

4 Microarhitectura procesoarelor

Arhitectura de bază a microprocesorului Pentium (fig. 4.1) include următoarele structuri componente:

a) Două unități de execuție pentru operații cu numere întregi (U și V) asimilate unor benzi de asamblare.

Pentru cele două unități de execuție se decodifică simultan două instrucțiuni, iar execuția lor se realizează tot simultan (cu condiția că rezultatul celei de-a doua instrucțiuni să nu depindă de rezultatul primei instrucțiuni). Aceasta conduce la o creștere a performanței cu circa 30%, Intel numind-o tehnologie superscalara, baza noii microarhitecturi (microarhitectură este metoda de realizare a arhitecturii setului de instrucțiuni (ASI) prin hard-ul microprocesorului).

Pentium folosește două memorii cache (de câte 8 kB), pentru coduri și date, care sunt folosite pentru a stoca cele mai frecvent folosite coduri și date. În acest fel se elimină, de câte ori este posibil, accesul la memoria microcalculatorului, acces care necesită un timp mai mare.

b) Buffer de decodificare anticipată a instrucțiunilor.

Codul din memoria cache este testat pentru a sesiza din timp eventualele instrucțiuni de salt anterior încărcării acestora în pipeline; decodificarea instrucțiunilor se realizează deci anticipat și ulterior sunt transmise unităților de execuție. Transmiterea din memoria cache se realizează pe o magistrală de 256 de biți, dimensiunea mare a acesteia permițând aducerea secvențelor de instrucțiuni cu o viteză mai mare chiar decât a capacității de procesare.

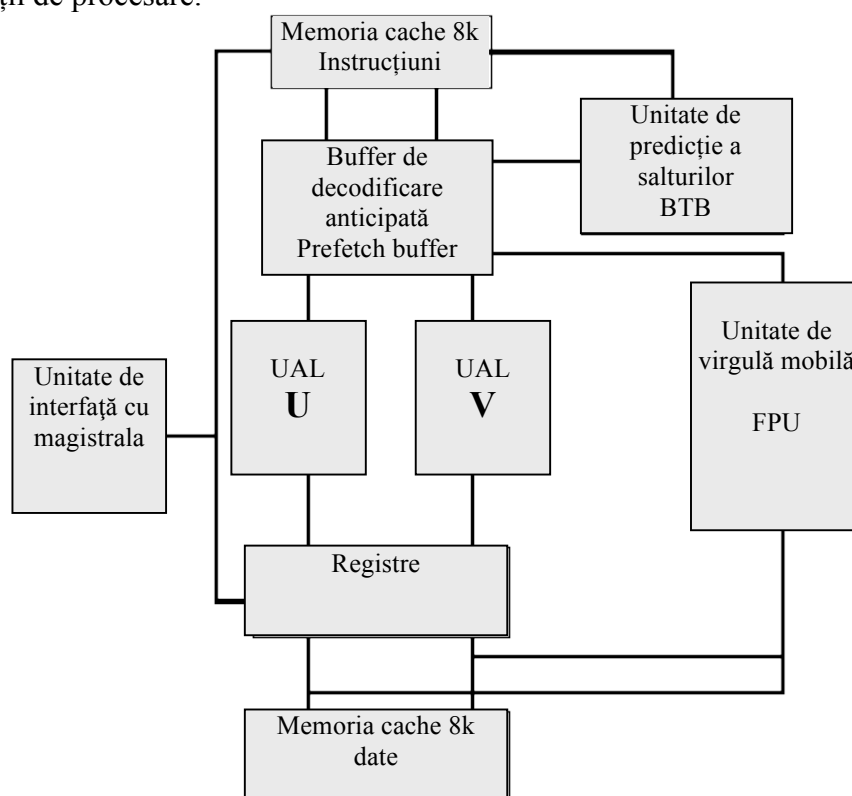


Figura 4.1

c) Procesorul *Pentium* utilizează predicția salturilor pentru a crește performanțele construcțiilor care utilizează bucle de program. Procesorul Pentium folosește *mecanismul predicției adreselor salturilor* (BTB-Branch Target Buffer). Conform statisticii, un program tipic la fiecare 6-8 instrucțiuni conține o instrucțiune de salt. Deci, la fiecare 6-8 instrucțiuni va fi necesară reactualizarea benzii de asamblare (BA) conform adresei de salt, și avantajele folosirii BA se pierd. Pentium folosește o memorie BTB în care se stochează informația ultimelor 256 de salturi. Majoritatea programelor conțin bucle de program, în decursul cărora se decide, ieșirea din buclă sau salt la începutul ei. În acest caz se memorează instrucțiunea ce controlează ramificarea, adresa saltului și presupunerea ce subprogram va fi executat. Unitatea BTB presupune, că ramificarea va fi repetată și furnizează instrucțiunea de salt benzii de asamblare. Statistica predicțiilor corecte este de 80% (conform Intel).

Registrele generale sunt tot de 32 de biți, dar s-au adăugat căi interne de date de 128 și 256 de biți pentru a crește viteza transferurilor interne, iar magistrala externă de date a fost extinsă la 64 de biți.

TLB (*Translation Lookaside Buffer*) translatează adresa liniară în adresa fizică.

Procesorului i s-a adăugat un controler avansat de întreruperi (*Advanced Programmable Interrupt Controller* – APIC) pentru a permite realizarea sistemelor cu mai multe procesoare.

Următoarea generație de procesoare este Intel P6. Microarhitectura procesoarelor din familia P6 este una superscalară cu trei benzi de asamblare. Aceasta înseamnă că, prin utilizarea tehnicilor de prelucrare paralelă, procesorul poate decodifica și executa în medie *trei instrucțiuni în fiecare ciclu de ceas*. Pentru a obține această rată de execuție a instrucțiunilor, procesoarele din familia P6 utilizează o cale de date de tip *pipeline* cu 12 segmente, care permite execuția instrucțiunilor într-o ordine diferită de cea specificată în program.

Figura 4.2 prezintă structura acestei căi de date, care este împărțită în patru unități de prelucrare:

- unitatea de extragere și decodificare,
- unitatea de expediere și execuție,
- unitatea de retragere și rezervorul de instrucțiuni.

Instrucțiunile și datele sunt transmise acestor unități prin unitatea de interfață cu magistrala.

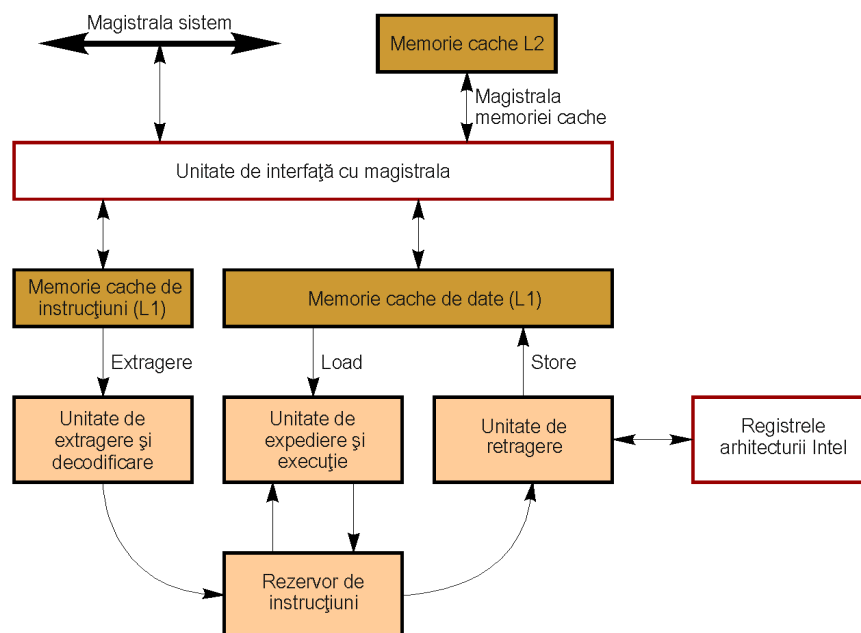


Figura 4.2 Unitățile de prelucrare din microarhitectura procesoarelor din familia P6 și interfața acestora cu subsistemul de memorie

Pentru a asigura furnizarea constantă a instrucțiunilor și datelor la unitățile de execuție din calea de date, microarhitectura utilizează două nivele de memorie *cache*.

Memoria *cache* L1 este împărțită într-o memorie *cache* de 8 KB pentru instrucțiuni și o memorie *cache* de 8 KB pentru date, ambele integrate în calea de date. Memoria *cache* L2 este o memorie RAM statică de 256 KB, 512 KB sau 1 MB, care este conectată la calea de date printr-o magistrală de 64 de biți, funcționând la frecvența de ceas a procesorului.

Microarhitectura procesoarelor din familia P6 utilizează *execuția dinamică* a instrucțiunilor. Mecanismul de execuție dinamică cuprinde trei concepte:

- Predicția salturilor;
- Analiza dinamică a fluxului de date;
- Execuția speculativă.

Predicția salturilor este un concept întâlnit la majoritatea arhitecturilor performante de calculatoare și a microprocesoarelor de viteză ridicată. Acest concept permite procesorului decodificarea instrucțiunilor de după cele de salt pentru a nu goli calea de date *pipeline* la execuția instrucțiunilor de salt. Unitatea de extragere și decodificare a instrucțiunilor utilizează un algoritm de predicție optimizat pentru anticiparea direcției fluxului de instrucțiuni prin nivele multiple de salturi, apeluri de proceduri și reveniri din proceduri.

Analiza dinamică a fluxului de date implică analiza în timp real a fluxului datelor prin procesor pentru a determina dependența datelor și a registrelor, și pentru a detecta posibilitatea execuției

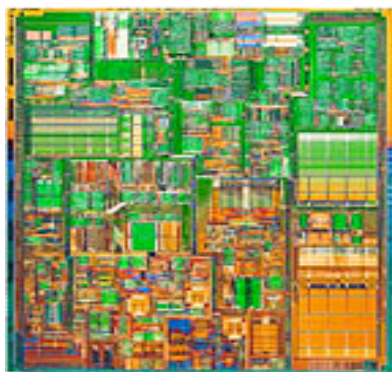
instrucțiunilor într-o ordine diferită de cea specificată în program. Unitatea de expediere și execuție a instrucțiunilor poate monitoriza simultan mai multe instrucțiuni și poate executa aceste instrucțiuni într-o ordine în care se optimizează utilizarea unităților multiple de execuție ale microarhitecturii, menținând în același timp integritatea datelor asupra cărora se operează. Această ordine de execuție asigură ocuparea unităților de execuție chiar și atunci când apar dependențe între datele instrucțiunilor.

Execuția speculativă se referă la posibilitatea procesorului de a executa instrucțiuni aflate înaintea instrucțiunii adresate de contorul de program, și de a furniza rezultatele în ordinea șirului inițial de instrucțiuni. Pentru ca execuția speculativă să fie posibilă, microarhitectura familiei P6 separă expedierea și execuția instrucțiunilor de producerea rezultatelor. Unitatea de expediere și execuție a instrucțiunilor utilizează analiza fluxului de date pentru a executa toate instrucțiunile din rezervorul de instrucțiuni și a memora rezultatele în registre temporare. Unitatea de retragere a instrucțiunilor caută apoi instrucțiunile care au fost executate și pentru care nu mai există dependențe de date cu alte instrucțiuni sau predicții nerezolvate ale salturilor. Atunci când se găsesc instrucțiuni care au fost executate, unitatea de retragere a instrucțiunilor depune rezultatele acestor instrucțiuni în memorie sau în registrele microarhitecturii (cele opt registre ale procesorului și opt registre ale unității de calcul în virgulă mobilă), în ordinea specificată în program, și retrage aceste instrucțiuni din rezervorul de instrucțiuni.

Aceste 3 concepte și sunt diferențele față de structura benzii de asamblare (fig.3.10), menționate anterior.

Schemele structurale ale microprocesoarelor, din următoarele generații, la nivel de microarhitectură devin tot mai complexe. Ne vom limita la descrierea noilor tehnologii implementate în microprocesoare.

Următorul reprezentant al microprocesoarelor, din generația Intel, este Pentium 4 cu microarhitectura NetBurst.

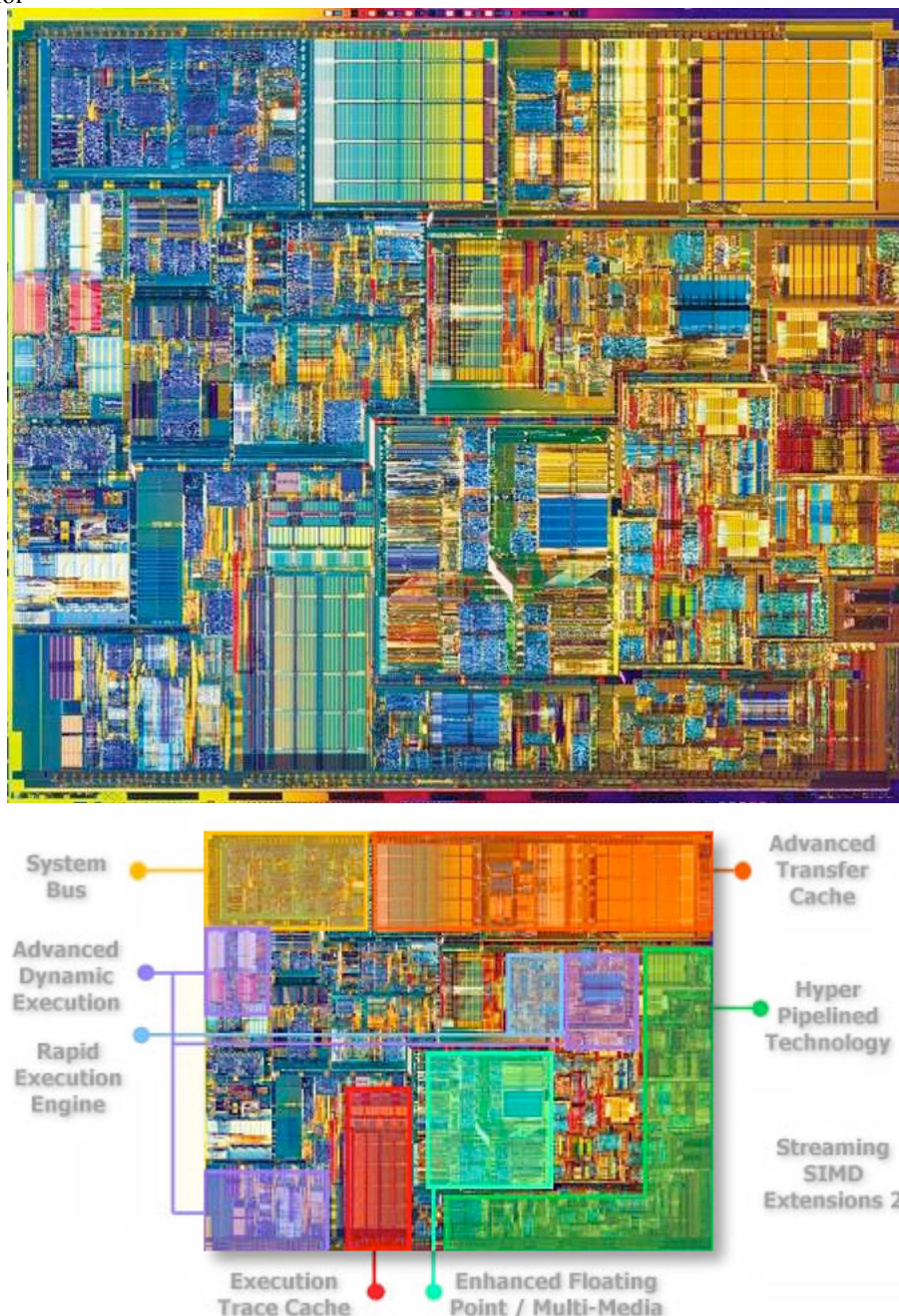


În microprocesor a fost implementată tehnologia *Hyper-threading*, ce permite execuția a două fluxuri de instrucțiuni de un singur nucleu fizic. Sistemul de operare, ce susține tehnologia *Hyper-threading*, determină existența a două procesoare logice (Logical procesor). Fizic, fiecărui procesor logic îi sunt alocate setul său de registre și un controler de întreruperi (APIC), iar restul componentelor ale procesorului sunt utilizate în comun. Potrivit Intel, utilizarea acestei tehnologii, majorează viteza de execuție cu 15-30%.

Următoarele elemente structurale implementate:

- adoptarea unei arhitecturi Hyper Pipeline cu 20 de segmente de procesare a instrucțiunilor (dublu față de varianta P6), fapt ce a majorat frecvența procesorului – de la 1,4 Ghz (3,2 Ghz în Pentium 4 Extreme Edition);
- a fost extinsă memoria cache L2 - 512 KB și L3 – până la 2MB (Pentium 4 Extreme Edition), iar transferurile între memoriile cache au ajuns la o rată de transfer de 48 GB/s;
- a fost introdusă o memorie cache pentru microinstrucțiuni, volum 12000 microinstrucțiuni;
- setul de instrucțiuni SSE2 a fost extins prin adăugarea a 144 instrucțiuni SIMD noi și extinderea lărimii datelor prelucrate la 128 biți (16 octeți prelucrați în paralel);
- a fost majorată rata de transfer între microprocesor și controlerul memoriei la 3,2 GB/s (La P6 – maximum 1,06 GB/s).
- proces tehnologic utilizat de 130 nm.

Pe următoarele figuri sunt prezentate nucleele procesorului Pentium 4.



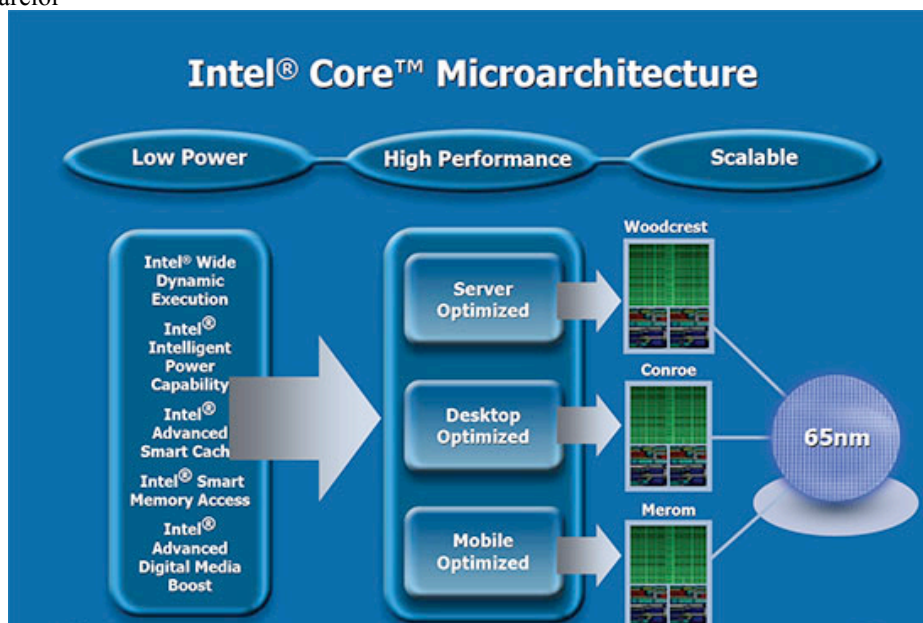
Dificultăți. Funcționarea procesorului la frecvențe înalte este asociată cu temperaturi ridicate ale nucleului său. Microprocesoarele Pentium 4 (nucleul Cedar Mill) au fost capabile să funcționeze la frecvențe mai mari de 7 GHz folosind răcire extremă (utilizând azot lichid). Din cauza dificultăților apărute la răcirea procesoarelor, frecvența maximă la Pentium 4, produs în serie, a fost limitată la 3800 MHz. Intel a fost nevoit să treacă de la o structură de mononucleu (single-core) la multinucleu (multi-core).

Compania Intel, în 2006, a prezentat noua microarhitectura Intel Core, care este o arhitectură multi-core. Ea încorporează mai multe nuclee, suportă tehnologia de virtualizare (Intel VT) și modul Intel 64 (64-bit mode), un set suplimentar de instrucțiuni SSE3.

Procesoarele pentru servere și stații de lucru au fost fabricate sub numele de brand Xeon, iar pentru utilizatorii de calculatoare variantele desktop și mobile - Core 2.



Microarhitectura Intel Core încorporează o bandă de asamblare cu 14 segmente (Pentium 4 "Prescott" - 31 segmente). Fiecare nucleu poate procesa, executa simultan până la patru instrucțiuni (microarhitectura NetBurst doar trei instrucțiuni).



În microarhitectura Intel Core a fost introdus un set de tehnologii care au primit numele Intel Wide Dynamic Execution. Cele mai importante sunt:

- **Advanced Smart Cache** – noua arhitectură este optimizată pentru procesoare cu 2 nuclee (dual-core). Memoria cache de nivelul L2 este partajată de ambele nuclee. Datele din memoria cache de nivelul L1 obligatoriu se află și în memoria cache de nivelul L2. Memoria cache de nivelul L2 este dinamic alocată ambelor nuclee pentru performanță maximă. Rata maximală de transfer între memorii - 96 GB/s (cu frecvența nucleului de 3 GHz).
- Tehnologia **Macro Fusion** constă în fuziunea dintre două instrucțiuni x86 (și câteva microinstrucțiuni - micro-ops Fusion), într-o singură instrucțiune (macroinstrucțiune). Unele perechi de instrucțiuni (de exemplu, instrucțiunea de comparare și de salt condiționat) la decodificare pot fi fuzionate și utilizate ulterior ca o singură microinstrucțiune. Ignorând tehnologia Macro Fusion procesorul (cu 4 unități de execuție) poate decodifica doar patru instrucțiuni. Utilizând tehnologia MacroFusion, procesorul poate decodifica cinci instrucțiuni (prin fuziunea a două din ele).

În luna iunie 2009 compania a anunțat că schimbă varietate de nume de brand, cum ar fi Core, Core 2 Duo, Core 2 Quad, Core 2 Extreme, în favoarea a trei nume: Core i3, Core i5 și Core i7. Dezvoltarea acestor microprocesoare, bazate pe microarhitecturi noi, a dus la divizarea lor pe generații:

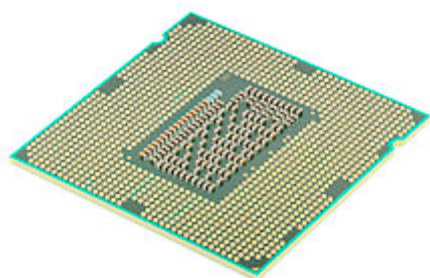
- Prima generație – microprocesoarele Core i3, Core i5 și Core i7 bazate pe microarhitectura Nehalem;
- A doua generație – microprocesoarele Core i3, Core i5 și Core i7 bazate pe microarhitectura Sandy Bridge;
- A treia generație – microprocesoarele Core i3, Core i5 și Core i7 bazate pe microarhitectura Ivy Bridge.

Prezentăm succint inovațiile microarhitecturii Nehalem:

- 2, 4 sau 8 nuclee;
- Tehnologia **SMT (Simultaneous Multi-Threading)**, permite execuția simultană a două fluxuri de instrucțiuni de un singur nucleu (2 nuclee logice din unul fizic);
- Memoria cache pe 3 nivele: cache L1 volum de 64 KB pe nucleu, Cache L2 volum de 256 KB pe nucleu, memoria Cache L3 partajată de toate nuclee, volum până la 24 MB;
- Controlerul de memorie încorporat, ce suportă câteva canale de memorie DDR3 SDRAM;
- Proces tehnologic de 45nm;
- Posibilitatea de încorporare în microprocesor a nucleului grafic;
- Magistrală de mare viteză **QPI (Quick Path Interconnect)** (înlocuind FSB) cu topologia punct-la-punct, destinată interconectării procesorului cu chipset-ul și a procesoarelor în sistem multiprocesor;
- Tehnologia **Turbo Boost** - permite automat, ca nucleele procesorului să poată funcționa cu o frecvență mai mare decât frecvența de bază (nominală), doar dacă procesorul funcționează în

cadrul limitelor de putere, curent și temperatură, conform TDP (TDP - *thermal design power*).

- Structura pe module;
- Set suplimentar de instrucțiuni SSE4.2.



Următoarea microarhitectură **Sandy Bridge**.

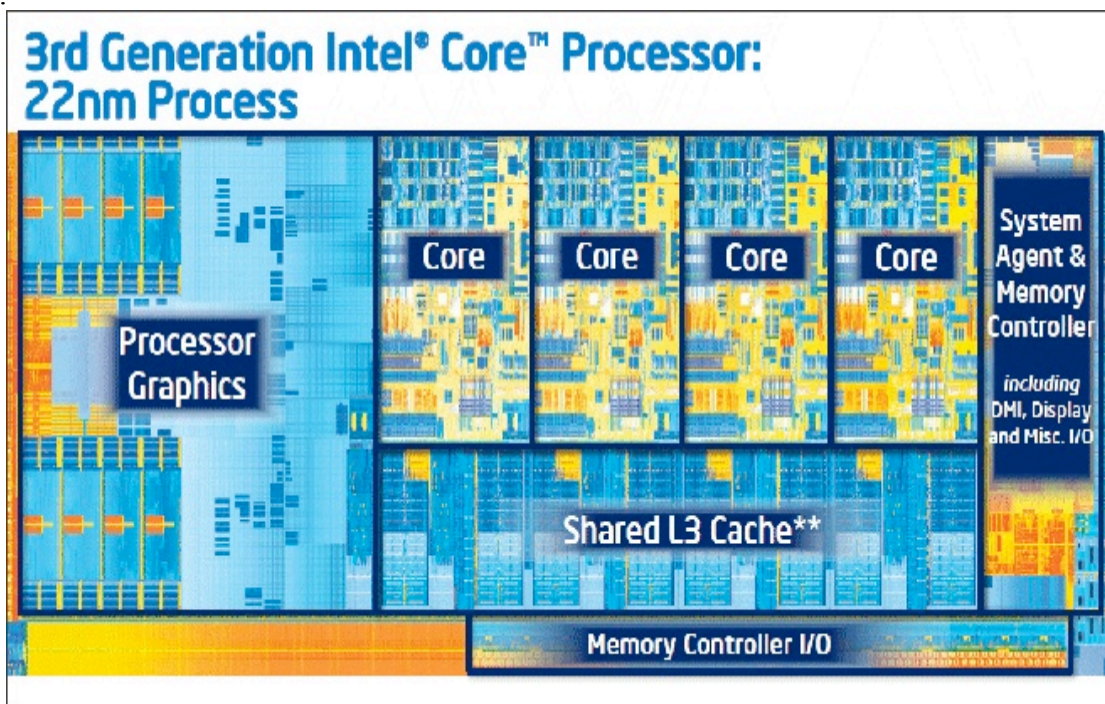
Principalele îmbunătățiri.

Circuitul microprocesorului Sandy Bridge încorporează principalele componente:

- Nucleele microprocesorului;
- Nucleul grafic (HD Graphics 3000, cu 12 unități de execuție);
- Memoria cache L3;
- „Puntea de nord” (System Agent).

Toate aceste componente sunt interconectate prin intermediul noii versiuni a tehnologiei QPI, o magistrală inelară pe 256-biți. Proces tehnologic de 32nm.

Următoarea microarhitectură **Ivy Bridge**. Ivy Bridge — este o versiune a microarhitecturii Sandy Bridge cu proces tehnologic de 22nm. În aprilie 2012 au fost prezentate procesoarele cu microarhitectura menționată.



Ivy Bridge include următoarele îmbunătățiri față de platforma precedentă (Sandy Bridge):

- Tehnologia « Tri-gate transistor » (“3D” sau pe 3 căi) care la același nivel de performanță ca tranzistorii 2-D au un consum de energie mai redus cu până la 50%;
- Suport PCI Express 3.0;
- Intel HD Graphics 2500/4000 cu suport a 3 monitoare independente;
- GPU (procesorul grafic integrat) are 16 unități de execuție (EUs) în comparație cu Sandy Bridge -12;
- DDR3L (“L” desemnează low-voltage = tensiune scăzută) și TDP configurabil pentru procesoarele mobile.

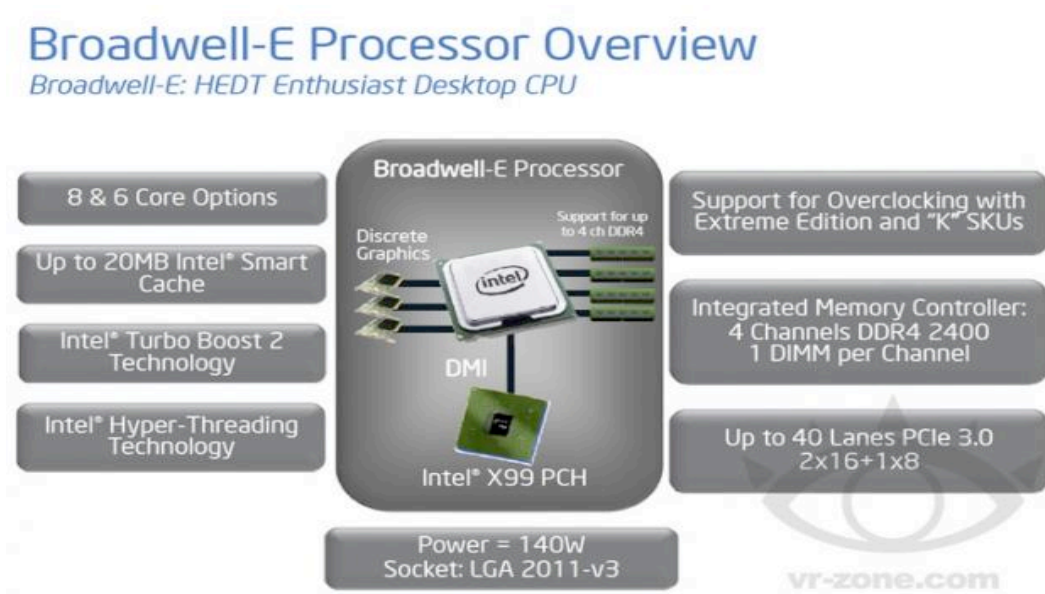
Este sugestivă următoarea comparație de performanță cu generația Sandy Bridge:

- 5% până la 15% creștere în performanța generală a procesorului,
- 20% până la 50% creștere în performanța procesorului grafic integrat (GPU).

22nm Ivy Bridge , 32nm Sandy Bridge



Pe următoarea figură sunt prezentate unele tehnologii susținute de microarhitectura Broadwell. Proces tehnologic de 14 nm.



Pe următoarea figură sunt prezentate dimensiunile comparative ale microprocesoarelor Broadwell și Haswell.

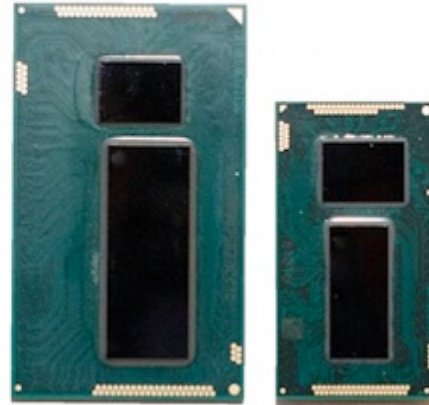
Broadwell Y Platform Enabled Board Area Reduction of ~25% Compared to Haswell

50% Smaller XY

30% Smaller Z

Key Enablers:

- 0.63x scaling due to 14nm
- 0.5mm ball pitch
- 200um PKG Core
- 170um thin die
- 3DL



HSW U/Y
40x24x1.5mm

BDW-Y
30x16.5x1.04mm

