



énergie atomique • énergies alternatives

Le HPC dans le monde

(quelques leçons pour l'Europe ?)

Pierre LECA

Chef de Département

CEA Direction des Applications Militaires

Un indicateur macroscopique: la liste du Top500



energie atomique • energies alternatives

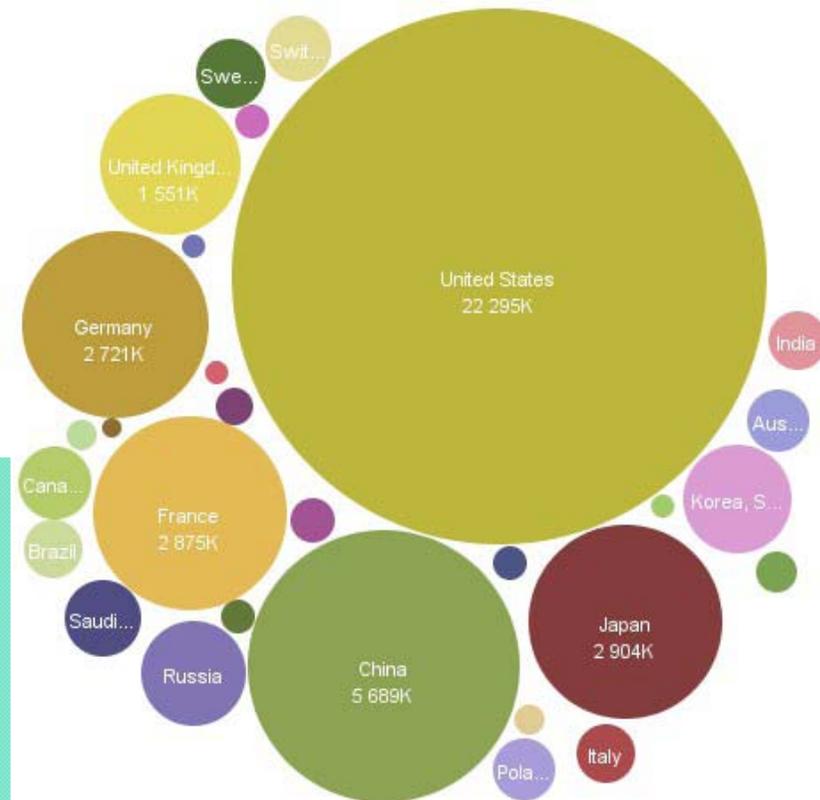
Unités de calcul :
Teraflops (1000 milliards flop/s)
Petaflops (1000 Tera)

Unités de stockage :
PetaOctets

TOP 10 Systems - 11/2010

1	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C
2	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz
3	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU
4	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows
5	Hopper - Cray XE6 12-core 2.1 GHz
6	Tera-100 - Bull bullx super-node S6010/S6030

Rmax Sum (GF)
Disks colored by Countries



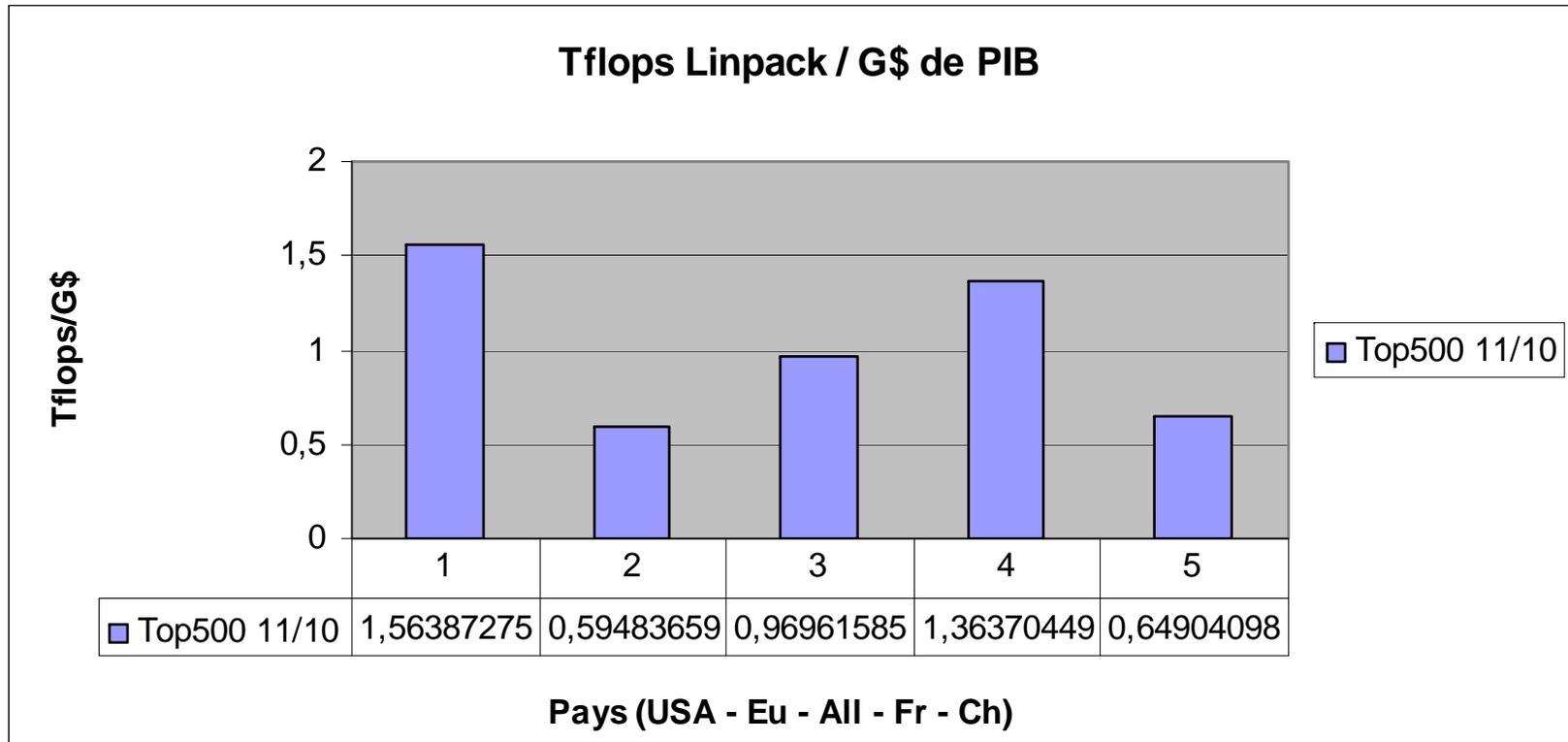
Plus de 44 Pflops (Linpack) recensés dans le monde par le Top500 en novembre 2010

- Etats-Unis : 51 % - Europe : 22 %
- Chine : 14 % - Japon : 6,5 %
- France : 6,3 % - Allemagne : 6,1 %

Le Top500 sous un autre angle ...



energie atomique • energies alternatives



Deux remarques :

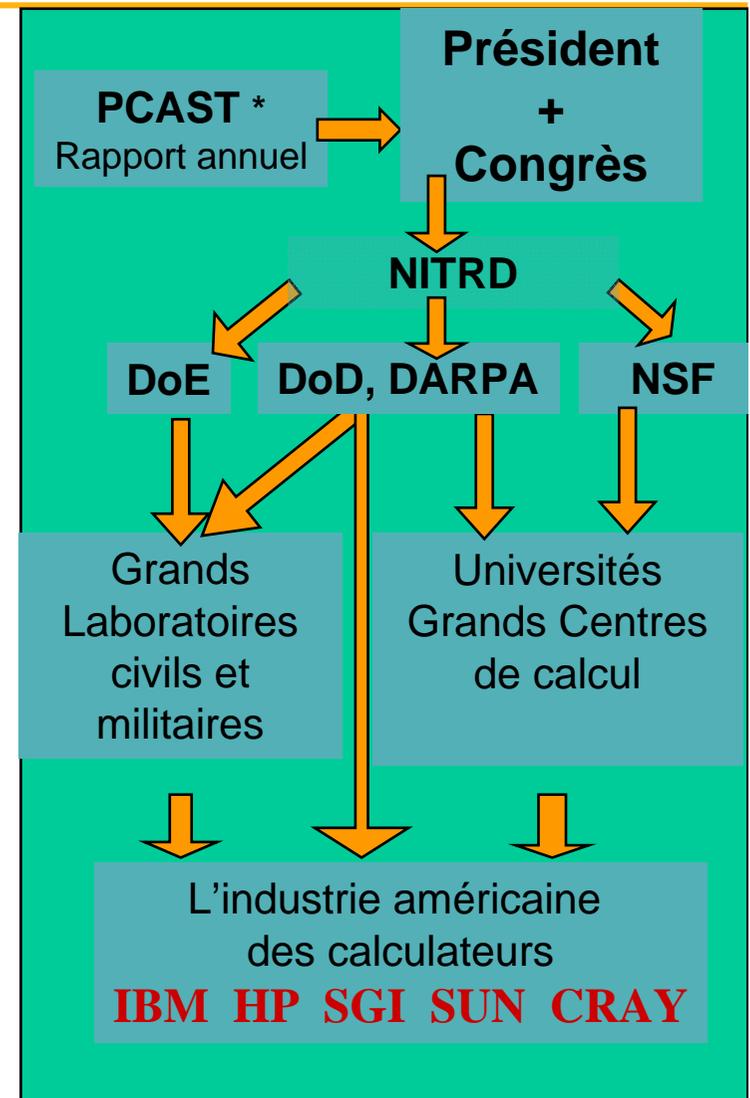
- la France, un exemple pour l'Europe ?
- en phase de rattrapage, la Chine possède encore une énorme marge de progression.

USA : une vision stratégique à long terme



energie atomique • énergies alternatives

- Le calcul intensif bénéficie d'un soutien au plus haut niveau (ex. « *High Perf. Comp. Act* » - 1991/1993/2007 et *Dpt. Of Energy High-End Computing Revitalization Act* – 2004)
- Un programme fédéral, coordonné par le Bureau du *Networking and Information Technology Research and Development* (NITRD), est décliné par différentes agences (DARPA, DOE, NSF, NASA, DoD..) :
 - financement des grands centres de calcul et des infrastructures.
 - soutien à l'industrie informatique
 - développements de logiciels
 - structuration des communautés disciplinaires
- Une présence dans tous les secteurs
- Une communication scientifique et stratégique colossale



(* PCAST : President's Council of Advisors on Science and Technology)

Plusieurs agences fédérales impliquées..



energie atomique • énergies alternatives



- **Defense Advanced Research Agency (DARPA)** : soutien aux développements technologiques – programmes « High Productivity Computing Systems » (pour le xPetaflops) puis UHPC (pour l'Exaflops)
- **Department of Energy (DoE)** : programme « Advanced Scientific Computing Research » (ASCR)
- **National Scientific Foundation** : recherche académique – « Directorate for Computer and Information Science and Engineering » (CISE) et « Office of CyberInfrastructure » (OCI)
- Mais aussi : NASA (Aéronautique et Espace), DoD (Défense), NIH (Santé) ...

Une dépense cumulée annuelle d'environ 1,3 G\$ (hors NSA)

Le dispositif législatif américain (ou le protectionnisme masqué)



énergie atomique • énergies alternatives

- High Perf. Comp. Act (1991-1993-2007): *coordination des efforts et orientation des marchés publics*
« Fostering United States competitiveness in high-performance computing and related activities »

- DOE HEC Revitalization Act (2004): *mission R-D et transfert de techno confiée au DOE*

- Bayh-Dole Act (1980): *oriente les transferts de techno vers l'industrie US*
Autrement connue sous le nom de « University and Small Business Patent Procedures Act »
Preference for United States industry

Synthèse - Ambassade de France (février 2006)



énergie atomique • énergies alternatives

« ..Si IBM apparait comme l'acteur majeur actuel, en étant le constructeur de près de la moitié des machines classées au Top500, il ne s'agit pas de la réussite d'une société privée isolée, mais bien d'un effort fortement appuyé par le gouvernement.... »

« ...Il faut noter que, si les considérations de sécurité de défense (que ce soit au sens du contrôle à l'export de ces technologies ou de leur emploi pour les besoins de défense) restent présentes, c'est l'avance technologique et la compétitivité économique qui sont le plus souvent mises en avant.Au demeurant un certain protectionnisme, sous l'influence du Congrès, est de mise en matière d'acquisitions. ... »

Asie : deux acteurs majeurs



énergie atomique • énergies alternatives

- **Japon :**

- un acteur historique bien connu grâce au développement de « l'Earth Simulator »
- une supervision par le Ministère de l'Education, de la Culture de la Science et de la Technologie
- une étroite coopération industrie-recherche
- une nouvelle étape, le projet **Kei-Soku** doté de **820 M€** sur la période (2006-2012)

- **Chine :**

- une volonté de maîtriser l'ensemble de la chaîne technologique du processeur jusqu'à l'intégration de systèmes
- un programme de R&D lancé en mars 86 (programme 863) sous l'égide du Ministère de la Science et de la Technologie

Japon - Le projet « Kei-Soku Keisan-Ki »



energie atomique • énergies alternatives

- **Objectifs :**

- reprendre le leadership dans la conception et la réalisation des supercalculateurs
- décliner à grande échelle les technologies développées (microélec.)
- faire des avancées en simulation en insistant sur des domaines non traditionnels (nanosciences et sciences de la vie – 12 applications sur les 21 sélectionnées)
- cible : **un calculateur de 10 Petaflops en 2012** (1^{ère} étape 1 Pflops crête démontré par Fujitsu fin 2010)

- **Démarche :**

- des technologies « matérielles » de base jusqu'au centre de calcul
- des logiciels « système » jusqu'aux applications

- **Pilotage :**

- confié au « Next Generation Supercomputer R&D Center » de l'Institut de Recherche RIKEN
- partenariat industriel (NEC, Fujitsu, Hitachi)



CHINE : un programme à long terme



energie atomique • énergies alternatives



Une planification par périodes de 5 ans avec une sélection d'objectifs d'usage dual (programme 863) :

- en 2005 le « quotidien du peuple » annonce que le 11^{ième} plan quinquennal prévoit le développement d'un ordinateur de 1 Petaflops à l'horizon 2010
- confirmation en avril 2007 du lancement du projet sur le site Web du MOST (www.most.gov.cn)

Sur les 41 ordinateurs chinois du Top500 de nov. 2010, les 4 plus puissants :

- # 1 - National University of Defense Technology (NUDT) TH-1A (cluster hybride CPU-GPU de 4,7 Pflops – 2,5 Linpack)
- # 3 - National Supercomputing Centre Shenzhen Dawning (cluster hybride de 2,98 Pflops – 1,27 Linpack)
- # 12 - National University of Defense Technology (NUDT) TH-1 (cluster hybride de 1,2 Pflops – 0,5 Linpack)
- # 28 – Académie des sciences – (cluster hybride de 1,1 Pflops – 0,2 Linpack)

Une démarche cohérente: MIPS, ARM...

- « Chinese IC developer uses ARM's Osprey for laptop designs – sept 2010 »
- Semiconductor Manufacturing International Corporation based in Shanghai (x) Résultat d'un accord ST – ICT de 2007

- 41 ordinateurs dans le Top500 de novembre 2010

- Les ordinateurs chinois les plus puissants ont été développés et intégrés en Chine (sociétés Dawning, Lenovo et Université nationale de la défense)

- Un programme de développement de microprocesseurs bien avancé :

- 6^{ième} génération de microprocesseur Godson (initialement basé sur l'architecture MIPS) - 200 chercheurs travaillent sur ce projet.
- Godson 3B à 8 cœurs (+ extension vectorielle) récemment annoncé – 40 W, 1 Ghz, techno 65nm STmicro (x) puis SMIC – émulation X86 - Utilisé par Dawning (cluster de blades 300 Tflops)
- Godson 3C sur les rails – techno 28 nm en 2012

Russie : faire émerger un acteur local



energie atomique • énergies alternatives



Discours du Pdt Medvedev au Conseil de Sécurité de Russie (juillet 2009):

«.. It's no secret that the majority of the most developed and advanced nations are focusing on this. It is obvious that the large-scale use of high technology data processing increases the effects of research many times over, radically reduces the cost of designing the most advanced and complex types of products, naturally increases the quality of industrial products, and streamlines business processes. It is precisely for these reasons that the entire world is working on this. Any country that makes headway in relation to creating supercomputers has, of course, advantages in terms of competitiveness, increasing its defence capacities, and strengthening security... »

En juin 2010: cluster de 414 Tflops installé par T-Platforms à U. Moscou – **13^{ème} calculateur mondial** (18^{ème} en novembre)

11 calculateurs dans le Top500

- « Boris Gryzlov, Pdt de la Douma, a souligné que la réalisation des programmes russo-biélorusses SKIF et SKIF-GRID, lancés en 2000, permettait de développer, indépendamment des Occidentaux, des superordinateurs qui seront à même de résoudre les problèmes les plus complexes (2008) »

- « **T-Platforms** is the leading Russia-based developer ... The company provides solutions for any sectors of national economy, including the state management, industry, science, education, and telecommunications.. »

- Ont été annoncés (Agence Novosti):
 - le déploiement de minisupercalculateurs, développés par les ingénieurs de Sarov, dans les centres « nucléaires de défense »
 - le développement de technologies de refroidissement dans le cadre d'un projet visant **5 Pflops**
 - en 2011 pour le tricentenaire de la naissance de M. Lomonossov, + 800 Tflops pour passer à 1,3 Pflops (avec GPU) à l'université de Moscou et 1 Pflops (X86) installé à Sarov (centre nucléaire)

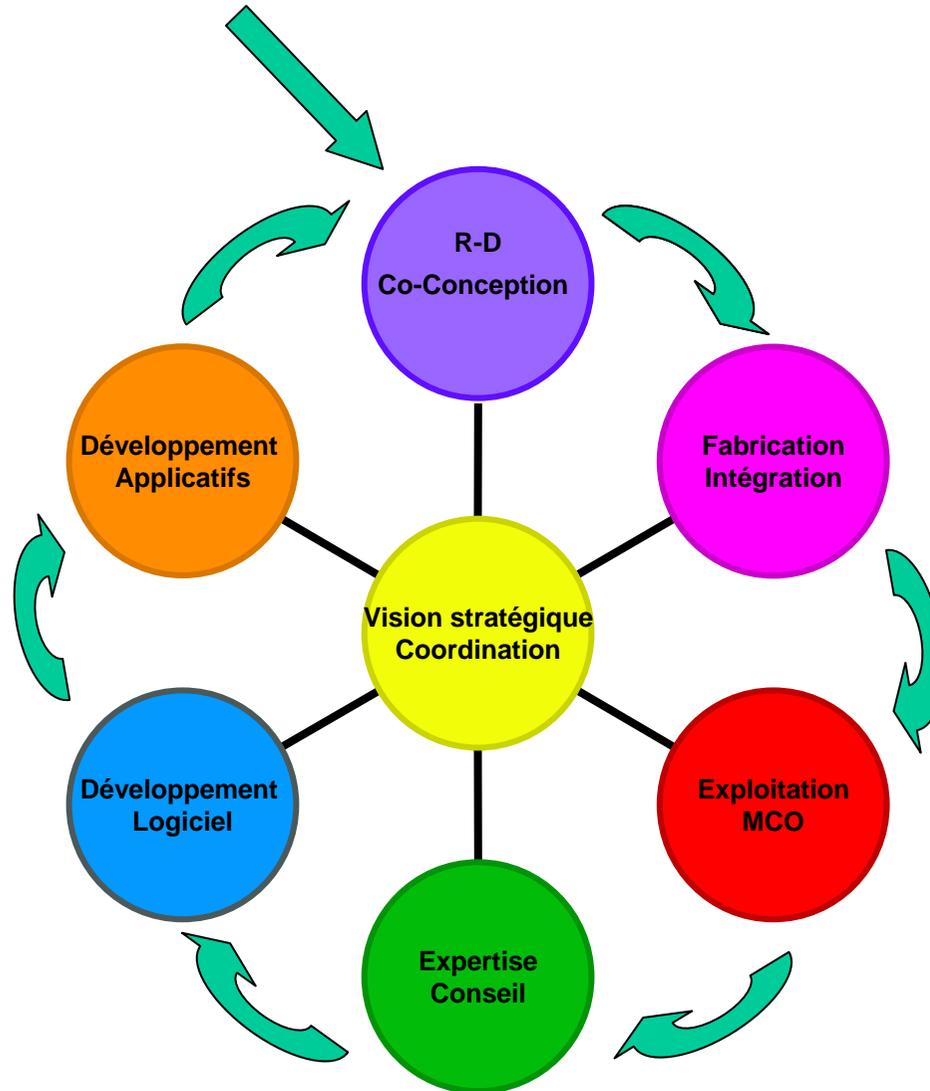
Il faut fermer le cercle...



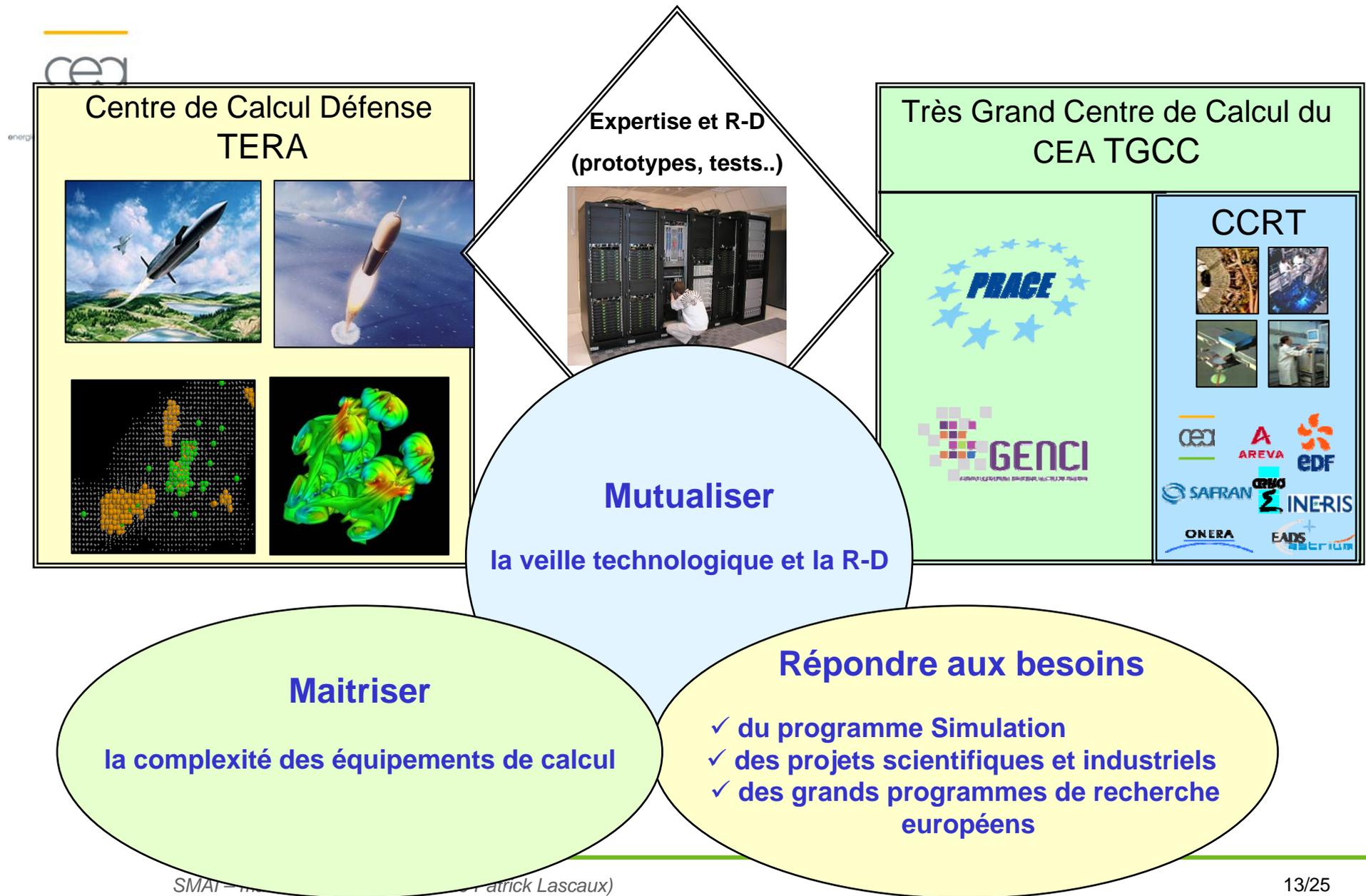
energie atomique • energies alternatives

Le chaînon manquant ! Sortir de la relation client-fournisseur

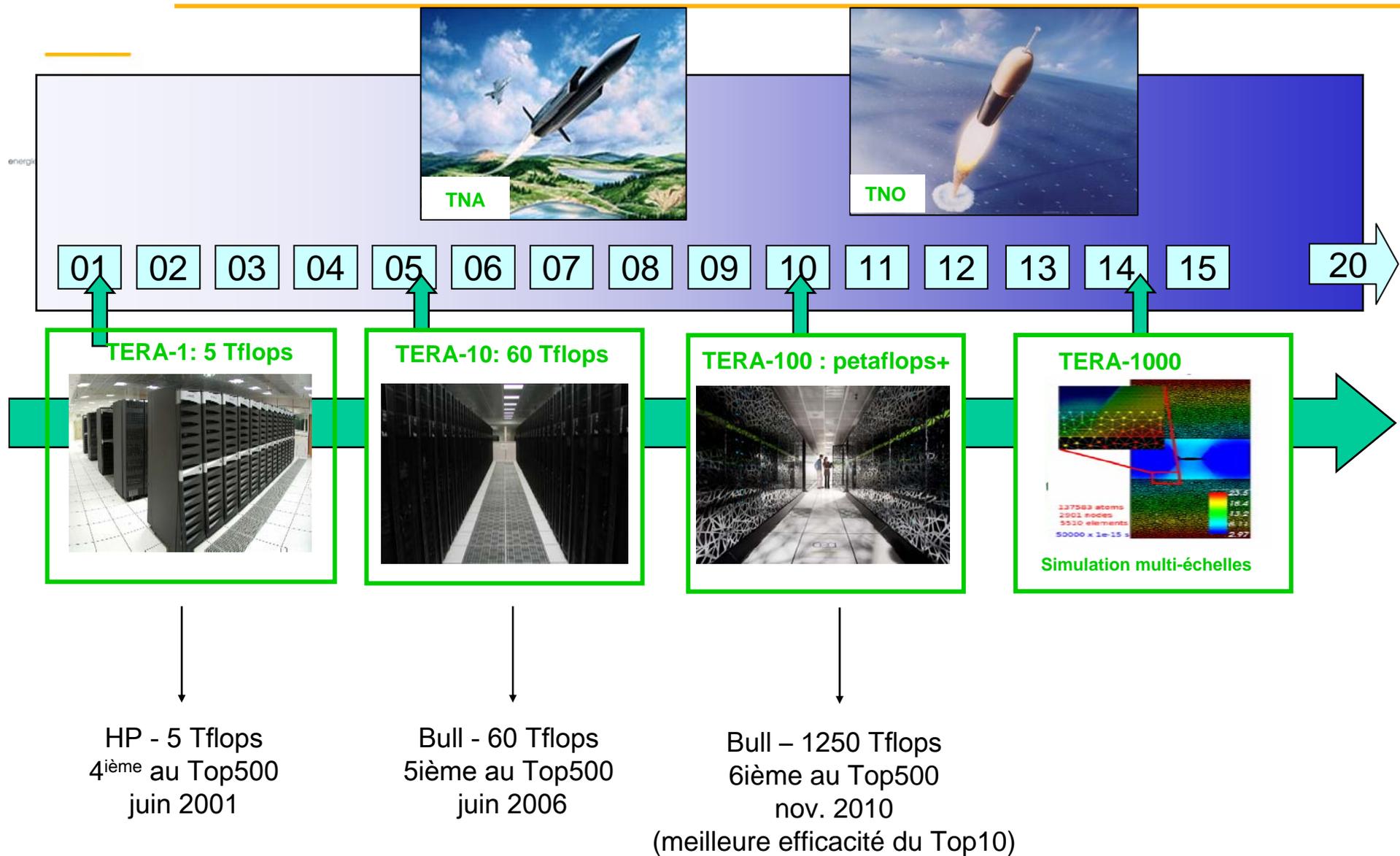
- Préparer en partenariat la génération N+1 alors que la génération N entre en service !
- Prendre en compte les différentes échelles de temps



Le Complexe de Calcul Scientifique du CEA



CEA-DAM : la feuille de route Tera



Tera100 – un partenariat pour le pétaflops « généraliste »



energie atomique • energies alternatives

Thématiques de R-D :

1. Architecture
2. Déploiement et évolutivité
3. Maintien en Condition Opérationnelle
4. Coût Total de Possession
5. Extensibilité
6. Liaison avec les infrastructures

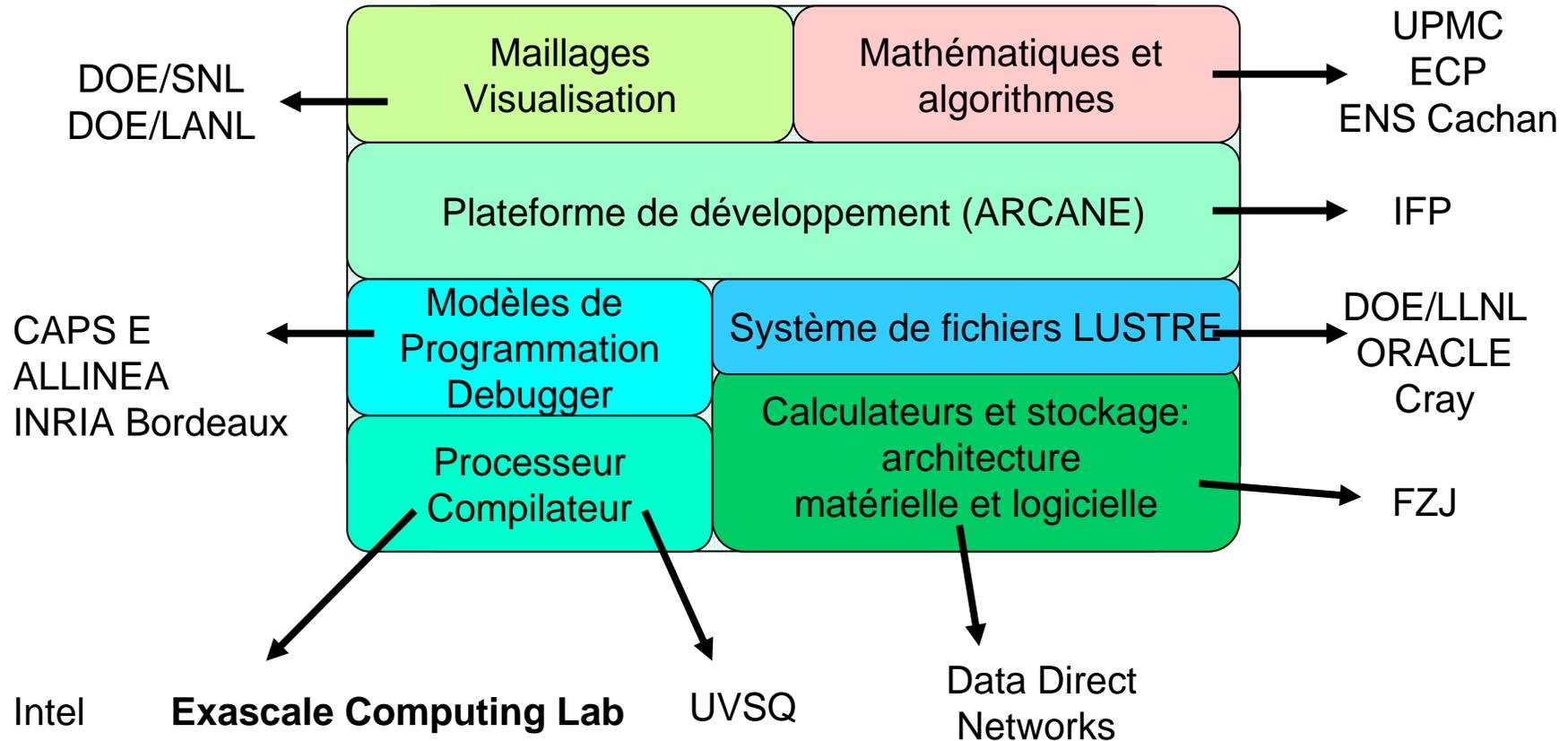
(avec partage de la PI – développements « Open Source » mis à disposition de la communauté - brevet commun)

The screenshot shows a Microsoft Internet Explorer browser window displaying the HPCwire website. The address bar shows the URL: http://www.hpcwire.com/topic/applications/Bull_CEA_Partener_on_Petaflops_Supercomputer.html. The page title is "HPCwire: Bull, CEA Partner on Petaflops Supercomputer". The main content area features the HPCwire logo and a navigation menu with options like Home, News, Features, Blogs, HPC Markets, Whitepapers, and Multimedia. The article title is "Bull, CEA Partner on Petaflops Supercomputer" dated July 29, 2008. The article text mentions that the Military Applications Department (DAM) at the French Atomic Energy Authority (CEA) and Bull have signed a collaboration contract to design and build Tera 100, a future supercomputer. It lists two phases: an initial R&D stage and a second phase for implementation. The article also mentions a significant investment in research and development.

Ecosystème CEA-Bull – vue générale



energie atomique • énergies alternatives



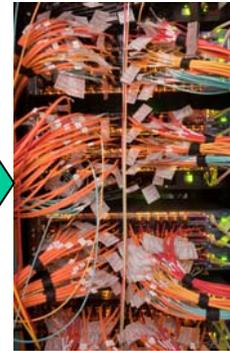
Contributions à : Linux - LUSTRE - MPC - SLURM - OpenIB – UNICORE - VTK

Le principe des poupées russes

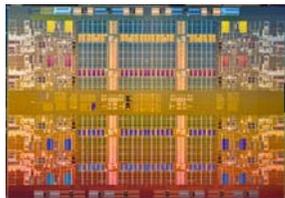


energie atomique • energies alternatives

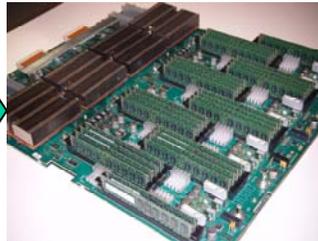
24 serveurs /
armoie



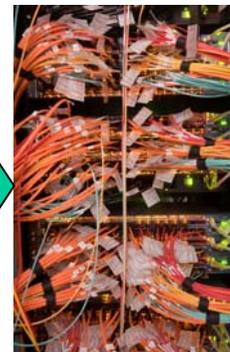
Le calculateur
Tera-100 –
gamme BullX



Processeur
Intel X7560



Serveur Bull



220 armoires



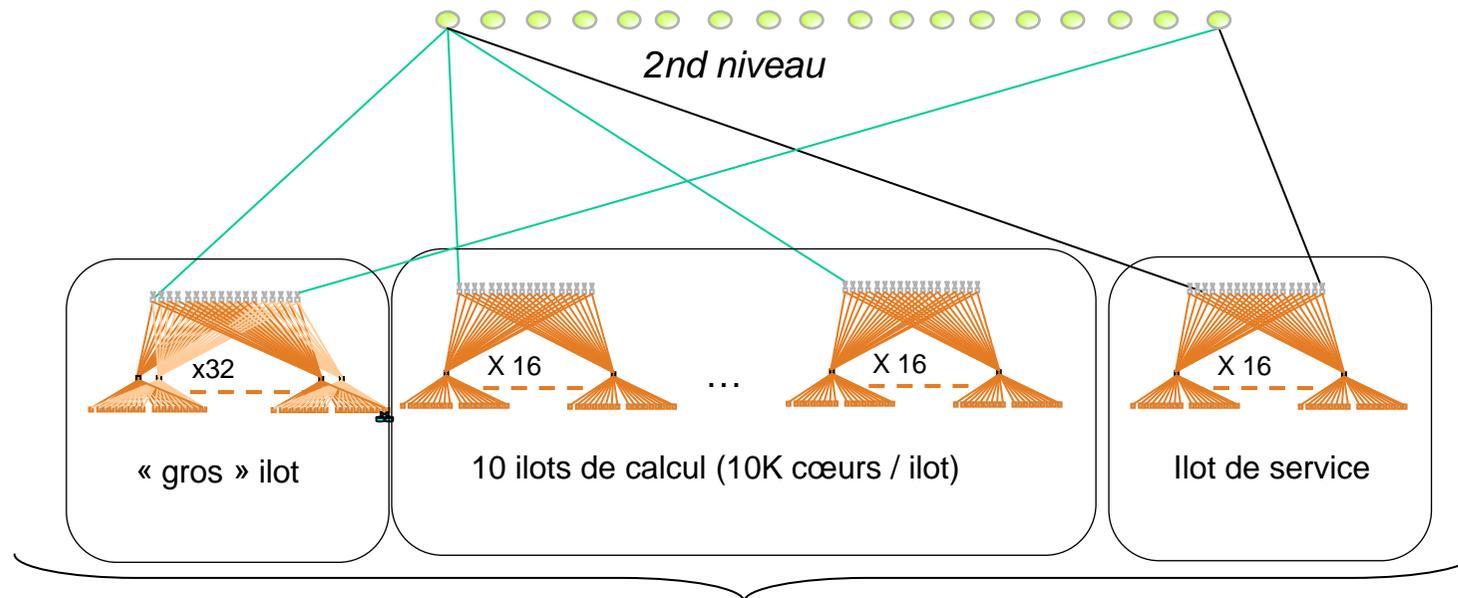
Réseau
d'interconnexion
/ switch

Tera100 : un cluster hiérarchique



energie atomique • energies alternatives

Un réseau d'interconnexion à 2 niveaux



caractéristiques principales

- 1250 Tflops performance max.
- 4 322 nœuds (198 nœuds de service) - SMP - 4 procs (32 cores)
- 17 288 Intel Xeon X7560 procs (Nehalem-EX 8 cœurs @ 2.266 Ghz)
- 138 304 cœurs – 2GO/cœur (5% des nœuds à 4 GO/cœur)
- 291 TO mémoire
- 500 GO/s débit vers les PFS
- 20 PO disques

Tera100 (nov. 2010) - numéro 1 en Europe

Premier supercalculateur européen à passer la barre du Pétaflops soutenu

1,05 Pflops Linpack pour 1,25 Pflops max. (efficacité de 83,7 %)

Le test du *Linpack* (*Linear Algebra Package* - calcul de référence pour le classement au Top500) est un des tests les plus exigeants de la recette du calculateur
Il a mis en jeu toute la machine (processeurs et interconnexion) pendant 22 h en continu



Numéro 6 mondial :
l'architecture la plus
efficace du Top500

Retombées industrielles de Tera100



BullX
« world best HPC server »

**HPCwire Award SC09
conference**



HPCWIRE March 31, 2010

Cray and Bull Team Up on SMP Offering

Nehalem EX has also spurred Bull and Cray to jump on the SMP bandwagon. In this case though, it's a two-for-one deal. Both vendors are using the same Bull-designed hardware for these systems and adding their own branding, software stacks, and custom support on top.. According to Ian Miller, senior vice president of the productivity solutions group and marketing at Cray, they very much liked the Bull technology. And since each company has more or less staked out different geographical territories -- Bull in much of Europe and Cray everywhere else -- the partnership just made sense to everyone involved.



Cray CX1000-C™ Chassis Cray CX1000-SC™ Server

- Two high frequency Intel® Xeon® 5600 series processor
- Large memory capacity (up to 48GB memory per building block)
- One SATA HDD or one SSD drive or diskless
- 1.5U, four socket "Building Block" form factor
- Coherency switch – a proprietary feature based on ability from a single "building block" of 32 cores up with 128 cores in 6U
- Up to 1TB of memory (with 8GB DIMMS)
- Non-uniform memory access (ccNUMA)



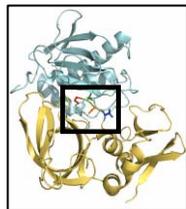
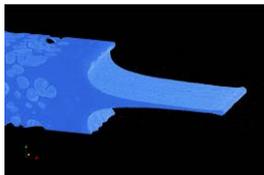
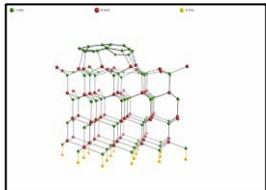
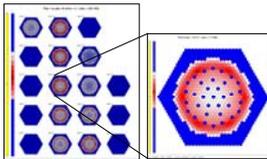
Quelques Grands Challenges TERA 100



energie atomique • energies alternatives

En particulier, profiter de la phase de pré-production de la machine pour passer une sélection de calculs « invités » :

- **Calcul massivement parallèle en transport 3D d'un réacteur rapide sodium avec APOLLO3**
R. Baron, C. Calvin, J. Dubois, J-J. Lautard, S. Salmons
CEA/DEN/DANS/DM2S/SERMA
Jusqu'à 30000+ cœurs
- **Simulation de la croissance de graphène sur le carbure de silicium (BigDFT)**
Thierry Deutsch, Luigi Genovese, Pascal Pochet, Eduardo Machado-Charry
CEA/DSM/INAC – Grenoble
- **Simulations de dynamique moléculaire classique (STAMP)**
L. Soulard et al. , CEA/DAM/DPTA
Jusqu'à 10000 cœurs
- **Conception d'enzymes artificiels (STAMPS)**
 - G. Debret, M. Masella, P. Cuniasse, G. Collet
 - CEA/DSV/iBiTecS

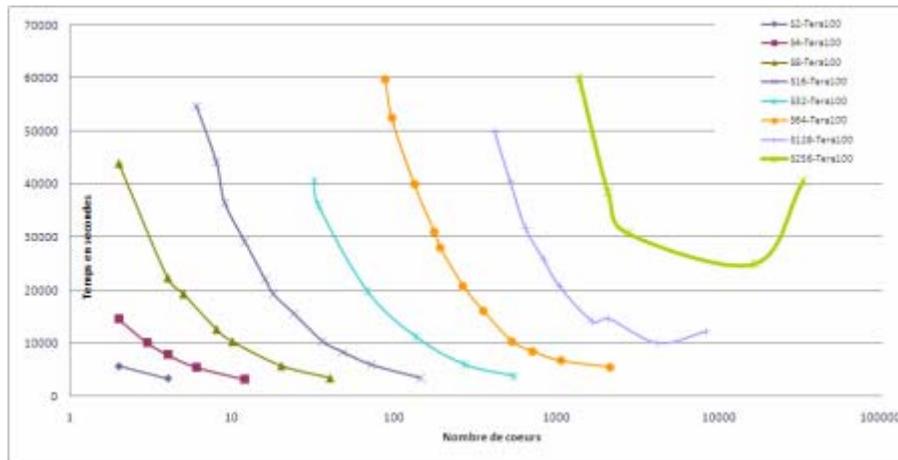
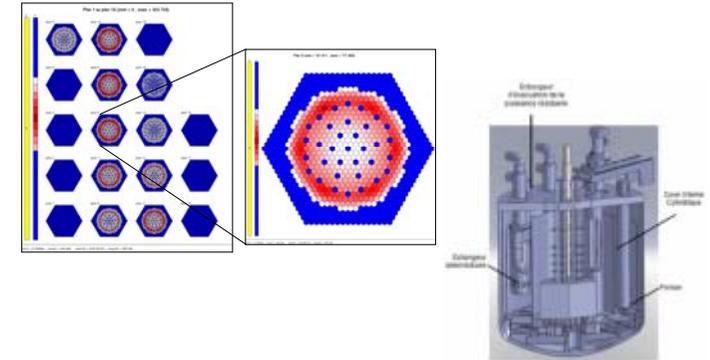


APOLLO3 au Pétaflop sur TERA-100 pour GENIV



energie atomique • energies alternatives

« Calcul massivement parallèle en transport 3D d'un réacteur rapide sodium avec APOLLO3 »

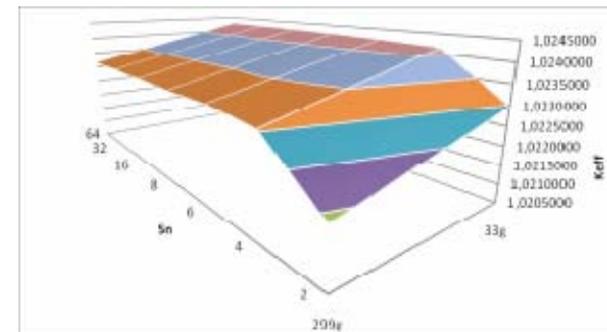


• Détermination d'une solution de référence – Minimisation des incertitudes

- Résolution de l'équation du transport par une méthode Sn (solveur MINARET)
- Ordre de méthode de S2 à S256 → utilisation jusqu'à **33.000** cœurs de calcul

• Etude paramétrée – corrélation entre options de calcul et précision

- 48 cas de calcul sur **6.000** cœurs de calcul en 12h (6 mois en séquentiel)

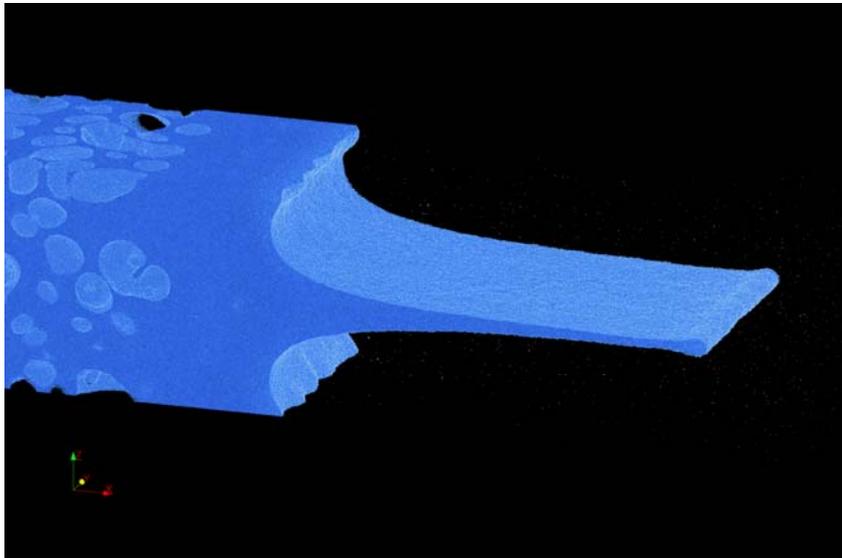


Dynamique moléculaire classique sur TERA-100

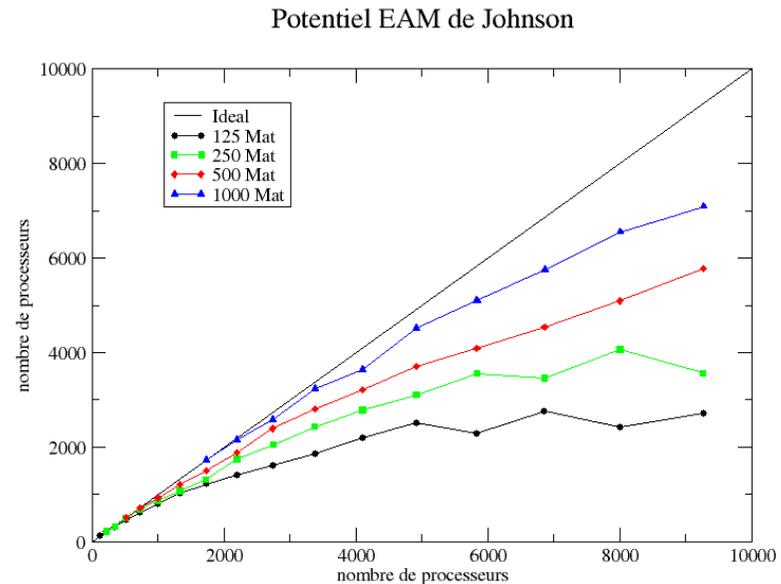


energie atomique • energies altern

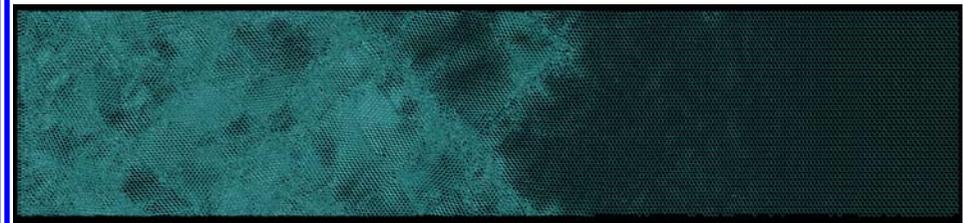
La dynamique résout l'équation fondamentale de la dynamique pour un ensemble de N particules en interaction mutuelle. Pourvu que N soit suffisamment grand, elle peut permettre de faire le lien entre le calcul quantique et la mécanique des milieux continus.



Ejection de matière lors de la réflexion d'un choc sur une surface libre. 125 millions d'atomes, 4000 cœurs, potentiel interatomique assez simple mais temps physique de simulation très long (pour la dynamique moléculaire): de l'ordre de la nanoseconde.

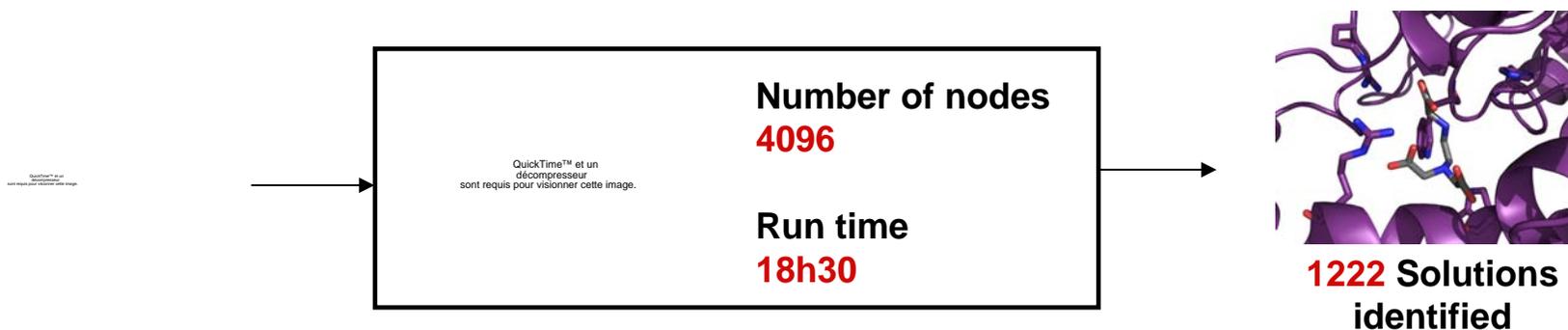


Performances du code Stamp sur TERA-100 entre 250 millions et un milliard d'atomes, et jusqu'à 10 000 cœurs.



Plastification du diamant au passage d'une onde de choc. Relativement peu d'atomes (1,3 millions, 6400 cœurs) mais potentiel interatomique très complexe.

Design of artificial enzymes



Cavazza et al.
Nature Chem (2010)

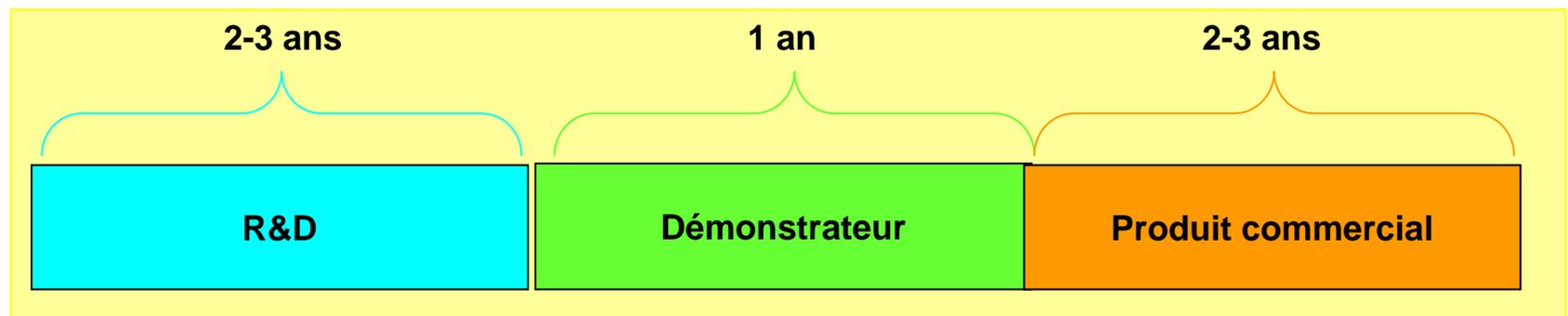
Applications : **Biofuel, Green Chemistry...**

Conclusion



énergie atomique • énergies alternatives

- La compétition s'élargit et de nouveaux pays/acteurs apparaissent
- Il est possible de retrouver des degrés de liberté dans ce domaine tout en étant compétitif sur le plan mondial
- Mais, à l'instar des USA, de la Chine, de la Russie ou du Japon, il faut sortir de la relation client-fournisseur
- Une vision à plus de 5 ans est nécessaire pour définir un programme de R-D en partenariat public-privé



Déclinaison DARPA : le programme HPCS



energie atomique • energies alternatives

- Soutien pluriannuel à la R-D industrielle pour le développement de systèmes « pétaflopiques » économiquement viables
- Phase I (2002) : étude des concepts (IBM, Cray, SUN, SGI) – 20 M\$
- Phase II (2003-2005) : prototypes (IBM, Cray, SUN) – 170 M\$
- Phase III (2006-2010) : développement de systèmes – (~500 M\$)
 - Annoncés en novembre 2006 : un contrat de 250 M\$ à la société Cray et un second de 244 M\$ à la société IBM

Budget HPCS * : ~80 M\$ + / an

** Complété par du budget en provenance de NSA et du DoE pour financer la Phase III du programme*



DARPA : le programme UHPC (2010 – 2014)



énergie atomique • énergies alternatives

3 consortia financés sur 4 ans (~25 M\$ chacun) pour le développement de

briques technologiques en vue de l'Exaflops :

- NVIDIA, CRAY, Oak Ridge lab. et 6 universités
- Intel, SGI, U. Delaware, U. Illinois ...
- Sandia lab., Micron Technology, LexisNexis Special Services

Les applications visées sont celles du DOD :

- acquisition et traitement des masses de données
- analyse de graphes
- décision et planification
- simulation numérique (aéronautique, radar..)

En août dernier la DARPA vient de lancer un nouvel appel à projets pour compléter UHPC : le programme Omnipresent HPC (OHPC)

Déclinaison DoE - ASCR : une vision globale



energie atomique • energies alternatives

- **Une vision à 20 ans énoncée en 2003**
- Programme « **Scientific Discovery through Advanced Computing** » (SciDAC-2 2007-2011)
 - 17 projets « applicatifs » (développer les grands codes de simulation – 26,1 M\$)
 - 9 Centres Technologiques (développer les outils pour les applications : bibliothèques, visualisation, gestion des données.. – 24,3 M\$)
 - 4 Instituts pour la formation impliquant 13 Universités (8,2 M\$)
- Soutien à la recherche en «computer science »
- Préparation de l'avenir : la simulation à l'échelle de l'Exaflops avant 2020
- Mise en place de grands moyens/centres de calcul (NERSC, Argonne, Oak Ridge)

Budget ASCR : 350 M\$ / an



DoE-NNSA : capacity vs capability clusters



energie atomique • energies alternatives

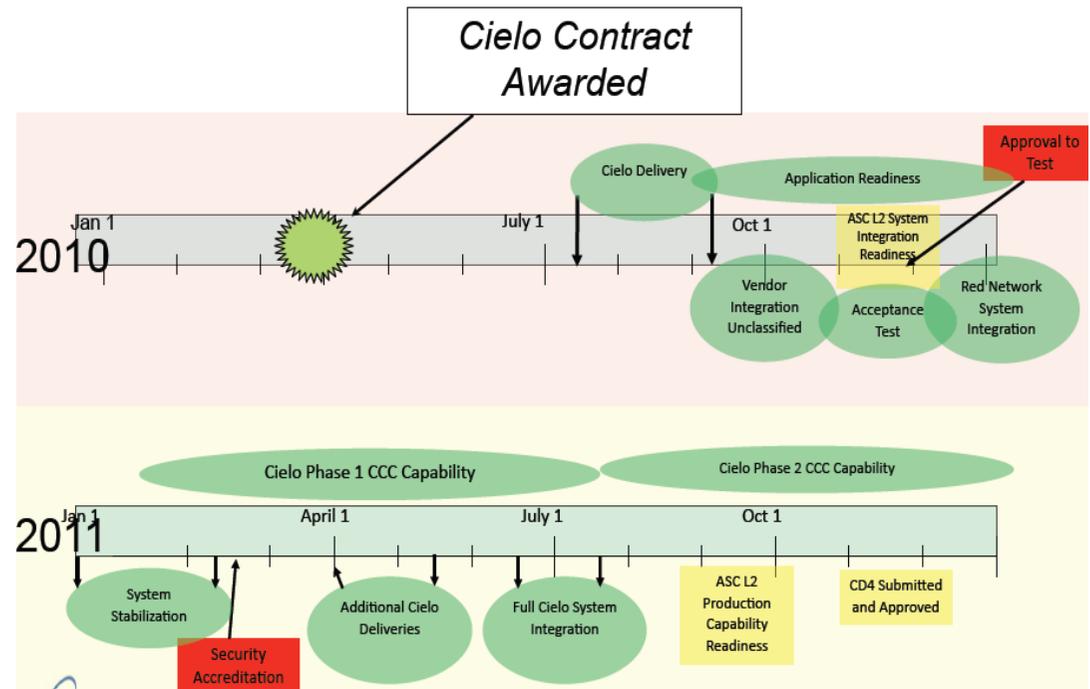
« Capacity »: la ressource cachée

- LLNL (2009) : 250 Teraflops répartis sur 4 clusters en environnement classifiés.
- Résultat d'un contrat global au 3 labs, passé en 2007, pour l'achat de 8 clusters (total de 436 Tflops).
- Un appel d'offres de renouvellement des « capacity clusters » sur les 3 labs sera lancé d'ici la fin de l'année.

« Capability »: ceux que l'on verra dans le Top 500

- LLNL : 20 Pflops en 2012 (projet IBM/SEQUOIA)
 - 1,5 M cœurs
 - 3^{ème} génération de machine BlueGene
- LANL et SNL : 1,03 Pflops (2010) puis 1,37 Pflops (2011) – CRAY/Cielo

*En cours de définition :
un programme DoE pour l'Exaflops*



Déclinaison NSF – Office of Cyberinfrastructure



energie atomique • énergies alternatives

- Un effort particulier pour la mise en place d'une infrastructure de calcul – environ **330 M\$ / an** (y compris les applications et la recherche)
- Fédération des ressources par une « grille » (TeraGrid)

Annoncé en août 2007 :

208 M\$ sur 4,5 ans pour un système multi-Petaflops installé à UIUC-NCSA et opérationnel en 2011.

Machine IBM Blue Waters, issue des travaux financés par la DARPA.

Etat à fin 2010 : cluster de Power7, 10 Pflops prévus pour 2012



Université de Moscou – la salle machine et le calculateur



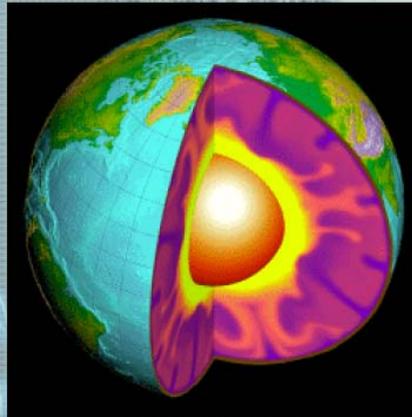
U. Moscou: exemples d'applications



energie atomique • energies alternatives

3D nonstationary magnetic hydrodynamics

Studying magnetic field generation in the Earth core. One of the results is explaining the change of orientation of Earth magnetic field.



Cryptography-oriented number theory problems

Factorization of large integer numbers requires ultra-teraflop processing power.
Factorization: $N = p q$, where p and q are prime numbers.

$N=2799783391122132787082946763872260162107044678695$
542853756000992932612840010760934567105295536085606
182235191095136578863710595448200657677509858055761
3579098734950144178863178946295187237869221823983

$p=35324619344027701212726049781984643686711974001976$
25023649303468776121253679423200058547956528088349
 $q=79258699544783330333470858414800596877379758573642$
19960734330341455767872818152135381409304740185467

