

Vers des simulations de Dynamique Moléculaire Quantique sur plusieurs milliers de cœurs CPU avec le code ABINIT

Unités CEA impliquées dans le projet :

Laboratoire : Service de Physique de la Matière Condensée

Centre : CEA-DIF, Bruyères-le-Châtel

Responsables: Laurent Colombet (laurent.colombet@cea.fr)
Marc Torrent (marc.torrent@cea.fr)

Laboratoire : Laboratoire de Simulation Atomistique

Centre : CEA-INAC, Grenoble

Responsable: Damien Caliste (damien.caliste@cea.fr)

Introduction : le code ABINIT

Le code ABINIT (<http://www.abinit.org>) est un package dont le programme principal permet de calculer, à partir des premiers principes, la structure électronique d'un système d'électrons et de noyaux (molécules et solides). Les théories implémentées dans le code sont la Théorie de la Fonctionnelle de la Densité (DFT) qui donne accès à l'énergie du système ainsi qu'aux forces s'exerçant sur les atomes, la Théorie de la Fonctionnelle de la Densité Perturbée (DFPT) qui permet de calculer les réponses du système à des perturbations (phonons, constantes élastiques), la Théorie de Perturbation à Plusieurs Corps (MBPT) qui donne accès aux états excités des électrons. Avec ABINIT, il est possible d'optimiser la géométrie du système physique à partir des forces issues de la DFT, ainsi que de réaliser des simulations de Dynamique Moléculaire.

ABINIT est distribué sous la licence GNU-GPL.

Il est utilisé par beaucoup d'équipes du CEA pour de multiples applications : CEA-IRAMIS-SRMP-Saclay (calculs de dynamiques moléculaire, études oxydes d'actinides), CEA-DEC-SESC-Cadarache (études des défauts dans les actinides ; études des produits de fission), CEA-INAC-Grenoble (études de surfaces et de molécules), CEA-LETI-Grenoble (calculs d'états excités dans les semi-conducteurs), CEA-IRAMIS-LSI (calculs de spectroscopies), CEA-DAM (calculs de Dynamique Moléculaire, de phonons), etc.

ABINIT est développé dans un cadre collaboratif par plusieurs laboratoires, principalement en Europe. Le CEA est le principal contributeur à ce développement. Deux groupes participent activement à son développement : le Service de Physique de la Matière Condensée (CEA-DAM), ainsi que le laboratoire de Simulation Atomistique (CEA-INAC). Le CEA-DAM est, en particulier — et en plus de multiples développements de propriétés physiques —, spécialisé dans les développements visant à adapter le code aux ordinateurs massivement parallèles. Ses travaux ont permis de faire tourner efficacement le code sur architecture de type *Tier1* (quelques centaines de cœurs), grâce à l'introduction de plusieurs niveaux de parallélisme en utilisant une technologie « à mémoire distribuée » (MPI). Le CEA-INAC-L_Sim est le principal développeur du code « BigDFT » (code de DFT qui s'appuie sur l'utilisation de fonction « ondelettes »). BigDFT est en interaction étroite avec ABINIT puisqu'il partage avec lui de nombreuses bibliothèques et qu'il est automatiquement incorporé dans ABINIT lors de la compilation. Citons, par exemple, le projet ANR « NEWCASTLE », qui a pour ambition de tirer le meilleur de ABINIT et BigDFT pour obtenir un code capable de s'exécuter sur architecture de type *Tier0* (plusieurs dizaines de milliers de processeurs).

Tout ceci fait d'ABINIT un code utilisé de manière transversal au CEA. Son développement et son adaptation aux nouveaux environnements de Calcul Haute Performance impactent de manière significative les travaux de beaucoup de groupes au CEA, et, en particulier, ceux impliqués dans le programme « nanoscience ».

Depuis la mise en place du consortium « *Prace* » (PaRtnerchip for Advanced Computing in Europe), le choix des architectures des supercalculateurs *Tier0* se décide à un niveau plus large et ne prend pas nécessairement en compte les besoins particuliers des codes de Science de la Matière ; les codes — comme ABINIT — doivent donc s'adapter à ces architectures. ABINIT et BigDFT font, par ailleurs, partie d'une liste de codes soutenus par *Prace* dans le cadre de sa seconde phase d'implémentation (2IP). Dans ce contexte, les codes de calcul doivent s'adapter aux supercalculateurs pour utiliser efficacement les moyens de calcul. C'est le cas pour ABINIT/BigDFT qui sont destinés à être utilisés sur les supercalculateurs du *Très Grand Centre de Calcul du CEA* (TGCC).

L'objectif du présent projet est de garantir une utilisation efficace de ABINIT sur architecture de type *Tier0* (plusieurs milliers de processeurs). Pour cela, il s'avère nécessaire d'adapter le cœur du code (un algorithme de résolution d'un problème aux valeurs propres), en le modifiant pour une utilisation sur architecture hybride (mémoire distribuée et partagée).

Contexte scientifique

Les calculs à l'échelle quantique (« *ab initio* ») sont aujourd'hui d'un intérêt confirmé pour l'étude des matériaux. Parmi leurs différents apports, citons, en particulier les deux suivants :

- Ils permettent une compréhension des phénomènes physiques à l'échelle atomique en traitant, à la fois, les atomes et les électrons, c'est à dire, en prenant en compte la chimie des liaisons électroniques.
- Ils peuvent être intégrés dans des modèles plus larges, multi-échelles.

Ce second aspect prend une part de plus en plus importante dans les projets ; en particulier au Service de Physique de la Matière Condensée du CEA-DAM. Citons, comme exemples, deux applications :

- 1- L'intégration de barrières d'énergie calculées en Théorie de la Fonctionnelle de la Densité permet de simuler la cinétique de réactions entre molécules et surfaces grâce à des codes de Monte-Carlo cinétique. Dans ce type de calculs, la performance du code de DFT influe directement sur la taille de la surface, ainsi que sur le nombre et la taille des molécules, que l'on peut traiter. De plus, le système traité étant fortement inhomogène (il est constitué, à la fois, de vide, de matière cristalline et de molécule), il s'agit d'un très bon test de l'efficacité d'un code.
- 2- Les transitions de phases dans les matériaux « corrélés » peuvent être simulées dans des codes à l'échelle mésoscopique (approche « champ de phase ») s'appuyant sur des propriétés microscopiques calculées à l'échelle quantique, en incluant les formalismes permettant de traiter des fortes corrélations électroniques (DFT+U ou DMFT). Le formalisme DMFT, qui est actuellement le plus sophistiqué pour traiter les fortes corrélations électroniques s'appuie sur la résolution d'un algorithme de Monte-Carlo quantique inséré dans le calcul de DFT. Ce type de calcul est très consommateur en ressources informatiques.

Ces deux applications sont, actuellement réalisées en utilisant le code ABINIT. D'autres études, dans d'autres laboratoires du CEA, s'appuient sur ABINIT pour effectuer des calculs dans un formalisme multi-échelle.

Il est donc d'un grand intérêt, pour le CEA, de disposer d'un code de calcul *ab initio* complet et très performant avec un temps de rendu le plus faible possible.

ABINIT est, en effet, reconnu d'une très grande précision, puisqu'il s'appuie sur l'approche dite « Projector Augmented-Wave » (PAW) qui permet d'obtenir la meilleure précision possible en DFT, dite « Tous Electrons » (Ce formalisme a été implémenté dans ABINIT par le CEA-DAM en 2005). Dès lors qu'il pourra utiliser plusieurs milliers de cœurs CPU, il sera également très rapide et, surtout, permettra de simuler des systèmes de plus grande taille. Par ailleurs, parmi les différents codes de DFT, ABINIT est l'un des codes proposant le plus grand choix de propriétés à calculer (état fondamental, état excité, réponse à perturbation, dynamique, spectroscopie, etc.).

Dans ce contexte, le développement d'ABINIT doit se poursuivre à un bon rythme et son adaptation aux environnements de calculs hauts performances — en constante évolution — permettra de nouvelles avancées, en particulier pour les nanosciences.

Programme de travail

Actuellement, la plupart des simulations réalisées avec ABINIT se font avec la version MPI du code. Il est évident que, malgré ses bonnes performances, cette version n'est plus adaptée à la nouvelle génération de machines parallèles. En effet, l'évolution de leurs architectures va vers un parallélisme massif avec des caractéristiques spécifiques : beaucoup de cœurs avec peu de mémoire par cœur, une mémoire très hiérarchique et des processeurs hybrides (cartes *GP-GPU* ou *Intel MIC*). D'ores et déjà, des machines de 300.000 à 1.000.000 de cœurs sont annoncées. Cette brusque multiplication du nombre de cœurs et la complexité des nœuds de calcul a des conséquences importantes au niveau applicatif, aussi bien pour les codes existants et en production que pour les codes en cours de développement. En particulier, l'approche la plus communément mise en œuvre de nos jours, consistant à utiliser uniquement la bibliothèque de passage de message (MPI) pour paralléliser une application, ne sera plus adaptée à ces nouvelles architectures et va généralement conduire à des extensibilités très limitées. Tirer parti de cette multitude de cœurs disponibles est un vrai défi et va imposer l'utilisation de nouveaux paradigmes de parallélisation. Il est donc impératif de faire évoluer le parallélisme du code ABINIT pour permettre à ses utilisateurs d'aborder des problèmes toujours plus complexes, et de plus grandes tailles sur ces machines.

Une approche hybride consistant à mixer au sein d'un même code une parallélisation sur plusieurs niveaux, un niveau « *multi-thread* » au sein d'un nœud à mémoire partagée et un niveau « MPI » entre les différents nœuds de calcul, est un paradigme dès à présent efficace. Dans cette optique, une première version *OpenMP* (<http://www.openmp.org>) du code est en cours de réalisation pour permettre une

évolution de la structure d'une partie du code. Il est néanmoins nécessaire de réaliser une véritable version hybride du code ABINIT.

Une première étape consistera à étudier les méthodes numériques impliquées dans la réalisation des simulations numériques nécessaires aux études physiques décrites dans le précédent paragraphe. Cela permettra de restreindre la programmation hybride de la nouvelle version du code au strict nécessaire. Les avantages liés à cette programmation sont, outre une meilleure adéquation du code aux spécificités matérielles de l'architecture cible, un gain mémoire et une extensibilité améliorée. De plus, il sera possible de réaliser des essais avec des directives *OpenACC* et *HMPP* (<http://www.caps-entreprise.com/hmpp.html>) pour mieux exploiter les cartes *GPU* ou *Intel MIC* tout en conservant la programmation par directives *OpenMP* pérenne.

Lors de la réalisation de cette phase de développement, il peut s'avérer nécessaire de revoir méthodes mathématiques plus adaptées à un parallélisme localement massivement *multi-threads*.

Le travail réalisé pourra être testé sur les différentes architectures de super calculateurs disponibles à la DIF (<http://www-hpc.cea.fr>) ainsi que la machine curie disponible au TGCC (<http://www-hpc.cea.fr/fr/complexes/tgcc.htm>).

Pour ce qui concerne les applications « cible », dans le cadre du présent projet, il apparaît — comme mentionné ci-dessus — que l'étude de l'interaction d'une molécule avec une surface cristalline est un excellent cas-test.

Par ailleurs, la réalisation du projet ne pourra avoir lieu qu'en étroite collaboration entre physiciens, informaticiens et spécialistes de mathématiques appliquées. Le laboratoire du CEA-DAM entretient à cet effet de nombreuses collaborations comme par exemple avec le laboratoire Exascale (http://www.teratec.eu/technopole/Lab_Exascale.html).

Résumé du sujet de post-doc

Pour mener à bien ce projet, nous envisageons de proposer un stage postdoctoral pour lequel le chercheur recruté aura en charge les différents aspects du projet :

- Analyse des algorithmes actuellement utilisés pour la résolution du problème aux valeurs propres (algorithme le plus consommateur d'un code DFT).
- Mise en place des solutions de programmation hybride pérennes, c'est à dire utilisable sur les moyens de calcul actuels mais aussi dans l'avenir.
- Modification des méthodes numériques pour les adapter à un traitement par une architecture « *many-core* ».

Il sera encadré par une équipe constituée de physiciens spécialistes de la DFT, ainsi que de numériciens confirmés. Il aura accès aux ordinateurs du CEA-DAM-DIF, actuellement classés dans la catégorie des calculateurs « *Tier0* » et il pourra encadrer des stagiaires de M2 travaillant sur le sujet.