

神经形态器件的特性与发展

曹 震^{1,2}, 张浴轩^{1,2}, 李灵蕾^{1,2}, 郭 璋^{1,2}, 孙 琦^{1,2}, 曹荣荣³, 侯 彪^{1,2}, 焦李成^{1,2}

(1. 西安电子科技大学人工智能学院, 陕西西安 710126; 2. 西安电子科技大学杭州研究院, 浙江杭州 311231;
3. 国防科技大学电子科学学院, 湖南长沙 410073)

摘 要: 随着大数据时代的到来和人类对大脑系统研究的日渐深入, 类脑计算领域的研究取得突破性进展, 有希望在根源上打破传统计算机的性能瓶颈. 神经突触是对人体脑部进行记忆能力训练和处理数据的重要基础单元, 因此开发新材料、新结构, 研究基于新型人造材料与光电子器件的神经突触可塑性, 对神经形态器件研究和类脑硬件设计的实现都有着重要意义. 本文首先指出目前“冯·诺依曼架构”的主要性能瓶颈, 引出类脑计算的概念, 提出神经形态器件的主要性能优势, 并梳理神经形态器件发展历史; 然后在忆阻器领域, 阐述与分析忆阻类型、忆阻结构与忆阻机理, 比较出几种忆阻器的特性, 举例说明忆阻器在不同领域的应用; 接着以神经形态器件为基础, 选取磁性隧道结、新型浮栅管和铁电晶体管, 介绍其结构、工作原理与应用; 最后总结目前神经形态器件发展的成果和方向, 并对行业发展前景进行预测.

关键词: 类脑计算; 神经形态器件; 忆阻器; 磁性隧道结; 人造突触

基金项目: 国家自然科学基金(No.62104176, No.62171347, No.62004220, No.62301395); 中国博士后科学基金(No.2019M663637, No.2020M673696); 西安电子科技大学杭州研究院概念验证基金(No.GNYZ2023XJ0409-1); 陕西省自然科学基金基础研究计划(No.2022JQ-661); 中央高校基本科研业务费专项资金项目(No.XJSJ23083, No.XJS222215)

中图分类号: TP183

文献标识码: A

文章编号: 0372-2112(2023)12-3619-24

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230458

A Comprehension Survey on the Characterization and Development of Neuromorphic Devices

CAO Zhen^{1,2}, ZHANG Yu-xuan^{1,2}, LI Ling-lei^{1,2}, GUO Zhang^{1,2}, SUN Qi^{1,2},

CAO Rong-rong³, HOU Biao^{1,2}, JIAO Li-cheng^{1,2}

(1. School of Artificial Intelligence, Xidian University, Xi'an, Shaanxi 710126, China;

2. Hangzhou Research Institute of Technology, Xidian University, Hangzhou, Zhejiang 311231, China;

3. Electronic Engineering Institute of PLA, National University of Defense Technology, Changsha, Hunan 410073, China)

Abstract: With the advent of the big data era and the increasingly in-depth study of the human brain system, the field of neuromorphic computing has made breakthrough progress, offering hope to break through the performance issue of traditional computers at the root level. Neural synapses are important basic units for memory training and data processing in the human brain, therefore, it is of great significance for research on neural morphological devices and the implementation of neuromorphic hardware design to develop new materials and structures to study the plasticity of neural synapses based on novel artificial materials and optoelectronic devices. This paper firstly points out the main performance issue of Von Neumann architecture, draws forth the concept of brain-like computing, puts forward the main performance advantages of neuromorphic devices, and sorts out the development history of neuromorphic devices. Then in the field of memristors, the types of memristors, memristor structures and memristor mechanisms are described and analyzed, the advantages and disadvantages of several types of memristors are compared, and the examples of applications of memristors in different fields are presented. Next, based on neural morphological devices, the structures, working principles and applications of magnetic tunnel junctions, new floating gate transistors, and ferroelectric transistors are selected to introduce. Finally, this paper summarizes the achievements and directions of the current development of neural morphological devices and predicts the development prospects of the industry.

Key words: brain-like computing; neuromorphic device; memristor; magnetic tunnel junction; artificial synapse

Foundation Item(s): National Natural Science Foundation of China (No.62104176, No.62171347, No.62004220, No.62301395); China Postdoctoral Science Foundation Project (No.2019M663637, No.2020M673696); Proof-of-concept Fund Project of Hangzhou Research Institute of Xidian University (No.GNYZ2023XJ0409-1); Basic Research Program of Natural Sciences of Shaanxi Province (No.2022JQ-661); Special Fund Project for Basic Scientific Research Business Funds of Central Universities (No.XJSJ23083, No.XJS222215)

1 引言

1.1 冯·诺依曼瓶颈

在计算机诞生之前,人们只能依靠手工计算来完成,然而这种方法的计算精度和数量无法保证.因此,当时对计算机的需求十分迫切.20世纪50年代,被誉为“现代计算机之父”的冯·诺依曼提出了计算机的基本组成,包括输入设备、输出设备、控制器、运算器和存储器,其中,运算器与控制器共同组成了中央处理器(Central Processing Unit, CPU).计算机的执行过程是二进制程序指令通过输入设备输入到存储器中,接着运算器与存储器进行程序指令的顺序读取与执行,并将执行结果存入存储器中,最终运行结果通过存储器输出到输出设备中,整个过程是通过控制器来控制运行的^[1],整个运行流程如图1所示.以二进制逻辑和程序存储顺序为主的执行逻辑构成了冯·诺依曼体系结构,这一理论被广泛应用于现代计算机的制造.

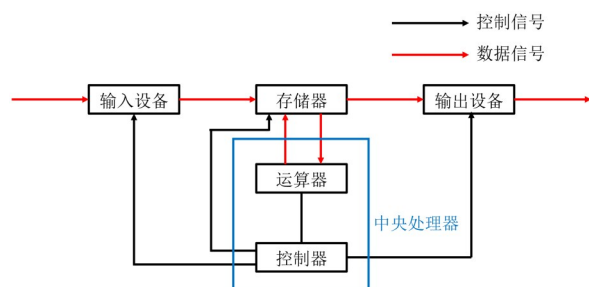


图1 冯·诺依曼体系结构

但是,经过几十年的发展,冯·诺依曼架构的计算机性能已经难以再有重大突破,提高其性能变得极其困难.面对当今智能化日益增强、追求极致能效比的发展趋势,该架构已经无法满足未来实际应用的需要.冯·诺依曼体系结构中的各组成部分独立工作,这从根本上决定了其性能难以提升,存储器和运算器之间的频繁传输所带来的性能代价严重,导致了“冯·诺依曼瓶颈”^[2]和现代计算机的效能问题.计算机智能化需要根据外界与需求变化自我演化提升,但由于冯·诺依曼架构的计算模式是程序预存储,采用此架构设计的计算机往往自适应性较差、智能化水平较低.人工智能也是模拟脑的一项主要应用,这种生物学简化的深度学习模型有它独特的优势,数学解释性强.虽然这种模型

在需要大量数据和计算的问题上表现优于人类,大数据和深度学习的发展更是将这种优势进一步扩大,但这种模型对一些只有人类才能解决的困难问题束手无策.

1.2 类脑计算与神经形态器件

图2为生物大脑神经细胞的结构,神经细胞由树突、轴突和胞体组成,树突负责接收神经递质传来的电信号,当神经元兴奋,电信号会从树突沿着轴突到达突触,接着突触释放神经递质,信号传向下一个神经细胞.人类大脑的结构极其复杂,其神经元数量级达到 10^{11} ,每个神经元中突触的数量级达到 10^4 .尽管人类大脑的工作频率仅为10 Hz左右,工作耗能仅为20 W左右,但在解决复杂问题时,它仍然表现出极高的效率^[3].人脑的部分神经元在不使用时可以关闭,这种功能极大地提高了工作的能效比,有利于人脑在极低能耗下实现高级智能.因此,人们从人脑的结构和工作方式中获得启发,希望模拟人脑建立高效的神经网络模型,从而提升计算机处理数据和解决问题的能力.

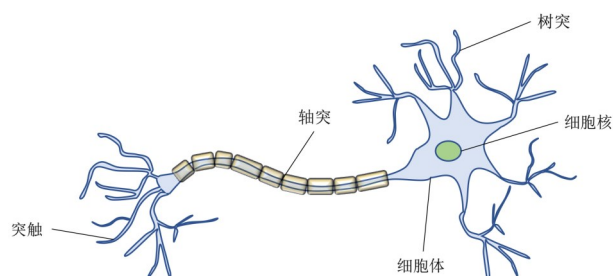


图2 生物大脑神经细胞的结构示意图

类脑计算就是基于人脑的结构和工作方式提出的.类脑计算系统的拟神经网络架构通过增加硬件实现高度并行运算,用极低的代价换取了传统计算机难以达到的高级智能,解决了“冯·诺依曼瓶颈”和现代计算机无法突破算力瓶颈的问题.

从模拟人类大脑神经网络的功能入手研究类脑计算,可以避免等待大脑的工作机制被完全揭示的漫长过程.如果想实现与人类大脑类似的功能,就要以模拟生物神经网络构建体系架构、模拟生物神经元功能和神经突触功能研制器件三个方面为研究方向^[4,5].针对模拟生物基本信息处理单元功能的目标所研发的器件

统称为“神经形态器件”,可分为突触器件、神经元器件两大类^[6].

突触可塑性是人类大脑实现多种高级智能的基础,因此现有的神经形态器件为实现复杂神经网络计算、降低功耗等提供了支撑,在研究更智能的低功耗类脑计算系统方面具有重大的价值.

1.3 神经形态器件的探索历程

在 20 世纪 40 年代, M-P 神经元模型和 Hebb 学习规

则被相继提出后,以神经网络为代表的类脑计算算法不断涌现. 1989 年, Mead^[7] 提出神经形态系统这一概念,即利用集成芯片模仿生物大脑神经系统功能此后人们越来越关注能实现类脑功能的神经形态器件的研究,并取得了一系列成果.

(1) 关于如何研制神经形态器件,人们首先尝试了使用传统 CMOS 技术进行各种类脑功能的仿真. 经过几十年的发展,国内外研究学者研制出了众多基于 CMOS 技术的神经形态器件,其发展历程如图 3 所示.

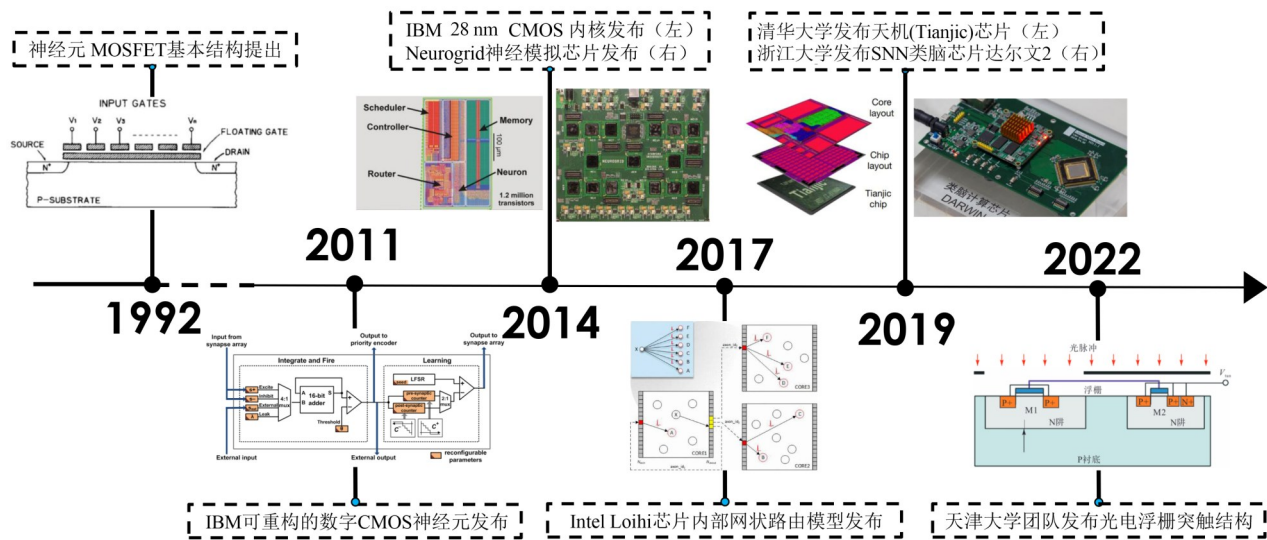


图 3 基于 CMOS 技术的神经形态器件发展历程^[8-15]

1992 年, Shibata 和 Ohmi^[8] 提出了一种神经元金属氧化物半导体场效应晶体管 (Metal Oxide Semiconductor Field Effect Transistor, MOSFET), 采用浮栅 MOS 器件实现了简单的生物神经元功能, 该器件由一个浮动栅极和多个输入栅极组成, 在栅极计算所有输入信号的加权和, 并根据这种加权和运算的结果控制晶体管的“开”和“关”. 2011 年, IBM^[9] 采用 45 nm SOI-CMOS 开发出一款高度可配置的芯片, 该芯片由 256 个神经元和 64 K 二进制突触组成, 具有基于脉冲时序依赖性的可塑性片上学习能力. 2014 年, IBM^[10] 发布了 TrueNorth 芯片, 该芯片由 54 亿个晶体管芯片构成, 其中 4 096 个神经突触内核通过芯片内网络互连, 该网络集成了 100 万个可编程脉冲神经元和 2.56 亿个可配置突触. 同年, 斯坦福大学的 Benjamin 等^[11] 公开了 Neurogrid 芯片, 该芯片模拟了除具有共享电子电路的体细胞以外的所有神经元, 实现了除轴突以外的所有电子电路, 同时在树突网络中互连神经阵列, 最大化了突触连接的数量, 最大限度提高了能源效率和吞吐量. 2018 年, Intel^[12] 研制出 Loihi 芯片, 该芯片具有一个众核网格, 包括 128 个神经形态内核、3 个嵌入式 x86 处理器内核和片外通信接口, 可将网格在四个平面方向上分层扩展到其他芯片,

异步片上网络 (Network on Chip, NoC) 以数据包化消息的形式传输内核之间的所有通信, 集成了硅中脉冲神经网络领域的分层连接、树突状隔室、突触延迟等功能, 还实现了可编程突触学习规则. 2019 年, 清华大学的研究团队^[13] 开发出天机 (Tianjic) 芯片, 该芯片由 156 个 FCores 组成, 包含大约 40 000 个神经元和 1 000 万个突触, 采用众核架构、可重构构建块和混合编码方案的流线型数据流, 不仅可以容纳基于计算机科学的机器学习算法, 还可以轻松实现类脑电路和多种编码方案. 同年 8 月, 浙江大学的研发团队研发完成脉冲神经网络 (Spiking Neural Networks, SNN) 类脑芯片达尔文 2, 该芯片采用 55 nm 标准 CMOS 工艺, 单芯片由 576 个内核组成, 每个内核支持 256 个神经元, 神经突触超过 1 000 万, 通过系统级扩展, 可构建千万级神经元类脑计算系统^[14]. 2022 年, 天津大学的研究团队^[15] 基于 TSMC 180 nm CMOS 工艺, 设计了一种由两个 PMOS 晶体管构建的光电浮栅突触结构, 使用光敏二极管感应光信号, 利用浮栅电荷调节突触权重, 实现了突触的光增强、电抑制和复位功能.

(2) 由于传统 CMOS 技术存在一些集成度、功耗等技术缺陷, 逐渐无法满足人们对器件性能和功耗的要

求,人们开始采用新型器件来模拟突触功能. 目前很多双端器件已经被提出并被用来设计神经形态器

件,其中最具有代表性的是忆阻器,其发展历程如图4所示.

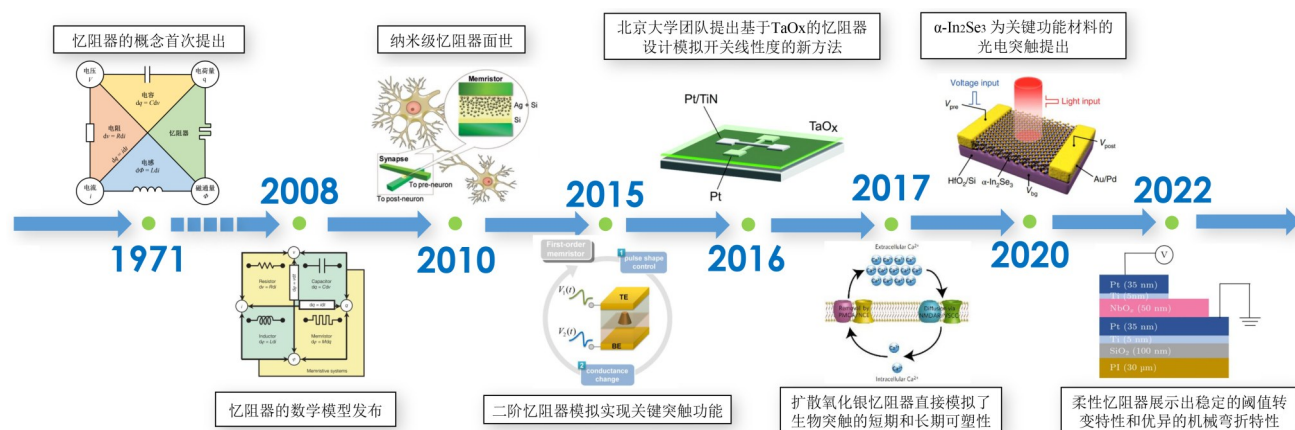


图4 基于忆阻器的神经形态器件发展历程^[16-25]

1971年,华裔科学家蔡少棠^[16]提出了忆阻器的概念. 2008年, Strukov等^[17]构建了忆阻器的数学模型,他们使用一个简单的分析示例来证明忆阻效应存在于纳米级系统中,其中固态电子和离子传输在外部偏置电压下耦合. 2010年,美国密歇根大学的研究团队^[18]通过实验展示了纳米级硅基忆阻器器件,并表明由互补金属氧化物半导体神经元和忆阻突触组成的混合系统可以支持重要的突触功能,例如脉冲时序依赖可塑性. 2015年, Kim等^[19]通过建模二阶忆阻器,并使用仅基于脉冲活动的简单脉冲形式实际实现关键突触功能,表明忆阻器提供的不仅仅是简单的电阻变化,其内部状态变量的动态演变允许基于氧化物的忆阻器表现出类似 Ca^{2+} 的运动,自然地原生编码定时信息并调节突触权重. 2016年,北京大学的团队^[20]提出并展示了一种在基于 TaO_x 的忆阻器中设计模拟开关线性度的新方法,即通过在 TiN/TaO_x 界面引入离子扩散限制层(Diffusion Limited Layer, DLL)来均匀化细丝生长/溶解速率,并且使用这些优化的设备成功实现了生物大脑中重要的突触学习规则,例如脉冲时序依赖可塑性. 2017年, Wang等^[21]开发了扩散氧化银忆阻器,其刺激期间和之后的时间响应类似于突触 Ca^{2+} 运动,直接模拟了生物突触的短时程和长时程可塑性. 2020年, Lv等人^[22]讨论了量子点在忆阻器中用于改进和优化记忆技术的有利特征; Ji等人^[23]基于忆阻神经形态系统构建人工智能感知系统,该系统表现出小尺寸、低功耗、3D堆叠能力和高密度的突出特点. 同年,北京大学的黄如院士-杨玉超教授课题组^[24]构建了基于 $\alpha\text{-In}_2\text{Se}_3$ 的多功能光电忆阻器件,利用其光电耦合的动力学特性,实现了多模态融合、多时间尺度的储备池计算. 2022年,朱佳雪等人^[25]制备的柔性忆阻器展示了稳定的阈值转变特性和优异的机械弯折特性,其弯

折半径可达1.5 mm,弯折次数可达 10^4 次. 基于此器件构建的神经元电路实现了神经元的关键积分放电特性,且其频率与输入电压关系具有整流线性单元相似性,可实现基于转换法的脉冲神经网络中神经元的非线性处理功能.

相变存储器(Phase Change Memory, PCM)和磁隧道结(Magnetic Tunnel Junction, MTJ)也是被广泛应用于神经形态器件研发的双端器件,其发展历程如图5所示.

1968年, Ovshinsky^[26]在各种类型的无序半导体材料中观察到了由电场影响的高电阻和导电状态之间的快速和可逆跃迁,提出了相变存储器的概念. 1975年, Julliere^[27]通过 $\text{Co}/\text{Ge}/\text{Fe}$ 磁性隧道结发现了隧穿磁电阻(Tunnel Magneto Resistance, TMR)效应,利用自旋电子器件中磁矩进动产生的TMR可实现仿生突触和神经元特性. 2011年, Suri等人^[28]介绍了可用于模拟大规模神经系统的PCM的通用行为模型,并演示了在2层脉冲神经网络中使用PCM突触从真实世界数据中提取复杂视觉模式. 2012年, Kuzum等人^[29]报告了一种基于GST($\text{Ge}_2\text{Sb}_2\text{Te}_3$)相变材料的新型纳米级电子突触,模拟了突触短时程可塑性到长时程可塑性的过渡. 2016年, IBM苏黎世研究中心^[30]基于硫族化物的相变材料创建人工神经元,其中膜电位由纳米级相变器件的相构型表示,通过利用可逆无定形到晶体相变的物理学,表明突触后电位的时间积分可以在纳秒时间尺度上实现. 同年, Srinivasan等人^[31]提出了一种由磁隧道结和重金属组成的异质结构作为随机二元突触,通过MTJ电导状态的随机切换来实现了突触可塑性; Sengupta等人^[32]展示了皮层中锥体神经元的概率脉冲性质与存在热噪声的磁隧道结的随机切换行为的映射,基于这种概率神经元的神经形态系统在存

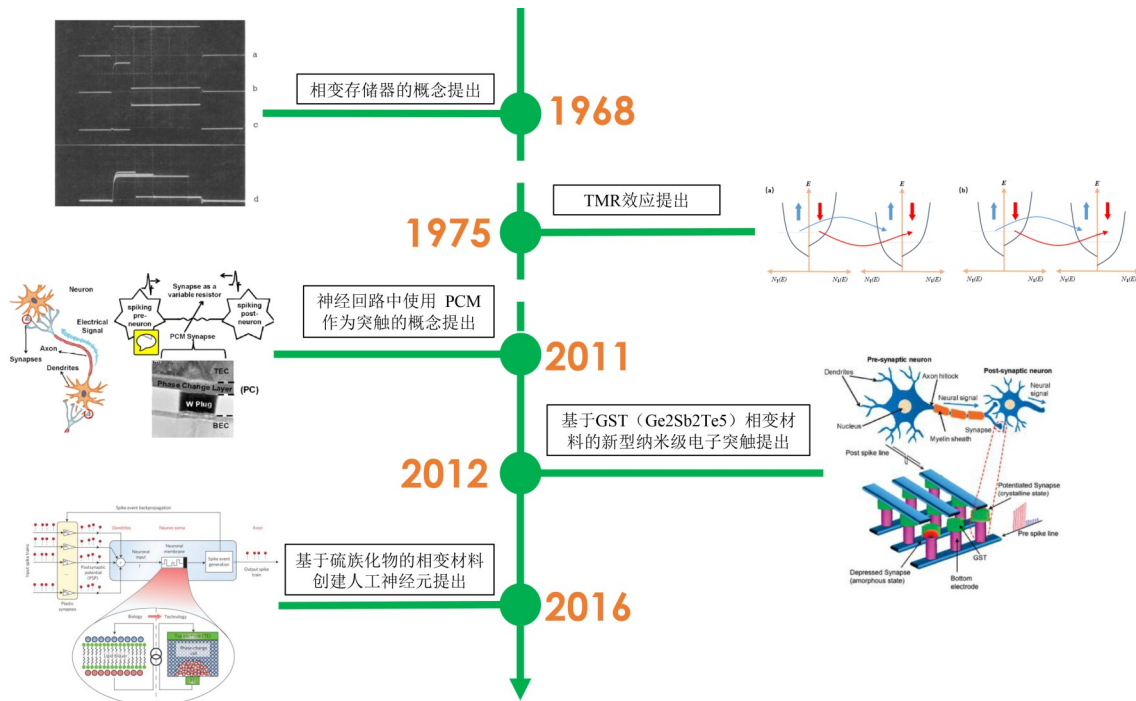


图5 基于相变存储器和磁隧道的神经形态器件发展历程^[26-32]

在横向抑制和体内平衡的情况下执行模式识别任务的效率。

(3) 进入 21 世纪, 比两端器件性能更好的三

端/多端晶体管被应用于神经形态器件研发, 其中最具代表性的是铁电晶体管, 其发展历程如图 6 所示。

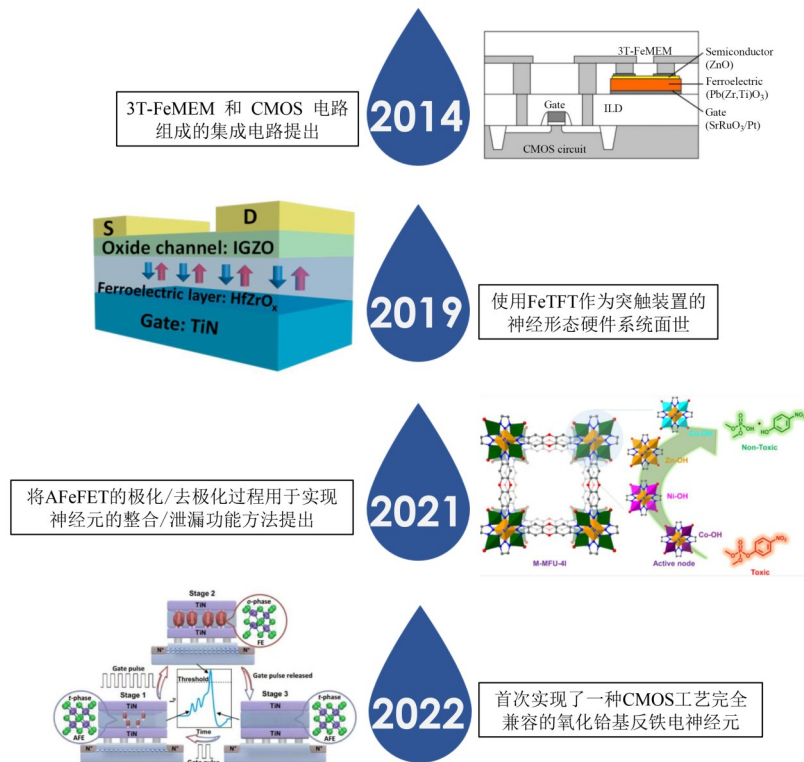


图6 基于铁电晶体管的神经形态器件发展历程^[33-37]

2014年, Kaneko 等人^[33]使用多个三端铁电忆阻器(3T-FeMEM)作为突触, 构建了脉冲神经网络模型对多阴影灰度图像进行模式识别, 使用3T-FeMEM的模拟和非易失电导变化作为突触权重, 在脉冲时序依赖性塑性学习规则之后学习基质模式. 2019年, Kim 等人^[34]演示了具有纳米级铁电材料和氧化物半导体的铁电薄膜晶体管(FeTFT)中的模拟电导调制行为, 通过精确控制纳米级铁电层的极化变化诱导电导调制, 证明了FeTFT的线性增强和抑制特性, 提供了一种实现使用FeTFT作为突触装置的神经形态硬件系统的方法. 同年, Jang 等人^[35]展示了一个超薄人工突触(总厚度约500 nm), 它具有独立的铁电有机神经形态晶体管(Ferroelectric Organic Neuromorphic Transistors, FONT), 可以在没有基板或封装层的情况下独立存在, FONT表现出出色而可靠的突触功能, 可以通过不同的电刺激和突触前和突触后脉冲之间的相对时间或时间顺序进行调节, 即

使在折叠条件下, FONT也可保持超过6 000个输入脉冲的突触可塑性. 2021年, Mian 等人^[36]基于AFeFET设计神经元, 将AFeFET的极化/去极化过程用于实现神经元的整合/泄漏功能, 这种神经元具有CMOS兼容、可调发射频率和超低功耗的特点. 2022年, 国防科技大学电子科学学院徐晖教授课题组与复旦大学芯片与系统前沿技术研究院刘琦教授课题组合作, 实现了一种CMOS工艺完全兼容的氧化铪基反铁电神经元^[37]. 研究表明, 所设计的反铁电神经元避免了大尺寸电容和复位电路的使用, 大幅降低了神经元电路硬件开销, 有效提升了神经形态计算芯片集成度, 其构建的双层神经网络具有高达96.8%的手写体识别率.

双电层/电化学晶体管也是一种被广泛应用于神经形态器件制备的三端/多端晶体管, 根据其使用的无机半导体材料不同可实现多种类脑功能, 其发展历程如图7所示.

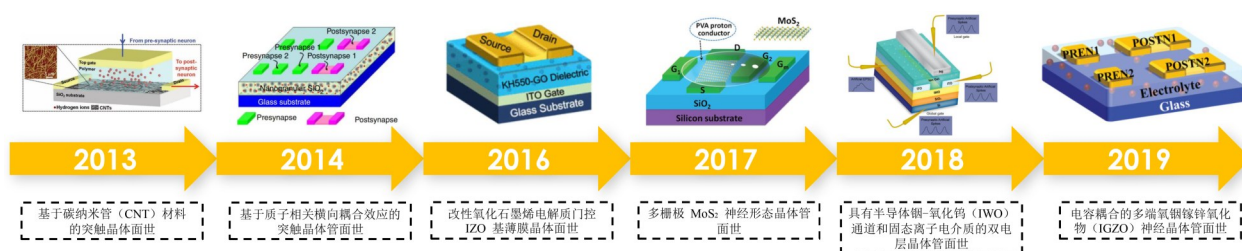


图7 基于双电层/电化学晶体管的神经形态器件发展历程^[38-44]

2013年, Kim 等人^[38]使用碳纳米管(Carbon Nano Tubes, CNT)制作了突触晶体管, 通过电化学电池中CNT和氢离子之间的相互作用模仿生物突触的动态逻辑, 学习和记忆功能. 2014年, Zhu 等人^[39]提出了一种基于质子相关横向耦合效应的突触晶体管, 他们将基于钽锌氧化物(IZO)的质子/电子混合晶体管在室温下自组装在磷(P)掺杂的纳米颗粒SiO₂质子导电薄膜上, 模拟了短期可塑性行为, 包括配对脉冲促进、动态滤波和时空相关信号处理. 2016年, Yang 等人^[40]以3-三乙氧基硅基丙胺改性氧化石墨烯(KH550-GO)电解质作为栅极电介质制备了基于氧化物的双电层(Equivalent Double-Layer, EDL)薄膜晶体管, 模拟了长时程突触可塑性基本学习规则之一的脉冲时序依赖性可塑性(Spike Timing Dependent Plasticity, STDP), 成功再现了包括长期记忆、突触时间整合和动态过滤器在内的突触功能. 2017年, Jiang 等人^[41]使用聚乙烯醇作为横向耦合的质子传导电解质制备了一种2D MoS₂突触/神经元晶体管, 成功模拟了兴奋性突触后电流、成对脉冲促进和用于生物突触信息传输的动态滤波器等生物突触基本功能. 同年, Feng 等人^[42]使用具有非常高饱和场效应迁移率的印刷双栅极碳纳米管薄膜晶体管制作了双电

层突触, 实现了生物突触配对脉冲促进(Paired-Pulse Facilitation, PPF)和信号滤波特性的仿生. 2018年, John 等人^[43]以双门控电双层神经质构型实现了具有半导体钨-氧化物(IWO)通道和固态离子电介质的双电层晶体管, 成功模拟了生物突触的高通滤波特性. 2019年, He 等人^[44]提出了一种电容耦合的多端钽镓锌氧化物(IGZO)神经晶体管用于时空信息处理, 模拟了不同时空输入序列的树突动态判别, 可以作为基本皮层计算的时空信息处理隔室, 通过构建基于这种多端神经晶体管的人工神经网络来模拟人类大脑的声音定位功能.

在神经形态器件的研究领域, 为实现模拟更多更复杂的生物神经网络的功能和行为, 进一步研发高精度神经形态器件是当前的研究重点. 在类脑计算系统中, 在器件层面实现越多越复杂的仿生功能和行为, 越能提高整个系统表现出的智能化水平. 本文从神经形态器件的三大类型, 即忆阻器、MOS管和磁隧道结入手, 并以当前最热门的忆阻器为主介绍具有代表性的神经形态器件, 剖析相关产业需求, 总结神经形态器件方向的发展规律, 并为此方向发展提供技术参考.

2 忆阻器的特性分析与应用

2.1 忆阻器的发现及其性质

1971年,华裔科学家蔡少棠教授首次定义了电荷 q 与磁通量 ϕ 之间的关系,证明了电压、电流、电荷量与磁通量这四个常用的电路变量彼此之间存在完备关系,其关系如图8所示,进而提出第四种电路元件,并称之为忆阻器^[16].然而由于蔡少棠教授仅仅提出了猜想与理论,缺乏实物模型,因此在之后的30余年里,研究人员没有足够地关注与重视关于忆阻器的研究.

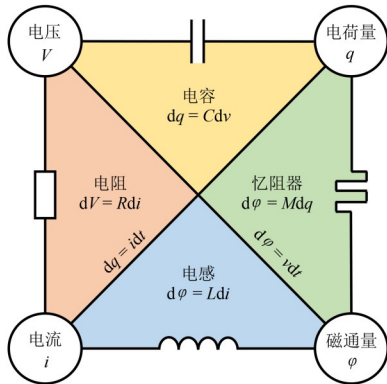


图8 电荷量、电流、电压和磁通量之间的关系

忆阻器是一种结构相对简单的新型的二端器件,可以看作一个简单的三层堆叠式结构,由两层金属层和中间的绝缘介质层构成,呈经典的“金属-绝缘体-金属”结构,其电阻值改变的诱因也有许多,如外部激励的类型、强度和持续时间等.忆阻器除了具有低电压、低功耗、速度快、易于集成等优点以外,还表现出了自然的记忆特性、极快地切换转换机制、阻值的连续可变性等,这些都使之具有模拟神经元与突触的天然条件,在许多领域都具有很高的应用潜力.作为第四种基本电路元件,忆阻器和其他三种电路配件有着无法隔断的关系^[45],但同时忆阻器器件也具有与电阻、电容和电感都不同的电气性能,且无法被这三者或三者的组合所取代,是模拟突触的最佳选择,具有断电不丢失信息、阻值连续且可调节、功耗低等优良特性.

2.2 数字型忆阻器与模拟型忆阻器

忆阻器的不同位置由不同的材料组成.其中,金属层分别作为顶电极和底电极,绝缘介质层一般由过渡金属的氧化物构成,也被叫作氧化物忆阻器,可以分为数字型忆阻器和模拟型忆阻器^[46].

数字型忆阻器的阻值变化是离散的,随着电压的变化其阻值呈现出可互相转化的高阻与低阻态.如果把这两种阻态看为“0”和“1”,那么数字型忆阻器适合应用于非易失存储领域.

与数字忆阻器不同,模拟忆阻器的离子迁移/扩散

过程与生物突触中的 Ca^{2+} 离子动力学过程高度匹配,因此它主要被应用于人工突触模拟.模拟忆阻器的电阻模式可以随着电压扫描而连续变化,器件的电阻可能随施加的电压而变化,并且可以记住变化的状态.自从2008年HP实验室在Nature杂志上首次报道了其团队制备的基于 TiO_2 材料的固体忆阻器^[47],几个研究小组已经发现了各种氧化物材料中的膜阻行为,例如 $In-GaZnO(IGZO)$, WO_x , $SrTiO_3$ 等.

2.3 忆阻器的基本模型

二端器件忆阻器(其基本结构如图9所示)是一个简单的三层堆叠结构^[48],由两层金属层和中间的绝缘介质层组成,不同材料的组合决定了忆阻器的特性.其中,金属层作为忆阻器的顶部电极和底部电极,而绝缘介质层一般由阻抗很高的过渡金属的氧化物构成.忆阻器件的特点在于其阻值不是固定的,而是和电流、磁通量等因素密切相关,这种特性被称为忆阻.

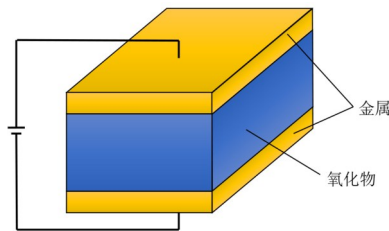


图9 二端器件忆阻器的结构

2008年,Strukov等人^[17]提出了忆阻器的基本数学模型.如果将流经忆阻器的电荷 q 看作自变量,将磁通量 ϕ 看作因变量,将得到相对应的电荷与磁通间的关系式如下:

$$\phi = f(q) \quad (1)$$

通过式(1)两边对时间 t 求导,得

$$\frac{d\phi}{dt} = \frac{df(q)}{dq} \cdot \frac{dq}{dt} \quad (2)$$

进一步,基于基本公式 $v=d\phi/dt$ 以及电流 $i=dq/dt$,可推得

$$\frac{v}{i} = \frac{df(q)}{dq} = \frac{d\phi}{dq} \quad (3)$$

如果将忆阻值 M 定义为 $M(q)=d\phi/dq$,则上式可改写为

$$M(q) = \frac{v}{i} \quad (4)$$

根据式(4)可知忆阻器满足经典欧姆定律,表现出电阻的特性,但与普通电阻不同的是,忆阻器的阻值取决于流经它的电荷量.因此,通过测量忆阻器的阻值,可以确定流经它的电荷量,说明忆阻器具有记忆电荷的特性.因此,这种忆阻器被称为电荷控制忆阻器模型.

后来,Strukov等人^[17]又提出的忆阻阈值开关模型证实了忆阻器具有阈值效应.这种效应常见于纳米级别的忆阻器件中,表现在当激励电压超过阈值电压时,忆阻器状态值才会发生改变,是忆阻器件的一个十分重要的特性,表达式为

$$v(t) = \left(R_{\text{ON}} \cdot \frac{w(t)}{2} + R_{\text{OFF}} \cdot \left(1 - \frac{w(t)}{2} \right) \right) \cdot i(t) \quad (5)$$

其中, D 模型总长度; w 为掺杂厚度(w 会根据电荷的变化而变化); R_{ON} 为忆阻器件全部为低阻区时的阻值; R_{OFF} 为忆阻器件全部为高阻区时的阻值,该模型区分正负极.

将忆阻器的绝缘介质层(中间层)分为低阻区和高阻区,建立边界漂移模型,其模型如图10所示,低阻区为高浓度掺杂区域,具有低电阻 R_{ON} ,高阻区为低浓度掺杂区域,具有高电阻 R_{OFF} .因此,忆阻器件的阻值随着低阻区和高阻区边界离子的移动而变化,忆阻器的等效电路如图11所示.

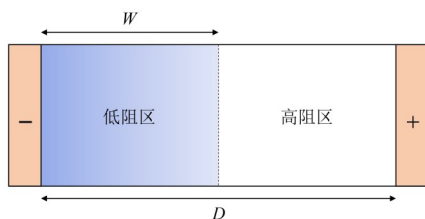


图10 边界漂移模型示意图

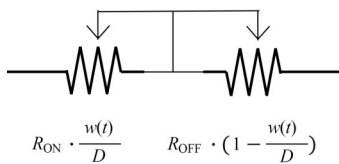


图11 忆阻器的等效电路

2.3.1 HP忆阻器模型

作为第四种基本电路元件的忆阻器首次亮相是在2008年,惠普公司的科研人员在《自然》杂志上第一次展示了它,其表现出的忆阻特性使之成为天然的非易失性存储器.

经典的惠普忆阻器的物理模型是四层堆叠式结构,其中最外面的两层是两个独立的铂(Pt)片电极层,中间的由两层二氧化钛(TiO_2)构成,其中一层掺杂了氧空位而另一层未掺杂.掺杂了氧空位的那一层由于缺失了一部分氧原子,因此呈现出低阻抗即半导体特性,另一层由于没有氧空位故呈现出高阻抗特性^[47](选用 TiO_2 材料是由于威廉的研究团队发现 TiO_2 被夹在电极中间会产生奇特的电子特性,整块的二氧化钛被分为了掺杂未掺杂两部分,表现出各自不同的特性).整个器件的阻值等于掺杂与未掺杂层阻值之和,当外界施

加偏压时,掺杂层中的氧空位会往氧含量正常的一侧漂移,从而引起层间边界的移动,对于整块材料,可以说是改变了忆阻器当前的阻值^[49].

惠普忆阻器阻值 M 表示为

$$M(t) = R_L \cdot x(t) + R_H \cdot (1 - x(t)) \quad (6)$$

$$x(t) = \frac{w(t)}{D} \in [0, 1] \quad (7)$$

其中, R_L 和 R_H 是忆阻器忆阻值的极小和极大值, x 为状态变量(在当前已经实现的神经形态器件中,忆阻器内部的导电细丝尺度、温度、相变过程、电荷等提供了物理层面的状态变量),这里的状态变量取掺杂层厚度 w 与总厚度 D 的比值.

然而,真实情况下的忆阻器由于其纳米级尺寸的特点会因为施加的一个极其微小的电压便产生极大的电场.这也使忆阻器的各特性表现出较为明显的非线性特性,因此在公式的右边乘了一个非线性窗函数,来更好地模拟内部离子的非线性运动,可得:

$$\frac{dx}{dt} = ki(t)f(x), \quad k = \frac{\mu_v R_L}{D^2} \quad (8)$$

其中, μ_v 表示平均的离子迁移率; i 表示流经忆阻器的电流; k 为状态变量的变化速度与流经电流之间的比例因子; $f(x)$ 表示窗函数.

2.3.2 阈值开关模型

在相关领域中,阈值开关模型有很多种类,这里选取的是基于量子隧穿(像电子等微观粒子穿入或穿越位势垒的量子行为)的研究而提出的模型^[50],等效模型如图12所示.

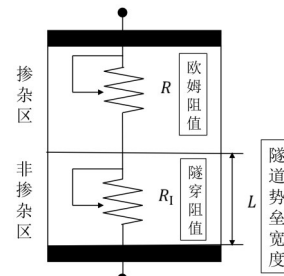


图12 忆阻器阈值开关模型

忆阻值 R_t 的表达式为

$$R_t = f_0 \cdot \frac{e^{2L}}{L} \quad (9)$$

其中, R 表示电阻阻值; R_t 表示忆阻值; L 表示隧道势垒的宽度,同时也是状态变量; f_0 是模型拟合参数. L 的表达式为

$$L = L_0 \left(1 - \frac{\kappa}{r} \right) \quad (10)$$

其中, L_0 是隧道势垒宽度的最大值; κ 的大小决定了 L 的取值范围,故其也可以叫作拟合参数; r 与电压有关.

2.3.3 VTEAM 忆阻器模型

TEAM (ThrEshold Adaptive Memristor) 是一种具有非易失性、内部状态依赖电荷的非线性以及电流越高改变幅度越大的忆阻器模型。但它是基于电流阈值的模型,结合忆阻器的发展历程与研究现状,大多数使用忆阻器的逻辑电路和神经形态网络结构中,更多的是需要由电压阈值调控的忆阻器模型 (Voltage ThrEshold Adaptive Memristor, VTEAM)^[51],故电压阈值忆阻器模型应运而生,并由于其较为广泛的通用性和优秀的忆阻特性而被更广泛地使用在各类电路和网络中。值得一提的是,以上选取的 TEAM 和 VTEAM 模型是基于西蒙斯忆阻器模型简化而来的^[52]。西蒙斯模型虽然很精确,但其表达式比较复杂,且只适用于特定的忆阻器,计算的效率不高,不利于工作的理论分析和模拟。故基于该模型设计了相对简化的 TEAM 模型,该模型不仅能有效模拟西蒙斯的特性,还具有更加简化的方程,在运算中更加方便与高效。

2.3.4 自旋忆阻器模型

1972年3月,Wang等人^[53]在美国《电子器件快报》的文章引起了广泛的关注,他们设计的模型结构如图13所示。

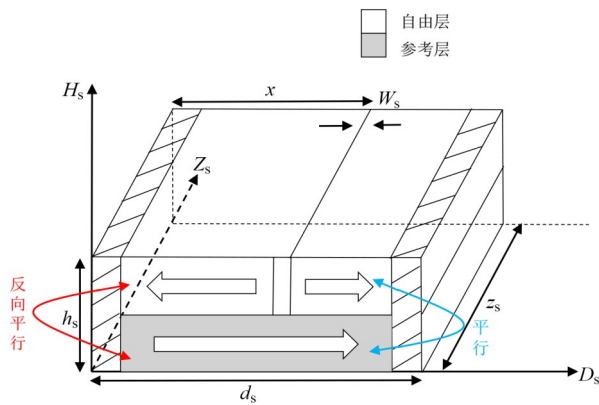


图13 自旋忆阻器模型^[53]

文章中阐述了三种忆阻器实例,均基于电子自旋。这种新型电子自旋忆阻器与惠普忆阻器件相比,可以在不同的速率下完成对电阻值的转换,且具有更强的抗干扰能力,独特的阈值电流密度使其能有效地避免噪声的产生,且器件的制作成本低,更容易制造。具有多种物理结构,有基于磁畴壁运动的、基于随机翻转的、基于自旋矩纳米振荡器的以及基于新物理现象的自旋神经形态器件。自旋忆阻器的阻值可以表示为^[54]

$$M(x) = r_H \cdot x + r_L \cdot (d_s - x) \quad (11)$$

其中, r_H 和 r_L 表示忆阻器处于低阻态和高阻态时每单位长度上的阻值; d_s 表示自旋忆阻器的长度; x 表示畴壁移动的距离。

2.4 忆阻模型比较与特性分析

2.4.1 忆阻模型对比

研究发现主要有三种代表性的忆阻模型。

(1) 导电前端移动模型: 基于导电前端的移动实现阻值的调控。上面讨论的 HP 忆阻器模型就属于这一类,其结构为四层堆叠式结构。其中最外面的两层是两个独立的铂 (Pt) 片电极层; 中间的由两层二氧化钛 (TiO_2) 构成, 其中一层掺杂了氧空位, 一层未掺杂, 前者由于缺失了部分氧原子, 故呈现出低阻抗即半导体特性; 后者由于没有氧空位故呈现出高阻抗特性, 通过压缩 undoped 区域的宽度从而改变器件的电阻。

(2) 导电细丝数量调控模型。除了双层结构的忆阻器之外, 单层氧化物材料也能实现忆阻行为。密歇根大学 Lu 实验组基于 WO_3 薄膜构建了忆阻器^[55], 这是一种除了具有基本记忆特性外, 还具有特殊遗忘效应的记忆电阻器, 同时也能实现电阻的连续调节。忆阻器的机制是通过增加或减少膜内导电丝的数量来实现对阻值的调节。

(3) 肖特基势垒 (具有整流特性的金属-半导体界面) 调制模型: 这里直接改变调节的对象为调制电极与绝缘层之间势垒, 而不是调节本身电阻, 也能实现忆阻行为。模型如图 14 所示, 图片的左侧为肖特基接触, 图片右侧为欧姆接触。

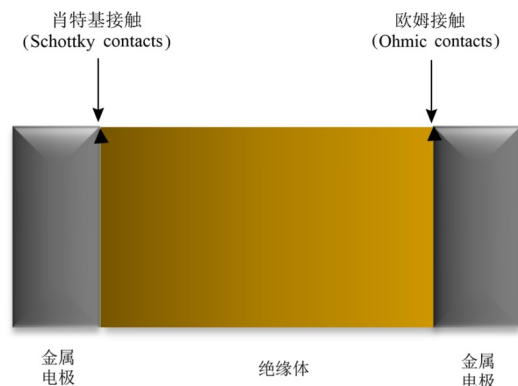


图14 肖特基势垒忆阻模型示意图

忆阻器的导电前端移动模型具有普遍适用性, 无论基于什么材料体系或漂移离子均能稳定地实现忆阻行为, 其电阻的调控基于外电场的驱动实现。忆阻器的导电细丝数量调控模型基于单层氧化物实现忆阻行为, 因为最先提出的理想模型导电前端移动模型无法解释基于单层氧化物如 WO_3 的忆阻行为, 故科学家们经过研究发现, 单层氧化物实现忆阻行为是基于调控氧化物内部氧空位导电细丝的数量实现的。肖特基势垒调制模型基于调制上下电极层与绝缘层之间势垒, 在绝缘层中施加外电压, 调整势垒层氧空位浓度从而改变势垒层高度, 进而改变器件阻值。

2.4.2 忆阻材料对比

近几年里,随着对忆阻器的研究越来越深入,越来越多的材料也被发现具有忆阻特性,从而被作为介质层实现阻态的转变.主要的忆阻器材料如下.

(1) 固体电解质,并掺杂有活性金属,如 Ag, Cu 等.

(2) 多元金属氧化物,如, SrTiO_3 , SrZrO_3 等三元金属氧化物及 $\text{LaCa}_{1-x}\text{MnO}_3$ 等四元金属氧化物.

(3) 二元金属氧化物,在忆阻器材料选择中占大多数,已被报道发现的具有忆阻特性的材料有很多,包括 SiO_x , TiO_x , AlO_x , ZnO_x 等.

(4) 低维纳米材料,如石墨烯、纳米线、量子点^[56]以及 BP 等新材料.

(5) 有机物材料.近些年来有机物凭借其工艺较简单、成本较低和良好的延展性等优点进入研究人员的视野,目前已报道的有机物材料主要有聚苯乙烯等聚合物材料,鸡蛋清、蚕丝蛋白等天然材料.

(6) 氮化物材料.近些年来氮化物材料在半导体的研究中应用较为广泛,却较少地被应用在忆阻器层级之中.这并不是因为不适合,恰恰相反,氮化物材料实现的阻变功能在某些方面更为优秀.目前研究人员正以 AlN 为主对氮化物材料进行更深入的探索,并在 Si_3N_4 和 Cu_xN 等材料中发现优异的阻变性能.忆阻器的应用材料有很多,不同的材料也对应着不同的研究方向.基于资料与实验材料间的优劣,具体如表 1 所示.

表 1 忆阻器应用材料对比

忆阻器材料	优势	劣势
固体电解质	工作电流低,开关速度快,操作电压低,数据存储时间长	与 CMOS 工艺较难兼容
多元金属氧化物	非易失,高速,低功耗	制备工艺的复杂,元素比例控制难度大,无法与 CMOS 工艺兼容
二元金属氧化物	低廉的价格,简单的制备工艺,与 CMOS 工艺兼容	阈值不高
低维纳米材料	独特的结构以及在尺寸、柔性、透明、超薄、导电 ^[56] 或绝缘特性的优势	成本较高
有机物材料	低成本、易于解决的工艺,灵活的延展性,创造性的分子设计,合成策略可调的电子性能,与 CMOS 工艺兼容	发展处于探索阶段
氮化物材料	较宽的带隙,高电阻率,高导热率,大的开/关比,是良好的热稳定性器件	成本较高,探索较少

2.4.3 忆阻器忆阻机制对比

新型非易失性神经形态器件的主要发展方向有可变电阻式存储器(Resistive Random Access Memory, ReRAM)、相变存储器(Phase Change Memory, PCM)、铁电存储器(Ferroelectric Random Access Memory, FeRAM)、磁性随机存储器(Magnetic Random Access Memory, MRAM)以及离子栅控晶体管等,然而不同的忆阻机制的忆阻性能也有很大差别,总结如表 2 所示,常见性能对比如图 15 所示.

阻抗存储器结构简单,但目前阻变参数呈离散型分布,需要研究人员对其机理有更深入的理解;相变存储器的阻变过程基于材料晶态与非晶态的转变,故其功耗与工作时间之间难以兼顾;铁电存储器擦写速度快,写入电压较低,但其制作工艺较为复杂,人工成本较

表 2 不同忆阻机制存储器常见性能对比

存储器类型	ReRAM	PCM	FeRAM	MRAM
CMOS 兼容性	好	好	好	较差
写入电压/V	2~3	≈1	<2	10~12
写入时间/ns	≈100	50~500	30	30
读取时间/ns	≈10	10~50	30	30
多量化可行性	有	有	无	无
撰写次数(循环)	$10^5 \sim 10^6$	10^9	10^6	10^{15}
Cell 面积/ F^2	≈4	4~16	15~30	~20

高;磁存储器由于其磁性,相邻存储单元间干扰较大,当器件密度较高时,磁场重叠现象严重^[57].

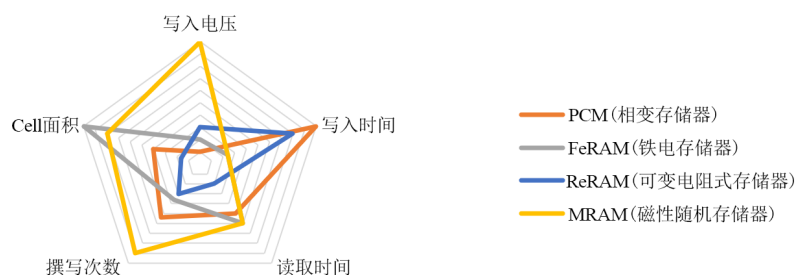


图 15 不同忆阻机制存储器常见性能对比

2.5 忆阻器的应用场景

2.5.1 忆阻器在传统行业中的应用

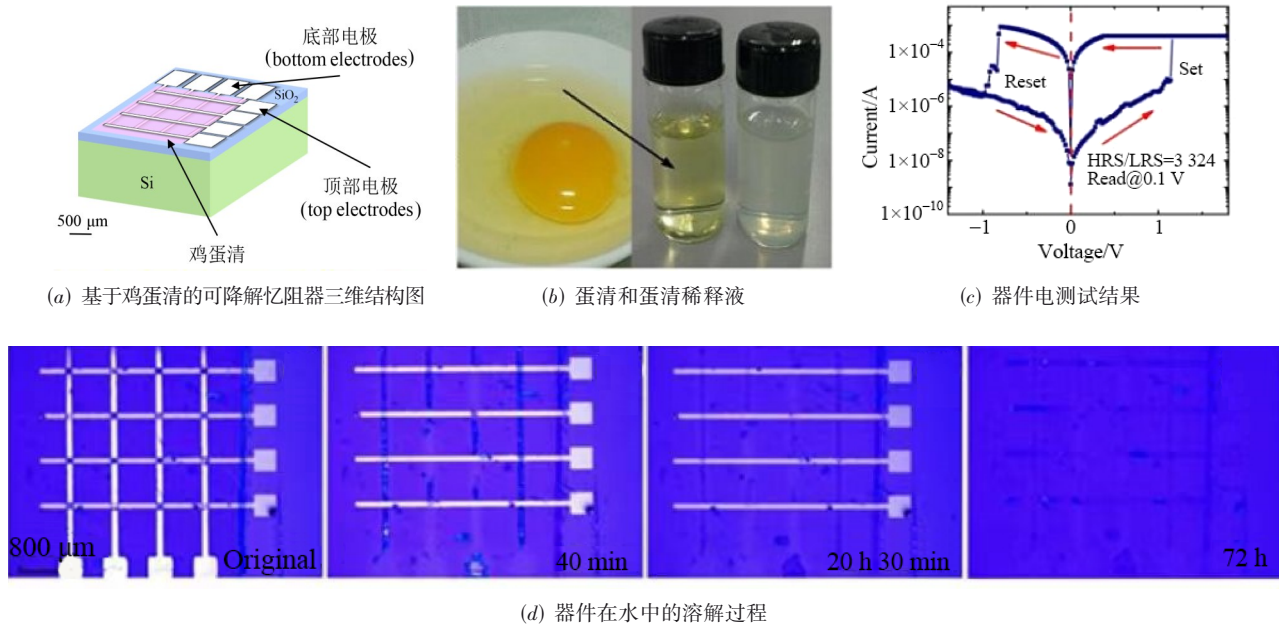
忆阻器在传统场景中有许多实际化的应用。

忆阻器由于其断电非易失性从而可以制作成可变电阻式存储器 (Resistive Random Access Memory, RRAM) 应用于手机电脑等电子设备, 由忆阻器制成的芯片相比闪存保存信息速度更快, 耗电以及占用空间更少。

忆阻器对于模拟式计算机的开发具有辅助作用,

可使计算机对收集到的数据有更深层次的理解, 从而使计算机能基于特定程序模拟大脑的决策功能, 甚至是学习功能, 该技术可被用于人脸识别等领域。

忆阻器还可以作为微型电子器件实现植入式医疗。浙江大学和英国剑桥大学研究团队制造出一种忆阻器, 其原料为鸡蛋清和可降解金属, 如图 16 所示。该忆阻器不仅可读写数百次, 还可以在 3 天之内被人体体液溶解吸收, 从而实现体征监测、疾病预警、伤口愈合跟踪等植入式医学功能, 干燥情况下也可储存信息 3 个月。



(d) 器件在水中的溶解过程

图 16 基于鸡蛋清的可降解忆阻器

2.5.2 忆阻器在类脑计算中的应用

忆阻器由于外部结构与电阻类似, 因此可以花较低成本实现较高密度的基层。但作为一种非线性的无源二端器件, 它具有与电阻不一样的特性, 在电流或电压的激励下它的阻值会发生变化。忆阻器能够对神经突触功能进行模拟, 主要原因是它在结构和功能两方面与神经突触相似, 结构上忆阻器包含上电极层、绝缘层和下电极层, 分别对应了神经突触结构的突触前膜、突触间隙和突触后膜。

目前, 大量的实验表明, 忆阻器在模拟神经网络中的突触方面, 具有很大的前景优势^[58-65]。忆阻器的离子移动过程可以模拟神经递质在突触神经间的传递过程, 并且作为一种纳米级的无源器件, 它不仅模拟不断变化的突触权重, 而且还可以做到存算一体、高度集成, 可以提高卷积神经网络的并行性, 实现更加多样化的功能。

2.5.2.1 多层感知机

多层感知机模型 (又被称为人工神经网络), 是单

层感知机的应用扩展。由于单层感知机的局限性在于只能解决线性问题, 因此为解决非线性问题, 例如基本的异或问题, 会在输入与输出层之间加入隐藏层。以简单的三层网络结构为例, 如图 17 所示, 网络由输入层、一层隐藏层与输出层组成, 相邻层之间是全连接的, 输入向量 X 由输入层进入, 与隐藏层之间的权重矩阵 W 进行矩阵计算, 其结果作为隐藏层的输出向量 H , 再与输出层之间的权重矩阵做运算, 得到最后结果向量 Y 。权重更新主要采用的是反向传播算法, 通过对输出值 Y 与目标值 Z 的差进行梯度下降运算, 来更新各层之间的误差, 使其收敛到最小, 从而达到训练模型的目的。

在训练过程中, 比较消耗资源的是输入向量与权重矩阵的乘法运算, 因此可以利用忆阻器交叉阵列来存储突触权重矩阵, 每个交叉点处的忆阻器的电导值被用来表示一个突触的权重, 从而实现并行一步向量矩阵的乘法运算, 这可以大大减少硬件化神经网络的能耗^[66]。

2017 年, 清华大学吴华强教授团队^[67]在 *Nature*

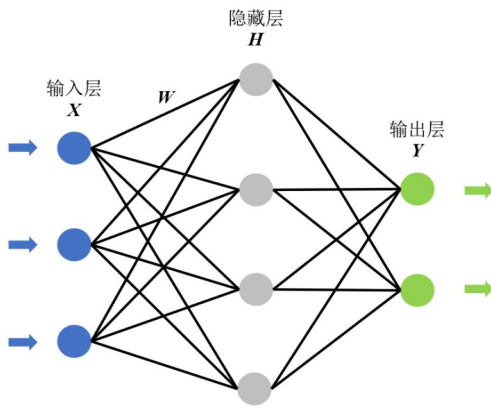


图 17 多层感知机结构示意图

Communication 期刊发表文章,通过利用大小为 1 K 的 1T1R 器件单元来模拟突触,实现了双向的器件电导调制与 3 层全连接多层感知机,并利用感知机学习识别了耶鲁人脸数据库中的灰度人脸图像,在测试集中对 9 000 个添加噪声的图片识别率达到了 88.08%。其感知机耗能仅为 Intel Xeon Phi 处理器的千分之一,仅为片外计算高能的 20 分之一。

2018 年,美国加利福尼亚大学圣芭芭拉分校 Strukov 教授团队^[68]设计制备了 20×20 的金属氧化物交叉阵列。其底电极为三角形,不仅可以更高覆盖下电极,还可以降低顶电极的接触电阻。其中间功能层采用 TiO_{2-x} 与 Al₂O₃ 叠层,Al₂O₃ 叠层作为阻挡层使基本的 I-V 变得非线性,有利于忆阻器 OT1R 阵列中漏电流问题的减轻。在此基础上,该团队进一步设计了更加复杂的多层感知机,可以达到 97% 以上的离线学习分类准确率。

2.5.2.2 卷积神经网络

卷积神经网络(Convolutional Neural Network, CNN)对图像的处理拓宽了神经网络的使用场景。CNN 通过卷积、池化等操作降低了参数的复杂度,并保留了一定的图像特征,图像进行旋转、镜像或变换位置的处理后依然可以有效地进行识别。如图 18 所示,典型的 CNN 由卷积层、池化层和全连接层三部分组成,它们的作用分别是:卷积层负责通过卷积操作提取图中的局部特征;池化层通过取最大或平均来显著降低参数的数量级;全连接层负责实现分类与输出结果。相比多层感知机,卷积神经网络的应用更广泛,泛化能力更强大。

利用忆阻器构建卷积神经网络有两个作用:卷积操作部分可以利用忆阻器的交叉阵列存储卷积核值来达到一步完成卷积计算的目的,提高计算效率;全连接操作部分可以如前文对多层感知机的实现一样实现并行计算^[69,70]。

法国的 Garbin 等人^[71]通过展示多个并联二值 HfO₂ 忆阻器实现一个突触功能来搭建脉冲 CNN,证实了基于氧化物的忆阻器具有低能耗与高耐久性,同时也说

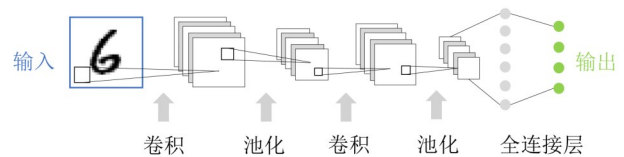


图 18 卷积神经网络结构示意图

明了忆阻器作为 CNN 中的突触器件是可行的。

IBM 团队 Gokmen 等人^[72]在此基础上进一步探讨了基于忆阻器的卷积神经网络可行性,研究了在忆阻器上进行 CNN 训练的方法与训练准确性问题,提出了利用硬件的并行性将卷积层映射到忆阻阵列的方法。

2.5.2.3 长短期神经网络

长短期记忆网络(Long Short-Term Memory, LSTM)是一种时间递归神经网络,解决了循环神经网络中(Recurrent Neural Network, RNN)“梯度消失”的问题,适用于时间间隔和延迟较长的事件的处理与预测,网络结构如图 19 所示。LSTM 的“门”结构具有增减信息到单元状态的功能。第一阶段是遗忘门,其功能是决定单元状态中信息的遗忘与否;第二阶段是输入门,其功能是将输入的信息存放到单元状态中;第三阶段是输出门,其功能是确定输出到下个网络的值。这种“门”结构能够使之记住较长时间前的信息,避免了长期依赖的问题。基于以上特点,利用忆阻存算一体以及阵列实现突触功能,可以使 LSTM 具有高并行低功耗的巨大优势。

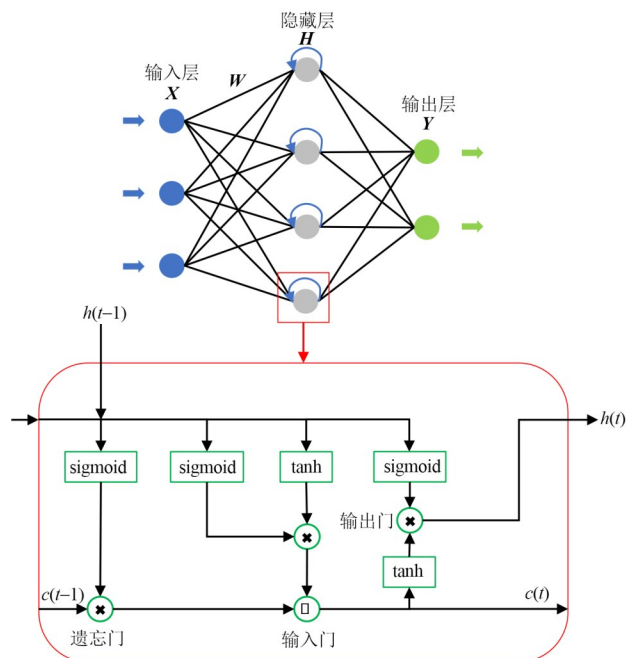


图 19 长短期记忆网络结构示意图

美国马萨诸塞大学杨建华团队^[73]验证了忆阻阵列作为边缘推断平台,利用其低延时、低功耗的特点运行 LSTM 的可行性。他们在 128×64 的 1T1R 阵列上实现了

LSTM 神经网络,并且将该网络应用于全球航空旅客人数预测与人类步态识别等系统.

IBM 公司的 Burr 团队^[74]实现了 LSTM 的前向推断,并且首次在 2.5 M 大小的阵列中实现了近似软件的文本预测准确率.

3 磁隧道结与前沿 MOS 管

3.1 磁隧道结原理与特性

3.1.1 磁隧道结基本原理

目前,用于构建神经形态器件的材料有许多,基于材料的不同,忆阻器也可以分为许多种类. 现如今主要材料包括电阻、铁电、相变、二维和自旋电子材料. 自旋电子学应用中的核心单元是自旋阀或磁隧道连接. 利用诸多物理场,调节核心单元中铁磁层的磁化方向,器件的电阻由 GMR 效应或 TMR 效应引起. 磁隧道结的核心是一种层叠式结构,由铁磁性材料/非磁性绝缘材料/铁磁性材料构成. 这种结构的几何形状类似于忆阻器的基本结构. 该结构包括两层铁磁层,外层相同,但内层不同. 内层的铁磁层中磁化钉扎固定,因此称为钉扎层;而外层铁磁层中的磁矩可以自由转动,因此称为自由层;中间还有一层非磁性绝缘材料构成的绝缘层,为势垒层.

如图 20 所示,两个铁磁层的磁化方向决定了 MTJ 表现出的电阻状态. MTJ 为低电阻状态,两铁磁层完全平行排列;MTJ 为高电阻状态,两个铁磁层的磁化方向完全反向平行排列. 由于磁性隧道结的阻值随自由层

和钉扎层的平行比例决定,故我们可以通过调整注入的磁畴壁的位置来实现平行比例的改变,进而实现对器件电阻值的连续调控.

这种电阻现象即为 TMR 效应(1975 年,Juliere^[27]在 Co/Ge/Fe 磁性隧道结中发现),称为隧穿磁电阻效应.

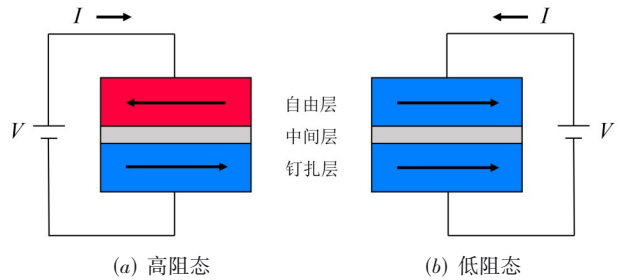


图 20 磁隧道结结构示意图

往往利用高电阻和低电阻的差值与低电阻的比值来计算磁电阻变化率:

$$TMR = (R_{AP} - R_P) / R_P \times 100\% \quad (12)$$

图 21 是隧穿磁阻效应的原理图. 其中一侧铁磁电极中多数自旋子带的电子将被反向散射,使电子能量增加,电流的流动受到抑制,呈现高电阻状态. 少数自旋子带的电子仍会隧穿到另一侧的少数自旋子带的空态. 因为隧穿过程中涉及较少的自旋电子数,所以具有较低的隧穿电导,上述过程中将有较少的电子隧穿入另一侧,因此隧穿电导较低,呈现高电阻状态.

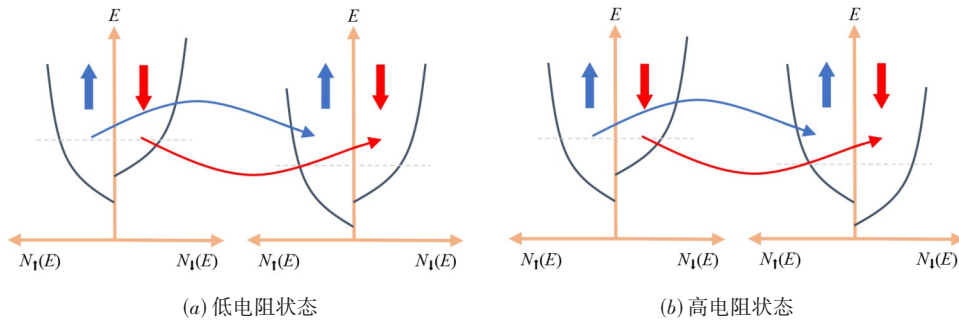


图 21 隧穿磁电阻效应原理示意图

3.1.2 隧道的随机翻转特性

随着磁性随机存储器(Magnetic Random Access Memory, MRAM)的发展,人们对磁矩一定条件下的随机翻转特性的研究不断加深. 磁隧道结之所以能发出随机信号,是因为自由层和钉扎层磁矩是按照垂直方向排列的:当有电流经过时,自旋转移力矩会和磁场力产生碰撞使得磁矩能量不断增加;当磁矩能量接近或高于翻转势垒时,磁矩会由于热扰动而在高低电阻态间随机地发生变化,输出“0”或“1”的二进制随机信号^[75],如

图 22 所示.

磁隧道结独特的随机翻转特性与人脑中神经元的随机性高度匹配,使其在众多前沿研究中有了实际应用. 2013 年,日本 Yuasa 等人^[76]提出,随机数生成装置可以利用磁性随机存储器在一定条件下的磁隧道自由层中磁矩的随机反转特性. 这种随机数发生器具备绝对随机的特性,因为它的随机性基于磁性材料的内禀物理随机性. 同时,磁随机数发生器的制备工艺和应用与磁性随机存储器(MRAM)非常相似,可以通过成熟

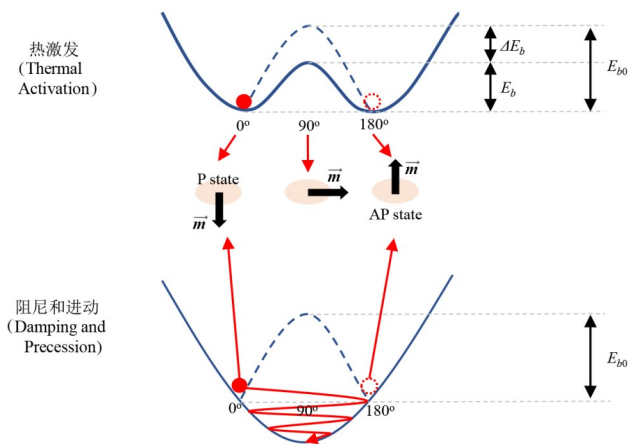


图22 基于电场调控的随机翻转特性示意图

的磁随机存取存储器相关技术,来省略必要的研究历程,具有很大的研究价值和应用前景.2014年,日本 Ando 等人^[77]基于磁隧道结的随机翻转特性设计了一种磁性随机数发生器,通过8个磁隧道结搭建的电路实现8位随机数的生成,并具备实现逻辑“或”和“或非”的功能;2015年,法国的 Querlioz 等人^[78]利用随机翻转特性模拟了大脑中的突触可塑性(根据保留时间的不同,突触可塑性可以简单地分为短时程可塑性 STP 和长时程可塑性 LTP),实现了图像识别功能;2016年,美国的 Sengupta 等人^[79]基于随机翻转特性,利用磁隧道结代替人工神经网络(Artificial Neural Network, ANN)中的神经元,由于磁隧道结随机翻转概率与传统常用的神经元激活函数 Sigmoid 函数相似,故磁隧道结有望替代 ANN 中的传统神经元,在继承传统神经元优越性能的同时发挥出神经形态器件低功耗的优势.

3.1.3 磁隧道结实现神经形态器件的模型

磁隧道结实现神经形态器件的功能,是基于自旋电子器件.21世纪初,美国 Wang 等人^[53]首次提出了基于自旋电子器件的忆阻器理论模型.当时他们一共设计了三种模型,这里选用的磁性隧道结就是第三种模型,具体模型如图23所示,基本原理是将磁畴壁注入磁性隧道结中间,由于磁性隧道结的阻值随自由层和钉扎层的平行比例决定,故可以通过调整注入的磁畴壁的位置 x 来实现平行比例的改变,进而实现对器件电阻值的连续调控.

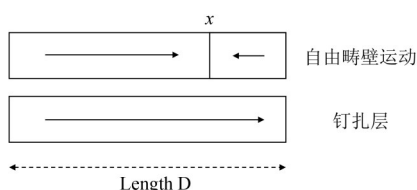


图23 基于磁畴壁运动的理论模型

3.1.4 磁隧道结在类脑计算中的应用

2022年,中国科学技术大学 Luo 团队^[80]设计并建立了一种基于 $\text{Ag}/\text{PbZr}_{0.52}\text{Ti}_{0.48}\text{O}_3$ (PZT, (111)-oriented)/ $\text{Nb}:\text{SrTiO}_3$ 铁电隧道结(Ferroelectric Tunnel Junction, FTJ)的高性能突触器件,通过对铁电畴形态和翻转动力学的设计,在铁电量子隧道结中实现了亚纳秒电脉冲下电导态可非易失连续调控的类脑突触器件,可用于构建人工神经网络类脑计算系统.此系统的优点包括:具有多种铁电开关状态、超细铁电畴和较小的矫顽压;具有高精度、可重现性、线性和对称的权重更新;具有大于 10^9 个周期的良好耐久性和超低的写入能耗.实验结果显示,卷积神经网络模拟在识别 Fashion-MNIST 图像方面实现了约 94.7% 的高在线学习准确率,接近基于浮点运算的卷积神经网络软件计算出的约 95.6% 的结果,如图24所示.

3.2 前沿 MOS 管

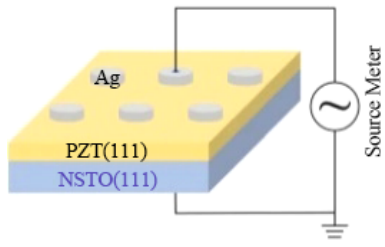
3.2.1 传统 CMOS 技术在当今类脑计算中应用的局限性

2004年以来,各国发现“类脑计算”的巨大价值并且传统的 CMOS 技术可以实现这种“类脑计算”,因此纷纷斥巨资开展相关“脑计划”的研究.这一系列的研究计划大大推动了基于传统 CMOS 技术的神经形态芯片的更深层研究.如图25所示,2013年曼彻斯特大学的 SpiNNaker 芯片^[81],2014年美国 IBM 研制的 TrueNorth 数字神经元^[82],中国研制的天机芯^[13]和达尔文2代等,都没采用传统的冯·诺依曼架构.虽然这些神经形态芯片的功能已经十分强大,但由于这类芯片的基本神经元和突触是基于 CMOS 的数电结构和数模混合电路构成的,需要多个 CMOS 器件才能满足模拟单个神经元或突触的需要,传统 CMOS 技术存在一系列技术瓶颈问题,诸如集成度、功耗和功能模拟准确度等,哪怕是 CMOS 的尺寸已经缩小到极致,所构建的人工神经网络的性能与人脑仍存在巨大差距.

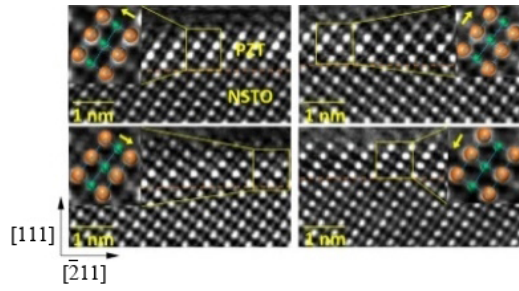
3.2.2 新型浮栅管

3.2.2.1 浮栅管工作原理

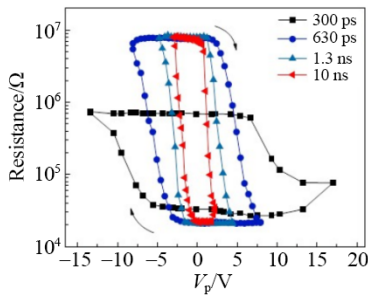
浮栅管的结构与普通场效应管不同,其主要特点在于具有两个栅极的场效应管结构.其中一个栅极被称为“选择栅”或“控制栅”,通过导线引出.这个栅极与普通场效应管的栅极相同.另一个栅极被称为“浮栅”,不与其他部分相连,独立地浮在二氧化硅中.浮栅 MOS 管具有电导可调性、低功耗、多种制造工艺兼容、可变阈值和可以使电路结构简单等特性,因此特别适用于搭建和组成人工神经网络并且已经被广泛应用于类脑神经网络.比较著名的闪存技术(Flash)便使用了浮栅场效应管.在类脑神经网络的构成方面,浮栅管不仅适用于构建神经元完成其功能,还可以作为元件来组成



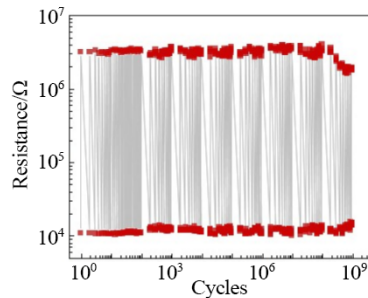
(a) 铁电隧道结类脑突触器件结构示意图



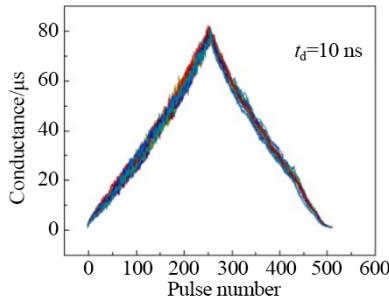
(b) 铁电隧道结不同铁电畴区的原子分辨像



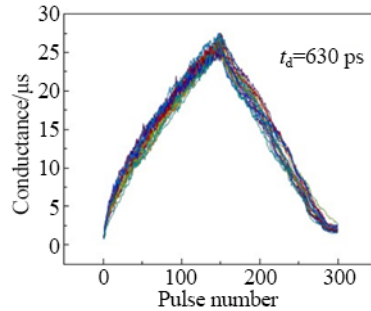
(c) 不同脉宽的电脉冲下隧道结电阻的连续调控



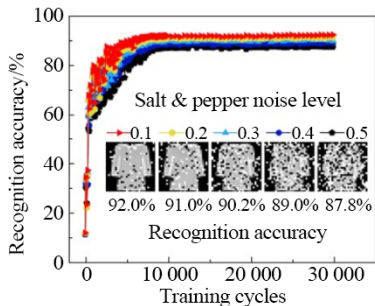
(d) 隧道结类脑突触器件的翻转耐受性



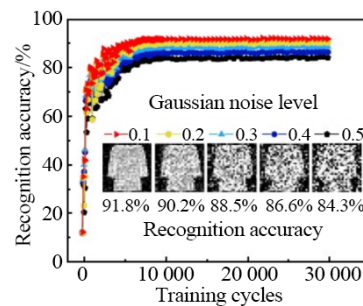
(e) 10 ns 和 630 ps 脉宽的电脉冲下隧道结类脑突触器件的长时程增强和抑制模拟



(f) 10 ns 和 630 ps 脉宽的电脉冲下隧道结类脑突触器件的长时程增强和抑制模拟



(g) 基于隧道结类脑突触器件实验性能,仿真构建的卷积神经网络在线训练识别带椒盐噪声和高斯噪声的Fashion-MNIST图片的结果



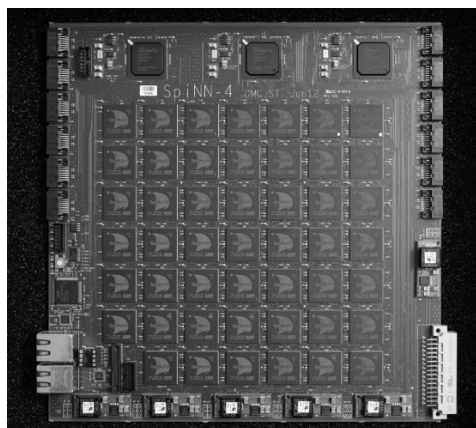
(h) 基于隧道结类脑突触器件实验性能,仿真构建的卷积神经网络在线训练识别带椒盐噪声和高斯噪声的Fashion-MNIST图片的结果

图 24 铁电隧道结类脑突触器件和仿真卷积神经网络

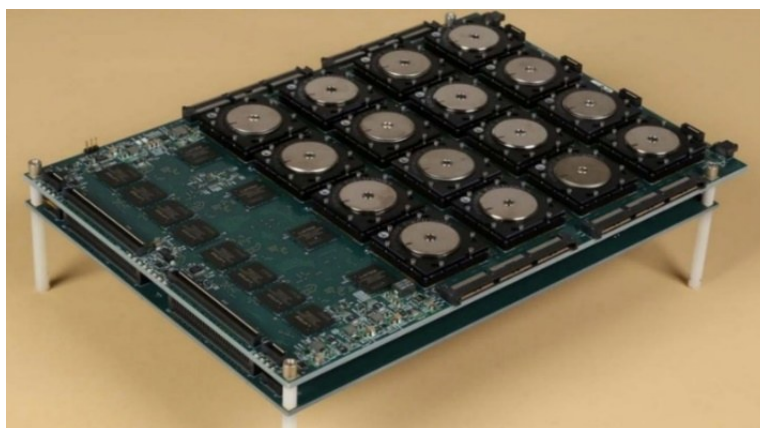
神经突触。

浮栅管的工作原理简单来说就是利用了 MOS 管阈值电压与外部特性的关系,通过改变施加存储在浮栅上的电荷量来进行 MOS 管特性的改变,从而使之工作。

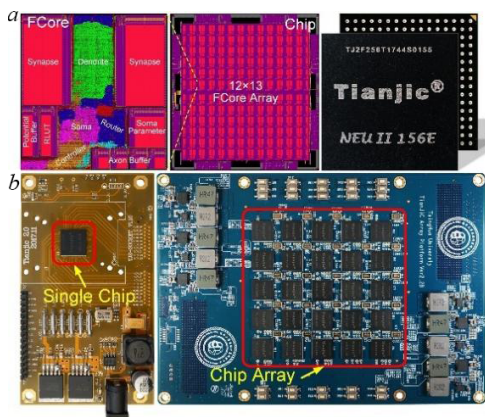
当 MOS 管端口被给予约 20 V 高电压时,源极接地,漏极浮空,此时产生的大量高密度的高能电子,由于其密度过大,一些电子不免会运动到二氧化硅层(在衬底与浮栅端口之间)。若此时的选择栅处于高电压,高能



(a) 曼彻斯特大学的第四代 SpiNNaker 系统^[81]



(b) 集成 16 块 TrueNorth 芯片的 IBM 电路板^[82]



(c) 清华大学研制的天机芯片和测试板^[13]



(d) 浙江大学研制的类脑芯片达尔文 2 代测试版

图 25 基于 CMOS 管的非冯·诺依曼架构的神经形态芯片举例

电子会隧穿二氧化硅层到达浮栅端口,同时移除外部电压,电子没有回路可以放电则会保留在浮栅端口,衬底会被感应成正电荷,这使 MOS 管中的导通电压继续变高;相反,若此时的选择栅处于接地状态,衬底被加上了高电压、源极和漏极开路,电子此时就会被从浮栅端口放电吸出, MOS 管导通电压就会变低。

3.2.2.2 Neu 浮栅管

图 26 所示为神经元 MOS^[83],其本质是一个多输入 Neu 浮栅管。它把电容加入到门极输入端与浮栅之间,从而组成了强耦合的多个栅极输入端。MOS 管导通条件与这多个输入端之和有关。若和大于等效阈值 V_2 电压,则 MOS 导通,否则不通。通过以上机制 Neu 浮栅管可以实现阈值可编程。

3.2.2.3 EFS 浮栅管

EFS 浮动栅极管是一种可调电导浮动栅极单元,具有高度优化的特点。它可以用于组成闪存器件,被应用在 CMOS 集成电路中。同时,它也可以用于对单个器件的状态进行精确调整,并且在低功率亚阈值条件下,可以用于神经网络上的模拟矩阵乘法。类脑神经网络中

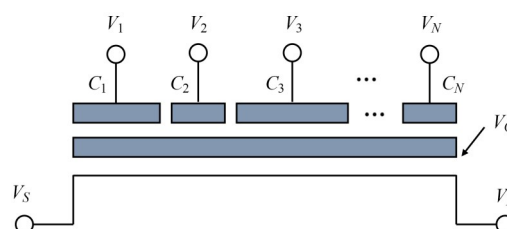


图 26 Neu 浮栅管结构

经常使用模拟电路来模拟神经突触,因为相比成相同功能的数字电路,模拟电路能够单独调整每个设备,并且在电路密度、速度和效率等方面的表现也远超数字电路。2017 年,Guo 等人^[84]就设计制造并测试了这样一种基于 EFS NOR 闪存架构单元的模拟向量矩阵乘法器,可以做到对每个单元进行高精度模拟调谐,其调节误差低于 1%,并且每个矩阵单元仅有 $0.33 \mu\text{m}^2$,单元状态随时间推移具有恒定性,精度约为 2%。

综上所述,浮栅 MOS 管能控制多输入信号,可调节阈值和电导,既能实现神经元的功能,也能实现突触的功能,非常适合作为神经元或突触构建类脑神经网络

络^[85],在神经形态器件的应用会有较高灵活性,在提升电路效率、缩小电路面积方面会有较大提升.

3.2.3 铁电 HfO₂ 场效应晶体管

近年来,铁电场效应晶体管(FeFET)引起了人们的广泛关注.在掺杂了 HfO₂后,科学家们发现了它具有铁电性、良好的兼容性和可扩展性.由于铁电器件的场驱动运行,FeFET具有无损读出、高编写/擦除速度和低功耗等优点.通常情况使用多晶硅作为通道材料,但多晶硅存在一些无法避免的局限性,如多余的电压损失以及亚阈值摆幅(Subthreshold Slope,SS,越小意味着开关速率越快)的退化.最近,受3D-NAND闪存的启发,一种采用IGZO作为通道材料的3D垂直FeFET被以Mo等人^[86]提出并成功演示.同时,该晶体管采用IGZO作为该垂直FeFET的材料.IGZO是一种很有前途的晶体管沟道材料,于2004年首次被报道.由于极薄的尺寸和较高的流动性,IGZO可以作为一种合适的三维垂直FeFET材料,同时也会带来更少的电压损失,具体结构如图27所示.

在器件设计过程中Mo等人发现5nm的IGZO沟道会让FET由于IGZO膜的不均匀性和低迁移率,产生显著的低漏极电流.于是他们别出心裁选用了8nm的IGZO沟道,使器件具有良好的亚阈值摆幅,并且不会造成漏极电流衰减.

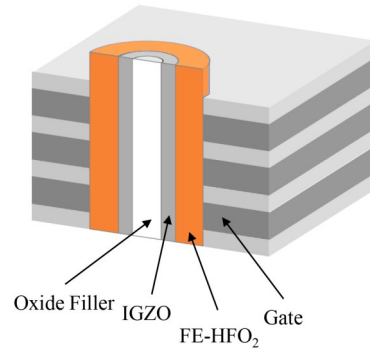


图27 超薄IGZO通道三维垂直场效应晶体管示意图

该场效应管属于铁电材料的较前卫研究,由于IGZO的使用有助于铁电相的形成,金属/IGZO/HZO/金属堆叠的结构更能表现出铁电性,使该器件表现出理想的SS、较高的通道迁移率以及较好的耐久性和保持性,突出了铁电HfO₂场效应晶体管在高密度、低功耗存储领域的应用潜力.

4 潜在的发展方向及公开问题

4.1 潜在发展方向

学术界和产业界针对忆阻器的探索与实用化做出了很多努力,图28为忆阻器发展的重要里程碑.

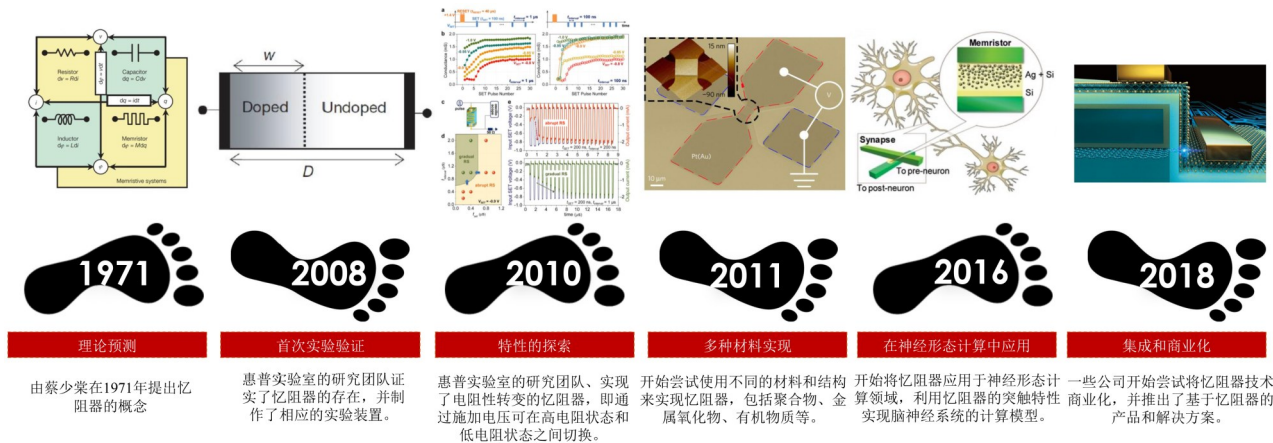


图28 忆阻器发展的重要里程碑

研究聚焦于提高器件性能和解决规模集成的挑战,包括提高工作速度、提高开关比、提高耐久性和数据保持时间以及减小器件的随机差异.在架构方面,交叉阵列(Crossbar Array;图29)结构简单,可大规模集成多层阵列,具有可扩展性,是大规模集成忆阻器的较佳结构^[87],但存在漏电流和接线电阻的问题.目前解决思路包括设计非线性器件、在存储单元中增加选择器、通过系统层面补偿读取电流差异等.

忆阻器能模拟神经突触的可塑性和神经元的积累-发射等行为,天然地具有仿生元件的特征.交叉阵列形

成的神经网络中,每一个节点相当于一个突触,忆阻器的电导起到权重作用,脉冲电压相当于前突触输入的信号,输出的电流相当于后突触的信号.这种天然的存算一体架构消除了数据从存储到计算的搬运,有助于打破“冯·诺依曼瓶颈”,提升系统效率和降低功耗;并且可直接实现向量-矩阵乘法运算,具有很强的并行计算能力.仿生神经网络对数据保持时间和器件耐久性的要求也可以降低.目前,基于脉冲的脉冲神经网络已经发展起来.由于和生物神经系统的高度相似性,在SNN中使用脉冲被认为是实现低能耗计算的关键,基

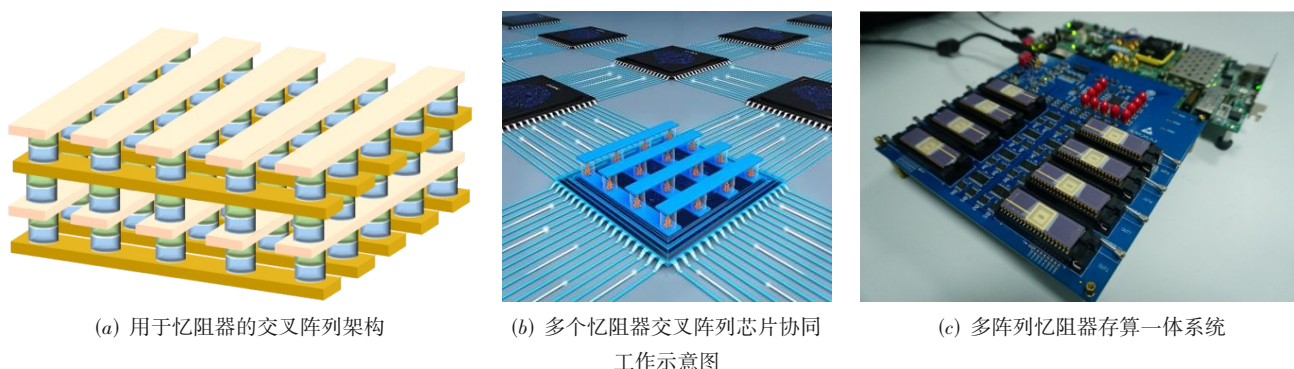


图29 忆阻器的交叉阵列架构

于忆阻器有可能发展出全新的“脉冲电子学”,实现高效率、低功耗的数据传输和处理. 忆阻器的未来研究有以下3个方向.

4.1.1 基于材料工程的高性能忆阻器

二维材料具有原子级厚度,可保证载流子扩散短距离,使快速且低功耗的阻变切换成为可能. 研究表明采用基于二维材料的忆阻器所制备的神经突触功耗可达到飞焦每脉冲的量级. 原子级厚度也提供了器件缩微的良好机会. Mo等人^[88]展示了原子级厚度 MoS_2 单个缺陷尺度上的阻变现象,并称之为Atomristor. 基于层间范德华力的堆垛可将不同功能层集成到同一结构体系中^[89],并构建高密度交叉阵列. 二维材料的高比表面积和平面结构也有利于缺陷工程和导电细丝调控,形成空间位置确定的导电通道,减小器件性能的随机性和差异性. 由于其超薄厚度和优良的力学特性,基于二维材料的忆阻器在柔性可穿戴设备领域也有发展空间.

4.1.2 超低功耗忆阻器

高能忆阻器是大规模集成的前提条件,其关键性能指标在于降低编程所需的电流和电压、提高器件耐久性、降低漏电流,以及克服器件-器件和循环-循环的统计差异性,在这些方面仍然需要进行深入研究. 对于神经网络的训练,关键在于编程和矩阵运算精确性,目前在基于节点相干性的神经网络加速器方面的研究已取得重要进展.

4.1.3 可扩展的忆阻器硬件系统

要实现实用化,器件需要规模化集成. 技术路径有以下几条. (1)增加忆阻器网络中的节点数量,这方面需要提升器件的制备工艺集成度. (2)使用三维架构,交叉阵列在这方面提供一个很好的架构,可以通过多层堆叠方式,增加运算维数;具有可堆垛性的二维材料和交叉结构的完美结合是很有潜力的方向. (3)器件多功能化,使单一器件能实现不同功能,如数据存储、运算操作、动态重构等. (4)与CMOS技术融合集成,利用CMOS极高的集成度,实现大规模的集成忆阻器要求高

密度的局部互联,这在近期的实验进展中已被证明是可行的^[90].

4.2 公开问题

在生物神经系统中,神经元作为最基本的计算单元,能有效整合信息并以低功耗运行,而连接神经元的突触则是重要的学习组件. 突触可塑性代表学习和记忆的重要特征^[91]. 因此,构建高度仿生的神经元和突触是实现大规模神经形态硬件和网络的核心问题. 目前主流的CMOS晶体管不具有内在的生物相似性,并且不适用类脑芯片的进一步发展,因此通过新材料和新器件去构建神经元和突触是未来的重点研究方向. 人们对神经形态技术的前景寄予厚望. 迄今为止,大多数神经形态研究都集中在该技术对边缘应用的承诺上,但新的发展表明,神经形态计算还可以为大型、复杂的计算问题提供价值,通过从根本上进行学习,实现实时处理,更好地解决问题.

神经形态器件种类繁多,各有用武之地. 基于传统成熟的CMOS技术虽然有发展阈值,但其成熟的技术基础和完善的电路结构可以实现绝大部分当下的实际需求. 我国近几年在以CMOS技术为基础的领域取得不少优秀的业绩,使我国芯片技术达到全新的高度,但仍有的可以前进的空间. 基于新型非易失性存储器件和新材料构建的新型存储器件,具有更低功耗、更高耐久性、更快读写速度等明显优势,被各国视为重要科研发展领域,发展与日俱进,也延伸出了各个分支,诸如ReRAM, FeRAM, MRAM以及PCM等,但仍可算作处于探索研究阶段.

5 总结与展望

本文首先通过“冯·诺依曼架构”的性能瓶颈,引出类脑计算与神经形态器件的优势,并梳理神经形态器件发展历史;接着详细介绍了忆阻器,列举了基本忆阻器种类和几种常用的忆阻器模型,从模型、材料和忆阻机制三个方面对现有忆阻器进行了对比分析,展示了

目前使用忆阻器在传统领域与类脑计算领域取得的一些研究成果;然后对磁隧道结与前沿 MOS 管作了概述,在概述中,一方面对磁隧道结基本原理及其独特的随机翻转特性作了简要说明,另一方面介绍了浮栅 MOS 管的工作原理并列了几种新型浮栅管;最后对未来忆阻器的研究方向和神经形态器件的研究前景作了预测和展望。

忆阻器在仿生器件方面的应用也是一个值得深入研究和发掘的方向。最近的研究成果显示基于脉冲的学习方式和生物中的感知现象具有惊人的相似性。目前,仿生器件大多是基于后处理的方式来实现仿生功能的,而未来主要研究方向是通过全硬件或主要通过硬件来模仿人脑的丰富功能^[92]。总之,忆阻器在超越冯·诺依曼架构的后摩尔计算系统中已经展现出巨大潜力。未来将会在新的功能材料和阻变机制的发现、超低功耗、高可拓展性、高稳定性、高耐久性等方面不断推进,或许能带来新一轮的产业化革命。

目前人工突触、神经元和硬件网络的开发都已取得显著进展,但神经形态计算仍处于起步阶段,在器件和系统层面都存在诸多挑战需要应对^[93]。在器件层面,主要障碍是技术不成熟,严重影响阵列的良率、稳定性和均匀性。在已经报道的人工突触中,可用于实现大规模阵列的器件非常有限,大多数演示都集中在使用单个器件进行概念验证,忽略了器件阵列的良率和均匀性。在系统层面,经由未选择单元的潜行电流会导致错误操作并限制阵列集成度,使具有自校正功能的器件在大规模阵列中具有紧迫性。虽然仿生突触和神经元已经展示了良好的发展前景,但还有诸多潜能等待发掘。

致谢 感谢西安电子科技大学杨银堂教授对本文的悉心指导。

参考文献

- [1] 邵博文. 冯·诺依曼的逻辑和计算机思想[J]. 数字通信世界, 2017(10): 249.
SHAO B W. Von Neumann's logic and computer thought [J]. Digital Communication World, 2017(10): 249. (in Chinese)
- [2] 龚庆悦, 董海艳, 冒宇清. 医学信息工程概论[M]. 南京: 南京大学出版社, 2019.
GONG Q Y, DONG H Y, MAO Y Q. Introduction to Medical Information Engineering[M]. Nanjing: Nanjing University Press, 2019. (in Chinese)
- [3] MARKRAM H. The Human Brain Project A Report to the European Commission[R]. The HBP Report, 2012.
- [4] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [5] TENENBAUM J B, KEMP C, GRIFFITHS T L, et al. How to grow a mind: Statistics, structure, and abstraction [J]. Science, 2011, 331(6022): 1279-1285.
- [6] 殷明慧, 杨玉超, 黄如. 神经形态器件现状与未来[J]. 国防科技, 2016, 37(6): 23-30.
YIN M H, YANG Y C, HUANG R. Progress and outlook of neuromorphic devices[J]. National Defense Science & Technology, 2016, 37(6): 23-30. (in Chinese)
- [7] MEAD C. Neuromorphic electronic systems[J]. Proceedings of the IEEE, 1990, 78(10): 1629-1636.
- [8] SHIBATA T, OHMI T. A functional MOS transistor featuring gate-level weighted sum and threshold operations [J]. IEEE Transactions on Electron Devices, 1992, 39(6): 1444-1455.
- [9] SEO J S, BREZZO B, LIU Y, et al. A 45nm CMOS neuromorphic chip with a scalable architecture for learning in networks of spiking neurons[C]//2011 IEEE Custom Integrated Circuits Conference (CICC). Piscataway: IEEE, 2011: 1-4.
- [10] MEROLLA P A, ARTHUR J V, ALVAREZ-ICAZA R, et al. Artificial brains. A million spiking-neuron integrated circuit with a scalable communication network and interface[J]. Science, 2014, 345(6197): 668-673.
- [11] BENJAMIN B V, GAO P R, MCQUINN E, et al. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations[J]. Proceedings of the IEEE, 2014, 102(5): 699-716.
- [12] DAVIES M, SRINIVASA N, LIN T H, et al. Loihi: A neuromorphic manycore processor with on-chip learning [J]. IEEE Micro, 2018, 38(1): 82-99.
- [13] PEI J, DENG L, SONG S, et al. Towards artificial general intelligence with hybrid Tianjic chip architecture[J]. Nature, 2019, 572(7767): 106-111.
- [14] 张玲, 刘国柱, 于宗光. 人工神经形态器件发展现状与展望[J]. 电子与封装, 2021, 21(6): 5-18.
ZHANG L, LIU G Z, YU Z G. Recent advances in neuromorphic devices[J]. Electronics & Packaging, 2021, 21(6): 5-18. (in Chinese)
- [15] 谢生, 高旭斌, 毛陆虹. 基于标准 CMOS 工艺的光电浮栅突触研究[J]. 天津大学学报(自然科学与工程技术版), 2022, 55(7): 701-705.
XIE S, GAO X B, MAO L H. Photoelectric floating-gate synapse based on standard CMOS process[J]. Journal of

- Tianjin University (Science and Technology), 2022, 55 (7): 701-705. (in Chinese)
- [16] CHUA L. Memristor-The missing circuit element[J]. IEEE Transactions on Circuit Theory, 1971, 18(5): 507-519.
- [17] STRUKOV D B, SNIDER G S, STEWART D R, et al. The missing memristor found[J]. Nature, 2008, 453 (7191): 80-83.
- [18] JO S H, CHANG T, EBONG I, et al. Nanoscale memristor device as synapse in neuromorphic systems[J]. Nano Letters, 2010, 10(4): 1297-1301.
- [19] KIM S, DU C, SHERIDAN P, et al. Experimental demonstration of a second-order memristor and its ability to bio-realistically implement synaptic plasticity[J]. Nano Letters, 2015, 15(3): 2203-2211.
- [20] WANG Z W, YIN M H, ZHANG T, et al. Engineering incremental resistive switching in TaOx based memristors for brain-inspired computing[J]. Nanoscale, 2016, 8(29): 14015-14022.
- [21] WANG Z R, JOSHI S, SAVEL'EV S E, et al. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing[J]. Nature Materials, 2017, 16 (1): 101-108.
- [22] LV Z Y, WANG Y, CHEN J R, et al. Semiconductor quantum dots for memories and neuromorphic computing systems[J]. Chemical Reviews, 2020, 120(9): 3941-4006.
- [23] JI X L, ZHAO X Y, TAN M C, et al. Artificial perception built on memristive system: Visual, auditory, and tactile sensations[J]. Advanced Intelligent Systems, 2020, 2(3): 1900118.
- [24] LIU K Q, ZHANG T, DANG B J, et al. An optoelectronic synapse based on α -In₂Se₃ with controllable temporal dynamics for multimode and multiscale reservoir computing[J]. Nature Electronics, 2022, 5(11): 761-773.
- [25] 朱佳雪, 张续猛, 王睿, 等. 面向神经形态感知和计算的柔性忆阻器基脉冲神经元[J]. 物理学报, 2022, 71(14): 338-349.
- ZHU J X, ZHANG X M, WANG R, et al. Flexible memristive spiking neuron for neuromorphic sensing and computing[J]. Acta Physica Sinica, 2022, 71(14): 338-349. (in Chinese)
- [26] OVSHINSKY S R. Reversible electrical switching phenomena in disordered structures[J]. Physical Review Letters, 1968, 21(20): 1450-1453.
- [27] JULLIERE M. Tunneling between ferromagnetic films [J]. Physics Letters A, 1975, 54(3): 225-226.
- [28] SURI M N, BICHLER O, QUERLIOZ D, et al. Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction [C]//2011 International Electron Devices Meeting. Piscataway: IEEE, 2012: 4.4.1-4.4.4.
- [29] KUZUM D, JEYASINGH R G D, LEE B, et al. Nano-electronic programmable synapses based on phase change materials for brain-inspired computing[J]. Nano Letters, 2012, 12(5): 2179-2186.
- [30] TUMA T, PANTAZI A, LE GALLO M, et al. Stochastic phase-change neurons[J]. Nature Nanotechnology, 2016, 11(8): 693-699.
- [31] SRINIVASAN G, SENGUPTA A, ROY K. Magnetic tunnel junction based long-term short-term stochastic synapse for a spiking neural network with on-chip STDP learning[J]. Scientific Reports, 2016, 6: 29545.
- [32] SENGUPTA A, PANDA P, WIJESINGHE P, et al. Magnetic tunnel junction mimics stochastic cortical spiking neurons[J]. Scientific Reports, 2016, 6: 30039.
- [33] KANEKO Y, NISHITANI Y, UEDA M. Ferroelectric artificial synapses for recognition of a multishaded image [J]. IEEE Transactions on Electron Devices, 2014, 61(8): 2827-2833.
- [34] KIM M K, LEE J S. Ferroelectric analog synaptic transistors[J]. Nano Letters, 2019, 19(3): 2044-2050.
- [35] JANG S, JANG S, LEE E H, et al. Ultrathin conformable organic artificial synapse for wearable intelligent device applications[J]. ACS Applied Materials & Interfaces, 2019, 11(1): 1071-1080.
- [36] MIAN M R, CHEN H Y, CAO R, et al. Insights into catalytic hydrolysis of organophosphonates at M-OH sites of azolate-based metal organic frameworks[J]. Journal of the American Chemical Society, 2021, 143(26): 9893-9900.
- [37] CAO R R, ZHANG X M, LIU S, et al. Compact artificial neuron based on anti-ferroelectric transistor[J]. Nature Communications, 2022, 13: 7018.
- [38] KIM K, CHEN C L, TRUONG Q, et al. A carbon nanotube synapse with dynamic logic and learning[J]. Advanced Materials, 2013, 25(12): 1693-1698.
- [39] ZHU L Q, WAN C J, GUO L Q, et al. Artificial synapse network on inorganic proton conductor for neuromorphic systems[J]. Nature Communications, 2014, 5: 3158.
- [40] YANG Y, WEN J, GUO L Q, et al. Long-term synaptic plasticity emulated in modified graphene oxide electro-

- lyte gated IZO-based thin-film transistors[J]. *ACS Applied Materials & Interfaces*, 2016, 8(44): 30281-30286.
- [41] JIANG J E, GUO J J, WAN X A, et al. 2D MoS₂ neuromorphic devices for brain-like computational systems[J]. *Small*, 2017, 13(29): 1700933.
- [42] FENG P, XU W W, YANG Y, et al. Printed neuromorphic devices based on printed carbon nanotube thin-film transistors[J]. *Advanced Functional Materials*, 2017, 27(5): 1604447.
- [43] JOHN R A, TIWARI N, CHEN Y Y, et al. Ultralow power dual-gated subthreshold oxide neuristors: An enabler for higher order neuronal temporal correlations[J]. *ACS Nano*, 2018, 12(11): 11263-11273.
- [44] HE Y L, NIE S, LIU R, et al. Spatiotemporal information processing emulated by multiterminal neuro-transistor networks[J]. *Advanced Materials*, 2019, 31(21): e1900903.
- [45] 董哲康. 基于忆阻器的电路分析及其在神经形态系统中的应用[D]. 杭州: 浙江大学, 2019.
DONG Z K. Memristor based Circuit Analysis and its Applications in Neuromorphic Systems[D]. Hangzhou: Zhejiang University, 2019. (in Chinese)
- [46] 刘益春, 林亚, 王中强, 等. 氧化物基忆阻型神经突触器件[J]. *物理学报*, 2019, 68(16): 119-137.
LIU Y C, LIN Y, WANG Z Q, et al. Oxide-based memristive neuromorphic synaptic devices[J]. *Acta Physica Sinica*, 2019, 68(16): 119-137. (in Chinese)
- [47] YANG J J, PICKETT M D, LI X M, et al. Memristive switching mechanism for metal/oxide/metal nanodevices [J]. *Nature Nanotechnology*, 2008, 3(7): 429-433.
- [48] 顾野. 忆阻器件的特性与电路应用研究[D]. 成都: 电子科技大学, 2015.
GU Y. Research on Memristor Device Characteristics and Circuit Applications[D]. Chengdu: University of Electronic Science and Technology of China, 2015. (in Chinese)
- [49] DONG Z K, LI C Y, QI D L, et al. Multiple memristor circuit parametric fault diagnosis using feedback-control doublet generator[J]. *IEEE Access*, 2016, 4: 2604-2614.
- [50] VOURKAS I, SIRAKOULIS G C. A novel design and modeling paradigm for memristor-based crossbar circuits [J]. *IEEE Transactions on Nanotechnology*, 2012, 11(6): 1151-1159.
- [51] KVATINSKY S, RAMADAN M, FRIEDMAN E G, et al. VTEAM: A general model for voltage-controlled memristors[J]. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2015, 62(8): 786-790.
- [52] KVATINSKY S, FRIEDMAN E G, KOLODNY A, et al. TEAM: ThrEshold adaptive memristor model[J]. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2013, 60(1): 211-221.
- [53] WANG X B, CHEN Y R, XI H W, et al. Spintronic memristor through spin-torque-induced magnetization motion [J]. *IEEE Electron Device Letters*, 2009, 30(3): 294-297.
- [54] HU M, LI H, CHEN Y R, et al. Geometry variations analysis of TiO₂ thin-film and spintronic memristors[C]//16th Asia and South Pacific Design Automation Conference. Piscataway: IEEE, 2011: 25-30.
- [55] CHANG T, JO S H, KIM K H, et al. Synaptic behaviors and modeling of a metal oxide memristive device[J]. *Applied Physics A*, 2011, 102(4): 857-863.
- [56] YAN X B, PEI Y F, CHEN H W, et al. Self-assembled networked PbS distribution quantum dots for resistive switching and artificial synapse performance boost of memristors[J]. *Advanced Materials*, 2019, 31(7): 1805284.
- [57] SLAUGHTER J M. Materials for magnetoresistive random access memory[J]. *Annual Review of Materials Research*, 2009, 39: 277-296.
- [58] WANG Z R, JOSHI S, SAVEL'EV S, et al. Fully memristive neural networks for pattern classification with unsupervised learning[J]. *Nature Electronics*, 2018, 1(2): 137-145.
- [59] CAI F X, CORRELL J M, LEE S H, et al. A fully integrated reprogrammable memristor-CMOS system for efficient multiply-accumulate operations[J]. *Nature Electronics*, 2019, 2(7): 290-299.
- [60] IELMINI D, AMBROGIO S, MILO V, et al. Neuromorphic computing with hybrid memristive/CMOS synapses for real-time learning[C]//2016 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway: IEEE, 2016: 1386-1389.
- [61] CHEN P Y, YU S M. Partition SRAM and RRAM based synaptic arrays for neuro-inspired computing[C]//2016 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway: IEEE, 2016: 2310-2313.
- [62] KIM S G, HAN J S, KIM H, et al. Recent advances in memristive materials for artificial synapses[J]. *Advanced Materials Technologies*, 2018, 3(12): 1800457.
- [63] TSAI H, AMBROGIO S, NARAYANAN P, et al. Recent progress in analog memory-based accelerators for deep

- learning[J]. *Journal of Physics D: Applied Physics*, 2018, 51(28): 283001.
- [64] SUNG C, HWANG H, YOO I K. Perspective: A review on memristive hardware for neuromorphic computation [J]. *Journal of Applied Physics*, 2018, 124(15): 151903.
- [65] CRISTIANO G, GIORDANO M, AMBROGIO S, et al. Perspective on training fully connected networks with resistive memories: Device requirements for multiple conductances of varying significance[J]. *Journal of Applied Physics*, 2018, 124(15): 151901.
- [66] LI C, BELKIN D, LI Y N, et al. Efficient and self-adaptive in situ learning in multilayer memristor neural networks[J]. *Nature Communications*, 2018, 9: 2385.
- [67] YAO P, WU H Q, GAO B, et al. Face classification using electronic synapses[J]. *Nature Communications*, 2017, 8: 15199.
- [68] BAYAT F M, PREZIOSO M, CHAKRABARTI B, et al. Implementation of multilayer perceptron network with highly uniform passive memristive crossbar circuits[J]. *Nature Communications*, 2018, 9: 2331.
- [69] KWAK M, PARK J, WOO J, et al. Implementation of convolutional kernel function using 3-D TiOx resistive switching devices for image processing[J]. *IEEE Transactions on Electron Devices*, 2018, 65(10): 4716-4718.
- [70] YAKOPCIC C, ALOM M Z, TAHA T M. Memristor crossbar deep network implementation based on a Convolutional neural network[C]//2016 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2016: 963-970.
- [71] GARBIN D, VIANELLO E, BICHLER O, et al. HfO₂-based OxRAM devices as synapses for convolutional neural networks[J]. *IEEE Transactions on Electron Devices*, 2015, 62(8): 2494-2501.
- [72] GOKMEN T, ONEN M, HAENSCH W. Training deep convolutional neural networks with resistive cross-point devices[J]. *Frontiers in Neuroscience*, 2017, 11: 538.
- [73] LI C, WANG Z R, RAO M Y, et al. Long short-term memory networks in memristor crossbar arrays[J]. *Nature Machine Intelligence*, 2019, 1(1): 49-57.
- [74] TSAI H, AMBROGIO S, MACKIN C, et al. Inference of Long-Short Term Memory networks at software-equivalent accuracy using 2.5M analog Phase Change Memory devices[C]//2019 Symposium on VLSI Technology. Piscataway: IEEE, 2019: T82-T83.
- [75] 蔡佳林. 基于自旋电子器件的神经形态器件研究[D]. 合肥: 中国科学技术大学, 2019.
- CAI J L. *Research on Neuromorphic Devices Based on Spintronics*[D]. Hefei: University of Science and Technology of China, 2019. (in Chinese)
- [76] YUASA S, FUKUSHIMA A, YAKUSHIJI K, et al. Future prospects of MRAM technologies[C]//2013 IEEE International Electron Devices Meeting. Piscataway: IEEE, 2014: 3.1.1-3.1.4.
- [77] FUKUSHIMA A, SEKI T, YAKUSHIJI K, et al. Spin dice: A scalable truly random number generator based on spintronics[J]. *Applied Physics Express*, 2014, 7(8): 083001.
- [78] VINCENT A F, LARROQUE J, LOCATELLI N, et al. Stochastic memristive synapses from spin-transfer torque magnetic tunnel junctions[C]//2015 IEEE International Magnetics Conference (INTERMAG). Piscataway: IEEE, 2015: 1-1.
- [79] SENGUPTA A, PARSA M, HAN B, et al. Probabilistic deep spiking neural systems enabled by magnetic tunnel junction[J]. *IEEE Transactions on Electron Devices*, 2016, 63(7): 2963-2970.
- [80] LUO Z, WANG Z J, GUAN Z Y, et al. High-precision and linear weight updates by subnanosecond pulses in ferroelectric tunnel junction for neuro-inspired computing [J]. *Nature Communications*, 2022, 13: 699.
- [81] PAINKRAS E, PLANA L A, GARSIDE J, et al. SpiNNaker: A 1-W 18-core system-on-chip for massively-parallel neural network simulation[J]. *IEEE Journal of Solid-State Circuits*, 2013, 48(8): 1943-1953.
- [82] AKOPYAN F, SAWADA J, CASSIDY A, et al. TrueNorth: Design and tool flow of a 65 mW 1 million neuron programmable neurosynaptic chip[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2015, 34(10): 1537-1557.
- [83] SHIBATA T, OHMI T. An intelligent MOS transistor featuring gate-level weighted sum and threshold operations [C]//International Electron Devices Meeting 1991 [Technical Digest]. Piscataway: IEEE, 2002: 919-922.
- [84] GUO X, BAYAT F M, PREZIOSO M, et al. Temperature-insensitive analog vector-by-matrix multiplier based on 55 nm NOR flash memory cells[C]//2017 IEEE Custom Integrated Circuits Conference (CICC). Piscataway: IEEE, 2017: 1-4.
- [85] JACKSON B L, RAJENDRAN B, CORRADO G S, et al. Nanoscale electronic synapses using phase change devices

[J]. ACM Journal on Emerging Technologies in Computing Systems, 2013, 9(2): 12.

- [86] MO F, TAGAWA Y, JIN C J, et al. Low-voltage operating ferroelectric FET with ultrathin IGZO channel for high-density memory application[J]. IEEE Journal of the Electron Devices Society, 2020, 8: 717-723.
- [87] 徐丽莹, 杨玉超, 黄如. 基于忆阻器的非易失逻辑研究前沿[J]. 中国基础科学, 2019, 21(2): 1-11, 27, 63.
XU L Y, YANG Y C, HUANG R. Research frontier at memristor-based nonvolatile logic[J]. China Basic Science, 2019, 21(2): 1-11, 27, 63. (in Chinese)
- [88] GE R J, WU X H, KIM M, et al. Atomristor: Nonvolatile resistance switching in atomic sheets of transition metal dichalcogenides[J]. Nano Letters, 2018, 18(1): 434-441.
- [89] 李锲, 曹荣荣, 孙毅, 等. 基于忆阻器的感存算一体技术研究进展[J]. 微纳电子与智能制造, 2019, 1(4): 87-102.
LI K, CAO R R, SUN Y, et al. Research progress on the fused technology of sensing, storage and computing based on memristor[J]. Micro/Nano Electronics and Intelligent Manufacturing, 2019, 1(4): 87-102. (in Chinese)
- [90] 张晨曦, 陈艳, 仪明东, 等. 基于忆阻器模拟的突触可塑性的研究进展[J]. 中国科学: 信息科学, 2018, 48(2): 115-142.
ZHANG C X, CHEN Y, YI M D, et al. Recent progress in memristors for stimulating synaptic plasticity[J]. Scientia Sinica (Informationis), 2018, 48(2): 115-142. (in Chinese)
- [91] 张鑫, 夏天, 叶葱, 等. 基于忆阻器实现多样化 STDP 学习规则的突触电路设计[J]. 中国科学: 技术科学, 2021, 51(1): 89-98.
ZHANG X, XIA T, YE C, et al. Realizing diverse STDP learning rules in synaptic circuit based on memristor[J]. Scientia Sinica (Technologica), 2021, 51(1): 89-98. (in Chinese)
- [92] 陆骐峰, 孙富钦, 王子豪, 等. 柔性人工突触: 面向智能人机交互界面和高效神经网络计算的基础器件[J]. 材料导报, 2020, 34(1): 1022-1049.
LU Q F, SUN F Q, WANG Z H, et al. Recent advances in flexible artificial synapses towards intelligent human-machine interface and neuromorphic computation systems [J]. Materials Reports, 2020, 34(1): 1022-1049. (in Chinese)
- [93] 王洋昊, 刘昌, 黄如, 等. 神经形态器件研究进展与未来趋势[J]. 科学通报, 2020, 65(10): 904-915.
WANG Y H, LIU C, HUANG R, et al. Progresses and

outlook in neuromorphic devices[J]. Chinese Science Bulletin, 2020, 65(10): 904-915. (in Chinese)

作者简介



曹 震 男, 1989年3月出生于山东省微山县. 现为西安电子科技大学人工智能学院华山学者准聘副教授, 硕士生导师. 主要研究方向为微电子器件设计与人工智能优化方法前沿交叉、脉冲神经网络 FPGA、智能芯片设计与智能系统等.

E-mail: caozhen@xidian.edu.cn



张浴轩 男, 1999年4月出生于辽宁省沈阳市. 2022年于西安邮电大学计算机学院获学士学位, 现为西安电子科技大学人工智能学院硕士研究生. 主要研究方向为神经网络 FPGA 硬件边缘部署.

E-mail: 22171214686@stu.xidian.edu.cn



李灵蕾 女, 2002年2月出生于山西省太原市. 现为西安电子科技大学人工智能学院智能科学与技术专业硕士研究生.

E-mail: linglei_li@163.com



郭 璋 男, 1991年8月21日生于陕西省渭南市. 现为西安电子科技大学前沿交叉研究院华山学者准聘副教授, 硕士生导师. 主要研究方向为忆阻器与神经形态器件、类脑计算与人工智能、状态监测与故障诊断等.

E-mail: guozhang@xidian.edu.cn



孙 琦 男, 2002年3月出生于湖北省孝感市. 2023年于西安电子科技大学大学人工智能学院获学士学位, 现为西安电子科技大学人工智能学院硕士研究生. 主要研究方向为人工智能半导体交叉.

E-mail: 23171214665@stu.xidian.edu.cn



曹荣荣 女, 1990年6月出生于山东省临沂市. 现为国防科技大学电子科学学院副教授、硕士生导师, 青年托举人才、湖南省湖湘英才. 主要研究方向为新原理智能信息器件开发及其类脑计算.

E-mail: caorongrong@nudt.edu.cn



侯 彪 男,1974年11月出生于陕西省宝鸡市.现为西安电子科技大学人工智能学院华山学者特聘教授、博士生导师.主要研究方向为遥感图像解译与目标识别、无人系统协同感知、人工智能芯片及系统等.

E-mail: avcodec@163.com



焦李成 男,1959年10月出生于陕西省白水县.现为西安电子科技大学人工智能研究院华山学者杰出教授,欧洲科学院外籍院士、俄罗斯自然科学院外籍院士、IEEE Fellow.主要研究方向为智能感知与图像理解、深度学习与类脑计算、进化优化与遥感解译.

E-mail: lchjiao@mail.xidian.edu.cn