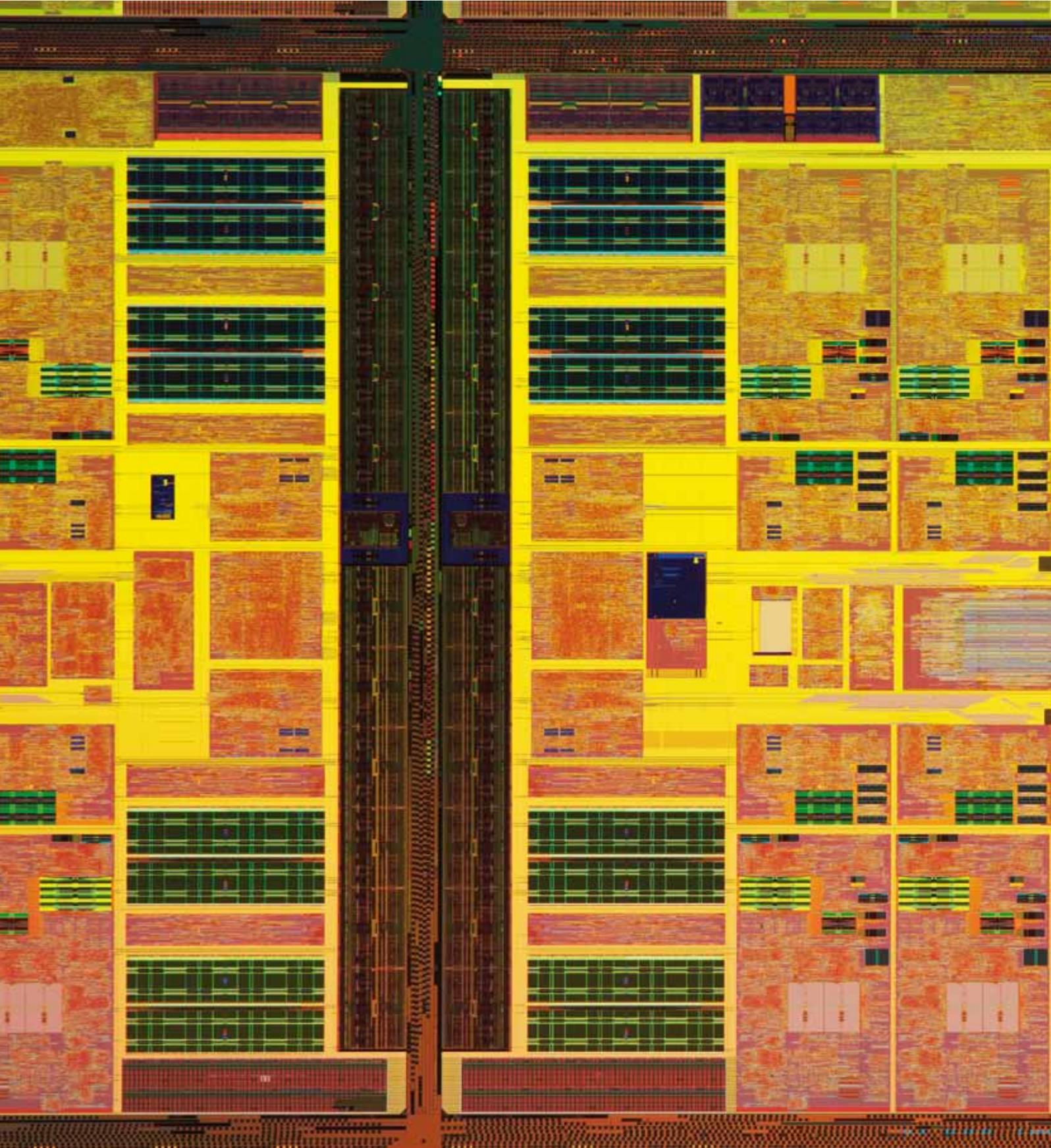


# LA SCIENCE «IN SILICO»



ISTOCK

---

Les chercheurs, y compris en sciences humaines, font de plus en plus souvent appel aux grands calculateurs qui offrent une nouvelle approche de leur discipline

---

Les machines les plus puissantes du monde ont dépassé il y a une année le seuil du petaflop, c'est-à-dire du million de milliards d'opérations par seconde

---

En plus d'un supercalculateur basé à l'EPFL, l'UNIGE dispose d'un grand nombre de grappes d'ordinateurs. Petite visite en sciences pharmaceutiques, en biologie moléculaire, en climatologie, en astronomie et en physique des particules

Dossier réalisé par Anton Vos

# LE CALCUL DE HAUTE PERFORMANCE À L'UNIGE

L'Université de Genève possède un superordinateur qu'elle partage avec l'Université et l'École polytechnique fédérale de Lausanne. Plusieurs groupes disposent également de «grappes d'ordinateurs» de petite à moyenne puissance. Les sciences humaines font également leur entrée dans ce champ dominé par les sciences naturelles et médicales

Il faut vivre avec son temps. De plus en plus, les chercheurs, quelle que soit leur spécialité, se tournent vers les supercalculateurs, ces nouveaux outils dont la puissance de calcul augmente de manière vertigineuse d'année en année et dont l'usage se répand progressivement dans les universités. Autrefois réservés à quelques chercheurs spécialisés, les superordinateurs offrent en effet une

nouvelle approche méthodologique dans de nombreux domaines, et pas seulement celui des sciences exactes. Même les sciences humaines devraient franchir le cap. C'est en tout cas l'avis de Bastien Chopard, professeur au Centre universitaire d'informatique (CUI), pour qui «modéliser» et «simuler» sont des termes qui devraient entrer dans le langage courant des académiciens, tout comme l'a fait le vocabulaire spécifique aux ordinateurs personnels et à Internet. Pour lui, les ordinateurs de «très haute performance», dont les meilleurs sont capables d'effectuer plus d'un million de milliards d'opérations par seconde, permettent d'aborder la science autrement, grâce à la restitution virtuelle des phénomènes étudiés. Une approche que l'on appelle *in silico*, en référence à la matière dans laquelle sont taillés les microprocesseurs.

## LE PLUS RAPIDE DE L'OUEST HELVÉTIQUE

L'Université de Genève, comme ses consœurs, se fraye un chemin dans cette direction. Preuve en est la création récente de Cadmos (*Center for Advanced Modeling Science*), dont Bastien Chopard préside le Comité de direction. Il s'agit d'une coopération entre les Universités de Genève et de Lausanne ainsi que l'École polytechnique fédérale de Lausanne (EPFL). Cette structure possède depuis août 2009 l'ordinateur le «plus rapide de l'Ouest» helvétique, un BlueGene/P d'IBM (lire ci-contre) dont les performances spectaculaires sont à la disposition des scientifiques des trois institutions lémaniques.

Afin d'encourager ses chercheurs à développer cette nouvelle approche, l'Université de Genève a nommé le 1er juin dernier Jean-Luc Falcone, collaborateur scientifique au

CUI, au poste d'«analyste d'application HPC» (pour *High Performance Computing*). En langage commun, cela signifie qu'il jouera le rôle de conseiller et d'intermédiaire pour toutes les personnes intéressées par l'utilisation du BlueGene ou d'une autre «grappe d'ordinateurs» (*cluster*) de l'Université.

«L'idée est d'aider les chercheurs qui désirent se lancer dans la modélisation, explique-t-il. Il s'agit de leur fournir l'accès à cette méthodologie, les clés d'entrée d'une telle machine, les familiariser avec le code informatique, évaluer les véritables besoins en calcul de leur projet, les informer des limites du système, etc.»

Jean-Luc Falcone ne manquera sans doute pas de travail tant est grand le potentiel de croissance du nombre d'utilisateurs. A l'heure actuelle, même si l'Université de Genève compte dans ses rangs des pionniers et de fréquents utilisateurs des supercalculateurs, ils ne forment qu'une petite minorité. Sur la bonne quinzaine de projets retenus pour bénéficier de temps de calcul sur le BlueGene, seuls trois viennent ainsi du bout du lac. Les autres ont été proposés par des chercheurs de l'EPFL. Il faut dire que ces derniers sont habitués à considérer des applications qui nécessitent des ressources de calcul impor-

«A l'heure actuelle, même si l'Université de Genève compte dans ses rangs des pionniers et de fréquents utilisateurs des supercalculateurs, ils ne forment qu'une petite minorité»



# NCE SE POPULARISE

tantes et à utiliser les ordinateurs puissants disponibles sur le campus.

L'Université de Genève, elle, n'est pas dotée d'une puissance de calcul exceptionnelle. En plus du BlueGene, qui est installé à l'EPFL, elle possède des clusters de taille modeste à moyenne. Leur puissance de calcul suffit cependant aux besoins des groupes de recherche qui en disposent.

## CENTRALISER LES ORDINATEURS

«La politique actuelle de l'Université en la matière est de laisser chaque groupe acheter son cluster si nécessaire et l'équiper avec les logiciels de son choix, explique Bastien Chopard. Cette stratégie n'est cependant pas toujours la plus simple, puisqu'elle exige de la place. Tout le monde ne possède pas une ou plusieurs salles pour accueillir ces machines. C'est pourquoi la réflexion se poursuit sur la meilleure manière de procéder. Peut-être qu'il vaudrait mieux centraliser toutes ces grappes d'ordinateurs à un endroit. La distance entre les utiliza-

teurs et le calculateur ne joue pas de rôle, puisque l'Université dispose de connexions rapides.»

Le cluster le plus puissant prévu à l'Université est celui de la Faculté de médecine dont le but sera de renforcer la plateforme VITAL-IT, pour les applications liées à la bio-informatique (lire en page 21). Ce cluster comptera pas moins de 1000 cœurs – autrement dit, 1000 microprocesseurs travaillant en parallèle. Le Département d'informatique, quant à lui, vient d'acquérir un calculateur de 240 cœurs, dont l'usage sera réservé aux simulations de systèmes complexes. «Sa puissance est ajustée aux besoins de la plupart de nos projets», précise Bastien Chopard. Autres traditionnels amateurs de grande puissance de calcul, le Département d'astronomie et celui de physique nucléaire et corpusculaire possèdent chacun leur calculateur propre. A cela s'ajoutent plusieurs groupes disposant dans leur laboratoire de petits clusters contenant entre 8 et 32 cœurs. ►



Le BlueGene/P, copropriété de l'UNIGE, est installé à l'Ecole polytechnique fédérale de lausanne (EPFL). Il s'agit d'un superordinateur massivement parallèle. Il possède 4096 nœuds de calculs, soit autant de microprocesseurs Power PC. Cette machine est capable d'effectuer 56 mille milliards d'opérations par seconde et offre une capacité de stockage de 16 terabytes, soit 16 mille milliards d'unités d'informations. Actuellement, une quinzaine de projets, issus des trois instituts universitaires lémaniques, se partagent l'essentiel du temps de calcul.

## Glossaire

► **Flops.** Les flops sont une unité mesurant la performance d'un ordinateur. Il s'agit de l'acronyme anglais pour *Floating Point Operations per Second*, ou opérations à virgule flottante par seconde. Ces dernières comprennent toutes les opérations (additions, multiplications, soustractions et divisions) impliquant les nombres réels. En 1960, les performances des meilleurs ordinateurs atteignaient 250 kiloflops. Aujourd'hui, le record est de 1,76 petaflop, soit 7 milliards de fois plus.

► **Cluster.** Un *cluster*, ou grappe d'ordinateurs (parfois aussi appelé *farm/ferme*), est un ensemble d'ordinateurs reliés par des connexions très rapides fonctionnant comme un seul superordinateur ayant une plus grande capacité de stockage et une plus grande puissance de calcul. C'est la configuration idéale pour réaliser des calculs en parallèle. Chaque serveur d'une telle grappe, appelé nœud, peut effectuer une partie du calcul de son côté avant de communiquer ses résultats aux autres.

► **Cœur.** Jusqu'en 2005, les microprocesseurs n'étaient composés que d'un seul cœur: ils ne pouvaient lire et effectuer qu'une seule instruction à la fois. Aujourd'hui, les fabricants conçoivent des microprocesseurs contenant deux, quatre, six voire huit cœurs, c'est-à-dire autant d'unités de calcul indépendantes, reliées entre elles et compactées dans une seule puce.

► **CPU.** Le *Central Processing Unit* (CPU) est le composant de l'ordinateur qui exécute les programmes informatiques. Il s'agit donc d'un microprocesseur.

► **GPU.** Le *Graphics Processing Unit* (GPU) est un microprocesseur présent sur les cartes graphiques au sein d'un ordinateur ou d'une console de jeux vidéo. Il se charge des opérations d'affichage et de manipulation de données graphiques. Moins cher et moins gourmand en énergie, le GPU est cependant plus limité et plus spécialisé qu'un microprocesseur conventionnel. Il entre néanmoins de plus en plus souvent dans la conception des superordinateurs les plus rapides du monde.

# Un nuage plane sur le parc informatique

Relier les ordinateurs personnels entre eux pour mettre en commun leur puissance de calcul est une stratégie très répandue et très prometteuse. L'Université de Genève participe à cet effort qui devrait aboutir à un «nuage de PC» genevois, voire suisses

L'Université de Genève (UNIGE), en collaboration avec la Haute Ecole du paysage, d'ingénierie et d'architecture de Genève (Hepia), a pour projet de mettre en réseau la puissance de calcul de tous leurs ordinateurs personnels, ceux qui trônent dans les bureaux administratifs, les laboratoires et les salles informatiques des étudiants. Au total, plusieurs milliers de machines sont concernées.

L'idée n'est pas nouvelle. Ces appareils ne fonctionnant que lors des heures de bureau, le reste du temps, la nuit ainsi que les samedi et dimanche, leur capacité de calcul dort alors qu'elle pourrait servir à l'avancement de la science. C'est pour exploiter ce potentiel que l'EZ-GRID, une infrastructure de calcul distribué, a été développée par le Centre universitaire d'informatique (CUI), la Division informatique de l'UNIGE et l'Hepia.

## CENT MILLE ORDINATEURS

Le principe est le même que le *Worldwide LHC Computing GRID* (WLCG) élaboré par le CERN afin de stocker, traiter et analyser les montagnes de données fournies par les détecteurs de l'accélérateur de particules LHC. Le WLCG exploite en continu les performances de pas moins de plus de 100 000 microprocesseurs répartis sur 130 sites, eux-mêmes distribués dans 34 pays (lire en page 26).

Cette stratégie est particulièrement à la mode puisqu'une compagnie internationale très active dans la vente par correspondance de livres et de musique sur Internet a également décidé de mettre à disposition les ressources de ses énormes serveurs de commandes qu'elle possède à travers le monde. Ces derniers sont en effet inutilisés à des moments précis (la nuit, certaines heures de la journée lorsque le nombre de commandes est faible, etc.). Le service, dans ce cas, n'est en revanche pas gratuit.

## MATURITÉ SUFFISANTE

L'EZ-GRID, lui, est bien plus modeste. Il permet également d'utiliser les ressources de toutes les machines qui font partie du réseau dès qu'elles sont en veille. Le potentiel est tout de même assez important, puisque plusieurs milliers de PC sont répartis entre l'UNIGE et l'Hepia. Ce qui correspond à un supercalculateur comptant des milliers de microprocesseurs travaillant en parallèle.

«Le système a atteint un stade de maturité suffisant et nous avons réalisé des projets tests sur quelques centaines de machines partagées entre l'Université et la Haute Ecole», précise Bastien Chopard, professeur au CUI. Si les résultats sont satisfaisants, il sera souhaitable que l'EZ-GRID se déploie sur le parc informatique de l'alma mater. Par ailleurs, toutes ces machines sont aussi reliées à un

projet de réseau encore plus grand qui couvrirait toute la Suisse.» Le Swiss Multiscience Computing GRID (SMSCG) vise en effet à connecter des ordinateurs d'un grand nombre d'universités et hautes écoles du pays ([www.smsgg.ch](http://www.smsgg.ch)).

## QUELQUES LIMITATIONS

Une telle infrastructure, bien que très puissante, bute néanmoins sur quelques limitations. «Dans le cadre de l'EZ-GRID, chaque ordinateur personnel est capable de faire des morceaux de calcul, mais il ne peut pas communiquer facilement avec les autres», précise Bastien Chopard. Ce genre de réseau est donc destiné à réaliser des calculs qui peuvent se séparer en un grand nombre de tâches indépendantes les unes des autres. Il ne peut dès lors pas se substituer à un superordinateur parallèle classique dans lequel les microprocesseurs échangent sans cesse des informations entre eux.»

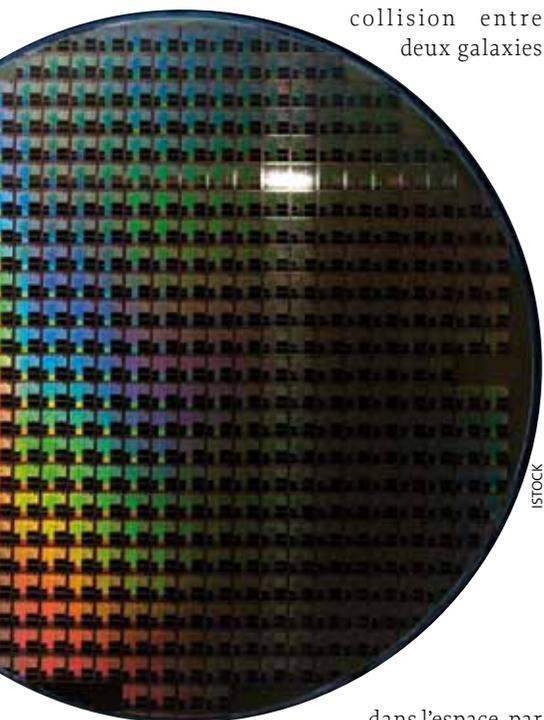
En d'autres termes, s'il s'agit de simuler l'effet d'un million de molécules différentes sur un même récepteur cellulaire, il serait judicieux de passer par l'EZ-GRID, chaque calcul étant indépendant des autres et pouvant être réalisé par une machine différente. En revanche, si l'on veut modéliser un écoulement sanguin dans une artère, où chaque élément de calcul a sans cesse besoin du résultat de plusieurs autres éléments de calcul, il est préférable de faire appel à un superordinateur ou une grappe d'ordinateurs

(cluster) avec un réseau d'interconnexions très rapides.

Il y a d'autres difficultés liées au déploiement d'une infrastructure de calcul distribué. Si, a priori, l'exploitation à des fins de calcul scientifique des parcs informatiques destinés aux étudiants ne pose pas de problèmes (puisque les appareils n'appartiennent à personne), la gestion administrative de ces postes et des salles où ils sont installés est souvent peu compatible avec les besoins de disponibilité requise pour EZ-GRID. En effet, l'électricité est parfois coupée dans certaines salles durant le week-end ou la nuit, ce qui empêche le «réveil à distance» des PC arrêtés. L'utilisation des appareils du personnel de l'Université serait dans ce sens plus aisée.

En revanche, il n'est pas sûr que tout le monde soit d'accord de confier facilement son ordinateur aux techniciens de l'EZ-GRID. Certaines personnes pourraient craindre, en effet, que cette infrastructure n'entraîne des dysfonctionnements ou un ralentissement de leur machine, sans même parler du sentiment d'atteinte à leur sphère privée. Et ce, même si le projet contient, selon Bastien Chopard, toutes les garanties en matière de confidentialité et de sécurité. ■

Toutes ces (plus ou moins super-) machines offrent aux chercheurs une nouvelle manière de faire de la science. Elles permettent de simuler des phénomènes très divers et de réaliser des prédictions sans devoir passer par la réalité parfois trop coûteuse, parfois inaccessible. Le film accéléré d'une collision entre deux galaxies



ISTOCK

en plus réduit et avec une précision de plus en plus grande, de simuler des phénomènes climatiques et météorologiques, l'écoulement du sang dans un vaisseau sanguin, le comportement d'une poche magmatique sous un volcan, l'effet probable d'une molécule sur une cible thérapeutique, les mécanismes d'une avalanche ou de l'explosion d'une arme à destruction massive.

dans l'espace, par exemple, n'est pas un événement que l'on peut observer à l'envi. Un supercalculateur permet de le réaliser (bien qu'il faille réduire le nombre d'étoiles contenues dans les galaxies en question). Il permet également de proposer des scénarios probables pour expliquer les images figées que les astronomes observent aujourd'hui avec leurs télescopes.

De la même manière, grâce aux superordinateurs, il est possible, en un temps de plus

Cela dit, les grandes puissances de calcul n'intéressent pas seulement les chercheurs en sciences naturelles et médicales (lire les pages 20 à 27). La finance est depuis longtemps une grande utilisatrice des supercalculateurs. Au début des années 2000, durant dix-huit mois, la machine la puissante de Suisse appartenait d'ailleurs à Credit Suisse.

Plus récent est l'intérêt des sciences humaines pour le calcul de haute performance. Luka Nerima, chargé d'enseignement au Laboratoire d'analyse et de traitement du langage (Faculté des lettres), par exemple, a participé l'année dernière à un projet international, PASSAGE 2, visant à évaluer un analyseur syntaxique du français. Pour l'analyse d'un corpus comptant plus de 200 millions de caractères (articles de presse et tirés de Wikipédia, l'encyclopédie en ligne ainsi que des textes littéraires), le chercheur a utilisé une dizaine d'ordinateurs fonctionnant en parallèle. Le travail a duré un jour au lieu de dix s'il n'avait utilisé qu'une seule machine.

«Jusqu'à récemment, ce genre d'ordinateurs de haute performance était trop cher pour les sciences humaines, estime Luka Nerima. Maintenant que leur usage se démocratise, nous les utilisons plus souvent. Et nous sommes intéressés à aller plus loin encore dans cette direction. Si nous avions accès à un Grid plus important, nous pourrions réduire encore le temps de calcul de nos analyses syntaxiques, ce qui est essentiel pour pouvoir améliorer nos outils informatiques.» ■

## Les cartes graphiques: rapides et bon marché

Une des stratégies suivies par les constructeurs de supercalculateurs pour augmenter la puissance de leurs appareils consiste à exploiter les performances exceptionnelles des cartes graphiques. Développées à l'origine pour les jeux vidéo, ces cartes peuvent compter plusieurs centaines de microprocesseurs très puissants (GPU, pour Graphics Processing Unit). Une seule d'entre elles peut aujourd'hui atteindre une vitesse de calcul d'un teraflop (mille milliards d'opérations) par seconde.

La différence entre ces petites cartes et les énormes supercalculateurs massivement parallèles, c'est que les premières sont spécialisées dans un type de calcul simplifié et ne peuvent de loin pas réaliser ce dont un microprocesseur traditionnel et moins rapide est capable. Il faut dire que leur usage premier est le traitement d'images graphiques, de vidéos et de rendus en trois dimensions.

Certains chercheurs et fabricants de superordinateurs tentent d'exploiter ces propriétés à des fins de recherche scientifiques. C'est notamment le cas de la firme chinoise Dawning, qui a construit le superordinateur Nebulae, qui pointe à la deuxième place du dernier classement des ordinateurs les plus rapides du monde publié en juin 2010 ([www.top500.org](http://www.top500.org)). Cette machine possède probablement (les informations ne viennent pas directement du fabricant, pas très bavard, mais de spéculations de spécialistes) un tiers de GPU pour deux tiers de microprocesseurs conventionnels.

# LE JAGUAR AMÉRICAIN TALONNÉ PAR LA NÉBULEUSE CHINOISE

Deux fois par année, une liste des 500 superordinateurs les plus puissants du monde est publiée. La dernière, datant du mois de juin, a placé pour la première fois une machine chinoise à la deuxième place

La définition d'un superordinateur varie avec le temps, et très rapidement. Le record mondial de vitesse de calcul est ainsi multiplié par un facteur d'environ 10 tous les cinq ans. En juin 1993, lorsque la Conférence internationale de calcul scientifique à Mannheim en Allemagne a publié sa première liste des 500 meilleures machines du monde, la première place était occupée par un appareil capable d'effectuer 60 milliards d'opérations (giga-flops) par seconde. Il appartenait à une institution de recherche en Californie spécialisée dans l'armement nucléaire (*Lawrence Livermore National Laboratory*). Cette performance est aujourd'hui à la portée d'un très bon ordinateur de bureau.

Depuis une année, le record est détenu par un superordinateur Cray, le Jaguar, installé au *Oak Ridge National Laboratory* au Tennessee. Ce caïd, qui occupe 400 m<sup>2</sup> au sol, peut réaliser 1,76 million de milliards d'opérations (petaflops) par seconde. Ses domaines de prédilection sont les simulations du climat, de matériaux supraconducteurs, de réactions chimiques, d'astrophysique ou encore de processus de combustion.

## ASCENSION INEXORABLE

Le Jaguar est talonné, c'est une première, par un concurrent chinois, le Nebulae, installé à Shenzhen, près de Hong Kong. Cette machine confirme l'ascension inexorable de la Chine dans ce domaine technologique. Il y a un an, elle possédait une machine occupant la cinquième position et il est probable que la première place lui revienne dans un proche avenir.

Le 500<sup>e</sup> mondial, selon le dernier palmarès publié en juin 2010 (le prochain doit paraître en novembre 2010), dépasse les 24 mille milliards d'opérations (teraflops) par seconde. Le BlueGene/P installé à l'Ecole polytechnique



Le superordinateur Jaguar installé au Oak Ridge National Laboratory au Tennessee, Etats-Unis.



Le superordinateur Monte Rosa est le plus puissant en Suisse. Il est installé au Centre de Calcul scientifique à Monna au Tessin. Vingt-septième dans le classement mondial des supercalculateurs du mois de juin 2010, cette machine est capable de 168 mille milliards d'opérations par seconde.

fédérale de Lausanne et dont l'Université de Genève est copropriétaire, lui, se porte honorablement, puisqu'il dépasse les 47 teraflops par seconde (117<sup>e</sup> mondial).

La Suisse possède cependant quelques machines plus rapides. La plus puissante est un Cray, baptisé Monte Rosa et installé au Centre suisse de calcul scientifique à Monna au Tessin. Pointant à la 27<sup>e</sup> place mondiale, il atteint les 168 teraflops par seconde (lire ci-contre).

### COURSE EN AVANT

La course en avant des supercalculateurs bute aujourd'hui sur un certain nombre d'obstacles techniques. Le principal est la surchauffe. Plus les microprocesseurs réduisent leurs composants et augmentent leur cadence de travail, plus ils dégagent de chaleur. Le défi consiste à refroidir suffisamment la machine pour qu'elle puisse continuer à fonctionner. Le Jaguar est doté d'un système très sophistiqué basé sur le changement de phase d'un liquide de refroidissement (le R134A). Le BlueGene lémanique, lui, est refroidi à l'air et à l'eau.

Par ailleurs, la miniaturisation extrême des puces provoque la fuite d'électron d'un circuit vers un autre. La seule façon de gagner encore en performance, avec la technologie actuelle, c'est d'augmenter le nombre de mi-

croprocesseurs contenus dans un supercalculateur. Mais cela pose ensuite le problème de l'interconnexion et du développement de logiciels capables de gérer une telle complexité.

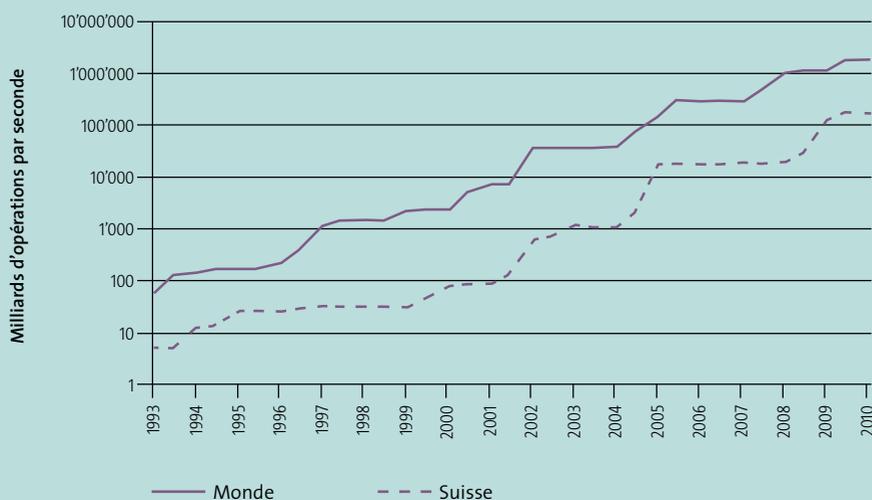
### SCIENCE-FICTION

Pour ces diverses raisons, d'aucuns pensent que l'évolution des superordinateurs est dans sa dernière ligne droite. Mais c'est compter sans l'inventivité des professionnels de la branche. Les ingénieurs des grandes entreprises ou, plus souvent encore, de petites start-up, ont jusqu'à maintenant à chaque fois trouvé des moyens de contourner les obstacles. Résultat: le rythme de progression de la rapidité des machines se maintient sans faillir depuis bientôt vingt ans.

D'ailleurs, les constructeurs affirment déjà être dans la course pour l'exaflop (un milliard de milliards d'opérations) par seconde. L'objectif devrait, selon eux, être atteint dans dix ans (sans qu'ils sachent exactement comment ils y parviendront). Certains parlent déjà du zettaflop (mille fois plus rapide). ■

[www.top500.org](http://www.top500.org)  
[www.cscs.ch](http://www.cscs.ch)  
[www.cadmos.org](http://www.cadmos.org)  
[www.nccs.gov/jaguar](http://www.nccs.gov/jaguar)

## Performance des meilleurs superordinateurs



## La Chine prend la première place

Au moment de mettre sous presse, la dernière liste des 500 supercalculateurs les plus puissants du monde, prévue pour la mi-novembre, n'était pas encore publiée. Plusieurs annonces dans la presse laissent néanmoins entendre que la Chine aurait raflé la première place avec le Tianhe-1A dont la puissance de calcul atteindrait les 2,507 petaflops, soit deux fois plus rapide que le Jaguar. Pour toutes informations supplémentaires sur cette compétition frénétique: [www.top500.org](http://www.top500.org).

# LA PHARMA VIRTUELLE: UNE APP DE LA RÉALITÉ

La simulation de médicaments par ordinateur est devenue chose courante. Mais l'amélioration de la puissance de calcul va de pair avec une complexification des simulations créées par les chercheurs. Pour s'approcher chaque fois un peu plus de la réalité

*«C'est une histoire sans fin: plus nous disposons d'ordinateurs puissants, plus nous complexifions les problèmes que nous voulons résoudre. Et nous avons de la marge puisque les modèles que nous créons, bien qu'ils soient suffisants pour obtenir des résultats utilisables, ne représentent encore qu'une très grande approximation de la réalité.»* Pierre-Alain Carrupt, professeur à l'École de pharmacie Genève-Lausanne et président de la Section de pharmacie (Faculté des sciences), sait de quoi il parle: cela fait plus de trente ans qu'il utilise des ordinateurs parmi les plus rapides que l'on puisse trouver pour simuler l'action de médicaments, ou plus précisément de composés bioactifs sur des cibles thérapeutiques.

La simulation par ordinateur est bien implantée en sciences pharmaceutiques. Les chercheurs ne se contentent plus, en effet, d'identifier des substances qui pourraient posséder un pouvoir curateur. Ils les construisent sur ordinateur, prédisent leur efficacité à se lier à leur cible, tentent de simuler leur effet chimique, avant même de les synthétiser ou les tester en conditions réelles. C'est ce qu'on appelle le *drug design*. Une technique qui entre aujourd'hui, à un degré plus ou moins élevé, dans la conception de n'importe quel nouveau médicament.

La dorzolamide, un médicament indiqué pour le traitement contre une pression intraoculaire élevée, est le premier traitement né sur ordinateur à être commercialisé en 1995. Plus connu est le cas de l'oseltamivir (vendu par Hoffman-la-Roche sous le nom de Tamiflu), un médicament contre la grippe et dont l'action bloque une enzyme du virus

impliquée dans sa réplication. Dans ces deux cas, la cible thérapeutique était connue et les chimistes ont façonné par ordinateur des composés capables de la neutraliser.

Car la simulation ne peut pas tout. Avant de faire chauffer les processeurs, le chercheur doit d'abord avoir une idée assez précise de la problématique. Dans le cas d'une maladie, il est indispensable d'avoir identifié une cible d'intérêt thérapeutique, qu'il s'agisse d'une enzyme, d'une autre protéine ou encore d'un acide nucléique. Il doit ensuite déterminer sa structure moléculaire et spatiale afin de la recréer sur un

écran. Le site actif de cette cible forme souvent une cavité et c'est pourquoi les chercheurs la surnomment la «poche». Le composé capable de s'y fixer, lui, est nommé le ligand.

*«Il est préférable de connaître déjà un ligand à l'avance, précise Pierre-Alain Carrupt. Cela permet de donner une idée de la forme que pourrait prendre notre futur médicament. La dopamine, par exemple, est un neurotransmetteur impliqué, entre autres, dans la maladie de Parkinson. Cette molécule et son récepteur associé qui se trouve à la surface des neurones s'imbriquent selon une géométrie très précise. Si on la connaît, on peut commencer à chercher, par simulation, une substance qui remplacerait la dopamine. L'ordinateur permet ainsi de tester très rapidement de nouvelles substances. Sans lui, ce travail prendrait des années.»*

## SITUATION STATIQUE

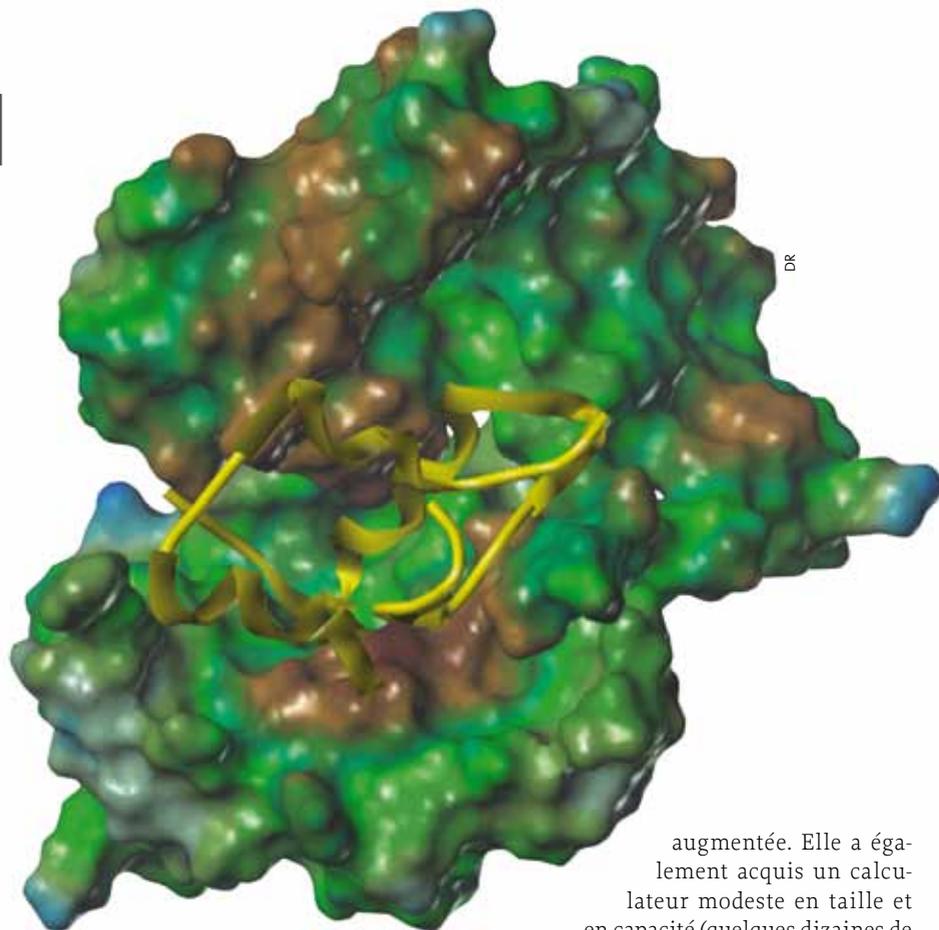
Le degré de difficulté le plus élémentaire, dans la simulation pharmaceutique, est la situation statique qui ne fait appel qu'à la mécanique moléculaire. Dans ce cas, le chercheur tente simplement d'ajuster le ligand pour qu'il se fixe dans la poche de la façon souhaitée. Cette phase ne demande pas énormément de temps de calcul. Les ordinateurs personnels modernes sont suffisants.

Le problème se complexifie lorsqu'on ajoute un solvant, principalement de l'eau, mais parfois d'autres substances présentes dans le milieu biologique telles que les ions inorganiques. On entre alors dans une situation dynamique. Quand le ligand entre dans la poche, par exemple, il faut bien que l'eau puisse sortir de cet espace et se redistribuer autour du complexe ligand-cible nouvellement formé.

«Les choses se corrent véritablement lorsqu'on veut tenir compte des réactions chimiques entre le composé actif et sa cible»

# ROXIMATION

Image tirée d'une simulation moléculaire d'un facteur de croissance (le IGF-1 en jaune) qui est associé à des maladies lorsque son taux sanguin n'est pas optimal. Il est lié à un complexe dont la surface moléculaire est colorée en fonction de ses propriétés chimiques.



Les simulations se déroulent donc durant un certain laps de temps. Les supercalculateurs représentent dès lors une aide précieuse, surtout lorsque les chercheurs, dans un souci de coller au mieux à la réalité, augmentent la taille du système (nombre de molécules d'eau, par exemple) et le temps de simulation.

Les choses se corsent véritablement lorsqu'on veut tenir compte des réactions chimiques entre le composé actif et sa cible. En effet, certaines liaisons peuvent se casser ou se modifier quand le ligand se fixe dans la poche. Les chercheurs ne peuvent alors plus se contenter d'un modèle simple. Ils doivent faire intervenir les notions de chimie quantique, la discipline qui décrit les propriétés et les processus chimiques des molécules. Le degré de précision de la modélisation devient si important qu'il intègre les atomes entrant en jeu dans la réaction – moyennant de nombreuses approximations. Même si le nombre d'atomes pris en compte ne représente qu'un faible pourcentage de la quantité totale d'atomes impliqués dans la réalité, de telles simulations commencent à devenir très gourmandes en calcul. A ce stade, le supercalculateur devient très utile.

## UN PAS PRESQUE INSIGNIFIANT

Le top, aujourd'hui, est de modéliser une telle réaction chimique, à l'aide de la mécanique quantique, et en y ajoutant un solvant. Si l'on rallonge la durée de la simulation et que l'on augmente la taille du système, le supercalculateur devient alors indispensable.

«Cela dit, même avec tous les progrès accomplis en matière de puissance de calcul des ordinateurs

et de techniques de modélisation, nous n'avons réalisé qu'un pas presque insignifiant vers la description parfaite de la réalité, souligne Antoine Daina, maître assistant à la Section des sciences pharmaceutiques. Les simulations, par exemple, ne tiennent compte que d'un seul ligand et d'un seul récepteur alors qu'un cachet d'aspirine (500 mg) contient environ  $2 \times 10^{20}$  molécules de principe actif.»

A cela s'ajoute le fait que, du point de vue physique, seules quelques molécules simples comme celle d'hydrogène n'ont plus de secrets pour les chercheurs. Toutes les autres ont encore une part plus ou moins importante de mystère. Sans même parler du fait que les méthodes de simulation de la chimie quantique ont été développées pour copier les réactions se déroulant en phase gazeuse. Elles sont encore imprécises pour la modélisation des réactions en phase liquide et non pertinentes si l'on considère un milieu biologique complexe.

«Toutes les simulations par ordinateurs, quel que soit le sujet étudié, comportent des approximations, admet Pierre-Alain Carrupt. Et de toute façon, n'importe quel résultat obtenu *in silico* est une prédiction qui doit être vérifiée expérimentalement. On ne peut plus se passer des ordinateurs dans la fabrication des médicaments, mais on ne se passera jamais non plus des essais cliniques.»

L'équipe de Pierre-Alain Carrupt dispose aujourd'hui d'un petit réseau de stations de calcul dont la puissance est en train d'être

augmentée. Elle a également acquis un calculateur modeste en taille et en capacité (quelques dizaines de fois plus puissant qu'un PC de bureau). Ces appareils suffisent néanmoins à la plupart des besoins des chercheurs en matière de calcul et de visualisation des résultats en trois dimensions. Pour l'instant.

## TEMPS PARTAGÉ

«Nous allons bientôt collaborer avec une équipe de l'Ecole polytechnique fédérale de Lausanne (EPFL) sur des projets de mécanique quantique dynamique, souligne Pierre-Alain Carrupt. Nous utiliserons donc vraisemblablement dans un avenir proche le BlueGene (lire en page 15) installé à Lausanne.»

Toutefois, selon le professeur, ces énormes machines comportent quelques désavantages. La plupart du temps, ce sont des ordinateurs partagés, c'est-à-dire que beaucoup de chercheurs les utilisent en même temps. Ce qui divise d'autant leur puissance de calcul.

«Il y a trente ans, lorsqu'il fallait encore manipuler des cartes perforées, nous étions déjà quelques dizaines de chercheurs en compétition pour le même calculateur, l'un des premiers de l'EPFL, se souvient-il. Je devais louer la machine et travailler environ tous les trois week-ends pour avoir plus de temps de calcul et, ainsi, avancer plus rapidement. Aujourd'hui, la compétition s'est exacerbée car de nombreuses disciplines nouvelles font appel à des techniques complexes de simulation. C'est pourquoi, nous avons choisi d'acheter une infrastructure modeste et de nous tourner vers le supercalculateur seulement en cas de réel besoin.» ■



# LA GÉNOMIQUE CROULE SOUS LES DONNÉES

Les biologistes produisent plus de données qu'il n'y a de serveurs pour les stocker ou les traiter. Les fermes d'ordinateurs, dotées d'une grande puissance de calcul et d'une vaste mémoire, se multiplient pour faire face. Entretien avec Antoine Geissbühler, directeur du Département de radiologie et informatique médicale et président de la commission informatique de la Faculté de médecine

**La Faculté de médecine se dote d'une ferme d'ordinateurs parmi les plus puissantes de l'Université de Genève. Pourquoi avez-vous besoin d'un tel supercalculateur?**

ANTOINE GEISSBÜHLER: Cela fait quelque temps déjà que les chercheurs en médecine fondamentale et, plus généralement, en sciences de la vie se sont rendu compte qu'ils pouvaient aborder leur discipline d'une nouvelle ma-

nière. Après les études classiques *in vivo* (sur un organisme vivant) puis *in vitro* (en éprouvette dans les laboratoires), voici que l'on assiste à l'émergence des études *in silico* (sur ordinateur). De nombreuses activités ont déjà été développées dans cette troisième voie, qu'il s'agisse de la modélisation de processus biologiques, d'actions de molécules sur leur cible thérapeutique, etc. Toutes ces recherches

demandent de plus en plus de ressources en calcul. Mais il est un aspect qui domine actuellement les sciences de la vie, c'est la profusion de données.

**Comment cela?**

Les laboratoires de génomiques disposent de séquenceurs dernier cri, dits d'ultra haute performance. Ces machines déchiffrent des

codes génétiques à des vitesses très élevées. Elles sont capables de produire des milliards de données par seconde. On se retrouve donc avec une avalanche d'informations qu'il nous faut stocker, traiter et analyser au fur et à mesure. Du coup, les chercheurs en sciences de la vie sont confrontés à une explosion inédite de leurs besoins en puissance informatique. Il faut bien l'admettre: notre situation est critique. Sans une capacité de calcul et de mémoire exceptionnelle, nous ne pouvons pas utiliser correctement nos séquenceurs.

#### **Est-ce que cela signifie que vous ne pouvez pas, actuellement, traiter toute l'information produite par les séquenceurs?**

Dans le monde, l'accroissement de la production de données biologiques est plus rapide que la loi de Moore qui prédit avec beaucoup d'exactitude l'augmentation de la puissance des microprocesseurs. Cela signifie que si cela continue, nous allons bientôt créer plus d'informations que nous ne pouvons en traiter. Notre problème ressemble sous certains aspects à celui rencontré par le LHC, le dernier accélérateur du CERN: nous recherchons des aiguilles dans une meule de foin gigantesque. En biologie, cependant, il nous faut, en plus, garder précieusement toutes les données récoltées. En effet, nos connaissances en génomique (l'étude des gènes), en protéomique (l'étude des protéines) ou encore en transcriptomique (l'étude des ARN messagers) s'accroissent régulièrement. Certaines informations ne pourront être exploitées que plus tard, quand nous serons à même de les comprendre.

#### **La recherche «in silico», dans le domaine des sciences de la vie, est-elle dépassée par sa propre puissance?**

Disons plutôt que, d'un point de vue philosophique, l'on peut se poser des questions sur la capacité du cerveau humain à continuer à avancer intelligemment dans cet océan de données inexploitées. Le scientifique, qui est un être humain comme un autre, a tendance, face à cette débauche d'informations, à chercher ses clés sous le réverbère. Dans l'abon-

«Nous recherchons  
des aiguilles  
dans une meule  
de foin  
gigantesque»

dance, il utilisera les données qui lui semblent utiles et négligera les autres. Ce n'est pas fréquent, en science, que de disposer d'une telle manne.

#### **C'est pourquoi il est intéressant de laisser travailler l'ordinateur à sa place...**

En effet. Les superordinateurs sont capables de trouver des associations que l'on ne suspecterait pas. Mais le problème demeure. Il faut davantage de puissance de calcul et de stockage. D'où la nécessité de construire une ferme d'ordinateurs à la Faculté de médecine qui puisse répondre à ce défi, en tout cas localement.

#### **De quelle puissance allez-vous disposer avec ce nouvel outil?**

La ferme d'ordinateurs est en cours de construction. Elle devrait être terminée d'ici à la fin de l'année. Au total, nous disposerons de mille cœurs (ou processeurs) de calcul et

d'une capacité de stockage de plus de 200 terabytes, ou 200 mille milliards de bytes (la plus petite unité adressable d'un ordinateur). Des performances de rêve. Les plus importantes de l'Université de Genève à l'heure actuelle. Cela dit, nous avons rencontré une contrainte inattendue. Ce n'est ni le manque de place ni l'alimentation électrique qui nous a freinés, même si ces appareils sont gourmands en énergie (jusqu'à 100 kilowatts). En réalité, c'est le dégagement de chaleur qui nous a limités. La ferme d'ordinateurs chauffe tellement que le refroidissement du bâtiment (le Centre médical universitaire) n'arrive pas à suivre. En plus, ces ordinateurs, qui ressemblent à de grosses armoires noires où une multitude de lumières clignotent, sont lourds. Nous avons beau nous diriger vers une société dématérialisée, l'informatique continue de peser. Nous avons donc dû tenir compte aussi de la résistance des planchers du bâtiment. Tout cela pour dire que l'augmentation de notre capacité informatique a finalement nécessité une réflexion sur un partenariat plus large.

#### **A qui vous êtes-vous adressé?**

Tout d'abord, la commission informatique a décidé, il y a deux ans déjà, de mieux coordonner ses efforts en la matière et de mutualiser les ressources disponibles localement. La ferme de calcul, une fois terminée, jouera le rôle de navire amiral. Elle découle en fait d'un partenariat avec Vital-IT, le centre de supercalcul basé à l'Université de Lausanne. Cette initiative commune à plusieurs institutions lémaniques, dont l'Université de Genève et l'Institut suisse de bioinformatique (lire ci-contre), offre des capacités de calcul et de stockage importantes pour les projets dans les sciences de la vie (notamment une vitesse de calcul de pointe atteignant 2 teraflops, ou mille milliards d'opérations par seconde). Notre ferme d'ordinateurs sera connectée à ce centre et jouera le rôle de poste avancé de Vital-IT à Genève. Nous pourrions donc exploiter notre puissance de calcul locale et, si nécessaire, faire appel aux ressources plus importantes installées à Lausanne. Et surtout à leur savoir-faire en la matière. ■

# Des puces dédiées aux sciences de la vie

Le Centre informatique de haute performance Vital-IT est une collaboration entre les institutions universitaires de l'Arc lémanique. Il met des ressources de calcul à la disposition des chercheurs actifs en sciences de la vie

Des milliers de milliards de données. C'est ce que produisent chaque semaine les différents séquenceurs à haut débit fonctionnant actuellement dans les Universités de Genève et de Lausanne. Ces machines dernier cri décryptent des chaînes d'ADN à très haute vitesse. En attendant

Pour les aider, le centre informatique de haute performance Vital-IT met à la disposition des scientifiques, depuis 2004, des ressources de calcul, de stockage de données et, surtout, une équipe d'experts capable de conseiller, aider et guider les chercheurs dans leurs démarches. Basé sur le site de

«La particularité de notre centre est qu'il est dédié exclusivement à la recherche dans les sciences de la vie, précise Ioannis Xenarios, responsable de Vital-IT. Nous proposons, pour l'instant, environ 1000 cœurs de calculs (ou processeurs) et pas moins d'un demi-million de milliards de bytes de volume de stockage sur nos serveurs. Un byte est la plus petite unité adressable d'un ordinateur et est composée de huit «1» et «0». Mais ces ressources augmentent d'année en année.»

## NOMBREUX DOMAINES

Les domaines des sciences de la vie faisant appel à l'informatique de haute performance sont nombreux. Les compétences de Vital-IT sont exploitées par des équipes actives dans la génétique, la génomique, la métagénomique, la génétique des populations, la protéomique, la phylogénie, la biologie structurale ou encore l'évolution (la liste n'est pas exhaustive). Certains réalisent des simulations de processus biologiques ou étudient le comportement de colonies de fourmis, d'autres analysent le contenu génétique du chromosome 21 ou le génome du poulet, d'autres encore traitent des centaines de milliers d'images de protéines fluorescentes évoluant dans des cellules.

«Vital-IT possède 500 bases de données différentes provenant de chercheurs du monde entier, poursuit Ioannis Xenarios. Il peut s'agir du génome de n'importe quel organisme

vivant, de l'expression des gènes ou de la présence de protéines dans un tissu tumoral humain. Les possibilités sont infinies.»

## LE DÉFI DU STOCKAGE

Le grand défi actuel de la biologie est le stockage. C'est pourquoi Vital-IT a adopté une approche développée notamment par les physiciens et ingénieurs du CERN qui ont à traiter depuis longtemps des avalanches de données fournies par les détecteurs de particules: le HSM, ou Hierarchical Storage Management.

«Il s'agit simplement de ranger les informations en fonction de leur utilité immédiate, explique Ioannis Xenarios. Les informations les plus importantes aux yeux du chercheur sont stockées sur les disques durs, là où elles sont le plus rapidement disponibles. Celles dont on ne sait pas quoi faire dans l'immédiat mais qui pourraient se révéler intéressantes à l'avenir, sont enregistrées sur des bandes magnétiques.»

L'autre point fort de Vital-IT est l'expertise. Celle des membres de l'équipe, mais aussi celle d'un vaste réseau de compétences en Suisse et à l'étranger. «Nous disposons de près de 1000 algorithmes développés ces dernières années pour réaliser des simulations ou du traitement de données dans les différentes disciplines des sciences de la vie, souligne Ioannis Xenarios. Et nous suivons l'évolution de chacun de ces logiciels.» ■

[www.vital-it.ch](http://www.vital-it.ch)  
[www.isb-sib.ch](http://www.isb-sib.ch)

L'«*Arabidopsis thaliana*», ou arabettes des dames, est une plante appréciée des biologistes. L'étude de son génome, de l'expression de ses gènes, de l'activité de son ARN ou de ses protéines exige parfois des ressources informatiques supérieures.

la génération suivante d'appareil déjà en phase de conception. En d'autres termes, les sciences de la vie subissent actuellement une mutation technologique impressionnante et elles produisent plus d'informations que les chercheurs ne peuvent interpréter.

Dorigny dans le Centre intégratif de génomique et géré par l'Institut suisse de bioinformatique, Vital-IT est une collaboration entre les Universités de Genève et de Lausanne, l'Ecole polytechnique fédérale de Lausanne et le Ludwig Center for Cancer Research.



# LE CLIMAT MIS EN BOÎTE

Les sciences climatiques sont exigeantes en puissance de calcul. Les résultats des simulations, parfois décriés sur la scène politique, sont pourtant difficiles à remettre en cause. Explications avec Stéphane Goyette, chercheur à l'Institut des sciences de l'environnement

## Campus: La climatologie utilise-t-elle les supercalculateurs?

STÉPHANE GOYETTE: Oui. Les sciences climatiques sont très gourmandes en puissance de calcul. Les modèles de simulation à l'échelle de la planète qui intègrent de nombreux paramètres et qui couvrent des périodes de plusieurs siècles demandent des supercalculateurs très rapides et une équipe sachant les piloter.

## Réalisez-vous ce genre de simulation à Genève?

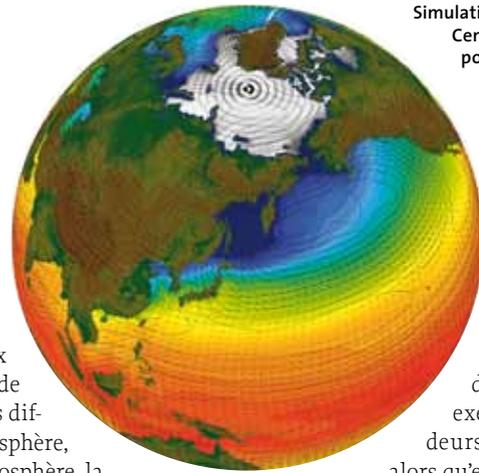
Non, nous n'effectuons pas de simulations à l'échelle globale. D'ailleurs, il n'existe pas beaucoup d'équipes qui ont les ressources techniques et humaines nécessaires pour le faire. Il s'agit le plus souvent de projets internationaux. Un tel projet pourrait néanmoins voir le jour à Genève par le biais du projet ICES (*International Center for Earth Simulations*), une initiative du fondateur et ancien directeur de la firme Silicon Graphics, qui vise une puissance de calcul inédite dans ce domaine. Cela dit, comme tous les autres chercheurs actifs en climatologie, nous avons accès aux résultats produits par ces modèles globaux. Il nous arrive donc de travailler sur une partie des données extraites des grandes simulations pour répondre à des questions spécifiques. Nous utilisons également des modèles limités à des zones géographiques plus petites (modèles régionaux du climat) et sur des laps de temps plus courts, mais qui sont basés sur les mêmes lois de la physique que les modèles globaux. Nous disposons d'une puissance informatique qui nous permet par exemple de faire tourner un modèle climatique adapté aux conditions de la Suisse avec une résolution d'un kilomètre. Plus précisément, mon travail consiste à perfectionner les modèles existants.

## Quel degré de précision atteignent les meilleurs modèles climatiques?

Depuis moins de dix ans, avec l'augmentation de la puissance des supercalculateurs, la complexité des modèles climatiques globaux et régionaux a explosé. On intègre de plus en plus de modules différents en plus de l'atmosphère, à savoir les océans, la biosphère, la cryosphère, etc. On les couple ensuite les uns avec les autres, ce qui demande un travail informatique gigantesque. On intègre également de plus en plus de processus physiques et chimiques qui se déroulent dans l'ensemble de ces modules et qui jouent un rôle dans l'évolution du climat. Curieusement, durant tout ce temps, la résolution des modèles globaux est restée sensiblement la même, aux alentours de 200 kilomètres. Cette maille commence néanmoins à se rétrécir et, dans un avenir proche, on pourra même tenir compte des lacs. Ces derniers influencent le climat notamment par leur capacité à emmagasiner l'énergie provenant du Soleil durant l'été et à la restituer plus tard dans la saison ainsi qu'à réfléchir la radiation solaire s'ils sont couverts de glace.

## Les prévisions dont ces modèles sont capables jouent un grand rôle dans la politique mondiale en matière de changements climatiques. Jusqu'à quel point peut-on s'y fier?

Les prévisions climatiques basées sur ces modèles sont des résultats scientifiques. Cela signifie qu'elles comprennent toujours des incertitudes qui sont mesurées et rapportées dans les articles scientifiques. Il est rarement fait mention de ce fait. Par ailleurs, et cela est valable pour n'importe quelle simulation, ces prévisions sont le résultat de modèles numériques qui ne sont qu'une approximation de la réalité. Elles ont été obtenues en posant



Simulation du climat réalisée par le Centre national des Etats-Unis pour la recherche atmosphérique. IMAGE: NCAR

un certain nombre d'hypothèses concernant les paramètres que nous ne pouvons pas maîtriser. La croissance économique et démographique, par exemple, sont des grandeurs difficiles à prévoir, alors qu'elles jouent un rôle déterminant dans les émissions de gaz à effet de serre. C'est pourquoi le GIEC (Groupe d'experts intergouvernemental sur l'évolution du climat) a planché sur une quarantaine de scénarios possibles en ce qui concerne ces émissions. Ces scénarios sont désormais imprimés aux modèles et fournissent un éventail de résultats possibles en ce qui concerne, par exemple, l'augmentation de la température globale ou le niveau des mers.

## Que répondez-vous aux climato-sceptiques qui mettent en doute la validité de ces modèles?

Je leur dis que les bases scientifiques de la climatologie numérique sont solides. Bien sûr que les modèles sont imparfaits. La perfection, dans ce domaine, demanderait un temps infini à atteindre. On ne peut pas non plus faire dire à ces simulations autre chose que ce qu'elles sont capables de dire, ce qui signifie que l'on ne peut pas interpréter leurs résultats au-delà d'une certaine limite. Quoi qu'il en soit, il en ressort clairement que les gaz à effet de serre émis par les activités humaines contribuent de manière prépondérante aux changements climatiques et que ces derniers vont déboucher sur une augmentation, plus ou moins importante, de la température mondiale et du niveau des océans. Je n'ai pas vu un seul modèle, aussi simple ou complexe soit-il, qui n'ait jamais prédit le contraire. ■

# QUAND LES GALAXIES TOURNENT

Les simulations numériques permettent de voir en moins d'une minute comment les galaxies se forment au cours de milliards d'années. Des prouesses qui demandent de grandes puissances de calcul

L'astronomie est une discipline gourmande en calcul numérique. Si l'on veut se faire une idée de la dynamique des galaxies sur des millions, voire des milliards d'années, par exemple, il n'y a pas d'autres moyens que tenter de se refaire le film en accéléré sur un écran d'ordinateur. On a beau savoir qu'une étoile comme le Soleil tourne autour

du centre de la Voie lactée à une vitesse de 220 kilomètres par seconde, personne n'a jamais vu une galaxie tourner dans le ciel. Elles semblent désespérément statiques. Toutefois, le souci, pour le chercheur qui souhaite reproduire le mouvement d'une galaxie, c'est qu'un tel objet compte au bas mot quelques centaines de milliards d'étoiles. Sans parler du gaz, de la poussière et de l'hypothétique matière noire.

«Nous ne sommes pas capables de simuler une galaxie de 100 milliards d'étoiles, précise Yves Revaz, chercheur à l'Observatoire astronomique de l'Université de Genève. Notre meilleure performance est une galaxie de 32 millions de «points», simulée sur 7 milliards d'années. Et chacun de ces points représente plusieurs centaines de milliers d'étoiles.»

Les besoins de calcul de telles simulations augmentent très vite. Il s'agit en fait d'un

système physique à  $n$  corps qui évolue, dans une première approximation, sous l'effet de la seule gravité. Chaque point exerce une force gravitationnelle sur l'ensemble des autres, mais est en même temps soumis à  $n-1$  attractions différentes. L'ordinateur doit donc effectuer à chaque instant un nombre de calculs égal à  $n^2$ . En utilisant des approximations raisonnables, certains algorithmes permettent de réduire ce nombre d'opérations. Mais il demeure important.

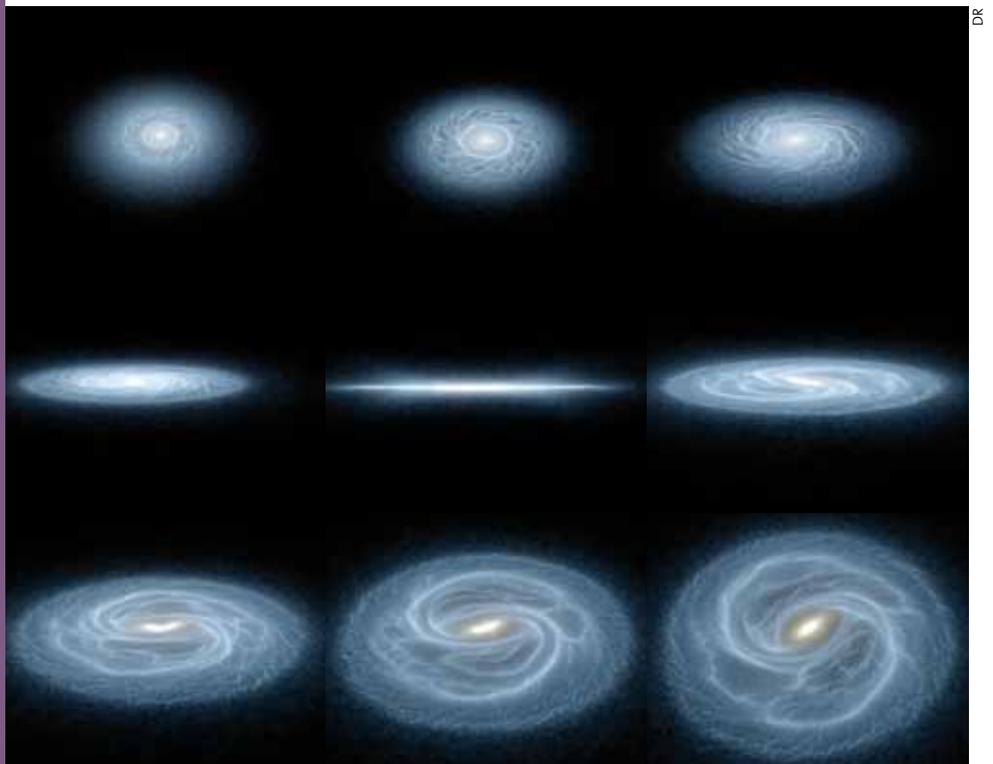
## UN MILLION DE POINTS

Ainsi, si l'on choisit de découper la galaxie en un million de points, il est nécessaire de réaliser plusieurs millions de calculs pour se faire une idée de la forme de l'objet astronomique à un instant donné. Tous les résultats doivent ensuite être gardés en mémoire pour la simulation de l'instant suivant.

## Bras noirs cosmiques

La dynamique des galaxies est une spécialité genevoise depuis plusieurs décennies, sous l'impulsion de Daniel Pfenniger, professeur à l'Observatoire de Genève. L'une des dernières études de son groupe, publiée dans la revue *Astronomy & Astrophysics* du mois de juillet 2009, a consisté à simuler la formation d'une galaxie en spirale dont la particularité est que les bras, essentiellement formés de gaz, s'étendent sur une très grande distance (jusqu'à 130 000 années-lumière du centre). Ce genre de structure, observée dans l'univers, est très difficile à reproduire sur ordinateur. Le comportement du gaz s'apparentant à celui d'un banc de brouillard, les chercheurs ne parviennent pas à faire apparaître des bras stables. A moins d'introduire de la matière noire.

De nature inconnue, impossible à détecter directement à n'importe quelle longueur d'onde, la matière noire n'existe que par le fait que sa masse, importante, permet d'expliquer la dynamique des étoiles au sein des galaxies et de ces dernières à très grande échelle. En l'introduisant dans un modèle reproduisant la formation d'une galaxie, les astrophysiciens genevois ont vu juste. Selon leurs calculs, la matière sombre, assez massive, permet de construire la structure observée. Le gaz, attiré par la force gravitationnelle de ces bras invisibles, joue alors le rôle de marqueur lumineux.



Extraits d'une simulation de la galaxie NGC 6946 (<http://obswww.unige.ch/~revaz/n6946/n6946-x264.avi>).

# T COMME DES TOUPIES

## Sur la trace des exoplanètes noyées dans le **bruit de fond**

La recherche des planètes extrasolaires se raffine de plus en plus. Elle ne se contente plus de dénicher les planètes les plus massives tournant assez près de leur étoile. Elle cherche désormais, noyées dans des systèmes multiples, des compagnons de quelques masses terrestres évoluant si possible dans la zone habitable. Cela signifie que les chercheurs butent contre une limite inhérente à la mesure. Le signal qu'ils cherchent dans les mouvements d'une étoile et qui trahirait la présence d'un compagnon ne se distingue plus clairement du «bruit de fond». C'est pourquoi, les chasseurs d'exoplanètes font depuis peu de

temps, appel aux supercalculateurs.

*«Les perturbations de la vitesse radiale des étoiles provoquées par la présence d'une planète sont parfois très faibles et cachées par plusieurs autres compagnons et par le bruit stellaire, précise Damien Ségransan, chercheur à l'Observatoire astronomique de l'Université de Genève. Pour l'extraire de tout ce bruit de fond, il nous faut traiter les données par ordinateur. Avec une machine conventionnelle, cela peut prendre dix jours pour une seule étoile. Ce n'est plus satisfaisant. Surtout que l'on doit traiter des centaines de cas.»*

Le groupe exoplanètes, dirigé par le professeur Stéphane Udry, a

donc décidé d'acquérir une grappe d'ordinateurs (cluster) assez puissante pour faire le même travail en beaucoup moins de temps. Pour ce faire, le *Data and Analysis Center for Exoplanets* (DACE), dont le but sera de réaliser du calcul de haute performance, mais aussi de stocker les données et d'en permettre l'accès aux chercheurs, a été créé. Il possède à l'heure actuelle un cluster comprenant 72 microprocesseurs «conventionnels» (ils ressemblent à ceux des ordinateurs de bureau) et 1000 microprocesseurs GPU (*Graphics Processing Unit*). Ces derniers, dérivés des cartes graphiques utilisés notamment dans les jeux vidéo,

sont extrêmement rapides, communiquent bien entre eux, mais se contentent des calculs les plus simples.

*«Nous allons tester ce matériel durant un an pour voir ce qui nous est le plus utile, précise Damien Ségransan. Les GPU sont certainement très rapides, mais ne conviennent pas forcément à toutes les simulations. Il nous faut aussi vérifier si leur programmation ne pose pas trop de difficultés. Cela dit, une fois que notre choix sera arrêté, nous avons l'intention, d'ici à quelques années, de monter la puissance de notre cluster à environ une dizaine de teraflops, soit une dizaine de milliers de milliards d'opérations par seconde.»* ■

*«La détermination de l'état du système à un moment donné peut remplir, dans le meilleur des cas, un dixième du disque dur d'un ordinateur de bureau, poursuit Yves Revaz. La lecture de l'ensemble de ces données à l'aide de ce même support prend trois ou quatre minutes. Et comme une simulation est généralement découpée en un millier d'instants», le temps total de lecture, de calcul et d'écriture durerait plus d'une dizaine de jours. Il nous serait impossible de travailler dans ces conditions. L'intérêt des simulations est de pouvoir les répéter à volonté en faisant varier un ou plusieurs paramètres.»*

Pour y remédier, l'Observatoire astronomique possède une grappe d'ordinateurs (cluster) représentant 220 «cœurs» de calcul (ou microprocesseurs), l'une des plus importantes de l'Université. Le tout est relié par un réseau extrêmement rapide capable de délivrer près de 20 milliards d'informations par seconde. Les cœurs et la mémoire de stockage des informations fonctionnent en parallèle. C'est-à-dire que tous les calculs et écritures sont effectués localement avant d'être rassemblés à la fin de la simulation.

Sans être le dispositif le plus puissant de Suisse, il convient parfaitement aux besoins des astrophysiciens, dont les modèles se complexifient progressivement. En effet, une galaxie n'est pas constituée que d'étoiles. Elle renferme aussi du gaz et de la poussière, des matières nettement plus diffuses que les astres lumineux. Du coup, en plus de la gravitation, les chercheurs doivent aussi tenir compte des forces hydrodynamiques.

### INTÉGRER LE GAZ

*«Nous avons réussi à intégrer le gaz dans nos simulations, mais pas encore la poussière, note Yves Revaz. Cela dit, la résolution de ces équations spécifiques complique et rallonge déjà beaucoup nos calculs.»* Mais ce n'est pas tout. Une fois que le gaz a poussé la porte de la modélisation, les chercheurs doivent aussi tenir compte de son refroidissement selon la région galactique où il se trouve.

Les températures ambiantes peuvent en effet varier des environs du zéro absolu à des dizaines de millions de degrés si l'on se trouve dans un amas de galaxies. Ces fortes

variations sont également gourmandes en temps de calcul.

L'un dans l'autre, malgré les difficultés, la simulation rend un résultat raisonnablement fiable et, surtout, peut résumer 10 milliards d'années d'histoire galactique en un petit film de 30 secondes. *«Ces simulations nous permettent d'étudier des phénomènes impossibles à observer dans la réalité, note Yves Revaz. On peut ainsi comprendre comment, entre autres, les galaxies adoptent leur forme.»*

La méthode numérique permet aussi de réaliser des prédictions. Il y a vingt ans, les théoriciens avaient ainsi remarqué, à l'aide de leurs modèles, que les galaxies spirales développaient systématiquement une barre en leur centre, c'est-à-dire un allongement de la région centrale aux extrémités duquel sont accrochés les bras. Problème: les observations d'alors contredisaient ce résultat. Il fallut un perfectionnement des moyens d'observation pour se rendre compte qu'en effet la majorité des galaxies spirales sont barrées. La Voie lactée aussi. La prédiction numérique s'est donc avérée correcte. ■