用于抗噪脉冲神经网络的随机畴壁磁隧道结人工神经元

Thomas Leonard, [1, b)] Samuel Liu, [1, b)] Harrison Jin, [1] 和 Jean Anne C. Incorvia[1, a)]

[1]电气与计算机工程系，德克萨斯大学奥斯汀分校，美国得克萨斯州奥斯汀市 78712

脉冲神经网络（SNNs）中神经元行为的时空特性，使SNNs在要求高能效的边缘应用中具有良好的应用前景。为了

引言

神经形态计算能够实时分析复杂程度超出传统计算能力的非结构化信息。目前，神经形态计算的数字实现依赖大量 CMOS 晶体管来模拟人工神经网络（ANN），在面积和能耗方面都效率低下[1]。对于能够从高能效实现中获益

已有多种磁性器件被提出用于模拟脉冲神经元。这些方法通常侧重于利用在数十微米尺度上制备的霍尔条，通过渐进式开关来模拟神经元行为的漏积分发放（LIF）模型[9–12]。其中一种方法采用了畴壁（DW）自旋织构，但其

a) 通讯作者。电子邮箱：incorvia@austin.utexas.edu b) Thomas Leonard 和 Samuel Liu

将MTJ读出与缩放后的DW自旋纹理相结合，以表示神经元状态。这些DW-MTJ器件提供了一种高度可调的方式，可在单片平台上模拟多种神经形态功能。我们展示了使用单个MTJ实现的多权重（MW）人工突触，以及通过改变

本文提出了一种具有随机行为的二值DW-MTJ人工神经元，该神经元采用与此前展示的多权重突触相同的材料制成[2]，为实现具有完全自旋电子矩阵运算的脉冲神经网络提供了可能。该器件如图1所示，由一条矩形DW轨道组

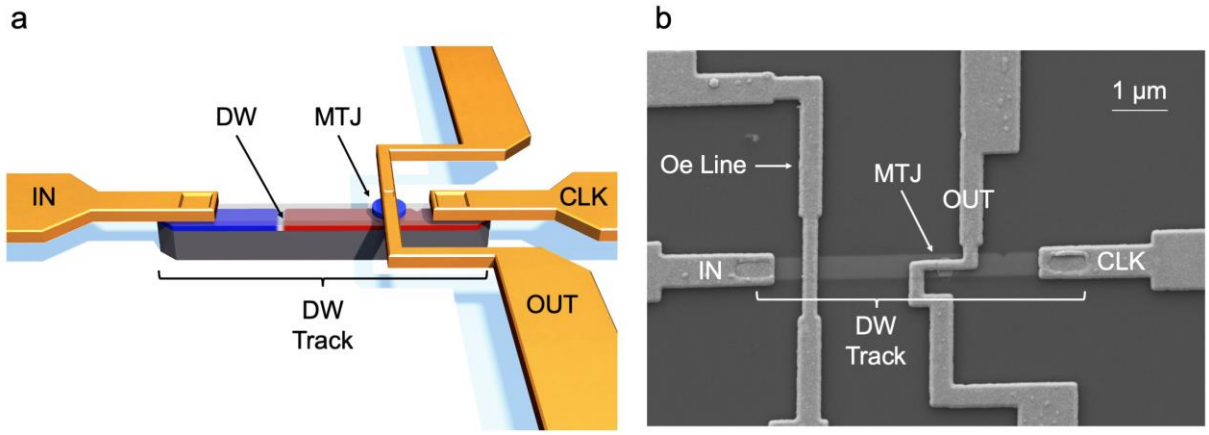


图1. 器件示意图。(a) 二值DW-MTJ随机神经元的3D渲染图。蓝色/红色分别表示向上/向下的磁畴，白色表示磁畴壁（DW）。灰色表示钽底层。三个关键端子分别为IN、CLK和OUT。(b) 二值DW-MTJ随机神经元的SEM图像。右侧缺口清晰可见；左侧缺口位于Oe线下方。

当隐藏层规模或时间戳数量减少时，所提出的二值神经元的开销变得可以忽略不计。这表明，随机二值神经元可应用于边缘计算，其中空间和能耗受到严格限制。当使用相同的网络执行推理时，这一点变得更加明显。在加入高斯

实验数据

图1(a)展示了所研究的二值DW-MTJ人工神经元示意图。与文献18中的逻辑器件相比，区别在于MTJ沿DW轨道的相对位置。这一点见扫描电子显微镜（SEM）图像图1(b)。材料堆叠结构（详见补充图S1）和制备工艺与文献2、18相同。该器件利用钽底层产生的自旋轨道力矩（SOT）驱动DW运动；由于MTJ的顶层由合成反铁磁层钉扎，MTJ的电阻取决于DW在轨道中的位置。MTJ两侧各有一个通过光刻形成的缺口，用于将DW稳定在轨道上的两个位置之一。器件工作时，在CLK与IN之间施加电压脉冲，以成核、解钉扎并驱动DW沿轨道移动。随后从IN至OUT读取MTJ电阻。测试期间未使用奥斯特（Oe）线，因为写入过程中DW成核具有概率性，而Oe线会降低随机性。

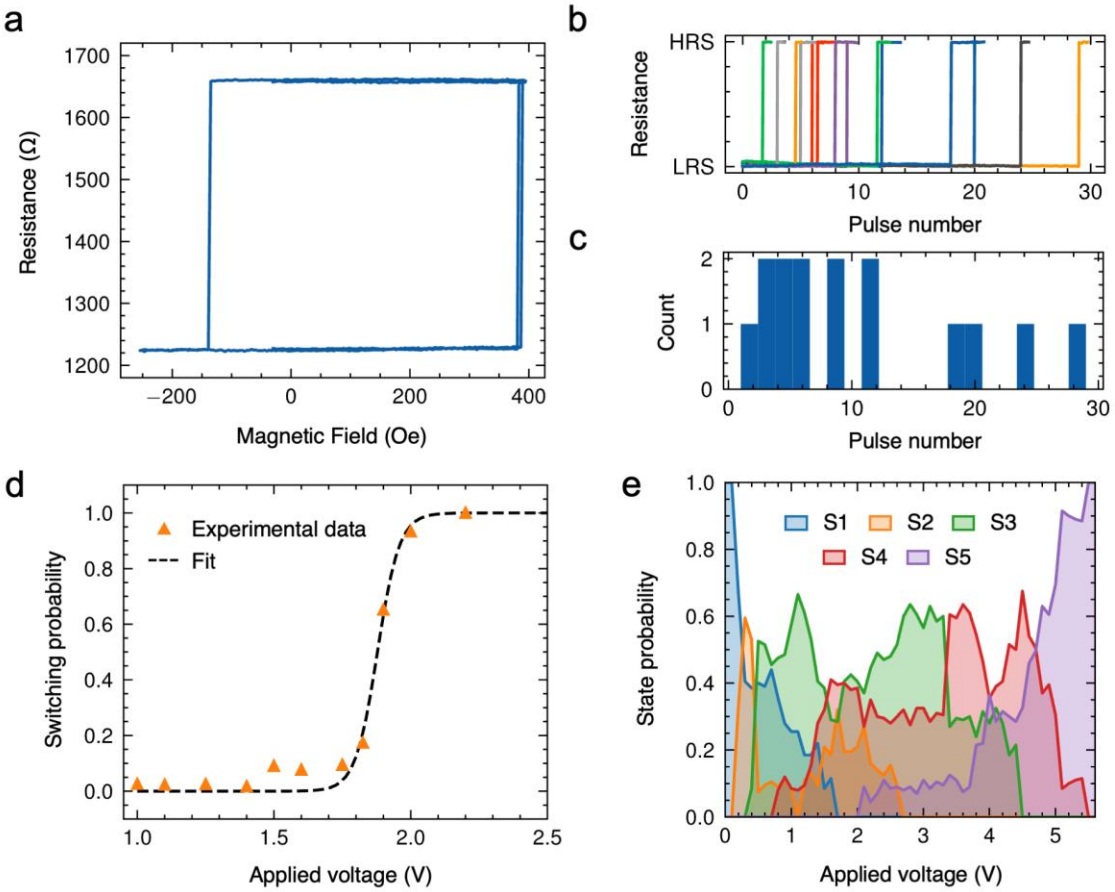


图 2. 实验结果。(a) 二元神经元 MTJ 电阻（从 IN 到 OUT 的两点电阻）随面外磁场的变化。(b) 对 CLK 至 IN 端重复施加 1.75 V 恒定幅值脉冲 15 个周期时，MTJ 电阻随脉冲次数的变化。(c) 在恒定幅值为 1.75 V 时，各脉冲次数使二元 DW-MTJ 神经元发生切换的次数。(d) 对 11 种电压幅值获得的相应 (b-c) 数据中，切换概率随施加电压的变化。(e) 引自参考文献 2 的 MW DW-MTJ 神经元的类似测量数据。该神经元具有 5 个光刻缺口，因此对于给定的施加电压脉冲幅值，其处于五种状态 S1-S5 之一的概率各不相同。

二值DW-MTJ随机神经元的数据如图2所示。图2(a)中的次磁场回线展示了高电阻反平行态和低电阻平行态。器件通过−200 Oe的外部磁场设为平行态。饱和后，施加350 Oe的偏置场以辅助SOT开关。所有实验中这些场强均保

写入脉冲所需的能量耗散可根据该分布估算。对于约 50% 的切换概率，在重金属导线上施加 1.8 V 电压，其两点电阻为 1029 Ω，器件的四点电阻为 365 Ω，表明 DW轨道上的电压降为 0.64 V。因此，估算得到每个样本的能量

图 2(e) 为我们先前对参考文献 2 中所示多权重（MW）DW-MTJ 器件的测量数据，用于与图 2(a-d) 中测量的二态 DW-MTJ 进行比较。由于本文展示的二态器件仅有两种状态，因此人工神经元在器件运行过程中无法模拟积分。

脉冲神经网络仿真

在本节中，我们展示二元随机DW-MTJ神经元如何用于构建面向边缘设备训练和推理的抗噪声脉冲神经网络。该网络架构为一个具有两个隐藏层的多层感知机（MLP），隐藏层采用随机神经元激活函数，如图 3(a) 所示。输出层为泄漏积分层，以实现精确的 softmax激活。为了调节各层的脉冲频率，在每个隐藏层之后添加了一个批归一化层。该脉冲神经网络在 Norse 框架中实现[24] ，使用自定义神经元模块，这些模块采用由二元DW-MTJ（图2(d)）和 MWDW-MTJ（图 2(e)）随机神经元采集的实验数据构成的查找表。为展示有监督训练性能，将脉冲MLP 应用于 Fashion-MNIST 服装类别分类任务[30]。Fashion-MNIST数据集采用泊松脉冲编码方案在频域中进行编码，将最亮像素的最大脉冲频率限制为网络的采样速度，该速度选定为10 MHz。网络的损失函数通过取每个输出在各时间步上的最大值来计算

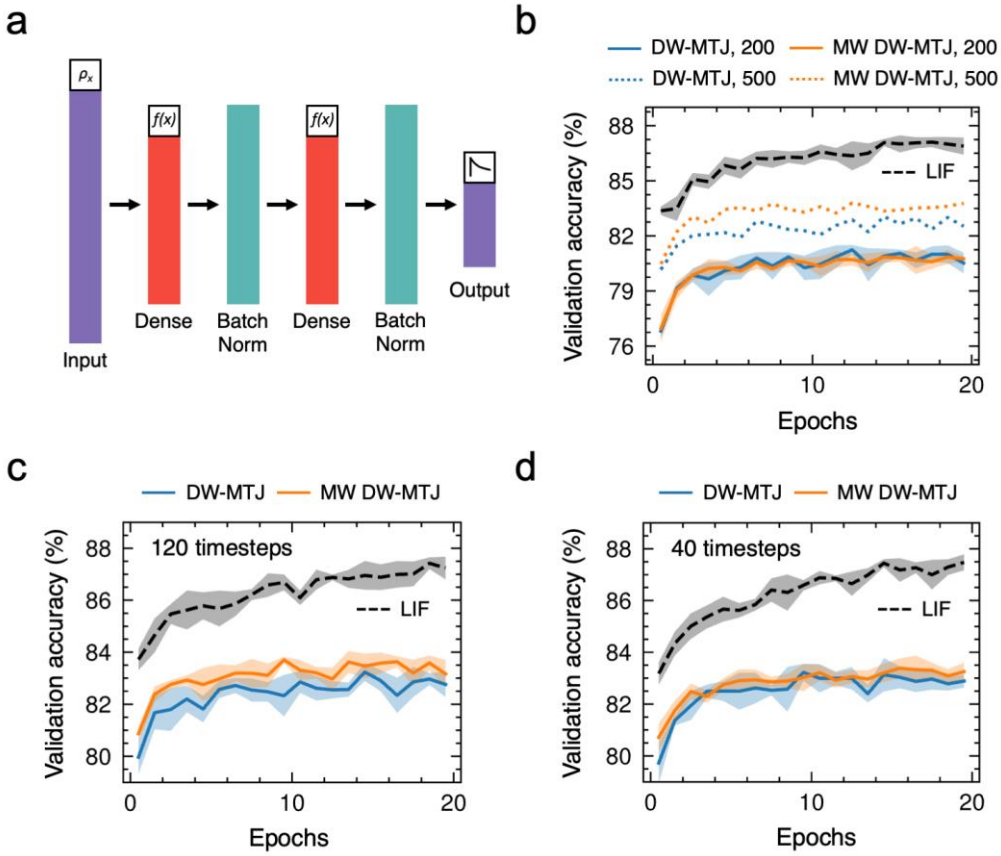


图 3. Fashion-MNIST 上的脉冲神经网络架构与训练性能。(a) 具有 2 个隐藏层的多层感知器示意图。(b) 由二值 DW-MTJ（蓝色）和多级 DW-MTJ（橙色）随机神经元构成的网络验证准确率，隐藏层大小分别为 200 个单元（实线）和 500 个单元（虚线）。同时显示 LIF 神经元 SNN 的验证准确率（黑色虚线）。阴影区域表示 5 个随机种子下准确率的标准差。显示了三种网络在每张图像训练 (c) 120 个时间步和 (d) 40 个时间步、隐藏单元数为 500 时的验证准确率。

通道，然后应用softmax 进行归一化。使用 Adam 优化器[31] 计算权重更新，使用一个

学习率为 0.001。训练时的批量大小为 100 张图像。

我们评估了采用不同激活函数的多个网络，将二元 DW-MTJ 和多级 DW-MTJ（MW DW-MTJ）随机神经元的训练性能与理想 LIF 神经元基线进行比较。图 3b 展示了由三种类型神经元组成的脉冲神经网络，每个隐藏层分别包含 200 个（虚线）和 500 个单元（实线）。网络在 80 个时间步内进行采样。基线 LIF 网络的训练性能结果来自每个隐藏层包含 500 个单元的网络，不过，将每个隐藏层的单元数减至 200 个并未显著影响验证准确率。由于 MW DW-MTJ 具有多个电平，因此我们模拟了量化的积分发放行为。这种额外的表达能力增加了激活函数的复杂度，使得每个隐藏层包含 500 个单元时，MW DW-MTJ 神经元的表现略优于更简单的二元 DW-MTJ 神经元。然而，对于每个隐藏层包含 200 个单元的网络，两者的训练性能没有显著差异。这表明，MW DW-MTJ 器件额外电平带来的表达能力并不能显著提升较小网络的训练性能。因此，对于资源受限的边缘计算而言，采用制造工艺更简单的二元 DW-MTJ 神经元仍具有优势。

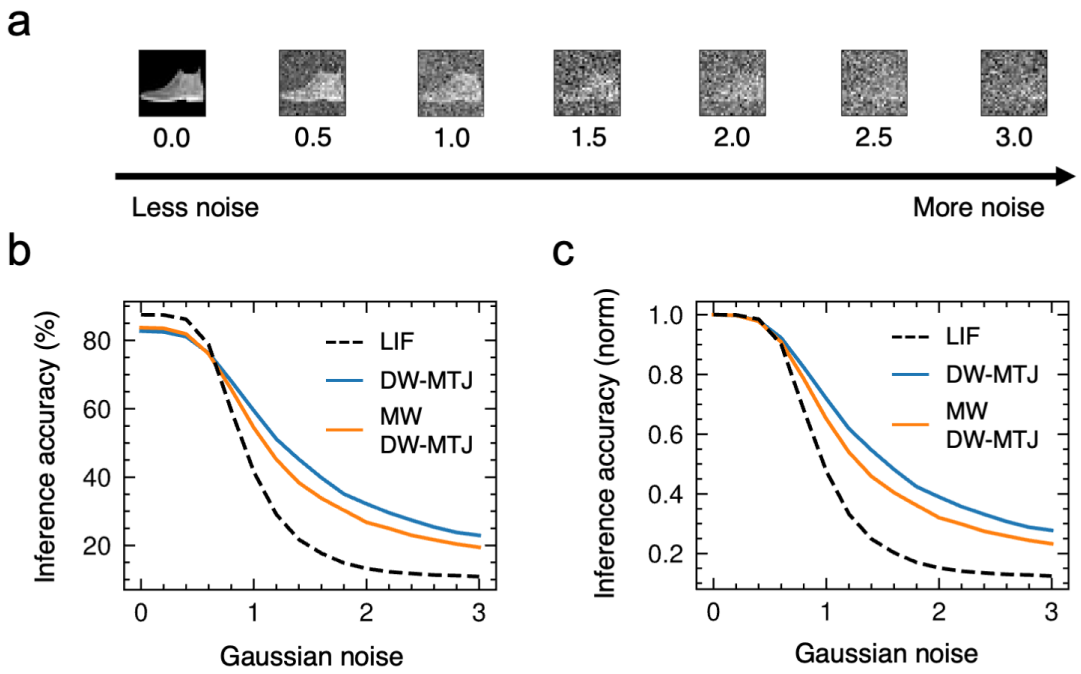


图4.在添加（a）高斯噪声的 Fashion-MNIST数据上的推理准确率。（b）原始测试准确率和（c）归一化测试准确率，针对标准差从0到3的高斯噪声。

验证准确率结果进一步证实了这一点：每层包含500 个隐藏单元的网络分别进行了120个（图 3(c)）和40个（图 3(d)）时间步的采样。在采样120个时间步的网络中，MW DW-MTJ器件网络的验证准确率略高于二值 DW-MTJ网络。然而，当网络采样 40个时间步时，两种DW-MTJ器件的验证准确率相近。因此，对于采样时间步较少的小型网络，选择二值 DW-MTJ而非 MW器件可以降低制造复杂度。

文献中众所周知，在训练阶段引入噪声后，所得网络受到含噪输入数据的影响会更小[32]。这是因为优化算法会在训练过程中找到一种能够以抗噪方式使误差最小化的网络状态。由于开关具有概率性，研究假设这两类 DW-MTJ 神经元都能够实现增强网络抗噪能力的相同功能。为对其进行表征，我们向 Fashion-MNIST 测试集中的 10,000 张图像添加了不同程度的高斯噪声，如图 4(a) 所示。随后，使用上一节中在无噪声数据集上训练的各类型准确率最高的模型，对含噪数据集进行推理。图 4(b) 的结果显示，尽管基于 LIF 的网络初始准确率高于另外两个网络，但当高斯噪声幅度达到 0.7 时，其准确率骤降至低于随机 DW-MTJ 神经元网络的水平。此时，由 MW DW-MTJ 器件构成的网络，其推理准确率也降至低于更简单的二值 DW-MTJ 器件网络。在高斯噪声幅度为 3.0 时，LIF 网络的准确率相较于随机猜测（~10%）已微不足道，而基于 DW-MTJ 器件的两个网络仍然能够完成

推理准确率高于随机猜测。当查看相对于网络在无噪声数据上的准确率进行归一化后的数据时，DW-MTJ随机神经元的优势变得更加明显（图4(c)）。在实际实现中，这种噪声鲁棒性可能不是一种可接受的准确率权衡。然而，在面向边缘应用的深度神经网络中，一种有前景的实现方式是在其中一层集成随机神经元，以引入噪声鲁棒性，同时通过其他层中的LIF神经元保持较高的分类准确率。

结论

总而言之，我们首次制备了尺寸缩小的DW-MTJ人工神经元，并对其随机开关行为进行了表征。对于二值DW-MTJ器件，所得脉冲概率随输入电压变化的分布近似呈S形，单个器件每个样本的能量耗散仅为44.9 pJ。我们还提出了一种替代性的二值化器件架构，通过设置多个缺口来实现量化积分状态。随后，将两种器件的实验数据映射到查找表中，以模拟脉冲神经网络中多层感知机的监督学习。在线学习结果表明，使用MW DW-MTJ进行训练时，由于额外缺口带来了更强的表达能力，其验证准确率略高（~1%）。然而，在网络规模较小且时间步数较少时，二值DW-MTJ器件能够达到与MW DW-MTJ器件相当的性能。在含噪数据集上进行推理时，DW-MTJ器件得益于使用随机器件训练的网络所具备的噪声鲁棒性，性能优于理想的LIF神经元。随着数据集中的噪声增加，由LIF神经元组成的网络的推理准确率迅速下降，而由两种类型DW-MTJ神经元组成的网络均能够保持合理的性能，并且高于LIF网络。这些结果表明，所提出的器件非常适合用于边缘端能效高且资源受限的神经形态计算硬件，也进一步证明了采用自旋电子权重和激活的模拟DNN加速器具有应用价值。

致谢

作者感谢美国国家科学基金会通过 CCF 项目（奖项编号 1910997）提供的研究经费支持，以及该基金会研究生研究奖学金项目通过资助编号 2020307514 和 2021311125 提供的支持。本研究在德克萨斯纳米加工设施完成，该设施由 NSF 资助（编号 NNCI- 1542159）；研究也在德克萨斯材料研究所（TMI）开展。作者感谢德克萨斯大学奥斯汀分校德克萨斯高级计算中心（TACC）提供的计算资源（http://www.tacc.utexas.edu）。

数据可用性

支持本研究发现的数据可在提出合理请求后向通讯作者获取。

参考文献

1. V. Milo, G. Malavena, C. M. Compagnoni & D. Ielmini. 用于神经形态计算的忆阻器与 CMOS 器件。《材料》13，166（2020）。 2. T. Leonard, S. Liu, M. Alamdar, H. Jin, C. Cui, O. G. Akinola, L. Xue, T. P. Xiao, J. S. Friedman, M. J. Marinella, C. H. Bennett & J. A. C. Incorvia. 用于神经形态计算的形状依赖型多权重磁性人工突触。《先进电子材料》8，2200563（2022）。 3. D. Kireev, S. Liu, H. Jin, T. Patrick Xiao, C. H. Bennett, D. Akinwande & J. A. C. Incorvia. 用于提升神经形态计算准确性的超低能耗、生物相容性石墨烯人工突触晶体管。《自然·通讯》13，4386（2022）。 4. G. Zhou, Z. Wang, B. Sun, F. Zhou, L. Sun, H. Zhao, X. Hu, X. Peng, J. Yan, H. Wang, W. Wang, J. Li, B. Yan, D. Kuang, Y. Wang, L. Wang & S. Duan. 用于神经形态计算的易失性与非易失性忆阻器件。《先进电子材料》2101127（2022）。 5. S. Kumar, X. Wang, J. P. Strachan, Y. Yang & W. D. Lu. 面向更高复杂度神经形态计算的动态忆阻器。《自然·材料评论》7，575–591（2022）。 6. D. Ielmini & H.-S. P. Wong. 利用电阻开关器件进行存内计算。《自然·电子学》1，333–343（2018）。 7. S. Liu, T. P. Xiao, J. Kwon, B. J. Debusschere, S. Agarwal, J. A. C. Incorvia & C. H. Bennett. 基于磁性隧道结的概率存内计算贝叶斯神经网络。《纳米技术前沿》4，（2022）。 8. J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami & M. D. Stiles. 神经形态自旋电子学。《自然·电子学》3，360–370（2020）。 9. Q. Yang, R. Mishra, Y. Cen, G. Shi, R. Sharma, X. Fong & H. Yang. 用于神经形态计算的随机自旋电子积分-发放-复位神经元。《纳米快报》22，8437–8444（2022）。 10. S. Yang, J. Shin, T. Kim, K. W. Moon, J. Kim, G. Jang, D. S. Hyeon, J. Yang, C. Hwang, Y. J. Jeong & J. P. Hong. 通过人工自旋突触和自旋神经元构建集成神经形态计算网络。《NPG 亚洲材料》13，（2021）。 11. C. Wang, C. Lee & K. Roy. 基于多畴自旋电子器件的抗噪声漏积分 - 发放神经元。《科学报告》 12，8361 （2022）。 12. D. Wang, R. Tang, H. Lin, L. Liu, N. Xu, Y. Sun, X. Zhao, Z. Wang, D. Wang, Z. Mai, Y. Zhou, N. Gao, C. Song, L. Zhu, T. Wu, M. Liu & G. Xing. 具有自复位和胜者全得功能、用于神经形态计算的自旋电子漏积分-发放脉冲神经元。《自然·通讯》14，1068（2023）。 13. A. Sengupta, P. Panda, P. Wijesinghe, Y. Kim & K. Roy. 磁性隧道结模拟随机皮层脉冲神经元。《科学报告》6，30039（2016）。 14. P. Jadaun, C. Cui, S. Liu & J. A. C. Incorvia. 利用具备情境感知能力的灵活神经元实现自适应认知，面向下一代人工智能。《PNAS Nexus》1，（2022）。 15. M. A. Azam, D. Bhattacharya, D. Querlioz & J. Atulasimha. 基于固定磁斯格明子的共振发放神经元。《应用物理学杂志》124，（2018）。 16. H. Vakili, J.-W. Xu, W. Zhou, M. N. Sakib, M. G. Morshed, T. Hartnett, Y. Quessab, K. Litzius, C. T. Ma, S. Ganguly, M. R. Stan, P. V. Balachandran, G. S. D. Beach, S. J. Poon, A. D. Kent & A. W. Ghosh. 斯格明子电子学——基于磁体拓扑激发的计算与存储技术。《应用物理学杂志》130，070908（2021）。 17. S. Liu, T. P. Xiao, C. Cui, J. A. C. Incorvia, C. H. Bennett & M. J. Marinella. 具有缺口几何结构的畴壁 - 磁性隧道结人工突触，可用于深度神经网络的精准高效训练。《应用物理快报》118，202405（2021）。 18. M. Alamdar, T. Leonard, C. Cui, B. P. Rimal, L. Xue, O. G. Akinola, T. Patrick Xiao, J. S. Friedman, C. H. Bennett, M. J. Marinella & J. A. C. Incorvia. 用于存内计算的畴壁-磁性隧道结自旋轨道力矩器件与电路。《应用物理快报》118，112401（2021）。 19. W. H. Brigner, X. Hu, N. Hassan, C. H. Bennett, J. A. C. Incorvia, F. Garcia-Sanchez & J. S. Friedman. 梯度各向异性诱导的磁畴壁漂移：一种人工自旋电子漏积分-发放神经元。《IEEE 探索性固态计算器件与电路期刊》5，19–24（2019）。 20. W. H. Brigner, N. Hassan, X. Hu, C. H. Bennett, F. Garcia-Sanchez, C. Cui, A. Velasquez, M. J. Marinella, J. A. C. Incorvia & J. S. Friedman. 基于形状配置激活函数的畴壁漏积分-发放神经元。《IEEE 电子器件学报》69，2353–2359（2022）。 21. N. Hassan, X. Hu, L. Jiang-Wei, W. H. Brigner, O. G. Akinola, F. Garcia-Sanchez, M. Pasquale, C. H. Bennett, J. A. C. Incorvia & J. S. Friedman. 具有侧向抑制的磁畴壁神经元。《应用物理学杂志》124，（2018）。

22. C. Cui, O. G. Akinola, N. Hassan, C. H. Bennett, M. J. Marinella, J. S. Friedman & J. A. C. Incorvia. 面向神经形态计算，在成对磁畴壁赛道中实现最大化侧向抑制。《纳米技术》31, 294001 (2020). 23. S. Liu, C. Bennett, J. Friedman, M. Marinella, D. Paydarfar & J. A. Incorvia. 面向任务自适应计算的畴壁–磁隧道结人工神经元的可控复位行为。《IEEE 磁学快报》12, 1–5 (2021). 24. C. Pehle & J. E. Pedersen. Norse——用于脉冲神经网络的深度学习库。(2021). doi:10.5281/zenodo.4422025 25. K. Garello, F. Yasin, H. Hody, S. Couet, L. Souriau, S. H. Sharifi, J. Swerts, R. Carpenter, S. Rao, W. Kim, J. Wu, K. K. V. Sethu, M. Pak, N. Jossart, D. Crotti, A. Furnemont & G. S. Kar. 用于无场开关自旋轨道转矩磁性随机存取存储器的可制造 300 毫米平台解决方案。载于《2019 年超大规模集成电路研讨会》T194–T195（IEEE，2019）。26. C. Frenkel, M. Lefebvre, J.-D. Legat & D. Bol. 一款采用 28 纳米 CMOS 工艺的 0.086-mm^2 12.7-pJ/SOP、64k 突触、256 神经元在线学习数字脉冲神经形态处理器。《IEEE 生物医学电路与系统汇刊》13, 1–1 (2018). 27. M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C.-K. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, et al. Loihi：一款具有片上学习功能的神经形态众核处理器。《IEEE Micro》38, 82–99 (2018). 28. G. S. D. Beach, M. Tsoi & J. L. Erskine. 电流诱导畴壁运动。《磁学与磁性材料杂志》320, 1272–1281 (2008). 29. K.-S. Ryu, L. Thomas, S.-H. Yang & S. Parkin. 磁畴壁处的手性自旋转矩。《自然·纳米技术》8, 527–533 (2013). 30. H. Xiao, K. Rasul & R. Vollgraf. Fashion-MNIST：用于机器学习算法基准测试的新型图像数据集。(2017). 31. D. P. Kingma & J. Ba. Adam：一种随机优化方法。(2014). 32. H. Song, M. Kim, D. Park, Y. Shin & J.-G. Lee. 利用深度神经网络从噪声标签中学习：综述。《IEEE 神经网络与学习系统汇刊》1–19 (2022).