



DMIF, Università di Udine

Tecnologie Digitali per il Cibo e la Ristorazione

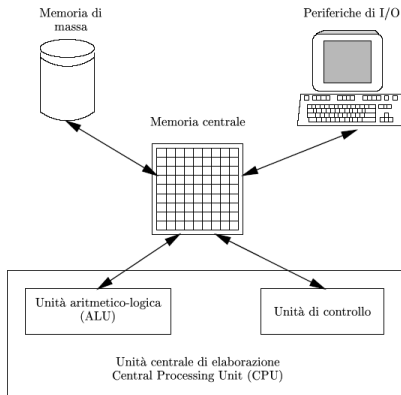
Architettura del calcolatore

Andrea Brunello

andrea.brunello@uniud.it

A.A. 2021–2022

- I calcolatori moderni sono basati sull'architettura detta macchina di von Neumann (1945)

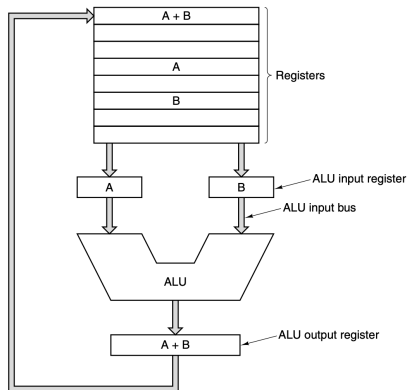




- **CPU**, che costituisce il “cervello” del calcolatore e contiene:
 - Unità di controllo
 - Unità aritmetico-logica + registri
- **Memoria centrale** (RAM) che immagazzina dati e istruzioni dei programmi
- **Unità di input** (es. tastiera, mouse)
- **Unità di output** (es. monitor, memoria di massa)
- **Bus**, un canale che collega tutti i componenti fra loro

- La CPU esegue i programmi memorizzati nella memoria principale, prelevandone le istruzioni, esaminandole, ed eseguendole in sequenza
- La CPU ha diversi componenti:
 - L'**unità di controllo** (CU) trasferisce dati e istruzioni da/verso la memoria principale
 - L'**unità aritmetico-logica** (ALU) esegue semplici operazioni aritmetiche e booleane
 - I **registri** costituiscono una memoria di lavoro interna alla CPU estremamente veloce:
 - Program Counter (PC): punta alla prossima istruzione che deve essere recuperata per l'esecuzione dalla memoria
 - Instruction Register (IR): memorizza l'istruzione correntemente in esecuzione



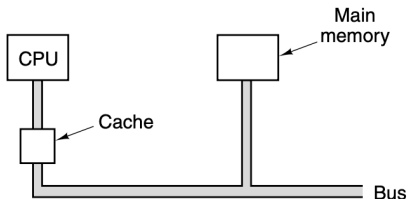


Le moderne CPU hanno diverse ALU che operano in parallelo, e specializzate per compiti diversi

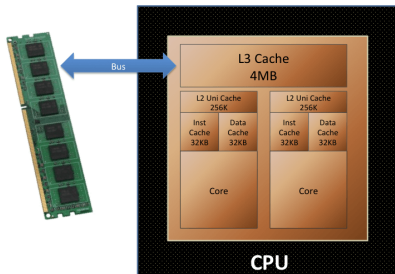


- La CPU esegue ciascuna istruzione seguendo una piccola serie di passi:
 - 1 Recupera l'istruzione da eseguire dalla memoria (puntata dal PC), e memorizzala nell'IR
 - 2 Cambia il PC affinché punti alla successiva istruzione da eseguire
 - 3 Determina il tipo di istruzione recuperata dalla memoria
 - 4 Se l'istruzione necessita di dati (parole) contenuti in memoria, determina dove si trovano
 - 5 Recupera i dati, se necessari, e memorizzali nei registri della CPU
 - 6 Esegui l'istruzione
 - 7 Ritorna al passo 1

- Le CPU sono da sempre più veloci della memoria principale
- Quando la CPU richiede dei dati in memoria, possono trascorrere diversi cicli (ciclo = tempo necessario per l'esecuzione di una semplice operazione) prima che tali dati vengano trasferiti sui registri
- Soluzione: creare una memoria intermedia, piccola e molto veloce, la **cache** (circa 3–5 volte più lenta dei registri)



- Tipicamente la cache è organizzata su più livelli: L1, L2, L3, ...
- Al crescere del livello, aumenta la capienza, aumenta la “distanza” dal processore, e diminuisce la velocità
- Tipicamente si va da qualche decina di KB per la cache di livello L1, all’ordine di MB per la cache di livello L3





- Quali dati portare in memoria cache? Intuitivamente, quelli che verranno utilizzati dalla CPU
- Principio di **località**: se la CPU ha richiesto dei dati in memoria, sul breve periodo probabilmente utilizzerà dei dati a loro vicini (temporalmente e/o spazialmente)
- Quando è necessario un dato, la CPU lo cercherà prima in cache L1; se non disponibile, in cache L2; e così, via, fino eventualmente a richiederlo alla memoria

| Core i7 Xeon 5500 Series Data Source Latency (approximate) | | [Pg. 22] |
|--|-------------------------------------|----------|
| local L1 CACHE hit, | ~4 cycles (2.1 - 1.2 ns) | |
| local L2 CACHE hit, | ~10 cycles (5.3 - 3.0 ns) | |
| local L3 CACHE hit, line unshared | ~40 cycles (21.4 - 12.0 ns) | |
| local L3 CACHE hit, shared line in another core | ~65 cycles (34.8 - 19.5 ns) | |
| local L3 CACHE hit, modified in another core | ~75 cycles (40.2 - 22.5 ns) | |
| remote L3 CACHE (Ref: Fig.1 [Pg. 5]) | ~100-300 cycles (160.7 - 30.0 ns) | |
| local DRAM | ~60 ns | |
| remote DRAM | ~100 ns | |



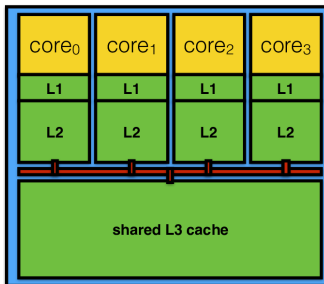
- Dato c il tempo di accesso alla cache (supponiamo vi sia una cache singola), m il tempo di accesso in memoria, e h l'hit ratio (frazione di volte in cui un dato richiesto viene trovato in cache), il tempo medio per accedere ad un dato sarà:

$$mean_access_time = c + (1 - h) * m$$

- Esempio: $c = 21\ ns$, $m = 80\ ns$, $h = 0.95$

$$mean_access_time = 21 + (1 - 0.95) * 80 = 25\ ns$$

- Numero di core (ciascuno contenente i propri registri, CU, ALUs, ...)
- Frequenza di clock, misurata in GHz \approx miliardi operazioni al secondo: utile per comparare processori appartenenti alla stessa famiglia
- Dimensione (in MB) e numero di cache: L1, L2, L3



La memoria principale

- La memoria principale (RAM, Random Access Memory) memorizza dati e istruzioni
- Dialoga direttamente con la CPU
- Capienza limitata (tipicamente 4–16 GB)
- Veloce (anche se 10–100 volte più lenta della CPU cache)
- Volatile: perde il suo contenuto allo spegnimento della macchina
- Può essere implementata con flip-flop (SRAM); oggi tipicamente si sfruttano altre tecnologie (DRAM, SDRAM)



- Come abbiamo visto, l'unità base di memoria è il **bit**
- La RAM consiste di un insieme di **celle**, ognuna delle quali memorizza un insieme di bit (tipicamente 8, = 1 byte)
- Ciascuna cella è caratterizzata da un numero, detto **indirizzo**, che la identifica
- Alle celle vengono assegnati indirizzi contigui, es., in una memoria con n celle, queste ultime avranno indirizzi da 0 a $n - 1$

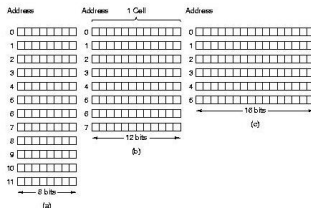


Figure 2-9. Three ways of organizing a 96-bit memory.



- A loro volta, i byte vengono raggruppati in **parole**
- Le parole sono tipicamente costituite da 32 o 64 bit (4 o 8 byte)
- Il calcolatore tipicamente opera su parole intere: i registri nella CPU conterranno 32 o 64 bit, e la ALU effettuerà operazioni su sequenze di 32 o 64 bit
- Dunque, le parole sono le unità fondamentali trattate all'interno della CPU, nonché le unità di informazioni trasferite da/verso la RAM

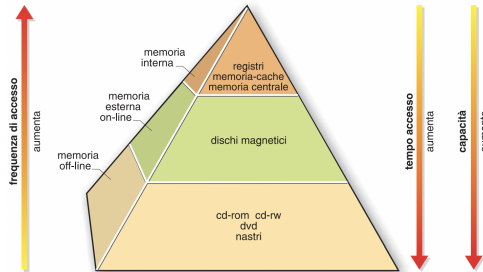
- Dimensioni: tipicamente 4 – 32 GB
- Velocità:

| RAM Type | Speed (MHz) | Peak Transfer Rate* |
|----------|-------------|---------------------|
| DDR2 | 533 | 4.27 GB/s |
| | 667 | 5.33 GB/s |
| | 800 | 6.4 GB/s |
| DDR3 | 1066 | 8.5 GB/s |
| | 1333 | 10.6 GB/s |
| | 1600 | 12.8 GB/s |
| | 1866 | 14.9 GB/s |
| DDR4 | 2133 | 17 GB/s |
| | 2400 | 19.2 GB/s |
| | 2666 | 21.3 GB/s |
| | 3200 | 25.6 GB/s |

- Non importa quanto è grande la memoria principale; non sarà mai sufficiente a contenere tutta l'informazione che un utente vuole memorizzare
- **Idea:** organizzare la memoria in una gerarchia
 - si va dai registri sulla CPU, alla CPU cache, alla memoria principale, memoria secondaria, memoria esterna
 - con il procedere nella gerarchia:
 - aumentano i tempi di accesso
 - aumentano le dimensioni
 - diminuisce il costo per byte

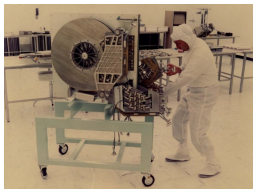
La memoria secondaria

La gerarchia della memoria



| Tipo di memoria | Tempo di accesso | Capacità |
|---------------------|------------------|-----------------|
| Registri di memoria | 1 - 3 ns | < 1KB |
| Memoria cache | 3 - 10 ns | 512 KB - 4 MB |
| Memoria centrale | 50 - 200 ns | 1 - 4 GB |
| Disco magnetico | 20 - 30 ms | 50 GB - 1 TB |
| Nastro | > 1 s | 4 GB - 300 GB |
| Dischi ottici | > 1 s | 650 MB - 4,7 GB |

- Un disco magnetico (o hard disk, storicamente per distinguerlo dal floppy disk) è costituito da uno o più piatti di alluminio ricoperti da materiale magnetizzabile
- Inizialmente, i piatti potevano essere 50 cm in diametro; oggi, tipicamente variano dai 3 ai 9 cm



- Il disco ruota e sulla sua superficie è posta una testina, che può muoversi verso l'interno o l'esterno del disco
- La testina fluttua ad un'altezza di circa 0.5 micron
- Le testine scrivono i dati sulla superficie del disco modificando la carica del materiale magnetizzabile, tramite l'applicazione di una corrente positiva o negativa
- Allo stesso modo, la carica del materiale magnetizzabile induce una corrente positiva o negativa nella testina; in questo modo si effettua la lettura



- La sequenza circolare di bit incontrata dalla testina durante una rotazione completa del disco è detta **traccia**
- Ciascuna traccia è costituita da un insieme di **settori**, tipicamente contenenti 512 byte di dati; inoltre, sono presenti gap e informazioni accessorie
- Ciò spiega perché la capacità di un disco formattato è circa il 15% inferiore rispetto al dato dichiarato dal produttore
- In un centimetro possono esserci anche 50.000 tracce

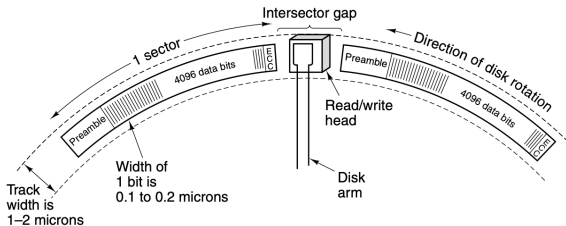


Figure 2-19. A portion of a disk track. Two sectors are illustrated.

- La maggior parte dei dischi è costituita da 1 a 12 piatti, impilati verticalmente
- Ciascuna superficie ha la propria testina
- I dischi ruotano solidalmente, e le testine si muovono in sincrono
- L'insieme delle tracce poste alla stessa distanza radiale è detto **cilindro**

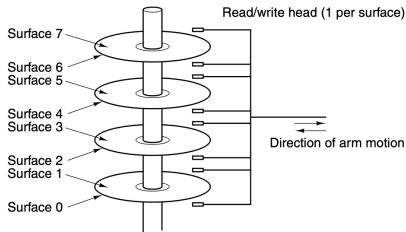


Figure 2-20. A disk with four platters.



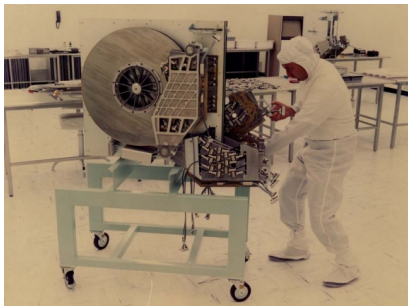
- Il dato fondamentale è la **velocità di rotazione**, tipicamente 5.400, 7.200, o 10.800 RPM
- Il tempo di accesso all'informazione su disco è dato da *tempo di seek + latenza di rotazione*
- I tempi medi di seek (posizionamento della testina) sono di 5–10 millisecondi
- Il tempo di rotazione è 6–12 millisecondi, dunque la latenza media di rotazione è 3–6 millisecondi
- È inoltre importante distinguere fra:
 - **maximum burst rate**: velocità di lettura una volta che la testina si trova sopra al primo bit da leggere della sequenza
 - **maximum sustained rate**: prende in considerazione anche i tempi di seek e di rotazione

- Gli hard disk (HD) rappresentano forse il caso più emblematico dell'evoluzione tecnologica a cui sono stati (e sono tutt'ora) soggetti i calcolatori
- Il primo HD è stato sviluppato da IBM nel 1956 (IBM Model 350 Disk). Conteneva 24 dischi, ciascuno di 60 centimetri di diametro. Capacità di memorizzazione: 5 MB



Breve storia degli hard disk

- Nonostante la densità di memorizzazione dei dati incrementasse rapidamente anno dopo anno, i dischi rigidi (principalmente per uso mainframe) continuarono ad essere costruiti seguendo la filosofia del “bigger is better” fino agli anni '80
- Nell'immagine, è possibile vedere un disco del 1979, con una capacità di 250 MB



- IBM introduce il primo disco in grado di superare 1 GB di capacità nel 1980 (IBM 3380, 2.52 GB)
- L'intero dispositivo ha le dimensioni di un congelatore, ed un peso di 250 chili



Breve storia degli hard disk

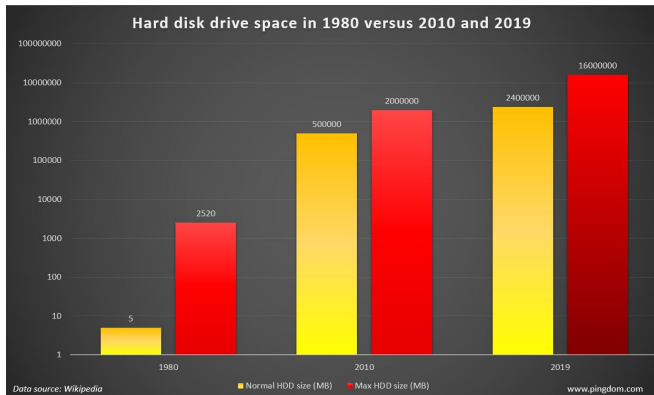
- I dischi di piccole dimensioni iniziano ad apparire negli anni '80, per equipaggiare i PC
- I primi drive installati su tali dispositivi avevano una capacità di 5 MB, e dischi del diametro di 13 cm



Figure: Riduzione delle dimensioni degli HD, dagli anni '80 ad oggi

Breve storia degli hard disk

- Ci sono voluti 51 anni per produrre il primo disco in grado di contenere 1 TB di dati, nel 2007
- Nel 2009, si ha il primo disco con una capacità di 2 TB
- Nel 2019, la capacità massima sale a 15 TB





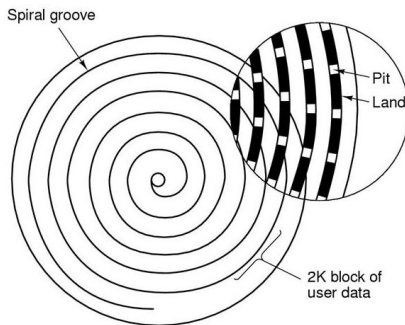
La memoria secondaria

Dischi allo stato solido

- I dischi allo stato solido (SSD, solid-state disk) sono costituiti da celle di memoria flash che memorizzano l'informazione tramite speciali transistor
- A differenza degli hard disk, non hanno parti mobili:
 - Maggiore resistenza agli urti e alle sollecitazioni
 - Maggiore velocità di accesso e trasferimento dei dati (> 15 volte risp. hard disk)
- Svantaggi:
 - Costo superiore
 - **Durata**: ciascuna cella può sopportare tipicamente 100.000 operazioni di scrittura. Tecnica di **wear leveling** per ottimizzare la durata delle celle distribuendo le scritture uniformemente sulle celle



- Sviluppati nel 1980 da Philips/Sony per rimpiazzare i vinili (red book), durata stimata 100 anni
- Standard per la memorizzazione di dati generici nel 1984 (yellow book)
- I dati sono memorizzati su una traccia a spirale, che inizia dal centro del disco e va verso l'esterno
- La traccia presenta avvallamenti (pit) e parti piane (land), e l'informazione è codificata come passaggio da pit/land o land/pit (1), o assenza di passaggio (0)



- Velocità di rotazione variabile da 530 (interno) a 200 RPM (esterno) per mantenere velocità angolare costante
- La traccia è lunga 5.6 km!
- Spazio a disposizione: inizialmente 650 MB, poi 700 MB

- DVD
 - Tracce più fini e più fitte
 - Laser a lunghezza d'onda più corta
 - 4.7 GB (single layer, single side)
 - Single/Double side, single/double layer (fino a 17 GB)
- Blu-ray
 - Laser blu a lunghezza d'onda ancora inferiore
 - Permette di mettere a fuoco con più precisione
 - Ulteriore riduzione della dimensione delle tracce, ed aumento della loro densità
 - 25 GB (single side), 50 GB (double side)