

## Progresses and outlook in neuromorphic devices

王洋昊<sup>1,†</sup>, 刘昌<sup>1,†</sup>, 黄如<sup>1,2,\*</sup> and 杨玉超<sup>1,2,\*</sup>

Citation: [科学通报 65, 904 \(2020\); doi: 10.1360/TB-2019-0739](#)

View online: <https://engine.scichina.com/doi/10.1360/TB-2019-0739>

View Table of Contents: <https://engine.scichina.com/publisher/scp/journal/CSB/65/10>

Published by the [《中国科学》杂志社](#)

---

### Articles you may be interested in

#### [Spintronic devices for neuromorphic computing](#)

SCIENCE CHINA Physics, Mechanics & Astronomy **63**, 277531 (2020);

#### [Neuromorphic computing with memristive devices](#)

SCIENCE CHINA Information Sciences **61**, 060422 (2018);

#### [Research progresses of SOI optical waveguide devices and integrated optical switch matrix](#)

Science in China Series F-Information Sciences **48**, 234 (2005);

#### [Recent progresses on the new condensed forms of single-walled carbon nanotubes and energy-harvesting devices](#)

Chinese Science Bulletin **57**, 181 (2012);

#### [Research progresses of dynamic random access memory devices](#)

SCIENTIA SINICA Physica, Mechanica & Astronomica **46**, 107304 (2016);

---

# 神经形态器件研究进展与未来趋势

王洋昊<sup>1†</sup>, 刘昌<sup>1†</sup>, 黄如<sup>1,2\*</sup>, 杨玉超<sup>1,2\*</sup>

1. 北京大学信息科学技术学院微纳电子学系, 北京 100871;

2. 北京大学人工智能研究院, 北京 100871

† 同等贡献

\* 联系人, E-mail: [rhuang@pku.edu.cn](mailto:rhuang@pku.edu.cn); [yuchaoxy@pku.edu.cn](mailto:yuchaoxy@pku.edu.cn)

2019-11-14 收稿, 2019-12-19 修回, 2019-12-25 接受, 2019-12-26 网络版发表

国家自然科学基金(61925401, 61674006, 61927901, 61421005)、国家重点研发计划(2017YFA0207600)和北京智源人工智能研究院智源青年科学家项目资助

**摘要** 大数据时代的信息爆炸、摩尔定律的逐渐减缓和“万物互联”的最终愿景使得发展高性能的非传统计算迫在眉睫。21世纪以来, 神经形态计算以高度的并行、极低的功耗和存算一体的特征受到了广泛的关注。其中, 具有独特物理机制的神经形态器件是神经形态计算硬件的基本组成单元, 对新型非冯·诺依曼架构芯片的研发乃至类脑智能的最终实现都具有重要意义。本文重点介绍了神经形态器件的研究进展和未来研发的趋势。研究低功耗的神经形态器件与集成方法, 提高突触器件的线性度、对称性和开关比以及从动力学角度模拟生物启发的神经系统是该领域的研究热点。

**关键词** 神经形态器件, 类脑计算, 人工智能, 人工突触, 人工神经元, 忆阻器

20世纪80年代末, 加州理工学院的Mead<sup>[1]</sup>首次提出了“神经形态计算”的概念: 利用超大规模集成的芯片, 采取模拟信号仿照生物大脑中的神经系统来处理信息。当前的集成电路产业面临着摩尔定律的逐渐放缓, 尺寸缩小过程中能耗的急剧升高, 以及存算分离导致的冯·诺依曼瓶颈等问题, 必须采用新的方法、新的硬件来补充与发展现有的计算技术。而神经形态计算具有高并行、低功耗以及存算融合的特征, 能在更小面积上更快、更高效地解决包括非结构化数据处理在内的计算任务, 并在可靠性、可扩展性与容错率方面相较数字系统具有天然的优势<sup>[2]</sup>。目前工作中, 人们利用互补金属氧化物半导体(complementary metal oxide semiconductor, CMOS)技术在人工神经元和突触电路搭建方面取得了很大进展, 并证明了大规模构筑脑启发电路系统的可行性<sup>[3]</sup>, 如IBM的TrueNorth芯片<sup>[4]</sup>、斯

坦福大学的Neurogrid芯片<sup>[5]</sup>、Intel的Loihi芯片<sup>[6]</sup>等都是近年来富有重要意义的尝试。

随着材料科学的发展和新型结构器件的发明, 越来越多的新器件被用来构筑神经形态电路。这些器件依靠独特机理, 能够直接从物理层面模拟突触和神经元的行为, 因此被称作神经形态器件<sup>[7]</sup>。大多数神经形态器件都是从新型存储器件逐渐发展而来, 能直接用于构建存储与计算在同一个物理位置的神经形态系统。目前, 神经形态器件可以模仿生物神经元和突触的很多已知功能, 集成到阵列中可以直接依靠物理定律<sup>[8]</sup>来执行大规模并行的高效存内计算, 例如加速人工神经网络训练和推断中的向量矩阵乘法<sup>[9]</sup>, 目前已经实现了信息编码、数据分类和强化学习等经典的机器学习任务。基于这种原理制成的专用芯片可以与GPU互为补充, 共同加速处理信息时代下的海量数据。

**引用格式:** 王洋昊, 刘昌, 黄如, 等. 神经形态器件研究进展与未来趋势. 科学通报, 2020, 65: 904–915

Wang Y H, Liu C, Huang R, et al. Progresses and outlook in neuromorphic devices (in Chinese). Chin Sci Bull, 2020, 65: 904–915, doi: [10.1360/TB-2019-0739](https://doi.org/10.1360/TB-2019-0739)

然而,仅仅在表象上实现神经形态功能是远远不够的。为了与现有的CMOS技术竞争、与大脑中存在的突触和神经元媲美,人工神经形态器件必须具有高度集成的潜力、超低的功耗以及在原理层面逼真地模拟突触可塑性和神经元发放行为的能力。近几年来,人工突触器件的尺寸显著缩小,最低可以达到2 nm<sup>[10]</sup>。突触器件的集成方式大多采用十字交叉阵列,既可平面集成,也可三维堆叠,每一个交叉节点即为一个人工突触器件。器件结构简单、加工流程简便、集成密度高,显示出未来工业化的巨大潜力。器件的运算过程直接采用基于物理定律与材料特性联合作用的模拟计算,而不是基于逻辑门的电路二值运算,从而能够带来功耗上的大幅降低,在物联网时代可用于超低功耗的智能边缘计算。基于物理原理模拟生物功能的人工神经元拥有比CMOS更为简洁高效的电路,利用材料的物理特性可直接制作出更为复杂的振荡神经元、变阈值神经元、混沌神经元等,运用到计算中可具有更加强大的性能。近5年的研究还表明,可在物理层面直接根据生物中钙离子、钾离子、钠离子等微观组分迁移的原理模拟生物突触可塑性,诸如突触晶体管、异源性突触器件和多端物理演化器件等大量涌现,为人类在芯片上验证与开发大脑功能提供了更加丰富的选择。总而言之,这些器件在降低单个器件功耗的同时,可以更好地模拟生物中观测到的、多样化的生物信息处理机制,为计算神经科学中算法的实现提供了重要的硬件基础,未来将是人类理解大脑并在芯片上实现大规模类脑计算功能的重要元器件基础。

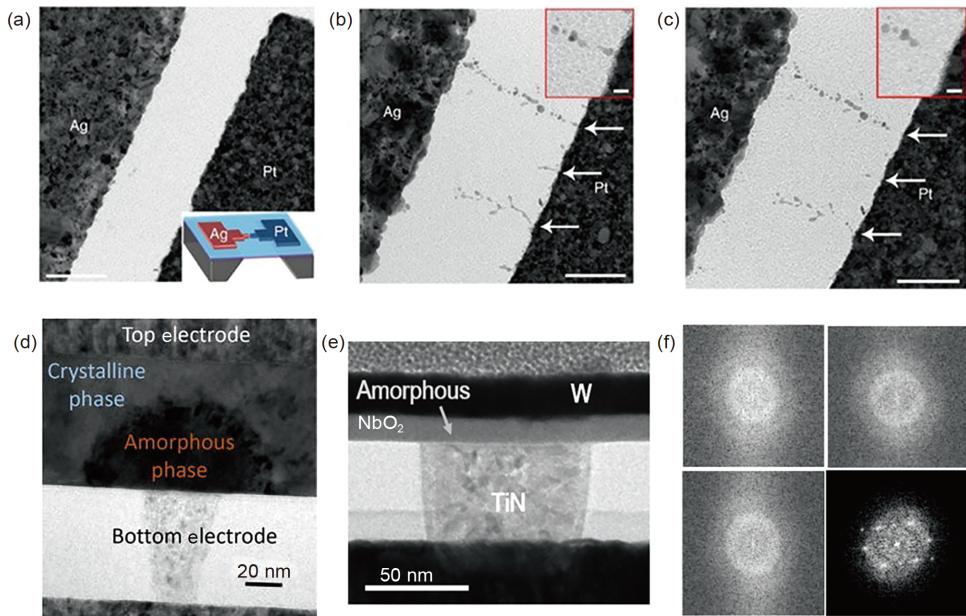
## 1 神经形态器件物理机制

在过去的十几年里,各种新型的神经形态器件被开发出来,按物理机制划分,神经形态器件主要有离子迁移器件、材料相变器件、电子迁移器件和磁性器件等。不同的物理机制从根本上决定了神经形态器件的行为,也指导了新结构、新材料神经形态器件的设计。下文将主要介绍与现有CMOS工艺兼容的离子迁移器件和材料相变器件。

离子迁移器件通常采用金属-绝缘体-金属(metal-insulator-metal, MIM)结构。离子迁移的基本原理是利用外部刺激,例如电场,来驱动离子在材料内部移动,并伴随不断的氧化还原反应生长出纳米尺度的导电细丝,最终导致可测量的材料电学特性变化<sup>[11]</sup>,例如器件的电阻降低;而在施加反向刺激后,导电细丝将逐渐溶

解,使得器件的电阻重新升高,从而可以通过外部的电学信号有效地调节器件的电阻值,器件高低阻态转变的时间尺度能够达到亚纳秒量级。**图1(a)~(c)**展示了在透射电子显微镜(transmission electron microscope, TEM)观测下纳米尺度导电细丝的生长、溶解过程。根据驱动离子种类的不同,离子迁移型器件可进一步分为由活性金属阳离子(如Ag<sup>+</sup>、Cu<sup>x+</sup>)输运导致的电化学金属化器件,以及由氧阴离子输运导致的价态变化器件。然而,导电细丝的形状大小、组分和生长过程的随机性均会影响离子迁移器件电阻值调控的精确性。例如,绝缘层中缺陷的类型与分布将影响导电细丝的循环变化过程,进而导致器件的噪声<sup>[12]</sup>。影响导电细丝生长的因素还有材料中的离子迁移速率、氧化还原反应速率、界面粗糙度和接触条件等<sup>[13]</sup>。为解决这些问题,研究人员在理论分析与大量实验的情况下,发展了基于HfO<sub>x</sub>、TaO<sub>x</sub>材料的器件,以获得更好的调控性能与非易失存储能力;此外,还可以通过引入离子扩散限制层来均匀化导电细丝的生长速度,进而改善电导调制的线性度<sup>[14]</sup>。最近研究发现,温度也是离子迁移型器件一个重要的可以调控的隐状态变量。例如,在基于TaO<sub>x</sub>/Ta<sub>2</sub>O<sub>5</sub>的二阶忆阻器中,器件的温度可以看作影响器件电导改变的隐变量,通过施加脉冲序列调控器件内部温度,可实现时间编码的计算任务<sup>[15]</sup>。这些发现增加了我们对离子迁移器件的理解,但需要更进一步的研究来表征导电细丝的微观结构,建立导电丝生长/溶解动力学模型以分析生长/溶解的速率限制,以帮助我们更好地选择材料体系与调控电压脉冲参数,进而制作出稳定且性能优良的器件。

相变器件是一类历史悠久、工艺较为成熟的神经形态器件,其外加刺激一般为电流脉冲,利用焦耳热引发的相变过程来改变器件的电导。相变器件,根据相变材料的机理可以分为两种,一种是基于Ge<sub>2</sub>Sb<sub>2</sub>Te<sub>5</sub>等材料、利用晶态-非晶态之间转变实现电导变化的相变存储器(phase change memory, PCM)<sup>[16]</sup>,其结构为金属/相变材料/金属;另一种是基于NbO<sub>2</sub>和VO<sub>2</sub>材料利用金属到绝缘体的相变制成的莫特(Mott)器件<sup>[17]</sup>,其结构为金属/氧化物/金属。对于PCM,晶态为高电导,非晶态为低电导。若材料初始处于高电导的结晶态,当施加高幅度的电流脉冲时,由于焦耳热,很大一部分相变材料熔化,一旦脉冲突然中断,熔融材料会淬火成低电导的非晶相,从而转变为低电导。若再施加电流脉冲加热材料达到结晶的温度却低于熔点,将导致部分非晶区域重



**图1** (网络版彩色)器件内部发生动力学过程表征. (a)~(c) 在透射电子显微镜下离子迁移型器件导电细丝的生长和溶解<sup>[11]</sup>, Copyright © Springer Nature; (d) 在透射电子显微镜下PCM器件相变过程<sup>[16]</sup>, Copyright © AIP Publishing; (e) 在透射电子显微镜下NbO<sub>2</sub>莫特器件的结构<sup>[17]</sup>, Copyright © AIP Publishing; (f) 莫特器件转变过程下的电子束衍射图<sup>[18]</sup>, Copyright © Springer Nature

**Figure 1** (Color online) Characterization of the dynamic process occurring inside the device. (a)–(c) Growth and dissolution of conductive filaments in the ion transfer device under the TEM<sup>[11]</sup>, Copyright © Springer Nature; (d) phase transition process in the PCM device under the TEM<sup>[16]</sup>, Copyright © AIP Publishing; (e) structure of NbO<sub>2</sub> Mott device under the TEM<sup>[17]</sup>, Copyright © AIP Publishing; (f) selected electron diffraction pattern in the transition process of the Mott device<sup>[18]</sup>, Copyright © Springer Nature

结晶。结晶程度取决于脉冲的幅度和持续时间。通过施加连续的脉冲使非晶区域逐渐结晶，可以实现连续的电导变化。**图1(d)**展示了在透射电子显微镜下PCM晶体的生长。对于莫特器件，其可同时表现出由电流控制和温度控制的负微分电阻效应。**图1(e)**展示了莫特器件的截面结构。**图1(f)**用电子束衍射的方式展示了绝缘层中由电流控制与温度控制的负微分电阻状态。惠普公司的研究人员<sup>[19]</sup>在基于NbO<sub>2</sub>的莫特器件中把器件与电容并联，基于器件温度和电容电荷两个状态量构造的振荡器呈现出混沌行为。

## 2 神经形态器件计算原理

神经形态器件是模拟生物神经系统中基本信息处理单元的器件，可分为人工突触器件和人工神经元器件两大类。生物中，神经元是大脑神经系统的基本组成细胞，它能在膜电位上整合各个接收到的脉冲信号加权后的总和，达到阈值时刻将发放出新的脉冲信号，实现信息传递中级联的过程。突触是一个神经元的脉冲信号传递给另一个神经元的枢纽，突触强度代表突触后神经元接收到突触前神经元发放的脉冲信号后膜电

位的增量，并且突触强度会随着前后神经元的发放历史而改变，实现主体的计算功能。目前，CMOS电路制作单个人工神经元或突触，往往需要使用几十个晶体管，利用数模混合的复杂电路在数据表象上等效突触与神经元的计算功能函数，难以兼顾集成度与性能，并且难以切换表现函数以适应不同的工作环境。人类的大脑中存在约10<sup>11</sup>的神经元以及10<sup>15</sup>的突触，这些神经元和突触的形态性能各异，功耗却只有20 W，利用CMOS技术模拟人脑的代价之大、实现之复杂我们难以承受。而对于神经形态器件，它们可直接凭借单个器件的内在物理机制建立对应关系，模拟生物信息处理过程，在功耗与集成度方面具有显著优势，并可进一步利用材料和器件的特性实现工作机制的变换，未来有望用于新一代类脑网络中。

人工突触器件需要同时具备记忆和计算功能，在网络设计中占据了大部分的面积和功耗，是类脑计算研究的关键。现阶段，国内外研究机构在通过离子迁移器件、相变存储器和自旋转移矩磁存储器等新型器件模拟突触的功能方面取得了重要进展。上述器件虽然物理机制各异，但具有一个共同的特征，即可以通过外

加的电学脉冲改变器件的等效电阻值并存储在器件中，一般将这些器件统称为“忆阻器”<sup>[20,21]</sup>。图2(a)展示了忆阻器作为第4个基本电路元件的理论对称性；图2(b)展示了忆阻器的伏安特性图。若将生物中神经元发放的脉冲信号在电路中对应为电压信号，把忆阻器的电导值视为突触的权重，则根据欧姆定律电压信号经过忆阻器调制后转成电流信号，进而可通过忆阻器交叉阵列结构(图2(c))，借助基尔霍夫定律的加法实现电压向量与电导矩阵的乘积，对应神经元膜电位的整合功能；同时，脉冲式的电压信号可在到达忆阻器的同时改变忆阻器的电导，实现连续的学习过程，此即生物中存在

的学习模式。而在人工神经网络算法的简化模型中，进一步地把信息编码为脉冲信号的幅值或宽度，在需要计算时，电压信号采取较小的“读”电压值，施加于阵列上，于是可直接通过外部电路计算神经元的输出电流；根据实际输出、目标输出与相应的学习算法，在下一个时钟周期内施加一个较大幅值的“写”电压脉冲，调控阵列中忆阻器的电导，从而起到训练的效果。如图2(d)所示，此即当前人工神经网络在神经形态器件阵列上实现的基本原理。其中，向量矩阵乘法通常是人工神经网络计算最密集的部分，这部分运算可完全在忆阻器交叉阵列中大规模并行，依靠物理定律高效地完成，

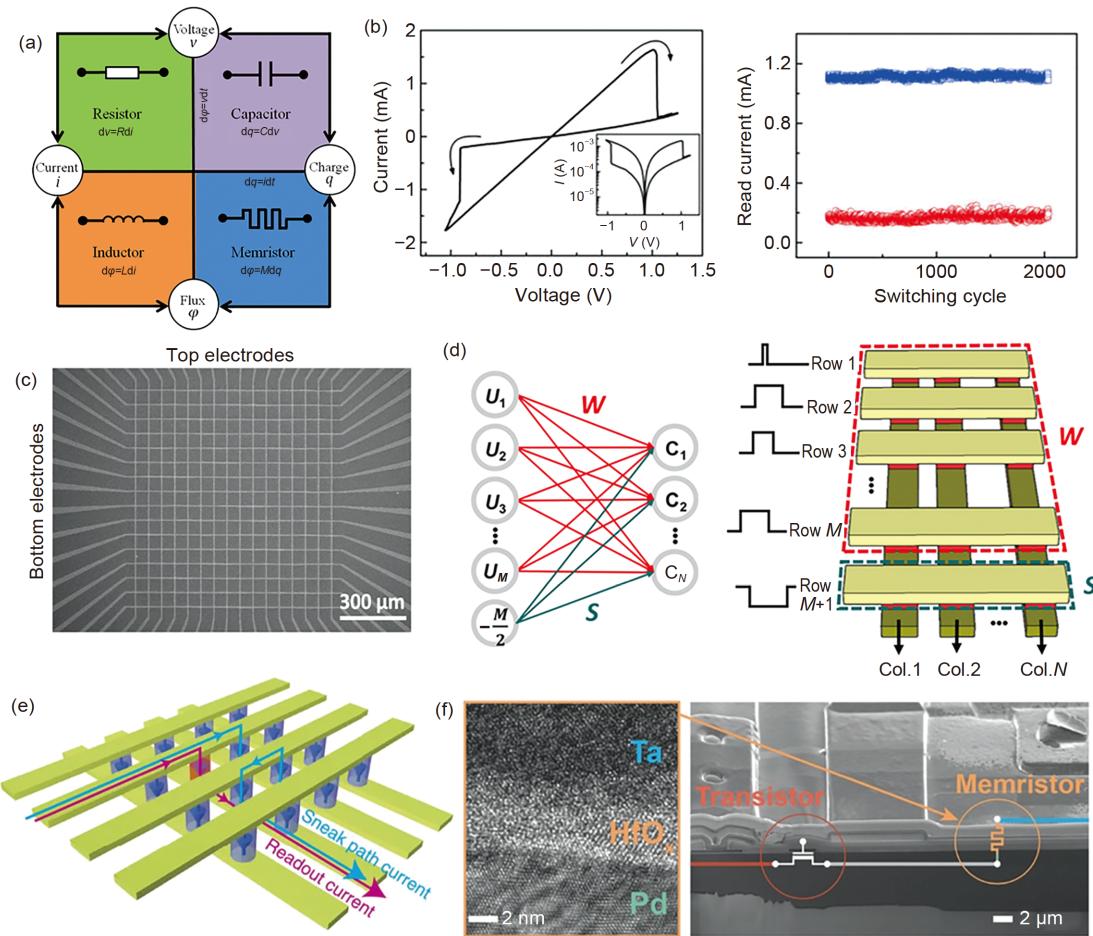


图2 (网络版彩色)忆阻器的计算原理。(a) 忆阻器作为第4个基本电路元器件;(b) 忆阻器的典型I-V特性图, 在不同扫描次数下高低阻态的稳定性<sup>[22]</sup>, Copyright © AIP Publishing; (c) 扫描电子显微镜下基于TaO<sub>x</sub>的忆阻器交叉阵列; (d) 人工神经网络映射到阵列计算的示意图<sup>[23]</sup>, Copyright © American Chemical Society; (e) 潜行电流的影响<sup>[25]</sup>, Copyright © Springer Nature; (f) 隧道电子显微镜下1T1R的单元结构<sup>[26]</sup>, Copyright © John Wiley and Sons

**Figure 2** (Color online) Computing principle of memristor. (a) Memristor as the fourth basic circuit component; (b)  $I$ - $V$  characteristic diagram of memristor, and the endurance characteristics of memristor<sup>[22]</sup>. Copyright © AIP Publishing; (c) crossbar structure of TaO<sub>x</sub> based memristor under SEM (scanning electron microscopy); (d) schematic diagram of array calculation mapping on crossbar<sup>[23]</sup>, Copyright © American Chemical Society; (e) influence of the sneak path<sup>[25]</sup>, Copyright © Springer Nature; (f) 1T1R structure under TEM<sup>[26]</sup>, Copyright © John Wiley and Sons

实现了硬件加速人工神经网络，并且具有超低的功耗。器件的存储能力体现在“写”操作时的电导调节与非易失特性，计算能力体现在“读”操作时候的向量矩阵乘法，从而达到存算一体。当前器件上的研究主要关注在“写”操作过程中电导调制的线性度与对称性，以及“读”操作过程中的权值精度、开关比、一致性等。此外，当交叉阵列规模较大时，电极电阻和寄生电流等寄生效应的存在会严重影响阵列的性能<sup>[24]</sup>(图2(e))。部分寄生效应可通过加入选择器改善，例如加入晶体管作为选择器与忆阻器串联制成1T1R结构(图2(f))，通过对晶体管栅压的调控可以提高读写操作的稳定性<sup>[27]</sup>；也可以使用不对称的电阻转换层来制备出互补式忆阻器，如Pd/Ta<sub>2</sub>O<sub>5-x</sub>/TaO<sub>y</sub>/Pd<sup>[22]</sup>使得单元的阻值恒为高阻，独立于存储的内容从而有效避免漏电流的存在。

当前用忆阻器阵列实现基础的人工神经网络算法已经卓有成效。按神经网络的训练方法分，在无监督学习上已实现主成分分析<sup>[28]</sup>、稀疏编码<sup>[29]</sup>、联想学习<sup>[30]</sup>、生成对抗网络<sup>[31]</sup>等；在监督学习上已实现单层感知机、多层感知机<sup>[32]</sup>、卷积神经网络<sup>[26]</sup>、长短期记忆网络<sup>[33]</sup>、循环卷积神经网络<sup>[34]</sup>等。研究者们还在阵列上利用不同方式实现了强化学习<sup>[35,36]</sup>。2019年，密歇根大学研究团队制作出通用可编程忆阻器：CMOS混合架构加速器芯片<sup>[37]</sup>，阵列规模为54×108，推动了忆阻器的商用化进程。但是，构筑深度神经网络加速硬件需要大规模的神经形态器件，若不考虑片上资源复用，仅仅实现LeNet<sup>[38]</sup>就需要约34万人工突触和1万人工神经元，其中单层阵列规模的最大要求为100×1516。若需要处理更加复杂的机器视觉与自然语言处理等任务，硬件代价则变得尤为庞大。因此，在保证单器件性能的情况下，提高忆阻器件的集成度是目前将忆阻器投入实际应用中的主要挑战之一。

### 3 神经形态器件研究进展

神经形态器件的另一个重要意义就是应用离子动力学逼真地模拟生物突触可塑性与神经元的工作模型，为未来的类脑算法提供硬件基础。对于突触可塑性，其可分为长时程可塑性与短时程可塑性。长时程可塑性中突触连接强度不会随着时间而变化，短时程可塑性中连接强度将随时间而改变。具体地，突触可塑性又可分为长时程增强(long-term potentiation, LTP)、长时程抑制(long-term depression, LTD)、短时程增强(short-term potentiation, STP)和短时程抑制(short-term depression,

STD)，具有代表性的包括脉冲时间依赖可塑性(spike timing dependent plasticity, STDP)。STDP是一种特殊的长时程可塑性，它表示突触的连接强度会随突触前后神经元发放脉冲的间隔呈指数增加或衰减，但增加与衰减的作用是长时间存在的。从神经科学的角度来看，大脑中的记忆分布在由突触和神经元组成的神经网络中，存储在神经网络的等效能量最低点处<sup>[39]</sup>；而大脑在感知和学习的过程中工作在混沌的边缘<sup>[40]</sup>，具有强大的全局搜索能力和实时计算能力<sup>[41]</sup>。上述功能理论上完全可以凭借具有时间动力学特性的突触与神经元动态地实现。脑中的记忆根据时间长短的不同，可分为短时记忆和长时记忆两种，两者的可逆转换取决于突触短时程可塑性与长时程可塑性的转换，这种转换同样也是生物个体适应环境的关键因素。此外，短时程可塑性被认为是大脑中注意力机制产生的根源<sup>[42]</sup>，注意力产生启动效应，使得大脑更容易编码与学习时空信息。

当前的人工神经网络仅间接采用了LTP和LTD的思想，并未涉及时间因子的直接作用，而时间的引入将使得神经网络具有更强的信息蕴含能力与处理能力<sup>[43]</sup>。同时，基于CMOS的神经形态计算硬件，例如IBM公司的TrueNorth，仅仅只能从表象上体现脉冲神经系统的行为，并未深入到硬件原理层面的模拟。因此，新型神经形态器件的研究至关重要。2010年，美国密歇根大学研究团队<sup>[7]</sup>基于银掺杂非晶硅的忆阻器率先实现了生物突触的长时程可塑性模拟，并精心设计神经元脉冲的波形，实现了STDP功能(图3(a))。2011年，法国的研究团队<sup>[44]</sup>使用基于相变原理的PCM同样实现了长时程可塑性，并将其用于神经网络的分类任务中。2015年，密歇根大学研究人员<sup>[45]</sup>以温度作为隐状态变量来编码时间，在基于TaO<sub>x</sub>的二阶忆阻器中通过低压脉冲调节器件温度从而影响高压“写”脉冲的作用能力，更加自然地实现了生物突触的STDP功能。2016年，北京大学研究人员<sup>[14]</sup>在基于TiN/TaO<sub>x</sub>的器件中引入离子扩散限制层，控制导电细丝生长的速率，实现了高线性度的突触长时程可塑性模拟。2017年，美国马萨诸塞大学研究人员<sup>[46]</sup>开发了基于银离子的扩散忆阻器，模仿生物突触中Ca<sup>2+</sup>的行为，从而实现了短时程可塑性，并首次将扩散忆阻器和非易失性忆阻器串联，实现了突触的STDP功能；英国牛津大学研究人员<sup>[47]</sup>创新地利用相变材料和氮化硅波导结合，构建了片上可集成的光子突触，它具有超快的速度，几乎无限的带宽，并且无互联功率的损耗；同年，密歇根大学研究人员<sup>[48]</sup>使用基于

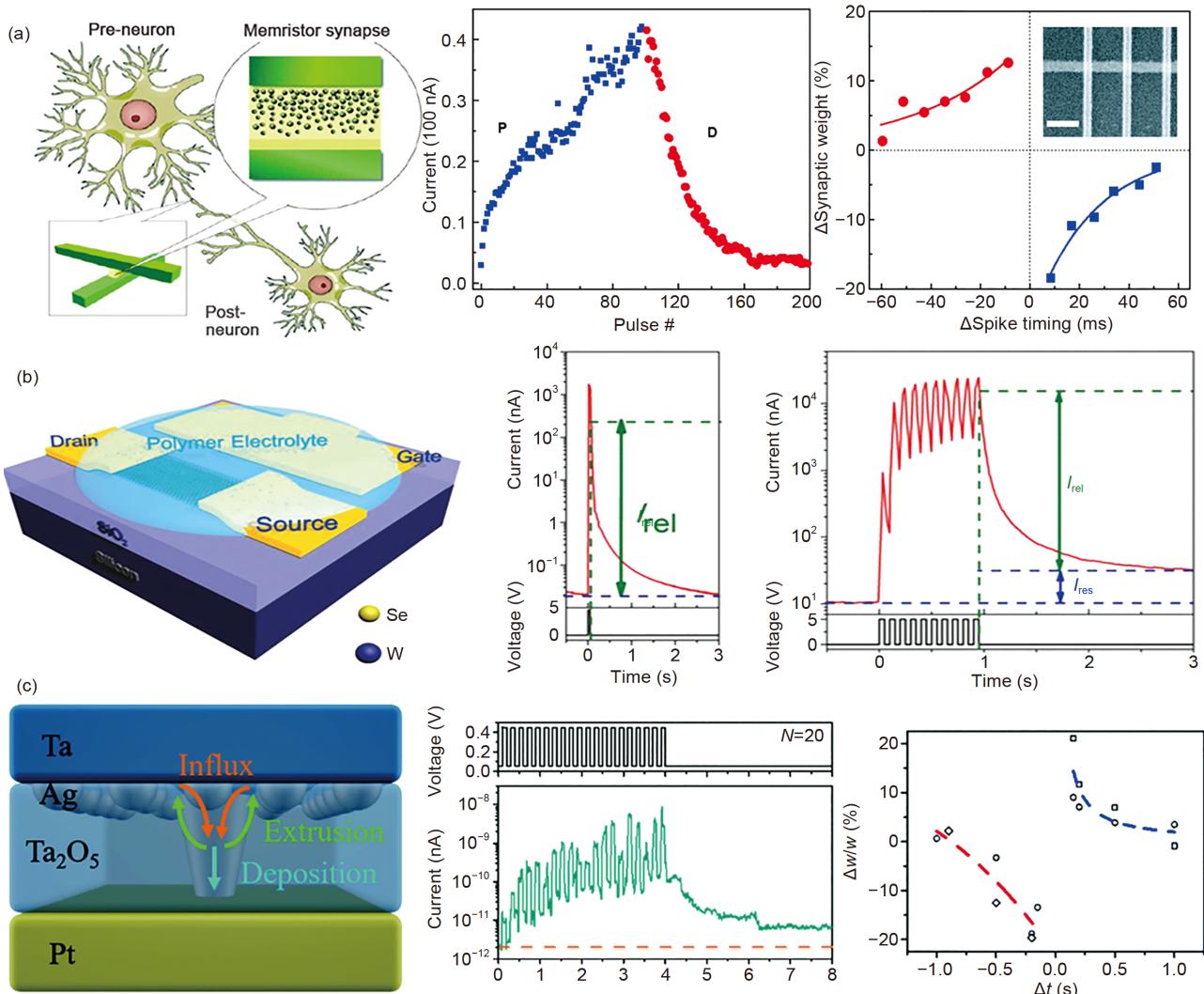


图3 (网络版彩色)人工突触器件. (a) 生物突触与忆阻器的对应, 调制脉冲波形实现LTP、LTD与STDP<sup>[7]</sup>, Copyright © American Chemical Society; (b) 利用突触晶体管实现LTP与STP的相互转换<sup>[49]</sup>, Copyright © John Wiley and Sons; (c) 利用银离子团簇的迁移实现LTP与STP的共存, 以及脉冲时间依赖可塑性<sup>[52]</sup>, Copyright © Royal Society of Chemistry

**Figure 3** (Color online)Artificial synaptic devices. (a) The correspondence between biological synapse and memristor, and the modulation of pulse waveform to realize LTP, LTD & STDP<sup>[7]</sup>, Copyright © American Chemical Society; (b) LTP and STP realized using synaptic transistors<sup>[49]</sup>, Copyright © John Wiley and Sons; (c) coexistence of LTP and STP and pulse time dependent plasticity using silver cluster migration<sup>[52]</sup>, Copyright © Royal Society of Chemistry

WO<sub>x</sub>的易失性忆阻器, 利用短时程可塑性将一定时间内的脉冲序列通过STP与其时间衰减因子的相互作用进而实现区分, 构造了大脑中存在的回声机/库网络计算模型; 2018年, 北京大学科研人员<sup>[49]</sup>展示了基于WSe<sub>2</sub>沟道和聚合物电解质的突触晶体管, WSe<sub>2</sub>沟道发生的离子吸附和嵌入过程模仿生物中的钙离子运输过程, 分别导致了短时程和长时程可塑性, 并通过调节脉冲信号频率可以同时精确调控这两种可塑性(图3(b)). 在高度模拟生物突触可塑性的同时, 该器件还在LTP和LTD过程中的线性度、对称性和能耗等方面达到了国

际领先水平. 2019年, 密歇根大学研究人员<sup>[50]</sup>使用基于Li<sub>x</sub>MoS<sub>2</sub>的二维材料器件, 实现了突触长时程增强和抑制功能, 并从器件层面上模拟了突触间的竞争与协作; 此外, 密歇根大学研究人员进一步拓展了短时程可塑性下的回声机/库网络模型, 通过引入反馈信号, 实现了语音识别与混沌序列预测<sup>[51]</sup>; 同年, 北京大学研究人员<sup>[52]</sup>利用银离子的迁移过程, 研发出突触与神经元功能一体化器件, 通过改变脉冲幅度实现了突触长短时程可塑性的相互转换(图3(c)), 并可用于实现泄漏累积释放型(leaky integrate-and-fire, LIF)神经元.

此外，科研人员还受真实生物大脑中突触形态的多样性启发，拓展出多端的神经形态器件。神经科学的观测发现，异源突触可塑性中神经元之间的突触强度可以通过附加的中间神经元调制，即中间的神经元可以调控另外两个神经元之间突触连接强度的变化，这在脑皮层的分层网络中具有重要应用。2015年，密歇根大学研究人员<sup>[53]</sup>利用银纳米团簇在电介质中的自组织原理，提出了可模拟异源突触可塑性的物理演化网络(图4(a))。2017年，北京大学研究人员<sup>[54]</sup>使用基于 $\text{HfO}_x$ 和 $\text{TaO}_x$ 的三端器件模拟了异源突触可塑性(三端器件结构，图4(b))，通过在第三端加入调制的电压信号，可以影响LTD与LTD的变化强度(图4(c))，并演示了变学习率的神经网络，这类三端器件非常适合神经形态硬件中的学习。目前的神经网络都是人为调整设定神经网络中的超参数，其先决条件是已经用软件进行过大量的仿真，但是智能化应用更期待一个可以进行自主学习的片上神经网络，利用三端器件结合learning-to-learn算法<sup>[55]</sup>，可以在硬件上自主推断出神经网络学习所需要的各个超参数进而运用到各式学习任务中，因此三端器件可变学习率概念的提出具有重要的意义。此外，基于第三端调制信号的算法被认为是具有替代反向传播的潜力，进而将演化出生物中可行的神经网络。2018年，美国西北大学研究人员<sup>[56]</sup>进一步将异源突

触可塑性扩展到了基于单层 $\text{MoS}_2$ 的多端薄膜晶体管中，获得了更好的调控性能与可扩展性。

目前为止，新型神经形态器件已经数目众多，但在应用层面上突触的动态计算能力仍有待进一步发掘，研究人员需要在大量新器件中做抉择，思考包含具体器件特性的计算功能。最新研究发现，树突也具有强大的计算功能，可执行数模混合的复杂运算<sup>[57]</sup>，将人类对大脑计算能力的认知刷新为原先的近100倍，其可作为数据大规模预处理与编码的模块，将极大地提高神经网络信息处理的能力，理论上可以用忆阻器实现。此外，发展生物上可行并且性能优异的脉冲神经网络训练算法也是实现新型类脑神经网络研究的重中之重。欧洲研究人员<sup>[58]</sup>提出使用资格迹表示突触内部的分子亚稳态，在液态机的理论上建立了e-prop算法来适应神经形态硬件中循环神经网络的计算<sup>[59]</sup>，彻底抛弃了原先循环神经网络按时间反向传播、缺乏生物可信度的训练算法，并在软件模拟中达到了相近的精度。

作为神经形态计算的另一类基本元件，人工神经元的硬件实现也具有重要意义。目前利用计算机软件模拟脉冲神经元的行为需要求解微分方程，运算时间慢并且能耗巨大。而在硬件设计中使用CMOS电路模拟生物神经元往往需要数十个晶体管，功耗和集成密度都受到极大的限制。2012年，惠普实验室研究团队<sup>[60]</sup>

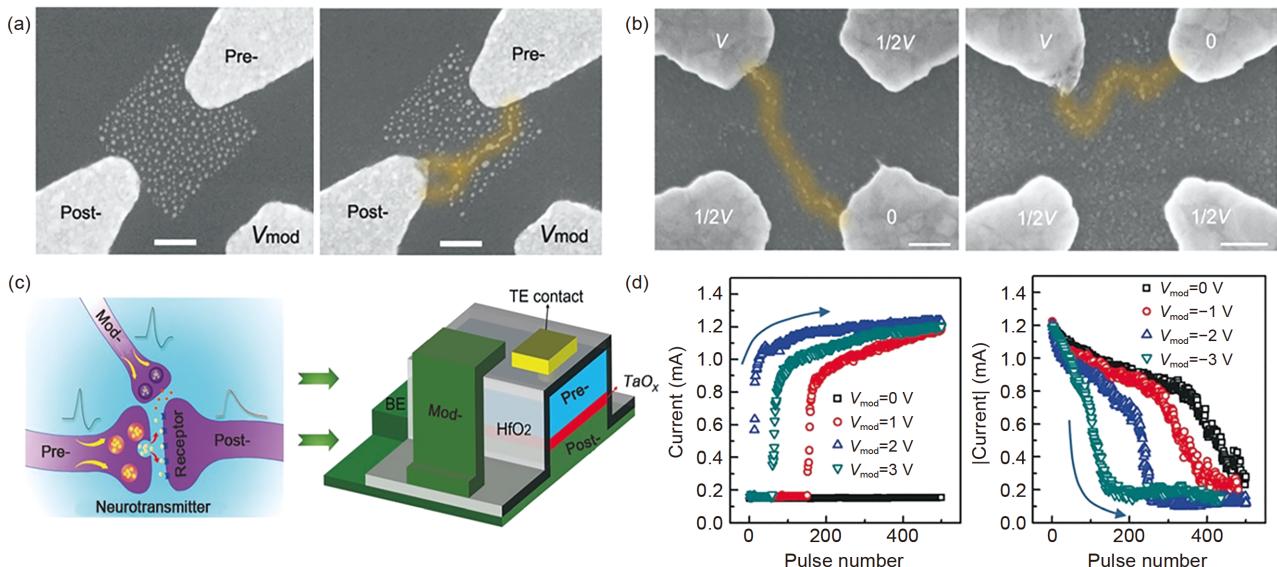


图 4 (网络版彩色)异源突触可塑性器件。 (a), (b) 通过调节端口电压可以实现银纳米团簇的自组织<sup>[53]</sup>; (c) 用于模拟异源突触可塑性的基于 $\text{HfO}_x$ 和 $\text{TaO}_x$ 的三端器件<sup>[54]</sup>; (d) 通过调制三端器件栅极的电压，可以实现不同速率的LTP与LTD过程<sup>[54]</sup>，Copyright © John Wiley and Sons  
**Figure 4** (Color online) Heterosynaptic plasticity device. (a), (b) Self-organization of silver nanoclusters realized by adjusting the port voltages<sup>[53]</sup>; (c) three terminal device based on  $\text{HfO}_x$  and  $\text{TaO}_x$  used to simulate heterosynaptic plasticity<sup>[54]</sup>; (d) heterosynaptic plasticity can be realized by three terminal device, where different I-V characteristic curves corresponding to different gate voltages<sup>[54]</sup>, Copyright © John Wiley and Sons

首次报道了仅使用两个莫特忆阻器和电容、电阻搭建的Neuristor电路，模拟了Hodgkin-Huxley(HH)神经元模型中的钠离子和钾离子通道，该电路不含任何晶体管，在保证性能的情况下极大地提高神经元的集成度，并显著降低了功耗。2016年，IBM苏黎世研究中心<sup>[61]</sup>利用PCM非晶态到晶态的转变过程模拟神经元膜电位的整合发放过程，并利用材料回到非晶态过程中的原子结构重构，实现了生物神经元的固有随机性，这种随机神经元可以检测输入信号的时间相关性，捕获脉冲信号的关键特征。2017年，惠普实验室的研究团队在基于NbO<sub>2</sub>的莫特器件研究中把器件与电容并联构造出LIF(leaky integrate-and-fire)神经元模型，并将温度与电荷作为独立变量，利用莫特相变的负微分电阻特性从理论上解释了这种神经元处于持续的混沌振荡中，它能处理反馈结构下的神经网络信号，具有从硬件层面上显著改善神经网络算法的能力。例如，在Hopfield网络中利用器件混沌特性能进行全局的搜索优化<sup>[19]</sup>。2018年，美国麻省大学研究人员<sup>[62]</sup>利用银纳米团簇的Pt/SiO<sub>x</sub>N<sub>y</sub>:Ag/Pt扩散忆阻器与电容并联，实现了简化后的LIF神经元模型，并将其与忆阻器突触阵列集成，首次搭建了完全由忆阻器和电容组成的双层卷积神经网络，可用于无监督学习下的数据分类；美国佐治亚理工学院研究人员<sup>[63]</sup>开创性地提出并演示了使用经典晶体管与PZT(lead zirconate titanate)铁电场效应晶体管构建人工神经元，利用栅极铁电材料的累积极化翻转效应可实现整合与发放功能，并通过仿真验证了它在脉冲神经网络无监督学习下的计算能力与超低功耗；美国HRL实验室研究人员<sup>[64]</sup>巧妙地利用基于VO<sub>2</sub>的Mott器件的负微分电阻特性，构造出了LIF神经元的钠钾离子通道，使得仅仅调控电学参数就可以实现23种神经元的特征发放模式，论证了忆阻器件构筑生物神经元的强大优越性(图5)。2019年，北京大学研究团队<sup>[65]</sup>展示了由基于IGZO(indium gallium zinc oxide)的肖特基二极管构成的具有随机动力学性质的人工神经元，该器件可以基于随机漫步算法进行全局极小值计算，并具有很好的鲁棒性；此外，北京大学研究团队<sup>[66]</sup>开发了基于部分结晶Hf<sub>0.5</sub>Zr<sub>0.5</sub>O<sub>2</sub>的新型铁电神经元，它具有基本的泄漏整合和发放功能，硬件成本仅需一个晶体管与电阻器，并可进一步在电路层面拓展到混合L-FeFET-CMOS神经元，实现复杂的神经元功能，如尖峰频率适应性。最近，北京大学研究团队<sup>[67]</sup>在该项研究的基础上进一步拓展了铁电神经元的计算能力，利用铁电材料

的定向极化特性来模拟神经元的兴奋性与抑制性输入，进而使基于铁电神经元的脉冲网络实现自组织映射学习的聚类算法与赢者通吃的推理算法。但是，目前神经元器件研究中还缺乏输出可以用于驱动下一级的电路的人工神经元，纯忆阻型神经网络的实际搭建仍面临器件与工艺方面的困难。如果搭建起基于忆阻器件的突触和神经元联合的脉冲式神经网络，直接利用器件关于时间的动力学进行计算，将带来与向量矩阵乘法完全不同的全新计算加速原理，并从物理层面上真正模拟生物大脑中的信息处理方式。

#### 4 未来发展趋势与展望

神经形态器件作为研究热点已经有十余年的历史，并取得了丰硕的成果。迄今为止，离子迁移型和材料相变型器件在科研与产业上已日趋成熟。但是，这些器件仍然存在局限性：离子迁移型器件离子输运与导电细丝生长过程的本征随机性与多样性带来了器件特性的涨落，相变型器件的不对称转变机制与电导漂移限制其对高精度模拟计算的处理。此外，大规模阵列集成中仍然面临线阻和寄生电流等问题，需要进一步的深入研究来改善并解决这些问题。同时，磁性器件、电子迁移型器件和离子栅控晶体管等新型神经形态器件也需要进一步发展，以满足可扩展性、高稳定性等要求。

在神经形态计算的应用层面上，短期来看，神经形态器件将主要用于实现人工神经网络的专用加速器，加速计算最密集的向量矩阵乘法。相较于CPU(central processing unit)，它采取模拟信号的大规模并行计算，能在提高容错率的同时提高运算速度；而相较于GPU(graphics processing unit)，它将带来能耗上的大幅度降低，这种超低功耗的特性非常适合边缘计算，将在物联网(Internet of things, IoT)时代发挥重要作用；在体系结构层次，由神经形态器件制作的存储器将兼具RAM和Flash存储器的优势：非易失、擦写速度高、断电后仍能保存，并且存内计算的方式将会带来芯片设计架构的重新布局。长期来看，模拟实验神经科学所揭示的生物中存在的突触离子动力学，是进一步发展人工突触器件的一个巨大挑战，这是硬件实现复杂人脑系统的一个具有极大潜力的途径，研究过程中也将促进我们对大脑自身的理解。然而，当前在硅基芯片上仅仅忠实地模拟观测到的生物神经系统以达到器件层面的仿生，似乎不是一个最优的选择，更需要计算神经科学方面的研究，针对神经形态器件的简化计算模型提出数学

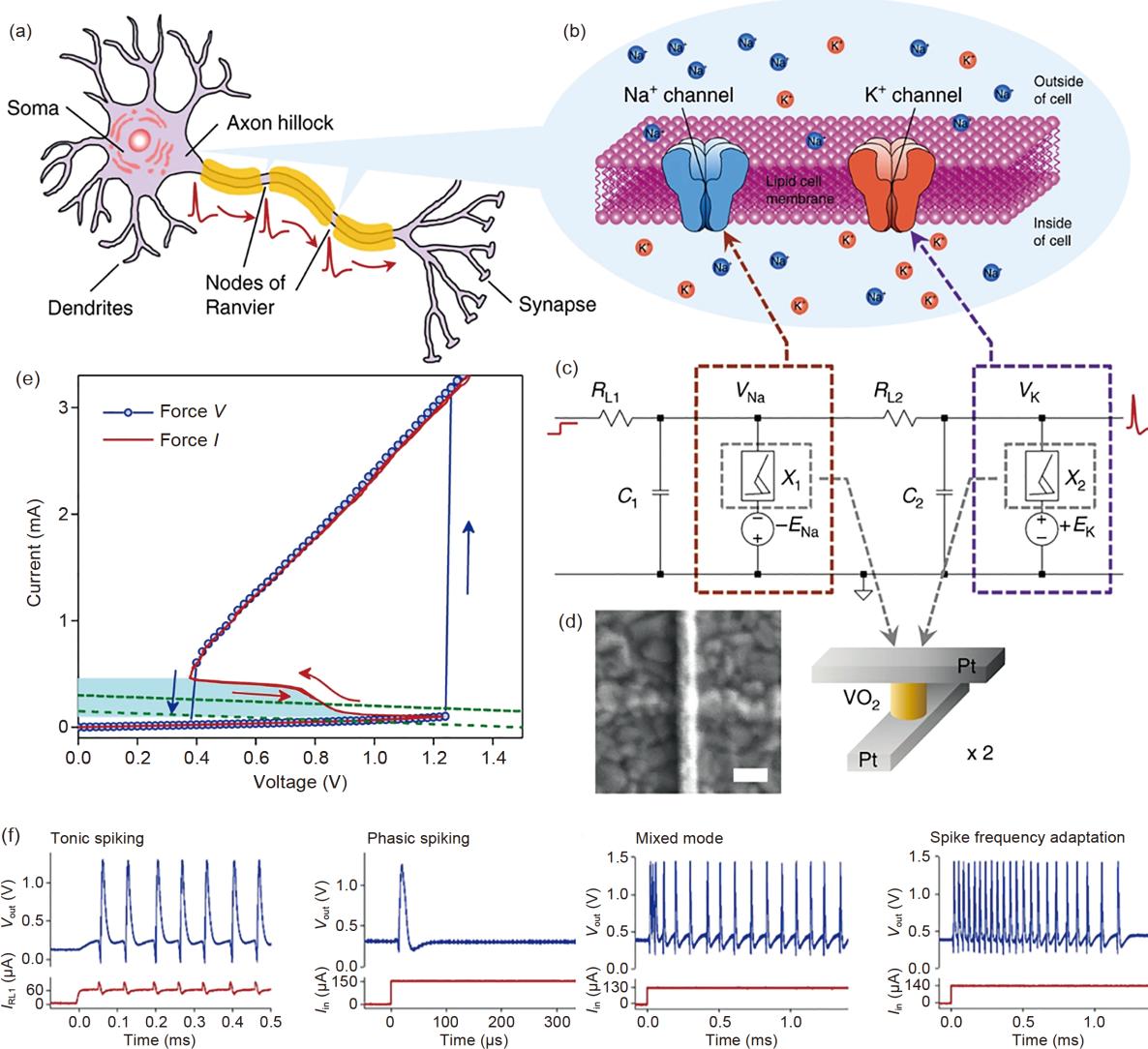


图 5 (网络版彩色)基于忆阻器的人工神经元器件<sup>[64]</sup>. (a) 生物神经元的结构示意图; (b) 神经元细胞膜中的 $\text{Na}^+$ ,  $\text{K}^+$ 门控; (c) Mott忆阻器组成LIF神经元离子通道的电路示意图; (d) Mott忆阻器的结构示意图与扫描隧道显微镜下的表征; (e) Mott忆阻器的I-V特性曲线, 呈现负微分电阻效应; (f) 基于忆阻器的神经元的部分仿生电脉冲发放模式. Copyright © Springer Nature

**Figure 5** (Color online) Artificial neuron device based on memristor<sup>[64]</sup>. (a) Structure diagram of biological neuron; (b)  $\text{Na}^+$ ,  $\text{K}^+$  ion channel in neuron cell membrane; (c) circuit diagram of LIF neuron based on Mott memristor; (d) structure diagram of Mott memristor and characterization under scanning tunneling microscope; (e) I-V characteristic curve of Mott memristor, showing negative differential resistance effect; (f) some kinds of bionic electric pulse patterns of neurons based on memristor. Copyright © Springer Nature

指导并赋予其网络意义。例如，经典计算机模拟脉冲神经网络的成本巨大且难以训练，而神经形态器件具备独特的动态特性和可重构特性，它不仅能在突触器件层面上直接实现多样化的、与时间相关的突触可塑性，还能天然地实现人工神经网络与脉冲神经网络的兼容，可在当前的深度学习与计算神经科学中各取所长，根据任务信息的流动在不同工作状态之间切换算法。进一步地，一旦算法上能抽象出脑启发式脉冲网络与对

应的训练算法，将有助于基于神经形态器件构建出生物启发的下一代神经网络标准硬件，软件和硬件的相辅相成将开创下一个人工智能的时代。

尽管前景光明，但神经形态器件的研究仍存在若干问题有待解决：如何设计和选择神经形态器件的材料、机理以制备出高精度的器件？材料的选择将直接影响神经形态器件的性能，需要筛选出能够支撑高性能神经形态器件的材料。机理的澄清有利于对神经形

态器件进行针对性的优化和设计。我们期待在材料科学甚至在物理机理水平上的突破，能大大提高神经形态器件的性能。最终，在不同的应用方面针对性地设计

神经形态器件将是未来信息产业发展的重要趋势。软件与硬件的真正融合将为实现神经形态计算乃至类脑智能奠定基础。

## 参考文献

- 1 Mead C. Neuromorphic electronic systems. *Proc IEEE*, 1990, 78: 1629–1636
- 2 Schuman C D, Potok T E, Patton R M, et al. A survey of neuromorphic computing and neural networks in hardware. 2017, arXiv: 1705.06963.
- 3 Markram H. The blue brain project. *Nat Rev Neurosci*, 2006, 7: 153–160
- 4 Merolla P A, Arthur J V, Alvarez-Icaza R, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 2014, 345: 668–673
- 5 Benjamin B V, Gao P, McQuinn E, et al. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proc IEEE*, 2014, 102: 699–716
- 6 Davies M, Srinivasa N, Lin T H, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 2018, 38: 82–99
- 7 Jo S H, Chang T, Ebong I, et al. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett*, 2010, 10: 1297–1301
- 8 Yang J J, Strukov D B, Stewart D R. Memristive devices for computing. *Nat Nanotechnol*, 2012, 8: 13–24
- 9 Prezioso M, Merrikh-Bayat F, Hoskins B D, et al. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature*, 2015, 521: 61–64
- 10 Pi S, Li C, Jiang H, et al. Memristor crossbar arrays with 6-nm half-pitch and 2-nm critical dimension. *Nat Nanotechnol*, 2018, 14: 35–39
- 11 Yang Y, Gao P, Gaba S, et al. Observation of conducting filament growth in nanoscale resistive memories. *Nat Commun*, 2012, 3: 732
- 12 Ambrogio S, Balatti S, Cubeta A, et al. Statistical fluctuations in  $HfO_x$  resistive-switching memory: Part II—Random telegraph noise. *IEEE Trans Electron Devices*, 2014, 61: 2920–2927
- 13 Yang Y, Huang R. Probing memristive switching in nanoionic devices. *Nat Electron*, 2018, 1: 274–287
- 14 Wang Z, Yin M, Zhang T, et al. Engineering incremental resistive switching in  $TaO_x$  based memristors for brain-inspired computing. *Nanoscale*, 2016, 8: 14015–14022
- 15 Zidan M A, Jeong Y J, Lu W D. Temporal learning using second-order memristors. *IEEE Trans Nanotechnol*, 2017, 16: 721–723
- 16 Sebastian A, Le Gallo M, Burr G W, et al. Tutorial: Brain-inspired computing using phase-change memory devices. *J Appl Phys*, 2018, 124: 111101
- 17 Cha E, Park J, Woo J, et al. Comprehensive scaling study of  $NbO_2$  insulator-metal-transition selector for cross point array application. *Appl Phys Lett*, 2016, 108: 153502
- 18 Kumar S, Wang Z, Davila N, et al. Physical origins of current and temperature controlled negative differential resistances in  $NbO_2$ . *Nat Commun*, 2017, 8: 658
- 19 Kumar S, Strachan J P, Williams R S. Chaotic dynamics in nanoscale  $NbO_2$  Mott memristors for analogue computing. *Nature*, 2017, 548: 318–321
- 20 Chua L. Memristor—The missing circuit element. *IEEE Trans Circuit Theor*, 1971, 18: 507–519
- 21 Strukov D B, Snider G S, Stewart D R, et al. The missing memristor found. *Nature*, 2008, 453: 80–83
- 22 Yang Y, Sheridan P, Lu W. Complementary resistive switching in tantalum oxide-based resistive memory devices. *Appl Phys Lett*, 2012, 100: 203112
- 23 Jeong Y J, Lee J, Moon J, et al. K-means data clustering with memristor networks. *Nano Lett*, 2018, 18: 4447–4453
- 24 Jeong Y J, Zidan M A, Lu W D. Parasitic effect analysis in memristor-array-based neuromorphic systems. *IEEE Trans Nanotechnol*, 2017, 17: 184–193
- 25 Li C, Han L, Jiang H, et al. Three-dimensional crossbar arrays of self-rectifying Si/SiO<sub>2</sub>/Si memristors. *Nat Commun*, 2017, 8: 15666
- 26 Hu M, Graves C E, Li C, et al. Memristor-based analog computation and neural network classification with a dot product engine. *Adv Mater*, 2018, 30: 1705914
- 27 Li C, Hu M, Li Y, et al. Analogue signal and image processing with large memristor crossbars. *Nat Electron*, 2018, 1: 52–59
- 28 Choi S, Shin J H, Lee J, et al. Experimental demonstration of feature extraction and dimensionality reduction using memristor networks. *Nano Lett*, 2017, 17: 3113–3118
- 29 Sheridan P M, Cai F, Du C, et al. Sparse coding with memristor networks. *Nat Nanotechnol*, 2017, 12: 785–789
- 30 Hu S G, Liu Y, Liu Z, et al. Associative memory realized by a reconfigurable memristive Hopfield neural network. *Nat Commun*, 2015, 6: 7522
- 31 Lin Y, Wu H, Gao B, et al. Demonstration of generative adversarial network by intrinsic random noises of analog RRAM devices. IEEE International Electron Devices Meeting (IEDM), San Francisco, 2018. 8614483
- 32 Li C, Belkin D, Li Y, et al. Efficient and self-adaptive *in-situ* learning in multilayer memristor neural networks. *Nat Commun*, 2018, 9: 2385
- 33 Li C, Wang Z, Rao M, et al. Long short-term memory networks in memristor crossbar arrays. *Nat Mach Intell*, 2019, 1: 49–57
- 34 Wang Z, Li C, Lin P, et al. *In situ* training of feed-forward and recurrent convolutional memristor networks. *Nat Mach Intell*, 2019, 1: 434–442

- 35 Wang Z, Li C, Song W, et al. Reinforcement learning with analogue memristor arrays. *Nat Electron*, 2019, 2: 115–124
- 36 Berdan R, Marukame T, Kabuyanagi S, et al. In-memory reinforcement learning with moderately-stochastic conductance switching of ferroelectric tunnel junctions. 2019 Symposium on VLSI Technology (VLSI), Kyoto, 2019. 8776500
- 37 Cai F, Correll J M, Lee S H, et al. A fully integrated reprogrammable memristor—CMOS system for efficient multiply-accumulate operations. *Nat Electron*, 2019, 2: 290–299
- 38 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proc IEEE*, 1998, 86: 2278–2324
- 39 Hopfield J J. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA*, 1982, 79: 2554–2558
- 40 Chua L. Memristor, Hodgkin-Huxley, and edge of chaos. *Nanotechnology*, 2013, 24: 383001
- 41 Bertschinger N, Natschläger T. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Comput*, 2004, 16: 1413–1436
- 42 Grossberg S. How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vis*, 1999, 12: 163–185
- 43 Maass W. Networks of spiking neurons: The third generation of neural network models. *Neural Netw*, 1997, 10: 1659–1671
- 44 Suri M, Bichler O, Querlioz D, et al. Phase change memory as synapse for ultra-dense neuromorphic systems: Application to complex visual pattern extraction. 2011 International Electron Devices Meeting (IEDM), Washington, 2011. 6131488
- 45 Kim S, Du C, Sheridan P, et al. Experimental demonstration of a second-order memristor and its ability to biorealistically implement synaptic plasticity. *Nano Lett*, 2015, 15: 2203–2211
- 46 Wang Z, Joshi S, Savel'ev S E, et al. Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing. *Nat Mater*, 2017, 16: 101–108
- 47 Cheng Z, Rios C, Pernice W H P, et al. On-chip photonic synapse. *Sci Adv*, 2017, 3: e1700160
- 48 Du C, Cai F, Zidan M A, et al. Reservoir computing using dynamic memristors for temporal information processing. *Nat Commun*, 2017, 8: 2204
- 49 Zhu J, Yang Y, Jia R, et al. Ion gated synaptic transistors based on 2D van der Waals crystals with tunable diffusive dynamics. *Adv Mater*, 2018, 30: 1800195
- 50 Zhu X, Li D, Liang X, et al. Ionic modulation and ionic coupling effects in MoS<sub>2</sub> devices for neuromorphic computing. *Nat Mater*, 2019, 18: 141–148
- 51 Moon J, Ma W, Shin J H, et al. Temporal data classification and forecasting using a memristor-based reservoir computing system. *Nat Electron*, 2019, 2: 480–487
- 52 Li J, Yang Y, Yin M, et al. Electrochemical and thermodynamic processes of metal nanoclusters enabled biorealistic synapses and leaky-integrate-and-fire neurons. *Mater Horizons*, 2020, 7: 71–81
- 53 Yang Y, Chen B, Lu W D. Memristive physically evolving networks enabling the emulation of heterosynaptic plasticity. *Adv Mater*, 2015, 27: 7720–7727
- 54 Yang Y, Yin M, Yu Z, et al. Multifunctional nanoionic devices enabling simultaneous heterosynaptic plasticity and efficient in-memory boolean logic. *Adv Electron Mater*, 2017, 3: 1700032
- 55 Bohnstingl T, Scherr F, Pehle C, et al. Neuromorphic hardware learns to learn. *Front Neurosci*, 2019, 13: 483
- 56 Sangwan V K, Lee H S, Bergeron H, et al. Multi-terminal memtransistors from polycrystalline monolayer molybdenum disulfide. *Nature*, 2018, 554: 500–504
- 57 Moore J J, Ravassard P M, Ho D, et al. Dynamics of cortical dendritic membrane potential and spikes in freely behaving rats. *Science*, 2017, 355: eaaj1497
- 58 Gerstner W, Lehmann M, Liakoni V, et al. Eligibility traces and plasticity on behavioral time scales: Experimental support of neohebbian three-factor learning rules. *Front Neural Circuits*, 2018, 12: 53
- 59 Bellec G, Scherr F, Hajek E, et al. Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. 2019, arXiv: 1901.09049.
- 60 Pickett M D, Medeiros-Ribeiro G, Williams R S. A scalable neuristor built with Mott memristors. *Nat Mater*, 2013, 12: 114–117
- 61 Tuma T, Pantazi A, Le Gallo M, et al. Stochastic phase-change neurons. *Nat Nanotechnol*, 2016, 11: 693–699
- 62 Wang Z, Joshi S, Savel'ev S, et al. Fully memristive neural networks for pattern classification with unsupervised learning. *Nat Electron*, 2018, 1: 137–145
- 63 Wang Z, Crafton B, Gomez J, et al. Experimental demonstration of ferroelectric spiking neurons for unsupervised clustering. 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, 2018. 13.3.1–13.3.4
- 64 Yi W, Tsang K K, Lam S K, et al. Biological plausibility and stochasticity in scalable VO<sub>2</sub> active memristor neurons. *Nat Commun*, 2018, 9: 4661
- 65 Dang B, Liu K, Zhu J, et al. Stochastic neuron based on IGZO Schottky diodes for neuromorphic computing. *APL Mater*, 2019, 7: 071114
- 66 Chen C, Yang M, Liu S, et al. Bio-inspired neurons based on novel Leaky-FeFET with ultra-low hardware cost and advanced functionality for all-ferroelectric neural network. 2019 Symposium on VLSI Technology (VLSI), Kyoto, 2019. T136–T137
- 67 Luo J, Yu L, Liu T, et al. Capacitor-less stochastic leaky-FeFET neuron of both excitatory and inhibitory connections for SNN with reduced hardware cost. 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, 2019. 6.4.2–6.4.4

Summary for “神经形态器件研究进展与未来趋势”

## Progresses and outlook in neuromorphic devices

Yanghao Wang<sup>1†</sup>, Chang Liu<sup>1†</sup>, Ru Huang<sup>1,2\*</sup> & Yuchao Yang<sup>1,2\*</sup>

<sup>1</sup> Department of Micro/nanoelectronics, School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China;

<sup>2</sup> Academy for Artificial Intelligence, Peking University, Beijing 100871, China

† These authors contributed equally to this work

\* Corresponding authors, E-mail: [rughuang@pku.edu.cn](mailto:rughuang@pku.edu.cn); [yuchao.yang@pku.edu.cn](mailto:yuchao.yang@pku.edu.cn)

In this information era with bursting data, there are three fundamental problems limiting the further development of computing power: The gradually slowing down of Moore's law, the rapidly increasing energy consumption while scaling down, and the restriction of data transfer between separated memory and processor known as "Von Neumann bottleneck". To deal with massive data and make all things in our daily life interconnected, it is imperative to develop new generation of computing paradigms, for example, neuromorphic computing. Inspired by the human brain, neuromorphic computing has drawn extensive attention in recent years due to its high parallelism, low power consumption and in-memory computation, especially when the tide of artificial intelligence is sweeping across the globe now. Novel neuromorphic devices are key to the construction of neuromorphic computing systems, providing an efficient implementation of artificial neural network on chip. Besides, by simulating the behavior of biological neurons and synapses at the physical level, neuromorphic devices can enable a brand new computing method, which is thought to be an essential and promising way to build a brain-like system on chip. This paper focuses on the current research progresses and future research trends of neuromorphic devices. It summarizes the physical mechanisms of neuromorphic devices, according to which they can be divided into ion-migration device, material phase change device, electron-migration device, magnetic device and so on. It further details the inner dynamics happened in a single device by characterizing the changing process in transmission electron microscope. This paper also explains how these neuromorphic devices can be used for computing, where memristor, as the missing fourth basic circuit component, is used as a theoretical support. Through making up a crossbar structure, these devices can directly calculate vector-matrix multiplication by Ohm's law and Kirchhoff's law. It is an elegant way that can accelerate artificial neural network in parallel by computing in memory. In addition, exploiting the intrinsic dynamics in the devices can realize complex and interesting functions of biological neural networks, such as long-term plasticity, short-term plasticity and spike timing dependent plasticity. For example, a synaptic transistor based on two-dimensional materials WSe<sub>2</sub> can simulate the biologically transport process of calcium ion, leading precisely regulated coexistence of long-term and short-term memory. Remarkably, a significant advance is realizing heterosynaptic plasticity, a general mode in brain's cortical network, in a single multi-terminal device, which is crucial for biologically plausible supervised learning on hardware. Furthermore, artificial neurons composed of memristors and capacitors have some natural advantages compared to traditional circuits, containing rich dynamics for neuromorphic computing, such as stochasticity, adaptive threshold and chaotic oscillation. Recently, materials with new mechanisms, like ferroelectric materials, are also being explored to realize complex neuron functions, contributing to less power and smaller area. By devices and algorithms co-design, it is time to explore the next generation of neural network, transforming current processing unit to a more efficient and intelligent brain-inspired style. As a conclusion, the outstanding challenges and trends in the field of neuromorphic devices are discussed in this paper. These researches are promising for building a neuromorphic computer in the future, which will be complementary to classical computer and outperform in many tasks.

**neuromorphic devices, brain-inspired computing, artificial intelligence, artificial synapse, artificial neuron, memristor**

doi: [10.1360/TB-2019-0739](https://doi.org/10.1360/TB-2019-0739)