

6 Speicherglieder

6.1 Begriffe und Kenngrößen

6.1.1 Speicherelement

Ein Speicherelement muss **zwei physikalische Zustände** besitzen, die den binären Werten 0 und 1 zugeordnet werden können. Beide Werte müssen von außen verändert werden können (Schreiben bzw. Speichern). Außerdem müssen die Speicherelemente nach bestimmten Kriterien wieder aufgefunden werden können (z.B. durch Adresse oder Informationsinhalt).

6.1.2 Speicherkapazität

Die Speicherkapazität gibt die Anzahl der Speicherelemente oder Zellen im Speicher an. Die gebräuchlichen Einheiten sind **Bit** und **Byte**, wobei 8 Bits 1 Byte entsprechen. Größere Speicherkapazitäten werden mit Größenfaktoren **Ki** (kibi), **Mi** (mebi) und **Gi** (gibi) bezeichnet. Hierbei gilt, abweichend vom Dezimalsystem (Kilo = 10^3 , Mega = 10^6 , Giga = 10^9):

$$\text{Ki} = 2^{10} = 1024^1 = 1.024$$

$$\text{Mi} = 2^{20} = 1024^2 = 1.048.576$$

$$\text{Gi} = 2^{30} = 1024^3 = 1.073.741.824$$

Die Abweichung beruht auf der dualen Speicheradressierung.

6.1.3 Speicherorganisation

Die Speicherorganisation beschreibt die **Speicherkapazität** und die **Auswahlmöglichkeit**.

6.1.4 Arbeitsgeschwindigkeit

Die Arbeitsgeschwindigkeit kann je nach Speichertyp durch die Größen **Zugriffszeit**, **Zykluszeit** und **Datenrate/Übertragungsgeschwindigkeit** charakterisiert werden.

Zugriffszeit t_z

Die Zugriffszeit ist die Zeitspanne zwischen dem Adressieren einer Speicherzelle bis zum Anstehen der gesuchten Daten am Ausgang.

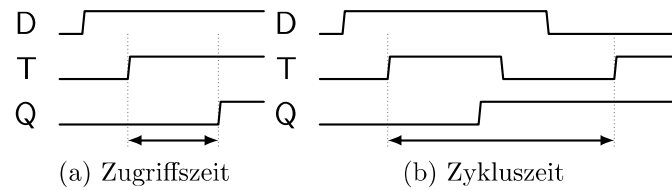


Abbildung 6.1: Ein-Bit-Speicherzelle

Zykluszeit t_c

Die Zykluszeit entspricht dem minimalen Zeitabstand mit dem zwei aufeinanderfolgende Daten eingeschrieben bzw. ausgelesen werden können. Im Grenzfall kann die Zykluszeit gleich der Zugriffszeit werden. Im Allgemeinen ist sie aber größer.

Datenrate oder Übertragungsgeschwindigkeit

Die Datenrate entspricht der Geschwindigkeit, mit der Daten ausgelesen und eingeschrieben werden können (meistens **Kehrwert der Zykluszeit**).

6.1.5 Weitere Kenngrößen

Neben den Kosten pro Bit (Bitpreis) wird Speicher auch nach der Speicherdichte klassifiziert, wobei letzteres in Bit pro mm^2 angegeben wird. Zudem gibt es dann noch die Zuverlässigkeit des Speichers. Diese wird durch die Ausfallrate und das Signal-Rauschverhältnis (engl.: Signal-to-Noise-Ratio [SNR]) angegeben.

6.2 Klassifizierung digitaler Speicher

6.2.1 Einteilung nach Arbeitsgeschwindigkeit und Kapazität

Die Hauptparameter **Arbeitsgeschwindigkeit**, **Kapazität** und **Kosten** können nicht gleichzeitig optimiert werden. Hohe Arbeitsgeschwindigkeit bedeutet beispielsweise hohe Kosten und wird meist mit Speichern kleinerer Kapazität realisiert. Es ist daher je nach Anwendungsfall ein Kompromiss zu schließen. Für Rechenanlagen wird ein **hierarchisches Speichersystem** eingesetzt.

6.2.2 Einteilung nach der Art des Zugriffs

Informationen können im Wesentlichen durch zwei Arten wiederaufgefunden werden. Zum einen kann durch Adressierung eines ortsadressierten Speichers der Inhalt erreicht werden. Die andere Möglichkeit ist das Auffinden von Dateninhalten in inhaltsadressierten Speichern (engl.: content addressed memory [CAM]).



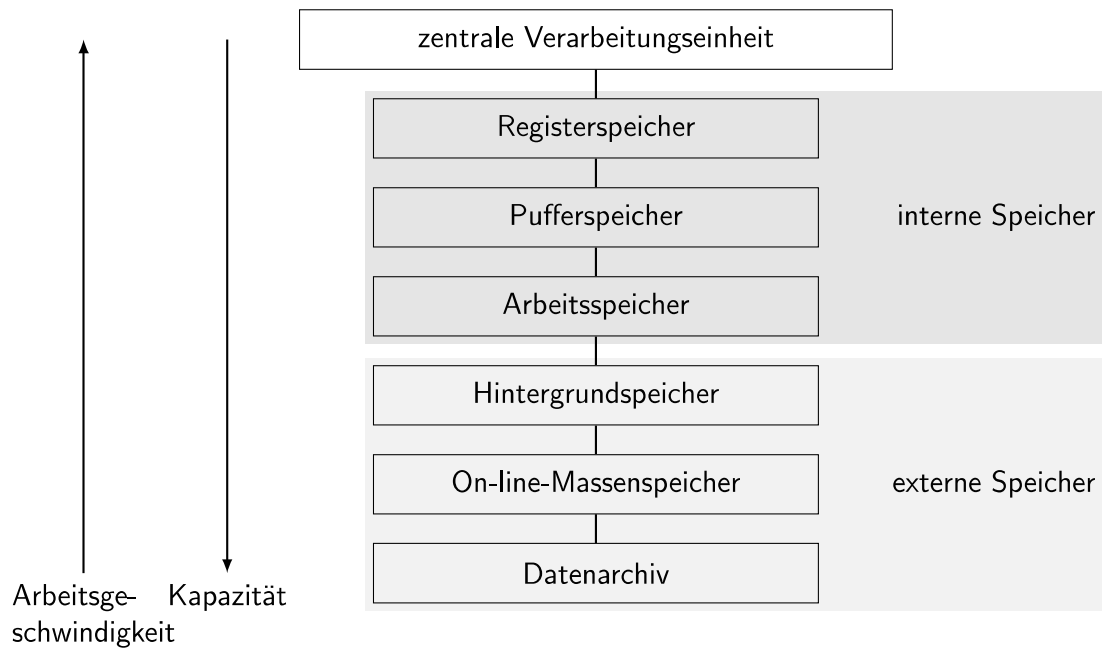


Abbildung 6.2: Aufbau der Speicherhierarchie

Adressierung

Beim Wiederauffinden der Information durch Adressierung wird zwischen **wahlfreiem Zugriff** (random access) und **seriellem Zugriff** (serial access) unterschieden.

Beim Random-Access kann jede Speicherzelle unabhängig von den Anderen adressiert, beschrieben und gelesen werden. Diese Zugriffsart ermöglicht die kürzeste Zugriffszeit. Sie findet beispielsweise im Hauptspeicher von Rechner Verwendung.

Der seriellen Zugriff ist abhängig von einer bestimmter Reihenfolge, so dass der Speicherplatz nicht direkt erreichbar ist. Der Durchlaufspeicher ist ein Beispiel dafür. Die zuerst angekommenen Daten werden auch wieder zuerst ausgegeben (first-in-first-out-Prinzip [FIFO]).

Dateninhalt

Typische Vertreter dieser **Assoziativspeicher** sind beispielsweise Caches und Renaming-Register wie sie in superskalaren Prozessoren verwendet werden aber auch das menschliche Gehirn.

6.2.3 Einteilung nach Art des Datenverkehrs

Schreib-Lese-Speicher

Die Daten können **mehrfach** gespeichert (Schreiben) und wieder abgerufen (Lesen) werden.

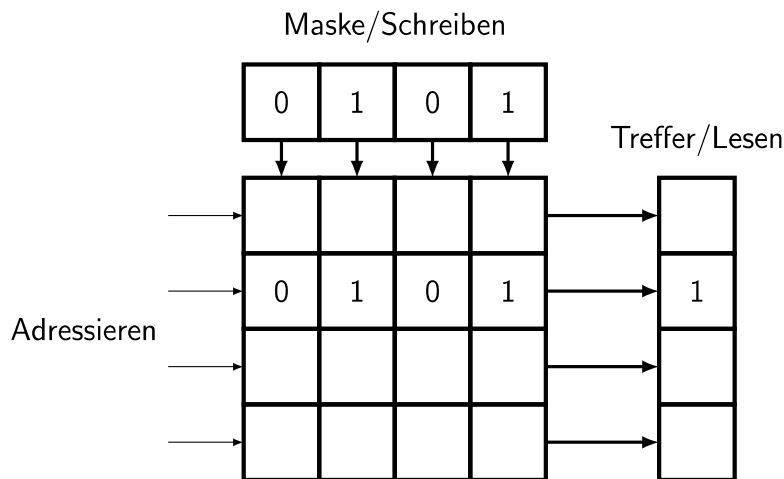


Abbildung 6.3: Zugriff auf inhaltsadressierte Speicher

Lesespeicher/Festwertspeicher

Das Einschreiben erfolgt einmalig, während des Herstellungsprozesses. Diese Speicher nennt man (nur) Lesespeicher (engl.: read only memory [ROM]). Ist diese einmalige Einspeicherung durch den Anwender möglich, so spricht man von programmierbaren Lesespeichern (programmable ROM, PROM, EPROM). Wenn die Einspeicherung ein reversibler Prozess ist, also mehrmals möglich ist, dann werden solche Speicher als EAROM, EEROM (electrical erasable ROM) bezeichnet (z. B. Flash-Speicher).

6.3 Halbleiterspeicher

6.3.1 Einleitung

Überblick

Der Bedarf große Datenmengen zu speichern führte in den letzten 30 Jahren zur Entwicklung zahlreicher Technologien. Bei Halbleiterspeichern konnten große Fortschritte bei der **Kapazitätserhöhung** und der **Kostensenkung pro Bit** erzielt werden. Die Kosten je Bit haben sich seither auf unter ein 1/100 reduziert und liegen 2007 etwa bei 10 Euro je GiBit. Die Kapazität hat 2GiByte je Chip erreicht.

Ein weiterer Grund für den Einsatz von Halbleiterspeichern ist die Kompatibilität zwischen Speicher und Prozessor. Hierbei sind die jeweiligen Betriebsspannungen und Pegel gleich, da die Speicher und Prozessoren in der gleichen Technologie hergestellt sind. Dadurch entfallen teure und störanfällige Interfaceschaltungen. Zusätzlich besteht die Möglichkeit größere Speichersysteme modular aus kleineren Bausteinen aufzubauen.

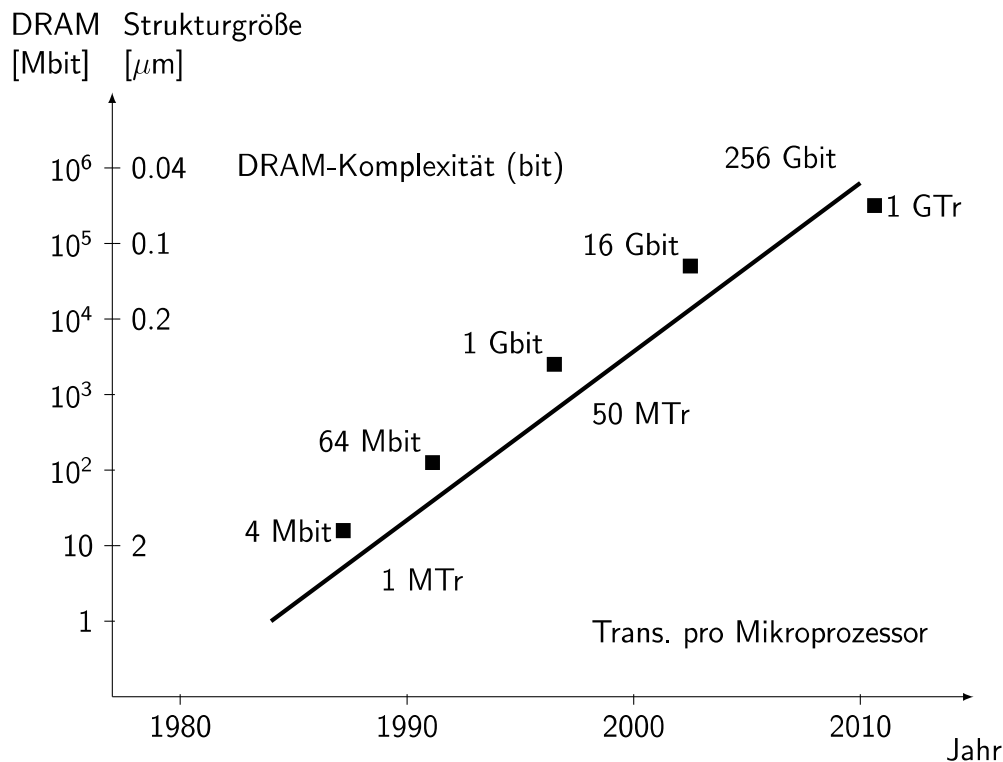


Abbildung 6.4: Überblick der Technologie-Generationen

Realisierung unterschiedlicher Organisationsformen

Die **Schreib-Lese-Speicher** können **statisch** oder **dynamisch** ausgeführt sein. Die statischen Zellen halten die Information solange, wie Versorgungsspannung anliegt. Die dynamischen Zellen erfordern in bestimmten Zeitabständen sogenannte Refresh-Impulse, damit die Daten nicht verloren gehen. **Festwertspeicher** sind **nichtflüchtige Speicher**.

Statische Halbleiterspeicher

Statische Halbleiterspeicher besitzen **kurze Zugriffszeiten**. Im Vergleich zu dynamischen Speichern ist aber für die Realisierung ein **höherer Schaltungsaufwand** notwendig. Anwendung finden die statische Halbleiterspeicher hauptsächlich in Speicherbausteinen kleiner und mittlerer Kapazität. Ihr Einsatz wird vorrangig durch eine kurze Zugriffszeit bestimmt. Realisiert werden diese Speicher in **Bipolar- und MOS-Technik**.

Dynamische Halbleiterspeicher

Dynamische Halbleiterspeicher kommen in komplexen Speichersystemen wegen ihrer hohen Integrationsdichte zum Einsatz. Hergestellt werden diese Speicher nur in **MOS-Technik**.

6.3.2 Bistabile Kippstufe

Funktionsprinzip der statischen Speicherung



Realisiert werden statische Speicher mit zwei Invertern, deren Ausgänge auf die Eingänge der jeweils anderen Stufe zurückgeführt sind. Diese Schaltung mit zwei kreuzgekoppelten Invertern wird als **bistabile Kippstufe** oder als **Flipflop** bezeichnet.

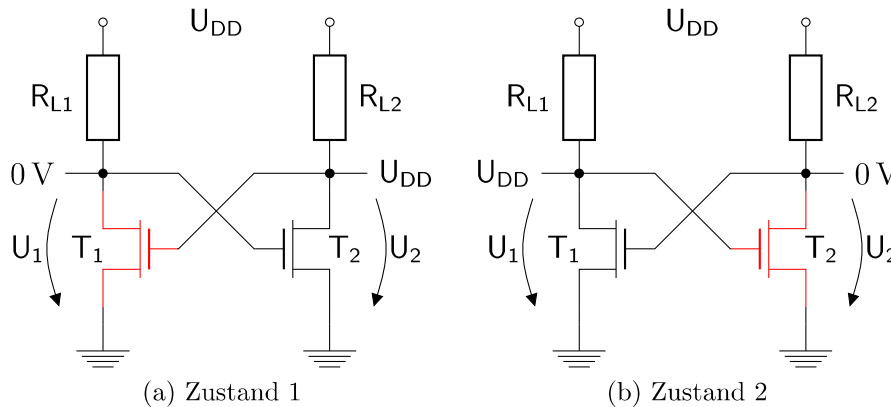


Abbildung 6.5: Bistabile Kippstufe

Funktionsweise der Kippstufe (Abbildung 6.5)

Für die Anfangsbetrachtung sei T_1 leitend, U_1 wird auf etwa 0 V gezogen. Diese Spannung liegt gleichzeitig am Gate von T_2 an, der somit sperrt. Wegen des Widerstandes R_{L2} stellt sich die Spannung U_2 auf U_{DD} ein. Da diese Spannung auch am Gate von T_1 anliegt, bleibt T_1 leitend.

6.3.3 Dynamischer MOS-Speicher

Funktion

Die einfachste dynamische Speicherzelle ist das 1-Transistorelement. Über den Transistor T wird eine Kapazität C geladen und die Spannung an ihr wieder abgefragt.

Schreiben

Der Transistor T wird mit einer „1“ auf der Wortleitung leitend und die Kapazität C wird auf den Pegel der Datenleitung aufgeladen.

Lesen

Der Transistor T wird mit einer „1“ auf der Wortleitung leitend und die Ladung der Kapazität C wird abgefragt. War eine „0“ gespeichert, befand sich keine Ladung auf C .

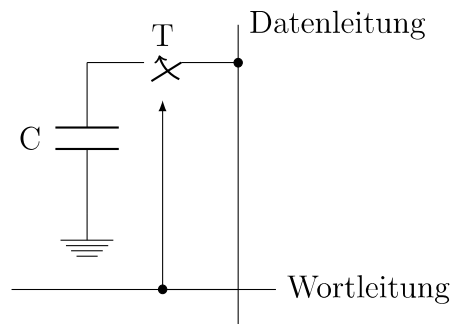


Abbildung 6.6: Dynamische 1-Transistor-Zelle

Da beim Lesen der Ladung von C entnommen wird, muss sie anschließend regeneriert werden.

6.3.4 Ein-Bit-Flipflopspeicher

Asynchrones RS-Flipflop

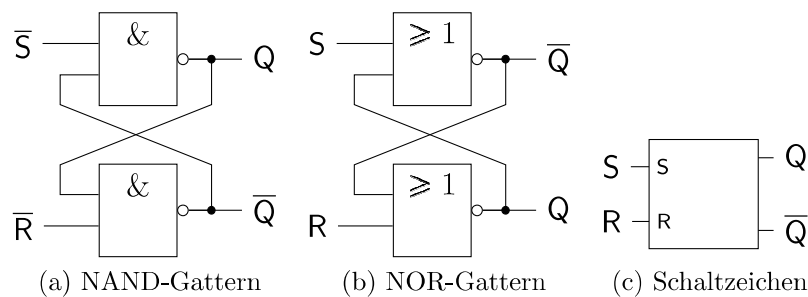


Abbildung 6.7: RS-Flipflop realisiert mit unterschiedlichen Gattertypen

Als Speicherglied wird eine bistabile Kippstufe aus zwei kreuzgekoppelten NAND- oder NOR-Gattern eingesetzt. Jeweils ein Eingang der beiden Gatter wird für die Rückkopplung verwendet, die freien Eingänge dienen zum **Setzen (S)** bzw. **Rücksetzen (R)**. In der Praxis hat dieses RS-Flipflop als Einzelelement zur Speicherung wenig Bedeutung. Der Grund dafür liegt im asynchronen Verhalten, denn die Information wird geladen bzw. gelöscht, sobald am Setz- oder Rücksetzeingang ein Signal anliegt. Ein weiterer Nachteil ist durch die verbotenen Eingangskombinationen (NAND-Gliedern: $\bar{R} = \bar{S} = 0$, NOR-Gliedern $R = S = 1$) gegeben.

Unzulässige Eingangskombination (NOR)

Der Zustand der Ausgänge Q und \bar{Q} ist nicht mehr negiert zueinander. Beide Eingangstransistoren werden für $R = S = 1$ leitend und schalten den „L“-Pegel auf beide Ausgänge.

R	S	Q	\overline{Q}
0	0	Q	\overline{Q}
0	1	1	0
1	0	0	1
1	1	*	*

Abbildung 6.8: Wahrheitstabelle für NOR-Flipflop

Die mögliche Eingangskombination $R = 1, S = 1$ ergibt eine mehrdeutige Ausgangssituation $Q = 0, \overline{Q} = 0$ und ist deshalb **unzulässig**.

Getaktetes RS-Flipflop

Bei synchronen Schaltwerken soll eine bereits anliegende Information erst zu einem bestimmten Zeitpunkt in das Flipflop übernommen werden. Dieses Verhalten wird durch die jeweilige Verknüpfung des R- und S-Einganges mit einem **Steuertakt T** erzielt.

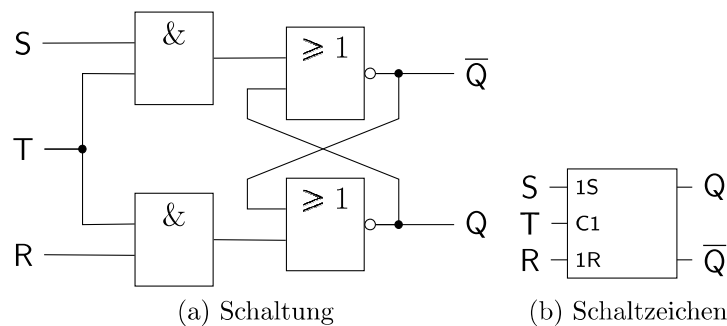


Abbildung 6.9: Getaktetes RS-Flipflop

Die Wahrheitstabelle bleibt dieselbe wie für das asynchrone RS-Flipflop. Sie wird lediglich durch die Bedingung erweitert, dass sie nur für den Zeitraum $T = 1$ gilt.

Getaktetes D-Flipflop (Latch)

Der unzulässige Zustand $R = S = 1$ (bei NOR) bzw. $\overline{R} = \overline{S} = 0$ (bei NAND) lässt sich durch eine Beschaltung des Rückstelleingangs mit negiertem Signal des Setzeinganges umgehen. Die Kombination $\overline{R} = \overline{S} = 0$ bzw. $R = S = 1$ wird so prinzipiell vermieden. Dieses Flipflop bezeichnet man als **Delay-Flipflop** oder **Latch**.

Master-Slave-D-Flipflop

Das einfache D-Latch weist den Nachteil auf, dass während $T = 1$, sich jede Änderung des D-Einganges auf den Ausgang überträgt (**Transparenz**). Dies kann durch Anwendung des **Master-Slave-Prinzips** vermieden werden. Es werden zwei Flipflops kaskadiert, wobei das zweite mit negiertem Takt angesteuert wird. Dadurch wird gewährleistet, dass immer nur eines der beiden Flipflops Daten übernehmen kann.

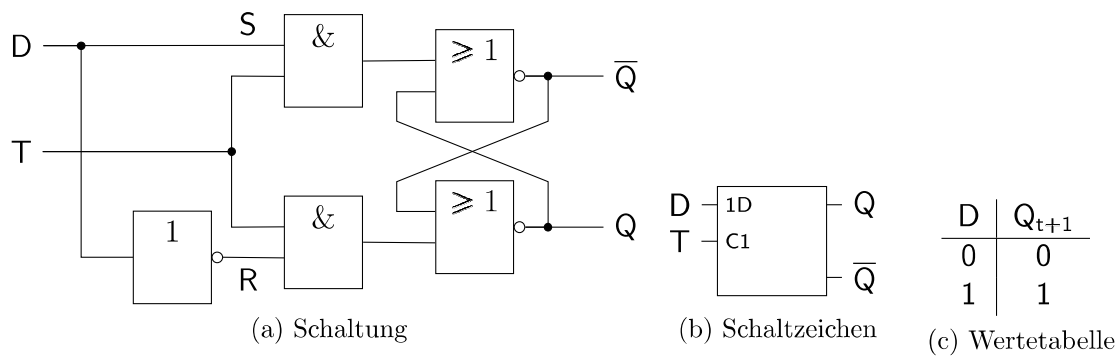


Abbildung 6.10: Getaktetes D-Flipflop

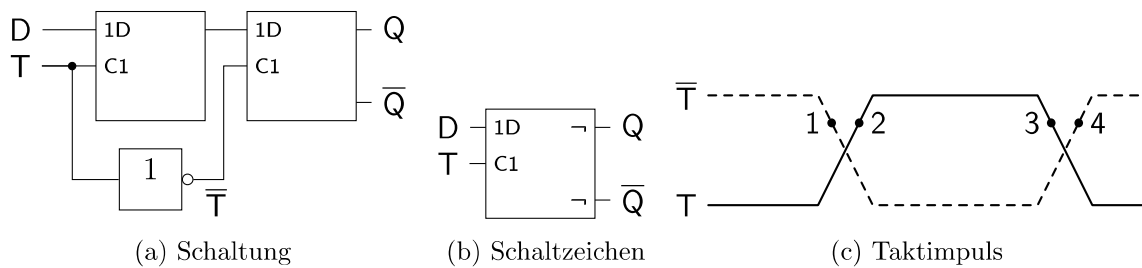


Abbildung 6.11: Master-Slave-D-Flipflop

Erläuterung der Arbeitsweise des MS-Flipflops am Verlauf der Taktflanken

Punkt	Beschreibung
1	Sperren der Eingänge des Slave-Flipflops
2	Übernahme der Information in das Master-Flipflop
3	Sperren des Dateneingangs des Master-Flipflop
4	Übernahme der Information vom Master-Flipflop in das Slave-Flipflop

Zwischen dem Setzen des Master-Slave-Flipflops und dem Anliegen der Daten an den Ausgängen vergeht eine halbe Taktperiode. Da die Übernahme vom Master zum Slave während einer Taktflanke erfolgt, werden Master-Slave-Flipflops als **taktflankengesteuerte Flipflops** bezeichnet.

6.3.5 Speicherorganisation eines SRAMs

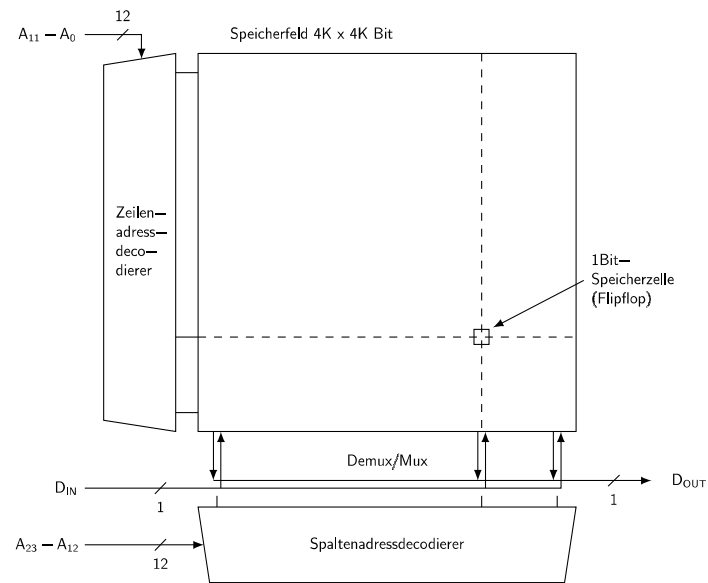


Abbildung 6.12: Speicherorganisation eines SRAMs