

Technische Informatik

Speicher

Thorsten Thormählen

15. Dezember 2022

Teil 8, Kapitel 1

Dies ist die Druck-Ansicht.

Aktiviere Präsentationsansicht

Steuerungstasten

- nächste Folie (auch Enter oder Spacebar).
- ← vorherige Folie
- d schaltet das Zeichnen auf Folien ein/aus
- p wechselt zwischen Druck- und Präsentationsansicht
- CTRL + vergrößert die Folien
- CTRL - verkleinert die Folien
- CTRL 0 setzt die Größenänderung zurück

Notation

Typ	Schriftart	Beispiele
Variablen (Skalare)	kursiv	a, b, x, y
Funktionen	aufrecht	$f, g(x), \max(x)$
Vektoren	fett, Elemente zeilenweise	$\mathbf{a}, \mathbf{b} = \begin{pmatrix} x \\ y \end{pmatrix} = (x, y)^\top,$ $\mathbf{B} = (x, y, z)^\top$
Matrizen	Schreibmaschine	$\mathbf{A}, \mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$
Mengen	kalligrafisch	$\mathcal{A}, \mathcal{B} = \{a, b\}, b \in \mathcal{B}$
Zahlenbereiche, Koordinatenräume	doppelt gestrichen	$\mathbb{N}, \mathbb{Z}, \mathbb{R}^2, \mathbb{R}^3$

Inhalt

Speicher

- Speicherhierarchie

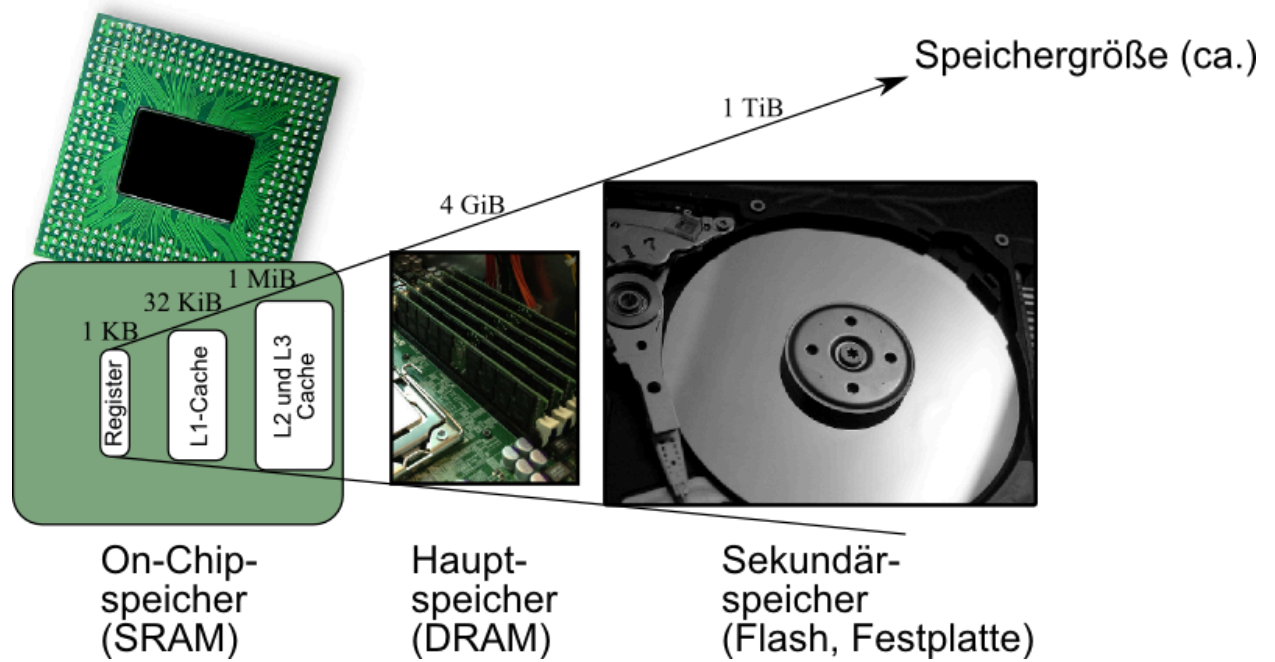
- SRAM

- DRAM

- Flash-Speicher

- Festplattenlaufwerk

Speicherhierarchie



[Bildquelle: Zusammenstellung aus: [CPU](#), Flickr user: VIA Gallery; [DRAM](#), Flickr user: Keith McDuffee; [HDD](#), Flickr user: ingenero.creativo; [Creative Commons License](#)]

Thorsten Thormählen 5 / 32

Speicherhierarchie

Schneller Speicher (d.h. Speicher mit geringer Zugriffszeit) hat hohen Aufwand pro Bit und damit einen hohen Preis pro Bit

In einem typischen Computersystem gilt: Je schneller der Speicher, desto weniger steht davon zur Verfügung

Um eine insgesamt möglichst schnelle Verarbeitung zu erreichen, sollten häufige Speicherzugriffe immer auf möglichst schnellem Speicher ausgeführt werden:

Speicherhierarchie:

Operand oder Ergebnis einer Rechenoperation → Register

Daten werden wahrscheinlich bald wieder verwendet → Cache

Daten werden gerade bearbeitet → Hauptspeicher

Daten sollen nicht-flüchtig gespeichert werden → Sekundärspeicher

Zugriffszeiten

Mittlere Zugriffszeiten flüchtiger Speicher

Register: 0,25 bis 0,5 ns

Cache (SRAM): 0,5 bis 5 ns

Hauptspeicher (DRAM): 10 bis 50 ns

Mittlere Zugriffszeiten nicht-flüchtiger Speicher

Flash-Speicher: 10 bis 250 μ s

Festplatten: 3 bis 20 ms

Diese Zeiten sind nur ungefähre Größenordnungen und werden ständig geringer
Ebenfalls kommt es teilweise stark auf die Art des Zugriffs an. Konsekutiver Zugriff ("best access") auf Speicheradressen ist häufig schneller:

Cache (SRAM) 40 - 700 GiB/s

Hauptspeicher (DRAM): 10 GiB/s

Flash-Speicher: 2 GiB/s

Festplatten: 0.3 GiB/s

[Quelle: [Wikipedia: Solid-State-Drive](#)]

Thorsten Thormählen 7 / 32

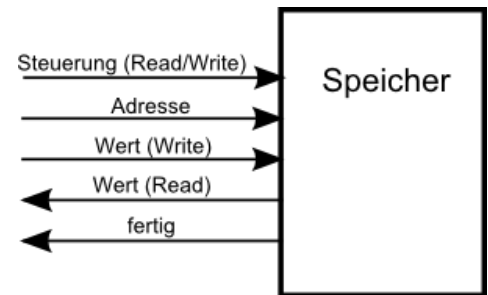
Abstrakte Sicht auf Speicher

Zuordnung: Name \rightarrow Wert

Die Namen entsprechen in der Regel Byteadressen

Werte besitzen häufig eine Größe von 2^n Bytes

(z.B. 4 oder 8 Byte für die Gleitkommadarstellung einer Zahl)



Folge von Lese- und Schreibzugriffen

Schreiben (engl. Write) bindet einen Wert an eine Adresse

Lesen (engl. Read) einer Adresse liefert den zuletzt an diese Adresse gebundenen Wert zurück

SRAM

SRAM (Static Random Access Memory) wird häufig zur Realisierung von schnellem Cache-Speicher verwendet

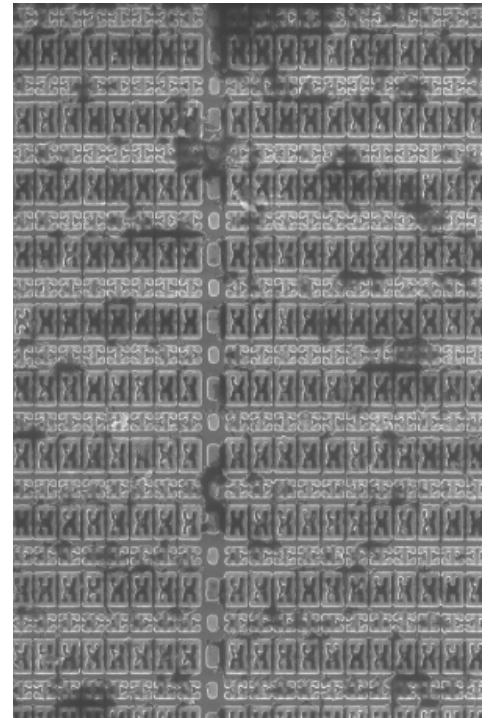
Es besteht im Prinzip aus vielen Flipflops

Im Gegensatz zu einem einzelnen Register, benötigt ein SRAM zusätzlich einer Dekodierlogik, die die Adresse in Steuersignale umsetzt, damit die richtigen Flipflops angesprochen werden

Auf der folgenden Folie wird eine mögliche Realisierung eines 4×4 -Bit SRAMs mit Hilfe von Logikgattern gezeigt

In der Realität kann eine Speicherzelle z.B. bereits aus einer Schaltung von 6 Transistoren gebildet werden

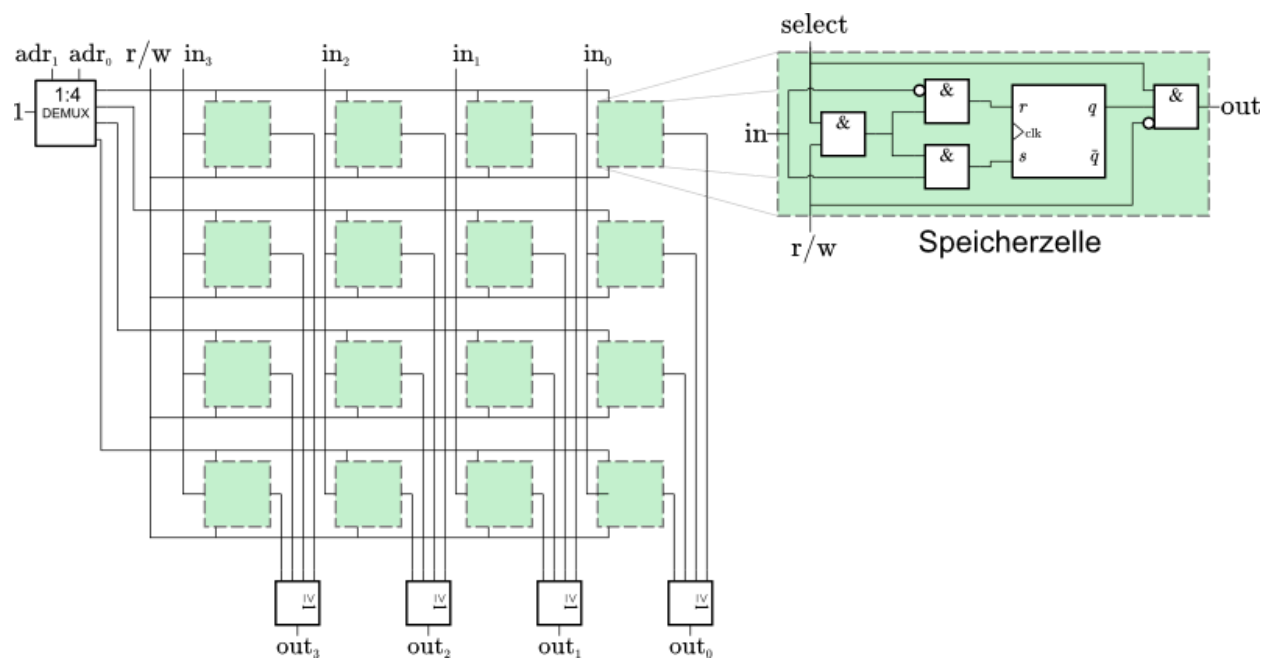
Bei Vergrößerung (rechts) durch ein Rasterelektronenmikroskop wird die regelmäßige Struktur des SRAMs sichtbar



[Bildquelle: [Die image of a STM32F103VGT6 ARM Cortex-M3 MCU](#) , Autor: ZeptoBars, [Creative Commons License](#)]

Thorsten Thormählen 9/32

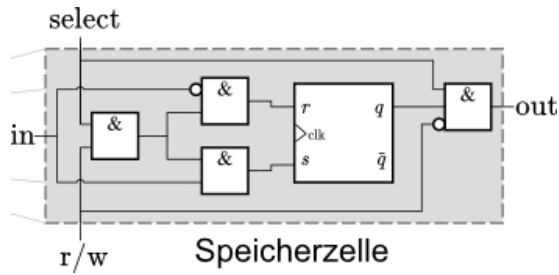
SRAM Speichermatrix



[Bildquelle: basierend auf: D. W. Hoffmann: Grundlagen der Technischen Informatik, 2. Auflage, Hanser 2009, Abb. 9.25, S. 329ff]

Thorsten Thormählen 10/32

SRAM Speicherzelle



select	r/w	in	out	q^{t+1}
0	\times	\times	0	q^t
1	0	0	q^t	q^t
1	0	1	q^t	q^t
1	1	0	0	0
1	1	1	0	1

Wenn $\text{select}=0$:

Die Speicherstelle wird nicht angesprochen

Der Ausgang out ist 0

Der gespeicherte Wert q ist unverändert

Wenn $\text{select}=1$:

Beim Lesen (" $\text{r/w}=0$ ") wird der Speicherinhalt ausgegeben und bleibt unverändert

Beim Schreiben (" $\text{r/w}=1$ ") ist der Ausgang 0 und der gespeicherte Wert q des Flipflops ändert sich gemäß dem Eingang in

DRAM

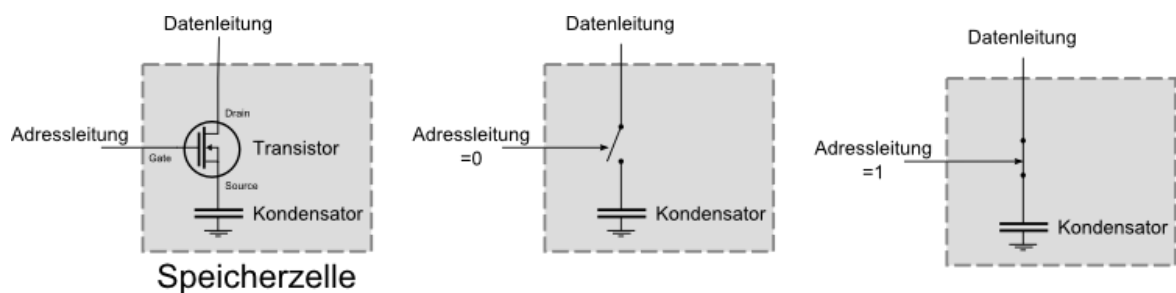
Beim DRAM (engl. Dynamic Random-Access Memory) wird eine Speicherzelle durch nur einen Kondensator und einen Transistor realisiert und benötigt daher sehr wenig Chipfläche

Damit ist es möglich, große Speicher kostengünstig herzustellen

Der Transistor wird an die Adressleitung angeschlossen und verhält sich wie ein Schalter:

Adressleitung=1, Schalter ist geschlossen, Kondensator übernimmt die Spannung von der Datenleitung (lädt sich entweder mit Ladungsträgern auf oder nicht)

Adressleitung=0, Schalter ist geöffnet, Kondensator behält sein Spannungspotential (Ladungsträger bleiben erhalten)



Thorsten Thormählen 12 / 32

DRAM: Grundlagen Transistoren

Transistoren sind spannungsgesteuerte Schalter



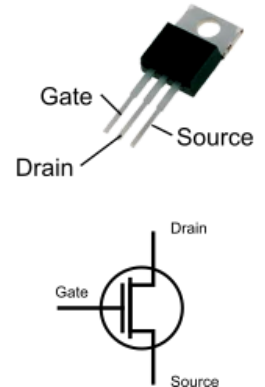
Ein Feldeffekttransistor hat drei Anschlüsse: Source, Drain und Gate.

Der Strom zwischen Source und Drain wird über die Spannung zwischen Gate und Source gesteuert

Niedrige Spannung am Gate = die Elektronen können das p-dotierte Gebiet nicht überwinden = kein Strom fließt

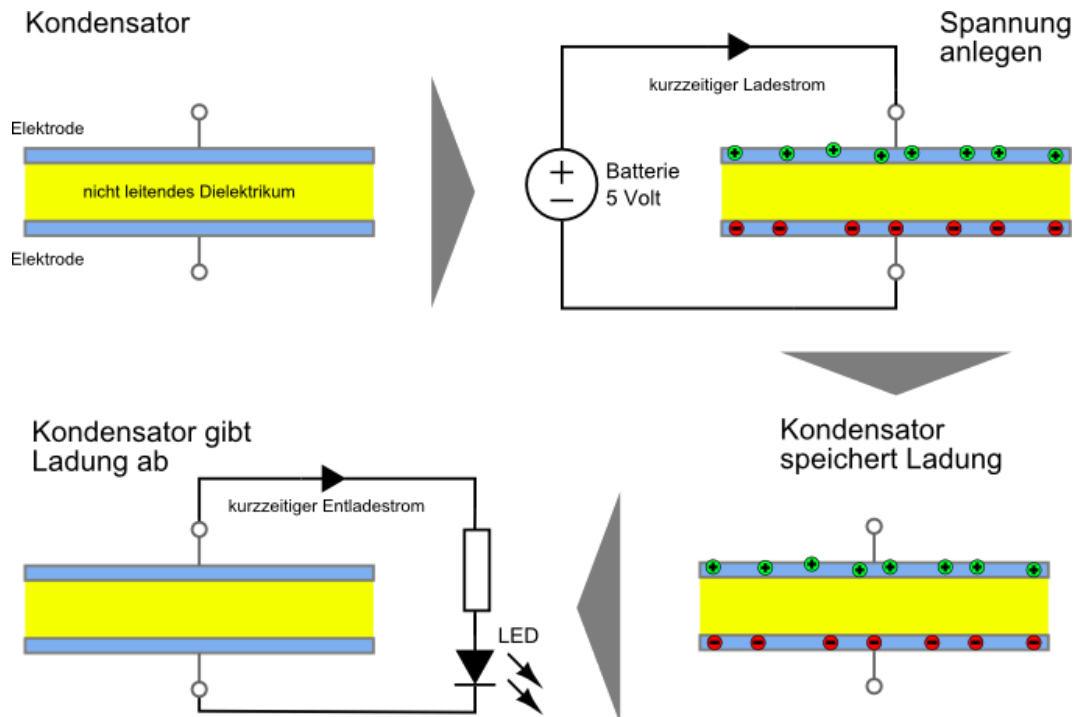
Ausreichend hohe Spannung am Gate = Elektronen reichern sich unterhalb des Gate an (n-Kanal) = Elektronen fließen durch Kanal = Strom fließt

Feldeffekttransistor



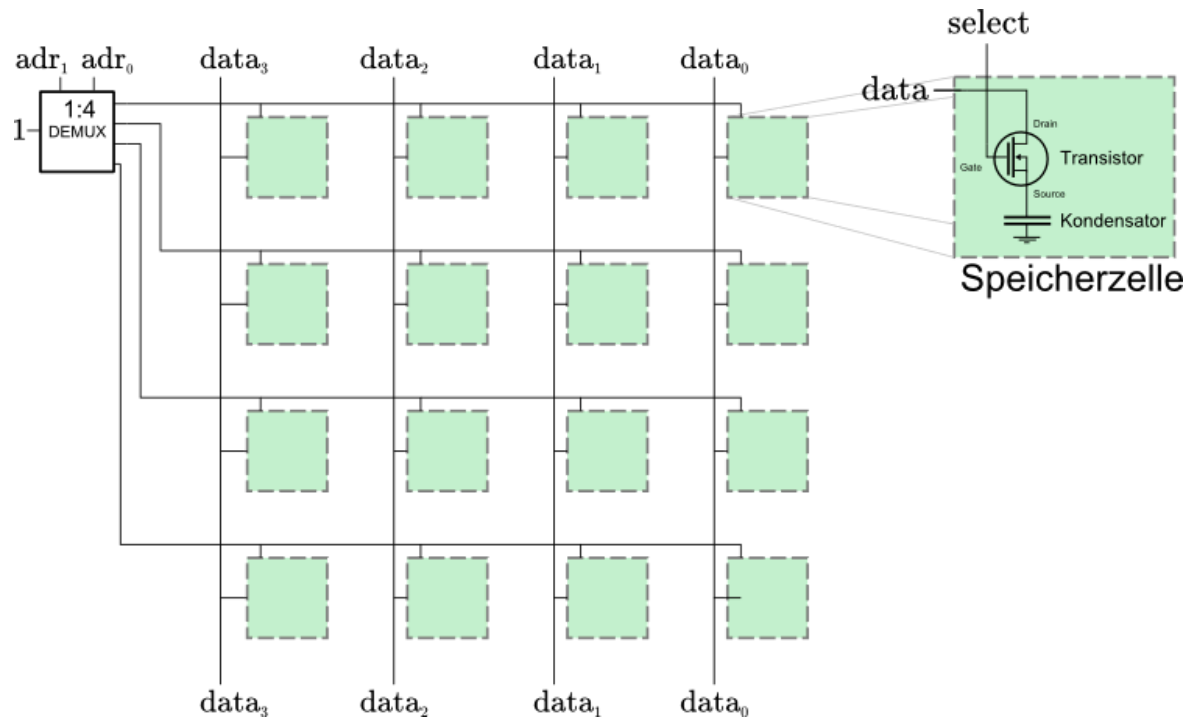
Thorsten Thormählen 13 / 32

DRAM: Grundlagen Kondensator



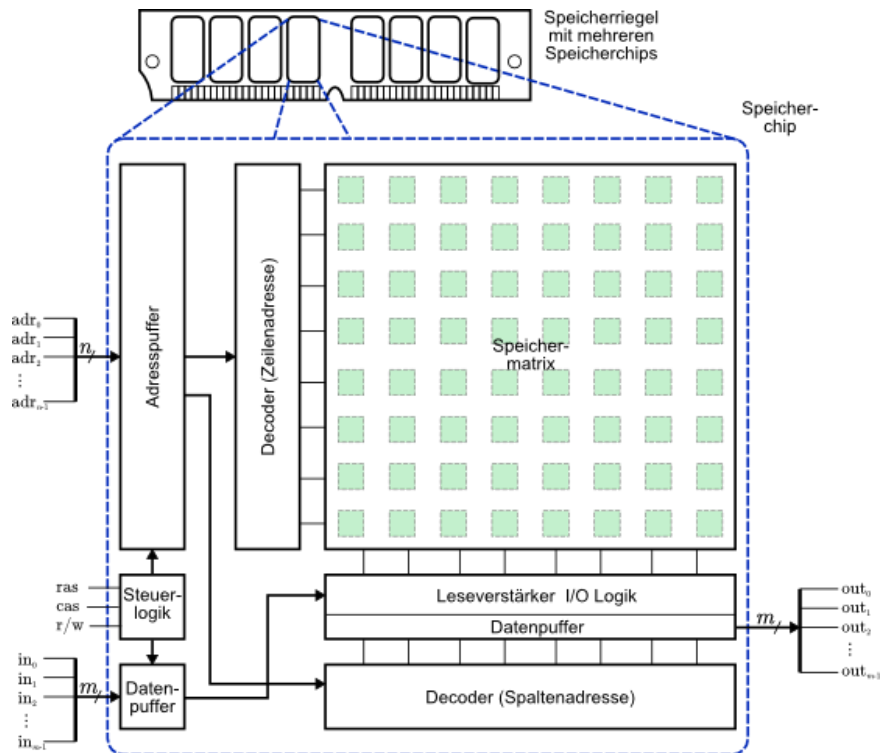
Thorsten Thormählen 14/32

DRAM Speichermatrix



Thorsten Thormählen 15/32

DRAM Speicherriegel mit mehreren Speicherchips



[Bildquelle: basierend auf: D. W. Hoffmann: Grundlagen der Technischen Informatik, 2. Auflage, Hanser 2009, Abb. 9.28, S. 332ff]

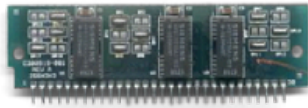
Thorsten Thormählen 16 / 32

Verschiedene Bauformen von DRAM Speicherriegeln



DIP

Dual In-Line



SIPP

Single In-Line Pin Package



SIMM 30 pin

Single Inline Memory Module



SIMM 72 pin

Single Inline Memory Module



DIMM (168-pin)

Dual Inline Memory Module



DDR DIMM (184-pin)

Dual Inline Memory Module

[Bildquelle: [Verschiedene RAM Typen](#) , Autor: Topory, [Creative Commons License](#)]

Thorsten Thormählen 17 / 32

DRAM Refresh

Der Kondensator in den Speicherzellen der DRAMs kann seine Ladung nicht sehr lange aufrechterhalten

Es gibt Leckströme über den Transistor, die zur Entladung führen

Ist der Kondensator zu stark entladen, kann nicht mehr festgestellt werden, ob eine logische 1 oder 0 gespeichert war

Daher werden alle Speicherzellen der Matrix periodisch "aufgefrischt" (engl. "refresh")

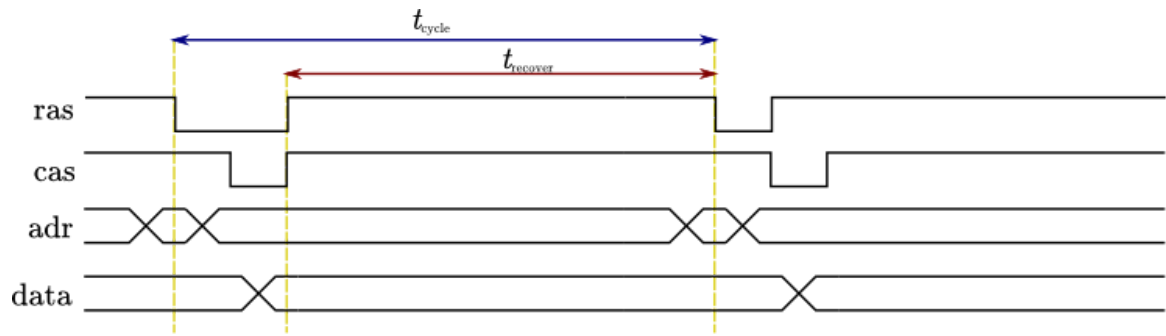
Eine typische Zeit bis zu einem Refresh sind 64 ms

Die Notwendigkeit, einen Refresh durchzuführen, gibt den DRAMs ihren Namen, da sie "dynamisch" betrieben werden müssen

Um ein Refresh durchzuführen, wird eine Zeile der Speichermatrix über den Leseverstärker in den Ausgangspuffer transferiert und anschließend wieder zurückgeschrieben

Frühere Chips benötigten dazu externe Ansteuerung, um die Adressen und Steuersignale zu generieren, heutige Chips haben dazu eigene Logik auf dem Chip, die den Refresh selbstständig durchführt

DRAM Zeitverhalten



Bei DRAM Bausteine wird häufig "Adressmultiplexing" betrieben

D.h. um Adressleitungen zu sparen, werden diese aufgeteilt und jeweils nur ein Teil der Adresse übertragen:

Ist $\text{ras}=0$, wird die Zeilenadresse gesendet ("Row Address Strobe")

Ist $\text{cas}=0$, wird die Spaltenadresse gesendet ("Column Address Strobe")

Die minimale Zeit t_{cycle} zwischen zwei Speicherzugriffen wird beim RAM maßgeblich geprägt durch die Erholzeit t_{recover} , die benötigt wird, um die Daten nach einem Zugriff wieder zurückzuschreiben, und die Bitleitungen für einen erneuten Zugriff vorzubereiten

DRAM Zeitverhalten

Um trotz langer Erholzeit t_{recover} schnell Daten zu liefern, werden die Daten derart auf die einzelnen Speicherchips eines Speicherriegels verteilt, dass bei konsekutivem Zugriff jeweils andere Speicherchips Daten liefern müssen

Damit kann sich ein Chip in der Erholzeit befinden, während die anderen Chips des Speicherriegels Anfragen beantworten

Des Weiteren gibt es den so genannten "Page" bzw. "Nipple" Modus, um direkt hintereinander liegende Speicherstellen auszulesen

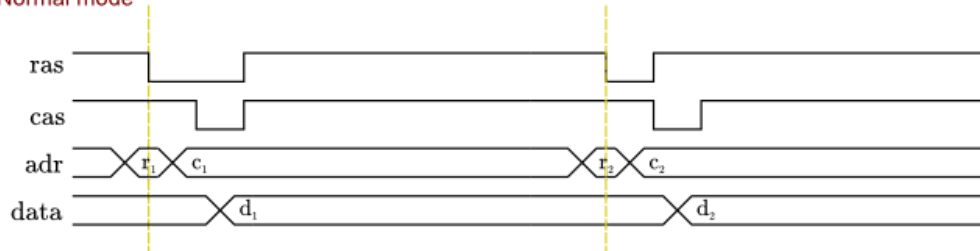
Dies funktioniert allerdings nur, solange Daten mit der gleichen Zeilenadresse benötigt werden

Im "Page"-Mode können unterschiedliche Spaltenadressen übergeben werden

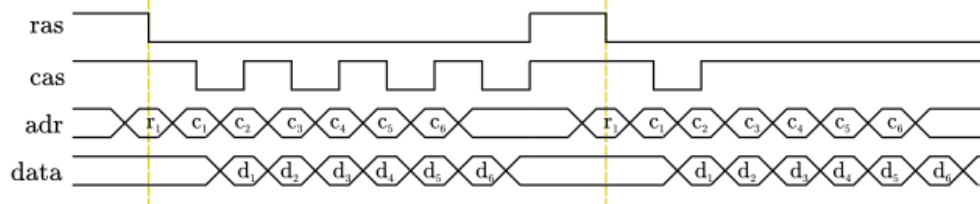
Im "Nibble"-Mode werden mit jeder fallenden Flanke vom cas-Signal konsekutive Daten der gleichen Zeile ausgegeben

DRAM Zeitverhalten

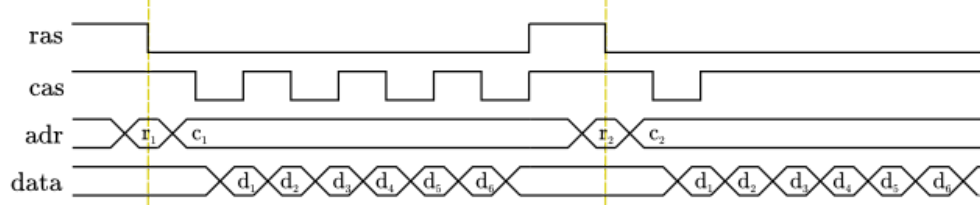
Normal mode



Page mode



Nibble mode



[Bildquelle: basierend auf: D. W. Hoffmann: Grundlagen der Technischen Informatik, 2. Auflage, Hanser 2009, Abb. 9.30, S. 334ff]

DRAM Entwicklung

Jahr der Einführung	Größe	Preis pro MB	Zugriffszeit neue Zeile bzw. Spalte	Zugriffszeit gleiche Zeile
1980	64 KB	1500 USD	250 ns	150 ns
1983	256 KB	500 USD	185 ns	100 ns
1985	1 MB	200 USD	135 ns	40 ns
1989	4 MB	50 USD	110 ns	40 ns
1992	16 MB	15 USD	90 ns	30 ns
1996	64 MB	10 USD	60 ns	12 ns
1998	128 MB	4 USD	60 ns	10 ns
2000	256 MB	1 USD	55 ns	7 ns
2002	512 MB	0,25 USD	50 ns	5 ns
2004	1 GB	0,10 USD	45 ns	3 ns
2012	8 GB	0,01 USD	24 ns	0,25 ns

[Quelle: basierend auf: David A. Patterson, John L. Hennessy, *Computer Organization and Design*, 2005 by Elsevier Inc., S. 490, adaptiert gemäß [Memory Prices \(1957-2013\)](#), John C. McCallum]

Thorsten Thormählen 22 / 32

Flash-Speicher

Flash-Speicher ist ein nicht-flüchtiger Datenspeicher, d.h. die Information bleibt auch ohne Stromversorgung erhalten

Im Gegensatz zu Festplatten oder optischen Laufwerken (CD, DVD, Blue-Ray) hat Flash-Speicher keine mechanisch-beweglichen Teile und kann damit in kleiner Bauform realisiert werden

Die Kosten pro Byte liegen allerdings höher als bei z.B. Festplatten

Anwendungen:

- Speicherkarten in mobilen Geräten (Kamera, Mobilfunkgerät, etc.)

- USB-Sticks

- Solid-State-Drive (kurz SSD), als schnelle Systemplatte in PCs und Laptops

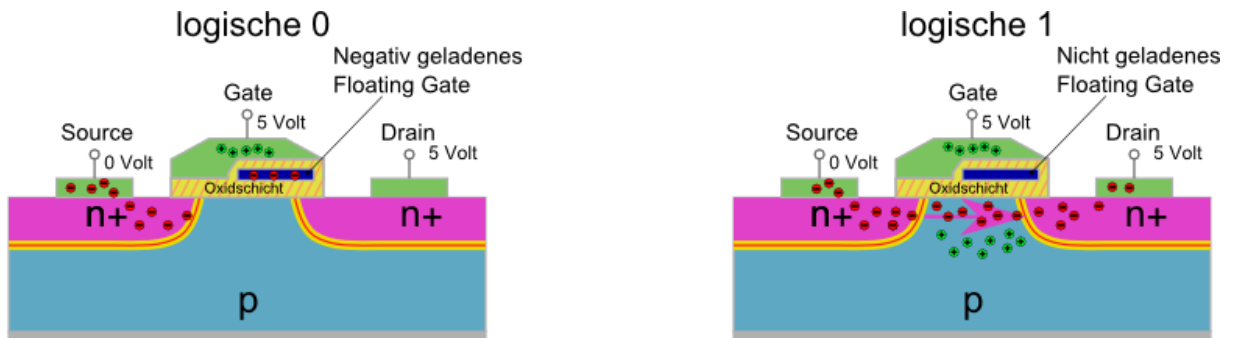
Flash-Speicher

Flash-Speicher basiert auf dem Floating-Gate-Transistor

Dies ist ein normaler Feldeffekttransistor mit zusätzlichem "Floating Gate"

Die elektrische Ladung auf dem Floating Gate kodiert, ob die Speicherzelle eine logische 0 oder 1 speichert

Das Floating Gate ist vom Dielektrikum umgeben, d.h. es behält die aufgebrachte Ladung permanent (> 1 Jahr)



Thorsten Thormählen 24 / 32

Flash-Speicher

Lesen

Ist das Floating Gate negativ geladen, verhindert es, dass der Transistor normal betrieben werden kann = kein Strom fließt = Kodiert logische 0

Ist das Floating Gate nicht geladen, kann der Transistor normal arbeiten = Strom fließt = Kodiert logische 1

Schreiben

Logische 0: Es wird eine hohe positive Spannung (ca. 15V) angelegt. Elektronen "tunneln" aus dem Substrat auf das Floating Gate.

Logische 1: Es wird eine hohe negative Spannung (ca. -15V) angelegt. Elektronen "tunneln" vom Floating Gate in das Substrat.

Degeneration

Bei jedem Lese- oder Schreibvorgang wird die Oxidschicht leicht beeinträchtigt, so dass ein Transistor nach ca. 10.000 bis mehreren 100.000 Zyklen unbrauchbar wird.

Auf dem Chip gibt es Logik, die fehlerhafte Speicherzellen erkennt, sich deren Position in der Speichermatrix merkt, und die betroffene Speicherzelle durch andere ersetzt.

Thorsten Thormählen 25 / 32

Festplattenlaufwerk

Eine Festplatte (engl. Hard Disk Drive, HDD) ist ein magnetisches Speichermedium für die nicht-flüchtige Speicherung großer Datenmengen (ca. 1 bis 4 TB in heutigen Standard-PCs)

Typischerweise besteht ein Laufwerk aus mehreren Platten, die auf einer gemeinsamen Spindel durch einen Elektromotor gedreht werden

Rotationsgeschwindigkeiten bewegen sich im Bereich von ca. 5.400 bis 15.000 Rotationen pro Minute

Auf einem gemeinsamen Actuatorarm befinden sich die Schreib-Lese-Magnetköpfe für die einzelnen Platten

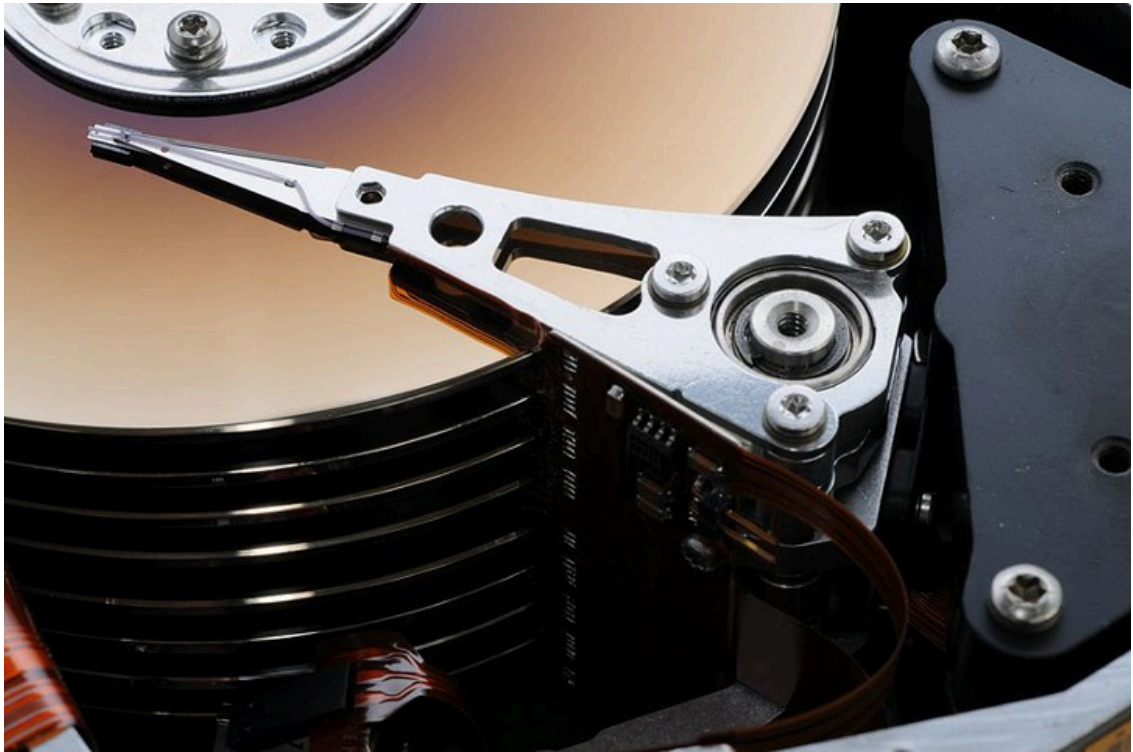
Die Magnetköpfe schweben mit einem sehr geringen Abstand von wenigen Nanometern über den Platten, die eine hart-magnetischen Beschichtung besitzen, und übertragen die magnetische Information während sich die Platte am Kopf vorbei dreht



[Bildquelle: [An HDD](#), Autor: Ervins Strauhmanis, [Creative Commons License](#)]

Thorsten Thormählen 26 / 32

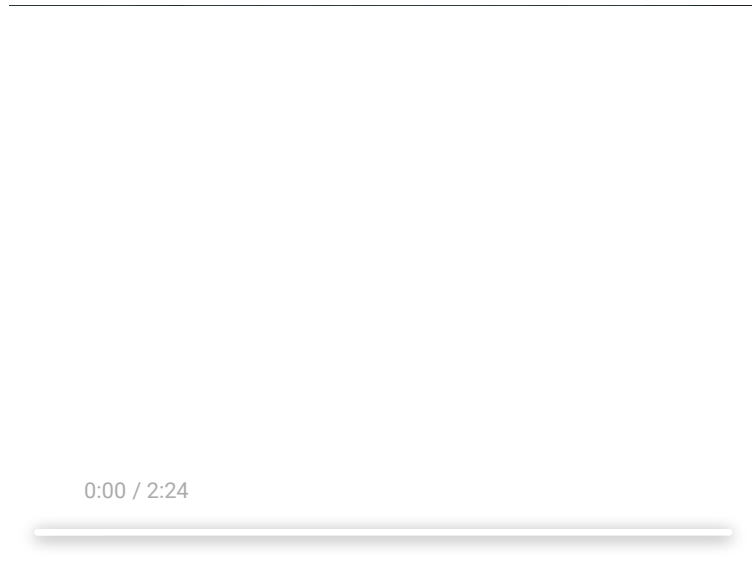
Festplattenlaufwerk



[Bildquelle: [Festplatte Seagate ST4702N](#), Autor: Hubert Berberich, public domain]

Thorsten Thormählen 27 / 32

Festplattenlaufwerk



Video einer geöffneten Festplatte

[Quelle: [Hard Drive Operating Without Cover](#), Autor: Nick Oakman]

Thorsten Thormählen 28 / 32

Festplattenlaufwerk

Daten werden blockweise adressiert. Ein Block hat z.B. 512, 2048 oder 4096 Bytes.

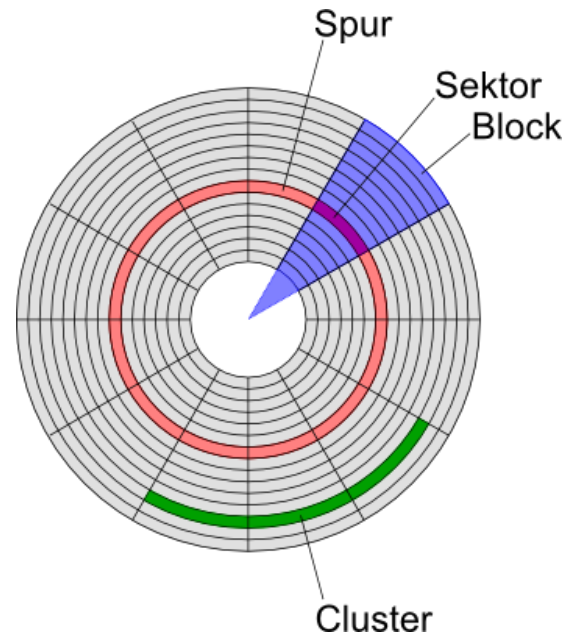
Eine Spur auf der Platte entspricht einer Kreisbahn, auf der der Magnetkopf läuft

Da typischerweise mehrere Magnetköpfe im Einsatz sind, werden die gemeinsam angesteuerten Spuren der verschiedenen Platten "Zylinder" genannt

Ein Block entspricht einem kleinen Winkelbereich auf einer Spur

Alle Blöcke mit dem gleichen Winkelbereich werden "Sektor" genannt

Cluster sind Gruppen von benachbarten Blöcken auf einer Spur, die gemeinsam angesprochen werden können



Thorsten Thormählen 29 / 32

Festplattenlaufwerk

Platten können hohe kontinuierliche Übertragungsraten besitzen (heutzutage ca. 300 MB/s)

Allerdings ist die mittlere Zugriffszeit geringer (5-20 ms), da die Köpfe erst mechanisch an die richtige Stelle gebracht werden müssen und gewartet werden muss, bis der richtige Block vorbei rotiert

Die Blöcke enthalten nicht nur die eigentlichen Daten, sondern auch Zusatzinformationen für die Fehlererkennung

Aufgrund der langsamen Zugriffszeit von Festplatten, werden Daten typischerweise immer erst in den Hauptspeicher (DRAM) geladen und von dort aus vom Prozessor verarbeitet

Thorsten Thormählen 30 / 32

Quiz

Frage: Eine 3,5 Zoll Festplatte dreht mit 7200 Umdrehungen pro Minute. Wie groß ist ca. die Geschwindigkeit auf der äußersten Spur?

Antwort 1: 120 km/h

Antwort 2: 1200 km/h

Antwort 3: 12000 km/h

Am Online-Quiz teilnehmen durch Besuch der Webseite:
www.onlineclicker.org

Gibt es Fragen?



Anregungen oder Verbesserungsvorschläge können auch gerne per E-mail an mich gesendet werden: [Kontakt](#)

[Weitere Vorlesungsfolien](#)

[[Impressum](#), [Datenschutz](#)]

Thorsten Thormählen 32 / 32

