

Demande pour le Centre de Calcul en région Centre

- Cadre de la demande:

Cette présente proposition de participation au GE, portée par GENCI, vient du projet Calcul Scientifique et Modélisation Orléans-Tours (CaSciModOT). CaSciModOT a pour objectif principal de donner des occasions et moyens de contacts et d'échanges pour tous ceux qui font de la modélisation et du calcul numérique et de favoriser les interactions pluridisciplinaires entre les membres du projet. CaSciModOT est une structure ouverte qui a vocation à rassembler tous les acteurs de la modélisation et du calcul scientifique sur l'ensemble des thématiques des laboratoires (Universités Orléans et Tours, CNRS, INRA, BRGM et le CEA-le Ripault, à partir de 1012) de la région Centre et établir des contacts avec les entreprises de la région Centre.

Le PPF CaSciModOT a été créé en 2004 conjointement pour les universités d'Orléans et de Tours et réunissait 10 laboratoires et fédérations de ces universités. À son renouvellement demandé au quadriennal 2008-2011, le PPF CaSciModOT comptait 19 membres dont les INRA d'Orléans, de Tours ainsi qu'une équipe du BRGM d'Orléans. Le projet CaSciModOT a été labellisé « cluster » de la Région Centre en 2010, indiquant l'intérêt qu'elle porte aux activités de ce projet, sur le contrat d'objectifs suivant : animation scientifique, pérennisation du centre de calcul, relation avec les PME. Pour le futur quadriennal des universités d'Orléans et Tours, la création d'une Structure Fédérative est demandée. Elle reprend le contrat d'objectifs du cluster en les déclinant en :

- Maison de la modélisation
 - Animation scientifique (journées, colloques, workshops...)
 - Coordination des formations (continue, doctorale, master...)
 - hôtel à projets (incubateurs de projets ANR ou ERC, collaboration industrielle)
- Centre de Calcul Scientifique en région Centre (CCSC) :
 - atelier développeurs : formation et/ou autoformation des membres de CaSciModOT au calcul haute performance (K. Hinsen)
 - équipe administrateurs (E. Le Trong)
 - mutualisation des moyens de calculs
- Cellule valorisation
 - actions vers les PME : annuaire, enquête, rencontres
 - offre de collaborations : stages, thèse CIFRE, expertise, projet « sécurité informatique »;
 - lien avec les pôles de compétitivité (DREAM, S2E2....)
 - mise en place d'une structure de pilotage avec les structures chargées de valorisation des différents partenaires (Universités d'Orléans et de Tours, INRA, CNRS (délégation régionale), BRGM, CEA Le Ripault) et avec l'ARITT (Agence Régionale pour l'Innovation et le Transfert de Technologie). Cette structure intégrera également des représentants de PME (en particulier GeoHyd) ou encore l'ADIRC.

Ces objectifs sont en continuité, pour les deux premiers, avec des actions menées par le PPF.

- 1) Description du mésocentre et de l'équipement demandé

1a) Description des équipements actuels (puissance, stockage, type de matériel)

Le cluster fourni par IBM se compose pour la partie calcul proprement dite de 3 BladeCenters de 14 serveurs lames chacun (42 serveurs en tout). Ils sont équipés de deux processeurs quadri-cœurs Intel X5450, 24 Go de RAM et de disques de scratch SAS 10.000 tr/m, pour une puissance théorique cumulée de 4 TFlops pour 336 cœurs. Les services (système de fichiers parallèles GPFS, sauvegarde, réseau, Ordonnanceur de tâches) sont gérés par 4 serveurs IBM xSeries, reliés en fibre optique à une baie SAN de 6 To en RAID 5 et à un robot de sauvegarde à cassettes. Le réseau de calcul est en Infiniband 4X DDR, celui d'administration en Gigabit ethernet. Le lien avec les utilisateurs se fait via un serveur SSH dédié, équipé de tous les outils de développement qui leur sont nécessaires.

1b) Description de l'équipement demandé

La présente demande du CCSC a pour but d'amener la puissance de calcul de ce centre au niveau régional comme indiqué par le schéma directeur de GENCI. Avec des machines européennes au petaflops, des machines nationales à la centaine de terraflops, notre ambition est de monter la puissance du CCSC à plus de 31 terraflops (26 Tflops au moins, compte-tenu de la partie GPU, provenant du présent projet pour 5 Tflops disponibles au CCSC à la suite de l'upgrade prévu en 2011), hors moyens des laboratoires, dont le bilan révèle une puissance de 4 Tflops et dont certains sont ouverts sous condition à la communauté. Une part du futur équipement sera constituée de GPU, d'une part pour répondre au projet « Stockage et traitement des données de l'observatoire des sciences de l'univers en région centre (OSUC) de l'université d'Orléans. Centre de données radio basse-fréquence » (cf. point 2b), mais aussi pour permettre à la communauté de se frotter à de nouvelles architectures. La machine aura une forte capacité de stockage environ 50 TO pour répondre aux projets liés aux bases de données.

1c) Description du mode de fonctionnement du mésocentre

Administration : Le CCSC va bénéficier d'un poste d'administrateur pérenne mis au concours par l'INSU et qui sera pourvu en décembre prochain. Un administrateur, en CDD, est en poste au CCSC. Cependant, afin de répartir les connaissances de l'administration de la machine sur plus d'une personne, nous reprendrons la structure actuelle. Elle comporte un responsable technique, un responsable scientifique et une équipe d'administrateurs. L'administrateur du CCSC sera le pivot de cette équipe. L'idée est de partager les connaissances du fonctionnement de la machine sur plusieurs personnes qui peuvent à la fois s'épauler et palier au départ de l'une d'entre-elles. Les administrateurs, autre que celui du CCSC, sont des administrateurs de parcs de machines des laboratoires faisant partie du projet CaSciModOT. Ils constituent ainsi autant de relais auprès des utilisateurs du CCSC dans les laboratoires.

Ouverture au CCSC : L'ensemble des membres du projet CaSciModOT peuvent avoir accès au CCSC. CaSciModOT se veut un projet ouvert et accueillera tout laboratoire qui voudra rejoindre le projet. Le CCSC est aussi ouvert au monde des entreprises. Cela implique une attention particulière à la sécurité et la confidentialité des données. Cet aspect est pris en compte par l'équipe d'administrateurs et par les chercheurs du LIFO. À cet égard, l'équipe des administrateurs compte le responsable sécurité du CRI de l'université d'Orléans. La liaison avec le monde industriel se fait pour l'heure principalement par le biais des stages des étudiants de master. Si ce projet bénéficie des fonds du grand emprunt, le CCSC sera également ouvert à une communauté plus large dans des conditions qui restent à définir avec GENCI.

Attribution des heures de calcul. La volonté du centre régional est de favoriser l'accès au moyen de calcul haute performance. Aussi, dans cet esprit, il n'y a pas a priori d'attribution d'heures de calcul par projet. Il est pourtant demandé à chaque porteur de projet de rédiger quelques lignes le décrivant et mentionner dans les publications qui en sont issues qu'elles ont bénéficié du soutien du CCSC. Ce soutien ne se réduit pas nécessairement aux heures de calcul, mais peut s'étendre à l'aide au développement ou portage de code, via l'atelier développeurs. A posteriori, si le bilan annuel montre qu'un projet consomme un nombre d'heures trop important, les porteurs de ce projet seront aidés à porter leur demande vers les centres nationaux, voire internationaux, en collaboration avec GENCI.

Gouvernance et Evaluation scientifique : Le CCSC est une entité de CaSciModOT qui est soumise à son mode de fonctionnement : CaSciModOT possède un bureau exécutif et un comité de pilotage composé de 12 représentants de ses partenaires couvrant aussi l'ensemble de ses thématiques scientifiques. Ce comité qui se réunit deux fois par an au moins est l'instance de décision. Dans sa structure fédérative, CaSciModOT est doté d'un conseil scientifique, composé de Laurent Desbat, Jacques Laskar, Jean-Philippe Nominé, qui donne les orientations stratégiques.

1d) Capacité d'accueil des infrastructures

La salle qui accueille la machine du CCSC est sous-dimensionnée pour la future machine. Nous cherchons actuellement une solution avec l'aide du chargé de mission des locaux de l'université d'Orléans. Plusieurs solutions sont à l'étude, notamment avec le Centre de Ressources Informatiques de l'université d'Orléans. Il nous semble aussi important de réfléchir à l'installation de ces machines en termes de green computing.

- 2) Projet(s) scientifique(s)

2a) Environnement scientifique du projet (axes de recherche)

Depuis sa création, le PPF organise des **rencontres biannuelles**. À ce jour 12 journées ont été organisées. Ces conférences, qui se déroulent alternativement sur les sites d'Orléans et de Tours, ont réuni en moyenne de 25-30 personnes en début de période à 50-60 sur les dernières sessions. Les participants proviennent de communautés très diverses. Ces journées ont donné lieu, sur la période de référence, à environ 70 présentations ou démonstrations de solutions logicielles. Le nombre de participants cumulés (et comptés autant de fois que leurs participations) est donc d'environ 350 sur la période.

L'ensemble des supports de conférences et des comptes-rendus des comités de pilotage est accessible à l'adresse <http://fdpoisson.fr/cascimodot/>.

La **liste de diffusion** du projet qui comptait environ 150 personnes au début du quadriennal a dépassé les 300 abonnés en juin 2010. Il y a environ une dizaine de messages envoyés par an et ils sont archivés. Notons que nous n'avons pas souhaité mettre en place de liste de discussions pour privilégier plutôt l'utilisation de la liste du groupe calcul (qui comporte désormais plus de 800 abonnés) et qui permet aux utilisateurs d'échanger directement des informations sur toutes les questions liées au calcul (choix de méthodes, d'algorithmes, de solutions logicielles ou matériels, offres de stage, d'emploi, informations sur les formations, les écoles...). Cette liste, créée en 2003, fonctionne bien et il nous a semblé contre-productif d'en proposer une autre au niveau local. Reste que ces chiffres donnent une idée du nombre d'utilisateur potentiel du CCSC.

Les projets qui utilisent actuellement le centre de calcul sont les projets suivants (ces projets sont détaillés dans le rapport 2010 du CCSC). Certains d'entre-eux (MILES, Climatologie) ont pu être profilé sur le cluster avant d'être lancé sur les machines de l'IDRISS ou du CINES:

- Projet Volcanologie (A. Burgisser, ISTO)
- Projet Climatologie (V. Marécal, LPCE)
- Modélisation sismique et risques naturels (J. Vayron, F. Dupros, BRGM)
- Modèle numérique de terrain (Société GeoHyd)
- Méthode : Modélisation de l'Écoulement sur une Topographie avec des Hétérogénéités Orientées et des Différences d'Échelles. (S. Cordier, MAPMO, F. Darboux, INRA)
- Pierre mise en oeuvre et patrimoine (O. Rozenbaum ISTO, M. Bergounioux MAPMO)
- Détonations dans milieu gazeux (D. Davidenko, ICARE)
- Écoulements compressibles, multi espèces et réactifs : approche MILES (Y. Fédioun, ICARE)
- Étude des propriétés de matériaux inorganiques ou hybrides organiques inorganiques.
- Conception de bibliothèques et outils pour le parallélisme structuré (LIFO)
- Développement systématique de programmes parallèles fonctionnels (LIFO)

2b) Originalité et caractère novateur du projet (ambition vis-à-vis du nouvel équipement)

La création du CCSC est assez récente (la machine est en fonctionnement nominal depuis septembre 2009) mais montre une progression d'utilisation constante jusqu'à sa saturation en février 2010 qui ne s'est pas démentie depuis. L'accueil et la formation de nouveaux chercheurs au calcul intensif, y compris des industriels, ou d'utilisateurs hors région, passe donc nécessairement par une augmentation significative de la puissance de la machine. Tel que le projet est proposé, il inclut des frais d'infrastructure (estimés à 200k€) qui sont indispensables, à la fois pour accueillir la machine et permettre son évolution dans le futur.

Outre les projets mentionnés auparavant qui continuent d'utiliser les ressources du CCSC, les projets suivants sont en émergences :

- **Stockage et traitement des données de l'Observatoire des Sciences de l'Univers en région Centre (OSUC) de l'université d'Orléans. Centre de données radio basse-fréquence** : un centre de traitement de données radioastronomiques basse-fréquence, au bénéfice de la communauté scientifique nationale, accompagnant l'exploitation des instruments actuels et futurs dans ce domaine. Ce centre sera un lieu de stockage et de distribution des données, et d'expertise dans

l'usage des logiciels de traitement adaptés aux instruments du futur (LOFAR SuperStation, SKA...). En plus des calculateurs traditionnels, le centre s'attachera également à généraliser le portage sur GPU (processeurs graphiques, déjà utilisés opérationnellement pour le traitement des données pulsars du RadioTélescope de Nançay) de ces logiciels. Il pourra servir de support à l'exploitation des instruments de la station de Nançay (Radiotélescope décimétrique, Radiohéliographe), tout en augmentant considérablement le retour scientifique des observations. Le caractère innovant réside dans la mise en place d'un service d'observation solaire, qui permettra de développer des outils destinés à la météorologie de l'espace, c'est-à-dire notamment la surveillance de l'activité solaire.

- Mise au point de matériaux innovants : CEMTHI, GREMI.

Simulations multiéchelles des procédés de dépôt par pulvérisation plasma et de la croissance de couches minces catalytiques sur support poreux (P. Brault, J.-M. Bauchire, GREMI)

Ce projet, qui fait également l'objet d'une thèse à compter du 15 septembre 2010, vise à faire un traitement complet de la croissance de films catalytiques par pulvérisation plasma magnétron en traitant à la fois la phase plasma - création et transport des espèces actives - et la croissance de couches minces qui en résulte - dynamique moléculaire et Monte-Carlo. L'originalité de ce travail est de faire le couplage entre la phase plasma et la croissance de sorte que la modélisation de la croissance dispose de conditions initiales réalistes. De ce fait, la comparaison avec les expériences menées au GREMI bénéficiera d'aller-retour simulations - diagnostics pour optimiser les procédés de dépôts. Cette confrontation alimentera aussi les efforts de modélisation qui sont menés actuellement sur le transport des espèces qui se déposent dans le milieu poreux hôte, en collaboration avec le MAPMO (F. James) et l'IJLRA (C. Josserand).

Le code existant est un code de dynamique moléculaire parallèle (sous MPI) impliquant des potentiels d'interaction analytiques qui ne sont pas de paire (issus des liaisons fortes approximation du second moment). Ceci est très coûteux en temps de calculs (3 semaines sur 14 processeurs sur une machine Xeon quadri-processeurs quadruple cœur pour 10000 particules en interaction et injectées successivement pour simuler un flux incident). Le portage sur CCSC ne posera pas de problème a priori.

La partie phase plasma est à concevoir entièrement. Il est prévu qu'elle soit traitée sous COMSOL multiphysics. Le traitement réaliste du problème nécessitera la compilation en mode parallèle.

Production scientifique associée → *Articles revues à comité de lecture* :

1/ T. H. Vo Thi, J.-L. Rouet, P. Brault, J. M. Bauchire, S. Cordier, Ch. Josserand, *A continuous nonlinear shadowing model of columnar growth*, J. Phys. D **41** (2008) 022003

2/ A Caillard, C Charles, R Boswell, A Meige and P Brault, *Deposition of platinum catalyst by plasma sputtering for fuel cells: 3D simulation and experiments*, Plasma Sources Sci. Technol. **17**, 035028 (2008)

3/ P. Brault, Ch. Josserand, J.-M. Bauchire, A. Caillard, Ch. Charles, R. W. Boswell, *Anomalous diffusion mediated by atom deposition into a porous substrate*, Phys. Rev. Lett. **102** (2009) 045901

4/ D. B. Graves, P. Brault, *Molecular dynamics for low temperature plasma-surface interaction studies*, J. Phys. D **42** (2009) 194011, (Topical Review, 27 pages)

Modélisation moléculaire et calculs de réponses spectroscopiques des milieux fondus

(Catherine Bessada, Louis Maksoud, Georges Moussaed, Didier Zanghi, François Vivet, CEMHTI)

Le projet proposé s'intéresse à la description structurale de milieux fondus fluorés afin de comprendre les mécanismes régissant les propriétés de transport observées dans ces systèmes. L'originalité de ce projet articulé autour de deux thèses (bourses région Centre et ANR Milifox) couvrant la période 2010-2013, repose sur le couplage de spectroscopies (RMN, EXAFS) in situ à haute température et la modélisation par dynamique moléculaire des milieux fondus. L'aspect simulation est indispensable pour décrire la nature des complexes formés, leur proportion mais aussi accéder à la dynamique de leurs mouvements dans le liquide fondu.

L'utilisation du code de calcul PIM (développé par Paul Madden, université d'Oxford) dans sa version parallèle optimisée pour tourner avec 8 cœurs, a permis de valider avec succès cette démarche lors d'une précédente thèse (Olivier Pauvert) achevée en 2009. En se basant sur cette expérience, nous estimons que le temps machine nécessaire sur le cluster Phoebus pour mener à bien ce projet est d'environ 300 jours répartis sur les trois années à venir pour un volume de données générées de l'ordre du téraoctet.

Modélisation moléculaire et calcul de réponse en résonance magnétique nucléaire (RMN) dans les matériaux inorganiques et hybrides organiques-inorganiques (Dominique Massiot, Pierre Florian, Franck Fayon, Sylvian Cadars, CEMHTI)

Les approches couplant mesures expérimentales, notamment avec l'apport des hauts champs magnétiques (le CEMHTI est partie prenante du réseau Très Grands Instruments de Recherche - TGIR - RMN Très Hauts Champs, FR3050 CNRS), et calculs par la théorie de la fonctionnelle de la densité (DFT) permettent de repousser de plus en plus loin les limites de l'analyse de la structure moléculaire de matériaux complexes. Nos études portent sur l'exploration de la capacité du calcul à prédire avec précision la réponse en résonance magnétique nucléaire (RMN) dans de nouveaux solides inorganiques cristallins, et sur l'étude des différents degrés et des différentes formes

(structure, composition) d'ordre et/ou du désordre des échelles atomique à nanométrique dans les matériaux. Ceux-ci incluent notamment des hybrides organiques-inorganiques pour la catalyse hétérogène dans le cadre de l'ANR Aluborosil (ANR-09-BLAN-0383 - CSD 3, programme blanc international 2009) ou des matériaux inorganiques impliqués par exemple dans la formation de verres et vitrocéramiques, ou des bio-matériaux (ANR Nanoshap, programme blanc 2009). Le code utilisé est CASTEP (achat d'une licence pour le CCSC en cours de négociation), qui effectue du calcul parallèle (8 à 64 processeurs). Les besoins au cours des trois prochaines années sont estimés à environ 200 000 heures CPU. Le volume de données total devrait chiffrer à environ 1 TO.

Modélisation moléculaire et calcul d'activité spectroscopiques des matériaux d'intérêt pour le nucléaire (Guillaume Guimbretière, Patrick Simon, Marie-France Barthe, François Vivet, CEMHTI)

Les besoins en calculs de structures viennent en complément de la mise en œuvre (projet ANR Ramiris) d'une plateforme instrumentale originale de caractérisation in situ, par des spectroscopies optiques (diffusion Raman, photo- et cathodo-luminescence), de matériaux sous irradiation, en contact avec un fluide ou à haute température. Les matériaux étudiés sont variés, allant du combustible nucléaire (Oxyde d'Uranium) aux verres de stockage (Borosilicates d'alcalins), en passant par les matériaux de structure des réacteurs (Oxyde de Fer, Carbure de Silicium, ...). L'aspect simulation est alors indispensable pour pouvoir différencier les modifications structurales induites par les différents facteurs environnementaux (irradiation, eau, température ...). Cela est d'autant plus impératif lorsqu'il s'agit d'étudier des matériaux présentant une activité radioactive et sur lesquels l'expérimentation est longue et rare. Les codes utilisés actuellement sont VASP et PHONON via l'interface MEDEA, qui effectue du calcul parallèle (8 à 64 processeurs). Le calcul de la densité d'états vibrationnel, appliqué à des clusters d'atomes lourds comme les actinides nécessite de longues séquences de calculs. Pour les trois prochaines années, les besoins en h/CPU et volume de donnée sont alors estimés respectivement à 250 000 et 1 TO.

- astrophysique et physique théorique : LPC2E, LMPT, Nancay,

Développement d'un code dédié au comportement des poussières dans les disques (proto)-planétaires (M. Tagger, LPCEE)

Des développements récents en collaboration avec le laboratoire APC (Paris) ont permis de montrer le rôle important que pourraient jouer les tourbillons dans la coagulation des poussières menant à la formation des planètes. Ces travaux ont été menés avec des simulations hydrodynamiques 3D utilisant le code VAC (Versatile Advection Code), développé à Louvain (Belgique). Ce code moderne (Godunov, parallélisé en MPI) fournit une base solide grâce à ses nombreuses possibilités : utilisable en géométrie cylindrique (adaptée à la description des disques), il existe une version à maille adaptative (AMR) dont nous testons actuellement l'application dans ce même problème. De plus il s'agit d'un code magnéto-hydrodynamique, permettant une extension de ces travaux vers la description des disques magnétisés et des jets.

Au stade actuel, les poussières ne peuvent être ajoutées à la description hydrodynamique que d'une manière approximative. En collaboration avec l'université de Louvain, un nouveau module sera donc développé pour ce code : utilisant des développements algorithmiques récents, ce module permettra de décrire les poussières comme un gaz sans pression, interagissant par friction avec le gaz ordinaire, et décrit avec les mêmes méthodes que celui-ci. Ce module servira en tout premier lieu à des travaux sur la formation des planètes, mais sera utilisable dans de nombreux domaines étudiés au LPC2E impliquant les poussières dans l'environnement des planètes et des petits corps du système solaire.

Pour le développement et les tests, en préparation de runs complets sur les centres de calcul nationaux, ainsi que pour le post-traitement des données 3D, nous envisageons un budget d'environ 30 000 heures/an.

- Vivant : dynamique des protéines, drug design, nutrition : CBM, ICOA, INRA Tours, LMPT

Structure et dynamique des macromolécules biologiques - expériences virtuelles. (G. Kneller, CBM) Depuis les premières simulations de liquides simples dans les années 1960, la simulation moléculaire est devenue un pilier indispensable entre expérience et théorie pour l'étude de systèmes moléculaires complexes. En mimant une partie représentative d'un système macroscopique, elle permet en particulier de mieux comprendre les données expérimentales provenant des techniques spectroscopiques qui sondent la structure et la dynamique de macromolécules biologiques à l'échelle nanométrique [1-8]. On cite ici la diffusion de neutrons, la spectroscopie par résonance magnétique nucléaire, la (bio)crystallographie et la spectroscopie THz. Une approche particulièrement intéressante est d'aller au delà de la simulation d'une expérience « idéale » et d'inclure d'autant plus que possible l'environnement expérimental dans la simulation. Dans cette optique de simulation d'expériences virtuelles, une machine locale sera un élément clé pour les chercheurs au Centre de Biophysique Moléculaire qui s'intéressent au fonctionnement de macromolécules biologiques, en particulier dans l'optique des applications pharmacologiques, car sa puissance de calcul permettra d'aborder des simulations de plus en plus réalistes de ces systèmes. Dans ce contexte, le projet ANR *Simulation d'expériences pour l'étude de la structure et de la*

dynamique de protéines (G. Kneller), dont les partenaires sont l'Institut Laue-Langevin à Grenoble (M. Johnson), l'ENS Paris (D. Abergel) et le Jülich Supercomputer Center en Allemagne (G. Sutmann), pourrait jouer un rôle moteur dans ce contexte

- [1] G. Kneller, "Quasielastic neutron scattering and relaxation processes in proteins: Analytical and simulation-based models," *Physical Chemistry Chemical Physics*, vol. 7, pp. 2641 - 2655, 2005.
- [2] G. Kneller and V. Calandrini, "Estimating the influence of finite instrumental resolution on elastic neutron scattering intensities from proteins," *J. Chem. Phys.*, vol. 126, p. 125107, 2007.
- [3] V. Calandrini and G. Kneller, "Influence of pressure on the fractional relaxation dynamics in proteins: A simulation study," *J. Chem. Phys.*, vol. 128, no. 6, p. 065102, 2008.
- [4] V. Calandrini, V. Hamon, K. Hinsén, P. Calligari, M.-C. Bellissent-Funel, and G. Kneller, "Relaxation dynamics of lysozyme in solution under pressure: Combining molecular dynamics simulations and quasielastic neutron scattering," *Chem. Phys.*, vol. 345, pp. 289-297, 2008.
- [5] R. Fourme, I. Ascone, and G. R. Kneller, "New trends in high-pressure molecular biophysics," *Synchrotron Radiation News*, vol. 22, no. 5, pp. 39-41, 2009.
- [6] G. Kneller and K. Hinsén, "Quantitative model for the heterogeneity of atomic position fluctuations in proteins: A simulation study," *J. Chem. Phys.*, vol. 131, no. 4, p. 045104, 2009.
- [7] P. Calligari, G. Kneller, A. Giansanti, A. P., A. Porrello, and A. . Bocedi, "Group-1 and group-2 neuraminidases: Screw motion analysis reveals fine structural changes induced by binding with oseltamivir.," *Biophysical Chemistry*, vol. 141, no. 1, pp. 117-123, 2009.
- [8] G. Kneller and V. Calandrini, "Self-similar dynamics of proteins under hydrostatic pressure - computer simulations and experiments," *Biochimica et Biophysica Acta*, vol. 1804, pp. 56-62, 2010

- **Traitement de grands volumes de données (LI, LIFO, Nancy, ...)**

Traitement de SIG volumineux (Limet, LIFO et société GéoHyd)

Dans le cadre du projet eXtenGIS labellisé par le pôle DREAM, le LIFO travaille sur la mise au point d'outils permettant d'exploiter des Systèmes d'Information Géographique (SIG) de très grande taille sur des machines parallèles type grappe d'ordinateurs. Des premiers travaux réalisés au sein du LIFO ont permis de mettre en place des algorithmes permettant de traiter des modèles numériques de terrains de très grandes tailles (plusieurs giga-octets). Dans la poursuite de ces travaux il est prévu une thèse en collaboration avec l'entreprise Géo-Hyd portant sur la distribution des données massives sur une architecture distribuée.

La généralisation de l'utilisation des appareils de mesures extrêmement précis, comme les lasers aéro-portés, permet de construire des modèles numériques de terrain (MNT) de plus en plus fins. À titre d'exemple, pour des raisons pratiques de capacité de traitement, les MNT couramment utilisés en géomatique travaillent au pas de 50 mètres alors que l'on est capable de générer des MNT au pas de 1 mètre. Cela signifie que l'on passe, pour un territoire de 100 km², d'un modèle à 4 millions de points à un modèle à 10 milliards de points. Cette explosion des volumes de données à traiter nécessite des moyens de calculs bien plus importants que ceux utilisés jusqu'alors.

Les machines multi-cœurs, qui ont fait leur apparition récemment, ne permettront pas de traiter de tels volumes de données à cause du manque d'extensibilité de leur mémoire vive. Ceci signifie que ce genre de données devra être traité sur des machines distribuées telles que les grappes d'ordinateurs.

Par conséquent les traitements des SIG doivent être re-pensés pour permettre l'utilisation de machines parallèles extensibles en fonction des besoins en calcul, ce qui signifie une répartition des calculs sur la machine cible, mais aussi et surtout une répartition adéquate des données. Au vu des volumes de données à traiter, ce dernier point s'avère crucial pour l'efficacité des traitements à effectuer. Les travaux entrepris dans le cadre de cette recherche vont nécessiter des expérimentations traitant de très gros volumes de données sur des machines puissantes afin de pouvoir valider les solutions proposées.

- **3) Stratégie de valorisation des résultats du projet**

Une collection Hal pour mettre en valeur les travaux dans le cadre du cluster et plus particulièrement ceux qui utilisent le CCSC sera ouverte à la rentrée 2010.

Le projet CaSciModOT, via sa cellule de valorisation, veut établir des points de contact avec les entreprises de la région Centre. Un annuaire des compétences des membres de CaSciModOT est actuellement à l'étude

Actions de formations.

Elles sont cruciales à la fois pour l'utilisation du CCSC, mais aussi pour préparer les chercheurs à l'utilisation des moyens nationaux voire européens. Elles se déclinent suivant 3 axes :

- *Atelier développeur* . L'atelier-développeurs réunit les "développeurs de codes", intéressés aussi par les architectures des machines, de leurs évolutions ainsi que des nouveaux outils et des nouvelles technologies. Cet atelier permet des échanges entre les chercheurs et ingénieurs qui développent des logiciels scientifiques et qui sont souvent isolés dans leurs laboratoires. Sa création date de juin 2008. À ce jour, 7 présentations réunissant une dizaine de participants en moyenne ont été faites à Orléans. Les thèmes abordés vont de la parallélisation aux outils de gestion de version. La liste des présentations est disponible à l'adresse suivante : <http://www.fdpoisson.fr/cascimodot/developpeurs.php>
- *Action de formation à l'utilisation du centre de calcul* : Afin de permettre que la machine soit au mieux utilisée par de plus en plus de chercheurs, plusieurs actions de formations à l'utilisation de la machine du centre de calcul ont été organisées autour de travaux pratiques. Elles permettent à des utilisateurs néophytes d'acquérir les éléments nécessaires à la connexion au centre. Ces formations donnent un aperçu des calculs parallèles et paramétriques.
- *Formation initiale* : En ce qui concerne la formation initiale, les compétences rassemblées par le PPF en termes de modélisation et de calcul scientifique sont mises à la disposition des doctorants via un module de l'école doctorale (cf. <http://www.univ-orleans.fr/ed/st/cours2007/d21.pdf>). De plus Il est proposé dans le contrat quadriennal 2011-2015 la création d'une spécialité de master orienté modélisation et calcul haute (master MoCaHP, Modélisation et Calcul Haute Performance) aux mentions des masters et de mathématique et d'informatique.

Valorisation et Relation avec les entreprises :

Les relations avec les entreprises du CCSC se feront via la cellule valorisation du cluster Cascimodot. Cette cellule sera opérationnelle au 1^{er} janvier 2012. Elle sera dotée d'une structure de pilotage associant le bureau exécutif du cluster et des représentants des les structures chargées de valorisation des différents partenaires (Universités d'Orléans et de Tours, INRA, CNRS (délégation régionale), BRGM, CEA Le Ripault) et de l'ARITT (Agence Régionale pour l'Innovation et le Transfert de Technologie). Cette structure intégrera également des représentants de PME (en particulier la société GeoHyd) ou encore l'ADIRC (Association des Directeurs Informatiques de la Région Centre). Ces actions seront naturellement menées en concertation avec les pôles de compétitivité (DREAM, S2E2....). Les actions prévues sont les suivantes :

- Cette cellule sera chargée des actions vers les PME. Nous allons mettre en place un annuaire de compétences (qui devrait être réalisé dès l'automne 2010) et qui sera communiqué aux entreprises de la région en liant cette information (sur les compétences/expertises présentes parmi les partenaires du projet CaSciModOT) à une enquête afin de connaître leurs besoins. Cette opération fera appel à un prestataire pour la mise en place et le suivi de l'enquête (relances par mail et téléphonique). À l'issue de cette enquête, une série de rencontres ou de visites d'entreprises les plus intéressées sera organisée.
- L'objectif de cette cellule n'est pas de gérer les collaborations mais de les susciter, d'améliorer la visibilité de la recherche académique et des domaines d'expertise présents dans les équipes du cluster. Nous espérons que la visibilité donnée par la labellisation régionale en tant que cluster et la clarification offerte en offrant
- d'une part un « guichet unique » pour les entreprises (et notamment les plus petites) ayant des questions pour lesquels la modélisation ou le traitement informatique peut s'avérer un élément concurrentiel déterminant
- et, d'autre part, en proposant un accès aux moyens de calcul (qui pourra être gratuit lorsque les volumes de calcul et de données seront modestes) et un accompagnement pour l'utilisation du CCSC (via les formations initiales à l'utilisation du centre ou en encourageant la formation continue de personnel des entreprises pour des projets plus conséquents).
- Les offres de collaborations pourront prendre de nombreuses formes et seront ensuite gérées par les partenaires qui auront été mis en relation par le cluster comme par exemple : stages, thèses CIFRE, expertise ou consultance ponctuelle.
- Citons également un projet « sécurité informatique » qui consiste à proposer aux entreprises un accès sécurisé au centre de calcul. Ce projet s'appuie sur des compétences existant au LIFO et notamment liée à la filière SSL du master d'informatique d'Orléans.

- 4) Premiers éléments budgétaires

Demande de financement en équipement

Contributions des utilisateurs (ANR, ERC ...)			100 KE
Participation CaSciModOT			50 KE
Financement Région Centre			250 KE
Mise à niveau en 2016			100 KE
Demande GE			850 KE
Total achat et infrastructures			1350 KE

Modalités du budget de fonctionnement

Administration 1poste + 1/2 ETP	60kE/an	5 ans	300 KE
Fluides, bâtiment	56kE/an	5 ans	280 KE
Coût utilisateur IE calcul, chercheurs 12 ETP	1200KE	5 ans	6000 KE