PROGRAMME INSTITUTS ET INITIATIVES

Appel à projet – campagne 2022

Proposition de projet de recherche doctoral (PRD)

SCAI - Sorbonne Center of Artificial Intelligence

Intitulé du projet de recherche doctoral (PRD):

Speeding-up quantum chemistry: Fast evaluation of two-electron integrals with Machine Learning

Directeur de thèse porteur du projet (titulaire d'une HDR) :

NOM : **SISOURAT** Prénom : **Nicolas**

Titre : Maître de Conférences des Universités

e-mail : Nicolas.Sisourat@sorbonne-universite.fr

Adresse professionnelle : 4 Place Jussieu, Tour 43-44, 1er étage

Unité de Recherche :

Intitulé : Laboratoire Chimie Physique Matière et Rayonnement

Code: UMR 7614

École Doctorale de rattachement de l'équipe (future école doctorale du.de la doctorant.e) : ED388-ChimiePhysique&ChimieAnalytique ParisCentre

Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse: 1

Co-encadrant :

NOM : THOME Prénom : Nicolas

Titre : Professeur des Universités HDR ☒

e-mail : Nicolas.thome@cnam.fr

Unité de Recherche : CEDRIC

Intitulé : Centre d'études et de recherche en informatique et communications

Code: EA 4629

École Doctorale de rattachement : Choisissez un élément :

Ou si ED non Alliance SU :

Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse : 3

Cotutelle internationale : ☒ Non ☐ Oui, précisez Pays et Université :

Selon vous, ce projet est-il susceptible d'intéresser une autre Initiative ou un autre Institut ?
☒ Non ☐ Oui, précisez Choisissez l'institut ou l'initiative :

Description du projet de recherche doctoral *:*

Introduction, scientific context and objectives

Quantum chemistry (QC) is the application of quantum mechanics to compute the physicochemical properties of atomic and molecular systems. There are various QC methods, which have been efficiently implemented in numerous QC packages [1]. In most QC approaches, the electronic wave function is expressed in terms of Slater determinants. Computing atomic and molecular properties requires the evaluation of one- and two-electron integrals over these determinants. A common bottleneck of QC approaches is computing and handling an enormous number of these two-electron integrals. Therefore, there is a continuous search for always faster methods to compute these integrals (see [2] for a review on standard techniques that have been developed and implemented for the evaluation of these integrals).

The aim of the proposed project is to test and employ state-of-the-art machine learning (ML) algorithms to compute two-electron integrals. The ML approaches should speed-up the calculations of many QC methods by several orders of magnitude, allowing to treat atomic and molecular systems of large size and with an accuracy beyond what is currently possible. Since QC is an essential tool in many disciplines, including physics, chemistry, biology, materials science and engineering, the proposed project should have a significant and very broad impact.

Several approximations have been developed to reduce the computational costs of evaluating these two-electron integrals (see e.g. [3] and references therein). The use of these approximations allows to consider atomic and molecular systems that would be out of reach otherwise, but, obviously decreases the accuracy in predicting the properties of these systems. In our project, we will employ these approximations and correct them using advanced ML techniques (see next section for details) in order to obtain both efficiency and accuracy.

It should be mentioned that ML techniques are already employed in QC (see the recent special issues [4,5] for an overview of how data-driven methodologies are employed in QC). However, in general ML techniques are used to predict one specific property (e.g. the energy of the system). One has therefore to employ several techniques and run numerous calculations to obtain a more detailed description of the atomic and molecular systems. With our proposed approach, one will be able to obtain the electronic wave function and thus all physicochemical properties within the same calculations.

SORBONNE UNIVERSITÉ

Research methodology and PhD project

In this project we will rely on approximated QC models to predict two-electron integrals for large molecular systems and to augment them with a data driven component. As a starting point these integrals will be approximated with the Schwarz bound [6]: this model provides estimates of the complete set of integrals using only a small number of integrals that can be efficiently computed and stored. The PhD student will explore how ML techniques can correct the QC models. As a starting point, they will adapt the method in [7] to complete the simplified models and thus to accurately and efficiently compute these two-electron integrals. More approximate QC models (as in [3]) and ML techniques will then be tested to improve further the efficiency.

The PI (Nicolas SISOURAT) team has already developed and trained neural networks (NNs) to predict two-electron integrals of small molecular systems. As shown in Fig. 1, NNs can provide accurate integrals. The evaluation of these integrals with NNs is however 100 times faster with NNs than with standard QC techniques. It is therefore possible to compute the integrals on-the-fly, which avoids storing them. Note that in a typical QC calculations, storing the integrals take several hundreds of gigabytes which can represent the limiting factor of the calculations. While these preliminary results demonstrate that ML techniques can be succesfully used to compute two-electron integrals, the NN approach that has been developed is limited to molecular systems of medium size. To go further, more advanced ML techniques are required as proposed in this project.

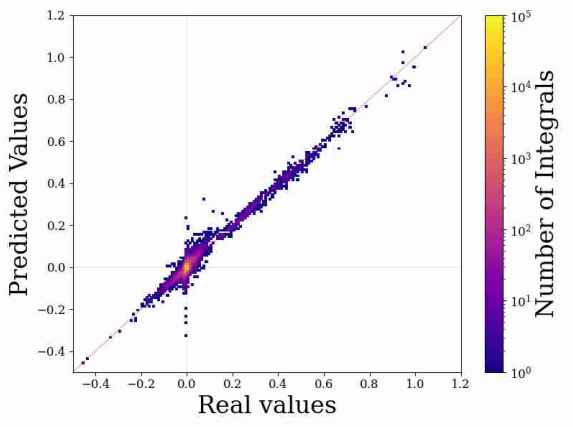

Fig. 1: Two-electron integrals predicted by a NN as functions of the real values of the integrals. The red line indicates an exact prediction. The color code represents the number of integrals in a given range of values.

The co-advisor (Nicolas THOME) has developed the so-called APHYNITY framework [7]. In this approach, an incomplete physical model (as the QC models mentioned above) is augmented with a deep data-driven component accounting for errors of the physical model. APHYNITY has proven to efficiently correct approximate physical models on several simple use cases. The PhD student will first adapt this method for the fast evaluation of two-electron integrals. They will then explore other related ML techniques.

Originality and innovative aspects of the proposal

As mentioned above, QC plays a central role in nearly every fields of natural science: from

physics to engineering. QC can be used to predict properties, help the interpretation of experiments and provide input data for further modeling. Recognizing the growing importance of QC, funding agencies across the world have invested heavily in the development of QC software. Our project will improve significantly the efficiency and accuracy of these codes. While ML techniques are already used in QC for specific tasks, we propose a general solution to one of the major bottlenecks of QC methods.

Nicolas Sisourat has over 10 years of experience in developing and using QC codes. He will supervise the PhD student in the use of the approximate QC models. Nicolas Thome is an expert in ML techniques and the main developer of the APHYNITY framework. He will advise the PhD student to adapt APHYNITY and other ML techniques in order to correct the QC models. To reach the goal of the project a trans-disciplinary approach is mandatory; the complementary of the two teams is therefore a strong asset of the project.

Relevance of the project for SCAI

The proposed project is essentialy interdisciplinary and brings together two communities, namely the QC and AI ones. Furthermore, the results of our project will significantly contribute to a broad range of scientific applications. Our project will therefore federates communities and promotes AI to new challenges, which are the main goals of SCAI.

References

[1] C. D. Sherrill, D. E. Manolopoulos, T. J. Martinez, and A. Michaelides; Electronic structure software, J. Chem. Phys. **153**, 070401 (2020).

[2] S. Reine, T. Helgaker and R.Lindh; Multi-electron integrals, WIREs Comput. Mol. Sci. **2,** 290 (2012).

[3] T. Yamada, R. P. Brewster and S. Hirata; Asymptotic expansion of two-electron integrals and its application to Coulomb and exchange lattice sums in metallic, semimetallic, and nonmetallic crystals. J. Chem. Phys. **139**, 184107 (2013).

[4] M. Ceriotti, C. Clementi and O. A. von Lilienfeld; Machine learning meets chemical physics, J. Chem. Phys. **154**, 160401 (2021)

[5] Machine Learning in Physical Chemistry; https://pubs.acs.org/page/vi/machine-learning

[6] J. L. Whitten; Coulombic potential energy integrals and approximations, J. Chem. Phys. **58**, 4496 (1973).

[7] Y. Yin, V. L. Guen, J. Dona, E. de Bézenac, I. Ayed, N. Thome, and P. Gallinari; Augmenting physical models with deep networks for complex dynamics forecasting, International Conference on Learning Representations (2021).

Publications of the advisors that are relevant for the present proposal

B. Casier, S. Carniato, T. Miteva, N. Capron, N. Sisourat; *Using principal component analysis for neural network high-dimensional potential energy surface*, J. Chem. Phys. **152**, 234103 (2020).

https://github.com/sisourat/qp2_plugins_nsisourat

V. Le Guen, Y. Yin, J. Dona, I. Ayed, E. De Bézenac, N. Thome, P. Gallinari; *Augmenting*

Physical Models with Deep Networks for Complex Dynamics Forecasting, Journal of Statistical Mechanics: Theory and Experiment **12 (**2021). : http://cedric.cnam.fr/~thomen/papers/APHYNITY_JSTAT.pdf

V. Le Guen, N. Thome; *Disentangling Physical Dynamics from Unknown Factors for Unsupervised Video Prediction,* CVPR (2020) : http://cedric.cnam.fr/~thomen/papers/PhyDNet-CVPR20.pdf

Profile of the PhD Applicant

The PhD applicant should have a master degree in data science or related. She/He should have good competence in programming and numerical methods. Knowledge in theoretical chemistry approaches will be an asset.