

Fondamenti di Informatica
Corso di laurea in Psicologia
Università di Roma “La Sapienza”
a.a. 2000/2001

Dispensa 1
Il computer

Riccardo Rosati

VERSIONE PRELIMINARE

Indice

1	Il computer: generalità	4
2	Rappresentazione binaria dell'informazione	4
2.1	Rappresentazione digitale e analogica	5
2.2	Il codice binario	6
2.3	Rappresentazione binaria di informazione numerica	7
2.4	Rappresentazione binaria di informazione analogica	8
3	La macchina di von Neumann	10
4	La CPU	12
4.1	Caratteristiche fisiche della CPU	12
4.2	La frequenza di clock	13
4.3	Il set di istruzioni della CPU	13
5	Memoria centrale	14
5.1	La RAM	14
5.2	La ROM	15
5.3	La memoria cache	16
6	Il bus	16
7	Le unità periferiche	17
7.1	Periferiche di input	18
7.2	Periferiche di output	18
7.3	Memorie di massa	19

Sommario

L'obiettivo di questa dispensa è quello di introdurre i concetti di computer e di rappresentazione binaria dell'informazione. A questo scopo, vengono dapprima introdotte alcune nozioni di base dell'informatica, e successivamente viene illustrata la rappresentazione digitale e binaria dell'informazione. Viene quindi presentata a grandi linee l'architettura di un computer (macchina di von Neumann) ed il suo funzionamento. Infine, vengono brevemente presentati i principali componenti fisici del computer: CPU, memoria centrale, memorie di massa, unità periferiche.

Riguardo agli argomenti trattati in questa dispensa:

- le memorie di massa e le unità periferiche vengono ripresi e trattati in maggiore dettaglio nella dispensa 4;
- i formati usati per la rappresentazione binaria dei vari tipi di informazione vengono ripresi e descritti in maggiore dettaglio nella dispensa 7.

1 Il computer: generalità

Cos'è un computer? Un computer è una macchina in grado di elaborare informazione. Più precisamente, un computer elabora e memorizza *automaticamente* informazioni, rappresentate in un particolare formato detto *digitale* o più precisamente *binario*, sulla base di istruzioni memorizzate nel computer. Un insieme di tali istruzioni è detto *programma*.

La caratteristica fondamentale dei computer è quella di essere delle macchine *programmabili*: in altre parole, è possibile cambiare il tipo di elaborazione svolta dalla macchina, e quindi la funzione che svolge la macchina, cambiando il programma contenuto nella macchina. La programmabilità è la caratteristica che rende i computer delle macchine estremamente potenti e versatili.

Possiamo quindi immaginare il computer come una macchina che:

1. legge un certo insieme di informazioni, detto *dati di input* o di ingresso
2. elabora tali informazioni in base ad un insieme di istruzioni (programma)
3. restituisce le informazioni così trasformate, dette *dati di output* o di uscita

Esistono vari tipi di computer, classificati in base alla loro capacità elaborativa e ai tipi di funzioni che svolgono: supercomputer, mainframe, server, personal computer. Il *personal computer*, che è quello che esamineremo in dettaglio nel seguito, ha un impiego fondamentalmente individuale, cioè è di solito usato da una singola persona, o utente.

Come già accennato, un computer elabora automaticamente informazione sulla base di insiemi di istruzioni detti programmi. I programmi costituiscono quindi una componente importante del computer. Accanto ad essi, anche i dati che vengono elaborati automaticamente dal computer, e che possono venire memorizzati in modo permanente dal computer stesso, sono una componente necessaria per l'utilizzo di tali macchine.

Tuttavia, a differenza dei componenti fisici degli elaboratori, questi componenti non si possono toccare con mano. Si distingue quindi la componente fisica del computer, detta *hardware*, dal *software*, cioè i vari tipi di informazioni (programmi, dati da elaborare, dati memorizzati) presenti nel computer.

Si noti, comunque, che molto spesso nel gergo informatico il termine software viene usato per indicare soltanto i programmi, non i dati.

2 Rappresentazione binaria dell'informazione

Nel computer l'informazione è rappresentata in forma numerica, più precisamente in forma binaria. In questa sezione esaminiamo le nozioni di rap-

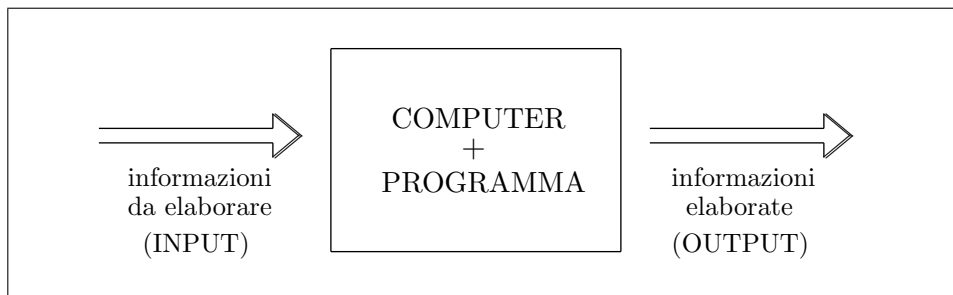


Figura 1: Il computer come elaboratore di informazione

presentazione analogica, digitale (o numerica) e binaria dell'informazione, e introduciamo i concetti di bit e byte di informazione.

2.1 Rappresentazione digitale e analogica

L'informazione che viene elaborata all'interno di un computer è rappresentata in forma *digitale*, cioè numerica. La rappresentazione digitale dell'informazione si distingue dalla rappresentazione *analogica*. In pratica, nella rappresentazione digitale, l'informazione riguardante una certa grandezza è rappresentata in modo *discreto*, nel senso che in tale rappresentazione la grandezza può assumere soltanto un numero finito di valori diversi, mentre nella rappresentazione analogica la grandezza è rappresentata in modo *continuo*, vengono cioè rappresentati un numero infinito di valori diversi della grandezza.

Più precisamente, una grandezza si dice continua quando, dato un qualunque intervallo di valori (identificato da un valore minimo e da un valore massimo), la grandezza può assumere un numero infinito di valori diversi all'interno di tale intervallo. Tutte le grandezze fisiche sono in realtà continue: si pensi allo spazio, al tempo, all'intensità e frequenza della luce, all'intensità e frequenza delle onde sonore.

Se l'informazione che deve essere rappresentata riguarda una grandezza di tipo analogico, e vogliamo adottare una rappresentazione digitale, si pone il problema di come rappresentare in modo *finito* tale grandezza. Se ad esempio dobbiamo rappresentare il valore della temperatura di una certa località, e decidiamo di adottare solo 1 cifra decimale per rappresentare la parte frazionaria del valore della temperatura, rappresenteremo come 18,2 gradi tutte le temperature comprese nell'intervallo tra 18,2 e 18,3 gradi (oppure, per minimizzare l'errore massimo commesso nella rappresentazione, tutte le temperature tra 18,15 e 18,25 gradi), commettendo quindi un errore di approssimazione nella rappresentazione di tale valore. E anche aumentando a piacere il numero di cifre utilizzato per rappresentare la parte frazionaria, introdurremo comunque un errore di approssimazione, per quanto piccolo. Ovviamente, per poter elaborare in modo automatico l'informazione, oc-

corre che questa sia rappresentata in modo finito, quindi nelle applicazioni informatiche la rappresentazione digitale di informazione relativa a grandezze continue (e quindi analogiche) è sempre approssimata, mai esatta. Questo implica, ad esempio, che immagini, suoni e filmati sono sempre rappresentati in modo approssimato all'interno del computer.

D'altra parte, esistono altri tipi di informazione che *non* sono analogici: un esempio classico è l'informazione contenuta in una pagina di testo scritta in lingua italiana. Tutta l'informazione contenuta in tale tipo di testo può essere rappresentata fedelmente (cioè senza introdurre errori) in modo digitale, associando un diverso numero *finito* a ogni carattere dell'alfabeto e a ogni simbolo di punteggiatura o di spaziatura, e quindi rappresentando il testo attraverso una sequenza di numeri, in base a tale associazione. La differenza con il caso della grandezza continua, è che l'alfabeto italiano (così come l'alfabeto di qualunque lingua esistente) è un insieme finito di simboli, mentre l'insieme dei valori che la grandezza continua può assumere è infinito.

La differenza tra rappresentazione analogica e digitale può essere illustrata dal seguente esempio. Uno strumento di misurazione di una grandezza fisica che visualizza la misura fatta attraverso una lancetta, effettua una rappresentazione analogica, in quanto la lancetta può assumere un numero infinito di posizioni diverse, mentre uno strumento con visualizzatore a numeri adotta una rappresentazione digitale, in quanto può visualizzare solo un numero finito di valori diversi (ad esempio, un visualizzatore a tre cifre decimali può visualizzare al più 1000 valori diversi).

2.2 Il codice binario

L'informazione che viene elaborata all'interno di un computer è rappresentata in formato *binario*. In pratica, viene utilizzato un linguaggio o codice il cui alfabeto è costituito da due soli simboli (per convenzione si denotano con 0 e 1): ogni informazione contenuta nel computer è pertanto rappresentata da una sequenza di 0 e 1. Tutto il software è dunque scritto in codice binario: programmi, dati da elaborare, dati memorizzati, sono tutti rappresentati da sequenze di 0 e 1.

Ogni singolo elemento di una sequenza binaria viene detto *bit*: un bit può quindi assumere il valore 0 oppure il valore 1. Il bit è, pertanto, l'unità elementare di informazione.

Oltre al bit, l'unità di misura dell'informazione immediatamente seguente è il *byte*, corrispondente ad una sequenza di 8 bit. Come vedremo nel seguito, il byte (ed i suoi multipli) è in realtà l'unità di misura dell'informazione maggiormente utilizzata in informatica, in quanto all'interno del computer in genere l'informazione binaria viene elaborata dai dispositivi fisici non un bit alla volta, ma considerando contemporaneamente gruppi di otto bit o di multipli di 8 bit (16, 32 o 64 bit). Pertanto molto spesso l'in-

formazione binaria viene logicamente vista considerando il byte (invece del bit) come unità elementare di informazione.

La ragione per cui nei computer l'informazione è codificata in forma numerica, in particolare in binario, e non in forma analogica, è puramente tecnica. All'interno dei computer infatti le informazioni viaggiano sotto forma di segnali elettrici, che sono soggetti ad interferenze ed errori: la codifica binaria è quella che garantisce la maggiore tolleranza alle interferenze, in quanto è necessario rappresentare solo due diversi valori, che vengono rappresentati da due diversi livelli di una grandezza fisica, cioè la tensione elettrica, alla massima distanza tra loro. Perché una interferenza faccia commettere un errore di trasmissione (cioè perché uno 0 venga interpretato come 1 o viceversa), è necessario che tale interferenza trasformi il livello massimo di tensione nel livello minimo, o viceversa. Per una rappresentazione di tipo analogico, ogni interferenza causerebbe un errore, mentre qualunque altro codice numerico con più di due simboli costringerebbe a ridurre la differenza di valore tra due livelli corrispondenti a due simboli diversi, con conseguente maggiore sensibilità alle interferenze.

2.3 Rappresentazione binaria di informazione numerica

Ogni informazione di tipo numerico non binario può essere facilmente rappresentata in forma binaria. La rappresentazione binaria di un numero si ottiene attraverso una sequenza di 0 e 1 via via più lunga quanto più è grande il numero da rappresentare, così come nel sistema di rappresentazione decimale dei numeri (quello più familiare a noi esseri umani) viene utilizzata una sequenza di cifre da 0 a 9 via via più lunga quanto più è grande il numero da rappresentare. Ad esempio, usando 4 bit, cioè quattro cifre binarie, siamo in grado di rappresentare i numeri da 0 a 15 come segue:

Rappresentazione decimale	Rappresentazione binaria
0	0000
1	0001
2	0010
3	0011
4	0100
5	0101
6	0110
7	0111
8	1000
9	1001
10	1010
11	1011
12	1100

13	1101
14	1110
15	1111

Per rappresentare numeri più grandi di 15 occorrono più di quattro bit. Ad esempio, il numero 16 si rappresenta come 10000, cioè con cinque bit.

In generale, k bit permettono di rappresentare 2^k numeri. Ad esempio, 4 bit permettono di rappresentare $2^4 = 16$ numeri, cioè i numeri da 0 a 15: ci sono infatti solo 16 diverse sequenze di 0 e 1 lunghe 4 bit. Per rappresentare un qualunque numero n occorre un numero di bit k tale che $2^k > n$. Ad esempio, per rappresentare il numero 1000 sono sufficienti 10 bit, perché $2^{10} = 1024 > 1000$: d'altra parte, 9 bit non sono sufficienti, infatti $2^9 = 512 < 1000$.

Dato un numero intero positivo n , un modo per trovare la sua rappresentazione binaria è il seguente: si divide il numero per due: se il resto della divisione è 1 (cioè se il numero era dispari) si scrive 1, altrimenti si scrive 0, e si ripete la divisione per due finché il numero non diventa 0. I resti di tutte le divisioni successive, presi nell'ordine inverso a quello di esecuzione delle divisioni, corrispondono alla rappresentazione binaria del numero.

Ad esempio:

13 diviso 2 = 6 e resto 1
6 diviso 2 = 3 e resto 0
4 diviso 2 = 2 e resto 0
2 diviso 2 = 1 e resto 0
1 diviso 2 = 0 e resto 1

La rappresentazione binaria di 13 è infatti 1101.

Oltre ai numeri interi positivi, è possibile rappresentare numeri negativi e numeri non interi, cioè numeri con parte frazionaria, estendendo opportunamente la rappresentazione vista in precedenza per gli interi positivi.

2.4 Rappresentazione binaria di informazione analogica

Affrontiamo ora il problema della rappresentazione binaria di informazione analogica. Come fare a rappresentare informazione di tipo non numerico, cioè analogico, in forma binaria? essendo quella binaria un caso particolare di rappresentazione digitale, occorre necessariamente passare per una approssimazione dell'informazione, come visto in precedenza.

Nel caso della rappresentazione di un valore di una grandezza analogica, si passa dapprima ad una rappresentazione numerica finita, approssimando il valore vero v al valore più vicino al valore reale della grandezza, rispetto ad un insieme finito di valori. Quindi, si rappresenta in binario tale numero.

Supponiamo ad esempio di dover rappresentare in binario il valore v di una temperatura che può assumere un valore compreso tra 0 e 100 gradi

centigradi, scegliendo come insiemi di valori le temperature corrispondenti a numeri interi, cioè

$$0, 1, 2, \dots, 99, 100$$

Si procede come segue:

1. dapprima si approssima v al valore di temperatura più vicino a v nella “griglia” dei valori $0, 1, 2, \dots, 99, 100$. Se ad esempio v è compreso tra 21,5 e 22,5 gradi centigradi, approssimiamo v al valore 22;
2. si rappresenta quindi il numero usato per rappresentare v in binario. Applicando il metodo di calcolo della rappresentazione binaria di un intero positivo illustrato in precedenza, otteniamo in questo caso che la rappresentazione binaria di 22 corrisponde a 10110.

Un ulteriore livello di approssimazione va inoltre introdotto nei casi in cui si deve rappresentare una grandezza in un intervallo di spazio o di tempo. Ad esempio, una fotografia in bianco e nero rappresenta una informazione relativa all'intensità luminosa in un intervallo spaziale bidimensionale, così come l'andamento della temperatura di una località nell'arco di 24 ore rappresenta una informazione sulla temperatura in un intervallo temporale. Dato che spazio e tempo sono anch'esse grandezze continue, qualunque loro intervallo contiene un insieme infinito di valori. Ad esempio, qualunque intervallo di tempo, per quanto piccolo, contiene al suo interno un insieme infinito di istanti di tempo.

Come fare allora per rappresentare in modo finito tali tipi di informazioni? si fa ricorso al *campionamento* della grandezza da misurare: si sceglie cioè un insieme finito di rappresentanti dell'intervallo continuo in cui la grandezza deve essere misurata. Tali rappresentanti vengono detti *campioni* della grandezza fisica. Ad esempio, nel caso della misura dell'andamento della temperatura di una località nell'arco di 24 ore, si decide di rappresentare il valore che assume tale temperatura solo in un insieme finito di istanti di tempo, ad esempio ogni minuto. Avremo in questo modo 60 campioni per ogni ora. Si rappresenta così l'andamento nell'intervallo per mezzo di $24 \cdot 60 = 1440$ valori della temperatura. Tali valori vengono quindi rappresentati in binario come visto in precedenza, dapprima approssimando il loro valore ad un numero compreso in una griglia di valori e quindi rappresentando tale numero in binario.

Come è facile immaginare, sia la scelta dell'insieme finito di valori (griglia) usato per rappresentare un singolo valore della grandezza, sia la scelta del numero dei campioni della grandezza in un intervallo continuo, sono aspetti cruciali rispetto al grado di attendibilità della rappresentazione binaria dell'informazione. Da una parte, minore è il numero di valori della griglia (o il numero dei campioni), maggiore è l'approssimazione della rappresentazione binaria; d'altra parte, aumentando il numero di valori della

griglia (o il numero dei campioni) aumenta la dimensione della rappresentazione binaria dell'informazione, che è un aspetto di estrema importanza nelle applicazioni informatiche. Occorre pertanto trovare un compromesso tra le opposte esigenze di precisione e di compattezza della rappresentazione.¹

In conclusione, apriamo una parentesi per far notare come la rappresentazione binaria dell'informazione analogica sia un problema centrale per molti ambiti applicativi dell'informatica, in particolare per le applicazioni *multimediali*, cioè le applicazioni che riguardano l'elaborazione automatica di immagini, suoni e filmati. Per comprendere l'importanza che ora riveste questo tipo di applicazioni per l'informatica basti pensare che, insieme alla diffusione di Internet, l'elaborazione di informazioni multimediali è la causa principale della diffusione di massa delle tecnologie informatiche, iniziata nella seconda metà degli anni '90 e ancora in corso.

3 La macchina di von Neumann

Dal punto di vista fisico, i principali dispositivi che costituiscono un computer sono i seguenti:

- CPU o microprocessore
- memoria centrale (RAM)
- unità periferiche
- canali di comunicazione tra dispositivi (bus)

L'architettura di base di un computer è sintetizzata nella *macchina di von Neumann*. La macchina di von Neumann venne ideata da von Neumann, Burks e Goldstine nel 1947 allo "Institute for Advanced Studies" di Princeton, USA. Lo scopo di tale macchina era quello di svolgere con grande velocità operazioni di calcolo.

Nonostante le architetture dei computer si siano molto evolute negli anni, presentano ancora molte analogie con il modello originale. Pertanto, la macchina di von Neumann è ancora oggi adatta a rappresentare la struttura astratta di un computer.

La macchina di von Neumann presenta l'architettura mostrata in figura 2. Uno schema di massima del funzionamento della macchina di von Neumann, e quindi del computer, è il seguente.

- L'unità centrale, o CPU (Central Processing Unit) elabora i dati eseguendo i programmi.

¹Questi aspetti verranno ripresi in una dispensa successiva dedicata ai formati di rappresentazione digitali.

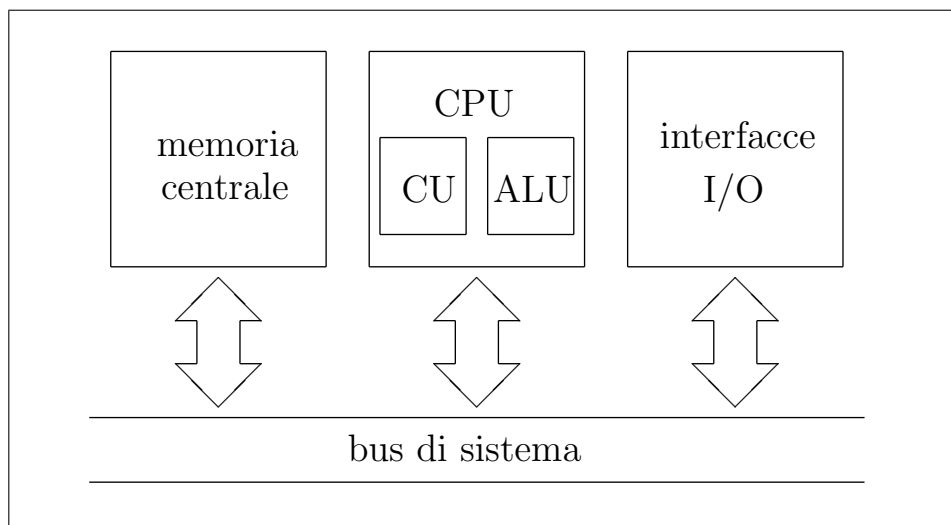


Figura 2: La macchina di von Neumann

- Sia i dati da elaborare che il programma da eseguire risiedono nella memoria centrale, che viene letta e modificata dalla CPU.
- La CPU comanda inoltre dei dispositivi chiamati *interfacce di input/output (I/O)*, che sono a loro volta connesse con le *unità periferiche*. Le unità periferiche sono i dispositivi che permettono il trasferimento dei dati dal computer al mondo esterno e viceversa. Compito delle interfacce di I/O è quello di gestire la comunicazione tra le unità periferiche e la macchina. Pertanto, secondo il modello della macchina di von Neumann, le unità periferiche sono considerate dispositivi esterni al computer.

Più precisamente, nella CPU possono essere distinti due blocchi funzionali: l'*unità di controllo* (CU, Control Unit) e l'*unità logico-aritmetica* (ALU, Arithmetic Logical Unit). L'unità di controllo esegue le istruzioni dei programmi, comanda il funzionamento degli altri dispositivi della CPU e controlla il flusso di istruzioni tra CPU e memoria, mentre l'unità logico-aritmetica effettua le operazioni aritmetiche e le operazioni logiche di confronto sui dati.

Esistono due tipi di istruzioni: istruzioni di *elaborazione* dei dati, come ad esempio il calcolo della somma di due dati memorizzati in memoria centrale, e istruzioni di *trasferimento* dei dati da un dispositivo ad un altro della macchina, come ad esempio lo spostamento di un dato residente in memoria centrale in una interfaccia di I/O.

Unità di controllo, unità logico-aritmetica, memoria centrale e interfacce di I/O partecipano al seguente *ciclo della macchina*, che permette l'esecu-

zione di una singola istruzione. Ad esempio, nel caso dell'esecuzione di una istruzione di elaborazione dati, il ciclo della macchina è il seguente:

1. *caricamento*: l'unità di controllo preleva dalla memoria centrale la prossima istruzione da eseguire;
2. *decodifica*: l'unità di controllo interpreta l'istruzione da eseguire e trasferisce i dati a cui tale istruzione si riferisce dalla memoria centrale all'unità logico-aritmetica;
3. *esecuzione*: l'unità logico-aritmetica esegue l'elaborazione sui dati;
4. *memorizzazione*: i risultati dell'elaborazione vengono memorizzati nella memoria centrale o in un registro dell'unità logico-aritmetica.

Iterando il precedente ciclo, la macchina è in grado di eseguire un programma, cioè un insieme di istruzioni, memorizzato nella memoria centrale.

4 La CPU

Come visto in precedenza, l'unità centrale di elaborazione o *CPU* (dall'inglese Central Processing Unit), è il “cervello” del computer. Dal punto di vista funzionale, la CPU comanda il funzionamento del computer. In particolare, la CPU interpreta i programmi contenuti nel computer, esegue tali programmi, elaborando le informazioni in base alle istruzioni contenute nei programmi, e comanda gli altri dispositivi fisici del computer (memoria centrale, unità periferiche).

4.1 Caratteristiche fisiche della CPU

Dal punto di vista fisico, nei personal computer la CPU è costituita da un *microprocessore*, cioè un circuito elettronico ricavato da una lamina di silicio (chip). Il chip contenuto nel microprocessore contiene milioni di transistor.

I parametri che danno una indicazione della capacità elaborativa della CPU sono:

- numero di transistor nel processore
- frequenza di clock

Il numero di transistor contenuti in un singolo chip dà una misura della potenza elaborativa delle CPU in quanto, maggiore è tale numero, più complesse sono le operazioni che la CPU è in grado di compiere a parità di tempo impiegato. Tale numero è in costante aumento: la *legge di Moore*, dovuta a Gordon Moore, uno dei fondatori della Intel, prevede che il numero di transistor contenuti in un chip raddoppi ogni anno e mezzo. La legge fu enunciata nel 1965, e fino ad oggi (anno 2001) è stata rispettata.

4.2 La frequenza di clock

La *frequenza di clock* è in pratica il “ritmo” a cui si svolge il funzionamento delle operazioni nella CPU. Ad esempio, una frequenza di clock di 800 Megahertz (Mhz) sta ad indicare che il ciclo di lettura, decodifica ed esecuzione di una istruzione viene ripetuto 800 milioni di volte al secondo.

La frequenza di clock della CPU è una misura, anche se approssimata, delle prestazioni della CPU. Più è alta tale frequenza, maggiore è il numero di istruzioni al secondo eseguite dalla CPU. Ad esempio, un Pentium III a 800 Mhz è più veloce di un Pentium III a 600 Mhz. Si noti però che la frequenza di clock non è un buon parametro per confrontare le prestazioni di due *diversi* tipi di CPU, in quanto CPU diverse sono in grado di leggere, decodificare ed eseguire diversi insiemi di istruzioni, e quindi il semplice numero di istruzioni eseguite in un secondo non caratterizza completamente le reali capacità della CPU.

Tra le più diffuse CPU nei personal computer attuali, citiamo il Pentium III, Pentium IV e Celeron della Intel, il PowerPC della Motorola, l'Athlon della AMD.

4.3 Il set di istruzioni della CPU

Ogni CPU è in grado di riconoscere e interpretare un insieme preciso di istruzioni, detto *set di istruzioni*. Il set di istruzioni della CPU costituisce il cosiddetto *linguaggio macchina* della CPU.

Le istruzioni del linguaggio macchina che una CPU è in grado di interpretare sono molto semplici, come ad esempio:

- istruzioni di lettura da memoria centrale e scrittura in memoria centrale di un dato;
- istruzioni aritmetiche (operazioni su dati numerici);
- istruzioni di “salto” ad una certa istruzione del programma.

In base al set di istruzioni riconosciuto (che a sua volta dipende dalla loro architettura interna), le CPU si distinguono in *CISC* (Complex Instruction Set Computer) e *RISC* (Reduced Instruction Set Computer). Una CPU di tipo CISC ha un set di istruzioni molto esteso (in genere qualche centinaio di istruzioni), mentre una CPU di tipo RISC ha un set di istruzioni più ristretto (in genere intorno al centinaio di istruzioni), e comprende solo istruzioni estremamente semplici. Tra le CPU per personal computer più diffuse, il processore Pentium è di tipo CISC, mentre il processore PowerPC è di tipo RISC.

Una ulteriore misura delle prestazioni della CPU è data dal numero di istruzioni al secondo (IPS) che è in grado di eseguire. Attualmente, l'unità di misura adottata al riguardo è il MIPS, corrispondente ad un milione di

istruzioni al secondo. Si noti, comunque, che la misura in MIPS dà solo una stima indicativa della reale velocità dell'elaboratore, in quanto il set di istruzioni (cioè il tipo di istruzioni eseguite) varia da CPU a CPU, e può quindi rendere relativamente significativa la misura delle prestazioni in MIPS, specialmente quando si confrontano tra loro una CPU RISC con una CPU CISC.

5 Memoria centrale

La memoria centrale del computer si distingue in RAM, ROM e memoria cache.

5.1 La RAM

La memoria centrale, o *RAM*, del computer, è il dispositivo dove il computer mantiene i dati che vengono elaborati e i programmi che sono in esecuzione. La RAM ad ogni istante contiene quindi sia le informazioni che vengono elaborate che le istruzioni per eseguire l'elaborazione stessa. La CPU è il dispositivo che “usa” la RAM, scrivendo nuovi dati all'interno di essa oppure leggendo dei dati precedentemente memorizzati o delle istruzioni che la CPU dovrà eseguire. Pertanto la RAM costituisce la “memoria di lavoro” in cui la CPU tiene nota delle informazioni necessarie per eseguire l'elaborazione corrente.

La RAM è realizzata mediante uno o più circuiti integrati (chip). Il significato dell'acronimo RAM è “Random Access Memory”, cioè memoria ad accesso casuale. Il termine “casuale” sta a significare che la CPU può accedere direttamente ogni elemento di memoria, detto *locazione*, della RAM (per leggere e scrivere dati), senza dover prima leggere il contenuto di altre locazioni della memoria. In questo senso, l'aggettivo “casuale” si contrappone a “sequenziale”. Si pensi alla differenza d'uso tra una audiocassetta (dispositivo ad accesso sequenziale) e un CD (ad accesso casuale): per ascoltare la quinta canzone registrata nella audiocassetta bisogna prima scorrere il nastro fino a raggiungere la posizione cercata, mentre nel caso del CD l'accesso alla quinta canzone è diretto.

Quando il computer viene spento, i dati e i programmi presenti nella RAM vengono cancellati. Questo è il motivo per cui la RAM è detta memoria *volatile* o temporanea. Inoltre, anche quando il computer è acceso, la CPU modifica di continuo il contenuto della RAM. Questa caratteristica di volatilità rende necessaria la presenza di memorie permanenti all'interno del computer, le memorie di massa, che analizzeremo nel seguito.

Esistono due principali tipi di RAM: la RAM statica (SRAM) e la RAM dinamica (DRAM). La DRAM è il tipo di memoria centrale più diffuso. Il suo nome deriva dal fatto che in essa le celle di memoria mantengono i dati solo per breve tempo e devono essere quindi aggiornate varie centinaia di

volte al secondo. Tra i tipi di DRAM attualmente in uso, in ordine crescente di velocità di accesso, citiamo la EDO (Extended Data Out) DRAM, la SDRAM (Synchronous DRAM) e la RDRAM (Rambus DRAM). La SRAM invece non ha bisogno di essere aggiornata per mantenere i dati memorizzati. Per questa ragione è una memoria più “veloce” della DRAM (nel senso che la CPU è in grado di reperire i dati più velocemente da una SRAM che da una DRAM), ma ha lo svantaggio di essere più costosa della DRAM, anche in termini di consumo di energia.

La quantità di memoria che la RAM mette a disposizione è una importante grandezza per il funzionamento del computer: maggiore infatti è la capacità di memoria della RAM, maggiore è il numero di programmi che il computer è in grado di tenere in memoria, e quindi di eseguire, contemporaneamente. Inoltre, per molti programmi la dimensione della RAM è direttamente proporzionale alla dimensione massima dei dati di input che il computer è in grado di gestire.

La capacità di memorizzazione della RAM si misura in multipli di byte. In particolare, viene usata la seguente terminologia:

- KiloByte (KB) = 2^{10} byte = 1024 byte
- MegaByte (MB) = 2^{10} KB = 2^{20} byte = 1048576 byte
- GigaByte (GB) = 2^{10} MB = 2^{30} byte = 1073741824 byte
- TeraByte (TB) = 2^{10} GB = 2^{40} byte = 1099511627776 byte

Le RAM dei personal computer attualmente prodotti hanno dimensioni che variano dai 64 ai 256 MegaByte.

5.2 La ROM

Oltre alla RAM, c'è un altro tipo di memoria presente all'interno del computer, la *ROM* che, come la RAM, è realizzato mediante uno o più circuiti integrati (chip). Il significato dell'acronimo ROM è “Read-Only Memory”, cioè memoria a sola lettura, infatti, a differenza della RAM, il contenuto della ROM può essere letto ma non modificato. Inoltre, la ROM è una memoria *non volatile*, nel senso che quando il computer viene spento, la ROM non viene cancellata.

In realtà, il contenuto della ROM viene predisposto al momento della sua fabbricazione, e il suo contenuto non può essere modificato dall'utente. La ROM viene quindi usata quando è necessario tenere all'interno del computer una copia permanente di un insieme di dati e/o programmi (ad esempio il BIOS, che verrà presentato nel seguito).

5.3 La memoria cache

La memoria cache è una particolare area di memoria, ad alta velocità, che permette di accelerare il funzionamento del computer.

Per le sue caratteristiche fisiche, la memoria cache permette di accedere ai dati da essa memorizzati in tempi più veloci rispetto alla memoria RAM. Pertanto, per velocizzare le operazioni della CPU, una parte dei dati memorizzati dalla RAM vengono copiati sulla memoria cache. Quindi, quando la CPU ha bisogno di accedere ad un certo dato, controlla per prima cosa se questo dato è disponibile nella memoria cache: se tale dato non è presente in memoria cache, la CPU è costretta ad accedere la più lenta memoria RAM.

Essendo molto costosa, la memoria cache ha dimensioni molto minori della memoria RAM. Pertanto, occorre scegliere quali dati vanno copiati dalla RAM alla memoria cache. Per ottimizzare l'efficienza, occorre che i dati residenti in memoria cache siano tra quelli più acceduti dalla CPU, ma identificare esattamente tali dati non è in generale possibile, e si utilizzano allo scopo varie strategie che approssimano la scelta ottimale.

Nei personal computer attuali si utilizzano molto spesso due diversi livelli di memoria cache. La *cache di primo livello* è un dispositivo che fa fisicamente parte della CPU, ed ha dimensioni molto piccole (tipicamente 256 KB o al massimo 512 KB), mentre la *cache di secondo livello* è un dispositivo esterno alla CPU e di dimensioni maggiori della cache di primo livello (può superare la dimensione di 1MB). Nell'architettura a due livelli di cache, la CPU cerca il dato da accedere dapprima nella cache di primo livello, che essendo interna alla stessa CPU è accessibile in modo estremamente veloce. Se il dato non è presente, viene cercato nella cache di secondo livello. Infine, se tale ricerca non ha avuto esito positivo, la CPU accede al dato in memoria RAM.

6 Il bus

Il bus del computer è un insieme di collegamenti elettrici che permette la comunicazione tra tutti i componenti fisici del computer.

Un bus è costituito da un insieme di linee di comunicazione che permettono ai segnali di viaggiare all'interno del computer. Nell'architettura classica esiste un unico bus, che connette tra loro tutti i componenti del computer. Le varie linee di comunicazione del bus vengono distinte in base al tipo di informazione trasportato: possono infatti trasportare informazione relativa ai dati (bus dati), agli indirizzi (bus indirizzi) o a segnali di controllo per i dispositivi (bus controlli).

Nell'architettura classica del computer, esiste un unico bus, al quale tutti i componenti fisici del computer sono connessi. In realtà, per migliorare le prestazioni, nei personal computer attuali esistono molti bus: un *bus di sistema* dedicato al collegamento tra CPU e RAM (e al quale quindi si

affacciano solo questi due dispositivi), e uno o più *bus locali* dedicati al collegamento delle unità periferiche con la CPU e la RAM. Ad esempio, in molti personal computer che adottano la CPU Intel Pentium esiste un bus locale chiamato AGP (Accelerated Graphics Port) dedicato alla comunicazione tra la periferica video (il monitor) e la CPU e la RAM.

7 Le unità periferiche

Le unità periferiche sono così chiamate perché, nello schema classico del computer, non fanno parte del “nucleo” dell’elaboratore, costituito da CPU, RAM, bus e interfacce con le unità periferiche. Queste unità sono insomma concettualmente alla “periferia” dell’elaboratore.

Le unità periferiche vengono anche dette *dispositivi di input-output* (I/O) in quanto il loro scopo è permettere lo scambio di dati, sia in ingresso che in uscita, tra il computer ed il resto del mondo. A seconda della loro capacità di trasmettere dati dal computer all’esterno o viceversa, le unità periferiche si dividono in:

- periferiche di input
- periferiche di output
- memorie di massa

Le principali periferiche di input sono:

- tastiera
- mouse
- scanner
- microfono
- telecamera

Le principali periferiche di output sono:

- monitor (o schermo)
- stampante
- casse acustiche

Le principali memorie di massa sono:

- hard disk
- floppy disk
- CD-ROM
- CD-RW

7.1 Periferiche di input

Le principali periferiche di input sono:

- tastiera
- mouse
- scanner
- microfono
- telecamera

La *tastiera* è il dispositivo che permette di immettere dati nel computer sotto forma di caratteri alfanumerici. A tale proposito, ricordiamo che nel computer le informazioni sono sempre rappresentate in formato binario. Pertanto, i caratteri alfanumerici vengono convertiti dalla tastiera in sequenze di bit (caratteri diversi vengono rappresentati da sequenze diverse): tali sequenze vengono quindi trasmesse dalla tastiera al computer sotto forma di impulsi elettrici.

Il *mouse* è un dispositivo di puntamento: permette cioè, usando lo schermo, di puntare un elemento visualizzato sullo schermo attraverso un *puntatore*, rappresentato da una figura anch'essa visualizzata sullo schermo e che si muove in base ai movimenti del mouse su un piano. Mediante i pulsanti associati al mouse è possibile fare *clic* su di un elemento visualizzato, il che permette di trascinare l'elemento da un punto all'altro dello schermo e di compiere operazioni su di esso.

Infine, lo *scanner* è un dispositivo che permette di inserire, in formato digitale, immagini come dati di input per il computer. Il *microfono* permette invece di inserire suoni, e la *telecamera* consente l'inserimento di filmati.

7.2 Periferiche di output

Le principali periferiche di output sono:

- monitor (o schermo)
- stampante
- casse acustiche

Il *monitor* o *schermo* è il dispositivo che permette di visualizzare i dati provenienti dal computer sotto forma di immagini. Ogni immagine è nel computer scomposta in un insieme di punti luminosi, detti *pixel*: l'intensità luminosa ed il colore di ogni pixel corrisponde ad un numero, rappresentato in formato binario nel computer. Tale numero viene tradotto dal monitor nel segnale luminoso corrispondente.

La *stampante* è un dispositivo che permette di stampare i dati provenienti dal computer, e che rappresentano testo e/o immagini, su carta. Nel caso di informazioni di tipo testuale, la stampante svolge un'operazione inversa della tastiera, ovvero converte le sequenze di bit provenienti dal computer in caratteri alfanumerici, quindi stampa tali caratteri su carta.

Infine, le *casse acustiche* convertono sequenze di bit in suoni, ovvero svolgono l'operazione inversa del microfono. Come nel caso delle immagini, anche i suoni nel computer sono scomposti in una sequenza di campioni, dove ogni campione è un suono di durata istantanea; le caratteristiche sonore di ogni campione sono rappresentate da un numero (in formato binario). Un suono viene così trasformato in una sequenza di numeri binari, cioè una sequenza di bit.

7.3 Memorie di massa

Le *memorie di massa* sono dispositivi per la memorizzazione permanente di programmi e dati all'interno del computer.

Nella maggior parte dei casi, le memorie di massa possono essere considerate dispositivi sia di input che di output, in quanto quasi sempre possono essere sia lette che scritte dalla CPU. Le principali memorie di massa sono:

- hard disk
- floppy disk
- CD-ROM
- CD-RW

Tra queste, solo il CD-ROM è un dispositivo di solo input, in quanto non può essere scritto dal computer, mentre tutti gli altri sono dispositivi di input e output.

Tali tipi di memorie di massa sono costituiti da un dispositivo di lettura/scrittura e da un supporto di memorizzazione, formato da uno o più dischi rotanti, ricoperti da un supporto magnetico o ottico in grado di memorizzare informazione.

I due parametri principali che misurano le qualità una memoria di massa sono:

- capacità di immagazzinare dati
- velocità di lettura/scrittura dati

Come per la RAM, la capacità di memorizzazione è misurata in multipli di byte (MegaByte, GigaByte). La velocità di lettura/scrittura dati è rappresentata dal tempo medio di accesso ad un singolo dato, e dal numero di byte al secondo che la memoria di massa è in grado di leggere/scrivere.

La memoria di massa più importante è l'hard disk, che ha grandi capacità di memorizzazione (dai 10 ai 100 GigaByte nei personal computer attuali) e grandi velocità di accesso ai dati. Tuttavia, l'hard disk ha lo svantaggio di non essere removibile, cioè non può essere facilmente rimosso dal computer e spostato su un altro computer. Il floppy disk ha il vantaggio che il supporto di memorizzazione è removibile, ma ha bassa capacità di memorizzazione (1.44 MegaByte) e bassa velocità di accesso ai dati. Il CD-ROM ha anch'esso supporto removibile, ha una capacità di 650 MegaByte, ma è un dispositivo di sola lettura. Tale limitazione è superata dal CD-RW, o CD riscrivibile.