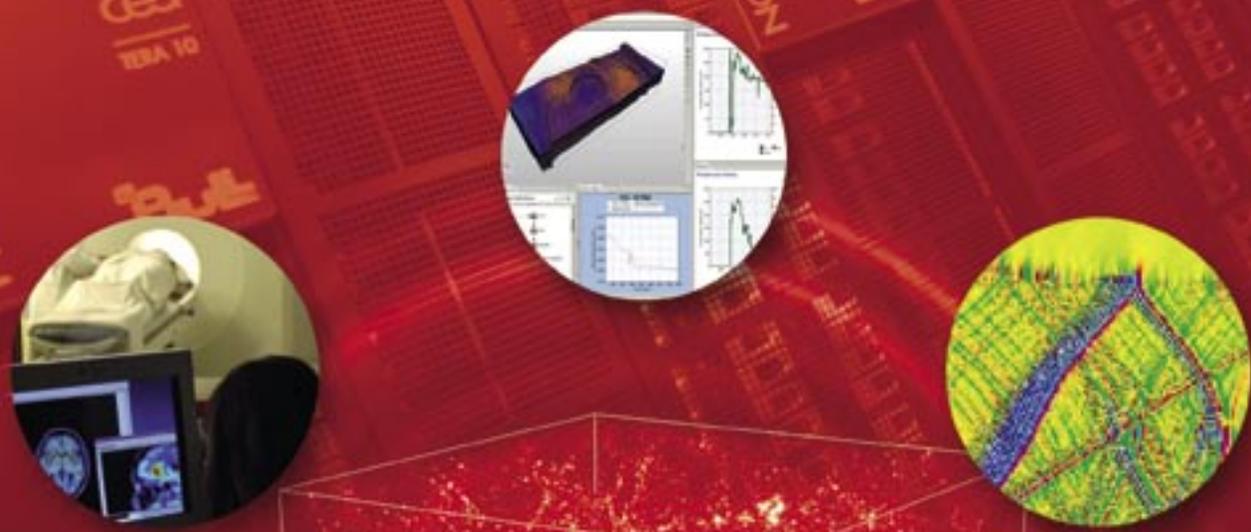


La RECHERCHE



Les **défis** des supercalculateurs



Méthodes Numériques pour les Modèles des Milieux Continus

<http://www.cmla.ens-cachan.fr/master/mn2mc>

MASTER RECHERCHE 2^e Année



La simulation numérique, sortie des laboratoires de recherches universitaires, est désormais omniprésente dans tous les secteurs de l'économie : industrie mais aussi services. Autant les grandes entreprises que les PME et PMI (Bureaux d'études, entreprises de hautes technologies, ...) sont concernées.

De nombreux débouchés, tant dans les organismes de recherche publique (Universités, INRIA, CNRS, CEA, ...) que dans les équipes de Recherche et Développement des entreprises de toutes tailles, s'offrent aux docteurs issus de la formation doctorale dont dépend ce master. A ceci s'ajoutent de nombreuses opportunités à l'international.



Ce parcours propose une formation avancée en méthodes numériques pour les modèles des milieux continus. Il propose d'étudier de nombreux aspects de la recherche actuelle en méthodes numériques, en analyse asymptotique, en mécanique des fluides numérique, en mécanique non linéaire des solides, en optimisation... Il est ouvert aux étudiants en Master 1^{ère} année des disciplines scientifiques (Universités et Grandes Ecoles).

Un stage long en général en entreprise permet d'appliquer et d'approfondir les connaissances acquises.

Possibilités (sous conditions) de financement au cours de la formation.
Inscriptions de mai à septembre : formulaire disponible sur le site du master

Les défis des supercalculateurs



ÉDITORIAL par Bernard BIGOT P. 3

SUPERCALCULATEURS P. 6

Course à la puissance de calcul :
trois **révolutions** en marche



INTERVIEW Jean Gonnord P. 10

« **2007** sera l'année
du calcul intensif en Europe »

AVIONIQUE P. 12

Simulation aérodynamique :
une révolution permanente

GESTION DE BASE DE DONNÉES P. 15

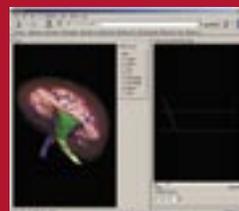
Xedix, un système de
gestion de... masses de données

MOTEUR D'HÉLICOPTÈRE P. 18

Maîtriser les turbulents
phénomènes de **combustion**

SANTÉ P. 20

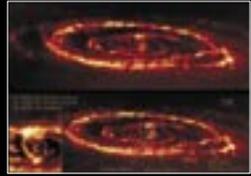
Cancérologie : première
simulation complète d'un examen



PORTFOLIO

Nombres et lumières dans le ciel

La simulation numérique est devenue un complément indispensable de l'observation pour les astronomes. Ils peuvent d'expérimenter l'évolution d'un système, sans perturbation de leur part, de résultats souvent obtenus dans le cadre de programmes de grands relevés astronomiques, tel celui des chercheurs du Service d'Astronomie de l'Observatoire de Paris et de ses partenaires (Université Paris 7, CNRS, Observatoire de Paris). Ces outils numériques nous donnent un aperçu de la contribution de la simulation à l'avancement des connaissances en astronomie à travers les siècles.



Andromède se remet tout juste d'une collision

Andromède est la galaxie spirale la plus proche de la nôtre. La dernière fois qu'elle a subi une collision avec une autre galaxie, elle le fut il y a environ 250 millions d'années. Elle est aujourd'hui en train de se remettre de cette collision. Elle est en train de se remettre de cette collision. Elle est en train de se remettre de cette collision.

DÉTONIQUE

Explosion sur ordinateur

EXPLOITATION PÉTROLIÈRE

Décrypter les mystères du sous-sol

ESSAIS NUCLÉAIRES

La simulation pour surveiller les explosions nucléaires

SANTÉ

Révéler les dessous du « mauvais » prion

INTERVIEW Philippe Miltin

« Les supercalculateurs, un gage de souveraineté et de compétitivité »

TECHNOPOLE

ter@tec, Une référence du calcul intensif en Europe

TRIBUNE François Goulard

WEB

P. 22

P. 30

P. 32

P. 35

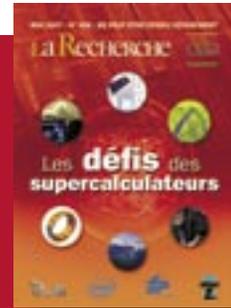
P. 36

P. 38

P. 40

P. 41

P. 42



Société d'Éditions Scientifiques

Bernard Wouts, directeur de la publication
Stéphane Khémis, président du comité éditorial
Olivier Postel-Vinay, conseiller de la direction

74, avenue du Maine - 75014 Paris

Tél. : 01 44 10 10 10

e-mail rédaction : courrier@larecherche.fr

ABONNEMENTS/ANCIENS NUMÉROS/RELIURES

Adresse e-mail : la_recherche@presse-info.fr

La Recherche Service Abonnement
B 604, 60732 Sainte-Geneviève Cedex

Tarif France : 1 an 11 n°, 52,60 € ;

1 an 11 n° + 4 hors-série, 72 €

Tarif international : nous contacter.

Suisse : Edigroup, Case postale 393, 1225 Chêne Bourg.

Belgique : Edigroup, Bastion Tower-Etage 20,

place du Champ-de-Mars, 5, 1050 Bruxelles

Canada : Express Mag, 8155, rue Larrey, Anjou Québec H1J 2L5.

Directeur scientifique Jean-Michel Ghidaglia

Rédactrice en chef Aline Richard

Assistante Juliette Quenin (44 33)

Rédacteur en chef adjoint

Luc Allemand (44 23)

Rédactrice en chef adjointe du cahier 2 : Isabelle Bellin

Directeur artistique du cahier 2 : Philippe Lewit

Iconographie du cahier 2 : Julie Épessé

Maquettiste du cahier 2 : Carol Müller

Secrétaire de rédaction du cahier 2 : Stéphane Demazure

Correction du cahier 2 : Catherine Legrand

Conception et réalisation du Cahier 2

Éditorial et Régie :

22, rue Chauchat 75009 Paris - Tél. 01 53 34 12 22

Ont collaboré à ce numéro

Isabelle Bellin, Didier Gout, Léo Gerat,

Dominique Ritman, Anne Lefèvre-Balleydier,

Pierre Vandeginste, Yves Sciana.

Directeur délégué : Frédéric Texier

Marketing direct et abonnements

Directrice : Virginie Marliac (44 16)

Chargé du marketing : Arnaud Cobo

Diffusion (diffuseurs/dépositaires)

Céline Balthazard (44 11) n° Vert : 0 800 30 76 02

Responsable gestion : Isabelle Parez (44 95)

Comptabilité : Marie-Françoise Chotard (01 44 10 13 43)

Diffusion librairies

DIFPOP. Tél. : 01 40 24 21 31 Fax : 01 40 24 15 88

Chef de projet développement : Grégory Luneau

Webmaster : Lymédias (01 53 36 06 01)

Fabrication : Christophe Perrusson (44 64)

PUBLICITÉ : Le Point Communication (44 54)

Chef de publicité (marché scientifique)

Laurent Allègre (44 57)

Assistante commerciale et technique

Françoise Hullot (f.hullot@interdeco.fr)

La Recherche est publiée par la SES, filiale

de La Financière Tallandier. Président-directeur général

et directeur de la publication : Bernard Wouts.

Directeur général : Stéphane Khémis.

Les titres, les intertitres, les textes de présentation et les légendes sont établis par la rédaction du

mensuel. La loi du 11 mars 1957 interdit les copies ou reproductions destinées à une utilisation

collective. Toute représentation ou reproduction intégrale ou partielle faite sans le consentement

de l'auteur, ou de ses ayants droit ou ayants cause, est illicite (article L122-4 du Code de propriété

intellectuelle). Toute copie doit avoir l'accord du Centre français du droit de copie (CFC, 20, rue des

Grands-Augustins, 75006 Paris. Tél. : 01 44 07 47 70. Fax : 01 46 34 67 19). L'éditeur s'autorise à refuser

toute insertion qui semblerait contraire aux intérêts moraux ou matériels de la publication.

Cahier 2 de La Recherche - Commission paritaire : 0909 K05863 ISSN 0029-5671

Impimerie Canale, Borgaro (Italie) Dépôt légal 2^e trimestre 2007

© 2007 Société d'Éditions Scientifiques. IMPRIMÉ EN ITALIE. PRINTED IN ITALY

Course à la puissance de calcul :

Aujourd'hui, que ce soit pour les besoins de la recherche ou de l'industrie, simulation rime souvent avec supercalculateur. Trente ans après l'avènement du calcul intensif, le paysage a bien changé, tant du côté des architectures matérielles que logicielles. Et ce n'est pas fini...

Pierre Leca est chef du département sciences de la simulation et de l'information CEA/DAM Île-de-France

Une nouvelle discipline scientifique est née dans les années 1970, avec l'arrivée des calculateurs CRAY : la « simulation numérique par le calcul intensif » – ou HPC pour High Performance Computing – à la croisée des chemins entre informatique, modélisation physique et mathématiques. Depuis, le calcul intensif a pris une place essentielle dans la plupart des domaines de la recherche et de l'industrie. C'est désormais un secteur capital, voire stratégique, pour certains pays qui soutiennent des politiques à long terme.

Un constat inquiétant pour l'Europe

La lecture du Top 500, la liste des 500 plus puissants calculateurs au monde remise à jour tous les 6 mois depuis 1993, en dit long sur l'écrasante domination américaine dans le domaine du calcul intensif : en novembre 2006, les États-Unis possédaient plus de 60 % des supercalculateurs et autant de la performance totale. L'Europe, quant à elle, ne détenait qu'environ 20 % de la puissance de calcul.

Ce constat est inquiétant pour la compétitivité de notre recherche et de notre industrie, d'autant plus que la part du secteur industriel ne cesse d'augmenter : elle est passée de 20 % en 1993 à près de 50 % en 2006. De fait, le calcul intensif est de moins en moins réservé au secteur public (recherche, défense, organismes gouvernementaux...) comme ce fut le cas dans un passé proche.

Mais le plus frappant dans l'analyse des évolutions du Top 500 est la croissance exponentielle, non démentie depuis l'origine, de la puissance de calcul que représente l'ensemble de ces 500 calculateurs. En effet, la puissance installée a été plus que

décuplée tous les 4 ans, passant de 1 téraflops⁽¹⁾ (soit 10^{12} opérations/sec) en 1993 à plus de 3 pétaflops⁽²⁾ (soit 3×10^{15} opérations/sec) en 2006. Cette croissance est supérieure à ce que l'on pouvait attendre des conséquences de la loi de Moore (cette « loi » s'applique au microprocesseur et prévoit que le nombre des transistors sur une puce est multiplié par deux tous les 18 mois ; nous y reviendrons). Grâce à l'augmentation de la fréquence des processeurs, elle aurait seulement dû conduire à une croissance d'un facteur 8 tous les 6 ans.

D'où provient cette accélération de puissance des calculateurs, supérieure à la puissance des microprocesseurs qui les composent ? Du développement des architectures parallèles, autrement dit de l'interconnexion et de la coordination de plusieurs processeurs au sein du calculateur. Ainsi d'une moyenne d'environ 100 processeurs par machine du Top 500 en 1993, on est passé à une moyenne bien supérieure à 1 000 en 2006.

Un déluge de données

Cependant, considérer la croissance de la puissance de calcul du Top 500 ne suffit pas à appréhender toutes les évolutions en matière de calcul intensif. En particulier, cela ne rend pas compte de l'accroissement phénoménal des flots de données à traiter. Deux nouveaux facteurs sont en effet à l'œuvre depuis quelques années. D'une part, la réduction du coût des supports de stockage de l'information : un disque de 1 téraoctets (10^{12} octets) coûte aujourd'hui moins de

En novembre 2006, les États-Unis possédaient plus de 60 % des supercalculateurs et autant de la performance totale.

1 000 euros, ce qui rend économiquement viable la mise en place de dispositifs permettant de stocker plusieurs pétaoctets (10^{15} octets). Pour fixer un ordre de grandeur, la numérisation de la Bibliothèque François

trois **révolutions** en marche



Mitterrand ne représenterait que 30 à 40 téraoctets. D'autre part, l'évolution des applications du calcul intensif aboutit à une consommation et une production de volumes de données considérables. À titre d'exemple, les simulations effectuées sur le supercalculateur Tera-10 du CEA/DAM, qui mettent en œuvre des modèles tridimensionnels de phénomènes stationnaires, produisent depuis un an plus de 10 téraoctets de données par jour. Dans un autre domaine d'utilisation, l'analyse des données produites par de grands instruments,

comme le détecteur *Babar* de l'accélérateur linéaire de Stanford (États-Unis), conduit au traitement de 1 pétaoctets de données.

Loi de Moore et architecture des calculateurs

Mais l'accroissement de la performance a aussi des conséquences de plus en plus préoccupantes : elle a conduit à une augmentation de la puissance électrique consommée et de la chaleur dissipée. Un microprocesseur de dernière génération peut consommer

Le supercalculateur Tera-10 développe une puissance de plus de 50 téraflops. Cette machine est constituée de 4352 processeurs double cœur Intel® Montecito reliés par un réseau d'interconnexion haute performance Quadrics.

Pierre Leca,
chef du
département
sciences de la
simulation et de
l'information
CEA/DAM
île-de-France.

jusqu'à 130 watts, ce qui se traduit par un dégagement de chaleur équivalent à celui d'une grosse ampoule électrique sur quelques centimètres carrés ! Cet accroissement de la consommation électrique et donc du dégagement de chaleur par puce ne peut se poursuivre faute de dispositif d'évacuation de la chaleur. Cela a conduit depuis quelques années à une remise en question fondamentale de l'architecture des microprocesseurs au cœur de tous les systèmes, avec des conséquences considérables.

En effet si, d'après les technologues, la loi de Moore n'est pas remise en cause pour les 10 ans à venir, ce sont les conséquences qui doivent être revues. Jusqu'à présent, la dynamique d'amélioration de la performance était obtenue par l'augmentation de la fréquence d'horloge, associée à l'optimisation de l'exécution des instructions et la mise en œuvre de mémoires caches. Cet accroissement « gratuit » de la performance n'est aujourd'hui plus de mise. Diminuer la finesse de gravure des transistors se traduit depuis deux ans par la mise en œuvre simultanée de « cœurs multiples » au sein du microprocesseur (*i.e.* plusieurs processeurs sur la puce) et de la capacité à gérer simultanément plusieurs flots d'exécution (mécanisme d'« *hyperthreading* »). Ainsi, il faut considérer qu'à présent la loi de Moore va se traduire au cours du temps davantage par l'augmentation du nombre de cœurs que par l'augmentation de la fréquence.

La révolution du parallélisme à tous les étages

Depuis novembre 1997 il n'y a plus eu de machine monoprocesseur dans la liste du Top 500 et dès cette date, l'ensemble de la communauté scientifique du calcul intensif a dû prendre en considération la contrainte du parallélisme. Dix ans plus tard, avec la diffusion généralisée des microprocesseurs multi-cœurs dans tous les équipements informatiques, du PC au supercalculateur, c'est l'ensemble de la communauté du logiciel qui doit impérativement tenir compte de ce nouveau paradigme.

De plus, la non-linéarité dans la relation entre fréquence et consommation électrique peut encore accroître cette dynamique. L'exemple de l'architecture BlueGene d'IBM est révélateur de cette tendance poussée à l'extrême. Le constructeur américain a choisi d'interconnecter un très grand nombre de processeurs moins puissants que la technologie ne le permettrait. En effet, BlueGene ne comporte pas moins de 130 000 processeurs à 700 mégahertz, ce qui lui permet d'atteindre plus de 360 téraflops pour



© C. DUPONT/CEA

une consommation de seulement 1,5 mégawatt. Un autre exemple plus représentatif de supercalculateurs généralistes est la gamme des calculateurs Tera installés au CEA/DAM dans le cadre du programme Simulation : le calculateur Tera-1 installé en 2001 possédait 2 500 microprocesseurs mono-cœur pour une puissance de calcul de 5 téraflops. Pour atteindre une puissance 12 fois supérieure, le calculateur Tera-10 construit par Bull, installé 4 ans plus tard, utilise près de 5 000 microprocesseurs bi-cœurs, soit près de 10 000 cœurs de calcul.

La révolution du logiciel et des algorithmes

Tout indique qu'une machine d'une puissance de 1 pétaflops pourra être construite avant 2010 : ce calculateur utilisera entre 100 000 et 1 000 000 de cœurs selon le compromis qui sera effectué entre fréquence du microprocesseur et consommation électrique. De plus ces dizaines de milliers de cœurs seront disposés dans le calculateur de manière hiérarchique, tout d'abord au sein du microprocesseur, puis à l'intérieur d'un multiprocesseur à mémoire partagée, enfin au sein de grappes de multiprocesseurs.

À titre d'exemple voici la configuration possible d'un tel supercalculateur en 2010 : à partir d'un microprocesseur à 4 cœurs cadencés à 2 gigahertz, capables d'effectuer 4 instructions par cycle, on obtient, en intégrant 16 de ces microprocesseurs au sein d'un multiprocesseur, une puissance de calcul de 0,5 téraflops. L'interconnexion de 2 000 de ces multiprocesseurs permettrait d'obtenir une machine de 1 pétaflops utilisant 128 000 cœurs. Il s'agit bien entendu là d'un potentiel de puissance qui ne sera exploitable qu'au prix d'une adaptation des algorithmes et des méthodes numériques.

Un microprocesseur de dernière génération peut consommer jusqu'à 130 watts

Une adaptation imposée par ces nouvelles architectures matérielles, comme cela a toujours été le cas : rappelons-nous les évolutions induites au cours des années 1980 par l'avènement des premiers calculateurs vectoriels.

Pour certains, la diffusion du parallélisme dans toute la pyramide de l'informatique représente le plus grand changement depuis la révolution de la programmation orientée objets. Pour les spécialistes du calcul scientifique qui seront au cœur de la tourmente, la prise en compte de contraintes multiples telles que parallélisme massif, hiérarchie dans le calculateur, gestion des flots de données, va conduire obligatoirement à la composition de plusieurs approches tant du point de vue des méthodes numériques que des modèles de programmation ou des architectures logicielles.

Combien cela pourrait-il coûter de posséder un tel supercalculateur en considérant une durée d'utilisation de 5 ans ? Basons-nous sur un coût d'acquisition de 80 millions d'euros. Une estimation de la consommation électrique est obtenue de la manière suivante : tout d'abord les 32 000 microprocesseurs contribueront pour près de 4 mégawatts. Il faut y ajouter la consommation de la mémoire et des disques estimée à environ 3 mégawatts si l'on veut une architecture équilibrée. Soit un total de 7 mégawatts à multiplier par un facteur 2 pour tenir compte de la consommation des équipements de refroidissement et d'autres équipements complémentaires. Au coût actuel de l'électricité, cela représente 7 millions d'euros par an, soit 35 millions d'euros pour 5 ans d'utilisation. Ajoutons à cela le budget de maintenance, estimé sur la base du coût d'acquisition à 25 millions d'euros. Finalement en comptabilisant d'autres coûts annexes, tels que ceux relatifs à la nécessaire adaptation de la salle machine et des infrastructures, une estimation réaliste démontre que le simple fait de faire fonctionner un tel supercalculateur durant 5 ans a un coût proche de son coût d'acquisition.

La révolution des communautés scientifiques : participer à la course ou mourir

Ce « coût de possession » mais aussi tous les bouleversements que ce type de calculateurs introduit dans la communauté du calcul intensif, en particulier l'intégration de la puissance de calcul avec la gestion du flot de données et la nécessité de capitaliser la connaissance dans du logiciel pérenne (« *les calculateurs passent, le logiciel reste* »), conduisent à considérer la mise en place de ces grands moyens de calcul, leur utilisation et leur administration comme celle des très grands équipements. Les communautés

d'utilisateurs devront s'organiser en équipes pluridisciplinaires autour de ceux-ci.

De plus, les avancées de la simulation promises par les ordinateurs de grande puissance ne pourront être obtenues qu'au prix d'une maîtrise de la complexité : complexité de la modélisation, complexité des méthodes mathématiques et algorithmiques, complexité de l'informatique, des techniques de vérification, de validation et de production du logiciel. Ce qui fera demain les leaders du domaine, c'est une telle approche intégrée et c'est le véritable défi auquel les communautés du calcul en Europe sont aujourd'hui confrontées.

Ce travail exige la collaboration d'ingénieurs et de chercheurs spécialistes de la modélisation, des mathématiques, du génie logiciel et des architectures informatiques. Le domaine du calcul intensif est en effet à la croisée de plusieurs disciplines, un domaine où l'existence et l'accroissement d'une base industrielle et technique sont fondamentaux pour en anticiper les évolutions.

Le succès du projet ayant abouti à l'installation du calculateur Tera-10 au CEA/DAM, la montée en puissance de l'association Ter@tec (lire l'article de Ter@tec p. 40), la dynamique du pôle de compétitivité System@tic et l'inscription de la thématique du calcul intensif dans le 7^e Programme cadre de recherche de la Communauté européenne, ont créé un contexte favorable pour enclencher un cycle vertueux impliquant recherche, développement, innovation et objectifs industriels. À l'instar des États-Unis, de la Chine ou du Japon, il existe maintenant en Europe tous les éléments pour relever les nouveaux défis du calcul intensif.

P. L.

Flops :
acronyme de Floating Point Operations Per Second, unité de vitesse de traitement d'un processeur, exprimée en nombre d'opérations par seconde.

(1) Un téraflops : mille milliards d'opérations par seconde.

(2) Un pétaflops : un million de milliards d'opérations par seconde.

Liens Web

Top 500 : www.top500.org
Les nouvelles du HPC : www.hpcwire.com
www.hoise.com

Le supercalculateur Tera-10, construit par Bull, a été livré au CEA/DAM fin 2005.



© C. DUPONT/CEA

Jean Gonnord : « 2007 sera l'année du calcul intensif en Europe »

Questions à Jean Gonnord, chef du projet simulation numérique et informatique au CEA/DAM Île-de-France. La France mais aussi l'Europe seraient-elles en passe de rattraper leur retard en matière de calcul intensif ?

La Recherche. Revenons d'abord sur l'année 2006, une année riche en événements en matière de calcul intensif au CEA...

Jean Gonnord. Oui. Une année jalonnée de succès : le supercalculateur Tera-10, le plus puissant en Europe (60 téraflops), premier à avoir été conçu par un constructeur européen (Bull), a été livré à l'heure, fin 2005, et mis en service opérationnel en avance sur le planning, en juin 2006. Dédié au programme Défense, Tera-10 s'est placé d'emblée en 5^e position du Top 500⁽¹⁾. Six mois plus tard, au Top 500 de novembre 2006, la machine n'était déjà plus qu'au 7^e rang mondial : c'est dire la course à la puissance. Néanmoins, notre vocation n'est pas de battre des records, mais de faire des calculs ! Nous avons évalué nos besoins en puissance de calcul sur la base de l'amélioration progressive attendue de nos modèles, soit un facteur dix tous les quatre ans. Le programme d'évolution de nos calculateurs suit cette progression : Tera-1 (5 téraflops) installé fin 2001 par HP a cédé la place à Tera-10 (60 téraflops). Tera-100 (0,5 pétaflops) sera installé en 2010. En 2018, nous prévoyons une capacité de 50 pétaflops.

L. R. Excepté le calculateur japonais Earth Simulator, les plus grosses machines mondiales sont clairement réservées à la défense. Cela va-t-il continuer ?

J. G. Effectivement, suite aux décisions d'arrêt des essais nucléaires, les programmes de défense ont été les premiers à porter la R&D dans le domaine du calcul intensif, avec des simulations numériques d'envergure. Mais la plupart des grands projets pour 2010 concernent maintenant des applications civiles dont les demandes en calcul intensif explosent. Les grands calculateurs sont d'ores et déjà une des principales clés de la compétitivité mondiale tant pour la recherche que pour l'industrie. Le CEA a d'ailleurs initié depuis plusieurs années une politique d'ouverture de ses moyens. Cela s'est en particulier traduit par la création du Complexe de calcul scientifique de Bruyères-le-Châtel (Essonne) rassemblant, autour des équipes d'experts du CEA les moyens défense (Tera-10), les moyens civils partagés entre le CEA et ses partenaires dans le cadre du CCRT⁽²⁾, et les nécessaires moyens d'expérimentation. De même, avant que Tera-10 ne soit totalement dédié aux besoins de la défense, nous avons ouvert la plus puissante machine d'Europe à de grands défis scientifiques et industriels, calculs inédits à l'échelle mondiale qui font l'objet de ce numéro spécial. Enfin, moins d'un an après l'installation de Tera-10, un calculateur de 43 téraflops, conçu par Bull, d'architecture voisine et de puissance comparable, est en train d'être installé au CCRT. Il devrait entrer en service opérationnel en septembre 2007.

L. R. La technopole Ter@tec, symbole de cette synergie entre défense, industrie et recherche, a été créée autour du Complexe de calcul scientifique. Quels sont les projets en cours ?

J. G. C'est effectivement la meilleure illustration de cette politique d'ouverture. Le

premier succès, précurseur de ces collaborations, a donné naissance en 2004 au serveur NovaScale commercialisé par Bull, dont une version équipe Tera-10. La technopole fédère des laboratoires de recherche et des industriels, tous acteurs et promoteurs de la simulation numérique pour promouvoir les applications du calcul intensif et développer les technologies futures des puissants calculateurs (voir l'article sur Ter@tec p. 40). Aujourd'hui, elle compte plus de 40 partenaires. Cette année, le CNRS, l'Inria, Airbus, Total... nous ont rejoint. Selon leur finalité, les projets proposés au sein de Ter@tec sont soutenus soit par le CEA et les partenaires concernés, soit par l'Agence nationale pour la recherche (ANR⁽³⁾), soit dans le cadre du pôle de compétitivité francilien System@tic⁽⁴⁾ dont Ter@tec est un élément essentiel, soit par la Communauté européenne. Trois des cinq grands projets labellisés par System@tic sont des projets à l'initiative de partenaires de Ter@tec.

L. R. A l'échelle nationale aussi, l'année 2006 aura été charnière pour la communauté du calcul scientifique...

J. G. C'est indéniable. Elle a vu le deuxième appel à projet de la ligne « calcul intensif » de la toute nouvelle ANR : 14 projets ont été labellisés en 2006 pour un total de 8 millions d'euros. Ils sont répartis sur toute la France. Trois projets sont issus de Ter@tec. C'est un coup de pouce très important, essentiel. On peut seulement regretter qu'aucun projet ne concerne le développement de technologies et architectures matérielles, domaine abandonné depuis plus de 20 ans, un manque auquel il faudrait rapidement remédier pour recréer les compétences indispensables dans ce



domaine. Autre bonne nouvelle : l'annonce par le gouvernement, en juillet dernier, de la création de Genci (Grand équipement national pour le calcul intensif), société civile qui doublera les investissements alloués jusqu'à maintenant au calcul intensif, les coordonnera et sera l'interlocuteur unique de la France pour l'Europe (lire article sur Genci, p. 41).

L. R. Qu'en est-il justement en Europe ?

J. G. Nous saluons, avec le lancement du 7^e programme cadre de recherche (2007-2013) le retour du calcul intensif dans les objectifs européens, une préoccupation qui avait disparu depuis 10 ans ! L'Europe s'est donné deux grands objectifs : revenir au niveau des grands pays industriels, en matière de mise à disposition de puissance informatique et maîtriser les technologies de ces grands calculateurs. Pour répondre au défi technologique, nous venons de mettre en place une alliance baptisée TALOS (Technologies for Advanced Large Scale Open Supercomputing) entre Bull, Quadrics, la filiale allemande d'Intel, le CEA et HLRS (un des plus grands centres de calcul scientifique et d'expertise allemand, à l'université de Stuttgart). En ce qui concerne la mise à disposition de puissance informatique, l'Europe propose de participer au financement de machines pétaflopiques à l'horizon 2009/2010. Le Complexe de calcul scientifique du CEA est bien entendu candidat pour accueillir et gérer une telle infrastructure.

Propos recueillis par Isabelle Bellin

(1) **Top 500** : palmarès des 500 plus puissants calculateurs au monde. www.top500.org

(2) **CCRT** : Centre de calcul pour la recherche et la technologie. Les partenaires du CEA dans le CCRT sont EDF, le groupe Safran au travers de ses filiales, Snecma, Turbomeca et Techspace Aéro.

(3) **ANR** : Agence nationale pour la recherche. <http://www.agence-nationale-recherche.fr/>

(4) **System@tic Paris Région** est un pôle de compétitivité créé en 2005 : sa vocation est de donner à l'Île-de-France une visibilité mondiale dans le domaine des systèmes complexes sur quatre marchés prioritaires : automobile et transports, sécurité, télécommunications et ingénierie de conception. <http://www.systematic-paris-region.org/>

Simulation aérodynamique une révolution permanente

Le Falcon 7X, modèle haut de gamme de Dassault Aviation, est le premier avion d'affaires à avoir été entièrement optimisé par calcul numérique. Des modélisations sur le supercalculateur Tera-10 ont permis à l'avionneur français de tester les capacités des futurs moyens de simulation numérique.

Didier Gout
est journaliste
scientifique

Avec un supercalculateur comme Tera-10, la conception aérodynamique est plus précise et plus rapide et, à raison du doublement tous les dix-huit mois des capacités des machines, on peut prévoir que d'ici quelques années une telle puissance sera devenue chez nous un outil classique » : chez Dassault Aviation, on n'est pas avare de compliments pour Tera-10, le supercalculateur livré au CEA en 2006, et les perspectives qu'une machine de cette puissance permet d'envisager. Elle a été utilisée au premier semestre de l'année 2006 dans le cadre d'une des opérations dites « Grand Challenge » du supercalculateur. L'objectif était d'évaluer les possibilités d'améliorations des processus de conception aérodynamique d'un avion complet, grâce à ce type de calculateur. Car le dernier-né de l'avionneur, le Falcon 7X, nouvel avion d'affaires haut de gamme déjà commandé à plus de 125 exemplaires et dont la certification est prévue au printemps, a déjà bénéficié de l'utilisation des calculateurs les plus pointus, en interne (des IBM SP et Bull NovaScale) : des simulations qui ont contribué aux performances aérodynamiques remarquables de ce triréacteur capable de franchir 11 000 kilomètres sans escale.

Le recours au calcul haute performance n'est, en effet, pas nouveau dans le domaine des avions, en particulier chez Dassault, qui a toujours eu un rôle de lea-

der grâce entre autres à son logiciel CATIA dont le concept est né au sein du bureau d'étude de l'avionneur. Et les performances des avions d'affaires de la société, en particulier en matière de consommation de carburant, sont le fruit direct des calculs aérodynamiques et de l'expertise de ceux qui les mettent en œuvre.

La simulation numérique a révolutionné le design aérodynamique des avions

Le calcul haute performance s'est développé depuis quelques années dans l'aéronautique à tous les niveaux de la conception et en particulier pour la partie aérodynamique. La complexité de l'écoulement de l'air autour d'un avion ainsi que le niveau des performances recherchées imposent le recours à des logiciels de simulation très avancés. Ces logiciels

mique :



résolvent les équations classiques de la mécanique des fluides (équations de Navier-Stokes), avec une modélisation des phénomènes de turbulence. Le haut degré de précision requis demande une grande puissance informatique pour obtenir les résultats de calcul dans des temps raisonnables (environ une journée), compatibles avec les délais du cycle de conception de l'avion. La combinaison d'une modélisation fine de la physique des écoulements et d'une puissance de calcul élevée a permis de révolutionner le design aérodynamique. Celui de la voilure du Falcon 7X est un processus itératif qui implique l'utilisation intensive de logiciels d'optimisation automatique et de simulation numérique. Ces outils sont déployés pour générer

« Ces outils sont déployés pour générer plusieurs dizaines de voilures différentes »

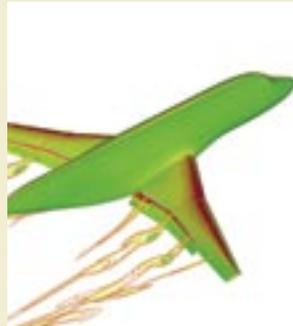
à moindre coût plusieurs dizaines de voilures différentes, dont l'analyse par calcul permet d'obtenir le meilleur compromis entre les différents objectifs de performance. Les meilleures formes sont ensuite testées en soufflerie afin de valider les résultats. Les écarts éventuels constatés entre les essais et les calculs sont par la suite analysés pour alimenter ce processus itératif. L'utilisation combinée de l'analyse théorique la plus poussée et de la confrontation immédiate aux résultats expérimentaux les plus modernes permet de réduire considérablement le nombre total d'essais de mise au point avant la caractérisation des formes finales retenues pour l'avion. Quel intérêt représente l'utilisation d'une machine

La combinaison d'une modélisation fine de la physique des écoulements et d'une puissance de calcul élevée a permis de révolutionner le design aérodynamique.

CALCUL NUMÉRIQUE DES CARACTÉRISTIQUES TECHNIQUES D'UN FALCON 7 X



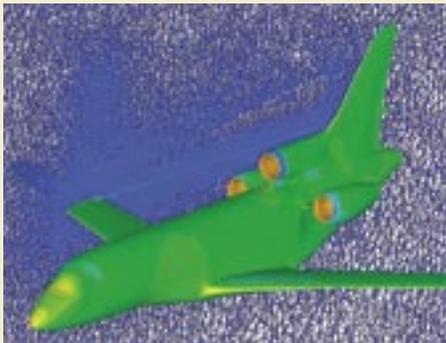
Le calcul numérique permet d'optimiser la forme des ailes (ici d'un Falcon). Au-dessus de l'aile, on observe en rouge la zone de basse pression, caractéristique de portance de l'avion. La visualisation du sillage permet d'analyser la traînée de l'avion. On note aussi la trace du tourbillon issu de l'extrémité de la voilure.



Calcul numérique d'un Falcon en configuration d'atterrissage :

on distingue les becs de bord d'attaque à l'avant et les volets à l'arrière.

La maîtrise de la conception de tels dispositifs permet aux avions d'atterrir sur des pistes de faible longueur. Les couleurs renseignent sur les champs de pression sur l'aile.



Lors des simulations aérodynamiques réalisées avec Tera-10, l'espace fluide qui entoure l'avion a été discrétisé en un maillage très fin, de 115 millions d'éléments.

ments, là où les simulations courantes actuelles utilisent environ 30 millions d'éléments contre 10 millions au début de la conception du Falcon 7X. À raison de sept valeurs (pression, vitesses, température, énergie cinétique et taux de dissipation de la turbulence...) à définir pour chaque nœud d'un élément de maillage (qui en comprend quatre), on aboutit aujourd'hui à 140 millions d'équations à résoudre pour l'ensemble des champs aérodynamiques autour de l'avion, contre 14 millions il y a quelques années. Ce sont donc des niveaux d'équation dix fois plus compliqués que la machine Tera-10 permet d'appréhender.

Concrètement, avec Tera-10, les ingénieurs de Dassault Aviation ont évalué la traînée du Falcon 7X à son allure de croisière (on parle de « point de croisière », le point de vol le plus classique, Mach 0,8 pour le Falcon 7X). Cela a permis de dimensionner les besoins informatiques correspondant à cette problématique. Maintenant, l'avionneur envisage d'appliquer cette puissance de calcul pour améliorer la conception des dispositifs hyper-sustentateurs (des becs et des volets) utilisés pour atterrir et décoller de pistes aussi courtes que possible, en tirant parti de ce meilleur raffinement du maillage (ou discrétisation).

Dassault Aviation peut aussi envisager d'autres applications, qui demandent non pas plus de discrétisation mais de temps de calcul plus long. C'est le cas des calculs des phénomènes « instationnaires » qui varient en fonction du temps ; typiquement, ce sont les écoulements observés aux extrémités du domaine de vol, notamment pour des taux de descente ou de vitesse maximums. Ultérieurement, une autre application de ce type d'architecture informatique sera l'optimisation automatique des formes, qui est un processus logiciel facilitant la recherche de formes. Cette technique, en cours de développement, devrait s'appliquer aux prochains avions. Plus généralement, Dassault Aviation, fort de l'expérience réussie de l'utilisation de la chaîne de calcul aérodynamique sur l'architecture massivement parallèle de Tera-10, envisage dans les prochaines années le recours de manière intensive à cette échelle de calcul, sur les productions civiles et militaires de la société que développe le bureau d'étude : les Falcon du futur mais aussi les avions de combat habités ou non (par exemple les drones de combat dont un démonstrateur technologique devrait voler à l'horizon 2011 dans le cadre du projet européen nEUROn). Ce sont donc tous les produits d'avenir de Dassault Aviation qui devraient bénéficier des progrès aérodynamiques, et donc économiques, que permet la révolution désormais permanente des supercalculateurs.

© Dassault Aviation

comme Tera-10 ? « Avec de telles puissances de calcul, on peut imaginer que la conception aérodynamique d'un avion par simulation jusqu'à ses formes finales, qui demande plus d'un an actuellement, pourrait être réduite à terme de quelques mois, tout en bénéficiant d'une précision plus fine », estime-t-on chez Dassault Aviation. C'est une conclusion qui ressort des 200 000 heures de calcul effectuées sur la machine du CEA, des calculs réalisés en 200 heures seulement grâce à mille processeurs parallèles de Tera-10. Cela représente l'équivalent d'une centaine de calculs, sachant que classiquement une seule boucle totale de conception et de validation des formes d'un avion requiert quelques centaines de ces mêmes calculs.

Tera-10 permet des calculs dix fois plus complexes

Avec Tera-10, ces calculs ont été réalisés à partir d'un maillage plus fin, qui est de plus de 100 millions d'élé-

D. G.

Xedix, un système de gestion de... masses de données

Au départ, il s'agissait d'organiser plus de mille milliards d'octets racontant 40 ans d'essais nucléaires. À l'arrivée, Xedix est un nouveau type de système de gestion de bases de données, taillé pour les plus gigantesques « entrepôts de données ».



Robot de stockage des données Storage Tek au centre CEA/DAM de Bruyères-le-Châtel.

Léo Gerat est journaliste scientifique

Treize février 1960, oasis de Reggane, dans le Sahara algérien : un enfer se déchaîne sur terre. La France vient de procéder à son premier essai nucléaire. 27 janvier 1996, atoll de Fangataufa, dans le Pacifique : pour la dernière fois, une ogive nucléaire française a explosé... réellement. Depuis cette date, la France teste ses armes atomiques par simulation numérique, et s'est dotée pour cela

de puissants moyens de calcul. Mais la simulation est bien obligée de s'arrimer au réel, elle ne vaut que parce qu'elle fait tourner des « modèles » eux-mêmes validés, « calés » à partir de mesures expérimentales. Autrement dit, l'information issue des 210 essais réels effectués en près de 40 ans est irremplaçable. Telle est l'origine du projet Xedix du CEA/DAM (Direction des applications militaires), né de cette volonté de préserver 40 ans d'expérience nucléaire

Le plus grand télescope du monde (projet Seti@home à Arecibo, Porto-Rico) écoute en temps réel les signaux radio de l'espace pour y détecter d'éventuels messages extra-terrestres. Une masse considérable de données à gérer, application potentielle pour Xedix.

et de les mettre à disposition des concepteurs d'armes qui allaient dorénavant travailler par simulation. Quarante ans de documents de toutes sortes, de mesures en tout genre, de clichés, de films et de vidéos... au total 400 gigaoctets⁽¹⁾ (Go), croyait-on. Didier Courtaud et Pierre Brochard, qui ont mené à terme cette ambitieuse opération, savent aujourd'hui qu'il s'agissait de 1,3 téraoctet⁽²⁾ (To).

S'agissant d'une masse d'informations de nature majoritairement documentaire, on pense très tôt à une représentation reposant sur SGML. Le «Standard Generalized Markup Language» est une norme ISO depuis 1986, il est à l'époque déjà largement employé pour normaliser la documentation technique. Formellement, SGML est un puissant métalangage⁽³⁾ permettant de créer des langages de balisage. C'est-à-dire des langages permettant d'enrichir un texte à l'aide de «balises», pour le structurer (titre, chapitres, notes...) et le mettre en forme (italique, centrer...). Le plus connu est html, employé pour décrire les pages du World wide web.

Reste à rendre cette masse d'informations exploitable. Il faut pour cela en faire une «base de données», dans laquelle des utilisateurs agréés retrouveront ce dont ils ont besoin en lui adressant des «requêtes».

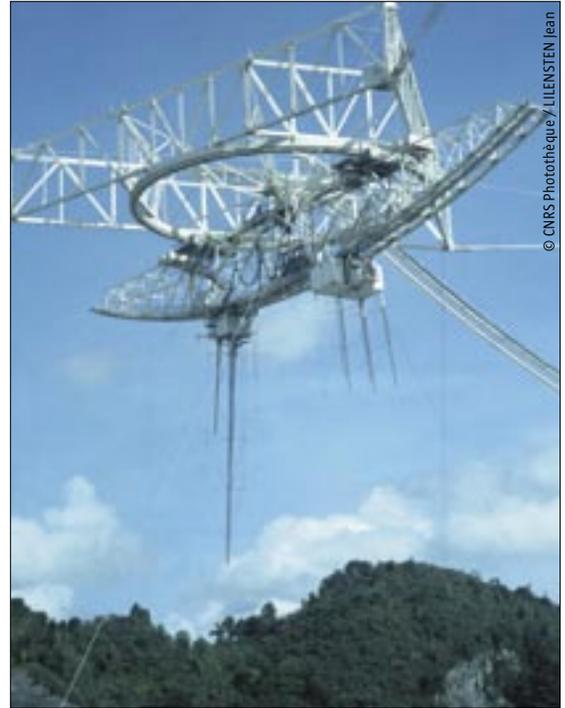
Des bases de données « XML natives »

À cette époque (1996), le paradigme de base de données en vogue est le modèle relationnel. Mais personne n'a jamais manipulé une telle masse d'information documentaire et multimédia de cette manière-là. Un nouveau concept semble prometteur : la base de données « orientée objet », néanmoins pas spécialement adapté au problème posé. On se dirige d'abord de ce côté-là, en se rapprochant du projet O2 né d'une collaboration entre l'Inria et le Laboratoire de recherche en informatique de l'université d'Orsay.

Quelques années plus tard, on passe de SGML à XML (Extended Markup Language), une simplification du premier qui s'impose comme une traînée de poudre et dans laquelle on voit un brillant avenir pour toutes sortes d'applications, notamment sur le web.

En 2000, le projet O2 s'éteint. Il est devenu clair alors que l'on ne peut compter sur aucune catégorie classique de base de données pour bien résoudre le problème, mais qu'il faut au contraire se lancer dans une approche « Native XML Database » (NXD), autrement dit concevoir un Système de gestion de bases de données (SGBD) totalement pensé pour et par XML. Xedix sera un SGBD « XML natif ».

De 2001 à 2003, les chercheurs fournissent un gros travail de développement, aboutissant à un outil complet, autonome, susceptible de prendre en charge



© CNRS Photothèque / LILLENSTEN Jean

toutes sortes d'applications, de type documentaire et multimédia notamment, et tout particulièrement si elles manipulent des masses de données considérables. Xedix est, entre autres, capable de gérer et d'indexer des bases de données physiquement réparties sur de multiples machines.

Le moment est venu de mettre le bébé à l'épreuve. Ce sera le premier banc d'essai de Xedix, appelé Xtera 1, en 2003. Une base de données d'un téraoctet, fictive mais plausible, est créée et testée sur une grappe de 16 PC dotés de disques de 80 Go, coordonnée par un PC maître. On constate alors que le temps de réponse à une requête est presque indépendant de la taille de la base et essentiellement proportionnel au nombre d'éléments XML retournés en réponse à la requête. Dans un cas type pénalisant (recherche d'un mot présent dans 60 % des documents), le temps de réponse moyen est de l'ordre de 10 secondes. Ce premier test démontre la capacité de Xedix à prendre en charge des bases de données d'un téraoctet sur une architecture matérielle à la portée d'une PME.

Décembre 2005 : nouvelle mise à l'épreuve, en plaçant la barre à 10 téraoctets, un test baptisé Xtera 10. Il s'agit cette fois de montrer que Xedix sait tirer parti d'architectures matérielles parallèles à haute performance. Le supercalculateur TeraNova⁽⁴⁾ est mis à contribution. On confie à 300 de ses processeurs une base d'essai occupant 10 To. À nouveau, on constate que le temps de réponse à une requête est quasi proportionnel au nombre de réponses. Une réponse type est obtenue cette fois en 2 à 8 secondes. Un troisième banc d'essai (Xtera 100) est prévu. Il doit confirmer,

(1) Gigaoctet (Go) : milliard d'octets.

(2) Téraoctet (To) : millier de milliards d'octets.

(3) Métalangage : formalisme permettant de décrire des langages.

(4) Supercalculateur TeraNova : cluster de serveurs Bull NovaScale, installé sur la technopole Ter@tec (voir l'article p. 40).

Puissance : 1,3 téraflops.
(5) Inist : Institut de l'information scientifique et technique, rattaché au CNRS.

au printemps 2007, que Xedix maîtrise des bases de données de 100 To. Pour fixer les idées, c'est trois fois la masse de données que doit manipuler la Très Grande Bibliothèque (TGB, Paris).

De l'après Mururoa à la start-up

Aujourd'hui, Xedix n'est plus seulement un projet visant à préserver et rendre accessible « 40 années d'essais nucléaires ». La question est réglée. Xedix est maintenant surtout un produit et une entreprise que le CEA s'apprête à laisser voler de ses propres ailes. Cette dernière sera dirigée par Jean-Claude Sabattier, du CEA/DAM, et on y retrouvera à la direction technique les mêmes Pierre Brochard et Didier Courtaud.

Le marché des SGBD est en forte croissance (8 % par an). Le cabinet Gartner prévoit un marché de 15 milliards de dollars en 2009. Dont 10 % sous XML. Xedix prétend prendre sa place sur ce segment de marché. Elle revendique pour son produit la position de première implémentation industrielle du concept de base de données XML native. Et affiche des résultats de tests en vraie grandeur sur une base de 10 To. Du jamais vu, affirme-t-on.

Un tel outil serait particulièrement adapté à la prise en charge de vastes collections de documents de

type documentaire, mais pas seulement. Des collaborations dans le cadre du pôle de compétitivité System@tic ont déjà permis de tester ce potentiel sur une application liée au projet international « Physiome » (voir encadré) et par une collaboration avec l'Institut national des télécommunications sur l'accès à de vastes bases de fichiers vidéo.

Au-delà des applications clairement documentaires ou multimédia, relevant de la sphère de la bibliographie ou des médias (TGB, Inist⁽⁵⁾...), Xedix espère intéresser des secteurs tels que celui des grands instruments scientifiques (accélérateurs, télescopes...) engendrant des masses considérables d'informations, des télécoms (qui doivent garder la trace de chaque communication) ou encore des industries qui produisent des montagnes de documents, comme l'avioni-

« Xedix est maintenant un produit et une entreprise que le CEA s'apprête à laisser voler de ses propres ailes. »

que (que l'on songe à la série de tests subis par l'A380) ou la pharmacie (même remarque à propos des documents décrivant l'histoire d'un médicament jusqu'à son autorisation). Un indice intéressant à propos de ce dernier exemple : la FDA (Food and Drug Administration, organisme américain d'autorisation des mises sur le marché des médicaments) exige désormais que cette documentation soit fournie en XML...

L. G.

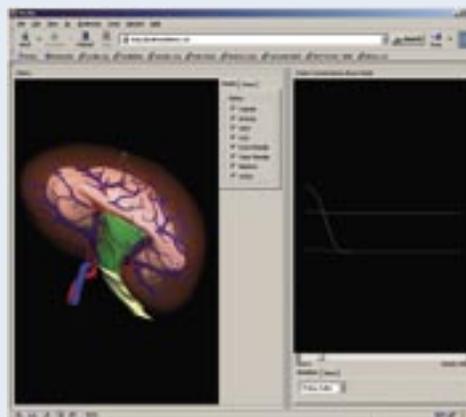
Une application remarquable : le Physiome

Le projet mondial connu sous le nom de « Physiome »⁽¹⁾ affiche une folle ambition. De même que le projet « Génome » visait au décodage intégral du patrimoine génétique de l'homme, il veut « mettre en boîte » un modèle physiologique complet de l'homme. Le temps passant, les chercheurs impliqués dans ce projet se sont fixés un objectif intermédiaire plus raisonnable : la constitution d'une banque de données multimédia, mondiale et répartie, de la physiologie du corps humain, à toutes les échelles. Cela signifie que des équipes réparties dans le monde entier pourront mettre en commun leurs dernières découvertes sur le fonctionnement de chaque organe, tissu, cellule, molécule...

En France, au Laboratoire Ibisc⁽²⁾ du CNRS et de l'université d'Evry, Randy Thomas est coordinateur au niveau mondial pour le Physiome rénal. Il gère la base QKDB⁽³⁾ (Quantitative Kidney Database), qui a pour vocation de devenir la partie « rein » du projet Physiome.

Dans le cadre du projet FAME 2⁽⁴⁾ du pôle de compétitivité francilien System@tic, et en collaboration avec le CEA/DAM, Randy Thomas et Fariza Tahî développent un entrepôt de données sous Xedix pour le Physiome.

Il intégrera des données de tous types (textes, images, vidéos...) et à toute échelle, allant du génome à l'organisme entier. L. G.



© C. STEVENS, A. LONIE, SR THOMAS

Exemple d'interface reposant sur un modèle 3D qui devrait permettre d'accéder de manière interactive à l'information disponible dans le Physiome. Il s'agit d'un rein de rat virtuel en 3D zoomable, orientable, démontable.

(1) Projet Physiome : <http://physiome.org>

(2) Ibisc : Informatique, biologie intégrative et systèmes complexes.

(3) Base de données QKDB : <http://physiome.ibisc.fr/qkdb>

(4) Projet FAME 2 : <http://www.fame2.org>

Maîtriser les turbulents phénomènes de **combustion**

Des calculateurs parallèles de la puissance de Tera-10 ont permis de simuler l'allumage complet d'une chambre de combustion d'hélicoptère.

Dominique Ritman
est journaliste scientifique

L'aéronautique, à l'instar de l'industrie automobile, est l'un des secteurs industriels qui a connu les plus grands bouleversements avec l'avènement de la simulation numérique puis du calcul parallèle. Le Centre européen de recherche et de formation avancée en calcul scientifique (CERFACS) en sait quelque chose : il compte deux leaders mondiaux de l'aéronautique et de l'aérospatial parmi ses actionnaires, la firme EADS et le groupe Safran (auquel appartiennent notamment Snecma et Turbomeca). L'une des thématiques prioritaires, chez tous les constructeurs de moteurs aéronautiques, concerne aujourd'hui la conception des chambres de combustion. La simulation d'un allumage de chambre complète d'un turbomoteur d'hélicoptère Turbomeca, réalisée par le CERFACS en 2006, est en la matière une première mondiale.

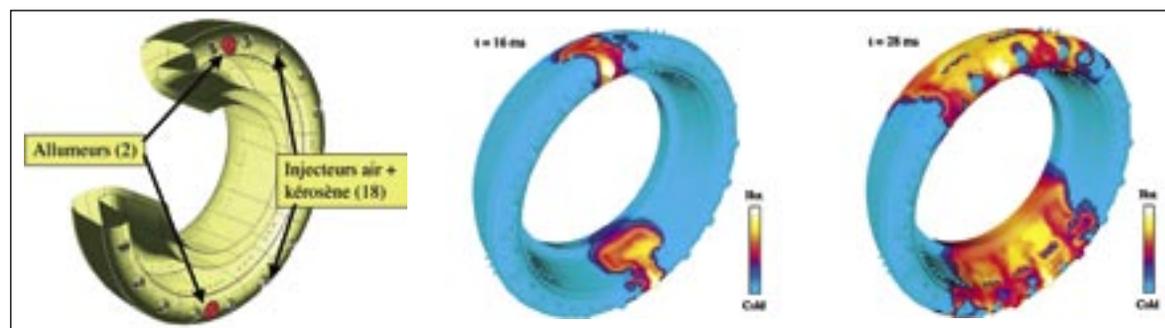
La technologie des chambres de combustion, de plus en plus sophistiquée, doit permettre de répondre aux nouvelles contraintes économiques (prix élevé du kérosène) et d'anticiper les réglementations futures (diminution des polluants tels que l'oxyde d'azote et des émissions de gaz à effet de serre comme le gaz carbonique). Dans ce contexte, la simulation numé-

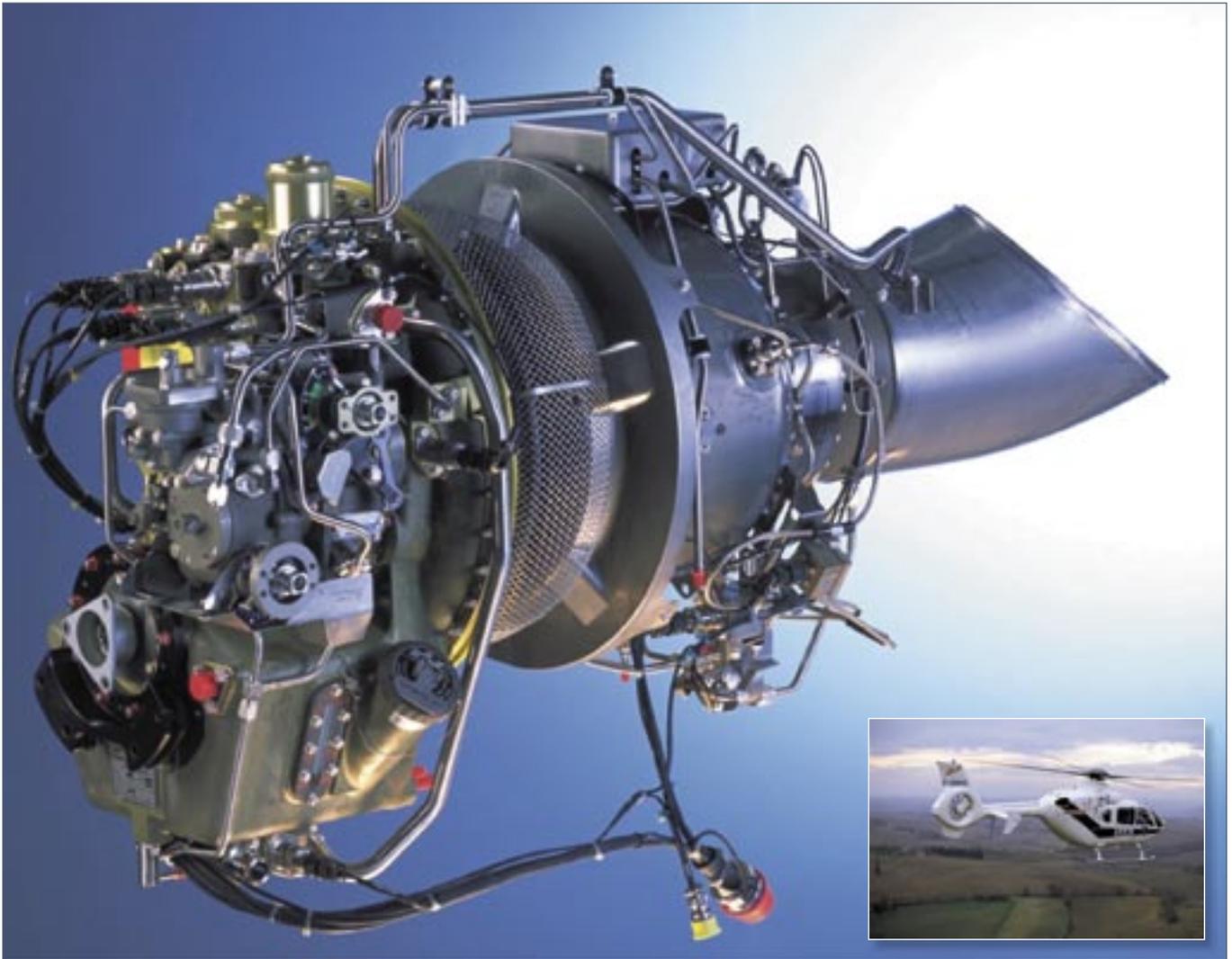
rique est devenue le passage obligé, pour des raisons de coût et de temps de conception.

Simuler sans trop simplifier

Le fonctionnement de ces chambres est régi par des équations bien connues : celles de la mécanique des fluides (équations de Navier-Stokes) couplées à celles de la combustion des différentes espèces chimiques présentes (hydrocarbures, oxygène, oxydes d'azote, oxyde de carbone...). Mais pour modéliser dans son intégralité une telle chambre, les scientifiques se trouvent conduits à résoudre plusieurs millions de fois (correspondant aux millions d'itérations du calcul en fonction du temps) un système d'équations comportant une quinzaine de variables sur 40 millions de cellules. Un système particulièrement complexe qui, il y a encore peu de temps, n'était résolu, qu'au prix d'un certain nombre de simplifications.

Le problème, explique Thierry Poinot, directeur de recherche CNRS et consultant au CERFACS, est que ces modélisations simplifiées ne donnent accès qu'à des valeurs moyennes des paramètres physiques, que ce soit de la température des gaz de combustion, de la pression... Or l'un des objectifs, notamment des industriels, est de mieux comprendre les mécanismes contrôlant la combustion turbulente et d'être ainsi





© TURBOMECA

en mesure d'anticiper tout dysfonctionnement. Plus généralement, les questions concernent les conditions d'allumage du moteur, de son réallumage en cas d'extinction imprévue, de son extinction ou encore l'étude de ses instabilités.

Les scientifiques ont utilisé la méthode de simulation dite LES⁽¹⁾ (Large Eddy Simulation, ou simulation aux grandes échelles), développée depuis une dizaine d'années à peine, tandis qu'ils ont mis au

point le code AVBP permettant de faire tourner ce modèle sur une machine massivement parallèle. AVBP a été testé avec succès sur plusieurs machines : sur le BlueGene d'IBM, aux États-Unis, sur un Cray XT3 et, au premier semestre 2006, sur le Tera-10 du CEA. Alors qu'il aurait nécessité 150 000 à 200 000 heures sur un ordinateur classique, le calcul a pris quelques jours sur ces machines.

« Au lieu de calculer un seul brûleur sur un secteur de chambre, comme cela se fait souvent pour des raisons d'économie, il a été possible de simuler le comportement des 18 brûleurs », souligne Thierry Poinsot. Le code AVBP est le fruit de la collaboration du CERFACS avec un grand nombre de laboratoires publics (principalement du CNRS) et privés (IFP, Turbomeca, Snecma, Peugeot, Renault...) associés au sein d'un consortium. Reste à passer au stade industriel : « Cela suppose un investissement important de la France dans l'acquisition de machines massivement parallèles », ajoute le chercheur.

(1) La simulation aux grandes échelles (Large Eddy Simulation, LES), par opposition à la simulation dite directe (Direct Numerical Simulation, DNS), repose sur une séparation des échelles : elle consiste à calculer les grandes échelles en modélisant l'action des petites échelles.

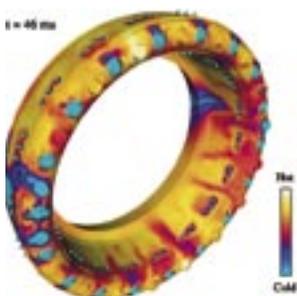


SCHÉMA DES ÉTAPES DE LA COMBUSTION

© Infographie CERFACS

D. R.

Cancérologie :

première simulation complète d'un examen

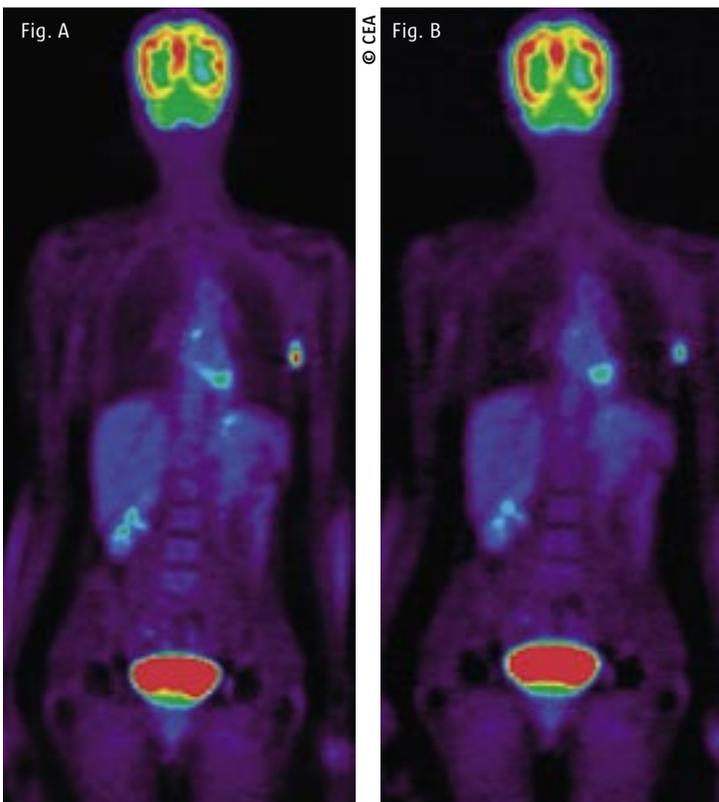
Les techniques d'imagerie nucléaire, comme la tomographie par émission de positons, devraient bénéficier des récentes simulations numériques d'examens médicaux.

Anne Lefèvre-Balleydier est journaliste scientifique

Fig A : Résultat d'un examen réel TEP : les zones en couleur sont celles où le métabolisme est accru, les tumeurs sont ici sous l'aisselle gauche du patient et dans le foie.

Fig. B : Résultat obtenu par simulation sur Tera-10

(1). Mis en place en 2002 par l'équipe de l'École polytechnique de Lausanne, le projet Open GATE, qui a pour but de valider un simulateur dédié à la TEP, regroupe plus de 20 laboratoires. En France, la collaboration réunit des équipes de l'Inserm, du CNRS et du CEA.



Ces deux images n'ont pas la même origine : l'une vient d'un examen médical de plus en plus courant en cancérologie la tomographie par émission de positons (TEP), l'autre est le résultat de sa simulation numérique. Mais elles se ressemblent terriblement : on repère dans les deux cas des tumeurs au niveau du foie et sous l'aisselle gauche. Une ressemblance qualitative, mais aussi quantitative : le volume

du ganglion présent sous l'aisselle est de 23,7 ml (millilitres) dans l'image issue de l'examen de TEP, et de 22 ml dans celle produite par simulation, soit un écart de seulement 6 %. Pour Sébastien Jan, chercheur au CEA et coordinateur technique depuis avril 2003 d'un projet international de simulations dédiées à la TEP (1), c'est sans l'ombre d'un doute une « très belle réussite ». Qui, dans un horizon de 5 à 10 ans, devrait faire avancer à grands pas le domaine de l'imagerie nucléaire et, spécifiquement, la TEP. Particulièrement utilisée en cancérologie, la TEP renseigne sur l'aspect fonctionnel de la maladie, et présente donc un intérêt tant dans le diagnostic que dans le suivi des traitements.

Donner une image réaliste des tumeurs...

Cette technique d'imagerie, dont le nombre d'appareils a récemment augmenté grâce au plan Cancer (environ un appareil par million d'habitants), permet de repérer les cellules tumorales en visualisant la consommation de glucose. « En pratique, on injecte au patient un radio-pharmaceutique qui, dans le cadre d'un examen en oncologie, est une molécule analogue au sucre (le fluoro-désoxyglucose, FDG) marquée avec un traceur radioactif à demi-vie courte (deux heures), le Fluor-18, explique Sébastien Jan. Il se répartit dans tout le corps

et se concentre là où le métabolisme est accru, comme c'est notamment le cas dans les tumeurs. Ce traceur émet des positons qui, dans les tissus environnants, vont s'annihiler avec des électrons : pour chaque annihilation, il naîtra deux photons gamma, émis suivant deux directions opposées. » C'est leur détection simultanée, sur les détecteurs d'une couronne tournante encadrant le malade (cf. image examen), qui permet ensuite de construire une image de distribution du traceur dans le corps. Et pour simuler cet examen de TEP sur le corps entier, il faut donc reproduire tous ces cheminements, depuis l'injection de sucres marqués jusqu'à la reconstruction d'image en trois dimensions, en passant par les interactions entre particules et matière...

...en simulant le parcours des particules

Cette simulation s'appuie sur les méthodes de Monte-Carlo : on part des propriétés physiques des différentes particules (telles que l'énergie, l'impulsion, la charge...), mais aussi des tissus et des matériaux qu'elles traversent (comme leur arrangement atomique, leur densité...), et l'on évalue le parcours de chaque particule en fonction des probabilités qu'elle a d'interagir avec eux. L'objectif ? Se rapprocher tant que faire se peut des processus réels, pour optimiser l'extraction du signal pertinent, autrement dit l'activité métabolique d'une tumeur. Car avant qu'on n'y applique toutes sortes de corrections, l'image issue d'un examen TEP est en général très bruitée. D'une part, parce que le sucre n'est pas fixé par les seules cellules cancéreuses. D'autre part, parce que la détection des deux photons émis par l'annihilation positon-électron ne va pas toujours se faire dans de bonnes conditions : les photons peuvent diffuser dans les tissus, ou y être atténués, ou provenir de deux annihilations différentes provoquant une coïncidence d'impact fortuite. Or, dans la simulation, tous ces événements sont connus, et l'on peut donc en tenir compte pour corriger les données. Mieux : on peut aussi mettre en œuvre des corrections du mouvement, pour tenir compte par exemple de la respiration du patient, qui a tendance à lisser les signaux issus des poumons. Le problème est que tout cela demande énormément de temps de calcul.

« Celui-ci est directement corrélé à la dose de traceur injecté et au temps d'acquisition, c'est-à-dire au nombre de particules qu'il faut simuler », souligne Sébastien Jan. Pour un examen de TEP classique, où la dose radioactive injectée est de l'ordre de



© C. Boultze/CEA

L'acquisition des données d'un examen TEP dure environ une heure, pendant laquelle le patient doit rester immobilisé dans cette couronne.

300 MBq (millions de becquerels, unité de mesure de la radioactivité) et le temps d'acquisition des données de 40 minutes, on génère environ 600 milliards de positons, et donc deux fois plus de photons. La modélisation s'appliquant à chacune des particules, elle requiert au total un minimum de 20 000 heures de calcul, soit environ 800 jours sur un ordinateur standard. En développant une plateforme de simulation (baptisée GATE, Geant4 Application for Tomographic Emission), portée sur le supercalculateur Tera-10, les chercheurs ont pu considérablement réduire ce laps de temps : dans l'examen corps entier qu'ils ont choisi de simuler, la dose atteignait 264 MBq, et le temps d'acquisition 49 minutes, mais avec 7 000 processeurs, il n'a fallu que 3 heures de calcul.

« C'était une première. Mais d'ici 5 à 10 ans, je pense qu'avec des archi-

tectures de calcul plus distribuées, on pourrait voir des simulations de ce type couplées à des examens réels à l'hôpital, ce qui permettrait de personnaliser les examens de TEP en optimisant, entre autres, les doses de traceur, les temps d'acquisition et plus généralement les protocoles d'acquisition. » L'équipe de Sébastien Jan y travaille, en particulier pour optimiser le rapport signal sur bruit dans l'image et permettre une meilleure détection tumorale. L'objectif est aussi de réduire le temps de calcul, pour réaliser ces simulations avec beaucoup moins de processeurs. Car ce qui intéresse les chercheurs, ce n'est pas tant la « beauté » de leur simulation, mais le fait qu'elle puisse devenir un outil accessible à tous...

A.-L. B.

« D'ici 5 à 10 ans, on pourrait voir des simulations de ce type couplées à des examens réels à l'hôpital. »

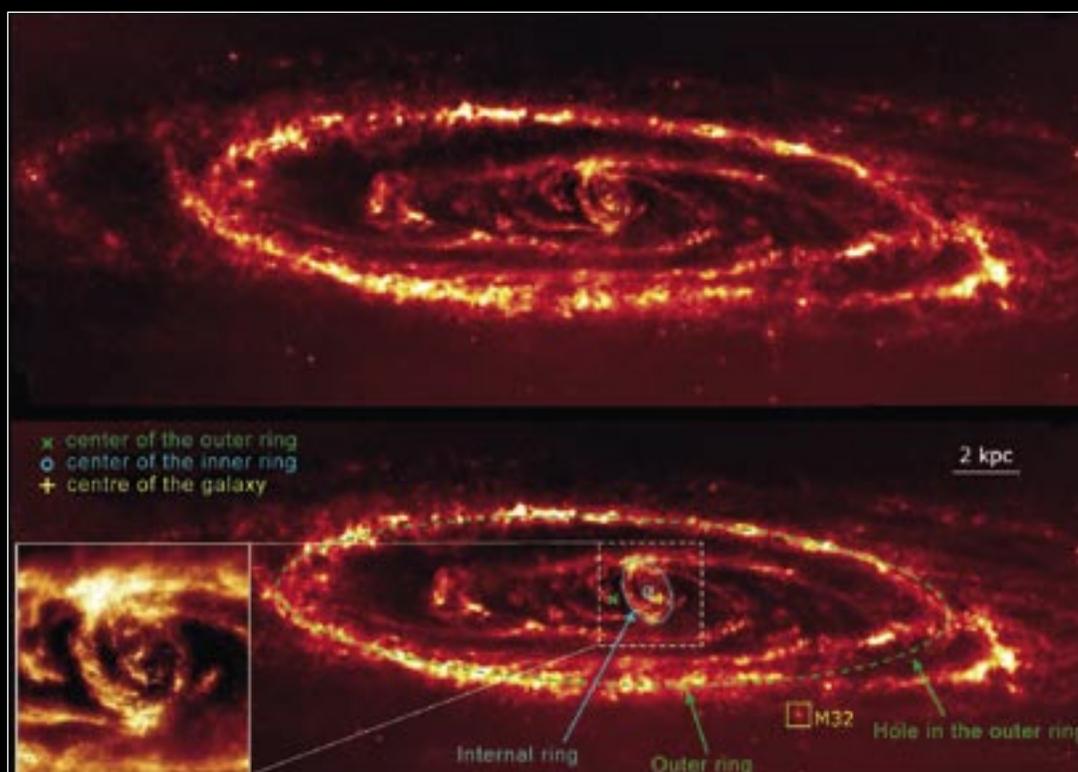


Fig. 1A. Andromède, photographiée en lumière visible (mont Palomar, Californie, 1958), avec ses deux compagnons, M110 et M32. Ce dernier (à gauche) est soupçonné d'avoir percuté Andromède il y a 210 millions d'années.

Nombres et lumières dans le ciel

La simulation numérique est devenue un complément indispensable du télescope pour les astrophysiciens. Elle permet d'explorer l'espace mais aussi le temps. Nous présentons ici une série de résultats récents obtenus dans le cadre du projet COAST (Computational astrophysics), qui réunit des chercheurs du Service d'astrophysique du CEA-Saclay et de ses partenaires (université Paris 7, CNRS, Observatoire de Paris). Ces clichés spectaculaires donnent un aperçu de la contribution de la simulation à l'avancement des connaissances en astrophysique à toutes les échelles.

Pierre Vandeginste est journaliste scientifique



© F. Bournaud, F. Combes/CNRS/OP, P. Barmby/CEA/NASA/JPL

Fig. 1B. Andromède vue par le satellite Spitzer, en infrarouge. On y voit un double anneau décentré au cœur d'Andromède ainsi qu'une lacune dans son anneau externe.

(1) <http://www.dapnia.cea.fr/Projets/COAST/index.htm>

Andromède se remet tout juste d'une collision

Andromède est la galaxie spirale la plus proche de la nôtre, la Voie lactée. Située à deux millions d'années-lumière, elle est également la seule visible à l'œil nu dans l'hémisphère nord et la plus massive de notre groupe local de galaxies. D'apparence régulière en lumière visible, Andromède dévoile une géométrie plus troublante

sur l'image (1B) issue du télescope spatial Spitzer. Lancé par la Nasa en 2003, ce dernier observe dans l'infrarouge, ce qui permet notamment de voir la poussière interstellaire. Ce cliché révèle un double anneau décentré au cœur d'Andromède et une lacune dans son anneau extérieur. Quel événement a pu ainsi défigurer Andromède ? Très certaine-

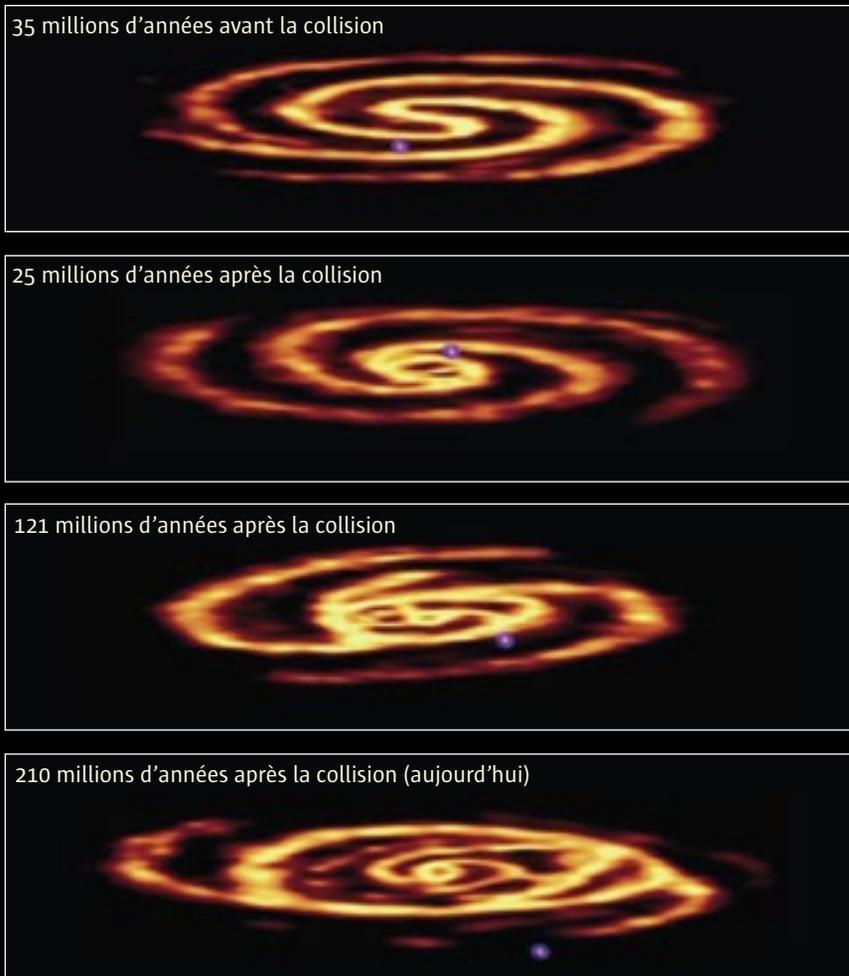


Fig 1C. Simulation de la collision de M32 avec Andromède. Les 4 images correspondent aux dates suivantes : 35 millions d'années (MA) avant la collision, 25 MA après la collision, 121 MA après et 210 MA après, (aujourd'hui). La dernière image est très proche de l'observation faite par le satellite Spitzer.

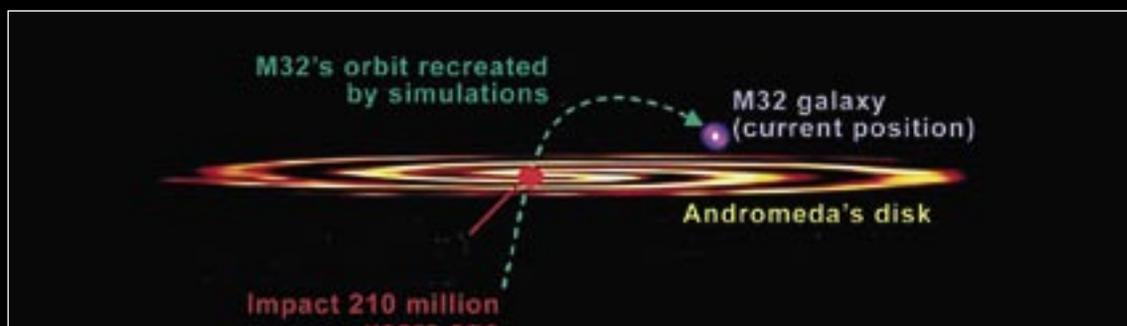
(2) www-dapnia.cea.fr/Sap/Actualites/Breves/bournaud061019/hpage.shtml

(3) *An almost head-on collision as the origin of two off-centre rings in the Andromeda galaxy*, D.L. Block, F. Bournaud, F. Combes, R. Groess, P. Barmby, M. Ashby, G. Fazio, M. Pahre, S. Willner, Nature, 19/10/06

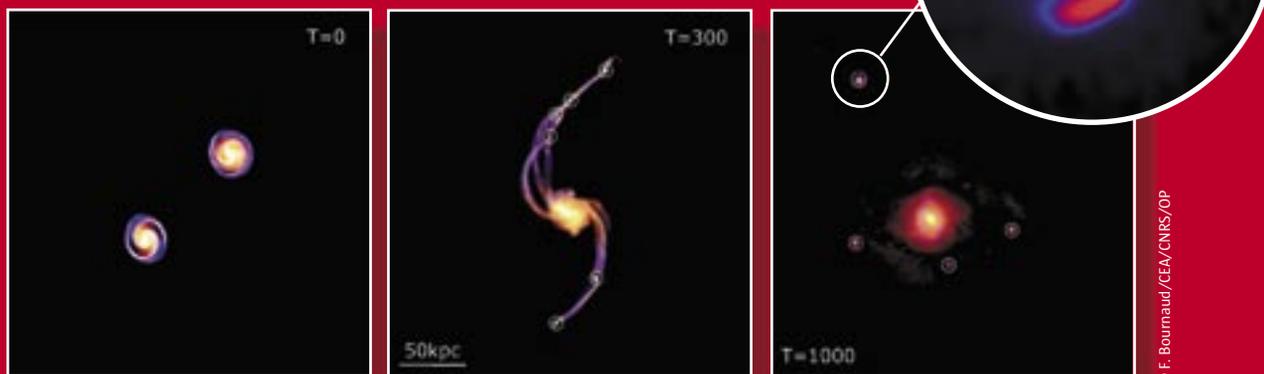
ment une collision il y a 210 millions d'années² avec une galaxie plus petite, son compagnon M32, expliquait en octobre dernier une équipe internationale dans le magazine *Nature*³, s'appuyant sur une série de simulations numériques réalisées par Frédéric Bournaud (CEA-Saclay, Observatoire de Paris) sur un supercalculateur vectoriel, le NEC-SX6 du CCRT (Centre de Calcul Recherche et Technologie). Le modèle employé, représentant

qui a percuté Andromède. Le cliché 1B (voir p.23), réalisé dans l'infrarouge par le satellite Spitzer, laisse voir ses deux anneaux décentrés et une lacune dans l'anneau externe. Enfin la séquence 1C (ci-dessus) montre le déroulement de la collision, obtenu par simulation numérique, tandis que la figure 1D (ci-dessous) montre la trajectoire de la galaxie naine M32 qui lui a fait percuter et même traverser la galaxie spirale Andromède.

Fig 1D. Reconstitution de la trajectoire de M32, la galaxie naine qui a percuté et traversé Andromède il y a 210 millions d'années.



Deux galaxies spirales se rencontrent... et font des bébés



© F. Bournaud/CEA/CNRS/OP

D'où viennent les « galaxies naines », qui réunissent moins d'un milliard d'étoiles, quand des galaxies géantes comme la nôtre en comptent cent fois plus ? On en trouve justement beaucoup en orbite autour de ces mêmes galaxies géantes. Quatorze, aux dernières nouvelles, autour de la Voie lactée. On croyait jusqu'à présent que la plupart avaient été créées très tôt dans l'histoire de l'univers et que l'on observait aujourd'hui celles qui n'ont pas encore fusionné pour former de plus grandes galaxies. Les simulations numériques⁴ réalisées par une équipe du Service d'Astrophysique du CEA/DAPNIA (Laboratoire de

recherche sur les lois fondamentales de l'univers) et de l'Observatoire de Paris amènent à réviser un peu cette vision. Elles montrent que les galaxies naines sont un sous-produit banal des collisions entre galaxies géantes, qu'il y aurait donc une production continue de naines.

Frédéric Bournaud et Pierre-Alain Duc ont simulé une centaine de collisions de galaxies spirales sur le supercalculateur vectoriel NEC-SX6 du CCRT. Leur modèle numérique représentait chaque galaxie par un nuage de trois à six millions de particules et simulait la dynamique des étoiles, du gaz interstellaire et de la matière noire.

En moyenne, cinq heures de calcul étaient nécessaires pour effectuer chaque simulation.

Lors des collisions de galaxies spirales, de puissants effets de marée ont pour résultat de projeter de longues traînées de part et d'autre de la nouvelle galaxie en formation. Dans ces « antennes » se forment des « galaxies naines de marée ». Les simulations ont montré que près du quart de ces naines nées d'une collision survivaient au moins deux milliards d'années. On peut donc penser qu'une petite fraction des galaxies naines observées ont pour origine une collision, au moins dans certains environnements.

Fig. 2A. Simulation d'une collision-fusion de deux galaxies spirales. Les trois premiers clichés correspondent à trois dates : $T = 0$, $T = 300$ MA (millions d'années) et $T = 1000$ MA. Le quatrième cliché montre un détail du troisième : une galaxie naine de marée résultant de la collision. Le modèle utilisé ici comporte 80 millions de particules et a demandé 40 heures de calcul vectoriel sur le NEC-SX6 du CCRT.

(4) http://www-dapnia.cea.fr/Sap/Actualites/Breves/paduc060604/page_fr.shtml



© F. Bournaud/CEA/CNRS/OP

Fig. 2B. Une autre collision simulée de deux galaxies spirales, au ralenti. Cette fois il s'écoule 30 millions d'années entre deux images, la séquence dure 330 MA en tout.

La formation des grandes structures de **l'univers enfant**

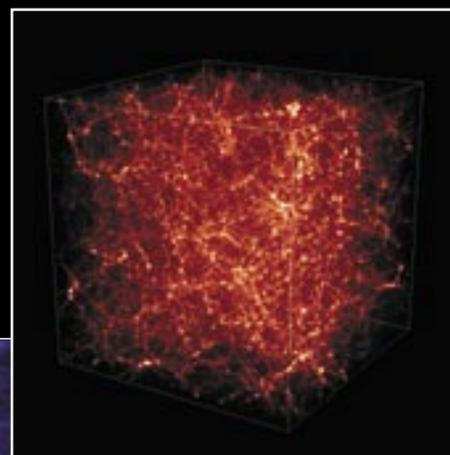


Fig. 3A. Simulation cosmologique de la formation de grandes structures de matière noire dans un fragment d'univers âgé d'un milliard d'années. Les zones claires correspondent aux halos de matière noire, au sein desquels les galaxies et leurs étoiles se forment.

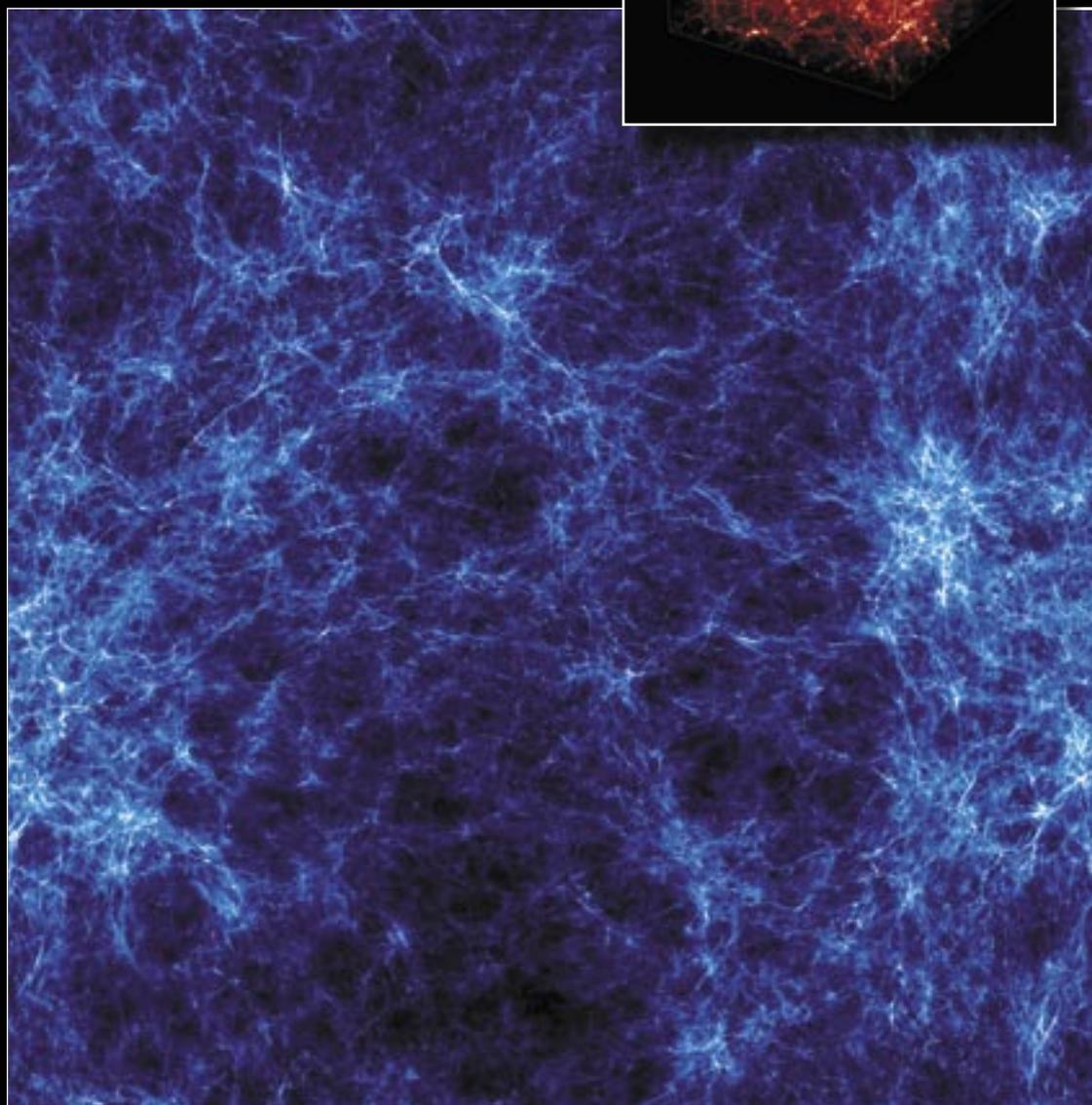
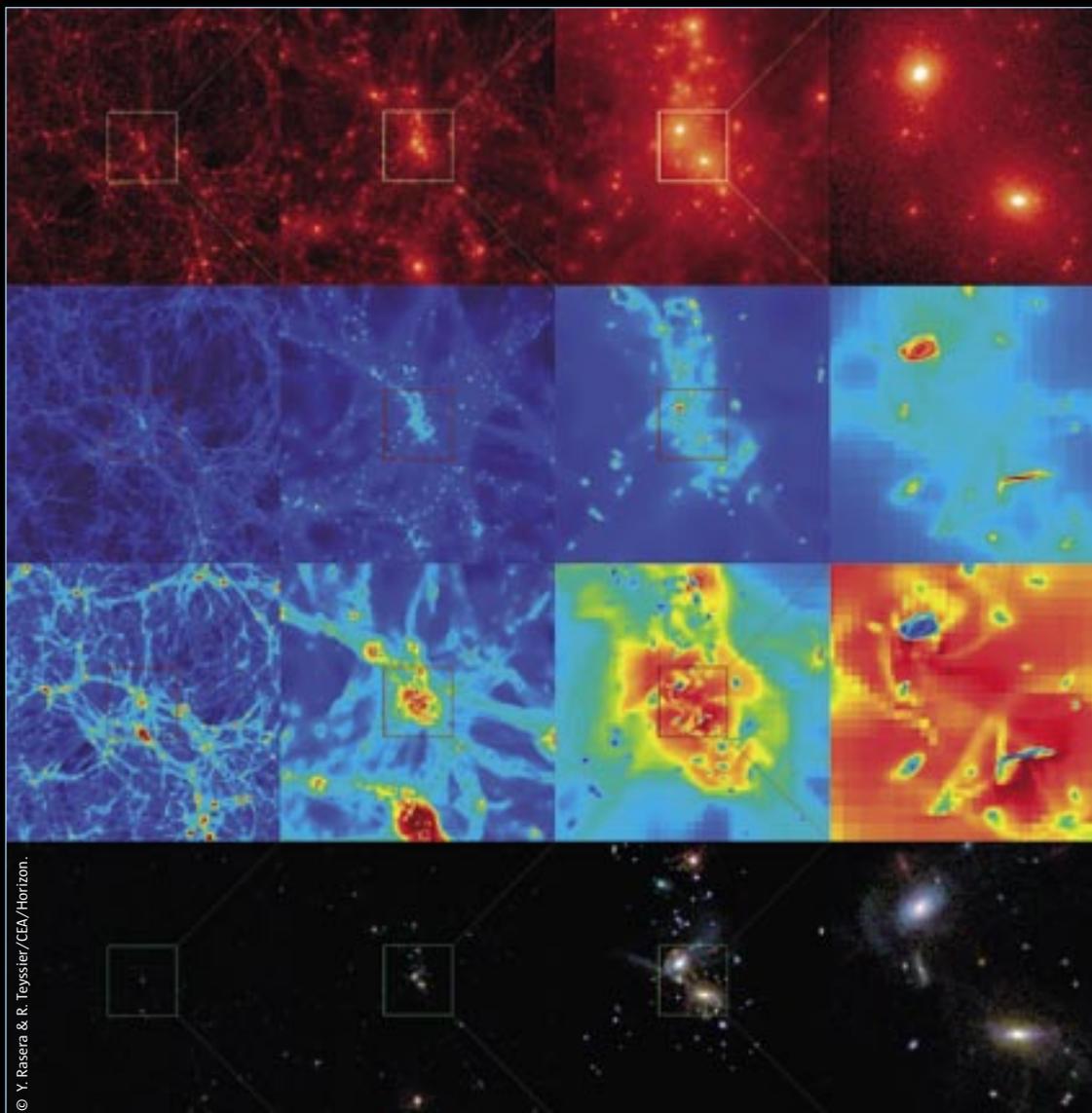


Fig. 3B (en haut à droite). Douze milliards d'années plus tard... une tranche de l'univers actuel (simulé), cette fois en volume. Avec des régions encore plus denses, correspondant aux amas de galaxies actuels.

© R. Teyssier/CEA/Horizon



© Y. Rasera & R. Teyssier/CEA/Horizon.

Fig. 3C. Même type de simulation qu'en 3A, mais à 2 milliards d'années après le big-bang, sur une boîte de 10 millions d'années-lumière de côté. L'image montre 4 niveaux de zoom de gauche à droite. Et de haut en bas la matière noire, le gaz, la température du gaz et les étoiles.

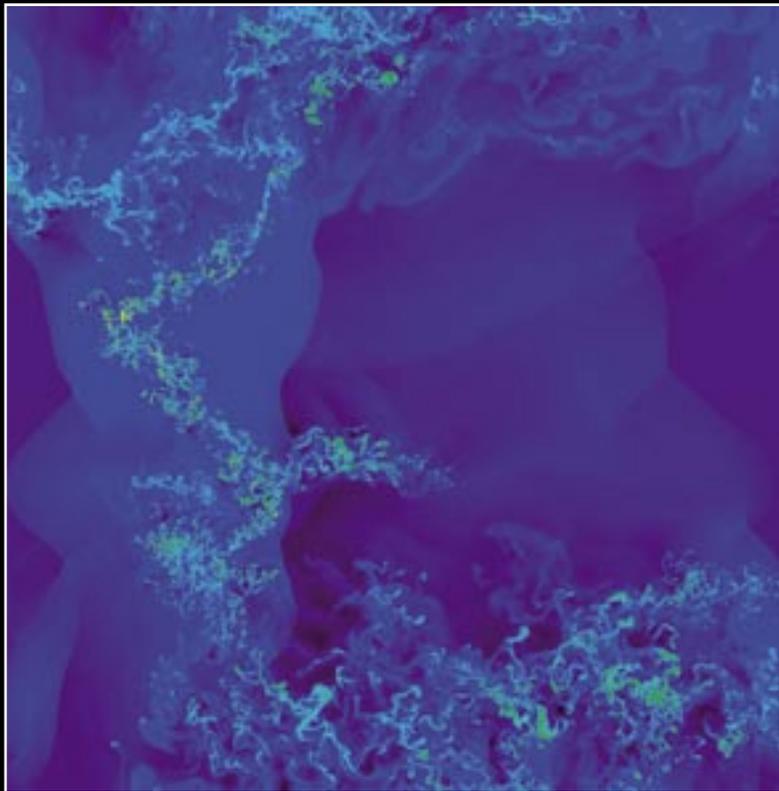
La simulation numérique remonte le temps. Un ou deux milliards d'années, c'est l'enfance, pour un univers qui pourrait souffler aujourd'hui treize milliards de bougies. À cette époque, on assiste à la formation de ses grandes structures : des filaments, des amas. L'image 3A (voir p. 26) montre l'univers à un milliard d'années. Les masses claires correspondent à des halos de matière noire, au sein desquels se forment des étoiles et donc des galaxies. On voit ainsi au centre droit un gros amas de galaxies en train de se former. Le coin d'univers représenté ici mesurait à l'époque 150 millions d'années-lumière de côté, mais du fait de l'expansion de l'univers, son contenu s'étend aujourd'hui sur un milliard d'années-lumière. Le modèle employé⁵ met en scène quelque dix milliards de particules. Il a été

exploité sur MareNostrum, le supercalculateur de type IBM-Cluster du Barcelona Supercomputing Center. À deux milliards d'années (photo ci-dessus), l'univers est un peu plus mûr, et comporte des galaxies déjà plus évoluées. On voit ici un petit cube de dix millions d'années-lumière, visualisé sous plusieurs aspects (verticalement) et « zoomé » trois fois (de gauche à droite). De haut en bas les images mettent en valeur successivement la matière noire (1^{re} ligne), le gaz, en densité de masse (2^e ligne), la température du gaz (3^e ligne) et les étoiles, ou, ce qui revient au même, la lumière visible (4^e ligne). À cette époque, le gaz « coule » le long de filaments vers le centre des halos de matière noire et y forme des disques « froids », les futures galaxies, où s'allument des étoiles.

(5) <http://www-dapnia.cea.fr/Phys/Sap/Actualites/Breves/teyssier070115/page.shtmlfr.shtml>

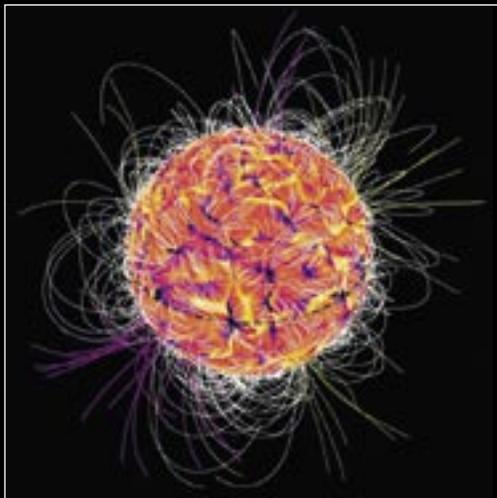
Très turbulent, le milieu interstellaire !

Cette simulation montre un « petit » morceau de galaxie de 60 années-lumière de côté. Elle met en scène les phénomènes de turbulence dans le gaz interstellaire, au sein duquel se forment les étoiles. Ces simulations précises du comportement du gaz interstellaire, et notamment de la turbulence qui peut s'y développer, permettent de mieux comprendre les premières phases de la formation des étoiles. Le parallélisme massif permet ici d'atteindre une haute résolution.

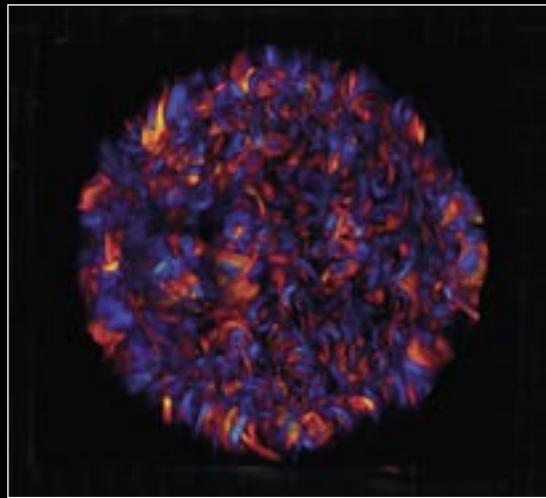


© E. Audit/CEA

Les caprices magnétiques du soleil



© A.S. Brun/CEA



© A.S. Brun/CEA

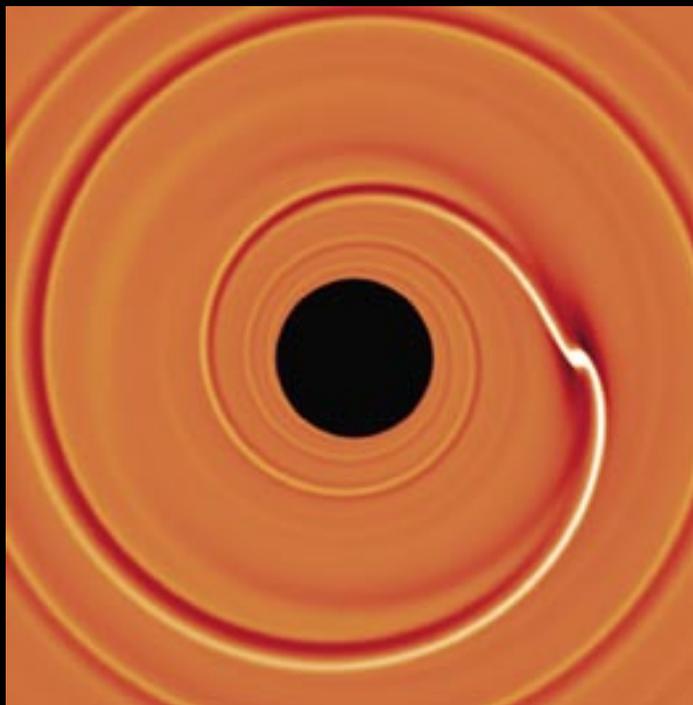
Ces images montrent la reconstruction du champ magnétique de la couronne solaire à partir d'une simulation tridimensionnelle haute résolution du magnétisme et de la dynamo solaires. Ce type de simulation ⁶ permet de mettre à jour les mécanismes

physiques à l'origine de la rotation différentielle du soleil, de sa circulation méridienne et de son magnétisme. Les couleurs représentent le champ magnétique à la surface et les traits sur le cliché de gauche, les lignes de champ magnétique dans la couronne solaire.

(6) <http://www-dapnia.cea.fr/Phys/Sap/Actualites/Breves/sbrun051115/page.shtml>

Naissance et migration des planètes

Il pourrait s'agir de notre système solaire lorsqu'il est né, il y a plus de 4 milliards d'années. Le soleil (au centre, non représenté) est alors entouré d'un anneau de matière, « protoplanétaire ». Une première planète (au centre de la tache claire, sur la droite) perturbe par sa présence (et sa masse) l'ensemble du disque, qui était initialement uniforme, mais dans lequel une structure spirale s'est développée, avec des régions sur-denses (en clair) et des régions sous-denses (sombres). Cette réponse du disque rétroagit sur la planète en exerçant sur elle des forces gravitationnelles qui vont accélérer ou ralentir sa rotation, et par là accroître ou diminuer son rayon orbital : c'est la « migration planétaire ». De telles simulations permettent d'expliquer l'emplacement actuel des planètes du système solaire, qui ont dû commencer leur formation ailleurs que sur leur orbite actuelle.

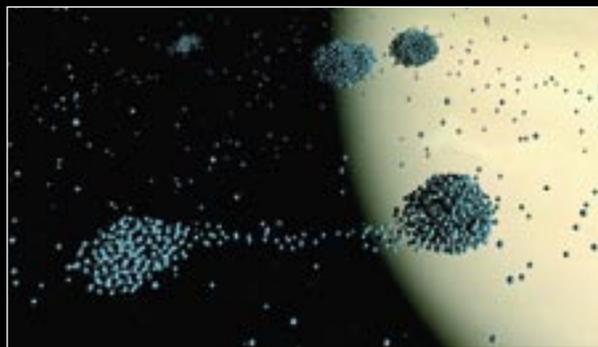


© F. Masset/CEA

Filaments et agrégats dans les anneaux de Saturne

On simule ici la physique des anneaux de Saturne à l'échelle de quelques dizaines de mètres. La matière des anneaux, des « glaçons » de quelques centimètres à un mètre, se rassemble tantôt en « filaments » tantôt en « agrégats ». Ces derniers sont de petits satellites de Saturne en formation, que des collisions peuvent aussi détruire. Les simulations numériques calculent donc le résultat de l'équilibre complexe entre la gravité qui tend à rassembler la matière, et les effets de marée qui tendent à étirer et déchirer tout objet qui se forme. Cassini (sonde Nasa/Esa en orbite autour de Saturne depuis 2004) a très récemment mis en évidence, dans les anneaux, les structures filamentaires prédites par les simulations.

Il est intéressant de noter que le mécanisme qui crée des filaments et des satellites dans les anneaux de Saturne (appelé « onde de Jeans ») est le même que celui qui, à beau-



© S. Chamois/Univ. Paris 7 & AMM/CEA. Post-production Frédéric Durillon/Animae

coup plus grande échelle, forme des étoiles dans les nuages moléculaires, et aussi les bras spiraux dans les galaxies. On voit donc bien ici la dimension « multi-échelle » de la physique de l'univers abordée par les simulations.

Explosion sur ordinateur

La puissance croissante des supercalculateurs ouvre la voie à des simulations de plus en plus réalistes de la physique des explosifs.

Dominique Ritman
est journaliste
scientifique

(1) Les équations de la dynamique moléculaire décrivent l'évolution, en fonction du temps, d'un système de particules interagissant entre elles.

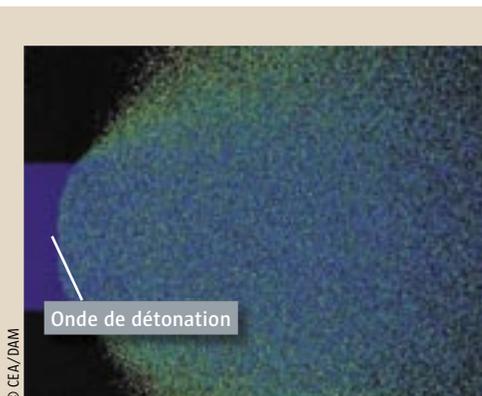
Il n'est guère d'expériences plus destructives que celles destinées à étudier l'onde de détonation se propageant dans un explosif. Par ailleurs, les mécanismes qu'elles engendrent se déroulent sur un temps très court et mettent en jeu des pressions et des températures extrêmement élevées, ce qui rend leur observation particulièrement délicate. C'est dire l'intérêt des simulations numériques réalisées à la Direction des applications militaires (DAM) du CEA, à Bruyères-le-Châtel (Île-de-France). Une centaine d'heures ont par exemple suffi à un millier de processeurs de Tera-10, le supercalculateur installé en 2006 au CEA-DAM, pour modéliser, sur la base d'équations de dynami-

que moléculaire⁽¹⁾ classique, la rencontre de deux ondes de détonation se propageant dans un milieu dense composé de millions de molécules en interaction mutuelle.

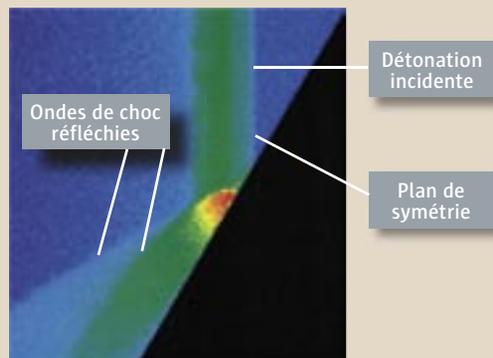
«La simulation du fonctionnement d'une arme impose donc une connaissance très fine de la physique des explosifs.»

Le fonctionnement d'un engin nucléaire nécessite une source d'énergie initiale très puissante et peu encombrante. Le choix s'est porté immédiatement sur les explosifs condensés. La simulation du fonctionnement d'une arme impose donc une connaissance très fine de la physique des explosifs. Principe de l'explosif :

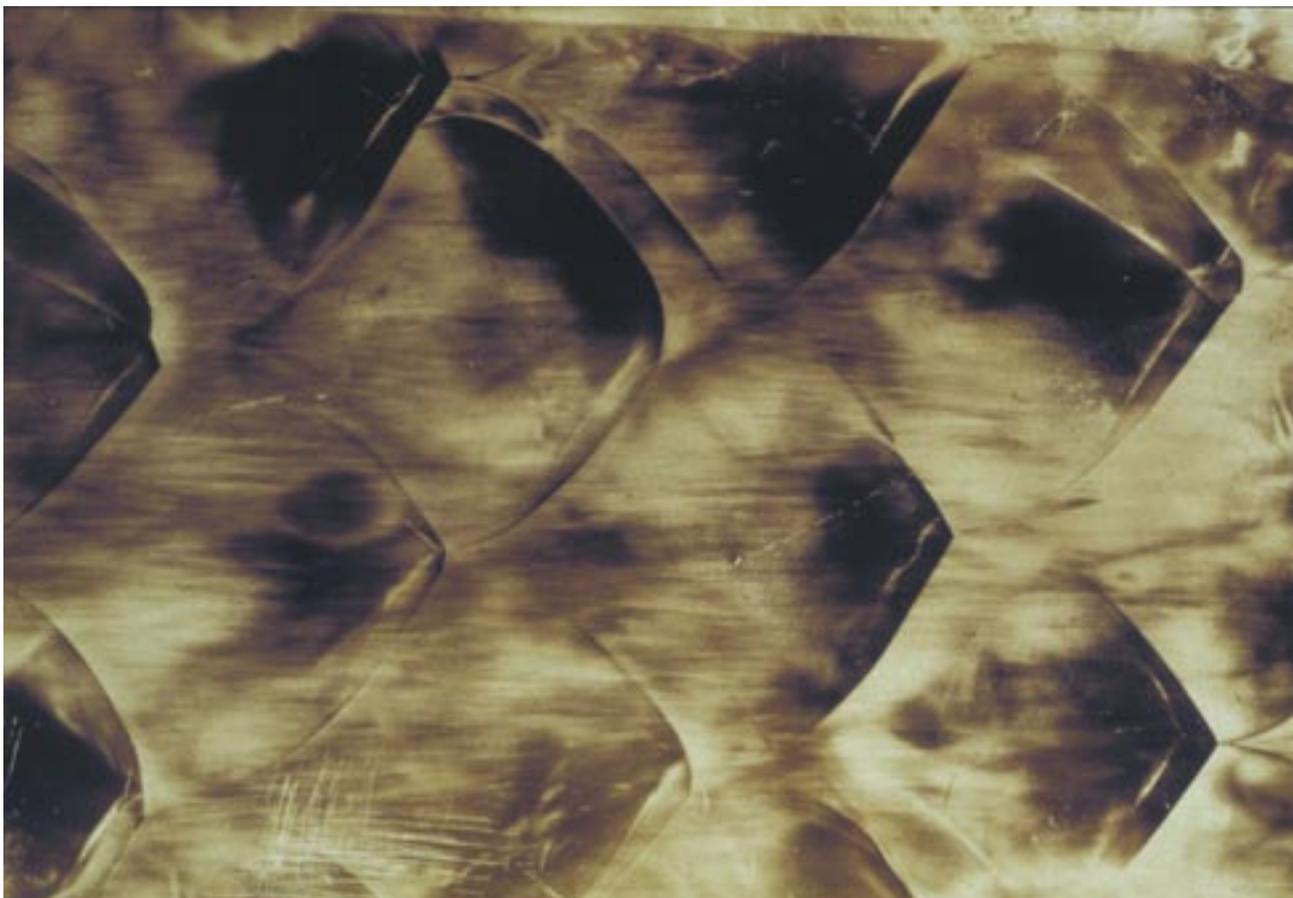
une onde de détonation (assimilable à une onde de choc réactive) se propage dans un matériau réagissant chimiquement et libérant par là même une grande quantité d'énergie, laquelle entre-



Propagation d'une onde de détonation dans une cartouche cylindrique sans confinement : la courbure de l'onde est une caractéristique importante, bien reproduite par une simulation de dynamique moléculaire.



Réflexion symétrique d'une onde de détonation plane. Dans ce cas complexe, deux ondes de choc sont réfléchies. Cette simulation de 10 millions de molécules a été réalisée sur 1 000 processeurs de la machine Tera 10.



© CNRS

tient la propagation de l'onde. « *Nous étudions notamment deux points cruciaux : d'une part les processus chimiques dans la zone où se produisent les réactions de décomposition de l'explosif (appelée zone de réactions), et d'autre part le comportement du gaz résultant de la combustion de l'explosif. La propulsion résulte de l'expansion de ce gaz* », explique Laurent Soulard, ingénieur à la DAM. Alors que les expériences réelles ne permettent d'évaluer que des grandeurs macroscopiques (pression, température, vitesse de la matière), les simulations réalisées à Bruyères-le-Châtel donnent accès aux mécanismes de décomposition et aux propriétés thermodynamiques de la zone de réactions à l'échelle atomique.

Quand deux ondes de détonation se rencontrent

Les simulations par dynamique moléculaire s'appuient sur la résolution des équations du mouvement, qui suppose le choix d'une fonction d'énergie potentielle pour décrire les interactions entre particules. La durée de simulation dépend de la complexité de cette fonction et du nombre de molécules. Dans un premier temps, les scientifiques ont choisi une fonction simple et ont validé la méthode sur des géométries planes ne nécessitant la prise en compte que

de quelques milliers de molécules. Ils ont ensuite réalisé des simulations plus proches de situations réelles, utilisant des géométries de l'explosif de type cylindrique et un nombre de molécules d'un ordre de grandeur supérieur.

Dernière simulation en date : celle de la rencontre de deux ondes de détonation, réalisée sur Tera-10, qui met en jeu une dizaine de millions de molécules. Ces calculs nécessitent un code de dynamique moléculaire massivement parallèle : c'est dans cette optique que le code STAMP a été développé à la DAM. De telles simulations offrent deux grands avantages, souligne Laurent Soulard. Elles permettent de connaître précisément et à chaque instant les valeurs de la température, de la pression... dans la zone de réactions. Elles assurent de plus que le résultat obtenu est la solution exacte du problème, s'affranchissant des problèmes délicats de maillage inhérents aux méthodes de la mécanique des milieux continus.

Reste à s'assurer que la simplicité de la fonction d'énergie potentielle choisie n'engendre pas de biais rédhibitoire. « *Nous travaillons aujourd'hui sur des fonctions plus réalistes*, souligne Laurent Soulard, *il faudra peut-être attendre la prochaine génération d'ordinateurs pour simuler des systèmes de type industriel. Mais nous préparons d'ores et déjà les outils qui seront alors nécessaires.* »

D. R.

Cellules d'ondes de détonation se propageant dans un mélange gazeux d'oxygène et d'argon : on observe que contrairement aux figures d'interférences classiques d'ondes lumineuses (dont la propagation obéit à un système linéaire décrit par les équations de Maxwell), les ondes de détonations, solutions d'équations non linéaires, conduisent à des structures géométriques complexes

Décrypter les mystères du sous-sol

La modélisation du sous-sol et la simulation des écoulements des fluides dans les réservoirs de pétrole sont des outils de premier ordre pour les compagnies pétrolières.

Yves Sciama
est journaliste
scientifique

Quarante millions de dollars en moyenne, et jusqu'à 65 millions de dollars ! Le prix d'un forage pétrolier en *offshore* profond en dit long sur l'importance des investissements aujourd'hui associés à l'exploitation pétrolière. L'époque pionnière où un forage rudimentaire, effectué à quelques dizaines de mètres sous la surface, faisait jaillir un geyser de pétrole appartient à jamais au passé. « *Aujourd'hui, le pétrole facile a fait place au pétrole technologique* », résume Stéphane Requena, spécialiste du calcul haute performance, en charge de l'évolution des moyens de calcul à l'Institut français du pétrole (IFP), à Rueil-Malmaison. Il s'agit désormais d'aller chercher l'or

noir dans des gisements très profonds, à haute température et haute pression, dans des sites particulièrement difficiles d'accès, ou encore de se tourner vers des hydrocarbures complexes à exploiter, par exemple des pétroles très visqueux.

Les énormes investissements que cela suppose engendrent des risques financiers considérables. Le taux de succès de l'exploration est de l'ordre de 33 %. Il n'a pas augmenté de manière significative mais les progrès de la sismique, avec une résolution d'une dizaine de mètres aujourd'hui, permettent de découvrir des réservoirs dans des zones géologiquement de plus en plus complexes, de plus en plus éloignées et de plus en plus profondes. Des incertitudes non négligeables subsistent cependant sur la quantité d'hydrocarbu-



Les compagnies pétrolières explorent désormais en offshore profond jusqu'à 2 250 mètres d'eau

res qui pourra en être retirée et sur l'étalement dans le temps de sa production.

Quelques explications sur la façon dont on extrait le pétrole permettent de comprendre le problème. Un gisement se résume, au fond, à un ensemble de roches poreuses gorgées d'hydrocarbures – le réservoir – coiffées de couches de roches imperméables – la couverture. En fonction des conditions de pression et de température, on trouve au sommet du gisement des hydrocarbures gazeux surmontant des hydrocarbures liquides – « l'huile », c'est-à-dire le pétrole – puis l'aquifère où la roche poreuse ne contient plus que de l'eau. Un gisement présente souvent une géométrie complexe, avec une

inclinaison globale des couches, des systèmes de failles, des hétérogénéités de la roche, etc.

La récupération se fait en forant un ou plusieurs puits dans la couche imprégnée d'huile, puits qui permettent de commencer à récupérer les hydrocarbures simplement sous l'effet de la pression élevée qui règne dans le réservoir. Dans un second temps, il faut mettre en place de nouveaux puits par lesquels sont injectés de l'eau ou du gaz, ces fluides « balayant » le gisement et « poussant » en quelque sorte les hydrocarbures vers les puits extracteurs. Éventuellement, pour améliorer encore la récupération, il est possible (moyennant un coût élevé) de forer de nouveaux puits dans les zones mal « balayées », ou bien d'injecter

des produits chimiques, du CO₂ ou encore de la chaleur sous forme de vapeur d'eau. Au final, le taux d'extraction atteint aujourd'hui en moyenne 33 % du pétrole initialement présent. Ce taux, relativement faible, a une importance cruciale : une variation de quelques pour cents représente des volumes colossaux et des sommes en proportion.

Avoir une vision dynamique de l'exploitation

Grâce notamment aux progrès de la sismique, on dispose désormais d'images assez fines des réservoirs en exploitation. Les modèles géologiques qui représentent ces structures le font à l'aide de mailles de seulement quelques dizaines de mètres de côté, et parfois de moins d'un mètre de hauteur. Un gisement s'étendant sur quelques kilomètres de longueur est alors modélisé à l'aide de maillages de plusieurs millions de cellules. Mais ce qui est décisif pour une bonne exploitation, c'est d'anticiper la façon dont les différents fluides vont se déplacer à l'intérieur du réservoir dès lors qu'il sera mis en production. Avec à la clef des questions très concrètes : où faut-il positionner les puits (injecteurs et producteurs) et combien faut-il en mettre pour optimiser la récupération ? Quels fluides faut-il choisir d'injecter, dans quelles couches et à quel débit (injecter trop vite a souvent pour effet d'abaisser le taux de récupération) ? Comment va évoluer dans le temps la production du gisement ? La difficulté est donc de passer de la vision fine mais statique que fournissent les géologues à une vision dynamique liée à l'exploitation.

Dès la fin des années 1960 apparaissent les premiers logiciels visant cet objectif. Depuis, ils n'ont cessé de s'améliorer, sous l'effet de l'augmentation de la puissance de calcul disponible et de l'amélioration de la compréhension des gisements. Le principe de ces modèles est simple : le réservoir est divisé en mailles ou cellules (plus grossières que celles qui constituent le modèle géologique) où les valeurs de porosité et de perméabilité sont supposées constantes, alors qu'elles varient bien sûr d'une maille à l'autre. Il s'agit ensuite de relier ces mailles entre elles par des équations simulant les écoulements en milieu poreux pour modéliser les déplacements des fluides. Si ce principe fondamental a peu évolué, les modèles actuels sont devenus des outils impressionnants à la fois par leurs performances et par la diversité des possibilités qu'ils offrent.

Modélisation en temps réel

Le nombre de mailles et la finesse des phénomènes représentés se sont considérablement accrus. « Aujourd'hui, nous sommes capables d'effectuer des simulations à plusieurs millions de mailles, avec de

Résultat d'une simulation dynamique de réservoir pétrolier, obtenu avec le logiciel First : vue 3D du maillage (à gauche), outils de gestion de « calage d'historique » et résultats de production aux puits (à droite, courbes de pression, débits, ratios).

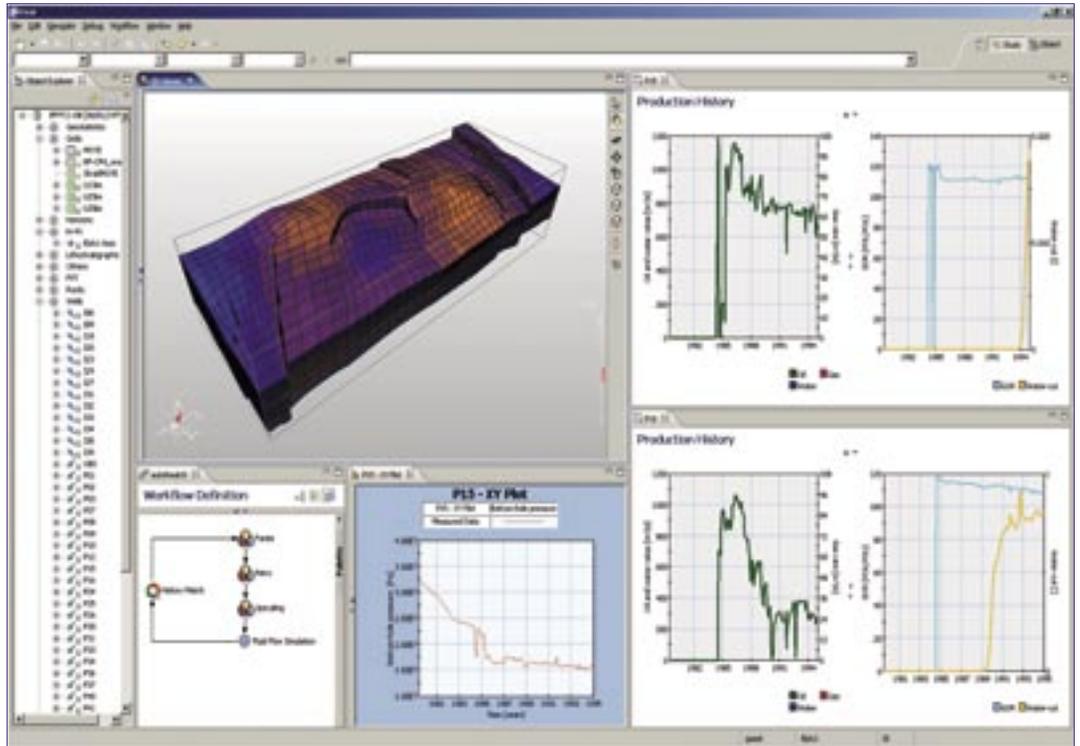


ILLUSTRATION : BEICIP FRANLAB

(1) First : plateforme de simulation de réservoirs développée par l'IFP. Des précisions sur le site de Beicip Franlab : <http://www.beicip.com>
 (2) Projet FAME 2 : www.fame2.org
 Pour en savoir plus : <http://www.ifp.fr>

3 à 7 inconnues par maille en moyenne, en seulement quelques heures, grâce à des ordinateurs parallèles multiprocesseurs », indique Patrick Lemonnier, chef du département « Simulation des écoulements et transferts en milieu poreux » à l'IFP. En outre, ces modèles, qui atteignent le million de lignes de code, évoluent désormais sans cesse par l'intégration permanente de données nouvelles. Ainsi, à mesure que l'exploitation du réservoir avance, les mesures réalisées dans les puits sont comparées aux prévisions du logiciel et utilisées pour rapprocher la simulation de la réalité (ce qu'on appelle le calage d'historique). De même,

on assiste depuis quelques années au développement de la sismique dite 4D : des campagnes sismiques sont désormais menées sur les mêmes gisements, tous les deux ans, afin de suivre l'évolution des fluides. Les modèles actuels intègrent également les données fournies par ces campagnes au fur et à mesure qu'elles sont produites, les données fournies en temps réel par certains capteurs (de pression par exemple) dont sont équipés les puits ainsi que les informations obtenues lors du forage de nouveaux puits. Cette évolution générale vers le temps réel et vers l'intégration de données très diverses est bien sûr très gourmande en puissance de calcul. Une puissance qui, par exemple à l'IFP, a été multipliée par 45 depuis 2003, et qui va continuer à croître rapidement.

« Aujourd'hui, le pétrole facile a fait place au pétrole technologique »

« Nous développons actuellement une plateforme de simulation, baptisée First⁽¹⁾, distribuée à partir d'avril 2007 par notre filiale Beicip Franlab, qui permettra d'accroître sensiblement la finesse du modèle réservoir, explique Jean-François Magras, spécialiste du calcul haute performance appliqué aux écoulements des fluides en milieux poreux à l'IFP.

Et nous participons au projet FAME 2⁽²⁾ de la technopole Ter@tec (lire article sur Ter@tec p. 40) qui devrait nous permettre d'ici quelques années de simuler de manière industrielle des modèles à 100 millions de mailles, avec plusieurs centaines de processeurs, ce qui serait une première mondiale. »

Aujourd'hui, avec un taux de récupération du pétrole en progression constante et de nouvelles architectures de puits (puits horizontaux, multi-branches...) qui permettent de multiplier la production par un facteur 10 par rapport à des puits verticaux, la modélisation de réservoir a gagné ses galons. Elle est devenue un élément clé dans la prise de décision, et l'accroissement futur de la puissance des ordinateurs permet d'espérer l'amélioration continue de ses performances. Il reste que la complexité des hydrocarbures dont on envisage aujourd'hui l'exploitation est porteuse de nouveaux défis qui ne seront pas simples à relever.

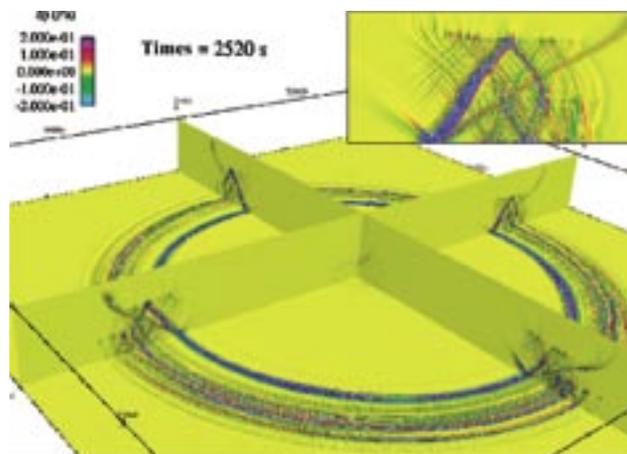
Y.S

La simulation pour surveiller les explosions nucléaires

Capteurs d'infrasons, modèles de propagation et supercalculateurs : trois conditions pour localiser les essais nucléaires.

Les infrasons sont perçus par certains animaux (éléphants, dauphins...) mais inaudibles par notre espèce humaine : ils sont bien trop graves, de fréquence inférieure à 20 hertz. En revanche, leur aptitude à se propager sur de longues distances les rend particulièrement précieux pour détecter un événement sonore se produisant à plusieurs centaines de kilomètres, voire à plusieurs milliers. D'où l'idée de s'en servir, en se dotant de capteurs adéquats, pour veiller au respect du Traité d'interdiction complète des essais nucléaires (TICE). Mais comment, à partir de signaux enregistrés, identifier et localiser une explosion prohibée, parfois très lointaine ? La réponse nécessite des modèles tridimensionnels de propagation du son, et surtout des calculateurs très puissants compte tenu de la taille de ces modèles. Au CEA, la progression d'infrasons d'un dixième de hertz a pu être simulée en 10 heures avec un peu moins de 2 000 des processeurs de Tera-10.

La taille des calculs résulte du grand nombre de paramètres (la propagation des infrasons est fortement dépendante des conditions météorologiques, dont la vitesse des vents), mais aussi de la finesse du maillage de la région étudiée. Les chercheurs de la Direction des applications militaires (DAM) se sont appuyés sur des données recueillies à l'occasion d'une expérience américaine, réalisée en 1984 au Nouveau-Mexique. « Le défi était de retrouver, par des méthodes numériques, ce qu'avaient enregistré les capteurs



Simulation de la propagation d'infrasons sur Tera-10 : l'onde de pression (infrasons) créée au centre de l'image s'est propagée tout autour, ici sur 2 000 km. On voit, en coupe, qu'elle se réfléchit à environ 120 km d'altitude : l'agrandissement montre la riche structure fréquentielle et la légère dissymétrie due au vent.

il y a une vingtaine d'années », souligne Bruno Després, directeur de recherche au Département sciences et simulation de l'information (DSSI) de la DAM. « Les résultats ont été au-delà des attentes, la précision s'est avérée excellente », ajoute-t-il.

Les modélisations doivent tenir compte des vents

Le volume sur lequel a été effectuée la simulation s'étendait sur une région de superficie grande comme huit fois la France jusqu'à une hauteur de 200 kilomètres. Quant au maillage, il a été déterminé de manière à accéder à des détails de l'ordre de 400 mètres, d'où pas moins 10 milliards de mailles. La simulation a notamment permis de vérifier l'importance des vents : « À 100 km d'altitude, l'atmosphère se comporte un peu comme un guide d'onde : les infrasons émis au sol rebondissent, en quelque sorte, et redescendent alors vers le sol, ce qui contribue au fait qu'on puisse les écouter à de gran-

des distances », explique encore Bruno Després. Les signaux recueillis par chaque capteur dépendent donc fortement des vents atmosphériques, ce dont le modèle a très bien rendu compte.

Autre résultat remarquable : la qualité des méthodes numériques et du code, fruit d'une collaboration entre le DSSI et le Département analyse et surveillance de l'environnement (DASE) de la DAM, ainsi que de son optimisation pour la machine Tera-10. Les chercheurs ont en effet vérifié qu'un doublement du nombre de processeurs se traduisait presque exactement par une réduction d'un facteur deux du temps de calcul : ce critère caractérise ce que l'on nomme la scalabilité, ici voisine de 100 %. Et au-delà, note Bruno Després, rien ne s'oppose, dans les principes, à l'utilisation de telles méthodes dans d'autres applications, par exemple pour simuler la propagation dans le sol d'ondes acoustiques et élastiques lors de tremblements de terre. **D.R**

Révéler les dessous du « mauvais » prion

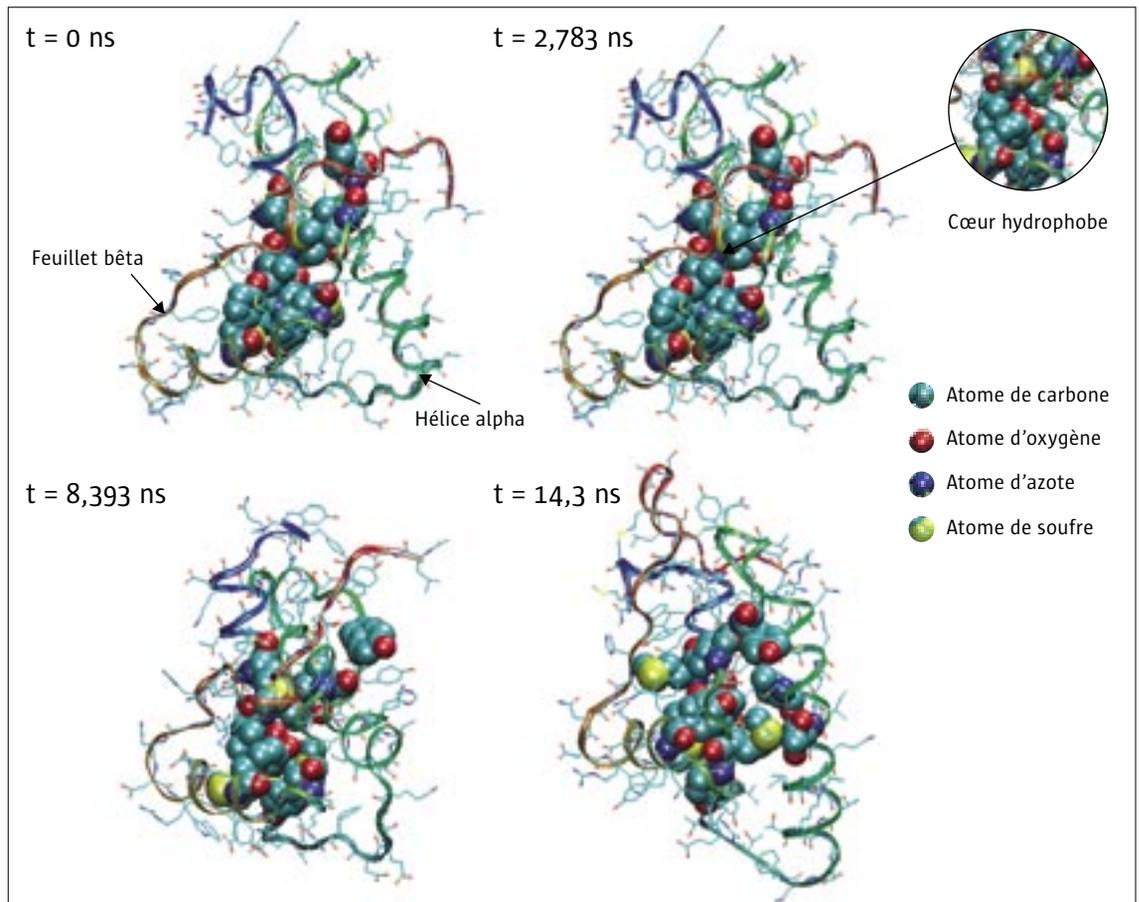
La simulation numérique permettra-t-elle un jour de révéler la structure en trois dimensions d'un prion pathogène ?

Anne Lefèvre-Balleydier
est journaliste scientifique

Maladie de la vache folle, tremblante du mouton, nouveaux variants de Creutzfeld Jacob... un même agent infectieux est en cause dans tous ces syndromes : une forme anormalement repliée de la protéine dite prion (PrP), qui s'agrège en fibres puis en plaques dans le cerveau, et conduit à la dégénérescence des neurones. Pour lutter contre ce prion pathogène (PrP^{sc}), il faut connaître sa structure tridimensionnelle, autrement dit la façon dont la lon-

gue chaîne d'acides aminés se replie en boucles, hélices et feuillets. Et c'est là que le bât blesse. « Pour déterminer la structure d'une protéine de façon expérimentale, il faut pouvoir en disposer sous forme de cristal ou en solution (on peut alors l'analyser respectivement par diffraction des rayons X et par résonance magnétique nucléaire, NDLR). Mais le prion pathogène est amorphe et insoluble », explique Vincent Croixmarie, un ancien docteur du service de physique de la matière condensée du CEA, aujourd'hui chercheur chez Servier.

Simulation du dépliement du prion normal, du pas de temps zéro jusqu'à 14,3 nanosecondes (ns) à 177°C. La protéine se déplie bien : son cœur se disloque et les orientations relatives de ses structures secondaires changent. Mais pas de trace des nombreuses structures en feuillet bêta, caractéristiques des prions pathogènes.



DR

Mettre à jour l'étape dans laquelle le prion pathogène reste « coincé »

Il y a néanmoins moyen de contourner le problème. Car comme il le souligne, « une protéine est comme une pelote de laine que l'on peut dérouler puis enrrouler : en la dépliant, on peut donc visualiser les étapes de son repliement ». Durant la thèse⁽¹⁾ qu'il a soutenue au printemps 2005, Vincent Croixmarie a ainsi simulé le dépliement de la protéine prion normale en autant d'étapes que possible, pour mettre au jour l'étape intermédiaire dans laquelle le prion pathogène reste « coincé », et révéler ainsi sa structure. Seul indice à sa disposition : les uniques données expérimentales sur la structure du prion pathogène, qui font état d'un grand nombre de feuillettes dans la protéine avant la formation d'agrégats.

« Lâchons le mot, s'exclame Vincent Croixmarie, ce fut un échec. Mais nous avons beaucoup appris sur le comportement des hélices et feuillettes qui composent une protéine ». Le jeune chercheur est donc parti de la protéine prion normale, qui étant soluble, a une structure tridimensionnelle connue. Recourant à la simulation par dynamique moléculaire⁽²⁾, il a modélisé cette protéine comme un assemblage d'atomes (définis, entre autres, selon leur masse et leur charge électrique), liés entre eux par des constantes de raideur décrivant les forces qui s'exercent. Puis Vincent Croixmarie a choisi d'imposer un certain nombre de conditions – par exemple, des températures plus ou moins élevées pour déplier la protéine – et d'observer ce qui en résultait en découpant le temps en tranches ultrafines (10⁻¹⁵ secondes), pour voir « le film » au ralenti. Il n'y serait jamais parvenu sans la puissance d'un supercalculateur, en l'occurrence Tera-1, le premier calculateur de la génération « Tera » installé en

2001 au CEA. Car la portion de prion qu'il a étudiée comportait 102 acides aminés, soit environ 1 700 atomes. Chaque atome étant défini selon ses trois coordonnées, le nombre de variables à faire évoluer est énorme. D'autant que dans les cellules, la protéine est entourée d'eau : il faut donc inclure dans la simulation suffisamment de molécules d'eau.

Au total, la trajectoire de dépliement de la protéine la plus longue comportait 66 millions de pas de temps, soit 66 nanosecondes de dépliement : il lui a fallu un an pour l'obtenir. Cela paraît considérable. Mais c'était encore insuffisant. « Nous

pensions voir disparaître des hélices dites alpha et obtenir un taux important de certains feuillettes appelés bêta, que l'on sait préalables à la formation des agrégats du prion pathogène, commente Vincent Croixmarie. Or notre protéine s'est révélée d'une très grande stabilité : à très haute température, la chaleur l'a détruite et sa structure est devenue aléatoire, et à température moins élevée, les hélices sont restées en place. » La protéine s'est donc dépliée trop vite, sans laisser le temps aux feuillettes bêta d'apparaître. Autrement dit, le film de ce dépliement a été trop accéléré. « Pour avoir une chance de capturer l'instant où se forment ces fameux feuillettes, il faudrait utiliser une température de simulation plus basse, ajoute-t-il. Mais cela reviendrait à augmenter encore la quantité de calculs... »

D'autres pistes à explorer

Vincent Croixmarie pense aussi à une autre piste : « Comme il y a bien plus d'eau que de protéine, la simulation des molécules d'eau mobilise 80 à 90 % du temps de calcul. Or il existe des moyens de résoudre ce problème, en simulant la présence d'eau par ses propriétés : cette technique de l'eau " implicite ", encore en développement, devrait permettre de diminuer le temps de calcul nécessaire à l'obtention d'un pas de temps ». Bien qu'il mette d'autres bémols à la réussite de la simulation, il est plutôt optimiste. D'après lui, on peut aussi imaginer que le prion normal

ne se soit pas virtuellement transformé en prion pathogène, parce que le processus a été trop simplifié : dans le cerveau, le repliement anormal ne se produit peut-être qu'avec un environnement particulier.

Reste que si ses recherches n'ont pas révélé la structure de la protéine pathogène, elles ont en revanche permis d'en savoir plus sur le processus de

formation des fibres caractéristiques du prion pathogène, similaires à celles retrouvées dans certains types de diabètes, ou encore dans la maladie d'Alzheimer. Partant d'un fragment de la protéine prion normale avec une structure en feuillet, Vincent Croixmarie a simulé l'assemblage de plusieurs d'entre eux en une fibre ce qui lui a permis d'observer quels atomes et quelles forces de liaisons étaient en jeu. À terme, cela pourrait ouvrir de nouvelles pistes thérapeutiques. D'ici là, le développement des méthodes de simulation et des puissances de calcul auront peut-être permis de mettre enfin au jour les dessous du « mauvais » prion...

A.-L. B.

(1) Vincent Croixmarie, thèse de doctorat, Paris 11 Orsay : Etude du dépliement de la protéine PrP^c réduite par simulation de dynamique moléculaire. Modélisation de fibres amyloïdes peptidiques, 2005.

(2) La dynamique moléculaire simule les mouvements des atomes de systèmes moléculaires et l'évolution de leur configuration spatiale. À partir d'équations de la mécanique classique, elle décrit les trajectoires des atomes en fonction du temps.

Philippe Miltin :

« Les supercalculateurs, un gage de souveraineté et de compétitivité »

Le vice-président de Bull Produits et Systèmes considère les supercalculateurs comme un passage obligé pour l'innovation des entreprises européennes et pour la recherche. Ce sont des outils stratégiques que l'Europe doit soutenir.

La Recherche. Avec Tera-10 installé au CEA fin 2005 et la nouvelle machine commandée par le CCRT⁽¹⁾, Bull pose un pied dans la cour des grands fabricants de supercalculateurs. Quel est pour Bull l'enjeu du marché du calcul scientifique et technique ?

Philippe Miltin. Il est essentiel. Bull souhaite continuer à développer sa position de constructeur européen innovant. Le calcul haute performance est l'un des éléments clés de notre stratégie. Le marché du calcul scientifique et technique est celui qui connaît une des plus fortes croissances en informatique avec 11,5 milliards d'euros prévus en 2010 contre 8 milliards en 2006, soit une progression de 9 %. Notre innovation technologique se situe dans la partie *hardware* de nos plateformes : nous les développons en intégrant des nouvelles technologies qui servent également dans nos solutions pour les infrastructures et les applications du monde de la gestion.

De plus, de nouveaux besoins en fortes puissances apparaissent, qui ne se limitent pas uniquement à l'ingénierie. Ainsi, les banques demandent à leur tour davantage de puissance afin de réaliser de purs calculs financiers, pour mesurer et gérer les risques à partir de modèles mathématiques. La simulation numérique permet aussi de concevoir de nouveaux produits, de les mettre plus rapidement sur le marché, avec des coûts réduits de développement, et des économies radicales à la clef, comme dans l'automobile ou le pétrole.

Notre objectif est de prendre 10 % du marché européen. Nous voulons augmenter fortement notre chiffre d'affaires direct dans le calcul scientifique et pénétrer de nouveaux comptes grâce à notre stratégie « d'architectes d'un monde ouvert » basée sur l'utilisation de composants standards autour de nos serveurs NovaScale⁽²⁾. Ils sont à base de processeurs Intel (Itanium et Xeon). Nos réseaux d'interconnexion entre les différents serveurs pour les fédérer en grappes (*clusters*) sont également basés sur des technologies du marché. Notre valeur ajoutée tient à cette architecture de plateforme, à son intégration et aussi à sa composante logicielle. Nous avons ainsi développé un savoir-faire logiciel qui nous permet d'utiliser au mieux ces plateformes, d'optimiser, de rationaliser et de simplifier l'emploi des applications de nos clients.

L. R. Quel a été l'apport du contrat Tera-10 pour la société ?

P. M. Avec Tera-10, première plateforme européenne par sa puissance, nous avons gagné en crédibilité. Nous souhaitons maintenant déployer ce savoir-faire dans l'ensemble des industries comme l'automobile, le pétrole, les sciences de la nature, la météorologie, la finance... La commande de décembre dernier par le CCRT d'un supercalculateur de 43 téraflops⁽³⁾ est une étape importante de ce déploiement. Cette machine prévue pour être opérationnelle en 2007 sera parmi les plus puissantes

au monde. Elle sera à la disposition de la communauté scientifique et industrielle. D'autres clients nous font déjà confiance : l'université de Manchester en Grande-Bretagne, Dassault Aviation, l'université polytechnique de Valence, celles d'Utrecht et de Reims, le centre océanographique de Southampton, Pininfarina en Italie. Tera-10 a été un projet structurant. Nous vendons désormais ses briques de base à différents clients dans les domaines industriels et universitaires.

Au-delà des enjeux de souveraineté nationale et des aspects économique et industriel, il est important de considérer aussi les enjeux de compétitivité européenne en matière de recherche. Seuls 15 % des supercalculateurs du Top 500⁽⁴⁾ sont situés en Europe et 49 % de leurs utilisateurs sont des industriels. Les autres sont issus du domaine académique, de la recherche et de la défense.

« Nous souhaitons maintenant déployer ce savoir-faire dans l'ensemble des industries »

Le retard de l'Europe dans la pratique du calcul intensif et dans ses équipements est donc d'autant plus inquiétant. Les États-Unis ont pris conscience que le calcul scientifique est fondamental pour la compétitivité des entreprises et la sécurité nationale : ils financent largement leurs industriels, avec deux projets de la Darpa (l'agence américaine de la défense) de 250 millions de dollars chacun, l'un pour Cray et l'autre pour IBM, pour produire des machines pétaflops⁽⁵⁾ d'ici



© Jean-Michel RILLON

(1) CCRT : Centre de calcul pour la recherche et la technologie (Bruyères-le-Châtel, Essonne).

(2) Serveurs NovaScale : <http://www.bull.com/fr/novascale/index.php>

(3) Téraflops : mille milliards d'opérations par seconde.

(4) Top 500 : classement des 500 calculateurs les plus puissants au monde.

(5) Pétaflops : un million de milliards d'opérations par seconde.

2010 d'une puissance près de vingt fois supérieure à celle de Tera-10. Il est ainsi critique que Bull, seul industriel européen à développer une technologie dans le domaine des supercalculateurs scientifiques, puisse conserver son avantage compétitif. Certes, la Communauté européenne a également des projets de supercalculateurs. Mais la Chine et le Japon ont des projets équivalents, voire supérieurs... Les enjeux du calcul scientifique dépassent largement le niveau des entreprises, ils sont nationaux et supranationaux.

L. R. Dans cette compétition internationale où vous affrontez des entreprises telles qu'IBM, Cray, SGI, Dell... quels sont vos atouts ?

P. M. Il existe trois grandes contraintes à l'installation de grands systèmes : la surface au sol importante qu'ils nécessitent, une forte consommation électrique et la dissipation thermique. Bull apporte un

savoir-faire dans le refroidissement, dans l'optimisation de la gestion de l'énergie et dans la densité de la plateforme. Notre valeur ajoutée se situe également dans l'intégration de composants standard du marché autour d'une architecture serveur que nous avons développée (notre prochaine génération sera autour du bus CSI d'Intel) et dans le développement de la propriété intellectuelle autour de la pile logicielle. Le fait d'utiliser Linux et des logiciels *open source* permet à nos clients de travailler sur le code source de la suite logicielle que nous développons. Nos concurrents ont davantage de technologies propriétaires, disposent de leurs propres microprocesseurs, avec la lourde charge de développer leur technologie, avec parfois leur propre réseau d'interconnexion et leur propre système d'exploitation. Notre concentration sur le marché européen est aussi un avantage. Nous

avons enfin une réelle valeur ajoutée sur des éléments qui s'ajoutent aux briques de base, par exemple les accélérateurs dédiés ou le *packaging* de nos serveurs afin de réduire la densité thermique. Ces accélérateurs sont spécialisés dans le calcul d'algorithmes spécifiques tels les FFT (Fast Fourier Transforms) très utilisés en sismique ou les transpositions et inversions de matrices en mécanique des fluides. Ils permettent de calculer 10 à 100 fois plus vite que les microprocesseurs usuels, pour une consommation électrique 2 à 5 fois moindre. Notre différenciateur se situe et se situera de plus en plus, dans la qualité de l'intégration de tous les composants matériels et logiciels des supercalculateurs ainsi que dans la qualité de nos équipes qui accompagnent nos clients en optimisant leurs applicatifs sur nos serveurs.

Propos recueillis par Didier Gout

Ter@tec

Une référence du calcul intensif en Europe

Christian Saguez
est président
de Ter@tec

L'ensemble des acteurs scientifiques, économiques et industriels reconnaissent le rôle stratégique des technologies numériques et tout particulièrement de la simulation haute performance. La technopole Ter@tec et ses 43 partenaires⁽¹⁾, sociétés et organismes de recherche, a constitué, depuis sa création en août 2005, un environnement unique pour répondre à ces défis en fédérant l'ensemble des compétences françaises en calcul intensif autour de grands projets R&D collaboratifs. Ter@tec fournit un accès à des moyens de calcul et de stockage d'informations parmi les plus puissants d'Europe. Elle génère une dynamique économique fortement créatrice d'emplois. Cela doit permettre à la France et à l'Europe de retrouver, dans une compétition mondiale extrêmement active, le rôle de premier plan qu'elles n'auraient jamais dû quitter.

Des projets de recherche d'envergure

Membre du pôle de compétitivité System@tic, Ter@tec en est un acteur majeur dans le cadre de grands projets comme FAME 2 (Flexible Architecture and Multiple Environment) qui vise à développer une nouvelle génération de serveurs pour le calcul intensif et le traitement de l'information ; CARRIOCAS (Calcul réparti sur réseau Internet optique à

capacité surmultipliée) pour le développement et la mise en place de réseaux grande vitesse à 40 gigabits par seconde ; IOLS (Infrastructure et outils logiciels pour la simulation). Ces trois projets doivent permettre de disposer de l'ensemble des outils de base pour le calcul haute performance. Deux nouveaux projets démarrent : POPS (Péta opérations par seconde) et EHPOC (Environnement haute performance pour l'optimisation et la conception).

De nombreux projets sont soutenus par l'ANR⁽²⁾. Citons PARA (Parallélisme et amélioration du rendement des applications), LN3M (Logiciels nouvelle génération pour la modélisation multi-échelle des matériaux), NUMASIS (Optimisation des performances des applications en sismologie) et la plateforme SCOS qui vise à fédérer les initiatives françaises dans le domaine des logiciels libres en calcul intensif. Par ailleurs, ont vu le jour des laboratoires communs de recherche entre le CEA, l'université Versailles Saint Quentin (UVSQ), l'École centrale Paris et l'ENS Cachan.

Des infrastructures uniques

Avec l'achat par la communauté de communes de l'Arpajonnais (CCA) et le CEA d'une zone d'activités à Bruyères-le-Châtel, Ter@tec va développer une activité économique importante. Un

vaste ensemble immobilier sera construit pour accueillir entreprises et laboratoires ainsi qu'un centre de calcul pour l'installation dès 2010 de machines pétaflopiques, créant ainsi la première technopole européenne du calcul intensif. Dès 2007, la puissance de calcul disponible dépassera les 50 téraflops avec les centres CCRT (Centre commun recherche et technologies), Bull et HP et devrait atteindre le pétaflop en 2011, en synergie avec les initiatives européennes.

Des actions de formation et de promotion

Ter@tec a mis en place un accord de collaboration avec trois masters spécialisés en simulation haute performance : MN2MC (Méthodes numériques pour les modèles en milieux continus), M2S (Modélisation et simulation) et COSY-AHP (des concepts aux systèmes – architecture haute performance). Des activités de formation continue démarrent aussi cette année, notamment à destination des PME-PMI pour lesquelles la maîtrise des outils de calcul intensif est devenue une nécessité. Par ailleurs Ter@tec a, dès 2006, lancé un « Forum européen du calcul haute performance » autour d'un ensemble de conférences et d'une vaste exposition. L'édition 2007 aura lieu le 20 juin à l'université de Versailles Saint Quentin.

C. S.

(1) Ter@tec regroupe des entreprises informatiques (Aria Technologies, Bull, Cluster Vision, CS, Data Direct Networks, Distène, ESI Group Eurobios, Fluent, Fujitsu, HP, Intel, Numtech, OpenCascade, Oxalya, Serviware, SGI, Sun, Transtec), des industriels (Airbus, Bertin Technologies, CSTB, Dassault Aviation, EDF, Principia, Snecma, STMicroelectronics, Total), des universités et laboratoires de recherche (CEA, Cenaero, Cerfacs, CNRS, École centrale Paris, ENS Cachan, IFP, Inria, INT, École des mines Paris, Supélec, UVSQ) et des collectivités locales (CCA, mairie d'Ollainville, mairie de Bruyères-le-Châtel).

(2) ANR : Agence nationale pour la recherche.

En juillet dernier, le gouvernement décidait de créer une société civile dédiée au calcul intensif et baptisée Genci. François Goulard, ministre chargé de la recherche nous en décrit les objectifs.

Genci

Grand équipement national en calcul intensif

Le calcul intensif est essentiel pour bénéficier des avancées de la modélisation et de la simulation numérique. A l'instar de la théorie et de l'expérimentation, la simulation numérique est incontournable aussi bien dans le domaine de la recherche que dans celui de l'industrie. C'est un levier déterminant du progrès dans les domaines du climat, de l'astrophysique, de la combustion, de l'aéronautique, de la production pétrolière, des technologies nucléaires, des nanotechnologies, de la chimie, de la pharmacie...

Afin que la France retrouve une puissance de calcul compatible avec son rang de puissance scientifique et industrielle, et se maintienne dans la compétition internationale, nous avons mis en place une organisation permettant de fixer les orientations stratégiques et de réaliser les investissements prioritaires.

Aux côtés du CEA (20 %), du CNRS (20 %) et des universités (10 %), l'Etat est fondateur de la société civile Genci et lui apportera 50 % de son budget annuel. Premier pilier de ce dispositif pour rationaliser la politique française d'investissement en grands calculateurs et augmenter significativement les moyens de la France dans ce secteur, Genci est chargé de mettre en place et d'assurer la coordination des principaux

équipements des grands centres nationaux civils. Ses équipements seront ouverts à toutes les communautés scientifiques intéressées, académiques ou industrielles. Les coopérations entre les équipes publiques et privées seront privilégiées. Son budget annuel sera de 25 millions d'euros doublant ainsi les moyens mobilisés par la France pour le calcul intensif.

Le choix d'une société civile, déjà adopté pour d'autres grandes infrastructures scientifiques comme le synchrotron Soleil dans l'Essonne ou l'Institut Laue-Langevin à Grenoble, assure une gestion plus souple et permet, notamment, de prendre des participations dans les infrastructures internationales ou européennes de même nature. Il facilite aussi les partenariats de recherche en simulation numérique avancée avec

les collectivités régionales ou territoriales et les grandes sociétés de technologie avancée.

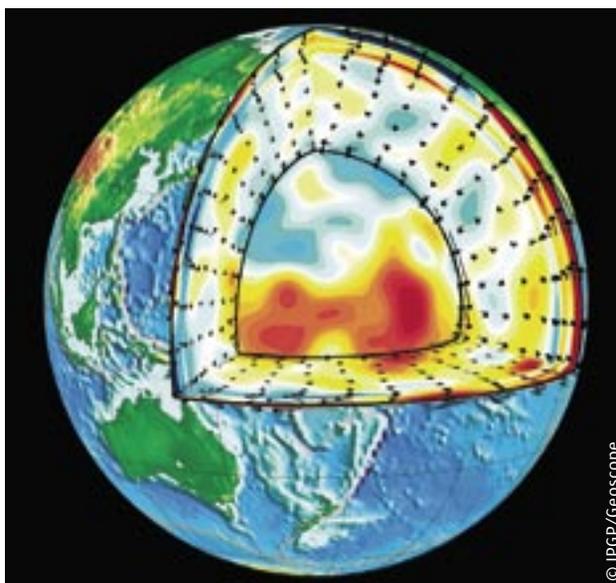
Parallèlement, le deuxième pilier du dispositif, le comité stratégique du calcul intensif sera prochainement mis en place. Il a pour objectif de favoriser la consultation prospective sur les besoins de calcul de l'ensemble de la communauté scientifique française, de nourrir ainsi l'élaboration de la politique de l'Etat dans ce domaine, de fédérer l'ensemble des intervenants et de coordonner les investissements.

Avec un plan d'investissements de 25 millions d'euros/an sur 4 ans, la France sera au niveau des meilleurs pays européens et pourra s'insérer harmonieusement dans la politique communautaire en cours d'élaboration.

F. G.



François Goulard
Ministre délégué à l'Enseignement supérieur et à la Recherche



© IPGP/Geoscope

Les simulations en sciences de la Terre nécessitent des puissances de calcul croissantes. Sur cette coupe schématique du globe, centrée sur le Pacifique, les hétérogénéités de vitesses dans le manteau terrestre ont été modélisées par tomographie (zones lentes en rouge, rapides en bleu). On en déduit les mouvements de convection (champ de vitesse représenté par les flèches)

Associations

ORAP

www.irisa.fr/orap

Association qui vise à diffuser l'usage du calcul parallèle par le biais de manifestations ou de lettres d'information.

TERATEC

www-dam.cea.fr/statique/ouverture/teratec.htm

Avec 43 membres, Teratec est l'association de référence des utilisateurs et fournisseurs de calcul haute performance.

Sites d'information

CLUSTER-HPC.COM

www.cluster-hpc.com

Quotidien web d'information sur le calcul haute performance. Ce site français propose également une newsletter hebdomadaire.

CLUSTER MONKEY

www.clustermonkey.net

Site de ressources techniques complet sur le clustering. Une bonne source d'information en anglais.

TOP 500 DES SUPERCALCULATEURS

www.top500.org/

Classement des 500 calculateurs les plus puissants au monde.

GROUPE CALCUL DU CNRS

<http://calcul.math.cnrs.fr>

Ensemble de ressources sur le calcul haute performance en France. À souligner le lien vers le projet CIEL (Codes Informatiques en Ligne).

Acteurs de référence

BULL

www.bull.com/fr/hpc

Découvrez les gammes HPC du constructeur français. Avec ses serveurs NovaScale, il a livré le plus puissant calculateur français.

CERFACS

www.cerfacs.fr

Présentation des recherches du Centre européen de recherche et de formation avancée en Calcul Scientifique : algorithmique parallèle, traitement de l'image et du signal et leurs applications associées.

COMMISSARIAT A L'ÉNERGIE ATOMIQUE

www.cea.fr

Au travers du site principal du CEA, retrouvez des dossiers sur le nucléaire, la fusion contrôlée, les nanotechnologies, le climat et une description des grands thèmes de recherche de l'organisme.

CRAY

www.cray.com

Acteur historique du calcul avec ses plateformes massivement parallèles.

HP

www.hpl.hp.com/techreports

Retrouvez publications et offres HPC d'HP, qui fournit plus de 30% du marché mondial.

IBM

www.research.ibm.com

Mine d'informations sur les programmes de recherche HPC menés par IBM.

INRIA

www.inria.fr

Des dossiers thématiques sur la modélisation du vivant, les neurosciences et l'algorithmique, le Grid-computing, et une description des programmes de l'institut liés à la simulation haute performance (Climé, Gamma...)

INSTITUT FRANÇAIS DU PÉTROLE

www.ifp.fr

Présentation des activités de l'IFP, très actif sur les problématiques du calcul haute performance, et impliqué dans de nombreux projets collaboratifs : IOLS, FAME2, Scilab, SCOS, ...

MICROSOFT

www.microsoft.com/france/hpc

Découvrez la nouvelle offre de calcul Microsoft disponible depuis 2006 : Windows CCS (Compute Cluster Server).

ONERA

www.onera.fr/synindex/calcul-hautes-performances.html

Moyens de calcul et informations sélectionnées par l'ONERA.

OXALYA

www.oxalya.fr

Fournisseur de technologies pour le HPC : clusters, solutions d'administration et portails d'accès distants.

SUN

www.sun.com/servers/hpc

Retrouvez les gammes de serveurs et les références de SUN.

Grands projets

MARE NOSTRUM

www.bsc.es

Premier centre de calcul européen (n° 5 mondial), le centre pour le supercalcul de Barcelone offre l'opportunité aux chercheurs de s'attaquer aux grands défis des sciences de la terre, de la vie ou de l'informatique.

FAME2

www.fame2.org

Le projet FAME2 concerne une nouvelle génération de serveurs pour le calcul intensif et le traitement de l'information. Dirigé par Bull, le projet compte 13 partenaires.

SCILAB

www.scilab.org

Scilab est un logiciel de calcul numérique open source avec un puissant environnement de développement pour les applications scientifiques et l'ingénierie. Le consortium mené par l'INRIA compte 23 membres.

SCOS

www.oscos.org

Le projet SCOS standardise et normalise les plates-formes de calcul scientifique, et fédère les initiatives de développement de codes de calcul, notamment les codes Open Source. Emmené par Oxalya, le consortium compte 22 partenaires.

100 000 milliards d'opérations
par seconde pour repousser les limites
de la recherche européenne

Bull, partenaire du CEA et du CCRT

Grâce aux serveurs NovaScale de Bull, le complexe de calcul scientifique du CEA dispose de plus de 100 Teraflops, 100 000 milliards d'opérations par seconde. Avec Bull, la recherche mise sur les standards et les environnements ouverts. L'Europe peut gagner la bataille du calcul haute performance en s'appuyant sur les experts du seul constructeur européen, fournisseur de nombreux centres de recherches et industries. Connectez-vous sur www.bull.fr

BULL

Architect of an Open World™

*Architecte d'un monde ouvert



Innover les énergies

L'IFP est un organisme public de recherche et de formation, à l'expertise internationalement reconnue, dont la mission est de développer les énergies du transport du XXI^e siècle. Il apporte aux acteurs publics et à l'industrie des solutions innovantes pour une transition maîtrisée vers les énergies et matériaux de demain, plus performants, plus économiques, plus propres et durables.

Pour remplir sa mission, l'IFP poursuit 5 objectifs stratégiques complémentaires :

- Repousser les limites du possible dans l'exploration et la production du pétrole et du gaz
- Transformer le maximum de matière première en énergie du transport
- Développer des véhicules propres et économes en carburant
- Diversifier les sources de carburants
- Capturer et stocker le CO₂ pour lutter contre l'effet de serre

Son école d'ingénieurs, partie intégrante de l'IFP, prépare les générations futures à relever ces défis.



IFP (Siège social) - 1 et 4 avenue de Bois-Préau – 92852 Rueil-Malmaison Cedex – France
Tél. : +33 1 47 52 60 00 – Fax : +33 1 47 52 70 00 – www.ifp.fr

IFP-Lyon - BP 3 - 69390 Vernaison - France
Tél. : +33 4 78 02 20 20 - Fax : +33 4 78 02 20 15 – www.ifp.fr