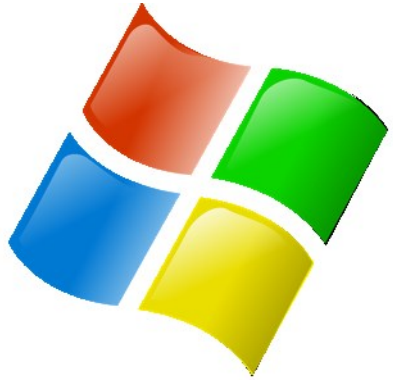
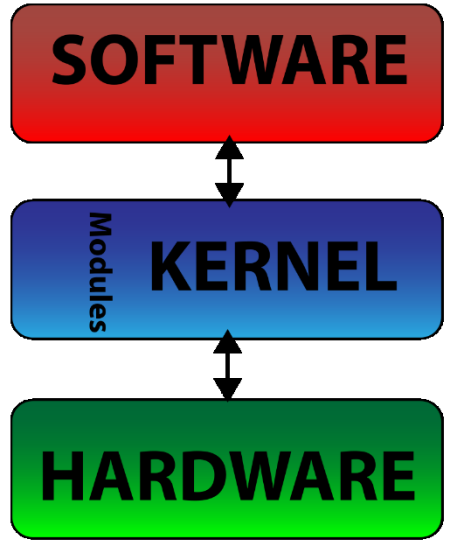
**UNIVERSITÀ DEGLI STUDI DELLA BASILICATA**



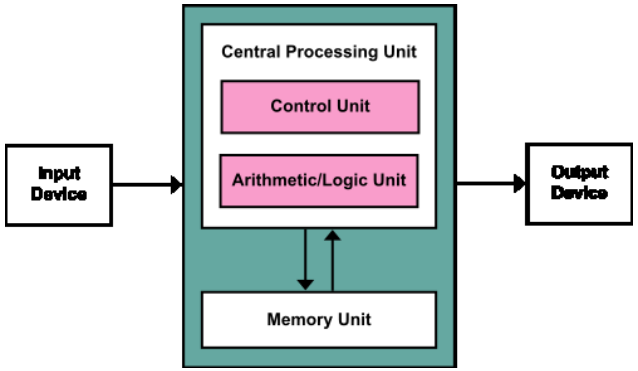
*Corso di Sistemi Operativi*
*A.A. 2019/20*

# Memoria di massa




Docente:
Domenico Daniele Bloisi

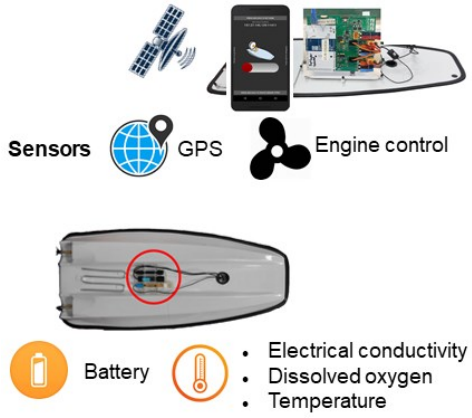
*Dicembre 2019*

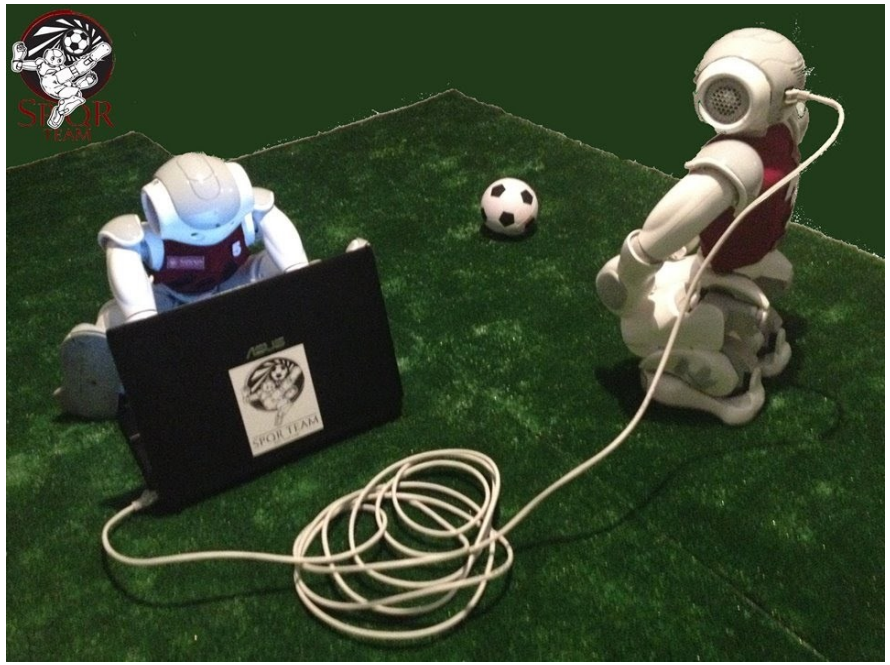



# Domenico Daniele Bloisi

- Ricercatore RTD B
  Dipartimento di Matematica, Informatica ed Economia
  Università degli studi della Basilicata
  http://web.unibas.it/bloisi

- SPQR Robot Soccer Team
  Dipartimento di Informatica, Automatica e Gestionale Università degli studi di Roma “La Sapienza”
  http://spqr.diag.uniroma1.it



# Ricevimento

- In aula, subito dopo le lezioni
- Martedì dalle 11:00 alle 13:00 presso:
  Campus di Macchia Romana
  Edificio 3D (Dipartimento di Matematica, Informatica ed Economia)
  II piano, stanza 15

Email: domenico.bloisi@unibas.it

# Programma – Sistemi Operativi

- Introduzione ai sistemi operativi
- Gestione dei processi
- Sincronizzazione dei processi
- Gestione della memoria centrale
- Gestione della memoria di massa
- File system
- Sicurezza e protezione

# Memorie di massa

I dischi magnetici e i dispositivi di memoria non volatile (nonvolatile memory, NVM) costituiscono i supporti fondamentali di memoria secondaria nei computer attuali

- Dischi magnetici (HDD - Hard Disk)
- Dischi a stato solido (SSD - Solid State Disk)

## Memorie terziarie

- Un solo drive molti dispositivi rimovibili
- Pen drive (flash)
- CD, DVD, Blu-ray

# Dischi magnetici

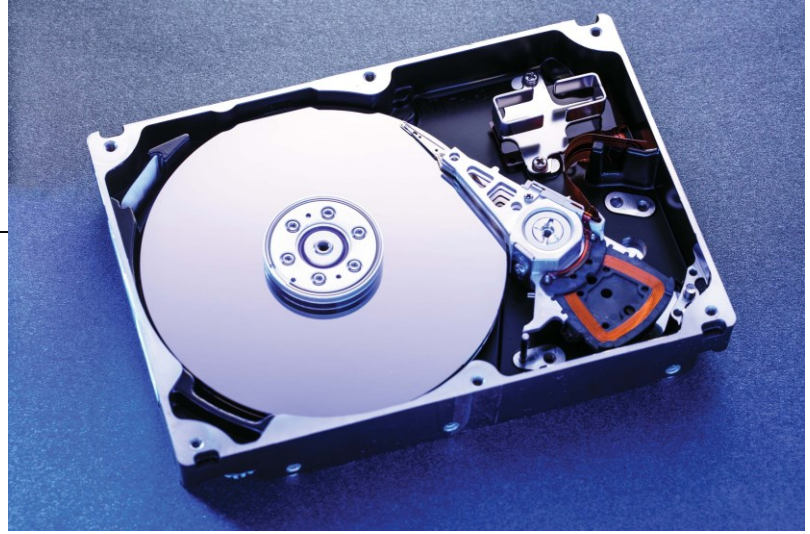
I dischi magnetici rappresentano ancora oggi un mezzo molto diffuso per la memorizzazione di massa

- Rivestiti con materiale magnetico (ossido di ferro), erano originariamente in alluminio
- La tecnologia attuale, viceversa, è orientata all'utilizzo del vetro:
  - Superficie più uniforme → maggiore affidabilità (errori di lettura/scrittura meno frequenti)
  - Più rigido e più resistente agli urti
  - Permette di ridurre la distanza della testina dalla superficie

# Dischi magnetici


https://youtu.be/kdmLvl1n82U

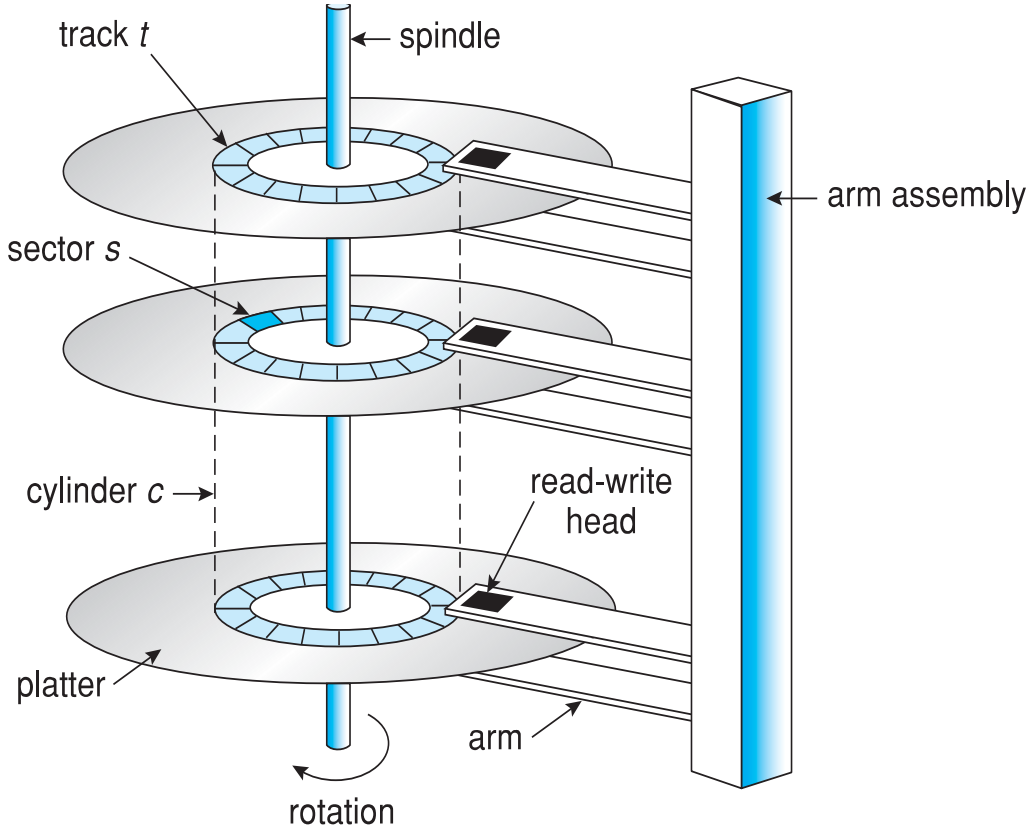
# Dischi magnetici



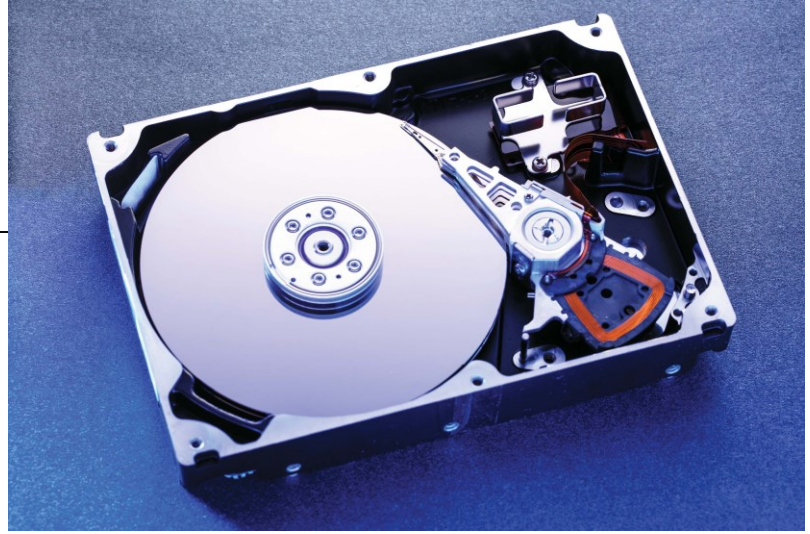
**Piatto (platter)**
un disco rigido si compone di uno o più dischi paralleli, in cui ogni superficie, detta “piatto” e identificata da un numero univoco, è destinata alla memorizzazione dei dati

**Traccia (track)**
su ogni piatto, vi sono numerosi anelli concentrici, detti tracce, ciascuno identificato da un numero univoco

**Cilindro (cylinder)**
l’insieme di tracce poste alla stessa distanza dal centro e relative a tutti i dischi; corrisponde a tutte le tracce con lo stesso numero, ma giacenti su piatti diversi



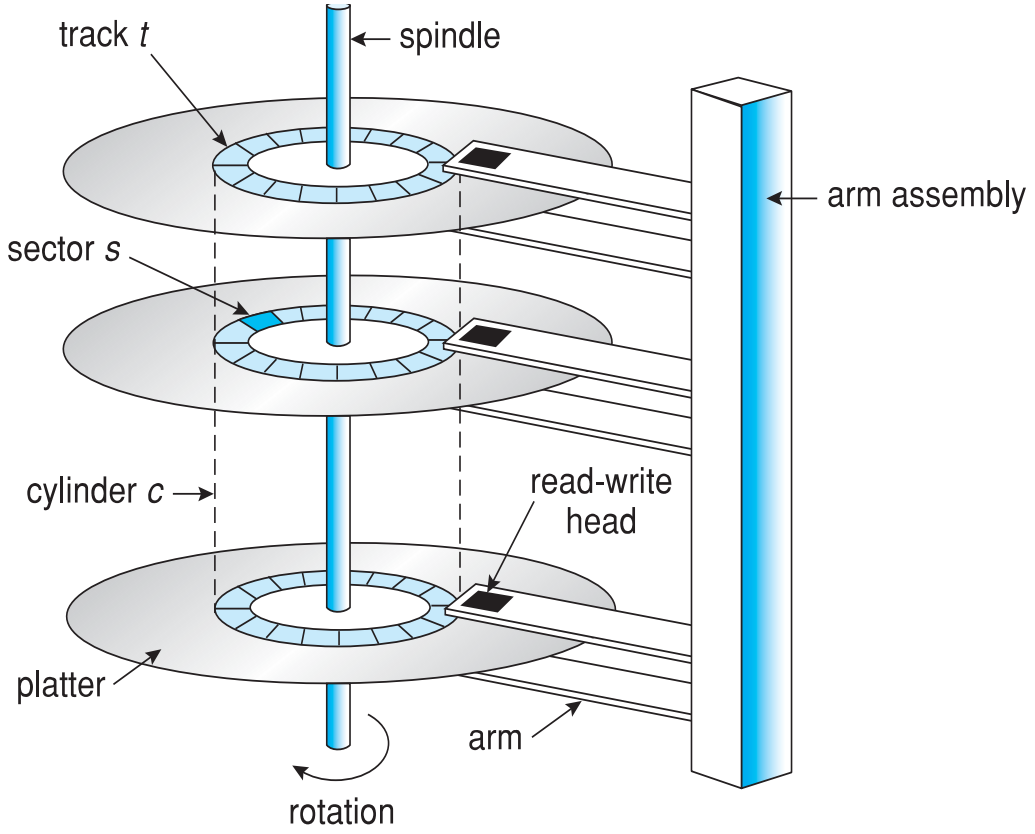
# Dischi magnetici



**Settore (sector)**
ogni traccia è suddivisa in settori, cioè in “spicchi” uguali, ciascuno identificato da un numero univoco

**Blocco (block)**
l’insieme dei settori posti nella stessa posizione in tutti i piatti

**Testina (read-write head)**
su ogni piatto è presente una testina di lettura/scrittura; la posizione di tale testina è solidale con tutte le altre sui diversi piatti: se una testina è posizionata sopra una traccia, tutte le testine saranno posizionate sul cilindro a cui la traccia appartiene



# Dischi magnetici – Tempi di accesso

- I dischi ruotano ad una velocità compresa tra i 60 e i 250 giri al secondo
- La velocità di trasferimento è la velocità con cui i dati fluiscono dall'unità a disco alla RAM
- Il tempo di posizionamento è il tempo necessario a spostare il braccio del disco in corrispondenza del cilindro desiderato (seek time), più il tempo necessario affinché il settore desiderato si porti sotto la testina (latenza di rotazione)
- Il crollo della testina, normalmente sospesa su un cuscinetto d'aria di pochi micron, corrisponde all'impatto della stessa sulla superficie del disco - di solito comporta la necessità di sostituire l'unità a disco
- I dischi possono essere rimovibili

# Dischi magnetici – Lettura/scrittura

Memorizzazione e recupero dell'informazione tramite bobina conduttiva detta testina (head)

- Durante la lettura/scrittura, la testina è stazionaria, mentre il disco ruota
- Scrittura
  - la corrente, che fluisce nella bobina (nelle due possibili direzioni) produce un campo magnetico
  - le particelle aciculari dell'ossido di ferro si orientano in base al campo magnetico prodotto (0 e 1 memorizzati su disco)
- Lettura
  - Il campo magnetico presente sul disco, muovendosi rispetto alla testina, induce corrente nella bobina

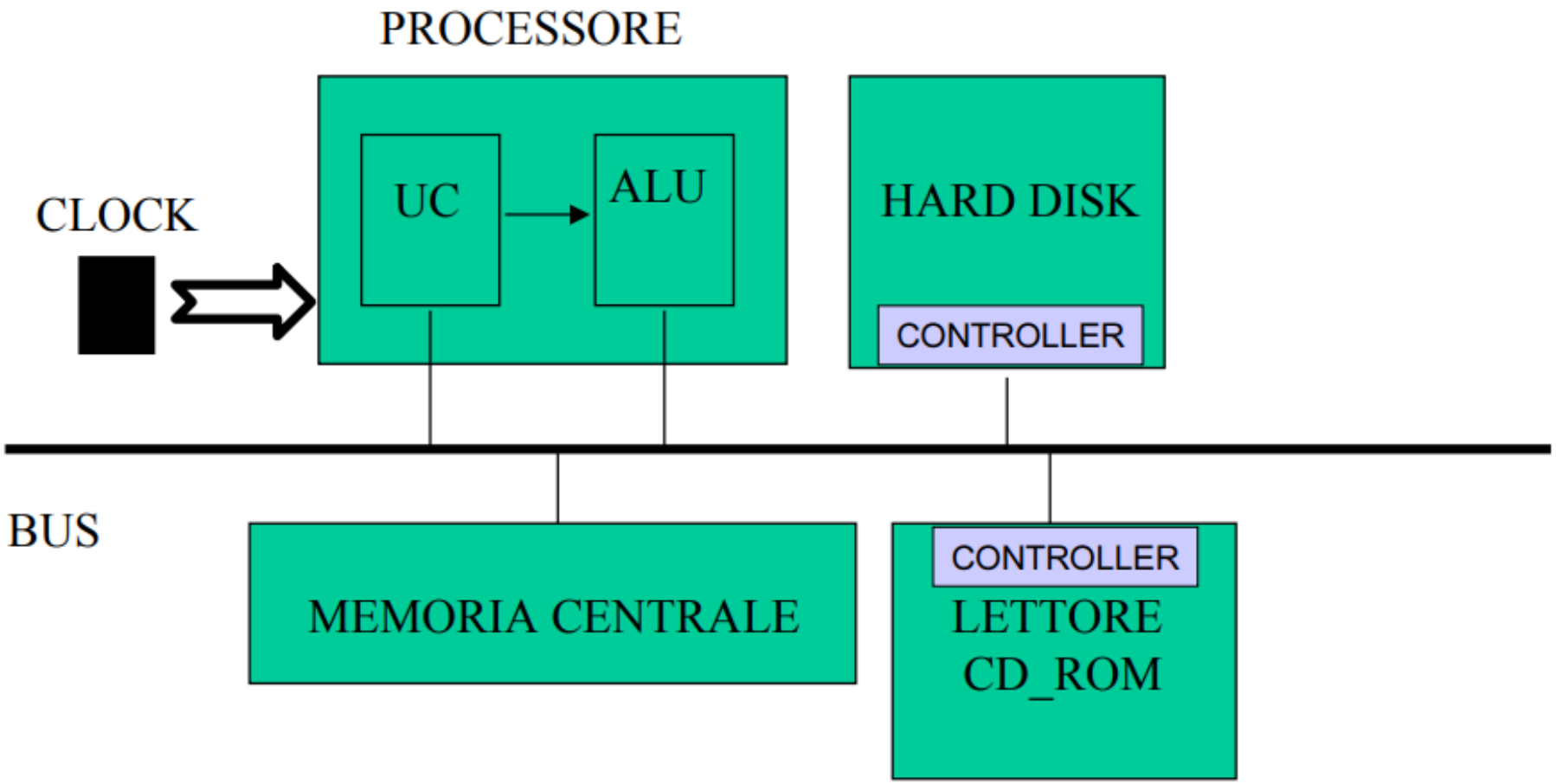
# Dischi magnetici – Connessione

L’unità a disco è connessa al calcolatore per mezzo del bus di I/O

- Diversi tipi: ATA (Advanced Technology Attachment), SATA (Serial ATA), USB (Universal Serial Bus), Fiber Channel, SCSI (Small Computer System Interface)

- Il trasferimento di dati in un bus è eseguito da speciali unità di elaborazione, dette controllori: gli adattatori sono i controllori posti all’estremità del bus relativa al calcolatore, i controllori dei dischi sono incorporati in ciascuna unità a disco

# Dischi magnetici – Connessione




http://www.di.unito.it/~rossano/DIDATTICA/suism-0203/lezione7.pdf

# Dischi magnetici – Controller


https://commons.wikimedia.org/wiki/File:Serial_ATA_hard_disk_connected.jpg

# Dischi magnetici – I/O

- Per eseguire un'operazione di I/O, si inserisce il comando opportuno nell'adattatore, generalmente mediante porte di I/O mappate in memoria
- L'adattatore invia il comando al controllore del disco, che agisce sugli elementi elettromeccanici dell'unita per portare a termine il compito richiesto
- Il trasferimento dei dati nell'unità a disco avviene tra la superficie del disco e la cache incorporata nel controllore
- Il trasferimento dei dati tra la cache e l'adattatore avviene alla velocità propria dei dispositivi elettronici

# Dischi magnetici – Caratteristiche

- Il raggio dei piatti variava, storicamente, fra 14 e 85 pollici
- I formati attualmente più comuni sono 3.5”, 2.5”, e 1.8”
- La capacità attuale dei dischi si attesta fra 30GB e 15TB

- Performance
  - Velocità di trasferimento (teorica): 6Gb/sec
  - Velocità di trasferimento (effettiva): 1Gb/sec
  - Seek time compreso fra 3msec e 12msec (9msec in media per i dischi presenti nei PC)
  - Tempo di latenza calcolato in base alla velocità di rotazione
    1 / (RPM / 60) = 60 / RPM
  - Latenza media = ½ giro

# Dischi magnetici – Tempi di accesso

- Tempo di accesso medio = seek time medio + latenza media

  Per i dischi più veloci → 3msec + 2msec = 5msec
  Per dischi lenti → 9msec + 5.55msec = 14.55msec

- Tempo medio di I/O = tempo medio di accesso + (quantità di dati da trasferire / velocità di trasferimento) + overhead

# Dischi magnetici – Tempi di accesso

Per esempio, per trasferire un blocco da 4KB su un disco con una velocità di rotazione pari a 7200 RPM, tempo medio di ricerca pari a 5msec, velocità di trasferimento di 1Gb/sec e con un overhead dovuto al controllore di 0.1msec, si ottiene:

Tempo di trasferimento = 4KB/1Gb/s = 0.031 msec

Tempo medio di I/O per un blocco da 4KB =
= 5msec + 4.16msec + 0.1msec + tempo di trasferimento
= 9.27msec + 0.031msec = 9.301msec

# Il primo HD commerciale

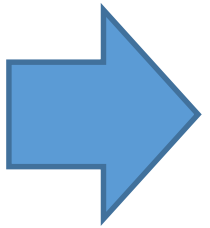
1956

Il computer IBM 305 RAMAC includeva il primo disco magnetico nella storia dei calcolatori

- 5 milioni di caratteri da 7 bit più parità
- 50 dischi da 24”
- Tempo di accesso circa 600 msec

# Dispositivi NVM

- Spesso inseriti in chassis simili agli HDD e perciò denominati dischi a stato solido (SSD)
- Altre forme includono unità USB (pen drive, unità flash) e DRAM dotate di batteria di backup
- Negli smartphone, le NVM sono montate sulla scheda madre e rappresentano il dispositivo primario di archiviazione
- Le NVM possono essere più affidabili degli HDD
- Hanno costo al MB più elevato
- Possono avere vita più breve



# Dispositivi NVM

- Hanno minore capacità, ma sono molto più veloci e consumano meno energia
- Nessuna parte meccanica in movimento, quindi nessun tempo di ricerca o latenza di rotazione e maggiore resistenza a sollecitazioni e urti (minor rumore e minore dispersione termica)
- I bus standard possono essere troppo lenti
- Collegamento al bus PCI di sistema con tecnologia NVM express

# Dispositivi NVM

Le caratteristiche dei semiconduttori NAND aprono a nuove sfide per l'affidabilità

Le NAND si deteriorano ad ogni ciclo di cancellazione e dopo circa 100.000 cicli le celle non sono più in grado di mantenere l'informazione

- Durata misurata in numero di scritture al giorno (DWPD, Drive Writes per Day)
- Su una NAND da 1 TB di classe 5DWPD si possono scrivere teoricamente 5 TB al giorno senza errori

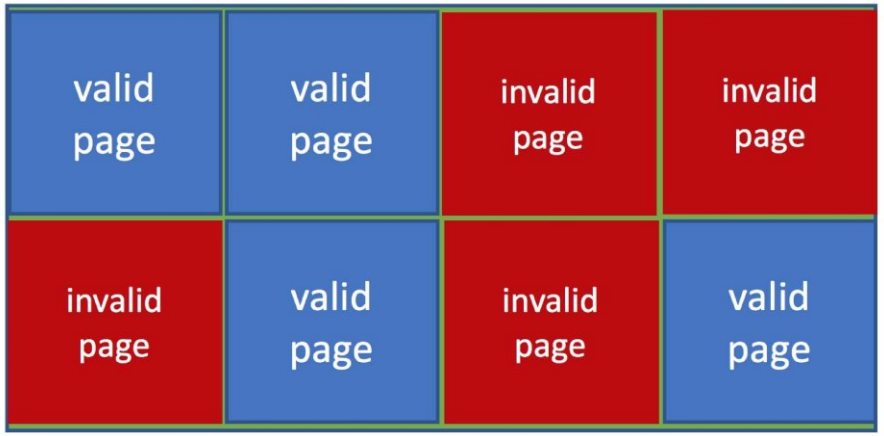
# Dispositivi NVM

Letture e scritture avvengono con granularità di “pagina” (analogo del settore)

- Impossibilità di cancellazione per “sovra-scrittura”
- Il contenuto della pagina deve prima essere cancellato e le cancellazioni avvengono per “blocchi” (della dimensione di diverse pagine) → operazione costosa

# Dispositivi NVM

Senza sovrascrittura, i blocchi sono costituiti da un mix di pagine valide e non valide




- Per tenere traccia dei blocchi logici validi, il controllore mantiene la tabella FTL (Flash Translation Layer)

- Implementa anche la garbage collection per liberare spazio

# Garbage collection

- Per cancellare dati non validi, un controller SSD di norma deve prima copiare tutti i dati validi (quelli che dovranno essere ancora utilizzati in futuro) nelle pagine vuote di un altro blocco

- Quindi deve cancellare tutte le celle del blocco da liberare (eliminando sia i dati da cancellare che quelli nel frattempo copiati per poter essere riutilizzati) e solo a quel punto può iniziare a scrivere nuovi dati nel blocco che è stato così liberato

# Overprovisioning

- In alternativa, assegna un overprovisioning (7-20%) per fornire spazio di lavoro per la garbage collection
- Ogni cella ha durata di vita limitata, quindi il livello di usura deve essere mantenuto uniforme
- L'overprovisioning è anche funzionale a garantire un numero adeguato di celle sostituibili a quelle che raggiungono il limite del ciclo di programmazione e cancellazione, così da prolungare la vita dello stesso SSD

# Memoria volatile

- DRAM usata frequentemente come dispositivo di archiviazione di massa
  - Tecnicamente, non si può definire archiviazione secondaria perché volatile, ma può contenere file system, da utilizzare come storage secondario molto veloce
- Infatti, le unità RAM possono essere utilizzate come dispositivi a blocchi non formattati, ma più spesso contengono un file system
  - Supportate dai principali sistemi operativi
    Linux: **`/dev/ram`**, **`/tmp`** (con file system temporaneo)
    MAC OS: **`diskutil`**

# Blocchi logici

- Le unità a disco vengono indirizzate come giganteschi vettori monodimensionali di blocchi logici, dove il blocco logico rappresenta la minima unità di trasferimento

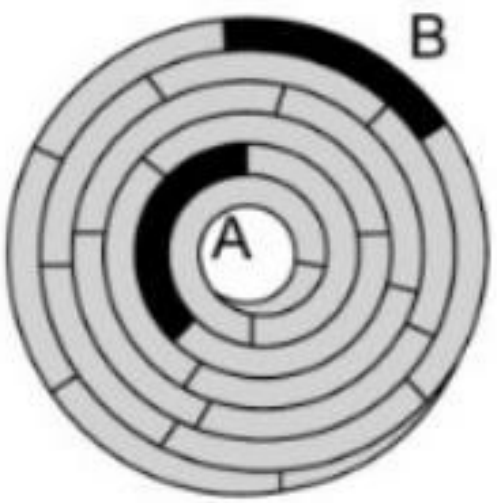
- I blocchi logici sono creati all'atto della formattazione di basso livello

# Mappatura degli indirizzi

CLV (Constant Linear Velocity): densità dei bit per traccia uniforme

- Tracce più lontane dal centro del disco sono più lunghe e contengono un maggior numero di settori (fino al 40% in più rispetto alle tracce vicine al centro di rotazione)
- La velocità di rotazione aumenta spostandosi verso l'interno ($v = \omega r$), per mantenere costante la velocità lineare e, quindi, la quantità di dati che passano sotto le testine nell'unità di tempo
- CD e DVD
- Talvolta, si ha un'unica traccia a spirale

Disco CLV con un'unica traccia a spirale

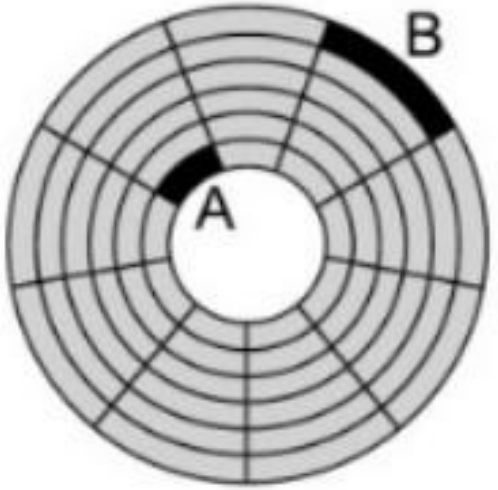


# Mappatura degli indirizzi

CAV (Constant Angular Velocity): velocità di rotazione costante

- La densità dei bit decresce dalle tracce interne alle più esterne per mantenere costante la quantità di dati che passano sotto le testine nell'unità di tempo
- Dischi magnetici

Disco CAV a tracce concentriche



# Scheduling del disco

Il SO è responsabile dell'uso efficiente dell'hardware: per i dischi ciò significa garantire tempi di accesso contenuti e ampiezze di banda elevate

Il tempo di accesso al disco si può scindere in due componenti principali:

- Tempo di ricerca (seek time) → è il tempo impiegato per spostare la testina sul cilindro che contiene il settore desiderato
- Latenza di rotazione (rotational latency) → è il tempo necessario perché il disco ruoti fino a portare il settore desiderato sotto la testina

# Scheduling del disco

Per migliorare le prestazioni si può intervenire solo sul tempo di ricerca e si tenta quindi di minimizzarlo

Il seek time è proporzionale alla distanza di spostamento fra le tracce

Seek time ≈ seek distance

# Bandwidth

L’ampiezza di banda (bandwidth) del disco è il numero totale di byte trasferiti, diviso per il tempo trascorso fra la prima richiesta e il completamento dell’ultimo trasferimento

# Operazioni di I/O su disco

Quando un processo (utente o di sistema) deve effettuare un'operazione di I/O relativa ad un'unità a disco, effettua una chiamata al SO

La richiesta di servizio contiene:

- Specifica del tipo di operazione (immissione/emissione di dati)
- Indirizzo su disco relativamente al quale effettuare il trasferimento
- Indirizzo nella memoria relativamente al quale effettuare il trasferimento
- Numero di byte da trasferire

# Algoritmi di scheduling del disco

Una richiesta di accesso al disco può venire soddisfatta immediatamente se unità a disco e controller sono disponibili; altrimenti la richiesta deve essere aggiunta alla coda delle richieste inevase per quella unità

Esistono diversi algoritmi di scheduling del disco per gestire la coda di richieste alla memoria secondaria

# Algoritmo di scheduling del disco FCFS

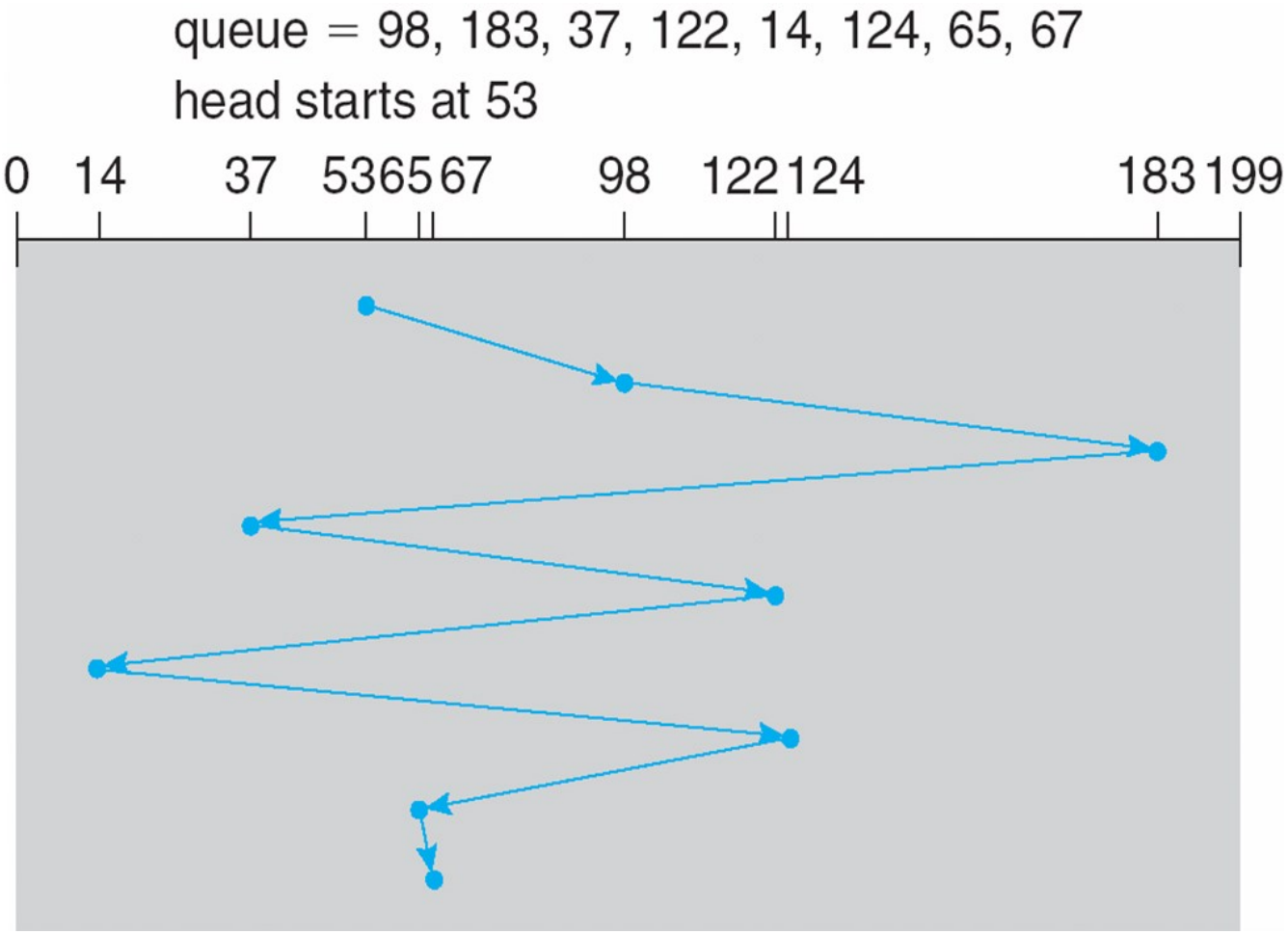
FCFS → First Come First Served

Lista di richieste da esaudire:
98, 183, 37, 122, 14, 124, 65, 67
Testina correntemente sul cilindro 53, cilindri totali 200 (0-199)

La distanza (in cilindri) da coprire per la testina sarà
$|53 – 98| + |98 – 183| + |183 – 37| + |37 – 122| +$
$|122 – 14| + |14 – 124| + |124 – 65| + |65 – 67| =$
$45+85+96+35+108+110+59+2 = 640$

# Algoritmo di scheduling del disco FCFS

FCFS → First Come First Served




Viene generato un movimento totale della testina pari a 640 cilindri

# Algoritmo di scheduling del disco SCAN

Il braccio della testina si muove da un estremo all'altro del disco, servendo sequenzialmente le richieste; giunto ad un estremo, inverte la direzione di marcia e, conseguentemente, l'ordine di servizio.

È chiamato anche algoritmo dell'ascensore

# Algoritmo di scheduling del disco SCAN

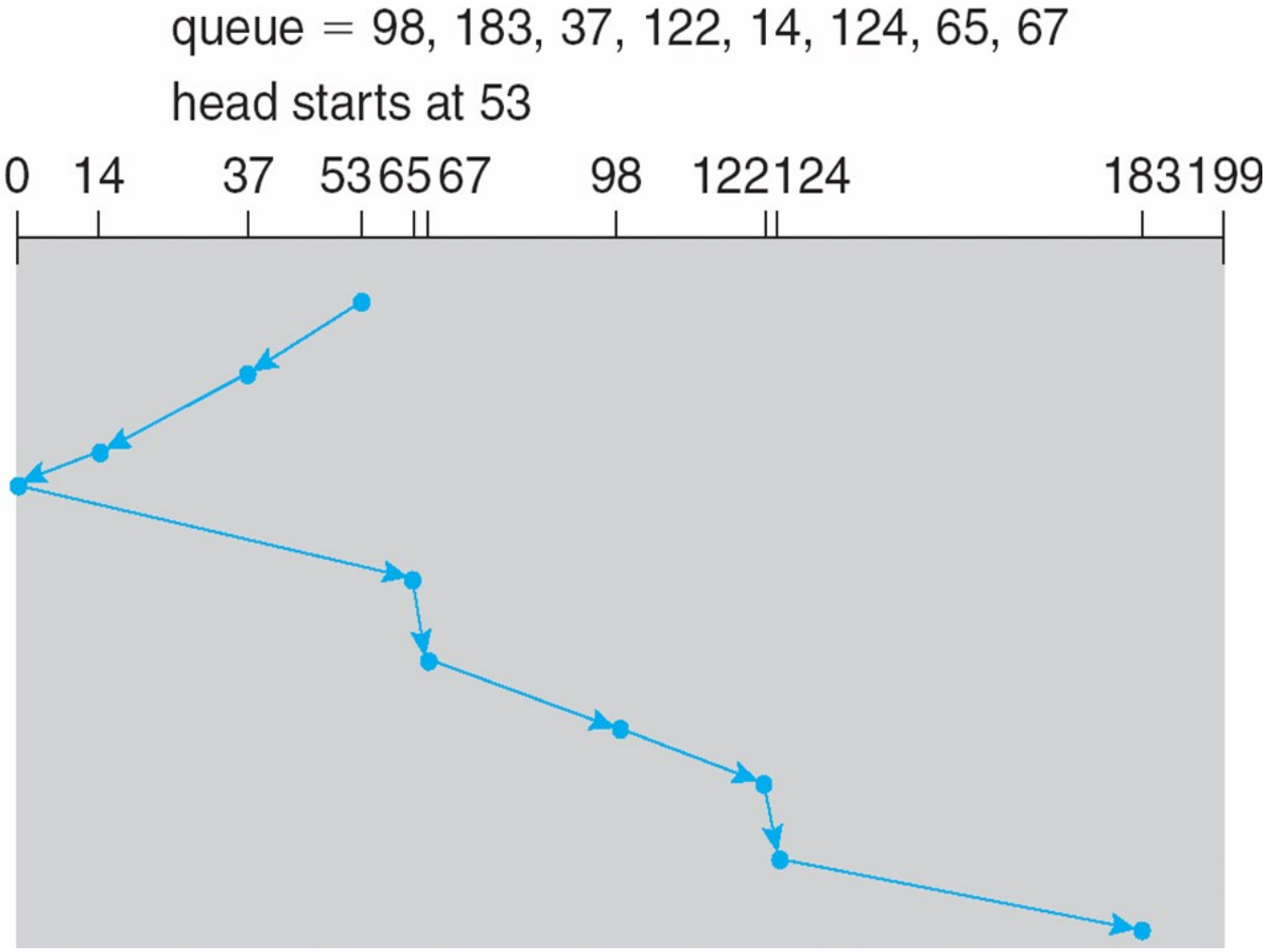
Lista di richieste da esaudire:
98, 183, 37, 122, 14, 124, 65, 67
Testina correntemente sul cilindro 53, cilindri totali 200 (0-199)

Scheduling SCAN (supponendo movimento verso il cilindro 0)
37, 14, 0, 65, 67, 98, 122, 124, 183

La distanza (in cilindri) da coprire per la testina sarà
$|53 - 37| + |37 - 14| + |14 - 0| + |0 - 65| + |65 - 67| +$
$|67 - 98| + |98 - 122| + |122 - 124| + |124 - 183| =$
$16+23+14+65+2+31+24+2+59 = 236$

# Algoritmo di scheduling del disco SCAN

SCAN



# Algoritmo di scheduling del disco SCAN

Se gli accessi sono distribuiti uniformemente, quando la testina inverte il proprio movimento, la maggior densità di richieste si ha all'estremo opposto del disco

Tali richieste avranno anche i tempi più lunghi di attesa di servizio

# Algoritmo di scheduling del disco C-SCAN

La testina si muove da un estremo all'altro del disco servendo sequenzialmente le richieste

Quando raggiunge l'ultimo cilindro ritorna immediatamente all'inizio del disco, senza servire richieste durante il viaggio di ritorno

Considera i cilindri come organizzati secondo una lista circolare, con l'ultimo cilindro adiacente al primo

C-SCAN garantisce un tempo di attesa più uniforme rispetto a SCAN

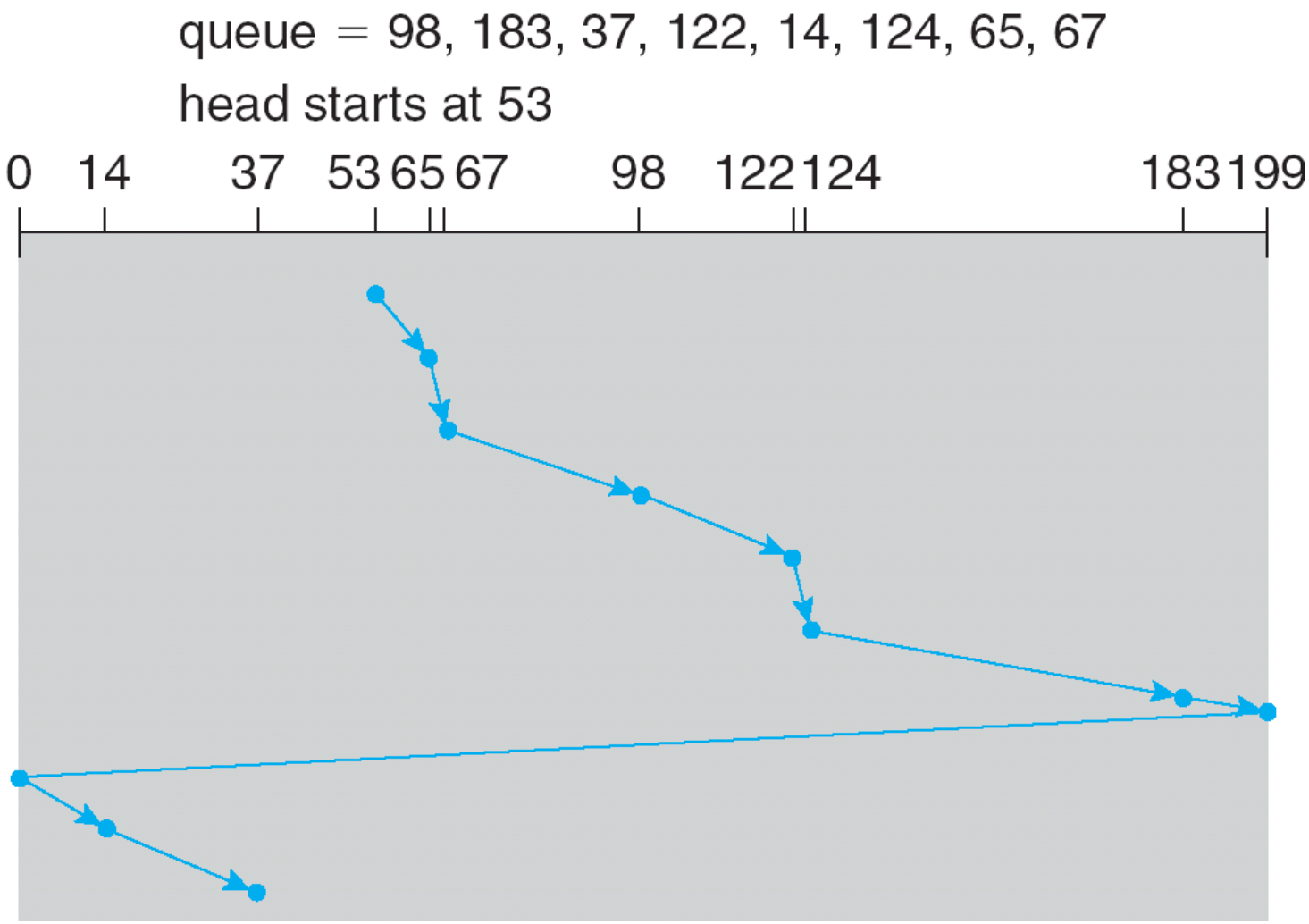
# Algoritmo di scheduling del disco C-SCAN

Lista di richieste da esaudire:
98, 183, 37, 122, 14, 124, 65, 67
Testina correntemente sul cilindro 53, cilindri totali 200 (0-199)

Scheduling C-SCAN (supponendo movimento verso il cilindro 199)
65, 67, 98, 122, 124, 183, 199, 0, 14, 37

# Algoritmo di scheduling del disco C-SCAN

C-SCAN



# Scelta di un algoritmo di scheduling del disco

SCAN e C-SCAN forniscono buone prestazioni in sistemi che utilizzano intensamente le unità a disco

Le prestazioni dipendono comunque dal numero e dal tipo di richieste

Le richieste di I/O per l'unità a disco possono essere influenzate dal metodo di allocazione di file e directory

Nelle unità NVM, dove non esistono parti mobili, si utilizza di solito una politica FCFS

L'unica ottimizzazione possibile riguarda il servizio combinato di richieste (di lettura) relative a indirizzi logici adiacenti

# Rilevamento e correzione di errori

- Il rilevamento degli errori determina se si è verificato un problema (ad esempio un bit flipping)
- A fronte del verificarsi di un errore, il sistema può interrompere l'operazione prima che l'errore venga propagato
- Rilevazione eseguita frequentemente tramite bit di parità
- La parità è una forma di checksum che utilizza l'aritmetica modulare per calcolare, archiviare, confrontare valori su parole a lunghezza fissa

# Rilevamento e correzione di errori

- Un altro metodo di rilevamento degli errori comune nelle reti è il controllo di ridondanza ciclica (CRC) che utilizza una funzione hash per rilevare errori su più bit

- Il codice di correzione degli errori (ECC) non solo rileva, ma può correggere alcuni errori
  → Errori soft correggibili, errori hard rilevati ma non corretti

# Formattazione

Formattazione di basso livello o fisica → Si suddivide il disco in settori che possono essere letti e scritti dal controllore del disco

Salvataggio su disco di una struttura dati per ogni settore (intestazione/coda/ECC)
→Dimensione standard pari a 512 byte

# Partizionamento

- Per poter impiegare un disco per memorizzare i file, il SO deve mantenere le proprie strutture dati sul disco

- Si partiziona il disco in uno o più gruppi di cilindri, ognuno dei quali rappresenta un “disco logico”

- Formattazione logica o “creazione di un file system”

- Per migliorare le prestazioni, la maggior parte dei file system accorpa i blocchi in gruppi, detti cluster
  - I/O su disco fatto per blocchi
  - I/O via file system fatto per cluster

# Partizione di boot

La partizione di boot contiene il SO; altre partizioni possono contenere altri SO, altri file system o essere partizioni raw

- Viene montata all'avvio del sistema
- Altre partizioni possono essere montate automaticamente o manualmente (al boot o successivamente)

Al momento del montaggio, si verifica la coerenza del file system (controllando la correttezza dei metadati)

- Si aggiorna la tabella di montaggio
- Il blocco di avvio può puntare al volume di avvio o all'insieme di blocchi contenenti il caricatore di avvio, ovvero codice sufficiente per caricare il kernel dal file system

# Partizione di boot

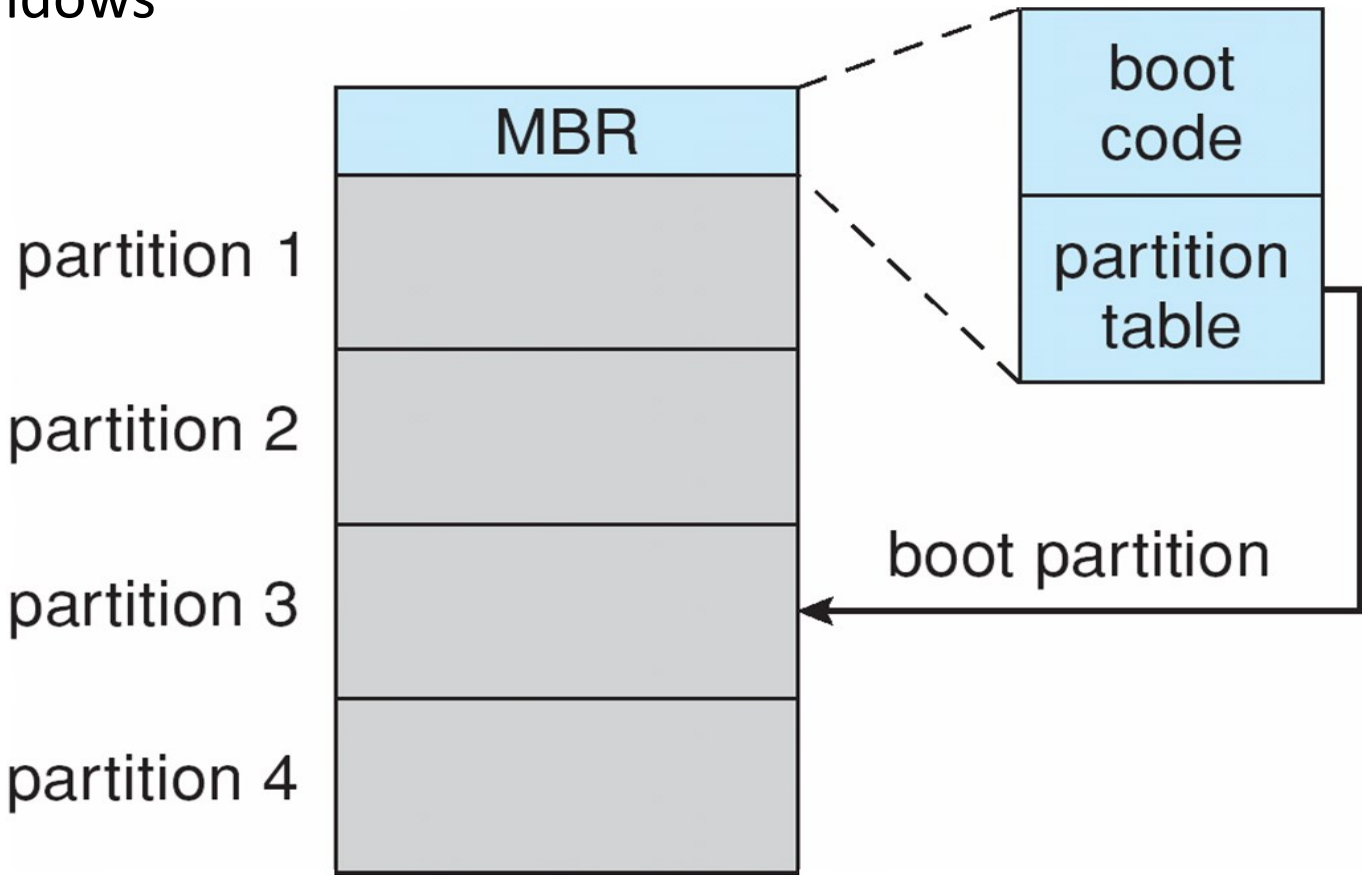
Nel boot block sono contenute le informazioni necessarie all'inizializzazione del sistema

Il bootstrap loader è memorizzato nella ROM

Il bootstrap completo è memorizzato in posizione fissa nell'hard disk → per esempio, nel primo settore del disco di avviamento (o disco di sistema)

# Boot in Windows

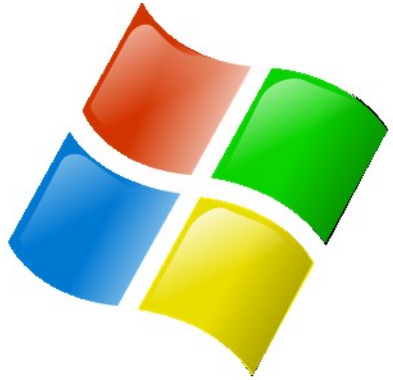
Booting from secondary storage in Windows



# Accantonamento dei settori

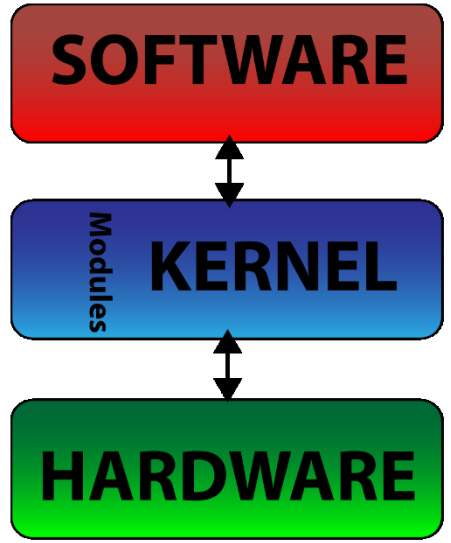
- Si impiega l'accantonamento dei settori come modalità di gestione dei blocchi difettosi
- Durante la formattazione fisica si mantiene un gruppo di settori di riserva non visibili al SO
- Il controllore "è istruito" per sostituire, dal punto di vista logico, un settore difettoso con uno dei settori di riserva inutilizzati

**UNIVERSITÀ DEGLI STUDI DELLA BASILICATA**



*Corso di Sistemi Operativi*
*A.A. 2019/20*

# Memoria di massa




Docente:
Domenico Daniele Bloisi

*Dicembre 2019*