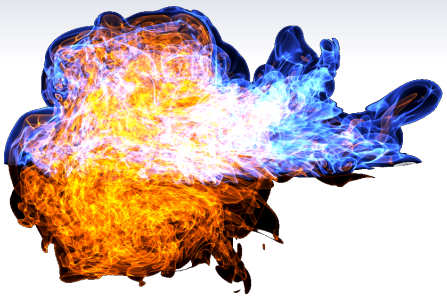


Architecture des ordinateurs

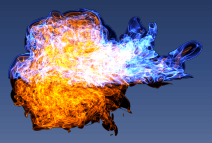
Décoder la relation entre l'architecture et les applications

Violaine Louvet,

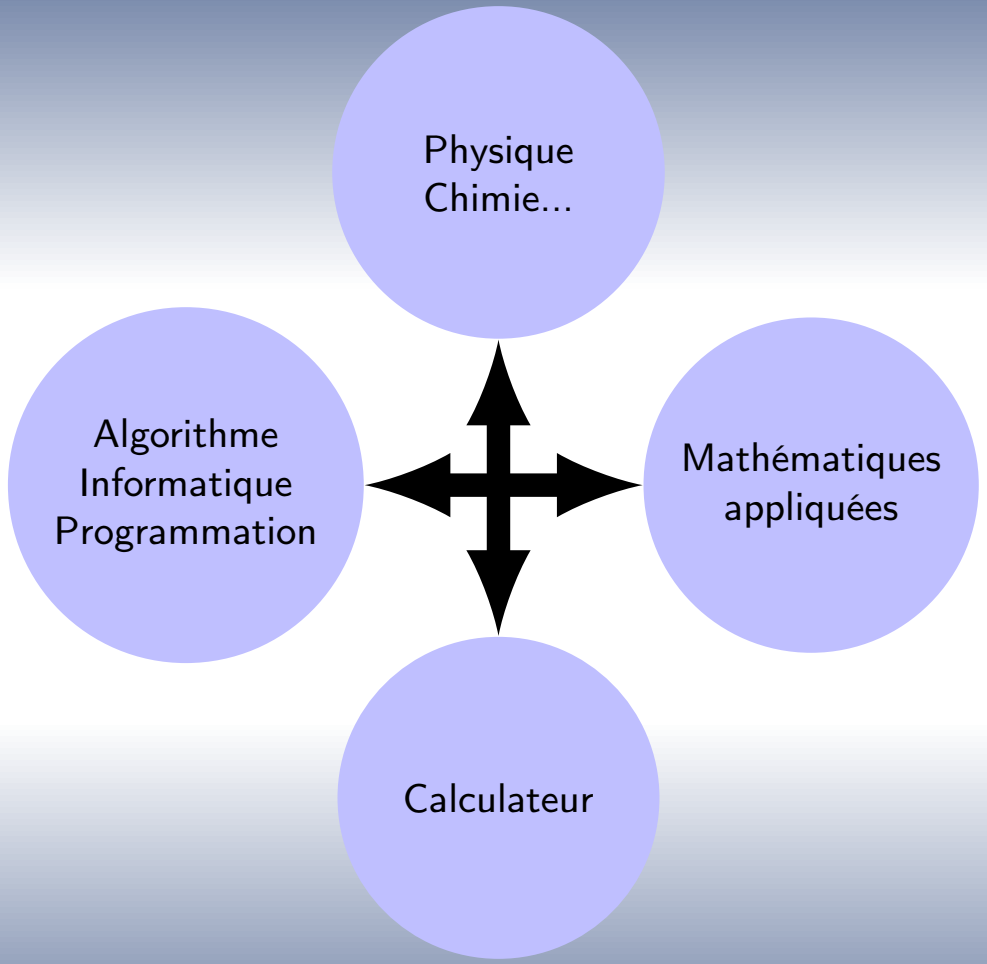
Institut Camille Jordan

CNRS & Université Lyon 1

Ecole « Découverte du Calcul » 2013



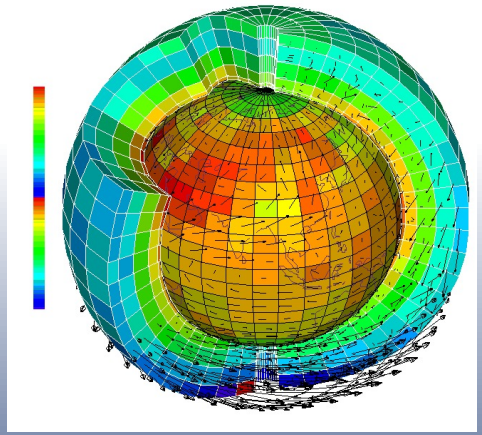
Simulation numérique ...

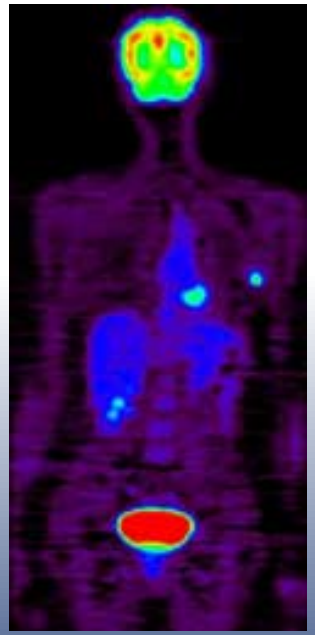


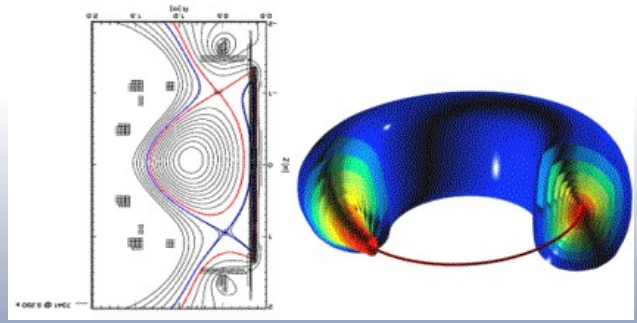
Simulation numérique ...

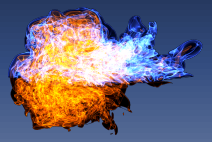
Complexification des modèles, augmentation de la taille et de la précision des domaines de calcul, nouvelles disciplines ... les besoins en simulation numérique et en calculs ne font qu'augmenter :

- Climat
- Médecine
- Sismologie
- Fusion nucléaire ...



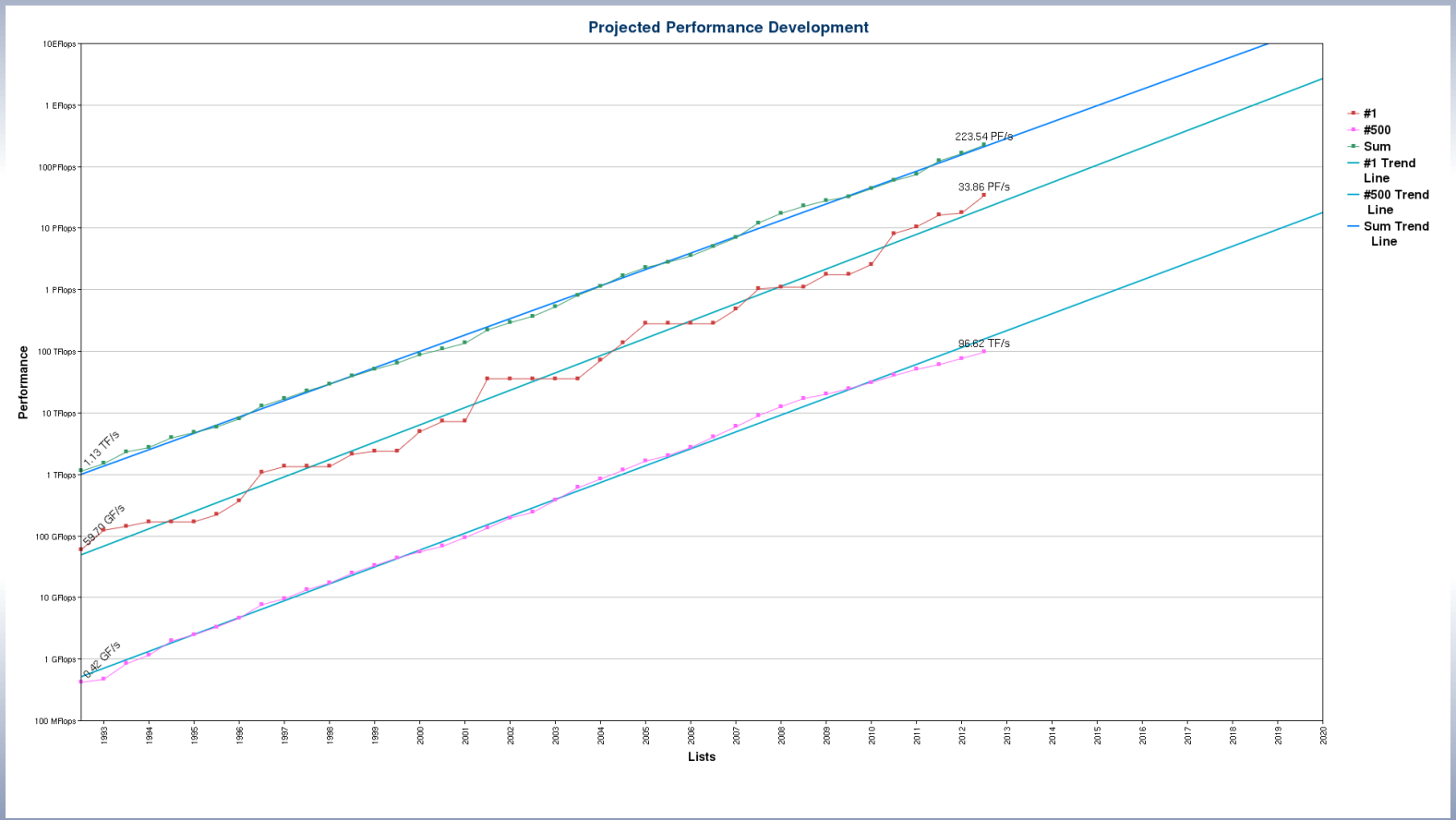






et puissance de calcul ...

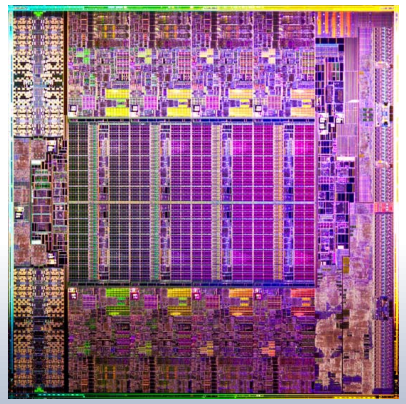
Parallèlement, la puissance de calcul poursuit sa croissance



Importance des aspects matériels

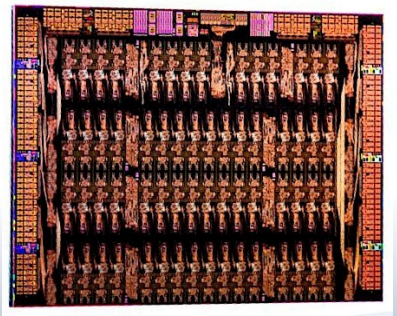
La connaissance des infrastructures et architectures de calcul permet de :

- Choisir son architecture de calcul en fonction de ses besoins
- Adapter les méthodes numériques, les algorithmes et la programmation
- Comprendre le comportement d'un programme
- Optimiser les codes de calcul en fonction de l'architecture

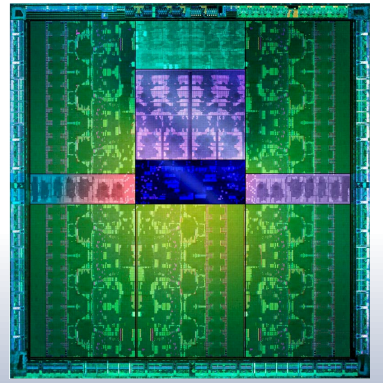


Intel Sandy-Bridge

IBM Blue Gene/Q



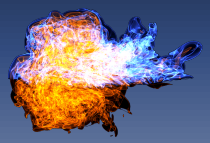
Intel Xeon Phi



Nvidia Kepler

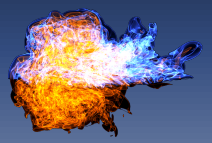
Sommaire

1. **Architectures de calcul actuelles**
2. **Architecture d'un serveur**
   - Processeur
   - Mémoire
   - Réseaux, stockage
3. **Evolutions technologiques et problématiques associées**
   - GPU
   - Many-cœurs
4. **Moyens de calcul disponibles**
   - Micro éclairage sur le HPC mondial
   - Pyramide des moyens de calcul : paysages européen et français
   - Les grilles de calcul
5. **Conclusions**



Sommaire

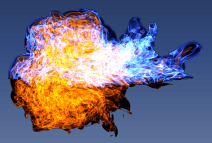
1. **Architectures de calcul actuelles**
2. **Architecture d'un serveur**
   - Processeur
   - Mémoire
   - Réseaux, stockage
3. **Evolutions technologiques et problématiques associées**
   - GPU
   - Many-cœurs
4. **Moyens de calcul disponibles**
   - Micro éclairage sur le HPC mondial
   - Pyramide des moyens de calcul : paysages européen et français
   - Les grilles de calcul
5. **Conclusions**



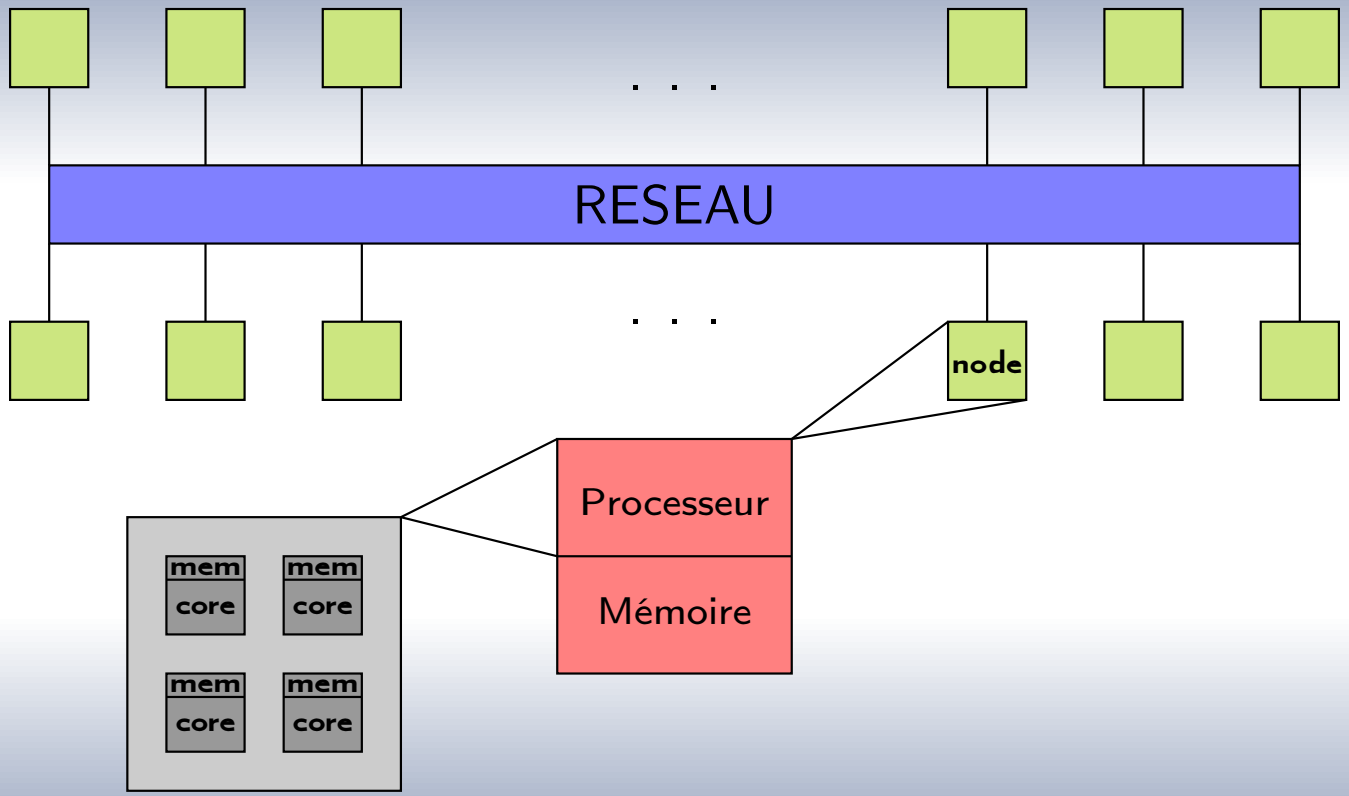
Architectures actuelles

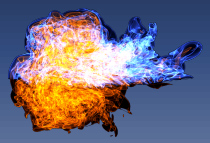
Majoritairement des clusters : regroupement de serveurs connectés par un réseau

- Nombre et type des unités de calcul (processeurs, cœurs, GPU ...)
- Taille de la mémoire
- Type du réseau (infiniband, ethernet ...)



Hiérarchie des architectures





Sommaire

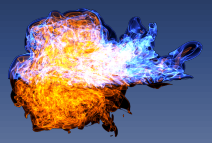
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

5 Conclusions

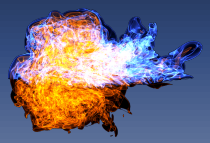


Bits, octets & co

- bit (b) : binary digit, plus petite unité d'information d'un composant en informatique. Vaut 0 ou 1
  - 1 bit permet donc d'avoir 2 états (0 ou 1), 2 bits permettent d'en avoir 4 (2^2) (00,01,10,11),... , n bits permettent ainsi d'en avoir 2^n
- octet (o) ou byte (B) = composé de 8 bits

Usage

- Processeurs et mémoires opèrent sur des octets
- Débits des réseaux et des bus exprimés en bits par seconde



Sommaire

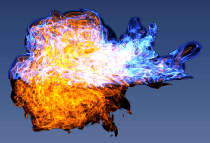
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

5 Conclusions

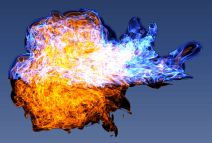


Caractéristiques d'un processeur

- Fréquence d'horloge (MHz) : vitesse de fonctionnement du processeur = nombre de millions de cycles que le processeur est capable d'effectuer par seconde
  - Cycle = plus petite unité de temps au niveau du processeur. Chaque opération/instruction nécessite au minimum un cycle, et plus souvent plusieurs
  - $1 GHz = 10^9 Hz = 10^9 \mathrm{cycle/s}$
- Largeur (32 ou 64 bits) : notamment du bus de données et des registres internes. Bon indicateur de la quantité d'information que celui-ci peut gérer en un temps donné
- Jeu d'instructions : ensemble des opérations qu'un processeur peut exécuter, plus ou moins complexes

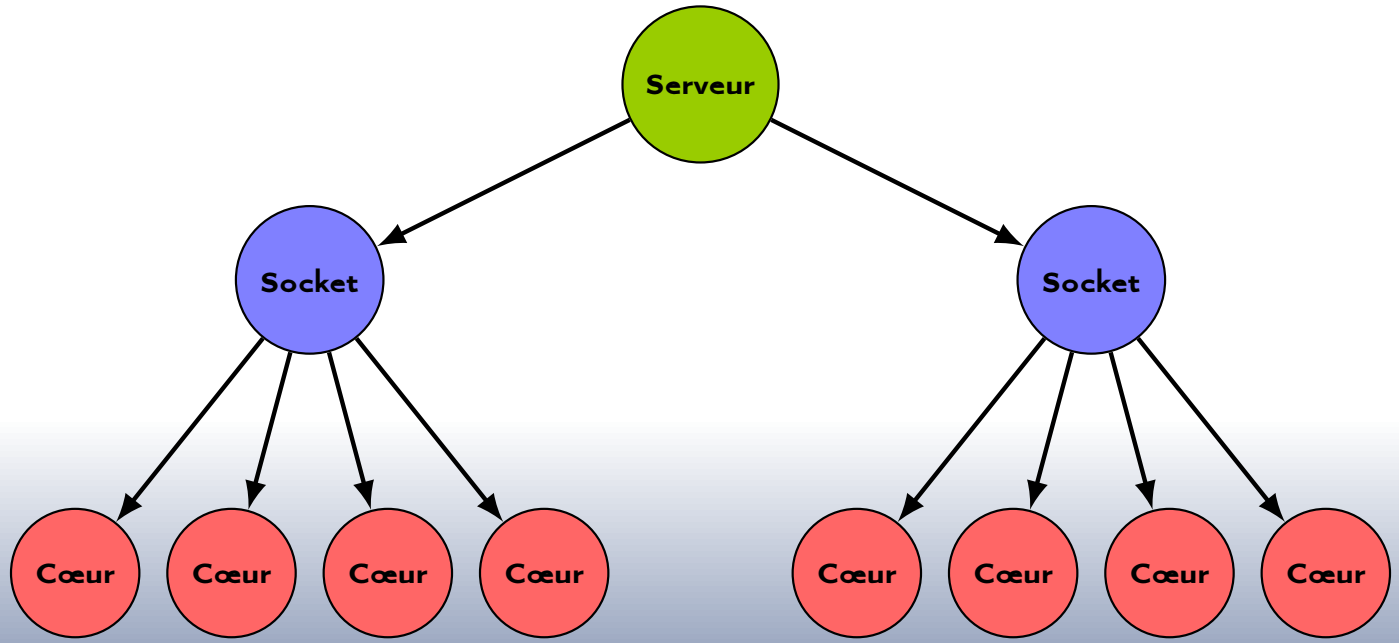
Adressage mémoire

Les processeurs 32 bits ne peuvent pas adresser plus de 4 gibioctets (2^{32}) de mémoire centrale, tandis que les processeurs 64 bits peuvent adresser 16 exbioctets (2^{64}) de mémoire.

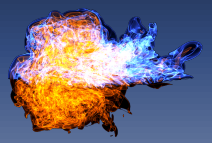


Cœurs, sockets

On évite de parler de CPUs : il faut distinguer le support (socket) de l'unité de calcul elle-même (cœur).
Le socket ou slot est le connecteur qui interfère entre la carte mère d'un ordinateur et le processeur lui-même.

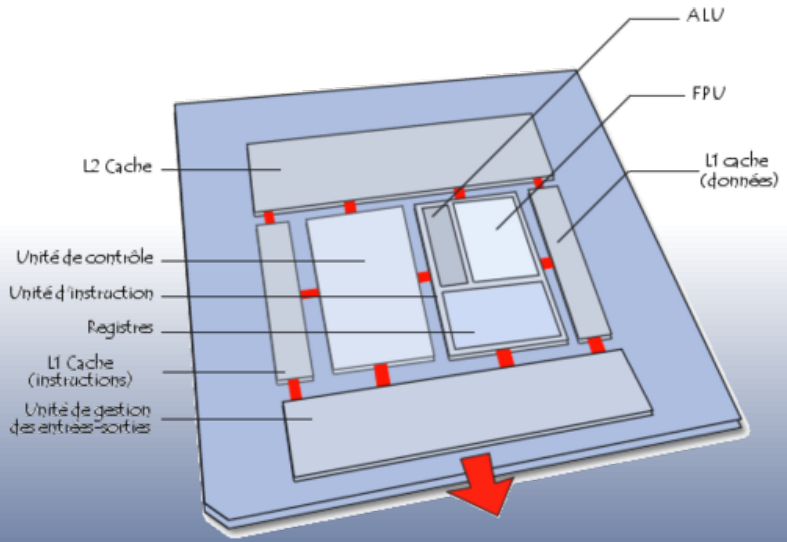


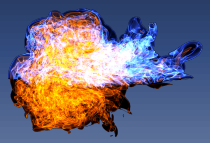
Serveur bi-sockets quadri-cœurs



Composition d'un processeur

- Plusieurs Unité Arithmétique et Logique (UAL), qui prennent en charge notamment les calculs arithmétiques élémentaires ce qui permet de traiter plusieurs instructions en même temps
- Registres : mémoires de petite taille (quelques octets) rapides
- Unité de calcul en virgule flottante (en anglais Floating Point Unit - FPU), qui permet d'accélérer les calculs sur des nombres flottants
- Mémoires caches, qui diminuent les temps d'accès à la mémoire

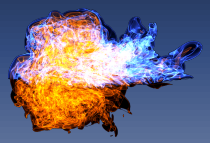




Fonctionnement

- Le processeur est rythmé par une horloge défini par sa fréquence (GHz)
- Pour schématiser, on peut dire que le traitement d'une instruction passe par 5 étapes fondamentales dont la durée est d'au minimum un cycle d'horloge :
  - lecture de l'instruction (IF, Instruction Fetch)
  - décodage de l'instruction (ID, Instruction Decode)
  - exécution de l'instruction (EX, Execute)
  - écriture ou chargement depuis la mémoire en fonction du type de l'instruction (MEM, Memory)
  - stockage du résultat dans un registre (WB, Write Back)
- Plus la fréquence est élevée et plus le processeur peut traiter les instructions rapidement

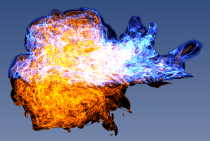
5 cycles au minimum sont nécessaires pour accomplir une instruction



Pipeline

IF	ID	EX	MEM	WB										
	IF	ID	EX	MEM	WB									
		IF	ID	EX	MEM	WB								
			IF	ID	EX	MEM	WB							
				IF	ID	EX	MEM	WB						

- Chaque « travailleur » effectue le même travail sur des objets différents
- Si il y a 5 étapes dans la chaîne, 5 objets peuvent être traités simultanément
- Et si ces étapes prennent le même temps, la chaîne est continuellement occupée

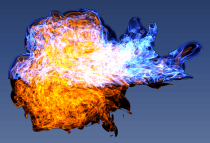


Jeux d'instructions (ISA, Instructions Set Architecture)

- Ensemble des opérations élémentaires qu'un processeur peut accomplir
- Plusieurs familles de processeurs possédant chacune un jeu d'instructions propre :
  - 80x86 : le « x » représente la famille. On parle ainsi de 386, 486, 586, 686, etc.
  - ARM
  - IA-64
  - PowerPC ...

Attention

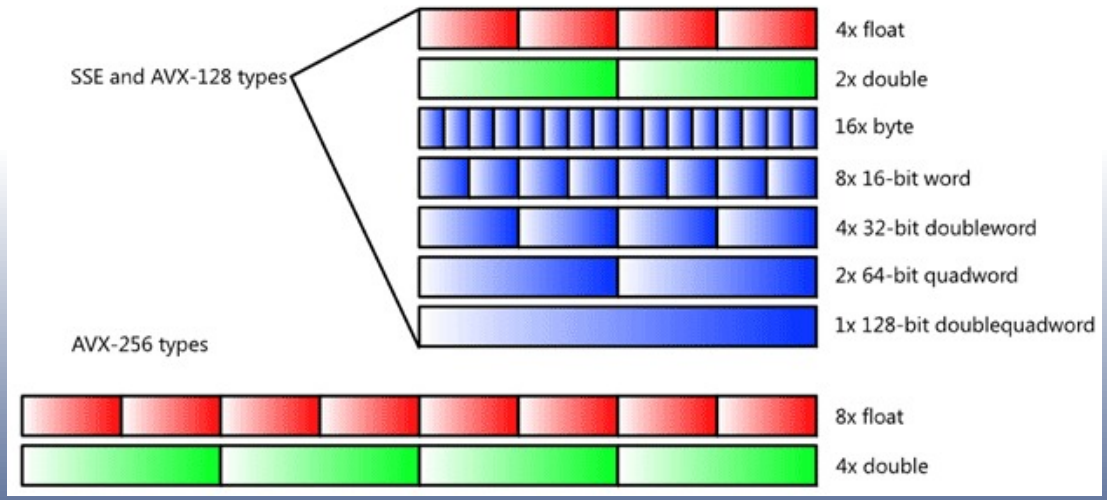
Un programme réalisé pour un type de processeur ne peut pas fonctionner directement sur un système possédant un autre type de processeur.

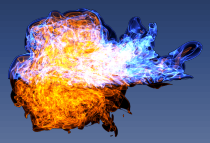


Parallélisation de données

SIMD (Single Instruction Multiple Data) : appliquer la même instruction simultanément à plusieurs données

- Jeux d'instructions :
  - SSE (Streaming SIMD Extensions) chez Intel
  - AVX (Advanced Vector Extensions) chez Intel et AMD (prise en charge de registres vectoriels plus grands)
- Capacités vectorielles des processeurs : possibilité d'exécuter une opération arithmétique sur un registre vectoriel (contenant deux DP ou quatre SP en général).





Parallélisation de données

Exemple SSE

Algorithme :

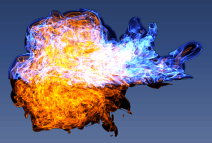
```
for i = 1,n
    x[i] = sqrt(x[i])
```

Travail du processeur :

```
for i = 1,n
    load x[i] to the floating-point register
    calculate the square root
    write the result from the register to memory
```

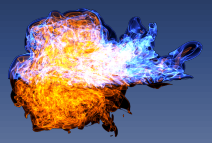
Avec SSE :

```
for {i1,i2,i3,i4} in {1:n}
    load x[i1],x[i2],x[i3],x[i4] to the SSE register
    calculate 4 square roots in one operation
    write the result from the register to memory
```



Flops et performance du processeur

- Opérations (additions ou multiplications) à virgule flottante par seconde = FLoating point Operations Per Second
- Puissance crête (point de vue théorique) : mesure les performance des unités de calcul en virgule flottante (FPU) contenues dans le cœur.
  - Processeur Intel Sandy Bridge six core, 3.2 GHz : 4 opérations flottantes possibles par cycle en DP soit 8 en SP, 6 cœurs par nœud : $8 \times 6 \times 3.2 = 153.6 GFlops$.
  - Processeur PowerPC A2 (Blue Gene/Q), 1.6 GHz : 4 opérations flottantes possibles par cycle en DP, soit 8 en SP, 16 cœurs par nœud : $8 \times 16 \times 1.6 = 204.8 GFlops$.
- D'un point de vue pratique, la puissance d'une machine dépend de l'ensemble de ses composants : fréquence du processeur, accès mémoire, vitesse des bus, complexité de l'architecture ... mais aussi charge de la machine, système d'exploitation ... On est souvent loin de la puissance théorique ...



Performances théoriques et réelles

Exemple : résolution du système linéaire dense $n \times n$ de type $Ax = b$ par élimination de Gauss. La complexité de l'algorithme est de l'ordre de $\frac{1}{3}n^3$ opérations flottantes.

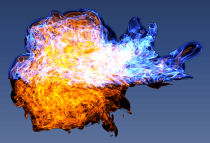
Pour un calculateur de 48 GFlops $= 48 \times 10^9$ Flops :

n	Nombre d'opérations	Temps
1000	3.3×10^8	0.0068 s
10 000	3.3×10^{11}	6.875 s
100 000	3.3×10^{14}	6875 s = 1h 44 mn 35 s
1 000 000	3.3×10^{17}	79j

Mais ...

L'architecture intrinsèque du processeur peut impacter négativement la performance :

- Partage de certaines ressources
- Gestion des dépendances entre instructions
- Prédiction de branchement ...



Sommaire

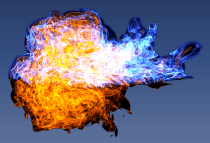
1 Architectures de calcul actuelles

2 Architecture d'un serveur
- Processeur
- Mémoire
- Réseaux, stockage

3 Evolutions technologiques et problématiques associées
- GPU
- Many-cœurs

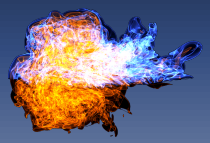
4 Moyens de calcul disponibles
- Micro éclairage sur le HPC mondial
- Pyramide des moyens de calcul : paysages européen et français
- Les grilles de calcul

5 Conclusions

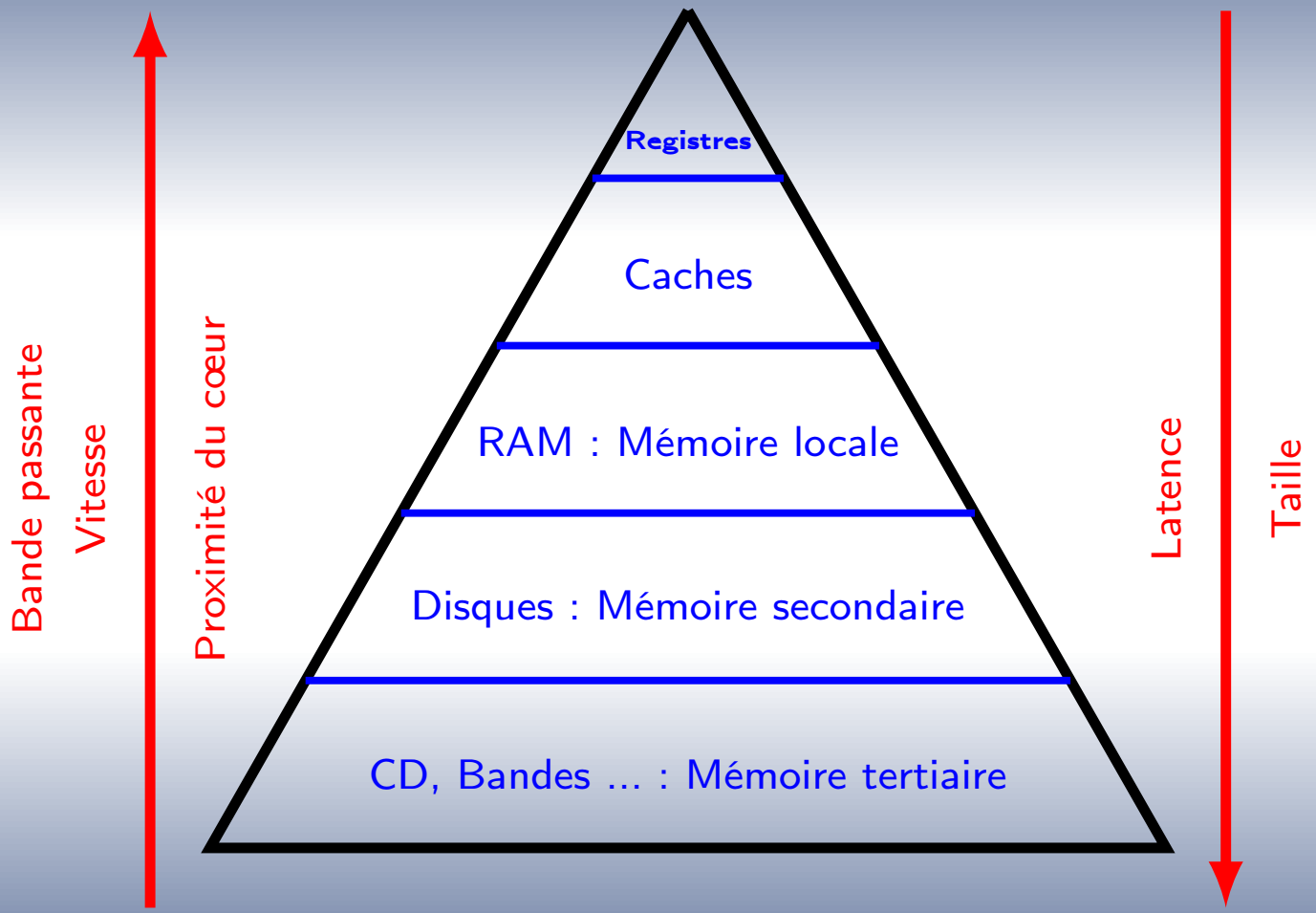


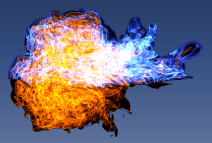
Quelques définitions

- Bande passante = débit d'informations ; d'un périphérique : mémoire, disque dur ... ou d'un médium de communication : réseau, bus ....
  - Mesurée généralement en octets (byte) par seconde (o/s, ou B/s) ou en bits par seconde (bit/s ou bps)
  - Exemple : USB 2.0 $\longrightarrow$ 480 Mbits par seconde. Le transfert d'une vidéo de 900 Mo mettra donc 15 secondes ($900Mo = 8 \times 900Mbits = 7200Mbits$; $7200/480 = 15s$).
- Latence = temps minimum d'établissement de la connexion : indépendant de la quantité de données à transporter
  - Latence de l'ADSL : 30 ms, latence de l'infiniband QDR : 100 ns



Hiérarchie Mémoire

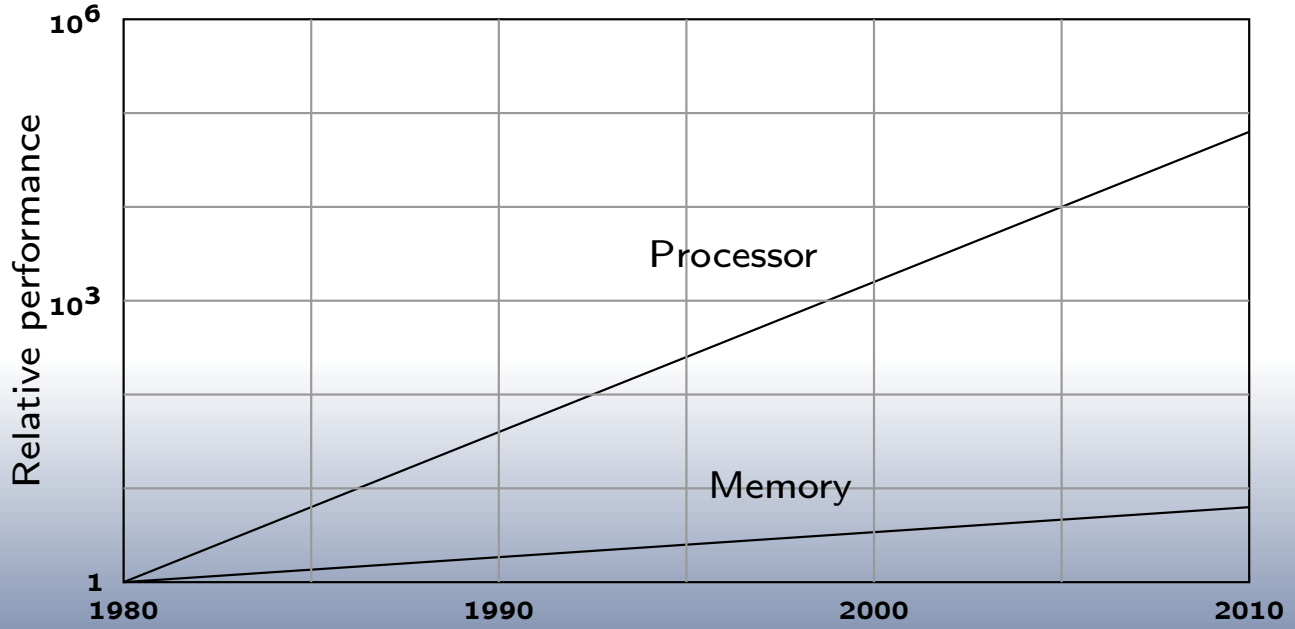


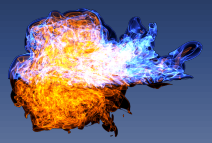


Problématique

« Memory Wall »

$$Vitesse_{CPU} \gg Vitesse_{Memory}$$



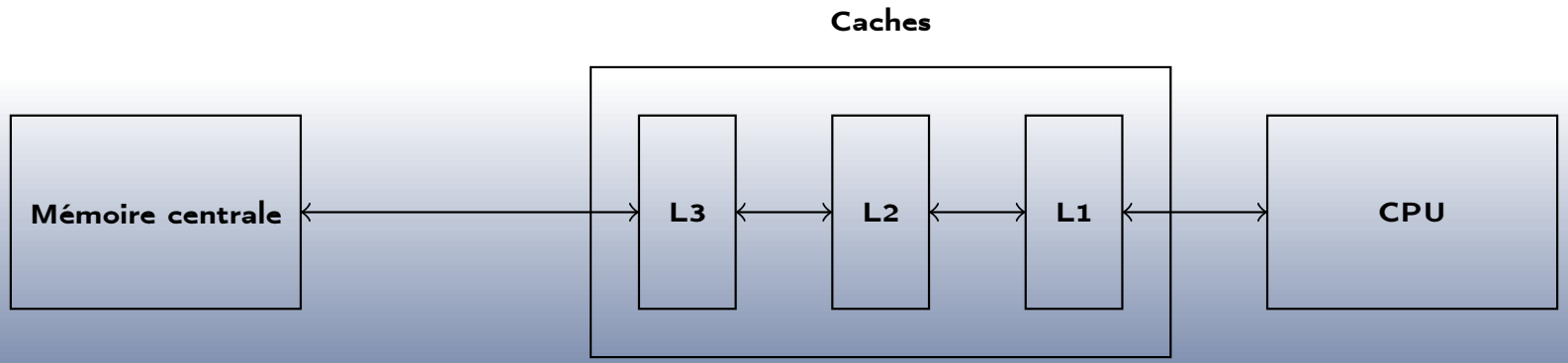


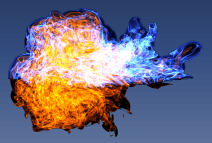
Accès mémoire : problématique

Quels facteurs influencent les performances des accès mémoire ?

- Localisation : cache, RAM, disque ?
- Manière dont elles sont accédées :
  - au travers d'un chipset (jeu de composants électroniques qui permet le contrôle des échanges d'information : Northbridge par exemple)
  - directement par le processeur
  - via le processeur voisin ...

Plus une donnée est proche du processeur, plus elle est accédée rapidement.





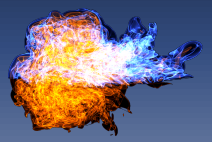
Principes de localité

Localité

- Localité spatiale : lorsqu'un programme accède à une donnée ou à une instruction, il est probable qu'il accédera ensuite aux données ou instructions voisines
- Localité temporelle : lorsqu'un programme accède à une donnée ou à une instruction, il est probable qu'il y accédera à nouveau dans un futur proche

```
subroutine sumVec(vec,n)
    integer :: n
    integer :: vec(n)
    integer :: i,sum=0
    do i = 1,n
        sum = sum + vec(i)
    end do
end subroutine
```

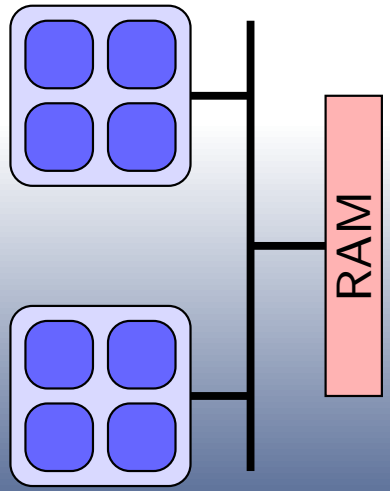
- Bonne localité spatiale des données du tableau `vec` : accès en séquence
- Bonne localité temporelle de la donnée `sum` : accès fréquent



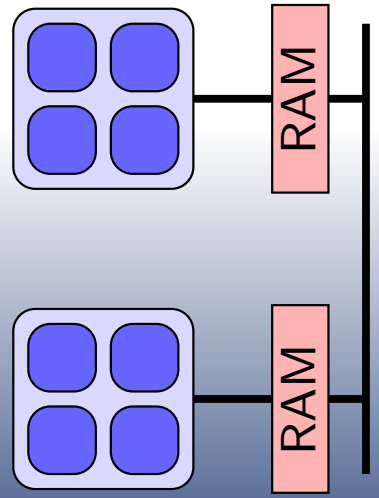
UMA/NUMA

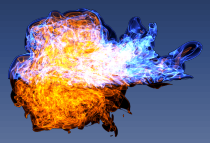
- Modèle SMP : relie plusieurs processeurs à une seule mémoire centralisée. Architecture UMA (Uniform Memory Access) : le temps nécessaire pour accéder à un emplacement quelconque en mémoire est le même pour tous les processeurs.
- Modèle NUMA (Non Uniform Memory Access) : les processeurs sont capables d'accéder à la totalité de la mémoire du système mais les temps d'accès mémoire sont non uniformes en fonction de la distance mémoire/cœur.

Architecture SMP



Architecture NUMA



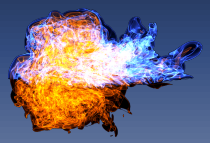


Fonctionnement des caches

- Le cache est divisé en lignes (ou blocs) de mots
- 2 niveaux de granularité :
  - le CPU travaille sur des mots (par ex 32 ou 64 bits)
  - les transferts mémoire se font par ligne (ou bloc, par ex 256 octets)
- Les lignes de caches sont organisées en ensembles à l'intérieur du cache, la taille de ces ensembles est constante et appelée le degré d'associativité.
- Exploitation de la localité spatiale : le cache contient des copies des mots par lignes de cache
- Exploitation de la localité temporelle : choix judicieux des lignes de cache à retirer lorsqu'il faut rajouter une ligne à un cache déjà plein

Lorsque le processeur tente d'accéder à une information (instruction ou donnée)

Si l'information se trouve dans le cache (hit), le processeur y accède sans état d'attente, sinon (miss) le cache est chargé avec un bloc d'informations de la mémoire



Sommaire

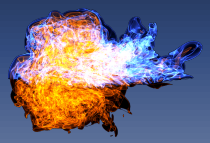
1 Architectures de calcul actuelles

2 Architecture d'un serveur
- Processeur
- Mémoire
- Réseaux, stockage

3 Evolutions technologiques et problématiques associées
- GPU
- Many-cœurs

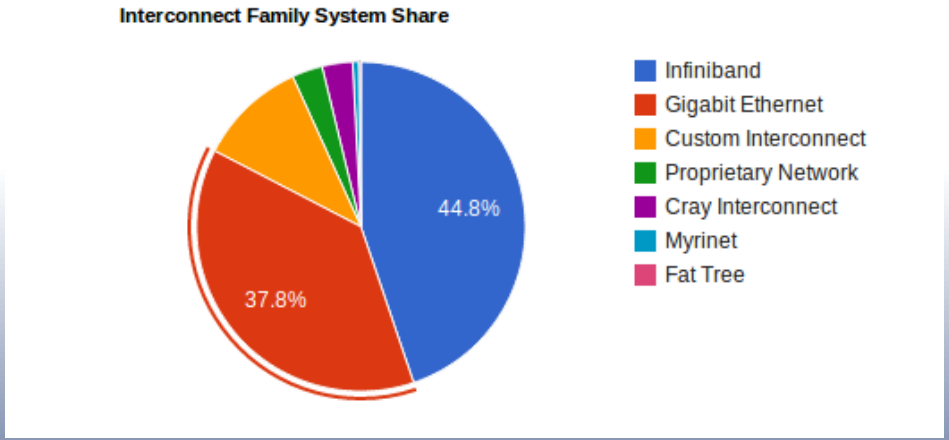
4 Moyens de calcul disponibles
- Micro éclairage sur le HPC mondial
- Pyramide des moyens de calcul : paysages européen et français
- Les grilles de calcul

5 Conclusions

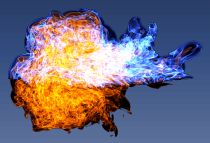


Caractéristiques des réseaux

- La bande passante (bandwidth) est le débit maximal.
  - La bande passante effective est généralement inférieure à la bande passante physique (souvent à cause de l'overhead du protocole)
  - Ce qui intéresse l'utilisateur est la bande passante MPI
  - Les performances bas niveaux sont intéressantes pour le stockage
- Principales technologies :
  - Ethernet (Gb, 10 Gb, 100 Gb)
  - Myrinet
  - Infiniband



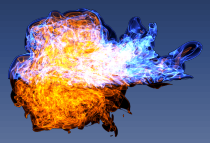
Répartition des familles d'interconnexion, top 500, nov 2012



Stockage : le goulot d'étranglement

Le débit global est dépendant de toute la chaîne d'I/Os (baie, serveur, interconnexion ...).

- La capacité de stockage augmente beaucoup plus vite que la vitesse d'accès
- Le débit vers les disques augmente moins vite que la puissance de calcul
- La quantité de données générées augmente avec la puissance de calcul
- L'approche classique un fichier/processus génère de plus en plus de fichiers
- Trop de fichiers = saturation des systèmes de fichiers (nombre maximum de fichiers et espace gaspillé à cause des tailles de bloc fixes)
- Le temps passé dans les I/O croît (surtout pour les applications massivement parallèles)



Sommaire

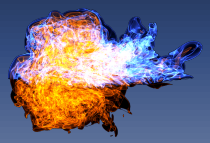
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

5 Conclusions



Kilo, Méga, Giga, Téra, Péta, Exa ...

Top 500 Bench Linpack	Puissance soutenue en Gflops	Puissance crête en Gflops	Nombre de processeurs ou de cœurs
Juin 1993	59.7	131	1024
Juin 2011	8 262 000	8 773 630	548 352

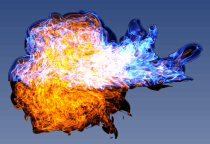
1997 : année du Teraflops (10^{12} Flops)

↓

2008 : année du Petaflops (10^{15} Flops)

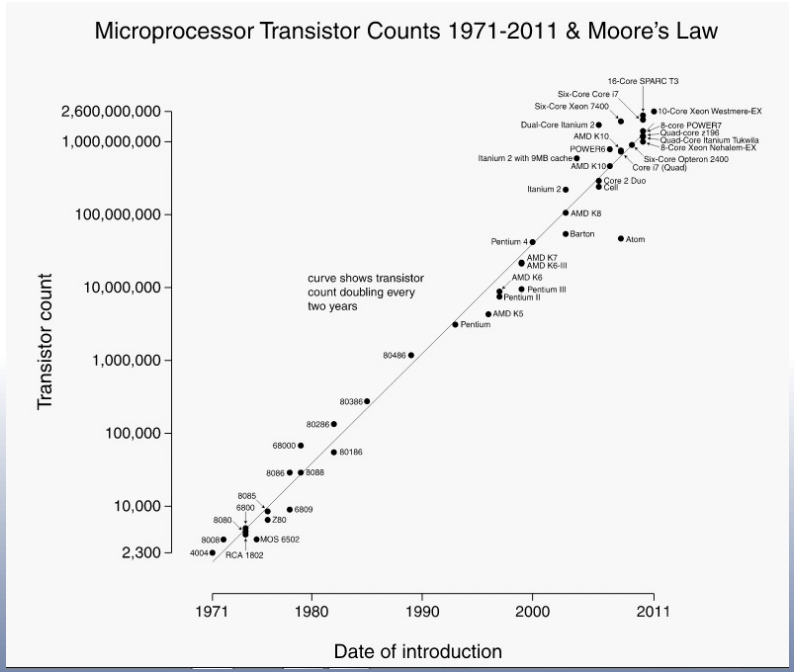
↓

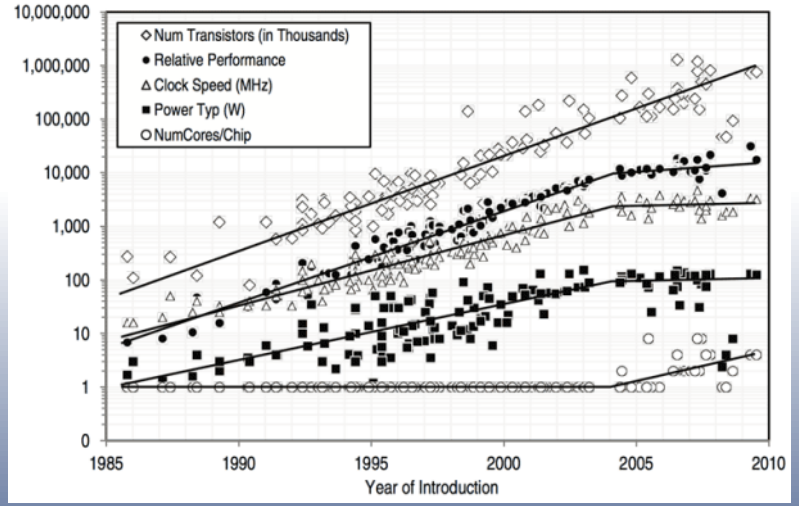
2017 : année de l'Exaflops (10^{18} Flops) ? ? ?

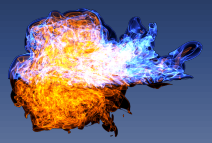


Comment faire des calculateurs plus rapides ?

1. Améliorer la vitesse du processeur : augmenter la fréquence d'horloge.
2. Augmenter le nombre de transistors, donc augmenter le nombre de cœurs.
3. Utiliser des processeurs spécialisés ou des cartes accélératrices.







Les principales limitations

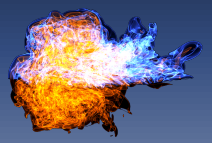
Consommation électrique

- Augmente de façon exponentielle en fonction de la fréquence d'horloge :

$$P \propto f^3$$

  - P : puissance (Watt)
  - f : fréquence

- Mais pas que ! ! tous les éléments sont consommateurs : mémoires, carte mère, alimentation inefficace ...
- *Pour fonctionner à une vitesse d'un exaflops avec ses composants actuels, le super-calculateur Tianhe nécessiterait une puissance de 1,6 milliard de watts, soit un peu plus que la puissance de la centrale nucléaire de Belleville dans le Cher !*



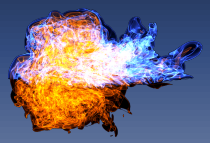
Les principales limitations

Dissipation thermique

- Directement liée à la puissance consommée : plus on augmente la fréquence, plus la dissipation thermique est importante
- Problème du refroidissement au niveau des chips, et au niveau de l'infrastructure

Finesse de gravure

- Actuellement 22 nm. Feuille de route Intel : 4 nm en 2022.
- Plus de transistors donc plus de cœurs, et moins de dissipation thermique (plus c'est petit moins ça dissipe)
- Mais des défis technologiques pour les fondeurs et des technologies de plus en plus coûteuses !
- Jusqu'à quand ? ?



Sommaire

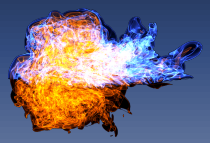
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

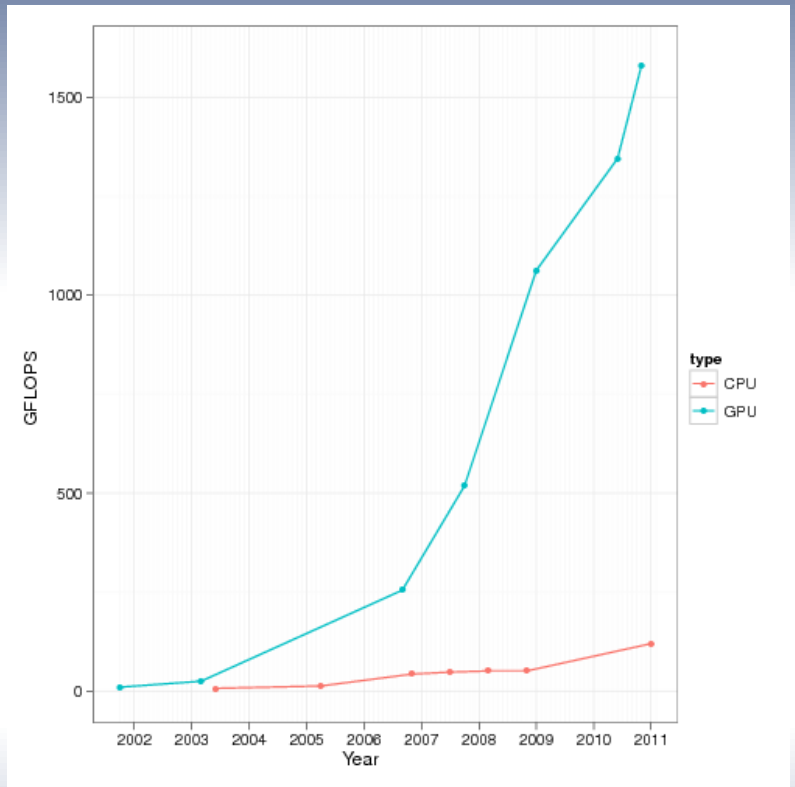
3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

5 Conclusions

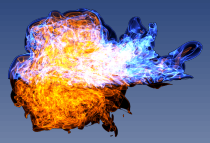


Pourquoi les GPU ?

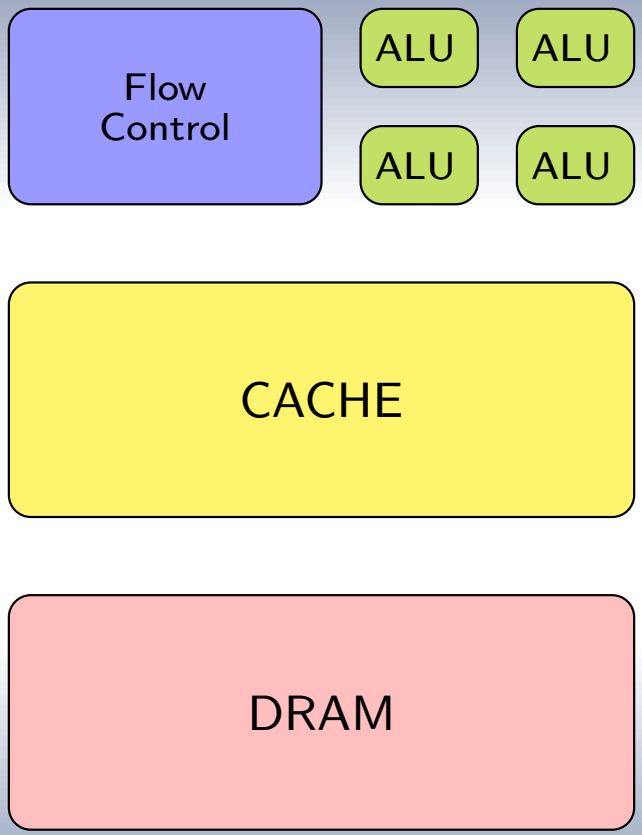


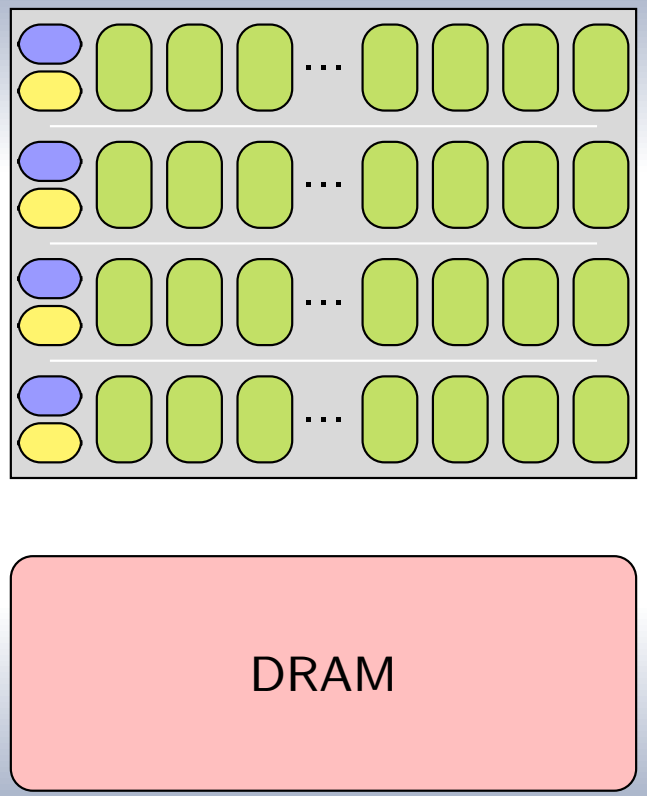
A performance théorique égale, les plateformes à base de GPU :

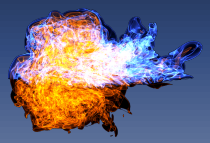
- occupent moins de place
- sont moins chères
- sont moins consommatrices d'électricité



Architecture comparée d'un CPU et d'un GPU







Caractéristiques d'un GPU

- Beaucoup de cœurs : 2496 dans le Nvidia Tesla K20 (Kepler) avec une fréquence de 705 MHz.
- Support IEEE double précision en virgule flottante
- Mémoire : jusqu'à 6 GB

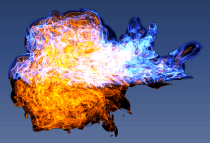
Exploitation d'un GPU

- Programmation de type SIMD, gestion de milliers de thread
- Transfert CPU-GPU très couteux
- API de programmation (CUDA, OpenCL)

→ Complexe à exploiter, toutes les applications ne s'y prêtent pas

Principaux fournisseurs

- Nvidia
- Depuis peu, AMD avec sa carte graphique serveur FirePro S10000



Sommaire

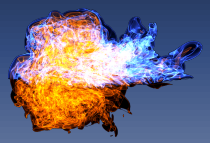
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

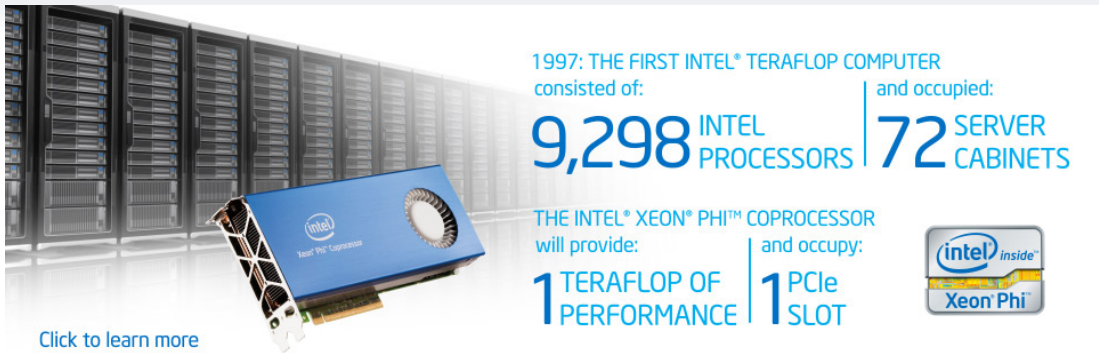
4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

5 Conclusions



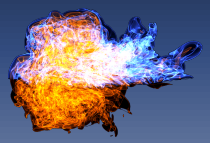
Evolution actuelle

- Apparition sur le marché de cartes Many Integrated Core : les Xeon Phi©
- Intel Xeon Phi 5110P : 60 coeurs à 1053 MHz, 8 Go de mémoire



Principal avantage

- Les Xeon Phi supportent un environnement de programmation complètement standard.



Sommaire

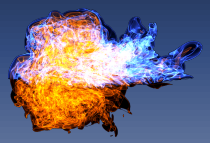
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

5 Conclusions



Sommaire

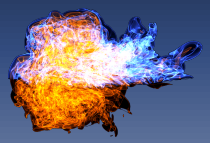
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

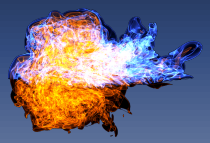
5 Conclusions



Micro éclairage sur le HPC mondial

Top 500

- Projet de classification des 500 premiers supercalculateurs connus au monde.
- Depuis juin 1993, liste mise à jour tous les six mois
- LinPack : créé par Jack Dongarra, il mesure le temps mis par un ordinateur pour résoudre un système de n équations à n inconnues dense, la solution étant obtenue par une utilisation partielle du pivot de Gauss. La performance est ensuite calculée en divisant le nombre d'opérations par le temps mis, donc en FLOPS.



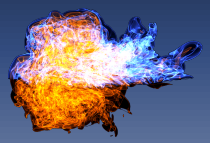
Top 500, novembre 2012

Quelques statistiques

- Cartes accélératrices : 87% sans, 10% Nvidia, 1.4% Xéon Phi.
- Pays : 50.2% USA, 14.4% Chine, 6.4% Japon, 4.8% UK, 4.2% France, 3.8% Allemagne

IBM BlueGene

- Augmenter le nombre de coeurs en réduisant leur fréquence.
- Diminution de l'écart entre les débits et les latences mémoires, diminution de l'écart entre les débits et les latences réseaux $\longrightarrow$ machine plus équilibrée, une puissance de calcul élevée à consommation électrique équivalente
- Mais chaque coeur est lent, un très grand nombre de coeurs, peu de mémoire par coeur



Sommaire

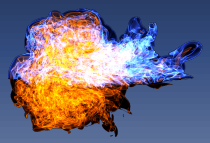
1 Architectures de calcul actuelles

2 Architecture d'un serveur
Processeur
Mémoire
Réseaux, stockage

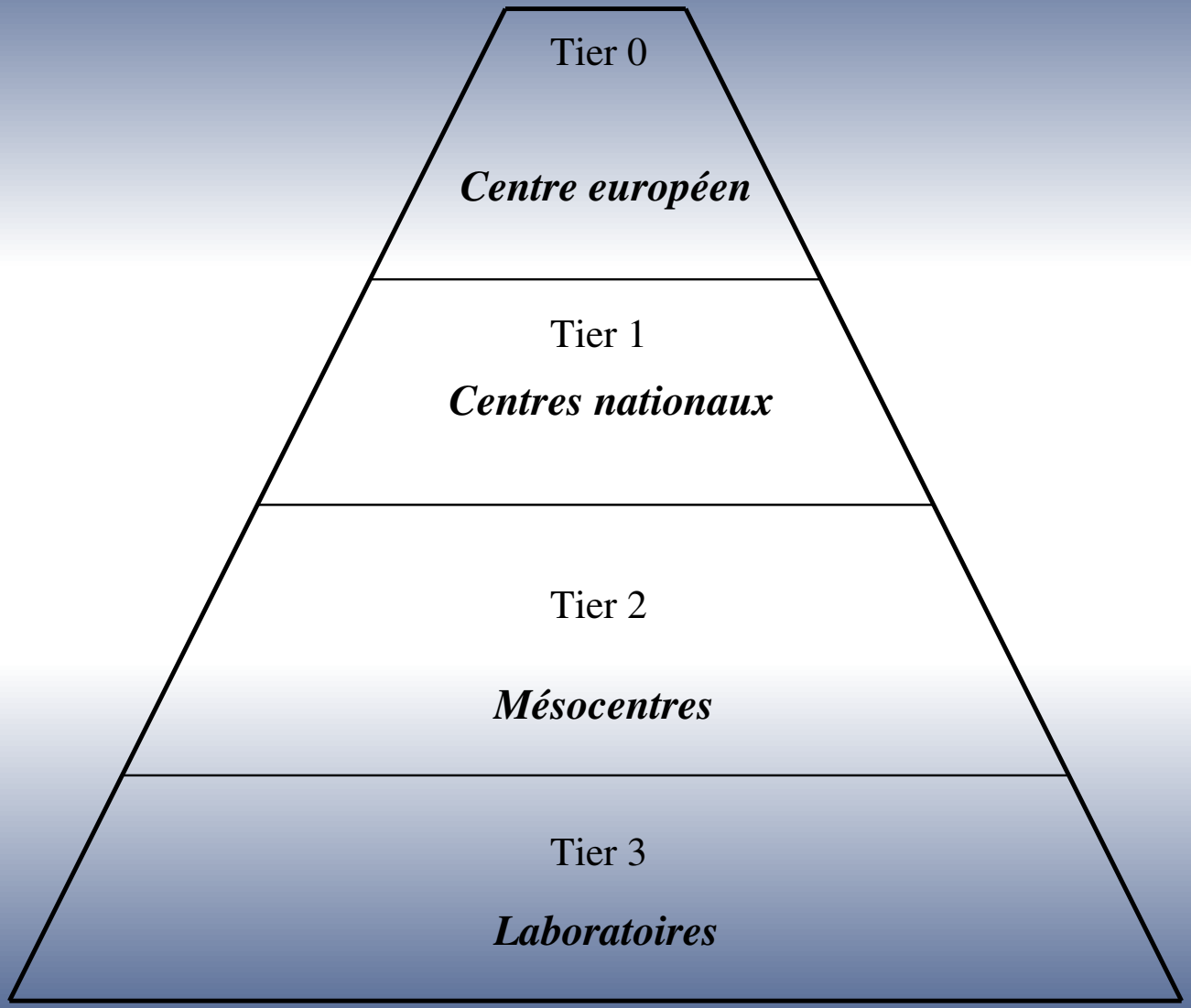
3 Evolutions technologiques et problématiques associées
GPU
Many-cœurs

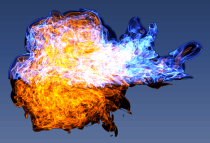
4 Moyens de calcul disponibles
Micro éclairage sur le HPC mondial
Pyramide des moyens de calcul : paysages européen et français
Les grilles de calcul

5 Conclusions

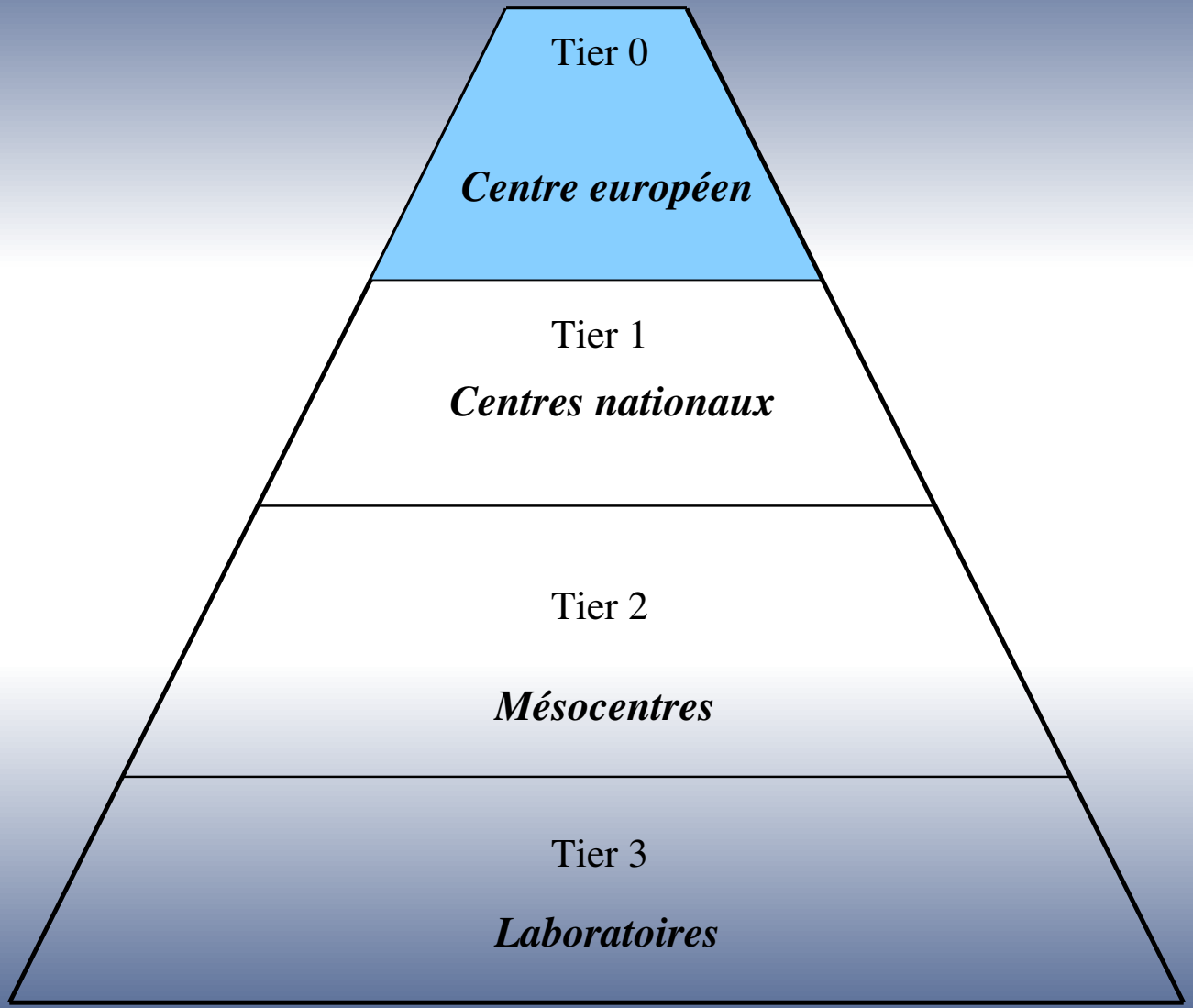


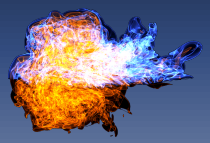
Paysages européen et français





Paysage européen



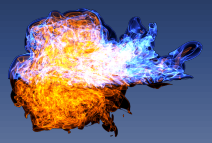


PRACE et l'échelle européenne

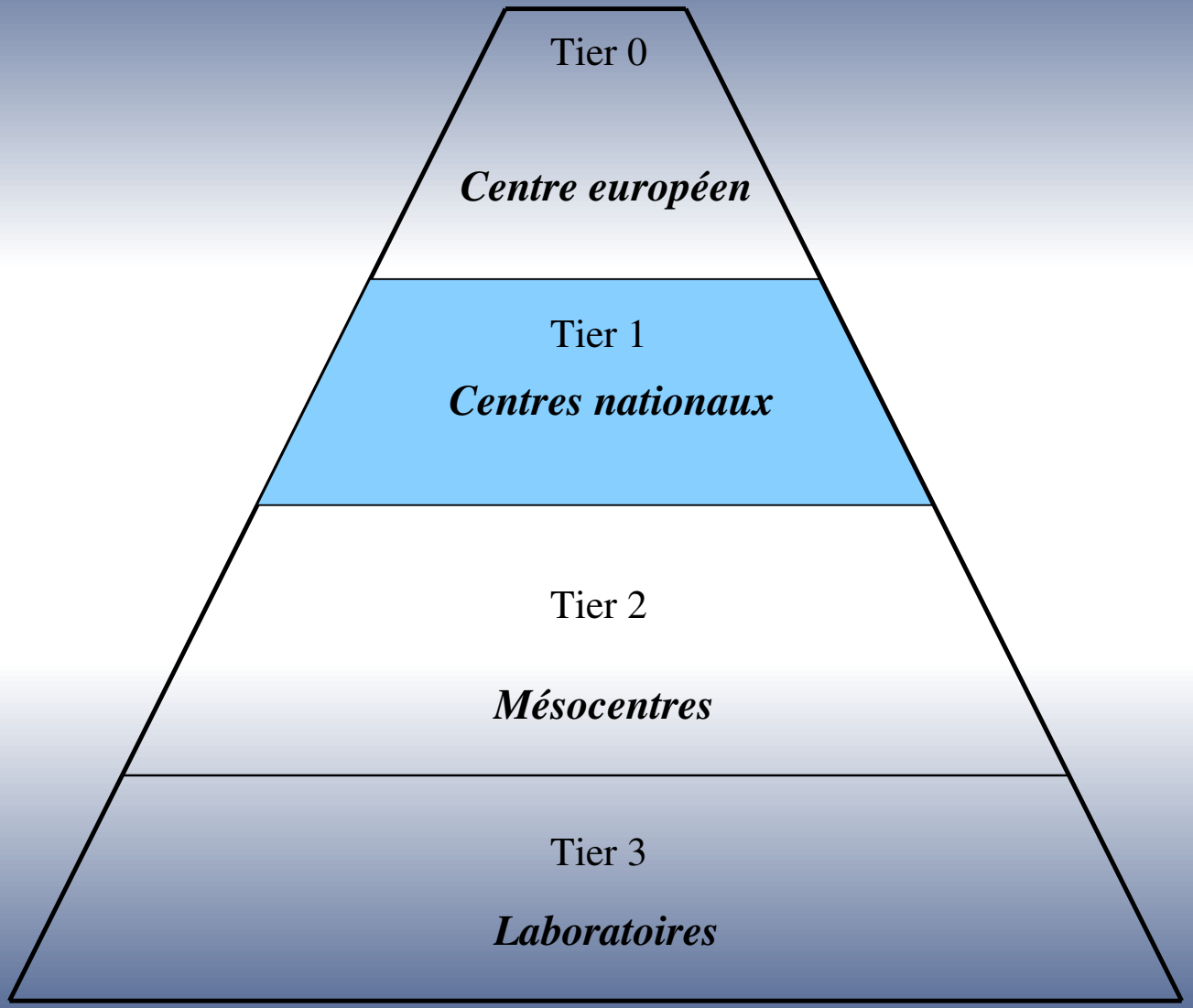
PRACE : Partnership for Advanced Computing in Europe

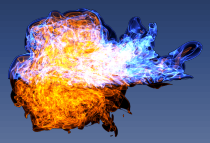
- Regroupement de 25 pays (dont Allemagne, Espagne, Grande-Bretagne, France ...)
- Accès à 6 supercalculateurs aux architectures complémentaires, d'une puissance crête globale de près de 15 petaflops, localisés en Allemagne, en Espagne, en France (CURIE) et en Italie.

http://www.prace-ri.eu/



Paysage national français





GENCI et l'échelle nationale

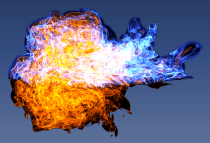
GENCI, Grand Equipement National de Calcul Intensif

Société de droit civil détenue à 49% par l'Etat représenté par le Ministère de la Recherche et l'Enseignement Supérieur, 20% par le CEA, 20% par le CNRS , 10% par les Universités et 1% par l'INRIA.

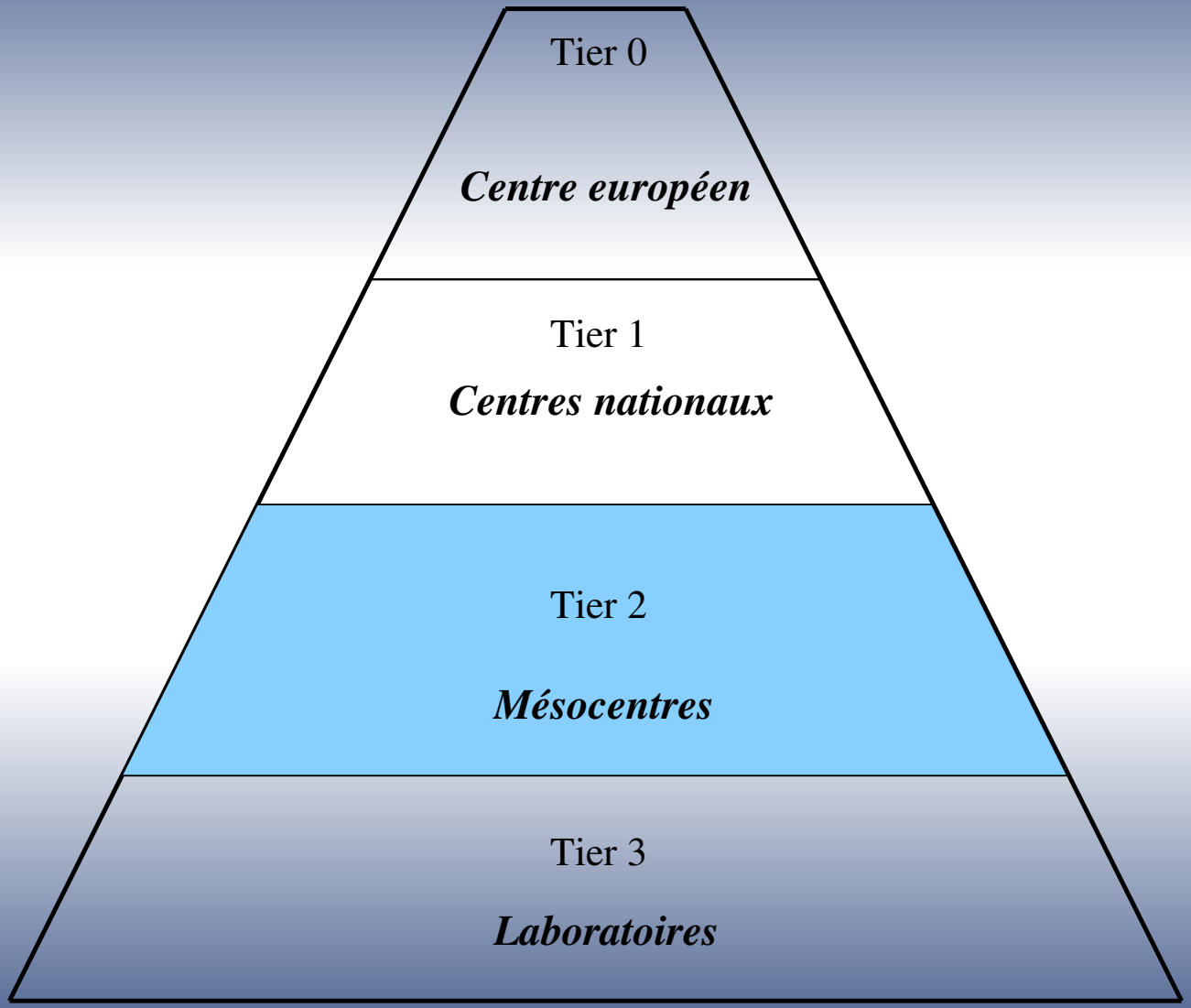
- TGCC/CCRT au CEA
- CINES
- IDRIS

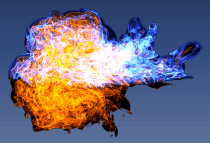
GENCI
GRAND EQUIPEMENT NATIONAL POUR LE CALCUL INTENSIF

`http://www.genci.fr`



Paysage régional





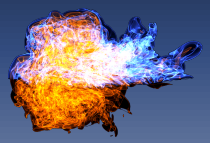
Les mésocentres et l'échelle régionale

Définition d'un mésocentre

- Un ensemble de moyens humains, de ressources matérielles et logicielles à destination d'une ou plusieurs communautés scientifiques, issus de plusieurs entités (EPST, Universités, Industriels) en général d'une même région, doté de sources de financement propres, destiné à fournir un environnement scientifique et technique propice au calcul haute performance.

FLMSN et CIMENT

- Fédération Lyonnaise de Modélisation et des Sciences Numériques à Lyon
  `http://www.flchp.univ-lyon1.fr/`
- CIMENT à Grenoble
  `https://ciment.ujf-grenoble.fr`



Sommaire

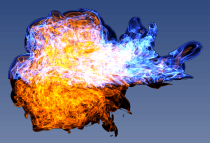
1 Architectures de calcul actuelles

2 Architecture d'un serveur
- Processeur
- Mémoire
- Réseaux, stockage

3 Evolutions technologiques et problématiques associées
- GPU
- Many-cœurs

4 Moyens de calcul disponibles
- Micro éclairage sur le HPC mondial
- Pyramide des moyens de calcul : paysages européen et français
- Les grilles de calcul

5 Conclusions

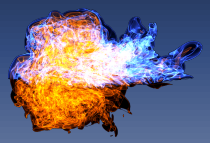


Les grilles de production et de recherche

France Grille

- Infrastructure de grille de production française
- Contribue à l'infrastructure de grille européenne EGI
- Favorise les échanges entre grilles de production et grilles de recherche

http://www.france-grilles.fr



Sommaire

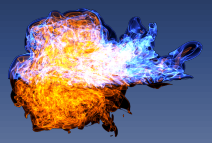
1 Architectures de calcul actuelles

2 Architecture d'un serveur
- Processeur
- Mémoire
- Réseaux, stockage

3 Evolutions technologiques et problématiques associées
- GPU
- Many-cœurs

4 Moyens de calcul disponibles
- Micro éclairage sur le HPC mondial
- Pyramide des moyens de calcul : paysages européen et français
- Les grilles de calcul

5 Conclusions

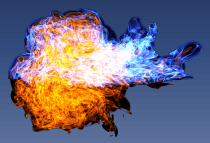


Conclusions

Fréquence d'horloge, nombre de cœurs, nombre d'unités flottantes, caches, stockage, réseaux ...
Tous les éléments ont leur importance ; l'architecture doit être équilibrée pour que l'ensemble soit performant

Au niveau technologique

- Beaucoup de cœurs
- Des cœurs hybrides CPU / cartes accélératrices
- Une différence entre la vitesse de calcul et les transferts mémoire et I/O



Conclusions

Au niveau applicatif

- Nécessité d'une bonne connaissance des architectures, notamment au niveau de la mémoire.
- Plus de croissance des performances au niveau du cœur $\Longrightarrow$ nécessité d'élaborer des algorithmes et des programmes capables d'exploiter un grand nombre de processeurs.
- Nécessité d'exploiter le parallélisme aux différents niveaux du hardware.
- Programmation sur des architectures hybrides avec différents types de processeurs.