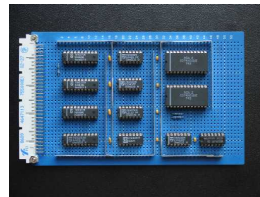


# Digital- och datorteknik



## Föreläsning #21

Biträdande professor Jan Jonsson

Institutionen för data- och informationsteknik  
Chalmers tekniska högskola

# Minnestyper

## Teknologier och hierarkier

Vi har hittills betraktat primärminnesmodulen i FLIS-processorn på ett ganska abstrakt plan. När man studerar minnesmoduler mer generellt, och i mer detalj, finner man att det finns många aspekter som kan beaktas. Vi fokuserar här på teknologier och hierarkier.

- Minnesteknologin anger hur data kan skrivas till, lagras i och läsas från minnesmodulen.
- Minneshierarkin beskriver hur man använder teknologierna för olika syften utgående från deras relativa snabbhet och lagringskapacitet.

# Minnestyper

## Teknologier och hierarkier

Ett sätt att betrakta minnesmoduler är att undersöka hur lätt det är att komma åt innehållet på en viss minnesadress. Här brukar man använda åtkomsttiden som måttstock, d v s den tid det tar från det att processorn begärt läsning av data på en viss adress till dess processorn fått dessa data.

- Hur förutsägbar ett minnes åtkomsttid är beror på teknologin. Exempelvis har ett sekventiellt minne (magnetband) olika åtkomsttid beroende på var i minnet data finns. Som kontrast kan man nämna att ett random access-minnes åtkomsttid alltid är samma oavsett var i minnet data finns.

# Minnestyper

## Teknologier och hierarkier

Ett sätt att betrakta minnesmoduler är att undersöka hur lätt det är att komma åt innehållet på en viss minnesadress. Här brukar man använda åtkomsttiden som måttstock, d v s den tid det tar från det att processorn begärt läsning av data på en viss adress till dess processorn fått dessa data.

- Storleken på åtkomsttiden för ett minne avgör på vilken nivå i minneshierarkin som minnet bör användas. Exempelvis bör minnesmodulerna i ett cacheminne generellt sett ha en mycket kortare åtkomsttid än en minnesmodul i ett primärminne.

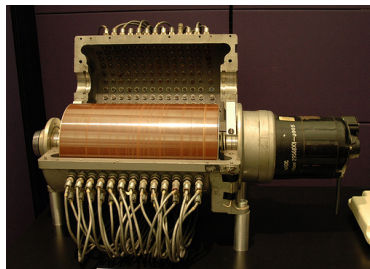


# Minnesteknologi

## Historia

Innan halvledartekniken slog igenom på allvar användes ett antal olika metoder för att lagra data. Här är några exempel:

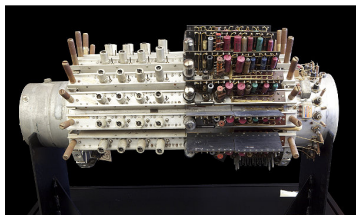
Trumminne



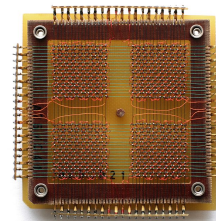
Williams-Kilburn tube



Mercury delay line

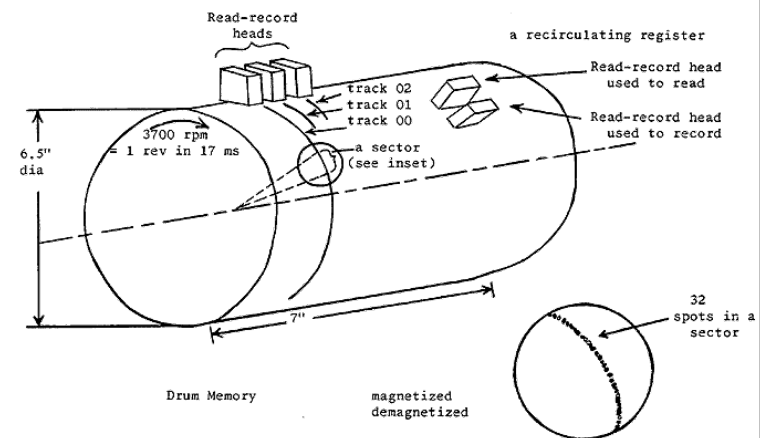
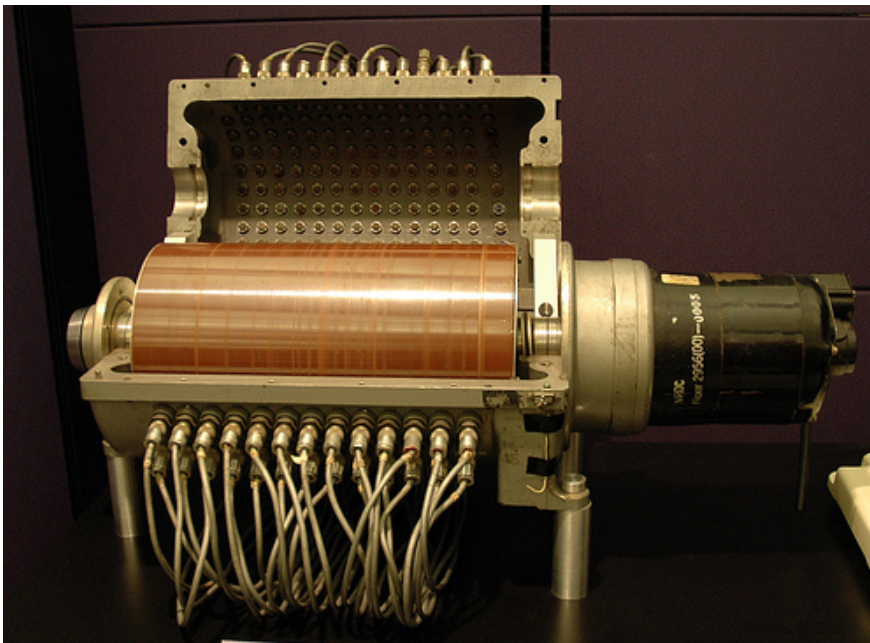


Magnetiskt kärnminne



# Minnesteknologi

## Historia: trumminne ("drum memory")



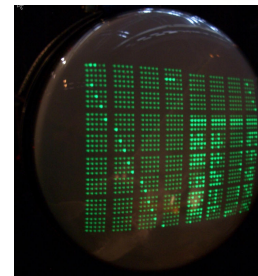
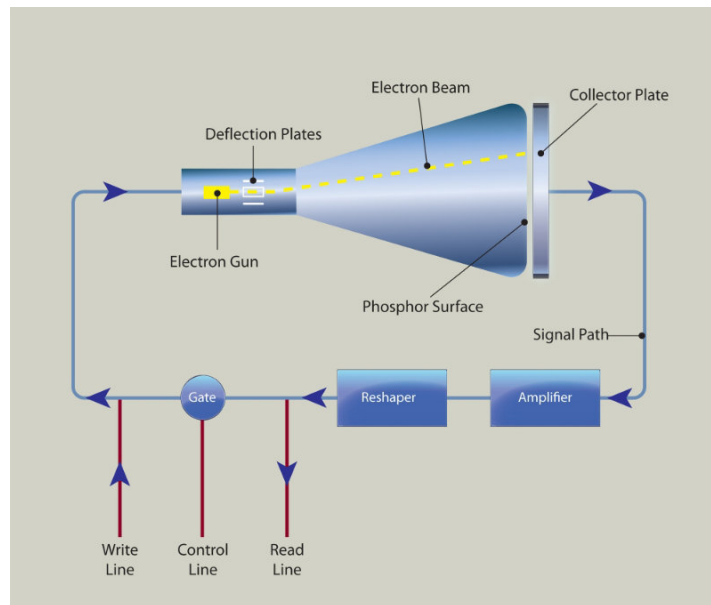
Föregångare till hårddisken. Bevarar lagrat data även vid avstängt system.

Tidsepok: 1940-talet till 1960-talet

Lagringskapacitet: 10–100 kbyte. Åtkomsttid: 5–50 ms

# Minnsteknologi

## Historia: Williams-Kilburn tube



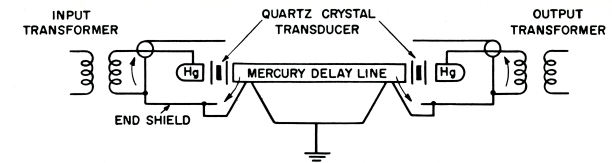
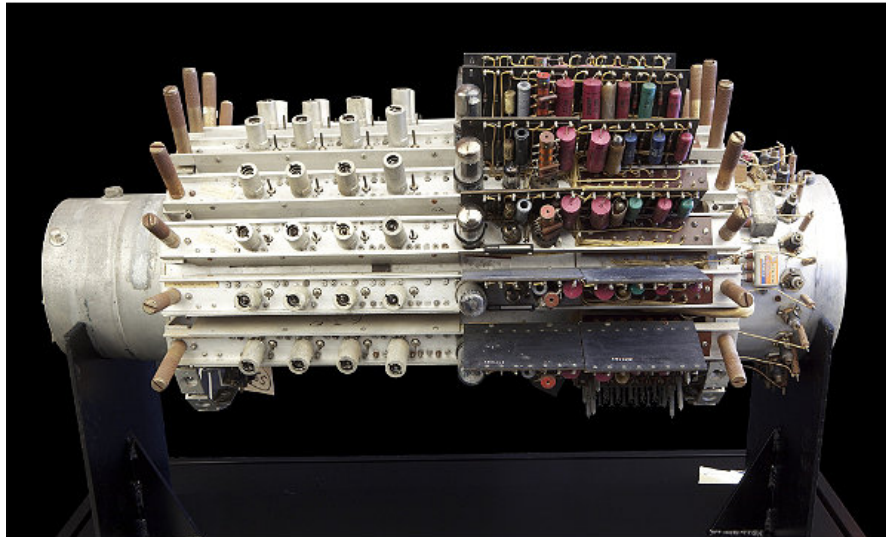
Bevarar data (via recycling) endast så länge systemet är påslaget.

Tidsepok: 1940-talet

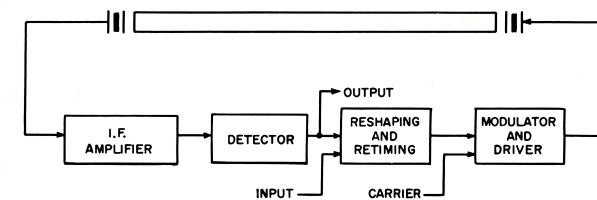
Lagringsskapacitet: 256 byte. Åtkomsttid: 10–50  $\mu$ s

# Minnsteknologi

## Historia: mercury delay line



*Schematic diagram of circuit connections to the acoustic delay line used in NBS mercury memory.*



*Block diagram of the mercury memory system.*

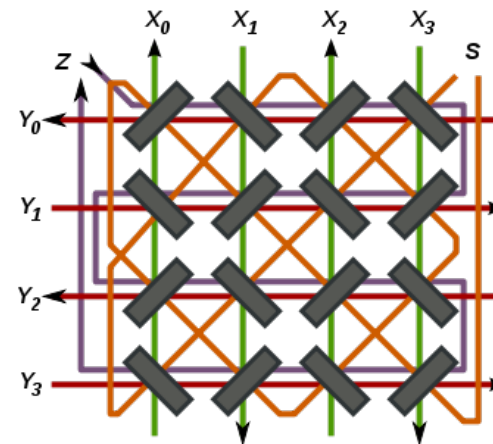
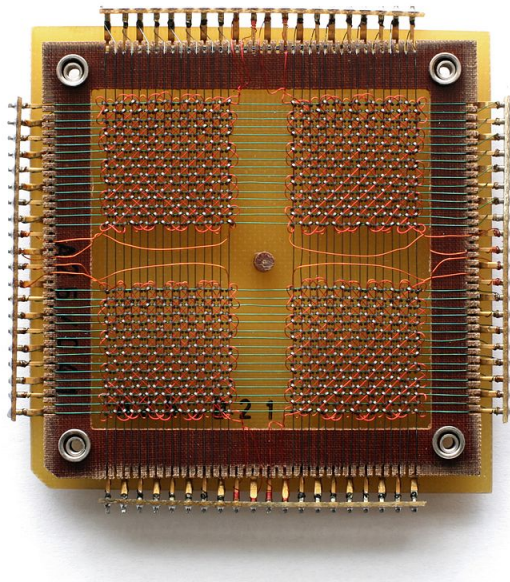
Bevarar data (via recycling) endast så länge systemet är påslaget.

Tidsepok: 1940-talet

Lagringskapacitet: 1–10 kbyte. Åtkomsttid: 100–200  $\mu$ s

# Minnesteknologi

Historia: magnetiskt kärnminne ("core memory")



Bevarar lagrat data även vid avstängt system.

Tidsepok: 1950-talet till 1970-talet

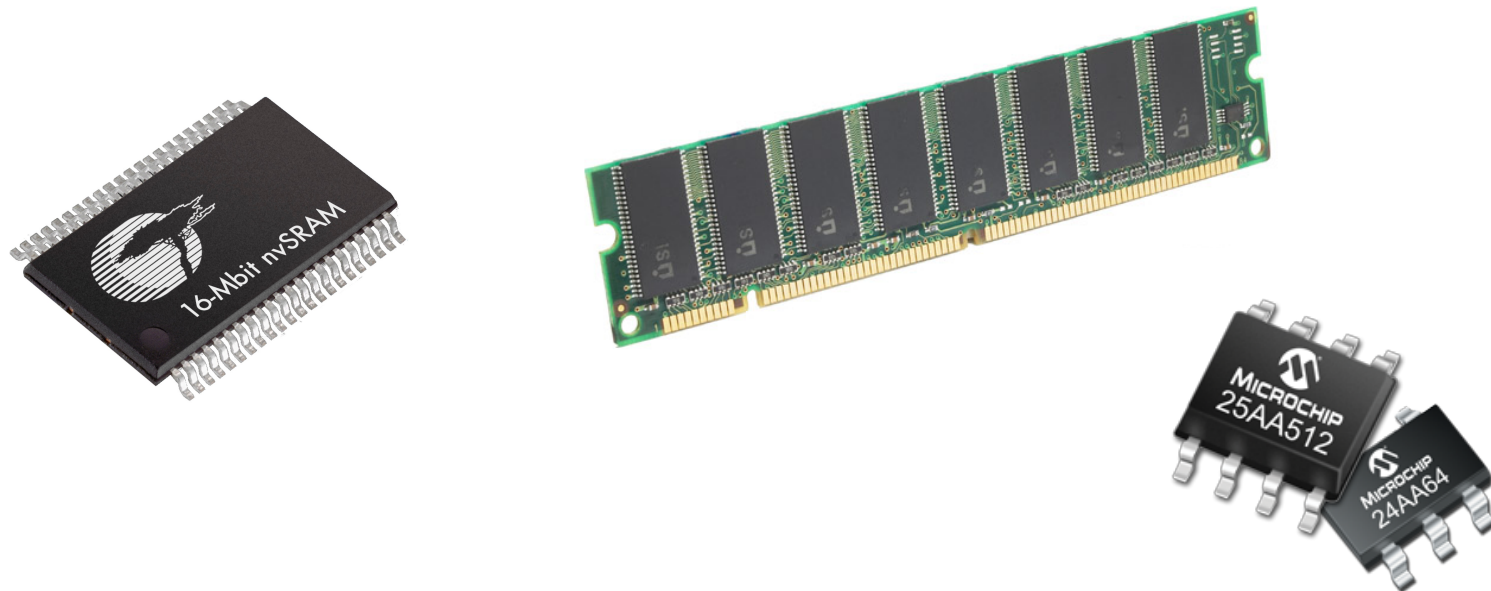
Lagringskapacitet 1–10 kByte. Åtkomsttid: 1–5  $\mu$ s



# Minnesteknologi

## Random access-minne (RAM)

För denna typ av minnen är åtkomsttiden alltid den samma oavsett vilken del av minnet som adresseras. Denna egenskap krävs hos minnen som skall kunna hantera data i samma takt som processorn arbetar.



# Minnesteknologi

## Random access-minne (RAM)

För denna typ av minnen är åtkomsttiden alltid den samma oavsett vilken del av minnet som adresseras. Denna egenskap krävs hos minnen som skall kunna hantera data i samma takt som processorn arbetar. Minnen av typen RAM förekommer i två varianter: flyktiga ("volatile") och icke-flyktiga ("non-volatile").

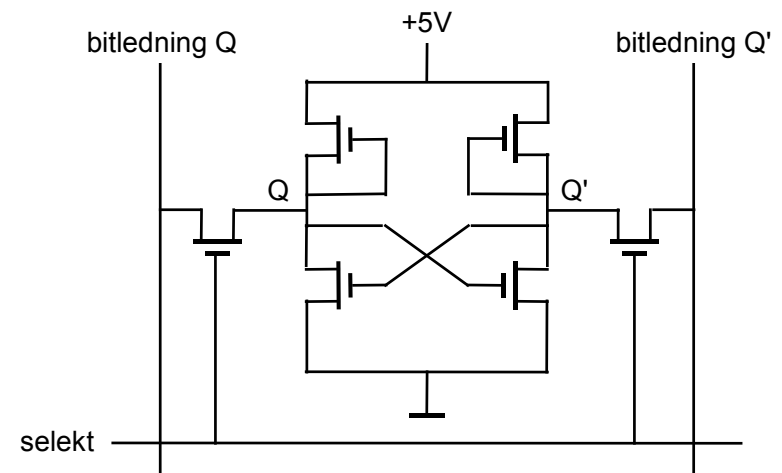
- De flyktiga minnena förlorar sitt informationsinnehåll när systemets matningsspänning slås av. Denna variant kallas också för RWM (read write memory). Exempel på sådana minnen är SRAM och DRAM.
- De icke-flyktiga minnena behåller sitt informationsinnehåll när systemets matningsspänning slås av. Exempel på sådana minnen är PROM, EEPROM och Flashminnen.

# Minnesteknologi

## Statiskt RAM (SRAM)

Hos SRAM består en bitcell av en latch som går att ettställa och nollställa. Latchen byggs vanligtvis upp med sex nMOS-transistorer. Cellen kallas för statisk då den behåller sitt värde så länge den har matningsspänning.

Innehållet i cellen kan läsas eller ändras genom att aktivera selektledningen och därefter på lämpligt sätt använda bitledningarna.



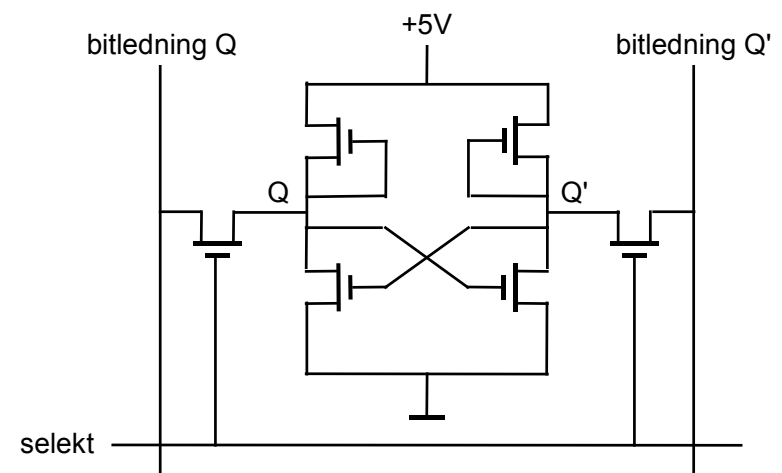


# Minnesteknologi

## Statiskt RAM (SRAM)

Hos SRAM består en bitcell av en latch som går att ettställa och nollställa. Latchen byggs vanligtvis upp med sex nMOS-transistorer. Cellen kallas för statisk då den behåller sitt värde så länge den har matningsspänning.

Åtkomsttiden hos SRAM ligger i dagsläget på 1–10 ns, vilket gör minnestypen väl lämpad att hantera data för de snabbaste minnesoperationerna hos en modern mikroprocessor.



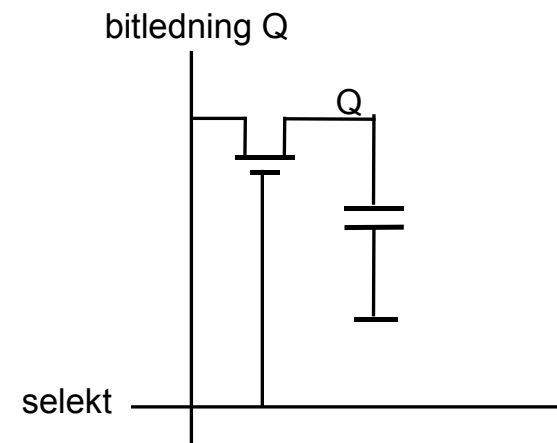
# Minnesteknologi

## Dynamiskt RAM (DRAM)

Hos DRAM består en bitcell endast av en nMOS-transistor och en kondensator. Cellen kallas för dynamisk då den kontinuerligt behöver återladdas ("refresh") för att behålla sitt värde.

När kondensatorn är bortkopplad från bitledningen så "läcker" den långsamt, d v s laddningen försvinner långsamt.

Funktionen som bitcell bygger därför på att kondensatorn används ofta, d v s laddas upp (med "1") eller ur (med "0") ofta.



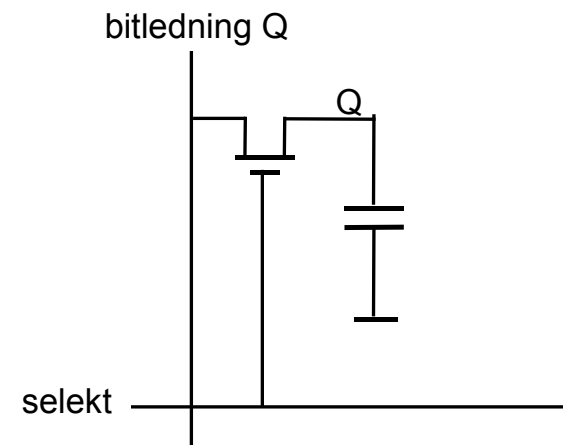
# Minnesteknologi

## Dynamiskt RAM (DRAM)

Hos DRAM består en bitcell endast av en nMOS-transistor och en kondensator. Cellen kallas för dynamisk då den kontinuerligt behöver återladdas ("refresh") för att behålla sitt värde.

nMOS-transistorn fungerar som switch och möjliggör att man kan komma åt kondensatorn. Via selekt kan då kondensatorn kopplas till eller från bitledningen.

Vid läsning laddas kondensatorn ur, och det lästa värdet måste alltså återskrivas i cellen, d v s kondensatorn måste återladdas.



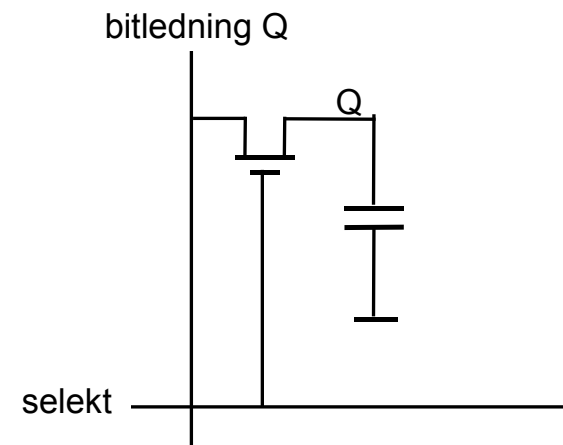
# Minnesteknologi

## Dynamiskt RAM (DRAM)

Hos DRAM består en bitcell endast av en nMOS-transistor och en kondensator. Cellen kallas för dynamisk då den kontinuerligt behöver återladdas ("refresh") för att behålla sitt värde.

Åtkomsttiden hos DRAM ligger typiskt på 50–150 ns, vilket gör att minnestypen inte kan hantera data i den takt som de allra snabbaste mikroprocessorerna vill arbeta.

Å andra sidan erbjuder DRAM en betydligt högre packningstäthet för bitcellerna, jämfört med SRAM (som har 6 gånger fler transistorer per cell.)



# Minnesteknologi

## Läsminnen ("read-only memory", ROM)

Denna typ av minnen kännetecknas av att (1) deras innehåll läses betydligt oftare än det behöver uppdateras och (2) att de behåller sitt innehåll även efter bortfall av matningsspänning. Beroende på tillämpningens behov av att uppdatera minnets innehåll använder man olika varianter av läsminnen.

- Programmerbara ROM (PROM): Läsminne som bara kan programmeras en gång och därefter aldrig ändra sitt innehåll.
- Raderbara PROM (EPROM, EEPROM, Flashminnen): Läsminne som kan raderas och omprogrammeras ett stort antal gånger.

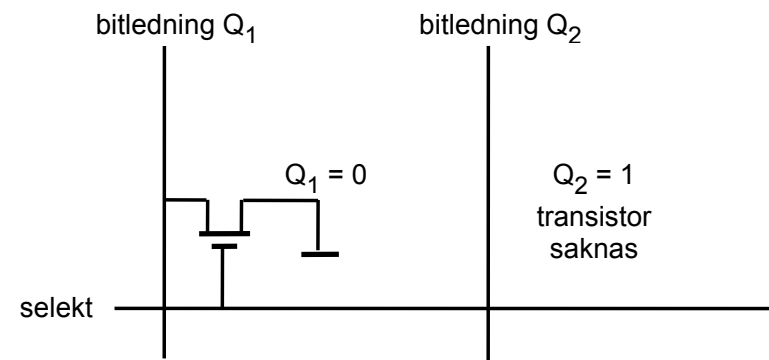
Åtkomsttiden hos läsminnen ligger typiskt på 100–150 ns.

# Minnesteknologi

## Programmerbart ROM (PROM)

Hos PROM består en bitcell av en enstaka nMOS-transistor ifall innehållet i cellen skall vara "0". Om innehållet i cellen skall vara "1" saknas det en transistor. Innehållet i cellen kan läsas genom att aktivera selektledningen. Värdet på bitledningen är "1" såvida det inte finns en transistor som lägger värdet "0" på ledningen.

Programmering av minnet sker med en speciell utrustning där man med en hög spänning "bränner bort" transistorer i de bitceller som skall ha innehållet "1".



# Minnesteknologi

## Raderbart PROM ("Erasable PROM", EPROM)

Hos EPROM består en bitcell av en speciell transistor där närvaro av en elektrisk laddning på transistorns styrgrind bestämmer om läsning av cellen skall ge värdet "1" eller "0".

Programmering av minnet sker med en speciell utrustning där man med en hög spänning placerar en elektrisk laddning på styrgrindarna i de bitceller som skall ha innehållet "0".

De elektriska laddningarna kan tas bort genom att belysa minnet (via ett fönster på kretsens kapsel) med UV-ljus under en längre period.



# Minnesteknologi

## Elektriskt raderbart PROM (EEPROM)

EEPROM liknar på många sätt EPROM i sin uppbyggnad, och programmeras på ett liknande sätt. Skillnaderna mellan de två typerna av minnen är:

- EEPROM raderas på ett liknande sätt som de programmeras, d v s genom att applicera en hög spänning. Belysning med UV-ljus behövs alltså inte.
- EEPROM kan programmeras ”på plats” i datorn, eftersom de nödvändiga spänningarna genereras internt i kretsen. Någon speciell programmeringsutrustning behövs alltså inte.
- EEPROM behöver inte programmeras om i sin helhet, utan man kan välja att enbart ändra i utvalda delar av minnet.



# Minnesteknologi

## Flashminnen

Ett flashminne är i stort sett ett EEPROM. Skillnaden mellan de två typerna av minnen ligger huvudsakligen i hur små de minnesareor kan vara som skall raderas och omprogrammeras.

- EEPROM kan radera och omprogrammera minnesareor som kan vara så små som ett fåtal byte. Detta gör EEPROM speciellt lämpligt för lagring av t ex systemkonfigurationer där enstaka delar av konfigurationen kan behöva ändras ofta.
- Flashminnen kan normalt sett bara radera och programmera om minnesareor med en minsta storlek av ett "block" (typiskt 512 bytes). Detta gör flashminnen lämpliga som primärminne för lagring av programkod, eller som sekundärminne för lagring av filsystem.

# Minnesteknologi

## Blockminnen

För denna typ av minnen gäller att data endast är tillgängligt i form av block. Vanliga blockstorlekar är 512 byte och 4096 byte.

Exempel på blockminnen är hårddiskar, i form av roterande skivor av magnetiskt media, eller Solid State Discs (SSD) vilka är uppbyggda kring flashminnen.

Hårddiskens block ligger längs spår på skivorna. Åtkomsttiden blir därför inte konstant utan utgörs av den tid det tar att flytta läs/skrivarmen till rätt spår plus den tid det tar att vänta på och läsa in avsett block.

Genomsnittlig åtkomsttid hos en modern hårddisk är ca 10ms.



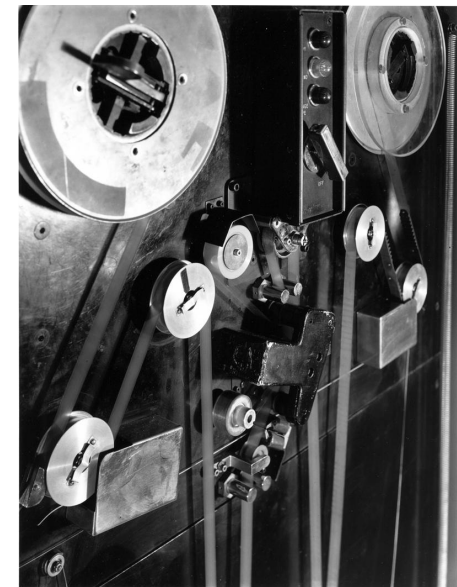
# Minnesteknologi

## Sekventiella minnen

Även för denna typ av minnen gäller att data är tillgängligt i form av block med ord. Blocken lagras i en lång sekvens i minnet.

Typexemplet på ett sekventiellt minne är en bandstation med ett magnetband. För att hitta ett speciellt block med data måste en rad läsningar, och fram- eller bakspolningar, ofta utföras.

Åtkomsttiden hos en bandstation kan sålunda variera väldigt mycket beroende på var det block som skall läsas ligger på magnetbandet.



# Minneshierarki

## Behovet av olika typer av minnen

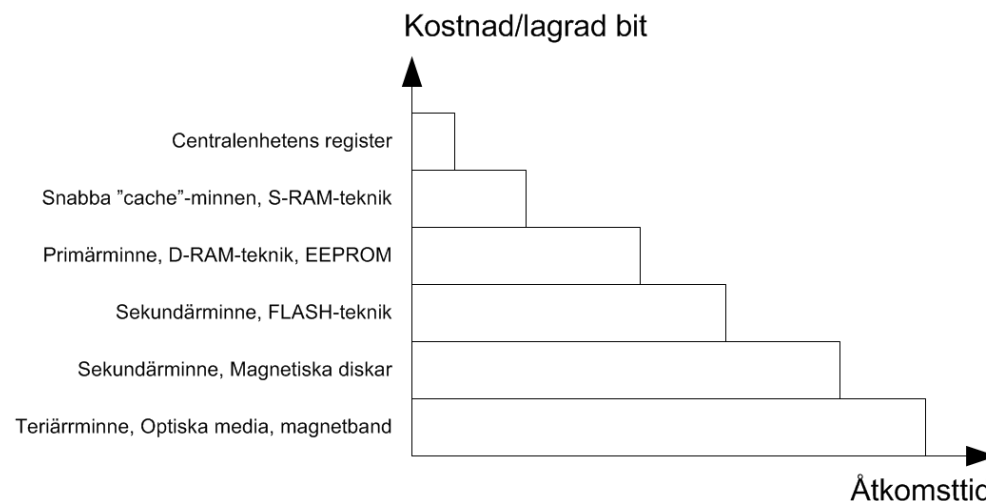
I en modern dator finner man behov av följande minnen:

- Cacheminne: ett snabbt minne (SRAM) som innehåller kopior av frekvent använda instruktioner och data från primärminnet.
- Primärminne: ett minne (DRAM, EEPROM) där processorns program och data lagras.
- Sekundärminne: blockminne för lagring av filsystem samt lagring av aktiv programkod och data som inte ryms i primärminnet (s k virtuellt minne).
- Tertiärminne: sekventiellt minne för långtidslagring av data, t ex säkerhetskopiering.

# Minneshierarki

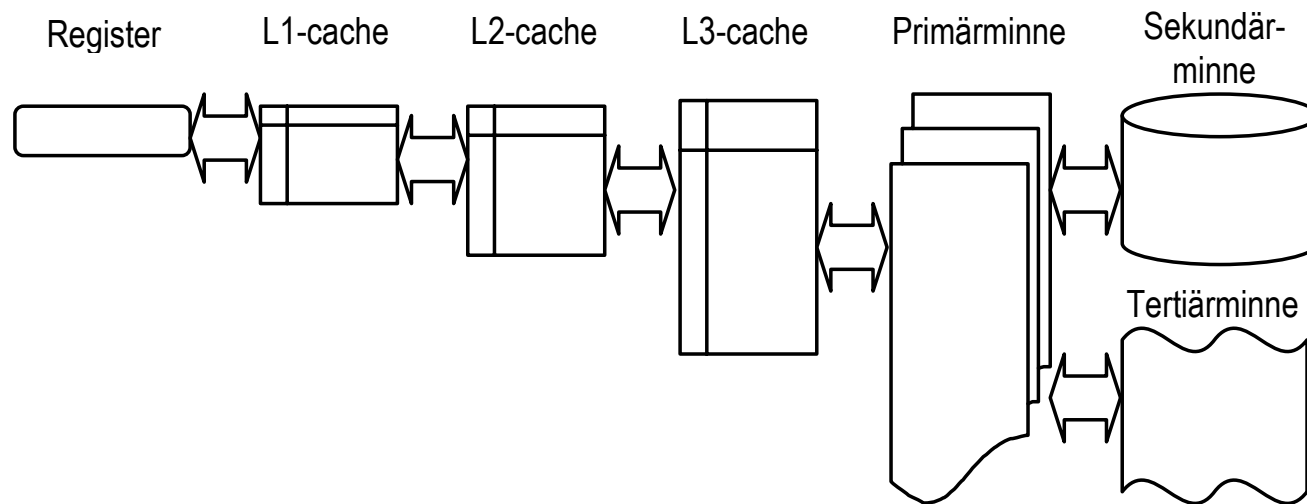
## Åtkomsttid vs lagringskapacitet vs kostnad

Allmänt gäller att ett minne där åtkomsttiden är kort kostar mer per lagringsposition än ett minne med längre åtkomsttider. Det leder till en minneshierarki där det finns ett relativt litet antal lagringspositioner i de snabbaste minnena och ett betydligt större antal lagringspositioner i de långsammare minnena.



# Minneshierarki

Exempel:



Karakteristika för de ingående minnestyperna:

Minnestyp	Register	L1-cache	L2-cache	L3-cache	Primärminne	Sekundärminne	Tertiärminne
Storlek	1 KB	64 KB	256 KB	2-4 MB	4-16GB	4-16 TB	1000 TB
Åtkomsttid	300 ps	1 ns	3-10 ns	10-20 ns	50-100 ns	5-10 ms	-minuter