

Pentium II: storia e geografia

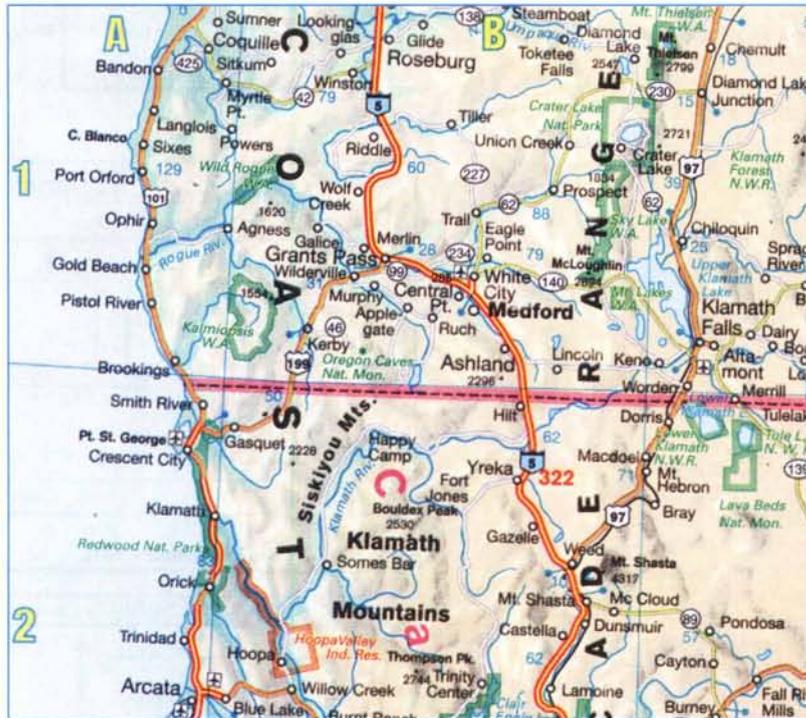
di Leo Sorge

Mettiamoci d'accordo sui nomi. In codice, il Pentium originale attualmente prodotto si chiama P54C, e lo chiameremo Pentium. Chiameremo invece P55C il Pentium MMX. Il Pentium Pro lo indicheremo come PPro mentre P2, senza riferimenti storico-sociali alla loggia Propaganda 2, sarà il modo con cui indicheremo il Pentium Pro MMX, in codice Klamath, o ora giunto sul mercato col nome di Pentium II. Usare sempre lo stesso nome (a parte che nei titoli) renderà il pezzo più noioso ma più chiaro.

Un po' di storia

P2 è quindi un figlio di PPro, non di Pentium come il nome lascerebbe intendere. In effetti Pentium e PPro sono due macchine completamente diverse. Il primo è un progetto nato con l'intenzione di sfruttare vantaggi sia architetturali che tecnologici rispetto al 486, ma finito per avvantaggiarsi solo dalle maggiori velocità di clock. Le scelte fatte in sede di progetto del Pentium, infatti, mettevano nel compilatore la complessità necessaria per gestire le due unità di esecuzione interne, grande novità per l'epoca (negli x86: i RISC già li avevano). Ma la tecnologia dei compilatori non era ancora sufficientemente avanzata, per cui questo grande svantaggio sfumò. Restava il fatto che Pentium poteva andare più veloce degli altri chip, e questo ne ha portato in su le prestazioni. Come vedremo è una situazione che si ripresenta anche in P2 su PPro.

Nel frattempo gli avversari di Intel, molto meno burocratici in fase di progetto, capirono che a queste velocità la gestione di più unità andava fatta diret-



tamente in hardware anche se con politiche più semplici e meno ottimizzate di quelle possibili in software. E' per questo che sia Cyrix che NexGen tirarono fuori dei prodotti competitivi con il Pentium già in una prima fase, ed intermedi tra Pentium e PPro nella seconda: il progetto del Cyrix 6x86, oggi chiamato processore di sesta generazione per confrontarlo con il PPro, era nato per surclassare il Pentium; analogamente ciò che oggi chiamiamo AMD K6 è in realtà una versione leggermente modificata del NexGen Nx686. Delle origini di questi aspetti parliamo a lungo già su MC 146 (dicembre 1994), da pagina 160, in un pezzo intitolato "x86 vs' cercando".

Già che ci siamo, mettiamoci d'accordo anche sulle generazioni di processori. Tutto ruota intorno ad Intel. L'8086/8088 del PC prima maniera è detta prima generazione, l'80286 del PC-AT seconda generazione, l'80386 poi semplificato in 386 è la terza generazione, il 486 è una quarta (generazione, non misura), anche se il Cyrix 386/486 è già un prodotto intermedio e

a ben vedere i compatibili AMD perdono anche il 20% rispetto agli analoghi Intel. Poi iniziano i problemi, perché il nascere di prodotti con prestazioni intermedie ha complicato la cosa. La proposta "salva-capra-ecavoli" è venuta da NexGen, che propone un indice di prestazioni relative al Pentium a clock fisso. Per loro era necessario far così, avendo per primi un'architettura assolutamente innovativa: una macchina RISC che traduce in hardware le istruzioni x86. Non era proprio possibile fare altri confronti. Da allora Pentium e PPro sono diventati il metro di valutazione relativa.

Dal Pentium al Pentium Pro

Il Pentium era sostanzialmente un doppio 486 con cache di 8K+8K in tecnologia BiCMOS, ovvero i transistor erano di due tipi, bipolare e C-MOS. Si tratta di un processo portato avanti per un certo tempo e poi mantenuto solo da Intel e Texas: rispetto al CMOS puro, che è lo standard per le memorie (quindi il più economico), ha bisogno di qualche passo in più ma garantiva alcuni risultati migliori, almeno finché il transistor è stato più largo di 0,50 micron. Tutto ciò che riguardava memoria, quindi dai registri alla cache, era essenzialmente identico al 486.

La tecnologia del silicio andava avanti con incrementi spaventosi. Per avere incrementi di prestazioni che sfruttassero tale crescita, bisognava rendere regolare l'esecuzione delle istruzioni. La storia dei chip x86 mostrava che la compatibilità con il passato portava con sé la presenza di dati ed istruzioni di

lunghezza diversa, così come l'esecuzione delle istruzioni poteva andare da pochi cicli di clock ad alcune centinaia.

Quella di arrivo nel 486 era assolutamente irregolare, ed andava rifondata in toto. La scelta di Intel è stata quella più pura e più avanzata da un punto di vista filosofico: una macchina RISC, che esegue istruzioni non x86 in modo regolare (quindi incrementabili linearmente con la potenza), ed un traduttore *on-chip* che rende digeribili le istruzioni x86. All'aumentare della velocità di clock, poi, si sarebbe fatta corrispondere una modifica nella struttura della cache.

Un confronto tra Pentium desktop, Pentium Mobile e Pentium Pro è in "Pentium: come, dove, perché", su MC 167, da pagina 232. Si arriva quindi al PPro, la cui storia prosegue nelle migliori del P55C per giungere al P2.

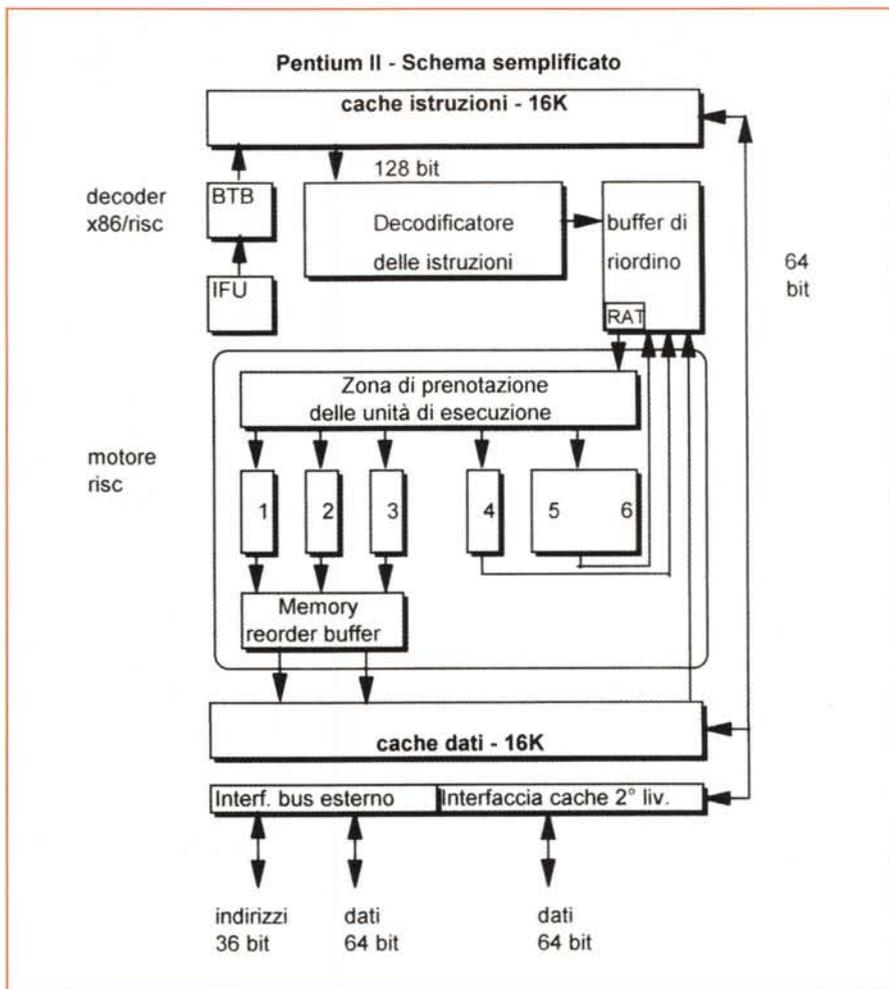
Cos'è il Pentium II

Le due grandi innovazioni di casa Intel, l'MMX e il CMOS puro, confluiscono nel PPro che diventa P2. Oltre a queste viene risolta una serie di problemi vuoi tecnici, vuoi commerciali. Procediamo per gradi, benché per sommi capi.

A partire dal PPro, alla Intel i chip vengono progettati da *équipe* differenti. Da Santa Clara, sede tradizionale dei progetti, si è passati ad Hillsboro (in Oregon), dai progettisti dell'i960, un controller RISC. Il P55C è stato sviluppato in Israele.

Il PPro era nato in BiCMOS per eseguire codice a 32 bit, ma aveva ereditato la cache del 486: 8K dati, 8K istruzioni. Per non toccarla si era ricorsi ad un artificio, una cache di secondo livello da 256K che va alla velocità del chip, tra 150 e 200 MHz. A queste velocità un oggetto del genere dovrebbe essere compreso nello stesso chip del microprocessore, ma tecnologicamente non era possibile per due motivi: la cache è di 15 milioni di transistor, il PPro di altri 5,5 milioni. Va precisato che le memorie sono più dense dei microprocessori, tanto che a parità di processo il PPro occupava 306 mm² (densità 5,5M/306 = 18.000 transistor/mm²) mentre la cache solo 202 mm² (15M/202=74.000 transistor/mm²). Per di più li faceva Intel in BiCMOS, piuttosto costoso per le memorie.

Quindi la cache di secondo livello andava alloggiata affiancata al microprocessore. Erano già disponibili delle soluzioni di questo tipo, i *multichip module* o MCM. Roba costosissima e difficile



da alloggiare: saliva il costo del PPro, saliva il costo delle piastre madri. Roba da ricchi.

Erano evidenti due sviluppi futuri: A) questa cache di secondo livello sarebbe diventata un *optional* per casi particolari; B) il MCM sarebbe stato abbandonato.

Irrompe il Pentium MMX

Passiamo alla storia recente. Il P55C ha introdotto nel panorama Intel ben quattro innovazioni: due architetturali, una tecnologica, una sistemistica.

1) L'architettura è stata estesa con l'MMX (che pare stia per *Matrix Multiplying eXtension*, con riferimento alla singola nuova istruzione più importante, e non per *MultiMedia eXtension*), del

Ecco per sommi capi la struttura del P2, ovviamente molto simile a quella del PPro. Le parti rinnovate sono le due cache con l'interfaccia per la L2C, l'MMX e la tavola di alias dei registri.

Legenda:
BTB Branch Target Buffer
IFU Instruction Fetch Unit
RAT Register Alias Table

Unità di esecuzione:
1 Store data
2 Store address
3 Load address
4 ALU interi ed MMX
5 FPU
6 Interi ed MMX

quale trovate trattazione su MC 171, pagine 192 e seguenti, in *Intel Pentium MMX*.

2) La nuova cache di primo livello, che ora è di 16K dati + 16K istruzioni. Per la cronaca P55C è l'ultimo processore di quinta o successiva generazione a raggiungere la dimensione complessiva di 32K: il Cyrix MGx ha 16K, l'M2 avrà 64K, per AMD vedete l'articolo di

Ecco nei dati Intel un confronto tra i vari chip. Solo gli ultimi due danno rilevanza al multimedia accelerato da MMX. Si noti che il PPro ha solo 256K di L2C contro i 512 del P2.

Nelle prove di P55C effettuate da MCmicrocomputer (n° 171 pag. 198 e segg.) le prestazioni (elenate a pag. 212) si mostrano conformi a quanto dichiarato da Intel. Si noti che la scala dell'articolo citato è relativa al Pentium a 200 MHz senza MMX, quindi circa 1,6x156 (che è il valore del Media Benchmark assegnato al P200) da 249,6, un valore molto prossimo a quello dichiarato.

Pentium: prestazioni a confronto

		Intel IComp 2	Unix - Spec Int95	Unix - Spec fp_base95	Ziff-Davis CPUmark32	Norton Multimedia	Intel Media BM
P2/266	L2C: 512K	295	10,8	6,4	693	19,4	351
P2/233	L2C: 512K (stima)	260	9,5	5,9	606	17,2	310
PPro/200	L2C: 256K	220	8,2	5,5	540	11,6	194
P55C/200		182	6,4	3,9	427	13,8	255
Pentium/200		142	5,1	3,1	389	9,6	156

Luca Angelelli pubblicato su questo stesso numero fra poche pagine (ma si parte dai 24K del K5).

3) Il processo è ora un CMOS puro. Tra l'altro Intel misura la dimensione lineare dei suoi processi in modo diverso da quello degli altri, in quanto lei misura l'areola metallica del contatto, mentre in genere si pensa al transistor: quello che lei chiama 0,35 micron è in realtà uno 0,28 degli altri.

4) La CPU (ovvero microprocessore, cache e chipset) sono alloggiati in un modulo tipo SEC (Single Edge Contacts). In futuro l'upgrade dei sistemi, per sostituzione o aggiunta, avverrà in maniera molto semplice, lasciando intatto tutto il resto.

Finalmente il Klamath!

Lasciamo la storia e passiamo alla geografia. Infatti il Klamath è un fiume che nasce in Oregon (dov'è stato progettato il PPro) ma sfocia nella California del nord, dando nome ad una regione e ad una città. Analoga provenienza, ovvero fiumi californiani, hanno i nomi Deschutes (nome in codice del successore del PPro) e anche Merced, che dovrebbe essere o il P7 oppure direttamente il chip a 64 bit di Intel e HP.

P2 è un PPro ottimizzato con una nuova architettura di cache e l'MMX. Mette in 203 mm² i suoi 7,5 milioni di transistor in tecnologia 0,28 micron (che Intel misura come 0,35) con alimentazione a 2,8V. La configurazione è all'interno del cosiddetto Slot 1, che permette di avere piastre madre non troppo sofisticate e manutentori non necessariamente espertissimi, come già accade nel resto dell'elettronica di consumo, ad esempio i televisori. La conseguenza di mercato più immediata è che per avere P2 bisogna acquistare una nuova piastra madre. Inoltre tale

scelta tiene la cache di secondo livello abbastanza vicina al microprocessore, in una architettura detta *closely coupled*. Il chipset 440 FX, e anche la velocità del bus esterno (66 MHz), restano invariati.

Da quanto detto è chiaro che le ipotesi A) e B) si sono verificate, e anche che i punti da 1) a 4) sono stati ereditati dal P2. Ma non basta, perché le modifiche sono ancora tante. Vediamole per sommi capi.

Esecuzione del codice a 16 bit

A che serve questa cosa, direte voi? Ai nostalgici di Windows 3? No, certo. Gli è che buona parte del codice di Windows 95 è ancora a 16 bit, così come buona parte delle applicazioni. Gli esperti mostrarono subito che una scelta progettuale di accesso al *reorder buffer*, la parte che nutre il motore RISC, penalizzava l'esecuzione di questo tipo di software. Poiché P2 diventerà tra breve la scelta preferenziale su notebook di fascia alta, desktop e server di fascia bassa, il problema andava risolto. Tecnicamente parlando c'è stata una modifica alla RAT, *Register Alias Table*, che adesso supporta il *register renaming* in modo corretto e senza interferire con l'MMX.

Ripariamo della cache

Raddoppiando la cache si ha un vantaggio che è funzione della dimensione finale e della velocità di clock; inoltre tale vantaggio è diverso per i dati e le istruzioni. Si può dire che più veloce è il clock, meglio viene sfruttata una cache più grande. Infatti ogni volta che un dato non è sul chip va caricato dalla me-

moria, il che rallenta l'esecuzione di un numero di colpi di clock che dipende proprio dalla velocità di clock: semplificando, se il chip internamente va a 100 MHz batte un colpo ogni 10 ns, quindi una RAM da 60 ns lo ferma per 6 colpi; se il chip va a 200 MHz lo ferma per 12 colpi, penalizzandolo maggiormente.

Con una cache di primo livello da 32K, il P2 può scegliere quella di secondo livello. Infatti ne supporta tre tipi: a 1/3 del clock del microprocessore; alla metà; a velocità piena. La prima modalità serve oggi per progetti meno costosi, ma un domani, al crescere della velocità interna, allungherà la vita del P2. La seconda è la scelta standard, mentre l'ultima riguarda progetti ad elevate prestazioni.

Mettendo *off-module* la L2C (Level-2 Cache, ovvero cache di secondo livello) si ha un degrado delle prestazioni. La maggiore velocità di clock, che parte da 233 e 266 MHz, compensa egregiamente le piccole perdite di potenza imposte in condizioni normali dalla nuova architettura di cache.

Chi ci rimette?

Se con il PPro potevano lavorare insieme fino a quattro chip senza elettronica esterna di coordinamento, con P2 il numero è stato ridotto a due soli. Questo fatto limita l'immediata applicazione dell'oggetto ai server di fascia estremamente alta, che peraltro ben poco vantaggio avrebbero dall'altra innovazione, l'MMX.

La più lenta cache di secondo livello è un problema per le workstation: le applicazioni tecnico-scientifiche accedono con gran frequenza alla L2C, per cui lo svantaggio è notevole. Ovviamente in attesa che le applicazioni medesime vengano ricomilate per MMX.

MC