

MRAM/PCM/RRAM/ECRAM 应用于CIM技术的优点与挑战：

- **MRAM：**

- **优点：**

- 某些种类的MRAM，如**SOT-MRAM**能将**电阻**维持在较高的水平，同时具有较小的**器件电流**，这有利于减少交叉点阵列中因字线与位线电阻的存在而引起的电压降（IR drop），使同一行或同一列的器件单元的读写尽可能不受所处位置的影响，保证器件内部的均匀性。

Cosemans S, Verhoef B, Doevenspeck J, et al. Towards 10000TOPS/W DNN inference with analog in-memory computing—a circuit blueprint, device options and requirements[C]//2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019: 22.2. 1-22.2. 4.

- 某些种类的MRAM，如**SOT-MRAM**被设计为**三端器件**，因而可实现读写操作的分离，防止读操作影响器件的电阻状态。

Doevenspeck J, Garelo K, Verhoef B, et al. SOT-MRAM based analog in-memory computing for DNN inference[C]//2020 IEEE Symposium on VLSI Technology. IEEE, 2020: 1-2.

- MRAM在状态改变时（即进行写操作时），原子发生的位移较小，因此具有**较快的编程速度**，**较低的编程能耗**与**较高的耐久性**。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- 由于MRAM**只具有2个电导状态**，因此其单个器件在编程时出现噪声/随机误差的可能性极低。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- **挑战：**

- MRAM**只具有2个电导状态**，因而只能存储1bit信息。单个器件无法实现多位计算（MLC）与模拟内存计算（AiMC），要实现AiMC中的权重存储必须使用多个器件，构成差分电导对等单元，同时也需要对模拟权重进行量化。

Doevenspeck J, Garelo K, Verhoef B, et al. SOT-MRAM based analog in-memory computing for DNN inference[C]//2020 IEEE Symposium on VLSI Technology. IEEE, 2020: 1-2.

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- 相对其它器件而言，**器件体积较大**，同时**器件的开关比（on/off ratio）较低**，但这只是劣势，并不会对器件本身在CIM中的应用产生阻碍。

Cosemans S, Verhoef B, Doevenspeck J, et al. Towards 10000TOPS/W DNN inference with analog in-memory computing—a circuit blueprint, device options and requirements[C]//2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019: 22.2. 1-22.2. 4.

- **PCM:**

- **优点:**

- 制作工艺较为简单（如使用ALD技术沉积GST材料），**花费较低**，同时具有**良好的扩展性**。

Perniola L, Sousa V, Fantini A, et al. Electrical behavior of phase-change memory cells based on GeTe[J]. IEEE Electron Device Letters, 2010, 31(5): 488-490.

Kim W, Bruce R L, Masuda T, et al. Confined PCM-based analog synaptic devices offering low resistance-drift and 1000 programmable states for deep learning[C]//2019 Symposium on VLSI Technology. IEEE, 2019: T66-T67.

- PCM器件在晶相与非晶相切换时，原子发生的位移较大，电阻状态间难以在常温下自发地进行转换，因此PCM具有**优良的保持特性**，同时PCM的**电阻状态数也相对较多**。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- **二端器件**，器件结构简单，可以进行小尺寸缩放，进而可以制作高密度的阵列。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- **挑战:**

- 传统的PCM器件在电阻权重更新时具有**较低的线性与对称性**，难以实现**矩阵向量乘法器**用于NN推理。这一挑战通过使用具有薄金属衬套的PCM（**Proj-PCM**）器件被基本解决。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- **二端器件**，读写操作共用一条路径，在读操作时可能会影响PCM的电阻状态（因为传统PCM器件电阻状态的场依赖性较强），该挑战同样通过使用Proj-PCM被基本解决。

Giannopoulos I, Sebastian A, Le Gallo M, et al. 8-bit precision in-memory multiplication with projected phase-change memory[C]//2018 IEEE International Electron Devices Meeting (IEDM). IEEE, 2018: 27.7. 1-27.7. 4.

- 由于是二端器件，因此需要增加**选择器 (selector)** 以实现对其特定器件单元的选择与权重更新（可以是FET, OTS等），这会增加架构设计的复杂程度，引入的选择器同时也会**增加器件尺寸**。

Cosemans S, Verhoef B, Doevenspeck J, et al. Towards 10000TOPS/W DNN inference with analog in-memory computing—a circuit blueprint, device options and requirements[C]//2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019: 22.2. 1-22.2. 4.

- 传统的PCM器件在多轮脉冲后会因**材料的元素偏析**或**材料中出现空隙**而导致PCM故障，因此具有**较低的耐久性/较大的电阻漂移**，该挑战同样通过使用Proj-PCM被基本解决。

Kim W, BrightSky M, Masuda T, et al. ALD-based confined PCM with a metallic liner toward unlimited endurance[C]//2016 IEEE International Electron Devices Meeting (IEDM). IEEE, 2016: 4.2. 1-4.2. 4.

- 基于PCM相对较多的电阻状态晶相与非晶相切换时较大的原子位移，PCM器件编程时具有**较大的随机性**，同时**编程能耗较大**。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- **RRAM:**

- **优点:**

- **二端器件**，制作工艺上非常简单，同时可以进行**小尺寸缩放**以及**3D堆叠**，制作**高密度阵列**。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

Esmanhotto E, Brunet L, Castellani N, et al. High-density 3D monolithically integrated multiple 1T1R multi-level-cell for neural networks[C]//2020 IEEE International Electron Devices Meeting (IEDM). IEEE, 2020: 36.5. 1-36.5. 4.

- 可以与多种器件（**选择器**）共同构成器件单元以提升阵列性能，如1T1R/1S1R/1T4R等，具有**良好的可扩展性**。

Robayo D A, Sassine G, Lopez J M, et al. Reliability and Variability of 1S1R OxRAM-OTS for High Density Crossbar Integration[C]//2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019: 35.3. 1-35.3. 4.

Hsieh E R, Giordano M, Hodson B, et al. High-density multiple bits-per-cell 1T4R RRAM array with gradual SET/RESET and its effectiveness for deep learning[C]//2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019: 35.6. 1-35.6. 4.

Esmanhotto E, Brunet L, Castellani N, et al. High-density 3D monolithically integrated multiple 1T1R multi-level-cell for neural networks[C]//2020 IEEE International Electron Devices Meeting (IEDM). IEEE, 2020: 36.5. 1-36.5. 4.

- **电阻状态数很多**，这意味着RRAM在理想状态下可以实现高精度的运算与高密度的存储。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- RRAM器件具有**良好的非易失性**，取决于其编程的物理原理，原子发生的位移较大。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

- RRAM器件具有**较快的编程速度与较低的能耗**，但这针对的是现有的最优成果，实际上RRAM器件的编程速度与能耗取决于选择的器件材料、施加的编程电压、温度等多种因素。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

Kuzum D, Yu S, Wong H S P. Synaptic electronics: materials, devices and applications[J]. Nanotechnology, 2013, 24(38): 382001.

- RRAM器件在CIM技术中的应用较为广泛，比如可用于**随机数生成器**，或利用其基于缺陷的器件原理构建**物理不可克隆函数（PUF）**用于硬件安全。

Ielmini D, Wong H S P. In-memory computing with resistive switching devices[J]. Nature electronics, 2018, 1(6): 333-343.

○ 挑战：

- RRAM器件本质上是一种基于缺陷设计的器件，因此**器件与器件间的不均匀性，器件编程的随机性**均较强，在电阻权重更新时也不具有**较好的线性与对称性**，难以实现矩阵向量乘法器（或需要增加大量外围电路）用于NN推理。

Wang Z, Wu H, Burr G W, et al. Resistive switching materials for information processing[J]. Nature Reviews Materials, 2020, 5(3): 173-195.

Kim S, Todorov T, Onen M, et al. Metal-oxide based, CMOS-compatible ECRAM for deep learning accelerator[C]//2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019: 35.7. 1-35.7. 4.

- 传统的RRAM器件在经历多次编程后可能导致氧化层的永久性击穿，这使得RRAM具有**有限的耐久性**。

Grossi A, Vianello E, Sabry M M, et al. Resistive RAM endurance: Array-level characterization and correction techniques targeting deep learning applications[J]. IEEE Transactions on Electron Devices, 2019, 66(3): 1281-1288.

- **二端器件**，与PCM一样需要选择器，且读写操作共用一条路径，可能造成的潜在影响已在前文提及。

• ECRAM：

○ 优点：

- 一般是基于**离子嵌入**原理的RAM，编程时，整个器件离子源内的离子均发生相同趋势的移动，相对来说具有**较小的随机性**。

Li Y, Fuller E J, Sugar J D, et al. Filament-Free Bulk Resistive Memory Enables Deterministic Analogue Switching[J]. Advanced Materials, 2020, 32(45): 2003984.

- 取决于ECRAM内部较多的离子数量，**电阻状态数很多**，可以实现实际意义上的模拟权重存储，进而实现**高精度的运算**，而无需如其余器件一样使用较少的电阻状态对模拟权重进行低精度量化。

Tang J, Bishop D, Kim S, et al. ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing[C]//2018 IEEE International Electron Devices Meeting (IEDM). IEEE, 2018: 13.1. 1-13.1. 4.

- 能够实现**高速编程**，同时**编程功耗较低**，在电阻权重更新时也具有**良好的线性与对称性**。

Tang J, Bishop D, Kim S, et al. ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing[C]//2018 IEEE International Electron Devices Meeting (IEDM). IEEE, 2018: 13.1. 1-13.1. 4.

Kim S, Todorov T, Onen M, et al. Metal-oxide based, CMOS-compatible ECRAM for deep learning accelerator[C]//2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019: 35.7. 1-35.7. 4.

Onen M, Emond N, Wang B, et al. Nanosecond protonic programmable resistors for analog deep learning[J]. Science, 2022, 377(6605): 539-543.

- **三端器件**，读写操作相互独立，同时可以加入FET对器件电流进行调控，使电阻权重更为精确。

Tang J, Bishop D, Kim S, et al. ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing[C]//2018 IEEE International Electron Devices Meeting (IEDM). IEEE, 2018: 13.1. 1-13.1. 4.

○ **挑战：**

- 器件尺寸较大，制作工艺较为复杂，同时由于其基于离子嵌入的编程原理，在**长脉宽电压编程**时可能导致器件退化。

Onen M, Emond N, Wang B, et al. Nanosecond protonic programmable resistors for analog deep learning[J]. Science, 2022, 377(6605): 539-543.
