

**Algorithmes d'apprentissage en chimie quantique
et applications au screening (sélection) de matériaux pour cellules photovoltaïques**
Sujet de projet de 3e année M2 Maths (2018–2019)

D. Rocca ¹, J. Unterberger ²

La matière condensée à notre échelle (molécules, solides à température ordinaire) est modélisée par ce qu'on appelle le *problème à N corps*, à savoir un système de $N \approx \mathcal{N}_A \approx 10^{23}$ équations de Schrödinger couplées, dont la solution analytique est impossible à obtenir. Deux développements récents, concomitantes à l'amélioration exponentielle des capacités des ordinateurs et du calcul parallèle, ont néanmoins permis *deux révolutions* de portée considérable. La *première* (à partir des années 80) a été l'utilisation avec un succès croissant des techniques dites de DFT ("density functional theory"), mélange d'arguments théoriques rigoureux et d'approximations de plus en plus raffinées, permettant d'obtenir au niveau de précision dit chimique (coïncidant avec le niveau de précision expérimental) le niveau d'énergie du fondamental ainsi que les premiers états excités. La *deuxième* est consécutive à l'explosion des méthodes de machine learning (apprentissage assisté par ordinateur) au milieu de la décennie 2000; elle a ouvert la voie à la fouille systématique de l'espace moléculaire (espace de toutes les molécules possibles) par extrapolation des données considérables obtenues grâce aux expériences et à la DFT. Le couplage DFT/machine learning est un domaine en plein essor et très prometteur, notamment pour les applications technologiques et industrielles. Celles que nous avons en vue concernent la sélection par ordinateur ("screening") de matériaux pour cellules photovoltaïques avec le meilleur rendement possible.

L'objectif de ce projet est, partant d'une base de données publique récente, de tester différents algorithmes d'apprentissage pour la prédiction numérique de propriétés physico-chimiques quantitatives de petites molécules.

L'article de M. Rupp (cf. bibliographie) donne une bonne introduction au sujet; la méthode particulière qu'on y trouve exposée pourra être appliquée aux données produites par Ramakrishnan, Dral, Rupp et von Lilienfeld. L'étudiant pourra s'initier à Python et à sa bibliothèque d'outils pour le machine learning (Scikit) sur le cours en ligne du CNAM. S'initier aux bases de la DFT afin de mieux comprendre les tenants et aboutissants du projet et de pouvoir aller plus loin (en stage) serait un plus, si le temps le permet. Une connaissance minimale des principes de base de la mécanique quantique est souhaitée.

¹LPCT (Labo Physique et Chimie théorique), Nancy, Université de Lorraine

²IECL (Institut Elie Cartan de Lorraine), laboratoire de recherches en mathématiques, Nancy

Bibliographie:

- Cours en ligne du CNAM: Python et machine learning en Python avec Scikit,
<http://cedric.cnam.fr/vertigo/Cours/ml/tpIntroductionPython.html>
<http://cedric.cnam.fr/vertigo/Cours/ml2/tpArbresDecision.html>
- The Harvard Clean Energy Project, <http://nrs.harvard.edu/urn-3:HUL.InstRepos:13478847>
- N. W. Ashcroft, N. D. Mermin. *Solid state physics*, Saunders College Publishing (1976).
- J. Nelson. *The physics of solar cells*, Imperial College Press (2003).
- R. Ramakrishnan, P. Dral, M. Rupp, O. Anatole von Lilienfeld. *Quantum chemistry structures and properties of 134 kilo molecules*, Scientific Data **1** (2014), open data consultables sur <http://www.nature.com.bases-doc.univ-lorraine.fr/articles/sdata201422>
- C. E. Rasmussen, C. K. I. Williams, *Gaussian processes for machine learning*, MIT Press (2006).
- Numéro de novembre 2017 du magazine "La Recherche", *L'intelligence artificielle transforme les sciences*, <https://www.larecherche.fr/parution/mensuel-529>
- M. Rupp. *Machine learning for quantum mechanics in a nutshell*, International Journal of Quantum Chemistry **115** 1058—1073 (2015).
- C. A. Ullrich. *Time-dependent density functional theory*, Oxford University Press (2012).