

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

* 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: zhoulfc@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. 视网膜内感光细胞和神经节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, 以此完成边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 10^4 .

由 HfO_2 ^[26], TaO_x ^[27], TiO_2 ^[28], SrTiO_3 ^[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

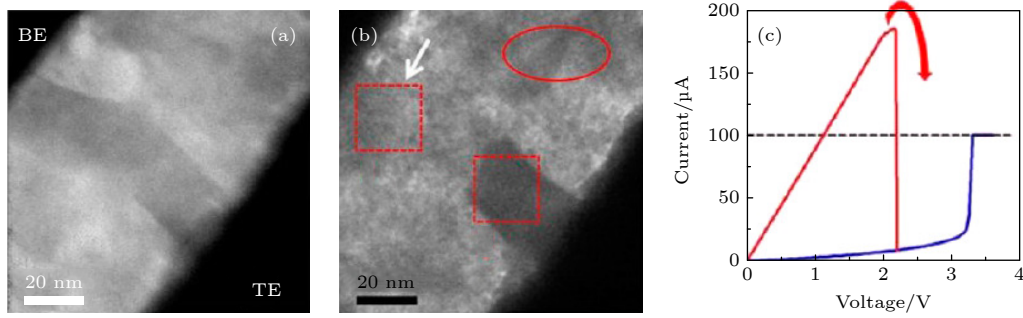


图 1 氧空位导电丝型 RRAM^[31] (a) Pt/ZnO/Pt RRAM 内生成的氧空位导电丝; (b) Pt/ZnO/Pt RRAM 内导电丝的断裂; (c) 导电丝生成 (蓝)/断裂 (红) 过程中器件的 I - V 特性曲线

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. $\text{TiN}/\text{TaO}_x/\text{HfO}_x/\text{TiN}$ 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. HfO_2 阻挡层/石墨烯浮栅层/ Al_2O_3 隧穿层/ MoS_2 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 WSe_2 沟道/WCL 浮栅通道 (具有高线性长程增强特性) 和 MoS_2 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. $\text{PEO-LiClO}_4/\alpha\text{-MoO}_3$ 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 ($< 75 \text{ nS}$), 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 $\text{Pt}/\text{Al}_2\text{O}_3/\text{TiO}_{2-x}/\text{Ti}/\text{Pt}$ RRAM 的 20×20 1R 阵列. 基于两个这样的阵列实现的 $16 \times 10 \times 4$ 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_x 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 $\text{TiN}/\text{HfAl}_y\text{O}_x/\text{TaO}_x/\text{TiN}$ 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 HfO_2 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

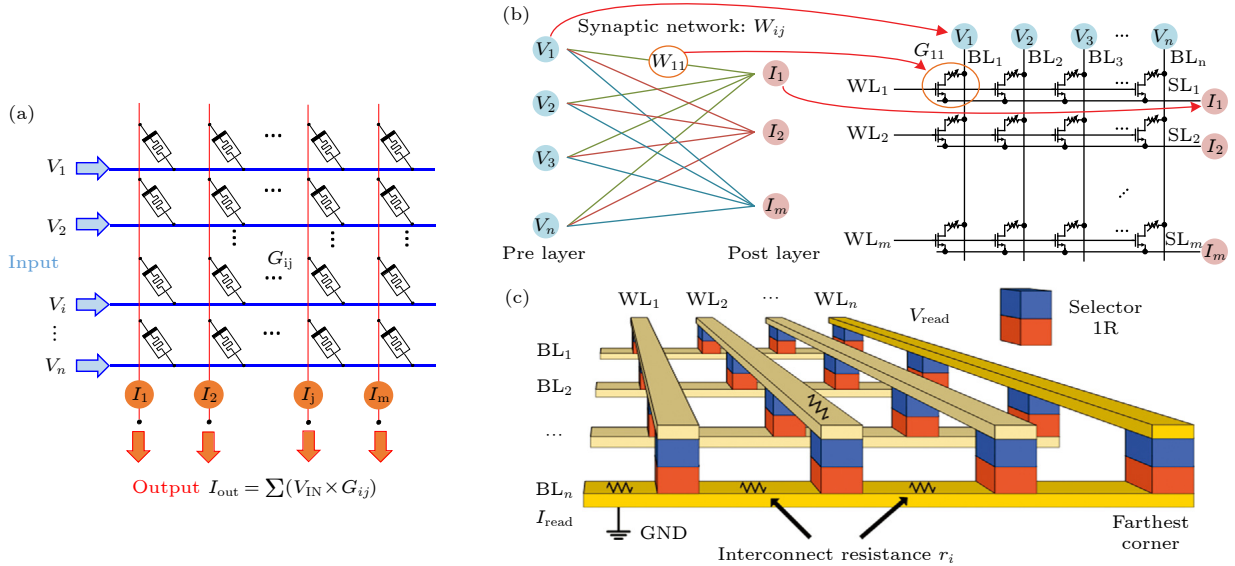


图2 应用于ANN的神经形态RRAM阵列 (a) 1R阵列的VMM运算示意图^[36]; (b) 1T1R阵列实现ANN的方式^[40]; (c) 1S1R阵列结构^[42]

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/TiO₂: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 和 τ 分别为强度和时间常数^[56]. 由于 ZrO_2 和 WS_2 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 $\text{Al}/\text{Ti}_3\text{C}_2\text{T}_x/\text{Pt}$ 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. $\text{Pt}/\text{KNbO}_3/\text{TiN}$ 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 $\text{Au}/\text{SiO}_x\text{N}_y/\text{Ag}/\text{Au}$ 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 Ca^{2+} 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/ WO_3 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/ WO_3 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

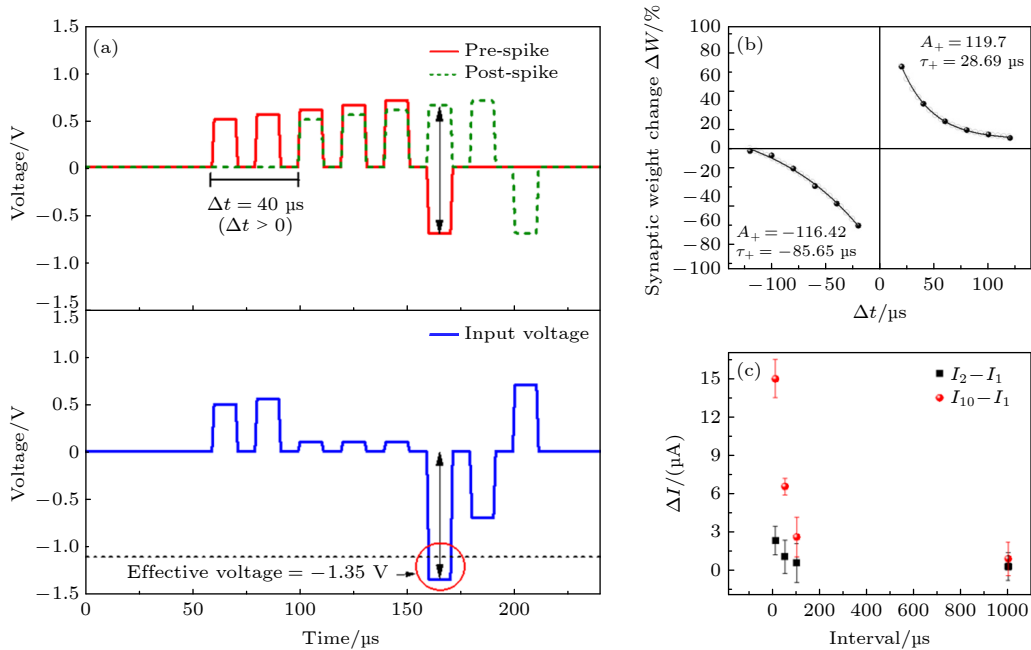


图 3 $\text{Pt}/\text{KNbO}_3/\text{TiN}$ 神经形态 RRAM^[58] (a) 具有 40 μs 时间差的突触前脉冲 (红) 和突触后脉冲 (绿), 以及对应的等效输入脉冲 (蓝); (b) 器件的 STDP 特性; (c) STP ($I_2 - I_1$) 和 PTP ($I_{10} - I_1$) 特性

Fig. 3. $\text{Pt}/\text{KNbO}_3/\text{TiN}$ neuromorphic RRAM^[58]: (a) Presynaptic pulse (red) and postsynaptic pulse (green) with 40 μs 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_xSiO_2 电解质层/ Nb_2O_5 沟道层结构的 RRAM 也具有 STDP 特性^[61]. 32×32 的 $\text{Li}_x\text{SiO}_2/\text{Nb}_2\text{O}_5$ RRAM 阵列与 CMOS LIF 神经元的系统可用于实现 25×8 的 SNN. RbAg_4I_5 电解质层/PEO 阻挡层/ P_3HT 沟道层结构的低功耗 RRAM 具有 STDP, STP 和 PPF 特性^[62]. 得益于 RbAg_4I_5 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]. $\text{Ag}/\text{SiO}_2/\text{Au}$ RRAM 的 TS 特性表现为外加电压大于 $V_{\text{th}2}$ 时, Ag 导电细丝生成, RRAM 从 HRS 切换到 LRS. 外部电压小于 $V_{\text{th}1}$ 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 $V_{\text{th}2}$ 之前, 输出电流可以忽略. 两端电压到达 $V_{\text{th}2}$ 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. $\text{Ag}/\text{SiO}_2/\text{Au}$ 器件电位下降到 $V_{\text{th}1}$ 时, 器件回到初始 HRS, 进入下一轮整合过程.

$\text{Pt}/\text{Ti}/\text{NbO}_x/\text{Pt}/\text{Ti}$ 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 $\text{Pt}/\text{Ta}/\text{Ta}_2\text{O}_5/\text{Pt}/\text{Ti}$ 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 $\text{Pt}/\text{Ag}/\text{TiN}/\text{HfAlO}_x/\text{Pt}$ 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 $\text{HfO}_2/\text{Al}_2\text{O}_3$ 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 (10^{12})、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 $\text{Pt}/\text{Ta}_2\text{O}_5/\text{TaO}_x/\text{Pt}$ 电容和 $\text{Ag}/\text{SiO}_x/\text{Ag}/\text{Ag}/\text{Pt}$ 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 $\text{Ag}/\text{V}_2\text{C}/\text{W}$ 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 $\text{Ag}/\text{SiO}_x/\text{TiN}$ 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 $\text{Si}/\text{SiO}_2/\text{Si}_3\text{N}_4/\text{SiO}_2/\text{Si}$ 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 Si_3N_4 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 Si_3N_4 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

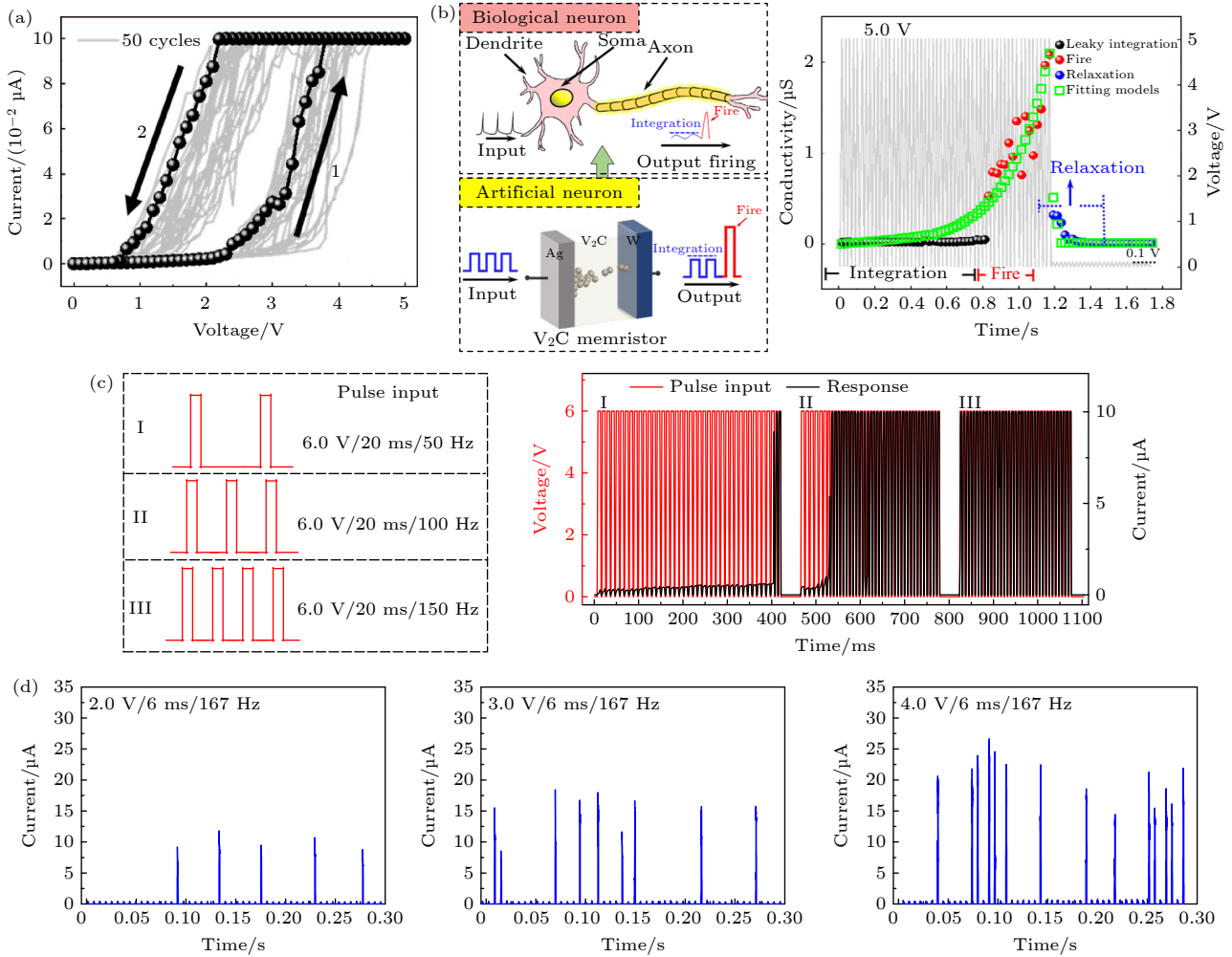


图4 两端 Ag/ V_2C /W 型 RRAM 器件^[68] (a) RRAM 的 TS 特性; (b) RRAM 作为人工 LIF 神经元 (左) 和神经元的 LIF 行为 (右); LIF 人工神经元输出脉冲频率受 (c) 输入脉冲频率和 (d) 输入脉冲幅值调控

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^[77]设计了 ITO/ZnO/Nb-SrTiO₃(NSTO) 结构的 ORRAM。该器件具有光可控的 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。

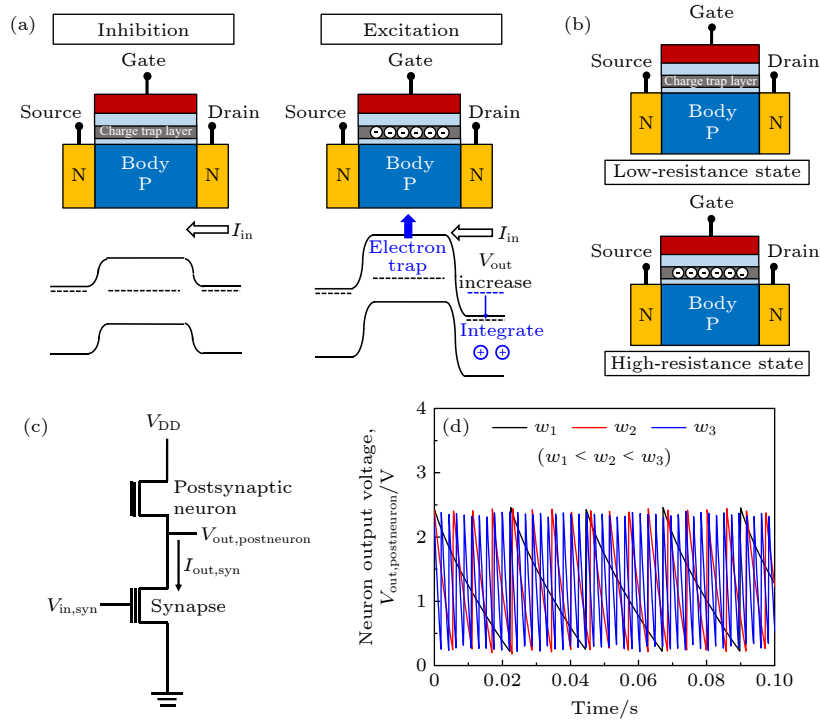


图 5 Si/SiO₂/Si₃N₄/SiO₂/Si 浮栅型神经形态 RRAM^[70] (a) 作为人工神经元; (b) 作为人工突触; (c) 人工突触和突触后神经元连接方式; (d) 人工神经元 LIF 行为产生的输出脉冲频率与所连接的人工突触权重大小的关系

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/WS₂ 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 WS₂ 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 (1.1×10^6)、阻态数 (> 128)、阻态保持时间 ($> 4.5 \times 10^4$)、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/WS₂/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 WS₂ 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/WS₂ 表面势垒, 离开 WS₂ 层, 器件电导率

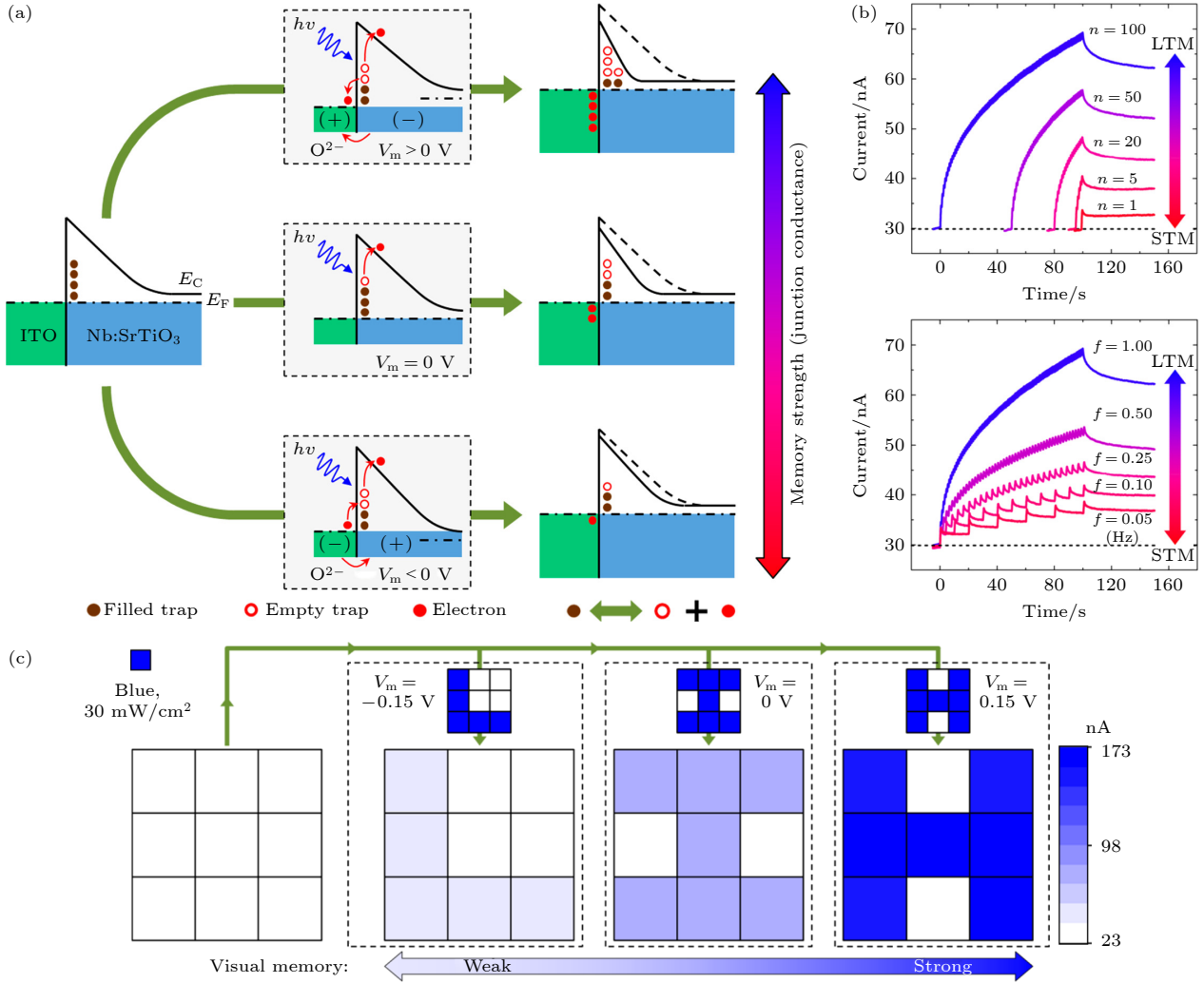


图 6 ITO/Nb-SrTiO₃/Ag 结构的神经形态 ORRAM^[76] (a) 光电调控的阻变机理; (b) 通过改变输入光脉冲频率或数量实现的 STP 和 LTP 特性之间的转换; (c) 器件阵列记忆强度随输入电压幅值增加而增强的特性

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 (5.1×10^7 A/W) 和高探测率 (2×10^6 Jones, 1 Jones = $1 \text{ cm} \cdot \text{Hz}^{1/2} \cdot \text{W}^{-1}$), 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 ($1 \mu\text{W}/\text{cm}^2$) 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%.

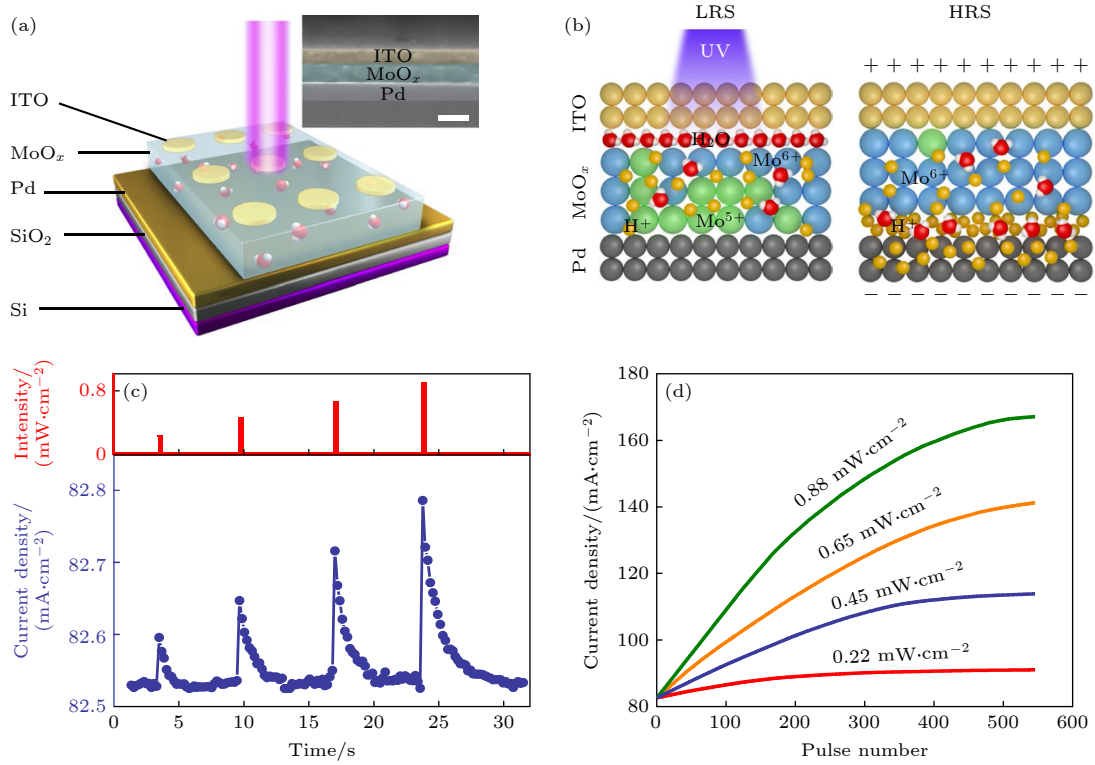


图 7 ITO/MoO_x/Pd 神经形态 ORRAM^[78] (a) 器件结构; (b) 基于 Mo 离子价态转变的电阻调控机理; (c) ORRAM 的 STP 特性; (d) ORRAM 的 LTP 特性

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.

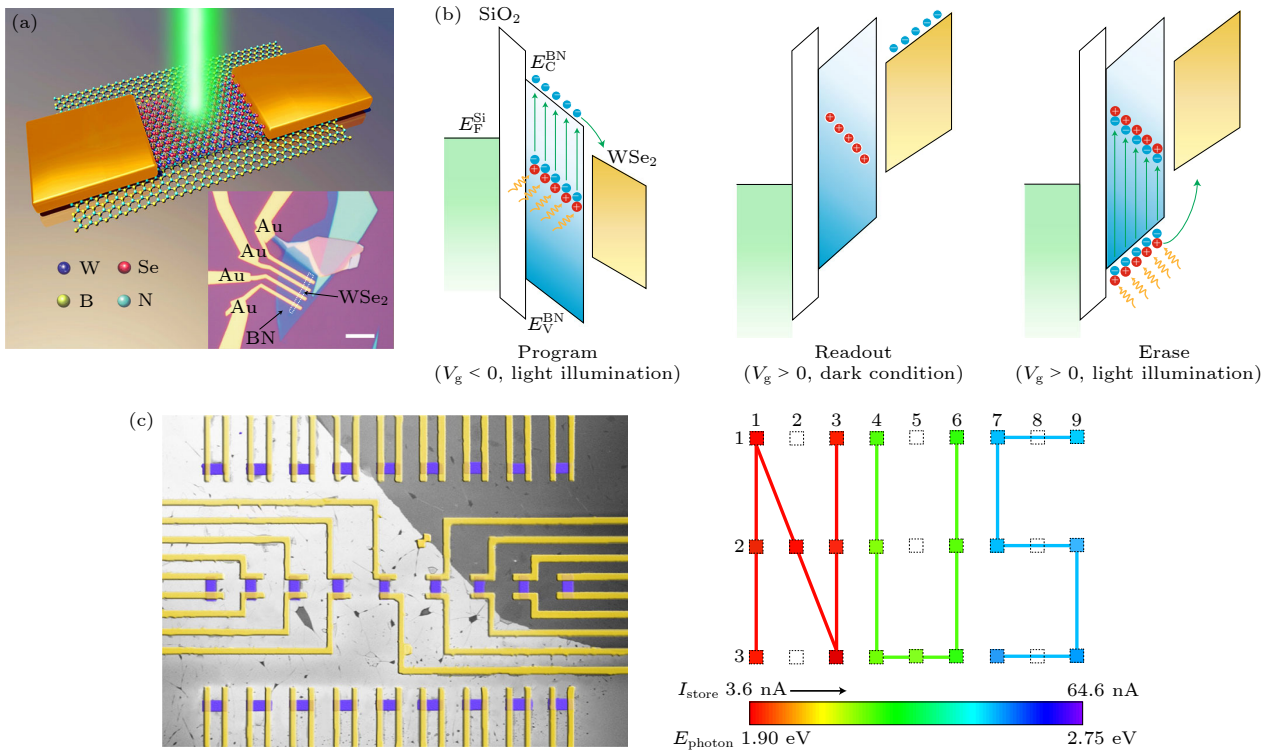


图 8 基于 BN/WS₂ 异质结结构的三端 ORRAM^[81] (a) 器件结构; (b) 光电调控的阻变原理; (c) ORRAM 组成的阵列对不同波长光输入的不同存储效应

Fig. 8. Three-terminal ORRAM device based on BN/WS₂ heterostructure^[81]: (a) Device structure; (b) switching mechanisms; (c) different storage levels resulted from different light wavelengths in ORRAM array.

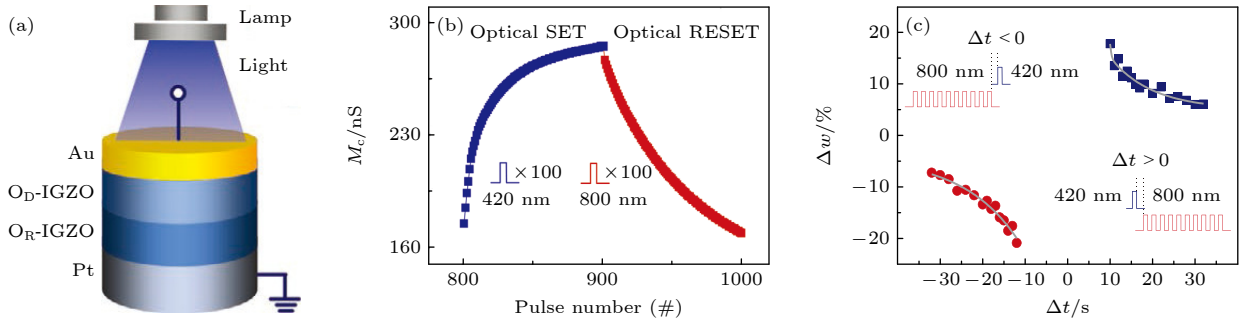


图9 Au/富氧IGZO/缺氧IGZO/Pt结构的ORRAM^[85] (a) 器件结构; (b) 可见光脉冲(420 nm)使器件电导率上升和近红外光脉冲(800 nm)使器件电导率降低的过程; (c) 光调控的STDP特性

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

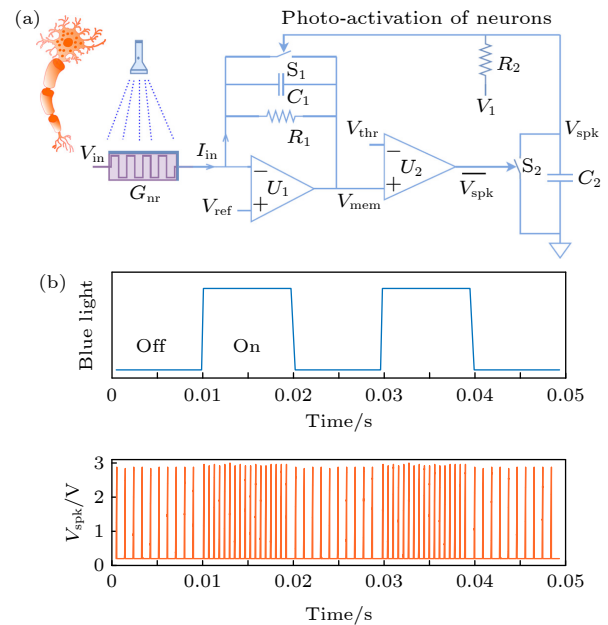


图10 基于 ReS₂ ORRAM 与 CMOS LIF 神经元构建的光可控神经元^[89] (a) 光可控神经元结构; (b) 光可控神经元输出脉冲频率在光照下增加的行为

Fig. 10. Light tunable artificial neuron based on ReS₂ ORRAM and CMOS LIF neuron^[89]: (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, M_L 和 M_U 控制当前单元与左/右单元连接模拟蚂蚁行进方向, M_{DD} 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.

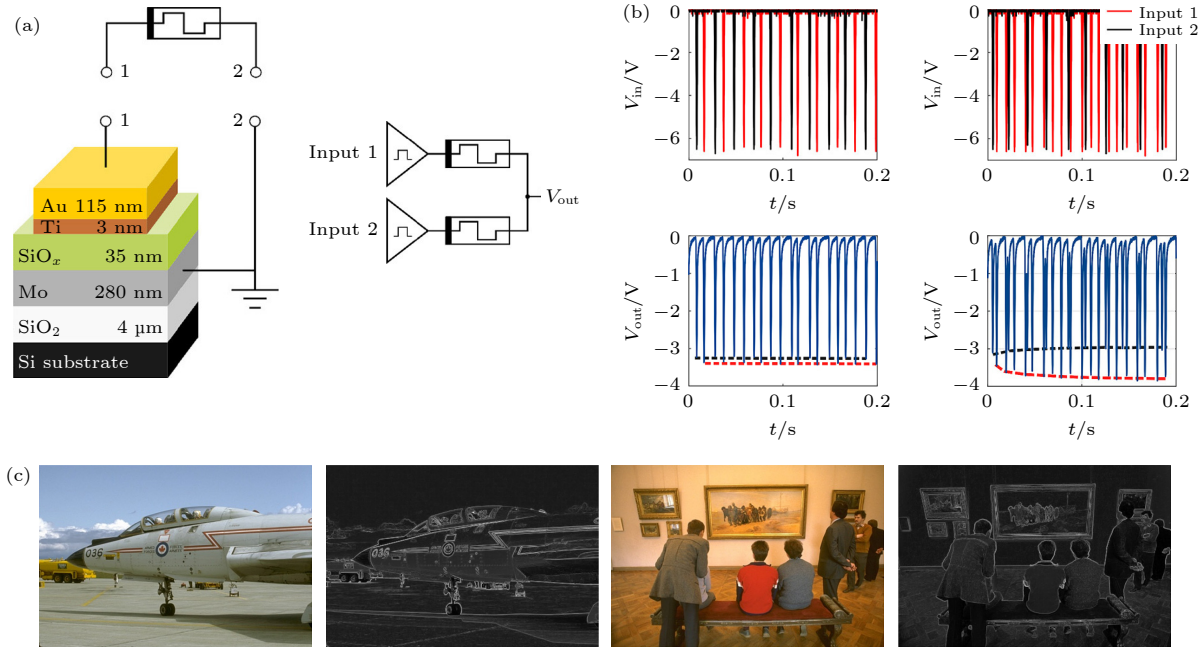


图 11 基于神经形态阻变器件频率差检测电路实现的图像边缘提取^[94] (a) 基于 RRAM 分压器的频率差检测电路 (右) 和所使用的器件结构 (左); (b) 两组输入脉冲频率相同 (左) 和不同 (右) 时频率差检测电路的输出; (c) 原图和频率差检测电路提取的图片边缘

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, t_1 时刻的 4 个图像像素 (x_{1-4}) 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 $t_1 - \Delta t$ 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 t_1 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 $\text{h-BN}/\text{WSe}_2/\text{Al}_2\text{O}_3/\text{BP}$ 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

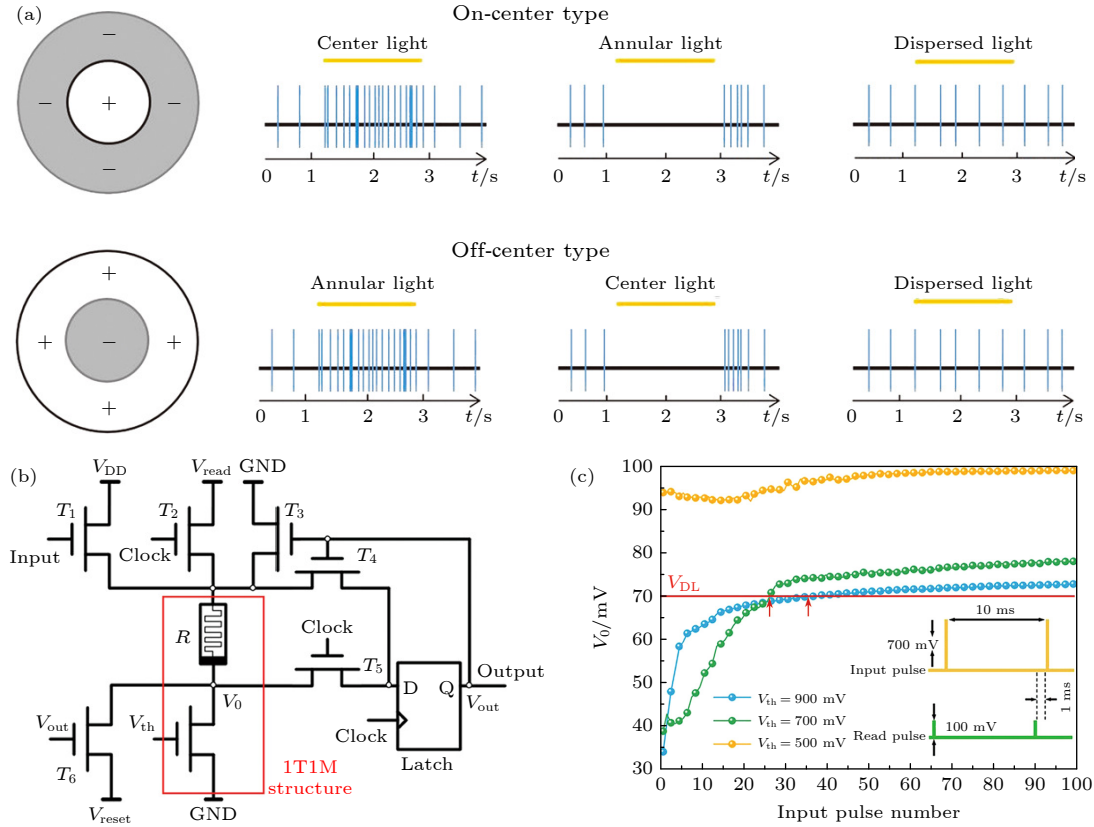


图 12 基于 RRAM 和 CMOS 晶体管人工视网膜单元实现的边缘提取^[99] (a) 生物视网膜系统 (光感受-双极-神经节细胞) 对不同输入光照的不同输出脉冲频率; (b) 人工视网膜单元结构; (c) 人工视网膜单元输出信号 V_0 随 V_{th} 端口输入信号和 input 端口输入信号的变化

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_{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, 随背栅电压降低, 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)).

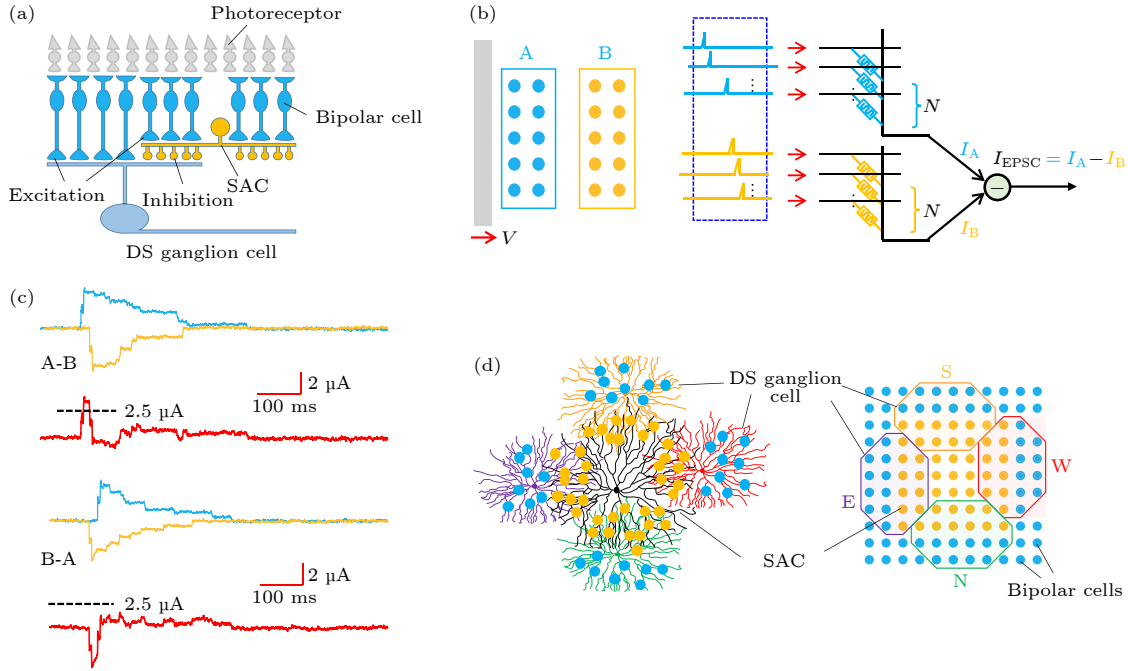


图 13 基于 Ag/HfO₂/C RRAM 人工神经节细胞实现的运动检测^[11] (a) 具有给光/撤光反应机制的生物视网膜系统结构; (b) 人工神经节细胞结构; (c) 人工神经节细胞工作原理; (d) 包含 4 个人工神经节细胞的 RRAM 阵列

Fig. 13. Artificial ganglion cell based on Ag/HfO₂/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.

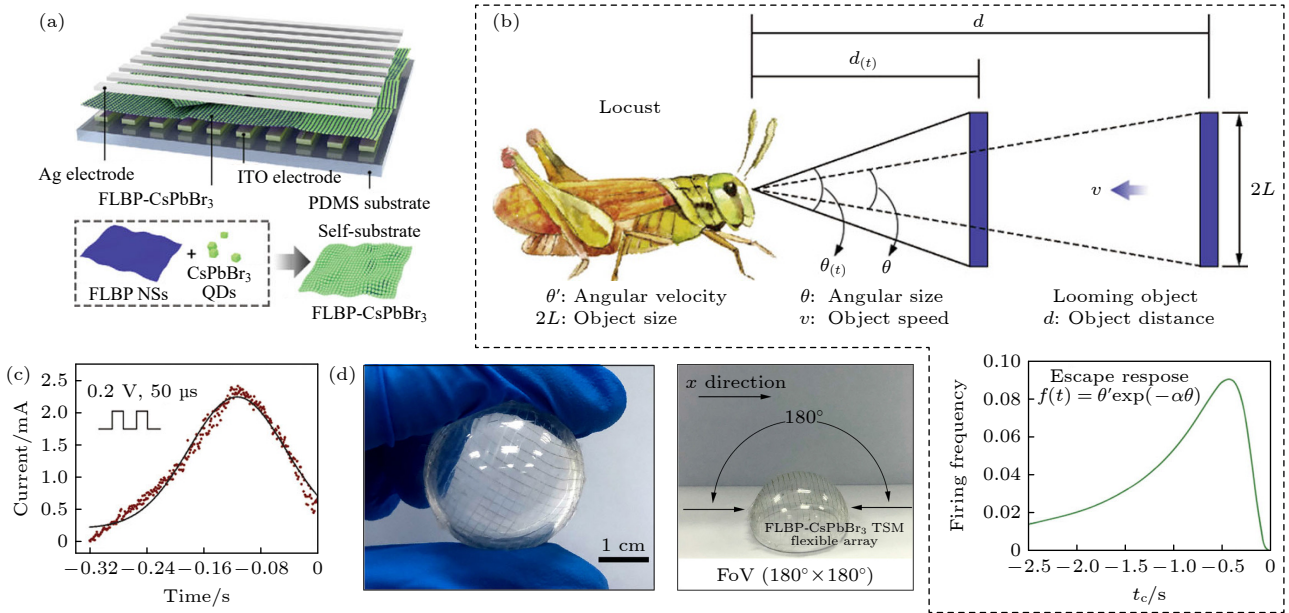


图 14 基于 Ag/FLBP-CsPbBr₃/ITO ORRAM 类眼球形阵列实现的运动检测^[105] (a) 单个器件的结构; (b) 生物 LGMD 细胞输出脉冲频率对接近物体的非线性反应; (c) 基于 Ag/FLBP-CsPbBr₃/ITO ORRAM 实现的人工 LGMD 神经元对生物 LGMD 神经元非线性响应特性的模仿; (d) 柔性 Ag/FLBP-CsPbBr₃/ITO ORRAM 构建的类眼球形阵列

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].

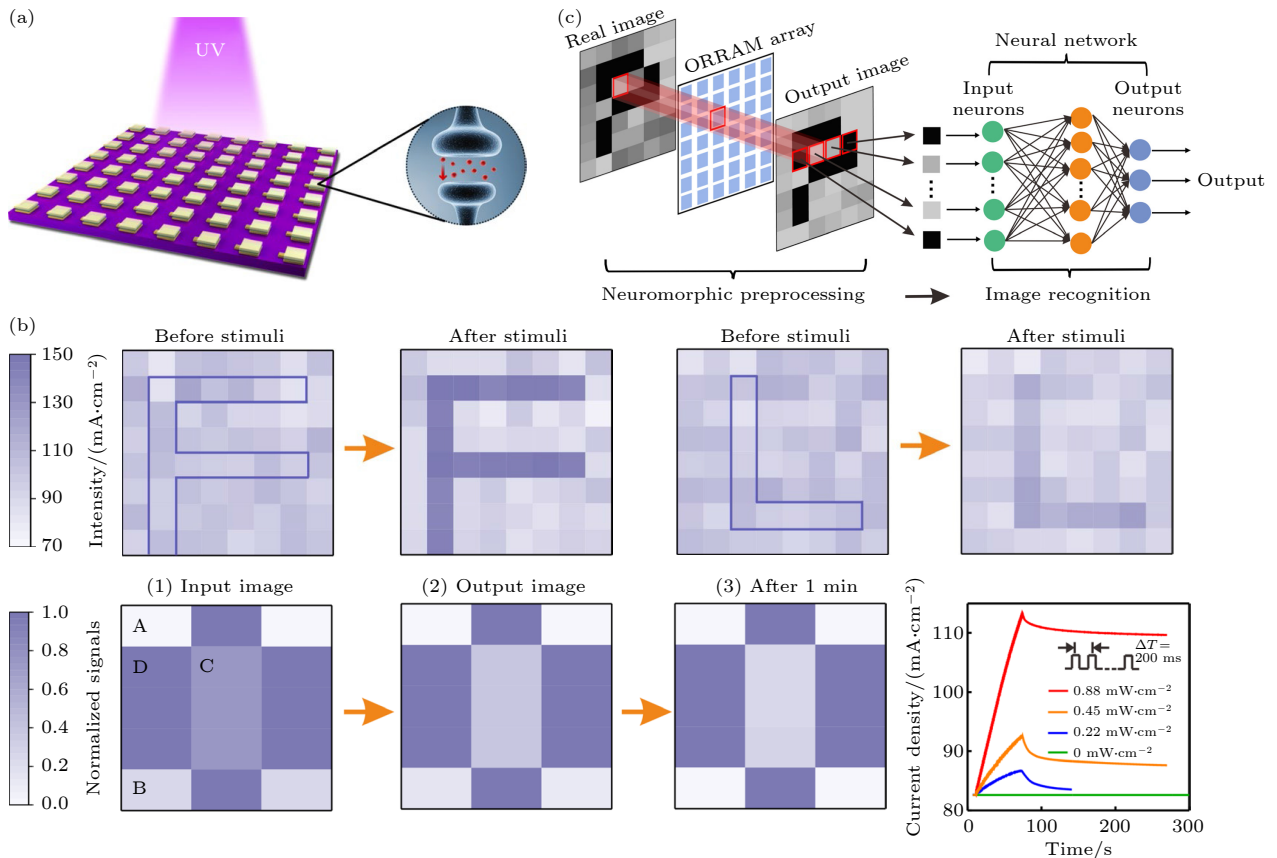


图 15 基于 ITO/MoO_x/Pd ORRAM 阵列实现的图像锐化^[78] (a) 8×8 ITO/MoO_x/Pd ORRAM 阵列; (b) ORRAM 阵列的非线性阻变特性; (c) 基于 ORRAM 图像锐化阵列和图像识别神经网络的人工视觉系统

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 $\text{SiN}_x/\text{n-Si/p-Si/Al}$ 太阳能电池光探测器和 $\text{ITO/CsPbBr}_2\text{I/P}_3\text{HT/Ag}$ RRAM 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 $\text{TiN/TaO}_x/\text{HfO}_x/$

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 HfO_2 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法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 $\text{WSe}_2/\text{BN}/\text{Al}_2\text{O}_3$ 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 $\text{Pt/SiO}_x\text{N}_y\text{:Ag/Pt}$ RRAM 的 RC LIF 人工神经元和基于 $\text{Pd/HfO}_2/\text{Ta}$ 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/AlO₃/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法运算. 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

- [1] 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, February 26–27, 2014 p1
- [4] Chaki J, Dey N 2018 *A Beginner's Guide to Image Preprocessing Techniques (Vol. 1)* (Boca Raton: CRC Press)
- [5] 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
- [7] 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* **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, Prabakaran 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, Kvatinisky 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, Protachevich 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 Vllora 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

- [74] 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, Mehonc 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, Papanikolaopoulos 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 preprocessing and higher-level image processing algorithms, but also to implement highly efficient biomimetic 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