科学通报 2025 年 第 70 卷 第 13 期: 1874 ~ 1883
面向人工智能的物理神经网络专题 进 展
《中国科学》杂志社 SCIENCE CHINA PRESS

基于自旋的智能器件与物理神经网络进展

刘祥语[1,2], 文辉[1,2], 雷坤[1,2], 兰修凯[1], 王开友[1,2*]

1. 中国科学院半导体研究所半导体超晶格国家重点实验室, 北京 100083
2. 中国科学院大学材料科学与光电技术学院, 北京 100049
* 联系人, E-mail: kywang@semi.ac.cn

2024-08-30 收稿, 2024-11-10 修回, 2024-11-20 接受, 2024-11-21 网络版发表
国家重点研发计划(2022YFA1405100)和国家自然科学基金(12241405, 12204473, 12274406, 12104449)资助

摘要 在信息爆炸的时代, 解决计算中不断升级的能源需求变得至关重要. 传统的计算机系统面临性能限制, 促使人们寻找新的计算范式. 自旋电子器件以其非易失、速度快、高能效等特点, 有望解决传统电子器件所面临的困境. 本文从自旋人工电子突触、自旋振荡器神经元、自旋人工神经网络和概率神经网络四个方面综述了基于自旋的智能器件和物理神经网络的最新研究进展. 自旋电子器件的蓬勃发展及其在智能计算和信息存储等领域展现出的应用潜力有望激发新一轮的信息技术革命.

关键词 自旋电子器件, 物理神经网络, 自旋轨道矩, 存算一体, 神经形态计算

近些年, 人工智能(artificial intelligence, AI)、物联网、5G/6G通讯、智能驾驶等新兴产业的迅猛发展, 伴随着算力需求和能源压力的日益增长. 基于传统互补金属-氧化物-半导体(complementary metal oxide semiconductor, CMOS)晶体管的计算系统普遍具有如图1(a)所示的存算分离的架构, 在执行AI等数据密集型运算任务时面临高延迟、高能耗的性能瓶颈. 与之相对的, 人脑由大量具有非易失特性的突触和神经元组成的神经网络处理信息, 在拥有高度智能的同时, 功耗仅有数十瓦. 近年来, 人们尝试使用电子器件和集成电路模拟人脑的功能[1]. 然而, 由于存算分离的架构, 利用CMOS构建神经网络需要耗费大量的硬件资源和极高的功耗. 得益于自旋电子学的蓬勃发展, 自旋电子器件因其非易失性和高速、低功耗、与CMOS兼容等优点而被视为实现新型电子神经网络的有力载体, 受到人们的广泛关注[2,3].

自旋电子器件依靠调控磁性物质的磁化状态实现信息的处理和存储功能, 通常具备非易失性. 目前, 电学方法控制自旋翻转主要采用自旋转移矩(spin-transfer torque, STT)[4]和自旋轨道矩(spin-orbit torque, SOT)[5~8]效应, 它们的读写方式如图1(b)所示. 首先, 自旋转移矩依靠自旋极化电流转移角动量来改变自旋电子器件的磁化排列. 例如, 在一个具有铁磁/非磁绝缘体/铁磁三明治结构的磁隧穿结(magnetic tunnel junction, MTJ)器件中, 电流可以通过流经自旋有序排列的局域磁矩(或通过局域磁矩对电子的反射)来形成自旋极化电流. 自旋极化电流可以将其携带的角动量传递给磁自由层, 如果施加的电流足够大, 自旋极化电流携带的角动量足够多, 可以通过控制电流加载方向的不同来控制自由层的磁化方向, 从而改变磁自由层和磁钉扎层的相对磁化排列方向. 相较于自旋转移矩[9], 自旋轨道矩调控磁化状态具有翻转速度更快(<1 ns)、能耗更低和器件耐久更高(>10^{15})等优势[10,11]. 通常, SOT-MTJ在磁自由层邻近设置一个具有强自旋轨道耦合(spin-orbit coupling, SOC)相互作用的自旋源层(典型材料为重金属). 当对自旋源层施加电场时, 电荷流通过源

引用格式: 刘祥语, 文辉, 雷坤, 等. 基于自旋的智能器件与物理神经网络进展. 科学通报, 2025, 70: 1874–1883
Liu X, Wen H, Lei K, et al. Progress in spin-based intelligent devices and physical neural networks (in Chinese). Chin Sci Bull, 2025, 70: 1874–1883, doi: 10.1360/TB-2024-0923

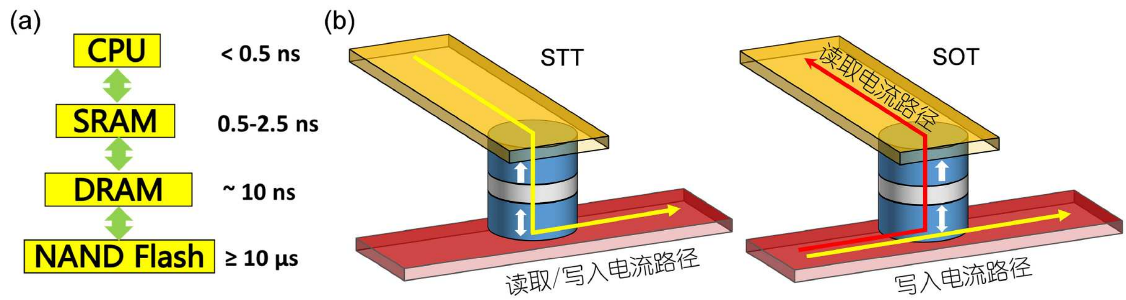

图 1 (网络版彩色)存算分离架构与电控自旋电子器件示意图. (a) 存算分离的传统计算架构; (b) STT和SOT调控自旋电子器件的读写方式

Figure 1 (Color online) Schematic diagrams of the conventional computational architecture and electronically controlled spintronic devices. (a) Conventional computational architecture with separation of computation and storage; (b) read/write of the spintronic device modulated by spin-transfer torques and spin-orbit torques

层中或界面处的自旋轨道耦合作用转化为自旋流并注入到磁自由层中, 利用自旋流携带的角动量诱导磁自由层发生磁化方向改变, 从而实现对器件的写入. 目前, 基于STT的磁性随机存储器(magnetic random access memory, MRAM)已经进入商业化阶段, 而针对SOT-MRAM的研究仍处在实验室和集成工艺开发阶段[12~22].

人脑是一个复杂而精妙的客观存在, 它由数以亿计的神经元构成, 这些神经元之间通过突触形成了错综复杂的网络连接, 承载了人脑强大的计算和处理能力. 以新型非易失电子器件模拟神经元和突触的基本工作原理被视为构建高效人工神经网络的基础. 利用自旋电子器件模拟人脑中突触和神经元的基本工作原理, 基于对磁矩的精准控制, 可以构建存算一体的高效自旋人工神经网络[23,24]. 目前, 利用自旋电子器件构建人工突触、神经元乃至神经网络的研究方兴未艾, 本文将简单介绍相关研究的最新进展.

1 自旋轨道矩人工突触

脉冲时间依赖可塑性(spike timing dependent plasticity, STDP)是神经科学中描述神经元之间突触连接强度变化的一种机制, 其中突触连接强度的变化与突触前神经元和突触后神经元发放脉冲的精确时间差紧密相关. STDP被认为是大脑中学习和信息存储的重要机制之一. 通过调整神经元之间的连接强度, STDP可以塑造神经网络的结构和功能, 使其能够学习和适应外部环境的变化.

人工电子器件实现STDP需要有足够多的可调节状态变化来模拟突触的权值变化. 一般地, 具有垂直磁各向异性的自旋轨道矩器件只有磁化向上和向下两种稳定状态. 然而, 随着研究深入, 研究者们采用层间交换耦合[25]、交换偏置[26]、磁场调控[27]等方法实现了自旋轨道矩器件的稳定磁化多态翻转, 这为人工模拟突触可塑性提供了新途径, 如图2(a~c)所示. 这些多态器件可制备为类脑计算中的仿生突触, 通过电流脉冲设置垂直磁化状态模拟突触权重更新, 实现基于自旋电子器件的兴奋性/抑制性突触后电位和STDP等类突触功能. 在性能上, 以补偿亚铁磁材料[28,29]作为自旋轨道矩器件的磁自由层可以观测到纳秒级的SOT诱导模拟突触响应, 有望实现超快类脑计算(如图2(d)所示). 另外, Li等人[30]通过数值模拟证明, 依靠设置前后神经元间电流脉冲间隔, 调控器件中的焦耳热效应, 实现了两态自旋轨道矩器件的概率翻转可调, 这种翻转概率与电流脉冲间隔的关系也与STDP功能一致. 此外, 自旋轨道矩器件中的拓扑磁结构如斯格明子(skyrmion)也被证明可用于人工突触的模拟[31,32], 且其线性权重变化更加理想, 因为它取决于累积的斯格明子的数量.

相较于基于CMOS晶体管的人工突触, 自旋人工突触具有非易失性和简单紧凑的器件结构, 有望应用在人工神经网络和类脑计算中, 实现低功耗、低延迟的原位智能计算.

2 自旋振荡器神经元

人脑神经元是神经系统中负责接收、传导和整合信息的基本结构和功能单元, 有着非常复杂的非线性动力学特征. 仿生学计算有助于我们认识大脑的运行规律, 同时, 基于它的新计算范式应用会加速我们对认知问题的求解. 大脑中的神经元表现为一个非线性振荡器, 产生节律性活动并相互作用来处理信息, 而纳米尺度的自旋电子振荡器, 由于其优良的非线性动力学

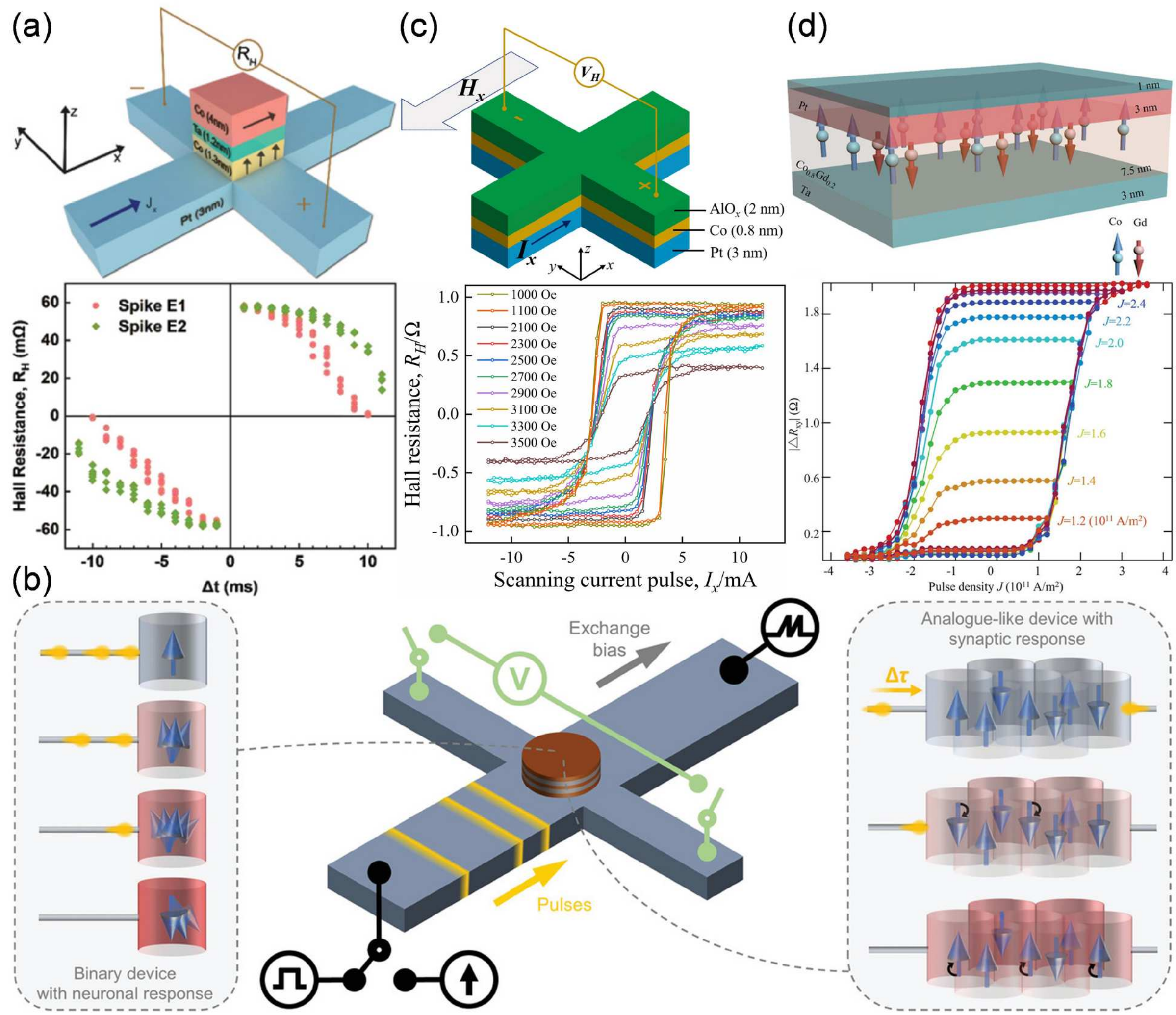

图 2 (网络版彩色)自旋轨道矩人工突触. (a) 利用层间交换耦合的多态翻转自旋电子器件[25], Copyright © 2019 WILEY-VCH Verlag GmbH & Co. KGaA; (b) 利用交换偏置的多态翻转自旋电子器件[26], Copyright © 2019 WILEY-VCH Verlag GmbH & Co. KGaA; (c) 磁场调控下的多态翻转[27]; (d) 补偿亚铁磁CoGd的多态翻转[28], Copyright © 2021 Wiley-VCH GmbH

Figure 2 (Color online) Spin-orbit torque artificial synapses. (a) Multilevel spintronic device utilizing interlayer exchange coupling[25], Copyright © 2019 WILEY-VCH Verlag GmbH & Co. KGaA; (b) multilevel spintronic device using exchange bias[26], Copyright © 2019 WILEY-VCH Verlag GmbH & Co. KGaA; (c) multistate switching with magnetic field modulation[27]; (d) multistate switching with compensated ferrimagnet CoGd[28], Copyright © 2021 Wiley-VCH GmbH

特征和大规模集成的潜力，有望用来实现类似于人脑的高效计算模式.

自旋纳米振荡器分为自旋转矩纳米振荡器(spin-torque nano-oscillator, STNO)[33~36]和自旋霍尔纳米振荡器(spin-Hall nano-oscillator, SHNO)[37]两种. 前者STNO利用自旋转移矩(STT)效应来实现微波信号的振荡. 当给一个STNO器件输入的直流信号强度达到或超过阈值时，STT能够补偿磁阻尼，使自由层磁矩围绕有效场持续进动，进而通过磁电阻效应转化为射频输出(radiofrequency, RF)信号，因此STNO器件通常采用MTJ为核心结构，如图3(a)左图所示. 后者SHNO则基于自旋霍尔效应(spin Hall effect, SHE)来产生磁矩振荡. 当输入电流强度达到或超过阈值时，SOT能够补偿磁阻尼，使磁自由层磁矩围绕有效场持续进动. 目前SHNO的研究通常采用重金属/铁磁金属双层平面结构，如图3(a)右图所示，这种由磁矩进动产生的微波输出信号可以通过布里渊光散射显微镜[37]、频谱分析仪[38]等方式测量.

Torrejon等人[33]将单个STNO应用到循环神经网络(recurrent neural network, RNN)中，利用动态振幅来实

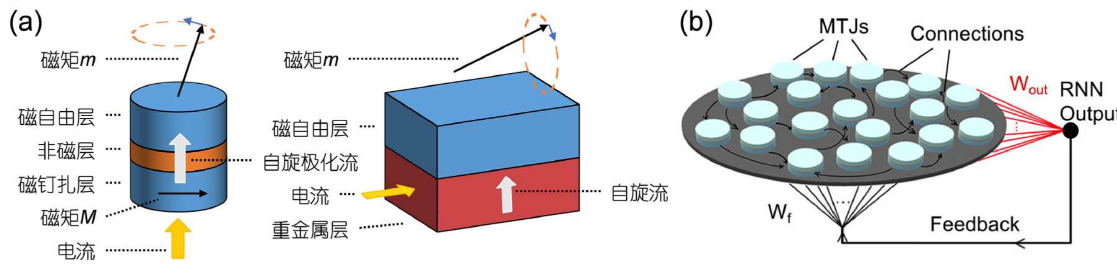

图 3 (网络版彩色)自旋振荡器神经元及神经网络. (a) 自旋转矩纳米振荡器和自旋霍尔纳米振荡器的结构示意图; (b) 由40个相互耦合的自旋转矩纳米振荡器构成的阵列[35]

Figure 3 (Color online) Spin oscillator neurons and neural networks. (a) Schematic diagram of STNO and STHO structures; (b) an array of 40 STNOs coupled to each other[35]

现复杂的神经形态计算，从而完成了高效率的数字语音识别. 后续Romera等人[34]利用4个STNO之间的耦合效应, 实现了多个振荡器的同步进动, 其能够学习并准确分类不同的元音模式，表明了小规模的自旋电子振荡网络在复杂的模式分类问题上的求解能力. Jiang等人[39]通过数值仿真，构建了由斯格明子磁畴型忆阻器与24个STNO器件构成的自旋储备池神经网络, 演示了图像分类和复杂未知非线性动态问题的求解，如非线性动态系统的预测. Zheng等人[35]通过对40个STNO器件的仿真，如图3(b)所示，考虑到MTJ各自的和相互之间的非线性动力学过程，成功实现了时间序列的产生和识别. 此外, 自旋振荡器还可以将输入的射频信号转化为直流电信号, 即整流效应[40]. Ross等人[36]分别利用具有整流效应和自旋振荡效应的STNO器件构建人工突触和神经元, 制备了全自旋神经网络, 成功实现了非线性分类任务. 相比STNO, SHNO可以表现出更高的频率和更宽的频率调谐范围，在高频应用中具有更广阔的应用空间. Zahedinejad等人[37]利用自旋霍尔效应及磁偶极子相互作用，在10 GHz工作频率下构建了至多8×8的二维SHNO同步阵列，是STNO阵列工作频率的25倍[34]; 同时，他们验证了这些二维SHNO阵列可用于处理更复杂的神经形态计算任务，如图像分割和边缘检测.

3 自旋人工神经网络

正如真正的突触和神经元共同构成生物神经网络,基于自旋器件的人工突触和神经元可以组建自旋人工神经网络. 这种神经网络具有自旋器件的低功耗、非易失等优势，相比传统CMOS电路更适合大规模的神经网络运算，在AI技术充分发展的当下具有广阔的应用前景.

神经网络运算主要由乘加(multiply-accumulate, MAC)运算构成, 具有非易失性的自旋器件阵列天然地与这种计算方式相匹配，通过将权重存储在阵列单元中，将行和列分别作为输入和输出，MAC运算得以原位地执行. 2022年，三星利用集成化的STT-MRAM阵列实现了二值化的神经网络运算[41]，实现了最高405 TOPS/W (tera operations per second per watt, 万亿操作每秒每瓦特)的能效，其阵列结构如图4(a)所示. 随后,台积电采用类似方案实现了高能效的神经网络计算,且加入了物理不可克隆(physically unclonable function, PUF)模块，提升了安全性，可用于边缘计算等场景中[42]. 图4(b)左图展示了一种基于PtW合金的全电控SOT器件. 由于PtW合金中的Pt和W元素分别具有相反的自旋霍尔角，该合金层在电场的作用下能够产生竞争自旋流，打破对称性，驱动铁磁层实现无外磁场翻转[12]. 通过两个PtW器件构筑如图4(b)右图所示的同或逻辑电路，可以将SOT器件用于二值化神经网络运算[43,44], 它们具有更低的能耗和更快的写入速度, 是自旋神经网络的有力候选者. 自旋电子器件不仅可以用于二值化神经网络，一些经过特殊设计的自旋电子器件还可进行模拟权重值的神经网络计算. 例如, 采用具有多个MTJ单元的SOT器件[45]、磁畴壁MTJ器件[46]等方案，如图4(c, d), 都可以实现更高精度的神经网络训练和推理. 这些研究进展证实了自旋器件应用于人工神经网络的可行性, 特别是在某些特定场景, 诸如边缘计算和物联网中，具有替代传统CMOS芯片的巨大潜力.

4 自旋概率神经元和神经网络

除了上述较为精确和确定性的神经网络模型之外,随机性在大脑中也扮演着重要角色. 神经元和突触的

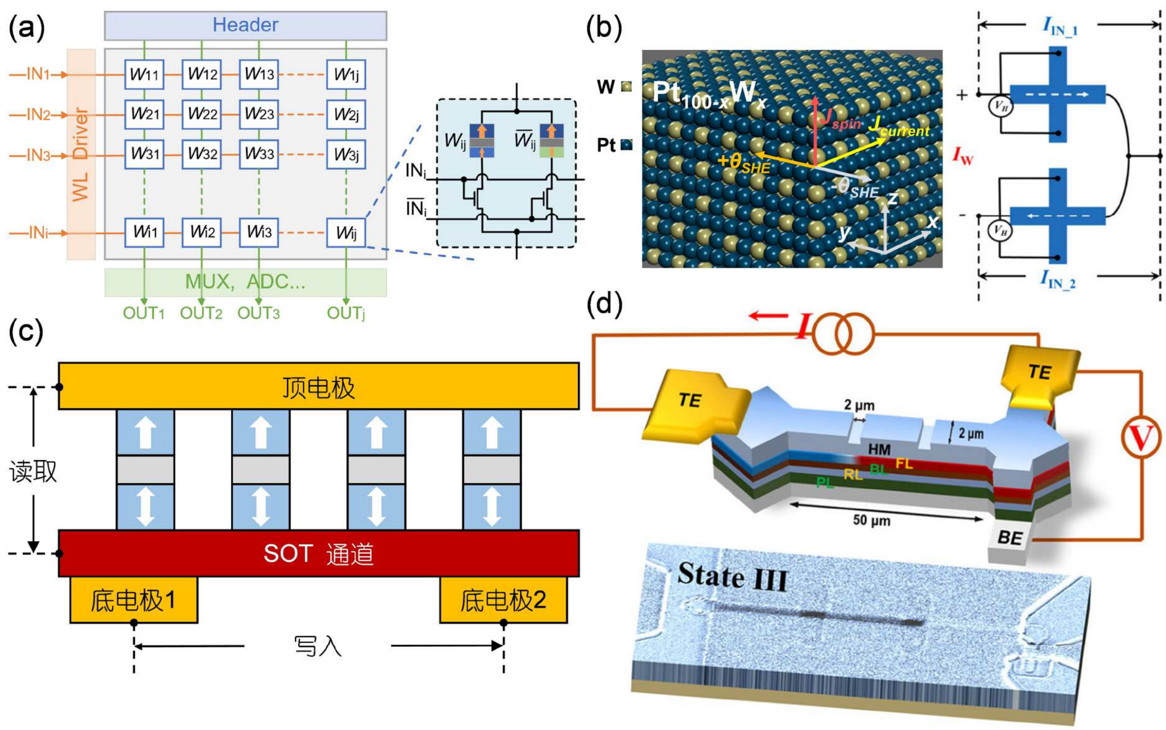

图 4 (网络版彩色)自旋人工神经网络. (a) 基于STT-MRAM的二值化神经网络阵列结构示意图; (b) 基于无外场SOT器件的二值化神经网络单元结构示意图[12,43], Copyright © 2023 IEEE, Copyright © 2023 AIP Publishing; (c) 用于模拟权重神经网络计算的多MTJ单元SOT器件示意图; (d) 用于模拟权重神经网络计算的磁畴壁MTJ器件示意图[46]

Figure 4 (Color online) Spin-based artificial neural networks. (a) Schematic diagram of the array structure for binary neural networks based on STT-MRAM; (b) schematic diagram of the unit cell of binary neural networks based on field-free SOT devices[12,43], Copyright © 2023 IEEE, Copyright © 2023 AIP Publishing; (c) schematic diagram of the multi-pillar SOT device for neural network calculations with multi-level weights; (d) schematic diagram of the domain wall MTJ device for neural network calculations with analog weights[46]

随机性使得大脑能够处理更加复杂和多变的信息[47]. 如图5(a)所示, 通过降低MTJ器件的能垒(≤15k_BT, 其中k_B为玻尔兹曼常数, T为温度), 由于室温条件下热噪声的影响, 自由层磁矩将在两个稳定态之间随机波动[48]. 改变施加在器件两端的电压, 电流产生的STT将会调控器件随机输出为某一状态的概率, 实现0~100%的翻转概率调控. 这种器件通常被称作概率比特, 可以看作一种具有可调随机性的概率神经元. 类似地, 利用压控磁各向异性(voltage-controlled magnetic anisotropy, VCMA)效应降低自由层的能垒, 可以通过施加不同的外部磁场实现对器件翻转概率的调控[49].

通常, 对于一个伊辛网络, 其哈密顿量可以表示为

$$H_{\text{Ising}} = \sum_{i,j} J_{ij} s_i s_j - \sum_i h_i s_i, \tag{1}$$

其中, s_i代表伊辛网络中每个自旋的状态(通常为0或1), J_{ij}为相互作用矩阵元, 代表自旋间的相互作用强度, h_i代表施加在每个自旋上的外场强度. 如果将伊辛网络映射到自旋神经网络中, 其中概率MTJ器件用于表示伊辛网络中的自旋s_i, 器件间的相互耦合及器件受到的外部激励分别代表伊辛网络中的J_{ij}和h_i, 自旋神经网络获得特定组态S的概率满足玻尔兹曼统计:

$$P(S) \propto \mathrm{e}^{-\frac{H_{\text{Ising}}}{k_B T}}. \tag{2}$$

观察(2)式可以发现, 哈密顿量H_{Ising}越小, 对应的自旋组态S出现的概率越高, 其中基态出现的概率最高. 因此, 由自旋概率神经元组成的神经网络能够在室温下高效执行伊辛模型的基态搜索等复杂计算问题[50,51], 且具备可扩展性、与CMOS兼容等优点, 在密码安全、资源优化、药物研发、人工智能等领域具有重要应用. 然而, 基于低热稳定性的概率神经元牺牲了自旋器件的非易失性, 对通过应用存内计算的方式进一步降低概率神经网络的能耗造成了挑战. 如图5(b)所示, 在高能垒SOT器件中, 重置初始状态后施加临界翻转电流

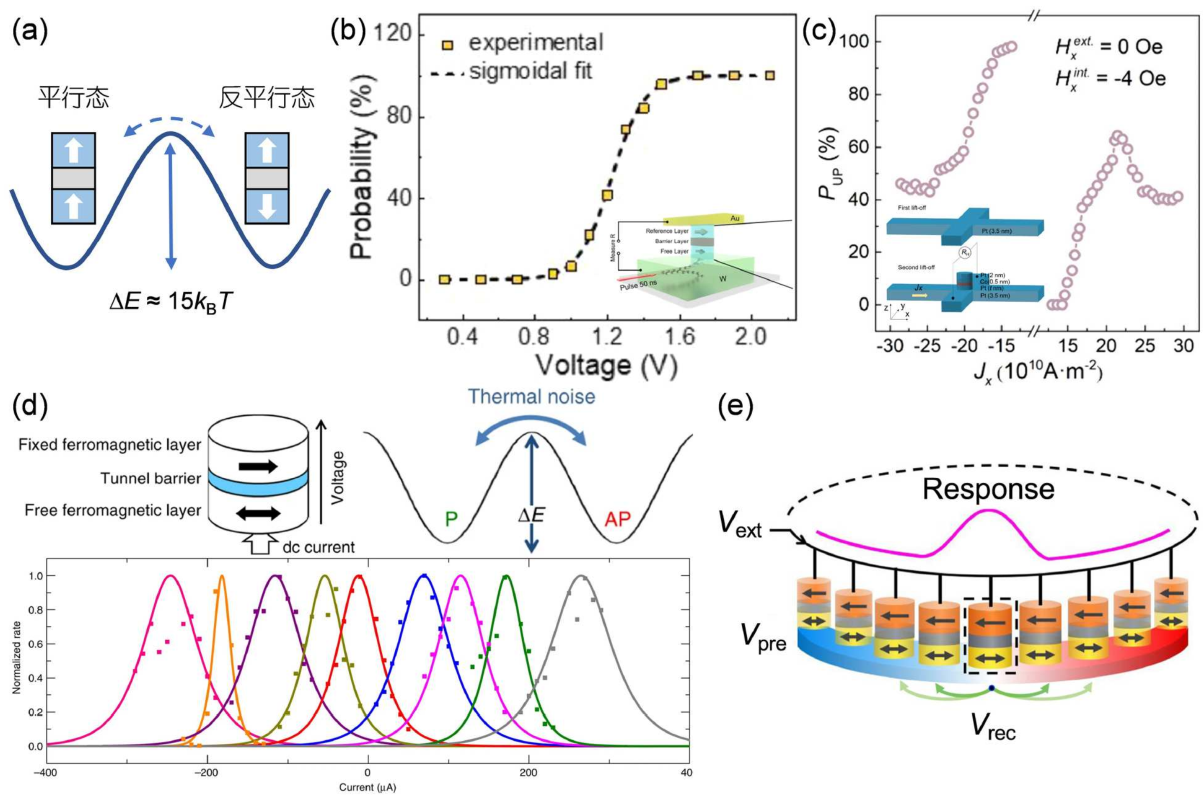

图 5 (网络版彩色)自旋概率神经元和神经网络. (a) 低能垒STT-MTJ器件的随机性示意图; (b) 高能垒SOT-MTJ器件的可调翻转概率[53], Copyright © 2024 American Chemical Society; (c) 无初始化的全电控高能垒SOT概率比特[54], Copyright © 2024 American Chemical Society; (d) 自旋概率神经元及群体编码[56]; (e) 基于自旋概率神经元的连续吸引子神经网络[59], Copyright © 2024 American Physical Society

Figure 5 (Color online) Spin-based probabilistic neurons and neural networks. (a) Schematic diagram of stochasticity of the low-energy barrier STT-MTJ device; (b) tunable switching probability in the high-energy barrier SOT-MTJ device[53], Copyright © 2024 American Chemical Society; (c) initialization- and field-free high-energy barrier SOT p-bit[54], Copyright © 2024 American Chemical Society; (d) spin-based probabilistic neurons and population coding[56]; (e) continuous-attractor neural network based on spintronic probabilistic neurons[59], Copyright © 2024 American Physical Society

密度J_c(通常被定义为50%翻转概率对应的电流密度[52])附近的驱动电流, 也可以实现保留非易失性的自旋概率神经元[53]. 最近, Ren等人[54]演示了一种利用高能垒SOT器件后跳效应的无初始化全电控自旋概率比特(图5(c)), 避免了高能垒SOT概率比特初始化的问题, 且垂直各向异性的器件结构具有比面内各向异性更高的集成密度. 此外, 理论计算表明, 基于纳米级斯格明子的螺旋自由度在受挫磁体中的量子行为, 可以构建由电场和自旋电流控制的单量子比特门和双量子比特门等, 有望实现通用的量子计算, 进而高效地执行伊辛问题求解等功能[55]

除了自旋概率比特, 低能垒MTJ器件还可以应用于群体编码[56]. 群体编码是大脑中常见的编码方式, 通过多个神经元编码同一个函数功能, 能够大大提高神经网络的容错性[57]. 如图5(d)显示, 在低能垒MTJ中, STT电流大小不仅能调控器件的翻转概率, 而且还能调控器件的翻转频率. 器件翻转频率和STT电流大小的关系近似为一条高斯曲线. 对多个器件施加不同的偏置电流, 归一化的器件翻转频率可以覆盖更大范围的输入电流. 这组器件的翻转频率-STT电流关系可以组成一组基函数, 任意函数$H(I)$可以利用这组基函数表示为

$$H(I)=\sum_i w_i r_i(I), \tag{3}$$

其中, $r_i(I)$为器件翻转频率与电流的对应关系曲线, w_i为每个器件的权重, I为输入电流. Mizrahi等人[56]基于多个自旋概率神经元的群体编码实现了手写字体生成等功能. 此外, Zheng等人[58]基于此类具有高斯响应的低能垒MTJ器件构建了连续吸引子神经网络(如图5(e)所示), 通过模拟和理论分析, 验证了自旋概率神经元可

用于实现高鲁棒性的多感官信息整合系统.

5　总结与展望

近年来，人们通过电场对磁矩实现精准或概率性质的调控，实现了对人脑中突触或神经元功能的模拟，并初步构建了基于自旋的物理神经网络．这一前沿探索不仅挑战了传统计算与信息处理模式的边界，还预示着智能科技领域的一场深刻变革．在自旋电子器件中，磁矩的方向、强度和稳定性等多方面指标能够被稳定精确地调控，研究者们有望由此设计出模拟生物神经元间复杂计算机制的智能电路和芯片．这些电路和芯片可以像大脑中的神经网络一样，根据外部刺激动态调整其连接强度，实现信息的传递、整合与处理，从而在硬件层面上模拟出人脑的部分认知与学习能力．这一技术的突破，不仅为神经形态计算、类脑智能等前沿领域提供了强有力的支撑，也为解决当前人工智能在计算效率、能耗及自主适应性等方面的瓶颈问题开辟了新的方向.

参考文献

1 Pei J, Deng L, Song S, et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. Nature, 2019, 572: 106–111
2 Cao Y, Xing G, Lin H, et al. Prospect of spin-orbitronic devices and their applications. iScience, 2020, 23: 101614
3 Zhou C Q, Qin R D, Li X. Research progress of nuero-inspired computing using non-volatile memory (in Chinese). J Funct Mater Devices, 2021, 27: 505–513 [周琮泉, 秦瑞东, 李鑫. 基于非易失性存储器件的类脑计算研究进展. 功能材料与器件学报, 2021, 27: 505–513]
4 Slonczewski J C. Current-driven excitation of magnetic multilayers. J Magn Magn Mater, 1996, 159: L1–L7
5 Miron I M, Garello K, Gaudin G, et al. Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. Nature, 2011, 476: 189–193
6 Liu L, Pai C F, Li Y, et al. Spin-torque switching with the giant spin Hall effect of tantalum. Science, 2012, 336: 555–558
7 Cai K, Yang M, Ju H, et al. Electric field control of deterministic current-induced magnetization switching in a hybrid ferromagnetic/ferroelectric structure. Nat Mater, 2017, 16: 712–716
8 Cao Y, Sheng Y, Edmonds K W, et al. Deterministic magnetization switching using lateral spin–orbit torque. Adv Mater, 2020, 32: 1907929
9 Lee T Y, Lee J M, Kim M K, et al. World-most energy-efficient MRAM technology for non-volatile RAM applications. In: Proceedings of the 2022 International Electron Devices Meeting (IEDM). New York: IEEE, 2022, https://ieeexplore.ieee.org/abstract/document/10019430
10 Beek S V, Cai K, Yasin F, et al. Scaling the SOT track—A path towards maximizing efficiency in SOT-MRAM. In: Proceedings of the 2023 International Electron Devices Meeting (IEDM). New York: IEEE, 2023, https://ieeexplore.ieee.org/abstract/document/10413749
11 Grimaldi E, Krizakova V, Sala G, et al. Single-shot dynamics of spin–orbit torque and spin transfer torque switching in three-terminal magnetic tunnel junctions. Nat Nanotechnol, 2020, 15: 111–117
12 Liu X, Lan X, Bekele Z A, et al. High-efficiency field-free spin-orbit switching based on PtW alloy layer. IEEE Trans Nanotechnol, 2023, 22: 576–580
13 Sheng Y, Edmonds K W, Ma X, et al. Adjustable current-induced magnetization switching utilizing interlayer exchange coupling. Adv Elect Mater, 2018, 4: 1800224
14 Sheng Y, Wang W, Deng Y, et al. Electrically function-switchable magnetic domain-wall memory. Natl Sci Rev, 2023, 10: nwad093
15 Bekele Z A, Li R, Li Y, et al. Tuning the high-efficiency field-free current-induced deterministic switching via ultrathin PtMo layer with Mo content. Adv Elect Mater, 2021, 7: 2100528
16 Deng Y, Li W, Lan X, et al. Field-free switching of spin crossbar arrays by asymmetric spin current gradient. Adv Funct Mater, 2024, 34: 2307612
17 Li W, Zhu W, Zhang G, et al. Room-temperature van der Waals ferromagnet switching by spin-orbit torques. Adv Mater, 2023, 35: 2303688
18 Yang M, Deng Y, Wu Z, et al. Spin logic devices via electric field controlled magnetization reversal by spin-orbit torque. IEEE Electron Device Lett, 2019, 40: 1554–1557
19 Li R, Li Y, Sheng Y, et al. All-electrical multifunctional spin logics by adjusting the spin current density gradient in a single device. ACS Appl Electron Mater, 2021, 3: 2646–2651
20 Zhang N, Cao Y, Li Y, et al. Complementary lateral-spin–orbit building blocks for programmable logic and in-memory computing. Adv Elect Mater, 2020, 6: 2000296
21 Deng Y, Liu X, Chen Y, et al. All-electrical switching of a topological non-collinear antiferromagnet at room temperature. Natl Sci Rev, 2023, 10: nwac154
22 Liu X, Feng Q, Zhang D, et al. Topological spin textures in a non-collinear antiferromagnet system. Adv Mater, 2023, 35: 2211634
23 Bai Y, Yin J L, Guo Z X, et al. Technology roadmap and expectation of low-power spintronic devices (in Chinese). Micro Nano Electron Intell

Manuf, 2021, 3: 104–128 [白月, 殷加亮, 郭宗夏, 等. 低功耗自旋电子器件技术路线及展望. 微纳电子与智能制造, 2021, 3: 104–128]
24 Liu R H, Li L Y, Chen L N, et al. Nonlinear dynamics and applications of spin Hall nano-oscillators (in Chinese). Prog Phys, 2020, 40: 195–210 [刘荣华, 李丽媛, 陈丽娜, 等. 自旋霍尔纳米振荡器的非线性动力学及其应用. 物理学进展, 2020, 40: 195–210]
25 Cao Y, Rushforth A W, Sheng Y, et al. Tuning a binary ferromagnet into a multistate synapse with spin–orbit-torque-induced plasticity. Adv Funct Mater, 2019, 29: 1808104
26 Kurenkov A, DuttaGupta S, Zhang C, et al. Artificial neuron and synapse realized in an antiferromagnet/ferromagnet heterostructure using dynamics of spin–orbit torque switching. Adv Mater, 2019, 31: 1900636
27 Lan X, Cao Y, Liu X, et al. Gradient descent on multilevel spin–orbit synapses with tunable variations. Adv Intelligent Syst, 2021, 3: 2000182
28 Liu J, Xu T, Feng H, et al. Compensated ferrimagnet based artificial synapse and neuron for ultrafast neuromorphic computing. Adv Funct Mater, 2022, 32: 2107870
29 Li W, Lan X, Liu X, et al. Switching plasticity in compensated ferrimagnetic multilayers for neuromorphic computing. Chin Phys B, 2022, 31: 117106
30 Li H, Li L, Zhou K, et al. Stochastic spin-orbit-torque device as the STDP synapse for spiking neural networks. Sci China-Phys Mech Astron, 2023, 66: 257512
31 Song K M, Jeong J S, Pan B, et al. Skyrmion-based artificial synapses for neuromorphic computing. Nat Electron, 2020, 3: 148–155
32 Huang Y, Kang W, Zhang X, et al. Magnetic skyrmion-based synaptic devices. Nanotechnology, 2017, 28: 08LT02
33 Torrejon J, Riou M, Araujo F A, et al. Neuromorphic computing with nanoscale spintronic oscillators. Nature, 2017, 547: 428–431
34 Romera M, Talatchian P, Tsunegi S, et al. Vowel recognition with four coupled spin-torque nano-oscillators. Nature, 2018, 563: 230–234
35 Zheng Q, Zhu X, Mi Y, et al. Recurrent neural networks made of magnetic tunnel junctions. AIP Adv, 2020, 10: 025116
36 Ross A, Leroux N, De Riz A, et al. Multilayer spintronic neural networks with radiofrequency connections. Nat Nanotechnol, 2023, 18: 1273–1280
37 Zahedinejad M, Awad A A, Muralidhar S, et al. Two-dimensional mutually synchronized spin Hall nano-oscillator arrays for neuromorphic computing. Nat Nanotechnol, 2020, 15: 47–52
38 Zahedinejad M, Fulara H, Khymyn R, et al. Memristive control of mutual spin Hall nano-oscillator synchronization for neuromorphic computing. Nat Mater, 2022, 21: 81–87
39 Jiang W, Chen L, Zhou K, et al. Physical reservoir computing using magnetic skyrmion memristor and spin torque nano-oscillator. Appl Phys Lett, 2019, 115: 192403
40 Sharma R, Ngo T, Raimondo E, et al. Nanoscale spin rectifiers for harvesting ambient radiofrequency energy. Nat Electron, 2024, 7: 653–661
41 Jung S, Lee H, Myung S, et al. A crossbar array of magnetoresistive memory devices for in-memory computing. Nature, 2022, 601: 211–216
42 Chiu Y C, Khwa W S, Yang C S, et al. A CMOS-integrated spintronic compute-in-memory macro for secure AI edge devices. Nat Electron, 2023, 6: 534–543
43 Lan X, Liu X, Bekele Z A, et al. Field-free spin–orbit devices via heavy-metal alloy with opposite spin Hall angles for in-memory computing. Appl Phys Lett, 2023, 122: 172402
44 Song M Y, Chen K L, Chen K M, et al. High RA dual-MTJ SOT-MRAM devices for high speed (10 ns) compute-in-memory applications. In: Proceedings of the 2023 International Electron Devices Meeting (IEDM). New York: IEEE, 2023, https://ieeexplore.ieee.org/abstract/document/10413832
45 Doevenspeck J, Garello K, Rao S, et al. Multi-pillar SOT-MRAM for accurate analog in-memory DNN inference. In: Proceedings of the 2021 Symposium on VLSI Technology. New York: IEEE, 2021, https://ieeexplore.ieee.org/abstract/document/9508714
46 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
47 Pouget A, Beck J M, Ma W J, et al. Probabilistic brains: Knowns and unknowns. Nat Neurosci, 2013, 16: 1170–1178
48 Borders W A, Pervaiz A Z, Fukami S, et al. Integer factorization using stochastic magnetic tunnel junctions. Nature, 2019, 573: 390–393
49 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
50 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
51 Singh N S, Kobayashi K, Cao Q, et al. CMOS plus stochastic nanomagnets enabling heterogeneous computers for probabilistic inference and learning. Nat Commun, 2024, 15: 2685
52 Zhang C, Fukami S, Sato H, et al. Spin-orbit torque induced magnetization switching in nano-scale Ta/CoFeB/MgO. Appl Phys Lett, 2015, 107: 012401
53 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

54 Ren R, Cao Y, Wang C, et al. Initialization-free and magnetic field-free spin–orbit p-bits with backhopping-like magnetization switching for probabilistic applications. Nano Lett, 2024, 24: 10072–10080
55 Xia J, Zhang X, Liu X, et al. Universal quantum computation based on nanoscale skyrmion helicity qubits in frustrated magnets. Phys Rev Lett, 2023, 130: 106701
56 Mizrahi A, Hirtzlin T, Fukushima A, et al. Neural-like computing with populations of superparamagnetic basis functions. Nat Commun, 2018, 9: 1533
57 Pouget A, Dayan P, Zemel R. Information processing with population codes. Nat Rev Neurosci, 2000, 1: 125–132
58 Zheng Q, Mi Y, Zhu X, et al. Anticipative tracking with the short-term synaptic plasticity of spintronic devices. Phys Rev Appl, 2020, 14: 044060
59 Zhao L, Zhang M, Zhang Y, et al. Neuroscience-inspired information-integration system based on stochastic magnetic tunnel junctions. Phys Rev Appl, 2024, 21: 064040

Summary for "基于自旋的智能器件与物理神经网络进展"

Progress in spin-based intelligent devices and physical neural networks

Xiangyu Liu[1,2], Hui Wen[1,2], Kun Lei[1,2], Xiukai Lan[1] & Kaiyou Wang[1,2*]

[1] *State Key Laboratory for Superlattices and Microstructures, Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China*
[2] *College of Materials Science and Opto-Electronic Technology, University of Chinese Academy of Sciences, Beijing 100049, China*
* Corresponding author, E-mail: kywang@semi.ac.cn

In the era of information explosion, addressing computational bottlenecks and escalating energy demands has become crucial. Conventional computer systems face performance limitations, prompting the search for novel computational paradigms. Spin-based electronic devices, renowned for their non-volatility, fast response and high energy efficiency, offer promising solutions. This review outlines recent progress in spin-based intelligent devices and physical neural networks, focusing on spintronic synapses, spin oscillator neurons, spin-based artificial neural networks (ANNs), and probabilistic neural networks.

Spin-orbit torque (SOT) devices are emerging as attractive candidates for artificial synapses due to their non-volatility and tunability. By exploiting methods like interlayer exchange coupling, exchange bias, and magnetic field modulation, researchers have achieved stable multi-state switching in SOT devices, mimicking synaptic weight changes. This enables functionalities like spike-timing-dependent plasticity (STDP), critical for learning and information storage. Devices utilizing topological magnetic structures like skyrmions show linear weight updates, enhancing performance.

Inspired by the nonlinear dynamics of biological neurons, spin-torque nano-oscillators (STNOs) and spin-Hall nano-oscillators (SHNOs) exhibit similar behavior. STNOs utilize spin-transfer torque, while SHNOs leverage spin Hall effect for microwave oscillations. Applications in recurrent neural networks (RNNs) for speech recognition and radiofrequency (RF) signal identification have demonstrated their potential in neuromorphic computing.

Building on the low power consumption and non-volatility of spin devices, spin-based ANNs have emerged as viable alternatives to CMOS-based networks. Arrays of spintronic devices facilitate in-memory computing, significantly reducing energy consumption. Studies using STT-MRAM and SOT devices have achieved binary and analog weight neural network computations, demonstrating their suitability for edge computing and IoT applications.

Recognizing the role of stochasticity in biological neural networks, spintronic devices with low energy barriers exhibit intrinsic randomness, enabling probabilistic computation. By manipulating the switching probability of magnetic tunnel junctions (MTJs), probabilistic bits mimic neurons capable of solving complex problems like Ising model optimization. High-energy barrier SOT devices retain non-volatility while allowing tunable stochasticity, advancing the field.

Recent advancements in spin-based devices have paved the way for the development of physical neural networks that mimic the functionalities of biological neural systems. By precisely controlling magnetic moments, researchers are designing circuits and chips that can dynamically adjust their connections, mirroring the cognitive and learning abilities of the brain. This groundbreaking technology promises profound impacts on neuromorphic computing, brain-inspired AI, and addressing the bottlenecks in current AI systems related to efficiency, energy consumption, and adaptability.

spintronic devices, physical neural networks, spin-orbit torque, in-memory computing, neuromorphic computing

doi: 10.1360/TB-2024-0923