

Speichertechnologien



Mario Putz

IT 233

Inhalt

Speichertechnologien	3
Flüchtiger Speicher	3
Permanente Speicher	3
Semi- Permanente Speicher	3
Flüchtiger Speicher (Volatile Memory)	3
Nicht flüchtiger Speicher (Non-Volatile Memory)	3
RAM (Random Access Memory)	4
ROM (Read-Only Memory)	4
EPROM (Erasable Programmable Read-Only Memory)	4
EEPROM (Electrically Erasable Programmable Read-Only Memory)	5
Flash-Speicher	5
SSD (Solid State Drive)	5
HDD (Hard Disk Drive)	5
Aufbau einer HDD	6
Arbeitsspeicher	7
Module	7
SIMM	7
RIMM	7
SDR-DIMM	7
DDR-DIMM	7
SO-SIMM	8
SO-RIMM	8
SO- DIMM	8
Micro DIMM	8
Technologien	9
DDR-SDRAM	9
DDR-SDRAM	9
DDR2-SDRAM	9
DDR3-SDRAM	10
DDR4-SDRAM	10
DDR5-SDRAM	10
Cache	11
Arten von Cache Memory	11
Der L1-Cache	11
Der L2-Cache	11
Der L3 Cache	11

Speichertechnologien

Generell unterscheidet man zwischen 3 verschiedenen Speichertechnologien

Flüchtiger Speicher

- Benötigt Strom, um die gespeicherten Daten zu erhalten
- DRAM
- SRAM

Permanente Speicher

- Informationen werden permanent und unveränderbar gespeichert
- ROM
- PROM

Semi- Permanente Speicher

- Informationen werden permanent und veränderbar gespeichert
- EPROM
- EEPROM
- Flash- EEPROM
- FRAM
- MRAM
- Phase-change RAM
- SSD
- HDD

Flüchtiger Speicher (Volatile Memory)

Ein flüchtiger Speicher, oder Volatile Memory, ist eine Art von Speicher, der seine Daten nur behält, solange das Gerät mit Strom versorgt wird. Wenn die Stromversorgung aus irgendeinem Grund unterbrochen wird, gehen die Daten verloren. In einem Computer wird das volatile Memory in der Regel für den Arbeitsspeicher (RAM) des Systems verwendet, sowohl für den Hauptspeicher als auch für den L1-, L2- und L3-Cache des Prozessors.

Nicht flüchtiger Speicher (Non-Volatile Memory)

Ein nichtflüchtiger Speicher, Non-Volatile Memory (NVM), ist ein programmierbarer Schaltkreis, der seine einmal eingegebenen Daten permanent und ohne zusätzliche Energiezufuhr speichert. Die Daten können gelöscht und erneut gespeichert werden. Nichtflüchtige Speicher unterscheiden sich in ihrer Speicherkapazität, dem Speicherverfahren, den Zugriffszeiten und der Integrationsdichte.

RAM (Random Access Memory)

Der RAM ist eine Komponente welche Daten kurzfristig für einen schnelleren Zugriff speichern.

Er arbeitet gemeinsam mit der CPU und dem Speicherlaufwerk (SSD, HDD, ...) um auf Daten zuzugreifen um diese wiederum schneller bearbeiten bzw. ausführen zu können.

Unterschieden wird beim RAM zwischen DRAM (Dynamisches RAM) und SRAM (Statisches RAM), am häufigsten wird DRAM verwendet. Der dynamische Teil kommt von den Daten, die ständig aktualisiert werden müssen. SRAM ist schneller, aber auch teurer.

ROM (Read-Only Memory)

Ein ROM-Speicher ist ein Datenspeicher, der im normalen Betrieb nur lesbar und nicht beschreibbar ist. Er behält im stromlosen Zustand seine Daten. In Computersystemen ist beispielsweise das BIOS für Grundfunktionen zum Starten des Rechners im ROM abgelegt. Es existieren verschiedene Arten von ROM-Speichern, die sich unter anderem in der Art der Programmierung unterscheiden. Erste ROM-Speicher konnten nur zum Fertigungszeitpunkt programmiert werden und waren anschließend unveränderbar. Mittlerweile existieren verschiedene Arten von programmierbaren ROM-Speichern wie PROMs, EPROMs oder EEPROMs.

EPROM (Erasable Programmable Read-Only Memory)

EPROM ist ein löschbarer programmierbarer Lesespeicher. EPROM ist eine Art nichtflüchtiger ROM-Speicher, der Daten nach dem Ausschalten speichert. EPROM-ICs werden in Schaltkreisen verwendet, um Speicherfähigkeiten hinzuzufügen. Ein EPROM enthält eine Matrix aus Speicherzellen, in denen jeweils ein Transistor ein Bit repräsentiert. Eine Speicherzelle besteht aus einem MOSFET-Transistor mit einer zusätzlichen Gate Elektrode zwischen Gate und Kanal, die jedoch keinen Anschluss besitzt. Es kann daher frei ein Potential annehmen und wird deshalb Floating Gate genannt.

Zum Löschen wird üblicherweise kurzwellige Ultraviolettstrahlung verwendet. Dadurch werden durch den äußeren photoelektrischen Effekt Fotoelektronen angeregt, die ausreichende Energie haben, die Isolierbarriere zu überwinden – die Floating Gates werden entladen. Das Bitmuster ist dadurch gelöscht und das EPROM in seinen ursprünglichen Zustand zurückversetzt.

EEPROM (Electrically Erasable Programmable Read-Only Memory)

Ein EEPROM ist ein nichtflüchtiger, elektronischer Speicherbaustein, dessen gespeicherte Information elektrisch gelöscht werden kann. Er ist verwandt mit anderen löschbaren Speichern, wie dem durch UV-Licht löschbaren EPROMs und dem ebenfalls elektrisch löschbaren Flash-Speicher. Er wird verwendet zur Speicherung kleinerer Datenmengen in elektrischen Geräten, bei denen die Information auch ohne anliegende Versorgungsspannung erhalten bleiben muss oder bei denen einzelne Speicherelemente bzw. Datenworte einfach zu ändern sein müssen. Zur Speicherung größerer Datenmengen wie z. B. dem BIOS in PC-Systemen sind meist Flash-Speicher ökonomischer.

Flash-Speicher

Flash-Speicher sind digitale Speicherbausteine für eine nicht flüchtige Speicherung ohne Erhaltungsenergieverbrauch. Die genaue Bezeichnung dieses Speichertyps lautet Flash-EEPROM. Im Gegensatz zu gewöhnlichem EEPROM-Speicher lassen sich hier Bytes nicht einzeln löschen oder überschreiben.

SSD (Solid State Drive)

Eine SSD ist ein nichtflüchtiger Datenspeicher, welche Großteils die klassischen HDD-Festplatten ersetzen. Die Bauform und die elektrischen Anschlüsse können den Normen für Laufwerke mit magnetischen oder optischen Speichermedien entsprechen, müssen dies aber nicht. Sie können zum Beispiel auch als PCIe-Steckkarte ausgeführt sein.

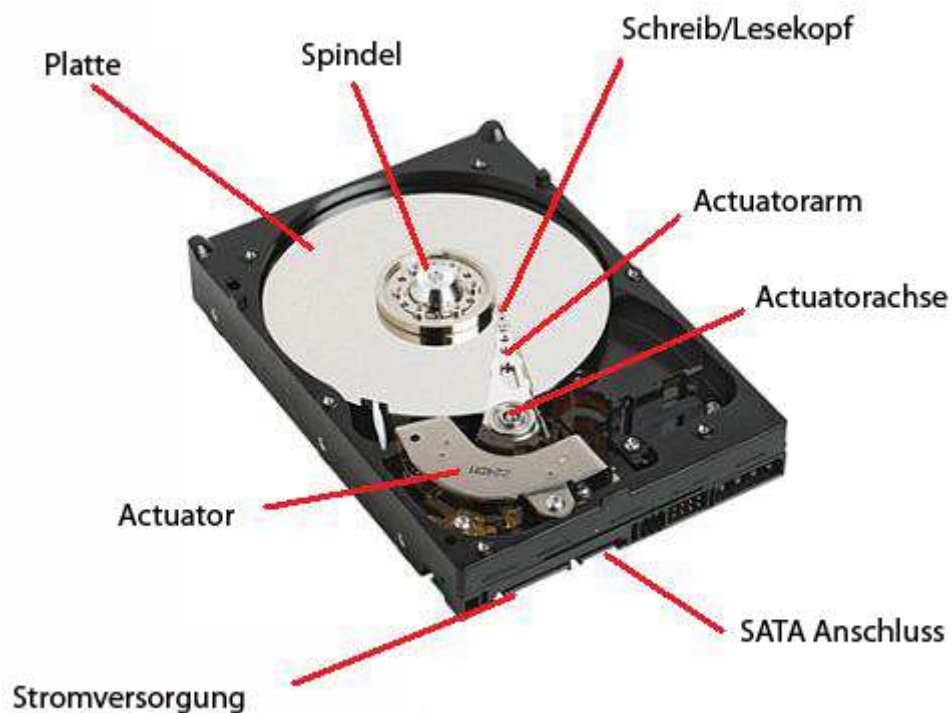
SSDs funktionieren auf dem Prinzip des Flash-Speichers, außerdem haben sie keine beweglichen Teile und sind daher unempfindlich gegen Stöße, Erschütterungen und Vibrationen. Sie haben außerdem kürzere Zugriffszeiten und arbeiten geräuschlos.

HDD (Hard Disk Drive)

Eine HDD auch oft als Festplatte bezeichnet, ist ein magnetisches Speichermedium, bei welchem Daten auf die Oberfläche rotierender Scheiben (Platter) geschrieben werden. Zum Schreiben wird die hartmagnetische Beschichtung der Scheibenoberfläche entsprechend der aufzuzeichnenden Information berührungslos magnetisiert. Durch die verbleibende Magnetisierung erfolgt das Speichern der Information. Das Lesen der Information erfolgt durch berührungsloses Abtasten der Magnetisierung der Plattenoberfläche.

Aufbau einer HDD

- Platten - magnetischen Scheiben (auch Platten/Platter)
- Spindel - eine Welle auf der die Platten übereinander montiert werden
- Elektromotor - für den Antrieb der Platte(n)
- Schreib/Lesekopf - "schreibt" durch Magnetisierung kleinster Bereiche Daten auf die Platten
- Actuatorarm - "Schwinge" an dem der Schreib/Lesekopf befestigt ist
- Actuator - Motor der über die Actuatorachse den Schreib/Lesekopf in Position über der Platte bringt
- SATA-Anschluss - Schnittstelle für den Datenaustausch
- Stromanschluss - zur Stromversorgung

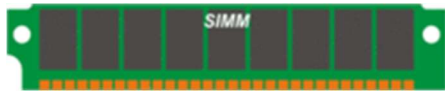


Arbeitsspeicher

Module

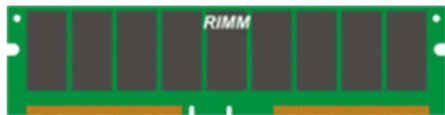
SIMM

Single Inline Memory Module



RIMM

Rambus Inline Memory Module



SDR-DIMM

Synchronous Dynamic Random Access Memory- Dual Inline Memory Module



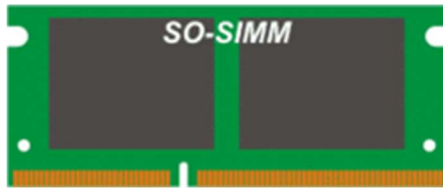
DDR-DIMM

Double Data Rate- Dual Inline Memory Module



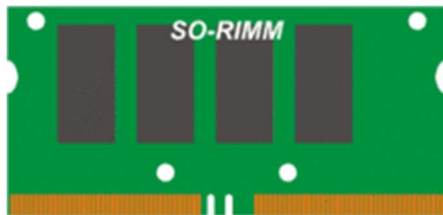
SO-SIMM

Small Outline- Single Inline Memory



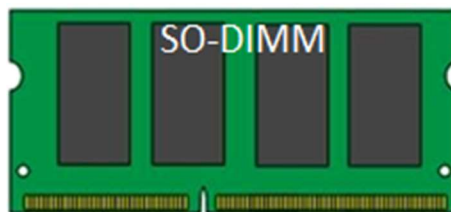
SO-RIMM

Small Outline- Rambus Inline Memory



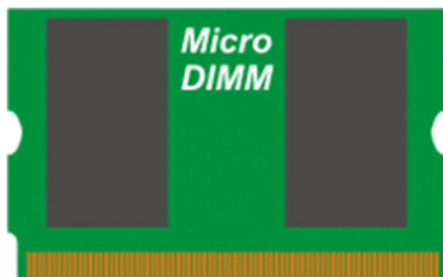
SO- DIMM

Small Outline- Dual Inline Memory Module



Micro DIMM

Micro Dual Inline Memory Module



Technologien

DDR-SDRAM und die Varianten DDR2, DDR3, DDR4 und DDR5 sind Halbleiterspeicher für beliebige Computersysteme. DDR-SDRAM wird als Arbeitsspeicher bzw. Hauptspeicher verwendet. Dieser Halbleiterspeicher wird nicht nur in Computern, sondern auch in Kraftfahrzeugen, Netzwerken, Kommunikationstechnik, medizinischen Apparaten und in der Unterhaltungselektronik eingesetzt. Damit sich die Rechenleistung vieler Prozessorkerne entfalten kann, müssen Daten schnell genug fließen. Das führt dazu, dass die Datentransferrate des Arbeitsspeichers und die Anzahl der Kanäle steigen muss.

DDR-SDRAM entspricht dem normalen SDRAM, jedoch mit einer kleinen Modifikation: Bei der Übertragung der Daten wird nicht nur die ansteigende Flanke, sondern auch die abfallende Flanke des Taktsignals zur Datenübertragung genutzt. In der Praxis entspricht das einer Taktverdopplung. Rein rechnerisch entsteht so eine Verdopplung der Übertragungsrate.

Double Data Rate bedeutet, dass pro Übertragungszyklus (eine Taktrate) zweimal bzw. die doppelte Menge an Daten übertragen werden. Damit die beteiligten Komponenten des Speicherbusses synchron arbeiten, orientieren sie sich am Speichertakt. Ein Speicherbus, der mit 100 MHz arbeitet würde bei DDR rein rechnerisch mit 200 MHz arbeiten.

DDR2-SDRAM wurde entwickelt, um den Energiebedarf zu reduzieren und die Signalqualität und damit die Systemstabilität zu verbessern. Beides wird durch die Senkung der Frequenz, beispielsweise von 200 MHz auf 100 MHz, und der Spannung, von 2,4 V auf 1,8 V, erreicht.

Das Senken der Taktfrequenz war wegen der schlechten Signalqualität und dem zunehmenden Rauschen notwendig. Der Zeitraum, in dem ein Datensignal als 1 oder 0 erkannt werden kann, reduziert sich bei hoher Frequenz deutlich. Das Signal ist kürzer und deshalb die Erkennung anfälliger für Fehler.

DDR2 überträgt die Daten genauso wie DDR1 mit steigender und fallender Taktflanke. Die DDR2-Technik nennt sich QDR und überträgt vier Datenworte pro Takt. Durch ein 4-Bit-Prefetching liefert der interne Speicher vier, anstatt zwei Bit pro Taktschritt an ein Ein-/Ausgabe-Puffer. Somit bleibt die maximale Bandbreite von DDR400 und DDR2-400 mit 3,2 GBit/s gleich.

DDR3-SDRAM wurde entwickelt um der Nachfrage nach einem schnellen Speicher der wenig Strom verbraucht gerecht zu werden. Die Speicherchips werden mit einer Spannung von 1,5 V betrieben. Dadurch wird Verlustleistung reduziert. Und die Speicherchips eignen sich noch besser für höhere Taktraten.

Die Latenzzeiten der Zugriffe sind bei DDR3 bei gleichem Takt etwas höher als bei DDR2. Höhere Latenzen ermöglichen höhere Taktraten. Durch einen höheren Takt werden die höheren Latenzen wieder ausgeglichen.

Unterscheiden muss man auch bei DDR3 die Taktfrequenz des Speicher-Interfaces und die Speicher-interne Taktrate. Im Speicher beträgt die Taktfrequenz nur ein Viertel des nominellen Takts. Um die Daten trotzdem für die hohe Bandbreite aus den Speicherzellen lesen zu können, sind die Speicherzellen von DDR3-SDRAM gegenüber DDR1-SDRAM mit einem vierfach so breiten Interface angebunden. Nur ein Bruchteil eines einzelnen Speicherchips wird als Speicher verwendet. Der Großteil sind I/O-Einheiten.

DDR4-SDRAM-Speichermodule haben verschiedene Verbesserungen gegenüber DDR3-SDRAM. Dazu zählen höhere Taktfrequenzen, verbesserte DRAM-Chips und Speicher-Controller. Die Taktfrequenzen sind auf 800, 933, 1.066 und 1.200 MHz festgelegt. Das bedeutet, es gibt DDR4-1600, -1866, -2133 und -2400-Chips. Die Latenzparameter reichen von 10-10-10 bis 18-18-18. Höhere Datentransferraten und eine niedrigere Leistungsaufnahme sind das Ziel. Die Speichermodule (DIMM) laufen mit 1,2 Volt und haben 288 (Desktop und Server) bzw. 256 (Notebooks) Pins. Um höhere Taktfrequenzen zu erreichen bringt DDR4 ein paar technische Neuerungen.

Zu den Neuerungen gehört das Übertragungsverfahren POD12 (Pseudo Open-Drain Interface mit 1,2 Volt Nominalspannung). Die Referenzspannungen zur Unterscheidung von High- und Low-Pegeln sind nicht mehr festgelegt, sondern die handeln die Chips untereinander aus. Damit kompensieren sie äußere Einflüsse durch wechselnde Temperaturen, unterschiedlich lange Busleitungen und unterschiedliche SDRAM-Chips.

DDR5-SDRAM schafft die Voraussetzungen, um die Datentransferrate pro Speichermodul um fast 80 Prozent im Vergleich zu DDR4 zu steigern. Beispielsweise von 23,5 GByte/s bei DDR4-2933 auf 41,6 GByte/s bei DDR5-5200.

Um die Leistungsaufnahme und Abwärme in Grenzen zu halten, reduziert sich die Spannung von 1,2 auf 1,1 Volt. Mehrere Maßnahmen verbessern Signalübertragung und Fehlerkorrektur.

Cache

Der Cache-Speicher wird manchmal auch als CPU-Speicher bezeichnet, da er normalerweise direkt in den CPU-Chip integriert oder auf einem separaten Chip platziert ist, der über eine separate Busverbindung mit der CPU verfügt. Daher ist er für den Prozessor leichter zugänglich und kann die Effizienz steigern, da er sich physisch in der Nähe des Prozessors befindet.

Um in der Nähe des Prozessors zu sein, muss der Cache-Speicher viel kleiner als das Main Memory sein. Folglich hat er weniger Speicherplatz. Außerdem ist er teurer als der Hauptarbeitsspeicher, da es sich um einen komplexeren Chip handelt, der eine höhere Leistung erbringt.

Was er an Größe und Preis verlangt, macht er an Geschwindigkeit wieder wett. Cache-Speicher arbeitet 10 bis 100 Mal schneller als RAM und benötigt nur wenige Nanosekunden, um auf eine CPU-Anforderung zu reagieren.

Arten von Cache Memory

Cache-Speicher ist schnell und teuer. Traditionell wird er in „Ebenen“ (Levels) kategorisiert, die seine Nähe und Zugänglichkeit zum Mikroprozessor beschreiben. Es gibt drei allgemeine Cache-Ebenen:

Der L1-Cache oder primäre Cache ist extrem schnell, aber relativ klein und wird normalerweise als CPU-Cache in den Prozessor-Chip eingebettet.

Der L2-Cache oder sekundäre Cache ist oft umfangreicher als der L1-Cache. Der L2-Cache kann in die CPU eingebettet sein, oder er kann sich auf einem separaten Chip oder Ko-Prozessor befinden und über einen alternativen Hochgeschwindigkeits-Systembus verfügen, der den Cache und die CPU verbindet. Auf diese Weise wird er nicht durch den Verkehr auf dem Hauptsystembus verlangsamt.

Der L3 Cache ist ein spezialisierter Arbeitsspeicher, der entwickelt wurde, um die Leistung von L1 und L2 zu verbessern. L1 oder L2 können wesentlich schneller sein als L3, obwohl L3 normalerweise doppelt so schnell wie DRAM ist. Bei Mehrkernprozessoren kann jeder Kern über einen dedizierten L1- und L2-Cache verfügen, aber sie können sich einen L3-Cache teilen. Wenn ein L3-Cache auf eine Anweisung verweist, wird er normalerweise auf eine höhere Cache-Ebene angehoben.

In der Vergangenheit wurden L1-, L2- und L3-Caches mit kombinierten Prozessor- und Mainboard-Komponenten erstellt. In letzter Zeit geht der Trend dahin, alle drei Ebenen der Memory-Caches auf der CPU selbst zu konsolidieren.