神经拟态的类脑计算研究

顾宗华 潘 纲
浙江大学

关键词：类脑计算 脉冲神经网络 脑机融合

引言

基于冯·诺伊曼体系结构的处理器性能随着摩尔定律的持续而不断提高。2004~2005年，登纳德缩放比例定律(Dennard Scaling)[1]失效，功耗与散热等问题使得芯片设计者开始将目光转向多核和众核。但现代的大型众核芯片会出现暗硅(dark silicon)问题，功耗与散热问题导致众核芯片上的大部分内核不能全部同时开启运行。有研究者[1]认为，功耗问题将终结多核的可伸缩性，单个芯片上的内核个数将不再增加。摩尔定律在持续50年后，由于技术与经济两方面的原因，被普遍认为将于2020年左右走到尽头[7]。随着摩尔定律逐步走向终结，半导体行业在后摩尔定律时代将如何发展？不同领域的研究者有不同的答案。针对这一问题，本文将介绍神经拟态的类脑计算方面的主要思想、现有工作与未来面临的挑战。

主要思想与发展历史

神经拟态的类脑计算的基本思路是将生物神经网络的概念应用于计算机系统设计，针对智能信息处理的特定应用来提高性能与降低功耗。冯·诺伊曼体系结构的基本特征是内存与计算单元分离，优点是软件化、可编程，同一硬件平台可以通过存储不同的软件来实现不同的功能；缺点是存储单元与计算单元之间的通信延迟可能成为性能瓶颈，出现“内存墙”问题。而生物神经网络的基本特征是“内存与计算单元合二为一”。神经元既有计算功能，又有存储功能，从根本上解决了冯·诺伊曼体系架构的“内存墙”问题。人脑是一部极其高效的“计算机”，其特征与优势包括：通过与外界交互自主学习（无须显式编程）、高度容错（容忍大量神经元的死亡而不影响其基本功能）、高度并行性（约10^{11}个神经元）、高度连接性（约10^{15}个突触）、低运算频率（约100Hz）、低通信速度（每秒钟几米）、低功耗[2]（约20瓦）。有研究者[18]提出了一个用于比较人脑与计算机性能的指标——一个大型随机图上“每秒穿越的边的个数”(Traversed Edges Per Second, TEPS)，其基本思路是神经元之间通信的一次脉冲(spike)，类似于在图上穿越一个边。基于TEPS指标，人脑的运算速度是当今最快的超级计算机的30倍。

类脑计算的愿景是实现新一代计算机体系结构，设计出可以与人脑相媲美的、高效节能的通用计算机系统。与电脑相比，人脑在精确数值的计算方面性能较差，但是在智能模糊信息处理方面（例如图像中的物体识别、视频音频理解、自然语言处理等应用）具有独特的优势。IBM的深蓝电脑早在1997年就击败了国际象棋冠军加里·卡斯帕罗夫(Garry Kasparov)，其工作原理是采用并行处理器的暴力搜索。但

[1] 登纳德缩放比例定律，指晶体管的尺寸在不断缩小的同时其功耗密度大致保持不变。
[2] 基于冯·诺伊曼体系结构建造一个与人脑复杂程度相当的计算机，需要将近100兆瓦的功耗。

是直到今天，电脑也无法与人脑在围棋上较量，因为围棋的搜索空间太大，使用暴力搜索不可行，而必须依赖人脑的直觉。随着大数据的兴起，人工智能（尤其是深度学习技术）在近年来得到了迅猛发展，有了长足的进步。例如，2011年，IBM的沃森(Watson)电脑在问答游戏“Jeopardy!”中击败了上届冠军布拉德·鲁特(Brad Rutter)与肯·詹宁斯(Ken Jennings)，赢得了100万美元奖金；2015年，大规模深度神经网络的性能已经可以在ImageNet测试集的分类算法上超越人脑。这表明电脑在一些特定领域已经可以与人脑相媲美。但是在通用性、自主非监督学习等方面，目前的电脑无法超越人脑。

与经典的人工神经网络(Artificial Neural Network, ANN)不同，生物神经网络属于脉冲神经网络(Spiking Neural Network, SNN)：一个神经元接受输入脉冲，导致细胞体(soma)的膜电压升高，当达到特定阈值时，发出一个输出脉冲到轴突(axon)，并通过突触发送神经递质与后续神经元树突(dendrite)上的接受体相结合来改变其膜电压。因此生物神经元之间的通信机制是膜电压升降的脉冲，而非人工神经网络中的数值运算。机器学习算法，尤其是人工神经网络硬件加速方面的工作由来已久，最新的研究进展包括中国科学院计算技术研究所的寒武纪系列神经网络处理器[6]，通过硬件体系结构的创新来大幅度提高性能并降低功耗。脉冲神经网络在响应速度、功耗等方面具有的独特优势，使其成为未来有潜力的发展方向。

本文首先回顾类脑计算的简史：

1. 1989年，加州理工学院的卡弗·米德(Carver Mead)提出了“类脑工程”概念，并撰写了《模拟VLSI与神经系统》(*Analog VLSI and Neural Systems*)一书，采用亚阈值模拟电路来仿真脉冲神经网络，其应用是仿真视网膜。

2. 1990~2003年，摩尔定律持续发展，基于冯·诺伊曼架构的处理器主频与性能持续增长，而类脑计算则沉寂十余年。

3. 2004年前后，单核处理器主频停止增长，设计者开始转向多核，学术界开始寻求冯·诺伊曼架构的替代技术。类脑计算经过十多年的小众研究，开始成为热点。

4. 2004年，斯坦福大学教授夸贝纳·波尔汉(Kwabena Boahen)研制出基于模拟电路的类脑芯片Neurogrid。

5. 2005年，英国曼彻斯特大学开始研制基于ARM的多核超级计算机SpiNNaker。

6. 2005年，欧盟启动FACETS (Fast Analog Computing with Emergent Transient States)项目，研制基于模拟混合信号(Analog Mixed-Signal, AMS)的类脑芯片。

7. 2005年，美国国防部高级研究计划署(DARPA)启动SyNAPSE(Systems of Neuromorphic Adaptive Plastic Scalable Electronics)项目，支持IBM与多家单位联合研发类脑芯片。瑞士洛桑联邦理工学院(EPFL)研究者亨利·马克拉姆(Henry Markram)与IBM合作启动了蓝脑项目，采用IBM Blue Gene/L超级计算机模拟大规模神经网络。

8. 2008年，惠普公司实现忆阻器原型，可作为类脑计算基本元件，并展示了首个忆阻器与硅材料的混合电路。

9. 2011年，欧盟启动BrainScaleS (Brain-inspired multiscale computation in neuromorphic hybrid systems)项目，作为FACETS的后续项目，研发大规模并行类脑计算机。

10. 2012年，蓝脑项目所模拟的最大神经网络包括100万个神经元与10亿个突触，其规模相当于蜜蜂的大脑，仿真速度是实时速度的1/300。

11. 2013年，欧盟启动人脑计划，由亨利·马克拉姆牵头，总金额10亿欧元，包括神经信息学、医学信息学、脑仿真、高性能计算、类脑计算与神经机器人6个平台。

12. 2013年，美国启动BRAIN (Brain Research through Advancing Innovative Neurotechnologies)倡议，总金额1亿美元。BRAIN并不直接涉及类脑计算，但是它将推动对生物脑机理的深入理解，为计算领域的研究者提供大量的实验数据与相关理论。

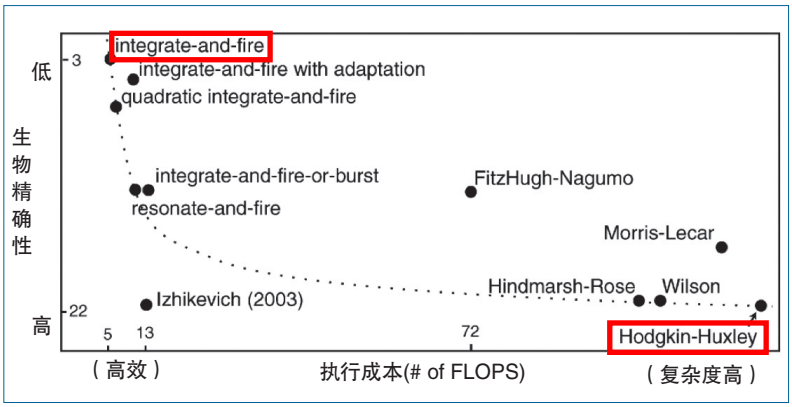

图 1 不同抽象层次的脉冲神经网络模型与其所需的运算能力[21]。“# of FLOPS”代表仿真1ms的单个神经元运行所需的浮点运算数

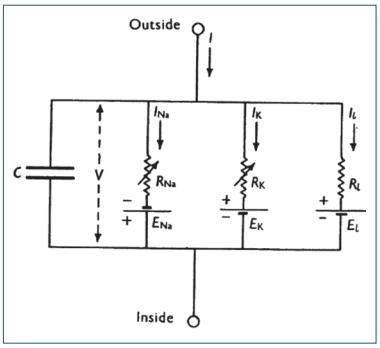

图 2 Hodgkin–Huxley 模型电路图[9]

13. 2014 年，达曼德拉·莫德哈 (Dharmendra Modha) 领导的 IBM SyNAPSE 项目推出了 True-North 芯片，包含 54 亿个晶体管，是 IBM 迄今为止制作的最大的一款芯片，其功耗只有 70mW，是传统 CPU 功耗的 1/5000（晶体管数量相等情况下）。后来实现的用于视觉显著性的应用系统[13]，可以实时处理 30 帧 / 秒的彩色视频流，包含 300 万个神经元，而其功耗只有 200mW。

从 1989 年沉寂，到 2004~2005 年悄然复兴，类脑计算在近年成为一个研究热点。主要原因是冯·诺伊曼架构遇到的内存与功耗瓶颈。

脉冲神经元建模

从生物精确性最高的 Hodgkin-Huxley (HH) 模型到最简化的 Leaky Integrate & Fire(LIF) 模型，脉冲神经网络的建模可以有多种抽象层次（如图 1 所示）。越精确的模型运算复杂度越高。

Hodgkin–Huxley(HH) 模型

HH 模型[9] 是一组描述神经元细胞膜的电生理现象的非线性微分方程，由艾伦·劳埃德·霍奇金 (Alan Lloyd Hodgkin) 与安德鲁·赫胥黎 (Andrew Huxley) 提出，并因此获得了 1952 年的诺贝尔生理学或医学奖。

HH 模型所对应的电路图如图 2 所示，其中 C 代表脂质双层 (lipid bilayer) 的电容；R_{Na}, R_K, R_l 分别代表钠离子通道、钾离子通道与漏电通道的电阻（R_{Na}, R_K 上的斜箭头表明其为随时间变化的变量，而 R_l 则为一个常数）；E_l, E_{Na}, E_K 分别代表由于膜内外电离子浓度差别所导致的漏电平衡电压、钠离子平衡电压、钾离子平

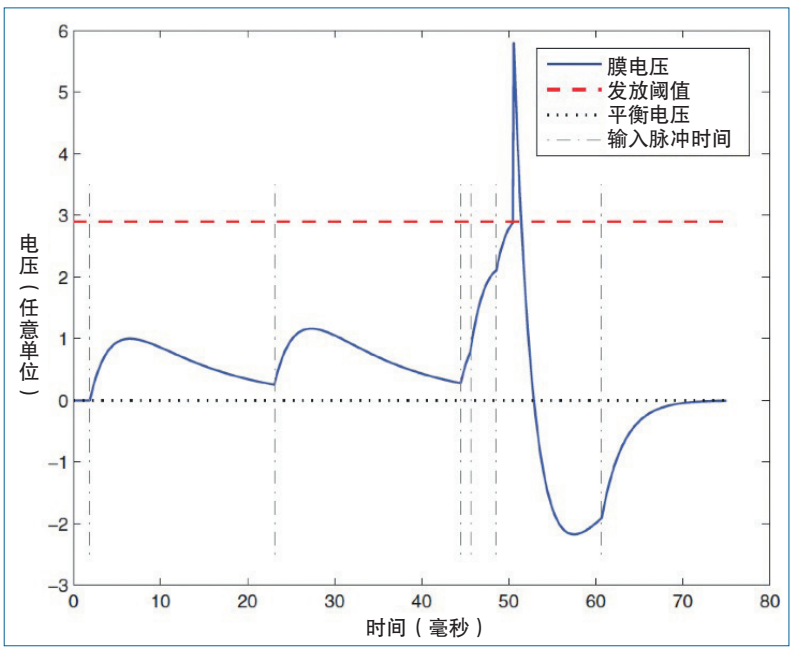

图 3 HH神经元膜电压随输入脉冲变化的示意图[28]

衡电压；膜电压 V 代表神经膜内外的电压差，可以通过 HH 模型的仿真来得到 V 随时间变化的曲线。HH 模型可以对一个点状神经元进行建模，即用一组 HH 模型公式来描述整个神经元，也可以对多个区间的详细神经元进行建模，即用一组 HH 模型公式来描述一个区间。一个完整的神经元由很多区间组成，不同区间有不同的电化学特性。HH 模型可以精确地重现不同种类生物神经元的电生理实验数据，但是运算量较高，难以实现大规模神经网络的高效仿真。基于超级计算的蓝脑项目采用的就是 HH 模型，也并非实时仿真。

HH 神经元的膜电压 V 随输入脉冲变化的示意图如图 3 所示。图中共有 6 个输入脉冲（垂直虚线所示），每个脉冲触发膜电压 V 的快速上升。如果输入脉冲之间的时间间隔较长（例如在 1ms 与 22ms 到达的 2 个脉冲之间，由于漏电通道的作用，没有新的输入脉冲），膜电压 V 就会随着时间逐渐降低至平衡电压 E_l。如果有多个输入脉冲在短时间内连续到达（例如在 45~50ms 之间的 3 个脉冲），那么膜电压 V 会上升至发放阈值 V_{th}（红色水平虚线所示）而触发一个输出脉冲。之后 V 被重置为低于平衡电压 E_l 的 V_{reset}，然后逐渐回升至平衡电压 E_l。神经元的行为与输入的时间特性密切相关，一组脉冲如果在短时间内连续到达，可以触发神经元的脉冲；但是同样数量的一组脉冲如果在较长时间段内分散到达，那么膜电压的漏电效应便不会产生脉冲。此特性可以用来实现基于精确时间的脉冲序列的模式识别，也对生物的一些实时感官功能至关重要。例如谷仓猫头鹰 (barn owl) 利用声音信号的细微时间差来定位声音来源，时间精度达数百微秒级别。

Leaky Integrate and Fire (LIF) 模型

LIF 神经元的建模有多个变种，其中最简单的是基于电流的 LIF 模型。LIF 也是类脑计算中最常用的模型。

图 4 描述了一个 LIF 神经元的膜电压 V 随着输入脉冲变化的趋势。每个输入脉冲通过输入电流项 I 触发膜电压 V 的垂直上升。当膜电压 V 上升至发放阈值 V_{th} 而触发一个输出脉冲后，V 立刻被重置为 V_{reset}，并且之后有一个无反应期。图 4 是图 3 的一种近似情况，有利于提高运算效率，尤其是在基于数字电路的仿真方面。

AdEx IF 模型

AdEx IF(Adaptive Exponential Integrate and Fire) 模型 [4] 的生物精确性与运算复杂度介于 LIF 与 HH 模型之间。当膜电压升至发放阈值 V_{th} 时，会触发一个脉冲，之后重置 V。AdEx IF 模型会同时降低膜电压的响应速度，其结果是在恒定电压的刺激下，神经元的脉冲频率会逐渐降低。

Izhikevich模型

Izhikevich 模型由研究者 Eugene M. Izhikevich 提出 [21]，其生物精确性接近于 HH 模型，运算

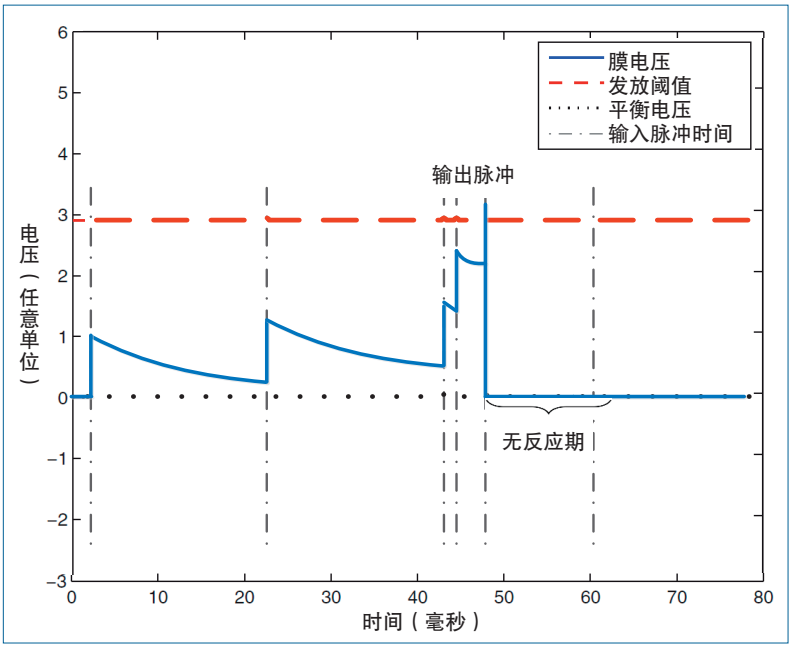

图 4　LIF 神经元膜电压随输入脉冲变化的示意图

复杂度接近于 LIF 与 AdEx IF 模型，如图 1 所示。

神经网络建模所选用的抽象层次应当与具体应用目标密切相关。如果目标是为科学家提供高效仿真工具来进行计算神经科学研究，就应当在运算能力允许的前提下尽量选用生物精确性高的模型，例如 HH 模型；如果目标是开发出具备强大实时信息处理能力的算法与产品，用于机器人、自主驾驶等应用领域，就应当尽量选用简化的模型，如 LIF 模型。2009 年，IBM 研究者达曼德拉·莫德哈采用包含 14 万 7456 个处理器的 IBM Blue Gene/P 超级计算机仿真了 16 亿个基于 Izhikevich 模型的神经元与 10^{13} 个突触，与猫的大脑规模相当，获得了超级计算领域的戈登·贝尔 (Gordon Bell) 奖。由亨利·马克拉姆牵头的瑞士洛桑联邦理工学院的蓝脑项目则采用 HH 模型进行生物神经网络仿真，希望为神经科学家提供更快、更大规模的仿真工具，以更好地理解大脑与脑疾病。莫德哈的研究目标与马克拉姆十分不同，其目的是建造低功耗的智能信息处理系统（这也体现在他后来牵头研制的基于 LIF 模型的 TrueNorth 芯片），因此 Izhikevich 或 LIF 模型是其较为合适的选择。

生物神经网络在树突脉冲、胶质细胞、神经递质触发的蛋白质信号通路、神经递质触发的基因调控、电突触、解释人脑意识方面具有高度复杂性，因此搭建一个大规模仿真系统来复制生物大脑的所有功能是极其困难的。一些哲学家与未来学家提出了全脑模拟 (whole brain emulation) 与意识上传 (mind uploading)，即将人脑神经网络所包含的所有信息上传至一个电脑仿真系统中作为备份，从而可以在一个人的身体死亡之后还能有意识的延续，以此达到永生。这种科幻式的构想真正实现起来恐怕也是遥遥无期的。

脉冲神经网络的训练算法

脉冲神经网络的训练与学习算法可以划分为以下几类：

非监督学习 该算法基于赫布法则 (Hebbian Rule) 而设。STDP（Spike Timing Dependent Plasticity，脉冲时序相关的可塑性）是赫布法则在脉冲神经网络中的表现，如果输出神经元 A 的脉冲总是发生在输入神经元 B 的脉冲之前的一个很短的时间窗口内，这就意味着 A 与 B 之间存在相关联的触发，它们之间的突触就会被增强，神经元之间的连接权重增大；反之，如果输出神经元 A 的脉冲总是发生在输入神经元 B 的脉冲之后的一个很短的时间窗口内，它们之间的突触就会被减弱。STDP 是神经科学家通过生物体电生理实验所发现的学习规则，具有很强的生物学依据。

监督学习 基于人工神经网络中常用的反向传播训练算法的思想，从所犯的错误中学习。有研究者提出了多种脉冲神经网络的监督学习训练算法，包括 SpikeProp[3], Tempotron[19], Remote Supervised Method (ReSuMe)[38], Chronotron[14], Spike Pattern Association Neuron (SPAN)[32], Precise-Spike-Driven Synaptic Plasticity (PSD) [44] 等。与 STDP 不同，这些基于反向传播的监督学习算法没有生物学依据，是研究者设计出来的。

强化学习 在与外界环境的交互过程中，基于奖赏与惩罚来选择性能较好的参数配置。

演化算法 保持一个种群，基于自然选择原理（选择、交叉与变异）来选出最优的参数配置。

由于脉冲神经网络的训练算法不太成熟，一些研究者提出了将传统的人工神经网络转化为脉冲神经网络的算法，利用较为成熟的人工神经网络训练算法来训练基于人工神经网络的深度神经网络，然后通过触发频率编码 (firing rate encoding) 将其转化为脉冲神经网络 [12]，从而避免了直接训练脉冲神经网络的困难。这些工作目前局限于前馈神经网络。基于这种转换机制，美国休斯实验室 (HRL Labs) 的研究者 [5] 将卷积神经网 (Convolutional Neural Network, CNN) 转换为 Spiking CNN，在常用的物体识别测试集 Neovision2 与 CIFAR-10 上的识别准确率接近卷积神经网；瑞士研究者将深度信念网 (Deep Belief Network, DBN) 转换为 Spiking DBN，在手写数

字识别测试集 MNIST 上的识别准确率接近深度信念网。目前还没有人尝试将脉冲神经网络应用于最具挑战性的图像物体识别测试集 ImageNet，可能是因为 ImageNet 需要巨大的深度神经网络，转换后的脉冲神经网络的软件仿真所需的运算超出了一般台式机的能力。

还有一种脉冲神经网络架构——液体状态机 (Liquid State Machine, LSM)[29] 也可以避免直接训练脉冲神经网络。液体状态机与基于人工神经网络的回声状态机 (Echo State Machine, ESM) 类似，神经元之间的连接与权重是随机产生且固定的，神经网络形成一个“水池”，其作用是将外界输入映射到一个高维状态空间以便于分类，因此脉冲神经网络本身不需要训练，只须训练一个输出层分类算法。只要脉冲神经网络的规模足够大，理论上就可以完成任意复杂输入的分类任务。由于液体状态机属于回归神经网络，神经元之间的连接是有反馈的，这使其具有记忆能力，可以有效处理时序信息。新西兰研究者尼可拉·卡萨博夫 (Nikola Kasabov) [25] 基于液体状态机的基本思想提出了 NeuCube 系统架构，用于处理时序与空间信息，例如脑电图 (ElectroEncephaloGraph, EEG) 神经信号的解码。NeuCube 在训练阶段采用 STDP、量子启发的遗传算法等来训练脉冲神经网络；在运行阶段，脉冲神经网络与输出层分类算法的参数是动态变化的，称为演化联结系统 (Evolving Connectionist Systems , ECOS)。这赋予了 NeuCube 系统很强的自适应能力。

国际上代表性类脑计算项目

表 1 列举了国际上具有代表性的类脑计算项目（基于一些大规模 CPU 或 GPU 集群的仿真系统，例如基于 IBM 超级计算机的蓝脑项目与基于日本的 K COMPUTER 超级计算机的仿真系统未列入）：

1. 实现技术 基于多核处

表1 主要类脑计算项目与成果概述

项目/单位	实现技术	神经元模型	学习算法	神经元个数	突触个数
SpiNNaker, 英国曼彻斯特大学[15]	18核ARM芯片，片上网络互联	LIF, Izhikevich	STDP	每个ARM核上1K个神经元；核的个数可达百万级别	每个ARM核上1M个突触
TrueNorth, 美国IBM公司[31]	数字电路	LIF	无	每个神经突触核上256个神经元；每个芯片上4096个核	每个神经突触核上256K个突触
HICANN, 德国海德堡大学[40]	模拟混合电路，晶片级集成	AdEx IF	STDP	每个芯片上512个神经元；每个晶片上448个芯片	每个芯片上115K个突触
Neurogrid, 美国斯坦福大学[2]	亚阈值模拟混合电路	QIF	无	每个神经核上65K个神经元；每个芯片上16个神经核	每个芯片上375M个突触
ROLLS处理器, 瑞士苏黎世理工学院 [39]	亚阈值模拟混合电路	AdEx IF	STDP	每个芯片上256 个神经元	每个芯片上128K个突触
BlueHive, 英国剑桥大学[33]	数字电路，多FPGA集群	Izhikevich	无	每个FPGA上64K个神经元	每个FPGA上64M个突触
EMBRACE, 爱尔兰阿尔斯特大学与国立大学[35]	模拟混合电路，分级片上网络互联	LIF	遗传算法	每个处理单元上32个神经元(16个输入+16个输出)	每个输入神经元144个突触；每个输出神经元17个突触
IFAT, 美国加州大学圣地亚哥分校 [36]	模拟混合电路	LIF	无	每个芯片上65K个神经元	每个芯片上65M个突触
Zeroth NPU, 美国高通公司	模拟混合电路	LIF	未知	未知	未知
Si Elegans, 欧洲多家研究机构 [26]	数字电路, 多FPGA集群（最多330个）	多种类型，包括HH, LIF	无	每个FPGA上1个神经元	全连接

理器、数字逻辑电路或模拟混合电路。

2. 神经元模型 包括 LIF, AdEx IF, Izhikevich, QIF (Quadratic Integrate and Fire) 等。

3. 学习算法 是否支持片上学习算法，是否可以动态调整突触的强度。

4. 所支持的神经元与突触个数 这决定系统所能支持的神经网络规模。大量的突触，而非神经元，往往是占用片上硬件资源的主要因素。

由表1我们总结出以下几点：

1. 线虫 (C Elegans) 体内共有302个神经元，约8000个突触连接，其神经系统的连接组 (connectome) 已被神经科学家了解透彻，因此对其神经系统的仿真比对哺乳动物更加现实。欧盟第七框架的 Si Elegans 项目旨在模拟线虫的神经系统，基于多 FPGA（Field Programmable Gate Array，现场可编程门阵列）集群，每个 FPGA 来仿真单个神经元，提供多种抽象程度的神经元模型库供研究者选用。

2. 除了曼彻斯特大学的 SpiNNaker 系统是采用多核 ARM 平台运行软件外，其他芯片均是基于硬件电路设计。采用数字电路来仿真，涉及到微分方程的递归求解，因此所需的运算能力与功耗较高。而采用模拟混合电路设计，可以利用模拟电路的物理特性来直接仿真神经元的连续性动态行为，其运算效率与功耗远远优于数字电路，因此大多数类脑芯片采用了模拟混合电路技术。要想做到超低功耗，则需要采用亚阈值 (sub-threshold) 模拟混合电路。IBM TrueNorth 数字电路芯片之所以能做到超低功耗，应当归功于采用了异步电路设计技术来充分利用脉冲神经网络的事件触发特性，使得只有接受脉冲的神经元才会被激活，而其他神经元则处于睡眠状态。

3. 大部分芯片是实时运行的，可以与外界进行实时交互，适用于机器人、手持设备等嵌入式系统领域。海德堡大学的 HICANN(High Input Count Analog Neural Network) 系统的仿真速度是实时速度的1万倍。其电路设计采用较高电压（非亚阈值）、比 LIF 更为精确的 AdEx 模型，并且通过晶片级集成，将多个芯片集成在一个未经切割的晶片上来实现大规模并行计算，其主要目标是为神经科学家提供用于大规模脉冲神经网络仿真加速的超级计算机，而非低功耗嵌入式系统。

4. 一些芯片没有片上的在线学习功能（例如 IBM TrueNorth），神经元之间的连接权重在运行时是固定不变的，这意味着神经网络的训练必须离线完成。可以采用基于 CPU 或 GPU 的大规模并行计算平台，将训练好的神经网络拓扑与参数下载到芯片上执行。这个限制大大简化了芯片设计，但同时也限制了芯片的动态自适应能力，任何改动都需要人工干预或重启系统。

5. 高通公司的 Zeroth 神经处理器 (Neural Processing Unit, NPU) 与 CPU、GPU、DSP 等处理单元并列作为一个协处理器，目标应用领域是手持移动设备。从2013年尚处于研发阶段的媒体报道来看，猜测是采用了 AMS 实现 LIF SNN 模型；2015年，高通宣布将 Zeroth 作为协处理器用于蛟龙处理器产品 (Snapdragon 820) 中，采用数字电路实现基于人工神经网络的深度卷积神经网，并且未来将引入基于人工神经网络的回归神经网络。

6. 类脑计算的产业化还处于初步阶段。IBM TrueNorth 是由美国国防部高级研究计划署提供资金支持研发的（共计5300万美元），IBM 并没有投入任何资金。与 IBM 的沃森认知计算系统在医疗健康领域已经取得的商业成功相比，TrueNorth 的产业化路线尚不明朗。其他项目也尚无成功的产业化案例。

类脑传感器

类脑视觉与听觉传感器

传统摄像头基于周期性的视频帧，帧频越高，视频质量越好，但视频码流所需的带宽也就越大。而基于视网膜原理的类脑摄像头，例如瑞士苏黎世大学研发的动态视觉传感器 (Dynamic Vision Sensor, DVS)[24]，基于事件驱动原理来检测图像中像素的亮度变化：当某个像素的亮度变

化超过某一阈值时（从亮变暗或从暗变亮），会输出一个脉冲；如果图片静止不动，没有像素的变化，摄像头就不会有任何输出。脉冲的编码采用地址事件表示 (address event representation)，包含发出脉冲的时间戳与像素地址。这种类脑摄像头的时间分辨率可达微秒级，可以实现对高速移动物体的跟踪，而其所需的码流带宽比传统的高速摄像头要低很多。由于动态视觉传感器的输出是一系列脉冲，而不是传统的基于像素矩阵的图像帧，所以传统的信号与图像处理算法并不适用，需要设计新的后端处理算法。很直接的一个思路是采用脉冲神经网络来实现后端处理算法。动态视觉传感器研究团队已经与IBM TrueNorth 团队展开合作，将 TrueNorth 芯片用于动态视觉传感器的后端处理。动态视觉传感器的低带宽使其在机器人视觉领域具有天然优势，已应用于自主行走车辆与自主飞行器中。类脑耳蜗 [23] 是基于类似原理的类脑听觉传感器，可以用于声音识别与定位。

类脑嗅觉传感器

欧盟项目 NEUROCHEM[27] 研发了基于动物嗅觉系统仿真的类脑嗅觉系统，其前端是基于导电聚合物的大规模传感器阵列，后端是基于 x86 处理器的脉冲神经网络软件模型，用于仿真昆虫嗅觉中枢或脊椎动物嗅觉中枢进行气味识别。台湾研究者 [20] 研发了一个基于模拟电路的小型专用类脑芯片，作为一个商用电子鼻产品的后端，通过仿真脊椎动物的嗅觉中枢来实现气味识别。海德堡大学研究者 [37] 采用类脑芯片 Spikey（HiCANN 芯片的前身）来仿真昆虫嗅觉系统。这些类脑嗅觉传感器的主要优点是功耗低，而其灵敏度与传统电子鼻相比并不具有明显优势。

脑机融合领域的应用

脑机接口技术在生物（人或动物）脑（或者脑细胞的培养物）与外部设备（电脑）之间建立了直接通信通路，在残疾人康复等领域有着重要应用。闭环脑机接口可以在生物脑与电脑之间建立感知与控制的双向通信机制，生物脑与电脑之间相互协作，形成一个脑机融合的混合智能系统 [43]。类脑芯片的超低功耗特性，使其在脑机接口（尤其是植入式脑机接口）领域，具有广泛的应用前景。

现有的一些用于采集生物脑电信号的植入式无线传感器，例如加州大学伯克利分校的 neural dust[41] 与华盛顿大学的 neurochip[45]，仅负责数据采集与无线传输，并不具备强大的处理能力。如果我们将类脑芯片植入到生物体内，与生物体形成一个闭环脑机接口系统来进行神经解码与编码，要面临很多研究挑战，包括：芯片要与生物体兼容，必须具备超低功耗与散热功能；要长期植入而不换电池，最好采用能量收集技术从环境中获取能量；要实时处理大量电生理数据，需要较高性能。代表性的工作包括：

1. 约翰·霍普金斯大学的研究者研制了一个基于模拟混合信号的低功耗芯片 SiCPG，包含 10 个 LIF 神经元与 190 个连接突触，用于仿真动物脊柱中的中央模式生成器 (Central Pattern Generator, CPG)。该芯片可控制一只瘫痪猫的腿部神经系统，使其自主行走。该项目的最终目标是用来帮助脊柱损伤病人实现自主行走。

2. 斯坦福大学研究者 [10] 采用滑铁卢大学克里斯·伊利亚史密斯 (Chris Eliasmith) 的神经工程框架 (Neural Engineering Framework, NEF)，搭建了一个基于 2000 个 LIF 神经元的脉冲神经网络模型，用于取代现有的卡尔曼滤波算法来解码猴子运动皮层的电生理数据，用于基于脑机接口的神经假体 (neural prosthesis)，用脑电信号来控制外部机械手。其未来工作是基于斯坦福大学的类脑芯片 Neurogrid 以完成此脉冲神经网络模型。

3. 南加州大学研究者 [16] 研制了一个基于模拟电路的超低功耗神经信号处理芯片，用于大鼠海马体的认知假体 (cognitive prosthesis)，修复一定程度的记忆功能。其实现是采用经典信号处理算法来对脑电信号进行函数拟合。我们想象可以采用一个基于脉冲神经网络的类脑芯片来仿真部分脑区的功能，与其他脑区

紧密交互，来修补一些缺失或者出故障的脑功能。如果需要仿真的脑区范围较小、功能较为简单、神经元个数不多，这也许可行。

展望

尽管类脑计算近年来有了不少进展，但是依然面临诸多挑战。

训练算法 脉冲神经网络训练算法的理论发展还不够成熟，特别是不能很好地训练包含多个隐层的深度神经网络。针对工业界常用的标准测试集，例如手写数字识别测试集 MNIST，脉冲神经网络在分类算法精确度方面的性能通常会略低于基于人工神经网络的深度神经网络。

低功耗运算 类脑计算的主要优势是功耗低，但是基于人工神经网络的数字电路硬件实现也可以通过体系结构层面的创新来做到低功耗，例如寒武纪系列芯片[6]、用于卷积神经网加速的协处理器 NeuFlow[34] 与 Synopsys 公司的视觉处理器系列 DesignWare EV 等。

编程模型 基于亚阈值的模拟混合信号芯片可以做到超低功耗，但是基于模拟混合信号的应用开发十分困难，需要有经验模拟电路设计者，并给予高薪待遇，因此在可编程性方面输给了数字电路技术。IBM TrueNorth 芯片提供了一套编程模型 corelet，但它是为 TrueNorth 系统量身定制的，而非通用的类脑计算编程模型。

类脑计算技术离工业界的实际应用还有较大差距。但这些挑战也为研究者提供了新的研究方向与机遇。我们认为基于脉冲神经网络的类脑计算在未来 5~10 年内将会是一个重要的研究题目。而其产业化前景是否能够被工业界广泛接受，则取决于研究者在此期间是否能够在某一方面取得突破性进展。■

顾宗华
CCF高级会员。浙江大学副教授。主要研究方向为实时嵌入式系统与软硬件协同设计。
zgu@zju.edu.cn

潘 纲
CCF高级会员。浙江大学教授。主要研究方向为普适计算、计算机视觉、智能系统等。gpan@zju.edu.cn

参考文献

[1] Hadi Esmaeilzadeh, Emily R. Blem, Renée St. Amant, Karthikeyan Sankaralingam, Doug Burger: Power challenges may end the multicore era. *Commun. ACM* 56(2): 93~102 (2013).

[2] Benjamin B V, Gao P, McQuinn E, et al. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations[J]. *Proceedings of the IEEE*, 2014, 102(5): 699~716.

[3] Bohte S M, Kok J N, La Poutré J A. SpikeProp: backpropagation for networks of spiking neurons[C]// ESANN. 2000: 419~424.

[4] Brette R. and Gerstner W.. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol*., vol. 2005; 94: 3637~3642.

[5] Cao Y, Chen Y, Khosla D. Spiking Deep Convolutional Neural Networks for Energy-Efficient Object Recognition[J]. *International Journal of Computer Vision*, 2014: 1~13.

[6] 陈云霁. 体系结构研究者眼中的神经网络硬件. 中国计算机学会通讯. 2015年第7期.

[7] Robert Colwell, The Chip Design Game at the End of Moore's Law, the 25th Hot Chips Conference, 2013.

[8] Cruz-Albrecht J M, Derosier T, Srinivasa N. A scalable neural chip with synaptic electronics using cmos integrated memristors[J]. *Nanotechnology*, 2013, 24(38): 384011.

[9] Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. Cambridge: MIT Press.

[10]Dethier, Julie, et al. Design and validation of a real-time spiking-neural-network decoder for brain–machine interfaces. *Journal of neural engineering* 10.3 (2013): 036008.

[11]Tim Dettmers, The Brain vs Deep Learning, https://timdettmers.wordpress.com/2015/07/27/brain-vs-deep-learning-singularity/.

[12]Diehl P U, Neil D, Binas J, et al. Fast-Classifying, High-Accuracy Spiking Deep Networks Through Weight and Threshold Balancing[J]. 2015.

[13]Andreopoulos A, Taba B, Cassidy A S, et al. Visual saliency on networks of neurosynaptic

cores[J]. IBM Journal of Research and Development, 2015, 59(2/3): 9: 1~16.

[14]Florian R V. The chronotron: a neuron that learns to fire temporally precise spike patterns[J]. PloS one, 2012, 7(8): e40233.

[15]Furber S B, Galluppi F, Temple S, et al. The spinnaker project[J]. Proceedings of the IEEE. 2014;102(5): 652~665.

[16]Ghaderi V, Song D, Berger T. Nonlinear Cognitive Signal Processing in Ultra-Low-Power Programmable Analog Hardware[J]. IEEE Transactions on Circuits and Systems—II: Express Briefs, Vol. 62, No. 2, 2015.

[17]Gomes L., Facebook AI Director Yann LeCun on His Quest to Unleash Deep Learning and Make Machines Smarter, Interview by IEEE Spectrum, 2015.

[18]Katja Grace, http://aiimpacts.org/brain-performance-in-teps/.

[19]Gütig R, Sompolinsky H. The tempotron: a neuron that learns spike timing–based decisions[J]. Nature neuroscience, 2006, 9(3): 420~428.

[20]Hsieh H Y, Tang K T. VLSI implementation of a bio-inspired olfactory spiking neural network[J]. Neural Networks and Learning Systems, IEEE Transactions on, 2012, 23(7): 1065~1073.

[21]Izhikevich E., Which model to use for cortical spiking neurons? Neural Networks, IEEE Transactions on, ,2003; 15 (5), 1063~1070.

[22]Indiveri G, Linares-Barranco B, Hamilton T J, et al. Neuromorphic silicon neuron circuits[J]. Frontiers in neuroscience, 2011, 5.

[23] Li, Cheng-Han, Tobi Delbruck, and Shih-Chii Liu. Real-time speaker identification using the AEREAR2 event-based silicon cochlea. Circuits and Systems (ISCAS), 2012 IEEE International Symposium on. IEEE, 2012.

[24]Liu S C, Delbruck T. Neuromorphic sensory systems[J]. Current opinion in neurobiology, 2010, 20(3): 288~295.

[25]Kasabova N, Scotta N, Tua E, et al. Evolving Spatio-Temporal Data Machines Based on the NeuCube Neuromorphic Framework: Design Methodology and Selected Applications[J].

[26]Machado P, Wade J, McGinnity T M. Si elegans: FPGA hardware emulation of C. elegans nematode nervous system[C]//Nature and Biologically Inspired Computing (NaBIC), 2014 Sixth World Congress on. IEEE, 2014: 65~71.

[27]Marco S, Gutiérrez-Gálvez A, Lansner A, et al. A biomimetic approach to machine olfaction, featuring a very large-scale chemical sensor array and embedded neuro-bio-inspired computation[J]. Microsystem technologies, 2014, 20(4-5): 729~742.

[28]Masquelier T, Guyonneau R, Thorpe S J. Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains[J]. PloS one, 2008, 3(1): e1377.

[29]Maass W, Natschläger T, Markram H. Real-time computing without stable states: A new framework for neural computation based on perturbations[J]. Neural computation, 2002, 14(11): 2531~2560.

[30]Mayr C, Partzsch J, Noack M, et al. A Biological-Realtime Neuromorphic System in 28 nm CMOS using Low-Leakage Switched Capacitor Circuits[J]. arXiv preprint arXiv:1412.3233, 2014.

[31]Merolla P A, Arthur J V, Alvarez-Icaza R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface[J]. Science, 2014, 345(6197): 668~673.

[32]Mohemmed A, Schliebs S, Matsuda S, et al. Span: Spike pattern association neuron for learning spatio-temporal spike patterns[J]. International Journal of Neural Systems, 2012, 22(04): 1250012.

[33]Moore S W, Fox P J, Marsh S J T, et al. Bluehive-a field-programable custom computing machine for extreme-scale real-time neural network simulation[C]// Field-Programmable Custom Computing Machines (FCCM), 2012 IEEE 20th Annual International Symposium on. IEEE, 2012: 133~140.

[34]NeuFlow, http://www.neuflow.org/.

[35]Pande S, Morgan F, Smit G, et al. Fixed latency on-chip interconnect for hardware spiking neural network architectures[J]. Parallel computing, 2013, 39(9): 357~371.

[36]Park J, Ha S, Yu T, et al. A 65k-neuron 73-Mevents/s 22-pJ/event asynchronous micro-pipelined integrate-and-fire array transceiver[C]//Biomedical Circuits and Systems Conference (BioCAS), 2014 IEEE. IEEE, 2014: 675~678.

[37]Pfeil T, Grübl A, Jeltsch S, et al. Six networks on a universal neuromorphic computing

substrate[J]. Frontiers in neuroscience, 2013, 7.

[38]Ponulak F, Kasinski A. Supervised learning in spiking neural networks with ReSuMe: sequence learning, classification, and spike shifting[J]. Neural Computation, 2010, 22(2): 467~510.

[39]Qiao, Ning, et al. A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses. Frontiers in neuroscience 9 (2015).

[40]Schemmel J, Bruderle D, Grubl A, et al. A wafer-scale neuromorphic hardware system for large-scale neural modeling[C]//Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. IEEE, 2010: 1947~1950.

[41]Seo D, Carmena J M, Rabaey J M, et al. Neural dust: An ultrasonic, low power solution for chronic brain-machine interfaces[J]. arXiv preprint arXiv:1307.2196, 2013.

[42]Wang R, Hamilton T J, Tapson J, et al. An FPGA design framework for large-scale spiking neural networks[C]//Circuits and Systems (ISCAS), 2014 IEEE International Symposium on. IEEE, 2014: 457~460.

[43]Zhaohui Wu, Gang Pan, José Carlos Príncipe, Andrzej Cichocki: Cyborg Intelligence: Towards Bio-Machine Intelligent Systems. IEEE Intelligent Systems 2014; 29(6): 2~4.

[44]Yu Q, Tang H, Tan K C, et al. Precise-spike-driven synaptic plasticity: Learning hetero-association of spatiotemporal spike patterns[J]. 2013.

[45]Zanos S, Richardson A G, Shupe L, et al. The Neurochip-2: an autonomous head-fixed computer for recording and stimulating in freely behaving monkeys[J]. Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 2011, 19(4): 427~435.

CCF普适计算专委会完成换届选举

9月11日，CCF普适计算专委会在辽宁葫芦岛召开换届会议。CCF秘书长**杜子德**代表CCF总部监督换届选举。到场的普适计算专委委员通过不记名投票方式，差额选举出新一届专委会主任**吴朝晖**（浙江大学教授），秘书长**李石坚**（浙江大学教授），副主任**陈益强**（中科院计算所研究员）、**陈渝**（清华大学教授）、**於志文**（西北工业大学教授）、**张大庆**（北京大学教授）。

CCF人机交互专业组完成换届选举

9月11日，CCF人机交互专业组在辽宁葫芦岛召开换届会议。CCF秘书长**杜子德**代表CCF总部监督换届选举。到场的人机交互专业组委员通过不记名投票方式，差额选举出新一届专业组主任**戴国忠**（中科院软件所研究员），秘书长**田丰**（中科院软件所研究员），副主任**史元春**（清华大学教授）、**汪国平**（北京大学教授）、**王涌天**（北京理工大学教授）、**王茜莺**（联想用户研究中心高级总监）。