

Il processore



muscoli e cervello dell'elaboratore elettronico

di Aldo Ascenti

Alla domanda "che computer usi?" la risposta più breve è quasi sempre il modello e la frequenza di funzionamento del microprocessore. Eppure, razionalmente, sappiamo bene che la CPU non è che una delle componenti del nostro amico elettronico, e, talvolta, nemmeno la più costosa. Il motivo di questa identificazione computer/CPU è che le tecnologie e le conoscenze che riguardano il processo produttivo del processore si sono evolute più rapidamente di qualsiasi altra area di sviluppo dell'informatica. Così se un hard disk attuale è circa 100 volte più capiente e 10 volte più veloce di uno costruito 10 anni fa, un Pentium III a 550 MHz ha prestazioni che sono 150-200 volte quelle di un 80386 a 33 MHz ed è venduto ad un prezzo decisamente inferiore a quello del suo antenato nel lontano 1989.

Tutto ebbe inizio quando, nei primi anni '70, divenne realtà la tecnologia denominata LSI (Large Scale Integration) che consentiva l'integrazione di migliaia di transistor su un'unica fettina di silicio. Questa innovazione portò alla presentazione, avvenuta nel 1971, del primo microprocessore: l'Intel 4004 che, con i suoi 2300 transistor, una frequenza di funzionamento di 108 kHz e la gestione di parole di 4 bit, raggiungeva la potenza di calcolo di circa 0,06 MIPS (milioni di istruzioni per secondo) ed è stato il capostipite di quella dinastia di processori Intel che a tutt'oggi è ben lontana dall'esaurirsi. Dopo quasi 30 anni di rapidissima evoluzione, oggi si parla di ULSI (Ultra Large Scale Integration) che permette al Pentium II di stipare nella piastrina che ospita il circuito oltre 7,5 milioni di transistor, mentre Intel promette,

per i prossimi anni, ulteriori sviluppi che si concretizzeranno in un processore da 200 milioni di transistor che potrebbe vedere la luce intorno al 2002.

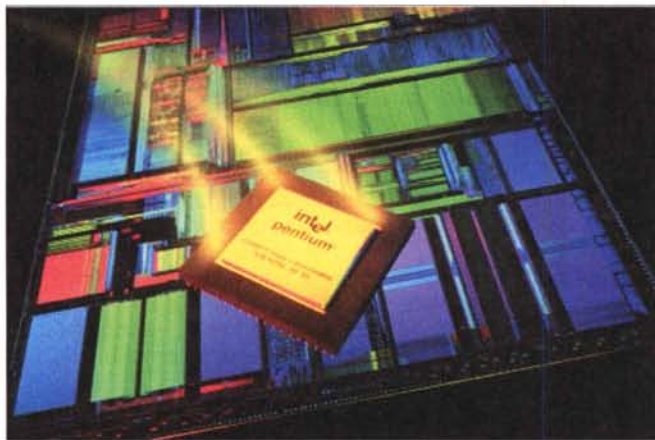
Come si realizza un processore

Malgrado l'abbattimento dei costi di produzione, legato soprattutto alla maggiore quantità di prodotti venduti, la fabbricazione di un microprocessore dell'ultima generazione resta appannaggio di poche multinazionali americane e giapponesi a causa dell'estrema complessità e delicatezza dell'intero processo produttivo.

Nella realizzazione di chip esistono due tecnologie alternative: la bipolare,

che porta alla produzione di integrati meno densi (con un minor numero di transistor) ma molto veloci, e la CMOS (Complementary Metal Oxide Semiconductor) che viene utilizzata in tutti gli attuali processori per PC e consente livelli altissimi di integrazione.

I microcircuiti digitali vengono impressi su dischi di spessore inferiore a due decimi di millimetro ricavati "affettando" un cilindro di silicio purissimo monocristallino. Il disegno della circuitazione viene fissato su diversi strati (tanto che la fetta di silicio è denominata "wafer") mediante un processo di fotolitografia



Il primo Pentium introdusse l'architettura superscalare in un processore CISC (fonte Intel).

che impiega la luce ultravioletta, filtrata attraverso apposite maschere, per fissare i tracciati e bagni di sostanze chimiche per eliminare le parti superflue. Tutto il procedimento deve avvenire in un ambiente ad atmosfera controllata, affinché sia mantenuto il massimo grado di purezza del materiale; infatti i transistor sono ricavati proprio "drogando" il silicio con infinitesime quantità di altri materiali che ne alterano le caratteristiche di conducibilità elettrica, ricavando zone adiacenti di conduttori, isolanti e semiconduttori. Una grande importanza rivestono anche le fasi di controllo dei chip prodotti, poiché solo una percentuale limitata dei processori così realizzati risulta dotata di un'affidabilità sufficientemente elevata da garantire il funzionamento alle frequenze previste. Il livello di miniaturizzazione raggiunto durante la lavorazione ed il numero di transistor impiegati sono elementi fondamentali per valutare le doti di un microprocessore. L'attuale Pentium II è realizzato con una tecnologia a 0.25 micron che pone dei limiti alla frequenza raggiungibile da questo processore, mentre per il Pentium III si parla già di un processo produttivo a 0.18

micron e sono in cantiere realizzazioni con un livello di miniaturizzazione che dovrebbe raggiungere 0.13 micron, permettendo frequenze dell'ordine dei gigahertz.

Il più semplice

La CPU si occupa di eseguire tutte le operazioni e i calcoli necessari allo svolgimento dei programmi, prelevando i dati e trasferendo i risultati mediante l'accesso alla memoria centrale; pertanto la sua forma più semplice deve prevedere

almeno tre componenti: una serie di registri, che non sono altro che celle di memoria speciali, estremamente veloci e accessibili solo al processore, un'unità di calcolo denominata ALU (Arithmetic/Logic Unit) che svolge le operazioni matematiche necessarie all'elaborazione, e un circuito di controllo, che si occupa di interpretare le istruzioni operative del programma, di reagire agli input esterni e di canalizzare

il flusso di esecuzione in modo da ottimizzare le prestazioni. In pratica il programma è una sequenza di istruzioni, semplici o complesse, e di operandi; ogni istruzione che raggiunge il microprocessore (fase detta fetch) viene memorizzata in un registro apposito, quindi passa attraverso l'unità di controllo dove subisce il processo di interpretazione, che consiste nell'individuare la natura dell'istruzione confrontandola con un elenco predefinito che costituisce il ba-

gaglio di operazioni direttamente eseguibili dalla CPU. L'unità di controllo si occupa poi di richiamare gli eventuali operandi, conservandoli in altrettanti registri, e di ordinare all'ALU di eseguire su di essi i calcoli necessari. Ciascuno di questi compiti viene portato a termine col



Una maschera di metallizzazione usata nella fabbricazione del primo microprocessore mai realizzato: l'Intel 4004 (fonte Intel).

ritmo imposto da un oscillatore, realizzato con un cristallo di quarzo analogo a quello presente nei comuni orologi da polso, che genera segnali alternati, periodici, di alto/basso voltaggio. Ogni sequenza di alto/basso voltaggio è denominata ciclo di clock ed il susseguirsi dei cicli scandisce, come un metronomo, l'esecuzione delle operazioni. Le CPU attuali non condividono la stessa frequenza di clock del sistema che le ospita, ma funzionano ad una velocità più elevata, ottenuta moltiplicando la frequenza di base per un fattore costante. Questo fatto rende ogni comunicazione con l'esterno un evento che ha l'effetto di rallentare significativamente l'attività

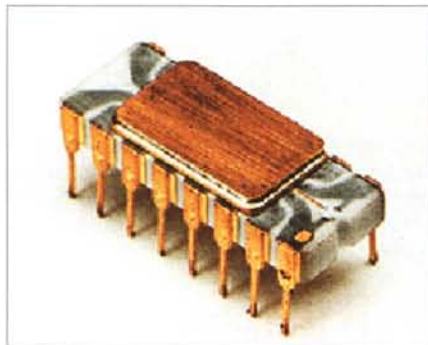


Due tecnici alle prese con lo schema tecnico di un processore (fonte Intel).

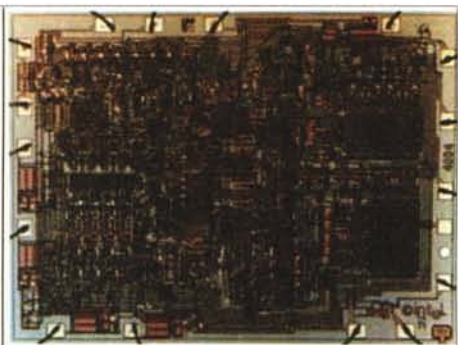
del processore. Siccome il ricorso al bus di sistema da parte del processore avviene quasi esclusivamente per operazioni di lettura e scrittura sulla memoria centrale, si è cercato di ridurre il numero di queste operazioni integrando nella CPU un certo quantitativo di memoria ad alta velocità detta cache di primo livello (L1). Questa memoria fa da tramite tra la RAM di sistema e il processore, che in molti casi troverà ciò che gli serve per completare i suoi compiti senza dover cercare altrove. La cache di secondo livello è un ulteriore stadio interposto tra quella di primo livello e la memoria centrale. Ha dimensioni maggiori di quella L1 e può essere integrata nel chip e sfruttare la frequenza della CPU, come avviene nel Celeron Mendocino e nel K6 III, o, più semplicemente, operare sul bus di sistema, come nei PC basati sull'AMD K6 II, avvantaggiandosi rispetto alla comune memoria solo grazie all'impiego di componenti più performanti e di una più efficiente logica di gestione. Il Pentium II si avvale di una soluzione intermedia per la sua cache L2, che si trova su integrati separati da quello della CPU e lavora ad una frequenza che è la metà di quella del processore.

L'elevato grado di integrazione e la conseguente straordinaria velocità di elaborazione stanno spostando verso la CPU un quantitativo sempre maggiore di compiti nell'ambito di un sistema di elaborazione. Il primo dei componenti, una volta esterni, ad essere realizzato internamente al processore è stato il coprocessore matematico o FPU (Floating Point Unit). Questo circuito deve la sua denominazione inglese, che si traduce in "unità a virgola mobile", all'abilità nell'esecuzione di calcoli su numeri reali molto grandi, rappresentati con la notazione a mantissa ed esponente. Si tratta quindi di una naturale estensione dell'ALU, che è progettata per gestire principalmente numeri interi e rallenta sensibilmente quando deve occuparsi di operandi di tipo diverso; l'FPU è parte integrante di tutte le CPU dal 80486DX in poi.

I moderni processori custodiscono al loro interno quantità sempre maggiori di memoria cache e si parla di integrare circuiti per la compressione e decompressione dei dati, per la cifratura dei codici, per funzionalità anti-



Il 4004, presentato all'inizio degli anni '70, integrava 2300 transistor e raggiungeva una potenza di calcolo di circa 0,06 Mips (fonte Intel).



virus e molto altro ancora.

Il superamento delle architetture tradizionali

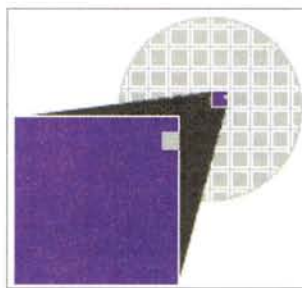
A definire l'architettura di un processore concorrono principalmente due parametri: la larghezza di banda, intesa come la dimensione in bit dei dati e delle istruzioni che la CPU è in grado di gestire, e la natura delle istruzioni, che possono essere poche e semplici o numerose e complesse. Da quest'ultima distinzione scaturisce la nota divisione in categorie dei processori, dalla quale ha avuto origine un interminabile dibattito su quale dei due sistemi meritasse di prevalere.

CISC è l'acronimo di Complex Instruction Set Computer, ovvero di computer con un insieme complesso di istruzioni. I processori CISC sono in grado di elaborare direttamente centinaia di funzioni matematiche come il calcolo di una radice quadrata o la valutazione del coseno di un angolo, laddove una CPU RISC (Reduced Instruction Set Computer) è costretta a smembrare un'istruzione complessa in numerose operazioni più

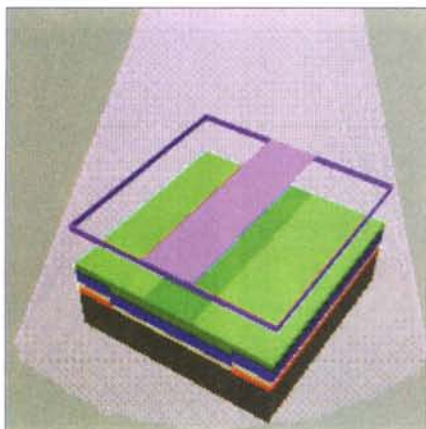
semplici. Nelle prime implementazioni, i processori RISC non potevano effettuare nemmeno moltiplicazioni e divisioni, che dovevano essere ricavate da una sequenza di somme, complementazioni e operazioni logiche. Per contro i processori CISC erano più complessi da realizzare e quindi molto più lenti nell'esecuzione di operazioni semplici, che sono di gran lunga più frequenti di quelle complicate. Altro grosso limite dei processori CISC era dovuto al fatto che la frequenza operativa di un processore deve essere tarata sulla più lenta delle operazioni che questo è in grado di svolgere, rendendo di fatto inutili le prestazioni dei componenti più veloci, assoggettati ad operare con gli stessi tempi di quelli più lenti.

Oggi la distinzione CISC-RISC ha perso molto del suo significato, poiché i processori moderni, qualunque sia stata la filosofia progettuale originaria, raccolgono il meglio delle due tecniche, superandone i limiti in molti modi. In particolare l'architettura RISC pone, ormai, l'accento non già sul numero di istruzioni disponibili, bensì sulla riduzione della complessità della circuitalizzazione. Così, ecco che il set di istruzioni dei processori PowerPC consiste in oltre 220 diverse operazioni, ma mantiene un'invidiabile semplicità di progetto grazie all'accorgimento di rendere le istruzioni il più

omogenee possibile per lunghezza e durata, in modo che possano essere portate a termine in un solo ciclo di clock. D'altra parte, l'architettura del Pentium II, che trae origine da una filosofia CISC, sfrutta tecniche di parallelismo per suddividere un'operazione complessa in una serie di operazioni più semplici, da eseguire contemporaneamente su istruzioni diverse. La tendenza alla semplificazione dell'HW come metodo efficace per aumentare le prestazioni, si manifesta con la ripresa degli studi sulla tecnologia VLIW (Very Long Instruction Word), in auge alla fine degli anni '80 e poi abbandonata per la difficoltà di ricompilare il software. Le architetture basate su



Su ogni "fetta" di silicio purissimo vengono realizzati contemporaneamente più chip, di cui molti non risponderanno a severi criteri di affidabilità e verranno scartati alla fine del processo (fonte Intel).



Il processo produttivo prevede tecniche di fotolitografia, che sfruttano i raggi ultravioletti, filtrati da maschere semiriflettenti, per "disegnare" i componenti sui vari strati del wafer di silicio (fonte Intel).

VLIW prevedono l'assemblaggio, durante la fase di compilazione, di più comandi in un'unica macro istruzione (dell'ordine di 80 byte) che viene digerita in un colpo solo dal processore, consentendo di ridurre di un terzo i circuiti necessari alla CPU.

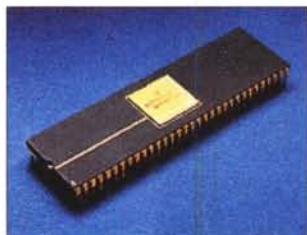


Il diverso aspetto esteriore dei processori non è sempre indice di differenze architetturali, ma più spesso riflette l'impiego a cui la CPU è destinata e le esigenze di montaggio della macchina sulla quale verrà installata (fonte Intel).

La ricerca delle prestazioni

La contesa per ottenere una sempre maggiore potenza di elaborazione ha portato le CPU a svilupparsi su più fronti. Prima di tutto crescono le velocità di clock, grazie al miglioramento dei processi produttivi operato nel grado di miniaturizzazione (riduzione delle distanze tra i componenti) e impiegando materiali migliori, come il rame in luogo dell'alluminio nella definizione dei percorsi elettrici all'interno del chip. Il secondo aspetto sul quale i progettisti lavorano assiduamente è l'aumento del grado di parallelismo dell'architettura del processore, vale a dire che si cerca di far lavorare contemporaneamente, ad ogni ciclo di clock, parti differenti della CPU su più operazioni necessarie a diverse istruzioni. Il parallelismo nella gestione dei dati si concretizza in un approccio denominato SIMD (Single Instruction Multiple Data) che definisce una peculiarità di molti processori, come l'attuale generazione di CPU Intel, di effettuare operazioni su numerosi operandi contemporaneamente, grazie a unità di calcolo progettate appositamente. Il pipelining è invece una tecnica per accrescere il parallelismo nella gestione delle diverse fasi in cui è scomponibile l'esecuzione di ogni istruzione: un processore dotato di una pipeline a più stadi può, ad esempio, operare la decodifica di un'istruzione e contemporaneamente eseguire i calcoli necessari ad un'altra e memorizzare i risultati relativi a una terza. I processori attuali sono dotati di più pipeline e sono pertanto detti super-scalari.

Un altro sistema per rendere un processore più efficiente è quello di dotarlo di nuove istruzioni, attentamente selezionate per rispondere alle mutate esigenze del software più recente, così da rispondere direttamente alle richieste degli utilizzatori analizzate sta-



Le capacità del 68000, prodotto da Motorola alla fine degli anni '70, fecero la fortuna dei primi Apple Macintosh e del Commodore Amiga.

tisticamente. A questa categoria appartengono le MMX di Intel, che sono 57 istruzioni dedicate alla grafica, alla riproduzione di filmati, alla sintesi vocale e alla trasmissione dati. Come è noto anche AMD ha sviluppato un set di istruzioni denominato 3DNow! con analoghe finalità. Proprio ad un nuovo gruppo di istruzioni deve la sua notorietà il Pentium III, arricchito dai suoi progettisti delle SSE (Streaming SIMD Extensions) e di una nuova modalità operativa che gli consente di sfruttare contemporaneamente istruzioni MMX, SSE e la FPU, avvalendosi, inoltre, di 8 nuovi registri a 128 bit. Grazie alle 70 SSE il Pentium III si candida al ruolo di unico componente intelligente di un sistema di elaborazione, accentrando su di sé molti compiti svolti dai chip della scheda grafica, di quella audio, dei modem e degli adattatori di rete.

La velocità: una misura difficile

Cosa vuol dire che un processore è migliore di un altro? Lo ammetto, la domanda è tendenziosa, ma i fattori che decretano il successo di una CPU sono talmente tanti da non consentire una risposta immediata e definitiva. I tempi di esecuzione dei programmi, da parte delle diverse architetture dei processori, possono variare bruscamente a cau-

sa delle condizioni di impiego, delle abitudini degli utenti e della qualità della programmazione e della compilazione del software. Anche i benchmark, per quanto onestamente studiati per offrire valutazioni attendibili, condizione, questa, che spesso non si è verificata, soffrono dei limiti di qualsiasi analisi statistica, basandosi forzatamente sull'utente medio, che svolge un'attività media con le sole applicazioni più diffuse.

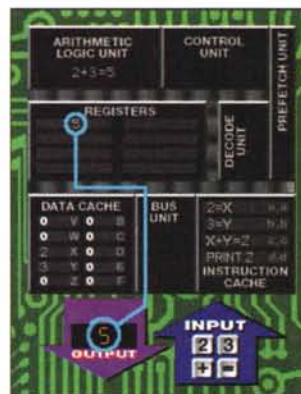
I MIPS (milioni di istruzioni per secondo) sono originariamente definiti come rapporto tra il numero di istruzioni che compongono un programma e il tempo di esecuzione.

Si tratta di numeri direttamente proporzionali alla frequenza di elaborazione e presentano tre grosse limitazioni: per cominciare non tengono conto dell'istruzione set dei diversi processori e perdono progressivamente di validità con l'aumentare delle differenze architetturali, rendendoli, di fatto, inadatti a confrontare CPU che funzionano in modo sostanzialmente diverso; inoltre dipendono fortemente dai programmi utilizzati nella valutazione e non forniscono valori univoci nemmeno nell'ambito di uno stesso processore; infine, l'uso con differenti compilatori può condurre a valori contraddittori e fuorvianti, che possono avere un andamento contrario alle prestazioni effettive.

Particolarmente erratico è il valore in MIPS di picco, che fornisce un'indicazione arbitraria delle migliori prestazioni di una CPU, ottenute in particolari condizioni di funzionamento, che potrebbero verificarsi così raramente da non influire affatto sul comune utilizzo.

Alternativi ai MIPS sono i megaFLOPS (million floating-point operations per second) che valutano il numero di operazioni in virgola mobile eseguite da un processore ogni secondo di esecuzione di un dato programma, dal quale dipendono fortemente. Valutando operazioni e non istruzioni, i MFLOPS sono più adatti dei MIPS a confrontare macchine diverse, ma sono comunque limitati da come le diverse architetture gestiscono i dati in virgola mobile.

Insomma, un modo preciso per far gareggiare i processori non è ancora stato inventato e spetta al mercato eleggere, di volta in volta, un vincitore, che spesso è quello sul quale è stata condotta la migliore campagna pubblicitaria.



Uno schema semplificato dei componenti di un moderno processore che intervengono nell'esecuzione di una semplice operazione algebrica (fonte Intel).