

Le Memorie

Giuseppe Pozzi

Impianti di Elaborazione
Facoltà di Ingegneria dell'Informazione
Politecnico di Milano

giuseppe.pozzi@polimi.it
- versione del 30 settembre 2003 -

La memoria principale (RAM)

• Bibliografia:

Tanenbaum A. S., Goodman J. R.,
"Architettura dei computer - Un approccio strutturato",
Prentice Hall International, 2000 - paragrafo 2.2

• Sommario:

- Concetto di bit.
- Indirizzi di memoria.
- Memoria cache.
- Memoria virtuale.

30 settembre 2003

Impianti di Elaborazione - Memorie e gerarchie di memorie

2

Il concetto di bit

- Un bit può contenere uno 0 o un 1 e rappresenta l'**unità di base** della memoria.
 - Per memorizzare informazione bisogna avere un dispositivo che possa assumere **almeno due** stati distinti;
 - i dispositivi che hanno **solo due** stati sono più **affidabili** e **veloci** di quelli con tre o più stati;
- A volte si parla di dispositivi che usano l'**aritmetica decimale** intendendo con ciò la codifica binaria delle cifre decimali (**BCD - Binary Coded Decimal**)
 - con **quattro bit** (16 combinazioni) si codificano le cifre **da 0 a 9**, che poi vengono utilizzate per comporre un numero;
 - $1944_{\text{dieci}} = 0001\ 1001\ 0100\ 0100_{\text{BCD}} = 11110011000_{\text{due}}$;
 - 16 bit in formato BCD codificano 10000 numeri (0-10'000);
 - 16 bit in formato binario codificano 65536 numeri (2^{16}).

30 settembre 2003

Impianti di Elaborazione - Memorie e gerarchie di memorie

3

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.

30 settembre 2003

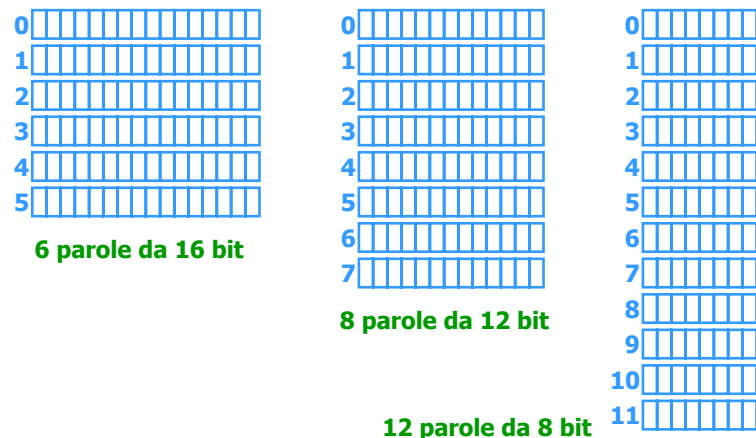
Impianti di Elaborazione - Memorie e gerarchie di memorie

4

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 ($12 \cdot 8 = 96$).

Organizzazione della memoria



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**.

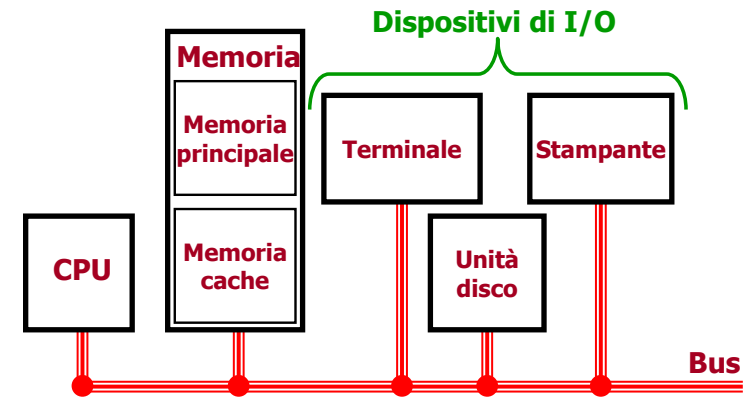
La memoria centrale

- 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

Memoria centrale – DRAM

- Fast Page Mode DRAM (**FPM DRAM**)
 - TA=70-60ns
 - 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
 - Non richiede la disattivazione della colonna e del buffer di uscita.
- 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)

La memoria cache



Memoria cache: SRAM

- Interna (cache di livello 1)
 - stessa frequenza della CPU
- Esterna (cache di livello 2)
 - frequenza inferiore a quella della CPU
- Esterna (cache di livello 3)
 - generalmente saldata sulla motherboard
 - sempre meno usata
 - TA dell'ordine dei nsec, frequenza inferiore alla frequenza della CPU

La memoria virtuale

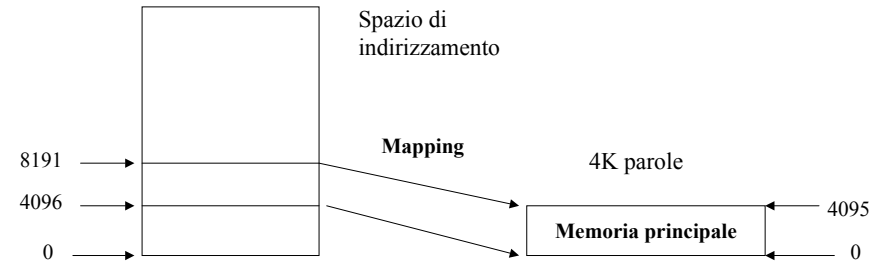
- **Separare il concetto di spazio di indirizzamento e locazioni di memoria**
- Spazio di indirizzamento = numero di parole indirizzabili
 - Dipende esclusivamente dal numero di bit dell'indirizzo e non dal numero di parole di memoria effettivamente disponibile.
 - Se la dimensione della memoria principale è minore dello spazio indirizzabile, senza la memoria virtuale una parte degli indirizzi non sarebbe utilizzabile a causa della corrispondenza uno a uno tra indirizzi e locazioni di memoria.

Spazio di indirizzamento

- Separazione tra spazio di indirizzamento e locazioni di memoria
- Esempio:
 - 16 bit di indirizzo, memoria principale di 4K parole
 - In ogni istante di tempo è possibile accedere a sole 4096 parole di memoria, ma non necessariamente corrispondenti agli indirizzi 0-4095.

Traduzione degli indirizzi

Quando si accede all'indirizzo 4096, questo viene automaticamente tradotto nell'indirizzo 0 della memoria principale, 4097 nell'indirizzo 18191 in 4095.



Memoria virtuale

- Cosa succede se un programma chiede di saltare dall'indirizzo 2500 all'indirizzo 8190?
- Con il meccanismo di memoria virtuale:
 - I contenuti della memoria principale vengono salvati su disco
 - Le parole agli indirizzi da 4096 a 8191 vengono copiate in memoria principale tra gli indirizzi 0 e 4095
 - L'esecuzione del programma prosegue come se non fosse successo nulla
- I blocchi di parole hanno una dimensione fissa (pagina), generalmente una potenza di 2 (ad es. 4K).

Memoria virtuale

- Indirizzi a cui fa riferimento il programma in esecuzione: **indirizzi virtuali**
- Indirizzi di memoria centrale: **indirizzi fisici**
- Meccanismo di traduzione automatico degli indirizzi virtuali in indirizzi fisici
- Con la memoria virtuale, la CPU produce un *indirizzo virtuale*, che viene tradotto da una combinazione di elementi hardware e software (MMU, memory management unit) in un *indirizzo fisico*, che può essere utilizzato per accedere alla memoria principale.

Obiettivi della memoria virtuale

- consentire una condivisione efficiente della memoria da parte di più programmi
- rimuovere il carico di lavoro del programmatore provocato dalla limitata quantità di memoria disponibile
 - La paginazione è trasparente al programmatore
 - Il programmatore può scrivere i propri programmi pensando di avere a disposizione l'intero spazio di indirizzamento virtuale indipendentemente dalle dimensioni effettive della memoria centrale

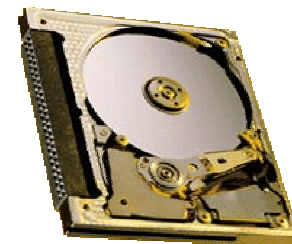
La memoria secondaria (di massa)

• Bibliografia:

Tanenbaum A. S., Goodman J. R.,
"Architettura dei computer - Un approccio strutturato",
Prentice Hall International, 2000 - paragrafo 2.3

• Sommario:

- Gerarchie di memoria.
- Dischi magnetici
- Dischi ottici



Caratteristiche dei diversi livelli

	Capacità	Velocità (TA)	€/MByte
registri	~1KB	~1ns	NA
cache	256 ÷ 1024 KB	~1ns	150-N/A
RAM	128 ÷ 2048 MB	~4ns	0.33
HD	60 ÷ 200 GB	~10ms	0.005
nastri/CD	~GB per unità	~100ms	0.005

Una gerarchia di memoria

E' tipicamente costituita da:

- **registri** contenuti nella CPU (qualche KB)
- **cache** (da circa 32KB a circa 1024KB)
- **memoria principale** (da circa 64MB a qualche GB)
- **dischi fissi** (da qualche GB a qualche TB)
- **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.**

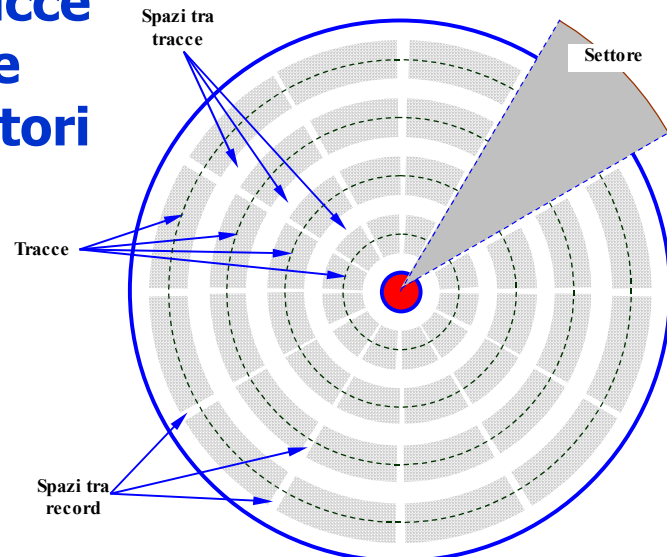
Prestazioni dei dischi

- **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

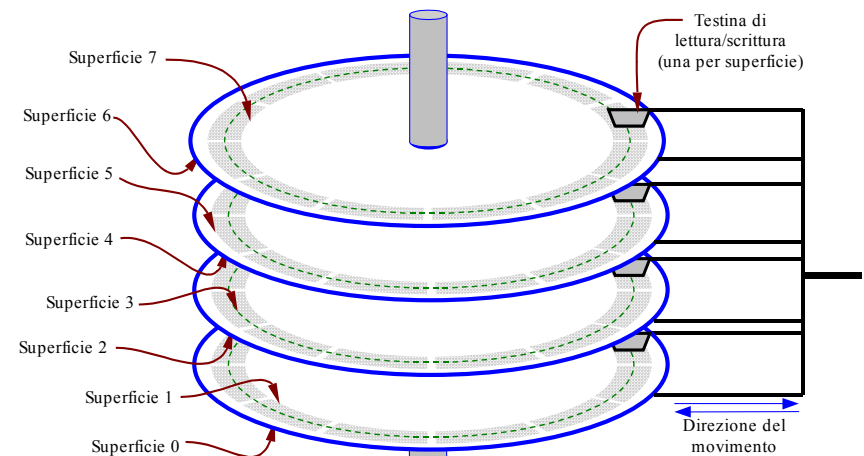
Tracce e settori

- **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 (5-10 μ m 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 0,1-0,2 μ m 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.

Tracce e settori



Schema di uno Hard Disk



Le tracce in grigio formano un "cilindro"

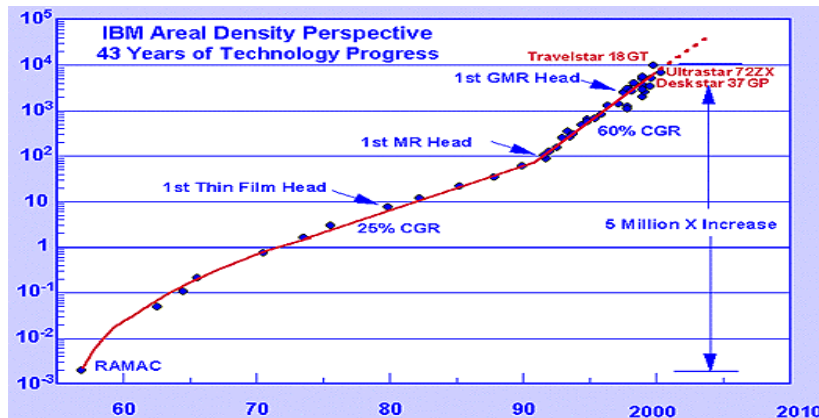
Dischi RAID

- RAID: Redundant Array of Inexpensive Disks.
- Vantaggi:
 - Utilizzo di più dischi per replicare le informazioni in modo ridondante, tenendole tutte aggiornate **contemporaneamente**;
 - Utilizzo di dischi di costo ridotto;
 - Nel caso di guasto (failure) di un hard disk, vengono utilizzate copie, presenti sugli altri hard disk, dei dati persi;
 - Differenti livelli di ridondanza e di sicurezza. Il livello più sicuro è RAID 5.

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**

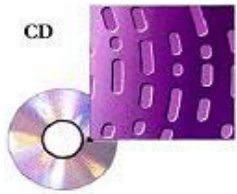
Andamento densità HD [by IBM]



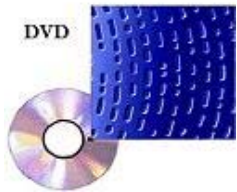
Compact Disk - CD

- 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:
 - laser ad alta potenza che brucia fori di 0,8 μm in un **disco master** (le depressioni si chiamano **pit** e le aree fra pit si chiamano **land**);
 - dal master si ricava uno **stampo**;
 - nello stampo viene iniettata una resina liquida di **policarbonato** che forma un CD con la stessa sequenza di fori del master,
 - sul policarbonato viene depositato uno strato molto sottile di **alluminio riflettente**,
 - copertura con uno strato **protettivo** e infine con **un'etichetta**.

Digital Versatile Disk (DVD)



- 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).

Nastri

- Nastri magnetici:
 - generalmente in cassetta, non più su bobine;
 - tecnologia DAT.
- Vantaggi:
 - costo ridotto;
 - utilizzati esclusivamente per il back-up;