Architettura dei Calcolatori
Sistema di memoria
parte prima

Ing. dell'Automazione

A.A. 2011/12

Gabriele Cecchetti

Sistema di memoria – *parte prima*

- Sommario:
  - Banco di registri
  - Generalità sulla memoria
  - Tecnologie di memoria
- Riferimenti
  - C. Hamacher, "Introduzione all'architettura del Calcolatore", cap. 9, sez. 9.1, 9.2, 9.3.
  - G. Corsini "Dalle porte AND OR NOT al sistema calcolatore: un viaggio nel mondo delle reti logiche": cap. "Reti Sequenziali Asincrone"

alcune nozioni fondamentali
banco di registri

BANCO DI REGISTRI

Banco di Registri

- Spesso occorre utilizzare un certo numero di registri, tutti aventi le medesime dimensioni e funzioni.
- Invece di avere tanti registri paralleli indipendenti, è preferibile organizzarli in una struttura a matrice, chiamata banco di registri (*register file*).
- Il banco di registri permette anche di ridurre i collegamenti necessari.

Funzionamento del Banco (1/3)

- Si consideri un banco di 8 registri da 16 bit ciascuno, ovvero un banco 8×16.
- Ogni registro è identificato da un indirizzo:
  - un numero compreso tra 0 e 7, inclusi
- Gli 8 registri sono pertanto chiamati:
  - R0, R1, …, R7
- Ogni registro contiene 16 bit.

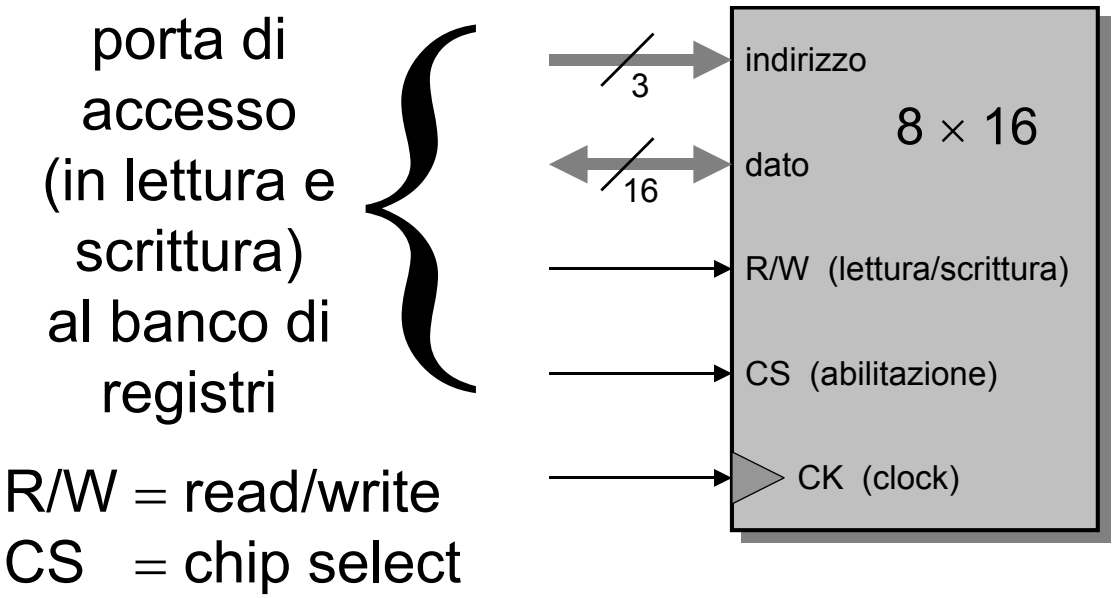
Funzionamento del Banco (2/3)

- Per specificare l'indirizzo di un registro occorrono 3 bit, sufficienti a codificare i numeri interi nell'intervallo 0, 1, …, 7.
  (con $r \geq 1$ registri occorrono $\lceil \log_2 r \rceil$ bit)
- Le operazioni eseguibili sul banco sono:
  - lettura, ovvero si acquisiscono i 16 bit memorizzati nel registro indirizzato
  - scrittura, ovvero si caricano 16 bit nel registro indirizzato

Funzionamento del Banco (3/3)

- Il banco dei registri ha i controlli seguenti:
  - 3 ingressi di indirizzo, dove si manda l'indirizzo del registro da leggere o scrivere
  - 16 uscite/ingressi dati, dove si riceve/invia il contenuto del registro da leggere/scrivere
  - un ingresso di comando:
    - lettura o scrittura
  - un ingresso di abilitazione
  - e naturalmente il segnale di clock

Rappresentazione del Banco

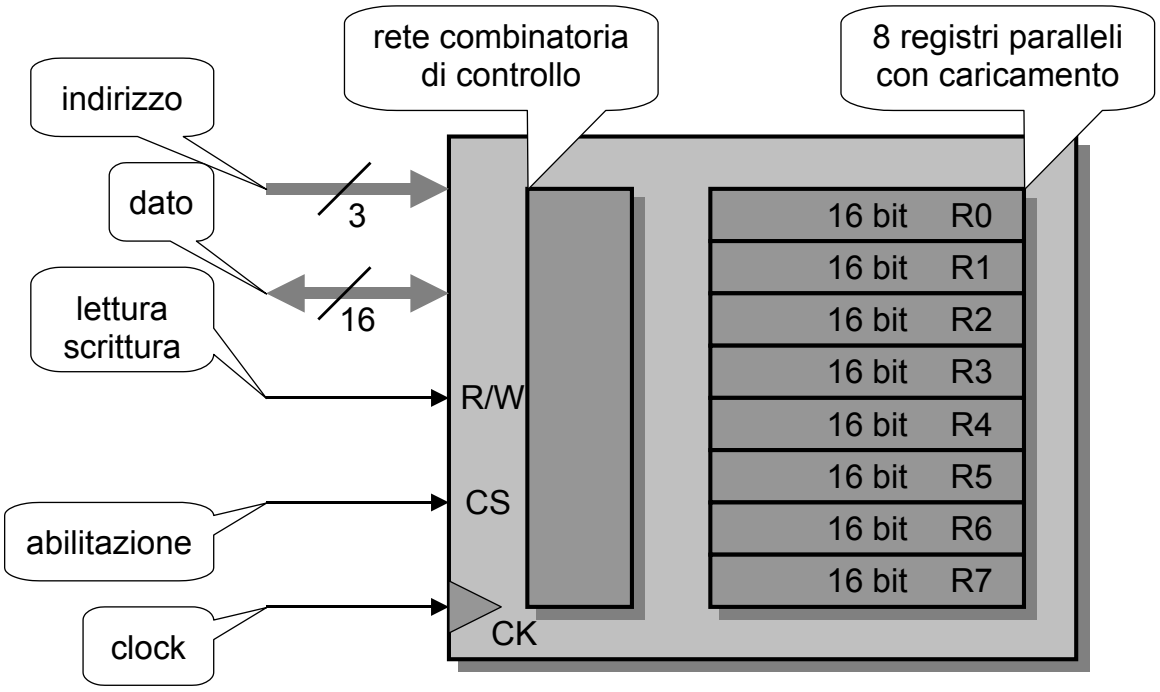


R/W = read/write
CS = chip select

Operazioni del Banco

- Se l'abilitazione è attiva (CS = 1):
  - lettura:
    - all'inizio del ciclo si inviano l'indirizzo del registro e il comando di lettura (R/W = 1),
    - verso la fine del ciclo il contenuto del registro viene emesso;
  - scrittura:
    - all'inizio del ciclo si inviano l'indirizzo del registro, il dato da scrivere e il comando di scrittura (R/W = 0),
    - verso la fine del ciclo il dato viene memorizzato nel registro.
- Se l'abilitazione è inattiva (CS = 0), il banco mantiene il contenuto.

Struttura del Banco



Varianti e Integrazioni

- I banchi di registri sono disponibili in svariate dimensioni:

  8×8 , 16×16 , 32×32 e altre ancora

- I banchi più sofisticati hanno più porte di accesso distinte, in lettura/scrittura, o in sola lettura o scrittura, per operare in parallelo su più registri.

- In caso di più porte di scrittura, sono in grado di gestire i conflitti di scrittura.

Uso Caratteristico

- **Il banco dei registri è un blocco funzionale** essenziale del processore.

- *È usato per realizzare i registri interni del processore, che contengono gli operandi di tipo intero delle istruzioni aritmetico-logiche del processore.*

- Esistono anche banchi di registri adattati per operandi di tipo reale (standard aritmetico IEEE-754 per virgola mobile).

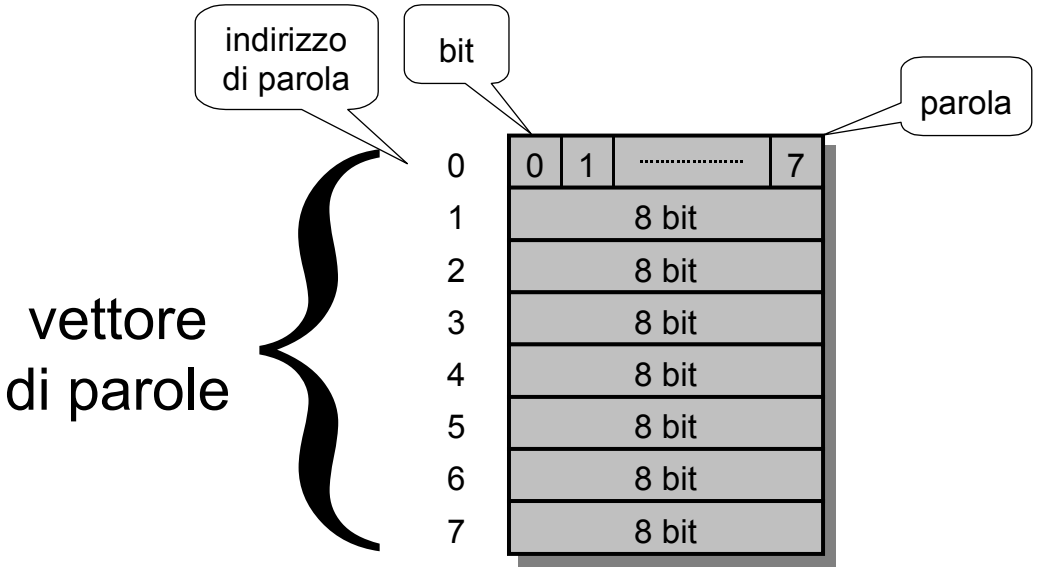
Alcune nozioni fondamentali
Interfaccia di memoria

GENERALITÀ SULLA MEMORIA

Memoria Centrale

- La memoria è un blocco funzionale di tipo sequenziale complesso.
- Serve per mantenere a tempo indefinito dati e programmi, e per permetterne l'accesso, in lettura o in scrittura.
- Ha una struttura a vettore, i cui elementi sono le parole di memoria.
- Ogni parola di memoria è una sequenza di bit, in numero fissato ≥ 1.

Organizzazione a Parole



capacità: 8 parole $\times$ 8 bit per parola $=$ 64 bit

Caratteristiche della Memoria

- Si caratterizza un componente integrato (chip) di memoria specificandone:
  - ***capacità***, misurata in numero totale di bit memorizzabili:
    - di solito si esprime la capacità come prodotto del numero di parole per il numero di bit contenuti nella parola;
  - ***funzione***: lettura e scrittura, solo lettura;
  - ***numero di porte di accesso***;
  - ***tempo necessario per l'accesso***.

Interfaccia di Memoria (1/3)

- **Il contenuto della memoria viene letto o scritto una parola per volta**, *in un ciclo di clock* (più cicli in memorie lente).
- Si accede a una parola di memoria (byte o più lunga) tramite la porta di accesso alla memoria.
- **La porta di accesso può funzionare in lettura e scrittura** (caso più frequente), *solo in lettura* e teoricamente anche solo in scrittura (caso poco frequente).

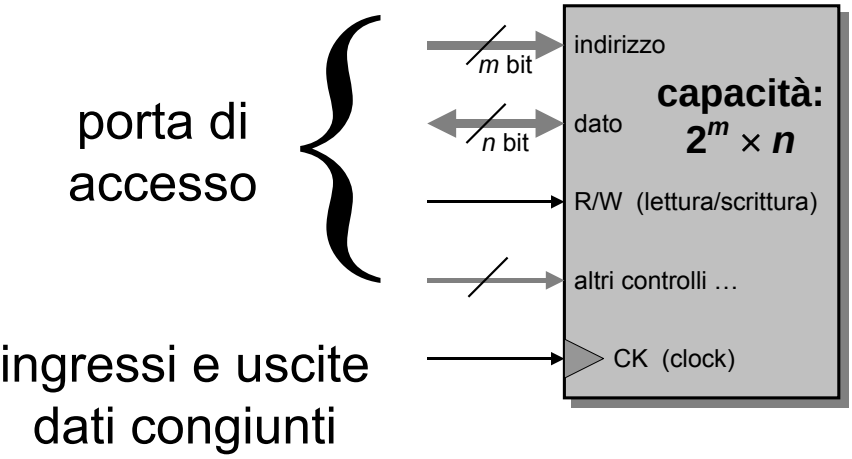
Interfaccia di Memoria (2/3)

- La porta di accesso alla memoria è costituita dai segnali seguenti:
  - ***ingressi di indirizzo***, che codificano in binario l'indirizzo della parola dove si deve operare:
    - se la memoria ha capacità di $2^m \geq 1$ parole, occorrono $\lceil \log_2 2^m \rceil = m$ ingressi di indirizzo;
  - ***uscite / ingressi di dato***, che servono per leggere / scrivere una parola:
    - se la parola ha dimensione di $n \geq 1$ bit, occorrono esattamente n uscite / ingressi di dato.

Interfaccia di Memoria (3/3)

- La porta di accesso alla memoria è costituita dai segnali seguenti:
  - comando di lettura / scrittura, R/W (*read/write*):
    - R/W = 1 lettura,
    - R/W = 0 scrittura;
  - comando di abilitazione, CS (*chip select*):
    - CS = 1 chip attivo, si può accedere al contenuto,
    - CS = 0 chip in stato di riposo, non si può né leggere né scrivere;
  - comando di abilitazione dati, OE (*output enable*):
    - OE = 1 le uscite dati sono funzionanti
    - OE = 0 le uscite dati sono isolate.

Componente di Memoria



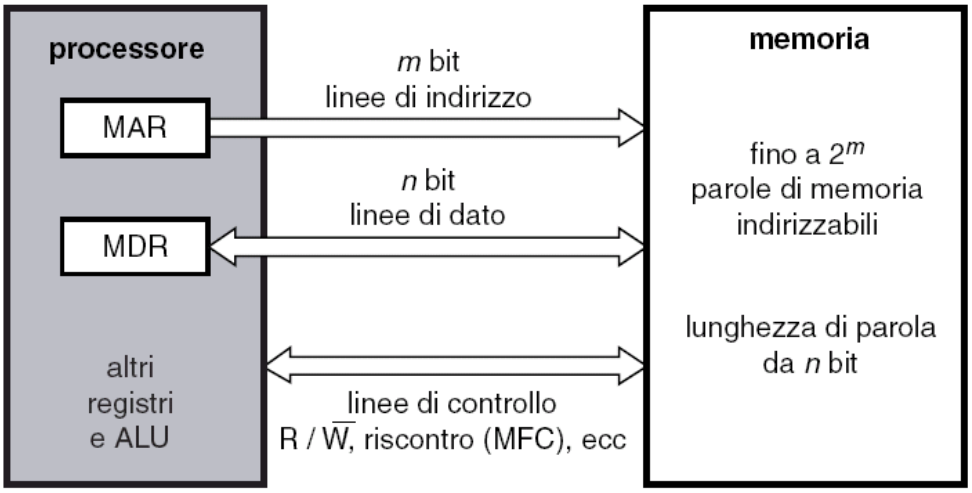
- Componente integrato (chip) di memoria.
- Capacità: 2^m parole $\times$ n bit per parola.
- Spesso il clock non serve o è interno.

Collegamento al Processore (1/2)

- Il processore (o l'unità MASTER) pilota l'operazione di accesso a memoria.
- Esso ha due registri interni per pilotare il bus:
  - registro **MAR** (*memory address register*), che contiene l'indirizzo della parola di memoria
  - registro **MDR** (*memory data register*), per scambiare il dato con la memoria

- L'unità di controllo del processore pilota anche i segnali di controllo del bus.
- *Si ritroveranno tali registri e comandi quando si esaminerà la struttura interna del processore*.

Collegamento al Processore (2/2)



schema tipico di collegamento memoria / processore, registri interni del processore con funzioni di supporto (e segnali di controllo ausiliari)

Cella di memoria RAM statica e dinamica
Cella di memoria ROM
Matrice di memoria

TECNOLOGIA DI MEMORIA I

Struttura della Memoria (1/3)

- Le strutture interne dei componenti integrati di memoria sono molto varie, e pertanto sono disponibili numerose tecnologie di memoria.
- Il modo più naturale per realizzare una memoria (sebbene non adatto a memorie di grande capacità) è costruire una matrice di bistabili, completandola con reti combinatorie di controllo per gestire l'accesso alle parole:
  - memoria RAM statica (SRAM, static RAM)
- RAM sta per Random Access Memory (memoria ad accesso casuale, cioè libero).

Struttura della Memoria (2/3)

- La matrice di bit è strutturata in righe e colonne.
  - Una riga della matrice di bit corrisponde a una parola della memoria.
  - Ogni riga della matrice di bit ha un indirizzo, che corrisponde alla parola.
- Per leggere una parola, si legge la sequenza di bit memorizzata in una riga.
- Per scrivere una parola, si scrive la sequenza di bit memorizzandola in una riga.

Struttura della Memoria (3/3)

- *Per selezionare la riga dove leggere o scrivere, si invia l'indirizzo* (emesso dal processore e presente sul bus) *a un decodificatore*, che attiva la linea di riga.
- *Per operare sulla riga selezionata* (in lettura o scrittura), *si usa un circuito apposito di lettura e scrittura*, sensibile al comando emesso dal processore e presente sul bus.
- *Per isolare le uscite*, si usano porte *tri-state* (ad alta impedenza), in genere integrate nel circuito di lettura e scrittura.

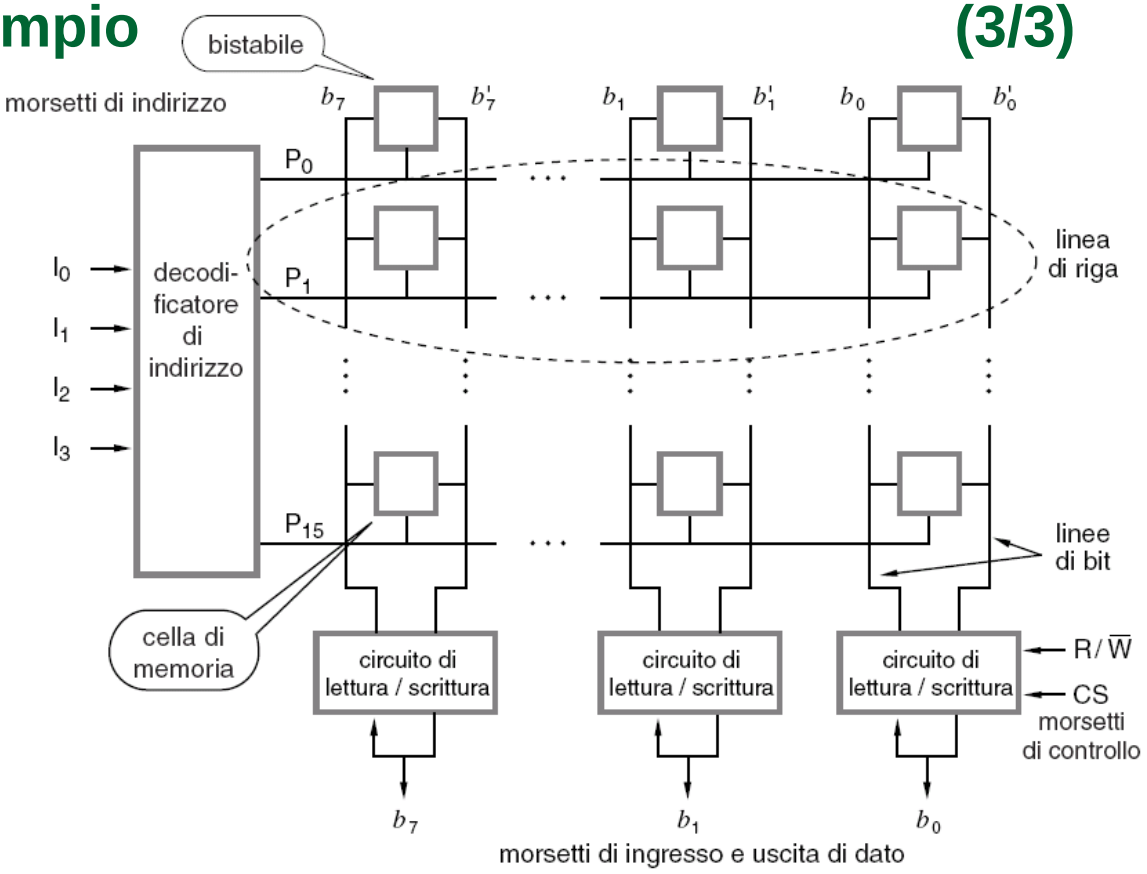
Esempio (1/3)

- Si consideri una memoria semplicissima con capacità 16 parole × 8 bit per parola.
- Essa ha $log_2 16 = 4$ bit di indirizzo.
- Ha ingressi e uscite dati separati (solo per comodità, normalmente ingressi e uscite dati sono congiunti).
- Ha ingressi e uscite dati da 7 bit.
- E ha i due comandi:
  - CS (*Chip Select*), e
  - R/W (*Read/Write*).

Esempio (2/3)

- La memoria in questione contiene i blocchi funzionali seguenti:
  - matrice 16 righe × 8 colonne di bistabili di tipo D, a sincronizzazione sul fronte
  - decodificatore a 4 ingressi e $4^2 = 16$ uscite
  - porte logiche accessorie (o “glue logic”), per la funzione di lettura e scrittura
- Il decodificatore attiva la riga (parola) indirizzata, e i circuiti di controllo interpretano ed eseguono il comando (lettura/scrittura).
- Si veda lo schema logico seguente.

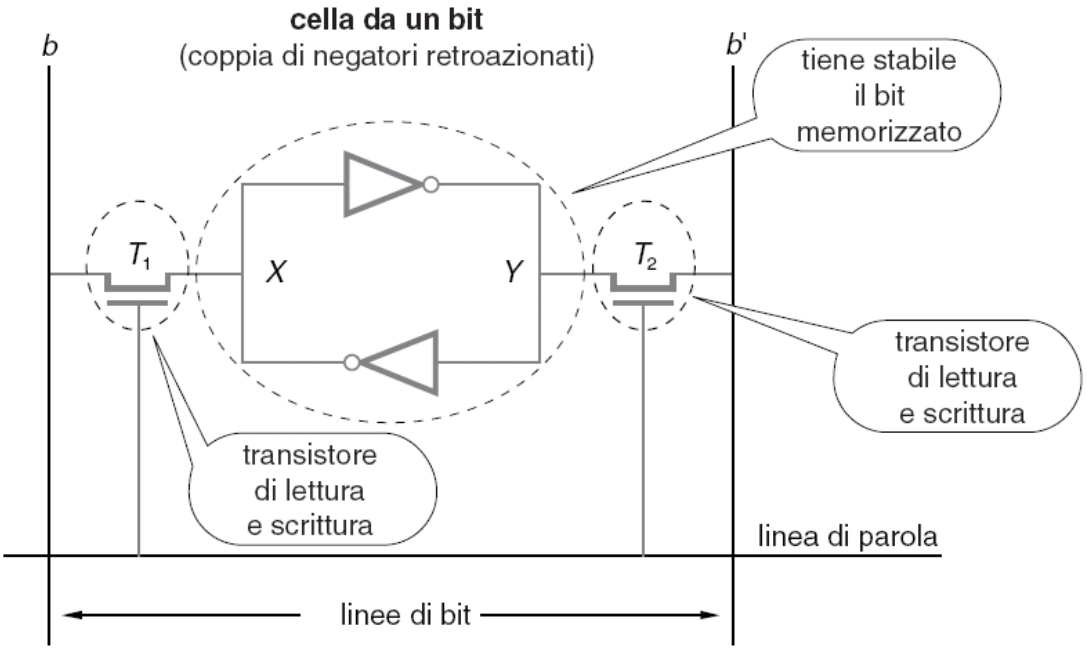
Esempio (3/3)



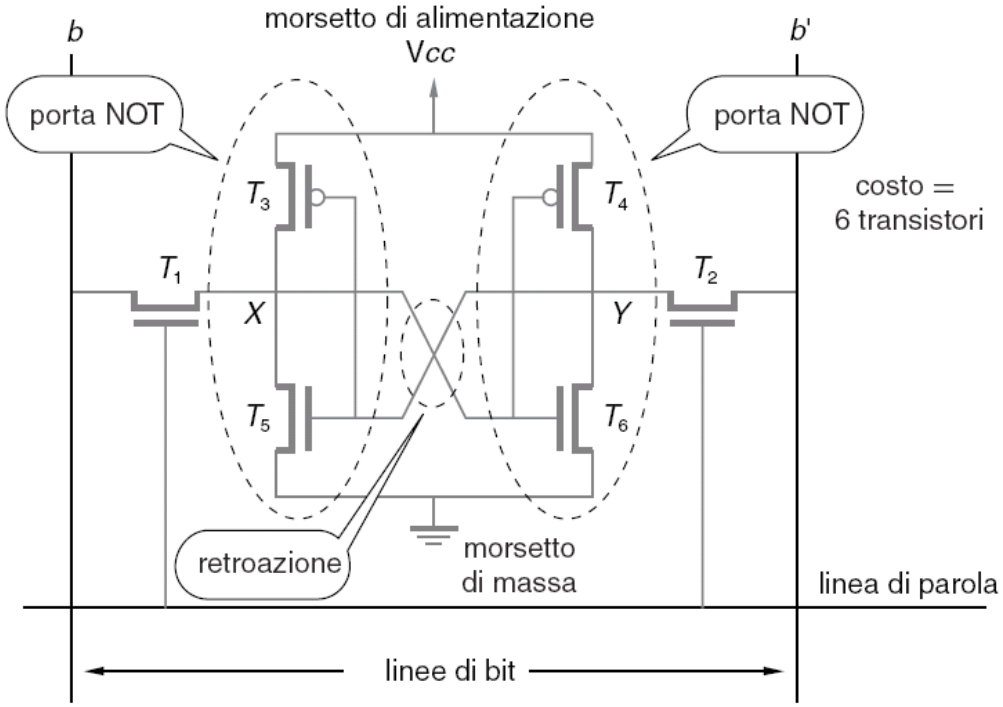
Cella di Memoria RAM Statica (1/3)

- In linea di principio la cella (da un bit) di memoria SRAM è un bistabile di tipo D (sincrono o asincrono, secondo i casi).
- In realtà, spesso la cella è realizzata risparmiando un certo numero di transitori.
- Di base la cella è costituita da due negatori retroazionati, che realizzano due stati stabili, e una coppia di transistori (= interruttori) per leggere e scrivere il bit.

Cella di Memoria RAM Statica (2/3)



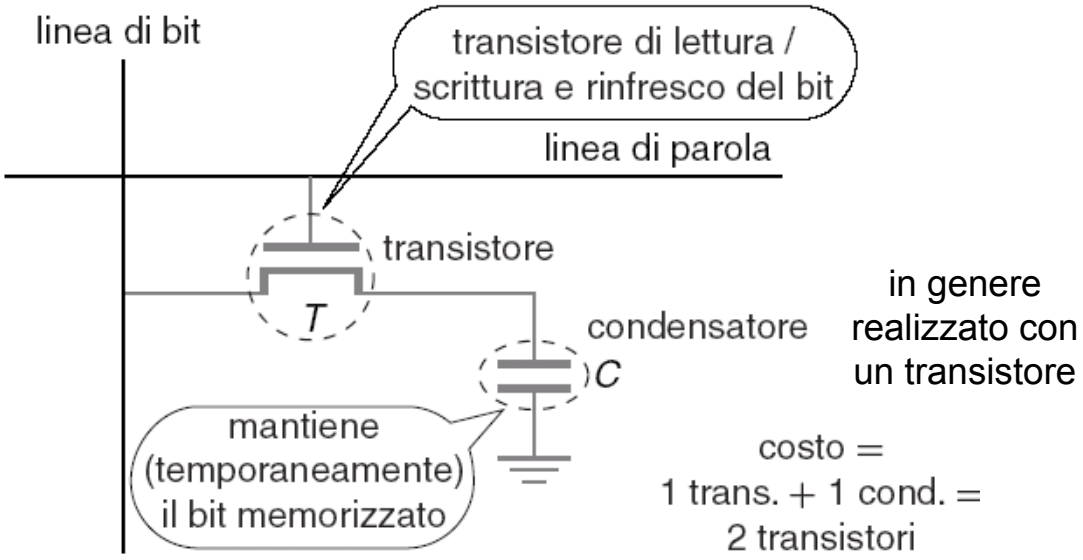
Cella di Memoria RAM Statica (3/3)



Cella di Memoria RAM Dinamica (1/2)

- *La cella SRAM* (da un bit) *consuma parecchi transistori* (circa 6 transistori per bit).
- **La cella di RAM dinamica** (DRAM, Dynamic RAM) ne consuma **uno solo**, o poco di più.
- Il bit è memorizzato come stato di carica del transistore.
- Lo stato della cella è temporaneo (dura solo qualche millisecondo, poi si altera in modo non riconoscibile).
- Lo stato della cella va rinfrescato (riletto e subito riscritto) periodicamente per mantenerlo in modo persistente.
- L'operazione di rinfresco è effettuata da un circuito apposito, integrato nel componente di memoria.
- La memoria DRAM consuma in media solo circa 1,5 transistori per bit (0,5 è per il circuito di rinfresco).

Cella di Memoria RAM Dinamica (2/2)

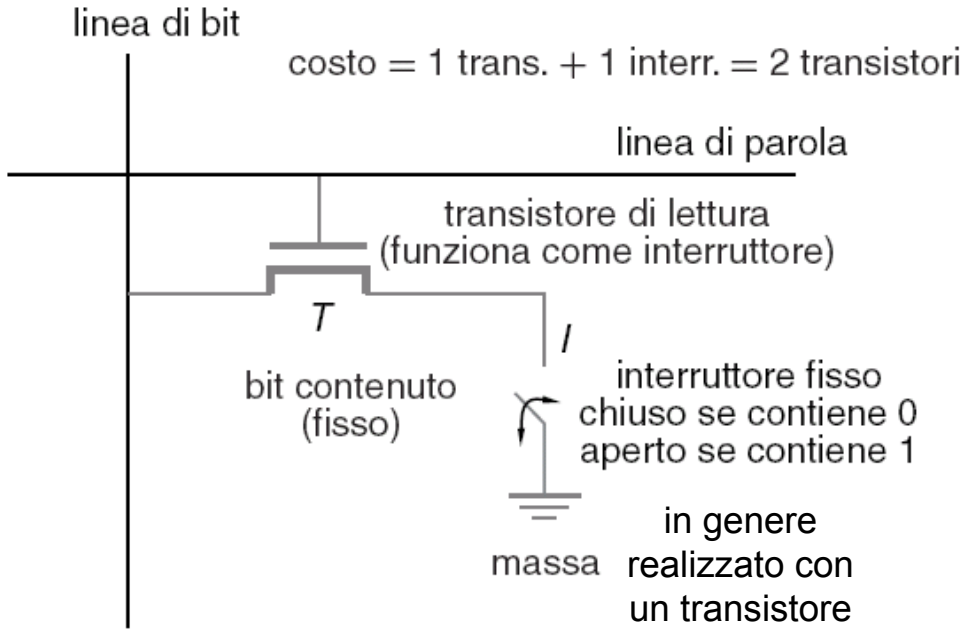


il circuito di rinfresco qui non è mostrato

Cella di Memoria ROM (1/2)

- **La memoria ROM** (*Read Only Memory*) **ha contenuto fisso**, *non modificabile*.
- La cella di memoria ROM è una variante semplificata della cella RAM dinamica.
- In sede di fabbricazione (o in seguito, secondo la tecnologia), lo stato del transistore che memorizza il bit viene definito in modo fisicamente persistente e non più modificabile.
- Ciò equivale a fissare lo stato dell'interruttore (aperto o chiuso), in modo definitivo.
- Esistono numerose tecnologie di tipo ROM.

Cella di Memoria ROM (2/2)



l'interruttore viene fabbricato in uno stato definito (aperto o chiuso)

Decodifica di indirizzo
Banco di memoria
Tecnologie di memoria

TECNOLOGIA DI MEMORIA II

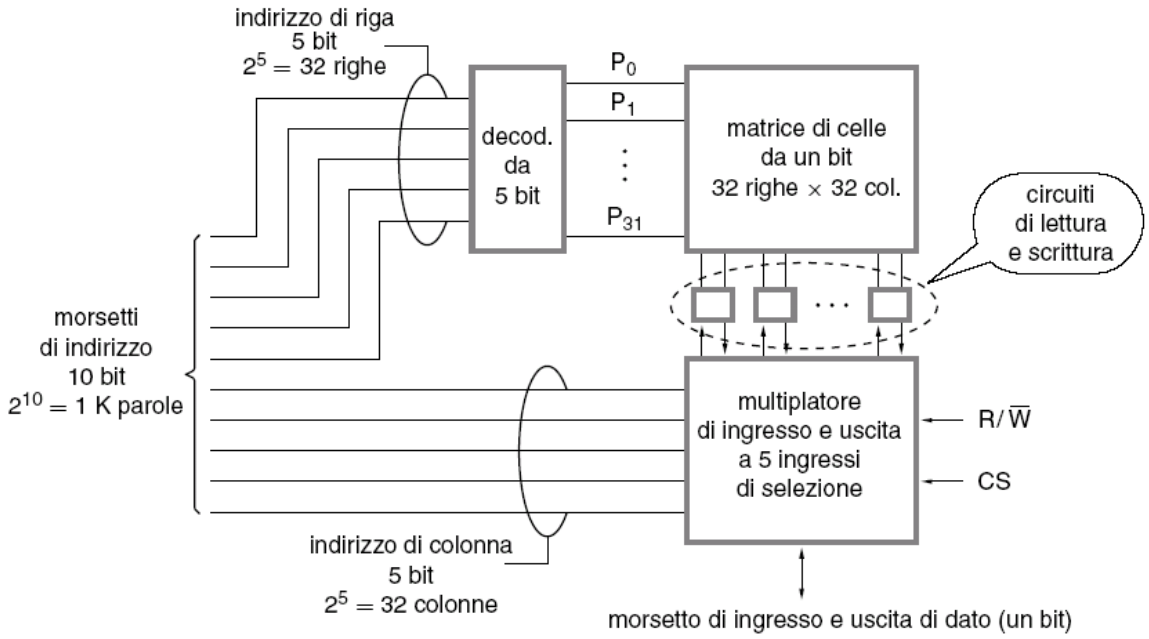
Decodifica di Indirizzo

- Per costruire memorie di capacità elevata, si aggregano componenti integrati di memoria con capacità inferiore.
- Come primo passo, la matrice di celle da un bit può avere organizzazione diversa da quella dichiarata per la memoria ed esternamente visibile dall'utilizzatore.
- L'indirizzo può essere decodificato in maniera parziale, o suddivisa.

Esempio

- Struttura esterna apparente della memoria:
  - dimensione indirizzo: 10 bit
  - dimensione dato (parola): 1 bit
  - capacità totale: 1024 × 1 bit = 1024 bit
- Struttura interna effettiva della memoria:
  - matrice di 32 × 32 bit = 1024 bit
- I 5 bit più significativi dell'indirizzo vengono decodificati e usati per accedere a una riga della matrice di bit (tra le 32 righe esistenti).
- I 5 bit meno significativi sono usati per selezionare (tramite mux) uno dei 32 bit di riga.

Decodifica Parziale



organizzazione interna effettiva della memoria e vista esterna

Altre Ottimizzazioni

- La decodifica dell'indirizzo può anche essere effettuata in due momenti diversi:
  - indirizzo di riga per primo,
  - indirizzo di colonna per secondo.
- Se si opera su parole di memoria collocate a indirizzi consecutivi, l'indirizzo di riga (o di colonna) non varia per diverse operazioni, lo si può tenere fisso e la decodifica è più veloce.
- Occorrono segnali di controllo aggiuntivi per gestire questi schemi di decodifica più sofisticati (segnali RAS e CAS, si veda il testo).

Banco di Memoria (1/2)

- Sono disponibili componenti integrati di memoria di svariate capacità (e velocità). Per esempio:

  64 K $\times$ 8, 1 M $\times$ 8, 1 M $\times$ 1, 256 M $\times$ 1, …

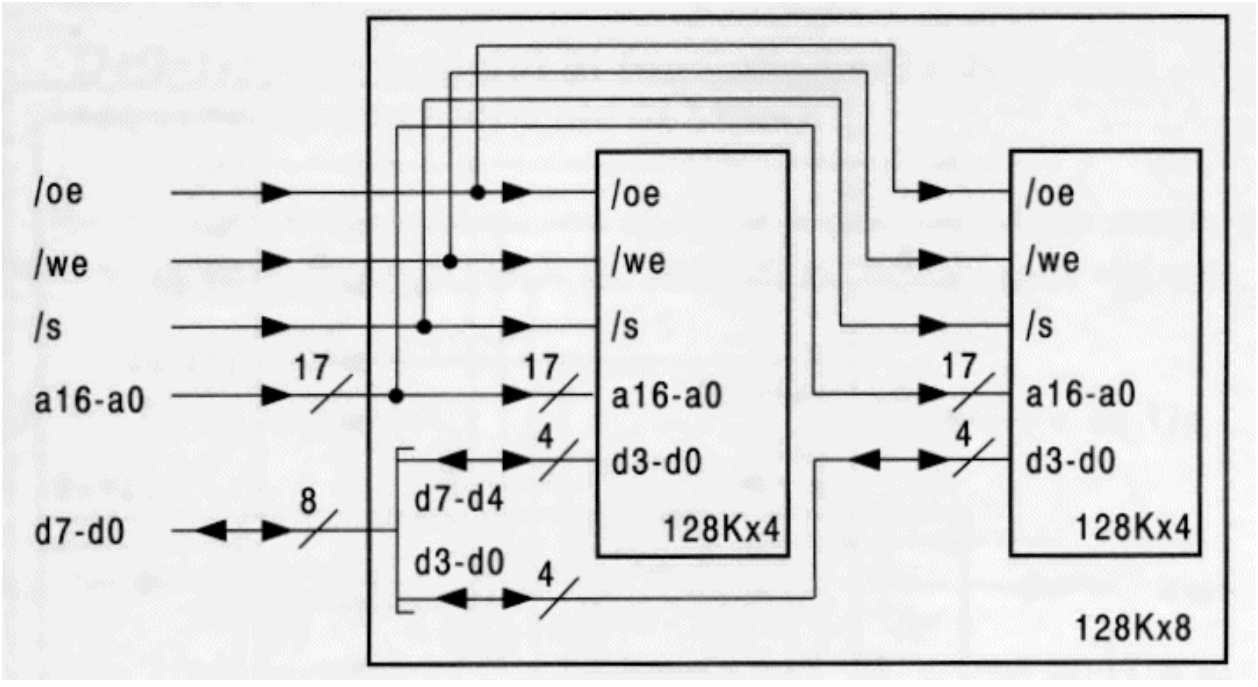
- Tuttavia, per ottenere memorie di capacità elevata occorre aggregare più componenti integrati di memoria (chip), realizzando un cosiddetto banco di memoria.

- Il banco di memoria deve avere una struttura possibilmente standard, ed essere scalabile (cioè estensibile per composizione).

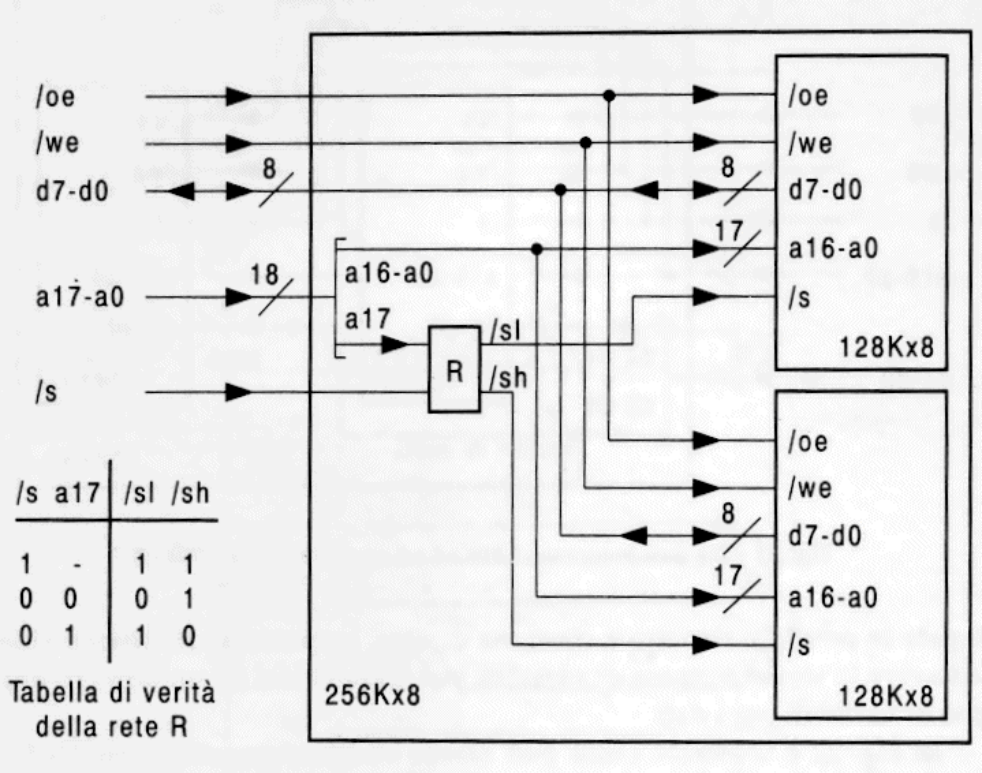
Banco di Memoria (2/2)

- I banchi di memoria hanno una struttura a matrice di componenti integrati (chip).
- ***Per aumentare la lunghezza della parola di memoria*, si compone una riga di chip di memoria, da usare in parallelo.**
- ***Per aumentare il numero di parole della memoria*, si compone una colonna di chip di memoria, da usare in esclusione.**

Banchi di memoria: espansione bus dati



Banchi di memoria: espansione bus indirizzi



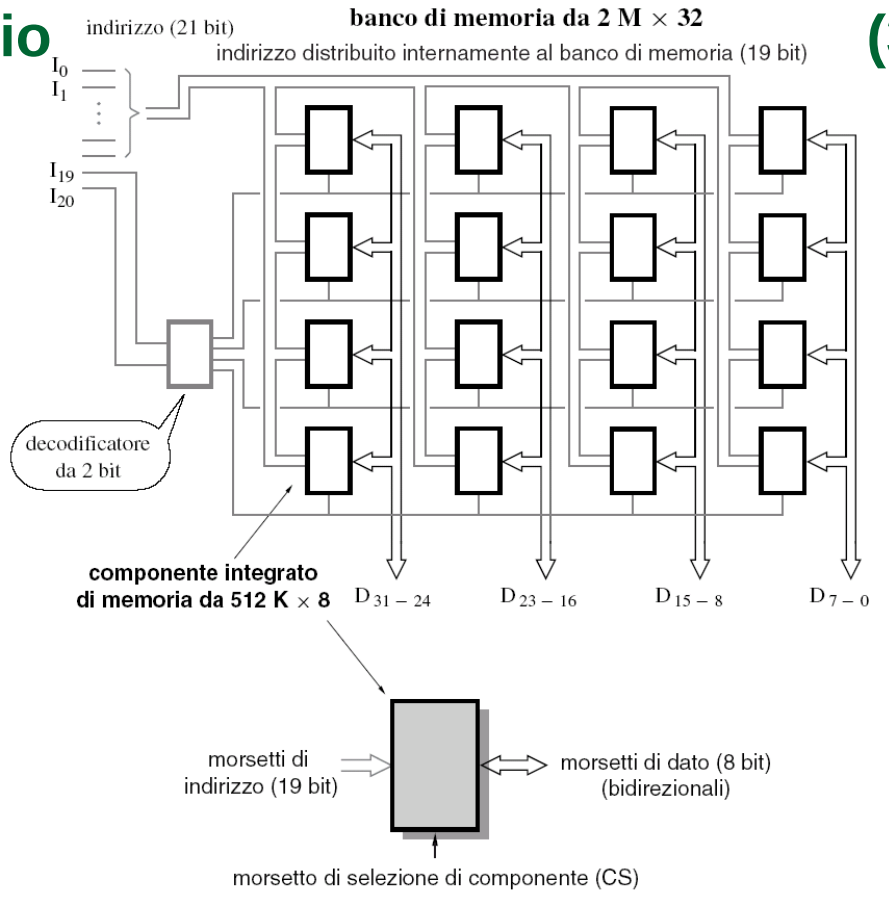
Esempio (1/3)

- Si supponga di disporre di 16 chip di memoria da 512 K × 32 bit ciascuno.
- Si desidera ottenere un banco di memoria da 2 M × 8 bit complessivi.
- Il numero di chip è sufficiente:
  - 16 (512 K × 8) = 64 Mbit
  - 2 M × 32 = 64 Mbit
- Occorre comporre una matrice 4 × 4:
  - 4 righe di 4 chip di memoria ciascuna

Esempio (2/3)

- L'indirizzo del banco è da 21 bit, perché $2^{21} = 2$ M.
- I 2 bit più significativi dell'indirizzo vanno al decodificatore, che genera in esclusione il segnale di attivazione (CS) di una delle 4 righe di componenti.
- I 19 bit meno significativi dell'indirizzo vanno in parallelo alle 4 righe, ma solo quella attiva li prende in esame.
- Le linee dati e il comando di lettura/scrittura vanno in parallelo alle 4 righe, ma solo quella attiva li prende in esame:
  - il comando viene interpretato ed eseguito,
  - la parola passa sulle linee dati (in lettura o scrittura).

Esempio (3/3)



Memoria Interallacciata

- Il banco di memoria, costituito da vari componenti integrati, può anche avere struttura interallacciata (*interleaved*).
- L'idea di fondo è che parole consecutive di memoria stiano in componenti integrati di memoria differenti, secondo vari schemi possibili di interallacciamento.
- Se si legge (o scrive) una successione di *n* parole consecutive (per esempio scandendo sequenzialmente un vettore di parole, numeri, ecc), le *n* letture (o scritture) hanno luogo in componenti diversi.
- È allora possibile effettuare gli *n* accessi tutti in parallelo, o almeno sovrapporli in parte temporalmente.
- Come risultato l'accesso sequenziale a un blocco di *n* parole è accelerato, rispetto al caso senza interallacciamento.

Tecnologie di Memoria

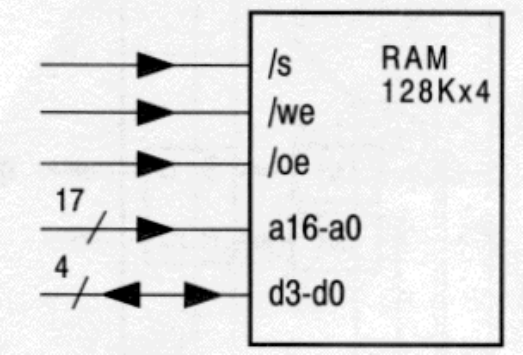
- Esistono svariate tecnologie di memoria.
- Tali tecnologie dipendono da vari fattori:
  - capacità della memoria;
  - tempo di accesso a una parola (in lettura o scrittura);
  - politica di accesso alla parola:
    - lettura e scrittura,
    - sola lettura,
    - programmabilità sul campo,
    - immodificabilità;
  - stabilità del contenuto: volatile o persistente;
  - e costo del componente.

RAM Statica (SRAM)

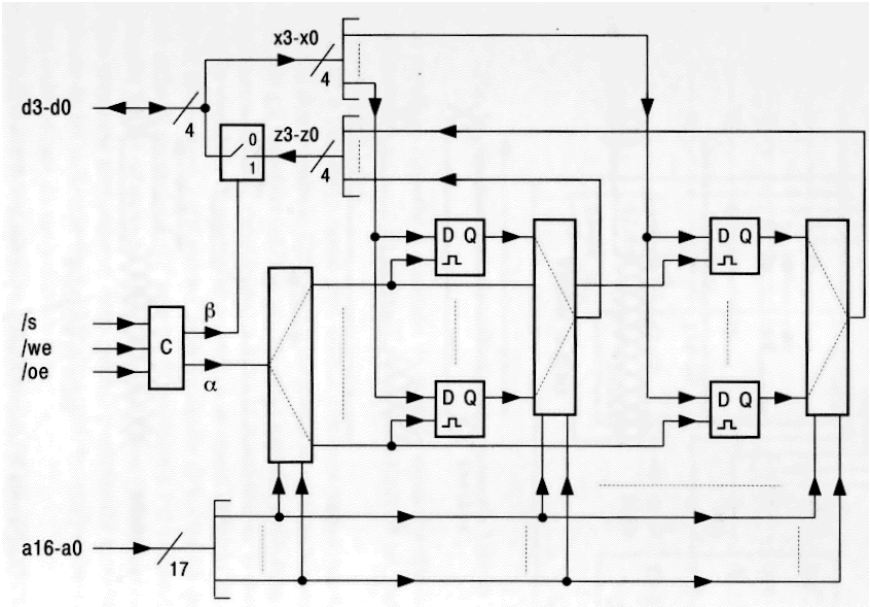
- Memoria RAM (*Random Access Memory*) realizzata come matrice di bistabili.
- Capacità medio-piccola (da K a M parole).
- Tempo di accesso molto breve (1 ns o meno).
- Funziona sia in lettura sia in scrittura (non necessariamente alla stessa velocità).
- Volatile: senza alimentazione il contenuto della memoria svanisce (si tampona con batteria).
- Usi: svariati, in particolare come banco di registri interni di processore e come cache.

RAM Statica: piedinatura

- /s → selezione
- /we → abilitazione scrittura
- /oe → abilitazione lettura
- a16-a0 → indirizzo
- d3-d0 → dati (var. bidirezionali)



RAM Statica: sintesi

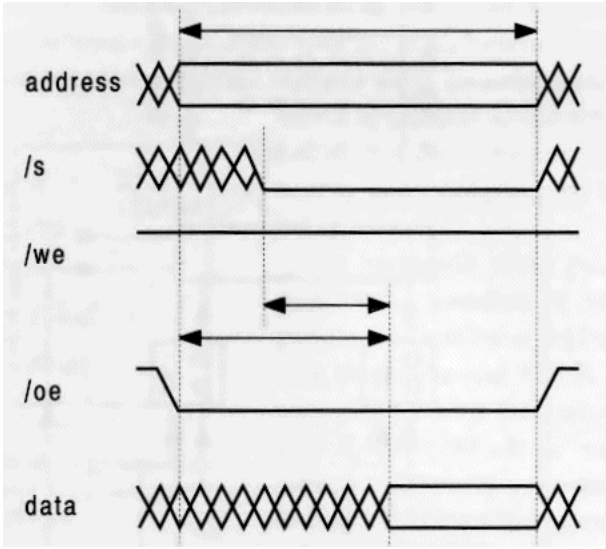


/s	/we	/oe	α	β
1	-	-	0	0
0	1	1	0	0
0	1	0	0	1
0	0	1	1	0
0	0	0	-	-

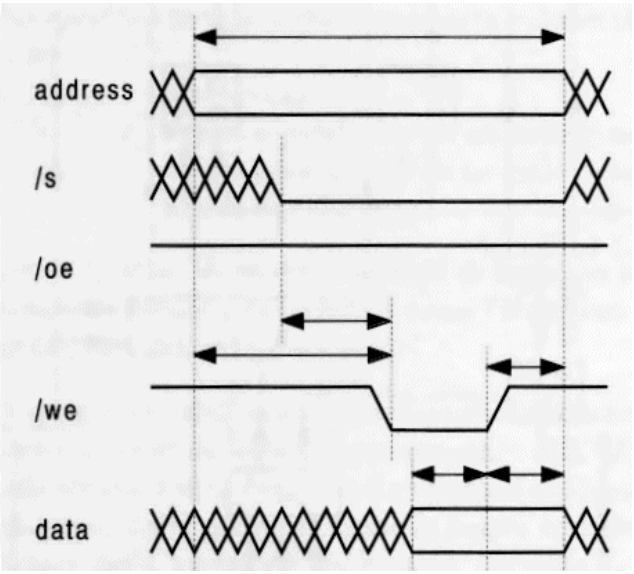
Tabella di verità della rete C

- Matrice di D-latch con logica combinatoria e porte 3-state per selezionare le locazioni di memoria e distinguere le operazioni di lettura da quelle di scrittura.

RAM Statica: diagramma di temporizzazione



Ciclo di lettura



Ciclo di scrittura

RAM Dinamica (DRAM) (1/2)

- La memoria SRAM consuma parecchi transistori per bit memorizzato (circa 6 transistori per bit).
- La tecnologia DRAM usa solo circa 1,5 transistori per bit memorizzato.
- Essa sfrutta il fenomeno dell'accumulo temporaneo di carica sul transistore.
- Internamente il componente di memoria DRAM contiene un circuito di rinfresco che rigenera le cariche sui transistori prima che queste svaniscano.
- La tecnologia memoria DRAM ha struttura molto regolare, e può raggiungere una densità di transistori estremamente elevata.

RAM Dinamica (DRAM) (2/2)

- Memoria RAM realizzata come matrice di transistori, ad altissima densità.
- Capacità grande-grandissima (da M a G parole), non conviene farla piccola.
- Tempo di accesso medio (10 ns - 100 ns).
- Funziona sia in lettura sia in scrittura (non necessariamente alla stessa velocità).
- Volatile: senza alimentazione il contenuto della memoria svanisce (batteria tampone).
- Usi: numerossisimi, la memoria centrale dei calcolatori normalmente è DRAM.

ROM

- Memoria ROM (Read Only Memory) realizzata come matrice di transistori in stato di apertura o chiusura fisso.
- Capacità grande (da M a G).
- Tempo di accesso medio (10 ns - 100 ns).
- Funziona in sola lettura.
- Persistente: il contenuto permane anche in assenza di alimentazione.
- Usi: per memorizzare programmi permanenti, non modificabili (software di base, e simili).
- Valida per grandi volumi di produzione.

Memoria Programmabile sul Campo (1/2)

- Serie di tecnologie di memoria derivate dalla tecnologia ROM, con raffinamenti vari.
- Capacità e tempo di accesso simili a ROM.
- Funzionano in sola lettura e sono persistenti.
- Ma sono programmabili sul campo (*field programmable*), cioè il contenuto della memoria è definibile dopo la fabbricazione.
- In alcune di tali tecnologie, si può riprogrammare la memoria più volte.
- Usi: piccoli volumi di produzione, prototipi.

Memoria Programmabile sul Campo (2/2)

- La programmazione è un'operazione speciale:
  - si arresta il funzionamento normale della memoria
  - si sostituisce l'intero contenuto (non parte di esso)
- La programmazione va fatta tramite una macchina programmatrice apposita:
  - PROM: programmabile una volta sola
  - EPROM: cancellabile più volte con raggi UV
  - EEPROM: cancellabile più volte elettricamente (si può anche scrivere un solo byte per volta)
- In EEPROM il circuito di programmazione è integrato nella memoria stessa.

Memoria FLASH

- Tecnologia di memoria abbastanza recente con tempo di accesso simile a quello della tecnologia DRAM (o solo di poco inferiore).
- Funziona sia in lettura sia in scrittura (la scrittura però è a blocchi di byte).
- Persistente: il contenuto permane anche in assenza di alimentazione (per alcuni anni).
- Molti usi e in rapido aumento:
  - dati multimediali (immagini statiche, sequenze audio e video)
  - programmi fissi ma periodicamente aggiornabili

Tabella Riassuntiva

tipo	categoria	modalità di cancellazione	scrittura byte	volatile	usi specifici
SRAM	lett/scritt	-	si	si	cache
DRAM	lett/scritt	-	si	si	mem. centrale
ROM	sola lett	nessuna	no	no	grandi vol.
PROM	sola lett*	nessuna	no	no	piccoli vol.
EPROM	sola lett*	luce UV	no	no	prototipi
EEPROM	sola lett*	elettrica	si (lenta)	no	prototipi
FLASH	lett/scritt	elettrica	a blocchi	no	multimedia

*Le memorie cancellabili vengono talvolta qualificate come “memorie prevalentemente a sola lettura” (*read-mostly*), invece che “a sola lettura” (read-only).