



北京大學

软件与微电子学院
集成电路与智能系统系

《集成电路前沿技术导论》

课程报告

报告题目： 神经形态器件研究报告

姓 名： 胡成成

学 号： 2101210578

日 期： 2021 年 11 月 01 日

神经形态器件研究报告

摘要

随着人工智能等领域的迅速发展，对高效计算、低功耗、智能化的芯片需求越来越多，在芯片器件的设计上的要求更高。而神经形态器件主导“类脑计算”突破了冯诺依曼计算范式的限制。本文针对神经形态器件的技术背景和意义展开，分析其研究现状和在某些器件上面临的挑战。并调研了基于 ReRAM、MRAM、PCRAM、FeRAM 和 Flash 浮栅型晶体管的设计思路，并对这些方法进行思考，得出提升神经形态器件集成度的方案需要从化学材料和器件架构设计上改进的结论。

引言

随着物联网、云计算、大数据、人工智能以及 5G 等新兴技术应用需求的快速增加，微电子产业在国民经济和国防建设中的战略地位与各方面需求日益凸显。众所周知，自集成电路发明至今，其工艺技术一直遵循着摩尔定律不断进步。如今，步入后摩尔时代，集成电路的工艺上要求新的器件成果来支撑，包括神经形态器件、Beyond CMOS 器件、面向全方位智能感知的传感器件、以及功率与射频器件。其中，神经形态器件随着人工智能的快速发展展现出强大的生命力。

研究背景

人工智能的迅猛发展渐渐崭露出对专用的硬件的需求，人工神经网络已经展现出长足的优势，在医疗诊断、环境科学、智能机器、物联网和安全等多个领域产生了深远的影响。但是，现有的算法实现都基于传统的冯诺依曼计算架构来设计与实现，这导致现有的计算面临着低智能、高能耗、低容错等问题，无法再满足智能大数据应用场景快、准、智的响应需求。于是，“类脑计算”应运而生。

“类脑计算”借鉴大脑的体系结构和信息处理方式，基于神经形态器件构建逻辑与存储相融合的计算硬件，旨在实现更为灵活和智能的信息处理与计算模式。类脑神经网络主要是模拟人脑突触、神经元以及连接方式来实现神经形态计算，神经形态器件便是在这样的神经元模型和神经信号处理的基础上进行设计的，因

此,针对模拟生物基本信息处理单元功能的目标所研发的器件统称为神经形态器件,可分为突触器件、神经元器件两大类。通过借鉴人脑神经网络的结构可知,构建神经网络硬件需要数量庞大的突触器件,因此具有记忆和学习功能的“人造突触”器件是类脑硬件研究中的关键问题,并且神经形态突触器件应具备尺寸方面的优势,国内外研究团队在通过忆阻器、磁隧道结 (Magnetoresistive Tunnel Junction, MTJ)、相变存储器 (Phase Change Memory, PCM)、CMOS 晶体管、双电层晶体管等器件实现突触的功能方面已经取得了重要进展。

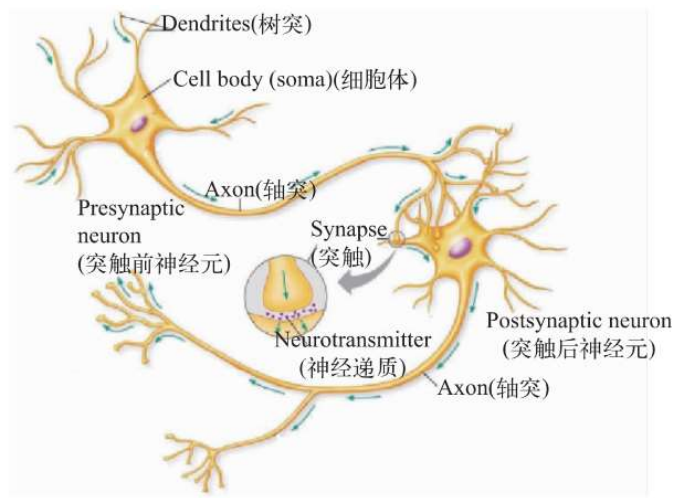


图 1 生物大脑神经网络基本组织结构 (图源自文献[5])

研究意义

类脑计算是目前人工智能研究领域的重点,受人脑存算一体、并行处理启发而构建的包含突触与神经元的神经形态计算架构,可以有效地降低人工智能中计算工作的能耗,记忆元件在神经形态计算的硬件实现中展现出巨大的应用价值。目前有关类脑神经网络的热点课题一般从两个不同领域开展相关研究。首先人们从模拟生物神经网络系统及结构出发,提出了以脉冲神经网络(Spiking Neural Network, SNN)模型为主的研究方案,涉及拟态神经元及其突触的器件兑现方案与网络结构;其次,大部分有关人工智能的神经网络算法是基于冯·诺依曼结构的计算机编程实现的,但随着大数据信息处理复杂性不断增加,数据存储效率和计算处理能耗已经达到极限,人们开始意识到需要建立起基于存储—计算一体化类脑的信息处理方式,以实现成熟智能算法的神经网络。

因此,神经形态器件的研究意义主要体现在:(1)降低传统计算的能耗,贯彻可持续发展理念;(2)提高计算与存储的信息处理效率,增加计算的容错能力;

(3) 提高智能处理信息的能力，推动智能化的发展。

神经形态器件研究现状与技术挑战

研究现状

人工智能正引领时代的变革，人工神经形态技术在图像、语音识别等领域取得突破性进展，但距离人脑智能或者具有自我意识的强人工智能还存在很大的差距。同时，目前对大脑的研究尚未完成，但类脑神经网络的实现可从模拟神经形态网络的结构和功能入手，让类脑神经网络具备自主学习能力、获取新信息和新技能的能力，以及进行交互、推理从而做出决策的能力，同时具有高并行性、高容错性、高连接性、存算融合和低功耗特性。我们知道，模拟生物基本信息处理单元功能的目标所研发的器件统称为神经形态器件，可分为突触器件、神经元器件两大类。目前多种模拟突触功能的 VLSI (Very Large Scale Integration) 电路被提出，神经网络向深度发展，突触器件的尺寸也越来越小。大量新型纳米尺度器件被研究与开发出来，这些器件可以近似模拟生物突触功能，在这些新型的基于纳米尺度的突触模拟器件中，忆阻器获得了广泛的研究。

忆阻器是一种非易失性记忆元件，其状态/电阻可以通过施加足够强的电压脉冲而改变，是脉冲神经网络权重元件中最合适的候选器件。另一方面，忆阻器作为第四个基本电路元件被提出，它具有阻值连续可调、断电非易失、低功耗等特性，是模拟突触的最佳选择。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。

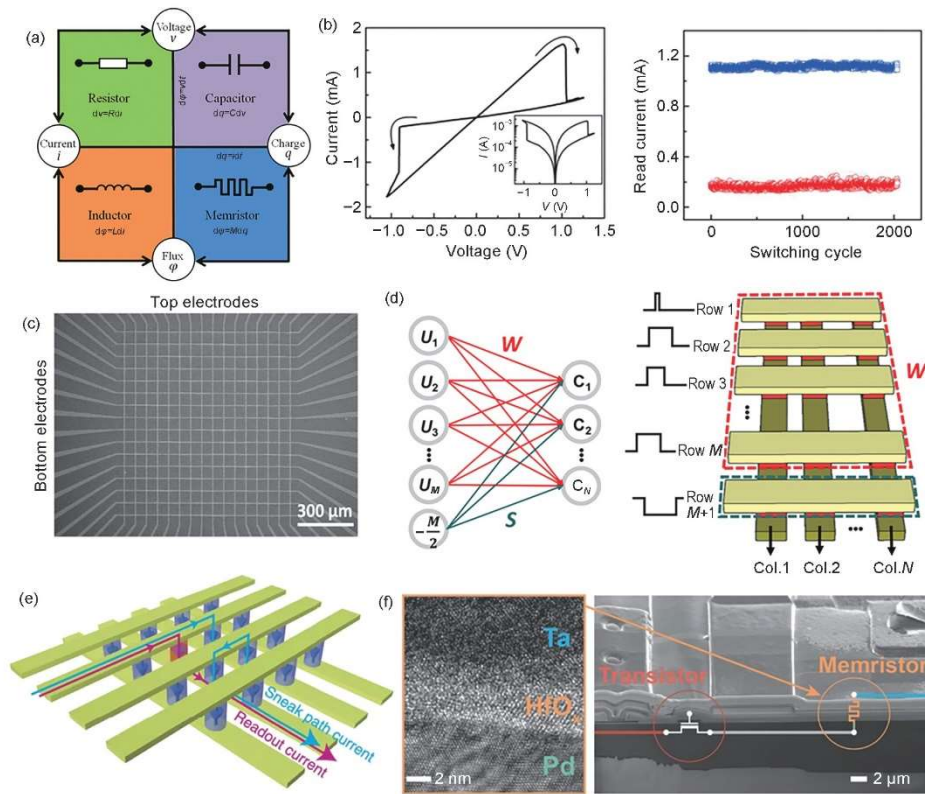


图 2 忆阻器的计算原理. (a) 忆阻器作为第 4 个基本电路元器件 (b) 忆阻器的典型 I-V 特性图, 在不同扫描次数下高低阻态的稳定性 (c) 扫描电子显微镜下基于 TaOx 的忆阻器交叉阵列 (d) 人工神经网络映射到阵列计算的示意图[23] (e) 潜行电流的影响 (f) 隧道电子显微镜下 1T1R 的单元结构 (图源自文献[4])

相对忆阻器的设计,神经形态器件的另一个重要意义在于模拟生物突触可塑性,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)。

除此之外,有学者提出人工神经形态器件的发展主要包括 2 条技术路线:一种是基于传统成熟的 CMOS 技术的 SRAM 或 DRAM 构建,其原型器件在信息存储方面属于易失型;另一种是基于非易失性 FLASH 器件或新型存储器件、新材料构建.从传统 CMOS 技术角度来看,各大公司和高校的实验室研发类脑芯片,2019 年,清华大学研制成功了全球首款异构融合类天机芯(Tianjic)类脑芯

片；同年 8 月，浙江大学基于 55 nm 工艺技术研制出了达尔文类脑芯片等等。这类芯片虽然区别于传统集成电路和人工神经网络加速 AI 芯片，但芯片中基本人工神经元和突触依然是采用数字电路或者数模混合电路搭建，存在集成度、功耗和功能模拟准确度等瓶颈技术问题。从非易失性神经形态器件的角度来看，这种非易失性神经形态器件就是一种具有人工神经形态特性的忆阻器（Memristor），前面我们已经提到过，其具有独特的非线性特性，并已成为新型模拟生物神经元与突触的基本信息处理单元，得到了人们的广泛研究。目前新型非易失性神经形态器件的研究主要集中在机理和功能模拟两个方面。现阶段国内外在这些新型器件实现了模拟突触和神经元的功能并取得了重要进展，但尚处在探索研究阶段。

技术挑战

神经形态器件研究在各种器件中表现出不同的挑战与难点，下面以常见的两种神经形态器件——忆阻器和忆容器面临的挑战进行分析。

首先，对于模拟突触的关键性器件忆阻器这几年发展迅速，但是商用化进程还有一定的距离，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 就需要约 34 万人工突触和 1 万人工神经元。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。

其次，实现复杂人脑系统需要我们加深对大脑自身的理解，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。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。

最后，忆容器作为新型非线性无源器件，是与忆阻器等共同构建超低能耗神经网络的理想器件。对于基于整体等效介电常数变化的忆容器而言，器件尺寸的缩减会伴随着器件漏电流增大、击穿电压降低的现象；当忆容介质层的厚度减小到一定程度时会导致器件电导激增，电容值下降。在大规模集成方面，这是忆容器与忆阻器、CMOS 晶体管等器件相比的不足之处。如何获得同时具有微纳尺寸

与高绝缘性的忆容器仍然是未来研究的难点与重点之一。

解决方案与思路

针对上文中提出的挑战，这里针对在保证单器件性能的情况下，提高忆阻器件的集成度这一挑战进行解决方案调研与思考。

基于 ReRAM 的人工神经形态器件研究

ReRAM 是由金属 / 阻变材料薄膜 / 金属构成，利用阻变材料电阻在外加电场激励作用下具有高阻态（High Resistive State, HRS）与低阻态（Low Resistive State, LRS）之间可逆改变特性实现对数据信息的存储功能。该器件具有结构简单、非易失、功耗低、速度快、多阻态、尺寸小、易三维集成等诸多优点，同时，由于 ReRAM 具有模拟生物神经突触功能，已被广泛应用于硬件神经网络系统的研究。

基于 MRAM 的人工神经形态器件研究

1975 年，Julliere 通过 Co/Ge/Fe 磁性隧道结（Magnetic Tunnel Junctions, MTJ）发现了隧穿磁电阻（Tunneling Magnetoresistance-TMR）效应，其机理是利用电流诱导的自旋转移力矩（Spin-Transfer Torque, STT）驱动磁矩变化，从而产生隧穿磁电阻变化的响应，可以进行信号输出与功能表达。利用电子的磁性和电学性质的自旋电子纳米器件，可以提高能量效率并缩小神经形态电路的面积，而磁性隧道结作为神经形态计算元件特别受关注，因为它们与标准集成电路兼容，可以支持多种功能。

基于 PCRAM 的人工神经形态器件研究

相变存储器（Phase Change Random Access Memory, PCRAM）是利用相变材料作为存储介质（如 Sb-Te、Ge-Te、Ge-Sb-Te 等硫系材料），在电流的焦耳热作用下，实现相变材料在结晶相态（即低阻态）和非晶相态（即高阻态）之间快速可逆转换，利用这一电阻率随晶态转化的特性来实现信息存储的 NVM 技术。PCRAM 因具有高集成度、高响应速度，也被广泛应用于高密度忆阻神经-突触阵列，构建神经形态芯片。

基于 FeRAM 的人工神经形态器件研究

FeRAM 是一种非易失存储器，具有高速、高密度、低功耗和抗辐射等优点，目前主要的铁电材料有钙钛矿结构系列，包括 $\text{PbZr}_{1-x}\text{Ti}_x\text{O}_3$ 、 $\text{SrBi}_2\text{Ti}_2\text{O}_9$ 和 $\text{Bi}_{4-x}\text{La}_x\text{Ti}_3\text{O}_{12}$ 等，而传统的 FeRAM 存在尺寸收缩和集成度问题，经过科学家不断地探索，发现铁电隧道结（FTJ）具有尺寸小、易于集成、灵敏度高、使用寿命长等优点，可以很好地解决传统的 FeRAM 瓶颈技术问题。

基于 Flash 浮栅型晶体管的人工神经形态器件研究

NOR Flash 浮栅器件具有工艺成熟度高、集成度高、成本低等优点，从非易失性、高密度和可实现多位编程特性等方面看，Flash 器件也非常适合深度学习和人工智能应用，尤其在端侧 AIoT 领域。

针对解决方案的思考

可以看出，针对提高忆阻器件的集成度的问题上，研究者们大多从化学材料的角度出发，从减小器件尺寸，增大密度的方式来提高集成度，每种工艺的研究都有他的优势和潜在的问题所在。

针对神经形态器件，我想在提高器件的集成的角度来看，除了从材料的角度来看，是否可以从模型设计的角度来解决这些问题，让器件的功能更加集成，让每个单独的器件承载更多的功能，而不是靠堆砌器件来集成。甚至从类脑的角度来看，神经元-突触式的建模是否可以稍微改动，引入“中枢”的概念来建模等等。参阅一些文献后，发现一些对自选电子器件、高密度 TaO_x 忆阻器设计，其中在外文文献上能够找到有关架构的设计，例如通过电解质门控实现全局连接性的神经形态器件设备等等。

可以看出，针对神经形态器件的解决方案的主要有从化学材料和架构设计角度来设计解决现有的集成问题。同时，随着人工智能场景的不断扩大，对底层集成电路的性能要求也越来越多样化，使得研究者在思考器件设计上有了更多元的方式。

总结与展望

本文针对神经形态器件的发展和现状做了简单的调研学习。如今，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，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。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。

同时,类脑计算机与类脑智能的实现毫无疑问将成为科学的巅峰。尽管当前对于大脑功能与机制的有限认知给“类脑计算”研究提出了很大的挑战,但也为研究者们提供了发挥创造力和探索精神的广阔天地。如今我们对于大脑的奥秘还知之甚少,但是随着神经科学家、认知科学家对于大脑机制研究的进一步深入,通过计算机科学、智能科学、材料科学、微电子学等研究领域的学者开展更加密切的交流与合作,“类脑计算”的实现将不再是梦想。

当然,在目前 ReRAM、PCRAM、MeRAM 等工艺、器件尚处于探索性研究与应用阶段,存在基准工艺尚未成熟、制作成本高等问题,如阻变机理不完善,阻变开关比,多阻态线性度、对称性、一致性与稳定性等,需要从新材料、器件结构、原理等方面进一步优化与提升。神经形态器件并有望成为大数据及智能化时代应对海量实时数据的颠覆性计算范式,对新型非冯·诺依曼架构芯片的研发乃至类脑智能时代的最终实现都具有重要意义。有理由相信随着类脑计算的发展,将会引领人工智能走向新高度。

参考文献

- [1]. 朱进宇,闫峥,苑乔,张少真.集成电路技术领域最新进展及新技术展望[J].微电子学,2020,50(02):219-226.
- [2]. 武俊齐,赖凡.后摩尔时代新兴计算芯片进展[J].微电子学,2020,50(03):384-388.
- [3]. 殷明慧,杨玉超,黄如.神经形态器件现状与未来[J].国防科技,2016,37(06):23-30.
- [4]. 王洋昊,刘昌,黄如,杨玉超.神经形态器件研究进展与未来趋势[J].科学通报,2020,65(10):904-915.
- [5]. 张玲,刘国柱,于宗光.人工神经形态器件发展现状与展望[J].电子与封装,2021,21(06):5-18.
- [6]. 邓亚彬,王志伟,赵晨晖,李琳,贺珊,李秋红,帅建伟,郭东辉.类脑神经网络与神经形态器件及其电路综述[J].计算机应用研究,2021,38(08):2241-2250.
- [7]. 任宽,张珂嘉,秦溪子,任焕鑫,朱守辉,杨峰,孙柏,赵勇,张勇.基于忆容器件的神经形态计算研究进展[J].物理学报,2021,70(07):21-38.

-
- [8]. 周振宇. 神经形态忆阻器特性及光调控研究[D].河北大学,2021.
- [9]. 张志成. 晶圆尺寸石墨炔/石墨烯异质结的制备及其在神经形态器件领域的应用[D].天津理工大学,2021.
- [10]. 蔡佳林. 基于自旋电子器件的神经形态器件研究[D].中国科学技术大学,2019.
- [11]. Gkoupidenis, P., Koutsouras, D. & Malliaras, G. Neuromorphic device architectures with global connectivity through electrolyte gating. *Nat Commun* 8, 15448 (2017).