



Institute of Physics, CAS

面向类脑计算的物理电子学专题编者按

Preface to the special topic: Physical electronics for brain-inspired computing

引用信息 Citation: Acta Physica Sinica, 71, 140101 (2022) DOI: 10.7498/aps.71.140101 在线阅读 View online: https://doi.org/10.7498/aps.71.140101 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

柔性电子专题编者按 Preface to the special topic: Flexible electronics 物理学报. 2020, 69(17): 170101 https://doi.org/10.7498/aps.69.170101

电介质材料和物理专题编者按

Preface to the special topic: Dielectric materials and physics 物理学报. 2020, 69(12): 120101 https://doi.org/10.7498/aps.69.120101

机器学习与物理专题编者按

Preface to the special topic: Machine learning and physics 物理学报. 2021, 70(14): 140101 https://doi.org/10.7498/aps.70.140101

太赫兹自旋光电子专题编者按

Preface to the special topic: Terahertz spintronic optoelectronics 物理学报. 2020, 69(20): 200101 https://doi.org/10.7498/aps.69.200101

非厄米物理前沿专题编者按 Preface to the special topic: Frontiers in non-Hermitian Physics 物理学报. 2022, 71(13): 130101 https://doi.org/10.7498/aps.71.130101

光学超构材料专题编者按 Preface to the special topic: Optical metamaterials 物理学报. 2020, 69(15): 150101 https://doi.org/10.7498/aps.69.150101 专题: 面向类脑计算的物理电子学

面向类脑计算的物理电子学专题编者按

DOI: 10.7498/aps.71.140101

类脑计算技术作为一种脑启发的新型计算技术,具有存算一体、事件驱动、模拟并行等特征,为 智能化时代开发高效的计算硬件提供了技术参考,有望解决当前人工智能硬件在能耗和算力方面的 "不可持续发展"问题.硬件模拟神经元和突触功能是发展类脑计算技术的核心,而支持这一切实现 的基础是器件以及器件中的物理电子学.根据类脑单元实现的物理基础,当前类脑芯片主要可以分 为数字 CMOS 型、数模混合 CMOS 型以及新原理器件型三大类.IBM 的 TrueNorth、Intel 的 Loihi、清华大学的 Tianjic 以及浙江大学的 Darwin 等都是数字 CMOS 型类脑芯片的典型代表,旨 在以逻辑门电路仿真实现生物单元的行为.数模混合型的基本思想是利用亚阈值模拟电路模拟生物 神经单元的特性,最早由 Carver Mead 提出,其成功案例有苏黎世的 ROLLs、斯坦福的 Neurogrid 等.以上两种类型的类脑芯片虽然实现方式上有所不同,但共同之处在于都是利用了硅基晶体管的 物理特性.此外,以忆阻器为代表的新原理器件为构建非硅基类脑芯片提供了新的物理基础.它们 在工作过程中引入了离子动力学特性,从结构和工作机制上与生物单元都具有很高的相似性,近年 来受到国内外产业界和学术界的广泛关注.鉴于硅基工艺比较成熟,当前硅基物理特性是类脑芯片 实现的主要基础.忆阻器等新原理器件的类脑计算技术尚处于前沿探索和开拓阶段,还需要更成熟 的制备技术、更完善的系统框架和电路设计以及更高效的算法等.

为促进本领域国内同行交流,应《物理学报》编辑部邀请,我们邀请部分活跃在利用物理器件 实现类脑计算领域第一线的中青年科学家,组织出版了本专题.探讨不同物理机制的器件实现、算 法优化、架构设计以及应用.器件层面上,刘琦老师报道了一种柔性忆阻器基神经元器件及电路,徐 文涛老师报道了一种基于 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 的离子迁移型人造突触,缪向水老师综述了主流忆阻 器的器件结构、物理机制并比较分析了它们的性能特性,万昌锦老师和万青老师介绍了多类柔性神 经形态晶体管的研究进展以及在仿生感知领域中的应用;算法层面上,曾中明老师和袁喆老师利用 磁性隧道结可调控的随机动力学实现了群体编码,尚大山老师提出了一种基于存内计算范式的储池 计算硬件实现方法;系统层面上,刘洋老师报道了一种基于忆阻器的脉冲神经网络硬件加速器架构 设计方法,郭新老师提出了一种仿生生物感官的感存算一体化系统,康晋锋老师综述了用于实现存 内计算的非挥发存储器件及其性能特征;感知应用层面上,王中强老师从器件物理角度、周菲迟老 师和柴扬老师从应用角度分别综述了面向感存算功能一体化的光电忆阻器研究进展,诸葛飞老师综 述了光电神经形态器件及其应用的最新研究进展,韩素婷老师综述了应用于感存算一体化系统忆阻 器的研究方向和研究进展.

本专题从不同角度描述了面向类脑计算的器件及物理特性的进展,反映了此领域的一些现状, 希望对读者了解此前沿课题有所帮助,可以吸引更多学者尤其是年轻学者的关注和加入,为我国在 本领域的蓬勃发展增添新生力量.

(客座编辑: 刘琦 复旦大学)

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Preface to the special topic: Physical electronics for brain-inspired computing

DOI: 10.7498/aps.71.140101

物理学报Acta Physica Sinica





Institute of Physics, CAS

忆阻类脑计算

温新宇 王亚赛 何毓辉 缪向水

Memristive brain-like computing

Wen Xin-Yu Wang Ya-Sai He Yu-Hui Miao Xiang-Shui

引用信息 Citation: Acta Physica Sinica, 71, 140501 (2022) DOI: 10.7498/aps.71.20220666 在线阅读 View online: https://doi.org/10.7498/aps.71.20220666 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

NbO、忆阻神经元的设计及其在尖峰神经网络中的应用

Design of NbO, memristive neuron and its application in spiking neural networks

物理学报. 2022, 71(11): 110501 https://doi.org/10.7498/aps.71.20220141

N型局部有源忆阻器的神经形态行为

Neuromorphic behaviors of N-type locally-active memristor 物理学报. 2022, 71(5): 050502 https://doi.org/10.7498/aps.71.20212017

氧化物基忆阻型神经突触器件

Oxide-based memristive neuromorphic synaptic devices 物理学报. 2019, 68(16): 168504 https://doi.org/10.7498/aps.68.20191262

基于二维材料MXene的仿神经突触忆阻器的制备和长/短时程突触可塑性的实现

Fabrication of synaptic memristor based on two-dimensional material MXene and realization of both long-term and short-term plasticity

物理学报. 2019, 68(9): 098501 https://doi.org/10.7498/aps.68.20182306

基于水热法制备三氧化钼纳米片的人工突触器件 Artificial synapse based on MoO₃ nanosheets prepared by hydrothermal synthesis

物理学报. 2020, 69(23): 238501 https://doi.org/10.7498/aps.69.20200928

专题: 面向类脑计算的物理电子学

忆阻类脑计算*

温新宇¹⁾²⁾ 王亚赛¹⁾²⁾ 何毓辉^{1)2)†} 缪向水^{1)2)‡}

1) (华中科技大学集成电路学院, 信息存储材料及器件研究所, 武汉 430074)

2) (江城实验室, 武汉 430205)

(2022年4月11日收到; 2022年4月27日收到修改稿)

随着深度学习的高速发展,目前智能算法的飞速更新迭代对硬件算力提出了很高的要求. 受限于摩尔定 律的告竭以及冯·诺伊曼瓶颈, 传统 CMOS 集成无法满足硬件算力提升的迫切需求. 利用新型器件忆阻器构 建神经形态计算系统可以实现存算一体, 拥有极高的并行度和超低功耗的特点, 被认为是解决传统计算机架 构瓶颈的有效途径, 受到了全世界的广泛关注. 本文按照自下而上的顺序,首先综述了主流忆阻器的器件结 构、物理机理,并比较分析了它们的性能特性. 然后, 介绍了近年来忆阻器实现人工神经元和人工突触的进 展,包括具体的电路形式和神经形态功能的模拟. 接着, 综述了无源和有源忆阻阵列的结构形式以及它们在 神经形态计算中的应用, 具体包括基于神经网络的手写数字和人脸识别等. 最后总结了目前忆阻类脑计算从 底层到顶层所遇到的挑战, 并对该领域后续的发展进行了展望.

关键词: 忆阻器, 人工神经元, 人工突触, 神经形态计算 PACS: 5.45.-a, 07.05.Mh, 84.35.+i, 87.18.Sn

DOI: 10.7498/aps.71.20220666

1 引 言

自德国提出"工业 4.0"的战略部署以来,第四 轮工业革命——智能化革命已经席卷全球,人类正 式进入人工智能时代.人工智能时代的发展需要数 据、算法和算力.而随着近年来深度学习的兴起, 深层大规模的深度神经网络模型 (deep neural networks, DNNs)在人工智能识别任务上达到了 非常高的准确性,但是其较高的计算复杂程度也给 硬件算力带来了挑战^[1].一直以来,硬件计算性能 的提升依赖于工艺制程的进步和体系架构的创新. 然而,由于物理因素和热耗散的限制,当前 CMOS 工艺节点的缩小成本大大提高,摩尔定律 (Moore's law)^[2]和 Dennard scaling^[3]达到瓶颈.同时,传统 的冯·诺伊曼架构中,存储器件和计算单元是分开 的,数据在 CPU 和内存单元之间的来回传输消耗 掉大量的能量,限制了硬件能量效率和信号带宽的 继续提高,这也被称为"冯·诺伊曼瓶颈"^[4].针对传 统 CMOS 架构的这些问题, 受生物大脑架构和工 作原理启发产生的类脑神经形态计算是未来可能 的发展方向 6. 哺乳动物大脑的规模巨大, 每个人 类大脑由大约850亿个神经元组成,通过1015个 突触相连接,集成度非常高,但相对于当前专用计 算芯片成千上万瓦特的功耗,出色完成日常智能任 务的人脑功耗仅为 20 W^[6]. 与存算分离的冯·诺依 曼架构体系不同,人脑是高并行度的存算一体系 统,即在数据存储单元上直接进行计算,避免了数 据传输过程中的能量损耗[7]. 然而, 传统晶体管的 设计初衷是为了实现数字逻辑,基于传统 CMOS 器件构建出的神经形态计算芯片需要复杂的外设 电路,无法完全地发挥存算一体的潜力.因此,神经

© 2022 中国物理学会 Chinese Physical Society

^{*} 国家重点研发计划 (批准号: 2019YFB2205100) 和国家自然科学基金 (批准号: 92064012, 61974051, 61874164, 51732003) 资助的课题.

[†] 通信作者. E-mail: heyuhui@hust.edu.cn

[‡] 通信作者. E-mail: miaoxs@hust.edu.cn

形态计算架构需要新型的存储器件,它们在器件物 理和材料方面与传统 CMOS 有根本上的不同,以 实现高通量、节能和局域高效的信息处理^[8].

忆阻器^[9,10]就是这样的一种新型器件. 忆阻器 通常具有简单的双端结构, 总共三层——两个发送 和接收电信号的电极和中间的一层"存储"层. 该器 件的外部结构与电阻器类似, 使得它可以进行低成 本的高密度集成. 与静态电阻器不同的是, 因为忆 阻器独特的物理机理, 在电信号的刺激下器件电阻 可以动态地发生改变, 并且这种变化在电信号撤去 后依然存在, 具有非易失性^[10,11]. 因此, 忆阻器不 仅可以同时支持存储和计算操作, 还可以通过改变 电阻状态实现有效的数据处理^[12]. 目前, 基于忆阻 器件的存算一体系统在世界范围内得到广泛关注, 有效地解决了上述的两大瓶颈.

在人脑的生物神经回路中,神经元和突触在内部分子过程的驱动下进行信息处理,使人脑有着非常高效的记忆和信息处理能力.利用忆阻器的特殊阻变物理机制去近似生物突触和神经元的神经动力学,可以使得忆阻阵列的神经形态计算更加符合生物现实^[13].同时,利用忆阻器构建 crossbar 阵列可以很好地拟合生物系统中的存算一体特性,即神经元整合输入信号,并通过相应的突触(忆阻器)向下游神经元产生输出信号^[14].这种存算一体架构既可以支持传统的机器学习算法,也可以支持更加仿脑的脉冲神经网络(spiking neural networks, SNNs),并且记忆设备的电导调制和保持特性可以用于实现在线或离线学习.

本文的整体结构如图1所示.本文首先根据材 料和开关机制的不同对主流忆阻器进行分类,并对 针对它们的特性进行具体讨论与分析.然后讨论不



图 1 综述各节内容总括

Fig. 1. Schematic organization of each section of the review work.

同的突触和神经元的生物特性的实现,重点是利用 忆阻器件内部动力学进行物理化学机制上的近似. 最后调研忆阻器阵列在神经网络中的应用,并总结 了其未来的发展与挑战.

2 忆阻器模型分类

根据材料和物理机理,可以将忆阻器件分为阻 变随机存储器 (resistive random access memory, RRAM), 相变存储器 (phase-change random access memory, PCRAM), 磁随机存储器 (magnetic random access memory, MRAM)和铁电随机存储器 (ferroelectric random access memory, FeRAM) 四种.为更好地实现存算一体系统,在单个器件层 面上主要是针对以下参数的优化: 可区分状态数 (distinguishable states)、开关速度 (switching speed)、 开关功耗 (switching energy)、开关窗口比 (on-off ratio)、波动性 (cycle-to-cycle variation 与 deviceto-device variation)、保持特性 (retention) 和耐擦 写特性 (endurance) 等. 本文对四种不同类型的忆 阻器结构和机理进行介绍,并对它们各自的优势特 性及发展情况进行分析.由于篇幅原因,本文对有 机材料忆阻器件和光致材料忆阻器件介绍较少,这 两类器件在本领域也有非常丰富的应用,在其他优 秀综述中有着更详细的介绍[15].

2.1 阻变随机存储器 (RRAM)

从物理机理上来讲, RRAM 属于离子迁移型 器件. 这类器件介电材料内部的离子运动可以由外 部刺激, 如电场、温度变化或光脉冲来驱动. 在外 部驱动作用下, 活化离子经历一系列的化学和物理 过程, 最终通过形成/断裂导电丝或改变器件掺杂 的方式来导致器件电导的变化. 去除外部刺激后, 离子大部分在吉布斯自由能或浓度梯度的驱动力 下仍能自发扩散, 保持非易失性^[13]. 除 RRAM 外, 电解液型阻变存储器也属于离子迁移型器件. 电解 液型阻变存储器是三端器件, 利用栅极电压来调控 电解液和沟道间的离子交换, 本文主要讨论的是忆 阻器件, 故不做更多介绍, 具体介绍可参照其他优 秀综述^[16,17].

RRAM 主要包括电化学金属存储器 (electrochemical metallization memories, ECM) 和价变 化存储器 (valence change memory, VCM) 两大类, 根据电化学反应的不同对象来划分,下面具体介绍.

2.1.1 电化学金属存储器 (ECM)

ECM 也被称为导电桥型随机存储器 (conductive bridging RAM, CBRAM), 主要是依靠电驱 动下活泼金属电极 Ag 和 Cu 等的电化学溶解和沉 积完成电阻改变. ECM 的 MIM 结构主要包括:活 泼电极、固体电解质层和惰性对电极. 活泼金属电 极需要对电压比较敏感, 在施加电压下较容易地进 行电化学溶解成为离子并穿过绝缘层导电. 迄今为 止,绝大部分的 ECM 都是使用 Ag, Cu 两种金属 之一来作为活泼电极. 固体电解质层主要决定了器 件的整体性能, 是 ECM 单元的主要研究对象, 也 有一些工作调整惰性电极以提高器件性能^[18].

目前,已经有不少研究通过高分辨率的透射电 子显微镜观察到了金属阳离子在固体电解质层形 成的导电丝,对ECM导电丝机理进行了验证^[19,20]. 但在导电丝的形成和破坏的微观过程、组成和形状 等关键问题上仍存在重大争议^[21,22]. 意大利 Ielmini 团队^[23,24]提出了一种导电丝表面离子受限自扩散 的机制,给出了与爱因斯坦关系相关联的理论计算 公式,解释了纳米级导电丝阻变器件由阻值转变到 阈值转变的原因. Chekol 团队^[25]探究了施加脉冲 的幅值和宽度对导电丝的形成时间和弛豫时间的 影响,分析了 Ag/HfO₂/Pt 基忆阻器导电丝置态过 程中的速率限制因素. 他们发现,通过对置态脉冲 的设置可以控制阻变存储器到选通管的转变,在具 体电路中实现不同的功能,如图 2 所示.中国科学 院刘明团队^[26]制作出两种含不同形式缺陷的石墨



图 2 根据 t_{set}和 t_r顺序排列的 v-ECM 型扩散忆阻器的不同应用领域^[25]

Fig. 2. Different areas of applications of v-ECM-type diffusive memristors arranged according to exemplary requirements in $t_{\rm set}$ and $t_{\rm r}$ ^[25].

烯,一种是纳米孔形式的集中缺陷,另一种是离散 原子尺度的分散缺陷,将制作的石墨烯放入活泼金 属电极处,形成 Ag/DG/SiO₂/Pt 型器件.利用石 墨烯的不渗透性,可以选择性地调控导电丝直径, 最终用该器件同时实现了低工作电流 (≈1 μA)的 存储器和高驱动电流 (≈1 mA) 的选通管.

ECM 型器件的优势在于有非常大的开关比, 主要研究集中在导电桥型选通管和 CBRAM 上. 选通管器件可以用作神经元,或作为高非线性元件 与忆阻器件串联形成阵列.华中科技大学缪向水团 队^[27]构建了一个带小孔结构的 TiW/CuS/GeSe/ Pt 型选通管,达到 10⁹的高开关比、高电流驱动 (600 μA)、超低电流 (100 fA)、极陡的开关斜率 (低于 1.4 mV/dec).他们在后续继续优化材料,使 选通管能容纳更高的电流密度^[28].其他研究中导 电桥型选通管也往往达到至少 10⁵的开关比^[29,30]. CBRAM 的开关比一般能够达到 10³—10⁶,但可擦 写次数往往较低,仅在 10⁴ 数量级^[31,32].

2.1.2 价态变化存储器 (VCM)

VCM 也被称为氧化物阻变随机存储器 (oxide-RRAM),其 MIM 结构相对更为简单,为两层惰性 金属电极中夹杂一个氧化物绝缘层,如图 3(a) 所 示. 不同于基于活性金属的电化学反应的 ECM 器 件, VCM 器件的阻变机理是基于其氧化物层固有 的氧相关缺陷的电化学反应,即在外界条件作用下 氧离子/空位迁移形成导电丝^[33]. 图 3(b) 和图 3(c) 给出了 VCM 的两种阻态切换 I-V特性图. 单极性 开关意味着开关方向只取决于施加电压的幅度,不 取决于施加电压的极性,即置态和重置可以在相同 的极性下发生. 如果单极性开关可以在正电压和负 电压下对称发生,那么也被称为非极性开关.双极 性开关意味着开关方向取决于施加电压的极性.因 此,置态只能在一个极性发生,而重置只能在相反 的极性发生^[34]. 两种开关模式对应不同的物理机 理,导电丝的热熔解模型可以解释单极性[35],而离 子迁移模型可以解释双极性.

首尔大学 Kwon 等^[36]首先进行了 VCM 导电 丝机理的验证. 在单极开关 Pt/TiO₂/Pt器件中, 他们利用透射电子显微镜 (TEM) 观察到了成分 为 Ti₄O₇ 的圆锥柱状纳米晶导电丝. 后续一些工作 相继在 X 射线吸收光谱^[37]、电子-能量损失光谱^[38] 中确定了不同氧化层中产生导电丝的成分. 利用



图 3 (a) Oxide RRAM 的 MIM 结构原理图; (b) 非极性和 (c) 双极性不同原理的阻变 LV特性曲线图^[34]

Fig. 3. (a) Schematic of MIM structure for metal-oxide RRAM, and schematic of metal-oxide memory's *I-V* curves for (b) unipolar and (c) bipolar^[34].

仪器表征验证机理的同时,通过研究电场、温度、 氧空位浓度梯度对 VCM 器件的影响,已经有一些 工作实现了对其动态开关行为的精确建模,并通过 预测实验证明了建模的准确性^[39,40].基于这些建 模,Kim 等^[41]进一步构建了将温度作为二阶状态 变量的忆阻器模型,并实际做出相应的器件完成了人 工突触的突触可塑性实现,将在第3节中具体介绍.

相对于 ECM, VCM 虽然开关比往往较低, 只 能达到 10—10³ 数量级, 但在其他器件特性上拥有 巨大的优势. 不同材料的 VCM 器件可以达到 115 fJ 的开态功耗^[42], 85 ps 的开关速度^[43], 10 年以上的 保持特性^[44]及 10¹² 的可擦写次数^[45], 是一种非常 有潜力的存储器件^[8]. 此外, VCM 中离子漂移、扩 散、热电泳以及缺陷的产生和重组过程所提供的离 子动力学与生物神经形态动力学近似^[13], 故 VCM 非常适合作为神经形态计算的器件实现.

2.2 相变型器件

相变型器件是指受到外部刺激局部结构 (物理 结构、电子结构) 发生改变以导致阻变的器件, 主

要包括相变随机存储器 (PCRAM) 和金属绝缘体 转换器件 (MIT 型).

2.2.1 相变随机存储器 (PCRAM)

PCRAM 主要由两个金属电极和一层夹在其 中的相变材料组成.相变材料的非晶状态是一种热 不稳定的状态,在温度较高但低于熔化温度时可以 快速结晶. 通过给电极施加如图 4(a) 所示的窄而 高或宽而低的脉冲进行重置和置态过程,相变功能 层在晶态和非晶态间转换,产生电阻率和折射率的 变化,从而进行"0"与"1"的切换.相变材料是 PCRAM 性能的主要决定因素, 如图 4(b) 所示, 目 前的研究热点主要集中在以 GeSbTe 为基础的两 个系列上: 第一个是沿着 GeTe-Sb₂Te₃线的合金, 它们重结晶的效果好; 第二个是在 Sb₂Te 上进行 掺杂, 通常是 Ag₅In₅Sb₆₀Te₃₀ (AIST), 常用于光存 储中[46]. 相变随机存储器的切换速度和数据保留 能力都与结晶动力学直接相关, Salinga 等^[47]研究 发现这两个系列的相变材料结晶方式不同,并用 TEM 观察到了相应的现象. GST 系列为成核驱动



图 4 (a) PCRAM 相变的脉冲操作示意图^[46]; (b) 目前主流合金材料的三元 Ge:Sb:Te 相图^[48]

Fig. 4. (a) Schematic of pulse operation leading to PCRAM phase transition^[46]; (b) the most popular alloy line in ternary Ge:Sb:Te phase diagram^[48].

型, 成核过程中凝结核随机形成, 后围绕凝结核生 长进行结晶, 成核速度更快. AIST 系列为生长驱 动型, 成核过程中由晶体基质包围的非晶区的结晶 在晶体-非晶界面上快速进行, 在生长过程中的短 时间尺度内没有稳定的小尺寸晶核形成, 成核速度 较慢^[47].

PCRAM 作为忆阻器存在以下优势:因为相变 材料晶态和非晶态的电阻率差异很大,且晶态的电 阻较高,使得其功耗较低.而通过提高 PCRAM 的 晶化速度、提高结晶温度等手段,可以使 PCRAM 拥有高开关速度和良好的数据保持能力(正常工作 温度下保持 10 年).与此同时,PCRAM 具有高擦 写次数(至少能够循环 10⁶ 个读写周期)、低循环间 变化等特点^[49].

但是, PCRAM 也存在一些应用上的问题, 一 是 PCRAM 非晶态下会因为弛豫效应产生电阻漂 移的现象. 电阻漂移的产生使得高阻态的器件电阻 逐渐降低,影响其保持特性. Raty 等⁵⁰提出了一 个基于 Peierls 畸变的结构模型,其表现与 GST $(Ge_xSb_yTe_{1+x+y})$ 所有的实验数据一样. 这个模型 强调了 PCRAM 非晶相的成键特殊性, 从物理上 解释了其电阻漂移的具体原因. Ding 等^[51]提出一 种新型的相变存储器架构——变异质结构 (PCH), 由交替堆叠的相变和约束纳米层组成,成功抑制噪 声和漂移,为高性能神经启发计算提供可靠的迭代 重置和置态操作. 另一个问题是 PCRAM 的成核 时间具有随机性,结晶时间会有一个大区间的波 动,这会对 PCRAM的开关速度造成影响^{52]}. 成核 随机性与以下事实有关: 在室温下非晶态 GST 中 的结晶前驱体,即 Abab 正方形和立方体,在温度 升高的情况下不稳定^[53]. Rao 等^[54] 提出一种通过引 入合金加速晶体结晶的方法. 他们设计的碲化钪 (Sc_{0.2}Sb₂Te₃)化合物写入速度仅为 700 ps, 且无需 在大型传统 PCRAM设备中进行预编程. 这种方式 引入了几何匹配且坚固的碲化钪 (ScTe) 化学键, 使无定形状态下晶体前驱体更加稳定,降低了成核 的随机性,从而大大缩短了结晶时间.

2.2.2 金属绝缘体转换器件 (MIT 型)

金属绝缘体转换器件 (MIT 型,即 metal-insulator Transition) 是一种利用 MOTT 绝缘体相变 的易失性忆阻器,通常结构为金属电极夹在较薄的 一层 MOTT 材料两边. 1949 年, Mott^[55] 对过渡金 属氧化物中的 NiO 在常温下为绝缘体而与理论预 测不符的问题做出了解释,提出体系中轨道最外层 电子之间有强库仑相互作用,形成的能隙大于电子 的迁移能,使得常温下 NiO 中的电子无法迁移,为 绝缘体. 1963年, Hubbard^[56]在此基础上, 进一步 将电子相互作用引入紧束缚模型中,认为电子间的 强相互作用会使得能级分层,根据能量最低原理, 电子首先填充到下能带,使得费米面无可填充电 子,形成绝缘体,称其为 Mott-Hubbard 绝缘体, 简称 Mott 绝缘体. 对于 Mott绝缘体材料, 施加外 界扰动,如加热、外加电场、光照或引入外界应力 等达到一定的阈值时,其内部电子迁移能增大或者 电子库仑相互作用减小,使其转变为金属,形成金属 绝缘体转变 (MIT). MIT 转变具有迟滞效应, 且迟 滞窗口内材料的阻值不发生很大变化[57]. 目前, 主要 研究的 Mott 绝缘体材料包括 VO2^[58-60], V2O3^[61], NbO_r^[62] 等.

由于 MIT 型材料为易失性忆阻器, 无法用作 存储器,通常可以作为神经元,且发放脉冲后会自 动复位,这将在第3部分重点讲解.同时,还能将 其用作阈值转换开关器件,即选通管,用于阵列中 器件单元的选通,以控制阵列中的 sneak-path,具 体在第4部分介绍.对于 MIT 型选通管,目前对 VO2的研究较多. Zhou 等[58] 提出了一种 Pt/(211 向) VO₂/ITO 的结构, 可以达到 23 ns 的开关速 度, 109个循环的擦写次数, 143的开关比, 且器件 间差异小.可以看出, VO,选通管的电致相变开关 比小,原理存在争议,这也是目前该器件的最大问 题. Shi和 Chen^[59] 通过相场模型证明了电流可以 通过电子诱导的电子相关衰减等温地驱动 MIT 相 变,给后续电致相变的研究带来重要参考.为更好 地在阵列中集成,提高器件开关比以及加强其与传 统 CMOS 的工艺兼容度是目前的 MIT 型器件下 一阶段的研究目标.

2.3 磁随机存储器 (MRAM)

MRAM 是一种目前已经广泛商用的非易失存储器.其中,自旋转移扭矩磁随机存储器 (spintransfer torque-MRAM, STT-MRAM)目前技术 最为成熟,其基本单元为 MTJ (magnetic tunnel junctions).如图 5(a) 所示, MRAM 一般由两个金 属电极和中间至上而下的自由层 (free layer)、隧 穿层 (通常为 MgO)和固定层 (fixed layer) 组成.



图 5 (a) STT-MRAM和 SOT-MRAM的单元结构图^[63]; (b) STT-MRAM的置态和重置操作原理图^[64] Fig. 5. (a) Bit-cell for STT-MRAM and SOT-MRAM^[63]; (b) schematic illustration of the STT switching mechanism to achieve 1 and 0 states^[64].

其中,自由层和固定层都是铁磁性金属,隧穿层为 非常薄的非铁磁性绝缘层,方便电子隧穿^[63]. MTJ 的"0""1"态的产生依赖于隧道磁阻效应 TMR (tunnel magnetoresistance),即自由层与固定层的 磁化方向为反平行和平行时,电子的自旋依赖性导 致其隧穿的概率不同,磁阻大小不同.如图 5(b)所 示,STT-MRAM 器件的编码原理是通过两端施加 不同极性的电压,使得穿过固定层或由其反射的极 化电子通过自由层时施加自旋扭矩从而改变它的 磁化极性^[64].

STT-MRAM 的优势在于低于 10 ns 的高切换 速度,并且具有 5 × 10¹⁴ 的高耐擦写次数,同时具 有很好的保持特性,但其 TMR 通常只能达到 100%—200%^[65-67]. STT-MRAM的开关比太小, 限制了它的集成度. 针对这个问题, 有一些研究通 过铁磁性材料的变换对 TMR 进行了提高, 但也仅 仅使其超过了 200%^[68,69]. Jain 等^[70] 基于 STT-MRAM 构建了一个存内计算的阵列,可以执行基 本的布尔逻辑、算术和复杂矢量运算,一定程度上 解决了 STT-MRAM 的集成问题. Jung 等^[71] 提出 了另一种将器件电阻串联求和以形成更多态的结 构, 解决了低电阻的问题, 制作了一个 64 × 64 的 crossbar 阵列. 另一个问题是传统的 STT-MRAM 由于读取电流总是会穿过超薄 MgO 隧道屏障,加 速其老化,它的可重复循环性不够好,且编程电流 较大.针对这个问题,自旋轨道扭矩磁随机存储器 (spin-orbit torque-MRAM, SOT-MRAM) 被提 出,它通过增加一块重金属电极使器件变为3端, 如图图 5(a) 右所示. 通过这个方式, 编程过程不需 经过 MgO 屏障,同时解决了老化和能耗问题^[72]. 尽管 SOT-MRAM 性能优异,但它的工作机制还 处于讨论之中,是未来 MRAM 的重要发展方向^[73].

2.4 铁电随机存储器 (FeRAM)

FeRAM 也可以作为一种忆阻器件, 其主要单 元是铁电隧穿结 (ferroelectric tunnel junctions, FTJ). FTJ 通常由两个金属电极中间夹杂一个几 纳米厚的铁电材料势垒层组成, 如图 6(a) 所示. 纳 米级的铁电膜层可以自发极化, 通过施加外部电场 使得其在至少两个稳定取向之间切换. 表现在晶体 结构上, 以 BaTiO₃ (BTO 族) 为例, 极化取向即 为 Ti 原子在晶体中心对称位置的偏移程度. FTJ 的特点是隧穿电阻 (tunnel electroresistance, TER)



图 6 (a) FTJ 结构及铁电层 BaTiO₃ 晶体结构示意图; TER 效应: (b) *P-E* 特性曲线; (c) *I-V* 特性曲线^[77]

Fig. 6. (a) FTJ structure and schematic diagram of the ferroelectric layer BaTiO₃ crystal structure; TER effect: (b) *P*-E characteristic curve; (c) *I*-*V* characteristic curve^[77].

效应,即在外部电压使得铁电层极化反转的同时, 隧道阻态发生变化,达到 10⁴的开关比^[74-76].经总 结,铁电材料极化翻转时铁电势垒层高度或宽度发 生变化,从而使得 FTJ 发生阻态转变,如图 6(b) 和图 6(c) 所示,红线代表高阻态,蓝线代表低阻 态,由于剩余极化的作用,FTJ 具有非易失性^[77].

作为忆阻器, FeRAM 的优势在于纳秒级别的 读写速度, 且切换能耗极低 (pJ/bit 级别), 同时存 储密度很高^[78].使用 Hf_{0.5}Zr_{0.5}O₂(HZO) 作为铁电材 料层, FeRAM 能拥有比较好的CMOS 兼容性^[79,80]. 应用 HZO 作为铁电隧穿层, 可以达到常温下 10 年 以上的保持时间特性, 且可重复循环操作 10¹¹ 次, 但开关比只有 10 倍左右^[81–83].应用其他铁电材料 可以得到不同的特性, FTJ 也可以达到非常高的开 关比, Wu 等^[84]使用层状铜铟硫代磷酸 (CuInP₂S₆) 作为铁电势垒, 将石墨烯和铬作为不对称接触的铁 电隧道结, 得到了 10⁷ 的 TER变化. 通过构建异质 结构 Pt/BaTiO₃/Nb:SrTiO₃, Wen 等^[76] 得到了 10⁴ 的开关比, 同时达到了 10 年的保存时间, 但可重 复循环操作次数较低.

2.5 总结对比

本节根据物理机理的不同对忆阻器进行了具体的分类,并进行了细致的介绍.最后,在表1中列出了近年来每种忆阻器接近最好的单项特性,包括切换速度、开关比、保持特性、耐擦写次数和开关功耗等参数指标,方便读者有一个直观的对比.

表	1 不同类型忆阻器件参数指标比较
Table 1.	Comparison of parameter specifications of
different	types of memristors.

	51				
参数指标		RRAM	PCRAM	MRAM	FeRAM
切换速度/ps		$85^{[43]}$	$700^{[54]}$	$210^{[85]}$	$4000^{[86]}$
	开关比	$10^{7[87]}$	$10^{6[88]}$	$< \! 10^{[67, 89]}$	$10^{7[84]}$
	室温保持特性/a	$> 1000^{[44]}$	$> 10^{[90]}$	$> 10^{[91]}$	$> 10^{[76,81]}$
	耐擦写次数	$10^{12[45]}$	$10^{11[92]}$	$5 imes 10^{14 [66]}$	$10^{11[82]}$
	开关功耗/fJ	$115^{[42]}$	$50^{[93]}$	$10^{[94]}$	$10^{[95]}$

3 忆阻神经形态器件

近年来,神经形态器件对生物神经过程模拟已 经取得了突破性进展.本节首先简要回顾大脑的学 习机制,分别从人工突触和神经元两方面给出对忆 阻器件的特性要求,然后重点介绍目前忆阻器件仿 脑的物理机理和实现形式.

3.1 忆阻器作人工突触

当前的神经科学研究得出,改变神经元之间突 触连接的强度是记忆被编码和存储在中枢神经系 统中的机制^[96].获取经验,无论是上课学习,遇到 压力事件还是精神发生波动,都通过改变特定神经 回路的活动和组织来影响大脑.经验产生的神经活 动改变大脑功能的主要机制是突触传递的修饰,这 就是突触可塑性.一个多世纪以来,学术界认为突 触可塑性在大脑将瞬时经验纳入持久记忆的功能 中发挥核心作用.突触传输可以通过神经元活动增 强或抑制,并且突触变化持续时间的跨度很大,范 围从毫秒到小时、几天,甚至可能更长.此外,哺乳 动物大脑中几乎所有的兴奋性突触都能表现出许 多不同形式的突触可塑性^[97].使用器件的物理机 理去仿照各种形式的突触可塑性,有利于从硬件上 实现人工突触,从而实现类脑计算,如图 7(a)所示.

突触器件与新出现的非易失性忆阻器有许多 共同的特性,因为这两种器件都需要进行编程、读 取以及良好的保持特性.同时,新兴的非易失忆阻 器 (第2节介绍) 作为突触设备, 还具备许多其他 优秀特性,如低编程能量、高切换速度、良好的可 扩展性等[98]. 作为突触器件, 对忆阻器性能有一些 特别的要求,如人工突触器件在使用过程中往往需 要更多的状态 (8 bit 或 256 个状态) 以达到学习所 需精度等.突触器件的可擦写次数和保持特性也更 加依赖于应用. 如一个在线训练并实时进行权重更 新的神经网络需要良好的可擦写次数,对保留特性 的要求不那么严格. 另一方面, 已通过离线学习得 到突触权重的神经网络则需要更好地保持特性,而 较少强调可擦写次数^[99].为了用器件生物性地实 现一个神经形态系统,忠实地模拟突触的功能是至 关重要的,即突触的功效和可塑性.本节将重点讨 论忆阻器的动态行为,并表明忆阻器中的物理机理 可以与生物突触中的神经机制类比,利用现实器件 实现不同的突触功能.

3.1.1 长时程突触可塑性 (LTP)

通过增加或减少突触前神经元释放的神经递 质量或通过增加或减少突触后神经元存在的 AMPA 受体的量,可以持续改变生物突触的作用能力,这 被称为突触长期可塑性 (long-term synaptic plasticity). 具体而言,突触作用效能的增加和减少分别 称为长期增强 (long term potentiation, LTP) 和



图 7 LTP 的忆阻器实现 (a) 使用忆阻器作为神经元之间的突触的概念示意图; (b) 忆阻器对编程脉冲的响应^[101]; (c) Pt/LiAlO_x/ TiN 忆阻器在不同初始电导状态下的电导调制性能 textsuperscript^[102]; (d) DW-MTJ 人工突触的侧视和俯视图; (e) DW-SOT 和 DW-STT 器件的测试更新线性度和对称性^[103]

Fig. 7. Memristor implementation of LTP: (a) Schematic illustration of the concept of using memristors as synapses between neurons; (b) memristor response to programming pulses^[101]; (c) conductance modulation performance at different initial conductance states of $Pt/LiAlO_x/TiN$ memristor^[102]; (d) side and top profile of DW-MTJ artificial synapse; (e) update linearity and symmetry with experimental data from DW-SOT and DW-STT devices^[103].

长期抑制 (long-term depression, LTD)^[100]. 突触 长期可塑性往往被认为是人脑进行学习和记忆的 关键^[96]. 表现在忆阻器件的实现上,就是在置态和 重置过程中,电导调制应该是非易失性和渐进式 的,并且需要有足够多的态. Jo等^[101]首先设计了 一个多层结构的忆阻器,如图 7(a)所示,其功能层 由 Ag/Si 共同溅射,以一定的 Ag/Si 梯度比形成 一个富 Ag (高电导率) 区域和一个贫 Ag (低电导 率) 区域. 溅射形成的功能层中, Ag+稳定存在,通 过施加置态和重置电压使 Ag+在富 Ag 区和贫 Ag 区迁移以改变器件电导,以形成 LTP/LTD 过程, 如图 7(b)所示. 但由于是推动 Ag+运动,电压脉冲 的幅值在 3 V 左右,且持续时间为 300 μs, 功耗和 速度都需要继续提升.

在后续的突触发展中,一些非理想效应,包括 有限的电导态、不对称因子 (asymmetry factor, AF) 和电导变化的非线度 (nonlinearity, NL),限制了 忆阻网络水平上记忆突触的进一步发展^[98].Fu 等^[102] 构建了 Pt/LiAlO_x/TiN 忆阻器,通过精细地调整 初始电导状态,使得导电丝尖端和电极之间的肖特 基势垒高度 (SBH) 降低,如图 7(c) 所示.最终得 到了 120 级高度线性的连续可调电导状态,另外, 通过计算,AF = 0.04,NL = 0.53/-0.01,且保持 特性良好.Liu 等^[103]提出了一种含磁畴壁的 MTJ 结构模型,通过电压推动磁畴壁改变电导值,NL = 0.07/-0.15,且循环间变化值低至 0.77%,如图 7(d) 和图 7(e)所示.Liu 等^[104]通过插入一个 2 nm 的 BiFeO₃层,构造了一个 Pt/BiFeO₃/HfO₂/TiN 型 VCM 忆阻器,实现了开关比达 10⁴ 的大存储窗口 和 10⁸ 个周期的高耐擦写次数.基于分子设计方法 对忆阻结构优化也可以得到更稳定的 LTP 响应. Wang 等^[105]采用金属卟啉 MTPP (metalloporphyrin)分子作为功能层,其配位位点可以调节氧 迁移,使制备的器件表现出平滑、逐渐变化的持久 记忆响应.

3.1.2 短时程突触可塑性 (STP)

STP 指几十毫秒到分钟的时间尺度上暂时的 突触传递效率变化,随后恢复到初始状态,在大脑 中的信息编码和处理中发挥着重要作用^[106].具体 而言,它包含以下几个具体机制:短时程增强 (short-term potentiation, STP)/短时程抑制 (short-term

depression, STD), 指突触权重暂时性地增强与抑制; 双脉冲易化 (paired-pulse facilitation, PPF)/双脉冲抑制 (paired-pulse depression, PPD), PPF 指数百毫秒尺度上两个连续的脉冲刺激, 第二个脉冲的刺激强度相对第一个脉冲增强, 且间隔时间越短, 增强效果越大, PPD则相反; 强直性后增强 (post-tetanic potentiation, PTP)/强直性后抑制 (post-tetanic depression, PTD), PTP 和 PTD 与 PPF 类似, 但是作用时间尺度更长, 可以到1 min 左右, 适用于一连串的脉冲^[106]. 长期可塑性被认为与学习和记忆功能有关, 而短期可塑性往往与生物系统中时空信息处理的关键计算功能联系紧密^[107].

STP 在器件中的实现, 主要原理是在电压条

件下建立一个不稳定的电导态,经历一个弛豫过程 后恢复原有稳定状态.如第2节所述,ECM型 RRAM 具备类似的特性,即随着电压增大建立不 稳定的导电丝增加电导,随后导电丝断裂恢复原有 的低导态.Ohno等^[108]首先制备了一种 Ag/Ag₂S/ 电极的 ECM 型忆阻器,通过间歇性地输入脉冲, 形成不稳定的 Ag 导电丝,电导自发衰减,从而实 现了 STP 特征.同时,Chang等^[109]提出了一种 Pt/WO_x/W/Si 结构的 VCM 型忆阻器,利用氧空 位在底电极处的分布不同改变电导值实现了 STP, 如图 8(a) 和图 8(b)所示.并且实验证明,通过调 整该器件脉冲输入的时间间隔和总数,可以将 STP 转化为 LTP,如图 8(c)所示.Yang 等^[110]也



图 8 STP 的忆阻器实现 (a)测试过程中通过忆阻器的连续电流变化;(b)图 (a)中矩形区域的特写图;(c)电导转换速率与刺激速率的关系图,在不同脉冲间隔条件下,每次刺激脉冲后通过忆阻器的电流^[109];(d) ITO/PVPy-Au NPs/Al RRAM 器件的结构和 Au NPs 的 HRTEM 图像;在 (e) 2 次和 (f)10 次不同脉冲间隔的脉冲之间的器件电流变化图^[11]

Fig. 8. Memristor implementation of STP: (a) The corresponding current through the memristor data recorded continuously throughout the test. (b) A close-up view of the rectangular area in panel (a). (c) Dependence of the transition efficiency on stimulation rate. Current through the memristor recorded after each stimulation pulse, at different pulse interval conditions^[109]. (d) The structure of ITO/PVPy–Au NPs/Al RRAM device and the HRTEM image of Au NPs. Current change between (e) two pulses and after (f) 10 pulses with different pulse intervals^[111].

利用 WO_{3-x} 功能层得到类似的结论. Zhang 等^[111] 提出一种基于固体电解质聚乙烯吡咯烷酮 (PVPy)-金纳米颗粒 (nanoparticle, NP) 混合的忆阻器, 如 图 8(d) 所示. 在施加电压下 PVPy 宽带隙的势垒 高度发生改变, 正负压下发生复合膜中电子的捕获 和复位, 从而发生电导的改变, 最终实现图 8(e) 和 图 8(f) 所示的类 PPF, PTP 效应. Shi等^[112]使用 垂直金属/h-BN/金属细胞制作了电子突触, 这些 细胞同时显示易失性和非易失性阻变特征, 模拟了 几种 STP 和 LTP 突触行为, 包括 PPF, PPD 和 STDP. 该忆阻器弛豫特性良好, 在 500 个循环中 变化很小, 且在易失性 STP 过程中, 突触的静态和 动态功耗分别仅为 0.1 fW 和 600 pW.

3.1.3 脉冲频率依赖可塑性 (SRDP)

人类神经网络中的信息传递模式与动作电位 的平均发射率有关,神经元之间的突触权重可以通 过放电频率来改变,这就是脉冲频率依赖可塑性 (spike-rate-dependent plasticity, SRDP). SRDP 通常依赖于 Bienenstock-Cooper-Munro (BCM) 学习规则.根据 BCM 规则,频率大于阈值频率θm 的突触前脉冲会导致 LTP, 而低于θm 的突触前脉 冲会导致 LTP, 如图 9(a)所示^[113-115].其中θm是 一个历史性的活动值,与之前突触的活动相关.部 分研究提出了阈值漂移效应,其中阈值频率根据学 习经验而变化,从而实现依赖于历史的突触适应. 这个过程将使大脑处于一个有效的动态平衡状态, 在大脑的认知行为中发挥重要作用^[116,117].

正如前面所述,基于 WO_x 的忆阻器件依赖底

电极氧空位的运动,可以响应脉冲频率的变化给出 不同的电导调制效果^[109,110].根据该特性,Du等^[118] 制作了基于 WO_x的忆阻器件并提出了相应的二阶 忆阻器模型,并通过实际的测试证明了该器件的 类 SRDP 特性与忆阻模型的可靠性.在测试实验 中,他们发放了一系列的脉冲序列,每个脉冲序列 由 5 个相同的编程脉冲 (1 V,1 ms)组成,并记录 了忆阻器电导的变化,如图 9(b)和图 9(c)所示. 在步骤 1 中,施加了具有 200 Hz 刺激频率的第一 个脉冲序列,发现通过忆阻器的电流增加.随后, 在步骤 2 中,一个 10 Hz 的脉冲序列导致忆阻器电 流下降.另一方面,在步骤 3 中的 1 Hz 脉冲序列 之后,步骤 4 中同样的 10 Hz 脉冲序列反而导致了 忆阻器电流的增加. 10 Hz 脉冲序列的不同响应表 现了阈值频率 θ_m 的活动依赖性.

3.1.4 脉冲时间依赖可塑性 (STDP)

1949年, Hebb^[119]提出"fire together, wire together",即重复和持续共同兴奋的互联神经元应该 增加它们之间的连接强度,以作为存储记忆的手 段.近年来,细胞学习中出现了一个新概念,强调时 间顺序而不是频率,这种新的学习范式被称为脉冲 时间依赖可塑性 (spike-timing-dependent plasticity, STDP),它结合了优雅的简单性、生物学的合 理性和计算能力.一般来说,如果突触前脉冲在突 触后脉冲前几毫秒或更短时间内发生,则引发 LTP,而时间顺序相反则引发 LTD,且间隔时间越 短,突触变化强度越大,如图 10(a)实线所示^[120]. 在一些皮质突触上,LTD 的时间窗 (灰色虚线) 被



图 9 SRDP 的忆阻器实现 (a) 具有 SRDP 特性的生物突触示意图^[116]; (b) WO_x 基忆阻器对不同频率下连续编程脉冲序列 (1 V, 1 ms, 蓝线) 的响应; (c) 在经历了不同程度的激活后, 忆阻器电流随刺激频率的变化. 由 5 个不同频率的脉冲 (1.2 V, 1 ms) 组成的不同频率的脉冲序列对忆阻器进行编程^[118]

Fig. 9. Memristor implementation of SRDP: (a) Schematic diagram of a biological synapse with SRDP activities^[116]; (b) WO_x based memristor response to consecutive programming pulse trains (1 V, 1 ms, blue lines) at different frequencies; (c) memristor current change as a function of the stimulation frequency after the memristor has been experienced to different levels of activities. Pulse trains consisting of five pulses (1.2 V, 1 ms) with different repetition frequencies were used to program the memristor^[118]. 扩展^[121,122]. 这些时间窗口也通常依赖于神经元活动, LTP 在低频脉冲的条件下可能会消失 (灰色连续线)^[122,123].

为了实现上述 STDP 功能, 突触器件通常需要 满足电导逐渐变化和对单个峰值快速响应的要求^[125]. 对于器件电路, STDP 可以通过构造不同的脉冲使 其重叠达到阈值以发生电导调制的方式来实现. Yu 等^[124]首先构建了一个 TiN/HfO_x/AlO_x/Pt 型 RRAM, 使用时分复用 (time-division multiplexing, TDM) 方法来设计脉冲的形状, 以实现 STDP 行 为. 他们从 RRAM 的两端分别输入图 10(b) 所示 的突触前脉冲和突触后脉冲, 相叠加后的脉冲达到 阈值,产生图 10(c) 所示的 LTP 和LTD. 也有一些研究利用类似的方式用其他器件,如 PCRAM^[126], MRAM^[44] 实现了 STDP. 但这种方式需要产生特定的脉冲,会带来一定的外围电路硬件代价.

在神经生物学中,脉冲之间的相对时间信息不 是依赖于保持时间信息的外部因素,而是自然嵌入 的,如通过 Ca²⁺水平的自然衰减,神经元对突触响 应强度变化,提供了一种内部的时间机制^[127].很多 研究通过构建二阶漂移 RRAM 忆阻器模型,利用 热耗散^[41]或者氧空位迁移率的改变^[118]来近似化 学突触中 Ca²⁺的动力学,以此体现脉冲的时间影 响以实现 STDP.如图 10(d) 所示,从忆阻器同一



图 10 STDP 的忆阻器实现 (a) STDP 特性展示图^[120]; (b) 利用 TDM 和脉冲幅度调制的 STDP 实现方案, 突触前脉冲振幅分 别为–1.4, 1, 0.9, 0.8, 0.7 和 0.6 V, 突触后脉冲振幅分别为–1, 1.4, 1.3, 1.2, 1.1 和 1 V; (c) 利用图 (b) 中的方法实测的器件 STDP 曲线^[124]; (d) 忆阻权重的变化与突触前后脉冲相对时间的关系, $\Delta t = t_{\text{post}} - t_{\text{pre}}$; (e) 脉冲相对时间影响忆阻器的原理示意图^[118]

Fig. 10. Memristor implementation of STDP: (a) Defining spike-timing-dependent plasticity^[120]; (b) STDP realization schemes developed with TDM and pulse amplitude modulation. The pulse amplitudes for the prespike are -1.4, 1, 0.9, 0.8, 0.7, and 0.6 V, consecutively, and for the postspike, they are -1, 1.4, 1.3, 1.2, 1.1, and 1 V, consecutively. (c) Measured STDP curve of the memristors utilizing method described in panel (b)^[124]. (d) Memristor weight change as a function of the relative timing between the pre- and postsynaptic pulses, $\Delta t = t_{\text{post}} - t_{\text{pre-}}$ (e) Simulation results illustrating how relative timing of the pulses affects memristor weight^[118].

端输入极性相反的脉冲以模拟突触前脉冲和突触 后脉冲,前一个脉冲作用下会产生氧空位迁移率的 积累变化 (参数 wm 变化), 使得后一个脉冲的作用 效果更强,从而产生如图 10(e) 所示的 STDP效应. Yan 等^[128]提出了一种 Ag/Ga₂O₃/NQDs/Pt型 RRAM (NQD: networked QDs). 利用自组装的硫 化铅 (PbS) 量子点 (QDs) 的有序排列, 可以有效 地引导导电灯丝的生长方向,提高 RRAM 开关参 数的均匀性.由于量子点的快速响应能力,该突触 器件 STDP 的响应速度得到大大提升,达到纳秒 级. Wang 等^[20] 制作了一种基于 SiO₂N₂: Ag 的忆 阻器,并构建了 Ag+扩散的 RRAM 模型,实现了 STDP. 通过高分辨率透射显微镜和纳米颗粒的动 力学方程推导, Wang 等首次揭示了利用 Ag+扩散 忆阻器的阈值开关和弛豫特性,提供了忆阻器模拟 突触的理论基础.

3.2 忆阻器作人工神经元

由细胞体、轴突和树突组成的神经元是在人体 中传递生物信号的基本结构和功能单元^[129],如 图 11(a)所示.一个神经元通过树突接收来自前神 经元的信号,然后通过轴突将它们传递到后神经元. 神经元的细胞体根据电位信号的兴奋性和抑制性 决定其电反应(即离子通道的开启/关闭).图 11(b) 给出了具有兴奋或抑制电位的神经元的膜电位.膜 电位大于阈值时,离子通道打开,产生一个动作电位 (脉冲),通过轴突传递后,向外释放离子,回到初始 状态(静息态).模电位低于阈值电位,神经元不产 生动作电位,信号电荷逃逸,回到静息状态^[100,130,131]. 模拟生物神经元的这些一系列行为是人工神经元 实施的关键因素.

目前已经有各种模型来解释神经元的行为和

实现人工神经元,具体有 H-H (Hodgkin-Huxley), Izhikevich,漏电积分点火 (leaky integrate-and-fire, LIF)和脉冲响应 (spike response model, SRM)模 型等^[132].其中,LIF 模型在脉冲神经网络中被广泛 地采用,它极大简化了动作电位过程,但保留了实 际神经元膜电位的泄露、积累以及阈值激发这3个 关键特征.该模型主要针对阈下电位的变化规律进 行描述,公式如下^[133]:

$$\tau_{\rm m} \frac{\mathrm{d}V}{\mathrm{d}t} = V_{\rm rest} - mV + R_{\rm m}I,$$

其中, *τ*_m表示膜时间常数, *V*_{rest}表示静息电位, *R*_m和 *I*分别表示细胞膜的阻抗与输入电流. LIF 模型因 为其简洁的数学表达可以保证较低的实现代价^[132]. 目前, 主要从忆阻阻变器件和阈值转变器件两方面 进行人工神经元尤其是 LIF 神经元的实现.

3.2.1 非易失性忆阻器实现人工神经元

IBM 实验室^[134]在 2016年首先提出了基于硫 系化合物的相变材料创建人工神经元的方式.在这 个人工神经元中,膜电位由纳米级相变器件的相 位结构表示.利用可逆非晶-晶体相变的物理学, 作者证明了突触后电位的时间积分可以在一个纳 秒的时间尺度上实现.除了相变忆阻器,也有一些 利用 RRAM 实现人工神经元的研究.Mehonic 和 Kenyon^[135]利用 SiO₂型 RRAM 建立了一个 LIF 模型,其电路和脉冲发放情况如图 12(a),(b)所 示.频繁施加脉冲会使得电容电压慢慢升高,直到 到达阈值使得忆阻器被置态,发放出脉冲.Lashkare 等^[136]制造了一个 PCMO RRAM 器件作为神经元, 将 70 nm 厚的 PCMO 层插入到 Ti 电极和 W 电 极之间,如图 12(c) 所示.在给器件施加正电压时,



图 11 (a) 生物神经元的结构; (b) 基于兴奋性和抑制性电位的神经元膜电位变化[130]

Fig. 11. (a) Structure of the biological neuron; (b) membrane potential change of the neuron depending on the excitatory and inhibitory potentials^[130].

器件电导缓慢增加,直到达到置态电压后器件电导 迅速上升以产生一个电流脉冲,随后对其施加一个 重置电压使其回到初始态,构成 IF 神经元,如 图 12(d) 和图 12(e) 所示.

如上面所述,利用非易失性阻变器件实现人工 神经元,往往需要器件在达到低阻态发放脉冲后重 新施加一个重置脉冲使器件返回高阻态.这样的操 作增加了人工神经元电路的复杂度,且这种方式每 次产生脉冲的形状不同,不利于脉冲神经网络的学 习.故更多的研究集中在易失性忆阻器对 LIF 模 型的实现上.

3.2.2 易失性忆阻器实现人工神经元

这里易失性忆阻器包含易失性高阶忆阻器和 阈值转换忆阻器两种. 阈值转换忆阻器往往又被称 为选通管 (selector), 是一种易失性器件, 被置态的 选通管两端电压在小于保持电压 Vhold 后会直接被 重置.目前主流的选通管依据机理分为三种,除了 第2节中提到的 CBTS 器件和 MIT 器件外, 还有 OTS (ovonic threshold switching) 器件, 在文献 [137] 中有详细的介绍. HP 实验室^[138]在 2013 年首先提 出了将 MIT 选通管作为神经元的模型, 并展示了 两个基于 NbO2 MIT 器件与电容并联的神经元电 路. 该神经元电路能够发放相同的脉冲, 且发放脉 冲与信号增益关联,可以依据输入调整脉冲发放频 率. 之后, Zhang 等^[139] 提出一种基于 Ag/SiO₂/Au 阈 值转变器件的积分-点火神经元模型,如图 13(a)— 图 13(c) 所示, 实现了动作电位的脉冲一致性、脉 冲阈值驱动性、不应期和输入幅度调制频率响应等 生物神经元特征. 该电路主要通过 Rs 的阻值设计 使得 TSM 高阻态时, 电容充电速度远大于放电速 度,形成积分过程.而 TSM 达到阈值变为低阻态



图 12 非易失器件实现神经元 (a)漏电积分点火神经元的模型展示图; (b)输入间隔 640 ms的兴奋性脉冲序列时得到的输出 电流图^[135]; (c) PCMO RRAM 的器件结构图; (d) 施加-2.3 V 置态电压时显示出的 3 个不同阶段的瞬时电流值; (e) 应用预设脉 冲序列的瞬态实验电流值^[136]

Fig. 12. Neurons implemented by nonvolatile Devices: (a) Basic representation of leaky integrate-and-fire neuronal model; (b) the output current measured after excitatory input pulse with the time separated of 640 ms^[135]; (c) device schematic of PCMO RRAM; (d) SET current transient at -2.3 V showing 3 regions of operation; (e) experimental Current transient for the applied sequence of SET pulses^[136].

时,电容快速放电,发放脉冲,形成点火过程.该电路是阈值转换器件实现神经元的经典电路,对选通管的耐久度要求较高.利用类似的方式,基于FeO_x^[140]和HfAlO_x^[30]的易失性忆阻器实现了更低功耗的LIF神经元.

Huang 等^[141]利用易失性 W/WO₃/poly 器件 构建了一个更为完整且可调的 LIF 神经元电路, 如图 13(d) 所示. 在神经元电路中, 输入信号流入 使得神经元 M1 的电导提高, 局部分级电压 (local graded potential, LGP) 分压上升, 模拟积分过程. 未达到设定阈值时, LGP 的泄漏使得 M1 阻态回 升,模拟漏电过程.当LGP达到设定阈值后,振荡器的脉冲通过 M2发放脉冲,模拟点火过程.整个神经元电路设计复杂,硬件代价较大,但能够较好地调制所发放的脉冲,如图 13(e)和图 13(f)所示.为了简化模型,Hua等^[143]直接将选通管和一个RRAM 串连,构成 1S1R 结构,形成一个神经元.适当的电阻值使得施加电压超过选通管阈值时,选通管表现出自振荡特性,且输出脉冲振荡频率与输入的脉冲电压和 RRAM 的电导率成正比.

除了寻常的仿生神经元外,利用选通管的阈 值漂移特性,也有部分研究集中于随机点火的神



图 13 易失型器件实现神经元 (a) 所提出的神经元电路的示意图; (b) 电容上的电压变化图; (c) 输出神经元的发放脉冲具有 相应的不应期和积分时间^[139]; (d) 带有两个 W/WO₃/PEDOT: PSS/Pt 忆阻器件的具体神经元电路; (e) 利用电路得到的单脉冲 生物积分点火; (f) 利用电路得到的连续脉冲生物积分点火^[141]; (g) 神经元电路原理图, 输入电压脉冲来自信号发生器; (h) 基于 CuS/GeSe 的神经元电路在脉冲幅度为 2 V、脉冲宽度为 7.5 ms 的输入电压脉冲序列下的随机脉冲发放事件^[142]

Fig. 13. Neurons implemented by volatile Devices: (a) Schematic illustration of the proposed neuron circuit; (b) the voltage variation across the capacitor; (c) the output neuron spike with the corresponding refractory period and integration moment^[139]; (d) the electrical circuit with two W/WO₃/PEDOT:PSS/Pt memristive devices; (e) spatial integration and bioinspired fire realized with the circuit; (f) temporal integration and bioinspired fire realized with the circuit^[141]; (g) schematic of neuronal circuit where the input voltage pulses originate from the signal generator; (h) the experimentally measured stochastic spike events of the CuS/GeSe based neuronal circuit under an input voltage pulse train with pulse height 2 V and duration 7.5 ms^[142]. 经元以用硬件实现神经网络训练时的 dropout. Wang 等^[142]利用 CuS/GeSe 导电桥型选通管制作 了随机神经元电路, 如图 13(g) 所示. 他们利用 CBTS 在物理机理上的阈值漂移, 使得神经元发放脉冲的 阈值变化, 形成概率发放脉冲的神经元. 图 12(h) 给出了随机神经元的脉冲发放情况, 神经元膜电位 越高, 发放脉冲的概率越高. 最后利用概率神经元, 在脉冲神经网络中提出了一种给出乳腺癌肿瘤诊 断不确定性的方法, 大大提高了诊断的保真性. 利 用 Ag/Ta₂O₅/Pt 型 CBTS 选通管与运算放大器, Huang 等^[144]实现了一种可调控概率的 dropout 神经元, 并利用此神经元在 MNIST数据集实现了 不错的识别准确率.

4 忆阻阵列实现类脑计算

忆阻器件提供了许多良好的特性,如结构简 单、高可扩展性、后端工艺兼容性和低能耗等,可 用于构建大规模的神经形态系统^[13].工业界将忆 阻器 crossbar 阵列,以无源和有源阵列的形式与传 统 CMOS 电路集成,以执行广泛的应用,包括机器 学习和神经形态计算等.

有源器件阵列通过晶体管的控制,很好地避免 了信号的串扰.相比于有源器件,无源器件需要具 备很好的自整流特性或非线性,才能在阵列操作中 不造成串扰以制备出大规模交叉阵列,且操作过程 必须具备良好的一致性.但是无源阵列可以在无外 加电源的情况下工作,并且集成密度更高,在小尺 寸下仍能保持良好的器件特性.下面对业界的主要 相关工作进行讨论.

4.1 无源集成阵列

无源阵列指不包含诸如晶体管等的无源器件 组成的阵列.无源阵列结构不需要通常为3端且需 要单晶硅衬底来制造的有源器件,提供了非常高的 密度,并具有多层三维集成的潜力^[13].但无源阵列 架构存在一个严重的问题,即存在潜行电流路径 (sneak path)导致的串扰.图14(a)是典型的 crossbar 阵列的示意图,其中通过打开状态相邻的存储 单元(2,3和4)的潜行电流干扰了关闭状态单元 1的读取.特别是相邻单元格的存储单元都处于低 阻态时,问题即更加严重^[14].目前,一种很好的解 决方式是将高非线性度元件(比如选通管)与存储 单元串联,构成比如 1S1R (one selector one resistor)的形式.

美国加州大学圣巴巴拉分校 Strukov教授课题 组^[145]首先在无源 RRAM 阵列上进行了探索.基 于优化后高一致性的 Pt/Ti/TiO_{2-x}/Al₂O₃/Pt/Ta, 他们建立了一个 12×12 的 crossbar 阵列,构成一 个单层感知机网络,如图 14(b) 所示,实现了对 3 × 3 像素的黑白字母"z""v""n"的正确分类. 该工作 利用了如图 14(c)的差分对结构, 通过权重的减法 尽可能地减小了共模串扰的影响,使得训练成功进 行. 后续, Strukov 教授课题组^[146] 进一步提高了阵 列的规模,制作了两个 20 × 20 的 crossbar 阵列, 分别作为输入层和隐藏层,构成了一个如图 14(d) 的含隐藏层的全连接神经网络,将之前网络的复杂 度提升了 10 倍. 美国密歇根大学的 Wei Lu 教授 研究团队^[147,148] 也进行了相关的研究. 基于 Pt/Au/ WO_r/W 型 RRAM 器件, 他们成功制备了 32×32 的无源 RRAM 阵列并用其完成了稀疏编码算法, 表明了该网络在自然图像分析等方面的应用价值. 在此应用中,阵列中的忆阻器仅用于做反复的前向 和反向推断,并不需要对器件进行更新,故读取串 扰的影响较小[147].

IBM 团队^[149]选用 PCRAM 作为阵列的突触, 构建了一个基于 1S1P(one selector one PCRAM) 的含 164885 个突触的无源阵列,如图 15(a)所示. 他们选用的 PCRAM 器件的置态过程往往是缓慢 的,但重置过程是突变的,对称性很差,故选用 2 个 PCRAM 器件作差分对形成一个突触,如图 15(b) 所示,以实现双向连续变化.通过该阵列中他们展 示了具有两个隐藏层的 ANN 网络对手写数字的 识别,其中训练模式为如图 15(b)所示的片上反向 传播训练和原位更新.

无源阵列的一个巨大优势是可以进行三维 (3D)阵列的延伸,从而大大提升集成密度.Strukov 教授课题组^[150]在之前工作^[145,146]的基础上提出基 于 Pt/Al₂O₃/TiO_{2-x}/TiN/Pt 忆阻器的 3D crossbar 结构,如图 16(a)所示.该阵列由具有共享的中间 电极的两个无源的 10 × 10 crossbar 组成,且优化 的器件可以在动态范围内精细地调控电导.利用类 似如图 16(b)的结构,有许多工作成功构造出 3D 1S1R 阵列^[151–153].

三维卷积神经网络 (convolutional neural network, CNN) 由于其固有的三维特性,已广泛应用 于磁共振成像 (MRI) 和视频识别等医学图像分析. 中国科学院刘明院士课题组 ^[154] 基于 HfO_2/TaO_x 器件构成的 1S1R 结构, 制作了如图 16(c)—(e) 所 示的三维阵列, 将其称为 3D 垂直电阻式随机存取 存储器 (vertical resistive random access memory, VRRAM). 通过将 VRRAM 作为 3D 卷积核, 该工 作对 16 × 16 × 16 像素的手写数字成功进行了边 缘检测,实现了 3D CNN 加速器低功耗、高容量的 需求. 美国马萨诸塞大学的 Yang 教授和 Xia 教授 课题组^[155]提出了一个不同的 3D CNN 概念结构, 构建了一个由八层单体集成的记忆器件组成的三 维阵列. 在该三维结构中,通过垂直排列的输入和 输出实现了卷积核的斜向映射,极大提高了卷积神 经网络的并行能力.



图 14 无源忆阻阵列神经形态计算 (a) 典型的二维 Crossbar 阵列潜行电流读取扰动问题示意图 (细蓝线表示读取电流、粗红 线表示潜行电流)^[14]; (b) 利用 10 × 6 忆阻器 Crossbar 实现的单层感知机网络; (c) 针对特定输入图片 (程式化的字母 "z") 的分类 操作示例^[145]; (d) 两个 20 × 20 的 crossbar 阵列实现双层神经网络的示意图^[146]

Fig. 14. Passive memristive arrays for neuromorphic computing: (a) A schematic diagram of the typical 2D Crossbar array showing the read disturbance problem by the presence of sneak current (The thin blue line represents reading current, and the thick red line represents sneak current)^[14]; (b) an implementation of a single-layer perceptron using a 10×6 fragment of the memristive crossbar; (c) an example of the classification operation for a specific input pattern (stylized letter 'z')^[145]; (d) a schematic diagram of two 20×20 crossbar arrays implementing a two-layer neural network^[146].



图 15 1S1P 无源忆阻阵列神经形态计算 (a) 存算一体的 1S1P 结构对 crossbar 阵列实现形式; (b) 差分对结构示意图; (c) 神经 网络的输入前向传播过程; (d) 神经网络更新示意图^[149]

Fig. 15. 1S1P passive memristive array for neuromorphic computing: (a) In-memory computing implemented using dense crossbar arrays of 1S1P pairs; (b) structure diagram of differential pairs; (c) the input forward propagation process of the neural network; (d) schematic diagram of neural network update^[149].

4.2 有源集成阵列

在有源阵列中, 忆阻器件与 CMOS 晶体管串 联, 形成 1T1R(one transistor one resistor) 结构. 晶体管能够对每个忆阻进行单独的控制, 从而减弱 了潜行路径的影响, 允许精确的读写. 与无源结构 相比, CMOS 晶体管成熟的设计和制造技术使构 建大型 1T1R 阵列成为可能. 然而, 晶体管集成的 代价是器件单元面积会更大. 此外, 对单晶硅衬底 的依赖限制了多层有源阵列的发展, 这进一步限制 了阵列的存储密度^[13].

IBM 团队^[149]利用 1S1P 器件并未得到理想的 结果,开始了有源相变存储阵列 (one transistor one PCM, 1T1P)的探索.引入有源结构单元 1T1P 会 牺牲器件的单元面积,但是可以增加器件的可控 性,能够减小阵列集成中的"IR drop"问题,易于大 规模阵列集成.另外,利用 PCMO RRAM,他们也 实现了对手写数字的高精度分类^[156]. Yang教授课 题组^[157]也同样致力于1T1R有源阵列的研究.该 阵列主要基于 Ta/HfO₂/Pt RRAM 与晶体管的串 联, 阵列规模为 128 × 164, 如图 17(a) 所示. 通过 将整个阵列拆成两部分,他们构建了一个 64 × 54 × 10 的 3 层 ANN 网络, 运用与文献 [149] 类似的训 练方式实现了对手写数字更准确的识别. 储备池 计算 (reservoir computing, RC) 是一种新兴的神 经形态计算算法.相较于传统的循环神经网络 (recurrent neural network, RNN), RC 只需训练 读出层的权重,在训练时间成本上进一步优化[158]. 中国科学院微电子所尚大山课题组[159]首次使用 具有瞬态去极化特性的超薄 (3.5 nm) 铁电隧穿结 (FTJs)构建了高维的储备池单元. 然后, 他们将 FTJ 储备池与基于 1T1R 的二值化读出层相结合, 给出了高能效 (35 pJ)、高处理速度 (500 ns) 和高 识别精度 (92.3%) 的数字序列分类.



图 16 3D 无源忆阻阵列神经形态计算 (a) 两个 Pt/Al₂O₃/TiO_{2-x}/TiN/Pt 型忆阻器堆叠结构的等效电路^[150]; (b) 共享位线结构的 3D Crossbar 阵列^[14]; (c) 整个电路利用 FPGA 控制的继电器矩阵实现 Crossbar 的自动控制测试; (d) 新的三维 VRRAM 结构的高分辨率透射显微镜图象; (e) 三维 VRRAM 架构中一次卷积操作的电流方向原理图^[154]

Fig. 16. 3D passive memristive array for neuromorphic computing: (a) Equivalent circuit for two $Pt/Al_2O_3/TiO_{2-x}/TiN/Pt$ memristors in the stacked configuration^[150]; (b) a schematic diagram showing the shared bit line structure in cross-line type 3D Crossbar array^[14]; (c) FPGA-controlled relay matrix to achieve test automation; (d) HRTEM image of the novel 3D VRRAM structure; (e) the schematic of the 3D VRRAM architecture and current flow for one convolution operation^[154].

清华大学吴华强教授和钱鹤教授课题组[160] 同 样一直在进行有源忆阻阵列的研究. 他们首先优化 器件,选择双向连续可调的 TiN/TaO_x/HfAlO_x/ TiN RRAM 作为阵列的突触. 如图 17(b)—图 17(e) 所示, 他们构造了 128 × 8 的 1T1R 阵列, 将人脸 像素利用脉冲个数编码输入,提出了一种基于写验 证 (write-verify programming) 的原位更新模式, 最终实现了对 Yale Face Database 中人脸数据的 高效分类. 后续, 通过进一步的研究, 他们构建了 一个具有 8 × 2048 个单元的 1T1R 阵列用作卷积 神经网络计算[161]. 该阵列拥有目前业界最大的 集成规模之一,这归因于单个忆阻器件的高性能和 高一致性. 通过该阵列, 他们对 MNIST 数据集 进行了图像识别,最终得到了96%以上的高识别 准确性. 由于传统忆阻器件非线性和非对称性等非 理想因素严重影响了系统性能, IBM 研究团队^[162] 另辟蹊径,引入晶体管和电容 (three transisotrs one capcitor, 3T1C) 作为片上训练的基本单元, 通 过电容上蓄积的电压进一步调控晶体管的工作 状态.

5 总结与展望

本文主要从忆阻器的材料与物理机理分类、人 工突触和神经元的忆阻器实现和忆阻阵列的神经 形态计算三个方面回顾了近期忆阻神经形态计算 领域内的重大进展. 忆阻器件根据材料和物理机理 可分为阻变随机存储器 RRAM、相变存储器 PCRAM、磁随机存储器 MRAM和铁电随机存储 器 FeRAM 四种. 总体来说, 它们展现出良好的突 触特性——10⁹的开关比^[28]、115 fJ的开态功耗^[42], 85 ps的开关速度^[43], 10 年以上的保持特性^[44]及10¹² 的耐擦写次数^[45]. 近年来, 以忆阻器件为核心的神 经形态芯片已经成功实现了存算一体架构, 规避了 冯·诺伊曼瓶颈和 Dennard 缩放比例定律的限制, 展现出了远高于 CMOS 芯片的能量利用效率. 然 而, 忆阻类脑计算的研究仍处于初级阶段, 面临着 很多挑战, 进一步的发展可以从以下三个方向展开.

1) 新材料和新机制的探索

目前, 忆阻器件的阻值转换机制依旧存在争



图 17 1T1R有源忆阻阵列神经形态计算 (a)用于原位学习的忆阻平台.从左到右分别是:带有晶体管阵列的晶圆、芯片特写 图、1T1R单元的显微镜图像、1T1R单元的 SEM 图像、Ta /HfO₂/Pt 忆阻器的横截面 TEM 图像^[157].(b)单层神经网络在 1T1R 阵列上的映射.(c)使用 CMOS 兼容制造工艺制备的 1024 1T1R 单元的阵列显微镜图.(d)训练过程流程图.(e)模型映射到 输入及并行读取操作的原理图^[160]

Fig. 17. 1T1R passive memristive array for neuromorphic computing: (a) Memristive platform for in situ learning. From left to right are: A wafer with transistor arrays, close-up of chip image, microscope image of 1T1R cell, SEM of an individual 1T1R cell, cross-sectional TEM image of the Ta/HfO₂/Pt memristor^[157]. (b) Mapping of a one-layer neural network on the 1T1R array. (c) The micrograph of a fabricated 1024-cell-1T1R array using fully CMOS compatible fabrication process. (d) The training process flow chart. (e) The schematic of parallel read operation and how a pattern is mapped to the input^[160].

议,当涉及纳米级原子快速迁移排列时,利用仪器 表征存在很大的困难.通过对机制的进一步研究探 索,可以进行相应的紧凑模型建模来定量描述器件 开关特性,从而实现更高性能的器件构建与仿真. 为更好地实现存算一体系统,需要理想的模拟型阻 变器件,这对中间状态数、线性度、对称性、阻变窗 口、良率、波动性、保持特性和耐擦写特性都提出 了需求^[98].目前,并未发现足够优秀的器件同时将 所有特性做到极致."界面工程"、功能层中掺杂以 及结合新型二维材料是目前较为有效的器件改进 方式^[33],但新型材料的引入会提高制备流程的复 杂度,使得器件大规模集成成为挑战.如何平衡好 两者,也是未来值得研究的课题.另外,对忆阻器 一些难以避免的非理想特性加以利用也是当前的 一个重要研究方向.如忆阻器件内在的开关随机 性,已被用于构建随机数发生器,在信息安全方面 有着重要应用.研究出能够有效利用这些内在器件 特性的新算法和架构,也能为神经形态计算带来新 的发展.

2) 全忆阻神经网络的构建

由人工神经元和突触组成的脉冲神经网络阵列可以显著降低或消除 ADC (analog to digital converter) 能量消耗,从而控制模拟 IMC (internal model control) 系统的功率和面积,是目前神经形

态计算领域中的一个重要方向^[13]. 经过最近几年的探索研究, 忆阻器件在突触与神经元的功能模拟 上取得了许多重要成果. 然而, 目前主流的脉冲神 经网络芯片却主要依赖于传统 CMOS 技术. 这主 要是因为忆阻器件的研究目前还处于初级过程, 阻 变机制的不明确和制备工艺的不成熟使得不同器 件和同个器件不同循环过程中的差异性较大, 使得 全忆阻神经网络的实现十分具有挑战性. 针对这个 问题, 杨建华团队^[163]和缪向水团队^[164,165]已经取 得了较大进展, 分别完成了阵列和仿真级别的全忆 阻神经网络识别任务. 在未来, 成功构造高准确率 的全忆阻神经网络, 模拟 IMC 硬件系统中的许多 挑战将迎刃而解.

3) 更复杂的神经形态计算

通过 3D 堆叠技术, 可以极大地提高忆阻阵列 的存储与计算密度,是近年来阵列方面的一个重大 突破. 然而, 目前的 3D 阵列所做的往往只是提高 计算并行度,并未合理地利用多出来的第三个存储 和计算维度. 这使得当前主流的神经形态计算任务 往往都是二维的图片识别等,更加复杂的视频动作 识别成为当前的一个重要攻关目标. 除此之外, 目 前的类脑芯片总是集中在一些特定的识别应用上, 但实际上真正的智能应该是能够主动地学习以适 应环境的通用智能.引入深度强化学习算法是近年 来提出的一种解决办法. 但是采样低效导致训练缓 慢、试错成本高以及可复现性低等问题成为目前难 以突破的瓶颈[166],直至目前也没有一款低功耗的 高效强化学习芯片成功问世. 在未来, 如果出现更 有效的强化学习算法,那人们又将朝智能时代前进 一大步.

总之,在人工智能革命的浪潮下,忆阻类脑计 算的研究将带来重大的发展机遇.探索更优异的器 件,建立更大规模的忆阻阵列以及最后真正实现类 脑计算,需要研究者们共同努力,克服困难.也许 今后的某一天,世界将真正进入新的智能时代!

参考文献

- Sze V, Chen Y H, Yang T J, Emer J S 2017 Proc. IEEE 105 2295
- [2] Moore G E 1998 *Proc. IEEE* 86 82
- [3] Dennard R, Gaensslen F, Yu H N, Rideout V, Bassous E, LeBlanc A 1974 IEEE J. Solid-State Circuits 9 256
- [4] Di Ventra M, Pershin Y V 2013 Nat. Phys. 9 200
- [5] Mead C 1990 Proc. IEEE **78** 1629

- [6] Furber S 2016 J. Neural Eng. 13 051001
- [7] Ielmini D, Wong H S P 2018 Nat. Electron. 1 333
- [8] Wang Z, Wu H, Burr G W, Hwang C S, Wang K L, Xia Q, Yang J J 2020 Nat. Rev. Mater. 5 173
- [9] Chua L O, Kang S M 1976 Proc. IEEE 64 209
- [10] Strukov D B, Snider G S, Stewart D R, Williams R S 2008 Nature 453 80
- [11] Lee J, Lu W D 2018 Adv. Mater. **30** 1702770
- [12] Zidan M A, Strachan J P, Lu W D 2018 Nat. Electron. 1 22
- [13] Wu Y, Wang X, Lu W D 2022 Semicond. Sci. Technol. 37 024003
- [14] Seok J Y, Song S J, Yoon J H, Yoon K J, Park T H, Kwon D E, Lim H, Kim G H, Jeong D S, Hwang C S 2014 Adv. Funct. Mater. 24 5316
- [15] Bian H, Goh Y Y, Liu Y, Ling H, Xie L, Liu X 2021 Adv. Mater. 33 2006469
- [16] Zhu J, Zhang T, Yang Y, Huang R 2020 Appl. Phys. Rev. 7 011312
- [17] Rivnay J, Inal S, Salleo A, Owens R M, Berggren M, Malliaras G G 2018 Nat. Rev. Mater. 3 1
- [18] Valov I, Waser R, Jameson J R, Kozicki M N 2011 Nanotechnology 22 254003
- [19] Sun H, Liu Q, Li C, Long S, Lv H, Bi C, Huo Z, Li L, Liu M 2014 Adv. Funct. Mater. 24 36
- [20] Wang Z, Joshi S, Savel'ev S E, Jiang H, Midya R, Lin P, Hu M, Ge N, Strachan J P, Li Z, Wu Q, Barnell M, Li G L, Xin H L, Williams R S, Xia Q, Yang J J 2017 Nat. Mater. 16 10
- [21] Yan P, Li Y, Hui Y J, Zhong S J, Zhou Y X, Xu L, Liu N, Qian H, Sun H J, Miao X S 2015 Appl. Phys. Lett. 107 083501
- [22] Liu N, Yan P, Li Y, Lu K, Sun H, Ji H, Xue K, Miao X 2018 Appl. Phys. A 124 1
- [23] Wang W, Wang M, Ambrosi E, Bricalli A, Laudato M, Sun Z, Chen X, Ielmini D 2019 Nat. Commun. 10 81
- [24] Wang W, Covi E, Lin Y H, Ambrosi E, Ielmini D 2019 IEEE International Electron Devices Meeting (IEDM) San Francisco, CA, USA, December, 2019 pp32.3.1–32.3.4
- [25] Chekol S A, Menzel S, Ahmad R W, Waser R, Hoffmann-Eifert S 2021 Adv. Funct. Mater. 32 2111242
- [26] Zhao X, Ma J, Xiao X, Liu Q, Shao L, Chen D, Liu S, Niu J, Zhang X, Wang Y, Cao R, Wang W, Di Z, Lv H, Long S, Liu M 2018 Adv. Mater. 30 1705193
- [27] Lin Q, Li Y, Xu M, Cheng Q, Qian H, Feng J, Tong H, Miao X 2018 IEEE Electron Device Lett. 39 496
- [28] Lin Q, Feng J, Yuan J, Liu L, Eshraghian J K, Tong H, Xu M, Wang X, Miao X 2021 J. Mater. Chem. C 9 14799
- [29] Song B, Xu H, Liu S, Liu H, Li Q 2018 IEEE J. Electron Devices Soc. 6 674
- [30] Lu Y F, Li Y, Li H, Wan T Q, Huang X, He Y H, Miao X 2020 IEEE Electron Device Lett. 41 1245
- [31] Yu S, Chen P Y 2016 IEEE Solid State Circuits Mag. 8 43
- [32] Vianello E, Thomas O, Molas G, Turkyilmaz O, Jovanovic N, Garbin D, PalmaG, Alayan M, Nguyen C, Coignus J, Giraud B, Benoist T, Reyboz M, Toffoli A, Charpin C, Clermidy F, Perniola L 2014 *IEEE International Electron Devices Meeting(IEDM)* San Francisco, CA, USA , December, 2014 pp6.3.1–6.3.4
- [33] Wang H, Yan X 2019 Phys. Status Solidi RRL 13 1900073
- [34] Wong H S P, Lee H Y, Yu S, Chen Y S, Wu Y, Chen P S, Lee B, Chen F T, Tsai M J 2012 Proc. IEEE 100 1951
- [35] Russo U, Ielmini D, Cagli C, et al. 2007 IEEE International Electron Devices Meeting(IEDM) Washington, DC, USA, 2007 pp775–778

- [36] Kwon D H, Kim K M, Jang J H, Jeon J M, Lee M H, Kim G H, Li X S, Park G S, Lee B, Han S, et al. 2010 Nat. Nanotechnol. 5 148
- [37] Kumar S, Wang Z, Huang X, Kumari N, Davila N, Strachan J P, Vine D, Kilcoyne A D, Nishi Y, Williams R S 2016 ACS Nano 10 11205
- [38] Cooper D, Baeumer C, Bernier N, Marchewka A, La Torre C, Dunin-Borkowski R E, Menzel S, Waser R, Dittmann R 2017 Adv. Mater. 29 1700212
- [39] Kim S, Choi S, Lu W 2014 ACS Nano 8 2369
- [40] Larentis S, Nardi F, Balatti S, Gilmer D C, Ielmini D 2012 IEEE Trans. Electron Devices 59 2468
- [41] Kim S, Du C, Sheridan P, Ma W, Choi S, Lu W D 2015 Nano Lett. 15 2203
- [42] Strachan J P, Torrezan A C, Medeiros-Ribeiro G, Williams R S 2011 Nanotechnology 22 505402
- [43] Choi B J, Torrezan A C, Strachan J P, Kotula P, Lohn A, Marinella M J, Li Z, Williams R S, Yang J J 2016 Adv. Funct. Mater. 26 5290
- [44] Jiang H, Han L, Lin P, Wang Z, Jang M H, Wu Q, Barnell M, Yang J J, Xin H L, Xia Q 2016 Sci. Rep. 6 1
- [45] Lee M J, Lee C B, Lee D, Lee S R, Chang M, Hur J H, Kim Y B, Kim C J, Seo D H, Seo S, et al. 2011 Nat. Mater. 10 625
- [46] Le Gallo M, Sebastian A 2020 J. Phys. D: Appl. Phys. 53 213002
- [47] Salinga M, Carria E, Kaldenbach A, Bornhöfft M, Benke J, Mayer J, Wuttig M 2013 Nat. Commun. 4 1
- [48] Xu K, Miao X, Xu M 2019 Phys. Status Solidi RRL 13 1800506
- [49] Noé P, Vallée C, Hippert F, Fillot F, Raty J Y 2017 Semicond. Sci. Technol. 33 013002
- [50] Raty J Y, Zhang W, Luckas J, Chen C, Mazzarello R, Bichara C, Wuttig M 2015 Nat. Commun. 6 7467
- [51] Ding K, Wang J, Zhou Y, Tian H, Lu L, Mazzarello R, Jia C, Zhang W, Rao F, Ma E 2019 *Science* 366 210
- [52] Zhang W, Mazzarello R, Wuttig M, Ma E 2019 Nat. Rev. Mater. 4 150
- [53] Kalikka J, Akola J, Jones R 2016 Phys. Rev. B 94 134105
- [54] Rao F, Ding K, Zhou Y, Zheng Y, Xia M, Lv S, Song Z, Feng S, Ronneberger I, Mazzarello R, Zhang W, Ma E 2017 *Science* 358 1423
- [55] Mott N F 1949 Proc. Phys. Soc. A 62 416
- [56] Hubbard J 1963 Proc. Math. Phys. Eng. Sci. 276 238
- [57] You Zhou, Ramanathan S 2015 Proc. IEEE 103 1289
- [58] Zhou X, Gu D, Li Y, Qin H, Jiang Y, Xu J 2019 Nanoscale 11 22070
- [59] Shi Y, Chen L Q 2019 Phys. Rev. Appl. 11 014059
- [60] Del Valle J, Salev P, Tesler F, Vargas N M, Kalcheim Y, Wang P, Trastoy J, Lee M H, Kassabian G, Ramírez J G, et al. 2019 Nature 569 388
- [61] Sultan M, Ignatova K, Thorsteinsson E, Arnalds U 2022 Thin Solid Films 742 139048
- [62] Zhao X, Chen A, Ji J, Wu D, Gan Y, Wang C, Ma G, Lin C Y, Lin C C, Liu N, Wan H, Tao L, Wang B, Chang T C, Wang H 2021 IEEE Trans. Electron Devices 68 2255
- [63] Oboril F, Bishnoi R, Ebrahimi M, Tahoori M B 2015 IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 34 367
- [64] Bhatti S, Sbiaa R, Hirohata A, Ohno H, Fukami S, Piramanayagam S 2017 Mater. Today 20 530
- [65] Carboni R, Ambrogio S, Chen W, Sid[dik M, Harms J, Lyle A, Kula W, Sandhu G, Ielmini D 2016 IEEE International Electron Devices Meeting (IEDM) San Francisco, CA, USA ,

2016-12 pp21.6.1-21.6.4

- [66] Sato H, Honjo H, Watanabe T, Niwa M, Koike H, Miura S, Saito T, Inoue H, Nasuno T, Tanigawa T, Noguchi Y, Yoshiduka T, Yasuhira M, Ikeda S, Kang S Y, Kubo T, Yamashita K, Yagi Y, Tamura R, Endoh T 2018 *IEEE International Electron Devices Meeting (IEDM)* San Francisco, CA, December, 2018 pp27.2.1–27.2.4
- [67] Sakhare S, Perumkunnil M, Bao T H, Rao S, Kim W, Crotti D, Yasin F, Couet S, Swerts J, Kundu S, Yakimets D, Baert R, Oh Hr, Spessot A, Mocuta A, Kar G S, Furnemont A 2018 IEEE International Electron Devices Meeting (IEDM) San Francisco, CA, December, 2018 pp18.3.1–18.3.4
- [68] Yuasa S, Hono K, Hu G, Worledge D C 2018 MRS Bull. 43 352
- [69] Song Y J, Lee J H, Han S H, Shin H C, Lee K H, Suh K, Jeong D E, Koh G H, Oh S C, Park J H, Park S O, Bae B J, Kwon O I, Hwang K H, Seo B Y, Lee Y K, Hwang S H, Lee D S, Ji Y, Park K C, Jeong G T, Hong H S, Lee K P, Kang H K, Jung E S 2018 *IEEE International Electron Devices Meeting (IEDM)* San Francisco, CA, Decemeber, 2018 pp18.2.1–18.2.4
- [70] Jain S, Ranjan A, Roy K, Raghunathan A 2018 IEEE Trans. VLSI Syst. 26 470
- [71] Jung S, Lee H, Myung S, Kim H, Yoon S K, Kwon S W, Ju Y, Kim M, Yi W, Han S, et al. 2022 *Nature* 601 211
- [72] Wang M, Cai W, Zhu D, Wang Z, Kan J, Zhao Z, Cao K, Wang Z, Zhang Y, Zhang T, Park C, Wang J P, Fert A, Zhao W 2018 Nat. Electron. 1 582
- [73] Ikegawa S, Mancoff F B, Janesky J, Aggarwal S 2020 IEEE Trans. Electron Devices 67 1407
- [74] Zhuravlev M Y, Sabirianov R F, Jaswal S S, Tsymbal E Y 2005 Phys. Rev. Lett. 94 246802
- [75] Kohlstedt H, Pertsev N A, Rodríguez Contreras J, Waser R 2005 Phys. Rev. B 72 125341
- [76] Wen Z, Li C, Wu D, Li A, Ming N 2013 Nat. Mater. 12 617
- [77] Velev J P, Burton J D, Zhuravlev M Y, Tsymbal E Y 2016 NPJ Comput. Mater. 2 16009
- [78] Huang W, Zhao W, Luo Z, Yin Y, Lin Y, Hou C, Tian B, Duan C G, Li X G 2018 Adv. Electron Mater. 4 1700560
- [79] Wen Z, Wu D 2020 Adv. Mater. 32 1904123
- [80] Sunbul A, Ali T, Mertens K, Revello R, Lehninger D, Muller F, Lederer M, Kuhnel K, Rudolph M, Oehler S, Hoffmann R, Zimmermann K, Biedermann K, Schramm P, Czernohorsky M, Seidel K, Kampfe T, Eng L M 2022 IEEE Trans. Electron Devices 69 808
- [81] Lyu J, Song T, Fina I, Sánchez F 2020 Nanoscale 12 20
- [82] Shekhawat A, Walters G, Yang N, Guo J, Nishida T, Moghaddam S 2020 Nanotechnology 31 39LT01
- [83] Goh Y, Hwang J, Lee Y, Kim M, Jeon S 2020 Appl. Phys. Lett. 117 242901
- [84] Wu J, Chen H Y, Yang N, Cao J, Yan X, Liu F, Sun Q, Ling X, Guo J, Wang H 2020 Nat. Electron. 3 466
- [85] Garello K, Yasin F, Couet S, Souriau L, Swerts J, Rao S, Van Beek S, Kim W, Liu E, Kundu S, Tsvetanova D, Croes K, Jossart N, Grimaldi E, Baumgartner M, Crotti D, Fumemont A, Gambardella P, Kar G S 2018 *IEEE Symposium on VLSI Circuits* Honolulu, HI, June, 2018 pp81–82
- [86] Francois T, Coignus J, Makosiej A, Giraud B, Carabasse C, Barbot J, Martin S, Castellani N, Magis T, Grampeix H, Van Duijn S, Mounet C, Chiquet P, Schroeder U, Slesazeck S, Mikolajick T, Nowak E, Bocquet M, Barrett N, Andrieu F, Grenouillet L 2021 *IEEE International Electron Devices*

Meeting~(IEDM)San Francisco, CA, USA , December 11, 2021 pp33.1.1–33.1.4

- [87] Wu X, Ge R, Chen P A, Chou H, Zhang Z, Zhang Y, Banerjee S, Chiang M H, Lee J C, Akinwande D 2019 Adv. Mater. 31 1806790
- [88] Zhang F, Zhang H, Krylyuk S, Milligan C A, Zhu Y, Zemlyanov D Y, Bendersky L A, Burton B P, Davydov A V, Appenzeller J 2019 Nat. Mater. 18 55
- [89] Ikeda S, Hayakawa J, Ashizawa Y, Lee Y, Miura K, Hasegawa H, Tsunoda M, Matsukura F, Ohno H 2008 Appl. Phys. Lett. 93 082508
- [90] Cheng H Y, Chien W C, BrightSky M, Ho Y H, Zhu Y, Ray A, Bruce R, Kim W, Yeh C W, Lung H L, Lam C 2015 *IEEE International Electron Devices Meeting (IEDM)* Washington, DC, USA, December 2015 pp3.5.1–3.5.4
- [91] Golonzka O, Alzate J G, Arslan U, et al. 2018 IEEE International Electron Devices Meeting (IEDM) San Francisco, CA, 2018-12 pp18.1.1–18.1.4
- [92] Kim I S, Cho S L, Im D H, Cho E H, Kim D H, Oh G H, Ahn D H, Park S O, Nam S W, Moon J T, Chung C H 2010 Symposium on VLSI Technology Honolulu, USA, June, 2010 pp203–204
- [93] Yang Z, Li B, Wang J J, Wang X D, Xu M, Tong H, Cheng X, Lu L, Jia C, Xu M, et al. 2022 Adv. Sci. 9 2103478
- [94] Grezes C, Ebrahimi F, Alzate J, Cai X, Katine J, Langer J, Ocker B, Khalili Amiri P, Wang K 2016 Appl. Phys. Lett. 108 012403
- [95] Grenouillet L, Francois T, Coignus J, Vaxelaire N, Carabasse C, Triozon F, Richter C, Schroeder U, Nowak E 2020 IEEE Silicon Nanoelectronics Workshop (SNW) Honolulu, USA, June 2020 pp5–6
- [96] Martin S J, Grimwood P D, Morris R G M 2000 Annu. Rev. Neurosci. 23 649
- [97] Citri A, Malenka R C 2008 Neuropsychopharmacol 33 18
- [98] Chen P Y, Peng X, Yu S 2017 IEEE International Electron Devices Meeting (IEDM) San Francisco, USA , December, 2017 pp6.1.1–6.1.4
- [99] Wan Q, Sharbati M T, Erickson J R, Du Y, Xiong F 2019 Adv. Mater. Technol. 4 1900037
- [100] Kandel E R, Schwartz J H, Jessell T M, Siegelbaum S, Hudspeth A J, Mack S, et al. 2000 *Principles of Neural Science* (Vol. 4) (New York: McGraw-hill) pp512–514
- [101] Jo S H, Chang T, Ebong I, Bhadviya B B, Mazumder P, Lu W 2010 Nano Lett. 10 1297
- [102] Fu Y, Dong B, Su W C, Lin C Y, Zhou K J, Chang T C, Zhuge F, Li Y, He Y, Gao B, et al. 2020 Nanoscale 12 22970
- [103] Liu S, Xiao T P, Cui C, Incorvia J A C, Bennett C H, Marinella M J 2021 Appl. Phys. Lett. 118 202405
- [104] Liu L, Xiong W, Liu Y, Chen K, Xu Z, Zhou Y, Han J, Ye C, Chen X, Song Z, Zhu M 2020 Adv. Electron Mater. 6 1901012
- [105] Wang Z, Wang L, Wu Y, Bian L, Nagai M, Jv R, Xie L, Ling H, Li Q, Bian H, et al. 2021 Adv. Mater. 33 2104370
- [106] Zucker R S, Regehr W G 2002 Annu. Rev. Physiol. 64 355
- [107] Abbott L F, Regehr W G 2004 $Nature \ 431$ 796
- [108] Ohno T, Hasegawa T, Tsuruoka T, Terabe K, Gimzewski J K, Aono M 2011 Nat. Mater. 10 591
- [109] Chang T, Jo S H, Lu W 2011 ACS Nano 5 7669
- [110] Yang R, Terabe K, Yao Y, Tsuruoka T, Hasegawa T, Gimzewski J K, Aono M 2013 Nanotechnology 24 384003
- [111] Zhang S R, Zhou L, Mao J Y, Ren Y, Yang J Q, Yang G H, Zhu X, Han S T, Roy V A L, Zhou Y 2019 Adv. Mater. Technol. 4 1800342

- [112] Shi Y, Liang X, Yuan B, Chen V, Li H, Hui F, Yu Z, Yuan F, Pop E, Wong H S P, Lanza M 2018 Nat. Electron. 1 458
- [113] Rachmuth G, Shouval H Z, Bear M F, Poon C S 2011 Proc. Natl. Acad. Sci. 108 E1266
- Bienenstock E L, Cooper L N, Munro P W 1982 J. Neurosci.
 2 32
- [115] Cooper L N, Bear M F 2012 Nat. Rev. Neurosci. 13 798
- [116] Ren Z Y, Zhu L Q, Guo Y B, Long T Y, Yu F, Xiao H, Lu H L 2020 ACS Appl. Mater. Interfaces 12 7833
- [117]~ Shouval H Z 2011 Proc. Natl. Acad. Sci. 108~19103
- [118] Du C, Ma W, Chang T, Sheridan P, Lu W D 2015 Adv. Funct. Mater. 25 4290
- [119] Hebb D O 2008 The Organization of Behavior; A Neuropsychological Theory (Vol. 1)(New York: Wiley) p335
- [120] Markram H, Gerstner W, Sjöström P J 2011 Front. Synaptic Neurosci. 3 4
- [121] Feldman D E 2000 Neuron 27 45
- [122] Sjöström P J, Turrigiano G G, Nelson S B 2001 Neuron 32 1149
- [123] Markram H, Lübke J, Frotscher M, Sakmann B 1997 Science 275 213
- [124] Yu S, Wu Y, Jeyasingh R, Kuzum D, Wong H S P 2011 IEEE Trans. Electron Devices 58 2729
- [125] Ielmini D, Wang Z, Liu Y 2021 APL Mater. 9 050702
- [126] Li Y, Zhong Y, Zhang J, Xu L, Wang Q, Sun H, Tong H, Cheng X, Miao X 2014 Sci. Rep. 4 1
- [127] Graupner M, Brunel N 2012 Proc. Natl. Acad. Sci. 109 3991
- [128] Yan X, Pei Y, Chen H, Zhao J, Zhou Z, Wang H, Zhang L, Wang J, Li X, Qin C, Wang G, Xiao Z, Zhao Q, Wang K, Li H, Ren D, Liu Q, Zhou H, Chen J, Zhou P 2019 Adv. Mater. 31 1805284
- [129] Gerstner W, Kistler W M 2002 Spiking Neuron Models: Single Neurons, Populations, Plasticity (Cambridge: Cambridge University Press) pp3–5
- [130] Lee G, Baek J H, Ren F, Pearton S J, Lee G H, Kim J 2021 Small 17 2100640
- [131] Bear M, Connors B, Paradiso M A 2020 Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain (Burlington: Jones & Bartlett Learning) pp:65–68
- [132] Yi F H, Guo Q L, Yu J W, Lei D 2021 J. Control. Decis. 36 1
- [133] Dayan P, Abbott L 2001 Theoretical Neuroscience (Cambridge: MIT Press) pp162–166
- [134] Tuma T, Pantazi A, Le Gallo M, Sebastian A, Eleftheriou E 2016 Nat. Nanotechnol. 11 693
- [135] Mehonic A, Kenyon A J 2016 Front. Neurosci. 10 57
- [136] Lashkare S, Chouhan S, Chavan T, Bhat A, Kumbhare P, Ganguly U 2018 IEEE Electron Device Lett. 39 484
- [137] Lee D, Kwak M, Moon K, Choi W, Park J, Yoo J, Song J, Lim S, Sung C, Banerjee W, Hwang H 2019 Adv. Electron. Mater. 5 1800866
- [138] Pickett M D, Medeiros-Ribeiro G, Williams R S 2013 Nat. Mater. 12 114
- [139] Zhang X, Wang W, Liu Q, Zhao X, Wei J, Cao R, Yao Z, Zhu X, Zhang F, Lv H, Long S, Liu M 2018 IEEE Electron Device Lett. 39 308
- [140] Zhang Y, He W, Wu Y, Huang K, Shen Y, Su J, Wang Y, Zhang Z, Ji X, Li G, et al. 2018 *Small* 14 1802188
- [141] Huang H, Yang R, Tan Z, He H, Zhou W, Xiong J, Guo X 2019 Adv. Mater. **31** 1803849
- [142] Wang K, Hu Q, Gao B, Lin Q, Zhuge F W, Zhang D Y, Wang L, He Y H, Scheicher R H, Tong H, Miao X S 2021 Mater. Horiz. 8 619

- [143] Hua Q, Wu H, Gao B, Zhang Q, Wu W, Li Y, Wang X, Hu W, Qian H 2019 *Glob. Chall.* **3** 1900015
- [144] Huang H, Xiao Y, Yang R, Yu Y, He H, Wang Z, Guo X 2020 Adv. Sci. 7 2001842
- [145] Prezioso M, Merrikh-Bayat F, Hoskins B D, Adam G C, Likharev K K, Strukov D B 2015 Nature 521 61
- [146] Bayat F M, Prezioso M, Chakrabarti B, Nili H, Kataeva I, Strukov D 2018 Nat. Commun. 9 2331
- [147] Sheridan P M, Cai F, Du C, Ma W, Zhang Z, Lu W D 2017 Nat. Nanotechnol. 12 784
- [148] Cai F, Correll J M, Lee S H, Lim Y, Bothra V, Zhang Z, Flynn M P, Lu W D 2019 Nat. Electron. 2 290
- [149] Burr G W, Shelby R M, Sidler S, di Nolfo C, Jang J, Boybat I, Shenoy R S, Narayanan P, Virwani K, Giacometti E U, Kurdi B N, Hwang H 2015 *IEEE Trans. Electron Devices* 62 3498
- [150] Adam G C, Hoskins B D, Prezioso M, Merrikh-Bayat F, Chakrabarti B, Strukov D B 2017 *IEEE Trans. Electron Devices* 64 312
- [151] Wang I T, Chang C C, Chiu L W, Chou T, Hou T H 2016 Nanotechnology 27 365204
- [152] Yang H, Hao X, Wang Z, Malmhall R, Gan H, Satoh K, Zhang J, Jung D H, Wang X, Zhou Y, Yen B K, Huai Y 2017 IEEE International Electron Devices Meeting (IEDM) San Francisco, CA, USA, 2017-12 pp38.1.1–38.1.4
- [153] Kim S, Liu X, Park J, Jung S, Lee W, Woo J, Shin J, Choi G, Cho C, Park S, Lee D, Cha E jun, Lee B H, Lee H D, Kim S G, Chung S, Hwang H 2012 Symposium on VLSI Technology (VLSIT) Honolulu, HI, USA, 2012-06 pp155–156
- [154] Huo Q, Song R, Lei D, Luo Q, Wu Z, Wu Z, Zhao X, Zhang F, Li L, Liu M 2020 IEEE Electron Device Lett. 41 497
- [155] Lin P, Li C, Wang Z, Li Y, Jiang H, Song W, Rao M, Zhuo Y, Upadhyay N K, Barnell M, Wu Q, Yang J J, Xia Q 2020 *Nat. Electron.* **3** 225

- [156] Jang J W, Park S, Burr G W, Hwang H, Jeong Y H 2015 IEEE Electron Device Lett. 36 457
- [157] Li C, Belkin D, Li Y, Yan P, Hu M, Ge N, Jiang H, Montgomery E, Lin P, Wang Z, Song W, Strachan J P, Barnell M, Wu Q, Williams R S, Yang J J, Xia Q 2018 Nat. Commun. 9 2385
- [158] Du C, Cai F, Zidan M A, Ma W, Lee S H, Lu W D 2017 *Nat. Commun.* 8 1
- [159] Yu J, Li Y, Sun W, Zhang W, Gao Z, Dong D, Yu Z, Zhao Y, Lai J, Ding Q, et al. 2021 2021 Symposium on VLSI Technology (IEEE) pp1–2
- [160] Yao P, Wu H, Gao B, Eryilmaz S B, Huang X, Zhang W, Zhang Q, Deng N, Shi L, Wong H S P, Qian H 2017 Nat. Commun. 8 15199
- [161] Yao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Yang J J, Qian H 2020 Nature 577 641
- [162] Ambrogio S, Narayanan P, Tsai H, Shelby R M, Boybat I, di Nolfo C, Sidler S, Giordano M, Bodini M, Farinha N C P, Killeen B, Cheng C, Jaoudi Y, Burr G W 2018 Nature 558 60
- [163] Wang Z, Joshi S, Savel'ev S, Song W, Midya R, Li Y, Rao M, Yan P, Asapu S, Zhuo Y, et al. 2018 Nat. Electron. 1 137
- [164] Fu Y, Zhou Y, Huang X, Dong B, Zhuge F, Li Y, He Y, Chai Y, Miao X 2022 Adv. Funct. Mater. 32 2111996
- [165] Fu Y, Zhou Y, Huang X, Gao B, He Y, Li Y, Chai Y, Miao X 2021 IEEE International Electron Devices Meeting (IEDM) San Francisco, USA, December 11, 2021 pp12.6.1-12.6.4
- [166] Hessel M, Modayil J, Van Hasselt H, Schaul T, Ostrovski G, Dabney W, Horgan D, Piot B, Azar M, Silver D 2018 AAAI Conference on Artificial Intelligence New Orleans Louisiana, USA Febrary, 2018 pp3215–3222

SPECIAL TOPIC—Physical electronics for brain-inspired computing Memristive brain-like computing^{*}

Wen Xin-Yu¹⁾²⁾ Wang Ya-Sai¹⁾²⁾ He Yu-Hui^{1)2)†} Miao Xiang-Shui^{1)2)‡}

 (Institute of Information Storage Materials and Devices, School of Integrated Circuits, Huazhong University of Science and Technology, Wuhan 430074, China)
 (Hubei Yangtze Memory Laboratories, Wuhan 430205, China)

(Received 11 April 2022; revised manuscript received 27 April 2022)

Abstract

With the rapid development of deep learning, the current rapid update and iteration of intelligent algorithms put forward high requirements for hardware computing power. Limited by the exhaustion of Moore's law and the von Neumann bottleneck, the traditional CMOS integration cannot meet the urgent needs of hardware computing power improvement. The utilization of new device memristors to construct a neuromorphic computing system can realize the integration of storage and computing, and has the characteristics of extremely high parallelism and ultra-low power consumption. In this work, the device structure and physical mechanism of mainstream memristors are reviewed in bottom-to-top order firstly, and their performance characteristics are compared and analyzed. Then, the recent research progress of memristors to realize artificial neurons and artificial synapses is introduced, including the simulation of specific circuit forms and neuromorphic functions. Secondly, in this work, the structural forms of passive and active memristive arrays and their applications in neuromorphic computing, including neural network-based handwritten digits and face recognition, are reviewed. Lastly, the current challenges of memristive brain-like computing from the bottom to the top, are summarized and the future development of this field is also prospected.

Keywords: memristor, artificial neuron, artificial synapse, neuromorphic computing PACS: 5.45.–a, 07.05.Mh, 84.35.+i, 87.18.Sn DOI: 10.7498/aps.71.20220666

^{*} Project supported by the National Key R&D Program of China (Grant No. 2019YFB2205100) and the National Natural Science Foundation of China (Grant Nos. 92064012, 61974051, 61874164, 51732003).

[†] Corresponding author. E-mail: heyuhui@hust.edu.cn

[‡] Corresponding author. E-mail: miaoxs@hust.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

柔性神经形态晶体管及其仿生感知应用

蒋子寒 柯硕 祝影 朱一新 朱力 万昌锦 万青

Flexible neuromorphic transistors and their biomimetric sensing application Jiang Zi-Han Ke Shuo Zhu Ying Zhu Yi-Xin Zhu Li Wan Chang-Jin Wan Qing 引用信息 Citation: Acta Physica Sinica, 71, 147301 (2022) DOI: 10.7498/aps.71.20220308 在线阅读 View online: https://doi.org/10.7498/aps.71.20220308 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

用于触觉感知的自供能可拉伸压电橡胶皮肤电子器件

Stretchable self-powered epidermal electronics from piezoelectric rubber for tactile sensing 物理学报. 2020, 69(17): 178701 https://doi.org/10.7498/aps.69.20200664

pH敏感有机电化学晶体管I-V特性及其电压依赖性

I-V characteristics and voltage dependence of pH-sensitive organic electrochemical transistors 物理学报. 2022, 71(13): 138501 https://doi.org/10.7498/aps.71.20220241

柔性基周期性厚度梯度薄膜的应变效应

Strain effects of periodic thickness-gradient films on flexible substrates 物理学报. 2021, 70(1): 016801 https://doi.org/10.7498/aps.70.20201008

双极型晶体管总剂量效应的统计特性

Statistical characteristics of total ionizing dose effects of bipolar transistors 物理学报. 2021, 70(13): 136102 https://doi.org/10.7498/aps.70.20201835

铁电负电容场效应晶体管研究进展

Recent research progress of ferroelectric negative capacitance field effect transistors 物理学报. 2020, 69(13): 137701 https://doi.org/10.7498/aps.69.20200354

基于蛋清栅介质的超低压双电层薄膜晶体管

Ultralow-voltage albumen-gated electric-double-layer thin film transistors 物理学报. 2018, 67(23): 237302 https://doi.org/10.7498/aps.67.20181539

专题: 面向类脑计算的物理电子学

柔性神经形态晶体管及其仿生感知应用*

蒋子寒 柯硕 祝影 朱一新 朱力 万昌锦† 万青‡

(南京大学电子科学与工程学院,南京 210093)

(2022年2月21日收到; 2022年3月22日收到修改稿)

生物感知系统具有高并行、高容错、自适应和低功耗等独特优点.采用神经形态器件实现生物感知功能的仿生,在脑机接口、智能感知、生物假体等领域具有重大应用前景.与其他神经形态器件相比,多端口神经 形态晶体管不仅可以同时实现信号的传输和训练学习,还可以对多路信号进行非线性的时空整合与协同调 控.然而,传统刚性神经形态晶体管很难实现弯曲变形以及和人体密切贴合,限制了神经形态器件应用范围. 所以,具有良好弯曲特性的柔性神经形态晶体管的研究成为了最近的研究重点.本文首先介绍了多种柔性神 经形态晶体管的研究进展,包括器件结构、工作原理和基本功能;另外,本文还将介绍上述柔性神经形态晶体 管在仿生感知领域中的应用;最后给出上述研究领域的总结和简单展望.

关键词:神经形态晶体管,仿生感知,突触晶体管,柔性电子学 PACS: 73.40.Qv, 85.30.Tv, 87.18.Sn, 87.85.fk

/ ±

DOI: 10.7498/aps.71.20220308

1 引 言

人类感知系统中具有无数的感觉受体,这些受体负责检测各种体内和环境的信息,并将其转化为低频电脉冲并传递到专门的认知感觉神经系统^[1]. 认知神经系统通过修改神经元之间的连接强度,实时地收集、整合、处理和存储信息,将海量的感官 信息转化为综合的认知和意识,这使得我们能够比 较准确和实时地认知外部世界并做出复杂的互动^[2,3]. 自然界优化后的生物感知系统,具有并行性、容错 性、自适应性和低功耗等优点^[4–9].生物突触和神 经元具有许多独特的特性,比如突触可塑性和时空 信息整合,这些特性被称为神经形态特性.模仿神 经形态特性的神经形态器件及其仿生感知可以极 大地推动脑机接口^[10,11]、智能感知^[12–15]、生物假 体^[16–18]等领域的发展.

人工突触和人工神经元分别能够对信号进行

© 2022 中国物理学会 Chinese Physical Society

记忆和整合,是仿生感知系统的重要组成部分.目 前研究人员已经提出了多种类型的人工突触器件, 例如基于两端孔忆阻器的人工突触[19-29]和基于三 端/多端晶体管的人工突触[30-37].神经形态晶体管 不仅可以实现信号的并行传输和训练学习,还可以 实现多信号的非线性时空整合[38,39].此外,通过调 控端的引入研究人员还可以实现对神经形态特性 的协同调控^[40,41]. 然而, 任何旨在与人体交互的设 备都需要与人体适当集成来确保设备具有长期稳 定的性能^[42]. 传统的刚性突触器件很难密切贴合 柔软、弯曲的人体,通常会在超低应变(约1%)下 断裂[43], 无法承受器官和组织的高生理应变, 严重 限制了该类器件的使用范围. 所以, 需要开发能够 与皮肤、器官、大脑等生物组织集成的、具有良好 机械柔性、延展性和生物相容性的柔性神经形态晶 体管.

本文将介绍三类柔性神经形态晶体管(电解质栅 突触晶体管、铁电栅突触晶体管、浮栅突触晶体管)

^{*} 国家自然科学基金 (批准号: 62074075, 62174082) 资助的课题.

[†] 通信作者. E-mail: cjwan@nju.edu.cn

[‡] 通信作者. E-mail: wanqing@nju.edu.cn

物理学报 Acta Phys. Sin. Vol. 71, No. 14 (2022) 147301

	1	51		1		
器件类型	衬底/栅介质/沟道	弯曲半径/mm	弯曲次数	生物相容性	神经功能	文献
	PET/Al ₂ O ₃ /IWO	20			EPSC、PPF、高通滤波	[44]
	$\rm PET/LiClO_4/MoSSe$	3	1000		图像识别、存储和处理	[45]
	—/离子凝胶/P3 CT	_			LIF、语音识别	[46]
中四手棚目上体	$\rm PEN/LiClO_4{+}PEO/SnO_2$	8	1000		学习规则、痛觉感知	[47]
电解原惯面体官	$\rm PEN/SiO_2 + PVA/CNTs$	3	1000		STP,LTP,LTD	[48]
	PI/壳聚糖/ITO	20		是	STDP、学习规则	[49]
	PET/蛋白质/ITO	10	3500	是	神经递质释放动力学	[50]
	PDMS/离子凝胶/P3 HT	_		_	学习规则	[51]
	—/P(VDF-TrFE)/并五苯	2.5	100		SRDP、STDP、图像识别	[52]
铁电晶体管	—/P(VDF-TrFE)/并五苯	$5{ imes}10^{-5}$			STDP	[53]
	云母/PZT/IGZO	4	400		数字识别	[54]
	PET/Al ₂ O ₃ +PMMA:C ₆₀ /并五苯	10	500	_	STP、LTP	[55]
巡押目休德	纸/Au+SiO _x /CNTs	_			STDP、图像识别	[<mark>56</mark>]
仔伽前件信	$\rm PET/ZrO_2+Al_2O_3/MoS_2$	10		_	学习规则	[57]
	PET/黑磷-QD+Al ₂ O ₃ /MoSSe	5	1000		学习规则	[58]

表 1 不同类型柔性突触晶体管比较 Table 1. Comparison of different types of flexible synaptic transistors.

的最新研究进展,包括器件的结构、原理和功能, 并将进一步介绍上述柔性神经形态晶体管在仿生 感知领域中的应用.最后,本综述还给出了上述研 究领域的总结和简单展望.表1显示了近期3种柔 性突触晶体管的代表性工作,并对其实现的神经功 能和机械柔性进行简要比较.由于生物相容性对于 突触晶体管在生物医疗等领域的应用具有重要意 义,所以表1中特别指出了使用具有生物相容性栅 介质的突触晶体管.

2 生物突触和生物神经元

大脑中的突触和神经元共同构成学习和记忆的最小单元^[50],神经元可以简单地认为是用来处理信息的,突触是神经元之间在功能上发生联系的部位,也是信息传递的关键部位^[60].如图 1(a)所示,含有细胞核的胞体是神经元的主体,树突是胞体延伸出来的树状凸起,负责接收并整合来自其他神经元的信息.轴突通常从胞体上延伸一段距离,是信息传递的主要通道.在轴突的末端会分裂出许多细枝,这些细枝能够与其他神经元的树突或者胞体形成特殊的纳米间隙 (20—40 nm),被称为突触,如图 1(b)所示.每个神经元通过突触与其他成千上万个神经元相连构成了神经网络^[61].

生物神经元通过突触接收来自其他生物神经 元的时空脉冲信号,并对接收到的信号进行整合.

如果整合后的信号强度达到阈值,神经元就会输出 动作电位,动作电位沿着轴突传播到其末端,刺激 轴突末端释放神经递质,神经递质通过扩散穿过突 触间隙与突触后膜上的受体结合,改变了突触后神 经元的膜电位,从而实现信号的传递,如果积分未 达到阈值,神经元的膜电位就会逐渐衰减至静态电 位. 根据突触后膜上受体的不同, 可将突触分为抑 制性突触和兴奋性突触两种. 在兴奋性突触中, 突 触前神经元释放的神经递质会使突触后神经元产 生兴奋性突触后电流/电位 (excitatory post-synaptic current/potential, EPSC/EPSP). 相反, 在抑 制性突触中,突触前神经元释放的神经递质会使突 触后神经元产生抑制性突触后电流/电位 (inhibitory postsynaptic current/potential, IPSC/IPSP). 至今为止,已经提出了一些神经元计算模型来模拟 神经元信号的整合和产生过程,比如 Hodgkin-Huxley (H-H)模型、integrate and fire (IF)模型、leaky integrate and fire(LIF) 模型^[62-64]. H-H 模型虽然 可以高精度地模拟生物神经元的行为,但是较为复 杂,包含了4个方程和数十个参数,很难在实际中 应用[65]. IF 模型和 LIF 模型通过将膜电位与阈值 进行比较来决定是否触发尖峰,忽略了神经元树突 的空间结构^[67].在LIF模型中,膜电位低于阈值会 在短时间内漏出,更接近真实的神经元.IF 模型是 LIF 模型的简化版本, 没有泄漏行为. 只要膜电位 继续累积超过阈值, 就会触发放电.



图 1 生物神经元 (a) 和生物突触 (b) 的结构示意图^[66] Fig. 1. Schematic diagram of biological neuron (a) and biological synapse (b)^[66].

神经元间的连接强度称为突触权重,突触权重 的重新配置 (即突触可塑性), 被认为是神经系统学习 和记忆的基础^[68]. 突触可塑性分为短程塑性 (shortterm plasticity, STP)和长程塑性. 短程塑性对应 的是刺激后突触权重的短暂改变,持续时间为几十 毫秒到几分钟,是神经网络实现时空相关计算功能 的生理基础^[69,70]. 双脉冲易化/抑制 (paired-pulse facilitation/depression, PPF/PPD) 是短程塑性的 一种重要表现形式. 由于短期塑性可以在短时间内 改变突触权重,因此突触可以通过增强或抑制突触 权重来充当滤波器[71,72]. 长程可塑性包括长程增强 (long-term potentiation, LTP) 和长程抑制 (longterm depression, LTD), 是突触权重持续数小时、 数年甚至更长时间的改变,会给神经网络带来永久 性的变化,为生物系统学习和记忆提供了生理基础[73]. 短程塑性可以通过充分训练或持续的神经元活动 转化为长程塑性,这种转变是突触连接结构变化的 结果[74,75].

时间尖峰依赖的可塑性 (spike-timing-dependent plasticity, STDP) 是 Hebbian 学习规则的基 本类型之一, 它描述了两个尖峰之间的时间间隔和 顺序对突触权重的影响, 在神经网络中具有将时间 信息转化为记忆存储的潜力^[76].人们普遍认为 STDP 是生物神经网络中无监督学习的基本机制^[77].频 率依赖的可塑性 (spike-rate-dependent plasticity, SRDP) 通过控制突触前脉冲的频率来调节突触权 重.根据 SRDP 学习规则, 高频的突触前脉冲会增 强突触后神经元的反应, 而低频的突触前脉冲会抑 制这种反应^[78-80].

3 柔性电解质栅突触晶体管

电解质栅晶体管 (electrolyte-gate transistor,

EGT) 具有低工作电压和与突触、神经元类似的动 力学行为等优点, 在神经形态电子学中引起了极大 的关注. EGT 的结构和场效应晶体管相同, 只是用 电解质替代了传统的栅介质, 在栅极电场的作用 下, 电解质中的阴阳离子发生定向移动, 并且分别 在电解质和栅极、电解质和沟道的界面处积累, 形 成具有μF/cm² 量级的巨大双电层 (electric-doublelayer, EDL) 电容^[81,82]. 当施加到栅极的电压相对 较低时, 由双电层电容对沟道电导进行易失性调 控. 当施加到栅极的电压足够高时, 电解质中的离 子可能穿过电解质和沟道的界面, 对沟道进行电化 学掺杂或者与沟道发生电化学反应, 从而实现对沟 道电导的非易失性调控^[42].

目前,常用电解质材料大致可以分为3类:1)无 机固态电解质,例如纳米颗粒SiO₂^[83-86]、Al₂O₃^[87,88]; 2)离子液或离子凝胶^[46,89-92];3)聚合物电解质或 聚电解质,例如LiClO₄/聚环氧乙烷 (polyethylene oxide, PEO)^[93-95]、壳聚糖^[96-99].

3.1 无机固态电解质

无机固态电解质栅晶体管能够和标准工艺兼容,并且具有较低的制备温度,在柔性电子领域具有巨大的应用前景.但是较低的制备温度也容易产生粉尘,可能会对设备造成损害.2013年,Wan等^[85]通过等离子体增强化学气相沉积法在铟锡氧(ITO)覆盖的聚对苯二甲酸乙二醇酯(polyethylene terephthalate, PET)衬底上沉积了一层掺磷的纳米颗粒SiO₂,通过简单的自组装方法制备了铟锌氧(IZO)源漏电极和沟道,成功地模拟了EPSC, PPF, LTP等重要突触塑性.这是对柔性人工突触的一次积极尝试,对构建神经形态系统非常有帮助.受到生物神经元中树突整合和尖峰操作的启发,他们还在柔性衬底上制备了用于 PH 传感器的多输入柔性氧

化物神经形态晶体管^[86].器件的结构示意图和柔 性神经形态晶体管的电容网络示意如图 2(a)所示. IZO 沟道中的载流子密度由传感栅和控制栅所有 输入加权和调制.器件工作在准静态双栅协同传感 模式时,对 pH 的灵敏度最高可达约 105 mV/pH. 该研究为生化检测提供了一种超低功耗、高灵敏 度、快速响应的新概念传感平台.



离子液是室温下呈液相的盐类,仅由阴阳离子 组成.离子凝胶通常是由嵌段共聚物溶解在离子液 中并凝胶化获得^[100],其丰富的空间网状结构有利 于离子的移动.离子液和离子凝胶具有相对更高的 比电容和更短的极化时间,能够实现更高的开关速



图 2 (a) 基于多栅 IZO 神经形态晶体管的柔性 pH 传感器的示意图^[86]; (b) 神经纤维-OECT 的装置结构示意图和 OECT-神经 纤维的照片, 插图: 离子在可渗透半导体中的掺杂机制示意图; (c) P3CT-神经纤维的 PSC 作为施加电压尖峰之间的时间间隔 (Δt) 的函数 ($V_{GS} = -0.7$ V, 100 ms); (d) P3CT-和 P3HT-神经纤维中超过 45 个周期的 LTP 和 LTD 循环测试; (e) 生物神经网络 和神经纤维晶体管网络示意图 (左), 10 × 10 P3CT-神经纤维阵列的照片 (右)^[46]

Fig. 2. (a) Schematic illustration of the flexible pH sensor based on an IZO neuromorphic transistor with multiple gate electrodes^[86]; (b) schematic of the device architecture for neurofiber-OECT and photograph of OECT-neurofiber, inset: schematic of the doping mechanism by ions in a permeable semiconductor; (c) PSC of a P3CT-neurofiber as a function of the time interval (Δt) between applied voltage spikes ($V_{\rm GS} = -0.7$ V, 100 ms); (d) cycle test of LTP and LTD in P3CT- and P3HT-neurofibers over 45 cycles; (e) schematic of biological neural network and neurofiber transistor network (left), photograph of a 10 × 10 array of P3CT-neurofibers (right)^[46].

度和更快的工作频率[101]. Kim 等[46] 提出了一种可 实现树突网络的神经纤维有机电化学晶体管,并特 别提出以羧酸官能化聚噻吩-聚 [(3-(6-羧己基) 噻 吩)-2, 5-二基] (poly[3-(6-carboxyhexyl) thiophene-2, 5-diyll, P3CT) 为沟道来增强记忆保留和循 环耐受性,器件结构如图 2(b) 所示.该器件不仅模 拟了 EPSC, LTP, LTD 等基本突触功能, 而且模 拟了神经元的 LIF 行为, 如图 2(c) 所示. 与聚 (3-己 基噻吩) [poly(3-hexylthiophene-2, 5-diyl), P3HT] 相比, P3CT 在长期稳定性和循环耐受性上具有显 著优势, 如图 2(d) 所示. 为了证实 P3CT-神经纤 维在纺织人工神经网络系统中的应用可行性,又展 示了一个包含 100 个不同突触的阵列, 连接了 10 个 突触前神经元和10个突触后神经元.如图2(e)所 示,这个神经网络阵列是由 10 个 P3CT-神经纤维 与10个栅级微纤维编织而成,成功演示了基于时 域的迭代尖峰神经网络学习的语音识别, 平均识别 准确率高达 88.9%.

延展性对于柔性突触晶体管至关重要,目前具 有延展性的柔性突触晶体管多数以离子凝胶为电 解质栅. 2018年, Lee 等^[102]制备了以离子凝胶作为 电解质栅、P3HT 纳米线为沟道的可拉伸突触晶体 管,这种突触晶体管即使沿着沟道长度或宽度方向 拉伸 100%,仍然能够保持稳定的性能. 2019年, Bao 等^[103] 制备了以离子凝胶为电解质栅、碳纳米管为 沟道的突触晶体管阵列.虽然在沿沟道长度方向拉 伸 20% 时,观察到了源漏电流的少许退化.但是在 沿沟道宽度方向拉伸 20% 时,器件的源漏电流基 本保持不变. 2022年, Liu 等^[104] 制备了以离子凝 胶为电解质栅、P3HT/PEO 纳米纤维为沟道的可 拉伸神经形态晶体管,即使在沿着沟道长度方向拉 伸 50% 的条件下,器件仍然具有稳定的突触功能. 所以,以离子凝胶为电解质栅的柔性突触晶体管在 仿生感知、软机器人、生物假肢等领域具有巨大的 应用潜力.

3.3 聚合物电解质或聚电解质

单纯的绝缘聚合物不能传输电子,是很好的栅介质材料^[106-108].但是不包含离子的纯聚合物在施加偏压时不会形成 EDL,所以通常将无机盐溶解在离子配位聚合物中形成聚合物电解质^[105,109,110]



图 3 (a) LiClO₄ 溶解在 PEO 中作为栅极电解质的突触晶体管的结构示意图; (b) 双脉冲易化, 插图: PPF 指数被绘制为两个脉 冲之间时间间隔的函数; (c) 由 40 个突触前脉冲触发的 EPSC; (d) LTP 和 LTD 的可重复性^[105]

Fig. 3. (a) Schematic of synaptic transistors with $LiClO_4$ dissolved in PEO as gate electrolyte; (b) paired-pulse facilitation, inset: PPF index is plotted as a function of time interval between the two pulses; (c) EPSC triggered by 40 presynaptic pluses; (d) repeatability of LTP and $LTD^{[105]}$. 或使用含有离子或可电离基团重复单元的聚电解 质[111-113] 作为电解质栅. 聚合物电解质中研究最多 的就是锂盐与 PEO 所构成的聚合物电解质, 例如 LiClO₄/PEO. 2020年, Zhu 等^[105] 在聚酰亚胺 (polyimide, PI) 衬底上制备了以 In₂O₃ 为沟道、LiClO₄/ PEO 为栅介质的柔性突触晶体管,具有良好的柔 韧性和机械可靠性,器件结构如图 3(a) 所示,成功 地模拟了突触的短程可塑性和短程可塑性到长程 可塑性的转换,如图 3(b)和图 3(c)所示.随后又 通过在栅极上施加一系列连续脉冲来改变沟道电 流,如图 3(d) 所示,这表明该突触晶体管具有可接 受的稳定性和循环耐受性. 最后, 为了进一步可视 化 In₂O₃ 突触晶体管中的短期和长期记忆, 使用 了 5×5 阵列来研究图像记忆. 为使用环保溶液制 备的非晶金属氧化物半导体制造人工突触器件提 供了一种新方法.

壳聚糖等天然聚合物的衍生物是用于 EGT 的 常见聚电解质,具有无毒、可生物降解、生物兼容 等优势^[114–117]. Ke 等^[97]制备了以壳聚糖/氧化石

墨烯复合薄膜为栅介质、铟镓锌氧 (IGZO) 为沟道 的自支撑神经形态晶体管,器件结构如图 4(a) 所示, 在设计智能警报系统和人工眼睛方面具有较大的 应用潜力. 光刺激角膜伤害感受器 (photoexcited corneal nociceptor, PCN) 的示意如图 4(b) 所示, 通过以光脉冲作为突触前刺激、IGZO 沟道中的电 流作为突触后响应,成功地模拟了这一生物功能. 光响应的原理如图 4(c) 所示. EPSC 的幅值被视 为 PCN 对外界刺激的响应,将 10 nA 定义为 PCN 的阈值. 当光刺激触发的 EPSC 达到或超过 10 nA 时, PCN会"活化"并触发警报程序,此时的光功 率称为触发阈值功率 (P_T). 用 50 个光脉冲来产生 受伤条件,受伤前后不同功率的单个光脉冲产生的 光电流如图 4(d) 所示. 可以看出, 在受伤情况下, PCN 会产生更大的光电流响应并且 PT 向较低处 移动,成功地模拟了异常性疼痛和痛觉过敏.此外, 还利用调控端 V_G 对 IGZO 沟道的横向调控, 成功 地模拟了中枢敏化和中枢镇痛作用,如图 4(e),(f) 所示.



图 4 (a) 自支撑光电神经形态晶体管示意图; (b) 光刺激角膜伤害感受器示意图; (c) IGZO 晶体管中光学响应的能带图; (d) PCN "受伤"前后实验测量的光电流; (e) 利用 $V_{\rm G} = 0.1$ V 模拟的中枢敏化, $P_{\rm T}$ 降至 4.98 nW/µm²; (f) 利用 $V_{\rm G} = -0.1$ V 模拟的镇痛 作用, $P_{\rm T}$ 增大到 17.62 nW/µm²^[97]

Fig. 4. (a) Schematic diagram of the freestanding photoelectric neuromorphic transistor; (b) schematic illustration of photoexcited corneal nociceptor; (c) energy-band diagrams of optical responses in IGZO-based transistor; (d) experimentally measured photocurrents of the PCN before and after "wounded"; (e) central sensitization simulated by $V_{\rm G} = 0.1$ V with $P_{\rm T}$ reduced to 4.98 nW/µm²; (f) analgesic effect simulated by $V_{\rm G} = -0.1$ V with $P_{\rm T}$ increased to 17.62 nW/µm²[97].

EGT 具有丰富的离子动力学特性,与其他器件结构相比,可以更好地模拟突触功能,其低工作电压特性也为超低能耗突触器件的实现提供了可能性.然而,其小动态范围和有限的保留特性限制了其进一步应用,此外,电解质的不稳定性(例如:离子液)也可能成为 ECT 实际应用的主要限制.

4 柔性铁电突触晶体管

柔性铁电场效应晶体管 (ferroelectric fieldeffect transistor, FeFET) 具有无损读出、低功耗 和高运行速度等优点,在非易失性存储、人工突触 等领域得到广泛应用^[118–122]. FeFET 使用铁电材 料作为栅介质,铁电材料可以随着电场的转换在两 种极化状态之间转换,这两种极化状态可用作存储 器的两种数字态^[123].由于铁电极化和载流子之间 的库仑作用,可以通过施加栅极电压来控制铁电栅 的极化状态,进而实现对载流子浓度的非易失性控 制^[124]. 2011 年,Müller 等^[125] 报道了 Hf_{0.5}Zr_{0.5}O₂ (HZO) 薄膜的铁电特性. HfO₂ 基铁电薄膜因其具 有良好的 CMOS 兼容性被广泛研究,然而,氧化物 铁电材料通常需要较高的结晶温度,极大的限制了 他们在柔性电子学领域的应用[126].

2020年, Li 等^[54]提出了一种由云母衬底、Sr-RuO₃ (SRO) 栅电极、PbZr02Ti08O3 (PZT) 铁电栅 介质和 IGZO 沟道组成的全无机 FeFET 来构建突 触器件,器件结构如图 5(a) 所示. 50 μs 的突触前 脉冲能够导致沟道电导急剧的增大,但无法维持, 这对应于短期可塑性,如图 5(b) 所示. 这种器件不 仅能实现编程电压脉冲对沟道电导接近线性的调 制,还具有优异的机械柔性和高温可靠性.弯曲半 径降至4mm、弯曲重复次数高达400次、弯曲持 续时间长达 7200 s 等条件对器件的 LTP, LTD 影响 很小,对网络的识别精度影响也很小,如图 5(c)— (h) 所示. 受益于器件的全无机结构, 在 100 ℃ 时, 器件的突触行为仍然完好无损,并且在不同温度下 的性能并没有明显变化.此外,由于该突触器件具 有线性、可重复和稳定的权重更新特性, MNIST (Modified National Institute of Standards and Technology) 数据集手写数字的识别准确率高达 94.4%. 这种柔性、稳定的全无机突触器件在未来 需要忍受环境干扰的高性能神经形态计算中具有 极大地应用潜力.



图 5 (a) 全无机柔性 FeFET 示意图; (b) 不同脉宽的突触前脉冲电压触发的 EPSC; 不同弯曲半径 (c)、不同弯曲循环次数 (d)、 不同弯曲时间 (e) 下的 LTP 和 LTD; (f)—(h) 对应的 MNIST 数字识别准确率^[54]

Fig. 5. (a) Schematic of the all-inorganic flexible FeFET; (b) EPSC triggered by presynaptic voltage pulse with different spike widths; LTP and LTD with the different bending radius, (c) different bending cycles (d), and different bending durations (e); (f)–(h) the corresponding MNIST digit recognition accuracy^[54].



图 6 (a) 以 P(VDF-TrFE) 薄膜为栅介质的自支撑有机神经形态晶体管的结构示意图; (b) 贴合在大脑形状模型 (上图) 和弯曲 半径为 50 μm 的 FONTs(下图) 照片; (c) 在 6000 次突触前脉冲期间, 折叠 FONTs 的 LTP 和 LTD 的重复转换, 上左、上右图分别 代表最初和最后的 10 个 LTP, LTD 循环^[53]

Fig. 6. (a) Schematic diagram of freestanding ferroelectric organic neuromorphic transistors with a P(VDF-TrFE) film as the dielectric layer; (b) photo images of the FONTs on the brain-shaped mold and folded FONTs with a bending radius of 50 μ m (lower panel); (c) repetitive transition between the LTP and LTD in the folded FONTs during 6000 spikes of presynaptic pulses (± 30 V for 500 ms), the left and the right in upper graph shows the LTP and LTD during the initial and final 10 cycles, respectively^[53].

聚偏二氟乙烯 (polyvinylidene fluoride, PVDF) 及其共聚物是应用最广泛的铁电聚合物,其中又以 聚 (偏二氟-三氟乙烯)[poly(vinylidene fluoride-cotrifluoroethylene), P(VDF-TrFE)] 的应用最为广 泛^[127-131]. Kim 等^[53] 报道了第1个基于自支撑铁电 有机神经形态晶体管 (ferroelectric organic neuromorphic transistors, FONTs) 的超薄人工突触,该 器件以 P(VDF-TrFE) 薄膜为栅介质、并五苯为沟 道,器件结构如图 6(a) 所示,总厚度只有 500 nm 左右. 通过简单的干剥离和粘贴方法,制备的器件 可以稳定地转移到各种不平整的衬底上. 另外,通 过精确调节 P(VDF-TrFE) 的剩余极化,成功地模 拟了 EPSC, LTP, LTD, STDP 等重要的突触性 能,并且 FONTs 在如图 6(b) 所示的恶劣情况下, 施加 6000 次突触前脉冲依然能获得稳定的 LTP, LTD 转换,结果如图 6(c) 所示,展现出在可穿戴 智能电子领域的巨大应用潜力.2020年,Lee 等^[132] 提出了一种能够感知、存储和学习各种触觉信息的 人工触觉学习铁电皮肤 (artificial tactile learning ferroelectric skin, ATFES),器件结构如图 7(a) 所 示,展示了 10000 次电脉冲输入期间 LTP, LTD 的 稳定转变、3.18% 的低循环变化性以及与压力脉冲数 量和幅度相关的突触可塑性等基本突触功能.此 外,构建了 4×4 的 ATFES 阵列用于对 3 种不同风 格手写字母的识别,如图 7(b) 所示.即使在考虑 10% 噪声的情况下,也有高达 99% 以上的识别准确率. 这种 ATFES 为设计具有容错触觉感知学习能力 的高精度人工智能电子皮肤提供了一条新的途径.

虽然铁电突触晶体管具有稳定性高、开/关比 大、权重更新曲线变化小等优点^[133].但是,制备大


图 7 (a) 生物触觉感知系统示意图 (左) 和人工触觉学习铁电皮肤的器件结构示意图 (右); (b) 三种不同手写风格 (N₁, N₂ 和 N₃) 的 "N" 图案示意图 (左) 和用于识别手写图案的单层神经网络的组成部分 (右)^[132]

Fig. 7. (a) Schematic of the biological tactile perception system (left) and schematic device structure of the artificial tactile learning ferroelectric skin (right); (b) schematic illustrations of "N" patterns with three different handwriting styles (N_1 , N_2 , and N_3) (left) and constituents of a single-layer neural network used to recognize handwriting pattern (right)^[132].

规模铁电突触阵列所需的高质量铁电薄膜是困难 的.此外,它们还难以实现出色的短程突触可塑性. 因此,需要进一步研究以解决这些问题.

5 柔性浮栅突触晶体管

浮栅场效应晶体管具有和传统场效应晶体管 相似的器件结构,区别在于浮栅场效应晶体管的栅 介质中间夹了一层存储功能层,称之为"浮栅".在 编程过程中,当栅极电压足够大并且隧穿层足够薄 时,可以通过量子隧穿效应或热发射将电荷注入到 浮栅上^[55,134,135].由于电荷阻挡层和隧穿层的存在, 浮栅中的电荷可以被非易失的存储,进而对沟道电 导进行非易失性的调制^[136].2009年,Someya等^[137] 在聚萘二甲酸乙二醇酯 (polyethylene-naphthalate, PEN) 衬底上制备了第一个柔性非易失性存储器 阵列,AlO_x(4 nm)和烷基膦酸自组装单层 (2 nm) 的复合层作为阻挡层和隧穿层,Al(20 nm)为浮栅. 得益于超薄的栅介质, 仅需较小的写/擦电压 (≤6 V) 就能产生大的、可逆的、非易失性的阈值电压偏移.

浮栅晶体管能够对沟道电导进行调制并且长 期保持的能力,可以用来有效地记录突触权重,因 此成为最流行的突触结构之一^[138-141]. Zhang 等^[57] 在柔性的 PET 衬底上制备了一种基于 MoS, 的光 电双调控异突触,器件结构如图 8(a) 所示. 该突触 器件具有超快的操作速度和超低的功耗 (LTP 中每 个尖峰消耗 18.3 aJ, LTD 中每个尖峰消耗 28.9 aJ). 除了 EPSC, LTP, LTD 和短程记忆到长程记忆的 转化以外,还模拟了学习-遗忘-再学习行为,如图 8(b) 所示. 与单独的电调制相比, 光电协同调制可以增 强 LTP 的高阶相关性,获得更大范围的突触权重. 但是,由于光和电刺激的协同调制,抑制效应减弱, 可以通过额外的电脉冲将突触权重恢复到初始状 态以进行对称调制,如图 8(c) 所示.为了研究该突 触器件的机械稳定性,测量了该器件在弯曲半径 为 10 mm 和 7.5 mm 条件下的转移曲线, 阈值电

压没有明显变化,并且成功模拟了在弯曲半径为 10 mm 的情况下的 EPSC, IPSC, LTP, LTD 等突触 可塑性. 应该注意的是,在弯曲状态下, LTP 和 LTD 没有任何退化,甚至观察到了一些增强,如图 8(d) 所示,证明了这种柔性人工突触装置在可穿戴应用 方面具有巨大的应用潜力.

基于连续型浮栅的突触晶体管不可避免地存 在电荷泄漏、耦合比低、电荷保持能力差等局限性^[132]. 使用分散的、不连续的浮栅,如纳米颗粒^[137,140,141]、 量子点材料^[142-144],可以显著地抑制浮栅中电荷的 横向泄漏,从而延长电荷保持时间. Han 等^[55] 首次 使用 C₆₀/聚 (甲基丙烯酸甲酯)[poly (methyl methacrylate), PMMA] 复合层作为栅介质,展示了能够 同时表现出信号传输和学习功能的人工柔性有机 突触晶体管,如图 8(e) 所示. C₆₀ 纳米颗粒通过溶 液法均匀分散在 PMMA 中,能够通过对两种载流 子电荷的捕获来实现阈值电压的双向迁移. 这种突 触晶体管具有 2.95 V 的存储窗口、大于 10³ 的电



图 8 (a) 光电双调控的柔性人工异突触示意图; (b) 由两个连续光脉冲序列模拟的学习-遗忘-再学习行为; (c) 光照条件下, 电脉冲产生的 LTP 和 LTD, 并且通过单独的电脉冲获得进一步的抑制; (d) 在平坦状态和弯曲状态 (*R* = 10 mm) 下, PSC 作为突触前脉冲数的函数^[57]; (e) C₆₀ 浮栅突触晶体管的示意图 (左) 和横截面 SEM 形貌图像 (右)^[55]

Fig. 8. (a) Schematic diagram of flexible artificial heterosynapse with photoelectric dual modulation; (b) learning, forgetting and relearning behaviors emulated by two sequences of consecutive light pulses; (c) electrical pulses induced the LTP and LTD under illumination of light, further depression was obtained by electrical pulse independently; (d) PSC as a function of pre-synaptic pulse number in a flat states and curved state $(R = 10 \text{ mm})^{[57]}$; (e) schematic representation (left) and cross-sectional SEM topography image (right) of a C₆₀ floating gate synaptic transistor^[55]. 流开/关比、超过 500 次的写入/擦除循环耐受性, 成功地模拟了包括 EPSC, PPF, PPD, LTP 和重 复学习过程在内的突触功能, 推动了基于纳米颗粒 的柔性浮栅人工突触晶体管的发展.

6 仿生感知应用

使用神经形态器件构建人工神经系统可以有 力地推动脑机接口、智能感知、生物假体等领域的 发展.而构建人工神经系统需要开发能够实时感知 外界刺激、对传感信息进行处理和存储,并做出反 应的智能仿生感知系统^[142–144].目前的仿生感知系 统一般由传感器、神经形态器件和转换组件构成, 并用以模仿特定的感知功能.通过将外部刺激(比 如:声音、压力、光、气体)转化为电信号^[145–150],然 后利用人工突触的权重调节对信号的进行整合或 记忆,产生与生物感知系统相似的响应,实现对外 界环境的仿生感知.与传统的感知系统相比,基于 柔性突触晶体管的仿生感知系统具有更简单的电 路、更低的功耗,同时又赋予系统可拉伸、可降解 等生物学特性[66,102,151,152].

人类的视觉系统对于生存和学习都至关重要, 大约80%的外部刺激是通过视觉感知获得的^[154,155]. 这是一个极为高效的过程,在大脑进行更复杂的行动之前,

视网膜对光的检测和图像信息的预处理是并 行操作的. 2021 年, Zhu 等^[153]报告了一种高集成 密度、对光具有非凡灵敏度的 32×32 柔性传感器 阵列.光电传感器同时充当光感受器和生物突触, 可以直接响应光刺激并进行预处理,器件结构如 图 9(a) 所示.碳纳米管 (carbon nanotubes, CNTs) 和 钙钛矿 CsPbBr₃(CPB)量子点 (quantum dot, QD) 组成的沟道在光生载流子的分离和传输中起关键 作用,实现了高响应度 5.1×10⁷ A/W 和超高比检 测 率 2×10¹⁶ Jones (1 Jones = 1cm · Hz^{1/2} · W⁻¹) 的超高比检测率.图 9(b) 描述了观察到陌生和熟 悉的面孔时,人类的视觉系统印象,该现象被传感 器阵列成功模拟.此外,还成功地演示了神经形态 强化学习功能,如图 9(c) 所示,理想输入图片和训 练的权重图之间的计算精度随着训练脉冲数量的



图 9 (a) 以 CNTs/CsPbBr3-QDs 为沟道的光电晶体管示意图; (b) 当观察到陌生和熟悉的面孔时, 人类视觉系统印象的示意图; (c) 不同训练脉冲数的训练权重结果; (d) 模拟人脸的学习过程^[153]

Fig. 9. (a) Schematic diagram of the phototransistor with a CNTs/CsPbBr3-QDs channel; (b) schematics illustration of the impression of human visual systems when unfamiliar and familiar faces are observed; (c) training weight results with different number of training pulses; (d) simulation of the learning process of a human face^[153]. 增加而增大.最后,又进行了人脸学习过程的模拟, 如图 9(d) 所示,随着训练脉冲数量的增加,可以学 习更多的面部特征.光感受器、存储元件和计算节 点组件在阵列中共享相同的物理空间,并行和实时 的处理信息,这为开发人工视觉系统提供了动力.

生物感觉系统的一个主要优点是能够整合两 种或多种感觉模式[156,157],这将有利于执行更复杂 的识别或决策任务. Wan 等[158] 报道了一种具有视 觉-触觉融合的双模人工感觉神经元 (bimodal artificial sensory neuron, BASE). 该 BASE 由 4 个核 心组件组成:电阻式压力传感器、基于钙钛矿的光 电探测器、基于水凝胶的离子电缆和突触晶体管, 如图 10(a) 所示. 光电探测器和压力传感器分别作 视网膜和皮肤中的受体,负责将外部的触觉和视觉 刺激转化为电信号,两种电信号通过离子电缆传输 到突触晶体管,以进行整合并转换为 EPSC. 他们 还制造了一种生物混合神经肌肉接头,用于传输来 自 BASE 的信号并支配骨骼肌管,进而模仿基于 视觉-触觉融合的身体运动控制, 如图 10(b) 所示. 如果位于机械手附近的网球可以(或不)传递视觉 或触觉反馈来触发机械手抓球,则将其标记为"是" (或"否"), 视觉反馈和触觉反馈分别用于区分在 *z* 轴和 *y* 轴上"是"或"否", 如图 10(c) 所示.

只有当球在两个方向都处于"是"位置时才能 抓住球.与单种感觉相比,融合感觉能够为机械手 提供更多维的信息,从而做出更合适的动作.随后 又使用融合的视觉和触觉线索来模拟多透明图像 识别,只有视觉-触觉融合矩阵才能充分提取形状 和透明度,如图 10(d)所示 (V:视觉; H:触觉; VH: 视觉-触觉融合).这项工作为在神经形态感知和神 经形态计算中开发人工多感觉整合神经提供了新 的见解,对未来的人机交互、机器人假肢和神经机 器人系统具有重要意义.

脑机接口是指在大脑和外部设备之间建立的 直接通信通道,这种通信通道不需要肌肉运动就可 以完成对用户神经活动的记录和分类,甚至可能参 与用户的神经活动.由于神经形态器件具有和生物 突触相似的信号传输,所以使用神经形态器件构建 的人工神经系统可以根据生物神经系统对生物信 号处理后得到的电输出对人工突触的突触权重进 行调节.这种突触权重的改变又可以通过人工突触 的突触后电流/电位反馈给生物神经系统,影响生



图 10 (a) 视觉-触觉融合的双模人工感觉神经元示意图; (b) 用于肌肉和机械手驱动的视觉-触觉融合示意图; (c) 视觉 (顶部, 粉红色) 和触觉 (底部, 蓝色) 反馈分别用于在 z 轴和 y 轴上推断 "是"或 "否"; (d) 单感觉模式和双感觉模式各自的识别率^[158] Fig. 10. (a) Schematic illustration of bimodal artificial sensory neuron with visual-haptic fusion; (b) schematic diagram of visual-haptic fusion for muscle and robotic hand actuation; (c) visual (top, pink) and haptic (bottom, blue) feedback used to infer "YES" or "NO" in z-axis and y-axis respectively; (d) the recognition rates of unimodal and bimodal modes, respectively^[158].

物神经系统对生物信号的下一步处理,实现了人工 神经系统和生物神经系统的双向信号传输,有希望 完美的实现脑机接口功能.所以,由神经形态器件 构成的仿生感知系统在生物医学工程、仿生机器人 等领域具有极大地应用前景.

7 结论与展望

为克服传统冯·诺依曼计算架构局限性[159-161], 在过去十多年中,新概念神经形态器件引起了研究 人员的极大关注. 以电解质为栅介质层的突触晶体 管具有丰富的离子动力学特性,可以很好地模拟生 物突触的基本特性. 与固态电解质栅相比, 液态电 解质栅具有更强的耦合能力和更短的响应/恢复时 间,也意味着液态电解质栅晶体管具有更低的工作 电压和更快的响应速度,但是液态电解质难于集成 和封装,在电路集成上可能会有更多的困难.不过 随着对突触晶体管研究的不断深入,未来可能出现 既具有较快响应速度和较低的工作电压又便于集 成和封装的优质电解质材料. 铁电晶体管具有稳定 性高、开/关比大、编程速度快、权重更新曲线变化 小等优点. 但是, 制备大规模铁电突触阵列所需的 高质量铁电薄膜是一个难点,这可能需要铁电薄膜 的制备工艺取得突破. 此外, 该类器件还难以实现 优良的短程可塑性,这可以通过减小施加在栅极脉 冲的脉宽、幅度和频率来有效缓解. 浮栅晶体管通 常具有可控且稳定的沟道电导和较大的开/关比, 可以很好地模拟突触的长程塑性,但也限制了其在 短程可塑性上的应用,使用超薄的隧穿层和窄带隙 材料作为浮栅可能是一个解决办法.此外,浮栅晶 体管通常还需要较大的工作电压,未来可能还需要 探索性能更加优异的高 κ 材料作为阻挡层和隧穿 层来减小工作电压.

目前对于柔性神经形态晶体管的研究仍然局 限在单个器件或小规模阵列,开发大规模集成的类 脑芯片来处理实际的人工智能任务仍然是个挑战, 这对器件的一致性、可靠性、可扩展性都提出了更 高的要求.此外,由于生物神经网络是一个高度复 杂的三维网络.因此,类脑芯片要实现类似生物神 经网络的复杂度,可能需要三维集成技术.具有良 好可弯曲特性的柔性神经形态器件为未来智能感 知、神经修复、软机器人等领域的发展带来了新的 机遇.目前的仿生感知系统还处在比较原始的实验 室研究阶段,还只能初步模拟生物对外界环境的传 感和响应过程,未来我们还需要进一步优化神经形 态器件特性并寻找能够实现多感知融合与集成的 技术方案,从而实现超低功耗智能感知系统的实际 应用.

参考文献

- [1] Jung Y H, Park B, Kim J U, Kim T I 2019 Adv. Mater. 31 1803637
- [2] Lumpkin E A, Caterina M J 2007 Nature 445 858
- [3] Abraira Victoria E, Ginty David D 2013 Neuron 79 618
- [4] Wan C J, Zhu L Q, Liu Y H, Feng P, Liu Z P, Cao H L, Xiao P, Shi Y, Wan Q 2016 Adv. Mater. 28 3557
- [5] Indiveri G, Chicca E, Douglas R J 2009 Cognit. Comput. 1 119
- [6] James C D, Aimone J B, Miner N E, Vineyard C M, Rothganger F H, Carlson K D, Mulder S A, Draelos T J, Faust A, Marinella M J, Naegle J H, Plimpton S J 2017 *Biol. Inspired Cogn. Archit.* **19** 49
- [7]~ Ho V M, Lee J A, Martin K C 2011 Science $\mathbf{334}$ 623
- [8] Machens Christian K 2012 Science 338 1156
- [9] Kuzum D, Yu S, Wong H S 2013 Nanotechnology 24 382001
- [10] Khodagholy D, Gelinas J N, Thesen T, Doyle W, Devinsky O, Malliaras G G, Buzsáki G 2015 *Nat. Neurosci.* 18 310
- [11] Viventi J, Kim D H, Vigeland L, Frechette E S, Blanco J A, Kim Y S, Avrin A E, Tiruvadi V R, Hwang S W, Vanleer A C, Wulsin D F, Davis K, Gelber C E, Palmer L, Van der Spiegel J, Wu J, Xiao J, Huang Y, Contreras D, Rogers J A, Litt B 2011 Nat. Neurosci. 14 1599
- [12] Xu L, Gutbrod S R, Bonifas A P, Su Y, Sulkin M S, Lu N, Chung H J, Jang K I, Liu Z, Ying M, Lu C, Webb R C, Kim J S, Laughner J I, Cheng H, Liu Y, Ameen A, Jeong J W, Kim G T, Huang Y, Efimov I R, Rogers J A 2014 Nat. Commun. 5 3329
- [13] Lipomi D J, Vosgueritchian M, Tee B C K, Hellstrom S L, Lee J A, Fox C H, Bao Z 2011 Nat. Nanotechnol. 6 788
- [14] Wang C, Hwang D, Yu Z, Takei K, Park J, Chen T, Ma B, Javey A 2013 Nat. Mater. 12 899
- [15] Kim D H, Ghaffari R, Lu N, 'et al. 2012 Proc. Natl. Acad. Sci. U. S. A. 109 19910
- [16] Moore D R, Shannon R V 2009 Nat. Neurosci. 12 686
- [17] Minev Ivan R, Musienko P, Hirsch A, et al. 2015 Science 347 159
- [18] Luo Y H L, da Cruz L 2016 Prog. Retinal Eye Res. 50 89
- [19] Terabe K, Hasegawa T, Nakayama T, Aono M 2005 Nature 433 47
- [20] Aono M, Hasegawa T 2010 Proc. IEEE 98 2228
- [21] Hasegawa T, Ohno T, Terabe K, Tsuruoka T, Nakayama T, Gimzewski J K, Aono M 2010 Adv. Mater. 22 1831
- [22] Zhao M, Li R, Xue J 2020 AIP Adv. 10 045003
- [23] Jeon Y R, Choi J, Kwon J D, Park M H, Kim Y, Choi C 2021 ACS Appl. Mater. Interfaces 13 10161
- [24] Tuma T, Pantazi A, Le Gallo M, Sebastian A, Eleftheriou E 2016 Nat. Nanotechnol. 11 693
- [25] Hua L, Zhu H, Shi K, Zhong S, Tang Y, Liu Y 2021 IEEE Trans. Circuits Syst. I, Reg. Papers 68 1599
- [26] Yao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Yang J J, Qian H 2020 *Nature* 577 641
- [27] Wen S, Wei H, Yan Z, Guo Z, Yang Y, Huang T, Chen Y

2020 IEEE Trans. Netw. Sci. Eng. 7 1431

- [28] Du F, Lu J G 2021 Appl. Math. Comput. 389 125616
- [29] Grollier J, Querlioz D, Camsari K Y, Everschor-Sitte K, Fukami S, Stiles M D 2020 Nat. Electron. 3 360
- [30] Jiang J, Guo J, Wan X, Yang Y, Xie H, Niu D, Yang J, He J, Gao Y, Wan Q 2017 *Small* **13** 1700933
- [31] Jiang S S, He Y L, Liu R, Zhang C X, Shi Y, Wan Q 2021 J. Phys. D-Appl. Phys. 54 185106
- [32] Beck M E, Shylendra A, Sangwan V K, Guo S, Gaviria Rojas W A, Yoo H, Bergeron H, Su K, Trivedi A R, Hersam M C 2020 Nat. Commun. 11 1565
- [33] Taube Navaraj W, Garcia Nunez C, Shakthivel D, Vinciguerra V, Labeau F, Gregory D H, Dahiya R 2017 *Front. Neurosci.* 11 501
- [34] Cho Y, Lee J Y, Yu E, Han J H, Baek M H, Cho S, Park B G 2019 Micromachines 10 32
- [35] Dai S, Wu X, Liu D, Chu Y, Wang K, Yang B, Huang J 2018 ACS Appl. Mater. Interfaces 10 21472
- [36] John R A, Liu F, Nguyen Anh C, Kulkarni M R, Zhu C, Fu Q, Basu A, Liu Z, Mathews N 2018 Adv. Mater. 30 1800220
- [37] Kim M K, Lee J S 2019 Nano Lett. 19 2044
- [38] Nishitani Y, Kaneko Y, Ueda M, Morie T, Fujii E 2012 J. Appl. Phys. 111 124108
- [39] Sun J, Oh S, Choi Y, Seo S, Oh M J, Lee M, Lee W B, Yoo P J, Cho J H, Park J H 2018 Adv. Funct. Mater. 28 1804397
- [40] Wang J, Chen Y, Kong L A, Fu Y, Gao Y, Sun J 2018 Appl. Phys. Lett. 113 151101
- [41] Jiang J, Hu W, Xie D, Yang J, He J, Gao Y, Wan Q 2019 *Nanoscale* 11 1360
- [42] Park H L, Lee Y, Kim N, Seo D G, Go G T, Lee T W 2020 Adv. Mater. 32 1903558
- [43] Rogers J A, Someya T, Huang Y 2010 Science 327 1603
- [44] Tiwari N, Rajput M, John R A, Kulkarni M R, Nguyen A C, Mathews N 2018 ACS Appl. Mater. Interfaces 10 30506
- [45] Meng J, Wang T, Zhu H, Ji L, Bao W, Zhou P, Chen L, Sun Q Q, Zhang D W 2022 Nano Lett. 22 81
- [46] Kim S J, Jeong J S, Jang H W, Yi H, Yang H, Ju H, Lim J A 2021 Adv. Mater. 33 2100475
- [47] Wei H, Ni Y, Sun L, Yu H, Gong J, Du Y, Ma M, Han H, Xu W 2021 Nano Energy 81 105648
- [48] Wang Y, Huang W, Zhang Z, Fan L, Huang Q, Wang J, Zhang Y, Zhang M 2021 Nanoscale 13 11360
- [49] Yu F, Zhu L Q, Xiao H, Gao W T, Guo Y B 2018 Adv. Funct. Mater. 28 1804025
- [50] Li Z Y, Zhu L Q, Guo L Q, Ren Z Y, Xiao H, Cai J C 2021 ACS Appl. Mater. Interfaces 13 7784
- [51] Wang X, Yan Y, Li E, Liu Y, Lai D, Lin Z, Liu Y, Chen H, Guo T 2020 Nano Energy 75 104952
- [52] Ham S, Kang M, Jang S, Jang J, Choi S, Kim T-W, Wang G 2020 Sci. Adv. 6 eaba1178
- [53] Jang S, Jang S, Lee E H, Kang M, Wang G, Kim T W 2019 ACS Appl. Mater. Interfaces 11 1071
- [54] Zhong G K, Zi M F, Ren C L, Xiao Q, Tang M K, Wei L Y, An F, Xie S H, Wang J B, Zhong X L, Huang M Q, Li J Y 2020 Appl. Phys. Lett. 117 092903
- [55] Ren Y, Yang J Q, Zhou L, Mao J Y, Zhang S R, Zhou Y, Han S T 2018 Adv. Funct. Mater. 28 1805599
- [56] Kim S, Choi B, Lim M, Yoon J, Lee J, Kim H D, Choi S J 2017 ACS Nano 11 2814
- [57] Wang T Y, Meng J L, He Z Y, Chen L, Zhu H, Sun Q Q, Ding S J, Zhou P, Zhang D W 2020 Adv. Sci. 7 1903480
- [58] Meng J L, Wang T Y, Chen L, Sun Q Q, Zhu H, Ji L, Ding S J, Bao W Z, Zhou P, Zhang D W 2021 Nano Energy 83

105815

- [59] Cao G M, Meng P, Chen J G, Liu H S, Bian R J, Zhu C, Liu F C, Liu Z 2021 Adv. Funct. Mater. 31 2005443
- [60] Zhu J D, Zhang T, Yang Y C, Huang R 2020 Appl. Phys. Rev. 7 011312
- [61] He Y L, Zhu L, Zhu Y, Chen C S, Jiang S S, Liu R, Shi Y, Wan Q 2021 Adv. Intell. Syst. 3 2000210
- [62] Hodgkin A L, Huxley A F 1952 J. Physiol. 117 500
- [63] Abbott L F 2008 Neuron **60** 489
- [64] Abbott L F 1999 Brain Res. Bull. **50** 303
- [65] Izhikevich E M 2004 IEEE Trans. Neural Netw. 15 1063
- [66] Burkitt A N 2006 Biol. Cybern. 95 1
- [67] Sun F, Lu Q, Feng S, Zhang T 2021 ACS Nano 15 3875
- [68] Choquet D, Triller A 2013 Neuron 80 691
- [69] Reyes A D 2011 Hear. Res. 279 60
- [70] Rotman Z, Deng P Y, Klyachko V A 2011 J. Neurosci. 31 14800
- [71] Fortune E S, Rose G J 2002 J. Physiol. Paris 96 539
- [72] Fortune E S, Rose G J 2000 J. Neurosci. 20 7122
- [73] Bliss T V P, Collingridge G L 1993 Nature 361 31
- [74] Kim M K, Lee J S 2018 ACS Nano 12 1680
- [75] Lamprecht R, LeDoux J 2004 Nat. Rev. Neurosci. 5 45
- [76] Fu Y, Kong L A, Chen Y, Wang J, Qian C, Yuan Y, Sun J, Gao Y, Wan Q 2018 ACS Appl. Mater. Interfaces 10 26443
- [77] Alibart F, Pleutin S, Bichler O, Gamrat C, Serrano-Gotarredona T, Linares-Barranco B, Vuillaume D 2012 Adv. Funct. Mater. 22 609
- [78] Yang Y, Wen J, Guo L, Wan X, Du P, Feng P, Shi Y, Wan Q 2016 ACS Appl. Mater. Interfaces 8 30281
- [79] Bear M F, Malenka R C 1994 Curr. Opin. Neurobiol. 4 389
- [80] Law C C, Cooper L N 1994 Proc. Natl. Acad. Sci. U. S. A. 91 7797
- [81] Yuan H, Shimotani H, Ye J, Yoon S, Aliah H, Tsukazaki A, Kawasaki M, Iwasa Y 2010 J. Am. Chem. Soc. 132 18402
- [82] Zhang L L, Zhao X S 2009 Chem. Soc. Rev. 38 2520
- [83] Wang J, Li Y, Liang R, Zhang Y, Mao W, Yang Y, Ren T L 2017 IEEE Electron Dev. Lett. 38 1496
- [84] Wan C J, Zhu L Q, Zhou J M, Shi Y, Wan Q 2014 Nanoscale 6 4491
- [85] Zhou J M, Wan C J, Zhu L Q, Shi Y, Wan Q 2013 IEEE Electron Dev. Lett. 34 1433
- [86] Liu N, Zhu L Q, Feng P, Wan C J, Liu Y H, Shi Y, Wan Q 2015 Sci Rep 5 18082
- [87] Wang X L, Shao Y, Wu X, Zhang M N, Li L, Liu W J, Zhang D W, Ding S J 2020 RSC Adv. 10 3572
- [88] Liang X, Li Z, Liu L, Chen S, Wang X, Pei Y 2020 Appl. Phys. Lett. 116 012102
- [89] Okaue D, Tanabe I, Ono S, Sakamoto K, Sato T, Imanishi A, Morikawa Y, Takeya J, Fukui K I 2020 J. Phys. Chem. C 124 2543
- [90] Ono S, Seki S, Hirahara R, Tominari Y, Takeya J 2008 Appl. Phys. Lett. 92 103313
- [91] Yang J T, Ge C, Du J Y, Huang H Y, He M, Wang C, Lu H B, Yang G Z, Jin K J 2018 Adv. Mater. 34 1801548
- [92] Eguchi K, Matsushita M M, Awaga K 2019 Phys. Chem. Chem. Phys. 21 18823
- [93] Sharbati M T, Du Y, Torres J, Ardolino N D, Yun M, Xiong F 2018 Adv. Mater. 30 1802353
- [94] Nikam R D, Kwak M, Lee J, Rajput K G, Hwang H 2020 Adv. Electron. Mater. 6 1901100
- [95] Qin J K, Zhou F, Wang J, Chen J, Wang C, Guo X, Zhao S, Pei Y, Zhen L, Ye P D, Lau S P, Zhu Y, Xu C Y, Chai Y 2020 ACS Nano 14 10018

- [96] Zhang C X, Li S, He Y L, Chen C S, Jiang S S, Yang X Q, Wang X R, Pan L J, Wan Q 2020 IEEE Electron Dev. Lett. 41 617
- [97] Ke S, He Y L, Zhu L, Jiang Z H, Mao H W, Zhu Y X, Wan C J, Wan Q 2021 Adv. Electron. Mater. 7 2100487
- [98] Chen C, He Y, Zhu L, Zhu Y, Shi Y, Wan Q 2021 *IEEE Trans. Electron Dev.* 68 3119
- [99] He Y L, Zhu Y, Chen C S, Liu R, Jiang S S, Zhu L, Shi Y, Wan Q 2020 IEEE Trans. Electron Dev. 67 5216
- [100] Kim S H, Hong K, Xie W, Lee K H, Zhang S, Lodge T P, Frisbie C D 2013 Adv. Mater. 25 1822
- [101] Cho J H, Lee J, Xia Y, Kim B, He Y, Renn M J, Lodge T P, Daniel Frisbie C 2008 Nat. Mater. 7 900
- [102] Lee Y, Oh J Y, Xu W, Kim O, Kim T R, Kang J, Kim Y, Son D, Tok J B H, Park M J, Bao Z, Lee T W 2018 Sci. Adv. 4 eaat7387
- [103] Molina-Lopez F, Gao T Z, Kraft U, Zhu C, Öhlund T, Pfattner R, Feig V R, Kim Y, Wang S, Yun Y, Bao Z 2019 *Nat. Commun.* 10 2676
- [104] Liu L, Xu W, Ni Y, Xu Z, Cui B, Liu J, Wei H, Xu W 2022 ACS Nano 16 2282
- [105] Zhu Y, Liu G, Xin Z, Fu C, Wan Q, Shan F 2020 ACS Appl. Mater. Interfaces 12 1061
- [106] Yao B W, Li J, Chen X D, Yu M X, Zhang Z C, Li Y, Lu T B, Zhang J 2021 Adv. Funct. Mater. 31 2100069
- [107] Ji D, Li T, Zou Y, Chu M, Zhou K, Liu J, Tian G, Zhang Z, Zhang X, Li L, Wu D, Dong H, Miao Q, Fuchs H, Hu W 2018 Nat. Commun. 9 2339
- [108] Stucchi E, Dell'Erba G, Colpani P, Kim Y H, Caironi M 2018 Adv. Electron. Mater. 4 1800340
- [109] Grey P, Pereira L, Pereira S, Barquinha P, Cunha I, Martins R, Fortunato E 2016 Adv. Electron. Mater. 2 1500414
- [110] Singaraju S A, Baby T T, Neuper F, Kruk R, Hagmann J A, Hahn H, Breitung B 2019 ACS Appl. Electron. Mater. 1 1538
- [111] Zhao D, Fabiano S, Berggren M, Crispin X 2017 Nat. Commun. 8 14214
- [112] Zhang W, Zhao Q, Yuan J 2018 Angew. Chem. Int. Ed. 57 6754
- [113] Kim Hyeong J, Chen B, Suo Z, Hayward Ryan C 2020 Science 367 773
- [114] Ali A, Ahmed S 2018 Int. J. Biol. Macromol. 109 273
- [115] Elgadir M A, Uddin M S, Ferdosh S, Adam A, Chowdhury A J K, Sarker M Z I 2015 J. Food Drug Anal. 23 619
- [116] Negm N A, Hefni H H H, Abd-Elaal A A A, Badr E A, Abou Kana M T H 2020 Int. J. Biol. Macromol. 152 681
- [117] Ways T M M, Lau W M, Khutoryanskiy V V 2018 Polymers 10 267
- [118] Ali T, Mertens K, Kuhnel K, Rudolph M, Oehler S, Lehninger D, Muller F, Revello R, Hoffmann R, Zimmermann K, Kampfe T, Czernohorsky M, Seidel K, Van Houdt J, Eng L M 2021 Nanotechnology 32 425201
- [119] Choi Y, Kim J H, Qian C, Kang J, Hersam M C, Park J H, Cho J H 2020 ACS Appl. Mater. Interfaces 12 4707
- [120] Tsai M F, Jiang J, Shao P W, Lai Y H, Chen J W, Ho S Z, Chen Y C, Tsai D P, Chu Y H 2019 ACS Appl. Mater. Interfaces 11 25882
- [121] Seo M, Kang M H, Jeon S B, Bae H, Hur J, Jang B C, Yun S, Cho S, Kim W K, Kim M S, Hwang K M, Hong S, Choi S Y, Choi Y K 2018 *IEEE Electron Dev. Lett.* **39** 1445
- [122] Jung S W, Koo J B, Park C W, Na B S, Oh J Y, Lee S S, Koo K W 2015 J. Vac. Sci. Technol. B 33 051201

- [123] Hoffman J, Pan X A, Reiner J W, Walker F J, Han J P, Ahn C H, Ma T P 2010 Adv. Mater. 22 2957
- [124] Nishitani Y, Kaneko Y, Ueda M, Fujii E, Tsujimura A 2013 Jpn. J. Appl. Phys. 52 04CE06
- [125] Müller J, Böscke T S, Bräuhaus D, Schröder U, Böttger U, Sundqvist J, Kücher P, Mikolajick T, Frey L 2011 Appl. Phys. Lett. 99 112901
- [126] Dai S L, Zhao Y W, Wang Y, Zhang J Y, Fang L, Jin S, Shao Y L, Huang J 2019 Adv. Funct. Mater. 29 1903700
- [127] Narayanan Unni K N, de Bettignies R, Dabos-Seignon S, Nunzi J M 2004 Appl. Phys. Lett. 85 1823
- [128] Kang S J, Park Y J, Bae I, Kim K J, Kim H C, Bauer S, Thomas E L, Park C 2009 Adv. Funct. Mater. 19 2812
- [129] Kim K L, Lee W, Hwang S K, Joo S H, Cho S M, Song G, Cho S H, Jeong B, Hwang I, Ahn J H, Yu Y J, Shin T J, Kwak S K, Kang S J, Park C 2016 Nano Lett. 16 334
- [130] Tian B B, Zhong N, Duan C G 2020 Chin. Phys. B 29 097701
- [131] Park C, Lee K, Koo M, Park C 2021 Adv. Mater. 33 2004999
- [132] Lee K, Jang S, Kim K L, Koo M, Park C, Lee S, Lee J, Wang G, Park C 2020 Adv. Sci. 7 2001662
- [133] Yu S 2018 Proc. IEEE 106 260
- [134] Van Tho L, Baeg K J, Noh Y Y 2016 Nano Converg. 3 10
- [135] Zhang H, Zhang Y T, Yu Y, Song X X, Zhang H T, Cao M X, Che Y L, Dai H T, Yang J B, Yao J Q 2017 ACS Photonics 4 2220
- [136] Cho S W, Kwon S M, Kim Y H, Park S K 2021 Adv. Intell. Syst. 3 2000162
- [137] Sekitani T, Yokota T, Zschieschang U, Klauk H, Bauer S, Takeuchi K, Takamiya M, Sakurai T, Someya T 2009 Science 326 1516
- [138] He Y L, Liu R, Jiang S S, Chen C S, Zhu L, Shi Y, Wan Q 2020 J. Phys. D: Appl. Phys. 53 215106
- [139] Yang X X, Yu J R, Zhao J, Chen Y H, Gao G Y, Wang Y F, Sun Q J, Wang Z L 2020 Adv. Funct. Mater. 30 2002506
- [140] Zhao T S, Zhao C, Xu W Y, Liu Y N, Gao H, Mitrovic I Z, Lim E G, Yang L, Zhao C Z 2021 Adv. Funct. Mater. 31 2106000
- [141] Park E, Kim M, Kim T S, Kim I S, Park J, Kim J, Jeong Y, Lee S, Kim I, Park J K, Kim G T, Chang J, Kang K, Kwak J Y 2020 Nanoscale 12 24503
- [142] Feng G, Jiang J, Zhao Y, Wang S, Liu B, Yin K, Niu D, Li X, Chen Y, Duan H, Yang J, He J, Gao Y, Wan Q 2020 Adv. Mater. 32 1906171
- [143] Basbaum A I, Bautista D M, Scherrer G, Julius D 2009 Cell 139 267
- [144] Lee Y, Lee T W 2019 Accounts Chem. Res. 52 964
- [145] Hou Y X, Li Y, Zhang Z C, Li J Q, Qi D H, Chen X D, Wang J J, Yao B W, Yu M X, Lu T B, Zhang J 2021 ACS Nano 15 1497
- [146] Li Y, Xuan Z H, Lu J K, Wang Z R, Zhang X M, Wu Z H, Wang Y Z, Xu H, Dou C M, Kang Y, Liu Q, Lv H B, Shang D S 2021 Adv. Funct. Mater. 31 2100042
- [147] Jang Y W, Kang J, Jo J W, Kim Y H, Kim J, Park S K 2021 Sens. Actuator B: Chem. 342 130058
- [148] Yang J C, Mun J, Kwon S Y, Park S, Bao Z, Park S 2019 Adv. Mater. 31 1904765
- [149] Wan C J, Liu Y H, Feng P, Wang W, Zhu L Q, Liu Z P, Shi Y, Wan Q 2016 Adv. Mater. 28 5878
- [150] Kim Y, Chortos A, Xu W, Liu Y, Oh J Y, Son D, Kang J, Foudeh A M, Zhu C, Lee Y, Niu S, Liu J, Pfattner R, Bao Z, Lee T W 2018 *Science* 360 998

- [151] Jiang S, He Y, Liu R, Chen C, Zhu L, Zhu Y, Shi Y, Wan Q 2021 IEEE T. Electron Dev. 68 415
- [152] Ling H, Koutsouras D A, Kazemzadeh S, van de Burgt Y, Yan F, Gkoupidenis P 2020 Appl. Phys. Rev. 7 011307
- [153] Zhu Q B, Li B, Yang D D, Liu C, Feng S, Chen M L, Sun Y, Tian Y N, Su X, Wang X M, Qiu S, Li Q W, Li X M, Zeng H B, Cheng H M, Sun D M 2021 Nat. Commun. 12 1798
- [154] Gao S, Liu G, Yang H, Hu C, Chen Q, Gong G, Xue W, Yi X, Shang J, Li R W 2019 ACS Nano 13 2634
- [155] Wang G, Wang R, Kong W, Zhang J 2018 Cogn. Neurodynamics 12 615

- [156] Yamamoto Y, Harada S, Yamamoto D, Honda W, Arie T, Akita S, Takei K 2016 Sci. Adv. 2 e1601473
- [157] Wu X, Li E, Liu Y, Lin W, Yu R, Chen G, Hu Y, Chen H, Guo T 2021 Nano Energy 85 106000
- [158] Wan C, Cai P, Guo X, Wang M, Matsuhisa N, Yang L, Lv Z, Luo Y, Loh X J, Chen X 2020 Nat. Commun. 11 4602
- [159] Shastri B J, Tait A N, de Lima T F, Pernice W H P, Bhaskaran H, Wright C D, Prucnal P R 2021 Nat. Photonics 15 102
- [160] Song S, Kim J, Kwon S M, Jo J W, Park S K, Kim Y-H 2021 Adv. Intell. Syst. 3 2000119
- [161] Komar M S 2017 Autom. Control Comp. Sci. 51 701

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Flexible neuromorphic transistors and their biomimetric sensing application^{*}

Jiang Zi-Han Ke Shuo Zhu Ying Zhu Yi-Xin

Zhu Li Wan Chang-Jin[†] Wan Qing[‡]

(School of Electronic Science & Engineering, Nanjing University, Nanjing 210093, China)
 (Received 21 February 2022; revised manuscript received 22 March 2022)

Abstract

Biological perception system has the unique advantages of high parallelism, high error tolerance, selfadaptation and low power consumption. Using neuromorphic devices to emulate biological perceptual system can effectively promote the development of brain-computer interfaces, intelligent perception, biological prosthesis and so on. Compared with other neuromorphic devices, multi-terminal neuromorphic transistors can not only realize signal transmission and training learning at the same time, but also carry out nonlinear spatiotemporal integration and collaborative regulation of multi-channel signals. However, the traditional rigid neuromorphic transistor is difficult to achieve bending deformation and close fit with the human body, which limits the application range of neuromorphic devices. Therefore, the research of flexible neuromorphic transistor with good bending characteristics has become the focus of recent research. Firstly, this review introduces the research progress of many kinds of flexible neuromorphic transistors, including device structure, working principle and basic functions. In addition, the application of the flexible neuromorphic transistor in the field of bionic perception is also introduced. Finally, this review also gives a summary and simple prospect of the above research fields.

Keywords: neuromorphic transistors, bionic perception, synaptic transistors, flexible electronicsPACS: 73.40.Qv, 85.30.Tv, 87.18.Sn, 87.85.fkDOI: 10.7498/aps.71.20220308

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 62074075, 62174082).

[†] Corresponding author. E-mail: cjwan@nju.edu.cn

[‡] Corresponding author. E-mail: wanqing@nju.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

应用于感存算一体化系统的多模调控忆阻器

张宇琦 王俊杰 吕子玉 韩素婷

Multimode modulated memristors for in-sensor computing system Zhang Yu-Qi Wang Jun-Jie Lü Zi-Yu Han Su-Ting 引用信息 Citation: Acta Physica Sinica, 71, 148502 (2022) DOI: 10.7498/aps.71.20220226 在线阅读 View online: https://doi.org/10.7498/aps.71.20220226 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

一种适用于大规模忆阻网络的忆阻器单元解析建模策略

An analytic modeling strategy for memristor cell applicable to large-scale memristive networks 物理学报. 2021, 70(17): 178505 https://doi.org/10.7498/aps.70.20210116

高温压电材料、器件与应用

Review of high temperature piezoelectric materials, devices, and applications 物理学报. 2018, 67(20): 207701 https://doi.org/10.7498/aps.67.20181091

蛋白质基忆阻器研究进展

Research progress of protein-based memristor 物理学报. 2020, 69(17): 178702 https://doi.org/10.7498/aps.69.20200617

电极材料及偏压极性对氧化物介质击穿行为的影响及机制

Effects of electrode materials and bias polarities on breakdown behaviors of oxide dielectrics and their mechanisms 物理学报. 2021, 70(8): 087302 https://doi.org/10.7498/aps.70.20201262

基于二维材料MXene的仿神经突触忆阻器的制备和长/短时程突触可塑性的实现

Fabrication of synaptic memristor based on two-dimensional material MXene and realization of both long-term and short-term plasticity

物理学报. 2019, 68(9): 098501 https://doi.org/10.7498/aps.68.20182306

专题: 面向类脑计算的物理电子学

应用于感存算一体化系统的多模调控忆阻器*

张宇琦1)# 王俊杰2)# 吕子玉2) 韩素婷2)†

(深圳大学微纳光电子学研究院,深圳 518060)
 (深圳大学电子与信息工程学院,深圳 518060)

(2022年2月2日收到; 2022年3月4日收到修改稿)

交互式人工智能系统的构建依赖于高性能人工感知系统和处理系统的开发. 传统的感知处理系统传感器、存储器和处理器在空间上是分离的, 感知数据信息的频繁传输和数据格式转换造成了系统的长延时与高能耗. 受生物感知神经系统的启发, 耦合感知、存储、计算功能的感存算一体化技术为未来感知处理领域提供了可靠的技术方案. 具有感知光、压力、化学物质等能力的忆阻器是应用于感存算一体系统的理想器件. 本文从器件层面综述了应用于感存算一体化系统忆阻器的研究方向和研究进展, 包括视觉、触觉、嗅觉、听觉和多感官耦合类别, 并在器件、工艺与集成、电路系统架构和算法方面指出现阶段的挑战与展望, 为未来神经形态感存算一体化系统的发展提供可行的研究方向.

关键词: 感存算一体化, 忆阻器, 人工突触, 传感器 **PACS**: 85.35.-p, 73.40.Rw, 84.30.-r, 87.19.lt

1 引 言

人工智能领域现阶段取得的成功引发了学习 和模仿生物感知和处理系统来构建未来交互式智 能系统的热潮.生物神经系统主要分为两部分:中 枢神经系统和外围神经系统.中枢神经系统负责完 成高阶任务,如学习记忆,外围神经系统也被称为 感知系统,负责感知外界刺激(化学物质、光、压力 等)并将信息传到中枢神经系统和身体其他部位^[1]. 感知系统可以直接进行一些低级的处理,如本能反 应,即感知信号不需要发送到大脑系统就可做出响 应.感知系统的边缘处理不仅能快速响应外部刺激 以维持正常的生理活动,还能减少大脑的计算负担^[2]. 神经网络具有存内计算和大规模并行处理等特性^[3]. 这些特性使我们在面对现实世界的事件时,能够以 稳健容错的方式做出合适的反应.受生物中枢神经 **DOI:** 10.7498/aps.71.20220226

系统和感知系统处理模式的启发,一些科研人员将研究重点放在构建仿生物的集感知、存储、处理功能于一体(感存算一体化)的交互式系统^[4-7].

在传统的系统架构中,由于不同的功能和制造 工艺,传感器、存储器和处理器在空间上是分离的. 传感器收集现实环境中的模拟信号,模拟信号经过 预处理转换为数字信号,随后输入到冯·诺依曼型 数字计算架构中的存储器和处理器进行存储和后 续处理^[8,9].随着识别、分类等任务复杂度的提高, 传统感知系统会产生大量的未处理原始数据,这加 重了处理系统的工作负荷,而且数据的频繁传输和 格式转换极大地限制了集成系统的工作性能^[10]. 按感存算一体化系统的理念,系统边缘基本单元应 耦合感知、存储、处理功能于一体,并应具有低功 耗、小面积、高响应速度等特性^[9].感存算一体化系 统基本单元的设计、制造和应用的研究对于实现交 互式人工智能系统是至关重要的.

^{*} 国家自然科学基金 (批准号: 62122055, 62074104, 61974093) 资助的课题.

[#] 同等贡献作者.

[†] 通信作者. E-mail: sutinghan@szu.edu.cn

^{© 2022} 中国物理学会 Chinese Physical Society

忆阻器是近年来备受关注的一类新型存储器 件,具有操作速度快、功耗低、集成密度高等优势. 其电阻状态可以由所施加过的电激励历史控制,数 据信息可以以电阻的形式存储.此外,这种基于电 阻的存储器本质上就通过物理法则(如欧姆定律和 基尔霍夫定律)来处理信息,融合了存储和处理功 能(存内计算)[11-15]. 忆阻器与生物突触在结构、物 理行为和功能上有着极高的相似性, 被广泛应用于 构建可模拟生物突触可塑性的人工突触. 基于忆阻 器的人工突触是人脑启发的硬件类脑神经网络电 路的基本存储处理单元,神经形态的忆阻器突触网 络具备处理复杂高阶任务的能力[16-19].显然,将传 感功能引入已具备存算一体特性的忆阻器突触是 一种实现感存算一体化单元的策略. 将仿生感知技 术和神经形态工程技术相结合来实现感存算功能 一体化,将会开启人工智能的新时代.近年来,国 内外许多研究组专注于开发传感器与忆阻器突触 的集成,或融合传感功能的忆阻器突触.多模(光、 压力、化学物质等)调控的忆阻器突触为实现感存 算一体系统提供了切实可行的解决方案^[20-25].

本文首先介绍感存算一体化系统的工作机制 和应用场景、忆阻器及人工突触,然后从器件层面 综述了应用于感存算一体化系统的忆阻器突触的 研究方向和研究进展,包括视觉、触觉、嗅觉、听觉 和多感官耦合的类别.最后在器件、工艺与集成、 电路系统架构和算法4个方面指出现阶段的挑战 与展望,为未来神经形态感存算一体化系统的发展 提供可行的建议.

2 感存算一体化系统

仿生机器人系统、无人驾驶汽车、可穿戴医疗 设备等众多新兴领域的出现加速了人工智能技术 的发展,同时也对感知网络系统提出了更高的要 求.感知系统是从复杂环境中获取信息不可或缺的 部分,应用于人工感知网络的传感器节点的个数正 在飞速增长,伴随而来的是大量的原始非结构化的 冗余数据.传统的感知系统包含接收信号、数据转 化预处理、数据传输功能,数据被传输至当地计算 系统或云平台进行后续的高阶计算.传感大多发生 在有噪声的模拟信号域,而计算通常采用传统的 冯·诺依曼计算体系结构进行数字化执行,其对应 的架构如图 1(a)所示.由于功能需求和制造技术 的差异,感知系统与计算系统在物理上分离.这种基于数据转换和传输的处理策略极大影响了系统的能耗、响应速度、通信带宽和安全性等方面⁹.

相比而言,生物感知处理系统具有完备的层次 结构、计算和存储一体化特性以及复杂的神经网 络,能够高效地处理来自复杂环境的信息.视觉、 触觉、听觉、嗅觉和味觉这5种基本感觉通过人体 大脑中的神经网络相互作用产生信息(图 1(b)), 使人们能够探索、学习和适应世界. 受生物系统的 启发,科研人员致力于深度挖掘生物感知系统的工 作原理,应用其原理至半导体器件中实现新型的传 感功能器件,进而取代或者加强传统的传感器^[26]. 科研人员提出一种有潜力的方案是构建感知、存 储、计算一体化的人工智能系统,在感存算一体系 统中,感知单元融合了存储计算功能,具有额外的 初步学习和认知能力,其框架图如图1(c)所示.神 经形态感知系统能够从捕获的原始数据中提取有 用的感知信息并减少冗余数据的传输,可以降低系 统功耗并提供了一种高效的并行计算方法来实时 处理高吞吐量的时空数据. 感知端的边缘计算包含 低级处理功能和高级处理功能. 低级处理功能可以 优化难以识别的原始和非结构化数据中的特征,包 括噪声抑制、滤波、特征增强等,如图 1(d) 所示. 高级处理功能可以抽象输入的数据,通过特定算法 得到计算结果,完成如识别、分类、定位等任务.为 了加速神经网络学习算法的硬件实现,即感存算一 体单元阵列直接执行高阶任务, 传感器阵列应具有 可重构响应度特性. 如图 1(e) 所示, 系统边缘有 $m \times n$ 个感存算单元, 感知激励输入为 I 矢量, R为响应度矩阵,则最终的电流输出矩阵 O 可根据 基尔霍夫定律叠加得到. 在这里, 每个基本单元对 于输入刺激源的响应度是可调控更新的,类似于神 经网络中的突触权重,外界刺激输入至训练好的感 存算单元阵列即可得到乘累加的高阶操作结果.

如图 1(f) 所示, 感存算一体化技术在需要交互 式人工智能系统的领域具有重要的应用场景, 如自 动驾驶、可穿戴柔性电子设备、仿生机器人、医疗 检测等, 用于感知、存储、计算的神经形态器件的 开发极大地促进了人工智能领域的发展. 该应用领 域需要更多的跨学科合作, 要实现感存算一体系统 关键的第一步即是从材料/器件层级设计出符合预 期功能的基本工作单元.



图 1 (a) 传统的感知处理系统架构; (b) 人体五感示意图; (c) 感存算一体化系统架构; (d) 低级感官处理功能; (e) 用于神经网络计算的可重构响应度的感存算一体单元阵列; (f) 感存算一体化技术的应用领域

Fig. 1. (a) Traditional architecture of sensing and processing; (b) schematic of human sensory system; (c) in-sensor computing architecture; (d) low-level sensory processing functions; (e) in-sensor computing units with reconfigurable responsivity for neural network computing; (f) application fields of in-sensor computing technology.

3 忆阻器突触

3.1 忆阻器

非线性电子元件忆阻器的概念最初是由加州 大学伯克利分校蔡少棠教授提出的,而后惠普实验 室成功制备了忆阻器原型器件^[27].典型的两端忆 阻器具有金属-介质层-金属的结构,外部激励场可 重构介质层的物理结构状态,调控的状态取决于所 施加过的外部激励历史(图 2(a)).具有不同物理结 构的忆阻器呈现出不同的电阻状态, 忆阻器在施加 激励的过程中表现出电阻变化行为, 这赋予了器件 基本存储器的特性^[14].基础材料和器件研究表明, 忆阻器中的物理重构过程主要是由内部离子迁移 再分布引起的, 介质层中导电细丝的形成与断裂改 变了器件的导电性. 该物理过程根据响应时间尺度 可分为突发的或渐进的, 分别对应于数字型忆阻器 和模拟型忆阻器.数字型忆阻器具有两个区分度明 显的离散阻态, 以存储器的角度看, 低电阻状态与 高电阻状态可分别对应逻辑 1 和逻辑 0(图 2(b)).



图 2 (a) 两端忆阻器示意图; (b) 数字型忆阻器的典型电压-电流曲线; (c) 模拟型忆阻器的典型电压-电流曲线; (d) 忆阻器常见 机理; (e) 数字型和模拟型忆阻器的应用

Fig. 2. (a) Schematic of a two-terminal memristor; (b) typical I-V curve of digital memristor; (c) typical I-V curve of analog memristor; (d) three main mechanisms of memristors; (e) application of analog and digital memristor.

开启操作定义为器件从高阻态转变为低阻态, 而关闭操作定义为器件从低阻态恢复为高阻态. 在电压扫描下模拟型忆阻器具有连续型变化的阻态, 可被看作多比特存储器 (图 2(c)).

如图 2(d) 所示, 忆阻器的物理机理可主要分 为三大类^[28]: 一是导电细丝型^[29-31], 电场可以驱动 阳/阴离子在介质层中迁移进而控制连接两端电极 的导电细丝的形成与断裂, 器件的电阻状态取决于 导电细丝的尺寸和稳定程度; 二是非导电细丝型^[32,33], 器件的电阻状态由介质层中载流子捕获/解捕获或 离子迁移调制的界面肖特基/隧穿势垒决定; 三是 晶相变化^[34,35], 与常见的忆阻器中局部离子迁移物 理机量不同, 外部电刺激产生的热效应会使相变忆 阻器的相变层在非晶相 (高阻态) 和结晶相 (低阻 态) 之间切换. 基于其他效应 (如铁电效应、磁电效 应) 的器件也表现出典型的忆阻特性, 然而这些类 型的器件仍需大量的科学研究去改进其性能.

鉴于其快操作速度、结构简单、低功耗、集成 度高等优势,忆阻器在下一代存储计算技术领域展 现出巨大的潜力.在传统的冯·诺依曼计算机架构 中,数据处理单元与存储单元在空间上是分离的, 数据需要在处理器与存储器之间的反复传输极大 地限制了计算机的处理能力.存内计算设计的理念 是应用集存储与计算功能一体的单元构建高并行 度高性能的计算系统.忆阻器除了数据存储能力, 本质上就可以通过物理法则(欧姆定律、基尔霍夫 定律)对信息计算处理.因此,忆阻器是一种模糊 了存储与计算的边界的存储器技术,器件本身展示

出了类人脑存内计算的概念,是实现了存算一体化 功能的基本单元. 如图 2(e) 所示, 数字型忆阻器和 模拟型忆阻器有不同的应用领域. 数字型忆阻器常 被用于数字逻辑门的实现. 布尔逻辑的状态 (1 和 0) 映射为忆阻器的电阻状态 (低阻态和高阻态), 即输入和输出用忆阻器的电阻状态表示,基于数字 型忆阻器的逻辑电路具备逻辑门和锁存器的特性. 模拟型忆阻器阵列则被应用于在单个计算周期内 实现乘累加计算. 电压刺激输入至忆阻器阵列的行 后,模拟型忆阻器单元充当了矩阵点乘运算的权重 值,由欧姆定律可知每个忆阻单元的电流为输入电 压与器件电导的乘积 (乘法运算), 由基尔霍夫定律 可知忆阻器阵列的列输出电流为对应节点的累加 电流 (累加运算). 高并行度以及高吞吐量的特性使 得模拟型忆阻器非常适合加速计算密集型应用(如 硬件人工神经网络)[12].

3.2 人工突触

为了突破冯·诺依曼瓶颈,研究人员致力于开 发模拟人脑处理信息、学习、记忆的智能处理系统. 脑神经系统的高效性主要突显在大脑皮质中大型 复杂的神经元互连网络,其中包含了约1011个神 经元及1015个神经突触.在大脑系统中,神经元负 责从前神经元收集并处理输入信号,随后输出动作 电位到后神经元;突触则是前神经元与后神经元之 间的间隙,它是神经系统中调控信号传递的基本单 元. 突触权重是衡量神经元间的连接强度的参数, 其作用体现在一次信号传输过程中控制着突触前 膜的神经递质囊泡的数量及大小. 突触的性质 (功 能、权重等)发生变化的现象被称为突触可塑性, 是大脑学习与记忆的神经分子基础,模拟该生物功 能是实现神经形态电路的关键部分. 突触可塑性按 记忆时间长短可分为短时程可塑性 (short-term plasticity, STP) 和长时程可塑性 (long-term plasticity, LTP), 还有放电时间依赖可塑性 (spikingtiming-dependent plasticity, STDP)、放电速率依赖 可塑性 (spiking-rate-dependent plasticity, SRDP)、 经验学习等[36]. 在生物学中, 突触权重参数的增大 和减小对应着增强和抑制. 人脑的短时记忆一般持 续毫秒到数分钟,突触的 STP 行为通过刺激被短 暂地维持,突触权值短暂增加或减少随后迅速恢复 到初始状态. 然而, 重复输入脉冲刺激会产生一个 永久的变化,称为 LTP 状态. 这种依赖于外界活动的 调制特性促进了信息在人脑中的处理和存储. STP 和 LTP 是大脑中最主要的两种突触可塑性的形式. STP 往往是短期或一些突发性活动引起,可以帮助大脑过滤掉部分非必要的信息,而 LTP 则是在短时记忆的基础上经过反复训练形成的永久性记忆, STP 到 LTP 的转化需要充分的训练^[37].

基于互补金属氧化物半导体 (complementary metal-oxide-semiconductor transistor, CMOS) 技 术来构建单个神经形态计算单元需要超过10个晶 体管及电容、电阻等其他电子元件,功耗及面积远 超人脑,因而传统 CMOS 器件并不是构建高能效 智能类脑系统的理想技术方案.相比而言,忆阻器 的出现大大促进了硬件神经形态电路飞速发展. 忆 阻器与生物突触有着极高的相似性:在结构上,忆 阻器的两端电极对应着突触前膜和突触后膜, 忆阻 器中的介质层对应着突触;在物理行为上,外部刺 激使得忆阻器内部的离子迁移引起介质层物理重 构,生物突触则是通过释放神经递质在突触间隙中 传递;在功能上,模拟型忆阻器的电阻状态可以被 外场刺激调控,对应着生物突触的可塑性.鉴于忆 阻器的诸多优势,许多科研人员致力于构建基于忆 阻器的硬件形态神经网络电路系统[16].

4 感存算一体单元

近年来,大量的科学研究专注于开发出多模式 调控的忆阻器突触将其应用于未来感知、存储、计 算一体化的研究方向.本节介绍应用在各种人工感 知系统的忆阻器单元,包括视觉^[38-60]、触觉^[61-68]、 嗅觉^[69-75]、听觉^[76-79]和多模感知领域^[41,80,81].

4.1 视觉感存算一体单元

视觉是人类重要的一种感官,近一半的大脑皮 层忙于处理视觉信息,通过视觉可以判断物体的大 小、形状、颜色、亮度、距离、位置、光滑度、粗糙度 等.传统机器视觉系统通常由3个独立单元组成, 包括图像传感器(光电探测器)、存储器和处理单 元.相比之下,人类视觉系统具有传感和处理功能 相结合的能力.在人类视觉系统中,视网膜在接受 光信号的同时会对图像进行初步预处理,随后将提 取到的精简信息传递到视觉皮层进行更复杂的信 息处理(图3(a)).为了模拟该生物视觉成像过程, 科研人员首先研究开发光调控的忆阻器突触器 件.光调控忆阻器突触在直接响应光学刺激的同时,



图 3 (a) 人类视觉系统示意图; (b) 突触、神经元和制备的忆阻器示意图; (c) 大脑 STP 和 LTP 行为的示意图; (d) 人工突触在 红光和紫外光刺激下电流响应对比图^{[42}]; (e) 可见光/紫外光调控突触可塑性示意图; (f) 人工突触在可见光脉冲刺激下的电流响 应; (g) 人工突触在紫外光脉冲刺激下的电流响应; (h) 可见光调控的突触 STDP 功能模拟; (i) 基于忆阻器阵列的视觉感存算一 体系统低级处理和高级处理功能示意图^[56]

Fig. 3. (a) Schematic of the human visual system; (b) schematic diagrams of the synapse, neuron, and two-terminal memristor; (c) schematic diagram of STP and LTP behavior; (d) comparison of current response of artificial synapses under red light and ultraviolet light^[42]; (e) diagram of synaptic plasticity regulated by visible/ultraviolet light; (f) current response of artificial synapses stimulated by visible light pulses; (g) current response of artificial synapses stimulated by ultraviolet light pulses; (h) simulation of synaptic STDP function regulated by visible light; (i) schematic diagram of low-level and high-level processing functions of visual insensor computing system based on memristor array^[56].

可以对视觉感知信息进行记忆和实时处理. 光调控 忆阻器突触可作为视觉感存算一体系统中的基本 器件, 能实现人类视觉系统从视网膜到视觉皮层的 感知和处理过程. 理想的感存算一体系统中, 光响 应忆阻器的光响应度是可调控的, 则感知阵列可以 直接组成感知处理神经网络实现低级与高级任务. Mennel 等^[82]构建了一个由基于二硒化钨的光电 晶体管组成的神经网络,每个晶体管的光响应度可 以单独调节.改变光电二极管的光响应度就可以改 变神经网络中的连接强度,即突触权重.该系统将 光学传感与神经形态计算相结合,可以执行简单的 计算任务.基于晶体管结构的视觉感存算一体化技 术发展程度比忆阻器结构的要高,光调控忆阻器突 触阵列仅能做一些低级任务处理,如何设计光响应 度可调控的光控忆阻器突触是跨入高级任务处理 的重要研究方向.

2019年, Zhou 等^[39] 设计了具有 Pd/MoO_x/ITO 结构的光电价变忆阻器突触器件.器件在 365 nm 紫外光刺激后从高阻态转化为低阻状态且具有非 易失特性,当施加负压到-2.13 V 时器件复位.其 机理可解释为:在 UV 照射后, MoO_x薄膜中产生 电子和空穴,光生空穴和水分子之间反应产生质 子 (H+). 光生电子、质子导致 Mo 的价态从 6+变 为5+,从而影响了介质层的导电性.在复位过程 中,电场驱动质子漂移使器件恢复为高阻态 (开关 比约为40). 和传统传感器相比该光电忆阻器的输 出电流具有时间依赖性, 输入光脉冲强度和宽度影 响着输出电阻状态,即器件可以模拟突触功能,且 具有光可调的突触可塑性. 在较高的光强下, 器件 脉冲电流增加得更快、保持时间更长.相应地,亮 度越高的像素积累效果越强,器件实现了图像的对 比度增强功能. Zhou 等^[39] 选取字母"P", "U", "C" 进行图像识别,使用忆阻器阵列突出字母特征,平 滑背景噪声. 忆阻器阵列预处理后的图像输入网络 经过 1000 次训练后, 网络的识别率高达 0.986, 这 说明前端图像预处理功能的实现有效地提高了后 续处理任务的处理效率和精度. 然而由于材料限 制,器件只对UV有响应,而人类视觉系统需要对 外界较宽的波长范围响应,若要更真实地模拟人类 视觉系统,器件的响应波长范围还需要进一步拓 宽. Wang 团队^[44]将具有 ITO/ZnO/Ag 结构的非 易失电化学光电忆阻器应用于可感知白光的感存 算一体化系统. 白光诱导器件内部产生电子-空穴 对,光生电子吸引电极上的 Ag+移动到功能层形成 导电细丝, 忆阻器在白光照射下从高阻态转变为低 阻态. ZnO 忆阻器的电导状态在光刺激时增大,在 电刺激时减小,模拟生物突触 LTP/LTD 特性.团 队使用光电忆阻器来模拟构建了一个用于感存算 一体的人工视觉系统,网络在1000次训练后,人 脸识别准确率达到 86.7%. 然而白光包含多波长的 光刺激,该项工作并未对不同波长光源输入展开研 究,区分不同波长的感知功能的研究有利于更加真 实的场景应用.

如图 3(b) 所示, Yang 及其团队^[42]利用真空 沉积法制备了模拟光子突触功能的两端人工突 触器件,器件结构为 ITO/SnO₂/CsPbCl₃/TAPC/ TAPC:MoO₃/MoO₃/Ag/MoO₃.紫外光照射下SnO₂ 纳米粒子和 CsPbCl₃钙钛矿界面中的载流子捕获 和释放使得器件的电导率可受光调控,也使器件在 光驱动下可以模拟生物突触的 STP 和 LTP 等行 为(图 3(c)). 在强紫外光较长时间照射下, 突触的 响应电流在开始时增加随后逐渐下降,器件的这种 特性成功模仿了人眼虹膜在强光照射下自动控制 入射光量的行为.此外,TAPC:MoO3薄膜的存在 使器件具有检测深红光的能力,但由于它不会触发 光载流子的捕获/释放,因此器件在红光刺激下不 具有记忆和存储功能,如图 3(d) 所示. 虽然红光的 引入并不能激发器件的突触行为,但也使器件具有 在紫外光和红外光下双模式的工作能力.在 2021年, Shan 等^[56]利用表面等离子体共振和光激发原理 设计了一种具有 Au/Ag-TiO₂/FTO 结构的等离子 体光电忆阻器,器件具有完全光调控的突触可塑 性, 且响应光波范围为 300-800 nm(紫外光及可 见光范围), 如图 3(e) 所示. 在可见光的照射下, Ag 纳米粒子的表面由于等离子体共振效应发生光氧 化而产生热电子, 热电子使界面肖特基势垒降低从 而提高了器件的导电性. 该效应允许器件具有可见 光诱导的突触长时程增强效应 (图 3(f)). 而 UV 光 辐照可以导致 Ag+的光还原, 器件具有紫外光诱导 的突触长时程抑制效应 (图 3(g)). 基于光调控 LTP 和 LTD 特性, 忆阻器阵列实现了对图像预处理的 功能.相比传统电学预处理操作,团队利用器件全 光调制的优势,使用紫外光刺激来降低长期噪声 点,从而愈加突出了图像的主要特征.进行预处理 后的图像传输到光电忆阻器神经网络进行训练和 识别, 基于 STDP 权重更新法则 (图 3(h)), 300 次 训练后网络可拥有 98% 的识别率. 在这项工作中, 基于全光调制的人工忆阻器突触实现了感存算一 体系统的低级和高级功能,如图 3(i) 所示.

4.2 触觉感存算一体单元

皮肤下的触觉感受器能够接受外部压力刺激, 产生的响应信号经神经系统传入大脑形成了触觉 (图 4(a)),触觉信号被神经系统存储下来便成为触 觉记忆使我们更好地对外界环境作用,指导我们日 常生活中对物体的握力和互动.在人机交互、柔性 机器人等领域,所感知到的触觉应能被反馈来检 测/操纵目标,否则这些设备在面对熟悉的物体时 仍然会感到僵硬和生疏.像人类一样进行触觉感知 和处理的仿真对于未来的智能交互系统是至关重



图 4 (a) 生物触觉感知系统示意图; (b) 压力传感器和 Nafion 忆阻器集成的人工触觉感知系统; (c) 触觉系统在不同按压力度 下的电流响应图; (d) 对采集到的数据进行 K 邻近分类网络算法处理^[61]; (e) 集成触觉传感器和 HfO₂ 基忆阻器的触觉感觉神经; (f) "SOS"和"TEAM"莫斯电码信号刺激人工触觉神经元的电流响应^[66]; (g) MXene 传感器、ADC-LED 电路、光电忆阻器构成的 神经系统; (h) 光调控的突触 PPF 模拟^[64]

Fig. 4. (a) Schematic illustration of the biological haptic perception system; (b) artificial haptic perception system consisting of pressure sensor and Nafion-based memristor; (c) current response of tactile system at different pressing magnitudes; (d) schematic of processing by *K*-nearest neighbors algorithm^[61]; (e) tactile sensory nerve consisting of haptic sensor and HfO₂-based memristor; (f) current response of artificial tactile neuron under "SOS" and "TEAM" Morse code signals stimulus^[66]; (g) artificial afferent nerve system integrating MXene sensor, ADC-LED circuit and optoelectronic memristor; (h) simulation of photo-tunable synaptic PPF behavior^[64].

要的, 而基于忆阻器的触觉感存算一体单元的开发 更是具有举足轻重的意义.

目前科研人员主要应用多种功能器件集成来 搭建感存算一体化触觉系统. Zhang 等^[61]将金包 覆金字塔结构的压阻传感器与基于 Nafion 的忆阻器连接,模拟了人工触觉感知系统,其系统示意图如图 4(b) 所示.压阻传感器将压力刺激转换为电脉冲,然后电刺激输送至忆阻器.Nafion 忆阻器中

质子的量受电刺激调控,器件具有连续可调的电导 状态,实现了基本的生物突触可塑性如括双脉冲 抑制 (paired-pulse depression, PPD)、双脉冲易化 (paired-pulse facilitation, PPF)、STDP 的模拟. 突触后电流受压力幅度(图 4(c))、作用次数、频率 和持续时间调控,基于此特性,团队将人工触觉系 统的输出电流输入至 K 邻近分类网络用于识别不 同手写字母 (图 4(d)). 类似地, Xia 等[66] 搭建了碳纳 米管/聚二甲基硅氧烷与 HfO2 忆阻器结合的电子 皮肤系统, 如图 4(e) 所示. 基于芦苇叶模板的 PDMS 衬底与碳纳米管活性层结合构成的压阻传感器作 为电子皮肤将外界物理刺激转换为电信号.电压刺 激作用于 Pt/HfO₂/TiN 忆阻器, 使得 HfO₂活性 层中氧空位导电细丝形成和断裂进而影响着器件 电阻状态,器件的响应电流对应着生物突触后电 流. 传感器独特的多尺寸传感层表面保证了系统具 有较宽的压力探测范围和超高的灵敏度和线性度. HfO₂忆阻器被用于模拟生物突触功能,存储并处 理输入的压力信息. 该人工触觉感知神经元系统通 过识别外部压力实现了莫斯码的准确解码(图 4(f)), 具有保持性能优越和稳定性好的特点.

在传感-处理的架构下, Tan 等[64] 对人工触觉 系统做了进一步的性能优化. 与传统压电转换处理 方式不同的是,他们采用压电转换-电光转换的编 码模式,将光脉冲序列作为携带数据信息的形式. 在系统中,压力刺激经过模数转换器和发光二极管 耦合电路,被编码为光脉冲序列再输送至具有 ITO/ ZnO/NSTO 结构的光电忆阻器中 (图 4(g)), 忆阻 器成功模拟了突触的基本可塑特性 (图 4(h)). 仿 生的尖峰序列编码方式比电压幅值编码方式更加 稳定,因为电压幅值会因后电路的寄生电阻衰减, 而且尖峰序列形式允许频率编码和时间编码等多 种编码策略.此外,光作为信息载体可以非接触地 作用于光电忆阻器突触,即光信息与器件的通信不 局限于点对点,可以是多位点信息源耦合作用于忆 阻器突触,光通信方式的功耗、速度与灵活性都优 于纯电场调控形式. 该仿生触觉系统不仅能够检测 和识别多个压力输入,还可以识别莫斯码、盲文和 物体运动.光电忆阻器突触可以降维提取图像特征 值,使得系统能够以较高效率识别和记忆手写字母 和单词.在上述工作中,忆阻器突触仅实现了存储 和计算功能,系统的传感功能依赖于压力传感器实 现,即忆阻器突触尚未融合传感功能,不能原位处 理感知信息.

理想的触觉感存算一体化边缘单元应具备原 位感知压力后存储并处理信号的能力,这对于典型 的两端结构的忆阻器是难以实现的. Wang 等 [68] 采用了衍生的伪3端忆阻器突触实现了触觉类别 感存算一体化功能. 忆阻器具有 Au/ZnO/Au 的水 平结构,聚乙烯醇 (polyvinyl alcohol, PVA) 和氯 化钙 (calcium chloride, CaCl₂) 聚合物离子胶通过 旋涂成膜于 ZnO 膜上, ZnO/PVA 杂化异质结赋 予了器件在电和压力作用下模拟生物突触可塑性 的能力.在电刺激下,外加电场的施加与撤销控制 着 PVA 膜中带电离子的分离与聚合, 从而形成内 建电场. 外加电场与内建电场协同作用下 PVA/ZnO 层间的电荷耦合效应影响着 ZnO 沟道中电子的浓 度,即器件的电导性,器件成功模拟了电场调控的 生物突触可塑性. 压力作用于 PVA 膜表面时, 两 电极上方的薄膜位置与中间位置之间的厚度差逐 渐减小,导致 PVA 膜不同位置下的介电常数发生 改变. 根据介电常数、电压、距离和电场强度之间 的关系两电极间的有效电场受压力影响,受压力调 控的有效电场进而影响着 ZnO 沟道的导电性,器 件成功模拟了压力调控的生物突触可塑性.团队应 用 10 × 10 的忆阻器阵列检测外加压力分布后产 生触觉图像,并将预处理后信息传入三层的神经网 络进行学习训练. 在该项工作中, 采用创新结构的 器件真正意义地集成了触觉感知、存储、处理功能 于一体,具有优越的多功能性.然而,水平结构的 忆阻器可能不利于以后的大规模立体集成工艺,器 件的结构优化依然是一项重要的研究方向.

4.3 嗅觉感存算一体单元

生物嗅觉系统能感知和辨别气味,这对神经内 分泌调节、情绪反应、捕食或躲避等方面都至关重 要.人体的嗅觉系统示意图如图 5(a)所示.受生物 嗅觉器官的启发,科研人员开发了人造化学传感器 应用于食物质量鉴别、违禁品或爆炸物检测、疾病 诊断等方面.传统的化学传感器在概念和技术上都 面临着重大的挑战,如信号载体波动、不稳定性、 灵敏度低、响应时间长,以及对生物气味嗅觉特征 了解不深入.与视觉和触觉感知系统相比,嗅觉感 知是一个更为复杂的过程.嗅觉感存算技术起步较 晚,由于信号兼容性、响应速度和应用范围等限制, 嗅觉感存算一体系统的实现难度较大.



图 5 (a) 生物嗅觉感知系统示意图; (b) 人工嗅觉推理系统原理图; (c) W/WO₃/PEDOT:PSS/Pt 忆阻器在脉冲下刺激下的电流 相应; (d) 所用忆阻器突触真实和理想的电导调制曲线^[73]; (e) 气敏忆阻器机理示意图; (f) SnO₂ 气敏忆阻器对不同浓度一氧化氮 气体的电流响应; (g) 由 Ta₂O₅, HfO₂ 和 SnO₂ 忆阻器组成的气体感知阵列^[71]

Fig. 5. (a) Schematic of biological olfactory system; (b) schematic of artificial olfactory inference system; (c) current response of memristor with $W/WO_3/PEDOT:PSS/Pt$ structure under pulse stimulus; (d) experimental and ideal conductance modulation curves of the memristive synapse^[73]; (e) schematic of the gas sensing mechanism; (f) current response of SnO₂ based gas-sensing memristor depending on NO gas concentration; (g) schematic diagram of the gas-sensing array consisting of Ta₂O₅, HfO₂, and SnO₂-based memristors^[71].

受生物嗅觉系统工作机理启发, Lu 等^[75] 开发 了由气体传感器、柔性振荡器和人工突触集成的气 体感知系统.在该工作中, NiO 基气体传感器实现 生物受体功能,当其电阻值在接收到气体刺激时会 发生改变.柔性振荡器根据传感器阻值编译产生不 同频率的电压脉冲信号并输入基于还原氧化石墨 烯和壳聚糖的忆阻器.在连续电压脉冲刺激下, 忆 阻器表现出连续电导变化的特性,该特性可用于模 拟生物突触可塑性功能,器件机理为壳聚糖提供的 质子与还原氧化石墨烯纳米片中的缺陷和官能团 之间相互作用.电压刺激后忆阻器电导的变化率和 斜率作为特征输入传送至神经网络进行训练,训练 后的神经网络可完成识别硫化氢气体浓度的高级 任务.该仿生嗅觉系统虽然实现了对硫化氢气体浓

度的感知识别功能,但单一气体感知系统应用面 窄,有较大的局限性,且电路系统占用空间大.2021年, Wang 及他的团队^[73]开发了基于忆阻器的人工嗅 觉系统,可以实现复杂环境中4种气体(乙醇、甲 烷、乙烯和一氧化碳)在10个不同浓度下的识别和 推理功能. 气体传感器阵列感知气体后输出电脉冲 信号,脉冲序列输入至由W/WO3/PEDOT:PSS/Pt 易失性忆阻器器件组成的储蓄池计算系统(图 5(b)). 忆阻器在脉冲下的电流响应特性如图 5(c) 所示, 在连续电脉冲刺激下,器件的电导随着脉冲数量增 加而增加. 电导的变化与输入电脉冲的数量满足一 定的关系,即器件的电导可以看作是对电输入进行 处理后的输出参数. 传统储蓄池计算系统中大量的 非线性函数节点可以被该忆阻器替代.基于忆阻器 的储蓄池计算系统处理电脉冲序列的时空信息,提 取出高维空间的特征变量,系统输出不同的电导状 态对应着不同的特征信息.特征信息随后传入基 于 Pd/W/WO3/Pd 非易失忆阻器 (电导调制曲线 如图 5(d) 所示) 搭建的人工神经网络进行训练学 习,神经网络人工突触的权重被调节进而进行训练 识别. 结果表明系统具有良好的识别速度, 且对复 杂气体种类、浓度的识别率较高.

上述工作的忆阻器仅具有存储计算能力,没能 原位感知气体.有一些实验组应用忆阻器完成了气 体感知功能,但受限于响应速度和机理,尚未有工 作实现相应的突触行为模拟. Kim 实验组^[71]应用 SnO₂, Ta₂O₅和 HfO₂薄膜制备了忆阻器基的 NO 气体传感器. 在电刺激下, 连接顶电极和底电极的 氧空位导电细丝会在介质层形成,使得原始器件从 高阻态转换成低阻态 (图 5(e)). 由于高电负性, 注 入的 NO 气体分子会从 SnO2 中捕获电子并附着于 介质层表面. 随后, 带电负性的 NO 分子与带正电 的氧空位结合呈中性状态,导致了导电细丝的断裂 从而使器件的电阻增大. 电流的变化程度决定了器 件的气体响应范围和灵敏度. 在恢复过程中, 他们 应用电压刺激可以解吸附在介质层表面的 NO 分 子,直接使器件复位至原始的低电阻状态.如图 5(f) 所示,所制备的基于 SnO2 忆阻器的传感器具有快 的反应/恢复速度 (<1 s/<90 ns), 能够在低浓度 的 NO 气体环境下正常工作. 他们提出了在混合气 体环境下应用传感阵列来检测目标气体的策略 (图 5(g)). 不同响应特性的材料体系相互耦合可实 现多功能检测系统,极大地提高了检测精确度和检 测范围,该策略可以为气体传感技术提供技术创 新,可用于安全、医疗和环境监测等多个领域.该 工作仍有值得深入研究的方面,如忆阻器如何能恢 复到同样的电流状态,接触面积如何影响性能,导 电细丝的状态如何影响电流变化的程度等.

目前人工嗅觉感觉系统的工作相对简单,处于 初步阶段,气体种类差距小、感知周期长、精度低 等问题都需要更深入的研究来解决.

4.4 听觉感存算一体单元

声音定位和识别是生物听觉系统中非常重要 的功能. 传播的声波以一定的频率和幅度振动耳 膜,物理振动通过听骨传递到耳蜗毛细胞后转换成 生物电信号.神经系统对声音信号的组成进行分 析,完成声音的定位和识别.为实现人工听觉系统, 听觉感受器需要在较宽的振动频率探测范围内具 有超高的灵敏度. 基于仿生时空信息处理算法, 听 觉处理器处理事件驱动的信息完成对应的定位识 别功能. 具备响应输入声波频率和幅度的三大类材 料是压电材料、摩擦电材料和电磁材料,开发合适 材料体系和特定结构的听觉传感器和处理器是至 关重要的. 目前较少研究工作将忆阻器应用于听觉 感存算一体系统,现阶段人工听觉系统普遍由基 于 CMOS 技术的大规模集成电路搭建而成, 声波 信息的复杂度、时间依赖性、时空编码等特点使得 构建听觉感存算一体单元极富挑战性.

声音定位的两种工作机制包括通过两耳时差 和通过两耳水平差进行检测. 2018年, Wang 等^[79] 通过设计基于 HfO2 忆阻器的一个晶体管/一个电 阻 (1 transistor 1 resistor, 1T1R) 结构构建了人工 突触尖峰神经网络 (spiking neural network, SNN) 来进行人脑神经形态的时空信息处理. HfO。忆阻 器器件内部导电细丝的尺寸可被电压刺激调控,器 件呈现出的不同的电阻状态对应着不同的生物突 触权重. SNN 根据输入脉冲时间间隔模拟人脑对 声音位置检测的功能. 输入端包含两个突触前神经 元,分别模拟人的左耳和右耳,突触后神经元产生 内部电压信号,网络根据两个后神经元间电压信号 的差异可准确识别发声位置. 声源探测基本功能的 实现仅需 2 × 2 个忆阻器突触, 这证明了基于时空 信息计算的 SNN 提高了神经形态硬件电路的能量 和信息效率. 该系统未完成原位感知声音的传感功 能,模拟声音源是直接以电信号的形式输入系统

的,如何搭建传感、存储、计算一体化的听觉系统 仍需深入的研究.

4.5 多模调控感存算一体单元

协同综合多感官信息是人类感知系统的基础 功能.人类的大脑可以整合来自包括视觉、听觉、 触觉、嗅觉等多个感官系统的输入,这有助于在单 模感知信息不足的复杂环境下更快速、更准确地做 出反应.由于信号本身的随机性和噪声,人工感知 系统凭借单一的感官信息做决策通常会导致不可 避免的不确定性.参考人类的多模感知协调作用, 人工感知系统可将多个单模信号协同耦合来实现 更加先进智能的认知功能.为了搭建超智能化机器 人,人工感觉系统需要具备高级的认知感知和多模态环境信息处理能力.科研人员致力于开发能够处理多感官耦合信号的多功能感存算一体化系统,而 多模调控的忆阻器件为实现该系统提供了潜在可行的策略.

2021 年 Wang 等^[41] 在柔性 PDMS 衬底上制备 了具有 ITO/MXene-ZnO/Al 结构的可多模式调 控的双极型非易失性忆阻器 (图 6(a)). 紫外光和环 境湿度两个外场刺激可同时调控器件的工作性能. 器件开启电压随着入射光强度的增加而减小,且高 强度紫外光照射可使器件从高阻态向低阻态切换, 这归咎于 MXene 具有长时程捕获光生电子的能 力,光敏效应诱导氧空位导电细丝的形成 (图 6(b)).



图 6 (a) 柔性 MXene-ZnO 忆阻器示意图; (b) 器件在不同紫外光照强度下的 *I-V*曲线; (c) MXene-ZnO 忆阻器受光和湿度调控 的电流分布图; (d) 应用光和电脉冲实现突触 LTP 和 LTD 行为的模拟; (e) 基于光和湿度调控的忆阻器突触搭建的神经网络示 意图^[11]; (f) 多模脉冲感知处理系统工作流程图^[81]

Fig. 6. (a) Schematic structure of the flexible MXene-ZnO-based memristive device; (b) I-V curves of device under UV irradiance with different intensities; (c) current profile of MXene-ZnO memristor regulated by light and humidity; (d) simulation of synaptic LTP and LTD behaviors by UV light and electrical pulses; (e) schematic of neural network based on MXene-ZnO-based memristive synapses^[41]; (f) operational diagram of the multimode spiking perception and processing system^[81].

自组装的 ZnO 纳米颗粒增大了异质结构在两电极 之间的总接触面积,这不仅影响了电导率、态密度 等性质,而且改善了离子吸附和扩散行为.在高湿 度环境下,水分子会通过双氢键吸附在 MXene-ZnO 异质结上. 表面官能团的水解增加了质子的浓度, 质子和氧空位之间的静电吸引限制氧空位导电丝 的生长,破坏了氧空位导电细丝的稳定状态.光辅 助氧空位导电丝形成和湿度诱导的氧空位导电细 丝断裂之间存在竞争关系, 调控着器件的响应电流 以及电导状态 (图 6(c) 和图 6(d)). 忆阻器阵列被应 用于对图片进行感知和预处理,模拟视网膜功能. 在相对湿度在 0%-20% 和 40%-60% 的情况下, 网络对图像的识别准确率分别为 75.44% 和 82.96%, 这说明基于该忆阻器阵列的人工神经网络具有适 应环境的图像预处理功能.此外,该团队进一步探 讨了使用基于 MXene-ZnO 的忆阻器作为突触来 实现高级处理的感存算一体系统的权重更新行为, 突触权重受光、电、湿度调控.训练 60000 次后神 经网络系统具有高识别准确率 (图 6(e)). 基于多模 态忆阻器的感存算一体系统具有降低传统视觉系 统电路复杂性的潜力.

2021年, Tan 等^[64]在触觉感存算一体化系统 研究的基础上提出一种人工多模尖峰神经系统,该 系统对5种人工感官(视觉、触觉、嗅觉、听觉、味 觉)进行多模态感知.多个种类的传感器被应用于 感知,感知信息被编码为光脉冲序列.光敏忆阻器 阵列在硬件层面上对感知信息进行解释、过滤、集 成和记忆,其原位记忆和信息过滤特性有助于神经 网络的学习和训练,如图 6(f) 所示. 通过交叉模态 学习系统实现机器人识别和想象功能. 在这项工作 中,光尖峰序列作为数据载体,单个光电忆阻器可 以同时处理多种感知信息,这为多模调控感存算一 体化提供了新的思路. 然而在该系统中, 嗅觉和味 觉的模拟并未真正从现实环境中感知信息,而是使 用电脉冲模拟刺激,这与人类多模感知系统还存在 一定差距,要解决这一问题,需要在进一步探究人 类感知机制的基础上来改进感知系统.应用于感存 算一体化系统的忆阻器的性能对比如表1所列.

5 结论与展望

随着人们对生物传感过程认识的加深和神经 形态忆阻器件的发展,神经形态忆阻器件在感存算

领域的应用也随之应运而生. 感存算一体架构的发 展还处于初始阶段,还有许多分支领域有待开拓. 目前的研究处理数据量级较低、工作任务较为简 单,器件仅具有简单的感知存储性能,或感知存储 一体化加简单处理的性能,尚未形成真正意义的感 存算一体化. 此外, 除了常见的感官感知器件的研 究,一些研究组研究将忆阻器用于液体中羟基离子 浓度检测[83]、重金属检测[84]、伽马射线检测[85]、温 度探测[86]等,但研究还仅限于感知领域,相应的原 位存储处理功能仍需深入开发,目前仍存在一系列 科学技术难题亟待解决. 要搭建感存算一体化技术 平台,材料、器件、工艺与集成、电路系统架构和算 法等不同方面都存在瓶颈,从工程方面来说,目前 最大的挑战主要在于器件层面. 传统忆阻器的性能 问题依然是阻碍其商业用途的主要因素,如器件的 均一性、稳定性、耐受性等. 忆阻器简单的结构是 其作为新型存储器的一大优势,要在满足简单器件 结构的前提下实现复杂的感存算一体功能,即如何 权衡器件结构复杂性和多功能性是非常有挑战性 的,这需要科学家们更深入地探究生物单元的潜在 机理后设计器件结构,进而确定合适的工作机理来 模拟生物功能性,目前所报道的器件在性能和技术 成熟度等方面存在较大差异,这增加了选择、优化 和迭代的难度. 科研人员需要对器件的传感、存储 和处理性能进行取舍,如为了提高感存算一体化器 件传感的响应速度,其数据保持性能相比传统存储 器会有所下降.此外,一些感存算器件对不同传感 刺激源都有所响应,这影响了器件对特定感官信息 的选择性. 在器件性能评估方法方面也存在一定的 困难, 传统的传感器/存储器需要在输入刺激下进 行长达数万个周期操作来评估其耐受性,然而当涉 及感存算一体单元时,将器件暴露于外部刺激后执 行大量的循环周期的测试方案就变得具有挑战性. 本文从器件、工艺与集成、电路系统架构和算法 4个方面进行思考和展望.

器件层级: 感存算一体化系统要求基本单元具 有原位感知、存储与处理的功能. 传统的传感、存 储和计算单元是基于不同的材料组合、器件结构、 异构集成技术来组合的, 要将 3 个功能集成于单个 器件必须要从材料及器件的角度出发进行设计. 目 前可用于不同感知源 (化学物质、辐射、温度、压 力)的材料依然非常有限, 设计合适的材料体系是 构建感存算基本单元的前提. 传统的传感器的评估

			1	11			1	0.		
	忆阻器结构	响应类型	阻变机理	开启/关闭 电压/V	开关比	PSC	STP	LTP	具体实现功能	文献
视觉	Ag/CH ₃ NH ₃ PbI ₃ (OHP)/ITO		碘空位导电细丝	0.32/-0.52	1×10^4	\checkmark		\checkmark	数字识别分类	[47]
	Ni/Al ₂ O ₃ /Au	UV	金属导电细丝	1.7/-1.6	1×10^2				图像记忆	[3 8]
	$\mathrm{Pd}/\mathrm{MoO}_{x}/\mathrm{ITO}$	UV	界面效应	-2.13	40	\checkmark	\checkmark	\checkmark	图像预处理	[39]
	Ag nanowire/TiO_2	visible light (vis)	界面效应	—		\checkmark	\checkmark		广角感知、处理存储	[50]
	$\frac{\rm glass/ITO/ZnO/PbS}{\rm ZnO/Al}$	UV/infrared ray (IR)	氧空位导电细丝			\checkmark	\checkmark		数字识别分类	[45]
	$\rm ITO/Nb:SrTiO_3$	vis	界面效应			\checkmark	\checkmark	\checkmark	自适应光电突触	[48]
	ITO/PEDOT:PSS/C uSCN/CsPbBr ₃ PNs/Au	UV	界面效应	—		\checkmark	\checkmark		回溯记忆功能的图 像记忆	[51]
	$\begin{array}{c} \mathrm{ITO}/\mathrm{SnO_2}/\mathrm{CsPbCl_3}/\\ \mathrm{TAPC}/\mathrm{TAPC:MoO_3}/\\ \mathrm{MoO_3}/\mathrm{Ag}/\mathrm{MoO_3} \end{array}$	UV/red light	界面效应	—		\checkmark	\checkmark		双模式图像检测记 忆	[42]
	RGO/GO- NCOD/graphono	UV	氧空位导电细丝				\checkmark		图像识别	[53]
	ITO/CsPbBr ₂ I/P ₃ HT /Ag	vis/NIR	卤素空位导电细 丝	0.4/-0.4	> 10	\checkmark	\checkmark		图像预处理	[46]
	ITO/PCBM/MAPbI ₃ : Si NCs/Spiro- OMeTAD/Au	$\rm UV/NIR/vis$	界面效应	—		\checkmark			图像预处理	[54]
	Au/Ag-TiO ₂ /FTO	vis/UV	表面等离子体共 振效应/金属导 电细丝	3.4/-1.8	1×10^{3}		\checkmark	\checkmark	图像预处理及识别	[56]
	$Ag/Cu_3P/ITO$	$\lambda = 660~\mathrm{nm}$	金属导电细丝	_	$1{ imes}10^4$	\checkmark			回溯记忆功能的图 像记忆	[57]
	Ni/p-NiO/n-ZnO/Ni	UV	界面效应			\checkmark			图像记忆	[40]
	ITO/MXene-ZnO/Al	UV	氧空位导电细丝	-0.5/1.2	$1{ imes}10^4$	\checkmark	_	\checkmark	图像预处理及数字 识别分类	[41]
	$\rm ITO/ZnO/Ag$	白光	金属导电细丝	2/-2		\checkmark	\checkmark	\sim	人脸识别	[44]
	$\rm NiO/TiO_2/FTO$	UV	界面效应	_	> 10	\checkmark	\checkmark	\sim	识别分类图像	[59]
触觉	Au/Nafion/ITO	压力	质子迁移			\checkmark	\checkmark		手写字母识别	[<mark>61</mark>]
	NiO/ZnO/ITO/PET	应变	界面效应	—		\checkmark	\checkmark		外部应变的时空信 息处理	[62]
	$\rm Si/NbO_x/TiN$	压力	晶体NbO2通道	$\begin{array}{l} V_{\mathrm{TH}} = \\ 2.05 \ \mathrm{V} \\ V_{\mathrm{H}} = 1.53 \ \mathrm{V} \end{array}$			_		将压力模拟信号转 换为动态振荡频率	[63]
	ITO/ZnO/NSTO	压力	界面效应	—	1×10^4	\checkmark	\checkmark	_	识别和记忆手写字 母和单词	[64]
	$\rm Al/TiO_2/Al$	压力	氧空位导电细丝	_	14.2	\checkmark		\checkmark	压力实时感知、学习 /推理、反馈可视化 图像	[65]
	$\rm Pt/HfO_2/TiN$	压力	氧空位导电细丝	0.9 - 1.1 / -1	> 100	\checkmark			触觉记忆学习	[<mark>66</mark>]
	ZnO/PVA基忆阻器	压力	界面效应	$V_{\rm TH} =$ 3.25 V	1×10^3	\checkmark	\checkmark	\checkmark	识别压力分布, 触觉 可视化	[6 8]
嗅觉	$\rm Pd/W/WO_3/Pd$	乙醇、甲烷、乙 烯、一氧化碳	氧空位导电细丝				_		气体识别	[73]
	Ti/rGO-CS/Au	H_2S	界面效应	_		\checkmark	\checkmark		气体识别	[75]

表 1 应用于感存算一体化系统的忆阻器的性能比较 Table 1. Performance comparison of memristors applied to in-sensor computing systems.

参数有响应范围、分辨率、灵敏度、选择比、操作速度/响应时间等,而传统的存储设备的评估参数有操作电压、保留特性和耐受性等.以综合的参数来评估感存算基本单元,设计人员改进器件时需要从各方面进行权衡取舍,如何兼顾传感、存储和计算

性能是一项重大的挑战.目前的感存算一体器件的 研究工作大多是集中在单模感官系统模拟 (如触 觉、视觉、嗅觉)和简单处理,处理能力有限.但实 际生物感知记忆系统处于更为复杂的环境,极小的 感受单元可以同时多模式感官并进行信息处理使 生物对外部事件做出准确反应,依靠多个感官通道 的系统会更可靠.一方面,人工感存算一体系统的 性能远不如人类,结合多种感官刺激有利于提高整 体的敏感度和准确性.另一方面,单一的感知输入 信息会带来很高的不确定性,在实际应用中容易导 致重大误判.实现多模式感知融合和多元化处理功 能的器件体系是未来感存算系统的发展方向.在功 耗和器件尺寸方面,人工感存算系统与生物感官系 统之间存在着巨大的差距.这些挑战需要科研人员 对生物传感过程加深认识获取灵感、对神经形态忆 阻器件机理深入研究、发展新型纳米电子制造技术 来解决.进一步研究感存算器件的工作机理,也将 促进器件的成品率、均一性以及可靠性的提高.

工艺与集成:对于需要集成的感知与计算系 统,目前许多研究只是基于规模较小的分立式器件 单元阵列的简单互连,没有发挥集成阵列高效并行 运算的优势.集成方法包括三维单片集成、平面 SoC 集成、三维异构集成、2.5D 异构集成等. 不论 是上述哪种集成方式,都涉及多个功能层级/芯片 以及不同材料的整合,即系统构建需要考虑各层级 的工艺兼容性,这给集成工艺条件带来了挑战.一 些成膜过程 (如外延生长) 需要高温工艺, 为了避 免其对系统的性能和稳定性造成影响,开发基于低 温工艺的感存算一体化系统是非常有必要的. 如在 三维堆叠芯片中,热膨胀系数不匹配导致的高内建 应力会引起可靠性问题. 超薄低维半导体低温下可 以转移到任意衬底上,但考虑到大规模和高质量的 材料增长以及加工兼容性,集成工艺仍然是一个挑 战. 就集成技术而言, 可靠性是一个亟需解决的问 题. 如在三维集成技术中, 传感器和处理单元距离 很近,系统工作时所产生的热量会导致热噪声和降 低识别精度,这限制了系统性能和能源效率的提 高. 发展可靠的集成技术是未来大规模集成感存算 一体运算系统的关键研究方向.

算法层级:在人类多感知系统中,感知单元将 环境信息转化为电位变化,并将电位变化编码为脉 冲序列,脉冲序列在大脑皮层解码进行进一步信息 处理.对比幅值编解码形式,脉冲序列编码形式可 使系统更加灵活地处理时空感官信息,这需要科研 人员开发出相应的脉冲神经网络的训练学习算法. 设计多种感官信息耦合处理方式是搭建多模式感 存算一体化系统的重要一环,对生物传感系统机制 认识不足也是限制仿生处理算法开发的主要瓶颈. 电路系统架构:算法需要配合对应的硬件电路 部署才能实现.感存算一体器件对原始信息预处理 后还需配合系统级架构完成更为高阶的任务.针对 不同的应用场景,电路架构需要作相应的调控.如 嗅觉传感器通常对湿度和温度都很敏感,系统需要 额外的信号管理子电路以保证灵敏度和准确性.目 前感存算电路系统的开发基本都是针对特定场景, 这限制了系统的可移植性和可扩展性,未来需要深 入研究如何构建适用多场景的通用性和可重构性 强的硬件平台.

感存算一体化系统是一个涉及多个学科的研 究领域,涵盖材料、化学、生物、机械学、电子学等, 近年来已经成为一个重要的战略研究领域,感存算 一体系统有可能成为一个颠覆性的技术平台. 忆阻 器突触在感存算一体系统中有着巨大的应用潜力. 本综述从器件单元层级回顾了当前应用于感存算 一体化系统的忆阻器突触的研究,讨论了该领域的 主要研究和进展,同时也提出了目前的一些挑战以 及对未来的展望. 未来研究人员需要从材料、器 件、算法、电路系统等多个层面协同创新,开发高 能效的新型感存算一体化系统.

参考文献

- [1] Lee Y, Lee T W 2019 Acc. Chem. Res. 52 964
- [2] Zeng M, He Y, Zhang C, Wan Q 2021 Front. Neurosci. 15 690950
- [3] Wan C, Cai P, Wang M, Qian Y, Huang W, Chen X 2020 Adv. Mater. 32 1902434
- [4] Zhou F, Chai Y 2020 Nat. Electron. 3 664
- [5] Wan T, Ma S, Liao F, Fan L, Chai Y 2022 Sci. China Inf. Sci. 65 141401
- [6] Liao F Y, Chai Y 2021 Physics 50 378 (in Chinese) [廖付友, 柴扬 2021 物理 50 378]
- [7] Kim Y, Chortos A, Xu W, Liu Y, Oh J Y, Son D, Kang J, Foudeh A M, Zhu C, Lee Y, Niu S, Liu J, Pfattner R, Bao Z, Lee T W 2018 Science 360 998
- [8] Shi W, Cao J, Zhang Q, Li Y, Xu L 2016 IEEE Internet Things 3 637
- [9] El-Atab N 2021 Phys. Status Solidi A 219 2100528
- [10] Phong Truong T, Toan Le H, Thi Nguyen T 2020 J. Phys. : Conf. Ser. 1432 012068
- [11] Li Y, Wang Z, Midya R, Xia Q, Yang J J 2018 J. Phys. D:Appl. Phys. 51 503002
- [12] Wang Z, Wu H, Burr G W, Hwang C S, Wang K L, Xia Q, Yang J J 2020 Nat. Rev. Mater. 5 173
- [13] Sebastian A, Le Gallo M, Khaddam-Aljameh R, Eleftheriou E 2020 Nat. Nanotechnol. 15 529
- [14] Ielmini D, Wong H S P 2018 Nat. Electron. 1 333
- [15] Wang J, Lv Z, Xing X, Li X, Wang Y, Chen M, Pang G, Qian F, Zhou Y, Han S T 2020 Adv. Funct. Mater. 30 1909114

- [16] Zidan M A, Strachan J P, Lu W D 2018 Nat. Electron. 1 22
- [17] Zhang Y, Wang Z, Zhu J, Yang Y, Rao M, Song W, Zhuo Y, Zhang X, Cui M, Shen L, Huang R, Yang J J 2020 Appl. Phys. Rev. 7 011308
- [18] Sun K, Chen J, Yan X 2020 Adv. Funct. Mater. 31 2006773
- [19] Lv Z, Wang Y, Chen J, Wang J, Zhou Y, Han S T 2020 *Chem. Rev.* **120** 3941
- [20] Li K, Cao R, Sun Y, Liu S, Li Q, Xu H 2019 Micro/nano Electronics and Intelligent Manufacturing 1 87 (in Chinese)
 [李锟, 曹荣荣, 孙毅, 刘森, 李清江, 徐晖 2019 微纳电子与智能 制造 1 87]
- [21] Ji X, Zhao X, Tan M C, Zhao R 2020 Adv. Intell. Syst. 2 1900118
- [22] Sun F, Lu Q, Feng S, Zhang T 2021 ACS Nano 15 3875
- [23] Carrara S 2021 *IEEE Sens. J.* **21** 12370
- [24] Zhang Z, Li C, Han T T, Xu A, Cheng X, Liu G, Xie G J
 2021 Journal of Electronics and Information Technology 43
 1498 (in Chinese) [张章, 李超, 韩婷婷, 许傲, 程心, 刘钢, 解光
 军 2021 电子与信息学报 43 1498]
- [25] Zhu Y, Zhu Y, Mao H, He Y, Jiang S, Zhu L, Chen C, Wan C, Wan Q 2021 J. Phys. D:Appl. Phys. 55 053002
- [26] Tripathy A, Nine M J, Losic D, Silva F S 2021 Mater. Sci. Eng. R. Rep. 146 100647
- [27] Strukov D B, Snider G S, Stewart D R, Williams R S 2008 Nature 453 80
- [28] Zhao M, Gao B, Tang J, Qian H, Wu H 2020 Appl. Phys. Rev. 7 011301
- [29] Zhang Y, Mao G Q, Zhao X, Li Y, Zhang M, Wu Z, Wu W, Sun H, Guo Y, Wang L, Zhang X, Liu Q, Lv H, Xue K H, Xu G, Miao X, Long S, Liu M 2021 Nat. Commun. 12 7232
- [30] Kim S J, Kim S B, Jang H W 2021 Science 24 101889
- [31] Tsai S C, Lo H Y, Huang C Y, Wu M C, Tseng Y T, Shen F C, Ho A Y, Chen J Y, Wu W W 2021 Adv. Electron. Mater. 7 2100605
- [32] Arndt B, Borgatti F, Offi F, Phillips M, Parreira P, Meiners T, Menzel S, Skaja K, Panaccione G, MacLaren D A, Waser R, Dittmann R 2017 Adv. Funct. Mater. 27 1702282
- [33] Herpers A, Lenser C, Park C, Offi F, Borgatti F, Panaccione G, Menzel S, Waser R, Dittmann R 2014 Adv. Mater. 26 2730
- [34] Le Gallo M, Sebastian A 2020 J. Phys. D: Appl. Phys. 53 213002
- [35] Sebastian A, Le Gallo M, Eleftheriou E 2019 J. Phys. D:Appl. Phys. 52 443002
- [36] Zhang C, Chen Y, Yi M, Zhu Y, Li T, Liu L, Wang L, Xie L, Huang W 2018 Sci. Sin. Inf. 48 115
- [37] Ho V M, Lee J A, Martin K C 2011 Science 334 623
- [38] Chen S, Lou Z, Chen D, Shen G 2018 Adv. Mater. 30 1705400
- [39] Zhou F, Zhou Z, Chen J, Choy T H, Wang J, Zhang N, Lin Z, Yu S, Kang J, Wong H S P, Chai Y 2019 Nat. Nanotechnol. 14 776
- [40] Zhang L, Yu H, Xiao C, Si J, Xu H, Zhu W, Wang L 2020 Adv. Electron. Mater. 7 2000945
- [41] Wang Y, Gong Y, Yang L, Xiong Z, Lv Z, Xing X, Zhou Y, Zhang B, Su C, Liao Q, Han S T 2021 Adv. Funct. Mater. 31 2100144
- [42] Yang L, Singh M, Shen S W, Chih K Y, Liu S W, Wu C I, Chu C W, Lin H W 2020 Adv. Funct. Mater. 31 2008259
- [43] Vasileiadis N, Ntinas V, Sirakoulis G C, Dimitrakis P 2021 Materials 14 5223
- [44] Wang T Y, Meng J L, Li Q X, He Z Y, Zhu H, Ji L, Sun Q Q, Chen L, Zhang D W 2021 Nano Energy 89 106291
- [45] Li H, Jiang X, Ye W, Zhang H, Zhou L, Zhang F, She D, Zhou Y, Han S T 2019 Nano Energy 65 104000

- [46] Yang X, Xiong Z, Chen Y, Ren Y, Zhou L, Li H, Zhou Y, Pan F, Han S T 2020 Nano Energy 78 105246
- [47] Ham S, Choi S, Cho H, Na S I, Wang G 2019 Adv. Funct. Mater. 29 1806646
- [48] Gao S, Liu G, Yang H, Hu C, Chen Q, Gong G, Xue W, Yi X, Shang J, Li R W 2019 ACS Nano 13 2634
- [49] Zhao L, Fan Z, Cheng S, Hong L, Li Y, Tian G, Chen D, Hou Z, Qin M, Zeng M, Lu X, Zhou G, Gao X, Liu J M 2019 Adv. Electron. Mater. 6 1900858
- [50]~ Kumar M, Lim J, Kim S, Seo H 2020 ACS Nano ${\bf 14}$ 14108
- [51] Ma F, Zhu Y, Xu Z, Liu Y, Zheng X, Ju S, Li Q, Ni Z, Hu H, Chai Y, Wu C, Kim T W, Li F 2020 Adv. Funct. Mater. 30 1908901
- [52] Wu Z, Lu J, Shi T, Zhao X, Zhang X, Yang Y, Wu F, Li Y, Liu Q, Liu M 2020 Adv. Mater. 32 2004398
- [53] Lin Y, Wang Z, Zhang X, Zeng T, Bai L, Kang Z, Wang C, Zhao X, Xu H, Liu Y 2020 NPG Asia Mater. 12 64
- [54] Huang W, Hang P, Wang Y, Wang K, Han S, Chen Z, Peng W, Zhu Y, Xu M, Zhang Y, Fang Y, Yu X, Yang D, Pi X 2020 Nano Energy 73 104790
- [55] John R A, Acharya J, Zhu C, Surendran A, Bose S K, Chaturvedi A, Tiwari N, Gao Y, He Y, Zhang K K, Xu M, Leong W L, Liu Z, Basu A, Mathews N 2020 Nat. Commun. 11 3211
- [56] Shan X, Zhao C, Wang X, Wang Z, Fu S, Lin Y, Zeng T, Zhao X, Xu H, Zhang X, Liu Y 2021 Adv. Sci. 9 2104632
- [57] Liu Y, Wu L, Liu Q, Liu L, Ke S, Peng Z, Shi T, Yuan X, Huang H, Li J, Ye C, Chu P K, Wang J, Yu X F 2021 Adv. Funct. Mater. 32 2110900
- [58] Hu G, An H, Xi J, Lu J, Hua Q, Peng Z 2021 Nano Energy 89 106282
- [59] Kumar M, Lim J, Seo H 2021 Nano Energy 89 106471
- [60] Wang S, Wang C Y, Wang P, Wang C, Li Z A, Pan C, Dai Y, Gao A, Liu C, Liu J, Yang H, Liu X, Cheng B, Chen K, Wang Z, Watanabe K, Taniguchi T, Liang S J, Miao F 2021 *Natl. Sci. Rev.* 8 nwaa172
- [61] Zhang C, Ye W B, Zhou K, Chen H Y, Yang J Q, Ding G, Chen X, Zhou Y, Zhou L, Li F, Han S T 2019 Adv. Funct. Mater. 29 1808783
- [62] Kumar M, Singh R, Kang H, Kim S, Seo H 2020 Nano Energy 73 104756
- [63] Zhang X, Zhuo Y, Luo Q, Wu Z, Midya R, Wang Z, Song W, Wang R, Upadhyay N K, Fang Y, Kiani F, Rao M, Yang Y, Xia Q, Liu Q, Liu M, Yang J J 2020 Nat. Commun. 11 51
- [64] Tan H, Tao Q, Pande I, Majumdar S, Liu F, Zhou Y, Persson P O A, Rosen J, van Dijken S 2020 Nat. Commun. 11 1369
- [65] Kim S H, Baek G W, Yoon J, Seo S, Park J, Hahm D, Chang J H, Seong D, Seo H, Oh S, Kim K, Jung H, Oh Y, Baac H W, Alimkhanuly B, Bae W K, Lee S, Lee M, Kwak J, Park J H, Son D 2021 Adv. Mater. 33 2104690
- [66] Xia Q, Qin Y, Zheng A, Qiu P, Zhang X 2021 Adv. Mater. Interfaces 8 2101068
- [67] Kumar M, Park J Y, Seo H 2021 Small Methods 5 2100566
- [68] Wang D, Wang L, Ran W, Zhao S, Yin R, Yan Y, Jiang K, Lou Z, Shen G 2020 Nano Energy 76 105109
- [69] Shulaker M M, Hills G, Park R S, Howe R T, Saraswat K, Wong H S P, Mitra S 2017 Nature 547 74
- [70] Vidiš M, Plecenik T, Moško M, Tomašec S, Roch T, Satrapinskyy L, Grančič B, Plecenik A 2019 Appl. Phys. Lett. 115 093504
- [71] Lee D, Yun M J, Kim K H, Kim S, Kim H D 2021 ACS Sens.
 6 4217
- [72] Ban C, Min X, Xu J, Xiu F, Nie Y, Hu Y, Zhang H, Eginligil

M, Liu J, Zhang W, Huang W 2021 Adv. Mater. Technol. 6 2100366

- [73] Wang T, Huang H M, Wang X X, Guo X 2021 InfoMat. 3 804
- [74] Gao Z, Chen S, Li R, Lou Z, Han W, Jiang K, Qu F, Shen G 2021 Nano Energy 86 106078
- [75] Lu Q, Sun F, Dai Y, Wang Y, Liu L, Wang Z, Wang S, Zhang T 2021 Nano Res. 15 423
- [76] Vanarse A, Osseiran A, Rassau A 2016 Front. Neurosci. 10 115
- [77] Wang L, Wang Z, Lin J, Yang J, Xie L, Yi M, Li W, Ling H, Ou C, Huang W 2016 Sci. Rep. 6 35273
- [78] Sun L, Zhang Y, Hwang G, Jiang J, Kim D, Eshete Y A, Zhao R, Yang H 2018 Nano Lett. 18 3229
- [79] Wang W, Pedretti G, Milo V, Carboni R, Calderoni A, Ramaswamy N, Spinelli A S, Ielmini D 2018 Sci. Adv. 4 eaat4752

- [80] Rahman M A, Walia S, Naznee S, Taha M, Nirantar S, Rahman F, Bhaskaran M, Sriram S 2020 Adv. Intell. Syst. 2 2000094
- [81] Tan H, Zhou Y, Tao Q, Rosen J, van Dijken S 2021 Nat. Commun. 12 1120
- [82] Mennel L, Symonowicz J, Wachter S, Polyushkin D K, Molina-Mendoza A J, Mueller T 2020 Nature 579 62
- [83] Mohamad Hadis N S, Abd Manaf A, Ngalim S H, Herman S H 2017 Sens. Bio-Sens. Res. 14 21
- [84] Pawar A V, Kanapally S S, Kadam K D, Patil S L, Dongle V S, Jadhav S A, Kim S, Dongale T D 2019 J. Mater. Sci. : Mater. Electron. 30 11383
- [85] Abdul Hadi S, Humood K M, Abi Jaoude M, Abunahla H, Shehhi H F A, Mohammad B 2019 Sci. Rep. 9 9983
- [86] Song Y G, Suh J M, Park J Y, Kim J E, Chun S Y, Kwon J U, Lee H, Jang H W, Kim S, Kang C Y, Yoon J H 2021 Adv. Sci. 9 2103484

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Multimode modulated memristors for in-sensor computing system^{*}

Zhang Yu-Qi^{1)#} Wang Jun-Jie^{2)#} Lü Zi-Yu²⁾ Han Su-Ting^{2)†}

1) (Institute of Microscale Optoelectronics, Shenzhen University, Shenzhen 518060, China)

2) (College of Electronics and Information Engineering, Shenzhen University, Shenzhen 518060, China)

(Received 2 February 2022; revised manuscript received 4 March 2022)

Abstract

To develop future interactive artificial intelligence system, the construction of high-performance human perception system and processing system is vital. In a traditional perceptual and processing system, sensors, memory and processing units are physically separated because of their different functions and manufacture conditions, which results in frequent shuttling and format transformation of data resulting in long time delay and high energy consumption. Inspired by biological sensory nervous system, one has proposed the concept of in-sensor computing system in which the basic unit integrates sensor, storage and computing functions in the same place. In-sensor computing technology can provide a reliable technical scheme for the area of sensory processing. Artificial memristive synapse capable of sensing light, pressure, chemical substances, etc. is one type of ideal device for the application of in-sensor computing system. In this paper, at the device level, recent progress of sensory memristive synapses applied to in-sensor computing systems are reviewed, including visual, olfactory, auditory, tactile and multimode sensation. This review points out the challenge and prospect from the aspects of device, fabrication, integrated circuit system architecture and algorithms, aiming to provide possible research direction for future development of in-sensor computing system.

Keywords: in-sensor computing, memristors, artificial synapses, sensors

PACS: 85.35.-p, 73.40.Rw, 84.30.-r, 87.19.lt

DOI: 10.7498/aps.71.20220226

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 62122055, 62074104, 61974093).

 $^{^{\#}\,}$ These authors contributed equally.

[†] Corresponding author. E-mail: sutinghan@szu.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

神经形态阻变器件在图像处理中的应用

江碧怡 周菲迟 柴扬

Application of neuromorphic resistive random access memory in image processing Jiang Bi-Yi Zhou Fei-Chi Chai Yang

引用信息 Citation: Acta Physica Sinica, 71, 148504 (2022) DOI: 10.7498/aps.71.20220463 在线阅读 View online: https://doi.org/10.7498/aps.71.20220463 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

基于忆容器件的神经形态计算研究进展

Research progress of neuromorphic computation based on memcapacitors 物理学报. 2021, 70(7): 078701 https://doi.org/10.7498/aps.70.20201632

氧化物基忆阻型神经突触器件

Oxide-based memristive neuromorphic synaptic devices 物理学报. 2019, 68(16): 168504 https://doi.org/10.7498/aps.68.20191262

超声图像中复合材料褶皱形态的Mask-RCNN识别方法

Mask-RCNN recognition method of composite fold shape in ultrasound images 物理学报. 2022, 71(7): 074302 https://doi.org/10.7498/aps.71.20212009

N型局部有源忆阻器的神经形态行为

Neuromorphic behaviors of N-type locally-active memristor 物理学报. 2022, 71(5): 050502 https://doi.org/10.7498/aps.71.20212017

基于随机放电神经元网络的彩色图像感知研究

Color image perception based on stochastic spiking neural network 物理学报. 2022, 71(7): 070501 https://doi.org/10.7498/aps.71.20211982

专题: 面向类脑计算的物理电子学

神经形态阻变器件在图像处理中的应用*

江碧怡¹⁾²⁾ 周菲迟^{1)†} 柴扬^{2)‡}

1) (南方科技大学深港微电子学院, 深圳 518000)

2) (香港理工大学应用物理学系,香港 999077)

(2022年3月15日收到; 2022年4月5日收到修改稿)

随着搭载于边缘终端上的图像与视频等数据密集型应用的日益增长,基于传统冯·诺依曼架构的互补金 属氧化物半导体 (complementary metal oxide semiconductor, CMOS) 硬件系统正面临着能耗、速度和尺寸等 多方面的挑战.神经形态器件包括具有存算一体特性的电学阻变器件和具有感存算一体特性的光电阻变器 件,因其具有与生物神经系统的高相似度,及其高能效、高集成度、宽带宽等优势,在图像处理应用方面展现 出巨大发展潜力.这类器件不仅能够用于加速传统图像低阶预处理和高阶处理中的大量运算,且能用于实现 仿生物视觉系统的高效图像处理算法.本文介绍了最近的电学及光电神经形态阻变器件,并结合图像处理算 法综述了神经形态阻变器件在图像处理方面的硬件实施和挑战,并对其发展前景提出了思考.

关键词:神经形态阻变器件,图像预处理,图像识别 **PACS**: 85.35.-p, 95.75.Mn, 87.18.Sn, 84.30.-r

DOI: 10.7498/aps.71.20220463

1 引 言

随着人工智能、大数据与 5G 时代的到来,图 像和视频等应用变得普遍且占据了很大比重,尤其 在移动和嵌入式系统、自动驾驶、机器人、医学图 像分析和工业制造业等应用中.图像及视频处理是 一种数据密集型应用,其数据量和任务复杂度不断 增加,纵使视觉处理算法正在快速优化,大量的计 算负载仍对实施算法的硬件系统构成了挑战,尤其 是在资源有限的移动及嵌入式系统中(如手机人脸 识别^[1]、无人机自动驾驶^[2]、智能机器人交互^[3]等). 与固定设备不同,大多数移动设备的可用空间较 小,更加限制了图像处理模块的计算能力.此外, 在这些由电池供电的移动设备中,电池中的有限能 量对计算单元的低功耗提出了更为严格的要求.因 此,为降低响应时间、功耗和通信带宽,迫切需要 将智能、高效、强大的处理能力集成到图像传感系 统中以进行高能效的实时决策.

图像处理通常可以分为低阶预处理 (如边缘检 测、锐化、运动检测等) 和高阶处理 (如图像认知和 识别等)^[4]. 低级预处理通常涉及在原始和非结构化 数据中对数据的降维、去噪声和特征提取,而高阶 处理涉及图像的抽象表示及认知学习的过程,如识 别、分类和定位. 预处理操作中完成的图像特征初 步提取或图像质量初步提升可有效提高后期高阶 处理的表现. 基于互补金属氧化物半导体 (complementary metal oxide semiconductor, CMOS) 器 件的冯·诺依曼计算系统,如大规模可编程阵列 (field-programmable gate array, FPGA),中央处 理器 (central processing unit, CPU),图形处理器 (graphics processing unit, GPU),专用集成电路

© 2022 中国物理学会 Chinese Physical Society

^{*} 国家自然科学基金 (批准号: 62104091, 62174074)、广东省自然科学基金 (批准号: 2022A1515011064)、广东省青年创新人才基金 (批准号: 2021KQNCX077) 和深圳市南山 5G 前沿项目 (批准号: K21799131, K21799128) 资助的课题.

[†] 通信作者. E-mail: zhoufc@sustech.edu.cn

[‡] 通信作者. E-mail: ychai@polyu.edu.cn

(application-specific integrated circuit, ASIC)等已被广泛用于实现不同的图像低阶及高阶处理^[5-8],然而,冯·诺依曼架构硬件系统中处理、存储和传感功能的分离导致的数据在不同单元之间的频繁传输和转换限制了带宽、功耗及延迟性能的进一步提升,不利于海量传感数据的高能效低功耗处理的需求.

在人类的感官系统中,超过 70% 的信息是由 高效率的视觉感知系统获得^[9].视网膜光感知细胞 将接收到的光信号转化为电信号,并通过视网膜系 统内部的突触和神经元对特征进行初步处理.视网 膜内感光和撤光细胞组成的同心圆区域不仅能传 递输入图像的明暗信息,并能以中心和周围细胞发 射率的差异传递图像对比度的信息,以此完成缘检 测和运动检测等功能^[10,11].此外,生物视网膜系统 对不同频率光输入的不同响应使其具有增强彩色 图像的能力^[12].完成预处理的图像电信号通过视 神经被传递到大脑皮层,进行进一步的记忆和处理^[13], 最终能够从复杂的环境图像中高效识别出大部分 所需特征.因此,模拟生物视觉系统感知是设计高 效能图像处理系统的关键步骤.

神经形态电学阻变器件 (resistive random access memory, RRAM) 和神经形态光电阻变器件 (optoelectronic resistive random access memory, ORRAM) 因其与生物视觉系统功能的高相似度而 常用于构成人工视觉系统,实现多种图像预处理和 高阶处理. 在神经形态 RRAM 中, 电阻值随外部 电压刺激而发生改变的过程与生物突触权重的改 变过程非常相似. 此外, 一些神经形态 RRAM 还可 以展现出与生物神经元相似的脉冲输出特性.因此 其阵列的存算一体特性能有效模仿视觉系统的处 理和记忆功能. 感存算一体的神经形态 ORRAM 阻 值可直接响应于光信号,与高度集成的生物视觉系 统相似度更高,可进一步应用于神经形态视觉传感 器,能够有效避免传感、存储及计算单元间大量冗 余数据的传输及转换,进一步提高了带宽和集成度, 降低了延迟与功耗,同时增强数据的保密性[14-17].

本文首先从材料、器件结构和原理以及神经网络应用等方面分别综述了可用于图像处理的神经形态 RRAM 和 ORRAM;接着文章结合多种图像预处理(包括边缘检测、运动检测和图像增强)和高阶处理(包括图像识别)等算法,综述了基于这些器件及其阵列的硬件实现;最后总结了该领域存在的问题,并提出未来可能的研究方向.

2 电学神经形态阻变器件

2.1 应用于人工神经网络

人工神经网络 (artificial neural network, ANN) 在图像增强^[18]、边缘提取^[19]和图像识别^[20]等图像 处理任务中展现出了优越的性能.然而,不断增加 的 ANN 任务规模和复杂性给基于冯·诺依曼架构 的传统计算机带来了能耗与速度等方面的巨大挑 战.这使得对模仿生物神经系统工作方式的神经形 态器件的研究成为必要. RRAM, 因其能够以阻变 机制模拟生物突触的长程突触可塑性 (long term plasticity, LTP)^[21],常被用作 ANN 的人工突触器 件. LTP 特性包括长程增强特性和长程抑制特性.

以金属导电丝的生成和断裂作为 LTP 机制的 RRAM 通常由活性金属电极 (例如 Cu, Ag 和 Ni)/ 介质层/非活性金属电极 (如 Pt, Pd, W, Au 和 TiN) 组成^[22].在外加电场的影响下,活性金属溶解产生 金属离子向着惰性电极迁移,在此过程中金属离子 逐渐被还原,形成金属导电丝,器件电导率上升, 实现了长程增强特性.反向的电场使导电丝溶解并 断裂,电导率下降,实现了长程抑制特性^[23].2017年 Yuan 等^[24]进一步观察到导电丝的形状、尺寸和阻 态保持时间与电极的尺寸密切相关.为缓解导电丝 生成的随机性对器件性能稳定性的影响,Choi等^[25] 利用 SiGe 中的位错将 Ag 导电金属丝限制在一维 的确定通道中,器件周期性误差和器件间误差分别 仅为 1% 和 4.9%,且开关比可达到 10⁴.

由 HfO₂^[26], TaO_x^[27], TiO₂^[28], SrTiO₃^[29]和ZnO^[30] 等氧化物和非活性金属电极组成的 RRAM 通常由 氧空位导电丝实现 LTP 特性. Chen 等^[31]用透射 电子显微镜技术观测了 Pt/ZnO/Pt RRAM 中氧 空位导电丝的生成和断裂 (图 1(a), (b)); 导电丝生 成和断裂过程中器件的 *I-V*特性曲线如图 1(c) 所 示.导电丝生成过程中,外加电场作用下氧空位的 迁移导致导电丝处的 ZnO 转变为 ZnO_{1-x},器件从 高阻态 (high resistance state, HRS) 转变为低阻 态 (low resistance state, LRS). 而反向电压使氧离 子迁移到底电极附近与氧空位复合,导电细丝断 裂,器件回到 HRS. 为进一步提高器件的多阻态和 线性度, Wu 等^[32]在 HfO_x介质层和底电极之间插 入了 AlO_x阻挡层,将导电丝的破裂和连接限制在 了层间界面上. 此外, TaO_x热增强层也可用于增



图 1 氧空位导电丝型 RRAM^[31] (a) Pt/ZnO/Pt RRAM 内生成的氧空位导电丝; (b) Pt/ZnO/Pt RRAM 内导电丝的断裂; (c) 导电丝生成 (蓝)/断裂 (红) 过程中器件的 *I-V*特性曲线

Fig. 1. Oxygen vacancy conductive filament in RRAM^[31]: (a) Oxygen vacancy conductive filament formed in Pt/ZnO/Pt RRAM; (b) rupture of conductive filament in Pt/ZnO/Pt RRAM; (c) *I-V* characteristic curves of the device during conductive filament formation (blue) and rupture (red).

加存储态个数. TiN/TaO_x/HfO_x/TiN 氧空位导电 丝型 RRAM 阻态数的增加来自热增强层热效应导 致生成的多个弱导电丝.

浮栅型和电解质栅控三端 RRAM 因其额外的 端口而具有较好的稳定性和控制能力,有利于提高 实现算法的运算准确率. HfO₂ 阻挡层/石墨烯浮栅 层/Al₂O₃ 隧穿层/MoS₂ 沟道层结构的浮栅型 RRAM 具有高线性 LTP 特性^[33]. 其 LTP 特性源于浮栅 层在外加电场作用下存储/释放来自沟道的隧穿电 子.恰当的材料介电常数和厚度导致的高层间电容 耦合度提高了电子隧穿效率,实现了 LTP 特性的 高线性度. Seo 等^[34]结合 WSe₂ 沟道/WCL 浮栅通 道 (具有高线性长程增强特性) 和 MoS2 沟道/WCL 浮栅通道 (具有高线性长程抑制特性)构成的双通 道浮栅型 RRAM 同样具有非常高的线性度. PEO-LiClO₄/α-MoO₃电解质栅控型 RRAM 的 LTP 特 性主要来源于 Li 离子的调控^[35]. 该 RRAM 同时具 有高线性 LTP 特性以及 HRS 时的超低通道电导 值 (< 75 nS), 有利于减小器件的漏电流, 提高其 能效.

为实现大规模及高分辨的图像应用,开发基于 RRAM的大规模阵列及稳定性调控至关重要. 乘加运算 (multiply and accumulation, MAC) 是许 多图像处理算法的核心计算单元,如图像识别神经 网络中的向量-矩阵乘法 (vector-matrix multiplication, VMM),算子边缘提取法中卷积核和输入 图片的卷积运算,以及图像平滑算法中滤波器和图 片的卷积运算等.如图 2(a)所示,基于 RRAM (one resistor, 1R)的人工突触阵列可以用于实现 VMM 运算^[36]. 阵列中每个行列交叉点的电流是输入电 压和对应 RRAM 电导的乘积,即利用欧姆定律实 现了乘法运算.此外,根据基尔霍夫电流定律,每 一列输出总电流是每个交叉点的电流之和,即利 用阵列本身的结构实现了加法运算^[36]. 2018年, Bayat 等^[37] 制造了基于 Pt/Al₂O₃/TiO_{2-r}/Ti/Pt RRAM 的 20 × 20 1R 阵列. 基于两个这样的阵列 实现的 16×10×4 全连接神经网络经过在线训练, 于识别三组 4×4 输入字母图片的任务上可达到 70% 的准确率. 此外, Sheridan 等^[38]制造了基于 WO_r RRAM 的 32 × 32 1R 阵列, 并基于此阵列实现了 图像的稀疏编码.稀疏编码的原理是利用从预定义 的特征集中选出的部分特征来取代原图片. 选取特 征时的判定标准为该特征与图片的相似度.1R阵 列中的 RRAM 突触器件被用于存储这些特征,并 接收原图片作为阵列的输入. 编码过程中, 相连突 触存储的特征与输入相似度较低的神经元被抑制, 以此完成输入图片的稀疏编码.

然而, 1R 阵列中的漏电流使计算准确率下降, 因此阵列规模通常较小, 难以应用于大规模、高分 辨的图像处理任务中^[39]. 采用 RRAM 和晶体管相 连 (one transistor one resistor, 1T1R) 的结构能 有效缓解此问题. CMOS 技术的成熟使 1T1R阵列 的制造难度较低, 一些具有较大规模的 1T1R已被 制造并应用于图像处理. 如 Yao 等^[40] 制造了基于 TiN/HfAl_yO_x/TaO_x/TiN RRAM 器件的 128 × 8 1T1R 阵列, 此阵列具有实现 ANN 的能力, 可用于 人脸识别 (图 2(b)); 利用 128 × 64 1T1R的 HfO₂ RRAM 阵列实现的循环型 ANN 被应用于 USF-NIST 数据库中的步态识别, 准确率达 79.1%^[41]; 由选通器件和 RRAM 组成的单元 (one selector



图 2 应用于 ANN 的神经形态 RRAM 阵列 (a) 1R 阵列的 VMM 运算示意图^[36]; (b) 1T1R 阵列实现 ANN 的方式^[40]; (c) 1S1R 阵列结构^[42]

Fig. 2. Neuromorphic RRAM arrays applied to ANN: (a) Schematic diagram of the VMM operation of 1R array^[36]; (b) method of implementing ANN with 1T1R array^[40]; (c) structure of 1S1R array^[42].

one transistor, 1S1R)也可用于抑制漏电流.得益 于选通器件与 RRAM 堆叠的结构,1S1R单元尺寸 通常比1T1R单元更小,因此1S1R 阵列(图 2(c))^[42] 面积开销通常更低.选通器件的机理包括导电丝^[43]、 隧道结^[44]、离子传输^[45]、金属-绝缘体转换^[46]以及 肖特基势垒^[47]等.

2.2 应用于脉冲神经网络

脉冲神经网络 (spiking neural network, SNN) 作为第三代神经网络,更接近于生物神经系统的行 为,其使用脉冲信号传递信息.在该过程中,漏电 整合放电 (leaky integrate and fire, LIF) 神经元^[48] 接收到足够的兴奋信号产生脉冲,通过突触传递给 下一个神经元. 在网络学习过程中, SNN 不仅能够 进行有监督学习,其突触的脉冲时序依赖可塑性 (spiking time dependent plasticity, STDP)^[49] 还 支持了 SNN 的无监督学习行为. 此外, 突触的短程 突触可塑性 (short term plasticity, STP) 有利于模 仿神经系统的遗忘特性,加速网络中较弱的噪声信 号的衰减,进一步提高神经系统相似度,获得更高 性能的 SNN^[50,51]. 神经元的 LIF 特性和突触的 STDP, STP 特性使得 SNN 在处理大量时空信息 上具有优势,因此可用于图像分割^[52]、运动检测^[11]、 图像识别^[53]和图像压缩^[54]等场景. 许多 RRAM 都展现出了 STDP 和 STP 特性, 且可以与电容组 合实现 LIF 特性,因此常被用于 SNN 人工突触器 件和 SNN 人工神经元器件.

两端 RRAM 人工突触器件与生物突触的结构 及功能都具有极高的相似度,能够展现出实现 SNN 所需的 STDP, STP, PPF 特性, 有利于实现较小 面积的二维阵列或高度集成的三维堆叠结构. 2018年 Yan 等^[55] 设计了 Ag/TiO₂:Ag/Pt 结构的 两端 RRAM, 在不同的电压下, 此 RRAM 在金属 导电丝和 Ag 纳米团簇间的电子隧穿行为这两种 导电机制之间切换,实现了电导率的调控.器件的 高速 STDP 行为表现为输入突触的激活脉冲早于 (晚于)输出的激活脉冲时,器件的电导率分别增加 (减少), 且输入输出脉冲到达时间差越小, 电导率 幅值变化越大. 该器件可以通过增加脉冲幅值或 脉冲数量实现从 STP 到 LTP 的转换过程. 双脉冲 易化 (paired pulse facilitation, PPF) 是 STP 的 一种表现形式,即相近的两个激活/抑制脉冲可引 发电导率短暂提升/降低,常由 PPF 因子描述 $(PPF = (G_2 - G_1)/G_1 = C_1 e^{\frac{-\Delta t}{\tau_1}} + C_2 e^{\frac{-\Delta t}{\tau_2}}, G_1, G_2$ 分别为第1个和第2个输入脉冲后的电导率, τ_1, τ_2 为特征衰减时间). 得益于输入脉冲后 Ag 离 子迁移率的残留增强效应, $Ag/TiO_2:Ag/Pt$ RRAM 也展现出了 PPF 行为. Ag/ZrO₂/WS₂/Pt 金属导电丝 RRAM 被证明具有 STDP 和 PPF 特 性,器件的 STDP 特性可由指数函数拟合,即

 $\Delta W = A \times e^{\frac{-\Delta t}{\tau}}, A \pi \tau 分别为强度和时间常数^[56].$ 由于 ZrO₂ 和 WS₂ 层具有不同的离子传输速率,Ag 导电丝的断裂/生成被限制在层间界面上,从而使这一器件具备低功耗和高稳健性的额外优势.

基于 Al/Ti₃C₂T_x/Pt 结构, Yan 等^[57] 报道了具 有 STDP 和 STP 特性的氧空位导电丝型 RRAM, 能够以极窄的脉冲 (10 ns) 进行高速电导率调控. 器件 STP 到 LTP 的转换可通过增加输入脉冲的 数量、频率或幅值实现. Pt/KNbO₃/TiN RRAM 也可展现出 SNN 所需的 STDP(图 3(a), (b)), STP 和强直后增强 (post-tetanic potentiation, PTP) (图 3(c))特性^[58]. PTP 为 STP 的另一种表现, 表 征第 10 个输入脉冲信号产生的输出电流增加程度. 此外, 其导电细丝的生长/溶解主要受到氧离子氧 化还原过程的影响, 氧离子扩散的影响可忽略不 计, 因此还具有高线性度的 LTP 特性. 改变输入 脉冲信号的数量或频率可调控器件的电导率.

扩散型 Au/SiO_xN_y:Ag/Au RRAM 具有 STDP, 尖峰速率依赖可塑性 (spike rate dependent plasticity, SRDP), STP, PPF 以及双脉冲抑制 (paired pulse depression, PPD) 特性^[59]. 基于 Ag 纳米团 簇的阻变机制与生物突触基于 Ca²⁺的权重调制具 有非常高的相似度. 在初始状态下,两组较大的 Ag 纳米团簇分别聚集于顶电极和底电极. 外加正 电压产生的焦耳热使顶电极处的团簇分裂,在电场 作用下逐渐形成导电通路与底电极处团簇连接,电 导率升高. 而电压一旦撤去,团簇将自发扩散重新 聚集到电极附近,电导率下降,团簇的自发扩散以 及在电场下的分解和漂移使扩散型 RRAM 具有 STDP, SRDP, STP 和 PPF 特性. 而其 PPD 特性 来源于过高电压下团簇倾向聚集于底电极的行为.

相比两端结构, 三端结构的 SNN 型 RRAM 人工突触器件可同时利用栅电极和沟道电流分别 进行阻态调控和信号传输操作. 同时, 调控通道和 传输通道的分离可减小阻态调控信号对存储阻态 的干扰, 有利于提高阵列的稳定性和精确度. 基于 DEME-TFSI 离子液体/WO₃ 结构, Yang 等^[60] 设 计了具有 STDP, LTP, PPF 和 PTP 特性的三端 电解质栅控型 RRAM. 离子液体/WO₃ RRAM 的 STP 特性来自低栅压下沟道表面积聚/自发结合的 离子生成的屏蔽电场, 而其 LTP 特性来自于高栅 压下水分子分解产生氢离子嵌入/嵌出沟道的行 为. 栅压调控的 STP 和 LTP 之间的切换可用于模 仿生物神经系统基于注意力的长短记忆机制. 基于



图 3 Pt/KNbO₃/TiN 神经形态 RRAM^[58] (a) 具有 40 µm 时间差的突触前脉冲 (红) 和突触后脉冲 (绿), 以及对应的等效输入 脉冲 (蓝); (b) 器件的 STDP 特性; (c) STP(*I*₂ - *I*₁) 和 PTP(*I*₁₀ - *I*₁) 特性

Fig. 3. Pt/KNbO₃/TiN neuromorphic RRAM^[58]: (a) Presynaptic pulse (red) and postsynaptic pulse (green) with 40 μ m time difference, and the equivalent input pulse (blue) of the RRAM; (b) STDP characteristic; (c) STP ($I_2 - I_1$) and PTP ($I_{10} - I_1$) characteristics.

Li 离子调控机理的 Li_xSiO₂ 电解质层/Nb₂O₅ 沟道 层结构的 RRAM 也具有 STDP 特性^[61]. 32×32 的 Li_xSiO₂/Nb₂O₅ RRAM 阵列与 CMOS LIF 神经 元的系统可用于实现 25×8的 SNN. RbAg₄I₅ 电 解质层/PEO 阻挡层/P₃HT 沟道层结构的低功耗 RRAM 具有 STDP, STP 和 PPF 特性^[62]. 得益于 RbAg₄I₅ 的超离子导电性和高介电常数, 每次阻态 变化仅需 2.0 pJ 能量.

对于人工神经元器件, 传统的 CMOS LIF 神 经元通常由复位电路、电容和比较器组成^[63]. 输入 脉冲可使电容电位逐渐增加,当比较器检测到电容 电位超过阈值时,复位电路将存储的电荷清零,形 成脉冲输出.为了减少面积开销,具有阈值开关 (threshold switching, TS) 特性的 RRAM 常用于 替代比较器和复位器,与电容并联形成 RC LIF 人 工神经元^[64]. Ag/SiO₂/Au RRAM 的 TS 特性表现 为外加电压大于 Vth2 时, Ag 导电细丝生成, RRAM 从 HRS 切换到 LRS. 外部电压小于 V_{th1} 时, 导电 细丝断裂, RRAM 回到 HRS. 其组成的 RC LIF 人 工神经元在整合过程中,输入脉冲通过充电回路 CL 给 RC LIF 人工神经元中的电容充电, RRAM 两 端电压到达 V_{th2}之前, 输出电流可以忽略. 两端电 压到达 V_{th2}之后器件切换到 LRS, 电容通过放电 回路 DL 放电,产生输出脉冲. Ag/SiO₂/Au 器件 电位下降到 V_{th1}时,器件回到初始 HRS,进入下 一轮整合过程.

Pt/Ti/NbO_x/Pt/Ti RRAM 也可以与电容并 联形成 RC LIF 人工神经元^[65]. Duan 等^[65] 制备 了 4×4 的 Pt/Ta/Ta₂O₅/Pt/Ti RRAM 人工突触 阵列, 阵列每行连接 1 个 NbO_x 人工神经元, 以此 实现了具有 4 个神经元的 SNN, 能够分辨输入图 片的模式. 为了降低 RC LIF 人工神经元的功耗, Lu 等^[66] 制备了 Pt/Ag/TiN/HfAlO_x/Pt 结构, 具 有临界开关特性的 RRAM. 由于 TiN 缓冲层对 Ag 扩散的限制作用和 HfO₂/Al₂O₃ 多层结构下导电丝 的非均匀生长, 形成的导电丝较弱, 器件具有极高 的关断电阻 (10¹²)、较小的 LRS 阈值电压 (0.4 V) 和高低阻之间的快速开关速度 (50 ns), 因此与电 容并联后可以实现高效率 LIF 特性.

不同于 RC LIF 人工神经元中电容和 RRAM 并联结构, Wang 等^[67]设计了 Pt/Ta₂O₅/TaO_x/Pt 电容和 Ag/SiO_x:Ag/Ag/Pt 扩散型 RRAM 串联的 电容式 LIF 神经元. 此 RRAM 在开/关状态下的 本征电容值不同,导致了阻态转换后电荷在本征电容和串联电容之间的重新分配,以此实现了 LIF 特性.同时,Wang 等^[67]将电容式 LIF 神经元中的 RRAM 替换为非易失 RRAM,实现了人工突触的功能,最终利用所设计的电容式 LIF 神经元和人工突触实现了完整的脉冲神经网络.

单个 RRAM 也可表现出 LIF 特性. Wang 等^[68] 设计的两端 Ag/V₂C/W RRAM 具有本征 LIF 特性. 器件的漏电、整合、放电特性分别源于无外加 电压时 Ag 离子的扩散, Ag 导电细丝的生长过程, 以及器件的 TS 特性 (图 4(a)). 该 RRAM 实现的 LIF 神经元和其 LIF 特性如图 4(b) 所示. 其输出 脉冲的频率可由输入脉冲的频率和幅值所调控 (图 4(c), (d)). 此外, 基于氧空位导电丝的 Ag/SiO_x/ TiN RRAM 也可以展现出本征 LIF 特性^[69]. 这一 器件的输出脉冲频率不仅与输入脉冲的频率和幅 值成正比, 同时还受 SiO_x 中氧空位的密度调控.

Han 等^[70]设计的 Si/SiO₂/Si₃N₄/SiO₂/Si 浮 栅型神经形态 RRAM 能同时实现 SNN 人工神经 元所需的 LIF 特性和 SNN 人工突触所需的 STDP 特性.器件作为人工神经元时需要撤去外加的正向 栅压 (图 5(a)). 由于 Si₃N₄ 浮栅层的存储能力, 器 件初始处于 HRS, 随后输入漏极的脉冲信号使电 荷积累在器件沟道中,电荷量到达一定阈值后沟道 导通, 电荷被释放形成输出脉冲. 而作为人工突触 (图 5(b))时需要利用栅极接收来自神经元的脉冲 信号, 控制 Si₃N₄ 层内捕获电荷的密度, 从而调控 沟道阈值电压,改变器件的输出电流.人工神经元 和人工突触的连接方式由图 5(c) 所示. 人工神经 元 LIF 行为产生的输出脉冲频率随人工突触权重 的增加而增加(图 5(d)). 基于单一浮栅型神经形 态 RRAM 构建的阵列可实现 SNN, 并应用于字母 图案识别和人脸图像识别的图像处理任务中.

3 光电神经形态阻变器件

不同于传统电学调控 RRAM, ORRAM 的阻 值可直接响应于不同光刺激. ORRAM 具有光可 调控的 LTP, STP, PPF, STDP 和 SRDP 等特性, 不仅可用于模拟生物突触的功能,还能与 LIF 神 经元组合形成光可调控的人工光电神经元器件, 因此可作为高效能人工视觉系统的组成单元. ORRAM 组成的阵列能够应用于感存算一体化



图 4 两端 Ag/V₂C/W 型 RRAM 器件^[68] (a) RRAM 的 TS 特性; (b) RRAM 作为人工 LIF 神经元 (左) 和神经元的 LIF 行为 (右); LIF 人工神经元输出脉冲频率受 (c) 输入脉冲频率和 (d) 输入脉冲幅值调控

Fig. 4. Two-terminal $Ag/V_2C/W$ type RRAM^[68]: (a) TS characteristic of RRAM; (b) RRAM as an artificial LIF neuron (left) and the corresponding LIF behavior (right); modulation of LIF artificial neuron output frequency by (c) the input pulse frequency and (d) the input pulse amplitude.

智能图像传感器,集成图像感知、存储以及计算功能,具有高运行速度、高带宽与低互连功率损耗的优势^[71],可用于图像锐化、降噪、边缘检测、运动目标检测及图像识别^[72-75]等场景.

两端 ORRAM 因其较简单的结构而具有低 功耗,低面积开销和良好的可扩展性等优势,有利 于集成大规模阵列,实现较复杂的图像处理算法. Gao 等^[76]报道了基于 ITO/Nb-SrTiO₃/Ag 结构的 ORRAM. 如图 6(a)所示,光照使界面陷阱电荷脱 离,在正向电压的作用下向顶电极方向移动,留下 带正电的界面氧空位,导致界面肖特基势垒高度和 宽度降低,对应器件电导率升高.而在负电压下, 光生电子与氧空位复合,界面肖特基势垒的宽度和 高度重新增加,器件电导率随之降低.该器件具有 可见光区宽光谱响应、输入光脉冲频率和数量调控的 STP 和 LTP 之间的转换 (图 6(b))及 PPF 等特性,因而适合用于构建人工视觉系统,且能以外加电压幅值的调控模仿人类关注度决定记忆强度的行为 (图 6(c)).

基于 ZnO/NSTO 界面肖特基势垒变化机理, Tan 等^[7]设计了 ITO/ZnO/Nb-SrTiO₃(NSTO) 结 构的 ORRAM. 该器件具有光可调控的 STP, LTP, PPF 和 SRDP 特性. 基于此 ORRAM 的人工传入 神经具有提取手写字母的特征. 前端的压力传感器 和紫外 LED 根据输入压力的强度输出不同频率的 光脉冲信号,末端的 ORRAM 将多组输入的光脉 冲信号融合为一组电脉冲输出信号,实现了特征的 降维提取.



图 5 Si/SiO₂/Si₃N₄/SiO₂/Si 浮栅型神经形态 RRAM^[70] (a) 作为人工神经元; (b) 作为人工突触; (c) 人工突触和突触后神经元 连接方式; (d) 人工神经元 LIF 行为产生的输出脉冲频率与所连接的人工突触权重大小的关系 Fig. 5. Si/SiO₂/Si₃N₄/SiO₂/Si floating gate neuromorphic RRAM device^[70]: (a) As artificial neuron; (b) as artificial synapse; (c) connection of artificial synapse and postsynaptic artificial neuron; (d) effects of connected synaptic weight on the artificial LIF neuron output frequency.

随后, 基于光照作用下 Mo 离子的价态转变机 理, Zhou 等^[78]设计了有紫外光可塑性的 ORRAM (图 7(a)), 并将其首次应用于图像处理. 由图 7(b) 所示, 紫外光照下光生空穴和水分子反应产生 H+ 离子. H+离子和 MoO_x反应产生具有导电特性的 H_yMoO_x, 实现 Mo⁶⁺到 Mo⁵⁺的价态转换, 器件从 HRS 转换到 LRS. 相对地, 负向电压能使 H+离子 从 MoO_x向 Pd 电极漂移, Mo⁵⁺转变回 Mo⁶⁺, 回 到 HRS. 此器件的 STP (图 7(c))和 LTP (图 7(d)) 特性使 ORRAM 人工突触阵列同时具备图像传 感、图像记忆和图像预处理功能^[78].

光照影响导电细丝的断裂和形成这一特性同 样可用来构建 ORRAM. Liu 等^[79] 设计了 Al/TiS₃/ ITO 结构的 ORRAM. 光照条件下, 光生电子抑制 了 Al 原子的氧化, 使导电细丝的断裂更加困难, 实现了电阻值的调控. 受益于 TiS₃ 的宽光谱吸收 能力, 此器件具有宽光谱响应能力. 在相同条件下, 短波长光照下产生的氧化抑制电子数量更多, 电导 值更大. 此外, 该 ORRAM 器件不仅有良好的突触 可塑性, 较大的开关比率, 还具有 STDP 特性.

得益于二维材料所具有的独特光电性质,基于

二维材料的三端 ORRAM 具有优越的光可调控性和宽光谱响应等优点,同时可支持多信号同时输入的优势^[80]. Xiang 等^[81]设计了基于 BN/WSe₂ 异质结结构的三端 ORRAM (图 8(a)). 如图 8(b)所示,负栅压下光生电子从 BN 漂移到 WSe₂ 沟道层,电导率上升,剩下的正电荷储存在 BN 层中^[81].这些正电荷产生的屏蔽电场降低了沟道的阈值电压,实现了器件的存储功能.而正栅压下,光生空穴向沟道漂移,电导率下降.该 ORRAM 在输出电流比率 (1.1×10⁶)、阻态数 (> 128)、阻态保持时间(> 4.5×10⁴)、循环耐久性(> 200 次)上展现出了优良的表现. 此外,具有宽带频谱上的高分辨率响应,这一特性为它在 3×9 阵列中直接检测和存储彩色图像的应用提供了支持 (图 8(c))^[81].

Zhang 等^[82] 报道了 h-BN/WSe₂/Al₂O₃/BP 结构的三端 ORRAM 人工突触,同时具有高线性度和高对称性正光电导和负光电导特性.正栅压使WSe₂层内的电子隧穿到达顶电极,层内余留的空穴产生作用于 BP 沟道的电场,器件电导率上升.随即施加的负向电压和光照使空穴获得能量越过h-BN/WSe₂表面势垒,离开 WSe₂层,器件电导率



图 6 ITO/Nb-SrTiO₃/Ag 结构的神经形态 ORRAM^[76] (a) 光电调控的阻变机理; (b) 通过改变输入光脉冲频率或数量实现的 STP 和 LTP 特性之间的转换; (c) 器件阵列记忆强度随输入电压幅值增加而增强的特性

Fig. 6. ITO/Nb-SrTiO₃/Ag neuromorphic ORRAM^[76]: (a) Optoelectronic resistive switching mechanism; (b) transition between STP and LTP characteristics by changing the frequency or number of input optical pulses; (c) enhanced memory characteristics in the array with increased input voltage amplitude.

下降,产生负光电导.产生正光电导的过程与之相反.WSe₂的宽光谱吸收能力使 ORRAM 能够响应不同波长的光照,因此被用于彩色小车的动作检测和识别.

有机材料因其良好的延展性、灵活性和低制造 成本等优势而被认为是实现 ORRAM 人工突触的 选择之一. Wang 等^[83]设计了有机材料 PTCDA/ 二维材料 MoS₂结构的 ORRAM. 光照条件下,大 量光生电子从 PTCDA 隧穿入 MoS₂沟道,电导率 上升,光照撤去后,电子逐渐回到 PTCDA 层,电 导率降低. 通过控制栅压的幅值或输入光脉冲的 数量可实现 STP 和 LTP 之间的转换. 该 ORRAM 还展现出 SRDP 和 PPF 特性. 此外,零维量子点 材料 (QDs)具有良好的电荷捕获性能和光电响 应能力,因而有利于实现高响应率的 ORRAM. Zhu 等^[84]利用 CNT/CsPbBr₃-QDs 结构在可见光 区实现了具有高光敏度的 ORRAM.在正向栅压 条件下,光照产生的电子被 CsPbBr₃-QDs 所捕获, 而空穴漂移到 CNT 沟道层,器件电导值升高.外 加负向栅压可以使电导值重新降低.此器件对光 输入具有高响应率 (5.1 × 10⁷ A/W)和高探测率 (2 × 10⁶ Jones, 1 Jones = 1 cm·Hz^{1/2}·W⁻¹),同时 也具有 LTP, STP 和 PPF 等生物突触特性.基于 此 ORRAM 制造的 32 × 32 阵列可以模仿学习增 强记忆这一生物特性.在输入阵列的弱光脉冲 (1 μ W/cm²) 从 0 个增加到 200 个的过程中,阵列 权重中存储的图片与输入目标图片的相似度从 约 65% 逐渐提高到 95%.


图 7 ITO/MoO_x/Pd 神经形态 ORRAM^[78] (a) 器件结构; (b) 基于 Mo 离子价态转变的电阻调控机理; (c) ORRAM 的 STP 特性; (d) ORRAM 的 LTP 特性

Fig. 7. $ITO/MoO_x/Pd$ neuromorphic ORRAM^[78]: (a) Device structure; (b) resistive switching mechanism based on change of Mo ion valence state; (c) STP characteristic of ORRAM; (d) LTP characteristic of ORRAM.



图 8 基于 BN/WSe₂ 异质结结构的三端 ORRAM^[81] (a) 器件结构; (b) 光电调控的阻变原理; (c) ORRAM 组成的阵列对不同 波长光输入的不同存储效应

Fig. 8. Three-terminal ORRAM device based on BN/WSe_2 heterostructure^[81]: (a) Device structure; (b) switching mechanisms; (c) different storage levels resulted from different light wavelengths in ORRAM array.





Fig. 9. Au/oxygen-deficient IGZO/oxygen-rich IGZO/Pt ORRAM^[85]: (a) Device structure; (b) conductivity increasing realized by visible light pulses (420 nm) and conductivity decreasing realized by near-infrared light pulses (800 nm); (c) light modulated STDP characteristic.

当前大多数 ORRAM 主要基于光学和电学的 共同调控 (如光学或电学复位),而纯光学 ORRAM 的电阻调控仅依赖于光照,不需要电信号辅助.这 类 ORRAM 构成的人工视觉系统所需的连接更少, 具有高带宽、高计算效率和低互扰等优势,因此被 认为能更好地模拟生物视觉系统的行为.Hu 等^[85] 基于 Au/缺氧 IGZO/富氧 IGZO/Pt 结构实现了 纯光学调控的二端 ORRAM(图 9(a)).如图 9(b) 所 示,施加波长较短的可见光和波长较长的近红外光 分别可以使器件电导率上升和下降,这是由于在短 波长条件下,界面氧空位的电离占主导地位,产生 的电离氧空位带正电,使界面势垒变窄;而在长波 长条件下,由隧穿电子与电离氧空位发生的中和反 应占主导地位,使界面势垒重新增宽^[85].这一器件 还具有光调控的 STDP 特性 (图 9(c))^[85].

Hou 等^[86]设计的两端 Pyr-GDY/石墨烯/PbS 量子点 ORRAM 在栅压为零时,输入 450/980 nm 的光脉冲信号可以得到负/正光电流响应.这一特 性来源于石墨烯和 Pyr-GDY 之间的功函数差异, 短波长光照下, Pyr-GDY 的光响应占主导地位,大 量光生电子进入石墨烯层导致其电导率降低.相 反,长波长光照下, PbS 量子点的光响应占主导地 位,光生空穴进入石墨烯层,电导率升高.此外,通 过集成在红色光 (635 nm)照射下具有正光电导 的 Bi₂O₂Se 材料和在紫外光 (365 nm)照射下具有 负光电导的石墨烯材料,也可以实现光波长调控 的 ORRAM^[87].与生物突触相似,这一器件具有长 程记忆、短程记忆和 PPF 特性.

此外,除了利用光电突触器件,Pei等^[88]将TiN/PbS-QDs/ITO ORRAM人工突触和基于 Ag/MoO_x/

Ag RRAM 的 RC LIF 神经元电路相连接,构造了 人工光电神经元.光照强度增强时,ORRAM 人工 突触电导率上升,输入 LIF 神经元的电流增加,导 致人工光电神经元的输出脉冲频率上升.这一特性 使其可实现于汽车自动驾驶中的碰撞检测.两辆无 人驾驶汽车距离越近,来自对方车灯的光照越强, 人工光电神经元输出脉冲频率升高,汽车速度随之 降低.John 等^[89]设计的基于 ReS₂的 ORRAM 与 CMOS LIF 神经元相连接后 (图 10(a)) 也可以



图 10 基于 ReS₂ ORRAM 与 CMOS LIF 神经元构建的 光可调控神经元^[89] (a) 光可调控神经元结构; (b) 光可调 控神经元输出脉冲频率在光照下增加的行为

Fig. 10. Light tunable artificial neuron based on ReS₂ OR-RAM and CMOS LIF neuron ^[80]: (a) Structure of light tunable artificial neuron; (b) increasing of light tunable artificial neuron output frequency in response to light illumination. 展现出输出脉冲频率受光调控的性质,且可以接受不同波长的光作为输入(图 10(b)).

4 阻变神经形态器件在图像处理的 应用

描述型的传统图像处理算法通常使用人为定 义的特征来判定图像类别,因此很难应对实际应用 中可能出现的未定义场景和噪声干扰. 然而, 具有 自我学习能力的神经网络 (全连接 ANN、卷积神 经网络 (convolutional neural network, CNN) 和 SNN 等) 能够发现输入数据之间的基础规律, 预测 到所有可能的干扰因素, 使实际结果和预测结果之 间的误差最小,因而具有更好的准确率和稳定性[90]. 然而,随着所需处理的数据量和任务复杂度的不断 增加,所需神经网络的规模也在不断增加,大大提 高了功耗和计算时间. 包括运动检测、边缘检测和 图像增强 4 等操作的图像预处理因其计算复杂度 通常低于神经网络的计算复杂度,且同样具有增强 图像中的相关信息,抑制冗余噪声的功能^[91],从而 可以用于减小后端神经网络的计算压力,有助于提 高其效率和准确率. RRAM 和 ORRAM 常被应用 于实现图像预处理和图像识别.相比具有分立传 感,信号转换,计算和存储等模块的传统 CMOS 电 路,基于神经形态器件的系统具有更高的集成度和 并行运算能力,有利于实现更高能效、更低延迟的 图像处理.

4.1 边缘检测

图像的像素值包含了图像的明暗信息,因此, 图像的边缘可以看作像素值阶跃变化的像素点集 合,即像素值的导数较大的位置集合.图像边缘检 测可以通过比较并判定相邻像素值幅值的差别是 否大于特定阈值来实现.随着输入图像维数的增 加,传统的 CMOS 边缘检测电路将在处理器和存 储器之间的数据交换上消耗大量能量.2019年, Chakraborty等^[92]提出用 RRAM 存算一体阵列 并行运算的优势来加速这一差别判定算法,通过训 练得出 RRAM 阵列的一种权重模式,使阵列以这 种模式工作时的输出与差别判定算法的输出具有 最高的相似度.训练所得的 RRAM 阵列可实现 BSD500 数据库的边缘提取.

Pannu 等^[93] 改进了基于 RRAM 阵列的差别

判定边缘检测法, 原始的二输出差别判定算法被改进为三输出 (存在/不存在/不重要)算法. 比起模仿二输出算法的阵列提取的边缘图像, 模仿三输出算法的 8×8 HfO₂ RRAM 阵列提取的边缘图像噪声更少且所需能耗更低. Mannion 等^[94]将输入像素值映射到脉冲频率, 设计了基于 RRAM 分压器的频率差检测电路 (图 11(a))来实现差别判定边缘检测法. 如图 11(b) 所示, 若输入频率不同, RRAM 两端的输入信号无法相互抵消, 两个器件的电导值分别上升/下降, 使输出信号包含两组具有不同幅值的脉冲序列. 输入频率差越大, 幅值的差别越大. 基于 RRAM 分压器的频率差检测电路不需要额外的训练过程, 识别过程中也不依赖于电源和控制信号, 极大地减少了电路复杂程度. 基于这一电路提取的图片边缘如图 11(c) 所示.

算子边缘提取法利用差分算子和图像卷积的 结果估计图像像素值的梯度,检测图像的边缘.每 个差分算子中包括两个卷积核,分别用于估计横向 和纵向梯度,最终输出为两组卷积结果的平方和. RRAM 阵列常用于加速此卷积运算.Li等^[95]制造 了基于 Ta/HfO₂/Pd RRAM 的1T1R 阵列来加速 Sobel 算子的卷积运算.此器件具有高线性度、多 阻态和高产量 (99.8%)的优点,因此其阵列可以达 到较大规模 (128×64),支持多个卷积核并行运算. Lin 等^[96]设计的三维 HfO₂ RRAM 阵列可用于加 速基于 Prewitt 算子的边缘提取运算.

蚁群边缘检测算法是一种高效的生物启发式 算法. 通过将路径长度设置为像素值之间的差, 即 利用蚁群通过信息素的更新选择最短路径的行为 来进行边缘检测. 而该行为与 RRAM 电导值受输 入电流影响而更新的趋势具有很高的相似度[97]. 在实现蚁群优化边缘检测算法的阵列中,每一个 RRAM 单元对应一个像素值, ML 和 MU 控制当前 单元与左/右单元连接模拟蚂蚁行进方向, Mpp 用 于更新电导值 (模拟信息素更新). 图片边缘提取完 毕后, 以 M_{read} 和 M_{ini} 分别控制读取结果和重新置 位. Yu 等^[98] 通过在图像外部填充对称像素, 定义 了蚂蚁运动的边界,进一步提高了基于 RRAM 的 蚁群边缘检测法的精确度,所设计的以外部控制电 路和 1R 阵列组成的 RRAM 蚁群边缘检测电路具 有更小的面积开销.改进的蚁群边缘检测法提取的 边缘比 Sobel 算子提取的边缘更加清晰.



图 11 基于神经形态阻变器件频率差检测电路实现的图像边缘提取^[94] (a) 基于 RRAM 分压器的频率差检测电路 (右) 和所使 用的器件结构 (左); (b) 两组输入脉冲频率相同 (左) 和不同 (右) 时频率差检测电路的输出; (c) 原图和频率差检测电路提取的图 片边缘

Fig. 11. Edge detection based on frequency difference circuit implemented by neuromorphic RRAM^[94]: (a) Frequency difference detection circuit (right) and the adopted RRAM (left); (b) output of the frequency difference detection circuit when two sets of the input pulses are at the same frequency (left) and different frequencies (right), respectively; (c) original image and extracted edges by frequency difference detection circuit.

光感受细胞、双极细胞和输入神经节细胞组成的生物视网膜系统,可以将光信号中心强度和边缘强度之间的差别编码为不同频率的脉冲输出(见图 12(a)),以此完成边缘提取图像预处理操作^[99]. Bao等^[99]基于此设计了如图 12(b)所示的基于HfO_x RRAM和 CMOS 晶体管的人工视网膜单元,其中晶体管 T与 RRAM 组成了人工双极细胞,剩余晶体管实现了神经节细胞的 LIF 特性.如图 12(c)所示,人工视网膜单元输出脉冲的频率同时与 V_{th}端口的输入(模拟光信号输入)和 input 端口的输入(模拟来自其他神经元的生物脉冲输入)相关^[99]. 以该人工视网膜单元组成的人工视网膜网络因其高度并行的处理模式而具有很高的效率,有利于实现移动终端中的实时边缘提取功能.

4.2 运动检测

运动检测是从时间图像序列中检测物体运动 行为的过程,常用于智能监控^[100]、交通情况检测^[101] 和运动物体追踪^[102]等场景. RRAM 具有的存算一 体特性使前一帧图片信息的存储和部分计算可以 在同一器件内完成,有利于减少器件的面积和能 耗. ORRAM 具有的感存算一体特性可进一步提 高运算效率. 差别判断是常见的运动检测方案之 一, 通过检测连续两帧图片之间的差别, 并将差值 与特定阈值相比较,可以判断运动行为是否发生. Maan 等^[103]设计了基于 RRAM 的二元可变电阻 阈值逻辑单元来进行差别判断. 输入图像的每4个 像素连接到一个四输入的逻辑单元. 在检测过程 中, t₁时刻的4个图像像素(x1-4)分别输入到存储 着 $t_1 - \Delta t$ 时刻图像信息 (w_{1-4}) 的 4 个 RRAM 中. Δt 为前后两帧图片的时间差. 当 t_1 与 t_1 - Δt 时刻的 输入有显著不同时, RRAM 的输出电流之和超过 末端反相器的阈值, 使输出翻转. 完成当前时刻的 检测后, 位于器件前端的训练电路将 t₁ 时刻的输 入信息写入权重, 以便 $t_1+\Delta t$ 时刻的判断. 实验证 明该二元可变电阻阈值逻辑单元所组成的阵列可 以实现运动目标的追踪.

Zhang 等^[82] 设计了一个基于三端 h-BN/WSe₂/ Al₂O₃/BP ORRAM 阵列的神经形态运动检测电路. 两个正光电导和负光电导 (*W*/-*W*) ORRAM 阵



图 12 基于 RRAM 和 CMOS 晶体管人工视网膜单元实现的边缘提取^[99] (a) 生物视网膜系统 (光感受-双极-神经节细胞) 对不 同输入光照的不同输出脉冲频率; (b) 人工视网膜单元结构; (c) 人工视网膜单元输出信号 V₀ 随 V_{th} 端口输入信号和 input 端口 输入信号的变化

Fig. 12. Unit of artificial retinal system based on RRAM for edge extraction^[99]: (a) Different output frequencies of the biological retinal system (photoreceptor cells-bipolar cells-ganglion cells) in response to different light pulse inputs; (b) structure of artificial retinal system unit; (c) change of the artificial retinal system unit output signal V_0 with respect to input signals from $V_{\rm th}$ port and input port.

列分别用于感应先后两帧图片 $(X_{t_1}/X_{t_1+\Delta t})$, 阵 列输出的叠加 $(WX_{t_1}-WX_{t_1+\Delta t})$ 即为运动检测的 结果. 随后,为进一步完成目标识别, ORRAM 被 用于构建 ANN. h-BN/WSe₂/Al₂O₃/BP ORRAM 的高线性度和对称性 LTP 特性使构建的 ANN 表 现出良好的抗噪声能力.

基于图像帧的传统运动检测机制比起生物视 网膜光感受-双极-神经节细胞给光/撤光反应运动 检测机制,其能耗大且延迟高.Wang等^[11]利用 Ag/HfO₂/C RRAM 阵列成功模拟了给光/撤光反 应机制(图 13(a),(b)),实现了目标运动方向的高 效检测.如图 13(c)所示,Ag/HfO₂/C RRAM 的 STP 特性使人工神经节细胞可以捕获输入信号的 时空特性,从给光区运动到撤光区(A 到 B)和从 撤光区运动到给光区(B 到 A)时输出电流脉冲的 方向相反,以此判断运动方向^[11].图 13(d)示意了 包含基于 4 个人工神经节细胞的 RRAM 阵列,其 可实现全方位方向检测.对这一检测结果的进一步 识别可以由 ANN 完成, 当方向精度为 15°时, ANN 识别准确率达到 83.92%^[11].

除此之外,受蝗虫小叶巨型运动检测神经元 (lobula giant movement detector, LGMD) 启发, Jayachandran 等^[104] 设计了一个基于 RRAM 和光 感受器堆叠的人工 LGMD 神经元, 具有进行碰撞 检测的功能.物体靠近时,这一神经元对接收到的 兴奋(源于物体靠近)和抑制(源于背景)信号做非 线性数学运算. 其输出与时间呈非线性关系, 呈现 出单个尖峰,这一尖峰即为触发逃跑反应的信号. 人工 LGMD 神经元器件电导率随光脉冲增加,随 背栅电信号降低,以及输出电流随时间的变化,分 别与 LGMD 神经元的兴奋, 抑制和信号输出过程 相似 (见文献 [104] Fig. 1(c)-(i)). 因此, 该器件可 以模拟 LGMD 的碰撞检测行为. 人工 LGMD 神 经元不仅能对具有不同速度的接近物体进行碰撞 检测,还可通过调整背栅电压来调整器件对高速/ 低速物体检测的精度.

为了进一步缩减人工 LGMD 神经元的面积开 销, Wang 等^[105] 利用如图 14(a) 所示的 Ag/FLBP- CsPbBr₃/ITO ORRAM 模拟了生物 LGMD 细胞 输出脉冲频率随时间的非线性行为 (图 14(b)).



图 13 基于 Ag/HfO₂/C RRAM 人工神经节细胞实现的运动检测^[11] (a) 具有给光/撤光反应机制的生物视网膜系统结构; (b) 人工神经节细胞结构; (c) 人工神经节细胞工作原理; (d) 包含 4 个人工神经节细胞的 RRAM 阵列

Fig. 13. Artificial ganglion cell based on $Ag/HfO_2/C$ RRAM for motion detection^[11]: (a) Structure of biological retinal system with both excitation and inhibition response to optical input; (b) structure of artificial ganglion cells; (c) working principle of artificial ganglion cells; (d) RRAM array realized with four artificial ganglion cells.



图 14 基于 Ag/FLBP-CsPbBr₃/ITO ORRAM 类眼球形阵列实现的运动检测^[105] (a) 单个器件的结构; (b) 生物 LGMD 细胞输 出脉冲频率对接近物体的非线性反应; (c) 基于 Ag/FLBP-CsPbBr₃/ITO ORRAM 实现的人工 LGMD 神经元对生物 LGMD 神经 元非线性响应特性的模仿; (d) 柔性 Ag/FLBP-CsPbBr₃/ITO ORRAM 构建的类眼球形阵列

Fig. 14. Ag/FLBP-CsPbBr₃/ITO ORRAM array based biometric compound eye for motion detection^[105]: (a) Structure of single device; (b) nonlinear response to approaching objects regarding output spike frequency of biological LGMD cell; (c) emulation of the nonlinear response properties in biological LGMD neuron by artificial LGMD neuron based on Ag/FLBP-CsPbBr₃/ITO ORRAM; (d) flexible Ag/FLBP-CsPbBr₃/ITO ORRAM array as biometric compound eye.

连续施加的光脉冲可用于模拟物体的靠近,物体距 离较远时,光照产生焦耳热效应导致器件温度升 高,合理的温度促进Ag导电丝的形成,电导率增 加;物体距离过近时,过高的温度使导电丝熔解断 裂,电导率降低,形成尖峰(图 14(c))^[105].该柔性器 件组成的类眼球形阵列(图 14(d))比起单个器件 具有更大的接受角,且可以分辨接近物体的方向和 速度.Ag/FLBP-CsPbBr₃/ITO器件也可以与电 容组合形成人工 RC LIF LGMD 神经元.这一人 工神经元随后被成功地应用于小车避障任务中.

4.3 图像增强

图像锐化和图像平滑均可用于增强图像质量. 图像锐化强调对比度的增强,而图像平滑强调噪声的衰减^[106].平均值滤波器、高斯滤波器、圆形均值 滤波器、拉普拉斯滤波器等平滑滤波器与图像的卷 积可以实现图像平滑. RRAM 或 ORRAM 组成的 阵列常用于加速此卷积运算.例如,Li等^[95]设计 的基于 Ta/HfO₂/Pd RRAM 的 1T1R 阵列可实现. 为进一步提高运算效率, Wang 等^[107] 设计了基于 WSe₂/BN/Al₂O₃ ORRAM 的传感器阵列,利用 此 ORRAM 阻态由背栅电压和光照共同决定的特 性, 拉普拉斯滤波器权重和输入图片像素值分别被 映射到背栅电压和光照强度上, 以此完成卷积运 算. 感存算一体的特性使这一阵列具有高效率和低 电路复杂度的优点.

RRAM和ORRAM的本征非线性阻变特性 常被用于图像锐化. 基于 ITO/MoO_x/Pd ORRAM, Zhou 等^[78]构建了 8×8 ORRAM 阵列 (图 15(a)), 并证明了该阵列具有图像锐化能力.图 15(b)展示 了 ITO/MoO_x/Pd ORRAM 阵列的非线性阻变特 性.阵列输出电流与输入光脉冲强度不呈线性关 系,且器件衰减时间的非线性进一步扩大了输出电 流之间的差值^[78].以该阵列实现的图片锐化处理, 可以提高后续神经网络的识别准确率和识别速率 (图 15(c))^[78].





Fig. 15. $ITO/MoO_x/Pd$ ORRAM array for image sharpening^[78]: (a) 8×8 ITO/MoO_x/Pd ORRAM array; (b) nonlinear resistance switching characteristics of the ORRAM array; (c) an artificial vision system based on ORRAM image sharpening array and image recognition neural network.

Yang 等^[108]设计的自供电光电突触具有相似的本征非线性阻变特性,由 SiN_x/n-Si/p-Si/Al太阳能电池光探测器和 ITO/CsPbBr₂I/P₃HT/AgRAM人工突触组成.光电突触阵列对输入强度和输入波长均具有非线性输出和非线性衰减特性,因此可用于彩色图像的图像增强.彩色手写图片数据库经过图像增强处理后,神经网络识别准确率上升了 5.31%,识别速率上升了 85.71%.用于复杂的MNIST 数据库时,图像增强的作用更加显著,神经网络准确率从 77.31% 上升至 85.45%.

基于 Ni/NiO/Ni RRAM 阵列的两步非线性图 像锐化算法可达到与传统锐化算法相似的性能^[109]. 无控制信号条件下, 忆阻器的本征非线性只能增加 输入信号中间亮度范围内的对比度, 而添加了控制 信号的的两步非线性可以拓展图片全亮度范围内 的对比度, 包括过亮/过暗部分.运算过程中, 灰化 的原始图片每4个像素最暗像素作为控制信号输 入 RRAM 阵列, 再将原始图片像素值作为输入信 号输入同一阵列, 阵列输出电流即为锐化后的图片 像素值.与传统锐化算法相比 (21624 kB, 0.542 s), 具有相似性能的 Ni/NiO/Ni RRAM 阵列能耗更 低, 速度更快 (6816 kB, 0.047 s).

扩散型 RRAM 的 STP 和阈值阻态变换特性 使其具有本征阈值开关特性,可以实现用于图像平 滑的均值滤波器^[110]. 阵列大小与输入图片像素数 相同,其中每种扩散型 RRAM 的 LRS 阈值都被设 置为对应像素周围 8 个像素的平均值. 此外,为缓 解因均值滤波器滤除原图边缘造成的图像质量 下降问题,另一扩散型 RRAM 阵列被用于边缘增 强.以扩散型 RRAM 阵列均值滤波器、扩散型 RRAM 阵列边缘增强滤波器与漂移型 RRAM 阵列组成的预处理-ANN 网络,在受白噪声影响的 MNIST 数据库识别任务中准确率可达 91.55%.

4.4 图像识别

在不同类型的 ANN 中, CNN 常被认为在图 像识别任务上具有更好的表现^[111]. 它能通过卷积 捕捉相邻像素之间的关系, 同时减少网络中的数据 量, 因此有利于处理高维图像信息, 在一些复杂的 图像识别任务中, 能以更低的功耗取得比全连接 ANN 更好的表现.

RRAM 组成的阵列可用于加速 CNN 中大量的卷积运算. Yao 等^[112]用基于 TiN/TaO_x/HfO_x/

TiN RRAM 阵列的硬件系统实现了 5 层结构的 CNN,支持卷积核间并行卷积运算.所用的离线和 在线训练相结合的训练方式使实现的 CNN 对于 器件本身产生的噪声的容忍能力更强,在手写数字 识别任务 (MNIST 数据库)上的准确率从 93.86% (仅离线训练)上升至 95.83%(离线和在线相结合). 这一硬件系统消耗的能量仅为实现同样 CNN 的 Tesla V100 GPU 的1/110.以该硬件系统实现的 更大规模的 CNN (ResNET-56)可用于更复杂的 图像识别任务 (CIFAR10 数据库)中,准确率为 95.57%,仅比理想情况低 1.49%.

Lin 等^[96]设计的具有像素级并行操作能力的 三维 HfO₂ RRAM 阵列不仅支持多个卷积核之间 的并行运算,且支持每个卷积核与整个图像卷积的 并行运算.卷积核的每行权重被映射到一组以楼梯 形电极连接的 RRAM 上.因此,相同像素和卷积 核不同行权重之间的多次乘加操作可以通过支持 三维叠加的楼梯形 RRAM 连接结构并行完成.三 维 RRAM 阵列的运行速度远快于 TPU(V1) 和 GPU(RTX 6000).以此阵列实现的四层 CNN 在 手写识别任务上可达到与理想情况几乎一致的准 确率 (理想 98.11%/三维阵列 98.10%).

Wang 等^[107] 基于 WSe₂/BN/Al₂O₃ ORRAM 的感存算一体传感器阵列实现了图像识别 CNN. 3个 ORRAM 阵列组成了具有 3个卷积核的单层 CNN. CNN 的 3个输出分别对应输入图片为字母 n, j, u 的可能性. 实现的 CNN 在 10 轮训练后可 达到 100% 的准确率.

SNN 可直接处理来自基于事件的视觉传感器的脉冲输入,且因其低计算复杂度而具有低延迟和低功耗的优点^[113],因此也被用于图像识别任务中. Boybat 等^[114]利用多 RRAM 并联的人工突触结构缓解了单个 GST RRAM 作为人工突触时的阻态高度非线性问题,并基于改进的人工突触构建了单层 SNN 进行手写字母识别 (MNIST 数据库). 当单个人工突触中并联的 GST RRAM 数量达到 7 个时,构建的 SNN 准确率高于 77%,与理想情况 (77.02%) 十分相近. Wang 等^[115]利用基于 Pt/ SiO_aN_y:Ag/Pt RRAM 的 RC LIF 人工神经元和 基于 Pd/HfO₂/Ta 氧空位导电丝型 RRAM 的人 工突触实现了 SNN,并成功利用该 SNN 分辨输入 的字母图片.

为提高 SNN 识别的准确率, Li 等 [116] 用 Pt/

TaO_x/AlO_δ/Al RRAM 模仿了生物神经元中树突的噪声抑制和非线性整合功能. 该人工树突器件只有在输入信号大于阈值 (3 V)后才会开启, 进行整合, 因此可以阻止幅值较小的噪声信号通过, 放大图片信息. 人工树突器件与基于 TiN/HfO_x/TaO_x/TiN RRAM 的人工突触阵列、基于 NbO_x RRAM 的 LIF 人工神经元一起实现了生物相似度更高的SNN. RRAM 人工树突的加入使实现的 SNN 在门牌号数字图片识别任务上 (SVHN 数据库)准确率从 80.1% 上升到 88.5%.

5 结 论

神经形态 RRAM 和 ORRAM 因其存算一体 或感存算一体特性,在图像预处理和高阶处理应用 中具有实现低功耗,高集成度人工视觉系统的潜 力.然而,这一领域的研究工作仍处于发展阶段, 算法和硬件方面仍亟需进一步提升.

在器件层面上,尽管 ORRAM 人工突触被证 明在实现高效能人工视觉系统上占有优势,但研究 人员对这一类器件的工作机理和塑性调控机制尚 未深入研究,尤其是可实现纯光学电阻调控的 ORRAM 限制未来器件性能及面向不同图像应用 的进一步提升. 基于有机或二维材料实现的吸收特 性及可调控性不利于未来大规模集成及高分辨图 像应用. 而基于氧化物实现的 ORRAM 仍需从器 件结构设计和机理上突破. 进一步提升 ORRAM 的 集成度、改善光反应的灵敏度以及优化 ORRAM 的结构,是改进 ORRAM 的 3 个重要策略.目前 对 RRAM 的研究相对更加成熟,已有一些研究汇 报了较大规模阵列的实现. 然而, RRAM 的非线性 度和不稳定性使其相比于传统 CMOS 器件实现的 人工视觉系统表现较差,不利于需要较高精确度的 应用场景. 进一步缩小 RRAM 性能与需求的差距 仍然是未来热门的研究方向.

在算法层面上,神经形态器件在实现图像预处 理算法方面的潜力尚未充分挖掘.目前大部分应用 集中在利用神经形态器件组成的阵列加速传统预 处理算法中的卷积或乘加运算.而考虑到生物视觉 系统的高效性,更多模仿生物图像预处理机制,且 可以 RRAM 和 ORRAM 的阻变特性所模拟的生 物图像预处理算法应当被设计出来,实现软硬件协 同发展. 在系统层面上,当前外围电路设计的缺乏不利 于实现信号在系统内部的高效调度和不同模块之 间的分工协作.更多可应用于实际场景的全硬件 RRAM或ORRAM人工视觉系统仍待开发.

参考文献

- Ma Y, Wu J, Long C, Lin Y B 2021 IEEE Internet Things J. 9 2802
- [2] Machida F, Andrade E 2021 2021 IEEE 5th International Conference on Fog and Edge Computing (ICFEC) Melbourne, Australia, May 10–13, 2021 p66
- [3] Pilli S K, Nallathambi B, George S J, Diwanji V 2015 2014 2nd International Conference on Electronics and Communication Systems (ICECS) Coimbatore, India, Feburary 26–27, 2014 p1
- [4] Chaki J, Dey N 2018 A Beginner's Guide to Image Preprocessing Techniques (Vol. 1) (Boca Raton: CRC Press)
- Zhang J F, Lee C E, Liu C, Shao Y S, Keckler S W, Zhang Z 2019 2019 Symposium on VLSI Circuits Kyoto, Japan, June 9–14, 2019 pC306
- [6] Kinget P, Steyaert M S J 1995 IEEE J. Solid-State Circuits 30 235
- Yin S, Ouyang P, Zheng S, Song D, Li X, Liu L, Wei S 2018
 2018 IEEE Symposium on VLSI Circuits Honolulu, HI, USA, June 18–22, 2018 p139
- [8] Rao M V G, Kumar P R, Prasad A M 2016 2016 International Conference on Microelectronics, Computing and Communications (MicroCom) Durgapur, India, January 23-25, 2016 p1
- [9] Treichler D 1967 Film and AV Communication 1 14
- [10] Róka A, Csapó Á, Reskó B, Baranyi P 2007 Acta Polytech. Hung. 4 31
- [11] Wang W, Covi E, Milozzi A, Farronato M, Ricci S, Sbandati C, Pedretti G, Ielmini D 2021 Adv. Intell. Syst. 3 2000224
- [12] Webster M A 1996 Netw. Comput. Neural Syst. 7 587
- [13] Sabesan R, Schmidt Brian P, Tuten William S, Roorda A 2016 Sci. Adv. 2 e1600797
- [14] Cheng Z, Ríos C, Pernice W H P, Wright C D, Bhaskaran H 2017 Sci. Adv. 3 e1700160
- [15] Zhu X, Lu W D 2018 ACS Nano 12 1242
- [16] Liao F, Zhou F, Chai Y 2021 J. Semicond. 42 013105
- [17] Chai Y 2020 Nature **579** 32
- [18] Li C, Guo J, Porikli F, Pang Y 2018 Pattern Recognit. Lett. 104 15
- [19] Khan M Z, Harous S, Hassan S U, Ghani Khan M U, Iqbal R, Mumtaz S 2019 IEEE Access 7 72622
- [20] Ni L, Huang H, Liu Z, Joshi R V, Yu H 2017 ACM J. Emerg. Technol. Comput. Syst. 13 1
- [21] Rajendran B, Alibart F 2016 IEEE J. Emerg. Sel. Top. Circuits Syst. 6 198
- [22] Shi T, Wang R, Wu Z, Sun Y, An J, Liu Q 2021 Small Struct. 2 2000109
- [23] Yang Y, Gao P, Gaba S, Chang T, Pan X, Lu W 2012 Nat. Commun. 3 732
- [24] Yuan F, Zhang Z, Liu C, Zhou F, Yau H M, Lu W, Qiu X, Wong H S P, Dai J, Chai Y 2017 ACS Nano 11 4097
- [25] Choi S, Tan S H, Li Z, Kim Y, Choi C, Chen P Y, Yeon H, Yu S, Kim J 2018 *Nat. Mater.* **17** 335
- [26] Chandrasekaran S, Simanjuntak F M, Saminathan R, Panda

D, Tseng T Y 2019 Nanotechnology $\mathbf{30}$ 445205

- [27] Zhao X, Zhang K, Hu K, Zhang Y, Zhou Q, Wang Z, She Y, Zhang Z, Wang F 2021 *IEEE Trans. Electron Devices* 68 6100
- [28] Dash C S, Sahoo S, Prabaharan S R S 2018 Solid State Ionics 324 218
- [29] Nili H, Ahmed T, Walia S, Ramanathan R, Kandjani A E, Rubanov S, Kim J, Kavehei O, Bansal V, Bhaskaran M, Sriram S 2016 Nanotechnology 27 505210
- [30] Simanjuntak F M, Chandrasekaran S, Lin C C, Tseng T Y 2019 APL Mater. 7 051108
- [31] Chen J Y, Hsin C L, Huang C W, Chiu C H, Huang Y T, Lin S J, Wu W W, Chen L J 2013 Nano Lett. 13 3671
- [32] Wu W, Wu H, Gao B, Deng N, Yu S, Qian H 2017 IEEE Electron Device Lett. 38 1019
- [33] Park E, Kim M, Kim T S, Kim I S, Park J, Kim J, Jeong Y, Lee S, Kim I, Park J K, Kim G T, Chang J, Kang K, Kwak J Y 2020 Nanoscale 12 24503
- [34] Seo S, Kang B S, Lee J J, Ryu H J, Kim S, Kim H, Oh S, Shim J, Heo K, Oh S, Park J H 2020 Nat. Commun. 11 3936
- [35] Yang C S, Shang D S, Liu N, Fuller E J, Agrawal S, Talin A A, Li Y Q, Shen B G, Sun Y 2018 Adv. Funct. Mater. 28 1804170
- [36] Zhang W, Pan L, Yan X, Zhao G, Chen H, Wang X, Tay B K, Zhong G, Li J, Huang M 2021 Adv. Intell. Syst. 3 2100041
- [37] Bayat F M, Prezioso M, Chakrabarti B, Nili H, Kataeva I, Strukov D 2018 Nat. Commun. 9 2331
- [38] Sheridan P M, Cai F, Du C, Ma W, Zhang Z, Lu W D 2017 Nat. Nanotechnol. 12 784
- [39] Cassuto Y, Kvatinsky S, Yaakobi E 2013 2013 IEEE International Symposium on Information Theory Istanbul, Turkey, July 7–12, 2013 p156
- [40] Yao P, Wu H, Gao B, Eryilmaz S B, Huang X, Zhang W, Zhang Q, Deng N, Shi L, Wong H P, Qian H 2017 Nat. Commun. 8 15199
- [41] Li C, Wang Z, Rao M, Belkin D, Song W, Jiang H, Yan P, Li Y, Lin P, Hu M, Ge N, Strachan J P, Barnell M, Wu Q, Williams R S, Yang J J, Xia Q 2019 Nat. Mach. Intell. 1 49
- [42] Li Y, Tang J, Gao B, Sun W, Hua Q, Zhang W, Li X, Zhang W, Qian H, Wu H 2020 Adv. Sci. 7 2002251
- [43] Zhao X, Ma J, Xiao X, Liu Q, Shao L, Chen D, Liu S, Niu J, Zhang X, Wang Y, Cao R, Wang W, Di Z, Lv H, Long S, Liu M 2018 Adv. Mater. 30 1705193
- [44] Choi B J, Zhang J, Norris K, Gibson G, Kim K M, Jackson W, Zhang M X, Li Z, Yang J J, Williams R S 2016 Adv. Mater. 28 356
- [45] Ohba K, Yasuda S, Mizuguchi T, Sei H, Tsushima T, Shimuta M, Shiimoto T, Yamamoto T, Sone T, Nonoguchi S, Kouchiyama A, Otsuka W, Aratani K, Tsutsui K 2018 2018 IEEE International Memory Workshop (IMW) Kyoto, Japan, May 13–16, 2018 p1
- [46] Kim W G, Lee H M, Kim B Y, Jung K H, Seong T G, Kim S, Jung H C, Kim H J, Yoo J H, Lee H D, Kim S G 2014 2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers Honolulu, HI, USA, June 9–12, 2014 p1
- [47] Lu D, Zhao Y, Anh T X, Yu Y H, Huang D, Lin Y, Ding S J, Wang P F, Li M F 2014 *IEEE Trans. Electron Devices* 61 2294
- [48] Farsa E Z, Ahmadi A, Maleki M A, Gholami M, Rad H N 2019 IEEE Trans. Circuits Syst. II Express Briefs 66 1582
- [49] Hu D, Zhang X, Xu Z, Ferrari S, Mazumder P 2014 14th

IEEE International Conference on Nanotechnology Toronto, Canada, August 18–21, 2014 p873

- [50] Lameu E L, Borges F S, Iarosz K C, Protachevicz P R, Antonopoulos C G, Macau E E N, Batista A M 2021 Commun. Nonlinear Sci. Numer. Simul. 96 105689
- [51] Tsodyks M V, Markram H 1997 Proc. Natl. Acad. Sci. USA 94 719
- [52] Meftah B, Lezoray O, Benyettou A 2010 Neural Process. Lett. 32 131
- [53] Iakymchuk T, Rosado Muñoz A, Guerrero Martínez J F, Bataller Mompeán M, Francés Víllora J V 2015 Eurasip J. Image Video Process. 2015 4
- [54] Cho S G, Beigne E, Zhang Z 2019 2019 IEEE Custom Integrated Circuits Conference (CICC) Austin, TX, USA, April 14–17, 2019 p1
- [55] Yan X, Zhao J, Liu S, Zhou Z, Liu Q, Chen J, Liu X Y 2018 Adv. Funct. Mater. 28 1705320
- [56] Yan X, Qin C, Lu C, Zhao J, Zhao R, Ren D, Zhou Z, Wang H, Wang J, Zhang L, Li X, Pei Y, Wang G, Zhao Q, Wang K, Xiao Z, Li H 2019 ACS Appl. Mater. Interfaces 11 48029
- [57] Yan X, Wang K, Zhao J, Zhou Z, Wang H, Wang J, Zhang L, Li X, Xiao Z, Zhao Q, Pei Y, Wang G, Qin C, Li H, Lou J, Liu Q, Zhou P 2019 Small 15 1900107
- [58] Lee T H, Hwang H G, Woo J U, Kim D H, Kim T W, Nahm S 2018 ACS Appl. Mater. Interfaces 10 25673
- [59] Wang Z, Joshi S, Savel'ev S E, Jiang H, Midya R, Lin P, Hu M, Ge N, Strachan J P, Li Z, Wu Q, Barnell M, Li G L, Xin H L, Williams R S, Xia Q, Yang J J 2017 Nat. Mater. 16 101
- [60] Yang J T, Ge C, Du J Y, Huang H Y, He M, Wang C, Lu H B, Yang G Z, Jin K J 2018 Adv. Mater. 30 1801548
- [61] Li Y, Lu J, Shang D, Liu Q, Wu S, Wu Z, Zhang X, Yang J, Wang Z, Lv H, Liu M 2020 Adv. Mater. 32 2003018
- [62] Mukherjee A, Sagar S, Parveen S, Das B C 2021 Appl. Phys. Lett. 119 253502
- [63] Liang F X, Wang I T, Hou T H 2021 Adv. Intell. Syst. 3 2100007
- [64] Zhang X, Wang W, Liu Q, Zhao X, Wei J, Cao R, Yao Z, Zhu X, Zhang F, Lv H, Long S, Liu M 2018 IEEE Electron Device Lett. 39 308
- [65] Duan Q, Jing Z, Zou X, Wang Y, Yang K, Zhang T, Wu S, Huang R, Yang Y 2020 Nat. Commun. 11 3399
- [66] Lu Y F, Li Y, Li H, Wan T Q, Huang X, He Y H, Miao X 2020 IEEE Electron Device Lett. 41 1245
- [67] Wang Z, Rao M, Han J W, Zhang J, Lin P, Li Y, Li C, Song W, Asapu S, Midya R, Zhuo Y, Jiang H, Yoon J H, Upadhyay N K, Joshi S, Hu M, Strachan J P, Barnell M, Wu Q, Wu H, Qiu Q, Williams R S, Xia Q, Yang J J 2018 *Nat. Commun.* 9 3208
- [68] Wang Y, Chen X, Shen D, Zhang M, Chen X, Chen X, Shao W, Gu H, Xu J, Hu E, Wang L, Xu R, Tong Y 2021 Nanomaterials 11 2860
- [69] Bousoulas P, Panagopoulou M, Boukos N, Tsoukalas D 2021 J. Phys. D:Appl. Phys. 54 225303
- [70] Han J K, Oh J, Yun G J, Yoo D, Kim M S, Yu J M, Choi S Y, Choi Y K 2021 *Sci. Adv.* 7 eabg8836
- [71] Wan T, Ma S, Liao F, Fan L, Chai Y 2022 Sci. China Inf. Sci. 65 141401
- [72] Wang T Y, Meng J L, Li Q X, He Z Y, Zhu H, Ji L, Sun Q Q, Chen L, Zhang D W 2021 *Nano Energy* 89 106291
- [73] Meng J, Wang T, Zhu H, Ji L, Bao W, Zhou P, Chen L, Sun Q Q, Zhang D W 2022 Nano Lett. 22 81

- Seo S, Jo S H, Kim S, Shim J, Oh S, Kim J H, Heo K, Choi J W, Choi C, Oh S, Kuzum D, Wong H P, Park J H 2018 Nat. Commun. 9 5106
- [75] Liao F, Zhou Z, Kim B J, Chen J, Wang J, Wan T, Zhou Y, Hoang A T, Wang C, Kang J, Ahn J H, Chai Y 2022 Nat. Electron. 5 84
- [76] Gao S, Liu G, Yang H, Hu C, Chen Q, Gong G, Xue W, Yi X, Shang J, Li R W 2019 ACS Nano 13 2634
- [77] Tan H, Tao Q, Pande I, Majumdar S, Liu F, Zhou Y, Persson P O A, Rosen J, van Dijken S 2020 Nat. Commun. 11 1369
- [78] Zhou F, Zhou Z, Chen J, Choy T H, Wang J, Zhang N, Lin Z, Yu S, Kang J, Wong H S P, Chai Y 2019 Nat. Nanotechnol. 14 776
- [79] Liu L, Cheng Z, Jiang B, Liu Y, Zhang Y, Yang F, Wang J, Yu X F, Chu P K, Ye C 2021 ACS Appl. Mater. Interfaces 13 30797
- [80] Zhou F, Chen J, Tao X, Wang X, Chai Y 2019 Research 2019 9490413
- [81] Xiang D, Liu T, Xu J, Tan J Y, Hu Z, Lei B, Zheng Y, Wu J, Neto A H C, Liu L, Chen W 2018 Nat. Commun. 9 2966
- [82] Zhang Z, Wang S, Liu C, Xie R, Hu W, Zhou P 2022 Nat. Nanotechnol. 17 27
- [83] Wang S, Chen C, Yu Z, He Y, Chen X, Wan Q, Shi Y, Zhang D W, Zhou H, Wang X, Zhou P 2019 Adv. Mater. 31 1806227
- [84] Zhu Q B, Li B, Yang D D, Liu C, Feng S, Chen M L, Sun Y, Tian Y N, Su X, Wang X M, Qiu S, Li Q W, Li X M, Zeng H B, Cheng H M, Sun D M 2021 Nat. Commun. 12 1798
- [85] Hu L, Yang J, Wang J, Cheng P, Chua L O, Zhuge F 2021 Adv. Funct. Mater. 31 2005582
- [86] Hou Y X, Li Y, Zhang Z C, Li J Q, Qi D H, Chen X D, Wang J J, Yao B W, Yu M X, Lu T B, Zhang J 2021 ACS Nano 15 1497
- [87] Yang L, Singh M, Shen S W, Chih K Y, Liu S W, Wu C I, Chu C W, Lin H W 2020 Adv. Funct. Mater. 31 2008259
- [88] Pei Y, Yan L, Wu Z, Lu J, Zhao J, Chen J, Liu Q, Yan X 2021 ACS Nano 15 17319
- [89] John R A, Acharya J, Zhu C, Surendran A, Bose S K, Chaturvedi A, Tiwari N, Gao Y, He Y, Zhang K K, Xu M, Leong W L, Liu Z, Basu A, Mathews N 2020 Nat. Commun. 11 3211
- [90] Egmont Petersen M, De Ridder D, Handels H 2002 Pattern Recognit. 35 2279
- [91] Rao D H, Panduranga P P 2006 2006 IEEE International Conference on Industrial Technology Mumbai, India, December 15–17, 2006 p2821
- [92] Chakraborty D, Raj S, Fernandes S L, Jha S K 2019 IEEE J. Emerging Sel. Top. Circuits Syst. 9 580
- [93] Pannu J S, Raj S, Fernandes S L, Chakraborty D, Rafiq S, Cady N, Jha S K 2020 IEEE Trans. Circuits Syst. II Express Briefs 67 961
- [94] Mannion D J, Mehonic A, Ng W H, Kenyon A J 2019 Front. Neurosci. 13 1386

- [95] Li C, Hu M, Li Y, Jiang H, Ge N, Montgomery E, Zhang J, Song W, Dávila N, Graves C E, Li Z, Strachan J P, Lin P, Wang Z, Barnell M, Wu Q, Williams R S, Yang J J, Xia Q 2017 Nat. Electron. 1 52
- [96] Lin P, Li C, Wang Z, Li Y, Jiang H, Song W, Rao M, Zhuo Y, Upadhyay N K, Barnell M, Wu Q, Yang J J, Xia Q 2020 Nat. Electron. 3 225
- [97] Pajouhi Z, Roy K 2018 IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. 37 1774
- [98] Yu Y, Deng Q, Ren L, Tashi N 2020 Neural Process. Lett. 51 1891
- [99] Bao L, Kang J, Fang Y, Yu Z, Wang Z, Yang Y, Cai Y, Huang R 2018 Sci. Rep. 8 13727
- [100] Hu W C, Yang C Y, Huang D Y 2011 J. Visual Commun. Image Represent. 22 543
- [101] Somasundaram G, Sivalingam R, Morellas V, Papanikolo poulos N 2013 IEEE Trans. Intell. Transp. Syst. 14 69
- [102] Huang K, Wang L, Tan T, Maybank S 2008 Pattern Recognit. 41 432
- [103] Maan A K, Kumar D S, Sugathan S, James A P 2015 IEEE Trans. Very Large Scale Integr. VLSI Syst. 23 2337
- [104] Jayachandran D, Oberoi A, Sebastian A, Choudhury T H, Shankar B, Redwing J M, Das S 2020 Nat. Electron. 3 646
- [105] Wang Y, Gong Y, Huang S, Xing X, Lv Z, Wang J, Yang J Q, Zhang G, Zhou Y, Han S T 2021 Nat. Commun. 12 5979
- [106] Russo F 2002 IEEE Trans. Instrum. Meas. 51 824
- [107] Wang C Y, Liang S J, Wang S, Wang P, Li Z a, Wang Z, Gao A, Pan C, Liu C, Liu J, Yang H, Liu X, Song W, Wang C, Cheng B, Wang X, Chen K, Wang Z, Watanabe K, Taniguchi T, Yang J J, Miao F 2020 Sci. Adv. 6 eaba6173
- [108] Yang X, Xiong Z, Chen Y, Ren Y, Zhou L, Li H, Zhou Y, Pan F, Han S T 2020 Nano Energy 78 105246
- [109] Zhu R, Tang Z, Ye S, Huang Q, Guo L, Chang S 2021 IEEE Trans. Electron Devices 68 602
- [110] Tang Z, Zhu R, Hu R, Chen Y, Wu E Q, Wang H, He J, Huang Q, Chang S 2021 *IEEE Trans. Cognit. Dev. Syst.* 13 645
- [111] Xin M, Wang Y 2019 Eurasip J. Image Video Process. 2019 40
- [112] Yao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Yang J J, Qian H 2020 Nature 577 641
- [113] Bouvier M, Valentian A, Mesquida T, Rummens F, Reyboz M, Vianello E, Beigne E 2019 ACM J. Emerging Technol. Comput. Syst. 15 1
- [114] Boybat I, Le Gallo M, Nandakumar S R, Moraitis T, Parnell T, Tuma T, Rajendran B, Leblebici Y, Sebastian A, Eleftheriou E 2018 Nat. Commun. 9 2514
- [115] Wang Z, Joshi S, Savel'Ev S, Song W, Midya R, Li Y, Rao M, Yan P, Asapu S, Zhuo Y, Jiang H, Lin P, Li C, Yoon J H, Upadhyay N K, Zhang J, Hu M, Strachan J P, Barnell M, Wu Q, Wu H, Williams R S, Xia Q, Yang J J 2018 Nat. Electron. 1 137
- [116] Li X, Tang J, Zhang Q, Gao B, Yang J J, Song S, Wu W, Zhang W, Yao P, Deng N, Deng L, Xie Y, Qian H, Wu H 2020 Nat. Nanotechnol. 15 776

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Application of neuromorphic resistive random access memory in image processing^{*}

Jiang Bi-Yi¹⁾²⁾ Zhou Fei-Chi^{1)†} Chai Yang^{2)‡}

1) (School of Microelectronics, Southern University of Science and Technology, Shenzhen 518000, China)

2) (Department of Applied Physics, The Hong Kong Polytechnic University, Hong Kong 999077, China)

(Received 15 March 2022; revised manuscript received 5 April 2022)

Abstract

With the increasing demands for processing images and videos at edge terminals, complementary metal oxide semiconductor (CMOS) hardware systems based on conventional Von Neumann architectures are facing challenges in terms of energy consumption, speed, and footprint. Neuromorphic devices, including resistive random access memory with integrated storage-computation characteristic and optoelectronic resistive random access memory with highly integrated in-sensor computing characteristic, show great potential applications in image processing due to their high similarity to biological neural systems and advantages of high energy efficiency, high integration level, and wide bandwidth. These devices can be used not only to accelerate large numbers of computational tasks in conventional image processing algorithms. In this paper, we first introduce the state-of-the-art neuromorphic resistive random access memory and optoelectronic neuromorphic resistive random access memory, then review the hardware implementation of and challenges to image processing based on these devices, and finally provide perspectives of their future developments.

Keywords: neuromorphic resistive random access memory, image preprocessing, image recognition PACS: 85.35.-p, 95.75.Mn, 87.18.Sn, 84.30.-r DOI: 10.7498/aps.71.20220463

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 62104091, 62174074), the Natural Science Foundation of Guangdong Province, China (Grant No. 2022A1515011064), the Guangdong Youth Innovation Talent Fund, China (Grant No. 2021KQNCX077), and NSQKJJ Fund, China (Grant Nos. K21799131, K21799128).

[†] Corresponding author. E-mail: zhoufc@sustech.edu.cn

[‡] Corresponding author. E-mail: ychai@polyu.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

光电神经形态器件及其应用

沈柳枫 胡令祥 康逢文 叶羽敏 诸葛飞

Optoelectronic neuromorphic devices and their applications Shen Liu-Feng Hu Ling-Xiang Kang Feng-Wen Ye Yu-Min Zhuge Fei 引用信息 Citation: Acta Physica Sinica, 71, 148505 (2022) DOI: 10.7498/aps.71.20220111 在线阅读 View online: https://doi.org/10.7498/aps.71.20220111 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

基于忆容器件的神经形态计算研究进展

Research progress of neuromorphic computation based on memcapacitors 物理学报. 2021, 70(7): 078701 https://doi.org/10.7498/aps.70.20201632

氧化物基忆阻型神经突触器件

Oxide-based memristive neuromorphic synaptic devices 物理学报. 2019, 68(16): 168504 https://doi.org/10.7498/aps.68.20191262

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

N型局部有源忆阻器的神经形态行为

Neuromorphic behaviors of N-type locally-active memristor 物理学报. 2022, 71(5): 050502 https://doi.org/10.7498/aps.71.20212017

混合时钟驱动的自旋神经元器件激活特性和计算性能

Activation function and computing performance of spin neuron driven by magnetic field and strain 物理学报. 2021, 70(20): 207502 https://doi.org/10.7498/aps.70.20210611

专题: 面向类脑计算的物理电子学

光电神经形态器件及其应用*

沈柳枫1)2) 胡令祥2) 康逢文2) 叶羽敏1) 诸葛飞2)3)†

1) (宁波大学材料科学与化学工程学院, 宁波 315211)

2) (中国科学院宁波材料技术与工程研究所, 宁波 315201)

3) (中国科学院脑科学与智能技术卓越创新中心,上海 200031)

(2022年1月16日收到; 2022年2月22日收到修改稿)

传统冯·诺依曼计算机在并行性计算和自适应学习方面效率较低,无法满足当前飞速发展的信息技术对 高效、高速计算的迫切需求.受脑启发的神经形态计算具有高度并行性、超低功耗等优势,被认为是打破传 统计算机局限性,实现新一代人工智能的理想途径.神经形态器件是实施神经形态计算的硬件载体,是构建 神经形态芯片的关键.与此同时,人类视觉系统与光遗传学的发展为神经形态器件的研究提供了新的思路. 新兴的光电神经形态器件结合了光子学与电子学各自的优势,在神经形态计算领域展露出巨大潜力,受到了 国内外研究人员广泛关注.本文对光电神经形态器件及其应用的最新研究进行了总结.首先综述了人工光电 突触与人工光电神经元,内容包括器件结构、工作机制以及神经形态功能模拟等方面.然后,对光电神经形态 器件在人工视觉系统、人工感知系统、神经形态计算等领域中的潜在应用作了阐述.最后,总结了当前光电 神经形态器件所面临的挑战,并对其未来的发展方向进行了展望.

关键词:光电神经形态器件,光电突触,光电神经元,神经形态计算 PACS: 85.60.-q,72.40.+w,85.35.-p,87.18.Sn **DOI:** 10.7498/

DOI: 10.7498/aps.71.20220111

1 引 言

自从 1936 年图灵提出图灵机计算模型概念以 来^[1], 计算机经历了突飞猛进的发展. 然而, 现有计 算机主要是基于传统冯·诺依曼架构, 受制于处理 单元与内存单元的物理分离, 存在有限的数据传输 效率、高能耗等问题, 无法满足智能化时代迫切的 高算力需求^[2,3]. 相比之下, 人脑是一个具有超低功 耗且高效的生物计算系统. 凭借事件驱动的运行模 式和高度的并行性计算, 人脑在执行认知、学习与 决策等复杂任务方面远远优于现代计算机^[4-6]. 因 此, 受人脑启发的神经形态 (类脑) 计算被认为是 实现高效人工智能最有效的途径之一, 受到越来越 广泛的关注^[7,8].相比传统架构计算系统,神经形态 系统从物理层面模仿人脑神经结构与运行模式,具 有高度的并行性与高效的计算能力,无需额外的内 存单元用于存储和检索数据,以实现存算一体化.

神经形态系统主要是由人工突触与人工神经 元组成. 自从 20 世纪 80 年代加州理工学院 Mead^[9] 提出"神经形态"概念以来,神经形态硬件系统在 过去几十年得到了深入研究. 其中,基于互补金属氧 化物半导体 (complementary metal oxide semiconductor, CMOS) 技术构造的人工神经元与人工突触 电路取得了重要进展,诸多性能优异的神经形态芯 片相继被报道. 例如, IBM 在 2014 年推出的 True-North 芯片内置了 100 万个人工神经元与 2.56 亿 个人工突触,使用了 54 亿个晶体管,能够以较低

© 2022 中国物理学会 Chinese Physical Society

^{*} 国家自然科学基金 (批准号: U20A20209, 61874125)、中国科学院战略性先导专项 (批准号: XDB32050204)、浙江省自然科学基金 (批准号: LD19E020001, LQ22F040003) 和宁波市自然科学基金 (批准号: 2021J139) 资助的课题.

[†] 通信作者. E-mail: zhugefei@nimte.ac.cn

功耗实现高速运算^[10].2015年浙江大学推出了由 30万个晶体管构成的达尔文芯片,其可以模拟2048 个神经元与400万个神经突触^[11].然而,作为神经 形态计算的关键,芯片中单个神经元与突触功能的 实现往往依赖于由电容及多个晶体管器件组成的 电路模块,导致电路结构复杂化,增大了系统能耗, 极大限制了芯片的集成规模与密度.因此,基于传 统 CMOS 电路开发的神经形态芯片并不是未来实 现类脑计算最有效的方案.

近年来,随着新型存储技术的深入研究,多种 新型存储器件被相继开发,包括忆阻器、Flash晶 体管、相变存储器、磁存储器、铁电存储器等.这些 具备记忆存储功能器件的发展使得利用单个器件 实现神经元与突触功能成为可能(注:上述能够模 拟神经元或突触功能的器件被称为神经形态器 件[12]). 目前,神经形态器件大多采用电信号驱动[12-14]. 相比电信号,光具有超高速、宽带宽和低串扰等优 势. 光遗传学的研究表明利用光可以有效调控大脑 行为[15],也为构建光电神经形态系统奠定了生物 学基础.因此,光电神经形态器件在实现超低功 耗、超高速神经形态计算方面具有独特的优势,逐 渐成为该领域的研究热点,例如,光电神经形态器 件能耗可以低至飞焦量级[16],显著低于传统电子 学器件. 维也纳工业大学 Mueller 团队^[17] 报道了 一种基于二维材料的神经形态器件,实现了超快的 机器视觉技术,构建的人工神经网络能够以每秒 2000万个二进制文件的处理速度对图像进行分类 与编码.

人脑中有大约 10¹¹ 个神经元和 10¹⁵ 个突触, 神经元之间依靠突触进行连接和信号传递,以此形 成复杂庞大的神经网络^[4-6].其中,神经元是神经 系统传递信息的基本结构与功能单元,可以接收来 自前神经元释放的信号并在胞体中进行整合.当整 合的信号强度超过一定阈值,神经元的轴突末梢会 释放神经递质,向后神经元发送电化学信号,以实 现信息的处理与传递过程^[18].类似地,光电神经元 器件接收编码的光脉冲,进行光电转换,并对转换 后的电信号进行整合,当累积的电信号超过器件转 变阈值时,则会向下一个神经元发射信号,以此完 成信号传递.光电神经元器件通常需要同时具备光 信号探测与神经元累积-发射功能. 丰桥技术科学 大学 Yonezu 等^[19] 早在 1989 年报道了一种权重可 调的光电神经元电路,实现了神经元间的光学互 联. 此后, 诸多研究工作相继报道了各种光电耦合 的神经元电路^[20-22],推动了光电神经元的发展.不 同于传统的光电神经元电路,光电神经元器件能够 在单个器件中实现光电转换与神经元的功能,显著 降低了功耗与信号传输中延时的问题. 例如, 韩国 科学技术院 Choi 团队^[23] 报道了一种基于单个晶 体管的光电神经元. 另一方面, 突触是实现神经元 间或神经元与效应器间信号传递的基础. 生物突触 可以通过电化学的相互作用引起突触强度(权重) 的改变,称为突触可塑性.突触可塑性是大脑实现 信息编码、学习与记忆的基础^[24].因此,开发人工 突触器件,实现突触可塑性模拟对构建神经形态芯 片至关重要.目前人工突触器件主要是依靠纯电信 号驱动, 而光电突触器件主要是依赖于光信号或光 电组合信号实施突触功能模拟^[25].突触的权重可 以由器件电导或电阻来表示. Agnus 等^[26]在 2010年 开发了一种基于碳纳米管的光电晶体管,模拟了生 物突触功能,为人工光电突触器件的研究奠定了重 要基础.此后,基于忆阻器^[27-30]、薄膜晶体管^[31-34]、 相变存储器[5,35]和铁电存储器[36,37]等不同结构的 器件相继被用于光电突触的模拟.

值得一提的是,人类从外界获取的信息超过 70%来自视觉系统^[38],感光功能在生物感知系统 中具有举足轻重的地位.光电神经形态器件兼具感 光和信号处理功能,对于构建人工视觉系统具有广 阔的应用前景.因此,光电神经形态器件在过去几 年中受到了国内外研究人员越来越多的关注,并取 得了系列重要的研究进展.

本文对光电神经形态器件的研究进展进行了 梳理与总结.主要内容如图1所示,包括:1)生物 神经元与生物突触的工作机制;2)光电突触器件 研究进展;3)光电神经元器件研究进展;4)光电神 经形态器件在人工视觉系统、人工感知系统和神经 形态计算三方面的应用.基于这些内容,最后本文 总结了当前光电神经形态器件所面临的挑战,并展 望了其未来的前景和发展方向.

2 生物神经元与突触

如图 2(a) 所示, 生物神经元通常由树突、胞体、轴突和轴突末梢组成. 其中, 树突是生长于胞体上的树状突起, 负责接收其他神经元的刺激信号并传递至胞体; 胞体是神经元细胞的代谢中心, 同



图 1 光电神经形态器件的研究进展概述^[16,23,29,33,39-43]





图 2 (a) 生物神经元和突触结构示意图; (b) 生物神经元随不同神经递质信号产生的膜电位变化, 其中, 黑线、红线和蓝线分别 表示神经元动作电位、兴奋性突触后电位和抑制性突触后电位^[47]

Fig. 2. (a) Schematic illustration of the structure of biological neurons and synapses; (b) membrane potential of biological neurons with different neurotransmitter signals, where the black, red and blue curves denote the neuronal action potential, excitatory post-synaptic potential respectively^[47].

时可对传递来的刺激信号进行时空整合;而轴突与 轴突末梢是胞体中延展出来的管状纤维组织,负责 神经元信号的输出^[18].神经元先是通过树突接收 前神经元释放的信号,然后引起膜电位改变;当累 积的局部膜电位超过阈值时,神经元细胞先会产生 动作电位(脉冲),随后通过轴突将脉冲传递至下 一个神经元,最终完成神经元与神经元之间的信号 传递^[44].突触是不同神经元之间彼此连接的关键 部位,其结构如图 2(a) 中插图所示. 当突触前神经 元的动作电位到达轴突末梢时,位于突触前膜上由 电位控制的钙离子 (Ca²⁺) 通道则会打开,从而诱 导 Ca²⁺内流并导致前膜内 Ca²⁺浓度急剧增高,进 而使得含有神经递质的囊泡从轴突末梢脱离,最终 将神经递质释放至突触间隙^[45]. 根据所释放的神 经递质作用机制不同,突触可分为兴奋性突触和抑 制性突触. 对于前者,神经递质与突触后膜受体相 结合时,位于后膜上特定的离子通道打开,带正电的钠离子 (Na⁺) 内流透过性显著提高,由此降低突触后膜的膜电位,进而产生局部的去极化,最终引起兴奋性突触后电位.与上述情形不同的是,抑制性突触释放的神经递质可显著提高钾离子 (K⁺)的外流及氯离子 (Cl⁻)的内流,由此提升突触后膜的膜电位,从而产生超极化并引起抑制性突触后电位^[45,46].图 2(b)所示的是生物神经元随不同神经递质信号产生的膜电位变化^[47].细胞膜在去极化后,K⁺外流速率逐渐超过 Na⁺的内流速率,导致膜电位复极化,恢复至静息状态^[18].

突触可塑性是人脑学习与记忆的基础,对实现生 物神经系统的高级功能具有重要作用^[24].突触可塑 性可分为短时程可塑性 (short-term plasticity, STP) 与长时程可塑性 (long-term plasticity, LTP)^[14,48]. 其中,短时程可塑性表现为在受到刺激后突触权重 仅能保持数秒至数分钟的范围内,随后会逐渐恢复 至初始状态,这是生物神经系统处理相关时空信息 的关键功能^[49]. 突触权重变化包括暂时性增强与抑 制,称为短时程增强(short-term potentiation, STP) 与短时程抑制 (short-term depression, STD)^[50]. 兴 奋性突触后电流 (excitatory post-synaptic current, EPSC)、双脉冲易化 (paired-pulse facilitation, PPF) 与双脉冲抑制 (paired-pulse depression, PPD) 是 反映突触短时程可塑性的重要功能. PPF 指的是 对于2个连续的脉冲刺激,突触对第2个刺激比对 第1个刺激的反应增强的现象. PPD 则与之相反^[48]. 长时程可塑性表现为在受到刺激后突触权重能够 保持数小时至数天不等,其与生物学习与记忆功能 密切相关.长时程可塑性通常分为长时程增强 (longterm potentiation, LTP) 与长时程抑制 (long-term depression, LTD)^[51]. 基于长时程可塑性的脉冲时 间依赖可塑性 (spike-timing-dependent plasticity, STDP) 规则与脉冲频率依赖可塑性 (spike-ratedependent plasticity, SRDP) 规则在突触进行学 习行为时起着重要作用[46,52].

3 光电突触器件

模拟突触可塑性功能是实现神经形态计算的 关键,其中基于长时程可塑性 (LTP/LTD) 的突触 权重可逆调控对构建人工神经网络至关重要^[53,54]. 然而目前报道的突触器件主要是利用电信号驱动, 面临功耗高、稳定性差等问题,严重制约了神经形态计算的发展.基于光信号驱动的光电突触器件在功耗、速度与稳定性方面具有很大的优势,有望解决上述挑战.近年来,研究人员基于不同材料、器件结构以及调控模式研发出各种新型光电突触.本部分将以光电突触器件的结构进行划分,重点阐述器件各自优缺点,并对其工作机制进行归纳总结.

3.1 光电忆阻器

忆阻器是除了电容、电感和电阻之外的第4种 基本电路元件,由加利福尼亚大学 Chua 教授^[55] 于 1971年首次提出. 忆阻器电阻状态可以通过流 经介质层的电荷量来调控,并具有独特的非线性电 阻转变特性,被广泛用于生物突触功能模拟. 目前 报道的忆阻器通常是利用电信号调控器件电阻,相 比于经过几十年的发展并且已取得了巨大进步的 电控忆阻器,光电忆阻器的研究仍处于起步阶段. 基于器件电导调控模式的不同,光电忆阻器可进一 步划分为光电协同型和全光型.

3.1.1 光电协同型

3.1.1.1 光增强-电抑制

光电协同型忆阻器是指器件电导调控需通过 光信号与电信号的共同作用才能实现. 华中科技大 学 Guo 团队^[27] 在 2018 年成功获得一种基于氧化 物异质结的光电忆阻型突触器件.研究人员利用溅 射沉积的方法在 Al 衬底上生长 ZnO 薄膜, 由于 Al 会夺取 ZnO 中的氧, 两者界面处形成一层 AlO_w 进而获得了 ZnO_{1-r}/AlO₁异质结, 如图 3(a) 所示. 该器件表现出缓变忆阻开关特性和持续光电导 (persistent photoconductivity, PPC) 特性, 并在 光信号作用下模拟了多种突触的可塑性功能. PPC 现象是指器件在光信号移除之后出现的光电 导长时间保持的现象. 通过进一步研究和分析, 该 团队发现器件在 310 nm 紫外光照射下所产生的 PPC 效应源于光载流子在 ZnO_{1-x} /AlO_y界面处 内建电场作用下的累积和捕获. 当紫外光照射时, 光生电子被激发至导带,器件电导增加,光生空穴 在内建电场的作用下在 ZnO1-x /AlOy 界面处不断 累积并被 AlO_u 层所捕获. 当光照移除后, 被捕获 的光生空穴难以在短时间内释放,进而阻碍了与光 生电子的复合,使得器件电导可以长时间保持,从 而产生 PPC 效应. 上述过程非常类似于生物突触



图 3 基于忆阻器实现的光电协同型突触器件 (a) 基于 ITO/ZnO₁₋₄/AlO₄/Al 的忆阻器件结构示意图, 插图为器件横截面的透射电子显微镜 (TEM) 图像^[27]; (b) EPSC 随刺激脉冲发生的光增强与电抑制过程^[27]; (c) 基于 MAPbI₃ 的平面型忆阻器结构示意 图; (d) 光照抑制碘空位形成和加速碘空位湮灭的过程^[28]; (e) 基于 MAPbI₃ 的平面型忆阻器在黑暗与光照条件 (可见光, 1.29 μW/cm²)下的 LTP 与 LTD 行为^[28]; (f) 基于 MAPbI₃ 的垂直型忆阻器结构示意图^[20]; (g) 基于 MAPbI₃ 的垂直型忆阻器在光照下内部工作 机制示意图^[20]; (h) 基于 MAPbI₃ 的垂直型忆阻器在光照与黑暗条件下的电增强与电抑制过程^[20]; (i) 基于 InAs 量子点的光电忆阻器电导在电压辅助下的光增强与光抑制过程^[33]

Fig. 3. Optoelectronic cooperative synaptic devices based on memristor: (a) Structural illustration of memristive device based on ITO/ZnO_{1-x}/AlO_y/Al, and the corresponding transmission electron microscope (TEM) image ^[27]; (b) photonic potentiation and electrical depression of stimulated pulses-dependent EPSC^[27]; (c) structural illustration of planar memristor based on MAPbI₃^[28]; (d) schematic illustration for illustrating how the light inhibits the formation (upper) and accelerates the annihilation (down) of iodine-related vacancies^[28]; (e) LTP and LTD behaviors of planar memristor based on MAPbI₃ under dark condition and upon illumination with a visible light at a power output of 1.29 μ W/cm², respectively^[28]; (f) structural illustration of MAPbI₃-based vertical memristor ^[29]; (h) dependence of electrical potentiation and depression of MAPbI₃-based vertical memristor on electrical pulses under dark and light illumination ^[20]; (i) structural illustration of InAs quantum dots (QDs)-based optoelectronic memristor^[63]; (j) photonic potentiation and depression of the conductance of InAs QDs-based optoelectronic memristor with the assistance of voltage^[63].

的长时程可塑性,因此常被用于该功能模拟. 图 3(b) 所示的是突触的光增强和电抑制过程. 在紫外光脉 冲的照射下,器件电导随时间持续增强,对应突触 LTP 行为;在施加电脉冲情况下,器件电导随时间 持续减小,对应突触 LTD 行为. 除此之外,在 In₂O₃/ ZnO^[56], ZnO/WS₂^[57], TiN_xO_{2-x}/MoS₂^[58] 与 MA-PbI₃^[59] 忆阻器中,研究人员也相继利用器件的 PPC 效应模拟了突触长时程可塑性功能.

3.1.1.2 电增强-光抑制

密歇根大学 Lu 团队^[28] 在 2018 年基于 MAPbI₃ 材料制备了一种横向结构的光电忆阻器,并通过光 照实现了器件电导降低 (即光抑制),器件结构如 图 3(c) 所示.该团队认为,通过增加 MAPbI₃ 中碘 空位 (*V_I*/*V_I*[×],即带正电或中性的空位)的形成能, 光照可抑制电场诱导碘空位的产生,并促进碘空位 的自发湮灭,如图 3(d) 所示.因此,在光照作用下, 器件可以实现光抑制过程,模拟 LTD 行为,而 LTP 行为的模拟可通过电脉冲实现,如图 3(e) 所示.除 此之外,研究人员在基于 MoSe₂/Bi₂Se₃ 异质结^[60] 和 HfO₂^[61] 的光电忆阻器中也发现了类似的"电增 强-光抑制"协同调控器件电导的特性,并将其用 于突触可塑性模拟.

3.1.1.3 光辅助-电调控

同样用 MAPbI3 材料, 高丽大学 Wang 团队^[29] 在 2019 年制备了垂直结构光电忆阻器, 如图 3(f) 所示. 在该器件中, 光照对内部离子迁移的影响与 图 3(c) 中平面结构忆阻器相反. 这是因为在对顶 电极施加正偏压时,光照所产生的光生电场 (E_{nb}) 与外部电场 (E_{Ext}) 具有相同的方向. 如图 3(g) 所 示,光生电场可以加速碘空位迁移.图 3(h)表明该 器件在光照下施加 0.15 V 电压脉冲时,器件电导 连续增加. 然而, 在无光照下施加-0.28 V 电压脉 冲时,器件电导则会逐渐降低.由此可见,与无光 条件下电导增加和降低的过程相比, 施加光照可起 到辅助降低器件编程电压的作用,进而减少运行功 耗. 另外, 基于黑磷纳米片制备的透明光电忆阻器 可在光照下降低器件操作电压,有利于实现器件的 低功耗操作.同时光照还可以降低器件电导态,进 而增大器件开关比.由此可见,光照是改善忆阻器 性能的有效调控手段[62].

3.1.1.4 电辅助-光调控

维尔茨堡大学 Hartmann 团队^[63] 在 2016 年报 道了一种具有"电辅助-光调控"的光电忆阻器, 如 图 3(i) 所示. 该器件主要是基于 GaAs/AlGaAs 异 质结和 InAs 量子点 (QDs) 制备,其中 InAs 量子 点生长在 GaAs/AlGaAs 异质结固定位置的的孔 洞内. 该器件具有 4 个端口,与传统两端忆阻器具 有较大差异. 器件两侧增加了栅极,并将栅极与器 件漏极相连栅极和漏极用于施加电压,源极接地. InAs 量子点的引入,改变了异质结界面的能带结 构. GaAs/AlGaAs 异质结在光照下会产生电子-空 穴对,实现量子点在不用偏压作用下充电或放电, 从而器件在光照下表现出独特的光响应特性.例 如,在-1.8 V 的偏置电压辅助下,器件可以在红外 LED 的光脉冲下实现电导的竭小,如图 3(j) 所示.

3.1.2 全光型

对于理想的光电忆阻器,器件电导应该是全光 可逆调控,即电导的增加和降低均通过光信号实 现.然而,光电协同型忆阻器在光信号作用下只能 实现突触权重的单向调控,要实现权值的可逆调控 必须依赖额外的电信号刺激,导致操作复杂,功耗 升高.此外,电信号刺激会引起器件微结构改变和 焦耳热产生,进而引起器件稳定性恶化.然而,受 限于半导体材料自身所固有的光电效应,通常情况 下光信号只能增大器件的电导,想要实现器件电导 的全光可逆调控是国内外公认的挑战.

针对上述挑战,中国科学院宁波材料技术与工 程研究所 Zhuge 团队^[30]在 2020 年采用制备工 艺较为成熟的四元氧化物半导体材料铟镓锌氧 (InGaZnO, IGZO), 在国际上率先研发出全光控忆 阻器,并成功应用于突触功能模拟.该忆阻器仅仅 通过改变入射光信号的波长,就可实现电导态的可 逆调控,并且调节后的电导态具有非易失性. 通过 深入研究表明,全光控忆阻器电导可逆调控机制源 于光诱导双层氧化物界面势垒宽度的可逆变化. 当 施加短波光时,氧空位的电离作用占主导,导致界 面处电离氧空位浓度增大,界面势垒变窄,隧穿电 流增大,进而器件电导增大;此后施加长波光,氧 空位的中和作用大于电离,电离氧空位浓度减少, 界面势垒变宽, 隧穿电流减小, 进而器件电导降低. 该全光控忆阻器的工作模式如 4(a) 所示, 在蓝光 和近红外光脉冲下实现低电导态到高电导态 (光 SET 过程)和高电导态到低电导态的可逆转变 (光 RESET 过程). 图 4(b) 中顶部图所示的是器件



图 4 基于忆阻器实现的全光型突触器件 (a) 基于 IGZO 全光控忆阻器的工作模式^[30]; (b) 基于 IGZO 全光控忆阻器电导可逆 调控特性及循环稳定性^[30]; (c) 基于 IGZO 全光控忆阻器的电导态保持特性,分别通过光 SET 和光 RESET 获得^[30]; (d) 基于 Ag-TiO₂ 纳米复合材料的忆阻器在可见光刺激下产生的 LTP 行为^[64]; (e) 基于 Ag-TiO₂ 纳米复合材料的忆阻器在紫外光刺激下产生的 LTD 行为^[64]

Fig. 4. All-optical synaptic devices based on memristor: (a) Working mode of all-optically controlled memristor based on IGZO^[30]; (b) reversible regulation characteristics of conductance (upper) and cycle stability (down) ^[30]; (c) retention characteristics of memconductance states after optical SET (upper) and optical RESET (down) operations^[30]; visible light-induced LTP (d) and UV lightinduced LTD (e) of the Ag-TiO₂ nanocomposite-based memristor^[64].

电导在全光信号调控下的可逆转变过程,在波长 420 nm 的光脉冲刺激下实现电导增加,在波长 800 nm 的光脉冲刺激下实现电导减小.得益于纯 电子的忆阻转变机制,该器件表现出非常优越的稳 定性,如图 4(b)中底部图所示.此外,为了验证该 器件电导全光可逆调节的非易失性,分别测试了 SET 和 RESET 操作后电导状态的保持特性,如图 4(c) 所示.器件电导表现出先缓慢衰减然后相对稳定, 且在 10⁴ s 后各电导状态之间仍可清楚区分,表明 器件具有良好的非易失性.该全光控忆阻器工作所 需光信号的功率密度非常低 (~20 μW/cm²),工作 机制不涉及微结构改变和焦耳热的产生,为解决忆 阻器的稳定性难题提供了一条全新的途径.此外, 该器件采用的 IGZO 材料,具有优异的热化学稳定 性,而且为非晶结构,在制备过程中能够保证优越 的均一性,有利于器件后续的大规模集成应用.值 得注意的是,大数据时代传统人工视觉系统由于视 觉信息的探测、处理和存储分离,使其在实时处理 视觉信息方面面临着体积大、速度慢以及功耗高等 挑战,而全光控忆阻器具有独特的光电响应,有望 利用同一器件实现视觉信息的感/存/算一体化,可 用于构建新一代人工视觉系统.

东北师范大学 Liu 团队 [64] 在 2021 年基于 Ag-TiO₂纳米复合材料设计了一种基于表面等离激元效 应的光电忆阻器. 该器件在可见光和紫外光照射下 可实现全光信号调制的突触可塑性模拟. 如图 4(d) 和图 4(e) 所示,器件在可见光照射下电流增大,光 照撤除后, 电流开始快速衰减直至趋于稳定, 稳定 后的电流值大于初始电流值,器件表现出一定的 LTP 特性;器件在紫外光照射下电流同样先增大, 然而当光撤除后, 电流会快速衰减到小于初始电流 的值,器件表现出一定的 LTD 特性. 该器件的工 作机制主要是源于可见光诱导的 Ag 纳米颗粒表 面等离激元效应与紫外光诱导的 TiO2 光电导效 应, 两种效应的共同作用会导致 Ag 纳米颗粒发生 氧化/还原,从而引起突触权重的变化.此外,基于 ZnO/PbS QDs 异质结的光电忆阻器, 可在紫外光 照射下模拟突触 LTP 行为, 在红外光照射下模拟 突触的 STP 特性^[65].

3.2 光电晶体管

晶体管的研究最早可追溯到 20 世纪 40 年代. 美国贝尔实验室发明了第一个实用的晶体管,为微 电子领域的发展奠定了基石. 1996 年, Mead 教授^[66] 首次演示了晶体管在神经形态计算方面的应用. 神 经形态晶体管在模拟生物突触功能时表现出许多 优势,例如:可同步进行信息读取与学习,突触权 重可进行高精度调控结构设计上具有更高的灵活 性. 此外,多栅极的突触晶体管在结构上与生物树 突更为相近,有望实现生物突触的高级功能^[67]. 光 电突触晶体管在传统电控突触晶体管的基础上引 入了光信号, 拓宽了器件的应用领域. 根据器件电 导调控模式的不同, 光电突触晶体管可分为光电协 同型突触晶体管和全光型突触晶体管.

3.2.1 光电协同型

深圳大学 Han 团队^[31]在 2019 年报道了一种 基于硒化物纳米片的光电晶体管,其结构如图 5(a) 所示. MoSe₂/Bi₂Se₃/PMMA 复合膜用作光活性电 荷俘获层, p型并五苯作为晶体管沟道材料.器件 可以在波长 790 nm 的近红外光照射下表现出明 显的光电流响应. 通过 KPFM 与原位 EFM 的测 试,验证了光诱导电子俘获机制.器件在近红外光 照射下可以提高 MoSe₂/Bi₂Se₃/PMMA 俘获层的 电子俘获能力,因而表现出优异的电导保持特性. 图 5(b) 所示的是器件在不同的近红外光强度照射 下所表现出的长/短时程可塑性功能.其中,在波 长 790 nm 的近红外光脉冲下,该器件表现出 LTP 行为, 而在 Vss 施加电脉冲时, 其能够实现 LTD 行 为, 如图 5(c) 所示. 此外, 该团队在基于硫化钼的 浮栅晶体管中也观察到近红外光调控的突触可塑 性行为[68].

除栅极电压外,漏源极电压 (V_{ds}) 也可用于调 控晶体管电导. 中弗罗里达大学 Thomas 团队^[32] 在 2020 年报道了基于石墨烯 (Gr) 和钙钛矿量子点 (PQDs)的光电晶体管,利用光照和漏源电压实现 了器件电导的可逆调控. 图 5(d) 是该器件在蓝光 照射和黑暗条件下的输出特性曲线,器件在光照下 的电流明显高于黑暗环境.图 5(e) 为器件能带示 意图, Gr与PQDs存在界面势垒,产生由PQD指 向 Gr 的内建电场. PQDs 在光照下吸收光子能量 产生电子-空穴对,电子-空穴对在内建电场作用下分 离. 其中, 光生空穴被转移到 Gr 中, 电子被输运 至 PQDs 内部. 随着电子在 PQDs 不断累积, 引起 光栅效应, 使 Gr 中产生更多空穴, 导致器件电流 增加. 此光电晶体管表现出了优异的电荷传输性 能、高光响应度,以及良好的电导保持特性.图 5(f) 展示了器件电导在光脉冲刺激下的连续增加,模拟 突触 LTP 行为, 在源极电脉冲作用下, 电导逐渐 降低,模拟突触 LTD 行为.

除此之外,研究人员基于酞菁铜/对六苯基^[69]、 氧化石墨烯^[70]、硼掺杂的硅纳米晶体^[71]、铟锌氧化 物^[72]等材料设计了具有不同光电特性的光电晶体 管,并且通过光照与电压协同作用成功实现了器件 电导的可逆调控.



图 5 光电协同型突触晶体管 (a) 基于 MoSe₂/Bi₂Se₃ 光电晶体管的器件结构^[31]; (b) 在 0.15 mW/cm² (i) 和 1.65 mW/cm² (ii) 功率密度的光脉冲刺激下突触后电流的变化^[31]; (c) 基于 MoSe₂/Bi₂Se₃ 光电晶体管电导的光增强与电抑制过程^[31]; (d) 基于 Gr-PQDs 的光电晶体管在黑暗与光照 (440 nm) 条件下的输出特性曲线, 插图为光电晶体管的示意图^[32]; (e) 基于 Gr-PQDs 光电晶体管在光激发 (i) 与光栅效应 (ii) 下的能级图, 其中 VB 和 CB 分别表示价带与导带^[32]; (f) 基于 Gr-PQDs 光电晶体管在光电协 同作用下的 LTP 与 LTD 行为^[32]

Fig. 5. Optoelectronic cooperative synaptic transistors: (a) Schematic illustration of the structure of $MoSe_2/Bi_2Se_3$ -based phototransistor^[31]; (b) dependence of the change of post-synaptic current on the time after continuously stimulating with the photonic pulses at the light intensity of 0.15 mW/cm² (i) and 1.65 mW/cm² (ii) ^[31]; (c) photonic potentiation and electrical depression of the conductance of $MoSe_2/Bi_2Se_3$ -based phototransistor^[31]; (d) output characteristic curve of the Gr-PQDs-based phototransistor under dark condition and 440 nm light illustration, where the phototransistor structure, as seen in the inset, is also given here^[32]; (e) schematic illustration of the energy band diagram for Gr-PQDs-based phototransistor under consideration of photoexcitation (i) and photogating effect (ii), where the VB and CB denote valence band and conduction band, respectively^[32]; (f) LTP and LTD behaviors of Gr-PQDs-based phototransistor under optoelectronic cooperation^[32].

3.2.2 全光型

墨尔本皇家理工大学 Walia 团队^[33]在 2021年 报道了一种基于黑磷 (BP)的光电晶体管,实现了 全光信号控制下的突触可塑性功能,器件结构如 图 6(a)所示.该器件可在波长 280 nm 深紫外光照 射下实现电导的增大,在波长 365 nm 的紫外光照 射下实现电导的减小,如图 6(b)和图 6(c)所示.BP 在机械剥离过程中,不可避免会吸附空气中的氧气 和水蒸气,导致 BP 表面部分氧化,形成光电子的 捕获中心,导致器件在紫外光照射下电导降低;然 而,由于波长 280 nm 深紫外光具有更高能量,器 件在该波段照射下产生大量光生载流子,对光电流 的影响占据主导作用,从而使器件电导增加.基于上 述特性,Walia及其合作者^[33]利用波长 280 nm 和 365 nm 的紫外光脉冲模拟了突触的 LTP 和 LTD 行为,如图 6(d)所示.此外,研究人员基于该器件实 现了视觉记忆、图像处理和机器学习等功能.同年, 天津工业大学 Zhang 团队^[34] 报道了一种基于芘基 石墨炔/石墨烯/硫化铅量子点 (Pyr-GDY/Gr/ PbS-QD) 的光电突触器件, 可以在可见光和近红 外作用下实现器件电导的可逆调节,如图 6(e) 所 示. 器件中 Pyr-GDY 和 PbS-QD 分别作为吸光 层和电荷捕获层, 单层 Gr 作为沟道层. 由于 Gr 与 Pyr-GDY, 以及 PbS-QDs 与 Gr 间的功函数不 同,3种材料界面处能带发生弯曲.器件在蓝光照 射下,大部分入射光被顶层的 Pvr-GDY 吸收,产 生大量电子-空穴对,光生电子在内建电场驱动下 转移到 Gr, 空穴被 Pyr-GDY 捕获. 虽然部分入射 光会被底层的 PbS-QDs 吸收, 并向 Gr 沟道中注 入空穴,同时留下的电子被 PbS-QDs 捕获,然而 Pyr-GDY 捕获的空穴数量远远大于 PbS-QDs 捕 获的电子数量.因此, Pyr-GDY 引起的正光栅效应



图 6 全光型突触晶体管 (a) BP 基光电晶体管结构示意图^[33]; (b), (c) BP 基光电晶体管在 280 nm 与 365 nm 波长光脉冲刺激 下的光电响应^[33]; (d) BP 基光电晶体管 LTP 与 LTD 突触行为模拟^[33]; (e) Pyr-GDY/Gr/PbS-QD 基光电晶体管结构示意图^[34]; (f) Pyr-GDY/Gr/PbS-QD 基光电晶体管在 450 nm 与 980 nm 波长光照射下的能带图^[34]; (g) Pyr-GDY/Gr/PbS-QD 基光电晶体 管 LTP 与 LTD 突触行为模拟^[34]

Fig. 6. All-optically controlled synaptic transistors: (a) Schematic illustration of the structure of fully light-controlled optoelectronic transistor based on BP^[33]; (b), (c) optoelectronic response of BP-based optoelectronic transistor upon stimulation with 280 nm (b) and 365 nm (c) light pulses^[33]; (d) LTP and LTD behaviors of BP-based optoelectronic transistor upon stimulation with 280 nm and 365 nm light pulses^[33]; (e) schematic illustration of the structure of Pyr-GDY/Gr/PbS-QDs-based optoelectronic transistor^[34]; (f) mechanistic illustration for the bandgap change of Pyr-GDY/Gr/PbS-QD-based optoelectronic transistor upon illumination with the light wavelengths of 450 nm (left) and 980 nm (right)^[34]; (g) LTP and LTD behaviors of the Pyr-GDY/Gr/PbS-QD-based optoelectronic transistor^[34].

比 PbS-QDs 引起的负光栅效应强得多,导致器件 电导降低.相反,器件在近红外光照射下,底层的 PbS-QDs 吸收率更高,大量的光生空穴注入到 Gr 沟道,从而使电导增大.在此基础上,该器件可在近 红外光脉冲照射下模拟突触 LTP 行为,在可见光 脉冲照射下,模拟突触 LTD 行为,如图 6(g) 所示.

3.3 其他光电器件

长庚大学 Lai 团队^[73] 在 2020 年报道了一种 基于二维 Bi₂O₂Se/和 Gr 的全光调节突触器件. 该 器件是由能够产生正向光电导的 Bi₂O₂Se 基两端 器件与能够产生负向光电导的 Gr 基两端器件串联 而成, 如图 7(a) 所示. 器件在 635 nm 波长的红光 照射下, Bi₂O₂Se 的正向光电导效应占主导地位, 费米能级升高, 电导增大. 器件在 365 nm 波长的 紫外光照射下发生两种效应, 即热辐射效应和等离 子体激元效应. 一方面, 费米能级会因热辐射效应 而降低,导致热载流子从 Bi₂O₂Se 注入到 Gr 中, 产生正向光电导效应;另一方面,Gr 表面吸附的氧 离子因等离子体激元效应形成氧气 ($O^2 \rightarrow e^- + O_2$), 导致空穴浓度降低,造成 Gr 的电导降低,器件产 生负向光电导效应.该器件可基于正负光电导效应 分别模拟生物突触的 EPSC 与抑制性突触后电流 (inhibitory post-synaptic current, IPSC),如图 7(c) 所示.此外,该器件在 635 nm 和 365 nm 波长的光 脉冲刺激下也可实现突触 LTP 与 LTD 行为模拟 (图 7(d)).

中国科学院北京纳米能源与系统研究所 Wang 团队^[39]在 2021 年报道了一种机械力与光电协同 作用的人工突触器件,如图 7(e) 所示.该器件是基 于 Gr/MoS₂ 异质结的光电晶体管与摩擦纳米发电 机 (triboelectric nanogenerator, TENG) 集成得到, 通过两种器件之间的协同合作, 实现了突触行为的 模拟.将 TENG 连接到晶体管的栅极, TENG 移动

产生摩擦电势可作用于突触晶体管,如图 7(f)所示. 当摩擦层分离时,Gr 费米能级 (*E*_F)降低,MoS₂在光 照下产生的光生电子会注入到 Gr 中, Gr 电阻升高, 导致器件电导降低. 当摩擦层靠近时, 会抑制 MoS₂



图 7 (a) Bi₂O₂Se/Gr 基突触结构图^[72]; (b) Bi₂O₂Se/Gr 基突触在红光 (i) 和紫外光 (ii) 照射下的能带图^[72]; (c) Bi₂O₂Se/Gr 基突 触在红光和紫外光照射下的突触后电流^[72]; (d) Bi₂O₂Se/Gr 基突触在同样的红光和紫光光脉冲下实现突触 LTP 与 LTD 行为模拟^[72]; (e) Gr/MoS₂基突触结构图,其中插图为光电晶体管的扫描电子显微镜 (SEM) 图像^[30]; (f) Gr/MoS₂基突触与 TENG 分离状态 (i) 与接触状态 (ii) 的工作原理及相应的能带图^[30]; (g) V_{TENG} 随位移变化曲线,其中插图为 V_{TENG} 输出的等效电路图^[30]; (h) Gr/MoS₂ 基的突触在光脉冲与 TENG 位移脉冲共同作用下实现的电流增加与降低过程^[30]

Fig. 7. (a) Structural illustration of synaptic device based on Bi₂O₂Se/Gr heterojunction^[72]; (b) mechanism illustration for the bandgap change of this Bi₂O₂Se/Gr-based synaptic device upon illumination with red (i) and UV light (ii), along with the corresponding post-synaptic current (c) as well as LTP and LTD behaviors (d) stimulated by the same red and UV light ^[72]; (e) schematic illustration of the structure of artificial synapse based on Gr/MoS₂ heterostructure and the scanning electron microscope (SEM) image of a phototransistor (inset) ^[30]; (f) working mechanistic principle and the corresponding bandgap illustration for this artificial synapse based on Gr/MoS₂ heterostructure at (i) separation state and (ii) contact state with TENG^[30]; (g) dependence of the variation of V_{TENG} value on the displacement, together with the equivalent circuit illustration for V_{TENG} output (inset) ^[30]; (h) current depression and potentiation of the artificial synapse based on Gr/MoS₂ heterojunction^[30]. 中光生电子注入 Gr, 从而有效地控制 Gr 与 MoS₂ 层之间的电荷转移与交换, 实现光电突触行为的调 控.图7(g) 给出了 TENG 的输出电压 (*V*_{TENG}) 与 摩擦层位移 (*D*) 关系, 以及 *V*_{TENG} 输出的等效电 路图 (见插图). TENG 每移动-0.1 mm 约相当于 5 V 的 *V*_{TENG} 输出,器件能够产生的等效栅压范 围在-52 V 至 60 V 间.当摩擦层距离固在 1.5 mm 时,器件电导在光脉冲刺激下逐渐减小,模拟了突 触 LTD 行为;在光脉冲移除后,通过控制 TENG 移动,器件电导会随着位移的连续增加而增加,模 拟了突触 LTP 行为.

深圳大学 Han 团队^[74] 在 2020 年报道了一种 基于光伏器件和忆阻器组成的自供电人工视觉感 知系统.光伏器件用于感受外界光信号,将光信号 转变为电信号,然后电信号作用于传统电控忆阻 器,引起器件内部导电细丝的形成和断裂,以此调 控突触权重变化.这种自供电系统避免了外部驱动 电源,减少了额外的能耗.通过将光电突触器件与 多种不同功能的器件集成而获得的新型系统显著 拓展了人工突触器件的应用领域与发展方向.

4 光电神经元

神经形态芯片的构建,除了需要突触器件以外,人工神经元器件也非常重要.近年来,研究人员在神经元器件研究方面取得了不少成绩,提出了多种神经元模型^[75,76].通常情况下,传统计算机模拟神经元行为需求解大量微分方程,会占用大量的处理与内存单元,而且耗时耗能.相比之下,基于

硬件实现人工神经元能大幅提升运算效率. 大规模 电控神经形态系统在计算过程中面临巨量的信号 传输,面临传输带宽、散热、功耗等问题. 光信号具 有速度快、超高带宽以及低串扰等优点,并拥有更 多的自由度 (含频率、偏振、模式指数、强度、统 计、相干性等)^[77]. 因此,光电神经元在神经网络信 号的传输与处理方面具有巨大的潜力,有望在高 效、低功耗的神经形态系统中发挥重要作用. 相比 光电突触器件,光电神经元的研究相对较少,我们 根据神经元间信息传输的信号,将神经元分为两 类: 1) 光输入-光输出型神经元; 2) 光输入-电输出 型神经元,并对这两种类型的神经元的器件结构、 工作机制以及功能模拟进行阐述.

4.1 光输入-光输出型

加州理工学院 Psaltis 团队^[40]在 1991 年提出 了基于砷化镓 (GaAs) 材料制备的光电神经元电 路,其结构如图 8 所示.每个神经元单元由发光二 极管 (LED)、驱动 LED 的场效应晶体管、负载晶 体管以及光电晶体管组成.用于驱动 LED 的场效 应晶体管是由光电晶体管与负载晶体管所组成的 开关电路控制.当光电晶体管检测到一定强度的输 入光时,施加在驱动晶体管上的栅极电压会相应增 大,从而促使晶体管开启并驱动 LED 产生输出光. 输出光功率随输入光功率的增加呈现非线性增加. 该神经元电路具有优异的光探测灵敏度,不仅能够 在较弱的输入光作用下产生高增益的光输出,而且 可以通过调节负载晶体管的栅极电压对光输出进 行调控.这种非线性输出特性与神经网络算法中非





Fig. 8. Schematic illustration of the structure of photoelectric neuron based on $GaAs^{[40]}$.

线性函数极为相似. 该团队在上述研究的基础上, 通过优化神经元电路的结构,实现了更为复杂的神 经元非线性输出函数.

美国国家标准与技术研究所 Shainline 等^[22] 也提出了一种由半导体发光二极管与超导纳米线 单光子探测器结合的光电神经元电路. 该硬件系统 能够通过对输入光的调控,实现非线性光的输出, 有望在人工神经网络和大规模神经形态计算中起 到重要作用.

4.2 光输入-电输出型

韩国科学技术院 Choi 团队^[23]在 2020 年报道 了基于单个硅基晶体管的光电神经元,该硅基晶体 管是由浮体、漏源极和栅极组成的垂直柱状结构, 如图 9(a) 所示. 神经元工作原理是基于单晶体管 锁存效应, 如图 9(b) 所示. 由于漏极寄生电容的存 在,当输入电流 (Iin) 施加到漏极时,正电荷会不断 地聚集在电容器中,导致输出电压 (Vout) 随电荷 增加而增加,呈现出线性累积过程.然而,一旦 Vout 超过锁存电压 (V_{latch}), 原本存储于电容器中的电 荷就会突然从漏极逸出并注入到浮体中,实现神经 元发射过程,最终恢复到初始静息状态.与此同时, 光照产生的电子-空穴对在这个过程中会在浮体中 不断积累,进而降低 Vlatch,使神经元脉冲幅值与 频率发生改变. 图 9(c) 所示的是神经元在不同光 功率下所输出脉冲随时间的变化关系.虽然该神经 元的输出脉冲可以被不同波长的红绿蓝光调控,但 其不能被光子能量小于硅带隙的红外光调控.此 外,在光源固定的情况下,栅压也可用于调控神经 元的输出. 基于这种光调控的神经元输出特性, 研 究人员构建了一个单层感知器, 演示了其在图像模 式识别上的应用.

在生物感知神经系统中,伤害感受器是必不可 少的关键组成部分,它能够在感测到伤害刺激时向 中枢神经系统发出快速警示.在模拟生物感觉神经 系统方面的研究中,仁川国立大学 Kim 团队^[78]在 2019 年报道了一种全氧化物基的透明光子伤害感 受器.该器件在紫外光脉冲作用前期不会出现明显 的光电流响应,然而随着紫外光脉冲刺激累积到一 定时间,器件光电流急剧变大并最终达到饱和.这 种光电流阈值的转变特性与人眼的感觉神经元功 能非常相似.

西安电子科技大学 Wang 团队与中国科学院

微电子研究所 Li 团队^[41]在 2020 年合作设计了一种由 IGZO₄ 紫外传感器和 NbO_x 振荡神经元串联 而成的人工视觉神经元,器件结构如图 9(d) 所示. 该器件不仅可以感知紫外光信息而且可将这些光 信息编码为电脉冲. 由于 NbO_x器件具有易失性的 阈值转变特性,基于其所设计的 Pearson-Anson 电路可进行模拟人工神经元的振荡行为. IGZO₄器 件具有良好的紫外光响应,电阻值随着紫外光波长 的减小而减小. 图 9(e) 所示的是人工视觉神经元 的电路结构. 该器件在不同波长的紫外光刺激下会显示出 4 种稳定的尖峰频率,如图 9(f) 所示. 基于 该人工神经元,研究人员构建了一种脉冲神经网络,根据振荡频率的不同,可对复杂背景图像进行分 割,并演示了人工视觉系统信息编码功能.

5 光电神经形态器件的应用

5.1 人工视觉系统

光电神经形态器件可以直接感受外界光信号, 并在光电转换过程中完成信息处理.这种集信息感 知与处理一体化的工作模式与人类视觉系统极为 相似.因此,目前有关光电神经形态器件应用的主 要集中于人工视觉系统,用于视觉信息的感知与存 储、预处理以及后处理等功能.

中国科学院半导体研究所 Shen 团队 [79] 在 2018 年实现了一种视觉信息探测与存储一体的人工视 觉系统. 该系统由基于氧化铟的图像传感阵列和基 于氧化铝的忆阻器组成,如图 10(a) 所示.其中,图 像传感器负责视觉信息探测, 忆阻器负责图像记 忆. 中央大学 Park 团队^[80] 在 2019 年实现了一种基 于 IGZO 突触晶体管和光电传感器的光控突触电 路. 该突触电路可根据环境中光强度的变化调整负 载栅极电压,模拟人眼视网膜明适应与暗适应功 能,如图 10(b) 所示.此外,研究人员利用基于卤化 物钙钛矿材料的光电器件也实现了明暗适应功能 的模拟[81,82]. 人眼除了能够感知外界光信号, 还可 对不同颜色的光进行识别. 成均馆大学 Park 团队^[16] 在 2018 年设计了一种基于 h-BN/WSe2 异质结的 人工光电突触器件,利用该器件在不同波长光照射 下的光电响应差异,模拟了人眼的颜色识别功能, 如图 10(c) 所示. 除上述功能外, 神经形态器件还 可用于图像预处理功能. 传统图像传感器在探测外 界信息时会产生大量冗余数据,影响图像处理效



图 9 光电神经元器件 (a) 硅基光电神经元的 TEM 图^[23]; (b) 神经元器件在加光与撤光条件下的光电响应^[23]; (c) 神经元器件 在不同功率光照射下的光电响应^[23]; (d) 基于 IGZO₄ 紫外传感器和 NbO_x振荡器的人工视觉神经元结构示意图^[41]; (e) 人工视觉 神经元在不同光照下的工作模式示意图^[41]; (f) 人工视觉神经元在黑暗和不同波长紫外光照射件下的 4 种发射行为^[41]

Fig. 9. Optoelectronic neuron devices: (a) TEM image of silicon-based optoelectronic neuron^[23]; (b) optoelectronic response of neuron under light ON and light OFF^[23]; (c) optoelectronic response of neuron upon stimulation with different light intensity^[23]; (d) structural illustration for artificial visual neuron composed of IGZO₄-based UV sensor and NbO_x-based oscillator^[41]; (e) working mode of artificial visual neuron under different light illumination^[41]; (f) four different firing behaviors of artificial visual neuron in dark and upon stimulation with different wavelength UV light^[41].

率.图像预处理技术可以实现原始图像的初级处理,包括图像去噪、背景扣除、特征提取及边缘增强等.相比于未处理的原始图像,预处理技术能减少冗余数据,提高图像在人工神经网络中的识别精度和迭代收敛速度.香港理工大学 Chai 团队^[83]

在 2019 年报道了一种基于 Pd/MoO_x/ITO 的光 电阻变存储器 (optoelectronic resistive random access memory, ORRAM), 如图 10(d) 所示.器件 电导在强光照射下可表现出良好保持性, 在弱光照 射下, 器件电导保持性较差.基于上述原理, 研究



图 10 光电神经形态器件在人工视觉系统中的应用 (a)由 In₂O₃ 基图像传感器与 Al₂O₃ 基阻变存储器构建的人工视觉系统^[79]; (b)具有明适应与暗适应功能的人工视觉系统^[80]; (c)具有颜色识别功能的 h-BN/WSe₂基光电突触器件^[16]; (d) MoO_x 基 ORRAM 结构示意图,其中插图为器件横截面的 SEM 图^[83]; (e) 基于 ORRAM 阵列构建的人工视觉系统^[83]; (f) 基于二维 WSe₂ 的 光电二极管结构示意图^[17]; (g) 基于 WSe₂ 光电二极管实现的分类器与自编码器应用^[17]

Fig. 10. Optoelectronic neuromorphic devices for artificial vision system: (a) Artificial vision system integrated by image sensor based on In_2O_3 and resistive random access memory based on $Al_2O_3^{[79]}$; (b) artificial visual system having the functions of light and dark adaptation^[80]; (c) h-BN/WSe₂ heterojunction-based optoelectronic synaptic device with the function of color recognition^[16]; (d) schematic illustration for ORRAM structure based on Pd/MoO_x/ITO, in which the inset shows the SEM image of the cross section of the device^[83]; (e) artificial vision system constructed by ORRAM array^[83]; (f) schematic illustration of photodiode based on two-dimensional (2 D) WSe₂ materials^[17]; (g) applications of 2D WSe₂-based photodiode for classifier and autoencoder^[17].

人员利用 ORRAM 阵列实现了图像的去噪和增强 对比度等预处理功能,并通过人工神经网络验证了 图像预处理效果.图 10(e)展示了基于 ORRAM 阵 列的图像预处理过程和用于图像识别的人工神经 网络结构图.此外,研究人员在基于钙钛矿的光电 忆阻器与基于 IGZO 的光电晶体管中也实现了图像 对比度增强与去噪等预处理功能^[74,84]. Wu 等^[85] 基于二阶电导转变特性的忆阻器与光敏元件集成 制备了一个光学预处理单元,实现了昏暗环境下带 噪点二维码图像的预处理. Wang 等^[86] 基于二维材 料范德瓦耳斯异质结制备了具有光响应度可调特性的突触器件,基于卷积操作实现了图像预处理.

图像预处理后的图像还需经过后处理过程,以 实现图像的分类、识别等高级功能.目前图像后处 理主要是基于深度学习算法的人工神经网络实现. 光电神经形态器件的一个重要应用方向便是基于 硬件构建人工神经网络,以提高图像后处理效率. 光电突触器件是构建人工神经网络的基本单元,在 本文第3节中我们对光电突触器件进行了总结.目 前报道的光电突触器件主要是利用电导可逆调控 实现人工神经网络中权重的更新,以实现学习功 能. 例如, Pradhan 等^[32]利用光电协同型突触晶体 管构建了脉冲神经网络,并基于无监督式机器学习 实现了人脸识别功能. 此外, 哈佛大学 Ham 团队[87] 在 2020 年报道了一种基于二维 MoS2 的大规模光 电晶体管阵列,可用于图像预处理与后处理.基于 此阵列构建的卷积神经网络,实现了对 MNIST 手 写体数字库识别, 准确率达 94%. 同年, 维也纳工业 大学 Mueller 团队^[17]在 2020 年制备了基于二维 WSe2 的光电二极管阵列, 如图 10(f) 所示. 该光电 二极管通过控制栅压的极性和大小可获得连续可 调的光响应灵敏度,并以此作为突触权重,构建人 工神经网络,实现了视觉信息的实时探测与识别, 获得了超快的信息处理速度. 图 10(g) 展示了基于 该光电晶体管阵列的两种人工神经网络,即分类器 与自编码器. 基于该硬件人工神经网络在识别精度 方面与基于软件的网络处理效果接近.

5.2 人工感知系统

除视觉系统外,人体还存在诸如触觉、听觉以 及嗅觉等感知系统.近年来光电神经形态器件在其 他人工感知系统方面也受到广泛关注.阿尔托大学 Dijken 团队^[88]在2020年设计了一种由压力传感器、 模数转换器 (analog-to-digital converter, ADC)、 发光二极管 (light emitting diode, LED)与光电 突触器件构成的人工触觉系统,如图 11(a)所示. 其中,压力传感器可根据压力的变化产生不同幅值 的电压信号,电压信号通过模数转换器 (analogto-digital converter, ADC) 后作用于发光二极管, 以此将压力信号转换为光信号.光信号作用于光电 突触器件,用于调控突触权重.基于此人工触觉 系统,可以实现手写字母识别.在此基础上,该 团队还报道了一种多模态集成的人工感知系统. 该系统将多种传感器与光电突触集成,实现了视觉、触觉、听觉、嗅觉和味觉信号的多模态信息处理应用^[42].

除上述之外, 延世大学 Cho 团队^[89] 在 2021 年 设计了一种由光电二极管、双电层突触晶体管、人 工神经元电路和机器手臂所组成的人工感知系统, 如图 11(b) 所示. 该系统可以根据外界光信号的改 变自发控制机械手臂,完成指令动作.当该系统接 收到外界光刺激时,光电二极管会将光信号转变为 电信号,并作用于突触晶体管,完成信号处理与输 出,当人工神经元接收到此电信号就会进行决策, 并对机械臂发出指令,完成相应动作.此外,Karbalaei Akbari 等^[90] 通过将光电突触器件与驱动器相 结合,模拟光刺激诱导的动作行为. Zhu 等^[91]设计 了一种发光忆阻器,通过压力传感器发出的刺激信 号使忆阻器人工突触发出光信号,并作用在后端机 械臂,诱发产生动作行为. Wan 等 [92] 将光电探测 器与压力传感器两者的信号作为刺激源,并作用于 突触晶体管,模拟了生物中的多模态刺激行为.

5.3 神经形态计算

5.3.1 非联想学习

生物体可以通过不断调整个体行为来适应外 界变化的环境,这种凭借经验的调节行为被认为是 一种高级学习功能. 经验式调节行为可分为两种形 式,即非联想学习与联想学习.其中非联想学习是 生物中不可或缺的自我学习与适应性行为之一,不 需要在外界刺激与机体反应之间形成特定关联[93]. 对于某种重复性刺激,非联想学习会表现出两种不 同的反应——习惯化和敏化. 习惯化指生物体对反 复施加的无害刺激反应逐渐减弱的行为^[94].相反, 敏化指生物体对反复施加的有害刺激反应增强的 行为^[93]. 华中科技大学 Guo 团队^[95] 在 2018 年基于 MoS₂制备的光电忆阻器实现了非联想学习中的习 惯化与敏化行为模拟.如图 12(a) 所示,器件在-8 V 电脉冲刺激下电流不断减小,模拟生物体在无害刺 激下的习惯化行为;在紫外光照射下,器件电流逐 渐增大,模拟生物体敏化行为,生物体中的伤害感 受器可在受到强烈的有害刺激时,进入敏化状态,以 此来避免环境中的伤害. 仁川国立大学 Kumar 等[78] 在 2019 年开发了一种基于 ZnO 薄膜的光电器件. 该 器件可在有害光信号刺激下发生敏化行为. 当再次 受到伤害后,器件可对有害光信号刺激具有更高的 敏感性,呈现出超敏 (allodynia) 和过敏 (hyperalgesia) 行为,见图 12(b).此外,有研究人员在基于 ZnO 纳米线^[96]、CeO_{2-x}^[97]、TiO₂/Ga₂O₃^[98]、Azo-Au 纳米颗粒^[99]等材料制备的光电器件中也发现了类 似的习惯化或敏化等非联想学习行为.



图 11 光电神经形态器件在人工感知系统中的应用 (a)由压力传感器与光电突触组成的人工神经系统^[88]; (b)由光电二极管、 突触晶体管以及机械臂组成的控制系统^[89]

Fig. 11. Optoelectronic neuromorphic devices for artificial sensing system: (a) Artificial system composed of pressure sensors and optoelectronic synapses^[88]; (b) control system composed of photodiodes, synaptic transistors and robotic arms^[89].



图 12 光电神经形态器件在非联想学习中的应用 (a) MoS₂基光电忆阻器模拟习惯化与敏化行为^[95]; (b) ZnO 基光电器件模拟 敏化行为^[78]

Fig. 12. Nonassociative learning based on optoelectronic neuromorphic devices: (a) Simulation of habituation and sensitization behaviors using the MoS_2 -based optoelectronic memristor^[95]; (b) simulation of sensitization behavior using the ZnO-based optoelectronic device^[78].

5.3.2 联想学习

联想学习属于大脑中的一个高阶学习行为,大脑通过学习 2 个或 2 个以上事件,建立它们之间的关系,并产生与之相对应的反应.巴普洛夫条件反射是一种典型的联想学习行为^[100].近年来,研究人员发现光电神经形态器件可以很好地模拟这种经典联想学习行为.中国科学院深圳先进技术研究所 Wang 团队^[101]在 2021 年报道了一种基于二维过渡金属硫化物的光电忆阻器,利用光调控阻变特性成功模拟了巴普洛夫实验.如图 13(a)所示,使用脉宽为 50 ms 和间隔为 450 ms 的电脉冲刺激作为条件反射信号 (摇铃),波长为 400 nm 的持续光刺激作为非条件发射信号 (喂食).单独的电脉冲刺激产生的电流在刚开始阶段无法达到阈值,对应小狗不会产生分泌唾液的行为;同时施加电脉冲和

光照一段时间后,发现单独施加电脉冲刺激可产生 超过阈值的电流,对应小狗因发生条件反射而分泌 唾液;经过一段时间的无条件刺激之后,电脉冲无 法产生超过阈值的电流,对应小狗对铃声不会产生 条件反射,恢复至初始状态.除此之外,Ahmed等^[102] 基于 BP 的光电突触器件在全光信号控制下模拟 了巴普洛夫实验,如图 13(b)所示.研究人员分别 使用波长为 280 nm 和 660 nm 的光脉冲作为非条 件刺激和条件刺激.单独施加 660 nm 的光脉冲信 号作为摇铃刺激时,产生的光电流无法超过阈值, 对应小狗不会产生流涎反应;单独施加 280 nm 的 光脉冲信号作为食物刺激时,产生的光电流超过阈 值,对应小狗会产生流涎反应;同时施加摇铃刺激 与食物刺激,经过一段时间训练后,小狗对单独的 摇铃刺激也会产生条件反射.



图 13 光电神经形态器件在联想学习模拟中的应用 (a)光电协同刺激实现的巴普洛夫实验^[101]; (b) 全光刺激实现的巴普洛夫 实验^[102]

Fig. 13. Associative learning based on optoelectronic neuromorphic devices. Paplov's experiment realized by optoelectronic co-stimulation^[101] (a) and by all-optical stimulation^[102] (b).

5.3.3 STDP

STDP 作为赫布学习规则的一种, 表示的是突 触连接强度变化 ($\Delta \omega$) 与前后神经元发射脉冲时 间间隔 (Δt) 的关系^[103]. 大量的生物学实验证明了 STDP 与人脑学习规则相似^[104,105]. STDP 学习规 则的函数表达式为^[106]

$$\Delta \omega = \begin{cases} A_+ e^{-|\Delta t|\tau_+}, & \Delta t > 0, \\ A_- e^{-|\Delta t|\tau_-}, & \Delta t < 0, \end{cases}$$

式中 A_{\pm} 与 τ_{\pm} 分别代表的是函数缩放因子与时间 常数.根据STDP学习规则,如果前神经元脉冲出 现在后神经元脉冲的峰值之前, Δt 为正, $\Delta \omega$ 的极 性为正,突触权重表示为增强效果;如果前神经元 脉冲出现在后神经元脉冲的峰值之后, Δt 为负, $\Delta \omega$ 的极性为负,突触权重表示为抑制效果.

光电神经形态器件可用于 STDP 学习规则模 拟,并可以通过改变光脉冲信号的宽度、个数及光 功率密度实现 STDP 学习规则的有效调控.中兴 大学 Lin 团队^[43]在 2021 年设计了基于 ReSe₂/h-BN/Gr 异质结的光电晶体管,通过光信号与电信号 的协同控制模拟了 STDP 学习规则 (见图 14(a)). 此外,中国科学院宁波材料所 Zhuge 团队^[30] 制备 了基于 IGZO 的全光控忆阻器,仅通过调控光信号 波长和脉冲个数,即可实现 STDP 学习规则的模 拟, 如图 14(b) 中图 (i) 所示. 该忆阻器实现 STDP 模 拟的操作机制如图 14(b) 中图 (ii), (iii) 所示, 其中 蓝光脉冲作为突触前刺激, 近红外光脉冲作为突触 后刺激. 图 (ii) 在蓝光脉冲刺激后施加近红外光脉 冲的刺激, 即 $\Delta t > 0$, 相应的突触权重表现出增强 效果; 图 (iii) 在施加近红外光脉冲刺激后施加蓝 光脉冲刺激, 即 $\Delta t < 0$, 突触权重表现出抑制效果.

6 总结与展望

人工突触与神经元器件是构建神经形态芯片 的基本单元,为实现低功耗、高效的神经形态计算 提供了新的思路.然而,现有神经形态器件的研究 主要集中于纯电控突触和神经元,在功耗及功能应 用方面还同人脑具有很大差距,限制了神经形态芯 片的发展.新兴的光电神经形态器件可以兼具光子 学、电子学及生物学的研究特色,在降低器件功 耗、拓宽应用领域方面具有独特的优势,成为近年 来研究的热点.本文综述了近年来光电突触器件与 光电神经元器件取得的系列研究进展,从器件结 构、材料体系、工作模式、机理解释等方面进行了 重点阐述,并详细介绍了其在人工视觉系统、人工 感知系统与神经形态计算等领域的应用.光电神经 形态器件经过近几年的发展取得了诸多重要进展,



图 14 光电神经形态器件在 STDP 学习规则模拟中的应用 (a) 基于 ReSe₂/h-BN/Gr 光电晶体管实现的四种 STDP 学习规则^{[43}; (b) 基于全光控忆阻器实现的 STDP 学习规则^[30]

Fig. 14. STDP learning rules based on optoelectronic neuromorphic devices: (a) Four STDP learning rules based on ReSe2/h-BN/Gr phototransistors^[43]; (b) STDP learning rules based on all-optically controlled memristor^[30].

然而仍处于研究的初期阶段,在光敏材料合成、器件/ 阵列构筑以及功能应用等方面面临诸多关键科学 问题和技术挑战.

首先在材料制备方面.目前用于光电神经形态 器件的光敏材料主要分为氧化物半导体材料、钙钛 矿材料及新型二维材料.其中氧化物半导体材料具 有优异的热化学稳定性、良好的光电特性及较为成 熟的制备工艺,有利于器件的大规模集成.因此, 基于氧化物半导体材料的光电神经形态器件研究 最为广泛,取得了系列重要研究进展.然而,因受 其带隙限制,氧化物半导体往往只对紫外光具有高 效响应,而在可见及红外波段响应较差或者不响 应,限制了器件的应用场景.因此,开发具有宽光 谱高效响应的氧化物半导体材料,是未来该领域发 展的重要方向之一.除了传统氧化物半导体材料, 以钙钛矿和低维材料为代表的新型光敏材料因其 具有优异的光电特性,在光电神经形态器件上也得 到广泛关注.虽然新型光敏材料在一定程度上提高 了器件光响应度、拓宽了响应波长范围,然而由于 其制备工艺不成熟, 难以保证器件的稳定性和重复 性,给未来器件大规模集成应用带来很大挑战.

然后是器件/阵列构筑方面. 光电神经形态器 件经过近几年的发展,虽然在突触与神经元功能模 拟上取了许多重要成果,然而绝大部分研究还停留 在单一器件的改性以及简单类脑功能的模拟上,很 少涉及包含多个器件的硬件网络,这些工作离构建 最简单的神经形态芯片还有很远的距离. 限制器件 集成应用的因素主要有以下几个方面.1)器件稳 定性.光电神经形态器件的大规模集成应用,需要 器件与器件之间以及单个器件不同循环之间具有 优异的均一性和稳定性. 然而目前光电神经形态器 件的研究还处于初级阶段,器件工作机制不明确、 制备工艺不成熟,使得器件性能还无法满足大规模 集成的要求. 2) 器件能耗. 人类大脑功能强大, 功 耗只有 20 W 左右, 单个突触行为的能耗更是仅约 为 10 fJ, 然而目前开发的光电神经形态器件能耗 普遍在微焦量级,远远大于生物突触和神经元,高 能耗问题严重制约了器件的集成密度.3) 全光调 控. 对于理想的光电神经形态器件应该是可全光调 控的,即可以直接利用外界光信号驱动实施功能模 拟,以此简化操作方式、降低能耗.然而目前报道 的光电神经形态器件往往需要光信号和电信号的 共同作用,不仅增加了操作复杂性、带来了高能耗,

而且会产生大量焦耳热,影响器件稳定性.虽然近 期报道了有关全光控忆阻器、晶体管的研究工作, 但全光控技术依然存在诸多挑战, 需进一步深入研 究. 4) 光信号引入方案. 光电神经形态器件在工作 过程中往往需要实现各个器件的独立光信号控制, 例如基于光电突触器件的人工神经网络学习过程 中,需要根据目标不断利用光信号调控每个突触器 件的权值. 然而目前报道的实施方案, 主要是利用 LED、氙灯、激光器等光源非集成式地照射器件, 只能用于对单个器件以及多个器件组成的简单阵 列的控制,而无法用于大规模阵列.因此,如何实 现大规模阵列中每个器件的独立光信号控制具有 很大的挑战. 值得庆幸的是, 各种新型显示技术以 及光波导技术的快速发展,给这一问题提供了新的 思路. 实现光波导或 LED 与光电神经形态器件的 无缝集成是解决光信号引入的有效方案,值得进一 步研究.

最后是功能应用方面.光电神经形态器件虽然 在视觉模拟、人工感知系统以及神经形态计算等方 面得到初步应用,然而在应用功能上还很简单,主 要是基于突触和神经元的部分功能实现简单图像 的探测、预处理与记忆等,要实现较为复杂功能 (例如,图像识别)必须借助软件算法和传统电路. 器件稳定性以及器件之间的性能差异等所带来的 诸多非理想因素会影响了器件功能模拟的精度,也 成为限制器件复杂应用的重要原因.此外,对于理 想的光电神经形态器件不仅可以完成光信息的探 测,而且实时处理探测到的信号,并将处理后的结 果原位存储,这一特性对构建感/存/算一体化的新 型人工视觉系统具有重要意义.然而目前报道的光 电神经形态器件只能实现一些初级的感/存/算一 体化,有待进一步探索.

总之,在人类社会智能化技术飞速发展的今 天,光电神经形态器件的研究即面临重大发展机 遇,又存在严峻挑战.构建性能优异、功能多样以 及超低能耗的光电神经形态器件,实现器件在神经 形态芯片中的大规模集成应用需要国内外研究者 的共同努力.希望在不久的将来,这一目标可以实现.

参考文献

- [1] Turing A M 1937 Proc. London Math. Soc. 42 230
- [2] Von Neuman J 1993 IEEE Ann. Hist. Comput. 15 27
- [3] Zidan M A, Strachan J P, Lu W D 2018 $\it Nat. \ Electron. \ 1$ 22

- [4] Attwell D, Laughlin S B 2001 Cereb. Blood Flow Metab. 21 1133
- [5] Chen Z, Ríos C, Pernice W H P, Wrigh C D, Bhaskara H 2017 Sci. Adv. 3 e1700160
- [6] Drachman D A 2005 Neurology 64 2004
- [7] Indiveri G, Linares-Barranco B, Legenstein R, Deligeorgis G, Prodromakis T 2013 Nanotechnology 24 384010
- [8] Indiveri G, Liu S C 2015 *Proc. IEEE* 103 1379
- [9] Mead C 1990 Proc. IEEE 78 1629
- [10] Merolla P A, Arthur J V, Alvarez-Icaza R, et al. 2014 Science 345 668
- [11] Shen J C, Ma D, Gu Z H, Zhang M, Zhu X L, Xu X Q, Xu Q, Shen Y J, Pan G 2016 *Sci. China Inf. Sci.* **59** 023401
- [12] Jo S H, Chang T, Ebong I, Bhadviya B B, Mazumder P, Lu W 2010 Nano Lett. 10 1297
- [13] Prezioso M, Merrikh-Bayat F, Hoskins B D, Adam G C, Likharev K K, Strukov D B 2015 Nature 521 61
- [14] Ohno T, Hasegawa T, Tsuruoka T, Terabe K, Gimzewski J K, Aono M 2011 Nat. Mater. 10 591
- [15] Han W, Tellez L A, Rangel M Jr, Motta S C, Zhang X, Perez I O, Canteras N S, Shammah-Lagnado S J, van den Pol A N, de Araujo I E 2017 *Cell* 168 311
- [16] Seo S, Jo S H, Kim S, Shim J, Oh S, Kim J H, Heo K, Choi J W, Choi C, Oh S, Kuzum D, Wong H P, Park J H 2018 Nat. Commun. 9 5106
- [17] Mennel L, Symonowicz J, Wachter S, Polyushkin D K, Molina-Mendoza A J, Mueller T 2020 Nature 579 62
- [18] Han J S 2009 Neuroscience (Beijing: Peking University Medical Press) pp139–145 (in Chinese) [韩济生 2009 神经科 学 (北京大学医学出版社) 第139—145页]
- [19] Yonezu H, Miho A, Himeno T, Pak K, Takano Y 1989 *Electron. Lett.* 25 670
- [20] Krishnamoorthy A V, Yayla G, Esener S C 1992 IEEE T. Neural Networ. 3 404
- [21] Wen Z, Frahat N H, Lin S Y 1994 Opt. Lett. 19 1394
- [22] Shainline J M, Buckley S M, McCaughan A N, Chiles J, Jafari-Salim A, Mirin R P, Nam S W 2018 J. Appl. Phys. 124 152130
- [23] Han J K, Geum D M, Lee M W, Yu J M, Kim S K, Kim S, Choi Y K 2020 Nano Lett. 20 8781
- [24] Bliss T V P, Collingridge G L 1993 Nature 361 31
- [25] Zhuge X, Wang J R, Zhuge F 2019 Phys. Status Solidi-R 13 1900082
- [26] Agnus G, Zhao W, Derycke V, Filoramo A, Lhuillier Y, Lenfant S, Vuillaume D, Gamrat C, Bourgoin J P 2010 Adv. Mater. 22 702
- [27] Hu D C, Yang R, Jiang L, Guo X 2018 ACS Appl. Mater. Interfaces 10 6463
- [28] Zhu X, Lu W D 2018 ACS Nano 12 1242
- [29] Ham S, Choi S, Cho H, Na S I, Wang G 2019 Adv. Funct. Mater. 29 1806646
- [30] Hu L X, Yang J, Wang J R, Cheng P H, Chua L O, Zhuge F 2020 Adv. Funct. Mater. 31 2005582
- [31] Wang Y, Yang J, Ye W B, She D H, Chen J R, Lv Z Y, Roy V A L, Li H L, Zhou K, Yang Q, Zhou Y, Han S T 2019 Adv. Electron. Mater. 6 1900765
- [32] Pradhan B, Das S, Li J, Chowdhury F, Cherusseri J, Pandey D, Dev D, Krishnaprasad A, Barrios E, Towers A, Gesquiere A, Tetard L, Roy T, Thomas J 2020 Sci. Adv. 6 eaay5225
- [33] Ahmed T, Tahir M, Low M X, Ren Y, Tawfik S A, Mayes E L H, Kuriakose S, Nawaz S, Spencer M J S, Chen H, Bhaskaran M, Sriram S, Walia S 2021 Adv. Mater. 33 e2004207

- [34] Hou Y X, Li Y, Zhang Z C, Li J Q, Qi D H, Chen X D, Wang J J, Yao B W, Yu M X, Lu T B, Zhang J 2021 ACS Nano 15 1497
- [35] Feldmann J, Youngblood N, Wright C D, Bhaskaran H, Pernice W H P 2019 Nature 569 208
- [36] Luo Z D, Xia X, Yang M M, Wilson N R, Gruverman A, Alexe M 2020 ACS Nano 14 746
- [37] Guo F, Song M L, Wong M C, Ding R, Io W F, Pang S Y, Jie W J, Hao J H 2021 Adv. Funct. Mater. 32 2108014
- [38] Wang G Z, Wang R B, Kong W Z, Zhang J H 2018 Cogn. Neurodyn. 12 615
- [39] Yu J R, Yang X X, Gao G Y, Xiong Y, Wang Y F, Han J, Chen Y H, Zhang H, Sun Q J, Wang Z L 2021 Sci. Adv. 7 eabd9117
- [40] Psaltis D, Lin S 1991 Proc. SPIE 1562 204
- [41] Wu Q T, Dang B J, Lu C Y, Xu G W, Yang G H, Wang J W, Chuai X C, Lu N D, Geng D, Wang H, Li L 2020 Nano Lett. 20 8015
- [42] Tan H W, Zhou Y F, Tao Q Z, Rosen J, van Dijken S 2021 Nat. Commun. 12 1120
- [43] Tsai M Y, Lee K C, Lin C Y, Chang Y M, Watanabe K, Taniguchi T, Ho C H, Lien C H, Chiu P W, Lin Y F 2021 Adv. Funct. Mater. 31 2105345
- [44] Nicholls J G, Martin A R, Fuchs P A, Brown D A, Diamond M E, Weisblat D A (translated by Yang X L) 2014 From Neuron to Brain, Fifth Edition (Beijing: Science Press) p10 (in Chinese) [尼克尔斯 J G, 马丁 A R, 福克斯 P A, 布朗 D A, 戴蒙特 M E, 韦斯勃拉脱 D A 著 (杨雄里 译) 2014 从神经 元到脑: 第5版 (北京: 科学出版社) 第10页]
- [45] Kandel E R, Squire L R 2000 Science 290 1113
- [46] Bi G, Poo M 1998 J. Neurosic. 18 10464
- [47] Lee G, Baek J H, Ren F, Pearton S J, Lee G H, Kim J 2021 Small 17 e2100640
- [48] Fioravante D, Regehr W G 2011 Curr. Opin. Neurobiol. 21 269
- [49] Abbott L F, Regehr W G 2004 Nature **431** 796
- [50] Hennig M H 2013 Front Comput. Neurosci. 7 154
- [51] Montgomery J M, Madison D V 2004 Trends Neurosci. 27 744
- [52] Rachmuth G, Shouval H Z, Bear M F, Poon C S 2011 Proc. Natl. Acad. Sci. U. S. A. 108 E1266
- [53] Zhao Y H, Liu B, Yang J L, He J, Jiang J 2020 Chin. Phys. Lett. 37 088501
- [54] Ren K, Zhang K J, Qin X Z, Ren H X, Zhu S H, Yang F, Sun B, Zhao Y, Zhang Y 2021 Acta Phys. Sin. 70 078701 (in Chinese) [任宽, 张珂嘉, 秦溪子, 任焕鑫, 朱守辉, 杨峰, 孙柏, 赵勇, 张勇 2021 物理学报 70 078701]
- [55] Chua L 1971 IEEE Trans. Circuit Theory 18 507
- [56] Kumar M, Abbas S, Kim J 2018 ACS Appl. Mater. Interfaces 10 34370
- [57] Kumar M, Ban D K, Kim S M, Kim J, Wong C P 2019 Adv. Electron. Mater. 5 1900467
- [58] Wang W X, Gao S, Li Y, Yue W J, Kan H, Zhang C W, Lou Z, Wang L L, Shen G Z 2021 Adv. Funct. Mater. 31 2101201
- [59] Zhao X N, Wang Z Q, Li W T, Sun S W, Xu H Y, Zhou P, Xu J Q, Lin Y, Liu Y C 2020 Adv. Funct. Mater. 30 1910151
- [60] Wang Y, Yang J, Wang Z P, Chen J R, Yang Q, Lv Z Y, Zhou Y, Zhai Y B, Li Z X, Han S T 2019 Small 15 e1805431
- [61] Zhou Y, Yew K S, Ang D S, Kawashima T, Bera M K, Zhang H Z, Bersuker G 2015 Appl. Phys. Lett. 107 072107
- [62] Zhou Y, Liu D N, Wang J H, Cheng Z Q, Liu L, Yang N,

Liu Y X, Xia T, Liu X Y, Zhang X, Ye C, Xu Z, Xiong W, Chu P K, Yu X F 2020 ACS Appl. Mater. Interfaces 12 25108

- [63] Maier P, Hartmann F, Rebello Sousa Dias M, Emmerling M, Schneider C, Castelano L K, Kamp M, Marques G E, Lopez-Richard V, Worschech L, Höfling S 2016 Appl. Phys. Lett. 109 023501
- [64] ShanX Y, Zhao C Y, Wang X N, Wang Z Q, Fu S C, Lin Y, Zeng T, Zhao X N, Xu H Y, Zhang X T, Liu Y C 2021 Adv. Sci. 9 2104632
- [65] Li H L, Jiang X T, Ye W B, Zhang H, Zhou L, Zhang F, She D H, Zhou Y, Han S T 2019 Nano Energy 65 104000
- [66] Diorio C, Hasler P, Minch A, Mead C A 1996 IEEE Trans. Electron Devices 43 1972
- [67] Zhu L Q, Xiao H, Liu Y H, Wan C J, Shi Y, Wan Q 2015 Appl. Phys. Lett. 107 143502
- [68] Zhai Y B, Zhou Y, Yang X Q, Wang F, Ye W B, Zhu X J, She D H, Lu W D, Han S T 2020 Nano Energy 67 104262
- [69] Qian C, Oh S, Choi Y, Kim J H, Sun J, Huang H, Yang J, Gao Y, Park J H, Cho J H 2019 *Nano Energy* 66 104095
- [70] Sun J, Oh S, Choi Y, Seo S, Oh M J, Lee M, Lee W B, Yoo P J, Cho J H, Park J H 2018 Adv. Funct. Mater. 28 1804397
- [71] Yin L, Han C, Zhang Q T, Ni Z Y, Zhao S Y, Wang K, Li D S, Xu M S, Wu H Q, Pi X D, Yang D R 2019 Nano Energy 63 103859
- [72] Wang J X, Chen Y, Kong L A, Fu Y, Gao Y L, Sun J 2018 Appl. Phys. Lett. 113 151101
- [73] Yang C M, Chen T C, Verma D, Li L J, Liu B, Chang W H, Lai C S 2020 Adv. Funct. Mater. 30 2001598
- [74] Yang X Y, Xiong Z Y, Chen Y J, Ren Y, Zhou L, Li H L, Zhou Y, Pan F, Han S T 2020 Nano Energy 78 105246
- [75] Mcculloch W S, Pitts W 1990 Bull. Math. Biol. 52 99
- [76] Hodgkin A L, Huxley A F 1990 Bull. Math. Biol. 52 25
- [77] Shainline J M, Buckley S M, Mirin R P, Nam S W 2017 Phys. Rev. Applied 7 034013
- [78] Kumar M, Kim H S, Kim J 2019 Adv. Mater. 31 e1900021
- [79] Chen S, Lou Z, Chen D, Shen G Z 2018 Adv. Mater. 30 1705400
- [80] Kwon S M, Cho S W, Kim M, Heo J S, Kim Y H, Park S K 2019 Adv. Mater. 31 e1906433
- [81] Chen Q L, Zhang Y, Liu S Z, Han T T, Chen X H, Xu Y Q, Meng Z Q, Zhang G L, Zheng X J, Zhao J J, Cao G Z, Liu G 2020 Adv. Intell. Syst. 2 2000122
- [82] Hong S, Choi S H, Park J, Yoo H, Oh J Y, Hwang E, Yoon D H, Kim S 2020 ACS Nano 14 9796
- [83] Zhou F C, Zhou Z, Chen J W, Choy T H, Wang J L, Zhang N, Lin Z Y, Yu S M, Kang J F, Wong H P, Chai Y 2019 *Nat. Nanotechnol.* 14 776
- [84] Qiu W J, Huang Y L, Kong L A, Chen Y, Liu W R, Wang Z, Sun J, Wan Q, Cho J H, Yang J L, Gao Y L 2020 Adv. Funct. Mater. 30 2002325
- [85] Wu L D, Wang Z W, Wang B W, Chen Q Y, Bao L, Yu Z

Z, Yang Y F, Ling Y T, Qin Y B, Tang K C, Cai Y M, Huang R 2021 Nanoscale **13** 3483

- [86] Wang C Y, Liang S J, Wang S, Wang P F, Li Z A, Wang Z R, Gao A Y, Pan C, Liu C, Liu J, Yang H F, Liu X W, Song W H, Wang C, Cheng B, Wang X M, Chen K J, Wang Z L, Watanabe K, Taniguchi T, Yang J J, Miao F 2020 Sci. Adv. 6 eaba6173
- [87] Jang H, Liu C Y, Hinton H, Lee M H, Kim H, Seol M, Shin H J, Park S, Ham D 2020 Adv. Mater. 32 e2002431
- [88] Tan H W, Tao Q Z, Pande I, Majumdar S, Liu F, Zhou Y F, Persson P O A, Rosen J, van Dijken S 2020 Nat. Commun. 11 1369
- [89] Kim S, Roe D G, Choi Y Y, Woo H, Park J, Lee J I, Choi Y, Jo S B, Kang M S, Song Y J, Jeong S, Cho J H 2021 Sci. Adv. 7 eabe3996
- [90] Karbalaei Akbari M, Zhuiykov S 2019 Nat. Commun. 10 3873
- [91] Zhu Y B, Wu C X, Xu Z W, Liu Y, Hu H L, Guo T L, Kim T W, Chai Y, Li F S 2021 *Nano Lett.* 21 6087
- [92] Wan C J, Cai P Q, Guo X T, Wang M, Matsuhisa N, Yang L, Lv Z S, Luo Y F, Loh X J, Chen X D 2020 Nat. Commun. 11 4602
- [93] Yang X, Fang Y C, Yu Z Z, Wang Z W, Zhang T, Yin M H, Lin M, Yang Y C, Cai Y M, Huang R 2016 Nanoscale 8 18897
- [94] Rankin C H, Abrams T, Barry R J, Bhatnagar S, Clayton D F, Colombo J, Coppola G, Geyer M A, Glanzman D L, Marsland S, McSweeney F K, Wilson D A, Wu C F, Thompson R F 2009 *Neurobiol. Learn. Mem.* **92** 135
- [95] He H K, Yang R, Zhou W, Huang H M, Xiong J, Gan L, Zhai T Y, Guo X 2018 *Small* 15 1800079
- [96] Zhao B, Xiao M, Shen D Z, Zhou Y N 2020 Nanotechnology 31 125201
- [97] Gong G D, Gao S, Xie Z L, Ye X Y, Lu Y, Yang H L, Zhu X J, Li R W 2021 Nanoscale 13 1029
- [98] Akbari M K, Hu J, Verpoort F, Lu H L, Zhuiykov S 2020 Nano-Micro Lett. 12 83
- [99] Zhou L, Zhang S R, Yang J Q, Miao J Y, Ren Y, Shan H Q, Xu Z X, Zhou Y, Han S T 2020 Nanoscale 12 1484
- [100] Hawkins R D, Byrne J H 2015 Cold Spring Harb. Perspect. Biol. 7 a021709
- [101] Liu L, Cheng Z Q, Jiang B, Liu Y X, Zhang Y L, Yang F, Wang J H, Yu X F, Chu P K, Ye C 2021 ACS Appl. Mater. Interfaces 13 30797
- [102] Ahmed T, Kuriakose S, Mayes E L H, Ramanathan R, Bansal V, Bhaskaran M, Sriram S, Walia S 2019 Small 15 e1900966
- [103] Feldman D E 2012 Neuron 75 556
- [104] Abbott L F, Nelson S B 2000 Nat. Neurosci. 3 1178
- [105] Caporale N, Dan Y 2008 Annu. Rev. Neurosci. **31** 25
- [106] Li Y, Zhong Y P, Zhang J J, Xu L, Wang Q, Sun H J, Tong H, Cheng X M, Miao X S 2014 *Sci. Rep.* 4 4906

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Optoelectronic neuromorphic devices and their applications^{*}

1) (School of Materials Science and Chemical Engineering, Ningbo University, Ningbo 315211, China)

2) (Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, China)

3) (Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai 200031, China)

(Received 16 January 2022; revised manuscript received 22 February 2022)

Abstract

Conventional computers based on the von Neumann architecture are inefficient in parallel computing and self-adaptive learning, and therefore cannot meet the rapid development of information technology that needs efficient and high-speed computing. Owing to the unique advantages such as high parallelism and ultralow power consumption, bioinspired neuromorphic computing can have the capability of breaking through the bottlenecks of conventional computers and is now considered as an ideal option to realize the next-generation artificial intelligence. As the hardware carriers that allow the implementing of neuromorphic computing, neuromorphic devices are very critical in building neuromorphic chips. Meanwhile, the development of human visual systems and optogenetics also provides a new insight into how to study neuromorphic devices. The emerging optoelectronic neuromorphic devices feature the unique advantages of photonics and electronics, showing great potential in the neuromorphic computing field and attracting more and more attention of the scientists. In view of these, the main purpose of this review is to disclose the recent research advances in optoelectronic neuromorphic devices and the prospects of their practical applications. We first review the artificial optoelectronic synapses and neurons, including device structural features, working mechanisms, and neuromorphic simulation functions. Then, we introduce the applications of optoelectronic neuromorphic devices particularly suitable for the fields including artificial vision systems, artificial perception systems, and neuromorphic computing. Finally, we summarize the challenges to the optoelectronic neuromorphic devices, which we are facing now, and present some perspectives about their development directions in the future.

Keywords: optoelectronic neuromorphic devices, optoelectronic synapse, optoelectronic neuron, neuromorphic computing

PACS: 85.60.-q, 72.40.+w, 85.35.-p, 87.18.Sn

DOI: 10.7498/aps.71.20220111

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. U20A20209, 61874125), the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB32050204), the Natural Science Foundation of Zhejiang Pvovince, China (Grant Nos. LD19E020001, LQ22F040003), and the Ningbo Natural Science Foundation, China (Grant No. 2021J139).

[†] Corresponding author. E-mail: zhugefei@nimte.ac.cn
物理学报Acta Physica Sinica





Institute of Physics, CAS

基于非挥发存储器的存内计算技术

周正 黄鹏 康晋锋

Non-volatile memory based in-memory computing technology Zhou Zheng Huang Peng Kang Jin-Feng 引用信息 Citation: Acta Physica Sinica, 71, 148507 (2022) DOI: 10.7498/aps.71.20220397 在线阅读 View online: https://doi.org/10.7498/aps.71.20220397 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

过渡金属元素X(X=Mn,Fe,Co,Ni)掺杂对ZnO基阻变存储器性能的影响 Effect of transition metal element X (X=Mn, Fe, Co, and Ni) doping on performance of ZnO resistive memory 物理学报. 2018, 67(6): 063101 https://doi.org/10.7498/aps.67.20172459

55 nm硅-氧化硅-氮化硅-氧化硅-硅闪存单元的 γ 射线和X射线电离总剂量效应研究 Total ionizing dose effects of γ and X-rays on 55 nm silicon-oxide-nitride-oxide-silicon single flash memory cell 物理学报. 2019, 68(3): 038501 https://doi.org/10.7498/aps.68.20181661

氧分压对Ni/HfOx/TiN阻变存储单元阻变特性的影响

Influences of different oxygen partial pressures on switching properties of Ni/HfOx/TiN resistive switching devices 物理学报. 2018, 67(5): 057301 https://doi.org/10.7498/aps.67.20172194

尺寸调控SnO2量子点的阻变性能及调控机理

Size-controlled resistive switching performance and regulation mechanism of SnO2 QDs

物理学报. 2021, 70(19): 197301 https://doi.org/10.7498/aps.70.20210608

铁电存储器60Co γ射线及电子总剂量效应研究

Total ionizing dose effect of ferroelectric random access memory under Co-60 gamma rays and electrons 物理学报. 2018, 67(16): 166101 https://doi.org/10.7498/aps.67.20180829

铁电存储器中高能质子引发的单粒子功能中断效应实验研究

Experimental study about single event functional interrupt of ferroelectric random access memory induced by 30-90 MeV proton 物理学报. 2018, 67(23): 237803 https://doi.org/10.7498/aps.67.20181225

专题: 面向类脑计算的物理电子学

基于非挥发存储器的存内计算技术

周正 黄鹏 康晋锋†

(北京大学集成电路学院,北京 100871)

(2022年3月5日收到; 2022年6月10日收到修改稿)

通过在基本单元上集成存储和计算功能,存内计算技术能够显著降低数据搬运规模,被广泛认为是突破 传统冯·诺依曼计算架构性能瓶颈的新型计算范式.非挥发存储器件兼具非易失特性和存算融合功能,是实现 存内计算的良好功能器件.本文首先介绍了存内计算范式的基本概念,包括技术背景和技术特征.然后综述 了用于实现存内计算的非挥发存储器件及其性能特征,包含传统闪存器件和新型阻变存储器;进一步介绍了 基于非挥发存储器件的存内计算实现方法,包括存内模拟运算和存内数字运算.之后综述了非挥发存内计算 系统在深度学习硬件加速、类脑计算等领域的潜在应用.最后,对非挥发型存内计算技术的未来发展趋势进 行了总结和展望.

关键词:存内计算,非挥发存储器,闪存,阻变存储器 PACS: 85.35.-p,07.05.Mh,84.35.+i,85.30.Tv

DOI: 10.7498/aps.71.20220397

1 引 言

随着集成电路技术的发展,电子设备逐渐推广 到日常生活的各个方面^[1].电子设备互联互通,产 生的数模规模不断攀升.预计 2025 年,全球的实 时数据规模将达到 47 泽字节 (1 泽字节 = 2⁷⁰字 节)^[2].由此催生出众多数据依赖的信息处理任务, 如模式识别、数字孪生等.然而,由于摩尔定律接 近物理极限,集成电路等比例缩小趋势逐步放缓^[3], 传统计算平台性能进一步提升面临严峻挑战.同 时,传统计算平台受限于冯·诺依曼瓶颈^[4]和存储墙^[5] 等问题,难以应对海量数据搬运和处理的实际需 求.存内计算技术通过在基本单元上集成计算和存 储功能,打破了传统冯·诺依曼瓶颈,能够显著降低 数据搬运,被认为是未来计算架构的重要发展趋势 之一^[6-8].

非挥发存储 (non-volatile memory, NVM) 器件是实现存内计算的优良器件. 首先, NVM 器件

具备断电数据保持特性,工作状态下无需预加载数 据,待机状态下无需多余的能耗开销,从而具备低 功耗特性;其次,NVM 器件具备多值/模拟存储特 性⁹, 能够实现高密度的数据存储和信息处理能力; 最后, NVM 器件能够在器件层次上集成存储和计 算功能^[10],实现存内计算的基本单元结构简单、集 成度高,具备良好的等比例缩小能力^[11]. NVM器 件根据技术成熟度,可分为传统 NVM 器件和新型 NVM 器件^[12]. 闪存 (flash) 器件是典型的传统 NVM 器件,具备工艺成熟、性能稳定等优势.研究人员 拓展了传统 flash 器件的存内计算功能,实现了诸 如卷积神经网络加速^[13]和稳态逻辑计算^[14]等应 用. 新型 NVM 器件类型众多, 具备高集成度、低 功耗和响应迅速等预期优势,被认为能够在存储金 字塔中弥补传统存储和内存之间的存储级内存[15]. 进一步地,新型 NVM 器件表现出优良模拟双向调 制特性[16],具备高密度存内计算应用潜力.

基于 NVM 器件, 研究人员拓展出多种存内计 算模式. 根据运算类型, 可分为存内模拟运算和存

[†] 通信作者. E-mail: kangjf@pku.edu.cn

^{© 2022} 中国物理学会 Chinese Physical Society

内数字运算^[17].存内模拟运算主要利用 NVM 器件 及其阵列结构的器件响应特性和信号调制能力进 行运算,在深度学习加速^[18]以及线性方程组求解^[19] 等方面具备显著优势.存内数字运算则利用多个 NVM 器件的存储状态,通过激励信号和器件之间 的相互作用进行逻辑运算,形成了随机计算和布尔 逻辑^[20]等发展方向.总体来看,近年来随着非挥发 型存内计算技术研究工作的不断深入,逐步形成了 非挥发存内计算器件功能开发、存内计算运算模式 设计实现和存内计算系统应用开发逐步递进的研 究体系.本文将聚焦在非挥发存内计算技术路线, 从 NVM 器件、存内计算运算模式和存内计算应用 场景三个层面,回顾近年来研究人员在非挥发型存 内计算技术领域所取得的最新研究进展,总结当前 面临的关键问题,展望未来发展趋势与前景.

2 存内计算范式

根据存储模块与计算模块的相对关系,可以将 硬件系统的计算范式分为存算分离范式、近存计算 范式和存内计算范式(又称为"存算一体"或"存算 融合")^[21,22], 如图 1 所示. 传统冯·诺依曼架构不仅 是典型的存算分离的计算范式,还是当前主流处理 器 CPU 和 GPU 等计算平台的架构基础. 其特征是 计算部分与存储部分相互独立,并通过总线连接[23], 具备"程序存储、共享数据、顺序执行"的特点;此 外以计算单元为中心执行任务,严重依赖存储元件 的数据交互能力.可把近存计算范式看作是对存算 分离范式的优化设计. 通过平衡系统中存储体系各 部分速度和容量、引入高带宽存储模块、增加片上 存储容量等手段,降低数据搬运延时、提高数据带 宽,从而提升系统性能. 张量处理器 (tensor processing unit, TPU)^[24]和网络处理器 (neural-network process unit, NPU)^[25] 等新型计算平台是近存计算 的典型代表.存内计算范式^[26]则从根本上改变了

存储和计算的关系,其特征是在基本单元内同时实现计算和存储功能,模糊了存储和计算的界限,从 根本上缓解了数据搬运的问题.

存内计算作为一种新型计算范式,对硬件单 元、运算逻辑和系统架构都提出了全新的需求和挑 战,并需要针对应用场景进行定制化开发.首先, 存内计算基本单元是构成存算融合的基础, 需同时 具备存储和计算两种功能.存内计算将基本单元的 存储状态视为逻辑计算的输入或输出,在基本单元 内部或附近实现计算功能,由此减少输入或输出的 数据搬运.构成存内计算基本单元的可以是复杂的 逻辑电路^[27],也可能是器件组合^[28]甚至单个器件^[29]. 从集成角度看,单个器件或器件组合能够显著降低 存内计算单元的开销.其次,在基本单元的基础上 需构建恰当的运算逻辑. 与互补金属氧化物半导 体 (complementary metal-oxide-semiconductor, CMOS) 逻辑门不同,存内计算不再局限在布尔逻 辑范畴内, 而是适当地引入硬件兼容的算数运算作 为基本运算功能,如加法、乘法等,从而在有限的 硬件资源条件下,提高特定任务的计算效率.最后, 受输入输出数据的形式和基本运算类型的影响,存 内计算的系统架构面临诸多挑战. 诸如算数运算的 输入输出模拟信号处理问题^[30]、运算过程的级联 和中间数据的存储问题[31] 以及硬件资源复用问题[32] 等. 针对实际任务进行定向开发是存内计算技术的 重要推动力量.因此,存内计算的研究工作,往往 是根据任务的运算需求和数据形式,如深度学习的 多层网络结构、状态逻辑的级联特点等,进行定制 化的基本单元设计和优化,进而研究基本运算的实 现方式,提出对应的存内计算系统架构.

3 非挥发型存算一体功能器件

根据数据是否断电易失,可以将存内计算分为 挥发型和非挥发型.以静态随机存取存储器 (static



图 1 传统计算范式分类 (a) 存算分离计算; (b) 近存计算; (c) 存内计算

Fig. 1. Classification of the traditional computational paradigms: (a) Separated memory and logic computing; (b) near memory computing; (c) in-memory computing.

random-access memory, SRAM) 为代表的易失器 件可以构成典型的易失型存内计算系统[33,34]. 该系 统具备配置灵活、性能稳定和技术成熟等优势,能 够显著缓解存储墙等问题. 然而, 由于缺乏断电数 据保持特性,需要进行初始化数据搬运,从而无法 脱离存储模块而独立工作.同时,基于 SRAM 的存 算融合基本单元结构复杂,超大规模集成存在挑 战. 非易失型存算融合系统主要依托于非易失型存 算一体功能器件,包括以 flash 为代表的传统非易 失存储器和以阻变存储器 (resistive random access memory, RRAM) 为代表的新型非易失存储器件. 由于具备非易失特性,系统运行过程不仅无需外部 数据的预加载,还具备低待机功耗、快速响应等性 能优势. 另外, 由于非易失存算功能器件在器件层 级上就实现了存储和计算功能的集成,从而具备更 好的等比例缩小能力,在集成大规模系统方面具有 天然优势.

3.1 传统 NVM 存算功能器件: Flash

Flash 器件具有工艺成熟、性能稳定和产业化 程度高等显著优势,是存算融合系统应用实现的优 良硬件载体^[35].如图 2(a)所示, flash 器件的浮栅 结构位于栅极氧化层和控制层之间,由具备电子缺陷的绝缘层和两侧 SiO₂ 材料层构成^[36].电荷存储在电子缺陷层不连续的缺陷状态中,可以通过热电子注入和 Fowler-Nordheim (F-N) 隧穿效应,改变电荷的存储状态,从而实现数据的写入和擦除^[36],见图 2(b).由于需要克服界面势垒,数据写入需要较高电压 (>10 V) 和较长时间 (>10 µs),循环擦写次数约为10⁵—10⁶ 次.根据 flash 单元的排列方式,主流 flash 阵列结构可分为 NOR 型和 NAND 型.在 NOR 型 flash 阵列中,如图 2(c) 所示,相同位线 (bit line)的 flash 呈并联结构; NAND 结构中,如图 2(d) 所示, flash 通过控制管与位线和源线 (source line) 相互串联^[36].

基于 flash 的存内计算研究涵盖了器件、阵列 和系统架构等方面. 2015年,加州大学圣塔芭芭拉 分校 (UCSB)课题组^[37]在商用 ESF1NOR型 flash 的基础上开发了高精度的编程算法,实现了高于 10位的精确存储能力.为了降低功耗,利用 Flash 亚阈值区域,限制单元电流在 10⁻¹⁰—10⁻⁶A范围 内. 2017年, UCSB 课题组^[38]进一步实现了首款基 于 180 nm NOR型 flash 工艺的全连接神经网络加 速芯片. 2018年,北京大学 Han 等^[39]提出基于 NOR



图 2 传统 NVM 器件 flash^[36] (a) 典型器件结构; (b) flash 操作模式与物理机制; (c) NOR 型阵列结构; (d) NAND 型阵列结构 Fig. 2. Flash, a traditional NVM device^[36]: (a) Typical flash device structure; (b) the operation scheme and physical mechanism; (c) NOR flash array; (d) NAND flash array.

型 flash 的卷积计算加速方案,并基于 65 nm flash 工艺节点进行了流片验证,结果表明单位功耗下处 理器可执行运算操作 14.5 次,即能效比达到 14.5. 在 2019 年,明尼苏达大学 Kim 等^[40]研发了基于 eNAND 型 flash 的矩阵向量乘法器,并协同压控 振荡器实现了 LeNet-5 网络. 2019 年,旺宏公司成 功演示了基于 64 GB 容量、SLC 存储密度和 3D 堆叠结构的 NAND 型 flash 颗粒,研制出矩阵向量 乘积运算加速器^[41].

3.2 新型 NVM 存算功能器件: RRAM

为了进一步提升存算融合系统性能,相继提出 了一系列新型非挥发存算功能器件.如以相变效应 形式存储信息的相变存储器^[42],即利用电流产生 的焦耳热促使相变材料在晶态和非晶态之间发生 转变,非晶态为高阻、晶态为低阻;以自旋转移力 矩形式存储数据的磁阻随机存储器^[43],由自由磁 层、隧穿层和固定磁层组成,当自由磁层和固定磁 层的磁场方向平行,表现为低阻态,反之为高阻态; 以自发极化铁电效应形式存储信息的铁电器件^[44], 即利用铁电材料在不同电场作用下,晶体中原子产 生位移而导致正负电荷的中心位置发生偏移,形成 极化向上和向下两种稳定状态;以细丝导电通道通 断效应存储信息的 RRAM^[45].其中,RRAM 具备 低功耗、高密度集成等优势,是新型 NVM 器的典 型代表.RRAM 结构简单,单元面积可达 4 F^2 (F为光刻工艺所能达到的最小特征尺寸),具备三 维集成能力.同时,RRAM 可实现约 85 ps 的电阻 转变、擦写次数大于 10¹²、擦写电流小于 15 μA、阻 变窗口大于 100、高温 (150 °C)数据保持时间超 过 10 年.

典型的 RRAM 工作过程可分为 3 个阶段, 如 图 3(a) 所示: 1) 导线细丝初始化过程 (forming), 一般是利用较大幅度的电压或电流作用器件, 形成



图 3 新型 NVM 器件 RRAM (a) 常见阻变行为^[46]; (b) 双极型 RRAM 的典型 *I-V* 特性曲线; (c) 基于氧空位的阻变物理机制模型^[46]; (d) 常见 RRAM 单元结构包括 1R, 1S1R 和 1T1R

Fig. 3. RRAM, a novel NVM device: (a) Typical resistive switch behavior^[46]; (b) the typical *I-V* curve of bipolar RRAM device; (c) the physical mechanism of oxide-based RRAM^[46]; (d) the typical basic unit based on RRAM include 1R, 1S1R and 1T1R.

软击穿,使其从初始阻态不可逆地转变为低阻态; 2) 复位过程 (reset), 利用脉冲或直流信号, 在阻变 器件两端形成反向电势差,使其从低阻态转变为高 阻态; 3) 置位过程 (set), 同样利用脉冲或直流信 号,在器件两端形成正向电势差,使其从高阻态转 变至低阻态. 双极型 RRAM 的典型 I-V特性曲线 中(图 3(b)),置位和复位是非破坏性的可逆过程^[46]. 双极型阻变器件的电阻转变过程受外加电压极性 的控制,置位操作和复位操作施加的电压极性相反 且方向固定. RRAM 的物理机制如图 3(c) 所示, 在初始化/置位过程外场作用下,处于格点上的氧 离子被激发为游离态,同时在原位上留下氧空位缺 陷^[46]. 游离态氧离子在电场作用下逆着电场跳跃 到电极和氧化物的界面处,被电极吸附并存储在电 极中.伴随着更多的氧空位产生,形成了一条连通 上下两个电极的导电通道. 在复位过程中, 电场方 向反转,存储在电极附近的游离态氧离子被释放, 并跳跃到氧空位的俘获截面内与氧空位复合,导致 导电通道发生断裂,器件从低阻态转变为高阻态. 常见的 RRAM 单元结构如图 3(d) 所示, 包括 1R, 1S1R 和 1T1R 结构. 1R 具备 4F² 的器件集成密度 优势, 1T1R 限制了阵列的串扰能够实现较大阵列 规模.1S1R 兼具二者优势,但面临选通管器件工 艺不成熟, 1S1R 协同设计难度大等问题.

虽然在 20 世纪 60 年代已经发现了阻变现象, 但 RRAM 真正开始得到学术界和工业界的广泛关 注,是从 2004 年三星公司在国际电子器件大会上 发布基于 NiO 的 RRAM 器件开始^[47].此后,基于 不同材料体系的忆阻行为相继被发现,包括氧化 物、电解液、低维材料等无机材料体系和纳米纤维 素、聚合物等有机材料体系^[48–50].导电前段移动模 型、导电细丝数量调控模型、肖特基势垒模型等物 理机制模型^[9]也相继被提出用以解释阻变行为.与 此同时,基于阻变器件的存内计算应用研究相继展 开,逐渐发展出深度学习加速、类脑计算、状态逻 辑等新兴研究方向^[51].

4 非挥发型存内计算模式

根据运算不同的模式,可以将存内计算模式分 为存内模拟运算和存内数字运算.存内模拟运算是 利用器件的模拟特性和信号调制能力,结合器件阵 列的结构特征,实现如乘法、加法等基本算术运 算,从而在存内计算单元阵列上完成模拟运算.存 内数字运算的主要特点是利用固定的外界激励信 号,通过多个存内计算单元之间的相互作用和单元 的存储状态,以满足布尔逻辑的方式实现存内运算 功能.

4.1 存内模拟运算

4.1.1 向量-矩阵运算模式

如图 4(a) 所示,存内计算单元在二维空间呈 阵列分布,并通过交叉结构相互连接.利用存内计 算单元的存储特性,可将矩阵元素映射至单元中. 在阵列的每行,同时输入激励信号来表示输入向 量.同行存算单元同时对激励信号做出响应,将同 列存算单元的响应累积起来构成列向量输出,即向 量矩阵乘积结果.以两端器件 RRAM 为例,利用 欧姆定律同列器件可同时实现输入电压与电导的 乘积操作.同时利用基尔霍夫电压定律,RRAM 阵 列能够在一个周期内完成矢量与矩阵的乘累加运 算.对应的数学表达式为

$$I_j = \sum_{i=1}^n V_i G_{ij},\tag{1}$$

其中Vi为第 i行的输入电压, Gii 为存储器阵列中 第*i*行、*j*列的电导值, *I_i*为第*j*列的输出电流值. 为了完善矩阵向量乘积方法, Pedretti 等^[52]探索 了适配各类存内计算单元存储能力的均匀编码、位 编码等矩阵表示方案,如图 4(b) 所示,与之对应的 映射误差见图 4(c). 同时,研究人员也探索了输入 向量的信息编码方法,提出了幅值、脉宽等时空编 码方案等[52,53],能够有效地提高数据传输密度和运 算效率.为了实现完备的正负输入和正负权重的运 算, Park 等^[54] 设计了如图 4(e) 的运算方法; Li 等^[55] 探索了输出信息的模数转化方法. 然而, 在实际情 况下,向量矩阵乘积仍面临较多挑战.例如,通常 存内计算器件并非理想的欧姆器件,电压响应呈现 非线性特性^[56],见图 4(f);交叉阵列的线阻影响激 励信号传输,会引起运算偏差^[57],见图 4(g);输出 向量的模数转化方法实现方法仍不完善等.

4.1.2 向量-向量运算模式

如图 5(a) 所示, 基于交叉阵列的互连结构, 将 行和列作为输入, 利用存内计算单元对外界施加的 组合激励信号的响应作为输出结果, 直接存储在存



物理学报 Acta Phys. Sin. Vol. 71, No. 14 (2022) 148507

图 4 向量-矩阵运算模式 (a) 基本原理; (b) 矩阵编码模式^[52]; (c) 器件状态波动性对编码的影响^[52]; (d) 向量编码模式^[53]; (e) 正负输入和权重的运算方法^[54]; (f) 器件 *I-V* 非线性^[56]; (g) 交叉阵列互联电阻^[57]

Fig. 4. Vector-matrix operation mode: (a) The basic principle; (b) matrix encode (mapping) scheme^[52]; (c) impact of device variation on the matrix $encode^{[52]}$; (d) input vector encode schemes^[53]; (e) the operation method of positive and negative input and weight^[54]; (f) the nonlinearity *I-V* behavior of device^[56]; (g) the interconnect resistance of cross-bar $array^{[57]}$.

内计算单元内,从而实现了行向量与列向量的运算 功能.仍以两端器件 RRAM 为例,利用两端信号 的叠加作用,阵列中每个单元均受其所在行和列的 激励信号的影响.具体数学表达式为

$$S_{ij}^{t+1} = f(V_i^1, V_j^2, S_{ij}^t), \tag{2}$$

其中, V_i¹为第 *i*行的电压, V_j²为第 *j*列的电压, S^t_{ij} 为第 *t*时刻位于第 *i*行、*j*列的 NVM 器件的存储 状态, *f*(·)为 NVM 器件在不同存储状态下, 对激 励信号的影响关系. 利用这一特性 Liao 等^[58]提出 了向量-向量乘积方案, 如图 5(b)所示, 通过将运算 矩阵拆解为向量的形式来实现完整的矩阵运算. 2019年, Ambrogio 等^[59]实现了时间依赖可塑性 学习规则, 如图 5(c)所示, 用于类脑计算的学习法 则; 另外, 可将器件的累积特性视为加法功能, 从 而可以实现基于存内计算的半加器^[60], 见图 5(d). 实际情况下, 向量-向量运算模式依然面临诸多问 题,如存储状态对单元激励响应的影响,激励信号的编码方案、互联电阻的影响,存内计算单元的存储容量限制(图 5(e))^[61]等.

4.2 存内数字运算

计算机硬件基于二进制数据的表示和处理,布 尔逻辑在硬件体系结构中结构的描述、构建和优化 过程方面扮演着十分重要的角色,也是存内数字运 算的研究目标.常见的布尔逻辑如图 6(a)所示,其 中 NAND, NOR 和蕴含逻辑 (implication, IMP) 是完备的逻辑形式,仅利用任意一种逻辑,通过组 合就可实现其他所有逻辑类型,成为存内数字运算 硬件实现的首要目标.如图 6(b), (c)所示,利用传 统挥发型存储器 SRAM/动态随机存取存储器 (dynamic random access memory, DRAM) 均可实现 完备的布尔逻辑功能.然而,面向大数据应用,直



图 5 向量-向量运算模式 (a) 基本原理; (b) 向量形式的矩阵-矩阵乘积运算^[58]; (c) 尖峰时间依赖可塑性学习规则^[59]; (d) 一种 基于 RRAM 的半加器实现方式^[60]; (e) 典型 NVM 器件存储状态饱和限制^[61]

Fig. 5. Vector-vector operation mode: (a) The basic principle; (b) the matrix-matrix multiplication based on vector form^[58]; (c) the spike time dependent plasticity learning rule^[59]; (d) the half-adder implementation approach based on RRAM^[60]; (e) the saturation limited states range of typical NVM device^[61].

接在非易失存储器内实现逻辑运算更具优势. 根据 输入输出物理量的不同,可以将存内数字运算分 成 V-R 型、R-V型、V-V型和 R-R 型这 4 种逻辑形 式. 以 RRAM 为例, 如图 6(d), V-R 型逻辑的输 入A和B由施加在 RRAM 两端电极的电压高低 表示,逻辑结果为 RRAM 的存储状态. 2011年, 亚琛工业大学[62] 通过利用双极型阻变器件实现了 V-R型逻辑.在此基础上,华中科技大学^[63]利用 1T1R 单元的栅、源、漏和 RRAM 的阻态作为逻辑 输入,实现了广义上的 V-R 逻辑,能够在两步操作 内实现任意 16 种布尔逻辑. 由于 V-R 逻辑的输出 结果为存内计算单元的存储状态,具备原位存储特 征,是一种高效的逻辑实现方法.然而,由于输入 输出变量统一,级联需借助额外的信号转换电路. 如图 6(e) 所示, R-V型逻辑将 RRAM 的存储状态 作为逻辑输入^[64],通过对A和B同时施加读电压, 比较公共节点与基准值,实现组合逻辑.通过设置 不同的基准值,可以实现 NAND, AND, OR, NOR, XOR 和 XNOR 这 6 种常见逻辑. 与 V-R型逻辑 类似, R-V逻辑的输入输出物理量依然不统一, 也 面临着无法直接级联的问题. V-V型逻辑如图 6(f) 所示, 输入输出物理量均为电压信号[65]. 与 R-V型 逻辑的固定读电压不同, V-V型逻辑通过是否施加

读电压来代表输入1和0,在输出端同样使用基准 值进行比较,得到输出电压值^[65].可见,V-V型逻 辑具备的直接级联的优势,但其输出结果是易失 的.*R-R*型逻辑是利用 RRAM 件与辅助电阻的分 压原理实现逻辑运算,如图 6(g)所示,其输入输出 物理量均为 RRAM 的存储状态^[66].*R-R*型逻辑是 最典型的非易失逻辑,具备直接级联的优势.同时, *R-R*型逻辑的输入输出均为 RRAM 的存储状态, 因此具备非易失特性.

尽管存内数字运算研究进展显著,但距离实际 应用仍存在许多关键科学问题需要解决.在器件层 次,存内计算单元的一致性和擦写次数有待提高, 以满足逻辑运算的准确性和频繁性;在阵列层次, 泄漏电流和线阻等问题有待解决,以推动存内数字 运算向大规模电路的方向发展;在逻辑级联层次, 单步逻辑的可靠性问题被进一步放大,实现复杂逻 辑运算面临较大的挑战.

4.3 存内计算模式的综合比较

表1对比了上述存内计算模式,可以得出以下 结论.第一,存内计算的计算模式呈多样化发展趋势,这种多样化趋势归根于应用场景中运算的多样 化需求.由于利用存储器件实现各类运算功能存在



物理学报 Acta Phys. Sin. Vol. 71, No. 14 (2022) 148507

图 6 存内数字运算模式 (a) 常见逻辑真值表; 基于 SRAM (b) 和 DRAM (c) 的逻辑实现方案示例^[26]; 基于 NVM 器件的逻辑 (d) *V-R*型, (e) *R-V*型, (f) *V-V*型, (g) *R-R*型

Fig. 6. In-memory digital computing mode: (a) The true value table of typical logic; SRAM (b) and DRAM (c) based logic implementation^[26]; logics based on NVM device: (d) V-R type, (e) R-V type, (f) V-V type, (g) R-R type.

典型差异,从而催生出不同类型的存内计算模式. 各类计算模式均在其特定的应用场景下发挥着重 要作用. 第二, 存内模拟运算充分利用了存储器件 的多值和模拟特性,使单位面积的计算密度得到显 著提升. 通过利用器件及其互联结构的电学特性, 存内模拟计算能够在模拟域进行诸如乘法、加法的 代数逻辑运算,从而突破布尔逻辑门的限制.然而, 在功能器件的有限动态范围、编程精度等客观条件 限制下,存内模拟计算更加适合非精确计算的应用 场景. 第三, 存内数字运算本质上是利用功能器件 实现存算一体的完备布尔逻辑.由于具备原位的计 算和存储能力,存内数字运算相较传统布尔逻辑, 具备更短距离的数据搬运特点和分布式计算的特 征,能够在数据密集型应用中承担数据预处理能 力,补充传统计算架构的不足.然而,受限于器件 有限的鲁棒性和波动性, 在较为严苛的布尔逻辑运 算的需求面前,存内数字运算的可靠性有待进一步 提升. 第四, 得益于器件的非易失特性、优秀的等 比缩小能力和高并行度的阵列拓扑结构,存内模拟 运算和存内数字运算在各自擅长的领域内,在能耗 开销、运算效率和系统集成度等方面均具备一定的 性能优势和发展潜力.

表 1 存内计算模式的特征

Table 1. Feature of in-memory computing modes.

	存内模拟计算	存内数字运算
功能	布尔逻辑,代数运算	布尔逻辑
优势	高运算密度,高并行度, 缓解数据搬运	精确计算,高并行度, 缓解数据搬运
挑战	运算精度,模数转化	器件鲁棒性、波动性
应用	深度学习、类脑计算等	逻辑电路、嵌入式存储

5 非挥发型存内计算系统应用

5.1 深度学习硬件加速器

深度学习是第三代人工智能算法的典型代表, 在模式识别和自然语言处理等领域具备显著的性能优势^[67].反向传播算法是深度学习的理论基础^[68],其特点是基于随机梯度下降方法,利用大量 的样本和标签信息,通过损失函数不断对多层网络的权重矩阵进行逐步更新,不断趋近理想权值.通过不断加深网络层数和权重参数规模,深度学习在高维特征变换、信息过滤等方面取得了显著进步.由于深度学习算法包含海量的网络权重,算法执行过程包含大量的特征图像与权重矩阵的乘积求和运算,十分契合存内计算模拟运算,被认为是存内计算的典型应用场景^[69].

如图 7(a) 所示, 深度学习算法种类繁多, 囊括 了全连接网络、卷积神经网络和循环神经网络等众 多拓扑结构[70]. 其中, 全连接结构是深度学习的基 本结构,其前后神经元相互连接,等价于交叉阵列 结构构成的行列互连结构^[50],见图 7(b);卷积神经 网络是在全连接结构下,进行结构化稀疏,通过选 通特定的行列也可在交叉阵列结构中实现;循环神 经网络则是引入时间因素,将输出结果作为部分或 全部信息,再次输入到互联的网络内进行迭代计 算,同样将交叉结构作为基础.如图7(c)所示,深 度学习硬件加速器根据功能不同,可分为推理功能 加速和训练功能加速. 实现推理功能时, 需将算法 权重映射至存内计算单元内,利用存内计算模拟运 算的性能优势,加速网络的前传计算能力;实现训 练功能时,不仅需要推理功能实现前传和反传运 算,同时还要求存内计算单元能够实现原位的权重 更新.

针对深度学习硬件加速研究,研究人员分别在 器件层次、运算核层次和系统层次展开了深入研 究. 在器件层次, 深度学习硬件加速器面临着器件 性能的有限表现与算法权重较高需求的不适配问 题,主要表现为器件有限的动态范围、编程精度、 存储状态波动、信号响应非线性和非对称等非理想 因素.为克服这一系列问题,研究人员提出了各类 算法在不同运算精度下的权重映射方法. 2019年, Huang等^[71]研究了利用非线性非对称器件进行权 重更新的方法 (图 7(d)); 2021 年, Feng 等^[72] 研究 了器件的权重调制方法、优化了器件的状态保持 特性 (图 7(e), (f)). 同时, 深度学习算法也朝着轻 量化方向发展,衍生出硬件更加友好的二值化、少 值化的算法变体. 在运算核层次, 面临着利用存储 阵列实现高效数据传输和模数混合运算的挑战,不 仅要克服阵列结构本身的串扰、线阻等问题,也需 要解决外围控制电路的设计实现问题. 相关的研究 工作在数据搬运方法、后处理模数转化实现方案、 无需模数转化的多级直连传输方法等方面取得了 一定的研究进展^[73,74].在系统层次,如图 7(g)所 示,需在保证运算效率的前提下解决有限硬件资源 的分配问题^[75].研究人员提出了图 7(h)所示的流 水线式网络映射方法,研究了运算核的协同工作模 式以及多核存内计算架构^[76,77]等问题.同时,根据 训练和推理的不同功能特点,需进行适当的加速器 架构研究.研究人员基于 NVM 器件提出了兼容多 种算法的存内计算系统框架^[78,79],并验证了存内计 算硬件加速研究依然面临着图 7(i)所示的稀疏神 经网络的适配性问题,也需要解决图 7(j)所示的输 入信息的数据搬运问题^[74].

5.2 类脑计算

当前的人工智能技术仍然存在一定局限性,相 比人脑的功能多样性和复杂度仍存在明显差距.为 了更接近人脑功能,人们提出了神经形态计算的概 念.预期的神经形态计算系统,具备在功能上模拟 脑、性能上趋近脑、规模上超越脑的典型特征,这 被认为是未来人工智能的发展方向,也将是存内计 算技术的重要应用场景.仿脑(brain-like)和类脑 (brain-inspired)神经网络都属于神经形态硬件系 统的研究范畴.前者侧重模仿人脑神经网络的工作 模式,注重模拟生物神经元、突触等基本单元的功 能,以期望更接近人脑的工作模式;后者偏向在生 物的基础上抽象数学模型,构建基本单元的数学模 型并发展相应的算法理论.仿脑为类脑提供了硬件 基础,类脑拓展了仿脑的发展空间,二者相辅相成, 构成了当前神经形态硬件系统的发展方向.

如图 8(a)—(g) 所示,在仿脑领域研究人员利 用存内计算功能器件实现生物突触的长程可塑性 (long-time plasticity, LTP)、短程可塑性(shottime plasticity, STP)和长短程可塑性的转变,实 现了生物突触的双脉冲易化特性^[80],实现了尖峰 脉冲时间依赖和频率依赖可靠性^[80-82]等一系列突 触基本功能;如图 8(h),(i)所示,Li等^[83]利用存 内计算功能器件实现了生物神经元的阈值特性、非 线性信号调制能力和信号激励特性等.在类脑领 域,Lashkare等^[84]利用存内计算功能单元实现了 积分触发和泄漏积分触发神经元,实现了基于尖峰 时间和频率依赖可塑性的学习法则.Milo等^[85]利 用全连接网络演示了无监督学习能力(图 8(k));



图 7 存内计算加速深度学习 (a)常见深度学习算法分类^[70]; (b)存内计算加速深度学习的基本原理^[50]; (c)深度学习各功能的存内计算实现方式; (d)利用二值神经网络算法克服器件非线性的影响^[71]; (e)器件操作优化方案^[72]; (f)利用激励信号波形抑制器件波动性^[72]; (g)基于存内计算的深度学习加速器的典型架构^[73]; (h)流水线硬件实现方法加速网络运算效率; (i)神经网络稀疏性表现形式,结构化和非结构化; (j)减少输入信息搬运的数据调用方案^[74]

Fig. 7. In-memory computing based deep learning accelerator: (a) The classes of deep learning algorithms^[70]; (b) the basic principle of in-memory computing accelerates deep learning algorithm^[50]; (c) in-memory computing implementation of deep learning functions; (d) solve the impact of device non-linearity switch behavior by binarized neural network^[71]; (e) the optimized programming scheme of device^[72]; (f) improve the device reliability by optimizing the stimulus signal^[72]; (g) typical architecture of deep learning accelerators based on in-memory computing^[73]; (h) pipeline weight mapping approach to speed up network computing efficiency; (i) the sparsening of neural network: structured and unstructured; (j) data call scheme to reduce input information handling^[74].

图 8(1) 实现了霍普菲德网络和联想学习功能^[86]等. 同时,图 8(m)为 Larkum^[87]探索得更复杂的仿生 神经网络模型.基于存内计算的脉冲神经网络也在 不断推进^[88,89],如图 8(n)所示的可直接级联的脉冲神经网络^[75].然而,当前的神经形态计算研究仍存在巨大挑战.在仿脑领域,仍面临着如何完整揭



图 8 基于存内计算技术的类脑计算研究 (a) 生物突触结构;利用 NVM 器件实现突触的 (b) LTP, (c) STP 和长短程可塑性转变; (d) 双脉冲易化响应特性^[80]; (e) 尖峰脉冲频率依赖可塑性^[80]; (f) 尖峰脉冲时间依赖可塑性^[80]; (g) RRAM 中的 Bienenstock-Cooper-Munro 权重更新规则^[82]; (h) 生物神经元结构^[83]; (i) 基于 RRAM 的神经元树突非线性调制功能^[83]; (j) 神经元积分触发功能^[84]; (k) 基于尖峰脉冲频率依赖可塑性的脉冲神经网络非监督学习功能^[85]; (l) 霍普菲德网络学习规则^[86]; (m) 生物神经网络理论模型^[87]; (n) 脉冲神经网络实现方案^[75]

Fig. 8. Neuromorphic computing based on in-memory computing: (a) The biological synapse; (b) LTP, (c) STP and the conversation between STP and LTP of NVM device based artificial synapse; (d) double pulse facilitated response characteristics^[80]; (e) the spike rate dependent plasticity (SRDP) ^[80]; (f) the spike-time dependent plasticity (STDP) ^[80]; (g) the Bienenstock-Cooper-Munro weight update rules in RRAM^[82]; (h) the principle of biological neural^[83]; (i) the signal modulation capability of the RRAM based artificial dendrite^[83]; (j) neuron integration-fire function^[84]; (k) unsupervised online training follows the spike rate dependent plasticity based spike neural network learning rule^[85]; (l) the Hopfield eLearning rules^[86]; (m) the model of biological neural network^[87]; (n) implementation of spiking neural network^[75]. 示完备的人脑工作原理、如何在单一元件内集成多种仿生功能、如何在系统层次上融合各类生物功能等重要问题.在类脑领域,则面临着理论模型不完善,网络算法功能单一、性能不足等显著问题,仍然难以在结构和功能上模拟生物神经网络的完整功能.二值化、少值化的脉冲神经网络是探索类脑计算系统功能的重要手段,研究工作利用二值突触实现了低复杂度的硬件友好的脉冲神经网络^[90].

5.3 非易失型布尔逻辑

2010年,惠普公司利用 RRAM 在特定电压脉 冲下发生阻值变化的特点,实现了一种状态逻辑计 算功能^[91].这种逻辑计算的基础是实质蕴含逻辑, 数学形式是"*p*IMP*q*"(图 9(a)).结合 FALSE逻辑, 实质蕴含逻辑可以实现完整的二值数字逻辑函数, 从而形成了一种在存储器内实现完整逻辑计算的 技术路线.2014年,以色列理工学院利用 RRAM

的分压关系进行逻辑运算,实现了 NOR 逻辑,从 而提出了一种名为 MAGIC 的 RRAM 辅助逻辑操 作方案^[92]. 2016年,北京大学 Huang 等^[93] 通过改 变 RRAM 上的电压信号施加方法实现了一种布尔 代数逻辑,即一步操作实现 NAND 和 AND 逻辑, 具备逻辑重构功能,并建立了图 9(b) 所示的存内 计算系统架构. 这三种实现方案各有优劣. 从结构 上看,实质蕴含逻辑和布尔代数逻辑均需负载电阻 参与逻辑操作,影响存内计算的集成密度.从操作 步骤看,实质蕴含逻辑和 RRAM 辅助逻辑均只能 实现一种基础逻辑操作,故实现完整逻辑需较多操 作步数,复杂度较高.从输入输出状态变化看,实 质蕴含逻辑的一个输入与输出共享同一个器件,会 改变输入信息. RRAM 辅助逻辑需要在满足特定 条件的阻变器件中才能实现,否则输入器件的存储 状态也可能发生改变. 为了提升稳态逻辑的系统可 靠性,北京大学 Shen 等 [94] 在 2019 年提出了基于



图 9 非易失状态逻辑 (a) IMP 逻辑实现方案^[91]; (b) 状态逻辑运算核架构^[93]; (c) 基于 1T1R 的状态逻辑^[94]; (d) 利用 1T1R 结构抑制交叉阵列串扰^[94]; (e) 利用寄生电容替代辅助 RRAM 的状态逻辑^[95]

Fig. 9. Non-volatile stateful logic: (a) implementation scheme based IMP logic^[91]; (b) the architecture of stateful logic process core^[93]; (c) 1T1R based stateful logic^[94]; (d) reduce the impact of sneak path by 1T1R structure^[94]; (e) parasitic capacitor assisted RRAM based stateful logic^[95].

1T1R 的状态逻辑实现方式,如图 9(c) 所示,用以 抑制阵列中的泄漏电流.2020年,Shen 等^[95]提出 了一种基于寄生电容的状态逻辑实现方案,从而省 去了辅助电阻,形成了更易于加工的阵列制备方 案,见图 9(e).

5.4 内容可寻址存储器

内容可寻址存储器 (content addressable memorv, CAM) 是一种面向超高速数据搜索应用的 一种存储系统, 如图 10(a) 所示. CAM 的工作原理 与随机存取存储器相反,输入为存储信息,输出为 存储地址.其工作过程为将输入信息依次或并行 与 CAM 内部信息比较, 输出匹配的存储地址. 根 据 CAM 的单元存储状态, 分为二态内容可寻址存 储器 (BCAM) 和三态内容可寻址存储器 (TCAM). TCAM 除了 0/1 外, 还包含"don't care"状态, 即 该状态是否匹配不影响比较结果. 传统 CAM 通常 使用 SRAM 作为基本存储单元, 见图 10(b). 然而, 尽管器件尺寸仍在不断缩小, SRAM 的面积开销 和泄漏电流问题仍限制着 CAM 的进一步发展.存 内计算单元由于具备高密度存储能力、低读取功耗 和 NVM 等特点, 是发展 CAM 技术的一种潜在方 向. 2011年, IBM 利用 PCM 器件演示了 CAM 和 TCAM 的基本功能^[96]. 相较传统基于 SRAM 的实现方式,基于 PCM 的 CAM 在存储密度和能耗开销方面展现出超过 5 倍的性能优势. 匹兹堡大学 Yan 等^[97]提出了基于 STT-MRAM 的 TCAM 模块,并设计演示了 Dual-N 型和 P-N 型方案,改善了搜索延时性能 (见图 10(c)). 2018 年,Grossi 等^[98]研制出基于 RRAM 的 128 bit TCAM 宏,搜索速度与传统 TCAM 齐平,并取得了良好的系统可靠性表现 (图 10(d)). 图 10(e) 为北京大学 Yang 等^[99]基于 3D NAND 型 flash 提出了一款超低功耗、高存储密度的 TCAM 实现方案.结果如图 10(f) 所示,通过比较读出电流与阈值电压,可以分辨出搜索信息与存储信息的匹配与否.预期每次搜索每比特能耗能够达到 0.298 fJ (64 bit word)和大于 582 倍的存储密度 (96 layer).

5.5 线性方程组求解器

线性方程组是科学计算领域极其重要的运算 需求,广泛应用于天气预报、半导体器件仿真等实 际场景.在绝大多数情况下,解析求解线性方程组 都是不现实的.通常利用数字求解的方法,将解空 间近似为离散网格,通过高精度运算不断迭代,得 到满足实际精度需求的近似解.矩阵-向量乘积是



图 10 (a) CAM 基本实现方式^[91]; (b) 基于 SRAM 的 TCAM 基本单元^[93]; (c) 基于 STT-RAM 的 TCAM 基本单元^[97]; (d) 基于 RRAM 的 TCAM 基本单元^[98]; (e) 基于 NAND 型 flash 的 TCAM 实现方式^[99]; (f) 匹配与失配的输出结果示例^[99]

Fig. 10. (a) Typical structure of CAM^[91]; (b) TCAM basic unit based on SRAM^[93]; (c) TCAM basic unit based on STT-RAM^[97];
(d) TCAM basic unit based on RRAM^[98]; (e) based implementation of TCAM based on NAND flash^[99]; (f) example of the matched and mismatched results^[99].

迭代计算的关键步骤,涉及大量的数据搬运和乘加 运算.存内模拟计算技术恰好在矩阵运算方面具备 天然优势,因此研究人员提出了基于存内计算技术 的偏微分方程求解器. 2018年, 如图 11(a) 所示, 密西根大学 Zidan 等^[100] 提出了利用位切片技术 的 16 bit 全 RRAM 偏微分方程求解器. 为了提升 求解器性能, IBM 整合了高速低功耗的存内计算 求解器和高精度的数字求解器的优势,提出了混合 精度架构,齐平了 CPU、GPU 的处理能力 [101],见 图 11(b). 2019年,米兰理工大学 Sun 等^[19]利用 RRAM 和负反馈阵列接连结构, 演示了一步求解 线性方程, 如图 11(c)-(e) 所示. 之后, 这种技术方 案又被扩展到求解薛定谔方程和其他代数问题^[102]. 2021年,山东大学 Feng 等^[103]提出了存内计算技术 加速浮点数尾数乘法的方法,设计实现了 32 bit 浮 点数求解器.综上所述,器件稳定性、一致性以及 阵列的非理想因素等问题仍不可忽略,依然影响着 求解器的性能表现,相应的算法和架构仍有较大改 讲空间.

5.6 其他应用场景

随机计算 (stochastic computing, SC) 是一种

低成本的计算形式^[104],如图 12(a).其工作原理是 将信息量化为随机分布的 0/1 数据流,利用数据流 中的 0/1 比例表示实际信息,由此可将乘法等运算 简化为 AND 等基本逻辑操作,相较传统计算具备 较高的误差容忍度、运算逻辑简答等优势^[104].存内 计算功能器件是实现 SC 的重要技术路线^[104,105]. 一方面,存内计算功能器件是良好的随机信号发生 器,具备不同尺度下的噪声来源,如图 12(b),(c); 另一方面,存内计算功能在实现简单逻辑计算方面 具备显著优势,且具备高集成程度、高存储密度的 特点.

物理不可克隆函数 (physically unclonable functions, PUF) 是一种利用某种物理内在机制构建的唯一性标识, 如图 12(d) 所示, 输入任意激励都 会输出唯一且不可预测的响应, 常用作信息密钥、防伪等应用场景^[106].存内计算功能器件, 特别是 RRAM 等新型 NVM 器件, 由于具备本质的不可预测的随机信号源, 是实现 PUF 的良好硬件基础^[107,108]. 相关的系统架构设计如图 12(e) 所示, 其工作原理 是利用 NVM 器件本征的不可预测、不可复制的器件波动性, 对给定外界激励信号形成不可克隆的、可重复触发的唯一响应^[107].



图 11 线性方程组求解器 (a)高精度线性方程求解器实现方法^[100]; (b)混合精度求解器架构^[101]; (c)正权重矩阵求逆^[19]; (d)求解特征向量方程 *Ax* = λ*x*^[19]; (e)混合矩阵求逆^[19]

Fig. 11. Linear equations solver: (a) The implementation of high-precision linear equation solver^[100]; (b) the mixed-precision solver architecture^[101]; (c) inverting a positive weight matrix^[19]; (d) solve eigenvector equation $Ax = \lambda x$ ^[19]; (e) inverting a mixed matrix^[19].



图 12 (a) SC 乘法工作原理^[104]; (b) RRAM 阻变过程的随机性^[104]; (c) 随机电报噪声特性; (d) 基于 PUF 的射频识别工作原理^[106]; (e) 基于 RRAM 器件的 PUF 架构^[107]; (f) HDC 分类原理^[109]; (g) 基于 NVM 器件交叉阵列的稀疏编码^[113] Fig. 12. (a) The multiplication operation realized by SC^[104]; (b) the random behavior of RRAM^[104]; (c) noise characteristics of random telegram; (d) the operating principle of PUF based radio frequency identification^[106]; (e) the architecture of RRAM based PUF^[107];

(f) classification overview with $HDC^{[109]}$; (g) sparse coding in NVM device based crossbar array^[113].

此外,超维计算 (hyperdimensional computing, HDC) 是一种新兴的计算方法,如图 12(f) 所示,通 过在超维空间计算超维向量化的特征图像的相对 距离,实现分类识别功能^[109].为方便理解,可以将 HDC 的工作原理归纳为编码、搜索两个阶段.在 编码阶段,利用算法将训练库数据的特征图像提取 出来,并编码为特征向量.在搜索阶段,将新产生 的特征向量与存储器内部的特征向量进行对比,剔 除重复的冗余信息,记忆具备显著区别的特征向 量.HDC 可以使人工智能系统对过去感知的事情 形成记忆,以便更好地完成未来类似的任务.存内 计算由于具备高密度存储和运算功能,能够同时实现 HDC 编码器和关联存储器,在实现超维计算方面具有显著优势^[110,111].利用存内计算功能器件和交叉阵列结构的矩阵结构,如图 12(g)所示,研究人员开发出新型的图像压缩功能^[112]、稀疏编码^[113]等非易型存内计算技术的多种新型应用场景.

6 结 论

存内计算技术通过在基本单元上集成存储和 运算功能,能够显著减少数据搬运,是突破传统 冯·诺依曼瓶颈和存储墙的一种新型计算范式. NVM 器件是实现存内计算的理想硬件载体,不仅 具备非挥发、低功耗等性能优势,而且可以在器件 层级实现存算功能融合,从而构建高集成度、低功 耗的存内计算硬件系统.从工艺成熟的 flash 器件 到潜力巨大的 RRAM 器件, 一脉相承的非挥发型 存内计算技术体系,在短期内具备成熟硬件开发能 力,在长期将拥有广阔的应用拓展空间,是未来计 算技术的重要发展方向. 基于 NVM 器件的存内计 算各类实现方式均拥有各自的性能优势与适用场 景,为未来计算形态的发展提供了多种可能性.总 体来看,存内模拟计算充分开发了 NVM器件的性 能潜力,在非精准运算场景下具备显著的低功耗、 高集成度的性能优势.其中向量-矩阵模式适用于 以深度学习为典型代表的数据流驱动的运算任务; 向量-向量模式则更加契合类脑计算中前馈、反馈 等多激励耦合的应用场景.相对而言,存内数字计 算的高可靠性和存算融合能力,更加兼容当前主流 的计算平台,能够作为协处理器来增强传统计算平 台处理数据密集型任务的能力. 当然, 当前的非挥 发型存内计算技术仍然面临诸多挑战.

1)存内计算运算宏.随着非挥发存内计算功 能器件的性能逐步稳定、运算模式的设计日趋完 善,构建存内计算运算宏成为当前重要的研究课 题.存内计算运算宏涉及功能定义、电路设计等方 面,既需克服模数转化问题、工艺兼容性问题等硬 件相关科学问题,也需进行运算任务分解方法、数 据分配机制等配套算法研究.

2) 系统架构设计.存内计算系统在硬件上拥 有存算融合的典型特征,在处理任务的过程中,需 在大量存储数据中进行特定的运算操作,其数据存 储格式、中间数据搬运过程等与传统计算系统存在 显著差异.因此,该系统架构的研发面临多方面的 创新需求,包括算法映射方法、多核协作机制等关 键技术,以及与传统计算体系互动互通的实现方法.

3) 硬件规模扩展与应用落地. 尽管非挥发型 存内计算技术在深度学习、线性方程求解等诸多 应用上展现出显著的性能优势. 然而, 受限当前存 内计算系统有限的硬件规模, 其计算能力、系统功 能仍十分有限, 难以承接实际的应用任务, 也就无 法与主流计算平台进行技术竞争. 扩大非挥发型存 内计算系统硬件规模、推动应用落地, 将是促进存 内计算技术良性发展的重要途径.

参考文献

- [1] Shah S H, Yaqoob I 2016 IEEE SEGE 1 381
- [2] Reinsel D, Gantz J, Rydning J 2017 Don't Focus on Big Data 1 2
- [3] Waldrop M M 2016 Nat. News 530 144
- [4] Backus J W 1978 Comm. ACM. **21** 613
- [5] McKee S A 2004 Proceedings of the 1st Conference on Computing frontiers Ischia Italy, April 14–16, 2004 p162
- [6] Indiveri G, Liu S C 2015 *Proceedings of the IEEE* 103 1379
 [7] Yang J J, Strukov D B, Stewart D R 2013 *Nature Nanotech.*
- [7] Yang J J, Strukov D B, Stewart D R 2013 Nature Nanotech.
 8 13
- [8] Chen B, Cai F, Zhou J, Ma W, Sheridan P, Lu W D 2015 *IEEE IEDM* Washington, December 7–9, 2015 p17.5.1
- [9] Zahoor F, Zulkifli T Z A, Khanday F A 2020 Nanoscale Res. Lett. 15 1
- [10] Ma Y, Du Y, Du L, Lin J, Wang Z 2020 GLSVLSI'20 Virtual Event, China, September 7–9, 2020 p265
- [11] Xu X, Luo Q, Gong T, Lv H, Long S, Liu Q, Chuang S S, Li J, Liu M 2016 *IEEE VLSI* Honolulu, April 25–27, 2016 p1
- [12] Chen A 2016 Solid State Electron. **125** 25
- [13] Tao L, Xu R, Tian T, Xiang Z, Li Y, Jin X, Ren J, Li Z, Li C 2019 *MobiSys'19* Seoul, June 17–21, 2019 p612
- [14] Lee J, Park B G, Kim Y 2019 *IEEE EDL.* 40 1358
- [15] Freitas R F, Wilcke W W 2008 IBM J. Res. Dev. 52 439
- [16] Wu W, Wu H, Gao B, Yao P, Zhang X, Peng X, Yu S, Qian H 2018 IEEE VLSI Honolulu, June 18–22, 2018 p103
- [17] Wang Z, Wu H, Burr G W, Hwang C S, Wang K L, Xia Q, Yang J J 2020 Nat. Rev. Mater. 5 173
- [18] Xiang Y, Huang P, Zhao Y, Zhao M, Gao B, Wu H, Qian H, Liu X, Kang J F 2019 *IEEE Trans. Electron. Devices* 66 4517
- [19] Sun Z, Pedretti G, Ambrosi E, Bricalli A, Wang W, Ielmini D 2019 Pro. Nat. Acad. Sci. 116 4123
- [20] Zanotti T, Puglisi F M, Pavan P 2020 IEEE Trans. Device Mater. Reliab. 20 278
- [21] Yan B, Li B, Qiao X, Xue C X, Chang M F, Chen Y, Li H 2019 Adv. Intell. Syst. 1 1900068
- [22] Schuiki F, Schaffner M, Gürkaynak F K, Benini L 2018 IEEE Trans. Comput. 68 484
- [23] McClanahan C 2010 A Survey Paper 9 1
- [24] Wang Y E, Wei G Y, Brooks D 2020 Pro. Mach. Learning Syst. 2 30
- [25] Liu S, Du Z, Tao J, Han D, Luo T, Xie Y, Chen Y, Chen T 2016 ACM/IEEE ISCA Seoul, June 18–22, 2016 p393
- [26] Sebastian A, Le Gallo M, Khaddam-Aljameh R, Elefheriou E 2020 Nature Nanotech. 15 529
- [27] Si X, Chen J J, Tu Y N, Huang W H, Wang J H, Chiu Y C, Wei W C, Wu S Y, Sun X, Liu R, Yu S 2019 *IEEE J Solid-State Circuits* 55 189
- [28] Zhou Z, Huang P, Xiang Y C, Shen W S, Feng Y L, Gao B, Wu H Q, Qian H, Liu L F, Zhang X, Liu X Y, Kang J F 2018 *IEEE IEDM* San Francisco, November 29–December 07, 2018 p20.7.1
- [29] Jerry M, Chen P Y, Zhang J, Sharma P, Ni K, Yu S, Datta S 2017 *IEEE IEDM* San Francisco, December 4–6, 2017 p6.2.1
- [30] Peng X, Chakraborty W, Kaul A, Shim W, Bakir M S, Datta S, Yu S 2020 *IEEE IEDM* San Francisco, December 10–18, 2020 p30.4.1
- [31] Xiang Y C, Huang P, Zhou Z, Han R Z, Jiang Y N, Shu Q M, Su Z Q, Liu Y B, Liu X Y, Kang J F 2019 *IEEE ISCAS* Sapporo Convention Center, May 26–29, 2019 p1

- [32] Jiang H, Huang S, Peng X, Yu S 2020 IEEE ISCAS Spain, May 17–20, 2020 p1
- [33] Khwa W S, Chen J J, Li J F, Si X, Yang E Y, Sun X, Liu R, Chen P Y, Li Q, Yu S, Chang M F 2018 *IEEE ISSCC* San Francisco, February 4–8, 2018 p496
- [34] Guo R, Liu Y, Zheng S, Wu S Y, Ouyang P, Khwa W S, Chen X, Chen J J, Li X, Liu L, Chang M F, Wei S, Yin S 2019 *IEEE VLSI* Kyoto, June 9–14, 2019 p120
- [35] Wang P, Xu F, Wang B, Gao B, Wu H, Qian H, Yu S 2018 IEEE TVLSI 27 988
- [36] Bez R, Camerlenghi E, Modelli A, Visconti A 2003 Pro. IEEE 91 489
- [37] Bayat F M, Guo X, Klachko M, Do N, Likharev K, Strukov D 2016 74th Annual Device Research Conference (DRC) Newark, June 19–22, 2016 p1
- [38] Guo X, Bayat F M, Bavandpour M, Klachko M, Mahmoodi M R, Prezioso M, Likjarev K K, Strukov D B 2017 *IEEE IEDM* San Francisco, December 4–6, 2017 p6.5.1
- [39] Han R, Huang P, Xiang Y, Liu C, Dong Z, Su Z, Liu Y, Liu L, Liu X, Kang J F 2019 *IEEE TCAS-I* 66 1692
- [40] Kim M, Liu M, Everson L, Park G, Jeon Y, Kim S, Lee S, Song S, Kim C H 2019 *IEEE IEDM* San Francisco, December 9–11, 2019 p38.3.1
- [41] Lue H T, Hsu P K, Wei M L, Yeh T H, Du P Y, Chen W C, Wang K C, Lu C Y 2019 *IEEE IEDM* San Francisco, December 9–11, 2019 p38.1.1
- [42] Tyagi V V, Buddhi D 2007 Renewable and sustainable energy reviews 11 1146
- [43] Khvalkovskiy A V, Apalkov D, Watts S, Chepulskii R, Beach R S, Ong A, Tang X, Driskill-Smith A, Butler W H, Visscher P B 2013 J. Phys. D 46 074001
- [44] Mikolajick T, Dehm C, Hartner W, Kasko I, Kastner M J, Nagel N, Moert M, Mazure C 2001 Microelectron. Reliab. 41 947
- [45] Wong H S P, Lee H Y, Yu S, Chen Y S, Wu Y, Chen P S, Lee B, Chen F, Tsai M J 2012 *Pro. IEEE* 100 1951
- [46] Huang P, Liu X Y, Chen B, Li H T, Wang Y J, Deng Y X, Wei K L, Zeng L, Gao B, Du G, Zhang X, Kang J F 2013 *IEEE TED* 60 4090
- [47] Baek I G, Lee M S, Seo S, Lee M J, Seo D H, Suh D S, Park J C, Park S O, Kim H S, Yoo I K, Chuang U I, Moon J T 2004 *IEEE IEDM* San Francisco, December 13–15, 2004 p587
- [48] Wang H, Yan X 2019 Phys. Status Solidi (RRL) 13 1900073
- [49] Rehman M M, Rehman H M M U, Gul J Z, Kim W Y, Karimov K S, Ahmed N 2020 Sci. Technol. Adv. Mat. 21 147
- [50] Zhang Y, Huang P, Gao B, Kang J K, Wu H Q 2020 J. Phys. D 54 083002
- [51] Shi T, Wang R, Wu Z, Sun Y, An J, Liu Q 2021 Small Structures 2 2000109
- [52] Pedretti G, Ambrosi E, Ielmini D 2021 IEEE IRPS Monterey, March 21–25, 2021 p1
- [53] Guo W, Fouda M E, Eltawil A M, Salama K N 2021 Front. Neurosci. 15 212
- [54] Park J, Kwak M, Moon K, Woo J, Lee D, Hwang H 2016 *IEEE EDL* 37 1559
- [55] Li W, Sun X, Jiang H, Huang S, Yu S 2021 IEEE ESSCIRC Grenoble, September 13–22, 2021 p79
- [56] Sung C, Padovani A, Beltrando B, Lee D, Kwak M, Lim S, Larcher L, Marca V D, Hwang H 2019 IEEE J Electron Devi. 7 404
- [57] Han L X, Xiang Y C, Huang P, Yu G H, Han R Z, Liu X Y,

 Kang J F 2021 *IEEE IRPS* Monterey, March 21–25, 2021 p1
 Liao Y, Wu H, Wan W, Zhang W, Gao B, Wong H S P, Qian H 2018 *IEEE VLSI* Honolulu, June 18–22, 2018 p31

- [59] Ambrogio S, Balatti S, Milo V, Carboni R, Wang Z, Calderoni A, Ramaswamy N, Ielmini D 2016 *IEEE VLSI* Honolulu, June 14–16, 2016 p1
- [60] Ma W, Zhou Z, Zhu D, Liu L 2016 Electron. Lett. 52 1073
- [61] Chen Z, Gao B, Zhou Z, Huang P, Li H, Ma W, Zhu D, Liu L, Liu X, Kang J F 2015 *IEEE IEDM* Washington, December 7–9, 2015, p17.7.1
- [62] Rosezin R, Linn E, Kugeler C, Bruchhaus R, Waser R 2011 IEEE EDL 32 710
- [63] Gao S, Zeng F, Wang M, Wang G, Song C, Pan F 2015 Sci. Rep. 5 15467
- [64] Xie L, Du Nguyen H A, Yu J, Kaichouhi A, Taouil M, AlFailakawi M, Hamdioui S 2017 *IEEE ISVLSI* Bochum, July 3–5, 2017 p176
- [65] Gao L, Alibart F, Strukov D B 2013 *IEEE T Nanotechnol.* 12 115
- [66] Li H, Gao B, Chen Z, Zhao Y, Huang P, Ye H, Liu L, Liu X, Kang J F 2015 *Sci. Rep.* 5 1
- [67] LeCun Y, Bengio Y, Hinton G 2015 Nature 521 436
- [68] Van Ooyen A, Nienhuis B 1992 Neural Networks 5 465
- [69] Tsai H, Ambrogio S, Narayanan P, Shelby R M, Burr G W 2018 J. Phys. D 51 283001
- [70] Leijnen S, Veen F 2020 Multidisciplinary Digital Publishing Institute Proceedings 47 9
- [71] Huang P, Zhou Z, Zhang Y, Xiang Y, Han R, Liu L, Liu X, Kang J 2019 APL Mater. 7 081105
- [72] Feng Y, Huang P, Zhao Y, Shan Y, Zhang Y, Zhou Z, Liu L, Liu X, Kang J F 2021 *IEEE EDL* 42 1168
- [73] Peng X, Liu R, Yu S 2019 IEEE ISCAS Sapporo Convention Center, May 26–29, 2019 p1
- [74] Xiang Y, Huang P, Han R, Li C, Wang K, Liu X, Kang J F 2020 IEEE TED 67 2329
- [75] Peng X, Huang S, Luo Y, Sun X, Yu S 2019 IEEE IEDM San Francisco, December 9–11, 2019 p32.5.1
- [76] Prabhu N L, Raghavan N 2021 IEEE Access 9 168093
- [77] Hsu T H, Lue H T, Hsu P K, Yeh T H, Du P Y, Lee G R, Chu C J, Wang K C, Liu C Y 2020 *IEEE IEDM*, San Francisco, December 10–18, 2020 p6.3.1
- [78] Chi P, Li S, Xu C, Zhang T, Zhao J, Liu Y, Wang Y, Xie Y 2016 ACM SIGARCH Computer Architecture News 44 27
- [79] Cheng M, Xia L, Zhu Z, Cai Y, Xie Y, Wang Y, Yang H 2017 ACM/EDAC/IEEE DAC Austin, June 19–23, 2017 p1
- [80] Du C, Ma W, Chang T, Sheridan P, Lu W D 2015 Adv. Funct. Mater. 25 4290
- [81] Huang P, Li Z, Dong Z, Han R, Zhou Z, Zhu D, Liu L, Liu X, Kang J F 2019 ACS Appl. Electronic Mater. 1 845
- [82] Wang Z, Zeng T, Ren Y, Lin Y, Xu H, Zhao X, Liu Y, Ielmini D 2020 Nature Comm. 11 1
- [83] Li X, Tang J, Zhang Q, Gao B, Yang J J, Song S, Wu W, Zhang W, Yao P, Deng N, Xie Y, Qian H, Wu H 2020 *Nature Nanotech.* 15 776
- [84] Lashkare S, Chouhan S, Chavan T, Bhat A, Kumbhaew P, Ganguly U 2018 IEEE EDL 39 484
- [85] Milo V, Pedretti G, Carboni R, Calderoni A, Ramaswamy N, Ambrogio S, Ielmini D 2016 *IEEE IEDM* San Francisco, December 3–7, 2016 p16.8.1
- [86] Milo V, Ielmini D, Chicca E 2017 IEEE IEDM San Francisco, December 04–06, 2017 p11.2.1
- [87] Larkum M 2013 Trends Neurosci. 36 141
- [88] Zhou Z, Liu C, Shen W, Dong Z, Chen Z, Huang P, Liu L,

Liu X, Kang J F 2017 Nanoscale Res. Lett. 12 1

- [89] Majdabadi M M, Shamsi J, Shokouhi S B 2021 Analog Integr. Circ. S 107 249
- [90] Tang H, Kim H, Kim H, Park J 2019 IEEE T. Biomed. Circ. S 13 1664
- [91] Borghetti J, Snider G S, Kuekes P J, Yang J J, Stewaert D R, Williams R S 2010 Nature 464 873
- [92] Talati N, Gupta S, Mane P, Kvatinsky S 2016 IEEE T. Nanotechnol. 15 635
- [93] Huang P, Kang J, Zhao Y, Chen S, Han R, Zhou Z, Chen Z, Ma W, Li M, Liu L, Liu X 2016 Adv. Mater. 28 9758
- [94] Shen W, Huang P, Fan M, Han R, Zhou Z, Gao B, Wu H, Qian H, Liu L, Liu X, Zhang X, Kang J F 2019 *IEEE EDL* 40 1538
- [95] Shen W, Huang P, Wang X, Feng Y, Xu W, Gao B, Wu H, Qian H, Liu L, Zhang X, Kang J F 2020 *IEEE EDTM* Penang, April 6–21, 2020 p1
- [96] Rajendran B, Cheek R W, Lastras L A, Franceschini M M, Breitwisch M J, Schrott A G, Li J, Montoye R K, Chang L, Lam C 2011 *IEEE IMW* Monterey, May 22–25, 2011 p1
- [97] Yan B, Li Z, Chen Y, Hai L 2016 NVMTS Pittsburgh, November 17–19, 2016 p1
- [98] Grossi A, Vianello E, Zambelli C, Royer P, Noel J P, Giraud B, Perniola L, Olivo P, Nowak E 2018 *IEEE T. VLSI Syst.* 26 2599
- [99] Yang H Z, Huang P, Han R Z, Xiang Y C, Feng Y, Gao B, Chen J Z, Liu L F, Liu X Y, Kang J F 2020 *IEEE SNW* Honolulu, June 13–14, 2020 p29
- [100] Zidan M A, Jeong Y J, Lee J H, Chen B, Huang S, Kushner M J, Lu W D 2018 Nat. Electron. 1 411

- [101] Gallo M L, Sebastian A, Mathis R, Manica M, Giefers H, Tuma T, Bekas C, Curioni A, Eleftheriou E 2018 Nat. Electron. 1 246
- [102] Sun Z, Ambrosi E, Pedretti G, Bricalli A, Ielmini D 2020 IEEE TED 67 1466
- [103] Feng Y, Chen B, Liu J, Sun Z H, Hu H Y, Zhang J Y, Zhan X P, Chen J C 2021 *IEEE IEDM* San Francisco, December 13–15, 2021 p12.1.1
- [104] Shen W, Huang P, Fan M, Zhao Y, Feng Y, Liu L, Liu X, Kang J F 2020 IEEE TED 68 103
- [105] Suri M, Querlioz D, Bichler O, Palma G, Vianello E, Vuillaume D, Gamrat C, DeSalvo B 2013 IEEE TED 60 2402
- [106] Devadas S, Suh E, Paral S, Tom Z, Vivek K 2008 IEEE International Conference On RFID Las Vegas, April 16–17, 2018 p58
- [107] Liu R, Wu H, Pang Y, Qian H, Yu S 2016 IEEE HOST McLean, May 3–5, 2016 p13
- [108] Mahmoodi M R, Nili H, Strukov D B 2018 IEEE VLSI Honolulu, June 18–22, 2018 p99
- [109] Ge L, Parhi K K 2020 IEEE Circ. Syst. Mag. 20 30
- [110] Karunaratne G, Le Gallo M, Cherubini G, Cherubini G, Benini L, Rahimi A, Sebastian A 2020 Nat. Electron. 3 327
- [111] Liu J, Ma M, Zhu Z, Wang Y, Yang H 2019 IEEE ICECS Genoa, November 27–29, 2019 p4
- [112] Xu J, Feng D, Hua Y, Tong W, Liu J, Li C, Zhou W 2017 IEEE ICCD Boston Area, November 5–8, 2017 p573
- [113] Sheridan P M, Cai F, Du C, Ma W, Zhang Z, Lu W D 2017 Nat. Nanotechnol. 12 784

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Non-volatile memory based in-memory computing technology

Zhou Zheng Huang Peng Kang Jin-Feng[†]

(School of Integrated Circuits, Peking University, Beijing 100871, China)
(Received 5 March 2022; revised manuscript received 10 June 2022)

Abstract

By integrating the storage and computing functions on the fundamental elements, computing in-memory (CIM) technology is widely considered as a novel computational paradigm that can break the bottleneck of Von Neumann architecture. Nonvolatile memory device is an appropriate hardware implementation approach of CIM, which possess significantly advantages, such as excellent scalability, low consumption, and versatility. In this paper, first we introduce the basic concept of CIM, including the technical background and technical characteristics. Then, we review the traditional and novel nonvolatile memory devices, flash and resistive random access memory (RRAM), used in non-volatile based computing in-memory (nvCIM) system. After that, we explain the operation modes of nvCIM: in-memory analog computing and in-memory digital computing. In addition, the applications of nvCIM are also discussed, including deep learning accelerator, neuromorphic computing, and stateful logic. Finally, we summarize the current research advances in nvCIM and provide an outlook on possible research directions in the future.

Keywords: in-memory computing, non-volatile memory, flash, resistive random access memoryPACS: 85.35.-p, 07.05.Mh, 84.35.+i, 85.30.TvDOI: 10.7498/aps.71.20220397

[†] Corresponding author. E-mail: kangjf@pku.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

面向感存算一体化的光电忆阻器件研究进展

单旋宇 王中强 谢君 郑嘉慧 徐海阳 刘益春

Recent progress in optoelectronic memristive devices for in-sensor computing Shan Xuan-Yu Wang Zhong-Qiang Xie Jun Zheng Jia-Hui Xu Hai-Yang Liu Yi-Chun 引用信息 Citation: Acta Physica Sinica, 71, 148701 (2022) DOI: 10.7498/aps.71.20220350 在线阅读 View online: https://doi.org/10.7498/aps.71.20220350 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

基于忆容器件的神经形态计算研究进展

Research progress of neuromorphic computation based on memcapacitors 物理学报. 2021, 70(7): 078701 https://doi.org/10.7498/aps.70.20201632

N型局部有源忆阻器的神经形态行为

Neuromorphic behaviors of N-type locally-active memristor 物理学报. 2022, 71(5): 050502 https://doi.org/10.7498/aps.71.20212017

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

专题: 面向类脑计算的物理电子学

面向感存算一体化的光电忆阻器件研究进展*

单旋宇1) 王中强1)2)† 谢君1) 郑嘉慧1) 徐海阳1)2) 刘益春1)2)

1) (东北师范大学,紫外光发射材料与技术教育部重点实验室,长春 130024)

2) (物理学国家级实验教学示范中心 (东北师范大学), 长春 130024)

(2022年2月28日收到; 2022年3月28日收到修改稿)

脑启发神经形态计算系统有望从根本上突破传统冯·诺依曼计算机系统架构瓶颈,极大程度地提升数据 处理速度和能效.新型神经形态器件是构建高能效神经形态计算的重要硬件基础.光电忆阻器作为新兴的纳 米智能器件,因具备整合光学感知、信息存储和逻辑计算等功能特性,被认为是发展类脑视觉系统的重要备 选.本文将综述面向感存算功能一体化的光电忆阻器研究进展,包括光电忆阻材料与机制、光电忆阻器件与 特性、感存算一体化功能及应用等.具体将根据机制分类介绍光子-离子耦合型和光子-电子耦合型光电忆阻 材料,根据光电忆阻特性调节方式介绍光电调制型和全光调制型光电忆阻器件,根据感存算一体化功能介绍 其在认知功能模拟、光电逻辑运算、神经形态视觉功能、动态探测与识别等方面的应用.最后总结光电忆阻 器的主要优势以及所面临的挑战,并展望光电忆阻器的未来发展.

关键词: 感存算一体化, 光电忆阻器, 神经形态计算 **PACS**: 87.18.Sn, 73.40.Rw, 85.25.Hv, 85.60.-q

DOI: 10.7498/aps.71.20220350

1 引 言

随着现代信息技术的持续高速发展,计算机系 统开始进入"云计算"和"大数据"时代,海量的信息 存储与数据处理要求计算机软硬件具有高速度、低 功耗及智能化等特点.然而,传统的冯·诺依曼计算 架构由于存储单元与运算单元结构分立,而且两者 沿着各自的目标发展,导致不同硬件之间存在较大 性能差距与集成方式问题,难以高效完成非结构化 数据的深度处理^[1,2].相比之下,生物大脑具有高效 率、低功耗、并行运算、自主认知等优点.脑启发的 神经形态计算成为突破冯·诺依曼瓶颈的重要方向, 是发展智能计算机的前沿交叉领域^[3-5].因此,开 发类脑神经形态器件单元是发展类脑智能系统的 核心,也是构筑类脑智能计算机的必要途径^[6]. 外的第4种基本电子元件,其具有电阻连续调节的 记忆特性,与大脑认知的突触可塑性功能高度相似77. 相较于传统电子元件, 忆阻器具有自主学习能力, 被视为发展类脑神经形态硬件单元的理想选择[8,9]. 同时, 忆阻器具有运行功耗低、读写速度快、集成 密度高等优势,在信息存储、逻辑运算、类脑计算 等领域展现了极大应用潜力[10,11]. 从广义上讲, 忆 阻器根据其电阻转变机理可以划分为离子-电子耦 合型忆阻器、铁电型忆阻器、相变型忆阻器、自旋 型忆阻器、电荷填充型忆阻器等多种类型[12,13].本 文将着重综述离子-电子耦合型忆阻器,其工作机 理是: 电场诱导忆阻材料中的离子发生迁移/扩散 等过程,导致缺陷态的数量变化或重新分布,进而 实现对电子输运性质的调制.从可迁移离子类型 看,离子-电子耦合型忆阻材料主要分为阴离子化合 价变化 (valence change memory, VCM) 和阳离子 电化学金属化 (electrochemical metallization me-

忆阻器 (memristor) 是除电阻、电容、电感之

* 科技部国际科技合作重点专项 (批准号: 2018YFE0118300) 和国家自然科学基金 (批准号: 11974072, U19A2091) 资助的课题.

© 2022 中国物理学会 Chinese Physical Society

[†] 通信作者. E-mail: wangzq752@nenu.edu.cn

mory, ECM)两种类型,包括金属氧化物、硫化物、碳基材料^[14-16],有机材料^[17],有机-无机杂化钙钛矿^[18], 生物兼容材料、新型二维材料等多种材料体系^[19]. 目前,忆阻器在构筑人工神经突触和人工神经元等 神经形态硬件方面得到了广泛发展^[20,21].通过对忆 阻器中离子动力学过程的调控,研究者实现了对多 种神经突触功能和神经元功能的仿生模拟,包括长 时/短时突触可塑性、脉冲时序依赖可塑性(STDP)、 脉冲频率依赖可塑性(SRDP)、树突滤波/整合功 能、神经元信号整合触发功能等等^[9,21-23].

基于忆阻器的数据信息存储以电阻值的形式 呈现,这一存储原理本质上可以通过基尔霍夫定律 等物理原理实现对数据信息的处理,从而其在发展 存储计算一体化方面具有独特优势.目前,国际上 多个研究组已经在忆阻器存算一体化研究方面取 得了诸多进展,包括美国惠普实验室、意大利米兰 理工大学、华中科技大学等^[13,24,25].进一步,在忆阻 器的存储与运算功能基础上,引入感知功能,研制 兼具感知、存储、运算功能的感存算一体化将有助 于发展更加智能化、高能效的计算系统,有望开启 人工智能的新时代^[26].因此,忆阻器感存算一体化 成为当前该领域的研究前沿.从感知功能角度,多 种传感功能均可以与忆阻器相集成,包括触觉传感功能、嗅觉传感功能、视觉传感功能等^[27-29].其中,视觉系统是大脑获取外界信息的主要途径^[10],本 文将主要综述面向人工视觉系统的忆阻器感存算 一体化领域的研究进展.将具有更多操作维度的光 场引入忆阻器构成"光电忆阻器",有望突破传统 CMOS 图像处理技术的瓶颈,集感、存、算功能于 一体,避免无效的数据传输同时解决计算效率低、 功耗高等缺点,从而提高系统效能^[30].同时,通过 光子-电子-离子的多维、多源、多尺度耦合能够丰 富离子调控手段,有望推进光控离子传输、光电逻 辑运算、人工视觉系统等领域的发展.

本文综述了基于感存算一体化的光电忆阻器 件的最新进展,首先介绍光子-离子耦合型和光子-电子耦合型两种光电忆阻材料和忆阻机制;然后依 据忆阻器的光电特性,介绍光电调制型和全光调制 型光电忆阻器件;接着综述了感存算一体化光电忆 阻器件在认知功能模拟、光电逻辑运算、神经形态 视觉、动态探测与识别等方面的应用方向和研究进 展(见图 1).最后在材料制备、器件工艺以及系统 架构等方面总结光电忆阻器发展面临的主要挑战 并展望其未来发展前景.



图 1 综述框架示意图,主要包括光电忆阻材料与机制,忆阻器件与特性、功能与应用三部分^[14,19,30,41,44,64,72,95,96] Fig. 1. Schematic illustration of optoelectronic memristor in terms of materials and mechanism, device and characteristics, functions and applications^[14,19,30,41,44,64,72, 95,96].

2 光电忆阻材料与机制

材料是器件的基础,光电忆阻材料的优选在构 筑高性能光电忆阻器方面十分重要.光电忆阻器不 仅要求电学信号能够调制器件阻态,而且光学信号 同样要有调制作用^[31].在光电忆阻材料选择中,如 何引入光子作用实现光子-离子耦合、光子-电子耦 合乃至三者协同作用是发展光电忆阻材料的根本 问题.一方面,多数光电忆阻材料自身具有良好的 光敏特性,能够有效利用光学信号调制忆阻特性, 例如多种半导体材料^[14,15]、光致变色材料^[32]、有机-无 机杂化钙钛矿材料^[18]、新型二维材料^[19]等.另一方 面,一些光电忆阻材料则需要通过复合材料体系、器 件结构设计等方案实现其光电忆阻特性^[26,31].本节 将围绕光电忆阻材料的优选与设计进行介绍,包括 光子-离子耦合型和光子-电子耦合型两个主要方面.

2.1 光子-离子耦合型光电忆阻材料

光子-离子耦合型光电忆阻器材料工作机理主 要是利用光学信号调制离子动力学特性,从而实现 光电忆阻器的阻态调节,其中主要包括光调制离子 化合价和光调制离子迁移型两种类型.

光致变色材料是一类通过光辐照诱导材料结 构发生改变的材料体系.其中,无机光致变色材料 的颜色改变通常伴随着氧化-还原反应以及离子化 合价变化,比如 MoO₃, WO₃等半导体材料^[30,33]. 基于光致变色过程中的光子-离子耦合作用,无机 光致变色材料能够通过光学信号诱导离子价态发 生变化,从而调节材料微观结构和电子输运特性, 是光电忆阻材料的重要备选体系之一. 香港理工大 学 Chai 研究组^[30] 基于 MoO_x 材料研制出了具有 光调制功能的光电阻变式存储器 (ORRAM), 器件 具有简单的 Pd/MoO_r/ITO 三明治结构 (图 2(a)). 在 365 nm 紫外光照射下, MoO_x基光电忆阻器件 能够从初始高阻态转变为低阻态,实现光信号对忆 阻器的存储开启操作;撤去光信号后仍然保持在低 阻态,展示出良好的信息保持特性,反之,器件在 负向电压作用下可以由低阻态转变至高阻状态,从 而获得电学信号对器件的信息关闭操作 (图 2(b)). 如图 2(c) 所示, MoO_r基光电忆阻器的工作机理 是:在光照过程中,光生空穴可以与薄膜吸附的水



图 2 (a) MoO_x光电阻变式存储器的结构示意图; (b) Pd/MoO_x/ITO 器件的脉冲开关特性; (c) 器件阻变机制示意图^[30]; (d) 基于 PDR1A 材料光学忆阻器结构示意图; (e) ITO/ZnO/PDR1 A/Al 结构器件的可逆电阻调制过程; (f) PDR1A 分子化学结构的示意图^[32]

Fig. 2. (a) Structural illustration of the MoO_x ORRAM; (b) pulse-switching characteristics of Pd/MoO_x/ITO device; (c) schematic of switching mechanism^[30]; (d) schematic of PDR1A based optical memristor; (e) conductance modulation of ITO/ZnO/PDR1A/Al device; (f) schematic of the PDR1A molecules^[32].

分子反应产生质子 (H⁺). 随后质子和光生电子会 导致钼离子价态从 Mo⁶⁺价转变为 Mo⁵⁺价, 形成 H_yMoO_x^[33-35]. 随着 H_yMoO_x的不断生成, 薄膜颜 色发生明显变化, 同时器件逐渐从高阻态转变为低 阻态. 在器件关闭过程中, 电场驱动质子向 Pb 电 极迁移, 钼离子的化学状态从 Mo⁵⁺价恢复到 Mo⁶⁺ 价, 器件从低阻态回到高阻态.

有机光致变色材料也是一类易于实现光电忆 阻行为的材料体系,其机理为光诱导有机分子结构 发生变化,如光诱导材料顺反式异构或开闭环调节[17,32]. 英国赫尔大学 Kemp 研究团队^[17]利用有机光致变 色材料光活性偶氮苯聚合物 PDR1 A(分散红1丙 烯酸酯)与 ZnO 纳米棒相结合,构建了可逆光开 关忆阻器. 如图 2(d) 所示, 器件结构包括 Al 顶电 极、忆阻层和 ITO 底电极. 实验结果表明, 圆偏振 光和线偏振光多次交替照射器件,聚合物 PDR1A 连续发生可逆顺反光化学异构化^{36]}.其中圆偏振 光照射过程使得 PDR1A 材料膨胀, 薄膜整体厚度 增大约 30%, 器件电阻升高; 相反当施加线偏振光 时,聚合物材料收缩,薄膜厚度下降并恢复到初始 状态,整体表现为电阻降低;由此实现对器件高阻 态 (HRS) 和低阻态 (LRS) 的可逆光调制 (图 2(e)). 除此之外,南京邮电大学 Huang 研究组^[32] 基于 BMThCE(二芳烯烷烃衍生物)制备了忆阻器件, 紫外和可见光的照射下 BMThCE 可在开环状态 o-BMThCE 和闭环状态 c-BMThCE 之间发生可 逆转换,从而实现了光辐照对忆阻器在一次写入多 次读取特性与双极性阻变特性之间的切换. 闭环状 态 c-BMThCE 有着更高的 HOMO 能级, 更有利 于空穴的注入和传输. 对于开环状态 o-BMThCE 器件, 施加正偏压时空穴从 ITO 电极注入并填充 临近陷阱从而形成导电细丝. 施加反向偏压时, 注 入的空穴仍然被困在深捕获位点中,器件无法关闭 到高阻态.紫外光照射后,闭环状态 c-BMThCE 和 ITO 之间的势垒降低, 捕获的空穴在反向电压作 用下能够释放,器件导电细丝断裂并恢复到高阻态 从而实现双极性阻变行为. 尽管光信号并没有直接 诱导器件电阻转变,但光信号对忆阻器特性产生了 影响,也为光电忆阻材料选择提供了参考.

电场作用诱导离子迁移是通常忆阻器产生阻变的离子动力学机制.利用光学信号影响离子迁移能力调制忆阻器电学特性,从而实现光子-离子耦合型光电忆阻器也是一种常见的研究途径^[37,38].有

机-无机杂化钙钛矿材料是沿该研究途径构建光电 忆阻器的典型代表,包括 CH₃NH₃PbI₃ (MAPbI₃), CH₃NH₃PbBr₃ (MAPbBr₃), 与 CH[NH₂]₂PbI₃ (FA PbI₃) 等^[39-43]. 有机-无机杂化钙钛矿材料由于其 独特的光电特性被广泛应用于多种光电器件,例如 发光二极管、光探测器、晶体管以及忆阻器[44-48]. 以 MAPbI3 材料为例, MAPbI3 基忆阻器的阻变机 制通常归因于电场诱导碘离子 (I-) 迁移, 产生大量 碘空位并形成导电通道^[49].美国密歇根大学Lu研 究组^[49]利用 MAPbI₃ 材料设计了光遗传学启发的 光电忆阻器件,其工作机理归因于光学信号对碘离 子迁移的抑制作用 (图 3(a)). 如图 3(b) 所示, 光辐 照过程 (光强为 1.29 µW/cm²) 可以增大 MAPbI₃ 材料中碘空位的形成能,从而抑制电场引起的碘空 位和碘离子形成并促进他们的自发湮灭 50. 由此 实现光信号对电导的抑制操作. 另一方面, 东北师 范大学 Liu 研究团队^[18] 发现在更高强度的光辐照 条件下 (250 mW/cm²), 光信号能够诱导 MAPbI₃ 材料发生晶格膨胀,降低碘离子的迁移势垒,从而 辅助电场驱动离子迁移. 据此, 他们提出了一种温 和的光辅助电初始化方案,极大地降低了忆阻器初 始化电压,提升了忆阻器性能.该方案也为光电忆 阻器的构建和性能优化提供了一种简单有效的方 法. 与之类似, 韩国高丽大学 Wang 研究团队 [51] 研 制了基于 Ag/CH₃NH₃PbI₃ (OHP)/ITO 垂直结构 的光突触器件. 该器件通过光照能够降低突触可塑 性开启阈值,从而模拟了多巴胺对突触活性的增强 作用. 如图 3(g) 所示, 单独的电学输入脉冲或光照 刺激下,器件仅表现出短时增强现象 (STP). 同时 施加光电信号可以得到长时增强 (LTP) 行为. 相 应的光电调节机制可以归结为:施加光照时产生的 光生电场与施加正偏压时的外界电场方向相同,可 以降低碘空位的激活能进而促进电场下的迁移过 程,其中增强光照强度可以引起 OHP 薄膜材料结 构转变或键合强度减弱,进一步降低激活能.

2.2 光子-电子耦合型光电忆阻材料

光子-电子耦合型光电忆阻材料工作机理主要 是利用光学信号调制材料的电子传输特性,并最终 实现对光电忆阻器电导的调节.光电导效应是半导 体材料中常见的光子-电子耦合效应.因此,多种半 导体材料都具备用于光电忆阻器的潜力,包括氧化 物半导体^[52,53]、Si 纳米晶材料^[54]、杂化钙钛矿材料^[55]、



图 3 (a) 基于 MAPbI₃ 材料的平面结构器件示意图; (b) 光照抑制 V_{I}/V_{I}^{\times} 形成加速 V_{I}/V_{I}^{\times} 湮灭^[49]; (c) MAPbI₃ 忆阻器结构图; (d), (e) 光照强度对 MAPbI₃ 器件开启电压和过充电流的影响^[18]; (f) Ag/CH₃NH₃PbI₃ (OHP)/ITO 结构光突触器件示意图; (g) 器 件在光电脉冲刺激下的响应; (h) 响应幅值随照射时间、频率和强度的变化^[51]

Fig. 3. (a) Schematic of MAPBI₃ based planar device; (b) light illumination inhibits the formation and accelerate the annihilation of $V_{\rm I}/V_{\rm I}^{\times [49]}$; (c) structural illustration of MAPbI₃ based memristor; (d), (e) the variations of $V_{\rm Forming}$ and $I_{\rm OV}$ (overshoot current) with light intensity^[18]; (f) structural illustration of Ag/CH₃NH₃PbI₃ (OHP)/ITO optoelectronic memristor; (g) current response under the stimulation of electrical and optical pulse; (h) current response depending on exposure time, frequency and intensity^[51].

新型二维材料^[56]等.但不同于常见的光电导现象, 光电忆阻器中光引起的电导变化需要在一定时间 内具有非易失性,而这通常与光电忆阻材料的缺陷 态紧密相关^[54].进一步,一些光电忆阻器还需要通过 构筑同质/异质结结构来获得长时光电导特性^[31,56].

氧化物半导体材料是一类代表性的光子-电子 耦合型光电忆阻器材料^[31].常见的氧化物光电忆阻 器工作机理是:借助光信号改变界面缺陷态捕获电 子的数量,从而调制材料费米能级或电极与介电层 界面处的势垒,进而改变光电忆阻器的阻值^[56-58]. 中国科学院宁波材料技术与工程研究所 Li 研究组^[14] 提出了基于 ITO/CeO_{2-x}/AlO_y/Al 结构的光电忆 阻器件.由于氧化铈和铝层表面费米能级差异,界 面处会形成肖特基势垒^[59,60].在 ITO 上施加正偏 置电压时,电子会在电场驱动下向 ITO 方向移动, 而带正电的氧空位 (Vo²⁺)向 CeO_{2-x}/AlO_y 界面处 迁移. 如图 4(b)所示,这些带正电的氧空位能够有 效降低界面势垒宽度.当施加光照时,临近 AlO_y 处 CeO_{2-x}层捕获的电子进一步被激发,并在界面处 留下更多带正电的氧空位,使得肖特基势垒宽度变 薄.因此,光照可以增强载流子输运过程,从而提 升 ITO/CeO_{2-x}/AlO_y/Al结构器件^[14,61]电导率. 与之类似,研究者又设计出了多种氧化物异质结构 用以研制光电忆阻器,包括 ITO/Nb:SrTiO₃异质 结^[15]、CeO_x/ZnO 异质结^[31]等.同时,光响应波长 在一定程度上能够通过设计不同的氧化物异质结 构进行调节.

硅纳米晶 (Si nanocrystals, Si NCs) 材料也具 有良好的光电导特性, 且其表面具有丰富的缺陷 态. 浙江大学 Yang 研究组^[54]利用硼掺杂的 Si NCs 材料构建了 ITO/Si NCs/Al 结构的光电忆阻突触器件,该器件的光响应波段涵盖紫外光至红外光范围.由于硅纳米晶表面含有大量的悬挂键,这些缺陷态能够捕获与释放光生电子,从而实现光学信号对器件电导的调制过程^[62-64].如图 4(d) 所示,光照条件下, Si NCs 的价带电子被激发至导带,引起

器件电导增大;撤去光信号后,一部分光生电子立 即与空穴复合,而另一部分电子则被悬挂键等缺陷 态所捕获.被捕获的电子需要通过热波动释放到导 带并最终与空穴复合,整个过程需要相对较长时 间.结果表明该器件具有一定的长时记忆特性,也 为模拟突触可塑性功能提供了基础.



图 4 (a) ITO/CeO_{2-x}/AlO_y/Al结构光电突触示意图; (b) 器件阻变特性机制图^[14]; (c) 生物突触和硅纳米晶器件结构图; (d) 硅 纳米晶能带结构和载流子输运示意图^[54]; (e) 光调制 BP@PS 忆阻器示意图; (f) BP@PS 器件阻变机制的能带模型^[66]; (g) 基于单 层 MoS₂ 的忆阻突触器件; (h) MoS₂/p-Si 结阻变示意图^[56]

Fig. 4. (a) Schematic diagram of optoelectronic synapse with $ITO/CeO_{2-x}/AlO_y/Al$ structure; (b) schematic energy band diagram demonstrating memristive characteristics^[14]; (c) schematic of biological synapse and Si-NC-based device; (d) schematic illustration of the band structure and charge carrier transport of Si NCs^[54]; (e) schematic of light modulation BP@PS memristor; (f) energy band diagram explaining RS mechanism^[66]; (g) schematic of memristive synapse based on monolayer MoS₂; (h) schematic illustration of the resistive switching^[56].

二维材料具有强光-物质耦合作用,因此在新 型光电神经形态器件方面极具潜力[65].中国科学 院深圳先进技术研究院 Wang 研究组^[66] 基于黑磷 和聚苯乙烯 (PB) 材料构建了多色光调节的透明忆 阻器 (图 4(e)). 对器件分别施加近红外光 (785 nm), 绿光 (500 nm), 紫外光 (380 nm) 时, 器件的关闭 电压逐渐降低,开关比逐渐增大.这种阻变性能的 提升可以归因于捕获位点俘获光生电子后引起的 肖特基势垒升高, 如图 4(f) 所示. 华中科技大学 Guo 研究团队^[56]基于单层 n-MoS₂/p-Si 异质结构搭建 了超薄忆阻突触 (图 4(g)). 当施加连续光脉冲时 (0.11 mW/cm², 1 s), 器件由于光生电子-空穴出现 缓慢的电导上升.反向的电压脉冲信号 (-8 V, 5 ms) 可以逐渐降低器件电导,从而实现光增强电抑制调 节过程 (图 4(h)). 其中界面处自然形成的 SiO₂ 层 直接影响器件的阻变行为,当施加负脉冲时 MoS₂ 中的电子转移到界面处并被 SiO2 层的捕获位点捕 获,器件电导下降.通过改变光脉冲/电脉冲的频 率,可以实现短时和长时增强/抑制行为.该超薄 忆阻器件适用于发展光电协同调控的神经形态功 能模拟. 除此之外, 多个研究组基于 WSe,[67], SnS[10] 等二维材料研制出了性能优异的光电忆阻器件.

3 光电忆阻器件与特性

光电忆阻器件对比于传统电学忆阻器可以将 光信号和电信号的优势相结合,有着高能量效率、 低电流串扰等特点^[68-71].根据器件的调制信号类 型,可将光电忆阻器件分为光电调制型和全光调制 型器件两类^[61,63].对于光电调制型器件,器件电导 调控过程通常需要光信号和电信号的共同作用^[49,51]. 而全光调制型器件可以在同一单元实现可逆的光 调控过程,有效降低操作的复杂程度,为实现高集 成度的神经形态计算系统提供硬件基础^[3,72].

3.1 光电调制型忆阻器件

对于光电调制型忆阻器件,光学信号和电学信号都可以对忆阻材料的电子输运特性产生影响.其中光辐照过程可以直接作为刺激信号改变器件的电导状态.福州大学 Li 团队和汉阳大学 Kim 团队^[55]合作研究了基于全无机钙钛矿纳米片材料的 Au/CH₃NH₃PbI₃/硫氰酸亚铜层/PEDOT:PSS/ITO

光电突触器件,器件结构如图 5(a) 所示. 施加紫外 光信号可以明显提升器件电流 (约 6.9 µA); 在撤 去光信号后,器件电路缓慢下降并可以长时间保持 在中间状态 (图 5(b)). 光信号主要对器件电导起 增强作用,而电导可逆调控则需要光信号和电信号 的协同作用. 如图 5(c) 所示, 在暗态下对器件施加 20个负电压脉冲 (-1.5 V, 100 ms) 时, 器件电流逐 渐减小进而实现光开启电关闭过程.紫外光照射 时,钙钛矿纳米片产生电子空穴对分离.PEDOT: PSS 和硫氰酸亚铜层界面可以捕获电子导致势垒 变化,从而实现器件阻值转变[55].其中光照强度直 接决定电子空穴对分离效率,进而影响 PEDOT:PSS 和硫氰酸亚铜层界面电荷捕获效率.相同照射时间 下,光强越强器件电流上升越快.光学信号不仅可 以直接调制光电忆阻器件的电导还可以作为门控 信号改变器件的忆阻特性. 东北师范大学 Liu 研究 团队[16] 基于氧化石墨烯-氧化钛纳米复合材料体 系构建新型光调控阻变存储器 (图 5(d)). 该器件 可以通过调整紫外光照射时间和氧化钛浓度来精 确控制薄膜中氧化石墨烯的光催化还原过程.紫外 光照射下,氧化钛纳米颗粒附近的氧化石墨烯会发 生光催化还原反应, 官能团数量减少进而生成还原 氧化石墨烯.相比较于传统的电初始化过程,该方 法有效抑制还原氧化石墨烯的过度生长,降低器件 随机性,提升阻变存储性能.如图 5(e)所示,经过 紫外光照射 15 min 后,器件不再需要电初始化过 程,有效抑制了过冲电流,并且开启电压降低至 0.52 V,器件稳定性明显提升.碳基材料通常有着 极佳的热稳定性和化学稳定性,但器件阻变行为通 常需要与其他材料相结合,这无疑限制了全碳器件 的发展. 东北师范大学 Liu 研究团队 [73] 在原有工 作基础上将氮掺杂的碳量子点和氧化石墨烯复合 材料作为阻变层,还原氧化石墨烯和石墨烯层作 为顶电极和底电极,构建了全碳忆阻突触器件 (图 5(g)). 在波长范围为 320—380 nm 紫外光照射 下,薄膜中的碳量子点产生电子空穴对.其中的光 生电子会与氧化石墨烯表面的含氧官能团反应,引 起局域还原.在适宜的光照时间 (10 min) 和碳点 浓度 (质量浓度 30%) 条件下, 器件可以由数字型 阻变行为转变为模拟型阻变行为.得益于碳基材料 良好的热稳定性,器件在 450 K 高温下仍然可以 表现出稳定的忆阻行为.



图 5 (a) 全无机钙钛矿基光电忆阻突触器件示意图; (b) 器件在紫外光下的光开关特性; (c) 器件在电学脉冲信号下的电导调制^[55]; (d) Al/GO-TiO₂/ITO 存储器件结构图; (e) 紫外光照射对器件电初始化和阻变的影响; (f) 紫外照射时间对开关电压的调节^[16]; (g) 生物突触及 RGO/GO-NCQDs/石墨烯全碳器件示意图; (h) GO-NCQD 复合材料的光致还原过程^[73]

Fig. 5. (a) Structural diagram of all-inorganic perovskite optoelectronic synapses; (b) optical switching characteristics under UV light; (c) potentiation and depression behaviors under electrical stimulation^[55]; (d) schematic illustration of Al/GO-TiO₂/ITO memory device; (e), (f) the effect of UV irradiation time on forming (e) and switching voltages^[16] (f); (g) schematic illustration of biological synapse and RGO/GO-NCQDs/graphene memristor; (h) photo-reduction process of GO-NCQDs film^[73].

3.2 全光调制型忆阻器件

全光调制型器件通常需要复杂的器件结构以 及繁琐的信号调制过程,这无疑限制其进一步的发 展.如何通过材料优化选择及器件结构设计在单一 器件实现全光可逆调制过程是领域内的研究热点. 中国科学院宁波材料技术与工程研究所 Zhuge 研 究组^[74]利用双层 InGaZnO 材料成功构建全光调 制型忆阻器件.如图 6(a)所示,其双层结构由缺氧 的 O_D-IGZO 层和含氧丰富的 O_R-IGZO 层构成. 单纯的 O_D-IGZO 材料器件在可见光的照射下无法 观察到明显的光电流;而 O_R-IGZO 材料作为中间 层在可见光下产生易失性光电流,对红外光没有响应^[74]. 对于蓝光照射后的双层器件,绿光 (800 nm)和红 光 (900 nm)照射过程使得器件电流增加;而与暗 态相比,近红外光照射可以引起更明显的电流下降 (图 6(b)). 如图 6(c) 所示, 通过增加近红外光的强 度能够提升器件关闭效率. 借助波长为 420 nm 和 800 nm 的光信号可以实现器件电导的可逆转变过 程 (图 6(d)). 对于双层结构器件的光开启过程如 图 6(e) 所示, 光可以诱导界面势垒处的中性氧空 位 (V_o) 变成离化氧空位 (V²⁺₀), 导致界面处势垒 高度降低, 器件电导上升; 而在光关闭过程中, 红 光照射能诱导界面势阱中的电子隧穿或越过势垒 进入 O_D-IGZO 层. 部分电子被离化的氧空位 (V²⁺₀) 捕获, 转变为 V_o, 界面处势垒增大, 阻碍电子的流 动, 器件电导下降.

此外,墨尔本皇家理工大学 Ahmed 实验组^[19] 基于层状黑磷材料构建了具有全光调制存储能力 的神经形态成像器件.图 6(f) 为器件结构示意图, 层状黑磷材料位于 SiO₂/Si 衬底上,顶端为 Cr/Au



图 6 (a) Au/O_D-IGZO/O_R-IGZO/Pt 结构的全光调控忆阻器件; (b) 不同波长光照射下 IGZO 器件响应电流; (c) 光照强度对光 关闭过程影响; (d) 光开启和光关闭特性; (e) 全光调制过程机制图^[74]; (f) 基于 BP 材料的光电晶体管; (g), (h) 器件在 280 nm 和 365 nm 光脉冲下的响应电流; (i) 器件的长时增强和抑制特性^[19]

Fig. 6. (a) Schematic diagram of the Au/O_D -IGZO/O_R-IGZO/Pt device structure; (b) current response depending on light of various wavelengths; (c) effect of power density on optical reset behavior; (d) reversible modulation of device conductance; (e) schematic illustrations of all-optically controlled device^[74]; (f) schematic of BP based device; (g), (h) transient photocurrent under 280 nm and 365 nm illumination; (i) LTP and LTD behaviors under consecutive pulse^[19].

源漏电极.器件的归一化瞬态光电流如图 6(g),(h) 所示,在 280 nm 光脉冲刺激下源漏电流表现为上 升趋势,而在 365 nm 光作用下呈现出相反的光电 流降低行为(值得注意的是整个过程不需要栅极电 压信号的调制).图 6(i)为器件在连续增强光脉冲 和抑制光脉冲下的长时增强(LTP)和长时抑制 (LTD)曲线.其工作机制可以归因于黑磷材料自 身的性质.黑磷材料在自然环境中很容易吸附氧气 从而形成氧化磷,这种表面自发氧化形成的非晶层 可以作为光生载流子的捕获位点^[75–80].在相对低 光子能量(365 nm)照射时,光生载流子会被氧缺 陷捕获,器件电导下降产生负向的光电流.相反高 光子能量(280 nm)激发过程可以有效诱导环境中 氢气和水分子分裂进而钝化氧化层中的缺陷^[81].

增大载流子浓度实现正向光电导.

除了将光子与材料固有缺陷构建联系实现全 光调制的方法外,东北师范大学 Liu 研究团队^[82] 设计了一种新型等离激元光电忆阻器件,实现可逆 的全光调制过程.如图 7(a)所示,该工作将 Ag 纳 米颗粒负载到氧化钛多孔薄膜中组成纳米复合材 料作为阻变层,Au 和 ITO 分别作为顶电极和底电 极.器件的全光调制过程如图 7(b)—(f)所示,施加 可见光器件表现出长时增强过程,相反施加紫外光 时,器件呈现长时抑制过程.增加照射过程的光强 和时间可以提升电导的变化范围.研究表明器件的 全光可逆调控机制主要归因于 Ag 和 TiO₂ 材料微 纳界面上的肖特基势垒.在可见光的照射下,Ag 纳米颗粒由于自身的局域表面等离子体共振效应



图 7 (a) 基于 Ag-TiO₂ 材料的全光调控忆阻器件; (b), (c) 可见光和紫外光脉冲刺激下器件的电流响应; (d), (e) 光照强度和时间对器件电流的影响; (f) 全光可逆调制过程; (g), (h) 器件在光电信号刺激下的运行机制^[82]

Fig. 7. (a) Fully light-modulated memristor based on Ag-TiO₂ nanocomposite; (b), (c) transient photocurrent under the illumination of visible and UV light; (d), (e) the response current depending on irradiation time and intensity; (f) fully light-modulated behaviors; (g), (h) operating mechanism of the Ag-TiO₂ based optoelectronic device^[82].

(LSPR)可以激发出电子并转移到 TiO₂导带^[83]. 如图 7(g) 所示,电子和银离子发生空间分离过程, 银纳米颗粒发生光致氧化 (Ag⁰→e⁻+Ag⁺),并降低 Ag/TiO₂ 界面肖特基势垒提升器件电导.相反紫 外光能够激发 TiO₂ 价带电子跃迁至导带并进一步 与银离子结合,实现光致还原过程 (Ag⁺+e⁻→Ag⁰). 该过程中 Ag/TiO₂ 微纳界面处的肖特基势垒逐渐 增大,器件电导降低从而实现可逆的光增强和光抑 制过程^[72].此外,不同光照过程引起的银纳米颗粒 化学状态变化直接影响后续电场下的迁移过程,从 而在单一器件上可以实现光电多功能集成.

4 感存算一体化功能及应用

光电忆阻器由于自身独特的光电结合调制方 式在突触功能模拟方面有着难以比拟的优势.得益 于多维度的调节方式,光电忆阻器在光遗传学启发的突触可塑性模拟,光电/全光逻辑运算,构建高效神经形态视觉系统等方面有良好的应用前景^[10,84]. 以下部分主要讨论器件在感存算一体化功能应用方面的最新进展.

4.1 光电逻辑运算及类脑功能模拟

突触可塑性是大脑学习和记忆功能的神经机制,发展人工突触模拟生物突触行为是构建高效神经形态计算系统的基础^[51,74,85].中国科学院材料科学与光电工程研究中心 Shen 研究组^[86]基于 TiN_xO_{2-x}/MoS₂异质结构建人工光电突触实现神经形态计算和视觉系统.器件的结构如图 8(a)所示,从上至下依次为 Al 电极、TiN_xO_{2-x}薄膜、MoS₂薄膜以及 ITO电极.对器件施加 365 nm 的紫外光 (3.15 mW/cm²)时,器件电导出现明显的上升过程.光照结束后,



图 8 (a) TiN_xO_{2-x}/MoS₂ 异质结光电突触器件; (b) 单个光脉冲引起的增强过程; (c) 对脉冲促进 (PPF) 功能; (d), (e) 不同光照 强度和时间下器件的电导响应; (f) 连续光脉冲引起的电导变化^[86]

Fig. 8. (a) Structural illustration of $\text{TiN}_x O_{2-x}/\text{MoS}_2$ heterostructure-based optoelectronic synapse; (b) optical potentiation process; (c) paired pulse facilitation function; (d), (e) conductance response depending on the illumination intensity and duration; (f) transient response under consecutive optical pulses^[86].

器件电导出现缓慢衰减而不是快速恢复到初始状态.这就是典型的兴奋性突触后电流 (excitatory postsynaptic current, EPSC) 现象, 施加负电压脉冲可以消除这一衰减效果, 使器件电导快速下降到初始状态.通过改变连续光脉冲的间隔可以实现对脉冲促进 (paired pulse facilitation, PPF) 功能.如图 8(b) 所示, 延长脉冲对的间隔时间会降低相应的 PPF 指数.此外器件在不同的光强级别下展现出强度依赖的电导响应特性.更高光强信号不仅可以引起更大的电导响应,还会使撤光后电导的衰减过程减缓.调整光照时间同样可以改变电导响应幅值及暗态下的弛豫过程 (图 8(c), (d)).这种光引起的电导响应是模拟高阶突触功能的基础^[87].

联合式学习是人脑的重要能力,而巴普洛夫实 验是该领域内的经典实验案例^[88].如图 9(a)所示, 中国科学院深圳先进技术研究所 Wang 研究组^[89] 利用二维 TiS₃材料构建了两端光电人工突触器 件.器件的光响应范围涵盖可见光和红外光区 域^[90].不同波长光信号下器件的光增强电抑制行为如 图 9(c)所示.基于以上光电调控特性模拟巴普洛 夫狗实验,其中 400 nm 光照信号作为食物刺激 (非条件刺激),电信号作为铃声刺激 (条件刺激), 器件电流高于阈值视为分泌唾液 (图 9(d)).单独 的电压脉冲序列下器件电流没有明显变化,即没有 分泌唾液.对器件施加 400 nm 光照时,器件电 流超过阈值,表明开始分泌唾液.对器件同时施 加光信号和电信号作为训练过程.经过训练后, 单独的电刺激可以让器件电流高于阈值,表明成功 将非条件刺激和条件刺激建立联系.最后器件电流 略微降低,与生物系统中联合式学习的遗忘过程 一致.

光电忆阻器由于能直接对光信号进行存储和 处理,在光电逻辑运算方面也有良好的应用前景[91]. 中国科学院宁波材料技术与工程研究所 Li 研究组^[14] 利用 CeO_x基光电忆阻器的阻变行为和持续光电 导效应实现了可重构存储逻辑功能.器件的操作原 理示意图如图 10(a) 所示. 其中光信号和电信号作 为逻辑运算的输入信号,器件电流作为输出信号, 且输出电流高于阈值使结果为1,低于阈值为0. 对于初始低电流状态器件, 单一的电信号或电信号 都无法使器件电流高于阈值,输出为0.只有同时 输入光信号和电信号时,器件输入为1,整体逻辑 功能为"与"运算. 对器件施加光脉冲进行光开启过 程,器件转变为"或"门.只有光信号和电信号都不 输入时,输出结果为0,否则输出为1.相关的实验 结果如图 10(b), (c) 所示. 值得注意的是, 电关 闭过程可以将"或"门重构,恢复到"与"门[92].因此



图 9 (a) Al/TiS₃/ITO 器件结构示意图; (b) 不同波长光照射下的电流-电压曲线; (c) 光电信号下的器件电导变化; (d) 巴普洛夫狗实验中经典条件反射模拟^[99]

Fig. 9. (a) Sandwich-like structure of the Al/TiS₃/ITO memristor; (b) RS behaviors modulated by different wavelengths; (c) conductance change under optical and electric signals; (d) simulation of classical conditioning in Pavlov's dog experiment^[89].



图 10 (a) 与门或门逻辑运算转换示意图; (b) 可重构与门、或门输出结果; (c) 真值表及输出电流值[4]

Fig. 10. (a) Logic operation switching of "AND" and "OR" gates; (b) reprogrammable outputs of memlogic "AND" and "OR"; (c) truth table and output current of "AND" and "OR" gate^[14].

器件借助自身的光开启电关闭过程可以实现可重 构逻辑运算功能.此外,将负电压信号作为第2个 电输入可以实现"非"逻辑运算. 当第2个电输入端 为0时,整体输出结果与上述的"与"及"或"门相同; 当引入第2个电信号后,无论之前的逻辑输入信号 如何改变,器件都将被关闭回高阻状态,即输出结 果为 0. 此外, 基于不同光强和电流的对应关系, 利 用两个独立的光信号可以实现加法器和数模转换 功能. 此外墨尔本皇家理工大学 Ahmed 实验组^[71] 利用层状黑磷材料构建新型多功能光电器件,借助 不同波长的紫外光信号能够实现正/负光电流.将 不同激发波长的光信号作为输入,可以实现或门、 异或门逻辑运算;韩国东国大学 Im 研究组^[93] 基于 二维 Te/ReS, 范德瓦耳斯异质结构新型光电调控 晶体管,通过控制光电输入信号可以实现包括与 门、或门在内的多种逻辑运算.

4.2 高效神经形态视觉系统

视觉系统是人类获取外界信息的重要途经,如 何构建能与人类视网膜相比拟的高效神经形态视 觉系统一直是国内外的研究热点^[94–98]. 传统的机 器视觉通常由感知、存储、处理单元组成,其分离 的硬件结构严重影响整体效率^[99,100]. 光电忆阻器 凭借感存算一体化架构有效避免了硬件分离带来 的高延迟和高功耗,为构建高效神经形态视觉系统 提供了理想硬件基础^[101]. 本节主要介绍光电忆阻 器在神经形态视觉系统方面的研究进展.

香港理工大学 Chai 研究组^[30] 利用 Pd/MoO_x/ ITO 忆阻阵列实现图像感知、存储和实时预处理功 能.其中的图像预处理功能可以有效提升传感数据 质量,大幅度提高处理效率和后续识别的准确性. 如图 11(a) 所示, Chai 研究组^[30] 通过构建 3×5 光电忆阻阵列来展示图像的预处理过程,主要包括



图 11 (a) 基于光电忆阻器的图像记忆及预处理功能; (b) 神经形态视觉系统图像识别模拟^[30]; (c) 突触光电晶体管光照示意图; (d) 未知彩色光识别功能^[104]

Fig. 11. (a) Image memorization and preprocessing functions based on optoelectronic memristor; (b) simulation of image recognition in artificial neural network^[30]; (c) device structure of 2D perovskite/organic heterojunction synaptic phototransistor; (d) simulating the recognition of unknow light^[104].

对比增强和噪音降低功能. 对阵列不同部分施加 4种不同强度的光信号,不同光强引起的器件响应 电流差异明显高于光信号自身变化,也就是高光强 可以引起更大的电流增长率,起到图像的对比增强 作用. 经过1 min 后, 低光强照射的像素点电流弛 豫更快, 高强度光信号引起的电流变化衰减相对较 慢,输出图像的对比度进一步增强.作者进一步利 用 6×7 光电忆阻阵列与三层人工神经网络构建神 经形态视觉系统,模拟模式识别过程,如图 11(b) 所示.对比发现,图像预处理过程可以明显提升整 体的识别速率和精度. 借助光电忆阻阵列的预处理 功能, 仅需 1000 次训练识别率就能达到 0.986^[30]. 此外,韩国基础科学研究所纳米颗粒研究中心 Kim 研究组^[102]利用 MoS₂/1, 3, 5-三乙烯基–1, 3, 5-三 甲基环三硅氧烷异质结构建曲面图像传感阵列,实 现高效图像学习识别过程.

人眼除了可以有效分辨光学信号强度外,还能 辨别物体颜色差异^[103].中国科学院化学研究所 Liu 研究组^[104],利用二维钙钛矿 PEA₂SnI₄/有机 材料 (Y₆) 异质结构建光电突触晶体管实现红绿蓝 及红外光识别, 如图 11(c) 所示. 其中 PEA₂SnI₄ 材料的吸收范围覆盖整个可见光区域,而有机物 Y₆ 在红外区域具有良好的光响应[105].由于异质结结 构对不同波长光吸收效率以及光电转化效率的差 异,在可见光区域器件对蓝光响应幅值最大,红光 最小. 在正栅压调控下可见光范围信号使得器件电 流逐渐减小,也就是抑制突触后电流现象 (IPSC). 相反近红外光信号能引起兴奋性突触后电流 (EPSC). 对于一个未知的光信号,器件能够将其转化为兴奋/ 抑制性突触后电流, 内部处理器根据响应类型和幅 值实现色彩分辨功能. Liu 等^[105] 进一步制备 12×5 柔性器件阵列并将其贴敷在人工眼球上,成功分辨 光信号颜色并展示 4 种颜色的 G, o, o, d 字母. 成 均馆大学 Park 研究组^[106] 基于 h-BN/WSe2 异质 结结构制备了具有光传感和突触功能的光学神经 突触器件. 如图 12(a) 所示, 器件由突触器件和感



图 12 (a) 人类视觉神经系统及 h-BN/WSe₂ 基突触器件示意图; (b) 不同光照条件下的长时增强和抑制行为; (c) 人工视觉神经 网络训练测试实例; (d) 不同训练次数后的识别率^[106]

Fig. 12. (a) Schematic illustration of the human optical nerve system; (b) schematic illustration of the human optical nerve system; (c) dataset consisted of colored and color-mixed number for training and testing; (d) dependence of recognition rate on training epochs^[106].

光器件构成,由此构建的光学神经网络能够模拟人 类视觉系统识别彩色和混色图案的能力.对于感光 器件,短波长光信号能引起更大的电流增长幅值, 其中蓝光增强效果是红光的 1000 倍以上(图 12(b) 所示).借助感光器件调节异质结突触器件的电荷 捕获过程实现光可调谐的突触可塑性.利用器件的 光电特性构建光神经网络进行彩色和混色模式识 别.其中阵列是由 28×28 个人工锥形细胞群构成, 单个细胞群包含 3 个单元,分别对红绿蓝色光进行 探测.图 12(c)是用于训练的单色数字图像示例以 及彩色测试图像.实验结果显示,传统的神经网络 对于特定彩色测试图案的识别率仅为 40%,而借助 感光特性构建光神经网络后,经过 50 次学习整体 识别率就超过 90% (图 12(d))^[106]. 由此表明, 光学 神经网络在识别颜色混合图案方面比传统神经网 络更具优势, 也为构建进行复杂模式识别任务的神 经网络提供基础.

4.3 动态探测与识别

近期,南京大学 Miao 研究组^[107] 利用 WSe₂/h-BN/Al₂O₃ 范德瓦耳斯异质结构建视网膜拟态传感器,实现图像感知和处理功能集成.图 13(a) 是 3×3 视网膜拟态传感器的光学图像.通过调控栅极电压极性实现开启和关闭两种截然不同的光响应,这与生物视网膜中双极细胞的响应相似(图 13(b))^[108,109].借助不同的开启关闭光响应器件可以在图像感知



图 13 (a) 基于 WSe₂/h-BN/Al₂O₃ 异质结的视网膜拟态传感器件; (b) 异质结器件相应的开关光响应; (c)—(e) 基于神经形态 视觉系统实现目标追踪^[107]

Fig. 13. (a) Retinomorphic sensor based on $WSe_2/h-BN/Al_2O_3$ heterostructure device; (b) the On and Off photo response of heterostructure device; (c)–(e) object tracking of neuromorphic vision system^[107].
的同时实现边缘增强等处理功能,进而大幅度地提 升识别的效率和精确度. 进一步将传感器与 Pt/Ta/ HfO₂/Ta 结构的 1T1R 忆阻阵列连接, 成功实现 脑启发的视觉感知应用. 该系统中的视网膜拟态传 感器在感知图像信息的同时进行预处理,去除冗余 数据并保留了关键信息. 随后信息传入忆阻神经网 络用于感知,整个过程有效避免了繁琐的模拟-数 字信号转化过程. 鉴于以上优异性能, 该神经形态 视觉系统在物体追踪方面也有着良好的应用前景. 物体的运动轨迹轮廓经由视网膜拟态传感器提取 轮廓位置后作为时空特征传入循环神经网络实现 物体追踪和预测[110-112].

除了将视网膜拟态传感器和忆阻器阵列连接 构建神经形态视觉系统的方案外,复旦大学 Zhou 研究组[113] 通过构建二维异质结视网膜拟态器件完 成感知记忆计算功能集成,并实现对运动物体的探 测和识别. 该视网膜拟态器件主要是基于 BP/Al₂O₃/ WSe₂/h-BN 异质结结构的浮栅晶体管. 通过设计

> (a) RGB h-BN WSea Al_2O_3

(c) Original image pixel distribution

Total: 462570

0.2

0.4

0.6

Brightness/arb. units

0.8

 10^5

 10^{2}

10

10 0

Frequency 10^{4} 10^{3} 光电信号能够实现非易失性的正向和负向光电流, 可以模拟视网膜中双极细胞的可逆调节和存储功 能,如图 14(c) 所示.进一步将物体的运动过程视 为不同时刻的图片序列.将某一时刻的图像亮度分 布与正向电导矩阵相乘,并将下一时刻的图像亮度 分布与负向电导矩阵相乘,最终将所有结果整合并 输出结果,从而实现运动物体识别.与传统动态探 测技术相比,该方法可以极大程度地降低周围环境 的影响,并减少冗余数据传输[113].

总结与展望 5

本综述主要从光电忆阻材料与机制,器件与特 性及感存算一体化应用3个方面回顾了近期光电 忆阻器件领域内的重要进展. 其中常见的光电忆阻 材料主要包括氧化物、碳基材料、有机材料、有机-无机杂化钙钛矿、二维材料等.而相关的光电忆阻 机制主要分为光子-离子耦合型和光子-电子耦合型



图 14 (a) 多功能视网膜形态器件结构示意图; (b) 不同电压信号调节下的电流响应; (c) 基于二维神经形态器件的动态探测示 意图[113]

Fig. 14. (a) All-in-one retinomorphic device; (b) current response under various drain voltage modulation; (c) motion detection based on 2 D retinomorphic device^[113].

两个类别.得益于数据传输速度快、低功耗高能效 等优点,光电忆阻器被认为是构建多功能神经形态 计算系统的理想硬件选择.其中光电多维度调制手 段可以大幅度提升光电忆阻器在突触功能模拟方 面的可控性,并实现更高的时空分辨率.此外基于 光电忆阻器的感存算一体化架构可以显著提升整 体算力,从而构建更高效的计算系统.

尽管近期光电忆阻器取得了诸多重要进展,但 目前仍然面临着一些重要的挑战. 1) 材料层面. 常 见的光电材料由于固有带隙限制,通常只对特定波 长范围的光信号有响应.这也促使我们重新思考如 何借助材料复合、能带结构设计等方法来实现全光 谱响应,构筑具有高量子效率的光电忆阻器.目前 通过对硅纳米晶材料进行硼元素掺杂,可以将器件 的响应范围拓展为紫外至近红外波段^[54];将石墨 烯和碳纳米管材料混合构建新型光刺激突触器件, 其对可见光和紫外光范围的信号都有较好响应[114]. 从材料设计角度出发可以从根本上优化器件性能, 提升响应速度的同时降低相关功耗. 此外现有的光 电/全光调制型忆阻材料的运行机制仍然不是十分 明确, 需要进一步探究. 2) 在器件制备和工艺方 面. 目前光电忆阻材料的主要制备方法包括化学气 相沉积,物理气相沉积(真空蒸镀,溅射镀膜,分子 束外延等), 溶液法, 机械剥离法等. 加工大面积高 质量光电忆阻材料薄膜是制备高集成度器件阵列 的关键.其中机制剥离转移方法可以有效解决异质 结晶格失配问题,但在材料大规模制备方面仍存在 一些不足; 低温溶液法可以避免高温对器件性能的 影响,但材料缺陷难以控制.现阶段许多光电忆阻 研究只是基于小规模分立器件单元的简单串联,没 有发挥出集感存算于一体的光电集成优势. 针对集 成工艺而言,需要考虑材料生产、器件制备和阵列 集成等各层级的工艺兼容性,这给半导体集成电路 工艺可靠性带来了挑战.3) 光电集成系统架构方 面.针对不同的应用场景,需要在光的强度、波长、 相位和偏振态等参数的调制下,实现集光电感知、 计算和存储于一体的集成系统,配合相关的算法程 序完成更高阶的数据信息任务.目前已有研究组基 于 WSe₂/ReSe₂ 异质结构建线偏振敏感的光电器 件以及利用 ReS2 材料构建偏振感知神经形态晶体 管[115,116]. 此外,利用现有的高速光、电通信技术将 各功能单元进行智能互联,基于多体协同配合可以 构建更智能的光电集成系统.

目前感存算一体化架构的发展仍然处于发展 初期,有很多分支领域需要进一步探索.尽管如此 光电忆阻器作为感存算一体化系统的重要组成部 分,已经在逻辑运算、类脑功能模拟等多个方面展 现出良好的应用前景.相信随着工艺技术的不断进 步,相关问题会逐步得到解决,届时光电忆阻器件 必定会取得更大突破.

参考文献

- Hasegawa T, Terabe K, Tsuruoka T, Aono M 2012 Adv. Mater. 24 252
- [2] Lee J, Lu W D 2018 Adv. Mater. **30** 1702770
- [3] Cheng Z G, Rios C, Pernice W H P, Wright C D, Bhaskaran H 2017 Sci. Adv. 3 e1700160
- [4] Xu W T, Cho H, Kim Y H, Kim Y T, Wolf C, Park C G, Lee T W 2016 Adv. Mater. 28 5916
- [5] Gholipour B, Bastock P, Craig C, Khan K, Hewak D, Soci C 2015 Adv. Optical Mater. 3 635
- [6] Mao J Y, Zhou L, Zhu X, Zhou Y, HanS T 2019 Adv. Optical Mater. 7 1900766
- [7] Chua L 1971 IEEE Trans. Circuit Theory 18 507
- [8] Wang Z, Zeng T, Ren Y, Lin Y, Xu H, Zhao X, Liu Y, Ielmini D 2020 Nat. Commun. 11 1510
- [9] Jo S, Chang T, Ebong I, Bhadviya B B, Mazumder P, Lu W 2010 Nano Lett. 10 1297
- [10] Sun L F, Wang Z R, Jiang J, Kim Y, Joo B, Zheng S, Lee S, Yu W J, Kong B S, Yang H 2021 Sci. Adv. 7 eabg1455
- [11] Lin Y, Zhang X, Shan X, Zeng T, Zhao X, Wang Z, Kang Z, Xu H, Liu Y 2020 J. Mater. Chem. C 8 14789
- [12] Kim S J, Kim S B, Jang H W 2021 Iscience 24 101889
- [13] Milo V, Malavena G, Monzio Compagnoni C, Ielmini D 2020 Material 13 166
- [14] Tan H, Liu G, Yang H, Yi X, Pan L, Shang J, Long S, Liu M, Wu Y, Li R 2017 ACS Nano 11 11298
- [15] Gao S, Liu G, Yang H L, Hu C, Chen Q, Gong G, Xue W, Yi X, Shang J, Li R 2019 ACS Nano 13 2634
- [16] Zhao X, Wang Z, Xie Y, Xu H, Zhu J, Zhang X, Liu W, Yang G, Ma J, Liu Y 2018 *Small* 14 1801325
- [17] Jaafar A H, Gray R J, Verrelli E, O'Neill M, Kelly S M, Kemp N T 2017 Nanoscale 9 17091
- [18] Zhao X, Wang Z, Li W, Sun S, Xu H, Zhou P, Xu J, Lin Y, Liu Y 2020 Adv. Funct. Mater. 30 1910151
- [19] Ahmed T, Tahir M, Low M X, Ren Y, Tawfik S, Mayes E, Kuriakose S, Nawaz S, Spencer S, Chen H, Bhaskaran M, Sriram S, Walia S 2021 Adv. Mater. 33 2004207
- [20] Wang Z, Xu H, Li X, Yu H, Liu Y, Zhu X 2012 Adv. Funct. Mater. 22 2759
- [21] Mu B, Guo L, Liao J, Xie P, Ding G, Lv Z, Zhou Y, Han S, YanY 2021 Small 17 2103837
- [22] Ohno T, Hasegawa T, Tsuruoka T, Terabe K, Gimzewski J K, Aono M 2011 Nat. Mater. 10 591
- [23] Wan C, Zhu L, Liu Y, Feng P, Liu Z, Cao H, Xiao P, Shi Y, Wan Q 2016 Adv. Mater. 28 3557
- [24] Zidan M A, Strachan J P, Lu W D 2018 Nat. Electron. 1 22
- [25] Hu D, Yang R, Jiang L, Guo X 2018 ACS Appl. Mater. Interfaces 10 6463
- [26] Chen S, Lou Z, Chen D, Shen G 2018 Adv. Mater. 30 1705400

- [27] Sun F, Lu Q, Feng S, Zhang T 2021 ACS Nano 15 3875
- [28] Ji X, Zhao X, Tan M C, Zhao R 2020 Adv. Intell. Syst. 2 1900118
- [29] Carrara S 2021 *IEEE Sens. J* 21 12370
- [30] Zhou F, Zhou Z, Chen J, Choy T, Wang J, Zhang N, Lin Z, Yu S, Kang J, Wong H, Chai Y 2019 Nat. Nanotechnol. 14 776
- [31] Zhou Z, Pei Y, Zhao J, Fu G, Yan X 2021 Appl. Phys. Lett. 118 191103
- [32] Ling H, Tan K, Fang Q, Xu X, Chen H, Li W, Liu Y, Wang L, Yi M, Huang R, Qian Y, Xie L, Huang W 2017 Adv. Electron. Mater. 3 1600416
- [33] Wang S, Fan W, Liu Z, Yu A, Jiang X 2018 J. Mater. Chem. C 6 191
- [34] Tsuruoka T, Terabe K, Hasegawa T, Valov I, Waser R, Aono M 2012 Adv. Funct. Mater. 22 70
- [35] Deb S, Chopoorian J 1966 J. Appl. Phys. 37 4818
- [36] Natansohn A, Rochon P 2002 Chem. Rev. 102 4139
- [37] Tsai H, Asadpour R, Blancon J, Stoumpos C, Durand O, Strzalka J, Chen B, Verduzco R, Ajayan P, Tretiak S, Even J, Alam M, Kanatzidis M, Nie W, Mohite A 2018 Science 360 67
- [38] Liu Y, Ievlev A V, Collins L, Borodinov N, Belianinov A, Keum J K, Wang M, Ahmadi M, Jesse S, Xiao K, Sumpter B G, Hu B, Kalinin S V, Ovchinnikova O S 2019 Adv. Opt. Mater. 7 1901451
- [39] Hossain A, Bandyopadhyay P, Karmakar A, Ullah A, Manavalan R, Sakthipandi K, Alhokbany K, Alshehri S, Ahmed J 2021 Ceram. Int. 48 7325
- [40] Walsh A, Stranks S D 2018 ACS Energy Lett. 3 1983
- [41] Guan X, Hu W, Haque Md A, Wei N, Liu Z, Chen A, Wu T 2018 Adv. Funct. Mater. 28 1704665
- [42] Yang J M, Kim S G, Seo J Y, Cuhadar C, Son D Y, Lee D, Park N G 2018 Adv. Electron. Mater 4 1800190
- [43] Yang K, Li F, Veeramalai C P, Guo T 2017 Appl. Phys. Lett. 110 083102
- [44] Zhao X, Xu H, Wang Z, Lin Y, Liu Y 2019 InfoMat 1 183
- [45] Ahmadi M, Wu T, Hu B 2017 Adv. Mater. 29 1605242
- [46] Hu J, Yan L, You W 2018 Adv. Mater. 30 1802041
- [47] Choi J, Han J S, Hong K, Kim S Y, Jang H W 2018 Adv Mater. 30 1870317
- [48] Manjappa M, Srivastava Y K, Solanki A, Kumar A, Sum T C, Singh R 2017 Adv. Mater. 29 1605881
- [49] Zhu X, Lu W D 2018 ACS Nano 12 1242
- [50] Zhu X, Lee J, Lu W D 2017 Adv. Mater. 29 1700527
- [51] Ham S, Choi S, Cho H, Na S, Wang G 2019 Adv. Funct. Mater. 29 1806646
- [52] Skorodumova N, Simak S, Lundqvist B, Abrikosov I, Johansson B 2002 Phys. Rev. Lett. 89 166601
- [53] Pan Z, Peng W, Li F, He Y 2018 Adv. Funct. Mater. 28 1706897
- [54] Tan H, Ni Z, Peng W, Du S, Liu X, Zhao S, Li W, Ye Z, Xu M, Xu Y, Pi X, Yang D 2018 Nano Energy 52 422
- [55] Ma F, Zhu Y, Xu Z, Liu Y, Zheng X, Ju S, Li Q, Ni Z, Hu H, Chai Y, Wu C, Kim T W, Li F 2020 Adv. Funct. Mater. 30 1908901
- [56] He H, Yang R, Zhou W, Huang H-M, Xiong J, Gan L, Zhai T, Guo X 2018 *Small* 14 1800079
- [57] Yang J, Pickett M, Li X, Ohlberg D, Stewart D, Williams R 2008 Nat. Nanotechnol. 3 429
- [58] You T, Du N, Slesazeck S, Mikolajick T, Li G, Burger D, Skorupa I, Stocker H, Abendroth B, Beyer A, Volz K, Schmidt O, Schmidt H 2014 ACS Appl. Mater. Interfaces 6

19758

- [59] Zhai Y, Yang X, Wang F, Li Z, Ding G, Qiu Z, Wang Y, Zhou Y, Han S T 2018 Adv. Mater. 30 1803563
- [60] Yang R, Huang H, Hong Q, Yin X, Tan Z, Shi T, Zhou Y, Miao X, Wang X, Mi S 2018 Adv. Funct. Mater. 28 1704455
- [61] Tan H, Liu G, Zhu X, Yang H, Chen B, Chen X, Shang J, Lu W D, Wu Y, Li R 2015 Adv. Mater. 27 2797
- [62] Zhou S, Ni Z, Ding Y, Sugaya M, Pi X, Nozaki T 2016 ACS Photonics 3 415
- [63] Ni Z, Ma L, Du S, Xu Y, Yuan M, Fang H, Wang Z, Xu M, Li D, Yang J, Hu W, Pi X, Yang D 2017 ACS Nano 11 9854
- [64] Ni Z, Pi X, Zhou S, Nozaki T, Grandidier B, Yang D 2016 Adv. Opt. Mater. 4 700
- [65] Ge R, Wu X, Kim M, Shi J, Sonde S, Tao L, Zhang Y, Lee J, Akinwande D 2018 Nano Lett. 18 434
- [66] Zhou Y, Liu D, Wang J, Cheng Z, Liu L, Yang N, Liu Y, Xia T, Liu X, Zhang X, Ye C, Xu Z, Xiong W, Chu P, Yu X 2020 ACS Appl. Mater. Interface 12 25108
- [67] Xiang D, Liu T, Xu J, Tan J, Hu Z, Lei B, Zheng Y, Wu J, Neto A, Liu L, Chen W 2018 *Nat. Commun.* 9 2966
- [68] Tran M D, Kim H, Kim J S, Doan M H, Chau T K, Vu Q A, Kim J H, Lee Y H 2019 Adv. Mater. 31 807075
- [69] Pei F, Yan L, Wu Z, Lu J, Zhao J, Chen J Liu Q, Yan X 2021 ACS Nano 15 17319
- [70] Wang Q S, Wen Y, Cai K M, Cheng R Q, Yin L, Zhang Y, Li J, Wang Z X, Wang F, Wang F M, Shifa T A, Jiang C, Yang H, He J 2018 Sci. Adv. 4 eaap7916
- [71] Lee J, Pak S, Lee Y W, Cho Y, Hong J, Giraud P, Shin H S, Morris S M, Sohn J I, Cha S, Kim J M 2017 Nat. Commun. 8 14734
- [72] Miller D A B 2009 Proc. IEEE 97 1166
- [73] Lin Y, Wang Z, Zhang X, Zeng T, Bai L, Kang Z, Wang C, Zhao X, Xu H, Liu Y. 2020 NPG Asia Mater. 12 1
- [74] Hu L X, Yang J, Wang J R, Cheng P H, O. Chua L, Zhuge F 2020 Adv. Funct. Mater. **31** 2005582
- [75] Ahmed T, Balendhran S, Karim M N, Mayes E L H, Field M R, Ramanathan R, Singh M, Bansal V, Sriram S, Bhaskaran M, Walia S 2017 NPJ 2 D Mater. Appl. 1 18
- [76] Walia S, Sabri Y, Ahmed T, Field M R, Ramanathan R, Arash A, Bhargava S K, Sriram S, Bhaskaran M, Bansal V, Balendhran S 2017 2 D Mater. 4 015025
- [77] Favron A, Gaufres E, Fossard F, Phaneuf-L'Heureux A L, Tang N Y W, Levesque P L, Loiseau A, Leonelli R, Francoeur S, Martel R 2015 Nat. Mater. 14 826
- [78] Kuriakose S, Ahmed T, Balendhran S, Bansal V, Sriram S, Bhaskaran M, Walia S 2018 2D Mater. 5 032001
- [79] Utt K L, Rivero P, Mehboudi M, Harriss E O, Borunda M F, SanJuan A A P, Barraza-Lopez S. 2015 ACS Cent. Sci. 1 320
- [80] Wood J D, Wells S A, Jariwala D, Chen K S, Cho E, Sangwan V K, Liu X, Lauhon L J, Marks T. J, Hersam M C 2014 Nano Lett. 14 6964
- [81] Ahmed T, Kuriakose S, Abbas S, Spencer M J S, Rahman M A, Tahir M, Lu Y R, Sonar P, Bansal V, Bhaskaran M, Sriram S, Walia S 2019 Adv. Funct. Mater. 29 1901991
- [82] Shan X Y, Zhao C Y, Wang X N, Wang Z Q, Lin Y, Zeng T, Zhao X N, Xu H Y, Zhang X T, Liu Y C 2021 Adv. Sci. 8 2104632
- [83] Colton R J, Guzman A M, Rabalais J W 1978 ACC Chem. Res. 11 170
- [84] Zhu Y, Wu C, Xu Z, Liu Y, Hu H, Guo T, Kim T, Chai Y, Li F 2021 Nano Lett. 21 14
- [85] Shan X, Wang Z, Lin Y, Zeng T, Zhao X, Xu H, Liu Y 2020

Adv. Electron. Mater. 6 2000536

- [86] Wang W, Gao S, Li Y, Yue W, Kan H, Zhang C, Lou Z, Wang L, Shen G 2021 Adv. Funct. Mater. 31 2101201
- [87] Yang L, Singh M, Shen S, Chih K, Liu S, Wu C, Chu C, Lin H 2020 Adv. Funct. Mater. 31 2008259
- [88] Fanselow M, Poulos A A 2005 Rev. Psychol. 56 207
- [89] Liu L, Cheng Z Q, Jiang B, Liu Y X, Zhang Y L, Yang F, Wang J H, Yu X F, Chu P K, Ye C 2021 ACS Appl. Mater. Interfaces 13 30797
- [90] Li Y, Long S, Liu Q, Lv H, Liu M 2017 Small 13 1604306
- [91] Zhang K, Meng D, Bai F, Zhai J, Wang Z L 2020 Adv. Funct. Mater. 30 2002945
- [92] Li Y, Zhong Y P, Deng Y F, Zhou Y X, Xu L, Miao X S 2013 J. Appl. Phys. 114 234503
- [93] Nguyen D A, Jo Y, Tran T U, Jeong M S, Kim H, Im H 2021 Small Methods 5 2101303
- [94] Shepherd R K, Shivdasani M N, Nayagam D A, Williams C E, Blamey P J 2013 Trends Biotechnol. 31 562
- [95] Kolb H 2003 Am. Sci. 91 28
- [96] Brady T F, Konkle T, Alvarez G A, Oliva A 2008 Proc. Natl Acad. Sci. USA 105 14325
- [97] Wang G, Wang R, Kong W and Zhang J 2018 Analysis. Cogn. Neurodyn. 12 615
- [98] Zhou F, Chen J, Tao X, Wang, X and Chai Y 2019 Research 2019 9490413
- [99] Radovic A, Williams M, Rousseau D, Kagan M, Bonacorsi D, Himmel A, Aurisano A, Terao K, Wongjirad T 2018 *Nature* 560 41
- [100] Sze V, Then Y, Emer J, Suleiman A, Zhang Z 2017 IEEE Custom Integrated Circuits Conference (CICC) 1
- [101] Xiang D, Liu T, Xu J, Tan J, Hu Z, Lei B, Zheng Y, Wu J, Neto A, Liu L, Chen W 2018 Nat. Commun. 9 1
- [102] Choi C, Leem J, Kim M S, Taqieddin A, Cho C, Cho K W, Lee G J, Seung H, Bae H J, Song Y M, Hyeon T, Aluru N R, Nam S W, Kim D H 2020 Nat. Commun. 11 1

- [103] McCollough C 1965 Science 149 1115
- [104] Huang X, Li Q Y, Shi W, Liu K, Zhang Y P, Liu Y W, Wei X F, Zhao Z Y, Guo Y L, Liu Y Q 2021 *Small* 17 2102820
- [105] Yuan J, Zhang Y, Zhou L, Zhang G, Yip H L, Lau T K, Lu X, Zhu C, Peng H, Johnson P A, Leclerc M, Cao Y, Ulanski J, Li Y, Zou Y 2019 Joule 3 1140
- [106] Seo S, Jo S H, Kim S, Shim J, Oh S, Kim J H, Heo K, Choi J W, Choi C, Oh S, Kuzum D, Wong H P, Park J H 2018 Nat. Commun. 9 5106
- [107] Wang S, Wang C Y, Wang P F, Wang C, Li Z A, Pan C, Dail Y T, Gao A Y, Liu C, Liu J, Yang H F, Liu X W, Cheng B, Chen K J, Wang Z L, Watanabe K J, Taniguchi T, Liang S J, Miao F 2020 Natl. Sci. Rev. 7 1
- [108] Euler T, Haverkamp S and Schubert T 2014 Nat. Rev. Neurosci. 15 507
- [109] van Hateren J H, Lamb T D 2006 Neuroscience 7 34
- [110] Wang Z, Li C, Lin P, Rao M Y, Nie Y Y, Song W H, Qiu Q R, Li Y N, Yan P, Strachan J P, Ge N, McDonald N, Wu Q, Hu M, Wu H Q, Williams R S, Xia Q F, Yang J J 2019 *Nat. Mach. Intell.* 1 434
- [111] Moon J, Ma W, Shin J H, Cai F X, Du C, Lee S H, Lu W D 2019 Nat. Electron. 2 480
- [112] Du C, Cai F, Zidan M A, Ma W, Lee S H, Lu W D 2017 Nat. Commun. 8 2204
- [113] Zhang Z H, Wang S Y, Liu C S, Xie R Z, Hu W D, Zhou P 2021 Nat. Nanotechnol. 17 27
- [114] Qin S, Wang F, Liu Y, Wan Q, Wang X, Xu Y, Shi Y, Wang X, Zhang R 2017 2 D Mater. 4 035022
- [115] Ahn J, Ko K, Kyhm J H, Ra H S, Bae H, Hong S, Kim D Y, Jang J S, Kim T W, Choi S, Kang J H, Kwon N, Park S, Ju B K, Poon T C, Park M C, Im S, Hwang D K 2021 ACS nano 15 17917
- [116] Xie D D, Yin K, Yang Z J, Huang H, Li X H, Shu Z W, Duan H G, He J, Jiang J 2022 Materials Horizons, 9 1448-1459

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Recent progress in optoelectronic memristive devices for in-sensor computing^{*}

> (Key Laboratory of UV-Emitting Materials and Technology of Ministry of Education, Northeast Normal University, Changchun 130024, China)

2) (National Demonstration Center for Experimental Physics Education, Northeast Normal University, Changchun 130024, China)

(Received 28 February 2022; revised manuscript received 28 March 2022)

Abstract

Neuromorphic computing system, inspired by human brain, has the capability of breaking through the bottlenecks of conventional von Neumann architecture, which can improve the energy efficiency of data processing. Novel neuromorphic electronic components are the hardware foundation of efficient neuromorphic computation. Optoelectronic memristive device integrates the functions of sensing, memorizing and computing and is considered as a promising hardware candidate for neuromorphic vision. Herein, the recent research progress of optoelectronic memristive device for in-sensor computing are reviewed, including optoelectronic materials and mechanism, optoelectronic memristive device/characteristics as well as functionality and application of in-sensor computing. We first review the optoelectronic materials and corresponding memristive mechanism, including photon-ion coupling and photon-electron coupling type. Then optoelectronic and all-optical modulated memristive device in cognitive function simulation, optoelectronic logic operation, neuromorphic vision, object tracking, etc. Finally, we summarize the advantages/challenges of optoelectronic memristor and prospect the future development.

Keywords: in-sensor computing, optoelectronic memristor, neuromorphic computing

PACS: 87.18.Sn, 73.40.Rw, 85.25.Hv, 85.60.-q

DOI: 10.7498/aps.71.20220350

^{*} Project supported by the fund from the Ministry of Science and Technology of China (Grant No. 2018YFE0118300) and the National Natural Science Foundation of China (Grant Nos. 11974072, U19A2091).

[†] Corresponding author. E-mail: wangzq752@nenu.edu.cn





Institute of Physics, CAS

仿生生物感官的感存算一体化系统

王童 温娟 吕康 陈健中 汪亮 郭新

Bio-inspired sensory systems with integrated capabilities of sensing, data storage, and processing Wang Tong Wen Juan Lü Kang Chen Jian-Zhong Wang Liang Guo Xin 引用信息 Citation: Acta Physica Sinica, 71, 148702 (2022) DOI: 10.7498/aps.71.20220281 在线阅读 View online: https://doi.org/10.7498/aps.71.20220281 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

N型局部有源忆阻器的神经形态行为

Neuromorphic behaviors of N-type locally-active memristor 物理学报. 2022, 71(5): 050502 https://doi.org/10.7498/aps.71.20212017

具有感觉记忆的忆阻器模型 Mathematical model of memristor with sensory memory 物理学报. 2019, 68(1): 018501 https://doi.org/10.7498/aps.68.20181577

铝基薄膜忆阻器作为感觉神经系统的习惯化特性

Al-based memristor applied to habituation sensory nervous system 物理学报. 2021, 70(6): 068502 https://doi.org/10.7498/aps.70.20201961

基于二维材料MXene的仿神经突触忆阻器的制备和长/短时程突触可塑性的实现

Fabrication of synaptic memristor based on two-dimensional material MXene and realization of both long-term and short-term plasticity

物理学报. 2019, 68(9): 098501 https://doi.org/10.7498/aps.68.20182306

专题: 面向类脑计算的物理电子学

仿生生物感官的感存算一体化系统*

王童 温娟 吕康 陈健中 汪亮 郭新†

(华中科技大学材料科学与工程学院,材料成形与模具技术国家重点实验室,武汉 430074)

(2022年2月17日收到; 2022年3月19日收到修改稿)

生物感官集感知、存储与运算为一体的架构使其可以高效并且实时地采集和处理外界信息,这样的感存 算一体化架构可为物联网时代面临的传感器数据爆炸问题提供很好的解决方案.为此,本文提出仿生生物感 官的感存算一体化系统,采用不同的传感器模拟生物感受器的功能,以获取环境信息,传感器输出的模拟信 号输入到模拟信号处理系统进行预处理,这样信号不需要在模拟域与数字域之间转换,可极大降低功耗和延 时;预处理后的信号输入类脑运算芯片中进行分析和决策,该芯片由基于忆阻器的人工突触及人工神经元组 成,通过控制突触与神经元的连接方式,可以实现不同的算法架构,如全连接脉冲神经网络、卷积脉冲神经网 络以及循环脉冲神经网络等;通过运行不同的神经网络,类脑运算芯片可以实现对不同传感器信号的识别、 预测以及分类等任务;更进一步,将多种仿生感觉系统的识别或预测结果结合起来,就可以实现多感官融合, 这样的系统架构可以用于自动驾驶及智能机器人等复杂的场景中.

关键词: 生物感官, 仿生感觉系统, 神经形态运算, 脉冲神经网络, 忆阻器
 PACS: 87.19.lt, 84.35.+i, 85.30.-z
 DOI: 10.7498/aps.71.20220281

1 引 言

人类的感觉主要有 8 种:视觉、听觉、触觉、嗅 觉、味觉、痛觉、温觉和本体感觉. David Julius 和 Ardem Patapoutian 发现了温度感受器和压力感 受器,为温度感知和机械感知提供了分子和神经学 基础,为此,他们荣获了 2021 年诺贝尔生理学或 医学奖. 生物的感觉过程可以大致概括为"接受刺 激-传递信号-大脑接收"3个步骤,感觉神经元的感 受器接受刺激,将环境信息转换为神经脉冲,神经 脉冲再通过突触传输到脑内的中枢神经系统,中枢 神经系统由复杂的神经网络组成,负责对神经脉冲 进行分析与计算. 生物感官把感知、存储以及运算 3个功能集成于一体,从而可以高效并且实时地采 集和处理外界信息^[1]. 因此,生物感官在处理很多 复杂任务时表现出低功耗、高效率等优势^[2].

单一的感官会造成感觉的局限与偏差,人体通 过不同感官中的感受器来感受不同的刺激,并将各 种感知信号融合在一起,从而获得对环境的多方面 体验,例如,在品尝美食时,至少需要同时使用视 觉、嗅觉和味觉来感受食物的"色、香、味".对于一 个特定的场景,通过不同的维度感知不同特性的模 态 (如图像-语音所对应的视觉-听觉),并进行数据 样本的收集和耦合,这就是多模态传感融合^[3].多 模态传感融合可以获得更全面准确的信息,增强感 知的可靠性与容错性.

相比于生物感官,现有的人工感觉和计算系统 采用了完全不同的架构,其中传感器、存储器和运 算单元是相互分离的,但是,这样的架构已经不再 适合万物互联的物联网时代⁴¹.一方面,随着物联 网的发展,越来越多的传感器被应用到生活的方方

* 国家重点研发计划 (批准号: 2018YFE0203802)、国家自然科学基金 (批准号: 61971202) 和华为技术合作项目 (批准号: YBN2019055139) 资助的课题.

© 2022 中国物理学会 Chinese Physical Society

[†] 通信作者. E-mail: xguo@hust.edu.cn

面面,这些传感器会产生巨大的、未经处理的原始 数据^[4].但是由于传感器与信号处理系统的分离, 这些由传感器产生的原始数据都需要首先转换为 数字信号,然后传递至本地或者云端的信息处理系 统.这样的架构会带来额外的功耗、较长的响应时 间、大量的数据存储、带宽以及数据安全等问题^[6]. 另一方面,存储与运算单元分离的冯·诺依曼架构,不 再适应大数据、智能化时代的计算要求.数据在内 存和数据处理单元之间的频繁存取操作带来了高 功耗、高延时等问题,使得基于冯·诺依曼架构的计 算系统运算效率远低于人脑^[6].因此,需要探索新 的神经形态计算方式来帮助神经网络应用到有功 耗限制的场景中去,如用于边缘计算的终端设备^[7].

为了模拟生物存储与运算一体的架构,研究人员探索了多种新型存储器件,如忆阻器、铁电存储器、磁存储器等.其中,忆阻器具有并行计算的能力,而且其性能与生物神经元和突触相类似,是构建存算一体化系统的有力之选,这样的新型存算一体化系统可以用来构建脉冲神经网络 (spiking neural network, SNN)^[8,9].为了模拟生物感知与运算一体的架构,将模拟信号处理单元集成到传感器

系统中,使其可以处理简单的任务,如降噪、特征 提取或特征增强等.模拟信号处理单元可以有效地 减少数据的传输,降低功耗和延时,简化后续信号 处理过程^[1].此外,为了实现实时环境检测与识别, 研究人员尝试将感知、存储与运算系统集成于一 体,使得终端设备可以具有类似生物感官系统的环 境感知能力^[10],目前,这样的处理系统还处于起步 阶段,对信息的处理能力也比较有限.

本文提出一种仿生生物感官的通用架构.如 图 1(a)所示,在生物感官中,感受器接受外界环境 刺激,引起感觉神经元膜电位变化,当外界刺激达 到阈值后,感觉神经元发出神经脉冲;感觉神经元 发出的神经脉冲经过突触传递至中枢神经网络进 行信息提取、识别等工作.图 1(b)给出了仿生生物 感官的感存算一体化系统,该系统模拟了生物感官 的工作方式,首先由不同的传感器来模拟生物感受 器,检测环境中的气体、光、震动、压力等信号;传 感器信号经模拟信号处理单元处理及编码后直接 输入类脑芯片,类脑芯片为通用人工智能架构,由 人工突触和人工神经元组成,通过不同的神经元与 突触连接方式,可以运行不同的神经网络模型,如



图 1 生物感官及仿生感觉系统 (a)生物感官包括感觉神经元、突触及生物神经网络.感觉神经元的感受器感知环境刺激,并 引起神经元膜电位变化,当神经元膜电位达到阈值后,产生神经脉冲,神经脉冲信号通过突触传递至神经网络进行处理.(b)仿 生感觉系统包含了传感器、模拟信号处理单元以及类脑运算芯片.传感器负责感知环境信息,模拟信号处理单元负责对传感器信 号进行预处理,类脑芯片负责特征提取、分类以及预测等高级任务

Fig. 1. Biological sensory organ and bio-inspired sensory system. (a) Biological sensory organ includes sensory neurons, synapses and neural network. The receptors of the sensory neurons capture the environmental information and cause the increase of the neuron potential. The neurons generate neural spikes when the neuron potential exceeds its threshold. The neural spikes are then transmitted to the neural network through synapses. (b) Bio-inspired sensory system includes sensors, analogue signal processing units and the brain-inspired chip. Sensors acquire the environmental information and convert the information into analogue electrical signals, which are then directly processed by the analogue signal processing units. Finally, the brain-inspired chip receives the pre-processed signals, and performs advanced tasks, such as feature extraction, classification and prediction. 全连接脉冲神经网络、卷积脉冲神经网络和循环脉 冲神经网络等.在复杂的环境中运行时,需要综合 考虑多种感官的输入信号之后做出决策,通过对各 种感官信号进行识别以及对预测结果进行综合处 理,可以模拟生物感官的信息融合,这样的系统可 用于自动驾驶、智能机器人、生物医疗等领域.

此外,有研究人员^[1]提出使用忆阻器本身的感知能力,如视觉(光信号)、嗅觉(气体)等,来实现传感器内计算(in-sensor computing),但目前还没有取得明显的进展."传感器内计算"对材料和器件结构有非常大的限制,这是因为忆阻器和传感器对材料和器件结构有不同、甚至是矛盾的要求.以嗅觉模拟为例,器件必须同时具有气敏性能和忆阻性能,那么可选的材料只有 WO₃,ZnO 等极少数几种,忆阻器要求氧化物薄膜致密,且厚度<100 nm,而传感器则要求氧化物薄膜疏松多孔,以提供更多的气体接触位点来进行气敏反应,厚度为数百微米,甚至更厚^[11].考虑这些因素,只有在某些特定的条件下才可能实现"传感器内计算".当然,"传感器内计算"具有极高的集成度,在功耗和器件体

积等方面具有无与伦比的优势, 也是研究人员积极 探索的方向.

2 传感器及模拟信号处理单元

如图 2 所示, 生物体中分布着不同的感觉系统, 这些系统帮助生物获取和处理信息、寻找食物、规避风险等. 生物个体通过分布在各器官中的感觉神经元将外界信号转变为神经脉冲信号, 并通过突触将神经脉冲信号传递至大脑皮层中的神经中枢进行运算和处理. 不同的感官拥有不同的感受器来感知环境信息, 在嗅觉系统中, 分布在鼻腔内的嗅觉感觉神经元的嗅感受器可以与不同的气体分子相结合, 将化学信号转变为细胞膜电位变化, 当刺激达到阈值之后, 该嗅觉神经元产生神经脉冲并通过突触传递至嗅球细胞^[12]; 在视觉系统中, 视网膜起到视觉感受器的作用, 将光信号转变为电信号, 随后通过视神经传递至大脑皮层, 此外, 人类视网膜还具备预处理功能, 可以提取输入视觉信息的关键特征, 消除无用的输入信息, 加速后续神经





Fig. 2. Sensors and analogue signal processing units. Sensors work like the receptors in a biological sensory organ, detecting the environmental information such as gases, light, audio and pressure, and convert them to electrical sensory signals. The sensory signals are then directly processed by the analogue signal processing units.

网络的信息处理进程^[13];在听觉系统中,耳蜗基底 膜上的柯蒂斯感受器可以将震动转化为电信号传 递到人的听觉神经中^[14];在触觉系统中,皮肤上分 布着不同的触觉感受器,它们对机械能某些特定的 特征敏感,从而将施加于皮肤上不同类型的机械能 转化为不同频率的神经脉冲^[15],由于传入到大脑 皮层的脉冲频率不同,生物可感知到物体的软硬程 度、大小、形状等信息^[10].

随着材料科学的进步,不同种类的传感器被开 发出来,用于检测环境信息,也可以用于模拟生物 感受器的功能.然而,传感器信号处理系统需要将 模拟信号转变为数字信号,再传输至信号处理系统 中进行处理及运算,这个过程需要采用模数转换 器 (analogue-digital convertor, ADC)^[1].这种将模 拟信号转变为数字信号再处理的方式会带来较大 的功耗及延时.而在生物感官中,感受器感受到环 境信息之后,产生动作电位直接引起细胞膜电位变 化,不需要将信息进行转化,而是直接被神经元处 理.因此,借鉴生物系统,可以设计模拟信号处理 单元,直接与传感器相连,在模拟域对信号进行预 处理,再由通用的类脑运算芯片完成信号识别及预 测等高级任务.

下面通过对嗅觉的模拟来解释环境信号的获 得与处理过程. 用气体传感器阵列来模拟嗅觉感受 器的功能,将气体分子的化学信号转换为模拟型的 电信号.为了获得气体传感器对不同气体的响应信 号,采用不同的材料制备气体传感器以增加输入信 号的区别,因此,气体传感器往往有着不同的基线 电阻,且不同器件对不同气体的响应呈现出数量级 的差别,此外,在工作过程中,气体传感器的基线 电阻会随时间漂移^[16].为此,可采用气体传感器基 线动态校准电路在模拟域中高效地处理气体传感 器信号17].针对光学传感器的信号处理,可采用模 拟电路实现类似视网膜的图像边缘提取及特征增 强功能,相比于传统的数字信号处理方式,这样的 模拟电路集成度更高,还可以把功耗降低至毫瓦级[18]. 针对语音传感器的信号处理方法,研究人员模仿耳 蜗的柯蒂斯感受器设计了带通滤波器来处理声音 传感器信号,不需要将传感器信号转变为数字信 号,再用傅里叶变换进行处理,这样的设计既降低 了功耗,又减小了电路面积[14].采用压力传感器可 以模拟人的触觉,感受环境中的压力信息,压力传 感器的信号一般需要采用半波整流的方式进行预 处理.另外,传感器信号还要进行统一的降噪、滤波、放大和编码等预处理后再输入到类脑运算芯片中^[1].

信号预处理也广泛地应用于现有的传感器信 号处理模块中,首先也需要对模拟信号进行降噪、 滤波、放大等预处理,经预处理的模拟信号经过模 数转换后,输入到中央处理器或者传输到云端设 备.在这个过程中,模数转换器需要消耗大量的功 耗和芯片面积.此外,大量的数据传输也会产生功 耗和延时.因此,本文提出一个完全工作在模拟域 的模拟信号处理单元,与现有的传感器信号处理系 统相比,模拟信号处理单元在功耗与延时等方面均 有优势.

3 可重构类脑芯片

人脑中有大量的神经元 (约 1011 量级) 及突触 (约1015量级),这些神经元与突触连接起来组成庞 大的神经网络^[7]. 相比于基于现有计算平台 (central processing unit, CPU; graphic processing unit, GPU)的人工神经网络,人脑中的生物神经网络可 以完成各种感官信号的处理、复杂的记忆及快速的 决策,且仅消耗较少的能量^[19].因此,人们提出了 脉冲神经网络,也被称为第三代人工神经网络.它 模拟生物神经网络的信息传递及处理方式,采用神 经元来接收并处理脉冲信号,采用突触来存储、处 理和传递脉冲信息,突触的权重可由前后神经元发 放的脉冲进行调制^[20].脉冲神经网络是基于事件 来运算的,当有事件发生的时候,神经元才发出脉 冲,而没有事件的时候神经元则处于静息状态.这 样的工作方式更加适合处理传感器相关的信号,因采 用离散的脉冲信号进行运算,会极大地降低功耗[21].

通过传统的 CMOS (complementary metal oxide semiconductor) 电路可以实现大规模的人工突 触及人工神经元,进而实现脉冲神经网络,例如, IBM 设计并研制了 TrueNorth 芯片, Intel 设计并 研制了 Loihi 芯片,它们模拟了神经元和突触对脉 冲的响应,并且具备并行运算的能力,可以降低推 理及训练所需的功耗^[22].但是,在这些芯片中,每 个神经元或者突触都需要采用数十个 CMOS 器件 来实现^[23],因此,人们尝试采用忆阻器来实现神经 元和突触的功能.忆阻器的电导可以随着施加的激 励而连续变化,可以用来模拟突触可塑性及神经元 膜电位的连续变化^[24].

3.1 基于忆阻器的人工神经元

在神经科学中,神经元接受上级神经元经突触 传入的兴奋性或抑制性突触电位, 脂质双分子层细 胞膜的电位则会发生相应变化, 当累积超过阈值时 产生动作电位,再经突触传向下一级神经元,这是 神经元处理和传递信息的过程,为了解释牛物神经 元的工作模式,科学家们相继提出了 Hodgkin-Huxlev(HH) 模型和漏电整合发放 (leaky integrate-andfire, LIF) 模型等^[25]. 忆阻器电导的变化与神经元 细胞膜上离子通道的电导类似,因此,可以基于这些 数学模型来模拟神经元功能. 美国 HRL Laboratories [26] 利用忆阻器、电容器和电阻搭建了简单的 忆阻神经元电路,其中忆阻器模拟离子通道,电容 器两端的电压模拟神经元细胞膜上的电位,他们利 用该电路成功地模拟了神经元中动作电位产生的 过程.美国南加州大学杨建华教授课题组^[23]利用 扩散型忆阻器 (diffusive memristor) 和电容器并联 后再串联一个电阻组成了神经元电路(图 3(a)),并 联电容器上的电压表示细胞的膜电位,有脉冲时, 电荷在电容器中累积,发生整合;没有脉冲时,电 容器发生自发的漏电.一旦电容器电压高于忆阻器 的阈值,该忆阻器切换到高电导态,此时电容器的 电压无法保持,开始发放电流脉冲,这成功模拟了 神经元漏电整合发放功能.此外,杨建华课题组[27] 还实现了一种基于扩散型忆阻器的人工痛觉感觉 神经元,不仅模仿了生物处理信号的方式来处理传 感器传来的信号,还成功地模拟了神经元的关键功 能,包括阈值、弛豫和自适应等,神经元阈值调节 的高级功能有助于实现深度脉冲神经网络, Zhang 等^[28] 不仅采用阈值转变忆阻器实现了 LIF 神经元 的功能,还设计了结构复杂的电路实现了阈值可调 功能: Shaban 等^[29] 采用了更加简洁的方式实现了 阈值调节功能并完成了单层脉冲神经网络的设计 验证.

3.2 基于忆阻器的人工突触

如图 3(b) 所示,突触是神经元之间的连接节 点,每个突触具有一定的连接权值,根据权值的大 小确定两个神经元之间的信号传输效率,突触连接 权值可以根据突触前/后神经元的活动进行调节, 这被称为突触可塑性 (synaptic plasticity)^[7],是大 脑学习的基础.单个忆阻器在结构和功能上都表现 出与生物突触的相似性.在结构上,具有金属/绝 缘体/金属三明治结构的忆阻器是典型的两端子器 件,而突触和相应的前后神经元也同样形成两端子 结构;在功能上,忆阻器的电导变化可代表突触连 接强度的变化,在持续的外部电刺激下,器件电导 表现出连续增加或降低,这类似于突触之间连接强 度的增强或减弱.通过对忆阻器材料与结构的设计 可以实现忆阻器电导态的连续可调,即器件电导在



图 3 类脑芯片 (a) 人工神经元:采用忆阻器与电容并 联的结构模拟神经元膜电位连续变化的特性,实现神经元 漏电流积分点火的功能; (b) 人工突触:非易失性忆阻器电 导的连续变化可以模拟生物突触权重的连续变化及突触 可塑性,此外,忆阻器阵列还可以实现存内运算; (c) 类脑 芯片架构:由多个神经元-突触核心组成,每个核心由基于 忆阻器的人工神经元及人工突触组成,每个核心通过总线 连接

Fig. 3. Brain-inspired chip. (a) Artificial neuron, consisting of a memristor and a parallel capacitor, emulates the continuous change of the neuron potential and realize the leaky integrate-and-fire function of a biological neuron. (b) Artificial synapses. Non-volatile memristors emulate the continuous change of the synaptic plasticity, in addition, the crossbar array based on memristors can realize the in-memory computing. (c) Architecture of the brain-inspired chip. A brain-inspired chip consists of many neuron-synapse cores based on memristors, and the cores are connected by bus lines. 外加电脉冲的作用下连续可调. 基于多种材料如金 属氧化物、二维材料和高分子材料的忆阻器都成功 模拟了突触的可塑性^[19,30]. 如图 3(b) 所示, 将忆阻 器制成阵列后, 不仅可以模拟单个突触的性能, 还 可以实现存内运算. 根据欧姆定律和基尔霍夫定 律, 输入的电压信号 (*V*₁, *V*₂, *V*₃, …, *V*_n) 与阵列中 忆阻器的电导 (*G*) 相乘, 得到输出的电流信号 (*I*₁, *I*₂, *I*₃, …, *I*_n), 因此, 这种交叉阵列的结构可以在一 个时间周期内完成矩阵运算, 从而提高人工神经网 络的运算速度^[31].

3.3 基于忆阻器的类脑芯片及其计算架构

类脑芯片架构如图 3(c) 所示, 芯片采用了多 分布式人工神经元-人工突触核心的设计,这些通 用的人工神经元-人工突触核心可以通过重构来处 理不同的任务,核心之间通过可重构的片上网络方 式进行连接,经过模拟信号处理单元预处理之后的 信号可以通过 I/O 口输入到芯片中,并分配对应 的核心进行运算.神经元-突触核心由基于忆阻器 的人工神经元、忆阻突触阵列、数据选择器 (multiplexer, MUX)、缓冲区、特殊功能单元和控制器构 成. 人工神经元根据漏电流积分点火的功能处理来 自传感器或突触阵列的模拟信号,并将模拟信号转 变为脉冲信号输出. 忆阻突触阵列根据自身权重传 递人工神经元发出的脉冲信号,其突触权重则根据 突触可塑性由前后神经元脉冲发放情况进行调节. 为了提高芯片的容错率,突触阵列及人工神经元应 增加冗余.数据选择器用来选择固定区域的突触阵 列及人工神经元来完成某项任务. 特殊功能单元 (special function unit, SFU)则提供突触训练所需 要的计算、神经元阈值调节、神经元特殊互连结构 等功能. 这种基于人工神经元-人工突触内核的计 算架构将存储和计算集为一体,消除了由存储器 (memory)和中央处理单元 (CPU) 之间的性能差 异带来的效率损失,且基于脉冲的信息编码方式也 能够大大减少运算过程中的能量损耗.

4 脉冲神经网络算法

传统人工神经网络所具备的神经元连接方式 也可以在脉冲神经网络中实现,将这些网络架构应 用到类脑运算芯片上,最终可以实现对不同传感器 信号的识别、分类、预测等任务,例如,采用全连接 脉冲神经网络可以实现气体分类等任务,而采用卷 积脉冲神经网络可以实现图片特征提取和分类等 任务^[32].

图 4(a) 给出了双层的全连接脉冲神经网络, 其中输入层为简单的 LIF 神经元, 它可以对输入 脉冲信号进行运算并产生新的脉冲信号;突触根据 自身权重传递脉冲信号到后级神经元;输出层的神 经元是阈值可调的 LIF 神经元, 它不仅具备 LIF 神经元的基本功能,还具备神经元不应期和侧向抑 制的高级功能,这些高级功能可以通过神经元-突 触核心中的特殊功能单元来实现. 如图 4(a) 所示, 神经元不应期是指当输出层第一个神经元被激发 之后, 它会传递给自己一个信号, 在短时间内即使 有较大的输入信号,也不会被再次激发.神经元不 应期可以给该层其他神经元提供公平的竞争机会, 使神经元尽可能学习到更多的特征. 侧向抑制功能 指的则是当神经元激发后,会给当前层其他神经元 一个抑制信号,短时间内调高其他神经元的阈值^[28]. 采用神经元侧向抑制功能可以实现无监督学习算 法, 例如, "赢者通吃 (winner-takes-all)"法则^[21]. 通 过读取输出层神经元的脉冲发放情况,就可以判断 分类和识别结果.

可以用脉冲时间依赖可塑性 (spike-timing dependent plasticity, STDP) 或者脉冲频率依赖 可塑性 (spike-rate dependent plasticity, SRDP) 来 训练全连接脉冲神经网络. 如图 4(b) 所示, STDP 指的是根据突触连接前后神经元的脉冲发放时间 来更改突触权重,如果突触前神经元发放脉冲早于 突触后神经元,则增强突触连接权重,反之则降低 突触连接权重. SRDP 则是指根据突触前后神经元 脉冲发放频率来对突触连接权重进行修改,如果突 触前后神经元脉冲发放频率较高,则增强突触的连 接权重;如果仅有突触前神经元或突触后神经元发 放脉冲的频率较高,则降低突触的连接权重;如果 突触前后神经元发放脉冲的频率都较低,则不对突 触权重进行修改.相比于传统的反向传播训练法 则,这种通过突触可塑性对网络进行训练的学习法 则在硬件电路中更容易实现,可以降低硬件电路的 复杂度[20].

脉冲神经网络可以实现在线 (online) 和离线 (offline) 的两种训练方式. 离线学习的具体实施方 式如下:首先通过仿真计算得到该脉冲神经网络所 需的突触权重,再通过施加脉冲将突触权重调节至



图 4 脉冲神经网络 (a) 全连接脉冲神经网络; (b) 基于突触可塑性(脉冲时间依赖可塑性(STDP)和脉冲频率依赖可塑性(SRDP)) 来训练网络,修改突触权重; c) 卷积脉冲神经网络,由卷积层、池化层和全连接层组成

Fig. 4. Spiking neural network (SNN): (a) Fully connected SNN; (b) learning rules of SNN. The synaptic weights in SNN are updated according to STDP and SRDP; (c) convolutional SNN consisting of convolutional layers, pooling layers and fully connect layers.

计算所需的状态. 在此, 该脉冲神经网络可以采用 STDP 或 SRDP 等兼容硬件的学习规则进行训练, 也可以采用反向传播等更加精确的训练算法获得 突触权重. 尽管离线学习可以获得更加精准的突触 权重, 但是这种方式无法使用忆阻突触阵列来为训 练过程加速. 因此, 我们更期望使用该系统去完成 online 的训练, 并提出采用与硬件兼容的 STDP 和 SRDP 学习规则. 以 SRDP 为例, 采用两个计数 器分别记录输入神经元和输出神经元的脉冲发放 次数, 进而计算脉冲发放频率. 如果输出神经元计 数器达到阈值, 则根据输入神经元的脉冲频率来修 改突触权重; 如果输入神经元脉冲高于阈值, 则施 加正向脉冲到突触器件上, 反之则施加负向脉冲到 突触器件. 经过学习后, 每个神经元会学到对应的 输入样本的特征.

对于比较复杂的问题,如图片特征提取及识别,可以采用卷积脉冲神经网络.如图 4(c)所示,卷 积脉冲神经网络包括卷积层、池化层及全连接层^[33]. 传感器信号经过编码后变为脉冲信号输入到卷积 层,在卷积层中,突触阵列根据自身权值传递输入 的脉冲信号至阈值可调的神经元,该神经元负责对 接收到的信息进行处理并发放新的动作电位.池化 层模仿人的视觉系统对数据进行降维,用更高层次 的特征表示输入的图像,常用的有最大值池化、均 值池化等.在卷积脉冲神经网络中可以对卷积层神 经元发放脉冲频率进行记录并进行池化操作,因 此,这里可以采用简单的数字电路来实现池化操 作.卷积层及池化层可以进行多层的堆叠以实现深 度神经网络,输入信号经过卷积层及池化层提取特 征之后,可以输入到全连接层进行识别和分类,得到 最终的结果.卷积脉冲神经网络也可以采用 STDP 及 SRDP 的训练方式逐层进行训练.

此外, 在处理语音信号时, 目前常用的网络算法 是循环神经网络 (recurrent neural network, RNN). 相比于前馈式的神经网络 (全连接神经网络以及卷 积神经网络), 神经元节点之间的循环连接可以有 效地处理不同时间节点的信息, 并将时间上的逻辑 关系展现出来. 在脉冲神经网络中, 也可以借鉴循 环神经网络的连接方式, 对时间信息进行处理^[34]. 除了采用神经元之间循环连接的架构, 还有学者提 出采用神经元自身的时域动态特性来对时间信号 进行处理, 这也验证了神经元的 LIF 功能可以等 价替代神经元的循环连接结构^[35]. 但是, 采用神经 元自身动态特性来处理时间信号的方式依然处于 理论验证阶段, 神经元对时序信息的处理仍然需要 更多的研究.

在算法层对硬件系统进行调用可以实现多感 官的融合.针对本文提出的仿生生物感官系统有两 种实现多感官融合的可能方式: 第一种是将多种感 官的信号经过模拟信号处理单元预处理及编码后 一起输入到脉冲神经网络中,在这种情况下,多种 感官的信号增加了输入信息的维度,也极大地增加 了神经网络的复杂度. 例如, 在机器人控制系统中, 可以将压力传感器信号与视觉传感器信号一起输 入到脉冲神经网络中进行训练,从而得到决策结 果. 第二种处理方式是将不同的传感器信号分别输 入到对应的神经网络中,这些神经网络已经针对各 自对应的传感器信号类型做了优化,处理后的信号 再输入到一个新的决策系统中,由决策系统综合不 同的感官信号给出最终的结果. 比较这两种实现多 感官融合的方式,第一种方式处理信息的能力更 强,但需要考虑信号同步、延时等问题,因此,所需 的脉冲神经网络规模会较大: 而第二种方式需要新 的决策系统,比如中央处理器 (CPU) 或者云端设 备来实现.

5 仿生嗅觉系统

作为实例,图 5 展示了我们近期研发的一个仿 生嗅觉系统^[36].具体来讲,首先采用气体传感器检 测环境中的气体分子,传感器中的敏感材料可以吸 附气体分子,从而电阻发生变化,并在一定时间后 达到稳态.因此,气体传感器可以将环境中气体分 子的化学信息转变为模拟型的电信号.模拟信号处 理系统负责对传感器信号进行预处理,首先需要采 用气体传感器基线动态校准电路来解决传感器基 线漂移的问题;然后采用运算放大电路对传感器信 号进行放大. 经过预处理的传感器信号最终传输至 类脑芯片中的神经元-突触核心中进行识别. 我们 采用双层全连接脉冲神经网络来处理传感器采集 到的信号,并识别出不同的气体.在类脑芯片中, 传感器信号 (电压信号)首先输入到输入层神经元 中, 输入层神经元为简单的 LIF 神经元, 可以将输 入的电压信号转变为不同频率的脉冲信号. LIF 神 经元的输出信号经过放大后输入至突触阵列中,突 触阵列根据自身权值对输入信号进行计算,根据欧 姆定律,在一个时间步长内,就可以得到输出电流. 电流信号首先被转变为电压信号,再经过放大后被 输入至输出层神经元中,输出层神经元为具有阈值 调节功能的 LIF 神经元, 可以处理输入的脉冲信 号并得到识别结果. 采用阈值可调的 LIF 神经元 可以实现无监督的学习算法,采用 STDP 或 SRDP 的方式来对突触阵列进行训练. 在突触权重更新过 程中,突触阵列与神经元之间的连接断开,由特殊 功能单元来完成突触训练脉冲的发放.

该仿生嗅觉系统展现出了强大的鲁棒性和泛 化能力,能够完全识别4种待测挥发性有机化合 物(甲醛、乙醇、丙酮和甲苯).仅采用24个气体训 练样本训练该网络,就能够完全识别4种气体的 64个测试样本.尽管传感器在获得信号的时候,基 线值存在漂移现象并且受到环境噪声的影响,该脉 冲神经网络依然获得了完美的识别结果^[36].采用 这样架构的感存算一体化的嗅觉系统可以高效地 完成对不同气体的检测与识别.由于采用事件驱使





Fig. 5. Bio-inspired olfactory system. In the system, sensor array detects the gas molecules in environment, converts the chemical information of gas molecules into electrical sensory signals, which are analogue. The sensory signals are directly processed by the analogue signal processing units which include baseline cancellation and amplifier circuits. After pre-processing, the sensory signals are sent to a neuron-synapse core of the brain-inspired chip to classify the gases.

的脉冲神经网络,因此,该系统具备低功耗的特点, 可以集成到终端系统之中.

6 总 结

对生物感官的模拟可以帮助我们构建集感知、 存储与运算为一体的系统,从而高效地处理传感器 信号,解决物联网时代传感器数据爆炸的问题.首 先采用不同的传感器模拟生物的鼻子、眼睛、耳朵 及皮肤中的感受器,感知环境中的气体、光、声音 震动及压力等信息.然后采用模拟信号处理单元直 接对传感器信号进行处理和编码,以获得类脑运算 芯片的输入信号.类脑运算芯片通过人工神经元与 人工突触不同的连接方式来实现不同的脉冲神经 网络算法,用于完成信号识别及预测等高级任务, 最终通过对各个仿生生物感官信号的综合处理,完 成多种感官信息的融合.多感官融合的仿生感觉系 统可以用于自动驾驶、智能机器人等需要多种感官 协调工作来处理复杂信号的环境中,且仿生感觉系 统的功耗可得到极大的降低.

参考文献

- [1] Zhou F, Chai Y 2020 Nat. Electron. 3 664
- [2] Wang T, Huang H M, Wang X X, Guo X 2021 $\mathit{InfoMat}$ 3 804
- [3] Pei J, Deng L, Song S, Zhao M, Zhang Y, Wu S, Wang G, Zou Z, Wu Z, He W, Chen F, Deng N, Wu S, Wang Y, Wu Y, Yang Z, Ma C, Li G, Han W, Li H, Wu H, Zhao R, Xie Y, Shi L 2019 Nature 572 106
- [4] LaValle S, Lesser E, Shockley R, Hopkins M S, Kruschwitz N 2011 MIT Sloan Manag. Rev. 52 21
- [5] Shulaker M M, Hills G, Park R S, Howe R T, Saraswat K, Wong H S P, Mitra S 2017 Nature 547 74
- [6] Cai F, Correll J M, Lee S H, Lim Y, Bothra V, Zhang Z, Flynn M P, Lu W D 2019 Nat. Electron. 2 290
- [7] Huang H M, Wang Z, Wang T, Xiao Y, Guo X 2020 Adv. Intell. Syst. 2 2000149
- [8] Prezioso M, Merrikh-Bayat F, Hoskins B D, Adam G C, Likharev K K, Strukov D B 2015 Nature 521 61
- [9] Shao Z Y, Huang H M, Guo X 2021 Solid State Ionics 370 115746
- [10] Zhang X, Zhuo Y, Luo Q, Wu Z, Midya R, Wang Z, Song W, Wang R, Upadhyay N K, Fang Y, Kiani F, Rao M, Yang Y, Xia Q, Liu Q, Liu M, Yang J J 2020 *Nat. Commun.* 11 51

- [11] Lee D, Yun M J, Kim K H, Kim S, Kim H D 2021 ACS Sens. 6 4217
- [12] Imam N, Cleland T A 2020 Nat. Mach. Intell. 2 181
- [13] Hartridge H 1947 Nature 160 538
- [14] Hasler P, Smith P D, Graham D, Ellis R, Anderson D V 2005 IEEE Sens. J. 5 1027
- [15] Kim Y, Chortos A, Xu W, Liu Y, Oh J Y, Son D, Kang J, Foudeh A M, Zhu C, Lee Y, Niu S, Liu J, Pfattner R, Bao Z, Lee T W 2018 Science 360 998
- [16] Fonollosa J, Fernández L, Gutiérrez-Gálvez A, Huerta R, Marco S 2016 Sens. Actuators, B 236 1044
- [17] Kucic M, Hasler P, Dugger J, Anderson D 2001 Proceedings 2001 Conference on Advanced Research in VLSI. ARVLSI 2001 Salt Lake City, UT, March 14-16, 2001 pp148–162
- [18] Rüedi P F, Heim P, Kaess F, Grenet E, Heitger F, Burgi P Y, Gyger S, Nussbaum P 2003 IEEE J. Solid-State Circuits 38 2325
- [19] Yang R, Huang H M, Guo X 2019 Adv. Electron. Mater. 5 1900287
- [20] Wang W, Pedretti G, Milo V, Carboni R, Calderoni A, Ramaswamy N, Spinelli A S, Ielmini D 2018 Sci. Adv. 4 eaat4752
- [21] Lobo J L, Del Ser J, Bifet A, Kasabov N 2020 Neural Networks 121 88
- [22] Zhang W, Gao B, Tang J, Yao P, Yu S, Chang M F, Yoo H J, Qian H, Wu H 2020 Nat. Electron. 3 371
- [23] Wang Z, Joshi S, Savel'ev S, Song W, Midya R, Li Y, Rao M, Yan P, Asapu S, Zhuo Y, Jiang H, Lin P, Li C, Yoon J H, Upadhyay N K, Zhang J, Hu M, Strachan J P, Barnell M, Wu Q, Wu H, Williams R S, Xia Q, Yang J J 2018 Nat. Electron. 1 137
- [24] Xia Q, Yang J J 2019 Nat. Mater. 18 309
- [25] Huang H M, Yang R, Tan Z H, He H K, Zhou W, Xiong J, Guo X 2019 Adv. Mater. 31 1803849
- [26] Yi W, Tsang K K, Lam S K, Bai X, Crowell J A, Flores E A 2018 Nat. Commun. 9 4661
- [27] Yoon J H, Wang Z, Kim K M, Wu H, Ravichandran V, Xia Q, Hwang C S, Yang J J 2018 Nat. Commun. 9 417
- [28] Zhang X, Lu J, Wang Z, Wang R, Wei J, Shi T, Dou C, Wu Z, Zhu J, Shang D, Xing G, Chan M, Liu Q, Liu M 2021 *Sci. Bull.* 66 1624
- [29] Shaban A, Bezugam S S, Suri M 2021 Nat. Commun. 12 4234
- [30] Wan C, Cai P, Wang M, Qian Y, Huang W, Chen X 2020 Adv. Mater. 32 e1902434
- [31] Yao P, Wu H, Gao B, Tang J, Zhang Q, Zhang W, Yang J J, Qian H 2020 Nature 577 641
- [32] Pfeiffer M, Pfeil T 2018 Front. Neurosci. 12 774
- [33] Roy K, Jaiswal A, Panda P 2019 Nature 575 607
- [34] Bellec G, Scherr F, Subramoney A, Hajek E, Salaj D, Legenstein R, Maass W 2020 Nat. Commun. 11 3625
- [35] Ponghiran W, Roy K 2021 arXiv: 2109.01905 [cs.NE]
- [36] Wang T, Wang X, Wen J, Shao Z Y, Huang H M, Guo X 2022 Adv. Intell. Syst. 2200047

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Bio-inspired sensory systems with integrated capabilities of sensing, data storage, and processing^{*}

Wang Tong Wen Juan Lü Kang Chen Jian-Zhong Wang Liang Guo Xin[†]

(State Key Laboratory of Material Processing and Die & Mould Technology, School of Materials Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

(Received 17 February 2022; revised manuscript received 19 March 2022)

Abstract

In current sensing-computing systems, sensors are used to acquire information from environments, such data are normally analogue, unstructured and even redundant. After the analogue-to-digital conversion (ADC), the data are transferred into digital computers for processing. In computers with the von Neumann architecture, memories and central processing units (CPUs) are physically separated. Such a separation of sensing terminals, memories and CPUs yields serious problems, such as high energy consumption, long response time, huge data storage, and stringent requirements for the communication bandwidth and security. However, time- and energy-efficient ways are urgently required to process information at where data are generated. On the other hand, biological sensory organs respond to external stimuli in real-time with high efficiency due to the integrated capabilities of sensing, memory and computing. Therefore, the problem of sensing units, memories and processing units can be solved by emulating biological sensory organs.

In this work, we propose bio-inspired sensory systems with integrated capabilities of sensing, data storage and processing. In such a system, different sensors are used to capture the environmental signals from e.g. gases, light, audio and pressure, then the sensory signals are processed by an analogue signal processor, so that the energy-consuming ADC is avoided, afterwards the sensory signals are processed by a brain-inspired chip which consists of neuron-synapse cores based on memristors. In the neuron-synapse cores, leaky integrate-and-fire (LIF) neurons can be implemented by memristors and capacitors, and adaptive LIF neurons are developed from the LIF neurons to realize unsupervised learning algorithms. The synapses are realized by memristor arrays which can also perform the in-memory computing. By changing the connection between the neurons, the braininspired chip can realize different spiking neural networks (SNNs), such as fully connected SNN, convolutional SNN, and recurrent SNN. The synaptic weight in SNNs can be updated according to the spike-timing dependent plasticity (STDP) or the spike-rate dependent plasticity (SRDP). As an example, a bio-inspired olfactory system is demonstrated. In a artificial olfactory system, a sensor array detects and transforms the chemical information about gas molecules into electrical sensory signals. Then the sensory signals are processed by the analogue signal processing unit. After pre-processing, the brain-inspired chip classifies gases by constructing a fully connected SNN with two layers. Such a bio-inspired olfactory system emulates the function of a biological nose, overcoming the low efficiency caused by the frequent sampling, data conversion, transfer and storage under the current sensing-computing architecture. More importantly, the approach of this work can be used to emulate almost all the biological perceptions, such as touch, sight, hearing and taste, through the integration with different types of sensors., Therefore, this work offers a brand new approach to realizing the artificial intelligence (AI).

Keywords: biological sensory organ, bio-inspired sensory system, neuromorphic computing, spiking neural network, memristor

PACS: 87.19.lt, 84.35.+i, 85.30.-z

DOI: 10.7498/aps.71.20220281

^{*} Project supported by the National Key R&D Program of China (Grant No. 2018YFE0203802), the National Natural Science Foundation of China (Grant No. 61971202), and the Huawei Technical Cooperation Project, China (Grant No. YBN2019055139).

[†] Corresponding author. E-mail: xguo@hust.edu.cn





Institute of Physics, CAS

基于忆阻器阵列的下一代储池计算

任宽 张握瑜 王菲 郭泽钰 尚大山

Next-generation reservoir computing based on memristor array Ren Kuan Zhang Wo-Yu Wang Fei Guo Ze-Yu Shang Da-Shan 引用信息 Citation: Acta Physica Sinica, 71, 140701 (2022) DOI: 10.7498/aps.71.20220082 在线阅读 View online: https://doi.org/10.7498/aps.71.20220082 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

蛋白质基忆阻器研究进展

Research progress of protein-based memristor 物理学报. 2020, 69(17): 178702 https://doi.org/10.7498/aps.69.20200617

一种适用于大规模忆阻网络的忆阻器单元解析建模策略

An analytic modeling strategy for memristor cell applicable to large-scale memristive networks 物理学报. 2021, 70(17): 178505 https://doi.org/10.7498/aps.70.20210116

具有感觉记忆的忆阻器模型 Mathematical model of memristor with sensory memory 物理学报. 2019, 68(1): 018501 https://doi.org/10.7498/aps.68.20181577

基于Au/TiO2/FTO结构忆阻器的开关特性与机理研究

Resistive switching characteristics and resistive switching mechanism of Au/TiO2/FTO memristor 物理学报. 2018, 67(15): 157302 https://doi.org/10.7498/aps.67.20180425

基于忆阻器的多涡卷混沌系统及其脉冲同步控制

Memristor-based multi-scroll chaotic system and its pulse synchronization control 物理学报. 2018, 67(11): 110502 https://doi.org/10.7498/aps.67.20180025

基于时延光子储备池计算的混沌激光短期预测

Short-time prediction of chaotic laser using time-delayed photonic reservoir computing 物理学报. 2021, 70(15): 154209 https://doi.org/10.7498/aps.70.20210355

专题: 面向类脑计算的物理电子学

基于忆阻器阵列的下一代储池计算*

任宽1)2)# 张握瑜1)3)# 王菲1)3) 郭泽钰1)3) 尚大山1)3)†

1) (中国科学院微电子研究所, 微电子器件与集成技术重点实验室, 北京 100029)

2) (西南交通大学超导与新能源研究开发中心, 磁浮技术与磁浮列车教育部重点实验室, 成都 610031)

3) (中国科学院大学,北京 100049)

(2022年1月12日收到; 2022年1月26日收到修改稿)

储池计算是类脑计算范式的一种,具有结构简单、训练参数少等特点,在时序信号处理、混沌动力学系 统预测等方面有着巨大的应用潜力.本文提出了一种基于存内计算范式的储池计算硬件实现方法,利用忆阻器 阵列完成非线性向量自回归过程中的矩阵向量乘法操作,有望进一步提升储池计算的能效.通过忆阻器阵列 仿真实验,在 Lorenz63 时间序列预测任务中验证了该方法的可行性,以及该方法在噪声条件下预测结果的鲁 棒性,并探究忆阻器阵列阻值精度对预测结果的影响.这一结果为储池计算的硬件实现提供了一种新的途径.

关键词:储池计算,忆阻器,存内计算,非线性向量自回归 PACS:07.05.Mh,84.35.+i,85.40.-e,87.15.A-

1 引 言

理解生物大脑中信息的加工、处理模式,并在 此基础上构建类脑计算硬件系统是现代信息科学 的前沿研究之一^[1].研究表明,生物大脑等效于一 个复杂神经网络动力学系统^[2],其处理外界信息的 机能依赖于神经网络的动力学过程^[3].如何理解大 脑的神经动力学过程、构建类脑动力学系统,是类 脑计算硬件系统实现的核心问题^[4].自然界中的信 息大部分是用时序数据来定义的.大脑的动力学系 统受外部时序信号刺激,并将刺激产生的数据进行 编码和存储^[5,6],进而形成各类认知过程.循环神经 网络 (recurrent neural network, RNN)^[7] 是一种 具有短时记忆能力的神经网络,其中的神经元通过 具有环路的网络结构,不仅可以接受其他神经元的 信息,也可以接受自身的信息,从而使网络具有处 理时序数据的能力,因此,更加适合模拟大脑的动力学系统.当前,RNN已经被广泛应用于语音识别、自然语言处理等任务中.然而,由于梯度消失和爆炸问题^[8],RNN需要的超参数多,而且训练过程复杂.因此,RNN在硬件系统实现上依然面临结构复杂、训练时间长和能耗高等问题^[9].

DOI: 10.7498/aps.71.20220082

储池计算 (reservoir computing, RC) 是 RNN 的一种简化形式. RC 概念最初的提出是为了模拟 生物大脑中具有大量循环连接的皮质纹状体系统 处理视觉空间序列信息的过程^[10]. 随后人们基于 RNN 的框架, 构建了统一的 RC 计算框架^[11-13] (如 图 1(a)). RC 的核心是一个被称为"储池"的循环神 经网络隐藏层. 该网络能够将时序输入信号转换到 高维空间中. 经过高维转换后, 输入信号的特征就 可以更容易地通过简单线性回归方法有效读出. 目 前, RC 在时序信号处理^[14]、混沌动力学系统预测^[15] 等动力学系统学习方面有良好的功能. 值得注意的

^{*} 国家重点基础研究发展计划 (批准号: 2018YFA0701500)、国家自然科学基金 (批准号: 61874138) 和中国科学院战略性先导科技 专项 (批准号: XDB44000000) 资助的课题.

[#] 同等贡献作者.

[†] 通信作者. E-mail: shangdashan@ime.ac.cn

^{© 2022} 中国物理学会 Chinese Physical Society

是, 与标准 RNN 相比^[16], RC 中只需要训练输出 层权重, 并且不需要反向传播算法, 有效避免了梯 度消失问题, 因此, 可以有效降低训练复杂度和训 练时间.



图 1 三种 RC 结构 (a) 传统 RC 结构; (b) 单节点延时 RC 结构; (c) 非线性向量自回归 RC 结构

Fig. 1. Three types of RC frameworks: (a) Conventional RC; (b) RC using a single nonlinear node reservoir with time-delayed feedback; (c) NGRC, which is equivalent to nonlinear vector autoregression.

传统 RC 主要包括由 Jaeger^[11] 提出的 ESN 模型(echo state network)和 Maass 等^[13] 提出的 LSM(liquid state machine)模型,他们的结构都 如图 1(a) 所示,包含输入、储池和输出三部分. ESN 模型的储池和 LSM 模型的储池都为基于 RNN 框架,由多个神经元随机连接而成的结构.所不同的 是,ESN 模型中的神经元是离散时间人工神经元,而 LSM 模型中的神经元是具有兴奋性和抑制性的 脉冲神经元.储池计算模型结构(以 ESN 为例)可 以用以下公式描述:

$$\boldsymbol{x}(n) = f(\boldsymbol{W}_{\text{in}}\boldsymbol{u}(n) + \boldsymbol{W}\boldsymbol{x}(n-1)), \qquad (1)$$

$$\boldsymbol{y}\left(n\right) = \boldsymbol{W}_{\text{out}}\boldsymbol{x}\left(n\right),\tag{2}$$

其中, u(n)为输入向量, n 为离散的时间, f 表示 储池层单元的非线性激活函数, Win 为输入连接储 池的权重矩阵, x(n)是所有离散时间人工神经元 的回声状态向量, W 为神经元间的连接矩阵, Wout 为储池连接输出的权重矩阵, y(n)为输出向量. 储 池计算只需要训练输出权重矩阵 Wout, 其性能取 决于储池神经元间的连接矩阵 W. 当W的谱半径 小于1时(即特征值的绝对值的最大项),对任意输入都可以得到对应的回声状态属性^[17].节点的回声状态属性等效于节点具备"衰退记忆"^[18].人们从理论上证明了,由离散时间"衰退记忆"节点构成的RC网络在输入有界的情况下具备动力学通用逼近能力^[19].

这类 RC 的硬件实现方法可大致分为两种: 一种方法是通过使用神经网络硬件或神经形态 计算技术实现,如用模拟电路^[20]、FPGA(fieldprogrammable gate array)^[21,22]、大规模集成电 路[23]、忆阻器[24-26]等直接构造储池中多个随机连 接的神经元.这种方法可灵活调整神经元间连接的 拓扑结构以改善性能,但是构造神经元需要的器件 众多,并且计算中每一个时间步都需要进行大量计 算以及存储大量神经元的状态. 另一种方法是采用 具备"衰退记忆"的物理节点代替随机连接的神经 元,构成储池的动力学系统,如纳米线网络[27,28]、 光学器件网络^[29]、易失性忆阻器网络^[30]等.这种 方法利用物理节点的"衰退记忆"特性进行计算.储 池的存与算在节点网络中同时进行. 然而由随机连 接的物理节点构成的动力学系统无法调整节点连 接的拓扑结构,故这种方法在面对不同任务时, RC 的性能具有一定的不稳定性.

为了提高 RC 的性能, 研究人员对 RC 结构进行了多种改进 (如多储池计算^[31]、进化储池计算^[32]等), 以及将 RC 与其他特征提取方法 (如卷积神经网络^[33]、强化学习^[34]、注意力学习^[35]等)相结合. 目前, 传统 RC 及其改进方法已经成功地被应用于众多领域, 如生物医学、声音识别、无线电等^[36]. 然而, 由于储池结构中神经元很多, 神经元状态存储以及更新需要大量的硬件资源, 并且由于神经元连接拓扑结构难以调整, 导致储池计算的参数优化困难.

RC 中神经元节点间相互作用产生的高维信号 可以通过延时动力学系统来实现. 其状态方程描述 为^[37]

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F\left(t, x\left(t\right), x\left(t-\tau\right)\right),\tag{3}$$

其中, t为连续时间信号, x(t)为系统的状态, F为 系统函数, τ 为延时时间. 2011年, Appeltant等^[38] 提出基于延时动力学系统的单延时节点 RC 结构 (见图 1(b)). 输入信号通过掩码函数 (masking) 进 行时间复用, 然后输入到单个物理节点在时间维度 上展开的虚拟节点中. 虚拟节点通过平等分割 τ 的 N个时间点上来设置.两个虚拟节点之间的时间间 隔为 $\theta = \tau/N$.所有虚拟节点 $x[t - (N - i)\theta], i =$ 1,…,N共同作为t时刻节点状态,并通过输出层 得到计算结果.

单节点延时 RC 的提出, 使得 RC 的硬件实现 变得更加便捷, 在一定程度上解决了传统储池硬件 实现中, 由于神经元数量多而导致的神经元状态存 储和更新硬件资源问题. 这种单节点延时储池已经 在光子器件^[39]、FPGAs^[40] 中得到硬件实现, 用于 语音识别、图像分类和混沌预测等任务中. 我们在 前期工作中, 利用铁电隧道结 (FTJ) 中超薄铁电 层的退极化效应产生的电流延时特性, 实现了单节 点延时储池计算功能^[41]. 为了拓展单节点延时 RC 功能, 我们采用了多个单延时节点储池并联的方 式, 提高了计算的维度, 实现了对动态数字序列的 识别功能^[41]. 然而, 由于虚拟节点是通过时间切分 获得, 所以其连接拓扑结构是按时间顺序固定的. 这意味着这种延时储池同样存在着参数优化困难 的问题.

最近, Gauthier 等^[42] 提出了一种新型 RC, 称 为下一代储池计算 (NGRC). NGRC 是一种特殊 的非线性向量自回归过程, 其等效于具有线性激活 节点的储池与一个非线性读出层的结合, 如图 1(c) 所示. NGRC 模型描述为

$$\boldsymbol{O}_{\mathrm{lin},i} = \left[\boldsymbol{X}\left(i\right), \boldsymbol{X}\left(i-s\right), \cdots, \boldsymbol{X}\left(i-\left(k-1\right)s\right)\right]^{\mathrm{T}},$$
(4)

$$\boldsymbol{O}_{\text{nonlin},i} = \boldsymbol{O}_{\text{lin},i} [\otimes] \boldsymbol{O}_{\text{lin},i}, \qquad (5)$$

$$\boldsymbol{O}_{\text{total},i} = c \bigotimes \boldsymbol{O}_{\text{lin},i} \bigotimes \boldsymbol{O}_{\text{nonlin},i}, \qquad (6)$$

$$\boldsymbol{Y}(i) = \boldsymbol{W}_{\text{out}} \boldsymbol{O}_{\text{total},i}, \qquad (7)$$

其中, i 为离散时间, $O_{\text{lin},i}$ 为线性特征向量, X(i)为第 i 时刻的输入向量, s 为时间间隔, k 为构成线 性特征向量的组数, $O_{\text{nonlin},i}$ 为第 i 时刻非线性特征 向量, $[\otimes]$ 功能为将符号两边项进行外积、并收集 外积结果的唯一单项式的运算符号, $O_{\text{total},i}$ 为第 i时刻总特征向量, c 为常数修正项, Y(i) 为输出 值, W_{out} 为储池连接输出的权重矩阵. NGRC 目前 被证实在完成短期动态预测、长期混沌预测、推断 动力学系统不可见数据等三个方面有很好的性能. 相对于传统 RC 和延时 RC, NGRC 使用更小的数 据集进行训练, 并避免了 RC 的参数优化困难问 题. 然而, 非线性向量自回归过程本身仍需要大量 硬件计算资源用于乘法计算操作.

忆阻器是近年广受关注的一种具有记忆功能的器件^[43].由忆阻器器件构成的交叉阵列^[44],可以通过欧姆定律和基尔霍夫定律,以存内计算的方式原位、并行、物理地完成矩阵向量乘运算,有效减少了计算过程中数据的搬运,从而具有功耗低、速度快的优点^[45-47].本文将 NGRC 过程通过矩阵向量乘法操作简化,提出了一种 NGRC 的存内计算硬件实现方法,并利用忆阻器阵列完成矩阵向量乘法操作.通过进行忆阻器阵列仿真完成了 Lorenz63时间序列预测任务,验证了该方法的可行性,并研究了忆阻器件电阻精度和波动性对 NGRC 预测精度的影响.这一结果为高能效 RC 提供了一种新的途径.

2 NGRC 的存内实现方法

传统 RC 过程中,每一个时间步都需要更新大 量具有"衰退记忆"特性的神经元的状态,然而具有 "衰退记忆"特性的线性神经元组成的储池与二次 非线性读出层组合,在数学上等效于一种特殊的非 线性向量自回归过程.NGRC 是对这种特殊的非 线性向量自回归过程的优化^[42].NGRC 过程与传 统储池计算过程的相同点在于都只需要训练输出 权重,但是在输入数据的选择和将输入数据进行高 维空间非线性转换的方式上有所不同.

输入数据方面,传统储池输入数据一般为当前 时刻的数据, 而 NGRC 的输入数据中, 除了当前时 刻的数据,还包括之前时刻所对应的数据.高维空 间非线性转换方式方面,传统储池的高维空间非线 性转换通过储池中具备"衰退记忆"神经元的非线 性激活函数达成. NGRC 结构储池的高维空间非 线性转换可分为3个过程(见图1(c)):1)选择不 同时刻输入数据构成线性特征向量Oin; 2) 由线性 特征向量构造非线性特征向量 O_{nonlin} ; 3) 由线性特 征向量与非线性特征向量构造总特征向量 O_{total}. 3个过程中,线性特征 Oim 向量是由选择的输入数 据直接拼接而成;总特征向量Ototal是由固定常数 c、线性特征向量 O_{in} 与非线性特征向量 O_{nonlin} 直接 拼接而成;而由线性特征向量Oim构造非线性特征 向量Ononlin 则需要经过一个非线性转换过程. NGRC 中的非线性转换过程将线性特征向量 Oim 通过外 积操作映射到一个高维空间中,并在高维空间中去

除对应映射向量的重复部分,得到非线性特征向量 Ononlin.

NGRC 的非线性转换过程虽然避免了传统储 池中随机连接的性能不确定性与需要同时更新多 个神经元状态的复杂性,但其向量间的外积操作与 除去高维向量重复部分的操作仍然需要大量硬件 开销与时间开销.我们注意到相同向量间的外积可 以用矩阵向量乘法 (matrix vector multiplication, MVM) 来表示,去除重复映射向量的操作可以通 过保留外积后固定位置元素的值 (保留元素操作) 来实现.硬件上,使用忆阻器阵列进行 MVM 操作, 使用忆阻器阵列的选择线电路进行保留元素操作.

2.1 线性特征向量的构建

图 2(a) 为三维空间时序数据预测任务的 NGRC 储池的存内实现结构. 其中, $t_i = i \times \Delta t$, Δt 为采 样间隔, i 为离散时间数, s 为时间间隔数, x(t), y(t), z(t)分别表示预测点在 t 时刻的 x 轴、y 轴、 z 轴三维空间坐标, k 为每个线性特征向量所取数 据的组数. 当已知 t_i 时刻及之前时刻点的轨迹坐 标,要预测 t_{i+1} 时刻点的坐标时,取 k = 2,即令 t_i 时刻和 t_{i-s} 时刻空间点的三维坐标构建第 i 个线性 特征向量 $O_{\text{lin},i}$,有

$$\boldsymbol{O}_{\text{lin},i} = [x_i, y_i, z_i, x_{i-s}, y_{i-s}, z_{i-s}]^{\mathrm{T}},$$
 (8)

其中, $[x_i, y_i, z_i]$ 与 $[x_{i-s}, y_{i-s}, z_{i-s}]$ 为 t_i 时刻与 t_{i-s} 时刻点的三维坐标.将构建的线性特征向量 $O_{\text{lin},i}$ 用时序电压脉冲幅值编码和电导编码,编码的电压序列矩阵 $V_{\text{lin},i}$ 为

$$\boldsymbol{V}_{\mathrm{lin},i} = \left[\boldsymbol{V}_{\mathrm{lin},i,1}, \boldsymbol{V}_{\mathrm{lin},i,2}, \boldsymbol{V}_{\mathrm{lin},i,3}, \boldsymbol{V}_{\mathrm{lin},i,4}, \boldsymbol{V}_{\mathrm{lin},i,5}, \boldsymbol{V}_{\mathrm{lin},i,6}\right]^{\mathrm{T}},$$
(9)

其中列向量 $V_{\text{lin},i,a}$ 的第 a行的值,为线性特征向量 $O_{\text{lin},i}$ 中第 a个元素对应的量化电压值,列向量中 的其他值为零.线性特征向量 $O_{\text{lin},i}$ 编码的电导矩 阵为

$$\boldsymbol{G}_{\mathrm{lin},i} = \begin{bmatrix} \boldsymbol{G}_{\mathrm{lin},i,1} & \boldsymbol{G}_{\mathrm{lin},i,2} & \boldsymbol{G}_{\mathrm{lin},i,3} & \boldsymbol{G}_{\mathrm{lin},i,4} & \boldsymbol{G}_{\mathrm{lin},i,5} & \boldsymbol{G}_{\mathrm{lin},i,6} \\ 0 & \boldsymbol{G}_{\mathrm{lin},i,2} & \boldsymbol{G}_{\mathrm{lin},i,3} & \boldsymbol{G}_{\mathrm{lin},i,4} & \boldsymbol{G}_{\mathrm{lin},i,5} & \boldsymbol{G}_{\mathrm{lin},i,6} \\ 0 & 0 & \boldsymbol{G}_{\mathrm{lin},i,3} & \boldsymbol{G}_{\mathrm{lin},i,4} & \boldsymbol{G}_{\mathrm{lin},i,5} & \boldsymbol{G}_{\mathrm{lin},i,6} \\ 0 & \cdots & 0 & \boldsymbol{G}_{\mathrm{lin},i,4} & \boldsymbol{G}_{\mathrm{lin},i,5} & \boldsymbol{G}_{\mathrm{lin},i,6} \\ \cdots & \ddots & \cdots & 0 & \boldsymbol{G}_{\mathrm{lin},i,5} & \boldsymbol{G}_{\mathrm{lin},i,6} \\ 0 & \cdots & 0 & 0 & 0 & \boldsymbol{G}_{\mathrm{lin},i,5} \end{bmatrix},$$
(10)

其中, *G*_{lin,*i*,*a*}为*O*_{lin,*i*}中的第*a*个元素对应的量化 电导值. 将电导序列使用差分编码存储到忆阻器阵 列中; 再将电压序列*V*_{lin,*i*}通过 Bitline 输入到忆阻 阵列中, 具体过程如图 2(b) 所示. 需要指出的是, 电 压脉冲幅值编码需要经过一个数模转换器 (DAC). DAC 的精度与忆阻器本身的精度影响着整体精度.

2.2 非线性特征向量的构建

构建非线性特征向量需要对线性特征向量进 行外积操作与保留元素操作. 忆阻器阵列中, 每通 过一个电压序列向量 *V*_{lin,*i*,*a*}, 能从阵列输出端 (SL) 得到一个电流向量 *I*_{lin,*i*,*a*}, 总电流矩阵由欧姆定律 和基尔霍夫定律可表达为

$$I_{\mathrm{lin},i} = V_{\mathrm{lin},i} \cdot G_{\mathrm{lin},i}$$

 $= \left[\boldsymbol{I}_{\text{lin},i,1}^{\text{T}}, \boldsymbol{I}_{\text{lin},i,2}^{\text{T}}, \boldsymbol{I}_{\text{lin},i,3}^{\text{T}}, \boldsymbol{I}_{\text{lin},i,4}^{\text{T}}, \boldsymbol{I}_{\text{lin},i,5}^{\text{T}}, \boldsymbol{I}_{\text{lin},i,6}^{\text{T}} \right]^{\text{T}}.$ (11) R $\mathbf{I}_{\text{lin},i}$ 矩阵中的非零元素, $\boldsymbol{I}_{\text{lin},i,a}$ 保留元素操作 后的向量为 *I*_{lin,*i*,*ap*},保留元素操作可通过只读取忆 阻器阵列电流输出中对应位置的输出实现,过程如 图 2(a) 中绿框部分所示,将选择读取的输出电流 组成非线性特征向量 *O*_{nonlin,*i*},可表达为

$$O_{\text{nonlin},i} = [I_{\text{lin},i,1p}, I_{\text{lin},i,2p}, I_{\text{lin},i,3p},$$
$$I_{\text{lin},i,4p}, I_{\text{lin},i,5p}, I_{\text{lin},i,6p}]$$
(12)

2.3 总特征向量的构建及输出

 t_i 时刻的总特征向量 $O_{\text{total},i}$ 是由固定常数c、 线性特征向量 $O_{\text{lin},i}$ 与非线性特征向量 $O_{\text{nonlin},i}$ 直接 拼接而成,表示为

$$\boldsymbol{O}_{\text{total},i} = [c, \boldsymbol{O}_{\text{lin},i}, \boldsymbol{O}_{\text{nonlin},i}].$$
 (13)

*t*_{*i*+1}时刻点的预测位置可直接由总特征向量乘以 输出权重得出:

$$[x_{i+1}, y_{i+1}, z_{i+1}] = \boldsymbol{O}_{\text{total},i} \cdot \boldsymbol{W}_{\text{out}}, \qquad (14)$$



图 2 基于忆阻阵列的 NGRC 储池结构 (a) 用于预测三维时序信号的 NGRC 储池结构. 输入为三维时序信号; 提取 t_i时刻 (红 色框) 和 t_{i s}(紫色框) 时刻信号的值组成线性特征向量 **O**_{lin}, 将第 i 个线性特征向量编码为时序电压和电导, 时序电压作为忆阻器 阵列的输入, 电导映射到忆阻器阵列上作为权重; 非线性特征向量 **O**_{nonlin} 由忆阻器阵列特定单元 (绿色方框) 的输出构成; 总特 征向量由 **O**_{lin} 与 **O**_{nonlin} 直接拼接而成. (b) 图 (a) 中的线性特征向量 **O**_{lin,i} 映射到忆阻器阵列的方式. **O**_{lin,i} 中的每一个值都由两 个忆阻器电导的差分 g⁺, g 表示

Fig. 2. Structure of the NGRC based on memristor-based crossbar. (a) Structure of the NGRC reservoir for three dimensional (3D) timing signals predicting. The input is a 3D timing signal. The linear feature vector O_{lin} is formed by extracting the signal values of t_i time (red box) and $t_{i,s}$ time (purple box). The *i*th linear feature vector is encoded as timing voltage and conductance, and the timing voltage is the input of the memristor array, and the conductance is mapped to the memristor array as weight. The nonlinear feature vector O_{nonlin} consists of the outputs of specific elements of the memristor array (green boxes). The total feature vector is directly spliced by O_{lin} and O_{nonlin} . (b) The way the linear feature vector $O_{\text{lin, }i}$ in panel (a) mapping to the memristor array. The g^+ and g^- represent the device conductance values for the positive and negative weights in the differential pair, respectively.

其中, [*x*_{*i*+1}, *y*_{*i*+1}, *z*_{*i*+1}]为所预测的*t*_{*i*+1}时刻点的三 维坐标, *W*_{out}为预先用岭回归方法训练好的输出 权重矩阵. 值得注意的是, 忆阻器电流值的读出需 要一个模数转换器 (ADC), ADC 的精度也会对最 终预测精度造成一定影响.

2.4 训练过程

储池训练过程只训练输出层 Wout,训练采用 岭回归方法,先用训练数据集得到由特征向量组成 的特征矩阵 Ototal,以及所有特征向量对应的输出 组成的结果矩阵 Y_d,岭回归方法表达为

$$\boldsymbol{W}_{\text{out}} = \boldsymbol{Y}_d \boldsymbol{O}_{\text{total}}^{\text{T}} \left(\boldsymbol{O}_{\text{total}} \boldsymbol{O}_{\text{total}}^{\text{T}} + a \boldsymbol{I} \right)^{-1}, \quad (15)$$

其中, **O**^T_{total} 是训练数据集的总特征向量的转置矩 阵, *a*为岭回归参数, **I** 为单位矩阵.

2.5 仿真平台

仿真平台基于 python 3.8, pytorch1.9.1(主机 GPU 型号 NVIDIA GeForce RTX 3080, CPU 型 号 i9-11980HK) 构建,其结构示意图如图 3 所示, 可分为输入部分、权重部分和输出部分.

输入部分模拟将外界信号转换为电压信号的 过程,最高转换精度为定点 32 bit.权重部分模拟 将外界信号映射到忆阻器阵列中(将带符号的权重 映射到一对忆阻器差分电导上)并进行运算的过 程,量化电导映射公式为

$$\Delta G = (G_{\text{max}} - G_{\text{min}})/2^{n-1}, \qquad (16)$$

$$G = [W/\Delta G] \times \Delta G, \tag{17}$$

其中 n 为权重精度 (单位 bit), G_{max}和G_{min}为忆阻器可变化的最大电导与最小电导, [x]为取整操作,

W为需要映射的权重, G为映射到忆阻器阵列对 应位置的电导值. 输出部分模拟忆阻器阵列输出经 过 ADC 转变为电脑可处理数据的过程, 输出精度 为所使用 ADC 的精度.



图 3 基于忆阻器阵列 (包括正、负列) 的矩阵乘法运算仿 真平台结构示意图, g_r 为忆阻器的电导, g_T 为晶体管电导 Fig. 3. Simulation platform of memristor array (including positive and negative arrays) as analog dot-product engine. The memristor conductance corresponds to g_r and the transistor conductance corresponds to g_T .

3 实验结果与讨论

动力学系统的短期预测能力与动力学系统的 长期预测能力通常被用来作为衡量 RC 性能的基 准,我们将用经典混沌动力学系统模型-Lorenz63 模型的短期预测任务与长期预测任务,验证基于忆 阻器阵列实现的 NGRC 结构的可行性及其对噪声 的鲁棒性. Lorenz63 是 1963 年洛伦兹^[48]提出来的 天气预测模型,由 3 个方程组成:

 $\frac{\mathrm{d}x}{\mathrm{d}t} = a(y-x), \ \frac{\mathrm{d}y}{\mathrm{d}t} = ax(b-z) - y, \ \frac{\mathrm{d}z}{\mathrm{d}t} = xy - cz, \ (18)$

其中状态 $X(t) \equiv [x, y, z]^{T}$ 是一个分量为 Rayleigh-Bénard 的对流可观测量的矢量, a = 10, b = 28, c = 8/3. Lorenz63 模型确定性的混沌行为体现在其对 初始条件的敏感依赖 (蝴蝶效应), 以及在相空间轨 迹形成奇异吸引子 (图 3).

动态系统的预测任务中,用原序列(由动态系统方程得到的序列)与预测序列(储池不断将此时刻输出值作为下一时刻的输入值进行预测得到的序列)之间的结构相似度来衡量预测效果.就 Lore-nz63 时间序列预测任务而言,归一化均方根误差(NRMSE)可在一定程度上衡量短预测期内的结构相似度,但难以反映长期预测的结构相似度.Lor-

enz63 时间序列预测的 z 回归图能直观地反映 z 变量的长期行为,比较原序列与预测序列的 z 回归图可以定性地比较两个序列长时段的结构相似度.在之后的 Lorenz63 时间序列预测任务中,NRMSE 衡量短期预测 (1 个李雅普诺夫周期)的结构相似度,通过比较量原序列与预测序列的 z 回归图衡量长期预测的结构相似度.

在基于忆阻器阵列实现的 NGRC 结构的可行 性验证实验中,维持系统的输入精度不变,通过改 变系统的权重精度 (忆阻阵列中忆阻器的量化映射 比特数)和输出精度 (忆阻器阵列输出 ADC 比特 数),研究不同权重精度和输出精度对预测结果的 结构相似度的影响.在基于忆阻器阵列实现的 NGRC 结构对噪声的鲁棒性验证实验中,维持输入精度和 输出精度不变,研究不同权重精度以及不同噪声大 小对预测结果的结构相似度的影响.

3.1 可行性验证实验

保持输入精度为定点 32 bit, 输出精度为定点 64 bit, 在权重精度为4, 6, 8, 16, 32 和 64 bit 情况 下进行预测实验. 800个时间步的预测结果及其 xz 截面图如图 4 所示. 可以看出, 权重精度在 4 和 6 bit 时无法产生混沌现象 (无洛伦兹吸引子); 当 权重精度达到 8 bit 及以上时, 开始产生明显的洛 伦兹混沌吸引子. 这一结果意味着权重精度对混沌 的产生有重要影响. 当忆阻器阵列对应的权重精度 达到一定值时, 基于忆阻器阵列实现的 NGRC 结 构构成的系统能由稳定状态过渡到混沌状态.

保持输入精度为定点 32 bit, 通过改变权重精 度以及输出精度,在达到混沌状态的前提下探究短 期预测结构相似度与权重精度的关系.图 5 为短期 预测 (1 个李雅普诺夫周期) 的 NRMSE 随不同权 重精度(8,16,32,64 bit)和不同输出精度(8,16, 32,64 bit)的变化.结果显示,当基于忆阻器阵列 实现的 NGRC 结构构成的系统达到产生混沌所需 的权重精度和输出精度时,短期预测的性能随着权 重精度、输出精度的增加而增加. 在权重精度不变 的情况下,当输出精度达到16 bit,输出精度的增 加对短期预测结构相似度几乎无影响;在输出精度 不变的情况下,短期预测结构相似度随着权重精度 的增加而变高 (NRMSE 变小), 8 bit 权重精度下 的 NRMSE 低于 0.05, 16 bit 权重精度下的 NRMSE 接近于 0, 当权重精度超过 16 bit 时, 权重精度的 增加对短期预测结构相似度几乎无影响.



图 4 输入精度为定点 32 bit, 输出精度为定点 64 bit, 不同权重精度下 800 个时间步的预测 XZ 截面图 (a) 64 bit; (b) 32 bit; (c) 16 bit; (d) 8 bit; (e) 6 bit; (f) 4 bit

Fig. 4. The XZ cross sections of 800 time steps with different weight precision, when input precision of integer is 32 bit and output precision of integer is 64 bit; (a) 64 bit; (b) 32 bit; (c) 16 bit; (d) 8 bit; (e) 6 bit; (f) 4 bit.

Lorenz63 系统的 *z* 分量在连续的局部极大值 之间具有函数关系,通过找到 *z* 分量的连续局部极 大值 *M_i*并根据 *M_{i+1}* 画出 *M_i*形成 *z* 回归图,可以 简洁地展现 *z* 变量的长期行为. Lorenz63 系统的 z回归图如图 6 所示,紫色点为真实序列 z回归图, 其他颜色为输入精度为定点 32 bit,输出精度为定 点 64 bit,权重精度分别为 8, 16, 32, 64 bit 的 z回 归图.结果显示,权重精度为 8 bit 时的 z回归图相 比真实序列的回归图有明显偏移;当权重精度在 16 bit 及以上时,预测的回归图几乎完全覆盖了真 实序列的回归图;随着权重精度的增加,预测的回 归图往真实序列的回归图收敛.



图 5 短期预测 (1 个李雅普诺夫周期) 的 NRMSE 随不同 权重精度 (8, 16, 32, 64 bit) 和不同输出精度 (8, 16, 32, 64 bit) 的变化

Fig. 5. The variation diagram of NRMSE for short-term prediction (1 Lyapunov cycle) with different weight precision (8, 16, 32, 64 bit) and different output precision (8, 16, 32, 64 bit).





Fig. 6. (a) The z return map of Lorenz63 (purple) overlaid with the z return map under different weight accuracy; (b) detail of the region marked in Fig. (a).

3.2 噪声鲁棒性验证实验

在保持输入精度为定点 32 bit, 输出精度为定 点 64 bit, 给权重 (即忆阻器电导 G) 添加高斯噪 声 Noise $N(0,\sigma)$, 其中 $\sigma = G \times 10^{-4} \times$ percent 为方 差, percent 表示噪声强度百分比. 短期预测结构相 似度的 NRMSE 随权重噪声强度变化的结果如图 7 所示. 当权重精度在 8 bit 时, 随着 σ 的增大, 短期预 测结构相似度的 NRMSE 会先降低后升高: 当权重 精度在 16 bit 时, 短期预测结构相似度的 NRMSE 也会先降低后升高,但降低点对应的噪声强度比权 重精度在 8 bit 时小; 当权重精度在 16 bit 以上时, 短期预测的 NRMSE 会随着 percent 的增大而增 大. 由于量化本身具备一定的抗噪声能力, 故权重 精度越低,噪声对短期预测结构相似度的 NRMSE 的影响越小;值得注意的是,一定程度的噪声有利 于提升短期预测性能,并且量化的比特数越高,能 带来增益的噪声强度越小.



图 7 短期预测结构相似度的 NRMSE 在不同权重精度条件下随权重噪声强度的变化

Fig. 7. The variation of NRMSE under different weight precision conditions for short-term prediction with increasing weight noise intensity.

3.3 讨论

基于忆阻器阵列的 NGRC 存内实现具备两方 面的优势:第一,就 NGRC 算法本身而言,相对于 传统储池计算和基于延时的储池计算, NGRC 的 储池具备更短的激活时间、更少的参数训练量以及 更快的训练和推理速度^[42];第二,就存内计算方面 而言, NGRC 中提取非线性特征向量的过程需要 大量的乘法操作,而忆阻器阵列相比传统 CMOS 电路,在矩阵向量乘法方面具备更快的计算速度和 更低的功耗^[49].然而,使用忆阻器阵列进行 NGRC 的过程中,每一次推理过程都需要在忆阻器阵列中 写入采样数据;同时,仿真结果表明,忆阻器阵列 中每个忆阻器精度达到 8 bit, Lorenz63 才能有较 好的预测结果.考虑到当前忆阻器还存在各种非理 想性因素,因此,如何进一步提高写入效率,同时 降低所需忆阻器的阻值精度还需进一步探索.

4 结 论

储池计算自提出至今可以分为传统储池计算、 延时储池计算和下一代储池计算三个阶段. 储池计 算性能上的优势不仅来自于算法自身,而且与硬件 的实现方式密切相关. 本文在总结储池计算发展历 程的基础上,提出一种基于存内计算范式的硬件实 现方法,将 NGRC 过程通过矩阵向量乘法操作简 化,并利用忆阻器阵列完成矩阵向量乘法操作. 忆 阻器阵列仿真实验验证了这一方法在 Lorenz63 三 维时间序列预测任务中的可行性. 仿真实验结果表 明, 预测效果与输出精度和权重精度密切相关. 当 输出精度达到 16 bit, 进一步提高输出精度对预测 效果的影响可忽略不计,并且具有良好的抗噪声能 力;当权重精度达到 8 bit,对 Lorenz63 三维时间 序列预测的短期预测 (1个李雅普诺夫时间) 就可 以有良好的预测效果 (NRMSE 小于 0.05), 并可以 在一定程度进行长期预测.这些结果为 NGRC 的 硬件实现提供了一种新的途径,同时也展现了忆阻 器阵列在开发基于储池计算的实时、低能耗边缘计 算系统方面的潜力.

参考文献

- [1] Guillem C, Jordi F 2015 Front. Psychol. 6 818
- [2] Dayan P, Abbott L F 2001 J. Cogn. Neurosci. 15 154
- [3] Vogels T P, Rajan K, Abbott L F 2005 Annu. Rev. Neurosci. 28 357
- [4] Tian Y, Li G, Sun P 2021 Phys. Rev. Res. 3 043085
- [5] Borst A, Theunissen F E 1999 Nat. Neurosci. 2 947
- [6] Amit D J, Gutfreund H, Sompolinsky H 1987 Phys. Rev. A 35 2293
- [7] Danilo P, Mandic J A C 2001 Recurrent Neural Networks Architecture (Hoboken: John Wiley & Sons Ltd) pp69–89
- [8] Choi E, Schuetz A, Stewart W F, Sun J 2016 J. Am. Med. Inform. Assoc. 24 361
- [9] Pascanu R, Mikolov T, Bengio Y 2013 Proceedings of the 30 th International Conference on Machine Learning Atlanta, Georgia, USA, June 16–21, 2013 p1310
- [10] Dominey P, Arbib M, Joseph J P 1995 J. Cogn. Neurosci. 7 311
- [11] Jaeger H 2001 German National Research Institute for Computer ScienceGerman National Research Centre for

Information Technology, GMD Technical Reports Bonn, Germany, January 01, 2001 p13

- [12] Jaeger H, Haas H 2004 Science **304** 78
- [13] Maass W, Natschlager T, Markram H 2002 Neural Comput. 14 2531
- [14] Kan S, Nakajima K, Takeshima Y, Asai T, Kuwahara Y, Akai-Kasaya M 2021 Phys. Rev. Appl. 15 024030
- [15] Pathak J, Hunt B, Girvan M, Lu Z, Ott E 2018 Phys. Rev. Lett. 120 024102
- [16] Chattopadhyay A, Hassanzadeh P, Subramanian D 2020 Nonlinear Processes Geophys. 27 373
- [17] Lukoševičius M, Jaeger H, Schrauwen B 2012 KI Künstliche Intelligenz 26 365
- [18] Boyd S, Chua L 1985 IEEE Trans. Circuits Syst. 32 1150
- [19] Grigoryeva L, Ortega J P 2018 Neural Networks 108 495
- [20] Zhao C, Li J, Liu L, Koutha L S, Liu J, Yi Y 2016 Proceedings of the 3 rd ACM International Conference on Nanoscale Computing and Communication New York, New York, USA, September 28–30, 2016 p1
- [21] Canaday D, Griffith A, Gauthier D 2018 Chaos: An Interdisciplinary Journal of Nonlinear Science 28 123119
- [22] Yi Y, Liao Y, Fu X 2016 Microprocess. Microsyst. 46 175
- [23] Bertschinger N, Natschlager T 2004 Neural Comput. **16** 1413
- [24] Yang X, Chen W, Wang F 2016 Analog Integr. Circuits Signal Process. 87 263
- [25] Merkel C, Saleh Q, Donahue C, Kudithipudi D 2014 5 th Annual International Conference on Biologically Inspired Cognitive Architectures (BICA) MIT Campus, Cambridge, Massachusetts, USA, November 7–9, 2014 p249
- [26] Donahue C, Merkel C, Saleh Q, Dolgovs L, Ooi Y, Kudithipudi D, Wysocki B 2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA) Verona, New York, May 26–28, 2015 p24
- [27] Demis E, Aguilera R, Scharnhorst K, Aono M, Stieg A, Gimzewski J 2016 Jpn. J. Appl. Phys. 55 1102B2
- [28] Lilak S, Woods W, Scharnhorst K, Dunham C, Teuscher C, Stieg A, Gimzewski J 2021 Front. in Nanotechnol. 3 1
- [29] Vandoorne K, Mechet P, van Vaerenbergh T, et al. 2014 Nat. Commun. 5 3541
- [30] Milano G, Pedretti G, Montano K, Ricci S, Hashemkhani S, Boarino L, Ielmini D, Ricciardi C 2021 Nat. Mater. doi: 10.1038/s41563-021-01099-9
- [31] Gallicchio C, Micheli A, Pedrelli L 2017 Neurocomputing 268 87
- [32] Qiao J, Li F, Han H G, Li W 2016 IEEE Trans. Neural Networks Learn. Syst. 28 391
- [33] Tong Z Q, Tanaka G 2018 24th International Conference on Pattern Recognition (ICPR) Beijing, China, August 20–24, 2018 p1289
- [34] Murakamli M, Kroger B, Birkholz P, Triesch J 2015 5th IEEE Joint International Conference on Development and Learning and on Epigenetic Robotics (IEEE ICDL-EpiRob) Providence, Rhode Island, August 13–16, 2015 p208
- [35] Sussillo D, Abbott L F 2009 Neuron $\mathbf{63}$ 544
- [36] Tanaka G, Yamane T, Heroux J B, et al. 2019 Neural Networks 115 100
- [37] Lepri S, Giacomelli G, Politi A, Arecchi F T 1994 Physica D 70 235
- [38] Appeltant L, Soriano M C, van der Sande G, et al. 2011 Nat. Commun. 2 468 468
- [39] Brunner D, Penkovsky B, Marquez B A, Jacquot M, Fischer I, Larger L 2018 J. Appl. Phys. 124 152004
- [40] Penkovsky B, Larger L, Brunner D 2018 J. Appl. Phys. 124 162101
- [41] Yu J, Li Y, Sun W, et al. 2021 Symposium on VLSI

Technology Kyoto, Japan, June 13–19, 2021 pl

- [42] Gauthier D J, Bollt E, Griffith A, Barbosa W A S 2021 Nat. Commun. 12 5564
- [43] Strukov D B, Snider G S, Stewart D R, Williams R S 2008 Nature 453 80
- [44] Li H, Wang S, Zhang X, Wang W, Yang R, Sun Z, Feng W, Lin P, Wang Z, Sun L, Yao Y 2021 Adv. Intell. Syst. 3 2100017
- [45] Li Y, Loh L, Li S, Chen L, Li B, Bosman M, Ang K W 2021

Nat. Electron. 4 348

- [46] Kim H, Mahmoodi M R, Nili H, Strukov D B 2021 Nat. Commun. 12 5198
- [47] Xiao T P, Bennett C H, Feinberg B, Agarwal S, Marinella M J 2020 Appl. Phys. Rev. 7 031301
- [48] Lorenz E N 2004 The Theory of Chaotic Attractors (New York: Springer New York) pp25–36
- [49] Zhang W, Gao B, Tang J, Yao P, Yu S, Chang M F, Yoo H J, Qian H, Wu H 2020 Nat. Electron. 3 371

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Next-generation reservoir computing based on memristor array^{*}

Ren Kuan^{1)2)#} Zhang Wo-Yu^{1)3)#} Wang Fei¹⁾³⁾

Guo Ze-Yu¹⁾³⁾ Shang Da-Shan^{1)3)†}

1) (Key Laboratory of Microelectronics Devices and Integrated Technology, Institute of

Microelectronics, Chinese Academy of Sciences, Beijing 100029, China)

2) (Key Laboratory of Magnetic Levitation Technologies and Maglev Trains, Ministry of Education, Superconductivity

and New Energy R&D Center, Southwest Jiaotong University, Chengdu 610031, China)

3) (University of Chinese Academy of Sciences, Beijing 100049, China)

(Received 12 January 2022; revised manuscript received 26 January 2022)

Abstract

As a kind of brain-inspired computing, reservoir computing (RC) has great potential applications in time sequence signal processing and chaotic dynamics system prediction due to its simple structure and few training parameters. Since in the RC randomly initialized network weights are used, it requires abundant data and calculation time for warm-up and parameter optimization. Recent research results show that an RC with linear activation nodes, combined with a feature vector, is mathematically equivalent to a nonlinear vector autoregression (NVAR) machine, which is named next-generation reservoir computing (NGRC). Although the NGRC can effectively alleviate the problems which traditional RC has, it still needs vast computing resources for multiplication operations. In the present work, a hardware implementation method of using computing-in memory paradigm for NGRC is proposed for the first time. We use memristor array to perform the matrix vector multiplication involved in the nonlinear vector autoregressive process for the improvement of the energy efficiency. The Lorenz63 time series prediction task is performed by simulation experiments with the memristor array, demonstrating the feasibility and robustness of this method, and the influence of the weight precision of the memristor devices on the prediction results is discussed. These results provide a promising way of implementing the hardware NGRC.

Keywords: reservoir computing, memristor, in-memory computing, nonlinear vector autoregression

PACS: 07.05.Mh, 84.35.+i, 85.40.-e, 87.15.A-

DOI: 10.7498/aps.71.20220082

^{*} Project supported by the National Basic Research Program of China (Grant No. 2018YFA0701500), the National Natural Science Foundation of China (Grant No. 61874138), and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB44000000).

 $^{^{\#}\,}$ These authors contributed equally.

[†] Corresponding author. E-mail: shangdashan@ime.ac.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

基于忆阻器的脉冲神经网络硬件加速器架构设计

武长春 周莆钧 王俊杰 李国 胡绍刚 于奇 刘洋

Memristor based spiking neural network accelerator architecture Wu Chang-Chun Zhou Pu-Jun Wang Jun-Jie Li Guo Hu Shao-Gang Yu Qi Liu Yang 引用信息 Citation: Acta Physica Sinica, 71, 148401 (2022) DOI: 10.7498/aps.71.20220098 在线阅读 View online: https://doi.org/10.7498/aps.71.20220098 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

NbO_忆阻神经元的设计及其在尖峰神经网络中的应用

 $\mathrm{Design}\ \mathrm{of}\ \mathrm{NbO}_{\mathrm{v}}$ memristive neuron and its application in spiking neural networks

物理学报. 2022, 71(11): 110501 https://doi.org/10.7498/aps.71.20220141

基于随机放电神经元网络的彩色图像感知研究 Color image perception based on stochastic spiking neural network 物理学报. 2022, 71(7): 070501 https://doi.org/10.7498/aps.71.20211982

铝基薄膜忆阻器作为感觉神经系统的习惯化特性

Al-based memristor applied to habituation sensory nervous system 物理学报. 2021, 70(6): 068502 https://doi.org/10.7498/aps.70.20201961

基于忆容器件的神经形态计算研究进展

Research progress of neuromorphic computation based on memcapacitors 物理学报. 2021, 70(7): 078701 https://doi.org/10.7498/aps.70.20201632

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

^{专题:面向类脑计算的物理电子学} 基于忆阻器的脉冲神经网络硬件加速器架构设计*

武长春 周莆钧 王俊杰 李国 胡绍刚 于奇 刘洋†

(电子科技大学电子科学与工程学院,成都 610054)

(2022年1月13日收到; 2022年5月29日收到修改稿)

脉冲神经网络 (spiking neural network, SNN) 作为第三代神经网络,其计算效率更高、资源开销更少,且 仿生能力更强,展示出了对于语音、图像处理的优秀潜能. 传统的脉冲神经网络硬件加速器通常使用加法器 模拟神经元对突触权重的累加. 这种设计对于硬件资源消耗较大、神经元/突触集成度不高、加速效果一般. 因此,本工作开展了对拥有更高集成度、更高计算效率的脉冲神经网络推理加速器的研究. 阻变式存储器 (resistive random access memory, RRAM)又称忆阻器 (memristor),作为一种新兴的存储技术,其阻值随电压变化 而变化,可用于构建 crossbar 架构模拟矩阵运算,已经在被广泛应用于存算一体 (processing in memory, PIM)、 神经网络计算等领域. 因此,本次工作基于忆阻器阵列,设计了权值存储矩阵,并结合外围电路模拟了 LIF (leaky integrate and fire)神经元计算过程. 之后,基于 LIF 神经元模型实现了脉冲神经网络硬件推理加速器 设计. 该加速器消耗了 0.75k 忆阻器,集成了 24k 神经元和 192M 突触. 仿真结果显示,在 50 MHz 的工作频率 下,该加速器通过部署三层的全连接脉冲神经网络对 MNIST (mixed national institute of standards and technology)数据集进行推理加速,其最高计算速度可达 148.2 frames/s, 推理准确率为 96.4%.

关键词:脉冲神经网络,阻变式存储器,存内计算,LIF 神经元,硬件推理加速器
 PACS: 84.35.+i, 85.40.-e, 95.75.Mn
 DOI: 10.7498/aps.71.20220098

1 引 言

近年来,随着深度学习的发展,神经网络已经 得到了广泛的应用.通过反向传播、梯度下降等方法 训练出神经网络可以协助人类完成复杂的工作^[1], 有时甚至可以做出优于人类的决策^[2].传统的人工 神经网络 (ANN)使用了 MP(McCulloch and Pitts) 神经元模型^[3],通过对前一层神经元信号的加权累 加和非线性激活输出来模拟神经元的行为,但是其 生物置信度差、计算效率低、硬件资源开销大.为 了构建一种具有更高生物置信度的神经元模型, 1952年,Hodgkin和Huxley^[4]基于神经元细胞膜 电位生理现象的非线性微分方程,提出Hodgkin-Huxley(HH)模型.HH模型可以更加精确地模拟 生物神经元膜电压变化和脉冲发放,但是由于其存 在大量的微分、积分操作,硬件实现难度大.相比 于 HH 模型, LIF(leaky integrate and fire) 神经元 模型^[5] 在保留一定程度仿生能力的同时,极大地降 低了计算复杂度,使之可以轻易部署到硬件平台 上,具有更强的硬件友好性.基于 LIF 神经元模型 的 SNN^[6,7] 作为第三代神经网络,相比于传统的单 层感知器和多层感知器具有更强的生物合理性及 处理时空信息时潜在的更优计算效率,同时具有相 对较低的硬件设计资源开销.

近年来,基于 SNN 的神经形态计算^[8,9] 受到 了广泛关注并得到极大发展.神经形态计算的一个 重要目标是通过硬件模拟生物神经网络计算.基于 冯·诺依曼架构的硬件计算平台 (CPU, GPU等) 是实现神经形态计算的重要途径之一.但是冯·诺 依曼架构存在存算分离导致的冯·诺依曼瓶颈^[10,11], 这使得其在实现神经形态计算时效率低下,且拥有

^{*} 国家自然科学基金 (批准号: 92064004) 资助的课题.

[†] 通信作者. E-mail: yliu1975@uestc.edu.cn

^{© 2022} 中国物理学会 Chinese Physical Society

极为庞大的功耗、面积开销. ASIC 设计是实现神 经形态计算的另一个重要途径. 相比于通用计算平 台, ASIC 设计具有高的专用性, 其计算效率更高^[12]、 功耗更低、硬件资源开销更少. 相比于基于模拟 ASIC 设计的类脑计算芯片^[13], 基于数字 ASIC 设 计的类脑计算芯片^[14-17]可以实现更高的集成度, 因此受到了更加广泛的关注. 为了解决冯·诺伊曼 瓶颈对计算效率的影响, 数字 ASIC 设计中往往使 用寄存器作为存算单元, 实现存算一体的神经形态 芯片设计. 但是随着设计规模的不断增加, 寄存器 阵列及其外围电路的复杂度不断上升、资源开销迅 速增长.

近年来,基于忆阻器的存算一体技术[18]由于 其资源开销低、计算效率高等特点受到了广泛的关 注. 相比于数字集成电路中基于寄存器的存算一体 架构,基于忆阻器的 crossbar 架构复杂度更低,且 集成度更高.同时,基于忆阻器 crossbar 架构的类 脑计算[19-22] 可以并行输入多组脉冲, 通过对输出 电流的统计、求和,从而实现对多路突触的并行计 算. 但是, 其输出往往是电流形式, 需要外加运算 放大器和 ADC 才能实现模拟电流到数字电压的 转换,其外围电路复杂度高且不易集成. 2020年, 国防科技大学忆阻器研究团队^[23]提出了使用 N位 等值逻辑比较电路实现忆阻器阵列的数字电压输 出,其在一定程度上降低了外围电路的复杂程度, 但是需要更多的时间步长进行计算.此外,随着突 触集成度的增加, 忆阻器阵列规模需要不断扩大, 资源开销随之增加. 通过突触复用的方式可以有效 地解决资源开销随设计规模增扩大而增加的问题, 但是这需要对忆阻器反复擦写来更新突触权值.尽 管这种设计方法降低了设计资源开销,但是其牺牲 了较多的计算效率.

在本次工作中,致力于解决硬件资源开销与计 算效率之间的矛盾,采用硬件设计语言(Verilog) 构建了基于忆阻器的 crossbar 阵列作为类脑计算 的权值存储阵列.基于 crossbar 阵列设计,提出了 基于权值共享技术的紧凑化权值存储阵列设计,并 通过神经元、突触复用技术减少忆阻器阵列的擦写 频率,仅使用 0.25k 忆阻器单元模拟 8k 神经元和 64M 突触计算.此外,通过在忆阻器阵列的输出端 进行串联电阻分压实现了忆阻器阵列输出从电流 到电压的转换. 忆阻器阵列的输出结果可以直接通 过数字电路采集.基于权值存储阵列的设计,构建 了紧凑化计算核心. 随后, 基于紧凑化计算核心设 计了推理加速器, 并使用通用验证方法学 (UVM) 搭建验证平台. 通过 GPU对 784×1024×1024×10 规模的全连接网络进行了训练, 并将其部署到加速 器平台对 MNIST 数据集进行推理验证实验. 实验 结果表明, 该加速器可以实现 148.2 frams/s 的图 像识别速率, 同时识别精度达到了 96.4%. 最后, 基 于硬件加速设计部署了不同规模的网络进行推理 测试, 总结了该设计对于不同规模网络的加速效果 并对结果进行了分析.

2 脉冲神经网络相关架构介绍

2.1 脉冲神经网络 LIF 模型结构介绍

LIF 神经元模型是实现脉冲神经网络的一个 重要神经元模型.其相比于传统的 MP 神经元模 型, LIF 神经元模型可以通过更少硬件资源来实现, 并且可以更好地模拟生物神经元工作原理.(1)式 所示为 LIF 神经元膜电位与输入脉冲的关系:

 $V_{mp(t)} =$

$$\begin{cases} V_{\rm mp(t-1)} + \sum_{i=0}^{N-1} O_i \times W_i + V_{\rm L}, & (|V_{\rm mp}| < V_{\rm th}), \\ V_{\rm reset}, & \text{otherwise}, \end{cases}$$
(1)

式中, Vmp是神经元膜电位; O_i是输入脉冲; W_i是 与输入脉冲对应的权重; V.是神经元的泄漏电压; V_{th}是膜电位的阈值; Vreset</sub>是神经元膜电位的复位 值; N是神经元个数.同时将膜电位恢复至,其膜 电位升高.在一个"time step"的计算中,神经元完 成所有突触的计算,其膜电位将与阈值电压作对 比.当神经元的膜电位达到阈值电压(V_{th})时,其 将发放一个脉冲,同时将膜电位恢复至静息值(Vreset); 否则,神经元将不会发放脉冲,并将当前膜电位与 泄漏电压 V.相加并保存,用于下一个"time step" 计算.

基于 LIF 神经元模型的四层全连 SNN 结构如 图 1 所示. 该网络由一个输入层、两个隐藏层和一 个输出层组成,输入层与各隐藏层分别由 5 个神经 元组成,输出层由两个神经元组成. 相比于传统的 ANN,该 SNN 网络中神经元的输入和输出都是脉 冲形式的,即 0/1 脉冲电平. 输出端统计两个输 出神经元的脉冲发放数量, 经过 n 个 time step" 的脉冲统计后,根据脉冲统计结果进行分类. 脉冲 发放最多的神经元表示其激活程度最高,即意味着 输入的脉冲序列最有可能对应该神经元所代表的 类别.

2.2 LIF 神经元的硬件设计方法

一个典型的 LIF 神经元模型的硬件结构如 图 2(a) 所示.其主要实现了脉冲输入、膜电位累 加、膜电位比较及脉冲发放.输入脉冲与对应权值 相乘并类形成膜电位,在膜电位计算完毕后,神经 元通过将之与阈值电压相比较判断是否发放脉冲 输出.其在硬件中的计算原理图如图 2(b) 所示.输入的脉冲序列由 1/0 脉冲串组成.输入脉冲通过多路选择器实现与权值的"乘"操作,选择结果作为突触的输入.神经元将初始膜电位 (即上一个"time step"结束后保存的膜电位) 与所有突触的输入相累加,形成总的膜电位,并与阈值电压相比较,如果膜电位超出阈值电压,则将膜电位清零并保存用作下一个"time step"的初始膜电位,同时发放一个脉冲输出;反之,膜电位将加上一个泄露电压并保存,同时不产生脉冲输出.





Fig. 1. Structure diagram of fully connected spiking neural network based on LIF model.





Fig. 2. (a) Neuron schematic diagram; (b) schematic diagram of neuron computation.





Fig. 3. The LIF neuron model architecture based on synapse multiplexing technology and weight sharing technology.

2.3 基于突触复用技术及权值共享技术的 LIF 神经元模型硬件设计方法

图 3 所示为基于突触复用技术及权值共享技术的 LIF 神经元模型架构. 与传统 LIF 神经元架构不同,其只有一个脉冲输入通道,通过分时复用的技术实现所有突触的计算. 所有突触共享 M 个权值,因此,每个突触的脉冲输入同时伴随一组权值索引输入,通过对应的权值索引来获取其对应的权值数据. 当神经元依次完成对所有突触的膜电位累加后,神经元会对膜电位进行判断并控制脉冲发放.

3 基于忆阻器阵列的脉冲神经网络 加速器设计

3.1 基于忆阻器阵列的权重矩阵设计方法

忆阻器由上下两块电极及其中间起隔离作用 的氧化层组成,如图 4(a) 所示.当其两端加正向电 压,氧化层中会出现氧空位,从而组成导电细丝, 使得忆阻器导电能力增强,呈现低阻特性,其阻 值为 *R*_L.当其两端加反向电压,导电细丝会断裂, 忆阻器呈现高阻特性,其阻值为 *R*_H.本次工作主要 利用了忆阻器的这两种工作状态,将其作为二值器 件使用,其中忆阻器的高阻与低阻的阻值相差千倍 以上.

基于忆阻器单元 A,设计了 crossbar 结构,如 图 4(b)所示. crossbar 的行数代表了权值的数量, 每一行代表了一个权值,每一个忆阻器代表权值 的一位,其高阻态代表逻辑 0,低阻态代表逻辑 1. 列数则代表了权值的位宽.当输入端有高电平脉冲 输入时,其对应的权值被激活,并以电流形式从输 出端输出.每一列的输出端串联一个分压电阻 R_D , 从而将输出电流转化成电压数据,其中串联电阻的 阻值 $R_D = \sqrt{R_L R_H}$.因为 R_D 远大于忆阻器的低阻 阻值且远小于忆阻器的高阻阻值,因此,当 R_D 对应 列被激活的忆阻器阻值为 R_L 时,即可认为输出端 是逻辑电平"1";当 R_D 对应列被激活的忆阻器阻值 为 R_H 时,即可认为输出端是逻辑电平"0".用这样 的方式即可通过数字电路直接对 crossbar 阵列的 输出进行结果采集,而不需要使用运算放大器、数 模转换器等模拟电路,降低了其外围电路的复杂 度,同时提升了计算效率.

3.2 权值共享技术

在神经网络结构中,一个神经元需要与多个突 触相连接,每个突触都对应有一个权值用来表示连 接关系的强弱.随着网络规模的增加,突触的数量 急剧增加,忆阻器阵列的资源开销随之增加.为了 减少权值数量从而降低硬件资源开销,本工作提出 了一种基于忆阻器阵列的权值共享技术.一层网络 中所有突触的权值被聚类成 16 个 16-bit 权值 W_i , 有 { $W_i \in W_0, W_1 \cdots W_{15}$ }.突触通过一个 4-bit 的 索引 S_j 对权值进行索引,有 { $S_i \in S_0, S_1 \cdots S_{15}$ }.



图 4 (a) 忆阻器的建立/擦除示意图; (b) 基于忆阻器的 crossbar 阵列 Fig. 4. (a) The set/reset operation of the memristor; (b) the crossbar structure based on the memristor.

本次工作基于 784 × 1024 × 1024 × 10 的全连 接网络开展了对权值共享技术的应用与分析. 实验 结果如图 5 所示,相比于聚类前,聚类后的权值占 用的存储开销从 28.4 Mb (784 × 1024 × 16 bit + 1024 × 1024 × 16 bit + 1024 × 10 × 16 bit)下 降至 0.75 kb,但网络推理精度仅下降了约 1%.同 时,由于聚类后权值精度固定为 16-bit,软件到硬 件的映射过程不存在截断误差或舍入误差造成的 精度损失,因此,权值聚类后的网络可以以同样的 精度从软件平台映射到硬件平台.





Fig. 5. The influence of weight sharing technology on accuracy.

权值量化、权值共享技术通过牺牲很少的网络 精度大幅度降低了权值数量与硬件平台资源开销, 是一种对硬件设计十分友好的技术.通过分析该技 术与网络精度的变化发现,聚类使得突触的权值向 不同方向偏移产生±ΔW,但是宏观上权值偏移量 存在互补,因此网络精度变化不明显.从另一方面 来讲,神经网络的连接关系相比于权值大小更为关键,在一定的网络架构下轻微调整权值大小不会对网络精度造成较大影响.

3.3 基于忆阻器阵列的计算核设计

基于忆阻器阵列的计算核架构如图 6 所示,其 中包含一个脉冲整形单元、一个多位的 D 触发器、 一个由忆阻器构成的权重矩阵、多位行波计数器和 一个忆阻器控制器.脉冲整形单元实现了对输入脉 冲进行计数,并依据每个脉冲对应的索引值,将输 入脉冲序列转换成 16 个通道的脉冲数据流,每个 通道内的脉冲对应相同的索引.16 个通道中的脉 冲数据通过 D 触发器传递给权重矩阵进行计算. 多位行波计数器统计每个突触的权重,并进行膜电 位累加.忆阻器控制器可以对忆阻器的阻值进行擦 写,同时也可以将每个"time step"计算后的膜电位 与阈值电压进行对比,从而控制输出脉冲的发放.

在一次计算过程中,脉冲整形器首先对输入脉冲计数,并将其整形并预存到不同通道对应的缓存 区间中.每一个通道的脉冲依次送入到权值矩阵中 进行权值索引,同时多位行波计数器对权值进行累加.直至所有脉冲输入完毕,忆阻器控制器会将行 波计数器的加结果与上一个"time sep"保存的膜电 位相加,并与阈值电压相对比,控制脉冲发放.如 果膜电位超过阈值电压,忆阻器控制器会发放一个 脉冲输出,同时清零并保存膜电位;如果膜电位小 于等于阈值电压,其会保存当前膜电位用于下一个 "time step"的计算,且不会发放脉冲.



图 6 基于忆阻器阵列的计算核架构

Fig. 6. Computing core architecture based on resistive random access memory matrix.



图 7 基于硬件加速器的应用架构图 Fig. 7. Application architecture diagram based on hardware accelerator.

3.4 整体结构

基于计算核的设计,构建了具有三个计算核的硬件加速器架构,并设计了完整的应用架构,如图7 所示.软件层通过泊松编码将原始 MNIST 数据集的 帧图像转换成脉冲图像,并送入硬件加速器进行计算.硬件加速器包含三个前文所述计算核心,可以同 时进行三层网络计算.硬件加速器在计算完成后,其 脉冲输出被采集、统计并分析,最终得出计算结果.

3.5 推理网络模型到硬件加速的映射

推理网络模型到硬件加速的映射关系示意图 如图 8 所示.示例中,输入的帧图像为"9",其通过 编码转换成 N个"time step"的脉冲图像,并输入 到硬件加速器中.硬件加速器中的三个计算核组成 了一个三层的计算网络,第一层接受输入的图像脉 冲数据并启动计算,最后一层将输出脉冲作为计算 结果输出至片外.软件层统计硬件加速器的输出脉 冲数,当 N个"time step"的脉冲数据计算完毕后, 软件层将输出脉冲数量做对比,选择累计发放脉冲 最多的神经元所对应的标签为本次输入对应的计 算结果.图8中,输出层对应标签为"9"的神经元 累计发放的脉冲数量最多,因此,软件层将本次计 算结果统计为"9",与输入脉冲图像标签相符,是 一次正确的计算.





Fig. 8. Mapping diagram of network to hardware accelerator.

4 仿真结果及分析

为验证加速器的效果,采用 Verilog 对模拟忆 阻器阵列设计,并实现整体硬件加速器平台设计, 通过 SPICE 仿真获取内部延时;采用 UVM 搭建 验证平台;采用 Synopsis 公司的 VCS2018 软件进 行编译、仿真,并对 MNIST 数据集中的 10000 张 图片进行识别.

4.1 识别精度

在软件层训练了 784×1024×1024×10 的全连 接 ANN 网络, 并通过脉冲转换的方法, 将其转换 成 SNN 网络. 软件仿真表明, 其在测试集推理精 度为 96.5%. 之后, 在软件层通过 3.2 节所述的权 值量化的方法, 分别将三层网络的突触权值量化 成 16 个 16-bit 的权值数据, 同时为每个突触生成 一个 4-bit 的索引数据. 忆阻器控制器通过对忆阻 器阵列的扫描, 将权值数据写入到忆阻器阵列中. 阈值电压、泄漏电压等配置参数被配置到忆阻器控 制器中. 软件仿真表明, 权值量化后的 SNN 网络 在测试集的推理精度为 96.4%, 相比于量化前, 精 度下降了 1%, 但是存储开销降低了约 99.997%.

4.2 识别速度

为验证加速器对不同规模网络的加速效果,在不考虑精度的前提下,本工作分别将4种不同规模的网络部署到硬件加速器平台,对MNIST数据集进行推理测试.4种不同的网络规模分别为784×1024×1024×10,784×1024×2048×10,784×1024×4096×10和784×1024×8192×10.

测试结果如图 9 所示. 实验结果表明, 加速器 对于 784 × 1024 × 1024 × 10 的网络计算速率为 148.2 frames/s, 随着网络规模的增长, 其计算速率 呈线性下降. 通过理论分析发现, 由于硬件加速器 的三个计算核心通过并行计算对网络推理进行加 速, 因此加速器对网络的加速效率取决于规模最大 的一层网络, 即计算量最大的一层网络. 硬件加速 器的加速效率随着神经网络中最大规模的一层网 络的规模增加而线性下降.

此外,该加速器对小规模网络表现出了优秀的 加速效果,而对大规模网络的加速效果不明显.因 此,该硬件加速器架构更适用于计算量相对较小、 资源相对匮乏的边缘计算.



图 9 硬件加速器对不同规模神经网络计算的加速效率 Fig. 9. Acceleration efficiency of hardware accelerator for different scale neuron networks.

4.3 与其他类似工作的对比

表1所列为本次工作与近年来其类似工作的 对比. 从表1中可以看出,本次工作相较于其他工 作,使用忆阻器实现了更为简单的 LIF 神经元模 型,通过更少的忆阻器资源开销 (0.75k) 集成了更 大规模的神经元/突触. 同时,在 MNIST 数据集的 识别工作中拥有相对较高的推理准确率 (96.4%).

表 1 本次工作与其他类似工作的对比 Table 1. Comparison of this work with other works.

	1			
	Burr et al. ^[21]	Peng et al. ^[24]	Huang et al. ^[25]	This work
神经元模型	MP	MP	MP	LIF
突触数	161k	1k	$\sim 4.8 \mathrm{M}$	192M
神经元数	385	128	~6.4k	24k
忆阻器开销	330k	1k	1.11M	0.75k
测试数据集	MNIST	_	MNIST	MNIST
准确率/%	94.0—97.0	91.7	93.4	96.4

5 总结与展望

在本次工作中, 对基于 LIF 神经元模型的脉 冲神经网络进行了详细阐述, 并基于忆阻器阵列及 其外围电路实现了 LIF 神经元的硬件设计. 基于 神经元/突触复用技术和权值共享技术的权值存储 阵列可以通过 0.25k 的忆阻器存储 64M 突触的权 值. 此外, 通过电阻分压的方式对忆阻器阵列输出 进行电流-电压转换, 并通过数字电路进行采集. 基 于忆阻器阵列设计, 提出了神经形态计算核设计. 计算核可以实现至多 8k 神经元、64M 突触的计算. 最终, 通过三个神经形态计算核构建了脉冲神经网 络推理加速器设计,并开展了仿真、验证工作.在 加速器中部署了784 × 1024 × 1024 × 10的三层 全连接脉冲神经网络,并对 MNIST 数据集进行了 推理验证.试验结果表明,该加速器设计可以实现 148.2 frames/s的推理计算,同时拥有96.4%的推 理准确率.

受限于忆阻器存储材料的工作频率,该架构的 硬件加速器最高工作频率为 50 MHz,未来随着工 艺的发展,忆阻器的工作主频将进一步提升,硬件 加速器的性能也能随之得以提升.此外,随着部署 在加速器中的网络规模不断提升,加速器的计算效 率逐渐下降,因此,该加速器设计适合部署在资源 匮乏、算力需求低的边缘设备中进行加速工作.在 算力需求较高的应用环境下,则需要扩大设计规 模,这也就意味着需要更大规模的忆阻器阵列和外 围电路.

参考文献

- Redmon J, Farhadi A 2017 30th IEEE Conference on Computer Vision & Pattern Recognition Honolulu, HI, July 21–26, 2017 pp6517–6525
- [2] Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, Chen Y, Lillicrap T, Hui F, Sifre L, Van Den Driessche G, Graepel T, Hassabis D 2017 Nature 550 354
- [3] McCulloch W S, Pitts W 1943 Bull. Math. Biophys. 5 115
- [4] Hodgkin A L, Huxley A F 1952 J. Physiol. 116 449
- [5] Gerstner W 1995 Phys. Rev. E:Stat. Phys. Plasmas Fluids Relat. Interdisciplin. Top. 51 738
- [6] Maass W 1997 Neural Networks 10 1659
- [7] Roy K, Jaiswal A, Panda P 2019 Nature 575 607
- [8] Chen Y R, Li H, Chen Y Z, Chen F, Li S C, Liu C C, Wen W J, Wu C P, Yan B N 2018 Artif. Intell. View 13 46 (in Chinese) [陈怡然, 李海, 陈逸中, 陈凡, 李思成, 刘晨晨, 闻武杰,

吴春鹏, 燕博南 2018 人工智能 13 46]

- [9] Schuman C D, Potok T E, Patton R M, Birdwell J D, Dean M E, Rose G S, Plank J S2017 arXiv:1705.06963
- [10] Mahapatra N R, Venkatrao B 1999 Crossroads 5 2
- [11] von Neumann J 1993 IEEE Ann. Hist. Comput. 15 27
- [12] Chen T, Du Z, Sun N, Wang J, Wu C, Chen Y, Temam O 2014 Acm Sigplan Notices 49 269
- [13] Benjamin B V, Gao P, Mcquinn E, Chou D Hary S, Chandrasekaran A R, Bussat J, Alvarez-Icaza R, Arthur J V, Merolla P A, Boahen K 2014 Proc. IEEE 102 699
- [14] Pei J, Deng L, Song S, Zhao M G, Zhang Y H, Wu S, Wang G R, Zou Z, Wu Z Z, He W, Chen F, Deng N, Wu S, Wang Y, Wu Y J, Yang Z Y, Ma C, Li G Q, Han W T, Li H L, Wu H Q, Zhao R, Xie Y, Shi L P 2019 Nature 572 106
- [15] Davies M, Srinivasa N, Lin T H, Chinya G, Cao Y, Choday S H, Dimou G, Joshi P, Imam N, Jain S 2018 *IEEE Micro* 38 82
- [16] Akopyan F, Sawada J, Cassidy A, Alvarez-Icaza R, Arthur J, Merolla P, Imam N, Nakamura Y, Datta P, Nam G J 2015 *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 34 1537
- [17] Furber S B, Galluppi F, Temple S, Plana L A 2014 Proc. IEEE 102 652
- [18] Li K, Cao R R, Sun Y, Liu S, Li Q J, Xu H 2019 Micro/nano Electron. Intell. Manuf. 1 87 (in Chinese) [李锟, 曹荣荣, 孙毅, 刘森, 李清江, 徐晖 2019 微纳电子与智能制造 1 87]
- [19] Xia Q F, Yang J J 2019 Nat. Mater. 18 309
- [20] Deng Y B, Wang Z W, Zhao C H, Li L, He S, Li Q H, Shuai J W, Guo D H 2021 Appl. Res. Comput. 38 2241 (in Chinese) [邓亚彬, 王志伟, 赵晨晖, 李琳, 贺珊, 李秋红, 帅建伟, 郭东辉 2021 计算机应用研究 38 2241]
- [21] Burr G W, Shelby R M, Sidler S, Nolfo C D, Jang J, Boybat I, Shenoy R S, Narayanan P, Virwani K, Giacometti E U 2015 *IEEE Trans. Electron Devices* **62** 3498
- [22] Moro F, Hardy M, Fain B, Dalgaty T, Clemencon P, De Pra A, Esmanhotto E, Castellani N, Blard F, Gardien F, Mesquida T, Rummens F, Eseni D, Casas J, Indiveri G, Payvand M, Vianello E 2022 Nat. Commun. 13 3506
- [23] Fang X D, Wu J J 2020 Comput. Eng. Sci. 42 1929 (in Chinese) [方旭东, 吴俊杰 2020 计算机工程与科学 42 1929]
- [24] Peng Y, Wu H, Gao B, Eryilmaz S B, Qian H 2017 Nat. Commun. 8 15199
- [25] Huang L, Diao J T, Nie H S, Wang W, Li Z W, Li Q J, Liu H J 2021 Front. Neurosci. 15 639526
SPECIAL TOPIC—Physical electronics for brain-inspired computing

Memristor based spiking neural network accelerator architecture^{*}

Wu Chang-Chun Zhou Pu-Jun Wang Jun-Jie Li Guo Hu Shao-Gang Yu Qi Liu Yang[†]

(School of Electronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China) (Received 13 January 2022; revised manuscript received 29 May 2022)

Abstract

Spiking neural network (SNN) as the third-generation artificial neural network, has higher computational efficiency, lower resource overhead and higher biological rationality. It shows greater potential applications in audio and image processing. With the traditional method, the adder is used to add the membrane potential, which has low efficiency, high resource overhead and low level of integration. In this work, we propose a spiking neural network inference accelerator with higher integration and computational efficiency. Resistive random access memory (RRAM or memristor) is an emerging storage technology, in which resistance varies with voltage. It can be used to build a crossbar architecture to simulate matrix computing, and it has been widely used in processing in memory (PIM), neural network computing, and other fields. In this work, we design a weight storage matrix and peripheral circuit to simulate the leaky integrate and fire (LIF) neuron based on the memristor array. And we propose an SNN hardware inference accelerator, which integrates 24k neurons and 192M synapses with 0.75k memristor. We deploy a three-layer fully connected network on the accelerator and use it to execute the inference task of the MNIST dataset. The result shows that the accelerator can achieve 148.2 frames/s and 96.4% accuracy at a frequency of 50 MHz.

Keywords: spiking neural networks, resistive random access memory, processing in memory, leaky integrate and fire model, hardware inference accelerator

PACS: 84.35.+i, 85.40.-e, 95.75.Mn

DOI: 10.7498/aps.71.20220098

^{*} Project supported by the National Natural Science Foundation of China (Grant No. 92064004).

[†] Corresponding author. E-mail: yliu1975@uestc.edu.cn





Institute of Physics, CAS

基于层状多元金属氧化物的人造突触

刘强 倪尧 刘璐 孙林 刘甲奇 徐文涛

Artificial synapses based on layered multi-component metal oxides Liu Qiang Ni Yao Liu Lu Sun Lin Liu Jia-Qi Xu Wen-Tao 引用信息 Citation: Acta Physica Sinica, 71, 148501 (2022) DOI: 10.7498/aps.71.20220303 在线阅读 View online: https://doi.org/10.7498/aps.71.20220303 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

基于水热法制备三氧化钼纳米片的人工突触器件 Artificial synapse based on MoO₃ nanosheets prepared by hydrothermal synthesis 物理学报. 2020, 69(23): 238501 https://doi.org/10.7498/aps.69.20200928

氧化物基忆阻型神经突触器件

Oxide-based memristive neuromorphic synaptic devices 物理学报. 2019, 68(16): 168504 https://doi.org/10.7498/aps.68.20191262

基于二维材料MXene的仿神经突触忆阻器的制备和长/短时程突触可塑性的实现

Fabrication of synaptic memristor based on two-dimensional material MXene and realization of both long-term and short-term plasticity

物理学报. 2019, 68(9): 098501 https://doi.org/10.7498/aps.68.20182306

有机金属卤化物钙钛矿中的离子迁移现象及其研究进展 Recent progress of ion migration in organometal halide perovskite 物理学报. 2019, 68(15): 158801 https://doi.org/10.7498/aps.68.20190853

钠离子层状氧化物材料相变及其对性能的影响

Phase transitions of Na-ion layered oxide materials and their influence on properties 物理学报. 2022, 71(10): 108801 https://doi.org/10.7498/aps.71.20220291

浅析电解质中离子输运的微观物理图像

Brief overview of microscopic physical image of ion transport in electrolytes 物理学报. 2020, 69(22): 226601 https://doi.org/10.7498/aps.69.20201519

专题: 面向类脑计算的物理电子学

基于层状多元金属氧化物的人造突触*

刘强 倪尧 刘璐 孙林 刘甲奇 徐文涛†

(南开大学光电子薄膜器件与技术研究所,天津市光电子薄膜器件与技术重点实验室,天津 300350)

(2022年2月19日收到; 2022年4月10日收到修改稿)

神经形态电子学的迅速发展为生物神经系统仿生与模拟提供了有力支持.具有三明治结构的两端人造 突触电子器件不仅在结构上模拟了生物突触,同时在类神经电脉冲信号的作用下可以完成对生物突触塑性 的模拟与调控.本文利用溶胶-凝胶法合成了具有层状结构的 P3 相 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 多元金属氧化物.借助 其晶体结构中 Na⁺易于嵌入/脱出的特性,设计并制备了基于 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 的离子迁移型人造突触,器件 在电脉冲信号的刺激下实现了对生物突触塑性的模拟,并通过调校类神经尖峰脉冲信号,成功对塑性行为进 行了调控.成功模拟了兴奋性突触后电流、双脉冲易化、脉冲数量依赖可塑性、脉冲频率依赖可塑性、脉冲电 压幅值依赖可塑性和脉冲持续时间依赖可塑性.同时,器件实现了对摩斯电码指令的准确识别与响应.

关键词:人造突触,离子迁移,多元金属氧化物,突触塑性 PACS: 85.35.-p,73.40.Mr,77.22.Ej

DOI: 10.7498/aps.71.20220303

1 引 言

近年来为进一步优化神经形态计算工作模式, 实现工作效率大幅提升,突触电子学应运而生^[1-3], 将电子器件的电学属性与生物突触的塑性行为紧 密联系,有望打破目前利用传统计算机设备模拟神 经网络时所面临的存储与计算难以并行、资源配置 严重失衡且发展缓慢的现状^[4,5].不同类别的人造 突触电子器件陆续被用以模拟生物突触,旨在实现 对突触塑性行为进行模拟,并通过主动干预的手段 对其进行调控,有望打破冯·诺伊曼瓶颈的桎梏,建 立新一代的神经形态计算模式.

两端人造突触电子器件具有与忆阻器等电子 元件相似的物理结构,工作机制与生物突触更为相 近,凭借响应时间短、超低功耗、易于集成、应用场 景丰富等特点受到了广泛关注^[6-9].常被设计为金 属电极-活性层-金属电极的三明治结构,其中活性 层的选取在一定程度上决定了器件整体的工作水 平. 在近年的研究中,利用卤化物钙钛矿[10,11]、过 渡金属氧化物^[12,13]、有机/无机半导体^[14]、无机盐^[15] 等多类光电半导体材料作为器件活性物质所构建 的人造突触,展现出了优异的塑性可调控能力,为 神经形态电子器件的发展提供了新思路. 而具有层 状结构的过渡金属氧化物,由于其在外界激励下易 实现电化学掺杂/脱掺杂[16-18]的特性,更接近于生 物突触电信号-化学信号-电信号的神经信号传递形 式,是电化学型人造突触电子器件中极具发展潜力 的活性功能材料. 具有 P3 相层状结构的多元金属 氧化物 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 作为一种优良的离子传 输材料, Na+在其晶格中可实现往复穿梭^[19-21]. 在正电压的作用下,位于外部电解质中的 Na+可 在电压作用下嵌入其晶格内部,占据过渡金属原子 层间的 Na 原子位点; 反之, 当撤去电压刺激后, Na_{2/3}Ni_{1/3}Mn_{2/3}O₂中所嵌入的 Na⁺脱出晶格后返 回电解质层^[22-24].

© 2022 中国物理学会 Chinese Physical Society

^{*} 国家杰出青年科学基金 (批准号: T2125005)、天津市杰出青年科学基金 (批准号: 19JCJQJC61000) 和深圳市科技计划 (批准号: JCYJ20210324121002008) 资助的课题.

[†] 通信作者. E-mail: wentao@nankai.edu.cn

本文通过溶胶-凝胶法合成了多元层状金属氧 化物 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂,基于其 Na⁺的电化学迁 移特性设计并制备了两端离子迁移型人造突触.在 类神经电脉冲信号的激励下,器件可对生物突触可 塑性进行模拟与调控,如兴奋性突触后电流 (excitatory postsynaptic current, EPSC)、双脉冲易化 (paired-pulse facilitation, PPF)、脉冲数量依赖可 塑性 (spike-number dependent plasticity, SNDP)、 脉冲频率依赖可塑性 (spike-frequency dependent plasticity, SFDP)、脉冲电压依赖可塑性 (spikevoltage dependent plasticity, SVDP)、脉冲持续时 间依赖可塑性 (spike-duration dependent plasticity, SDDP)等.此外,当对人造突触施加摩斯电 码制式的电信号激励时,器件可以准确识别摩斯电 码内容并做出响应.

2 实验方法

2.1 材料制备

活性材料 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 常通过溶胶-凝胶 法、固相法、共沉淀法、水热法等多种手段合成^[25,26]. 考虑到溶胶-凝胶法过程安全、易于操作且合成样 品结晶质量好等优势,本文采用此方法对目标材 料 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂进行制备.首先将称量好的 摩尔比例为 2:1:1 的乙酸钠、乙酸镍、乙酸锰置于 样品瓶中并溶于去离子水,室温条件下经磁力搅拌 后得到浅绿色均一澄清的溶液.向其中加入一定质 量的螯合剂一水合柠檬酸,并在 80 ℃ 时持续搅拌 使溶液逐渐稠化变为溶胶,之后持续加热实现溶胶 向凝胶的转变.将得到的凝胶置于鼓风烘箱内于 80 ℃ 烘干 8 h 后取出研磨并做压片处理.为除去 额外加入的一水合柠檬酸,压制所得的药片需在 400 ℃ 的条件下退火处理 3 h. 此外,为保证反应 充分进行,在除去螯合剂后药片仍需在 900 ℃ 条 件下二次退火处理 12 h. 经研磨后,便可得到结晶 良好的 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末.

2.2 器件设计与机制

突触作为神经系统中最基本的结构单元,承载 着神经信号传递的职能^[27].如图 1(a)所示,生物突 触由相互联结的前级神经元的轴突、后级神经元 的树突及二者间的间隙构成.在动作电位的刺激 下前级神经元轴突尾部的离子通道打开,神经元 内的离子开始自由扩散,导致突触前膜发生囊泡 的胞吐行为,向突触间隙中释放神经递质.最终, 神经递质与突触后膜上具有识别性的受体相结合, 将动作电位传递到后级神经元,形成了突触后电 流^[28-30].



图 1 (a) 生物神经元及突触结构示意图; (b) 人工突触电子器件结构示意图; (c) P3 相 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 结构示意图

Fig. 1. (a) Schematic diagram of biological neuron and synapse structure; (b) schematic diagram of artificial synaptic electronic device structure; (c) schematic diagram of the structure of P3 phase $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$.

结合生物突触的工作机制,设计了图1(b)所 示的具有三明治结构的两端人造突触电子器件用 来模拟生物突触.器件的制备在底电极铝箔上完成, 各组分制备顺序为: Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 涂层、PEO-Na电解质薄层、顶电极 Au. 首先, 将质量比例为 7:2:1的 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末、乙炔黑、质量分 数为4%的聚偏氟乙烯粘结剂(溶于N-甲基吡咯 烷酮) 置于研钵中均匀混合后, 采用规格为 15 µm 的涂布器在底电极铝箔上进行涂布.在 60 ℃ 的退 火处理之后, 溶剂充分挥发, 可以获得 Na2/3Ni1/3 Mn_{2/3}O₂ 涂层, 用以对前级神经元的结构与功能进 行模拟. 之后将铝箔裁剪为 2 cm×2 cm 的方形并 粘贴在与之大小相当的玻璃衬底上,以 3000 r/min, 30 s 的参数旋涂钠离子固态电解质溶液 (聚乙烯吡 咯烷酮 (PEO) 与 NaClO₄ 按照 8:1 的质量比溶于 乙腈构成质量分数为 10% 的 PEO-Na 电解质溶 液), 旋涂完成后按照 90 ℃ 的条件进行退火处理, 直至溶剂挥发完全后可得到 PEO-Na 电解质薄层, 实现对活性层表面的修饰[31,32]. 最后, 通过定制的 掩模版按照点状阵列的形状,在 PEO-Na 薄层上 进行蒸镀顶电极 Au. 根据所定制的掩模版尺寸, 点状电极的半径为 300 µm, 蒸镀过程中通过调整

参数可将电极厚度控制在 50 nm. 表面蒸镀了 Au 电极阵列的 PEO-Na 层对后级神经元进行了模拟. 金属氧化物活性层、PEO-Na 电解质层及其间界面 实现了对突触间隙的模拟^[33]. 图 1(c) 进一步形象 化地描述了具有 *P*3 相 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 活性层 的层状晶体结构, 每层过渡金属原子之间都有 Na 原子穿插其中.

3 实验结果与分析

3.1 材料物相结构及形貌测试

图 2(a) 为 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末的 X 射线 衍射测试图谱, 发现其存在 10 个衍射特征峰, 分 别在 2θ = 16.1°, 32.5°, 37.0°, 38.3°, 42.3°, 46.1°, 53.9°, 58.4°, 65.9°, 68.4°处. 将各特征峰与标准卡 片对比后可以发现其分别与 (003), (006), (101), (012), (014), (015), (107), (018), (110), (113) 晶 面相对应. 由此可知, Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 晶格为六 方晶系中具有 *P*3 相的层状结构^[34]. 在电脉冲信号 激励下, 处于过渡金属原子层之间的 Na 原子易于 在层间迁移并嵌入/脱出电解质薄层, 实现了层间 的电化学掺杂/脱掺杂, 与生物突触中间隙内信号 转变的过程类似.



图 2 (a) Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末 X 射线衍射测试图; (b) Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末扫描电子显微镜测试图; (c) Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末 X 射线能谱分析图; (d) Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 活性层扫描电子显微镜表面形貌测试图; (e) 底电极铝箔、Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 活性 层与 PEO-Na 电解质薄层扫描电子显微镜断面形貌测试图; (f) Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 活性层原子力显微镜测试图

Fig. 2. (a) X-ray diffraction test diagram of $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ powder; (b) scanning electron microscope test diagram of $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ powder; (c) EDS test diagram of $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ powder; (d) surface topography test diagram of $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ powder; (d) surface topography test diagram of $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ active layer scanning electron microscope ; (e) bottom electrode Al foil, $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ active layer and PEO-Na electrolyte thin layer scanning electron microscope cross-sectional morphology test diagram; (f) atom force microscope test diagram of $Na_{2/3}Ni_{1/3}Mn_{2/3}O_2$ active layer.

图 2(b) 为 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末在扫描电 子显微镜下的微观形貌,其X射线能谱分析的结 果如图 2(c) 所示. 不难看出, 粉末粒度均匀, 颗粒 尺度约为100 nm. 在能谱分析中, 除去粉末载体所 含的碳元素外,其余各元素所占权重之比与产物 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂的化学计量比相符,进一步证 明反应充分、合成产物可靠. 图 2(d) 为活性层扫描 电子显微镜表面形貌测试图,能清晰地看到涂层致 密、均一. 底电极铝箔、Na2/3Ni1/3Mn2/3O2活性层 与 PEO-Na 电解质薄层扫描电子显微镜断面形貌 测试图, 如图 2(e) 所示, 可以清晰地看到三层的层 堆叠结构, 经测量得底电极铝箔厚度约为 4 μm, Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 涂层厚度约为 7.5 µm, 与规格 为 10 μm 的涂布器误差范围相吻合, PEO-Na 电 解质薄层厚度约为 8.6 μm. 图 2(f) 为活性层在原 子力显微镜下所观测到的形貌, 对边长为 5 µm 的 方形区域进行采样,得到区域内的平均粗糙度为 84 nm,相较于活性层的厚度可忽略,由此可知所 制备的活性层较为均匀、平整.

3.2 人造突触电子器件塑性测试

由于神经元间相互联结且联结强度可调等特 点,生物突触产生了塑性行为,并称神经元间的联 结强度为突触权重. 在外界刺激下, 突触结构在动 作电位的驱动下进行塑性调节. 不同的刺激引发突 触权重的变化不同, 根据权重变化的特点, 将突触 的塑性行为分为短时程塑性 (STP) 和长时程塑性 (LTP) 两类. 其中, STP 是构建神经网络的核心基 础^[35-37], 其因塑性过程持续时间较短, 可应用于传 感、识别、计算等众多领域. EPSC 作为神经活动 中最基础的一种 STP, 是生物神经系统进行复杂 活动时所传递的基础信号之一^[38].

基于层状多元金属氧化物 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂的两端人造突触电子器件具有三明治结构, 底电极 铝箔与顶电极 Au 分别对突触前膜与后膜进行模 拟, 金属氧化物活性层与钠离子固态电解质薄层相 耦合,构成离子迁移系统对突触间隙进行模拟. 利 用金属探针将人造突触与外部测试设备相连,并对 器件进行阻变特性及塑性行为测试. 图 3(a) 为器 件在单次 0 V—5 V—0 V 电压扫描下的阻变特性 曲线,随着对器件施加电压的改变,器件的电阻发 生改变,即电解质层极化和活性层 Na⁺电化学掺杂 的程度随施加电压发生改变.同时可观察到,当施 加电压约为 2.5 V 时,电流响应明显增大,可解释 为 PEO-Na 电解质薄层与活性材料 Na_{2/3}Ni_{1/3} Mn_{2/3}O₂ 在施加电压为 2.5 V 时所发生的极化和



图 3 (a) 单次阻变特性测试; (b) 连续 50 次阻变特性稳定能力测试; (c) 对器件施加单个幅值为 0.2 V 的电脉冲信号所产生的 EPSC; (d) 对器件连续施加两个幅值为 0.2 V 的电脉冲信号所产生的 PPF; 对器件施加多对时间间隔不同幅值为 0.2 V 的电脉冲 信号所产生的 (e) PPF 以及 (f) PPF 指数

Fig. 3. (a) Single resistance characteristic test; (b) 50 consecutive tests of resistance characteristic stability; (c) EPSC generated by applying a single electrical pulse signal with an amplitude of 0.2 V to the device; (d) PPF generated by continuously applying two electrical pulse signals with an amplitude of 0.2 V to the device; (e) PPF and (f) PPF index generated by applying multiple pairs of electrical pulse signals with different amplitudes of 0.2 V to the device.

电化学掺杂现象格外强烈,引起器件阻态大幅改变. 图 3(b) 为器件在连续 50 次电压循环扫描下的 阻变特性稳定能力测试图,在 50 次连续的电压循环扫描下,器件的响应电流并未发生明显变化,体现了较好的电学稳定性.

当对铝箔底电极施加一恒定的读取电压 V_r (0.1 V), 对 Au 顶电极阵列施加瞬时突触前电压脉 冲 V_{Pre}时,器件表现出了较好的 STP. 当对顶电 极施加 V_{Pre}时, 顶电极端的 V_{Pre}大于底电极端的 Vr,存在由Au指向铝箔的电势差,电解质中的Na+ 在电势作用下发生极化,向活性层定向迁移,最终 嵌入到 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 的晶格中, 实现了电化 学掺杂. 根据其阻变特性曲线同样可以观测出脱 离 PEO-Na 的 Na+与活性层之间发生了电化学掺 杂,而并非只运动到两层材料的界面处. 当撤去电 压激励时,极化现象无法维系,电解质层发生去极 化,掺杂入活性层的部分 Na+会发生脱掺杂,返回 到 PEO-Na. 在电化学掺杂/脱掺杂过程中均存在 Na+的定向运动,因此导致电流的增加,但在撤去 激励后, 电流会逐渐衰减至初始水平^[39]. Na+由平衡 态产生极化至非平衡态,再经去极化恢复至平衡态 的运动过程所产生的离子电流即为 EPSC, 图 3(c) 为 $V_{\text{Pre}} = 0.2$ V 时的 EPSC, 突触后电流在 V_{Pre} 刺 激下瞬时增大到 5 nA, 撤去后逐渐衰减至 pA 水平.

PPF 作为一类典型的 STP 行为, 是完成学习 和信息处理等简单神经元任务的关键^[40],表现为 当两个连续的神经信号作用到前级神经元时,后级 神经元对第二个信号响应的突触后电流被放大,其 峰值总是会大于第一个神经信号. 如图 3(d) 所示, 当对顶电极施加幅值为 0.2 V 的两个连续 V_{Pre}时, 会产生两个突触后电流响应,所产生的第二个突触 后电流的峰值 A, 远大于前一个峰值 A1, 称 A2 与 A₁ 的比值为 PPF 指数,此时 PPF 指数为 123%. 当撤去第一个电脉冲时,离子交换系统发生去极 化, Na+恢复至平衡态的过程被施加的第二个电脉 冲信号破坏,导致系统再次极化,累积的离子浓度 变大,致使去极化时突触后电流增大.同时,改变 所施加两个电脉冲信号的时间间隔会影响易化率 大小,图 3(e)为 $V_{Pre} = 0.2$ V时,时间间隔分别为 50, 100, 150, 200 ms 时的 PPF 塑性行为. 可以看 出随着时间间隔变大,突触后电流峰值 A,逐渐变 小, 如图 3(f) 所示, 这表明 PPF 指数与时间间隔 呈负相关.

在生物神经系统中,由于神经信号数量庞大且 模式不同,难以简单归类处理,因此在塑性能力测 试过程中,通过调整 V_{Pre} 的数量、幅值、脉宽、频 率等参数来进一步实现对 SNDP, SVDP, SDDP, SFDP 等更为复杂的生物突触塑性行为的模拟十 分必要 ^[41,42].如图 4(a) 所示,当连续施加 10 个幅 值为 0.2 V 的 V_{Pre} 时,会产生 10 个峰值连续增大 的突触后电流响应,实现了对 SNDP 的模拟,最 后 1 个电流峰值 A_{10} 远大于第 1 个峰值 A_1 ,称 A_{10} 与 A_1 的比值为 n = 10 时的 SNDP 指数. 从图 4(b) 可知,所产生的突触后电流的峰值随电信号刺激 数量增加而增加. 随着脉冲数量的增加, SNDP 指数随之增大,但突触后电流的增益速度即曲线 斜率逐渐变小.

与之相似的是,当增加 VPre 的幅值、脉宽、频 率时,得到的突触后电流的峰值同样会随之变大, 与生物突触的 SVDP, SDDP, SFDP 等塑性行为 一致. 图 4(c) 为器件对 SVDP 的模拟, 当 V_{Pre} 的 幅值以 0.1 V 为步长, 按 0 V-4 V-0 V 的形式 循环 10 次时,突触后电流峰值会按照电脉冲幅值 的变化同步发生改变,由于器件在去极化与脱掺杂 过程中,所展现出较好的 STP,在 10 次脉冲信号 循环的过程中突触后电流的峰值变化呈现出较好 的重复性,体现了器件较好的循环保持能力.在对 SDDP 进行模拟过程中,设置 $V_{\rm Pre}$ 的幅值为 0.2 V, 脉宽从 50 ms 均匀增加到 500 ms时,突触后电流 的峰值逐渐增大,如图 4(d), (e) 所示. 规定脉宽为 50 ms的电信号刺激所产生的电流响应峰值为 A_1 , 随着脉宽逐渐均匀增加, 电流响应的峰值分别命名 为 A2-A10, 各峰值与 A1 的比值定义为 SDDP 指 数, SDDP 指数也随脉宽的增大而增大, 而电流的 增益速度即曲线斜率逐渐变小. SFDP 体现了突触 对神经信号的选频特性,如图 4(f) 所示,将连续 10个幅值为 0.2 V的 Vpre 的频率分别设置为 1.25, 2.50, 5.00, 10.00 Hz, 突触后电流峰值随脉冲 频率增大而增大,体现出突触器件对不同频率的分 辨能力.以上通过调整 VPre 的幅值、脉宽、频率等 参数而实现的塑性调控行为,是由于 Na+由非平衡 态恢复至平衡态的过程中被新的 VPre 打断, 再一 次发生极化,实现离子存储,因而增大了突触后电 流的响应峰值. 但因脉冲刺激强度有限且离子交换 系统中可激发的 Na+数量固定, 导致突触后电流增 长速度逐渐减小[43].



图 4 对器件连续施加 10 个幅值为 0.2 V 的电脉冲信号所产生的 (a) SNDP 以及 (b) SNDP 指数; (c) 对器件施加幅值从 0 V—4 V—0 V 变化的 10 组电脉冲信号循环所产生的 SVDP; 对器件施加多个脉冲宽度不同幅值为 0.2 V 的电脉冲信号所产生的 (d) SDDP 以及 (e) SDDP 指数; (f) 对器件连续施加多组频率不同幅值为 0.2 V 的电脉冲信号所产生的 SFDP

Fig. 4. (a) SNDP and (b) SNDP index generated by continuously applying 10 electrical pulse signals with an amplitude of 0.2 V to the device; (c) 10 groups of amplitudes varying from 0 V to 4 V to 0 V are applied to the device SVDP generated by electrical pulse signal cycle; (d) SDDP and (e) SDDP index generated by applying multiple electrical pulse signals with different pulse widths and amplitudes of 0.2 V to the device; (f) SFDP generated by continuously applying multiple groups of electrical pulse signals with the different frequencies and amplitudes of 0.2 V to the device.



图 5 对器件施加内容为 Na_{2/3} (a), Ni_{1/3} (b), Mn_{2/3} (c), O₂ (d) 的摩斯电码制式的电脉冲信号所产生的突触后电流响应 Fig. 5. Post-synaptic current response generated by applying Morse code electrical pulse signals with content of (a) Na_{2/3}, (b) Ni_{1/3}, (c) Mn_{2/3}, (d) O₂ to the device.

如图 5 所示, 基于层状多元金属氧化物 Na_{2/3} Ni1/3Mn2/3O2的两端人造突触电子器件,成功实现 了对摩斯电码的识别与响应[44,45]. 对摩斯电码的基 础字符"·"和"—"进行定义,将"·"编码为由单个脉 宽为 50 ms, 幅值为 0.2 V 的电压脉冲触发的信号; "—"为由连续3个脉宽为50ms,幅值为0.2V的 电压脉冲触发的信号.按照国际摩斯电码表,以活 性层材料名称"Na2/3Ni1/3Mn2/3O2"作为传递目标 信息,按照摩斯电码制式将携带其名称信息的脉冲 序列传递至突触器件,规定水平处于 5—6 nA 区 间的电流为"·"所触发,水平处于 8.5—9.5 nA 区间 的电流为"—"所触发. 通过划分突触后电流峰值的 水平所在区间的方法,可判断出脉冲信号所传输信 息的内容,准确读取传递信息中不同的英文字母、 数字和符号.这种以人造突触为媒介,利用生物神 经系统的工作模式解决问题的方式为神经形态电 子器实现类脑计算提供了新的应用场景[46],一定 程度上解决了传统冯·诺依曼计算机难以实现存算 一体且资源消耗巨大等问题.

4 结 论

本文通过溶胶-凝胶法合成了结晶良好的层状 多元金属氧化物 Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ 粉末,与钠离 子固态电解质 (PEO-Na) 薄层构成了离子迁移系 统,并将其制备为具有三明治结构的两端人造突触 电子器件.在类神经电脉冲信号的刺激下,Na⁺在 离子迁移系统中实现了电化学掺杂/脱掺杂,器件 实现了对生物突触中 EPSC, PPF, SNDP, SFDP, SVDP, SDDP 等塑性行为的模拟与调控,同时对 摩斯电码发出指令内容可以准确地识别与响应.

参考文献

- [1] Kuzum D, Yu S, Wong H P 2013 Nanotechnology 24 382001
- [2] Ling H, Koutsouras D A, Kazemzadeh S, Van De Burgt Y, Yan F, Gkoupidenis P 2020 Appl. Phys. Rev. 7 011307
- [3] Wang S, Zhang D W, Zhou P 2019 Sci. Bull. 64 1056
- [4] Wei H, Shi R, Sun L, Yu H, Gong J, Liu C, Xu Z, Ni Y, Xu J, Xu W 2021 Nat. Commun. 12 1
- [5] Choi D, Song M K, Sung T, Jang S, Kwon J Y 2020 Nano Energy 74 104912
- [6] Xia Q, Yang J J 2019 Nat. Mater. 18 309
- [7] Lu K, Li X, Sun Q, Pang X, Chen J, Minari T, Liu X, Song Y 2021 Mater. Horiz. 8 447
- [8] Sun J, Fu Y, Wan Q 2018 J. Phys. D: Appl. Phys. 51 314004
- [9] Gao J, Zheng Y, Yu W, Wang Y, Jin T, Pan X, Loh K P, Chen W 2021 Smart Mater. 2 88

- [10] Jeong B, Gkoupidenis P, Asadi K 2021 Adv. Mater. 33 2104034
- [11] Huang X, Li Q, Shi W, Liu K, Zhang Y, Liu Y, Wei X, Zhao Z, Guo Y, Liu Y 2021 *Small* 17 2102820
- [12] Wang C, Liu H, Chen L, Zhu H, Ji L, Sun Q Q, Zhang D W 2021 IEEE Electron Device Lett. 42 1555
- [13] Huang H, Liu L, Jiang C, Gong J, Ni Y, Xu Z, Wei H, Yu H, Xu W 2022 Neuromorph. Comput. Eng. 2 014004
- [14] Keene S T, Lubrano C, Kazemzadeh S, Melianas A, Tuchman Y, Polino G, Scognamiglio P, Cina L, Salleo A, van de Burgt Y, Santoro F 2020 Nat. Mater. 19 969
- [15] Ku B, Koo B, Sokolov A S, Ko M J, Choi C 2020 J. Alloys Compd. 833 155064
- [16] Yan Y, Chen Q, Wu X, Wang X, Li E, Ke Y, Liu Y, Chen H, Guo T 2020 ACS Appl. Mater. Interfaces 12 49915
- [17] Wang H, Zhao Q, Ni Z, Li Q, Liu H, Yang Y, Wang L, Ran Y, Guo Y, Hu W 2018 Adv. Mater. 30 1803961
- [18] Wei H, Yu H, Gong J, Li R, Han H, Ma M, Guo K, Xu W 2021 Mater. Chem. Front. 5 775
- [19] Gao J, Hao Y, Xu S, Rong X, Lu Q, Zhu C, Hu Y S 2021 *Electrochim. Acta* **399** 139421
- [20] Wang D, Xu S, Wang J, Rong X, Zhou F, Wang L, Bai X, Lu B, Zhu C, Wang Y, Hu Y S 2022 Energy Storage Mater. 45 92
- [21] Kong L, Tang C, Peng H J, Huang J Q, Zhang Q 2020 Smart Mater. 1 e1007
- [22] Liu Q, Hu Z, Chen M, Zou C, Jin H, Wang S, Gu Q, Chou S 2019 J. Mater. Chem. A 7 9215
- [23] Zhang S Y, Guo Y J, Zhou Y N, Zhang X D, Niu Y B, Wang E H, Huang L B, An P F, Zhang J, Yang X A 2021 Small 17 2007236
- [24] Xian L, Li M, Qiu D, Qiu C, Yue C, Wang F, Yang R 2022 J. Alloys Compd. 905 163965
- [25] Song T, Kendrick E 2021 J. Phys. :Mater. 4 032004
- [26] Yu M, Liu F, Li J, Liu J, Zhang Y, Cheng F 2021 Adv. Energy Mater. 12 2100640
- [27] Yang X, Specht C G 2019 Front. Mol. Neurosci. 12 161
- [28] Lu L, Jia Y, Kirunda J B, Xu Y, Ge M, Pei Q, Yang L 2019 *Nonlinear Dyn.* 95 1673
- [29] Beckstead M J, Grandy D K, Wickman K, Williams J T 2004 Neuron 42 939
- [30] Shipman S L, Nicoll R A 2012 Proc. Natl. Acad. Sci. 109 19432
- [31] Hayashi A, Masuzawa N, Yubuchi S, Tsuji F, Hotehama C, Sakuda A, Tatsumisago M 2019 Nat. Commun. 10 1
- [32] Wei H, Yu H, Gong J, Zhang J, Han H, Ma M, Ni Y, Du Y, Zhang S, Liu L, Xu W 2019 ACS Appl. Electron. Mater. 2 316
- [33] Wen Y, Wang B, Zeng G, Nogita K, Ye D, Wang L 2015 *Chem. Asian J.* **10** 661
- [34] Huang Q, Xu S, Xiao L, He P, Liu J, Yang Y, Wang P, Huang B, Wei W 2018 Inorg. Chem. 57 15584
- [35] Magee J C, Grienberger C 2020 Annu. Rev. Neurosci. 43 95
- [36] Li Y, Zhong Y, Zhang J, Xu L, Wang Q, Sun H, Tong H, Cheng X, Miao X 2014 Sci. Rep. 4 1
- [37] Van Rossum M C, Bi G Q, Turrigiano G G 2000 J. Neurosci.
 20 8812
- [38] Fang L, Dai S, Zhao Y, Liu D, Huang J 2020 Adv. Electron. Mater. 6 1901217
- [39] Yang K, Yang L, Wang Z, Guo B, Song Z, Fu Y, Ji Y, Liu M, Zhao W, Liu X 2021 Adv. Energy Mater. 11 2100601
- [40] López J C 2001 Nat. Rev. Neurosci. 2 307
- [41] Gong J, Yu H, Zhou X, Wei H, Ma M, Han H, Zhang S, Ni

Y, Li Y, Xu W 2020 Adv. Funct. Mater. 30 2005413

- [42] Guo K X, Yu H Y, Han H, Wei H H, Gong J D, Liu L, Huang Q, Gao Q Y, Xu W T 2020 Acta Phys. Sin. 69 238501 (in Chinese) [郭科鑫, 于海洋, 韩弘, 卫欢欢, 龚江东, 刘璐, 黄 茜, 高清运, 徐文涛 2020 物理学报 69 238501]
- [43] Zhang S, Guo J, Liu L, Ruan H, Kong C, Yuan X, Zhang B, Gu G, Cui P, Cheng G 2022 Nano Energy 91 106660
- [44] Lee Y, Oh J Y, Xu W, Kim O, Kim T R, Kang J, Kim Y, Son D, Tok J B H, Park M J 2018 Sci. Adv. 4 eaat7387
- [45] Shim H, Jang S, Jang J G, Rao Z, Hong J I, Sim K, Yu C 2022 Nano Res. 15 758
- [46] Yang F, Sun L, Duan Q, Dong H, Jing Z, Yang Y, Li R, Zhang X, Hu W, Chua L 2021 Smart Mater. 2 99

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Artificial synapses based on layered multi-component metal oxides^{*}

Liu Qiang Ni Yao Liu Lu Sun Lin Liu Jia-Qi Xu Wen-Tao[†]

(Key Laboratory of Photoelectronic Thin Film Devices and Technology of Tianjin, Institute of Photoelectronic Thin Film Devices and Technology, Nankai University, Tianjin 300350, China)

(Received 19 February 2022; revised manuscript received 10 April 2022)

Abstract

Neuromorphic electronics has received considerable attention recent years, and its basic functional units are synaptic electronic devices. A two-terminal artificial synapse with sandwiched structure emulates plasticity of the biological synapses under the action of nerve-like electrical impulse signals. In this paper, P3 phase Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ multi-element metal oxides with layered structure are synthesized by sol-gel process. Owing to the fact that Na⁺ is easy to embed/eject into its crystal structure, an ion-migrating artificial synapse based on Na_{2/3}Ni_{1/3}Mn_{2/3}O₂ is designed and fabricated. The device emulates important synaptic plasticity, such as excitatory postsynaptic current, paired-pulse facilitation, spike-number dependent plasticity, spike-frequency dependent plasticity, spike-voltage amplitude dependent plasticity and spike-duration dependent plasticity. The device realizes the identification and response to Morse code commands.

Keywords: artificial synapses, ion transport, multi-metal oxides, synaptic plasticity PACS: 85.35.-p, 73.40.Mr, 77.22.Ej DOI: 10.7498/aps.71.20220303

^{*} Project supported by the National Science Fund for Distinguished Young Scholars of China (Grant No. T2125005), the Tianjin Science Foundation for Distinguished Young Scholars, China (Grant No. 19JCJQJC61000), and the Shenzhen Science and Technology Project, China (Grant No. JCYJ20210324121002008).

[†] Corresponding author. E-mail: wentao@nankai.edu.cn

物理学报Acta Physica Sinica





Institute of Physics, CAS

面向神经形态感知和计算的柔性忆阻器基脉冲神经元 朱佳雪 张续猛 王睿 刘琦

Flexible memristive spiking neuron for neuromorphic sensing and computing Zhu Jia-Xue Zhang Xu-Meng Wang Rui Liu Qi 引用信息 Citation: Acta Physica Sinica, 71, 148503 (2022) DOI: 10.7498/aps.71.20212323 在线阅读 View online: https://doi.org/10.7498/aps.71.20212323 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

NbO_忆阻神经元的设计及其在尖峰神经网络中的应用

 Design of $\mathrm{NbO}_{\mathrm{v}}$ memristive neuron and its application in spiking neural networks

物理学报. 2022, 71(11): 110501 https://doi.org/10.7498/aps.71.20220141

N型局部有源忆阻器的神经形态行为

Neuromorphic behaviors of N-type locally-active memristor 物理学报. 2022, 71(5): 050502 https://doi.org/10.7498/aps.71.20212017

基于忆容器件的神经形态计算研究进展

Research progress of neuromorphic computation based on memcapacitors 物理学报. 2021, 70(7): 078701 https://doi.org/10.7498/aps.70.20201632

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用 Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

氧化物基忆阻型神经突触器件 Oxide-based memristive neuromorphic synaptic devices

物理学报. 2019, 68(16): 168504 https://doi.org/10.7498/aps.68.20191262

铝基薄膜忆阻器作为感觉神经系统的习惯化特性

Al-based memristor applied to habituation sensory nervous system 物理学报. 2021, 70(6): 068502 https://doi.org/10.7498/aps.70.20201961 专题: 面向类脑计算的物理电子学

面向神经形态感知和计算的柔性 忆阻器基脉冲神经元*

朱佳雪1)3) 张续猛2)4) 王睿1)3) 刘琦2)4)†

(中国科学院微电子研究所,微电子器件与集成技术重点实验室,北京 100029)
 2)(复旦大学,复旦大学芯片与系统前沿技术研究院,上海 200433)
 3)(中国科学院大学,北京 100049)
 4)(上海期智研究院,上海 200232)

(2021年12月16日收到; 2022年1月10日收到修改稿)

受人脑工作模式的启发,脉冲神经元作为人工感知系统和神经形态计算体系的基本计算单元发挥着重要作用.然而,基于传统互补金属氧化物半导体技术的神经元电路结构复杂,功耗高,且缺乏柔韧性,不利于 大规模集成和与人体兼容的柔性感知系统的应用.本文制备的柔性忆阻器展示出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 10⁴ 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-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.此外,基于电子传输机制和构建的核壳模型,对柔性忆阻器的工作机制 进行分析,提出了电场和热激发主导的阈值转变机制;进一步对忆阻器和神经元的电学特性进行电路仿真模 拟,验证了柔性忆阻器和神经元电路工作机制的合理性.本文对柔性神经元的研究可为神经形态感知和计算 系统的构建提供硬件基础和理论指导.

关键词:柔性忆阻器,脉冲神经元,神经形态感知和计算 PACS: 85.35.-p, 73.43.Cb, 73.63.-b

DOI: 10.7498/aps.71.20212323

1 引 言

近年来,人工智能、物联网、智能穿戴等领域 的发展,对海量数据的快速高效处理提出了更高的 要求.基于冯·诺依曼架构的传统计算系统,由于存 储器和处理器的物理分离导致了"存储器"瓶颈^[1], 难以实现智能系统的高效数理处理.受人脑工作模 式的启发,以脉冲神经元为计算单元的神经形态计 算架构体系应运而生,它具有并行信息处理、低功 耗、高容错性等优点^[2],使其在处理高维和非结构 化数据方面显示出巨大优势,为智能系统的设计提 供了备选方案.然而,当前神经形态芯片^[3-7]所使 用的神经元大多基于互补金属氧化物半导体 (complementary metal oxide semiconductor, CMOS) 技术,由于晶体管缺乏内在动力学特性和生物单元 相似性,其所构建的神经元电路不但结构复杂,集 成度低,而且学习能力欠缺^[8].因此,亟需寻找结构 简单、集成度高、内在动力学丰富的神经形态器件 构建紧凑的神经元电路,从而实现更高效的神经

© 2022 中国物理学会 Chinese Physical Society

^{*} 国家自然科学基金 (批准号: 61825404, 61732020, 61834009, 61821091, 61804167, 61851402, 62104044)、国家重大科技专项 (批 准号: 2017ZX02301007-001)、中国博士后科学基金 (批准号: 2020M681167) 和中国科学院战略重点研究发展计划基金 (批准号: XDB44000000) 资助的课题.

[†] 通信作者. E-mail: qi_liu@fudan.edu.cn

形态系统.

在众多的神经形态器件 (如阻变存储器⁹⁾、相 变存储器^[10]、磁自旋存储器^[11]、铁电存储器^[12]、突 触离子晶体管[13])中,阻变存储器(后续称为忆阻 器)因其结构简单,转变速度快,功耗低,动力学 特性丰富等优点[9,14],在存内计算[15]、突触功能模 拟^[16]和人工感知系统^[8,17,18]等领域备受研究人员 关注. 此外, 忆阻器所展现出的阈值转变特性还被 用于神经元电路的设计中,并在人工感知系统和神 经形态计算等领域获得应用^[19-22].例如, Zhang 等^[23] 利用 NbO_x 基忆阻器设计了漏电-积分-发射 (leaky integrate-and-fire, LIF) 神经元, 并实现了人工传 入神经的模拟. Lashkare 等^[24] 基于 PrMnO₃ 忆阻 器设计的神经元具备整流线性单元 (rectified linear unit, ReLU) 功能, 可用于脉冲神经网络的激 活函数模拟. 然而, 目前基于忆阻器所设计的神经 元均缺乏柔韧性,不利于智能皮肤、人工义肢、健 康监测等柔性感知方面的应用.此外,忆阻器基脉 冲神经元在神经形态计算的非线性变换功能领域 也有着重要应用^[20,23-28].因此,设计和开发一种基 于柔性忆阻器的脉冲神经元用于智能感知和计算 系统,对忆阻器基神经元的功能挖掘和应用推广意 义重大.

本研究组制备了一种柔性的阈值转变型忆阻器,该器件表现出了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 10⁴次.基于此器 件设计的柔性神经元电路具备阈值发放、"全"或 "无"、不应期和输入强度依赖的频率调制特性,可 将模拟的输入信号转换成脉冲频率信号,用于人工 感知系统的应用.此外,该神经元表现出的频率-电 压强度关系具有 ReLU 激活函数相似性,可用作神 经形态计算的非线性变换单元.针对柔性神经元表 现出的频率和阈值电压随输入信号强度的变化关 系, 对柔性忆阻器的电子传输机制和阈值转变机制 进行了分析, 并通过电路仿真技术进行了验证, 得 到了电场和焦耳热主导的阈值转变机制, 揭示了柔 性衬底的热积累效应对器件阈值电压和神经元放 电频率的影响. 最后, 基于构建的柔性神经元电路 搭建了两层神经网络 (784 × 100 × 10) 系统来识 别数字手写体, 得到了约 95.6% 准确率, 与理想情 况相当. 本研究为柔性神经元的设计和工作机制分 析提供了参考, 阐述了其在柔性感知和神经形态计 算中的应用潜力.

2 实 验

柔性 NbO_x 忆阻器的制备流程如图 1 所示. 首 先, 通过旋涂法在清洗后的 Si 片上涂布聚酰亚胺 (polyimide, PI) 溶胶, 之后将样品置于鼓风干燥箱 中加热干燥, 直至形成表面光滑且厚度均一的 PI 薄膜 (厚度约为 30 μm). 为保证后续薄膜的沉 积质量, 在 PI 薄膜表面上通过磁控溅射法生长 100 nm 厚的 SiO₂ 薄膜来获得更加光滑和致密的 表面. 随后, 利用紫外光刻技术和剥离法对后续薄 膜进行图案化, 结合磁控溅射法和电子束蒸发法 依次生长 5/35 nm 的 Ti/Pt 底电极、50 nm 的 NbO_x 阻变层和 5/35 nm 的 Ti/Pt 顶电极. 其中, 器件的 有效面积为 5 μm×5 μm. 最后将 PI 薄膜从 Si 衬 底上剥离, 于是, 柔性 PI/SiO₂/Ti/Pt/NbO_x/Ti/Pt 结构的忆阻器制备完成.

3 实验结果与讨论

3.1 柔性忆阻器基本电学特性

柔性忆阻器的结构如图 2(a) 所示,其中 PI 薄 膜为柔性衬底,在所有的电学测试中底电极 Pt 接



图 1 柔性 PI/SiO₂/Ti/Pt/NbO_x/Ti/Pt 结构忆阻器制备流程图

Fig. 1. Flow chart of the flexible $\rm PI/SiO_2/Ti/Pt/NbO_{\it x}/Ti/Pt$ structured memristor device.



图 2 柔性忆阻器基本电学特性 (a) 柔性 PI/SiO₂/Ti/Pt/NbO_x/Ti/Pt 忆阻器结构示意图; (b) 忆阻器基本 *I-V*曲线; (c) 50 次 电学循环下阈值电压和保持电压的累积分布函数; (d) 器件差异性表征

Fig. 2. Basic electrical characteristics of flexible memristor: (a) Structure diagram of $PI/SiO_2/Ti/Pt/NbO_x/Ti/Pt$ memristor device; (b) basic current-voltage (*I-V*) curve of memristor; (c) cumulative distribution function of threshold voltage and hold voltage under 50 cycles; (d) variation of device to device.

地, 而在顶电极 Ti/Pt 上施加电压. 器件经电激活 过程后呈现出如图 2(b) 所示的阈值转变行为,即 在正向的扫描电压下, 流过器件的电流随着电压的 增加而增大,直至电压达到阈值电压 (the threshold voltage, V_{TH}), 电流突然增加至限制电流 (compliance current, I_{CC}) 500 µA, 器件从高电阻状态 (the high resistance state, R_{OFF})转变为低电阻状 态 (the low resistance state, R_{ON}); 随着反向扫描 电压降低至保持电压 (the hold voltage, $V_{\rm H}$), 电流 突然降低,器件从低电阻状态转变为高电阻状态转 变,完成器件的关闭.从图 2(b) 中可以看出,在 50 次 电学操作下,器件的稳定性良好,从中提取 V_{TH} 和 V_H, 二者的累积分布函数如图 2(c) 所示, 分布 均一且离散性小,说明器件具有稳定的阈值转变行 为. 此外, 对不同器件间 V_{TH}和 V_H的差异性进行 了表征, 如图 2(d) 所示, 在测试的 10 个器件中, V_{TH} 和 V_H虽有一定的离散性, 但二者的分布范围无重 叠, 且器件的阈值转变窗口 $(V_{\text{TH}} - V_{\text{H}})$ 均大于 0.2 V, 可防止电压误读,具有较强的抗干扰能力,可用于 后续神经元电路的构建.

3.2 柔性忆阻器弯折能力表征

为表征柔性忆阻器的机械弯折能力,在不同弯 折半径和弯折次数下对器件的性能进行了测试.如 图 3(a)所示,将柔性忆阻器缠绕在不同半径的玻 璃棒表面,并将其固定在载玻片上进行电学测试, 获得了如图 3(b)所示的 *I-V*曲线.从图 3(b)中可 以看出,器件在不同弯折半径 (3.5—1.5 mm)下均 表现出阈值转变现象.在每个测试条件下,提取 50次电循环中的 *V*_{TH}和 *V*_H并作误差统计分布,如 图 3(c)所示.从图中可以看出,*V*_{TH}和 *V*_H在不同 弯折半径下的标准偏差 (σ/μ)分别为 2.3% 和 1.3%, 虽有一定的波动性,但二者并未有重叠,说明器件 在弯折条件下具有稳定性,表明了器件具有良好的 可弯折能力.

其次,对器件在弯折条件下的抗疲劳特性进行 了表征.如图 4(a) 所示,将柔性器件粘贴在丝杆滑 台组件的滑台上,滑台由步进电机带动做往复运 动,滑台距离固定端的最大距离可使器件展平,而 最小距离使得器件的弯折半径为 2.5 mm. 基于此 装置,对不同弯折次数下的器件性能进行了测试, 如图 4(b)所示,器件在不同弯折次数 (1—10⁴)下 均展现了稳定的阈值转变特性,且在往复弯折 10⁴ 次后,器件的 V_{TH} 和 V_H基本保持不变 (图 4(c)),表 明制备的柔性忆阻器具有良好的弯折耐受性.

3.3 柔性忆阻器基脉冲神经元

基于柔性忆阻器及其阈值转变特性,构建了如图 5(a)所示的忆阻器基脉冲神经元电路 (memristive spiking neuron, MSN). 该电路由一个负载 电阻 (the load resistance, R_L),一个柔性忆阻器,一 个串联电阻 (the series resistance, R_S)和一个电容器 (capacitor, *C*)组成. 其中电容器可为忆阻器的



图 3 不同弯折半径下器件的电学特性 (a) 不同弯折半径的测试图片; (b) 不同弯折半径下的 *I-V*曲线; (c) 不同弯折半径下 50 次电学循环的阈值电压和保持电压统计

Fig. 3. Electrical characteristics of devices at different bending radii: (a) Test image of different bending radii; (b) I-V curves at different bending radii; (c) $V_{\rm TH}$ and $V_{\rm H}$ statistics for 50 cycles at different bending radii.



图 4 不同弯折次数下器件的电学特性 (a) 不同弯折次数的测试图片; (b) 不同弯折次数下的 *I-V*曲线; (c) 不同弯折次数下 50 次电学循环的阈值电压和保持电压统计

Fig. 4. Electrical characteristics of devices after different cycles of bending: (a) Test image of different cycles of bending; (b) I-V curves after different cycles of bending; (c) $V_{\rm TH}$ and $V_{\rm H}$ statistics of 50 cycles after different cycles of bending.



图 5 柔性忆阻器脉冲神经元的关键特征 (a) 基于柔性忆阻器的脉冲神经元电路原理图; (b) 忆阻器脉冲神经元的振荡特性和脉冲输出特性; (c) 忆阻器脉冲神经元的全或无特性; (d) 忆阻器脉冲神经元的不应期特性

Fig. 5. Key features of flexible memristive spiking neuron: (a) Schematic diagram of spiking neuron circuit based on flexible memristor; (b) oscillation and output characteristics of memristive spiking neuron; (c) all or nothing characteristic of memristive spiking neuron; (d) refractory period characteristic of memristive spiking neuron.

内部电容、电路的寄生电容或外部接入的电容.值 得注意的是,这里为了定量表示电容值,以便后续 仿真,采用了外接的电容器,而在实际应用中,可 根据具体的需求,去掉外接的电容器,利用忆阻器 本身的寄生电容来维持神经元的工作,从而增加集 成度. 在本节测试中, $R_{\rm L}$, $R_{\rm S}$ 和 C的值分别为 30 kΩ, 100 Ω 和 1 nF, 而忆阻器的 R_{OFF} 和 R_{ON} 分别约为 200 kΩ和1 kΩ. 当对电路进行供电时,由于忆阻 器处于高阻态,此时"1-2-4-1"环路的 RC时间常数 $\tau_1 = R_L \times C$ 小于"2-3-4-2"环路的 RC 时间常数 $\tau'_1 =$ $(R_{OFF}+R_S) \times C$,所以先对电容器进行充电,直至节 点"2"处的电压,即电容电位达到忆阻器的阈值电 压, 忆阻器由 R_{OFF} 转变为 R_{ON}; 此时环路"2-3-4-2"处的 RC时间常数 $\tau_2 = (R_{ON} + R_S) \times C$ 小于 τ_1 , 电容器进行放电,直至节点"2"处的电压达到忆阻 器的保持电压, 忆阻器由 R_{ON}转变为 R_{OFF}, 电容 器进行新一轮的充放电. 如图 5(b) 所示, 当电路的 输入电压 (the input voltage, V_{IN}) 为 4 V 的固定 电压时,稳定工作状态下节点"2"处的电位(记为

神经元的输出电压)在 V_{TH}和 V_H之间进行振荡, 而从节点"3"处采集到的电流信号则为脉冲形式, 为神经元的脉冲输出. 当输入信号为图 5(c) 所示 幅值递增的脉冲信号时,只有电容电位达到 V_{TH} (约 2.0 V)时,才会有脉冲信号输出,表现出阈值 发放特性; 而当电容电位低于 V_{TH} 时, 则无脉冲输 出,体现了神经元的"全"或"无"特性.此外,神经 元在放电期间,表现出不应期特性,如图 5(d)所 示,当神经元开始发放脉冲时,随后施加在输入端 的脉冲并不能引起神经元的再次放电,直至神经元 完成一次放电行为,此段时间间隔称作不应期,与生 物中的不应期概念相对应. 这些结果表明本文所构 建的柔性忆阻器基脉冲神经元电路具备了生物神 经元的多种关键放电特征,能够把输入的模拟信号 转换为脉冲信号,可用作人工感知系统中的感觉神 经元, 而器件本身所具有的柔性使其特别适用于人 工义肢、神经界面和类人机器人等柔性感知应用.

另一方面,基于柔性忆阻器的神经元电路表现 出线性的输出频率-输入电压依赖关系,使其可用于

人工神经网络中 ReLU 激活函数的模拟. 如图 6(a) 所示,当神经元有正向输入时,ReLU神经元的输 出线性依赖于输入(图中斜线部分),而对于负向输 入, ReLU 神经元则无输出, 此特性可通过柔性忆 阻器神经元进行模拟.图 6(b)展示了神经元电路 在阶梯型的脉冲电压输入下神经元的振荡输出结 果. 对输入的电压脉冲的频率和幅值进行统计, 结 果见图 6(c), (d). 从图 6(c) 中可以看出, 当输入电 压大于等于 4 V 时, 神经元的输出频率随着输入电 压的增大而线性增加;而当输入电压小于4V时, 神经元无脉冲输出. 根据获得的实验数据进行线性 拟合,从图中可以看出,拟合后的曲线与 ReLU 神 经元的输出-输入关系具有相似性,经过归一化处 理,本文神经元的输出频率 (frequency, F) 和 $V_{\rm IN}$ 的关系可模拟 ReLU 神经元, 用于基于转化法的脉 冲神经网络的推理过程.此外,从图 6(b)的振荡电 压输出曲线中可以获得忆阻器的 V_{TH} 和 V_{H} 信息, 即 V_{TH} 对应于输出曲线的波峰,而 V_H 对应于输出 曲线的波谷,二者的统计结果如图 6(d) 所示.可以 看出,随着 V_{IN} 的增加, V_{TH} 有下降趋势,而 V_H 基 本保持不变.但直流测试发现,V_{TH} 和 V_H 为忆阻 器的内在特征,并不受扫描电压的影响.针对柔性 忆阻器神经元电路的 F-V_{IN} 线性依赖关系和 V_{TH} 在脉冲操作模式随着 V_{IN} 的增加而下降这一现象, 本文从柔性忆阻器的工作机制入手做进一步地分析.

4 机制分析与模型建立

4.1 电子传输机制

为解释柔性神经元的频率 *F* 和器件 *V*_{TH} 随输入 *V*_{IN} 变化的关系,首先对柔性忆阻器的电子传输机制进行分析.如图 7(a) 所示,将器件发生阈值转变之前的 *I-V* 曲线在双对数坐标下进行线性拟合可以发现,当扫描电压小于等于 0.2 V 时, *I-V* 曲



图 6 柔性忆阻器脉冲神经元在不同输入电压强度下的频率调制特性 (a) 整流线性单元对应的神经元输入输出关系; (b) 柔性 忆阻器脉冲神经元在不同输入电压下的脉冲输出特性, 内插图为虚框内的脉冲输出放大图; (c) 不同输入电压下的输出频率统计 及线性拟合; (d) V_{TH}和 V_H在不同输入电压下的统计

Fig. 6. Frequency regulation characteristics of flexible spiking neuron under different input voltage intensities: (a) Input and output relationship of neuron corresponding to rectified linear unit; (b) output characteristics of the flexible memristive spiking neuron under different input voltages, and the inset is the zoom in details of the output curves in the dashed windows; (c) output frequency statistics and linear fitting under different input voltages; (d) $V_{\rm TH}$ and $V_{\rm H}$ statistics at different input voltages.



图 7 柔性忆阻器的电子传输机制分析 (a) 阈值转变前 *I-V*曲线在双对数坐标下的线性拟合; (b) 阈值转变前 $\ln(I/E)$ 和 $E^{1/2}$ 的 线性拟合; (c) 欧姆传输机制下 $\ln(I)$ 和 1000/*T* 的线性拟合; (d) Poole-Frenkel 传输机制下 $\ln(I)$ 和 1000/*T* 的线性拟合 Fig. 7. Analysis of electron transport mechanism of flexible memristor: (a) Linear fitting of *I-V* curve in logarithmic coordinates before threshold switching; (b) linear fitting of $\ln(I/E)$ and $E^{1/2}$ in logarithmic coordinates before threshold switching; (c) linear fitting of $\ln(I)$ and 1000/T under ohmic conduction; (d) linear fitting of $\ln(I)$ and 1000/T under Poole-Frenkel transport mechanism.

线拟合的斜率约为 1.05, 说明在此段电压范围内, 电子传输符合欧姆传导机制^[29]. 当电压大于 0.2 V, 即对应的电场强度 (electric field intensity, E)大 于 4 MV/m 时, 如图 7(b)所示, $\ln(I/E)$ 对 $E^{1/2}$ 具 有线性依赖关系,表明在此电压范围内,电子传输 由 Poole-Frenkel(PF)传导机制主导^[30]. 基于以上 分析,对两种传输机制下的电子激活能进行了计 算,分别如图 7(c), (d) 所示. 在变温测试中,分别 提取了 0.1 V 和 0.6 V 所对应的电流数据,并根据 Arrhenius 公式对 $\ln(I)$ 和 1000/T 的关系进行了 线性拟合,从中得到欧姆传输机制和 PF 传输机制 下的电子激活能 E_{ae} 和 E_a 分别为 (0.2237±0.02) 和 (0.2251±0.05) eV,其中±0.02 和 ±0.05 eV 的 误差由器件的差异和读取电压选取的不同引起,可 为后续仿真模型的建立提供参考.

4.2 核壳模型与阈值转变机制

根据已有研究报道^[31-33]可知,NbO_x基忆阻器在电压模式下进行电激活会形成细丝通道,随后

在此区域内完成阈值转变过程.基于此,本文构建 了如图 8 所示核壳模型并结合电子传输机制对 器件的工作机制进行了分析.在该模型中,未经任 何电学操作的 NbO_x薄膜内部电学性质均一,由 Nb₂O_{5-x}构成,如图 8(a)所示;而经过电激活操作 后 NbO_x薄膜由 Nb₂O_{5-x}壳层区域和相对导电的 NbO₂ 细丝区域构成,如图 8(b)所示.根据电子传 输机制分析,由电场和热激发主导的 PF 机制可在 高电场和焦耳热的辅助下产生电流突增,因此本研 究组认为该机制主导了细丝区域的电子传输并结 合牛顿冷却定律^[34]引起了器件的阈值转变行为, 这与以往的报道结果一致^[30,35].其中,由 PF 机制 主导的细丝区域的电阻可简记为 *R*_{PF},其公式为

$$R_{\rm PF} = R_0 \exp\left[\frac{1}{k_{\rm B}T} \left(E_{\rm a} - q\sqrt{\frac{qE}{\pi \,\varepsilon_0 \varepsilon_{\rm r}}}\right)\right], \quad (1)$$

其中, R_0 为前置因子, E_a 为电子激活能, E 为电场 强度, T 为温度, k_B 为玻尔兹曼常数, ε_0 和 ε_r 分别 为真空介电常数和相对介电常数. 由牛顿冷却定律 主导的热传导公式为

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{R_{\mathrm{PF}}i_{\mathrm{PF}}^2 - \Gamma_{\mathrm{th}}\Delta T}{C_{\mathrm{th}}}, \qquad (2)$$

其中, $i_{\rm PF}$ 为流过细丝区域的电流; $\Gamma_{\rm th}$ 为有效热导率,其倒数为有效热阻 $R_{\rm th}$; $C_{\rm th}$ 为热容; $\Delta T = T - T_{\rm amb}$,为细丝区域与环境温度 $T_{\rm amb}$ 之间差值.另外,考虑到壳层区域面积相对较大,且在电场作用下也会产生漏电流,因此本研究组认为在低电场下



图 8 核壳模型 (a) 电激活操作前的器件结构示意图; (b) 电激活操作后的器件结构示意图, 其中 NbO_x 介质层 由 NbO₂ 细丝区域和 Nb₂O_{5-x}壳层区域构成

Fig. 8. Core-Shell mode: (a) Schematic diagram of the memristor before electroforming; (b) schematic diagram of the memristor after electroforming, in which the NbO_x dielectric layer consists of NbO₂filamentary region and Nb₂O_{5-x} shell region.



的漏电流主要来自于壳层电阻的贡献,并由欧姆机 制主导: $J \sim E \exp(-E_{ae}/(kT)^{[36]}, 其中 J 为壳层区域$ 的电流密度, <math>E 为壳层区域电场, E_{ae} 为电子激活 能, T 为壳层区域温度. 该壳层区域的电阻可记作 R_{Ohm} , 其表达式为

$$R_{\rm Ohm} = R_1 \exp\left(\frac{E_{\rm ae}}{k_{\rm B}T}\right). \tag{3}$$

其中, *R*₁为前置因子, 可由低电压下器件的电阻-温度曲线确定.

4.3 SPICE 电路仿真

为了验证提出机制的合理性,基于以上分析进 行了电路仿真 (simulation program with integrated circuit emphasis, SPICE). 仿真电路的组成如图 9 所示,其中图 9(a) 为细丝区域电阻 (表示为 Rpf) 及其热传导的仿真原理图^[37],图 9(b) 为忆阻器的 工作原理图,其中顶、底电极的电阻分别表示为 Rte(~350 Ω)和 Rbe(~350 Ω),壳层区域的电阻表 示为 Rohm, 拟合参数见表 1. 柔性忆阻器在直流 电压扫描下的仿真结果如图 10(a) 所示,从图中可 以看出,仿真曲线与实验曲线具有良好的一致性, 说明构建的 SPICE 模型具有合理性.

$$\begin{split} &. FUNC \ joule() = Rpf()*i(Rpf)**2 \\ &. FUNC \ Rpf() = R0*exp(q*(Ea-delta())/(k*v(T))) \\ &. FUNC \ delta() = pow(q*E()/(pi*e0*er), 0.5) \\ &. FUNC \ E() = dv()/d \\ &. FUNC \ dv() = v(1) - v(2) \end{split}$$

. FUNC Rohm() = R1*exp(q*Eae/(k*Tamb))

图 9 基于 NbO_x 忆阻器的 SPICE 仿真原理图 (a) 细丝区域的电阻和热传导仿真原理图; (b) 忆阻器的 SPICE 仿真模型 Fig. 9. Schematic diagram of SPICE simulation based on NbO_x memristor: (a) Resistor and heat conduction simulation of filament region; (b) SPICE simulation model of memristor.

		Table 1. Parameter	list for SPICE simulation	on.		
热容	有效热阻	前置	电阻	电子激活能		
$C_{ m th}/({ m J}{\cdot}{ m K}^{-1})$	$R_{ m th}/({ m K}{\cdot}{ m W}^{-1})$	R_0/Ω	R_1/Ω	$E_{\rm ae}/{\rm eV}$	$E_{\rm a}/{\rm eV}$	
5×10^{-15}	$1.39{ imes}10^6$	120	30	0.2237 ± 0.02	0.2251 ± 0.05	
室温	真空电荷	真空介电常数	相对介电常数	玻尔兹曼常数	薄膜厚度	
$T_{\rm amb}/{ m K}$	q/C	$e_0/({ m F}{\cdot}{ m m}^{-1})$	$e_{ m r}$	$k/({ m J}{\cdot}{ m K}^{-1})$	$d/{ m m}$	
298	$1.6{ imes}10^{-19}$	$8.85 imes 10^{-12}$	45	$1.38{ imes}10^{-23}$	5×10^{-8}	

表 1 用于 SPICE 仿真的参数列表 Table 1 Parameter list for SPICE simulati



图 10 柔性忆阻器及神经元的仿真结果 (a) 忆阻器在直流电压扫描下的 *I-V*; (b) 忆阻器脉冲神经元在阶梯电压下的输入输出 曲线; (c) 神经元在不同输入电压下的输出频率统计; (d) 神经元在不同输入电压下阈值电压和保持电压统计 Fig. 10. Simulation results of flexible memristor and neuron circuit: (a) *I-V* curve of memristor under DC voltage sweep; (b) input and output curve of memristive spiking neuron at stepped voltage pulses; (c) output frequency statistics of neuron under different

input voltages; (d) threshold voltage and hold voltage statistics of neuron under different input voltages.

基于忆阻器的 SPICE 模型和图 5(a) 所示的 神经元电路,对图6所示的神经元电路在不同输入 电压 (V_{IN}) 下的输出特性进行了仿真. 根据实验结 果 (图 6(d)) 可知, V_{TH} 随着 V_{IN} 的增加有下降的 趋势,这一变化与以往报道中NbO_x忆阻器 V_{TH}随 着温度的变化关系类似[28]. 在本工作中, 考虑到细 丝区域工作过程中产生的焦耳热对壳层的贡献以 及柔性衬底的散热性能的影响,引入了温度修正 项: $T_{\rm c} = 3.3 \times V_{\rm IN}$ -13.2 对壳层温度进行了修正. 修正公式由实验结果拟合得出,修正后壳层的温度 为 T+T_c. 引入温度修正项后对神经元的输出特性 进行仿真,得到了如图 10(b) 所示的振荡输出曲 线, 其中 R_L和 C的值分别为 30 kΩ和 1.15 nF(考 虑了实际电容值有测试偏差),该仿真结果与实验 具有高度的一致性, 对其输出频率 (F)、波峰 (V_{TH}) 和波谷 (V_H) 值进行统计, 得到了如图 10(c), (d) 所示的统计图. 从图 10(c) 中可以看出, 随着 $V_{\rm IN}$ 的升高,神经元的输出频率 F线性增大,与实验结

果一致.此外,从图 10(d) 中可以看出,V_{TH} 随着温度的升高逐渐下降,而 V_H 基本不变,这一变化规律也与实验数据匹配.由此说明,基于核壳模型、PF 传导、欧姆传导以及热传导公式建立的 SPICE 模型具有合理性,可对柔性忆阻器和基于此构建的神经元的电学行为进行解释,为神经元的应用和硬件设计提供指导.

5 基于柔性神经元的脉冲神经网络

与已报道的忆阻器脉冲神经元^[24,26,28]相对比, 本文工作中基于柔性忆阻器构建的神经元电路具 有更好的 *F-V*_{IN} 线性度,能够更好地模拟神经网络 中的 ReLU 函数,以实现基于人工神经网络 (artificial neural network, ANN)转换的 SNN 系统.为 验证其应用能力,构建了如图 11(a) 所示的转换脉 冲神经网络来执行数字手写体识别任务.该网络共 三层,分别为包含 784 个神经元的输入层、包含 100 个神经元的隐含层和包含 10 个神经元的输出层.



图 11 基于忆阻器脉冲神经元的脉冲神经网络仿真 (a) 用于 MNIST 手写体数据集识别的脉冲神经网络原理图; (b) 基于柔性 忆阻器脉冲神经元的识别结果

Fig. 11. Simulation of spiking neural network based on FMSN: (a) Schematic of spiking neural network based on FMSN for MNIST handwritten digit classification; (b) classification result of FMSN.

训练过程采用 ANN 中的反向传播算法 (backpropagation, BP), 隐含层和输出层分别使用 ReLU 和 Sigmoid 激活函数. 在推理过程中则采用 SNN 模 式,输入为脉冲频率信号,所有激活函数用所构建 的柔性忆阻器基脉冲神经元 (flexible memristive spiking neuron, FMSN) 实现. 在本工作中 FMSN 利用了电容器的充放电过程和忆阻器的阈值转变 特性来产生持续振荡,如果输入为脉冲形式,那么 输入频率越大,充电过程越快,输出频率越高.类 似地,对于固定的输入频率,脉冲幅值越大,输出 频率越高.因此在 SNN 中, 不仅可以用脉冲脉率 对输入编码,也可以通过单脉冲幅度或时间关系编 码, 而基于频率和幅度编码的 SNN 均可以利用本 文的神经元来实现. 图 11(b) 展示了最终的识别结 果,利用FMSN 所构建的 SNN 可以实现 95.6% 的 识别率,与理想情况下的 ANN 结果相当 (96%), 表明 FMSN用于神经形态计算中线性整流单元的 可行性,有利于实现高性能的转换 SNN 系统.

6 结 论

本文制备了一种 NbO_x基柔性忆阻器,该器件 具备稳定的阈值转变特性和良好的弯折特性,其中 最小的弯折半径可达 1.5 mm,弯折次数可达 10⁴ 次.基于柔性忆阻器,构建了脉冲神经元电路,该 电路具备阈值发放、"全"或"无"和不应期等神经元 放电特征,能够将模拟的输入信号转化为脉冲输 出,并应用于人工感受系统的传入神经元.为验证 该器件特性和电路行为的合理性,构建了由 PooleFrenkel传输机制和欧姆传输机制主导的核壳模型, 用来解释忆阻器的工作机制,并通过 SPICE 电路 仿真对器件机制和电路放电特性进行了验证.此 外,该神经元的频率-电压调制特性具有 ReLU 函 数相似性,可用来模拟神经形态系统中的 ReLU 激 活函数,实现非线性转换功能.最后,基于此神经 元搭建了基于转换法的 SNN 网络,在 MNIST 手 写体数据集上实现了 95.6%的识别率,与理想 ANN 识别率相当.本文为柔性脉冲神经元的设计提供了 参考,为其在柔性智能感知系统及神经形态计算方 面的应用奠定了基础.

参考文献

- Wang Z, Wu H, Burr G W, Hwang C S, Wang K L, Xia Q, Yang J J 2020 Nat. Rev. Mater. 5 173
- [2] Zhou F, Chai Y 2020 Nat. Electron. 3 664
- [3] Pei J, Deng L, Song S, Zhao M, Zhang Y, Wu S, Wang G, Zou Z, Wu Z, He W 2019 Nature 572 106
- [4] Merolla P A, Arthur J V, Alvarez-Icaza R, Cassidy A S, Sawada J, Akopyan F, Jackson B L, Imam N, Guo C, Nakamura Y 2014 Science 345 668
- [5] Furber S B, Galluppi F, Temple S, Plana L A 2014 Proceedings of the IEEE 102 652
- [6] Benjamin B V, Gao P, McQuinn E, Choudhary S, Chandrasekaran A R, Bussat J M, Alvarez-Icaza R, Arthur J V, Merolla P A, Boahen K 2014 *Proceedings of the IEEE* 102 699
- [7] Davies M, Srinivasa N, Lin T H, Chinya G, Cao Y, Choday S H, Dimou G, Joshi P, Imam N, Jain S 2018 *Ieee Micro* 38 82
- [8] Ji X, Zhao X, Tan M C, Zhao R 2020 Advanced Intelligent Systems 2 1900118
- [9] Pan F, Gao S, Chen C, Song C, Zeng F 2014 Mater. Sci. Eng. R-Rep. 83 1
- [10] Raoux S, Xiong F, Wuttig M, Pop E 2014 MRS Bull. 39 703
- [11] Doevenspeck J, Garello K, Verhoef B, Degraeve R, van Beek

S, Crotti D, Yasin F, Couet S, Jayakumar G, Papistas I 2020 2020 IEEE Symp. VLSI Technol. Honolulu, HI, USA, June 16–19, 2020 pp1–2

- [12] Vorotilov K A, Sigov A 2012 Phys. Solid State 54 894
- [13] Rivnay J, Inal S, Salleo A, Owens R M, Berggren M, Malliaras G G 2018 Nat. Rev. Mater. 3 1
- [14] Shi T, Wang R, Wu Z, Sun Y, An J, Liu Q 2021 Small Struct. 2 2000109
- [15] Ielmini D, Wong H S P 2018 Nat. Electron. 1 333
- [16] Yang R, Huang H M, Guo X 2019 Adv. Electron. Mater. 5 1900287
- [17] Wang M, Luo Y, Wang T, Wan C, Pan L, Pan S, He K, Neo A, Chen X 2021 Adv. Mater. 33 2003014
- [18] Jung Y H, Park B, Kim J U, Kim T i 2019 Adv. Mater. 31 1803637
- [19] Yi W, Tsang K K, Lam S K, Bai X, Crowell J A, Flores E A 2018 Nat. Commun. 9 4661
- [20] Stoliar P, Tranchant J, Corraze B, Janod E, Besland M P, Tesler F, Rozenberg M, Cario L 2017 Adv. Funct. Mater. 27 1604740
- [21] Zhang X, Wang W, Liu Q, Zhao X, Wei J, Cao R, Yao Z, Zhu X, Zhang F, Lü H 2017 IEEE Electron Device Lett. 39 308
- [22] Zhang X, Wu Z, Lu J, Wei J, Lu J, Zhu J, Qiu J, Wang R, Lou K, Wang Y 2020 2020 IEEE IEDM San Francisco, CA, USA, December 12–18, 2020 pp29.6.1–29.6.4
- [23] Zhang X, Zhuo Y, Luo Q, Wu Z, Midya R, Wang Z, Song W, Wang R, Upadhyay N K, Fang Y 2020 Nat. Commun. 11 1
- [24] Lashkare S, Bhat A, Kumbhare P, Ganguly U 2018 2018 NVMTS Sendai, Japan, October 22–24, 2018 pp1–4

- [25] Wu Q, Dang B, Lu C, Xu G, Yang G, Wang J, Chuai X, Lu N, Geng D, Wang H 2020 Nano Lett. 20 8015
- [26] Zhang X, Wang Z, Song W, Midya R, Zhuo Y, Wang R, Rao M, Upadhyay N K, Xia Q, Yang J J 2019 2019 IEEE IEDM San Francisco, CA, USA, December 7–11, 2019 pp6.7.1–6.7.4
- [27] Jerry M, Parihar A, Grisafe B, Raychowdhury A, Datta S 2017 2017 Symp. VLSI Technol. Kyoto, Japan, June 5–8, 2017 ppT186–T187
- [28] Wang P, Khan A I, Yu S 2020 Appl. Phys. Lett. 116 162108
- [29] Chiu F C 2014 Adv. Mater. Sci. Eng. 2014 578168
- [30] Slesazeck S, Mähne H, Wylezich H, Wachowiak A, Radhakrishnan J, Ascoli A, Tetzlaff R, Mikolajick T 2015 *RSC Adv.* 5 102318
- [31] Kumar S, Wang Z, Davila N, Kumari N, Norris K J, Huang X, Strachan J P, Vine D, Kilcoyne A D, Nishi Y 2017 Nat. Commun. 8 1
- [32] Nandi S K, Nath S K, El-Helou A E, Li S, Ratcliff T, Uenuma M, Raad P E, Elliman R G 2020 ACS Appl. Mater. Interfaces 12 8422
- [33] Nath S K, Nandi S K, Li S, Elliman R G 2019 Appl. Phys. Lett. 114 062901
- [34] Kumar S, Williams R S 2018 Nat. Commun. 9 2030
- [35] Nath S K, Nandi S K, El-Helou A, Liu X, Li S, Ratcliff T, Raad P E, Elliman R G 2020 Phys. Rev. Appl. 13 064024
- [36] Jung K, Kim Y, Im H, Kim H, Park B 2011 J. Korean Phys. Soc 59 2778
- [37] Radhakrishnan J, Slesazeck S, Wylezich H, Mikolajick T, Ascoli A, Tetzlaff R 2016 CNNA 2016: 15th International Workshop on Cellular Nanoscale Networks and their Applications Dresden, Germany, August 23–25, 2016 pp1–2

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Flexible memristive spiking neuron for neuromorphic sensing and computing^{*}

Zhu Jia-Xue¹⁾³⁾ Zhang Xu-Meng²⁾⁴⁾ Wang Rui¹⁾³⁾ Liu Qi^{2)4)†}

 (Key Laboratory of Microelectronics Device & Integrated Technology, Institute of Microelectronics of Chinese Academy of Sciences, Beijing 100029, China)

2) (Frontier Institute of Chip and System, Fudan University, Shanghai 200433, China)

3) (University of Chinese Academy of Sciences, Beijing 100049, China)

4) (Shanghai Qi Zhi Institute, Shanghai 200232, China)

(Received 16 December 2021; revised manuscript received 10 January 2022)

Abstract

Inspired by the working modes of the human brain, the spiking neuron plays an important role as the basic computing unit of artificial perception systems and neuromorphic computing systems. However, the neuron circuit based on complementary metal-oxide-semiconductor technology has a complex structure, high power consumption, and limited flexibility. These features are not conducive to the large-scale integration and the application of flexible sensing systems compatible with the human body. The flexible memristor prepared in this work shows stable threshold switching characteristics and excellent mechanical bending characteristics with bending radius up to 1.5 mm and bending times up to 10^4 . The compact neuron circuit based on this device shows the key features of the neuron, such as threshold-driven spiking, all-or-nothing, refractory period, and strength-modulated frequency response. The frequency-input voltage relationship of the neuron shows the similarity of the rectified linear unit, which can be used to simulate the function of rectified linear unit in spiking neural networks. In addition, based on the electron transport mechanism, a core-shell model is introduced to analyze the working mechanism of the flexible memristor and explain the output characteristics of the neuron. In this model, the shell region consisting of Nb_2O_{5-x} is subjected to ohmic conduction, while the core region consisting of NbO₂ is dominated by Poole-Frenkel conduction. These two mechanisms, combined with Newton's law of cooling, dominate the threshold switching behavior of flexible memristor device. Furthermore, the threshold switching characteristic of the memristor is simulated, verifying the rationality of the working mechanism of the flexible memristor. Considering the fact that the threshold voltage decreases with temperature increasing, a correction term is added to the temperature of the shell region. Subsequently, the output characteristics of the neuron regulated by the input voltage are simulated. The simulation results show that the frequency increases but the threshold voltage decreases with the input voltage increasing, which is consistent with the experimental result. The introduction of the correction term confirms the influence of the thermal accumulation effect of the flexible substrate on neuron output characteristics. Finally, we build a spiking neural network based on memristive spiking neurons to implement handwriting recognition, achieving a 95.6% recognition rate, which is comparable to the ideal result of the artificial neural network (96%). This result shows the potential application of the memristive spiking neurons in neuromorphic computing. In this paper, the study of flexible neurons can guide the design of neuromorphic sensing and computing systems.

Keywords: flexible memristor, spiking neuron, neuromorphic sensing and computing PACS: 85.35.-p, 73.43.Cb, 73.63.-b DOI: 10.7498/aps.71.20212323

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 61825404, 61732020, 61834009, 61821091, 61804167, 61851402, 62104044), the National Science and Technology Major Project of the Ministry of Science and Technology of China (Grant No. 2017ZX02301007-001), the China Postdoctoral Science Foundation (Grant No. 2020M681167), and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB44000000).

 $[\]dagger$ Corresponding author. E-mail: qi_liu@fudan.edu.cn





Institute of Physics, CAS

基于磁性隧道结的群体编码实现无监督聚类

张亚君 蔡佳林 乔亚 曾中明 袁喆 夏钶

Implementation of unsupervised clustering based on population coding of magnetic tunnel junctions Zhang Ya-Jun Cai Jia-Lin Qiao Ya Zeng Zhong-Ming Yuan Zhe Xia Ke 引用信息 Citation: Acta Physica Sinica, 71, 148506 (2022) DOI: 10.7498/aps.71.20220252 在线阅读 View online: https://doi.org/10.7498/aps.71.20220252 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

MgO基磁性隧道结温度-偏压相图的理论研究

Theoretical study on temperature-bias phase diagram of MgO-based magnetic tunnel junctions 物理学报. 2021, 70(10): 107302 https://doi.org/10.7498/aps.70.20201905

NbO、忆阻神经元的设计及其在尖峰神经网络中的应用

Design of NbO, memristive neuron and its application in spiking neural networks

物理学报. 2022, 71(11): 110501 https://doi.org/10.7498/aps.71.20220141

基于工艺偏差的电压调控磁各向异性磁隧道结电学模型及其在读写电路中的应用

Process deviation based electrical model of voltage controlled magnetic anisotropy magnetic tunnel junction and its application in read/write circuits

物理学报. 2020, 69(19): 198502 https://doi.org/10.7498/aps.69.20200228

基于工艺偏差的自旋转移矩辅助压控磁各向异性磁隧道结电学模型及其应用研究

Process deviation based electrical model of spin transfer torque assisted voltage controlled magnetic anisotropy magnetic tunnel junction and its application

物理学报. 2022, 71(10): 107501 https://doi.org/10.7498/aps.71.20211700

基于人工神经网络在线学习方法优化磁屏蔽特性参数

Online learning method based on artificial neural network to optimize magnetic shielding characteristic parameters 物理学报. 2019, 68(13): 130701 https://doi.org/10.7498/aps.68.20190234

基于深度神经网络的时空编码磁共振成像超分辨率重建方法

Super-resolved reconstruction method for spatiotemporally encoded magnetic resonance imaging based on deep neural network 物理学报. 2022, 71(5): 058702 https://doi.org/10.7498/aps.71.20211754

专题: 面向类脑计算的物理电子学

基于磁性隧道结的群体编码实现无监督聚类*

张亚君¹) 蔡佳林²) 乔亚¹) 曾中明²)[†] 袁喆¹)[‡] 夏钶³)

1) (北京师范大学物理学系,高等量子研究中心,北京 100875)

2) (中国科学院苏州纳米技术与纳米仿生研究所,苏州 215123)

3) (北京计算科学研究中心,北京 100193)

(2022年2月9日收到; 2022年4月13日收到修改稿)

利用新型材料器件发展类脑计算硬件研究的关键问题是发展出合适的算法,能够发挥新器件的特点和 优势.群体编码是生物神经系统常见的编码方式,能够有效去除噪音,实现短时程记忆及复杂的非线性映射 功能.本文选择自旋电子学器件中研究较多、工艺较成熟的磁性隧道结,应用其可调控的随机动力学实现群 体编码.作为一个应用的例子,超顺磁隧道结构建的二层脉冲神经网络成功完成了鸢尾花数据集的无监督聚 类.数值仿真表明基于磁性隧道结的群体编码可以有效对抗器件的非均一性,为类脑计算硬件研究提供重要 的参考.

关键词: 磁性隧道结, 群体编码, 脉冲神经网络, 无监督学习 PACS: 85.70.-w, 07.05.Mh

DOI: 10.7498/aps.71.20220252

1 引 言

得益于神经网络算法的进步和新型计算硬件 的发展,人工智能在近十多年取得了巨大的进步. 其中在图像识别、自动驾驶和自然语言处理等方面 的应用已经进入并逐步改变了人们的日常生活^[1]. 然而当前人工智能的发展仍然面临诸多问题.一方 面,随着摩尔定律的终结,处理器性能的增长开始 放缓;另一方面,传统冯·诺依曼架构采用的存算分 离系统,使数据在中央处理器和内存之间搬运消耗 了大量的时间和能量.为了解决这些问题,科学家 试图从大脑等生物神经系统寻求解决方案.大脑的 神经系统具有存算一体、高度并行、事件驱动等特 点^[2].目前已经有多款基于传统互补金属氧化物半 导体 (complementary metal oxide semiconductor, CMOS)工艺的类脑芯片问世^[3,4],展示了类脑芯片 在低能耗方面的巨大优势,但实现单个的脉冲神经 元和突触需要大量晶体管.通常类脑神经元之间所 需要的高连接度也限制了类脑芯片的规模和可扩 展性.因此,利用基于阻变材料、相变材料等新材 料器件实现类脑计算的基本单元成为研究的重点, 有望从根本上改进类脑芯片的计算架构和性能^[5–9]. 自旋电子学器件具有低能耗、高速和非线性磁动力 学及非易失性等特点,已经在硬盘、磁性随机存储 器等存储设备上实现了商业化应用.近年来,基于 自旋电子学的存内计算^[10]和类脑计算^[11–21]也获 得了广泛的关注.如何利用自旋电子学器件的天然 优势,结合新型类脑算法,实现软硬件的协同设计, 是基于自旋器件类脑计算的核心问题.

大脑中存在大量的神经元来执行日常的感知、 情感和记忆等高级功能.不同于计算机中精确的数 值表示,大脑中信息传递依赖于神经元之间的脉冲 发放,依靠神经元之间的连接形成的复杂网络执行

© 2022 中国物理学会 Chinese Physical Society

^{*} 国家自然科学基金 (批准号: 11734004, 12174028) 资助的课题.

[†] 通信作者. E-mail: zmzeng2012@sinano.ac.cn

[‡] 通信作者. E-mail: zyuan@bnu.edu.cn

相关的任务.尽管每个神经元细胞的信息传递和膜 电位变化过程中伴随大量噪音和随机动力学,但大 脑可以通过群体神经元共同编码信号来实现复杂 和精确的认知功能. 群体编码是生物系统在方向和 空间感知^[22]、颜色分辨、嗅觉识别和记忆^[23]等方 面的基础模式^[24],并具有很强的抗噪能力,对神经 细胞的异质性也具有很好的容忍度. 群体编码的优 势可以解决制造工艺的局限导致新材料类脑器件 的非均一性问题, 大幅降低新型类脑器件的制造要 求. 基于传统 CMOS 技术 [25]、相变器件 [26] 和自旋 器件[27]的群体编码系统都表现了对器件非均一性 的较强耐受能力,基于相变器件^[26]的群体编码利 用了脉冲发放的随机特性,其发放频率与输入信号 的脉宽呈线性关系: 而随机磁性隧道结的发放频率 与输入电压呈现单峰的非单调关系[27],更符合生 物系统的调谐曲线.

本文利用超顺磁磁性隧道结的随机翻转特性, 结合群体编码算法,提出一种基于自旋电子学器件 的脉冲神经网络方案.通过对鸢尾花数据集的无监 督分类研究,展示了该网络的鲁棒性和低能耗.首 先给出了在实验上观测到的磁性隧道结的随机翻 转特性及其电流调控,建立了基于超顺磁隧道结的 群体编码基础,同时验证了奈尔-布朗(Néel-Brown) 理论模型对器件的描述能力.研究了基于磁性隧道 结的群体编码,并结合延时脉冲发放和脉冲时序依 赖突触可塑性设计了基于自旋电子学器件的神经 网络.探究了磁性隧道结数目和器件参数对网络性 能的影响,并分析了该网络在能耗方面的优势.

2 利用超顺磁隧道结的群体编码

2.1 群体编码

在线性回归的数值计算中,我们的目标是由离 散的实验数据点得到直线的斜率和截距.由于实验 数据存在误差,我们期望通过更多的数据来降低数 据误差带来的影响.生物系统同样面临这样的问 题,单个神经元的信号噪声很大,为了获取更加精 准的感知信号,生物系统总是通过一组神经元的共 同响应来感知外界的刺激.这种利用一群神经元对 刺激的共同响应来编码信息的方式被称为群体编 码,其对于连续变量的编码尤为有效,例如颜色、 空间位置等.在群体编码中每个神经元对不同的外 界刺激输入信号具有一定的响应分布,称为该神经 元的调谐曲线;每个神经元对某个特定的外界输入 刺激会比较敏感,称为该神经元的偏好刺激.因此 通常可以用类似高斯函数的形式来描述调谐曲线, 峰值即对应偏好刺激.当神经元足够多,所有神经 元的偏好刺激可以覆盖全部输入刺激的取值范围 (感受野)时,群体编码即相当于把一个输入信号用 一组高斯型核函数展开,因此即使若干个神经元上 信息缺失也不影响整体工作效果.相比于用单个神 经元来处理精确的感知或运动控制,群体编码能够 显著提高准确性,并且具有去除噪音、实现短时记 忆和复杂非线性响应等功能^[24].

2.2 磁性隧道结的动力学

磁性隧道结是常见的自旋电子学器件,由两个 铁磁层夹着一个较薄的绝缘层组成,如图 1(a)所 示,其中上层的磁化方向固定,中间是隧穿层,下 层的磁化方向可以通过施加外界磁场或激励电流 控制翻转.磁性隧道结有两个稳定的状态 (能量极 小值),当自由层和固定层磁矩平行时称为平行态, 此时电阻较低,反之称为反平行态,具有高电阻. 当两个状态之间势垒高度 Δ*E*远大于热扰动 *k*_B*T* 时,磁性隧道结的状态是非常稳定的,适合作为磁 随机存储器中的非易失二值存储单元.而当势垒高 度与*k*_B*T* 相当时,磁性隧道结的状态会受外界的热 噪声影响而改变^[28].偏压或电流可以给自由层施 加自旋转移力矩^[20],从而调节隧道结在两个状态 上的概率分布,进而调控其作为神经元的调谐曲 线,实现基于磁性隧道结的群体编码系统.

本文使用的磁性隧道结多层薄膜通过磁控 溅射沉积,其核心结构自下而上为 PtMn(15 nm)/ Co₇₀Fe₃₀(2.5 nm)/Ru(0.85 nm)/Co₄₀Fe₄₀B₂₀(2.4 nm)/ MgO(0.80 nm)/Co₆₀Fe₂₀B₂₀(2.1 nm).薄膜沉积后 在1T磁场下,300 ℃ 退火 2 h.随后通过光刻、刻 蚀制备成 50 nm ×130 nm 的椭圆形器件.沿与磁 性隧道结器件难轴夹角 20°方向施加适当强度的磁 场,然后向器件注入直流偏置电流并通过示波器测 量器件两端电压.电流产生的自旋转移力矩与易轴 方向的磁场分量竞争使自由层磁矩在与钉扎层磁 矩平行和反平行方向翻转,通过示波器可以观察到 电压信号随机跳变,且随机跳变的频率随电流强度 变化.随机翻转测试在室温下进行.图1(b)展示了 实验上用电流调节磁性隧道结翻转频率的测量结 果.当电流为-60 µA 时,磁性隧道结处于高阻态



图 1 (a) 磁性隧道结示意图; (b) 不同电流下磁性隧道结电阻随时间的变化; (c) 磁性隧道结翻转频率与电流的函数关系; (d) 一组存在不同偏置电流的隧道结实现群体编码 (散点为 Néel-Brown 理论仿真结果, 实线对应隧道结的调谐曲线) Fig. 1. (a) Schematic of a magnetic tunnel junction; (b) measured electrical resistance of a magnetic tunnel junction as a function of time under different electrical currents; (c) switching frequency of a magnetic tunnel junction as a function of electrical current;

(d) population coding using a group of magnetic tunnel junctions with different bias currents (The dots are simulated data using

的概率较大;而当电流为-10 μA 时,磁性隧道结在 高阻态和低阻态之间随机翻转;当电流为 15 μA 时,磁性隧道结处于低阻态的概率较大.注意到在 不同电流下磁性隧道结的高阻态不同,这是由于磁 性隧道结的电流-电压曲线不是线性的,在高阻态 下呈抛物线型.如果将磁性隧道结在高低阻态之间 的一次翻转当作神经元的一次脉冲发放,就可以通 过调节输入电流来改变磁性隧道结的发放率.随着 输入电流的增大,磁性隧道结的发放率先增大后减 小,如图 1(c)所示.图中的每个数据点为 10 次独 立采样结果的平均值和标准差,每次独立采样 97300 次,大约 31 s.

the Néel-Brown theory and the solid lines are the corresponding tuning curves).

2.3 Néel-Brown 理论

Néel-Brown 理论可以很好地描述超顺磁隧道 结的随机动力学翻转^[30].处于平行态或反平行态 上的磁矩翻转的弛豫时间 $\tau_{P/AP}$ 由势垒高度 ΔE 和 经过隧道结的电流 I 共同决定,即

$$\frac{1}{\tau_{\rm P/AP}} = \phi_0 \exp\left[-\frac{\Delta E}{k_{\rm B}T} \left(1 \pm \frac{I}{I_{\rm c}}\right)\right],\qquad(1)$$

式中 ϕ_0 为尝试频率,与器件的翻转机制有关,本文 中取为10⁹ Hz. k_B为玻尔兹曼常数,T为环境温 度,I_c为磁性隧道结翻转的临界电流.在群体编码 中用输入电流的数值大小I来编码外界刺激信号. 磁性隧道结处于平行态或反平行态的概率表示为

$$P_{\rm P/AP} = 1 - \exp\left(-\frac{\Delta t}{\tau_{\rm P/AP}}\right),\tag{2}$$

其中 Δt 为采样时间, $\Delta t = 326.5 \,\mu s.$ 如果将连续两次采样中磁矩发生改变定义为该神经元发放一次,则 Néel-Brown 理论给出了超顺磁隧道结神经元的发放频率为

$$r\left(I\right) = \frac{\phi_0 \exp\left(-\frac{\Delta E}{k_{\rm B}T}\right)}{2\cosh\left(\frac{\Delta E}{k_{\rm B}T} \cdot \frac{I}{I_{\rm c}}\right)}.$$
(3)

可以用 (3) 式拟合实验数据, 结果见图 1(c) 实线, 可得 $\frac{\Delta E}{k_{\rm B}T}$ = 17.7, $I_{\rm c}$ = 293.15 μ A. 拟合得到的能量 势垒与临界电流与文献 [27] 中类似器件的参数可 比. 可以看出当电流 $I = -16.27 \mu$ A时发放率最大, 这是因为此时电流施加的自旋转移力矩刚好抵消 了隧道结自身的偏置场.

为了使各神经元的偏好刺激均匀分布在输入 刺激 I 的变化范围内,可对各个磁性隧道结施加等 间隔电流偏置 I_{bias},最终磁性隧道结的实际输入电 流为 I = I_{input} + I_{bias}.此时这一组神经元的调谐曲 线均匀排列,如图 1(d) 所示,实线为该组磁性隧道 结的调谐曲线,散点代表利用 Néel-Brown 方程模 拟的结果.由于磁性隧道结的随机发放特性,模拟 结果与调谐曲线存在一定偏差.

3 群体编码脉冲神经网络和无监督 聚类

3.1 脉冲神经网络模型

为展示基于超顺磁隧道结的群体编码方案,设 计了一个两层的脉冲神经网络,如图 2(a)所示.网 络前端是包含若干神经元的输入层,每个神经元将 一个输入信号编码为电流通入到下一层;第一层 是群体编码层,每个输入层神经元的输入信号传递 到一组超顺磁隧道结进行群体编码;第二层输出层 包含若干有适应性阈值的泄漏累积发放神经元^[31]. 群体编码层和输出层之间为全连接.输出层的神经 元具有横向抑制连接,保证该层神经元最多仅有一 个发放.

图 2(a) 的神经网络用来实现鸢尾花数据集的 无监督聚类. 鸢尾花数据集^[32]中收集了3个类别 的样本, 分别是 Setosa 鸢尾花, Versicolour 鸢尾花 和 Virginica 鸢尾花, 每个类别有 50 个样本, 共计 150条数据. 每条数据包含 4 个属性, 分别代表该 朵鸢尾花花瓣的长度和宽度,花萼的长度和宽度. 这4个数据通过4个输入神经元x_i分别把对应的 电流输入给中间层的群体编码神经元,即若干超顺 磁隧道结中. 输入电流大小决定了磁性隧道结的 翻转频率, 而磁性隧道结发放的脉冲会引起输出层 输出神经元yi膜电位的累积,当输出神经元的膜电 位达到该神经元阈值后将发放脉冲,代表该鸢尾花 样本属于发放的输出神经元所对应的类别. 由于输 出层神经元之间存在抑制型突触(紫色连线),某个 神经元发放后会抑制其他神经元的发放. 该抑制型 突触可以设置为 N × N 的连接矩阵 (输出层神经 元个数为 N),其对角元为零,非对角元为负值,这 里我们取值为-17.5. 中间层与输出层之间的突触 连接强度是可学习的,应用依赖于当前权重的 脉冲时序依赖可塑性 (spike-timing-dependent



图 2 (a) 群体编码脉冲神经网络示意图; (b) 网络训练过程示意图; (c) 鸢尾花数据集无监督聚类测试结果; (d) 用于编码一个数据的隧道结数目与输出神经元数量对网络聚类正确率的影响

Fig. 2. (a) Schematic of spiking neural network; (b) schematic illustration of the network learning process; (c) test results of the unsupervised classification of the iris data set; (d) the influence of number of magnetic tunnel junctions used in the population coding and number of output neurons. plasticity, STDP) 算法^[31] 训练. 训练规则如下: 当 突触前神经元发放时,

$$\Delta w = -\eta_{\rm pre} x_{\rm post} (w - w_{\rm min}); \tag{4}$$

当突触后神经元发放时,

$$\Delta w = \eta_{\text{post}} x_{\text{pre}}(w_{\text{max}} - w). \tag{5}$$

这里, Δw 为权重变化量; 前突触学习率 η_{pre} 和后 突触学习率 η_{post} 分别为 0.001 和 0.01; 权重 w最 大值 $w_{\text{max}} = 1$, 最小值 $w_{\text{min}} = 0$; x_{pre} 和 x_{post} 分别为 突触前后神经元的膜电位.图 2(b) 为网络运行的 流程图, 从左到右分别为群体编码的调谐曲线, 磁 性隧道结的电阻变化示意, 泄漏型积分发放神经元 的动力学示意.

3.2 鸢尾花样本的无监督聚类

首先用理想调谐曲线的算法进行数值仿真, 每个输入变量用 12 个磁性隧道结编码,输出神经 元的数量为 30. 在每一轮训练中随机通入 100 条 数据,注意这里的数据是随机挑选,并且由于是无 监督学习,所以我们并没有区分训练集和测试集. 输出神经元的标签是动态调整的,每完成一轮训练 后,将对输出神经元刺激最强的鸢尾花类别作为该 输出神经元的标签. 在每轮训练结束后,将 150 个 样本随机输入网络进行测试,测试得到的聚类正确 率如图 2(c)中的虚线所示. 经过 15 轮训练后网络 的聚类正确率达到了饱和值的 92.6%.

生物系统的神经元在不同时刻收到相同的刺激时,发放模式并不是完全重复调谐曲线,而是具有一定的随机性,但并不影响大脑精确地执行各项复杂的动作.同样,外界热噪声扰动下的磁性隧道结的动力学具有一定的随机性,因此超顺磁隧道结的群体编码可以很好地模拟生物系统中的神经元随机动力学.为了检验网络性能,我们分别进行了10次独立的试验,并把测试正确率的平均值和标准差画在图2(c)中.可以看出,随着训练轮次的增加,基于磁性隧道结的群体编码网络正确率最终也达到了使用理想调谐曲线的算法仿真极限值(图2(c)虚线).图2(c)中插图为15轮训练后测试结果的融合矩阵,其中的标签1,2,3依次对应Setosa 鸢尾花、Versicolour 鸢尾花和 Virginica 鸢尾花,红色的深度与正确预测的标签数目相对应.

随着群体编码使用隧道结 (神经元) 数量的增加,偏好刺激在感受野中的分布更加密集,可以使

群体编码包含的信息量更大, 鲁棒性更强^[33]. 同时, 随着磁性隧道结数目的增加, 输入数据将被投影到 更高维的空间, 该组磁性隧道结的发放脉冲包含的 信息随之增加, 输出层神经元能够解码出更多的信 息, 包含了数据的更多细节, 因此磁性隧道结数目 的增加有助于提高网络的性能. 图 2(d) 展示了随 着用于编码每个变量的磁性隧道结的增加, 网络性 能逐渐增强, 当磁性隧道结数目达到 12 个时, 无 监督聚类的正确率也达到饱和.

本文还探究了输出层神经元对网络性能的影 响.执行无监督聚类的网络事先并不知道样本数据 种类数量,因此输出层神经元的数量应当大于或 等于数据种类的数量. 输出层神经元的标签由引起 它发放次数最多的鸢尾花类别决定,因此,每一类 鸢尾花由一组输出层神经元共同标注. 当输出层的 神经元达到 20个时,正确率可以达到 90%;当输 出层的神经元达到 30个,正确率就可以达到 92.6%, 达到了网络性能的极限. 需要强调, 我们的 无监督学习中使用了 STDP 的学习规则, 在实际 训练过程中表现出一定的不稳定性,对权重初值较 为敏感, 而更多的输出神经元代表网络有更大的初 始参数空间,这给训练过程带来了更多的可能性, 更加有利于 STDP 的学习, 从而避免参数落入局 域最优解导致训练失败,这也解释了为什么更多的 输出神经元会提高网络的性能.

尽管我们仅测试了样本个数较少的鸢尾花数 据集,但该网络在软件层面已经成功应用于大规模 数据集的无监督分类^[31,34].因此,基于磁性隧道结 的群体编码硬件网络也同样具有可扩展性,能够推 广到更大规模的数据分类任务.此外,本文探究了 权重的精度对网络性能的影响,模拟结果表明在 鸢尾花聚类的任务中,3比特权重下网络的性能就 可以达到 90% 以上,这表明该网络的权重可以映 射到磁畴壁等自旋多态器件,有望实现全自旋脉冲 神经网络.

4 器件的非均一性与能耗

即使同一批生长的材料和相同的刻蚀工艺制造的磁性隧道结器件通常也会有所差异,这种器件的非均一性是类脑计算硬件不可避免的问题.具体到超顺磁隧道结,其非均一性表现在器件两个参数的差别,即势垒高度Δ*E*和临界电流*I*_c.这两个参

数的差别导致隧道结调谐曲线的高度和宽度都会 发生改变^[27].为了探究器件的不均一性对网络性 能的影响,我们测试了磁性隧道结的势垒 ΔE 存在 0—10% 涨落范围下无监督聚类的准确性.测试中 的群体编码均采用每组 8 个隧道结,输出层包含 50 个神经元. ΔE 增大 (减小) 会降低 (升高) 磁性 隧道结的翻转频率,因此 ΔE 的涨落影响器件调谐 曲线的高度.例如势垒过高的器件翻转 1 次所需的 时间很长,因此平均发放速率相对要低很多.图 3(a) 插图展示了 ΔE 的涨落为 6% 时一组磁性隧道结的 调谐曲线.将这组器件放入网络中进行群体编码, 最终实现无监督分类的正确率达到 80%.整体上随 着器件势垒高度涨落增大,网络的性能略有降低^[27], 如图 3(a) 所示.



图 3 磁性隧道结 (a) 势全和 (b) 翻转临界电流的不均一 性对网络性能的影响

Fig. 3. Population coding using the magnetic tunnel junctions with a fluctuation in their (a) energy barriers and (b) critical currents for switching.

我们进一步测试了磁性隧道结的临界电流 *I*。 存在 0—80% 涨落时网络的性能,结果如图 3(b) 所示. *I*。的变化会影响调谐曲线的宽度,图 3(b)的 插图展示了 *I*。存在 80% 涨落时的一组器件的调谐 曲线,此时网络的正确率依然能够达到 90%.因此 不同器件临界电流并不会影响网络的鲁棒性,势垒 高度的变化对网络性能的影响更加显著.

上述测试结果与生物系统中群体编码的特征 一致.在生物系统中,不同神经元在收到相同的刺 激时响应也不是完全相同的,这种异质性神经元的 群体编码并不影响生物神经系统的正常工作.

磁性隧道结的发放在每个采样时间窗口内是 随机的,我们可以改变观测时间,在1个观测时间 内对磁性隧道结进行多次采样,然后将多次采样的 结果转换为1次脉冲,这样可以在统计意义上消除 一定的随机性,从而提高网络的性能.但随着观测 时间的增加,每次运行时器件能耗也会随之增加. 为了探究器件耗能和网络性能的关系,我们使用每 组群体编码包含了16个磁性隧道结,模拟观测时 间与采样时间的比值对网络性能的影响.在一个采 样时间内磁性隧道结的最大输入电流为 100 μA, 电阻取高阻态与低阻态的均值 2.016 kΩ, 因此最 大能耗为6.58×10⁻⁹J. 将其乘以观测时间与采样 时间的比值,就可得到观测时间对应的最大能耗. 图 4 给出了不同观测时间的聚类正确率,其中上轴 表示每个观测时间对应的能耗. 当观测时间为 10次采样时间时网络达到正确率的饱和值,对应 的最大能耗为为6.58×10⁻⁸J. 随着实验工艺的进 步,降低采样时间可以有效地降低器件能耗,目前 实验上报道的隧道结采样时间可以低至 8 ns^[35],



图 4 不同观测时间 (以采样时间 dt 为单位) 对网络执行 无监督聚类正确率的影响 (上轴表示每个观测时间对应的 能耗)

Fig. 4. Classification accuracy of unsupervised clustering performed by the network as a function of the different observation time (The upper axis shows the energy consumption corresponding to each observation time). 相比于 CMOS 的群体编码系统^[25],采样速度提高 了 3 个量级,采用该器件实现群体编码有望将能耗 降到 4 × 10⁻¹² J. 在系统集成方面,磁性隧道结将 连续信号直接转换为二值信号,有利于后续数字信 号的处理,而基于 CMOS 的群体编码需要设计额 外的数模转换电路.因此,基于自旋电子器件的群 体编码在面积、能耗和速度方面都有突出的优势.

5 结 论

本文提出了一种使用超顺磁隧道结的群体编码方案,利用热扰动下超顺磁隧道结的随机翻转过程及其自旋转移力矩的调控,可以用带有偏置力矩的若干隧道结的偏好刺激编码整个神经元的感受野.作为一个简单且典型的例子,我们把该群体编码方案应用到一个三层的神经网络中,实现了对鸢尾花数据集的无监督聚类,探究了用于群体编码的隧道结个数等网络参数对聚类正确率的影响.群体编码方案在应用中可以很好地克服器件非均一性等技术和工艺问题,并且能够有效地降低神经网络的能耗.

参考文献

- [1] LeCun Y, Bengio Y, Hinton G 2015 Nature 521 436
- [2] Roy K, Jaiswal A, Panda P 2019 Nature 575 607
- [3] Davies M, Srinivasa N, Lin T H, et al. 2018 *IEEE Micro.* 38 82
- [4] Pei J, Deng L, Song S, et al. 2019 Nature 572 106
- [5] Ambrogio S, Narayanan P, Tsai H, Shelby R M, Boybat I, di Nolfo C, Sidler S, Giordano M, Bodini M, Farinha N C P, Killeen B, Cheng C, Jaoudi Y, Burr G W 2018 *Nature* 558 60
- [6] Torrejon J, Riou M, Araujo F A, et al. 2017 Nature 547 428
- [7] Yao P, Wu H Q, Gao B, Tang J S, Zhang Q T, Zhang W Q, Yang J J, Qian H 2020 Nature 577 641
- [8] Zhang X M, Zhuo Y, Luo Q, et al. 2020 Nat. Commun. 11 51

- [9] Zhang Y, Wang Z R, Zhu J D, Yang Y C, Rao M Y, Song W H, Zhuo Y, Zhang X M, Cui M L, Shen L L, Huang R, Yang J J 2020 Appl. Phys. Rev. 7 011308
- [10] Jung S, Lee H, Myung S, et al. 2022 Nature 601 211
- [11] Grollier J, Querlioz D, Camsari K Y, Everschor-Sitte K, Fukami S, Stiles M D 2020 Nat. Electron. 3 360
- [12] Lan X K, Cao Y, Liu X Y, Xu K J, Liu C, Zheng H Z, Wang K Y 2021 Adv. Intell. Syst. 3 2000182
- [13] Jiang W C, Chen L N, Zhou K Y, Li L Y, Fu Q W, Du Y W, Liu R H 2019 Appl. Phys. Lett. 115 192403
- [14] Zhang Y J, Zheng Q, Zhu X R, Yuan Z, Xia K 2020 Sci. China Phys. Mech. Astron. 63 277531
- [15] Zheng Q, Mi Y Y, Zhu X R, Yuan Z, Xia K 2020 Phys. Rev. Appl. 14 044060
- [16] Zheng Q, Zhu X R, Mi Y Y, Yuan Z, Xia K 2020 AIP Adv. 10 025116
- [17] Sengupta A, Roy K 2016 Phys. Rev. Appl. 5 024012
- [18] Yu W C, Xiao J, Bauer G E W 2021 Phys. Rev. B 104 L180405
- [19] Song K M, Jeong J S, Pan B, et al. 2020 Nat. Electron. 3 148
 [20] Kurenkov A, DuttaGupta S, Zhang C, Fukami S, Horio Y,
- Ohno H 2019 Adv. Mater. **31** e1900636 [21] Romera M, Talatchian P, Tsunegi S, et al. 2018 Nature **563**
- 230[22] Banino A, Barry C, Uria B, et al. 2018 Nature 557 429
- [23] Mi Y, Katkov M, Tsodyks M 2017 Neuron 93 323
- [24] Pouget A, Dayan P, Zemel R 2000 Nat. Rev. Neurosci. 1 125
- [25] Thakur C S, Wang R, Hamilton T J, Tapson J, van Schaik A 2016 IEEE Transactions on Circuits and Systems I: Regular Papers 63 211
- [26] Tuma T, Pantazi A, Le Gallo M, Sebastian A, Eleftheriou E 2016 Nat. Nanotechnol. 11 693
- [27] Mizrahi A, Hirtzlin T, Fukushima A, Kubota H, Yuasa S, Grollier J, Querlioz D 2018 Nat. Commun. 9 1533
- [28] Cai J L, Fang B, Zhang L K, Lv W X, Zhang B S, Zhou T J, Finocchio G, Zeng Z M 2019 Phys. Rev. Appl. 11 034015
- [29] Cai K M, Yang M Y, Ju H L, et al. 2017 Nat. Mater. 16 712
- [30] Li Z, Zhang S 2004 Phys. Rev. B 69 134416
- [31] Diehl P U, Cook M 2015 Front. Comput. Neurosci. 9 99
- [32] Fisher R A 1936 Annals of Eugenics 7 179
- [33] Dayan P, Abbott L F 2001 Theoretical Neuroscience (Cambridge, MA: MIT Press) pp108–112
- [34] Biswas A, Prasad S, Lashkare S, Ganguly U 2016 arXiv: 1612.02233
- [35] Hayakawa K, Kanai S, Funatsu T, Igarashi J, Jinnai B, Borders W A, Ohno H, Fukami S 2021 *Phys. Rev. Lett.* 126 117202

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Implementation of unsupervised clustering based on population coding of magnetic tunnel junctions^{*}

Zhang Ya-Jun¹⁾ Cai Jia-Lin²⁾ Qiao Ya¹⁾ Zeng Zhong-Ming^{2)†} Yuan Zhe^{1)‡} Xia Ke³⁾

1) (Center for Advanced Quantum Studies, Department of Physics, Beijing Normal University, Beijing 100875, China)

2) (Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences, Suzhou 215123, China)

3) (Beijing Computational Science Research Center, Beijing 100193, China)

(Received 9 February 2022; revised manuscript received 13 April 2022)

Abstract

Developing suitable algorithms that utilize the natural advantages of the corresponding devices is a key issue in the hardware research of brain-inspired computing. Population coding is one of the computational schemes in biological neural systems and it contains the mechanisms for noise reduction, short-term memory and implementation of complex nonlinear functions. Here we show the controllable stochastic dynamical behaviors for the technically mature spintronic device, magnetic tunnel junctions, which can be used as the basis of population coding. As an example, we construct a two-layer spiking neural network, in which groups of magnetic tunnel junctions are used to code input data. After unsupervised learning, this spiking neural network successfully classifies the iris data set. Numerical simulation demonstrates that the population coding is robust enough against the nonuniform dispersion in devices, which is inevitable in fabrication and integration of hardware devices.

Keywords: magnetic tunnel junction, population coding, spiking neural network, unsupervised learning PACS: 85.70.–w, 07.05.Mh DOI: 10.7498/aps.71.20220252

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 11734004, 12174028).

[†] Corresponding author. E-mail: zmzeng2012@sinano.ac.cn

[‡] Corresponding author. E-mail: zyuan@bnu.edu.cn

专题: 面向类脑计算的物理电子学

基于 3D-NAND 的神经形态计算

陈阳洋1)2)3) 何毓辉3)4) 缪向水3)4) 杨道虹1)2)3)†

(华中科技大学,博士后流动站,武汉 430074)
 (武汉新芯集成电路制造有限公司,博士后工作站,武汉 430205)

3) (江城实验室, 武汉 430205)

4) (华中科技大学集成电路学院, 武汉 430074)

(2022年5月16日收到; 2022年9月27日收到修改稿)

神经形态芯片是一种新兴的 AI 芯片. 神经形态芯片基于非冯·诺依曼架构, 模拟人脑的结构和工作方式, 相比冯·诺依曼架构的 AI 芯片, 神经形态芯片在效率和能耗上有显著的优势. 3D-NAND 闪存工艺成熟并且存 储密度极高, 基于 3D-NAND 的神经形态芯片受到许多研究者的关注. 然而由于该技术的专利性质, 少有基 于 3D-NAND 神经形态计算的硬件实现. 本文综述了用 3D-NAND 实现神经形态计算的工作, 介绍了其中前 向传播和反向传播的机制, 并提出了目前 3D NAND 在器件、结构和架构上需要的改进以适用于未来的神经 形态计算.

关键词:神经形态计算, 3D-NAND, 存算一体架构 **PACS**: 07.05.Mh, 85.35.-p, 84.30.-r, 87.18.Sn

DOI: 10.7498/aps.71.20220974

1 研究背景

1.1 神经形态计算是未来通用人工智能的 重要路径

数据、算法和算力是人工智能 (artificial intelligence, AI) 的三大要素, 未来 AI 的发展将面临算 力不足的瓶颈 (如图 1 所示^[1]). 一方面, 云端计算 中的通用 AI 模型性能强大, 但参数庞大, 通常采 用 CPU/GPU 硬件平台进行训练, 训练成本高昂 且难以普及. 另一方面, 随着 5G、物联网与工业 4.0 的发展, 越来越多的 AI 应用在边缘端设备中 设计和部署, 需要定制化 AI 芯片满足功耗和成本 限制下的计算需求, 例如阿里的"含光"、华为的 "昇腾"和寒武纪的"思源"等. 通用 CPU/GPU 以 及 AI 加速芯片, 均基于传统的冯·诺依曼ኺ颈"不可避 构, 计算和存储单元分离, "冯·诺依曼瓶颈"不可避 免 (如图 2 所示^[2-4]): 第一,数据在计算和存储单 元之间不停来回传输,消耗大部分的计算时间和功 耗; 第二,处理器和存储器之间运算速度的明显差 异限制了整体系统的计算效率.面对这一问题,存 储和计算融合是未来的发展趋势,新型的计算架构 逐渐兴起,其中包括近存计算 (near-memory computing)、存内计算 (in-memory computing) 以及神 经形态计算 (neuromorphic computing).受人脑智 能启发的神经形态芯片引起了学术和工业界的极 大兴趣.

人脑在复杂和陌生场景下的学习、推理和决策 能力远超过传统计算机.人脑有超过 10¹¹ 个神经 元和 10¹⁵ 个突触,功耗只有 20 W^[5].神经形态芯片 模拟人脑的结构和工作方式 (如图 3 所示^[6]):在结 构上,用互补金属氧化物半导体 (complementary metal-oxide-semiconductor, CMOS)或新型器件 模拟神经元和突触,并将神经网络映射到突触阵列

[†] 通信作者. E-mail: alan_yang@xmcwh.com

^{© 2022} 中国物理学会 Chinese Physical Society









图 2 限制芯片性能提升的冯·诺依曼瓶颈^[2-4]

Fig. 2. The von Neumann bottleneck limits chip performance promotion^[2–4].

中;在工作方式上,用基于权重模拟值计算的人工 神经网络 (artificial neural network, ANN) 或基于 脉冲计算的脉冲神经网络 (spiking neural network, SNN) 作为算法模型,输入的事件用电压脉冲编码, 经过突触后转化为电流输入到后级神经元进行积 分,后级神经元达到阈值电压后便向下一级神经元 发放电压脉冲.相比冯·诺依曼架构芯片,神经形态 芯片高并行、低功耗和存算融合的特性,有望成为 未来通用 AI 的理想硬件方案.

1.2 基于存储器的神经形态芯片

神经形态芯片根据实现的器件方案可分为基 于传统 CMOS 的神经形态芯片和基于存储器的神 经形态芯片两种类型.基于 CMOS 的神经形态芯 片代表性成果包括 TrueNorth, SpiNNaker, Brain-ScaleS, Loihi, 天机芯和达尔文等.这些芯片的突 触和神经元采用基于 CMOS 的数字电路或者数 模混合电路来搭建,模拟单个神经元或突触行为需 要靠多个 CMOS 晶体管组成的电路模块来实现,



图 3 神经形态计算的原理与芯片架构 6





图 4 基于非易失存储器的神经形态计算硬件方案^[7] (a) 几种作为突触的非易失性存储器件,其中包括 PCM, ReRAM, STT-MRAM, FeRAM 和 FeFET; (b) 突触和神经元构成 crossbar 阵列结构用于神经形态计算; (c) 神经形态计算芯片的架构

Fig. 4. Use of non-volatile memory devices as synaptic storage^[7]: (a) Non-volatile memory cell as artificial synapse including PCM, ReRAM, STT-MRAM, FeRAM, and FeFET; (b) the implementation of neuromorphic computation on crossbar array consists of artificial synapses and neurons; (c) typical architecture of the neuromorphic chip.

集成度和功耗受到限制,并且断电后信息无法保存.基于存储器的神经形态芯片从底层器件仿生的角度出发,用存储器件模拟神经元和突触,在功耗和硬件代价上有明显的优势.近几年,国内外研究机构展示了众多基于存储器的神经形态计算成果,硬件方案包括主流的闪存 (NOR/NAND Flash),以及阻变存储器 (ReRAM)、相变存储器 (PCM)、自旋转移矩磁存储器 (STT-MRAM)、铁电存储器 (FeRAM) 和铁电晶体管存储器 (FeFET)等新型存

储器.这些存储器的结构和工作原理如图 4 所示^[7]. 表1列出这几类存储器的性能指标,其中参数源于 已有阵列或芯片实现的研究工作.

1.3 NAND 用于神经形态计算的优点与局 限性

突触器件需要具有高集成度、低能耗、耐擦 写、CMOS工艺兼容以及模拟权重调制的特性.其 中模拟权重调制特性需要突触有多个权重、并且

			1	0 0		51	
	新型存储器					主流存储器	
-	PCM	ReRAM	STT-MRAM	FeRAM	FeFET	NOR flash	NAND flash
器件结构	1T-1R	1T-1R	1T-1R	1T-1C/ 1T-1FTJ	1T	1T	$1\mathrm{T}$
器件面积F ²	4 - 40	6-26	9-75	30 - 40	10 - 30	10	4
工艺节点/nm	$14^{[8,9]}$	$14^{[10]}$	$22^{[11]}$	$130^{[12]}$	$22^{[13]}$	40	15(2D) $80(3D)^{[14]}$
芯片容量	16—64 GB (Intel optane)	$1 \text{ MB} - 32 \text{ GB}^{[10,15]}$	16 kB $4 \text{ GB}^{[16,17]}$	16 kB— 128 MB ^[18,19]	64 kB— 32 MB ^[13,20]	1 MB—2 GB (Micron/Infineon/Macro- nix product manuals)	1 TB ^[21]
写电压/V	< 3	< 1	< 2	< 3	~ 1.5	10 - 15	15 - 25
写时间/ns	75 ^[9]	$1 - 10^{[22]}$	$3-14^{[23,24]}$	$1 - 10^{[25]}$	$1 - 10^{[25]}$	$10000^{[26]}$	$100000^{[27]}$
写功耗/(pJ·bit ⁻¹)	$1^{[28]}$	$\sim \! 10^{[29]}$	$4.5^{[30]}$	$0.1^{[25]}$	$1 - 10 \mathrm{fJ^{[25]}}$	$49^{[26]}$	$\sim \! 1000^{[31]}$
读时间	12—130 $ns^{[8,32]}$	2.9—21 $ns^{[33,34]}$	2—26 $ns^{[30,35]}$	1 25 ns	$1{-}10~\mathrm{ns}$	11 ns	$3 \ \mu s^{[27]}$
读功耗/(pJ·bit ⁻¹)	$2.47^{[36]}$	$1.76^{[29]}$	$0.7^{[30]}$	$9.8 - 19.2^{[37]}$	$0.27 \ {\rm fJ/bit^{[38]}}$	$2.2^{[26]}$	$\sim \! 100^{[31]}$
开关比	$> 10^4$	$> 10^3$	> 2	> 2	$> 10^8$	$> 10^8$	$> 10^8$
擦写次数	$> 10^{10} {}^{[22]}$	$>10^{12}$	$>10^{12}$	$>10^{12}$	$10^{5} - 10^{9}$	$< 10^6$	$< 10^5$
保持时间/a	> 10	> 10	> 10	> 10	> 10	> 10	> 10
存储比特	2	$6^{[39]}$	1	$3-4^{[39]}$	2 - 3	2	4

表 1 几种非易失性存储器的性能参数 Table 1. Benchmark table of performance of emerging memories and typical memories.





Fig. 5. Performance comparison of various memories (data was extracted from Table 1 and plotted into a radar diagram).

权重对称线性变化. 从图 5 可以看到, 新型存储器 ReRAM, PCM, STT-MRAM, FeRAM 和 FeFET 在读写速度、读写电压、擦写次数和功耗方面具有 突出的优势. 然而 NAND/NOR Flash 的器件特性 使得其在众多硬件方案中有不可替代的优势: 1) ReRAM, PCM, FeRAM 等二端结构器件为了防 止潜行通路 (sneak path) 的产生需要集成用于选 通的晶体管, NAND/NOR 的基本存储单元是三端 的 MOSFET, 栅电极自带选通功能, 硬件代价小; 2) NOR/NAND flash 工艺成熟, 器件的阈值电压 分布稳定,即权重的分布稳定,并且单元之间的性能高度一致,适合高精度的数值运算;3)得益于 NAND/NOR存储单元的栅控机制,即阈值电压由器件单元俘获层中俘获电荷数量决定,通过增量步进脉冲写入技术 (incremental step pulse program, ISPP)可获得稳定的阈值电压分布,权重 (电导率)的模拟权重调制特性更容易实现.

NOR 和 NAND 的存储单元分别为浮栅型闪存 (floating gate flash, FGF) 和电荷俘获型闪存 (charge trapping flash, CTF), 二者均利用栅介质
中的存储电荷调控阈值电压,主要区别在于电荷存 储层的材料、电荷俘获机制和阵列的结构,如图 6 所示. FGF 的浮栅层通常为掺杂的多晶硅, 利用热 电子注入效应实现电荷的存储:在源、漏电极和 栅、源电极之间施加高电压,电子在沟道中被源漏 电场加速后,被栅源之间的电场吸引,穿过隧穿氧 化层注入到浮栅层中. CTF 的电荷俘获层通常采用 氮化硅材料,利用 F-N 隧穿效应 (Fowler-Nordheim tunneling) 实现电荷的存储: 源、漏接地, 栅极施加 高电压, 源极电子通过 F-N 隧穿效应穿过隧穿氧 化层进入浮栅. 在阵列结构方面, NOR 的布局采 用并行结构,每个存储单元均有源线 (source line, SL) 和字线 (word line, WL) 引出, 随机读写的速 度快,但过多的布线和较大的器件尺寸 (10F2) 使 得其存储密度难以进一步提升.相比之下 NAND 采用串行结构,每个存储单元的源极无需单独引 线,因此具有更小的单元特征尺寸 (4F²) 和更高的 存储密度.利用三维集成技术,NAND的存储密度 提高至 TB 量级. 2022 年 5 月, 国内厂商长江存储 量产的 3D-NAND 已达到 192 层, 2022 年 8 月 SK Hynix 已宣布量产 238 层 3D-NAND, 如此高的集 成密度碾压 NOR 以及其他新型存储器.



图 6 NOR 和 NAND 的电路结构

Fig. 6. Circuit structure diagram of NOR flash and NAND flash.

NAND 的高存储密度为神经形态计算提供了 充足的硬件资源、更高的存储密度,意味着芯片能 分配更多的硬件资源用于映射更大规模、更多数量 和种类的神经网络,芯片性能也越强. 然而 3D-NAND 的存储单元,电荷俘获型晶体管 (transistor with charge trap layer, CTL) 的擦写次数不高 (如 表 1 所示,次数 < 10⁵),因此基于 3D-NAND 的 AI 芯片适用于权重更新不太频繁的场景.

AI系统通常涉及训练 (training) 和推断 (infer-

ence) 过程, 训练过程中需要输入大量的样本数据, 并且根据输出反馈不断调节芯片中权重的分布直 至输出达到预期的精度, 使芯片具备学习能力, 这 个过程涉及频繁的权重更新. 而推断过程是在已训 练好的芯片上输入新的数据, 完成指定任务的过 程, 权重更新频率低. 因此, 3D-NAND 芯片在执行 推断任务的场景中, 得益于其超高的存储密度, 相 对其他种类的存储器芯片具有显著的优势.

2 基于 3D-NAND 神经形态计算的 研究进展

2.1 3D-NAND 的结构和突触特性

随着 2D-NAND 工艺微缩到 14 nm 节点,每 个单元只能容纳少量的电子,并且单元之间电子的 串扰问题使得尺寸继续微缩变得愈加困难且不够 经济.在不降低工艺节点的前提下提高存储密度和 降低成本, 3D-NAND 技术成为了必然的选择. 铠 侠 (原东芝)、镁光、海力士和旺宏均提出了各自的 3D-NAND 技术方案,如图 7 所示^[40].这些技术方 案从 3D 堆叠方式上可分为两种:一种是栅极堆叠 (gate stack) 结构,沟道为垂直方向;另一种是沟道 堆叠 (channel stack) 结构,栅极为垂直方向,如 图 8 所示^[41,42].

其中存储层一般采用浮栅 (floating gate, FG) 或者电荷俘获层 (charge trap layer, CTL).一般 来说, FG采用掺杂的多晶硅,存储单元尺寸较大, 存储电荷量较多,阈值电压窗口较大,保持特性较 好.CTL 材料为氮化硅,存储电荷量相对较少,存 储单元的阈值电压窗口和保持特性略差,但CTL 器件的尺寸小,集成度更高.目前三星量产的96 层 V-NAND 技术便是基于栅极堆叠和CTL 层的 TCAT 结构.

神经形态计算芯片中突触是基本的结构单元, 在 3D-NAND 中用 CTL 器件作为突触 (图 9(a))^[43]. 突触的权重 (电导) 非易失且连续可调, 即模拟权 重调制特性.存储用 3D-NAND 中 CTL 器件的阈 值电压通过增量步进脉冲写入 (incremental step pulse program, ISPP) 机制进行调节 (图 9(b))^[44]. 用于神经形态计算的 3D-NAND 中, CTL 不仅可 用 ISPP 机制进行调节, 而且可用多次同脉冲写 入 (multiple identical pulses program, MIPP) 机制 进行调节以模拟权重调制特性 (图 9(c))^[45].



图 7 各种 3D-NAND 的技术方案^[40]





图 8 (a) 栅极堆叠的 TCAT 技术^[41]; (b) 沟道堆叠的 VG-NAND 技术^[42] Fig. 8. (a) TCAT technology with gate stack architecture^[41]; (b) VG-NAND technology with channel stack architecture^[42].

2.2 基于 3D-NAND 神经形态计算的相关 工作

神经形态计算包含3个过程:前向传播、反向 传播和权重更新.前向传播,即前级神经元发放的 电压脉冲经过突触阵列转化为电流脉冲传递给后 级神经元.把前级神经元发放的输入电压脉冲和输入到后级神经元的电流脉冲视为向量,突触的权重构成矩阵,那么前向传播过程等价为输入信号和突触权重进行乘加运算(multiply-and-accumulation, MAC)的累计,即向量矩阵相乘(vector matrix multiplication, VMM).反向传播过程中计算各层



图 9 (a) CTL 单元的结构示意图 (以栅极堆叠结构为例)^[43]; (b) 采用 ISPP 机制调控 3D-NAND 中 CTL 器件的阈值电压^[44]; (c) 3D-NAND 中 CTL 器件的模拟电导特性, 即模拟权重调制特性^[45]

Fig. 9. (a) Illustration of typical gate-stack type 3D-NAND^[43]; (b) ISPP modulation of threshold voltage in CTL device^[44]; (c) analog conductivity characteristics of CTL devices in 3D-NAND^[45].

突触的误差: 对输出神经元的结果与目标结果进行 比对, 从输出神经元往输入神经元方向逐层计算各 层突触权重的误差. 最后根据计算结果更新各层 突触的权重. 由于最近几年才被学术界关注, 3D-NAND 用于神经形态计算的相关报道并不多, 应 用多集中在前向传播和反向传播方面.

2.2.1 3D-NAND 用于前向传播

3D NAND 是多层 2D NAND 的堆叠.用于神 经形态计算时, 3D NAND 比 2D NAND 多了层间 选通的操作.为了使读者更方便地理解输入编码、器件选通、差分对突触和权重转置的概念,首先介 绍 Lee 等^[45,46] 基于 2D NAND 的神经形态计算的 工作,如图 10 所示.图 10(a) 为 2D NAND 的神经 形态计算法则,具体地:

1) 前级神经元 X_i 与后级神经元 Y_j 之间的权重 W_{ij} 用差分对实现,即 $W_{ij} = G_{ij}^+ - G_{ij}^-$,其中 G_{ij}^+ 为 正权重, G_{ij}^- 为负权重.假设第l-1层有N个神经 元 $X_{i=1,2,...,N}$, X_i 输出信号 a_i^{l-1} ,第l层具有M个 神经元 $Y_{j=1,2,...,M}$.

2) 前向传播. 后级神经元 Y_j 接收到的电流信 号为 $\sum_{i}^{N} W_{ij} a_i^{(l-1)}$, 硬件上用电压 V_i^{l-1} 作为输入 施加在 NAND string 的漏端上, 即 Y_j 接收到的电 流信号为 $\sum_{i}^{N} (G_{ij}^+ - G_{ij}^-) V_i^{(l-1)}$. 3) 反向传播.前向传播中得到的输出与预期的输出之间往往存在误差,为了使神经网络达到预期的识别率,需要计算出每层突触的误差然后进行权重更新.反向传播中将误差信号从最后一层神经元往第一层神经元传递,实现每层突触误差的计算.第*l*-1层中神经元*X_i*接收的误差信号为

$$\delta_j^{l-1} = \sum_j^M W_{ij} \delta_j^l f'\left(s_i^{l-1}\right),$$

其中 $f'(s_i^{l-1})$ 为神经元 X_i 激活函数的梯度.硬件上 将 δ_j^l 转化为对应的电压 V_j^l ,神经元激活函数多采 用 ReLu 函数 (一阶梯度值为 1),那么

$$\delta_j^{l-1} = \sum_j^M (G_{ji}^+ - G_{ji}^-) V_j^l f' \left(V_i^{l-1} \right) = \sum_j^M (G_{ji}^+ - G_{ji}^-) V_j^l$$

4) 权重更新. 根据已经得到的误差, 进行权重 更新, 权重的更新量 $\Delta W_{ij} = -\eta \delta_j^l f(s_i^{l-1})$, 其中 η 为学习率. 硬件上, CTL 器件的阈值电压随写入脉 冲的增大而上升, 器件电导 (权重) 减小. 因此, 当 $\Delta W_{ij} > 0$ 时, 需要增大突触的权重, 那么在负突触 上施加写入脉冲, 负突触权重减小 $|\Delta G_{ij}^-|$, 突触对 权重增加 $|\Delta G_{ij}^-|$. 当 $\Delta W_{ij} < 0$ 时, 需要减小突触对 的权重, 那么在正突触上施加写入脉冲, 正突触权 重减小 $|\Delta G_{ij}^+|$, 突触对权重减小 $|\Delta G_{ij}^+|$.



图 10 (a) 基于 NAND 的神经形态计算法则^[46]; (b), (c) 前向传播和反向传播过程中 NAND 的操作方法^[45] Fig. 10. (a) Learning rule of software-based and NAND-based neural network^[46]; 2D-NAND operation method in (b) forward and (c) backward propagation^[45].

前向传播和反向传播的具体操作方法如 图 10(b) 和图 10(c) 所示. 前向传播: 1) 输入 X;采 用幅值编码, 施加在 BL上. 2) NAND 的每个 page 等同于神经元 (X₀, X₁, …, X₇₈₅) 对一个后级 神经元 (Yi) 进行全连接的突触. 对目标 page 施加 V_{read} ,其他 page 施加 V_{pass} .其中 $V_{\text{read}} < V_{\text{pass}} <$ Vprogram, 施加 Vpass 的 CTL 器件处于导通状态, 可 视为导线. 输入的电压脉冲, 经过目标 page 后转化 为电流在 SL 上相加, 完成一次 VMM 过程. 3) 重 复上述操作,依次读取每个 page 的电流,完成一 次前向传播过程. 前向传播中, page 中 CTL 器件 为 785 × 2个, 一共有 51个 page. 由于 NAND 的 串行结构无法将误差信号从 SL 端输入,反向传播 过程中误差信号 (δ_1 δ_2 , …, δ_{51}) 仍从 BL 端输入, 但此时一个 page 中的 CTL 器件数为 51 × 2个, 共有 785 个 page. 因此反向传播过程要另外选取 硬件资源,映射权重时突触阵列的配置与前向传播 时的阵列互为转置,如图 10(c) 所示.

2019年佐治亚理工的余诗孟和清华大学的钱 鹤等^[47]提出了一种基于 3D-NAND 的 VMM 方案, 如图 11(a) 所示. 需要指出的是, 通常情况下 3D-NAND 中一个 WL 控制一个平面的器件, 但图 11(a) 中每一层器件沿 Y方向均有独立的 WL 作为输入端. 具体的 VMM 的过程为: 1) 输入信号采用二值编码, 通过地址解码器和传输门电路实现不同层的选通 (如图 11(b) 所示). 左侧传输门的输入端施加导通电压 V_{pass}, 地址解码器连接在传输门的 PMOS 栅上. 右侧传输门的输入端施加选通电压 V_{sel}, 地址解码器连接在传输门的 NMOS 栅上. 当输入信号为 0时, 左侧传输门开启, 右侧传输门关闭, 机上施加 V_{pass} 电压. 当输入信号为 1时, 左侧传输门打开, WL 上施加 V_{sel} 电压. 2) SL 上施加读电压 V_{read}, 经过目标层突触转化为电流在 BL 上读取. 从 XY平面上看, CTL 器件以 经典的 crossbar 形式排列.

余诗孟、钱鹤等^[47]提出的基于 3D-NAND 的 VMM 方案,并未用于神经形态计算.如果沿用此 方案进行神经形态计算,其优势在于同一平面内 的 CTL 单元为 crossbar 结构,反向传播可采用相 同的操作方法,误差信号 δ_i从 WL 输入, SL 上施加



图 11 一种基于 3D-NAND 的 VMM 方案^[47] (a) 3D-NAND 的电路结构和 VMM 的操作方法; (b) 用于 3D-NAND 层间选通的 外围电路结构

Fig. 11. A case of using 3D-NAND for VMM operation^[47]: (a) Circuit diagram and bias scheme of 3D-NAND array architecture for VMM; (b) peripheral circuitry for layer-to-layer selection.

V_{read}, BL上读取电流, 无需另外选用硬件资源. 并 且后续的权重更新可通过WL上施加电压脉冲实 现. 但进一步的工作需要考虑两个方面: 1)目前不 存在这种结构的 3D-NAND. 同一平面上制备独立 的WL将大大增加工艺难度、单元尺寸和引线的 复杂度, 硬件实现难度大; 2) 栅压有无作为输入 (二值编码), 那么所选通的器件必须工作在饱和区, CTL 只能有 1 bit 的存储态, 如果用一个 CTL 代 表一个突触, 那么神经网络的精度只有 1 bit. 要进 一步提高神经网络的精度, 可以用多个器件等效为 一个具有多比特精度的突触, 虽然硬件代价成倍增 加, 但 3D-NAND 的大容量可轻松满足其需求.

2019年 Lee 等^[48]研究了如何用 2D NAND 实现二值神经网络 (binary neural network, BNN)的前向传播,如图 12 所示.二值神经网络的输入、权重和输出均只有两种状态,运算过程较为简单,适合对精度要求不高的推断 (inference) 过程. BNN的运行只需要器件有两个稳定可区分的阻态,对硬件要求不高.图 12(a) 展示了用 SLC NAND 进行二值运算的原理 (single level cell, SLC 即存储态为1 bit 的 CTL 器件),输入信号施加在源端用于开关的晶体管的 WL 上 (即 bit line selector, BLS 或者select gate at drain side, SGD). 同一个 page 上两个相邻的 SLC 器件构成一个突触对,两个器件中有

且只有一个为写入状态. 左侧的 string 上的 BLS 有 电压,右侧无电压时,输入标记为+1,反之标记为 -1. 突触对左侧 SLC 器件为未写入状态 ($V_{\text{th, low}}$),右 例为写入状态 ($V_{\text{th, high}}$)时,突触权重标记为 W =+1,反之 W = -1.输入信号为+1,当 W = +1时, SL 端输出电流 I_{SL} ,如果将 I_{ref} 设为 0.5 × I_{SL} 时,电 流经过差分放大器,输出+0.5 × I_{SL} ,转化为正电 压,标记为+1.反之当 W = -1时, SL 端无电流输 出,电路输出-0.5 × I_{SL} ,读出负电压,标记为-1.同 理,当输入信号为-1 时,权重分别为+1和-1 时,输 出分别为-1和+1.这种输入和权重状态相同才有电 流输出的过程等效为 XNOR 逻辑运算.

BNN 前向传播的原理如图 12(b) 所示. 其中 一个 SLC 器件依次对应一个前级神经元对后级多 个神经元的突触. 前级神经元的输入视为一组向 量,同时输入到 NAND 的 BLS 上,如图 12(c) 所 示. 基于这种硬件方案, Lee 等^[48]设计了二值全连 接神经 网络以及二值卷积神经网络分别用于 MNIST 和 CIFAR-10 图片数据的识别, 网络的训 练次数与识别率的关系如图 12(d) 所示, 识别率分 别达到 98.12% 和 87.11%.

2019年 Lue 等^[49]提出一种用基于 SLC 3D-NAND 的卷积核映射和前向传播方案,用多个 BL 输入和 SLC 器件实现 4 bit 精度的输入和 4 bit



图 12 基于 2D NAND 的二值神经网络 BNN^[48] (a) 相邻的两个 CTL 器件组成差分对形式的突触, 与输入的信号进行同或运 算 (XNOR); (b), (c) BNN 前向传播中 NAND 的操作方法以及 NAND 电路示意图; (d) 采用二值全连接神经网络和二值卷积神经 网络分别用于 MNIST 和 CIFAR-10 图像库的识别性能

Fig. 12. A synaptic architecture based on 2D NAND for binary neural network (BNN) ^[48]: (a) NAND string structure for XNOR operation, in which two neighboring CTL device constructs a differential pair as one synapse; (b) operation scheme for forward propagation; (c) schematic diagram of synaptic array architecture; (d) the performance of binarized multi-layer and convolutional neural networks for MNIST and CIFAR-10 database recognition task respectively.

精度的权重,如图 13 所示,卷积神经网络 CNN 往 往采用多个卷积核,对图片进行卷积操作需要进行 大量的 MAC 运算, 原理如图 13(a) 所示. 图片中 每个像素对应一个输入信号,一次卷积的过程等同 于像素输入与卷积核中对应的权重进行 MAC 运 算(即前向传播). 卷积核中的权重具有多 bit 精度, 可以将权重拆分为高位和低位,存放在多个 SLC 中, 例如 2 个 SLC 可以实现 2 bit 存储状态. 2 bit 精度的输入和 2 bit 精度的权重的 MAC 过程 (2 bit input & 2 bit weight, 2I2W) 如图 13(b) 所示. 输 入信号的两位 X1 (0)和 X1 (1)先后施加在 BL1 上, 权重的两位分别存储在 2 个 SLC 器件上, 即 W(1-1, *i*) 和 W(1-2, *i*). MAC 过程的得到的总 电流为 $X_1(0) \times [W(1-1,j) + 2 \times W(1-2,j)] + 2X_1 \times$ $[W(1-1, j) + 2 \times W(1-2, j)], 分 4 次相乘后移位相加$ 获得. 依次类推, 4 bit 精度的 MAC 过程 (4I4W) 可拆分为2组2 bit 的输入和2组2 bit 的权重

进行相乘后移位相加,工作原理如图 13(c)所示. 图 13(d) 展示了卷积核电路的工作原理. 图中沿 BL 方向划分了 M个 block, 每个 block 代表不同 的通道,每个 block 中的每一层代表一个卷积核. 输入信号从 BL 进入, 3 根 BL 构成一个 2 bit的输 入,一个4 bit 精度的输入信号需要6根 BL(图中 只绘出一根). 3个 SLC 器件构成一个 2 bit 精度的 权重,4 bit 精度的权重需要 6个 SLC,对应地需 要6根SSL(图中只绘出2根),因此4I4W过程需 要 36 个 SLC. 对于更高精度的输入和权重, 以及 更多的图片像素 (文中不考虑输入端口复用), 需要 更多的 BL 和 SSL, 阵列具有相当的规模. Lue 等^[49] 并未从硬件上实现卷积核功能,他们测试了 64 GB SGVC NAND 中单元的电性能. 最后基于 4 bit 精 度的 3D-NAND 卷积核电路, 运行 VGG-16 卷积 神经网络对 CIFAR-10 数据库进行识别, 识别率达 到 90%.



图 13 采用 SLC 3D-NAND 实现 4 bit 精度的卷积神经网络方案^[40] (a) 卷积神经网络的工作原理示意图(上), 涉及大量的 MAC 过程(下); (b) 对于多 bit 权重的 MAC 过程, 用多个 SLC 器件构成一个多 bit 权重; (c) 4 bit 精度输入与 4 bit 精度权重的 MAC 原理; (d) 卷积核电路的工作原理; (e) VGG 16 神经网络的结构图以及用所设计的 3D-NAND 加速 VGG 16 的性能 Fig. 13. A case of SLC 3D-NAND for convolution neural network (CNN) with 4-bit resolution^[40]: (a) Flow schematic of CNN; (b) in a MAC array, plural SSLs to stand for multi-bit weight; (c) the arithmetic principle of MAC with 4-bit input and 4-bit weight (4I4W); (d) convolutional core circuit and working principle diagram; (e) schematic diagram of VGG 16 CNN and the simulated performance of 3D-NAND hardware implementation.

2019 年, Kim 等^[50] 提出了一种基于 3D-NAND 的卷积核映射方案,用 4 个 MLC (multi-level cell, 2 bit 存储态的 CTL 器件)构成了 8 bit 存储态的 突触,并实现了 8 bit 精度的 CNN 卷积运算,如 图 14 所示.图 14(a)中将 BiCS 结构的 3D-NAND 中的每个 block 按 BL 方向展开成 2D-NAND,每 个 SGD 线 (select gate at drain side, SGD)连接 一个前级神经元接收输入信号.具体的卷积核映射 方法如图 14(b)所示.一个 8 bit 的权重映射到一 个 string 上相邻的 4 个 MLC 中 (P₁₅/N₁₅, P₁₄/N₁₄, P₁₃/N₁₃, P₁₂/N₁₂).两个相邻的 block 中同一个 SGD 线控制的两个 string 上的权重分别标记为 正、负,构成差分对.差分对中权重为正时,负权重 设为 0,反之权重为负时,正权重设为 0. Kim 等^[50] 在 2021 年的报道中增加了对卷积运算具体过程的 阐述,如图 14(c) 所示.其过程为:1)两个输入信号的第一位和权重的前两位进行乘加运算后得到4,下一步两个输入信号的第一位和权重的下两位进行乘加运算得到结果3,两次运算结果进行移位相加,即3×2²+4×2⁰;2)图中有24根SGG线,支持24个神经元信号同时输入进行乘加运算,乘加运算32次后移位相加,得到24个前级神经元对同一个后级神经元的输出.同一个string上有16个 MLC,可存储4个8 bit 权重,图 14(b)的电路可以映射24个前级神经元和4个后级神经元之间全连接的突触.

由于 NAND 厂商禁止开放 WL 和 SGD 端口的 控制, Kim 等^[50] 用自研的 16 层 die 堆叠的 eNAND 等效 3D-NAND(具体的 eNAND 结构和工作原理 不在此赘述). 图 14(d) 中展示了基于 eNAND 的



图 14 基于 MLC 3D-NAND 的 8-bit 精度卷积方案^[50,51] (a) 基于 BiCS 结构的 3D-NAND 电路图^[51]; (b) 权重的映射方式, 正、 负权重存储在相邻的两个 block 中^[51]; (c) 2个 8 bit 精度的输入信号和 2个 8 bit 精度的权重的乘加运算过程^[51]; (d) 基于 eNAND 的卷积核电路, 有 7个 block, 28 个输入端口, 满足 5×5卷积核的功能^[51]; (e) 卷积过程的信号时序图^[51]

Fig. 14. A case of MLC 3D-NAND for CNN with 8-bit resolution^[50,51]: (a) Circuit diagram of BiCS type 3D-NAND, the 3D structure can be flattened into a 2D structure^[51]; (b) weight mapping method, positive and negative weight stored in two neighboring blocks^[51]; (c) MAC operation principle of 2 inputs and two weights with 8-bit resolution^[51]; (d) convolutional core circuit diagram with 7 blocks and 28 input ports can be used for 5×5 convolution operation^[51]; (e) timing diagram of 5×5 convolution operation with 8-bit data and 8-bit weights^[51].

卷积核电路,有7个卷积用的 Block 和一个用于修 复的 Block. 此电路具有28个输入端口,可以映射 5×5大小的卷积核,卷积的脉冲时序如图14(e)所 示. 基于此硬件方案,Kim 等^[50]设计了 LetNet-5 进行 MNIST 手写数字识别,识别率达到98.5%. 2021年, Kim 课题组^[52]提出了一种基于 3D NAND 的卷积核映射方案,用1 bit 的 MAC 算子进行分部相乘 (partial multiplication)并采用 Booth 编码映射正负权重,显著地降低了 VMM 过程中的电流,提高了芯片的能效,文中将这种芯片架构



图 15 基于 3D NAND 的 S-Flash 芯片用于卷积神经网络加速^[52] (a) 卷积过程示意图; (b) MAC 算子的比特对乘、加运算延迟时间的影响(左)和 MAC 算子的比特对累加运算延迟时间的影响(右); (c) S-Flash 芯片的架构; (d) S-Flash 中权重分布的示意图,用 16 个 SLC 构成一个差分结构的突触; (e) 通过 Booth 编码分配权重; (f), (g) 同时操作的 SSL 和 BL 增大 1 倍, MAC 次数缩减了 1/4 Fig. 15. 3D NAND-based CNN accelerator named as S-Flash^[52]: (a) Convolutional operation of CNN; (b) normalized latency for multiplication, accumulation (left) and MAC operation in various multiplication units (right); (c) overall S-FLASH architecture; (d) overall weight data layout, in which a differential synapse constructed with 16 SLC; (e) weight allocation by Booth coding; (f), (g) double the concurrently operated BLs and SSLs, 4 times faster the MAC operation speed.

命名为 S-Flash. 图 15(a) 为卷积运算的过程, 一个 卷积层由 *k* 个尺寸为 *K* × *K* × *N* 大小的卷积核构 成. 输入信号和卷积核的精度通常为 8 bit, 那么卷 积核在图像上进行一次滑动将产生 *k* × *K*² × *N* (8 bit×8 bit) 次 MAC 运算. 如何优化 8 bit×8 bit 计算过程是文献 [52] 研究的重点. 文中将输入信号 和权重拆分为低比特的算子进行分部相乘. 将输入 信号和权重的比特分别记为 (*B*_a, *B*_w), 当 (*B*_a, *B*_w) = (1,1)时,进行 64 次乘法运算.当(B_a, B_w) = (1,2) 或 (2,1) 时, 进行 32 次乘法运算. 图 15(b) 展示了 (B_{a}, B_{w}) 对乘法和积分对电路延时的影响, 随 (B_{a}, B_{w}) B_w) 增大, 乘法运算的延时略有降低, 但电流积分 的延时显著增大. 这是因为 MAC 运算的速率受限 于 ADC 精度, ADC 精度越高, 外围电路的开销、 延时和能耗越大. 采用一般精度的 ADC(如 3-bit/ 4-bit), (B_a, B_w) 越大, 电流积分的次数越多, 电流 积分时间显著延长. 文中采用 $(B_a, B_w) = (1, 1)$ 作为 MAC 运算的基本操作单位.图 15(c) 展示了 S-Flash 的架构, 神经网络的每一层突触映射到沿 WL 的每一层 SLC 中, 输入信号 A_1 — A_n 的 8 位信 号依次施加在 BL上, MAC 电流用 SSL 收集, n为卷积核的通道数,用一个 block 映射一个卷积 核的一个通道, $n \uparrow block$ 构成一个 array 映射一 个卷积核,多个 array 映射一个卷积层. 图 15(d) 展示了卷积核的映射过程,其中突触权重的精度 为8 bit, 16个 SLC 构成一个具有正、负权重的突 触对.图 15(e)展示了通过用 Booth 编码正、负权 重,以增加权重中的"0"位,提高 MAC 过程中的稀 疏性,从而降低 MAC 过程中的总电流.为了进一 步提高计算效率,将 $A_k \times W_k$ 对应的SLC阵列中 BL和SSL增大一倍,SLC单元增大2倍,原先 8 bit × 8 bit 的 MAC 次数从 64 缩减为 16, 如 图 15(f) 和图 15(g) 所示, 将 MAC 过程中具有 4 倍 SLC 规模但权重未经过 Booth 编码的 S-Flash 定 义为 S-Flash^{*}, 具有 4 倍 SLC 规模并且权重经过 Booth 编码的 S-Flash 定义为 S-Flash^{**}.

最后通过电路仿真研究了 S-Flash芯片在运 行卷积神经网络时的性能参数.分别用 S-Flash, S-Flash*和 S-Flash**运行 VGG-16 神经网络,计算了 前向传播中每层神经网络的能效,并以 GPU 方案 时的能效为标准进行归一化,结果如图 16(a) 所示. 从图 16(a) 可以看出, S-Flash, S-Flash*和 S-Flash** 的能效分别是 GPU 的 1.64, 6.43 和 13.49 倍, S-Flash*和 S-Flash**的能效分别是 S-Flash 的 3.9 倍 和 8.2 倍,所以正、负权重的分配对芯片能效的影 响最大. S-Flash**中各电路模块的面积和能耗对比 如图 16(b)所示,可以看到 ADC 面积占芯片面积 的 0.61%,但能耗占 90.67%.图 16(c) 中表格列出 了 S-Flash**和其他芯片的性能参数,包括存储密 度、单位面积的峰值吞吐量和能效,可以看到 S-Flash**在这 3 个方面均有明显优势.

2021年,余诗孟课题组^{53]}提出了一种基于 3D NAND的神经形态芯片架构,用于卷积神经 网络加速.图 17(a)所示为在 3D NAND 的一个



[†] The process node of all platforms is scaled to 22 nm

‡ The MAC operation is normalized to INT8 operation

图 16 S-Flash 电路仿真的结果^[52] (a) S-Flash/S-Flash*/S-Flash**运行 VGG-16 卷积神经网络时的能效; (b) S-Flash**电路模块 的面积对比和运行 VGG-16 时各电路模块的能耗对比; (c) S-Flash**与其他芯片的性能参数对比

Fig. 16. Simulation result of S-Flash^[52]: (a) Energy efficiency evaluation result of each VGG-16 layer accelerated by S-Flash^{*/} S-Flash^{**}; (b) area and energy breakdown of S-FLASH^{**}; (c) comparison with the other platform.

block 中进行 VMM 操作的方法: 沿 WL 方向的每 一层 SLC 器件对应神经网络的每一层突触, 输入 信号 X_N 施加在 BL 上, 权重 W_i 精度为 2 bit, 用 3 个 SLC 代表一个 W_i , 3 根 SSL 用于突触选通. X_N 采用十进制编码, 即精度为 *n* bit 的 X_N , 可用 2^n-1 根 BL 来表示. 用十进制编码的优势在于 block 中所有 string 上的电流可直接相加后进行模 数转换, 而无需移位操作. 图 17(b) 中展示了 2 bit 权重精度的卷积核映射方案, 卷积核尺寸为 $K = K_C \times K_W \times K_H, X_N$ 精度为 *n* bit, 卷积核中 一个单元的权重映射到 3 × (2ⁿ – 1) 个 SLC 中. 图 17(c) 为多个 block 构成的 subarray, 用于映射 一个卷积层. VGG-8 卷积神经网络中, 最大的卷积 层中有 16 个卷积核, W_i 精度为 2 bit, X_N 精度为 8 bit. 用一个 block 映射精度为 2 bit 的卷积核, block 的大小为 521 × 3 × 3 × (2² – 1) × (2² – 1) × 32 WL = 1.27 Mb. 考虑到 VGG-8 中权重的精度 为 8 bit, 采用 4 个 2 bit 精度的 block 通过分部相 乘实现, 那么 1 个 subarray 的大小为 16 × 4 × 1.27 Mb = 81 Mb. 图 17(d) 为芯片架构的示意图, 芯片中有 4 个 tile, 每个 tile 有 4 个 PE(process element), 每个 PE 有 4 个 subarray. VGG-8 神经 网络需要 110 Mb SLC, 所设计的芯片容量完全满 足需求. 用 HSPICE 计算了 3D NAND 运行 VGG-8



图 17 输入信号采用十进制编码的 3D NAND 芯片用于卷积神经网络加速^[53] (a)用 3D NAND 做 VMM 的操作方法; (b)卷积 核映射的方案; (c) subarray 的结构示意图; (d) 芯片的架构示意图; (e)用 3D NAND 芯片运行 VGG-8 神经网络用于 CIFAR-10 图片库识别时,各电路模块的能耗对比; (f) 3D NAND 与其他芯片的性能对比

Fig. 17. A 3D NAND CNN accelerator with decimal input coding^[53] (a) VMM operation method by using 3D NAND; (b) the mapping method of a CNN kernel; (c) designed subarray configuration; (d) hierarchy of the 3D NAND-based neuromorphic chip architecture; (e) energy breakdown of 3D NAND-based chip on VGG-8 network for the CIFAR-10 dataset; (f) comparison with other chips.

网络用于 CIFAR-10 图像识别任务的性能,电路中 各模块的能耗占比如图 17(e) 所示.图 17(f) 列出 了 3D NAND 的性能参数,并与 RRAM 和 SRAM 做比较,可以看到所设计的 3D NAND 在各项性 能指标上均有明显的优势.

2022年, 霍宗亮课题组^[54]利用 3D NAND 中 CTL 的模拟权重调制特性, 设计了两层全连接神 经网络, 用 Winner-takes-all(WTA) 非监督学习算 法实现 ZVN 图像的识别. 图 18(a) 为 3D NAND 的操作示意图, 不同于其他文献, 文中阵列的输入 信号施加在 WL 上, 输入信号为 V_{read} 和 V_{pass}, 分 别代表输入像素点的明暗两种状态,即1和0.突触映射在WL平面的CTL阵列中,采用差分对结构,输出电流在BL端收集并通过TSG (top select gate transistor)控制.图18(b)中为CTL器件的 模拟权重调制特性,可以看到CTL器件的长时程 增强 (long term potentiation, LTP)和长时程抑制 (long term depression, LTD)过程具有良好的 线性度.图18(c)展示了用3DNAND实现非监督 学习的训练过程.首先以相同的概率输入标准的 Z,V,N图像然后以WTA法则进行权重更新. WTA的权重更新法则:以图像"Z"为例,当输入





Fig. 18. A 3D NAND-based WTA neural network for unsupervised learning^[54]: (a) Schematic of the differential pair in 3D NAND flash array; (b) analog weight modulation of measured CTL device; (c) the training procedure of WTA neural network using 3D NAND array; (d) stylized letter clustering results before and after training.

X_i为1时,增强对应的3个权重 G_{i,j=1,2,3},即增 大G⁺_{i,j}值,减小G⁺_{i,j}值.然后将输出最大的后级神 经元标记为 winner,增强连接在 winner 神经元的 突触权重,即增大G⁺_{i,j}值,减小G⁻_{i,j}值,其他突触权 重不变.在训练了标准图像后,训练带有噪声的Z, V,N图像,即有一个像素点反转的图片,以提高神 经网络的容错率.测试时用一组随机图像做推断, 评估神经网络的性能.图18(d)中展示了训练之前 和训练之后,3个神经元的响应情况以及权重的分 部.经过42个训练周期后,3个神经元能准确识 别Z,V,N图像,并且输入像素的权重随训练次数 的增大,往正确的预测方向增大/减小.

上述关于用于前向传播的 3D NAND 均采用 二值或者幅值编码. 2020 年 Lee 等^[55] 设计了一种 基于脉宽编码的操作方案,在 3D-NAND 中实现 了前向传播过程,如图 19 所示.图 19(a) 展示了前 向传播的原理和操作方式:1) 前级神经元发放的 脉冲输入信号采用脉宽编码,通过脉宽调制电路 (pulse width modulation, PWM)发放施加在 SSL, 神经网络中每一层突触映射到 3D-NAND 中每一 层 CTL 器件上,并且采用差分对结构用两个 CTL 器件分别存储正、负权重; 2) SL 上施加驱动电压 V_{BL},选通层的 WL 上加较小的选通电压 V_{read},未 选通层的 WL 上加较大的导通电压 V_{pass}, V_{BL} 经 过选通层的正、负权重转化为电流进入后级神经元 电路; 3) 神经元电路的结构如图 19(b) 所示,输出 的电流经过电流镜构成的差分电路,相减后对电容 进行充电得到后级神经元的电压 V_c,其中 SSL 上 输入信号的脉宽决定了 SSL 上晶体管的开启时间, 即电流的积分时间. Lee 等^[55] 根据 3D-NAND 的 硬件方案设计了 3 层全连接的卷积神经网络,对 CIFAR-10 图片数据库进行识别任务,并且比较了 权重精度为 1 bit 和 4 bit 时网络的性能表现,如 图 19(c) 所示.

2.2.2 3D-NAND 用于反向传播

由于 CTL 器件擦写次数有限,因此 3D-NAND 通常用于权重更新不太频繁的应用场景.通常将已 训练好的权重映射到 3D-NAND 中,执行推断任



图 19 3D-NAND 中采用脉宽编码实现前向传播^[55] (a) 前向传播的操作方式、神经网络示意图和 WL 上读电压的时序图; (b) 左:前向传播的工作原理示示意图,即两个标记为正、负权重的突触对构成一个突触,输出电流相减后通过电容积分转化为 电压 V_c; 右:读电压 V_{read}、SSL 上输入脉冲 V_{SSL}和 V_c的时序图; (c) 3 层全连接 CNN 网络在权重精度为 4 bit(QNN)和 1 bit (BNN)条件下对 CIFAR-10 图片数据库的识别性能

Fig. 19. Forward propagation using 3D-NAND with pulse width modulation (PWM) scheme^[55]: (a) Operation scheme of forward propagation, schematic diagram of neural networks, the timing diagram of pulses applied to WLs; (b) Left: schematic diagram of synaptic string array consisting of synapses with positive weight (G^+) and synapses with negative weight (G^-); Right: timing diagram of V_{read} , V_{SSL} , and V_{c} ; (c) simulated classification accuracy of 4-bit QNN and BNN for CIFAR-10 images.

务, 2.2.1 节中介绍的工作均基于此应用场景. 但在 硬件上直接训练权重, 显然更智能且经济. 除了有 限的器件擦写次数, 在 3D-NAND 中做反向传播 面临两方面的挑战. 第一, 反向传播中突触矩阵和 前向传播中互为转置, 另选硬件资源映射转置后的 突触矩阵做反向传播 (如图 10(c) 所示), 训练时间 代价较高, 也容易出错. 第二, 突触往往采用差分 对结构, NAND 的串行结构天然不适合反向传播. 因此如何设计反向传播的操作方案是关键.

2021年, Lee 等^[56]在 2020年报道的方案上^[55]做出改进,可以在 3D-NAND 同一个 CTL 阵列中 同时实现神经形态计算中的前向传播和反向传播 过程,如图 20 所示.根据神经网络的工作原理,反 向传播过程中误差信号从最后一层神经元输入并 依次向前级神经元传递.对于 crossbar 结构的突触 阵列,反向传播时误差信号从原输出端输入,从矩 阵相乘角度看,前向传播和反向传播过程中突触阵 列的矩阵互为转置的关系,如图 20(a)所示.但如 果硬件采用差分对结构,相邻的器件分别为正、负 权重,前向传播和反向传播时突触的矩阵则为非转 置关系,如图 20(b)所示.为了实现反向传播,Lee 等^[56]将所有突触的正、负权重用两个阵列分开配置,如图 20(c)所示.

Lee 等⁵⁶ 用 3D-NAND 实现前向和反向传播 的原理如图 21 所示. 前向传播过程如图 21(a) 所 示:1) 前级神经元的输入信号用脉宽编码, 通过 PWM 模块发生并施加到 BL 上, SSL 控制 NAND string 的开关,即控制流入到对应的各个后级神经 元的电流. 2) 3D-NAND 中每一层突触器件对应神 经元中的每一层全连接突触,如图 21(c) 所示,信 号前向传播到第 i层时,在 3D-NAND 第 i层的 WL上施加选通电压 V_{read},在其他层施加导通电 压 V_{pass}; 3) 所有前级神经元输入的脉冲信号分别 经过正、负权重,转化为电流在输出神经元电路中 进行差分后积分转化为输出电压.反向传播的过程 如图 21(b) 所示: 1) 反向传播的误差信号同样采用 脉宽编码, 通过 PWM 电路发生并施加在 SSL 上. 由于读操作中 CTL 单元的电流受栅-源电压控制, 即 V_{read} 减去 CTL 单元源端的电位, 而源极的电 位与 pass 状态的 CTL 单元的分布情况有关. 如果 反向传播过程的误差信号施加在 SL 端, 那么 CTL 的源极电位与前向传播时不一致,即 CTL 栅源电



图 20 差分对突触阵列中将正、负权重分开放置可实现反向传播⁵⁶ (a) 前向传播和反向传播过程中对应的突触阵列,从矩阵运算角度上看互为转置结构; (b) 通常情况下,差分对结构中的正、负权重在同一个阵列中,突触阵列与前向传播过程中并非转置的关系,无法实现反向传播功能; (c) 将正、负权重分开置于不同的阵列中,可以实现反向传播过程

Fig. 20. Backward propagation can be implemented using a differential synaptic array where positive and negative weights are separated^[56]: (a) The matrix of synapse weight in forward and backward propagation are transposed; (b) synaptic array architecture consisting of two adjacent cells representing G^+ and G^- , the weights cannot be transposed; (c) synaptic array architecture where G^+ and G^- weights are separated in different arrays.



图 21 基于 3D-NAND 的前向传播和反向传播的工作原理^[56] (a) 前向传播中 3D-NAND 的操作方法; (b) 反向传播中 3D-NAND 的操作方法; (c) 具有 n 层全连接突触的神经网络结构; (d) 前向传播和反向传播中 WL 上选通电压 V_{read} 和导通电压 V_{pass} 的时序

Fig. 21. Forward and backward propagation using 3D-NAND with PWM scheme^[56]: (a) Synaptic array architecture based on NAND flash memory for forwarding propagation operation; (b) synaptic array architecture based on NAND flash memory for backward propagation operation; (c) schematic of neural networks consisting of n weight layers; (d) timing diagram of V_{read} and V_{pass} in forwarding propagation and backward propagation.

压不一致,读电流将会产生较大的误差.因此 Lee 等^[56]将误差信号从 SSL 端输入,既能保证权重的 转置也能避免 pass 单元造成的读误差.2) V_{read} 和 V_{pass} 依次施加各层 WL 上,施加的顺序与前向 传播相反.3) BL 连接前级神经元电路, SL 上施加 驱动电压 V_{BL},每个 NAND string 上的产生电流 在神经元电路中做差分后积分得到输出电压,通过 SSL 上误差信号的脉宽决定了电流在神经元电路中的积分时间得到对应的电压.

CTL 器件通过写脉冲可具有 32 个不同的 V_{th} , 即 5 bit 的权重态,如图 22(a) 所示.图 22(b) 中用 不同的读电压 $V_1 - V_5$ 可得到不同的权重态分布,



图 22 器件特性和神经网络性能⁵⁶ (a)器件的 *I*_{BL}-*V*_{BL}特性随写入脉冲数量的变化; (b)图 (a)中器件的归一化电导随写脉冲数量的变化; (c)导通电压 *V*_{pass}和写电压 *V*_{PGM} 对阈值电压的影响; (d)初始状态的器件和经历过擦写循环的器件对写入脉冲的响应; (e) 三层全连接神经网络的识别率; (f) 隐藏层数量对识别率的影响

Fig. 22. Device characteristics and neural network performance^[56]: (a) I_{BL} - V_{BL} curves with an increasing number of program pulses; (b) normalized conductance responses measured in (a); (c) I_{BL} - V_{WL} curves measured in a fresh, V_{pass} disturbed, and programmed cell; (d) conductance response of fresh and cycled cell; (e) recognition accuracy of 3-layer neural network; (f) recognition accuracy with the number of hidden layers.

	[45]	[47]	[48]	[49]	[50]	[52]	[53]	[54]	[55]	[56]
技术节点	N.A.	N.A.	N.A.	N.A.	$65 \ \mathrm{nm}$	N.A.	32 nm	N.A.	$26 \ \mathrm{nm}$	26 nm
芯片容量	N.A.	N.A.	N.A.	$64~\mathrm{GB}$	N.A.	N.A.	$1.13~\mathrm{GB}$	N.A.	N.A.	N.A.
器件类型	N.A.	SLC	SLC	SLC	SLC	SLC	SLC	Analog	MLC	PLC
输入端口	位线	字线	漏端选通 管的字线	位线	漏端选通 管的字线	位线	位线	字线	位线	位线
输入编码	脉冲幅值	N.A.	二值编码	数字编码	数字编码	数字编码	数字编码	二值编码	脉宽编码	脉宽编码
输入信 号精度	模拟	N.A.	1 bit	4 bit	8 bit	$8 \mathrm{\ bit}$	8 bit	1 bit	模拟	模拟
突触精度	N.A.	1 bit	1 bit	4 bit	8 bit	8 bit	8 bit	N.A.	4 bit/2 bit	6 bit
反向传 播方式	另选转置的 突触阵列, 误差信号在 位线上输入	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A	N.A.	误差信号施 加在源极选 通管位线上
神经网络	3层全连 接网络	N.A.	4层全连网络、 6层卷积 + 3层 全连接	VGG16	LeNet-5	VGG-16	VGG-8	两层全连 接网络	6层卷积 + 3层全连接	2—7层全 连接网络
神经网络的 突触数量	$0.3 \ \mathrm{MB}$	N.A.	N.A.	$138 \mathrm{~MB}$	N.A.	N.A.	$110 \mathrm{MB}$	27	N.A.	N.A.
数据集	MNIST	N.A.	MNIST, CIFAR-10	CIFAR-10	MNIST	N.A.	CIFAR-10	ZVN	CIFAR-10	MNIST
识别率	94.5%	N.A.	98.12%, 87.11%	90%	98.5%	N.A.	N.A.	N.A.	89.38% (4 bit 权重)、87.1% (2 bit权重)	95.65%
能效/ (TOPS·W ⁻¹)	N.A.	N.A.	N.A.	~ 40	N.A.	0.3	12.95	N.A	N.A.	N.A.

表 2	基于 3 D-NAND 的神经形态计算的各项工作比对					
	Table 2. A comparison of revie	wed works.				

其中读电压越大, 权重变化范围越小, 但分布越线性. V_{pass} 对 V_{th} 基本没有影响, 如图 22(c) 所示. 写/擦 脉冲循环对权重态的分布也基本无影响, 如图 22(d) 所示. CLT 器件权重的精度为 5 bit, 差分对构成的 突触精度达到 6 bit. 基于测试得到的几种权重态 分布和单元之间的权重误差分布, Lee 等^[56] 设计了 全连接神经网络用于 MNIST 图片识别任务, 网络 的性能表现如图 22(e) 所示. 其中权重态分布越线 性, 网络识别率越高. 将网络中隐藏层数量增加到 3 层, 可以略微提高网络识别率, 如图 22(f) 所示.

2.2.3 各项工作的比对

基于 3D-NAND 的神经形态计算的各项工作 比对如表 2 所列.

3 总结与展望

过去几年,具有存算一体特性的 AI 芯片不断 涌现,工艺节点涵盖了 14—180 nm,计算架构包括 了近存计算、存内计算和神经形态计算,应用场景 覆盖了边缘端到云端设备.在各种硬件方案中,基 于 3D-NAND 的神经形态芯片在芯片容量,CMOS 工艺兼容性和成本方面极具优势.本文首先介绍 了 3D-NAND 的基本结构和原理,以及用于神经 形态计算的优势和不足.然后详细梳理了近几年关 于 NAND 和 3D-NAND 用于神经形态计算的代表 性工作,重点介绍了其中的编码方式、前向传播原 理和反向传播过程.

基于现有的工作,考虑到 3D-NAND 的优势 与不足,如用于未来的神经形态计算, 3D-NAND 需要做的调整如下:

1)器件层面.用于数据存储的 3D NAND,器 件采用电荷俘获型晶体管 (CTL),通过在栅极施加 高幅值和长时程的脉宽 (>10 V, >100 μs),利用 Fowler-Nordheim 隧穿效应,在电荷俘获层中注入 或擦除电子以改变阈值电压 (V_{th}),实现存储功能. 随擦写次数的增加,隧穿绝缘层的晶格会被破坏 甚至失效,因此 CTL 的擦写次数有限.低功耗是 神经形态计算的特点,CTL 器件的操作功耗需 要进一步优化.目前国内外的一些研究机构,探索 了将氧化铪基铁电材料替代传统的氮化硅电荷俘 获层^[57,58],利用铁电效应实现了器件的存储功能. 如果能将铁电技术成功地应用到 3D NAND 中, 能大幅提高器件的擦写次数,并且降低操作功耗.

2) 结构层面. 1) CTL 晶体管是 3D NAND 的 基本单元,多个 CTL 器件组成一个 NAND string, 多个 string 组成一个 block, 多个 block 组成 3D NAND 结构. 在神经形态计算中, 突触和神经元是 神经网络的基本单元.2)突触可由一个或多个 CTL 器件构成. 对于低精度的计算, 可采用幅值或 者脉宽编码, 输入/输出均为模拟信号, 单个 CTL 突触即可满足模拟计算的需求,电路结构简单原理 直观. 对于高精度的计算,则采用二值编码,用多 个 SLC 构建一个多 bit 精度的突触, 采用二进制 计算方式. 3) 突触多采用差分对结构G=G+-G-, 为了避免正、负突触阈值电压达到最大而无法进一 步更新权重, 3D NAND 中通常需要定期进行块擦 除并重新赋予突触权重值. 2021年,首尔大学和 SK Hynix 合作开发了适用于神经形态计算的单个 CTL 器件的擦除方案, 避免了定期的块擦除⁵⁹.

3) 架构层面.存储用途的 3D NAND 只涉及 读、写、擦操作,计算由外部的 CPU 负责.读写按 block→string→CTL 的顺序串行操作.区别于存 储用途,在用于神经形态计算的 3D NAND 中,读 操作增加了 MAC 运算,外围电路需要配置大量 的 ADC/DAC 和移位加法器等单元.并且读写操 作按神经网络的映射规则执行,不一定按 block→ string→CTL 的顺序.

最后,由于 3D-NAND 的专利特性,厂商并未 开放用户对芯片颗粒端口的权限.目前的工作中, 前向传播过程和反向传播过程并未做硬件实现,多 数是基于厂商样片测得的存储单元特性以及读误 差分布,通过电路和软件层面上仿真得到的结果. 未来的工作应该考虑与厂商有更深入的交流合作, 在硬件层面执行前向传播、反向传播和权重更新, 更直接地展示 3D-NAND 在神经形态计算方面的 应用潜力.

参考文献

- Amodei D, Hernandez D, Sastry G, Clark J, Brockman G, Sutskever I https://openai.com/blog/ai-and-compute/ [2022-4-11]
- [2] Patterson D A, Hennessy J L 2021 Computer Organization and Design RISC-V Edition: the Hardware Software Interface (6th Ed.) (Amsterdam: Morgan Kaufmann) p44
- [3] Gai S https://pensando.io/dennard-scaling-and-other-powerconsiderations/ [2022-4-11]
- [4] Dally B https://www.cs.colostate.edu/~cs575dl/Sp2015/Lec

tures/Dally2015.pdf [2022-4-11]

- [5] Drachman D A 2005 Neurology **64** 2004
- [6] Zhang W, Gao B, Tang J, Yao P, Yu S, Chang M F, Yoo H J, Qian H, Wu H 2020 Nat. Electron. 3 371
- [7] Roy K, Jaiswal A, Panda P 2019 Nature 575 607
- [8] Khaddam-Aljameh R, Stanisavljevic M, Mas J F, et al. 2021 2021 Symposium on VLSI Technology Kyoto, Japan, June 13-19, 2021 p1
- [9] Narayanan P, Ambrogio S, Okazaki A, et al. 2021 2021 Symposium on VLSI Technology Kyoto, Japan, June 13–19, 2021 p1
- [10] Yang J, Xue X, Xu X, Lv H, Zhang F, Zeng X, Chang M F, Liu M 2020 2020 IEEE Symposium on VLSI Circuits Honolulu, HI, USA, June 16–19, 2020 p1
- [11] Chih Y D, Shih Y C, Lee C F, et al.2020 2020 IEEE International Solid-State Circuits Conference San Francisco, CA, USA, Feb 16–20, 2020 p222
- [12] Liu Y, Su F, Yang Y, Wang Z, Wang Y, Li Z, Li X, Yoshimura R, Naiki T, Tsuwa T, Saito T, Wang Z, Taniuchi K, Yang H 2019 *IEEE J. Solid-State Circuits* **54** 885
- [13] Dünkel S, Trentzsch M, Richter R, et al. 2017 2017 IEEE International Electron Devices Meeting San Francisco, California, USA, Dec 2–6, 2017, p19.17. 11
- [14] IRDSTM 2021 International Roadmap for Devices and Systems (2021 Ed.) (IEEE) from https://irds.ieee.org/images/files/ pdf/2021/2021IRDS_MM_Tables.xlsx [2022-4-11]
- [15] Liu T Y, Yan T H, Scheuerlein R, et al. 2013 2013 IEEE International Solid-State Circuits Conference San Francisco, CA, USA, Feb. 17–21, 2013 p210
- [16] Chen Y, Li H, Wang X, Zhu W, Xu W, Zhang T 2012 IEEE J. Solid-State Circuits 47 560
- [17] Rho K, Tsuchida K, Kim D, et al. 2017 2017 IEEE International Solid-State Circuits Conference San Franciso, CA, USA, Feb. 5–9, 2017 p396
- [18] Zwerg M, Baumann A, Kuhn R, et al. 2011 2011 IEEE International Solid-State Circuits Conference San Francisco, CA, USA, Feb. 20–24, 2011 p334
- [19] Takashima D, Nagadomi Y, Ozaki T 2011 IEEE J. Solid-State Circuits 46 681
- [20] Trentzsch M, Flachowsky S, Richter R, et al. 2016 2016 IEEE International Electron Devices Meeting San Francisco, CA, USA, December 3–7, 2016 p11.15.11
- [21] Lee J W, Na D, Kavala A, et al. 2020 2020 IEEE Symposium on VLSI Circuits Honolulu, HI, United States, June 16–19, 2020 p1
- [22] Mulaosmanovic H, Breyer E T, Dünkel S, Beyer S, Mikolajick T, Slesazeck S 2021 Nanotechnology 32 502002
- [23] Noguchi H, Ikegami K, Kushida K, et al. 2015 2015 IEEE International Solid-State Circuits Conference San Franciso, CA, USA, February 22–26, 2015 p1
- [24] Sato H, Honjo H, Watanabe T, et al. 2018 2018 IEEE International Electron Devices Meeting San Francisco, CA, USA, December 1–5, 2018 p27.22.21
- [25] Khan A I, Keshavarzi A, Datta S 2020 Nat. Electron. 3 588
- [26] Dong Q, Kim Y, Lee I, et al. 2017 2017 IEEE International Solid-State Circuits Conference San Francisco, CA, USA, February 5–9 2017 p198
- [27] Cheong W, Yoon C, Woo S, et al. 2018 2018 IEEE International Solid-State Circuits Conference San Francisco, California, USA, February 11–15, 2018 p338
- [28] Liang J, Jeyasingh R G D, Chen H, Wong H P 2011 2011 Symposium on VLSI Technology Kyoto, Japan, June 14–16, 2011 p100

- [29] Wu T F, Le B Q, Radway R, et al. 2019 2019 IEEE International Solid-State Circuits Conference San Francisco, CA, USA, February 17–21, 2019 p226
- [30] Dong Q, Wang Z, Lim J, Zhang Y, Sinangil M E, Shih Y C, Chih Y D, Chang J, Blaauw D, Sylvester D 2019 IEEE J. Solid-State Circuits 54 231
- [31] Yoon S, Youn Y, Kim S 2015 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems New York, NY, USA, August 24–26, 2015 p1045
- [32] Sandre G D, Bettini L, Pirola A, et al. 2010 2010 IEEE International Solid-State Circuits Conference San Francisco, California, USA, February 7–11, 2010 p268
- [33] Lo C P, Lin W Z, Lin W Y, Lin H T, Yang T H, Chiang Y N, King Y C, Lin C J, Chih Y D, Chang T Y J, Chang M F 2019 IEEE J. Solid-State Circuits 54 584
- [34] Xue X Y, Jian W X, Yang J G, Xiao F J, Chen G, Xu X L, Xie Y F, Lin Y Y, Huang R, Zhou Q T, Wu J G 2012 2012 Symposium on VLSI Circuits Honolulu, HI, USA, June 13–15, 2012 p42
- [35] Chang M, Shen S, Liu C, Wu C, Lin Y, King Y, Lin C, Liao H, Chih Y, Yamauchi H 2013 *IEEE J. Solid-State Circuits* 48 864
- [36] Chen J, Chiang R C, Huang H H, Venkataramani G 2012 SIGOPS Oper. Syst. Rev. 45 48
- [37] Qazi M, Clinton M, Bartling S, Chandrakasan A P 2012 IEEE J. Solid-State Circuits 47 141
- [38] Breyer E T, Mulaosmanovic H, Trommer J, Melde T, Dünkel S, Trentzsch M, Beyer S, Slesazeck S, Mikolajick T 2020 *IEEE J. Electron Devices Soc.* 8 748
- [39] Wang Z, Wu H, Burr G W, Hwang C S, Wang K L, Xia Q, Yang J J 2020 Nat. Rev. Mater. 5 173
- [40] Lee G H, Hwang S, Yu J, Kim H 2021 Appl. Sci. 11 6703
- [41] Jang J, Kim H S, Cho W, et al.2009 2009 Symposium on VLSI Technology Kyoto, Japan, June 15–17, 2009 p192
- [42] Wonjoo K, Sangmoo C, Junghun S, Taehee L, Park C, Hyoungsoo K, Juhwan J, Inkyong Y, Park Y 2009 2009 Symposium on VLSI Technology Kyoto, Japan, June 15–17, 2009 p188
- [43] Micheloni R 2016 3D Flash Memories (Dordrecht: Springer Netherlands) p89
- [44] Seo Y T, Kwon D, Noh Y, Lee S, Park M K, Woo S Y, Park B G, Lee J H 2021 *IEEE Trans. Electron Devices* 68 3801
- [45] Lee S T, Lim S, Choi N, Bae J H, Kim C H, Lee S, Lee D H, Lee T, Chung S, Park B G, Lee J H 2008 2018 IEEE Symposium on VLSI Technology Honolulu, HI, USA, June 18–22, 2018 p169
- [46] Lee S T, Lim S, Choi N Y, Bae J H, Kwon D, Park B G, Lee J H 2019 IEEE J. Electron Devices Soc. 7 1085
- [47] Wang P, Xu F, Wang B, Gao B, Wu H, Qian H, Yu S 2019 IEEE Trans. Very Large Scale Integr. VLSI Syst. 27 988
- [48] Lee S T, Kim H, Bae J H, Yoo H, Choi N Y, Kwon D, Lim S, Park B G, Lee J H 2019 2019 IEEE International Electron Devices Meeting San Francisco, CA, USA, December 7–11, 2019 p38.34.31
- [49] Lue H T, Hsu P K, Wei M L, Yeh T H, Du P Y, Chen W C, Wang K C, Lu C Y 2019 2019 IEEE International Electron Devices Meeting San Francisco, CA, USA, December 7–11, 2019 p38.31.31
- [50] Kim M, Liu M, Everson L, Park G, Jeon Y, Kim S, Lee S, Song S, Kim C H 2019 2019 IEEE International Electron

 $Devices\ Meeting\ San\ Francisco,\ CA,\ USA,\ December\ 7-11,\ 2019\ p38.33.31$

- [51] Kim M, Liu M, Everson L R, Kim C H 2022 IEEE J. Solid-State Circuits 57 625
- [52] Kang M, Kim H, Shin H, Sim J, Kim K, Kim L S 2022 IEEE Trans. Comput. 71 1291
- [53] Hsu P K, Du P Y, Lo C R, Lue H T, Chen W C, Hsu T H, Yeh T H, Hsieh C C, Wei M L, Wang K C, Lu C Y 2020 2020 IEEE International Memory Workshop Dresden, Germany, May 17–20, 2020 p1
- [54]~ Zhou W, Jin L, Jia X, Wang T, Xu P, Zhang A, Huo Z2022

IEEE Electron Device Lett. 43 374

- [55] Lee S T, Lee J H 2020 Front. Neurosci. 14 517292
- [56] Lee S T, Yeom G, Yoo H, Kim H S, Lim S, Bae J H, Park B G, Lee J H 2021 *IEEE Trans. Electron Devices* 68 3365
- [57] Kim M K, Kim I J, Lee J S 2021 Sci. Adv. 7 1341
- [58] Yoon S, Hong S I, Choi G, Kim D, Kim I, Jeon S M, Kim C, Min K 2022 2022 IEEE International Memory Workshop Dresden, Germany, May 15–18, 2022 p 1
- [59] Yoo H N, Back J W, Kim N H, Kwon D, Park B G, Lee J H 2022 2022 IEEE Symposium on VLSI Technology and Circuits Honolulu, HI, USA, June 12–17, 2022 p304

SPECIAL TOPIC—Physical electronics for brain-inspired computing 3D-NAND flash memory based neuromorphic computing

Chen Yang-Yang¹⁾²⁾³⁾ He Yu-Hui^{<math>3)4)}</sup></sup>

Miao Xiang-Shui³⁾⁴⁾ Yang Dao-Hong^{1)2)3)†}

1) (Post-doctoral Mobile Station, Huazhong University of Science and Technology, Wuhan 430074, China)

2) (Post-doctoral Work Station, Wuhan Xinxin Semiconductor Manufacturing Co., Ltd., Wuhan 430205, China)

3) (Hubei Yangtze Memory Laboratories, Wuhan 430205, China)

4) (School of Integrated Circuit, Huazhong University of Science and Technology, Wuhan 430074, China)

(Received 16 May 2022; revised manuscript received 27 September 2022)

Abstract

A neuromorphic chip is an emerging AI chip. The neuromorphic chip is based on non-Von Neumann architecture, and it simulates the structure and working principle of the human brain. Compared with non-Von Neumann architecture AI chips, the neuromorphic chips have significant improvement of efficiency and energy consumption advantages. The 3D-NAND flash memory has the merits of a mature process and ultra-high storage density, and recently it attracted many researchers' attention. However, owing to the proprietary nature of the technology, there are few hardware implementations. This paper reviews the present research status of neuromorphic computing by using the 3D-NAND flash memory, introduces the forward propagation and backward propagation schemes, and proposes several improvements on the device, structure, and architecture of 3D NAND for neuromorphic computing.

Keywords: neuromorphic computing, 3D-NAND, in-memory computing architecture

PACS: 07.05.Mh, 85.35.-p, 84.30.-r, 87.18.Sn

DOI: 10.7498/aps.71.20220974

[†] Corresponding author. E-mail: alan_yang@xmcwh.com





Institute of Physics, CAS

基于3D-NAND的神经形态计算

陈阳洋 何毓辉 缪向水 杨道虹

3D-NAND flash memory based neuromorphic computing Chen Yang-Yang He Yu-Hui Miao Xiang-Shui Yang Dao-Hong 引用信息 Citation: Acta Physica Sinica, 71, 210702 (2022) DOI: 10.7498/aps.71.20220974 在线阅读 View online: https://doi.org/10.7498/aps.71.20220974 当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

面向感存算一体化的光电忆阻器件研究进展

Recent progress in optoelectronic memristive devices for in-sensor computing 物理学报. 2022, 71(14): 148701 https://doi.org/10.7498/aps.71.20220350

光电神经形态器件及其应用

Optoelectronic neuromorphic devices and their applications 物理学报. 2022, 71(14): 148505 https://doi.org/10.7498/aps.71.20220111

忆阻类脑计算

Memristive brain-like computing 物理学报. 2022, 71(14): 140501 https://doi.org/10.7498/aps.71.20220666

仿生生物感官的感存算一体化系统

Bio-inspired sensory systems with integrated capabilities of sensing, data storage, and processing 物理学报. 2022, 71(14): 148702 https://doi.org/10.7498/aps.71.20220281

基于非挥发存储器的存内计算技术

Non-volatile memory based in-memory computing technology 物理学报. 2022, 71(14): 148507 https://doi.org/10.7498/aps.71.20220397

基于忆容器件的神经形态计算研究进展

Research progress of neuromorphic computation based on memcapacitors 物理学报. 2021, 70(7): 078701 https://doi.org/10.7498/aps.70.20201632

专题: 面向类脑计算的物理电子学

基于二维层状材料的神经形态器件研究进展*

李策 杨栋梁† 孙林锋‡

(北京理工大学物理学院,先进光电量子结构设计与测量教育部重点实验室,北京 100081)

(2022年7月16日收到; 2022年8月13日收到修改稿)

近年来,人工智能的发展对计算和存储的需求不断提升.但是,摩尔定律的放缓以及传统冯·诺依曼架构 中计算与存储单元的分离,导致了大量数据在搬运过程中功耗增加和时间延迟,致使集成电路以及芯片设计 面临越来越多的挑战.这迫切需要开发新型计算范式来应对这种挑战.而基于存算一体架构的神经形态器件, 可利用欧姆定律和基尔霍夫定律实现原位计算,从而有望克服传统冯·诺依曼架构瓶颈.通过调节具有"记忆" 功能的忆阻器阻值,实现类似生物大脑的人工神经网络,并对复杂网络信号进行处理,例如图像识别、模式分 类和决策执行等.二维材料由于其层状超薄特性和新奇的物理效应,为进一步缩小器件尺寸并实现感存算一 体提供了方案.本文综述了基于二维材料的神经形态器件中的物理效应和忆阻特性,并详细阐述了神经形态 器件对 LIF (leaky integrate and fire)模型、Hodgkin-Huxley模型等神经元模型以及长期可塑性、短期可塑 性、放电时间依赖可塑性和尖峰频率依赖可塑性的模拟.在此基础上,进一步介绍了基于二维材料的神经形 态器件在视觉、听觉以及触觉等领域的探索性应用.最后本文总结了当前研究领域面临的问题以及对未来应 用前景的展望.

关键词:神经形态器件,人工神经网络,忆阻器,二维材料 PACS: 85.35.-p,73.40.Rw,85.25.Hv,72.20.-i

DOI: 10.7498/aps.71.20221424

1 引 言

随着人工智能的发展以及大数据时代的到来, 社会发展对计算机的运算速度以及处理能效的要 求越来越高.而目前的计算机都是基于传统的 冯·诺依曼架构,这种物理上计算单元和存储单元 的分离,会导致数据在搬运过程中功耗增加和时间 延迟^[1,2].在著名的人机大战中,代表着传统冯·诺 依曼架构的人工智能产物 AlphGo 击败了世界冠 军柯洁,这显示了 AlphGo 优越的算法架构和运算 处理能力.但值得注意的是,AlphGo 的功耗是我 们生物大脑 (约 20 W)的数千倍^[3,4].因此,制造出 像人脑一样,具有存算一体、高效、精准和连续信 号处理等优势的类脑芯片成为近年来学术界和产 业界越来越关注的方向,各个国家也出台了相应的 脑计划,并且中国在"十四五"规划和 2035 年远景 目标中将"脑科学"列为国家重点前沿科技项目^⑤.

受人脑中神经元和突触计算模式的启发,神经 形态计算网络具有实现大规模并行计算、自适应以 及自学习等功能,并可通过欧姆定律和基尔霍夫电 流定律实现原位计算,这样构成的神经形态器件被 认为是最有潜力克服传统冯·诺依曼架构的电子元 件之一^[6-11].在众多用于实现神经形态计算的电子 元器件中,忆阻器凭借其高集成密度、超低功耗、 可实现存储与计算相融合以及模拟突触可塑性等 特点成为当前最具优势的选择.在1971年,Chua^[12] 从电路的完备性关系出发,首次提出了忆阻器的

* 北京市自然科学基金重点研究专题项目(批准号: Z210006)和国家自然科学基金青年科学基金(批准号: 12104051)资助的课题.

© 2022 中国物理学会 Chinese Physical Society

[†] 通信作者. E-mail: yangdl@bit.edu.cn

[‡] 通信作者. E-mail: sunlinfeng@bit.edu.cn

概念.根据忆阻器完成存内计算的方式不同,可以 把忆阻器分为数字型(突变型)忆阻器和模拟型 (渐变型)忆阻器^[13-17].其中,数字型忆阻器的高阻 态(HRS)以及低阻态(LRS)可以分别对应逻辑存 储单元中的"0"和"1",从而根据 HRS 和 LRS 之间 的转变实现数据的存储,该类型忆阻器通常用于实 现非易失性存储^[18].而模拟型忆阻器的阻值可随 电压扫描呈现连续变化特征,通常用于对信号进行 传输、变换和放大等处理^[19,20],并可进行编程模拟 电路设计以及仿真模拟类脑神经突触的权重调 节^[21].此外,利用模拟型忆阻器的交叉阵列可以一 步完成乘法-加法的矩阵运算,这比传统的计算过 程更加节时、节能.

忆阻器发展至今,国内外研究者发现了许多具 有忆阻特性的材料.传统阻变材料主要包括二元金 属氧化物^[22,23]、多元金属氧化物^[24-26]、固态电解 质^[27]和有机介质材料^[28,29]等.由传统阻变材料制 备的忆阻器已被广泛研究,器件性能稳定,制备工 艺成熟,但由于材料本身的限制,使得忆阻器尺寸 很难继续缩小.而二维(2D)材料由于其层状超薄 特性和丰富的物理效应受到了国内外众多研究人 员的关注^[30-38].

本文首先从经典理论和量子理论角度出发,介 绍了基于 2D 材料的神经形态器件的物理机制和 忆阻特性,并详细阐述了其对神经元模型和突触可 塑性的模拟.另外,本文列举了基于 2D 材料神经 形态器件在生物感知、声音定位以及触控模拟等领 域的应用.最后,总结了当前 2D 材料神经形态器 件在高密度、大规模阵列集成等方面存在的问题, 并对如何构建低能耗、存算一体的神经形态计算系 统进行了展望.

2 2D 材料忆阻器件中的物理效应

推动神经形态器件发展的关键在于理解和掌握器件的物理效应,而忆阻器作为实现神经形态计算的理想器件之一,研究其背后的物理机制变得尤为重要. 忆阻器一般是基于 HRS 和 LRS 的相互转化从而实现 SET 和 RESET,其开关机制与器件结构、忆阻材料、电极材料以及工艺流程有关.而对于 2D 材料制备而成的忆阻器来说,在纳米尺度下材料自身会伴随产生新奇的物理效应,这势必也会给器件带来一些奇异的微观特性. 随着研究的进

展, 2D 材料忆阻器的物理效应可以分为: 经典物 理效应以及量子物理效应. 经典物理效应主要为: 相变效应^[39-41] 以及铁电效应^[42-44]. 量子物理效应 可以分为: 导电细丝效应^[45-48]、氧空位效应^[49,50]、 隧穿效应^[51,52]、电荷捕获与释放^[53,54] 以及范德瓦 耳斯 (vdWs) 效应^[55-58], 如图 1 所示. 尽管在忆阻 器实际工作时, 往往会伴随多种效应协同进行, 为 了便于阐述, 接下来将分类单独介绍各个效应的物 理机理.



图 1 基于 2D 材料的忆阻器工作机制示意图,其中包括 经典效应,如相变效应、铁电效应,以及量子效应,如导电 细丝效应、氧空位效应、电荷捕获效应、隧穿效应、vdWs 效应等

Fig. 1. Schematic diagram of physical working mechanisms of memristor based on 2D materials, including classical effects, such as phase change effect, ferroelectric effect, and quantum effect, like conductive filament effect, oxygen vacancy effect, charge trapping effect, tunneling effect, vdWs effect, etc.

2.1 经典物理效应

2.1.1 相变效应

相变效应主要是依托相变材料在不同相态之间的可逆变换来调节忆阻器开和关.当相变材料受 到焦耳热^[59]、激光^[60]、电子束辐照^[61]以及离子掺 杂^[62]等因素影响时,材料的晶态就会发生可逆改 变,这种结构变化往往能在几纳秒内就可以完 成^[63].在相变材料发生结构可逆变换的同时,电导 率也会显著变化,电导率的强大反差可以作为数据 存储的"0"态和"1"态.

由于 Li⁺具有较小的离子半径,可以在 2D 材 料中进行层间运动,这使其可以通过离子掺杂 实现 2D 材料的相变. Zhu 等^[40]通过电场控制 Li⁺ 迁移,实现忆阻器中 MoS₂薄膜从 2*H*相到 1*T*相 的可逆转变,且器件具有良好的突触可塑性,并在 输入信号诱导下可产生突触竞争或突触合作等相 互作用.图 2(a)展示了由金电极和机械剥离的 MoS₂ 组成的器件结构示意图,当在 A 电极施加正向电 压的时候,Li⁺在电场驱动下会向 B 电极运动并积 累,这会导致 Li_xMoS₂转换为 1*T*相,电导率大大 增加,实现 SET 过程.反之,当在 A 电极施加负向 电压时,Li⁺在电场驱动下会向 A 极运动并积累, 导致 Li_xMoS₂转换为 2*H*相,实现 RESET 过程, 如图 2(b)所示.在实现开关特性的基础上,通过施 加脉冲电压,器件表现出了稳定的电导可调行为, 如图 2(c) 所示,这凸显了离子插层在实现 2D 材料 可逆相变以及模拟人工突触功能方面的潜力.此 外, Zhang 等^[41]利用 MoTe₂和 Mo_{1-x}W_xTe₂作为 相变记忆层,制作了垂直结构忆阻器.较为独特的 是,此工作通过横截面高角度环形暗场扫描透射电 子显微镜 (HAADF-STEM) 扫描,发现一种介于 2H与1T相态的2H_d相态,如图2(d)所示.为进 一步探究其电学特性,在-3 V→3 V→-3 V 的扫描 电压下,可在10 ns 内实现 HRS 和 LRS 之间的重 复切换,展现了良好的开关特性(图2(e)).该工作 证明2D 材料中的过渡金属二硫化物(TMD)只在 电场调控下也可实现可逆相变过程,使得具有高开 关比、超快开关速度的2D 材料相变忆阻器件的实 现成为可能.



图 2 相变效应 (a) Au/MoS₂/Au 器件结构示意图以及在电场作用下 Li+调控 MoS₂发生可逆相变过程的示意图^[40]; (b) Au/MoS₂/ Au 器件的 *I-V*特性曲线^[40]; (c) 通过脉冲编程电压改变电导增量^[40]; (d) MoTe₂在电场作用下的相变过程以及 MoTe₂在 2*H* (左下)和 2*H*_d (右下)态下的 STEM 图像^[41]; (e) 电形成过程前后的 *I-V* 扫描曲线^[41]

Fig. 2. Phase change effect: (a) Schematic diagram of Au/MoS₂/Au device structure and Li⁺ regulating the reversible phase change process of MoS₂ under the applied electric field^[40]; (b) typical *I-V* curve of Au/MoS₂/Au device^[40]; (c) conductance changing with continuous pulse programming voltage^[40]; (d) phase change process of MoTe₂ with electric field applied and the STEM images of 2H (bottom left) and $2H_d$ (bottom right) states of MoTe₂^[41]; (e) *I-V* curves of devices before and after forming processes^[41].

2.1.2 铁电效应

铁电忆阻材料在近些年来得到了广泛的研究^[42,64-67]. 在没有外加电场时,铁电材料具有稳定的自发极化性质,但极化取向并不规则,总体表现为无极化. 当施加外电场的时候,自发极化的方向可以由外加电场进行调控^[68,69],进而实现 LRS 和 HRS 之间的阻态转换.

目前, 对铁电场效应管 (FeFET)^[70,71] 以及铁 电隧道结 (FTJ)^[15] 的研究较为成熟. 然而传统的 FeFET 由金属-铁电-绝缘体-半导体组成, 这种组 合存在栅极漏电流等问题, 严重阻碍了实际应用. 基于此, Wang 等^[43] 利用 α-In₂Se₃ 的铁电性和半导 体特性制备了铁电半导体场效应晶体管 (FeSFET), 如图 3(a) 所示. 由于该器件极化转换过程发生在 α-In₂Se₃ 层而不是栅极介质内, 它可以潜在地解决 传统 FeFET 器件结构中漏电流问题. 为了探究其 忆阻机制, 图 3(b) 从铁电效应以及能带结构角度 出发进行了研究, 当背栅 (BG) 施加负向电压时, 其极化方向受电场影响在 α-In₂Se₃ 与 Al₂O₃ 的界 面附近产生向下的极化. 而在 α-In₂Se₃ 的下表面产 生一些正极化电荷,导致能带向下弯曲,进而使得 在沟道底表面出现电荷积累,载流子密度和沟道 电流大幅度增加,从而实现 SET 过程.反之,当在 BG 处施加正向电压时,实现 RESET.研究还发 现,栅极电压 (Vg)能够很好地调控忆阻窗口的大 小,如图 3(c)所示,随着 Vg 的增加,滞回曲线窗口 变大,这表明可以通过调整电场来精确控制极化过 程.同时,该器件表现出稳定的增强和抑制行为, 可实现 100 个可区分的电导状态,如图 3(d)所示. 这为基于 α-In₂Se₃ 的 FeSFET 在类脑智能系统中 开发神经形态器件提供了可能性.

此外,有研究表明在模拟生物突触功能上,二 阶忆阻器具有更大的优越性,有望进一步实现小型 化、高密度、低功耗的神经形态计算.Wang等^[44], 基于 2D 材料的 SnSe 铁电薄膜实现了二阶忆阻器. 通过逐渐改变 SnSe 铁电层的极化程度,实现了 Au/SnSe/NSTO(0.7%Nb 掺杂的 SrTiO₃)器件结 构电导的连续可调,展现了超低的功耗 (66 fJ).尽 管 FeFET 器件在实现特定的突触功能方面已经十 分成熟,但基于 2D 材料的 FeFET 在单个器件中



图 3 铁电效应 (a) 基于 α-In₂Se₃ 的 FeSFET 器件示意图^[43]; (b) α-In₂Se₃ 沟道材料向上和向下极化时的状态图示, 以及相应的 能带图^[43]; (c) 不同 V_g扫描下器件传输特性曲线^[43]; (d) 器件在 5 个连续的周期脉冲电压下稳定突触后电流 (PSC) 的增强、抑制 效果图^[43]

Fig. 3. Ferroelectric effect: (a) Schematic of the α -In₂Se₃ based FeSFET^[43]; (b) illustrations of the upward and downward polarized states of α -In₂Se₃ channel material and the corresponding energy band diagram^[44]; (c) device transfer characteristic curves under different scanning ranges of $V_{\rm g}^{[43]}$; (d) the potentiation and depression process of the post-synaptic-current (PSC) under 5 continuous periodic voltage pulses^[43].

同时实现逻辑计算和突触功能模拟却十分罕见. 基于此, Lou 等^[72]利用 MoS₂和 MoTe₂制备了双 栅 FeFET.由于铁电耦合效应的存在,使得器件表 现出静电多样化行为,从而成功实现了非易失性逻 辑门和人工突触等功能,同时也解决了数字和模拟 空间中的存内计算的问题.

2.2 量子物理效应

2.2.1 导电细丝

导电细丝效应是在原子尺度下,通过外电场调 控原子或离子的运动,从而形成纳米尺寸的导电细 丝,并会伴随量子效应,通常表现为器件会产生以 $G_0 = 2e^2/h$ 为单位的量子电导行为^[73].一般而言, 基于导电细丝效应工作的神经形态器件的电极材 料,一端为化学性质活泼的金属材料 (Ag, Cu, Ru), 另一端为惰性金属材料 (Pt, Pd, Ta, W)^[74,75].在 电场控制下,活性金属原子发生氧化还原反应在介 质层中形成导电细丝,通过改变电压的偏置来调控 细丝的形成与断裂,进而实现电阻开关.然而,随 着器件尺寸的不断缩小,虽然实现了更高的功能密 度、更低的编程电压,但可靠的电阻开关功能的维 持变得越来越困难^[76,77].

基于上述困难, Guo 等^[78] 采用 vdWs 与金属

电极集成的方法,利用 2D SnSe 构建了结构为 Ag/ $SnO_r/SnSe$ 的忆阻器,实现了超低工作电压 (0.4 V)、 高开关比 (>10³), 并具有优良的保持性和耐久性 等优势.其中,传统的热蒸镀和电子束蒸镀通常涉 及高能热原子团的轰击,可能导致电极和 2D 材料 之间的界面发生较大损坏,从而使忆阻器开关性能 恶化^[79,80]. 而 vdWs 金属集成法, 通过弱 vdWs 力集 成金属电极,可以实现电极和 2D 材料之间最小的 界面损伤,确保忆阻器具有可靠的电阻开关行为[81,82]. 器件结构以及界面的横截面 TEM 图像, 如图 4(a) 所示.为了进一步了解 Ag/SnSe 界面的化学性质, Guo 等^[78] 研究了界面的能量色散光谱 (EDS). 结 果显示,在没有转移 Ag 之前, SnSe 表面有 Se, Sn 和少部分O元素.其中O元素的存在,是由于SnSe 表面发生氧化形成了 SnO_x 但当转移 Ag 电极后, SnSe 表面存在了 Ag 元素, 这是由于 Ag 被 SnSe 表面氧化物诱导形成 Ag+导致的. 此外, 器件在连 续 4000 个周期的开关循环中保持 10³ 的开关比, 显示其具有一定的耐久性.并且在 10⁵ s 后仍然保 持较高的开关比,如图 4(b) 所示.其工作机制如 图 4(c) 所示, 当 Ag 电极接地, 在 SnSe 施加正/负 向电压的时候,调控导电细丝的形成/断裂,实现 SET/RESET 过程. 此外, 低开关电压不仅适用于



图 4 导电细丝效应 (a) Ag/SnO_x/SnSe 器件示意图以及器件的横截面 TEM 图像^[78]; (b) 忆阻器的保持性超过 10⁵ s^[78]; (c) Ag/SnO_x/SnSe 器件初始状态、导电细丝形成和断裂的示意图^[78]; 在 CDG (d) 和 DDG (e) 器件中导电细丝形成和断裂示意图^[83]

Fig. 4. Conductive filament effect: (a) Schematic of $Ag/SnO_x/SnSe$ device and the cross-sectional TEM image of the interface^[78]; (b) the retention of the device over 10⁵ s^[78]; (c) schematic of $Ag/SnO_x/SnSe$ device at initial state, conductive filament formation process and fracture state^[78]; schematic of conductive filament formation and rupture in CDG (d) and DDG (e) device^[83].

低功耗神经形态计算器件, 而且电压范围更加接近 生物动作电位, 也为与人脑神经网络的直接连接开 辟了途径^[46].

但是,导电细丝的形成过程存在随机性和不稳 定性. 特别是在低工作电流下, 忆阻器很难形成稳 定的细丝,但高工作电流下又会引起较大的功耗. 因此, Zhao 等^[83]利用石墨烯 (Gr) 缺陷工程, 通过 在 Ag/SiO₂/Pt 忆阻器中插入不同孔洞大小的石 墨烯 (DG) 层来集中或离散调控导线细丝的形成 和大小,从而使得在器件中形成稳定的导电细丝. 图 4(d) 和图 4(e) 展示了可控性细丝的形成机制, 在插入集中 DG 的 Ag/DG/SiO₂/Pt 器件 (CDG) 中,由于集中的 DG 会诱导限制在缺陷区域的阳离 子迁移,这使得即使很小的限流也会形成稳定的导 线细丝,并表现出非易失性特征.相反,插入离散 DG 的 Ag/DG/SiO₂/Pt 器件 (DDG) 由于离散 DG 的存在,使得即使很大的限流下形成的导电细丝也 不稳定,从而表现出易失性,上述缺陷工程为解决 导电细丝形成过程中的随机性以及稳定性问题,提 供了可行性方案.

2.2.2 氧空位

在一些过渡金属氧化物材料中,由于自身存在

与氧有关的缺陷空位, 在电场作用下可以发生氧化 还原反应, 形成氧空位导电通道来改变材料的阻 态. 尽管基于氧空位效应的忆阻器在性能改善方 面取得了显著进展, 但最常见的过渡金属氧化物 (TMOs, HfO_x, TaO_x等)功能层器件仍然无法满 足节能内存和计算任务的需求^[84]. 这是因为大多 数基于氧空位机制的忆阻器在较低的工作电流时, 很难提高数据保持性^[85].

而 2D 材料能够很好地解决上述由传统氧化物引起的问题. Liu 等^[50]利用氧等离子体技术调控 HfSe₂表面氧化过程,通过引入 HfSe_xO_y氧化层提高电阻,进而降低了工作电流,这避免了 HfSe₂高电导导致的高功耗和低开关比问题.并且,器件的 HRS 和 LRS 可以保持 1.5×10⁴ s 以上,在 40 个直流开关周期下仍具有较大的开关比 (10³). 器件的 SET 和 RESET 电压,分别以 2.32 V和-0.7 V 为中心呈现正态分布特征,显示出较好的稳定性.器件结构如图 5(a)所示,其中机械剥离的 2D 层状 HfSe₂纳米片的氧化层充当了阻变介质层,夹在顶部活性电极 Ti 和底部惰性电极 Au 之间.在扫描电压下,器件表现出可重复的双极特性并且可在超低电流下工作 (100 nA),如图 5(b)所示.这是



图 5 氧空位效应 (a) Ti/HfSe_xO_y/HfSe₂/Au 忆阻器示意图^[50]; (b) 器件的电学特性:在低工作电流 (100 nA) 下,器件的 *I-V* 曲线^[50]; (c) 锥形氧空位通道在电压调控下形成和断裂的过程^[50]; (d) Pd/WS₂/Pt 器件结构示意图^[49]; (e) 文献中报道的不同的编程或 SET 电流的比较^[49]

Fig. 5. Oxygen vacancy effect: (a) Schematic of Ti/HfSe_xO_y/HfSe₂/Au memristor; (b) electrical characteristics of the device: I-V curves of the device at low operating current (100 nA) ^[50]; (c) the formation and rupture of conical oxygen vacancy channels under voltage regulation^[50]; (d) schematic of the Pd/WS₂/Pt device^[49]; (e) comparison of various programs or SET currents reported in the literatures^[49].

由于活性电极 Ti 可以从 HfSe_xO_y 中吸收一些氧形 成 TiO_x, 界面附近产生丰富的氧空位, 从而使得器 件表现出了优异的性能.和无定形 HfSe_xO_y相比, HfSe₂具有更大的空位形成能, 从而充当"壁"的作 用以防止氧空位在阴极处累积.当 Ti 电极施加正 向偏压时, 带有正电的氧空位向底部电极移动, 但 被 2D HfSe₂层所阻挡, 在界面处开始形成氧空位 通道, 而由于 Ti 电极附近的 O 空位数量远高于 HfSe₂附近的氧空位数量, 因此形成稳定的"锥形" 导电通道^[86].相反, 当施加负电压时, 导电细丝会 因从圆锥体顶端回收氧空位而断裂, 这个过程仅需 要非常低的能量, 使得该器件工作电流可降低到 (100 pA), 能耗可以降低至 0.1 pJ 甚至 0.1 fJ, 这 在计算储存中能很好地实现节能效应, 如图 5(c) 所示.

与通过外界处理方式引入空位相比,利用材料 自身的空位不仅简化了工艺流程,也避免了引入空 位的不确定性和不可控性. 由于 WS, 薄膜中的具 有较多的 W 和 S 空位,并且在声子限制下电子迁 移率非常高,这使得其在实现低功率中很有应用前 景^[87]. Yan 等^[49] 基于 2D WS₂ 设计了 Pd/WS₂/Pt 结构的忆阻器, 如图 5(d) 所示. WS2 中 W 和 S 空 位的移动以及电子在空位间的迁移是该器件实现 忆阻行为的物理机制. 该器件与其他基于 2D 材料 的器件进行了比较,显示出了最低能耗 (如图 5(e) 所示), 凸显了其在低功耗的神经形态计算中的优 势. Yan 等^[49] 对有空位的 WS, 的电子结构进行了 研究,通过密度泛函理论 (DFT) 计算出态密度分 布情况,并与无空位的 WS2 的态密度进行了比较. 结果证实钨空位和硫空位形成的缺陷态处于较深 能级.因此,空位处电荷不容易发生泄漏,从而导 致器件表现出较低功耗.

为了在电调控基础上引入光调控,使得基于氧 空位的器件具有良好的光响应机制.与上述利用沟 道材料本身存在的或其表面氧化层产生的氧空位 方式不同的是, Chen 等^[88]通过沟道材料与富含氧 空位的材料进行 vdWs 异质结设计的方式引入氧 空位,设计了 MoS₂/BiFeO₃/SrTiO₃ 固态离子存储 器.其中 BiFeO₃ 中的氧空位不仅显著增强了 MoS₂ 的光响应,而且还很好地实现了光电存储特性.

2.2.3 隧穿效应

隧穿效应也是常见的一种量子效应,利用不同

材料电子亲和能以及功函数之间的差别, 在外电场 作用下, 通过调节材料的能带结构以及势全高度使 得电子隧穿, 从而实现阻态变换. 浮栅器件就是基 于隧穿效应设计而成的, 其中浮栅层与导电沟道之 间的电荷隧穿使得信息存储成为可能, 这使得浮栅 在互补金属氧化物半导体 (CMOS)体系结构中得 到了广泛的应用^[51,52]. 然而大部分传统浮栅器件却 存在较大的栅极电压, 在实际的神经形态计算的应 用中有较高的能耗, 这使得浮栅器件在应用方面有 很大的限制.

根据器件的隧穿机理,电子的隧穿概率与隧穿 势垒的高度和宽度有关. 降低工作电压的一种可能 性的方法是降低隧穿层的厚度. 然而, 薄的隧穿层 会造成电子在浮栅泄漏,影响器件的稳定性.基于 此, He 等^[55] 使用 2D MoS₂ 作为导电沟道层, 制作 了一种基于隧穿效应调制的多端器件, 通过优化隧 穿层 (h-BN) 的厚度 (15 nm), 降低了工作电压, 从而 提高了器件的可靠性,如图 6(a) 所示.其中,Au 的 功函数为 5.1 eV, 而 h-BN 具有较大的带隙 (5.2-5.9 eV) 和较小的电子亲和能 (2—2.3 eV)^[89,90]. 当 在漏极施加正向电压时,漏极和浮栅之间就会有较 大的电场存在,这使得原来平整的能带向漏极弯 曲,并使电子可以从浮栅隧穿到漏极.由于浮栅和 源极之间的电位差可以忽略,这可以防止它们之间 的电子泄漏. 电子从浮栅隧穿到漏极之后, 浮栅带 有正电, 而大部分电子在 MoS₂ 沟道中积累, 进入 LRS. 相反, 当在漏极施加负电位时候, 电子则从漏 极隧穿到浮栅,进入 HRS, 如图 6(b) 所示. 图 6(c) 和图 6(d) 展示了基于 MoS₂ 的多端器件在 V_g以 及 V_{ds}调节下的开关特性,与传统的由栅极电压调 控的浮栅存储器不同的是,浮栅中的充电和放电过 程也可以由 V_{ds}调节实现. 这种多端结构还可以用 来模拟突触可塑性,在不同 Vg下,器件电导变化 具有良好的可重复性和可调性,如图 6(e) 和图 6(f) 所示. 并且在 12 V, 50 µs 的脉冲下实现了高速开 关 (50 ns)、低能耗 (7.3 fJ) 突触的可调模拟性权重 更新,这为浮栅器件在低工作电压下实现低功耗神 经形态计算器件提供了指导.此外,降低工作电压 的另一种可能方法是设计浮栅的位置和形状, Wang 等^[91]利用 MoS₂ 作为导电沟道以及 h-BN 作为介 电层,通过特定的 vdWs 堆叠设计了双栅极结构的 浮栅器件,并在较低工作电压下(约5V)表现出



图 6 隧穿效应 (a) 基于 MoS₂ 的多端器件的示意图^[55]; (b) 浮栅/h-BN/漏极的能带图^[55]; (c) 基于 MoS₂ 的多端器件在 V_{ds} = 1 V 的传输特性^[55]; (d) 在不同 V_{ds}下的开关行为^[55]; (e) 基于三端器件的突触示意图^[55]; (f) 不同 V_g下多端突触器件重复性增强和抑制行为的对数图^[55]

Fig. 6. Tunneling effect: (a) Schematic diagram of the MoS_2 -based multi-terminal device^[55]; (b) band diagram of floating-gate/h-BN/drain^[55]; (c) transmission characteristics of multi-terminal device based MoS_2 at $V_{ds} = 1 V^{[55]}$; (d) switching behavior at different $V_{ds}^{[55]}$; (e) schematic diagram of a synapse based on a three-terminal device^[55]; (f) logarithmic plots of repetitive potentiation and inhibitory behavior of multiterminal synaptic apparatus under different $V_{g}^{[55]}$.

优异的性能,具有较长的保持时间 (10⁵ s) 和超低 关断漏电流 (10⁻¹³ A). 与传统的浮栅存储器中控制 栅和浮栅相对位置不同的是,该器件的控制栅和浮 栅并不在器件的同一侧,而是在器件导电沟道层的 两侧. 这样的结构会导致顶部浮栅和底部控制栅之 间存在电容耦合效应,这也是该器件能够降低工作 电压和实现良好非易失性存储的原因.

但是,对于浮栅器件来说,其读写需要的时间 比较长,一般在 ms 级别^[92-94].为进一步缩短器件 读写时间, Wu 等^[95]利用 2D vdWs 异质结设计了 InSe/h-BN/Gr 结构的浮栅存储器件.利用原子级 锐利的界面和增强的界面耦合特性,首次构筑了超 快、非易失性浮栅存储器,实现了纳秒级 (约 20 ns) 的读写操作以及极长的存储时间 (10 年以上).

从改变调控方式角度出发, Lai 等^[96]利用 2D Ruddlesden-Popper(2D-RPP)钙钛矿材料中优异 的电荷存储能力和敏感的光响应机制,设计了基 于 MoS₂/h-BN/2D-RPP vdWs 异质结的光电存储 器. 与传统浮栅器件中通过电刺激进行编程和擦除 操作不同,光不仅可以做到非接触式调控,还可以 避免重复的电压驱动增加的功耗,从而增加器件的 可靠性和稳定性. 这为光电融合神经形态器件的发 展提供了有力的支持.

2.2.4 电荷的捕获与释放

电荷捕获效应的本质是, 微观形态下基于电荷 量子化, 通过电场调控电荷的捕获与释放, 进而引 起宏观上阻态的改变. 与基于导电细丝、氧空位效 应的器件相比, 通过对电荷的捕获与释放对阻态进 行调节, 不会引起微观结构的变化^[97]. 这也使得基 于电荷捕获与释放的器件具有更稳定的性能以及 更广泛的应用价值^[98-102]. 随着 2D 材料研究的深 入, 人们基于 2D 材料中电荷捕获效应的神经形态 器件的成果展开了广泛的研究^[58,103-105].

但是,目前基于 2D 材料的电荷捕获效应的人 工突触仅限于在单个或数量较少的器件上实现.为 此,Xiang 等^[53]研究了基于富硅氮化硅 (sr-SiN_x) 衬底的多层 MoTe₂ 器件,并成功设计出基于 2D MoTe₂ 的突触阵列,如图 7(a)所示.图 7(b)显示 了器件具有典型的以 n 型为主的双极性行为,这是 由于富含 Si—Si 键的 SiN_x介电层具有优越的电荷 捕获能力^[106,107].当栅极加正电压时候,空穴被注 入到 sr-SiN_x介电层中,并被介电层中自身固有的 空穴捕获中心所捕获.这样介电层中带有正电的大 量空穴会使得 MoTe₂导电沟道中电子浓度和电流 显著增强,即使当栅极电压去除时,空穴仍然被存 储,器件处于 LRS,完成写入存储过程.在擦除过



图 7 电荷的捕获与释放 (a) 基于 2D MoTe₂ 的忆阻器结构示意图^[53]; (b) 在 V_g(-40 V→40 V) 扫描电压下器件的传输 特性曲线 (插图为在对数坐标下的 *I-V*曲线)^[53]; (c) 2D MoTe₂ 的忆阻器工作机制示意图^[53]; (d) 生物突触和基于 sr-SiN_x 的人工突 触器件的示意图; (e) 100 个周期内增强 (左) 和抑制 (右) 周期性电导的变化^[53]

Fig. 7. Charge trapping and de-trapping effects: (a) Schematic diagram of the memristor structure based on 2D MoTe₂^[53]; (b) the transfer characteristic curve of the device under the scanning voltage of $V_{\rm g}$ (-40 V \rightarrow 40 V \rightarrow -40 V) (the illustration is the same curve shown in logarithmic coordinates) ^[53]; (c) the working mechanism of the device^[53]; (d) schematic illustration of biological synapses and sr-SiN_x-based artificial synaptic device^[53]; (e) the conductance periodic changes in excitation (left) and inhibition (right) over 100 cycles^[53].

程时,通过在栅极施加负电压将捕获的空穴驱回 Si 中,从而去除有效存储电荷,如图 7(c)所示.图 7(d) 和图 7(e)展示了器件在人工神经网络应用中的能 力,研究了器件的基本突触功能,电导的聚集分布 在每个状态下都有很小的变化,循环稳定性很高, 展现出良好的增强和抑制效果.

对于基于电荷捕获效应的神经形态器件来说, 其电荷捕获能力主要取决于表面官能团和电子杂 化态,因此引入具有丰富杂化态的材料是提高器件 性能的关键^[108]. Wen 等^[109]利用 2D 石墨炔 (GDY) 中丰富的电子杂化态和不同的化学基团性质,设计 了基于 GDY/MoS₂ vdWs 异质结中的电荷捕获存 储器.该器件可以在光电两种模式下工作,并展现 出良好的数据存储能力,很好地显示了 GDY 作为 富态电荷捕获中心在光学存储器和人工突触等方 面的应用潜力.

2.2.5 vdWs

由于 2D vdWs 异质结在场效应管^[110]、光电探 测器^[111,112]、太阳能电池^[113]以及神经形态器件^[114] 等领域表现出新奇的物理特性, 受到研究人员越来 越多的关注与研究.其中, 2D 异质结在生长过程 中存在晶格对称性、晶格常数等晶格结构匹配的限 制,很难生长出任意需求的异质结^[115].但是, 2D vdWs 异质结可以通过转移方式进行搭建,在此过 程中它不受晶格匹配的限制,可以将具有不同晶格 结构的 2D 材料集成在一起^[116].所以仅需考虑不 同 2D 材料的能带结构匹配等因素,进行特定的异 质结构设计,为新型神经形态器件提供特定的功 能^[117–122].并且, 2D vdWs 异质结被认为是非易失 性光学存储器的良好选择,这是因为异质结的存在 使其具有多个光敏层,可以辅助宽光谱吸收,并且 它们具有强烈的光-物质相互作用^[123–125].此外, 2D vdWs 在结界面处建立了有效的势垒, 阻碍了随机载流子的输运, 并抑制了器件中的随机噪声^[126].

2D vdWs 异质结一般有两种结构, 横向结构 和垂直结构.其中,2D材料的横向异质结构由于 高质量的原子精度异质接口,为多功能高性能电子 器件的异质结构设计创造了新的范例[127,128]. 但是, 在 2D 材料的横向异质结构中,同时实现具有鲁棒 性的电阻开关性能和多栅调制的人工突触尚未得到 深入研究. 基于此, He 等^[129] 设计了基于 2D WSe2-WO3横向异质结构的多门栅控神经突触器件,成 功模拟了两种基于神经元的突触功能,并可通过四端 配置的栅极电压和可见光进行有效调制,如图 8(a) 所示. He 等发现改变栅极电压能有效调节开关特 性,当 V。电压从 0 更改为-20 V 时候, HRS 的电 流从 15 pA 增加到 122 pA (V_{ds} = 0.1 V), 但 LRS 的电流几乎保持不变,如图 8(b) 所示. 当进一步降 低 V。时, HRS 的电流增加到 820 pA, 而 LRS 的 电流仍然保持不变,这可以由 WSe2 中的多子空穴 来解释. 图 8(c) 显示了 WSe₂-WO₃ 异质结构的光 学图像.为了揭示 WSe2-WO3 异质结构中电阻开 关的物理本质, He 等测量了具有不同电极结构的 器件的电学特性, 如图 8(d) 所示, 只有 WSe₂-WO₃ 异质结构 (电极 3 和 4) 构成的器件在 *I-V*曲线中 呈现出典型的电阻开关迟滞回线.这些结果清楚地 表明, 中间过渡层 WSe_{2-x}和 WO_{3-x}在电阻开关中 起着关键作用.图 8(e) 和图 8(f) 展示了 WSe₂-WO₃ 异质结构引起阻变的原理, 这是由于 WSe₂ 有很好 的催化析氢能力, 能很好地吸收氢并与 WO₃反应. 在施加不同类型电压下, 驱动质子的移动, 使得 H_xWO_{3-y} 与 WO_{3-y}发生可逆转变, 进而实现 HRS 与 LRS 之间的相互转换.

对于垂直 vdWs 异质结, Wang 等^[130] 选取了 MoS₂和 Gr 分别用作忆阻器的沟道层和电极材料, 制备了垂直 vdWs 异质结构, 如图 9(a) 所示. 并发 现该器件具有优良稳定的开关性能, 其中可擦写次 数超过千万次, 即使在 340 ℃ 的高温下仍然可以 实现稳定工作并且保持优良的开关性能, 如图 9(b) 和图 9(c) 所示. 为了进一步探究器件热稳定性的 本质原因, Wang 等^[131] 采用透射电子显微镜 (TEM) 观察, 发现热稳定性来源于 MoS_{2-x}O_x 以及 Gr 层 尖锐的原子界面, 并揭示了器件的开关机制是由于 氧离子的迁移. 对于现阶段大部分忆阻器件来说,



图 8 横向 vdWs 异质结 (a) 基于 2D WSe₂-WO₃ 横向异质结构的器件示意图^[129]; (b) 由 Gate 1 调节的电阻开关特性^[129]; (c) WSe₂-WO₃ 横向异质结构的光学图像^[129]; (d) Pd-WSe₂-Pd(电极 4 和 5)、Pd-WO₃-Pd(电极 1 和 2) 和 Pd-WSe₂-WO₃-Pd(电极 3 和 4) 的 *I-V* 特性曲线^[129]; (e), (f) 开关原理的示意图,其中红色圆圈代表质子^[129]

Fig. 8. Lateral vdWs heterostructure: (a) Schematic diagram of the device based on the 2D WSe₂-WO₃ lateral heterostructure^[129]; (b) resistive switching characteristics regulated by Gate 1 voltage^[129]; (c) optical image of WSe₂-WO₃ lateral heterostructure^[129]; (d) *I-V* characteristic curves of Pd-WSe₂-Pd (electrodes 4 and 5), Pd-WO₃-Pd (electrodes 1 and 2) and Pd-WSe₂-WO₃-Pd (electrodes 3 and 4) ^[129]; (e), (f) schematic of the switching principle, where the red circles represent protons^[129].



图 9 垂直 vdWs 异质结 (a) 基于 MoS_{2-x}O_x/Gr 异质结的器件示意图^[130]; (b) 器件在不同温度下的开关曲线^[130]; (c) 器件在 340 和 160 ℃下的保持时间^[130]

Fig. 9. Vertical vdWs heterojunction: (a) Schematic diagram of the device based on $MoS_{2-x}O_x/Gr$ heterojunction^[130]; (b) switching curves of the device at different temperatures^[130]; (c) retention time of device at 340 and 160 °C ^[130].

在 200 ℃ 以上的温度环境下,器件可靠性将受到 严重的损害,而此工作实现了在高温环境下忆阻器 件依旧保持稳定的开关性能,为未来使用 vdW 异 质结构的器件在恶劣环境下稳定工作的电子设备 提供了一条途径.

3 神经形态器件的基本特征

3.1 神经元模型

在生物神经网络中,神经元通过离子的移动来 接收、处理和传输信号. 在神经元的脂质双层膜中, 离子通道调节细胞外液和细胞内液之间离子 (Na+, K⁺, Ca²⁺)的浓度. 随着突触信号的输入, 膜电位 会随着两种细胞液之间离子浓度的变化而变化.一 旦膜电位达到阈值,离子开始流过离子传导通道来 传输信号.之后,膜电位返回平衡状态.为了解释 神经元工作时电位的动态特性,神经生理学家建立 了许多模型,主要有5种神经元模型: Hodgkin-Huxley (H-H) 模型^[132,133], Izhikevich 模型^[134], LIF (leaky integrate and fire) 模型^[135,136], SRM (spike response model) 模型^[137]和 ANN (artificial neural network) 模型^[138]. 神经元通常在时域或频域表现 出丰富的动态和瞬态转换行为,这引起了越来越多 的研究人员的关注. 在过去的几年中, 人们在 2D 材料的光电突触器件方面开展了相当多的研究工 作. 尽管 2D 材料有望为结构紧凑、高密度集成、可 扩展和节能的人工神经元突触器件提供可行性条 件,但目前的实验探索仍处于起步阶段.在此将重 点介绍基于 2D 材料的神经形态器件实现对 LIF 以及 H-H 神经元模型的模拟.

3.1.1 LIF 神经元模型

LIF 神经元模型可以利用阈值开关忆阻器、电 容器、易失性无阈值忆阻器件和比较器组成的混合 电路来实现. Hao 等^[139] 通过引入离子迁移和电化 学反应, 在基于 2D MoS, 的平面器件中成功模拟 了基于 LIF 模型的人工神经元, Ag+离子在电场中 的行为模拟生物神经元中 Ca²⁺的流动, 如图 10(a) 所示. 通过在单层 MoS2 沟道中控制 Ag 导电细丝 的形成和断裂进而模拟 LIF 模型的易失性开关行 为. 且在输入连续电压脉冲后, MoS2器件表现出 LIF 神经元的电学行为, 如图 10(b) 所示. 图 10(c) 展示了通过施加单个脉冲,器件可以在 0.5 ms内 实现开关转换,这表明该器件具有良好的脉冲响 应. 此外, Dev 等^[140]利用化学气相沉积 (CVD) 生 长的 2D MoS₂ 实现了阈值电压忆阻器件,并通过 外接电路的方式,也成功实现了具有 LIF 特性的 人工神经元. 该电路模拟了生物膜电位的整合过 程,生物神经元的离子运动过程由阈值电压忆阻器 中的 Ag+来模拟, 如图 10(d) 所示. 电路示意图如 图 10(e) 所示. 当持续输入宽度为 100 us, 振幅为 1 V 的电压脉冲时, 可以得到稳定输出的多个电流 峰值,从而展现出 LIF 特性,如图 10(f) 所示.

3.1.2 H-H 神经元模型

H-H神经元模型的命名来源于 Hodgkin 和 Huxley 的合作,其中在 1952 年研究乌贼轴突电生 理活动时候,用数学模型的方式展示了神经元细胞 膜上的 *I-V*变化关系,即 H-H 模型.神经元具有可 刺激性,当受到微小的刺激后,电位发生波动后会 很快恢复到平衡状态,但当受到的刺激超过某一



图 10 LIF 模型神经元 (a) 平面器件 Ag/MoS₂/TiW 示意图^[139]; (b) 器件在连续脉冲序列下的泄漏-集成-发射的电学行为^[139]; (c) 器件在 1 ms、电压为 2.0 V 的单脉冲下的易失性开关行为^[139]; (d) 垂直器件 Ag/MoS₂/Au 的结构示意图和光学图片^[140]; (e) 上 图为 Ag/MoS₂/Au 人工神经元的电路图,下图为神经元的连续输出电流尖峰^[140]; (f) 上图为电路图节点 B 处的电压 V_B,下图为负载电阻 R_L 两端的电压 V_{RL}^[140]

Fig. 10. LIF model neurons: (a) Schematic diagram of planar device $Ag/MoS_2/TiW^{[139]}$; (b) leakage-integration-emission electrical behavior of device under continuous pulse trains^[139]; (c) volatile switching behavior of the device with a single pulse of 2.0 V at 1 ms^[139]; (d) schematic diagram and optical picture of the vertical device $Ag/MoS_2/Au^{[140]}$; (e) the top picture is the circuit diagram of the $Ag/MoS_2/Au$ artificial neuron, and the picture below is the continuous output current spike of the neuron^[140]; (f) the picture in the top panel shows the voltage V_B at node B of the circuit diagram, and the picture in the bottom panel shows the voltage V_{RL} across the load resistance $R_L^{[140]}$.

阈值时,神经元的电位将会有一个快速又陡峭的响 应.这时,神经元处于激发态.但 H-H 模型的人工 神经元通常需要一个复杂的电路来模拟神经元膜 中离子沟道的动态过程.

基于上述问题, Beck 等[141] 报道了一种可以通 过双栅极电压调控的高斯异质结晶体管 (GHeT), 并成功实现了 H-H 尖峰神经元, 器件的光学照片 以及示意图如图 11(a) 和图 11(b) 所示. 图 11(c) 显示了生物神经元的 H-H 模型的电路, 其中 Na+ 的注入会导致神经元膜电位 (V_m)产生尖峰, 而 K+ 离子的释放会重置 Vm. 图 11(d) 和图 11(e) 显示了 H-H 模型中 K⁺电导 (g_K) 和 Na⁺电导 (g_{Na}) 的时间 演变关系. gx 随时间的变化可以用 n 型金属氧化 物半导体晶体管 (NMOS) 的延迟开启来表示, 电 压 V_m 通过阻容负载施加到栅极. 而 Na⁺的行为更 复杂,需要峰值时间依赖响应,这会使得峰值电导 增加,但随着 V_m的增加,达到峰值电导的延迟会 减少. 图 11(f) 详细介绍了用单个 MoS₂-GHeT 器 件、NMOS(T1和T2)和一些无源元件(R1, R2, C1 和 C₂) 来模拟神经元的完整电路图. 电压源 V₃ 和

 V_5 连接在 T_1 和 T_2 的源电极处,使得场效应晶体 管的阈值电压具有可编程性. GHeT 和电路组件 T_1, R_1, C_1 模拟 g_{Na} ,而电路组件 T_2, R_2, C_2 模拟 g_K . 在施加突触电流 (I_{syn}) 之前,由于大的正栅极 偏置 (V_1) 使得 GHeT 处于关闭状态,当有足够大 的突触电流时, C_1 和 C_2 会整合 I_{syn} 以及 GHeT 的断开电流 (I_{OFF}) .当 V_m 超过 T_1 的阈值电压时, 施加到栅极的电压从 V_1 迅速下降到 0 V,从而产 生负的相对栅极电压 $(V_{TG}-V_m)$.这种情况驱动 GHeT 从关态到峰值开态,增加的电流使得 V_m 的 斜率急剧增加.当 V_m 达到 T_2 的阈值电压时,延迟 的 g_K 就能够支配和重置 V_m ,使其低于 T_1 的阈值 电压.当 $I_{syn}+I_{OFF}$ 足够高时,这种尖峰响应将重 置下去,如图 11(g) 所示.

3.2 突触可塑性

突触是将前一个神经元的冲动传输到下一个 神经元或者细胞中的结构.随着时间的推移,控制 和保持突触权重的能力被定义为突触可塑性,其包 括长期可塑性 (LTSP)、短期可塑性 (STSP)、峰值



图 11 H-H尖峰神经元 (a) 器件的光学图片示意图^[141]; (b) 器件结构示意图^[141]; (c) H-H 模型神经元的电路等效图^[141]; (d) H-H 模型中 g_K 的时间演变关系图^[141]; (e) H-H 模型中 g_{Na} 的时间演变关系图^[141]; (f) GHeT 神经元的完整电路图^[141]; (g) 图 (d) 中的 GHeT 神经元电路的前 30 s 的实验结果^[141]

Fig. 11. H-H spiking neurons: (a) Optical image of the device^[141]; (b) the schematic of device structure^[141]; (c) equivalent circuit diagram of H-H model neuron^[141]; (d) time evolution diagram of $g_{\rm K}$ in H-H model^[141]; (e) time evolution diagram of $g_{\rm Na}$ in H-H model^[141]; (f) complete circuit diagram of a GHeT neuron^[141]; (g) experimental results for the first 30 s of the GHeT neuron circuit in panel (d) ^[141].

时间依赖性可塑性 (STDP) 以及放电速率依赖可 塑性 (SRDP) 等^[142-144]. 它们是突触进行神经信号 处理、神经形态计算的基础.

3.2.1 LTSP 长期可塑性

LTSP 可以在兴奋性突触上获得,包括长时程 增强 (LTP) 和长时程抑制 (LTD),这表明在连续 刺激突触后,突触强度持续增加和减少.同时, LTSP 被认为是长期学习和记忆的生物学基础^[145].

但是,在器件工作时,LTP和LTD之间转换 的不稳定性对训练和识别过程有较大的负面影响. 基于此,Kwon等^[42]利用 2D 铁磁材料 SnS 研究 了基于 Pt/SnS/Pt 结构的 LTP 以及 LTD 等突触 特性,该器件具有高稳定的 LTP/LTD 的转换机 制,在 10000 个连续刺激下具有均匀的电导变化, 如图 12(a)和图 12(b)所示.而与两端突触相比, 三端突触结构由于其额外的输入端和改进的器件 结构已被证明更有利于实现复杂的突触功能^[146–148]. Zhu 等^[149]利用 2D 层状 WSe₂, NiPS₃和 FePSe₃ 设计 vdWs 异质结,制备了三端离子门突触晶体 管,如图 12(c) 所示,器件可以在生物真实性上模 拟人工突触功能,并且在单脉冲刺激下能耗低至 30 fJ.图 12(d)则展示了生物系统和离子门突触晶 体管离子浓度迁移和动态平衡的过程.而图 12(e) 展示了在一系列连续脉冲刺激下器件由 STSP 转 换到 LTSP 的过程.并且,该突触晶体管展现出显 著的线性和对称性,如图 12(f)所示,这可以在不 引入复杂的外围电路的同时实现高度可控和可预 测的学习过程^[150].

为了在不影响器件性能的基础上, 实现较低 工作电压下的突触可塑性, Wang 等^[151] 基于 2D α-In₂Se₃ 材料制备了铁电半导体沟道器件 (FeCTs), 该器件不但可以很好模拟生物上的 STSP, 而且 可以在施加小电压 (±0.5 V)的脉冲刺激下实现 LTP/LTD 的模拟. 同时, 该器件在单脉冲增强/抑 制中展现了超低的能耗.

3.2.2 STSP 短期可塑性

STSP 是指通过短期内平衡大脑皮层的增强 和抑制功能,来控制增强突触传递,实现神经活动



图 12 LTSP (a) 人工突触器件示意图^[42]; (b) 重复进行 LTP 和 LTD 操作, 一个周期是 100 个增强脉冲和随后的 100 个抑制脉冲. 左上图和右上图分别对应最初 10 个循环和最后 10 个循环的运行情况^[42]; (c) 离子门控突触晶体管的示意图^[149]; (d) 生物系统 (上) 和离子门控突触晶体管 (下) 中离子迁移和动态平衡过程^[149]; (e) 一系列电压脉冲 (5 V, 50 ms) 施加到栅极 (V_{ds} = 0.5 V) 时, 实现从 STSP 到 LTSP 的转换^[149]; (f) 对基于 WSe₂ 的突触晶体管使用增强 (1.2 V, 100 ms) 和抑制 (-0.4 V, 100 ms) 脉冲信号, 间隔 3 s, 显示出良好的线性、对称性和重复性^[149]

Fig. 12. LTSP: (a) Schematic diagram of artificial synapse^[42]; (b) the LTP and LTD operations were repeated with a cycle of 100 enhancement pulses followed by 100 inhibition pulses; the upper left and upper right diagrams correspond to the operation of the first 10 cycles and the last 10 cycles, respectively^[42]; (c) schematic of an ion-gated synaptic transistor^[149]; (d) ion migration and dynamic equilibrium in biological systems (top) and ion-gated synaptic transistors (bottom)^[149]; (e) when a series of voltage pulses (5 V, 50 ms) are applied to the gate ($V_{ds} = 0.5$ V), the transition occurs from STSP to LTSP^[149]; (f) by using excitatory (1.2 V, 100 ms) and inhibitory (-0.4 V, 100 ms) pulsed signals with 3 s intervals for WSe₂-based synaptic transistors, the device shows good linearity, symmetry, and reproducibility^[149].

的时间和空间特征,但其在记忆时间上往往只能维 持几分钟甚至几秒钟.STSP与LTSP相对应,也 同时具有短时程增强和短时程抑制,具体实现方式 包括双脉冲抑制 (PPD)、双脉冲易化 (PPF)以及 强直后增强 (PTP)^[56,152,153].如图 13(a)所示,在生 物突触中,PPD可以通过触发两个具有时间间隔 (Δt)的相同脉冲来实现.如果脉冲紧跟前一个脉 冲,在一定的时间范围内,它可以表现出时间突触 强度的减弱.而 PPF则与 PPD 相反,突触增强行 为会逐渐减弱.而 PTP和 PPF 类似,也属于短时 程增强的类别,但是与 PPF 不同的是, PTP 持续 时间可达十几秒甚至几分钟.

近来, Sun 等^[26] 基于 MoS₂ 设计的突触器件, 通过改变两脉冲的 Δt 来调节 MoS₂ 突触器件产生 的焦耳热进而改变器件的电导,最终成功模拟实现 了 PPF 和 PPD,如图 13(b)所示.正如预期的那 样,随着 Δt 从 10 ms 增加到 10 s, PPD 和 PPF 指数逐渐收敛到 100%,两条测量数据与拟合曲线 吻合良好.图 13(c) 和图 13(d)展示了在相同脉冲 刺激下, PPF 以及 PPD 的模拟效果.其中对于 PPF 来说,后脉冲产生了抑制前脉冲的效果, PSC 得到 抑制,导致抑制性突触后电流 (IPSC) 进一步减小,



图 13 STSP (a) 生物突触示意图; (b) PPD 指数和 PPF 指数与 Δt 的函数关系图 ^[26]; (c) IPSC 和 (d) EPSC 效果示意图 Fig. 13. STSP: (a) Schematic diagram of a biological synapse; (b) PPD index and PPF index are shown as the function of $\Delta t^{[26]}$; (c) IPSC and (d) EPSC effect diagram.

实现了 PPD 过程. 对于 PPD 来说, 后脉冲增强了前脉冲刺激突触强度, 兴奋性突触电流 (EPSC) 进一步增大. 此外, Gao 等^[154] 基于 2D MoTe₂ 设计了能够在弱电刺激下模拟 STSP 行为的 FeFET 器件. 通过增加 V_g脉冲的持续时间和幅度来实现 STSP 到 LTSP 转变, 并可以观察到更长的保持时间和更强的突触权重调制. 研究还表明, STSP 到 LTSP 的转变也可以通过改变突触前 V_g脉冲数或脉冲频率来模拟实现.

3.2.3 STDP 与 SRDP

STDP和 SRDP 与神经形态系统中信号传输 以及记忆学习相关. 二者符合 Hebbian 理论的学 习规律,可以为突触权重调整提供一种有效方法, 进而实现人工神经网络的学习功能. 在 Hebbian 理论中,当一个突触前神经元重复或持续地刺激 突触后神经元时,两个神经元之间的连接就会增 强. 也就是说,突触连接权重可以被加强或减弱, 这取决于从前神经元到后神经元的传入信号的强 度^[155]. 这里信号强度主要指的是输入信号的尖峰 时间和尖峰速率.

STDP 指的是突触权重根据施加峰值信号的 相对时间进行调控的过程.在 STDP 中,突触权重 由突触前脉冲和突触后脉冲之间的时间延迟来调 节.根据突触权重的变化和脉冲时间的间隔, STDP 可以分为四种类型^[156–159],如图 14(a)所展示的, 1)反对称 Hebbian 学习规则; 2)反对称反 Hebbian 学习规则 (原始 STDP); 3)对称 Hebbian 学习规 则; 4)对称反 Hebbian 学习规则.如果将突触前 (兴奋性)和突触后 (抑制性)脉冲的到达时间分别 定义为 t_1 和 t_2 ,则当 Δt 大于或小于 0 时, 会使得连 接强度增加或降低. Δt 越小, 突触权重的变化越 大,反之亦然.在一个典型的 STDP 中, 突触权重被 表示为一个双指数函数, 如图 14(b) 所示^[157,158,160].

SRDP 指的是突触权重由施加的峰值信号的 尖峰频率调控的过程. 它描述了突触可塑性对连续 动作电位之间 Δt 的依赖性, 在类脑神经网络中, 可以通过应用一系列 Δt 不同的电压脉冲模拟低频 或高频. 一般情况下, 增强型的 PSC 会随着两个电 压脉冲之间的 Δt 的缩短而增加, 而抑制型的 PSC 则会随着 Δt 的缩短而减小^[161]. 如图 14(c) 所示, 高频重复信号 (>10 Hz) 会引起 LTP, 而低频重复 信号 (<10 Hz) 会导致 LTD^[160,162,163].

由于 Hebbian 学习规则可以为调整突触权重 进而实现神经网络学习提供一种有效方法, Wang 等^[164] 基于单层 MoS₂ 设计了四端突触器件. 通过



图 14 STDP和 SRDP (a) 四种 STDP 模型突触权重变化的示意图; (b) 典型的 STDP 模型. 突触前电流 (兴奋性)和 IPSC 的 相对时间前后影响突触权重产生 LTP和 LTD; (c) 典型的 SRDP 模型, 尖峰频率的大小带来的突触权重的变化 Fig. 14. STDP and SRDP: (a) Schematic diagram of synaptic weight changes of the four STDP models; (b) typical STDP model. Influence synaptic weights to generate LTP and LTD on relative timing of presynaptic currents (excitatory) and IPSCs; (c) typical SRDP model; changes of synaptic weights caused by the magnitude of the spike frequency.

调节突触前和突触后脉冲的 Δt 来改变突触强度, 很好地模拟了 STDP. 其中,电流增加和减少分别 代表突触增强和抑制,并且电流随 Δt 的变化符合 反对称 Hebbian 学习规则,能够很好地实现神经 网络计算学习. 在相关性检测上,其决定系数 R² >0.9,这表现了其模拟结果与生物系统 STDP 有 良好的相关性. 此外, Liu 等^[165] 基于 SnS₂, h-BN 和 少层 Gr 设计了三端浮栅器件.利用浮栅器件成功 模拟了 SRDP,发现 SRDP 指数与刺激频率成正 比. 这一现象表明,频率较高的脉冲序列可以加强 突触前和突触后的相关性,从而产生更强的兴奋作 用,这也为人工突触器件中实现高效神经网络系统 的基本功能提供了条件. 简而言之, SRDP 和 STDP 的模拟使该器件能够更好地构建神经网络.

为了更加直观地对比不同材料的人工突触器

件性能,如表1所列,分别从2D材料^[44,53,166,167]、 氧化物材料^[168-176]、有机材料^[177-181]以及钙钛矿 材料^[182-185]出发,对人工突触器件的突触可塑性、 耐久性、保持性、能耗或功耗、刺激方式等方面 进行了比较.其中从表1可以看出,2D材料展 现出较好的光响应,因为基于2D材料的忆阻器 在具有良好的保持性和耐久性的同时,其突触器件 体现出较低的能耗;氧化物材料的研究相对成熟, 器件具有较好的保持性,可以很好地模拟突触可塑 性;有机材料在柔性应用方面有很大前景,可以在 低功耗的基础上实现光电耦合调控,但其在高温下 有机分子不稳定,易分解;基于钙钛矿材料的器件 功耗相对较低,对光、电有良好的响应机制,但其 在有氧以及高温环境下,数据保持性和耐久性 较差.
Table 1. Performance comparison of artificial synaptic devices based on different materials.							
	材料	突触可塑性	耐久性(循环)	保持性	功耗/能耗	刺激方式	文献
2D材料	MoS_2	LTP, LTD, PPF, SRDP	100	>11 h	4.5 fJ	电	[166]
	$\mathrm{h\text{-}BN}/\mathrm{WSe}_2$	LTP, LTD, STDP	—	_	66 fJ	光电	[167]
	SnSe	LTP, LTD, PPF, STDP	230	$> 10^4 { m \ s}$	$66 {\rm ~fJ}$	电	[44]
	$MoTe_2$	LTP, LTD, PPF, STDP	570	$> 10^4 { m \ s}$		电	[53]
氧化物材料	${ m TiO}_2$	LTP, LTD, PPF, STDP			$26 \mathrm{~pJ}$	电	[168]
	ZTO	LTP, LTD, STDP		$10^4 {\rm \ s}$		电	[169]
	IGZO	LTP, LTD			$160 \mathrm{~pJ}$	光电	[170]
	In_2O_3	LTP, LTD, PPF, SRDP	20			电	[171]
有机材料	C_8 -BTBT	LTP, LTD	10^{5}	$>3500 \mathrm{~s}$	$<5~{ m fJ}$	电	[177]
	PMMA	LTP, LTD, PPF, SRDP, STDP	_	$> 10^3 \mathrm{~s}$	10^{-8} W	电	[178]
	PVA	PPF, LTP, LTD, LTM		$600 \ s$		光	[179]
	P(VDF-TrFE)/P(VP- EDMAEMAES)	LTP, LTD, PPF			$75 \mathrm{ pJ}$	电	[180]
钙钛矿材料	BaTiO_3	STDP	10^{5}	$10^8 {\rm \ s}$	600 pJ	电	[182]
	$\mathrm{PdZr}_{0.52}\mathrm{Ti}_{0.48}\mathrm{O}_{3}$	SRTP, LSTP			$2.5 \ \mathrm{pJ}$	电	[183]
	${\rm BiFeO}_3$	STDP			$200~{\rm nW}$	电	[184]
	$(\mathrm{CH}_3\mathrm{NH}_3)_3\mathrm{Sb}_2\mathrm{Br}_9$	LTP, LTD, STDP	300	$10^4 { m \ s}$	117.9 fJ	电	[185]

表 1 不同材料的人工突触器件性能对比

4 2D 材料神经形态器件的应用

基于 2D 材料的神经形态器件可以将存算一体器件和传感器的功能集成到一个设备中,这为实现感存算一体的多功能神经形态计算,尤其是对于构建人工感知系统开辟了一条新途径.并且,基于 2D 材料的神经形态器件不仅在光电子学、生物学、声学以及热学等方面具有良好的物理性质,同时其在纳米尺度下会伴随新奇的物理性质^[186].这也使得基于 2D 材料的神经形态器件可以满足人们在生物感知(视觉识别^[161,187]、声音定位^[26]、触觉模拟^[188,189])等领域的应用需求.

4.1 视觉识别

人们在用突触器件模拟大脑神经网络方面已 经做了广泛的研究^[147,190-193].视觉作为人类感知 系统中重要的一部分,可以使人能够感知外界环境 中物体的方位、形状、大小等各种信息,并且人类 依赖视觉获得的信息比通过其他方式获得信息的 总和还要多.因此构建视觉识别系统,并对视觉信 息进行存储以及分类,是神经形态器件广泛应用的 重要部分.

然而,基于传统图像传感器、存储器的人工视

觉系统依赖复杂电路,这使得在器件集成和功耗 方面存在严重挑战. 基于此, Zhou 等^[194]利用 Pd/ MoO_r/ITO 结构设计了两端光电电阻随机存取存 储器 (ORRAM), 并发现此结构在波长为 365 nm、 功率密度为 150 mW·cm⁻² 的紫外激光器下可以实 现 HRS到 LRS 的转变, 展现出了非易失特性. 如 图 15(a)所示,视觉信息首先通过人眼中的视网膜 被检测和提取,然后该信息进一步通过视神经,并 在视觉皮层中得到处理. 在构建的神经形态视觉系 统中, ORRAM 阵列可以实现人类视网膜中的传 感和预处理功能. 在预处理后的图像经过三层人工 神经网络,可完成网络的训练和识别,如图 15(b) 所示. 图 15(c) 展示了通过 ORRAM 进行预处理 后,字母的主体特征被突出显示,背景噪声信号被 平滑.其中,图像识别率和效率在使用 ORRAM 预 处理后,有了明显的提高(图 15(d)).该方案展示 了利用 ORRAM 简化神经形态视觉系统电路的能 力,促进了基于简单电路的人工视觉系统的发展.

为进一步提高计算效率,降低成本,Sun等^[195] 采用储备池计算(RC),实现了在较低的计算成本 下简单快速地对数据进行训练,读出权重,使得学 习、计算成本大幅降低.Sun等通过 2D SnS 制备 了忆阻器阵列,并实现了大脑的视觉皮层功能对语 言的学习和记忆.图 15(e)展示了人脑视觉皮层神



图 15 视觉识别 (a) 人体视觉系统的示意图^[194]; (b) 基于 ORRAM 阵列的人工神经形态视觉系统,以及用于图像识别的人工 神经网络示意图^[194]; (c) 人工神经形态视觉系统预处理之前 (左) 和之后 (右) 的图像示例^[194]; (d) 有/无图像预处理的图像识别率 比较^[194]; (e) 在生物 RC 系统上执行的认知任务示意图^[195]; (f) 由电和光输入刺激的多功能忆阻器阵列示意图; (g) 对语言符号的 识别准确率^[195]

Fig. 15. Visual recognition: (a) Schematic diagram of the human visual system^[194]; (b) artificial neuromorphic vision system based on ORRAM array, and artificial neural network for image recognition^[194]; (c) images before (left) and after (right) preprocessing through an artificial neuromorphic vision system^[194]; (d) comparison of image recognition rates with and without image preprocessing^[194]; (e) schematic diagram of cognitive tasks performed on biological RC systems^[195]; (f) illustration of a multifunctional memristor array stimulated by various electrical and optical inputs; (g) recognition accuracy of language sign^[195].

经元之间的复杂连接和 RC 在权重训练后的输出 结果.语言符号作为光学信号输入到 RC 系统中, 不需要专门的图像传感器和相关的模式数据转换, 大大节省了能耗和时耗.并且在加入随机噪声干扰 后,对测试的语言符号的识别准确率仍然高达 91%,如图 15(f) 和图 15(g) 所示.该研究为边缘计 算达到低训练成本和信息的快速时效处理提供了 解决方案.

此外, Du 等^[196] 基于单层 MoS₂ 和 BaTiO₃ 铁 电薄膜设计了基于光电晶体管结构的神经形态视 觉传感器. 利用波长依赖的光学传感和多级光学存 储特性来实现传感器内神经形态视觉预处理, 减少 了冗余数据, 使得 MNIST 手写数据集的图像识别 率从 15% 提高到 91%.

4.2 声音定位与模式识别

人类的听觉可以检测、处理和存储声音信号, 这使得我们可以辨别发声物体的性质、方向和距 离等,进一步使得大脑更加准确地指导我们的行 为^[197,198].人类听觉通道中的神经中枢不仅仅可以 检测、转化电信号,还可以处理和存储声音^[199,200]. 因此,对于人工听觉系统来说,不仅仅要做到接受 外界电信号,同时还要对信号进行进一步分析处理 和存储. 其中人类听觉的一个重大作用就是可以实现 声音定位,其主要依靠双耳时差 (ITD) 和双耳水 平差 (ILD) 进行定位,如图 16(a) 所示.基于此,Sun 等^[26]利用单层 MoS₂设计了一种横向结构的忆 阻器,并可在超低能耗 (几十 fJ)下实现 STSP 的 人工突触计算.器件沟道中积累的焦耳热可以使其 电导发生改变,进而实现突触可塑性可调,如图 16(b) 和图 16(c)所示.为了抑制 ILD 对声音定位的干 扰,图 16(d)显示了通过抑制声音强度或 IPSC 来 检测双耳时间差,从而对声音源进行准确的定位.





Fig. 16. Sound localization and pattern recognition: (a) Schematic diagram of sound localization of ITD and ILD^[26]; (b) joule heat driven conductivity change under the sweep voltage^[26]; (c) the pulse intervals dependent PPF and PPD indexes stimulated by two consecutive pulses^[26]; (d) schematic diagram of the synaptic computing of ITD-based sound localization; "CA" refers to the cochlea shown in panel (a); the blue circles represent neurons; horizontal dashed lines represent potential thresholds for neuronal firing^[26]; (e) comparison of function and structure between biological synapses and vdWs hybrid synaptic devices^[167]; (f) acoustic pattern recognition rates based on three artificial neural networks with compared results achieved by the SW-NN recognition rate^[167].

这项工作在无需使用晶体管的情况下成功演示了 声音定位对声音源的识别过程.之后,Seo等^[167]开 发了一种具有线性和对称的电导更新特征的人工 vdWs 混合突触器件,成功实现了对声音模式的准 确识别 (90% 以上),如图 16(e)所示.Seo等通过 使用 WSe₂和 MoS₂杂化沟道实现了优异的电导可 控性,这两种沟道分别用于线性电导增强和抑制. 通过训练和推理模拟,证明了混合突触对硬件神经 网络 (HW-NN)的可行性,并提供了与软件神经网 络 (SW-NN) 相当的高识别率,如图 16(f)所示.

4.3 触觉模拟

触控操作是当前人与智能设备交互的主流方 式,其工作模式高度依赖于人体的触觉.人体皮肤 深层存在触觉小体,内有大量的神经细胞,当皮肤 感受到外界环境中压强、温度、湿度变化时就会发 出微小的电流信号,并随着神经纤维传输到大脑, 进而产生触觉^[20].

目前来说,开发高性能、制造工艺简单的人工 触控突触仍然是一个巨大的挑战.但随着压电电子 学的发展,研究者发现外界的机械刺激可以引起压 电极化和半导体传输特性之间的耦合效应,这种耦 合行为与人工神经元和传入神经中的耦合行为非 常相似^[201].因此,利用压电效应驱动的突触晶体 管成为实现并行感知和计算的重要途径.基于此, Chen 等^[202] 将压电纳米发电机 (PENG) 与离子凝 胶门控晶体管集成,报告了一种基于石墨烯的压电 突触器件.人体皮肤中的感受器接受机械刺激并 将其转化为突触前电位,并通过神经元和突触传递 到中枢神经系统,如图 17(a) 所示.其器件结构如 图 17(b) 所示, 由一个作为感觉组件的 PENG 和 一个作为人工突触的离子凝胶门控 Gr 场效应晶体 管组成.其中,信号传输过程是将机械应变触发的 压电势通过离子凝胶电容耦合到场效应晶体管来 实现的. 图 17(c)则进一步揭示了压电 Gr 人工感 觉突触的工作原理, 通过机械应变触发 PENG 产 生感应电场,使离子凝胶中的离子分布不均匀,从 而导致 Gr 沟道中的电导发生变化. 此外, 他们通 过对 PENG 施加的拉伸和压缩应变脉冲模拟了人 工突触可塑性,如图 17(d) 所示.这项工作为开发 神经形态计算中的压电电子突触器件提供了新的 思路,这对于未来研发具有人工智能的自供电电子 皮肤、神经机器人的神经形态接口、人机交互以及 智能压电晶体管等具有重要意义.



图 17 触觉模拟 (a) 生物触觉神经系统^[202]; (b) 带有离子凝胶门控晶体管的 PENG 示意图^[202]; (c) 压电 Gr 人工触觉突触工作 原理^[202]; (d) PSC 幅度与应变脉冲数的关系图^[202]

Fig. 17. Tactile mimicking: (a) Biological sensory nervous system^[202]; (b) schematic diagram of a PENG with iongel-gated transistors^[202]; (c) the working principle of piezoelectric Gr artificial sensory synapse^[202]; (d) PSC amplitude shown as the function of strain-pulse number^[202].

5 总结展望

2D 材料及其异质结的新奇物理特性, 使得神 经形态器件能够实现低工作电压以及高开关速度, 降低了能耗,提高了器件性能.另一方面,2D异质 结界面性质能够通过栅极电压进行调控,从而更好 地模拟神经元和突触功能.因此,2D材料及其异 质结对于构建具有多功能突触器件,设计合理的神 经形态集成电路,实现神经形态器件在柔性器件、 感知系统等方面的应用上显得至关重要.本文在综 述了神经形态器件工作的物理效应之外,还介绍 了 2D 材料在神经形态发展中的特点和优势. 另外, 对基于 2D 材料的神经形态器件实现人工突触功 能做了详细阐述,比如 STSP, LTSP 以及 STDP 等, 并且与传统神经形态器件相比,具有高集成度、低 功耗、权重可调、突触学习加速等优势.最后本文 对神经形态器件在视觉识别、声音定位、模式识别 及触觉模拟等方面的应用做了举例介绍.

虽然近些年基于 2D 材料的神经形态器件有 了重大研究进展,例如在单一忆阻器单元实现感 存算功能^[194]、超快感存算神经网络用于图像识 别^[203]、340 ℃ 高温极端环境下稳定性良好的异质 结忆阻器^[130]、高密度存储的忆阻器阵列等^[122],但 是对于实现人工智能神经形态计算应用体系走向 实际应用的角度来看,目前仍在材料制备、工艺兼 容、高密度集成、高效能算法构架等方面具有挑战 性^[204,205].

1)高质量均匀 2D 材料的晶圆级合成及无损 伤转移;大面积、高质量的 2D 晶体是开发下一代 光电器件的基础.而晶圆级 2D 材料的合成是产业 化应用的关键步骤.目前,CVD 可实现晶圆级合 成,并且操作相对简单,但是在 CVD 生长过程中 会引入不必要的缺陷,影响材料的质量,同时会使 得材料表面产生不可逆的粗糙度改变从而影响了 器件的性能^[116,206].此外,2D 材料之间清洁的界面 对器件性能有很大的限制.为了获得理想的 vdW 异质结,可以通过以下几种有效方法来获得洁净的 表面:在真空中退火,样品表面的残留物会显著减 少^[207],或者用氯仿溶解聚合物残留物^[208];另外, 通过原子力显微镜接触模式对样品表面机械清 洁^[209].在 vdW 异质结搭建过程中,层间气泡的产 生是很难避免的.但这可以使用热传导来抑制气泡 的产生,同时去除材料表面上的吸附物^[210].总之, 高质量均匀 2D 材料的晶圆级合成及无损伤转移 是器件制造过程的重要环节.

2) 与传统 CMOS 工艺的兼容性; 2D 材料及 其异质结能否集成到现有硅基器件在很大程度上 取决于它们与现有 CMOS 制造工艺的兼容性^[211]. 现有用于 2D 材料的 CVD 生长工艺需要 600— 900 ℃ 范围内的温度, 这限制了它们在已成型的 CMOS 基板上的直接生长^[212]. 尽管存在一些降 低生长温度的技术, 但所得材料的质量仍有待解 决^[213,214]. 因此, 解决神经形态计算器件的 CMOS 工艺兼容性问题, 是实现大规模集成并设计用于实 际应用的类脑芯片的基础.

3) 低功耗、高稳定性的器件开发; 突触器件的 能耗在神经形态计算器件的实现中起着至关重要 的作用. 沉积在 2D 材料上的金属电极会部分破坏 原子晶格中的共价键, 产生费米能级钉扎效应形成 肖特基接触, 导致接触电阻增加, 从而会引起功耗 的增加^[215,216]. 这需要通过改变掺杂或者接触方式 来降低 2D 材料与金属电极接触电阻产生的能耗. 另外, 通过降低操作电压, 以及降低漏电流、提高 电流的开关比也可明显地降低器件的功耗^[217].

因此,对于面向未来计算的神经形态器件集成 阵列,首先要制备出高质量、大面积、与传统 CMOS 工艺兼容的 2D 材料,并通过特定的物理效应设计 出高开关比、低功耗的神经形态器件,同时开发与 人类大脑相媲美的神经网络来存储和处理信息,最 后利用高密度的集成工艺构建出大规模、基于存算 一体架构的神经形态计算系统,延续摩尔定律,克 服冯·诺依曼瓶颈的限制,提高计算能力.

参考文献

- Liu C, Chen H, Wang S, Liu Q, Jiang Y G, Zhang D W, Liu M, Zhou P 2020 Nat. Nanotechnol. 15 545
- Horowitz M 2014 IEEE Int. Solid-State Circuits Conf. Digest Tech. Papers (ISSCC) San Francisco, USA, February 9–13, 2014 p10
- [3] Gibney E 2017 Nat. News **541** 142
- [4] Fukuda S 2020 World 2.0: From Working for Others to Working for Yourself (Berlin: Springer) p34
- [5] Zhang E, Wang W, Zhang C, Jin Y, Zhu G, Sun Q, Zhang D W, Zhou P, Xiu F 2015 ACS Nano 9 612
- [6] Zhou G D, Wang Z R, Sun B, Zhou F C, Sun L F, Zhao H B, Hu X F, Peng X Y, Yan J, Wang H M, Wang W H, Li J, Yan B T, Kuang D L, Wang Y C, Wang L D, Duan S K 2022 Adv. Electron. Mater. 8 2101127
- [7] Li H, Wang S, Zhang X, Wang W, Yang R, Sun Z, Feng W,

Lin P, Wang Z, Sun L 2021 Adv. Intell. Syst. 3 2100017

- [8] Sun L, Yu H, Wang D, Jiang J, Kim D, Kim H, Zheng S, Zhao M, Ge Q, Yang H 2018 2D Mater. 6 015029
- [9] Strukov D B, Snider G S, Stewart D R, Williams R S 2008 Nature 453 80
- [10] Shao N, Zhang S B, Shao S Y 2019 Acta Phys. Sin. 68 198502 (in Chinese) [邵楠, 张盛兵, 邵舒渊 2019 物理学报 68 198502]
- [11] Wang Z R, Rao M Y, Han J W, Zhang J M, Lin P, Li Y N, Li C, Song W H, Asapu S, Midya R, Jiang H, Yoon J H, Upadhyay N K, Qiu Q R, Williams R S, Xia Q F, Yang J J 2018 Nat. Commun. 9 3208
- [12] Chua L 1971 IEEE Trans. Circuit Theory 18 507
- [13] Yang D, Yang H, Guo X, Zhang H, Jiao C, Xiao W, Guo P, Wang Q, He D 2020 Adv. Funct. Mater. 30 2004514
- [14] Yang H, Wang Z, Guo X, Su H, Sun K, Yang D, Xiao W, Wang Q, He D 2020 ACS Appl. Mater. Interfaces 12 34370
- [15] Guo X, Wang Q, Lü X, Yang H, Sun K, Yang D, Zhang H, Hasegawa T, He D 2020 Nanoscale 12 4320
- [16] Jo S H, Chang T, Ebong I, Bhadviya B B, Mazumder P, Lu W 2010 Nano Lett. 10 1297
- [17] Liu Y C, Lin Y, Wang Z Q, Xu H Y 2019 Acta Phys. Sin.
 68 168504 (in Chinese) [刘益春, 林亚, 王中强, 徐海阳 2019 物理学报 68 168504]
- [18] Zhou Y, Li Y, Xu L, Zhong S, Sun H, Miao X 2015 Appl. Phys. Lett. 106 233502
- [19] Li C, Hu M, Li Y, Jiang H, Ge N, Montgomery E, Zhang J, Song W, Dávila N, Graves C E 2018 Nat. Electron. 1 52
- [20] Abuelma'atti M T, Khalifa Z J 2015 AEU-Int. J. Electron. C. 69 771
- [21] Wang Z, Wang L, Nagai M, Xie L, Yi M, Huang W 2017 Adv. Electron. Mater. 3 1600510
- [22] Yu Z Q, Liu M L, Lang J X, Qian K, Zhang C H 2018 Acta Phys. Sin. 67 157302 (in Chinese) [余志强, 刘敏丽, 郎建勋, 钱楷, 张昌华 2018 物理学报 67 157302]
- [23] Chang T, Jo S H, Lu W 2011 ACS Nano 5 7669
- [24] Xing Z W, Wu N J, Ignatiev A 2007 Appl. Phys. Lett. 91 052106
- [25] Wang Z Q, Xu H Y, Li X H, Yu H, Liu Y C, Zhu X J 2012 Adv. Funct. Mater. 22 2759
- [26] Sun L, Zhang Y S, Hwang G, Jiang J, Kim D, Eshete Y A, Zhao R, Yang H 2018 Nano Lett. 18 3229
- [27] Al-Saleh M H, Al-Anid H K, Husain Y A, El-Ghanem H M, Jawad S A 2013 J. Phys. D:Appl. Phys. 46 385305
- [28] Terabe K, Hasegawa T, Nakayama T, Aono M 2005 Nature 433 47
- [29] Hasegawa T, Terabe K, Tsuruoka T, Aono M 2012 Adv. Mater. 24 252
- [30] Pan C, Wang C Y, Liang S J, Wang Y, Cao T, Wang P, Wang C, Wang S, Cheng B, Gao A 2020 Nat. Electron. 3 383
- [31] Wang S, Pan X, Lyu L, Wang C Y, Wang P, Pan C, Yang Y, Wang C, Shi J, Cheng B 2022 ACS Nano 16 4528
- [32] Sun L, Yan J, Zhan D, Liu L, Hu H, Li H, Tay B K, Kuo J
 L, Huang C C, Hewak D W 2013 *Phys. Rev. Lett.* 111 126801
- [33] Lee C H, Lee G H, Van Der Zande A M, Chen W, Li Y, Han M, Cui X, Arefe G, Nuckolls C, Heinz T F 2014 Nat. Nanotechnol. 9 676
- [34] Lin Z, Liu Y, Halim U, Ding M, Liu Y, Wang Y, Jia C, Chen P, Duan X, Wang C 2018 Nature 562 254
- [35] Liu Y, Guo J, Zhu E, Liao L, Lee S J, Ding M, Shakir I, Gambin V, Huang Y, Duan X 2018 Nature 557 696

- [36] Chen S, Mahmoodi M R, Shi Y, Mahata C, Yuan B, Liang X, Wen C, Hui F, Akinwande D, Strukov D B 2020 Nat. Electron. 3 638
- [37] Wang M, Wang C Y, Wu C, Li Q, Pan C, Wang C, Liang S J, Miao F 2019 Adv. Electron. Mater. 5 1800853
- [38] Chen H, Xue X, Liu C, Fang J, Wang Z, Wang J, Zhang D W, Hu W, Zhou P 2021 Nat. Electron. 4 399
- [39] Fu X, Zhang L, Cho H D, Kang T W, Fu D, Lee D, Lee S W, Li L, Qi T, Chan A S 2019 *Small* **15** 1903809
- [40] Zhu X, Li D, Liang X, Lu W D 2019 Nat. Mater. 18 141
- [41] Zhang F, Zhang H, Krylyuk S, Milligan C A, Zhu Y, Zemlyanov D Y, Bendersky L A, Burton B P, Davydov A V, Appenzeller J 2019 Nat. Mater. 18 55
- [42] Kwon K C, Zhang Y, Wang L, Yu W, Wang X, Park I H, Choi H S, Ma T, Zhu Z, Tian B 2020 ACS Nano 14 7628
- [43] Wang L, Wang X, Zhang Y, Li R, Ma T, Leng K, Chen Z, Abdelwahab I, Loh K P 2020 Adv. Funct. Mater. 30 2004609
- [44] Wang H, Lu W, Hou S, Yu B, Zhou Z, Xue Y, Guo R, Wang S, Zeng K, Yan X 2020 Nanoscale 12 21913
- [45] Jang B C, Kim S, Yang S Y, Park J, Cha J H, Oh J, Choi J, Im S G, Dravid V P, Choi S Y 2019 Nano Lett. 19 839
- [46] Xu R, Jang H, Lee M H, Amanov D, Cho Y, Kim H, Park S, Shin H J, Ham D 2019 Nano Lett. 19 2411
- [47] Yan X, Qin C, Lu C, Zhao J, Zhao R, Ren D, Zhou Z, Wang H, Wang J, Zhang L 2019 ACS Appl. Mater. Interfaces 11 48029
- [48] Wu X, Ge R, Chen P A, Chou H, Zhang Z, Zhang Y, Banerjee S, Chiang M H, Lee J C, Akinwande D 2019 Adv. Mater. 31 1806790
- [49] Yan X, Zhao Q, Chen A P, Zhao J, Zhou Z, Wang J, Wang H, Zhang L, Li X, Xiao Z 2019 Small 15 1901423
- [50] Liu L, Li Y, Huang X, Chen J, Yang Z, Xue K H, Xu M, Chen H, Zhou P, Miao X 2021 Adv. Sci. 8 2005038
- [51] Vu Q A, Shin Y S, Kim Y R, Nguyen V L, Kang W T, Kim H, Luong D H, Lee I M, Lee K, Ko D S, Heo J, Park S, Lee Y H, Yu W J 2016 Nat. Commun. 7 12725
- [52] Liu C, Yan X, Song X, Ding S, Zhang D W, Zhou P 2018 Nat. Nanotechnol. 13 404
- [53] Xiang D, Liu T, Zhang X, Zhou P, Chen W 2021 Nano Lett. 21 3557
- [54] Jin T, Zheng Y, Gao J, Wang Y, Li E, Chen H, Pan X, Lin M, Chen W 2021 ACS Appl. Mater. Interfaces 13 10639
- [55] He C, Tang J, Shang D S, Tang J, Xi Y, Wang S, Li N, Zhang Q, Lu J K, Wei Z 2020 ACS Appl. Mater. Interfaces 12 11945
- [56] Wang S, Chen C, Yu Z, He Y, Chen X, Wan Q, Shi Y, Zhang D W, Zhou H, Wang X 2019 Adv. Mater. 31 1806227
- [57] Tran M D, Kim H, Kim J S, Doan M H, Chau T K, Vu Q A, Kim J H, Lee Y H 2019 Adv. Mater. 31 1807075
- [58] Qin S, Wang F, Liu Y, Wan Q, Wang X, Xu Y, Shi Y, Wang X, Zhang R 2017 2 D Mater. 4 035022
- [59] Ma Y, Liu B, Zhang A, Chen L, Fathi M, Shen C, Abbas A N, Ge M, Mecklenburg M, Zhou C 2015 ACS Nano 9 7383
- [60] Cho S, Kim S, Kim J H, Zhao J, Seok J, Keum D H, Baik J, Choe D H, Chang K J, Suenaga K 2015 *Science* 349 625
- [61] Lin Y C, Dumcenco D O, Huang Y S, Suenaga K 2014 Nat. Nanotechnol. 9 391
- [62] Kappera R, Voiry D, Yalcin S E, Branch B, Gupta G, Mohite A D, Chhowalla M 2014 Nat. Mater. 13 1128
- [63] Loke D, Lee T, Wang W, Shi L, Zhao R, Yeo Y, Chong T, Elliott S 2012 Science 336 1566
- [64] Jeon H, Kim S G, Park J, Kim S H, Park E, Kim J, Yu H Y 2020 Small 16 2004371

- [65] Wang Z, Liu X, Zhou X, Yuan Y, Zhou K, Zhang D, Luo H, Sun J 2022 Adv. Mater. 34 2200032
- [66] Yan M, Zhu Q, Wang S, Ren Y, Feng G, Liu L, Peng H, He Y, Wang J, Zhou P 2021 Adv. Electron. Mater. 7 2001276
- [67] Luo Q, Cheng Y, Yang J G, Cao R R, Ma H L, Yang Y, Huang R, Wei W, Zheng Y H, Gong T C, Yu J, Xu X X, Yuan P, Li X Y, Tai L, Yu H R, Shang D S, Liu Q, Yu B, Ren Q W 2020 Nat. Commun. 11 1391
- [68] Oh S, Hwang H, Yoo I K 2019 APL Mater. 7 091109
- [69] Upadhyay N K, Jiang H, Wang Z R, Asapu S, Xia Q F, Joshua Yang J J 2019 Adv. Mater. Technol. 4 1800589
- [70] Kim M K, Lee J S 2019 *Nano Lett.* **19** 2044
- [71] Oh S, Kim T, Kwak M, Song J, Woo J, Jeon S, Yoo I K, Hwang H 2017 IEEE Electron Device Lett. 38 732
- [72] Luo Z D, Zhang S, Liu Y, Zhang D, Gan X, Seidel J, Liu Y, Han G, Alexe M, Hao Y 2022 ACS Nano 16 3362
- [73] Long S B, Perniola L, Cagli C, Buckley J, Lian X J, Miranda E, Pan F, Liu M, Suñé J 2013 Sci. Rep. 3 2929
- [74] Yang Y C, Gao P, Gaba S, Chang T, Pan X Q, Lu W 2012 *Nat. Commun.* **3** 732
- [75] Sharath S U, Vogel S, Molina-Luna L, Hildebrandt E, Wenger C, Kurian J, Duerrschnabel M, Niermann T, Niu G, Calka P 2017 Adv. Funct. Mater. 27 1700432
- [76] Jeong D S, Kim K M, Kim S, Choi B J, Hwang C S 2016 Adv. Electron. Mater. 2 1600090
- [77] Sun L, Hwang G, Choi W, Han G, Zhang Y, Jiang J, Zheng S, Watanabe K, Taniguchi T, Zhao M 2020 Nano Energy 69 104472
- [78] Guo J, Wang L, Liu Y, Zhao Z, Zhu E, Lin Z, Wang P, Jia C, Yang S, Lee S J 2020 Matter 2 965
- [79] Wang L, Wang Z, Zhao W, Hu B, Xie L, Yi M, Ling H, Zhang C, Chen Y, Lin J 2017 Adv. Electron. Mater. 3 1600244
- [80] Li Q, Tao Q, Chen Y, Kong L, Shu Z, Duan H, Liao L, Liu Y 2021 Int. J. Extreme Manuf. 3 045103
- [81] Liu Y, Huang Y, Duan X 2019 Nature 567 323
- [82] Jariwala D, Marks T J, Hersam M C 2017 Nat. Mater. 16 170
- [83] Zhao X, Ma J, Xiao X, Liu Q, Shao L, Chen D, Liu S, Niu J, Zhang X, Wang Y 2018 Adv. Mater. 30 1705193
- [84] Yoon J H, Zhang J, Ren X, Wang Z, Wu H, Li Z, Barnell M, Wu Q, Lauhon L J, Xia Q 2017 Adv. Funct. Mater. 27 1702010
- [85] Lenser C, Kuzmin A, Purans J, Kalinko A, Waser R, Dittmann R 2012 Appl. Phys. Lett. 111 076101
- [86] Li Z, Tian B, Xue K H, Wang B, Xu M, Lu H, Sun H, Miao X 2019 IEEE Electron Device Lett. 40 1068
- [87] Zhang W, Huang Z, Zhang W, Li Y 2014 Nano Res. 7 1731
- [88] Chen J, Guo R, Wang X, Zhu C, Cao G, You L, Duan R, Zhu C, Hadke S S, Cao X 2022 ACS Nano 16 221
- [89] Lee G H, Yu Y J, Lee C, Dean C, Shepard K L, Kim P, Hone J 2011 Appl. Phys. Lett. 99 243114
- [90] Cumings J, Zettl A 2004 Solid State Commun. 129 661
- [91] Wang S, He C, Tang J, Lu X, Shen C, Yu H, Du L, Li J, Yang R, Shi D 2019 Adv. Electron. Mater. 5 1800726
- [92] Bertolazzi S, Krasnozhon D, Kis A 2013 ACS Nano 7 3246
- [93] Choi M S, Lee G H, Yu Y J, Lee D Y, Lee S H, Kim P, Hone J, Yoo W J 2013 Nat. Commun. 4 1624
- [94] Li D, Wang X, Zhang Q, Zou L, Xu X, Zhang Z 2015 Adv. Funct. Mater. 25 7360
- [95] Wu L, Wang A, Shi J, Yan J, Zhou Z, Bian C, Ma J, Ma R, Liu H, Chen J 2021 Nat. Nanotechnol. 16 882
- [96] Lai H J, Zhou Y, Zhou H B, Zhang N, Ding X D, Liu P Y,

Wang X M, Xie W G 2022 Adv. Mater. 34 2110278

- [97] Yoon J H, Kim K M, Song S J, Seok J Y, Yoon K J, Kwon D E, Park T H, Kwon Y J, Shao X, Hwang C S 2015 Adv. Mater. 27 3811
- [98] Yao Y, Li C, Huo Z L, Liu M, Zhu C X, Gu C Z, Duan X F, Wang Y G, Gu L, Yu R C 2013 *Nat. Commun.* 4 2764
- [99] Michalas L, Stathopoulos S, Khiat A, Prodromakis T 2018 Appl. Phys. Lett. 113 143503
- [100] Li Y, Zhong Y P, Xu L, Zhang J J, Xu X H, Sun H J, Miao X S 2013 Sci. Rep. 3 1619
- [101] Maehne H, Wylezich H, Hanzig F, Slesazeck S, Rafaja D, Mikolajick T 2014 Semicond. Sci. Technol. 29 104002
- [102] Kim H J, Zheng H, Park J S, Kim D H, Kang C J, Jang J T, Kim D H, Yoon T S 2017 Nanotechnology 28 285203
- [103] Seo S, Jo S H, Kim S, Shim J, Oh S, Kim J H, Heo K, Choi J W, Choi C, Oh S 2018 *Nat. Commun.* 9 1
- [104] Arnold A J, Razavieh A, Nasr J R, Schulman D S, Eichfeld C M, Das S 2017 ACS Nano 11 3110
- [105] He H K, Yang R, Zhou W, Huang H M, Xiong J, Gan L, Zhai T Y, Guo X 2018 *Small* 14 1800079
- [106] Pei Z, Chung A, Hwang H 2007 Appl. Phys. Lett. 90 223513
- [107] Kim E, Yim T, An S, Cho W J, Park K 2010 Appl. Phys. Lett. 97 222107
- [108] Lee J, Pak S, Lee Y W, Cho Y, Hong J, Giraud P, Shin H S, Morris S M, Sohn J I, Cha S, Kim J M 2017 Nat. Commun. 8 14734
- [109] Wen J, Tang W, Kang Z, Liao Q, Hong M, Du J, Zhang X, Yu H, Si H, Zhang Z 2021 Adv. Sci. 8 2101417
- [110] Britnell L, Gorbachev R, Jalil R, Belle B, Schedin F, Mishchenko A, Georgiou T, Katsnelson M, Eaves L, Morozov S 2012 Science 335 947
- [111] Massicotte M, Schmidt P, Vialla F, Schädler K G, Reserbat-Plantey A, Watanabe K, Taniguchi T, Tielrooij K J, Koppens F H L 2015 Nat. Nanotechnol. 11 42
- [112] Vu Q A, Lee J H, Nguyen V L, Shin Y S, Lim S C, Lee K, Heo J, Park S, Kim K, Lee Y H 2017 Nano Lett. 17 453
- [113] Furchi M M, Höller F, Dobusch L, Polyushkin D K, Schuler S, Mueller T 2018 NPJ 2D Mater. Appl. 2 3
- [114] Tian H, Cao X, Xie Y, Yan X, Kostelec A, DiMarzio D, Chang C, Zhao L D, Wu W, Tice J 2017 ACS Nano 11 7156
- [115] Duan X, Wang C, Shaw J C, Cheng R, Chen Y, Li H, Wu X, Tang Y, Zhang Q, Pan A 2014 Nat. Nanotechnol. 9 1024
- [116] Geim A K, Grigorieva I V 2013 Nature 499 419
- [117] Jin C, Kim J, Suh J, Shi Z, Chen B, Fan X, Kam M, Watanabe K, Taniguchi T, Tongay S 2017 Nat. Phys. 13 127
- [118] Zhong D, Seyler K L, Linpeng X, Cheng R, Sivadas N, Huang B, Schmidgall E, Taniguchi T, Watanabe K, McGuire M A 2017 *Sci. Adv.* **3** e1603113
- [119] Kunstmann J, Mooshammer F, Nagler P, Chaves A, Stein F, Paradiso N, Plechinger G, Strunk C, Schüller C, Seifert G 2018 Nat. Phys. 14 801
- [120] Li D, Chen M, Sun Z, Yu P, Liu Z, Ajayan P M, Zhang Z 2017 Nat. Nanotechnol. 12 901
- [121] Li H H, Xiong X L, Hui F, Yang D L, Jiang J B, Feng W X, Han J F, Duan J X, Wang Z R, Sun L F 2022 Nanotechnology 33 465601
- [122] Sun L F, Zhang Y S, Han G, Hwang G, Jiang J B, Joo B, Watanabe K, Taniguchi T, Kim Y M, Yu W J, Kong B S, Zhao R, Yang H 2019 Nat. Commun. 10 3161
- [123] Cai Z, Cao M, Jin Z, Yi K, Chen X, Wei D 2018 NPJ 2 D Mater. Appl. 2 21
- [124] Tran M D, Kim J H, Kim H, Doan M H, Duong D L, Lee Y

H 2018 ACS Appl. Mater. Interfaces $\mathbf{10}$ 10580

- [125] Pi L, Wang P, Liang S J, Luo P, Wang H, Li D, Li Z, Chen P, Zhou X, Miao F 2022 Nat. Electron. 5 248
- [126] Gbadamasi S, Mohiuddin M, Krishnamurthi V, Verma R, Khan M W, Pathak S, Kalantar-Zadeh K, Mahmood N 2021 *Chem. Soc. Rev.* 50 4684
- [127] Li M Y, Shi Y, Cheng C C, Lu L S, Lin Y C, Tang H L, Tsai M L, Chu C W, Wei K H, He J H 2015 Science 349 524
- [128] Huang C, Wu S, Sanchez A M, Peters J J, Beanland R, Ross J S, Rivera P, Yao W, Cobden D H, Xu X 2014 Nat. Mater. 13 1096
- [129] He H K, Yang R, Huang H M, Yang F F, Wu Y Z, Shaibo J, Guo X 2020 Nanoscale 12 380
- [130] Wang M, Cai S, Pan C, Wang C, Lian X, Zhuo Y, Xu K, Cao T, Pan X, Wang B 2018 Nat. Electron. 1 130
- [131] Chen C, Song C, Yang J, Zeng F, Pan F 2012 Appl. Phys. Lett. 100 253509
- [132] Pickett M D, Medeiros-Ribeiro G, Williams R S 2013 Nat. Mater. 12 114
- [133] Hodgkin A L, Huxley A F 1952 Philos. Trans. R. Soc. London, Ser. B 140 177
- [134] Izhikevich E M 2003 IEEE Trans. Neural Networks 14 1569
- [135] Fourcaud-Trocmé N, Hansel D, Van Vreeswijk C, Brunel N 2003 J. Neurosci. 23 11628
- [136] Brette R, Gerstner W 2005 J. Neurophysiol. 94 3637
- [137] Jolivet R, Gerstner W 2003 Joint International Conference ICANN/ICONIP Istanbul, Turkey, June 26–29, 2003 p846
- [138] Deng L, Wu Y, Hu X, Liang L, Ding Y, Li G, Zhao G, Li P, Xie Y 2020 Neural Networks 121 294
- [139] Hao S, Ji X, Zhong S, Pang K Y, Lim K G, Chong T C, Zhao R 2020 Adv. Electron. Mater. 6 1901335
- [140] Dev D, Krishnaprasad A, Shawkat M S, He Z, Das S, Fan D, Chung H S, Jung Y, Roy T 2020 *IEEE Electron Device Lett.* 41 936
- [141] Beck M E, Shylendra A, Sangwan V K, Guo S, Rojas W A G, Yoo H, Bergeron H, Su K, Trivedi A R, Hersam M C 2020 Nat. Commun. 11 1565
- [142] Wang Z, Joshi S, Savel'ev S E, Jiang H, Midya R, Lin P, Hu M, Ge N, Strachan J P, Li Z 2017 Nat. Mater. 16 101
- [143] Zhao H, Dong Z, Tian H, DiMarzi D, Han M G, Zhang L, Yan X, Liu F, Shen L, Han S J 2017 Adv. Mater. 29 1703232
- [144] Budiman F, Hernowo D G O, Pandey R R, Tanaka H 2018 Jpn. J. Appl. Phys. 57 03EA06
- [145] Ding G, Yang B, Chen R S, Mo W A, Zhou K, Liu Y, Shang G, Zhai Y, Han S T, Zhou Y 2021 Small 17 2103175
- [146] Zhu L Q, Wan C J, Guo L Q, Shi Y, Wan Q 2014 Nat. Commun. 5 3158
- [147] Van De Burgt Y, Lubberman E, Fuller E J, Keene S T, Faria G C, Agarwal S, Marinella M J, Alec Talin A, Salleo A 2017 Nat. Mater. 16 414
- [148] Sangwan V K, Jariwala D, Kim I S, Chen K-S, Marks T J, Lauhon L J, Hersam M C 2015 Nat. Nanotechnol. 10 403
- [149] Zhu J, Yang Y, Jia R, Liang Z, Zhu W, Rehman Z U, Bao L, Zhang X, Cai Y, Song L 2018 Adv. Mater. 30 1800195
- [150] Burr G W, Shelby R M, Sidler S, Di Nolfo C, Jang J, Boybat I, Shenoy R S, Narayanan P, Virwani K, Giacometti E U 2015 *IEEE Trans. Electron Devices* **62** 3498
- [151] Wang S Y, Liu L, Gan L R, Chen H W, Hou X, Ding Y, Ma S L, Zhang D W, Zhou P 2021 Nat. Commun. 12 53
- [152] Jiang J, Guo J, Wan X, Yang Y, Xie H, Niu D, Yang J, He J, Gao Y, Wan Q 2017 *Small* **13** 1700933
- [153] Bao L, Zhu J, Yu Z, Jia R, Cai Q, Wang Z, Xu L, Wu Y,

Yang Y, Cai Y 2019 ACS Appl. Mater. Interfaces 11 41482

- [154] Gao J, Lian X, Chen Z X, Shi S, Li E L, Wang Y A, Jin T Y, Chen H P, Liu L, Chen J S, Zhu Y, Chen W 2022 Adv. Funct. Mater. 32 2110415
- [155] Caporale N, Dan Y 2008 Annu. Rev. Neurosci. **31** 25
- [156] Sun L, Wang W, Yang H 2020 Adv. Intell. Syst. 2 1900167
- [157] Li Y, Zhong Y P, Zhang J J, Xu L, Wang Q, Sun H J, Tong H, Cheng X M, Miao X S 2014 *Sci. Rep.* 4 4906
- [158] Majumdar S, Tan H, Pande I, Van Dijken S 2019 APL Mater. 7 091114
- [159] Zappacosta S, Mannella F, Mirolli M, Baldassarre G 2018 PLoS Comput. Biol. 14 e1006227
- [160] Lee G, Baek J H, Ren F, Pearton S J, Lee G H, Kim J 2021 Small 17 2100640
- [161] Chen L, Wang L, Peng Y, Feng X, Sarkar S, Li S, Li B, Liu L, Han K, Gong X 2020 Adv. Electron. Mater. 6 2000057
- [162] Tang J, Yuan F, Shen X, Wang Z, Rao M, He Y, Sun Y, Li X, Zhang W, Li Y 2019 Adv. Mater. 31 1902761
- [163] Rachmuth G, Shouval H Z, Bear M F, Poon C S 2011 Proc. Natl. Acad. Sci. U.S.A. 108 E1266
- [164] Wang X, Wang B, Zhang Q, Sun Y, Wang E, Luo H, Wu Y, Gu L, Li H, Liu K 2021 Adv. Mater. 33 2102435
- [165] Liu Y, Li E, Wang X, Chen Q, Zhou Y, Hu Y, Chen G, Chen H, Guo T 2020 Nano Energy 78 105403
- [166] Feng X, Li Y, Wang L, Chen S, Yu Z G, Tan W C, Macadam N, Hu G, Huang L, Chen L 2019 Adv. Electron. Mater. 5 1900740
- [167] Seo S, Kang B S, Lee J J, Ryu H J, Kim S, Kim H, Oh S, Shim J, Heo K, Oh S, Park J H 2020 Nat. Commun. 11 3936
- [168] Yan X, Zhao J, Liu S, Zhou Z, Liu Q, Chen J, Liu X Y 2018 Adv. Funct. Mater. 28 1705320
- [169] Ryu J H, Kim B, Hussain F, Ismail M, Mahata C, Oh T, Imran M, Min K K, Kim T H, Yang B D 2020 IEEE Access 8 130678
- [170] Sun J, Oh S, Choi Y, Seo S, Oh M J, Lee M, Lee W B, Yoo P J, Cho J H, Park J H 2018 Adv. Funct. Mater. 28 1804397
- [171] Jin C, Liu W, Huang Y, Xu Y, Nie Y, Zhang G, He P, Sun J, Yang J 2022 Appl. Phys. Lett. 120 233701
- [172] Jin C, Liu W, Xu Y, Huang Y, Nie Y, Shi X, Zhang G, He P, Zhang J, Cao H 2022 *Nano Lett.* **22** 3372
- [173] Kwon S M, Cho S W, Kim M, Heo J S, Kim Y H, Park S K 2019 Adv. Mater. **31** 1906433
- [174] Qiu W, Huang Y, Kong L A, Chen Y, Liu W, Wang Z, Sun J, Wan Q, Cho J H, Yang J 2020 Adv. Funct. Mater. 30 2002325
- [175] Wang J, Chen Y, Kong L A, Fu Y, Gao Y, Sun J 2018 Appl. Phys. Lett. 113 151101
- [176] Chen Y, Qiu W, Wang X, Liu W, Wang J, Dai G, Yuan Y, Gao Y, Sun J 2019 Nano Energy 62 393
- [177] Fang L, Dai S, Zhao Y, Liu D, Huang J 2020 Adv. Electron. Mater. 6 1901217
- [178] Lee S H, Park H L, Kim M H, Kim M H, Park B G, Lee S D 2020 ACS Appl. Mater. Interfaces 12 51719
- [179] Jung K H, Yeon C, Yang J, Cheon Y J, Lim J W, Yun S J 2021 ACS Appl. Mater. Interfaces 13 8919
- [180] Wang H, Zhao Q, Ni Z, Li Q, Liu H, Yang Y, Wang L, Ran Y, Guo Y, Hu W 2018 Adv. Mater. **30** 1803961
- [181] Qian C, Sun J, Kong L A, Gou G, Yang J, He J, Gao Y, Wan Q 2016 ACS Appl. Mater. Interfaces 8 26169
- [182] Guo R, Zhou Y, Wu L, Wang Z, Lim Z, Yan X, Lin W, Wang H, Yoong H Y, Chen S 2018 ACS Appl. Mater. Interfaces 10 12862
- [183] Yoon C, Lee J H, Lee S, Jeon J H, Jang J T, Kim D H, Kim

Y H, Park B H 2017 Nano Lett. 17 1949

- [184] Boyn S, Grollier J, Lecerf G, Xu B, Locatelli N, Fusil S, Girod S, Carrétéro C, Garcia K, Xavier S, Tomas J, Bellaiche L, Bibes M, Barthelemy A, Saighi S, Garcia V 2017 Nat. Commun. 8 14736
- [185] Yang J M, Choi E S, Kim S Y, Kim J H, Park J H, Park N G 2019 Nanoscale 11 6453
- [186] Cao G, Meng P, Chen J, Liu H, Bian R, Zhu C, Liu F, Liu Z 2021 Adv. Funct. Mater. 31 2005443
- [187] Ji X, Zhao X, Tan M C, Zhao R 2020 Adv. Intell. Syst. 2 1900118
- [188] Liu S, Wang Y, Fardad M, Varshney P K 2018 IEEE Circuits Syst. Mag. 18 29
- [189] Yoon J H, Wang Z R, Kim K M, Wu H Q, Ravichandran V, Xia Q F, Hwang C S, Yang J J 2018 Nat. Commun. 9 417
- [190] Yu S, Wu Y, Jeyasingh R, Kuzum D, Wong H S P 2011 IEEE Trans. Electron Devices 58 2729
- [191] Prezioso M, Merrikh-Bayat F, Hoskins B, Adam G C, Likharev K K, Strukov D B 2015 Nature 521 61
- [192] Wang C, Liang S J, Wang C Y, Yang Z Z, Ge Y, Pan C, Shen X, Wei W, Zhao Y, Zhang Z 2021 Nat. Nanotechnol. 16 1079
- [193] Wang C Y, Liang S J, Wang S, Wang P, Li Z A, Wang Z, Gao A, Pan C, Liu C, Liu J 2020 Sci. Adv. 6 eaba6173
- [194] Zhou F, Zhou Z, Chen J, Choy T H, Wang J, Zhang N, Lin Z, Yu S, Kang J, Wong H S P 2019 Nat. Nanotechnol. 14 776
- [195] Sun L, Wang Z, Jiang J, Kim Y, Joo B, Zheng S, Lee S, Yu W J, Kong B S, Yang H 2021 Sci. Adv. 7 eabg1455
- [196] Du J, Xie D, Zhang Q, Zhong H, Meng F, Fu X, Sun Q, Ni
 H, Li T, Guo E J 2021 Nano Energy 89 106439
- [197] Admoni H, Scassellati B 2017 J. Hum. Robot Interact. 6 25
- [198] Yang G Z, Bellingham J, Dupont P E, Fischer P, Floridi L, Full R, Jacobstein N, Kumar V, McNutt M, Merrifield R, Nelson B J 2018 The Grand Challenges of Science Robotics. Sci. Robot. 3 eaar7650
- [199] Agmon-Snir H, Carr C E, Rinzel J 1998 Nature 393 268
- [200] Dabdoub A, Fritzsch B 2016 Audit. Neur. Their Net. 52 1
- [201] Wu W, Wang L, Li Y, Zhang F, Lin L, Niu S, Chenet D,

Zhang X, Hao Y, Heinz T F 2014 Nature 514 470

[202] Chen Y, Gao G, Zhao J, Zhang H, Yu J, Yang X, Zhang Q, Zhang W, Xu S, Sun J 2019 Adv. Funct. Mater. 29 1900959

- [203] Mennel L, Symonowicz J, Wachter S, Polyushkin D K, Molina-Mendoza A J, Mueller T 2020 Nature 579 62
- [204] Ren K, Zhang K J, Qin X Z, Ren H X, Zhu S H, Yang F, Sun B, Zhao Y, Zhang Y 2021 Acta Phys. Sin 70 078701 (in Chinese) [任宽, 张珂嘉, 秦溪子, 任焕鑫, 朱守辉, 杨峰, 孙柏, 赵勇, 张勇 2021 物理学报 70 078701]
- [205] Ren K, Zhang W Y, Wang F, Guo Z Y, Shang D S 2022 Acta Phys. Sin 71 140701 (in Chinese) [任宽, 张握瑜, 王菲, 郭泽钰, 尚大山 2022 物理学报 71 140701]
- [206] Haigh S J, Gholinia A, Jalil R, Romani S, Britnell L, Elias D C, Novoselov K S, Ponomarenko L A, Geim A K, Gorbachev R 2012 Nat. Mater. 11 764
- [207] Pirkle A, Chan J, Venugopal A, Hinojos D, Magnuson C, McDonnell S, Colombo L, Vogel E, Ruoff R, Wallace R 2011 *Appl. Phys. Lett.* **99** 122108
- [208] Cheng Z, Zhou Q, Wang C, Li Q, Wang C, Fang Y 2011 Nano Lett. 11 767
- [209] Goossens A, Calado V, Barreiro A, Watanabe K, Taniguchi T, Vandersypen L 2012 Appl. Phys. Lett. 100 073110
- [210] Fan S, Vu Q A, Tran M D, Adhikari S, Lee Y H 2020 2 D Mater. 7 022005
- [211] Salahuddin S, Ni K, Datta S 2018 Nat. Electron. 1 442
- [212] Cai Z, Liu B, Zou X, Cheng H M 2018 Chem. Rev. 118 6091
- [213] Jang J, Son M, Chung S, Kim K, Cho C, Lee B H, Ham M H 2015 Sci. Rep. 5 17955
- [214] Sun L, Yuan G, Gao L, Yang J, Chhowalla M, Gharahcheshmeh M H, Gleason K K, Choi Y S, Hong B H, Liu Z 2021 Nat. Rev. Methods Primers 1 5
- [215] Das S, Sebastian A, Pop E, McClellan C J, Franklin A D, Grasser T, Knobloch T, Illarionov Y, Penumatcha A V, Appenzeller J 2021 Nat. Electron. 4 786
- [216] Wang Y, Liu S, Li Q, Quhe R, Yang C, Guo Y, Zhang X, Pan Y, Li J, Zhang H 2021 *Rep. Prog. Phys.* 84 056501
- [217] Bernstein K, Cavin R K, Porod W, Seabaugh A, Welser J 2010 Proc. IEEE 98 2169

SPECIAL TOPIC—Physical electronics for brain-inspired computing

Research progress of neuromorphic devices based on two-dimensional layered materials^{*}

Li Ce Yang Dong-Liang[†] Sun Lin-Feng[‡]

(Key Laboratory of Advanced Optoelectronic Quantum Architecture and Measurement (MOE),
 School of Physics, Beijing Institute of Technology, Beijing 100081, China)
 (Received 16 July 2022; revised manuscript received 13 August 2022)

Abstract

In recent years, the development of artificial intelligence has increased the demand for computing and storage. However, the slowing down of Moore's law and the separation between computing and storage units in traditional von Neumann architectures result in the increase of power consumption and time delays in the transport of abundant data, raising more and more challenges for integrated circuit and chip design. It is urgent for us to develop new computing paradigms to meet this challenge. The neuromorphic devices based on the inmemory computing architecture can overcome the traditional von Neumann architecture by Ohm's law and Kirchhoff's current law. By adjusting the resistance value of the memristor, the artificial neural network which can mimic the biological brain will be realized, and complex signal processing such as image recognition, pattern classification and decision determining can be carried out. In order to further reduce the size of device and realize the integration of sensing, memory and computing, two-dimensional materials can provide a potential solution due to their ultrathin thickness and rich physical effects. In this paper, we review the physical effects and memristive properties of neuromorphic devices based on two-dimensional materials, and describe the synaptic plasticity of neuromorphic devices based on leaky integrate and fire model and Hodgkin-Huxley model in detail, including long-term synaptic plasticity, short-term synaptic plasticity, spiking-time-dependent plasticity and spiking-rate-dependent plasticity. Moreover, the potential applications of two-dimensional materials based neuromorphic devices in the fields of vision, audition and tactile are introduced. Finally, we summarize the current issues on two-dimensional materials based neuromorphic computing and give the prospects for their future applications.

Keywords: neuromorphic device, artificial neural network, memristor, two-dimensional material

PACS: 85.35.-p, 73.40.Rw, 85.25.Hv, 72.20.-i

DOI: 10.7498/aps.71.20221424

^{*} Project Supported by the Key Research Program of Beijing Natural Science Foundation, China (Grant No. Z210006) and the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 12104051).

[†] Corresponding author. E-mail: yangdl@bit.edu.cn

 $[\]ddagger$ Corresponding author. E-mail: <code>sunlinfeng@bit.edu.cn</code>

物理学报Acta Physica Sinica





Institute of Physics, CAS

基于二维层状材料的神经形态器件研究进展

李策 杨栋梁 孙林锋

Research progress of neuromorphic devices based on two-dimensional layered materials Li Ce Yang Dong-Liang Sun Lin-Feng

引用信息 Citation: Acta Physica Sinica, 71, 218504 (2022) DOI: 10.7498/aps.71.20221424

在线阅读 View online: https://doi.org/10.7498/aps.71.20221424

当期内容 View table of contents: http://wulixb.iphy.ac.cn

您可能感兴趣的其他文章

Articles you may be interested in

基于二维材料MXene的仿神经突触忆阻器的制备和长/短时程突触可塑性的实现

Fabrication of synaptic memristor based on two-dimensional material MXene and realization of both long-term and short-term plasticity

物理学报. 2019, 68(9): 098501 https://doi.org/10.7498/aps.68.20182306

基于二维材料的全光器件

All-optical devices based on two-dimensional materials 物理学报. 2020, 69(18): 184216 https://doi.org/10.7498/aps.69.20200654

新型忆阻器神经形态电路的设计及其在条件反射行为中的应用

Design of novel memristor-based neuromorphic circuit and its application in classical conditioning 物理学报. 2019, 68(23): 238501 https://doi.org/10.7498/aps.68.20191023

N型局部有源忆阻器的神经形态行为

Neuromorphic behaviors of N-type locally-active memristor 物理学报. 2022, 71(5): 050502 https://doi.org/10.7498/aps.71.20212017

基于忆容器件的神经形态计算研究进展

Research progress of neuromorphic computation based on memcapacitors 物理学报. 2021, 70(7): 078701 https://doi.org/10.7498/aps.70.20201632

氧化物基忆阻型神经突触器件

Oxide-based memristive neuromorphic synaptic devices 物理学报. 2019, 68(16): 168504 https://doi.org/10.7498/aps.68.20191262