



# L'ISA x86



Prof. Alberto Borghese  
Dipartimento di Informatica  
[alberto.borghese@unimi.it](mailto:alberto.borghese@unimi.it)

Università degli Studi di Milano

Patterson, sezioni 1.12.8, 2.17, 2.21.6



# Sommario



## Le architetture Intel

I registri ed il loro utilizzo

L'ISA degli Intel

La codifica delle istruzioni



## Intel Cortex I7



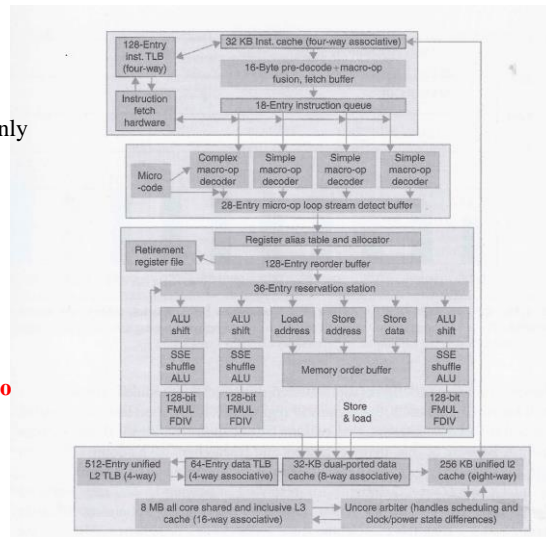
INTEgrated ELEctronics (1968), mainly memories.

Evoluzione dell'Intel 8086 (1978)

Simile all'evoluzione delle speci

Compatibilità a ritroso

**“This checked ancestry has led to an architecture that is difficult to explain and impossible to love”  
[Patterson Hennessy]**



## Le prime architetture Intel



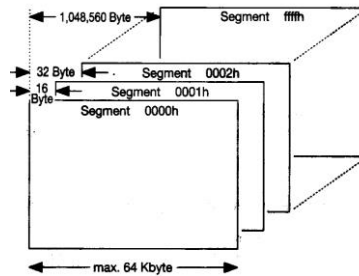
- **1978 – 8086**
  - Estensione del micro-processore 8080 utilizzato per applicazioni industriali (micro-controllore). Stessa ISA. **Architettura a 16 bit** con registri a 16 bit. Parte dei registri è dedicata a compiti specifici, **non è un'architettura con registri general-purpose.**
- **1980 – 8087**
  - Coprocessore matematico. Dedicato a velocizzare le operazioni in virgola mobile. Modifica nel modo di gestire gli operandi. Questi potevano essere **presi dallo stack (virgola mobile sempre) oppure dai (pochi) registri.**
  - Estensione degli operandi a 10 byte (80bit), *Extended Double Precision.*
    - E' il compilatore a potere dichiarare variabili su 10byte (Extended Double).
  - Ciclo di esecuzione di un'operazione aritmetica (dati in stack):
    - Push <operando\_1> in stack (esteso a 10byte); Push <operando\_2> in stack (esteso a 10byte).
    - Operazione.
    - Pop <risultato> da stack
- Limite nello spazio di indirizzamento: 1 Mbyte ( $2^{20}$ ).
  - Indirizzamento operato come: **Base shl 4 + Offset**
- Gestione di memoria ed I/O tramite **3 segnali di controllo: RD, WR, IO/MEM**



## Indirizzamento in modalità reale



Interleaving dei segmenti:  
Spazio di indirizzamento di 1Mbyte  
suddiviso in segmenti di 64kbyte



- Modalità **reale**
  - $\text{Indirizzo} = 16 * \text{Segmento} + \text{Offset}$
  - In tal modo posso indirizzare 64KB x 16 = 1 MByte di MM
  - Modalità **Virtual 8086**: Indirizzo e offset sono separati.

- CS = 0x80B8, IP = 0x019D  
→ indirizzo istruz. successiva:

$$\begin{array}{r}
 0x\ 80B8\ 0 + \\
 019\ D = \\
 \hline
 0x\ 80D1\ D
 \end{array}$$



## Gestione della memoria a segmenti



**Pagine:** segmenti di dimensioni fisse.

Molte architetture (a 64 bit) dividono lo spazio di indirizzamento in blocchi di pagine contigue di grandi dimensioni che vengono chiamati "segmenti", in modo improprio.

**I segmenti hanno:**

- Gli indirizzi fisici non vengono individuati in modo univoco (diverse combinazioni di indirizzo di segmento + offset) -> Estensione dello spazio di indirizzamento (complessità eccessiva, oggi abbandonata).
- Dimensione variabile (nelle architetture attuali)

Complessità per la protezione della memoria (numero di segmento + offset)

- Segmenti di dimensioni diverse -> offset può portare al di fuori di un segmento -> boundary check.



## Le architetture Intel avanzate



- **1982 – 80286: L'architettura diventa a 24 bit.** Viene utilizzata una modalità di utilizzo protetta che consente di mappare le pagine di memoria in indirizzi privati. Aggiunta di istruzioni specifiche.
- **1985 – 80386: Estensione a 32 bit.** Architettura, spazio di indirizzamento e registri a 32 bit. Nuove istruzioni, molto vicino ad un calcolatore con general-purpose registers. **Pre-fetching.** Paginazione della Memoria Principale.
- **1989-1992 – 80486:** Istruzioni per il trasferimento di dati condizionato. Cache. **Pipe-line singola.** Microprogrammazione per l'Unità di controllo (FSM).
- **1992-1995 – Pentium, PentiumPro:** Pipe-line multiple (super-scalare). Cache primaria e secondaria (separata, con bus dedicato). Introduzione dell'esecuzione condizionata. ISA quasi identico. Ampia ristrutturazione interna (=> **micro-architettura RISC:** micro-operazioni).
- **1997 – Pentium II:** Memorie cache a doppio accesso per ciclo di clock. Cache dei registri di segmento. PentiumPro + MMX (Multi-media extension, SIMD su dati piccoli: byte).
- **1999 – Pentium III:** "Internet Streaming single instruction multiple data extensions (SSE, calcolo vettoriale). Estensione dell'architettura MMX a istruzioni su **dati floating-point. L2 cache e CPU nello stesso integrato.** 8 registri da 128 bit per supporto ad esecuzione parallela (multiple-issue).



## Le architetture più recenti



- **2001 – Pentium 4:** Estensione del parallelismo e della Superscalarità. *Hyper-Threading Technology.* SSE2. I registri a 128 bit supportano anche virgola mobile in doppia precisione. **Register renaming e gestione separate delle code di esecuzione.** Massimizzazione delle prestazioni.
- **2003 – AMD:** Architettura a 64 bit. Aumento dello spazio di indirizzamento a 64 bit e registri a 64 bit. **16 registri GP + 16 registri a 128 bit.**
- **2004 – EM64T:** (Extended Memory 64 Technology). SSE3 che supporta numeri complessi e strutture. **Operazione atomica su blocchi di memoria di 128 bit.**
- **2006 – Virtual machines support:** SSE4 con operazioni quali somma di differenze in modulo, conteggio elementi, .... **Supporto alle macchine virtuali.**
- **2006 – Intel Itanium 2 9000 serie – Dual Core CPU.**
- **2007 – AMD.** Istruzioni su 3 operandi.
- **2007 – Intel.** Advanced Vector Extension. Registri da 128 a 256 bit.
- **2011 – Intel Core i7 Sandy Bridge. AVE.** Registri SSE fino 256 bit, con modalità di esecuzione vettoriale, **integrazione GPU.**
- **2013 – Canali multipli** per il trasferimento dalla memoria principale.
- **2017 – Intel Core i9.** Supporto alla crittografia.

### Trend:

«Fuse add and multiply»

Esecuzione vettoriale

Protezione dei dati



## Architettura x86



- CISC
  - Lunghezza istruzioni: 1 – 15 bytes
  - Operazioni direttamente in memoria
- Architettura condizionata dalla storia → necessità di compatibilità verso il basso
  - Real mode/Protected mode/Virtual 8086 Mode
- Nasce come architettura “not” [general-purpose register](#)
  - ogni registro è progettato per un uso specifico
  - dal 80386 in poi (IA-32) si definiscono 8 GP registers, ma non veri GP registers come in MIPS



## Modi di funzionamento IA-32



- **Modalità reale (Real mode)**
  - Modalità compatibile DOS
  - Max memoria indirizzabile: 1 MByte →  $2^{20}$
  - modo attivo all'accensione (power-on)
- **Modalità protetta (Protected mode)**
  - modalità “nativa” di IA-32
  - Memoria indirizzabile: 4 GByte →  $2^{32}$
  - Memoria protetta: evita corruzione memoria da parte di altri programmi
  - Memoria virtuale: permette ad un programma di disporre di più memoria di quella fisica disponibile
- **Modalità “8086 virtuale” (Virtual-8086 mode)**
  - “Real mode” simulato all'interno del “Protected Mode”
  - Esecuzione di programmi DOS in multitasking con altri programmi che girano in Protected Mode.



## Indirizzamento in modalità protetta



- Modalità **protetta**
  - Spazio di indirizzamento: 32 bit
  - I segmenti sono di fatto pagine di 64Kbyte e non più ogni 16 byte.
    - SALTO: indirizzo = composizione registri CS e IP
- Esempio:
  - CS = 0x80B8, IP = 0x019D
  - indirizzo istruz. Successiva mediante pipe dei due registri:  
**0x 80B8 || 019D**



## Architetture IA-64: caratteristiche



- 2 tipi di architetture a 64 bit: x64 (AMD) and IA64 (Intel). x64 è la più utilizzata. E' stata data in licenza a Intel che la ha sviluppata nell'architettura Itanium assieme ad HP.
- Modalità a 64 bit (**long mode**): è compatibile con le Architetture a 32 bit (**legacy mode**).
- Paginazione gerarchica su 4 livelli (**PAE – Physical Address Extension**). Segmentazione ignorata.
- Introduzione di un “**not executable**” bit nello stato di ogni pagina (protezione addizionale).
- Registri a 64 bit -> dati su 64 bit.
- **16 Registri general purpose**: rax, rbx .... + r0 ... r 15.
- Istruzioni registro-registro. E' un'architettura load-store a tutti gli effetti. La parte di load-store e' gestita in fase di pre-fetch.
- Le istruzioni sono organizzate in gruppi di istruzioni, **bundle**, con formato FISSO e designazione **esplicita** delle dipendenze (**issue slot**). Sono istruzioni che non hanno criticità (vere) tra loro e che sono separate da un gruppo successive da un separatore (stop).
- Architetture Itanium hanno una dotazione interna ricca di registri (**register renaming**).
- Istruzioni speciali e funzionalità per l'eliminazione dei salti (**loop unrolling**).
- Estensioni vettoriali: registri xmm, ymm, zmm (da 128 a 512 bit).



## Bundle IA-64



Le istruzioni sono codificate su 128 bit: bundle (VLIW). 2 bundle per ciclo di clock avviati a esecuzione (6 istruzioni / ciclo\_clock).

Ogni bundle ha un template (formato) di 5 bit + 3 istruzioni (ciascuna di 41 bit).

Il formato (template) specifica quale tra 5 differenti unità di esecuzione è richiesta da ciascuna delle 3 istruzioni (M-unit (memory instructions), I-unit (integer ALU, non-ALU integer, or long immediate extended instructions), F-unit (floating-point instructions), B-unit (branch or long branch extended instructions)).

Separatore di stop tra bundle: separa gruppi di bundle senza dipendenze.

Esecuzione predicativa:

if (p) espressione\_1

if (not\_p) espressione\_2

Se espressione\_i non deve essere eseguita, il risultato viene disabilitato (l'operazione viene sostituita con una nop) e i risultati nel commit buffer eliminati.



## 64 bits architectures



Operating mode	Operating sub-mode	Operating system required	Type of code being run	Default address size	Default operand size	Supported typical operand sizes	Register file	Typical GPR width
Long mode	64-bit mode	64-bit operating system or bootloader	64-bit code	64 bits	32 bits	8, 16, 32, or 64 bits	16 registers per file	64 bits
	Compatibility mode	64-bit operating system or bootloader	32-bit protected mode code	32 bits	32 bits	8, 16, or 32 bits	8 registers per file	32 bits
		64-bit operating system or bootloader	16-bit protected mode code	16 bits	16 bits	8, 16, or 32 bits	8 registers per file	32 bits
Legacy mode	Protected mode	32-bit operating system or bootloader, or 64-bit bootloader	32-bit protected mode code	32 bits	32 bits	8, 16, or 32 bits	8 registers per file	32 bits
		16-bit or 32-bit operating system or bootloader, or 64-bit bootloader	16-bit protected mode code	16 bits	16 bits	8, 16, or 32 bits	8 registers per file	16 or 32 bits
	Virtual 8086 mode	16-bit or 32-bit operating system	16-bit real mode code	16 bits	16 bits	8, 16, or 32 bits	8 registers per file	16 or 32 bits
	Real mode	16-bit or 32-bit operating system or bootloader, or 64-bit bootloader	16-bit real mode code	16 bits	16 bits	8, 16, or 32 bits	8 registers per file	16 or 32 bits

*From wikipedia*



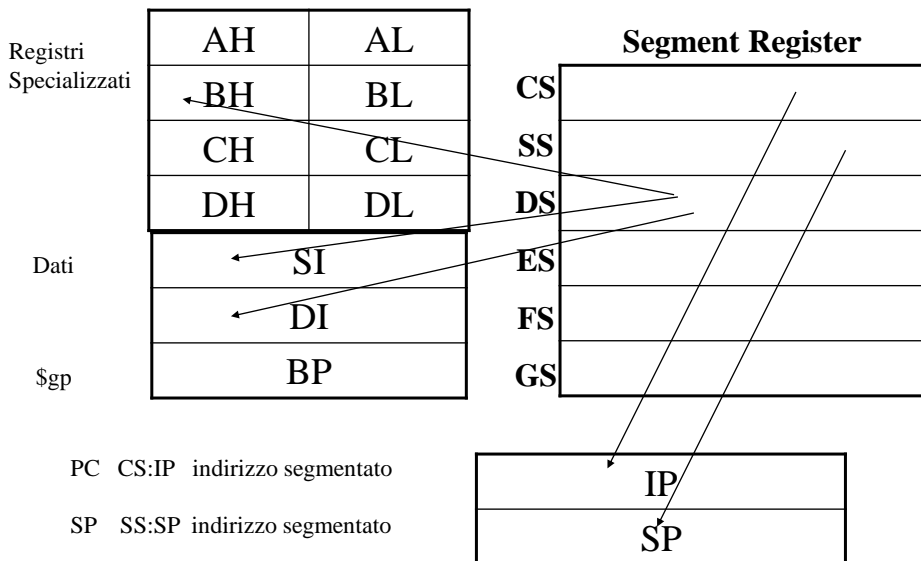
# Sommario



- Le architetture Intel
- I registri ed il loro utilizzo**
- L'ISA degli Intel
- La codifica delle istruzioni



# I registri dell'architettura 80286



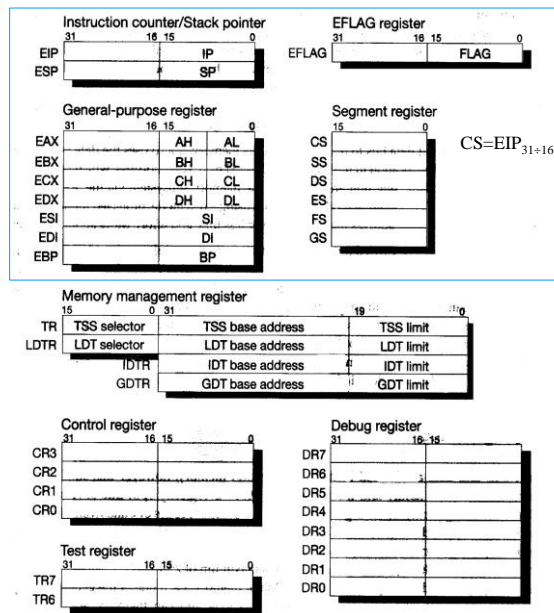




## I registri dell'architettura IA-32



- Estensione a 32 bit del set dei registri dell'80286.
- E.g. AH -> AX -> EAX
- ISA a partire dal 80386 (IA-32)
- 8 registri "general-purpose" a 32 bit (general purpose + ESP).
  - ma si deve potere accedere ad 1 o 2 byte al loro interno (compatibilità con 8086)
  - non sono poi così "general purpose"...
- I registri di segmento sono rimasti a 16 bit
  - usati come registri base
  - Possono essere utilizzati in alternativa ai registri estesi: (e.g. CS:IP -> EIP, SS:SP -> ESP, DS -> EBX).



A.A. 2019-2020

17/51

<http://borghese.di.unimi.it>



## Registri di Segmento (di memoria) a 16 bit



Nome	Descrizione	Utilizzo
CS    IP (EIP)	Code Segment	Puntatore al segmento di codice. Contiene l'indirizzo base dei (dati) ed istruzioni ad accesso immediato. Le istruzioni del segmento sono indirizzate tramite il registro <b>EIP</b> (Extended Instruction Pointer). Per modificare il registro CS occorre una <i>chiamata a procedura far o una far jump</i> oppure un interrupt ( <i>int</i> ). In modo protetto viene verificato se il nuovo segmento può essere utilizzato dal task.
DS (EBX)	Data Segment	Puntatore al segmento dati. <b>Contiene l'indirizzo base del segmento dati del programma (\$gp)</b> Molte istruzioni quali la <b>mov</b> utilizzano questo segmento. E' il segmento in cui sono contenuti i dati di un task. In modo protetto viene verificato se il nuovo segmento può essere utilizzato dal task.
SS    SP (ESP)	Stack Segment	Puntatore al segmento di stack. Quasi del tutto simile allo stack del MIPS. Cresce verso il basso. Contiene i dati locali delle procedure e gli argomenti di chiamata. Contiene anche <b>gli operandi per le operazioni aritmetiche</b> di tipo accumulatore. In modo protetto viene verificato se il nuovo segmento può essere utilizzato dal task.
ES, FS, GS	Extra Segments	Puntatori aggiuntivi al segmento dati. Principalmente utilizzati per operazioni su stringhe. Possono essere utilizzati in sostituzione di DS per accedere a dati al di fuori di DS. Il DOS ed il BIOS utilizzano spesso ES come buffer per le loro chiamate.

A.A. 2019-2020

Si sovrappongono ai registri a 32 bit (e.g. SS → [BP, DS] = EBP) [|borghese.di.unimi.it](http://borghese.di.unimi.it)



## Utilizzo dei registri IA-32



- **General Purpose Registers**
  - **General Data Registers**
    - **EAX**: accumulatore (ottimizzato per op. aritmetico-logiche) – unico registro per gli operandi nell'architettura 8080 -> operandi in stack.
    - **EBX**: base register (registro base nel segmento dati, indirizzo corrente, \$gp)
    - **ECX**: counter (ottimizzato per i cicli)
    - **EDX**: data register (contiene i dati)
  - **General Address Registers**
    - **EBP**: stack base pointer (base address dello stack, \$fp)
    - **ESI**: source index (ottimizzato per op. su stringhe, e.g. movs)
    - **EDI**: destination index (ottimizzato per op. su stringhe, e.g. movs)
- **Program counter e stack pointer**
  - **EIP**: extended instruction pointer (EIP = Program Counter = SS || IP)
  - **ESP**: stack pointer (\$sp) (SS || SP)
- **Floating-point Stack registers**
  - ST(0) – ST(7): 80 bit, accessibili come LIFO (stack, cf. registri \$fp)
- **Debug registers**
  - DR(0) – DR(7): 32 bit, debugging della CPU.
- **SIMD Registers**
  - MMX, SSE (SSE1, SSE2, SSE3, SSE4, SSE5)
- **Coordinamento tra: EBX e DS; tra EIP e SS; tra EBP, ESP e SS.**

.unimi.it\



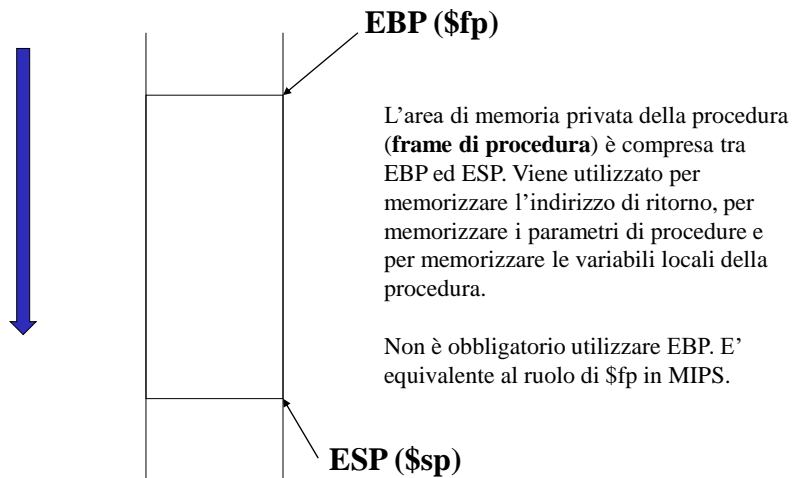
## Utilizzo dei registri “General Data”



Nome simbolico			Nome descrittivo	Funzioni
32 bit	16 bit	8 bit		
EAX	AX	AH, AL	Accumulator	Moltiplicazione/Divisione, I/O, shift veloce
<code>out 70h, al</code> ; Il contenuto di al viene trasferito alla porta 70h. <code>sb \$al, offset(\$k1)</code> ; In MIPS, accesso memoria > 2Gbyte				
EBX	BX	BH, BL	Base Register	Puntatore all'indirizzo base segmento dati
<code>mov edx, [ebx]</code> ;trasferisci quanto presente all'indirizzo 0(\$ebx) in edx <code>lw \$s0, 0(\$gp)</code>				
ECX	CX	CH, CL	Count Register	Indice di conteggio per cicli, rotazioni e shift
<code>move ecx, 10h</code> ; load ecx con 10h (=16), valore di inizio conteggio (associato a “loop”) <code>start: out 70h, al</code> ; Il contenuto di al viene trasferito alla porta 70h. <code>loop start</code> ; ritorna ad inizio ciclo, il quale verrà ripetuto 16 volte (fino a che ecx = 0)				
EDX	DX	DH, DL	Data Register	Moltiplicazione/Divisione, I/O
<code>mul edx</code> ; moltiplica edx con eax (implicito), il risultato è contenuto nella coppia edx:eax. ; in MIPS mult \$t1, \$t2 (risultato nella coppia di registri Hi:Lo)				



## Lo stack viene delimitato



## Registri di stack

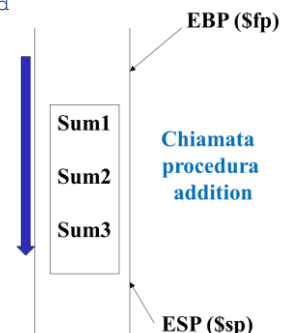
Nome simbolico			Nome descrittivo	Funzioni
32 bit	16 bit	8 bit		
ESP	SP	x,x	Stack Pointer	Stack Pointer
EBP	BP,SS	x, x	Base Pointer	Indirizzo base del segmento di Stack (32 bit)

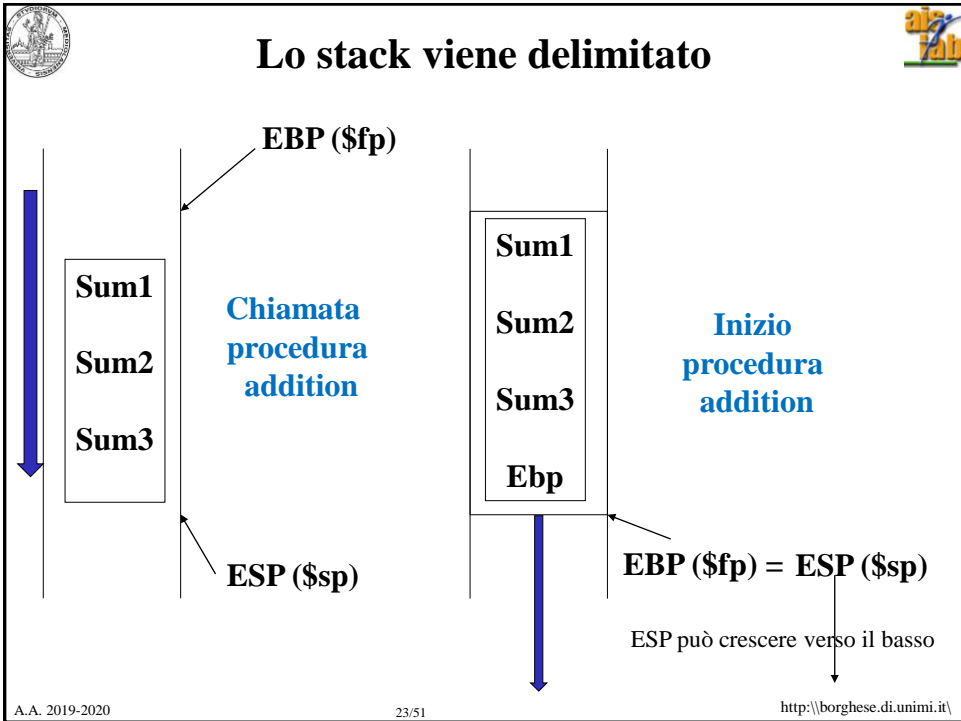
```

push sum1           ; create the first summand
                   ; addi $sp, $sp, -4; lw $s0, 0($sp)
                   ; (automaticamente ESP viene decrementato)
push sum2           ; create the second summand
push sum3           ; create the third summand
call addition       ; chiama procedura
                   ; addition (jal addition)
    
```

Situazione dello stack all'atto della chiamata  
Il frame di procedura del chiamante

Il frame per addition inizierà da ESP e crescerà verso il basso





## Registri di stack

Nome simbolico			Nome descrittivo	Funzioni
32 bit	16 bit	8 bit		
ESP	SP	x, x	Stack Pointer	Stack Pointer
EBP	BP, SS	x, x	Base Pointer	Indirizzo base del segmento di Stack (32 bit)

```

call addition ; chiama procedura addition (jal addition)

addition: proc near ; call near (inside 64k segment, cf. branch)
push ebp ; salva l'indirizzo base per il ritorno
; (inizio del record attivazione)
move ebp, esp ; copia lo StackPointer (esp) nel BaseP (individua
; l'inizio del frame di procedura per addition)
  
```

**EBP (\$fp)**

**Sum1**  
**Sum2**  
**Sum3**

**Chiamata procedura addition**

**ESP (\$sp)**

**Sum1**  
**Sum2**  
**Sum3**  
**Ebp**

**Inizio procedura addition**

**ESP (\$sp) = EBP (\$fp)**

ESP può crescere verso il basso



## Registri di stack



Architettura ad accumulatore

Nome simbolico			Nome descrittivo	Funzioni
32 bit	16 bit	8 bit		
ESP	SP	x,x	Stack Pointer	Stack Pointer
EBP	BP,SS	x, x	Base Pointer	Indirizzo base del segmento di Stack (32 bit)

```

push sum1          ; create the first summand
                   ; addi $sp, $sp, -4; lw $s0, 0($sp)
                   ; (automaticamente ESP viene decrementato)

push sum2          ; create the second summand
push sum3          ; create the third summand
call addition      ; chiama procedura addition (jal addition)

addition: proc near ; call near (inside 64k segment, cf. branch)
push ebp          ; salva l'indirizzo base per il ritorno
                 ; (inizio del record attivaz.)

move ebp, esp     ; copia lo StackP (esp) nel BaseP (individua
                 ; frame di procedura)

move eax, [ebp+12] ; carica sum1 in EAX
add eax, [ebp+8]   ; somma in eax sum1 + sum2
add eax, [ebp+4]   ; somma in eax sum1 + sum2 + sum3
pop ebp           ; recupera l'indirizzo base precedente
ret               ; ritorno al programma chiamante (cf. jr $ra)
addition endp

```

.unimi.it

**Indirizzi rispetto a EBP**



## IA-32 – operazioni logico-aritmetiche e confronto con il MIPS



Tipo operando 1	Tipo operando 2	Tipo risultato
Registro	Registro	Registro
Registro	Immediato	Registro
<b>Registro</b>	<b>Memoria</b>	<b>Registro</b>
<b>Memoria</b>	<b>Registro</b>	<b>Memoria</b>
<b>Memoria</b>	<b>Memoria</b>	<b>Memoria</b>

- Uno dei registri (memoria) deve fungere sia da operando che da registro destinazione (architettura ad accumulatore) E.g.: `add eax, [ebp + 8]` → `eax + [ebp + 8]` → `eax`.
  - MIPS permette di avere operandi e risultato in registri differenti
- Uno od entrambi gli operandi può provenire direttamente dalla memoria
  - Nel MIPS o nel PowerPC solo dai registri
- Gli operandi Immediati possono arrivare a 32 bit, gli altri ad 80 bit



## Dati di tipo diverso



- Possibilità di indirizzare dati da 8, 16, 32 (e 64 nell'estensione a 64 bit) bit.
- INTEL definisce un'ampiezza di default dei dati (16, 32 o 64 bit) .
- L'ampiezza di default viene impostata e modificata attraverso un **prefisso di 8 bit** inserito prima dell'istruzione (fino a 32 bit).

### Ruolo del prefisso (1 byte) nell'8086:

- modificare i registri di segmento
- impedire l'accesso al bus
- ripetere l'istruzione (fino a quando ECX = 0, pensato per la move di stringhe o blocchi)
- dimensione di default dei dati e degli indirizzi.



## Codifica istruzioni – campi **reg** e **w**

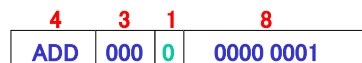


- Campo **reg** (cf. numero registro Register File in MIPS):

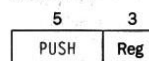
- la sua interpretazione dipende da **w**:
- **w** determina il tipo di dato (short o long: 8 o 16/32 bit).
- **w=0** → registri a 8 bit
- **w=1** → 16 o 32 bit (dipende dal setting dei dati di default)
  - IA-32: 32 bit di default

→ istruzioni di lunghezza diversa

- ADD AL, #1 → 16 bit
- ADD EAX, #1 → 40 bit



d.PUSH ESI



campo reg	w=0	w=1	
	8 bit	16 bit	32 bit
0	AL	AX	EAX
1	CL	CX	ECX
2	DL	DX	EDX
3	BL	BX	EBX
4	AH	SP	ESP
5	CH	BP	EBP
6	DH	SI	ESI
7	BH	DI	EDI

## Modalità di indirizzamento – dati

Modo	Descrizione	Restrizioni sui registri	Codice MIPS	Codice INTEL
<b>Ind. diretto (registro) Register Addressing</b>	Dato contenuto in un registro	No ESP e EBP	move \$s0, \$s1	mov eax, ebx
<b>Indiretta tramite registro</b>	Indirizzo del dato contenuto in un registro	No ESP e EBP	lw \$s0, 0(\$s1)	mov eax, [ebx]
<b>Base + offset (8 – 32bit)</b>	Indirizzo = contenuto del registro base più offset	No ESP	lw \$s0, 100(\$s1)	mov eax, 100 ([ebx] è sottinteso)
<b>Base + indice scalato</b>	Indirizzo = base + $2^{\text{scale}} * \text{indice}$	Base: GPR qualsiasi Indice: no ESP	mul \$t0, \$s2, 4 add \$t0, \$t0, \$s1 lw \$s0, 0(\$t0)	mov eax, [esi*scale] ([ebx] è sottinteso)
<b>Base + indice scalato + offset</b>	Indirizzo = base + $2^{\text{scale}} * \text{indice}$ + offset	Base: GPR qualsiasi Indice: no ESP	mul \$t0, \$s2, 4 add \$t0, \$t0, \$s1 lw \$s0, 24(\$t0)	mov eax, [esi*scale + 24] ([ebx] è sottinteso)

\$s1 – primo elemento del vettore --- 24 offset

Offset = Base + (Index \* Scale) + Displacement

**Potenza espressiva.** Esempio, caricamento in EAX (sottinteso) dell' *i*-esimo elemento di un vettore:

```

mov EBX, 0x0x224H ; indirizzo base
mov ESI, i
mov EAX, [ESI*4]

```

**Interno ciclo**

<http://borghese.di.unimi.it/>

## Modalità di indirizzamento – istruzioni “immediate”

Modo	Descrizione	Codice MIPS	Codice INTEL
<b>Indirizzamento immediato</b>	L'operando è in una parte dell'istruzione	li \$s0, 0x6a02	add eax, 0x6a02
<b>PC_relative addressing</b>	Indirizzamento relativo al Program Counter (edx sottinteso)	bne \$s0, \$s1, label	sub eax jnz label
<b>Pseudodirect addressing</b>	<b>MIPS:</b> indirizzo ottenuto cambiando i 26 bit dell'istruzione con i 28 LSB del PC (i 2 LSB sono 00) <b>INTEL:</b> modifica del registro Code Segment	j label	mov cs, 0x87ee mov ip 0x14fa = jump label = mov eip 0x87ee14fa

**Esecuzione condizionata.** Esegui un'operazione in modo condizionato. Il risultato viene scritto se il flag (registro EFLAG) assume un certo valore. sub EAX, scrivere anche il flag di «Zero».

Non c'è più il limite dell'allineamento ai 32 bit della parola

A.A. 2019-2020 30/51 <http://borghese.di.unimi.it/>



## Specifica dell'indirizzamento dati



w – specifica l'ampiezza (parte del codice operativo, 8, 16 o 32 bit)  
Reg – numero del registro.

Per alcune istruzioni il campo w e reg sono sufficienti.

Per altre istruzioni il codice operativo specifica direttamente anche il registro su cui si opera (e.g. ADD -> eax, ax) e la modalità di indirizzamento.

Per altre istruzioni esiste un post-byte (cf. campo function delle istruzioni di tipo R) per definire registry e modalità di indirizzamento. **Per la modalità di indirizzamento con indice scalato esiste un secondo post-byte.**

Schema che consente soluzioni diverse, molto utilizzato nelle architetture Intel.

### Post-byte 1:

- Reg (3 bit) – numero di registro
- Mod (2bit) – modalità di indirizzamento
- r/m (3 bit) – variante della modalità di indirizzamento.



## Sotto-campo reg



Indica uno degli 8 registri in forma compatta (8 bit) o in forma estesa (16/32 bit).

Equivalente al campo Reg.

reg	w = 0	w = 1	
		16b	32b
0	AL	AX	EAX
1	CL	CX	ECX
2	DL	DX	EDX
3	BL	BX	EBX
4	AH	SP	ESP
5	CH	BP	EBP
6	DH	SI	ESI
7	BH	DI	EDI





## Codifica indirizzi – campi **r/m** e **mod**



reg	w = 0	w = 1		r/m	mod = 0		mod = 1		mod = 2		mod = 3
		16b	32b		16b	32b	16b	32b	16b	32b	
0	AL	AX	EAX	0	addr=BX+SI	=EAX	same	same	same	same	same
1	CL	CX	ECX	1	addr=BX+DI	=ECX	addr as	addr as	addr as	addr as	as
2	DL	DX	EDX	2	addr=BP+SI	=EDX	mod=0	mod=0	mod=0	mod=0	reg
3	BL	BX	EBX	3	addr=BP+SI	=EBX	+ disp8	+ disp8	+ disp16	+ disp32	field
4	AH	SP	ESP	4	addr=SI	=(sib)	SI+disp8	(sib)+disp8	SI+disp8	(sib)+disp32	"
5	CH	BP	EBP	5	addr=DI	=disp32	DI+disp8	EBP+disp8	DI+disp16	EBP+disp32	"
6	DH	SI	ESI	6	addr=disp16	=ESI	BP+disp8	ESI+disp8	BP+disp16	ESI+disp32	"
7	BH	DI	EDI	7	addr=BX	=EDI	BX+disp8	EDI+disp8	BX+disp16	EDI+disp32	"

- **r/m (3 bit)** seleziona il registro usato come **registro base** dell'indirizzo
- **mod (2 bit)** seleziona la **modalità di indirizzamento** (+ offset, offset+displacement,...)

### mod = 3

Registro base dell'indirizzo = registro identificato dal campo reg.



## Sotto-campo **r/m** (3 bit) e **mod** (2 bit)



r/m	mod = 0		mod = 1		mod = 2		mod = 3
	16b	32b	16b	32b	16b	32b	
0	addr=BX+SI	=EAX	same	same	same	same	same
1	addr=BX+DI	=ECX	addr as	addr as	addr as	addr as	as
2	addr=BP+SI	=EDX	mod=0	mod=0	mod=0	mod=0	reg
3	addr=BP+SI	=EBX	+ disp8	+ disp8	+ disp16	+ disp32	field
4	addr=SI	=(sib)	SI+disp8	(sib)+disp8	SI+disp8	(sib)+disp32	"
5	addr=DI	=disp32	DI+disp8	EBP+disp8	DI+disp16	EBP+disp32	"
6	addr=disp16	=ESI	BP+disp8	ESI+disp8	BP+disp16	ESI+disp32	"
7	addr=BX	=EDI	BX+disp8	EDI+disp8	BX+disp16	EDI+disp32	"

### r/m = 0, 1, 2, 3, 7

mod = 0 indica il registro base dell'indirizzo utilizzato di norma

mod = 1 coincide con mod = 0 con la differenza di un offset su 8 bit (disp8)

mod = 2 coincide con mod = 0 con la differenza di un offset su 16 / 32 bit (disp16 / disp 32)

Registro a 32 bit per parole di 32 bit

Somma di due registri a 16 bit per parole di 16 bit (indirizzamento di segmento + offset)



## sotto-campo r/m = 5, 6



r/m	mod = 0		mod = 1		mod = 2		mod = 3
	16b	32b	16b	32b	16b	32b	
0	addr=BX+SI	=EAX	same	same	same	same	same
1	addr=BX+DI	=ECX	addr as	addr as	addr as	addr as	as
2	addr=BP+SI	=EDX	mod=0	mod=0	mod=0	mod=0	reg
3	addr=BP+SI	=EBX	+ disp8	+ disp8	+ disp16	+ disp32	field
4	addr=SI	=(sib)	SI+disp8	(sib)+disp8	SI+disp8	(sib)+disp32	"
5	addr=DI	=disp32	DI+disp8	EBP+disp8	DI+disp16	EBP+disp32	"
6	addr=disp16	=ESI	BP+disp8	ESI+disp8	BP+disp16	ESI+disp32	"
7	addr=BX	=EDI	BX+disp8	EDI+disp8	BX+disp16	EDI+disp32	"

### r/m = 5 – architettura a 16 bit

- Come per r/m = 0, 1, 2, 3, 7

### r/m = 5 – architettura a 32 bit

- mod = 0 → Offset su 32 bit
- mod = 1,2 → Seleziona EBP + spiazamento (8 / 32 bit)

### r/m = 6 – architettura a 16 bit

- mod = 0 → Offset su 16 bit
- mod = 1,2 → indirizzo come BP + spiazamento (8 / 16 bit)

### r/m = 6 – architettura a 32 bit

- Come per r/m = 0, 1, 2, 3, 7



## sotto-campo r/m = 4



r/m	mod = 0		mod = 1		mod = 2		mod = 3
	16b	32b	16b	32b	16b	32b	
0	addr=BX+SI	=EAX	same	same	same	same	same
1	addr=BX+DI	=ECX	addr as	addr as	addr as	addr as	as
2	addr=BP+SI	=EDX	mod=0	mod=0	mod=0	mod=0	reg
3	addr=BP+SI	=EBX	+ disp8	+ disp8	+ disp16	+ disp32	field
4	addr=SI	=(sib)	SI+disp8	(sib)+disp8	SI+disp8	(sib)+disp32	"
5	addr=DI	=disp32	DI+disp8	EBP+disp8	DI+disp16	EBP+disp32	"
6	addr=disp16	=ESI	BP+disp8	ESI+disp8	BP+disp16	ESI+disp32	"
7	addr=BX	=EDI	BX+disp8	EDI+disp8	BX+disp16	EDI+disp32	"

### architettura a 16 bit

- Come per r/m = 0, 1, 2, 3, 7

### architettura a 32 bit – scaled index mode (sib)

- mod = 0 → sib + 0
- mod = 1 → sib + offset 8 bit
- mod = 2 → sib + offset 32 bit

esi, scala, offset are specified in a second post-byte

Modo	Descrizione	Restrizioni sui registri	Codice MIPS	Codice INTEL
Base + indice scalato + offset	Indirizzo = base + 2 <sup>scala</sup> * indice + offset	Base: GPR qualsiasi Indice: no ESP	mul \$t0, \$s2, 4 add \$t0, \$t0, \$s1 lw \$s0, 100(\$t0)	mov eax, [esi*scala + 4] ([ebx] è sottinteso)



# Sommario



- Le architetture Intel
- I registri ed il loro utilizzo
- L'ISA degli Intel
- La codifica delle istruzioni



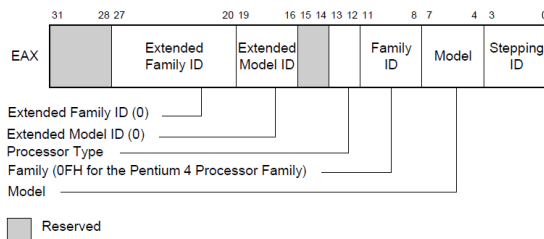
# Quale CPU viene utilizzata?

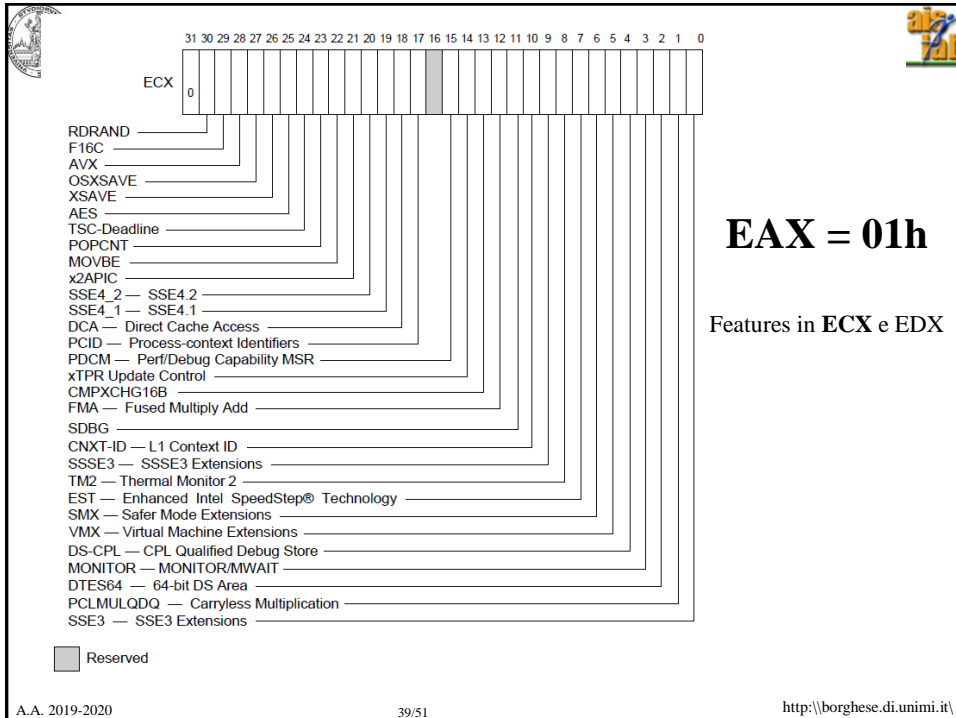


## CPUID—CPU Identification

Opcode	Instruction	64-Bit Mode	Compat/ Leg Mode	Description
0F A2	CPUID	Valid	Valid	Returns processor identification and feature information to the EAX, EBX, ECX, and EDX registers, as determined by input entered in EAX (in some cases, ECX as well).

CPUID has no arguments. La richiesta viene codificata in EAX. Disponibile dal Pentium (1993). Prima chiamata con EAX = 0h, restituisce il tipo di CPU:





Bit #	Mnemonic	Description
0	SSE3	Streaming SIMD Extensions 3 (SSE3). A value of 1 indicates the processor supports this technology.
1	PCLMULQDQ	A value of 1 indicates the processor supports PCLMULQDQ instruction.
2	DTES64	64-bit DS Area. A value of 1 indicates the processor supports DS area using 64-bit layout.
3	MONITOR	MONITOR/MWAIT. A value of 1 indicates the processor supports this feature.
4	DS-CPL	CPL Qualified Debug Store. A value of 1 indicates the processor supports the extensions to the Debug Store feature to allow for branch message storage qualified by CPL.
5	VMX	Virtual Machine Extensions. A value of 1 indicates that the processor supports this technology.
6	SMX	Safer Mode Extensions. A value of 1 indicates that the processor supports this technology. See Chapter 6, "Safer Mode Extensions Reference".
7	EST	Enhanced Intel SpeedStep® Technology. A value of 1 indicates that the processor supports this technology.
8	TM2	Thermal Monitor 2. A value of 1 indicates whether the processor supports this technology.
9	SSSE3	A value of 1 indicates the presence of the Supplemental Streaming SIMD Extensions 3 (SSSE3). A value of 0 indicates the instruction extensions are not present in the processor.
10	CNXT-ID	L1 Context ID. A value of 1 indicates the L1 data cache mode can be set to either adaptive mode or shared mode. A value of 0 indicates this feature is not supported. See definition of the IA32_MISC_ENABLE MSR Bit 24 (L1 Data Cache Context Mode) for details.
11	SDBG	A value of 1 indicates the processor supports IA32_DEBUG_INTERFACE MSR for silicon debug.
12	FMA	A value of 1 indicates the processor supports FMA extensions using YMM state.
13	CMPXCHG16B	CMPXCHG16B Available. A value of 1 indicates that the feature is available.
14	xTPR Update Control	xTPR Update Control. A value of 1 indicates that the processor supports changing IA32_MISC_ENABLE[bit 23].
15	PDCM	Perfmon and Debug Capability. A value of 1 indicates the processor supports the performance and debug feature indication MSR IA32_PERF_CAPABILITIES.
16	Reserved	Reserved
17	PCID	Process-context identifiers. A value of 1 indicates that the processor supports PCIDs and that software may set CR4.PCIDE to 1.
18	DCA	A value of 1 indicates the processor supports the ability to prefetch data from a memory mapped device.
19	SSE4.1	A value of 1 indicates that the processor supports SSE4.1.
20	SSE4.2	A value of 1 indicates that the processor supports SSE4.2.
21	x2APIC	A value of 1 indicates that the processor supports x2APIC feature.
22	MOVBE	A value of 1 indicates that the processor supports MOVBE instruction.
23	POPCNT	A value of 1 indicates that the processor supports the POPCNT instruction.
24	TSC-Deadline	A value of 1 indicates that the processor's local APIC timer supports one-shot operation using a TSC deadline value.
25	AES	A value of 1 indicates that the processor supports the AESNI instruction extensions.
26	XSAVE	A value of 1 indicates that the processor supports the XSAVE/XRSTOR processor extended states feature, the XSETBV/XGETBV instructions, and XCRO.
27	OSXSAVE	A value of 1 indicates that the OS has set CR4.OSXSAVE[bit 18] to enable XSETBV/XGETBV instructions to access XCRO and to support processor extended state management using XSAVE/XRSTOR.
28	AVX	A value of 1 indicates that processor supports AVX instructions operating on 256-bit YMM state, and three-operand encoding of 256-bit and 128-bit SIMD instructions.

**Features description**

A.A. http://borghese.di.unimi.it/



## Quale CPU viene utilizzata?



### CPUID—CPU Identification

Opcode	Instruction	64-Bit Mode	Compat/ Leg Mode	Description
0F A2	CPUID	Valid	Valid	Returns processor identification and feature information to the EAX, EBX, ECX, and EDX registers, as determined by input entered in EAX (in some cases, ECX as well).

### Eempio: istruzioni vettoriali

La maggior parte delle istruzioni di base dell'estensione AVX-512 lavorano sui registri zmm a 512 bit + 8 registri di stato (opmask).

Per sapere se il processore supporta le istruzioni AVX-512 viene chiamata CPUID con **EAX = 07H, ECX = 0** e viene analizzato il bit 16 del registro EBX.



## IA-32 Instruction Set



- Istruzioni **general purpose**
  - Istruzioni per lo spostamento dei dati
    - push, pop, utilizzo dello stack e della memoria dati
    - Tra registri
  - Istruzioni aritmetiche logiche, confronto e operazioni su interi e float
  - Istruzioni di controllo del flusso
    - basati sui flag (**condition codes**), allineamento al byte, salti condizionati e incondizionati, chiamata a procedura
  - Istruzioni della gestione delle stringhe (legacy mode)
  - Istruzioni di I/O
- Istruzioni **floating point** x87
  - funzioni trigonometriche, potenze di 2 ( $2^x$ ,  $\log_2 x$ )
- Istruzioni **SIMD**
  - MMX, SSE (SSE2, SSE3, SSE4, SSE5), 3DNow! (IA-32 by AMD)
- Istruzioni di sistema
  - Cambio di modo, Halt, Reset, ...



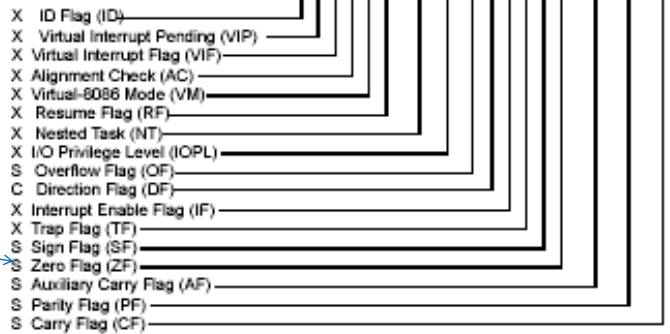
# Registro EFLAG



- I risultati notevoli vengono salvati in questo registro
  - Carry, zero, overflow, segno, parità, ...
  - Le istruzioni di branch si riferiscono sempre a EFLAG

**MIPS:**  
beq \$s0, \$s1, label

**IA-32:**  
sub eax  
jz label



S Indicates a Status Flag  
C Indicates a Control Flag  
X Indicates a System Flag



# Alcuni esempi di IA-32 Instruction Set



Instruction	Meaning
<b>Control</b>	<b>Conditional and unconditional branches</b>
jnz, jz	Jump if condition to EIP + 8-bit offset; JNE (for JNZ), JE (for JZ) are alternative names
jmp	Unconditional jump—8-bit or 16-bit offset
call	Subroutine call—16-bit offset; return address pushed onto stack
ret	Pops return address from stack and jumps to it
loop	Loop branch—decrement ECX; jump to EIP + 8-bit displacement if ECX ≠ 0
<b>Data transfer</b>	<b>Move data between registers or between register and memory</b>
move	Move between two registers or between register and memory
push, pop	Push source operand on stack; pop operand from stack top to a register
les	Load ES and one of the GPRs from memory
<b>Arithmetic, logical</b>	<b>Arithmetic and logical operations using the data registers and memory</b>
add, sub	Add source to destination; subtract source from destination; register-memory format
cmp	Compare source and destination; register-memory format
shl, shr, rcr	Shift left; shift logical right; rotate right with carry condition code as fill
cbw	Convert byte in eight rightmost bits of EAX to 16-bit word in right of EAX
test	Logical AND of source and destination sets condition codes
inc, dec	Increment destination, decrement destination
or, xor	Logical OR; exclusive OR; register-memory format
<b>String</b>	<b>Move between string operands; length given by a repeat prefix</b>
movs	Copies from string source to destination by incrementing ESI and EDI; may be repeated
lodsb	Loads a byte, word, or doubleword of a string into the EAX register



## Descrizione delle istruzioni



<b>JE, JZ:</b>	jump on equal - near ( $\pm 128$ byte)	(MIPS <code>beq...</code> , <code>beq \$0, reg, ...</code> )
<b>JMP:</b>	jump (near $\rightarrow$ uso CS; far $\rightarrow$ uso EIP)	(MIPS <code>j</code> )
<b>CALL:</b>	jump; SP=SP-4	(MIPS <code>jal</code> )
<b>MOVW:</b>	lettura di una parola dalla memoria	(MIPS <code>lw</code> )
<b>PUSH,POP:</b>	aggiornamento implicito SP	(MIPS <code>addi \$sp, \$sp, <math>\pm 4</math></code> )
<b>TEST:</b>	Carica nel flag il risultato di un'operazione (e.g <code>\$EDX AND 0x42</code> )	
<b>MOVSL:</b>	Sposta 4 byte da memoria a memoria e incrementa EDI ed ESI (stringhe/aree dati)	



## Sommario



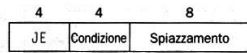
Le architetture Intel  
I registri ed il loro utilizzo  
L'ISA degli Intel  
**La codifica delle istruzioni**



# Codifica delle istruzioni

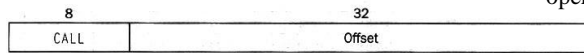


a. JE EIP + spiazamento

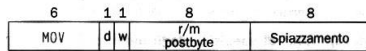


- Ampiezza variabile: 1-15 byte.
  - Codice operativo su 1 o 2 byte.
  - **CLC** (Clear Carry: 1 byte, non ha operandi)

b. CALL

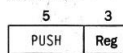


c. MOV EBX, [EDI + 45]

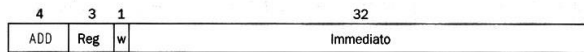


- **w** specifica se lavora sul **byte** o sulla parola a 16/32 bit (**word**)
- **d** specifica la **direzione** del trasferimento
- **Post\_byte r/m**, specifica la modalità di indirizzamento

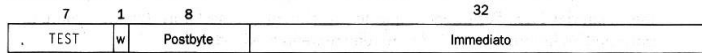
d. PUSH ESI



e. ADD EAX, #6765



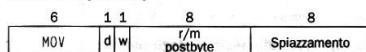
f. TEST EDX, #42



# Esempio



c. MOV EBX, [EDI + 45]

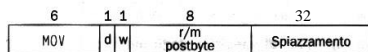


3 Byte

M[EDI + 45] -> EBX

- d = 1 (lettura)
- w = 1
- r/m = 7
- mod = 1**
- Spiazamento = 45

c. MOV EBX, [EDI + 45]



6 Byte

M[EDI + 45] -> EBX

- d = 1 (lettura)
- w = 1
- r/m = 7
- mod = 2**
- Spiazamento = 45

reg	w = 0	w = 1	r/m	mod = 0		mod = 1		mod = 2		mod = 3
	16b	32b		16b	32b	16b	32b	16b	32b	
0	AL	AX	EAX	0	addr=BX+SI =EAX	same	same	same	same	same
1	CL	CX	ECX	1	addr=BX+DI =ECX	addr as	addr as	addr as	addr as	as
2	DL	DX	EDX	2	addr=BP+SI =EDX	mod=0	mod=0	mod=0	mod=0	reg
3	BL	BX	EBX	3	addr=BP+SI =EBX	+ disp8	+ disp8	+ disp16	+ disp32	field
4	AH	SP	ESP	4	addr=SI = (sib)	SI+disp8	(sib)+disp8	SI+disp8	(sib)+disp32	"
5	CH	BP	EBP	5	addr=DI =disp32	DI+disp8	EBP+disp8	DI+disp16	EBP+disp32	"
6	DH	SI	ESI	6	addr=disp16 =ESI	BP+disp8	ESI+disp8	BP+disp16	ESI+disp32	"
7	BH	DI	EDI	7	addr=BX =EDI	BX+disp8	EDI+disp8	BX+disp16	EDI+disp32	"





## Esempio



f. TEST EDX, #42

7	1	8	32
TEST	w	Postbyte	Immediato

If (EDX == 42) Flag == 1

w = 1  
 r/m = 5  
 mod = 0  
 Immediato = 42

reg	w = 0	w = 1		r/m	mod = 0		mod = 1		mod = 2		mod = 3
		16b	32b		16b	32b	16b	32b	16b	32b	
0	AL	AX	EAX	0	addr=BX+SI =EAX	same	same	same	same	same	same
1	CL	CX	ECX	1	addr=BX+DI =ECX	addr as	addr as	addr as	addr as	as	as
2	DL	DX	EDX	2	addr=BP+SI =EDX	mod=0	mod=0	mod=0	mod=0	reg	reg
3	BL	BX	EBX	3	addr=BP+SI =EBX	+ disp8	+ disp8	+ disp16	+ disp32	field	field
4	AH	SP	ESP	4	addr=SI =(sib)	SI+disp8	(sib)+disp8	SI+disp8	(sib)+disp32	"	"
5	CH	BP	EBP	5	addr=DI =disp32	DI+disp8	EBP+disp8	DI+disp16	EBP+disp32	"	"
6	DH	SI	ESI	6	addr=disp16 =ESI	BP+disp8	ESI+disp8	BP+disp16	ESI+disp32	"	"
7	BH	DI	EDI	7	addr=BX =EDI	BX+disp8	EDI+disp8	BX+disp16	EDI+disp32	"	"



## Codifica istruzioni: osservazioni



- Architettura CISC
  - Lunghezza variabile istruzioni
  - Lunghezza variabile dei campi del Codice Operativo (primo Byte)
- La lunghezza dell'istruzione dipende dal contenuto di alcuni campi dell'istruzione stessa
  - mod, r/m, reg, w
  - Devo iniziare la decodifica dell'istruzione per sapere quant'è lunga → prima di terminare la fase di fetch
- Architettura complessa e poco razionale
  - prezzo da pagare per mantenere la compatibilità verso il basso (8, 16, 32 bit)
- **Periodicamente Intel considera di virtualizzare le architetture vintage**
- **Elenco delle istruzioni Intel** (alcune vengono mantenute solo per compatibilità):  
[https://en.wikipedia.org/wiki/X86\\_instruction\\_listings](https://en.wikipedia.org/wiki/X86_instruction_listings)



## Sommario



Le architetture Intel  
I registri ed il loro utilizzo  
L'ISA degli Intel  
La codifica delle istruzioni