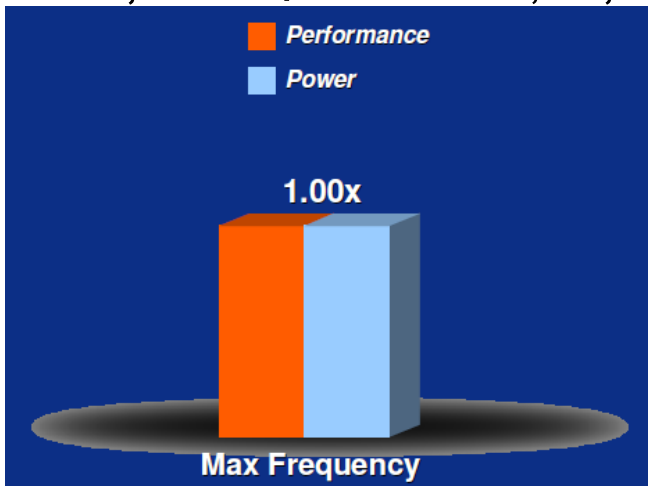
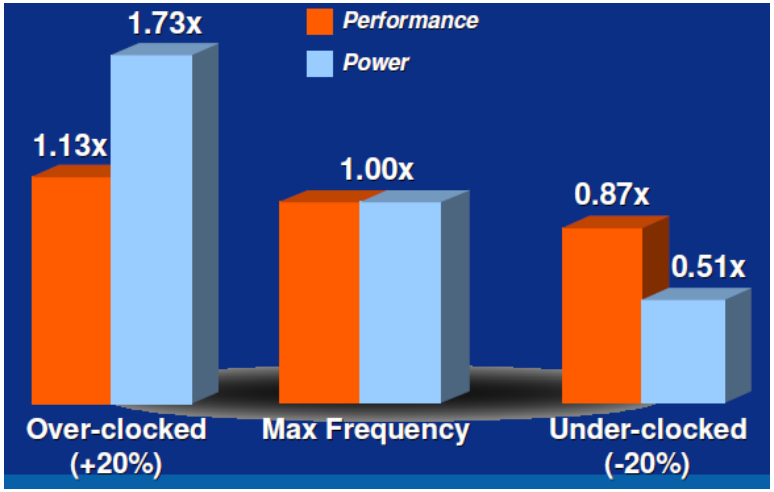
# Curs 13

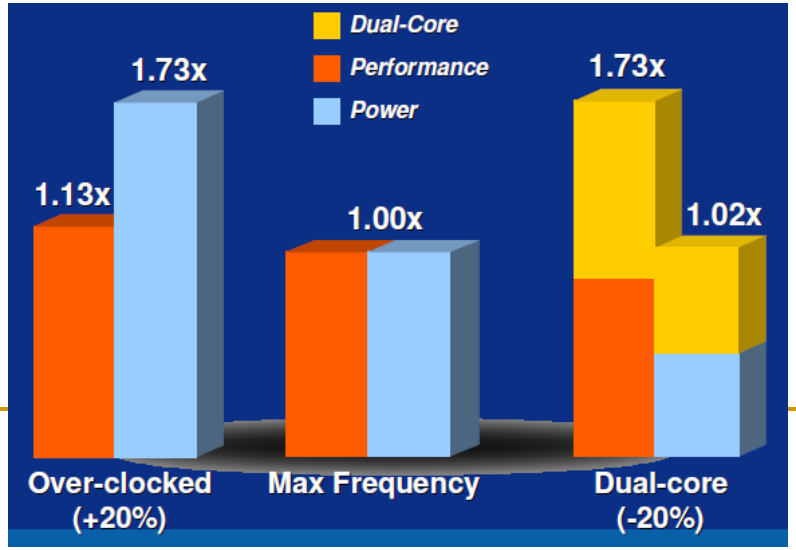
# 9.2. Microprocesoare multicore

## 9.2.1. Necesitatea apariției microprocesoarelor multicore

- Creșterea performanței și păstrarea/scăderea consumului



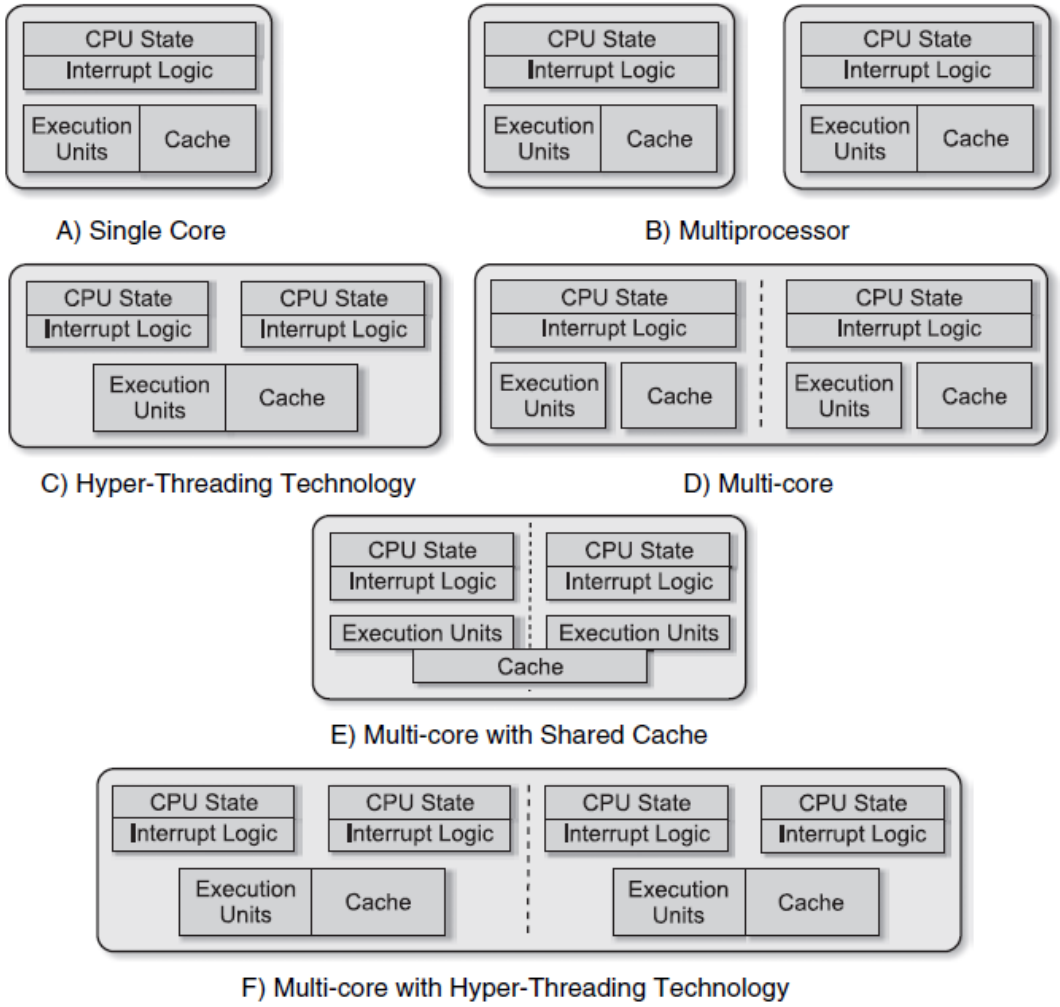



# Proiectarea cu Microprocesoare

- Evoluția
  - Microprocesoare unicore
    - Execuție secvențială a instrucțiunilor;
    - Facilități: pipeline, out-of-order execution (instruction level parallelilsm)
      - Performanța limitată de interdependențele între instrucțiuni;
      - Performanța pipeline-ului limitată de instrucțiunile care modifică secvențialitatea;
  - Multimicroprocesoare: sisteme cu mai multe microprocesoare + circuite, cu acces comun la resurse;
    - Problemă: accesul la resursele comune;
  - Microprocesoare multitasking (multithreading)
    - Un unic microprocesor fizic este văzut de software ca mai multe microprocesoare logice;
    - Fiecare microprocesor logic execută un task;
    - Problemă: comutarea taskurilor;
  - Microprocesoare multicore
    - Mai multe microprocesoare fizice (core (nucleu)) execută fiecare câte un task;

# Proiectarea cu Microprocesoare

- Reprezentarea evoluției



- Creșterea performanței prin paralelizare

  - Se măsoară cu factorul de accelerare (câștig de viteză) (FA):

  $Speedup\ (n) = T_{cel\_mai\_bun\_alg\_secvențial} / T_{exec\_paralelă\_n\_nuclee}$

  - Care este limita teoretică a creșterii performanței (vitezei) prin creșterea numărului procesoarelor?
  - Legea lui Amdahl: G. Amdahl a considerat că orice program paralel are o componentă secvenţială care nu poate fi paralelizată, *S* şi o parte complet paralelizabilă, *P*; timpul de execuţie al programului pe un procesor unicore va fi *S* + *P* iar timpul de execuţie pe un processor cu n nuclee va fi *S* + *P*/ *n*; atunci FA va fi:

  $$Speedup\ (n) = (S + P)/(S + P/n) = 1/((1 – P) + P/n) = 1/(S + P/n) \leq 1/S$$

  adică creșterea de viteză este limitată de porțiunea din cod neparalelizată, indiferent de numărul de nuclee;

# Proiectarea cu Microprocesoare

- Creşterea numărului de nuclee nu duce automat la creşterea performanţelor; FA este limitat de inversul fracţiunii din program care trebuie executată secvenţial;

- Dacă *S = 0,1* atunci *FA ≤ 10* indiferent de numărul de procesoare;

- La limită, dacă programul este complet secvenţial, adică *P = 0*, atunci *FA ≤ 1* adică durata de execuţie a programului pe un sistem multicore va fi egală sau mai mare decât pe un sistem unicore; degradarea acestui timp apare datorită comunicării între nuclee, în cazul sistemului multicore;

- Dacă jumătate din program este executat de un procesor cu 2 nuclee atunci factorul de accelerare va fi:

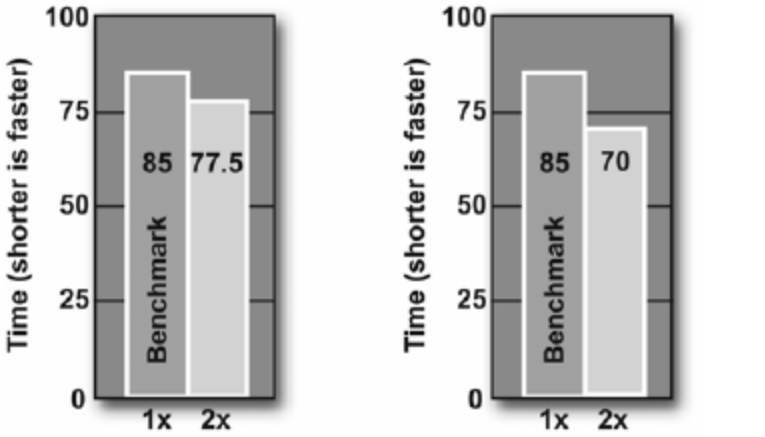
$$Speedup\ (2) = 1/(0,5 + 0,5/2) = 1/0,75 = 1,33$$

adică un câștig de performanță de 33%; la 8 nuclee creșterea este de 78%

# Proiectarea cu Microprocesoare

- Din legea lui Amdahl rezultă că FA este limitat superior de o valoare care nu depinde de numărul de nuclee sau de arhitectura procesorului;
- Dacă $n \rightarrow$ infinit, atunci rezultă aceeași concluzie:

  *Speedup = 1/(1 – P) = 1/S*
- Corolar: este mai eficientă creșterea porțiunii din cod paralelizată decât creșterea numărului nucleelor; ex.: cod paralelizat în raport de 30%, rulând pe un procesor cu 2 nuclee; dublând numărul nucleelor scade timpul de rulare de la 85 unități la 77,5 unități; dublând porțiunea din cod paralelizată, scade timpul de rulare de la 85 unități la 70 unități;




Performance benefit of doubling the number of processor cores

Performance benefit of doubling the amount of parallelism in code

Note: The advantage gained by writing parallel code

- În cazul procesoarelor multicore legea lui Amdahl se rescrie:

  *Speedup = 1/(S + (1 – S)/n + H(n))*

  unde *H(n)* = overhead;

  - Overhead are 2 componente:
    - Overhead la nivelul sistemului de operare;
    - Overhead datorat activităților inter-thread (sincronizare, comunicare etc.);
  - Dacă overhead este prea mare există riscul ca *Speedup* < *1* adică performanța scade comparativ cu procesorul unicore;
  - Overhead trebuie minimizat; se obține din arhitectura multithread;
- În cazul procesoarelor cu tehnologie Hyper-Threading:
  - Faptul că unele resurse sunt partajate de mai multe thread-uri afectează performanța;
  - Înăuntrul procesorului, fiecare thread rulează mai încet decât dacă ar fi deținut întregul procesor întrucât există resurse partajate (de ex. unitatea de execuție); întârzierea depinde de tip de aplicație; dacă se consideră o întârziere de 33% atunci legea lui Amdahl devine:

    *Speedup = 1/(S + 0,67((1 – S)/n + H(n)))*

    *H(n)* se determină empiric și valoarea sa depinde de la un tip de aplicații la altul;

# Proiectarea cu Microprocesoare

- Legea lui Gustafson:
  - Obs. la legea lui Amdahl:
    - Legea lui Amdahl se aplică la sistemele multicore cu memorie unică, partajată; la sistemele cu memorie distribuită nu se mai aplică întrucât se pot face accese concurente la memorie; un procesor multicore poate avea memorii cache separate ptr. fiecare nucleu, reducând drastic întârzierea în accesarea memoriei;
    - Algoritmul serial este cea mai bună soluție pentru o problemă; nu este întotdeauna adevărat, există probleme la care soluția paralelă este mai eficientă; ex.: aplicațiile SIMD;
    - Dependența scăderii timpului de rulare la creșterea numărului de nuclee este limitată: în practică creșterea numărului de nuclee a dus la scăderea continuă a timpului de rulare datorită fie unui algoritm secvențial neoptimal fie unor caracteristici hardware (de ex. datele unei probleme pot fi atât de multe încât să nu încapă în memoria unui număr mic de nuclee, ca urmare va fi necesară o memorie suplimentară care va degrada performanțele și atunci creșterea nr. de nuclee și partiționarea datelor între ele va asigura ca datele să încapă în memoria existentă);
  - Fie un procesor cu p nuclee, cu memorii cache distribuite şi considerăm execuţia unui algoritm; timpul pentru varianta paralelă este $S + P$ iar timpul pentru o execuţie secvenţială ar fi $S + pP$; factorul de accelerare devine:

$$SG_p = \frac{S + pP}{S + P} = \frac{S + P + (p-1)P}{S + P} = 1 + \frac{(p-1)P}{S + P} \geq \frac{(p-1)P}{S + P}.$$

- FA este limitat inferior, de data aceasta, de o valoare care depinde de numărul de nuclee; limita inferioară a FA se numeşte factorul lui Gustafson;

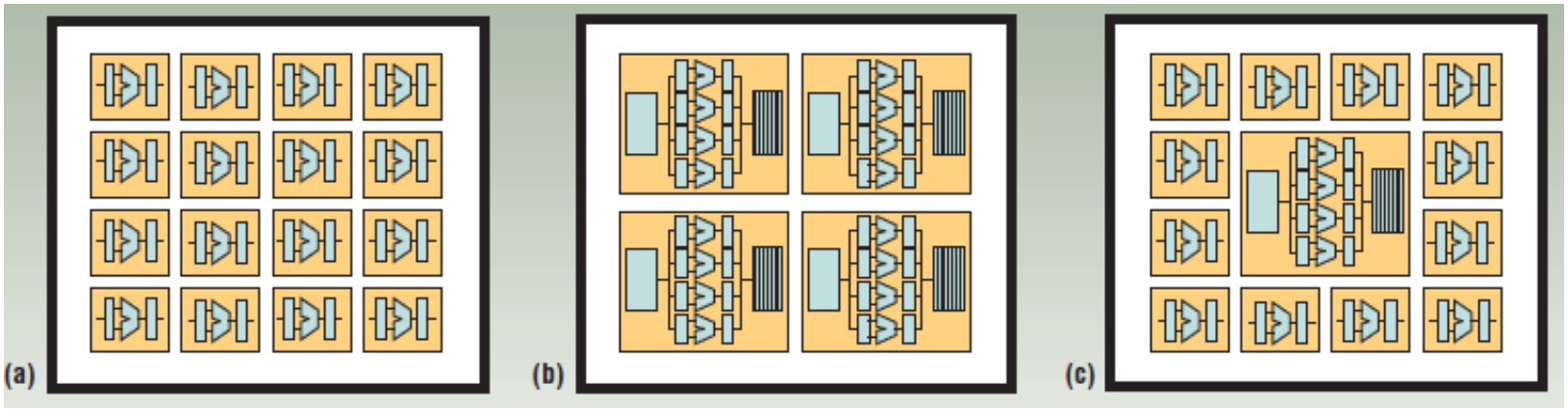
- Legea lui Minsky
  - Minsky a considerat execuţia programelor cu puncte de ramificare pe o structură SIMD; dacă un program are un punct de ramificare cu ramuri egale, jumătate din nuclee vor fi inactive la execuţie; fiecare ramură poate admite, ea însăşi, un punct de ramificare; în cazul ideal în care fiecare ramură din program admite p ramificări binare de aceeaşi lungime, programul va avea o structură arborescentă; fie $t_1$ timpul secvenţial total, rezultă timpul $t_1 / log\ p$ pentru fiecare din cele *log p* părţi din program; scriind timpul aferent execuției de către $2^i$ nuclee se obţine:

$$t_1 = \sum_{i=1}^{log\ p} \frac{t_1}{log_2\ p}; t_p = \sum_{i=1}^{log\ p} \frac{t_1}{2^i\ log_2\ p} \leq \frac{t_1}{log_2\ p}; S_p \leq log_2\ p$$

  - FA este limitat de numărul de nuclee.

# Proiectarea cu Microprocesoare

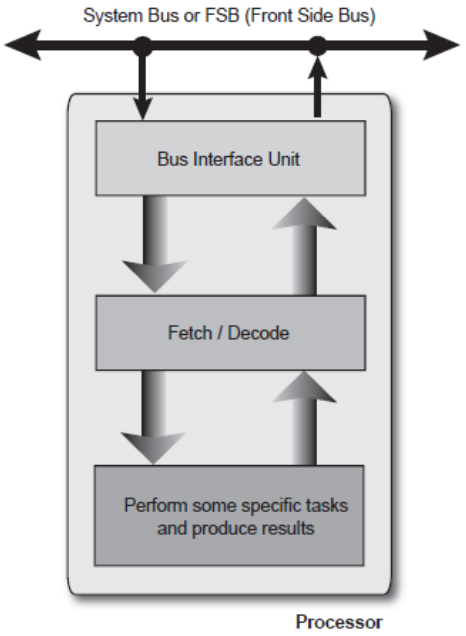
- Exemple de structuri multicore:




  - a: multicore simetric cu 16 nuclee cu 1 resursă;
  - b: multicore simetric cu 4 nuclee cu 4 resurse;
  - c: multicore asimetric cu 12 nuclee cu 1 resursă și 1 nucleu cu 4 resurse;
  - Procesoarele multicore pot fi:
    - Simetrice: toate nucleele sunt la fel (resurse, performanțe, cost);
    - Asimetrice: nuclee diferite (resurse, performanțe, cost);
    - Dinamice: nucleele își pot modifica caracteristicile în timpul rulării; sau:
    - Omogene: toate nucleele au aceeași funcționalitate;
    - Heterogene: o parte din nuclee au funcționalități diferite, sunt neconvenționale (GPU, FPGA etc.);

## 9.2.2. Structuri unicore vs. multicore

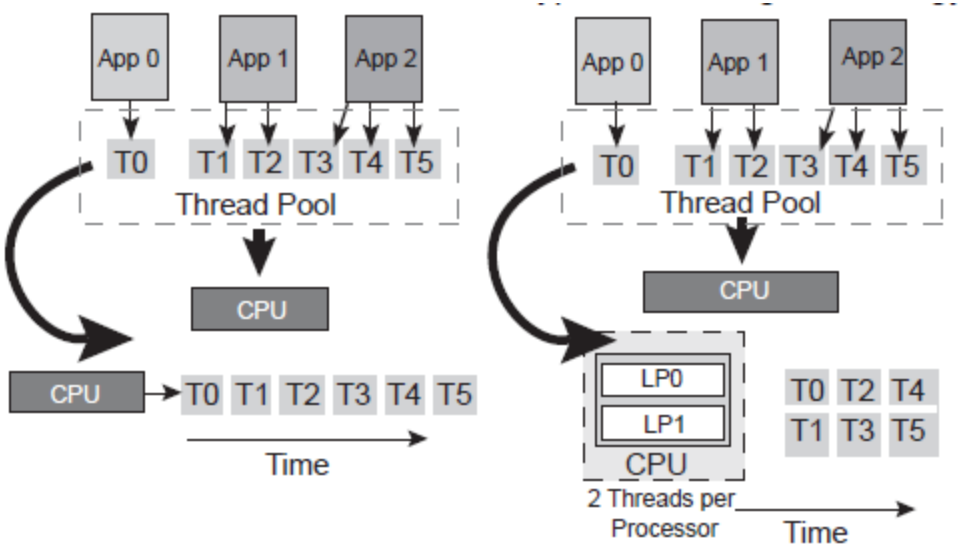
- Execuția la microprocesoare unicore
  - Secvențială:




  - Hyperthreading:
    - OS vede 2 procesoare logice, fiecare poate să execute un thread; instrucțiunile sunt executate simultan;
    - Multithreading/hyperthreading poate fi fine-grained sau coarse-grained;

# Proiectarea cu Microprocesoare

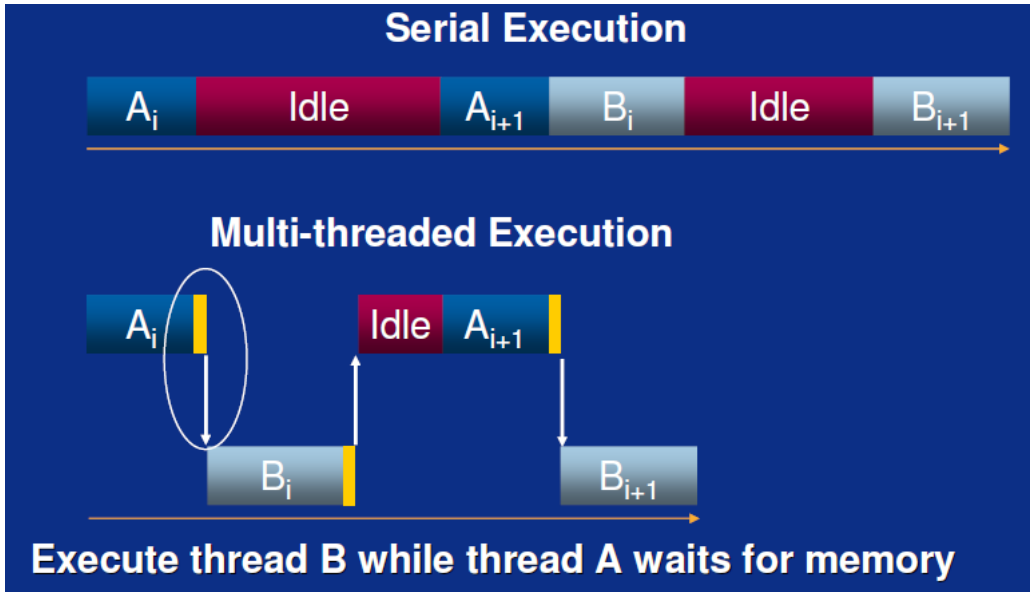
- Comparație între execuția mai multor thread-uri pe un procesor fără și cu tehnologie hyperthreading:

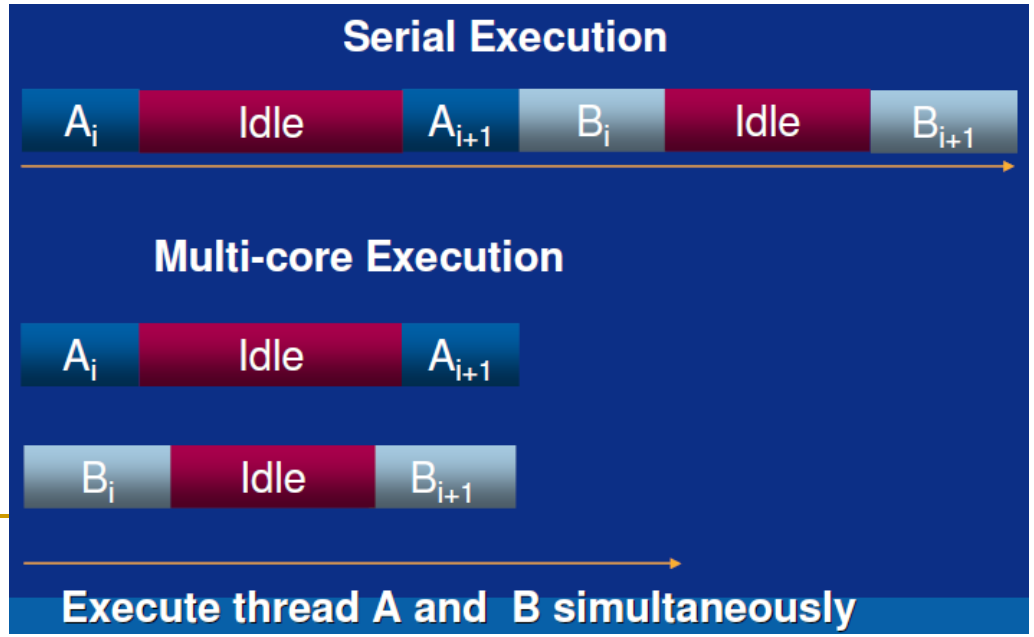



- Fine-grained multithreading/hyperthreading înseamnă deținerea resurselor de către un thread un timp egal cu cel pentru execuția unei instrucțiuni;
- Coarse-grained (swith on event) multithreading/hyperthreading înseamnă deținerea resurselor de către un thread un timp egal cu cel pentru execuția unui proces;
- Overhead mai mare la fine-grained multithreading/hyperthreading;

# Proiectarea cu Microprocesoare

- Execuția la micr. multicore
  - Hyperthreading vs. Multicore



# 9.3. Microprocesoare moderne

## 9.3.1. Caracteristici generale

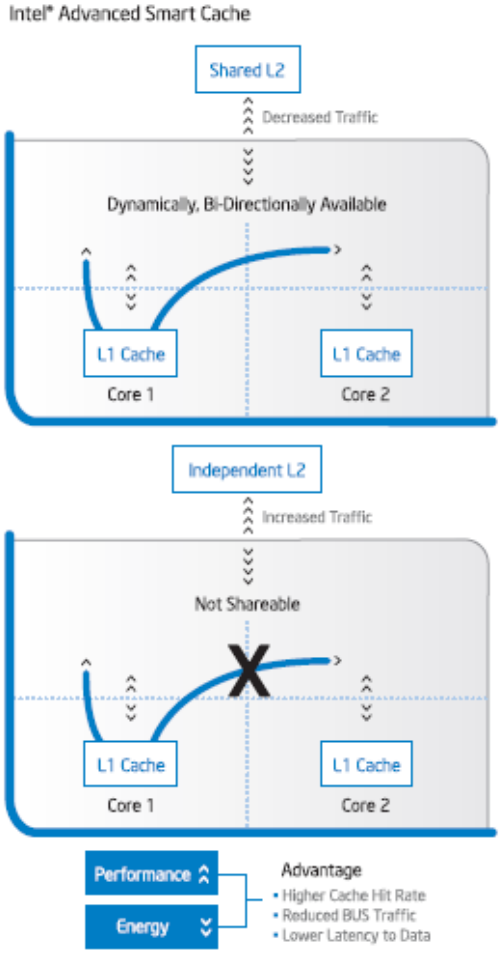
- Începând cu microarhitectura Intel Core (2006):
  - Intel Core Uno (Solo);
  - Intel Core Duo;
  - Intel Core 2 Duo;
  - Intel Core 2 Quad;
  - Intel Core i3, i5, i7, i9.
- Inovații ale microarhitecturii Intel Core: scopul este creșterea performanței (vitezei) și scăderea consumului:
  - Intel Wide Dynamic Execution;
  - Intel Intelligent Power Capability;
  - Intel Advanced Smart Cache;
  - Intel Smart Memory Access;
  - Intel Advanced Digital Media Boost.

# Proiectarea cu Microprocesoare

- Intel Wide Dynamic Execution:
  - O combinație de tehnici folosite la procesoarele Pentium: analiza fluxului de date, execuția speculativă, execuția out-of-order, superscalaritatea;
    - Scopul este eliminarea sau minimizarea timpilor de așteptare;
    - Permite creșterea numărului de instrucțiuni care pot fi executate în 1 ciclu;
    - Utilizând macro și microfuziunea numărul de instrucțiuni executate simultan de un nucleu poate crește;
  - Predicția performantă a salturilor și apelurilor condiționate;
  - Macrofuziunea:
    - Înseamnă unirea a 2 instrucțiuni în o singură secvență de microinstrucțiuni, exemplu: o instrucțiune de comparare urmată de o instrucțiune de salt condiționat;
    - Este posibilă doar între anumite perechi de instrucțiuni;
  - Microfuziunea:
    - Înseamnă unirea a 2 sau mai multe microoperații care alcătuiesc execuția unei instrucțiuni;
    - Este posibilă doar între anumite perechi de instrucțiuni;
- Intel Intelligent Power Capability
  - Gestionează puterea în mod dinamic;
  - Se bazează pe Power-Gating; întârzierea introdusă de etajul de comandă a alimentării este minimizată;

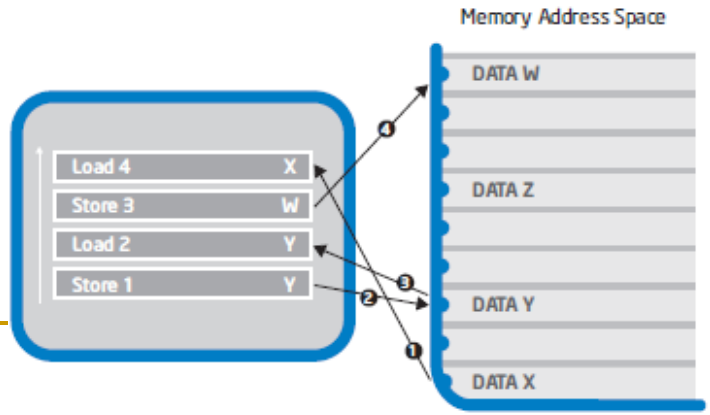
# Proiectarea cu Microprocesoare

- Intel Advanced Smart Cache:
  - Prin crearea unei memorii cache L2 comune, crește probabilitatea ca fiecare nucleu să acceseze informația dorită din memoria cache;
  - Dacă un nucleu accesează rareori memoria cache comună, alt nucleu poate să o utilizeze mult mai intens;



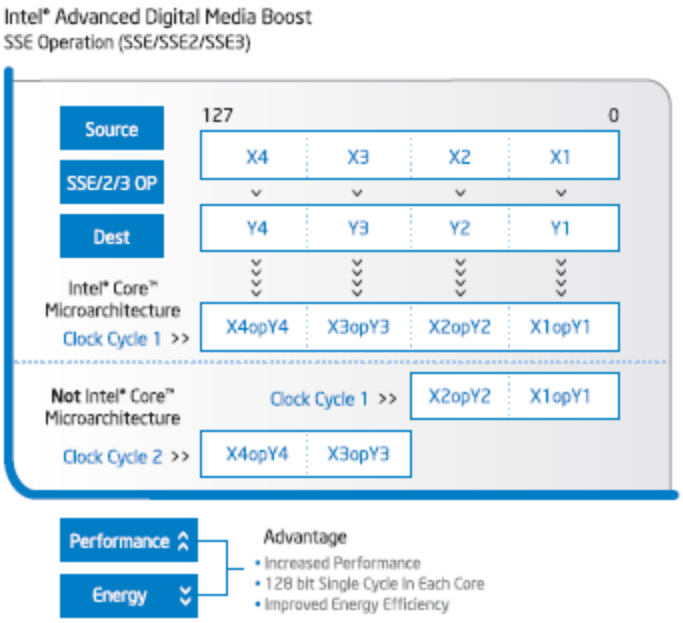
# Proiectarea cu Microprocesoare

- Intel Smart Memory Access:
  - Crește lățimea de bandă disponibilă pentru memorie și scade latența memoriei; scopul este ca memoria să poată fi accesată cât mai repede, să ofere/preia date cât mai repede și acestea să fie poziționate cât mai aproape de locul în care este nevoie de ele;
  - Există 2 unități de prefetch la nivelul memoriilor cache L1 și L2;
  - Memory disambiguation: crește eficiența procesării out-of-order:
    - Nucleele au posibilitatea de a executa speculativ operații de *load* înaintea celor de *store*;
    - Fără această caracteristică, se recomandă execuția operației de *load* după cea de *store* pentru a ține cont de eventualele dependențe între date;
    - Există operații de *load* care nu au legătură cu datele depuse de *store*, ca urmare pot fi executate înaintea de *store*; memory disambiguation permite nucleelor să detecteze aceste *load*-uri și să le execute; în rarele situații în care detecția nu a reușit, operația de *load* va fi invalidată și va fi executată după *store*;
    - Ex.: *Load 4* poate fi executat înaintea lui *Store 1* și *Store 3*:



# Proiectarea cu Microprocesoare

- Intel Advanced Digital Media Boost:
  - Permite creșterea performanței la execuția operațiilor de tip SIMD, cu instrucțiuni pe 128 biți (extensia SSE – Streaming SIMD Extensions); instrucțiunile pe 128 biți pot fi executate în un singur ciclu față de 2 cicluri în lipsa acestei caracteristici;
  - Util la aplicații multimedia, video, sunet, procesare foto, criptare, calcule inginerești etc.;



# Proiectarea cu Microprocesoare

## 9.3.2. Microprocesoarele Intel

- Perioada: noiembrie 1971 – 2019;
- Evoluția după nr. de biți: 4, 8, bit-slice, 16, 32, 64;
- Evoluția microarhitecturilor: x86, IA32, Pentium, Celeron, NetBurst, IA64, Xeon, Intel Core, →
- Evoluția după generații: 1 – 9 (după Intel Core, inclusiv);
- Evoluția după frecvența tactului: 740 KHz – 4 GHz (+ Turbo Boost);
- Evoluția după gradul de integrare: 2300 tranzistoare – 2,2 miliarde în 2011 – 7 miliarde la unele exemplare i9 în 2019;
- Evoluția după tehnologie: 10 μm – 10 nm;
- Evoluția după capacitatea de memorie adresabilă: ~ 5 Ko – $2^{64}$ o + ~ 25 Mo memorie cache;
- Evoluția după complexitate: 1 nucleu, 1 thread – 18 nuclee, 36 thread-uri;
- Evoluția după consum (putere): ~ 0,5 W – 165 W/nucleu;
- Evoluția după specializare: desktop, mobile, servere;
- Preț (actual, 2019): ~ 26 $ - 2000 $.

# Proiectarea cu Microprocesoare

- Microprocesoarele i3, i5, i7, i9 (microarhitecturi):

| | i3 | i5 | i7 | i9 |
|---|---|---|---|---|
| Generația 1 | Nehalem, Clarkdale, Westmere | Nehalem, Lynnfield, Clarkdale | Nehalem, Bloomfield, Lynnfield, Gulftown, Clarskfield | |
| Generația 2 | Sandy Bridge | Sandy Bridge | Sandy Bridge | |
| Generația 3 | Ivy Bridge | Ivy Bridge | Ivy Bridge | |
| Generația 4 | Haswell | Haswell | Haswell | |
| Generația 5 | Broadwell | Broadwell | Broadwell | |
| Generația 6 | Skylake | Skylake | Skylake | |
| Generația 7 | Kaby Lake | Kaby Lake, Skylake | Kaby Lake, Skylake | |
| Generația 8 | Coffee Lake | Kaby Lake, Coffee Lake, Amber Lake, Skylake | Kaby Lake, Coffee Lake, Amber Lake, Skylake | |
| Generația 9 | Cannon Lake | Coffee Lake Refresh, Skylake X | Coffee Lake Refresh, Skylake X | Coffee Lake Refresh, Skylake X |

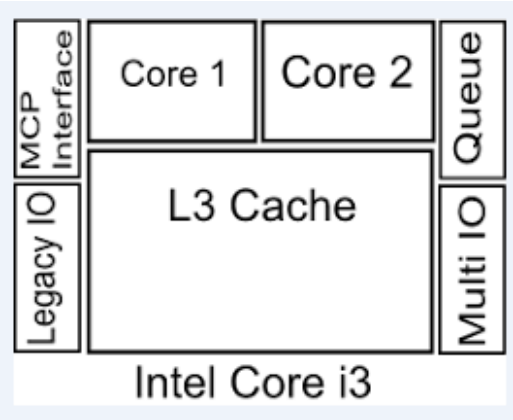
# Proiectarea cu Microprocesoare

- Caracteristici:

| | An | Tehnologie [nm] | Nuclee/ thread-uri | Mem, cache L1, L2, L3 | Frecvență | HT, TB | Include |
|---|---|---|---|---|---|---|---|
| I3, Clarkdale | 2010 | 32 | 2/4 | 32+32 K, 256 K, 4M | ~ 3 GHz | HT | HD GPU |
| I5, Sandy Bridge | 2011 | 32 | 4/4 | 32+32 K, 256 K, 6M | ~ 3 GHz | TB | GPU, HD GPU |
| I7, Ivy Bridge | 2012 | 22 | 4/8 | 32+32 K, 256 K, 8M | ~ 3 GHz | TB | HD GPU |
| I3, Haswell | 2013 – 2014 | 22 | 2/4 | 3 M L3 | ~ 3,5 GHz | HT | HD GPU |
| I5, Broadwell | 2015 | 14 | 4/4 | 4 M L3 | ~ 3 GHz | TB | GPU |
| I7, Skylake | 2015 | 14 | 4/8 | 8 M L3 | ~ 3,5 GHz | TB | HD GPU |
| I3, Kaby Lake | 2016 – 2017 | 14 | 2/4 | 32+32 K, 256 K, 3M | 2,4 – 3.9 GHz | HT | HD GPU |
| I5, Coffee Lake | 2017 | 14 | 8/6 | 9 M L3 | ~ 3,6 GHz | TB | HD GPU |
| I9, Coffee Lake, Extreme Edition | 2018 | 14 | 18/36 | 24,75 M L3 | 3 GHz | HT, TB | HD GPU |

# Proiectarea cu Microprocesoare

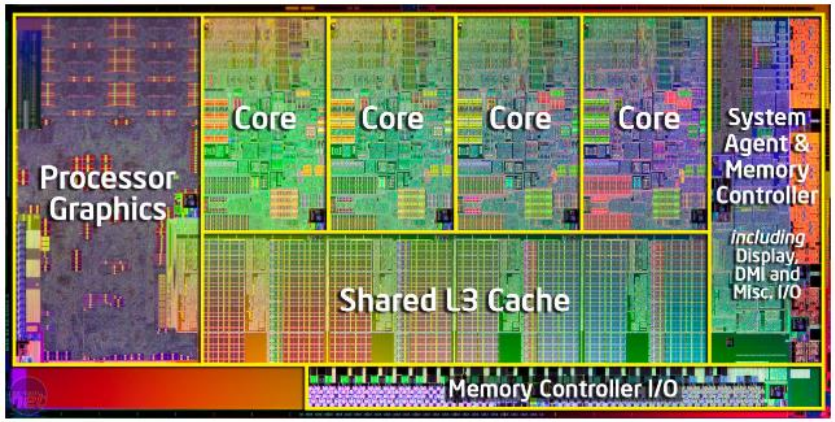
- Microprocesorul i3
  - Procesor folosit la desktop-uri și laptop-uri;
  - Procesor cu 2 nuclee, unele modele din generația a 8-a au 4 nuclee; prin implementarea HT, se obțin 4 thread-uri;
  - Structura:




  - Memorie cache L3 de capacitate 3 sau 4 Mo;
  - Variante:
    - Viteze mici, 1.30 GHz – 1.80 GHz, și putere/nucleu 11.5 W, 15 W sau 25 W; folosite în laptop-uri;
    - Viteze medii, 2.00 GHz – 2.50 GHz, și putere/nucleu 28 W, 35 W sau 37 W; folosite în laptop-uri și desktop-uri;
    - Viteze mari, 2.90 GHz – 3.90 GHz, și putere/nucleu 35 W, 37 W sau 54 W; folosite în desktop-uri;

# Proiectarea cu Microprocesoare

- Microprocesorul i5
  - Procesor folosit la desktop-uri și laptop-uri;
  - Procesor cu 4 nuclee, unele modele din generația a 9-a au 8 nuclee; nu au HT în schimb au TB;
  - Structura:




  - Memorie cache L3 de capacitate 3, 4, 6 sau 8 Mo;
  - Variante:
    - Viteze mici, 1.90 GHz – 2.30 GHz, și putere/nucleu 11.5 W; folosite în laptop-uri;
    - Viteze medii, 2.60 GHz – 3.10 GHz, și putere/nucleu 15 W, 25 W sau 28 W; folosite în laptop-uri și desktop-uri;
    - Viteze mari, 3.20 GHz – 3.80 GHz, și putere/nucleu 35 W, 37 W, 45 W, 47 W, 65 W sau 84 W; folosite în desktop-uri;

# Proiectarea cu Microprocesoare

- Alimentarea procesorului (inclusiv a procesorului grafic inclus)
  - Procesorul include un regulator de temperatură care este comandat prin semnalele VID0-7; este posibil ca valorile pentru semnalele VID0-7 să fie stabilite în faza de fabricație, ca urmare 2 procesoare similare pot avea praguri de alimentare diferite;
  - Tabelul următor arată legătura între configurațiile VID și nivelul tensiunii de alimentare:

| VID 7 | VID 6 | VID 5 | VID 4 | VID 3 | VID 2 | VID 1 | VID 0 | $V_{CC\_MAX}$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | OFF |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | OFF |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1.60000 |
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1.59375 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1.58750 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1.58125 |
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1.57500 |
| 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1.56875 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1.08750 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1.08125 |
| 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1.07500 |
| 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1.06875 |
| 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1.06250 |
| 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1.05625 |
| 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1.05000 |

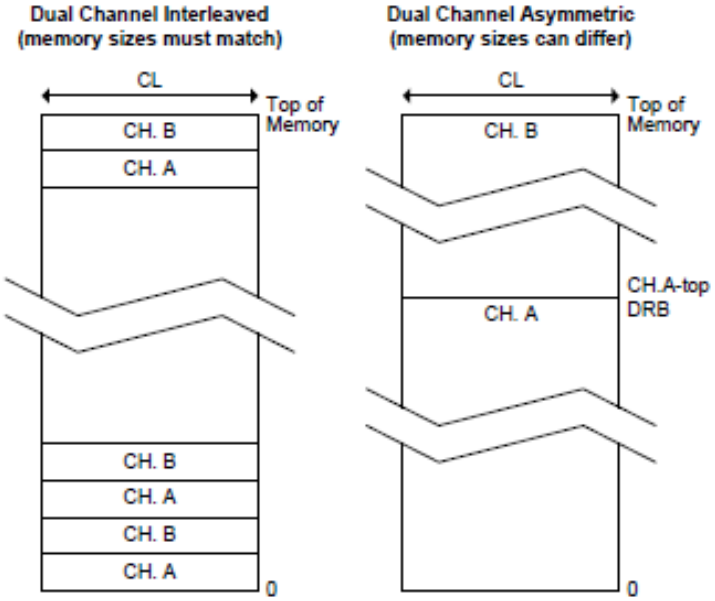
| VID 7 | VID 6 | VID 5 | VID 4 | VID 3 | VID 2 | VID 1 | VID 0 | $V_{CC\_MAX}$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1.04375 |
| 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1.03750 |
| 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1.03125 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1.02500 |
| 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1.01875 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1.01250 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1.00625 |
| 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1.00000 |
| . | | | | | | | | |
| . | | | | | | | | |
| . | | | | | | | | |
| 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0.51875 |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0.51250 |
| 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0.50625 |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0.50000 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | OFF |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | OFF |

- Comanda memoriei
  - IMC (Integrated Memory Controller) implementează protocoale specifice pentru 2 canale de memorie, pe câte 64 biți, având fiecare 1 – 2 module DIMM; tipul de memorie care se poate conecta depinde de setul de circuite;
  - Exemple de module DIMM ce pot fi conectate:

| Tip placă | Capacitate [GO] | Organizare | Nr. circuite | Nr. linii de adresă pe rând/col. |
|---|---|---|---|---|
| A | 1 | 128 M x 8 | 8 | 14/10 |
| B | 2 | 128 M x 8 | 16 | 14/10 |
| B | 4 | 256 M x 8 | 16 | 15/10 |
| C | 0,5 | 64 M x 16 | 4 | 13/10 |

  - IMC poate lucra în 2 moduri: single channel sau dual channel; modul dual channel poate fi simetric sau asimetric;
    - Modul single channel: comenzile sunt generate către un singur canal; se folosește atunci când un canal conține toate circuitele de memorie;

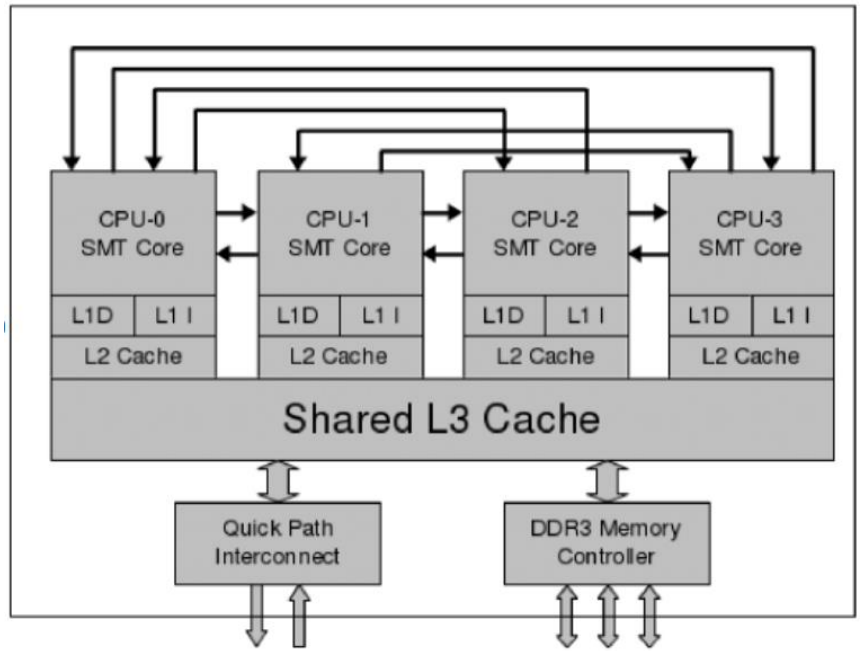
- Modul dual channel simetric: numit și mod cu întrețesere, adresarea se face alternativ la cele 2 canale, după fiecare linie a memoriei cache (limită de 64 octeți); dacă 2 comenzi succesive pot fi generate câte una la fiecare canal, atunci a 2-a comandă nu trebuie să aștepte încheierea primei comenzi, ele se pot suprapune; condiția este ca cele 2 comenzi să se adreseze la canale diferite;
- Modul dual channel asimetric: comenzile succesive se fac la același canal;




- Intel Flex Memory Technology: combină cele 2 moduri dual channel; se poate folosi atunci când cele 2 canale conțin capacități diferite; zona simetrică începe la cea mai mică adresă din fiecare canal și continuă până la începerea zonei asimetrice sau până când se atinge adresa cea mai mare a canalului având capacitatea cea mai mică; apoi accesul se face în mod single channel;

# Proiectarea cu Microprocesoare

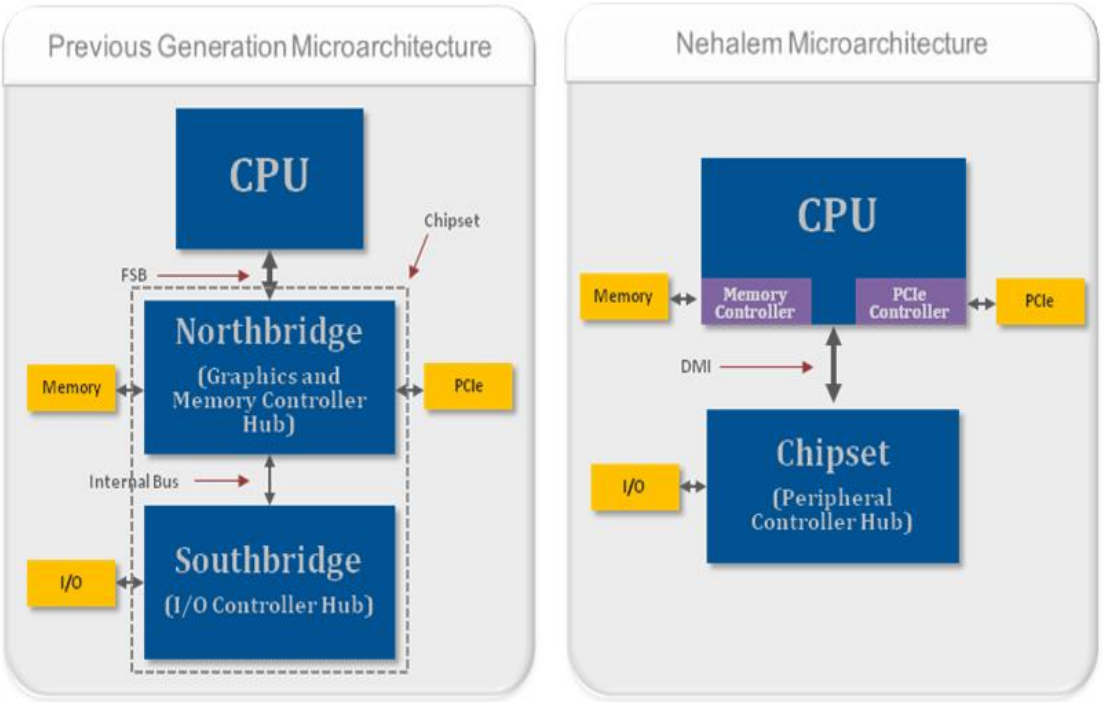
- Microprocesorul i7
  - Procesor performant folosit la servere, desktop-uri și laptop-uri;
  - Procesor cu 4 – 10 nuclee; au HT iar unele modele și TB;
  - Structura:




  - Memorie cache L1, 32 Ko pentru cod și 32 Ko pentru date, L2, 256 Ko și L3 de capacitate 6, 8 sau 12 Mo;
  - Viteză: 2.93 – 3.70 GHz, putere maximă/nucleu: 140 W;

# Proiectarea cu Microprocesoare

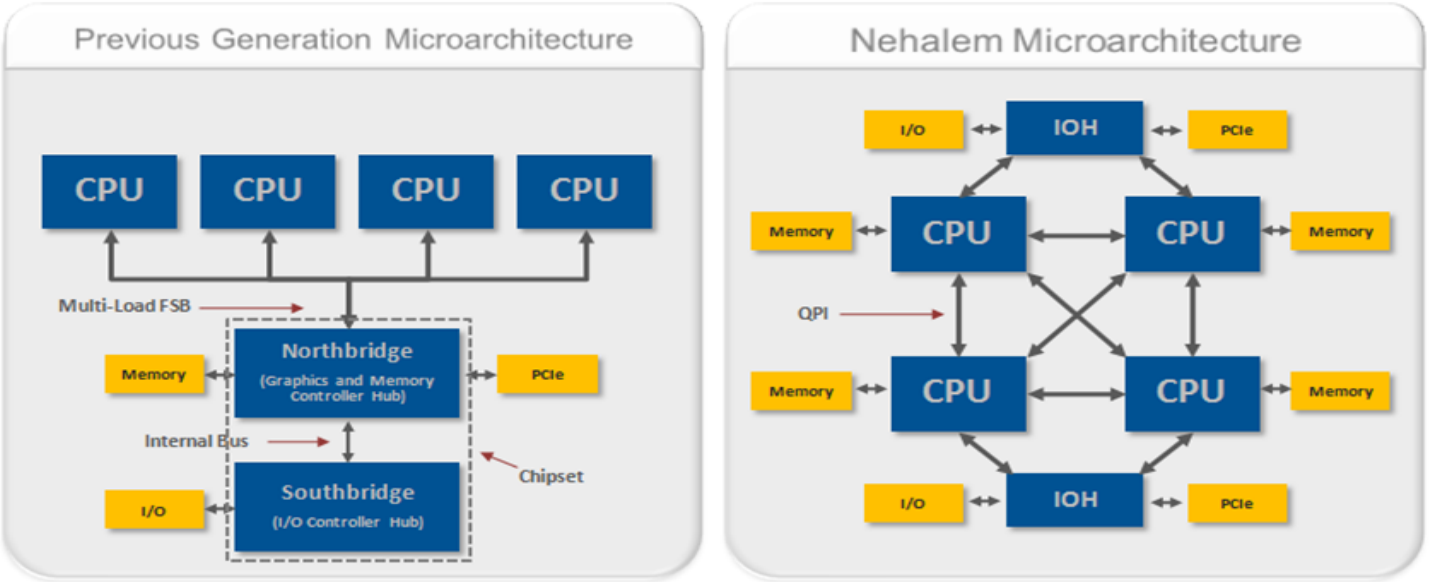
- Îmbunătățiri:
  - Modificarea microarhitecturii:




- În variantele anterioare un acces la memorie cerea parcurgerea traseului: procesor – FSB – Northbridge – controlerul de memorie – memorie; prin includerea controlerului de memorie în procesor scade latența și crește rata de transfer pentru că scade nr. acceselor exterioare la memorie; similar dacă se dorește un transfer prin PCIe;

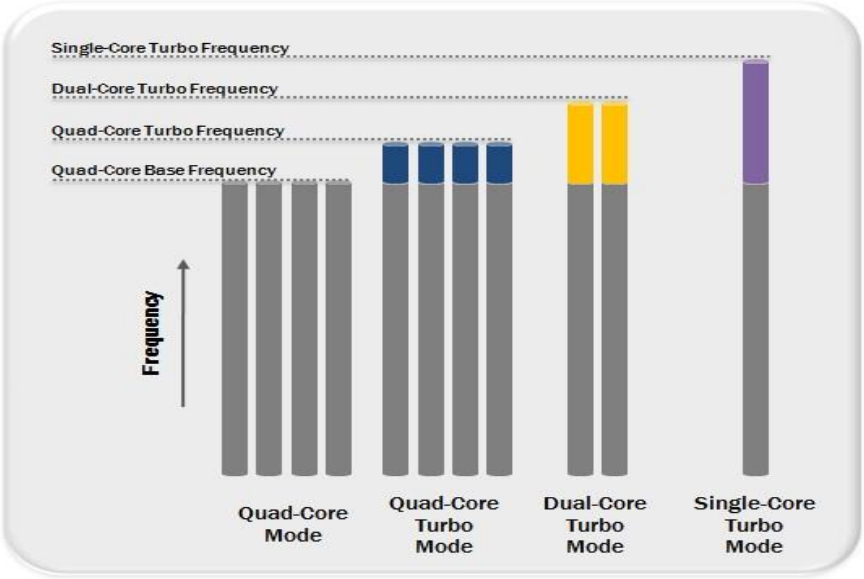
# Proiectarea cu Microprocesoare

- Modificarea rețelei de interconectare:




  - A fost introdusă o rețea de interconectare de tip punct – la – punct, Intel QuickPath Interconnect (QPI) și o nouă structură a memoriei, cea distribuită; QPI asigură o rată de transfer de 25.6 GO/s;
  - În variantele anterioare nucleele erau legate la o memorie comună prin o magistrală comună; în noua variantă, fiecare nucelu are propria memorie, se pot executa accese simultane, nuceleele pot comunica între ele mai rapid (pot avea loc comunicări simultane între perechi de nuclee), iar dacă un nucleu dorește acces la o altă memorie poate face accesul prin nucleul de care aparține respectiva memorie;

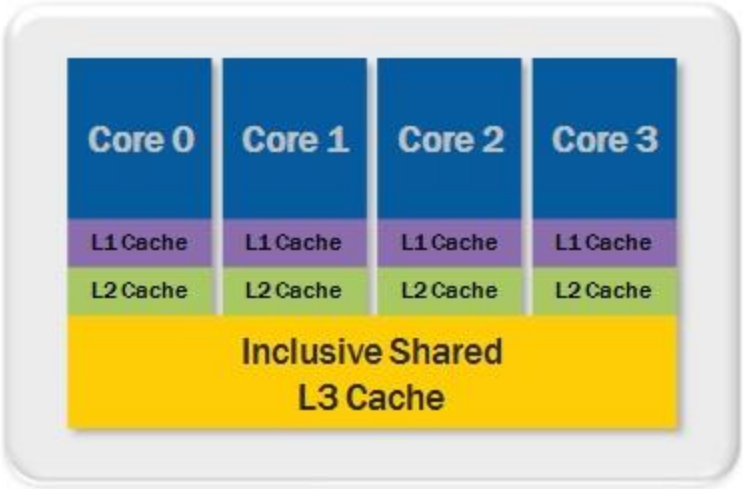
- Overclocking mai performant prin Intel Turbo Boost Technology:
  - Creșterea frecvenței nucleului, peste cea nominală, este activată de OS care cere creșterea performanței; în funcție de numărul nucleelor active, a numărului de thread-uri active, a duratei de timp necesară pentru lucrul cu frecvența mărită, a încărcării nucleelor și a temperaturii lor, facilitatea decide câte nuclee vor intra în turbo boost;




  - Dacă aplicația are 4 thread-uri active, toate cele 4 nuclee vor intra în turbo boost cu o creștere a frecvenței, dacă aplicația are 2 thread-uri active atunci 2 nuclee vor intra în mod low-power iar celelalte 2 nuclee vor intra în turbo boost cu frecvență mai mare ca în cazul anterior iar dacă aplicația va avea doar 1 thread activ atunci doar 1 nucleu va intra în turbo boost cu frecvența mai mare ca în celelalte cazuri;

# Proiectarea cu Microprocesoare

- Reordonarea memoriei cache:
  - Crearea unei memorii cache comune, de nivel L3 și de dimensiune mare (de ex. 12 MO);



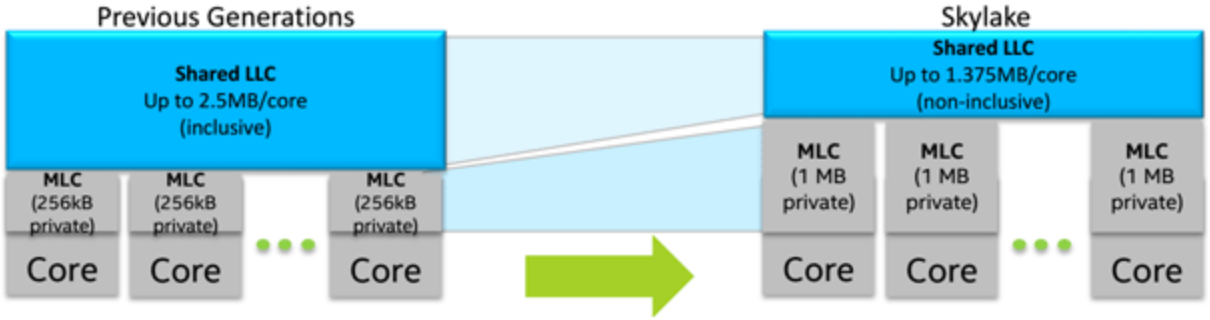

  - Justificare: atunci când un nucleu nu găsește informația dorită în propria memorie cache începe procesul de *cache snooping* prin care caută informația în memoriile cache ale celorlalte nuclee, prin intermediul acestora; dacă nu găsește informația trece la nivelul următor de memorie cache, dacă există, sau execută un ciclu la memoria externă (de evitat); existența unei memorii cache comune va reduce traficul de tip *snooping* iar dacă aceasta are capacitate mare scade riscul de *miss* și, implicit, riscul unul acces la memoria externă;

# Proiectarea cu Microprocesoare

- Microprocesorul i9 – seria Intel Core X
  - Procesorul cel mai performant, folosit la servere, desktop-uri;
  - Procesor cu 10 – 18 nuclee; au HT și TB;
  - Variante:
    - Core i9 Extreme Edition:
      - I9-7980XE: 2,6 GHz, 4,4 GHz TB, 18 nuclee, 36 thread-uri, 1999$;
    - Core i9:
      - I9-7960XE: 2,8 GHz, 4,4 GHz TB, 16 nuclee, 32 thread-uri, 1699$;
      - I9-7940XE: 3,1 GHz, 4,4 GHz TB, 14 nuclee, 28 thread-uri, 1399$;
  - Comparație:

| | I7-7800X | I7-7820X | I9-7900X |
|---|---|---|---|
| Nuclee/Th. | 6/12 | 8/16 | 10/20 |
| Tact | 3,5 GHz | 3,6 GHz | 3,3 GHz |
| Tact turbo | 4 GHz | 4,3 GHz | 4,3 GHz |
| Tact turbo max | | 4,5 GHz | 4,5 GHz |
| L3 | 8,25 Mo | 11 Mo | 13,75 Mo |
| Canale ptr. memorie | 4 | 4 | 4 |
| PCIe căi de comunicare | 28 | 28 | 44 |
| TDP | 140 W | 140 W | 140 W |
| Preț | 389 $ | 599 $ | 999 $ |

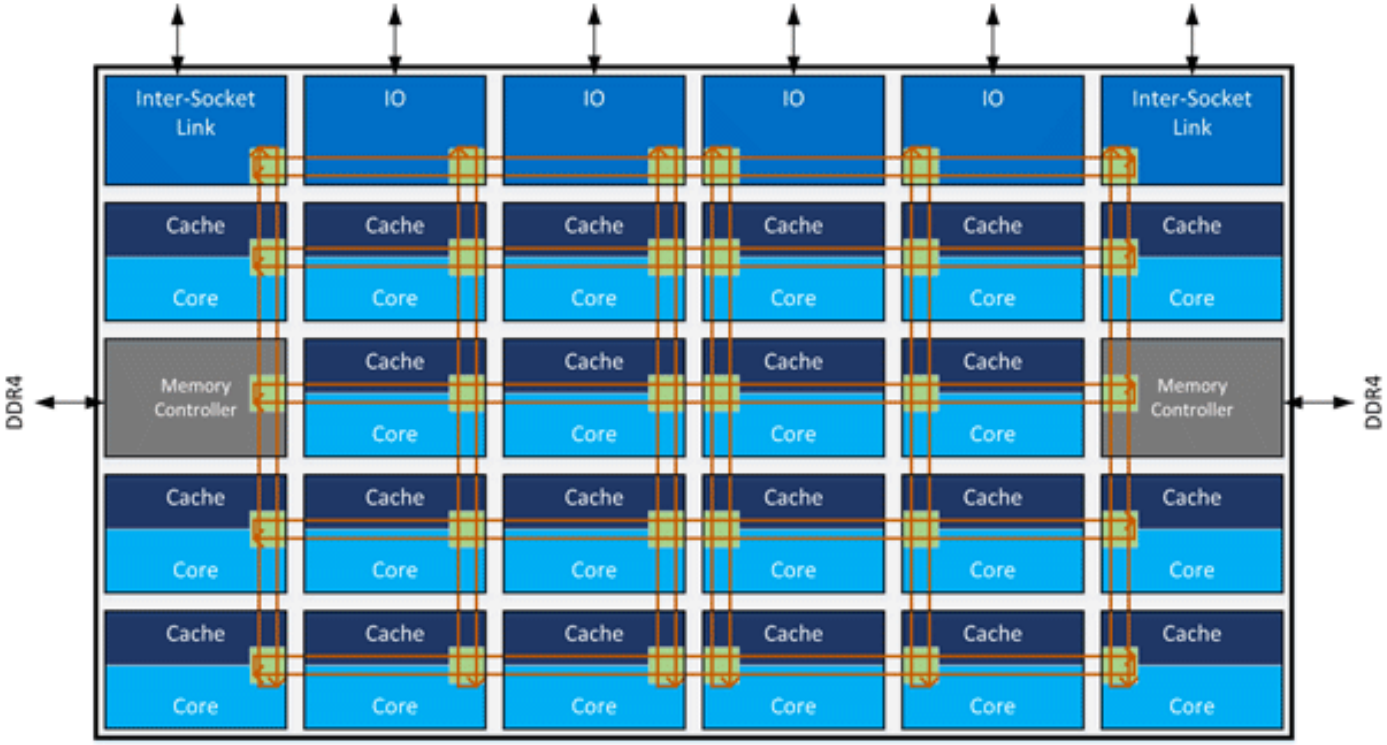
- Îmbunătățiri:
  - Bandă de asamblare cu mai multe nivele și algoritm de predicție mai performant;
  - Reordonarea memoriei cache: prin creșterea memoriei cache locale L2; în acest fel structurile de date vor fi mai apropiate de nucleu;
    - Quadruplarea capacității memoriei cache L2 a redus la jumătate rata de miss, păstrând asociativitatea;




  - Modificarea rețelei de interconectare: renunțarea la rețeaua de tip inel și utilizarea rețelei de tip mesh:
    - Rețeaua de tip mesh asigură latență mai mică și lățime de bandă mai mare în condițiile unei frecvențe și a unei tensiuni de alimentare mai scăzute;
    - Rețeaua de tip mesh permite inserarea unităților ca I/E, controlere de memorie, interconectări inter-socket într-o manieră modulară și scalabilă ținând seama și de creșterea numărului de nuclee;

# Proiectarea cu Microprocesoare

- Oferă suport pentru aplicații care necesită comunicare intensă între nuclee;
- Impactul diferențelor de latență la accesarea datelor între memoriile cache ale diferitelor nuclee scade; nucleele pot accesa uniform modulele de I/E și controlerul de memorie;




*Mesh architecture conceptual representation. Red lines represent horizontal and vertical wires connecting CPU components; green squares represent switches at intersections. Source: Intel*