类脑计算的研究进展与发展趋势

目 录

[摘 要 3](#_Toc459523467)

[**Abstract** 3](#_Toc459523468)

[**1 引言** 3](#_Toc459523469)

[**2 类脑计算：从理念与实践** 5](#_Toc459523470)

[**2.1 人工智能的四条路线** 5](#_Toc459523471)

[**2.2 类脑“新思维”** 8](#_Toc459523472)

[**2.3 类脑计算机** 10](#_Toc459523473)

[**3 国际研究现状** 13](#_Toc459523474)

[**3.1 各国“脑计划”** 13](#_Toc459523475)

[**3.1.1 美国脑计划** 13](#_Toc459523476)

[**3.1.2 欧洲人类大脑计划** 15](#_Toc459523477)

[**3.1.3 其他国家和地区的脑计划** 16](#_Toc459523478)

[**3.2 神经形态器件** 16](#_Toc459523479)

[**3.2.1 硅基神经形态器件** 17](#_Toc459523480)

[**3.2.2 新型神经形态器件** 18](#_Toc459523481)

[**3.3 神经网络芯片** 19](#_Toc459523482)

[**3.4 类脑计算模型** 21](#_Toc459523483)

[**3.4.1 脉冲神经元模型** 22](#_Toc459523484)

[**3.4.2 神经系统信息编码** 26](#_Toc459523485)

[**3.4.3 脉冲神经网络学习与训练** 27](#_Toc459523486)

[**3.4.4 神经形态认知计算架构** 29](#_Toc459523487)

[**3.5 仿生感知与类脑应用** 30](#_Toc459523488)

[**3.5.1 仿生感知** 31](#_Toc459523489)

[**3.5.2 基于脉冲时间的视觉识别** 32](#_Toc459523490)

[**3.5.3 基于脉冲神经网络的语音识别** 33](#_Toc459523491)

[**3.5.4 机器人仿脑空间感知和记忆** 34](#_Toc459523492)

[**4 国内研究现状** 35](#_Toc459523493)

[**4.1 我国“脑计划”与相关研究** 35](#_Toc459523494)

[**4.1.1 我国脑科学研究与“中国脑计划”概况** 35](#_Toc459523495)

[**4.1.2 部分省市 “脑计划”** 35](#_Toc459523496)

[**4.1.3 我国类脑计算相关研究机构** 36](#_Toc459523497)

[**4.2 神经形态器件** 37](#_Toc459523498)

[**4.3 神经网络芯片** 38](#_Toc459523499)

[**4.3.1 人工神经网络芯片** 38](#_Toc459523500)

[**4.3.2 脉冲神经网络芯片** 39](#_Toc459523501)

[**4.3.3 视觉处理芯片** 40](#_Toc459523502)

[**4.4 类脑计算模型与应用** 41](#_Toc459523503)

[**5 国内外研究进展比较** 43](#_Toc459523504)

[**5.1 神经形态器件** 43](#_Toc459523505)

[**5.2 神经网络芯片** 43](#_Toc459523506)

[**5.3 类脑计算模型与应用** 44](#_Toc459523507)

[**6 发展趋势与展望** 45](#_Toc459523508)

[**6.1 神经形态器件** 45](#_Toc459523509)

[**6.2 神经网络芯片** 46](#_Toc459523510)

[**6.3 类脑计算模型与应用** 46](#_Toc459523511)

[**致谢** 47](#_Toc459523512)

[**参考文献** 49](#_Toc459523513)

[**作者简介** 56](#_Toc459523514)

类脑计算的研究进展与发展趋势

中国计算机学会多媒体专业委员会

摘 要

类脑计算，是指仿真、模拟和借鉴大脑神经网络结构和信息处理过程的装置、模型和方法，其目标是制造类脑计算机和类脑智能。与经典人工智能符号主义、连接主义、行为主义以及机器学习的统计主义这些技术路线不同，类脑计算采取仿真主义：结构层次模仿脑（非冯·诺依曼体系结构），器件层次逼近脑（模拟神经元和神经突触的神经形态器件），智能层次超越脑（主要靠自主学习训练而不是人工编程）。类脑计算相关研究已经有二十多年的历史，本报告从脑科学相关研究、模拟生物神经元和神经突触的神经形态器件、神经网络芯片、类脑计算模型与应用等方面对国内外研究进展和面临的挑战进行介绍，并对未来的发展趋势进行展望。

**关键词**：类脑计算；类脑智能；神经形态器件；神经网络芯片；脉冲神经网络；脑计划

**Abstract**

Brain-like computing is the device, model or methodology which emulates, simulates or analogize the structure of the brain’s neural system and its information processing mechanism, with the goal to produce the brain-like computer and brain-like intelligence. Different from the classical artificial intelligence methodologies, including the symbolism, connectionism, behaviourism and statisticsism, brain-like computing follows the analogism or mimicism: mimics the brain at structure level (non von Neumann architecture), approximates the brain at device level (neuromorphic devices emulating the biological neurons and synapses), and surpasses the brain at intelligent level (mainly by autonomous learning and training rather than manual programming). This report reviews the progress and challenges of brain-like computing in the last two decades, in the world and in China, including related research in the brain science, the neuromorphic devices, neural network chips, brain inspired computing models and applications. The future development trend is also prospected.

**Keywords:** Brain-like computing, brain-inspired intelligence, neuromorphic device, neural network chip, spiking neural network, Brain Project

**1 引言**

长期以来，制造智能机器一直是人类的重大梦想，计算机是承载这一梦想的最新载体。现代计算机刚刚登上历史舞台的1956年，人工智能学科就诞生了。六十年来，人工智能经历了三次浪潮，基本思想可大致划分为四个流派：符号主义、连接主义、行为主义和统计主义。这四个流派从不同侧面抓住了智能的部分特征，在“制造”人工智能方面都取得了里程碑式的成就。近年来，深度学习风起云涌，在图像和语音识别等领域取得突破性进展，人工智能迎来第三次浪潮。

特别地，阿尔法围棋（AlphaGo）综合深度学习（连接主义）、特征匹配和线性回归（统计主义）、蒙特卡洛搜索（符号主义）和强化学习（行为主义）思想，利用高性能计算（CPU+GPU）和大数据（16万局人类对弈及三千万局自我博弈），2016年3月一举战胜围棋九段高手李世石，最近已经名列世界职业围棋第一位。机器智能的快速提升吸引了全球目光，世界各国纷纷推出政策或计划推进相关研究，产业界投入也急剧攀升，机器智能将超越人类甚至毁灭人类等伦理问题再次成为媒体热议话题，也成为学界关注和研究的严肃问题。

靠四条经典路线甚至“四剑合璧”就能设计出达到人类水平的类人智能或者具有自我意识的强人工智能吗？回答是否定的，即使有更高性能的计算平台和更大规模的大数据助力，也还只是量变，不是质变，其根本原因在于：传统人工智能的思维定式是在“理解智能”机理的基础上设计制造人工智能系统，而人类对自身智能的认识还处在初级阶段。在对人类智能的理解还极其有限的情况下，把人工智能和人类智能相提并论，这种对比本身就存在问题。在这种思维定式下，缺乏“跨越人工智能到类人智能鸿沟”的逻辑通道：类人智能彼岸模糊一片，跨到哪里都不清楚，却又要在“理解智能”的基础上设计类人智能，这实际上是把解决“制造智能”问题的前提设定为“理解智能”这个更难的问题，犯了本末倒置的错误。

然而，如果我们跳出传统思维定式的窠臼，先把对功能意义上“智能”的孜孜追求稍稍放松，回顾一下生物智能和人类智能的产生及其进化发展过程，就会发现通向“类人智能”还有一条“新”路线，这里称为“仿真主义”（Analogism或Mimicism）。说这是一条新路线，是因为它反转了“理解智能”和“制造智能”的前后关系：不再寻求先“理解智能”（因为这是何时能够解决都还不清楚的终极性难题），而是通过先进的大脑探测工具从结构上解析大脑（或称大脑“反向工程”），再利用工程技术手段“照葫芦画瓢”式地构造出仿脑装置，最后通过环境刺激和交互训练仿真大脑，实现类人智能，简言之，“先结构，后功能”。这里的“新（路线）”之所以加了引号，是因为“先结构，后功能”是数千年来人类工程技术实践惯用的路线，而“凡事先找科学原理”的思维定式是过去数十年才形成的僵化误区。

仿真主义可以说是符号主义、连接主义、行为主义和统计主义之后的第五个流派，但也和前四个流派有个千丝万缕的联系。特别地，仿真主义可以说是走向极致的连接主义：神经元、神经突触和神经环路都尽可能逼近生物对应物，从而从基元到网络都重现生物对应层次的功能。传统人工神经网络尽管也是连接主义，但是做法相反：靠简化的神经元模型和人工设计的网络结构尝试产生复杂功能，尽快也能实现某些功能，但并不知道什么时候才能产生人脑级别的真正智能。而仿真主义先从结构层次逼近生物神经网络，因而产生类脑功能的机会更大，成功之后再试图理解功能发生的原理，理解之后再考虑系统简化和优化问题，这和人类自古以来的工程技术实践是一脉相承的。

沿着仿真主义的路线实现类人智能，需要设计开发新的智能载体。经典计算机是数理逻辑的开关电路实现，采用冯·诺依曼体系结构，可以作为逻辑推理等特定智能实现载体，但要靠经典计算实现类脑智能，无异于缘木求鱼。实际上，冯·诺依曼提出这一结构时也没期望用这个结构仿真大脑，相反，冯·诺依曼倒是认真考虑了根据大脑神经网络结构设计新型计算机。尽管神经网络和现代计算机的概念几乎同期出现，但在摩尔定律作用下，经典计算机性能六十多年来保持指数增长，给人以无所不能的假象，掩盖了神经网络思想的光芒。实际上，即使用现在最强大的超级计算机来仿真人类大脑，也至少还存在两个数量级的差异，在功耗方面更是遥不可及。因此，要按仿真主义的路线“仿脑”，就必须设计全新的软硬件系统，这就是“类脑计算机”，或者更准确地称为“仿脑机”。

本报告将首先概要介绍“仿脑”（类脑计算）背后的理念和类脑计算机开发实践，然后介绍类脑计算国际国内发展现状并进行对比分析和展望，希望能为推动我国相关研究开发工作尽一臂之力。报告主体内容分为四个部分：

1. 国内外“脑计划”：大脑解析仿真为类脑计算提供神经科学基础，是世界各国“脑计划”的重要内容，这部分将介绍美国、欧洲、日韩和我国以及我国相关省市“脑计划”的基本情况，这是影响类脑计算未来发展的重要背景。
2. 神经形态器件：模拟生物神经元和神经突触等生物神经网络基本单元的构造、功能与行为的光电器件称为神经形态器件，其特征尺寸应该与生物对应物相当甚至更小，从而为研制高密度、超低功耗的仿脑神经网络奠定基础。
3. 神经网络芯片：基于神经网络的智能系统要走向普及，芯片是必要一环。传统人工神经网络和深度学习需要高效率、高能效的人工神经网络芯片，类脑计算机则需要脉冲神经网络芯片，面向视觉等专用功能的专用芯片也大有用武之地。目前可以采用传统集成电路工艺，未来需要基于神经形态器件。
4. 类脑计算模型与应用：模拟仿真生物大脑皮层的神经回路结构和功能是手段，实现类脑功能才是类脑计算的最终目的。本部分将介绍类脑计算模型，包括神经信息编码，突触学习算法，编码和学习集成模型，学习、记忆和具身认知。另外，还将介绍类脑计算在仿生感知、图像识别、语音识别和空间感知等方面的应用。

**2 类脑计算：从理念与实践**

**2.1 人工智能的四条路线**

长期以来，制造智能机器一直是人类的重大梦想，计算机是承载这一梦想的最新载体。计算机在计算、逻辑推理、信息存储和检索等方面的性能早已远超人类，因而也被形象地称为“电脑”。在现代计算机发展初期，人们就开始了尝试以它为平台“制造”智能。1956年夏，达特茅斯学院的约翰·麦卡锡(John McCarthy，1927-2011)、哈佛大学的马文·明斯基(Marvin Lee Minsky,1927-2016)、IBM的纳撒尼尔·罗彻斯特 (Nathaniel Rochester，1919-2001)、贝尔实验室的克劳德·艾尔伍德·香农(Claude Elwood Shannon，1916-2001)四人发起在达特茅斯举行“用机器模拟人类智能”研讨会[1]，为期两个月，邀请数学、神经生理学、精神病学、心理学、信息论和计算机科学领域的10名学者参加，其中包括艾伦·纽厄尔（Allen Newell，1927-1992）和赫伯特·西蒙(Herbert A．Simon，1916～2001)，会议认为“学习或者智能的任何其他特性的每一个方面都应能被精确地加以描述，使得机器可以对其进行模拟”，并首次提出了“人工智能”这一术语。

经典人工智能主张实现人工智能应从功能模拟入手，用逻辑来建立人工智能的统一理论体系，通过分析人类认知系统所具备的功能和机能，用计算机模拟这些功能。这一流派由于将智能视为符号处理过程，采用形式逻辑实现智能，故称为“符号主义（Symbolism）”或“逻辑主义(Logicism)”。符号主义学派初期过于乐观，甚至估计“二十年内，机器将能完成人能做到的一切工作”，但直到四十年后的1997年，“深蓝”才战胜国际象棋冠军卡斯帕罗夫。符号主义对能够形式化表达的问题是有效的，但很多事物（包括大多数人类常识）并不能显式表示，而且即使形式化了，也还存在与物理世界的对应问题。视听觉等智能的基本特征，看起来不如逻辑推理“高级”，但符号主义至今难以有效应对，想象、情感、直觉和创造等人脑特有的认知能力，符号主义更是遥不可及。

人工神经网络是经典人工智能的第二条路线：既然人脑智能是由神经网络产生的，那就通过人工方式构造神经网络，进而产生智能。这一流派强调智能活动是由大量简单的单元通过复杂的相互连接后并行运行的结果，因而称为“连接主义”（Connectionism）。人工神经网络的思想可追溯到1943年神经生理学家沃伦·麦克洛奇（Warren Sturgis McCulloch，1898-1969）和数理逻辑学家沃尔特·哈利·皮兹（Walter Harry Pitts，1923-1969）提出的神经元数学模型。在1956 年达特茅斯会议提出“人工智能” 概念一年之前，在洛杉矶召开的美国西部计算机联合大会(Western Joint Computer Conference) 上已经展开了“学习机讨论会”（Session on Learning Machine），讨论会的参加者中有两位参加了第二年的达特茅斯会议，他们是奥利弗·赛弗里奇(Oliver Selfridge) 和艾伦·纽厄尔(Alan Newell)。赛弗里奇发表了一篇关于模式识别的文章，而纽厄尔则探讨了计算机是否能下棋，分别代表两派观点。讨论会的主持人正是沃尔特·哈利·皮兹，他在总结时说“（一派人）企图模拟神经系统，而纽厄尔则企图模拟心智，……，但殊途同归”，这为随后数十年人工智能“结构”与“功能”两条路线的交织斗争埋下了伏笔[2]。

人工智能的第三大流派是“进化主义”（Evolutionism），也称行为主义（Behaviourism），其思想源头是控制论，认为智能并不只是来自计算引擎，而且也来自环境世界的情景、感应器内的信号转换以及机器人和环境的相互作用。行为主义的代表人物是麻省理工学院（MIT）的罗德尼·布鲁克斯（Rodney Brooks），他领导研制的各种机器人从上世纪90年代开始就走出实验室，登上火星，进入军用和家庭（吸尘和洗地），近年来万众瞩目的机器大狗BigDog也是这一流派的力作，由MIT另一名教授马克·雷波特（Marc Raibert）领导。

机器学习把人工智能的重心从如何“制造”智能转移到如何“习得”智能，专门研究机器怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。机器学习有很多分支，其中部分与人工智能三大流派的基本思想有千丝万缕的联系，例如强化学习与行为主义、深度学习与多层神经网络。统计学习是机器学习十分重要的一部分，它基于数据构建概率统计模型并运用模型对数据进行预测和分析，因而被称为“贝叶斯派”（Bayesians）”。相对于经典人工智能三大流派可视为第四个流派：统计主义（Statisticsism）。

新世纪以来，四大流派呈现交叉融合的新趋势，特别是深度学习的出现和发展，开启了人工智能的第三次浪潮。比人工智能前两次浪潮幸运的是，深度学习的出现恰逢其时。一方面，在摩尔定律数十年来持续翻番的指数作用下，计算机性能已经大幅提升，构造大规模神经网络具备了技术可行性；另一方面，计算机的广泛应用产生了大规模的数据，大数据浪潮亟待计算机自动从数据中发现规律。

在深度学习基础上，谷歌2014年收购的DeepMind团队开发了阿尔法围棋（AlphaGo），2016年3月战胜九段围棋选手。AlphaGo可以说是四种经典人工智能路线的集大成者：(1)采用了蒙特卡洛树搜索（Monte Carlo Tree Search）这一符号主义的经典算法，但是围棋搜索空间约250150，已知宇宙中的原子数才1080，不可能蛮力穷举，必须大幅度缩小搜索空间才能增加搜索深度，这就需要根据当前棋局评估下一手的模型，AlphaGo称之为“策略网络”（Policy Network）；(2)深度学习在图像识别方面取得了巨大成功，AlphaGo把围棋盘面视为一个19×19的图像，根据围棋常识定义各像素的特征，采用卷积神经网络，从围棋对战平台KGS获得16万局6-9段人类选手的对弈棋谱（共计3000万手），训练出“有监督学习策略网络”（Supervised Learning policy network），学到人类“棋感”；(3) 强化学习：人类下的每一手不一定导致最终输赢，为了找到对最终胜负起关键作用的“妙招”，AlphaGo通过自我博弈产生了3000万手（为了防止过拟合，来自3000万局不同的自我博弈棋局），作为输入数据训练出“估值网络”（Value Network），能够根据当前局面，判断白棋还是黑棋取胜的概率大；(4) 快速策略（Fast Policy）：神经网络效果好但速度慢，为了在有限时间内增加搜索深度，AlphaGo同时采用了传统的局部特征匹配和线性回归方法，虽然这是统计机器学习中最基本的方法，但对AlphaGo整体性能贡献很大。上述四个方面分别源自符号主义、连接主义、行为主义和统计主义，四者有机结合造就了AlphaGo的成功，这四个方面都还有提升空间， AlphaGo未来版本会更强大。

四条经典技术路线“四剑合璧”，再加上高性能计算和大数据的主流，已经成为 “制造”智能的新范式。那么，这种“制造”智能的新范式真能制造出“类人智能”吗？

一方面，人类会设计出更多更强大的人工智能系统。智能问题的突破和发展，会带来一系列问题的解决，给人类社会带来翻天覆地的变化。符号主义、连接主义、行为主义以及统计主义都从某个侧面抓住了智能的部分特征，这些技术路线的有机结合能够制造出更强的智能，就像AlphaGo战胜人类围棋高手一样，未来的人工智能系统在某些特定功能上可以把人类远远甩在后面，一定会出现更多更强大的人工智能系统。

另一方面，基于经典技术路线人工智能（包括将来的）仍然是“弱人工智能”，它可以在某些方面远远超越人类，但和人类智能存在天壤之别。人类智能是活体大脑表现出来的特有功能，我们似乎对自己的意识和智能很熟悉，但实际上背后的发生知之甚少，如果与物理学相比，脑科学处在牛顿之前的时期，甚至有的脑科学家认为人类永远不可能揭不开意识之谜。因此，在“人类智能”的内涵还远远不清楚的情况下，将人工智能与之相提并论，这种对比本身就存在问题。

传统人工智能的符号主义、统计主义和行为主义，本质上都是基于对智能的部分外在了解而进行的功能模拟。计算能力和逻辑推理是一种智能，计算机模拟的是演算和推导过程，符号主义是这一思路的自然延续。人类可以从经验中学习，机器也可以从数据中统计出规律，而且可以比人做得更快更准，面对浩如烟海的“大数据”，机器可以不知疲倦的学习下去。生物拥有感官和身体，通过与环境交互进化成长，行为主义模拟这个过程，开拓出“制造”智能的新空间。但是，上述对智能的理解和模拟都是针对“智能冰山”浮在水面上的部分，“智能冰山”水面下的部分我们还一无所知，因此也就谈不上功能模拟。

**2.2 类脑“新思维”**

理解意识现象和功能背后的发生机理（简称“理解智能”）是人类的终极性问题，也是挑战人类知识极限的“科学圣杯”。相应地，制造类似人脑的具有自我意识的智能机器（简称“制造智能”），是工程技术领域的“技术圣杯”。如果“理解智能”这个问题能够解决，“制造智能”也就不在话下，但是，“理解智能”这个终极性问题到底数十年还是数百年亦或数千年才能解决？甚至永远不能解决？都还是未知数。因此，把“制造智能”寄希望于“理解智能”，实质上把解决问题的方案建立在解决另一个更难问题的基础上，犯了本末倒置的错误。

那么，在智能（意识）发生机制最终得到科学解释之前，能够制造出类人智能呢？回答是肯定的，而且制造出这样的智能机器，可能正是揭开意识之谜最便捷的途径！[3]

大脑和意识的关系，机器和智能关系，正如飞机和飞行的关系，或者再简单一点，就像胶水和粘连、竹笛和笛声的关系。请问，制作胶水，是否需要先搞清楚“粘连”的化学原理？制作一支竹笛，是否要先解出笛腔内发出优美笛声的动力学方程？显然不是，千百年前甚至今天的制笛工匠也未必明白空气动力学，这一点恐怕大家都不反对。但要接着问，制造飞机是否必须学习空气动力学？回答可能就不那么干脆了。今天的飞机设计师当然要学习空气动力学，但1903年莱特兄弟发明飞机时，却无空气动力学作为指南，他们借鉴鸟类滑翔，设计出第一架飞机。1908年冯·卡门专赴巴黎，亲眼目睹飞行表演后才相信机器真的可以飞上天，并下决心“要不惜一切努力去研究风以及在风中飞行的全部奥秘。”30多年后，他才和学生钱学森建立起空气动力学，回答了飞机为什么会飞这个问题。

“理解智能”和“制造智能”的关系也是如此。要实现真正的类人智能，首先要分清大脑的结构（主要是皮层神经网络）和大脑的功能（智能、意识）这两个层次，尽管目标是实现智能这个功能，但不要就功能而论功能，而要回到在结构层次，尝试先制造出具有同样结构的机器，再测试是否能够产生预期功能。自古以来人类的很多工程实践都是采用这种技术路线，小到一支竹笛，大到一架飞机，都是如此。

人工智能的发展历史也是“功能主义”和“结构主义”交织斗争的历史。人工神经网络的路线方向是正确的，这也是深度学习能够成功的重要原因，但是至今为止的人工神经网络都过度简化。人脑拥有数百种、上千亿个神经元（1011），每个神经元通过数千甚至上万个神经突触和其他神经元相连接（连接数量达到1014）, 尽管结构复杂，但仍然是一个复杂度有限的物理结构，采用神经科学实验手段，从分子生物学和细胞生物学层次解析大脑神经元和突触的物理化学特性，理解神经元和突触的信号加工和信息处理特性，并无突破不了的技术障碍。2008年，美国工程院把“大脑反向工程”列为本世纪14个重大工程问题之一，近年来各国“脑计划”也将大脑高精度解析列为重要内容进行支持。因此，我们不应再纠结于意识原理是否突破，而应该在大脑解析的最新进展基础上，争取在结构仿脑层面先行突破。一旦仿脑机器能够产生部分功能，揭示大脑智能的奥秘也就为期不远了。

分清了结构仿脑和功能类脑这两个层次，也就很容易理解“类脑计算”的概念了，本文所述类脑计算，是指仿真、模拟和借鉴大脑神经网络结构和信息处理过程的装置、模型和方法，其目标是制造类脑计算机，或称“仿脑机”（Brain-like Machinery）。类脑计算机的“类”是从结构层次仿真入手，即采用光电微纳器件模拟生物神经元和神经突触的信息处理功能，网络结构仿照大脑神经网络。在仿真精度达到一定范围后，类脑计算机将具备生物大脑类似的信息处理功能和系统行为（包括自我意识和“灵感涌现”等高级智能）。简言之，类脑计算机不是等待理解智能或心智的机理后再进行模拟，而是绕过这个更为困难的科学问题，通过结构仿真等工程技术手段间接达到功能模拟的目的。

类脑计算的技术路线总体上可分为三个层次[4]：结构层次模仿脑，器件层次逼近脑，智能层次超越脑。近十年来，国内外在这三个层次已经取得不少阶段性成果。

所谓“结构层次模仿脑”，是将大脑作为一个物质和生理对象进行解析，获得基本单元（各类神经元和神经突触等）的功能及其连接关系（网络结构），这一阶段主要通过神经科学实验采用先进的分析探测技术完成。英国科学家生理学家霍奇金(Alan Hodgkin, 1914-1998)和赫胥黎(Andrew Huxley, 1917-2012)1952年提出了著名的Hodgkin-Huxley方程（HH方程），这一标准数学模型精确刻画了单个神经元放电的动力学过程，1963年获得诺贝尔奖。1998年，Tsodyks Misha和Henry Markram等人构建了突触计算模型，2005年又在瑞士洛桑联邦理工学院（EPFL）启动“蓝脑计划”，专注于逼近生物真实的神经元及其微环路精细建模。2013年1月，欧洲启动“人类大脑计划”，投入十亿欧元，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。不到三个月之后的4月2日，奥巴马高调宣布美国脑计划，计划投入45亿美元，12年内绘制出“人类大脑动态图”。相关进展表明，对人类大脑结构的解析有望在十年内获得重大突破。

所谓“器件层次逼近脑”，是指研制能够模拟神经元和神经突触功能的微纳光电器件，从而在有限的物理空间和功耗条件下构造出人脑规模的神经网络系统。这方面的代表性项目是美国国防先进研究项目局(DARPA)2008年启动的“神经形态自适应可塑性可扩展电子系统（简称SyNAPSE，即突触），其目标是研制出器件功能、规模与密度均与人类大脑皮层相当的电子装置，功耗为一千瓦（人脑为20瓦），IBM和多所大学获得一亿多美元资助。2014年8月7日，IBM在《科学》发表文章[6]，宣布研制成功TrueNorth神经形态芯片，内含一百万个神经元和2.56亿个突触，这项成果入选“2014年十大科学突破”。德国海德堡大学在神经形态芯片研制方面已有十多年积累，2015年3月，已在一个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”的类脑计算机已经成功运行，其神经元采用模拟电路实现，功能比IBM方案更接近生物神经元。由于大脑中突触数量是神经元的上万倍，因此突触模拟器件也成为国际竞争的热点，近期在忆阻器和光突触方面进展也很快。

类脑计算研究的目标是制造出类脑计算机。类脑计算机的硬件主体是大规模神经形态芯片，这种芯片主要包括神经元阵列和突触阵列两大部分，前者通过后者互联，一种典型联接结构是纵横交叉，使得一个神经元和上千乃至上万其他神经元联接，而且这种联接还可以是软件定义和调整的。类脑计算机基础软件除管理神经形态硬件外，主要实现各种神经网络到底层硬件器件阵列的映射，这里的“软件神经网络”可以复用生物大脑的局部甚至整体，也可以是经过优化的乃至全新设计的神经网络。

所谓“智能层次超越脑”，属于类脑计算机应用软件层次的问题，是指通过对类脑计算机进行信息刺激、训练和学习，使其产生与人脑类似的智能甚至涌现出自主意识，实现智能培育和进化。刺激源可以是虚拟环境，也可以来自现实环境的各种信息（例如互联网大数据）和信号（例如遍布全球的摄像头和各种物联网传感器），还可以是通过安装机器人身体在自然环境中探索和互动。这个层次，类脑计算采用的是行为主义（控制论）路线，只是智能平台换成了类脑计算机这个“可塑之才”，通过交互实现类脑神经网络的突触连接关系及连接强度调整，并实现学习、记忆、识别、会话、推理以及更高级的智能。

**2.3 类脑计算机**

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。

2005年前后，普遍认为摩尔定律在持续50年后，将于2020年左右走到尽头，人们开始反思冯·诺依曼体系结构的问题。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，因此，类脑计算替代经典计算机，也就计算技术发展的刚性需要。

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。

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 Machinery）[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，开始引起广泛关注。
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大学基于ARM开始研制支持脉冲神经网络的多核超级计算机SpiNNaker。
6. 2005年，欧盟启动FACETS (Fast Analog Computing with Emergent Transient States) 项目，由海德堡大学牵头研制基于模拟混合信号(Analog Mixed-Signal, AMS)的神经形态芯片。
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。
8. 2005年，瑞士洛桑联邦理工学院（EPFL）研究者Henry Markram与IBM合作启动了“蓝脑计划”（Blue Brain Project），在IBM Blue Gene/L超级计算机上开展尽可能逼近生物真实的大规模仿生神经网络模拟。
9. 2008年，惠普公司实现忆阻器（memristor）原型，能够模拟神经突触功能，并展示了首个忆阻器与硅材料的混合电路。全球人造突触热潮兴起。
10. 2011年，欧盟启动BrainScaleS (Brain-inspired multiscale computation in neuromorphic hybrid systems)项目，作为FACETS延续项目, 研发大规模并行类脑计算机。
11. 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。
12. 2013年，欧盟启动人脑计划（Human-Brain Project, HBP），由EPFL的Henry Markram牵头，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。
13. 2013年，美国启动BRAIN （Brain Research through Advancing Innovative Neurotechnologies）计划。BRAIN并不直接涉及类脑计算，但它将推动对于生物大脑的深入理解，这是类脑计算研究的重要基础。
14. 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，功耗只有70毫瓦，比半导体数量相当的传统CPU功耗低5000倍左右。实现的一个用于视觉对象检测的应用系统，包含三百万个神经元，功耗只有200mW。
15. 2015年3月，德国海德堡大学在一个8英寸硅片上集成了20万神经元和5000多万突触，采用这种“神经形态处理器”的类脑计算机成功运行。
16. 2016年3月，欧盟人类大脑计划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。
17. 2016年8月，IBM苏黎世研究院制造出脉冲神经元。与之前采用电阻、电容等模拟器件构造模拟神经元不同，这种人造神经元使用相变材料，特征尺寸达到纳米级别，未来可以小到14纳米。每个单元能稳定存储3比特数据，还能执行多种计算，如检测数据关联等，并且像生物神经元一样具有随机性。

近十年来，类脑计算相关研究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。就像1948年晶体管的发明引发计算机革命一样，纳米级别的人造突触和人造神经元的出现，有望掀起一场计算机的新革命，具有仿生视觉等部分类脑功能的装置不久有望进入应用，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。

**3 国际研究现状**

**3.1 各国“脑计划”**

### **3.1.1 美国脑计划**

1989年，美国在国防部等部门资助下，美国国立卫生研究院（NIH）和国家科学基金会（NSF）召集有关神经科学家、计算机科学家和预研项目负责人会议，集中讨论了“利用新的计算机技术构建脑的数据库和模型”问题。1992年，美国国立精神卫生研究院（NIMH）正式成立“人类脑计划”联邦协调委员会（FICC-HBP），当年得到了4个部15个联邦机构的联合资助，包括NIH、NSF、美国海军研究办公室（Office of Naval Research）、美国宇航局（NASA）和能源部等。1997年，“人类脑计划”正式在美国启动，20多所著名大学和研究所参加了这项科研大计划，30多位课题负责人获得了该计划项目的资助，该计划于2004年顺利结束。作为美国脑科学研究的主要资助机构，NIH在2004 年推出“神经科学研究蓝图(Blueprint for Neuroscience Research)”框架，年均投入超过55 亿美元，2011年公布“神经科学10年计划”。

2013年4月，美国总统[奥巴马](http://renwu.hexun.com/figure_873.shtml)宣布新的大脑研究计划BRAIN（Brain Research through Advancing Innovative Neurotechnologies，创新性神经技术推动的大脑研究）[10]，称要“通过10年努力绘制出完整的人脑活动图”。美国脑计划咨询委员会联合主管、神经生物学家科尼利亚·巴格曼曾对媒体说，最初获得资金时，科学家并不知道如何推进“脑计划”研究，后来达成了一个基本共识：第一步是技术创新和推动技术发展，只有这样才能推动神经科学向前发展；第二步是将这些新技术应用于探索大脑功能的工作机制。

NIH是美国脑计划预算最大的部门[11]，2014年6月率先发布“脑计划2025” 12年规划：前5年着重开发探知大脑的新技术，如功能性核磁共振、电子或光学探针、功能性纳米粒子、合成生物学技术，后5年力争用新技术实现脑科学的新发现，包括绘制堪比人类基因图谱的“人类大脑动态图”。计划在原有投入基础上（也就是每年50多亿美元），十年新增投入45亿美元，重点资助以下9个领域：

1. 大脑神经元和神经胶质细胞类型普查；
2. 创建大脑结构图谱；
3. 开发新的基于电极、光学、分子遗传学和纳米技术等的大型神经网络记录能力；
4. 开发基于光遗传学、药物遗传学、生物化学和电磁场等的神经回路操纵工具；
5. 利用虚拟现实、机器学习、微型录音设备等了解认知和行为的神经活动基础；
6. 加强神经科学实验研究与理论、建模、统计、计算等专业的结合；
7. 研究fMRI、DWI、EEG、MEG、PET等成像技术的细胞机制，提高人类大脑成像技术的空间分辨率和时间采样率；
8. 建立最高伦理道德标准的人体临床数据收集新机制；
9. 开设计算神经科学、统计学、影像学、电生理学、光遗传学等课程，加大神经科学技术知识的培训与传播。

除NIH外，美国脑计划还涉及多个部门或机构，从美国联邦政府2017财年“脑计划”相关预算可以大致看出这些机构的分工[12]。2017财年“脑计划”预算4.34亿美元，其中NIH预算1.95亿美元。能源部900万美元，重点发展新型生物传感器和监测设备以监测关键分子在大脑中的工作过程。DARPA 1.18亿美元，重点内容包括研发能够读写单个神经元功能的硬件原型等。NSF 7400万美元，用于设立“国家大脑观测站”，协调大规模国际脑研究项目，平衡在神经技术领域的多方投资[13]。美国情报高级研究计划署（IARPA）继续支持应用神经科学，IARPA在2015年1月曾发布大脑皮层网络机器智能（MICrONS）计划[14]，预算1亿美元，目标是对一立方毫米的大脑进行反向工程，分析大脑的计算方式，并运用这些发现改善机器学习和人工智能算法，项目规划五年。2016年3月，选定三个团队，负责人分别是哈佛大学的生物学家和计算科学家 David Cox，卡内基梅隆大学的计算科学家 Tai Sing Lee 和贝勒医学院神经科学家Andreas Tolias。

美国民间机构对BRAIN计划的投入几乎与公共机构投入相当[15]。艾伦（Allen）脑科学研究所每年将有超过6000万美元用于支持BRAIN 计划，主要致力于研究认知、决策和指挥行动的脑部活动。霍华休斯医学研究中心（Howard Hughes Medical Institute）每年投入3000万美元发展成像技术，并研究神经网络的信息存储和加工过程。科维理（Kavli）基金会未来十年每年投入4约00万美元，致力于研究脑部疾病发生的机制，并寻找治疗方法。 索尔克研究所（Salk Institute for Biological Studies）投入2800万美元，从单个基因到神经回路再到行为层次深入理解大脑。

### **3.1.2 欧洲人类大脑计划**

2013年1月，欧盟宣布开展为期十年的“人类脑计划”（Human Brain Project，HBP），提出将信息技术和生命科学结合，投入十亿欧元（其中一半左右由欧盟直接资助），作为未来新兴技术旗舰两大项目之一，十年内在认识脑、治疗脑疾病和类脑计算三个方面取得突破[16]。项目参加单位包括来自24个国家的112个单位，骨干研究人员180多人。

欧盟HBP的一个主要前期项目是2005年瑞士启动的“蓝脑计划”[17]，由瑞士洛桑联邦理工学院神经科学家马克拉姆（Henry Markram）负责，内容是以实验神经科学的数据为基础在IBM 蓝色基因超级计算机上实现仿真大脑，马克拉姆也担任HBP的负责人。

HBP涉及三大科学领域：一是神经科学，主要是收集和了解有关人脑的数据，并把它们集成起来建立模块做仿真，以更好地理解人脑功能：从基因到细胞、神经回路、认知以致行为层次；第二是医疗领域，要使用来自临床试验的大规模数据，更好地确定疾病指症和诊断方法；第三是计算领域，需要了解大脑工作原理，并且把这一原理用于开发新型计算机硬件，革新计算模式。HBP实施分三个阶段：最初两年半（“快速启动”阶段）重点是建立ICT（信息计算通信）平台基础版，以及策略性选择数据的传播；接下来的四年半（“运作阶段”）加强工作以推动产生策略性数据以及为平台添加新功能，同时展示平台对基础神经科学研究、医学应用以及未来计算技术的价值；最后三年（“可持续发展阶段”）继续上述活动，同时向可持续方向发展，确保所创造的能力和知识成为欧洲科学和工业的永久财富。

HBP共设置了13个子项目，其中5-10即为上述的ICP平台，具体包括：

1. 策略性鼠脑数据；
2. 策略性人脑数据；
3. 认知架构；
4. 脑研究数学与理论基础；
5. 神经信息学平台；
6. 大脑仿真平台；
7. 高性能计算平台；
8. 医疗信息学平台；
9. 神经形态计算平台；
10. 神经机器人平台；
11. 应用；
12. 伦理与社会；
13. 管理。

2014年，数百名科学家联名写信给欧盟抵制HBP计划，“这个耗资12亿欧元的大型计划没有得到妥善的管理，因此无法达成其模拟人脑内部运作的宏伟目标。”[18]其实，这些反对意见主要来自认知科学家，他们与马克拉姆以神经科学为基础进行大脑仿真的理念激烈冲突，子项目3被中途取消，另外一些神经科学家认为马克拉姆采用的HH神经元模型与真正的生物神经系统相距甚远，这种情况下去做大规模仿真对神经科学研究方面没有什么意义。我们认为，HBP的定位是基本合适的，认知科学和神经科学之间本来就存在巨大的鸿沟，系统仿真作为桥梁是弥补这一鸿沟的重要手段，对于认知科学和神经科学的发展是有利的，至于因此造成科研经费分配额度变化的问题，就超出科研本身了。不过，欧盟在2015年重启了因争议而被搁置的“人类大脑计划”，希望我国“脑计划”能够尽量避免这种问题。

### **3.1.3 其他国家和地区的脑计划**

日本1996年推出“脑科学时代”计划纲要，2008年启动“脑科学研究战略研究项目”(Strategic Research Program for Brain Sciences，SRPBS)。2009 年，日本科技理事会下设的脑科学委员会提出“融合脑”目标，并确定四大领域：脑科学与教育、社会，脑科学与身心健康，脑与信息产业以及基础技术开发。2014年9月11日，日本科学省宣布日本大脑研究计划Brain/MINDS (Brain Mapping by Integrated Neurotechnologies for Disease Studies)[19]，中心任务是通过对狨猴大脑的研究加快对人脑疾病的研究。日本脑计划分成三个组，第一小组由庆应大学Hideyuki Okano负责，用功能MRI等对大脑进行功能定位，并从细胞尺度对宏观大脑功能进行分析。第二小组由理化研究所大脑科学研究所Atsushi Miyawaki负责，这一小组分为17个独立小组，分别开发相关研究技术。东京大学Kiyoto Kasai负责第三小组，将收集和分析患者大脑影像等相关信息，将人类精神分裂症、神经疾病和脑血管疾病的信息与绒猴信息结合进行综合分析。

韩国1998年出台《脑科学研究促进法》，发布“脑科学研究推进计划（1998-2007）。2007年设立“第二轮脑科学研究推进计划（2008～2017）”，10年投资13.8亿美元。2016年5月30日，韩国科技信息与未来规划部发布《大脑科学发展战略》[20]，计划未来十年财政投入3400亿韩元（约合人民币19亿元），目标是到2023年发展成为脑研究新兴强国，计划在2023年前构建出大脑地图，并应用于脑部疾病治疗并促进人工智能技术发展。

此外，英国、法国、加拿大、俄罗斯、以色列等国也脑科学或神经科学相关研究计划。

**3.2 神经形态器件**

开发与神经网络算法相匹配的神经网络硬件系统，是一个全新的领域。现阶段，模拟生物神经元的行为与构造是主要研究方向，实现了仿生的电子器件被称为神经形态器件。人脑中，神经元是构成神经系统结构和功能的基本单位，它具有接受、整合和传递信息的功能，神经元之间由突触进行连接，神经元和连接突触构成了信息传递的基本通道与回路，被认为是神经系统的学习和适应等过程的关键环节。各种神经形态器件，即围绕神经元和突触，进行功能和行为上的模拟。基于人脑中大约有~1011个神经元和1014个突触，研制具有高密度和超低功耗的神经网络是该领域的主要目标。

### **3.2.1 硅基神经形态器件**

基于传统的成熟半导体工艺的神经元形态器件和突触形态器件首先被提出并应用于对神经网络的模拟。一般来说，硅基神经元形态器件主要由动态电导模块（即信息处理）、脉冲触发器、阈值和修复期模块、触发频率适应模块、以及树突轴突模块组成[2[1](#_ENREF_1)]。用硅基形态器件组成的神经网络系统，可以进行实时模拟，其速度并不依赖于神经元及其连接的数量，比计算机仿真能效高，适合实时及大规模的神经模拟。

在DARPA SyNAPSE的赞助下，IBM开发了一款基于数字电路的低功耗的模拟器，称为真北（TrueNorth）认知计算系统[2[2](#_ENREF_2)]。它由大量的神经突触核互相连接形成通信网络，每个核集成了计算（神经元）、记忆（突触）、核内通信（轴突）。每个核由“事件驱动”，可重配置，且功耗极低。其中，在经典的带泄漏积分触发神经元模型基础上，引入了如下几个方面的新功能：a）可配置、可再现的随机输入、状态和输出选项；b）可编程的4个泄漏积分模式；c）引入2个阈值模式，确定性和随机性的阈值，使得神经元即便积累同样的膜电位，可以呈现不同触发行为；d）提供6个不同的初始膜电位选项，实现不同的积分值触发。这些附加功能使该系统可以模拟50多种不同的神经元行为，包括Izhikevich提到过的触发神经元（spiking neuron）的20种最杰出的行为特征。此外，该系统中采用定点运算模式，通过简单的假发和复用的技术，避免了乘法/除法/取幂等复杂运算。另外，IBM的真北计算系统采用6个晶体管（6T）的静态随机存储器来代表突触的连接。当神经元进行信息汇总的时候，每个突触的数值由突触的连接状态、神经元参数和目标轴突经一定的权重加权共同决定。

曼彻斯特大学研究小组研制一款基于ARM处理器的芯片SpiNNaker，支持大规模的脉冲神经网络（SNNs）的仿真[2[3](#_ENREF_3), 2[4](#_ENREF_4)]。SpiNNake系统建立了了两种不同神经网络应用的映射，可实现简单的脉冲神经元模型和复杂的分层神经网络模型 （1输入层+1个输出层+1个隐藏层）；采用并行分布式处理器模型，每个神经元可以接受来自于前面层的许多神经元输入，经加权集成后输出。在该系统中，ARM选择了中等复杂度的Izhikevich模型，每个核可以仿真1000个触发神经元。突触是由128MB的移动式动态随机存取存储器（SDRAM）进行模拟，采用触发时差依赖式可塑性（STDP）的学习法则进行编程。

HP联合Boston大学共同开发了一款低功耗、全数字平台的硬件系统—Cog Ex Machina [2[5](#_ENREF_5), 2[6](#_ENREF_6)]。它可以建立与模拟真实环境实时交互的大脑模型。Cog用并行计算机线程和张量数据表示实现了类似生物的大脑，实现了诸如层级结构、并行处理。尽管该平台还没有实现智能行为以及生物规模的功耗与体积，但为研究者提供了一条成本相对较低地快速建模以验证新算法的途径。Cog系统采用计算更复杂、生物精度更高的Hodgkin-Huxley模型。

德国海德堡大学研制了一款名为“BrainScales”的神经网络模拟器[2[7](#_ENREF_7), 2[8](#_ENREF_8)]。它由模数混合的、多核阵列处理器组成，可以运行生物模型和抽象的神经网络算法，并具有优异的实时特性和微缩特性。

斯坦福大学开发了NeuroGrid系统，用以实时模拟的大尺度神经网络[2[9](#_ENREF_9), [30](#_ENREF_10)]。该系统基于数模混合电路构建，利用晶体管的亚阈值工作区间来实现突触和神经元的特殊功能。

另外，瑞士的苏黎世大学开发了一款定制的、模数混合的、大规模神经形态网络系统，用来探索计算神经科学的模型和开发人机交互系统[[31](#_ENREF_11), [32](#_ENREF_12)]。

### **3.2.2 新型神经形态器件**

硅基神经形态器件，特别是基于SRAM和DRAM的存储器，表现出较高的功耗，且在断电后，信息全部丢失。因此，研究人员将目光投向了新型的固态存储器。

**（1）基于忆阻器的突触形态器件**

惠普于2008年首度实现了由蔡少堂（Leon Chua）提出的忆阻器器件[[33](#_ENREF_13)]。该忆阻器器件使用二氧化钛（TiO2）作为阻变层，提出了一种基于金属/氧化物界面电势垒的改变和复原的物理模型[[34](#_ENREF_14)]。并首先提出利用忆阻器去模拟突触[[35](#_ENREF_15)]。展示了通过调制突触前、后神经元的脉冲宽度和频率，实现触发时差依赖可塑性（STDP）学习法则。跟之前的硅基突触相比，忆阻器突触可以更好的控制功耗，更简单的调制前后神经元的触发时差，减少突触集成面积和设计限制，为学习法则提供更多的自由度，以及可以对同一个或不同的神经元实施不同的学习法则。

文献[[36](#_ENREF_16)]使用掺银的非晶硅作为阻变层也展现了忆阻器特性，同时通过改变脉冲的宽度去模拟STDP学习法则。该忆阻器是是通过从阳离子漂移形成导电丝，从而产生阻值变化的存储器 （也被称作导电桥式阻抗存储器，CBRAM）。一般来说，它容易实现多阻态，且工作电压较低，但稳定性较差。

文献[[37](#_ENREF_17)]使用Ta2O5/TaOx双层阻变层实现忆阻器并展现STDP等学习法则。该忆阻器是依靠氧化还原反应，在氧化物的绝缘层中形成氧空穴导电丝的存储器。这类基于氧空穴导电丝的存储器件具有稳定性好，可擦写次数高，数据保存时间长和兼容当下半导体工艺等优点。

文献[[38](#_ENREF_18)]采用TiN/HfOx/AlOx/Pt的结构来实现忆阻器。可实现1000倍电阻变化，快编程速度~10ns，合理的擦写次数（105），同时展现了较低的功耗（6pJ），为大规模硬件实现神经网络打下基础。文献[[39](#_ENREF_19)]进一步提出了Pt/HfOx/TiOx/HfOx/TiOx/TiN多层结构，目的是使得Ti和Hf原子交互扩散，引起更多缺陷，从而导致更低的编程电压，甚至不需要forming过程。

文献[[40](#_ENREF_20)]利用氧化铁忆阻器，展现了良好的模拟器特性，可实现连续的阻态变化，拥有较好的数据保留能力。通过调制脉冲幅度，实现集时域（STDP）和频域（SRDP）学习法则于一体的复杂学习法则。

文献[[41](#_ENREF_21)]利用氧化钨忆阻器，在同一个器件上，同时模拟短期记忆和长期记忆，以及它们之间的转换。值得注意的是，在硅基形态器件中，从未有实现过这种短期记忆、长期记忆相互转换的行为。文献[[42](#_ENREF_22)]利用硫化银也实现了类似的行为,稀疏脉冲引出短时程可塑性（STP），而密集脉冲导致长时程可塑性（LTP）。

文献[[43](#_ENREF_23)]通过利用TiO2忆阻器短期记忆功能，对神经元的膜电压进行模拟、调制。

**（2）基于其它固态存储器的突触形态器件**

文献[[44](#_ENREF_24)]展示了基于相变存储器的突触形态器件，用较低的能耗演示了长期增强型和长期减弱型学习法则，并讨论了STDP的两种学习机制。同时，利用相变存储器进行Hopfield循环网络仿真，实现强健的联想学习和序列学习，并分析了器件阻值的变异对神经网络的影响[[45](#_ENREF_25)]。

惠普利用NbO2所具有的从绝缘体变到导体的莫特变换属性[[46](#_ENREF_26)]，开发了类神经元形态器件，用简单的电路即实现了对复杂的Hodgkin-Huxley神经元触发模型的模拟，展示重要的神经元all-or-nothing触发函数。

文献[[47](#_ENREF_27)] 利用自旋转移矩磁存储器(STT-MRAM) 非传统工作区间去模拟突触。该器件的架构与传统MRAM的略有不同，使耗电量更少，密度更高，具备快速读写和信息长久保存能力。提出利用不同的脉冲振幅和脉冲宽度，进行突触的编译操作，实现了简易版的STDP学习法则；并利用蒙特卡洛进行仿真，提出运用中等电流编程以平衡能耗和阻值异变。

**3.3 神经网络芯片**

随着智能计算和类脑计算的发展，神经网络芯片作为核心部件受到了学术界和工业界的广泛关注。神经网络芯片大致可以分为四大类：人工神经网络芯片、脉冲神经网络芯片、视觉处理芯片和类脑芯片工艺器件基础研究。人工神经网络芯片预期将在近期取得较广泛的实际应用进展；脉冲神经网络芯片尚处于探索性应用阶段；视觉处理芯片则专门用于完成图像和视频处理任务。

脉冲神经网络芯片的代表是IBM 2014年发布的“真北”（TrueNorth）[6]，其基本结构由硬件神经元和神经元之间的脉冲连接组成，硬件神经元接收输入脉冲，在累积到一定阈值后被激活产生输出脉冲。“真北”具有4096个处理核，每个内核包含256个硬件神经元，因此总共可以模拟100万个神经元和2.56亿个突触。其峰值性能达到了每秒266G定点运算速度。“真北”芯片包含有54亿晶体管，是迄今建造的最大的CMOS芯片之一。

表1列举了国际上具有代表性的类脑计算项目所采用的芯片或硬件平台[9]。基于大规模CPU 或GPU 集群的仿真系统未列入（例如瑞士基于IBM蓝色基因超级计算机的“蓝脑计划”与日本基于“京”超级计算机的仿真系统）。

表1国际上具有代表性的类脑计算项目及其所采用的芯片或硬件平台

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 项目/单位 | 核心芯片 | 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 | 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 | 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 | 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 | 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，瑞士苏黎世联邦理工学院 | 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 | 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，英国阿尔斯特大学/爱尔兰国立大学 | 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, 美国圣地亚哥加州大学 | 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, 英国阿尔斯特大学 | 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 |

说明：表中神经元模型和学习算法的进一步解释请见3.4.1和3.4.2节。

表中除了曼彻斯特大学的SpiNNaker系统是采用多核ARM平台运行软件，其它芯片均是基于硬件电路设计：数字电路或者模拟混合（Analog Mixed-Signal，AMS）。采用数字电路来仿真（Simulate）牵涉到微分方程的递归求解，因此所需运算能力与功耗较高。而采用模拟混合电路AMS设计，可以利用模拟电路的物理特性来直接模拟（Emulate）神经元的连续性动态行为，其运算效率与功耗远远优于基于数字电路的仿真（Simulation），因此大多数类脑芯片采用了AMS技术。如果想要做到超低功耗，那么则需要采用亚阈值AMS。IBM TrueNorth数字电路芯片之所以能做到超低功耗，应当归功于他们采用了异步电路设计技术来充分利用SNN的事件触发特性,使得只有接受脉冲时神经元才会激活，而其它神经元则可以处于睡眠状态。

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）系统的仿真速度比实时要快一万倍。他们的电路设计采用较高电压（非亚阈值），神经元模拟采用比LIF更为精确的AdEx模型，并且通过晶片级集成（wafer-scale integration）,即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，其主要应用目标是为神经科学家提供用于大规模SNN仿真加速的超级计算机，而非低功耗嵌入式系统。

一些芯片没有片上的在线学习功能，例如IBM TrueNorth，即神经元之间的连接权重在运行时是固定不可变的。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，然后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，但同时也局限了芯片的动态自适应能力，任何改动都需要人工干预与关机重启。

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。由于没有相关论文发表，并且配备Zeroth的产品尚未上市销售，因此大部分技术细节未知。从2013年尚处于研发阶段的媒体报道来看，是采用AMS实现LIF SNN模型；2015年，高通宣布将Zeroth用于蛟龙处理器产品（Snapdragon 820）中作为协处理器，其网站提到“Zeroth从基于SNN演化到同时支持ANN”，采用了数字电路实现基于ANN的深度卷积神经网，并且未来将会引入基于ANN的回归神经网络。目前尚不知道商业产品中的Zeroth是否真正支持SNN。

类脑计算的产业化还处于初步阶段。IBM TrueNorth是由DARPA提供资金支持研发的（共计5千3百万美元），IBM本身并没有任何资金投入。与IBM Watson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。其它大多数项目都是政府经费支持的学术研究，目前尚无成功的产业化案例。

忆阻器是具有记忆功能的非线性电阻器。虽然1971年就提出了忆阻器的概念，但直到2008年才由惠普公司创建了世界上首个忆阻器器件。由于忆阻器与生物神经元具有天然的相似性（具有记忆功能，能够同时提供存储和逻辑等），被认为是模拟神经元的重要器件。2015年，加州大学和纽约州立大学石溪分校首次用忆阻器创建出了神经网络芯片，该芯片通过无晶体管的金属氧化物忆阻器闩（Crossbars）创建，目前仅能够识别3×3像素黑白图像中的图案，距离真正人类智能还有较大差距。由于忆阻器是前沿技术，如果采用忆阻器来实现真正仿脑要求规模的神经网络，还需很长时间的努力。

**3.4 类脑计算模型**

类脑计算综合神经科学、认知科学和信息科学来探索生物神经系统是如何实现智能的，进而建造人工智能系统模拟生物神经系统。与传统人工智能、神经网络不同，类脑计算以大量的神经科学理论和实验结果为依据，包括感知信息处理、学习、认知与记忆等功能为一体的智能化计算平台[48]。与传统的人工神经网络（ANN）相比，类脑计算采用生物大脑所采用的脉冲神经网络，以异步的、事件驱动的方式进行工作，更易于在硬件上实现分布式计算与信息存储，能实时处理多感官跨模态等非精确、非结构化数据[49]。

经过近30年发展，虽然在神经电路的硬件模拟和简单类脑计算算法上，如视觉和听觉的信息处理和识别，取得了很多成就，但是在实现高层信息的表达和组织以及认知上，依然面临很多挑战。本部分将着重阐述研究现状、面临的问题、解决该问题的框架和算法原理。

### **3.4.1 脉冲神经元模型**

神经元建模可以有多种抽象层次，从最生物精确的Hodgkin–Huxley (HH) 模型，到最简化的Leaky Integrate & Fire（LIF）模型，有着多种介于二者之间的模型。如图 1 不同抽象层次的SNN模型与其所需运算能力[所示，越精确的模型，其运算复杂度也就越高，其中“# of FLOPS”代表仿真1ms的单个神经元运行所需的浮点运算操作数。我们介绍几个在类脑计算领域所常用的模型。

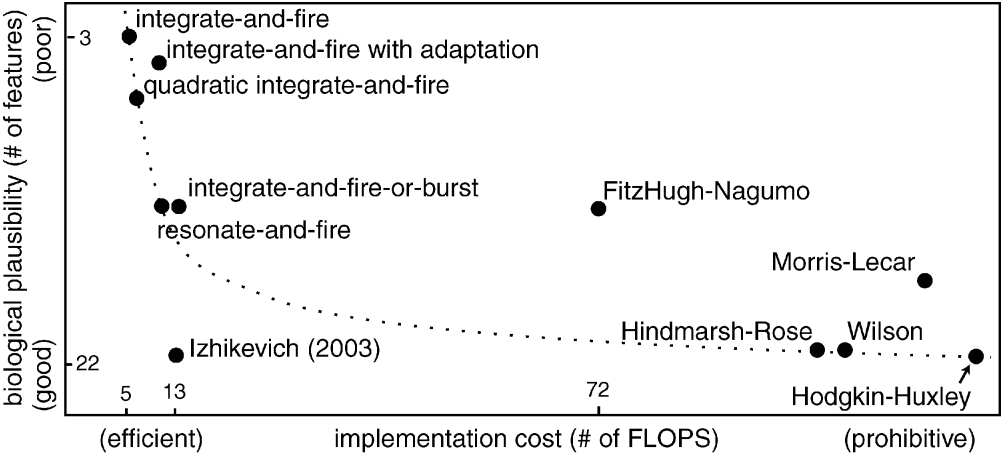


图 1 不同抽象层次的SNN模型与其所需运算能力[52]

**（1）Hodgkin–Huxley（HH）模型**

HH模型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[50]，由两位研究者Alan Lloyd Hodgkin与Andrew Huxley于1952年提出，获得了1963年诺贝尔生理学奖。



图 2 Hodgkin–Huxley（HH）模型电路图

图 2描述了HH模型所对应的电路图，其中C代表脂质双层（lipid bilayer）的电容; ，，分别代表钠离子通道，钾离子通道与漏电通道的电阻（，上的斜箭头表明其为随着时间变化的变量，而则为一个常数）；，，代表由于膜内外电离子浓度差别所导致的电化学电压，包括钠离子平衡电压，钾离子平衡电压，与漏电平衡电压；膜电压代表神经膜内外的电压差，可以通过HH模型的仿真来得到随时间变化的曲线。HH模型可以用于建模一个点状神经元，即用一组HH模型公式来描述整个神经元；也可以建模多个舱室（compartment）的详细神经元模型，即用一组HH模型公式来描述一个舱室，而一个完整的神经元则由很多舱室组成，不同舱室有不同的电化学特性，体现为HH模型中的不同参数。HH模型可以十分精确地重现出不同种类生物神经元的电生理实验数据，但是其运算量较高，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即采用HH模型建模，而且也并非实时仿真。



图 3 HH神经元膜电压随着输入脉冲的变化

图 3描述了一个典型HH神经元的膜电压。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的快速上升，称为兴奋性突触后电位（Excitatory PostSynaptic Potentials，EPSP）。如果输入脉冲之间的时间间隔较长，例如在时刻1与时刻22左右到达的2个脉冲，由于漏电通道的作用，如果没有新的输入脉冲，那么膜电压会随着时间逐渐降低至平衡电压。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，那么膜电压会上升至发放阈值（红色水平虚线所示）而触发一个输出脉冲。之后被重置为低于平衡电压的，称为脉冲后电压（spike-after potential），然后逐渐回升至平衡电压。（注意图3只是一个示意图，其中的各种参数并非生物精确的，例如生物神经元平衡电压一般为-70mV，而非图中所示的0mV。）

神经元的行为与输入的时间特性密切相关，类似一个巧合检测器（coincidence detector），即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使得其不会发生脉冲。这个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）通过双耳听到的声音信号的细微时间差别来定位声音来源，所需时间精度在数百微秒级别。

**（2）Leaky Integrate and Fire (LIF) 模型**

LIF神经元的建模有多个变种，其中最简单的一种，称为基于电流的LIF模型，其公式描述为[50]：, 



If {}

其中为膜电压；为膜电容；为漏电电导，即漏电电阻的倒数；为漏电平衡电压；为输入电流。可以看出，这就是一个基础电路理论中经典的RC电路。当膜电压升至发放阈值（firing threshold）时（），就会触发一个输出spike（又称为一个动作电位action potential)，发送给与它相连的下游神经元，即膜电压快速升高形成一个尖峰，然后立刻重置为（）。之后膜电压会有一个无反应期（refractory period），即对外界新的输入脉冲没有响应。LIF模型采用基于一个发放阈值参数的脉冲触发取代了HH模型中描述电导变量与变化的复杂公式，简化了运算公式，提高了运算效率，其代价是牺牲了较大的生物精确性。为了计算速度，到目前为止，LIF 仍然是类脑计算中最常用的模型。



图 4 LIF 神经元膜电压随着输入脉冲的变化

图 4描述了一个LIF神经元的膜电压。每个输入脉冲通过输入电流项触发膜电压的垂直上升（而非图 3所示的以一定坡度上升）。当膜电压会上升至发放阈值而触发一个输出脉冲后，被立刻重置为（图中设为等于平衡电压），并且在之后的有一个无反应期。可以看出，图 4是图 3的一种近似，有利于提高运算效率，尤其是基于数字电路的仿真。

**（3）Izhikevich模型**

由研究者Eugene M. Izhikevich从非线性动力系统的角度于2003年提出[52]，其模型结果可以呈现多种生物神经元的放电行为，运算复杂度则接近于LIF。公式描述为：



If () {}

其中的是一个辅助变量。通过调整a,b,c,d参数，Izhikevich模型可以重现出与HH模型十分类似的放电行为，但是，与HH模型中各参数具有明确的生理意义（诸如离子通道等）不同，这些模型参数不再具有十分明确的生理学意义。

**（4）Adaptive Exponential Integrate and Fire (AdEx IF) 模型**

AdEx IF模型是对Izhikevich模型的改进，并于2005年提出[51]与 Izhikevich模型类似，当膜电压升至发放阈值时，会触发一个脉冲后重置。但AdEx IF模型会同时减低膜电压的响应速度，称为脉冲触发适应（spike-triggered adaptation），其后果是在恒定电压刺激条件下，神经元的脉冲发放频率会逐渐降低,。可以直观地认为神经元在发出脉冲后会逐渐感到“疲倦”而反应变慢。这是AdEx模型的一个重要特征，从放电行为结果上更接近于HH模型的结果。

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，那么应当在运算能力允许的前提下尽量选用生物精确的模型，例如HH模型；但是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，那么则应当尽量选用简化的模型，例如LIF模型。

2009年，IBM研究者Dharmendra Modha采用包含147,456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1013个突触，与猫脑规模相当，获得了超级计算领域的Gordon Bell大奖，很多媒体报道IBM成功实现了猫大脑的仿真。牵头瑞士EPFL蓝脑项目的研究者Henry Markram写了一封公开信抨击Modha的工作，认为Izhikevich模型过于简单，他所采用的HH模型才是真正的生物神经网络仿真。但是后来Markram牵头的欧盟人脑计划（HBP）在2014至2015年受到数百名神经科学家通过给欧盟写公开信形式的抨击与抵制，认为他所采用的HH模型与真正的生物神经系统仍然相距甚远。Markram的研究目标是为神经科学家提供更快更大规模的仿真工具，用于更好的理解大脑与脑疾病，因此应当采用HH模型；而Modha的研究目标与Markram不同，是建造低功耗的智能信息处理系统（这也体现于后来他牵头研制的基于LIF模型的TrueNorth芯片），因此Izhikevich或LIF模型是较为合适的选择。

需要指出的是，生物神经网络的复杂程度远远超出本文所描述的这些模型，包括HH模型。下面举例阐述生物神经网络的复杂性[53]，而没有列出的还有很多：

1. 树突脉冲 （dendritic spike）：单个树突本身就可以产生脉冲并向神经元体（soma）传播，根据树突脉冲的发生的位置与时间，可以进一步触发或者不触发传统意义上的一个神经元脉冲。一个树突脉冲会影响附近树突的膜电压，使其都升至它们的最大输入电压，这个效果类似于卷积神经网中常用的max-pooling运算。一个神经元所具有的突触个数随着神经元的类型不同而差别很大，平均为一万个；有一种Purkinje神经元具有高达十万个突触，分布在一千个树突上。一个具有成千上万突触的神经元本身就类似一个传统意义上的神经网络，具备强大的信息处理能力。
2. 胶质细胞（glia cell）的作用：胶质细胞对神经元细胞提供一些辅助作用。神经科学研究近年来发现它们也会积极参与神经元之间的信息传递，影响神经网络的行为与结构，但是它被大多数类脑计算研究中的神经网络模型所忽略。
3. 神经递质触发的蛋白质信号通路（protein signaling cascade）：一个生物细胞，包括神经元细胞中，有大约一百亿个蛋白质。当一个突触的上游神经元发生脉冲时，其所释放的神经递质与下游神经元树突上的受体结合，会触发一系列的蛋白质化学反应，称为蛋白质信号通路，其效果是调整树突的电化学特性，使其具备一定的自适应与自学习能力。
4. 神经递质触发的基因调控(gene regulation): 一个生物体内的大部分细胞都具有同样的DNA,但是不同的神经细胞会具有与其它细胞不太一样的基因组。随着神经元每天都会改变其基因组来适应不同的信息处理任务的需求，这也是一种自适应能力。
5. 电突触（electrical synapses）：一些神经元之间的连接不是传统的基于电化学神经递质的（轴突—突触—树突）连接，而是不同神经元的树突之间的直接电突触连接，这种连接不能由本节提到的神经元模型描述。
6. 亚利桑那大学研究者Stuart Hameroff认为，人脑中意识（consciousness）的产生依赖于神经元中的“微管”（microtubules）内的量子振动，因此需要基于量子力学原理才能解释意识的产生。围绕这个主题，每年在亚利桑那有一个意识科学会议（The Science of Consciousness Conference）。但是这个假说并未受到主流神经科学家接受。

### **3.4.2 神经系统信息编码**

利用脉冲神经网络进行输入信息的分类、识别和学习，需要解决神经信息编码和学习算法这两个主要问题。不同的神经信息编码结果对学习性能将产生重要影响。神经信息编码包含两个过程：一个是特征提取，另一个是脉冲序列生成。感知神经系统首先对接收的感知信息进行关键信息表达或采样，对应于机器学习的特征提取过程，这方面神经计算的机理目前还没有成熟的理论和算法。在脉冲序列生成方面，研究人员常用的有两种方式：脉冲平均率编码（rate coding）和时间编码（temporal coding）。脉冲平均率编码以脉冲的频率来表达所有脉冲序列的信息，不能有效描述快速时变的感知信息。不同于平均率编码，时间编码考虑了精确定时的脉冲承载了有效的信息，因而可以更加准确地描述神经元的活动。神经科学实验都表明精确脉冲定时在视觉、听觉等感知信息处理中的重要作用[54,55]。

与采用平均率编码的传统ANN相比，SNN可以对快速时变信息进行特征提取和分析。Hopfield[56]首先提出了一种时间编码算法“时滞编码（Latency Coding）”，即神经脉冲的产生时间与刺激的强度呈负相关。由于其编码的简单和直观的特点，在SNN的研究中，时滞编码成为对视觉和听觉输入信息编码的主要方法。神经科学研究还发现了神经信息的群体编码（population coding），即通过神经元群而不是单个神经元的状态表达信息的方案[57,58]。由于神经系统的复杂性，单个神经元的活动容易受到影响，所以群体编码有利于提高信息编码的可靠性。

由神经科学实验提出的神经编码方法对信息的提取包含了大量冗余信息，因此，直接应用神经编码方法（如时滞编码）往往不能为SNN学习提供有效的信息表达。为了提高信息编码的有效性，研究人员尝试在编码过程中引入了不同的机制: [59]用Gabor 滤波器对输入图像预处理，提高了SNN的识别精度；[60]用自组织映射（SOM）先对原始音频数据提取特征，然后再进行脉冲编码，实现了对高噪声情况下的语音识别。

### **3.4.3 脉冲神经网络学习与训练**

由于神经元之间的可塑性使得生物神经系统具有强大的学习和适应环境的能力，因此在建模中考虑由外界环境变化和神经过程引起的神经突触的变化的调整能力是极为重要的。突触权重（Synaptic Weights）定义了两个神经元之间连接的强弱。赫布（Hebb）提出了第一个有关于神经突触权重修改的假设。基于此假设的学习算法可被总结为“同时激发的神经元连接在一起”。它被认为是学习跟记忆的根本机制并作为线性相关器被广泛应用于不同的神经网络模型中。虽然脉冲神经网络更为注重神经生理学的学习方法，但目前生物神经系统中完整的学习过程和机制仍然不清楚。

随着精确脉冲定时（Precise Spike Timing）和突触前激发和突触后激发之间的时间间隔被发现，毫秒级精度学习受到了热切关注。在过去的几十年里，科研人员从生物实验现象和结论中汲取灵感来探索神经突触可塑性（即学习）理论。通过引入突触前后脉冲之间的时间相关性（Temporal Correlation），毕国强和蒲慕明提出了脉冲时间依赖的可塑性（Spike-Timing Dependent Plasticity，STDP）机制并被推广到不同的脉冲学习机制[61,62]。STDP根据脉冲发放的先后顺序，调整神经元之间的连结强度：反复出现的突触前脉冲有助于紧随其后产生的突触后动作电位并将导致长期增强（Long-term Potentiation，LTP）；相反的时间关系将导致长期抑制（Long-term Depression，LTD）。

由于STDP原本是无监督学习机制，为了解决SNN的监督学习问题，人们将STDP机制与其他权值调整方法结合，这其中主要包含梯度下降和Widrow-Hoff 规则，以此推导出不同的监督学习算法。Gutig等人在《Nature Neuroscience》提出了基于梯度下降规则的Tempotron学习算法[63]，根据突触前后脉冲时间差和误差信号的共同作用来更新突触权值，从而决定突触后神经元产生或不产生脉冲。值得注意的是，Temptron算法只能训练突触后神经元（在特定时间范围内）产生或不产生脉冲，且不能决定脉冲发放的时间，因而不能作为计算的中间层传递信息。Ponulak等人[66]提出的ReSuMe学习方法从Widrow-Hoff 规则出发，从而避免了梯度下降算法中的梯度求解问题，突触的调整由两个过程决定（STDP和anti-STDP），可以实现对多脉冲输入的学习。[65]提出了SPAN算法，同ReSuMe类似，同样基于Widrow-Hoff规则，不同的是采用了脉冲卷积变换将脉冲转换成模拟值再进行运算，计算量较大且只能离线学习。Chronotron 算法[66]给出了基于梯度下降的E-Learning rule，通过最小化由实际输出脉冲序列和目标脉冲序列的差定义的误差函数来调整突触。两个脉冲序列的差可以由Van Rossum距离[67]或VP距离[68]来计算。在对神经元的单脉冲输出（如Tempotron）和多脉冲输出（如ReSuMe）的对比中，研究发现神经元的多脉冲输出可以大大提高分类精度和学习容量[69,70]。因此，采用多脉冲输入-输出映射的神经元作为计算单元是设计高效和大学习容量的SNN的基础。然而，目前脉冲神经元的主要学习算法包括ReSuMe、SPAN、Chronotron，但依然有很大的局限，虽然可以实现多脉冲（multi-spike）输入-输出映射，但只适用于单层的SNN。

相对于在单层SNN领域的大量的研究工作，多层SNN的学习算法研究却非常缺乏。主要原因是STDP机制的前向传播的单一性和神经脉冲的不连续性，给多层SNN的监督学习算法的构造带来很大的挑战。Bohte等人仿照ANN的误差反传（Error Back-Propagation）机制，首先提出了基于梯度下降的多层SNN学习算法SpikeProp，但是限制所有神经元只能发放一个脉冲，即只适用于单脉冲信息编码的问题。SpikeProp的改进和扩展算法如[71]、Multi-SpikeProp[72]放开了对神经元发放脉冲的限制，但是仍限制输出层神经元只能发放一个脉冲。虽然多层SNN及其学习算法已经显示了其优越的计算性能，如解决单层SNN无法解决的XOR问题和Iris数据集等分类问题，但由于其学习过程中隐藏层的不稳定性（不激发任何脉冲）（如[73]），造成较低的学习容量，限制了这些算法在复杂和大样本的学习问题中的应用[74]。国内也有不少研究者在SNN领域进行了相关研究工作，但主要集中于单层SNN和直接编码方法，对多层SNN的研究还处于起步阶段。

由于SNN的训练算法不太成熟，一些研究者提出了将传统的ANN转化为SNN的算法，利用较为成熟的ANN训练算法来训练一个基于ANN的深度神经网络，然后通过发放频率编码（firing rate encoding）将其转化为SNN[75]，从而避免了直接训练SNN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，HRL Labs研究者[76]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CNN；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DBN。目前还未见有人尝试将SNN应用于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SNN的软件仿真所需运算能力超出了一般台式机的运算速度。

还有一种SNN架构称为液体状态机（Liquid State Machine, LSM）[77]也可以避免直接训练SNN。LSM与基于ANN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的而固定的，神经网络形成一个“水池”，其作用将外界输入映射到一个高维状态空间以便于分类，因此SNN本身不需要训练，而只须训练一个输出层分类算法。只要SNN的规模足够大，理论上讲可以实现任意复杂输入的分类任务。由于LSM属于回归神经网络，即神经元之间的连接是有反馈的，这赋予其记忆能力，可以有效处理时序信息的分析。新西兰研究者Nikola Kasabov[78]基于LSM的基本思想提出了NeuCube系统架构，用于时序与空间信息处理，例如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SNN；在运行阶段，SNN的与输出层分类算发的参数也是在动态变化的，称为Evolving Connectionist Systems (ECOS)，这赋予了NeuCube系统很强的自适应能力。

### **3.4.4 神经形态认知计算架构**

认知计算以由神经元和神经突触组成的计算系统为重点，致力于推动对大脑中信息处理和神经突触通信相关的脑功能障碍的理解研究。通过将神经元和神经突触作为最基本的运算单元推动类脑计算科技的进步，将为神经处理器、信息处理技术和非冯·诺伊曼结构计算机的发展提供理论和实验基础。

学习和记忆是一切复杂认知过程能够实现的基础。组成大脑的学习与记忆的模块结构，包括组织信息预处理和识别的大脑皮层和记忆的海马体（Hippocampus）。海马体中的CA3和CA1被认为是存储和回忆检索自联想记忆(Auto-associative Memory)和异联想记忆(Hereto-associative Memory)的区域。 而反馈网络（吸引子网络）可以被用来模拟海马体中的CA1和CA3区域中的长期记忆。例如，图5所示的包含激励和抑制神经元的反馈网络可用来模拟实现CA3的部分功能，通过集成海马体中的椎体细胞(Pyramidal Neurons)、theta/gamma 震荡和锁相脉冲实现联想记忆的存储[79]。

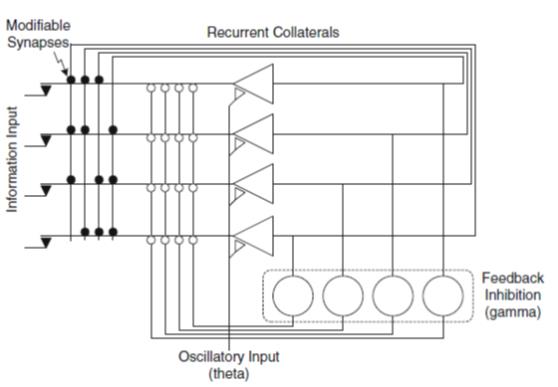


图5实现CA3的神经电路结构图[80]

除了用神经电路模拟实现大脑的某个区域的特定功能外，如图6所示的独立的可适应于大脑各区域计算原理的模块也可以用来探索和研究大脑的运算机制。该层级组织记忆模型体现了信息的分层传输，可以实现层内表达不同特征或信息的族群，以及激励神经元和抑制神经元的相互作用。

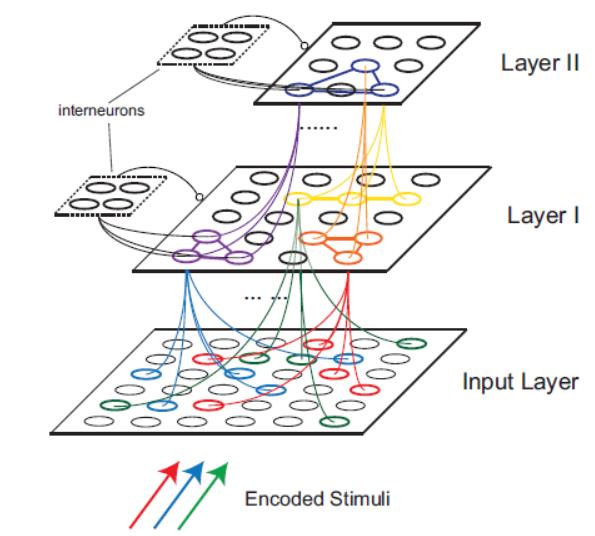


图6可适应于大脑各区域的学习和记忆模型

加拿大滑铁卢大学Eliasmith 团队的SPAUN 脑模拟器是多脑区协同计算领域标志性的工作。该团队早期曾提出神经工程框架理论(NEF)，通过定义功能函数并用神经网络逼近函数的思路来建立神经信息处理与认知功能实现之间的联系[81]。2012年，该团队基于早期的积累及新提出的语义指针架构统一网络，在其SPAUN 项目中将250 万个虚拟神经元组织为约10个模块化脑区，并在此基础上构建工作流式的脑区计算环路，发展出模拟笔迹、逻辑填空、简单视觉处理、归纳推理、强化学习等能力，实现了基于多脑区协同的多个特定功能神经网络[82]。然而该项目的问题在于为不同任务的实现人工构建了不同的工作流，脑区模型之间的协同并非自组织的，这与人脑的工作机制具有很大差异，即SPAUN 脑区计算环路并不具有真正的自适应性和通用性，而是根据不同任务人工组织的[83]。

由Hawkins 提出的分层时序记忆（Hierarchical Temporal Memory）模型更为深度借鉴了脑信息处理机制[84,85]，主要体现在借鉴了大脑皮层柱六层组织结构及不同层次之间的信息传递机制和皮质柱的信息处理原理等。该模型非常适用于处理带有时序信息的问题，并被广泛地应用于物体识别与跟踪、交通流量预测、人类异常行为检测等。

**3.5 仿生感知与类脑应用**

感知是人脑智能的重要组成部分，从进化角度看也是其他高级智能的源头，可作为实现仿脑智能的优先突破口。上个世纪七十年代，大卫·马尔（David Courtnay Marr，1945-1980）在对神经元群之间存储、处理、传递信息的神经计算建模研究基础上[153-156]，开辟了计算机视觉这个研究领域[157]。马尔的合作者托马索·波焦（Tomaso Armando Poggio，1947-）提出了模仿人类视觉信息处理通路的HMAX 模型[158,159]。波焦的学生克里斯托弗·科赫（Christof Koch，1956-）和弗朗西斯·克里克（Francis Harry Compton Crick，1916-2004）对意识神经基础的著名研究[160-162]也选择视觉作为突破口[163]。

受益于经典计算机算法成熟和性能的不断提升，基于经典计算平台的计算机视觉等机器感知研究汇聚了大量人力。HMAX模型尽管更“类脑”，也受到高度关注，但由于性能与深度学习等相比还有差距，因此还处于学界研究的边缘。然而，随着大规模视频监控、无人驾驶等应用需求的出现，对更灵敏、更智能和复杂度更低的机器感知算法是必然发展方向，以视频为例，每秒数十帧可以满足人眼观看的需要，但是远远不能满足自动驾驶的需要，脉冲阵列式感知信号表达和基于脉冲神经网络的检测跟踪识别算法必将成为未来的主流。本节将简要介绍基于类脑原则的传感器和相应的算法及应用，包括视觉、听觉、嗅觉与空间感知。

### **3.5.1 仿生感知**

传统摄像头是基于周期性的视频帧的，帧频越高，视频质量越好，时间敏感度越高，但视频码流所需带宽也就越大，事实上，每秒超过1000帧就难以在常规芯片和设备中应用了。受生物视网膜中神经元传输信息方式的启发，加州理工学院的Mahowald和Mead在上个世纪90年代初提出了一种称为AER（Address Event Representation，地址事件表达）的异步信号传输方式[86-88]：当像素阵列中某个像素发生“事件”时，将此像素的位置与“事件”输出。基于这一原理，瑞士苏黎世大学研发了动态视觉传感器（Dynamic Vision Sensor, DVS）[89]，是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），则输出一个脉冲；如果图片是静止不动的，没有像素的变化，那么摄像头就不会有任何输出。脉冲编码采用AER模式，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，可以实现高速移动物体的跟踪，而其所需码流带宽比传统的高速摄像头低很多。由于DVS的输出是一系列的脉冲，而不是传统的基于像素矩阵的图像帧，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。很直观的一个思路是采用SNN来实现后端处理算法，当然也可以采用其它算法。DVS研究团队已经与IBM TrueNorth团队展开合作，将TrueNorth芯片用于DVS的后端处理。DVS的低带宽优势使其在机器人视觉领域具有天然优势，已有工作将其用于自主行走车辆与自主飞行器中。新加坡南洋理工大学的陈守顺博士研制出了时间灵敏度达到25纳秒的异步传感芯片[90-92]。Garrick Orchard等介绍了一种用于对象识别的脉冲层级模型[93]，利用AER视觉传感器的输出中所蕴含的精准时域信息进行对象识别。类脑耳蜗[94]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。

欧盟项目NEUROCHEM[95]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，后端处理是基于x86处理器的SNN软件模型，用于仿真昆虫嗅觉中枢（触角叶，antennal lobe）或脊椎动物嗅觉中枢（嗅球，olfactory bulb）进行气味识别。台湾研究者研发了一个基于模拟电路的小型专用类脑芯片[96]，作为一个商用电子鼻产品的后端处理，通过仿真脊椎动物的嗅球原理来实现气味识别。海德堡大学研究者采用类脑芯片Spikey (HiCANN芯片的前身)来仿真昆虫嗅觉系统[97]。这些类脑嗅觉传感器的主要优点是低功耗，而其灵敏度性能与传统电子鼻相比并不一定具有明显优势。

脑机接口（Brain-Computer Interface，BCI）提供了生物（人或动物）脑（或者脑细胞的培养物）与外部设备（电脑）之间建立直接通信通路，在残疾人康复等领域有着重要应用。如果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，需要解决很多研究挑战，包括：芯片需要与生物体兼容，因此必须具备超级功耗与散热；需要长期植入而不换电池，因此最好采用能量收集技术从环境中获取能量；需要实时处理大量电生理数据，因此需要较高性能。这方面的相关工作不多，下面是几项较有代表性的工作：

1. 约翰霍普金斯大学研制了一个基于AMS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）来控制一只瘫痪猫的腿部神经系统，使其可以自主行走。其最终目标是用来帮助脊柱损伤病人实现自主行走。
2. 斯坦福大学研究者采用滑铁卢大学Chris Eliasmith的神经工程框架（Neural Engineering Framework，NEF)实现了一个基于2000个LIF神经元的SNN模型[98]，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BCI的神经假体（Neural Prosthesis），即用脑电信号来控制外部机械手。未来工作是基于斯坦福大学的类脑芯片Neurogrid来实现这个SNN模型。
3. 南加州大学研究者研制了一个基于模拟电路的超低功耗神经信号处理芯片[99]，用于大鼠海马体的认知假体（cognitive prosthesis）,可以实现一定程度的记忆功能修复。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，而并非基于类脑计算技术。但是我们可以想象可以采用一个基于SNN的类脑芯片来仿真部分脑区的功能，与其它脑区之间紧密交互，来修补一些缺失或者出故障的脑功能。尽管这听起来有些科幻色彩，但是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，那么这未必就不可行。

### **3.5.2 基于脉冲时间的视觉识别**

尽管脉冲神经网络已经被成功应用于模式识别问题，神经系统是如何从感觉编码到突触学习利用脉冲时间进行运算的问题仍值得研究。[100]提出了一个基于脉冲时间的模式识别模型，如图7所示。

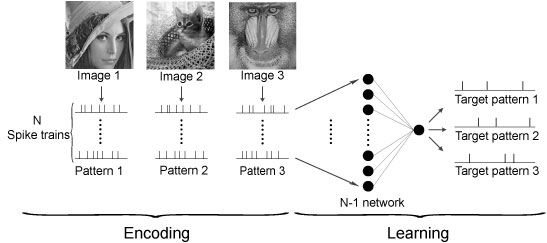


图7基于脉冲时间的模式识别模型结构

上述模型主要由编码和学习网络两个部分组成。由时滞编码（Latency Code）和相位编码（Phase Code）结合编码方法被用来将图像信息转化成为由神经脉冲序列组成的时空斑图（Spatiotemporal Pattern）。随后，一个单层脉冲神经网络被用来学习识别不同的输入。

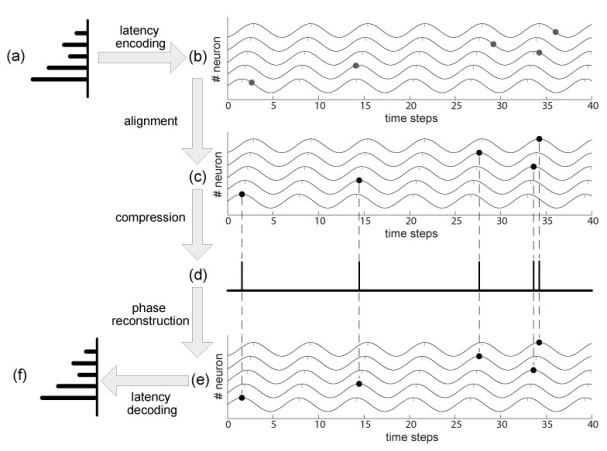


图8 时滞-相位编码的流程图[100]

受生物视觉系统的启发，时滞-相位编码被提出用以将视觉信号转化为时空脉冲信号，如图8所示。灰度图的像素值被理解为感光细胞（Photoreceptor Cell）接受的刺激强度。由于神经节细胞（Ganglion Cells，GCs）产生神经脉冲的延迟时间与其接收的感光细胞输出值的大小负相关，灰度信息因而可以被准确时间脉冲编码。此外，通过引入相位信息，同一个感受区（Receptive Field, RF）里的神经脉冲可以被压缩为一个神经脉冲序列（图4 c-d）。一副灰度图从而可以被编码成由若干个神经脉冲序列构成的时空斑图。经过学习，输出神经元可以在期望时间产生毫秒精度级的脉冲序列。当含有噪声的图像呈现给该网络时，根据其与目标脉冲序列的相似程度可以进行成功的识别。类似的基于脉冲时间的模式识别模型[101]被成功的应用于基于AER(Address Event Representation)传感器的分类系统，实现了对手写数字0-9的识别[102]。

### **3.5.3 基于脉冲神经网络的语音识别**

尽管许多以前的研究工作被认为是受生物逻辑启发的听觉前端，比如基于Gammatone滤波器组的语音识别，传统的模式识别方法Gammatone Cepstral Coefficients (GTCC)，但是并没有一个完整的生物系统。实验表明神经元之间的交互是通过短时脉冲的方式，这代表外部刺激是以时空脉冲序列的模式存在于大脑中。论文[103]从生物的角度出发，提出了一个基于局部频谱特征的声音识别系统（LSF-SNN系统），如图9所示。

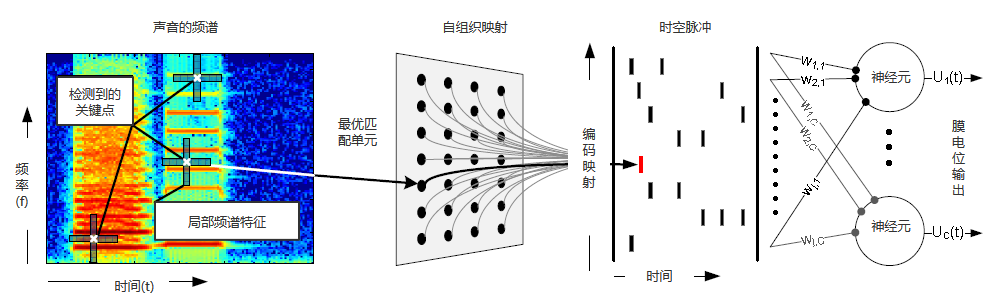


图9 基于局部频谱特征的声音识别系统

上述模型主要由特征提取、时间编码和学习算法三个部分组成。首先，使用一个简单的搜索局部时间-频率最大的方法，通过稀疏准则来剔除无意义的峰值，将得到的一系列关键点信息进行时间编码。在论文中，采用自组织映射神经网络（SOM）进行编码，通过寻找最好匹配单元对输入的关键点信息进行映射，从而形成一个时空脉冲模式来表达基本的声音信息。然后，利用Tempotron学习算法学习识别不同的输入。

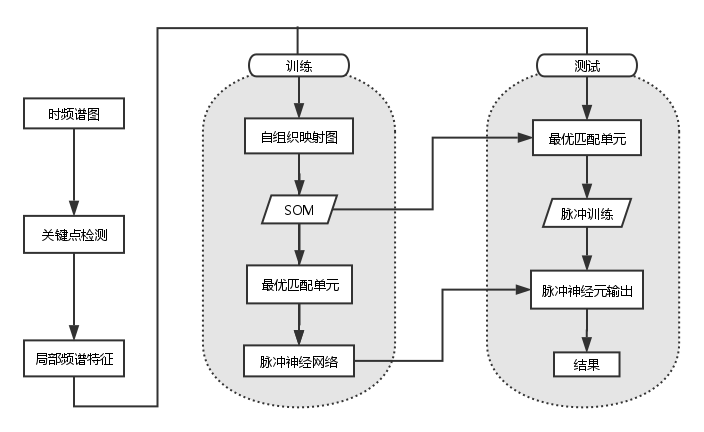


图10 LSF-SNN算法流程图

如图10所示，训练过程和测试过程被联系在了一起。声音信息被转换成时间序列才能被用于脉冲神经网络进行识别。结果说明基于脉冲神经网络的语音识别方法具有可行性。与传统的语音识别方法相比，该方法不但能正确识别语音数据，而且对噪声具有鲁棒性。

### **3.5.4 机器人仿脑空间感知和记忆**

受认知神经科学启发，通过集成认知地图(Cognitive Map)和情景记忆(Episodic Memory)研究人员提出了一个面向任务的机器空间认知架构[104]，如图11所示。其中，大脑皮层的网格神经元（Grid Cell）和位于海马体的位置神经元（Place Cell）通过对视觉输入信号的共同作用，构成基于大脑的空间定位和导航系统[105]。同时，结合海马体的记忆网络，可以实现对指定位置或动作序列的学习和记忆[106]。

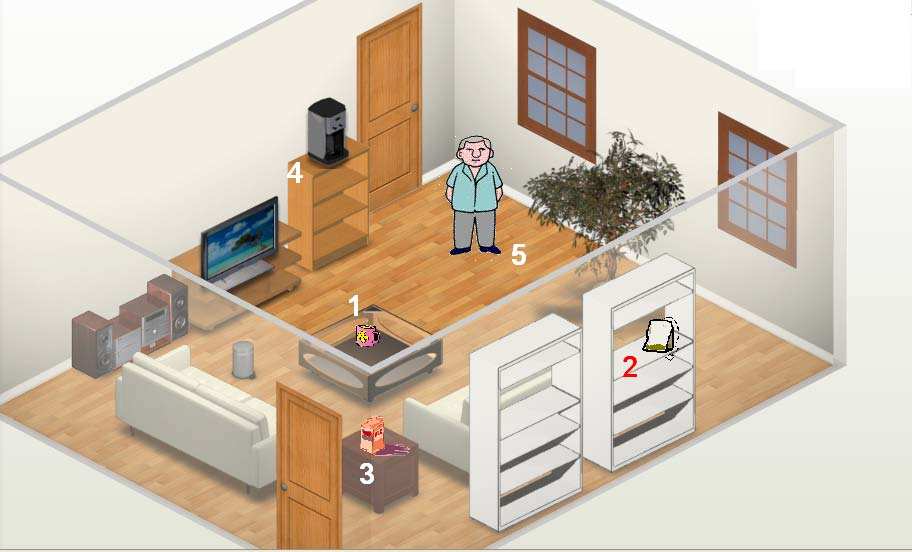


图11 空间感知和位置（动作）序列记忆示例[106]

机器人在图11中所示的环境中进行学习如何制作奶茶，并生成认知地图。

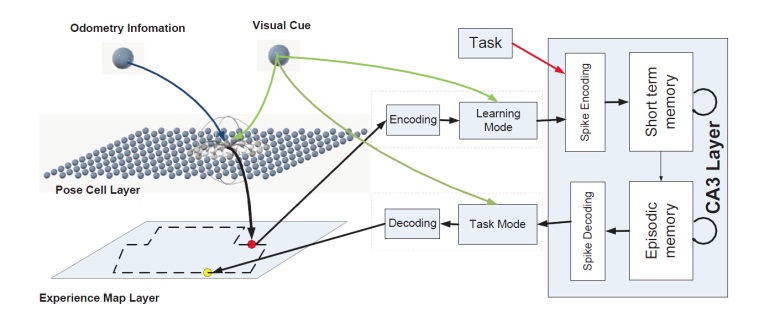


图12 基于大脑的空间感知和记忆的认知系统结构[106]

与此同时，任务信息和视觉提示被图12中的脉冲神经网络联系在一起并以情景记忆的形式存储。而在任务模式中，任务信息可以恢复机器人的情景记忆，并在视觉提示的帮助下顺序完成子任务。实验结果验证了该机器人导航系统通过结合认知地图功能和情景记忆可以成功解决现实生活中的实际问题。

**4 国内研究现状**

**4.1 我国“脑计划”与相关研究**

### **4.1.1 我国脑科学研究与“中国脑计划”概况**

我国在《国家中长期科学和技术发展规划纲要 (2006~2020)》中即将“脑科学与认知”列入基础研究科学前沿问题，2011 年发布的《国家基础研究发展“十二五”专项规划》也将“脑科学与认知科学”列入重点方向。国家重点基础研究发展计划(973 计划)先后启动了40多项脑科学与脑疾病相关的课题，投入十多亿元，国家自然科学基金委支持了 “视听觉信息的认知计算”等近200项重大、重点课题，总投入数亿元。2012 年，中国科学院启动 “脑功能联结图谱”科技先导专项，计划在10年内投入6 亿元，描述几种特殊脑功能(如感觉、情绪、记忆、学习等) 的神经网络联接方式和运作机制[108]。

“中国脑计划”的讨论开始于2013年，2015年3月多部委联席会讨论确定的计划名称为 “脑科学与类脑计算”（Brain Science and Brain-Inspired Intelligence Technology），基本架构为“一体两翼”，“一体”为“脑认知原理”，“两翼”为“脑重大疾病与健康”和“类脑计算与脑机智能”，分别服务于“人口健康与和谐社会”和“产业发展与国家安全”两大目标。2016年3月获得国务院层次的立项批准，规划15年（2016-2030），具体研究内容和规划尚在讨论中。

### **4.1.2 部分省市 “脑计划”**

北京在脑科学研究方面拥有全国最完整的学科布局、最强的研究队伍、国际一流的实验条件和技术资源。北京在类脑计算方面至少有十年的研究历史，例如北京大学微电子研究院从2006年开始研制神经形态器件研制，2012年实现了神经突触模拟器件，响应速度比生物突触快百万倍，单元体积只有生物突触的十万分之一。

2015年9月1日，北京市科学技术委员会正式发布 “北京脑科学研究”专项规划。主要依托首都脑科学研究优势，瞄准国际脑科学研究前沿，从“脑认知与脑医学”和“脑认知与类脑计算”两个方面进行布局[109]。 “脑认知与类脑计算”作为专项两大重点任务之一，沿着“结构仿真、器件逼近和功能超越”这条技术路线，布局了三个层次、九个方面的科研任务：建设支撑类脑计算基础研究开发和产业创新的四大基础性公共平台（大脑解析仿真平台、认知功能模拟平台、神经形态器件平台和类脑计算机系统平台），开发两款类脑计算处理器芯片（类脑处理器和机器学习处理器）并投入实际应用，研制类脑计算机软硬件系统，在视听感知、自主学习、自然会话三大类脑智能方向取得突破并实现规模应用。

2015年3月31日，上海市发布“科技创新行动计划”基础研究领域项目指南，“脑科学与类脑人工智能”列入基础研究项目专题，围绕脑学习记忆的神经基础，发展新型类脑智能算法，开发具有自适应、自组织、自学习功能的智能器件和设备，切实推动和加速类脑人工智能的研发与应用。2015年10月，复旦大学类脑人工智能科学与技术研究院首席科学家冯建峰教授介绍，经过一年左右的讨论，上海脑计划将由复旦大学、上海交通大学、华东师范大学、上海纽约大学四所大学牵头，联合一批高校和科研院所以及企业从四个方面开展：一是关于神经元的研究；二是发展智能算法，通过计算神经生物学建立大脑研究和人工智能之间的桥梁，帮助科学家更好地理解大脑；第三是根据人脑研究，发展人工智能以及可穿戴设备、芯片等；第四是应用，比如通过大数据对疾病进行研究等。

### **4.1.3 我国类脑计算相关研究机构**

我国类脑计算的研究起步较晚，但其科研价值和社会价值已受到政府和学术界广泛重视，是关系到我国未来信息科学发展和“脑计划”项目实施的重大科技问题。近年来，我国高校和研究机构在类脑计算研究方面十分活跃，成立了多个类脑计算或类脑智能方面的研究中心。

北京大学成立了由生命科学学院、信息科学学院、工学院、心理学系、基础医学部等开展学科交叉研究的“脑科学与类脑研究中心”。清华大学成立了跨七个院系的“类脑计算研究中心”，从基础理论、芯片、软件系统展开全方位研究。复旦大学联合浙江大学、华中科技大学、同济大学、上海交通大学等十几所高校及中科院研究所，成立了“脑科学协同创新中心”，推进脑科学研究和转化应用。中国科学院自动化研究所成立了 “类脑智能研究中心”，中国科学院计算技术研究所研制出世界上首款深度学习处理器芯片。中国科学院成立了中国科学院脑科学与智能技术卓越创新中心，由神经科学研究所、自动化研究所等机构共建，将深度实质性融合脑与神经科学、认知科学、人工智能、计算机科学等不同领域的研究，实现脑科学的深度探索与类脑智能的创新研究。中国传媒大学2013年5月成立“脑科学与智能媒体研究院”开展类脑仿真平台的创建研究。四川大学计算机学院（软件学院）成立了西南地区第一个类脑计算研究中心。电子科技大学神经信息教育部重点实验室联合相关单位两次举办“天府脑科学论坛”。2015年10月，重庆市成立“脑科学协同创新中心”，由第三军医大学牵头，联合重庆大学、西南大学、重庆医科大学、重庆邮电大学和中科院重庆绿色智能研究院等建设。

**4.2 神经形态器件**

近几年来，国内在神经形态器件的研究方面也取得了长足的进步。 特别在新型固态存储器的研究领域，取得了可喜的科研成果。

北京大学利用过渡金属氧化物忆阻器展示了STDP学习法则，在硬件上首次实现了异源性突触可塑性[110]；针对现有忆阻器中普遍存在的低线性度问题，将具有低离子扩散系数的介质材料作为界面插层引入器件当中，从而有效限制导电细丝生长/溶解动力学过程的速率，显著改善了忆阻器的线性度[111]；采用石墨烯取代传统金属作为忆阻器的电极材料，使器件获得抗串扰的非线性低阻态特性，从而成功将选通器件的开关特性整合到忆阻器器件中，提高了整体的集成度[[112](#_ENREF_30)]。另外，提出关于金属氧化物阻变效应的统一机制，阐明了金属氧化物忆阻器中阻变效应的微观物理机理；建立了适用于大规模电路仿真、能够精确描述表征器件特性与物理效应的集约模型[[113](#_ENREF_31)]。

清华大学系统研制和开发了多款用于模拟突触的忆阻器器件。文献[114,115]发现TaOx材料具有极佳的可靠性，可以反复循环擦写1011以上；通过界面工程的方法，将TaOx的阻变机制从氧空位通道阻变改变为界面势垒变化控制的界面阻变，从而获得了模拟型的连续电阻变化[116]。利用FeOx纳米忆阻器件，展示了集STDP和SRDP于一体的复杂学习法则；并构建了一款全新的仿生反馈神经网络，通过引入递归最小二乘算法（RLS），成功实现了复杂的时空模式学习[[117](#_ENREF_35)]。此外，发现了磁场可以引起硅器件电阻的巨大变化，并成功制备了硅磁电阻器件与逻辑器件，对于将自旋电子学引入神经形态器件领域具有重要作用[[118](#_ENREF_36)]。同时，利用石墨烯独特的双极型输运特性，首次实现了基于二维材料的类突触器件[[119](#_ENREF_37)]。

南京大学研制了基于氧化物双电层晶体管的人造突触/神经元，和具有高质子导电特性的自支撑壳聚糖膜，并在这两种器件上，实现了LTP，STP和STDP等学习法则[120[-123](#_ENREF_38)]；基于银锗硒（AGS）固体电解质忆阻器，提出忆阻“算盘”的新概念，即根据刺激依赖的突触权重变化量，执行四则运算[[124](#_ENREF_42)]；研制了一种基于BaTiO3铁电隧道结（FTJ）的认知器件，利用铁电隧道结的自发极化翻转使得势垒高度变化，可在两个阻态之间以低功耗进行快速度的翻转。同时，在编码过程中引入“训练”环节，实现记忆固化，可有效解决在检索时出现误码的问题[[125](#_ENREF_43)]。

中科院上海微系统与信息技术研究所在自主新型相变材料开发和相变存储器工程化方面取得了很好的进展。提出通过掺杂提升相变存储器在非线性区可控性，使其行为更为接近突触的特点；并提出通过字线位线信号转换建立递归神经网络，实现逻辑推理和预测的功能；开发出Ti-Sb-Te等自主新型相变材料[126,127], 发现Ti掺杂的Sb2Te3表现出与传统GST完全不同的相变机制，可将相变速度可提升一个数量级，大幅降低器件的Reset功耗[128]。此外，在12英寸40nm的工艺线上，开发出非标准流程的1D1R工艺，存储单元成品率>99.9%，为嵌入式应用神经网络芯片开发奠定了基础。

华中科技大学重点开发了针对相变存储器的神经形态器件，包括突触的模拟和神经元的模拟，实现了四种不同形式的STDP学习法则，和神经元在连续的脉冲刺激下超过阈值而产生积分触发的特性[[129](#_ENREF_47)]；利用基于AgInSbTe的忆阻器突触器件实现了依赖活动频率和幅值的Hebbian学习法则[[130](#_ENREF_48)]；还基于忆阻突触器件，提出了单器件、互补式结构、阵列结构、CMOS混合、电磁耦合等多种非易失性布尔逻辑的运算方法[[131](#_ENREF_49)]。另外，研究了神经网络的多稳定性[[132](#_ENREF_50)]、联想记忆性[[133](#_ENREF_51)]、电路实现的镇定性[134]和无源性[135]等问题。

国防科技大学基于Au/Ti/TiO2/Au忆阻器器件，研究了Forming过程中的限流对的阻变特性的影响，建立了一个基于体材效应和界面效应的忆阻器物理模型[[136](#_ENREF_54)]，并提出了三种电路级模拟程序来模拟忆阻器的特性[137,138]。另外，基于忆阻器的存储和计算融合的能力，提出了一种支持二值图像的存储与简单图像处理操作的体系结构，可以显著减少访存操作，提高系统的能效[139,140]。

另外，在固态器件研究的基础上，清华大学还在基于神经形态器件的类脑芯片的架构设计方面，进行了系统的研究，包括器件结构、神经网络算法、芯片的架构和系统功耗等多个层面。制备了最多集成16M个HfOx忆阻器突触的大规模的神经网络芯片，实现了大于99.9%的良率[141]；设计了一款集存储和在线学习于一体的单层神经网络系统，演示了图像识别的功能，可用于识别人脸，准确率高达99%[[142](#_ENREF_60)]；分析了阻值异变对识别率的影响[[143](#_ENREF_61)]。另外，在基元忆阻器件级和网络学习级,提出了基于忆阻器突触的能耗估计方法和降低能耗的有效措施，并在三种具有不同突触结构和学习策略的网络中进行测试；同时提出了基于忆阻器器件的基本构建模块MCore的概念，展示了系统级的设计框架[[144](#_ENREF_62)]。成功的开发了基于忆阻器的类脑智能芯片，该芯片的功耗是英特尔的超级电脑芯片Xeon Phi功耗的千分之一。

**4.3 神经网络芯片**

### **4.3.1 人工神经网络芯片**

人工神经网络芯片的基本结构由硬件神经元和神经元间数值连接组成。早期的人工神经网络芯片其硬件神经元的数目、神经元间的拓扑连接和连接权值均预先设定，可以完成传统浅层人工神经网络计算的功能。国内代表性的工作为中科院半导体所开发的“预言神”[[14](#_ENREF_62)5]。近年来，随着人工神经网络研究的不断深入，中科院计算所开发了“寒武纪”芯片[[14](#_ENREF_62)6]，能够由指令集来配置与调度硬件神经元数目、拓扑连接结构和连接权值，能够高效地完成任意规模的深度神经网络计算。

**（1）“寒武纪”**

中科院计算所和法国INRIA提出了国际上首个深度神经网络处理器架构，相关研究论文两次获得计算机体系结构顶级会议ASPLOS和MICRO最佳论文[[14](#_ENREF_62)6] [[14](#_ENREF_62)7]。在此基础上，中科院计算所研制了“寒武纪”深度神经网络处理器芯片，包含独立的神经元存储单元和权值存储单元，以及多个神经元计算单元。“寒武纪”芯片每秒能处理160亿个神经元和2.56万亿个突触运算，可达到每秒512G浮点运算速度，比Intel通用处理器性能和能效提高100倍[[14](#_ENREF_62)7]，可广泛适应各种智能处理应用。IBM、Intel、HP、微软、MIT、哈佛大学、斯坦福大学、UCLA、哥伦比亚大学和佐治亚理工等国外知名机构纷纷跟踪引用“寒武纪”，开展深度神经网络硬件探索。目前“寒武纪”研究团队已完成专利布局，成立了创业企业，开展产业化工作，正在与曙光合作研制智能云服务器，与多个核心手机芯片厂商合作研制面向下一代智能手机芯片的智能处理器。未来，“寒武纪”需要在超大规模人工神经网络芯片架构、智能处理指令集、智能编程语言和编译器、以及自主智能算法等方面取得进一步突破，力争形成具有小样本学习能力的智能软硬件。

**（2）“预言神”**

中科院半导体所于1995年研制了“预言神1号”（CASSANDRA-I）[[14](#_ENREF_62)5]，其中集成了反向训练、权值存储和计算单元，达到了每秒20M的神经网络正向运算速度，可用于完成典型人工神经网络的运算。2000年半导体所进一步提出了“预言神2号”（CASSANN-II）。与前者相比，其每个神经元的连接可以有两个权值，提高了对于特定领域（如语音识别）的处理器精度。此后该课题组进一步研制了“预言神3号”（CASSANN-III）和二值递归神经网络芯片（神经元只能接受两个不同的值）等。其中二值递归神经网络芯片可以达到每秒128M浮点运算速度。“预言神”系列主要是采用FPGA，面向的主要是传统的浅层人工神经网络以及二值递归神经网络，难以高效地处理深度神经网络（如卷积神经网络和长-短时记忆网络等）。

### **4.3.2 脉冲神经网络芯片**

脉冲神经网络借鉴了大脑的工作机理，神经元之间直接以脉冲的形式传递信息：神经元接收输入脉冲，导致膜电位升高，在达到一定阈值后将被激活产生输出脉冲。目前其硬件神经元数目、拓扑连接结构固定不变，且不具有指令集。国内的代表性工作主要有清华大学的“天机”和浙江大学的“达尔文”[[148](#_ENREF_62)]。

**（1）“天机”**

清华大学的施路平教授团队于2015年针对未来通用人工智能的发展要求提出类脑混合计算范式架构，并发展了基于CMOS的类脑芯片“天机”，既支持脉冲神经网络，也支持ANN（analog neural network），及各种混合集成网络，受邀在电子器件国际最顶级会议IEDM2015上做了特邀报告。单个芯片包含6个核，核间通过2×3的片上网络进行互联，可达到每秒153.6G定点运算速度。每个核包含神经元块、权值网络、路由器、同步器和参数管理等模块，支持256个简化脉冲神经元的计算，运行频率为100MHz。此外，清华大学已经设计完成两款基于忆阻器的大型神经网络芯片。其中一款是限制玻尔兹曼机，可以对图像等进行特征提取。另外一款是基于忆阻器的完整双层神经网络，能够实现手写体数字的识别。

**（2）“达尔文”**

浙江大学计算机学院、微电子学院及杭州电子科技大学联合研究团队主要面向低功耗嵌入式应用领域，于2015年研发了一款基于CMOS数字逻辑的脉冲神经网络芯片“达尔文”，支持基于LIF神经元模型的脉冲神经网络建模[148]，如图13所示。该芯片当前仅有单核版本，包括8个物理神经元，通过时分复用技术，最多可支持2048个神经元、400万个神经突触和15个不同突触延迟。由于采用的流片工艺较老（180nm CMOS工艺），目前仅支持最高运行频率70MHz，从而单核仅支持每秒0.56G乘积累加（MAC）定点运算操作。相比较地，IBM的TrueNorth芯片采用最先进的28nm工艺，但每个核仅有一个物理神经元，通过时分复用技术支持最多256个神经元。2016 IEEE CIS 计算智能相关的Summer School 还将达尔文芯片作为一个案例供所有参加人员编程实践与应用开发。

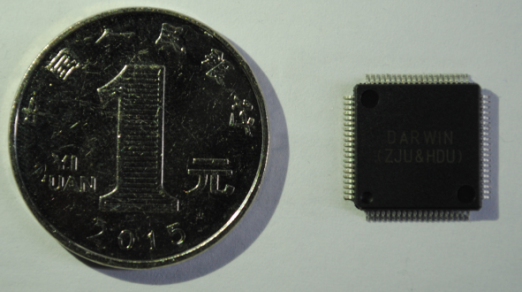
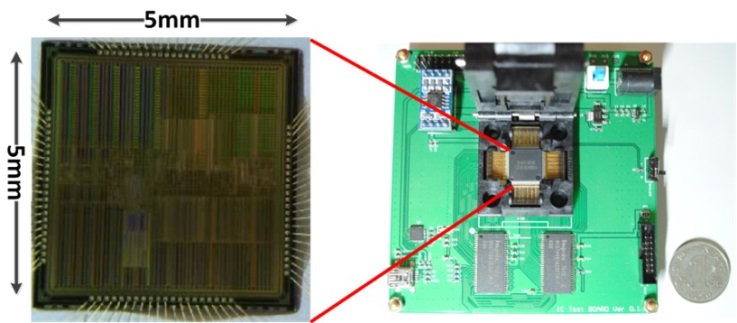
 

图13一款基于CMOS数字逻辑的脉冲神经网络芯片“达尔文”

### **4.3.3 视觉处理芯片**

视觉处理芯片主要针对视频和图像中的图片降噪、边缘检测、图像增强和目标追踪等问题进行优化，其基本结构仍然是传统的数字信号处理器，可包含多个多数据流运算单元，不具备硬件神经元概念。目前国内的代表性工作主要有中科院半导体所的视觉处理芯片[[14](#_ENREF_62)9]和中星微电子的“星光智能一号”嵌入式处理器。

**（1）中科院半导体所视觉芯片**

中科院半导体所的吴南健研究员及其团队长期从事视觉和图像处理片上系统的研究，2008年提出基于传统处理单元阵列（Processing Element Array）并行处理器架构的视觉芯片，2014年提出集成了处理器阵列和自组织映射神经网络的混合视觉芯片架构，专门针对各种图像处理任务，包含64×64的处理单元阵列和16×16的自组织映射神经网络，可达到每秒12G的定点运算速度。视觉芯片主要面向图像处理，难以应用到语音识别、文本处理和自然语言理解等其它智能处理领域。

**（2）“星光智能一号”**

中星微电子邓中翰院士团队于近日宣布“星光智能一号”嵌入式处理器已实现量产。该处理器在架构方面仍为传统的数字信号处理器架构，采用多核定点处理单元，共包含四个核，每个核包含4个内核，每个内核有2个流处理器，可达到每秒152G定点运算速度，实现了对卷积神经网络正向运算的支持，国际上已有类似的商用产品出现（如CEVA公司的XM4处理器、Cadence公司的Tensilica Vision P5处理器、Synopsys公司的EV处理器等）。从掌握的情况看，由于其系统架构和定点处理的限制，不支持神经网络反向训练，且能够支持的神经网络模型较少，难以实现在语音识别、文本处理和自然语言理解等更为广泛领域的应用，并不是完全意义上的神经网络处理器，主要面向嵌入式领域，而不适用于高性能服务器领域。

表2给出了前面提到的国内部分单位的智能芯片在性能、应用领域、所支持的网络类型和编程支持等方面的对比情况。

表2 国内神经网络芯片对比

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 类型 | 名称 | 性能 | 模式 | 应用领域 | 网络类型 | 编程支持及  指令集 |
| 人工神经网络芯片 | 寒武纪 | 512G | Flops | 图像、视频、语音、文本、自然语言理解、决策 | CNN/DNN/  RNN/LSTM/  RCNN/Fast RCNN等 | Caffe/Tensorflow  等主流编程框架  高性能库  调试和性能分析  神经网络指令集 |
| 预言神 | 0.128G | Flops | 语音 | 双权值浅层人工神经网络 | - |
| 脉冲神经网络芯片 | 达尔文 | 0.56G | Ops | 手写数字识别脑电波编码 | 基于SNN，不支持深度学习 | 自定义编程模型 |
| 天机 | 153.6G | Ops | 图像 | 混合计算范式，既支持SNN也支持模拟神经网络（ANN）及各种混合集成网络。 | 自定义编程模型 |
| 视觉处理芯片 | 半导体所视觉芯片 | 12G | Ops | 图像、视频 | SOM | - |
| 星光一号 | 152G | Ops | 图像、视频 | CNN | Caffe/Tensorflow等主流编程框架 |

备注：CNN-卷积神经网络；DNN-深度神经网络；RNN-递归神经网络；LSTM-长短时记忆神经网络；RCNN-基于区域神经网络；SNN-脉冲神经网络；SOM-自组织映射神经网络。

**4.4 类脑计算模型与应用**

从2015年7月起，北京市科学技术委员会一年内在“脑认知与类脑计算”方向已经分三批支持了十八个研究项目，涵盖了大脑解析仿真、类脑处理器、认知计算、神经形态器件等各个方面。以北京大学牵头的《大脑初级视觉系统解析仿真平台研究和应用验证》项目为例，在北京大学内部实现了信息科学和生命科学的深度交叉：视觉损伤与修复教育部重点实验室负责测量视网膜各类神经节细胞在多种视觉刺激的发放模型，数字视频编解码技术国家工程实验室负责开发高逼真视网膜仿真软件，生物医学中心采用双光子成像研究灵长类初级视皮层神经机制，定量生物中心则从神经网络仿真和动力学分析方面进行软件模拟，通过生物实验和计算机仿真进行交叉验证，实现对生物视觉系统视觉信息处理和加工过程的精确重构。预计这项研究不仅能深化对生物视觉系统信息处理过程的认识，还有望启发新的高效视频编码方法和对象检测识别算法，并通过新一代视觉芯片支持智能机器人和无人机等产业发展。

中国科学院自动化研究所将类脑智能作为所级战略方向，在类脑认知计算模型及其应用方面全面布局。在北京市科学技术委员会类脑计算专项和中科院先导专项的支持下，开展脑科学与类脑智能研究的深度融合。类脑认知计算模型方面，该所开展了不同尺度的类脑认知计算建模研究。在微观尺度，构建了超过30类不同类型的神经元与突触计算模型；在介观尺度，实现了由不同类型神经元构成的、完成不同基本单元级别认知任务的微环路和皮质柱计算模型。在宏观尺度，已构建了由213个脑区、约7100万神经元组成的哺乳动物多脑区协同计算模拟系统。在此基础上实现的类脑认知计算模型已经具备初步的多模态感知、自主学习、记忆、归纳推理、决策、动作模仿等认知能力，并应用于无人机和机器人平台。

北京师范大学认知神经科学与学习国家重点实验室在认知和计算神经科学领域有长期的研究基础。吴思教授领导的神经信息处理课题组致力于用数学建模和仿真方法来揭示生物神经系统处理信息的基本原理，为类脑计算奠定理论基础。在北京市脑计划中，吴思课题组主要关注高级视觉系统的大尺度建模，视觉系统动态处理信息的机制，以及由此发展的物体识别算法。

中科院半导体所高速电路与神经网络实验室从1990年起，一直在开展类脑计算机的研制。相关工作先后获得了95年中国电子十大科技成果奖、“八五”科技攻关重大科技成果、2001年北京市科学技术进步一等奖等。在“十一五”期间，半导体所团队研制成功了权值精度和神经元精度达到32位浮点精度、网络规模达到1024个神经元的半导体类脑神经网络硬件。以此为核心，研制成功了神经计算机系统，提供了可进行多种神经网络计算和应用研究的原型平台。“十二五”期间，半导体所在国家重点项目、中科院装备研制项目支持下，研制了神经计算机专用系统，能够解决国家重要工程中的超大规模计算问题，目前正在研制器件规模超过1亿只晶体管的专用类脑芯片，以支持更高性能的神经计算机专用系统。

中国传媒大学2013年5月成立“脑科学与智能媒体研究院”开展类脑仿真平台的创建研究，实现了对HH模型的快速计算方法，形成了在精度上接近HH模型，速度上接近LIF模型的新型类脑仿真模型。并在北京市科学技术委员会支持下，实现了基于GPU的高性能计算设备上达到百万神经元的近实时的类脑仿真效果。目前正在打造大规模近实时的类脑仿真平台，预计在2016年内向全球开放。

西安交通大学为了探索人类大脑的视觉分类和识别功能，提高视觉认知任务处理的效率，正在设计一款视觉认知脑计算网络并行仿真器。

浙江大学于2015年成立了脑科学与智能研究联盟，涵盖了神经生物学、系统神经科学、临床神经精神疾病、人工智能、计算机科学与技术、生物医学工程、光学等。目前引进的国际领先的7T功能核磁共振设备已开始运行，将有助于发现类脑运行机制与规律，为类脑计算模型研究提供基础与启发。脑机一体化智能系统是类脑计算的一个重要应用领域，浙江大学在此方面的工作独具特色，如猴子意念控制机械臂、听视觉增强的机械化大鼠等。

四川大学于2015年成立了“类脑计算研究中心”，目前建立了各类脑计算研究的软硬件平台，包括机器人、类脑处理器、大规模脑系统仿真工作站。该研究中心主要工作集中在神经形态计算和神经-认知机器人，提出了基于精确脉冲驱动的脉冲神经网络算法、集成时间编码与学习模型和仿脑GPS等代表性工作，获得MIT Technology Review专题报道。神经形态计算将为类脑芯片与类脑智能系统提供重要的底层神经电路和模型支撑，其在脉冲算法和复杂认知功能等关键问题上的突破将会有力促进类脑计算作为新一代信息技术与智能机器人技术的发展。

**5 国内外研究进展比较**

**5.1 神经形态器件**

近几年来，我国在神经形态器件研究方面，取得了可喜的成果。特别是基于固态存储器的突触模拟及开发，与国外差距在逐渐缩小。在各种类型的固态存储器上或电子层晶体管器件上，都成功的实现了对各种生物学习法则的模拟，包括STDP、SRDP、LTP、LTD，STP等等。但在相关专利布局方面，由于我国起步较迟，我们有较大的落后。另外，在固态存储器工程化的方面，相变存储器已实现大规模集成制造，为嵌入式应用芯片开发奠定了坚实基础；而忆阻器和磁存储器的大规模制造技术，与国外相比，我国还比较欠缺，应加大在这方面的投入。

另外，从系统角度来看，基于神经形态器件的新型类脑计算系统，特别是类脑计算架构设计方面，与国外相比，我们还是有一定的差距。

**5.2 神经网络芯片**

国际上在神经形态芯片方面已经开展很多研究，在大脑模拟仿真方面迈出了重要一步，但其智能处理能力仍然难以应对当前的实际应用场景（如大规模的图像识别场景）。此外，由于颠覆了传统的编程方式，“真北”等SNN芯片的编程还比较困难。

神经网络芯片主要分为脉冲神经网络和人工神经网络两种路线。国际上脉冲神经网络芯片主要以IBM的“真北”（TrueNorth）为代表，相比于国内脉冲神经网络的研究起步早，从2008年就开始投入数亿美元进行研究，是脉冲神经网络领域的领先者。而人工神经网络芯片则是以国内中科院计算所的“寒武纪”（DaDianNao）为代表，在该领域最先进行了寒武纪人工神经网络处理器体系结构的开创研究，并获得国际顶级会议最佳论文，引领了IBM、Intel、HP、微软、MIT、哈佛大学、斯坦福大学、UCLA、哥伦比亚大学和佐治亚理工等国外知名机构进行了跟随研究。

“真北”本质上是脉冲神经网络芯片，其基本结构由硬件神经元和神经元之间的脉冲连接组成，硬件神经元接收输入脉冲，在累积到一定阈值后被激活产生输出脉冲。最早的模拟芯片如Intel的ETANN，其中包含64个全连接的神经元和10240个权值连接。TrueNorth的每个核采用了Crossbar结构的SRAM，拥有256个输入，256个输出神经元。每个芯片拥有4096个芯片核，总计一百万神经元，256百万个权值连接，共计采用54亿个晶体管。单个芯片平均放电频率为20Hz，单个神经元放电功耗低至26pJ，功耗低至65mW，远远低于传统芯片。IBM表示今后将多个TrueNorth芯片连接起来还可以提供更加强大的脉冲神经网络处理能力。但是它在机器学习任务上的低精度使其一直在事实上缺乏说服力。Neoflow的作者Culurciello表示“从计算性能功耗比的角度上讲，和一般的数字芯片相比，TrueNorth并没有优势，反而处于劣势地位”。Culurciello认为TrueNorth之所以功耗低，仅仅是因为它的频率比主流硬件低约1亿倍。

寒武纪1号深度学习处理器在65nm工艺下面积仅为3mm2，速度和能效超过主流通用高性能CPU Intel Xeon E5-4620 100倍；而寒武纪2号深度学习处理器设计在28nm工艺下面积仅为67mm2，速度可达当前主流高性能GPU Nvidia K20M的21倍，而能效更是达到了330倍。自2014年中科院提出首个深度学习处理器“寒武纪”之后，深度学习处理器已经成为ISCA最关注的研究方向之一。ISCA 2016上有近1/6的论文（来自MIT、Stanford和UCSB等国际知名机构）引用“寒武纪”的工作来进行深度学习处理器探索。

目前看来，人工神经网络芯片的研究已经进入成熟期，正在进入产业化推广阶段，和相应的软件智能算法、操作系统相互促进，正在形成智能计算生态。脉冲神经网络芯片的研究还需要5-10年才能进入成熟期，需要相应的模型和基础元器件有所突破才能在精度等指标上达到目前人工神经网络的水平，其极低功耗等特点具有显著的优势，未来有很大发展潜力。

**5.3 类脑计算模型与应用**

类脑计算模型的相关研究国内起步相对较晚，虽然不同领域的学者也较早有分别从本领域出发的若干零星研究，但规模较小，尚未形成一个联系紧密的学术共同体。近年来，随着脑科学与人工智能的兴起，已越来越多的研究者开始关注并逐步投身于类脑计算领域的理论、技术、及应用研究，包括类脑计算模型、类脑学习算法、脑信息编解码等。中国计算机学会、中国人工智能学会等国家一级学会近年来持续举行过多个相关的专题论坛与研讨会，同时已成立或正在筹备相关的专业委员会。相信在不久的将来，在国内学者的努力下，很快能取得一批国际上有影响力的研究成果。

类脑计算模型的一个重要应用领域是类脑芯片，也被称为为神经形态芯片。浙江大学的达尔文芯片和清华大学的天机芯片都采用了脉冲神经网络计算模型，目前初步展示了在手写体识别、脑电波编解码和图像识别等问题中的应用。随着大规模脉冲神经网络芯片克服多层网络学习和扩展性等障碍，脉冲神经网络芯片将会在信息处理和计算方面取得更为广泛和深入的应用。

类脑计算的重要应用领域还包括智能硬件、类脑信息处理和认知机器人。四川大学类脑计算研究中心提出的脉冲编码和脉冲驱动学习算法在模式识别、语音识别和动态视觉传感器等问题中都取得了应用，尤其在噪音环境下的声音识别问题中体现了脉冲神经计算相比于传统机器学习算法的优势。而基于位置神经元和网格神经元的仿脑GPS建立了不同于现有的机器人依赖于精确物理空间信息的SLAM（同时定位与建图）算法，而是更加灵活和自适应环境变化的认知导航算法，将为机器人定位与导航提供新的思路与方法。

**6 发展趋势与展望**

**6.1 神经形态器件**

基于硅基的神经形态器件，如SRAM和DRAM，他们的优点是工艺成熟，动态功耗相对较低，缺点是需要持续供电以维持存储器的状态，静态功耗较高，一旦断电，信息全部丢失，需要重新写入，而且，他们在实现仿生学习法则方面，比较困难。因此，在该领域的研究方向主要是固态存储器，包括忆阻器（memristor），电桥式阻抗存储器（CBRAM），相变存储器（PCRAM），磁存储器（MRAM）等。

基于氧化物的忆阻器具有稳定性好、模拟性能好、可擦写次数高、数据保存时间长和兼容当下半导体工艺等优点，也是当下重点研究对象之一。在减少耗能，仿生学习等方面，取得巨大的进展。但其非线性不够理想，当大规模集成时，串扰将是一个需要解决的问题。另外，器件的稳定性和可控性仍有待提高，需要重视由器件自身及器件间的异变带来的对神经网络的负面影响。

电桥式阻抗存储器的耗能较低，具有较高的非线性，在模拟学习法则方面，具有独特的优势。但其异变较大，可重复性较差、不易控制，而且所用材料与半导体后端兼容性较差，在材料的选用和集成方面，需要更多的研究。

自旋转移矩磁存储器具有集成密度较高，耗电较低，可擦写次数极高等优点。但其工艺制造相对复杂，成本较高，而且开关比有限，在电路设计及应用方面，需要投入更多的研究来补偿限制。

相变存储器是所有固态存储器中研究最为成熟的、离商业化最近的器件。研究表明，相变存储器工作电流与器件底面积成正比，当底面电极直径减小到20nm时，Set和Reset的功耗都将随之降低。另外，据报道，用碳纳米管做电极的相变存储器，可达到更低的功耗（100 fJ）[150]。由于相变存储器的单向特性，可以用二极管做选择器件，很容易的实现三维交叉式堆栈结构，再加上优异的读写次数1012和超低的功耗[151]，是一种非常有潜力的、可以大规模应用的突触形态器件候选者，但其单向性在实现学习算法等方面，增加了一定的难度。

另外，在系统设计与集成方面，该领域仍处于前期研发阶段。其中最为重要的是整个系统架构还没有公认的技术方案，另外如何实现高密度和低功耗的大规模集成，并实现对仿生法则的凝练、选择、模拟，将是一项重大的挑战。

在所有问题之上，有一个最重要的问题，那就是模拟生物的形态器件，到底需要精确模拟到哪个层级？哪些仿生学习法则是有效的和必要的？在这个领域，我们亟需一个像牛顿定律或冯·诺依曼架构那样的、具有指导性的法则。

**6.2 神经网络芯片**

随着谷歌的AlphaGo程序在围棋比赛中战胜人类世界冠军李世石，英国《金融时报》认为，智能时代已经到来。而以智能手机、智能驾驶、智能制造、机器人和智能家居为代表的智能产品将会需要新的智能芯片，而目前的智能芯片主要是指各种神经网络芯片。从技术发展看，人工神经网络芯片预期将在近期取得较广泛的实际应用进展；脉冲神经网络芯片尚处于探索性应用阶段；视觉处理芯片则专门用于完成图像和视频处理任务。

目前我们已经在神经网络芯片、智能操作系统、智能算法方面具备了相应的研究基础，形成了研究链。神经网络芯片属于高端集成电路芯片制造。一方面，我们智能终端生产企业和互联网应用服务企业有着大量的神经网络芯片采购需求，如百度、科大讯飞、腾讯、阿里等大数据云计算龙头企业需要神经网络芯片提高服务质量降低运营成本，京东、华为等终端软硬件企业需要神经网络芯片提供服务内容升级。但是我国集成电路产业整体仍然落后于国际发达国家，整个产业从生产制造到市场受到跨国垄断企业明显压制。受行业特点和国内芯片制造技术发展水平的制约，高端芯片研发制造成本对于创新企业仍然高昂。

神经网络芯片属于整个智能产业链可以三个层次——应用层，技术层，芯片层的最底层。从上到下随着开发门槛的变化，能参与其中的公司越来越少。在应用层，会有海量的公司，主要是将智能应用在各个具体的行业中，例如智能+机器人，智能+客服，智能+制造，智能+商业分析，智能+手机等等。因此，在这个层面上，产值高，但是总体散，单个公司辐射面小。再下面的技术层，则是提供应用所需的核心共性智能算法、模型和方法。例如，不管机器人还是客服都需要语音的智能识别，那就需要语音的智能技术例如讯飞，图像视频的智能技术，自然语言理解的智能技术等。具备这个能力的公司就非常少了。最下面的是硬件基础层。无论应用和技术怎么变，最终都是要跑在芯片上的。而芯片开发门槛高，公司更少。像信息时代，也就是Intel，arm寥寥几家。最终主导神经网络芯片的公司也非常地少。因此掌握了神经网络芯片，就能从产业链的最源头全方位辐射整个智能产业。对于从事神经网络芯片开发的公司，应该予以支持。

综上所述，神经网络芯片设计制造属于高端制造业，要发展神经网络芯片技术，需要重点支持神经网络芯片开发公司，发展高端智能产品，制定产品标准、提供政策法规支持，支持智能芯片产业联盟。

**6.3 类脑计算模型与应用**

类脑计算计算经过几十年的发展，取得了令人欣喜的成果。然而，受限于我们对复杂神经系统的认知，我们只是揭开了其神秘面纱的一角。随着实验技术的进步与解剖学、生物学和生理学等相关领域研究的发展，我们必将对大脑的计算模式有更深、更多的了解，实现类脑计算将不是梦想。可以预见，类脑计算将是未来很长一段时间人类科学技术研究的热点和前沿。同时，随着人类对于智能系统、智能机器人的需求不断增加，类脑计算将具有非常广阔的应用前景[152]。我们将从以下几个方向进行展望：

（1） 硬件：基于亚阈值的模拟混合信号芯片可以做到超低功耗，但基于模拟混合信号的应用开发十分困难，在可编程性方面输给了数字电路技术。所以可在硬件方面进行器件的优化和改进，降低芯片功耗。类脑计算的主要优势是功耗低。如何更好的发挥类脑计算的优势将会成为研究的一个重要方向。其中，软硬件的有效结合将会推动类脑计算快速发展。

（2） 软件：为了在硬件层次提高芯片的性能 ，需要首先在软件层次进行仿真。所以需要开发一定的软件平台进行算法或模型的仿真和模拟。如果能够在软件仿真方面发现算法或模型取得突破性进展，加上硬件器的加速 ，就可以促进芯片性能的提升 。

（3） 计算模型：脉冲神经网络训练算法的理论发展还不够成熟，在计算神经科学和脉冲神经网络的研究中，人们往往孤立地研究神经信息编码或学习，很少有研究将编码和学习统一为一个完整的计算过程并综合考虑编码与学习之间的适应性问题。因此，尽管研究者针对脉冲神经网络提出了各种学习算法，其主要用于如何识别各种时空数据，而这些算法是否能有效识别实际的输入信息（如图像或语音）却依然是个未知数。因此，研究具有高效信息处理能力的基于完全时间编码的多层SNN（比如优化脉冲编码方法，寻找高效的学习算法），将为SNN在类脑计算领域中的应用提供有效的计算模型和理论基础

（4） 仿生感知：目前已有传感器可以采集到脉冲数据，这为类脑计算机的输入提供了有力支撑。随着大规模视频监控、无人驾驶等应用需求的出现，对更灵敏、更智能和复杂度更低的机器感知算法是必然发展方向，脉冲阵列式感知信号表达和基于脉冲神经网络的检测跟踪识别算法必将成为未来的主流。

（5） 应用：类脑计算技术的发展将推动图像识别、语音识别、自然语言处理等前沿技术的突破，机器人视听感知和自主学习是很有希望的应用方向，无人驾驶等领域应用需求也很迫切，类脑计算应用有望推动新一轮产业革命。

目前类脑计算离工业界实际应用还有较大差距，这也为研究者提供了重要研究方向与机遇。我们认为基于脉冲神经网络的类脑芯片在未来5~10年将会是一个重要的研究题目，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些方面取得突破进展，其中视听感知是最有希望率先突破的方向。

**致谢**

本报告的撰写得到北京大学、清华大学、中国科学院计算技术研究所、浙江大学、四川大学和华中科技大学相关研究团队研究人员的大力支持，特此致谢。

**参考文献**

1. 尼克. 人工智能的缘起.中国计算机学会通讯，2016，12（3）：38-43
2. Nilsson, Nils J.. The Quest for Artificial Intelligence: A History of Ideas and Achievements, Cambridge University Press, 2010.
3. 黄铁军.人类能够制造出“超级大脑”吗？ [N].中华读书报：2015年1月7日5版.
4. 黄铁军.类脑计算机的现在与未来[N]. 光明日报：2015年12月6日8版.
5. G E Hinton, R R Salakhutdinov (2006), “Reducing the dimensionality of data with neural networks”, Science (313)5786:504-507
6. Paul A. Merolla et al (2014), “A million spiking-neuron integrated circuit with a scalable communication network and interface”, Science (345)6197:668-673
7. John von Neumann (1958), “The Computer and the Brain”, Yale University Press
8. T. Huang, “Brain-Like Computing," Computing Now, vol.9, no.5, May. 2016, IEEE Computer Society [online]; https://www.computer.org/web/computingnow/archive/brain-like-computing-may-2016.
9. 顾宗华, 潘纲. 神经拟态的类脑计算研究. 中国计算机学会通讯. 2015, 11(10): 10-20.
10. 美国脑计划[EB/OL].[2016-07-25]. http://www.whitehouse.gov/share/brain-initiative
11. 美国脑计划NIH网站[EB/OL]. [2016-07-25].http://www.nih.gov/science/brain
12. 郭爽. 综述：美国“脑计划”这三年. 新华社[2016-04-27]
13. 美国NSF脑科学专题网站[EB/OL]. [2016-07-25]. http://www.nsf.gov/news/special\_reports/brain/
14. 美国情报高级研究计划署（IARPA）大脑皮层网络机器智能（MICrONS）计划[EB/OL].https://www.iarpa.gov/index.php/research-programs/microns
15. 游文娟. 美国如何推进BRAIN计划？生命科学研究快报. 2014年5月.
16. 欧盟人类大脑计划[EB/OL].[2016-07-25].www.humanbrainproject.eu/
17. 瑞士蓝脑计划[EB/OL].[2016-07-25]. http://bluebrain.epfl.ch/
18. Stefan Theil. Why the Human Brain Project Went Wrong—and How to Fix It. Scientific American. October 1, 2015.
19. Okano H, Miyawaki A, Kasai K. Brain/MINDS: brain-mapping project in Japan. Philosophical Transactions of the Royal Society B: Biological Sciences. 2015;370(1668):20140310. doi:10.1098/rstb.2014.0310.
20. 韩国《大脑科学发展战略》[EB/OL].[2016-07-25]. http://world.huanqiu.com/exclusive/2016-05/8992410.html
21. G. Indiveri, B. Linares-Barranco, T. J. Hamilton, A. Van Schaik, R. Etienne-Cummings, T. Delbruck, et al., "Neuromorphic silicon neuron circuits," Frontiers in neuroscience, vol. 5, p. 73, 2011.
22. F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, et al., "TrueNorth: Design and tool flow of a 65 mW 1 million neuron programmable neurosynaptic chip," IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, vol. 34, pp. 1537-1557, 2015.
23. M. M. Khan, D. R. Lester, L. A. Plana, A. Rast, X. Jin, E. Painkras, et al., "SpiNNaker: mapping neural networks onto a massively-parallel chip multiprocessor," in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 2849-2856.
24. S. B. Furber, D. R. Lester, L. A. Plana, J. D. Garside, E. Painkras, S. Temple, et al., "Overview of the SpiNNaker system architecture," IEEE Transactions on Computers, vol. 62, pp. 2454-2467, 2013.
25. G. Snider, R. Amerson, D. Carter, H. Abdalla, M. S. Qureshi, J. Léveillé, et al., "From synapses to circuitry: Using memristive memory to explore the electronic brain," Computer, vol. 44, p. 21, 2011.
26. G. Livitz, M. Versace, A. Gorchetchnikov, H. Ames, J. Léveillé, B. Chandler, et al., "Adaptive, brain-like systems give robots complex behaviors."
27. K. Meier, "A mixed-signal universal neuromorphic computing system," in 2015 IEEE International Electron Devices Meeting (IEDM), 2015, pp. 4.6. 1-4.6. 4.
28. J. Schemmel, J. Fieres, and K. Meier, "Wafer-scale integration of analog neural networks," in 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 431-438.
29. B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, et al., "Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations," Proceedings of the IEEE, vol. 102, pp. 699-716, 2014.
30. P. Gao, B. V. Benjamin, and K. Boahen, "Dynamical system guided mapping of quantitative neuronal models onto neuromorphic hardware," IEEE Transactions on Circuits and Systems I: Regular Papers, vol. 59, pp. 2383-2394, 2012.
31. G. Indiveri, F. Corradi, and N. Qiao, "Neuromorphic architectures for spiking deep neural networks," in 2015 IEEE International Electron Devices Meeting (IEDM), 2015, pp. 4.2. 1-4.2. 4.
32. G. Indiveri and S.-C. Liu, "Memory and information processing in neuromorphic systems," Proceedings of the IEEE, vol. 103, pp. 1379-1397, 2015.
33. D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," nature, vol. 453, pp. 80-83, 2008.
34. J. J. Yang, M. D. Pickett, X. Li, D. A. Ohlberg, D. R. Stewart, and R. S. Williams, "Memristive switching mechanism for metal/oxide/metal nanodevices," Nature nanotechnology, vol. 3, pp. 429-433, 2008.
35. G. S. Snider, "Spike-timing-dependent learning in memristive nanodevices," in 2008 IEEE International Symposium on Nanoscale Architectures, 2008, pp. 85-92.
36. S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," Nano letters, vol. 10, pp. 1297-1301, 2010.
37. S. Kim, S. Choi, and W. Lu, "Comprehensive physical model of dynamic resistive switching in an oxide memristor," ACS nano, vol. 8, pp. 2369-2376, 2014.
38. S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, and H.-S. P. Wong, "An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation," IEEE Transactions on Electron Devices, vol. 58, pp. 2729-2737, 2011.
39. S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H. S. P. Wong, "A low energy oxide‐based electronic synaptic device for neuromorphic visual systems with tolerance to device variation," Advanced Materials, vol. 25, pp. 1774-1779, 2013.
40. W. He, K. Huang, N. Ning, K. Ramanathan, G. Li, Y. Jiang, et al., "Enabling an integrated rate-temporal learning scheme on memristor," Scientific reports, vol. 4, 2014.
41. T. Chang, S.-H. Jo, and W. Lu, "Short-term memory to long-term memory transition in a nanoscale memristor," ACS nano, vol. 5, pp. 7669-7676, 2011.
42. T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, and M. Aono, "Short-term plasticity and long-term potentiation mimicked in single inorganic synapses," Nature materials, vol. 10, pp. 591-595, 2011.
43. R. Berdan, E. Vasilaki, A. Khiat, G. Indiveri, A. Serb, and T. Prodromakis, "Emulating short-term synaptic dynamics with memristive devices," Scientific reports, vol. 6, 2016.
44. D. Kuzum, R. G. Jeyasingh, B. Lee, and H.-S. P. Wong, "Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing," Nano letters, vol. 12, pp. 2179-2186, 2011.
45. D. Kuzum, S. Yu, and H. P. Wong, "Synaptic electronics: materials, devices and applications," Nanotechnology, vol. 24, p. 382001, 2013.
46. M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors," Nature materials, vol. 12, pp. 114-117, 2013.
47. A. F. Vincent, J. Larroque, N. Locatelli, N. B. Romdhane, O. Bichler, C. Gamrat, et al., "Spin-transfer torque magnetic memory as a stochastic memristive synapse for neuromorphic systems," IEEE transactions on biomedical circuits and systems, vol. 9, pp. 166-174, 2015.
48. G. Indiveri and S.-C. Liu,"Memory and information processing in neuromorphic systems,"Proceedings of the IEEE, 103(8): 1379–1397, 2015.
49. N. Dehorter, G. Ciceri, G. Bartolini, et al,"Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch," Science, 349(6253): 1216–1220, 2015.
50. Dayan, P., & Abbott, L. F. (2001). Theoretical Neuroscience. Cambridge: MIT Press.
51. Brette R. and Gerstner W., “Adaptive exponential integrate-and-fire model as an effective description of neuronal activity,” J. Neurophysiol., vol. 94, pp. 3637 – 3642, 2005.
52. Izhikevich E., Which model to use for cortical spiking neurons? Neural Networks, IEEE Transactions on, 15 (5), 1063 – 1070, 2003.
53. Tim Dettmers, The Brain vs Deep Learning, https://timdettmers.wordpress.com/2015/07/27/brain-vs-deep-learning-singularity/.
54. S. Panzeri, N. Brunel, N. K. Logothetis, C Kayser," Sensory neural codes using multiplexed temporal scales," Trends in Neurosciences, 33(3):111–120, 2010.
55. D. A. Butts, C. Weng, J. Jin et al," Temporal precision in the neural code and the timescales of natural vision," Nature, 449(7158):92-95, 2007.
56. J. J. Hopfield," Pattern recognition computing using action potential timing for stimulus representation," Nature, 376: 33-36, 1995.
57. J. M. Samonds, Z. Zhou et al," Synchronous activity in cat visual cortex encodes collinear and cocircular contours," Journal of Neurophysiology, 95(4): 2602–2616, 2006.
58. S. Leutgeb, J. Leutgeb, M.-B. Moser, and E. Moser," Place cells, spatial maps and the population code for memory," Current Opinion in Neurobiology, 15(6): 738–746, 2005.
59. Q. Yu, H. Tang, K. C. Tan, et al," Rapid feedforward computation by temporal encoding and learning with spiking neurons," IEEE Transactions on Neural Networks and Learning Systems, 24(10): 1539-1552. 2013.
60. J. Dennis, Q. Yu, H. Tang et al," Temporal coding of local spectrogram features for robust sound recognition," Proc of ICASSP 2013, IEEE, 803-807, 2013.
61. G. Bi, M. Poo," Distributed synaptic modification in neural networks induced by patterned stimulation," Nature, 401(6755): 792-796, 1999.
62. J. Gjorgjieva, C. Clopath, J. Audet, and J.-P. Pfister," A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations," Proceedings of the National Academy of Sciences, 108(48): 19383–19388, 2011.
63. R. Gutig, H. Sompolinsky," The Tempotron: A neuron that learns spike timing-based decisions," Nature Neuroscience, 9(3): 420–428, 2006．
64. F. Ponulak and A. Kasinski," Supervised learning in spiking neural networks with resume: Sequence learning, classification, and spike shifting," Neural Computation, 22(2): 467–510, 2010.
65. A. Mohemmed, S. Schliebs," SPAN: Spike pattern association neuron for learning spatio-temporal spike patterns," International Journal of Neural Systems, 22(4): 1250012, 2012．
66. R. V. Florian, "The Chronotron: A neuron that learns to fire temporally precise spike patterns," PLoS One, 7(8): e40233, 2012．
67. M. Rossum," A novel spike distance," Neural Computation, 13: 751–763, 2001.
68. J. D. Victor, K. P. Purpura," Metric-space analysis of spike trains: theory, algorithms and application," Network: Computation in Neural Systems, 8: 127–164, 1997.
69. B. Gardner, A. Gruning," Classifying patterns in a spiking neural network," In Proceedings of the 22nd European Symposium on Artificial Neural Networks (ESANN2014) (23–28). New York: Springer, 2014.
70. R. Gütig," To spike, or when to spike? ,"Current Opinion in Neurobiology, 25: 134–139, 2014.
71. S. Mckennoch, T. Voegtlin, L. Bushnell," Spike timing error backpropagation in theta neuron networks," Neural Computation, 21(1): 9–45, 2009.
72. S. Ghosh-Dastidar, H. Adeli," A new supervised learning algorithm for multiple spiking neural networks with application in epilepsy and seizure detection," Neural Networks, 22(10): 1419–1431, 2009.
73. I. Sporea, & A. Gruning," Supervised learning in multilayer spiking neural networks," Neural Computation, 25(2), 473–509, 2013.
74. Y. Xu, X. Zeng, S. Zhong," A new supervised learning algorithm for spiking neurons," Neural Computation, 25(6): 1472–1511, 2013.
75. Diehl P U, Neil D, Binas J, et al. Fast-Classifying, High-Accuracy Spiking Deep Networks Through Weight and Threshold Balancing[J]. 2015
76. 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.
77. 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.
78. 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].
79. E.Y. Cheu, J. Yu, C. H. Tan, H. Tang. Synaptic Conditions for Auto-Associative Memory Storage and Pattern Completion in Jensen et al.'s Model of Hippocampal Area CA3. Journal of Computational Neuroscience, vol. 33, no. 3, pp. 435-447, 2012.
80. W. Gerstner, W. Kistler. Spiking Neuron Models. Cambridge University Press, 2002.
81. Chris Eliasmith, Charles H. Anderson. Neural Engineering: Computation, Representation, and Dynamics in Neurobiological Systems. Cambridge: The MIT Press, 2003.
82. Chris Eliasmith, Terrence C. Stewart, Xuan Choo, Trevor Bekolay,Travis DeWolf, Yichuan Tang, Daniel Rasmussen. A Large-scale Model of the Functioning Brain. Science, 2012, 338(6111):1202-1205.
83. 曾毅，刘成林，谭铁牛. 类脑智能研究的回顾与展望. 计算机学报, 2016(01):212-222
84. Jeff Hawkins. On Intelligence. New York: Times Books, 2004.
85. Jeff Hawkins, Subutai Ahmad, D Dubinsky. Hierarchical Temporal Memory including HTM cortical learning algorithms. Technical Report, Numenta, Inc. 2011.
86. Mahar M. A. C., Deweerth S. P., Mahowald M. A., et al. Imaplementing neural architecture using analog VLSI circuits. IEEE T. on Circuits Systems, 1989, 36(5):643-652
87. Mahoward M. A., Mead C. A. The silicon retina, Scientic American, 1991,264(5):76-82
88. C. A. Mead. Analog VLSI and Neural Systems. Reading, MA:Addison-Wesley, 1989
89. Liu S C, Delbruck T. Neuromorphic sensory systems[J]. Current opinion in neurobiology, 2010, 20(3): 288-295.
90. Xinyuan Qian, Hang Yu, Shoushun Chen and Kay Soon Low, "A High Dynamic Range CMOS Image Sensor with Dual-Exposure Charge Subtraction Scheme,"  IEEE Sensors Journal, Volume 15 , Issue 2, pp. 661 - 662, 2015.
91. Bo Zhao, Shoushun Chen, Bernabe Linares-Barranco and Huajin Tang, “Feedforward categorization on AER motion events using cortex-like features and spiking neural network," IEEE Transactions on Neural Networks and Learning Systems (TNNLS), Volume 26 , Issue 9 , pp. 1963 - 1978, 2015.
92. Yifei Liu, Xiaoyu Yu, Shoushun Chen and Wei Tang, "Object Localization and Size Measurement Using Networked Address Event Representation Imagers," IEEE Sensors Journal, Volume 16 , Issue 9 , pp. 2894 - 2895, 2016.
93. Garrick Orchard, Cedric Meyer, Ralph Etienne-Cummings, Christoph Posch, Nitish Thakor, Ryad Benosman, "HFirst: A Temporal Approach to Object Recognition", IEEE Transactions on Pattern Analysis & Machine Intelligence, vol.37, no. 10, pp. 2028-2040, Oct. 2015, doi:10.1109/TPAMI.2015.2392947
94. 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.
95. 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.
96. 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.
97. Pfeil T, Grübl A, Jeltsch S, et al. Six networks on a universal neuromorphic computing substrate[J]. Frontiers in neuroscience, 2013, 7.
98. 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.
99. 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
100. J. Hu, H. Tang, K.C. Tan, H. Li, L. Shi. A Spike-Timing Based Integrated Model for Pattern Recognition. Neural Computation, vol. 25, no. 2, pp. 450-472, 2013.
101. [31] Q. Yu, H. Tang, K. C. Tan, and H. Li. Rapid Feedforward Computation by Temporal Encoding and Learning with Spiking Neurons. IEEE Trans. on Neural Networks and Learning Systems,vol. 24, no. 10, pp. 1539-1552, 2013.
102. B. Zhao, R. Ding; S. Chen, B. Linares-Barranco, and H. Tang. Feedforward Categorization on AER Motion Events using Cortex-like Features in a Spiking Neural Network. IEEE Trans. on Neural Networks and Learning Systems, DOI:10.1109/TNNLS.2014.2362542, 2015.
103. J. Dennis, Q. Yu, H. Tang et al," Temporal coding of local spectrogram features for robust sound recognition," Proc of ICASSP 2013, IEEE, 803-807, 2013.
104. M. Beyelera, N. D. Dutta, J. L. Krichmar. Categorization and decision-making in a neurobiologically plausible spiking network using a STDP-like learning rule. Neural Networks, vol. 48, pp. 109–124, 2013.
105. M. Yuan, B. Tian, V. A. Shim, H. Tang and H. Li.  An Entorhinal-Hippocampal Model for Simultaneous Cognitive Map Building. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-15), Austin, TX, USA, pp. 586-592, 2015.
106. B. Tian, W. Huang, C. H. Tan, , H. Tang, and H. Li. Robotic Spatial Cognition by Cognitive Map and Episodic Memory. Unpublished work.
107. 唐华锦, 胡隽. 神经形态认知计算. 中国计算机学会通讯. 2015年第10期.
108. 钱万强，朱庆平，沈建磊. 我国脑科学研究战略部署及发展情况浅析.《生命科学》, 2014(06):545-549
109. 北京脑计划专版. 人民日报. 2015年11月17日.
110. Y. Yang, B. Chen, and W. D. Lu, "Memristive Physically Evolving Networks Enabling the Emulation of Heterosynaptic Plasticity," Advanced Materials, vol. 27, pp. 7720-7727, 2015.
111. Z. Wang, M. Yin, T. Zhang, Y. Cai, Y. Wang, Y. Yang, et al., "Engineering incremental resistive switching in TaO x based memristors for brain-inspired computing," Nanoscale, 2016.
112. Y. Yang, J. Lee, S. Lee, C. H. Liu, Z. Zhong, and W. Lu, "Oxide Resistive Memory with Functionalized Graphene as Built‐in Selector Element," Advanced Materials, vol. 26, pp. 3693-3699, 2014.
113. J. Kang, B. Gao, P. Huang, H. Li, Y. Zhao, Z. Chen, et al., "Oxide-based RRAM: Requirements and challenges of modeling and simulation," in 2015 IEEE International Electron Devices Meeting (IEDM), 2015, pp. 5.4. 1-5.4. 4.
114. Y. Zhang, H. Wu, Y. Bai, A. Chen, Z. Yu, J. Zhang, et al., "Study of conduction and switching mechanisms in Al/AlOx/WOx/W resistive switching memory for multilevel applications," Applied Physics Letters, vol. 102, p. 233502, 2013.
115. Y. Bai, H. Wu, Y. Zhang, M. Wu, J. Zhang, N. Deng, et al., "Low power W: AlOx/WOx bilayer resistive switching structure based on conductive filament formation and rupture mechanism," Applied Physics Letters, vol. 102, p. 173503, 2013.
116. Y. Bai, H. Wu, R. Wu, Y. Zhang, N. Deng, Z. Yu, et al., "Study of multi-level characteristics for 3D vertical resistive switching memory," Scientific reports, vol. 4, 2014.
117. L. Deng, D. Wang, Z. Zhang, P. Tang, G. Li, and J. Pei, "Energy consumption analysis for various memristive networks under different learning strategies," Physics Letters A, vol. 380, pp. 903-909, 2016.
118. C. Wan, X. Zhang, X. Gao, J. Wang, and X. Tan, "Geometrical enhancement of low-field magnetoresistance in silicon," Nature, vol. 477, pp. 304-307, 2011.
119. H. Tian, W. Mi, X.-F. Wang, H. Zhao, Q.-Y. Xie, C. Li, et al., "Graphene dynamic synapse with modulatable plasticity," Nano letters, vol. 15, pp. 8013-8019, 2015.
120. L. Q. Zhu, C. J. Wan, L. Q. Guo, Y. Shi, and Q. Wan, "Artificial synapse network on inorganic proton conductor for neuromorphic systems," Nature communications, vol. 5, 2014.
121. C. J. Wan, L. Q. Zhu, Y. H. Liu, P. Feng, Z. P. Liu, H. L. Cao, et al., "Proton‐Conducting Graphene Oxide‐Coupled Neuron Transistors for Brain‐Inspired Cognitive Systems," Advanced Materials, vol. 28, pp. 3557-3563, 2016.
122. N. Liu, L. Q. Zhu, P. Feng, C. J. Wan, Y. H. Liu, Y. Shi, et al., "Flexible sensory platform based on oxide-based neuromorphic transistors," Scientific reports, vol. 5, 2015.
123. J. Zhou, N. Liu, L. Zhu, Y. Shi, and Q. Wan, "Energy-efficient artificial synapses based on flexible IGZO electric-double-layer transistors," IEEE Electron Device Letters, vol. 36, pp. 198-200, 2015.
124. H. Xu, Y. Xia, K. Yin, J. Lu, Q. Yin, J. Yin, et al., "The chemically driven phase transformation in a memristive abacus capable of calculating decimal fractions," Scientific reports, vol. 3, 2013.
125. H. Xu, Y. Xia, B. Xu, J. Yin, G. Yuan, and Z. Liu, "Encoding, training and retrieval in ferroelectric tunnel junctions," Scientific reports, vol. 6, p. 27022, 2016.
126. M. Xia, M. Zhu, Y. Wang, Z. Song, F. Rao, L. Wu, et al., "Ti–Sb–Te Alloy: A Candidate for Fast and Long-Life Phase-Change Memory," ACS applied materials & interfaces, vol. 7, pp. 7627-7634, 2015.
127. M. Xia, K. Ding, F. Rao, X. Li, L. Wu, and Z. Song, "Aluminum-centered tetrahedron-octahedron transition in advancing Al-Sb-Te phase change properties," Scientific reports, vol. 5, p. 8548, 2015.
128. F. Rao, Z. Song, Y. Cheng, X. Liu, M. Xia, W. Li, et al., "Direct observation of titanium-centered octahedra in titanium-antimony-tellurium phase-change material," Nature communications, vol. 6, 2015.
129. Y. Zhong, Y. Li, L. Xu, and X. Miao, "Simple square pulses for implementing spike‐timing‐dependent plasticity in phase‐change memory," physica status solidi (RRL)-Rapid Research Letters, vol. 9, pp. 414-419, 2015.
130. Y. Li, Y. Zhong, J. Zhang, L. Xu, Q. Wang, H. Sun, et al., "Activity-dependent synaptic plasticity of a chalcogenide electronic synapse for neuromorphic systems," Scientific reports, vol. 4, 2014.
131. Y. Zhou, Y. Li, L. Xu, S. Zhong, R. Xu, and X. Miao, "A hybrid memristor‐CMOS XOR gate for nonvolatile logic computation," physica status solidi (a), 2015.
132. Z. Zeng and W. X. Zheng, "Multistability of neural networks with time-varying delays and concave-convex characteristics," IEEE Transactions on Neural networks and learning systems, vol. 23, pp. 293-305, 2012.
133. G. Bao and Z. Zeng, "Analysis and design of associative memories based on recurrent neural network with discontinuous activation functions," Neurocomputing, vol. 77, pp. 101-107, 2012.
134. S. Wen, T. Huang, Z. Zeng, Y. Chen, and P. Li, "Circuit design and exponential stabilization of memristive neural networks," Neural Networks, vol. 63, pp. 48-56, 2015.
135. S. Wen, Z. Zeng, T. Huang, and C. Li, "Passivity and passification of stochastic impulsive memristor‐based piecewise linear system with mixed delays," International Journal of Robust and Nonlinear Control, vol. 25, pp. 610-624, 2015.
136. Z. Tang, L. Fang, N. Xu, and R. Liu, "Forming compliance dominated memristive switching through interfacial reaction in Ti/TiO2/Au structure," Journal of Applied Physics, vol. 118, p. 185309, 2015.
137. F. Xu-Dong, T. Yu-Hua, W. Jun-Jie, Z. Xuan, Z. Jing, and H. Da, "SPICE modeling of flux-controlled unipolar memristive devices," Chinese Physics B, vol. 22, p. 078901, 2013.
138. X. Fang, X. Yang, J. Wu, and X. Yi, "A compact SPICE model of unipolar memristive devices," IEEE Transactions on Nanotechnology, vol. 12, pp. 843-850, 2013.
139. X. Zhu, X. Yang, C. Wu, N. Xiao, J. Wu, and X. Yi, "Performing stateful logic on memristor memory," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 60, pp. 682-686, 2013.
140. J. Zhou, X. Yang, J. Wu, X. Zhu, X. Fang, and D. Huang, "A memristor-based architecture combining memory and image processing," Science China Information Sciences, vol. 57, pp. 1-12, 2014.
141. X. Ma, H. Wu, D. Wu, and H. Qian, "A 16 Mb RRAM test chip based on analog power system with tunable write pulses," in 2015 15th Non-Volatile Memory Technology Symposium (NVMTS), 2015, pp. 1-3.
142. H. W. Peng Yao, Guo Zhang, Ning Deng, and He Qian, "A Two-Layer Perceptron Network Implemented Based on 1T1R Memristor Array," International Conference on Computed Aided Design (ICCAD)，Austin, November 2015.
143. P. Yao, H. Wu, B. Gao, G. Zhang, and H. Qian, "The effect of variation on neuromorphic network based on 1T1R memristor array," in 2015 15th Non-Volatile Memory Technology Symposium (NVMTS), 2015, pp. 1-3.
144. L. Deng, G. Li, N. Deng, D. Wang, Z. Zhang, W. He, et al., "Complex learning in bio-plausible memristive networks," Scientific reports, vol. 5, 2015.
145. [王守觉](http://xueshu.baidu.com/usercenter/data/author?cmd=authoruri&wd=authoruri%3A%28c0c1f29c4b732d1c%29%20author%3A%28%E7%8E%8B%E5%AE%88%E8%A7%89%29%20%E4%B8%AD%E5%9B%BD%E7%A7%91%E5%AD%A6%E9%99%A2%E5%8D%8A%E5%AF%BC%E4%BD%93%E7%A0%94%E7%A9%B6%E6%89%80)，[鲁华祥](http://xueshu.baidu.com/usercenter/data/author?cmd=authoruri&wd=authoruri%3A%287fcb29cd46dedf52%29%20author%3A%28%E9%B2%81%E5%8D%8E%E7%A5%A5%29%20%E4%B8%AD%E5%9B%BD%E7%A7%91%E5%AD%A6%E9%99%A2%E5%8D%8A%E5%AF%BC%E4%BD%93%E7%A0%94%E7%A9%B6%E6%89%80%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C%E5%AE%9E%E9%AA%8C%E5%AE%A4)，[陈向东](http://xueshu.baidu.com/usercenter/data/author?cmd=authoruri&wd=authoruri%3A%28b9ce6d794bbd0d65%29%20author%3A%28%E9%99%88%E5%90%91%E4%B8%9C%29%20%E4%B8%AD%E5%9B%BD%E7%A7%91%E5%AD%A6%E9%99%A2%E5%8D%8A%E5%AF%BC%E4%BD%93%E7%A0%94%E7%A9%B6%E6%89%80). “[人工神经网络硬件化途径与神经计算机研究](http://xueshu.baidu.com/usercenter/data/paperhelp?cmd=paper_forward&longsign=74a786e86e6c65b61d4bd69c486c43b6&title=%E4%BA%BA%E5%B7%A5%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C%E7%A1%AC%E4%BB%B6%E5%8C%96%E9%80%94%E5%BE%84%E4%B8%8E%E7%A5%9E%E7%BB%8F%E8%AE%A1%E7%AE%97%E6%9C%BA%E7%A0%94%E7%A9%B6)”，深圳大学学报(理工版) ，1997
146. Tianshi Chen, Zidong Du, Ninghui Sun, Jia Wang, Chengyong Wu, Yunji Chen, and Olivier Temam, "DianNao: A Small-Footprint High-Throughput Accelerator for Ubiquitous Machine-Learning", in Proceedings of the 19th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'14), 2014.
147. Yunji Chen, Tao Luo, Shaoli Liu, Shijin Zhang, Liqiang He, Jia Wang, Ling Li, Tianshi Chen, Zhiwei Xu, Ninghui Sun, and Olivier Temam, "DaDianNao: A Machine-Learning Supercomputer", in Proceedings of the 47th IEEE/ACM International Symposium on Microarchitecture (MICRO'14), 2014.
148. [Juncheng Shen](http://dblp.uni-trier.de/pers/hd/s/Shen:Juncheng), [De Ma](http://dblp.uni-trier.de/pers/hd/m/Ma:De), [Zonghua Gu](http://dblp.uni-trier.de/pers/hd/g/Gu:Zonghua), [Ming Zhang](http://dblp.uni-trier.de/pers/hd/z/Zhang:Ming), [Xiaolei Zhu](http://dblp.uni-trier.de/pers/hd/z/Zhu:Xiaolei), [Xiaoqiang Xu](http://dblp.uni-trier.de/pers/hd/x/Xu:Xiaoqiang), [Qi Xu](http://dblp.uni-trier.de/pers/hd/x/Xu:Qi), [Yangjing Shen](http://dblp.uni-trier.de/pers/hd/s/Shen:Yangjing), Gang Pan，"Darwin: a neuromorphic hardware co-processor based on Spiking Neural Networks",[SCIENCE CHINA Information Sciences 59(2)](http://dblp.uni-trier.de/db/journals/chinaf/chinaf59.html#ShenMGZZXXSP16): 1-5 (2016)
149. [Yangfan Zhou](http://dblp.uni-trier.de/pers/hd/z/Zhou:Yangfan), [Zhongxiang Cao](http://dblp.uni-trier.de/pers/hd/c/Cao:Zhongxiang), [Qi Qin](http://dblp.uni-trier.de/pers/hd/q/Qin:Qi), [Quanliang Li](http://dblp.uni-trier.de/pers/hd/l/Li:Quanliang), [Cong Shi](http://dblp.uni-trier.de/pers/hd/s/Shi:Cong), Nanjian Wu,"A high speed 1000 fps CMOS image sensor with low noise global shutter pixels", [SCIENCE CHINA Information Sciences 57(4)](http://dblp.uni-trier.de/db/journals/chinaf/chinaf57.html#ZhouCQLSW14): 1-8 (2014)
150. F. Xiong, A. D. Liao, D. Estrada, and E. Pop, "Low-power switching of phase-change materials with carbon nanotube electrodes," Science, vol. 332, pp. 568-570, 2011.
151. G. W. Burr, M. J. Breitwisch, M. Franceschini, D. Garetto, K. Gopalakrishnan, B. Jackson, et al., "Phase change memory technology," Journal of Vacuum Science & Technology B, vol. 28, pp. 223-262, 2010.
152. 唐华锦, 胡隽. 神经形态认知计算. 中国计算机学会通讯. 2015年第10期.
153. Marr D. A theory for cerebral neocortex. Proceedings of the Royal Society of London B, 1970, 176:161-234.
154. Marr D. Simple memory:a theory for archicortex. Philosophical Transactions of the Royal Society B:Biological Sciences, 1971,262(841):23-81.
155. Marr D. Approaches to biological information processing. Science, 1975, 190:875-876.
156. Marr D, Poggio T. Cooperative computation of stereo disparity. Science, 1976, 194(4262):283-287.
157. Marr D. Vision:a computational investigation into the human representation and processing of visual information. New York, USA:W.H. Freeman and Company, 1982.
158. Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nature Neuroscience, 1999, 2(11):1019-1025.
159. Serre T, Oliva A, Poggio T. A feedforward architecture accounts for rapid categorization. Proceedings of the National Academy of Sciences, 2007, 104(15):6424-6429.
160. Francis Crick, Christof Koch, Towards a Neurobiological Theory of Consciousness. Seminars in the Neurosciences 2, (1990):263-275.
161. Christof Koch, Francis Crick, Understanding awareness at the neuronal level Behav Brain Sci , 14(04), 683-685 (1991)
162. Francis Crick & Christof Koch, A framework for consciousness, Nature Neuroscience 2003, (6)119-126
163. Christof Koch, The Quest for Consciousness: A Neurobiological Approach, Roberts & Company Publishers, 2004

**作者简介**

本报告的第1、2节和“脑计划”相关部分由北京大学黄铁军撰写，神经形态器件由清华大学的施路平撰写，神经网络芯片由中国科学院计算技术研究所陈云霁撰写，类脑计算模型和应用由浙江大学潘刚和四川大学唐华锦撰写，黄铁军负责统稿，华中科技大学于俊清负责策划和修改。以下以姓氏拼音排序。

|  |  |
| --- | --- |
|  | 陈云霁，博士，中国科学院计算技术研究所研究员，博士生导师，主要研究方向为计算机系统结构，CCF会员。 |
|  | 黄铁军，博士，北京大学信息科学技术学院教授，博士生导师，计算机科学技术系主任，数字媒体研究所所长，主要研究方向为视觉信息处理。CCF杰出会员，专委工作委员会执行委员。tjhuang@pku.edu.cn。 |
| C:\SkyDrive\MyResume\Photo-head\profile.small.jpg | 潘纲，博士，浙江大学计算学院教授、博士生导师，CCF高级会员，主要研究方向为计算机视觉、普适计算、智能系统 |
| http://www1.i2r.a-star.edu.sg/~htang/images/people/Huajin%20Tang.jpg | 唐华锦, 博士，四川大学计算机学院教授、博士生导师，类脑计算研究中心主任，主要研究方向为神经形态计算、认知机器人。 |
| **D:\my picture\SoftCopy-shi luping.jpg** | 施路平，博士，清华大学精仪系教授，博士生导师，清华大学类脑计算中心主任，国家光盘工程研究中心主任，主要研究方向类脑计算与信息存储。SPIE fellow。lpshi@tsinghua.edu.cn |
| D:\百度云\个人资料\Chinese\登记照\于俊清（2014.12.jpg | 于俊清，博士，华中科技大学计算机科学与技术学院，教授，博士生导师，数字媒体处理与检索实验室主任，主要研究方向为数字视频分析与检索、多核计算与流编译、网络安全大数据处理。中国计算机学会多媒体专业委员会副主任。yjqing@hust.edu.cn |