

## PROPOSITION DE SUJET DE THÈSE

**Laboratoire** : Laboratoire Interdisciplinaire Carnot de Bourgogne (ICB)

**Titre de la thèse** : Modélisation des spectres infrarouges du silane et du germane par calculs *ab initio* : apport de l'intelligence artificielle

**Direction et éventuelle co-direction de la thèse (grades, HDR)**

Michaël Rey (DR CNRS, HDR) – directeur de la thèse

Dominika Viglaska (MCF, non HDR) – co-encadrante de la thèse

**Adéquation scientifique avec les priorités de l'établissement**

Ce sujet s'inscrit pleinement dans les thématiques du groupe MARS (*Molecules, Atoms, Reactivity and Scattering*) du département ICQ du laboratoire ICB et de ses projets en cours sur la modélisation de l'absorption infrarouges de molécules pour les applications atmosphériques, planétologiques et industrielles. Il est aussi lié à l'arrivée dans l'équipe de M. Rey et D. Viglaska depuis le 1<sup>er</sup> janvier 2026 et vise à implanter de nouvelles méthodes de calcul innovantes. L'interaction avec le nouveau département « Systèmes Numériques » est fortement envisagée pour le recours aux méthodes liées à l'IA.

**Descriptif du sujet (enjeux scientifiques, applicatifs, sociétaux...)**

La spectroscopie moléculaire est un outil de diagnostic performant permettant d'identifier et de caractériser des structures complexes dans les spectres, afin de pouvoir modéliser différentes atmosphères planétaires, y compris celle de la terre. La construction de bases de données spectroscopiques précises est essentielle pour répondre aux besoins croissants des différentes communautés scientifiques. Depuis une trentaine d'années, la demande des planétologues et des astrophysiciens auprès des spectroscopistes n'a cessé d'augmenter, avec une accélération notable depuis le succès de la mission Cassini-Huygens. Les missions spatiales actuelles et futures, telles que JWST, PLATO ou ARIEL, contribuent également fortement à cette dynamique.

Dans ce contexte, l'apport de la théorie prend tout son sens, à travers des collaborations étroites avec les expérimentateurs et les planétologues/astrophysiciens, tant au niveau national qu'international, renforçant ainsi le caractère interdisciplinaire du domaine. Ces interactions se concrétisent aujourd'hui par des projets ambitieux, notamment sur l'étude des gaz chauds dans les milieux exo-planétaires. Toutefois, fournir des calculs d'opacité moléculaire précis reste un processus long et fastidieux, qui nécessite souvent la caractérisation d'états moléculaires fortement excités.

Les Hamiltoniens effectifs ont largement contribué au rayonnement de la spectroscopie moléculaire à haute résolution pour les systèmes présentant une structure énergétique bien définie, organisée en "polyades", c'est-à-dire en groupes de niveaux vibrationnels fortement

couplés. L'équipe ICQ/MARS possède une expertise solide dans la construction de modèles effectifs reposant sur l'ajustement de paramètres aux données expérimentales, en particulier pour les molécules toupies sphériques. Cependant, la spectroscopie moderne actuelle s'oriente de plus en plus vers l'étude d'espèces moléculaires complexes, avec la présence possible de bandes chaudes, y compris à température ambiante. La forte densité d'états rovibrationnels pour les énergies élevées ainsi que les nombreuses résonances rendent alors l'analyse des spectres infrarouges particulièrement difficile. La présence d'états dits « *dark* » — c'est-à-dire non observables expérimentalement — constitue l'un des principaux obstacles de l'approche effective traditionnelle. En effet, le manque d'information expérimentale peut conduire à des erreurs significatives dans la détermination de certains paramètres de résonance, avec des conséquences directes sur le calcul des intensités.

Une approche alternative et complémentaire consiste à résoudre l'équation de Schrödinger indépendante du temps afin de caractériser explicitement l'ensemble des résonances. Pour cela, une surface *ab initio* d'énergie potentielle obtenue par des calculs de chimie quantique est considérée et une approche variationnelle est utilisée pour résoudre le problème nucléaire. Cette approche globale est particulièrement bien adaptée aux applications (exo)planétaires pour lesquelles la complétude des données spectroscopiques est essentielle. L'équipe a développé un ensemble de méthodes algébriques et numériques dédiées au calcul des énergies et états propres moléculaires ; en contrepartie, ces calculs peuvent s'avérer très coûteux.

Pour pallier cette difficulté, l'équipe est désormais capable de construire des Hamiltoniens effectifs directement à partir de surfaces *ab initio*. Cette stratégie permet de prendre en compte un grand nombre de résonances selon un schéma en polyades défini par l'utilisateur. Elle fournit ainsi un jeu de paramètres initiaux physiquement pertinents et assiste le spectroscopiste dans l'analyse, en particulier lors de la phase d'attribution, étape clé pour pouvoir affiner les paramètres *ab initio*. Néanmoins, en présence de polyades très complexes, la procédure d'attribution peut encore s'avérer fastidieuse. Dans ce contexte, l'apport de l'intelligence artificielle et des méthodes d'apprentissage automatique pour l'attribution spectrale apparaît comme une piste particulièrement prometteuse.

Cette thèse a pour objectif de coupler les outils spectroscopiques aux calculs *ab initio* — *via* des Hamiltoniens et des moments dipolaires effectifs non empiriques — ainsi qu'à des méthodes d'apprentissage automatique, notamment afin d'automatiser la procédure d'attribution spectrale. Nous nous intéresserons pour cela aux molécules de silane ( $\text{SiH}_4$ ) et de germane ( $\text{GeH}_4$ ) qui, bien que beaucoup moins abondantes que des espèces telles que  $\text{H}_2\text{O}$ ,  $\text{CO}_2$  ou  $\text{CH}_4$ , présentent un intérêt planétaire et exo-planétaire réel, en particulier dans les atmosphères de Jupiter, de Saturne et des exoplanètes chaudes. L'équipe possède déjà une expertise avancée pour ces deux molécules, avec des prédictions théoriques fiables pour les premières polyades. Les surfaces d'énergie potentielle et de moment dipolaire *ab initio* sont également disponibles. L'objectif sera donc d'étendre les calculs actuels vers des polyades plus élevées, en construisant d'abord un modèle effectif *ab initio* complet pour chacun de ces deux systèmes, puis en développant des outils numériques dédiés à l'automatisation de la procédure d'attribution spectrale. Cette procédure sera d'abord validée sur des résultats déjà connus.

Plusieurs environnements d'apprentissage automatique développés en Python seront considérés et testés. Outre TensorFlow et Keras, d'autres bibliothèques de référence seront évaluées, telles que PyTorch ou scikit-learn. Cet écosystème permettra de comparer différentes stratégies d'apprentissage et d'identifier les approches les mieux adaptées à l'automatisation de l'attribution spectrale. Des interactions avec le Systèmes Numériques du laboratoire ICB sont envisagées. L'enregistrement de nouveaux spectres expérimentaux se fera en collaboration avec le LISA (Créteil).

### Contexte partenarial (cotutelle, partenariat avec un autre laboratoire, une entreprise...)

Une collaboration étroite se fera en interne, au sein de l'équipe MARS, avec Vincent Boudon (DR CNRS) et Cyril Richard (IR CNRS), qui possèdent une grande expertise dans les analyses de spectres de molécules toupies sphériques et dans la gestion de bases de données (base CaSDA de l'équipe ICQ/MARS). Concernant l'enregistrement de nouveaux spectres expérimentaux, cette thèse se fera en collaboration avec Fridolin Kwabia Tchana (PR) au LISA (Créteil).

### Impacts (scientifiques, technologiques, socio-économiques, environnementaux, sociétaux...)

Outre l'utilisation de calculs *ab initio* en spectroscopie moléculaire à haute résolution — domaine dans lequel seules quelques équipes au niveau international sont aujourd'hui capables de produire des listes spectroscopiques complètes à partir de surfaces *ab initio* (NASA Ames, ExoMol/UCL, Molpro/Stuttgart et ICB/Dijon) — cette thèse vise également à intégrer des méthodes d'intelligence artificielle dans la trousse à outils du spectroscopiste, même si de nombreux groupes explorent déjà cette voie, notamment pour la construction de surfaces d'énergie potentielle.

L'analyse des spectres moléculaires reste en effet une tâche longue et exigeante, où des erreurs d'attribution de raies peuvent entraîner des biais significatifs dans la détermination des paramètres de l'Hamiltonien. L'automatisation, même partielle, de cette phase grâce à l'IA constituerait donc une avancée majeure.

De plus, les données obtenues au terme de cette thèse seront directement intégrées dans la base de référence HITRAN (Harvard & Smithsonian), ainsi que notre base locale. Cette thèse a donc également pour but d'enrichir les données spectroscopiques de la littérature et de les rendre disponible.

### Programme de travail du doctorant (tâches confiées au doctorant)

#### Calendrier de réalisation

- Les premiers mois de la thèse (**0–6 mois**) seront l'occasion pour le/la doctorant(e) de se familiariser avec les méthodes de spectroscopie, notamment celles développées au sein de l'équipe. L'objectif sera de comprendre et d'assimiler les notions nécessaires à la construction d'un Hamiltonien moléculaire et d'un modèle effectif, ainsi qu'à la maîtrise de l'approche variationnelle, du calcul et de l'analyse de spectres. Les premiers spectres synthétiques seront alors calculés, tout d'abord pour le germane, puis pour le silane. L'effet de la température sur les spectres sera étudié, en particulier l'augmentation du nombre de bandes chaudes avec la température, ainsi que l'augmentation de la densité spectrale associée à des valeurs élevées du nombre quantique rotationnel  $J$ . La fin de la première année (**7–12 mois**) sera consacrée à un état de l'art des méthodes existantes dans le domaine du *machine learning* (réseaux de neurones profonds, etc.).
- La deuxième année (**13–24 mois**) sera pleinement dédiée au développement de codes informatiques en Python pour la mise en œuvre de techniques d'apprentissage en vue de l'attribution quasi automatique des spectres. Afin de valider ces codes, les premiers tests seront effectués sur des régions spectrales où la densité d'états moléculaires n'est pas très élevée et pour lesquelles des résultats de référence existent déjà (par exemple, la région des bandes fondamentales de pliage de  $\text{GeH}_4$  et  $\text{SiH}_4$ ). En parallèle, de nouveaux spectres expérimentaux seront enregistrés au LISA, et le/la doctorant(e) pourra éventuellement se déplacer à Créteil afin de participer à ces expériences. De plus, des listes théoriques seront construites pour  $\text{GeH}_4$  et  $\text{SiH}_4$  dans la gamme spectrale  $0–5000 \text{ cm}^{-1}$ , dans un premier temps à température ambiante.

- La troisième année (**25–31 mois**) sera consacrée à l'analyse des nouveaux spectres expérimentaux pour les polyades plus élevées de GeH<sub>4</sub> et SiH<sub>4</sub>, en testant les modèles d'intelligence artificielle sélectionnés par le/la doctorant(e). La fin de la thèse (**32–36 mois**) sera dédiée à la rédaction du manuscrit.

L'objectif de cette thèse est de permettre la publication de plusieurs articles dans des revues internationales à comité de lecture. Indépendamment de la partie consacrée à l'intelligence artificielle, une ou plusieurs publications pourront porter assez rapidement sur la modélisation des spectres infrarouges du silane et du germane à partir de calculs *ab initio* de haute précision, puisque les surfaces ainsi que les codes de calcul permettant la résolution de l'équation de Schrödinger nucléaire existent déjà.

**Accompagnement du doctorant / Fonctionnement de la thèse** (*accompagnement humain, matériel, financier, en particulier pour la prise en charge du fonctionnement de la thèse et des dépenses associées*)

Le/La doctorant(e) disposera de tous les moyens nécessaires au bon déroulement de la thèse. Un accompagnement humain sera assuré dès le démarrage, notamment pour l'ensemble des démarches administratives. Un bureau et un ordinateur lui seront fournis, ainsi qu'un accès au centre de calcul de l'UBE. Le/La doctorant(e) présentera ses résultats lors de congrès nationaux et internationaux et les frais correspondants seront pris en charge par l'équipe (sur crédits récurrents ou projets).