

Corso di Architettura degli Elaboratori

Valutazione delle prestazioni

DOCENTE Luigi Palopoli
AA. 2011/2012



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Valutare le Prestazioni

- La complessita' di un moderno calcolatore rende la valutazione delle prestazioni tutt'altro che banale
- Quando si parla di prestazioni, e' importante capire:
 - ★ Di quali prestazioni parliamo
 - ★ Come si misurano



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Definizione di Prestazioni

- Consideriamo i seguenti aereoplani

Aereoplano	Capacità passeggeri	Autonomia (km)	Velocità di crociera (km/h)	Capacità oraria/ numero di passeggeri trasportati all'ora (passeggeri × km/h)
Boeing 777	375	7400	980	367 500
Boeing 747	470	6640	980	460 600
BAC/Sud Concorde	132	6400	2160	285 120
Douglas DC- 8-50	146	13 950	870	127 020

- Supponiamo che la metrica di interesse sia la velocità'.



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Definizione delle Prestazioni

- Cosa intendiamo per velocità?
- Possibili definizioni:
 - ★ Trasportare un singolo passeggero da un luogo a un altro nel minor tempo possibile
(Vincitore *Concorde*)
 - ★ Trasportare 450 passeggeri nel minor tempo possibile
(Vincitore *Boeing 747*)



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Definizione delle Prestazioni

- Analogamente per un calcolatore desktop possiamo avere due tipi di prestazioni.
 - ★ Per il singolo utente interessa sapere quanto e' il tempo medio di risposta (tempo che intercorre tra avvio e terminazione di un task)
 - ★ Per il gestore di un centro di calcolo, interessa di piu' il *throughput*, cioe' quanti task sono in grado di completare nell'unita' di tempo



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Esempio

- Se scelgo un processore piu' veloce miglioro il throughput o il tempo di risposta?

Entrambi

- Se uso un multiprocessore miglioro il throughput o il tempo di risposta?

Il throughput
almeno a
parita' di processore

Tuttavia, se il sistema
e' molto carico,
diminuisco il tempo di
attesa, e quindi
il tempo di risposta



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Tempo di Esecuzione

- In questo corso, definiremo le prestazioni soprattutto in termini di tempo di esecuzione
- Per un calcolatore X

$$\text{Prestazioni}_X = \frac{1}{\text{Tempo di esecuzione}_X}$$

- Quindi, dati due calcolatori diremo:

$$\text{Prestazioni}_X \geq \text{Prestazioni}_Y \leftrightarrow$$

$$\text{Tempo di esecuzione}_X \leq \text{Tempo di esecuzione}_Y$$



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Tempo di Esecuzione

- Analogamente diremo che il calcolatore X e' n volte superiore (piu' veloce) del calcolatore Y se e solo se:

$$\frac{\text{Prestazioni}_X}{\text{Prestazioni}_Y} = n$$

- Ma cosa si intende per tempo di esecuzione?

Il tempo di esecuzione di un programma e' il suo tempo di risposta: cioe' quanto tempo occorre per il completamento di un task (mettendo insieme esecuzione, IO, latenze di sistema operativo, ecc.)

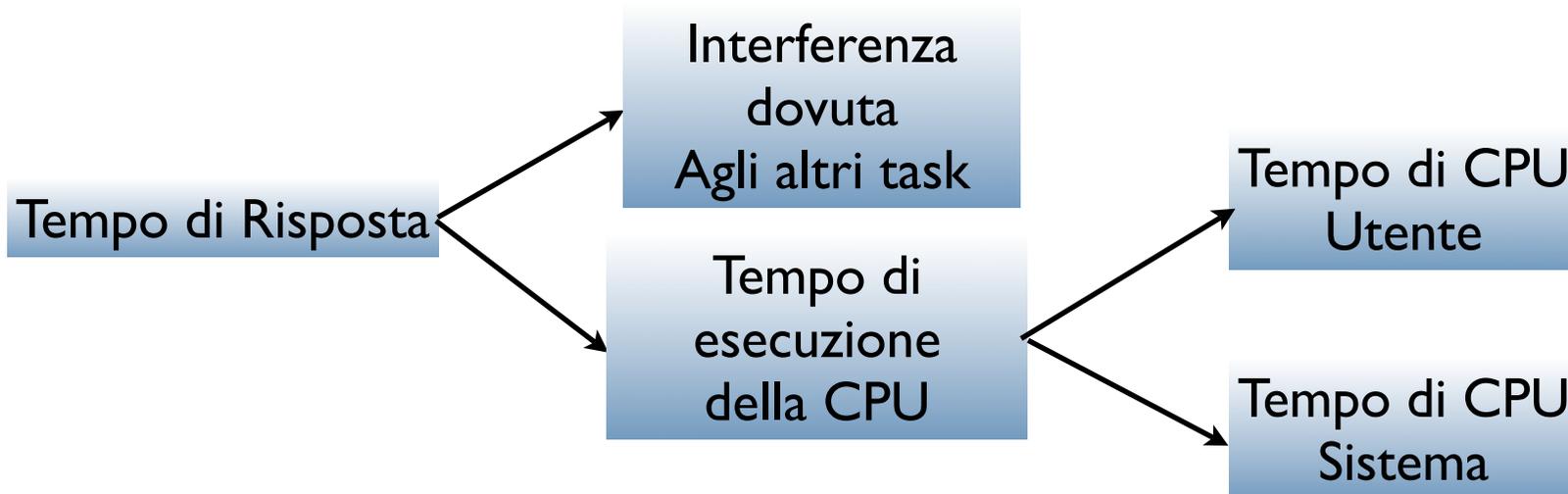


UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Tempo di Esecuzione

- In una macchina multiprogrammata il tempo di risposta dipende anche dagli altri task attivi e dalle loro priorità
- Il tempo di esecuzione della CPU tiene conto solo del tempo effettivamente speso per il task
- Tale tempo è in parte dedicato al programma *utente* e in parte al *sistema operativo*





UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Capire le Prestazioni

- I moderni processori, come vedremo, sono costruiti usando un segnale periodico che ne sincronizza le operazioni
- Il ciclo di clock è l'intervallo di tempo che intercorre tra due colpi di clock (la frequenza ne è l'inverso)
- Il ciclo di clock è misurato in secondi (o in frazioni di secondo), la frequenza in Hertz (o equivalentemente in cicli al secondo)
- Ad esempio un clock che va a un Gigahertz (10^9 Hertz) equivale a un periodo di clock pari a 10^{-9} secondi (un miliardesimo di secondo) 



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Calcolo del Tempo di CPU

- Il tempo di calcolo di un programma puo' essere espresso nella seguente maniera

$$\begin{aligned}\text{Tempo CPU Programma} &= \text{Cicli di clock della CPU} \times \text{Periodo del Clock} \\ &= \frac{\text{Cicli di clock della CPU}}{\text{Frequenza del Clock}}\end{aligned}$$



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Esempio

Il nostro programma esegue in 10s sul calcolatore A che e' dotato di un clock che funziona a 2GhZ. Un progettista vuole proporci una nuova architettura che puo' operare a frequenze significativamente piu' alte. Propone una modifica al processore che porterebbe ad aumentare di un fattore 1.2 i cicli macchina necessari a eseguire le varie istruzioni. A che frequenza dovrebbe operare la macchina per portare il nostro programma a eseguire in 6s?



Esempio

- Come prima cosa calcoliamo i cicli di clock sul computer A

$$\text{Tempo CPU Programma}_A = \frac{\text{Cicli di clock della CPU}_A}{\text{Frequenza del Clock}_A}$$

$$10s = \frac{\text{Cicli di clock della CPU}_A}{2 \times 10^9}$$

$$\text{Cicli di clock della CPU}_A = 20 \times 10^9 \text{ cicli}$$

- Ora possiamo calcolare la frequenza di clock necessaria

$$\text{Tempo CPU Programma}_B = \frac{1.2 \times \text{Cicli di clock della CPU}_A}{\text{Frequenza del Clock}_B} \quad \text{💬}$$

$$6s = \frac{1.2 \times 20 \times 10^9}{\text{Frequenza del Clock}_B}$$

$$\text{Frequenza del Clock}_B = \frac{1.2 \times 20 \times 10^9}{6s}$$

$$= 4\text{Ghz}$$



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Numero di cicli

- Il numero di cicli richiesto da un programma e' influenzato dal numero di istruzioni che compongono il programma
- Tale numero viene moltiplicato per il numero di cicli medio richiesto da ciascuna istruzione

Cicli di clock della CPU = Istruzioni del Programma \times
Numero medio di cicli per istruzione

- Il numero medio di cicli per istruzione viene denotato da CPI



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Esempio

Siano date due diverse implementazioni della stessa ISA. La prima (A) ha un ciclo di clock di 250ps, la seconda (B) di 500ps. Nel programma di riferimento che ha numero di istruzioni I il CPI è pari a 2.0 per A e a 1.2 per B. Quale delle due architetture esegue più rapidamente il nostro programma?



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Esempio

- **Architettura A**

$$\begin{aligned}\text{Tempo di CPU}_A &= \text{Cicli di Clock}_A \times \text{Periodo di Clock}_A \\ &= \text{Istruzioni del Programma} \times \text{CPI}_A \times \text{Periodo di Clock}_A \\ &= I \times 2.0 \times 250ps \\ &= I \times 500ps\end{aligned}$$

- **Architettura B**

$$\begin{aligned}\text{Tempo di CPU}_B &= \text{Cicli di Clock}_B \times \text{Periodo di Clock}_B \\ &= \text{Istruzioni del Programma} \times \text{CPI}_B \times \text{Periodo di Clock}_B \\ &= I \times 1.2 \times 500ps \\ &= I \times 600ps\end{aligned}$$



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Esempio

- Da questo concludiamo

$$\frac{\text{Prestaz. CPU}_A}{\text{Prestaz. CPU}_B} = \frac{\text{Tempo CPU}_B}{\text{Tempo CPU}_A} = \frac{600}{500} = 1.2$$



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Equazione Classica

- Siamo ora in grado di scrivere l'equazione classica delle prestazioni

$$\begin{aligned}\text{Tempo CPU} &= \text{Numero Istruzioni} \times \text{CPI} \times \text{Periodo Clock} \\ &= \frac{\text{Numero Istruzioni} \times \text{CPI}}{\text{Frequenza Clock}}\end{aligned}$$



Esempio di Uso

- Un progettista di compilatori deve scegliere tra due sequenze di codice per implementare un certo tipo di programma

CPI per tipo di Istruzione

Istr/ CPI	A	B	C
	1	2	3

Istruzioni di ciascun tipo

Sequenza	A	B	C
Seq. 1	2	1	2
Seq. 2	4	1	1



Soluzione

- Possiamo calcolare il numero di cicli richieste da ciascuna sequenza usando la formula:

$$\text{Cicli CPU}_i = \sum_{i=1}^n (CPI_i \times C_i)$$

$$\text{Cicli CPU}_1 = (2 \times 1) + (1 \times 2) + (2 \times 3) = 10$$

$$\text{Cicli CPU}_2 = (4 \times 1) + (1 \times 2) + (1 \times 3) = 9$$

$$CPI_1 = \frac{10}{5} = 2$$

$$CPI_2 = \frac{9}{6} = 1.5$$

- Quindi conviene l'architettura 2



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Tirando un po' di somme...

- La maniera migliore di valutare le prestazioni di una computer e' di misurare il tempo necessario per l'esecuzione di un programma
- Tale tempo e' il risultato di tre fattori:
 - ★ Numero istruzioni
 - ★ CPI
 - ★ Frequenza di clock

Componente delle prestazioni	Unità di misura
Tempo di esecuzione della CPU per un programma	Secondi per programma
Numero di istruzioni	Istruzioni eseguite per programma
Cicli di clock per istruzione (CPI)	Numero medio di cicli di clock per istruzione
Ciclo di clock	Secondi per ciclo di clock



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Componenti delle prestazioni

- Nessuna delle tre componenti può essere trascurata
 - ★ Non ha nessun senso dire che un processore è più veloce di un altro perché ha un clockrate più alto
- Ha dunque senso cercare di capire come i diversi componenti e gli strumenti usati nello sviluppo abbiano impatto sulle prestazioni (in particolare su ciascuna dei tre fattori)



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Algoritmo

- L'algoritmo adottato certamente influenza il *numero di istruzioni* ed eventualmente il *CPI*
- Perché?
 - ★ Un algoritmo più efficiente può essere strutturato in modo da risparmiare istruzioni
 - ★ Un algoritmo ben pensato (per una particolare architettura) andrà ad usare le istruzioni più efficienti (quelle con un basso



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Linguaggio di programmazione

- Il linguaggio di programmazione influenza il *numero di istruzioni* e il *CPI*
- Perché?
 - ★ I costrutti ad alto livello vengono tradotti in sequenze di istruzioni macchina
 - ★ Un linguaggio con molte chiamate indirette (es. Java) ha in generale un valore di CPI più alto



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Compilatore

- Il compilatore sicuramente influenza sia il numero di istruzioni che il CPI
- Perché?
 - ★ Un compilatore più o meno efficiente genera un numero di istruzioni macchina diverso per ogni istruzione
 - ★ Un compilatore ottimizzante può tenere conto di una serie di effetti piuttosto complessi per ridurre il CPI



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

ISA

- L'architettura del set di istruzioni (ISA) consiste nell'”interfaccia” che la macchina offre al software (in questo corso ne vedremo due)
- Essa ha impatto sul numero di istruzioni, sulla frequenza di clock sul CPI
- Perché?
 - ★ **Numero di istruzioni:** l'ISA può fornire istruzioni di alto basso livello (quindi più o meno istruzioni per eseguire un'operazione)
 - ★ **CPI:** il modo in cui un'ISA è progettata influenza il numero di cicli per eseguire ciascuna istruzione
 - ★ Un'ISA ben disegnata permette di avere frequenze più spinte

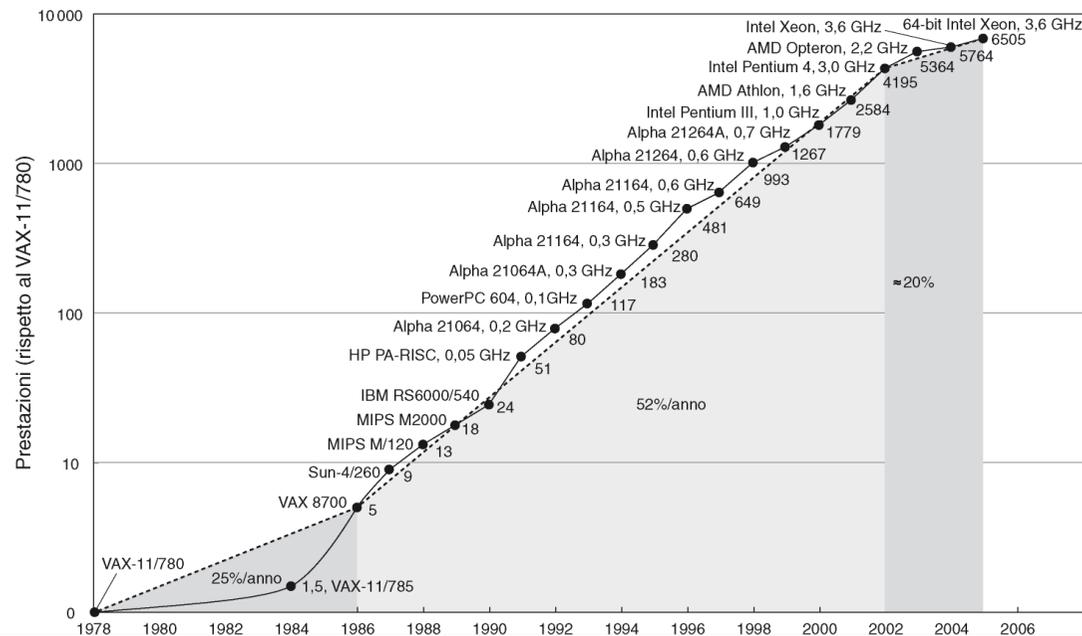


UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

La scalata delle prestazioni

- Negli scorsi anni le prestazioni dei calcolatori sono aumentati costantemente
- Di recente si e' assistito a una diminuzione dell'incremento tra una generazione e l'altra





UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

La barriera dell'energia

- Il nostro processore e' costituito di moltissimi interruttori che dissipano energia quando sono in fase di conduzione (commutazioni tra zero e uno)
- C'e' un limite alla capacita' di estrarre la potenza prodotta tramite ventole o radiatori (diciamo intorno ai 100W)
- Superato questo limite la refrigerazione diventa molto costosa e non e' attuabile in un normale desktop (per non parlare del laptop)
- La potenza e' data da:

Potenza = capacita' \times tensione² \times Frequenza di commutazione

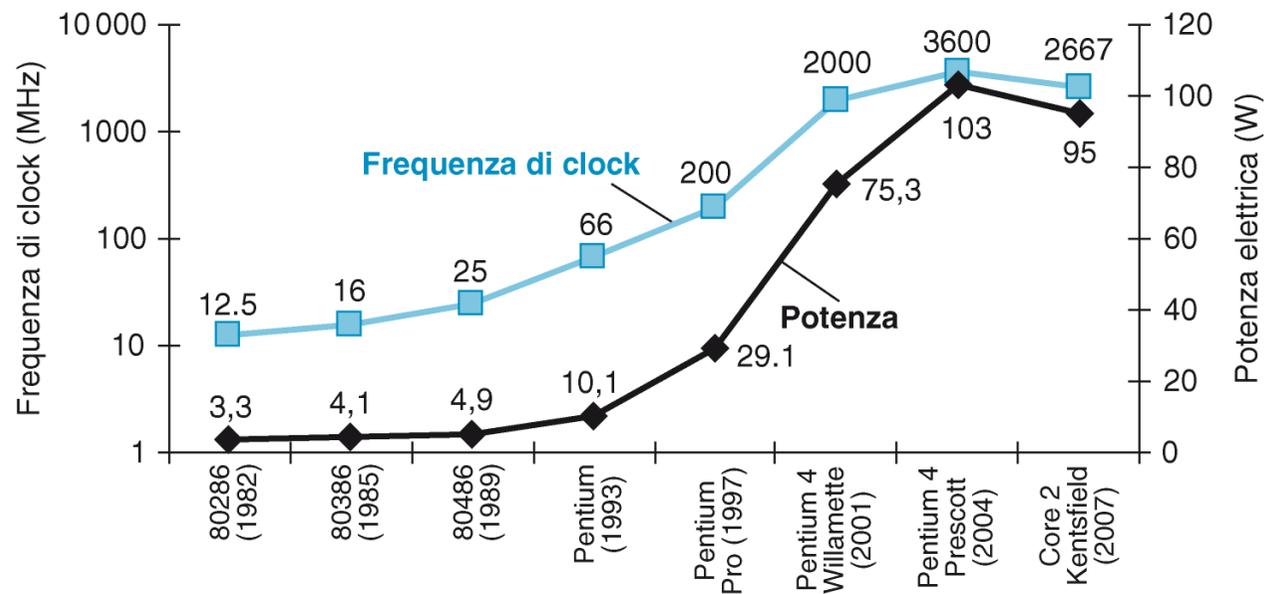


UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

La barriera dell'energia

- La frequenza di commutazione e' legata alla frequenza di clock ...
- Eppure negli scorsi anni si e' riusciti in un miracolo: aumentare in 20 anni la frequenza di 1000 volte a spese di un aumento di consumi di un fattore 30





UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Come hanno fatto?

- Il ‘trucco’ e’ stato quello di abbassare la tensione di alimentazione che agisce in maniera quadratica
- Si e’ passati in venti anni da tensioni di alimentazioni di 5V ai circa 1.2V attuali
- Questo ha permesso di incrementare la frequenza con impatti limitati sui consumi
- Sfortunatamente non ci si puo’ spingere oltre lungo questa direzione perche’ se si abbassa la tensione sotto il Volt ci sono dei fenomeni di scarica in condizioni statiche che aumentano la dissipazione anche lontani dalle fasi di commutazione



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Una nuova strada

- Nell'impossibilità di riuscire a drenare una grossa quantità di calore si è giunti a una sostanziale saturazione delle prestazioni del processore
- La nuova strada che allora viene esplorata è quella di aumentare il parallelismo (architetture multicore)
- Questo crea molte difficoltà al programmatore per i seguenti motivi:
 - ★ *correttezza*: è molto più difficile fare un programma che operi in maniera sequenziale rispetto a uno che opera in maniera "parallela"
 - ★ *Efficienza*: occorre che il carico di lavoro sulle CPU si mantenga bilanciato



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Pitfall 1

- Ci sono frequenti errori nell'analisi delle performance
- Primo errore:
 - ★ “Aumentando le prestazioni di un componente si riesce ad ottenere un aumento corrispondente delle prestazioni”
- Consideriamo ad esempio un programma che impiega 80s per operazioni di moltiplicazione e 20s in altre operazioni.
- Di quanto si deve migliorare la moltiplicazione per ridurre a 1/5 il tempo di calcolo?

Non si PUO'. Dovremmo scendere sotto i 20s ma 20s sono dovuti ad altre operazioni



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Pitfall 2

- **Errore**

- ★ “non tenere conto di tutti i termini dell'equazione delle prestazioni”

- Ad esempio spesso viene usato il MIPS (Mega instructions per second) per valutare le prestazioni

$$\text{MIPS} = \frac{\text{Numero Istruzioni}}{\text{Tempo di Esecuzione} \times 10^6}$$

- Si conclude che un calcolatore e' piu' veloce di un altro perche' esegue piu' istruzioni al secondo.



UNIVERSITÀ DEGLI STUDI DI TRENTO

Dipartimento di Ingegneria
e Scienza dell'Informazione

Pitfall 2

- Questo e' sbagliato!
- In primo luogo il MIPS non tiene conto della complessita' delle istruzioni
- ★ Puo' darsi che un'architettura esegua piu' istruzioni al secondo di un'altra ma che queste ultime siano piu' potenti
- ★ Anche nello stesso calcolatore i MIPS variano molto con il programma

$$\begin{aligned} \text{MIPS} &= \frac{\text{Numero Istruzioni}}{\text{Tempo di Esecuzione} \times 10^6} \\ &= \frac{\text{Numero Istruzioni}}{\frac{\text{Numero istruzioni} \times \text{CPI}}{\text{Frequenza di Clock}} \times 10^6} \\ &= \frac{\text{Frequenza di Clock}}{\text{CPI} \times 10^6} \end{aligned}$$

Il CPI puo' variare anche di un Ordine di grandezza per i vari benchmark. Cosi' il MIPS



Pitfall 2 (Esempio)

- Consideriamo i seguenti calcolatori che eseguono lo stesso programma

Misura	Calcolatore A	Calcolatore B
Numero Istruzioni	$10 \cdot 10^9$	$8 \cdot 10^9$
Frequenza di Clock	4Ghz	4 Ghz
CPI	1	1,1

- Quale dei due ha MIPS piu' alto?
- E quale prestazioni migliori?

MIPS: **A**

Prestazioni: **B**