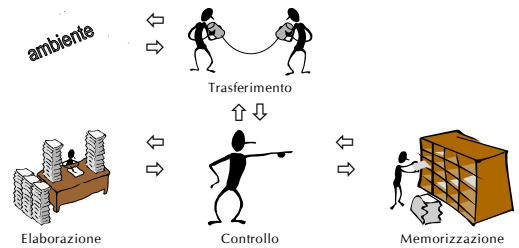


Capitolo 4 – Parte 1 Le infrastrutture hardware

Il processore
La memoria centrale
La memoria di massa
Le periferiche di I/O

Funzionalità di un calcolatore



04/04/02

Introduzione ai sistemi informatici

2

Caratteristiche dell'architettura

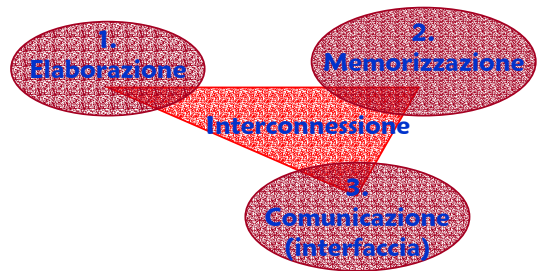
- **Flessibilità**
 - adatta a svolgere diverse tipologie di compiti
- **Modularità**
 - ogni componente ha una funzione specifica
- **Scalabilità**
 - ogni componente può essere sostituito con uno equivalente
- **Standardizzazione**
 - componenti facilmente sostituibili in caso di malfunzionamento
- **Riduzione dei costi**
 - grazie alla produzione su larga scala
- **Semplicità**
 - di installazione ed esercizio del sistema

04/04/02

Introduzione ai sistemi informatici

3

Il calcolatore: modello concettuale

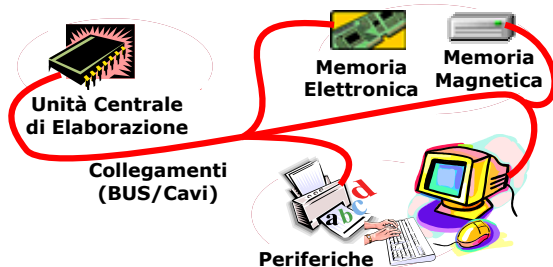


04/04/02

Introduzione ai sistemi informatici

4

Il calcolatore: modello architetturale

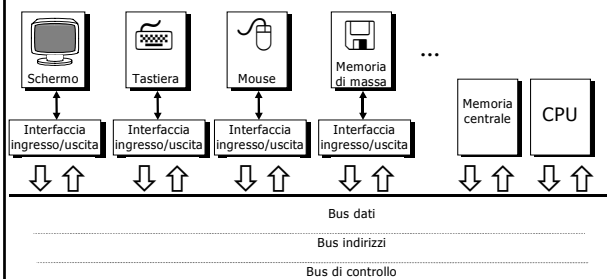


04/04/02

Introduzione ai sistemi informatici

5

Lo schema di riferimento



04/04/02

Introduzione ai sistemi informatici

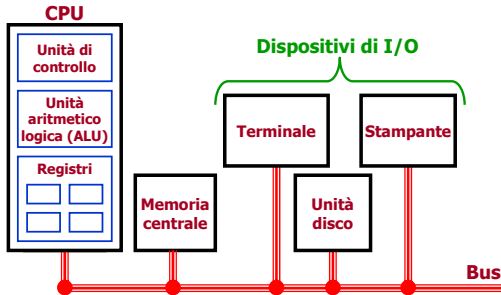
6

Caratteristiche del collegamento a BUS

- Semplicità
 - un'unica linea di connessione → costi ridotti di produzione
- Estendibilità
 - aggiunta di nuovi dispositivi molto semplice
- Standardizzabilità
 - regole per la comunicazione da parte di dispositivi diversi
- Lentezza
 - utilizzo in mutua esclusione del bus
- Limitata capacità
 - al crescere del numero di dispositivi collegati
- Sovraccarico del processore (CPU)
 - perchè funge da *master* sul controllo del bus

Unità centrale di elaborazione CPU

Organizzazione tipica di un calcolatore "bus oriented"



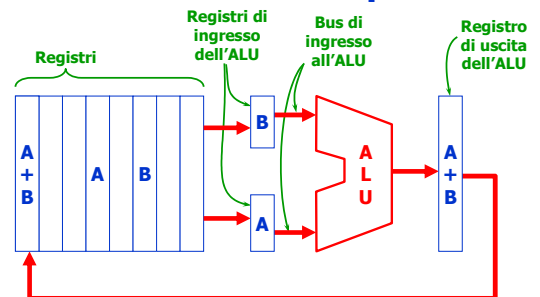
Tre tipologie di istruzioni

- Istruzioni aritmetico-logiche (Elaborazione dati)
 - Somma, Sottrazione, Divisione, ...
 - And, Or, Xor, ...
 - Maggiore, Minore, Uguale, Minore o uguale, ...
- Controllo del flusso delle istruzioni
 - Sequenza
 - Selezione semplice, a due vie, a n vie, ...
 - Ciclo a condizione iniziale, ciclo a condizione finale, ...
- Trasferimento di informazione
 - Trasferimento dati e istruzioni tra CPU e memoria
 - Trasferimento dati e istruzioni tra CPU e dispositivi di ingresso/uscita (attraverso le relative interfacce)

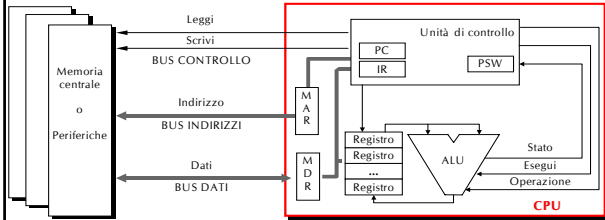
Elementi di una CPU

- **Unità di controllo**
 - legge le istruzioni dalla memoria e ne determina il tipo.
- **Unità aritmetico-logica**
 - esegue le operazioni necessarie per eseguire le istruzioni.
- **Registri**
 - **memoria ad alta velocità** usata per risultati temporanei e informazioni di controllo;
 - il **valore massimo** memorizzabile in un registro è determinato dalle **dimensioni** del registro;
 - esistono registri di uso generico e registri specifici:
 - **Program Counter (PC)** – qual è l'istruzione successiva;
 - **Instruction Register (IR)** – istruzione in corso d'esecuzione;
 - ...

Struttura del "data path"



La struttura della CPU



04/04/02

Introduzione ai sistemi informatici

13

Esecuzione delle istruzioni

➤ Ciclo **Fetch-Decode-Execute** (**leggi-decodifica-esegui**)

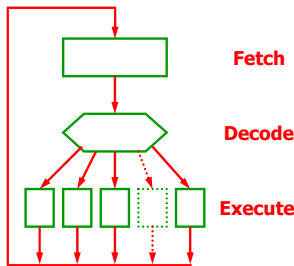
1. Prendi l'**istruzione corrente** dalla memoria e mettila nel **registro istruzioni (IR)**.
2. **Incrementa** il **program counter (PC)** in modo che contenga l'indirizzo dell'istruzione successiva.
3. Determina il tipo dell'istruzione corrente (**decodifica**).
4. Se l'istruzione usa una parola in memoria, determina dove si trova.
5. Carica la parola, se necessario, in un registro della CPU.
6. **Esegui** l'istruzione.
7. Torna al punto 1 e inizia a eseguire l'istruzione successiva.

04/04/02

Introduzione ai sistemi informatici

14

Ciclo **Fetch-Decode-Execute**



04/04/02

Introduzione ai sistemi informatici

15

CPU

- In grado di eseguire solo istruzioni codificate in **linguaggio macchina**
- Ciclo Fetch - Decode - Execute
 1. Prendi l'istruzione corrente dalla memoria e mettila nel registro istruzioni (IR) (**fetch**)
 2. Incrementa il Program Counter (PC) in modo che contenga l'indirizzo dell'istruzione successiva
 3. Determina il tipo di istruzione da eseguire (**decode**)
 4. Se l'istruzione necessita di un dato in memoria determina dove si trova e caricalo in un registro della CPU
 5. Esegui l'istruzione (**execute**)
 6. Torna al punto 1 e opera sull'istruzione successiva



04/04/02

Introduzione ai sistemi informatici

16

Evoluzione delle CPU



CPU	Anno	Frequenza (MHz)	Dimensione registri / bus dati	Numero di transistor
8086	1978	4.77 — 12	8 / 16	29 000
80286	1982	8 — 16	16 / 16	134 000
80386	1986	16 — 33	32 / 32	275 000
80386 SX	1988	16 — 33	32 / 16	275 000
80486	1989	33 — 50	32 / 32	1 200 000
Pentium	1993	60 — 200	32 / 64	3 100 000
Pentium II	1997	233 — 400	32 / 64	7 500 000
Pentium III	1999	450 — 1133	32 / 64	24 000 000
Pentium 4	2000	1600 — 2000	32 / 64	42 000 000

04/04/02

Introduzione ai sistemi informatici

17

Legge di Moore

Osservazione fatta da Gordon Moore nel 1965:

**il numero dei transistor per cm²
raddoppia ogni X mesi**

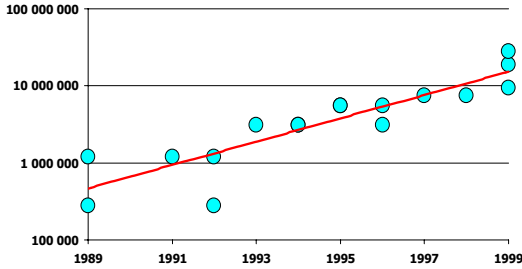
In origine X era 12. Correzioni successive hanno portato a fissare **X=18**. Questo vuol dire che c'è un incremento di circa **il 60% all'anno**.

04/04/02

Introduzione ai sistemi informatici

18

Transistor [CPU Intel]



Legge di Moore e progresso

- Il progresso della tecnologia provoca un **aumento del numero di transistor** per cm² e quindi per chip.
- Un maggior numero di transistor per chip permette di produrre **prodotti migliori** (sia in termini di prestazioni che di funzionalità) **a prezzi ridotti**.
- I prezzi bassi stimolano la nascita di **nuove applicazioni** (e.g. non si fanno video game per computer da milioni di \$).
- Nuove applicazioni aprono **nuovi mercati** e fanno nascere **nuove aziende**.
- L'esistenza di tante aziende fa **creștere la competitività** che, a sua volta, stimola il **progresso della tecnologia** e lo sviluppo di **nuove tecnologie**.

Approfondimento: incrementare le prestazioni con il parallelismo

Parallelismo

- La **frequenza di clock**
 - influenza direttamente il tempo di ciclo del data path e quindi le prestazioni di un calcolatore;
 - è limitata dalla tecnologia disponibile.
- Il **parallelismo** permette di migliorare le prestazioni senza modificare la frequenza di clock. Esistono due forme di parallelismo:
 - **parallelismo a livello delle istruzioni** (architetture **pipeline** o architetture **superscalari**);
 - **parallelismo a livello di processori** (**Array computer**, **multiprocessori** o **multicomputer**).

Architettura pipeline

- Organizzazione della CPU come una "**catena di montaggio**"
 - la CPU viene suddivisa in "**stadi**", ognuno dedicato all'esecuzione di un compito specifico;
 - l'esecuzione di un'istruzione richiede il **passaggio attraverso** (tutti o quasi tutti) **gli stadi della pipeline**;
 - in un determinato istante, **ogni stadio esegue la parte di sua competenza di una istruzione**;
 - in un determinato istante, esistono **diverse istruzioni contemporaneamente in esecuzione**, una per ogni stadio.

Esempio di pipeline

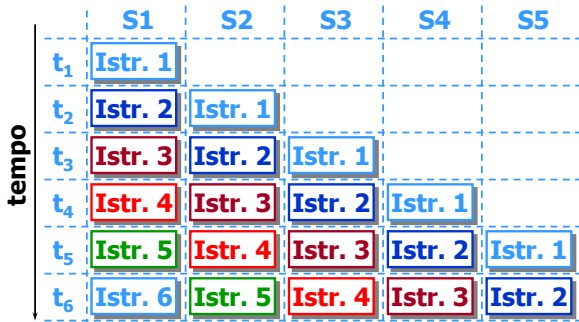
/1

Pipeline in **cinque stadi**:

- S1. **lettura istruzioni** dalla memoria e loro caricamento in un apposito buffer;
- S2. **decodifica** dell'istruzione per determinarne il tipo e gli operandi richiesti;
- S3. individuare e **recuperare gli operandi** dai registri o dalla memoria;
- S4. **esecuzione** dell'istruzione, tipicamente facendo passare gli operandi per il data path;
- S5. **invio dei risultati** al registro appropriato.



Esempio di pipeline /2

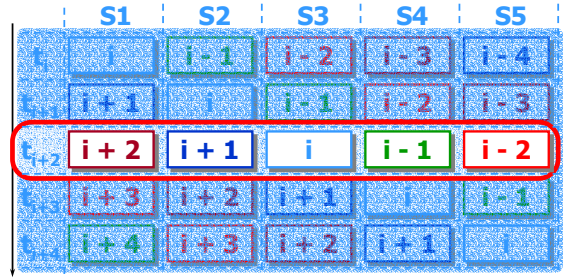


04/04/02

Introduzione ai sistemi informatici

25

Esempio di pipeline /3



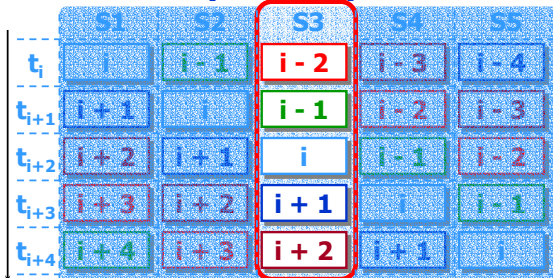
All'istante t_{i+2} ci sono 5 istruzioni in esecuzione

04/04/02

Introduzione ai sistemi informatici

26

Esempio di pipeline /4



Lo stadio S3 esegue la parte di sua competenza di istruzioni successive l'una all'altra.

04/04/02

Introduzione ai sistemi informatici

27

Prestazioni di una pipeline

- Il tempo di esecuzione (**latenza**) della singola istruzione non diminuisce, anzi **aumenta**
 - il tempo di attraversamento (latenza) della pipeline corrisponde al numero degli stadi (**N**) moltiplicato per il tempo di ciclo (**T**);
 - il tempo di ciclo è limitato dallo stadio più lento!
- **Aumenta** il numero di istruzioni completate nell'unità di tempo (**throughput**)
 - si completa **un'istruzione a ogni ciclo di clock**;
 - l'**incremento** di throughput è quasi **proporzionale al numero degli stadi**!

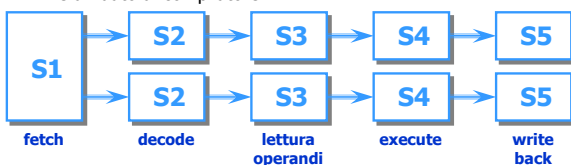
04/04/02

Introduzione ai sistemi informatici

28

Architetture superscalari

- Vista la disponibilità di un maggior numero di transistor si inseriscono **più pipeline** nella stessa CPU
 - aumenta il parallelismo perché è possibile eseguire contemporaneamente diversi **flussi di istruzioni**;
 - è necessario garantire che non ci siano **conflitti** tra le istruzioni che vengono eseguite insieme; di solito il controllo è affidato al compilatore.



04/04/02

Introduzione ai sistemi informatici

29

Architetture superscalari

- In alternativa si possono **replicare le unità funzionali**
 - rappresentano lo stadio più lento della pipeline (in genere richiedono diversi cicli di clock);
 - è più semplice evitare i **conflitti** tra le diverse istruzioni.



04/04/02

Introduzione ai sistemi informatici

30

Multiprocessori

> Diverse CPU condividono una **memoria comune**:

- le CPU debbono **coordinarsi** per accedere alla memoria;
- esistono diversi schemi di collegamento tra CPU e memoria, quello più semplice prevede che ci sia un **bus condiviso**;
 - se i processori sono veloci il **bus** diventa un **collo di bottiglia**;
 - esistono soluzioni che permettono di migliorarne le prestazioni, ma si adattano a sistemi con un **numero limitato di CPU** (<20).

> La **memoria condivisa** rende più semplice il **modello di programmazione**:

- si deve **parallelizzare l'algoritmo**, ma si può trascurare la **"parallelizzazione" dei dati**.



04/04/02

Introduzione ai sistemi informatici

31

Multicalcolatori

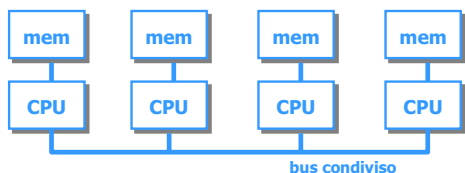
- > Sistemi composti da tanti calcolatori collegati fra loro
 - ogni calcolatore è dotato di una **memoria privata** e non c'è **memoria in comune**;
 - **comunicazione** tra CPU basata su **scambio di messaggi**.
- > Non è efficiente collegare ogni calcolatore a tutti gli altri, quindi vengono usate topologie particolari:
 - **griglie** a 2/3 dimensioni, **alberi** e **anelli**;
 - i messaggi, per andare da fonte a destinazione, spesso devono passare da uno o più calcolatori intermedi o **switch**.
 - **Tempi di trasferimento** dei messaggi dell'ordine di alcuni **microsecondi** sono comunque facilmente ottenibili.
- > Sono stati costruiti multicalcolatori con ~10.000 CPU.

04/04/02

Introduzione ai sistemi informatici

32

Struttura di un multicomputer



04/04/02

Introduzione ai sistemi informatici

33

La memoria

La memoria

- > **Supporto alla CPU**: deve fornire alla CPU dati e istruzioni il più rapidamente possibile
- > **Archivio**: deve consentire di archiviare dati e programmi garantendone la conservazione e la reperibilità anche dopo elevati periodi di tempo
- > Diverse esigenze:
 - **velocità** per il supporto alla CPU
 - **non volatilità** ed **elevate dimensioni** per l'archivio
- > Diverse tecnologie
 - **elettronica**: veloce, ma costosa e volatile
 - **magnetica** e **ottica**: non volatile ed economica, ma molto lenta

04/04/02

Introduzione ai sistemi informatici

35

Criteri di caratterizzazione di una memoria

- > **Velocità**
 - tempo di accesso (quanto passa tra una richiesta e la relativa risposta)
 - velocità di trasferimento (quanti byte al secondo si possono trasferire)
- > **Volatilità**
 - cosa succede quando la memoria non è alimentata?
 - per quanto tempo i dati vi rimangono immagazzinati?
- > **Capacità**
 - quanti byte può contenere? qual è la dimensione massima?
- > **Costo (per bit)**
- > **Modalità di accesso**
 - diretta (o casuale): il tempo di accesso è indipendente dalla posizione
 - sequenziale: il tempo di accesso dipende dalla posizione
 - mista: combinazione dei due casi precedenti
 - associativa: indicato il dato, la memoria risponde indicando l'eventuale posizione che il dato occupa in memoria.

04/04/02

Introduzione ai sistemi informatici

36

La memoria centrale

La memoria centrale (R.A.M.)

- Mantiene al proprio interno i **dati** e le **istruzioni** dei programmi in esecuzione
- Memoria ad accesso "casuale"
- Tecnologia elettronica
 - **veloce** ma **volatile** e **costosa**
- Due "eccezioni"
 - **R.O.M.**: elettronica ma permanente e di sola lettura
 - **Flash**: elettronica ma permanente e riscrivibile

04/04/02

Introduzione ai sistemi informatici

38

Indirizzi di memoria

- I bit nelle memorie sono raggruppati in **celle**:
 - tutte le celle sono formate dallo **stesso numero di bit**;
 - una cella composta da **k bit**, è in grado di contenere una qualunque tra **2^k combinazioni** diverse di bit.
- Ogni cella ha un **indirizzo**:
 - serve come accesso all'informazione;
 - in una memoria con **N celle** gli indirizzi vanno da **0** a **N-1**.
- **La cella è l'unità indirizzabile più piccola.** In quasi tutti i calcolatori è di **8 bit** (un **byte**).
- I byte vengono raggruppati in **parole** (che oggi sono di **32/64 bit**), su cui la CPU esegue le operazioni.

04/04/02

Introduzione ai sistemi informatici

39

Organizzazione della memoria

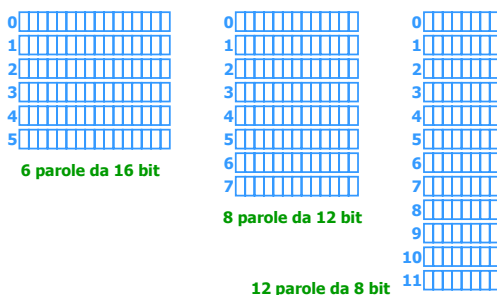
- Anche gli indirizzi della memoria sono rappresentati come numeri binari:
 - un indirizzo di **M bit** consente di indirizzare **2^M** celle;
 - per 6 o 8 celle bastano 3 bit, per 12 celle ne servono 4;
 - il **numero di bit nell'indirizzo** determina il **numero massimo di celle indirizzabili** nella memoria ed è indipendente dal numero di bit per cella (una memoria con 2^{12} celle richiede sempre 12 bit di indirizzo, quale che sia la dimensione di una cella).
- Una memoria può essere organizzata in diversi modi:
 - con 96 bit possiamo avere 6 celle di 16 bit ($6 \cdot 16 = 96$), o 8 celle di 12 bit ($8 \cdot 12 = 96$) o 12 celle di 8 bit ($12 \cdot 8 = 96$).

04/04/02

Introduzione ai sistemi informatici

40

Organizzazione della memoria



04/04/02

Introduzione ai sistemi informatici

41

Memoria vs. CPU

- Le **CPU** sono sempre state **più veloci** delle **memorie**
 - l'aumento di integrazione ha consentito di realizzare **CPU** pipeline e super scalari, molto efficienti e **veloci**;
 - nelle memorie è aumentata la **capacità** più che la **velocità**.
- L'**accesso** alla memoria passa **attraverso il bus**
 - la **frequenza** di funzionamento **del bus** è molto **più bassa** di quella della CPU;
 - il **bus** può essere **impegnato** ad effettuare trasferimenti controllati **da dispositivi** di I/O "**autonomi**" (e.g. DMA).
- **È difficile riordinare le istruzioni** in modo da poter sfruttare i tempi di attesa della memoria.
- È possibile fare **memorie molto veloci** se stanno nel chip della CPU, ma **sono piccole** e **costose**.

04/04/02

Introduzione ai sistemi informatici

42

Le memorie elettroniche

- Memorie di gran capacità, relativamente lente, economiche ed accessibili tramite il bus:
 - MGL ovvero **Memoria Grossa e Lenta**;
 - **dimensioni** pari a circa **10 unità**;
 - **tempo di accesso** (TA) di circa **10 unità**.
- Memorie veloci, integrate nello stesso chip della CPU, ma costose:
 - MPV ovvero **Memoria Piccola e Veloce**;
 - **dimensioni** pari a circa **1 unità**;
 - **tempo di accesso** pari a circa **1 unità**.
- Obiettivo: realizzare una memoria **grossa e veloce**
 - **dimensioni** pari a circa quelle della memoria **grossa**;
 - **prestazioni** pari a circa quelle della memoria **veloce**.

04/04/02

Introduzione ai sistemi informatici

43

Una gerarchia di memoria

- Memoria formata da **una MPV** e **una MGL**:
 - la MPV contiene una **copia** di **alcune celle** della MGL;
 - quando la CPU chiede una particolare cella di memoria la richiesta va ad **entrambe le memorie**:
 - se il dato si trova nella MPV, viene passato direttamente alla CPU;
 - se il dato si trova nella MGL, viene anche **caricato nella MPV**.
- Ipotesi: **distribuzione uniforme** delle richieste
 - la frequenza con cui si trova il dato cercato nella MPV (**hit ratio**) sarà in media il 10% (1/10), in questi casi il tempo di accesso (**hit time**) sarà pari a 1 unità;
 - la frequenza con cui è necessario accedere alla MGL (**miss ratio**) sarà in media il 90% (9/10), in questi casi il tempo di accesso (**miss penalty**) sarà pari a 10 unità;
 - il tempo medio di accesso sarà $0.1 \cdot 1 + 0.9 \cdot 10 = 9.1$ unità!

04/04/02

Introduzione ai sistemi informatici

44

Il principio di località

- **Località spaziale**: quando si accede all'indirizzo A, è molto probabile che gli accessi successivi richiedano **celle vicine ad A**.
 - le istruzioni del codice vengono in genere lette da locazioni consecutive della memoria;
 - gli accessi ad array o a strutture dati sono "vicini".
- **Località temporale**: quando si accede all'indirizzo A, è molto probabile negli accessi successivi si richieda **di nuovo** la cella **A**.
 - cicli di istruzioni accedono ripetutamente alle stesse locazioni di memoria;
 - istruzioni vicine tendono ad utilizzare le stesse variabili.

04/04/02

Introduzione ai sistemi informatici

45

Come si sfrutta la località

- Diversi approcci a seconda del tipo di località:
 - località **temporale**: i dati prelevati dalla MGL vengono conservati nella MPV **il più a lungo possibile**;
 - località **spaziale**: quando si copia un dato dalla MGL alla MPV, si copiano anche i dati vicini (**cache line** o **blocco**).
- La frequenza di successo (**hit ratio – h**) cresce fino a **superare il 99%**:
 - in effetti **h** dipende da due caratteristiche contrastanti:
 - **la dimensione dei blocchi**
un blocco grande sfrutta meglio la località spaziale;
 - **quanti sono i blocchi in memoria**
se c'è spazio per tanti blocchi un dato resta in memoria più a lungo e può sfruttare più a lungo la località temporale;
 - c'è anche il problema del **costo** della cache!

04/04/02

Introduzione ai sistemi informatici

46

L'effetto della località

Effetto del principio di località sull'esempio di prima:

- tempo di accesso alla cache pari a 1 unità (**TA_C = 1**);
- tempo di accesso alla memoria (detto anche miss penalty, ovvero penalità di fallimento) pari a 10 unità (**TAM = 10**);
- frequenza di successo (hit ratio, **h = 0.99**);
- frequenza di fallimento (miss ratio, **m = 1 – h = 0.01**);
- tempo di accesso medio pari a:

$$TA = h \cdot TA_C + m \cdot TA_M$$
$$TA = 0.99 \cdot 1 + 0.01 \cdot 10 = 1.09$$

04/04/02

Introduzione ai sistemi informatici

47

La memoria centrale

- Tecnologia elettronica (**veloce** ma **volatile**)
- Gerarchia di memoria:
ai **livelli più alti** corrispondono le **tecnologie più veloci** ma anche **più costose**
 - cache interna (Static RAM – SRAM)
 - cache esterna (SRAM)
 - memoria RAM (Dynamic RAM – DRAM e sue varianti)
 - area di swap su memoria di massa

04/04/02

Introduzione ai sistemi informatici

48

Memoria cache: SRAM

- Interna (L1) ➡ **stessa frequenza della CPU**
- Esterna (L2 e/o L3)
 - Tre diverse posizioni/configurazioni
 - **Saldato** sulla motherboard
 - Card Edge Low Profile (**CELP**) socket
 - **COAST** (Cache On A Stick) module
 - Diverse tipologie
 - Asynchronous SRAM (più economica), TA compreso tra 12 e 20ns, OK per bus tra 50 e 66 MHz, timing = 3-2-2-2
 - Synchronous Burst SRAM (**Synch SRAM**)
 - Bus fino a 66 MHz ➡ timing = 2-1-1-1
 - Bus oltre i 66 MHz ➡ timing = 3-2-2-2
 - Synchronous Pipelined Burst SRAM (**PB SRAM**)
 - TA compreso tra 4.5 e 8ns, OK per bus fino a 133 MHz, timing = 3-1-1-1

04/04/02

Introduzione ai sistemi informatici

49

Memoria centrale – DRAM /1

- Fast Page Mode DRAM (**FPM DRAM**)
 - TA=70-60ns ➡ timing = 5-3-3-3.
 - Per la lettura si attiva la riga, la colonna, si validano i dati, si trasferiscono i dati, poi si disattiva la colonna
 - I miglioramenti di velocità nascono dal progresso della tecnologia di integrazione.
- Extended Data Out DRAM (**EDO DRAM**)
 - TA = 70-50ns ➡ timing = 5-2-2-2
 - Non richiede la disattivazione della colonna e del buffer di uscita; 60ns è il minimo per bus a 66MHz
- Burst EDO DRAM (**BEDO DRAM**)
 - Evoluzione di EDO DRAM ➡ timing = 5-1-1-1 (pipeline + 2-bit burst counter)
 - Mai davvero supportata.

04/04/02

Introduzione ai sistemi informatici

50

Memoria centrale – DRAM /2

- Synchronous DRAM (**SDRAM**)
 - Sfrutta la sequenzialità delle richieste: una volta trovato il primo dato gli altri vengono recuperati velocemente.
 - Fornisce dati fino a 10ns (100MHz) con timing 6-1-1-1
- **PC133 SDRAM**
 - Evoluzione della SDRAM per bus a 133MHz
 - Trasferimento dati fino a 1.6GBps
- Double Data Rate DRAM (**DDR DRAM**)
 - Sfrutta entrambi i fronti del clock per trasferire dati: raddoppia la frequenza efficace non quella effettiva.
- Direct Rambus DRAM (**DRDRAM**)
 - Risultato della collaborazione tra Intel e Rambus
 - Nuova architettura: 600-800MHz (1000MHz nel 2001) con bus di sistema a 133MHz.
 - 1 canale arriva fino a 1.6GBps (4 canali 6.4 GBps)

04/04/02

Introduzione ai sistemi informatici

51

Packaging

- Fino all'inizio degli anni '90 la memoria veniva prodotta, acquistata e installata su chip singoli
 - densità variabili da 1 Kbit a 1 Mbit;
 - i PC avevano zoccoli vuoti dove inserire altri chip.
- Oggi si monta un gruppo di chip, tipicamente 8 o 16, su un piccola scheda stampata che si vende come unità minima installabile nei PC
 - **SIMM** (**Single Inline Memory Module**) se la fila di connettori si trova da un solo lato della scheda;
 - **DIMM** (**Dual Inline Memory Module**) se i connettori si trovano su ambedue i lati della scheda
- Sia **SIMM** che **DIMM** sono a volte dotate di un codice di rilevazione o di correzione dell'errore.

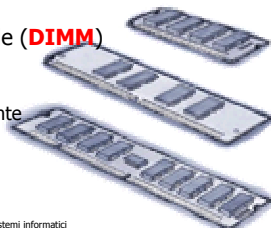
04/04/02

Introduzione ai sistemi informatici

52

Memoria centrale – Chip

- Single Inline Memory Module (**SIMM**)
 - 30/72 pin sullo stesso lato della scheda;
 - trasferimento dati a 8/32 bit per volta;
 - utilizzabili "a coppie".
- Dual In-line Memory Module (**DIMM**)
 - 168 pin su due lati;
 - 64 bit alla volta;
 - utilizzabili anche singolarmente
- **RIMM**
 - Moduli di RDRAM;
 - interfaccia DIMM 100MHz.



04/04/02

Introduzione ai sistemi informatici

53

La memoria di massa (magnetica)



Una gerarchia di memoria

Ottenuta per "generalizzazione" dell'applicazione del principio di **località** e tipicamente costituita da

1. **registri** contenuti nella CPU (qualche KB)
2. **cache** (da circa 32KB a circa 1024KB)
3. **memoria principale** (da circa 64MB a qualche GB)
4. **dischi fissi** (da qualche GB a qualche TB)
5. **nastri magnetici e dischi ottici** (da qualche GB a qualche TB per ogni supporto)

Man mano che ci si sposta verso il basso nella gerarchia aumenta il valore dei parametri fondamentali:

- **aumenta il tempo di accesso;**
- **aumenta la capacità** di memorizzazione;
- ma **diminuisce il costo per bit.**

Caratteristiche dei diversi livelli

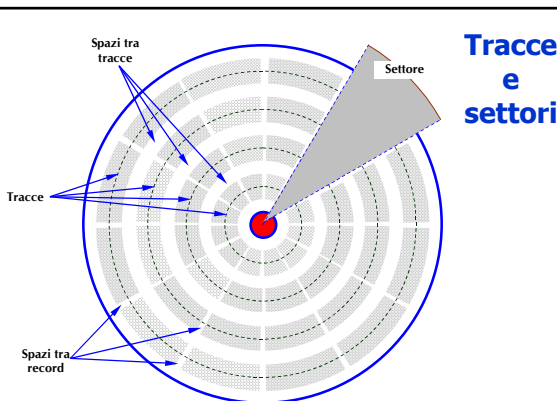
	Capacità	Velocità (TA)	€/MByte
registri	~1KB	~1ns	NA
cache	64 ÷ 1024 KB	~10ns	300
RAM	64 ÷ 2048 MB	~100ns	2
HD	8 ÷ 100 GB	~10ms	0.005
nastri/CD	~GB per unità	~100ms	0.005

Dischi magnetici

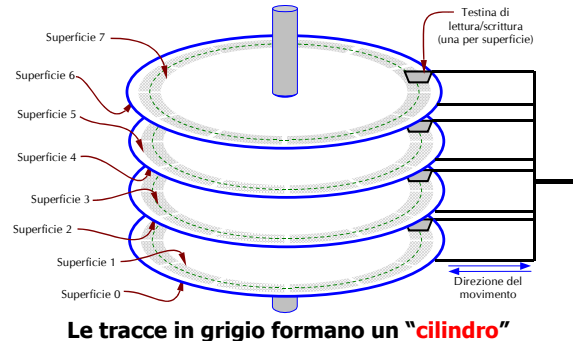
- Sono **piatti** d'alluminio (o di altro materiale) ricoperti di **materiale ferromagnetico**.
- **Fattore di forma** (diametro)
 - sempre più piccolo (consente velocità di rotazione maggiori);
 - 3.5 pollici per i sistemi desktop e fino a 1 pollice per i mobili.
- **Testina** di un disco (strumento di lettura/scrittura)
 - è sospesa appena sopra la superficie magnetica
 - **scrittura**: il passaggio di corrente positiva o negativa attraverso la testina magnetizza la superficie
 - **lettura**: il passaggio sopra un'area magnetizzata induce una corrente positiva o negativa nella testina.

Tracce e settori

- **Traccia (track)**: sequenza circolare di bit scritta mentre il disco compie una rotazione completa
 - la larghezza di una traccia dipende dalla dimensione della testina e dall'accuratezza con cui la si può posizionare; la densità radiale va da 800 a 2000 tracce per centimetro (5-10 µm per traccia);
 - tra una traccia e l'altra c'è un piccolo spazio di separazione (**gap**).
- **Settore (sector)**: parte di una traccia corrispondente a un settore circolare del disco
 - un settore contiene 512 byte di dati, preceduti da un preambolo, e seguiti da un codice di correzione degli errori;
 - la densità lineare è di circa 50-100kbit per cm (0.1-0.2 µm per bit);
 - tra settori consecutivi si trova un piccolo spazio (**intersector gap**).
- **Formattazione**: operazione che predispone tracce e settori per la lettura/scrittura
 - un 15% circa dello spazio disco si perde in gap, preamboli e codici di correzione degli errori.



Schema di un Hard Disk



Prestazioni dei dischi

- **Tempo di accesso** (ms o 10^{-3} s)
 - **Seek time**
 - la testina deve arrivare alla traccia giusta;
 - dipende dalla meccanica (5-15 ms, 1 per tracce adiacenti).
 - **Latency**
 - il disco deve ruotare fino a portare il dato nella posizione giusta;
 - dipende dalla velocità di rotazione (5400-10800 RPM \Rightarrow 2.7-5.4ms).
- **Transfer Rate** (MBps)
 - **Velocità di trasferimento del disco**
 - dipende dalla densità di registrazione e dalla velocità di rotazione;
 - un settore di 512 byte richiede fra 25 e 100 μ sec (5-20 MB/sec).
 - **Velocità di trasferimento del sistema di controllo**
 - SCSI vs. EIDE

Velocità burst vs. sustained

- Velocità **burst**
 - velocità di trasferimento dei dati una volta che la testina ha raggiunto il primo bit di dati;
 - velocità massima mantenuta per un **tempo limitato**.
- Velocità **sustained**
 - velocità media sostenibile per un certo numero di secondi;
 - velocità mantenibile per un **tempo illimitato**.
- La differenza è provocata dagli spazi di "**servizio**":
 - preamboli, ECC, spazi di intersezione, tempi di ricerca, ...
 - la rotazione dei dischi (60-120 giri/sec) ne provoca il riscaldamento e l'espansione: questi dischi debbono essere **ricalibrati** periodicamente.

Velocità lineare vs. angolare

- **Velocità angolare costante**
 - le tracce esterne sono **più lunghe** di quelle interne;
 - la velocità lineare è maggiore quando si leggono le tracce più esterne.
- Diverse soluzioni:
 - **densità lineare massima** sulla traccia più interna e densità dei bit decrescente sulle tracce più esterne: in un disco con 18 settori per ogni traccia, ogni settore occupa 20 gradi di arco, indipendentemente dal cilindro.
 - **cilindri divisi in zone** (tipicamente da 10 a 30 per ogni unità) e numero di settori per traccia aumentato in ogni zona man mano che si procede verso l'esterno. In questo modo aumenta la capacità dell'unità.

Floppy disk

- Funzioni:
 - **distribuzione** software su grande scala (avvento PC);
 - archiviazione dati.
- Struttura analoga a quella di un disco magnetico,
 - il disco si **ferma** quando non è operativo;
 - **l'avvio della rotazione** comporta un **ritardo** di **1/2 sec.**
- Caratteristiche tipiche di un floppy da 3.5"
 - Capacità di **1.44 MB**
 - Tracce x settori: **80 x 18**
 - RPM = **300**
 - velocità di trasferimento di **500Kbps**

Hard Disk IDE/EIDE

- Situazione originaria:
 - disco contenuto nel PC XT IBM, Seagate da 10 MB con 4 testine, 360 cilindri e 17 settori/traccia, il controllore era in grado di gestire due unità;
 - il SO inseriva parametri nei registri CPU e poi **chiamava il BIOS (Basic Input Output System)**.
- **IDE (Integrated Drive Electronics)**
 - controllore integrato nell'unità;
 - procedure di chiamata del BIOS immutate
 - 4 bit per la testina, 6 bit per il settore e 10 bit per il cilindro;
 - un'unità poteva avere al massimo 16 testine, 63 settori e 1024 cilindri per un totale di 1.032.192 settori (528 MB);
- **EIDE (Extended IDE)**
 - supportano lo schema **LBE (Logical Block Addressing)**,

ATA – IDE – EIDE

- **Integrated Drive Electronics – IDE (1986)**
 - Proposto da Western Digital & Compaq fu poi incluso nello standard ATA (AT Attachment).
 - Integra le funzioni di controllo sul drive (riduce i costi e migliora la compatibilità).
 - 16 bit - Max 2HD di 528MB ciascuno
- **Enhanced IDE – EIDE (1993)**
 - backward compatibility e DTR superiori
 - 4 dispositivi su due canali (master/slave x2)
 - Diversi standard di trasferimento
 - ATAPI per il supporto di periferiche diverse
 - PIO mode 3 & 4, DMA mode 1 & 2
 - Ultra DMA o Ultra ATA (1997)
 - 33MBps & Cyclical Redundancy Check (CRC).

Prestazioni EIDE

Mode	DTR (MBps)	Connector	Cable	CRC
Mode 3 PIO	11.1	40-pin IDE	40-way	No
Mode 4 PIO	16.6	40-pin IDE	40-way	No
Mode 1 DMA	13.3	40-pin IDE	40-way	No
Mode 2 DMA	16.6	40-pin IDE	40-way	No
Ultra ATA Mode 2	33.3	40-pin IDE	40-way	Yes
Ultra ATA Mode 4	66.6	40-pin IDE	80-way	Yes

Small Computer System Interface – SCSI (1986)

- Richiede un'interfaccia con il bus di sistema (**host adaptor**)
- Può controllare **8/16 dispositivi** (compreso l'host adaptor), HD, CD-ROM, scanner, ...
 - Ogni dispositivo è identificato da un ID
 - I connettori possono essere esterni o interni
 - Di solito l'ID num. 0 è riservato al disco di bootstrap

Prestazioni SCSI

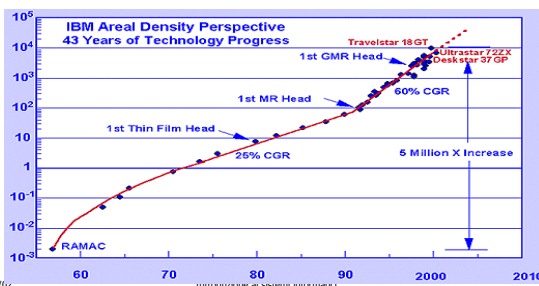
Versione SCSI	Frequenza (MHz)	Bus width (bit)	DTR max (MBps)	Max. num. dispositivi	Lungh. max. del cavo
SCSI-1	5	8	5	7	6m
SCSI-2	5	8	5	7	6m
Wide SCSI	5	16	10	15	6m
Fast SCSI	10	8	10	7	6m
Fast Wide SCSI	10	16	20	15	6m
Ultra SCSI	20	8	20	7	1.5m
Ultra SCSI-2	20	16	40	7	12m
Ultra2 SCSI	40	16	80	15	12m
Ultra160 SCSI	80	16	160	15	12m

Trend

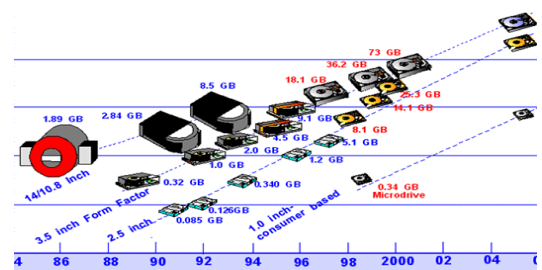
(1)

- **Densità**
 - Continua a crescere oltre le più rosee previsioni (35Gbits/in² in lab vs. 20GB/disco 3.5" in com)
- **Capacità**
 - Crescita accelerata (10MB nel 1981, oltre 10GB oggi e 100 GB entro l'anno prossimo)
- **Velocità di rotazione**
 - 7200 RPM è ormai lo standard (15000 RPM entro il 2000)
- **Form Factor**
 - Sempre più piccoli (oggi 3.5", domani 2.5")
 - Microdrive IBM (1999): 340 MB in un disco di un pollice di diametro e alto meno di ¼"

Andamento densità HD [by IBM]



Andamento capacità HD [by IBM]



Trend

(2)

> Prestazioni

- La velocità di trasferimento cresce più velocemente di quella di posizionamento (seek & latency).

> Affidabilità

- A livello di singolo dispositivo non sta crescendo come gli altri indici, anche perché la tecnologia viene sempre spinta al massimo.
- A livello di sistema è cresciuta grazie a sistemi RAID (**Redundant Arrays of Inexpensive Disks**)

> Interfaccia

- Praticamente invariata: IDE/ATA vs. SCSI

Approfondimento: RAID

RAID

> Le prestazioni dei dischi crescono più lentamente che quelle delle CPU

- accesso ai dischi migliorato di **5/10 volte** in 20 anni, frequenza di clock delle CPU raddoppia ogni 18 mesi;
- gap** di prestazioni sempre **più ampio**.

> Parallelizzazione per migliorare le prestazioni

- RAID - Redundant Array of Inexpensive Disks** vs. **SLED - Single Large Expensive Disk**.
- RAID = scatola piena di dischi
 - server di grosse dimensioni + controllore RAID
 - dal punto di vista del sistema operativo un RAID assomiglia ad uno SLED, ma fornisce prestazioni migliori e più affidabilità
 - realizzato in genere usando dischi SCSI
 - i dati vengono distribuiti fra le diverse unità, permettendo così il funzionamento in parallelo.

RAID livello 0 (striping without parity)

> Il disco singolo virtuale simulato dal RAID è diviso in strisce di k settori ciascuna:

- i settori da 0 a $k - 1$ costituiscono la striscia (**strip**) 0
- i settori da k a $2k - 1$ costituiscono la striscia 1
- ...

> La distribuzione dei dati su unità multiple si chiama **striping**.

> Diminuisce l'affidabilità.

> Cresce la velocità se le operazioni sfruttano il parallelismo.

Strip 0	Strip 1	Strip 2	Strip 3
Strip 4	Strip 5	Strip 6	Strip 7
Strip 8	Strip 9	Strip 10	Strip 11

RAID livello 1 (mirroring)

> Tutti i dischi sono **duplicati**

- con quattro dischi principali ci sono anche quattro dischi di backup;
- ogni scrittura viene eseguita due volte, la lettura può essere eseguita su una delle due copie:
 - le prestazioni in scrittura sono **uguali** a quelle di un'unità singola,
 - le prestazioni di lettura possono essere fino a **due volte** superiori.

> La **tolleranza agli errori** è eccezionale

- se un'unità smette di funzionare basta usare la copia;
- per la riparazione è sufficiente installare una nuova unità e copiarvi i dati di backup.

Strip 0	Strip 1	Strip 2	Strip 3	Strip 0	Strip 1	Strip 2	Strip 3
Strip 4	Strip 5	Strip 6	Strip 7	Strip 4	Strip 5	Strip 6	Strip 7
Strip 8	Strip 9	Strip 10	Strip 11	Strip 8	Strip 9	Strip 10	Strip 11

RAID livello 3

> Versione semplificata del RAID livello 2

- per ogni parola di dati viene calcolato un bit di parità che viene scritto in un'apposita unità;
- nota la posizione dell'errore, la parità ne consente la correzione.

Bit 0	Bit 1	Bit 2	Bit 3	Parity

La memoria di massa (ottica)

Dischi ottici

- Lettura ottica basata sulla riflessione (o sulla mancata riflessione) di un raggio laser.
- Densità di registrazione più alte dei dischi magnetici.
- Creati in origine per registrare i programmi televisivi, poi usati come dispositivi di memoria nei calcolatori.
- Diversi tipi/caratteristiche
 - CD-ROM
 - CD-R
 - CD-RW
 - DVD
 - DVD-RAM
 - ...

04/04/02

Introduzione ai sistemi informatici

80

Compact Disk - CD

- Proposto nel 1980 [da Philips e Sony] per sostituire i dischi in vinile per la musica.
- Standard internazionale IS-10149 [**libro rosso**].
 - diametro di **12 cm**, spessore di 1.2 mm con un foro di 15 mm in mezzo;
 - produzione:
 1. laser ad alta potenza che brucia fori di $0,8 \mu\text{m}$ in un **disco master** (le depressioni si chiamano **pit** e le aree fra pit si chiamano **land**);
 2. dal master si ricava uno **stampo**;
 3. nello stampo viene iniettata una resina liquida di **policarbonato** che forma un CD con la stessa sequenza di fori del master,
 4. sul policarbonato viene depositato uno strato molto sottile di **alluminio riflettente**,
 5. copertura con uno strato **protettivo** e infine con un'**etichetta**.

04/04/02

Introduzione ai sistemi informatici

81

Lettura di un CD

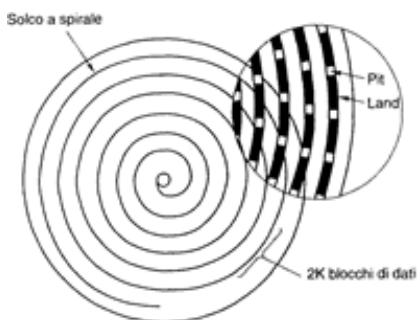
- Un **laser a bassa potenza** manda una luce infrarossa (lunghezza d'onda di $0,78 \mu\text{m}$) sul disco.
- I **pit** appaiono come **cunette** su una superficie piatta:
 - un pit è alto circa un quarto della lunghezza d'onda del laser,
 - la luce riflessa da un pit è sfasata di mezza lunghezza d'onda rispetto alla luce riflessa dalla superficie circostante,
 - l'interferenza negativa riduce l'intensità della luce riflessa.
- I passaggi **pit/land** o **land/pit** indicano un **1**, la loro assenza indica uno **0**.
- Pit e land sono scritti in una **spirale** unica che compie 22.188 giri attorno al disco (circa 600 per ogni mm).
- **Velocità lineare costante** (120 cm/sec):
 - all'interno è di 530 rpm, all'esterno deve scendere a 200 rpm;
 - l'unità è diversa da quella a velocità angolare costante usata per gli HD;
 - 530 rpm sono molti meno dei 3600/10440 rpm degli HD.

04/04/02

Introduzione ai sistemi informatici

82

Pit e land su un CD



04/04/02

Introduzione ai sistemi informatici

83

CD-ROM

- 1984: Philips e Sony pubblicano il **libro giallo**, in cui viene definito lo standard dei **CD-ROM (Compact Disc-Read Only Memory)**.
 - viene definita la struttura e il formato da utilizzare per memorizzare dati digitali invece che "semplice" musica.
- Rispetto ai CD audio i CD-ROM hanno
 - stesse **dimensioni**;
 - compatibilità **dell'ottica** e della **meccanica**;
 - stesso **processo produttivo**;
 - miglior capacità di **correggere errori**.
- Il **libro verde** [1986] aggiunge grafica e possibilità di mischiare audio, video e dati nello stesso settore.

04/04/02

Introduzione ai sistemi informatici

84

Velocità/capacità dei CD-ROM

- **Velocità base** (1x)
 - 75 settori/sec,
 - 153.6 KByte/sec (175.2 in modalità 2).
 - Velocità superiori crescono in proporzione
 - 32x corrisponde a 2400 settori/sec cioè quasi 5MB/sec
- **Capacità**
 - 74 minuti di musica = 681.984.000 byte = circa 650 MB;
 - 80 minuti di musica = circa 700 MB.
- **Tempo di accesso**
 - alcune **centinaia** di millisecondi.

File System

- Era necessario garantire la **compatibilità** con **diversi sistemi operativi** ➔ standard **"High Sierra"** (IS 9660);
- Tre livelli:
 - Livello 1 ("DOS")
 - nomi di 8 caratteri con estensione opzionale di 3 caratteri
 - i nomi possono contenere solo lettere maiuscole, numeri e " _ "
 - le cartelle possono essere annidate fino a otto livelli
 - i nomi di cartella non possono avere estensioni.
 - Livello 2
 - nomi con 32 caratteri.
 - Livello 3
 - file non contigui.
- Alcune estensioni di questo standard consentono di avere nomi molto lunghi (255 caratteri), e link simbolici.


CD Recordable (CD-R)

- CD che vengono scritti una sola volta (**WORM**):
 - utilizzati per backup, per produzioni in piccole serie, per la generazione di master, ...
 - standard definito nel **libro arancione**, dove si introduce anche il **CD-ROM XA** (CD-R scritti in modo incrementale);
 - stesse dimensioni dei CD-ROM
 - dischi di polycarbonato di 120 mm;
 - contengono un solco largo 0,6 mm (**guida** per il laser di scrittura).
- La riflettività di pit e land è simulata
 - c'è uno strato di colore fra il polycarbonato e lo strato riflettente: nello stato iniziale questo strato è trasparente;
 - per scrivere, un laser ad alta potenza colpisce un punto nello strato della superficie colorata, rompe un legame chimico e crea una macchia scura.

CD ReWriteable (CD-RW)

- Dischi ottici **riscrivibili**.
- Lo strato di registrazione utilizza una lega di argento, indio, antimonio e tellurio che ha **due stati stabili**:
 - lo stato **cristallino** con elevata capacità di riflessione (land);
 - lo stato **amorfo** con ridotta capacità di riflessione (pit).
- Si usa un **laser** con **tre potenze diverse**:
 - ad **alta potenza** il laser scioglie la lega e un raffreddamento rapido la porta dallo stato cristallino allo stato amorfo;
 - a **potenza media** la lega si scioglie e si raffredda tornando nel suo stato cristallino;
 - a **bassa potenza** si rileva solo lo stato del materiale.

Digital Versatile Disk (DVD)

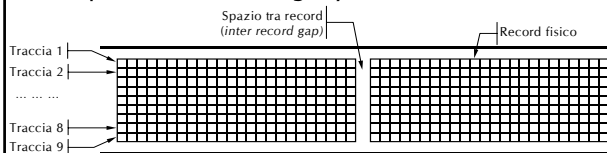
- 
- The image shows two optical discs. The top one is a CD with a purple surface and a spiral track. The bottom one is a DVD with a blue surface and a more densely packed spiral track. Labels 'CD' and 'DVD' are placed next to their respective discs.
- Evoluzione tecnologica ➔ maggior densità dei dati:
 - pit più piccoli (0.4 vs. 0.8 μm);
 - spirale più serrata (0.74 vs. 1.6 μm);
 - laser rosso (0.65 vs. 0.78 μm).
 - Caratteristiche dei DVD
 - capacità di 4.7 GB
 - 133 minuti di video fullscreen MPEG-2 ad alta risoluzione (720 x 480) con colonna sonora in 8 lingue e sottotitoli in altre 32;
 - 1x indica 1.4 MB/sec (vs. 150 KB/sec).

Diversi formati di DVD

- Esistono situazioni in cui servono **più di 4.7 GB**. Pertanto sono stati definiti quattro formati:
 1. Lato unico, strato unico (4,7 GB).
 2. Lato unico, strato doppio (8,5 GB).
 3. Due lati, strato unico (9,4 GB).
 4. Due lati, strato doppio (17 GB).
- Tecnologia dello strato doppio:
 - uno strato riflettente sul fondo coperto da uno strato semiriflettente; a seconda di dove viene indirizzato il laser, il raggio viene riflesso da uno strato o dall'altro;
 - lo strato inferiore ha pit e land leggermente più grandi, per cui la sua capacità è leggermente inferiore.

Nastri Magnetici e unità DAT

➤ Capacità di diversi GigaByte



- Accesso sequenziale
- Molto lenti
- Utili solo per operazioni di *backup*

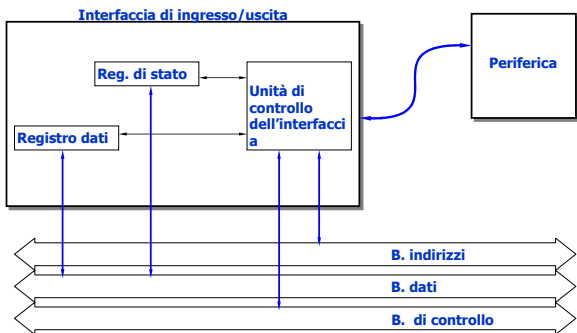
Gestione dell' I/O

I/O

➤ L' I/O può essere effettuato in 3 modalità:

- controllo da programma;
- interruzione (interrupt);
- DMA (Direct Memory Access).

Collegamento periferica-calcolatore



Esigenze

- Evitare perdite o duplicazioni di dati.
- Consentire comunicazioni asincrone.
- Nel caso di lettura da tastiera, le comunicazioni sono:
 - da tastiera a porta;
 - da porta a cella di memoria;
 - da cella di memoria a programma che utilizza il dato;
 indipendentemente dalle modalità di gestione di I/O adottata.

A controllo di programma

- Durante la sua normale esecuzione un programma esegue una istruzione di lettura della porta:
 - es Intel: `IN R0, INDIRIZZOPORTA`
 - es Motorola: `MOV INDIRIZZOPORTA, R0`
- Nella fase di esecuzione di questa istruzione il processore esegue il ciclo di bus di lettura della porta.
- Il programmatore ha deciso dove, nel programma, inserire questa istruzione.
- Il flusso dell'esecuzione del programma stabilirà **quando** l'istruzione verrà eseguita.

A interruzione

- La parte di programma che legge la porta (ad es. con la istruzione `IN R0, INDIRIZZOPORTA`) **NON** è nel programma ma è silente in memoria in una locazione convenuta.
- Quando l'interfaccia della periferica porta il dato alla porta di ingresso, con un segnale allerta il processore.
- Il processore interrompe l'esecuzione del programma in corso e salta automaticamente a eseguire la parte di programma che legge la porta. La lettura avviene come nel caso precedente.
- Al termine di questo, il processore riprende il programma interrotto.
- In pratica, la periferica ha deciso **quando** l'istruzione di lettura della porta deve essere eseguita.

DMA

- Quando l'interfaccia della periferica porta il dato alla porta di ingresso, manda un segnale al processore, imponendogli di lasciare libero il bus.
- Appositi circuiti generano un ciclo di bus che forza l'attivazione della porta, genera l'indirizzo in memoria dove deve finire il dato, comanda la memoria alla scrittura.
- Intanto, il processore non utilizza il bus.
- Terminato il ciclo, l'interfaccia della periferica manda un altro segnale al processore, lasciandolo libero di proseguire.
- In pratica, alcuni circuiti di I/O hanno scritto il dato in memoria, pochi nanosecondi dopo il suo arrivo.

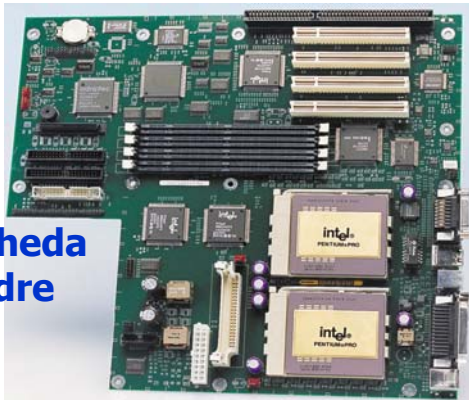
Dispositivi di Ingresso/Uscita (I/O)



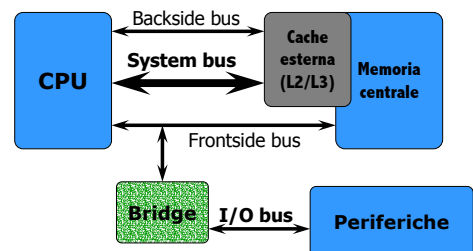
Struttura fisica di un PC

- Nella scatola (case) sono contenuti:
 - una **scheda madre** che contiene una CPU, alcuni connettori nei quali inserire moduli DIMM e vari chip di supporto;
 - uno o due **bus**, uno ad alta velocità (per schede moderne) e uno a bassa velocità (per schede più vecchie);
 - prese in cui si possono inserire i connettori delle **schede di I/O** che agiscono da **controllori** dei dispositivi di I/O, cioè ne **gestiscono l'accesso al bus**:
 - un controllore che legge o scrive dati verso e da una memoria senza interventi da parte della CPU effettua un accesso diretto alla memoria (**Direct Memory Access – DMA**)
 - completato il trasferimento, il controllore **effettua un interrupt**, la CPU sospende il programma in corso e inizia una procedura speciale, (**interrupt handler**); quando l'interrupt handler termina, la CPU continua con il programma.

La scheda madre



System & I/O Bus



Quanti bus di I/O

- Il bus di accesso alla memoria è condiviso dalla CPU e dai dispositivi di I/O: possono esserci dei **conflitti**
 - **arbitro del bus** decide a chi tocca;
 - i dispositivi di I/O hanno la precedenza sulla CPU.
- Problemi
 - il bus non regge il carico ed è il collo di bottiglia del sistema;
 - bisogna continuare a supportare le periferiche già disponibili.
- Soluzione: due bus
 - quello "vecchio" **ISA (Industry Standard Architecture)** o **EISA (Extended ISA)**;
 - un "nuovo" **PCI (Peripheral Component Interconnect)**.

Evoluzione dei bus I/O /1

- **Industry Standard Architecture** – ISA (1980)
 - 8/16 bit – 4.77/8 Mhz – fino a 5MBps effettivi
 - Introdotto fin dai primi PC IBM (PC/AT)
 - Presente praticamente su tutti i sistemi
 - In fase di estinzione
- **Micro Channel Architecture** – MCA (1987)
 - 32 bit – 10 Mhz – più di 20MBps – P&P
 - Incompatibile con ISA (**no backward compatibility**)
 - Architettura proprietaria IBM (!!)
- **Extended ISA** – EISA (1988)
 - 32 bit – 10 Mhz – più di 20MBps – P&P
 - Compatibile con ISA (backward compatibility)

Evoluzione dei bus I/O /2

- **VESA Local Bus** – VLB (1992)
 - Video Electronics Standards Association (VESA)
 - Strettamente accoppiato con il processore
 - Progettato per 486, difficile adattarlo ai successori
 - Non più di due dispositivi oltre i 33MHz
- **Peripheral Component Interconnect** – PCI
 - 32/64 bit – 33/66 MHz – **133/266 MB/sec** – **P&P**
 - Consente la condivisione degli indirizzi di interrupt IRQ
 - Tipico utilizzo per dischi, schede grafiche, ...
- **PCI-X** (by IBM, HP & Compaq)
 - 64 bit – 133MHz – **1.0 GBps** – P&P
 - Nato per GigabitEthernet, Ultra3SCSI, FiberChannel, ...

Evoluzione dei bus I/O /3

- **Accelerated Graphics Port** – AGP (1997)
 - Collega scheda video, processore e memoria
 - Permette di utilizzare la memoria di sistema quando quella della scheda grafica si esaurisce.
 - 32 bit – 66 MHz – 254.3/1017MBps
 - Libera il bus PCI dal traffico della scheda video
 - Richiede un sistema di prestazioni elevate per essere sfruttata appieno
 - Banda passante di un sistema:
 $64 \text{ bit} \times 133 \text{ MHz} = 1017 \text{ MBps}$

Terminali

- Composti di due parti: **tastiera** e **schermo**.
 - Nel mondo dei mainframe, sono integrati in un dispositivo singolo e collegati al calcolatore principale per mezzo di una linea seriale
 - Nel settore dei personal computer, sono dispositivi separati.
- **Tastiere**
 - molti tipi diversi, meccaniche o elettromagnetiche;
 - quando si preme un tasto viene generato un interrupt e viene avviato il gestore degli interrupt della tastiera, che legge un registro hardware all'interno del controllore della tastiera per avere il numero del tasto (da 1 a 102) premuto;
 - quando il tasto viene rilasciato si verifica un secondo interrupt.

Monitor CRT (Cathode Ray Tube)

- Un cannone spara un **raggio di elettroni** contro uno **schermo fosforescente** (per la riproduzione dei colori si usano tre cannoni, per il **rosso**, il **verde** e il **blu**).
- Il raggio viene **deflessa** in modo da coprire tutti i punti dello schermo, una riga per volta (**raster scan**).
- Un'immagine a schermo pieno viene completata **30/60** volte al secondo.
- Davanti allo schermo c'è una **griglia** che lo divide in **punti**:
 - quando la griglia ha una **carica positiva** gli elettroni vengono accelerati **raggiungono lo schermo**;
 - quando la griglia ha una **carica negativa** gli elettroni vengono respinti e il punto sullo **schermo rimane spento**.

LCD (Liquid Crystal Display)

- Schermi "piatti", **leggeri** e facilmente trasportabili.
- **Cristalli liquidi**: molecole organiche viscosi
 - **scorrono** come un liquido;
 - hanno una struttura **tridimensionale**, come un **cristallo**;
 - quando tutte le molecole sono allineate le proprietà ottiche del cristallo dipendono da **direzione** e **polarizzazione** della luce in ingresso;
 - un **campo elettrico modifica** l'allineamento molecolare e quindi **le proprietà ottiche**.

Struttura di un LCD

- Un LCD è compreso in **due lastre di vetro parallele** nella cui intercapedine sono contenuti i cristalli liquidi.
- Ogni lastra è "rivestita" da **elettrodi trasparenti**.
- Una **luce** (naturale o artificiale) situata dietro alla lastra posteriore **illumina lo schermo da dietro**.
- Gli **elettrodi** attaccati alle lastre di vetro vengono usati per **creare campi elettrici** nel cristallo.
- Le diverse parti dello schermo ricevono voltaggi diversi a seconda dell'immagine desiderata.
- Sulla parte anteriore e posteriore dello schermo vi sono dei **polarizzatori**, che servono a filtrare la luce che attraversa il cristallo.

Esempio di LCD: Twisted Nematic

- La **lastra posteriore** ha **scanalature orizzontali** e dietro lo schermo c'è un **polarizzatore orizzontale**.
- La **lastra anteriore** ha **scanalature verticali** e davanti allo schermo c'è un **polarizzatore verticale**.
- Se non c'è campo elettrico le molecole LCD tendono ad allinearsi con le scanalature: le molecole subiscono una rotazione di 90° e deviano di 90° la luce che le attraversa, in questo modo la luce passa!
 - In **assenza** di **campo elettrico** lo schermo LCD è completamente **luminoso**.
 - Applicando **una tensione** in alcuni punti della lastra si **distrugge la struttura** e si **blocca la luce**.

Schermi piatti (LCD)

- **Double-layer SuperTwist Nematic – DSTN**
 - Tecnologia LCD a **matrice passiva**;
 - sono anche chiamati "**dual-scan LCD**".
- **Thin Film Transistor – TFT**
 - Ogni **pixel** è controllato da 1-4 transistor;
 - sono anche detti LCD a "**matrice attiva**".
- **Equivalenza con CRT**
 - LCD di 13.5in = CRT di 15in (800 x 600)
 - LCD di 14.5in = CRT di 17in (1024 x 768)
 - LCD di 18.0in = CRT di 21in (1280 x 1024)

Confronto tra monitor

Caratteristica	PMLCD	AMLCD	CRT
Angolo visuale	50-90°	140°	180°
Contrasto	40:1	140:1	300:1
Risposta	300ms	25ms	NA
Luminosità	70-90	70-90	220-270
Potenza	45	50	180
Tempo di vita	60Kh	60Kh	anni

Terminali a caratteri

- Basati su una visualizzazione "**character map**" che riproduce il contenuto di una **memoria video**:
 - ogni carattere è associato a un **attribute byte** (colore, intensità, intermittenza e così via);
 - la scheda video richiede caratteri alla RAM video e genera i segnali necessari al funzionamento dello schermo.

Terminali grafici

- Visualizzazione "**bit map**": lo schermo è una matrice di **pixel indipendenti**
 - per indicare il **colore** di ogni pixel si usano fino a **32 bit** (8 bit per ogni colore fondamentale + 8 bit per la trasparenza);
 - per rappresentare un carattere si usa un rettangolo di pixel e si configurano i bit necessari per visualizzare il carattere (così si possono realizzare diversi **font**);
 - comodi per i **sistemi operativi a finestre**;
 - richiedono una **memoria video** di grandi dimensioni
 - VGA: 640 x 480 x 4 byte = 1.2 Mbyte
 - SVGA: 800 x 600 x 4 byte = 1.9 Mbyte
 - XGA: 1024 x 768 x 4 byte = 3.2 Mbyte
 - UXGA: 1600 x 1200 x 4 byte = 7.5 Mbyteriducibili grazie all'utilizzo di una "**palette**" (scelta di **2⁸=256** colori tra i **2³²** possibili).

Mouse

- Interfaccia "**point-and-click**" vs. "**command line**"
 - muovendo il mouse si sposta il cursore;
 - pressione di un tasto ⇒ invio di un comando;
 - il comando dipende dalla posizione del cursore.
- Diversi tipi di mouse
 - **meccanici**: movimento rilevato da sensori che controllano la rotazione di una pallina incastrata sotto il mouse;
 - **ottici** (vecchio tipo): un "LED" invia luce verso un "pad" che la riflette a un "fotorelettore", sul pad è disegnata una griglia di linee e il fotorelettore è in grado di rilevare il passaggio sopra una di queste linee;
 - **ottici** (nuovo tipo): una sorta di telecamera osserva il piano sotto il mouse e, confrontando le immagini riprese in istanti diversi, rileva il movimento
 - ...

Interazione mouse-computer

- Ogni volta che si **sposta**, il mouse invia una sequenza di 3 byte al calcolatore lungo una linea **seriale**:
 - un intero che indica lo **spostamento X**;
 - un intero che indica lo **spostamento Y**;
 - un intero che indica lo stato dei **pulsanti**.
- Il SO accetta queste informazioni e converte le indicazioni **relative** inviate dal mouse nella posizione **assoluta** del cursore.

Porte Standard

- **Interfaccia Seriale**
 - Trasporta un bit per volta.
 - Velocità massima di 115 kbps
 - Utilizzata per periferiche lente, come mouse e modem esterni
- **Interfaccia parallela**
 - Trasporta 8 bit alla volta.
 - Velocità di 150 KB/sec (2MB/s in modalità EPP)
 - Usata per stampanti, scanner e unità di backup (nastri, Zip).
- Direzione della comunicazione
 - **Simplex**: la linea trasmette solo in una direzione;
 - **Half-duplex**: la linea trasmette in entrambe le direzioni ma non contemporaneamente (una direzione per volta);
 - **Full-duplex**: la linea trasmette contemporaneamente in entrambe le direzioni.

Universal Serial Bus – USB

- Definito da un consorzio (Intel, Compaq, Microsoft, ...), con l'intento di **sostituire** le attuali **porte seriali** e **parallele**.
- Velocità di **12 MBit/sec**.
- Collega fino a **127** periferiche in cascata.
- Può **alimentare** direttamente le **periferiche a basso consumo** (e.g. tastiere e mouse).
- Completamente **Plug and Play** (anche per collegamento "**a caldo**").
- **USB 2.0** (1999) arriva fino a **360-480Mbps**.

Firewire 1394

- Bus seriale ad **alte prestazioni** per la connessione di periferiche.
- Connette **64 periferiche in cascata**.
- Supporta il **Plug and Play** e **connessione a caldo**.
- Velocità di trasferimento di **400/800 Mbps**.
- Adatto per videocamere e videoregistratori digitali, lettori DVD e periferiche audio.

Riassunto caratteristiche

Standard	Utilizzo	Burst DTR	Note
ATA/IDE	HD, CD, DVD	3.3 – 66.6	Standard per HD
SCSI	HD, dischi removibili, scanner	5 – 80	Standard per alte prestazioni
USB	Scanner, fotocamere digitali	12	Sostituisce porte parallela/seriale
IEEE 1394	Videocamere, dispositivi ad alte prestazioni	400	Diffusione nel 2000/01

Stampante ad aghi

> Funzionamento

- la testina di stampa contiene fra 2 e 24 aghi;
- ogni ago è azionato da un'elettrocalamita;
- mentre la testina si muove, l'azione combinata degli aghi compone i caratteri da stampare;
- la qualità di stampa dipende dal numero degli aghi e dalla sovrapposizione dei punti (che però influenza anche la velocità di stampa).

> Caratteristiche e utilizzo:

- **economiche** e molto **affidabili**
- **lente, rumorose** e con **grafica di bassa qualità**
- Tre applicazioni principali:
 1. stampa su **formulari** prestampati di grandi dimensioni (> 30 cm),
 2. stampa su **piccoli pezzi** di carta,
 3. stampa su **formulari continui** a più segmenti con **carta carbone**.

Stampante a getto d'inchiostro

> Funzionamento

- al posto degli aghi ci sono **ugelli** collegati a serbatoi d'inchiostro di diversi colori;
- mentre la testina si muove, gli ugelli spruzzano gocce d'inchiostro in modo da comporre i caratteri da stampare;
- la qualità di stampa dipende dalla dimensione delle gocce.

> Caratteristiche e utilizzo:

- risoluzioni che vanno da 300 a 1440 **dpi (dots per inch)**;
- **poco costose, silenziose** e di **buona qualità**;
- **lente**, usano **cartucce** d'inchiostro **costose** e producono documenti intrisi d'inchiostro;
- uso domestico, **SOHO** (small office, home office).

Stampante laser

> Funzionamento

1. il **tamburo** viene caricato fino a circa 1000 volt;
2. un raggio **laser** scorre sul tamburo e la sua modulazione produce una configurazione di punti chiari e scuri (i **punti** colpiti dal raggio **perdono** la loro **carica** elettrica);
3. la **rotazione** del tamburo permette di costruire le varie righe;
4. quando una riga si avvicina al **toner**, i punti carichi attirano la polvere d'inchiostro;
5. il tamburo ricoperto di toner viene premuto sulla **carta** e trasferisce la polvere nera sulla carta;
6. la carta passa attraverso dei rulli riscaldati che **fissano** il toner;
7. il tamburo viene **scaricato** e **ripulito** di eventuali residui.

> Caratteristiche

- alta **qualità**, eccellente **flessibilità**, buona **velocità** e costi contenuti;
- **bianco e nero**;
- tecnologia simile a quella delle **fotocopiatrici**.

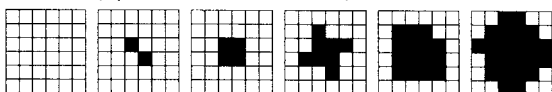
Stampante laser per foto B&W

> Fotografia letta a 600 dpi da stampare a 600 dpi:

- l'immagine letta contiene 600 x 600 dpi, ognuno dei quali si compone di un valore grigio da 0 (bianco) a 255 (nero);
- la stampante stampa con 600 dpi, ma ogni pixel stampato è nero (con toner) o bianco (senza toner).

> Mezzotono (halftoning)

- L'immagine viene suddivisa in celle di 6 x 6 pixel, ogni cella può contenere fra 0 e 36 pixel neri;
- i valori di grigio fra 0 e 255 vengono rappresentati dividendo questa gamma in 37 zone: i valori da 0 a 6 si collocano nella zona 0, quelli da 7 a 13 nella zona 1, ...



Stampa a colori

> Immagini a **luce trasmessa** (e.g. CRT)

- create dalla sovrapposizione lineare dei tre colori primari **additivi**: rosso, verde e blu (RGB).

> Immagini a **luce riflessa** (e.g. fotografia)

- create dalla sovrapposizione lineare dei tre colori primari **sottrattivi**: ciano, giallo e magenta (**CYM**);
- per produrre un buon nero i sistemi di stampa a colori utilizzano un quarto inchiostro: quello nero (black → **CMYK**).

> Problemi di **conversione** da schermo a stampa:

- gli **schermi** usano **luce trasmessa**, le **stampanti luce riflessa**;
- i **CRT** usano colori primari con **256 diverse intensità**, le **stampanti** a colori devono usare il **mezzotono**;
- gli **schermi** hanno **sfondo nero**; la **carta** ha **sfondo chiaro**;
- le **gamme** di colore di **RGB** e **CMYK** sono diverse.

Stampanti a colori /1

- **Getto d'inchiostro a colori**
 - buoni risultati per la grafica a colori
 - risultati mediocri per le fotografie.
- Per risultati migliori si usano inchiostri e carta speciali:
 - **inchiostri a base asciutta**
 - si basano su sostanze coloranti dissolte in un fluido
 - colori accesi che scorrono facilmente
 - scoloriscono se esposti ai raggi ultravioletti, come quelli del sole.
 - **inchiostri a base di pigmenti**
 - contengono particelle solide di pigmento sospese in un mezzo fluido che evapora dalla carta lasciando il pigmento
 - non scolorano con il tempo
 - sono meno intensi di quelli a base asciutta e le particelle di pigmento hanno la tendenza ad intasare gli ugelli (pulizia periodica)
 - la stampa di fotografie richiede carta speciale trattata o lucida.

Stampanti a colori /2

- **Stampanti a inchiostro solido**
 - **inchiostro speciale** a base di cera in quattro blocchi solidi;
 - tempo di **avviamento lungo** per sciogliere l'inchiostro;
 - l'inchiostro caldo viene spruzzato sulla carta, si rapprende e si fonde con la carta quando passa attraverso due rulli.
- **Stampante laser a colori**
 - funziona come il modello monocromatico ma genera **quattro immagini**: una per ogni colore **C, Y, M e K**;
 - un'immagine di 1200 x 1200 dpi per una pagina di 80 in² richiede **115 milioni di pixel**, con 4 bit/pixel la stampante ha bisogno di **55 MB** solo per la grafica, senza contare la memoria per processori interni, caratteri ecc
 - **costosa, veloce**, di alta **qualità** e crea **immagini stabili** nel tempo.

Stampanti a colori /3

- **Stampante a cera.**
 - un nastro di cera a quattro colori viene tagliato in bande delle dimensioni di una pagina;
 - elementi di riscaldamento sciolgono la cera mentre la carta passa al di sotto;
 - la cera viene fissata alla carta sotto forma di pixel.
- **Stampante a sublimazione**
 - un contenitore dei colori CMYK passa sopra una testina di stampa termica contenente migliaia di elementi di riscaldamento programmabili;
 - gli inchiostri vengono vaporizzati istantaneamente e assorbiti da una carta speciale;
 - ogni elemento di riscaldamento è in grado di produrre 256 temperature diverse;
 - è possibile ottenere colori quasi continui per ogni pixel, per cui non c'è bisogno del mezzotono.

Modem /1

- Connessione di calcolatori attraverso la rete telefonica (**analogica**).
- Velocità crescenti dal 1980 in poi
 - V.22bis, V.32 & V.32bis furono i primi standard per velocità di 2.4, 9.6 e 14.4Kbit/s.
 - V.34 (1994) supporta 28.8Kbit/s e corrisponde al minimo livello attualmente accettato
 - V.34+ (1996) arriva a 33.6Kbit/s
 - V.90 arriva a 56Kbit/s downstream e a 33.6Kbit/s upstream.
 - downstream indica dal digitale all'analogico
 - upstream indica dall'analogico al digitale

Modem /2

- La linea telefonica trasmette bene segnali tra 1000 e 2000 Hz ➡ si usano come portanti (**carrier**).
- Modulazione del carrier per portare un segnale digitale
 - Modulazione di **ampiezza** usa due voltaggi diversi per 0 e 1;
 - Modulazione di **frequenza (frequency shift keying)** tensione costante, ma cambia la frequenza
 - Modulazione di **fase** ampiezza e frequenza costanti, cambia la fase.
- Il numero di possibili cambiamenti di segnale al secondo si chiama **baud**.
 - È possibile **associare 2 o più bit a ogni segnale**, allora il **bit rate** è maggiore del **baud rate**.

Integrated Services Digital Network - ISDN

- Linea analogica sostituita da **linea digitale**
 - in realtà non viene sostituita la linea, ma solo le **attrezzature alle due estremità**.
 - **Uso domestico**: due canali digitali indipendenti, ognuno da 64'000 bit/sec, e un canale di segnalazione da 16'000 bit/sec (per un totale di **144'000 bps**)
 - **Uso commerciale**: 30 canali per uso commerciale.
- Caratteristiche
 - tempo di **setup** della connessione praticamente nullo (1 s);
 - non serve più un modem analogico (**connessione digitale-digitale**);
 - è molto più **affidabile** (meno errori) di una linea analogica.

Asymmetric Digital Subscriber Line - ADSL

- Funziona sul **doppino telefonico** tradizionale
- Usa tre canali (in frequenza) diversi sulla stessa linea
 1. Plain Old Telephone System (POTS)
 2. Upstream (64-640Kbps)
 3. Downstream (1.5-6.1Mbps)
- Appartiene alla famiglia di protocolli **xDSL**
 - Diverse velocità di download (fino a 52Mbit/s) e upload (da 64Kbit/s a più di 2Mbit/s)
 - In Italia (a oggi) viene offerta una connessione con 640 Kbps downstream e 128 Kbps upstream.
 - Altre varianti **xDSL**
 - high-bit rate (**HDSL**)
 - single-line (**SDSL**)
 - very-high-data-rate (**VHDSL**).

04/04/02

Introduzione ai sistemi informatici

133

Universal Mobile Telecommunications System

- Noto con l'acronimo UMTS
- Standard per i telefoni cellulari di terza generazione.
- Attivo commercialmente dal 2002
- UMTS potrà fornire ad ogni utente una banda fino a 2Mbit/sec
- Rende possibile la trasmissione attraverso la rete mobile di contenuti multimediali

04/04/02

Introduzione ai sistemi informatici

134

Tassonomia dei calcolatori

Quantità vs. qualità

- Cambiare di un ordine di grandezza la **quantità** significa cambiare anche la **qualità**:
 - un'auto in grado di raggiungere una velocità di 1000 km/h nel deserto del Nevada è una macchina **fondamentalmente diversa** da un'auto che fa 100 km/h sull'autostrada;
 - un grattacielo di 100 piani non è solo un edificio di 10 piani un po' più grande.
- Nei computer le differenze sono di **diversi ordini di grandezza**.
- I miglioramenti procurati dalla legge di Moore possono essere utilizzati in modi diversi:
 - costruire **calcolatori sempre più potenti a prezzo costante**;
 - costruire lo **stesso calcolatore a prezzi** ogni anno **più convenienti**.

04/04/02

Introduzione ai sistemi informatici

136

Calcolatori disponibili /1

Tipo	Prezzo (€)	Applicazione tipica
Calcolatore monouso	1	Biglietti di auguri
Calcolatore dedicato	10	Orologi, automobili, ...
Calcolatore per videogiochi	100	Videogiochi personali
Calcolatore per PC	1 K	PC da tavolo o portatile
Server	10 K	Server di rete
Reti di workstation	100 K	Centro di calc. dipartimentale
Mainframe	1 M	Database di una banca
Supercalcolatore	10 M	Previsioni del tempo

I prezzi sono solo indicativi.

04/04/02

Introduzione ai sistemi informatici

137

Calcolatori disponibili /2

- **Calcolatori monouso**:
 - chip singoli incollati all'interno dei biglietti di auguri;
 - si tratta in pratica di calcolatori usa e getta.
- **Sistemi embedded** (calcolatori dedicati):
 - calcolatori che si trovano in telefoni, televisori, forni, auto, ...
 - questi calcolatori contengono un processore, meno di un megabyte di memoria e qualche funzione di I/O.
- **Videogame**
 - normali calcolatori con particolari capacità grafiche, ma software limitato e poche possibilità di estensione; fanno parte di questa categoria anche i PDA;
 - contengono un processore, alcuni megabyte di memoria, un tipo di schermo (anche un televisore) e poco di più.

04/04/02

Introduzione ai sistemi informatici

138

Calcolatori disponibili /3

> **Personal computer** (PC) o **workstation**:

- dotati di alcune decine di megabyte di memoria, di un disco fisso contenente alcuni gigabyte di dati, drive CD-ROM, modem, scheda audio e altre periferiche;
- dotati di sistemi operativi elaborati, molte opzioni di espansione e una vasta gamma di software disponibile.

> **Server di rete**

- si tratta di PC o workstation potenziati utilizzati come server di rete sia per le reti locali che per Internet;
- esistono sia in configurazione con processore unico che con più processori, hanno alcuni gigabyte di memoria, molti gigabyte di spazio sul disco fisso e interfacce di rete ad alta velocità.

Calcolatori disponibili /4

> **NOW (Networks of Workstations)** o **COW (Cluster of Workstations)**

- composti da PC o workstation normali collegate con reti ad elevate prestazioni (qualche gigabit/sec) e funzionanti con software speciale, che permette a tutte le macchine di lavorare insieme su un unico problema;
- architetture sono facilmente scalabili (da alcune macchine ad alcune migliaia) e sono paragonabili a minisupercomputer.

> **Mainframe**

- calcolatori grandi come una stanza, in uso fin dagli anni '60;
- non sono più veloci di server potenti, ma solitamente hanno più capacità di I/O e sono dotate di grandi insiemi di dischi
- sono macchine estremamente costose, che vengono spesso mantenute per via dell'ingente investimento esistente in termini di software, dati, procedure operative e personale.

Calcolatori disponibili /5

> **Supercomputer**

- hanno CPU velocissime, molti gigabyte di memoria centrale, dischi e reti molto veloci.
- Recentemente molti supercomputer sono diventati macchine altamente parallele non molto diverse dai COW, ma con componenti più veloci e più numerosi.
- I supercomputer vengono utilizzati per risolvere problemi di calcolo molto complicati in campi scientifici e ingegneristici:
 - simulazione di uno scontro fra galassie,
 - sintesi di nuovi farmaci,
 - modelli del comportamento dell'aria attorno alle ali di un aereo.