SORBONNE UNIVERSITÉ

PROGRAMME INTITUTS ET INITIATIVES

Appel à projet – campagne 2021

Proposition de projet de recherche doctoral (PRD)

SCAI - Sorbonne Center of Artificial Intelligence

Intitulé du projet de recherche doctoral (PRD):
Nested Sampling for Nuclear Quantum Effects (QNEST)

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

NOM : **TRASSINELLI** Prénom : **Martino**
Titre : Chargé de Recherche ou
e-mail : martino.trassinelli@insp.jussieu.fr
Adresse professionnelle : Campus Pierre et Marie Curie, tour 22, couloir 22-12, porte 315
(site, adresse, bât., bureau)

Unité de Recherche :
Intitulé : Institut des NaoSciences de Paris
Code *(ex. UMR xxxx)* : UMR7588

École Doctorale de rattachement de l'équipe (future école doctorale du.de la doctorant.e) : **ED564-Physique en IdF**

Doctorant.e.s actuellement encadré.e.s par la.e directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) : 1: Perla Dergham, debout de thèse: octobre 2020, direction avec co-encadrement à 50%

Co-encadrant.e :

NOM : **FINOCCHI** Prénom : **Fabio**
Titre : Directeur de Recherche ou HDR ☒
e-mail : fabio.finocchi@insp.jussieu.fr

Unité de Recherche :
Intitulé : Institut des NaoSciences de Paris
Code *(ex. UMR xxxx)* : UMR7588

École Doctorale de rattachement : **ED397-Physique Chimie des Matériaux**
Ou si ED non Alliance SU :

Doctorant.e.s actuellement encadré.e.s par la.e co-directeur.rice de thèse (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) : 2: Erika Fallacara, debout de thèse

octobre 2019 en co-direction à 50%, Niccolò Avallone, debout de thèse janvier 2021 en co-direction à 50%.

Co-encadrant.e :

NOM : Prénom :
Titre : Choisissez un élément : ou HDR ☐
e-mail :

Unité de Recherche :
Intitulé :
Code *(ex. UMR xxxx)* :

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 (préciser le nombre de doctorant.e.s, leur année de 1e inscription et la quotité d'encadrement) :

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 *(en français ou en anglais) :*

Ce texte sera diffusé en ligne : il ne doit pas excéder 3 pages et est écrit en interligne simple.

Détailler le contexte, l'objectif scientifique, la justification de l'approche scientifique ainsi que l'adéquation à l'initiative/l'Institut.

Le cas échéant, préciser le rôle de chaque encadrant ainsi que les compétences scientifiques apportées. Indiquer les publications/productions des encadrants en lien avec le projet.
Préciser le profil d'étudiant(e) recherché.

State of the art and aim of the proposal
Although much heavier than electrons, light nuclei, mainly hydrogen, exhibit quantization of the vibrational levels, zero-point energy and tunneling. These so-called Nuclear Quantum Effects can have a large impact on the structure and the dynamics of materials. Here, we focus on systems with hydrogen bonds, as well as on materials in which protons or hydrogen can diffuse, such as solid fuel cells, which are relevant for energy harvesting and batteries. Proton diffusion is the key to transport phenomena which thus can be significantly impacted by Nuclear Quantum Effects [1,2,3] (tunneling in particular). The standard method to account for these effects when simulating the static properties at equilibrium is path integrals: the quantum partition function of a nucleus can be computed from the classical one for a polymer consisting of P beads (typically, $16 \leq P \leq 64$ for H at room temperature). As a counterpart, this method considerably increases the number of degrees of freedom, making the exploration of the potential energy landscape much harder. Moreover, a bottleneck of path-integral based methods regards the difficulty in evaluating dynamical properties. In principle, all relevant statistical properties can be derived from the partition function of the system. This requires a detailed knowledge of the low-energy part of the quantum potential energy surface, as given by path integrals, which can be quite different from the classical counterpart. Presently, few groups worldwide are working on the calculation of the quantum free energy by advanced sampling methods that are largely inspired from statistical mechanics.
Here, we aim at systematically exploring the quantum potential energy surface topology by combining path-integrals based techniques with the nested sampling exploration method coupled to machine learning algorithms and interpolations based on artificial neural networks.
The exploration will be performed by smart sampling of the quantum potential energy, by reducing as much as possible the number of the sampling points. Potential energy surfaces of real systems usually display several local minima and high dimensionality, which makes most sampling algorithms converge with difficulty. In this respect, our strategy will be twofold: first, we will reduce the number of degrees of freedom, by determining a small dimension base to represent the energy surfaces. We will employ principal components analysis and recent algorithms, such as the greedy-type approaches, like empirical interpolation [4]. The a-posteriori verification of the reduction effectiveness will be conducted via the Bayesian evidence, by comparing the model predictions and the full calculations. Second, on the reduced models we will apply nested sampling, which turns the multi-dimensional problem into a one-dimensional integral, thus considerably reducing the computational needs for sampling. The recognition via machine learning methods of cluster structures of the sampling points, corresponding to the function minima, will allow to focus on specific regions that mostly contribute to the energy landscape. Last but not the least, the sampling points that are obtained by nested fit will be used as input for an artificial neural network to

improve the accuracy on the description of the potential energy for a given number of evaluations. This two-step strategy is compulsory as the single energy (and gradient) evaluation is computationally cumbersome: the reduction of the number of degrees of freedom is unavoidable in real systems, and allows extracting the significant physical and chemical parameters.

The method and the algorithms

The ability of nested sampling to reduce the exploration of the multi-dimensional parameter space to one-dimensional integrals relies on the construction of an evolving set of sampling points called live points. For each step, a new sampling point is found, which by construction corresponds to a lower value of the object function. This procedure corresponds to search in a volume whose extension depends on the function value, which is a subset (nested) of the region that was sampled in the previous steps. This results in an integration with respect to the volume of the nested region only. Because of such drastic reduction of dimensions, nested sampling can outperform other integration methods, such as Markov-chain Monte Carlo. Moreover, it correctly probes the object function in critical cases where the integration domain changes dramatically with the value of the function, as in the case of first-order transitions [5,6]. However, in the presence of several local minima, the recursive search of new live points becomes inefficient and the algorithm hardly converges, possibly leaving un-explored regions out of the integration domain. In solid-state physics, global optimization and multi-particle exploration have been employed to study partition functions and potential energy surfaces. In data analysis applications of nested sampling, cluster recognition methods have been implemented by one of the promoters, via the mean shift method [7].

During the thesis, a new version of NestedFit (a nested sampling code developed by M. Trassinelli [7,8]) will be developed and interfaced to the path-integral code for specific cases (developed by F. Finocchi and collaborators [1,2,3]). Several machine learning methods using unsupervised algorithms will be also implemented and tested for an efficient cluster recognition of the nested sampling live points. In particular, we will use the mean shift approach, as well as other methods (k-nearest neighbor, x-means, DBSCAN, spectral clustering, etc.). Particularly promising is the conjunction of artificial neural network methods and nested sampling, which has been successfully employed for data analysis of gravitational waves [9]. A same accuracy of the likelihood function integral is obtained using 2-10 times fewer sampling points. In this project, we aim at applying the same strategy for the evaluation of the partition function; we remind that each energy and force computation implies the time-consuming solution of a Schrödinger-like equation. The Nested sampling algorithm, coupled to cluster recognition strategies, allows a drastic reduction of the number of calls to the Schrödinger solver. An additional reduction will be obtained with the use of the neural network, using as inputs the sampling point obtained from a crude nested sampling exploration.

The Fortran programming language will be preferred for its high computation speed, with calculation modules called by a Python handler. Due to the computationally intense path-integral simulations, we plan to make use of parallel algorithms, both when computing the object function and in the evolution of the live points.

In parallel to the characterization of the potential energy from the ab initio calculations, a modeling of the complex landscape will be carried out for specific problems that are described in the next section. The very first step will be based on the reduction of the system degrees of freedom by discrete empirical interpolation methods [10]. The identification of the more adapted model will be obtained by comparing the real data (that are extracted from the Schrödinger solver) and the model predictions via two complementary approaches: (I) the computation of Bayesian evidence (i.e., the marginal likelihood) and (II) model error estimates from greedy-type algorithms. Concerning (I), model predictions are compared to Schrödinger-solver data, to evaluate probability of the model itself from the computation of the Bayesian evidence with the NestedFit program. Concerning (II), the fitting steps in greedy algorithms consist in adding parameters iteratively to the current model to correct the error (which is here analogous to the difference between the model and the real data), while keeping the number of parameters as low as possible by discriminatory steps [11].

Benchmarks and Applications

The first developments and tests will be performed on Lennard-Jones (LJ) clusters with classical

nuclei. This system was the object of the first implementation of nested sampling on potential energy surfaces exploration [5] and represents a tough benchmark. In a second stage, we will include quantum effects in LJ clusters using the path integrals formalism, in analogy with a recent work employing nested sampling on simplified molecular models [12]. After extensive tests on LJ clusters, the developed methods and tools will be applied to one or several cases discussed below, depending on the expected specific speed-up.
- Water clusters: We will explore the quantum potential energy surface of (H2O)n water clusters (n≤8). The zero-point energy in water molecules is much larger than the thermal energy, and nuclear quantum effects are particularly significant in finite (H2O)n clusters due to the reduced connectivity with respect to bulk water.
- H diffusion: If successful, we will use the developed methods in bulk materials with hydrogen bonds, where protons can hop. As these bonds are generally described via the computationally demanding Density Functional Theory, the use of efficient sampling methods is crucial. Hydrated materials of geophysical interest for their capability to transport water in Earth's mantle such as AlOOH [1] or Mg(OH)2 [2] display yet not entirely understood quantum-driven phase transitions or protonic diffusion processes. In analogy with hydroxides under high pressures, proton transfer in perovskites (useful materials for solid fuel cells) is governed by distinct mechanisms below room temperature and ambient pressure.
- Analysis of atomic spectroscopy data: In addition to nuclear quantum effects studies, we plan to apply the developed codes for analyzing complex experimental data such as crystal high-resolution spectra of highly charged ions. To reach a very high-accuracy, the response function of the employed crystal X-ray spectrometer has to be modeled from simulations [8,13]. The poor statistics and possible presence of unresolved spectral components generally generate several maxima in the likelihood function. The study of its maxima will be performed using the same strategy than for the exploration of potential energy surfaces.

The Ph.D. student will benefit from a very open scientific and interdisciplinary environment. This project brings together the expertise of two physicists who, despite being in the same lab, have two distinct research domains (theory on solid state physics and Bayesian methods for atomic spectroscopy). He/She will benefit in addition from the collaboration with Julien Salomon at INRIA for his mathematical skills in model selection algorithms. We also expect to benefit from the SCAI environment, exchanging with researchers in applied mathematics and statistical learning. Their expertise could be crucial for choosing the most adapted machine learning methods and for the implementation of neural network for our specific problems, as there are only a couple of previous works that combined machine learning with nested sampling strategies. The ideal candidate has a background in quantum and statistical physics, applied mathematics and good programming skills.

[1] Y. Bronstein, P. Depondt and F. Finocchi, Thermal and nuclear quantum effects in the hydrogen bond dynamical symmetrization phase transition of δ-AlOOH, Eur. J. Mineral. 29 (2017), 1.
[2] S. Schaack, P. Depondt, M. Moog, F. Pietrucci, and F. Finocchi, How methane hydrate recovers at very highpressure the hexagonal ice structure, J. Chem. Phys. 152 (2020), 024504.
[3] S. Schaack, P. Depondt, S. Huppert, and F. Finocchi, Quantum driven proton diffusion in brucite-like mineralsunder high pressure, Sci. Rep. 10 8123 (2020).
[4] M. Barrault et al., An 'empirical interpolation' method: application to efficient reduced-basis discretization of partial differential equations, Comptes Rendus Mathematique 339 (2004), 667–672.
[5] J. Skilling, Nested sampling for general Bayesian computation, Bayesian Anal. 1 (2006), 833–859.
[6] L. B. Partay, A. P. Bartok, and G. Csanyi, Efficient sampling of atomic configurational spaces, J. Phys. Chem. B 114 (2010), 10502–10512.
[7] M. Trassinelli and P. Ciccodicola, Mean shift cluster recognition method implementation in the nested sampling algorithm, Entropy 22 (2020), 185.
[8] M. Trassinelli, Bayesian data analysis tools for atomic physics, Nucl. Instrum. Methods B 408 (2017), 301.
[9] P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby, BAMBI: blind accelerated multimodal Bayesian

inference, Mon. Not. R. Astron. Soc. 421 (2012), 169–180.
[10] S. Chaturantabut and D. C. Sorensen, Nonlinear model reduction via discrete empirical interpolation, SIAM J. Sci. Comput. 32 (2010), 2737–2764.
[11] Y. Maday and J. Salomon, A greedy algorithm for the identification of quantum systems, in Proceedings of the 48h IEEE Conference on Decision and Control (CDC) 375–379 (2009).
[12] B. Szekeres, L. B. Partay, and E. Matyus, Direct computation of the quantum partition function by path-integral nested sampling, J. Chem. Th. Comp. 14 (2018), 4353–4359.
[13] J. Machado, G. Bian, N. Paul, M. Trassinelli et al., Reference-free measurements of the 1s2s2p 2P1/2,3/2→1s2 2s 2S1/2 and 1s2s2p 4P5/2→1s2 2s 2S1/2 transition energies and widths in lithiumlike sulfur and argon ions, Phys. Rev. A 101 (2020), 062505.

Merci d'enregistrer votre fichier au <u>format PDF</u> et de le nommer :
«ACRONYME de l'initiative/institut – AAP 2021 – NOM Porteur.euse Projet »

Fichier envoyer simultanément par e-mail à l'ED de rattachement et au programme :
cd_instituts_et_initiatives@listes.upmc.fr avant le 20 février.