



Calcolatori Elettronici I

A.A. 2019-2020

La memoria centrale di un calcolatore

Prof. Roberto Canonico

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione

Definizione di memoria

- Sistema organizzato come un insieme di registri (nel senso generale di “contenitori d’informazione”) sui quali sono definite 3 operazioni:
 - Scrittura
 - Posizionamento di una cella in un determinato stato o registrazione dell’informazione
 - Lettura
 - Rilievo dello stato di una cella o prelievo dell’informazione
 - Selezione
 - Individuazione di una cella al fine di eseguire una delle operazioni precedenti

Parametri di una memoria

- Capacità
 - Numero di dati elementari che la memoria può contenere
 - Si esprime indicando il numero complessivo di registri e la dimensione in bit di un singolo registro: $N \times p$ bit
- Tempo di accesso
 - Tempo necessario ad eseguire un'operazione di lettura/scrittura
 - È composto in generale da un tempo di *selezione* più un tempo di *trasferimento*
- Tempo di ciclo
 - Tempo minimo tra due operazioni consecutive
 - Tipicamente, il tempo di ciclo è superiore al tempo di accesso
 - In alcune memorie, l'operazione di lettura è distruttiva e quindi, dopo aver effettuato una lettura, occorre una riscrittura automatica del dato appena letto, che non rende immediatamente disponibile la memoria per un'altra lettura

In base alla **modalità di accesso** ai dati:

- Sequenziali: tempo di accesso dipendente dalla posizione del dato
- Casuali: tempo di accesso indipendente dalla posizione del dato

In base alle **operazioni consentite**:

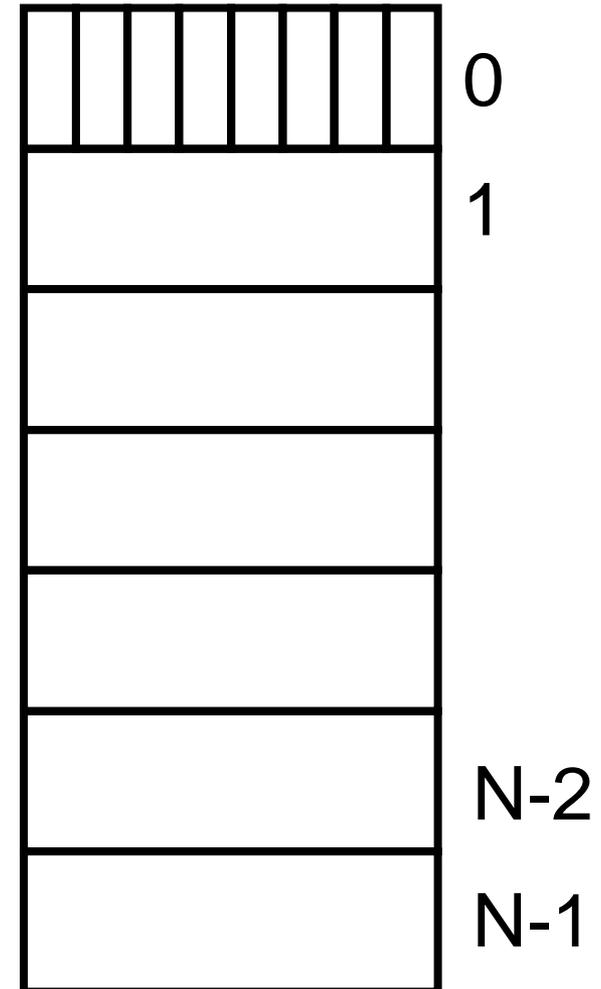
- Memorie a sola lettura (*Read Only Memory* - ROM)
- Memorie a lettura/scrittura (*Read Write Memory* - RWM)

In base alla “**stabilità**” dell’ **informazione memorizzata**:

- Volatili: perdono l’informazione memorizzata se non alimentate
- Non volatili: mantengono l’informazione anche se non alimentate

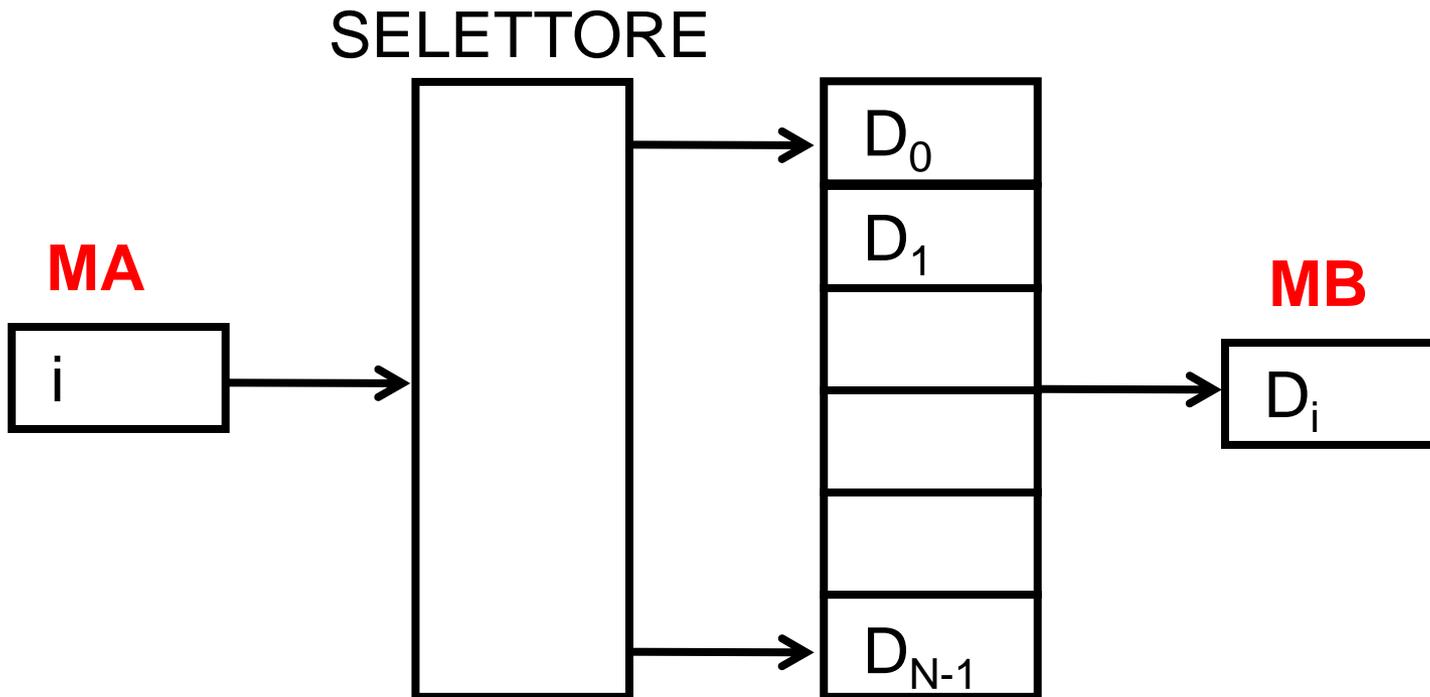
Memorie indirizzabili

- Ogni registro della memoria è univocamente individuato da un numero intero (*indirizzo*) che assume valori da 0 a $N-1$ (spazio di indirizzamento)
- I singoli registri di cui si compone la memoria vengono detti *locazioni di memoria*
- Ciascuna locazione contiene p bit ovvero è costituita da p celle elementari
- Il sistema di selezione è un decoder



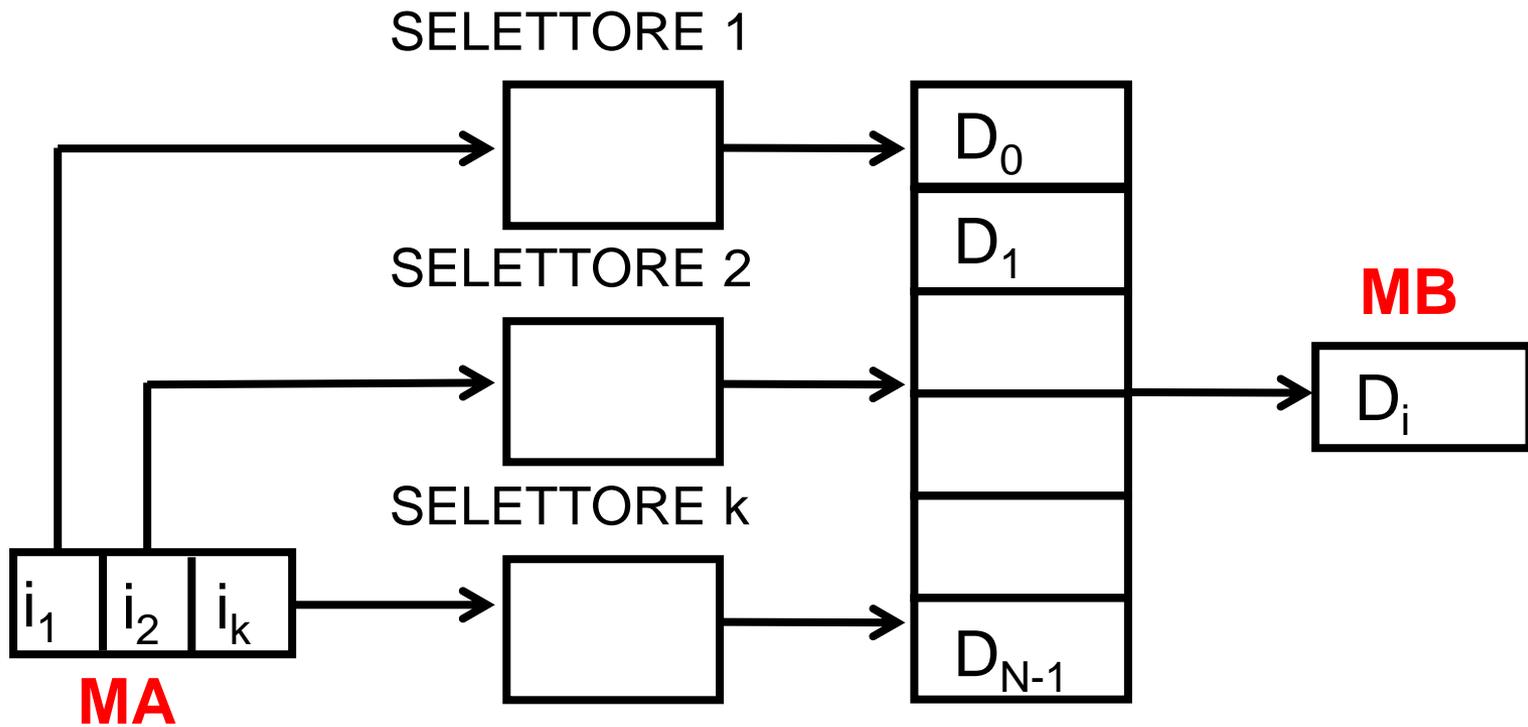
Selezione lineare

- Un unico sistema di selezione (un decoder) seleziona direttamente e singolarmente ciascuno degli $N-1$ registri



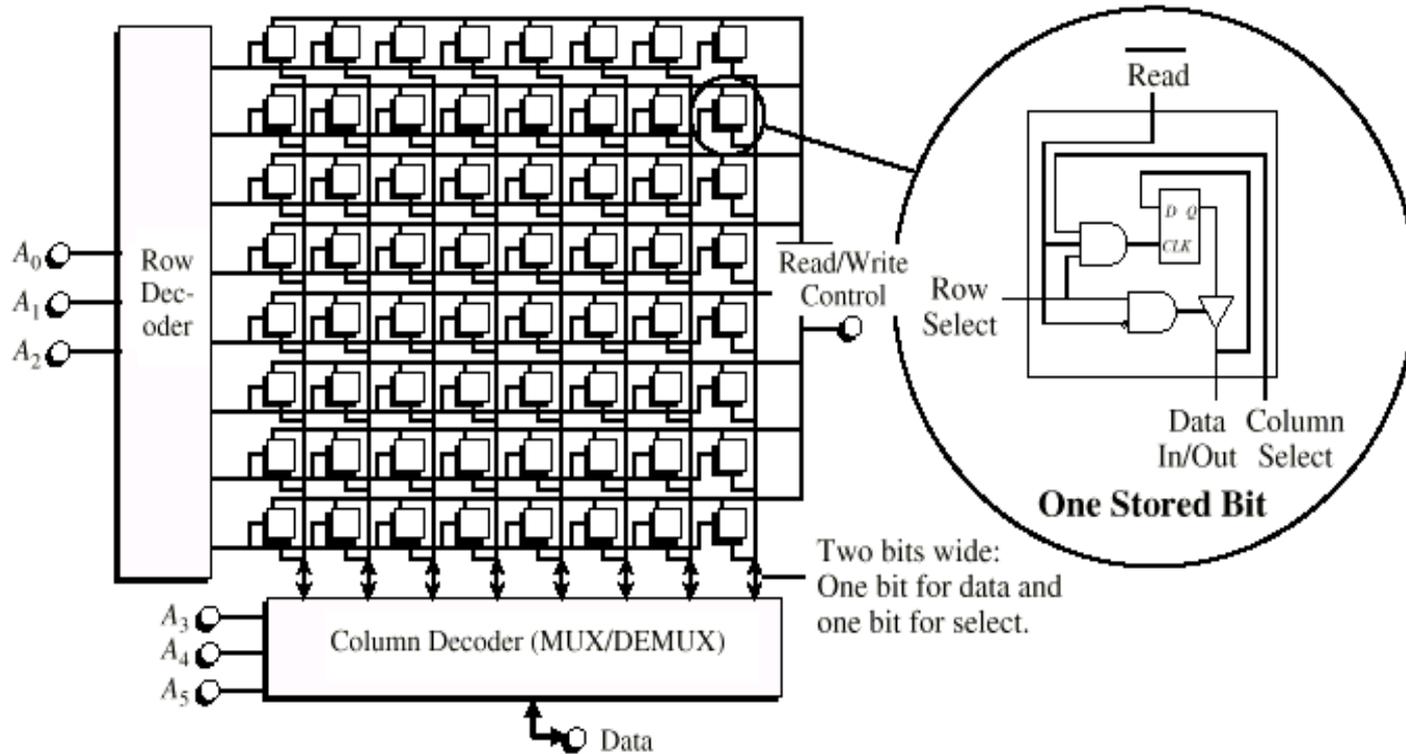
Selezione multidimensionale

- Il sistema di selezione è costituito da k sottosistemi, la cui azione combinata seleziona ciascuno degli $N-1$ registri



Semi-selezione (1)

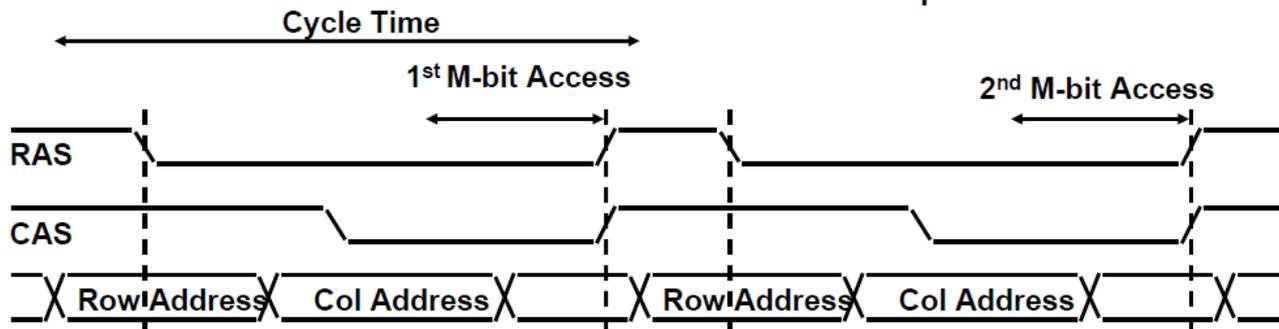
- E' un caso particolare di selezione multidimensionale con $k=2$



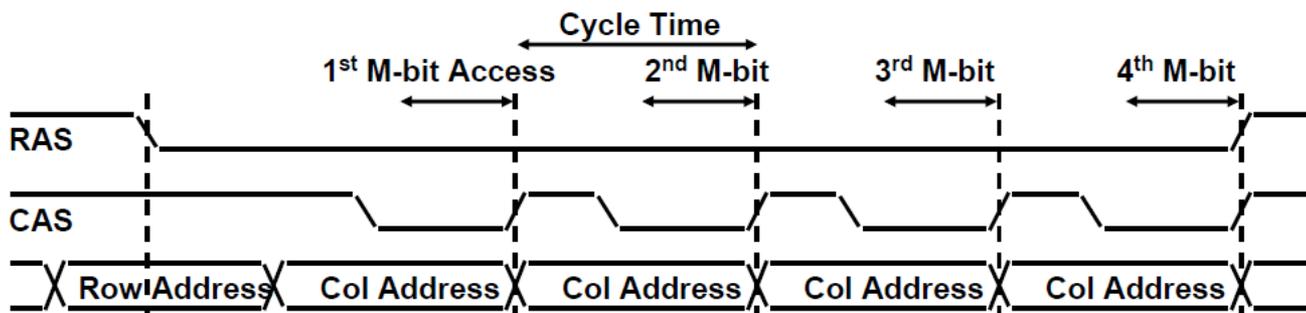
Modulo RAM da 64 word di un bit

Semiselezione (2)

- Gli indirizzi di riga e di colonna sono forniti alla memoria in due tempi successivi, attivando alternativamente una di due linee (RAS o CAS) che ne specificano il significato
 - RAS - *Row Access Strobe*
 - CAS - *Column Access Strobe*



- In alcune memorie, la semiselezione può avvenire in *page mode* per effettuare l'accesso a locazioni poste ad indirizzi consecutivi
 - L'indirizzo di riga è fornito una sola volta, e quello di colonna k volte



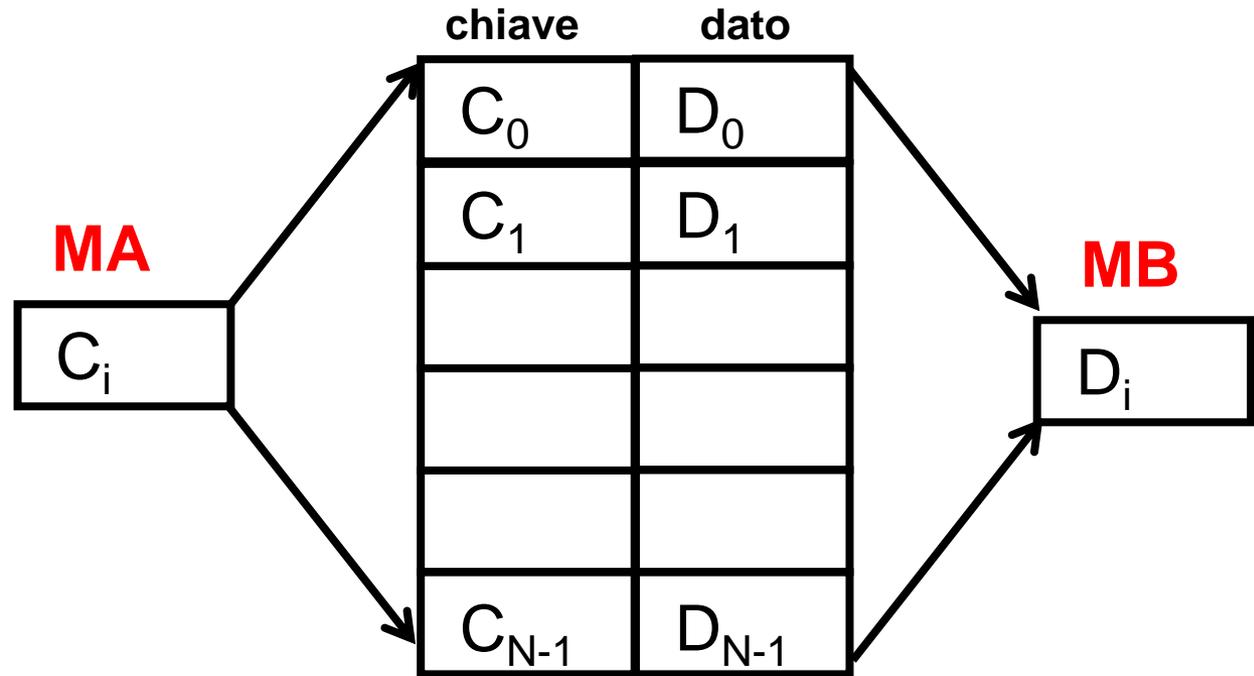
Memorie associative

- In alternativa alle memorie indirizzabili, in alcuni contesti (es. nelle memorie *cache*), si usano tipi di memorie, dette *associative*, in cui la selezione del registro avviene non tramite un indirizzo, bensì attraverso il valore di un particolare campo (*chiave*)
- Il valore del campo chiave deve essere tale da consentire di individuare univocamente uno specifico registro della memoria
- Anche dette CAM
 - *Content Addressable Memory*

chiave	dato

Selezione associativa – Schema di principio

- Il dato è selezionato dal valore della chiave



Evoluzione tecnologica delle memorie DRAM

- La tecnologia DRAM, utilizzata per realizzare la memoria centrale del calcolatore, è evoluta negli anni, consentendo la realizzazione di memorie di maggiore capacità
- All'incremento di capacità non è corrisposto un altrettanto significativo aumento della velocità

DRAM		
<u>Year</u>	<u>Size</u>	<u>Cycle Time</u>
1980	64 Kb	250 ns
1983	256 Kb	220 ns
1986	1 Mb	190 ns
1989	4 Mb	165 ns
1992	16 Mb	145 ns
1995	64 Mb	120 ns

Capacity 1000:1!

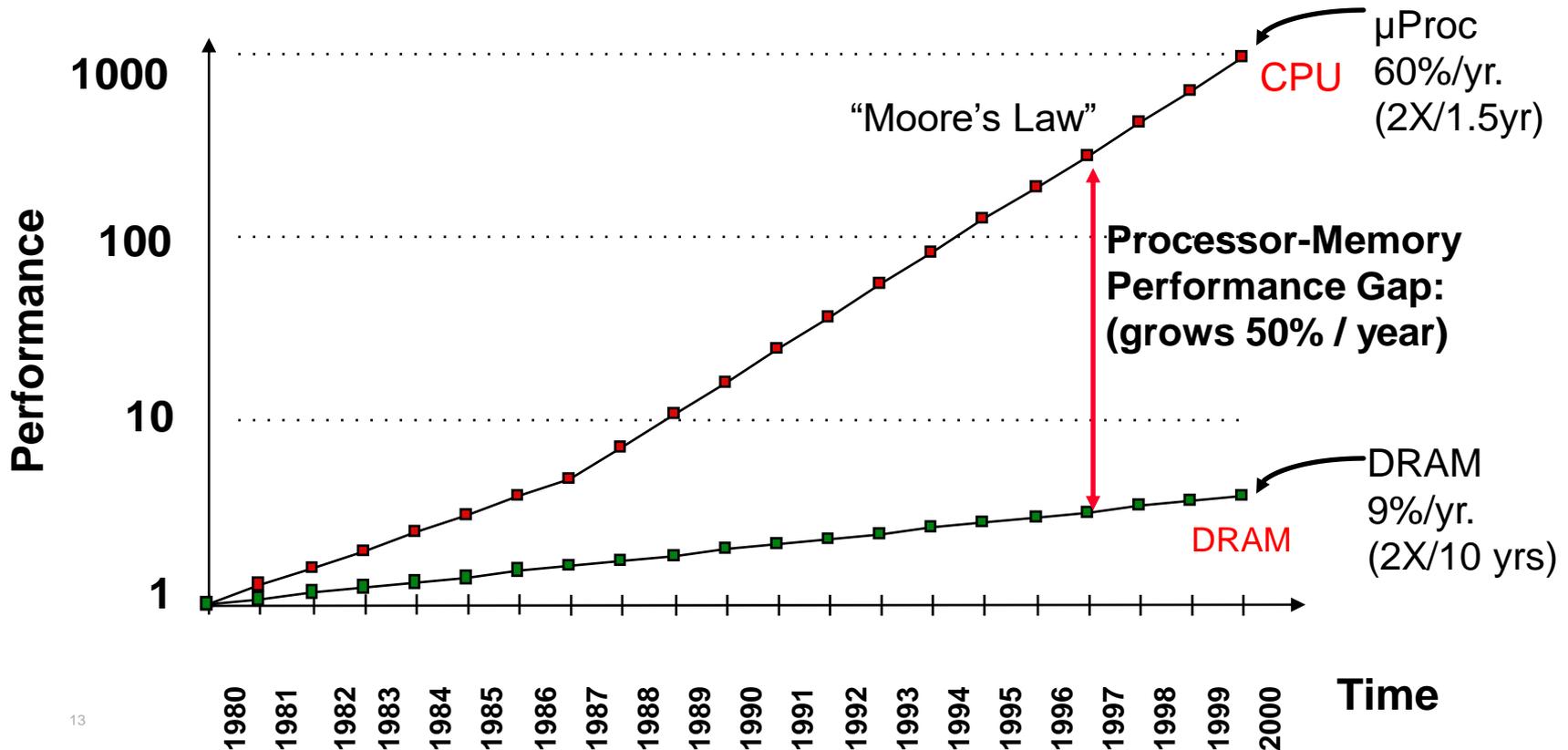
Speed 2:1!

Fonte: David A. Patterson (2000)

Gap di performance CPU-memoria

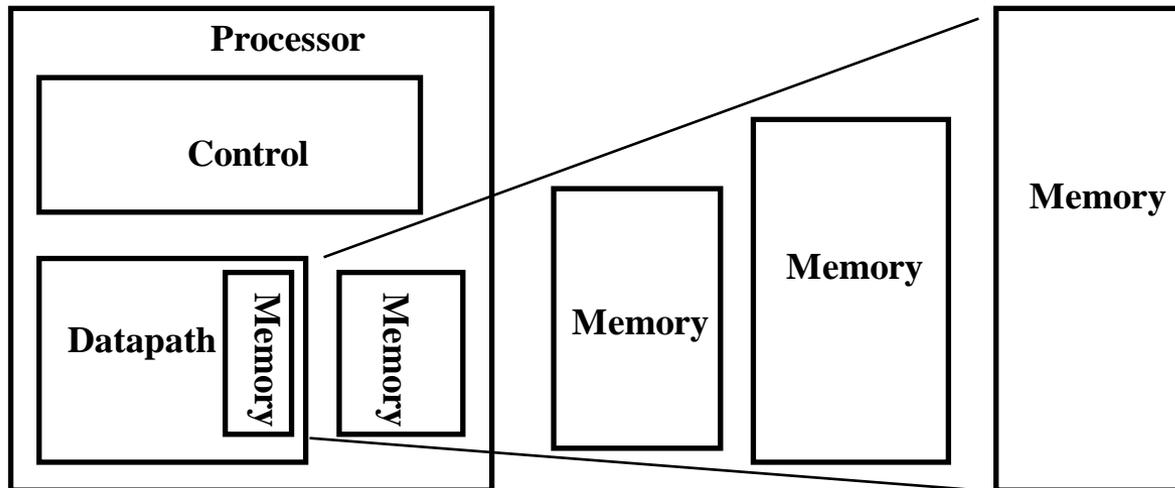
	Capacity	Speed (latency)
CPU:		2x in 1.5 years (Instructions per second)
DRAM:	4x in 3 years	2x in 10 years (Transfers per second)
Disk:	4x in 3 years	2x in 10 years (Transfers per second)

- Il differente tasso di aumento di velocità tra CPU e memoria ha creato una tensione che ha determinato la necessità di una diversa organizzazione della memoria rispetto al modello classico di von Neumann



Gerarchia di memorie nei calcolatori

- Nei calcolatori il sistema di memoria è organizzato gerarchicamente
- Il concetto di organizzazione gerarchica è sempre stato presente
- Le memorie di massa (non volatili) costituiscono il sistema di memorizzazione secondario (file system in linea) e terziario (backup)
- Le memorie di massa sono accedute dalla CPU attraverso il sistema di I/O



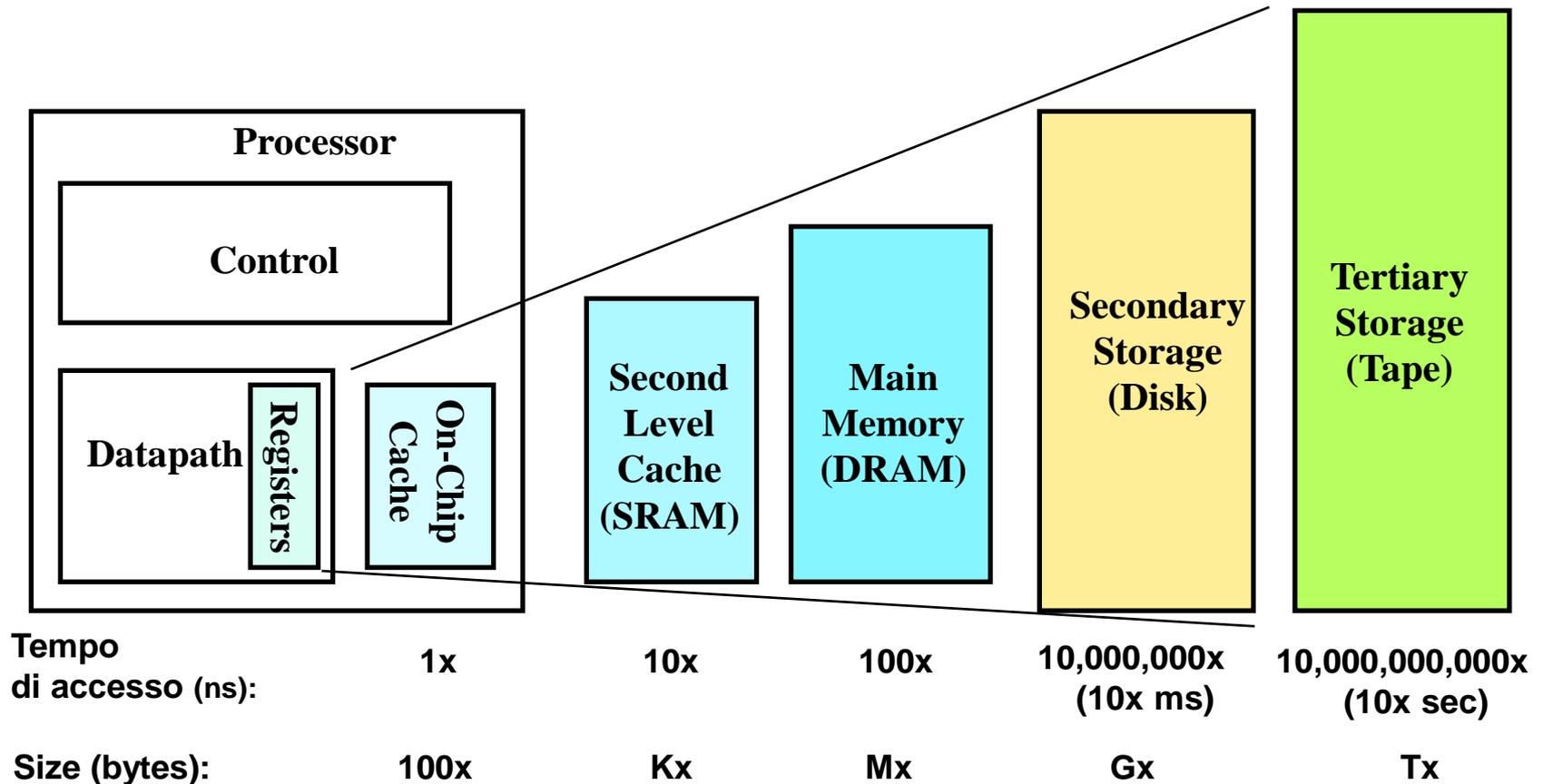
Speed: Fastest
Size: Smallest
Cost: Highest

Slowest
Biggest
Lowest

Fonte: David A. Patterson (2000)

La gerarchia di memorie nei calcolatori moderni

- Nei calcolatori moderni, tra la CPU e le memoria centrale si frappongono un certo numero di livelli di memoria cache realizzati in tecnologia SRAM
- La tecnologia SRAM consente la realizzazione di memorie di maggiore velocità ma di minore capacità rispetto alla tecnologia DRAM



Fonte: David A. Patterson (2000)

Confronto SRAM - DRAM

- DRAM sta per *Dynamic Random Access Memory*
 - Il termine Dynamic indica il fatto che l'informazione contenuta nelle celle di memoria necessita di un'operazione di "rinfresco" (*refresh*) periodica (ogni 8 ms)
 - Le celle di memoria sono organizzate a formare una matrice bidimensionale
 - La selezione avviene con la tecnica della semiselezione
- SRAM sta per *Static Random Access Memory*
 - Il termine Static indica che queste memorie non hanno necessità del refresh
 - La cella elementare SRAM è più complessa della cella DRAM (6 transistor contro 1)
 - Un chip SRAM ha una minore capacità di un chip DRAM
 - Il tempo di accesso di una memoria SRAM è di quasi due ordini di grandezza inferiore a quello di una DRAM

Static RAM (SRAM)

Tempo di accesso: 0.5ns – 3ns

Costo: \$2000 – \$5000 per GB

Dynamic RAM (DRAM)

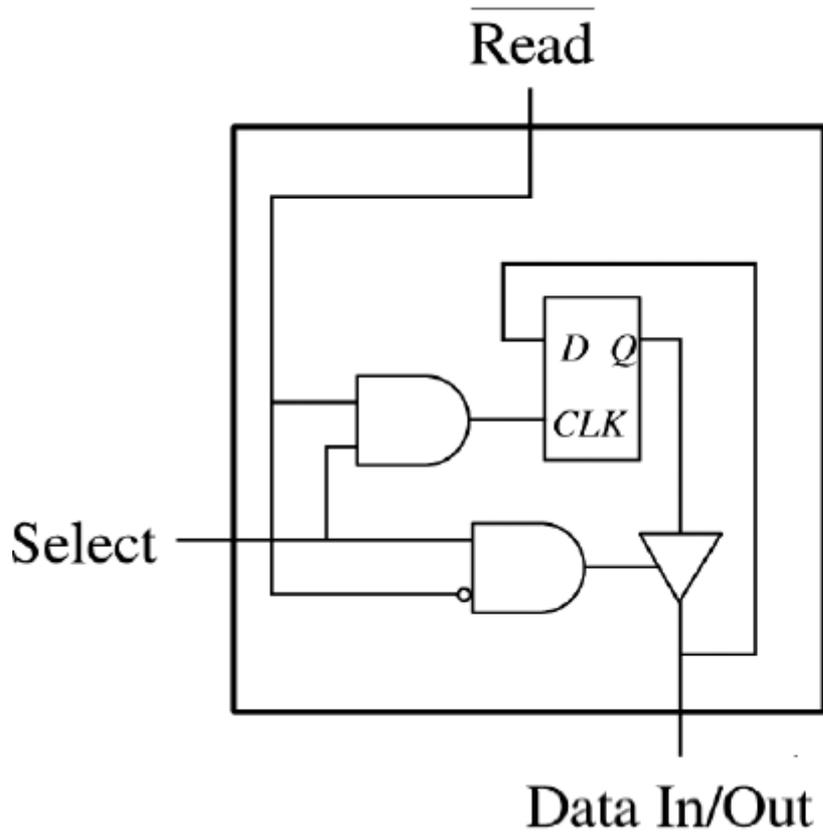
Tempo di accesso: 20ns – 50ns

Costo: \$20 – \$75 per GB

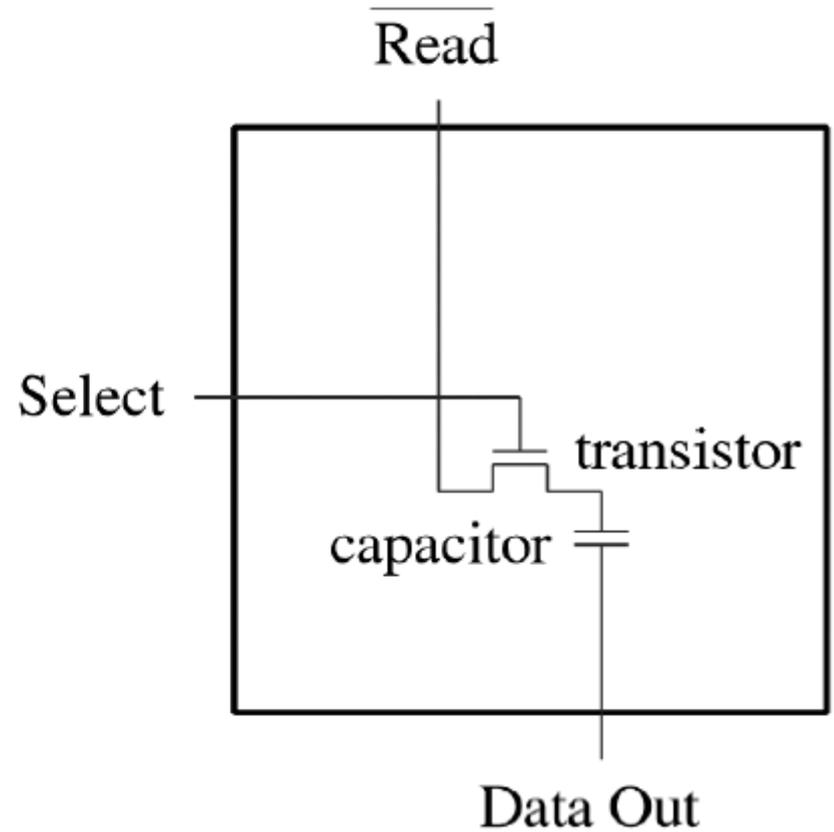
Dati del 2017

Come è fatta una singola cella di memoria RAM

Una cella di memoria RAM mantiene un singolo bit di informazione



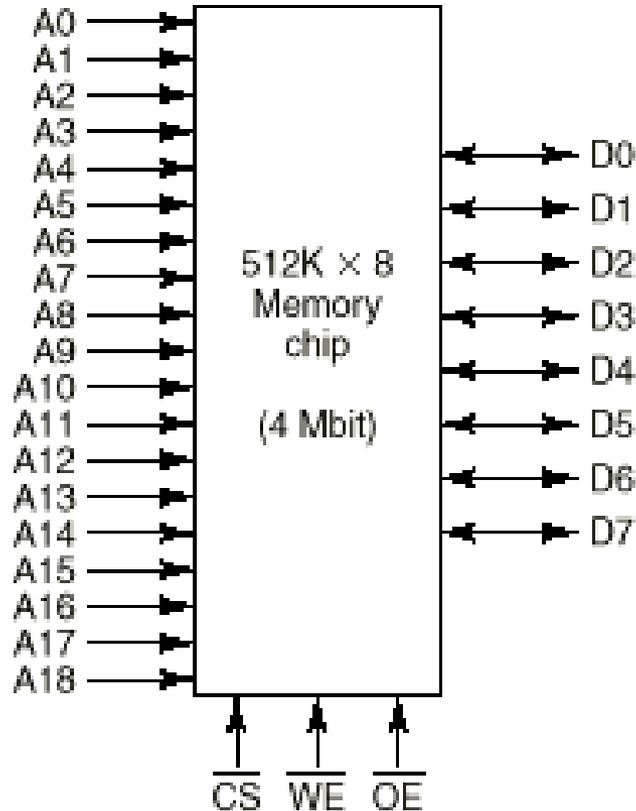
SRAM



DRAM

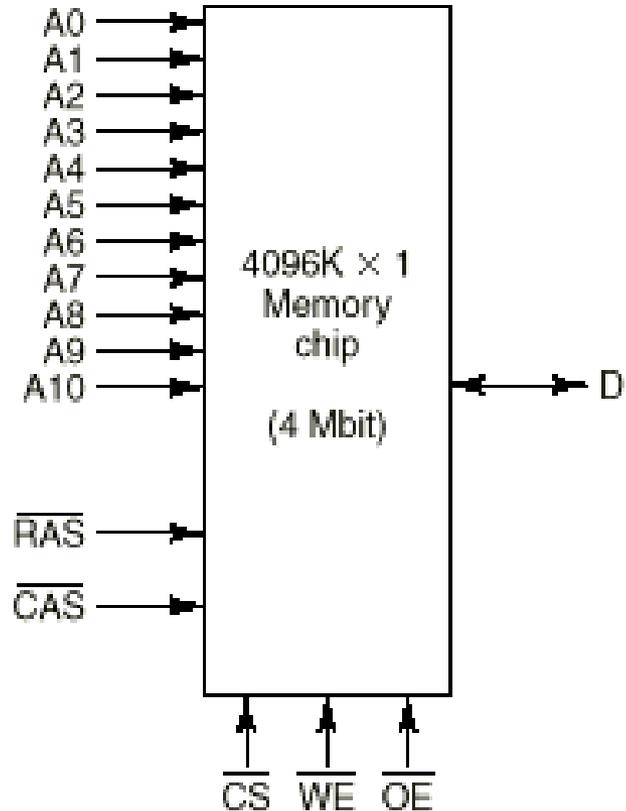
Esempi di chip di memoria

512 K * 8 bit



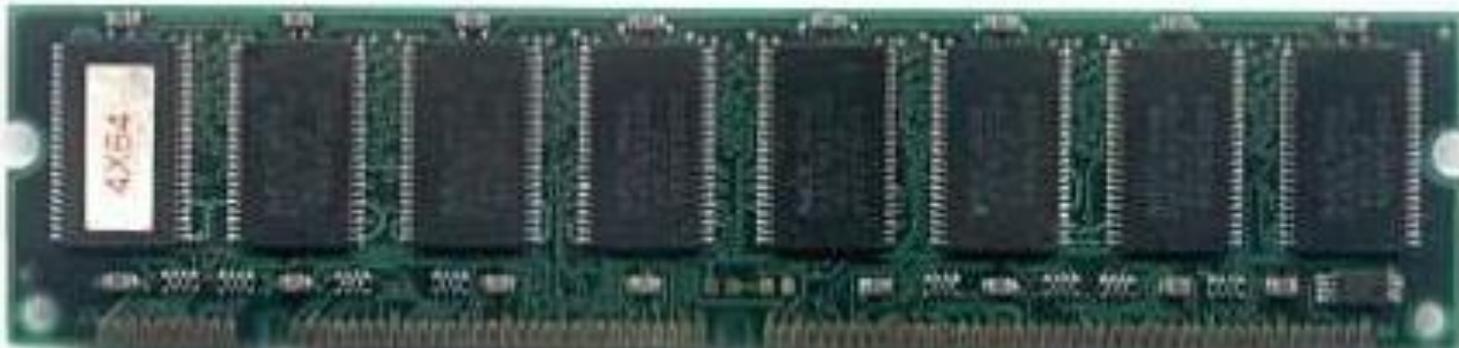
(a)

4096 K * 1 bit



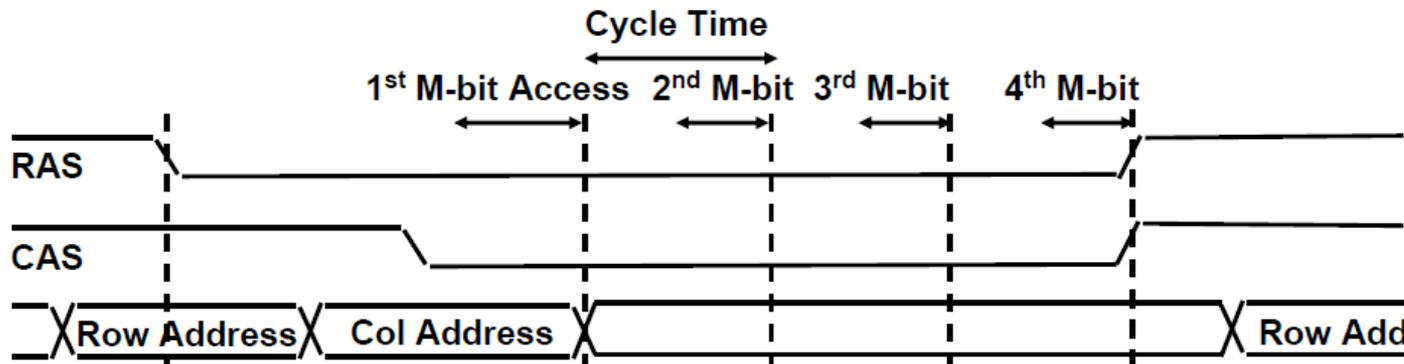
(b)

Moduli DRAM



Evoluzione della tecnologia DRAM: SDRAM

- Per far fronte all'esigenza di una maggiore velocità di trasferimento dei dati, la tecnologia DRAM è evoluta nella tecnologia SDRAM
- SDRAM sta per Synchronous DRAM
- Nelle memorie SDRAM le operazioni di accesso alla memoria sono regolate da un clock
- Trasferimenti multipli in lettura/scrittura su locazioni consecutive (*trasferimenti a blocchi*) avvengono mediante un'unica operazione di selezione preliminare ed una successiva fase di trasferimento dati
- Evoluzioni successive consentono due trasferimenti per singolo ciclo di clock
 - DDR – double data rate



Caratteristiche moduli SDRAM

- Organizzate in moduli DIMM (Dual In-Line Memory Modules)
- Bus da 64 bit

Lunghezza dei moduli	133,35 mm
Altezza dei moduli	38-43 mm
Pin	168
Bus	64 bit
Clock	66-100-133 MHz (PC66-PC100-PC133)
Bandwidth	533-800-1066 MB/s
Capacità dei singoli chip	8-32 MB
Capacità complessiva	32-64-96-128-192-256-512 MB
Tensione operativa	3,3V

Caratteristiche delle memorie DDR

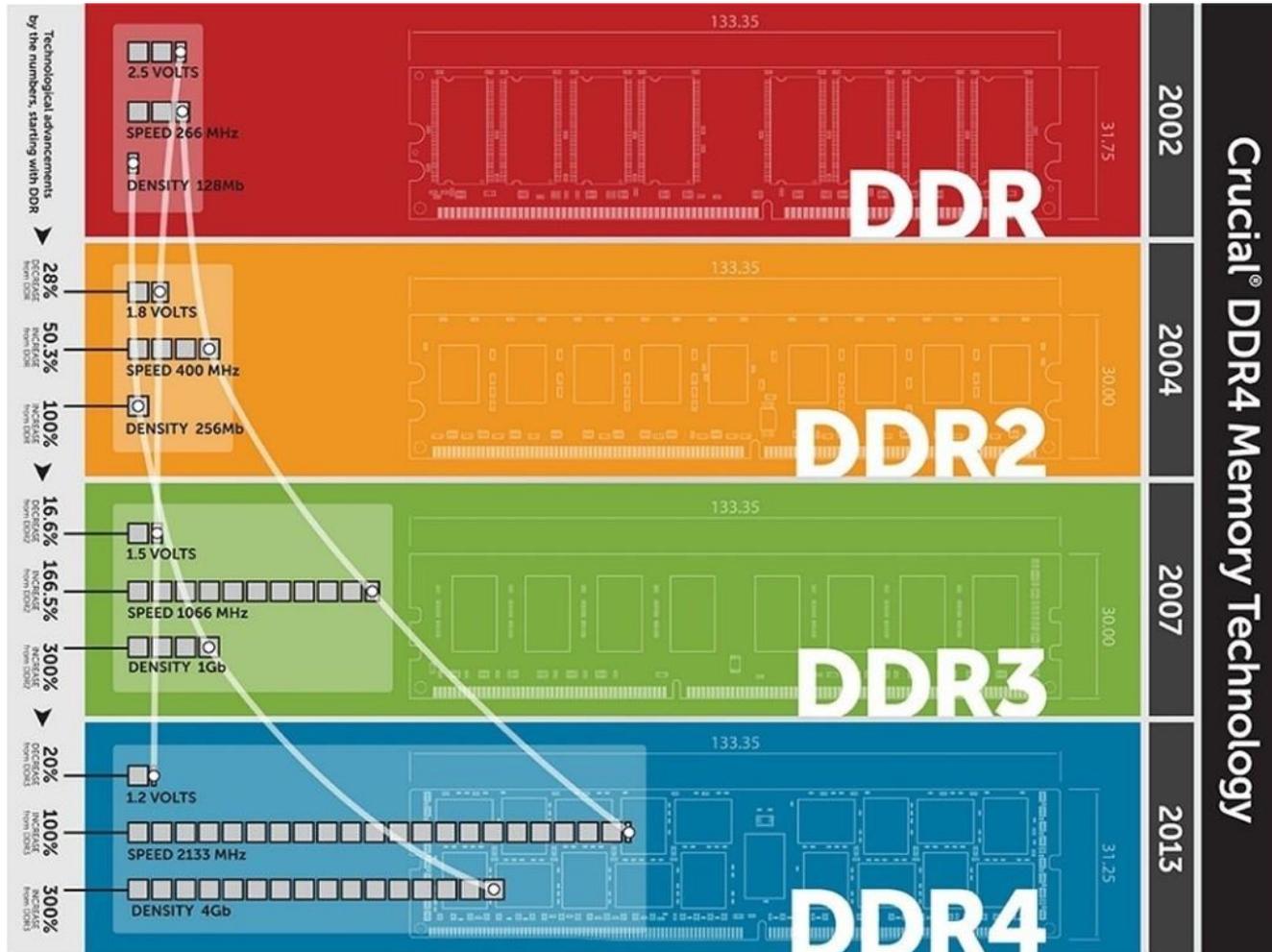
- Double Data Rate Synchronous Dynamic Random-Access Memory
- Evoluzione della tecnologia SDRAM (*sincrona*)
- La prima versione (DDR) realizzata su specifiche JEDEC JESD79 del 2001
- Ampiezza di bus di 64 bit: ogni trasferimento dati riguarda 8 byte
- Raddoppio del data rate rispetto alle SDRAM ottenuto mediante trasmissione dei dati sia sul fronte di salita che di discesa del clock
 - Accorgimento già adottato nelle memorie RDRAM (Rambus)
 - A parità di frequenza di clock (100 MHz) e di ampiezza di bus (64 bit), rispetto alle SDRAM PC-100, il tasso di trasferimento dei dati in MegaByte/s (*bandwidth*) raddoppia, passando da 800 MB/s a 1600 MB/s
- Tensione di alimentazione 2.5 V (inferiore rispetto ai 3.3 V delle SDRAM)
- Moduli da 184 pin

Standard JEDEC	Nome commerciale	Clock della memoria	Clock del bus I/O	Data Rate	Bandwidth	Tensione
DDR-200	PC-1600	100 MHz		200 MT/s	1.600 MB/s	2,5V
DDR-266	PC-2100	133 MHz		266,66 MT/s	2.133 MB/s	
DDR-333	PC-2700	166 MHz		333,33 MT/s	2.666 MB/s	
DDR-400	PC-3200	200 MHz		400 MT/s	3.200 MB/s	2,6V

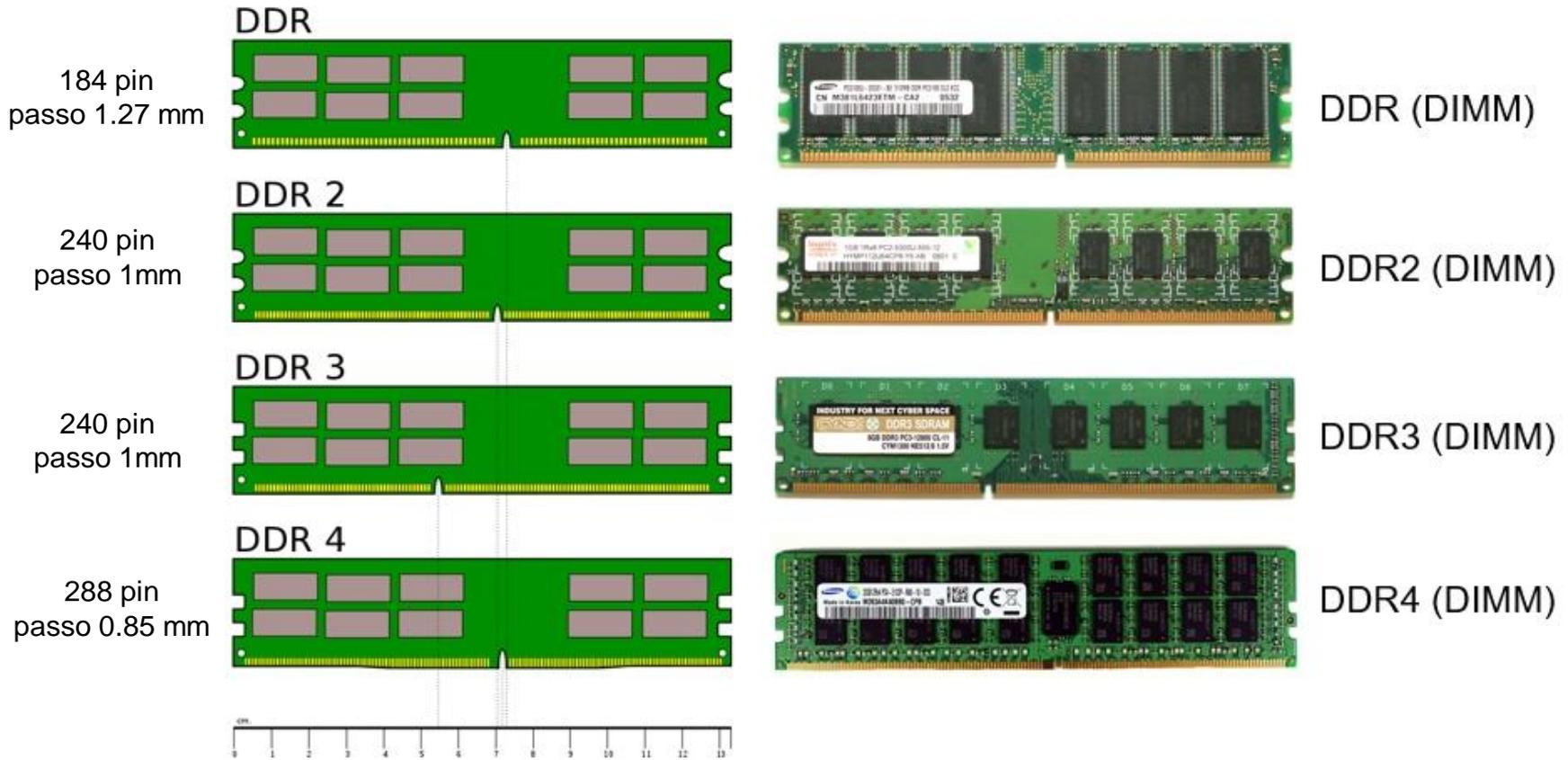


Evoluzione delle memorie DDR

- Double Data Rate Synchronous Dynamic Random-Access Memory



Memorie DDR: evoluzione del formato dei moduli





Standard JEDEC	Nome commerciale	Clock della memoria	Clock del bus I/O	Data Rate	Bandwidth	Tensione
DDR2-400	PC2-3200	100 MHz	200 MHz	400 MT/s	3.200 MB/s	1,8V
DDR2-533	PC2-4200	133 MHz	266 MHz	533 MT/s	4.267 MB/s	
DDR2-667	PC2-5300	167 MHz	333 MHz	667 MT/s	5.333 MB/s	
DDR2-800	PC2-6400	200 MHz	400 MHz	800 MT/s	6.400 MB/s	~2,0V
DDR2-1066	PC2-8500	267 MHz	533 MHz	1.066 MT/s	8.533 MB/s	~2,2V



DDR3

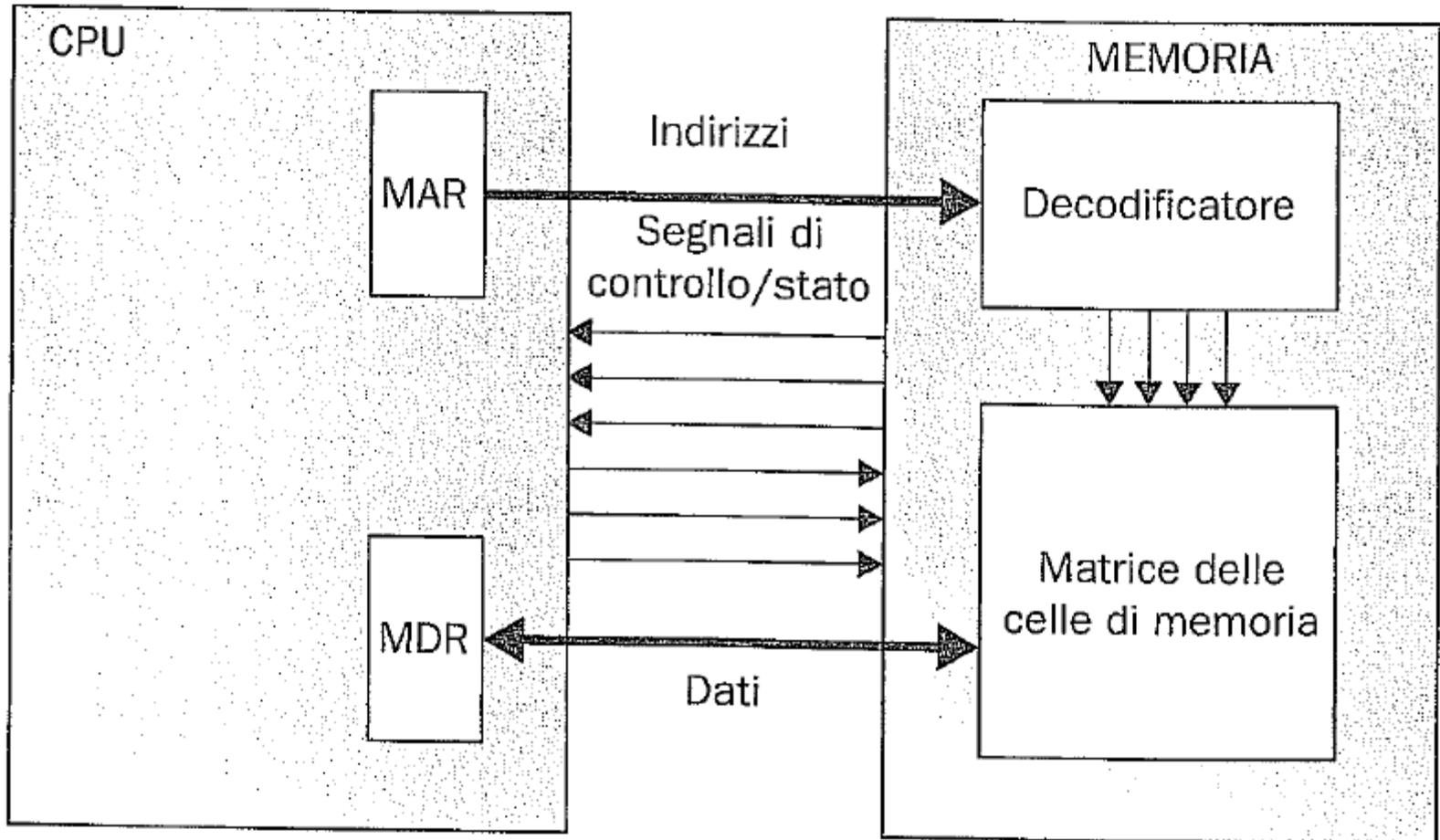
Standard JEDEC	Nome commerciale	Clock della memoria	Clock del bus I/O	Data Rate	Bandwidth	Tensione
DDR3-800	PC3-6400	100 MHz	400 MHz	800 MT/s	6.400 MB/s	1,5V
DDR3-1066	PC3-8500	133 MHz	533 MHz	1066 MT/s	8.533 MB/s	
DDR3-1333	PC3-10600	167 MHz	667 MHz	1333 MT/s	10.667 MB/s	
DDR3-1600	PC3-12800	200 MHz	800 MHz	1600 MT/s	12.800 MB/s	
DDR3-1866	PC3-14900	233 MHz	933 MHz	1866 MT/s	14.933 MB/s	~1,65V
DDR3-2133	PC3-17000	267 MHz	1066 MHz	2133 MT/s	17.067 MB/s	



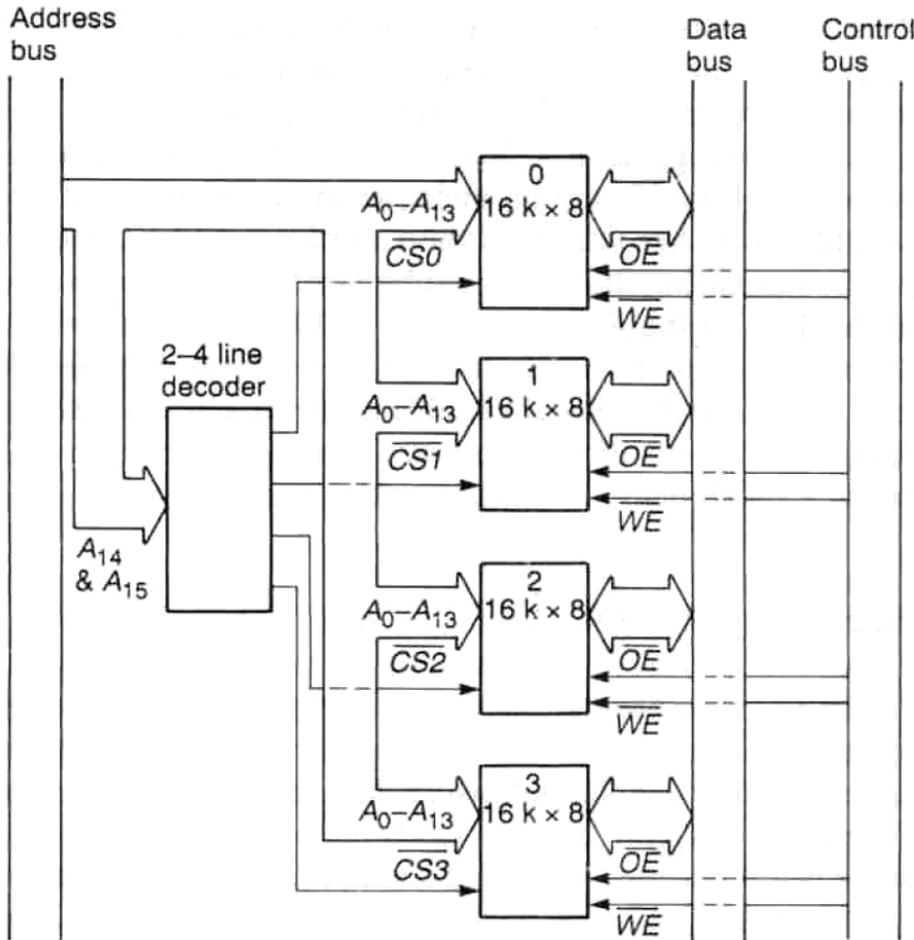
DDR4

Standard JEDEC	Nome commerciale	Clock della memoria	Clock del bus I/O	Data Rate	Bandwidth	Tensione
DDR4-1600	PC4-12800	200 MHz	800 MHz	1600 MT/s	12.800 MB/s	1,25V
DDR4-1866	PC4-14900	233 MHz	933 MHz	1866 MT/s	14.933 MB/s	
DDR4-2133	PC4-17000	266 MHz	1066 MHz	2133 MT/s	17.067 MB/s	
DDR4-2400	PC4-19200	300 MHz	1200 MHz	2400 MT/s	19.200 MB/s	
DDR4-2666	PC4-21300	333 MHz	1.333 MHz	2.666 MT/s	21.333 MB/s	1,2~1,35V
DDR4-3000	PC4-23400	366 MHz	1.467 MHz	2.934 MT/s	23.472 MB/s	
DDR4-3200	PC4-25600	400 MHz	1.600 MHz	3200 MT/s	25.600 MB/s	

Collegamento CPU-memoria (1)

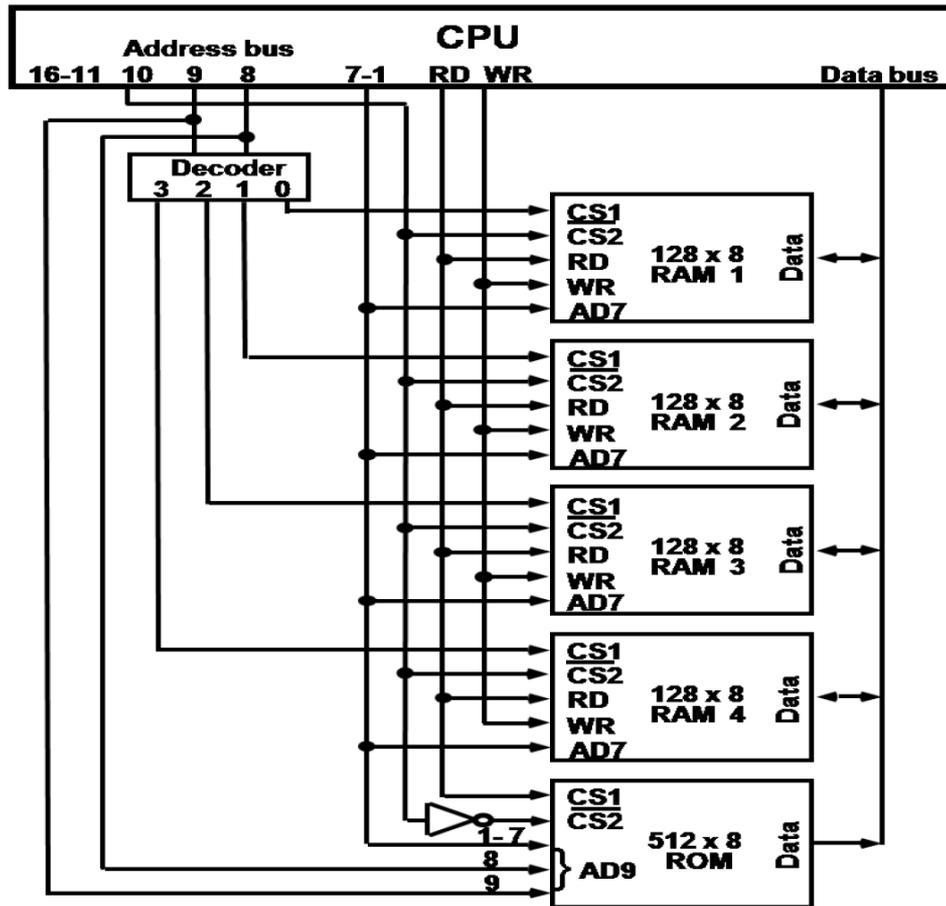


Collegamento CPU-memoria (2)



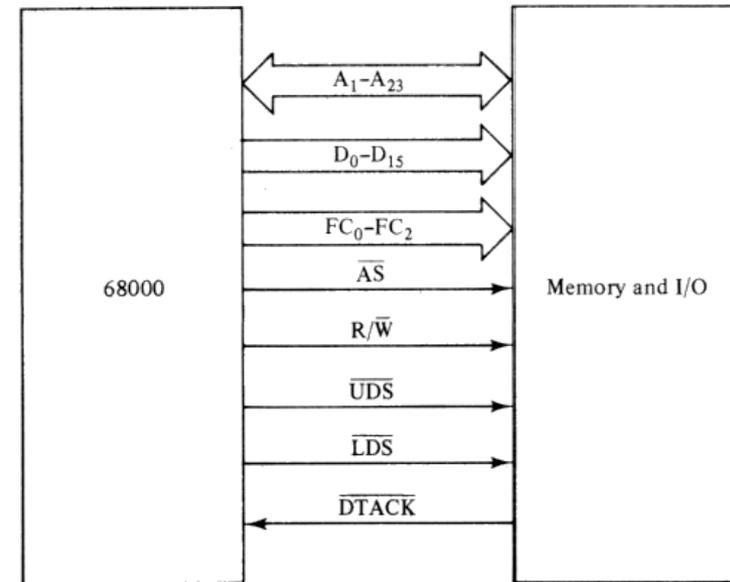
Address (hexadecimal)	Address (decimal)
FFFF	65 535
Memory device 3	
C000	49 152
BFFF	49 151
Memory device 2	
8000	32 768
7FFF	32 767
Memory device 1	
4000	16 384
3FFF	16 383
Memory device 0	
0	0

Collegamento CPU-memoria (3)

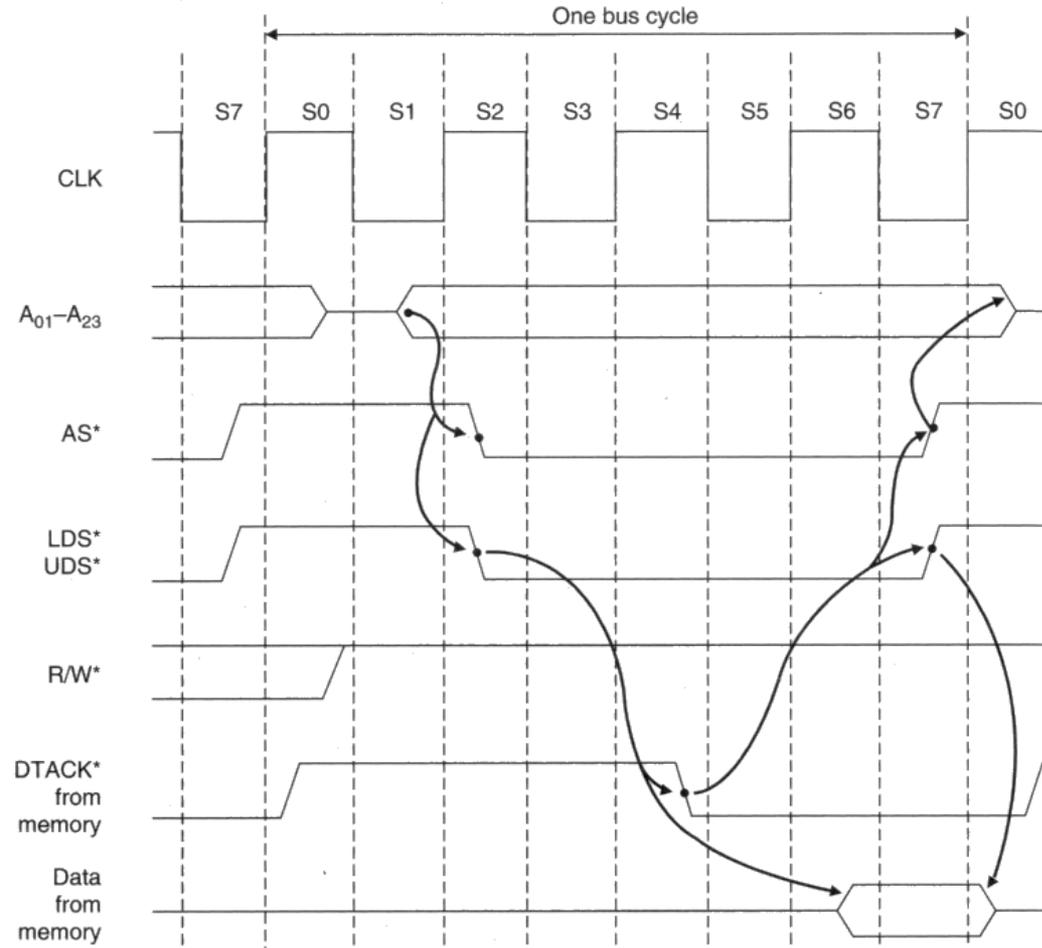


Interfacciamento asincrono CPU-memoria

- Il processore avvia un ciclo di lettura o scrittura in memoria (di un dato o una istruzione)
 - Mette un indirizzo su A1-A23
 - Attiva Address Strobe !AS
 - Attende la terminazione della operazione
- Il ciclo termina quando la memoria segnala la terminazione dell'operazione con un apposito segnale di riscontro !DTACK
- In questo modo il processore si adatta alla velocità della memoria
 - Il processore può interagire con memorie o periferiche di I/O lente
- Si osservi che gli stessi bus utilizzati per interfacciare CPU e memoria possono essere usati per interfacciare la CPU con le periferiche di I/O
 - I/O memory mapped



Tempificazione di un ciclo read



Interfacciamento sincrono CPU-memoria

- La memoria non usa più un segnale di riscontro !DTACK per indicare la terminazione di una operazione
- Un segnale di tempificazione E derivato dal clock di sistema determina la tempificazione del ciclo di lettura o scrittura che si sviluppa in un numero prefissato di cicli

