面向边缘智能的类脑人工智能：系统综述

Yingchao Cheng、Meijia Wang、Zhifeng Hao，*IEEE*高级会员、Rajkumar Buyya，*IEEE*会士

arXiv:2603.26722v1 [cs.NE] 19 Mar 2026

摘要——**尽管脉冲神经网络有望突破边缘智能在尺寸、重量和功耗（SWaP）方面的严苛限制，但该领域目前面临着一种“部署悖论”：将异步、事件驱动动态映射到传统冯·诺依曼基底时产生的低效率，往往抵消了理论上的能耗优势。本综述超越仅关注算法的片面视角，采用严谨的**系统级软硬件协同设计视角，**考察2020–2025年的发展轨迹，重点关注从量化方法到混合架构等“最后一公里”技术，探讨如何将生物合理性转化为硅上的现实。我们深入剖析训练复杂度（直接学习与转换的两难）、阻碍具有状态的神经元更新的“内存墙”瓶颈，以及神经形态编译工具链中关键的软件鸿沟之间的相互作用。最后，我们展望一条弥合根本性“同步-异步失配”的路线图，并提出开发标准化的**神经形态操作系统 **作为基础层，以实现无处不在、能源自治的**绿色认知基底。

关键词—**脉冲神经网络（脉冲神经网络）、硬件与软件协同设计、部署悖论、边缘智能、神经形态操作系统、事件驱动处理、绿色认知基底、工具链。**

I. 引言

物联网（IoT）设备呈指数级激增，引发了一场剧烈的范式转变，人工智能正从超大规模云基础设施迁移至资源受限的网络边缘 [1]–[3]。这一迁移并非仅仅出于偏好，而是迫于现实需要；我们正目睹全球范围内的“数据洪流”，感官数据生成的速度和规模已超过现有传输带宽。预计全球数据总量将达到泽字节级，传统的以云为中心的模式正面临严峻瓶颈：**带宽饱和** 以及不可预测的传输延迟。依靠集中式服务器处理未经处理的高保真数据流（例如4K视频监控或自动驾驶激光雷达）已不再可持续，因此亟需转向“边缘智能”，在这种模式下，数据会在本地进行处理 [4]–[7]。

然而，这一转变面临着一个根本性的物理困境：在电池供电的边缘设备上部署传统的深度神经网络（DNN）日益受到**冯·诺依曼瓶颈**的限制。DNN 对连续、稠密矩阵乘法的固有需求，使得存储器与处理单元之间必须进行大规模且频繁的数据传输，而二者在物理上彼此分离。这造成了“内存墙”，导致延迟和功耗指标往往高到实时、任务关键型应用难以承受 [8]–[10]。

为应对这些硬件限制，**脉冲神经网络（SNN）**已超越其作为生物学奇闻的起源，成为下一代边缘智能的实用计算必然选择。不同于传统DNN基于帧的同步处理方式，SNN 采用异步、事件驱动范式运行，模拟了生物神经基质的效率。通过将信息编码为离散的二进制脉冲，SNN 利用极致的**时空稀疏性**，将计算与时钟周期解耦，从而有效实现“计算简约性”。如图1所示，这一机制从根本上将主要工作负载由高功耗的乘加（MAC）操作转变为能效显著更高的累加（AC）操作 [11]–[14]，。因此，*SNN* 不仅代表着一种算法优化，也体现了人工智能模型与现实世界感官数据的稀疏、事件驱动特性之间的结构性对齐。

然而，弥合理论算法效率与实际硬件部署之间的鸿沟仍是一项关键挑战。目前，该领域受到**“部署悖论”**的制约：脉冲神经网络算法需要事件驱动型基底，而主流商用现成边缘硬件——例如 NVIDIA Jetson 系列或边缘张量处理单元——在架构上仍针对稠密、同步的矩阵运算进行了优化。这种低效率源于**单指令多数据（SIMD）**并行性上的根本不匹配。图形处理器通过密集线程的锁步执行和合并内存访问来实现峰值吞吐量；然而，脉冲神经网络的随机、不规则放电模式会造成严重的控制流分歧（破坏 SIMD 锁步执行），并产生非连续的内存请求。这种**内存访问粒度**不匹配意味着，为稀疏脉冲获取单个权重时，往往会触发整个缓存行的传输，导致有效带宽利用率低下。因此，将稀疏脉冲神经网络映射到这些通用加速器上，必须依赖低效的模拟层，而这些模拟层往往会抵消脉冲模型固有的能耗优势 [15]–[17]。

本研究部分得到了中国国家重点研发计划（2025YFC3410000）和广东省基础与应用基础研究基金（2023B1515120020）的资助。（通讯作者：Yingchao Cheng。）Y. Cheng 任职于广东省人工智能与数字经济实验室（深圳），中国深圳，518107。（电子邮箱：chengy-ingchao.dr@gmail.com）。M. Wang 任职于澳大利亚墨尔本大学计算与信息系统学院云计算与分布式系统（CLOUDS）实验室，澳大利亚维多利亚州墨尔本，3010；同时任职于广东省人工智能与数字经济实验室（深圳），中国深圳，518107。（电子邮箱：meijia.wang@student.unimelb.edu.au）。Z. Hao 任职于汕头大学数学与计算机科学学院，中国汕头，515063。（电子邮箱：haozhifeng@stu.edu.cn）R. Buyya 任职于澳大利亚墨尔本大学计算与信息系统学院云计算与分布式系统（CLOUDS）实验室，澳大利亚维多利亚州墨尔本，3010（电子邮箱：rbuyya@unimelb.edu.au）。

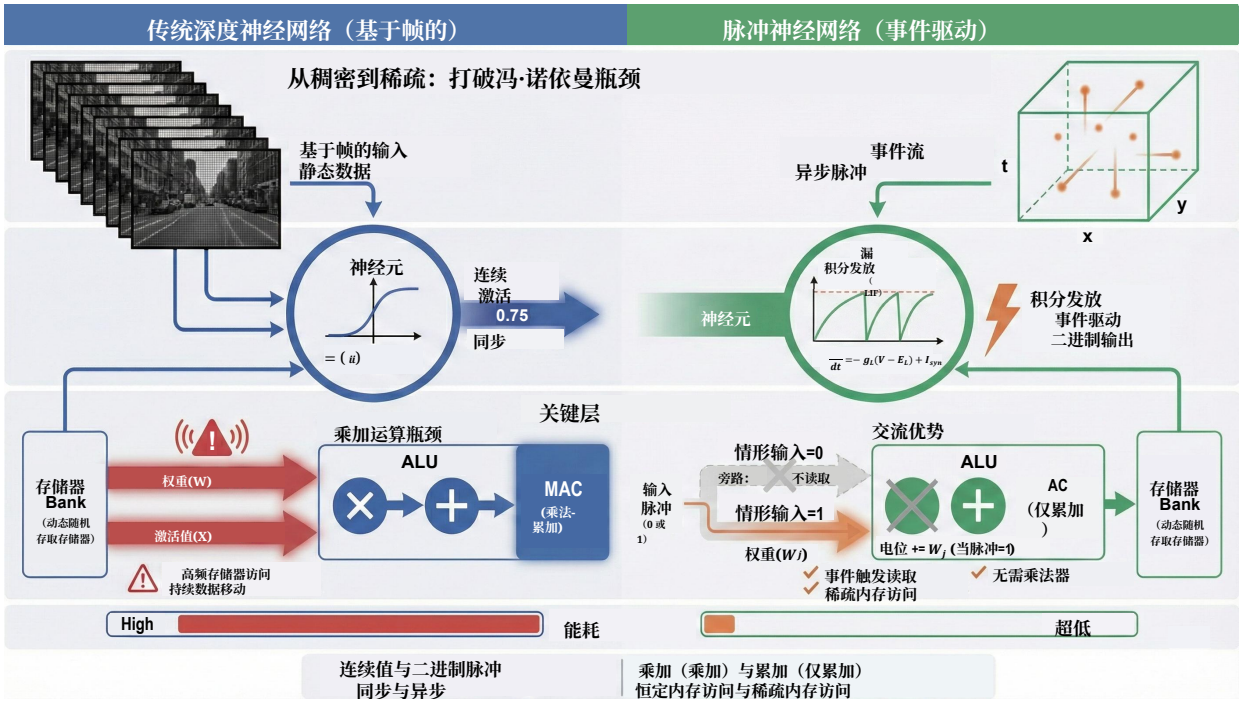

图1. **从传统深度神经网络到脉冲神经网络的范式转变概述。（上）数据表示：**深度神经网络处理冗余的基于帧的图像序列，而脉冲神经网络处理以时空体表示的稀疏异步事件流（x，y，t）。**（中）神经元动力学：**与深度神经网络中的静态激活函数（例如 Sigmoid函数）不同，脉冲神经网络采用符合生物学原理的漏积分发放（LIF）神经元，通过微分方程控制的信息整合来处理时间信息。**（下）能效：**从稠密乘加运算转向事件驱动的累加逻辑，可实现超低能耗，使脉冲神经网络非常适用于边缘智能。

相反，能够释放脉冲神经网络全部潜力的原生神经形态处理器（如 Intel Loihi 2 或 BrainChip Akida）虽然采用了所需的异步范式，却大多局限于研究原型或利基垂直市场。这种两极分化造成了显著的**部署差距**，先进算法缺乏适用且普及的硬件载体 [18]–[21]。归根结底，边缘脉冲神经网络目前难以广泛普及，原因并非算法能力不足，而是硬件生态系统仍受制于针对上一代稠密深度学习模型优化的惯性。

尽管此前已有综述试图描绘这一领域，但现有文献大多将其割裂开来，把脉冲神经网络算法与硬件实现视为彼此独立的领域。一类综述着重探讨算法创新——例如替代梯度学习或可塑性规则——却将部署平台的物理约束抽象化处理 [22]–[23]。相反，以硬件为中心的综述往往深入探讨新兴器件（如忆阻器）或电路设计，却未充分讨论在这些硬件上运行所需算法的可扩展性 [24]–[26]。这种割裂掩盖了部署差距中的实际复杂问题，使人们无法充分理解算法稀疏性如何与硬件原语相互作用。为弥合这一鸿沟，本文采用了整体性的**软硬件协同设计视角**。这一方法认为，边缘智能的效能不仅取决于突触精度，还取决于算法稀疏性与硬件架构能否无缝协同。

因此，本文阐述了三项主要贡献：

- **训练范式的批判性综合（2020–2025）：**我们对直接训练机制（例如替代梯度）的最新突破与人工神经网络到脉冲神经网络的转换技术进行了比较分析。我们并非孤立地评估这些方法，而是从边缘适用性的角度展开讨论，权衡训练延迟与推理能效 [27]–[31]。

- **以部署为中心的系统分析：**有别于泛泛的算法综述，本文探讨物理实现所面临的“最后一公里”挑战。我们仔细审视了常被忽视的瓶颈，包括训练后量化、异构映射工具链，以及制约实际性能的芯片间通信开销 [32]–[35]。

- **互联智能前瞻路线图：**本文的范围不局限于标准计算机视觉基准，而是进一步拓展至新兴前沿领域。我们阐述了脉冲神经网络在分布式边缘学习框架中的作用，以及其与下一代6G网络的潜在集成，并指出了迈向普及型超低功耗认知无线电系统的发展方向 [36]–[38]。

本综述的结构组织以**软硬件协同设计**原则为基础，具有鲜明特色。不同于现有文献中将算法稀疏性与神经形态基底孤立分析的做法，本研究提出，要打通边缘部署的“最后一公里”，就必须建立一个统一的

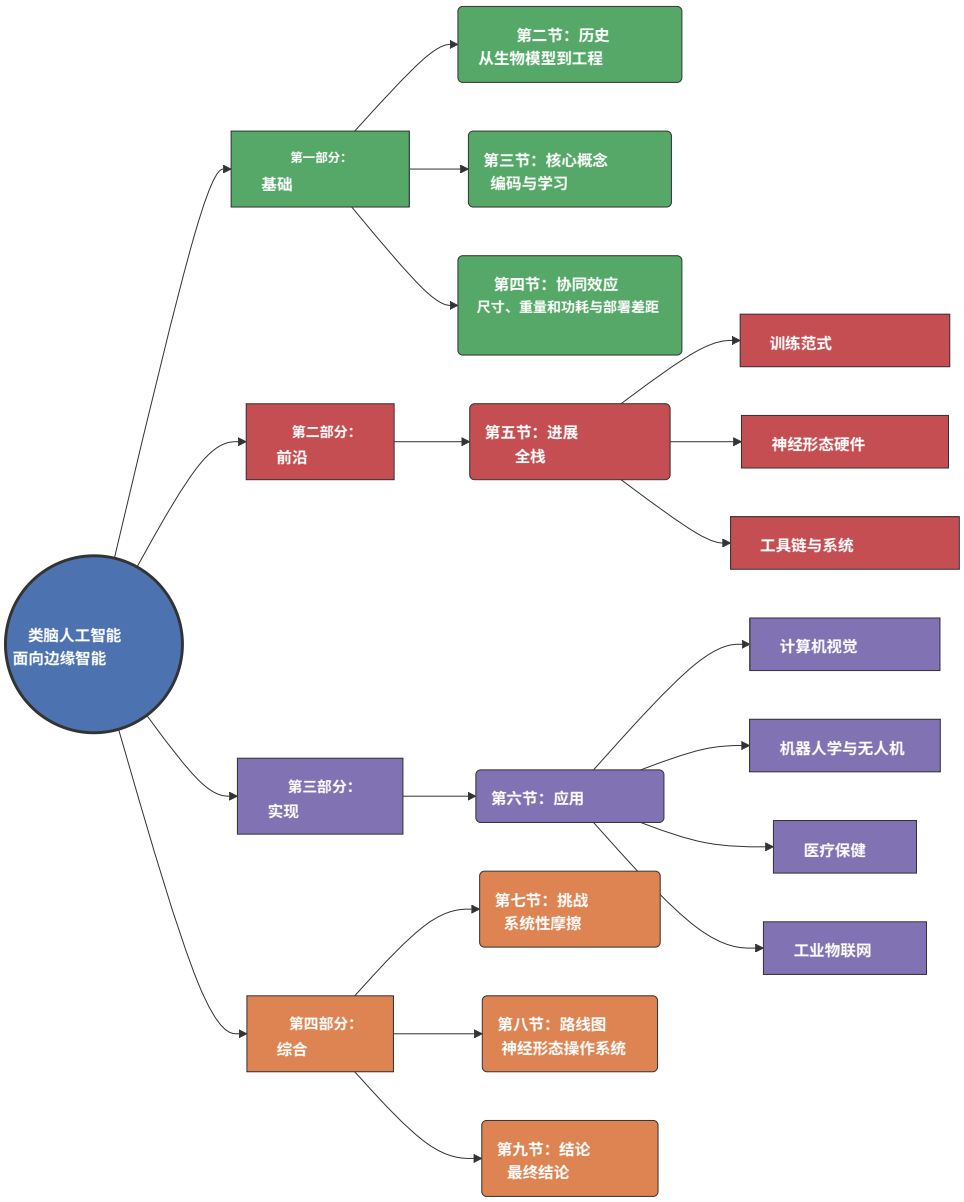

图2。**综述的层级组织**。本文分为四大部分：**第一部分**奠定理论基础并定义“部署悖论”；**第二部分**探讨算法、硬件和工具链的技术前沿（2020–2025）；**第三部分**考察边缘应用中的现实世界实现；**第四部分**综合各项挑战，并提出迈向统一神经形态操作系统的路线图。

分类体系。在这一框架下，算法参数（如稀疏度、神经元模型）被视为与硬件约束（如扇入限制、存储层次结构）密不可分，从而为普及边缘智能建立一条连贯的路线图 [39]–[41]。

为系统地梳理这些相互依存关系，本文余下部分安排如下（图2对此进行了直观概述）：

- **基础与研究动机（第二至第四节）：** 本文首先追溯脉冲神经网络从生物物理学根源发展到现代计算原语的演进历程（第二节）。随后，本文严谨地界定核心概念，包括编码方案，并将其与深度神经网络进行比较分析（第三节）。接着，第四节阐明了脉冲神经网络动力学与边缘环境中的能耗与时延约束之间的内在契合。
- **技术前沿（第五节）：** 本节构成核心技术综合（2020–2025），全面评估全栈的最先进水平：从新颖的训练范式和硬件感知优化，到弥合软硬件鸿沟所需的映射工具链。
- **实现与反思（第 VI–IX 节）：** 第 VI 节将理论讨论转向实践，探讨其在计算机视觉、机器人学和物联网中的部署。随后，第 VII 节以批判性视角审视可扩展性和训练方面长期存在的瓶颈。最后，第 VIII 节展望该领域迈向 6G 集成和伦理人工智能的发展方向，并在第 IX 节结束本综述。

归根结底，这一结构旨在引导读者从微观神经元动力学一路了解宏观系统级部署，并强调边缘人工智能的未来取决于对这两个尺度之间接口的优化。

II. 背景与基础：脉冲神经网络范式

脉冲神经网络（SNN）的理论发展轨迹由一个核心二分法所界定：追求生物同构性与满足工程可处理性的必要性之间的权衡。不同于第二代人工神经网络（ANN）的连续激活景观，SNN以离散、事件驱动的信息处理为前提。本节将阐述这一范式的演进，追溯其从严谨的生物物理建模到算法微分与硬件协同设计现代阶段的转变。

A. 建模动力学：从离子通道到抽象

将SNN归类为神经网络的第三代，依据在于它们能够通过脉冲的精确时序来编码信息，理论上比基于频率的同类网络具有更强的计算能力 [42]–[44]。然而，要实现这一潜力，就必须在建模粒度上做出根本性妥协。

神经元建模源于对离子通道电导所介导的动作电位生成过程的严谨描述，其典型代表是Hodgkin-Huxley形式 [45]–[47]。尽管这类模型在生物物理层面准确，但其高维微分方程对于网络级模拟而言，计算成本高得难以承受。这一计算瓶颈促使研究者转向现象学抽象。漏积分发放（LIF）模型由此成为高效模拟的标准，将动力学简化为线性膜积分过程，并通过硬阈值重置 [48]–[51]。

为弥合简单的 LIF 模型与复杂的 Hodgkin-Huxley 模型之间的差距，人们开发了混合方法。值得注意的是，通过降维相位系统中的分岔来捕捉多种放电模式（如爆发放电、颤动放电）的模型，在丰富动力学与简单映射的计算成本之间实现了关键的权衡[52]–[55]。

与这些算法抽象并行，硬件基底也经历了类似的理念转变。“神经形态”工程最初旨在利用亚阈值模拟VLSI的物理特性，直接模拟神经生物学离子流 [56]–[58]。随着时间推移，这一理念演变为数字架构和混合信号实现，优先考虑稀疏、事件驱动通信的能效，而非严格的模拟仿真。

神经元模型的发展体现出一种明显趋势：趋向功能极简主义。该领域基本已形成共识：对于神经形态计算而言，捕捉脉冲机制的“本质非线性”，比复现生物胞体中精确的离子电流更为重要。

B. 复兴：克服可微性障碍

尽管理论基础雄厚，脉冲神经网络在实际部署中历来受阻于脉冲生成函数不可微这一特性，导致标准反向传播无法奏效。

脉冲神经网络学习的早期阶段由基于生物学原理的局部学习规则主导，其中最著名的是脉冲时序依赖可塑性（STDP）和赫布变体[59]–[61]。这些无监督机制在浅层架构中的局部特征提取和模式识别方面表现出色，但难以扩展到现代卷积神经网络（CNN）所需的监督学习复杂度，也难以构建与之相当的深层层次表示 [62]–[65]。

替代梯度（SG）方法的正式确立，推动了脉冲神经网络的当代复兴，也促使其随后在边缘计算领域广泛应用。在反向传播过程中，以平滑的辅助函数近似不可微的狄拉克δ函数，使SG范式能够将梯度下降直接应用于时间动态，从而有效地将脉冲神经网络的能效与深度学习的训练效能联系起来 [66]–[70]。

这一算法突破恰逢工业级神经形态硬件日趋成熟 [71]–[74]。研究验证了事件驱动架构能够实现比冯·诺依曼系统低几个数量级的功耗，这为在资源受限的环境中部署这些深层脉冲架构提供了现实依据。

脉冲神经网络的复兴并非只是回归生物学启发，而是对现代计算中“内存墙”和功耗限制的务实回应。从纯粹的脉冲时序依赖可塑性转向基于梯度的学习，体现了一种折中：接受非生物学的训练方法，以在推理中实现生物水平的能效。

III. 脉冲神经网络的核心概念：工程视角

脉冲神经网络代表着一种范式转变：从连续值激活转向离散的事件驱动计算。不同于依赖**基于帧的、同步**处理的传统人工神经网络，脉冲神经网络利用脉冲序列的时间动态来编码信息。这一根本差异使异步处理成为可能，并显著降低能耗，尤其是在功耗预算受到严格限制的边缘场景中。

A. 脉冲神经元模型：保真度与成本

在神经形态工程中选择神经元模型，是在生物保真度与计算效率之间进行的战略性权衡。模型涵盖了从对离子通道的精确生物物理描述，到为大规模硬件实现而优化的简化抽象 [24], [39],[44], [75]。

1) 泄漏积分发放（*LIF*）：仿真标准：由于具有线性的阈下动态，漏积分发放模型是大规模脉冲神经网络仿真中广泛采用的抽象模型。它捕捉了基本的“积分”和“发放”

机制，而无需对复杂的通道动力学进行建模。从数学上讲，它可以描述为：

$$\tau_m \frac{dV(t)}{dt} = -(V(t) - V_{\text{rest}}) + R_m I_{\text{syn}}(t), \tag{1}$$

其中 $\tau_m = R_m C_m$ 是膜时间常数。尽管较为简单，漏积分发放模型为替代梯度学习提供了凸优化景观，因此成为深度脉冲神经网络 [76]的事实标准。

*2）Izhikevich*模型：计算桥梁：Izhikevich模型充当计算桥梁，能够捕捉通常与复杂生物物理模型相关的多种放电模式（例如爆发放电、颤动放电），同时保留耦合一阶微分方程的效率：

$$\begin{aligned} \frac{dv}{dt} &= 0.04v^2 + 5v + 140 - u + I, \\ \frac{du}{dt} &= a(bv - u). \end{aligned} \tag{2}$$

其中 v 是膜电位， u 是恢复变量， I 是输入电流。通过调整参数 a，b，c，d，该模型能够以与简单积分器相当的浮点运算次数（FLOPs）再现皮层动力学 [77]。

3）Hodgkin–Huxley（HH）：生物物理学的黄金标准：HH模型明确描述离子电导（Na^+、K^+）。尽管它仍是理解神经生理学的黄金标准，但其计算成本高昂（需要求解四个非线性微分方程），因此不适用于边缘智能应用，主要用作验证基准 [65], [78]。

脉冲神经网络设计的工程目标并非最大限度地实现生物真实性，而是最大限度地提高信息处理能力与能耗之比。因此，*LIF* 等简单模型主导着硬件实现，因为它们能最大限度地减少每个神经元所需的硅面积，从而实现大规模并行。

B. 编码分类体系：频率与时间编码

脉冲神经网络中的信息编码决定了系统的延迟和带宽效率。这些方案通常分为两类：基于频率的编码和时间编码 [79]–[82]。

*1）*基于频率的编码：频率编码源自经典神经生理学，通过时间窗口内的平均发放频率来表示信息。尽管这种编码具有很强的抗噪声和随机性能力，但其固有缺点是延迟较高（需要时间进行平均），且由于存在冗余脉冲发放，能效较低。

*2）*时间编码：时间编码方案利用脉冲的精确时序来承载信息，从而实现更高的稀疏度和带宽效率。**首脉冲时间（TTFS）**根据首个脉冲的延迟反向编码信息，使决策能够以极快的速度完成，这通常称为“延迟编码” [39], [83]。其他方法包括**相位编码**，即根据周期性背景振荡确定脉冲的时序；以及**增量调制**，该方法广泛用于异步传感器（例如 DVS相机），仅对强度变化进行编码，自然滤除静态冗余 [84], [85]。

从频率编码转向时间编码，标志着从“模拟生物机制”到“利用稀疏性”的关键飞跃。时间编码能够最大化每比特（脉冲）的信息含量，这是脉冲神经网络实现超低功耗的理论基础。

C. 学习动态：从局部到全局

脉冲神经网络中的学习分为两类：符合生物学原理的局部规则，以及以性能为导向的全局优化。

1) 局部无监督学习（脉冲时序依赖可塑性）：脉冲时序依赖可塑性（STDP）根据因果延迟 $\Delta t = t_{\text{post}} - t_{\text{pre}}$ 调整突触权重。它无需标注数据即可实现自主特征提取：

$$\Delta w = \begin{cases} A_+ e^{-\Delta t/\tau_+} & \text{if } \Delta t > 0 \text{ (LTP)} \\ -A_- e^{\Delta t/\tau_-} & \text{if } \Delta t < 0 \text{ (LTD)} \end{cases} \tag{3}$$

诸如奖励调制型 R-STDP 和镜像型 mSTDP 等变体应运而生，将监督信号引入这一局部过程 [26], [86], [87]。

2) 全局监督学习（替代梯度）：为克服二进制脉冲不可微的问题，现代脉冲神经网络采用替代梯度（SG）方法。这种方法在反向传播过程中对赫维赛德阶跃函数进行平滑处理，使脉冲神经网络能够通过时间反向传播（BPTT）进行训练，达到与深度神经网络相同的效果 [24], [62], [88]。

一种融合趋势正在形成：利用“局部”可塑性（脉冲时序依赖可塑性）实现高效预训练或硬件适配，同时通过“全局”梯度（替代梯度）微调网络，以提升任务特定准确率。这种混合方法兼顾了生物系统的自主性与工程控制的精确性。

D. 架构比较：脉冲神经网络与深度神经网络

表 I 和图 3 总结了架构差异。脉冲神经网络的关键优势在于以累加（AC）操作取代乘加（MAC）操作。由于脉冲是二进制的（$S(t) \in \{0, 1\}$），突触加权便转化为条件加法： $V_j(t) = V_j(t-1) + \sum_i w_{ij} \cdot S_i(t)$。这无需使用高精度乘法器，从而减少硅片占用面积和动态功耗 [89], [90]。

IV. 边缘端脉冲神经网络协同：架构契合

在边缘部署人工智能并非单纯的软件优化挑战，而是与尺寸、重量和功耗（SWaP）的严格限制进行的一场物理较量。传统方法试图压缩连续值深度神经网络（DNN），使其适应这些限制；相比之下，脉冲神经网络（SNN）在设计理念上截然不同。SNN 并非数据中心模型的缩小版近似，而是天然契合物理世界中零星出现、由事件驱动的特性。本节将从三个关键维度剖析这种协同关系：尺寸、重量和功耗的物理约束、常开传感的需求，以及消除处理流水线中的同步开销。

表 I 脉冲神经网络与深度神经网络的比较

特征	SNNs	DNNs
神经元模型	受生物学启发（例如 LIF、伊日克维奇）；离散脉冲	抽象化；连续激活值（例如 ReLU，Sigmoid函数）
信息编码	脉冲的时间模式（频率、延迟、相位、群体）	激活幅度
计算	事件驱动、异步、稀疏；主要进行累加（累加）操作	同步、稠密；主要执行乘加（MAC）ops
能效	潜在能效极高，尤其是在神经形态硬件上	通常较低，可能功耗较高
时间处理	具备处理时间数据的内在能力	通常需要循环架构（循环神经网络、长短期记忆网络）
生物合理性	较高；模拟神经动力学和类似的学习规则 STDP	较低；学习（例如反向传播）通常不生物合理的
训练复杂度	由于不可微脉冲而具有挑战性；脉冲时序依赖可塑性，替代梯度	已相当成熟（反向传播），但可能需要大量数据
硬件适用性	神经形态处理器、现场可编程门阵列	图形处理器、张量处理器、中央处理器
数据稀疏性处理	能够自然地利用输入稀疏性和激活稀疏性	除非经过专门设计，否则处理稀疏数据的效率较低
噪声鲁棒性	由于时间动态和群体，可能更高编码	可能对噪声输入敏感
延迟	由于脉冲快速传播和稀疏事件，延迟可能极低	可能更高，尤其是对于深层架构
主要应用场景（边缘）	实时传感器处理、低功耗可穿戴设备，事件视觉	通用人工智能任务，通常拥有更多可用资源

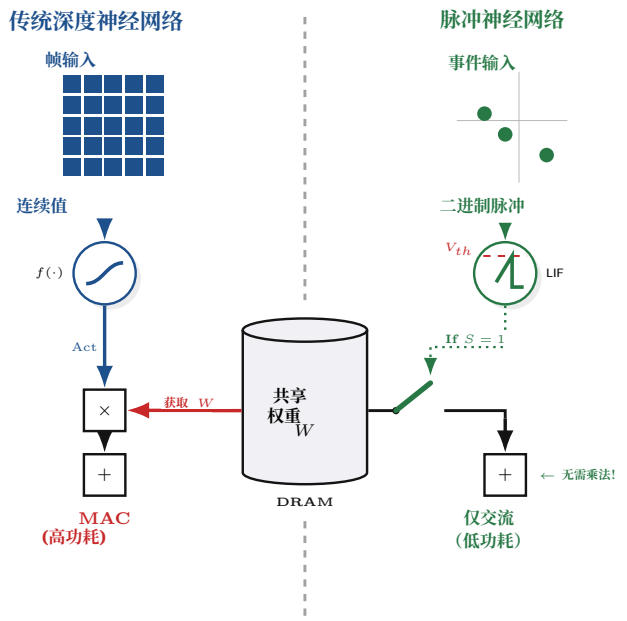

图 3. **数据流对比：连续与事件驱动。（左）深度神经网络数据流：**依赖稠密矩阵向量乘法（乘加），由于权重无条件读取，形成持续的“内存墙”。**（右）脉冲神经网络数据流：**采用条件触发机制，只有在脉冲到达时才会进行内存访问和计算。仅 在脉冲到达时进行的这种感知稀疏度的处理，显著降低了对片外存储带宽的需求。

A. 尺寸、重量和功耗的迫切要求：不止于能耗

边缘计算的约束涉及多个维度。评估脉冲神经网络的适用性，不能仅仅通过能耗视角，但更应通过尺寸（硅面积）、重量（电池约束）和功耗（热设计）这三者相互关联的维度来审视。如图 4 所示，这种多维比较表明，脉冲神经网络在受限环境中具有良好的架构契合度，能够以准确率的小幅下降换取效率指标的显著提升。

1）功耗与热约束：在传统的 CMOS 制程缩放过程中，业界面临“暗硅”现象：功率密度限制导致芯片上的所有晶体管无法同时运行，否则会引发热节流 [91]–[94]。依赖稠密矩阵乘法的标准深度神经网络加速器，存在超出热设计功耗（TDP）限制的风险。脉冲神经网络通过**激活稀疏性**缓解这一问题。只在显著事件发生时处理信息，使脉冲神经网络能够按需“点亮”暗硅，在不突破热限制的情况下维持较高的功能密度 [65], [77]。

2）重量与机动性：对于微型无人机或可穿戴健康监测设备等移动边缘智能体而言，首要约束往往是重量，尤其是维持持续运行所需的电池质量。算法效率与飞行时间或运行寿命直接相关。通过将推理能耗降低数个数量级（在神经形态实现中，降幅通常超过 100× [95], [96]），脉冲神经网络可显著减少电池载荷。这种储能需求的降低，直接提升了对重量敏感的自主系统的机动性，并延长其任务持续时间。

3）尺寸与硅片面积： 脉冲神经网络便于实现紧凑的物理结构。脉冲的二值特性消除——

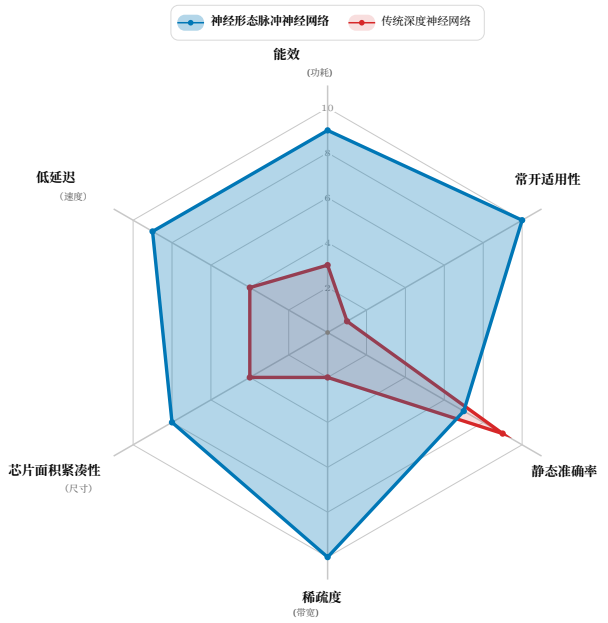

图 4. **雷达图展示了神经形态脉冲神经网络与传统边缘深度神经网络基线之间的多维权衡**。尽管深度神经网络在静态准确率方面略占优势，脉冲神经网络在受尺寸、重量和功耗约束的指标上表现更优，具体包括能效、稀疏度和常开适用性。这种架构契合使脉冲神经网络成为资源受限边缘部署的理想候选方案，尽管需要以准确率为代价。

除了对复杂浮点运算单元（FPUs）和大型乘法器的需求。相反，突触运算可简化为累加运算（AC），与深度神经网络中使用的传统乘加（MAC）单元相比，大幅减少每个神经元所需的逻辑门数量和硅面积。

B. “常开”感知范式

一种典型的边缘工作负载是“常开”监视器——这类系统必须持续分析环境，以检测特定触发信号（例如关键词检测、地震异常检测或安全监控），同时保持深度休眠状态。

1) 零输入、零能耗优势：在这类场景中，深度神经网络的静态功耗较高。即使处理的是静音片段或静态图像，传统卷积神经网络（CNN）也会针对每一帧执行固定数量的运算，以确认目标不存在。相比之下，脉冲神经网络具有**“零输入、零能耗”**特性。没有感官变化时，不会产生脉冲，动态功耗会降至接近零。

2）时间稀疏性：这使脉冲神经网络特别适用于唤醒控制器。在有效信号使输入神经元的膜电位累积足够电荷并越过发放阈值 [48]之前，网络始终处于静息状态，功耗可以忽略不计。这种事件驱动唤醒机制与传统人工智能的轮询式架构形成鲜明对比，确保只有在实际有信息输入时才会消耗能耗。

C. 传感器—算法—硬件流水线：端到端异步性

只有当算法成为贯穿传感器到硅基底的整体异步流水线的一部分时，脉冲神经网络的效率才能得到充分发挥。

1）无缝的感官整合：脉冲神经网络可与事件驱动传感器（如动态视觉传感器（DVS））自然对接 [97], [98]。基于帧的相机以固定间隔（例如 60 Hz）向处理器传送大量冗余像素数据，而 DVS 像素则以异步方式工作，仅在检测到强度变化时输出数据。由此形成“脉冲输入、脉冲输出”的数据路径，无需将稀疏事件转码为传统人工神经网络所需的稠密帧，避免了计算开销巨大的转码过程。

2）消除时钟开销：或许最深远的优势在于实现**端到端异步性**的潜力。在传统数字架构设计中，全局时钟分配网络最多可消耗芯片动态功耗的 30–40% [99], [100]。脉冲神经网络，尤其是在 Intel Loihi 2 等神经形态基底或异步 FPGA 架构上实现时 [76], [101], ，不需要全局同步，而是依靠局部握手运行。这样便消除了“冯·诺依曼瓶颈”：无需等待内存读取与时钟周期步调一致，使硬件能够模拟物理世界连续、异步的动态特性。

在边缘侧从深度神经网络向脉冲神经网络的转变，不仅意味着效率的提升；它还标志着从“高精度计算”（其目标是对输入进行精确的数值重构）向“可执行智能”的范式转变。通过舍弃冗余的时间数据并优先处理显著事件，脉冲神经网络使计算成本与所需行动的复杂度直接匹配，从而成为普及型自主边缘智能的关键推动因素。

V. 算法与优化的最新进展

面向边缘计算的脉冲神经网络研究已迅速从理论探索发展到系统级部署。该领域不再局限于简单的转换方法，而是经历了向硬件感知的直接训练、异构架构设计以及多维协同优化的范式转变。为系统地梳理这些多样化的进展，我们围绕图5所示的分类体系组织综述，将该领域划分为四大支柱：训练范式、优化策略、神经形态硬件和系统集成。

A. 训练范式：从梯度近似到快速转换

脉冲神经网络的训练格局已从生物仿真发展到高性能计算。我们将近期进展划分为两个不同的时代：基于梯度的时代，这一时代解决了脉冲的不可微性问题，实现直接学习；以及低延迟转换时代，这一时代重新定义了速度—精度权衡。

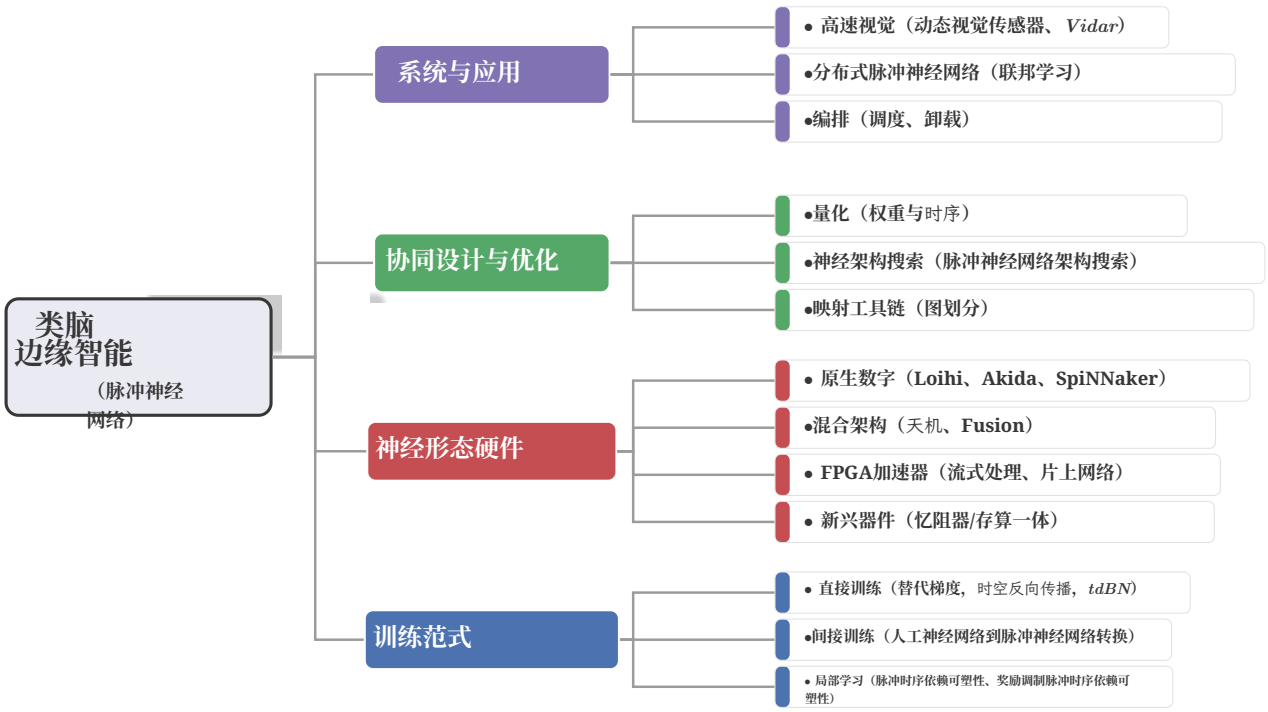

图5。**类脑边缘智能的分类体系**。本综述围绕四大支柱展开：（1）**训练范式**：着重介绍从转换到时空反向传播（**STBP** ）和 **tdBN** 等直接训练算法的转变；（2）**神经形态硬件**：涵盖原生数字芯片、**现场可编程门阵列** 加速器，以及天机等混合架构（**天机**）；（3）**协同设计策略**：聚焦于**神经架构搜索** 和时间量化；（4）**系统与应用**：重点关注高速视觉（**Vidar**）和边缘云编排。

1）基于梯度的时代：克服不可微性：训练深度脉冲神经网络的根本障碍在于，用于脉冲生成的赫维赛德阶跃函数不可微，导致反向传播时梯度无法流动。为解决这一问题，人们提出了**替代梯度（SG）**方法，在反向传播过程中使用连续的伪导数（例如 Sigmoid 函数或三角形函数）来近似脉冲函数的导数[23]，[62]，[67]，[69]。作为这一方向的开创性工作，**时空反向传播（STBP）**算法 [66]，[102]被提出，用于显式聚合空间和时间维度上的梯度。与常常忽略空间层间依赖关系的传统时间反向传播（BPTT）不同，STBP 能够直接在复杂数据集上训练高性能脉冲神经网络，并已成为现代脉冲神经网络优化的标准方法。

然而，将基于替代梯度的训练扩展到深层架构（例如 ResNet-50 ）带来了第二个挑战：脉冲传播的二值特性会导致梯度消失或梯度爆炸。为应对这一问题，**阈值依赖批归一化（tdBN）** [103] 应运而生。tdBN 沿时间维度对膜电位进行归一化，并采用与阈值相关的缩放因子，从而确保神经元维持适当的放电率，有效调节梯度幅值。为进一步优化学习动态，近期研究不再局限于固定的替代函数形状。Dspike 框架 [67]能够自适应地学习替代梯度的陡峭程度，动态平衡这一权衡即在准确的梯度估计与训练稳定性之间取得平衡。

针对替代梯度方法中随时间展开脉冲神经网络所带来的高存储器开销，研究正转向**事件驱动学习**。为弥合高性能反向传播与始终在线的边缘端约束之间的差距，研究人员开发了**脉冲时序依赖事件驱动（STD-ED）**算法 [90]。这些方法仅在脉冲事件发生时触发权重更新，而非在每个全局时间步都进行更新，从而将能耗降低约 30×。

2）低延迟转换时代：突破时间障碍：尽管直接训练在时间效率方面表现出色，但在大规模任务的绝对准确率上往往仍落后于人工神经网络。传统上，人工神经网络到脉冲神经网络的转换弥补了这一准确率差距，但存在严重局限：**极高的推理延迟**（数百个时间步），这是通过频率编码近似连续激活值所必需的。为缓解这一延迟瓶颈，近期研究的重点已从简单的放电率近似转向时间信息压缩 [104]–[106]。研究人员采用了校准前向时间偏差、引入脉冲簇等技术，以缩短推理窗口。为彻底消除转换过程的迭代特性，近期提出了一种**并行转换学习框架** [107] 。该方法通过建立并行神经元时间步与累积脉冲率之间的映射，仅用**4 个时间步**便在 ImageNet-1k（ResNet-34）上达到 72.90% 的 top-1 准确率，标志着基于转换的方法终于在低延迟优势上与直接训练相媲美。

不同范式正逐渐融合。“直接训练”与“转换”之间泾渭分明的界限正在消融，转而形成混合训练策略。近期的框架 *[108],[109]* 表明，脉冲神经网络已不再是孤立的模型，而是日益作为用于特征提取的“神经前端”部署，并与用于语义分类的人工神经网络后端相结合，或通过人工神经网络引导的知识蒸馏进行优化。未来的训练算法很可能会优先考虑这种异构协同，而非纯粹的独立性能。

B. 优化：硬件感知模型压缩

在资源受限的边缘设备上部署脉冲神经网络，需要在模型准确率、内存占用和能耗之间权衡。近期研究将这些目标统一纳入硬件感知压缩框架，着眼于三个关键维度：连接性（剪枝）、拓扑结构（神经架构搜索）和精度（量化）。

1) 剪枝：最大限度减少冗余连接：标准脉冲神经网络通常沿用稠密连接模式，对于稀疏的脉冲数据而言，这会造成计算冗余。为了解决这一问题，**自适应剪枝**策略应运而生，可根据神经活动动态调整剪枝率。例如，**动态回滚机制** [110] 已被用于在神经形态硬件上实现低于 μW 的功耗。然而，非结构化剪枝往往会导致不规则的内存访问模式，难以加速。因此，近期研究提出了**最优脉冲脑压缩（OSBC）** [111]，将最优脑压缩（OBC）理论应用于脉冲神经网络。OSBC 通过最小化膜电位上的损失，而非仅针对权重进行优化，在准确率损失可忽略不计的情况下实现了 97% 的稀疏度（例如，在 DVS128-Gesture 上准确率仅下降 $< 2\%$），证明脉冲神经网络的冗余度显著高于人工神经网络。

2）神经架构搜索（*NAS*）：自动化拓扑结构设计：手动启发式设计脉冲神经网络（例如简单地复制 VGG 或 ResNet）通常无法捕捉脉冲固有的时间动态。为克服这种人为偏差，**神经架构搜索（NAS）**被引入，用于自动发现脉冲友好型拓扑结构 [39], [112]。早期方法如 SNAS [113]将离散搜索空间松弛为连续域。然而，脉冲神经网络架构搜索（SNN-NAS）面临一个独特挑战，即拓扑结构搜索与训练所使用的替代梯度之间存在不匹配。为解决这一问题，**SpikeDHS** [114]提出了联合优化框架，同时搜索网络拓扑结构和替代梯度的最优形状。此外，随着部署目标日益多样化，该领域逐渐发展出**硬件感知神经架构搜索**。近期的神经进化方法[115]将硬件指标（如延迟和能耗）明确纳入搜索目标，确保所发现的架构位于目标设备的帕累托最优前沿上。

3）量化：压缩精度与时间：传统量化会降低权重位宽，而优化脉冲神经网络则提供了一个独特的额外维度：**时间量化**（减少仿真时间步 T）。减少T 可带来线性的延迟改善，但通常会降低准确率。为解决这一问题，**QUEST** 框架 [116] 提出了一种结合自适应精度与量化感知训练的协同设计方法。通过将低比特权重（2 位）与减少时间步相结合，QUEST 的能耗较同等人工神经网络提升 $\sim$93倍。同样，对于响应时间至关重要的自主智能体，**SNN4Agents**[117]表明，同时量化权重（至 10 位）并缩减注意力窗口，可实现 $4\times$ 的效率提升，证实多维量化对于实时边缘智能至关重要。

脉冲神经网络优化的发展轨迹正从单变量压缩（仅进行权重剪枝）转向多维协同优化。未来的框架可能会将权重稀疏性、比特精度和时间分辨率（时间步）视为相互耦合的变量，并对其进行联合优化，以适配非冯·诺依曼神经形态硬件的具体约束。

C. 神经形态硬件：计算基底的演进

在边缘部署脉冲神经网络，需要采用能够超越冯·诺依曼瓶颈所固有能耗—延迟权衡的硬件。神经形态硬件的发展已从简单的生物仿真转向复杂的领域专用架构。我们将近期进展归纳为四种截然不同的架构理念：数字异步原语、混合范式、扩展式系统，以及基于 FPGA 的可重构加速器。表 II 总结了这些平台，并将其容量、学习可塑性和边缘部署适用性与标准商用现成（COTS）硬件进行了对比。

1）数字异步架构：效率基准：神经形态工程的基础理念是通过事件驱动的异步通信优先实现能效。Intel Loihi 和 IBM TrueNorth 等典型实例采用细粒度电源门控，核心在脉冲事件发生前始终处于空闲状态 [71], [122], [123]。**与边缘图形处理器对比：**边缘图形处理器（如 NVIDIA Jetson）依赖批处理和连续的内存访问（静态功耗高），而异步架构在稀疏工作负载下可将动态功耗降低数个数量级。然而，它们严格遵循脉冲神经网络动力学，往往难以有效支持标准深度神经网络层，因此有必要转向混合设计。

2）混合架构：跨范式融合：为了解决纯神经形态设计的僵化问题，**混合架构**应运而生，旨在兼顾计算机科学（CS）的准确率与神经科学（NS）的效率。一项开创性突破是**天机**芯片 [73]，它引入了统一功能核心（FHC），可支持多种模式并发运行。通过集成可配置统一缓冲区和可适配轴突，天机使多种模型——从符合生物学原理的脉冲神经网络到标准卷积神经网络——能够共存于同一基底上。沿袭这一范式，一些新兴设计采用了这种异构策略，以支持在片上原生实现人工神经网络到脉冲神经网络的转换 [121], [124]–[126]。

表 II 与边缘脉冲神经网络相关的主要神经形态硬件平台（2023—2025年更新）

平台	Dev.	边缘脉冲神经网络的主要特征	容量	学习	Eff.	规模	边缘相关性	Ref.
Loihi 2	英特尔	可编程脉冲神经元，动态突触、异步处理	>1 M	Yes	●	●	机器人学、人工智能 control	[95]
Hala Point	英特尔	由 1,152 颗 Loihi 2 芯片组成的系统，大规模并行性	1.15 B	Yes	◐	●+	大规模人工智能 research	[118]
Darwin3	ZJU	机架级系统，专用脉冲神经网络指令集架构，DarwinOS 支持	>2 B	Yes	◐	●+	脑模拟，AGI	[119]
NorthPole	IBM	存内计算，仅推理的运算优化	N/A	No	●+	N/A	超低功耗 AI	[101]
Akida 2	BrainChip	事件驱动，IP for SoC 集成，张量化神经网络支持	可扩展	Yes	●+	●	IoT，可穿戴设备，智能家居	[120]
SpiNNaker2	曼彻斯特	ARM内核，定制脉冲路由	>1 M/bd	软件	○	●+	机器人学，嵌入式系统	[121]
天机	清华	混合式 ANN-SNN 架构	10^5	Yes	◐	●	自主驾驶	[73]

图例：能效 = 能效；**规模** = 可扩展性。评级： ●+ = 最佳（超低功耗/高可扩展性）； ●= 极高； ◐= 高； ○= 中等。注：M：百万个神经元，B：十亿个神经元。容量因具体实现而异。

与边缘中央处理器的对比：嵌入式 CPU 虽然具备通用灵活性，但缺乏实时传感器融合所需的大规模并行能力。混合芯片弥补了这一不足，使自主机器人等系统能够同时利用脉冲神经网络实现低延迟反射（障碍物规避），并利用人工神经网络实现高层认知（语义识别）。

3）横向扩展系统：从芯片到脑规模：边缘推理侧重于单芯片效率，与此同时，另一条研究路线致力于解决全脑模拟的可扩展性挑战。**Darwin** 系列 [127] 代表了这一方向上的一项重要努力。从早期原型到大型**Darwin3** 芯片 [119]的发展，体现出一种针对脉冲动力学和片上可塑性优化的专用指令集架构（ISA）。至关重要的是，研究重点已转向系统级集成。**Darwin Monkey（“悟空”）**系统通过高速专用互连集成了近 1,000 个 Darwin3 芯片，实现了超过 20 亿个神经元和 1,000 亿个突触的规模。该系统着力解决大规模神经形态集群中固有的互连瓶颈 [72], [74], [128], [129]，因此超越了包括近期 Intel Hala Point 配置在内的许多现有平台。**与数据中心 GPU 的对比：**与在稀疏脉冲交换中受通信开销困扰的 GPU 集群不同，Darwin Monkey 利用专用路由，高效支持类脑操作系统（DarwinOS）。

4) 基于 *FPGA* 的加速器：务实的折中方案：现场可编程门阵列（现场可编程门阵列）已成为一种至关重要的“折中方案”，兼具 ASIC 所不具备的可重配置能力，以及图形处理器所欠缺的确定性延迟。近期研究聚焦于克服脉冲神经网络在同步结构上的独特映射挑战 [76]。

突破内存墙：脉冲神经网络需要持久状态存储（膜电位、V_{mem}），从而造成显著的带宽压力。为缓解这一问题，提出了**S2N2** 流式处理架构 [130]，通过将状态更新限制在活跃神经元上来优化块 RAM 的使用。挑战事件驱动范式的**SyncNN** [131] 证明，在现场可编程门阵列上采用**同步数据流**实际上可以优于异步映射，使吞吐量提高**4.6**×，能效提高**2.3**×，相比最先进水平的事件驱动加速器而言。同样，权重固定—局部输出固定（WS-LOS）数据流 [132] 旨在最大限度地提高数据复用率，实现超低能耗（24.3 μ焦耳/图像）。

稀疏性感知流水线：针对脉冲的不规则稀疏性，**FireFly v2** [133] 引入了一种时空数据流，可消除存储冲突，并实现 600 MHz 的时钟频率。此外，**SpikeX** [134] 提出了一种面向非结构化稀疏性的软硬件协同优化框架。通过动态重新平衡脉动阵列，SpikeX 相较于稠密基线将能量延迟积（EDP）降低了超过 15×。

神经形态硬件的发展正经历一场根本性转变：从“严格的生物仿真”转向“功能异构性”。早期设计优先忠实复现神经元动力学，往往以牺牲计算灵活性为代价。然而，当前趋势更倾向于采用软件定义架构，使硬件资源能够在同步（类似人工神经网络）和异步（类似脉冲神经网络）模式之间动态重新分配。这一演进表明，边缘智能的未来并非彻底取代冯·诺依曼架构，而是通过增添领域专用的事件驱动加速器来增强它们，使其能够无缝集成到标准计算技术栈中。

D. 软件鸿沟：工具链与编译器

尽管神经形态硬件发展迅速，软件生态系统仍是阻碍其广泛应用的主要瓶颈。与面向图形处理器的成熟 CUDA 生态系统不同，脉冲神经网络领域目前存在严重的碎片化问题。由于缺乏统一的**中间表示（IR）**和标准化的编译器栈，开发者往往不得不针对特定硬件架构手动优化脉冲动力学。

1) 映射工具链：解决嵌入问题：将逻辑脉冲神经网络图（神经元和突触）映射到物理神经形态结构（核心和片上网络路由器）并非易事，这是一项组合优化问题，称为“映射问题” [22], [135], [136]。近期的进展着重通过自动化工具链弥合这一鸿沟，在通信带宽、能耗和延迟之间取得平衡。

NeuMap 通过将网络分区视为簇到核心分配问题，解决了将脉冲神经网络映射到多核架构的挑战。它通过分析脉冲通信模式和硬件约束，采用元启发式算法优化放置；与 SpiNeMap 等先前基准方案相比，能耗最多降低 84%，延迟改善 55%。

EdgeMap针对资源受限的边缘设备扩展了这一思路 [137]。它引入了一种基于流式处理的分区方法，严格考虑扇入/扇出限制。EdgeMap 利用基于 NSGA-II 的多目标优化，在能耗与通信成本之间权衡。基准测试结果彰显了其效能：与传统映射方案相比，执行时间提升 1225 倍，能耗降低 57%。

2）软件框架与碎片化：当前的开发环境分为厂商专用工具链和通用库两大类，如表 III 所示。

厂商专属生态系统：英特尔的 **Lava**代表着向模块化转变，为 Loihi 和其他神经启发式处理器提供开源框架。它支持超细粒度并行性和动态神经场（DNF），但其全部潜力仍与 Loihi 架构紧密耦合 [139]。同样，MetaTF 等框架也仍受 Akida 硬件限制。

通用库：在算法方面，基于 PyTorch 的库（如 **SpikingJelly** 和**SNNTorch**）通过将脉冲神经网络动力学融入标准深度学习工作流程，降低了研究人员的入门门槛 [24]。Brian2 和 Nengo 等成熟的模拟器则继续服务于计算神经科学领域 [143]–[145]。为理清这一碎片化生态系统，近期一项全面的多模态基准测试 [146]对包括 SpikingJelly、BrainCog 和 Lava 在内的框架进行了量化评估。该研究揭示了不同任务中能效与延迟之间的重要权衡，为选择适用于边缘部署的工具链提供了数据依据。至关重要的是，这些工具主要关注节点内算法保真度。为了弥合与现实部署之间的差距，**iFogSim** [147] 等系统级模拟器对于建模分布式边缘环境中固有的节点间延迟、带宽限制和能耗，正变得愈发重要。然而，如何将这些高级 Python 库无缝编译到异构神经形态芯片上，仍存在一个关键的“软件鸿沟”。

E. 系统级集成：云、边缘与物联网

脉冲神经网络算法的成熟推动了从独立加速向**分布式边缘智能**的转变。这一范式将重点从单芯片推理转向协作系统，使脉冲神经网络能够在更广泛的物联网和云生态系统中运行。

分布式无线智能：在分布式传感器节点上部署脉冲神经网络，会带来带宽和同步方面的独特挑战。近期研究将其建模为联合优化问题，在无线信道上严格限制脉冲丢失的同时，尽可能降低能耗 [15]。这确保了分布式脉冲神经网络即使在嘈杂的物联网环境中，也能保持较高的推理准确率。

边云协同： 为平衡边缘设备有限的处理能力与云端的计算能力，分层架构应运而生（图 6）。为促进这种分层协同，人们提出了诸如 **FogBus** [148] 这样的轻量级集成框架，用于管理资源受限的边缘节点与云端之间的异构性和数据流。基于这一范式，诸如**ECC-SNN**这样的框架运用知识蒸馏，由复杂的云端人工神经网络充当“教师”，指导轻量级边缘脉冲神经网络“学生” [108]。其中一项关键创新是基于置信度的卸载机制：边缘节点在本地处理实时事件流，仅将低置信度的“困难样本”传输至云端进行高精度分析。这一策略在保持全局系统准确率的同时，显著降低了通信开销。

脉冲神经网络联邦学习： 为解决数据隐私和“数据孤岛”问题，联邦学习（FL）已应用于脉冲域。诸如 **FedLEC**的方法针对神经形态网络中的非 IID（标签偏斜）数据问题。通过采用客户端内标签权重校准和客户端间蒸馏，这些方法能够缓解模型漂移，使脉冲神经网络在不共享原始事件数据的情况下协同学习 [149]–[152]。

脉冲神经网络软件的发展轨迹正接近一个转折点，类似于早期 *GPU* 计算的发展阶段。当前的碎片化局面——以彼此孤立的厂商工具链和互不关联的映射算法为特征——难以为继。未来的方向在于标准化。业界必须共同采用统一的神经形态中间表示（*IR*）和跨平台编译器栈（类似 *LLVM* 或 *TVM*）。只有当 *PyTorch* 中的高级描述能够无缝编译到任何底层神经形态基底——无论是 *Loihi*、*TrueNorth*，还是基于现场可编程门阵列的加速器——脉冲神经网络才能实现普及，从而满足下一代普适计算的需求。

表 III 精选的脉冲神经网络软件框架与映射工具链（2020–2025）

框架（开发者）	核心功能与用途	边缘相关能力	Ref.
硬件映射工具链			
NeuMap (国防科技大学)	将脉冲神经网络映射到多核硬件；通信模式计算；元启发式簇到核心映射	针对片上网络延迟和能耗进行优化，适用于多核边缘硬件	[138]
EdgeMap (上海交通大学)	针对边缘优化的映射；基于流式处理的分区tioning；多目标优化（NSGA-II）	显著降低延迟、能耗和边缘部署的通信成本	[137]
深度学习与部署框架			
Lava （英特尔）	神经启发式应用开发；模块化；平台无关；支持深度学习（lava-dl）	原生部署于 Loihi 芯片；针对能耗／速度效率进行优化	[139]
SNNTorch （加州大学圣克鲁兹分校）	基于 PyTorch 的训练；替代梯度；时间动态；GPU 加速	便于训练可部署的深度脉冲神经网络于边缘设备；无缝集成 PyTorch	[23]
SpikingJelly (北京大学)	基于 PyTorch；LIF/PLIF 神经元；直接训练（替代梯度）；丰富的编码方法	支持开发适用于用于直接边缘部署	[89]
仿生模拟与原型设计			
Nengo 艾伦脑科学研究所	大规模建模（神经工程框架）；支持多种后端——（Loihi、SpiNNaker、中央处理器/图形处理器）	在不同神经形态平台上验证的实时控制应用——多种神经形态平台	[140]
BindsNET （马萨诸塞大学）	生物学合理性模拟；STDP/R-STDP 规则；事件——驱动动力学	原型开发 具备设备端在线学习电位	[141]
Brian2 (索邦大学)	生物物理模拟；灵活的模型定义；代码生成	模拟 复杂的，生物学上 详细的在进行硬件映射之前的脉冲神经网络	[142]

VI. 应用

在边缘计算环境的严格约束推动下，脉冲神经网络正从理论模型迈向可部署的解决方案。脉冲神经网络并非只是提供人工神经网络的一种替代方案，还能应对特定的物理与运行瓶颈：**极低的功耗预算**（电池续航时间）、**通信带宽**（原始数据传输造成的饱和）、**延迟**（安全关键型反馈回路），以及**数据隐私**（本地处理）。本节将从这些约束的角度分析近期（2023年–2025年）的实现案例。表 IV 从准确率、能效和延迟方面，对这些应用与深度神经网络基线进行了定量比较。

A. 计算机视觉：高速与事件驱动传感

视觉系统在边缘产生的数据量最大，造成巨大的带宽和能耗负担。脉冲神经网络仅在场景发生变化时处理信息，从而缓解这一问题。

1）高速仿视网膜传感：标准帧式相机容易产生运动模糊，且数据冗余度高。动态视觉传感器（DVS）虽能解决这一问题，却往往会丢弃静态纹理信息。为弥合这一差距，研究者提出了**脉冲相机（Vidar）**架构[161], [162]。该范式采用类中央凹采样机制，像素根据亮度强度发出连续脉冲，实现了超高时间分辨率（最高可达 40,000 Hz）。这使脉冲神经网络能够重建高速运动（例如快速机械旋转）——这类运动传统相机无法捕捉——同时显著压缩数据流，与高速视频相比，数据量大幅减少 [163]–[165]。

2）低功耗手势与活动识别：对于电池供电的人机界面（HMI），首要限制因素是能耗。脉冲神经网络利用手势事件的空间稀疏性来最大限度地减少计算。神经形态硬件上的实现已展示出，在主动推理期间，功耗可低至 **178 mW**[166]。近期进展已将这一能力从简单的手势识别扩展到复杂的**人体活动识别（HAR）** [167]。值得注意的是，2025年2月的一项研究提出了一种多模态框架，将事件相机与脉冲图卷积网络（SGCN）相结合。该系统通过在本地处理骨架数据，避免传输涉及隐私的原始视频流，从而有效解决智能家居环境中的隐私问题，同时保持较高的识别精度 [111], [156]。

B. 超越视觉：听觉与嗅觉传感

非视觉模态天然具有稀疏性。脉冲神经网络利用这一特性，解决“始终在线”的功耗难题。

1) 神经形态听觉处理：在智能音箱和助听器中，唤醒词引擎必须持续运行。静音期间，脉冲神经网络可提供动态功耗为零的解决方案。基于 Intel Loihi 芯片的基准测试表明，在毫瓦级功耗范围内，脉冲神经网络在关键词检测（KWS）方面优于基于图形处理器的深度神经网络 [168]。为进一步推进这种仿生方法，循环脉冲神经网络架构已

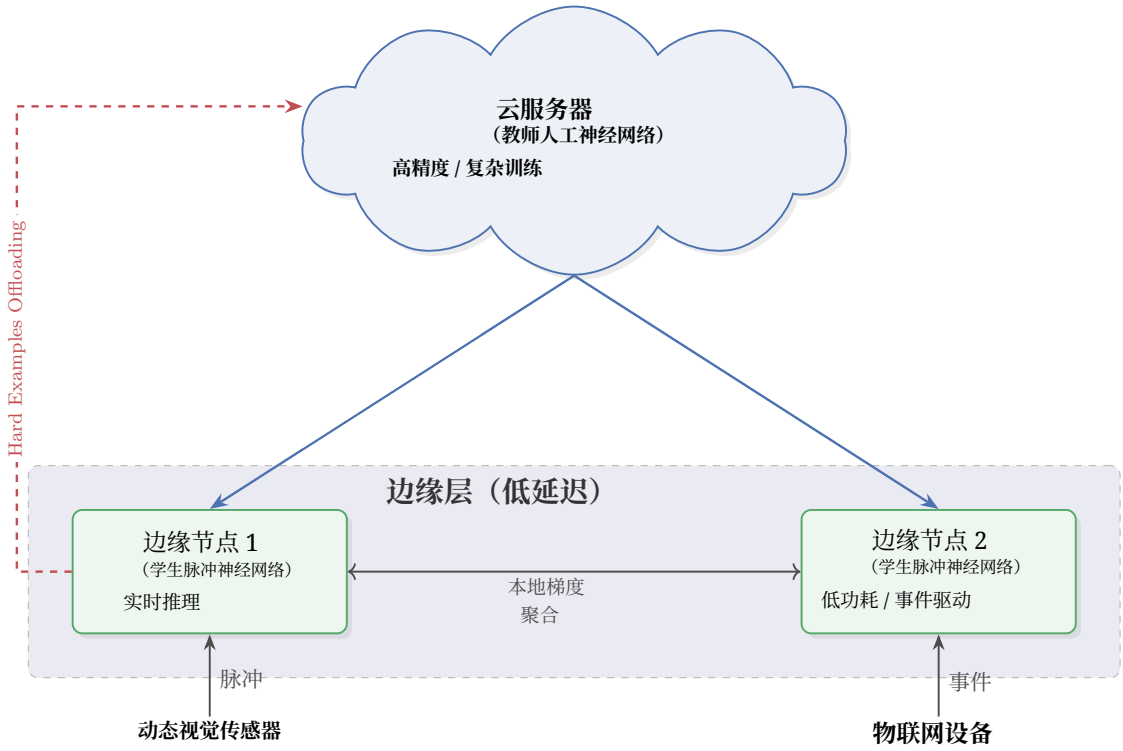

图 6. **提出的基于脉冲神经网络的云边协作框架**。云服务器部署了复杂的教师人工神经网络，通过知识蒸馏指导边缘节点上的轻量级学生脉冲神经网络。边缘节点处理实时事件流，仅将低置信度的“困难样本”卸载到云端，从而大幅减少带宽使用量。

用于提取细粒度时间特征，并达到最先进水平的准确率 [169]。值得注意的是，2025年4月的一项研究报告称，在商用神经形态处理器上，关键词识别的延迟为 **<1毫秒** [157]。除了语音之外，脉冲神经网络也在改变**环境声音分类（ESC）**。通过模仿人类听觉通路，仿生模型能够以极低功耗对背景场景进行分类，使声学传感器得以长期部署在偏远地区，无需频繁更换电池 [168], [170], [171]。

2） 神经形态嗅觉（电子鼻）：嗅觉数据维度高但稀疏，这给需要大量训练集的标准分类器带来了挑战。脉冲神经网络能够应对**数据匮乏**的限制。通过模仿哺乳动物嗅球，研究人员实现了神经回路，使其能够对危险化学品进行**“单样本学习”** [172]。与深度神经网络不同，这类系统只需接触一次，就能学习新的气味特征。进一步的工业验证证实，脉冲神经网络能够有效应对气体识别系统中的传感器漂移；这是传统算法常见的失效原因[173]。

C. 机器人与自主系统

在机器人领域，机载计算负载与飞行时间/运行时间之间的权衡至关重要。脉冲神经网络通过神经形态控制回路提供了一种解决方案。

1) 无人机与导航：对于无人机（UAVs），处理能力会直接与电池续航时间相竞争。**SNN4Agents框架**在NCARS数据集上实现了4.03倍的能效提升 [117]。此外，用于视觉地点识别（VPR）的模块化脉冲神经网络支持资源受限无人机上的可扩展映射 [155]，并能以亚毫秒级的速度响应障碍物——当地面通信中断时，这是一项至关重要的安全特征 [174], [175]。

2) 边缘云编排：边缘脉冲神经网络负责处理反射任务，而复杂查询则需要云端支持。**ECC-SNN框架**（2025年5月）展示了这一层级架构：通过动态将低置信度的“困难样本”卸载到云端，该框架在 CIFAR-10 上将准确率提高了 4.15% [108]。这种架构兼顾了本地能耗限制与全局服务质量（QoS）。

D. 医疗保健与可穿戴设备

医疗应用面临双重约束：植入式设备需要极高的能效，而可穿戴设备则必须严格保障数据隐私。

1） 生物信号监测：脉冲神经网络越来越多地应用于现场可编程门阵列上的心电图分类和心律失常检测[76]。2024年6月的一篇综述进一步印证了这一趋势，指出基于脉冲神经网络的实现正迅速缩小与深度神经网络之间的准确率差距，同时功耗仅占功耗预算的一小部分 [176]。通过在本地处理原始生物信号，这些系统确保敏感波形绝不会通过无线方式传输，从而显著缩小数据泄露的攻击面 [177], [178]。

表 IV 脉冲神经网络在边缘计算中的典型应用（2022–2025）

领域	Task	方法 / 模型	硬件	关键性能指标	Ref.
CV	神经渲染	Spiking-NeRF（混合脉冲神经网络）	图形处理器/模拟	**质量：** 峰值信噪比相当；**能耗：** 2.27× 相比神经辐射场降低	[153]
	手势识别	OSBC（单次剪枝）	DVS128-Gesture	**稀疏度：** 97%；**准确率损失：** 1.7%；**方法：** 训练后量化	[154]
机器人学	视觉地点识别	模块化脉冲神经网络集成	资源受限	紧凑模块（1.5k 个神经元）；序列匹配提升 R@1	[155]
	自主智能体	SNN4Agents（量化）	事件传感器	**累加：** 84.1%；**内存：** −68.8%；**效率：** 4.0× 倍提升	[117]
IoT	人体活动识别（多模态）	脉冲神经网络驱动的融合（事件 + 骨架）；SNN-Mamba	N/A	通过稀疏多模态融合实现高效率	[156]
	关键词识别	片上学习脉冲神经网络	商用神经形态处理器	**延迟：** < 1毫秒	[157]
医疗保健	神经解码	自适应剪枝脉冲神经网络	SENECA	**功耗：** 0.18 μW/步；**效率：** > 10×，相比稠密 SNNs	[110]
	心电图分类	FPGA上的脉冲神经网络	现场可编程门阵列 / 神经形态计算	**准确率：** 98.2%（MIT-BIH）；超低功耗 准确率与深度神经网络相当。	[158]
工业	缺陷检测	脉冲神经网络模型强化	工业数据集	**速度：** 比重新训练快 11–27× 倍的机器遗忘 重新训练	[159]
	故障诊断	深度脉冲残差网络/图神经网络	工业传感器	稳健的时空建模；高效率	[160]
边缘-云端	分类	ECC-SNN（联合训练）	边缘 + 云	**准确率：** +4.1%; **能耗：** −79%; **延迟：** −39%	[108]

2）可植入接口：对于皮层内神经解码，散热是防止组织损伤的限制因素。一项发表于2025年4月、关于脉冲神经网络自适应剪枝的研究展示了**低于μW 的功耗**（0.18 μW/时间步） [110]。这种热效率是标准冯·诺依曼架构无法实现的，也是长期脑机接口的关键使能技术。

E. 工业检测与智能基础设施

在大规模基础设施中，主要瓶颈是带宽饱和和数据主权。

1) 工业故障诊断（*IFD*）：由于带宽成本高昂，将数千台机器产生的高频振动数据传输到中心云通常并不现实。脉冲神经网络支持**边缘学习**，使模型能够在本地适应新的缺陷。近期方法展示了在边缘节点 [159], [179] 上直接进行表面缺陷检测的脉冲神经网络模型强化。这避免了将专有生产线数据暴露给外部网络。

2）智慧城市与公共安全：在交通管理中，传输原始视频流会使网络饱和。脉冲神经网络支持“语义压缩”，只传输元数据（例如“检测到事-故”），而非原始像素 [180]。除交通领域外，搭载脉冲神经网络的无人机在**灾害疏散监测** [181] 中也发挥着至关重要的作用。在这类场景中，依赖可能已受损的蜂窝基础设施存在风险，因此需要脉冲神经网络所提供的自主机载智能。此外，通过在传感器层面处理视频并即时丢弃个人身份信息，这些系统天然符合更严格的隐私标准[182], [183]。

在这些领域中，脉冲神经网络的适用性呈现出明显的规律。对于具有事件驱动动态、严格延迟要求和严苛功耗限制的场景（例如神经形态控制、常开传感），脉冲网络表现出显著优势。相反，在功耗预算较宽松、要求高精度静态模式识别的任务中（例如静态 *ImageNet*分类），脉冲神经网络相较于传统卷积神经网络的优势仍不明显。

VII. 重大挑战：系统性摩擦

尽管脉冲神经网络（SNNs）在边缘智能领域具有理论吸引力，但其从学术界的新奇研究走向主流部署的进程，却受到显著的系统性摩擦阻碍。这种摩擦不仅源于工程缺陷，更根植于当前深度学习范式、硅基硬件限制与脉冲动力学离散特性之间的根本性不匹配。我们将这些挑战归纳为三个关键维度：算法学习困境、软硬件鸿沟以及部署的实际情况。表V对这些障碍进行了系统分类，将具体的技术痛点（例如死亡神经元、内存墙）与当前研究方向及代表性文献相对应。

表V 边缘部署脉冲神经网络面临的主要挑战及当前研究方向

挑战领域	具体挑战	当前研究方法与方向	重要参考文献
训练复杂度	不可微的脉冲事件	**替代梯度 (SG)**；事件驱动学习 (STD-ED, MPD-ED)；可微脉冲 (Dspike)	[24], [90]
	梯度消失与“死亡神经元”	**可学习替代梯度**；自适应梯度规则；阈值相关批归一化（tdBN）；活动正则化	[24], [184]
	生物学合理的可扩展性（脉冲时序依赖可塑性）	混合学习（例如，奖励调制脉冲时序依赖可塑性）；监督式脉冲时序依赖可塑性变体；将脉冲时序依赖可塑性与反向传播相结合	[64], [89]
可扩展性	在大型数据集（例如，ImageNet）	**人工神经网络到脉冲神经网络的转换**；脉冲神经网络专用深度架构（脉冲化 Transformer）；混合模型	[107], [185]
硬件约束	内存墙与扇入/扇出限制	**协同设计框架**（例如 QUEST）；存内计算计算；SNN 专用剪枝/量化（OSBC）	[111], [116]
	芯片间通信与异构性	优化的片上网络设计；标准化中间表示（例如 NIR）；平台无关框架（Lava）	[109], [138]
能耗/延迟	准确率与效率的权衡	**联合优化**（例如，SNN4Agents）；边缘-云端协同（ECC-SNN）；提前退出机制	[108], [117]
标准化	碎片化生态系统与基准测试	**统一框架**（Lava、Nengo）；联邦脉冲神经网络学习（FedLEC）；神经形态基准测试套件	[109], [149]

A. 学习困境：离散世界中的优化

阻碍脉冲神经网络普及的首要障碍，是难以有效训练深层脉冲架构这一算法层面的挑战。

1）不可微性的根本分歧：深度学习中占主导地位的基于梯度的优化（例如反向传播）与脉冲神经网络的离散动力学之间存在根本分歧。脉冲的产生本质上是一个非此即彼的事件，通常用不可微的赫维赛德阶跃函数来建模。这种离散特性使直接求导变得不可能，破坏了标准反向传播所需的链式法则。尽管**替代梯度（SG）**——在反向传播过程中用连续函数（例如 Sigmoid函数或反正切函数）近似脉冲导数——已成为标准的变通方法，但它会引入梯度失配。随着网络深度增加，这种近似误差会不断累积，往往导致相较于人工神经网络更差的收敛效果。

2）训练稳定性：“死亡神经元”与时间反向传播：深度脉冲神经网络面临严重的稳定性问题。首先，**“死亡神经元”问题**意味着神经元可能由于初始化不佳或刺激不足而无法跨越发放阈值，从而保持沉默 [186]。与人工神经网络中的 ReLU 不同，沉默的脉冲神经元无法提供任何梯度信息，还会消耗时间步，却无法为推理做出贡献。其次，训练依赖**时间反向传播（BPTT）**来捕获时间依赖关系。时间反向传播会沿时间步展开网络，大幅增加内存占用，并加剧梯度消失/梯度爆炸问题，使超深脉冲神经网络（例如 ResNet-100+）的训练在计算上难以承受。

3) 可扩展性与转换之争：关于实现可扩展性的最佳途径，争论仍在持续：

- **人工神经网络到脉冲神经网络的转换：**这种方法利用成熟的人工神经网络训练流程，并将权重转换为脉冲神经网络。虽然稳定，但它往往存在延迟较高的问题（需要较长的仿真窗口来近似频率编码），且无法利用单个脉冲丰富的时间动态 [187]。
- **直接训练：**直接优化（通过替代梯度）能够捕获时间特征，并实现低延迟推理。然而，它难以解决上述可扩展性问题。

因此，尽管脉冲神经网络在神经形态基准测试（例如 CIFAR10-DVS）中表现出色，但若不付出过高的延迟代价，它们仍难以在 ImageNet 等大规模静态数据集上达到最先进水平的人工神经网络准确率。

B. 软硬件鸿沟：突破内存墙

脉冲神经网络的理论能效通常建立在理想化硬件的基础上。实际上，硬件实现受到存储层次结构和互连方面的严格限制。

1）状态存储器墙：双重瓶颈：与无状态的深度神经网络相比，脉冲神经网络面临一项独特挑战：必须在每个时间步 [188] 维护每个神经元的动态状态（膜电位、V_{mem}）。这造成了“双重瓶颈”：既要存储突触权重，也要存储神经元状态。

如图 7 所示，要解决这一问题，需要采用不同的架构：

- **冯·诺依曼瓶颈（图 7a）：**传统架构将处理单元（CPU/GPU）与存储器分离。数据必须经过系统总线传输，从而形成“内存墙”，其中能耗主要由数据移动

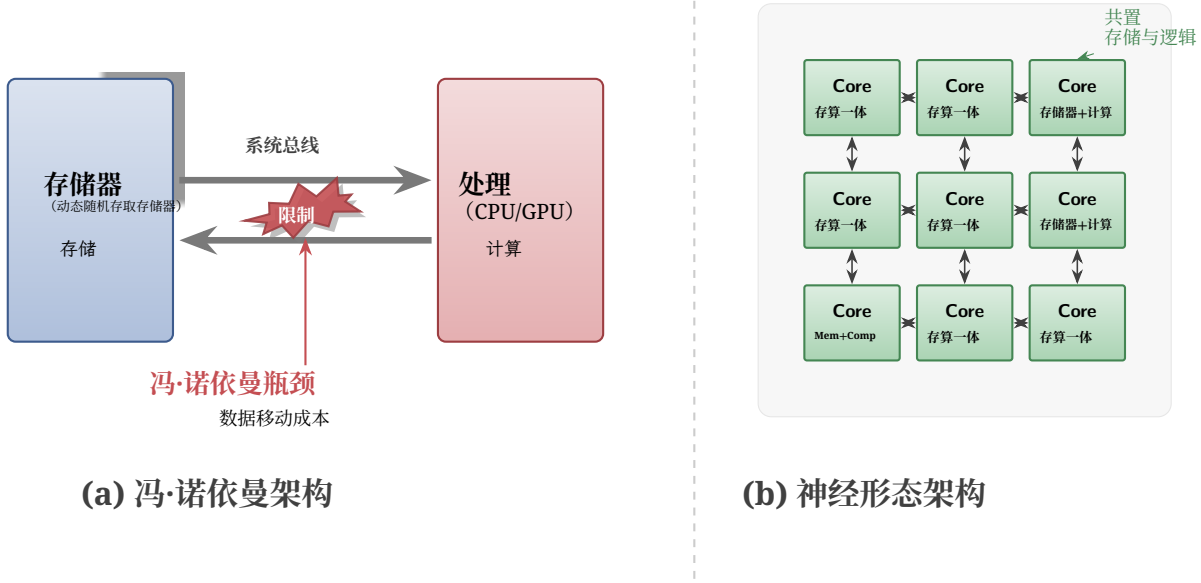

图 7。 **架构比较**。(a) 传统冯·诺依曼架构将存储器与处理分离。数据总线（以平行线表示）会成为瓶颈。(b) 神经形态架构将存储器与计算（M+C）整合到分布式核心中。

主导（100× ～ 1000×高于计算），而非算术逻辑。

- **神经形态局部性（图 7b）：** 神经形态芯片（例如 Intel Loihi、TrueNorth）采用非冯·诺依曼方式，在分布式核心中将存储器与计算置于同一位置。通过在本地存储状态（静态随机存取存储器），它们最大限度地减少了片外数据传输。

然而，这种局部性也带来了容量限制。高速片上静态随机存取存储器价格昂贵且容量有限 [72], [73],[189]–[191]。当深度脉冲神经网络模型超出片上容量时，系统就不得不访问片外动态随机存取存储器，导致能耗急剧增加，从而抵消事件驱动处理的优势。

2）连接性与碎片化：**扇入/扇出稀疏性：** 与理论模型中的全连接不同，物理芯片的路由资源有限。将稠密网络映射到稀疏硬件图上，需要采用“多跳”脉冲路由 [137], ，这会引入非确定性延迟和拥塞。

软件栈空缺： 或许最关键的障碍是缺乏类似 NVIDIA CUDA 的统一软件栈。整个领域被异构架构（异步、混合信号架构、数字架构）所割裂，阻碍了标准化编译工具链 [192]的开发。这迫使研究人员投入大量劳动进行硬件专用优化。

C. 部署现实：安全与运营权衡

随着脉冲神经网络从实验室走向暴露于风险中的边缘环境，新的运行漏洞随之出现。

1）对抗鲁棒性：时间脆弱性：过去，人们认为脉冲神经网络具有随机性和离散滤波带来的固有鲁棒性。然而，近期的审视揭示了这种安全感并不可靠。研究已证实，脉冲神经网络极易受到**时间扰动** [193]的影响。攻击者可以在不被察觉的情况下改变输入脉冲的时序，导致膜电位无法达到发放阈值。近期综述指出，基于梯度的攻击可以适配时间域，因此需要采用离散对抗训练等强健的防御机制 [194], [195]。

2）隐私与侧信道：尽管联邦神经形态学习（FNL）通过仅传输权重更新 [50],来保护隐私，硬件实现仍容易受到攻击。物理侧信道攻击可以监测脉冲事件独特的功耗特征，从而逆向推断输入刺激或模型架构 [196], ，因此需要在硬件层面采取掩蔽对策。

3）能耗—精度权衡：最后，部署面临着严峻的权衡。要实现脉冲神经网络所承诺的超低能耗，往往需要采用激进的量化策略（例如使用 4 位权重）或极短的仿真窗口，这可能会严重降低准确率。ECC-SNN 等框架尝试通过将难以处理的样本卸载到云端来缓解这一问题；然而，这又重新引入了对网络连接的依赖。

归根结底，*SNN* 的未来并不在于简单地用其替代 *DNN* 来完成所有任务，而在于识别并主导“时空稀疏性利基”——即涉及异步事件流、超低延迟控制环路和受限功耗预算的场景。“重大挑战”在于弥合软件灵活性与硬件

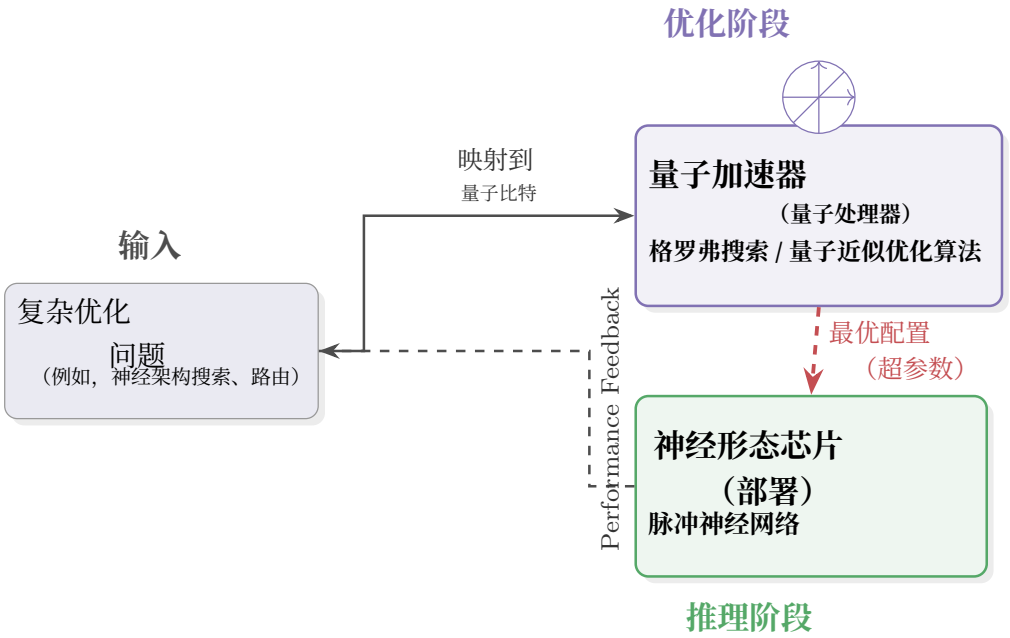

图 8. **量子-神经形态协同概念**。 量子算法（例如量子近似优化算法）用于解决脉冲神经网络设计中固有的 NP 难优化问题，例如神经架构搜索（NAS）或突触路由，从而能够在神经形态硬件上部署高度优化的脉冲神经网络。

基元之间的鸿沟，将理论上的效率转化为实际运行中的成果。

VIII. 未来路线图：迈向普适智能

展望当前的障碍之后，脉冲神经网络在边缘计算领域的发展轨迹，并非只是渐进式改进，而是代表着根本性的架构转变。如果上一节概述了各种阻碍，那么这份路线图便是下一代边缘智能的蓝图。我们将这一愿景分为三个紧密相连的支柱：算法前沿、软硬件连续体以及新兴技术的融合。

A. 算法前沿：从规则到推理

为了弥合生物效率与通用人工智能（AGI）之间的差距，我们必须超越简单的分类任务，转向复杂的时空推理 [197]。这一跨越需要学习规则与模型规模协同演进。

1）统一学习范式：必须解决生物合理性与基于梯度的优化之间的两难问题。未来的算法可能会采用混合策略：使用**高级替代梯度**对深层架构（例如脉冲变换器）进行离线元学习，同时利用生物启发规则（如高级脉冲时序依赖可塑性变体）实现快速、节能的**片上适应** [157]。目标是开发能够抵御梯度消失问题的“直接训练”方法，使脉冲神经网络能够捕获长程依赖，同时避免时间反向传播带来的过高计算成本 [185], [198]。

2）大规模脉冲模型：正如大语言模型革新了自然语言处理，大规模脉冲模型（LSM）的时代也即将到来。未来的模型将把时间移位模块和注意力机制等机制直接融入脉冲领域，从而在边缘端统筹复杂的多模态任务。这将使范式从静态数据集转向动态的事件流处理，而脉冲神经网络的稀疏度将在浮点运算次数削减方面带来显著优势 [186]。

B. 软硬件连续体

只有在底层基底（包括物理和虚拟基底）针对事件驱动动态进行优化时，算法的潜力才能得以实现。

1）新一代神经形态硬件：集成**新兴非易失性存储器（eNVM）**至关重要。诸如**忆阻器（RRAM）**和**相变存储器（PCM）**等器件，能够实现高扇出突触连接的物理仿真 [199]–[201]。与传统冯·诺依曼架构不同，这些器件支持模拟存内计算（*CIM*），使突触累加直接在存储阵列内进行，几乎消除了数据移动的能耗 [202]。

2）神经形态操作系统：编排与虚拟化：或许最关键、却尚未得到充分探索的前沿领域，是系统级软件栈。当前的云编排框架（例如 Kubernetes）存在根本性的**“同步-异步失配”**：它们将异步、细粒度的脉冲流强行装入低效的同步固定时间批次中。这种失配造成了**“因果性瓶颈”**，脉冲神经网络的延迟优势被编排层的等待时间抵消。

为了释放其全部潜力，未来的边缘系统必须演进为真正的**神经形操作系统**，并以三种核心机制为基础：

1) **事件驱动调度：**从基于帧的调度转向中断驱动策略，以遵循神经事件的因果关系 [72], [203]。

2) **动态异构性：**调度器必须量化转换成本 **(PoC)**——在脉冲（神经形态）与非脉冲（CPU/GPU）领域之间转码所产生的延迟和能耗。任务卸载决策应以动态最小化这一转换成本为依据[73], [198], [204]。

3)**服务级别协议感知的脉冲容器化：**为实现多租户，我们设想采用针对有状态神经元优化的轻量级虚拟化技术。这需要开发新机制，以保存/恢复膜电位，从而支持在不同脉冲神经网络模型之间切换上下文，且不降低性能 [205]–[207]。

C. 融合：量子、*6G*与伦理

脉冲神经网络并非孤立存在，而是作为认知引擎，融入更广泛的技术生态系统。

1）*6G*纽带：随着6G网络实现超低延迟和语义通信，脉冲神经网络有望成为网络边缘原生的处理方言 [208], [209]。全面的6G愿景和综述指出，神经形态计算是直接在基础设施层处理海量传感器数据的关键推动力 [210]。

2）量子-神经形态协同：尽管目前仍处于探索阶段，量子计算与脉冲神经网络的交汇为超越经典计算的局限提供了一条途径。如图8所示，我们提出一种协同范式，其中量子机器学习（QML）算法（例如量子近似优化算法）充当**离线加速器**。这些量子智能体可解决脉冲神经网络设计中固有的NP难优化问题，例如神经架构搜索（NAS）或复杂的突触路由。随后，将优化后的配置映射至神经形态硬件，以实现实时超低功耗推理 [211]。

3）合乎伦理且安全的智能：“始终在线”的边缘人工智能无处不在，因此必须设置严格的伦理护栏。除了提高效率，脉冲神经网络还必须通过联邦学习（例如 FedLEC [149]）来保护隐私，并增强抵御针对脉冲时序的对抗攻击的能力 [159], [212]。

最终，我们设想，脉冲神经网络与深度神经网络之间的严格界限将会消融。边缘智能的未来在于“统一神经形态连续体”，在其中，脉冲神经网络作为用于持续监测的“超低功耗档”，仅在需要更高认知分辨率时，才会无缝切换至深度神经网络的“高性能档”。这种无缝切换正是自适应边缘计算的终极目标。

IX. 结论

本综述首先探讨了神经形态计算面临的“部署悖论”：即仿生模型的理论效率与当代冯·诺依曼硬件在实践中的僵化限制之间的矛盾。在本综述中，我们证明，这一悖论并非固有缺陷，而是一种过渡性摩擦，可通过软硬件协同设计的视角有效化解。通过协调算法稀疏性与物理硅片的限制，我们表明，脉冲神经网络正从理论上的新奇概念，逐步转变为可靠的运行现实。

至关重要的是，脉冲神经网络的意义远不止节能。它们代表着一种根本性的范式转变：从处理数据——即静态、基于帧的分析这一传统方式——转向处理变化。这种事件驱动理念使计算与物理世界的动态特性相契合，为从根本上消除困扰传统基于帧的深度学习的时间冗余提供了一条途径。

然而，要实现这一潜力，仅靠渐进式优化远远不够。正如我们所论证的，当前领域亟须突破的前沿在于消除“同步-异步失配”。我们迫切需要开发统一神经形态操作系统，以便在现代计算系统的同步基础设施中编排异步脉冲工作负载。只有将这些硬件复杂性抽象化，才能推动神经形态解决方案的广泛应用。

归根结底，随着我们站在6G时代和万物互联的门槛上，对无处不在的低时延智能的需求比以往任何时候都更加迫切。脉冲神经网络有望超越小众技术的定位，成为数字基础设施的“绿色认知基底”——确保未来的边缘智能不仅无处不在，而且可持续发展。

参考文献

[1] W. Shi、J. Cao、Q. Zhang、Y. Li 和 L. Xu，“边缘计算：愿景与挑战”，*IEEE*物联网期刊，第3卷，第5期，第637–646页，2016年。[2] Z. Zhou、X. Chen、E. Li、L. Zeng、K. Luo 和 J. Zhang，“边缘智能：借助边缘计算铺平人工智能的最后一公里”，《*IEEE* 会刊》，第107卷，1762，2019年8，1738无。pp. –。[3] R. Buyya、S. N. Srirama、G. Casale、R. Calheiros、Y. Simmhan、B. Varghese、E. Gelenbe、B. Javadi、L. M. Vaquero、M. A. Netto 等，“未来一代云计算宣言：未来十年的研究方向”，*ACM*计算调查（*CSUR*），第51卷，第5期，第1–38页，2018年。[4] D. R.-J. G.-J. Rydning、J. Reinsel 和 J. Gantz，“从边缘到核心的世界数字化”，马萨诸塞州弗雷明汉：国际数据公司，第16卷，第1–28页，2018年。[5] J. Chen 和 X. Ran，“边缘计算中的深度学习：综述”，《*IEEE* 会刊》，第107卷，第8期，第1655–1674页，2019年。[6] M. Satyanarayanan，“边缘计算的兴起”，《计算机》，第50卷，第1期，第30–39页，2017年。[7] A. Botta、W. De Donato、V. Persico 和 A. Pescap´e，“云计算与物联网的融合：综述”，未来一代计算机系统，第56卷，第684–700页，2016年。[8] F. Dong、X. Si 和 M.-F. Chang，“基于SRAM的存内计算电路的设计方法与趋势”，载于*2022*年*IEEE*第*16*届国际固态与集成电路技术会议（*ICSICT*），2022年，第1–4页。[9] C. Ram´ırez、A. Castell´o、H. Mart´ınez 和 E. S. Quintana-Ort´ı，“面向RISC-V GAP8微控制器深度学习的基于GEMM卷积的通信规避融合”，*IEEE*物联网期刊，第11卷，第21期，第35 640–35 653页，2024年。

[10] Y. Sun、T. Chen、Q. V. H. Nguyen 和 H. Yin，“Tinyad：面向工业物联网时间序列数据的存储器高效异常检测”，IEEE工业信息学汇刊，第20卷，834，2024年1，824无。pp. –。[11] C. Fang、Z. Shen、Z. Wang、C. Wang、S. Zhao、F. Tian、J. Yang 和 M. Sawan，“具有三维计算阵列的能效型非结构化稀疏感知深层脉冲神经网络加速器”，IEEE固态电路期刊，第60卷，第3期，第977–989页，2025年。[12] F. Liu、Z. Wang、W. Zhao、N. Yang、Y. Chen、S. Huang、H. Li、T. Yang、S. Pei、X. Liang 和 L. Jiang，“利用时间展开并行性实现能效型脉冲神经网络加速”，IEEE并行与分布式系统汇刊，第35卷，1764，2024年10，1749无。pp. –。[13] S. Kim、S. Kim、S. Hong、S. Kim、D. Han、J. Choi 和 H.-J. Yoo，“C-DNN：采用异构卷积神经网络/脉冲神经网络核心架构的能效型互补深度神经网络处理器”，IEEE固态电路期刊，第59卷，第1期，第157–172页，2024年。[14] H. Chu、Y. Yan、L. Gan、H. Jia、L. Qian、Y. Huan、L. Zheng 和 Z. Zou，“采用脉冲驱动脉冲神经网络处理器的可穿戴心电图分类神经形态处理系统”，IEEE生物医学电路与系统汇刊，第16卷，第4期，第511–523页，2022年。[15] Y. Liu、Z. Qin 和 G. Y. Li，“面向无线边缘智能的能效型分布式脉冲神经网络”，IEEE无线通信汇刊，第23卷，第9期，第10 683–10 697页，2024年。[16] Z. Yan、Z. Bai、K. Tang 和 W.-F. Wong，“重新审视脉冲神经网络的能效”，arXiv预印本 arXiv:2409.08290，2024年。[17] M. Li、J. Ding、Z. Yue、Y. Li、Z. Luo、Z. Fan、W. Zhao、H. Min、F. Liu、X. Yang、X. Liang 和 L. Jiang，“用于深层脉冲神经网络训练的高能效多核神经形态架构”，arXiv预印本 arXiv:2412.05302，2024年。[18] G. Zhang、L. Feng、F. Zhou、Z. Yang、Q. Zhang、A. Saleh、P. K. Donta 和 C. K. Dehury，“智能边缘计算中的脉冲神经网络”，IEEE消费电子杂志，第14卷，第4期，第66–75页，2025年。[19] G. Li、L. Deng、H. Tang、G. Pan、Y. Tian、K. Roy 和 W. Maass，“类脑计算：系统综述与未来趋势”，《IEEE 会刊》，第112卷，第6期，第544–584页，2024年。[20] S. M. Meyer、P. Weidel、P. Plank、L. Campos-Macias、S. B. Shreshta、P. Stratmann、J. Timcheck 和 M. Richter，“用于高效流式序列处理的Loihi 2对角结构状态空间模型”，载于2025年神经启发计算元件（NICE），2025年，第1–9页。[21] S. Varadarajulu、M. O. K. Mendonca、G. Eappen、J. Querol 和 S. Chatzinotas，“采用时空注意力卷积自编码器和Akida BrainChip脉冲神经网络的增强型5G NTN解调器”，载于第41届国际通信卫星系统会议（ICSSC 2024），第2024卷，2024年，第99–104页。[22] C. D. Schuman、S. R. Kulkarni、M. Parsa、J. P. Mitchell、P. Date 和 B. Kay，“神经形态计算算法与应用的机遇”，《自然计算科学》，第2卷，19，2022年1，10无。pp. –。[23] J. K. Eshraghian、M. Ward、E. Neftci 等，“借鉴深度学习经验训练脉冲神经网络”，《IEEE 会刊》，第111卷，第9期，第1016–1054页，2023年。[24] C. Zhou、H. Zhang、L. Yu、Y. Ye、Z. Zhou、L. Huang、Z. Ma、X. Fan、H. Zhou 和 Y. Tian，“直接训练高性能深层脉冲神经网络：理论与方法综述”，《神经科学前沿》，第18卷，文章号1383844，2024年。[25] S. GHOSH-DASTIDAR 和 H. ADELI，“脉冲神经网络”，《国际神经系统杂志》，第19卷，308，2009年4，295无。pp. –。[26] S. Shen、R. Zhang、C. Wang、R. Huang、A. Tuerhong、Q. Guo、Z. Lu、J. Zhang 和 L. Leng，“进化脉冲神经网络：综述”，《膜计算杂志》，第1–12页，2024年。[27] M. Xiao、Q. Meng、Z. Zhang、D. He 和 Z. Lin，“脉冲神经网络的在线时间训练”，载于《神经信息处理系统进展》，S. Koyejo、S. Mohamed、A. Agarwal、D. Belgrave、K. Cho 和 A. Oh 编，第35卷。Curran Associates, Inc.，2022年，第20 717–20 730页。[28] T. Bu、W. Fang、J. Ding、P. Dai、Z. Yu 和 T. Huang，“面向高准确率和超低延迟脉冲神经网络的最优人工神经网络-脉冲神经网络转换”，arXiv预印本 arXiv:2303.04347，2023年。[29] S. Hwang、S. Lee、D. Park、D. Lee 和 J. Kung，“Spikedattention：无需训练且完全由脉冲驱动的Transformer到脉冲神经网络转换，采用面向赢家的脉冲偏移实现softmax运算”，载于《神经信息处理系统进展》，A. Globerson、L. Mackey、D. 第37卷。Curran Associates, Inc.。2024年。pp. 67721–67745。[30] R.-J. Zhu、Q. Zhao、G. Li 和 J. K. Eshraghian，《SpikeGPT：基于脉冲神经网络的生成式预训练语言模型》，arXiv预印本 arXiv:2302.13939，2023年。[31] Z. Zhou、K. Che、W. Fang、K. Tian、Y. Zhu、S. Yan、Y. Tian 和 L. Yuan，《Spikformer V2：凭借一张脉冲神经网络入场券跻身 ImageNet 高准确率俱乐部》，arXiv预印本 arXiv:2401.02020，2024年。[32] S. Snyder、S. R. Risbud 和 M. Parsa，《Lava 中的神经形态贝叶斯优化》，载于《2023年国际神经形态系统会议（ICONS）论文集》。美国纽约州纽约：美国计算机协会，2023年，页1–5。[33] J. E. Pedersen、S. Abreu、M. Jobst、G. Lenz、V. Fra、F. C. Bauer、D. R. Muir 等，《神经形态中间表示：用于互操作类脑计算的统一指令集》，《自然·通讯》，第15卷，第1期，文章编号8122，2024年。[34] D. Rasmussen，《Nengodl：结合深度学习与神经形态建模方法》，《神经信息学》，第17卷，628，2019 4, 611 no. pp. – 。[35] A. Carpegna、A. Savino 和 S. Di Carlo，《Spiker+：用于生成高效脉冲神经网络 FPGA 加速器、以在边缘侧进行推理的框架》，《IEEE 计算新兴主题汇刊》，2024年，在线优先发表。[36] K. R. KM、T. R. Babu、E. Albalawi、M. TR、S. B. Khan 和 O. Saidani，《面向工业5.0中6G超低延迟的脉冲神经动力学与联邦路由》，《传感器杂志》，第2025卷，文章编号5491672，2025年。[37] H. Huang 和 Y. Liu，《注意力感知神经形态语义通信》，载于《2024年 IEEE 第34届信号处理机器学习国际研讨会（MLSP）》。IEEE，2024年，页1–6。[38] M. Wang、J. Li、M. Ma 和 X. Fan，《SNN-SC：面向协同智能的脉冲语义通信框架》，《IEEE车辆技术汇刊》，第73卷，12 906，2024 9, 12 891 no. pp. – 。[39] K. Svoboda 和 T. Adegbija，《脉冲神经网络架构搜索：综述》，arXiv预印本 arXiv:2510.14235，2025年。[40] H. Shao、J. Lu、M. Wang 和 Z. Wang，《采用硬件—算法协同优化的高效 Transformer 训练加速器》，《IEEE超大规模集成（VLSI）系统汇刊》，第31卷，第11期，页1788–1801，2023年。[41] I. Aliyev、K. Svoboda、T. Adegbija 和 J.-M. Fellous，《面向脉冲神经网络的稀疏度感知软硬件协同设计：综述》，载于《2024年 IEEE 第17届嵌入式多核/众核片上系统国际研讨会（MCSoC）》。IEEE，2024年，页413–420。[42] W. Maass，《脉冲神经元网络：第三代神经网络模型》，《神经网络》，第10卷，第9期，页1659–1671，1997年。[43] W. Maass 和 H. Markram，《论脉冲神经元电路的计算能力》，《计算机与系统科学杂志》，第69卷，第4期，页593–616，2004年。[44] E. M. Izhikevich，《皮层脉冲神经元应采用哪种模型？》，《IEEE神经网络汇刊》，第15卷，1070，2004 5, 1063 no. pp. – 。[45] A. L. Hodgkin 和 A. F. Huxley，《对膜电流的定量描述及其在神经传导与兴奋中的应用》，《生理学杂志》，第117卷，第4期，页500–544，1952年。[46] H. Markram、E. Muller、S. Ramaswamy、M. W. Reimann、M. Abdellah、C. A. Sanchez、A. Ailamaki 等，《新皮层微电路的重建与模拟》，《细胞》，第163卷，第2期，页456–492，2015年。[47] P. Poirazi 和 A. Papoutsi，《通过计算模型揭示树突功能》，《自然评论·神经科学》，第21卷，第6期，页303–321，2020年。[48] W. Gerstner 和 W. M. Kistler，《脉冲神经元模型：单个神经元、群体与可塑性》。剑桥大学出版社，2002年。[49] C. Teeter、R. Iyer、V. Menon、N. Gouwens、D. Feng、J. Berg、A. Szafer 等，《广义漏积分发放模型可对多种神经元类型进行分类》，《自然·通讯》，第9卷，第1期，文章编号709，2018年。[50] N. Skatchkovsky、H. Jang 和 O. Simeone，《面向低功耗边缘智能的脉冲神经网络联邦神经形态学习》，载于《ICASSP 2020—2020年 IEEE 国际声学、语音与信号处理会议（ICASSP）》。IEEE，2020年，页8524–8528。[51] J. L. Lobo、J. Del Ser、A. Bifet 和 N. Kasabov，《脉冲神经网络与在线学习：概述及展望》，《神经网络》，第121卷，页88–100，2020年。[52] E. M. Izhikevich，《脉冲神经元的简单模型》，《IEEE神经网络汇刊》，卷 14，期 6，页 1569–1572，2003年。

[53] R. Brette 和 W. Gerstner，“自适应指数积分发放模型：对神经元活动的有效描述”，《神经生理学杂志》，卷 94，期 5，页 3637–3642，2005年。[54] R. Naud、N. Marcille、C. Clopath 和 W. Gerstner，“自适应指数积分发放模型中的放电模式”，《生物控制论》，卷 99，期 4-5，页 335–347，2008年。[55] R. Jolivet、R. Kobayashi、A. Rauch、R. Naud、S. Shinomoto 和 W. Gerstner，“简单神经元模型的定量评估基准测试”，《神经科学方法杂志》，卷 169，期 2，页 417–424，2008年。[56] C. Mead 和 M. Ismail 编，《神经系统的模拟 VLSI 实现》，克鲁维尔工程与计算机科学国际丛书。马萨诸塞州波士顿：施普林格，1989年，卷 80。[57] C. Mead，“神经形态电子系统”，《IEEE 会刊》，卷 78，期 10，页 1629–1636，1990年。[58] G. Indiveri、B. Linares-Barranco、T. J. Hamilton、A. van Schaik、R. Etienne-Cummings、T. Delbruck、S.-C. Liu 等，“神经形态硅神经元电路”，《神经科学前沿》，卷 5，页 73，2011年。[59] H. Markram、J. L¨ubke、M. Frotscher 和 B. Sakmann，“突触后动作电位与兴奋性突触后电位的同时发生对突触效能的调节”，《科学》，卷 275，期 5297，页 213–215，1997年。[60] S. Song、K. D. Miller 和 L. F. Abbott，“通过脉冲时序依赖性突触可塑性实现竞争性赫布学习”，《自然·神经科学》，卷 3，期 9，页 919–926，2000年。[61] G.-q. Bi 和 M.-m. Poo，“培养的海马神经元中的突触修饰：对脉冲时序、突触强度和突触后细胞类型的依赖性”，《神经科学》，卷 18，期 24，页 10 464–10 472，1998年。[62] E. O. Neftci、H. Mostafa 和 F. Zenke，“脉冲神经网络中的替代梯度学习：将基于梯度的优化的优势引入脉冲神经网络”，《IEEE 信号处理杂志》，卷 36，期 6，页 51–63，2019年。[63] B. A. Richards、T. P. Lillicrap、P. Beaudoin、Y. Bengio、R. Bogacz、A. Christensen、C. Clopath 等，“神经科学的深度学习框架”，《自然·神经科学》，卷 22，2019年，期 11，页 1761–1770。[64] A. Tavanaei、M. Ghodrati、S. R. Kheradpisheh、T. Masquelier 和 A. Maida，“脉冲神经网络中的深度学习”，《神经网络》，卷 111，页 47–63，2019年。[65] K. Roy、A. Jaiswal 和 P. Panda，“借助神经形态计算迈向基于脉冲的机器智能”，《自然》，卷 575，期 7784，页 607–617，2019年。[66] Y. Wu、L. Deng、G. Li、J. Zhu 和 L. Shi，“用于训练高性能脉冲神经网络的时空反向传播”，《神经科学前沿》，卷 12，页 331，2018年。[67] Y. Li、Y. Guo、S. Zhang、S. Deng、Y. Hai 和 S. Gu，“可微脉冲：重新思考训练脉冲神经网络的梯度下降”，载于《神经信息处理系统进展》，卷 34。Curran Associates, Inc.，2021年，页 23 426–23 439。[68] G. Bellec、F. Scherr、A. Subramoney、E. Hajek、D. Salaj、R. Legenstein 和 W. Maass，“脉冲神经元循环网络的学习困境解决方案”，《自然·通讯》，卷 11，期 1，页 3625，2020年。[69] F. Zenke 和 S. Ganguli，“SuperSpike：多层脉冲神经网络中的监督学习”，《神经计算》，卷 30，期 6，页 1514–1541，2018年。[70] S. B. Shrestha 和 G. Orchard，“SLAYER：时间中的脉冲层误差重分配”，载于《神经信息处理系统进展》，卷 31。Curran Associates, Inc.，2018年，页 1412–1421。[71] P. A. Merolla、J. V. Arthur、R. Alvarez-Icaza、A. S. Cassidy、J. Sawada、F. Akopyan、B. L. Jackson 等，“具有可扩展通信网络和接口的百万脉冲神经元集成电路”，《科学》，卷 345，期 6197，页 668–673，2014年。[72] M. Davies、N. Srinivasa、T.-H. Lin、G. Chinya、Y. Cao、S. H. Choday、G. Dimou 等，“Loihi：具有片上学习功能的神经形态众核处理器”，《IEEE Micro》，卷 38，期 1，页 82–99，2018年。[73] J. Pei、L. Deng、S. Song、M. Zhao、Y. Zhang、S. Wu、G. Wang、Z. Zou、Z. Wu、W. He 等，“借助混合式天机芯片架构迈向通用人工智能”，《自然》，卷 572，2019年，期 7767，页 206–211。[74] S. B. Furber、F. Galluppi、S. Temple 和 L. A. Plana，“SpiNNaker 项目”，《IEEE 会刊》，卷 102，期 5，页 652–665，2014年。[75] K. Yamazaki、V.-K. Vo-Ho、D. Bulsara 和 N. Le，“脉冲神经网络及其应用：综述”，《脑科学》，卷 12，期 7，页 863，2022年。

[76] M. Karamimanesh、E. Abiri、M. Shahsavari、K. Hassanli、A. van Schaik 和 J. Eshraghian，《基于现场可编程门阵列的脉冲神经网络：方法与最新进展综述》，《神经网络》，第 107256 页，2025年。[77] C. Iaboni 和 P. Abichandani，《用于目标检测的事件驱动脉冲神经网络：数据集、架构、学习规则与实现综述》，《IEEE Access》，第 12 卷，第 180 532–180 596 页，2024年。[78] R. Brette、M. Rudolph、T. Carnevale、M. Hines、D. Beeman、J. M. Bower、M. Diesmann、A. Morrison、P. H. Goodman、F. C. Harris 等，《脉冲神经元网络仿真：工具与策略综述》，《计算神经科学杂志》，第 23 卷，第 3 期，第 349–398 页，2007年，探讨了基于电导的模型所具有的数值刚性和高昂计算成本。[79] S. Thorpe、A. Delorme 和 R. Van Rullen，《基于脉冲的快速处理策略》，《神经网络》，第 14 卷，725，2001 6 7，715 期 - 页 – 。[80] J. Gautrais 和 S. Thorpe，《频率编码与时间顺序编码：理论方法》，《生物系统》，第 48 卷，第 1-3 期，第 57–65 页，1998年。[81] H. Mostafa，《基于脉冲神经网络时间编码的监督学习》，《IEEE神经网络与学习系统汇刊》，第 29 卷，第 7 期，第 3227–3235 页，2018年。[82] D. Auge、J. Hille、E. Mueller 和 A. Knoll，《脉冲神经网络信号处理编码技术综述》，《神经处理快报》，第 53 卷，第 6 期，第 4693–4710 页，2021年。[83] K. Malcolm 和 J. Casco-Rodriguez，《脉冲神经网络综合综述：解释、优化、效率与最佳实践》，arXiv 预印本 arXiv:2303.10780，2023年。[84] Y. Liang、W. Wei、A. Belatreche、H. Cao、Z. Zhou、S. Wang、M. Zhang 和 Y. Yang，《迈向精确的二值脉冲神经网络：采用自适应梯度调制机制进行学习》，载于《AAAI人工智能会议论文集》，第 39 卷，第 2 期，2025年，第 1402–1410 页。[85] W. Guo、M. E. Fouda、A. M. Eltawil 和 K. N. Salama，《脉冲神经网络中的神经编码：稳健神经形态系统的比较研究》，《神经科学前沿》，第 15 卷，638474，2021年，第 页。[86] S. H. Jo、T. Chang、I. Ebong、B. B. Bhadviya、P. Mazumder 和 W. Lu，《作为神经形态系统突触的纳米级忆阻器器件》，《纳米快报》，第 10 卷，第 4 期，第 1297–1301 页，2010年。[87] C. Frenkel，《自底向上与自顶向下的神经形态处理器设计：揭示通向嵌入式认知的道路》，博士学位论文，比利时鲁汶天主教大学，˙新鲁汶，比利时，2020年。[88] J. Gygax 和 F. Zenke，《阐明脉冲神经网络替代梯度学习的理论基础》，《神经计算》，第 37 卷，第 5 期，第 886–925 页，2025年。[89] B. Huo、F. Li、S. Peng、H. Chen、S. Xin 和 H. Wang，《基于生物合理性的脉冲神经网络学习算法与网络研究》，《IEEE Access》，第 13 卷，第 95 243–95 256 页，2025年。[90] W. Wei、M. Zhang、J. Zhang、A. Belatreche、J. Wu、Z. Xu、X. Qiu、H. Chen、Y. Yang 和 H. Li，《脉冲神经网络的事件驱动学习》，arXiv 预印本 arXiv:2403.00270，2024年。[91] H. Esmaeilzadeh、E. Blem、R. St. Amant、K. Sankaralingam 和 D. Burger，《暗硅与多核扩展的终结》，载于《第 38 届年度国际计算机体系结构研讨会论文集》，2011年，第 365–376 页。[92] M. B. Taylor，《暗硅有用吗？驾驭即将到来的暗硅末日的四骑士》，载于《第 49 届年度设计自动化大会论文集》，2012年，第 1131–1136 页。[93] N. Hardavellas、M. Ferdman、B. Falsafi 和 A. Ailamaki，《迈向服务器中的暗硅》，《IEEE Micro》，第 31 卷，第 4 期，第 6–15 页，2011年。[94] M. Shafique、S. Garg、J. Henkel 和 D. Marculescu，《暗硅时代的 EDA 挑战：温度、可靠性与变异性视角》，载于《第 51 届年度设计自动化大会论文集》，2014年，第 1–6 页。[95] M. Davies、A. Wild、G. Orchard、Y. Sandamirskaya、G. A. F. Guerra、P. Joshi、P. Plank 和 S. R. Risbud，《借助 Loihi 推进神经形态计算：成果综述与展望》，《IEEE 会刊》，第 109 卷，第 5 期，第 911–934 页，2021年。[96] N. Rathi 和 K. Roy，《Diet-snn：一种采用直接输入编码、泄漏与阈值优化的低延迟脉冲神经网络》，《IEEE神经网络与学习系统汇刊》，第 34 卷，第 6 期，第 3174–3182 页，2023年。[97] G. Gallego、T. Delbr¨uck、G. Orchard、C. Bartolozzi、B. Taba、A. Censi、S. Leutenegger 等，《事件视觉：综述》，《IEEE模式分析与机器智能汇刊》，第 44 卷，第 1 期，第 154–180 页

，2022年。

[98] S. H. Ahmed、J. Finkbeiner 和 E. Neftci，“高效的基于事件的目标检测：一种结合空间注意力和时间注意力的混合神经网络”，载于《IEEE/CVF计算机视觉与模式识别会议论文集》(CVPR)，2025年6月，页 13 970–13 979。[99] M. Horowitz，“1.1 计算能耗问题（以及我们能为此做些什么）”，载于《2014年IEEE国际固态电路会议论文集》（ISSCC）。IEEE，2014年，页 10–14。[100] J. M. Rabaey、A. Chandrakasan 和 B. Nikolic，《数字集成电路》。Prentice Hall，恩格尔伍德克利夫斯，2002年，第 2 卷。[101] D. S. Modha、F. Akopyan、A. Andreopoulos、R. Appuswamy、J. V. Arthur、A. S. Cassidy、P. Datta 等，“能源、空间和时间前沿的神经推理”，《科学》，第 382 卷，335，2023 6668，329 期，页 – 。[102] Y. Wu、L. Deng、G. Li、J. Zhu 和 L. Shi，“脉冲神经网络的直接训练：更快、更大、更优”，《AAAI人工智能会议论文集》，第 33 卷，1318，2019 01，1311 期，页 – 。[103] H. Zheng、Y. Wu、L. Deng、Y. Hu 和 G. Li，“通过直接训练更大型脉冲神经网络实现更深层网络”，载于《AAAI人工智能会议论文集》，第 35 卷，第 12 期，2021年，页 11 062–11 070。[104] J. Ding、Z. Yu、Y. Tian 和 T. Huang，“面向深层脉冲神经网络快速、准确推理的最优人工神经网络到脉冲神经网络的转换”，arXiv 预印本 arXiv:2105.11654，2021年。[105] T. Bu、M. Li 和 Z. Yu，“面向高性能和低功耗应用的人工神经网络到脉冲神经网络转换中的推理规模复杂度”，载于《计算机视觉与模式识别会议论文集》，2025年，页 24 387–24 397。[106] X. Song、A. Song、R. Xiao 和 Y. Sun，“具有线性复杂度的单步脉冲Transformer”，载于第三十三届国际人工智能联合会议论文集，2024年，页 3142–3150。[107] Z. Hao、Z. Yu 和 T. Huang，“更快、更强：人工神经网络到脉冲神经网络的转换遇上并行脉冲计算”，arXiv 预印本 arXiv:2412.13610，2024年。[108] D. Yu、C. Lv、X. Du、L. Jiang、W. Tong、Z. Liao、X. Zheng 和 S. Deng，“ECC-SNN：面向脉冲神经网络的高性价比边云协同”，载于第三十四届国际人工智能联合会议论文集（IJCAI-25），J. Kwok 编。国际人工智能联合会议组织，2025年8月，页 6904–6912。[109] J. Seekings、P. Chandarana、M. Ardakani、M. Mohammadi 和 R. Zand，“迈向混合脉冲神经网络在神经形态及边缘人工智能硬件上的高效部署”，载于2024年国际神经形态系统会议（ICONS）。IEEE，2024年，页 71–77。[110] F. Rivelli、M. Popov、C. S. Kouzinopoulos 和 G. Tang，“用于能效型皮层内神经解码的自适应剪枝脉冲神经网络”，arXiv 预印本 arXiv:2504.11568，2025年。[111] L. Shi、A. Li 和 B. Ward-Cherrier，“最优脉冲脑压缩：改进脉冲神经网络的一次性训练后剪枝与量化”，arXiv 预印本 arXiv:2506.03996，2025年6月。[112] B. Na、J. Mok、S. Park、D. Lee、H. Choe 和 S. Yoon，“Autosnn：迈向能效型脉冲神经网络”，载于国际机器学习会议。PMLR，2022年，页 16 253–16 269。[113] Y. Kim、Y. Li、H. Park、Y. Venkatesha 和 P. Panda，“脉冲神经网络的神经架构搜索”，载于欧洲计算机视觉会议。施普林格，2022年，页 36–56。[114] K. Che、L. Leng、K. Zhang、J. Zhang 和 J. Cheng，“用于脉冲神经网络的可微分分层与替代梯度搜索”，载于《神经信息处理系统进展》（NeurIPS），第 35 卷，2022年，页 24 975–24 990。[115] J. Yik、Z. Ahmed 和 A. G. Anderson，“面向硬件感知脉冲神经网络的神经进化学习”，《IEEE模式分析与机器智能汇刊》（TPAMI），第 46 卷，第 5 期，页 2845–2858，2024年。[116] S. Li、L. Chen、Y. Zhang、Z. Zhang、A. Du、B. Pan、Z. Wang、L. Wen 和 W. Zhao，“QUEST：面向多状态神经形态器件的量化能耗感知脉冲神经网络训练框架”，arXiv 预印本 arXiv:2504.00679，2025年。[117] R. V. W. Putra、A. Marchisio 和 M. Shafique，“SNN4Agents：用于开发能效型具身脉冲神经网络自主智能体的框架”，《机器人与人工智能前沿》，第 11 卷，页 1401677，2024年。[118] 英特尔实验室，“英特尔打造全球最大的神经形态系统，助力实现更可持续的人工智能”，2024年，新闻稿。网址：https://www.intel.com/。[119] D. Ma、X. Jin、S. Sun、Y. Li、X. Wu、Y. Hu、F. Yang、H. Tang 和 G. Pan，“Darwin3：一款采用新型指令集架构的大规模神经形态芯片以及片上学习”，《国家科学评论》，第11卷，第5期，页码 nwae102，2024年。[120] BrainChip Holdings Ltd，《Akida 神经处理器知识产权技术参考手册》，2024年，网址：https://brainchip.com/akida-neural-processor-soc/。[121] C. Mayr、S. Hoeppner 和 S. Furber，《SpiNNaker 2：用于脑模拟和机器学习的千万核心处理器系统——主题演讲》，载于《通信处理器架构》2017与2018。IOS Press，2019年，页277–280。[122] G. Orchard、E. P. Frady、D. B. D. Rubin、S. Sanborn、S. B. Shrestha、F. T. Sommer 和 M. Davies，《使用 Loihi 2 进行高效神经形态信号处理》，载于2021年IEEE信号处理系统研讨会（SiPS）。IEEE，2021年，页254–259。[123] F. Akopyan、J. Sawada、A. Cassidy、R. Alvarez-Icaza、J. Arthur、P. Merolla、N. Imam、Y. Nakamura、P. Datta、G.-J. Nam 等，《TrueNorth：65毫瓦、100万个神经元可编程神经突触芯片的设计与工具流程》，《IEEE 集成电路与系统计算机辅助设计汇刊》，第34卷，第10期，页1537–1557，2015年。[124] P. Yao、H. Wu、B. Gao、J. Tang、Q. Zhang、W. Zhang、J. J. Yang 和 H. Qian，《完全由硬件实现的忆阻器卷积神经网络》，《自然》，第577卷，第7792期，页641–646，2020年。[125] D. Zhao、Y. Li、Y. Zeng、J. Wang 和 Q. Zhang，《脉冲 CapsNet：一种在胶囊间采用生物学上合理路由规则的脉冲神经网络》，《信息科学》，第610卷，页1–13，2022年。[126] D. V. Christensen、R. Dittmann、B. Linares-Barranco、A. Sebastian、M. Le Gallo、A. Redaelli、S. Slesazeck、T. Mikolajick、S. Spiga、S. Menzel 等，《2022年神经形态计算与工程路线图》，《神经形态计算与工程》，第2卷，第2期，页022501，2022年。[127] D. Ma、J. Shen、Z. Gu、M. Zhang、X. Zhu、Q. Xu、Q. Shen 和 G. Pan，《Darwin：基于脉冲神经网络的神经形态硬件协处理器》，《系统架构期刊》，第77卷，页43–51，2017年。[128] C. Pehle、S. Billaudelle、B. Cramer、J. Kaiser、K. Schreiber、Y. Stradmann、J. Weis、A. Leibfried、E. M¨uller 和 J. Schemmel，《BrainScaleS-2：具有混合可塑性的加速神经形态系统》，《神经科学前沿》，第16卷，页795876，2022年。[129] L. Chen、X. Xiong 和 J. Liu，《基于脉冲神经网络的智能芯片设计研究综述》，《IEEE Access》，第10卷，页89 663–89 686，2022年。[130] A. Li、Y. Liu 和 R. Kastner，《S2N2：一种用于流式脉冲神经网络的现场可编程门阵列加速器》，载于2021年美国计算机协会/美国计算机学会数字设计特别兴趣组现场可编程门阵列国际研讨会论文集，2021年，页1–11。[131] S. Panchapakesan、Z. Fang 和 J. Li，《SyncNN：在现场可编程门阵列上评估并加速脉冲神经网络》，载于2021年第31届国际现场可编程逻辑与应用会议（FPL），2021年，页286–293。[132] L. Kang 和 X. Yang，《一种用于 DVS 手势识别的24.3 μj/图像脉冲神经网络加速器，采用 WS-LOS 数据流和稀疏方法》，《IEEE 电路与系统汇刊第二辑：简报》，第70卷，第10期，页3802–3806，2023年。[133] J. Li 等，《FireFly v2：通过时空现场可编程门阵列加速器推进对高性能脉冲神经网络的硬件支持》，《IEEE 集成电路与系统计算机辅助设计汇刊》（TCAD），第43卷，第9期，页1–14，2024年。[134] B. Xu 等，《SpikeX：面向稀疏脉冲神经网络的加速器架构与网络—硬件协同优化探索》，《IEEE 集成电路与系统计算机辅助设计汇刊》（TCAD），2025年，（在线优先发表/arXiv:2505.12292）。[135] A. Balaji、A. Das、Y. Wu、K. Huynh、F. G. Dell’Anna、G. Indiveri、J. L. Krichmar、N. D. Dutt、S. Schaafsma 和 F. Catthoor，《将脉冲神经网络映射到神经形态硬件》，《IEEE超大规模集成（VLSI）系统汇刊》，第28卷，86，2019 1，76第。页–。[136] G. Urgese、F. Barchi、E. Macii 和 A. Acquaviva，《优化密集互连众核神经形态平台上脉冲神经网络仿真的网络流量》，《IEEE 计算新兴主题汇刊》，第6卷，第3期，页317–329，2016年。[137] J. Xue、L. Xie、F. Chen、L. Wu、Q. Tian、Y. Zhou、R. Ying 和 P. Liu，《EdgeMap：面向边缘计算中脉冲神经网络的优化映射工具链》，《传感器》，第23卷，第14期，页6548，2023年。[138] C. Xiao、J. Chen 和 L. Wang，《面向边缘人工智能的脉冲神经网络到神经形态硬件的最优映射》，《传感器》，第22卷，第19期，页7248，2022年。

[139] W. G. Gomez、A. Pignata、R. Pignari、V. Fra、E. Macii 和 G. Urgese，《迈向 Lava-Loihi 神经形态生态系统微基准测试的初步探索》，载于2023年IEEE第16届嵌入式多核/众核片上系统国际研讨会（MCSoC），2023年，页462–469。[140] T. Bekolay、J. Bergstra、E. Hunsberger 等，《Nengo：用于构建大规模功能性脑模型的 Python 工具》，《神经信息学前沿》，第7卷，页48，2014年。[141] H. Hazan、D. J. Saunders、H. Khan 等，《BindsNET：面向机器学习的脉冲神经网络 Python 库》，《神经信息学前沿》，第12卷，页89，2018年。[142] M. Stimberg、R. Brette 和 D. F. Goodman，《Brian 2：直观且高效的神经模拟器》，eLife，第8卷，页e47314，2019年。[143] S. R. Kulkarni、M. Parsa、J. P. Mitchell 和 C. D. Schuman，《神经形态模拟器和脉冲神经网络模拟器的性能基准测试》，《神经计算》，第447卷，页145–160，2021年。[144] D. L. Manna、A. Vicente-Sola、P. Kirkland、T. J. Bihl 和 G. Di Caterina，《脉冲神经网络框架：面向数据科学的软件综述及 SpykeTorch 的扩展》，载于《神经网络工程应用国际会议》。施普林格，2023年，页227–238。[145] S. Al Abdul Wahid、A. Asad 和 F. Mohammadi，《用于运行人工智能算法的神经形态架构综述》，《电子学》，第13卷，第15期，页2963，2024年。[146] Y.-C. Cheng、W.-X. Hu、Y.-L. He 和 J. Z. Huang，《脉冲神经网络神经形态训练框架的综合多模态基准测试》，《人工智能工程应用》，第159卷，页111543，2025年。[147] H. Gupta、A. Vahid Dastjerdi、S. K. Ghosh 和 R. Buyya，《iFogSim：用于物联网、边缘计算和雾计算环境中资源管理技术建模与仿真的工具包》，《软件：实践与经验》，第47卷，1296，2017 9，第1275页。–。[148] S. Tuli、R. Mahmud、S. Tuli 和 R. Buyya，《FogBus：一种基于区块链的轻量级边缘与雾计算框架》，《系统与软件期刊》，第154卷，页22–36，2019年。[149] D. Yu、X. Du、L. Jiang 和 Y. Li，《FedLEC：一种在标签偏斜条件下结合脉冲神经网络的有效联邦学习算法》，arXiv 预印本 arXiv:2404.14488，2024年。[150] M. Abbasihafshejani、A. Maiti 和 M. Jadliwala，《纵向联邦学习中的脉冲神经网络：性能权衡》，载于NOMS 2025—2025年IEEE网络运维与管理研讨会。IEEE，2025年，页1–5。[151] D. Aksu、J. M. del Rincon 和 I. Alouani，《采用脉冲神经网络的联邦学习中的隐私》，arXiv 预印本 arXiv:2511 21181，2025年。[152] M. V. Nguyen、L. Zhao、B. Deng 和 S. Wu，《采用脉冲神经网络的稀疏化联邦学习：抵御拜占庭攻击并降低通信流量》，载于2025年IEEE第101届车辆技术会议（VTC2025春季会议）。IEEE，2025年，页1–5。[153] Z. Li、Y. Ma、J. Zhou 和 P. Zhou，《Spiking-NeRF：面向能效神经渲染的脉冲神经网络》，《新兴计算技术系统期刊》，第20卷，第3期，2024年8月。[154] L. Shi、A. Li 和 B. Ward-Cherrier，《脉冲脑压缩：探索脉冲神经网络的一次性训练后剪枝与量化》，载于OPT2025：第17届机器学习优化年度研讨会，2025年，arXiv:2506.03996。[155] S. Hussaini、M. Milford 和 T. Fischer，《脉冲神经网络在视觉地点识别中的应用》，《IEEE机器人学汇刊》，第41卷，页518–537，2025年。[156] N. Zheng 和 H. Xia，《通过事件相机和骨架数据融合实现由脉冲神经网络驱动的多模态人体动作识别》，arXiv 预印本 arXiv:2502.13385，2025年。[157] R. V. W. Putra、P. Wickramasinghe 和 M. Shafique，《在面向边缘人工智能的商用神经形态处理器上，借助片上学习实现脉冲神经网络的高效处理》，arXiv 预印本 arXiv:2504.00957，2025年。[158] M. Hizem、L. Bousbia、Y. Ben Dhiab、M. O.-E. Aoueileyine 和 R. Bouallegue，《面向医疗物联网应用的边缘设备可靠心电图异常检测》，《传感器》，第25卷，第8期，页2496，2025年。[159] D. Zhou、W. Chen、K. Chen 和 B. Mi，《面向工业应用的快速、准确脉冲神经网络模型强化》，《电子学》，第12卷，第18期，页3845，2023年。[160] H. Wang、Y.-F. Li 和 K. Gryllias，《用于工业故障诊断的类脑脉冲神经网络：综述、挑战与机遇》，arXiv 预印本 arXiv:2401.02429，2023年。[161] L. Zhu、S. Dong、T. Huang 和 Y. Tian，《利用中央凹式脉冲相机进行超高时间分辨率视觉重建》，IEEE模式分析与机器智能汇刊（TPAMI），第44卷第11期，第7785–7799页，2022年。[162] S. Dong、L. Zhu、D. Duan、T. Huang 和 Y. Tian，《脉冲相机：综述》，arXiv预印本 arXiv:2401.13490，2024年。[163] Y. Hu、Q. Zheng 和 G. Pan，《Bitsnns：重新审视能效型脉冲神经网络》，IEEE认知与发展系统汇刊，第16卷第5期，第1736–1747页，2024年。[164] X. Shi、J. Ding、Z. Hao 和 Z. Yu，《迈向能效型脉冲神经网络：一种非结构化剪枝框架》，载于《第十二届国际学习表征会议》，2024年。[165] V. Sudevan、F. Zayer、R. Kausar、S. Javed、H. Karki、G. De Masi 和 J. Dias，《利用卷积脉冲神经网络增强水下图像》，arXiv预印本 arXiv:2503.20485，2025年。[166] A. Amir、B. Taba、D. Berg、T. Melano、J. McKinstry、C. Di Nolfo、T. Nayak、A. Andreopoulos、G. Garreau、M. Mendoza 等，《一种低功耗、完全基于事件的手势识别系统》，载于《IEEE计算机视觉与模式识别会议（CVPR）会刊》，2017年，第7243–7252页。[167] H. Sabbella、A. Mukherjee、T. Kandappu、S. Dey、A. Pal、A. Misra 和 D. Ma，《脉冲神经网络在泛在计算中的前景：综述与新见解》，arXiv预印本 arXiv:2506.01737，2025年。[168] P. Blouw、X. Choo、E. Hunsberger 和 C. Eliasmith，《神经形态硬件上的关键词检测效率基准测试》，载于《第七届年度神经启发计算要素研讨会会刊》，2019年，第1–8页。[169] B. Yin、F. Corradi 和 S. M. Boht´e，《利用自适应脉冲循环神经网络实现准确高效的时域分类》，《自然·机器智能》，第3卷第10期，第905–913页，2021年。[170] S. Baek 和 J. Lee，《脉冲神经网络与声音：脉冲神经网络在声音领域的全面综述》，《生物医学工程快报》，第14卷第5期，第981–991页，2024年。[171] S. Y. A. Yarga 和 S. U. Wood，《采用脉冲密度调制微机电系统麦克风的神经形态关键词检测》，arXiv预印本 arXiv:2408.05156，2024年。[172] N. Imam 和 T. A. Cleland，《神经形态嗅觉电路中的快速在线学习与稳健回忆》，《自然·机器智能》，第2卷第3期，第181–191页，2020年。[173] J. Guo、X. Li、X. Li、Z. Liang、J. Cao 和 X. Wei，《基于神经网络的抗漂移气体检测算法》，IEEE仪器与测量汇刊，2024年。[174] A. Momp´o Alepuz、D. Papageorgiou 和 S. Tolu，《脑启发仿生机器人控制：综述》，《神经机器人学前沿》，第18卷，文章编号1395617，2024年。[175] Z. Bing、C. Meschede、F. R¨ohrbein、K. Huang 和 A. C. Knoll，《基于学习启发式脉冲神经网络的机器人控制综述》，《神经机器人学前沿》，第12卷，文章编号35，2018年。[176] H. Choi、J. Park、J. Lee 和 D. Sim，《面向低功耗环境的基于脉冲神经网络的心电图分类方法综述》，《生物医学工程快报》，第14卷，2024年5，917号。第941页–。[177] S. H. Choi，《用于生物医学信号分析的脉冲神经网络》，《生物医学工程快报》，第14卷第5期，第955–966页，2024年。[178] Y. Xing、L. Zhang、Z. Hou、X. Li、Y. Shi、Y. Yuan、F. Zhang、S. Liang、Z. Li 和 L. Yan，《基于脉冲神经网络和注意力机制的准确心电图分类，可在个人便携设备上实时实现》，《电子学》，第11卷第12期，文章编号1889，2022年。[179] A. Alzarooni、E. Iqbal、S. U. Khan、S. Javed、B. Moyo 和 Y. Ab- dulrahman，《工业应用中的异常检测：挑战、解决方案与未来方向综述》，arXiv预印本 arXiv:2501.11310，2025年。[180] S. H. H. DolatAbadi、S. M. H. Hashemi、M. Hosseini 和 M.-A. AliHosseini，《以人工智能驱动的机器视觉革新交通管理：迈向智慧城市》，arXiv预印本 arXiv:2503.02967，2025年。[181] O. K. Pal、M. S. H. Shovon、M. F. Mridha 和 J. Shin，《人工智能赋能的无人飞行器综合综述：趋势、愿景与挑战》，arXiv预印本 arXiv:2310.16360，2023年。[182] A. d. J. Pugliese Viloria、A. Folini、D. Carrion 和 M. A. Brovelli，《智慧城市中的人工智能——应用、障碍与未来方向：综述》，《智慧城市》，第7卷第3期，第1346–1389页，2024年。[183] A. d. J. Pugliese Viloria、A. Folini、D. Carrion 和 M. A. Brovelli，《利用机器学习与深度学习进行灾害易发性制图：文献综述》，《遥感》，第16卷第18期，文章编号3374，2024年。

[184] J. Jiang、L. Wang、R. Jiang、J. Fan 和 R. Yan，《利用膜电位动力学实现脉冲神经网络的自适应梯度学习》，arXiv预印本 arXiv:2505.11863，2025年。[185] K. Yu、T. Zhang、Q. Xu、G. Pan 和 H. Wang，《TS-SNN：用于脉冲神经网络的时间移位模块》，arXiv预印本 arXiv:2505.04165，2025年。[186] W. Maass，《要不要发放脉冲：这就是问题所在》，《IEEE 会刊》，第103卷第12期，第2219–2224页，2015年。[187] M. Schmolli、M. Baronig、R. Legenstein 和 O. Ozdenizci，《具有稀疏连接的抗对抗攻击脉冲神经网络》，载于《简约与学习会议（CPAL）会刊》，《机器学习研究会刊》系列，B. Chen、S. Liu、M. Pilanci、W. Su、J. Sulam、Y. Wang 和 Z. Zhu 编，第280卷，PMLR，2025年3月，第865–883页。[188] C. D. Schuman、T. E. Potok、R. M. Patton、J. D. Birdwell、M. E. Dean、G. S. Rose 和 J. S. Plank，《神经形态计算与硬件中的神经网络综述》，arXiv预印本 arXiv:1705.06963，2017年。[189] D. Le、B. Arig、M. Isik、I. C. Dikmen 和 T. Karadag，《神经形态计算中的内存墙综述》，载于《2025年IEEE第四届计算与机器智能国际会议（ICMI）》。IEEE，2025年，第1–6页。[190] M. Hellenbrand、I. Teck 和 J. L. MacManus-Driscoll，《新兴非易失性存储技术在工业中的进展》，《MRS通讯》，第14卷第6期，第1099–1112页，2024年。[191] J. Pan、Z. Wang、B. Zhao、J. Yin、P. Guo、Y. Yang 和 T.-L. Ren，《基于二维材料的非易失性存储器件最新进展》，《芯片》，第3卷第4期，第271–295页，2024年。[192] B. Vogginger、A. Rostami、V. Jain、S. Arfa、A. Hantsch、D. Kappel、M. Sch¨afer、U. Faltings、H. A. Gonzalez、C. Liu 等，《面向可持续人工智能数据中心的神经形态硬件》，arXiv预印本 arXiv:2402.02521，2024年。[193] L. Liang、X. Hu、L. Deng、Y. Wu、G. Li、Y. Ding、P. Li 和 Y. Xie，《利用兼容脉冲的梯度探索脉冲神经网络中的对抗攻击》，《IEEE神经网络与学习系统汇刊》（TNNLS），第34卷第5期，第2569–2583页，2023年。[194] B. K. Baghbaderani 和 H. Shahinzadeh，《脉冲神经网络中的对抗攻击与防御机制：全面综述》，载于《2025年第29届伊朗计算机学会国际计算机会议（CSICC）》。IEEE，2025年，第01–07页。[195] A. Ghosh 和 S. Ghosh，《特邀论文：迈向安全的传感器内智能：脉冲神经网络中的威胁与防御》，载于《2025年IEEE/ACM计算机辅助设计国际会议（ICCAD）》，2025年，第1–7页。[196] Q. Huang、Z. Yang、K. Ni、M. Imani、C. Zhuo、S. Yu 和 F. Koushanfar，《通过远程现场可编程门阵列侧信道分析推断神经网络结构》，IEEE信息取证与安全汇刊（TIFS），第16卷，第4377–4388页，2021年。[197] N. Rathi、I. Chakraborty、A. Kosta、A. Sengupta、A. Ankit、P. Panda 和 K. Roy，《探索基于脉冲神经网络的神经形态计算：从算法到硬件》，《ACM计算调查》，第55卷第12期，2023年3月。[198] A. Aydin、M. Gehrig、D. Gehrig 和 D. Scaramuzza，《面向低功耗、低延迟视觉感知的混合人工神经网络-脉冲神经网络架构》，载于《IEEE/CVF计算机视觉与模式识别会议（CVPR）会刊》，2024年，第5701–5711页。[199] G. S. Syed、M. Le Gallo 和 A. Sebastian，《用于存内计算的相变存储器》，《化学评论》，2025年。[200] X. Wang、Y. Zhu、Z. Zhou、X. Chen 和 X. Jia，《面向脑启发式感知与计算的基于忆阻器的脉冲神经形态系统》，《纳米材料》，第15卷第14期，文章编号1130，2025年。[201] S. Yu，《利用新兴非易失性存储器实现类脑计算》，《IEEE 会刊》，第106卷第2期，第260–285页，2018年。[202] J. Seekings、M. Ardakani、P. Chandarana、A. Eslami、M. Mohammadi 和 R. Zand，《混合神经形态系统的算法与硬件一体化设计》，《npj非常规计算》，第2卷第1期，文章编号20，2025年8月。[203] E. M¨uller、E. Arnold、O. Breitwieser、M. Czierlinski、A. Emmel、J. Kaiser、C. Mauch、S. Schmitt、P. Spilger、R. Stock 等，《一种在加速神经形态硬件上进行建模的可扩展方法》，《神经科学前沿》，第16卷，文章编号884128，2022年。[204] G. Wang、S. Ma、Y. Wu、J. Pei、R. Zhao 和 L. Shi，《在跨范式神经形态芯片上端到端实现多种混合神经网络》，《神经科学前沿》，第15卷，615279，2021年 p.。[205] Y. Xue、Y. Liu、L. Nai 和 J. Huang，《云平台中神经处理单元的硬件辅助虚拟化》，载于《2024年第57届IEEE/ACM微体系结构国际研讨会（MICRO）》。IEEE，2024年，第1–16页。

[206] A. Beloglazov、J. Abawajy 和 R. Buyya，《面向云计算数据中心高效管理的能量感知资源分配启发式算法》，未来一代计算机系统，卷 28，期 5，页 755–768，2012 年。[207] A. Shrestha 、 H. Fang 、 Z. Mei 、 D. P. Rider、Q. Wu 和 Q. Qiu，《神经形态计算综述：模型与硬件》，*IEEE*电路与系统杂志，卷 22，期 2，页 6–35，2022年。[208] E. Zeydan、C. D. Alwis、R. Khan、Y. Turk、A. Aydeger、T. R. Gadekallu 和 M. Liyanage，《面向5G及6G网络的量子技术：应用、机遇与挑战》，*IEEE*通信学会开放期刊，卷 6，页 6383–6420，2025年。[209] L. Barbieri、A. Menina、R. Bassoli 和 F. H. Fitzek，《6G中量子计算的安全性：从技术视角到伦理影响》，*arXiv*预印本 *arXiv:2504.10040*，2025年。[210] C.-X. Wang、X. You、X. Gao、X. Zhu、Z. Li、C. Zhang、H. Wang、Y. Huang、Y. Chen、H. Haas、J. S. Thompson、E. G. Larsson、M. D. Renzo、W. Tong、P. Zhu、X. Shen、H. V. Poor 和 L. Hanzo，《迈向6G之路：愿景、需求、关键技术与测试平台》，*IEEE*通信调查与教程，卷 25，期 2，页 905–974，2023年。[211] B. Narottama、Z. Mohamed 和 S. A¨ıssa，《面向下一代无线通信的量子机器学习：基础与未来发展方向》，*IEEE*通信学会开放期刊，卷 4，页 2204–2224，2023年。[212] X. Wang、Z. Tang、J. Guo、T. Meng、C. Wang、T. Wang 和 W. Jia，《赋能边缘智能：端侧人工智能模型综合综述》，*ACM*计算调查，卷 57，期 9，页 1–39，2025年。