

## 1 Lezione del 24-02-25

### 1.1 Introduzione al corso

Continuiamo lo studio di una particolare architettura per calcolatori, a partire da quanto detto riguardo alle reti logiche, introducendo i concetti di **interruzione**, **protezione** e **memoria virtuale**. Questi 3 strumenti ci permetteranno di realizzare il paradigma della **multiprogrammazione**, cioè di far eseguire ad una macchina con un singolo processore più programmi contemporaneamente. Non si pensi questo significhi avere più processori, in quanto il corso riguarda esclusivamente processori *single-threading*.

### 1.2 Architettura

L'architettura di riferimento è quella classica, composta da **CPU**, **memoria** e **I/O** interconnessi da un **bus**.

Durante lo studio di un'architettura è opportuno porsi la domanda "*chi fa cosa?*", che fornisce determinati *chi* ai determinati *cosa* forniti da un opportuno livello di astrazione (transistor, porte logiche, diagrammi funzionali, ecc...).

La domanda che potremo porci adesso è "*chi comanda?*" all'interno dell'architettura vista. La risposta più giusta è quella del **software**: l'architettura è fatta per *eseguire* software.

Per convincerci di questo possiamo sostituire la domanda "*chi fa cosa?*" con la domanda "*chi sa cosa?*".

- La **CPU** conosce lo stato corrente dei registri e l'istruzione in esecuzione. Fra un'istruzione e l'altra non c'è alcun bisogno di sapere cosa è accaduto finora, e cosa accadrà in futuro, ma solamente l'istruzione corrente. Quindi si può pensare che la CPU non *sa* qual'è l'obiettivo della computazione, ma si limita a portarla avanti.
- La **memoria** è un oggetto passivo, che contiene il programma, ma si limita a restituire i dati richiesti quando sono richiesti. Notiamo che le memorie che usiamo sono ad **accesso casuale**, ergo nessuno scorre alla ricerca di indirizzi, ma si può leggere e scrivere in posizioni arbitrarie in tempo pressoché costante. La memoria contiene **sempre** qualcosa, che questo sia significativo o meno, e la sua tipizzazione dipende solamente dalle intenzioni del programmatore.
- L'**I/O** è il componente più variegato dell'architettura. L'unica costante che rende la comunicazione con le periferiche più facile è la presenza di un interfaccia, che riduce tale comunicazione ad una semplice lettura o scrittura nello spazio di I/O. La differenza fra le letture e scritture nello spazio di I/O e lo spazio di memoria è la possibile presenza di **effetti collaterali**, cioè effetti non riconducibili alla sola variazione di stato di una locazione di memoria. Inoltre la CPU non è l'unica a scrivere nello spazio di I/O, in quanto questo può essere fatto anche dalle periferiche stesse.
- Il **bus** è un insieme di linee (*fili*), che trasportano ciò che ogni componente sta comunicando in un dato momento. Ogni componente vede ciò che viene scritto sul bus in qualsiasi momento, e l'indirizzamento di locazioni specifiche nello spazio di memoria o nello spazio di I/O viene fatto attraverso **maschere** di indirizzo.

### 1.2.1 Flusso di controllo

Abbiamo visto come la CPU si limita a prelevare ed eseguire istruzioni nel ciclo di **fetch-execute**. L'istruzione successiva alla corrente, il cui indirizzo viene scritto nell'**instruction pointer**, viene decisa dall'istruzione corrente stessa (si pensi alle istruzioni di salto). Il **flusso di controllo** è quindi deciso dall'istruzioni stesse, cioè dal programma.

### 1.2.2 Bootstrap

Il **bootstrap** è un processo secondo il quale si porta il sistema in un certo stato di esecuzione, apparentemente impossibile, o comunque molto difficile, da raggiungere. Ad esempio, il compilatore del linguaggio C è scritto esso stesso in linguaggio C. La domanda naturale è "*come è stato compilato il compilatore?*". La risposta è un processo di bootstrap, usando o un compilatore presistente, magari che implementa un sottoinsieme parziale del C, o scrivendo l'intero compilatore in linguaggio macchina, cioè assemblando codice assembler.

Il bootstrap si rende necessario anche all'avvio del calcolatore, per il caricamento del programma all'interno della memoria e l'inizio dell'esecuzione. Nei calcolatori moderni questo viene fatto attraverso la **ROM**, cioè una memoria a sola lettura che contiene un programma di bootstrap. All'avvio il processore è impostato in modo che al reset prenda come indirizzo proprio quello della ROM, e quindi inizi ad eseguire il programma di bootstrap. All'interno della ROM si trova, nei calcolatori moderni, il **BIOS** (o **UEFI**, nei sistemi moderni), che ha il solo compito di impostare alcune periferiche di base e caricare il sistema operativo.

Iniziamo quindi ad approfondire, uno per uno, i moduli dell'architettura.

## 1.3 Memoria

La memoria è un insieme contiguo di locazioni di memoria, che nelle architetture moderne sono rappresentate da byte. Storicamente, la memoria era indirizzata a *parole*, cioè insiemi di bit coincidenti in dimensioni coi registri del processore. Una parola poteva essere di più byte, mentre oggi le memorie sono accessibili ai singoli byte. Ad esempio, le memorie usate nell'architettura Intel x86 sono accessibili ad 1 byte (**MOVB**), 2 byte (**MOVW**), 4 byte (**MOVL**), e 8 byte (**MOVQ**).

### 1.3.1 Endianess

Notiamo che la posizione in memoria del byte più significativo di una parola (in questo caso consideriamo una "parola" da 8 byte, da cui si ricavano tutte le altre misure) determina l'*endianess* dell'architettura. In particolare, se l'ultimo byte sta in fondo nella memoria, si dice **big-endian**, mentre se viceversa l'ultimo byte viene per primo nella memoria, si dice **little-endian**.

L'architettura Intel x86 che andiamo a considerare è little-endian, come lo sono la maggior parte delle architetture moderne. Un esempio di utilizzo del big-endian è nella trasmissione di dati attraverso il protocollo IP, usato nelle comunicazioni Internet.

### 1.3.2 Allineamento

Indicheremo con **offset** la distanza in byte fra due locazioni di memoria, inteso come il numero di locazioni che vanno saltate per raggiungere un indirizzo a partire dall'altro. In questo ha senso parlare anche di offset *negativi*.

Visto che lo spazio di memoria è effettivamente ciclico, cioè si ha *wrap-around* ai suoi capi, si ha che gli offset rimangono validi **modulo** la dimensione dello spazio di memoria, che è sempre  $2^n$ , con  $n$  nel nostro caso uguale a 64.

Il *wrap-around* si comporta bene con gli offset, ma lo stesso non si può dire per quanto riguarda **intervalli** di byte. Preso un certo intervallo  $[x, y)$ , quindi, si ha che questo contiene gli indirizzi  $\{n \mid x \leq n < y\}$ , ammesso che  $x < y$ , cosa che risulta falsa nel caso di intervalli che hanno *wrap-around*. Decidiamo di non considerare intervalli di questo tipo. Questo rende necessaria un'eccezione per intervalli che comprendono l'ultimo byte: in questo caso è concesso  $[x, 0)$ , con 0 che indica il fondo dello spazio di memoria.

Veniamo quindi all'**allineamento**. Dire che un indirizzo è allineato ad un numero  $n$  significa dire che quell'indirizzo è un multiplo di  $n$ . Chiaramente, conviene scegliere  $n$  potenze di 2. In questo caso, per riconoscere se un indirizzo è allineato a  $2^k$ , basta guardare i suoi primi  $k$  bit.

Si dice spesso che oggetti sono *allineati alla parola*, ecc... Questo significa che sono allineati alla *dimensione* della parola specificata. Altrimenti, si può dire che un oggetto è allineato *naturalmente*, nel caso in cui sia allineato alla dimensione di stesso.

Infine, il **confine** di un oggetto è l'indirizzo che lo delimita dal resto dello spazio di memoria.

## 2 Lezione del 25-02-25

### 2.1 Interazione fra CPU e memoria

Nell'architettura Intel x86 la CPU interroga la RAM in due situazioni:

- Durante la lettura di un'istruzione;
- Durante la lettura di *eventuali* operandi in memoria richiesti dall'istruzione. Notiamo che per ogni istruzione è previsto un solo indirizzo esplicito di un operando in memoria (non è permesso scrivere qualcosa come `MOV (%RBP), (%RDI)`). indirizzo Alcune istruzioni possono però avere comunque più di un operando in memoria (ad esempio le istruzioni di stringa, `MOVS`, ecc... o la stessa istruzione di pila `POP`).

Dal punto di vista pratico, il collegamento fra CPU e RAM è rappresentato da:

- Un **bus dati** a 64 bit;
- Un certo numero di linee per il **numero di riga**. Questo non corrisponde all'indirizzo del primo byte contenuto in ogni riga, ma l'indice proprio di ogni regione (intesa come riga) da 64 bit all'interno della RAM. Si noti inoltre che queste non sono necessariamente  $2^{64}$ , o  $2^{57}$  (il massimo spazio indirizzabile secondo l'architettura x86), ma più spesso intorno alle  $2^{36}$ - $2^{37}$ ;
- Determinate **linee di controllo** che segnalano l'operazione in corso da parte del processore.
- 8 linee di **byte enable**, attive basse, che rappresentano i byte di interesse all'interno di ogni locazione da 64 bit della RAM. Dal punto di vista della lettura, queste linee non sono particolarmente utili in quanto tutta la locazione verrà comunque riportata sul bus dati, o comunque le locazioni non selezionate potranno essere invalide o in alta impedenza, senza avere effetto sulla CPU (che non le leggerà). Per quanto riguarda la scrittura, invece, la RAM lascerà inalterati i byte con byte enable alto.

### 2.1.1 Struttura della RAM

Modellizziamo un modulo di RAM come una rete provvista di:

- Una linea di **select**, attiva bassa;
- Le **linee di indirizzo**;
- Una linea di *memory read* e una linea di *memory write*, o comunque un certo numero di **linee di controllo** necessarie all'accesso in scrittura e lettura;
- Un **bus dati** di ingresso/uscita.

Dalla CPU arriveranno, come abbiamo detto, i **numeri di riga**, i **byte enable**, il **bus dati** e le **linee di controllo**.

I numeri di riga si collegano direttamente alle linee di indirizzo di ogni modulo, che rappresenterà un certo byte della locazione (avremo quindi, nell'architettura descritta, 8 moduli per 8 byte, quindi 64 bit). I byte enable dovranno quindi smistarsi nelle linee di select di ogni modulo di RAM, a selezionare il modulo corrispondente. Il bus dati verrà composto, analogamente, concatenando le linee di uscita da 8 bit di ogni modulo di RAM. Notiamo che avevamo chiamato questo montaggio **parallelo**.

Vorremo poter estendere la memoria disponibile oltre il numero di locazioni da un byte fornite da ogni modulo di RAM. Pensiamo di fare questo attraverso più banchi di memoria con locazioni da 64 bit. In questo caso avremo bisogno di montaggio in **serie**, e quindi di generare un segnale di select a partire non solo dalle linee di byte enable, ma anche da una **maschera** generata a partire dal numero di riga. Questo si potrà fare agevolmente mettendo il segnale di uscita della maschera in OR (ricordiamo segnali attivi bassi, quindi si applica De Morgan) con il byte enable di ogni modulo di RAM compreso nel banco di memoria associato a tale maschera.

### 2.1.2 Allineamento e RAM

Quanto discusso finora rende più chiaro l'importanza del corretto allineamento degli oggetti in memoria. Leggere un oggetto da 8 byte non allineato nel montaggio di RAM descritto, infatti, richiederà necessariamente 2 accessi, contro il singolo accesso necessario per un oggetto allineato. Inoltre, alcuni dei byte più significativi risulteranno invertiti di posto rispetto ai byte meno significativi, cioè si richiede un operazione di shift interna al processore.

Questa combinazione di operazioni, eseguite in **hardware**, rende gli accessi in memoria non allineati molto poco performanti, e quindi sconsigliati (anche se l'architettura Intel x86 li permette comunque).

Un problema che potrebbe interessarci è, data una regione di memoria  $[x, y]$  di dimensione  $b$  uguale a un singolo banco di RAM, ottenere gli indici della prima regione in cui cade l'intervallo, e la prima in cui non cade più.

Vediamo come calcolare la prima regione di appartenenza. In **hardware**, questo può essere calcolato semplicemente prendendo gli  $n - b$  bit più significativi dei numeri di riga  $x$  e  $y$ .

In **software**, questo equivarrà ad uno shift a destra che conservi i soli  $n - b$  bit più significativi.

Vediamo come calcolare l'offset di  $x$  o  $y$  all'interno delle rispettive regioni. Mascherando gli stessi bit, invece, si può ottenere l'indirizzo all'interno del banco del confine

della regione. Per la precisione, vogliamo una maschera fatta da  $n - b$  0 e  $b$  1. Questa si può ricavare agevolmente prendendo  $2^b$  come `1UL << b` e sottrandogli 1, ottenendo la maschera desiderata (si avranno borrow propagati dal bit in  $b$  fino al LSB).

Infine, vediamo come calcolare la prima regione di non appartenenza. In questo caso potremo calcolare la regione in cui cade  $y - 1$ , e aggiungervi 1 (tenendo conto di eventuali *wrap-around*). Il  $-1$  è richiesto dal fatto che  $y$  potrebbe cadere sul confine. In questo caso avremo  $((y - 1) >> b) + 1$ , considerata somma modulo  $n - b$ . Alternativamente, si può prendere  $y + b$  e calcolarne la regione di appartenenza.

## 2.2 Spazio di I/O

Veniamo quindi alla trattazione dello spazio di I/O e delle interfacce ivi connesse. L'accesso alle periferiche viene fatto attraverso le istruzioni **IN** e **OUT**, ammesso che non ci sia nessun sistema operativo in esecuzione, ma solo il nostro programma, e appositi sottoprogrammi di ingresso/uscita, la cui struttura non è al momento importante.

Le periferiche che studieremo, per semplicità di trattazione, derivano in parte da quelle disponibili sui PC **IBM AT** (famiglia *IBM 5170*). I PC di questa categoria (compresi tutti i vari *IBM compatible*) si basavano sullo standard per periferiche **ISA** (*Industry Standard Architecture*). Visto che i PC moderni derivano dai vecchi IBM compatible, anche oggi si cerca di emulare (almeno in parte) questo standard.

Le periferiche, nello specifico saranno:

- La **tastiera**;
- Il **video** su VGA;
- Il **timer**;
- Gli **hard disk**.

## 2.3 Tastiera

Dal punto di vista funzionale, la tastiera deve solo scoprire quali tasti sono premuti e comunicarlo al calcolatore. In particolare, noi studieremo tastiere IBM che trasmettono secondo lo standard PS/2.

Nei PC IBM il tasto non restituisce il carattere ASCII del carattere premuto, ma un codice associato ad ogni tasto che va convertito in software. Questo codice viene ottenuto per *scansione* dell'intero piano della tastiera. Dal punto di vista meccanico, ci sono **tracce** orizzontali e verticali disposte, rispettivamente, su ogni riga o colonna di tasti. La pressione di un tasto comporta una deformazione delle tracce che chiude un circuito fra la riga e la colonna del tasto corrispondente. Un **microcontrollore** (originariamente un Intel 8042) collegato sia alle tracce orizzontali che alle tracce verticali scansiona ciclicamente, con impulsi, o le righe leggendo le colonne, o le colonne leggendo le righe, cercando un circuito chiuso. Un cortocircuito viene quindi rilevato dal microcontrollore, che aggiorna una (piccola) memoria interna con il tasto premuto. Di conseguenza, invia al calcolatore un segnale che codifica quali tasti sono stati premuti rispetto al precedente istante temporale, e quali tasti sono stati rilasciati rispetto al precedente istante temporale.

La tastiera non restituisce solo pressioni di tasti, ma anche i loro rilasci, cosa che può essere utile per ottenere combinazioni di tasti, pressioni estese nel tempo, ecc... I codici di pressione si dicono **make code**, mentre i codici di rilascio si dicono **break code**. La

stessa pressione ripetuta di un tasto quando l'utente lo tiene premuto per un certo istante temporale era, nei PC IBM, realizzata direttamente nella tastiera (tecnologia *type-matic*), tra l'altro con periodo configurabile. Tramite il *type-matic*, su appositi tasti abilitati, si ha infatti una ripetizione dell'evento di *pressione* (non rilascio) di un tasto a frequenza costante dopo un intervallo di pressione continua.

Lato calcolatore, il segnale prodotto dal microcontrollore della tastiera viene letto da un interfaccia provvista dei seguenti registri:

- **RBR**, *Receive Buffer Register*;
- **TBR**, *Transmit Buffer Register*;
- **STR**, *Status Register*;
- **CMR**, *Command register*

RBR e TBR, come STR e CMR, condividono gli indirizzi, rispettivamente 0x60 e 0x64. Il RBR conterrà i make e break code, mentre l'STR conterrà i flag di stato sia per RBR che per TBR (rispettivamente ai bit 0 e 1).

Potremmo chiederci il significato di un registro di trasmissione TBR. Questo serve, ad esempio, a governare i led di stato per funzioni speciali quali Caps-Lock, Num-Lock, Scroll-Lock ecc... nonché a modificare le impostazioni del type-matic e, in maniera completamente slegata alla tastiera, a provocare il reset del PC, scrivendo 0xFE in CMR.

Vediamo quindi un programma C++ per l'interazione con l'interfaccia di tastiera. Notiamo che la libreria all'header `libce.h` definisce alcuni tipi (qui `natb`, un naturale su 8 bit, e `ioaddr`, un indirizzo nello spazio di I/O) e funzioni (qui `inputb`, ottieni byte dallo spazio di I/O, e `vi::char_write()`, stampa un carattere a schermo).

```

1 #include <libce.h>
2 #define NUM_CODES 28
3
4 // indirizzi porte tastiera
5 const ioaddr rbr_addr = 0x60;
6 const ioaddr str_addr = 0x64;
7
8 // tabella make code
9 natb make_codes[] = {
10     0x10, 0x11, 0x12, 0x13, 0x14, 0x15, 0x16, 0x17, 0x18, 0x19,
11     0x1e, 0x1f, 0x20, 0x21, 0x22, 0x23, 0x24, 0x25, 0x26,
12     0x2c, 0xd, 0x2e, 0x2f, 0x30, 0x31, 0x32,
13     0x1c, 0x39
14 };
15
16 // tabella caratteri minuscoli
17 char l_table[] = {
18     'q', 'w', 'e', 'r', 't', 'y', 'u', 'i', 'o', 'p',
19     'a', 's', 'd', 'f', 'g', 'h', 'j', 'k', 'l',
20     'z', 'x', 'c', 'v', 'b', 'n', 'm',
21     '\n', ' '
22 };
23
24 // tabella caratteri maiuscoli
25 char u_table[] = {
26     'Q', 'W', 'E', 'R', 'T', 'Y', 'U', 'I', 'O', 'P',
27     'A', 'S', 'D', 'F', 'G', 'H', 'J', 'K', 'L',
28     'Z', 'X', 'C', 'V', 'B', 'N', 'M',
29     '\n', ' '

```

```
30 };
31
32 // make code ESC
33 natb esc_code = 0x01;
34
35 // make code shfit
36 natb shift_down = 0x2A;
37 natb shift_up = 0xAA;
38
39 // gestione case
40 enum cas { lower, upper };
41 cas cur_cas = lower;
42
43 natb get_key() {
44     natb status;
45
46     // ciclo di lettura per il flag FI in str_addr
47     do {
48         status = inputb(str_addr);
49     } while(!(status & 0x01));
50
51     // FI alto, leggi da rbr_addr
52     return inputb(rbr_addr);
53 }
54
55 char get_char(natb make_code) {
56     // cerca il carattere per scansione lineare
57     for(int i = 0; i < NUM_CODES; i++) {
58         if(make_code == make_codes[i]) {
59             // trovato, controlla il case corrente
60             if(cur_cas == upper) {
61                 return u_table[i];
62             } else {
63                 return l_table[i];
64             }
65         }
66     }
67
68     // carattere nullo come default
69     return '\0';
70 }
71
72 void main() {
73     while(true) {
74         // ottieni make code
75         natb make_code = get_key();
76
77         // se ESC, esci
78         if(make_code == esc_code) {
79             break;
80         }
81
82         // gestisci shift
83         if(make_code == shift_down) {
84             cur_cas = upper;
85         }
86         if(make_code == shift_up) {
87             cur_cas = lower;
88         }
89     }
```

```
90 // ottieni carattere e stampa
91 char c = get_char(make_code);
92 vid::char_write(c);
93 }
94 }
```

Dal programma si evincono subito gli indirizzi dei registri di Receive Buffer (RBR) e di stato (STR), a cui i registri trasmettitore e comando (TBR e CMR) sono sovrapposti. Il funzionamento è quindi ottenuto attraverso una lettura ciclica dei make code dall'interfaccia, e una scansione per il rilevamento del carattere selezionato a partire dal make code stesso. Altro codice è usato per gestire il tasto shift, e il termine dell'esecuzione alla pressione del tasto ESC.

### 3 Lezione del 03-03-25

#### 3.1 Video

Il supporto principale al video è la **memoria video**, che lato software si comporta perlopiù come una normale memoria ad accesso casuale.

Questo è quindi il primo esempio di un oggetto che si trova nello spazio di memoria, senza necessariamente *essere* memoria: ciò che vi viene scritto non viene memorizzato, ma visualizzato sullo schermo.

Inoltre, la memoria video supporta un accesso *bidirezionale*: cioè vi si può accedere sia lato CPU che lato **adattatore video**, cioè la rete che si occupa di gestire tale memoria e visualizzarla sul *display*. Lo standard VGA usato dal PC IBM prevede che l'adattatore sia configurabile e utilizzabile in due modalità:

- **Modalità testo:** ogni locazione viene associata ad un carattere ASCII da visualizzare sullo schermo, diviso in 80 colonne  $\times$  25 righe (con un carattere di  $9 \times 16$  pixel, per una risoluzione totale di  $720 \times 400$  pixel a 70 Hz). È questa la modalità di default in cui si avvia l'adattatore.

In questo caso il compito dell'adattatore è quello di leggere i 4 KB di memoria, e convertire ogni codice nel carattere principale. Questo viene fatto consultando una ROM di caratteri che contiene quello che è effettivamente il carattere (*font*) dell'adattatore. Solitamente si può anche redirezionare la lettura in ROM ad una certa regione della RAM, modificando così il font.

La faccenda è veramente più complicata: si dedicano non 1 ma 2 byte ad ogni carattere, dove il byte più significativo rappresenta informazioni riguardo al **colore** del carattere:

- I 4 bit meno significativi rappresentano il colore del *foreground*;
- I 3 bit successivi rappresentano il colore del *background*;
- Il bit più significativo rappresenta il *blinking*, cioè indica all'adattatore di far lampeggiare quel carattere nel tempo.

La modalità testo non ha idea della posizione del cursore sullo schermo: attraverso registri si può indicare la posizione del cursore, e modificando la regione di memoria interessata si possono cambiare i caratteri in qualsiasi zona dello schermo. Il comportamento del cursore (spostamento, ritorno a capo, ritorno carrello, ecc...) è quindi gestito interamente lato software.



- **Modalità grafica:** programmando i registri dell'adattatore si possono ottenere diverse modalità grafiche, che permettono al programmatore di colorare singoli pixel sul display. Le modalità più popolari di questo tipo su schede VGA erano o a  $640 \times 400$  pixel a 70 Hz (fase di boot grafica), o a  $640 \times 480$  pixel a 60 Hz non interlacciati (la modalità di default di Microsoft Windows). Nella macchina virtuale usata incapsuliamo tale operazione di conversione in un'apposita libreria, e scriviamo pixel con colori su 8 bit (per 256 colori diversi). Nei sistemi moderni la memoria video non viene scritta dalla CPU, ma da un *coprocessore grafico* che esegue un suo programma, mentre la CPU può dedicarsi ad altro.

### 3.1.1 Indirizzamento dei registri dell'adattatore video

Vediamo nel dettaglio come si possono indirizzare i registri interni dell'adattatore video. Questo dispone infatti di una vasta gamma di registri, ma una sola linea di ingresso da un byte per indirizzamento e scrittura. Le scritture vengono quindi eseguite in serie:

- Prima specificando l'**indirizzo** del registro da aggiornare;
- Poi inserendo i **dati** da scrivere a tale indirizzo.

impulso

Vediamo quindi un programma di esempio che sfrutta l'adattatore grafico in modalità testo a  $25 \times 80$  caratteri. Dalla libreria `libce` si è importata la variabile `video`, che rappresenta l'intero buffer da 4 KB di memoria video (2 byte per cella) a disposizione dall'adattatore. Le scritture vengono fatte, come avevamo detto, sovrascrivendo dati in tale buffer, mantenendo *lato software* un cursore che ci indica la posizione sullo schermo.

```

1 #include <libce.h>
2 // dimensioni memoria video (2K)
3 #define SIZE 4000
4
5 namespace vid {
6     // dichiara l'array video di libce
7     extern volatile natw* video;
8 }
9
10 #define video vid::video
11
12 char mess[] = "- x86 rules ";
13 int cursor = 0;
14
15 // attributi carattere: 0x00001111-ASCII-
16 // significa bianco su sfondo nero
17 natl attr = 0x0F00;
18
19 // stampa una stringa
20 void prt_string(char* mess) {
21     while(*mess != '\0') {
22         video[cursor] = *mess | attr;
23         cursor = (cursor + 1) % SIZE;
24         mess++;
25     }
26 }
27
28 void main() {
29     for(int i = 0; i < 167; i++) {
30         prt_string(mess);

```

```

31 }
32
33 do {
34     // esci su ESC
35     natb k = kbd::get_code();
36     if(k == 0x01) break;
37 } while(true);
38 }

```

## 3.2 Timer

Il timer è realizzato come un interfaccia ad eventi, che riceve in ingresso un clock e aggiorna ciclicamente un registro contatore. Al raggiungimento di 0 da parte del contatore, si resetta e si invia un certo evento (un impulso).

Nel PC IBM in particolare troviamo 3 contatori:

- **Contatore 0:** è collegata al controllore delle interruzioni;
- **Contatore 1:** era storicamente usato per il refresh della RAM, oggi non viene più usato;
- **Contatore 2:** era collegato all'unico dispositivo audio presente sull'IBM, cioè il beeper speaker.

Il timer dispone di una vasta gamma di parametri e di solo 2 piedini di indirizzo, quindi 4 locazioni nello spazio di I/O, che sono **CWR**, un registro comune di configurazione delle interfacce, più 4 registri per timer collegati ad ognuna delle 3 porte rimanenti. L'accesso ai registri di un singolo timer va quindi fatto, cosa interessante del chip, in serie, modificando la stessa porta 1 o 2 volte consecutivamente.

### 3.2.1 Sonoro

Vediamo in particolare il lato sonoro del PC IBM. Essendo stato questo un calcolatore pensato per l'uso da ufficio, le capacità audio erano molto limitate: si disponeva di un beeper speaker a frequenza modulabile dal timer (contatore 2). Inoltre, un particolare registro in memoria era collegato direttamente in AND con l'uscita del contatore 2, permettendo la modulazione on/off del segnale allo speaker. La modulazione in volume del pc speaker, in particolare, viene fatta attraverso un ulteriore registro detto **SPR**.

Questo tipo di modulazione permetteva effettivamente di sfruttare, in maniera non prevista dalla IBM, per riprodurre segnali generici.

Mostriamo un esempio di un programma per la riproduzione di un brano musicale monofonico, inteso come una sequenza di note:

```

1 #include "song.h" // include libce
2 #define TIMESTEP 3
3
4 // indirizzi timer
5 const ioaddr timer0_addr = 0x40;
6 const ioaddr timer2_addr = 0x42;
7 const ioaddr cwr_addr = 0x43;
8 const ioaddr spr_addr = 0x61;
9
10 // frame corrente di esecuzione
11 natl frame = 0;
12

```

```
13 // imposta il divisore di un timer
14 void set_divisor(natw divisor, ioaddr addr) {
15     // metti divisor in addr in 2 passate
16     outputb(divisor, addr);
17     outputb(divisor >> 8, addr);
18 }
19
20 // leggi il conteggio di un timer
21 natl read_timer(ioaddr addr) {
22     // comando di latch
23     outputb(0x00, cwr_addr);
24
25     natb low = inputb(addr);
26     natb high = inputb(addr);
27     return (high << 8) | low;
28 }
29
30 // abilita lo speaker
31 void note_on() {
32     outputb(3, spr_addr);
33 }
34
35 // disattiva lo speaker
36 void note_off() {
37     outputb(0, spr_addr);
38 }
39
40 bool update_song() {
41     song_frame cur_frame = song[frame];
42
43     switch(cur_frame.mode) {
44         case 0:
45             // off
46             note_off();
47             break;
48         case 1: {
49             // on
50             note_on();
51
52             // divisore della nota
53             natl divisor = cur_frame.get_divisor();
54             set_divisor(divisor, timer2_addr);
55             break;
56         }
57         case 2: {
58             // legato
59             natl divisor = cur_frame.get_divisor();
60             set_divisor(divisor, timer2_addr);
61
62             break;
63         }
64     }
65
66     frame++;
67
68     if(frame == length + 1) return false;
69     return true;
70 }
71
72 void main() {
```

```

73 // imposta il timer 0
74 // outputb(0x36, cwr_addr); // modo 3
75 natl div = 0; // 0 significa 65536
76 set_divisor(div, timer0_addr);
77
78 // imposta il timer 2
79 outputb(0xB6, cwr_addr); // modo 3
80
81 // i tick svolti
82 natl tick = 0;
83 natl next_song_update = 0;
84
85 natl last_value = read_timer(timer0_addr);
86
87 while (true) {
88     natl current_value = read_timer(timer0_addr);
89
90     // se il valore corrente e' maggiore del valore precedente
91     // si e' fatto un salto
92     if (current_value > last_value) {
93         tick++;
94     }
95
96     // aggiorna se necessario
97     if(tick > next_song_update) {
98         bool res = update_song();
99         if(!res) break;
100
101         // imposta il prossimo aggiornamento
102         next_song_update += TIMESTEP;
103     }
104
105     last_value = current_value;
106 }
107 }

```

In particolare, il programma si basa sullo scorrimento di un array di strutture `song_frame` (definite in `song.h`, la cui struttura non ci interessa) ad intervalli regolari. Questi intervalli vengono ottenuti configurando il timer 0 per oscillare in modalità 2 (oscillazione a onda quadra) con un divisore di 65536, che per un clock a 1.19 MHz fa:

$$\frac{1.19 \text{ MHz}}{65536} \approx 17.158 \text{ Hz}$$

Lato software, si aspettano 3 di queste oscillazioni, per ottenere una frequenza di  $\approx 6\text{Hz}$  (abbastanza vicina a quella delle crome a 120 BPM).

Una volta ottenuto il frame, quindi, si aggiorna il timer 2 di conseguenza, sfruttando il registro SPR per il volume della nota (qui solo on/off) e il suo registro di scrittura per la nota stessa.

## 4 Lezione del 04-03-25

### 4.1 Hard disk

Gli **hard disk** (*dischi rigidi*) sono effettivamente, seppur memorie, **periferiche**, collegate al bus attraverso la loro interfaccia. La CPU non puo' eseguire programmi direttamente dall'hard disk, ma deve prima caricarli in memoria principale (memoria RAM).

Questo perchè letture e scritture in hard disk vengono effettuate per **blocchi** (storicamente di 512 byte), e richiedono molto più tempo di quanto sia possibile aspettare al prelievo di istruzioni o operandi.

Dal punto di vista elettromeccanico venivano realizzati attraverso dischi di materiale ferromagnetico imperniati ad un asse centrale, con testine mobili che scandivano il raggio dei dischi, rilevando o modificando la loro magnetizzazione per accedere all'informazione. Il complesso di dischi e testine viene detto **drive**.

L'informazione viene disposta su ogni disco in **settori** e **tracce**. Le tracce sono concentriche e i settori formano degli "spicchi" di ogni faccia. Notiamo che entrambe le facce di ogni disco possono memorizzare informazione. Un **blocco** è quindi formato dalla regione di una traccia compresa in un certo sensore.

I dischi vengono tenuti continuamente in rotazione (negli ordini delle centinaia/migliaia di RPM). Il tempo che la testina impiega a raggiungere una traccia viene detto **tempo di seek**,  $t_{seek}$ , il tempo che alla velocità di rotazione del disco l'informazione si trovi sotto la testina **latenza**  $t_{latency}$  e il tempo necessario ad effettuare l'operazione vera e propria **tempo di lettura/scrittura**  $t_{r/w}$ , per cui il tempo di lettura/scrittura complessivo risulta:

$$t_{seek} + t_{latency} + t_{r/w} \sim 1 \text{ ms}$$

nell'ordine del millisecondo, per la CPU estremamente (milioni di volte) più lento della RAM.

Quello che accade al tempo di lettura è che il blocco viene copiato in un buffer di memoria nell'interfaccia che viene poi reso disponibile alla CPU. Viceversa, al tempo di scrittura il buffer viene riempito dalla CPU, e l'interfaccia si occupa poi di copiarlo all'interno del settore giusto.

Per effettuare un operazione dobbiamo quindi sapere:

- Quale *testina* individuare;
- Quale *traccia* individuare;
- Quale *regione* (quindi quale *blocco*) individuare.

Storicamente queste informazioni erano gestite lato software, concedendo la possibilità di alterare la *formattazione* del disco. Oggi la formattazione è definita in fabbrica, e l'interfaccia offre una sua astrazione. In questa astrazione ogni blocco è quindi indirizzato da un indirizzo logico, il **Logical Block Address, LBA**.

#### 4.1.1 Interfaccia ATA

Nello standard PC AT gli hard disk usano interfacce **ATA** (capaci di gestire 2 drive, in configurazione *master/slave*). L'interfaccia ATA è dotata di diversi registri a 8 bit e uno a 16 bit:

- **Registri di selezione** del blocco:
  - **SNR** (Sector Number);
  - **CNL** (Cylinder Number Low);
  - **CNH** (Cylinder Number High);

- **HND** (Head And Drive): solo gli ultimi 4 bit di questo registro formano l'informazione sulla testina da utilizzare. Gli altri bit vengono usati diversamente, ad esempio per selezionare quale drive usare in configurazioni master/slave, o per abilitare il LBA, usando quindi i registri di selezione per specificare un indirizzo logico (su  $3 \cdot 8 = 4 = 28$  bit) anziché un'informazione geometrica sulla posizione del blocco desiderato.

Vediamo che dalla dimensione dell'LBA (assumiamo che per indirizzamento geometrico si trova la stessa cosa) si ha una dimensione del disco:

$$2^{28} \cdot 2^9 = 2^{37} = 128 \text{ GB}$$

Per questo si può abilitare la modalità **LBA48** (che non è un gruppo di idol giapponesi), dove ci si aspetta il LBA venga specificato in due passate, una da 24 bit e una da 20 bit sugli stessi registri.

- **SCR** (Section Counter): permette di specificare su quanti settori contigui a partire da quello specificato prima eseguire l'operazione;
- **BR** (Buffer Register): l'unico registro a 16 bit, permette di accedere al buffer 2 byte alla volta;
- **STS** (Status Register): il classico registro di stato che ci notifica se un'operazione è conclusa o si può effettuare;
- **CMD** (Command): serve a specificare l'operazione da effettuare (lettura, scrittura, ecc...).

Vediamo quindi un ultimo programma di esempio delle periferiche, che permette di scrivere un buffer di caratteri da 512 byte, stampandolo a schermo, e scriverlo/leggerlo su un settore di memoria ad un indirizzo LBA (prendiamo 1).

```

1 #include <libce.h>
2 #include "keyboard.h"
3 #include "video.h" // definisce il buffer video
4 #define BUF_SIZE 512
5
6 // registri disco
7 const ioaddr disk_buffer = 0x01F0;
8 const ioaddr disk_status = 0x01F7;
9 const ioaddr disk_sectors = 0x01F2;
10 const ioaddr disk_command = 0x01F7;
11
12 // registri indirizzo LBA (sarebbero SNR CNL CNH HND)
13 const ioaddr disk_lba0 = 0x01F3;
14 const ioaddr disk_lba1 = 0x01F4;
15 const ioaddr disk_lba2 = 0x01F5;
16 const ioaddr disk_lba3 = 0x01F6;
17
18 // dai indirizzo LBA al controllore disco
19 void give_lba(nat1 lba) {
20     // dividi in 4 byte
21     natb lba0 = lba;
22     natb lba1 = lba << 8;
23     natb lba2 = lba << 16;
24     natb lba3 = lba << 24;
25

```

```
26 // il byte piu' significativo deve attivare l'LBA,
27 // lba stava comunque su 28 bit
28 lba3 = (lba3 & 0x0F) | 0xE0; // 1110-LBA-
29
30 outputb(lba0, disk_lba0);
31 outputb(lba1, disk_lba1);
32 outputb(lba2, disk_lba2);
33 outputb(lba3, disk_lba3);
34 }
35
36 // dai comando al controllore disco
37 void give_command(natl lba, natb sectors, natb cmd) {
38     give_lba(lba);
39     outputb(sectors, disk_sectors);
40     outputb(cmd, disk_command);
41 }
42
43 // aspetta il disco
44 void wait_for_disk() {
45     natb s;
46     do {
47         s = inputb(disk_status);
48     } while ((s & 0x88) != 0x08);
49 }
50
51 // scrivi un settore sul disco
52 void write_sector(natb* sector) {
53     wait_for_disk();
54
55     // reinterpret_cast per mandare 2 byte per volta (ripetuti a 256 * 2 =
56     // 512)
57     outputbw(reinterpret_cast<natw*>(sector), 256, disk_buffer);
58 }
59
60 // leggi un settore dal disco
61 void read_sector(natb* sector) {
62     wait_for_disk();
63
64     // come sopra
65     inputbw(disk_buffer, reinterpret_cast<natw*>(sector), 256);
66 }
67
68 // make code salva (1) e carica (2)
69 const natb save_code = 0x02;
70 const natb load_code = 0x03;
71
72 // indirizzo lba disco
73 natl lba = 1;
74
75 // buffer testo
76 natb buffer[BUF_SIZE];
77
78 // svuota il buffer
79 void init_buffer() {
80     for(int i = 0; i < BUF_SIZE; i++) {
81         buffer[i] = 0x00;
82     }
83 }
84
85 // cursore buffer testo
```

```

85 natl cursor = 0;
86
87 // sposta il cursore senza uscire dal buffer
88 inline void mov_cursor(int d) {
89     if(cursor == 0 && d < 0) return;
90
91     cursor += d;
92     if(cursor >= BUF_SIZE) cursor = BUF_SIZE - 1;
93 }
94
95 // salva buffer testo
96 void save() {
97     give_command(lba, 1, hd::WRITE_SECT);
98     write_sector(buffer);
99 }
100
101 // carica buffer testo
102 void load() {
103     give_command(lba, 1, hd::READ_SECT);
104     read_sector(buffer);
105 }
106
107 void main() {
108     // inizia svuotando il buffer
109     init_buffer();
110
111     // vai in un ciclo di lettura
112     while(true) {
113         natb make_code = get_key();
114
115         if(make_code == esc_code) break;
116         if(make_code == back_code) mov_cursor(-1);
117
118         if(make_code == save_code) save();
119         if(make_code == load_code) load();
120
121         char c = get_char(make_code);
122         if(c != '\0') {
123             buffer[cursor] = c;
124             mov_cursor(1);
125         }
126
127         // aggiorna schermo
128         prt_screen(buffer, BUF_SIZE);
129         set_cursor(cursor);
130     }
131 }

```

Gli header `keyboard.h` e `video.h` contengono funzioni simili a quelle viste negli esempi precedenti per l'interfacciamento con tastiera e video (ci sono due funzioni video non viste, `prt_screen()` per la scrittura di tutto il buffer video, e `set_cursor()`, che imposta la posizione del cursore hardware agendo su registri specifici).

La scrittura viene effettuata alla pressione del tasto "1", e la lettura alla pressione del tasto "2". Entrambe le operazioni si riassumono fondamentalmente nell'invio di un comando (`give_command()`), che include la scrittura dell'indirizzo LBA (`give_lba()`), e nella successiva scrittura o lettura di un settore (`write_sector()` o `read_sector()`), che comprende di aspettare un certo bit di stato del disco (`wait_for_disk()`). Il bit particolare si può verificare consultando i manuali appositi.



Le funzioni viste finora su periferiche di I/O sono disponibili nella cartella `/code` degli appunti del corso, assieme a vari esperimenti e la definizione completa di tutti gli header usati. Si noti che questi non sempre corrispondono con `libce`, ma spesso riprendono, ridefiniscono o usano (probabilmente in maniera erranea) funzioni e oggetti ivi definiti.

## 4.2 Caching

Abbiamo detto che la memoria RAM è molto più veloce dei dischi rigidi. Questo è vero, ma non significa che non ci sia comunque un certo dislivello tra la velocità della CPU e la velocità della RAM: un'operazione può comunque richiedere nell'ordine dei  $\sim 100$  circa cicli di clock.

Per questo motivo si inframezzano fra la CPU e la RAM più memorie, relativamente piccole ma veloci, dette **memorie di cache**.

L'idea è che la RAM in sé è costituita da memoria dinamica (DRAM), quindi a condensatori, relativamente lenta e con tempo di refresh, mentre le memorie di cache vengono implementate con memorie statiche, più veloci ma più costose da realizzare su larga scala (per cui le dimensioni ridotte).

### 4.2.1 Principi di località

Le piccole dimensioni delle memorie vengono aidate dalla **località** del codice in memoria: istruzioni che compongono le stesse funzioni avranno istruzioni vicine fra di loro, le strutture definite dal programmatore conterranno dati locali, ecc... In particolare, potremo distinguere fra due **principi di località**:

- **Località temporale**: una volta visto un indirizzo, è probabile che questo o indirizzi ad esso vicini siano visti di nuovo;
- **Località spaziale**: solitamente si accede ad indirizzi vicini fra di loro.

La cache avrà quindi il compito di memoizzare i valori prelevati con frequenza dalla DRAM. Possiamo immaginare che la prima lettura di un dato richiederà il tempo completo di accesso, ma la lettura successiva, ammesso che quel dato sia stato salvato nella cache, richiederà un tempo di accesso significativamente minore.

L'importante è che questo processo sia **trasparente** per la CPU, cioè che questa non si debba preoccupare di quali indirizzi sono stati visti dalla cache e memoizzati e quali no. Il risultato finale è la velocizzazione di un qualsiasi programma senza dover agire in nessun modo sul programma stesso. Di contro, non è detto che il programmatore non possa sfruttare la presenza della memoria cache, cercando di sviluppare algoritmi e strutture dati che rispettano il più possibile i principi di località (tecniche *data driven*).

### 4.2.2 Cache ad indirizzamento diretto

Vediamo un primo esempio di memoria cache. Abbiamo che lato processore ci arriveranno le linee di byte enable (BE) e le linee di indirizzo (A). Inoltre avremo a disposizione un bus dati (D) di un certo numero di linee.

Vorremo porre fra CPU e DRAM una cache, connessa a quest'ultima dalle linee di indirizzo A. La memoria interna della cache, di dimensione complessiva 64 KB, sarà rappresentata da una serie di blocchi, o **cacheline** da 64 byte.

In fase di lettura, invece di leggere l'unica riga richiesta dal processore, si procederà alla lettura di un certo numero di righe (poniamo 8). Questo significa che per un tempo di lettura di riga di  $t$ , ci vorrà un tempo  $\sim 8t$  (solitamente meno). La speranza è che queste righe verranno lette successivamente dal processore.

Inoltre, ad ogni blocco di memoria letto dalla cache si dovrà associare dell'informazione riguardo alla posizione in memoria: questa viene contenuta in un'altra memoria, dette **memoria delle etichette**. E' quindi più conveniente leggere regioni relativamente più grandi di memoria, in modo da non sprecare *overhead* per piccole quantità di dati.

#### 4.2.3 Principio di funzionamento

La divisione della DRAM sulle cacheline è quindi realizzata giocando sulle scomposizioni degli indirizzi. Si divide ogni indirizzo in tre parti:

- L'**etichetta**, formata dai bit più significativi del bus;
- L'**indice**, formato dai 10 bit centrali (per indirizzare la totalità dei 64 KB di cache);
- L'**offset**, formato dai 3 bit meno significativi di A (per ottenere cache line da 64 byte, cioè  $8 = 2^3$  parole quaduple da 8 byte.).

Noto l'offset, l'**indice** verrà calcolato per indirizzare la totalità delle cacheline come stante su un numero di linee tali a:

$$\text{bit}_{\text{indice}} = \frac{\text{dimensione cache}}{\text{dimensione cacheline}}$$

Per ottenere la regione corrispondente ad un indirizzo (il numero di cacheline) si realizza una sorta di *funzione di hash*, prendendo l'etichetta e usandola come chiave per la regione di dati di indice corrispondente. Inoltre, alla regione selezionata si associa solitamente un singolo bit di validità. Un comparatore fra etichetta e gli  $n$  bit più significativi messo in AND a questo bit di validità ci assicurerà quindi la presenza nella cacheline del dato richiesto, detta **hit/miss**.

#### 4.2.4 Lettura

A questo punto, in fase di lettura, nel caso di hit basterà ricavare una linea di offset dai bit meno significativi di A, e leggere dalla memoria cache a tale offset, all'indice indicato dall'etichetta. Nel caso di miss si dovrà invece svolgere la lettura in memoria RAM, e poi riportare l'informazione nella cacheline di indice giusto della cache aggiornando l'etichetta.

#### 4.2.5 Scrittura

Per quanto riguarda le scritture invece, potremo muoverci in due strade: **write allocate** e **write no allocate**

- **Write allocate**: ci comportiamo in maniera simile alla lettura nel caso di hit. Nel caso di miss, invece, riportiamo il dato in cache.

A questo punto potremmo pensare di svolgere la scrittura in RAM e in cache contemporaneamente (regola *write-through*), mantenendo entrambe aggiornate.

Una tecnica più intelligente può invece essere quella di aggiornare il solo dato in cache, e rimandare la scrittura in RAM alla rimozione del dato dalla cache (per

l'introduzione di un nuovo dato allo stesso indice) (regola *write-back*). In questo caso dovremo dotarci di un nuovo bit nella memoria delle etichette, il bit *dirty*, che segnalerà il bisogno di ricopiare il dato in cache nella RAM in occasione del suo deallocaimento dalla cache. La difficoltà principale di questo metodo è l'avere un agente che non è la CPU che scrive in RAM, e come vedremo richiede soluzioni tecniche particolari.

- **Write no allocate:** in questo caso ignoriamo le scritture in cache e la sfruttiamo solamente per le letture.

Notiamo che questa cache soffre di problemi di **collisione**: infatti ci sarà un numero di regioni con lo stesso indice ed etichetta diversa, pari alla dimensione della RAM fratto la dimensione della cache.

## 5 Lezione del 07-03-25

Riprendiamo il discorso della memoria cache. Avevamo che questa è montata fra la CPU e lo spazio di memoria. Più propriamente, questa si trova fra la CPU e il bus.

In questo, può vedere non solo le operazioni sulla memoria, ma anche sullo spazio di I/O. In questo caso, però, dovrà ovviamente comportarsi sempre in maniera *read-through* e *write-through*, quindi effettivamente disattivarsi e lasciare che il processore interagisca direttamente con l'I/O.

Questo è dovuto al fatto che allo spazio di I/O potrebbero accedere e modificare dati dispositivi esterni alla CPU (le interfacce), operazione che invaliderebbe immediatamente qualsiasi cosa venga scritta in memoria cache.

Inoltre, ogni operazione di lettura può comportare di per sé un aggiornamento delle interfacce, che comporterà un aggiornamento della memoria, motivo per cui un'operazione di caching sarebbe superflua.

Operazione simile varrà effettuata per la memoria video (che non sta nello spazio di I/O). Questa facoltà verrà realizzata dalla cache attraverso, probabilmente, *maschere* o *tabelle*.

### 5.0.1 Cache associative ad insiemi

Avevamo visto come il difetto principale della cache ad indirizzamento diretto è quello delle *collisioni*. Presentiamo un metodo, quello delle **cache associative ad insiemi**, che risolve il problema permettendo di allocare più cacheline allo stesso indirizzo.

Duplichiamo quindi la struttura vista per la cache ad indirizzamento diretto (qui 2 volte), e sfruttiamo le uscite hit/miss delle singole memorie delle etichette per pilotare un multiplexer con in ingresso le linee dati delle memorie di cache corrispondenti.

In questo caso a letture allo stesso indice le cache potranno rispondere diversamente (magari la prima in miss e la seconda in hit), e il processore vedrà ritornarsi il dato corretto (in questo caso quello della seconda).

Compito di scegliere quale cache sfruttare nel caso di collisioni è quello del **controllore** di cache (nella cache ad indirizzamento diretto non c'era scelta). La scelta migliore possibile sarebbe quella di scegliere la cacheline al cui si accederà più tardi nel futuro (per mantenere i dati immediatamente utili nella cache).

Chiaramente, visto che non si può prevedere il futuro (o almeno non lo possono fare né la CPU né il controllore di cache), occorre adottare un'euristica. Una di queste

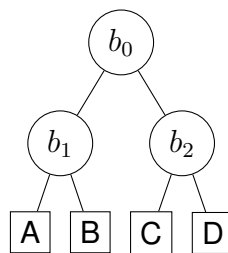
euristiche è la politica **LRU** (*Least Recently Used*), dove si sceglie la cacheline al quale non si accede da più tempo.

Per realizzare tale politica si sfrutta una memoria, che chiamiamo  $R$ . Con solo due vie, basterà memorizzare su  $R$  l'ultima via usata, e quella su cui scrivere sarà immediatamente l'altra.

Con più di due vie sarebbe necessario mantenere l'ordine degli accessi, cioè per  $n$  vie ricordare informazione necessaria a controllare  $n!$  diverse possibilità. Nella pratica, però, conviene usare politiche approssimate.

### 5.0.2 Pseudo-LRU dell'80486

Vediamo una di queste politiche approssimate, che gestiva 4 vie attraverso 3 bit  $b_0$ ,  $b_1$  e  $b_2$ . Si usava un albero binario per la selezione di una delle vie, disposto come:



dove i valori 1 sono i rami a destra, e viceversa i valori 0 sono i rami a sinistra.

In fase di rimpiazzamento, si sceglie la via seguendo l'albero. In fase di accesso, si modificano i  $b_i$  in modo da portare la via a cui si è fatto accesso in fondo all'ordinamento che si ottiene visitando l'albero. L'errore può essere dato dal fatto che la via che si trova nello stesso gruppo della via a cui si è fatto accesso potrebbe trovarsi ad un indice più alto del necessario, visto che si abbassa cumulativamente l'intero gruppo aggiornando  $b_0$ .

Per cache più grandi si sfruttano sempre algoritmi ad albero di questo tipo, magari tagliando i rami più bassi per lasciare spazio a scelte completamente casuali.

Notiamo poi che le memorie cache di questo tipo incontrano sempre difficoltà quando si fanno accessi ciclici ad indici che si ripetono con un modulo con il numero di vie diverso da zero: ad esempio se si leggono ciclicamente 5 indirizzi che corrispondono allo stesso indice, la cache non riuscirà mai a mantenere tutti e 5 in una delle cacheline delle vie, e quindi ogni accesso comporterà un miss.

### 5.0.3 Livelli di cache

Nei processori moderni si hanno solitamente più livelli di cache (3 o 4), che crescono in dimensioni e associatività più si vanno a disporre "lontano" dal processore e "vicini" alla RAM. Le cache di livello più basso saranno quindi più veloci ma più piccole, mentre le cache di livello alto saranno più lente ma più grandi.

Il controllore di cache provvederà a gestire i livelli di cache, effettuando gli accessi controllando a partire dal livello più basso (più veloce) per arrivare al livello più basso, fino alla RAM.

## 5.1 Interruzioni

La limitazione principale del processore studiato finora è che il flusso di controllo è completamente determinato dal programma in esecuzione. Attraverso il meccanismo

dell'interruzione, il sistema definisce  $e_1, \dots, e_n$  **eventi**, e il programmatore  $r_1, \dots, r_n$  **routine** per la gestione di tali eventi. Da qui in poi il processore continua ad eseguire il suo normale flusso di controllo, ma monitorando in qualche modo lo stato di questi eventi. Nel caso uno degli eventi  $e_i$  effettivamente si verifichi, la CPU provvederà a sospendere il flusso di controllo attuale e ad eseguire la routine  $r_i$ .

Un esempio classico dell'utilità di un meccanismo di questo tipo è dato dalle fasi di stampa che avevamo definito per dispositivi come le stampanti: attraverso l'approccio visto finora dovremmo controllare periodicamente un certo registro di stato per verificare la possibilità di scrivere un nuovo dato in un certo registro di buffer. Questo occupa la CPU con operazioni inutili, che potrebbe saltare se fosse la stampante stessa ad avvertirla di quando è pronta a ricevere un nuovo dato.

L'idea di base è quella di avere una nuova operazione da svolgere in fase di esecuzione di un'istruzione da parte della CPU, dopo l'esecuzione dell'istruzione stessa. Ad esempio, potremmo riportarci un bit di validità, `READY`, da parte della stampante, e controllarlo ad ogni istruzione per la chiamata di una routine di stampa. La chiamata sarà semplicemente un aggiornamento condizionato a `RIP`, con scrittura del contenuto attuale di `RIP` in pila (che è compatibile con le regole di chiamata dei sottoprogrammi a cui siamo abituati).

Un problema di questo approccio potrebbe essere che, se il bit che segnala l'evento non si aggiorna immediatamente, la CPU andrà in un ciclo continuo di arresto dell'esecuzione e inizio di una routine. Una soluzione potrebbe essere dotare della CPU di una *rete di accettazione* della richiesta: il bit di segnalazione dell'evento va in un generatore di impulsi che setta un SR flip-flop. A questo punto la CPU risponde (livello hardware, nella nuova fase di esecuzione appena descritta) con un segnale di reset nel momento in cui riesce a rilevare l'evento e spostarsi nella routine.

In verità la situazione è più complicata: ad esempio potremmo voler ignorare nuovi eventi quando stiamo già cercando di soddisfarne uno. Per questo i processori x86 prevedono un apposito flag, il flag **IF** (*Interrupt Flag*), che determina se le nuove interruzioni dovranno essere soddisfatte o meno. Il processore può essere quindi configurato per attivare automaticamente il flag IF in fase di risposta ad una richiesta di interruzione. Per effettuare il corretto ritorno, si usa la funzione `IRETQ`, che ripristina, oltre ad altre cose, lo stato dei flag (che era stato salvato in pila).

## 6 Lezione del 10-03-25

Torniamo sull'argomento delle interruzioni, specificando il modo in cui dobbiamo definire dei *gestori* per ogni interruzione.

Il calcolatore visto finora dispone di 4 interfacce:

- L'interfaccia *tastiera*, letta finora in controllo di programma, valutando la validità di un bit FI sul registro di stato;
- L'interfaccia *timer*, dotata di 3 singoli timer, di cui abbiamo detto il primo viene usato per generare interruzioni, il secondo non è più usato, e il terzo è connesso al *beeper speaker*;
- L'interfaccia a blocchi per *hard disk*, che accede ad un drive pilotando in base al suo stato un registro di stato (per noi era utile ad implementare la funzione di attesa del drive a controllo programma `wait_for_br()`).

Ignoriamo, per adesso, il video. Ognuna di queste interfacce può trarre beneficio dalla presenza di interruzioni:

- La tastiera potrebbe *avvertirci* dei nuovi tasti premuti, anziché costringerci a controllarli;
- Il timer ci deve avvisare, al termine del conteggio del timer 0, attraverso un interruzione;
- L'hard disk, come la tastiera, ci può avvisare con un interruzione quando è pronto ad una nuova scrittura.

Questo comportamento, delle cosiddette **interruzioni esterne**, è definito nella macchina studiata dal **controllore delle interruzioni**, che è l'Intel **APIC** (*Advanced Programmable Interruption Controller*). Questo scansiona periodicamente tutte le linee generatrici di interruzione ottenute dalle varie interfacce, e invia le interruzioni corrispondenti, una per volta, alla CPU.

Chiaramente, si rende necessario specificare un **tipo di interruzione**, su 8 bit (per 256 tipi) per ogni interruzione lanciata. A questo punto, l'APIC fornirà semplicemente la possibilità di assegnare un tipo di interruzione diverso ad ogni piedino di ingresso dall'interfacce, in modo che si possa assegnare ad ogni interruzione la routine di gestione più adatta.

La comunicazione fra CPU e APIC viene effettuata attraverso un *handshake* su due linee, **INTR** (*Interrupt Request*) e **INTA** (*Interrupt Acknowledge*), che comporta anche una lettura da parte della CPU di quanto l'APIC metterà sul bus (cioè il tipo di interruzione).

A questo punto, le routine vere e proprie verranno definite in una **IDT**, (*Interrupt Descriptor Table*), contenente in sequenza gli indirizzi delle prime istruzioni di ogni routine per ogni tipo di interruzione, e specificata a partire da un certo indirizzo indicato nel registro **IDTR**.

Come abbiamo visto, la reazione o meno della CPU ad una interruzione è data dall'attivazione del flag IF. Nel caso si passi effettivamente ad eseguire l'interruzione, ricordiamo che sia l'IP che lo stato dei flags verrà salvato in pila, e ripristinato a fine routine attraverso l'istruzione **IRET**.

### 6.0.1 Rilevamento di interruzioni da parte dell'APIC

Potremmo chiederci come fa il controllore APIC a capire quando un'interfaccia sta richiedendo una nuova richiesta.

Un primo approccio potrebbe essere di non rileggere il piedino di ingresso di quell'interfaccia, al ottenimento e successivo invio alla CPU di un interruzione, fino alla segnalazione, sempre da parte della CPU, di avvenuta gestione dell'interruzione. Questo può essere effettuato dotando l'APIC di un opportuno registro (**EOI**, *End Of Interrupt*), che la CPU andrà a modificare conclusa la gestione dell'interruzione.

Altri 2 registri di interesse sono i seguenti, entrambi su 256 bit (un bit per ogni tipo di interruzione):

- **IRR** (*Interrupt Request Register*): indica con bit alti quali interruzioni sono state inviate dalle interfacce attualmente;
- **ISR** (*Interrupt Service Register*): indica con bit alti a quali interruzioni sta rispondendo il processore attualmente. In un processore single-threaded come quello che studiamo al più uno solo dei suoi bit sarà alto in un dato momento.

### 6.0.2 Priorità delle interruzioni

Possiamo chiederci come l'APIC si comporta in caso di più richieste concorrenti. Un'idea potrebbe essere di assegnare una priorità ad ogni richiesta, e rispondere prima alle richieste di priorità più alta.

Ci rendiamo quindi conto che alcune richieste sono più importanti di altre: ad esempio, la pressione di un tasto su tastiera può essere ignorata, se ad esempio nel frattempo ci arriva la richiesta di interruzione da parte di un timer. La pressione del tasto non si ripeterà infatti in tempo utile, mentre il timer potrebbe inviarci nuove richieste mentre ancora non siamo pronti a riceverle, e continuerà a farlo a scadenze regolari (potremmo finire per gestire solo un sottoinsieme delle richieste che ci vengono effettivamente inviate).

Vediamo quindi nel dettaglio di come i codici delle interruzioni sono definiti. I primi 4 bit rappresentano la **classe di precedenza** dell'interruzione: a classi di precedenza maggiore abbiamo gestione prioritaria delle richieste di interruzione. Il trasferimento da IRR a ISR avverrà quindi prima per richieste di classe di precedenza più alta, e poi per quelle di classe di precedenza più bassa.

## 7 Lezione del 11-03-25

### 7.0.1 Interruzione di livello o di fronte

Vediamo un dettaglio sul comportamento dell'APIC: questo può rilevare, in base alla sua configurazione, i **livelli** o i **fronti** delle variabili in ingresso.

Questo può avere delle implicazioni diverse a seconda dell'interfaccia. Ad esempio, avevamo detto che il timer in modalità 2 genera un'onda quadra. Se si usa una routine lanciata dal timer a interruzione di programma, e si configura l'APIC per rilevare il livello, potrebbe essere che a routine concluse il livello del timer è sempre alto, e quindi l'interruzione viene lanciata nuovamente.

Questo è chiaramente diverso dal comportamento desiderato, ed è quindi opportuno configurare l'APIC per rilevare i soltanto fronti di salita.

### 7.1 Riassunto sui tipi di interruzioni

Abbiamo visto finora le **interruzioni esterne mascherabili**. Vediamo in verità che esistono altri tipi di interruzione, da cui la lista completa:

- **Interruzioni esterne:** il tipo che abbiamo visto finora, di cui notiamo le:
  - **Interruzioni esterne non mascherabili:** cioè che non possono essere mascherate, solitamente rappresentano eventi particolarmente gravi o comunque la cui gestione ha alta importanza;
- **Interruzioni interne (Eccezioni):** eventi che non arrivano dall'esterno, ma si generano all'interno del processore stesso. Questi sono particolari errori logici che potrebbe incontrare il processore, come ad esempio la divisione per 0, il tentativo di eseguire un'istruzione non riconosciuta, ecc...

Un'interruzione particolare è quella rappresentata da **INT3**, l'interruzione di *debug*. Attraverso questa, un *debugger* è capace di interrompere l'esecuzione di un programma ad un certo indirizzo del suo codice macchina.

Un'altra interruzione di debug è data dalla single step, che viene lanciata ad ogni istruzione quando è attivo un certo flag (appunto, il flag single step). Questo permette al debugger di eseguire il programma in modalità *passo singolo*, cioè eseguendo un'istruzione e interrompendo, permettendo al programmatore di osservare il suo andamento passo per passo.

Una differenza fra le interruzioni esterne e le eccezioni è che le eccezioni possono essere sollevate *durante* la lettura e esecuzione di un'istruzione, quindi ad esempio mentre si stava interpretando un codice operativo (si pensi all'interruzione di operazione non riconosciuta). In verità, per assicurare l'atomicità dei cicli di esecuzione, la CPU ripristina automaticamente il suo stato a prima del lancio dell'interruzione. In particolare, possiamo distinguere 3 tipi di eccezione:

- **Fault:** l'esecuzione non viene ancora eseguita, lo stato IP prima della sua esecuzione viene salvato (quindi si rimane alla stessa istruzione), e si può riprovare ad eseguirla dopo aver risolto l'errore;
- **Trap:** l'esecuzione ormai è stata eseguita, e si salva l'IP successivo.
- **Abort:** raggruppa degli eventi particolarmente disastrosi in cui l'esecuzione si arresta completamente (ad esempio la tripla eccezione).

Quando viene lanciata una *fault* o una *trap*, il processore cerca nella IDT se esiste un handler corrispondente (segnalato attraverso un bit nell'IDT stessa, alla riga della tabella corrispondente all'eccezione considerata). Nel caso questo non esista, si riprova con la fault di *doppia eccezione*, che quindi rappresenta una fault a sé. Nel caso nemmeno questo handler esista, viene lanciata una fault di *tripla eccezione*, che è di tipo *abort* e comporta quindi l'arresto del programma.

- **Interruzioni software:** interruzioni che vengono lanciate direttamente dal programma attraverso l'istruzione **INT**, la cui utilità verrà vista studiando il meccanismo della *protezione*.

## 8 Lezione del 17-03-25

### 8.1 Protezione

Tutti i programmi che abbiamo visto finora hanno il pieno controllo su la macchina su cui sono in esecuzione. Questo significa che possono impattare qualsiasi regione di memoria, incluso il loro stesso codice macchina, o i frame di stack di programmi lanciati prima di loro.

Un approccio di questo tipo non è ideale quando più programmi, magari di utenti diversi, vengono lanciati ed eseguiti *quasi* in contemporanea (*time-sharing*) sulla stessa macchina.

Un esempio di questa situazione può verificarsi nel caso di esecuzione *batch*, cioè di esecuzione successiva di più programmi, magari scritti da più utenti. Vorremmo massimizzare l'uso della CPU sospendendo un programma e iniziandone un altro nel caso il primo fra questi inizi un'operazione che richiede una quantità significativa di tempo (ad esempio un accesso a un dispositivo di I/O). In questo caso, visto che non possiamo fidarci della benevolenza degli utenti nell'inserire istruzioni esplicite per il cambio da un programma all'altro, vorremo agire sull'hardware per, ad esempio, vietare all'utente l'uso di certe istruzioni (qui **IN** e **OUT**) e costringerlo ad usare primitive messe a disposizione dal sistema.



Chiaramente, però, le primitive dovranno poter usare **IN** e **OUT** per fare l'I/O vero e proprio con i dispositivi. Per permettere questo doppio comportamento introduciamo l'idea di **protezione**.

### 8.1.1 Contesti di esecuzione

Il programma nella memoria potrà essere in esecuzione, in un momento qualsiasi, in uno di due **contesti**, o *modi* (vedremo nell'architettura x86 corrente, si parla di protezione a *ring*): il contesto **sistema** e il contesto **utente**. Le istruzioni di cui permetterà l'esecuzione saranno quindi determinate dal contesto corrente.

Forniamo allora il processore di un apposito registro, il **CS** (*Code Segment*), a 2 bit. I 2 bit sono necessari in quanto storicamente (il meccanismo descritto viene introdotto nell'architettura x86 a partire dal 286) si definivano quattro contesti, o **ring**:

CS	Ring	Tipo
00	Ring 0	Kernel (sistema)
01	Ring 1	Driver
10	Ring 2	/ /
11	Ring 3	Utente

Il nome CS deriva dal fatto che questo registro era pensato per gestire la *segmentazione* della memoria. Sia questo meccanismo, che i due ring interni (l'1 e il 2) sono pressoché inutilizzati nell'architettura x86-64 moderna, e quindi li ignoreremo, portandoci effettivamente alla situazione dove CS rappresenta un flag che distingue fra contesto *sistema* e contesto *utente*, come avevamo ipotizzato.

### 8.1.2 Transizioni fra contesti

Ipotizziamo quindi che all'avvio si parta in contesto sistema, e che si passi al contesto utente quando si esegue un programma utente. Per permettere all'utente di "accedere" alle istruzioni privilegiate, vogliamo che questo disponga di un modo di tornare al contesto sistema, ma lasciando il controllo al sistema operativo (altrimenti sarebbe inutile introdurre l'idea di un contesto utente in primo luogo). Di contro, vogliamo un modo per il sistema operativo di restituire in sicurezza il controllo al programma, previa transizione del processore in contesto utente.

Vediamo come il meccanismo dell'interruzione fornisce un metodo per gestire questa situazione.

Introdurremo un tipo di interruzione apposito, che restituisce il controllo al sistema operativo (semplicemente passando ad un gestore di interruzione definito dal sistema operativo) passando a contesto sistema. Il tipo di operazione che stiamo richiedendo al sistema operativo potrà essere passato in qualche registro specifico, solitamente `%EAX`. Il problema potrebbe essere chiaramente che l'utente ha la possibilità di modificare tutta la memoria, e quindi la stessa IDT e il gestore impostato.

### 8.1.3 Protezione di memoria

Si rende quindi necessario un meccanismo di gestione degli accessi in memoria. In contesto utente, quindi, oltre a permettere l'utilizzo di solo alcune istruzioni *non privilegiate*, il processore dovrà permettere l'accesso solo a determinate regioni di memoria. Visto

che non abbiamo ancora introdotto l'idea di *memoria virtuale*, modellizziamo temporaneamente questa configurazione con un apposito registro a controllo sistema che decide quali regioni di memoria sono o non sono accessibili.

Abbiamo quindi l'immagine completa del meccanismo della protezione, che avevamo introdotto per privilegiare le sole istruzioni, ma ci rendiamo adesso conto deve consistere in:

- Protezione delle **istruzioni** attraverso il loro privilegiamento al contesto sistema, come avevamo visto;
- Protezione della **memoria** definendo regioni accessibili in sola modalità sistema.

#### 8.1.4 Transizione da contesto utente a contesto sistema

Vediamo nel dettaglio come si passa dal contesto utente al contesto sistema. Per questo sfrutteremo l'istruzione x86 **INT**, che permette di generare un'interruzione software sulla base del tipo fornito come operando. Si potrà quindi avere una chiamata a sistema del tipo:

```
1 mov $0x00, %eax # tipo chiamata
2 int $0x80       # chiamata sistema (per x86, in x96-64 esiste syscall)
```

Questo si tradurrà a livello processore nel salvataggio dello stato corrente di esecuzione, la transizione al contesto sistema e lo spostamento in IP della prima istruzione di un apposito sottoprogramma di servizio atto a gestire l'eccezione (e quindi soddisfare, se possibile, la richiesta del programma per cui questo ha sollevato in primo luogo l'interruzione).

Per capire nel dettaglio cosa accade nel processore è necessario:

- Capire come è strutturata la Interrupt Descriptor Table (IDT) all'interno della memoria del sistema, che supponiamo essere privilegiata (altrimenti l'utente potrebbe manometterla);
- Capire come viene gestita un'interruzione software, cioè come si conserva lo stato al momento dell'interruzione, e come si inizia l'esecuzione del gestore in contesto sistema.

Vediamo questi dettagli in ordine.

#### 8.1.5 Struttura della IDT

Vediamo quindi nel dettaglio la struttura di un'entrata della IDT. Questa viene a trovarsi nella memoria privilegiata a partire da un indirizzo, come avevamo detto, contenuto nel registro IDTR. L'impostazione di questo registro si fa attraverso apposite istruzioni, sempre ad accesso privilegiato.

Le entrate dell'IDT si chiamano **gate IDT**, che si distinguono in 3 tipi, *Task Gate*, *Interrupt Gate* e *Trap Gate*, che al momento non vediamo. La struttura a livello di memoria contiene le seguenti informazioni:

- L'offset della routine di gestione dell'interruzione, in alcune modalità comprendente dell'indice di segmento, ecc...;
- **P**: un flag di **presenza**, indica se il descrittore è effettivamente abilitato;

- **L**: il livello di protezione (contesto sistema o utente) a cui deve essere eseguito il gestore. Notiamo che questa sembra essere una semplificazione del corso (il professore si è rivelato ombroso a riguardo). In verità, l'IDT mantiene un riferimento al CS dell'istruzione, che anche se ora abbiamo assunto come un semplice flag sistema/utente, rappresenta invece un riferimento al *segmento* vero e proprio all'interno del cui è allocata la routine. Informazioni riguardo al livello di ring di ogni segmento sono contenute in altre tabelle specifiche, dette **GDT** (*Global Descriptor Table*) e **LDT** (*Local Descriptor Table*). Il salto al livello L viene quindi fatto automaticamente in base al livello del segmento in cui è allocato il gestore (vediamo che con considerazioni simili si capisce come mai viene allocato, oltre a RIP, anche il CS corrente in fase di chiamata);
- **I/T**: il tipo di interruzione fra quelli sopra definiti.
- **DPL**: il livello minimo da cui si può accedere al gestore come interruzione interna (attraverso una **INT**). Questo non significa che tale gestore non possa essere lanciato da un eccezione.

#### 8.1.6 Gestione dell'interruzione software

A questo punto la chiamata di interrupt sta effettivamente nella transizione fra due **pile**: la separazione fra contesto utente e contesto sistema viene infatti resa possibile anche dalla presenza di due pile separate, di cui l'ultima chiaramente sta in memoria protetta. Il programma è normalmente in esecuzione nella pila utente: al momento del sollevamento di un interruzione software, si passa all'esecuzione (se alcune condizioni che vedremo fra poco sono rispettate) della routine di gestione definita dal sistema operativo. Questa imposta un nuovo frame sulla pila con i seguenti dati:

- L'Instruction pointer **RIP**, da dove si vorrà ripartire nell'esecuzione una volta gestita l'interruzione. Notiamo che in verità questo indirizzo, che è fra l'altro in memoria virtuale, è corredato a seconda del tipo di gate dall'**SS** (*Stack Segment*) o dal **TSS** (*Task State Segment*), utili alla memoria segmentata che come abbiamo visto non ci è di interesse. La caratteristica importante è che si conserva un riferimento a dove ripartire, in memoria, nell'esecuzione una volta gestita l'interruzione;
- Il contenuto attuale di **CS**, cioè il contesto al momento della chiamata, che chiaramente vorremo ristabilire in seguito;
- Come abbiamo visto, anche **EFLAGS** viene memorizzato, in quanto gli interrupt mascherabili vengono mascherati in fase di gestione di un interrupt sistema (attraverso il flag **IF**), e vogliamo resettare questo comportamento al termine della gestione.

Un caso particolare ma permesso è rappresentato dalla situazione dove **L**, il livello di destinazione, corrisponde allo stato attuale (ad esempio, sono permesse chiamate di interruzioni da contesto utente a contesto utente, o da contesto sistema a contesto sistema). In questo caso, chiaramente, tutta questa operazione verrà svolta su un'unica pila (sia questa la pila utente o la pila sistema). Noteremo fra poco come questa possibilità rivela delle falle di sicurezza che vanno gestite.

### 8.1.7 Transizione da contesto sistema a contesto utente

La transizione inversa a quella vista adesso viene fatta semplicemente ritornando dall'interruzione attraverso la `IRETQ`. In questo caso si preleva dalla pila sistema (utente se eravamo in un'interruzione a gestione livello utente) le informazioni che vi avevamo inserito al momento della chiamata dell'interruzione (`RIP`, `CS` ed `EFLAGS`) e si ristabilisce lo stato precedente al sollevamento dell'istruzione. Anche qui vi sono delle particolarità, che verranno spiegate, assieme a quelle annunciate in precedenza, nel paragrafo seguente.

### 8.1.8 Particolarità della gestione delle interruzioni software

Notiamo una particolarità riguardo alla transizione di contesto in fase di chiamata dell'interruzione (nota osservando il contesto attuale e l'`L` dell'interruzione lanciata), e riguardo alla transizione di contesto in fase di ritorno dall'interruzione (nota osservando il contesto attuale e il contesto salvato in pila).

Infatti, in fase di chiamata (quando si usa la `INT`), se `L` è minore del contesto corrente, viene lanciato un errore. La motivazione è principalmente una questione di simmetria nel meccanismo di chiamata delle interruzioni, piuttosto che una ragione di sicurezza: si vuole che le interruzioni ci portino in contesti maggiori o uguali del livello presente in `CS`.

Viceversa, se si prova a passare ad un livello superiore in fase di ritorno dall'interruzione (cioè quando si usa la `IRETQ`), viene lanciato un altro errore. La motivazione è che, visto che prevediamo nell'`IDT` il flag `L`, livello di destinazione, che permette di chiamare interruzioni in contesto utente, l'utente potrebbe impostare un frame di pila dove si richiede effettivamente l'accesso ad un livello di protezione superiore, e poi usare `IRETQ` per ritornare da tale frame di pila e passare quindi a tale livello di accesso.

## 9 Lezione del 18-03-25

### 9.1 Multiprogrammazione

Abbiamo accennato al funzionamento dei calcolatori in modalità *batch*, dove più programmi vengono eseguiti in sequenza, uno dopo l'altro.

Un paradigma sicuramente più piacevole per l'utente, e più diffuso al giorno d'oggi, è quello del **time-sharing**, dove il processore dà l'illusione agli utenti di portare avanti più attività contemporaneamente, mentre il tempo della CPU è in verità diviso in frammenti temporali ridotti dove si dedica a ogni attività singolarmente.

Il meccanismo stesso della protezione che abbiamo introdotto alla lezione precedente serve appunto a difendere i programmi l'uno dall'altro in caso di esecuzione "parallela" (da non confondere col *multithreading*). Infatti, anche se è un concetto nato nei *mainframe* a uso pubblico, la protezione si è subito diffusa anche nelle macchine personali degli utenti, in modo da difendere non più programmi di diversi utenti ma più programmi dello stesso utente, magari soggetti a bug che potrebbero corrompere lo stato di altri programmi o dell'intero sistema.

Oggi il meccanismo di protezione si trova in tutti i calcolatori moderni, dai telefoni cellulari ai supercomputer, ed è risparmiato solo nel caso dei microcontrollori più semplici.

La domanda che ci poniamo adesso è quindi quella di *come* realizzare un sistema capace di dare quest'illusione dell'esecuzione "parallela" di più programmi, che avevamo introdotto all'inizio del corso come **multiprogrammazione**.

### 9.1.1 Processo

Chiamiamo **processo** un programma in esecuzione. Ciò che vorremo eseguire in parallelo sono, più propriamente, non programmi ma *processi*.

Intendiamo quindi un processo non come il codice che definisce un programma, ma come il programma stesso una volta che viene messo in esecuzione nel calcolatore, quindi tutti gli stati di elaborazione (disposti nel tempo) del calcolatore nell'esecuzione di tale programma.

Il modo in cui andremo a definire il paradigma della multiprogrammazione è assumendo un processo come un insieme di operazioni **atomiche**, che possono essere interrotte al loro termine o prima del loro inizio, e che bastano insieme all'istruzione successiva del codice a determinare lo stato successivo di esecuzione del processo.

### 9.1.2 Contesto

Un altro concetto chiave nella multiprogrammazione sarà il **contesto** di un processo. Avevamo parlato di contesto in termini di protezine: adesso diamo un significato leggermente diverso. Ogni processo si aspetterà infatti di trovarsi nel *suo* contesto personale: le operazioni intaccheranno i suoi registri, che si aspetta essere l'unico a modificare, ecc... Il sistema operativo dovrà quindi essere in grado di fornire a ogni processo il suo contesto specifico.

Vediamo che questa idea si può tradurre già lato software. Il **cambio di contesto** può essere infatti effettuato, prendendo l'esempio dei soli registri, mantenendo una struttura dati che contiene un'entrata per ogni registro. Al momento del cambio basterà copiare l'insieme dei registri corrispondenti al contesto di un certo processo nei registri veri e propri del processore.

Un discorso analogo sarà quella della memoria: ogni processo si aspetterà che al suo contesto corrisponda una sua copia della memoria. Possiamo mantenere un'altra struttura dati, simile a quella posta per i registri, che si occupa di mantenere informazioni riguardo alle regioni di memoria corrispondenti ad ogni contesto, e caricare quindi queste in una sezione dedicata su e della memoria stessa, o per semplicità su e dall'hard disk (così erano i primi sistemi time-sharing). Vedremo più nel dettaglio questo aspetto quando introdurremo la *memoria virtuale*.

Facciamo un'ultima nota sulla *comunicazione* fra processi: nel caso più semplice, ogni processo non è al corrente dell'esistenza degli altri processi, e gestisce la sua *memoria privata*. Il sistema che studieremo dispone invece anche di una *memoria condivisa*, che permette ai processi di condividere informazioni fra di loro.

### 9.1.3 Kernel

Il programma che si occupa di effettuare queste operazioni di cambio di contesto si chiama **kernel** o *nucleo*. E' sempre in esecuzione in modo sistema e gestisce i contesti e le risorse assegnate ad ogni processo.

Immaginiamo quindi il kernel come un intermediario fra **processi** e **hardware**. Notiamo che questo non significa che kernel e processi sono *contemporaneamente* in esecuzione: questo è impossibile, in quanto la CPU è una sola. Kernel e processi sono infatti in

esecuzione singolarmente, l'uno alla volta, e l'unico modo in cui si restituisce il controllo al kernel da un processo è attraverso i 3 tipi di interruzioni:

- Interruzioni esterne (dai dispositivi);
- Eccezioni (errori e altri malfunzionamenti, non necessariamente dati da errori di programmazione);
- Interruzioni interne (sollevate dall'istruzione **INT**).

Nel caso dei sistemi in time-sharing di cui abbiamo brevemente parlato prima, il cambio di processo viene eseguito ad intervalli regolari sfruttando interruzioni esterne periodiche generate da un timer.

## 10 Lezione del 21-03-25

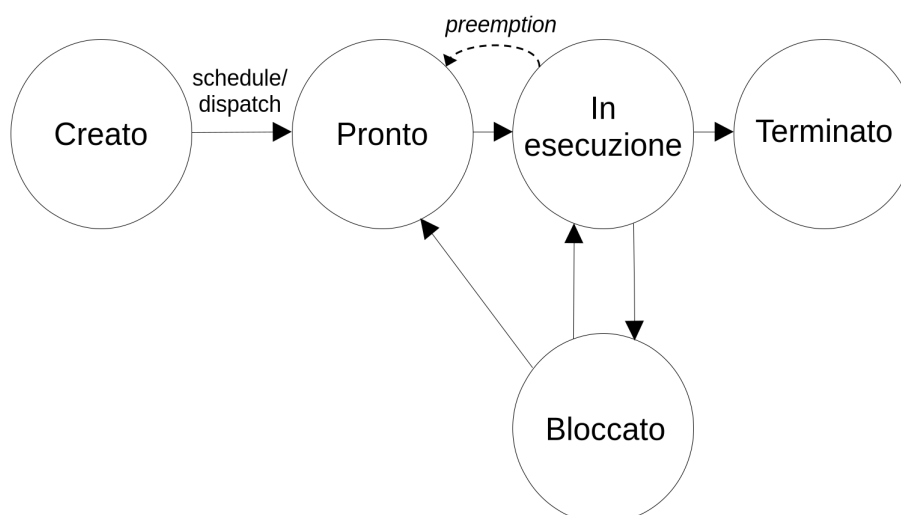
Andiamo a definire più nei dettagli la struttura di un processo e le modalità secondo le quali questi si possono creare e distruggere.

### 10.1 Descrittori di processo

Un processo è descritto fondamentalmente da un astrazione, detta **descrittore di processo**, idealmente contenuta in una qualche locazione contigua, assieme ad altri descrittori, in memoria.

- Il primo campo del descrittore sarà un **indice** numerico unico ad ogni processo;
- Dovremo poi tenere conto del **contesto** del processo, inteso come la copia di tutti i registri del processore.

Ogni processo ha un ciclo di vita che rispetta l'andamento del seguente grafico:



Vediamo nel dettaglio il significato delle diverse fasi. In fase di creazione del processo, il suo descrittore viene posto in una struttura dati che ne consente **schedulazione** e **dispatch**:

- **Schedulazione:** effettivamente la scelta che il kernel fa, assunto il controllo, su quale è il prossimo processo da portare in esecuzione (passaggio da processo **pronto** a processo in **esecuzione**);
- **Dispatch:** l'esecuzione effettiva di una serie di operazioni di tale processo.

I processi possono anche **bloccarsi**, cioè mettersi in attesa di qualche evento.

Infine, un processo può **terminare**, cioè sparire dal sistema (lui e il suo descrittore). Anche in questo caso il processo deve essere attualmente in esecuzione.

Una transizione che non è prevista da tutti i sistemi è quella di **preemption**, cioè di ritorno allo stato **pronto** a controllo dello scheduler. La maggior parte dei sistemi operativi supporta tale funzionalità, il nucleo che vedremo solo parzialmente.

## 10.2 Prima vista dell'esecuzione del kernel

Dopo il boot della macchina, il kernel si impadronisce della macchina e lancia il primo processo (il processo utente). Da qui in poi il kernel avrà il controllo solo fra un processo e l'altro, in caso di interruzioni (interne, esterne o eccezioni), e potrà restituirlo solo attraverso il ritorno da gestore con `IRETQ`.

Come abbiamo visto, ad ogni chiamata di gestore di interruzione lascia `RIP`, `CS`, `RFLAGS` e `RSP` al tempo di chiamata dell'interruzione (facendo le opportune distinzioni fra *fault* e *trap*) in pila. A questo punto il gestore fa una copia dei registri generali, e si ha a quel punto una "foto" del processore al momento di attraversamento del gate, che rappresenterà quindi il *contesto* del processo stesso al momento della chiamata dell'interruzione.

In questo, sfrutteremo delle routine (`salva_stato` e `carica_stato`) all'avvio e al termine di ogni gestore, che si occupano di salvare e caricare il contesto del processo attualmente in esecuzione. Per conoscere quale questo processo sia, si mantiene una variabile globale nel sistema, `esecuzione`, che punta al descrittore del processo (che è dove vogliamo mettere il contesto stesso).

I processi pronti staranno in una certa struttura dati (nel nucleo una linked list), ordinata per la **priorità** (un altro valore che manteniamo nel descrittore di processo) di ogni processo. Un gestore di interruzione di base, quindi, si potrebbe magari occupare di passare al contesto e all'esecuzione del processo di priorità più alta a intervalli regolari, magari regolato da un timer (cosiddetto *timeslicing*).

Altre situazioni, più vicine a noi, sono quelle del termine di una gestione di un interruzione esterna, o bloccaggio automatico di un processo, dove il kernel deve selezionare il prossimo processo da eseguire, scegliendo chiaramente quello a priorità più alta.

Inseriamo un processo fittizio, *dummy*, nella lista dei processi pronti con la priorità più bassa possibile. Questo ci assicurerà di non trovarci mai una situazione dove nessun processo è pronto all'esecuzione, e quindi avere sempre qualcosa a cui il kernel può passare (idealmente il processo dummy effettua solo un ciclo a vuoto).

Un ulteriore dettaglio è quello dello stato del processo alla sua creazione. Non è infatti realistico pensare di controllare se quel processo richiede inizializzazione ogni volta che si ritorna da un interruzione gestita a livello sistema. Alla creazione del processo, quindi, vogliamo svolgere le seguenti azioni in modo che il processo venga eseguito per la prima volta già in uno stato completo:

- Allocare una **pila sistema** dedicata al processo;
- Inizializzare la pila sistema. Questo consisterà nell'inizializzare a loro volta:

- IP alla prima istruzione del processo;
  - CS al segmento livello utente dove si trova il processo;
  - FLAG a quanto viene richiesto dallo standard C++ al momento di avvio (solitamente tutto a 0), con l'eccezione di IF a 1.
- Allocare il **descrittore** di processo, e mettere quel processo fra i processi pronti;
  - Inizializzare il descrittore. Questo consiste nell'inizializzare a loro volta:
    - Un puntatore alla pila sistema appena definita;
    - Il contesto del processo;
    - L'**argomento** di chiamata del processo, utile al debug;
    - L'**IOPL**, *IO Privilege Level*, che specifica la possibilità o meno del processo di accedere all'IO.

## 11 Lezione del 24-03-25

### 11.1 Primitive

Abbiamo introdotto il concetto di primitiva, cioè di routine svolte dal sistema al servizio di un dato programma. Queste verranno implementate come gestori di interruzioni, quindi non propriamente funzioni, in quanto implicano un passaggio di contesto. Ciò nonostante, in un linguaggio come il C++ le primitive saranno comunque rappresentate da funzioni, dette *funzioni interfaccia*, scritte in assembly e che hanno il solo compito di usare la funzione **INT** con i parametri necessari alla chiamata di una specifica primitiva primitiva.

#### 11.1.1 Primitiva di creazione di un processo

Abbiamo visto che la creazione di un processo consiste nell'inizializzazione dalla memoria ad esso dedicata in contesto utente e sistema, alla creazione del suo descrittore e all'inserzione di questo in pila "pronti". Se la pila è rappresentata come una linked list, l'operazione dovrà quindi essere quella di un *inserimento in testa*.

Notiamo che questa operazione non può essere divisa da altre interruzioni, in quanto richiede necessariamente almeno due passaggi, dove fra un passaggio e l'altro la lista viene lasciata in uno stato inconsistente:

- Prima si fa puntare il processo al resto della lista;
- Poi si fa puntare il puntatore della lista pronti al processo inserito.

anche invertendo l'ordine delle operazioni, dopo la prima la lista è inconsistente (in questo caso perché il processo inserito non viene effettivamente visto, nel caso opposto perché non vengono visti tutti gli altri).

Nel caso di routine di sistema basterà abbassare il flag **IF**, disabilitando effettivamente le interruzioni, durante tutta la durata della routine. A questo punto basterà evitare di generare eccezioni, e non usare mai l'istruzione **INT**, per ottenere una routine che viene eseguita dal processore nella sua interezza senza il rischio di interruzioni. Chiamiamo codice di questo tipo **codice atomico**. Il kernel Linux, ad esempio, *non* è atomico.



### 11.1.2 Disposizione delle primitive

La memoria del calcolatore conterrà in qualsiasi momento la tabella IDT, di cui abbiamo detto le prime 32 entrate rappresentano le eccezioni. Siamo quindi liberi di usare i gate dal 33 in poi per implementare le primitive. Per queste primitive dobbiamo impostare i parametri:

- **P:** 1, per attivare il gate;
- **L:** sistema, in quanto le primitive devono essere svolte a livello sistema;
- **DPL:** utente, in quanto le primitive devono essere accessibili all'utente;
- L'indirizzo effettivo della routine, implementata (in assembly, serve **IRET**), che deve trovarsi da qualche altra parte;
- **I/T:** tipo interrupt (interruzioni esterne mascherabili disabilitate).

Notiamo che l'interruzione esterna non mascherabile 2 è comunque in grado di bloccare le nostre istruzioni atomiche. Questo non è importante, in quanto abbiamo detto la useremo per casi particolarmente catastrofici (dove magari la salvaguardia dei dati dell'utente e del sistema è di maggiore priorità rispetto allo stato dei processi).

La struttura della routine sarà quindi tipicamente:

```
1 primitiva:
2     CALL salva_stato
3     CALL c_primitiva
4     CALL carica_stato
5     IRETQ
```

dove `c_primitiva` è una funzione, scritta in C++, che termina con una **RET** e lascia quindi che `primitiva` restituisca il controllo all'utente con **IRETQ**.

Per chiamare la primitiva da C++, come abbiamo detto, ci doteremo di una funzione di interfaccia del tipo:

```
1 primitiva_i:
2     INT $ tipo %il tipo di primitiva
3     RET
```

### 11.1.3 Passaggio di parametri alla primitiva

Supponiamo di voler passare dei parametri alla nostra primitiva. La funzione di interfaccia dovrà semplicemente essere modificata per accettare dati parametri (`primitiva_i(params...)`).

A questo punto la `primitiva_i` potrà svolgere il passaggio effettivo sfruttando i registri, solitamente il solo registro `%EAX` (in quanto `salva_stato` non modifica i registri).

### 11.1.4 Passaggio di parametri dalla primitiva

Per avere una restituzione di parametri da parte della primitiva la situazione è più complicata, in quanto abbiamo una chiamata a `carica_stato` prima del ritorno della primitiva per **IRETQ**.

Abbiamo però accesso al contesto di processo, nel descrittore di processo, e possiamo quindi modificare i registri che ci interessano direttamente lì.

## 11.2 Semafori

Per gestire l'accesso condiviso ad una risorsa, nel nostro kernel adotteremo il meccanismo dei **semafori**.

Introdotti da Dijkstra nel 1962, questi si possono meglio modellizzare come una scatola piena di gettoni: ogni utente può mettere un gettone o prelevare un gettone dalla scatola, con la condizione che questa operazione sia atomica: se si tenta di prendere un gettone che non esiste, si resta in attesa finché quel gettone non viene effettivamente immesso nella scatola.

I problemi che vogliamo risolvere sfruttando i semafori sono effettivamente due categorie:

- Problemi di **mutua esclusione**: assicurarsi che solo un processo possa accedere ad una risorsa in un dato momento.

In questo caso si associa un gettone alla risorsa: accedere alla risorsa significa prendere il gettone, restituire la risorsa significa reinserire il gettone. L'esistenza di un singolo gettone assicura che solo un processo abbia accesso alla risorsa in un dato momento. Al momento della reimmissione del gettone, il processo che ne vince l'accesso sarà nel nostro kernel quello a priorità più alta.

Notiamo inoltre che un processo che cerca di estrarre un gettone da una scatola vuota (tenta l'accesso ad una risorsa occupata o comunque non disponibile) dovrà aspettare che questa risorsa si renda disponibile: rappresenterà quindi il caso perfetto di **blocco** del processo, che può essere realizzato con **preemption** nei sistemi che la supportano;

- Problemi di **sincronizzazione**: esistono più attività, e ci interessa che alcune attività vengano fatte prime di altre (ordinamento *parziale*).

Prendiamo l'esempio di avere due processi, A e B, e di volerci assicurare che  $A \rightarrow B$ . In questo caso creiamo un semaforo associato al processo A, che parte vuoto. A mette il suo gettone nel semaforo quando finisce la sua esecuzione. A questo punto, B preleva il gettone ed esegue. Se B avesse provato ad entrare in esecuzione prima che A avesse terminato, non sarebbe riuscito a prelevare il gettone e avrebbe fallito.

Nel caso di 2 processi (sempre A e B, con A che scrive e B che legge) che devono scambiare dati fra di loro ciclicamente, potremmo usare 2 semafori per realizzare un *handshake*. Ad esempio, definiamo quelle che effettivamente sono due variabili logiche sfruttando i semafori, che intendiamo come "buffer scritto" e "buffer letto". Il processo A dovrà semplicemente attivare il semaforo "buffer scritto" in fase di scrittura, e il processo B attivare il semaforo "buffer letto" in fase di lettura. Abbassando questi semafori al termine delle rispettive operazioni, e assicurandosi, osservando l'altro semaforo, di poter effettivamente procedere ad una nuova operazione, potremmo realizzare il paradigma desiderato.

Dal punto di vista di implementazione, il kernel fornisce una primitiva `sem_ini(int val)` che inizializza un semaforo con `val` gettoni iniziali, restituendone l'indirizzo. Da qui in poi i processi hanno accesso alle primitive `sem_wait` e `sem_signal`, che si occupano rispettivamente di richiedere e restituire un gettone.