

Informatique Générale

Architecture des ordinateurs

Jacques Farré

(d'après Fabrice Huet, Wikipedia ... et bien d'autres)

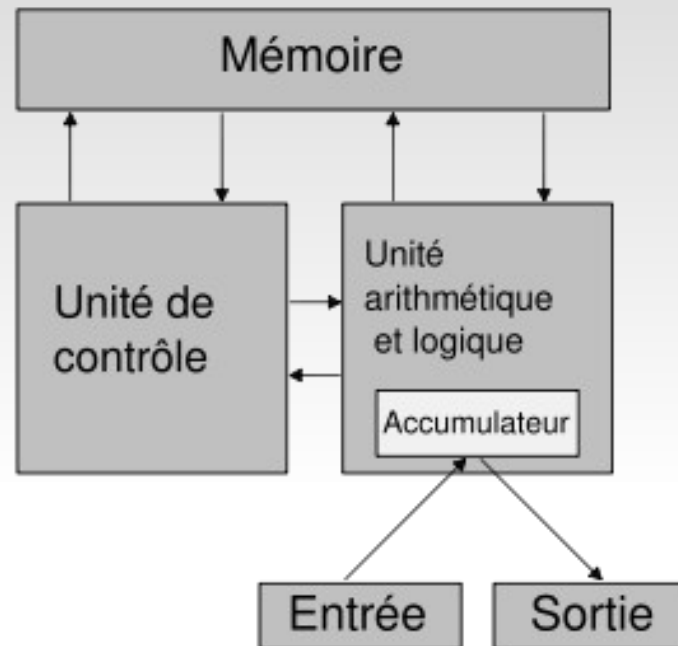
Jacques.Farre@unice.fr

<http://deptinfo.unice.fr/~jf/InfoGene>

Objectifs de cette partie du cours

- Comprendre ce qu'est un ordinateur
 - Composants nécessaires
 - Fonctionnement général
- Stockage d'informations
 - Hiérarchie des mémoires
 - Fonctionnement d'un disque dur
- Le processeur
 - Performance des processeurs
 - Architectures multi cœurs
 - GPGPU (General-purpose computing on graphics processing units)

Rappel sur l'architecture de von Neuman



En réalité plus compliqué, on peut s'en douter

Disséquons la bête pour voir ce qu'elle a dans le ventre

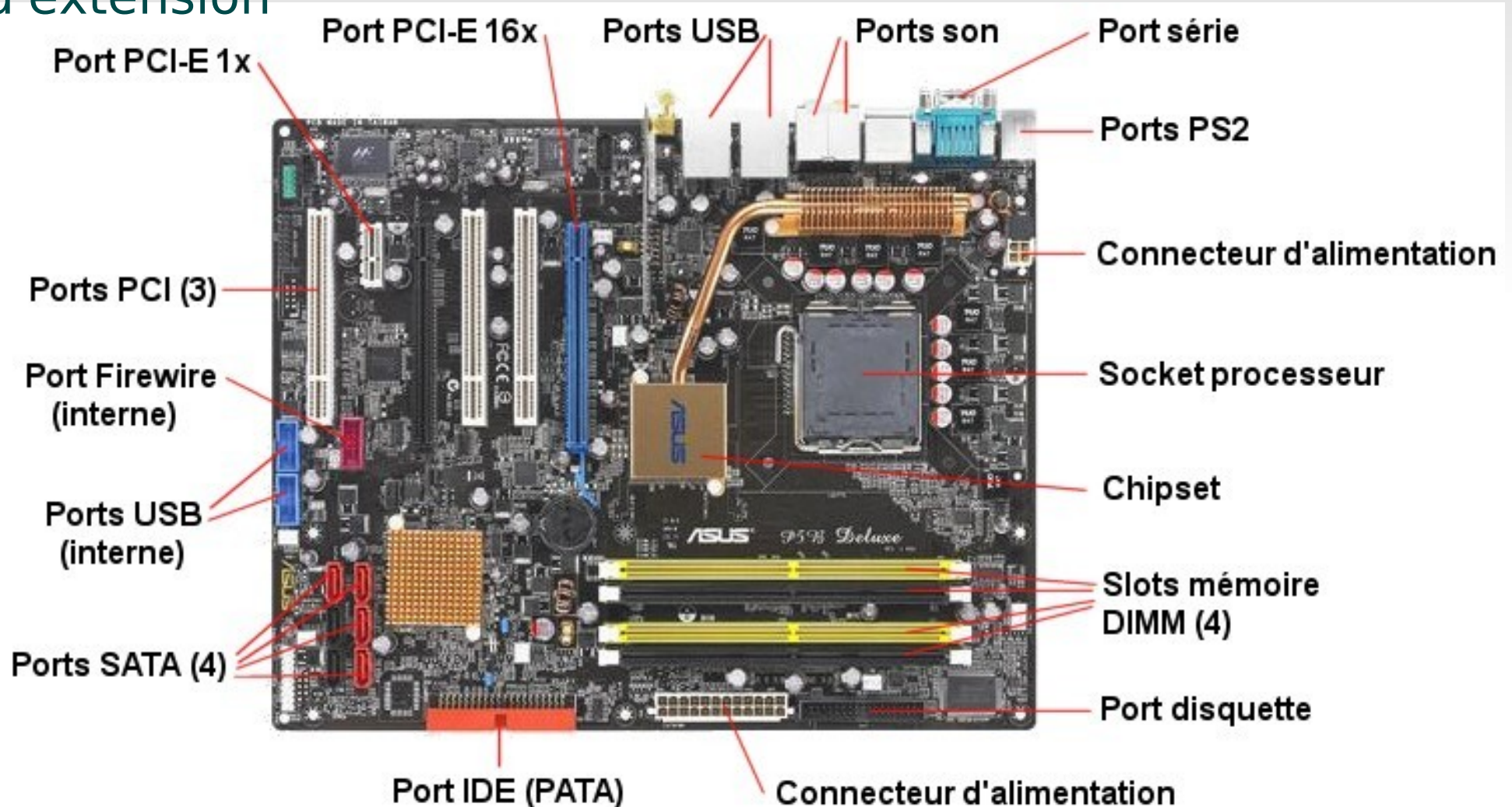
Composants d'un ordinateur

- Processeur (CPU)
- Mémoire (RAM)
- Stockage (disque dur, CD, clé USB, carte mémoire...)
- Périphériques de sortie (carte vidéo, écran, son...)
- Périphériques d'entrée (clavier, souris, capteur...)
- Tous ces éléments sont placés sur une **carte mère** et communiquent par des **ports** via des **bus**



Carte mère

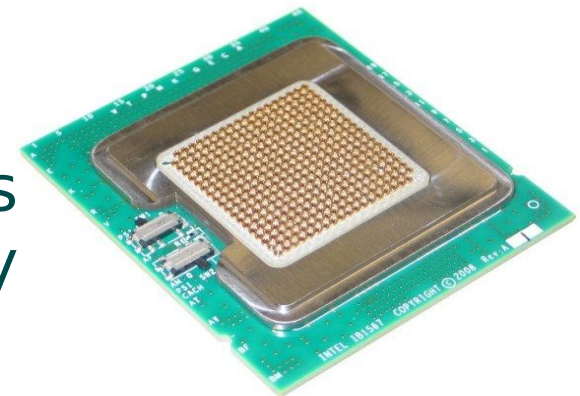
La carte mère permet d'interconnecter tous les composants de l'ordinateur : elle en conditionne les possibilités d'extension



source : <http://www.choixpc.com>

Quoi sert à quoi ? Le socket

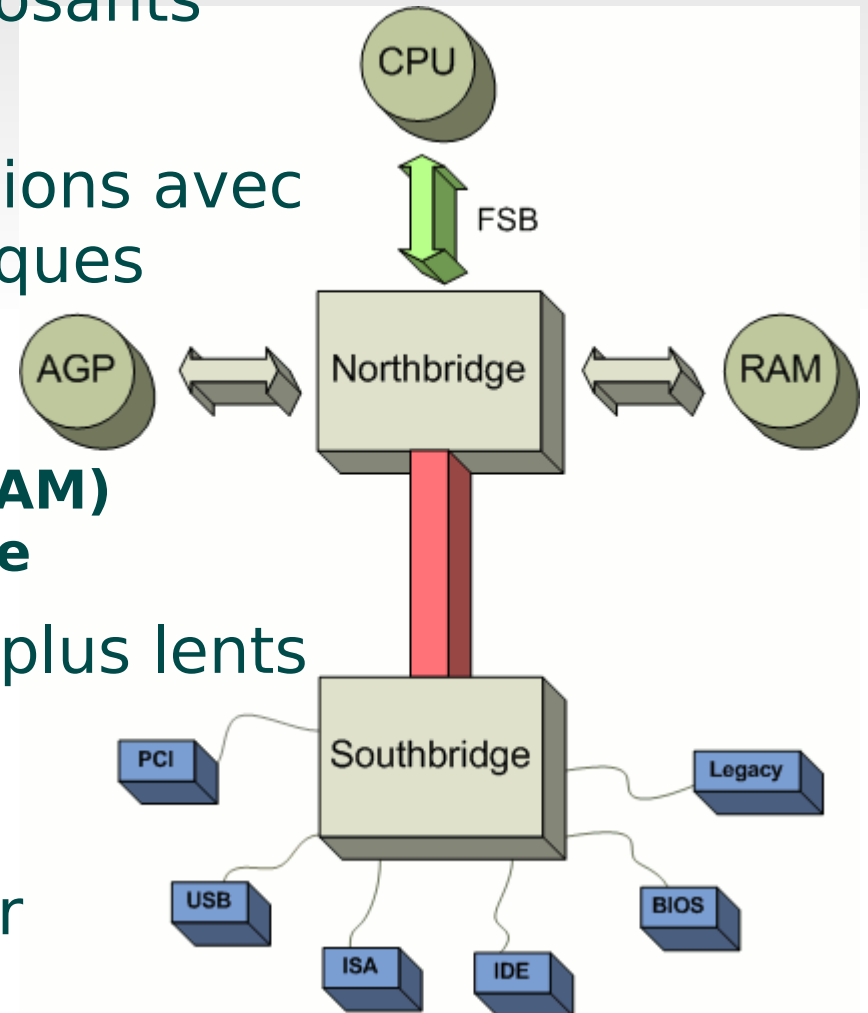
- C'est lui qui recevra le processeur et servira à faire l'interface avec la carte mère
 - permet de changer de processeur si compatibilité ; par exemple
 - S1 (AMD) avec Operon, Athlon, Sempron
 - LGA1156 (Intel) avec core i5 et i7
 - la plupart des sockets permettent d'y insérer les broches du processeur (architecture Pin Grid Array - PGA - matrice de broches)
 - on voit arriver maintenant des sockets à matrice de pastilles (Land Grid Array - LGA) permettant une plus grande densité de contacts



Quoi sert à quoi ? Le chipset

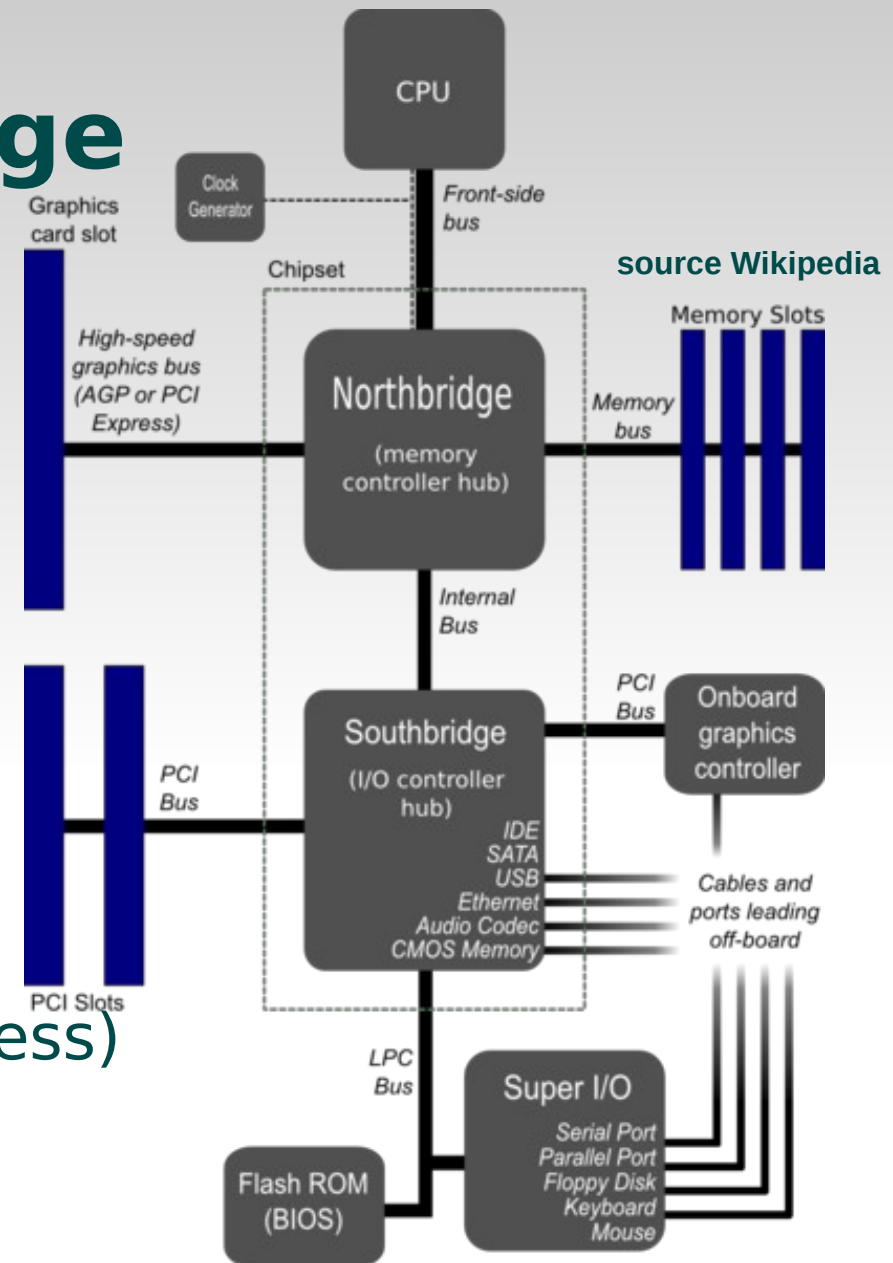
- Il gère les flux de données entre le processeur et les autres composants via le **Front Side Bus** (FSB)
- Le **Northbridge** gère les relations avec la mémoire et les ports graphiques (AGP - Advanced Graphics Port)
 - il détermine le type de processeur et de mémoire (RAM) compatible avec la carte mère
- Le **Southbridge** gère les flux plus lents
- Mais de plus en plus de leurs fonctionnalités tendent à être incorporées dans le processeur (AMD64, Intel core i7...)

source wikipedia



North et South Bridge

- Le north bridge gère les communications entre
 - Le CPU
 - La mémoire (RAM)
 - Le bus AGP
 - Le bus PCIe
- Le south bridge gère
 - Le bus PCI
 - Le DMA (Direct Memory Access)
 - IDE Parallel ATA, Serial ATA
- Exemples de chipsets :
 - Intel : i850 pour Pentium 4
 - Nvidia : nForce 700 pour Intel Core et AMD Phenom



Quoi sert à quoi ? Les slots mémoire

- Destinée à recevoir les barrettes de mémoire, on s'en doute
 - toute barrette n'est pas compatible avec toute carte mère
 - les principaux types utilisés sont Double Data Rate Synchronous Dynamic Random Access Memory : DDR-SDRAM)
 - synchrone avec l'horloge système avec transfert des données deux fois par cycles
 - DDR2-SDRAM : encore assez utilisées
 - DDR3-SDRAM : en train de devenir les plus utilisées ; débit de 6400 Mbits/s jusqu'à 10664 Mbits/s



Quoi sert à quoi ? Les principaux ports

- Port PCI (Peripheral Component Interconnect) : permet de connecter des cartes d'extension sur la carte mère via un bus PCI (cartes réseau, son, graphique) débit théorique 8 Go/s en 32 bits



Ports SATA (Serial Advanced Technology Attachment) pour connecter une mémoire de masse (typiquement un disque dur ou un lecteur de CD/DVD) à la carte mère : 6 Gbits/s pour SATA 3

- Ports série : transmettent des données en série (bit à bit), dont notamment le Universal Serial Bus (USB) : débit 4,8 Gbit/s pour USB 3

Le processeur (CPU)

- C'est le cœur de l'ordinateur
 - Il applique un programme sur des données
 - Un programme est une suite d'instructions en mémoire
 - Les données sont aussi situées en mémoire
- Principe général de l'exécution d'une instruction
 - Lecture de l'instruction en mémoire
 - Lecture des paramètres nécessaires en mémoire
 - Exécution de l'instruction
 - Stockage du résultat en mémoire

Famille de processeurs

- Classification des processeurs par familles
 - qui peuvent exécuter un même ensemble d'instructions de base
 - x86 (dans les PCs, xbox ...)
 - PowerPC (anciens Mac, PS3)
 - ARM (PDA ...)
 - Taille des données qu'ils peuvent manipuler
 - 4 bits, 8, 16, 32, 64....
- Chaque famille comporte énormément de modèles différents

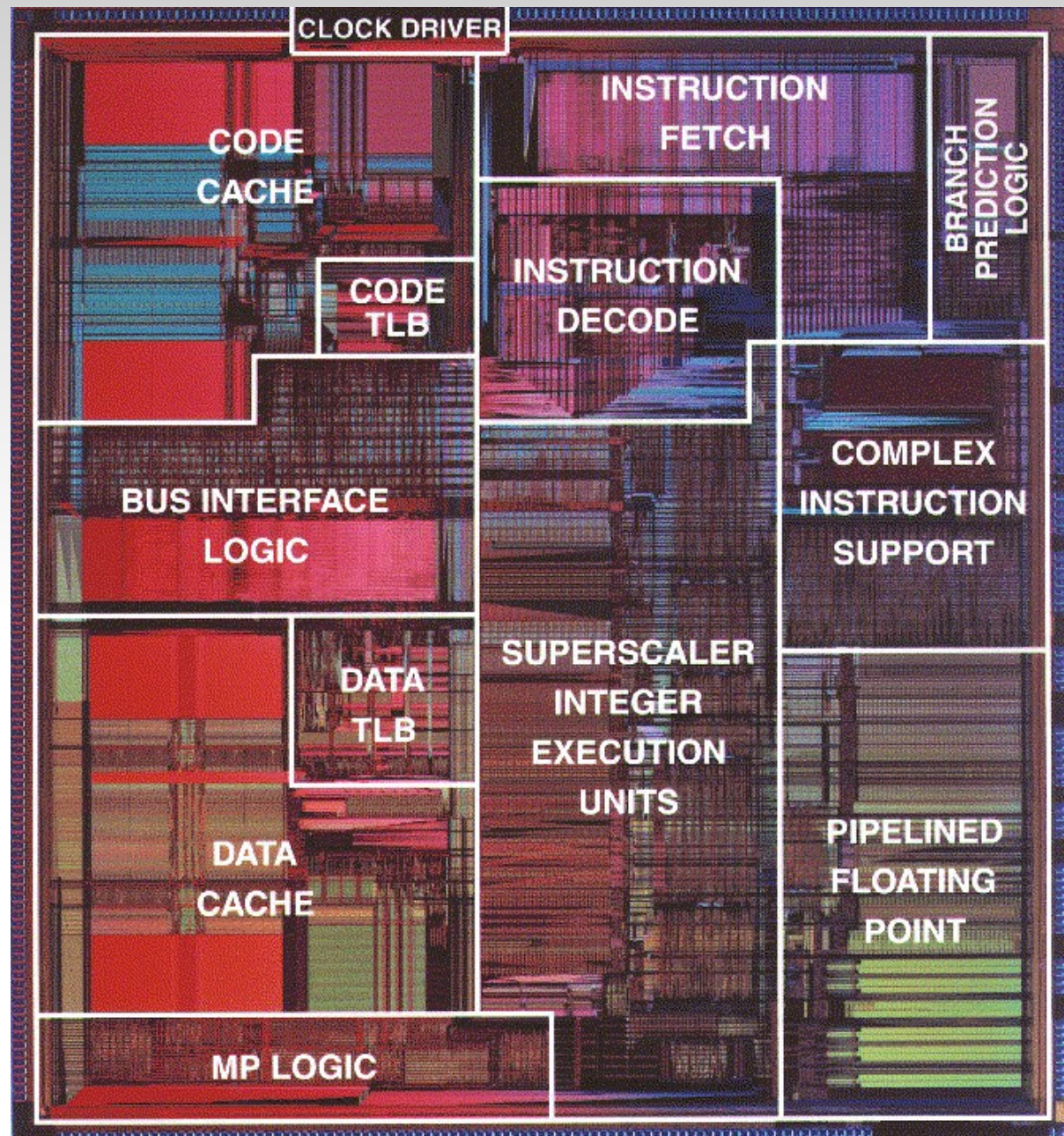
Famille x86 : quelques dates

- Créée en 1978 par Intel (le 8086, 16 bits),
- 1982 : 80286, 16 bits, avec MMU (Memory Management Unit), pour permettre d'adresser un plus grand espace mémoire
- 1985 : 80386, AMD386, 32 bits, avec mémoire paginée
- 1989 : 80486, AMD486, avec pipeline et mémoire cache
- 1993 : Pentium, 64 bits bus de donnée
- 1996 : Cyrix 6x86, superscalaire, compatibles Pentium
- 1999-2000 : AMD Athlon, Pentium 4, hyper-threading
- 2003-2004 : Athlon 64, Opteron, instructions 64 bits
- 2007 : AMD Phenom, 4 core, unité flottante de 128 bits ...
- Maintenant : Intel Core i7, Intel Atom (économique) ... et bientôt (nom de code) Ivy Bridge (transistors 3D)

Composants d'un microprocesseur (relativement) moderne

TLB (translation lookaside buffer) : transformation d'adresses mémoire virtuelles en adresses réelles

MP Logic : multiprocessor logic



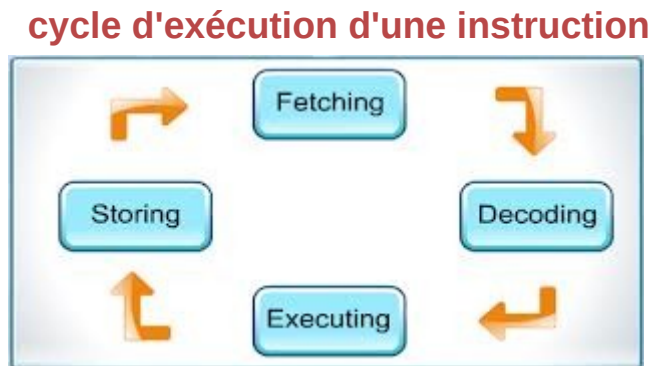
Intel Pentium

Contenu d'un processeur

- Un processeur est composé de transistors
 - Sorte d'interrupteurs, ouverts ou fermés
 - Ces transistors sont gravés sur une pastille de silicium
- Finesse de gravure \approx taille d'un transistor
 - Actuellement 22 nm ($22 \cdot 10^{-9}$ m) au mieux
 - taille d'un cristal de silicium de l'ordre de 10^{-9} m
 - Nombre de transistors par CPU
 - Core 2 Duo E6850 : 265 millions sur 143 mm^2
 - Phenom 9950 : 463 millions sur 288 mm^2

Fréquence du processeur

- Un processeur fonctionne à une certaine fréquence
 - Guidée par une horloge extérieure
 - Exprimée en MHz ou GHz
 - Nombre de basculements de transistors par seconde
 - Core 2 Duo E6850 : 3 GHz
 - Phenom 9950 : 2,6 GHz
- A chaque tic d'horloge le CPU peut effectuer du travail
 - Chaque instruction du CPU nécessite un certain nombre de cycles
 - Dépend de la complexité de l'instruction



Les processeurs multicoeurs

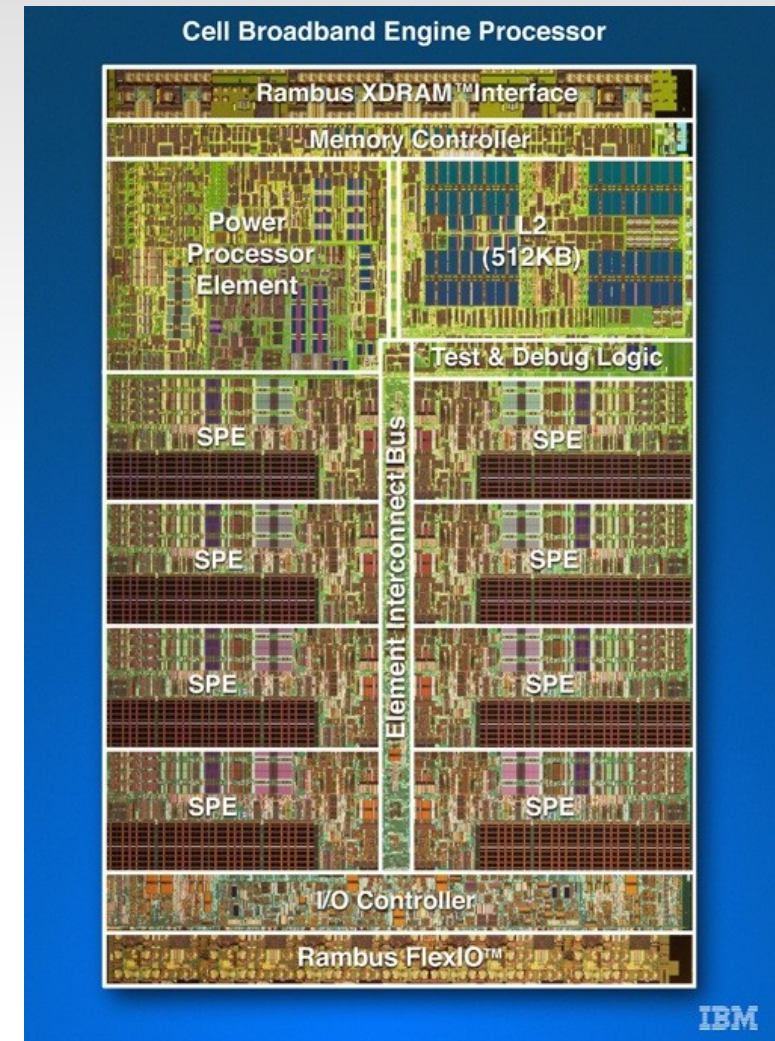
- Possèdent plusieurs unités de calcul qui travaillent en parallèle
 - à l'origine, coeurs homogènes : Power4 (IBM, 2001), puis Opterons (AMD, 2005)
 - permet d'augmenter la puissance de calcul sans augmenter la fréquence d'horloge (et donc de réduire la chaleur dissipée) en traitant en parallèle plusieurs tâches
 - pour bien profiter de ce parallélisme, les programmes doivent être adaptés (et notamment les systèmes d'exploitation)
 - la tendance est vers des coeurs spécialisés

Exemple de processeur multicoeurs spécialisés : le Cell d'IBM

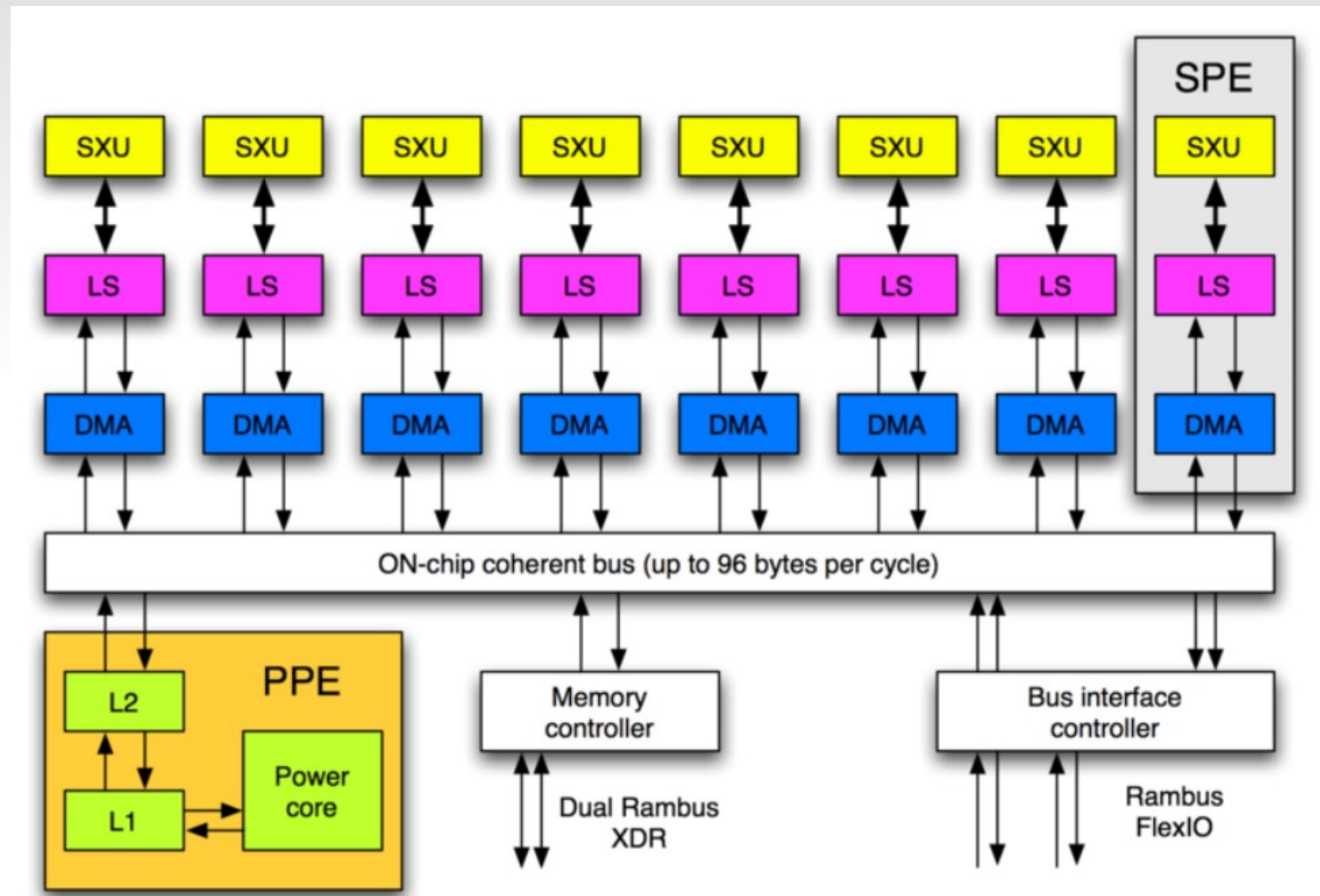
Les SPE (Synergistic Processing Element) sont constitués de 3 unités :

- une unité de calcul (SXU)
- un contrôleur DMA : le Memory Flow Controller (MFC)
- 256Ko de SRAM appelée Local Store (LS)

Ces trois unités forment un total de 21 millions de transistors (7 millions de logique et 14 millions de SRAM)



Le Cell d'IBM



Les super computers

- Apparus très tôt (Seymour Cray, années 60), car il y avait déjà besoin de beaucoup de puissance de calcul (météo, simulation d'aérodynamisme, de résistance des matériaux, d'explosion nucléaire ...)
 - processeurs vectoriels : une même instruction appliquée en parallèle sur plusieurs données (adapté à la résolution de systèmes linéaires)
- Actuellement, plutôt des grappes de milliers de processeurs multicoeurs
 - le plus rapide (juin 2013, <http://www.top500.org>) : Tianhe-2
 - plus de 3 millions de coeurs (Xeon)
 - 1 million de Goctets de mémoire, sous Linux
 - près de 34 petaflops (16.10^{15} opérations flottantes / seconde)
 - consommation de près de 18 Mw (comme une ville de 20000 habitants), occupe 300m²
 - Coût : près de 400 millions US dollars



La mémoire

- La mémoire est un ensemble de composants servant au stockage temporaire d'informations
 - Mémoire présente sur la carte mère (ROM, non volatile)
 - Accès aléatoire (Random Access Memory – RAM, volatile)
 - DRAM basée sur un transistor à effet de champ :
 - temps d'accès de l'ordre de 10ns
 - les informations ne sont conservées que pendant quelques millisecondes (évidemment, contenu perdu quand plus de courant)
 - le *contrôleur mémoire* lit régulièrement chaque cellule puis y réécrit l'information stockée (« rafraîchissement »)
 - type de mémoire très utilisé car bon marché (ne nécessite que peu de silicium)
 - Plusieurs types de mémoire non volatiles encore à l'état de recherches (FRAM, MRAM, NRAM, PRAM, RRAM)

Adresse mémoire

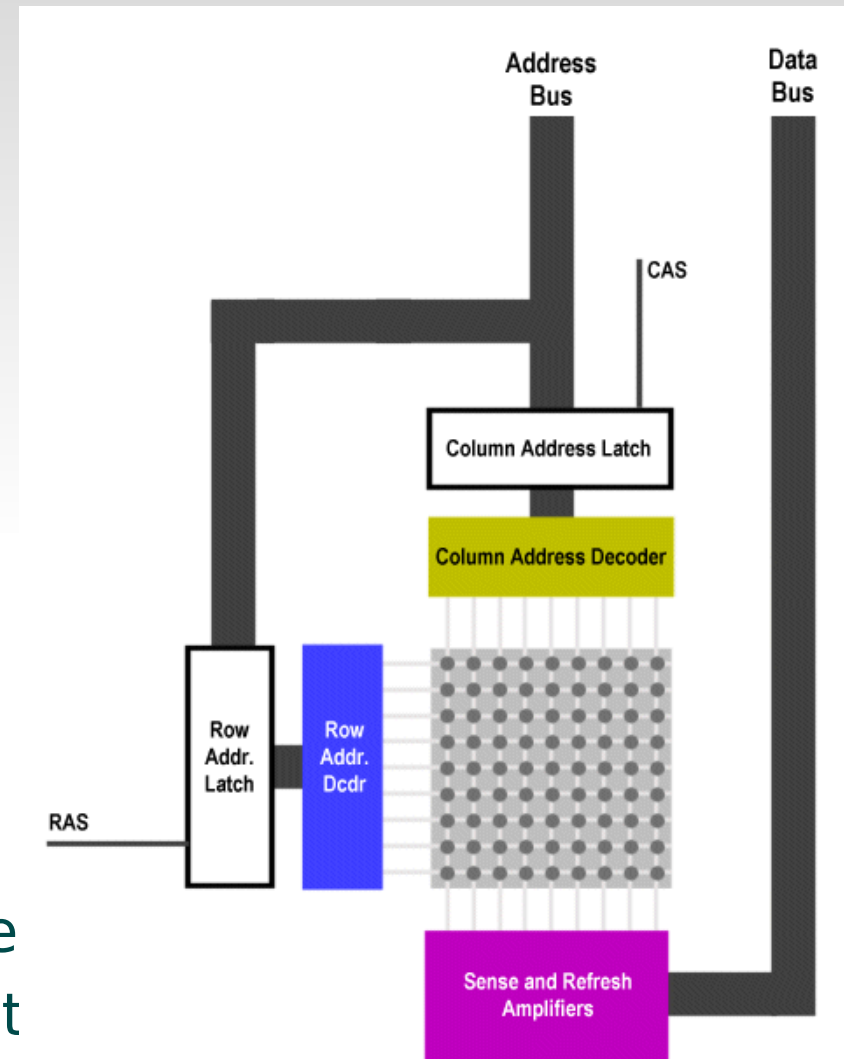
- La mémoire est divisée en un ensemble de cases ou octets (8 bits)
 - Chaque case est accessible indépendamment des autres soit en lecture, soit en écriture
- À chaque case mémoire est associée une **adresse** unique
 - c'est un entier codé sur 2^n bits, n dépendant du processeur (8, 16, 32 bits...)
 - Un processeur de bus d'adresse de 32 bits manipulera donc des adresses mémoire de 32 bits (mais peut avoir des données sur 64, par ex. i586)
 - Détermine la taille mémoire maximale adressable par le processeur
 - Processeur 32 bits, 2^{32} valeurs possibles, donc 4Go max

Lecture mémoire DRAM

- Une case mémoire est définie par un numéro de ligne et de colonne
 - ligne et colonne extraits de l'adresse mémoire
 - par exemple, sur 16 bits

ligne	colonne
1 octet	1 octet

- Pour lire en mémoire, il faut
 - Indiquer l'adresse mémoire de la case à lire sur le bus d'adre
 - Lire les données quand elles sont disponibles et les mettre sur le bu...



Latence mémoire et fréquence

- Il y a donc 2 sources d'attente
 - Indiquer l'adresse de la case (CAS/RAS) : **cycle time**
 - Lire le contenu : **latency time**
- Limite le nombre d'accès mémoire possibles par seconde
 - Notion de fréquence mémoire
 - De la mémoire à 100 Mhz peut lire/écrire 100 millions de cases par seconde (presque 100Mo)
 - Soit 40 secondes pour lire 4Go !
- Bande passante : quantité de mémoire (en bits ou en octets) lisible en 1 seconde

Lire/écrire plus

- Comment avoir de meilleures performances mémoire ?
 - Diminuer la latence en ayant des condensateurs plus rapides
 - Ça coûte cher et il y a des limites physiques
 - Lire/écrire plus qu'un octet à la fois
 - On a vu que les processeurs manipulaient des données de 8-16-32-64 bits (largeur de bus de données)
 - Permettre plusieurs lectures pendant un même cycle (pumping)
 - Autoriser la lecture de plusieurs adresses mémoire en une seule fois

DDR-SDRAM

- Double Data Rate Synchronous Dynamic Random Access Memory
 - Combine plusieurs techniques pour atteindre une bande passante élevée
 - Double Data Rate : possible de transférer 2x plus de données par cycle
 - Existe en 3 générations : DDR1, DDR2 et DDR3
 - Nom commercial : PC[generation]-BandePassante
 - Exemple : PC3-6400 \Rightarrow DDR3, 6400Mo/s
 - Exemples de performances en Mo/s si mémoires cadencées à 100 Mhz
 - DDR1 : 1600 (PC-1600)
 - DDR2 : 3200 (PC2-3200)
 - DDR3 : 6400 (PC3-6400)

Communication CPU-Mémoire : le FSB

- Résumons : Le CPU opère à une certaine fréquence, la mémoire opère à une autre fréquence
 - Par exemple, CPU à 2Ghz, mémoire à 100Mhz, c'est le drame...
- Pour lire/écrire en mémoire le CPU passe par le **Front Side Bus (FSB)**, caractérisé
 - par une fréquence (en Mhz)
 - une largeur de bus (en bits)
 - fréquence et largeur du FSB déterminent combien de bits sont transférables par cycle (bande passante)
 - un bus d'une largeur de 16 bits, cadencé à une fréquence de 133 MHz possède donc un débit égal à $16 \times 133 \times 10^6$ bit/s
 $= 2128 \times 10^6$ bit/s $= 266 \times 10^6$ octets/s $= 266$ Mo/s

Communication CPU-Mémoire : coefficient multiplicateur

- Fréquence des échanges
 - fréquence d'horloge (fréquence de base en gros) ?
 - ou fréquence de fonctionnement, qui n'est pas la même que la fréquence d'horloge ?
 - par exemple, les processeurs Intel utilisent un bus *quad pumped*, qui donne une fréquence de fonctionnement quatre fois supérieure à la fréquence d'horloge du FSB
- Le FSB détermine la fréquence du CPU !
 - par un coefficient multiplicateur :
 - fréquence CPU = fréquence FSB * coefficient
 - fréquence du FSB et coefficient sont des caractéristiques du CPU : par exemple Intel Core 2 : FSB max = 0,8Ghz, coeff 4, donc CPU 3,2Ghz max
- Il ne sert à rien d'avoir un processeur rapide si la mémoire ne suit pas

Bus de communication

- Un bus est un mécanisme qui permet le transfert de données entre éléments dans l'ordinateur
 - Exemple déjà vu : bus mémoire, composé d'un bus d'adresse et d'un bus de données
- Bus interne : connecte les composants internes de l'ordinateur (CPU, mémoire, graphique...)
- Bus externe : relie les périphériques externes
- Bus série : les données sont transmises les une à la suite des autres
- Bus parallèle : plusieurs données sont transmises en parallèle

Exemples de bus dans un PC

- Bus internes
 - Peripheral Component Interconnect (PCI) : parallèle
 - PCI express : série
 - Firewire : série
- Bus externes
 - Serial Advanced Technology Attachment (SATA) : série
 - Small Computer System Interface (SCSI) : parallèle
 - Universal Serial Bus (USB) : série
- Un bus série, transmettant les bits un par un, doit être plus rapide qu'un bus parallèle les transmettant 32 par 32 pour obtenir le même débit.
- Les bus parallèles sont limités en cadence par des difficultés techniques et physiques (à des fréquences élevées, production de plus d'interférences électromagnétiques qu'un bus série ce qui perturbe la qualité des signaux électriques transmis)

Cartes d'extension et périphériques

- Cartes d'extension pour ajouter des fonctionnalités au PC
 - Vidéo, son...
 - Se branchent sur un connecteur PCI ou PCIe
- Périphériques
 - Stockage (disque dur, lecteur/graveur DVD) sur IDE (PATA ou SATA)
 - Autres périphériques sur port USB (Joystick, disque externe en USB...)

Cartes video et son, autres périphériques

- Carte video, majoritairement en PCIe
 - Sert à l'affichage sur écran
 - Peut être parfois intégrée directement au chipset
 - Pas indispensable pour les applications courantes (quoi que ...) mais indispensable pour les jeux modernes
 - Effectue les calculs de rendu 3D
 - Extrêmement coûteux en temps de calcul
 - Indispensable pour les jeux modernes
 - 2 fabricants importants : Nvidia et ATI (AMD)
- Carte son, sur port PCI
 - Ajout de fonctionnalités sonores
 - De plus en plus intégrée dans le chipset
- Autres périphériques
 - Connectés en IDE ou SATA

Ce qu'il manque pour faire un PC complet

- Une alimentation
 - Doit fournir assez de puissance pour l'ensemble des périphériques
 - Gros consommateur : carte graphique (30-300W en charge)
- Radiateur/Ventilateur
 - Pour le CPU
 - Certifié pour type de cpu/socket
- Le boîtier
 - Les goûts et les couleurs...