

Halbleitertechnologie von A bis Z

Digitaltechnik

Warum neue Speichertechnologien?	3
<i>Die Grenzen von Flash</i>	3
<i>Die Lücke zwischen Arbeitsspeicher und Massenspeicher</i>	3

Warum neue Speichertechnologien?

Die Grenzen von Flash

Flash-Speicher (siehe Kapitel zur Flash-Speicherzelle) dominiert seit Jahrzehnten den nichtflüchtigen Speichermarkt, stößt aber zunehmend an technische Grenzen. Mit fortschreitender Skalierung der Floating-Gate-Zelle sinkt die Anzahl der gespeicherten Elektronen pro Zelle immer weiter – bei aktuellen Strukturgrößen sind es teils nur noch wenige Dutzend Elektronen, die über „0“ oder „1“ entscheiden. Das macht die Zelle anfälliger für Störeinflüsse benachbarter Zellen (Zellinterferenz durch parasitäre Kapazitäten), für Lese-Störungen (Read Disturb) und für Ladungsverlust über die Zeit (Charge Loss), da bereits der Verlust eines einzelnen Elektrons das Ausleseergebnis verfälschen kann. Bei Multi-Level-Cell- und Triple-Level-Cell-Verfahren (siehe Kapitel zur Flash-Speicherzelle), bei denen mehrere Bits über feine Abstufungen der Schwellspannung in einer Zelle codiert werden, verschärft sich dieses Problem zusätzlich, da die Spannungsfenster zwischen den einzelnen Ladungszuständen immer schmaler werden.

Hinzu kommt die begrenzte Endurance: Jeder Schreib-/Löschzyklus treibt Elektronen durch das dünne Tunneloxid der Zelle (Fowler-Nordheim-Tunneln bzw. Hot-Electron-Injection), was das Oxid graduell mit Fangstellen (Traps) durchsetzt und mechanisch beansprucht. Je nach Zelltyp übersteht NAND-Flash dadurch nur zwischen wenigen Hundert Zyklen (QLC, vier Bit pro Zelle) und rund 100.000 Zyklen (SLC, ein Bit pro Zelle), bevor die Fehlerrate der Zelle nicht mehr durch Fehlerkorrekturverfahren (ECC) kompensiert werden kann. Auch die Schreibgeschwindigkeit ist grundsätzlich limitiert: Anders als beim byteweisen Schreiben erfolgt das Löschen bei NAND-Flash blockweise (typischerweise mehrere Hundert Kilobyte auf einmal) und dauert mit teils mehreren Millisekunden deutlich länger als der eigentliche Schreibvorgang einer Seite.

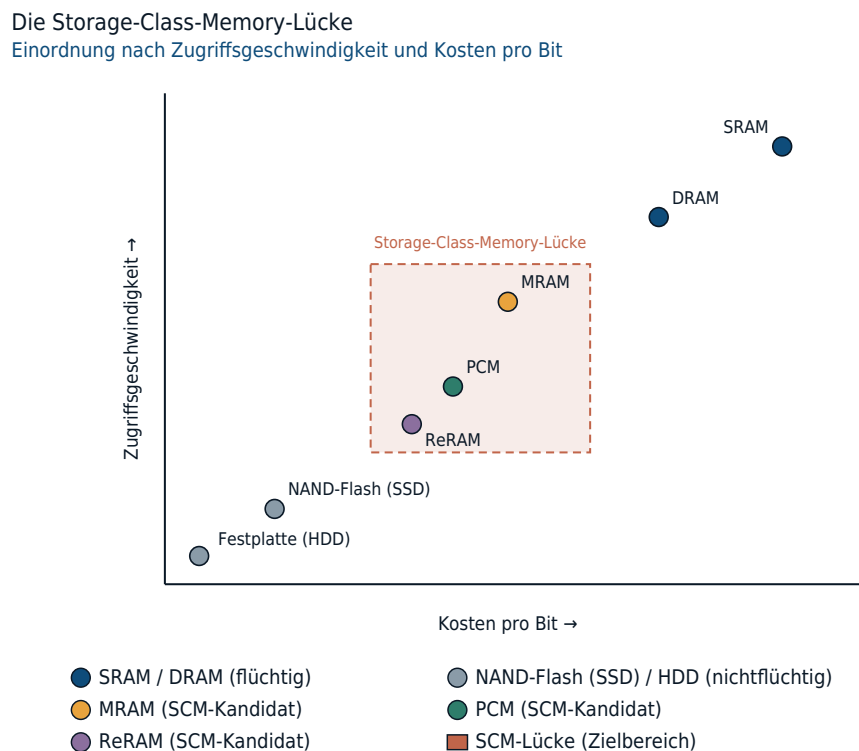
Die Lücke zwischen Arbeitsspeicher und Massenspeicher

Klassische Speicherarchitekturen kennen einen scharfen Bruch in der Speicherhierarchie. DRAM (siehe DRAM-Zelle-Kapitel) ist extrem schnell – Zugriffszeiten liegen im Bereich weniger Nanosekunden – aber flüchtig (der Inhalt geht ohne Refresh-Zyklus und letztlich ohne Stromversorgung verloren) und mit vergleichsweise hohen Kosten pro gespeichertem Bit verbunden, da jede Zelle einen eigenen Kondensator und Zugriffstransistor benötigt. Flash-basierte SSDs sind demgegenüber nichtflüchtig und deutlich günstiger pro Bit,

dafür aber um mehrere Größenordnungen langsamer (Mikrosekunden bis Millisekunden statt Nanosekunden) und in der Anzahl möglicher Schreibzyklen begrenzt.

Dazwischen klafft eine Lücke, die in der Fachliteratur als Storage-Class Memory (SCM) bezeichnet wird: ein Speichertyp, der die Geschwindigkeit von DRAM mit der Nichtflüchtigkeit und den Bitkosten von Flash verbindet. Ein solcher Speicher könnte in der Systemarchitektur direkt am Speicherbus betrieben werden, anstatt – wie SSDs heute – über einen vergleichsweise langsamen I/O-Controller (etwa NVMe über PCIe) angebunden zu sein, und würde damit ganze Softwareschichten der Betriebssystem- und Anwendungslogik überflüssig machen, die heute allein dazu dienen, den Geschwindigkeitsunterschied zwischen Arbeits- und Massenspeicher zu überbrücken (Caching, Paging, Write-Back-Strategien).

Einordnung nach Geschwindigkeit und Kosten pro Bit



Die in diesem Kapitel behandelten Technologien – MRAM, ReRAM und PCM – verfolgen jeweils grundlegend unterschiedliche physikalische Speicherprinzipien, um genau diese Lücke zu schließen: magnetische Zustände bei MRAM, resistive Filamente bei ReRAM, kristalline Phasenzustände bei PCM. Keine davon hat Flash als dominante Massenspeichertechnologie bislang

verdrängt, und angesichts der enormen Skalen- und Kostenvorteile, die NAND-Flash durch jahrzehntelange Prozessoptimierung und Massenproduktion erreicht hat, ist eine vollständige Ablösung auch mittelfristig unwahrscheinlich. Alle drei Technologien haben jedoch in spezialisierten Nischen bereits kommerzielle Relevanz erreicht: MRAM etwa als Cache-Ersatz und in Automotive- sowie Industrieanwendungen, PCM in Form von Intel Optane als schnelle Zwischenschicht vor NAND-SSDs, ReRAM vor allem in eingebetteten Anwendungen mit geringem Leistungsbedarf.