Halbleitertechnologie von A bis Z

Digitaltechnik

Vergleich & Einordnung

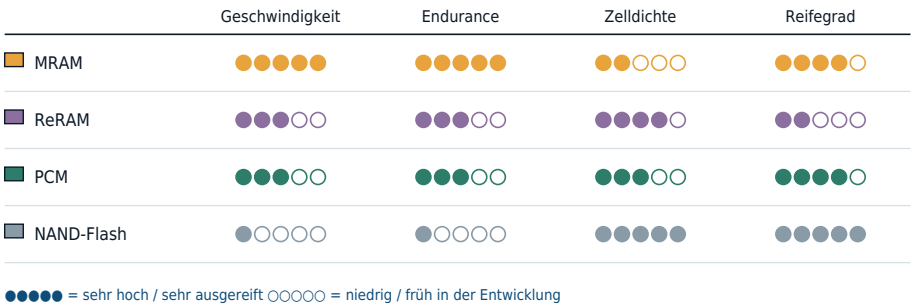
MRAM, ReRAM, PCM und Flash im direkten Vergleich

Alle drei Technologien verfolgen unterschiedliche physikalische Prinzipien – magnetische Ausrichtung bei MRAM, Filamentbildung durch Ionenmigration bei ReRAM, struktureller Phasenübergang bei PCM –, treten aber in der Praxis im direkten Wettbewerb zueinander und zum etablierten Flash-Speicher an, da sie sich in ähnlichen Anwendungsbereichen positionieren. Die wichtigsten Unterschiede im Überblick:

- Geschwindigkeit: MRAM ist am schnellsten und erreicht mit Schreib- und Lesezeiten im niedrigen Nanosekundenbereich eine Geschwindigkeit, die mit SRAM vergleichbar ist. PCM und ReRAM folgen mit Schaltzeiten im Bereich von einigen Zehn bis wenigen Hundert Nanosekunden, während NAND-Flash für Schreibvorgänge durch das blockweise Löschen deutlich langsamer ist und in den Millisekundenbereich vordringt.
- Endurance: MRAM übertrifft mit über 10^{12} Zyklen alle anderen Technologien deutlich, da der magnetische Schaltvorgang keinerlei strukturelle Materialermüdung verursacht. ReRAM und PCM liegen typischerweise im Bereich von 10^{6} bis 10^{9} Zyklen – bei PCM begrenzt durch mechanische Volumenänderungen des GST beim wiederholten Aufschmelzen, bei ReRAM durch graduelle strukturelle Veränderungen im Filamentbereich. Beide liegen damit dennoch um mehrere Größenordnungen über NAND-Flash.
- Skalierbarkeit und Zelldichte: ReRAM gilt aufgrund seiner einfachen Kreuzungsstruktur (Crossbar-Architektur) und der Möglichkeit zum mehrlagigen 3D-Stapeln als am dichtesten skalierbar, gefolgt von PCM, das ebenfalls vergleichsweise kompakte Zellen erlaubt. MRAM ist durch die komplexere MTJ-Schichtstruktur und die zusätzliche Fläche für den Zugriffstransistor in der erreichbaren Zelldichte am stärksten limitiert.
- Reifegrad: PCM (in Form von 3D XPoint/Optane, wenn auch mittlerweile eingestellt) und MRAM (als eingebetteter Speicher in kommerziellen Mikrocontrollern und SoCs) haben bereits kommerzielle Serienreife erreicht und werden in Millionenstückzahlen gefertigt. ReRAM befindet sich größtenteils noch in Entwicklung beziehungsweise wird bislang nur in spezialisierten Nischenanwendungen mit geringen Stückzahlen eingesetzt.

Bewertungsraster der vier Technologien

Speichertechnologien im Vergleich
Qualitative Einordnung von MRAM, ReRAM, PCM und NAND-Flash



Einsatzgebiete und Ausblick

Keine der drei Technologien wird Flash in absehbarer Zeit als dominante Massenspeichertechnologie ablösen – dafür ist NAND-Flash preislich pro Bit dank jahrzehntelanger Prozessoptimierung, extremer Skalierung und gewaltiger Fertigungsvolumina weiterhin uneinholbar günstig. Stattdessen etablieren sich MRAM, ReRAM und PCM zunehmend in komplementären, spezialisierten Rollen: MRAM als stromsparender, extrem schneller und praktisch unbegrenzt beschreibbarer Cache- und SRAM-Ersatz in eingebetteten Systemen, wo batteriegepuffertes SRAM bislang für Datenerhalt bei Stromausfall sorgen musste. PCM und ReRAM positionieren sich als potenzielle Bausteine für Storage-Class Memory in Rechenzentren, wo die Lücke zwischen DRAM und SSD-Speicher nach wie vor besteht.

Ein besonders zukunftsträchtiges Forschungsfeld ist das sogenannte In-Memory-Computing, bei dem Rechenoperationen – insbesondere die für neuronale Netze zentralen Matrix-Vektor-Multiplikationen – direkt im Speicher-Array ausgeführt werden, anstatt Daten erst zur separaten Recheneinheit zu transportieren. Da resistive und phasenwechselbasierte Speicherzellen ihren Zustand als analogen Widerstandswert codieren, lassen sie sich prinzipiell direkt zur gewichteten Summation von Signalen nutzen (Kirchhoffsches Stromgesetz an einer Crossbar-Struktur realisiert näherungsweise eine Matrixmultiplikation in Hardware). Dieser Ansatz verspricht, den sogenannten Von-Neumann-Flaschenhals zu umgehen – die fundamentale Limitierung klassischer Rechnerarchitekturen, bei denen Speicher und Recheneinheit physisch getrennt sind und der Datentransfer zwischen beiden einen Großteil des Energieverbrauchs moderner KI-Beschleuniger verursacht.