

SYNTHESE : ARCHITECTURE DES ORDINATEURS

Table des matières

Introduction.....	4
Rappel.....	4
Le débit.....	4
Unités	4
Comment l'augmenter ?	4
La latence	4
Qu'est-ce ?.....	4
Bridé	4
Le processeur central (CPU = Central Processing Unit).....	4
Composition	4
Architecture interne d'un cœur de CPU.....	5
Unité de contrôle.....	5
ALU (Arithmetic Logical Unit)	5
Registres	5
FPU (Floating Point unit)	6
Unité vectorielle	6
Mémoire cache L1	6
Mémoire cache L2	6
Entrées-sorties.....	7
Processeurs modernes	7
Exemple de processeur avec 2 cœurs	8
Architecture Big.LITTLE.....	8
Fabrication d'un processeur.....	8
Finesse de gravure.....	8
Le format du processeur (le socket).....	9
Overclocking	9
Problèmes.....	9
L'enveloppe thermique (ou TDP, Thermal Design Power)	9
But du parallélisme.....	9
CPU superscalaire.....	9
L'hyper threading (SMT ou Simultaneous Multithreading)	10
Estimation théorique de la vitesse d'un CPU	10
Jeu d'instruction	10

Attention au marketing	10
La mémoire vive (RAM)	11
A quoi sert la RAM ?	11
Définitions	11
DRAM vs SRAM.....	11
Caractéristiques.....	11
DDR (Double Data Rate) ?	11
Calculer la latence	11
Débit vs Latence	12
Mémoire ECC.....	12
Les disques dur et SSD.....	12
HDD	12
Les plateaux.....	13
Les caractéristiques techniques	13
Différence entre Half Duplex et Full Duplex.....	13
Calcul du temps d'accès	13
IOPS	13
Un HDD n'efface jamais !	13
Qu'est-ce que le SSD ?.....	14
Type de NAND Flash	15
SSD vs HDD	15
RAID	16
Types de RAID.....	16
Carte mère.....	16
Qu'est-ce ?.....	16
BIOS (Basic Input/Output System)	16
Le chipset.....	16
Circuits d'alimentation (VRM)	17
Format ATX.....	17
Le bloc d'alimentation.....	17
Que doit faire l'alimentation ?	17
Calculer l'alimentation d'une config	17
Rendement électrique.....	17
Alimentation modulaire	18
Comment choisir son alimentation ?	18
Le processeur graphique (GPU : Graphics Processing Unit).....	18

Formes de GPU.....	18
Intégré au CPU (peu performant).....	18
Soudé sur la carte mère	18
Sur une carte graphique	19
La carte graphique.....	19
Critères	19
Pour du jeu	19
Pour du calcul	19
Caractéristiques du GPU.....	19
Puissance de calcul théorique	20
La bande passante mémoire	20
Calculs.....	20
Interface et performances.....	21
Consommation (TDP vs TGP).....	21
Les boîtiers	21
Les écrans	21
Caractéristiques.....	21
Différence écran OLED et cristaux liquides	22
Résumé.....	22

Introduction

Rappel

1 Byte = 1 octet = 8 bits

Le débit

Débit (=Bande passante ou Bandwidth)

Unités

bps, b/s ou bit/s bit par seconde

Bps, B/s, Byte/s ou octet/s Byte ou Octet par seconde

Comment l'augmenter ?

- ⇒ Augmentation des **fréquences**
- ⇒ Augmentation du **nombre de canaux**

La latence

Qu'est-ce ?

Un délai dans les communications informatiques.

Le temps nécessaire à un paquet de données pour passer de la source à la destination à travers une connexion. (en ms, µs ou ns)

Parfois appelé **lag** ou **ping**

Bridé

Bridé par :

- ⇒ La vitesse de propagation des électrons (273.000 km/s)
- ⇒ La vitesse de propagation des photons dans une fibre optique

Le processeur central (CPU = Central Processing Unit)

Composition

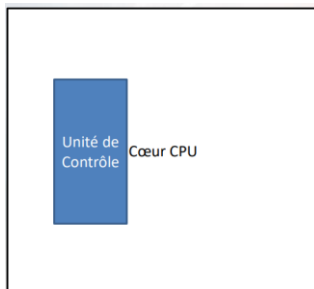
Circuit intégré composé de :

- Transistors (Pour faire les calculs)
- Mémoires (Pour fournir les données aux transistors et stocker le résultat)
- Bus de données (Pour entrer et sortir les données du CPU)

Architecture interne d'un cœur de CPU

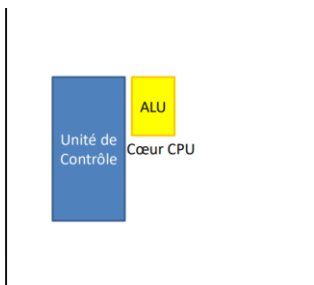
Unité de contrôle

- ⇒ Séquenceur
- ⇒ Synchronise les différents éléments du CPU
- ⇒ Prennent énormément de place
- ⇒ L'amélioration de ces unités est une voie d'optimisation importante.



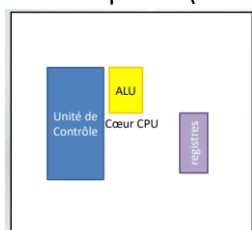
ALU (Arithmetic Logical Unit)

- ⇒ Fonctions basiques de calcul arithmétique et opérations logiques (ET, OU, OU exclusif, ...) sur les nombres entiers



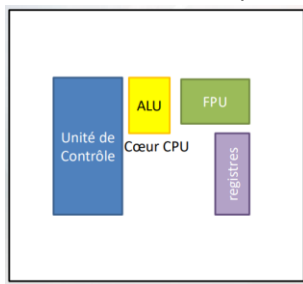
Registres

- ⇒ Emplacement mémoire interne quand le CPU exécute des instructions
- ⇒ Les données y sont temporaires
- ⇒ Mémoire la plus rapide
- ⇒ Cout de fabrication élevé
- ⇒ Faible capacité (moins d'1ko)



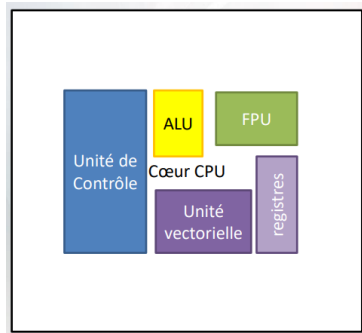
FPU (Floating Point unit)

- ⇒ Unité de calcul flottant effectue les calculs avec des nombres réels
- ⇒ En charge des opérations complexes (trigono, racines, etc.)
- ⇒ Unité fort sollicitée pour le multimédia



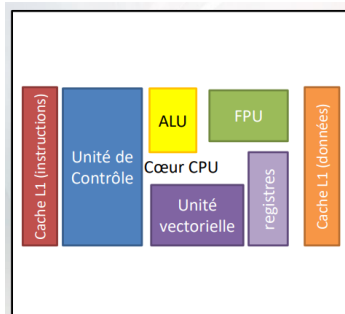
Unité vectorielle

- ⇒ Pour les jeux d'instructions complémentaires (MMX, 3DNow !, SSE, AVX, etc.)
- ⇒ Permettent souvent de traiter plusieurs données en même temps mais en exécutant la même opération.



Mémoire cache L1

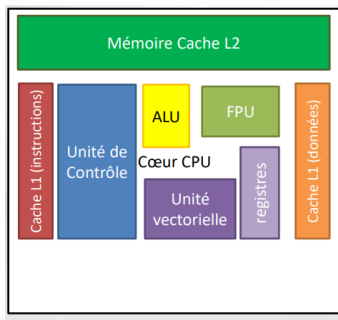
- ⇒ Stocke les prochaines instructions à exécuter
- ⇒ Mémoire très rapide mais plus lente que registres
- ⇒ L1 est généralement scindé en 2 :
 - Les prochaines instructions
 - Les données à utiliser
- ⇒ Exemple de taille : 2x32 Ko



Mémoire cache L2

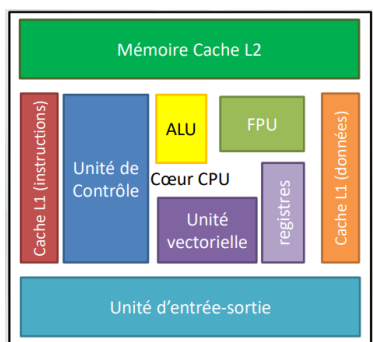
- ⇒ Même principe que la mémoire L1
- ⇒ Les instructions ne sont pas séparées des données
- ⇒ Plus grande quantité (256Ko à 1024Ko par cœur)

⇒ Plus lente que la L1

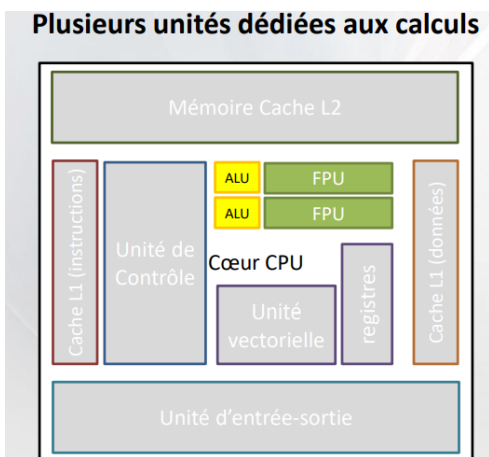


Entrées-sorties

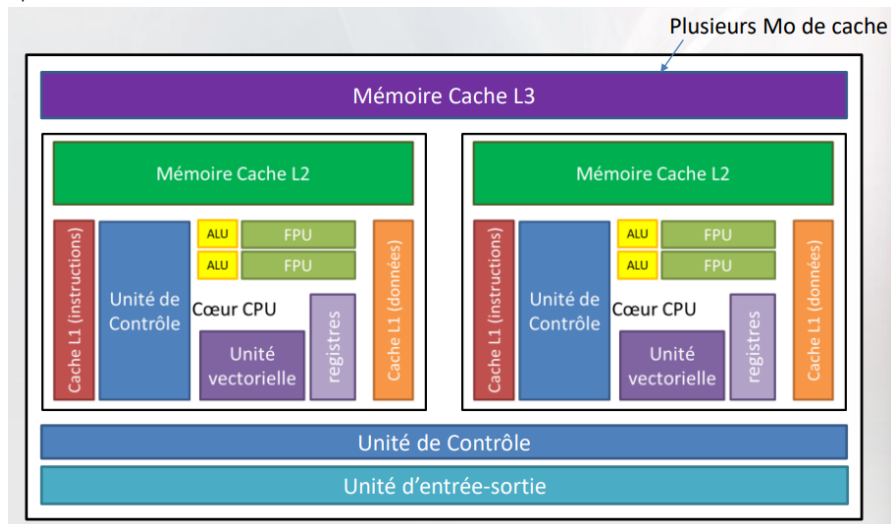
- ⇒ Gère les communications avec l'extérieur du CPU (RAM, CG, etc.)
- ⇒ Unité est de plus en plus grande car les CPU ont de plus en plus d'entrées et de sorties



Processeurs modernes



Exemple de processeur avec 2 cœurs



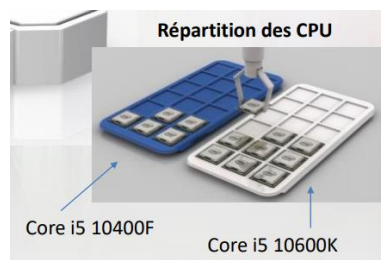
Architecture Big.LITTLE

Mélanger des gros cœurs (performants) avec des petits cœurs (à basse consommation)

(Quand la charge de calcul est faible, il n'y a que les petits cœurs qui fonctionnent)

Fabrication d'un processeur

Encapsulation (Soudure) → Phase de test → Répartition des CPU



Finesse de gravure

Précision avec laquelle les transistors sont gravés dans le silicium

Plus la finesse est élevée (valeur petite) plus les transistors sont petits.

Plus ils sont petits :

- Moins ils consomment
 - Moins ils chauffent
 - Permet d'augmenter la fréquence
- Moins ils prennent de place
 - Réduction des coûts (on en place sur une même surface)
 - Augmente le nombre de fonctionnalités
 - Plus ils chauffent
 - Nécessaire de diminuer la fréquence

Le format du processeur (le socket)

Un processeur au format Socket XYZ requiert une carte mère équipée d'un socket XYZ. (Ce n'est pas le seul critère pour garantir la compatibilité)

Overclocking

Modifier le coefficient multiplicateur et/ou la fréquence d'horloge de base pour augmenter la fréquence de fonctionnement (et donc les performances)

Problèmes

- ⇒ Chauffe plus
- ⇒ Marge de manœuvre de quelques % pour la fréquence de base
- ⇒ Coefficients verrouillés sur les nombreux modèles de CPU
- ⇒ Peut être nécessaire d'augmenter la tension → Chauffe encore plus

L'enveloppe thermique (ou TDP, Thermal Design Power)

Transfert thermique vers l'extérieur dont doit pouvoir bénéficier ce composant pour fonctionner correctement.

$$\text{TDP} = k.V^2.f$$

Plus il a une fréquence élevée, plus il va chauffer.

But du parallélisme

- Augmenter le nombre d'instructions par cycle (= Parallélisme séquentiel)
 - Dédoubler certaines unités de calcul
 - Pipelines
 - Unités vectorielles
- Permettre de traiter plus de tâches en même temps (= Parallélisme des CPU)
 - Multicoeur
 - Multiprocesseur
 - Multithreading simultané

CPU superscalaire

Dédoubler certaines unités de calcul pour pouvoir traiter plusieurs instructions en même temps

```
int a=2, b=3, c=0, d=0, e=0;
```

```
float x=32.5, y=0.0;
```

```
c=a*b;
```

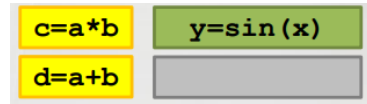
```
y=sin(x);
```

```
d=a+b;
```

```
e=b/a;
```

L'unité de contrôle constate que les calculs sont indépendants et va pouvoir les exécuter

simultanément en fonction des unités disponibles



L'hyper threading (SMT ou Simultaneous Multithreading)

Multiplier virtuellement le nombre de cœurs d'un CPU sans ajouter aucune unité de calcul.

Un CPU avec 4 cœurs et le SMT (de degré 2), va faire croire au système d'exploitation que c'est un CPU à 8 cœurs. On parle alors de CPU 4 cœurs / 8 threads.

Dans la pratique, le SMT (ou l'Hyper Threading) c'est une modification au niveau de l'unité de contrôle qui permet de charger les instructions de 2 programmes (ou tâches) différents en même temps. (-5% à +50% de perf)

Estimation théorique de la vitesse d'un CPU

FLOPS = nombre_{cœurs} * fréquence * (FLOP/cycle)

Cette mesure peut être effectuée sur des nombres flottants de :

8 bits (principalement utile en IA)

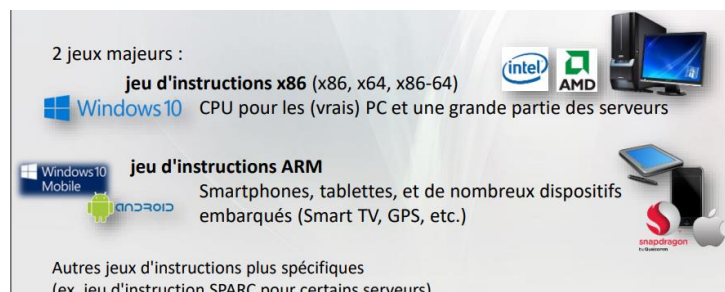
16 bits (principalement utile en IA)

32 bits (taille par défaut)

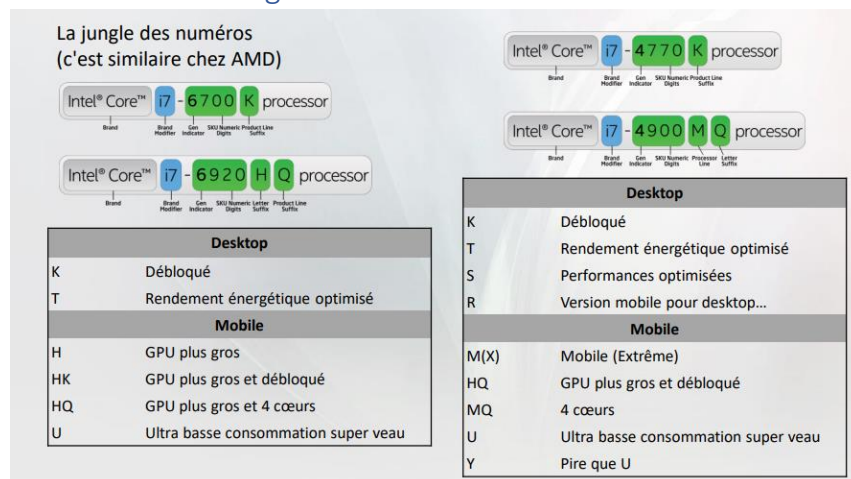
64 bits (TOP 500)

Jeu d'instruction

- Ensemble des instructions machines qu'un processeur peut exécuter
- Assure une compatibilité entre les applications et le matériel



Attention au marketing



La mémoire vive (RAM)

A quoi sert la RAM ?

- ⇒ Contient les données nécessaires au bon fonctionnement du CPU :
 - Données pour les calculs
 - Résultats des calculs
 - Programmes en cours d'exécution

Définitions

Mémoire vive : est sollicitée pendant les calculs du CPU

Mémoire volatile : les données sont perdues une fois l'alimentation électrique coupée

Mémoire RAM : Random Access Memory : L'information peut être accédée directement

DRAM vs SRAM

DRAM – Mémoire vive PC

SRAM – Mémoire cache (ex. CPU)

Caractéristiques

- Capacité (Go)
- Type de mémoire (DDRx)
- Format (DIMM, SO-DIMM)
- Fréquence (MHz)
- Timing (latence)
- Canaux
- Fonctionnalités
- Refroidissement, dimensions et soirée disco

DDR (Double Data Rate) ?

Optimisé pour les latences (CPU)



Calculer la latence

DDR4 4000 MHz CL18

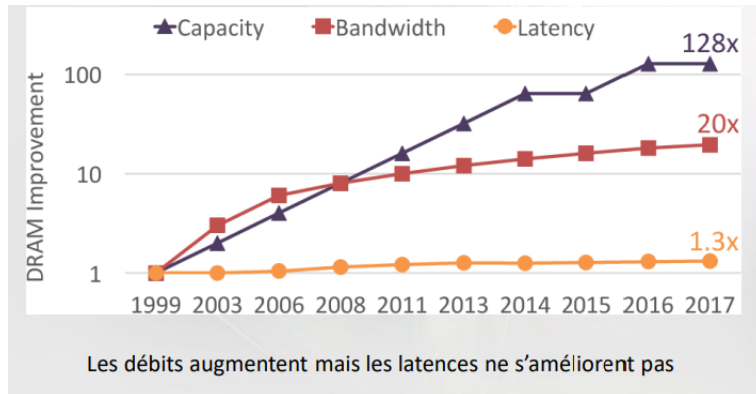
$$\text{Temps de cycle} = 1 / (4.10^9 / 2) = 0,5 \text{ ns}$$

$$t_{CAS} = 0,5 * 18 = 9 \text{ ns}$$

DDR 4 2400 MHz CL16

Temps de cycle = $1 / (2,4 \cdot 10^9 / 2) = 0,833 \text{ ns}$
 $t_{CAS} = 0,833 \cdot 16 = 13,333 \text{ ns}$

Débit vs Latence



Mémoire ECC

Sources d'erreurs en informatique :

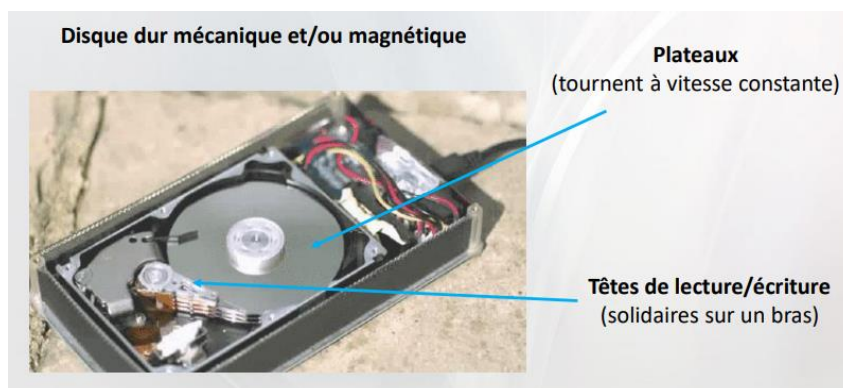
- L'utilisateur
- Les développeurs :
 - o Erreurs logiques
 - o Dépassements de capacité
 - o Précision des floats
- Panne et/ou usure matérielle
- Les bit-flips (interférences, rayons cosmiques, etc.)

Mémoire ECC (Error Correction Code)

Ajoute des bits dédiés à la détection et (éventuellement) correction d'erreur.
 Type de mémoire de prédilection des serveurs (ou grosses workstations)
 (Carte mère et CPU doivent être compatibles)

Les disques dur et SSD

HDD



2 types de format existent :

- Disque de 3,5" (Desktop)
- Disque de 2,5" (Laptop)

Les plateaux

Chaque plateau contiendra des données.

Tous les plateaux tournent ensemble autour d'un axe, à une vitesse de rotation constante

Au centre du plateau, le débit est moins grand qu'à l'extérieur.

A technologie identique, plus la vitesse de rotation est élevée :

- Meilleurs seront les débits (Mo/s)
- Meilleurs seront les latences (ms)
- Moins bonne sera la consommation électrique)

Les caractéristiques techniques

- Capacité (en Go ou To)
- Taille (2,5" ou 3,5")
- Vitesse de rotation (rpm)
- Nombre de plateaux
- Densité : Influence directe sur les débits
 - Radiale en tpi (track per inch)
 - Linéaire en bpi (bit per inch)
 - Surfacique en Bi²
- Technologie d'enregistrement (SMR ou PMR)
- Interfaçage (SATA ou SAS)
- Temps d'accès (ms)
- IOPS
- Mémoire cache (Mo)

Différence entre Half Duplex et Full Duplex

Half Duplex : Permet la communication dans les deux sens mais de manière alternée.

Full Duplex : Autorise la transmission simultanée dans les deux directions (permet de lire et d'écrire dans le disque en même temps)

Calcul du temps d'accès

Temps d'accès réel = temps d'accès moyen + temps de latence

Temps de latence = délai entre le moment où le disque trouve la piste et se synchronise sur les données = temps pour un demi tour = (60/rpm)/2

IOPS

IOPS signifie input/output operations per second

Un HDD n'efface jamais !

Une tête de lecture/écriture ne sait pas démagnétiser une piste.

Le disque contient une forme d'index qui permet de savoir si une donnée est toujours valide ou non. Effacer revient à invalider la donnée, et le disque pourra ainsi réécrire par-dessus si besoin.

Le seul moyen d'effacer un disque c'est de réécrire des données aléatoires sur l'intégralité de sa surface 7 fois.

Qu'est-ce que le SSD ?



SSD : Solid-State Drive

- Alternative au disque dur magnétique
- Disque purement électronique :
 - pas de pièces mécaniques
 - résiste mieux aux chocs et vibrations
 - moins de pannes
 - pas de nuisance sonore
 - faible consommation
- Utilisation massive de mémoire flash
- L'interface SATA est identique
- Prix plus élevé






www.heh.be

18

Type de NAND Flash




Type de NAND Flash




Type de mémoire		Taille d'une cellule	Tarif	Perfs	Durée de vie écritures max
SLC	Single Level Cell	1 bit	€€€€€€€	+++++	~ 100 000
MLC	Multi Level Cell	2 bit*	€€€€€	++++	~ 10 000
TLC	Triple Level Cell	3 bit	€€€	+++	~ 3 000
QLC	Quadruple Level Cell	4 bit	€€	+	~ 1 000

* Il faut vérifier !

Multi Level Cell ne signifie pas littéralement cellule à 2 niveaux ... certains fabricants utilisent cette appellation pour des cellules de plus de 2 bits

Il faut donc bien vérifier la doc... et si rien n'est spécifié... c'est que ce n'est probablement pas du vrai MLC



www.heh.be

19

SSD vs HDD

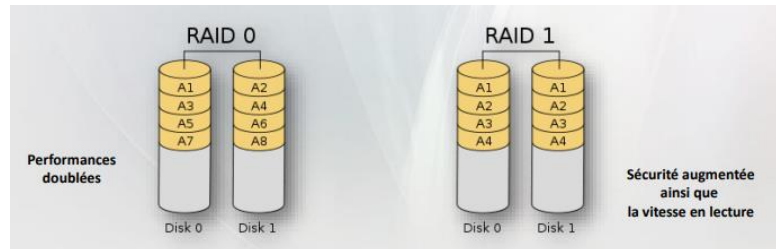
Caractéristique	SSD	Disque mécanique
Temps d'accès aléatoire	Environ 0,1 ms	8-16 ms
Vitesse de lecture/écriture	+ de 500 Mo/s (format disque) + de 7 Go/s (NVMe)	Jusqu'à 300Mo/s
Fragmentation	Aucun effet	Ralentissement de l'utilisation des fichiers
Bruit	Aucun	Variable
Vulnérabilités	Usure de la capacité de stockage	Chocs et vibration, sensibles aux champs magnétiques
Masse	Quelques dizaines de grammes	Jusqu'à près de 700 g
Durée de vie	Bonne pour les MLC 3D, à condition d'utiliser TRIM AFR 1%	MTBF 500.000 heures - 1 500 000 heures (idem SSD) AFR 1%
Temps de rétention	1 à 10 ans	minimum 5 ans
Rapport coût-capacité	<0,1 €/Go (QLC) - 0,2 €/Go (MLC)	~ 0,03 €/Go
Capacité de stockage	Jusqu'à 8 To pour le peuple (~1000 €)	Jusqu'à 18 To (~800 €)
Consommation	0,1 - 0,9 W (veille) jusqu'à 0,9 W (activité)	0,5 à 1,3 W (veille) 2 à 4 W (activité)

En retenir 5.

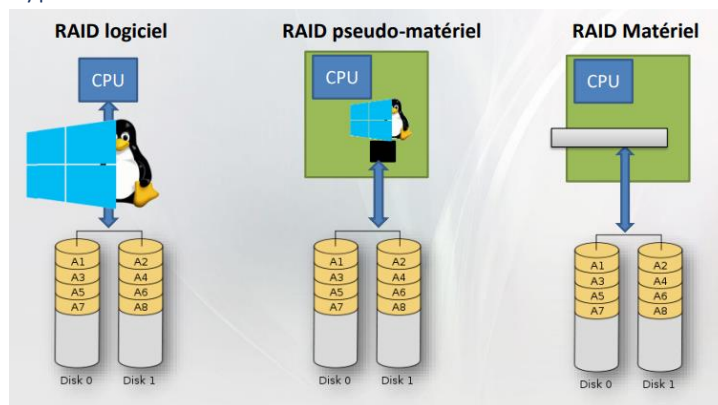
RAID

Assemblage de plusieurs disques physiques pour constituer un seul disque logique.

Différents systèmes (niveaux) RAID pour augmenter les performances et/ou la sécurité des données.



Types de RAID



Carte mère

Qu'est-ce ?

- Motherboard ou Mainboard (mobo dans le jargon)
- Élément crucial d'un ordinateur :
 - o Permet de connecter et de faire communiquer tous les composants
 - o Détermine complètement les possibilités et l'évolutivité
- Ne se remplace pas aussi aisément qu'une barrette de RAM

BIOS (Basic Input/Output System)

Programme stocké dans une ROM sur la carte mère et sert à :

- **Initialiser** tous les composants de la carte mère
- **Identifier** tous les composants qui lui sont connectés
- Initialiser l'**ordre de priorité** des périphériques de stockage
- **Démarrer le système d'exploitation**

A été remplacé par l'UEFI.

Le chipset

- Le chef d'orchestre de la carte mère
- Soudé et souvent affublé d'un dissipateur
- Une grande partie des fonctionnalités d'une carte mère dépend du chipset

Permet ou non :

- L'installation de certains CPU/périphériques
- L'overclocking
- Le RAID pseudo-matériel
- Etc.

Circuits d'alimentation (VRM)

A qualité de composant égale, plus le nombre de phases est élevé :

- Plus les tensions sont stables
- Moins les VRM s'échauffent pour une même puissance totale
- Meilleure est l'efficacité pour des hautes puissances

Format ATX

Le format ATX est le format standard d'une carte mère.

1.1 Choix d'une carte mère

1. Le socket
2. Le facteur d'encombrement (Format)
3. Le chipset

Le bloc d'alimentation

Que doit faire l'alimentation ?

PSU = Power Supply Unit

Composant indispensable et très important car il doit :

- Délivrer des tensions stables
- Pouvoir fournir une puissance élevée

Calculer l'alimentation d'une config

Formule

$$1,5 * (\text{TDP CPU} + \text{TGP GPU} + 100\text{W})$$

Rendement électrique

Au minimum prendre une Gold

Alimentation modulaire



Comment choisir son alimentation ?

1. Calculer la puissance nécessaire :

$$1,5 * (\text{CPU TDP} + \text{GPU TGP} + 100\text{W})$$

2. **Format (ATX, SFX,...) et dimensions de l'alimentation** (Attention aux tailles exotiques)
3. **Nombre de connecteurs** en suffisance (attention aux configs musclées | modulaire ?)
4. **Ventilation et nuisance sonore**

Ventilateurs thermo-régulés, taille de ventilateur, Odb, passif ?

5. **Rendements (Gold mini)**
6. **Marque et prix**

Le processeur graphique (GPU : Graphics Processing Unit)

- Intermédiaire entre ordinateur et écran
- Décharge le plus possible le CPU des calculs d'affichage
- Accélère les calculs graphiques
- **Le composant le plus important pour les performances en 3D**
- Pour les PC, principalement 3 constructeurs : AMD, NVIDIA et Intel
- Également utilisé comme **coprocesseur de calcul** (ex : calcul scientifique)

Formes de GPU

Intégré au CPU (peu performant)

- On parle d'**IGP** (Integrated Graphics Processor) pour les tout petits GPU
- Ou d'**APU** (Accelerated Processing Unit) pour les petits GPU
- Ou de **SoC** (System on Chip) quand le CPU contient également tout le reste (carte son, wifi, gestion des SSD, etc.), comme c'est le cas pour les smartphones et tablettes.

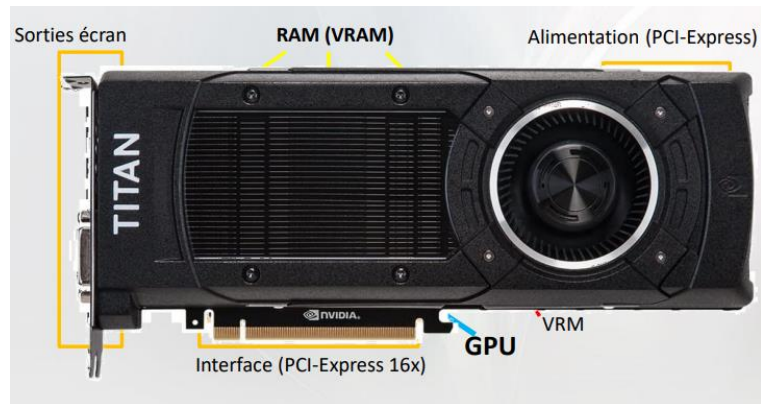
Soudé sur la carte mère

- Portables et certains serveurs

Sur une carte graphique

- Discrete/Dedicated GPU

La carte graphique



Critères

Pour du jeu

1. Le GPU
2. La bande passante mémoire (Go/s)
3. Consommation (TDP vs TGP)
4. Système de refroidissement

Pour du calcul

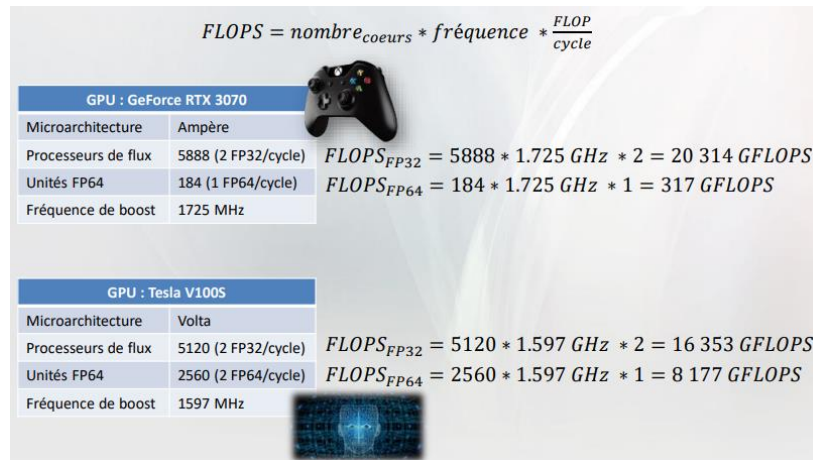
1. Le GPU
2. La bande passante mémoire (Go/s)
3. La quantité de mémoire
4. Interface
5. Support logiciel

Caractéristiques du GPU

Les processeurs de flux effectuent la majorité des calculs.

Puissance de calcul théorique

$FLOPS = \text{nombre}_{\text{cœurs}} * \text{fréquence} * (FLOP/cycle)$

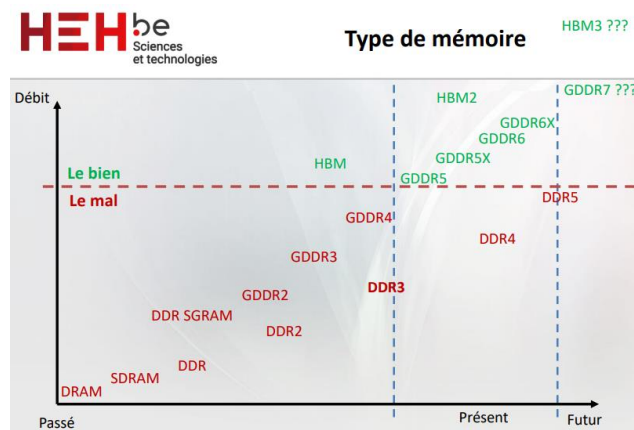


La bande passante mémoire

Les GPU peuvent avoir des milliers de cœurs de calculs, il faut donc pouvoir alimenter ces cœurs en données. ➔ Gros besoins en bande passante mémoire !

Types de mémoire -> Fréquence de fonctionnement

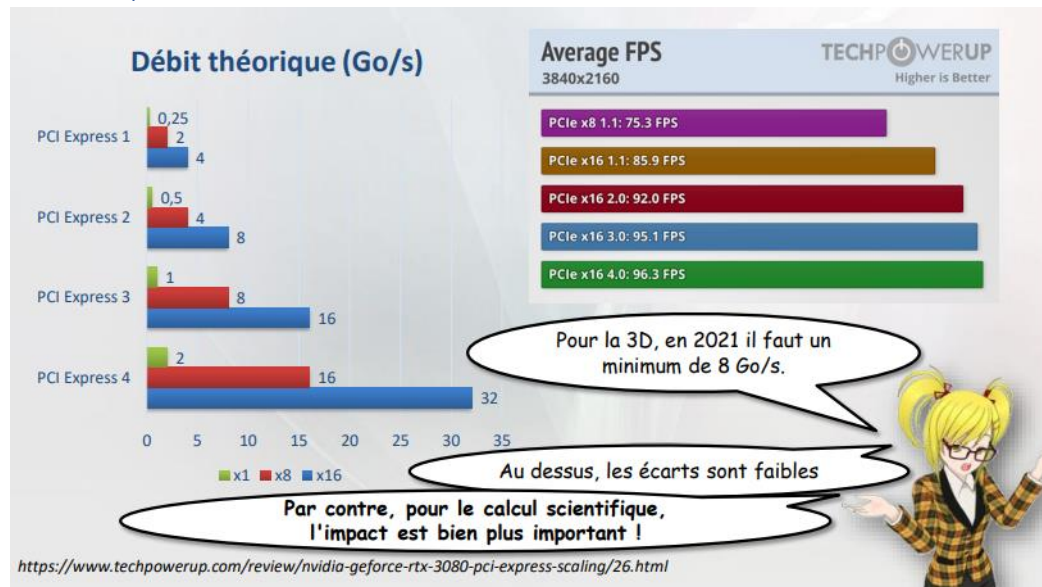
Fréquence de fonctionnement * largeur de bus = **bande passante mémoire**



Calculs

Bande passante en octets par seconde = Fréquence logique * (bus en bits/8)

Interface et performances



Consommation (TDP vs TGP)

TDP (Consommation du GPU)

TGP (Consommation de la carte complète)

TBP = Consommation de la carte complète + système de refroidissement et quelques LED)

Les boîtiers

Une bonne ventilation est primordiale.

Dimensions adéquates (120, 140)

Les écrans

Caractéristiques

Taille (diagonale en ")

HD	= 1280 x 720	= 720p	format 16/9
FullHD	= 1920 x 1080	= 1080p	format 16/9
QHD	= 2560 x 1440	= 1440p	format 16/9
UHD	= 3840 x 2160	= 2160p	format 16/9
4K	= 4096 x 2160	= 2160p ...	format 17/9
8K	= 7680 x 4320	= 4320p	format 16/9

Définition en pixels (L x H)

Luminosité (cd/m²) : Critère important si la luminosité ambiante est élevée

Taux de contraste (ratio) : rapport d'intensité lumineuse entre blanc et noir (attention tricheries)

Fréquence de rafraîchissement

Temps de réponse (attention tricheries)

Type de dalle

Différence écran OLED et cristaux liquides

Les écrans OLED (Organic Light-Emitting Diode) une couche organique qui produit sa propre lumière, le noir est **parfaitement** noir. Contrairement aux cristaux liquides.

Résumé



En résumé

Type de dalle	LCD			OLED	MicroLED
	TN	VA	IPS		
Pixel coupé	Blanc	Noir	Noir	Noir	Noir
Rendu des couleurs	de mauvais à moyen	bon	de bons à très bons	excellents	excellents
	Un peu meilleurs si QLED (et plus cher)				
Contrastes	de mauvais à moyens	de bons à très bons	de moyens à bons	parfaits	parfaits
	Un peu meilleurs si MiniLED (et beaucoup plus cher)				
Temps de réponse	de moyens à très bons	de mauvais à presque bons	de mauvais à bons	excellents	excellents
Angles de vision	mauvais	de moyens à bons	très bons	excellents	excellents
Luminosité	très bonne	très bonne	très bonne	de mauvaise à moyenne	excellente
Burn-in	Non	Non	Non	Oui	Non
Prix	peu cher	moyen	plus cher	cher	Hors de prix