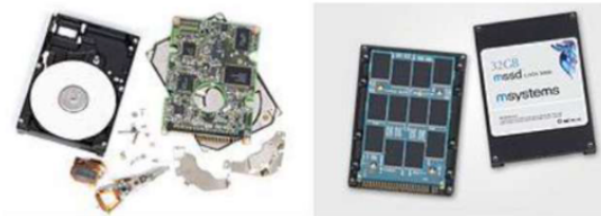


ARCHITETTURE DI CALCOLO LEZIONE 3

Linguaggi, livelli e macchine virtuali

Dischi SSD

Si tratta di dispositivi completamente elettronici, normalmente basati su memorie flash (NAND), senza nessuna parte in movimento. I dischi magnetici, per la brevissima distanza tra testina e disco, presentano il rischio di crash o caduta della testina, assente nelle schede SSD, risultando più fragili. I pro dei dischi SSD sono diversi:



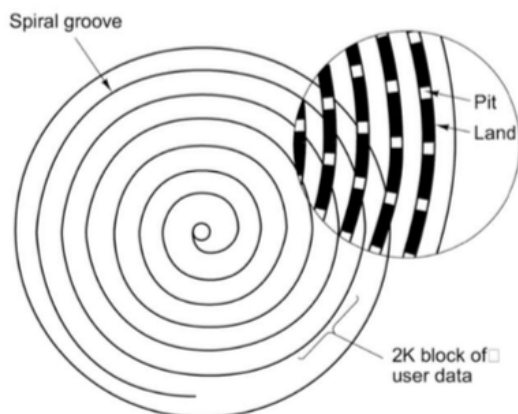
- Tempo di accesso (seek) ridotto
- Maggiore velocità di trasferimento dati (da 500 MB/sec a oltre 7 GB/sec)
- Minore possibilità di rottura e maggiore durata (rispetto agli hard disk)
- Rumorosità assente e minore produzione di calore
- Minori consumi durante le operazioni di lettura e scrittura o Maggiore capacità

Ovviamente sono individuabili anche dei contro:

- Maggiore prezzo per bit (4 volte il costo di un disco rigido tradizionale)
- Possibile minore durata dell'unità se usata per frequenti scritture

Qualsiasi tipo di dispositivo non può essere usato in modo illimitato: superato un valore soglia (statisticamente ogni cella può essere scritta 1 milione di volte) di scrittura/lettura del dispositivo, esso smetterà di funzionare. I dischi magnetici presentano da questo punto di vista una durata maggiore delle schede SSD; per questa ragione in caso di processi che prevedono un maggiore numero di cicli si preferiscono i primi.

Ritornando alla piramide della memoria (vedere sbobina precedente) osserviamo le cassette, oggi completamente in disuso. Erano usate per vecchi film e video.



Più moderni sono i CD (Compact Disc) o dischi ottici, che nascono con lo scopo di sostituire i dischi in vinile (oggi in realtà si vende di più il vinile perché il suono è più corposo/autentico rispetto a quello emesso dai CD, nel disco ottico abbiamo un suono campionato tramite un campionatore che viene poi quantizzato passando da un valore analogico a un valore digitale).

I CD sono nati negli anni 80 come supporto per la memorizzazione di musica.

Tale dispositivo si basa sulla codifica per mezzo di fori detti Pit alternati con zone piane dette Land lungo un'unica spirale. Un passaggio Pit-Land o Land-Pit codifica un 1, mentre l'assenza di variazioni codifica uno 0. Il tutto è letto tramite un raggio laser che viene riflesso diversamente al passaggio su pit e Land.

I CD vengono prodotti utilizzando uno stampo (ottenuto tramite erosione laser) su cui viene iniettata resina liquida che preserva incavi corrispondenti ai pit. Osservando il "tragitto" del laser lungo la spirale si osserva una variazione della circolazione; tuttavia, la velocità di lettura deve essere costante per riprodurre il suono nel modo corretto. Per risolvere tale problematica il CD ruota con velocità angolare variabile al fine di mantenere costante la velocità lineare. Successivamente ai CD troviamo: CD-ROM e CD-R.

I CD-ROM (Compact Disc-Read Only Memory) utilizzano la tecnologia dei CD per memorizzare dati informatici (a differenza degli altri CD che nascono come "contenitori" di musica). Nascono come dispositivi di sola memoria, cioè una volta prodotti contengono solo programmi per la lettura.

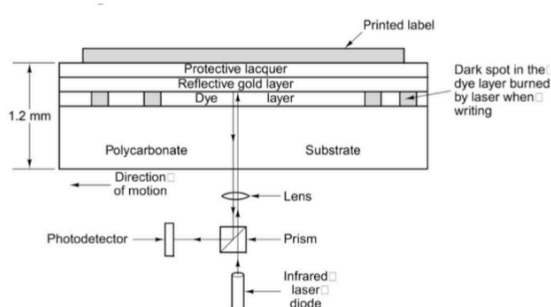
Al fine di evitare la perdita di bit durante la lettura, ogni byte viene codificato da un simbolo di 14 bit; nei bit in eccesso viene inserito un codice per la correzione dell'errore.

La differenza con gli altri CD risiede nel fatto che, mentre nel CD le tracce musicali sono sequenze di bit, nel CD-ROM abbiamo gruppi di bit così come troviamo le tracce o settori musicali nei dischi digitali moderni. Un gruppo di 42 simboli viene detto frame, mentre un gruppo di 98 frame viene denominato settore di CD-ROM, che si dispiega lungo la spirale. Il file system è un modo per organizzare dal punto di vista logico i file su un disco. Questo vale anche per i sistemi operativi (li vedremo nella seconda parte del corso di sistemi informativi).

Per motivi di compatibilità tra le diverse piattaforme, i CD-ROM utilizzano un proprio file system denominato High Sierra.

I CD-R (Compact Disc-Recordables), nati a metà anni 90 svolgono le stesse funzioni dei CD-ROM ma sono registrabili (tramite masterizzatore) dagli utenti senza l'utilizzo dello stampo.

I masterizzatori, che permettono di registrare i CD a casa, contengono due tipi di laser:



- Alta potenza: permette di "bruciare" il disco e quindi scriverlo
- Bassa potenza: è il laser di lettura
Osserviamo la presenza nella struttura del CD-R di:
- Strato protettivo
- Strato riflettente di oro
- Dye layer
- Strato plastico in policarbonato

Il raggio laser di scrittura va fino allo strato di policarbonato permettendo di scrivere sul dye layer, mentre quello di lettura si ferma allo strato in oro che lo riflette.

I CD-R avevano una memoria sufficiente (700 MByte) per contenere musica e film, ma alla fine degli anni 90 si sentì la necessità di aumentare lo spazio (sia per i film non compressi che per file vari) e nascono i DVD (Digital Video Disk o Digital Versatile Disk).

Nonostante le dimensioni siano molto simili a quelle di un CD, la tecnologia che li produce permette di creare spirali più serrate, contenenti pit più piccoli che vengono letti da un raggio

Poiché lo spazio a disposizione non è mai sufficiente (e per altri motivi legati al business del mercato cinematografico) sono stati introdotti 4 formati:

Formato	Capacità (GB)
Lato unico-Strato unico	4,7
Lato unico-Strato doppio	8,5
Lato doppio-Strato unico	9,5
Lato doppio-Strato doppio	17

laser rosso. A ciò si aggiunge la possibilità di scrivere su entrambi i lati del DVD, con conseguente raddoppio della memoria. Per eseguire questa procedura, si usavano due tipi di strati con sensibilità specifica per laser diversi così da prevenire la “bruciatura” di entrambi i lati contemporaneamente.

Si possono campionare i singoli strumenti di una canzone per raggiungere una qualità migliore della musica, tuttavia generalmente si tende a campionarli insieme.

Schede grafiche

Le schede grafiche dette anche schede video hanno il compito di tradurre la rappresentazione dell'immagine prodotta dal processore in un formato visualizzabile dal monitor.

Ogni volta che il video deve essere ridisegnato la scheda video legge la rappresentazione digitale dell'immagine dalla memoria e invia al monitor il segnale che permette di rappresentare il colore di ogni singolo pixel.

Fino a circa 20 anni fa la maggior parte dei monitor erano analogici, cioè caratterizzati dalla presenza di un tubo catodico (come anche la tv di casa fino a 15 anni fa). Ciò implica che la CPU, cui segnale prodotto è digitale, dovrà tradurre il suo segnale digitale in un segnale analogico da inviare al monitor. A questo scopo, le schede video presentano un'uscita detta VGA che si collega alla RAM-DAC (RAM Digital-Analog Converter) che ha il compito di convertire il segnale ricevuto dalla CPU in analogico. Oltre all'uscita VGA oggi si trova quella digitale come DVI (Digital Visual Interface), che non trasforma il dato in natura (analogico → digitale) ma in risoluzione (da una certa dimensione elaborata dalla CPU ad una compatibile con il monitor).

Oggi quasi tutti i monitor sono LCD e prevedono input digitale. Pertanto il ruolo primario della scheda grafica non è più quello di tradurre il segnale in formato analogico, ma di:

- Adattare l'immagine adeguando colori e risoluzione (attenzione al blurring!);
- Mettere a disposizione una memoria locale da utilizzare come buffer;
- Supportare l'accelerazione grafica.

La scheda grafica risulta quindi utile nel processo di ridimensionamento delle immagini tra dispositivi con standard di visualizzazioni diversi.

Posso collegare un cavo VGA al mio computer (digitale) ma eseguo un'inutile traduzione da digitale ad analogico che deteriora l'immagine.

Oggi i formati grafici più usati sono:

- DVI (Digital Visual Interface): ormai abbastanza vecchio
- HDMI (High-Definition Multimedia Interface): ha la capacità di trasportare sia video che audio
- DP (DisplayPort): principalmente per trasportare una sorgente video digitale a un monitor ma può trasportare anche audio e altri tipi di dati.

La scheda video, soprattutto negli ultimi anni, è diventata autonoma dalla CPU in quanto passare dalla CPU per l'elaborazione delle immagini implicherebbe un rallentamento di

Potenza di calcolo: la porzione di tempo di CPU dedicato al calcolo dell'immagine video è aumentato proporzionalmente al crescere dell'utilizzo di interfacce grafiche da parte delle applicazioni. Al fine di limitare lo "spreco" di tempo CPU le schede video sono state dotate di **processori dedicati** denominati **GPU** (Graphical Processing Unit). Il processore grafico può essere utilizzato sia per la **grafica 3D**, sia per la **grafica raster**: la CPU del calcolatore non calcola la posizione e il colore di tutti i pixel da disegnare, ma invia un comando all'acceleratore indicandogli cosa deve essere disegnato. Esempio disegna **rettangolo blu centrato in 100, 150 e di dimensioni 20x20**.

Per poter sfruttare le capacità di accelerazione delle schede grafiche è necessario che **le applicazioni conoscano le funzioni da esse implementate**. A causa dell'elevato numero e delle differenze tra le schede grafiche è stato necessario sviluppare delle **librerie standard** (OpenGL, Direct3D) che permettano ai programmatori di **prescindere dalla scheda grafica utilizzata**.

esecuzione del lavoro da parte della CPU stessa.

Così è stata introdotta la GPU (Graphical Processing Unit), una scheda video che è possibile

programmare (soprattutto in linguaggio UDA) che vengono usate nelle attività di calcolo in

quanto analoghi ai problemi di visualizzazione grafica affrontati dalla scheda video.

Le GPU attuali (ATI e Nvidia) sono così efficienti che spesso sono in grado di risolvere problemi più velocemente della stessa CPU. Per quanto riguarda il calcolo la GPU risulta più veloce ma è preferibile distinguere le due tipologie di componenti sia per la limitatezza delle funzioni della GPU (è più rapida solo nei calcoli rispetto alla CPU) che per le problematiche proprie delle schede video (esempio si surriscaldano velocemente).

Calcolatore "bus oriented"

Il computer che noi usiamo è "bus oriented", ossia secondo la teoria di Von Neumann le varie componenti sono collegate da una serie di cavi elettrici stampati su scheda madre detti bus che trasportano i segnali (possono essere dati, segnali di controllo ecc.). I bus si dividono in bus di controllo, dati ed indirizzi.

Consiglio di leggere quanto riportato di seguito osservando il powerpoint condiviso dal professore su teams.

(Link 02bis-Esempio esecuzione istruzioni.pptx)

La CPU internamente è costituita da tre componenti:

- Unità di controllo: comunica se la RAM deve leggere o scrivere
- ALU: riferisce l'operazione da eseguire (moltiplicazione, addizione, AND, OR, ecc.)
- Registri: general purpose o specializzati

La CPU legge le istruzioni della RAM e le esegue.

L' "esecutore"

» Un calcolatore basato sull'architettura di Von Neumann esegue un programma sulla base di questi principi:

- » dati e istruzioni sono memorizzati in una memoria unica che permette sia la scrittura sia la lettura
- » i contenuti della memoria sono indirizzati in base alla loro posizione, indipendentemente dal tipo di dato o istruzione contenuti
- » le istruzioni vengono eseguite in modo sequenziale

» Il linguaggio per cui la CPU si comporta da esecutore è detto **linguaggio macchina**. Le istruzioni scritte in **linguaggio macchina** sono piuttosto rudimentali:

- » il concetto di tipo di dato è quasi assente
- » il numero di operandi è limitato (in genere non più di due)
- » il numero di operazioni previste è ridotto

Struttura istruzione	codice operativo	dest	src1	src2
Linguaggio assemblatore	add	R01	R02	R03
Linguaggio macchina	000000 00000 100000	00001	00010	00011

Le istruzioni stanno, dunque, nella RAM insieme ai dati (innovazione).

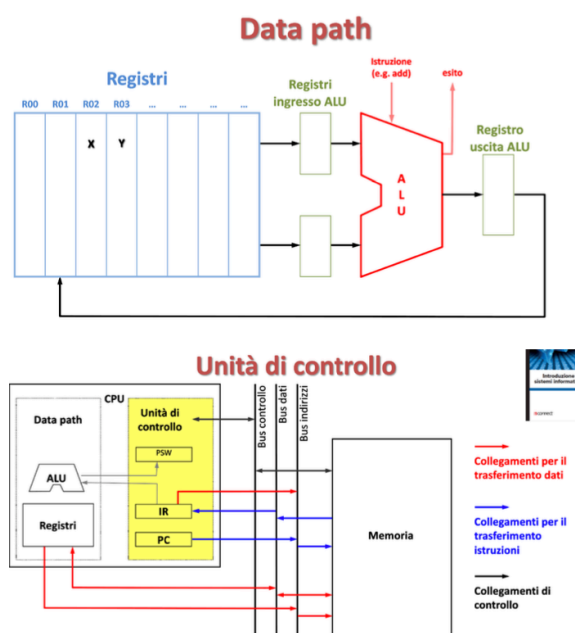
Le istruzioni sono dei numeri binari, ognuno dei quali corrisponde ad una azione particolare. Per esempio, questa riga a destra potrebbe essere contenuta in una cella della

RAM; osserviamo il codice operativo (stabilisce il tipo di operazione da eseguire) e gli argomenti dell'operazione (destinazione e sorgente 1 e 2).

Tramite la colonna add la ALU comprende che deve addizionare, se ci fosse stato moltiplicazione avrebbe capito moltiplica ecc.

C'è un acceso dibattito sul fatto se siano meglio macchine che eseguono più istruzioni semplici o macchine che eseguono meno istruzioni ma complesse. Entrambe le prospettive hanno pro e contro: se si esegue un numero troppo ridotto di istruzioni vuol dire che le istruzioni che vanno tradotte in alto livello saranno tradotte in un enorme numero di istruzioni di basso livello e serviranno più cicli di data path. Allo stesso tempo se si esagera con il numero di istruzioni in linguaggio di basso livello dovrò essere in grado di affrontare la difficoltà di programmazione e gestione delle istruzioni. C'è quindi sempre bisogno di un compromesso.

Il data path regola il percorso dei dati nella CPU quando si devono eseguire operazioni nella ALU:



- Registri in ingresso
 - ALU: esegue istruzione
 - L'esito viene inserito in un registro di uscita dell'ALU
 - L'esito va in memoria in un registro
- È necessaria la presenza di una precedente istruzione che abbia permesso il caricamento (load) dei valori (x,y) nei rispettivi registri della CPU e, per salvarli nella RAM, un'operazione di store. Ricordiamo che ci sono sia operazioni interne alla CPU che operazioni che includono il richiamo della RAM.
- IR: Instruction register, contiene l'istruzione in esecuzione in quell'istante
 - PC: program counter, contiene l'indirizzo dell'istruzione successiva
- Per passare dall'unità di controllo alla RAM e viceversa si usano i bus dati ed indirizzi

Slide: Una successione di istruzioni

Vediamo l'esempio di una successione di istruzioni: sommare i valori A, B e C e salvare il nuovo valore nel registro di A. Mentre in python tale istruzione si esegue in una sola operazione, in linguaggio macchina il tutto si complica: ad esempio, gli oggetti che possono interagire sono solo due per volta e bisogna considerare i trasferimenti tramite i bus tra CPU e RAM.

Nel nostro esempio le istruzioni sono contenute nella RAM nelle celle da 789 a 794 mentre i dati sono sempre nella RAM ma in celle diverse, 4000 (A) 4004 (B) 3 4008 (C). Per semplificazione si mette da parte la presenza della MDR e della MAR.

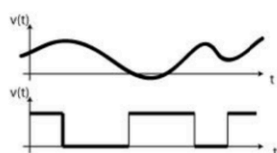
L'esecuzione delle istruzioni ha inizio con l'inserimento nella PC del numero 0789 che rimanda alla cella 0789 della RAM in lettura (secondo l'istruzione dell'unità di controllo);

viene inserita l'istruzione della cella 0789 nella IR, che è load R02, 4000 mentre il PC viene incrementato di 1.

L'istruzione della IR viene eseguita con il caricamento nel registro 2 (R02) del valore posto nella RAM nella cella 4000. Si passa così alla lettura dell'istruzione della cella 0790 (load R03, 4004) e la sua esecuzione (caricamento del valore posto nella cella 4004 della RAM nel registro 3) con aumento del PC. Segue l'istruzione add R01, R02, R03 che permette di sommare (ad opera della ALU) i valori contenuti nel registro 2 e nel registro 3 e conservarlo nel registro 1 (ricorda che il primo registro è la destinazione). Il tutto continua fino all'operazione 0794 store R01, 4000 che permette di conservare il risultato nella RAM tramite scrittura.

Rappresentazione digitale dei numeri

Segnale analogico: Un segnale è analogico quando i valori utili che lo rappresentano sono continui (infiniti) in un intervallo e non numerabili.



Segnale digitale: Un segnale è digitale quando i valori utili che lo rappresentano sono discreti e finiti.

I calcolatori moderni utilizzano due stati logici (binari), ma è possibile progettare sistemi digitali con logica a più stati (detta anche logica multi valori).

Un segnale è detto analogico quando i valori sono continui, cioè si muovono su un dominio infinito in un certo range. Ad esempio, in questa curva il valore della tensione assume qualunque valore nel range indicato.

Nel caso dei segnali digitali abbiamo, invece, un numero finito di valori utili (insieme discreto e numerabile).

I calcolatori moderni utilizzano due stati logici (binari) che per una questione di semplicità di elaborazione sono indicati con 0 = basso valore ed 1 = alto valore.

Inoltre, esistono rappresentazioni digitali multilivelli, ma sempre per una questione di semplicità, costo ed affidabilità (se ho un valore più vicino al valore basso allora si arrotonda ad esso, stessa cosa per un valore che si avvicina al valore alto; nel caso di più stadi invece si può avere incertezza nell'assegnazione del numero per arrotondamento ad un livello piuttosto che ad un altro) si preferisce il modello binario. Infine, un modello a due stadi permette l'uso dell'aritmetica dei calcolatori e dei booleani (verranno affrontati nella prossima lezione).

L'aritmetica usata dai calcolatori è diversa da quella comunemente utilizzata dalle persone per:

$$\sqrt{2} = \boxed{1 \quad . \quad 4 \quad 1 \quad 4 \quad 2 \quad 1} \quad 3 \quad 5 \quad 6 \quad 2$$

$$124 \Rightarrow 01111100$$

•**Precisione:** per esempio, una persona potrebbe non fermarsi mai nella ricerca della radice di 2 mentre il computer, avendo dei limiti, lo rappresenterebbe con 7 cifre per esaurimento dello spazio.

•**Rappresentazione:** è normalmente ottenuta utilizzando il sistema binario poiché più adatto a essere maneggiato dal calcolatore.

Numeri a precisione finita

I numeri a precisione finita sono quelli rappresentati con un numero finito di cifre una volta fissate delle caratteristiche.

Esempio: Numeri non rappresentabili con 3 cifre senza virgola e senza segno

Numeri superiori a 999

Numeri negativi

Frazioni

1	5	9
---	---	---

1	5	9
---	---	---

Le operazioni con i numeri a precisione finita causano errori ogni qualvolta il loro risultato non appartiene all'insieme dei valori rappresentabili:

- **Underflow:** si verifica quando il risultato dell'operazione è minore del più piccolo valore rappresentabile.
- **Overflow:** si verifica quando il risultato dell'operazione è maggiore del più grande valore rappresentabile.
- **Non appartenenza all'insieme:** si verifica quando il risultato dell'operazione, pur essendo troppo grande o troppo piccolo, non appartiene all'insieme dei valori rappresentabili.

Esempio: Numeri a precisione finita con 3 cifre senza virgola e senza segno

$600+600 = 1200 \Rightarrow \text{Overflow}$
 $300-600 = -300 \Rightarrow \text{Underflow}$
 $007/002 = 3.5 \Rightarrow \text{Non appartenenza all'insieme}$

Si dice che i numeri a precisione finita, a differenza dei numeri che siamo abituati a usare, non rispettano la chiusura rispetto alle operazioni di somma, sottrazione e prodotto.

Esempio: Risultati non rappresentabili con 3 cifre senza virgola e senza segno

Operazioni	Interi	Precisione finita
Somma: $600+600$	1200	Overflow
Sottrazione: $300-600$	-300	Underflow
Prodotto: 050×050	2500	Overflow

Anche l'algebra dei numeri a precisione finita è diversa da quella convenzionale. Poiché alcune delle proprietà non vengono rispettate:

Proprietà associativa: $a + (b - c) = (a + b) - c$

Proprietà distributiva: $a \times (b - c) = a \times b - a \times c$

Non sono rispettate poiché in base all'ordine con cui vengono eseguite le operazioni si può verificare o meno un errore

Esempio: Operazioni con numeri a precisione finita di 3 cifre senza virgola e senza segno

 $400 + (300 - 500) = (400 + 300) - 500$

$$400 + (-200) = 700 - 500$$

Underflow

$$50 \times (50 - 40) = 50 \times 50 - 50 \times 40$$

$$50 \times 10 = \text{2500} - \text{2000}$$

Overflow Overflow

ATTENZIONE: non confondere i numeri negativi con le operazioni di sottrazione

Questo problema dell'ordine dei valori nelle operazioni è utile non solo per scongiurare problemi di underflow ed overflow ma anche per l'approssimazione dei valori.