

## 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>

$\text{NbO}_x$  忆阻神经元的设计及其在尖峰神经网络中的应用

Design of  $\text{NbO}_x$  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  $\text{MoO}_3$  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<sup>1)2)</sup> 王亚赛<sup>1)2)</sup> 何毓辉<sup>1)2)†</sup> 缪向水<sup>1)2)‡</sup>

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<sup>[1]</sup>. 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)<sup>[2]</sup> 和 Dennard scaling<sup>[3]</sup> 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”<sup>[4]</sup>. 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<sup>[5]</sup>. 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  $10^{15}$  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<sup>[6]</sup>. 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<sup>[7]</sup>. 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

\* 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](mailto:heyuhui@hust.edu.cn)‡ 通信作者. E-mail: [miaoxs@hust.edu.cn](mailto: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<sup>[8]</sup>.

忆阻器<sup>[9,10]</sup>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<sup>[10,11]</sup>. 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<sup>[12]</sup>. 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<sup>[13]</sup>. 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<sup>[14]</sup>. 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

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 device-to-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<sup>[15]</sup>.

### 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<sup>[13]</sup>. 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<sup>[16,17]</sup>.

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.



图 1 综述各节内容总括

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

## 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/HfO<sub>2</sub>/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

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/SiO<sub>2</sub>/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 ( $\approx 1 \mu\text{A}$ ) 的存储器和高驱动电流 ( $\approx 1 \text{ mA}$ ) 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  $10^9$  的高开关比、高电流驱动 ( $600 \mu\text{A}$ )、超低电流 ( $100 \text{ fA}$ )、极陡的开关斜率 (低于  $1.4 \text{ mV/dec}$ ). 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  $10^5$  的开关比 [29,30]. CBRAM 的开关比一般能够达到  $10^3$ — $10^6$ , 但可擦写次数往往较低, 仅在  $10^4$  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, 导电丝的热溶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/TiO<sub>2</sub>/Pt 器件中, 他们利用透射电子显微镜 (TEM) 观察到了成分为 Ti<sub>4</sub>O<sub>7</sub> 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

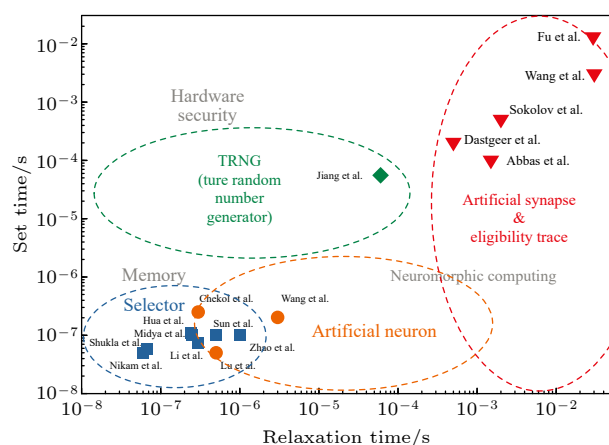


图 2 根据  $t_{\text{set}}$  和  $t_r$  顺序排列的 v-ECM 型扩散忆阻器的不同应用领域 [25]

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



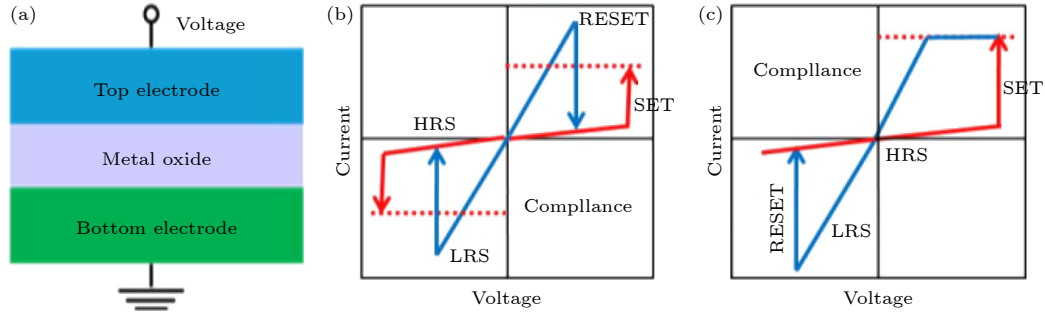

 图 3 (a) Oxide RRAM 的 MIM 结构原理图; (b) 非极性和 (c) 双极性不同原理的阻变  $I-V$  特性曲线图<sup>[34]</sup>

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<sup>[34]</sup>.

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<sup>[39,40]</sup>. 基于这些建模, Kim 等<sup>[41]</sup>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  $10-10^3$  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<sup>[42]</sup>, 85 ps 的开关速度<sup>[43]</sup>, 10 年以上的保持特性<sup>[44]</sup>及  $10^{12}$  的可擦写次数<sup>[45]</sup>, 是一种非常具有潜力的存储器件<sup>[8]</sup>. 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<sup>[13]</sup>, 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-Sb<sub>2</sub>Te<sub>3</sub> 线的合金, 它们重结晶的效果好; 第二个是在 Sb<sub>2</sub>Te 上进行掺杂, 通常是 Ag<sub>5</sub>In<sub>5</sub>Sb<sub>60</sub>Te<sub>30</sub> (AIST), 常用于光存储中<sup>[46]</sup>. 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, Salanga 等<sup>[47]</sup>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

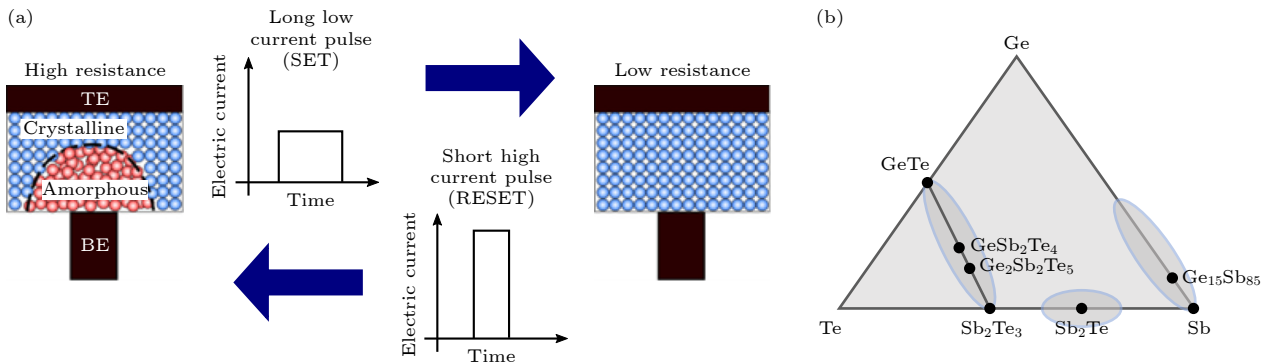

 图 4 (a) PCRAM 相变的脉冲操作示意图<sup>[46]</sup>; (b) 目前主流合金材料的三元 Ge:Sb:Te 相图<sup>[48]</sup>

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

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  $10^6$  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 ( $\text{Ge}_x\text{Sb}_y\text{Te}_{1+x+y}$ ) 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.他们设计的碲化铋( $\text{Sc}_{0.2}\text{Sb}_2\text{Te}_3$ )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铋(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  $\text{VO}_2$ [58–60],  $\text{V}_2\text{O}_3$ [61],  $\text{NbO}_x$ [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  $\text{VO}_2$  的研究较多. Zhou 等[58]提出了一种 Pt/( $\bar{2}11$  向)  $\text{VO}_2$ /ITO 的结构,可以达到 23 ns 的开关速度,  $10^9$  个循环的擦写次数, 143 的开关比,且器件间差异小.可以看出,  $\text{VO}_2$  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(spin-transfer 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.

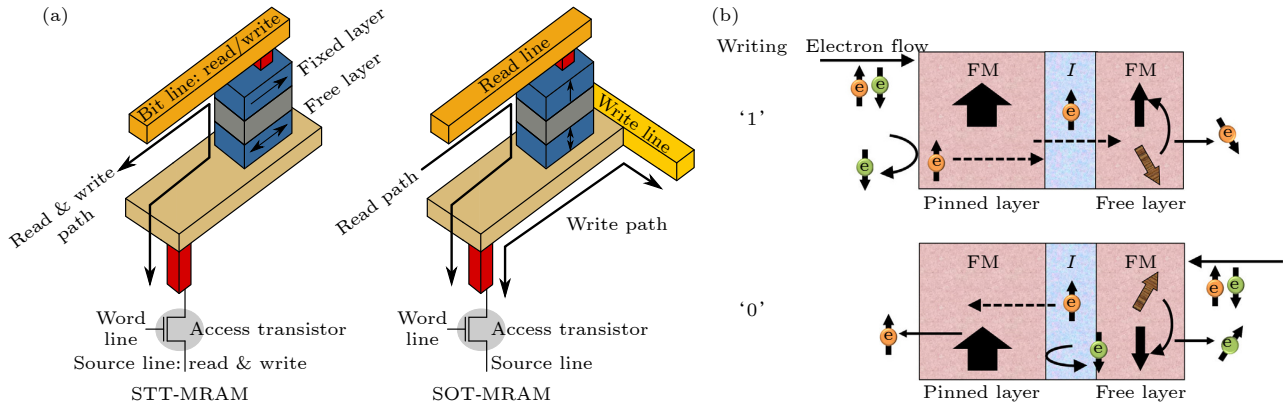


图 5 (a) STT-MRAM 和 SOT-MRAM 的单元结构图<sup>[63]</sup>; (b) STT-MRAM 的置态和重置操作原理图<sup>[64]</sup>

Fig. 5. (a) Bit-cell for STT-MRAM and SOT-MRAM<sup>[63]</sup>; (b) schematic illustration of the STT switching mechanism to achieve 1 and 0 states<sup>[64]</sup>.

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<sup>[63]</sup>. 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<sup>[64]</sup>.

STT-MRAM 的优势在于低于 10 ns 的高切换速度, 并且具有  $5 \times 10^{14}$  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%<sup>[65–67]</sup>. 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%<sup>[68,69]</sup>. Jain 等<sup>[70]</sup> 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<sup>[71]</sup> 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  $64 \times 64$  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 5(a) 右所示. 通过这种方式, 编程过程不需

经过 MgO 屏障, 同时解决了老化和能耗问题<sup>[72]</sup>. 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<sup>[73]</sup>.

## 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  $\text{BaTiO}_3$  (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)

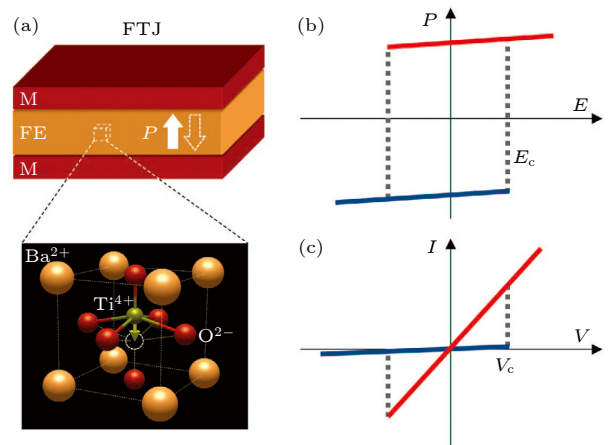


图 6 (a) FTJ 结构及铁电层  $\text{BaTiO}_3$  晶体结构示意图; TER 效应: (b)  $P$ - $E$  特性曲线; (c)  $I$ - $V$  特性曲线<sup>[77]</sup>

Fig. 6. (a) FTJ structure and schematic diagram of the ferroelectric layer  $\text{BaTiO}_3$  crystal structure; TER effect: (b)  $P$ - $E$  characteristic curve; (c)  $I$ - $V$  characteristic curve<sup>[77]</sup>.



效应, 即在外电压使得铁电层极化反转的同时, 隧道阻态发生变化, 达到  $10^4$  的开关比<sup>[74–76]</sup>. 经总结, 铁电材料极化翻转时铁电势垒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<sup>[77]</sup>.

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<sup>[78]</sup>. 使用  $\text{Hf}_{0.5}\text{Zr}_{0.5}\text{O}_2$  (HZO) 作为铁电材料层, FeRAM 能拥有比较好的 CMOS 兼容性<sup>[79,80]</sup>. 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  $10^{11}$  次, 但开关比只有 10 倍左右<sup>[81–83]</sup>. 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<sup>[84]</sup> 使用层状铜铟硫代磷酸 ( $\text{CuInP}_2\text{S}_6$ ) 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  $10^7$  的 TER 变化. 通过构建异质结构 Pt/BaTiO<sub>3</sub>/Nb:SrTiO<sub>3</sub>, Wen 等<sup>[76]</sup> 得到了  $10^4$  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.

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

## 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<sup>[96]</sup>. 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<sup>[97]</sup>. 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<sup>[98]</sup>. 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<sup>[99]</sup>. 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, 即突触的功效率和可塑性. 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



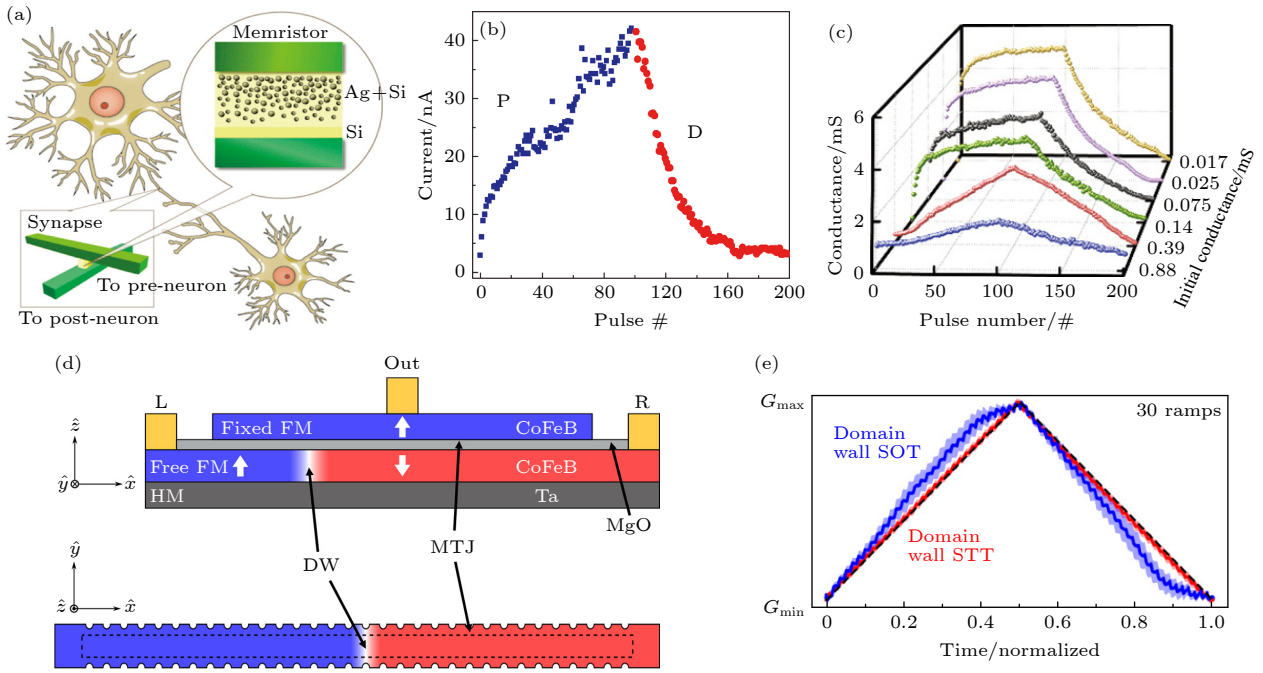


图7 LTP的忆阻器实现 (a) 使用忆阻器作为神经元之间的突触的概念示意图; (b) 忆阻器对编程脉冲的响应<sup>[101]</sup>; (c) Pt/LiAlO<sub>x</sub>/TiN忆阻器在不同初始电导状态下的电导调制性能<sup>[102]</sup>; (d) DW-MTJ人工突触的侧视和俯视图; (e) DW-SOT和DW-STT器件的测试更新线性度和对称性<sup>[103]</sup>

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<sup>[101]</sup>; (c) conductance modulation performance at different initial conductance states of Pt/LiAlO<sub>x</sub>/TiN memristor<sup>[102]</sup>; (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<sup>[103]</sup>.

长期抑制 (long-term depression, LTD)<sup>[100]</sup>. 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<sup>[96]</sup>. 表现在忆阻器件的实现上, 就是在置态和重置过程中, 电导调制应该非易失性和渐进式的, 并且需要有足够多的态. Jo等<sup>[101]</sup>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<sup>+</sup>稳定存在, 通过施加置态和重置电压使Ag<sup>+</sup>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<sup>+</sup>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) 和电导变化的非线性度 (nonlinearity, NL), 限制了忆阻网络水平上记忆突触的进一步发展<sup>[98]</sup>. Fu等<sup>[102]</sup>构建了Pt/LiAlO<sub>x</sub>/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<sup>[103]</sup>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<sup>[104]</sup>通过插入一个2 nm的BiFeO<sub>3</sub>层, 构造了一个Pt/BiFeO<sub>3</sub>/HfO<sub>2</sub>/TiN型VCM忆阻器, 实现了开关比达10<sup>4</sup>的大存储窗口和10<sup>8</sup>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<sup>[105]</sup>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<sup>[106]</sup>. 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<sup>[106]</sup>. 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<sup>[107]</sup>.

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<sup>[108]</sup> 首先制备了一种 Ag/Ag<sub>2</sub>S/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<sup>[109]</sup> 提出了一种 Pt/WO<sub>x</sub>/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<sup>[110]</sup> 也

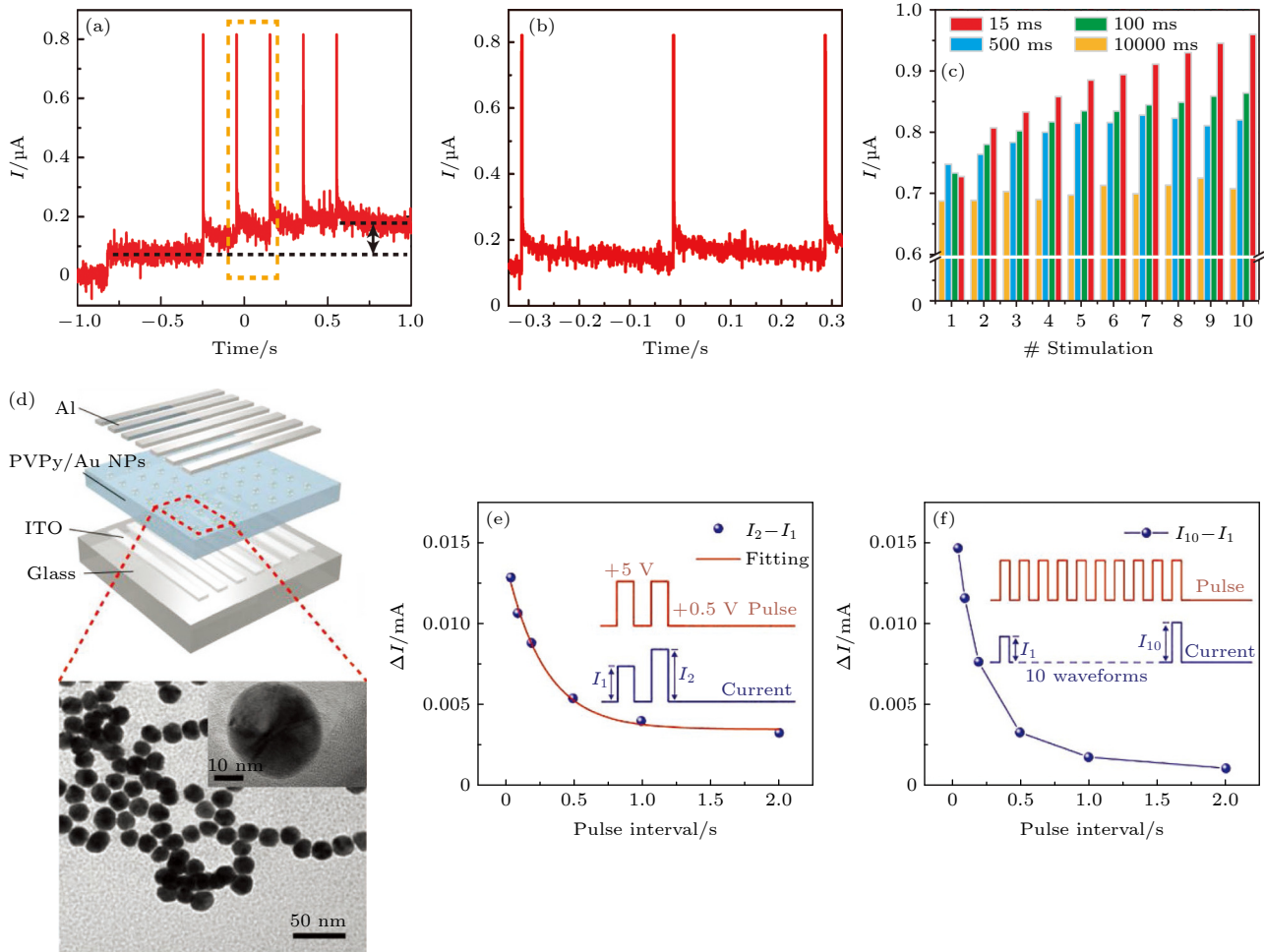


图 8 STP 的忆阻器实现 (a) 测试过程中通过忆阻器的连续电流变化; (b) 图 (a) 中矩形区域的特写图; (c) 电导转换速率与刺激速率的关系图, 在不同脉冲间隔条件下, 每次刺激脉冲后通过忆阻器的电流<sup>[109]</sup>; (d) ITO/PVPy-Au NPs/Al RRAM 器件的结构和 Au NPs 的 HRTEM 图像; 在 (e) 2 次和 (f) 10 次不同脉冲间隔的脉冲之间的器件电流变化图<sup>[111]</sup>

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<sup>[109]</sup>. (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<sup>[111]</sup>.

利用  $\text{WO}_{3-x}$  功能层得到类似的结论. Zhang 等<sup>[111]</sup> 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<sup>[112]</sup> 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  $\theta_m$  的突触前脉冲会导致 LTP, 而低于  $\theta_m$  的突触前脉冲会导致 LTD, 如图 9(a) 所示<sup>[113-115]</sup>. 其中  $\theta_m$  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<sup>[116,117]</sup>.

正如前面所述, 基于  $\text{WO}_x$  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<sup>[109,110]</sup>. 根据该特性, Du 等<sup>[118]</sup> 制作了基于  $\text{WO}_x$  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  $\theta_m$  的活动依赖性.

### 3.1.4 脉冲时间依赖可塑性 (STDP)

1949 年, Hebb<sup>[119]</sup> 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<sup>[120]</sup>. 在一些皮质突触上, LTD 的时间窗 (灰色虚线) 被

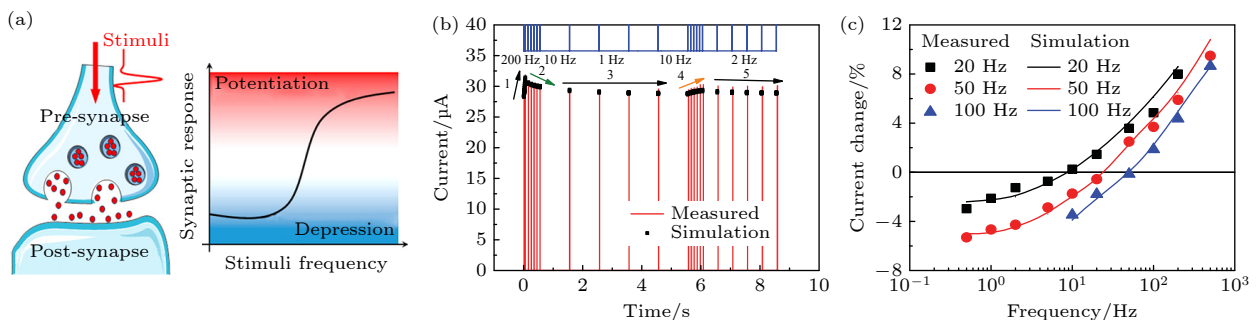


图 9 SRDP 的忆阻器实现 (a) 具有 SRDP 特性的生物突触示意图<sup>[116]</sup>; (b)  $\text{WO}_x$  基忆阻器对不同频率下连续编程脉冲序列 (1 V, 1 ms, 蓝线) 的响应; (c) 在经历了不同程度的激活后, 忆阻器电流随刺激频率的变化. 由 5 个不同频率的脉冲 (1.2 V, 1 ms) 组成的不同频率的脉冲序列对忆阻器进行编程<sup>[118]</sup>

Fig. 9. Memristor implementation of SRDP: (a) Schematic diagram of a biological synapse with SRDP activities<sup>[116]</sup>; (b)  $\text{WO}_x$  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<sup>[118]</sup>.



扩展<sup>[121,122]</sup>. 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)<sup>[122,123]</sup>.

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<sup>[125]</sup>. 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<sup>[124]</sup> 首先构建了一个 TiN/HfO<sub>x</sub>/AlO<sub>x</sub>/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<sup>[126]</sup>, MRAM<sup>[44]</sup> 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 Ca<sup>2+</sup>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<sup>[127]</sup>. 很多研究通过构建二阶漂移 RRAM 忆阻器模型, 利用热耗散<sup>[41]</sup> 或者氧空位迁移率的改变<sup>[118]</sup> 来近似化学突触中 Ca<sup>2+</sup>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

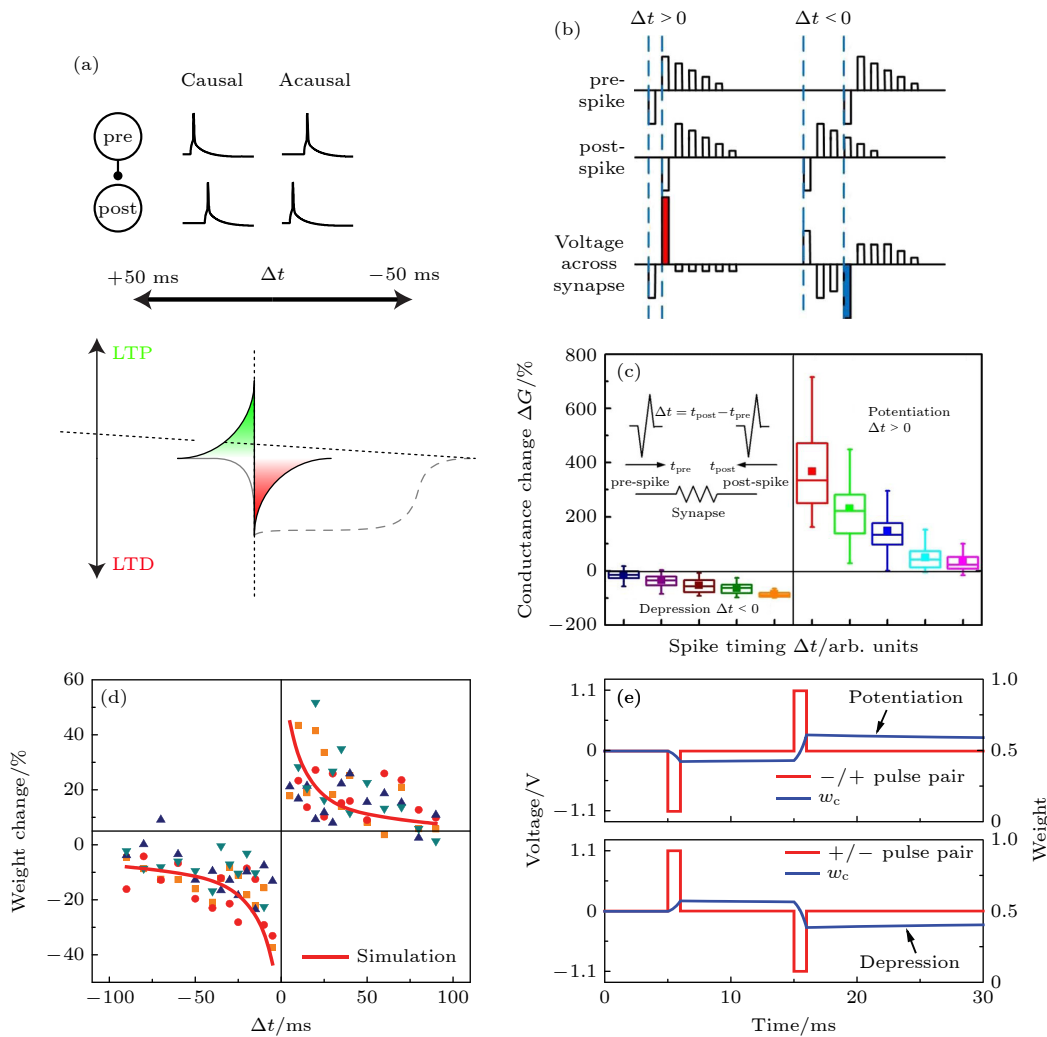


图 10 STDP 的忆阻器实现 (a) STDP 特性展示图<sup>[120]</sup>; (b) 利用 TDM 和脉冲幅度调制的 STDP 实现方案, 突触前脉冲振幅分别为 -1.4, 1, 0.9, 0.8, 0.7 和 0.6 V, 突触后脉冲振幅分别为 -1, 1.4, 1.3, 1.2, 1.1 和 1 V; (c) 利用图 (b) 中的方法实测的器件 STDP 曲线<sup>[124]</sup>; (d) 忆阻权重的变化与突触前后脉冲相对时间的关系,  $\Delta t = t_{\text{post}} - t_{\text{pre}}$ ; (e) 脉冲相对时间影响忆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

Fig. 10. Memristor implementation of STDP: (a) Defining spike-timing-dependent plasticity<sup>[120]</sup>; (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)<sup>[124]</sup>. (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<sup>[118]</sup>.

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  $w_m$  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/Ga<sub>2</sub>O<sub>3</sub>/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 SiO<sub>x</sub>N<sub>y</sub>:Ag 的忆阻器, 并构建了 Ag<sup>+</sup>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<sup>+</sup>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). 膜电位低于阈值电位, 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]:

$$\tau_m \frac{dV}{dt} = V_{\text{rest}} - mV + R_m I,$$

其中,  $\tau_m$  表示膜时间常数,  $V_{\text{rest}}$  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 SiO<sub>2</sub> 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,

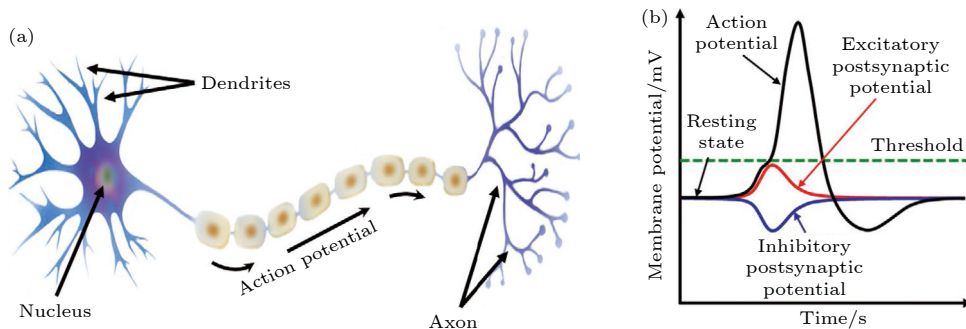


图 11 (a) 生物神经元的结构; (b) 基于兴奋性和抑制性电位的神元膜电位变化 [130]

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  $V_{\text{hold}}$  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  $\text{NbO}_2$  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  $\text{Ag}/\text{SiO}_2/\text{Au}$  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  $R_s$  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

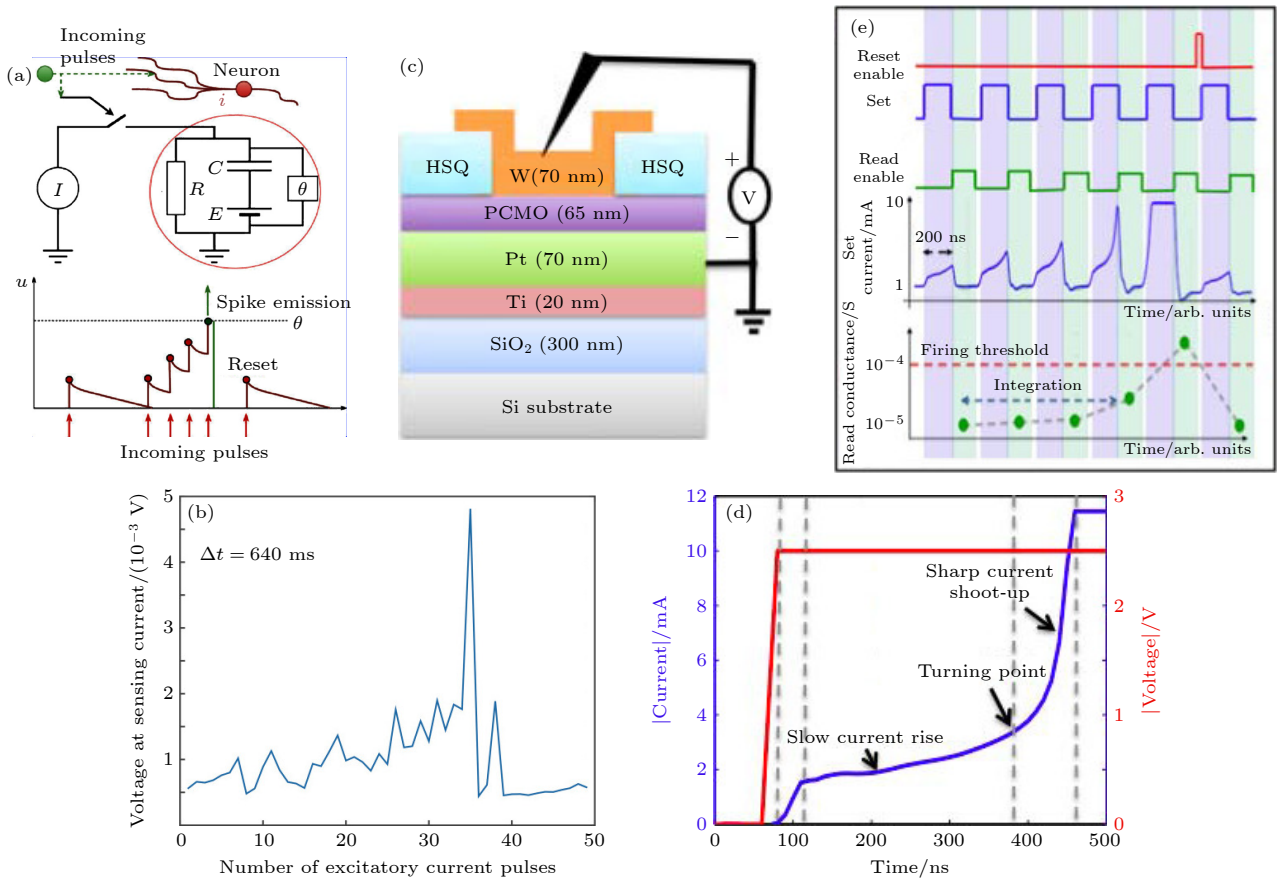


图 12 非易失器件实现神经元 (a) 漏电积分点火神经元的模型展示图; (b) 输入间隔 640 ms 的兴奋性脉冲序列时得到的输出电流图 [135]; (c) PCMO RRAM 的器件结构图; (d) 施加 -2.3 V 置态电压时显示出的 3 个不同阶段的瞬态电流值; (e) 应用预设脉冲序列的瞬态实验电流值 [136]

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  $\text{FeO}_x$ <sup>[140]</sup> 和  $\text{HfAlO}_x$ <sup>[30]</sup> 的易失性忆阻器实现了更低功耗的 LIF 神经元.

Huang 等<sup>[141]</sup> 利用易失性  $\text{W}/\text{WO}_3/\text{poly}$  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<sup>[143]</sup> 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

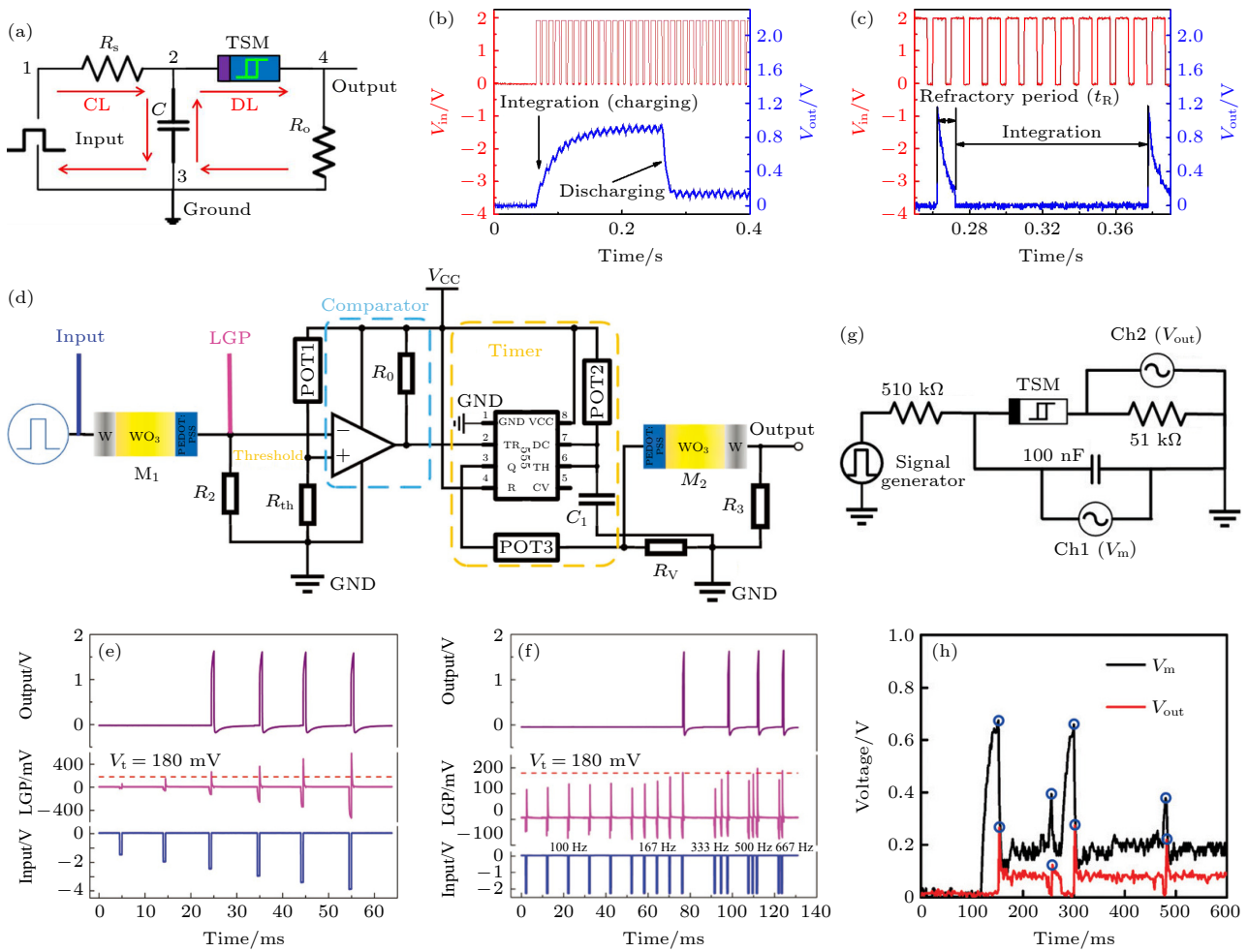


图 13 易失型器件实现神经元 (a) 所提出的神经元电路的示意图; (b) 电容上的电压变化图; (c) 输出神经元的发放脉冲具有相应的不应期和积分时间<sup>[139]</sup>; (d) 带有两个  $\text{W}/\text{WO}_3/\text{PEDOT:PSS}/\text{Pt}$  忆阻器件的具体神经元电路; (e) 利用电路得到的单脉冲生物积分点火; (f) 利用电路得到的连续脉冲生物积分点火<sup>[141]</sup>; (g) 神经元电路原理图, 输入电压脉冲来自信号发生器; (h) 基于  $\text{CuS}/\text{GeSe}$  的神经元电路在脉冲幅度为 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<sup>[142]</sup>

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<sup>[139]</sup>; (d) the electrical circuit with two  $\text{W}/\text{WO}_3/\text{PEDOT:PSS}/\text{Pt}$  memristive devices; (e) spatial integration and bioinspired fire realized with the circuit; (f) temporal integration and bioinspired fire realized with the circuit<sup>[141]</sup>; (g) schematic of neuronal circuit where the input voltage pulses originate from the signal generator; (h) the experimentally measured stochastic spike events of the  $\text{CuS}/\text{GeSe}$  based neuronal circuit under an input voltage pulse train with pulse height 2 V and duration 7.5 ms<sup>[142]</sup>.

神经元以用硬件实现神经网络训练时的 dropout. Wang 等<sup>[142]</sup>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/Ta<sub>2</sub>O<sub>5</sub>/Pt 型 CBTS 选通管与运算放大器, Huang 等<sup>[144]</sup>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<sup>[13]</sup>. 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<sup>[13]</sup>. 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<sup>[14]</sup>. 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<sup>[145]</sup>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<sub>2-x</sub>/Al<sub>2</sub>O<sub>3</sub>/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<sup>[146]</sup>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<sup>[147,148]</sup>也进行了相关的研究. 基于 Pt/Au/WO<sub>x</sub>/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<sup>[147]</sup>.

IBM 团队<sup>[149]</sup>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<sup>[150]</sup>在之前工作<sup>[145,146]</sup>的基础上提出基于 Pt/Al<sub>2</sub>O<sub>3</sub>/TiO<sub>2-x</sub>/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<sup>[151–153]</sup>.

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<sup>[154]</sup> 基于  $\text{HfO}_2/\text{TaO}_x$  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  $16 \times 16 \times 16$  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<sup>[155]</sup> 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.

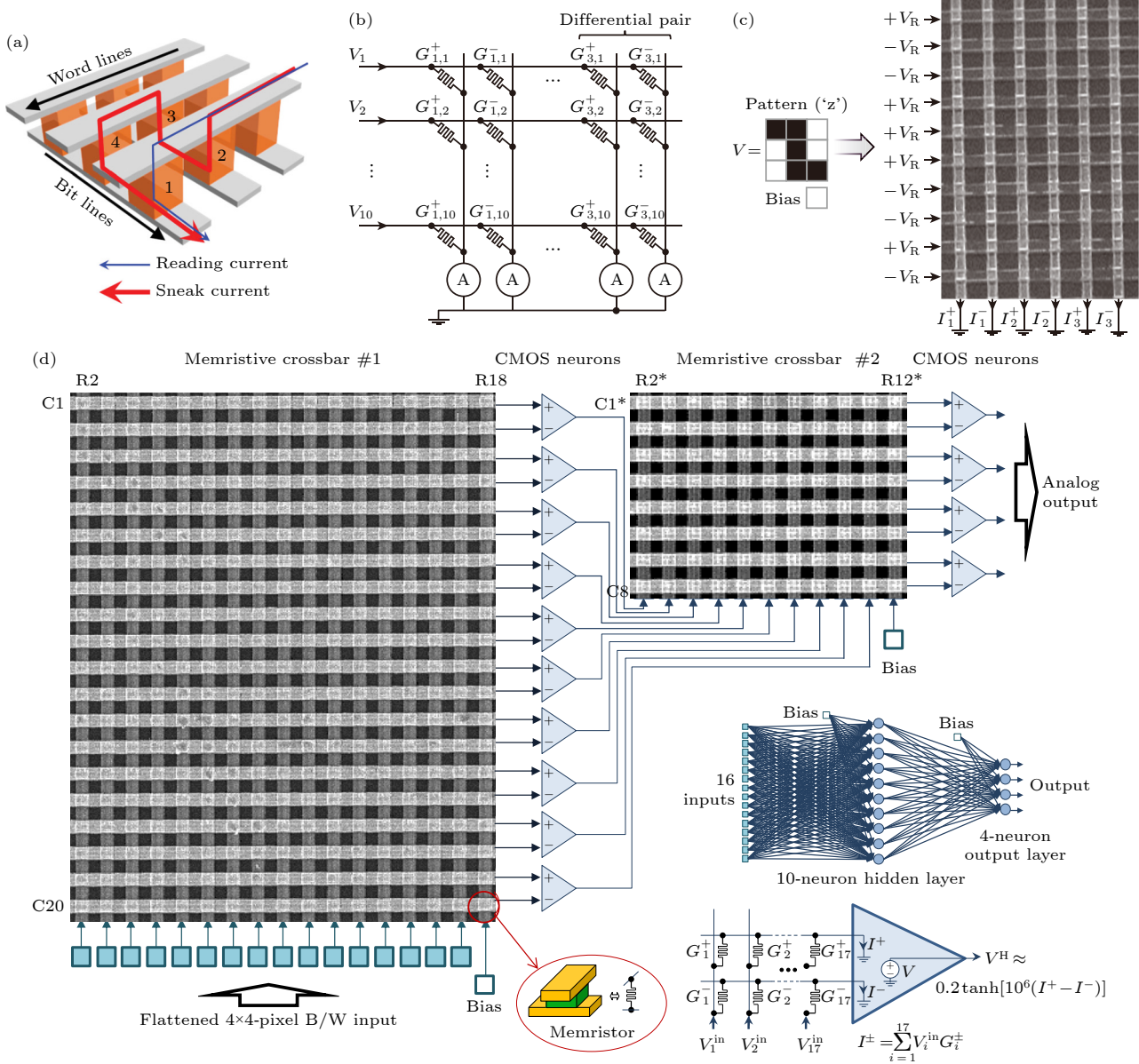


图 14 无源忆阻阵列神经形态计算 (a) 典型的二维 Crossbar 阵列潜行电流读取扰动问题示意图 (细蓝线表示读取电流、粗红线表示潜行电流)<sup>[14]</sup>; (b) 利用  $10 \times 6$  忆阻器 Crossbar 实现的单层感知机网络; (c) 针对特定输入图片 (程式化的字母“z”) 的分类操作示例<sup>[145]</sup>; (d) 两个  $20 \times 20$  的 crossbar 阵列实现双层神经网络的示意图<sup>[146]</sup>

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)<sup>[14]</sup>; (b) an implementation of a single-layer perceptron using a  $10 \times 6$  fragment of the memristive crossbar; (c) an example of the classification operation for a specific input pattern (stylized letter 'z')<sup>[145]</sup>; (d) a schematic diagram of two  $20 \times 20$  crossbar arrays implementing a two-layer neural network<sup>[146]</sup>.



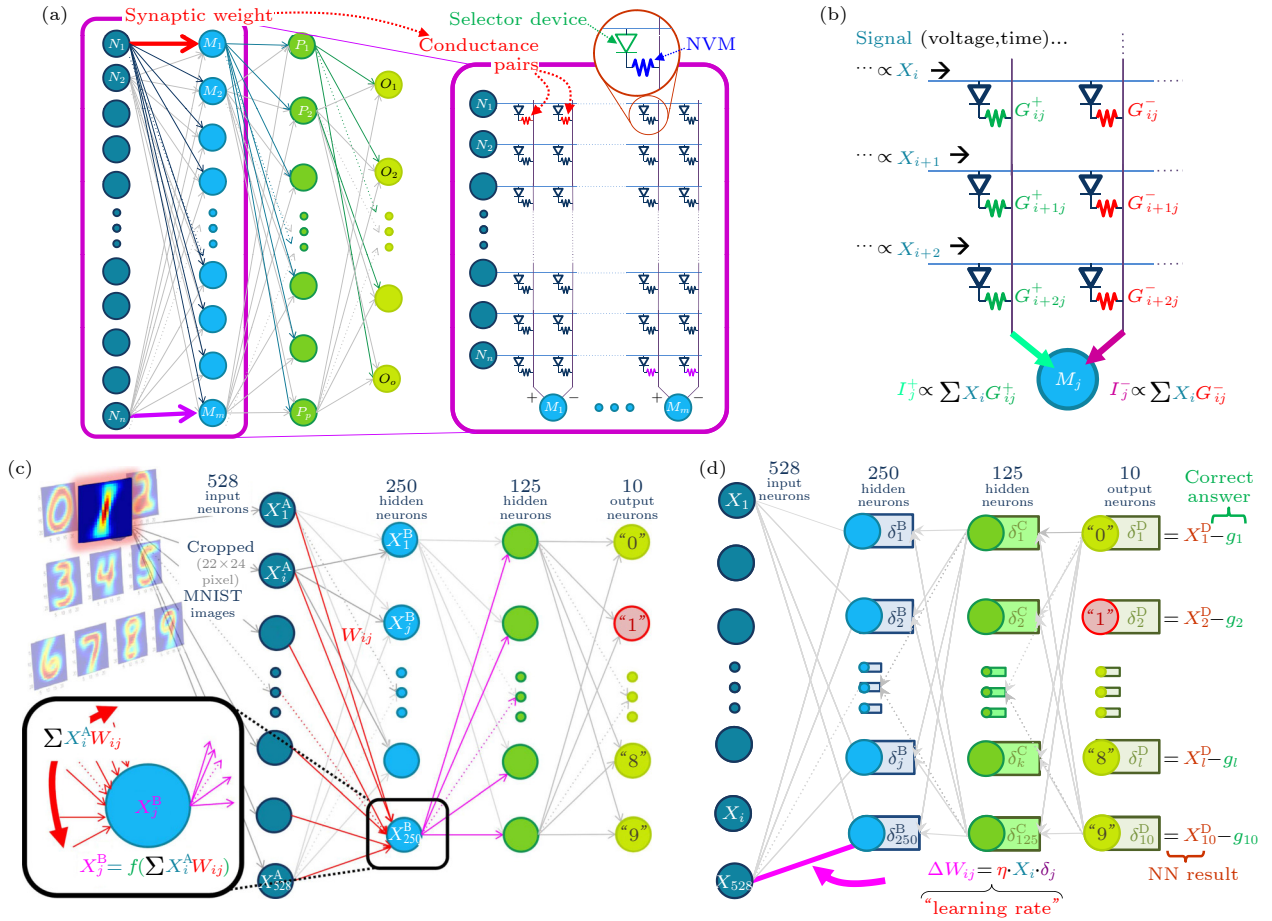


图 15 1S1P 无源忆阻阵列神经形态计算 (a) 存算一体的 1S1P 结构对 crossbar 阵列实现形式; (b) 差分对结构示意图; (c) 神经网络的输入前向传播过程; (d) 神经网络更新示意图<sup>[149]</sup>

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<sup>[149]</sup>.

## 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<sup>[13]</sup>.

IBM 团队<sup>[149]</sup> 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<sup>[156]</sup>. Yang 教授课题组<sup>[157]</sup> 也同样致力于 1T1R 有源阵列的研究. 该阵列主要基于 Ta/HfO<sub>2</sub>/Pt RRAM 与晶体管的串联, 阵列规模为  $128 \times 164$ , 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  $64 \times 54 \times 10$  的 3 层 ANN 网络, 运用与文献<sup>[149]</sup> 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<sup>[158]</sup>. 中国科学院微电子所尚大山课题组<sup>[159]</sup> 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.

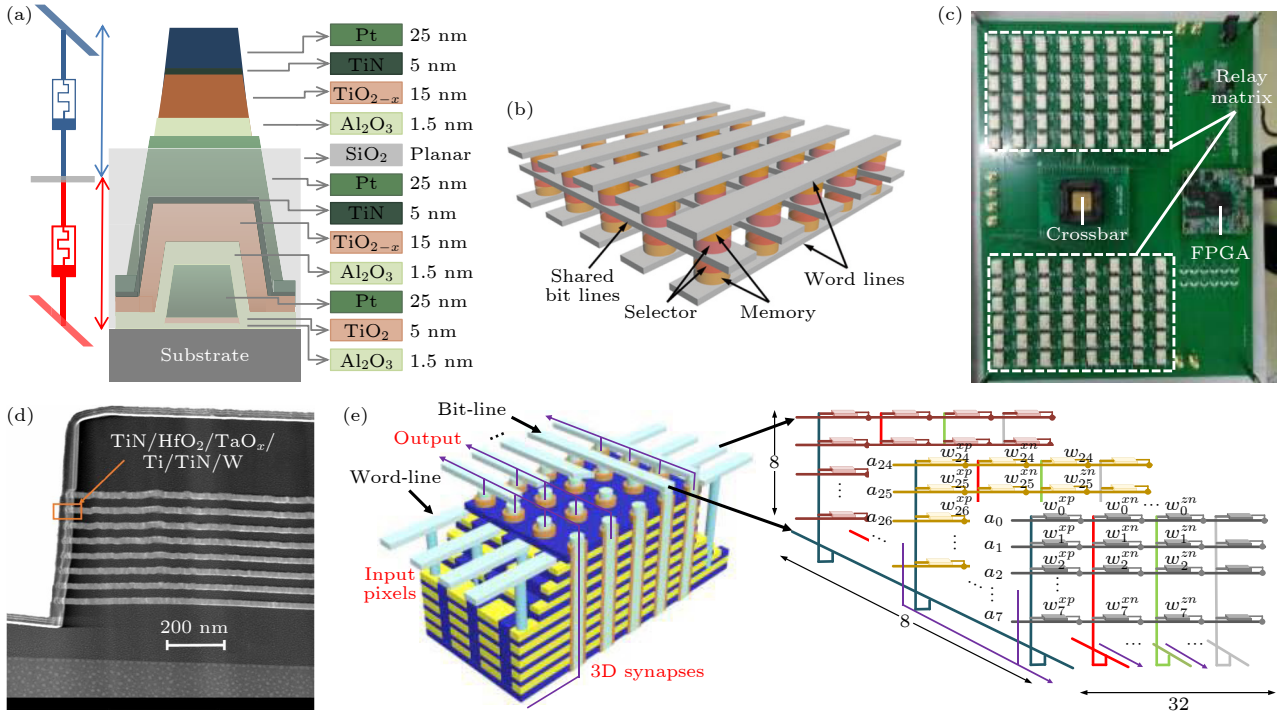


图 16 3D 无源忆阻阵列神经形态计算 (a) 两个 Pt/Al<sub>2</sub>O<sub>3</sub>/TiO<sub>2-x</sub>/TiN/Pt 型忆阻器堆叠结构的等效电路<sup>[150]</sup>; (b) 共享位线结构的 3D Crossbar 阵列<sup>[14]</sup>; (c) 整个电路利用 FPGA 控制的继电器矩阵实现 Crossbar 的自动控制测试; (d) 新的三维 VRRAM 结构的高分辨率透射显微镜图像; (e) 三维 VRRAM 架构中一次卷积操作的电流方向原理图<sup>[154]</sup>

Fig. 16. 3D passive memristive array for neuromorphic computing: (a) Equivalent circuit for two Pt/Al<sub>2</sub>O<sub>3</sub>/TiO<sub>2-x</sub>/TiN/Pt memristors in the stacked configuration<sup>[150]</sup>; (b) a schematic diagram showing the shared bit line structure in cross-line type 3D Crossbar array<sup>[14]</sup>; (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<sup>[154]</sup>.

清华大学吴华强教授和钱鹤教授课题组<sup>[160]</sup>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<sub>x</sub>/HfAlO<sub>x</sub>/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<sup>[161]</sup>. 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<sup>[162]</sup>另辟蹊径, 引入晶体管和电容 (three transistors one capacitor, 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——10<sup>9</sup> 的开关比<sup>[28]</sup>、115 fJ 的开态功耗<sup>[42]</sup>、85 ps 的开关速度<sup>[43]</sup>、10 年以上的保持特性<sup>[44]</sup>及 10<sup>12</sup> 的耐擦写次数<sup>[45]</sup>. 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 Demard 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

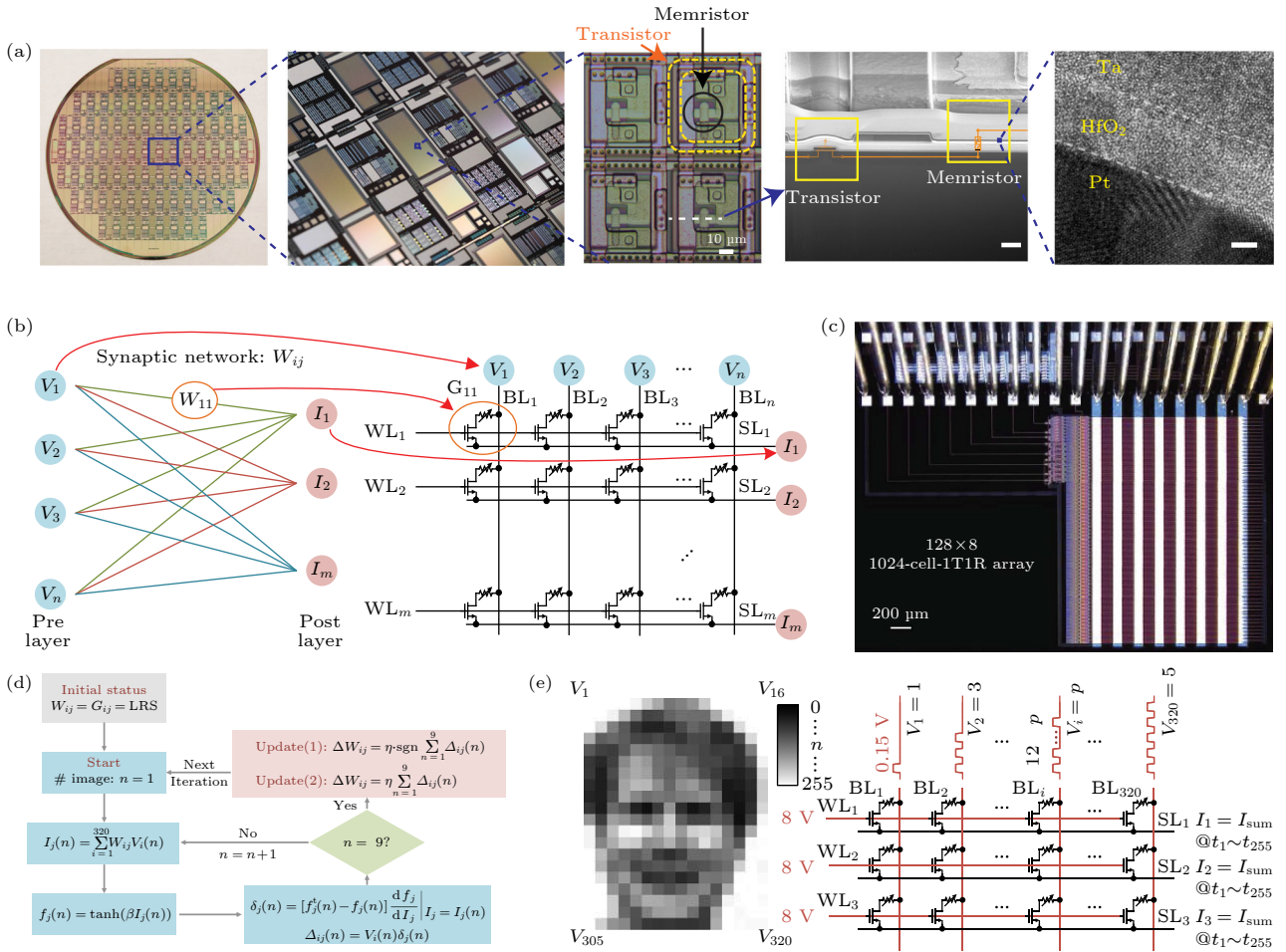


图 17 1T1R 有源忆阻阵列神经形态计算 (a) 用于原位学习的忆阻平台. 从左到右分别是: 带有晶体管阵列的晶圆、芯片特写图、1T1R 单元的显微镜图像、1T1R 单元的 SEM 图像、Ta /HfO<sub>2</sub>/Pt 忆阻器的横截面 TEM 图像<sup>[157]</sup>. (b) 单层神经网络在 1T1R 阵列上的映射. (c) 使用 CMOS 兼容制造工艺制备的 1024 1T1R 单元的阵列显微镜图. (d) 训练过程流程图. (e) 模型映射到输入及并行读取操作的原理图<sup>[160]</sup>

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/HfO<sub>2</sub>/Pt memristor<sup>[157]</sup>. (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<sup>[160]</sup>.

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<sup>[98]</sup>. 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<sup>[33]</sup>, 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<sup>[13]</sup>. 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<sup>[163]</sup>和缪向水团队<sup>[164,165]</sup>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<sup>[166]</sup>, 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

- [1] 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, Palma G, 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, Sidik 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, Perumkunnal M, Bao T H, Rao S, Kim W, Crotti D, Yasin F, Couet S, Swerts J, Kundu S, Yakimets D, Baert R, Oh H, 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, Lehniger 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] Garelo 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, Slesazek 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 Amir 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
- [114] 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, Merrih-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, Merrih-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<sup>\*</sup>

Wen Xin-Yu<sup>1)2)</sup> Wang Ya-Sai<sup>1)2)</sup> He Yu-Hui<sup>1)2)†</sup> Miao Xiang-Shui<sup>1)2)‡</sup>

1) (*Institute of Information Storage Materials and Devices, School of Integrated Circuits,  
Huazhong University of Science and Technology, Wuhan 430074, China*)

2) (*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

<sup>\*</sup> 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).

<sup>†</sup> Corresponding author. E-mail: [heyuhui@hust.edu.cn](mailto:heyuhui@hust.edu.cn)

<sup>‡</sup> Corresponding author. E-mail: [miaoxs@hust.edu.cn](mailto:miaoxs@hust.edu.cn)