

Tipi di Memoria

RAM (Random Access Memory) Tutte le memorie utilizzate come memoria centrale sono RAM. Tuttavia abitualmente per RAM si intende tipicamente un tipo di RAM (RAM dinamica) volatile, cioè che perde il proprio contenuto quando non è alimentata.

Da questo punto di vista si distingue dalle

ROM (Read Only Memory) memorie dal contenuto fisso, non riscrivibili, 'scritte' una volta per tutte in fase di fabbricazione.

PROM (Programmable ROM) si possono scrivere una sola volta.

EPROM (Erasable PROM) si possono scrivere e cancellare (ma non selettivamente) per essere riutilizzate esponendole ai raggi UV. ROM, PROM, EPROM mantengono il loro contenuto anche a computer spento.

Uso delle ROM

- Quando il calcolatore viene acceso, la memoria centrale è "vuota", quindi non contiene istruzioni da eseguire. I dati da elaborare che possano essere letti dalla CPU.
- La CPU deve quindi essere in grado di leggere dati dalla memoria di massa (permanente) per *caricare* (fase di *caricamento* o di *boot*) almeno il *sistema operativo*, cioè il programma che controlla il funzionamento di tutto il calcolatore.

Uso delle ROM

- Anche solo per accedere alla memoria di massa è tuttavia necessario un programma.
- In ogni computer c'è una ROM che contiene il cosiddetto BIOS (Basic Input/Output System), un insieme di programmi di base per l'accesso ai dispositivi di I/O.
- La ROM non è volatile e quindi la CPU può leggerne i dati anche immediatamente dopo l'accensione.

Memoria Centrale

■ Caratteristiche fondamentali:

- **Dimensione** (numero di celle, quindi di byte)
- **Tempo di accesso** (il ritardo dopo l'invio del comando di lettura con cui il dato è effettivamente disponibile per l'elaborazione)
- Tipicamente la dimensione della memoria centrale è di alcune centinaia di Mbyte.
- Il tempo di accesso è dell'ordine delle decine di nanosecondi.
(1 ns = 10^{-9} s = un miliardesimo di secondo)

Memoria Centrale

- La capacità di una memoria si misura in byte:

K (kilo)	$2^{10} \cong 1000$
M (mega)	$2^{20} \cong 1.000.000$
G (giga)	$2^{30} \cong 1.000.000.000$
T (tera)	$2^{40} \cong 1.000.000.000.000$

Memoria Cache

- Sebbene la RAM sia veloce non lo è abbastanza per i processori moderni (tempo di accesso \gg periodo di clock)
- il processore perde tempo perciò per aspettare i dati dalla RAM
- Soluzione: inserire tra processore e RAM una memoria particolarmente veloce dove tenere i dati usati più spesso: **memoria cache**
- Si fa accesso alla RAM solo quando la cache non contiene il dato desiderato

Memoria Cache

- Se abbiamo memorie così veloci perché non le usiamo come memoria centrale?
- ⇒ Costano troppo

Memorie di massa

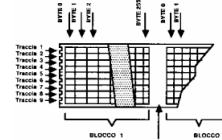
- Memorie di tipo *permanente*, cioè mantengono il loro contenuto anche dopo lo spegnimento del calcolatore.
- Memorie di tipo diverso rispetto alla memoria centrale, basate su fenomeni di tipo *ottico* o *magnetico* e non su fenomeni elettrici.
- Sono tipicamente organizzate in *dischi*, fissi o asportabili.
- Sono *più economiche* rispetto alle memorie RAM/ROM.
- Il loro *tempo di accesso* è tuttavia *maggiore*, cioè, a parità di dati da leggere, il processo di lettura avviene in modo più lento.

Memorie di massa

- Il primo tipo di memoria di massa sono state le *schede perforate*.
- Erano cartoncini che venivano forati. Ogni scheda codificava un certo numero di byte: un foro in una certa posizione codificava un 1 nel bit corrispondente o particolari codici.
- Il loro uso era particolarmente disagiata e poco affidabile a causa dell'ingombro e della facilità con cui potevano danneggiarsi.
- Il dispositivo di scrittura era analogo ad una macchina da scrivere, il lettore era di tipo opto-meccanico.

Memorie di massa

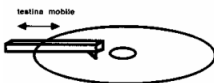
- Il primo tipo di memoria di massa magnetica sono stati i *nastri magnetici*



- L'informazione è divisa su 9 tracce, corrispondenti ciascuna ad un bit in una certa posizione (8 bit + 1 bit di *parità* per controllo degli errori).
- L'accesso ai dati è di tipo *sequenziale*.
- I dati sono divisi in *blocchi* separati da un *gap*.

Memorie di massa

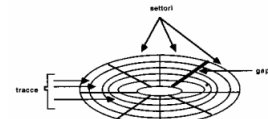
- Per ottenere una lettura random anziché sequenziale si possono utilizzare supporti magnetici di forma diversa, tipicamente dischi.



- Sommando il movimento di rotazione del disco e il movimento radiale della testina è possibile raggiungere rapidamente qualunque punto del disco. In pratica il tempo max per raggiungere una certa posizione è pari al maggiore fra il tempo richiesto per una rotazione e quello necessario alla testina per percorrere il raggio del disco.

Memorie di massa

- Invece che in blocchi rettangolari come i nastri, i dischi sono suddivisi in *settori* circolari.



- Poiché il disco ruota, lo stesso punto ripassa sotto la testina periodicamente. Quindi è sufficiente predisporre un unico gap come riferimento per ritrovare le altre posizioni sul disco.
- La *formattazione* del disco consiste nella creazione di questi riferimenti, che dipendono da sistema a sistema.

Dischi

- I dischi si dividono principalmente in 2 tipi:
 - **Dischi Magnetici**
 - floppy disk (portatili, estraibili, bassa capacità)
 - hard disk (generalmente fissi, pesanti, alta capacità)
 - altri formati proprietari (Zip, Jaz, ecc.)
 - **Dischi Ottici (tipicamente portatili)**
 - CD-ROM (capacità media, non riscrivibile)
 - CD-RW (come CD ma riscrivibile)
 - DVD-ROM (come CD ma capacità più elevata)
 - DVD-RW (come DVD-ROM ma riscrivibile)
 - altri formati proprietari

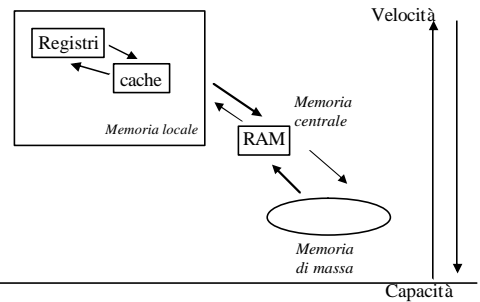
Dischi Ottici

- Sono i dischi che offrono la massima densità di memorizzazione (byte/cm²).
- Vengono scritti e letti con un laser. La superficie (normalmente riflettente) viene forata o alterata (in modo che non rifletta) in corrispondenza di un 1.
- Nei CD-ROM il processo di scrittura è irreversibile.
- Nei CD-RW è reversibile (è possibile riscrivere un certo numero di volte il disco).
- Nei DVD-ROM il disco è composto da più strati: il raggio laser può essere selettivamente focalizzato su uno di questi: è come avere più CD "impilati" uno sull'altro.

Parametri Caratteristici

- **Capacità** (Mb, Gb)
 - Floppy Disk 1.44 Mb (360 kB, 1.2 MB, 720 kB)
 - Hard Disk da 40 GB in su
 - CD-ROM, CD-RW 700 MB
 - DVD-ROM, DVD-RW circa 5 GB o circa 10 GB
- **Tempo di accesso** (ms)
 - Floppy Disk centinaia di millisecondi
 - Hard Disk alcuni millisecondi
 - CD/DVD decine di millisecondi
- **Transfer Rate** (kB/s)
quantità di dati che può essere trasferita da disco a memoria nell'unità di tempo.

Gerarchia delle Memorie



Le Unità Periferiche (dispositivi di I/O)

- Le unità periferiche permettono al calcolatore:
 - La comunicazione con il mondo esterno.
 - Il mantenimento di grandi quantità di informazioni non gestibili dalla memoria centrale (le memorie di massa sono dispositivi di I/O).

Le Unità Periferiche

- La comunicazione fra l'elaboratore ed una periferica avviene tramite un dispositivo chiamato **interfaccia**
- le interfacce sono molto diverse fra loro e dipendono dal tipo di unità periferica da connettere

Le Unità Periferiche

- Le unità periferiche si dividono in tre classi:
 - Dispositivi di Input
 - Dispositivi di Output
 - Dispositivi di Input/Output

Dispositivi di Input

- Le Unità di input permettono al calcolatore di ricevere informazioni dal mondo esterno.
- Sono basate su:
 - Trasduttori (di posizione, temperatura, pressione, luminosità,...).
 - Parte elettronica (per mantenere l'informazione, per codificarla, per segnalare lo stato dell'unità,...)
- Le principali periferiche di input sono:
 - Tastiera, mouse, scanner, microfono, telecamera, CD-ROM, DVD-ROM

Dispositivi di Output

- Le Unità di output permettono al calcolatore di inviare informazioni al mondo esterno.
- Sono basate su:
 - Parte elettronica (per mantenere l'informazione, per codificarla, per segnalare lo stato dell'unità,...)
 - Attuatori (trasformano il segnale elettrico digitale in: posizione, temperatura, pressione, luminosità,...)
- Le principali periferiche di output sono:
 - Monitor, stampante, casse acustiche

Dispositivi di Input/Output

- Le unità di input/output possono essere utilizzate sia come ingresso che come uscita.
- Le principali periferiche di input/output sono:
 - Hard-disk, floppy disk, CD-RW, DVD-RW, tape
- Le unità di input/output e le unità di input CD-ROM e DVD-ROM costituiscono le memorie di massa.

Terminali

- Dispositivi attraverso i quali avviene l'interazione uomo-macchina.
- Tipicamente composti da:
 - Tastiera
 - Monitor
 - Mouse
- Nei primi terminali (*telescriventi*) non c'era mouse ed il monitor era sostituito da una stampante.
- Il terminale principale di un sistema multiutente è di solito detto *console*.

Terminali

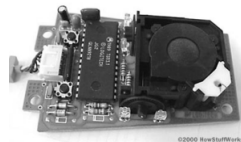
- La tastiera deriva da quella utilizzata nelle macchine da scrivere, con alcuni tasti aggiuntivi:
 - *Return (Invio)* conclude un comando o va a capo
 - *Insert (Ins)* cambia la modalità di inserimento del testo
 - *Delete (Del /Canc)* cancella il carattere sotto il cursore
 - *Backspace (<-)* cancella la lettera che lo precede
 - F1 - F12 (*Tasti funzione*) programmabili ed associati a funzioni diverse a seconda del programma.
 - *Escape (ESC)* termina l'attività corrente e torna alla modalità di base.
 - *Cursori (freccie)* per spostarsi all'interno di un comando o di un documento
 - *Control, Alternative (CTRL, ALT)* come shift aggiuntivi

Funzionamento Tastiera

- Quando si preme un tasto, la tastiera trasmette verso il calcolatore:
 - un codice diverso da 0 che corrisponde al codice del carattere che si è premuto (i codici da 1 a 127 corrispondono alla codifica ASCII dei caratteri)
 - Un codice uguale a 0 + un altro codice, quando si è premuto un tasto speciale (es. tasti funzione, cursori ecc.)
 - Per fare ciò, genera una *interruzione* (v. seguito)

Mouse

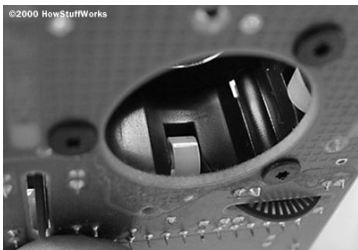
- Dispositivo opto-elettronico di *puntamento* utilizzabile con i sistemi operativi che prevedono una GUI (Graphic User Interface).



- Costituito, nel modello tradizionale, da una pallina che, scorrendo sul tavolo, fa muovere due dischi forati e ortogonali fra loro, il cui movimento è proporzionale allo spostamento orizzontale e verticale del mouse sul piano del tavolo.

Mouse

- La pallina (rivestita di gomma) aziona due rotelline ortogonali che si trovano sullo stesso asse di due dischi forati.



Mouse

- Ogni disco forato è "illuminato" da un LED a infrarossi da un lato; dall'altro lato è presente un sensore che rileva il fascio luminoso ad infrarossi.
- Ogni volta che un foro passa davanti al LED, il sensore riceve un impulso luminoso. Più veloce è il movimento, più frequenti saranno gli impulsi.
- Contando gli impulsi al secondo si ricava la velocità.

Monitor

- I monitor attuali sono tipicamente grafici. In essi ogni *pixel* (**picture element**), cioè ogni punto luminoso che contribuisce a formare l'immagine, può essere attivato singolarmente.
- Parametri fondamentali:
 - *Dimensione* (pollici): misura della diagonale dello schermo: il rapporto fra larghezza e altezza dello schermo è normalmente 4/3 (o 16/9).
 - *Risoluzione* (numero di pixel visualizzabili): tipicamente i monitor attuali hanno una risoluzione max variabile fra 1024x768 pixel a 1600x1200 pixel.
 - *Dot Pitch* (mm) è la minima distanza in mm che devono avere due pixel adiacenti sullo schermo per essere distinguibili.

Monitor

- Tecnologia di realizzazione:
 - CRT (Cathode Ray Tube), a tubo catodico, possono raggiungere dimensioni e risoluzione maggiore entro limiti di prezzo ragionevoli. Piuttosto ingombranti in profondità.
 - LCD (Liquid Crystal Display) a cristalli liquidi. Immagine più stabile e definita (matrice attiva), dimensioni limitate (sono piatti). Utilizzati nei portatili ma sempre più spesso anche nei PC fissi. A parità di dimensioni gli LCD costano circa il doppio dei CRT. Sono comunque ormai destinati a soppiantarli totalmente.

Scheda Video

- E' il dispositivo che 'pilota' il monitor.
- E' dotata di memoria RAM (*memoria video*, *VRAM*). Nella VRAM ogni pixel è codificato con 3 byte, corrispondenti al livello delle 3 componenti di base (Red, Green, Blue) che formano il colore del pixel. Per accendere un pixel quindi bisogna scrivere 3 valori nelle celle, di solito corrispondenti a 3 indirizzi consecutivi, corrispondenti a tale pixel.

Scheda Video

- Per una risoluzione max di 1280x1024 pixel la dimensione della VRAM deve quindi essere di almeno 3x1280x1024 byte. Di solito è maggiore, per contenere anche più immagini contemporaneamente.
- Ogni componente di colore assume valori da 0 a 255.
R=255 G=0 B=0 Rosso puro con la max intensità
R=G=B=0 nero R=G=B=255 bianco
R=G=B=[1,254] livelli di grigio intermedi

Stampanti

- Differiscono per il modo in cui l'inchiostro viene trasferito sulla carta.
- Si caratterizzano attraverso la *velocità di stampa* (caratteri o pagine al minuto) e la *risoluzione grafica* (punti per pollice, dpi (Dot Per Inch)).

Stampanti termiche

- Usate soprattutto nei fax e nei dispositivi portatili per la loro silenziosità e leggerezza. Usano carta chimica che si impressiona al contatto con una punta (o una matrice di punte) calda.
- Sono lente, hanno bassa risoluzione e la carta tende a cancellarsi se esposta al calore o alla luce.

Stampanti

Stampanti a impatto (ad aghi o a margherita)

- La struttura di trasporto della carta è simile a quella di una macchina da scrivere (un rullo di gomma, che rende possibile anche l'uso di moduli continui). La scrittura avviene per impatto di una testina su un nastro inchiostroato che va a contatto con la carta.
- Nelle stampanti *a margherita* l'impatto è causato da un percussore che spinge contro il foglio un carattere in rilievo che si trova sul bordo di un dispositivo circolare (margherita). Erano macchine da scrivere elettriche collegate al PC. Altissima risoluzione, ma stampano solo i caratteri presenti sulla margherita. Le stampanti laser hanno risoluzione paragonabile e possono essere usate anche in modo grafico.

Stampanti

Stampanti ad aghi

- Il carattere è generato da una testina composta da una colonna di aghi (di solito 9 o 24) che si sposta orizzontalmente sul foglio. Per ogni posizione della testina viene attivato un certo numero di aghi che compongono una colonna del carattere.
- Il movimento verticale sul foglio è dato dal rullo che trascina la carta.

Es.

○ ○ ● ○ ○
○ ● ● ○ ○
● ○ ● ○ ○
○ ○ ● ○ ○
○ ● ● ○ ○
○ ○ ○ ○ ○

Stampanti

Stampanti a getto di inchiostro

- Il carattere è generato da una testina con un numero molto elevato di ugelli, da cui inchiostro liquido contenuto in opportuni serbatoi viene spruzzato ad alta pressione sul foglio.
- Inchiostro e testina sono contenuti spesso in un'unica *cartuccia*. Se si hanno più serbatoi è possibile stampare a colori, di solito *in quadricromia*, utilizzando cioè 4 colori base per comporre tutti gli altri. Siccome il nero è il colore più utilizzato di solito ha un serbatoio separato e più grande rispetto agli altri colori.
- Costano molto poco, hanno elevata risoluzione (fino a 4800 dpi ed oltre), sono mediamente veloci ma hanno un elevato costo/pagina a causa del costo elevato dell'inchiostro.

Stampanti

Stampanti laser

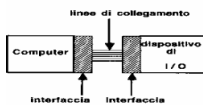
- Stesso principio delle fotocopiatrici: un rullo fotosensibile, colpito dalla luce, si carica elettrostaticamente ed attira microgranuli di inchiostro, che vengono poi trasferiti sulla carta che viene fatta scorrere su di esso.
- Nelle stampanti laser, come anche nelle fotocopiatrici più recenti, il fascio luminoso che eccita il rullo non è focalizzato su di esso attraverso un sistema di lenti ma è generato da un raggio laser.
- Hanno elevata risoluzione (fino 1200 dpi), sono veloci (fino a circa 20 pagine/minuto), più costose di quelle a getto di inchiostro ma con minore costo/pagina.
- A colori hanno costi ancora elevati per uso casalingo.

Plotter

- Utilizzati per disegno tecnico professionale, inizialmente per la maggiore risoluzione, in quanto disegnavano con una penna. Attualmente sono in pratica delle grosse stampanti a getto di inchiostro.
- Quelli a penna sono di due tipi: in quelli XY, un braccetto meccanico disegna con la penna su un foglio appoggiato su un piano. In quelli a rullo la carta viene spostata avanti e indietro dalla rotazione di un rullo su cui è appoggiata; la penna si muove in orizzontale sul foglio.
- I più recenti sono a rullo con una testina a getto di inchiostro al posto del portapenna.

Interfacce

- Schede elettroniche che permettono al calcolatore di comunicare secondo regole comuni (protocolli) con le periferiche, che possono invece essere progettate e costruite in modo molto differente le une dalle altre.
- Standard industriali definiscono protocolli di comunicazione fra PC e periferiche. Per adattare i segnali effettivamente usati all'interno di una periferica agli standard esistono le interfacce, circuiti che trasformano i segnali generati all'interno della periferica o del PC in segnali conformi allo standard.



Interfacce

- Le interfacce sono normalmente di due tipi:
 - seriale
 - parallelo
- Nelle interfacce seriali i dati sono trasmessi, un bit dopo l'altro, su una stessa connessione (cavo).
- Quindi, concettualmente, la seriale è il tipo di interfaccia più semplice possibile: un filo per la trasmissione, uno per la ricezione e uno di massa (GND). In realtà ci sono altre connessioni che servono per i segnali di controllo.
- Un tipo di seriale "evoluta" è l'interfaccia USB.

Interfacce

■ Interfaccia USB

- Standard definito da un consorzio (Intel Compaq Microsoft) con l'intento di sostituire sia le porte seriali che parallele
- Velocità 12MBit/sec
- Collega fino a 127 periferiche in cascata
- Completamente "Plug and Play"

Interfacce

- Nelle interfacce parallele i dati sono trasmessi contemporaneamente, su connessioni indipendenti per ogni bit. Su altre connessioni viaggiano anche i segnali di controllo.
- E' più complicata e costosa rispetto alla seriale, ma ha una maggiore velocità di trasferimento dati.



Interruzioni

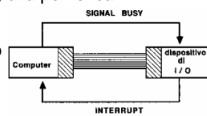
- I dispositivi di I/O:
 - elaborano dati ad una velocità normalmente molto inferiore rispetto alla capacità di elaborazione della CPU (es. stampanti).
 - Inviano dati verso la CPU in modo saltuario e 'imprevedibile' (es. tastiera); la CPU deve comunque essere pronta ad elaborare tali dati in modo tempestivo.
- E' quindi necessario realizzare un meccanismo per cui:
 - la CPU possa gestire, ad es., la stampa di un documento mentre compie altre attività.
 - La CPU non sia costretta a 'interrogare' continuamente le periferiche per sapere se sono pronte a ricevere o inviare nuovi dati.

Interruzioni

- Le periferiche vengono gestite attraverso un meccanismo detto *interruzione* (interrupt).
- Nel suo stato 'normale' di elaborazione la CPU ignora ciò che avviene nelle periferiche, cioè esegue soltanto i processi di elaborazione 'interni' (elaborazioni in memoria centrale, processi del sistema operativo ecc.).
- Nel momento in cui una periferica deve inviare o ricevere dati (ad es., se viene premuto un tasto sulla tastiera), viene attivata una linea che arriva alla CPU, che genera una interruzione.
- Quando si genera una interruzione, l'attività della CPU viene sospesa e il suo stato viene salvato in modo tale che l'elaborazione possa riprendere poi esattamente dal punto in cui era stata interrotta.

Interruzioni

- Durante l'interruzione la CPU esegue una *routine di servizio*, cioè un breve programma che esegue il trasferimento dati da/verso la periferica.
- Quindi quando i dati sono pronti per il trasferimento, durante l'esecuzione di un processo:
 - si genera una *interruzione*
 - lo stato della CPU viene salvato
 - il computer invia un segnale (*busy*) alla periferica
 - i dati vengono trasferiti
 - lo stato della CPU viene ripristinato
 - l'esecuzione del processo viene ripresa



Oltre la macchina tradizionale

- Problema: nella macchina di Von Neumann le operazioni sono strettamente sequenziali
- Nell'elaboratore tradizionale le operazioni sono strettamente sequenziali
- Altre soluzioni introducono forme di parallelismo
 - **processori dedicati** (coprocessori) al calcolo numerico, alla gestione della grafica, all'I/O
 - **esecuzione in parallelo** delle varie fasi di una istruzione: mentre se ne esegue una, si acquisiscono e decodificano le istruzioni successive (pipeline)
 - **architetture completamente diverse**: sistemi multi-processore, macchine dataflow, reti neurali