法在整数因子分解问题上的优势使得其在密码学领域具有实际影响，即量子计算的理论复杂度证明能够为其提供明确的价值支撑。相比之下，尽管已有研究表明在某些特定任务上神经形态硬件在性能或能耗方面具有优势，但仍缺乏类似的形式化复杂度证明，这几乎肯定影响了对神经形态计算硬件的投资。换句话说，量子计算的理论证明比神经形态计算的“存在性证明”（即人脑的计算能力）更具吸引力。

许多神经形态计算展现出潜力的任务，例如终生学习（Lifelong Learning）或持续学习（Continual Learning）[58,80]、在稀疏数据环境中的学习、在噪声和变化条件下的鲁棒性、基于实时传感器数据的超低功耗机器学习等，都难以通过数学上可证明的方式来拆解为一系列难度递增的基准（Benchmark）测试任务。此外要实现闭环基准测试，神经形态芯片需要能够与传感器和执行器无缝连接，因而迫切需要建立标准化的通信协议。通信协议需要涵盖从低级电子电路实现到更高层，确保不同类型的芯片（包括传感器、分布式自适应处理器、多核百万神经元系统等）能够无缝交互，且能够兼容非神经形态计算系统，形成统一的计算生态。

框2：硬件/软件协同设计（Hardware/Software Co-Design）

从本质而言神经形态计算是一个跨学科领域，它融合了神经科学、计算机科学、工程学和人工智能。由于这些学科的研究方法和最佳实践各不相同，神经形态系统的设计也可以采用多种方法。一种设计神经形态系统的方法是从大脑的可塑性和多层级学习机制中汲取灵感，将其转化为计算模型，然后在硬件和软件层面协同开发，最终部署成系统。

1. 自顶向下 (Top-Down) 方法: 研究人员在不涉及详细的神经元或突触模型 [15,16]的前提下, 抽象出大脑皮层的层次结构, 并利用这些模型来指导硬件和算法的设计。在此过程中需要识别大脑中的关键计算架构和连接模式, 以便构建可扩展的神经形态计算系统。
2. 自底向上 (Bottom-Up) 方法: 设计者利用底层器件的固有特性来推动新的算法进展, 从而优化计算架构和系统[145,146]。该方法依赖于新兴技术的突破, 例如存内计算 (In-Memory Computing) [147,148,149], 新型器件技术 (Emerging Device Technologies) [150,151], 低精度算术计算 (Low-Precision Arithmetic) [123,152]等。

然而, 无论采用自顶向下还是自底向上的方法, 都会因算法设计者和硬件架构师对彼此的技术细节缺乏了解而引入低效问题。潜在的解决方案是采用硬件/软件协同设计 (Co-Design) 方法 [80,123,153], 即无论采用何种设计范式, 在整个设计过程中都始终考虑系统的可扩展性 (Scalability) 以提高硬件与软件的匹配度。

跨平台神经形态计算软件

主流深度学习的成功在很大程度上归因于强大且易用的开源软件工具的可用性。这些工具为用户和应用开发者提供了高度抽象的编程接口, 使他们无需深入理解机器学习的理论细节即可使用。例如TensorFlow、Keras和 PyTorch等框架屏蔽了底层数学计算的复杂性 (如自动微分), 让用户能够专注于高层次的网络结构, 而不必直接处理低级计算细节。相比之下, 神经形态计算在这一方面仍然存在明显不足。当前可即插即用的神经形态计算工具仍非常有限, 并且主要由少数研究人员开发和维护, 例如 jAER[81] (Java版的地址事件表示框架)。尽管社区驱动的支持和开源工具的发展具有潜力, 但如何定义适当的抽象层级以开发这些计算工具仍然是一个开放性问题。

目前, 面向神经形态硬件的神经网络建模工具仍然相对落后, 其开发水平大致相当于寄存器传输级语言 (RTL, Register Transfer Language), 而非高层次综合 (HLS, High-Level Synthesis)。这一成熟度不足的原因例如研究社区的发展方向分散, 各团队采用不同的方法; 算法多样性较高, 缺乏统一规范; 不同神经形态计算平台 (模拟[82]、数字[23]或软件实现[4]) 各自具有独特的实现需求, 导致通用工具的开发变得困难。近年来, 神经形态计算框架的功能有所增强, 能够支持多种神经元模型、不同的学习规则、更丰富的后端计算平台以及从主流深度学习框架 (如 TensorFlow、PyTorch) 导入模型[59]。然而神经形态计算框架在研究社区的有限应用, 并非仅由于缺乏工具, 而还在于缺乏由社区主导的共享标准。当前许多工具的开发仍然受限于特定商业机构的主导, 而不是由整个研究社区共同推进, 这阻碍了工具的广泛采用和生态统一。近期, 开源框架和通用中间表示 (Intermediate Representation, IR) 的建立[52]为推动标准化迈出了重要一步。

研究社区对少量核心软件框架的广泛接受, 使得硬件公司能够专注于开发尖端加速器, 而无需为每款新芯片重新设计完整的软件栈。这一优化反过来也促进了新芯片的采用, 推动神经形态计算生态的进一步发展。未来, 我们预计神经形态计算工具将朝着更高层次的抽象发展, 包括采用通

用的中间表示层，增强跨平台兼容性，以及从功能模块库中组装应用程序，其中每个模块由脉冲神经元（Spiking Neurons）组成。然而这些功能模块应具备哪些特性、如何开发这些模块，目前仍然是一个开放性问题，需要进一步研究和探索。

社区准备度与生态

在成熟企业和初创公司积极参与所形成的增长型生态下，神经形态计算领域有望从长期的基础研究积累以及日益增长的商业兴趣中受益。尽管当前已有系统能够精确模拟大脑的自底向上计算过程，但大多数研究者和业界人士仍不愿意放弃现有解决方案，除非他们能够清晰地看到神经形态计算带来的优势。在此，我们概述神经形态计算生态系统的关键特征，并提出一系列促进研究社区为大规模系统发展做好准备的考量因素。

研发团队 (R&D Groups)

如图5所示，为研发团队提供早期且便捷的技术访问渠道，尤其是超越传统神经形态计算生态系统的研究群体，能够形成技术推广的乘数效应，加速该新兴计算范式的优势传播，例如低能耗和小尺寸等优势特性。早期技术访问指向广泛的利益相关方开放硬件原型（Hardware Prototypes），使他们能够测试和验证该技术。此外，为了保持研究社区的持续参与度，可以通过设立奖项、竞赛、统一的基准测试、新成员入门指导、长期技术支持、专题研讨会等措施激励研发团队。目前已有多个社区推动的活动在推进这些努力，包括Telluride 和 CapoCaccia 研讨会、NICE（神经信息处理与计算工程会议）、ICONS（国际神经形态系统会议）、Intel INRC（Intel 神经形态研究社区）等。这些研究活动与合作平台有助于加强社区建设，并推动神经形态计算的广泛采用。

易用通用的开源软件 (Easy, Common and Open-Source Software)

正如前文“跨平台神经形态计算软件”部分所讨论的，易用性在神经形态计算软件生态中的核心作用主要体现在易用的软件访问（Easy Software Access），即开发编译器（Compilers）使主流用户能够直接部署模型至神经形态计算平台，而无需理解底层硬件细节。应该开发通用跨编译器，以便轻松集成新的后端（Back-Ends），促进硬件厂商遵循统一标准。采用通用、广泛接受的软件框架（Common and Accepted Frameworks）有助于降低硬件厂商的个体开发成本，避免重复造轮子。在软件栈的各个层级使用宽松的开源许可协议（如 Apache License 2.0）则可以鼓励社区贡献，促进生态合作，并消除商业公司的顾虑，让它们在不受严格限制的前提下积极参与神经形态计算软件的发展，推动技术的实际落地。

基准测试平台 (Benchmarks)

研究社区广泛接受的软件栈对通用进展衡量标准的采用具有直接影响。然而，由于该领域尚未完全成熟，并且神经形态计算系统本质上具有异构性，在定义标准化的度量指标（Metrics）和基准测试方面面临诸多挑战。神经形态系统的多样性导致算法和硬件平台之间难以直接比较，因此需要多维度的度量指标和多样化的基准测试。目前已有部分技术已经成熟，例如基于事件驱动的接口标准、I/O 神经元等，对其进行正式规范化可以为公平的基准测试奠定基础。硬件无关的软件栈可以作为建立通用基准测试的参考架构。现有的基准测试示例包括：受高性能计算启发的系统级基准测试[83]、由工业界主导的脉冲神经网络（Spiking Systems）基准测试[84]和用于概率系统（Probabilistic Systems）的基准测试工具[85,86]。数据集方面，已有多个专门用于评估脉冲神经网络（SNN）性能的数据集[87,88,89,90,91,92,93]，但这些数据集尚未被广泛采用[30]。硬件性能度量方面，目前尚无研究社区达成共识的统一指标。常见的单一度量标准包括能耗、功率、延迟（Latency）、准确率（Accuracy）、芯片面积、鲁棒性（Robustness）、噪声容忍度（Noise Tolerance）等。但单独使用这些度量指标通常只有在比较高度相似的设备或系统时才具有实际意义。因此，研究人员更倾向于采用复合指标，例如能量延迟乘积（Energy-Delay Product）、每个突触操作的能耗（Energy per Synapse）、相对准确率（Relative Accuracy）等。尽管如此，标准、度量指标和基准测试的开发仍然是一个关键且活跃的研究领域，其目标是实现未来的广泛采用。目前研究社区正在努力改进现状，通过定义一系列基准测试和相关度量指标，以比较不同的神经形态计算算法和硬件系统[94]。此外，还开发了用于评估神经形态系统性能的工具，适用于神经科学建模和心理学建模[95,96]。然而，要定义一个通用于所有神经形态计算系统的单一基准测试和度量标准仍然是一项巨大挑战，特别是需要考虑到“神经形态计算关键特性”部分所描述的各种系统特性。

跨领域技术（Field-Crossing Technology）

神经形态计算系统在传统计算与神经形态应用程序接口（API）之间的双向转换过程中，必须应对额外的接口损耗（Interface Losses）。在某些情况下，这种损耗源于脉冲编码方案的转换问题。解决这些接口损耗问题至关重要，否则即使神经形态计算引擎具备高效的计算能力，其整体优势仍可能因接口效率低下而被抵消。

概念验证（Proofs of Concept, PoC）

行业特定的概念验证是研究大规模神经形态系统的能效和性能优势的重要手段。为了最大化投资的影响力，对于任何新技术而言选择合适的概念验证方向都至关重要。概念验证的影响力更大时，往往能够在特定领域产生“涟漪效应”（Ripple Effect）。例如，设计一个面向终端用户的能效优化概念验证，其影响力可能远不及设计一个面向大型对话引擎技术服务提供商的概念验证。技术服务提供商直接受到能耗的影响，因此他们的认可和采用能够推动更大范围的终端用户受益。在概念验证阶段构建大规模原型而非采用渐进式开发方式，将带来更大的实际收益。

倾听反馈（Listening to the Feedback）

递归式反馈循环（Recursive Feedback Loop）有助于改进研究方法、优化概念验证，并评估技术的采用程度。维持以用户为中心的方法能够促使自底向上的方法论考虑实际应用需求，并推动硬件/软件的协同设计（Hardware/Software Co-Design）。

展望

神经科学探索 Neuroscience exploration

在迈向脑规模化模拟的过程中神经形态计算系统的应用变得尤为重要。一个典型案例是虚拟脑（The Virtual Brain, TVB）[97]，一个包含复杂生物学细节的脑级别模型，但在GPU上进行如此规模的实时模拟仍然极具挑战性。虚拟脑模型在医学领域具有重要应用，诸如个性化阿尔茨海默病（Alzheimer’s Disease）检测[98]，能够加速神经形态计算系统在医疗诊断中的应用，实现快速执行，提高检测效率。另一个例子是Markram 细胞研究（Markram Cell Study），该研究使用Blue Brain IV 超级计算机进行模拟。在当时Blue Brain IV 是全球排名第100的超级计算机，但其仅能模拟大脑桶状皮层（Barrel Cortex）中的 31,000 余个神经元，相比而言人类大脑大约包含比这多 100 万倍的神经元，神经科学研究若要进行如此规模的模拟将需要更高效的计算资源。

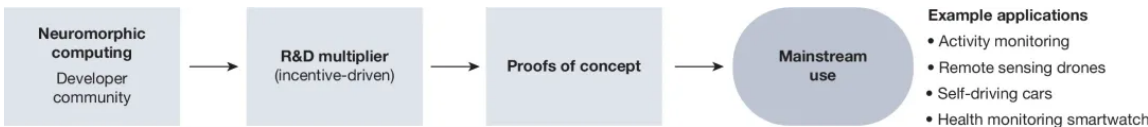
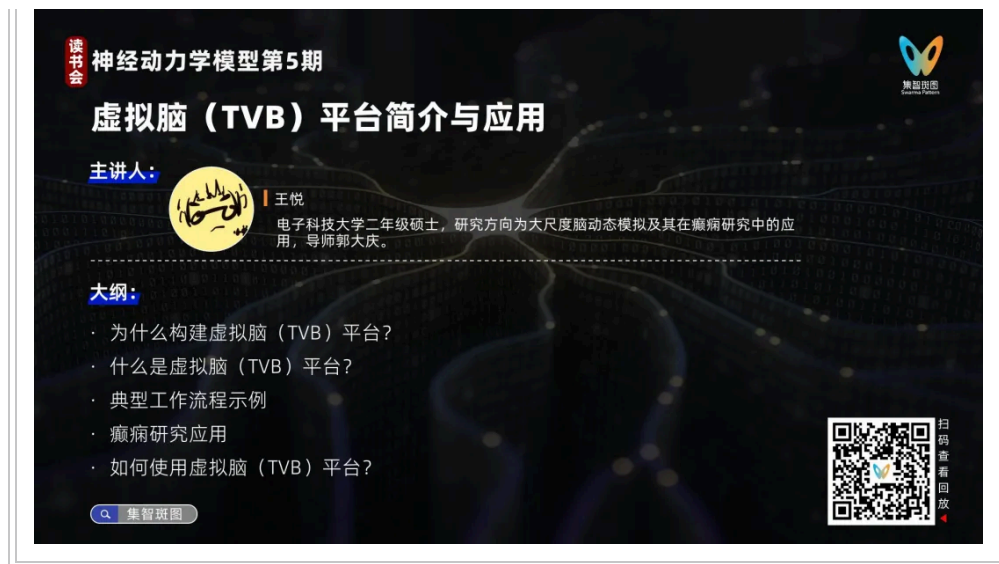


图5：实现社区准备度的关键考量

研发驱动的倍增包括开发行业特定的概念验证，向传统神经形态计算生态系统之外的研发团队提供早期且便捷的技术访问权限，从而形成技术传播的链式效应。概念验证的影响力更大，当所选的特定领域能够产生外溢效应，将最终促进主流应用场景的采用。此外，通过倾听早期采用者的反馈并激励他们积极参与，可以进一步放大神经形态计算生态系统的推动力。

电子科技大学郭大庆老师课题组也搭建了虚拟脑平台，扫码查看视频了解详情👉



一个潜在的神经形态计算硬件平台是 SpiNNaker2[99]，该芯片是人类大脑计划（Human Brain Project）的一部分，旨在支持神经科学探索。更重要的是，SpiNNaker2 与 Loihi 2 系统具备灵活的数值加速器（Flexible Numerical Accelerators），能够在多个层级上模拟神经行为，包括突触（Synaptic Computation）、树突计算（Dendritic Computation）[23,100]等精细建模，脉冲神经元（Spiking Point Neurons）、速率神经元（Rate Neurons）[23,99]等神经元层级建模，以及均场模型（Mean-Field Models）[29]、脑电信号（Electroencephalogram, EEG）行为模拟[23,101]等中尺度（Mesoscopic）建模。这些系统还可用于支持多尺度（Multiscale）脑模型，例如虚拟脑[97]，并且具备接近人类大脑规模的实时模拟能力。此外，越来越多的神经形态计算芯片正在进入市场，可以支持神经科学实验[23,39,102]。然而，要实现大规模、精细化的神经科学探索仍需要进一步的技术发展和优化。

机器学习启发 ML innovations

机器学习的目标是开发能够自主学习的算法，而不是依赖显式编程（Explicit Programming）来执行特定任务。近年来的算法进展不仅提升了神经网络的规模，还深入探索了连接模式（Connectivity）、信息表示（Information Representation）、学习机制（Learning Mechanisms）等方面。

除了算法优化，深度学习算法的计算需求也推动了计算架构的创新。然而，这些进步并非完全独立，其中一个显著现象是“硬件彩票（Hardware Lottery）”[103]，即某些计算方法的成功往往并非由于其本身的优越性，而是因为它们恰好匹配当前的硬件架构。这一现象在一定程度上阻碍了神经启发式（Neuro-Inspired）机器学习的创新，因为缺乏相应的硬件支持导致某些创新的算法无法高效运行。因此，我们认为大规模高效的神经形态计算系统能够推动神经启发式机器学习（Neuro-Inspired ML）[104,105]。图3中所识别的关键神经形态特性，可作为未来机器学习创新的蓝图。

最重要的是，这些技术进步的融合标志着计算范式的转变：传统冯·诺依曼（von Neumann）计算架构通常将算法和硬件作为相互独立的部分，而神经形态计算则强调算法与架构的互联则更类

似于大脑的设计方式，使得神经形态硬件的发展可能会在神经网络机器学习领域带来前所未有的变革。大脑的计算模式天然体现了学习算法的本质——神经元参数会随学习过程而自适应地调整功能，而传统硬件（如数据中心的GPU）采用的是固定的训练模式，与大脑的学习机制存在显著区别。相比之下，神经形态计算结合新兴器件技术能够提供一种克服当前机器学习计算与存储分离关键瓶颈的独特解决方案，从而推动人工智能向更高效、更接近生物学习模式的方向发展。

新兴器件与计算架构 Emerging devices and architectures

高效的存储技术对神经形态计算系统的可扩展性至关重要。从信息存储（Storage）、巩固（Consolidation）到检索（Retrieval），存储器构成了学习和基于经验解决问题的基础。新兴忆阻器（Memristive Devices）在神经形态计算领域尤其展现出巨大潜力，其非易失性（Non-Volatility）、高存储密度（High Density）和多状态存储特性（Multistate Properties）（详见框3），使其成为构建高效神经形态计算系统的关键技术之一。此外，忆阻器阵列（Memristor Arrays）能够直接执行乘-累加（Multiply and Accumulate, MAC）运算，而这正是神经网络的核心计算操作，其基于欧姆定律（Ohm's Law）实现紧凑且低功耗的计算加速[106,107]。要使新兴存储技术具备实际可行性，理想情况下应满足以下特性：低能耗： $\approx \text{fJ/bit}$ （飞焦耳每比特），低延迟，低工作电压（Low Operating Voltage）： $< 1\text{V}$ ，高耐久度： $> 10^{17}$ 次循环，高数据保持能力（High Data Retention），高可扩展性： $< 10\text{nm}$ ，兼容CMOS工艺[108]等。

英国伯明翰大学刘健教授课题组和中科院脑智中心、脑认知与类脑智能重点实验室张铁林研究员都在此领域做了诸多前沿探索，扫码查看视频了解详情👉

读书会

神经动力学模型第9期

集智班图

脉冲神经网络及其应用

主讲人:



刘健
伯明翰大学神经计算实验室副教授。实验室研究领域为计算神经科学，脑启发计算，类脑智能，神经形态计算。

大纲:

- 神经元细胞脉冲发放的特性
- 神经环路的脉冲信号数据特性
- 脉冲神经网络与人工神经网络的关系
- 脉冲神经网络的应用
- 脉冲神经网络的硬件实现

扫码查看回放



面向脑机接口的脉冲神经网络软硬件

主讲人:



张铁林

中科院脑智中心，脑认知与类脑智能重点实验室研究员，研究类脑脉冲神经网络、脑动力学模拟仿真、脑机接口类脑芯片。

大纲:

- 类脑脉冲神经网络(SNN)算法
- 物理知识嵌入神经网络算法
- 面向脑机接口应用的SNN算法
- 类脑SNN芯片及系统
- 总结

 集智班图扫码
查看
回放

框3：新兴存储技术及其挑战

近年来，许多新兴存储器件已被提出并应用于神经形态计算系统，包括阻变存储器（RRAM, Resistive Random-Access Memory）[107]、自旋电子器件（Spintronic Devices）[154]、铁电晶体管（Ferroelectric Transistors）[155]、相变存储器（PCM, Phase-Change Memory）[106]。补充表[1]对部分新兴存储器件与传统存储技术在多个相关指标上的性能进行了对比。尽管这些新兴器件在大规模系统中的潜力正在逐步被挖掘[156]，但仍面临多项挑战，主要包括器件的非理想特性（Non-Idealities）、与CMOS工艺的集成难度、漏电问题（Leakage）等。部分新兴存储器件已尝试通过某种形式的在线训练（Online Training）[121,146,157]来补偿器件非理想性，这一策略为在硬件内部集成补偿机制提供了潜在方案，但需要设计全新的存储器件和计算电路以适应该技术需求。此外，某些存储器件的物理非理想性也被转换为计算优势，例如循环间（Cycle-to-Cycle）和器件间（Device-to-Device）参数变化可用于贝叶斯计算（Bayesian Computation）[146,158]和自组织（Self-Organized）动态网络中的参数学习[145]。忆阻器的内部动态特性可被用于运动检测（Motion Detection）[159]和在线学习（Online Learning）[160]。目前这些方法在大规模模型和实际应用中的可扩展性仍然是一个活跃的研究方向。随着阻变存储器（RRAM）等新兴技术的持续成熟，其非易失性和存内计算能力使其在对尺寸、重量和功耗受限的应用场景中变得越来越具有吸引力。

总结

神经形态计算系统在多个领域具有巨大的潜在影响，当前正处于大规模创新的关键时刻。该领域已从学术机构开发的原型系统，发展到具备事件驱动处理、学习模型和设计工具，并与现实世界实验相结合的生产级系统。在此基础上，工业界已推动更先进的神经形态系统，并扩展其应用至多个领域，包括科学计算（Scientific Computing）、增强/虚拟现实（AR/VR,

Augmented/Virtual Reality)、可穿戴设备 (Wearables)、智慧农业 (Smart Farming)、智慧城市 (Smart Cities) 等。要继续推动该领域的发展, 需要大批工程师和科学家以及公共和私营机构在共享目标下开展合作。本研究提出了一份发展路线图, 总结了该领域的关键开放问题以激发未来的研究方向。

首先, 我们应全面审视神经形态计算系统所能支持的能力, 包括识别通用计算单元 (Common Primitives), 并以异构方式进行连接来整合不同组件, 开发多用途架构。人类大脑由多个不同但高度专业化的区域组成, 并以分布式方式连接, 但目前这些脑区模块与神经形态计算领域使用的智能组件 (在更精细的层面上) 并没有一一对应。在本文中, 我们提出了一组指导性特征用于识别计算单元, 以构建更强大、适应性更强的神经形态计算系统。这些系统应当支持广泛的学习机制, 例如在线持续学习、基于事件驱动传感器数据的实时决策、传感-运动融合算法 (Sensorimotor Fusion Algorithms)、多模态学习 (Multimodal Learning)、预测建模 (Predictive Modelling) 等。目前我们不应局限于单一硬件平台或部分特性, 而是应探索多样化的能力, 并明确不同应用场景下哪些神经形态特性至关重要。神经形态计算系统不应采用“一刀切”的解决方案, 相反, 不同应用场景可能需要具有不同特性的神经形态硬件解决方案。当前的神经形态计算系统主要基于单芯片或单板, 在大规模系统中可能并不适用。从这一问题出发, 我们需要研究两个方向:

1. 构建分布式、异步神经网络的鲁棒性模型和算法 (Robust Models and Algorithms for Distributed and Asynchronous Systems), 这需要更大规模的协作努力, 并非简单挑战。
2. 扩展学习模型至更大规模的网络和计算机 (Scaling Up Learning Models to Larger Networks and Machines), 可以借鉴AI社区的做法, 将大规模模型部署在云端 (Cloud) 或异构计算系统 (Heterogeneous Systems) 上。如果这一目标能够实现, 神经形态计算系统的采用率也将自然提升。

当前, 神经形态计算的进展往往受限于实验室中的小规模原型系统和研究资源的封闭性, 导致外部专家和非专家难以访问这些测试平台。这使得神经形态计算在可移植性 (Portability)、标准化和跨学科知识流通方面遇到挑战。为了解决这些问题, 我们可以推动开放测试平台和基准测试库的建设, 来开发新一代的大规模、自适应神经形态计算系统。当原型测试平台能够与传统计算系统集成时, 将促进研发的快速增长并加速面向应用的研究。此外, 这些开放资源需要具备可测量的性能指标或开放基准测试[94]。我们还可以借鉴深度学习领域成功的学术-工业合作, 进一步加速神经形态计算的研发周期。与此同时, 这也能为深度学习社区带来新的思路, 探索更类似人类学习方式的计算方法。

近年来, 用于快速原型开发或评估的软件和硬件设计工具呈增长趋势, 然而目前仍然缺乏一个完整的软件生态系统和社区论坛, 来同时支持专家和非专家的需求。虽然部分行业合作伙伴已经开始填补这一空白, 但大部分工具的开发仍然是孤立的, 缺乏系统性的整合与标准化 (见图4)。未来, 我们需要推动工具链的发展, 使其能够与主流深度学习工具兼容, 或直接受益于深度学习工具的成熟生态系统。此外, 我们还应关注大规模事件驱动系统的关键应用, 特别是在非认知计

算 (Non-Cognitive Applications)、边缘端分布式智能以及神经形态系统与深度学习系统集成
的特殊场景。

最后，我们提出未来可研究的开放问题，这些问题可分为短期内可研究的问题（框4）、中期探
索的问题（框5）以及长期待解的问题（框6）。当前正是投身于神经形态计算大规模发展的理想
时机，这一领域的发展将为未来的人工智能系统和自然智能系统带来突破性进展。

框4：短期研究问题

硬件/软件协同设计框架：我们是否能够设计用户友好的硬件/软件协同设计框
架，使非专家也能开发神经形态计算模型？此外，这种框架是否可以在大规模系
统中高效应用？

原型系统：面向大规模系统的原型测试平台应具备哪些特性？如何向更广泛的研究社区提供对这些原型的访问权限？

系统集成：如何提升神经形态计算系统与传统计算系统（如加速器）的集成便捷性？

软件工具：如何设计开放源码的软件工具，使其能够与主流深度学习框架互操作？是否可以开发更高抽象层次的软件工具，使其基于功能模块库，并具备通用中间表示层？

框5：中期研究问题 (Medium-Term Questions)

大规模测试平台：如何在通用测试平台上部署异构的、大规模神经形态计算系统演示？

基准测试：神经形态计算的通用基准测试套件应包括哪些内容？如何对这些基准测试进行有效评估？神经形态计算的开源协议应具备哪些特性？

终生学习：事件驱动的局部学习和可塑性如何支持具备终生学习能力的智能机器？是否存在适用于这些系统的统一架构？

动态建模：如何在未来的神经形态计算系统中引入复杂动力学？我们需要探索具备更复杂动力学的神经形态器件，如突触学习、树突处理 (Dendritic Processing)、更精细的神经连接建模。

框6：长期研究问题 (Long-Term Questions)

关键应用：如何将神经形态计算应用于任务关键型（Mission-Critical）或安全关键型（Safety-Critical）的场景？

参考文献

- [1] Abadi, M. et al. TensorFlow: large-scale machine learning on heterogeneous systems. <https://www.tensorflow.org/> (2015).
 - [2] Abreu, S. et al. Neuromorphic intermediate representation. Zenodo <https://doi.org/10.5281/zenodo.8105042> (2023).
 - [3] Aimone, J. B., Severa, W. & Vineyard, C. M. in Proc. International Conference on Neuromorphic Systems, 1–8 (ACM, 2019).
 - [4] Amir, A. et al. in Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition 7243–7252 (IEEE, 2017).
 - [5] Arthur, J. V. & Boahen, K. Learning in silicon: timing is everything. Adv. Neural Inf. Process. Syst. 18 (2005).
 - [6] Baby, S. A., Vinod, B., Chinni, C. & Mitra, K. in 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR) 316–321 (IEEE, 2017).
 - [7] Barnell, M., Raymond, C., Brown, D., Wilson, M. & Cote, E. in Proc. 2019 IEEE High Performance Extreme Computing Conference (HPEC) 1–5 (IEEE, 2019).
 - [8] Bergstra, J. et al. in Proc. 9th Python in Science Conference (eds van der Walt, S. & Millman, J.) 18–24 (2010).
 - [9] Cardwell, S. G. Achieving extreme heterogeneity: codesign using neuromorphic processors. Technical Report, Sandia National Laboratories (2021). A discussion on the need
- (参考文献可上下滑动查看)

计算神经科学读书会

人类大脑是一个由数以百亿计的神经元相互连接所构成的复杂系统，被认为是「已知宇宙中最复杂的物体」。本着促进来自神经科学、系统科学、信息科学、物理学、数学以及计算机科学等不同领域，对脑科学、类脑智能与计算、人工智能感兴趣的学术工作者的交流与合作，集智俱乐部联合国内外多所知名高校的专家学者发起神经、认知、智能系列读书会第三季——「计算神经科学」读书会，涵盖复杂神经动力学、神经元建模与计算、跨尺度神经动力学、计算神经科学与AI的融合四大模块，并希望探讨计算神经科学对类脑智能和人工智能的启发。读书会已完结，现在报名可加入社群并解锁回放视频权限。



聚焦神经编解码机制及其鲁棒性与非线性，认知功能组织特征及其适应性与稳定性

计算神经科学读书会

发起人：周昌松 弭元元 唐乾元 臧蕴亮 冯志聪 杜凯 李松挺 胡禹 张铁林 刘健 刘泉影 龚璞林（邀请报告）

- 复杂神经动力学：全景与基础
- 神经元建模与计算：鲁棒性与非线性
- 跨尺度神经动力学：适应性与稳定性
- 计算神经科学和AI的融合
- 圆桌讨论：多学科视角下的类脑仿真和类脑智能，我们在哪里？



详情请见：[计算神经科学读书会启动：从复杂神经动力学到类脑人工智能](#)

推荐阅读

1. [第三代神经网络模型：面向AI应用的脉冲神经网络](#)
2. [Sci. Adv. 速递：面向神经形态计算的生物合理可重构脉冲神经元](#)
3. [神经形态计算：生物启发式的人工神经网络设计 | 集智百科](#)
4. [游戏化科研——让我们突破内卷、共研涌现](#)
5. [探索者计划 | 集智俱乐部2025内容团队招募（全职&兼职）](#)
6. [加入集智，玩转复杂，共创斑图！集智俱乐部线下志愿者招募](#)

点击“阅读原文”，报名读书会

复杂科学前沿2025 · 目录

上一篇

百年谜题新解：胶体科学如何重塑熵的认知边界——从吉布斯佯谬到生命秩序的熵之启示

下一篇

Nature意识之争：两大理论首次正面对决

[阅读原文](#)

