

Méthodes mathématiques et analyse numérique pour la chimie moléculaire

La simulation numérique dans le domaine de la chimie moléculaire et dans les domaines connexes de la biologie moléculaire et de la science des matériaux, est un champ du calcul scientifique en pleine expansion. Les besoins sont énormes. Le tout début de l'histoire est en train de s'écrire. Seule une alliance de méthodes d'analyse mathématique pointues, de techniques d'algorithmique numérique adaptées et avancées et de moyens de calcul d'énorme puissance, permettra d'avoir un impact réel sur l'avancée de la science en ces domaines.

La chimie théorique, et plus précisément la chimie quantique, a pour but la modélisation des systèmes de taille moléculaire. L'objectif est de comprendre et prévoir les propriétés de la matière à l'échelle macroscopique à partir de la simulation du comportement de ses constituants à l'échelle microscopique. A cette échelle, les briques constitutives de la matière sont les électrons et les noyaux. L'échelle d'espace est l'angström (10^{-10} m), l'échelle de temps est la femtoseconde (10^{-15} s). Pour une grande gamme d'applications liées à la chimie, les noyaux peuvent être considérés comme des particules classiques alors que la nature quantique des électrons est essentielle pour la compréhension, notamment, de la réactivité chimique. En considérant la position des noyaux comme connue (elle peut être déterminée expérimentalement, ou numériquement par une boucle d'optimisation externe), l'état d'équilibre d'un système moléculaire constitué d'un assemblage de noyaux et d'électrons est déterminé par la résolution de l'équation de Schrödinger stationnaire pour la fonction d'onde électronique. Parallèlement, son évolution est dictée par la version dépendante du temps de cette équation. Nous nous focalisons ici sur la détermination de l'état stationnaire et n'aborderons les questions d'évolution que dans l'encadré 1.

La détermination de la structure électronique d'équilibre est un véritable défi pour la simulation numérique, la dimension de l'espace de configuration d'un système moléculaire, même de taille réduite et comportant par exemple une dizaine d'atomes, pouvant très vite atteindre plusieurs centaines. Pourtant, cette simulation numérique est aujourd'hui essentielle pour compléter l'observation

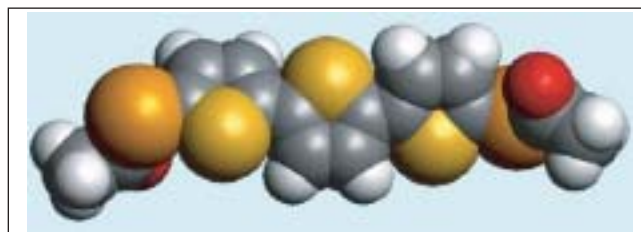


Figure 1 - Un système moléculaire.

expérimentale, car elle en fournit des moyens d'interprétation. Elle peut aussi remplacer cette observation expérimentale, quand cette dernière est impossible. Un exemple typique est l'étude des propriétés de la matière présente au centre de la terre : réaliser des pressions comparables en laboratoire est extrêmement difficile, alors que réaliser la simulation numérique est abordable. Comme le lecteur peut s'en douter au vu du nombre de degrés de liberté en jeu, et comme nous allons le voir, abordable ne veut pas pour autant dire facile.

UNE HIÉRARCHIE DE MODÈLES

Le problème central de la modélisation en chimie quantique moléculaire, et donc le cœur d'un code de simulation numérique en ce domaine, est la détermination de l'état fondamental électronique du système par la résolution de l'équation de Schrödinger stationnaire

$$H_N \Psi = \varepsilon \Psi \quad (1)$$

ou alternativement du problème de minimisation

– Claude Le Bris, CERMICS, École nationale des ponts et chaussées, 6 & 8 avenue Blaise Pascal, 77455 Marne-La-Vallée cedex 2
tél. : 01 64 15 35 73, lebris@cermics.enpc.fr

Je remercie Éric Cancès, Isabelle Catto, Mireille Defranceschi, Maria Esteban, Pierre-Louis Lions, Yvon Maday, Éric Séré et Gabriel Turinici pour plusieurs années de collaboration et d'interaction fructueuses sur le sujet.

Encadré 1

CONTRÔLE DES RÉACTIONS CHIMIQUES

La possibilité de simuler numériquement de manière efficace une évolution d'un système quantique par l'équation de Schrödinger sous sa forme dépendante du temps

$$i\hbar \frac{\partial \Psi(t, x)}{\partial t} = H(t) \Psi(t, x)$$

(où $\Psi(t, \cdot)$ désigne la fonction d'onde modélisant le système à l'instant t , et $H(t)$ le hamiltonien) ouvre la voie à la question du contrôle d'une telle équation. Parallèlement, il est aujourd'hui démontré expérimentalement qu'il est possible, avec un champ laser bien choisi, d'agir sur une molécule pour lui faire atteindre un état prescrit à l'avance. Cela peut par exemple permettre de favoriser, pour une molécule donnée, une voie de dissociation plutôt qu'une autre. La simulation numérique peut aider, par une procédure d'optimisation, à déterminer le champ laser approprié en fonction de la cible que l'on cherche à atteindre. Tout revient dans l'équation de Schrödinger à utiliser un hamiltonien de la forme $H(t) = H_0 + E(t)D$, où H_0 est le hamiltonien du système moléculaire libre, $E(t)$ la forme temporelle du champ laser et D l'opérateur moment dipolaire modélisant l'interaction entre le laser et la matière. Sur le plan mathématique, cela s'inscrit dans le contexte difficile du contrôle dit bilinéaire (le contrôle, $E(t)D$, multiplie l'état, Ψ), ce qui laisse augurer d'importantes difficultés théoriques et numériques. Sur le plan numérique, la recherche d'un meilleur champ laser, ou d'une famille de champs laser convenables, se fait dans ce contexte par des algorithmes stochastiques (pour éviter les pièges créés par la multitude de minima locaux). Ces algorithmes appartiennent le plus souvent à la famille des algorithmes

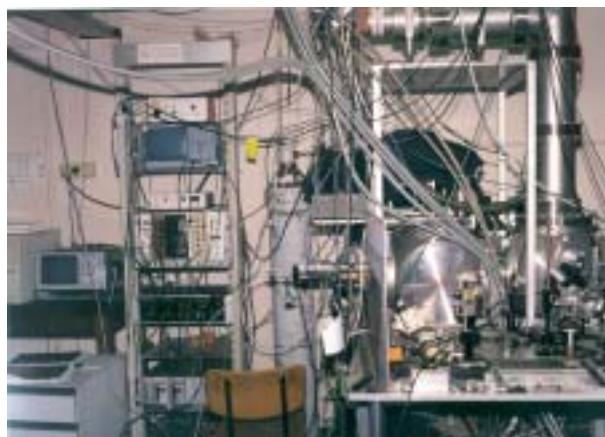


Figure 1 - Un dispositif expérimental de contrôle par laser.

génétiques. Dans ce domaine particulier où la réalisation de milliers d'expériences par minute est possible, une stratégie innovante consiste même à interfacer la boucle numérique d'optimisation avec l'expérience elle-même (plus rapide que sa simulation numérique !). En alliant modélisation physique, techniques expérimentales et simulations numériques, ce champ de la physique est en pleine explosion. Un défi scientifique en passe d'être réalisé est par exemple la réalisation d'un nanotransistor à l'échelle quantique : en actionnant le champ laser, on laisse ou non passer un micro-courant électrique à travers le système moléculaire. Une telle technologie pourrait être à la base d'une nouvelle génération d'ordinateurs.

$$\inf \{ \langle H_N \Psi, \Psi \rangle / \Psi \in L^2(\mathbf{R}^{3N}), \Psi \text{ antisymétrique, } \|\Psi\| = 1 \} \quad (2)$$

Dans ces deux formules, Ψ désigne la fonction d'onde du système et H_N le hamiltonien du problème à N corps (N désignant ici le nombre d'électrons de la molécule)

$$H_N = - \sum_{i=1}^N \Delta_{x_i} - \sum_{i=1}^N \left(\sum_{k=1}^K \frac{z_k}{|x_i - \bar{x}_k|} \right) + \sum_{1 \leq i < j \leq N} \frac{1}{|x_i - x_j|} \quad (3)$$

Dans ce hamiltonien, on a désigné par \bar{x}_k et z_k la position et la charge du k -ième des K noyaux de la molécule,

supposées connues, et par x_i la position du i -ième électron. On oublie ici, pour la simplicité de l'exposé, toutes les questions liées à la variable interne « spin » de l'électron.

Malheureusement pour la pratique numérique, le nombre d'électrons d'une molécule même petite est tel que l'on ne peut pas aborder brutalement la résolution du problème ci-dessus en raison de la complexité (exponentiellement croissante) de la discrétisation d'un espace comme $L^2(\mathbf{R}^{3N})$. On fait donc appel à des simplifications du problème de Schrödinger qui prennent toutes plus ou moins la forme d'un problème de minimisation du type

$$\inf \{ E(\varphi_1, \varphi_2, \dots, \varphi_N), \varphi_i \in H^1(\mathbf{R}^3), \int_{\mathbf{R}^3} \varphi_i \varphi_j = \delta_{ij} \} \quad (4)$$

où la fonctionnelle d'énergie E est dérivée de façon spécifique suivant les applications en vue pour « approcher »

l'énergie réelle au niveau Schrödinger. Les nouvelles fonctions φ_i recherchées (à partir desquelles on bâtera une approximation de la fonction d'onde Ψ globale du système) peuvent être pensées, *grosso modo*, comme les fonctions d'onde associées à chacun des N électrons. La plus célèbre des fonctionnelles d'énergie est celle dite de Hartree-Fock :

$$\begin{aligned}
 E(\varphi_1, \varphi_2, \dots, \varphi_N) = & \sum_{i=1}^N \int |\nabla \varphi_i|^2 \\
 & - \sum_{i=1}^N \int \left(\sum_{k=1}^K \frac{z_k}{|x - \bar{x}_k|} \right) \varphi_i^2(x) dx \\
 & + \frac{1}{2} \sum_{1 \leq i, j \leq N} \left[\iint \frac{\varphi_i^2(x) \varphi_j^2(y)}{|x - y|} dx dy \right. \\
 & \left. - \iint \frac{\varphi_i(x) \varphi_i(y) \varphi_j(x) \varphi_j(y)}{|x - y|} dx dy \right] \quad (5)
 \end{aligned}$$

L'équation d'Euler-Lagrange associée peut se ramener à

$$H_{\varphi_1, \dots, \varphi_N} \varphi_i = \varepsilon_i \varphi_i \quad (6)$$

ce qui est en fait un système d'équations non linéaires aux valeurs propres.

La simplification apparente qui a permis de réduire la taille de l'espace à discrétiser a résulté en une complication qualitative gigantesque du problème. Partant d'un problème de minimisation quadratique (2) (ou d'une équation linéaire [1]), la stratégie a abouti à la minimisation d'une fonctionnelle d'énergie non quadratique (5), et donc à une équation d'optimalité non linéaire (6).

Du point de vue théorique, les difficultés mathématiques sont celles des problèmes de minimisation non quadratiques, souvent non convexes, posés sur des ouverts non bornés. On doit donc éviter les pertes de compacité à l'infini et autres difficultés typiques de ces problèmes. Alternativement, leur considération *via* l'équation d'Euler-Lagrange conduit à l'étude d'équations aux dérivées partielles elliptiques, non linéaires et non locales. Mathématiquement, on peut dire aujourd'hui que les caractères essentiels de ces problèmes sont bien compris. Ce constat est bien sûr un peu schématique, car certaines questions restent encore ouvertes, et l'effort constant des chimistes théoriciens pour dériver de nouvelles et meilleures approximations du problème (2) ne manquera pas de fournir de nouveaux sujets d'étude.

Du point de vue numérique, les difficultés créées par la non linéarité présente dans toutes les approximations sont en revanche aujourd'hui le sujet d'études acharnées.

DES ALGORITHMES EFFICACES ADAPTÉS

La minimisation d'une énergie non quadratique sur un ensemble non convexe (ici essentiellement une sphère) n'est pas un exercice de calcul scientifique facile. Dans le contexte de la chimie moléculaire, les meilleurs algorithmes connus à ce jour utilisent un habile compromis entre la vision « minimisation d'une énergie » (5) et la vision « résolution de l'équation d'optimalité » (6), à savoir ici une équation non linéaire aux valeurs propres (brièvement dit, on fixe $(\varphi_1, \dots, \varphi_N)$, d'où la valeur de $H_{\varphi_1, \dots, \varphi_N}$, que l'on diagonalise pour résoudre (6), jusqu'à convergence). D'importants efforts sont en cours pour une résolution efficace à un nombre d'électrons N fixé. En fait, dans ce domaine l'objectif est double. D'abord, pour un petit nombre N d'électrons (disons moins d'une centaine), ce qui est l'apanage de l'étude théorique de la chimie des petits composés, l'objectif est de résoudre le plus précisément possible le problème. Mais aussi et surtout, un second objectif est de pouvoir aborder des nombres N grands (plusieurs milliers) pour pouvoir traiter les cas réellement intéressants pour des domaines comme la biologie et la science des matériaux. Cela se fera au prix d'approximations et de simplifications, que l'on souhaite pouvoir contrôler par une analyse numérique stricte. Dans cette optique, la donnée déterminante est la complexité algorithmique du problème. Ici, elle est particulièrement grande et parfois presque prohibitive. Disons seulement ici que, pour une méthode de type Hartree-Fock, la phase d'assemblage de la matrice (qui correspond au niveau algébrique à la transcription de la fonctionnelle d'énergie dans la base de discrétisation choisie pour les φ_i) coûte un temps de l'ordre de N^4 . Une fois l'assemblage effectué, la diagonalisation de la matrice coûte N^3 . Il faut réaliser ces deux phases à chaque itération de l'algorithme de minimisation ou à chaque résolution de l'équation aux valeurs propres non linéaire. On comprendra que pour N grand, de tels algorithmes ne sont plus abordables. Pire, des approximations plus précises que l'approximation Hartree-Fock présentent une complexité de N^7 , N^8 , voire plus.

Pour atteindre des valeurs de N pertinentes pour plus d'applications, de nombreuses pistes sont explorées. L'ambition annoncée est la construction d'algorithmes dits linéaires, c'est-à-dire dont la complexité algorithmique est de l'ordre de N .

Ainsi, l'assemblage de la matrice hamiltonienne peut se faire par des techniques dites multipôles, dans l'esprit de celles utilisées en électromagnétisme. Quant à sa diagonalisation, elle peut être évitée, au sens où le seul objet réellement nécessaire pour la détermination de l'état stationnaire est le projecteur sur les N états monoélectroniques de plus basses énergies (les φ_i dans [6]) et non ces états eux-mêmes. Or la détermination de ce projecteur peut se faire à un moindre frais que N^3 , et sous certaines bonnes conditions, en un temps N . Une autre solution technique peut consister à développer des techniques de

type décomposition de domaine, revenant à découper le système moléculaire en différentes parties, dont on peut déterminer plus facilement l'état fondamental, puis à assembler (itérativement) les états partiels obtenus. La boucle d'itération nécessitée par la non-linéarité de (6) doit elle aussi être réalisée au plus faible coût algorithmique. Les efforts développés par la communauté sur de tels sujets sont actuellement gigantesques et font appel aux techniques les plus avancées du calcul scientifique.

EFFETS D'ENVIRONNEMENT

Le système moléculaire que l'on désire simuler n'est bien sûr que rarement isolé. Il est le plus souvent soumis à des interactions avec son environnement. Avant tout, il peut être un constituant d'une phase condensée.

Pour simuler la présence d'une phase liquide, sujet d'étude privilégié de la chimie, une des techniques aujourd'hui la plus efficace est le couplage avec un modèle de l'échelle continue : la molécule étudiée est considérée comme le centre d'une cavité dont l'extérieur est rempli d'un continuum diélectrique de permittivité donnée. La présence de ce continuum se manifeste par la modification du potentiel coulombien d'interaction entre les charges électriques, transformant la loi d'interaction dans (3) et (5). La résolution numérique du problème fera alors appel à des méthodes intégrales, comme en électrostatique.

Pour la phase solide, cristalline par exemple, on prend en compte la répétition périodique du motif cristallin par l'intervention dans les équations du type (5) d'un paramètre « vecteur d'onde » \vec{k} . Il y aura donc une famille de problèmes du type (6) à résoudre, autant que de vecteurs \vec{k} pertinents. La célèbre théorie des bandes, expliquant la conduction des solides, est ainsi totalement explicitée dans une simulation numérique. Pour alléger le calcul, des simplifications peuvent souvent être faites, comme celle consistant à ne considérer explicitement dans le calcul que certains électrons (ceux des couches extérieures) dans chaque motif cristallin, agrégeant l'effet des autres (ceux des couches de cœur) dans des potentiels effectifs ajoutés.

Au-delà de la phase à laquelle le système moléculaire appartient, celui-ci peut subir des interactions d'autres

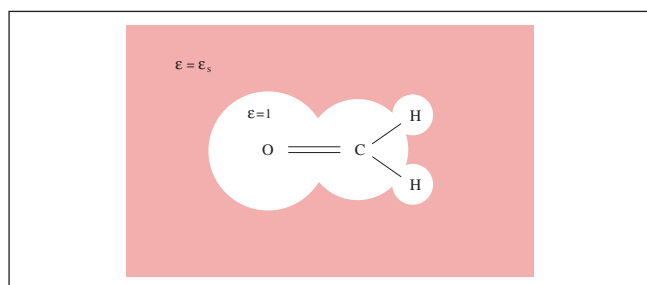


Figure 2 - Molécule plongée dans un solvant liquide.

origines : champs électriques, magnétiques... La simulation de ces interactions est par exemple un moyen de déterminer la polarisabilité d'un composé donné, son aptitude à présenter des propriétés optiques intéressantes, par exemple pour réaliser des doubleurs de fréquence (on éclaire un composé avec une fréquence ω et la lumière transmise ressort avec une fréquence 2ω , caractéristique d'un couplage non linéaire entre la lumière et la matière à l'échelle moléculaire). Les techniques mathématiques et numériques employées dans ce cadre sont le plus souvent celles des méthodes de perturbation.

VERS LES ÉCHELLES SUPÉRIEURES

La simulation numérique de modèles quantiques telle que nous l'avons décrite à ce stade est redoutablement précise mais, même mise en œuvre avec des méthodes numériques pointues, elle reste malheureusement limitée à ce jour à des composés de taille petite ou moyenne. Pour aborder des systèmes plus gros, composés de millions d'atomes, des simplifications sont nécessaires. La plus naturelle est de ne plus distinguer la sous-structure des atomes (électrons + noyaux) et de gérer ceux-ci comme des particules ponctuelles interagissant, *via* la mécanique newtonienne, par des champs de force. L'information de nature quantique peut être efficacement insérée dans ce modèle simplifié, par l'intermédiaire des champs de force qui peuvent être eux-mêmes précalculés par des modèles quantiques précis. Alternativement à ce couplage séquentiel, on peut aussi penser à coupler en parallèle un modèle de mécanique newtonienne avec des modèles quantiques. Ainsi, la modélisation quantique n'apparaît plus comme une discipline réservée aux petits systèmes, mais comme un moyen d'alimenter par une modélisation rigoureuse et contrôlée un niveau de description de la matière plus rudimentaire, mais aussi d'applicabilité plus large. Le célèbre champ de la dynamique moléculaire, qui est probablement le champ de la modélisation en chimie théorique ayant à ce jour le plus d'impact sur le monde expérimental et industriel (hormis les techniques de visualisation !) ne peut désormais se passer de paramètres calculés au niveau quantique ou même de couplage avec une modélisation quantique.

Du point de vue mathématique et numérique, les défis ouverts par la dynamique moléculaire ne sont pas moindres. Résoudre sur des temps longs un énorme système d'équations différentielles ordinaires

$$m_i \frac{d^2 x_i}{dt^2} = F(x_1, \dots, x_K) \quad (7)$$

présentant des échelles de temps et d'espace très disparates soulève des questions d'analyse numérique difficiles sur la symplecticité des schémas, le calcul simplifié des interactions à longue distance par schémas multi-pas, etc. C'est à ce prix que l'on peut simuler des phénomènes

SIMULATIONS MULTI-ÉCHELLES EN SCIENCE DES MATÉRIAUX

L'arrivée de matériaux de nouvelle génération et la nécessité de pousser des matériaux existants jusqu'à leurs limites peuvent poser un problème difficile de modélisation. L'établissement de lois constitutives représentatives du comportement mécanique de ces matériaux, dans le régime d'utilisation qui va être le leur et qui est souvent éloigné du champ connu, est parfois un véritable défi. L'alternative à l'expérimentation en laboratoire est le développement de simulations numériques multi-échelles, où est décrite la manière dont comportement macroscopique du matériau et comportement de ses constituants à l'échelle microscopique sont reliés. Ainsi, pour des matériaux cristallins, on peut faire intervenir dans la simulation la façon dont les atomes constitutifs du réseau cristallin sont déplacés par le champ de déplacement macroscopique subi par le matériau et comment, en retour, le déplacement de ces atomes, influe sur le champ de contraintes présent dans le matériau. Un modèle du champ de forces à l'échelle atomique, pouvant par exemple provenir lui-même de modèles de type quantique comme ceux décrits dans le corps du texte, reliant déplacement des atomes et forces qu'ils subissent, est donc couplé avec des modèles de mécanique du continuum pour décrire l'échelle supérieure. La figure présente un exemple schématique d'un calcul numérique résultant d'une telle modélisation : en chaque point de Gauss de la triangulation utilisée pour le calcul des éléments finis, le calcul du champ de contrainte provient de l'interaction entre les atomes déplacés de leur position d'équilibre. Une démarche analogue est adoptée en mécanique

des fluides où on peut décrire la rhéologie des fluides avec micro-structures (comme les suspensions ou les fluides polymériques) à l'aide de modèles à l'échelle atomique. Dans tous les cas, l'utilisation intensive des tels modèles multi-échelles pose une foule de questions mathématiques et numériques (dérivation rigoureuse des modèles variationnels, caractère bien posé des problèmes d'évolution, convergence des schémas numériques), dont l'étude n'est aujourd'hui qu'à peine amorcée.

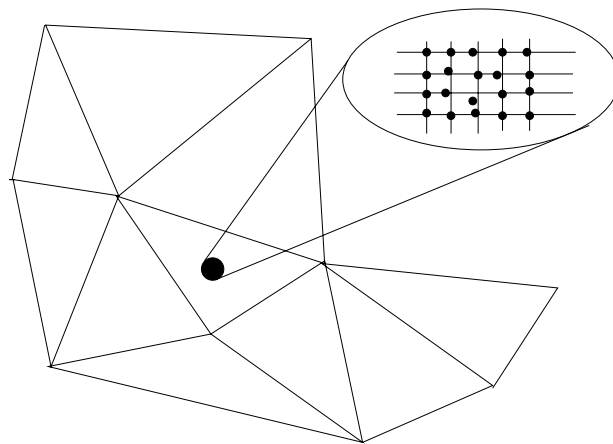


Figure - Une méthode d'éléments finis couplée à une simulation à l'échelle atomique. © Benoît Soep, LFP-CNRS

aussi complexes que le repliement d'un brin d'ADN en solution ou les dégâts causés par une cascade d'irradiation dans une structure cristalline.

L'AVENIR

La simulation numérique en chimie moléculaire et, plus généralement, à l'échelle atomique nécessite des efforts grandissants, à l'interface de plusieurs communautés scientifiques : physiciens, chimistes, mathématiciens, numériciens, informaticiens. Elle est en route pour devenir indispensable à double titre. En premier lieu,

avec la miniaturisation des systèmes technologiques, le traitement des échelles microscopiques, voire nanoscopiques, devient une nécessité. En second lieu, la complexification des systèmes de taille usuelle et la nécessité de les comprendre dans leurs plus fins détails pour les optimiser rendent indispensable le transfert de l'information provenant de l'échelle microscopique pour mener à bien correctement les simulations effectuées aux échelles supérieures, le plus souvent macroscopiques. Dans un tel contexte, les mathématiques appliquées ont un rôle moteur à jouer.

POUR EN SAVOIR PLUS

Cancès (E.), Le Bris (C.), « International Journal of Quantum Chemistry », *Volume 79*, 2000, 82.

Ciarlet (Ph.-G.), Le Bris (C.), Éditeurs, « Handbook of Numerical Analysis, special volume Computational Chemistry », *Elsevier*, 2003.

Defranceschi (M.), Le Bris (C.), Éditeurs, « Mathematical models and methods for ab initio Quantum Chemistry », *Lecture Notes in Chemistry, volume 74*, Springer, 2000.