



基于磁性隧道结的人工神经形态器件

王雨杰^{1,2}, 张黎可³, 罗仁涓^{1,2}, 鲁怡萌^{1,2}, 方彬^{1,2}, 曾中明^{1,2,4*}

1. 中国科学技术大学纳米科学与技术学院, 合肥 230026

2. 中国科学院苏州纳米技术与纳米仿生研究所纳米加工平台, 苏州 215123

3. 无锡学院电子信息工程学院, 无锡 214105

4. 江西省纳米技术研究院纳米器件及工艺研究部和南昌市先进封测重点实验室, 南昌 330200

* 联系人, E-mail: zmzeng2012@sinano.ac.cn

2024-08-31 收稿, 2025-01-15 修回, 2025-01-16 接受, 2025-01-17 网络版发表

国家自然科学基金(52371206, 12204357)、江西省双千计划(S2019CQKJ2638)和中国科学院“率先行动”引才计划资助

摘要 神经形态器件在构建未来人工智能芯片方面具有极大的应用潜力. 其中, 磁性隧道结作为自旋电子学的核心器件, 凭借其超长寿命、极低功耗、非易失性存储、高速运算以及多功能集成等特性, 在模拟人工神经元与突触功能、驱动神经网络高效计算等方面具有重要的应用前景. 然而, 虽然磁性隧道结神经形态器件的研究领域广泛而深入, 但是相关研究成果的系统梳理尚显不足. 本文通过概述人工神经形态器件的发展历程, 深入剖析了自旋忆阻器、自旋振荡器、自旋随机数发生器等不同功能的磁性隧道结器件在模拟人工神经突触与神经元方面的最新进展, 并对该领域面临的挑战与未来发展方向进行了展望.

关键词 磁性隧道结, 人工神经元, 人工突触, 自旋忆阻器, 自旋振荡器, 自旋随机数发生器

人工智能(artificial intelligence, AI)技术作为新一轮科技和产业革命的核心驱动力, 正在深刻地改变人类的生产生活方式, 实现社会生产力的不断提升. 在驱动人工智能技术发展的三驾马车——数据、算法和算力中, 算力作为底层硬件基础, 是人工智能的核心发展动力. 然而, 当前主流的以图形计算单元(graphics processing unit, GPU)为代表的人工智能芯片采用了传统的冯·诺依曼架构, 计算单元和存储单元分离, 存在着“存储墙”与“功耗墙”瓶颈, 在进行神经网络计算时, 处理单元和存储单元之间需要进行频繁的数据交互, 带来了算力提升困难和能量消耗大的问题^[1]. 相比之下, 神经形态芯片则是通过模拟人脑神经网络结构和功能构建的芯片^[2], 它的优势在于: (1) 存算一体架构, 即计算单元与存储单元一体化, 消除了冯·诺依曼架构中处理器-内存数据交互的时间和能量损耗; (2) 处理任务

时有极高的并行度. 因此, 通过开发具有类脑架构的神经形态芯片, 有望实现高性能、低功耗的认知智能计算. 神经形态芯片的基本单元是人工神经元和突触器件. 在生物系统中, 神经元的主要作用是对所有的输入信号进行处理, 当总信号超过一定阈值时, 神经元会被激活, 并将信号传递给下一级神经元^[3,4]. 对应地, 人工神经元器件便需要具有非线性的输入-输出特性. 突触的作用是连接上下级神经元, 通过突触权重的大小控制上下级神经元之间的连接强弱^[5]. 引申至硬件神经网络中, 则是利用器件具有的多电阻态等特性, 将权重信息存储在器件的物理态中.

目前, 适合于构建神经形态计算芯片的材料体系主要有相变材料^[6,7]、阻变材料^[8,9]、铁电材料^[10,11]、二维材料^[12]和自旋电子材料等^[13]. 相变材料主要是通过由焦耳热控制的晶体与非晶体之间的转变来模拟神

引用格式: 王雨杰, 张黎可, 罗仁涓, 等. 基于磁性隧道结的人工神经形态器件. 科学通报, 2025, 70: 1893–1907

Wang Y, Zhang L, Luo R, et al. Artificial neuromorphic devices based on magnetic tunnel junctions (in Chinese). Chin Sci Bull, 2025, 70: 1893–1907, doi: [10.1360/TB-2024-0931](https://doi.org/10.1360/TB-2024-0931)

经元的突触结构。阻变材料是通过电压控制的离子迁移实现材料电阻的变化来模拟突触。铁电材料的基本器件是由两片电极和中间的铁电材料构成的铁电隧道结,通过电场调控铁电极化来改变其阻值从而实现突触效果。二维材料通过电压调控材料晶界等特性来改变材料阻值。

与前四种材料相比,以磁性隧道结(magnetic tunnel junction, MTJ)为代表的自旋电子材料,利用电子的自旋特性来实现电阻可调,从而模拟神经元和突触特性,具有低功耗、高效率、高可靠性和明显的非线性动力学特征,是构建神经形态器件的优选方案之一^[14]。MTJ的核心结构为上、下铁磁层和中间的非磁性势垒层构成的三明治结构。其中一层铁磁层称之为参考层,其磁化方向固定为某一特定方向,另一层铁磁层为自由层,可以在外场(如:电场和磁场)作用下改变磁化方向。MTJ展示了丰富的物理效应和功能特性,包括隧穿磁电阻(tunnel magnetoresistance, TMR)效应^[15]、自旋转移矩(spin-transfer torque, STT)效应^[16]、自旋轨道矩(spin-orbit torque, SOT)效应^[17]、电压调控磁各向异性(voltage-controlled magnetic anisotropy, VCMA)^[18,19]以及随机翻转特性等。这些丰富的物理特性为构建神经形态芯片提供了物理基础。目前基于MTJ神经形态器件的研究按照器件种类可以大致分为三类:自旋忆阻器、自旋振荡器、自旋随机数发生器。

本文首先概述神经形态器件的发展,然后根据MTJ的神经形态器件的分类,详细介绍各种器件的结构、机理以及最新相关进展,最后分析相关神经形态器件发展过程中所遇到的问题和挑战,并展望一些潜在的解决方案。

1 神经形态芯片的发展与应用

神经形态芯片是一种从结构层面对人脑神经系统的连接方式和响应特性进行模拟,从而实现高性能、低功耗的神经形态计算的人工智能芯片。如图1所示^[20],在生物神经网络中,神经元以及神经突触作为最基本的单元结构组成多层神经网络,其中神经元之间依靠神经突触相互连接,实现电信号传递。神经突触的功能是存储突触权重信息,其需要突触权重可塑性。而神经元的功能则是接收来自连接突触的输入信号,并对这些信号进行响应和处理。基于这一基本过程,不同层级的神经网络分别对信息执行编码、过滤等处理过程,通过多层级协作实现各种高级智能功能。受人脑工

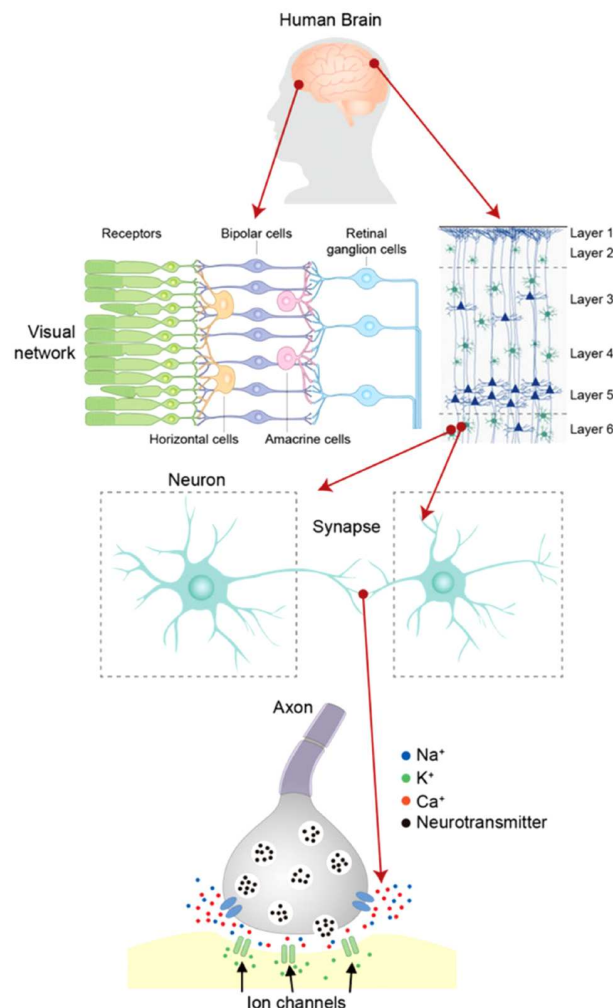


图1 (网络版彩色)人脑神经网络结构^[20]. Copyright©2019, WILEY-VCH Verlag GmbH & Co. KGaA

Figure 1 (Color online) The structure of the human brain neural network^[20]. Copyright©2019, WILEY-VCH Verlag GmbH & Co. KGaA

作机制的启发,人们期望通过开发具有类脑架构的神经形态芯片,以实现高性能、低功耗的仿生智能计算。

1.1 神经形态芯片的发展

神经形态计算芯片作为人工智能领域的前沿方向,代表着未来计算技术的发展方向,其相关的研究起源于20世纪80年代,加州理工学院的Mead教授^[21]最早提出了神经形态计算概念。1990年左右,来自斯坦福大学的Mahowald和Douglas^[22]首次提出了基于CMOS (complementary metal oxide semiconductor)电路的硅基神经元器件。2011年8月,IBM公司^[23]率先在神经形态芯片上取得进展,发布SyNAPSE芯片。这一芯片包含256个

神经元, 每个神经元对应1024个突触, 能够处理简单的游戏任务, 在一定程度上实现了对人类认知能力的模拟。之后随着半导体工艺制程的发展, 集成在芯片中的神经元和突触数量大大提高。2013年, IBM公司实现了具有万级神经元数和亿级突触数的神经形态芯片; 2021年, Intel^[24]发布了Loihi 2神经形态芯片, 使用更为先进的4 nm芯片制程, 神经元数量提升到了百万个。

国内神经形态芯片研究起步相对较晚, 但近年来逐渐取得一些重要进展。2019年, 清华大学施路平教授团队^[25]发布Tianjic神经形态芯片, 可以运行两种不同的神经网络算法, 并完成了自行车无人驾驶任务。2020年, 浙江大学发布的Darwin类脑芯片^[26], 其神经元数量已与小鼠相当, 并实现了训练机器人协同工作的任务。2022年, 清华大学施路平教授团队^[27]发布了TianjicX神经形态芯片, 实现了机器猫多模态识别避障并抓捕老鼠的任务。2023年, 中国科学院计算技术研究所尤海航等人^[28]研制出了超导神经形态芯片“SUSHI”, 获得每秒近1.4万亿次突触操作的峰值神经形态处理性能。2024年, 清华大学钱鹤教授团队^[29]发布了Tianmouc神经形态芯片, 通过借鉴人类视觉系统提出了互补双通路类脑视觉感知范式, 在同一芯片上同时实现了高分辨、高速、高动态特性, 突破了传统范式的瓶颈。从最初的理论探究到现在世界范围的广泛竞争, 神经形态计算芯片展现了巨大的发展潜力和速度, 大大推动了计算技术的不断进步提升。

1.2 神经形态器件的应用

上述的研究成果已经展现出神经形态芯片巨大的潜力, 但是随着任务复杂性的不断提升, 基于CMOS工艺的神经形态芯片的功耗和集成度逐渐显露瓶颈。为此, 人们开始研究基于新材料新物理的神经形态器件, 期望通过利用器件的本征物性直接模拟神经元和突触的工作特性。目前, 基于阻变材料体系的神经形态器件的研究较为成熟。2015年, 美国加州大学圣巴巴拉分校研究人员^[30]构建了一个包含100个突触的忆阻器神经网络, 实现了简单的人类视觉功能模拟。2017年, 清华大学钱鹤教授团队^[31]构建了可进行简单的人脸识别的128×8的忆阻器阵列。2018年, 美国麻省理工学院研究人员^[32]制备了第一个可用于模式识别的集成了神经元和突触的神经网络。2020年, 清华大学吴华强教授课题组^[33]提出了基于忆阻器阵列芯片卷积神经网络, 实现了全硬件构成的完整的存算一体系统, 高效准确地验

证了图像识别功能。2022年, 他们又提出了基于动态忆阻器的储层计算, 实现了对动态手势97.9%的识别准确率^[34]。2023年, 沙特阿卜杜拉国王科技大学朱凯辰等人^[35]提出了一种混合2D-CMOS微芯片的忆阻器, 并成功构建了脉冲神经网络的人工突触, 实现无监督学习功能。

神经形态器件的新材料体系仍在被继续探索。如铁电材料^[10,11,36,37]、相变材料^[6,7,38]、二维材料^[39,40]、光子器件^[41-44]以及自旋电子材料^[12,45]等纷纷被证明可构建具有神经形态特性的器件, 这些材料体系为神经形态器件的发展提供了机遇。

2 自旋神经形态特性研究

自旋电子器件通过调控与操纵电子自旋属性, 实现丰富的器件功能。在自旋电子器件中, 通过磁电阻效应、自旋霍尔效应等手段可以将器件中的电子自旋信息进行电信号读取。进一步利用力、热、电、光、磁等物理场, 可以对电子自旋进行调控, 从而引起器件输出电信号变化。利用自旋电子器件的自旋相关输运特性, 可以实现信息传感、信息存储^[46]、微波发射与微波探测^[47]等丰富的功能。利用多物理场驱动自旋电子器件产生的丰富的物理特性, 可以对突触和神经元特性进行模拟, 如利用微波探测功能的自旋振荡器阵列可以实现神经元的激活功能, 利用信息存储功能的自旋忆阻器件可以实现神经突触记忆和计算一体的功能。这些用于模拟神经突触和神经元功能的自旋电子器件构成了自旋神经形态器件, 并展示出潜在应用价值。

2.1 基于MTJ忆阻器的神经形态特性研究

1971年, 加州大学伯克利分校的蔡少棠教授首次提出第四类基本电子元件——忆阻器^[48], 并被美国惠普实验室首次在实验上证实^[49], 其特点是可以透过外界的电压或者电流调控自身的阻值, 并且在撤去调控手段之后仍能保持一段时间阻值不发生改变, 从而将信息储存在电阻中^[50]。忆阻器的记忆和计算功能使得存算一体成为可能, 并能够模拟突触特性, 从而实现神经形态计算。2009年, Wang等人^[51]提出基于自旋电子器件的忆阻器, 通过基于STT效应和磁畴壁变化来实现器件阻值的调控。其中基于MTJ的忆阻器模型有两种: 一种是通过STT效应调控MTJ的自由层磁矩, 实现自由层和参考层磁矩之间的取向不同, 从而改变隧道的阻值; 第二种是在MTJ的自由层中加入磁畴壁, 通

过输入的电导诱导自由层中磁畴壁发生运动, 调整自由层与参考层磁化方向反向平行的比例, 根据TMR效应实现调控MTJ的阻值。

2.1.1 基于MTJ忆阻器的人工突触特性研究

MTJ所具有的阻值可变、阻值连续可调的特性与神经突触调控权重的特性类似, 基于MTJ自旋忆阻器可以实现模拟神经突触的效果。2011年, 法国科学院Grollier等人^[52]首次制备出具有3个电阻状态的MTJ自由层磁畴壁运动的自旋忆阻器。该器件为半圆环MTJ, 通过垂直器件的磁场在自由层中产生磁畴壁。输入的电通过奥斯特磁场和STT驱动磁畴壁运动, 实现隧道结阻值的改变。固定磁场条件下的电阻变化, 上下为无磁畴壁时的TMR效应的阻值变化, 中间为存在磁畴壁时的阻值变化。但该方案的难点在于确定性调控磁畴壁的位置。要实现隧道结阻值连续变化的忆阻特性, 需要实现对磁畴壁的连续钉扎。蔡佳林等人^[53]提出使用宽度阶梯变化的条带结构的MTJ, 如图2所示, 观测到6个不同的电阻状态, 启发了通过器件形状调控阻值的思路。Samuel等人^[54]设计了具有锯齿的长条形MTJ, 实现了自由层磁畴壁的确定性调控。Leonard等人^[55]在此基础上进一步设计了梯形和长条形的带缺口的MTJ, 并分别实现了突触的可塑性和记忆特性。日本东北大学Ohno等人^[56]利用反铁磁层对铁磁层中磁畴壁的钉扎作用, 也实现了对MTJ阻值的连续调控。

除此之外, 还可以通过改变MTJ材料来实现阻值变化。2012年, 德国比勒费尔德大学Krzyszczko等人^[57]通过控制MTJ中MgO势垒层中的氧空位移动来模拟突触特性。山东大学颜世申教授团队^[58]通过电压调节CuO-ZnO势垒层中氧原子数量来改变MTJ的阻值以实现忆阻器效果, 并模拟了突触的长程增强、长程抑制以及脉冲时间依赖可塑性。北京航空航天大学赵巍胜教授团队^[59,60]先后于2018和2021年提出了两种MTJ忆阻器, 分别是添加阻变保护层的MTJ和垂直磁化的复合自由层MTJ。前者同时利用了阻变特性和TMR效应, 实现了隧道结的多阻态。后者通过向自由层添加钨元素, 退火使钨形成团簇, 增强了磁畴的钉扎效应, 实现了接近连续的阻态变化, 并且成功模拟了突触特性。2022年, 韩国三星高级技术研究所Seungchul Jung研究员^[61]开发出可用自旋转移矩磁阻随机存储器(magnetoresistive random-access memory, MRAM)的交叉开关阵列, 并用于存内计算。该阵列使用电阻求和进行模拟乘法计算, 实现了BNN (bayesian neural network) 网络以

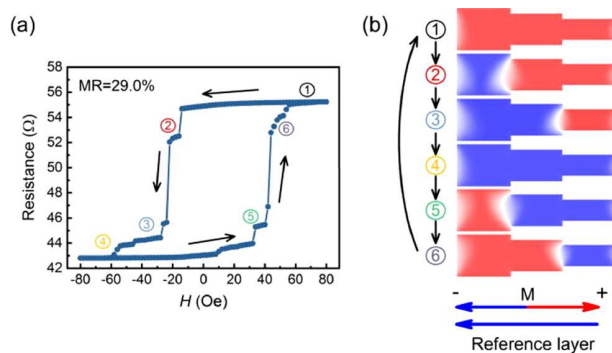


图2 (网络版彩色) (a) 具有6个电阻状态的自旋忆阻器。 (b) 不同电阻状态的微磁学仿真^[53]。 Copyright©2017, AIP Publishing
Figure 2 (Color online) (a) Spin memristor with six resistive states. (b) Micromagnetic simulations of different resistive states^[53]. Copyright©2017, AIP Publishing

对MNIST (mixed national institute of standards and technology)数据和人脸数据的高准确率识别, 分别达到93.2%和93.4%。存内计算与人类大脑的计算类似, 人脑的计算通常发生在记忆或突触网络中, 该MRAM网络将来可能会被用作模拟大脑突触的平台。2024年, 中国科学院微电子研究所邢国忠研究员团队与复旦大学刘明院士团队^[62]共同提出一种基于全电控磁畴壁(domain wall, DW)动力学特性的磁畴壁隧道结器件(DW-MTJ)的多态突触装置, 如图3所示。该装置利用自旋轨道耦合效应(spin-orbit coupling, SOC)和界面Dzyaloshinskii-Moriya相互作用(iDMI)在器件中产生稳定可控的多阻态, 并通过磁场或者电流脉冲实现调控器件的多阻态切换以模拟生物突触的可塑性, 从而实现制备可编程的人工突触器件。上述的研究成果展现了MTJ忆阻器在制备人工神经突触方面的长期的研究积累和巨大的研究潜力。MTJ忆阻器具备阻值记忆和阻值可调特性, 很好对应了神经突触调制权重的特性, 是制备人工神经突触的理想器件之一。

2.1.2 基于MTJ忆阻器的人工神经元特性研究

MTJ因其拥有的非线性输出特性, 也可以实现模拟神经元的效果。神经元行为遵循“泄漏-积累-触发”(leaky integrate and fire, LIF)机制。2013年, Roy等人^[63]制备出以MTJ磁畴壁为基础的一类神经元器件, 当输入电流达到阈值时, 器件电阻发生突变并输出脉冲信号, 从而实现模拟神经元LIF特性。Wesley等人^[64]基于形状各向异性制备出梯形的MTJ LIF神经元器件。但这些器件均需要复杂的外围电路并且仅局限于模拟研究。2022年, 中国科学院微电子研究所邢国忠研究员团

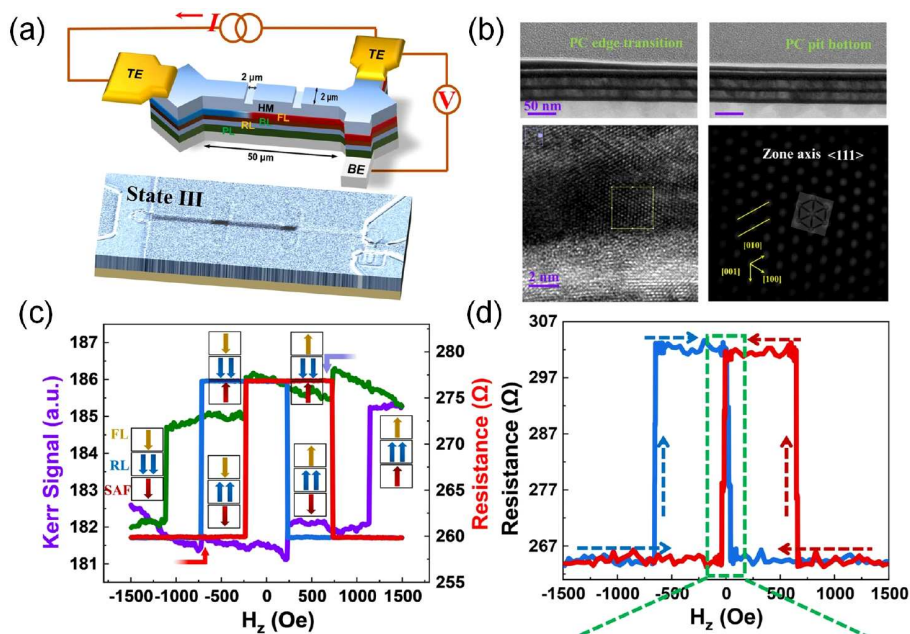


图3 (网络版彩色) (a) 基于DW-pMTJ的自旋突触装置. (b) 器件的TEM图. (c) Kerr-H随H_z的变化. (d) R-H随H_z的变化^[62]

Figure 3 (Color online) (a) Spin-synapse device based-on DW-pMTJ. (b) TEM Image of the device. (c) Kerr-H loops as a function of H_z. (d) R-H as a function of H_z^[62]

队^[55]使用SOT效应调控MTJ的磁畴壁运动, 实现器件阈值触发之后自动重置, 在实验上验证了LIF器件的可行性. 2023年, 他们与复旦大学刘明院士团队^[65]提出基于DW动力学特性, 通过(Ruderman-Kittel-Kasuya-Yosida, RKKY)交换作用和内建磁场相互竞争驱动磁畴壁往复运动的机制, 并制备出具有LIF特性的神经元器件以及阵列. 2024年, 他们再次与复旦大学刘明院士团队^[62]提出基于DW-MTJ器件的人工神经元, 该器件的钉扎中心(pinning center)呈现非线性分布, 从而可以实现由SOT和iDMI协同调控的非线性电阻分布, 其电阻变化与磁场或SOT电流之间的函数符合Sigmoid函数类型. 基于此, 如图4所示, 他们成功构建了具有16×16DW-MTJ突触和16×1的16态DW-MTJ神经元的ANN (artificial neural network)模型. 改变MTJ忆阻器的形状结构和材料组成, 可以很好地获得非线性的脉冲输出信号, 用于构建人工神经元, 体现了MTJ忆阻器在人工神经元方面的广泛应用. 但该方案仍存在复杂的外置电路、较为复杂的器件加工流程、较高能耗的问题, 距离更广泛的实际应用需要更进一步的研究.

2.2 基于MTJ自旋振荡器的神经形态特性研究

电流注入MTJ, 产生的自旋极化电流通过自由层,

在STT的作用下会诱导自由层磁矩发生进动, 自由层和参考层的磁化取向发生周期性的改变, 通过TMR效应使得MTJ电阻也发生周期性振荡^[66], 进而输出高频微波电压信号. 反之, 当注入与MTJ本征频率相同的微波信号时, 会发生共振自旋矩二极管整流现象^[67], 进而输出可以测量的整流电压. 基于MTJ的微波发射和微波探测特性的器件称为自旋振荡器, 具有广泛的应用场景. 利用自旋力矩诱导磁矩振荡特性, 已经发展出多种自旋微波器件, 包括纳米自旋振荡器^[68,69]、超高灵敏度的自旋转移矩微波探测器^[70~73]、宽频微波能量收集器^[74,75]. 除此之外, 利用自旋振荡器丰富的非线性和微波整流等特性, 可以构建人工神经形态器件和网络, 例如利用多个自旋振荡器耦合功能模仿大脑生物神经元^[76], MTJ丰富的自旋二极管整流特性也可以应用于人工神经突触研究.

2.2.1 基于MTJ自旋振荡器的人工突触特性研究

MTJ自旋振荡器的整流特性可以调制输出信号的线性函数, 可以应用于模拟神经突触可调权重的特性. 2021年, 法国的Grollier课题组^[77]通过理论计算证明, 利用MTJ的微波探测特性可实现人工突触所需的MAC (multiply and accumulate)运算功能. 即基于MTJ的自旋矩二极管效应作为神经突触, 将原本输入的微波信号

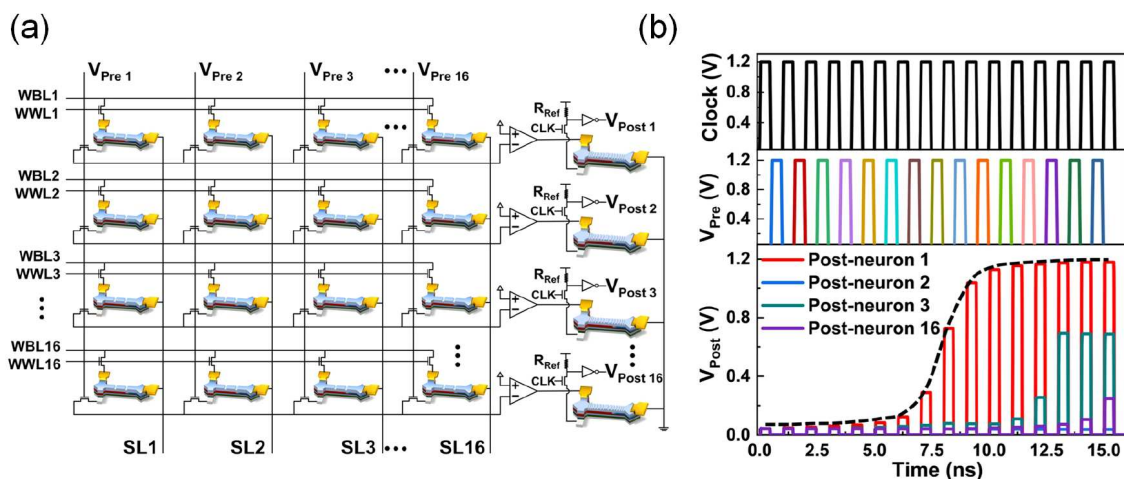


图4 (网络版彩色) (a) 16×16 DW-MTJ突触. (b) 16×16的16态DW-MTJ神经元ANN和电路模拟结果^[62]

Figure 4 (Color online) (a) ANN and circuit simulation results of 16×16 DW-MTJ synapses. (b) 16×16 16-state DW-MTJ neurons^[62]

频率与器件共振频率之间的差值作为突触权重, 实现将微波电流信号转变为可测量的整流电压信号. 随后, 他们通过测量两个互联的MTJ原型器件, 证明了这一推断^[78]. 他们还提出, 结合MTJ的微波发射和探测功能, 可构建大规模的卷积人工神经网络^[79], 用于微波信号的智能识别与处理. 2023年, 他们将多个MTJ器件连接起来^[80], 搭建了简单的神经网络系统, 该神经网络由两个层连接的9个磁隧道结组成. 这些器件在神经网络中借助磁化动力学实现模拟生物突触和神经元, 并通过处理、传输和接收射频信号进行通信. 实验验证其可以对非线性可分离射频输入进行分类, 准确率高达97.7%, 并且将功耗降低了几个数量级.

2024年, 曾柯心等人^[81]构建了一种由射频信号调制的人工突触, 并通过使用该突触执行人工神经网络(ANN), 实现了对8×8像素的手写数字图片的96%的分类识别率. 该器件通过MTJ铁磁自由层和非铁磁层之间的垂直各向异性PMA (perpendicular magnetic anisotropy)来调控谐振频率, 基于MTJ对射频与直流之间的相互转换来模拟突触, 并将输入的射频信号进行整流和加赋权重, 最终通过改变器件的本征频率实现调制权重值的效果, 其本征频率与输入的射频信号频率之间的差值视为权重. 他们利用垂直磁各向异性的自旋矩二极管器件作为神经突触, 实现对多个射频信号的并行处理, 并构建了基于自旋矩二极管微波神经突触的3×4×3的微波神经网络识别所需分类的射频信号, 最终实现了超过96%的识别准确率. 上述研究利用MTJ的微波整流特性成功地模拟了神经突触的可调权重特性,

并借此成功构建了可以实际执行识别任务的人工神经网络, 充分展现了MTJ自旋振荡器在神经形态器件中的发展潜力.

2.2.2 基于MTJ自旋振荡器的人工神经元特性研究

自旋振荡器的微波发射具有非线性阈值特性, 可以用于模拟人工神经元器件. 2015年, 美国普渡大学Roy等人^[82]提出基于自旋振荡器的同步耦合, 可以模拟神经元的激活特性, 从而实现神经形态计算. 2017年, 法国的Grollier课题组^[83]利用MTJ纳米振荡器的发射特性, 巧妙地与储备池计算相结合, 实验验证了纳米振荡器作为人工神经元的可行性, 并将单个器件应用于数字语音识别. 通过使用时分技术和储备池计算技术, 实现了用单个自旋振荡器模拟多个神经元并进行神经形态计算, 在1 μW的功率下实现了99.8%的识别率. 输入的电流会通过STT效应产生进动, 并通过TMR效应产生周期性的电阻变化. 该器件的输入输出信号之间存在非线性关系, 输出信号具有弛豫性, 这意味着其存在短时记忆特性. 自旋振荡器的非线性输出和短时记忆特性满足了储备池计算的条件, 充分地显示了自旋振荡器在构建神经形态器件方面的优势.

2018年, Grollier团队^[45]又展示了基于4个自旋振荡器耦合的语音识别系统, 通过电流和磁场控制自旋振荡器的频率来改变耦合模式, 从而实现模式识别任务. 2019年, 蔡佳林等人^[84]提出了基于自旋振荡器构建人工神经元的方案, 利用直流偏置对自旋振荡器微波探测电压的调控, 同样实现了对神经元激活功能的模拟. 2023年, 法国科学院Grollier教授团队的Farcis等人^[85]利

用具有双自由层结构的微波纳米振荡器,在电压驱动条件下构建尖峰神经元,并通过改变器件上的直流偏置电压来控制输出尖峰率,使其在无磁场条件下对脉冲信号具有尖峰状响应。如图5所示,在该结构的MTJ体系中,两个自由层磁矩均会受到STT效应产生进动,而两层磁矩的进动将会存在耦合。

2023年,张黎可等人^[86]基于面内磁化的MTJ模拟了稀疏神经元,利用器件的非线性整流特性模拟了类似ReLU激活函数的神经元,并用于手写数字的识别,识别率达93.53%。进一步地,他们利用器件的非线性整流特性模拟了神经元的单侧抑制特性,搭建了包含3层神经元的人工神经网络并用以执行图像识别任务,在注入微波功率为0.5 μW 的条件下实现了对手写数字高达94.92%的识别率^[87]。同年,涂华焱等人^[88]基于面内磁化的MTJ,在外加磁场下利用其对称微波响应曲线模拟生物神经元群体中单神经元的钟型调谐曲线,进行了神经元群体编码机制研究,构建了多种人工智能计算中常用的非线性激活函数(包括ReLU与Sigmoid),并在此基础上搭建了基于自旋神经元的人工神经网络,执行手写数字识别任务,识别率最高可达94.88%。MTJ的微波整流、微波探测特性可以很好地实现模拟神经元激活函数,上述的研究结果也印证了MTJ自旋振荡器在构架神经形态器件方面的丰富应用。但MTJ自旋振荡器仍存在着需要配置单独的高频信号发生和采集电路、需要外置磁场等问题,这增加了相关器件的能耗和面积,需要更进一步的研究。

2.3 基于MTJ随机数发生器的神经形态特性研究

一般情况下,MTJ具有高低两个阻态。当MTJ的翻转势垒较低时,外界极小的能量扰动即可引起磁矩发生随机转动,这种现象称作隧道结的随机翻转特性。2014年,Ando等人^[89]提出,MTJ自由层磁矩在磁场/电流与环境焦耳热的共同作用下消弭了翻转势垒的影响,使得自由层磁矩方向与参考层方向在平行和反平行之间随机跳动。根据隧穿磁阻效应,隧道结的阻值也随之发生随机跳变,从而输出二进制随机信号,可以作为随机数发生器,用于随机计算研究。由于随机翻转隧道结具有非线性动力学、低功耗、真随机等特性,十分适用于制备神经形态器件。近些年,随机翻转隧道结也在神经形态计算中引起了广泛的研究。

2.3.1 基于MTJ随机数发生器的人工突触特性研究

MTJ势垒层具有一定的记忆特性,因此施加脉冲

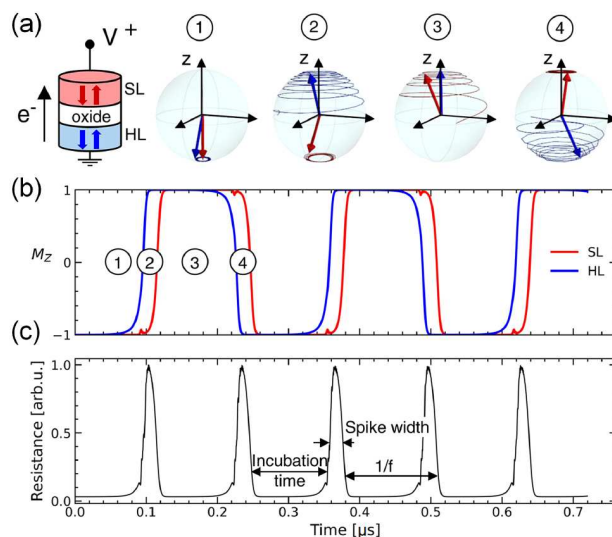


图5 (网络版彩色) (a) 双自由层垂直隧道结结构和三维宏自旋分布。(b) 沿 z 轴磁化强度的时间曲线。(c) 偏置电压下归一电阻的模拟曲线^[85]。Copyright©2023, American Chemical Society

Figure 5 (Color online) (a) Schematic of the dual free layer p-MTJ structure and 3D macrospin configurations. (b) Time trace of the magnetization projection along the z -axis. (c) Simulated normalized resistance under a bias voltage^[85]. Copyright©2023, American Chemical Society

信号频率不同时,引起的隧道结电阻变化大小也不同,这种脉冲之间相互影响的特性可以实现神经突触调节权重的功能。2016年,美国普渡大学Roy等人^[90]提出,可利用脉冲驱动的随机翻转来模拟突触的短程和长程可塑性(short-term and long-term plasticity)。2017年,法国科学院Querlioz等人^[91]发现缩小器件尺寸能够降低隧道结翻转势垒。这种器件无需施加外部信号,在室温扰动下便会自发产生磁矩的随机翻转,被称为超顺磁隧道结。这一成果为设计超低功耗的神经形态器件提供了思路。2021年, Jacob等人^[92]通过宏自旋模型提出,可以使用双电脉冲信号驱动MTJ随机翻转从而模拟突触脉冲时间依赖可塑性(spike-timing-dependent plasticity, STDP)特性,用于构建脉冲神经网络。其原理为第一个脉冲对自由层进行加热降低磁矩翻转势垒,第二个脉冲驱动磁矩发生翻转,两个脉冲信号之间的时间间隔影响第二个脉冲作用时的自由层温度,进而影响翻转概率,由此模拟突触特性。2022年,吕文星等人^[93]基于MgO覆盖自由层MTJ,如图6所示,利用其低导热率和双脉冲调控成功模拟了突触的STDP特性。目前基于MTJ随机翻转特性的神经突触研究较少,主要围绕着前后脉冲对MTJ翻转的影响开展。

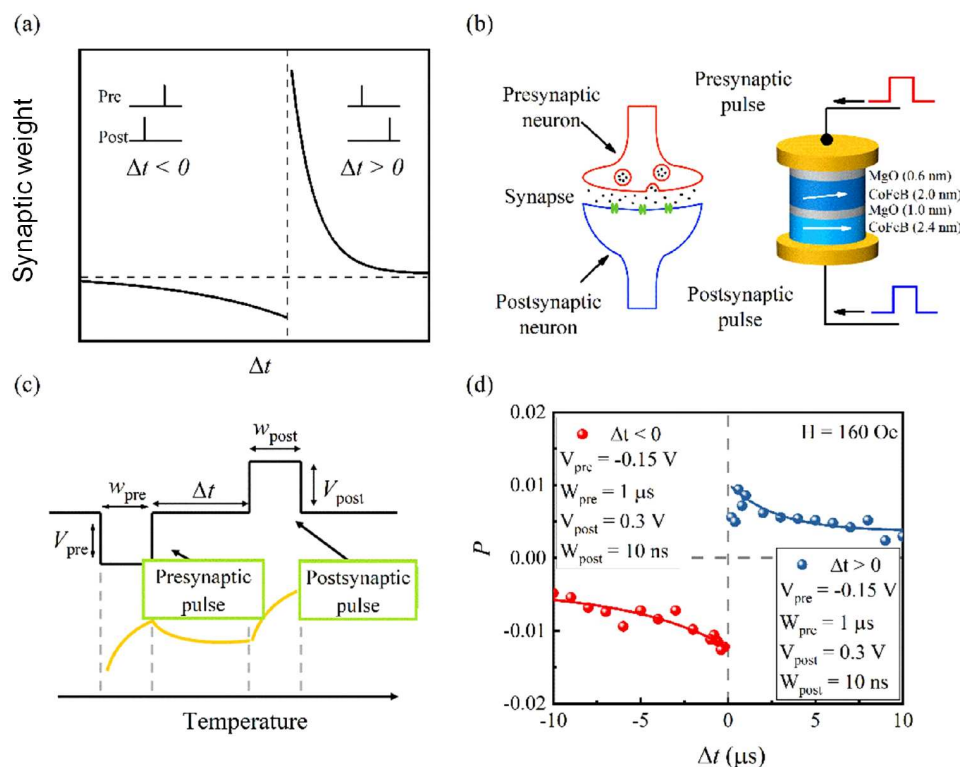


图6 (网络版彩色) (a) STDP特征. (b) MgO盖帽层MTJ的人工神经突触示意图. (c) 脉冲对器件温度影响. (d) 翻转概率曲线^[93]. Copyright©2022, AIP Publishing

Figure 6 (Color online) (a) STDP Characteristic. (b) Schematic of an artificial synaptic device based on MgO-capped MTJ. (c) Influence of synaptic pulses on thermal of device. (d) Switching Probability Curve^[93]. Copyright©2022, AIP Publishing

2.3.2 基于MTJ随机数发生器的人工神经元特性研究

以人工神经网络为代表的第二代深度神经网络,其结构是多层神经元以及与之相连接的突触构成的大规模神经网络,其中神经元作为非线性计算单元,主要的数学模型有Sigmoid、ReLU、softmax等.通过磁场、热量、电场等方式控制随机翻转隧道结的势垒高度,使得随机翻转的概率随着外加偏置的作用逐渐变化,相关研究表明这种变化规律符合人工神经网络中一种特殊的激活函数——Sigmoid函数,能够简单利用单个器件实现输入信号的非线性激活,而无需利用多个晶体管模拟复杂的非线性激活函数,被认为适用于制备人工神经网络中的Sigmoid神经元器件^[12,94-96].2015年, Park等人^[97]提出使用MTJ随机翻转特性模拟神经元的神经编码特性,输出类似神经元的脉冲信号.随后,美国普渡大学Roy等人^[98]提出基于SOT的MTJ随机数发生器. SOT只能提供自由层磁矩到达势垒顶部的能量,而不能跨越即不能决定磁矩的翻转方向.停止电流输入后,磁矩将随机选择一个方向达到稳态.该种随机

数发生器不需要电阻状态发生重置,因此能耗较低.2016年,该团队又提出MTJ随机数发生器可以用于模拟锥形神经元的概率脉冲特性,并构建了相应的神经网络实现图像识别功能^[90].2017年, Roy等人^[95]又在基于自旋轨道力矩驱动的隧道结中通过理论计算搭建隧道结模型,利用流过重金属层的电流脉冲大小不同改变翻转势垒,随机翻转的概率随电流脉冲大小的变化被用于制备Sigmoid神经元器件.

2018年,法国科学院Querlioz等人^[99]基于超顺MTJ的随机翻转特性模拟神经元的编码方案,可以用于高性能低功耗的神经形态计算器件.2024年,中国科学院物理研究所韩秀峰研究员团队^[100]提出基于SOT磁隧道结(SOT-MTJ)的真随机数发生器(true random number generator, TRNG). SOT-MTJ的翻转概率受电压连续调控,使其成为一种概率从0%到100%连续可调的0或1二元TRNG——伯努利TRNG.同时, SOT-MTJ的翻转概率和驱动电压也满足Sigmoid函数关系.该器件可以满足贝叶斯神经网络(BNN)、随机计算等应用.随后,他

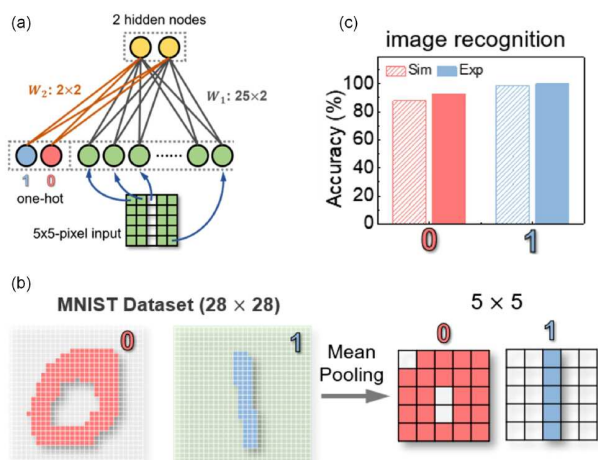


图7 (网络版彩色)基于SOT-MTJ-RBM的图像和语音识别示意图^[101]。(a) RBM模型结构。(b) MNIST数据集的“0”和“1”示例及训练数据。(c) 图像识别的精度。Copyright©2023, American Chemical Society

Figure 7 (Color online) schematic diagram of image and speech recognition based on SOT-MTJ-RBM^[101]. (a) A RBM model structure. (b) “0” and “1” examples of MNIST data sets and training data. (c) The accuracy of image recognition. Copyright©2023, American Chemical Society

们进一步将该SOT-MTJ器件应用到受限玻尔兹曼机(RBM)的吉布斯随机概率采样中^[101], 并进行了手写数字的图像识别和口述数字的语音识别, 如图7所示, 对于数字0(或1), 分别达到了100%(93%)和90%(95%)的识别准确率。

除SOT调控外, 2017年, 美国加州大学洛杉矶分校Wang等人^[102]设计出电场调控的MTJ随机数发生器, 该随机数发生器基于VCMA效应, 通过对MTJ施加一定的电压降低自由层磁矩翻转的势垒, 磁矩会向内排布。撤去电场之后, 翻转势垒恢复, 磁矩会向两侧翻转, 由此产生随机信号。该方案无需输入电流, 不产生电阻损耗, 功耗较低。2019年, 蔡佳林等人^[103]制备了自由层具有垂直磁各向异性的MTJ器件, 在施加偏压的条件下可以发生显著的电阻跳变。通过改变磁场调控器件的随机翻转概率, 获得了电压脉冲信号随磁场变化的S型曲线, 成功模拟了Sigmoid人工神经元, 并实现了对手写数字的97%识别率。他们同时探索了VCMA效应对器件随机翻转的调控作用, 电压对MTJ的随机翻转具有显著的调控作用, 推进了多手段调控MTJ的相关研究。

2024年, 雒雁翔等人^[104]构建了电压控制无磁场的磁隧道结随机计算神经网络, 并实现了95%的手写识别

率。随机计算神经网络使用随机数代替传统神经网络中的浮点数来实现识别任务。通过改变铁磁层厚度引入垂直各向异性降低能垒, 并通过VCMA效应调控切换概率, 实现了无磁场随机数发生器和随机神经元, 构建了一个三层随机计算神经网络, 将像素信息转换为随机比特, 输入到神经网络中进行识别计算, 并实现了对手写任务的95%的识别率。2024年, 新加坡国立大学Hyunsoo Yang教授团队^[105]提出了用一种特殊的自旋器件-超顺MTJ (superparamagnetic tunnel junction, SMTJ)来构造伊辛计算单元(Ising spin), 进而搭建伊辛计算机的方案。

该工作基于80个超顺MTJ实现了具有80个节点的伊辛机计算机, 并通过算法设计成功求解了70个城市的旅行商问题, 如图8所示。求解旅行商问题通常需要4761个节点, SMTJ伊辛机大大降低了所需的节点数, 与现有的伊辛机计算机相比具有更高的能量效率和计算速度, 是未来非常规智能计算的热门方向之一。MTJ随机数发生器具有概率可调特性, MTJ所输出的随机发展概率可以模拟Sigmoid、tanh等众多常见的激活函数, 可以很好地模拟神经元激活特性, 且MTJ随机翻转特性依靠室温扰动激发磁矩随机行为, 仅需要微小电流采集器件的电阻状态, 具有结构设计简单、功耗低的优势, 是构建低功耗神经形态器件的优秀方案之一, 有着广阔的应用场景。

3 总结与展望

3.1 总结

综上所述, 本文通过AI芯片发展过程中的“冯·诺依曼架构”瓶颈, 引出类脑计算和神经形态器件在未来AI技术发展中的优势, 并简要梳理了神经形态芯片的发展历程。通过对神经形态器件不同发展方向的介绍, 提出基于MTJ的神经形态器件的优势, 并列举出目前三个主要的研究方向: 忆阻器、振荡器和随机数发生器。围绕基于MTJ的人工神经元、人工突触以及由两者所构成的神经形态网络, 从器件基本原理以及研究结果进行详细的介绍, 展示了相关研究团队在基于MTJ神经形态器件领域取得的成果。目前在人工突触和神经元的模拟器件方面已经取得了显著的进展。然而, 对于神经网络计算研究仍处于初期阶段。对应的器件层面主要是稳定性和器件良率的问题, 目前可见的人工突触器件中可以用于大规模阵列的计算非常有限, 主要

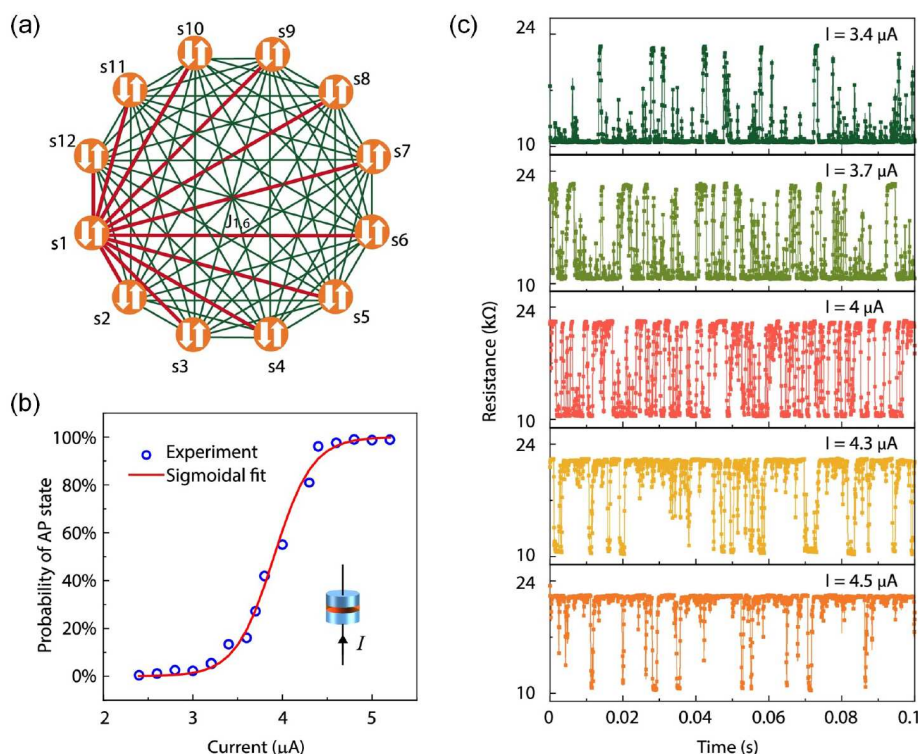


图8 (网络版彩色)基于超顺磁MTJ的伊辛机图示^[105]. (a) 全连接伊辛模型. (b) J 概率随电流的Sigmoid曲线拟合. (c) 不同输入电流下自旋转移矩超顺次MTJ的电阻时间依赖性

Figure 8 (Color online) Illustration of SMTJ-based Ising computer^[105]. (a) All-to-all connected Ising model. (b) Sigmoidal fit of probability under different input currents. (c) Time-dependent resistance of an SMTJ under different input currents

研究团队都是基于单个隧道结或者少数隧道结阵列实现神经网络计算. 一方面, 可以通过探索优化磁性材料生长和纳米器件的制造工艺, 实现高性能、高稳定性的MTJ器件. 另一方面, 可以通过结合神经网络计算理论, 借助实际应用场景功能的演示验证, 从而进一步探索优化MTJ在神经网络方面的应用.

3.2 挑战

尽管目前针对MTJ的神经形态器件的研究已取得长足的进展, 但MTJ神经形态器件仍然是一个很新的研究领域, 距离实际应用还有很多关键科学技术问题需要解决. 基于前文的内容, 我们在这里提出几个MTJ神经形态器件领域值得注意的问题和可能的解决方案.

(1) 自旋神经形态器件阵列化. 当前的神经形态器件研究主要局限于单个器件的时分模拟或少数器件构成的简单神经网络阵列, 虽然这些工作取得了非常好的效果, 但是单个或少数器件无法应对更大规模的应用场景, 实现更为复杂的功能, 且阵列规模提升后器件之间的相互影响和器件性能均一性也是需要解决的问题.

未来将MTJ神经形态器件与CMOS晶体管电路相结合, 可能是解决自旋神经形态器件阵列化的方案之一.

(2) 自旋神经形态器件用于神经形态计算. 自旋神经形态器件在模拟神经突触和神经元方面有了长足的进展, 但是自旋神经形态器件用于构建纯硬件的神经形态网络、实现神经形态计算仍缺乏相关的研究, 如随机计算等. 未来可以研究基于MTJ器件的有限状态机等神经网络基本器件, 探索MTJ神经形态器件纯硬件结构的神经网络, 从而解决MTJ神经形态器件的神经形态计算问题.

(3) 新材料体系和特有的电子自旋属性用于神经形态特性模拟. MTJ神经形态器件在传统材料的基础上构建了丰富的器件特性, 但是部分存在能耗较大、面积较大、需要外置磁场、外置电路等问题, 在未来的神经形态器件研究当中仍需要探索更多新材料、新特性来解决这些问题. 比如利用MTJ器件中的斯格明子结构来实现神经形态的功能, 可以实现低功耗和小尺寸的目标; 利用VCMA效应控制MTJ的随机翻转, 可以实现无需外置磁场和复杂电路的问题等.

3.3 展望

MTJ神经形态器件仍有广阔的发展空间, 结合更前沿的神经网络架构、更新兴的MTJ器件物理特性, 都将会不断促进MTJ神经形态器件的更进一步发展. 这里主要介绍两种潜在的研究方向, 例如基于MTJ的脉冲神经网络(spiking neural network, SNN)计算架构, 以及基于MTJ斯格明子磁结构在神经形态器件中的研究.

(1) 第三代SNN神经网络. 第三代人工神经网络SNN因其高效节能的特性, 近年来已经广泛地应用到模拟神经网络和神经形态计算, SNN也逐步成为MTJ神经形态计算的发展方向之一. SNN的基础包括: 动力学特征神经元、突触可塑性机制、基于脉冲序列的编码方式, 其硬件结构使用人工神经元和突触作为基本运算单元. 而MTJ在之前的研究中已经展现了LIF机制、脉冲时间相关的STDP、时序编码等特性, 可以满足硬件实现SNN的相关要求. 2023年, Tomas等人^[106]报道了一种磁畴壁磁隧道结(DW-MTJ)用于模拟SNN, 通过STT或者SOT调控DW移动来实现模拟神经元尖峰和LIF神经元等. 同年, Chen等人^[107]构建了基于STT效应的磁隧道结随机数发生器, 用于SNN系统的突触当中, 并最终实现了84%的手写识别率. 未来可以进一步结合生物神经网络的组织架构, 根据MTJ的特性开发

实现更合适、高效的SNN系统.

(2) 基于斯格明子磁结构的人工神经形态器件. 磁性斯格明子是具有拓扑性质的类粒子的自旋, 可以在磁隧道结中产生并被找到, 其类粒子性、小尺寸、低能耗、易于操控等特性逐渐成为神经形态计算的方向之一, 在过去的十年间引起广泛的关注和研究. 在基于斯格明子的自旋电子器件中, 斯格明子在外界激励的作用下可以发生有规律的产生、移动、放缩和湮灭现象. 从微观角度来看, 这些现象反映出器件中局域磁矩的变化, 在宏观上对应着器件磁阻的变化. 因此, 斯格明子器件可以通过调控外界条件而实现电阻的多级变化. 利用这一特点, 可以对生物神经元的LIF模型以及生物突触的长、短程突触可塑性进行高仿生度的模拟. 单个斯格明子在储备池计算中展现了优秀的特性, 其具有非线性特性, 可以增加系统的非线性, 提升储备池计算的效率. 在随机计算神经网络当中, 斯格明子所具有的扩散特性也可以为比特流添加随机性. 2017年, 北京航空航天大学赵巍胜教授团队^[108]提出一种基于斯格明子的人工突触器件, 能够模拟生物突触的长、短程可塑性. Li等人^[109]依据斯格明子的非线性动力学提出了斯格明子模拟神经元的可能. 斯格明子的优越特性有望实现下一代低功耗的计算架构.

参考文献

- 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
- Marković D, Mizrahi A, Querlioz D, et al. Physics for neuromorphic computing. *Nat Rev Phys*, 2020, 2: 499–510
- Chklovskii D. Theoretical neuroscience. *Neuron*, 2002, 34: 863–864
- Attwell D, Laughlin S B. An energy budget for signaling in the grey matter of the brain. *J Cereb Blood Flow Metab*, 2001, 21: 1133–1145
- Abbott L F, Nelson S B. Synaptic plasticity: taming the beast. *Nat Neurosci*, 2000, 3: 1178–1183
- Kuzum D, Jeyasingh R G D, Lee B, et al. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett*, 2012, 12: 2179–2186
- del Valle J, Salev P, Tesler F, et al. Subthreshold firing in Mott nanodevices. *Nature*, 2019, 569: 388–392
- Wei S T, Gao B, Wu D, et al. Trends and challenges in the circuit and macro of RRAM-based computing-in-memory systems. *Chip*, 2022, 1: 100004
- Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing. *Nat Rev Mater*, 2020, 5: 173–195
- Ma C, Luo Z, Huang W, et al. Sub-nanosecond memristor based on ferroelectric tunnel junction. *Nat Commun*, 2020, 11: 1439
- Luo Z, Wang Z, Guan Z, et al. High-precision and linear weight updates by subnanosecond pulses in ferroelectric tunnel junction for neuro-inspired computing. *Nat Commun*, 2022, 13: 699
- Grollier J, Querlioz D, Stiles M D. Spintronic nanodevices for bioinspired computing. *Proc IEEE*, 2016, 104: 2024–2039
- Grollier J, Querlioz D, Camsari K Y, et al. Neuromorphic spintronics. *Nat Electron*, 2020, 3: 360–370
- Locatelli N, Cros V, Grollier J. Spin-torque building blocks. *Nat Mater*, 2014, 13: 11–20
- Julliere M. Tunneling between ferromagnetic films. *Phys Lett A*, 1975, 54: 225–226

- 16 Berger L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys Rev B*, 1996, 54: 9353–9358
- 17 Garelo K, Miron I M, Avci C O, et al. Symmetry and magnitude of spin–orbit torques in ferromagnetic heterostructures. *Nat Nanotech*, 2013, 8: 587–593
- 18 Wang M, Cai W, Zhu D, et al. Field-free switching of a perpendicular magnetic tunnel junction through the interplay of spin–orbit and spin-transfer torques. *Nat Electron*, 2018, 1: 582–588
- 19 Kang W, Ran Y, Zhang Y, et al. Modeling and exploration of the voltage-controlled magnetic anisotropy effect for the next-generation low-power and high-speed MRAM applications. *IEEE Trans Nanotechnol*, 2017, 16: 387–395
- 20 Tang J, Yuan F, Shen X, et al. Bridging biological and artificial neural networks with emerging neuromorphic devices: fundamentals, progress, and challenges. *Adv Mater*, 2019, 31: 1902761
- 21 Mead C. Neuromorphic electronic systems. *Proc IEEE*, 1990, 78: 1629–1636
- 22 Mahowald M, Douglas R. A silicon neuron. *Nature*, 1991, 354: 515–518
- 23 Merolla P, Arthur J, Akopyan F, et al. A digital neurosynaptic core using embedded crossbar memory with 45 pJ per spike in 45 nm. In: Proceedings of the 2011 IEEE Custom Integrated Circuits Conference (CICC). New York: IEEE, 2011, <https://ieeexplore.ieee.org/abstract/document/6055294>
- 24 Orchard G, Frady E P, Rubin D B D, et al. Efficient neuromorphic signal processing with Loihi 2. In: Proceedings of the 2021 IEEE Workshop on Signal Processing Systems (SiPS). New York: IEEE, 2021, <https://ieeexplore.ieee.org/abstract/document/9605018>
- 25 Pei J, Deng L, Song S, et al. Towards artificial general intelligence with hybrid Tianji chip architecture. *Nature*, 2019, 572: 106–111
- 26 Ma D, Shen J, Gu Z, et al. Darwin: a neuromorphic hardware co-processor based on spiking neural networks. *J Syst Architecture*, 2017, 77: 43–51
- 27 Ma S, Pei J, Zhang W, et al. Neuromorphic computing chip with spatiotemporal elasticity for multi-intelligent-tasking robots. *Sci Robot*, 2022, 7: eabk2948
- 28 Liu Z, Chen S, Qu P, et al. Sushi: Ultra-high-speed and ultra-low-power neuromorphic chip using superconducting single-flux-quantum circuits. In: Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture. New York: IEEE, 2023. 614–627
- 29 Yang Z, Wang T, Lin Y, et al. A vision chip with complementary pathways for open-world sensing. *Nature*, 2024, 629: 1027–1033
- 30 Prezioso M, Merrih-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
- 31 Yao P, Wu H, Gao B, et al. Face classification using electronic synapses. *Nat Commun*, 2017, 8: 15199
- 32 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
- 33 Yao P, Wu H, Gao B, et al. Fully hardware-implemented memristor convolutional neural network. *Nature*, 2020, 577: 641–646
- 34 Zhong Y, Tang J, Li X, et al. A memristor-based analogue reservoir computing system for real-time and power-efficient signal processing. *Nat Electron*, 2022, 5: 672–681
- 35 Zhu K, Pazos S, Aguirre F, et al. Hybrid 2D–CMOS microchips for memristive applications. *Nature*, 2023, 618: 57–62
- 36 Saighi S, Mayr C G, Serrano-Gotarredona T, et al. Plasticity in memristive devices for spiking neural networks. *Front Neurosci*, 2015, 9: 51
- 37 Chanthbouala A, Garcia V, Cherifi R O, et al. A ferroelectric memristor. *Nat Mater*, 2012, 11: 860–864
- 38 Zhang W, Mazzarello R, Wuttig M, et al. Designing crystallization in phase-change materials for universal memory and neuro-inspired computing. *Nat Rev Mater*, 2019, 4: 150–168
- 39 Sangwan V K, Jariwala D, Kim I S, et al. Gate-tunable memristive phenomena mediated by grain boundaries in single-layer MoS₂. *Nat Nanotech*, 2015, 10: 403–406
- 40 Zhu X, Li D, Liang X, et al. Ionic modulation and ionic coupling effects in MoS₂ devices for neuromorphic computing. *Nat Mater*, 2019, 18: 141–148
- 41 Wetzstein G, Ozcan A, Gigan S, et al. Inference in artificial intelligence with deep optics and photonics. *Nature*, 2020, 588: 39–47
- 42 Shi Y, Ren J, Chen G, et al. Nonlinear germanium-silicon photodiode for activation and monitoring in photonic neuromorphic networks. *Nat Commun*, 2022, 13: 6048
- 43 Zhong C, Liao K, Dai T, et al. Graphene/silicon heterojunction for reconfigurable phase-relevant activation function in coherent optical neural networks. *Nat Commun*, 2023, 14: 6939
- 44 Cheng J, Huang C, Zhang J, et al. Multimodal deep learning using on-chip diffractive optics with in situ training capability. *Nat Commun*, 2024, 15: 6189
- 45 Romera M, Talatchian P, Tsunegi S, et al. Vowel recognition with four coupled spin-torque nano-oscillators. *Nature*, 2018, 563: 230–234
- 46 Tian Y, Bai L, Kang S, et al. Novel magnetoresistance explorations and multistate data storage applications. *Chin Sci Bull*, 2021, 66: 2071–2084
- 47 Feng T, Wang P, Wu D. Spin Hall magnetoresistance in metal/ferromagnetic insulator heterostructures. *Chin Sci Bull*, 2022, 67: 2993–3004
- 48 Chua L. Memristor-The missing circuit element. *IEEE Trans Circuit Theor*, 1971, 18: 507–519

- 49 Strukov D B, Snider G S, Stewart D R, et al. The missing memristor found. *Nature*, 2008, 453: 80–83
- 50 Mathur N D. The fourth circuit element. *Nature*, 2008, 455: E13
- 51 Wang X, Chen Y, Xi H, et al. Spintronic memristor through spin-torque-induced magnetization motion. *IEEE Electron Device Lett*, 2009, 30: 294–297
- 52 Chanthbouala A, Matsumoto R, Grollier J, et al. Vertical-current-induced domain-wall motion in MgO-based magnetic tunnel junctions with low current densities. *Nat Phys*, 2011, 7: 626–630
- 53 Cai J, Fang B, Wang C, et al. Multilevel storage device based on domain-wall motion in a magnetic tunnel junction. *Appl Phys Lett*, 2017, 111: 182410
- 54 Liu S, Xiao T P, Cui C, et al. A domain wall-magnetic tunnel junction artificial synapse with notched geometry for accurate and efficient training of deep neural networks. *Appl Phys Lett*, 2021, 118: 202405
- 55 Leonard T, Liu S, Alamdar M, et al. Shape-dependent multi-weight magnetic artificial synapses for neuromorphic computing. *Adv Elect Mater*, 2022, 8: 2200563
- 56 Fukami S, Zhang C, DuttaGupta S, et al. Magnetization switching by spin–orbit torque in an antiferromagnet–ferromagnet bilayer system. *Nat Mater*, 2016, 15: 535–541
- 57 Krzysteczko P, Münchenberger J, Schäfers M, et al. The memristive magnetic tunnel junction as a nanoscopic synapse-neuron system. *Adv Mater*, 2012, 24: 762–766
- 58 Li Q, Shen T T, Cao Y L, et al. Spin memristive magnetic tunnel junctions with CoO-ZnO nano composite barrier. *Sci Rep*, 2014, 4: 3835
- 59 Zhang Y, Lin X, Adam J, et al. Heterogeneous memristive devices enabled by magnetic tunnel junction nanopillars surrounded by resistive silicon switches. *Adv Elect Mater*, 2018, 4: 1700461
- 60 Zhang X, Cai W, Wang M, et al. Spin-torque memristors based on perpendicular magnetic tunnel junctions for neuromorphic computing. *Adv Sci*, 2021, 8: 2004645
- 61 Jung S, Lee H, Myung S, et al. A crossbar array of magnetoresistive memory devices for in-memory computing. *Nature*, 2022, 601: 211–216
- 62 Liu L, Wang D, Wang D, et al. Domain wall magnetic tunnel junction-based artificial synapses and neurons for all-spin neuromorphic hardware. *Nat Commun*, 2024, 15: 4534
- 63 Sharad M, Fan D, Roy K. Spin-neurons: a possible path to energy-efficient neuromorphic computers. *J Appl Phys*, 2013, 114: 234906
- 64 Brigner W H, Friedman J S, Hassan N, et al. Shape-based magnetic domain wall drift for an artificial spintronic leaky integrate-and-fire neuron. *IEEE Trans Electron Devices*, 2019, 66: 4970–4975
- 65 Wang D, Tang R, Lin H, et al. Spintronic leaky-integrate-fire spiking neurons with self-reset and winner-takes-all for neuromorphic computing. *Nat Commun*, 2023, 14: 1068
- 66 Kiselev S I, Sankey J C, Krivorotov I N, et al. Microwave oscillations of a nanomagnet driven by a spin-polarized current. *Nature*, 2003, 425: 380–383
- 67 Tulapurkar A A, Suzuki Y, Fukushima A, et al. Spin-torque diode effect in magnetic tunnel junctions. *Nature*, 2005, 438: 339–342
- 68 Zeng Z, Amiri P K, Krivorotov I N, et al. High-power coherent microwave emission from magnetic tunnel junction nano-oscillators with perpendicular anisotropy. *ACS Nano*, 2012, 6: 6115–6121
- 69 Zeng Z, Finocchio G, Zhang B, et al. Ultralow-current-density and bias-field-free spin-transfer nano-oscillator. *Sci Rep*, 2013, 3: 1426
- 70 Zhang L, Fang B, Cai J, et al. Ultrahigh detection sensitivity exceeding 105 V/W in spin-torque diode. *Appl Phys Lett*, 2018, 113: 102401
- 71 Zhang L, Cai J, Fang B, et al. Dual-band microwave detector based on magnetic tunnel junctions. *Appl Phys Lett*, 2020, 117: 072409
- 72 Miwa S, Ishibashi S, Tomita H, et al. Highly sensitive nanoscale spin-torque diode. *Nat Mater*, 2014, 13: 50–56
- 73 Fang B, Carpentieri M, Hao X, et al. Giant spin-torque diode sensitivity in the absence of bias magnetic field. *Nat Commun*, 2016, 7: 11259
- 74 Zhang L, Fang B, Cai J, et al. Enhanced broad-band radio frequency detection in nanoscale magnetic tunnel junction by interface engineering. *ACS Appl Mater Interfaces*, 2019, 11: 29382–29387
- 75 Fang B, Carpentieri M, Louis S, et al. Experimental demonstration of spintronic broadband microwave detectors and their capability for powering nanodevices. *Phys Rev Appl*, 2019, 11: 014022
- 76 Fan D, Maji S, Yogendra K, et al. Injection-locked spin Hall-induced coupled-oscillators for energy efficient associative computing. *IEEE Trans Nanotechnol*, 2015, 14: 1083–1093
- 77 Leroux N, Marković D, Martin E, et al. Radio-frequency multiply-and-accumulate operations with spintronic synapses. *Phys Rev Appl*, 2021, 15: 034067
- 78 Leroux N, Mizrahi A, Marković D, et al. Hardware realization of the multiply and accumulate operation on radio-frequency signals with magnetic tunnel junctions. *Neuromorph Comput Eng*, 2021, 1: 011001
- 79 Leroux N, De Riz A, Sanz-Hernández D, et al. Convolutional neural networks with radio-frequency spintronic nano-devices. *Neuromorph Comput Eng*, 2022, 2: 034002

- 80 Ross A, Leroux N, De Riz A, et al. Multilayer spintronic neural networks with radiofrequency connections. *Nat Nanotechnol*, 2023, 18: 1273–1280
- 81 Zeng K, Luo Y, Zhang L, et al. Radio-frequency-modulated artificial synapses based on magnetic tunnel junctions with perpendicular magnetic anisotropy. *Phys Rev Appl*, 2024, 21: 014020
- 82 Yogendra K, Fan D, Roy K. Coupled spin torque nano oscillators for low power neural computation. *IEEE Trans Magn*, 2015, 51: 1–9
- 83 Torreon J, Riou M, Araujo F A, et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature*, 2017, 547: 428–431
- 84 Cai J, Zhang L, Fang B, et al. Sparse neuromorphic computing based on spin-torque diodes. *Appl Phys Lett*, 2019, 114: 192402
- 85 Farcis L, Teixeira B M S, Talatchian P, et al. Spiking dynamics in dual free layer perpendicular magnetic tunnel junctions. *Nano Lett*, 2023, 23: 7869–7875
- 86 Zhang L, Luo Y, Tu H, et al. Highly sensitive spin-torque diodes based on in-plane magnetized magnetic tunnel junctions. *Appl Phys Lett*, 2023, 122: 232402
- 87 Zhang L, Tu H, Luo Y, et al. Ultralow-current density spin-torque diodes with high sensitivity. *Appl Phys Lett*, 2023, 122: 092405
- 88 Tu H, Zhang L, Luo Y, et al. Neural-like population coding based on spin-torque diode. *Appl Phys Lett*, 2023, 122: 122402
- 89 Fukushima A, Seki T, Yakushiji K, et al. Spin dice: a scalable truly random number generator based on spintronics. *Appl Phys Express*, 2014, 7: 083001
- 90 Sengupta A, Panda P, Wijesinghe P, et al. Magnetic tunnel junction mimics stochastic cortical spiking neurons. *Sci Rep*, 2016, 6: 30039
- 91 Vodenicarevic D, Locatelli N, Mizrahi A, et al. Low-energy truly random number generation with superparamagnetic tunnel junctions for unconventional computing. *Phys Rev Appl*, 2017, 8: 054045
- 92 Velarde H I, Nagaria J, Yin Z, et al. Intrinsic spike-timing-dependent plasticity in stochastic magnetic tunnel junctions mediated by heat dynamics. *IEEE Magn Lett*, 2021, 12: 1–5
- 93 Lv W, Cai J, Tu H, et al. Stochastic artificial synapses based on nanoscale magnetic tunnel junction for neuromorphic applications. *Appl Phys Lett*, 2022, 121: 232406
- 94 Sengupta A, Parsa M, Han B, et al. Probabilistic deep spiking neural systems enabled by magnetic tunnel junction. *IEEE Trans Electron Devices*, 2016, 63: 2963–2970
- 95 Liyanagedera C M, Sengupta A, Jaiswal A, et al. Stochastic spiking neural networks enabled by magnetic tunnel junctions: from non-telegraphic to telegraphic switching regimes. *Phys Rev Appl*, 2017, 8: 064017
- 96 Camsari K Y, Sutton B M, Datta S. p-bits for probabilistic spin logic. *Appl Phys Rev*, 2019, 6: 011305
- 97 Suh D I, Bae G Y, Oh H S, et al. Neural coding using telegraphic switching of magnetic tunnel junction. *J Appl Phys*, 2015, 117: 17D714
- 98 Kim Y, Fong X, Roy K. Spin-orbit-torque-based spin-dice: a true random-number generator. *IEEE Magn Lett*, 2015, 6: 1–4
- 99 Mizrahi A, Hirtzlin T, Fukushima A, et al. Neural-like computing with populations of superparamagnetic basis functions. *Nat Commun*, 2018, 9: 1533
- 100 Zhang R, Li X, Zhao M, et al. Probability-distribution-configurable true random number generators based on spin-orbit torque magnetic tunnel junctions. *Adv Sci*, 2024, 11: 2402182
- 101 Li X, Wan C, Zhang R, et al. Restricted Boltzmann machines implemented by spin-orbit torque magnetic tunnel junctions. *Nano Lett*, 2024, 24: 5420–5428
- 102 Lee H, Ebrahimi F, Amiri P K, et al. Design of high-throughput and low-power true random number generator utilizing perpendicularly magnetized voltage-controlled magnetic tunnel junction. *AIP Adv*, 2017, 7: 055934
- 103 Cai J, Fang B, Zhang L, et al. Voltage-controlled spintronic stochastic neuron based on a magnetic tunnel junction. *Phys Rev Appl*, 2019, 11: 034015
- 104 Luo Y, Wu Y, Zeng K, et al. Magnetic field-free stochastic computing based on the voltage-controlled magnetic tunnel junction. *Appl Phys Lett*, 2024, 124: 212403
- 105 Si J, Yang S, Cen Y, et al. Energy-efficient superparamagnetic Ising machine and its application to traveling salesman problems. *Nat Commun*, 2024, 15: 3457
- 106 Leonard T, Liu S, Jin H, et al. Stochastic domain wall-magnetic tunnel junction artificial neurons for noise-resilient spiking neural networks. *Appl Phys Lett*, 2023, 122: 262406
- 107 Chen X, Li H, Fan H, et al. MTJ-based random number generation and its application in SNN handwritten digits recognition. *AIP Adv*, 2023, 13: 105113
- 108 Huang Y, Kang W, Zhang X, et al. Magnetic skyrmion-based synaptic devices. *Nanotechnology*, 2017, 28: 08LT02
- 109 Li S, Kang W, Zhang X, et al. Magnetic skyrmions for unconventional computing. *Mater Horiz*, 2021, 8: 854–868

Summary for “基于磁性隧道结的人工神经形态器件”

Artificial neuromorphic devices based on magnetic tunnel junctions

Yujie Wang^{1,2}, Like Zhang³, Renjuan Luo^{1,2}, Yimeng Lu^{1,2}, Bin Fang^{1,2} & Zhongming Zeng^{1,2,4*}

¹ Nano Science and Technology Institute, University of Science and Technology of China, Hefei 230026, China

² Nanofabrication Facility, Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences, Suzhou 215123, China

³ College of Electronic Information Engineering, Wuxi University, Wuxi 214105, China

⁴ Division of Nano-Devices and Technologies & Nanchang Key Laboratory of Advanced Packaging, Jiangxi Institute of Nanotechnology, Nanchang 330200, China

* Corresponding author, E-mail: zmzeng2012@sinano.ac.cn

The development of artificial intelligence technology is intrinsically linked to advancements in hardware technology. Traditional Von Neumann architecture, which separates computational units and data storage units, faces significant bottlenecks in terms of processing speed and power consumption. These challenges, often referred to as the “storage wall” and the “power consumption wall”, pose substantial barriers to further progress in artificial intelligence technology. As the complexity and computational demands of artificial intelligence algorithms continue to grow, overcoming these limitations has become an urgent priority. Neuromorphic devices, which aim to mimic the low-power operation of the human brain, present a promising pathway for addressing these challenges. By emulating the fundamental structural and connectivity features of the brain’s neural systems, these devices enable hardware-based neuromorphic computing that can bypass many of the limitations of traditional architectures. Neurons and synapses, the basic building blocks of biological neural networks, serve as the core components simulated by neuromorphic devices. Among the various technologies being explored for this purpose, magnetic tunnel junctions stand out due to their numerous advantages. As core elements of spintronics, magnetic tunnel junctions offer long operational lifetimes, low power consumption, non-volatility, high speed, and multifunctionality, making them highly suitable for simulating artificial neurons and synapses in neuromorphic systems. Their unique ability to exploit the spin properties of electrons provides a versatile platform for memory and computation functions. Magnetic tunnel junctions can perform a variety of functions by manipulating the movement of magnetic domains. This manipulation adjusts the resistance of the device, enabling both memory and computational capabilities within a single memristor, which is crucial for simulating neural synapses. Furthermore, applying electrical current to magnetic tunnel junctions generates spin-polarized currents. These currents produce spin-transfer torques, which induce oscillations in the free-layer magnetic moments, resulting in high-frequency voltage signals. This property allows for the execution of weighted calculations and nonlinear function simulations based on resonance frequencies, facilitating the emulation of synaptic and neuronal functions.

Additionally, the barrier properties of magnetic tunnel junctions can be modulated through electrical and thermal means. This modulation causes the magnetic moments in the free and pinned layers to transition between parallel and antiparallel states. These resistance fluctuations generate random signals, which are particularly useful for true random number generation and probabilistic computations. By tailoring the applied conditions, these random signals can mimic various neuronal behaviors, further enhancing the functionality of neuromorphic systems.

This paper begins by examining the development challenges faced by current artificial intelligence hardware, focusing on the limitations of traditional architectures. It provides a concise overview of the progress in artificial neuromorphic devices and the materials employed in their fabrication. The fundamental structure and operating principles of spintronic devices, particularly magnetic tunnel junctions, are then introduced. The discussion subsequently delves into the specific roles of magnetic tunnel junctions-based neuromorphic devices, including spin memristors, spin oscillators, and spin random number generators, in simulating artificial neural synapses and neurons. Current research efforts in these areas are reviewed, highlighting the advances made as well as the remaining challenges. Finally, the paper explores the prospects for the development of spintronic-based neuromorphic devices, emphasizing their potential to revolutionize artificial intelligence hardware and pave the way for more efficient and powerful computing systems.

magnetic tunnel junction, artificial neuron, artificial synapse, spintronic memristor, spin oscillator, spin random number generator

doi: [10.1360/TB-2024-0931](https://doi.org/10.1360/TB-2024-0931)