Architettura degli Elaboratori - Organizzazione Elaboratori

Andrea Malvezzi

25 Settembre, 2024

Contents

1 Il modello Von Neumann

Tutti i PC moderni seguono un modello chiamato "Von Neumann",
basato sull'impiego della memoria non solo per salvare i dati ma anche per l'esecuzione di programmi.
Questi sono trasferiti attraverso il (sotto)bus dati e salvati in indirizzi di memoria, indicati seguentemente alla CPU dal bus indirizzi.

1.1 I bus

Un bus è un agglomerato di collegamenti elettrici paralleli
usati per trasportare informazioni da un punto A a un punto B in un sistema.

2 La CPU

La CPU è un componente comparabile al "cervello" di un PC, si occupa di eseguire i programmi nella memoria centrale.
Questa è composta da:

- **CU** (control unit) , che legge e interpreta le istruzioni;
- **ALU** (Arithmetic Logic Unit), che esegue le operazioni (es. *AND*, *addizioni*, ...);
- i **registri**, dove vengono salvati i risultati temporanei.

2.1 Alcuni registri importanti

- **PC** (Program Counter): indica la prossima istruzione in memoria;
- **IR** (Instruction Register): memorizza l'istruzione che si sta per eseguire;
- **MAR** (Memory Address Register): salva l'indirizzo in memoria della prossima Read/Write;
- **MDR** (Memory Data Register): comunica col bus dati;
- **PSW** (Program Status Word): "meta data" sull'ultima istruzione eseguita.

3 Il ciclo FDE

Il ciclo **Fetch-Decode-Execute** è diviso in tre fasi:

- acquisizione dalla memoria di un'istruzione. Qui:
  - il contenuto del registro PC viene scritto sul MAR e viene attivata la Read;
  - il contenuto nell'indirizzo scritto nella MAR viene scritto sul MDR;
- identificazione del tipo di operazione da eseguire. Qui:
  - il contenuto di MDR viene copiato su IR;
- effettuazione operazioni corrispondenti all'istruzione da eseguire. Qui:
  - l'istruzione viene eseguita dall'ALU;
  - gli eventuali operatori vengono caricati nei registri tramite MAR / MDR;
  - al termine dell'esecuzione il risultato ottenuto viene spostato sul registro di destinazione.
    Se serve si scrive in memoria con MAR / MDR e linea Write.

Al termine di questo ciclo, ricomincia da capo.

4 La Control Unit

Una CU si può realizzare puramente come HW con istruzioni prefissate (costoso) o come HW con una parte di microprogrammazione,
per rendere la CPU più versatile.

5 La Arithmetic Logic Unit

L'ALU risiede in una parte della CPU chiamata **Data path**, assieme ai suoi componenti I/O.
Questo agglomerato viene utilizzato durante la fase di *execute* del ciclo FDE, creando un nuovo loop detto "data path loop", governato da un **clock**.

5.1 Il clock

Un clock equivale ad $1/F$, con $F =$: Frequenza di lavoro della CPU, misurata in Hz.
La frequenza equivale al numero di cicli svolti in un secondo.
Il ciclo di clock equivale alla durata del ciclo di data path, il quale verrà eseguito un numero di volte pari alla quantità di istruzioni da svolgere in un programma.
Questa nuova durata viene chiamata **"durata dell'istruzione ISA"**.

6 Configurazioni CISC e RISC

La microprogrammazione permise di realizzare due tipi di Computer:

- **CISC**: Complex Instruction Set Computer;
- **RISC**: Reduced Instruction Set Computer, capace di eseguire istruzioni più semplici in meno tempo, talvolta evitando la microprogrammazione;

La prima configurazione mette enfasi sull'HW, mentre la seconda sul SW.
Ecco spiegato come si può realizzare una microarchitettura in entrambe le modalità.

7 Migliorare le prestazioni

7.1 Pipelining

Il pipelining è una tecnica che permette di eseguire in parallelo più istruzioni (cicli FDE), usando parti diverse della CPU.

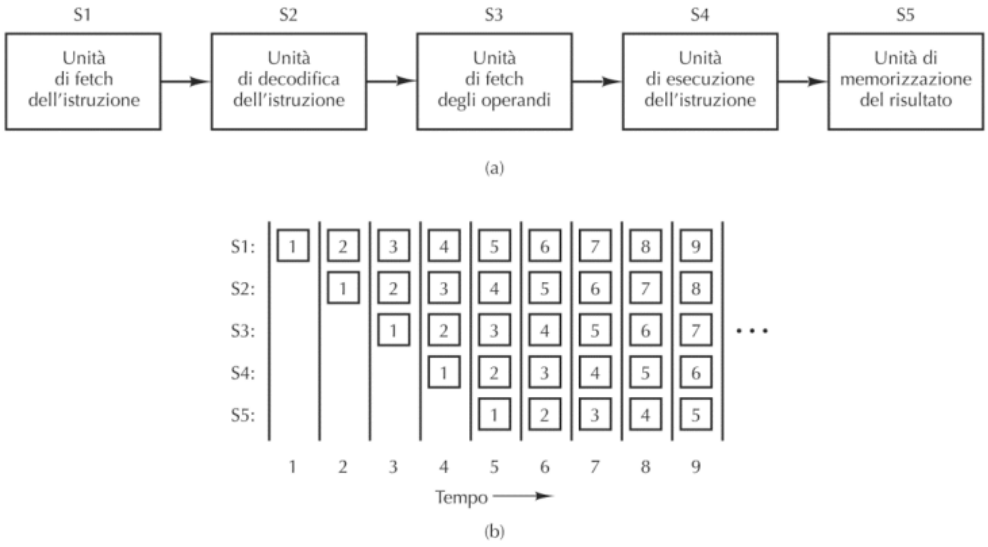

Figure 1: Un esempio di pipelining.

7.2 Multicore

In una CPU si possono simulare multiple CU e ALU per eseguire più istruzioni alla volta, **in parallelo**.

7.3 Diverse forme di parallelismo

Esistono 3 grandi tecniche per sfruttare il parallelismo, ovvero:

7.3.1 Gli array computer

Più processori eseguono la stessa istruzione su dati diversi.
Questa configurazione è detta **SIMD**: Single Instruction-stream, Multiple Data-stream.

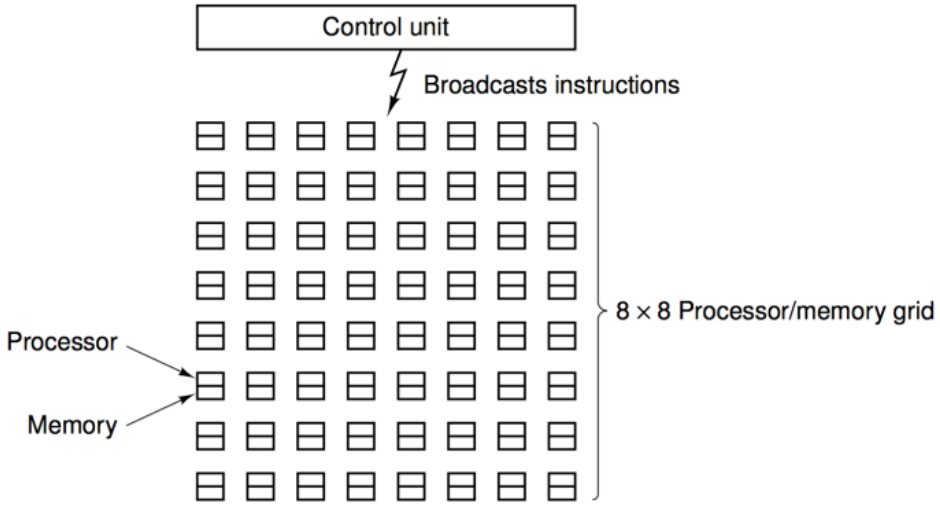

Figure 2: Negli Array Computer, ogni processore esegue la stessa istruzione, ma ha una sua memoria personale.

7.3.2 Multiprocessori

Molti processori condividono una memoria e possono eseguire istruzioni diverse tra loro.
Questa configurazione è detta **MIMD**: Multiple instruction-stream, Multiple data-stream.

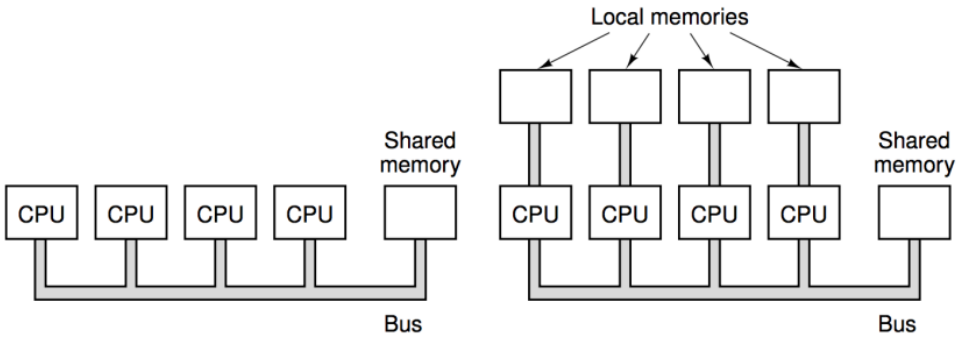

Figure 3: In questa configurazione i core possono avere o non avere una memoria locale, ma ne condivideranno sempre una.

7.3.3 Multicomputer

Infine ci sono i Multicomputer, dove si hanno svariati PC che non condividono una memoria e che comunicano scambiandosi messaggi.

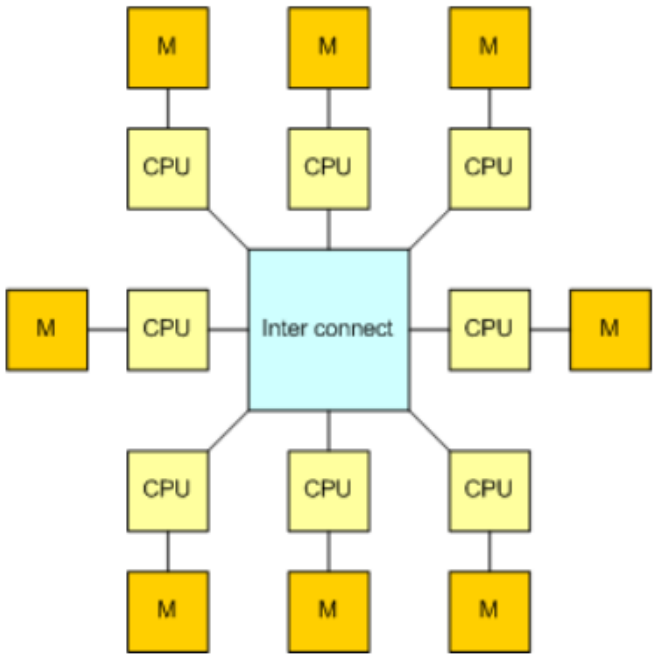

Figure 4: Esempio di configurazione Multicomputer.

8 Le memorie

Esistono tipi di memoria diversi:

- volatile: il dato rimane fino allo spegnimento del dispositivo;
- persistente: il dato rimane anche quando si spegne il dispositivo;
- on-line: i dati sono sempre accessibili;
- off-line: la memoria deve essere montata per accedere ai dati al suo interno;

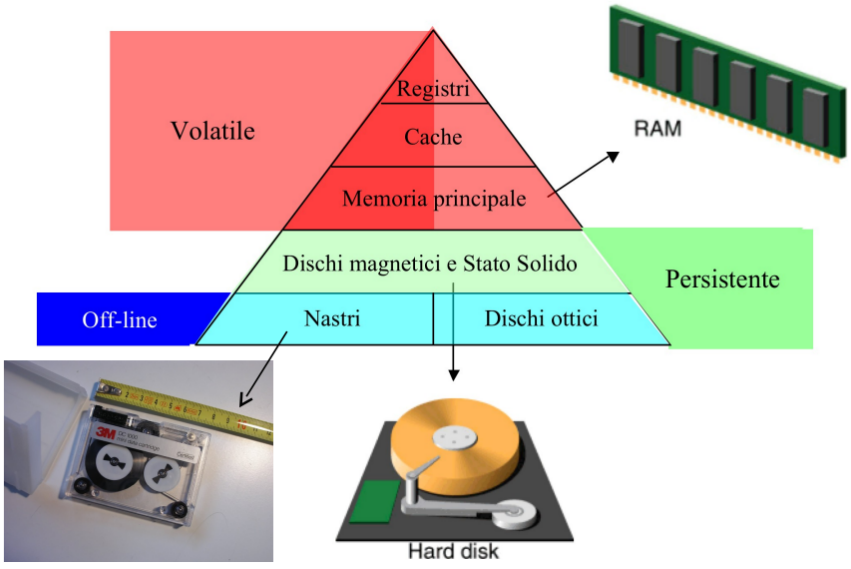

Figure 5: Salendo aumenta il costo per byte.

8.1 Salvare in memoria

La memoria è divisa in banchi da 8 byte ma ovviamente i computer lavorano con dati molto più grandi di così. Per salvare questi dati in memoria, li si "frammenta" nei loro byte e questi si salvano seguendo due filosofie:

- big endian: assegnando indirizzi da sx a dx (dal byte con valore inferiore a quello di valore maggiore);
- little endian: assegnando indirizzi da dx a sx (dal byte con valore maggiore a quello di valore inferiore);

8.2 La memoria cache

La cache è una memoria poco capiente ma veloce, che si usa per salvare dati a cui si accede di frequente.
Quando la CPU necessita di un dato, lo cerca nella cache e se non lo trova riprova nella memoria, per poi caricarlo se l'operazione termina con esito positivo.

8.2.1 Il principio di località

Quando si accede alla memoria più volte in uno span di tempo corto, è probabile che si stia accedendo a blocchi di memoria contigui.
Per questo motivo si caricano spesso zone di memoria contigue all'interno della cache: così facendo aumenta la probabilità di risparmiare tempo con gli accessi successivi.

8.3 Gli Hard Disk (HD)

Un HD è un dispositivo elettro-magnetico per salvare dati.
Possiede diverse parti:

- testina: componente montato su un braccio detto "di accesso", usato per scrivere sui piatti del disco;
- traccia: sequenza circolare di bit sui piatti;
- settore: porzione di traccia contenente n bit.

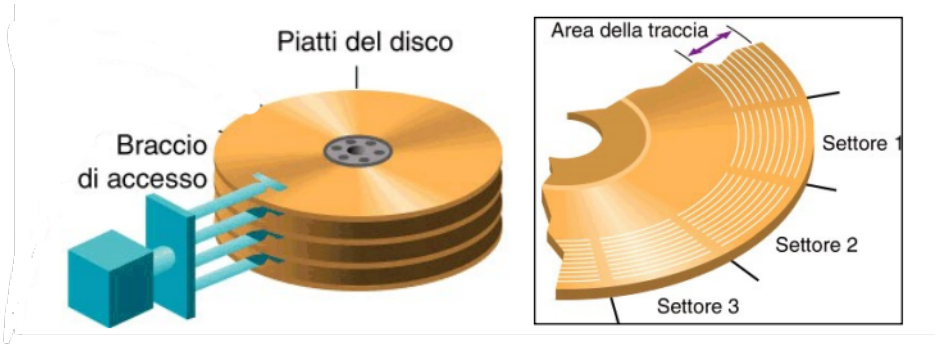

Figure 6: Composizione di un HD.

8.4 Memorie allo stato solido

Sono dispositivi completamente elettronici, quindi anche più resistenti (si ha un'assenza di componenti meccaniche) e veloci, ma meno capienti.

8.5 Memoria RAID

Per ridurre il gap di velocità tra CPU e memoria, si è pensato di utilizzare più dischi in parallelo.
Tale configurazione è detta RAID, e si suddivide in 5 livelli:

- livello 0: Non-reduntant data striping. Si hanno più dischi che salvano blocchi di dati (stripes) diversi, senza backup;
- livello 1: Redundant data striping. Come il livello 0, ma con backup;
- livello 2: Data striping at bit level. Spezza i dati in bytes piuttosto che in blocchi;
- livello 3: Bit-interleaved parity. Come il precedente, ma con un bit di parità per il controllo dell'errore;
- livello 4: Block-interleaved parity. Come il livello 0, ma con bit di parità;
- livello 5: Block-interleaved distributed parity. Come il precedente, ma i bit di parità sono disposti su più dischi per evitare errori.

8.6 CD

I CD sfruttano un principio ottico: sono composti da zone piane (land) e forate (pit) e in base alla sequenza con cui vengono lette dal PC, queste rappresentano 0 oppure 1.

9 Le schede grafiche

Le GPU possono essere programmate in appositi linguaggi (come CUDA-C) per diventare in grado di eseguire programmi non grafici (vedi *miners di Bitcoin*).

10 Conclusioni

Come dimostrato, i PC sono composti da più componenti, ognuno dei quali svolge un compito unico e comunica con gli altri mediante BUS, ai quali si

collegano grazie ad un **controller** e ad un **arbitro** (per evitare di utilizzarne uno al contempo stesso).