

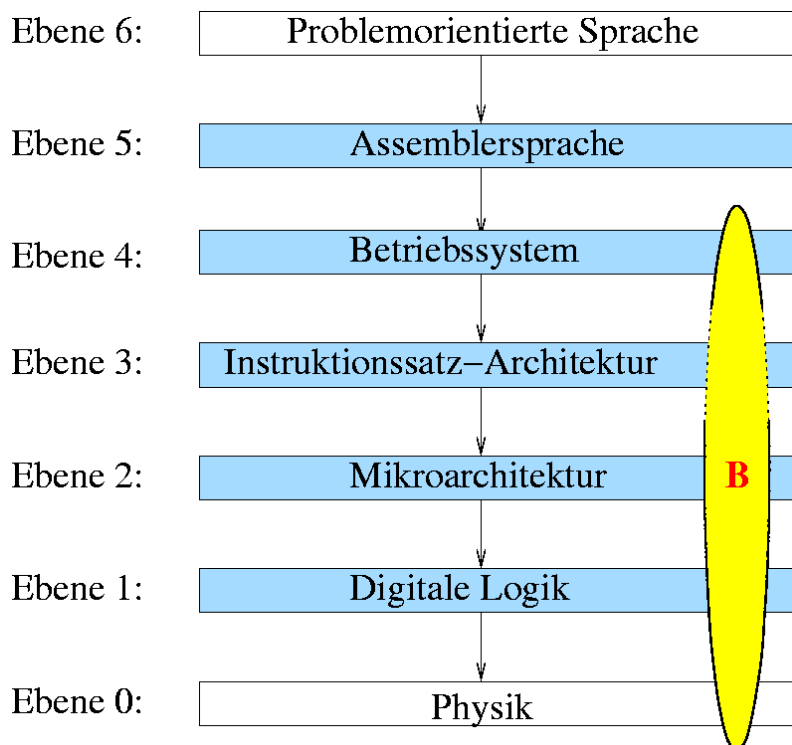
B Hauptspeicher und Cache

1. Begriffe
2. SRAM
3. DRAM
4. DRAM-Varianten: EDO-RAM, SDRAM, DDR-RAM, RAMBUS
5. Festwertspeicher: PROM, EPROM, EEPROM
6. Exkurs: Assoziativspeicher
7. Cache

1

B Hauptspeicher und Cache

- Einordnung in das Schichtenmodell:

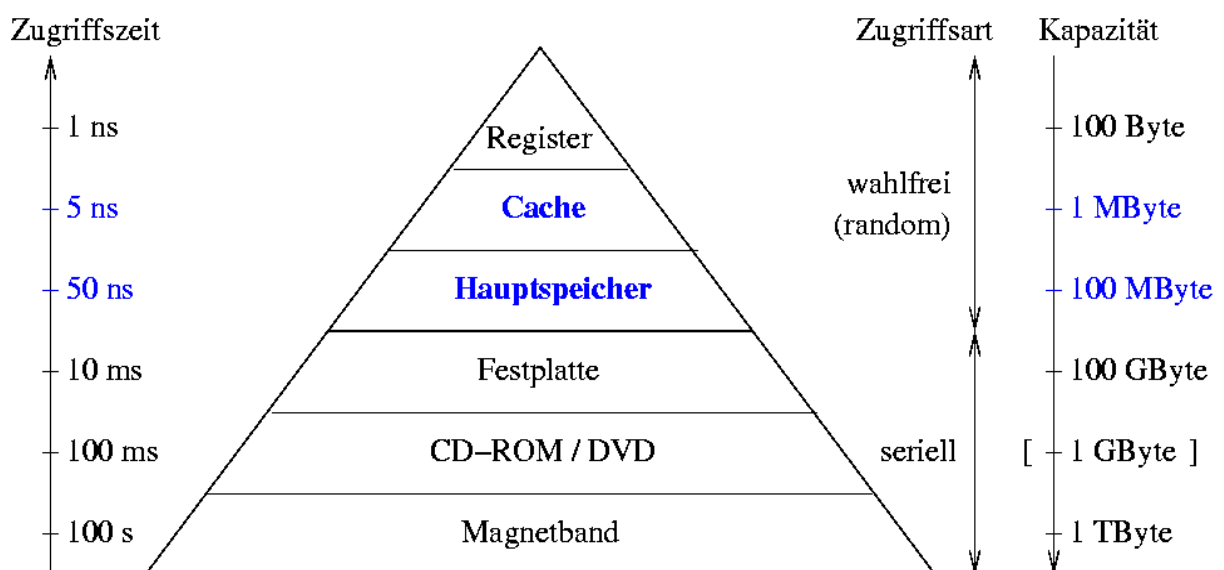


B Hauptspeicher und Cache

- im Laufe der Geschichte wurden verschiedene **Technologien** eingesetzt, um Informationen zu speichern:
 - Modifikation von Strukturen**: Lochkarte, Schallplatte
 - Rückkopplung**: Flip-Flops, SRAM
 - elektrische Ladungen**: Kondensator, DRAM
 - Magnetismus**: Magnetkernspeicher, Magnetband, Diskette, Festplatte
 - Optik**: Bar-Codes, CD-ROM, DVD
- Kriterien zum Vergleich von Speichertechnologien: Geschwindigkeit, Kapazität, Dichte, Energiebedarf, Robustheit, Kosten

B Hauptspeicher und Cache

- in heutigen Rechnersystemen findet man eine bis zu 6-stufige **Speicherhierarchie**:



1 Begriffe

- **RAM** (*Random Access Memory*)
Speicher mit wahlfreiem Zugriff auf beliebige Adressen
- **SRAM** (*Static RAM*)
statischer RAM-Speicher
- **DRAM** (*Dynamic Random Access Memory*)
dynamischer RAM-Speicher
- **SDRAM** (*Synchronous DRAM*)
synchroner dynamischer RAM-Speicher
- **DDR-SDRAM** (*Double Data Rate SDRAM*)
synchroner dynamischer RAM-Speicher mit doppelter Datenrate
- **RDRAM** (*Rambus DRAM*)
spezielle DRAM-Bausteine der Fa. Rambus

1 Begriffe (2)

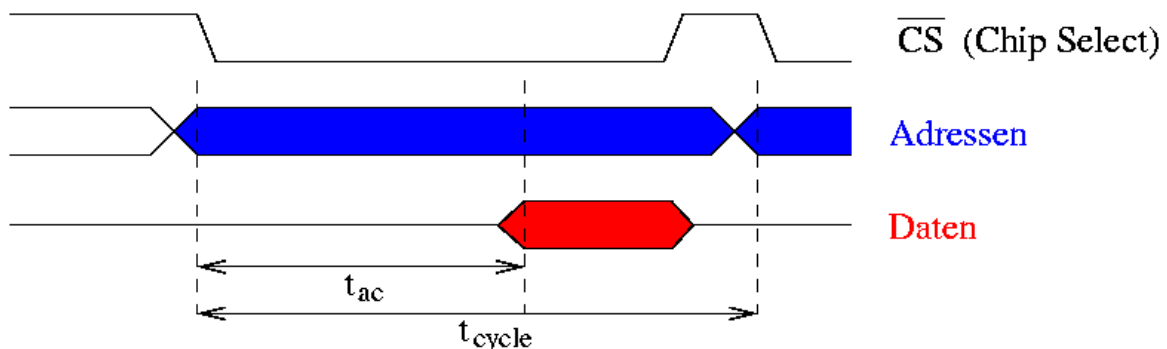
- **ROM** (*Read-Only Memory*)
- **PROM** (*Programmable ROM*)
einmalig programmierbarer Speicher, z.B. durch Masken-Programmierung bei der Chip-Herstellung
- **EPROM** (*Erasable PROM*)
mit UV-Licht löschares und elektrisch (durch zusätzliche Steuerleitungen und Überspannungen) programmierbares PROM
- **EEPROM** (*Electrically Erasable PROM*)
elektronisch löschares und danach wieder programmierbares PROM
- **Flash** (Eigenname), ähnliche Eigenschaften wie EEPROM
- **NV-RAM** (*Non-Volatile RAM*)
normalerweise SRAM mit eingebauten EEPROM: bei Stromverlust werden die Daten sofort in das EEPROM umkopiert; wenn wieder Strom vorhanden ist, werden Daten vom EEPROM in das SRAM zurückgeschrieben.
- **Batteriegepuffertes RAM**

1 Begriffe (3)

- Speicher werden eingeteilt in
 - 1) **flüchtige** Speicher: SRAM, DRAM, SDRAM, ...
Informationen gehen nach Ausschalten der Versorgungsspannung verloren!
 - 2) **nichtflüchtige** Speicher: PROM, EPROM, EEPROM, ...
Informationen bleiben auch ohne Versorgungsspannung über längere Zeit (typischerweise einige Jahre) erhalten!
- Die **Kapazität** von Speicherbausteinen wird in kByte (bzw. kBit), MByte (bzw. MBit) oder GByte (bzw. MBit) angegeben; es gilt:
 - 1 KByte = 1024 Byte = 2^{10} Byte
 - 1 MByte = 1024 KByte = 1024×1024 Byte = 2^{20} Byte = 1.048.576 Byte
 - 1 GByte = 1024 MByte = $1024 \times 1024 \times 1024$ Byte = 2^{30} Byte = 1.073.741.824 Byte

1 Begriffe (4)

- als **Zugriffszeit** t_{ac} bezeichnet man die Zeitspanne vom Anlegen einer Adresse bis zur Gültigkeit der ausgelesenen Daten
- als **Zykluszeit** t_{cycle} bezeichnet man die Zeitspanne vom Anlegen einer Adresse bis zum möglichen Anlegen der nächsten Adresse
- mögliches **Zeitdiagramm** eines Lesezyklus:



- Zykluszeit ist oft (z.B. bei DRAMs) größer als die Zugriffszeit!

1 Begriffe (5)

- zwei unterschiedliche Arten der Realisierung **flüchtiger** Speicher:

1) Statischer Speicher

- Speicherung durch 4 bis 6 Transistoren je Bit
- kein Refresh notwendig
- Bausteine: SRAM, Dual-ported SRAM
- Zugriffs- und Zykluszeit: ca. 10 ns als externe Bausteine (schneller, wenn sie direkt in den Prozessor integriert werden, z.B. als Cache).

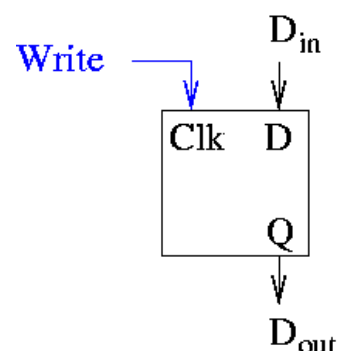
2) Dynamischer Speicher

- Speicherung durch einen Transistor und einen Kondensator je Bit
- Refresh notwendig, da der Kondensator im Verlauf der Zeit (Größenordnung: einige ms) seine Ladung verliert.
- sehr hohe Datendichte (Kapazität je mm² Chipfläche) möglich
- Bausteine: DRAM, SDRAM, DDR-SDRAM, RDRAM
- Zugriffszeit für erstes Datenwort: ca. 40 ns, Zykluszeit: ca. 60 ns

2 SRAM

- **Idee:** auf Flip-Flop Prinzip basierender Speicherbaustein, z.B. ein D-Flip-Flop für ein Bit:

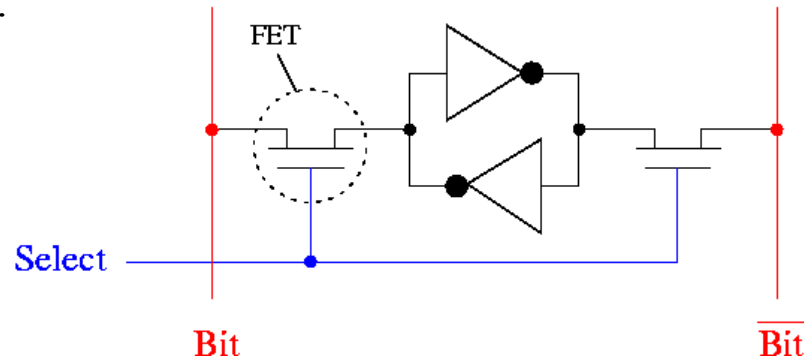
(vgl. Folie F-6 aus TI I)



- bei **Write = 1** wird Information von D_{in} gespeichert und steht am Ausgang D_{out} zur Verfügung
- Information bleibt gespeichert, auch wenn sich D_{in} bei **Write = 0** wieder ändern sollte
(solange Versorgungsspannung anliegt)

2 SRAM (2)

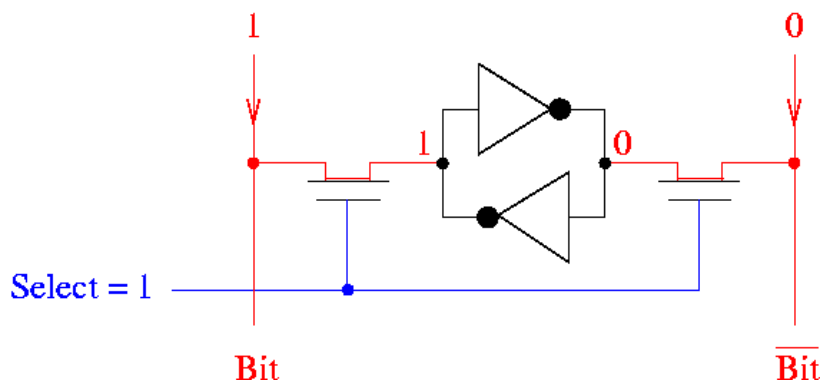
- SRAM-Bausteine basieren jedoch nicht auf vollständigen D-Flip-Flops, sondern auf einfacheren SRAM-Zellen
- vereinfachter Aufbau einer **typischen SRAM-Zelle** mit FET-Transistoren:



- zwei Inverter repräsentieren bistabile Kippstufe
- bei **Select = 1** wird die bistabile Kippstufe mit Leitungen **Bit** und **Bit-bar** verbunden

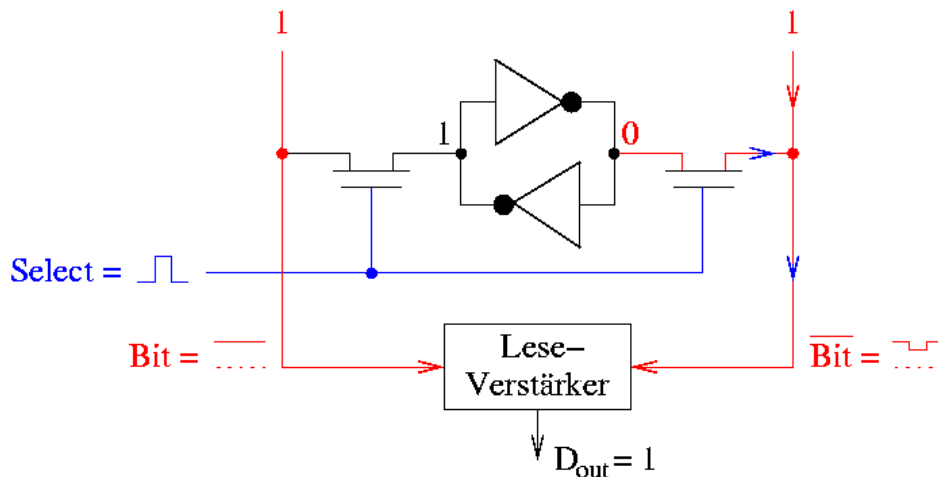
2 SRAM (3)

- **Beschreiben** einer SRAM-Zelle:
 - Bit-Leitungen werden mit **Bit = 1** und **Bit-bar = 0** (zum Speichern einer „1“), bzw. mit **Bit = 0** und **Bit-bar = 1** (zum Speichern einer „0“) geladen
 - Auswahl-Leitung wird auf **Select = 1** gesetzt
 - Kippstufe schwingt ggf. in den anderen stabilen Zustand
 - Rücksetzen der Pegel auf **Select**-Leitung und danach auf **Bit**-Leitungen
 - Zustand bleibt erhalten, solange Versorgungsspannung anliegt



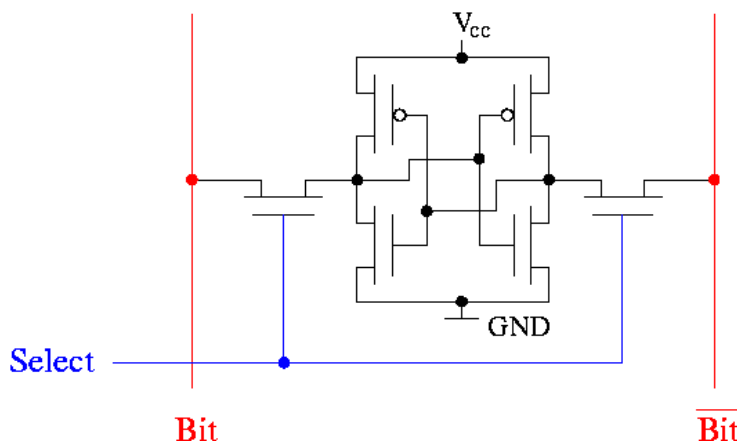
2 SRAM (4)

- **Auslesen** einer SRAM-Zelle:
 - Bit-Leitungen werden mit $\text{Bit} = 1$ und $\overline{\text{Bit}} = 1$ geladen („Precharging“)
 - auf Auswahl-Leitung wird ein kurzer Impuls gegeben
 - resultierender geringer Spannungsabfall auf Bit oder $\overline{\text{Bit}}$ wird durch Verstärkerschaltung erkannt, die eine entsprechende Ausgabe generiert



2 SRAM (5)

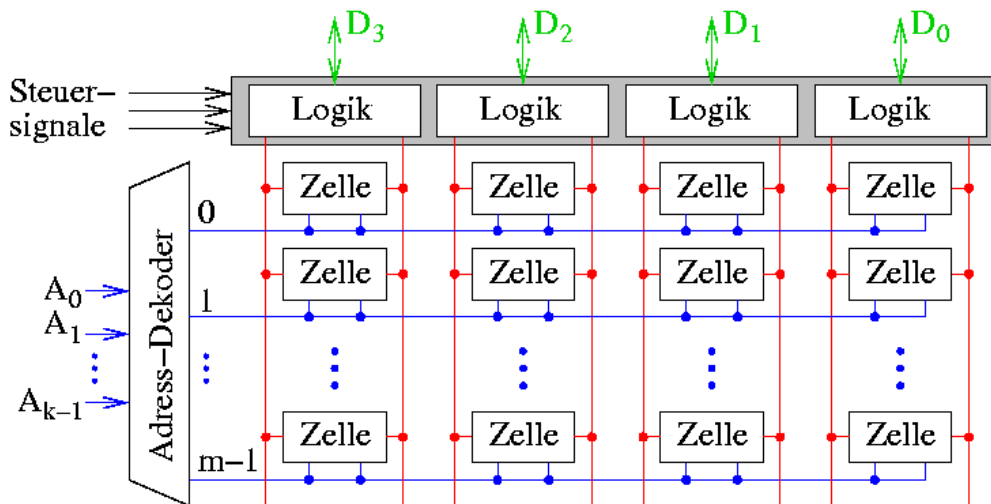
- Realisierung einer SRAM-Zelle mit insgesamt 6 Transistoren (mit 2 zusätzlichen FET-Transistoren je Inverter):



- heute i.a. Realisierung in CMOS-Technologie (Complementary Metal Oxide Semiconductor)

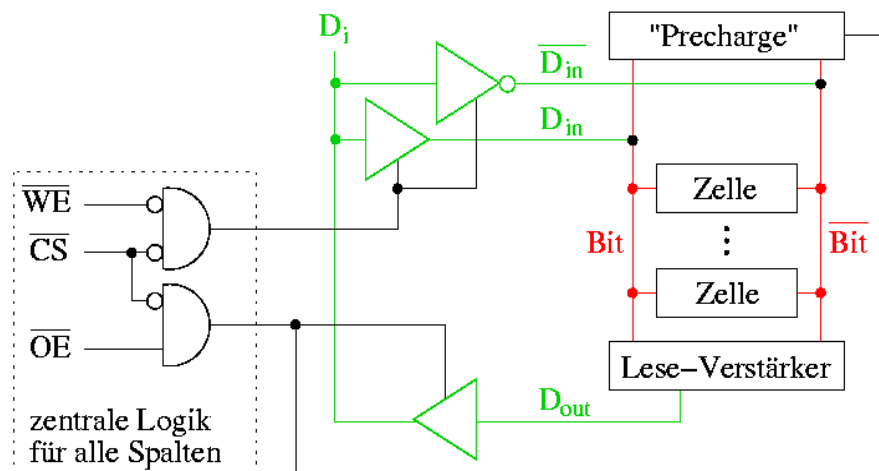
2 SRAM (6)

- **Organisation** eines SRAM-Bausteins: $m \times n$ Speichermatrix (m Speicherzeilen mit je n SRAM-Zellen, Auswahl einer der $m = 2^k$ Zeilen über k Adressleitungen und einen k -zu- m Adress-Dekoder, alle Zellen einer Spalte nutzen die gleichen Bit-Leitungen)
- vereinfachte **Architektur** eines $m \times 4$ SRAM-Bausteins:



2 SRAM (7)

- **Steuerlogik** eines SRAMs mit **Tristate-Puffern** und einigen weiteren Steuerleitungen (i.a. „low active“):
 - \overline{CS} (*Chip Select*) zur Auswahl und Aktivierung eines SRAM-Bausteins
 - \overline{WE} (*Write Enable*) zum Speichern eines Wertes
 - \overline{OE} (*Output Enable*) zum Lesen und Freischalten der Ausgänge
- Realisierung der Steuerlogik für SRAM-Spalte i :



2 SRAM (8)

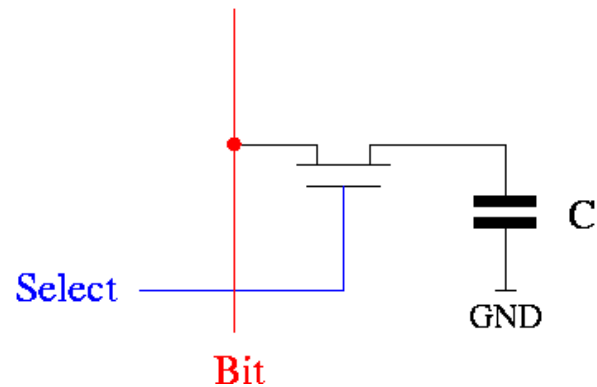
- SRAM-Bausteine gibt es in verschiedenen Organisationsformen, und Kapazitäten, z.B.:
 - **64k × 4** 64k Worte à 4 Bit (16 Adress-, 4 Datenleitungen, 32 KByte)
 - **128k × 8** 128k Worte à 8 Bit (17 Adress-, 8 Datenleitungen, 128 KByte)
 - **512k × 8** 512k Worte à 8 Bit (19 Adress-, 8 Datenleitungen, 512 KByte)
 - **1M × 16** 1M Worte à 16 Bit (20 Adress-, 16 Datenleitungen, 32 MByte)
- oft wahlweise als „*low power*“ oder als „*high speed*“ Version erhältlich
- auch synchrone (d.h. getaktete SRAMs) verfügbar
 - arbeiten synchron mit Prozessortakt
 - sehr kurze Zugriffszeiten (ca. 5 ns)
 - interne Inkrementenschaltung für Adressen erlaubt sehr schnelles getaktetes Schreiben und Lesen von vier Datenworten auf aufeinanderfolgenden Adressen (**Burst-Modus**, vgl. DRAMs)

2 SRAM (9)

- **Vor-/Nachteile** von SRAMs:
 - schneller Zugriff (weniger als 10 ns bei externem Baustein)
 - unempfindlich gegen elektromagnetische Strahlung
 - geringe Datendichte auf dem Chip
⇒ hoher Flächenbedarf (ca. 4-fach im Vergleich zu DRAM)
 - hoher Energiebedarf (bei hoher Geschwindigkeit)
 - hoher Preis
- Typischer **Einsatz** von SRAMs:
 - Netzwerkkomponenten
 - Weltraumgeräte
 - Höchstgeschwindigkeitsrechner (Vektorrechner)
 - L1 und L2 Cachespeicher (mit synchronen SRAMs)

3 DRAM

- benötigt zur Speicherung eines Bits nur einen Transistor und einen Kondensator:
- Aufbau einer **DRAM-Zelle**:



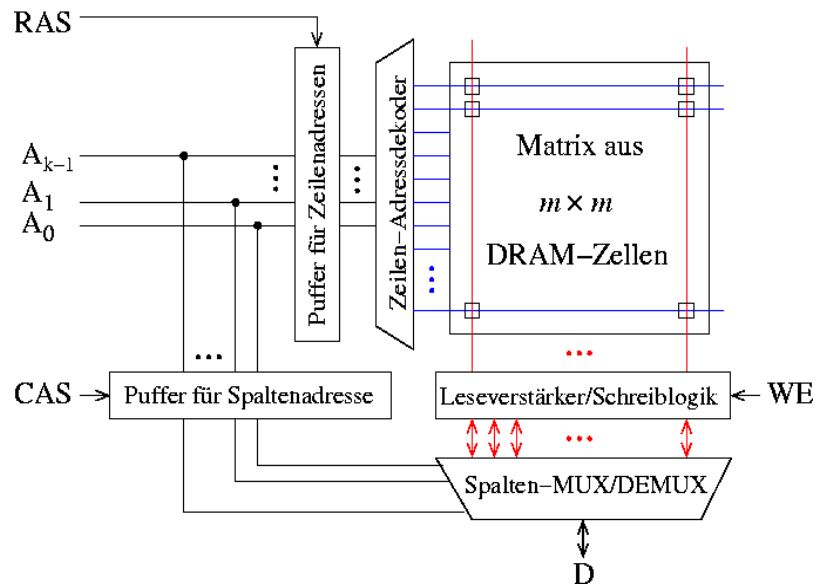
- extrem hohe Speicherkapazität durch geringen Flächenbedarf einer DRAM-Zelle: 128 MBit bis 1 GBit je Chip (Stand 2003)

3 DRAM (2)

- **Schreiben** einer DRAM-Zelle:
 - Bit-Leitung wird auf **Bit = 1** oder **Bit = 0** gesetzt
 - Auswahlleitung **Select = 1** setzen \Rightarrow Transistor leitet und Kondensator wird aufgeladen oder entladen
- **Auslesen** einer DRAM-Zelle:
 - Auswahlleitung **Select = 1** setzen \Rightarrow Transistor leitet
 - falls Kondensator aufgeladen war, wird ein kurzer Impuls auf **Bit**-Leitung erzeugt
 - Impuls kann durch **Leseverstärker** erkannt werden, der eine 1 am Ausgang erzeugt
 - der Kondensator wird jedoch beim Auslesen **entladen**; daher muß die DRAM-Zelle nach jedem Lesen wieder mit dem zuvor gelesenen Wert beschrieben werden!

3 DRAM (3)

- Organisation eines $m^2 \times 1$ DRAMs (vereinfacht):

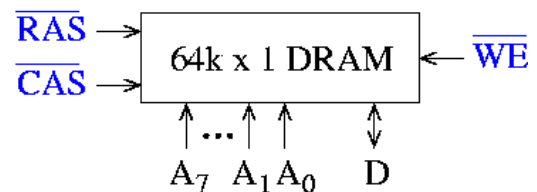


- **Multiplex-Interface:**

- jede $2k$ -Bit Adresse (mit $m = 2^k$) besteht aus Zeilen- und Spaltenadresteil
- um E/A-Pins einzusparen, werden beide Teile sequentiell über k Adressleitungen übertragen (\Rightarrow zusätzliche Steuersignale notwendig)

3 DRAM (4)

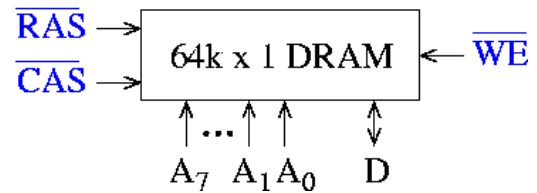
- Lesen eines Wertes aus einem $64k \times 1$ DRAM-Baustein:



- zuerst werden die Adressbits 15 bis 8 an A_7 bis A_0 angelegt
- das Signal **RAS** (*Row Address Strobe*) wird aktiviert und alle 256 Bits Speicherzeile werden von den Leseverstärkern ausgelesen
- dann werden die Adressbits 7 bis 0 an A_7 bis A_0 angelegt
- das Signal **CAS** (*Column Address Strobe*) wird aktiviert, wodurch aus den Ausgängen der 256 Leseverstärker das gewünschte Bit ausgewählt und am Pin D ausgegeben wird
- ein weiteres Bit aus der gleichen Zeile (*Page*) kann ausgelesen werden, indem **CAS** deaktiviert, eine andere Spaltenadresse angelegt wird, und **CAS** wieder aktiviert wird (\Rightarrow **Fast Page Mode, FPM**)
- nach Deaktivierung von **RAS** und **CAS** wird Zeile zurückgeschrieben

3 DRAM (5)

- Schreiben eines Wertes in einen 64k × 1 DRAM-Baustein:



- zuerst werden die Adressbits 15 bis 8 an A_7 bis A_0 angelegt
- die Signale **RAS** (*Row Address Strobe*) und **WE** werden aktiviert und alle 256 Bits der Speicherzeile werden von den Leseverstärkern ausgelesen
- dann werden die Adressbits 7 bis 0 an A_7 bis A_0 angelegt
- das Signal **CAS** (*Column Address Strobe*) wird aktiviert und das zu schreibende Bit wird an D angelegt; im Leseverstärker wird das ausgewählte Bit durch das Signal an D überschrieben
- in der ausgewählten Zeile können ggf. weitere Bits geschrieben werden (**Fast Page Mode**, vgl. vorige Folie)
- nach Deaktivierung von **RAS** und **CAS** wird modifizierte Zeile in die Speichermatrix zurückgeschrieben

3 DRAM (6)

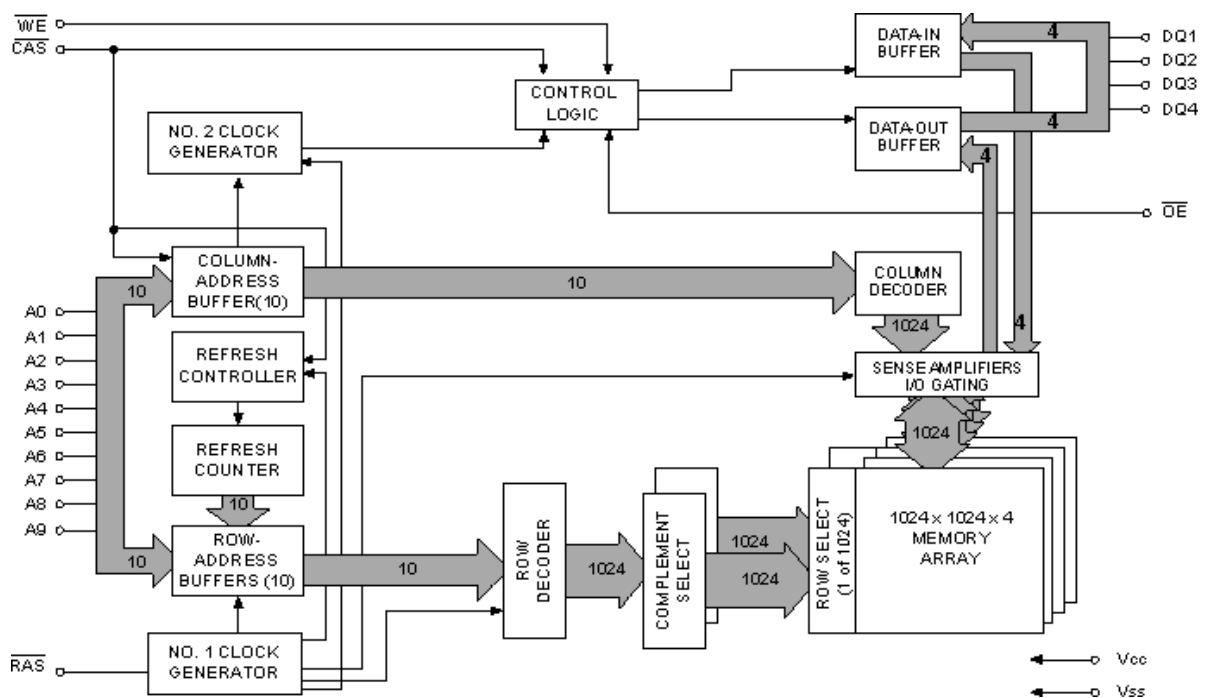
- Kapazität C des Kondensators einer DRAM-Zelle ist sehr gering (einige Femto-Farad, d.h. einige 10^{-15} Farad)
- Ladungsverlust nicht nur bei jedem Lesen, sondern auch
 - langsam mit der Zeit aufgrund eines geringen Leckstroms im Transistor (Größenordnung: einige ms)
 - durch elektromagnetische Strahlung
- periodischer **Refresh** erforderlich, um Zelleninhalt über längere Zeit zu speichern:
 - jede Zeile muss regelmäßig in Abständen von **8** bis **64 ms** gelesen werden, wodurch ihr Inhalt erneut geschrieben wird
 - dies wird implementiert durch einen in das DRAM integrierten **Zähler**, dessen Zählerstand die nächste aufzufrischende Zeilenadresse angibt
 - der Memory-Controller erzeugt periodisch (z.B. alle **15.6 μ s**, → BIOS) eine bestimmte Kombination der Steuersignale (z.B. *CAS-before-RAS*), durch die ein **Blindlesezyklus** ausgelöst und Zähler inkrementiert wird

3 DRAM (7)

- der Zugriff auf eine beliebige Speicheradresse benötigt ca. **60 ns** (da Auslesen der sehr geringen Ladung über Leitungen mit hohem Widerstand und das Verstärken eine gewisse Zeit benötigen)
- Folgezugriffe mit anderen Spaltenadressen in der gleichen Speicherzeile benötigen jeweils ca. **30 ns**
- DRAM-Bausteine sind wie SRAMs in unterschiedlichen **Organisationen** verfügbar (z.B. kann ein 16 MBit Speicher als $1M \times 16$, $512k \times 8$, $4M \times 4$ oder auch als $16M \times 1$ organisiert sein)
- bei einer Kapazität c und einer Wortbreite von n Bit benötigt man n Speichermatrizen aus $m \times m$ DRAM-Zellen mit $m = \sqrt{c / n}$
- erfüllen einzelne DRAM-Bausteine nicht die Anforderungen bzgl. Wordbreite oder Adreßraum, so können mehrere DRAM-Bausteine geeignet zusammengeschaltet werden

3 DRAM (7)

- detaillierter Aufbau eines $1M \times 4$ DRAM-Bausteins:

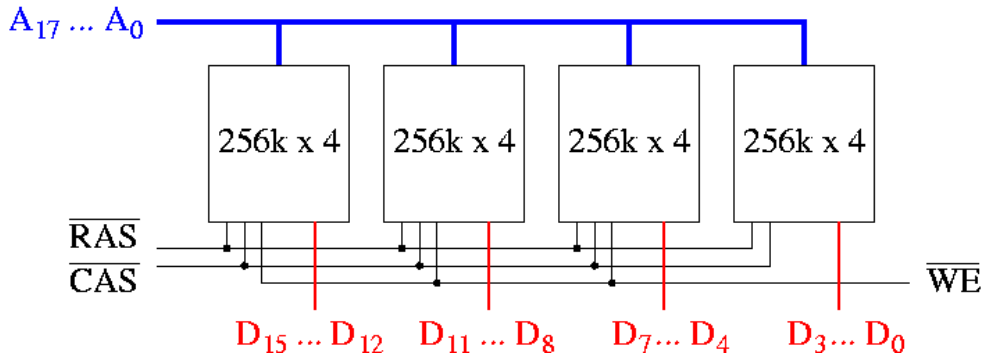


3 DRAM (8)

- Möglichkeiten der Zusammenschaltung von DRAM-Bausteinen:

1) Vergrößern der Wortbreite:

- identische Adress- und Steuer-Leitungen zu allen Speicherbausteinen, Datenleitungen werden aufgeteilt
- Beispiel: Speicher mit 256k Worten à 16 Bit

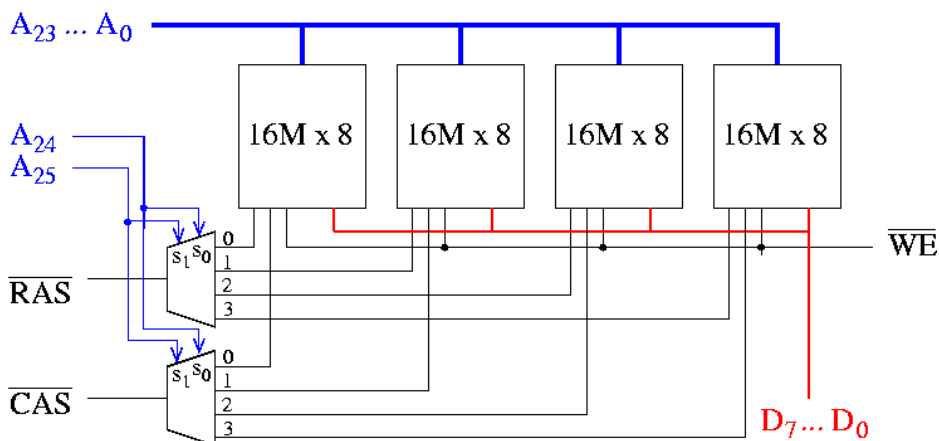


- auch die Speicherbandbreite (Anzahl übertragbarer Bytes je Sekunde zwischen CPU und Speicher) wird hierdurch erhöht!

3 DRAM (9)

2) Vergrößern des Adressraums:

- Datenleitungen und untere Adressleitungen an allen Bausteinen identisch; obere Adressleitungen dienen dem Multiplexen der Steuersignale
- Beispiel: Speicher mit 64M Worten à 8 Bit



- Möglichkeiten 1) und 2) werden i.a. kombiniert !