



UNIVERSITA DEGLI STUDI DI FOGGIA

Dipartimento di Agraria

Cdl in Ingegneria dei Sistemi Logistici per l'Agroalimentare

Corso integrato di Sistemi di Elaborazione

Modulo I

Prof. Crescenzo Gallo

crescenzo.gallo@unifg.it

La memoria principale

Organizzazione della memoria

La memoria principale è organizzata come un insieme di locazioni di uguale dimensione, ognuna delle quali è identificata tramite un numero progressivo ad essa associato, detto **indirizzo**, che rappresenta la posizione di quella locazione rispetto alla prima.

- ▶ Il contenuto delle locazioni non è immediatamente riconoscibile: non c'è distinzione esplicita tra istruzioni e dati e tra dati di tipo diverso.
- ▶ Una istruzione o un dato possono risiedere su più locazioni consecutive, se la dimensione della singola locazione di memoria non è sufficiente.
- ▶ Il parallelismo di accesso è definito dall'ampiezza della locazione.

0	01101101
1	10010110
2	00111010
3	11111101
1022	00010001
1023	10101001

Organizzazione della memoria

Operazioni sulla memoria principale

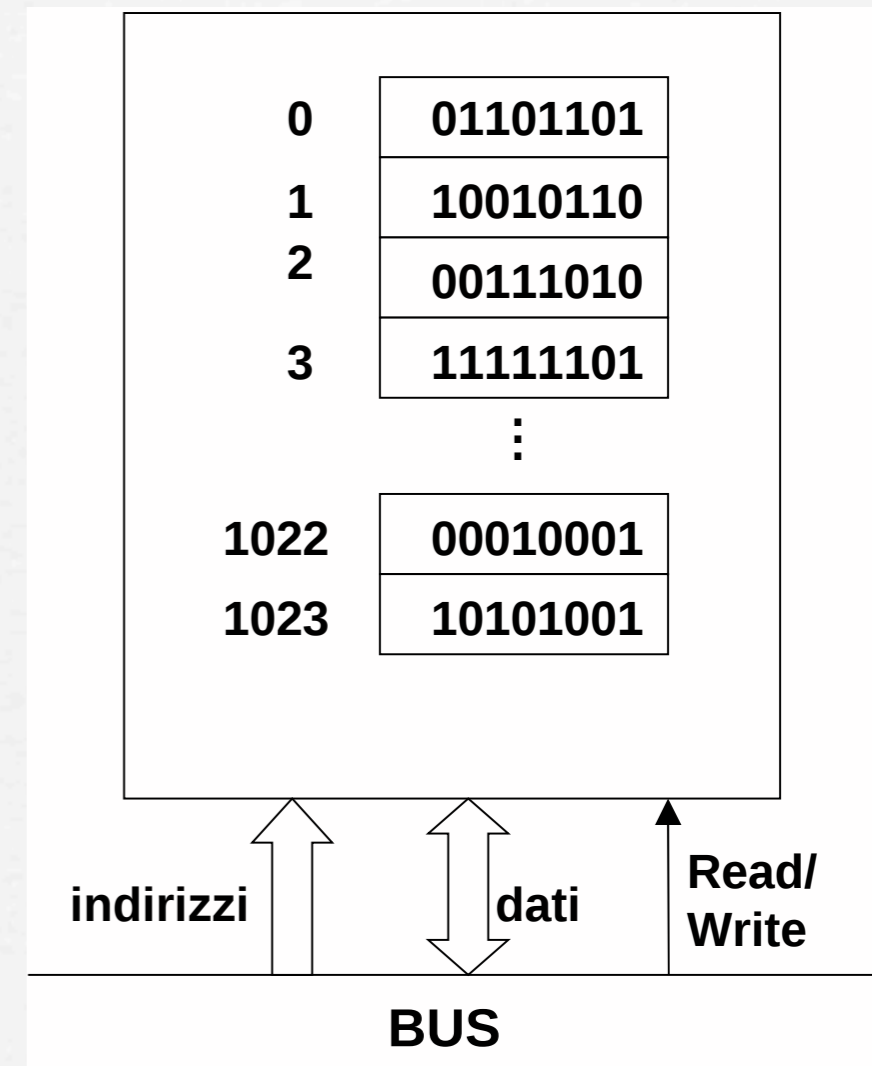
Il modulo di memoria principale è connesso al resto del sistema tramite il BUS.

In particolare, sono presenti tre gruppi di linee:

- linee **indirizzi**
- linee **dati**
- linee **read/write**

In ogni operazione è quindi necessario specificare:

- ▶ su quale registro si intende compiere l'operazione → *indirizzo*
- ▶ che tipo di operazione si intende realizzare → *Read/Write*
- ▶ in caso di scrittura, quale sia il *valore* da memorizzare

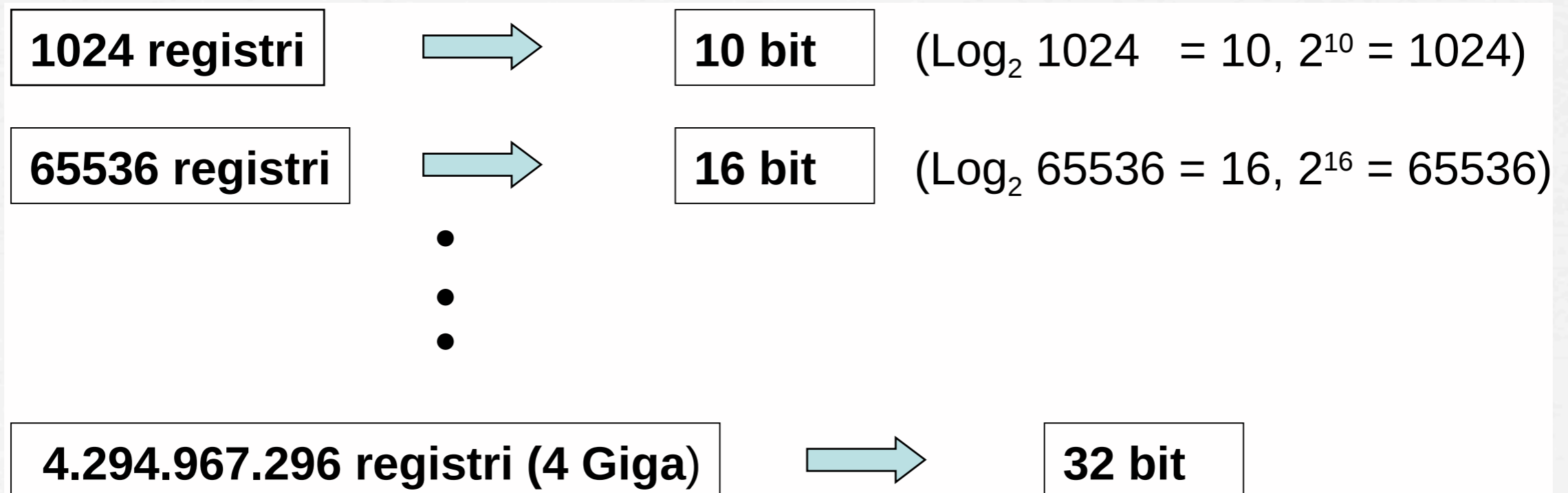


Organizzazione della memoria

Codifica degli indirizzi

- ▶ Se la memoria contiene N registri (locazioni) abbiamo bisogno di N indirizzi diversi.
- ▶ Di quanti bit ho bisogno per rappresentare N indirizzi diversi? Risposta: $\lceil \log_2 N \rceil$

Esempio:



Parametri della memoria principale

Capacità

Fornisce una misura della quantità di informazione che è possibile memorizzare. Questa dipende dall'ampiezza dei singoli registri e dal numero di registri contenuti. La capacità della memoria si misura in termini di byte (1 Megabyte = 2^{20} byte, 1 Gigabyte = 2^{30} byte, 1 Terabyte = 2^{40} byte).

Tempo di accesso

E' il tempo minimo che intercorre tra due operazioni (accessi) in memoria. Dipende dalla tecnologia di realizzazione della memoria.

Si misura in termini di secondi (1 nanosecondo = 10^{-9} secondi).

Canale DMA: un canale di accesso diretto alla memoria centrale da parte di una periferica.

Tipologie di memorie

Memorie RAM

RAM è l'acronimo di *Random Access Memory*.

Sta ad indicare che il tempo di accesso è costante per ogni locazione di memoria.

Hanno le seguenti caratteristiche:

- ▶ si possono realizzare operazioni sia di lettura che di scrittura;
- ▶ mantengono il loro contenuto finché è presente l'alimentazione (sono dette **memorie volatili**).



Tipologie di memorie

Memorie ROM

ROM è l'acronimo di *Read Only Memory*.

Sta ad indicare che il suo contenuto è inserito una volta per sempre all'atto della costruzione e non può più essere modificato o cancellato (a meno di particolari procedimenti — PROM, EPROM, EEPROM).

Hanno le seguenti caratteristiche:

- sono ***permanenti*** (NON volatili);
- anche in questo caso il tempo di accesso è costante.



Le memorie RAM

Esistono due tipi di memoria RAM:

RAM dinamica o DRAM (Dynamic Random Access Memory)

- Alta densità di integrazione, economica, lenta, bassa potenza di alimentazione
- *Dynamic*: è necessario rigenerare i contenuti periodicamente (refresh)



RAM statica o SRAM (Static Random Access Memory)

- Bassa densità di integrazione, costosa, veloce, alta potenza di alimentazione
- *Static*: il contenuto viene mantenuto finché è presente l'alimentazione



RAM ECC (Error Correction Code)

- Tentano di correggere localmente eventuali errori di scrittura e lettura

Il packaging delle memorie

- Fino all'inizio degli anni '90 le memorie venivano realizzate su chip singoli.
- Oggi si monta un gruppo di chip, tipicamente 8 o 16, su una piccola scheda stampata.
- Si parla di:
 - **SIMM** (Single In line Memory Module): la fila di connettori si trova da un solo lato della scheda
 - **DIMM** (Dual In line Memory Module): i connettori si trovano su ambedue i lati della scheda



Le memorie ROM

- Le memorie ROM vengono usate all'interno del calcolatore per memorizzare i programmi per l'avvio all'accensione (**bootstrap**), i quali devono rimanere memorizzati anche quando l'alimentazione viene a mancare. Questi sono, inoltre, programmi e dati che, una volta memorizzati, non devono essere più modificati.
- In generale questo tipo di memorie si usa per memorizzare il **firmware**, programmi e dati che sono memorizzati in maniera permanente su un qualunque dispositivo:
 - Cellulari
 - Lettore/masterizzatori DVD
 - Navigatori
- Allo stato attuale si usano le ROM programmabili, le EEPROM (Electrically Erasable Programmable Read-Only Memory), dette anche memorie **FLASH**.

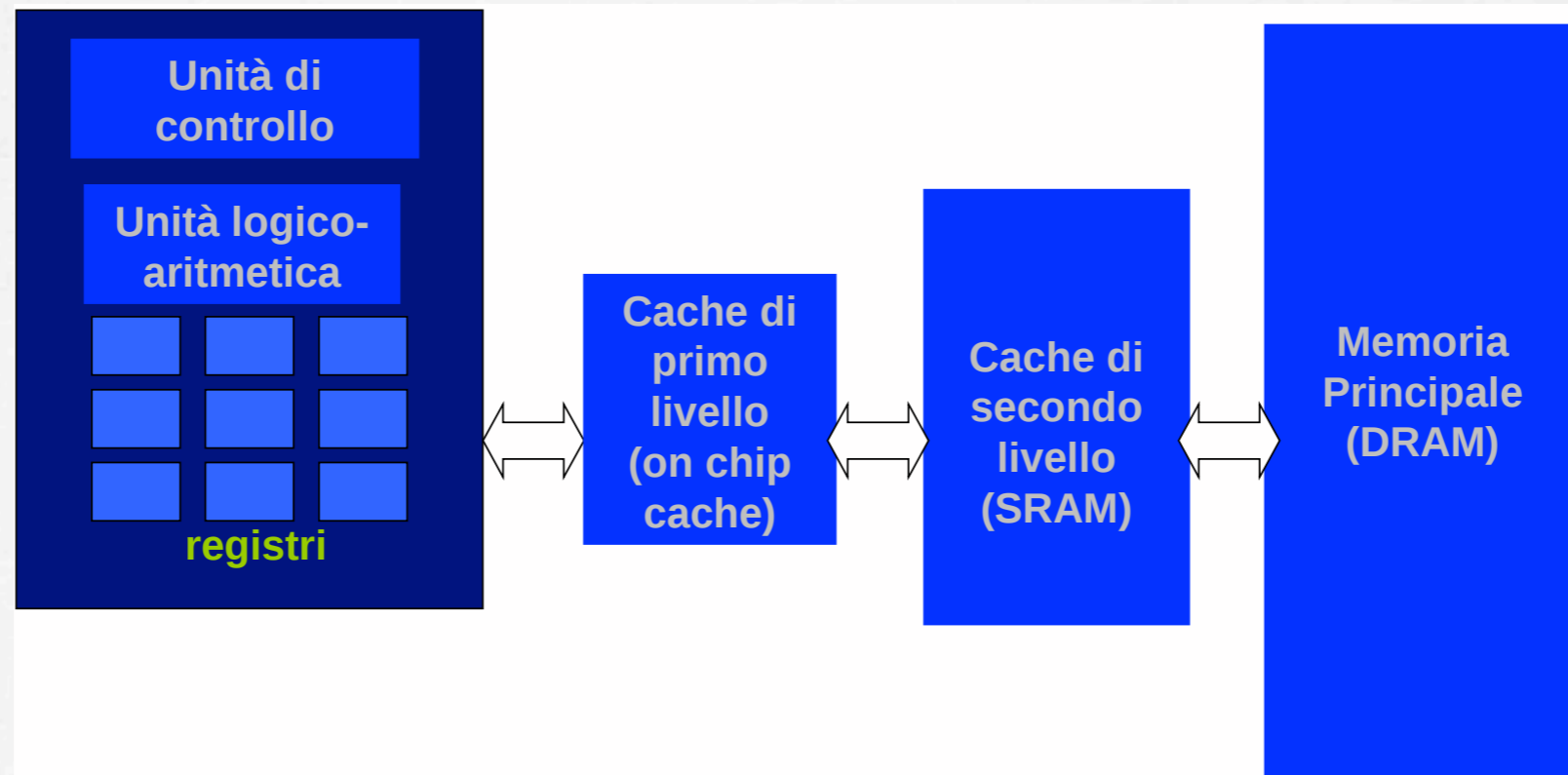


Organizzazione di un sistema di memoria

- Requisiti ideali di un sistema di memoria: **capacità** infinita, **velocità** infinita.
- Ma la situazione reale è:
 - le memorie capienti ed economiche (DRAM) sono lente;
 - le memorie veloci (SRAM) sono costose e meno integrabili.

Come realizzare un sistema di memoria che sia capiente, economico e veloce?

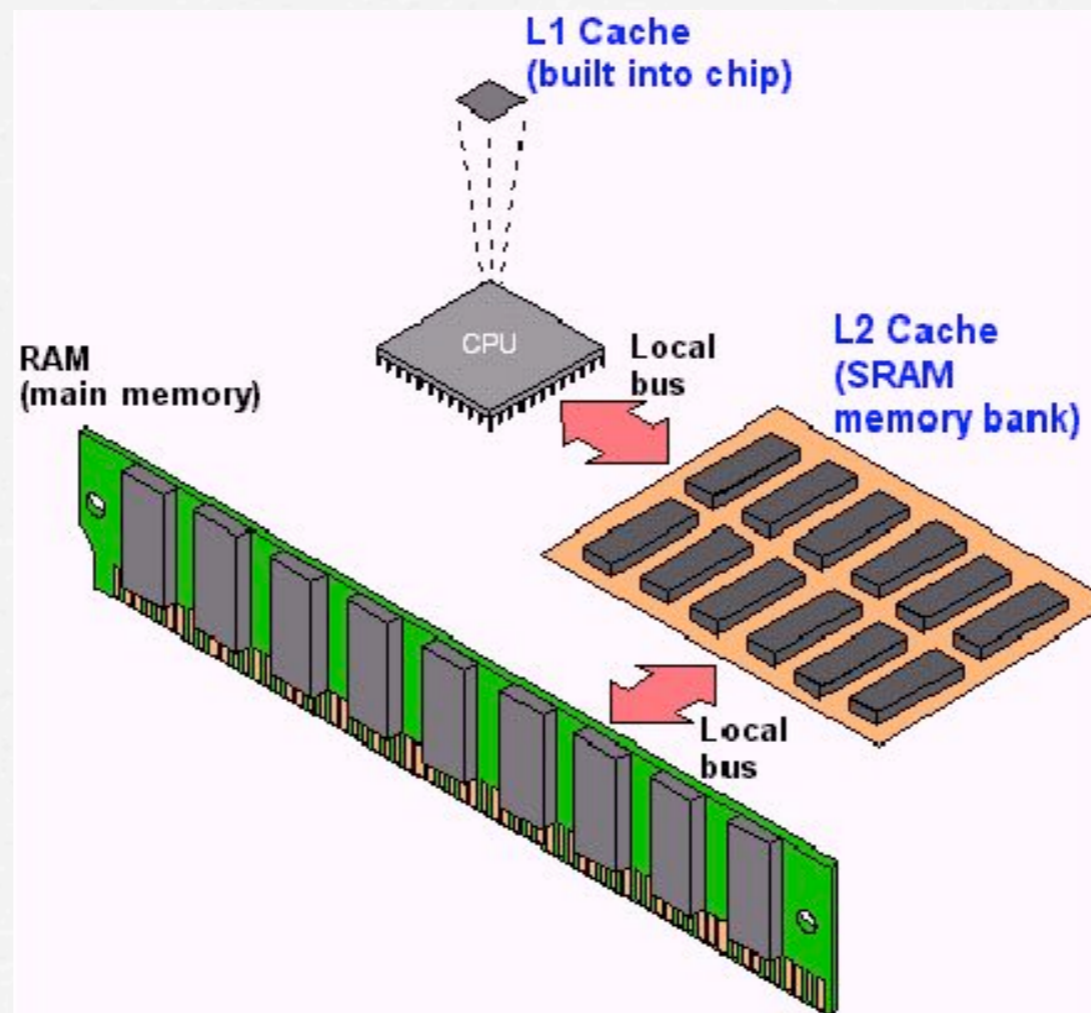
Un sistema basato su una gerarchia di memoria.



Realizzazione di un sistema di memoria

- Il sistema di memoria è composto da moduli di memoria con caratteristiche diverse e organizzati a livelli.
- Tra CPU e memoria principale viene posto un modulo di memoria intermedio (cache), ad accesso veloce, ma di capienza limitata.
- I dati memorizzati sono distribuiti sui vari moduli e possono essere trasferiti tra moduli adiacenti.

La distribuzione è realizzata in maniera da cercare di memorizzare i dati e le istruzioni richiesti più frequentemente nella cache, in modo che la CPU possa accedervi velocemente.



- Il tempo di propagazione del segnale (è un vincolo per il tempo di accesso) è minore.
- Una memoria grande ha bisogno di indirizzi grandi (maggiori tempi di decodifica).
- Il miglioramento delle prestazioni dovuto alla memoria cache si basa sul **principio di località del riferimento**:
i dati usati più di recente saranno utilizzati ancora nel recente, cioè nel prossimo futuro.

Funzionamento della cache

L'**algoritmo** seguito per la cache è il seguente:

1. Il dato viene cercato prima nella cache.
2. Se è presente abbiamo finito (*cache hit*).
3. Se non è presente, si legge in RAM e si mette una copia nella cache (*cache miss*).
4. Se non c'è spazio di solito si sovrascrivono i dati utilizzati meno di recente (strategia **LRU** *Least Recently Used*).
5. Per le scritture generalmente si scrive la RAM e si aggiorna la copia, se c'è.

