类脑计算系统

黄铁军，王韬，杜凯

1. 类脑计算理念

1.1 从图灵机到冯诺依曼体系结构

1.2人工神经网络

1.3 神经形态计算

1.4 类脑计算的必要性

2. 神经形态计算软件仿真平台

2.1 概述

2.2 NEURON

2.2 NEST

2.3 GENESIS

2.4 其他

3. 神经形态计算系统

3.1 概述

3.2 SpiNNaker

3.3 BrainScaleS

3.4 IBM TrueNorth系统

3.5 其他

4. 分析与展望

4.1 类脑计算的优势

4.2 类脑计算的挑战

4.3 类脑计算的未来

目录

1 类脑计算理念 2

1.1 从图灵机到类脑计算 2

1.2 类脑计算系统概述 5

2 神经形态计算软件仿真平台（杜凯） 8

3 类脑计算系统（王韬） 9

1.1 概述 9

1.2 SpiNNaker 10

1.3 BrainScaleS 12

3.4 IBM TrueNorth系统 14

3.5 其他 15

4 分析与展望 17

4.1 类脑计算的优势 17

4.2 类脑计算的挑战 17

4.3 类脑计算的未来 17

# 类脑计算理念

## 1.1 从图灵机到类脑计算

众所周知，现代计算机产生的数学基础是数理逻辑，物理基础是开关电路。数理逻辑的研究对象是证明和计算这两个直观概念符号化后的形式系统，1936年阿兰·麦席森·图灵(Alan Mathison Turing，1912-1954)为了研究“不可计算数”而提出了图灵机模型，这一看似简单的“思想实验”抓住了数理逻辑和抽象符号处理的本质：一台仅能处理0和1这样简单二元符号的机械设备，就能够模拟任意数学推理过程，这就是现代计算机的起源。1938年克劳德·艾尔伍德·香农开创开关电路理论，在数理逻辑和电路实现之间架起了桥梁。1946年首台计算机ENIAC研制成功，实际上是一个近1.8万个电子管作为开关的大型开关电路系统，同年冯·诺依曼（John von Neumann，1903-1957）提出存储和计算分离的存储程序结构，1952年采用这种体系结构的计算机EDVAC问世，只用2300个电子管，性能比ENIAC提高十倍。

冯·诺依曼体系结构一直沿用至今，虽然有各种优化，但无根本性变化。物理基础方面，1947年贝尔实验室成功制造出第一个晶体管，1954年贝尔实验室组装出第一台晶体管计算机TRADIC，晶体管代替电子管成为开关电路的基本单元，集成大规模晶体管的集成电路按照摩尔定律指数发展，推动计算机的计算能力和存储能力指数增长，时至今日，计算机已经渗透到社会的各个层面，计算机应用千变万化，令人眼花缭乱，但是背后的基本原理和体系结构并无根本变化。

2004-2005年前后，丹纳德尺度（Dennard Scaling，即在半导体的尺寸不断缩小的同时其功耗密度大致保持不变）失效，功耗、散热等问题使得芯片设计者开始转向多核以及众核。现代的大型众核芯片出现暗硅（Dark Silicon）问题，即功耗与散热问题导致众核芯片上的大部分核不能同时全部开启运行。研究者[Esmaeilzadeh13]认为功耗问题将终结多核尺度（multicore scaling），即单个芯片上的核的个数不会再持续增加。因此普遍认为，由于技术与经济两方面的原因，摩尔定律在持续50年后，将于2020年左右走到尽头。尽管不同人对于摩尔定律的具体终结时间有不同估算，但是工业界对于摩尔定律的终结本身并无争议。

除了开关电路规模遇到瓶颈外，冯·诺依曼体系结构也不是实现类脑智能的合适平台。冯·诺依曼体系结构的基本特征为内存与计算单元分离，优点是软件可编程，即同一硬件平台可以通过存储不同软件来执行不同功能；缺点是存储单元与计算单元之间的通信延迟成为性能瓶颈，造成所谓的“内存墙”问题。相比之下，生物神经网络的基本特征是“存储与计算合二为一”：神经元既有计算功能，又有存储功能，这就从根本上去除了冯·诺依曼体系架构的“内存墙”问题。

人脑是更为高效的智能平台，这个问题冯·诺依曼本人就认真就思考过。可惜的是，1956年人工智能先驱们举行夏季研讨时，冯·诺依曼已经因为脑癌住进医院，1957年去世。根据冯·诺依曼未完成的西列曼演讲整理而成的《计算机与人脑》一书在1958年出版[7]，其中超过一半的篇幅是讨论神经元、神经脉冲、神经网络以及人脑的信息处理机制。现在我们已经知道，人脑具备以下特征与优势：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1011个神经元）；高度连接性（约1015个突触）；低运算频率（约100Hz)；低通信速度（每秒钟几米）；低功耗（约20瓦），采用冯·诺依曼体系结构建造一个与人脑复杂程度相等的计算机，需要将近100兆瓦的功耗。性能方面，有研究者提出了一个用于比较人脑与计算机性能的指标[9]：在一个大型随机图上“每秒穿越的边的个数”（Traversed Edges Per Second，TEPS)，其基本思路是大脑的性能瓶颈是神经元之间的通信而不是运算，而用于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，人脑大约比当今最快的超级计算机要快30倍左右。

上文说过，经典的人工神经网络（ANN）与大脑神经网络有天壤之别。生物神经网络是脉冲神经网络（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（synapse）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ANN中的数值运算。

类脑计算机采用脉冲神经网络替代经典计算机的冯·诺依曼体系结构，采用微纳光电器件模拟生物神经元和突触的信息处理特性，或者说，类脑计算机是按照生物神经网络采用神经形态器件构造的新型计算机，更准确地应该称为“类脑机”或“仿脑机”（Brain-like Machine）[8]。

下面简要回顾一下类脑计算和类脑计算机的发展历史[9]：

1. 1989年，加州理工Carver Mead撰文提出了“类脑工程”的概念，并撰写了一本书, 题目为“模拟VLSI与神经系统”（Analog VLSI and Neural Systems)，采用亚阈值模拟电路来仿真SNN，其应用是仿真视网膜。
2. 1990-2003年，摩尔定律持续发展，基于冯诺依曼架构的处理器主频与性能持续增长，而类脑计算则十余年沉寂。
3. 2004年左右，单核处理器主频停止增长，设计者开始转向多核，同时学术界开始寻求非冯诺依曼架构的替代技术。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Kwabena Boahen (Carver Mead的学生) 研制出基于模拟电路的类脑芯片Neurogrid。
5. 2005年，英国Manchester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，美国国防研究局DARPA启动SyNAPSE（Systems of Neuromorphic Adaptive Plastic Scalable Electronics）项目，支持IBM与多家合作单位联合研发类脑芯片。瑞士EPFL研究者Henry Markram与IBM合作启动了蓝脑（Blue Brain）项目，采用IBM Blue Gene/L超级计算机模拟较为生物精确的大规模神经网络。
8. 2008年，惠普公司实现忆阻器（memristor)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一百万个神经元与十亿个突触，其规模相当于蜜蜂的大脑，仿真速度比实时慢约300倍。
11. 2013年，欧盟启动人脑计划（Human-Brain Project, HBP)，由Henry Markram牵头，总金额10亿欧元，包括六个平台：神经信息学平台, 医学信息学平台,脑仿真平台，高性能计算平台，类脑计算平台，与神经机器人平台。
12. 2013年，美国启动BRAIN （Brain Research through Advancing Innovative Neurotechnologies）。BRAIN并不直接涉及类脑计算，但是它将推动对于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。
13. 2014年，Dharmendra Modha领导的IBM SyNAPSE项目推出了TrueNorth芯片，包含54亿个半导体，从半导体数量来说是IBM迄今为止所制作的最大的一款芯片，而功耗只有70毫瓦，比半导体数量相当的传统CPU功耗要低5000倍左右。后来实现的一个用于视觉显著性的应用系统 [Andreopoulos15]（确认文献），可以实时处理30帧每秒的彩色视频流，包含三百万个神经元，功耗只有惊人的200mW。
14. 2015年3月，德国海德堡大学在一个8英寸硅片上集成了二十万神经元和五千万突触，采用这种“神经形态处理器”的类脑计算机成功运行。
15. 2016年3月，欧盟人类大脑计划HBP宣布把刚刚建成的两套类脑计算机通过互联网对外开放使用，以支持神经微回路模拟以及在机器学习和认知计算中应用类脑原理的相关研究。这两套系统即上面提到的德国海德堡大学的BrainScaleS系统和英国曼彻斯特大学的SpiNNaker系统。

过去二十多年特别是近十年来，类脑计算已经取得不少阶段性成果。近三年来，各国脑计划相继展开，大脑解析的步伐正在加快。两大潮流交相辉映下，具有部分类脑功能的类脑计算装置将很快进入实用，真正的类脑计算机极有可能在未来十年左右成为现实。类脑计算机成功应用后，人们能够更好地“理解智能”，并在此基础上进行简化或优化设计，进一步研制超越人脑的“超脑机”，这是后话。

## 1.2 类脑计算系统概述

仿真大脑的努力可以追溯到上世纪80年代。1972年美国生物学家杰拉尔德·艾德曼（Gerald Maurice Edelman，1929-2014）因发现免疫系统抗体的化学结构而获得诺贝尔生理学或医学奖，1978年左右转向意识研究，他的基本思想可概括为“神经群体选择理论（Neuronal Group Selection）”[12]，即神经系统产生感知、行为、学习、记忆以及意识的机理与生物适应自然环境的自然选择机制类似，因此被称为“神经达尔文主义”（Neural Darwinism）[12]。为了验证这一学说，从1981年开始，提出了统称为“综合神经建模（synthetic neural modeling）”的理论，即逼近真实解剖和生理数据的神经系统大规模仿真[14]，并研制了一系列名为“Darwin（达尔文）”的“仿脑机”（Brain-Based-Devices，BBD）[15-16]。仿脑机仿真的是不同脑区（如海马或小脑），通过从多种仿真神经回路中进行选择而实现学习。起初是软件，1992年开始采用硬件。以2005-2007年研制的达尔文10号和11号为例，仿真约50个脑区、10万神经元和140万突触连接，通过模拟啮齿类动物走迷宫的过程，理解大脑空间记忆的形成过程。基于BBD的足球机器人2004年至2006年参加RoboCup机器人足球公开赛，曾五局全胜卡内基梅隆大学的经典人工智能系统。

现代微电子学和大规模集成电路先驱、加州理工学院教授卡弗·米德（Carver Andress Mead，1934—）也是在上世纪80年代把兴趣转向了生物神经系统的，与艾德曼关注神经元群体和神经环路不同，米德的关注点在神经元的硬件实现，开创了“神经形态工程（Neuromorphic Engineering）”这个方向 [21-22]，提出采用亚阈值模拟电路来仿真脉冲神经网络，并提出了“神经形态处理器（Neuromorphic Processors）”的概念。1989年5月，米德在ISCAS（International Symposium on Circuits and Systems）会议期间组织了“模拟集成神经系统（Analog Integrated Neural Systems）”研讨会[23]，主要参会人员至今仍然活跃在这一领域。米德1989年招收的博士生博阿汉（Kwabena Boahen）2005年加入斯坦福大学，成立了“硅脑”（Brains in Silicon）实验室，2009年研制出了神经形态电路板Neurogrid，每块板16颗芯片，每颗芯片内集成了65536个神经元，每个神经元用340个亚阈值工作状态的晶体管模拟，这样一块Neurogrid板就支持100万个神经元和60亿个突触联结，能耗只有5瓦，在神经系统模拟方面可媲美能耗1兆瓦的超级计算机[24]。

使用超级计算机可以更为精确地仿真生物神经网络。2005年，瑞士洛桑联邦理工学院（EPFL）亨利·马克拉姆（Henry Markram，1962-）领导的“蓝色大脑计划”[17]获得瑞士政府资助，平台就是IBM的蓝色基因超级计算机。2007年，IBM Almaden研究中心认知计算研究组在超级计算机上开发了称为“皮层模拟器（Cortical Simulator）”的大脑模拟软件，牵头的“神经形态自适应可塑性可扩展电子系统SyNAPSE（Systems of Neuromorphic Adaptive Plastic Scalable Electronics）并得到DARPA资助， 2009年在蓝色基因超级计算机上实现了8.61T个神经突触的猫脑模拟，在哺乳动物大脑皮层上经常能观测到的阿尔法振荡（10Hz左右）和伽马振荡（~40Hz左右的信息整合现象）现象，也在这个“软件大脑”上出现了，在接受视觉模式刺激时，这个“软件大脑”还出现了哺乳动物视皮层上才能观察到的特定群体响应潜伏现象[18]。同样在2009年，马克拉姆团队在蓝色基因超级计算机上构造出出生两周大鼠的新皮质柱精细模型，包括1万个神经元和数千万个突触连接，同样出现了伽马振荡现象。2009年12月，马克拉姆在接受采访时表示：“从技术层面上讲，利用计算机和数据采集技术，有可能在10年内建立起人脑模型。”尽管对这一看法存在诸多争议，但这个前景实在诱人，由马克拉姆领衔的欧洲“人类大脑计划”于2013年1月获得欧盟批准，提出将信息技术和生命科学结合，整合从单分子探测到大脑整体结构解析，实现全脑仿真模拟[19]。

为了实现全脑仿真的目标，“人类大脑计划”支持了两台大型神经形态计算系统的研制：英国曼彻斯特大学的SpiNNaker系统和德国海德堡大学的BrainScaleS，2016年3月，两台阶段样机正式上线运行。SpiNNaker是一种受大脑结构和功能启发的大规模并行计算体系结构，正式启动于2005年的EPSRC项目，负责人是ARM处理器发明人史蒂夫·佛伯（Steve Furber，1953-）。SpiNNaker系统采用定制ARM处理器作为基本单元，分为五代，最初的102机使用了约102个ARM核，计划2020年完成的106机则集成了约106个ARM核，可实时精细仿真10亿生物神经元。SpiNNaker用数据包交换来模拟神经元之间的异步稀疏脉冲交换[25-26]，因此可以在物理连接大大少于大脑的情况下实现相同性能的信息交换。BrainScaleS由德国海德堡大学卡尔海因茨·迈耶（Karlheinz Meier，1955-）教授负责[27-28]，前身是2005到2010年的FACTES项目，强调生物神经系统和神经形态计算之间的比较研究，微观层面研究神经元的信号处理特性及模拟电路实现，介观层面研究突触可塑性及数字电路实现，在8英寸晶园上实现了20万神经元和5千万突触，晶圆内总线速度达每秒1T脉冲，晶圆间分布式通信速度每秒10G脉冲。在人类大脑计划支持下，2016年完成了20块晶圆、400万神经元和10亿突触的神经形态计算系统[29]，速度比生物系统快1万倍。2022年（也就是人类大脑计划结束前）目标系统现在尚未敲定，预计构造一个500块或5000块晶圆组成的大型系统，即使是500块方案，也能同时仿真5亿神经元，考虑比人脑速度高上万倍，因此将具备实时仿真人类大脑的能力。

IBM主导的SyNAPSE项目在超级计算机上进行大脑皮层仿真的基础上，为了突破规模瓶颈，也开发了神经形态芯片TrueNorth芯片[30]，2014年《Science》列入十大进展。TrueNorth采用成熟的CMOS集成电路工艺，神经元采用简单的LIF模型，每片集成4096个核，每核内有256个输入神经元和256个输出神经元，突触状态、神经元状态和参数、脉冲目的地址、轴突延迟等均用静态随机存储器记录。每片总计100万个神经元和2.56亿突触连接，耗费54亿个晶体管，单个芯片平均放电频率20Hz，单神经元放电功耗26pJ，芯片功耗低至65mW，比晶体管数量相当的传统CPU功耗低5000倍左右。基于这款芯片，IBM建立了Corelet编程模型、算法库和相应的软件开发环境，结合Compass模拟器，用户可以快速尝试不同的模型和参数，从中找出优化的方案，美国劳伦斯·利弗莫尔国家实验室和IBM公司近期公布了一款智能超级计算机。

如果说Neurogrid、BrainScaleS、SpiNNaker、TrueNorth等神经形态计算系统代表了神经计算机的今天，那么全新的神经形态器件将决定神经计算机的明天，要在一个有限的物理空间中以较低功耗实现大脑规模的神经计算机，必须研制尺度和功耗都与生物相当甚至更小的神经形态器件。在生物突触仿真方面，忆阻器[31]和相变存储器从2008年开始就持续引起广泛关注。2016年8月，IBM苏黎世研究院宣布用相变材料研制出世界第一个人工神经元[32]。9月，美国马萨诸塞大学阿默斯特分校采用扩散型忆阻器研制成功逼近神经突触的器件[33]，10月，普林斯顿大学宣布研制成功比光神经元。中国在神经形态器件方面已经有十多年的研究历史，北京大学、清华大学、南京大学、中科院上海微系统所、华中科技大学和国防科技大学等单位的成果表明中国很有可能对这一领域产生巨大影响。尽管这些器件应用于神经计算系统还有很长的路要走，它们将从根本上塑造神经计算机的未来，就像晶体管和集成电路对经典计算机的贡献一样。

# 神经形态计算软件仿真平台（杜凯）

# 类脑计算系统（王韬）

## 概述

在20世纪50年代，冯·诺依曼和图灵就都对脑启发机器（brain-inspired machines）进行了讨论。用定制硬件实现神经启发系统的想法早在计算机科学和计算机工程出现的时候就诞生了。计算机科学家一直都想要在计算机中复制生物神经系统，并做出了人工神经网络、机器学习等尝试。但是这些方法在发展中也都相应地遇到了一些问题，诸如缺乏足够的适应性和通用性，完成复杂任务伴随着高能耗的代价。Ailamaki 等人指出:“除人脑以外，没有任何一个自然或人工系统能够具有对新环境与新挑战的自适应能力、新信息与新技能的自动获取能力、在复杂环境下 进行有效决策并稳定工作直至几十年的能力。没有任何系统能够在多处损伤的情况下保持像人脑一样好的鲁棒性，在处理同样复杂的任务时，没有任何人工系统能够媲美人脑的低能耗性。” 近年来，类脑智能正在成为人工智能与计算科学领域的研究热点。

近30年来，类脑计算系统的提出和研究主要基于十大动机：与体系结构/微体系结构相关的有并行性、冯·诺依曼的瓶颈、可扩展性、低功耗、速度更快、在线学习，而实时性能、封装大小、高容错、神经科学启发则体现了类脑系统较之经典计算机系统的优势。最近出现了一系列大规模的神经形态项目，都具备前所未有的规模和能力，利用了现有的单个微芯片上的巨大的晶体管资源，在一种情况下，可以利用全硅晶片。该技术的功能与可扩展架构相结合，使得神经形态功能可以扩展到支持神经网络的规模，其中包含数百万个神经元，具有数十亿个突触。这些新的功能特性使得我们能够对从昆虫到小一点的哺乳动物的大脑或者人类大脑的大部分区域进行建模；同样的系统还提供能够支持新的认知结构尺度的平台。在进行建模的时候，对于神经元模型，我们可以选择各种模型，像HH、LIF、Izhikevich等；对于突触模型，我们也可以选择是否支持在线学习，是否有STDP算法；神经元和突触之间的互连形成的网络支持哪种拓扑结构，是否支持stochastic neural networks等问题也需要被考虑；选择不同模型，会对最后类脑计算系统的结构产生影响。现在已有的一些类脑计算系统中，研究者设计思路的不同，使得各大类脑系统的结构也不尽相同：BrainScaleS和SpiNNaker旨在理解和建模生物神经系统，IBM的TrueNorth的目的则在于支持实时认知应用和低功耗；SpiNNaker和TrueNorth底层实现方法是数字电路，而BrainScaleS采用的是模拟和数字混合电路；SpiNNaker支持LIF和Izhikevich神经元模型，且可简单编程，具有可塑性，TrueNorth支持LIF模型，灵活性较低，不可编程，BrainScaleS支持AdEx神经元模型，在片上采用STDP学习算法。不同的结构使得各大类脑计算系统各有所长，研究者可以让他们的特性互补，开发下一代类脑计算系统。

在一系列神经形态项目中，欧盟的“人类大脑计划”支持了两台大型神经形态计算系统的研制：英国曼彻斯特大学的SpiNNaker系统和德国海德堡大学的BrainScaleS，此外备受关注的还有IBM开发的神经形态芯片TrueNorth芯片。下面将主要介绍这三个类脑计算系统，还会提及在神经形态计算发展进程中的一些其他类脑计算系统。

## SpiNNaker

SpiNNaker是一种受大脑结构和功能启发的大规模并行计算体系结构，正式启动于2005年的EPSRC项目，负责人是ARM处理器发明人史蒂夫·佛伯（Steve Furber，1953-），研究小组为英国曼彻斯特大学的APT（Advanced Processor Technologies）研究组。SpiNNaker项目旨在提供一种大规模并行的百万核心计算机，其互连架构受到哺乳动物脑的连接特性的启发，适用于对于大规模脉冲神经网络进行生物实时建模。这种互连架构可以传输非常大量的小数据包，每个数据包明确地传达单个神经动作电位或“尖峰”的来源，同时隐含地传达时间。

每颗SpiNNaker芯片由18个简单处理器核心组成，采用定制ARM处理器作为基本单元，每个核心含有32KB指令内存与64KB数据内存，可以本地运行多种Neuron与Synapse模型；一颗芯片的18个内核共享128MB的内存，同时每个内核具有96kB的本地内存。每个处理器内核是通用的200-MHz 32-b整数处理器，没有浮点硬件，因此算术通常被实现为定点运算。每个内核可以模拟几百个点神经元模型，比如LIF(Leaky Integrate and Fire)和Izhikevich模型，每个神经元可以接受1000个突触的输入，但实际上由于各种不同的约束，每个处理器内核可以实时支持的神经元数量和连接数量会受到限制。

SpiNNaker机器是专为支持在大脑中发现的各种通信而设计的计算机，可以看做是一个同构的二维多指令，多个数据处理节点的数组，每个节点就是一颗SpiNNaker芯片，每个芯片连接相邻6个芯片。模拟大脑中的各种通信需要庞大的计算量，基于这一需求，SpiNNaker采用大规模的并行计算，其架构可容纳多达一百万个微处理器内核。它的关键创新在于轻量级的多播数据包路由机制，该机制是常规AER的延伸；与针对大型数据包的传统集群和高性能计算机通信系统相反，SpiNNaker的通信结构被优化以传输大量小型数据包。每个包携带一个单神经元“尖峰”事件，通过静态配置的多点传送路径将一个小型数据包传送到多个终点，这一点也与超级计算机通常采用的以动态路由分配点对点传输不同。通过用数据包交换来模拟神经元之间的异步稀疏脉冲交换，可以在物理连接大大少于大脑的情况下实现相同性能的信息交换。在SpiNNaker机器中，当处理器上的软件识别到神经元应该发射脉冲时，它只会发出一个识别被激发的脉冲神经元的数据包，该包被发到哪里完全由路由结构负责。每个节点包含一个包路由器，它会检查每个数据包，查看来源，用多播传输的方法将其相应地发送到节点内18个本地处理器的任何子集和/或其相邻六个节点的任何子集，发送路径由路由器表确定，该表会在应用被加载到机器中时被初始化。每个SpiNNaker节点选择其18个处理器核心中的一个作为“监视器处理器”（通常是最先被激活的那一个）。被选中的监视器将会被用作操作系统支持角色。 其余16个处理器则会被用作应用程序支持，第18个处理器保留为容错备用。芯片的通信网络支持更多芯片扩展成更大规模的系统。

不同尺寸的SpiNNaker系统是由一个或多个48节点PCB板组装而成的。每个48节点PCB板由48个SpiNNaker芯片之间有线连接，PCB和PCB之间使用高速串行链进行连接。最大的机器，包含超过一百万个ARM处理器内核，将在十个机房机柜中包含1200个48节点板，并将需要高达75 kW的电力（峰值）。SpiNNaker系统分为五代，从102到106机。最初的102机使用了约102个ARM核，是一个4节点板，用于训练、开发和移动机器人非常方便；103机则是上面提到的48节点板；计划2020年完成的106机则集成了约106个ARM核（1036800个），可实时精细仿真10亿生物神经元。

SpiNNaker软件可以分为在SpiNNaker系统本身运行的软件，以及在其他系统上运行的软件，其中一些可能与SpiNNaker进行交互。在SpiNNaker芯片上运行的大多数软件都是用C编写的。该软件可以分为控制软件（原始操作系统）和执行用户计算的应用软件。

SpiNNaker系统与外界的主要接口是以太网和基于IP的协议。每个SpiNNaker芯片都有一个以太网接口，每个PCB通常使用一个芯片来使用这个接口。这用于将代码和数据下载到SpiNNaker并从应用程序中收集结果。

小型SpiNNaker系统已被用于一系列实时应用，如机器人控制和视觉处理，也用于生物电路的非实时建模。迄今为止，还没有大型应用程序的例子。SpiNNaker系统目前已有被用来实时模拟神经网络，通过与Nengo接口，实现NEF（Neural Engineering Framework），对脉冲神经网络中的功能和动力系统进行编码，通过指定要计算的功能来创建神经网络。其中最值得注意的是250万个虚拟神经元的Spaun模型的实时执行，在一台有24GB的内存，一颗主频2.5GHz的四核处理器的电脑上运行这个模型，大概需要3小时才能模拟出Spaun 1秒的运行情况。

SpiNNaker系统有极高的灵活性和可编程性，但其大规模的系统的能耗相对非灵活类脑系统（如TrueNorth）较高，且采用ARM处理器作为核心过于简单，难于实现复杂模型与算法。

## BrainScaleS

BrainScaleS由德国海德堡大学卡尔海因茨·迈耶（Karlheinz Meier，1955-）教授负责[27-28]，前身是2005到2010年的FACTES项目和2011年到2015年的BrainScaleS项目。FACTES项目强调生物神经系统和神经形态计算之间的比较研究，微观层面研究神经元的信号处理特性及模拟电路实现，介观层面研究突触可塑性及数字电路实现。BrainScaleS项目旨在了解大脑信息处理中多空间和时间尺度的功能和相互作用。BrainScaleS的基本新方法在于体内生物实验和计算分析。 空间尺度范围从单个神经元到更大的神经元群体到整个功能性脑区域；时间尺度范围从与事件可塑性机制相关的毫秒级，到与学习和发展相关的数小时或数天。 在项目中将提取通用理论原则，以便人工合成皮层样认知技能。 为了这个目的，两者都将采用千万亿次超级计算机的数值模拟和基本不同的非冯诺依曼硬件架构。示范实验的规划和最终执行是BrainScaleS的关键方面。一种新型非冯诺依曼计算系统的工作重点是构建混合多尺度设备（HMF）的第一个完整的晶圆模块，包含以下组件的制造和测试：包含神经元和突触的模拟网络ASIC，用于主机和晶圆到晶片通信的数字网络ASIC，180nm CMOS技术的8英寸晶圆，为神经元提供通信通道的晶圆上的水平互连，包含FPGA的通信模块和组装系统的机电组件。

BrainScaleS的基本思路是使用超阈值模拟混合电路来实现神经元的物理模型，将生物神经元和电子电路中的离子电路做类比，在晶圆上集成超大规模突触，以降低通信代价，提高计算性能。这里用到的超阈值模拟电路比起斯坦福的Neurogrid用到的亚阈值电路要更快，速度比生物系统快10000倍。BrainScaleS使用晶圆级的集成，提供大量可以有效互连的模拟神经元，以适应10000倍的加速比。在BrainScaleS晶圆中，每片晶圆有48个reticle，其中每个reticle是八个HiCANN芯片，每个HiCANN芯片实现512个AdEx神经元模型和超过10万个突触，突触提供STDP学习功能。在8英寸的BrainSCaleS晶圆上实现了20万神经元和5千万突触，晶圆内总线速度达每秒1T脉冲，晶圆间分布式通信速度每秒10G脉冲。整个晶圆模块结构集成了晶圆，电源，通信和模拟支持电路。

BrainScaleS系统包含20个8英寸硅晶片，400万神经元和10亿可塑性突触。该系统不执行预编程代码，而是根据电子设备的物理特性进行演化，运行速度高达实时速度的一万倍。BrainScaleS系统中的主要通信层在晶圆内运行，高速串行通道每次将64个神经元的输出在连续时间从一个HiCANN芯片传输到另一个HiCANN芯片，如果两个同时输出的脉冲竞争相同的通道，则仅引入小的定时误差。这些高速通道通过交叉开关（其仅实现稀疏连接以减少通道上的电容负载）跨晶圆布线。为了支持跨多个晶圆的仿真，使用FPGA实现晶圆之间的高速串行通信以支持第二层通信。

BrainScaleS系统的极快的速度以为它很适合在生物学项目上花费很长时间的应用程序。比如说一些长期学习任务，像对童年发展的长达几年的建模，BrainScaleS的10000倍的加速比很可能使几年变成几个小时；不同的应用领域会带来非常大规模的参数搜索，其中高速批处理模式的操作在运行时方面具有直接的优势。

上面介绍的两大类脑计算系统SpiNNaker和BrainScaleS都得到了欧盟人类大脑计划HBP的支持。HBP的神经形态计算子项目有两个主要目标：一是构建两个大规模、独特的神经形态计算机器，而是制造下一代神经形态芯片。所要构建的大规模神经形态机器基于两个互补原则，其代表分别是SpiNNaker和BrainScaleS。这两个项目组都加入到HBP合作实验室中，并为两个系统的配置、操作和数据分析提供完整的软件支持；SpiNNaker系统实时运行，BrainScaleS作为10000倍加速比的加速系统。最近对神经科学和计算的研究表明，学习和发展是神经科学和认知计算的现实世界应用的关键方面。HBP这一子项目旨在通过专门的新颖硬件架构满足这一需求。 作为下一代大规模类脑计算机器基础的下一代神经形态芯片，SpiNNaker和BrainScaleS作为两个互补的方法，其第二代系统和芯片SpiNNaker2和BrainScaleS2的原型已经设计制作。两种芯片原型都都利用了固态制造技术的巨大进步，着重提高有效实施学习和发展的能力。

## 3.4 IBM TrueNorth系统

自2008年以来，IBM参与到美国国防部研究机构DARPA的SyNAPSE项目中，并于2014年开发了神经形态芯片TrueNorth芯片。SyNAPSE项目旨在实现一个非常密集且节能的平台，能够支持一系列的认知应用。2011年的时候，IBM首先推出了单核含256个神经元，256\*256个突触和256个轴突的芯片原型。当时的原型已经可以处理像玩Pong游戏这样复杂的任务。不过相对来说还是比较简单，从规模上来说，这样的单核脑容量仅相当于虫脑的水平。在2014年开发出来的TrueNorth芯片是一个28nm的CMOS芯片，实现技术为数字电路，每片集成了54亿个晶体管，拥有由4096个神经突触核心组成的片上网络，每核内有256个输入神经元和256个输出神经元，模拟的神经元模型为LIF模型；突触状态、神经元状态和参数、脉冲目的地址、轴突延迟等均用静态随机存储器记录。每片总计100万个神经元和2.56亿突触连接，单个芯片平均放电频率20Hz，单神经元放电功耗26pJ，实时操作功耗仅为70mW，比晶体管数量相当的传统CPU功耗低5000倍左右。

在内部异步和全局同步的情况下，TrueNorth通过二维网格上的地址事件表示（AER）通信来跨核发送神经尖峰事件。来自每个内核的神经元尖峰事件输出遵循单独配置的点对点路由，被发送到到另一个核作为输入，该核可以在同一个TrueNorth芯片上，也可以是另一个TrueNorth芯片。在神经元输出需要连接到两个或更多个神经突触核的地方，神经元被简单地复制在同一核内。每个核使用256个神经元有限状态机在自定时钟和一个256×256的二进制突触交叉阵列上实现256个脉冲神经元的LIF神经元模型的神经动力学。神经元输出的二进制突触通过单独可编程的8位因子进行缩放，从而在突触连接中提供更大的动态范围。TrueNorth还提供神经元级别的其他功能，例如在对神经突触内核编程时可以实例化不同的噪声来源和不同的神经激活函数。TrueNorth系统的行为是确定性的，但是有通过伪随机源产生随机行为的能力。这种随机性可以在软件仿真中精确复制。数字模型的确定性确保了所有的复制品都能产生相同的尖峰序列。

TrueNorth硬件由软件仿真器支持，软件仿真器利用硬件的确定性，可以依靠软件仿真器预测硬件的性能。突触的二进制性质使在线学习成为问题，所以在软件环境中进行的训练是脱机的。基于这款芯片，IBM建立了Corelet编程模型、算法库和相应的软件开发环境，结合Compass模拟器，用户可以快速尝试不同的模型和参数，从中找出优化的方案。

TrueNorth芯片可以直接连接在一起形成较大的系统，IBM已经开发了具有16个TrueNorth芯片的电路板，其中共包含1600万个神经元和40亿个突触。 可以通过将多个板连接在一起来组装较大的系统。

TrueNorth芯片的设计旨在解决从视觉（特别是使用AER视觉传感器）到听觉和多感官感觉融合这一范围内的问题。它为高维，多噪声的感官数据提供了低功耗的实时处理。像实时对象识别一类应用已被证明在TrueNorth上以极低的功耗水平运行。

## 3.5 其他

### 3.5.1 Neurogrid

米德1989年招收的博士生博阿汉（Kwabena Boahen）2005年加入斯坦福大学，成立了“硅脑”（Brains in Silicon）实验室，使用亚阈值模拟电路来模拟生物实时的神经元和突触动力学，在2009年研制出了神经形态电路板Neurogrid。Neurogrid使用多层数字分布和多级模拟聚合，构建具有数十亿突触连接的进行数字尖峰通信的神经形态系统。

Neurogrid系统包括用于配置和神经活动可视化的软件套件以及支持神经网络实时仿真的硬件平台。 硬件是一块包含16个Neurocore芯片和一些支持电路的电路板，其中每颗芯片集成了65536个亚阈值模拟神经元，一块Neurogrid板可以模拟100万个神经元和60亿个突触联结，能耗只有5瓦。每个Neurocore芯片都包括一个路由器，能够在其本地芯片，父芯片及其两个子芯片之间传送脉冲数据包。路由器支持多播树路由组织，其中脉冲数据被点对点传送到位于树中所有预期目的地之上的节点，然后到达所有目的地，在需要时被复制。

Neurogrid的实时操作特性使其适用于机器人控制，并且已经连接到机器人手臂，进一步的目标旨在利用该技术的低功耗性质开发可植入大脑的芯片来控制假肢，并开发无人机控制技术。

### 3.5.2 BrainStorm

来自斯坦福大学的Neurogrid的开发者Boahen研究组在2017年开发了新一代的神经形态芯片BrainStorm，这是一个百万神经元神经形态芯片装置，它将可以运行全脑模型。这一项目自 2013 年启动，由美国海军研究办公室资助，最后的成果将成为嵌入式应用和集群服务器上的计算芯片。

目前还没有相关论文解释该项目的细节，但Boahen指出Brainstorm 与其他已有神经形态芯片设计存在着很大不同；Boahen表示：“目前有很多神经形态设备正在使用超级计算机中使用的路由机制，就像网格一样。问题在于，在网格架构中你只能进行点对点信号传递。如果你想一次发出多个信号，系统就会锁死。”

Boahen 说 Brainstorm 是首个实现从高层次描述合成的脉冲神经网络的芯片，非 FPGA 编程将问题映射到设备上；这样的芯片能解决的一类问题是可被多维度非线形微分方程所数学描述的，或者说是基于现有的状态与输入如何随时间变化的一类问题。

### 3.5.3 天机

2015年06月，清华大学类脑计算研究中心首款类脑芯片—“天机芯”系列第一代芯片成功流片，该芯片为世界上首次将人工神经网络（Artificial Neural Networks, ANNs）和脉冲神经网络（Spiking Neural Networks，SNNs）进行异构融合，同时兼顾技术成熟并被广泛应用的深度学习模型与未来具有巨大前景的计算神经科学模型，可用于诸如图像处理、语音识别、目标跟踪等多种应用开发。在类脑“自行”车演示平台上，集成32个天机一号芯片，实现了面向视觉目标探测、感知、目标追踪、自适应姿态控制等任务的跨模态类脑信息处理实验。

# 分析与展望

## 类脑计算的优势

## 类脑计算的挑战

## 类脑计算的未来

参考文献：

1. H Markram, K Meier et al. *The Human Brain Project, A Report to the European Commission*. The HBP-PS Consortium, April 2012. 【技术报告】
2. 黄铁军. *人类能够制造出“超级大脑”吗？* .中华读书报：2015年1月7日5版. 【报纸】
3. Hodgkin A L, Huxley A. *A quantitative description of membrane current and its application to conduction and excitation in nerve*. The Journal of physiology, 1952, 117(4):500-544.
4. 饶毅,鲁白,梅林. *神经科学:脑研究的综合学科*. 《生理科学进展》1998年04期
5. M. Tsodyks, K. Pawelzik, H. Markram. *Neural networks with dynamic synapses*. Neural Computation, 10(4):821-835, 1998.
6. Guoqiang Bi, Muming Poo. *Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type.* J Neurosci 18: 10464-10472,1998.
7. Guoqiang Bi, Muming Poo.*Distributed synaptic modification in neural networks induced by patterned stimulation*. Nature, 401(6755): 792-796, 1999.
8. Sen Song, Kenneth D. Miller, L. F. Abbott. *Competitive Hebbian learning through spike-timing-dependent synaptic plasticity*. Nature Neuroscience 3:919-926, 2000.
9. Sen Song, L. F. Abbott. *Column and map development and cortical remapping through spike-timing dependent plasticity*. Neuron 32:339-50, 2001
10. J T Vogelstein, K Amunts et al. Grand Challenges for Global Brain Sciences. https://arxiv.org/ftp/arxiv/papers/1608/1608.06548.pdf, April 2016
11. R. Mizutani, R. Saiga, A. Takeuchi, K. Uesugi, Y. Suzuki. *Three-dimensional network of Drosophila brain hemisphere*. https://arxiv.org/ftp/arxiv/papers/1609/1609.02261.pdf. Sept 8, 2016
12. Gerald M. Edelman (1978). The Mindful Brain: Cortical Organization and the Group-selective Theory of Higher Brain Function. MIT Press. [ISBN](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [9780262050203](https://en.wikipedia.org/wiki/Special:BookSources/9780262050203).
13. Gerald M. Edelman (1987). Neural Darwinism. The Theory of Neuronal Group Selection (Basic Books, New York 1987). [ISBN 0-465-04934-6](https://en.wikipedia.org/wiki/Special:BookSources/0465049346)
14. G N Reeke, O Sporns, Gerald M. Edelman. *Synthetic neural modeling: the `Darwin' series of recognition automata.* Proceedings of the IEEE, vol. 78, no. 9, pp. 1498-1530, Sep 1990. doi: 10.1109/5.58327
15. Gerald M. Edelman. Learning in and from brain-based device [J]. Science, 2007, 318: 1103-1105.
16. Eugene M. Izhikevich and Gerald M. Edelman. Large-Scale Model of Mammalian Thalamocortical Systems. PNAS (2008) 105:3593-3598.
17. H. Markram. *The Blue Brain Project*. Nature Reviews Neuroscience, 2006, 7(2):153-160.
18. Dharmendra S. Modha et al(2011). *Cognitive Computing.* Communications of the ACM.(54)8:62-71
19. 欧盟人类大脑计划[EB/OL].[2016-07-25].www.humanbrainproject.eu/
20. Mead, Carver A.; Cohen, Shirley K. (July 17, 1996). *Interview with Carver A. Mead (1934–)*. Pasadena, California: Oral History Project, California Institute of Technology Archives.
21. Carver Mead. *Neuromorphic electronic systems*. Proceedings of the IEEE, vol. 78, no. 10, pp. 1629-1636, Oct 1990
22. Carver Mead(1989). *Analog VLSI and Neural Systems*. Addison-Wesley Pub. May 1989). ISBN: 0201059924
23. C Mead, M Ismail(ed). *Analog VLSI Implementation of Neural Systems.* (Dordrecht: Kluwer) 1989
24. B V Benjamin,…,K Boahen, Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations,*Proceedings of the IEEE,* vol 102, no 5, pp 699-716, 2014.
25. S B Furber et al. *The SpiNNaker project*. Proc. IEEE 102 652–65, 2014
26. A D Brown, S B Furber, J S Reeve, J D Garside, K J Dugan, L A Plana, S Temple. *SpiNNaker—Programming Model*. IEEE Transactions on Computers, vol.64, no. 6, pp. 1769-1782, June 2015
27. J Schemmel et al. *A wafer-scale neuromorphic hardware system for large-scale neural modeling*. Proc. Int. Symp. Circuits System pp 1947–50, 2010.
28. S Scholze S et al. *A 32 GBit s–1 communication SoC for a waferscale neuromorphic system Integr*. VLSI J. 45. 61–75, 2012.
29. K. Meier. *A mixed-signal universal neuromorphic computing system*. 2015 IEEE International Electron Devices Meeting (IEDM), pp. 4.6.1-4.6.4, 2015.
30. P A Merolla et al. *A million spiking-neuron integrated circuit with a scalable communication network and interface*. Science 345 668–73, 2014.
31. D B Strukov, G S Snider, D R Stewart, R S Williams. *The missing memristor found*. Nature, vol. 453, pp. 80-83, 2008.
32. T Tuma, A Pantazi, M L Gallo, A Sebastian, E Eleftheriou. *Stochastic phase-change neurons.* Nature Nanotechnology. doi:10.1038/nnano.2016.70, 2016
33. Zhongrui Wang, …, Joshua Yang. *Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing.* Nature Materials 16, 101–108 (2017) doi:10.1038/nmat4756.
34. Alexander N. Tait, …, Paul R. Prucnal. *Neuromorphic Silicon Photonics*. arXiv:1611.02272.
35. Steve Furber. *Large-scale neuromorphic computing systems.* Journal of Neural Engineering, 2016, 13(5):051001.
36. Catherine D. Schuman, …, James S. Plank. *A Survey of Neuromorphic Computing and Neural Networks in Hardware.* arXiv:1705.06963.
37. Ailamaki A, Alvandpour A, Amunts K, et al. *The human brain project: a report to the european commission.* Technical Report, 2012.
38. Eliasmith C. *To Build a Brain.* Oxford: Oxford University Press, 2013.