

# Calcolatori Elettronici II

## **Architettura, Organizzazione e Tecnologia**

*Fabrizio Angiulli*

# Calcolatore digitale

- Macchina in grado di risolvere problemi eseguendo un *programma*
  - sequenza di *istruzioni* in un *linguaggio* utilizzato per codificare un *algoritmo*
- **Linguaggio macchina**
  - linguaggio del calcolatore; molto *primitivo*: somma due numeri, verifica se un numero vale zero, copia porzioni di memoria, ecc.
  - difficile e tedioso da utilizzare
- Nel tempo i calcolatori sono stati strutturati su più livelli di astrazione, ognuno dei quali è costruito su quello sottostante e ne nasconde i dettagli

# Calcolatore come macchina multilivello

Livello del linguaggio orientato al problema

Livello dei linguaggio assemblativo

Livello del sistema operativo

Livello di architettura dell'insieme di istruzioni

Livello di microarchitettura

Livello logico digitale

Livello dei dispositivi

# Calcolatore come macchina multilivello

- Livello dei dispositivi
  - elettronica digitale, fisica stato solido; transistori, resistenze, capacità
- Livello logico digitale (porte logiche)
  - registri, memorie, circuiti combinatori/sequenziali
- Livello di microarchitettura
  - parte operativa: data path, ALU, ...
  - parte controllo:
    - microprogrammata: interprete per le istruzioni del livello ISA
    - cablata: esecuzione immediata in hw delle istruzioni del livello ISA

# Calcolatore come macchina multilivello

- Livello di architettura dell'insieme di istruzioni (ISA – Instruction Set Architecture level)
  - istruzioni del linguaggio macchina
- Livello del sistema operativo
  - ibrido; comprende istruzioni livello ISA e istruzioni eseguibili solo a questo livello; diversa organizzazione memoria, esecuzione concorrente di programmi, ecc
- Livelli di microarchitettura, ISA e di S.O. sono scritti da *programmatori di sistema* specializzati nella realizzazione di *macchine virtuali*

# Calcolatore come macchina multilivello

- Livello linguaggio assemblativo
  - forma simbolica del linguaggio macchina; assembler, linker
- Livello del linguaggio orientato al problema
  - linguaggi ad alto livello; compilatore, interprete (es. Java byte code)
- L'insieme dei tipi di dati, delle operazioni e delle funzionalità di ciascun livello è chiamato *architettura* (aspetti visibili ad un utente ad un certo livello)

# Evoluzione delle macchine multilivello

- Primi computer (anni '40)
  - Livello ISA e logico digitale in grado di **eseguire direttamente** le istruzioni (comprensibili dall'hw)
- Maurice Wilkes, Cambridge Univ. (1951)
  - proposta di computer a 3 livelli per semplificare l'hw (costituito allora da valvole soggette a guasti)
  - dotato di interprete non modificabile (**microprogramma**) in grado di eseguire le istruzioni ISA mediante interpretazione
- Nasce così il livello intermedio di microarchitettura che si è evoluto in varie forme

# Architettura e organizzazione

- **Architettura** (di un calcolatore)
  - caratteristiche del sistema che sono visibili al programmatore (in linguaggio macchina)
    - numero di bit utilizzati per rappresentare i dati, numero di registri macchina, modalità di indirizzamento, repertorio e formato istruzioni
- **Organizzazione** (di un calcolatore)
  - relazioni strutturali tra le unità funzionali e il modo in cui essere realizzano una data architettura (non è visibile al programmatore)
    - parte operativa e parte di controllo, memorie, tecnologia impiegata, periodo di clock



# Architettura e organizzazione

- Esempi:
  - Istruzione di moltiplicazione farà parte del repertorio di istruzioni? → scelta architetturale
  - Moltiplicazione implementata mediante unità di moltiplicazione o per somme successive? → scelta organizzativa
  - Numero di registri generali? → scelta architetturale
  - Unità di controllo in logica cablata o microprogrammata? → scelta organizzativa

# Architettura e organizzazione

- Concetto di architettura introdotto da IBM con il sistema IBM S/360 (1964)
  - L'utente poteva passare da un modello meno potente (e meno costoso) ad uno più potente (e più costoso) mantenendo inalterato il sw
- Vantaggi architettura:
  - utente: salvaguarda investimento sw
  - costruttore: adottare nuove tecnologie e soluzioni organizzative fornendo prodotti più appetibili ed a diverse fasce di prezzo
- **Retrocompatibilità:** chiave successo Intel
  - x86/8086/16-bit/1978 → IA-32 x86-32/80386/32-bit/1986 → Intel64 x86-64/AMD-2003 Pentium4-2004/64-bit (diversa è IA-64/Itanium/2001)

# Architettura e organizzazione: livelli e astrazioni



ARCHITETTURA E ORGANIZZAZIONE	Livello delle applicazioni	Strutture: programmi applicativi Componenti: sistema operativo, librerie, file system
	Livello linguaggio macchina	Strutture: programmi Componenti: modello di programmazione, repertorio istruzioni
	Livello funzionale (RTL)	Strutture: unità di controllo, modello di programmazione Componenti: registri, bus, memorie
	Livello della logica	Strutture: registri, contatori, unità aritmetiche, memorie Componenti: porte, flip-flop, clock
	Livello dei circuiti	Strutture: porte logiche, flip-flop, driver Componenti: transistori, resistenze, capacità

**Figura 1.4** - Schematizzazione a livelli di un sistema di elaborazione. Vengono evidenziati i livelli che interessano l'architettura e l'organizzazione dei calcolatori.

# Calcolatore nella società

- Mondo estremamente vitale
  - USA: 10% del PIL
  - Paragone con trasporti: da Londa a New York in circa 1 secondo al costo di pochi centesimi!
- Rivoluzione dell'Informazione
  - Cellulari
  - Automobili
  - Elettrodomestici
  - Genoma Umano
  - Internet
  - ...

# Evoluzione tecnologia dei calcolatori

- Evoluzione della tecnologia dei calcolatori dovuta a due diverse spinte; progressi:
  - progettazione del livello ISA e microarchitettura
  - progettazione dei circuiti integrati
- Per comprendere le linee evolutive dei livelli superiori, occorre esaminare come si è evoluta la tecnologia costruttiva dei circuiti integrati:
  - Scala di integrazione (tecnologia X metri)
  - Frequenza (e potenza/calore)

# Pietre miliari nella storia dei calcolatori

- Generazione zero
  - Computer meccanici (1642-1945)
- Prima generazione
  - Valvole (1945-1955)
- Seconda generazione
  - Transistor (1955-1965)
- Terza generazione
  - Circuiti integrati (1965-1980)
- Quarta generazione
  - Integrazione a larghissima scala (1980-)
- Quinta generazione
  - computer invisibili (*pervasive computing*)

# Legge di Moore

- Gordon Moore co-fondatore di Intel nel 1968 (insieme a Robert Noyce e Andy Grove)
- **Legge di Moore**
  - Aprile 1965 (rivista Electronics): capacità elaborativa (numero di transistori per chip) raddoppia ogni 12 mesi per tutti gli anni '70
  - 1975: raddoppia ogni 24 mesi
  - ***Fine '80: raddoppia ogni 18 mesi***
- Si tratta unicamente di una previsione, non di una legge (di natura)

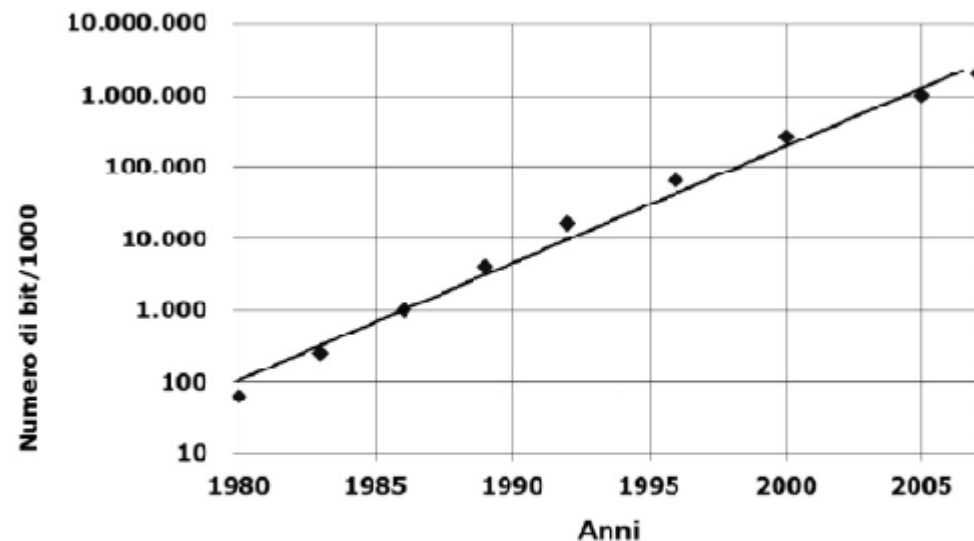
# Legge di Moore

- Prima o poi smetterà di valere, ma fonti autorevoli ritengono che la previsione varrà ancora per qualche decade
- Spiega in termini quantitativi lo spettacolare sviluppo della microelettronica
  - Crescita esponenziale: tra 18 mesi un progresso di entità pari a quello che si è avuto sino ad oggi!
  - Riduzione dei costi: pc di oggi 100 volte più potente di mainframe di 30 anni fa; costo ~1K vs ~1G dollari



# Evoluzione della capacità delle memorie RAM

Anno	Capacità (Mb)	Tempo di ciclo (ns)
1980	0,0625	250
1983	0,25	220
1986	1	190
1989	4	165
1992	16	145
1996	64	120
2000	256	100
2005	1000	10
2007	2000	2



**Figura 1.2** - Aumento della capacità per gli integrati DRAM. La capacità è misurata in Mb, il tempo di ciclo in ns. Le ultime due righe, che si riferiscono a integrati di tipo SDRAM, mostrano valori molto più bassi del tempo di ciclo.

# Tecnologia delle memorie RAM

- SRAM (Static RAM)
  - rapide, costose e meno dense; utilizzate per le memorie cache del processore
- DRAM (Dynamic RAM)
  - Lente, poco costose e più dense (maggiore capacità); principalmente utilizzate per la memoria centrale del computer
  - immagazzina ogni bit in un condensatore; se perde la carica l'informazione è perduta; la ricarica avviene periodicamente (da qui dinamica: *refresh*)

# Evoluzione della capacità delle memorie RAM

- Corrispondenza con la legge di Moore
  - ~100K (1980) a ~10G (2010)
  - fattore 16000 dal 1976 al 2007 !
- Prestazioni: tempo di ciclo (necessario per recuperare un dato dalla memoria)
  - crescita più lenta (7% annuo, ovvero raddoppio ogni 10 anni)
  - Notevole miglioramento dal 2000

# Evoluzione dei costi dei chip DRAM

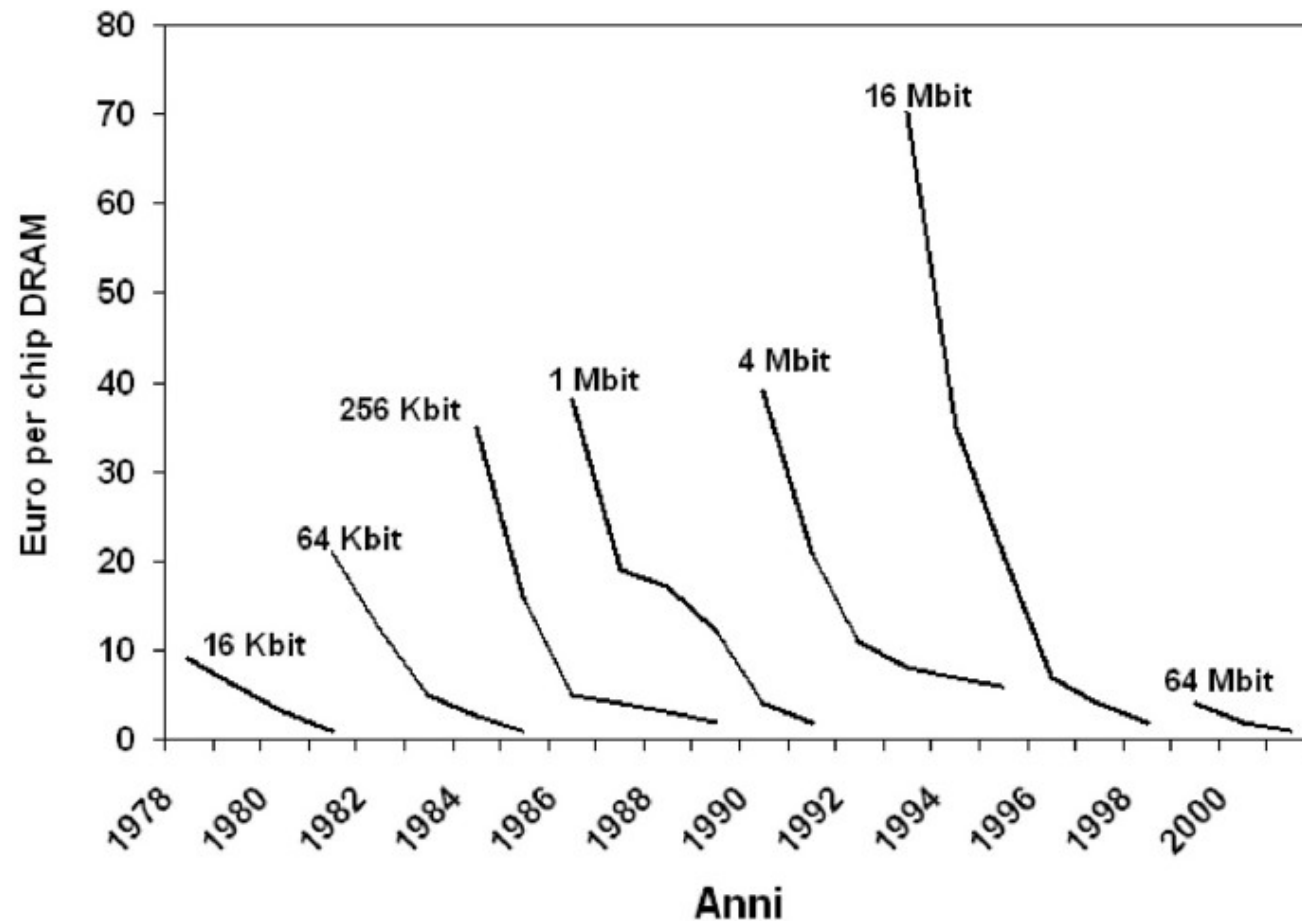


Figura 1.3 - Andamento dei costi per gli integrati DRAM.

# Evoluzione dei costi dei chip DRAM

- I costi dei chip DRAM seguono una curva di apprendimento (learning curve)
  - Costi diminuiscono col passare del tempo
- Quantità prodotta influisce sul costo
  - Ammortizzazione costi
  - Maggiore efficienza/affidabilità
  - Diminuzione del 10% di costo ogni raddoppio di quantità
- Nei primi 2 anni il costo di un chip DRAM varia di un fattore 5-10
  - 1MB: ~5K€ (1977) vs ~0,125€ (2001)

# Evoluzione del numero di transistor delle CPU Intel

Data di introduzione	Nome del chip	N. di transistor (/1000)	Tecnologia ( $\mu m$ )	Frequenza (MHz)
Novembre 1971	4004	2,3	10	0,108
Aprile 1972	8008	3,5	10	0,500
Aprile 1974	8080	4,5	6	2
Giugno 1978	8086	29	3	5
Febbraio 1982	80286	134	1,5	8
Ottobre 1985	80386	275	1,5	16
Aprile 1989	80486	1.200	1	25
Marzo 1993	Pentium	3.100	0,8	60
Novembre 1995	PentiumPro	5.500	0,6	150
Maggio 1997	Pentium II	7.500	0,35	233
Febbraio 1999	Pentium III	9.500	0,25	450
Novembre 2000	Pentium 4	42.000	0,18	1400
Marzo 2003	Pentium M	77.000	0,13	1300
Gennaio 2006	Core 2	291.000	0,065	1200
Gennaio 2008	Core 2 Quad	820.000	0,045	2500

**Tabella 1.2** - Aumento del numero di transistori delle CPU Intel. I dati riportati si riferiscono al modello di introduzione. Per i modelli introdotti in più versioni, la tabella riporta i dati relativi alla versione di più bassa capacità. Per esempio, il PentiumPro è stato introdotto in ben quattro versioni, di cui la meno potente (quella riportata) era tecnologia a  $0,6 \mu m$  e frequenza pari a 150 MHz, mentre la più avanzata era in tecnologia a  $0,35 \mu m$  e frequenza pari a 200 MHz.

# Evoluzione del numero di transistor delle CPU Intel

- Microprocessori Intel
  - Crescita del numero dei transistor al di sotto della legge di Moore (40% all'anno, ovvero 2,6 volte ogni 3 anni)
  - Crescita delle prestazioni (60% all'anno, ovvero 4 volte ogni 3 anni)
  - Aumentata frequenza e densità
  - Velocità di elaborazione dipende da diversi fattori: frequenza, velocità di trasmissione, parallelismo

# Produzione dei circuiti integrati

- *Silicio*: elemento naturale che è semiconduttore
- *Semiconduttore*: sostanza che non conduce bene l'elettricità
- Con processi chimici è possibile aggiungere materiali al silicio ottenendo minuscole aree:
  - eccellenti **conduttori** elettrici
  - eccellenti **isolanti** elettrici
  - **interruttori** comandati elettricamente (*transistor*)
- Circuito VLSI (Very Large Scale Integration)
  - milioni di elementi conduttori, isolanti e interruttori su singolo *chip* (piastrina di silicio)



# Produzione dei circuiti integrati

- **Lingotto di silicio** cristallino (cilindro silicio diametro 15-30 cm e lunghezza 30-60 cm)
- Viene tagliato in dischi sottili, detti **wafer**, dello spessore di ~2,5 mm
- **Mascheratura** deposita elementi conduttori, isolanti e interruttori su ogni wafer secondo un disegno prestabilito (20-40 passaggi)
- 1 falla nel wafer o 1 sbavatura in un passaggio della mascheratura possono introdurre **difetti** (virtualmente impossibile loro assenza)

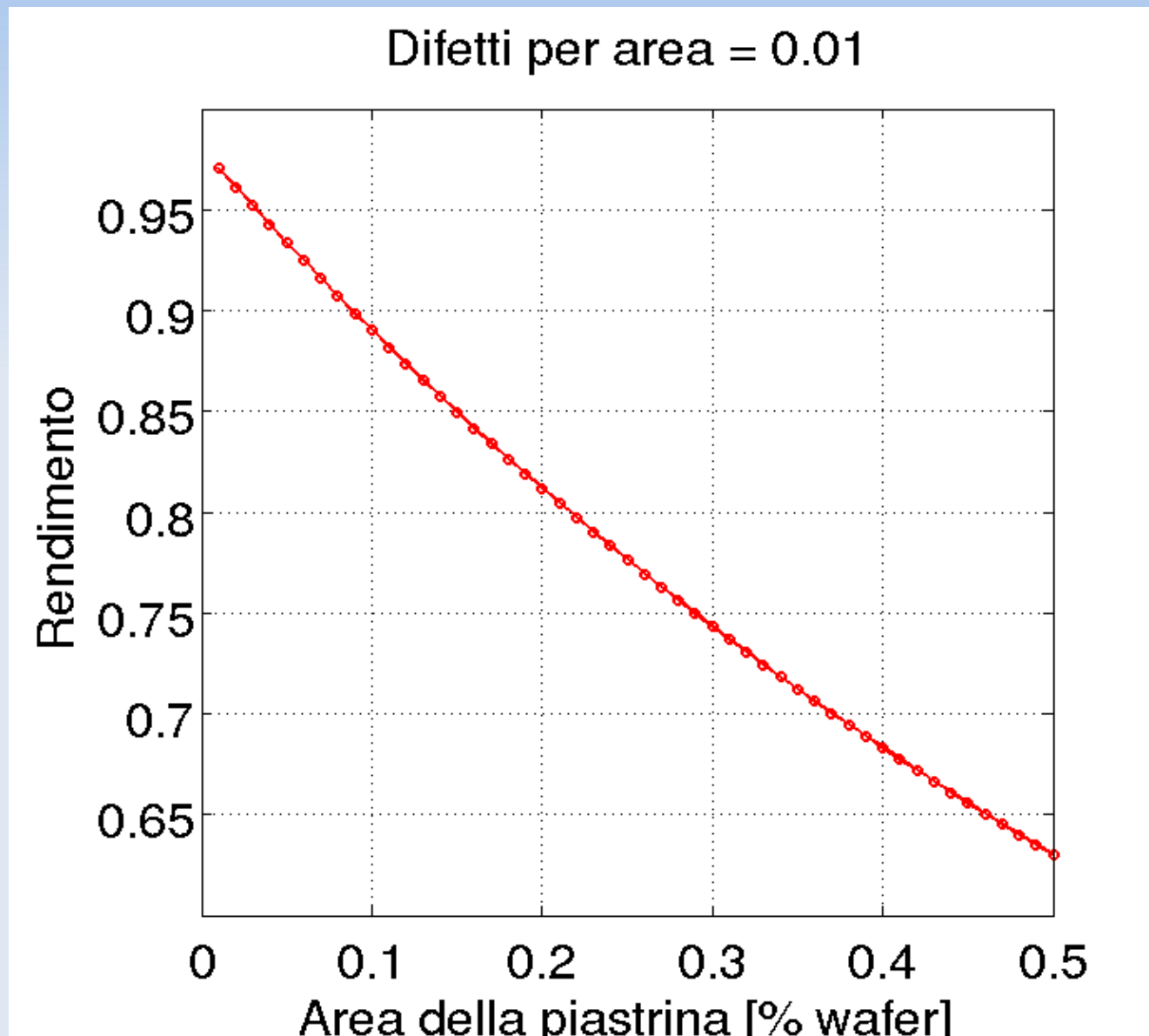
# Produzione dei circuiti integrati

- Per ovviare ai difetti, si pongono più componenti indipendenti sullo stesso wafer, detti **piastine o die o chip**
- I chip vengono **tagliati** e le piastine che contengono difetti trovate mediante un **collaudo** e scartate
- Le piastine funzionanti sono inserite nel loro contenitore e connesse ai piedini di I/O (**bonding o saldatura**)
- Infine si ha un ultimo **collaudo**
- Il **rendimento** è la percentuale di chip funzionanti rispetto ai chip totali sul wafer

# Produzione dei circuiti integrati

- Equazione del rendimento basata su osservazione empirica nelle fabbriche dei chip
  - i costi variano in maniera non lineare con la superficie della piastrina
- *Costo per piastrina = Costo per wafer / (Piastrine per wafer x Rendimento)*
- *Piastrine per wafer = Superficie del wafer / Superficie della piastrina*
- *Rendimento =  $1 / (1 + (\text{Difetti per area} \times \text{Area della piastrina} / 2))^2$*

# Produzione dei circuiti integrati



# Produzione dei circuiti integrati

- Costo circuito integrato cresce rapidamente con la dimensione della piastrina
  - rendimento più basso
  - minor numero di piastrine per wafer
- I costi della maggior scala di integrazione
  - vengono contenuti utilizzando il processo produttivo della generazione successiva che crea transistor e connessioni di dimensioni inferiori

# Evoluzione tecnologia di realizzazione dei circuiti integrati

- Il termine tecnologia X m indica approssimativamente la dimensione media di un transistor
- 2011-2012: 22 nm (22 nanometri); evoluzione del 32 nm (2010) utilizzato da Intel e AMD
  - Virus dell'HIV: ~120 nm; globulo rosso umano: ~6000-8000 nm; capello: quasi 80000 nm
- Previsioni:
  - 16 nm ~2013
  - 11 nm ~2015
  - 4 nm ~2022 ?

# Evoluzione del numero di transistor delle CPU Intel

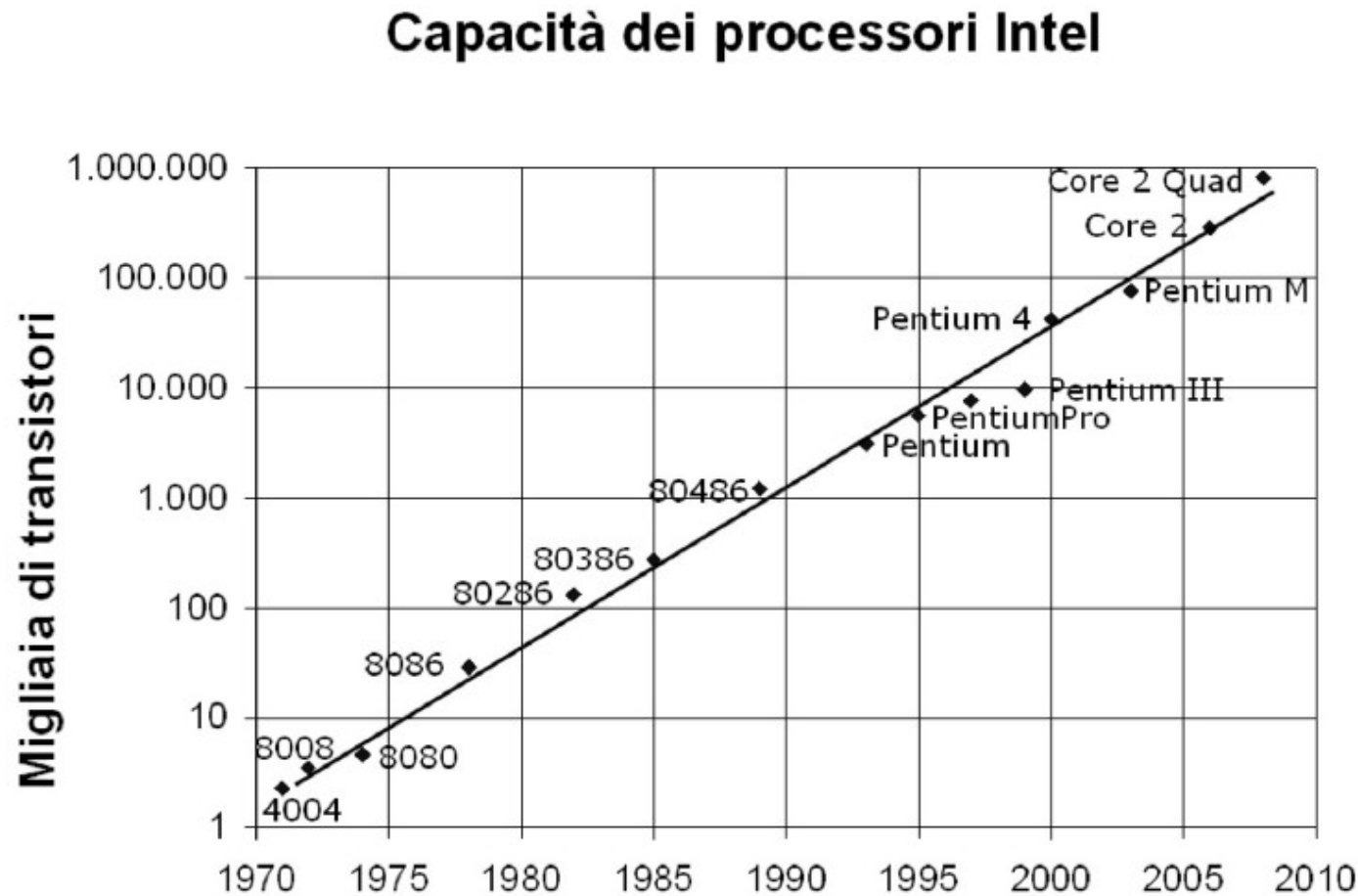


Figura 1.1 - Aumento del numero dei transistori nelle CPU Intel.

# Crescita delle capacità di elaborazione e memorizzazione

	Capacità		Velocità	
	Crescita	Tasso annuo	Crescita	Tasso annuo
Logica	2× in 3 anni	26%	4× in 3 anni	60%
DRAM	4× in 3 anni	60%	2× in 10 anni	7%
Dischi magnetici	4× in 3 anni	60%	2× in 10 anni	7%

**Tabella 1.3** - Tendenza di sviluppo delle tre principali tecnologie impiegate nei sistemi di elaborazione.

- Aumento della forbice tra prestazioni dei processori e delle memorie DRAM
  - introduzione delle memorie cache



# Prestazioni di un calcolatore

- Valutare le prestazioni è difficile
  - Fattori di difficoltà: dimensioni, complessità, ottimizzazioni presenti ed eterogeneità dei sistemi
- Cosa si intende per prestazioni?
  - Analogia con il trasporto aereo
    - Capacità: Boeing 747
    - Autonomia: Douglas DC-8
    - Velocità: Concorde

<b><i>Aereo</i></b>	<b><i>Passeggeri</i></b>	<b><i>Autonomia</i></b>	<b><i>Velocità</i></b>
		<b>(km)</b>	<b>(km/h)</b>
<i>Boeing 747</i>	470	6640	980
<i>Concorde</i>	132	6400	2160
<i>Douglas DC-8</i>	146	13950	870

# Prestazioni di un calcolatore

- **Tempo di esecuzione o risposta**
  - tempo fra l'inizio e il completamento del programma (task)
  - di interesse per l'utente
  - calcolatori desktop e embedded, minimizzare tempo di risposta

## **Throughput o bandwidth**

- numero di task completati nell'unità di tempo
  - gestore di centro di calcolo
  - Calcolatori server, massimizzare throughput
- Hanno bisogno di metriche e applicazioni diverse

# Prestazioni di un calcolatore

- **Tempo assoluto** (di esecuzione o risposta)
  - tempo totale richiesto per completare task, comprensivo del tempo richiesto dal sistema operativo per svolgere tutte le sue attività
  - percepito dall'utente
- Tuttavia calcolatori sono multitask ed il sistema operativo esegue i task in condivisione
- **Tempo di esecuzione della CPU**
  - Tempo effettivamente speso dalla CPU nel task, non comprende il tempo speso per l'esecuzione degli altri programmi

# Prestazioni di un calcolatore

- *Tempo CPU programma =*  
*Numero di Istruzioni x CPI x Periodo di Clock*
- *Numero di Istruzioni:*
  - problema, input, compilatore, ISA
- *CPI = cicli di clock per istruzione (media):*
  - programma, ISA
- *Periodo di clock (frequenza di clock):*
  - organizzazione, tecnologia

# Barriera dell'energia

<b>Anno</b>	<b>Processore</b>	<b>Frequenza (Mhz)</b>	<b>Potenza (W)</b>
1982	80286	12,5	3,3
1985	80386	16	4,1
1989	80486	25	4,9
1993	Pentium	66	10,1
1997	Pentium Pro	200	29,1
2001	Pentium 4	2000	75,3
2004	Pentium 4	3600	103
2007	Core 2	2667	95

- **Frequenza e potenza** cresciute velocemente negli anni, ma ultimamente si sono fermate
  - Sono cresciute in maniera proporzionale perchè correlate
  - Hanno smesso di aumentare perchè raggiunta massima potenza dissipabile dai sistemi di raffreddamento dei microprocessori

# Barriera dell'energia

- Tecnologia dominante circuiti integrati
  - $Potenza = Capacità \times Tensione^2 \times Frequenza$
  - Capacità dipende dal fanout e dalla tecnologia costruttiva
- Frequenza di clock aumentata di fattore 1000, mentre potenza di un fattore 30 !
  - Risultato ottenuto riducendo la tensione di alimentazione
  - -15% ad ogni generazione; in 20 anni da 5V ad 1V

# Barriera dell'energia

- Ulteriore riduzione della tensione di alimentazione
  - transistor si scarica troppo (analogia con un rubinetto che non può mai essere completamente chiuso)
- Soluzioni alternative
  - dispositivi di dissipazione del calore più grandi
  - altre tecniche di raffreddamento consentirebbero di arrivare a 300W, ma troppo costose per calcolatori desktop
  - meccanismi di controllo che spengono le parti di circuito inutilizzate in un ciclo di clock

# Crescita delle prestazioni dei processori (calcolatori)

- **Carico di lavoro (workload)**
  - insieme dei programmi eseguiti da un utente
- **Benchmark**
  - programmi *campione* che rappresentano tipici *carichi di lavoro*
  - es. calcoli aritmetica intera/floating-point, elaborazione di stringhe, compilatori, elaborazione video, giochi, compressione, ecc.
- **Benchmark SPEC**
  - System Performance Evaluation Cooperative

Periodo	Incremento annuo	Motivazioni
1978-1986	25,00%	Progressi tecnologica costruttiva
1986-2002	52,00%	Progressi organizzazione
2002-oggi	20,00%	Limitazioni potenza, parallelismo, memorie



# Metamorfosi da uniprocessore a multiprocessore

- Dal 2006 microprocessori con più processori su singolo chip
  - **core** (singolo processore)
  - **multicore** (microprocessore con più core)
    - dualcore (2 core), quadcore (4 core), ...
- Prestazioni: l'attenzione si è spostata
  - dal *Tempo di esecuzione*
  - al *Throughput*

Anno 2008	AMD Opteron x4	Intel Nehalem	IBM Power 6	Sun Ultra SPARC T2
Core per chip	4	4	2	8
Frequenza	2,5 Ghz	2,5 Ghz	4,7 Ghz	1,4 Ghz
Potenza	120 W	100 W	100 W	94 W

# Metamorfosi da uniprocessore a multiprocessore

- In passato:
  - i programmatori confidavano nelle innovazioni delle **tecnologie circuiti integrati**, **architetture**, **organizzazioni** e **compilatori** per raddoppiare le prestazioni dei loro programmi ogni 18 mesi, senza dover modificare il codice
- Oggi:
  - per sfruttare tutte le potenzialità del calcolatore ed ottenere significativi miglioramenti nel tempo di esecuzione dei loro programmi, i programmatori devono riscrivere il codice per sfruttare al meglio i diversi core a disposizione

# Parallelismo "a livello di istruzioni" vs "esplicito"

- Parallelismo a livello di istruzioni
  - calcolatore in grado di eseguire più istruzioni in maniera trasparente al programmatore
  - costringere i programmatori a parallelizzare il loro codice non è conveniente
- Parallelismo esplicito
  - compilatori hanno capacità limitate nel parallelizzare il codice
  - MIMD: multicore
  - SIMD: registri vettoriali, ISA specializzato (FLOPS)
  - industria dei calcolatori ha scommesso sul fatto che i programmatori si convertiranno alla programmazione parallela