



Offre de Stage Master 2

Utilisation du *transfert learning* pour améliorer la prédiction des spectres d'absorption moléculaires par *machine learning*.

Période de stage : 6 mois, début en février- ou mars 2025

Organisme d'accueil : Laboratoire CEISAM (UMR 6230) - Nantes Université

Description du sujet de stage

Contexte et enjeux

Le spectre d'absorption d'une molécule représente la quantité de lumière absorbée pour différentes longueurs d'onde de lumière incidente et indique donc les passages d'une molécule de son état électronique fondamental (GS : *ground-state*) à un état excité (ES : *excited state*). La compréhension des phénomènes d'excitation moléculaire permet de rationaliser et d'optimiser des processus chimiques, biologiques et physiques, avec des applications allant de l'identification de substances en chimie analytique à l'imagerie médicale et au développement de matériaux avancés.

Cependant, contrairement à leurs homologues de l'état fondamental, les états excités (ES) ont une durée de vie de courte durée (de quelques fs à quelques ns), ce qui rend leur caractérisation expérimentale précise difficile et souvent coûteuse. C'est l'une des raisons expliquant pourquoi les méthodes théoriques capables de fournir les propriétés des ES à un coût de calcul raisonnable restent d'actualité. Jusqu'à présent nous utilisons principalement au CEISAM des méthodes *ab initio* dites « statiques » qui ont été développées au cours des dernières décennies par les chimistes quanticiens, chacune de ces méthodes présentant un compromis entre précision, généralité, et coût.

L'utilisation de la *machine learning* en chimie quantique a permis l'avènement d'approches plus efficaces, et a requis le développement de nombreux *datasets*, comprenant des informations détaillées (structures, énergies et autres propriétés) sur des molécules. Depuis 2018 en collaboration avec l'équipe THEO du LCPQ de Toulouse, l'équipe ModES du CEISAM a développé un ensemble de valeurs de référence pour les ES qui a mené au *dataset* de référence dans le domaine : QUEST-DB¹. L'utilisation de QUEST-DB par l'équipe Google-DeepMind pour former une nouvelle approche Monte-Carlo/Neural-Network pour les ES³ en est une illustration.

Objectifs du stage

Dans ce stage de M2, nous proposons de prédire le plus précisément possibles les énergies d'excitations et les forces d'oscillateur associées aux ES de petites molécules en utilisant des approches dérivées du ML en utilisant les données de QUEST-DB.

Les Graph Neural Network (GNN) gagnent en popularité pour les tâches liées à la chimie en raison de leur capacité à traiter des structures complexes. Dans ce cadre, le stagiaire devra concevoir une architecture GNN adéquate pour prédire des propriétés d'excitation spécifiques, telles que les longueurs d'onde d'absorption ou les énergies d'excitation.

Dans un premier temps, elle préentraînera un premier modèle à l'aide d'un *dataset* de chimie quantique² ayant une plus large variété de données que QUEST-DB mais à un niveau de précision moindre. Dans un second temps, grâce au *transfert learning* qui s'est avéré efficace pour compenser le manque de données, le stagiaire pourra ajuster son modèle à l'aide de notre *dataset* QUEST-DB. Le stagiaire réalisera un *dataset* de validation de données issues de nos calculs théoriques non apprises pour évaluer cette approche combinée utilisant les GNN et le *transfert learning*.



Offre de Stage Master 2

Ce stage permettra donc d'obtenir une représentation des spectres d'absorption électronique de molécules de manière efficace et précise ; et nous pourrons déterminer comment elles absorbent la lumière à différentes longueurs d'onde sans avoir recours au moindre calcul de chimie quantique.

Mots-clés : Énergie, État excité, Chimie quantique, dataset, GNN, transfert learning, Python.

Candidat.e recherché.e

- Étudiant.e en Master 2 (M2) en Intelligence Artificielle, Data Science, Informatique, ou encore chimie.
- Bonnes compétences en communication, à la fois pour la documentation technique et la présentation des résultats.
- Bonne maîtrise de Python pour le traitement des données et le développement de modèles d'IA.
- Expérience avec des bibliothèques comme TensorFlow, Keras, PyTorch, Scikit-learn.
- Compétences en manipulation de données avec Pandas, NumPy, et gestion des datasets d'IA
- Curiosité pour la recherche scientifique
- La connaissance du domaine scientifique n'est pas nécessaire mais sera appréciée.

Informations pratiques

Lieu de déroulement du stage : Laboratoire CEISAM, UMR 6230 2 rue de la Houssinière BP92208 44322 Nantes cedex 3, Nantes Université

Encadrement : Aymeric Blondel (Ingénieur en calcul scientifique, CEISAM, équipe Modes), Pierre-François Loos (Directeur de recherche CNRS, LCPQ, Groupe Theo)

Financement : Rémunération conformément à la réglementation en vigueur du stage par le cluster FAISTOS

Candidature/contacts : candidature à envoyer avec CV et lettre de motivation , à l'adresse suivante : aymeric.blondel@univ-nantes.fr

Références

¹ M. Vénil, A. Scemama, M. Caffarel, F.Lipparini, M. Boggio-Pasqua, D. Jacquemin, P.F. Loos, *Wires Comput. Mol. Sci.* 2021, 11, e1517 https://lcpq.github.io/QUESTDB_website/

²A. Ullah, Y. Chen, P. O. Dral, *Molecular Quantum Chemical Data Sets and Databases for Machine Learning Potentials.* <https://arxiv.org/abs/2408.12058>

³D. Pfau, S. Axelrod, H. Sutterud, I. von Glehn, J. S Spencer, *J. Science*2024, 385, 6711.