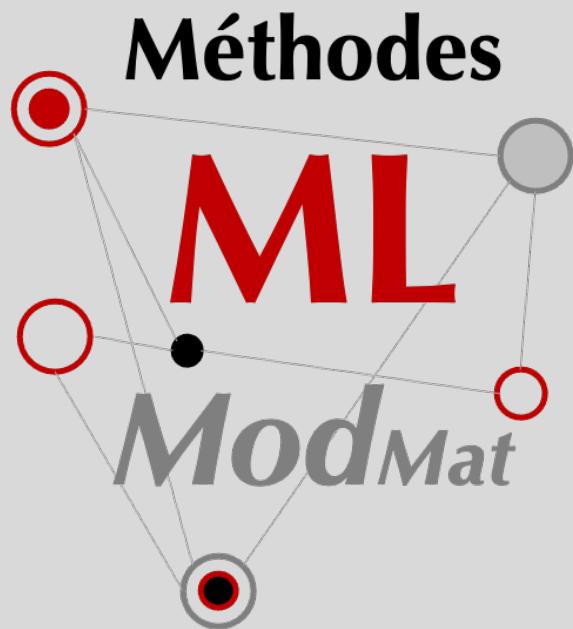


ATELIER DU RESEAU *ModMat*

METHODES MACHINE-LEARNING POUR LA MODELISATION DES MATERIAUX



Programme et Résumés

AU LAAS-CNRS, TOULOUSE
DU 22 AU 24 SEPTEMBRE 2021

INTRODUCTION

Résultat de l'avènement de nouvelles ressources informatiques toujours plus performantes et de la disponibilité de tous les calculs de structures électroniques précédemment menés, les méthodes machine-learning pour la modélisation des matériaux ont grandement gagné en popularité ces dernières années. D'une part, celles-ci ont ouvert des possibilités pour la construction de nouveaux potentiels d'interactions dont l'élaboration et l'utilisation sont peu coûteuses tout en retenant la précision des calculs à l'échelle électronique. D'autre part, l'analyse complexe de certaines structures notamment dans les matériaux cristallins ou les matériaux amorphes a aussi énormément bénéficié des apports des méthodes machine-learning.

Dans ce contexte, nous organisons cet atelier intitulé "Méthodes machine-learning pour la modélisation des matériaux" afin de regrouper à la fois les développeurs de ces méthodes et les chercheurs avides de les utiliser. Cet atelier se déroule sur 3 jours (2 jours complets) du 22 au 24 Septembre 2021 au LAAS-CNRS de Toulouse. L'objectif est de mieux faire connaître les différentes méthodes, partager les bonnes pratiques et les inconvénients liés à leur utilisation et finalement d'initier de nouvelles collaborations.

Dans le contexte sanitaire actuel, ces journées se tiennent selon un format dual : en présentiel dans la limite de 50 personnes (sur inscription), et en distancée par un système de streaming ou de visioconférence, ce qui assure la tenue de cet atelier.

Des sessions scientifiques (oraux et posters) sont planifiées, en mettant l'accent sur la participation des jeunes chercheurs et des doctorants, cet atelier devant être une occasion pour eux de créer des liens avec la communauté et de présenter leurs travaux.

Au plaisir de vous (re)voir à Toulouse,

Le Comité d'organisation

Anne HEMERYCK (LAAS-CNRS), Julien LAM (CEMES-CNRS), Céline MERLET (CIRIMAT), Magali BENOIT (CEMES), Marie BRUT (LAAS-CNRS), Romuald POTEAU (LPCNO-INSA Toulouse), Ruggero LOT (LAAS-CNRS).

Atelier « Méthodes machine-learning pour la modélisation des matériaux »

SPONSORS



CECAM-FR-GSO



PROGRAMME

SESSION 1 - Mercredi 22 septembre après midi

13h00-13h30	Introduction
13h30-14h10	Oratrice invitée : Clémence CORMINBOEUF (visio), EPFL, Lausanne Data-driven Approaches to the Discovery and Classification of Homogeneous Catalysts
14h10-14h40	O1- F.X. COUDERT (visio), IRCP, Paris Speeding Up the Discovery of Framework Materials with Targeted Properties by Machine Learning
14h40-15h10	O2- N. MINGO , CEA-LITEN, Grenoble Modeling oxide mixed phases via machine-learning enhanced ab initio calculations
15h10-15h30	<i>PAUSE café</i>
15h30-16h00	O3- B. CASIER , LPCT, Nancy Hybrid localized graph kernel for machine learning energy-related properties of molecules and solids
16h00-16h30	O4- G. FRAPPER , IC2MP, Poitiers Algorithme évolutionnaire auto-apprenant pour la prédiction de nouveaux matériaux : USPEX & ses applications
16h30-17h00	O5- G.M. DONGHO-NGUIMDO , IC2MP, Poitiers Prediction of novel Yttrium Nitride phases under pressure

SESSION 2 – Jeudi 23 septembre matin

9h-9h40	Orateur invité : Frank SMALLENBURG , LPS, Orsay Linking local structure and dynamics in glassy fluids using machine learning
9h40-10h10	O6- J. LAM , CEMES-CNRS, Toulouse Zinc oxide crystallization modeled with machine-learning interatomic potential
10h10–10h40	O7- R. LOT , LAAS-CNRS, Toulouse Evaluation of Neural Networks methods to model interface phenomena in solid phase epitaxy for Silicon
10h40-11h00	<i>PAUSE café</i>
11h00-11h30	O8- F. BRIX , IJL, Nancy Adsorption energies for atomic oxygen on the Al13Co4(100) surface: a machine learning approach
11h30-12h00	O9- M. CHAGAS DA SILVA , LPCT, Nancy Ab initio calculations of finite-temperature properties at multiple electronic structure levels made affordable: An effective combination of perturbation theory and machine learning
12h00-12h30	Intervention IDRIS - T. VERY , Orsay – (20 min) Flash session POSTER

12h30-14h00	Buffet et SESSION POSTERS
-------------	---------------------------

SESSION 3 – Jeudi 23 septembre après midi

14h00-14h40	Oratrice invitée : Geneviève DUSSON , LMB, Besançon An Atomic Cluster Expansion (ACE) to fit interatomic potentials with polynomials
14h40-15h10	O10- D. BISSUEL (visio), ILM, Lyon Machine-learned repulsive potentials for the Density Functional based Tight-Binding method
15h10-15h30	<i>PAUSE café et SESSION POSTERS</i>
15h30-16h00	O11- SH. LEE , CEA-LETI, Grenoble Neural Network Potential for Fast and Accurate Molecular Dynamics Simulations of SiGe
16h00-16h30	O12- A. CASTELLANO , CEA-DAM-DIF, Bruyères-Le-Châtel Machine-Learning Assisted Canonical Sampling
16h30-17h00	O13- P. GRIGOREV , CINaM, Marseille Synergistic coupling in QM/MM simulations of dislocations via machine learning

SESSION 4 - Vendredi 24 septembre matin

9h-9h40	Orateur invité : Mihai Cosmin MARINICA , CEA-SRMP, Gif-sur-Yvette Machine Learning for Atomistic Materials Science
9h40-10h10	O14- C. LAPOINTE , CEA-SRMP, Gif-sur-Yvette Modèles de régression de l'entropie vibrationnelle de formation et de migration de défauts ponctuels
10h10–10h30	O15- TD. SWINBURNE , CINaM, Marseille Learning how dense dislocation networks evolve
10h30-11h00	<i>PAUSE café</i>
11h00-11h30	O16- AM. GORYAEVA , CEA-SRMP, Gif-sur-Yvette Reinforcing materials modelling by encoding the structures of defects in crystalline solids into distortion scores
11h30-12h00	O17- MP. GAIGEOT (visio), LAMBE, Evry Graph Theory for Molecular Dynamics simulations
12h00-12h30	O18- G.D. FORSTER (visio), ICMN, Orléans Approche par deep learning pour déterminer les indices de chiralité des nanotubes de carbone à partir d'images HRTEM

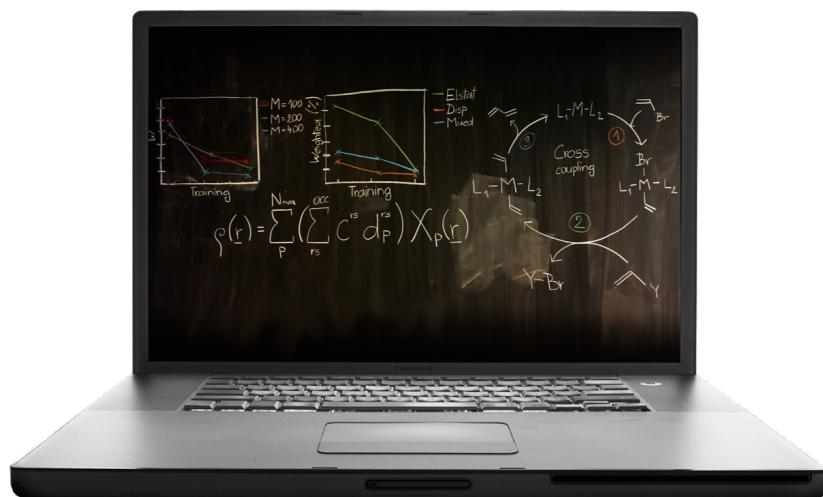
SESSION 1

I1. Data-driven Approaches to the Discovery and Classification of Homogeneous Catalysts

Clémence Corminboeuf¹

¹Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

With the goal of optimizing reactions, we exploit statistical learning techniques in tandem with quantum chemistry data to identify improved homogeneous catalysts. The appeal of machine learning (ML) is quite clear; the computational speed up provided allows the properties and energetics of thousands of prospective catalysts to be rapidly assessed while preserving quantum chemical accuracy [1]. The massive quantity of data generated can be compiled and mined into an interactive tool, which facilitates the analysis and assists in identifying the most compatible metal/ligand family combinations with the goal of establishing relationships between the intrinsic chemical properties of different catalysts and their overall catalytic performance. We focus on both prototypical classes reactions and challenging organic processes or target properties.



Références

- [1] Matthew D. Wodrich, Boodsrin Sawatlon, Michael Busch, Clemence Corminboeuf, Acc. Chem. Res. 2021, 54, 1107

O1. Speeding Up the Discovery of Frameworks Materials with Targeted Properties by Machine Learning

François-Xavier Coudert¹

¹ Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, Paris, France

The characterization of the electronic, mechanical and thermal properties of crystalline materials are nowadays considered a routine computational task in DFT calculations. However, their high computational cost still prevents them from being used in high-throughput screening methodologies. In my group, we have investigated various multi-scale simulations strategies, including quantum chemistry, classical molecular dynamics, and learning methods, in order to predict the physical properties of nanoporous materials.^{1,2}

Recently, we analyzed the behavior of about 600 000 hypothetical zeolitic structures at the classical level (a scale 3 orders of magnitude larger than previous studies), to highlight generic trends between mechanical properties and energetic stability.³ By comparing these results with DFT calculations on 991 zeolitic frameworks, we highlight the limitations of force field predictions, in particular for predicting auxeticity. We then used this reference DFT data as a training set for a machine learning algorithm, showing that it offers a way to build fast and reliable predictive models for anisotropic properties. The accuracies obtained are, in particular, much better than the current “cheap” approach for screening, which is the use of force fields. These results are a significant improvement over the previous work, due to the more difficult nature of the properties studied, namely, the anisotropic elastic response. It is also the first time such a large training data set is used for zeolitic materials.

Finally, we are working with other groups in the field in an attempt to elaborate, publicize and discuss some best practices for machine learning in chemical sciences.⁴

Mots-clés : multi-scale modeling, mechanical properties, thermal behavior, nanoporous materials

Références

¹ J. D. Evans and F.-X. Coudert, **Chem. Mater.**, 29, 7833–7839 (2017)

² S. Chibani and F.-X. Coudert, **APL Mater.**, 8, 080701 (2020)

³ R. Gaillac, S. Chibani and F.-X. Coudert, **Chem. Mater.**, 32, 2653–2663 (2020)

⁴ N. Artrith, K. T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain and A. Walsh, **Nature Chem.**, in press (2021), DOI: 10.1038/s41557-021-00716-z

O2. Modeling oxide mixed phases via machine-learning enhanced ab initio calculations

Quintin Meier¹, A. van Roekeghem¹, N. Mingo¹

¹ CEA-Grenoble

Mixed oxides containing several cation species are main ingredients for energy storage technologies, like Li-ion batteries, or thermochemical storage in concentrated solar power plants. To reach the 2050 worldwide zero-emissions target we need to improve the technologies in terms of capacity, durability, and cost, via tailored materials. Here, predictive computations can enable a more efficient search for structures and compositions with desirable properties. However, the inherent complexity of mixed oxides means that, to predict their thermodynamic and transport properties, one must go beyond standard DFT approaches in order to tackle the exceedingly large configurational space. We use moment tensor potentials to investigate the energetics and dynamics of oxide materials, and relate them to macroscopic properties. We will discuss the prediction of solubility gaps, finite temperature phonon densities, phase transition temperatures, and kinetics.

Mots clés : moment tensor potentials, oxides, phonons.

Références

¹ S. K. Wallace, A. S. Bochkarev, A. van Roekeghem, J. Carrasco, A. Shapeev, and N. Mingo, Phys. Rev. Materials, 5, 035402 (2021).

² S. K. Wallace, A. van Roekeghem, A. S. Bochkarev, J. Carrasco, A. Shapeev, and N. Mingo, Phys. Rev. Research, 3, 013139 (2021).

O3. Hybrid localized graph kernel for machine learning energy-related properties of molecules and solids

Bastien Casier¹, Mauricio Chagas Silva¹, Michael Badawi¹, Fabien Pascale¹, Tomáš Bučko^{2,3}, Sébastien Lebègue¹, Dario Rocca¹

¹ Université de Lorraine and CNRS, LPCT, UMR 7019, F-54000 Nancy, France

² Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Mlynská Dolina, Ilkovičova 6, SK-84215 Bratislava, Slovakia

³ Institute of Inorganic Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84236 Bratislava, Slovakia

Nowadays, the coupling of electronic structure and machine learning techniques serves as a powerful tool to predict chemical and physical properties of a broad range of systems ¹. With the aim of improving the accuracy of predictions, a large number of representations for molecules and solids for machine learning applications has been developed ²⁻⁴. In this work we propose a novel descriptor based on the notion of molecular graph. While graphs are largely employed in classification problems in cheminformatics or bioinformatics, they are not often used in regression problem, especially of energy-related properties. Our method is based on a local decomposition of atomic environments and on the hybridization of two kernel functions: a graph kernel contribution that describes the chemical pattern and a Coulomb label contribution that encodes finer details of the local geometry. The accuracy of this new kernel method in energy predictions of molecular and condensed phase systems is demonstrated by considering the popular QM7 and BA10 datasets ^{3,5}. These examples show that the hybrid localized graph kernel outperforms traditional approaches such as, for example, the smooth overlap of atomic positions ² (SOAP) and the Coulomb matrices ³.

Key-Words: energy-related properties, graph kernel, machine learning, QM7 and BA10 datasets, regression

References

- ¹ Noé, F. et al., *Annu. Rev. Phys. Chem.* **71**, 361 (2020).
- ² De, S. et al., *Phys. Chem. Chem. Phys.* **18**, 13754 (2016).
- ³ Rupp, M. et al., *Phys. Rev. Lett.* **108**, 058301 (2012).
- ⁴ Casier, B. et al., *J. Comput. Chem.*, 1-12 (2021).
- ⁵ Nyshadham, C. et al., *Npj Comput. Mater.* **5** (2019).

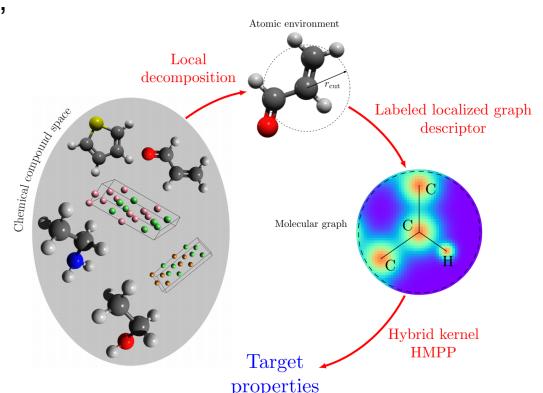


Figure 1. HMPP kernel

O4. Algorithme évolutionnaire auto-apprenant pour la prédiction de nouveaux matériaux : USPEX & ses applications

Gilles Frapper¹

¹Applied Quantum Chemistry group, E4 team, IC2MP UMR 7285, Université de Poitiers – CNRS, 4 rue Michel Brunet TSA 51106 - 86073 Poitiers Cedex 9, France

La détermination par simulation numérique de la structure cristalline d'un composé à partir de l'unique connaissance de sa composition chimique est un défi majeur en Sciences de la Matière. En effet, l'exercice est loin d'être trivial : il s'agit d'identifier les arrangements structuraux de plus basse énergie, donc les plus stables thermodynamiquement, parmi des millions de structures possibles. Pour illustrer ce défi, l'arrangement de vingt atomes dans une boîte –maille de répétition à la forme et au volume variables– peut générer *a priori* plus de 10^{21} structures potentielles localisées sur l'(hyper)surface d'énergie potentielle (PES). Imaginons qu'il faille 1h de temps de calcul pour déterminer numériquement l'énergie associée à chaque arrangement structural optimisé, le temps de calcul nécessaire dépasserait l'âge de l'univers.

La problématique est donc la suivante : comment accéder au puits de plus basse énergie (minimum global sur la PES) en monopolisant un minimum de ressource informatique ? Deux stratégies sont développées : l'une relative à l'élaboration d'une méthode efficace et robuste d'exploration de la PES – objet de cet exposé – et l'autre traite de la recherche d'une méthode de calcul d'énergie peu onéreuse mais précise... Ces deux stratégies peuvent s'appuyer sur le large domaine d'application du « *Machine learning - IA* » pour la modélisation des matériaux.

Cet exposé évoquera une méthode auto-apprenante d'exploration de la PES d'un composé cristallin, l'algorithme évolutionnaire implémenté dans le code USPEX. J'exposerai succinctement les fondements conceptuels de cet algorithme CSP basé sur les concepts de la théorie Darwinienne de l'évolution. Puis j'illustrerai son usage par la présentation de quelques résultats récents de travaux menés à l'IC2MP et en collaboration : exploration de diagrammes de phase binaires sous pression (ex. Sn-N¹), prédiction de métaux-pentazolates M(η^5 -N₅)_x de type polymères de coordination², proposition de nouvelles structures bidimensionnelles (ex. FeS₂ 2D³, CuX X=S, Se, Te), et interface USPEX/VASP dans le logiciel ouvert GDIS⁴. L'objet de cette communication orale est d'inciter à travers d'échanges nourris notre communauté à l'usage et au développement de telles approches, sous utilisées à ce jour en France dans le vaste domaine de la « modélisation des matériaux ».

Mots-clés : algorithme évolutionnaire, prédiction de matériaux, haute pression, 2D, CSP crystal structure prediction

Références

- 1 B. Wang *et al* **JPC C** 124/15, 8080 (2020)
- 2 B. Huang *et al* **Chemistry of Materials**, sous presse (2021)
- 3 B. Wang *et al* **JPC Letters** 11/20, 8861-8866 (2020)
- 4 H. Okadome Valencia *et al* **J. of Comput. Chem.** 1-15 (2021)

05. Prediction of novel Yttrium Nitride phases under pressure

GM Dongho-Nguimdo^{1,2}, Gilles Frapper¹

¹ Institut de Chimie des Milieux et des Matériaux de Poitiers, BP27, 4 rue Michel Brunet TSA 51106 - 86073 Poitiers Cedex 9, France

² College of Technology, University of Buea, PO Box 63, Buea, Cameroon

We used ab initio crystal search evolutionary simulations to predict stable structures and compositions in the binary Yttrium-Nitrogen (YxNy) diagram under pressure. We found that the well-studied YN (Fm-3m) structure undergoes a pressure phase transition at 122.5 GPa. The new phase YN polymorph with a P4/nmm remains stable up to 300 GPa. In addition, several Nitrogen-rich stable YxNy with different nitrogen networks emerge upon compression. After checking the thermodynamic, dynamic, and mechanical stability, the atomic and electronic structures are discussed on the basis of advanced orbital and electron density analyses, as well as by simple models (such as Zintl-Klemm, VSEPR, etc.). By analysing the thermal stabilities, we also found that the tripentazolate salt YN15 R3 stable above 40 GPa was quenchable, hence could be used as high energy-density material.

SESSION 2

I2. Linking local structure and dynamics in glassy fluids using machine learning

Frank Smallenburg¹

¹*Université Paris-Saclay, CNRS, Laboratoire de Physique des Solides, 91405 Orsay, France*

When a fluid is cooled down or compressed sufficiently rapidly, it avoids crystallization and its dynamics sharply slow down, until it eventually turns into an amorphous solid: a glass. One way of understanding this dynamical arrest is by considering that close to the glass transition, the particles that form the fluid become strongly caged by the neighbors in their local environment. Many recent studies have explored the possibility of predicting the future dynamics of particles in glassy fluids by looking at the local structure around them. In this talk, I highlight some of our recent work using both unsupervised and supervised machine learning to examine the link between local structure and glassy dynamics. We find that unsupervised learning can be a powerful tool for finding structural variations that turn out to be linked to dynamical heterogeneity. Additionally, we demonstrate how a simple and efficient supervised machine learning method can be used to provide excellent predictions for the future mobility of individual particles in glassy fluids.

O6. Zinc oxide crystallization modeled with machine-learning interatomic potential

Gaëtan Laurens¹, Jacek Goniakowski², Julien Lam³

¹ ILM, Université de Lyon and CNRS, 69622 Villeurbanne Cedex, France

² INSP, CNRS and Sorbonne Université, 4 Place Jussieu, 75005 Paris, France

³ CEMES, CNRS and Université de Toulouse, 29 rue Jeanne Marvig, 31055 Toulouse Cedex, France

Despite ZnO's key technological importance, its crystallization characteristics are currently relatively poorly understood. This is due to the overwhelming computational effort required for dedicated simulations, which precludes direct use of DFT-based methods and requires fine interatomic potentials, capable of correctly accounting for the structural diversity of ZnO. With the goal to tackle ZnO crystallization in an homogeneous liquid phase we have built and validated a robust Machine Learning Interatomic Potential (MLIP) suitable for large-scale simulations (nanosecond-long molecular dynamics of several hundreds of atoms). The potential is based on physically motivated mathematical formulation and a constrained Lasso-Lars method is used for identification and adjustment of its relevant parameters. The training database is composed of DFT-GGA results on a variety of ordered crystalline structures, but also includes disordered and amorphous configurations, such as those obtained in high-temperature molecular dynamics simulations. With the new MLIP, we have performed molecular dynamics simulations on the freezing of bulk ZnO liquid with particular emphasis on the singular role of cooling rate and applied pressure. Using a data-driven structural characterization approach to identify alternative ZnO polymorphs, we were able to access the birth of the initial zinc oxide crystal in the surrounding liquid and to track its behavior during subsequent growth.

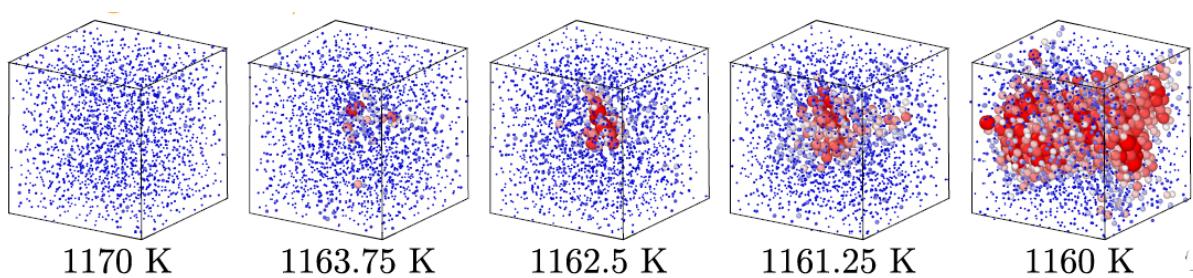


Figure 1. Snapshots along the slow nucleation and crystallization process from liquid ZnO (blue) to a constituted wurtzite crystal (red).

Mots-clés : machine learning, interatomic potential, zinc oxide, crystallization

07. Evaluation of Neural Networks methods to model interface phenomena in solid phase epitaxy for Silicon

Ruggero Lot^{1,2,3}, Layla Martin-Samos³, Anne Hémeryck¹, Stefano de Gironcoli²

¹LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

²CNR-IOM, Democritos National Simulation Center, Istituto Officina dei Materiali, Istituto Officina dei Materiali, c/o SISSA, via Bonomea 265, IT-34136 Trieste, Italy

³SISSA, via Bonomea 265, IT-34136 Trieste, Italy

Simulation of solid-phase epitaxy is a problem that requires an interatomic potential capable of correctly describe a large variety of atomic environments. In this work, we achieve such an accurate description by building a neural network potential [1] for silicon.

A training dataset has been constructed by sampling all the regions of the phase space relevant for the phenomena of interest. The training was carried out starting from a set of different empirical potentials and improved via a self-consistent cycle [2] until the expected accuracy with respect to a validation dataset was met. The Perdew-Burke-Ernzerhof functional was used in the density Functional Theory calculations assumed as ground truth.

The obtained potential has then been tested comparing several additional system properties with experimental data, such as the diamond phonon spectra, the structure factor of amorphous silicon, the surface reconstruction geometries and energies, the melting temperature. For this latter property our result is 1468 +/- 11 K, slightly underestimating the experimental result but in good agreement with Perdew-Burke-Ernzerhof values reported in literature [3].

After thus proving the reliability of our potential we are performing extensive simulations of solid phase epitaxy along the (100) direction at different temperatures in order to unveil the growth mechanisms and speed of the bulk phase. During the simulations, 3D geometric growth motives develop that suggest different growth rates along different directions, in agreement with experimental findings. Building upon these results, we will investigate the growth mechanism in order to identify the key events, and evaluate their rate, to provide a more accurate calibration of off-lattice kinetic Monte Carlo models used to simulate device fabrication with low thermal budget constraints.

References

- [1] R. Lot, F. Pellegrini, Y. Shaidu, and E. Kucukbenli, “PANNA: Properties from artificial neural network architectures” *Comput. Phys. Commun.* vol. 256, p. 107402, Nov. 2020.
- [2] Y. Shaidu, E. Kucukbenli, R. Lot, F. Pellegrini, E. Kaxiras, and S. de Gironcoli, “A systematic approach to generating accurate neuralnetwork potentials: The case of carbon,” *npj Comput Mater*, vol. 7, Apr.2021.
- [3] S. Yoo, S. S. Xantheas, and X. C. Zeng, “The melting temperature of bulk silicon from ab initio molecular dynamics simulations,” *Chem. Phys. Lett.*,vol. 481, pp. 88–90, Oct. 2009

08. Adsorption energies for atomic oxygen on the Al₁₃Co₄(100) surface: a machine learning approach

N. Boulangeot¹, F. Brix¹, T. T. Dorini¹, F. Sur², E. Gaudry¹

¹ Univ. Lorraine, CNRS, Institut Jean Lamour, 2 allée André Guinier, 54000 Nancy, France

² LORIA, CNRS, INRIA, Université de Lorraine, 54000, Nancy, France

When metals and alloys are exposed to air, an oxide layer can grow at their surfaces, thus altering their intrinsic surface properties. Metal / oxide interfaces can also be designed on purpose, to design functional systems in microelectronics, catalysis or coating technologies. Quasicrystalline Al-based approximant surfaces are renowned for their oxidation resistance. But the mechanisms at the atomic scale that drive the oxidation processes are still largely unknown. The investigation of the surface adsorption properties towards atomic oxygen is a first step towards the understanding of such processes, and the most popular method to draw the energy landscape of metallic surfaces is based on Density Functional Theory. This approach is however quite time consuming. Here, we investigate how regression models such as the kernel logistic regression and the support vector regression can accurately interpolate energy surfaces from a few DFT estimates. We focus on the Al₁₃Co₄(100) surface, a quasicrystalline approximant surface. Our study shows that machine learning network can lead to the determination of adsorption energies with a rather small root mean square error (Fig. 1). Further improvement of our approach will be discussed.

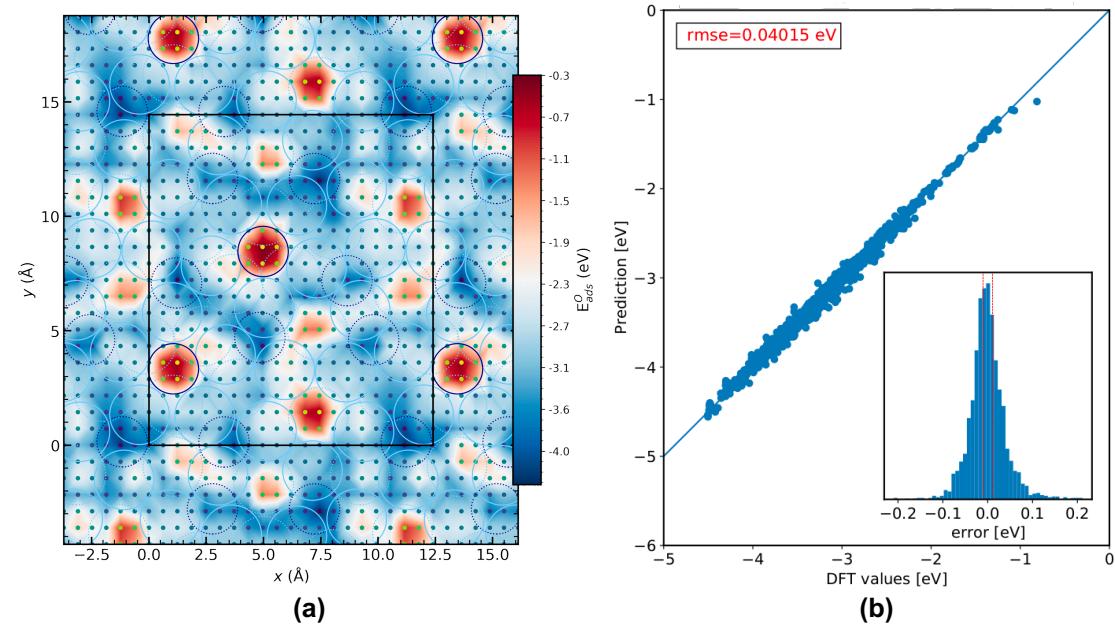


Figure 1. (a) Adsorption energy landscape by DFT: monoatomic oxygen adsorption energy on Al₁₃Co₄ (100) (b) Comparison of adsorption energies (DFT and Kernel Ridge Regression)

Mots-clés : machine learning, quasicrystalline surfaces, adsorption properties

09. Ab initio calculations of finite-temperature properties at multiple electronic structure levels made affordable: An effective combination of perturbation theory and machine learning

Maurício Chagas da Silva¹, Basile Herzog¹, Bastien Casier¹, Michael Badawi¹, Fabien Pascale¹, Tomáš Bučko^{2,3}, Sébastien Lèbegue¹, Dario Rocca¹

¹ Université de Lorraine and CNRS, LPCT, UMR 7019, Nancy, France

² Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Mlynská Dolina, Ilkovičova 6, SK-84215 Bratislava, Slovakia

³ Institute of Inorganic Chemistry, Slovak Academy of Sciences, Dúbravská cesta 9, SK-84236

Among the plethora of existing computational methods in materials science, density functional theory (DFT) provides a great balance between computational cost and accuracy for modeling material properties. However, widely used approximate functionals do not systematically reach chemical accuracy and have known shortcomings. To support conclusions of numerical simulations it is often highly desirable to compare different approximation or even to apply higher levels of theory that go beyond traditional DFT (e.g. the RPA). This is particularly challenging for finite-temperature properties, that require lengthy molecular dynamics simulations with possibly a prohibitive computational cost for the most sophisticated approximations. Recently, we introduced a new methodological approach which couples machine learning (ML) and thermodynamic perturbation theory (TPT) [1,2]. Starting from a molecular dynamics based on a certain production-level approximation, this MLPT method allows to obtain results at several different levels of theory by performing only a few (from a few tens up to a few hundreds) additional fixed-geometry calculations. However, because of its perturbative nature, this approach might depend on the production approximation. We will discuss the limitations of this approach, possible diagnostic tests to detect pathological cases, and a way of correcting MLPT simulations by performing a Monte Carlo resampling (MLMC) [3]. Starting from a semi-local PBE+D2 functional molecular dynamics we will show how it is possible to obtain finite-temperature properties at meta-GGA level of theory or to even attain the fifth rung of DFT approximations from only a few additional single-point calculations. To this purpose we will consider as an example the calculation of enthalpies of adsorption of a molecules in zeolites and we will compare our results with experimental data.

Keywords: machine learning, DFT, molecular dynamics, random phase approximation, zeolites, adsorption enthalpy.

References

- [1] B. Chehaibou, M. Badawi, T. Bučko, T. Bazhirov, and D. Rocca, *Journal of Chemical Theory and Computation* **15**, 6333-6342 (2019).
- [2] M. Gešvandnerová, T. Bučko, and D. Rocca, *Journal of Chemical Theory and Computation* **16**, 6049-6060 (2020).
- [3] M. Chagas da Silva, B. Herzog, B. Casier, M. Badawi, F. Pascale, T. Bucko, S. Lebègue, and D. Rocca, *in preparation*.

SESSION 3

I3. An Atomic Cluster Expansion (ACE) to fit interatomic potentials with polynomials

Geneviève DUSSON¹

¹*Laboratoire de Mathématiques de Besançon*

In this talk, I will present a derivation of interatomic potentials based on polynomials satisfying the true symmetries of the Potential Energy Surface, i.e. an invariance with respect to rotations and permutations of identical particles in the molecular or materials system under consideration. I will explain how such interatomic potential can be fitted from ab-initio data, hence combining the high-accuracy of quantum calculations and the low computational cost of interatomic potentials.

I will then present some numerical results to illustrate the performance of this method as well as provide a short demonstration of ACE.jl, a Julia code developed to perform such fits.

This is a joint work with Markus Bachmayr (Mainz), Gabor Csanyi (Cambridge), Ralf Drautz (Bochum), Cas van der Oord (Cambridge), and Christoph Ortner (UBC)

O10. Machine-learned repulsive potentials for the Density Functional based Tight-Binding Method

Dylan Bissuel¹, Thomas A. Niehaus ¹, Tristan Albaret ¹

¹ Univ Lyon, Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, F-69622 Villeurbanne, France

In the past decade, the landscape of materials simulation has seen the rise of numerous uses of machine-learning tools. As most teams obviously focus on rather varied applications, all of them still push in a similar direction: reshaping the tradeoff between accuracy and computational time. However, for purely numerical tools, the gain of performance is often linked to a loss of information (eg. electronic structure) or a heavy reliance on purely numerical tools, with few physical ground. Semi-empirical methods are older, such as the Density-Functional based Tight-Binding (DFTB) method which appeared at the end of the previous millenium.

Yet, they share the very same concern as ML-tools: partially relieve first-principle methods, such as Density Functional Theory (DFT), from their heavy quantum mechanics computations, without the characteristic accuracy drop implied. This is achieved by using some highly physical approximations, allowing to precompute expensive key-parameters while the information and accuracy losses are marginal.

In DFTB, the main approximation is the computation of the short-range repulsive potential as a simple sum of two-atoms polynomial interactions. This representation is limited by its functional form, while being the method's term restrained by its limited physical background.

In this talk, I will present my thesis work in which I attempt to improve the representativity and transferability of these repulsive potentials by considering it as many-body interactions, using flexible machine-learning tools to refine this already fully parametrized approximation. In particular, I will present the methodology used as well as results for the modeling of Silicon repulsive interactions.

Mots-clés : Density Functional based Tight-Binding method, repulsive potentials, Behler-Parrinello high-dimensional neural-networks

O11. Neural Network Potential for Fast and Accurate Molecular Dynamics Simulations of SiGe

Sung-Ho Lee¹, Valerio Olevano^{2,3}, Benoit Sklenard¹

¹ Univ. Grenoble Alpes, CEA, Leti, 38000 Grenoble, France

² Univ. Grenoble Alpes, F-38000 Grenoble, France

³ CNRS, Institut Néel, F-38042 Grenoble, France

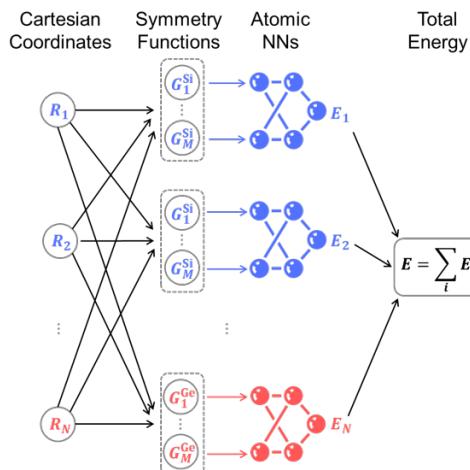


Figure 1. High-Dimensional Neural Network Potential (HDNNP) architecture

In recent years, machine learning (ML) has been finding more and more applications and has become an important area of research in many domains. In the context of computational physics, simulations using ML-powered interatomic potentials for materials were shown to overcome some major barriers that are present in traditional methods¹. Harnessing the power of neural network potentials, calculations can be performed many orders of magnitude faster than DFT, with near-*ab initio* level accuracies and with linear scaling with respect to the system size. This has important implications, namely that simulations that were previously prohibitively expensive for *ab initio* methods are now practicable.

Here, we present a neural network potential (NNP) for SiGe, a material that is of interest in many technological applications. The potential was trained with FFLearn, our in-house NNP package based on Behler's HDNNP model¹, capable of training a potential and running molecular dynamics simulations using the trained force field (via an interface to the LAMMPS code). The training is carried out with a Kalman Filter optimizer² which was shown to perform exceptionally for NNP applications. Both Behler's original Atom Centered Symmetry Functions (ACSF)¹ and its weighted variant (wACSF)³ were tested for this dataset.

Mots-clés : ML-IAP, molecular dynamics, NNP, HDNNP, neural network potential

Références

- ¹ J. Behler and M. Parrinello, *Phys. Rev. Lett.*, **98**, 146401 (2007).
- ² S. Singhal and L. Wu, *Advances in Neural Information Processing Systems* **1** (1988).
- ³ M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi and P. Marquetand, *J. Chem. Phys.*, **148**, 241709 (2018).

O12. Machine-Learning Assisted Canonical Sampling

Aloïs Castellano¹, François Bottin¹, Johann Bouchet²

¹ CEA, DAM, DIF, F-91297 Arpajon, France

² CEA, DES, IRESNE, DEC, Cadarache, F-13018 St Paul Les Durance, France

La Dynamique Moléculaire Ab Initio (AIMD) est une méthode puissante permettant d'échantillonner l'ensemble canonique afin d'étudier les effets de la température sur un matériau. Malheureusement, à cause de son fort coût en temps de calcul, l'AIMD est en général réservée aux systèmes les plus simples. Les progrès récents dans le domaine des Potentiels Interatomiques basés sur le Machine-Learning (MLIP) ont permis une forte accélération dans l'étude des effets de la température. Cependant, la création d'un MLIP est une tâche délicate qui requiert une sélection soigneuse de configurations atomiques et qui se fait généralement par AIMD.

Dans cette présentation, nous introduirons la méthode de Machine-Learning Assisted Canonical Sampling (MLACS). En se basant sur un principe variationnel, MLACS construit un MLIP de manière auto-cohérente dans le but de générer des configurations atomiques qui reproduisent au mieux l'ensemble canonique de la DFT, ce pour une fraction du coût de l'AIMD. Ainsi, une réduction allant jusqu'à deux ordres de grandeurs en temps humain et un ordre de grandeur en temps CPU sont accessibles, tout en gardant une précision proche de la DFT. De plus, en se basant sur des concepts issus de la physique statistique, la méthode permet de calculer des grandeurs thermodynamiques tel que l'énergie libre au niveau de la DFT, ainsi que des propriétés électroniques habituellement hors de portée des MLIP.

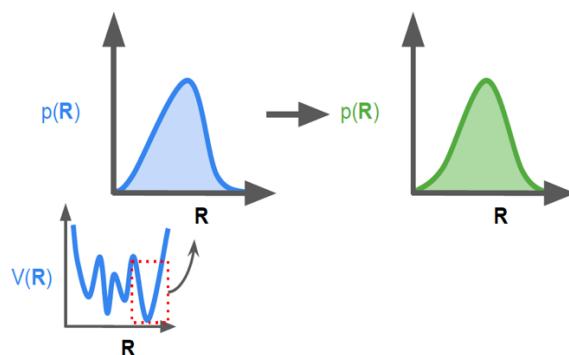


Figure 1. MLACS accélère le calcul de propriétés à température finie en remplaçant la distribution canonique de la DFT par celle d'un potentiel machine-learning.

Mots-clés : potentiels interatomiques, énergie libre, calcul de propriétés physiques

O13. Synergistic coupling in QM/MM simulations of dislocations via machine learning

Petr Grigorev¹, James Kermode, Mihai-Cosmin Marinica², Thomas D. Swinburne¹

¹CNRS, Centre Interdisciplinaire de Nanoscience de Marseille (CINaM), Université Aix-Marseille, France

²DES - Service de Recherches de Métallurgie Physique, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

Ab initio modelling of dislocations is essential to gain quantitative insight into a vast range of important material phenomena, from elementary glide or climb mechanisms to predictive models of solute solution strengthening or post irradiation annealing. The simulation data can be used directly in phenomenological micro-structural models or in the construction of empirical interatomic potentials in a multi-scale framework. With some notable exceptions, ab initio size limitations mean that most dislocations cannot be contained in periodic supercells, requiring the use of flexible boundary techniques which either couple to an elastic Green's function or an empirical interatomic potential in QM/MM simulations. In the latter case, the potential should ideally have identical elastic properties as the ab initio medium. A key limitation of these flexible boundary methods is that whilst the ionic positions and forces can be extracted, the total energy becomes ill-defined.

In recent work [1,2], we showed how this information could be used to extract energetic barriers for glide or segregation through the principle of virtual work. A very good agreement was found with special cases treatable with periodic supercells, whilst in general our ab initio results deviate significantly from predictions of empirical potentials, particularly for prismatic (edge) dislocations. In this work, we show how linear-in-descriptor machine learning (ML) potentials [3] can be used to exactly match ab initio elastic properties, typically leading to efficiencies over "traditional" empirical potentials during QM/MM relaxations. We then go further, using the ionic positions and forces to train the ML potential, showing that the resultant "classical" migration barriers are in excellent agreement with that obtained from QM/MM. Our work shows that QM/MM methods can qualitatively expand the range of atomic structures in the training database of machine learning interatomic potentials, with significant impact on their transferability, or predictive power, for materials modelling.

References:

- [1] Swinburne, T. D., & Kermode, J. R. (2017). Computing energy barriers for rare events from hybrid quantum/classical simulations through the virtual work principle. *Phys. Rev. B*, 96(14), 144102.
- [2] Grigorev, P., Swinburne, T. D., & Kermode, J. R. (2020). Hybrid quantum/classical study of hydrogen-decorated screw dislocations in tungsten: Ultrafast pipe diffusion, core reconstruction, and effects on glide mechanism. *Phys. Rev. Materials*, 4(2), 23601.
- [3] Dérès J, Goryaeva A. M., Lapointe C., Grigorev P., Swinburne, T. D., Kermode, J. R., Ventelon L., Baima J., and Marinica M.-C. Efficient and transferable machine learning potentials for the simulation of crystal defects in bcc Fe and W. Submitted to *Phys. Rev. Materials*

SESSION 4

14. Machine Learning for Atomistic Materials Science

A.M. Goryaeva¹, T. D. Swinburne², C. Lapointe¹, J. Baima¹, J. Dérès¹, Mihai-Cosmin Marinica¹

¹DES - Service de Recherches de Métallurgie Physique, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

²CNRS, Centre Interdisciplinaire de Nanoscience de Marseille (CINaM), Université Aix-Marseille, France

Multi-scale approaches in materials science face a traditional dichotomy in the choice of the atomistic force fields: robust, accurate and numerically expensive ab initio methods against less transferable but fast empirical methods. The Machine Learning (DL) / Deep Learning (DL) methods propose a third avenue that allows control of the balance between the accuracy and numerical efficiency. Moreover, the ML-based vision of fundamental concepts in materials science, such as structural defects, can augment and revise traditional interpretations. We will present recent advances in atomistic material simulations by means of machine learning and data-driven approaches.

Machine learning (ML) methods cannot fully replace traditional approaches in physics and/or materials science. The phase space in physics / materials science has a well-defined structure and is too vast and complex to be described only by the inherent statistical correlations within the data points. Statistical methods trained on the physical data can be of great help when the traditional approaches are limited and/or their direct application is hindered by factors such as high computational cost.

In metals, the interaction and transformation of crystal defect networks gives rise to an extraordinarily diverse range of defect morphologies [1]. Using the recently developed package MiLaDy (Machine Learning Dynamics) [2]: (i) we redefine the concept of defects in materials science [3]; (ii) we provide reliable force fields for complex defects such as interstitial, dislocation loops, dislocations; (iii) we are able to explore the atomistic free energy landscape of point defects in metals with ab initio accuracy up to the melting temperature [4], and, finally, (iv) we are able to propose surrogate models that bypass the traditional approaches [5]. We exemplify and discuss in the framework of experimental findings the case of energetic landscape of defects in body centered and face centered cubic metals.

References:

- [1] K. Arakawa , M.-C. Marinica et al. Nature Mat. 19, 508(2020) ; R. Alexander et al. Phys. Rev. B 94, 024103 (2016)
- [2] M.-C. Marinica, A. M. Goryaeva, T. D. Swinburne et al, MiLaDy - Machine Learning Dynamics, CEA Saclay, 2015-2021 ; A.M. Goryaeva, J.-B. Maillet, M.-C. Marinica. Comp. Mater. Sci. 166, 200 (2019)
- [3] A. M. Goryaeva et al. Nature Commun. 11, 4691 (2020)
- [4] C. Lapointe et al. (to be submitted). T.D. Swinburne, M.-C. Marinica, Phys. Rev. Lett. 120, 135503 (2018)
- [5] C. Lapointe, T. D. Swinburne et al. Phys. Rev. Materials 4, 063802 (2020); F. Bruneval et al. J. Chem. Theory Comput. 16, 4399 (2020)

O14. Modèles de régression de l'entropie vibrationnelle de formation et de migration de défauts ponctuels

Clovis Lapointe¹, Thomas D. Swinburne², Charlotte S. Becquart³, Laurent Proville³, Mihai-Cosmin Marinica¹

¹ DEN-Service de Recherches de Métallurgie Physique, CEA Saclay, 91191 Gif-sur-Yvette, France

² Aix-Marseille Université, CNRS, CINaM UMR 7325, Campus de Luminy, 13288 Marseille, France

³ Univ. Lille, CNRS, INRAE, Centrale Lille, UMR 8207 - UMET - Unité Matériaux et Transformations, F-59000 Lille, France

Les méthodes de type Machine Learning se sont démocratisées comme un outil efficace en science des matériaux. Elles permettent de construire des modèles de régressions précis pour un grand nombre d'observables associées à un paysage énergétique complexe, grâce à l'utilisation des descripteurs atomiques locaux.

Récemment, nous avons développé un modèle de régression de l'entropie vibrationnelle pour les défauts ponctuels et de leurs amas dans un cristal modèle de fer cubique centré¹. Le coût numérique de cette méthode évolue comme $O(N)$ (où N est le nombre d'atome du système) et permet d'évaluer l'entropie vibrationnelle pour des systèmes de grandes tailles dépassant ainsi largement les limites des méthodes standards telle que la diagonalisation de la matrice Hessienne qui évolue comme $O(N^3)$.

Nous proposons, ici, une généralisation non-linéaire à la proportionnalité entre une observable locale et un descripteur local. Cette généralisation sera utilisée pour prédire les variations d'entropies vibrationnelles de formations dans le cas d'une déformation. En utilisant cette approche, nous pouvons prédire l'entropie de formation d'une configuration déformée à partir uniquement de l'entropie de l'état non-déformé et les coordonnées du système déformé. Ce type de modèle est testé sur une base de données de fer cubique centré.

De plus, nous proposons un modèle de régression d'entropie vibrationnelle et d'énergie de migration de défauts. Ce modèle est mis en œuvre pour une base de données de silicium amorphe et présente un bon pouvoir prédictif. Enfin, nous mettons en avant une forme de généralisation de la loi phénoménologique de compensation de Meyer-Neldel dans l'espace des descripteurs. Dans l'espace des descripteurs, nos modèles présentent les mêmes corrélations que cette loi phénoménologique mais se basent uniquement sur des informations géométriques du système.

Cette étude ouvre un grand nombre d'opportunité pour les simulations multi-échelles car elle permet d'estimer « à la volée » des fréquences de migration associées aux mouvements des défauts dans les matériaux, permettant ainsi une approche précise de la cinétique de ces défauts et donc du vieillissement des matériaux.

Mots-clés : Température finies, défauts, contributions vibrationnelles, entropie, fréquences d'attaque

Références

¹ C. Lapointe*, T.D. Swinburne, L. Thiry, S. Mallat, L. Proville, C.S. Becquart et M-C Marinica, *Phys. Rev. Mat.* **4**, 063802(2020)

O15. Learning how dense dislocation networks evolve

Thomas D. Swinburne¹

¹ CNRS / CINaM, Aix-Marseille University

The defining characteristic of non-brittle fracture is dependency on the preexisting network of dislocations, topological line carriers of plasticity in crystalline materials. Under repeated mechanical loading or sustained irradiation, dislocations form dense tangles deep in a material's microstructure [1,2]. The evolution of such tangles under external loading plays a critical role in emergence of stress intensities, which later lead to failure. However, their analysis is challenging- highly deformed structures can lower barriers for local mass transport, enabling non-conservation of Burgers vector and other complex rearrangements that are incompatible with typical dislocation dynamics simulations. I will present some exploratory simulations which generated large ensembles of dense dislocation tangles in W[0-15%]Re evolving under an applied shear stress. I will show that with well-chosen descriptors, conditional distributions of microstructural changes during loading (e.g. stress, dislocation junction count or line length) can be reproduced by machine learning surrogate models, both shallow and not-so-shallow. In particular, marginal distributions can be sufficiently accurately reproduced (as measured by the cross-entropy) that subtle model or chemistry dependent changes in the evolution behaviour can be predicted. How these results can help data-driven microstructural models will be discussed.

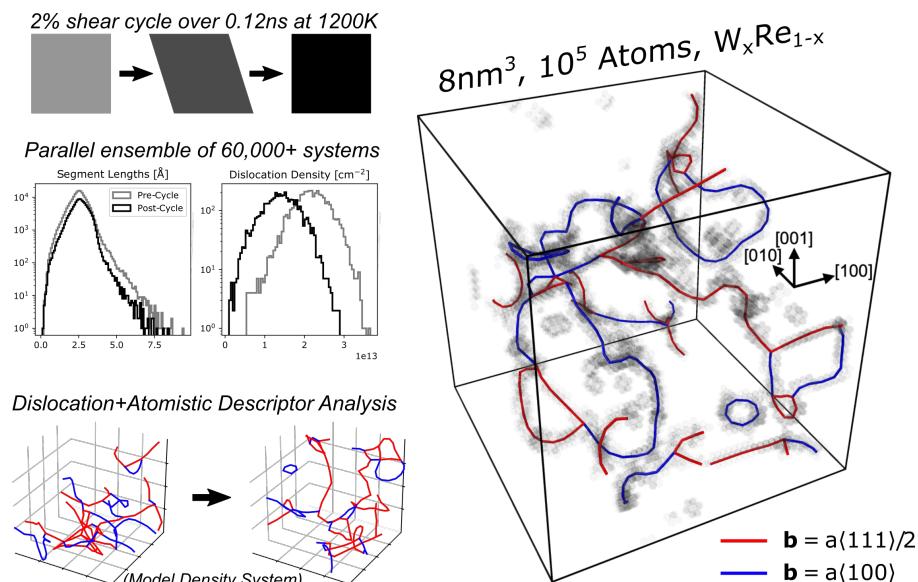


Figure 1. Illustration of simulation setup for annealing under shear in EAM-W[0-15%]Re

Mots clés : microstructural evolution, surrogate modelling

Références :

¹ TD Swinburne et al., Scientific Reports, 2016

² S Lavenstein et al., Science, 2020

O16. Reinforcing materials modelling by encoding the structures of defects in crystalline solids into distortion scores

A.M. Goryaeva¹, C. Lapointe¹, J.-B. Maillet², M.-C. Marinica¹

¹ Université Paris-Saclay, CEA, Service de Recherche de Métallurgie Physique, 91191, Gif-sur-Yvette, France

² CEA - DAM, DIF, F-91297 Arpajon Cedex, France

A perfect crystal is a purely theoretical concept. Real-world crystals contain imperfections, also called defects. Identification and characterization of defects provide the crucial information for interpretation of simulations and experiments that bridge the gap between atomic- and micrometer-scales. But what is actually a concept of defects at the atomic scale? How to define the spreading of the defect within the crystalline lattice and understand its impact on the energy landscape of materials? In this study, we try to answer these questions using machine learning (ML) methods and introduce a novel concept of the distortion scores of local atomic environments to reinforce the methods of atomic-scale materials modeling.

The distortion score of local atomic environments¹ describes a statistical distance from a reference (e.g., perfect crystal) structure in the descriptor space. Atomic descriptors, which span the descriptor space, are computed using MiLaDy package². Based on the distortion scores, we identify structural defects as atoms-outliers deviating from the reference structure. This score facilitates automatic localization of defects and enables their stratified description, which allows to distinguish the zones with different levels of distortion within the structure. The high sensitivity of the present approach to atomic displacements allows to detect the patterns produced by the elastic strain field around defects.

Beyond the localization of defects in the structure, the proposed concept of distortion scores has many advanced applications ranging from the surrogate concept for the energy per atom to the selection of the relevant structural information for the quantitative evaluation of energy barriers from the mean force. Moreover, this concept can serve for the design of robust interatomic ML potentials and high-throughput analysis of their databases.

In perspective, this approach can be further extended for amorphous structures as well as for interpretation of multivariate data provided by experimental techniques where the atomic coordinates are provided, such as atom probe tomography and transmission electron microscopy.

Mots-clés : Machine learning, defects, distortion score, minimum covariance determinant

Références

¹ A.M. Goryaeva *et al.* **Nature Comm.** **11**, 4691 (2020)

² A.M. Goryaeva, J.-B. Maillet, M.-C. Marinica **Comp.Mater. Sci.** **166**, 200-209 (2019)

O17. Graph Theory for Molecular Dynamics simulations

Prof Marie-Pierre Gaigeot¹

¹ LAMBE UMR 8587, Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université d'Evry val d'Essonne – Université Paris Saclay, France

Our recent developments in graph theory algorithms applied to molecular dynamics trajectories (DFT-based and classical MD) are presented. Most of our developments have been done for gas phase molecules, but they can be readily applied to materials, as will be discussed in this talk.

1 – Graph theory for the automatic recognition of 3D conformers along MD trajectories, at the atomistic granularity level, isomorphism being a key component together with keeping the atoms chemical nature. Isomorphism allows recognize conformations and provides graphs of transitions, the reduction in the isomorphism complexity has been achieved by the introduction of “orbits” and “reference snapshots.” The method provides the changes in conformations over time based on hydrogen bond(s), proton transfer(s), coordination number(s), covalent bond(s), fragmentation(s). We will show applications to gas phase molecules and clusters as well as direct transferability to condensed matter.

2 – Graph theory and deep learning for going from the graphs to 3D molecular structures, showing applications to peptides of increasing complexity.

3 – Graph theory for the assignment of anharmonic vibrational modes from MD simulations, building one graph/vibrational mode, without harmonic approximations. Graph vertices are made of the internal coordinates self-contributions to the mode while the graph edges contain the correlations between the internal coordinates. All coupling motions are therefore naturally included in the graph modes. The ultimate advantage of graphs is the capability for comparing graphs and extracting similarities, exactly what is needed for comparing vibrational modes between molecular systems, and this will be illustrated for peptides and their dimers.

4 – Gas phase polymers are highly flexible hence displaying complex dynamics, which analysis in terms of time-evolution of the 3D structures is rather complicated. We have developed graphs algorithms to analyze such trajectories directly in terms of the polymorphism of the H-Bonded rings over time. Illustrations will be given on Z-Ala₆ and Gramicidin peptides, to emphasize that this graph level granularity is the only one giving rise to a comprehensive view of the 3D structures time-evolution.

Works done in collaboration with Prof D. Barth, Dr S. Bougueroua, Dr D. Galimberti, Dr M. Bricage, Y. Aboulfath (PhD), V. Chantitch (PhD)

Our publications on these topics:

- [1] S. Bougueroua, R. Spezia, S. Pezzotti, S. Vial, F. Quessette, D. Barth, M.-P. Gaigeot, Graph theory for automatic structural recognition in molecular dynamics simulations. *J. Chem. Phys.* 149:184102-15 (2018)
- [2] D.R. Galimberti, A. Milani, M. Tommasini, C. Castiglioni, M.-P. Gaigeot Combining static and dynamical approaches for infrared spectra calculations of gas phase molecules and clusters *J. Chem. Theory Comput.*, 13:3802 (2017)
- [3] D. R. Galimberti, S. Bougueroua, J. Mahé, M. Tommasini, A. M. Rijs, M.-P. Gaigeot, Conformational assignment of gas phase peptides and their H-Bonded complexes using far-IR/THz: IR-UV ion dip experiment, DFT-MD spectroscopy, and Graph Theory for modes assignment. *Faraday Discussions, 'Advances in ion spectroscopy: From astrophysics to biology'* 217:67 (2019)

O18. Approche par deep learning pour déterminer les indices de chiralité des nanotubes de carbone à partir d'images HRTEM

G. Daniel Förster^{1,2,3}, Alice Castan^{3,4}, Annick Loiseau³, Jaysen Nelayah⁵, Damien Alloyeau⁵, Frédéric Fossard³, Christophe Bichara², Hakim Amara^{3,5}

¹ Interfaces, Confinement, Matériaux et Nanostructures, ICMN, Université d'Orléans, CNRS, Orléans

² CINaM - UMR 7325, CNRS - Aix Marseille Université, 13288 Marseille

³ Laboratoire d'Etude des Microstructures, ONERA-CNRS, Châtillon

⁴ Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA

⁵ MPQ, Université de Paris, CNRS, Paris

Notre objectif est d'identifier de façon fiable la structure des nanotubes à l'aide d'images obtenues par microscopie électronique à transmission de haute résolution. Il s'agit d'une tâche délicate et l'analyse d'un échantillon statistiquement significatif de tubes peut s'avérer fastidieuse si elle est effectuée « à la main ». Pour permettre une identification plus facile de la chiralité, nous avons conçu une procédure automatisée de reconnaissance d'images basée sur un réseau neuronal convolutionnel (CNN).

Notre base de données est alimentée par des images de microscopie simulées à partir de configurations atomiques obtenues lors de simulations de dynamique moléculaire¹. Sur la base de ces géométries, en utilisant la technique multislice, nous simulons des images HRTEM avec différents coefficients d'aberration et différentes orientations². Après l'entraînement du système de deep learning³, l'identification automatique de la structure des tubes se révèle rapide et fiable. En analysant les images simulées on atteint une précision de plus de 90%. Appliquée aux images HRTEM expérimentales, une analyse d'un nombre statistiquement significatif de nanotubes devient possible.

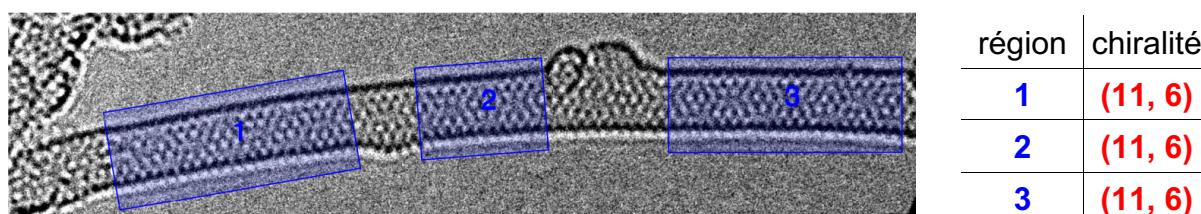


Figure 1. Image HRTEM expérimentale analysée par le système automatisée

Mots-clés : deep learning, nanotubes de carbone, HRTEM

Références

¹ J. Tersoff, *Phys Rev B* **39**, 5566 (1989).

² J. Barthel, *Ultramicroscopy* **193**, 1 (2018).

³ F. Chollet et al., <https://keras.io> (2015).

SESSION POSTERS

P1. Learning the dynamics of systems with memory: a maximum likelihood approach

Hadrien Vroylandt¹

¹ Sorbonne Université, institut des sciences du calcul et des données, ISCD, F-75005 Paris, France

The derivation of Generalized Langevin Equations for a set of coarse-grained variables requires the ability to compute the memory kernel from the underlying dynamics of the full system.

Various strategies have been proposed to achieve this goal, such as solving the integro-differential equation satisfied by the kernel or optimizing a (parametric or time-discretized) kernel to best reproduce time correlation functions.

Here, we will discuss an approach based on the probabilistic reconstruction of the noise history for the coarse-grained variables of interest, using extended dynamics with hidden variables.

Such extended dynamics have been promoted for the integration of Generalized Langevin Equations, by reintroducing additional degrees of freedom following Markovian dynamics.

We propose to reconstruct the dynamics of the hidden variables from the trajectory of the ones of interest (observed on the initial, non-coarse-grained system), following a maximum likelihood approach.

The subsequent analysis of the extended dynamics then provides statistical information on the noise history, including the corresponding memory kernel, as well as dynamical properties of the studied system.

Mots-clés : molecular dynamics, maximum likelihood, memory kernel

P2. Unsupervised learning of atomic-scale dynamics

Arthur France-Lanord¹

¹ Institut des sciences du calcul et des données, ISCD, Sorbonne Université, France

We introduce an unsupervised learning method able to capture atomic-scale dynamics, using molecular dynamics trajectories as an input. The local environment of atoms is described using graph convolutional neural networks, a representation which conveniently incorporates permutation and rotation invariances. The weights of the network are optimized in order to minimize the variational loss introduced in VAMPnets, which allows us to train a linear model of atomic dynamics. The method can be applied to rationalize many different kinetic phenomena, including for instance ion transport in crystalline and amorphous electrolytes, and atom dynamics at solid liquid interfaces. Finally, we apply this methodology to the description of ion association in liquid electrolytes. Analyzing the optimized neural network allows us to determine what makes a good structural descriptor for such a class of transformations.

Mots-clés : unsupervised learning, neural networks, dynamics, ion association

Références

¹ T. Xie, A. France-Lanord, Y. Wang, Y. Shao-Horn, J. C. Grossman, Nature communications 10, 2667 (2019).

² A. Mardt, L. Pasquali, H. Wu, F. Noé, Nature communications 9, 5 (2018).

P3. Supervised learning of DFT results to predict the formation enthalpy of the full set of configurations in complex phases: the sigma-phase as an example

Jean-Claude Crivello¹ Jean-Marc Joubert¹, Nataliya Sokolovska²

¹ Université Paris Est, ICMPE (UMR 7182), CNRS, UPEC, Thiais F-94320, France

² NutriOmics, INSERM, Sorbonne University Paris, France

Machine learning (ML) is becoming the state-of-the-art method in numerous domains, including material sciences. In this paper, we demonstrate how ML can be used to efficiently predict the heat of formation of a given complex crystallographic phase (here, the σ -phase, $tP30$, $D8b$).

Based on an independent and unprecedented large first principles dataset containing about 10,000 σ -compounds with $n=14$ different elements, we used a supervised learning approach to predict all the $\sim 540,000$ possible configurations within a mean absolute error of 23 meV/at (~ 2 kJ/mol) on the heat of formation and ~ 0.06 Å on the tetragonal cell parameters.

We show that deep neural network regression results in a significant improvement in the accuracy of the predicted output compared to traditional regression techniques. We also integrated descriptors having physical nature (atomic radius, number of valence electrons), and we observe that they improve the model precision.

We conclude from our numerical experiments that the learning database composed of the binary-compositions only, plays the major role in predicting the higher degree system configurations. Our results open a broad avenue to efficient high-throughput investigations of the combinatorial binary computations for multicomponent prediction of complex phases.

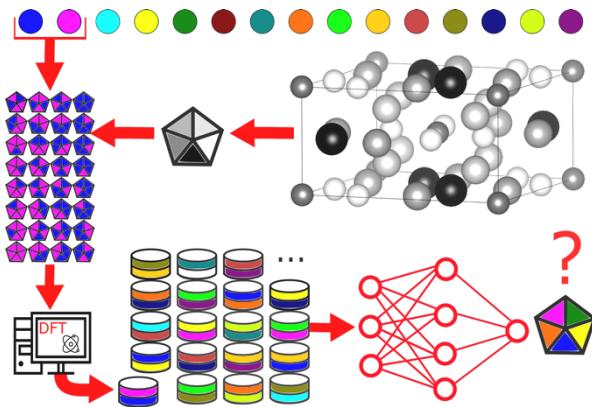


Figure 1. Schematic representation of properties prediction of the multicomponent σ -phase from binary properties learning.

Mots-clés : supervised machine learning, regression, intermetallic.

Références

¹ <https://arxiv.org/abs/2011.10883>

P4. Machine learning and Umbrella sampling to investigate chemical reactions in solution

Timothée Devergne¹, Théo Magrino¹, Fabio Pietrucci¹, A. Marco Saitta¹

¹ Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, IMPMC, Paris, France

L'étude des réactions chimiques en solution en utilisant des simulations de dynamique moléculaire est un vaste domaine. Une de ses applications est la chimie prébiotique : l'étude des réactions qui sont à l'origine de la vie sur terre. Dans ce cas, l'environnement joue un rôle très important dans la synthèse de nouvelles molécules. En effet, les différents catalyseurs utilisés en chimie moderne ne sont pas présents en milieu prébiotique : seulement certaines molécules qui ont servi de briques élémentaires étaient sur Terre, dans de l'eau ou sur des surfaces minérales. C'est pour cela qu'il faut prendre en compte le solvant dans les simulations. Il faut de plus prendre en compte les effets quantiques, ceci est fait grâce aux techniques dites *ab initio* qui sont très coûteuses en temps de calcul : il grandit avec le cube du nombre d'atomes. De ce fait, l'échelle de temps et d'espace accessible est de l'ordre de la nanoseconde et du nanomètre. Ici, nous présentons comment nous proposons d'ajouter du machine learning à un protocole d'échantillonnage déjà existant et qui a fait l'objet de plusieurs articles^{1,2} pour pouvoir étudier des systèmes plus grands. Pour cela, nous utilisons les données d'une étude précédente de l'équipe³ pour développer une méthode efficace de création d'ensemble d'entraînement permettant d'obtenir une surface d'énergie libre d'une qualité comparable à celle obtenue grâce à des simulations *ab initio* à moindre coût.

Mots-clés : Machine Learning, chimie prébiotique, chimie en solution, ab initio

Références

- ¹ Fabio Pietrucci, Antonino Marco Saitta, *PNAS* **112-49**, 15030-15035 (2015).
- ² Andrea Pérez-Villa, Fabio Pietrucci, A Marco Saitta, *Physics of life reviews* **34-35**, 105-135 (2020).
- ³ Théo Magrino, Fabio Pietrucci, A Marco Saitta, *J. Phys. Chem. Lett* **12-10**, 2630–2637 (2021)

P5. Modeling of Ni-Si interactions through a linearized machine learning potential

César Jara¹, Julien Lam², Antoine Jay¹, A. Hémeryck¹

¹LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

²CEMES, CNRS and Université de Toulouse, 29 rue Jeanne Marvig, 31055 Toulouse Cedex, France

Silicides are compounds of metals with silicon that play a significant role in modern device technology as efficient metallic source/drain junctions. Particularly, Ni-based alloys are interesting because of their low resistivity, low process temperature, and low Si consumption. Although the use of Ni-based silicides has proven to be very successful, the knowledge on their formation is still far from complete, and detailed knowledge of the process is required.

Although quantum methods (as density functional theory (DFT)) are considered the most accurate, their computational cost is high, limiting their use for the study of large-scale systems and long-time simulations. Machine-learning interaction potentials (MLIP) have been proposed in this scope in the recent years, and their success is recognized by the large variety of materials that have been successfully tackled [1].

In this work, we propose to model the Ni-Si interactions thanks to the development of a linearized machine learning potential [2]. For the development of a MLIP, the main procedure consists of using a very universal analytical formulation for the force field which is then parameterized to match a database of DFT calculations made of forces in various structures [1]. Our NixSiy MLIP has been developed for Ni-silicides phases (as Ni₃Si, Ni₂Si, and NiSi), pure Ni, as well as pure Si. The testing of the potential's accuracy was performed using molecular dynamics simulations, in order to compare with DFT calculations of energy and structural characteristics.

Références

- [1] Geka, P., Stoltz, G., Barati Farimani, A., Belkacemi, Z., Ceriotti, M., Chodera, J. D., Lelièvre, T. (2020). Machine learning force fields and coarse-grained variables in molecular dynamics: application to materials and biological systems. *Journal of Chemical Theory and Computation*, 16(8), 4757-4775.
- [2] Benoit, M., Amodeo, J., Combettes, S., Khaled, I., Roux, A., Lam, J. (2020). Measuring transferability issues in machine-learning force fields: the example of gold–iron interactions with linearized potentials. *Machine Learning: Science and Technology*, 2(2), 025003.

P6. Quantum dynamics and Monte Carlo simulations through machine learning

Julien Heu¹, Michele Casula¹, A. Marco Saitta¹

¹ Sorbonne Université, Muséum National d'Histoire Naturelle, UMR CNRS 7590, IMPMC, Paris, France

Lorsque l'on souhaite prendre en compte certains effets quantiques (comme la densité électronique) lors d'une simulation numérique, nous devons recourir à des simulations dites *ab initio*, comme la Density Functional Theory (DFT), Monte Carlo (MC) quantique, Coupled Clusters (CC), etc. Ces méthodes apportent un degré de précision supplémentaire par rapport à leur contrepartie classique, au prix d'un temps de calcul évoluant avec le cube du nombre d'électrons pour la DFT, jusqu'à la puissance sept pour la méthode des CC. Il n'est donc pas possible de simuler de grands systèmes (>1k atomes) sur de longues périodes (>1ns) avec cette précision quantique. Ces dernières années, des méthodes^{1,2,3} sont apparues permettant d'entraîner des réseaux de neurones à reproduire l'énergie potentielle et les forces d'un système selon sa configuration atomique. Ces méthodes permettent de réduire considérablement les temps de calcul des simulations *ab initio*³ et permettent aussi d'étudier des systèmes de plus grande taille avec un temps de calcul évoluant linéairement avec le nombre d'atomes^{1,3}. Nous avons développé un tel outil avec lequel nous avons pu étudier avec succès différents systèmes (ion Zundel, hexamère d'eau, eau bulk, silicone bulk, CO₂ bulk). Une fois nos réseaux de neurones entraînés, nous nous sommes particulièrement concentrés sur des simulations MC, contrairement à la majorité des approches usant de dynamique moléculaire. Nous avons mis au point un panel d'outils permettant de conserver la stabilité desdites simulations face à l'aspect stochastique du procédé (papier en cours de publication), de façon générique et indépendante du système étudié.

Mots-clés : Machine Learning, Monte Carlo, dynamique moléculaire, simulations ab initio

Références

- ¹ Behler, J., & Parrinello, M, *Phys. Rev. Lett.* **98**, 146401 (2007).
- ² Bartók, A. P., Kondor, R., & Csányi, G., *Phys. Rev. B* **87**, 184115 (2013).
- ³ Zhang, L., Han, J., Wang, H., Car, R., & E, W., *Phys. Rev. Lett.* **120**, 143001 (2018)

P7. Exploring machine learned collective variables in conjunction with rare events sampling methods in ab-initio molecular dynamics for catalytic reactions

T. Pigeon^{1,2}, M. Corral-Valero², A. Anciaux-Sedrakian³, G. Stoltz¹, P. Raybaud², T. Lelièvre¹

¹ CERMICS, Ecole des Ponts ParisTech, 6/8 av. Blaise Pascal - Citée Descartes, 77455 Marne-La-Vallée, FRANCE and Equipe-projet MATHERIALS, Inria Paris, 2 rue Simone Iff, CS 42112 75589 Paris Cedex 12

² IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP3, 69360 Solaize, France

³ IFP Energies nouvelles, 1-4 Avenue du Bois Préau, 92852 Rueil-Malmaison, France

Every elementary step of a catalytic chemical reaction can be described by a transition from one metastable state to another one. The transition paths and times are generally impossible to obtain by full *ab initio* molecular dynamics (AIMD). Indeed, given that the time step is $\sim 10^{-15}$ seconds for the numerical integration of AIMD, the direct simulation of events occurring once every microsecond or millisecond is out of reach. To overcome this bottleneck, our work explores a rare event sampling method called Adaptive Multilevel Splitting (AMS),^{1,2} combined with *ab initio* molecular dynamics (periodic plane-wave DFT).

The AMS algorithm requires the knowledge of the two metastable states and a one-dimensional collective variable discriminating them. Such a function can be determined by machine learning techniques given data only on the two metastable states.³ The efficiency of the AMS algorithm strongly depends on this function. In the present poster: (i) we compare different machine learning methods (SVM, LDA) to identify collective variables from AIMD simulations. For the analysis of chemical structures, different types of structural descriptors (SOAP, ACSF..) are used, (ii) we analyze the AMS behavior depending on the collective variable used, (iii) we benchmark the performance of the AMS algorithm with the learnt collective variable on various case studies of growing complexity: from simple 2D mathematical cases to real catalytic reaction such as dissociation of water on γ -alumina.⁴

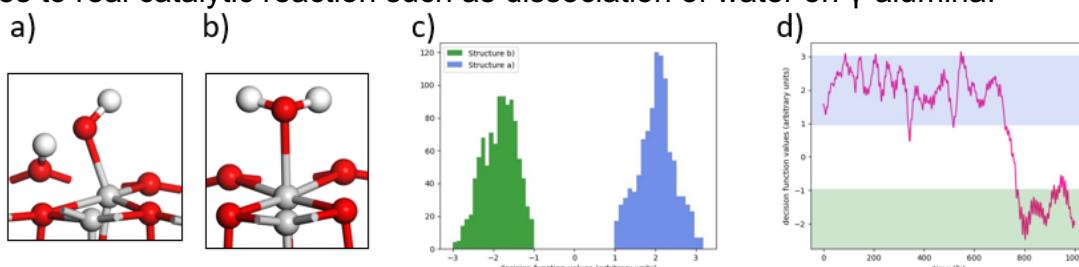


Figure 1. a) Reactant: adsorbed water dissociated into two surface hydroxyls. b) Product: adsorbed molecular water on alumina surface. c) Histogram of a SVM classifier decision function on its training set. d) Evolution of a SVM classifier decision function along a AIMD going from structure a) to b). Color legend: red: Oxygen; grey: Aluminum; white: Hydrogen

Keywords: Machine learning, Collective variables, AMS, AIMD, Catalytic reactions.

References

- ¹ F. Cérou, A. Guyader, *Stochastic Analysis and Applications* **25**, 417-443 (2007).
- ² L. J. S. Lopes, T. Lelièvre, *Journal of computational chemistry* **40**, 1198-1208 (2019).
- ³ M. M. Sultan, V. S. Pande, *The Journal of chemical physics* **149**, 094106 (2018).
- ⁴ M. Digne, P. Sautet, P. Raybaud, P. Euzen, H. Toulhoat, *Journal of catalysis* **226**, 54-68 (2004)

P8. RPA and coupled cluster finite-temperature properties from machine learning thermodynamic perturbation theory

Basile Herzog¹, Mauricio Chagas da Silva¹, Bastien Casier¹, Michael Badawi¹, Tomáš Bučko², Sébastien Lebègue¹, Dario Rocca¹

¹ Université de Lorraine and CNRS, LPCT, UMR 7019, Nancy, France

² Comenius University and Slovak Academy of Sciences, Bratislava, Slovakia

Correlated electronic structure methods, such as the random phase approximation or the couple cluster theory, are typically considered too computationally expensive to be systematically applied in simulations of realistic materials and condensed matter systems. This is particularly true if temperature effects are important and molecular dynamics or Monte Carlo simulations are required. In recent work we showed that finite-temperature enthalpies of adsorption in zeolites can be computed at the RPA level of theory by performing only as few as 10 fixed geometry RPA calculations [1]. This is achieved by coupling machine learning techniques and thermodynamic perturbation theory (MLPT method). Despite the potential of this approach, dependence of the final result on the starting point should be better understood. In this poster we will present new results based on a Monte Carlo resampling of the RPA trajectories that strongly decreases the starting point dependence and we show how the MLPT ideas can also be applied to perform coupled cluster finite-temperature simulations.

Mots-clés : machine learning, coupled cluster theory, random phase approximation (RPA), molecular dynamics, zeolites, adsorption enthalpy.

Références

- [1] B. Casier, M. Chagas da Silva, M. Badawi, F. Pascale, T. Bucko, S. Lebègue, and D. Rocca, *Journal of Computational Chemistry*, **15**, 6333 (2019).

P9. PANNNA a comprehensive toolkit for creating neural network models for atomistic systems: developer updates

Ruggero Lot^{1,2,3}, Franco Pellegrini⁴, Emine Kucukbenli^{3,5}, Yusuf Shaidu^{3,6}, Stefano De Gironcoli^{2,3}

¹*LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France*

²*CNR-IOM, Democritos National Simulation Center, Istituto Officina dei Materiali, Istituto Officina dei Materiali, c/o SISSA, via Bonomea 265, IT-34136 Trieste, Italy*

³*SISSA, via Bonomea 265, IT-34136 Trieste, Italy*

⁴*Laboratoire de Physique Statistique, Ecole Normal Supérieure, Université PSL, 75005 Paris, France*

⁵*John A. Paulson School of engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

⁶*The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy*

PANNA (Properties from Artificial Neural Network Architectures) is a package to train and validate fully connected deep neural network and build atomic potentials based on Behler-Parrinello or other local atomic environment descriptors.

These potentials can then be efficiently used to perform molecular dynamics with LAMMPS or similar packages.

The latest release of the code allowed researchers to perform such operations through python scripts configured with input files, leaving little space for power users to implement new features, explore different architectures or reuse the code to investigate the training behavior. Since then, a lot of effort has been put in converting the code to a library, allowing its usage in ipython and consequently on the widespread Jupiter platform. However, this functionality was still limited by the graph paradigm used in the development of the core parts of the package (inherited from the first version on Tensorflow, our neural network library of choice).

In this developer update, we discuss the main strategies that we are adopting during the migration from TF1.8 to TF2.* and their advantages. A strong simplification in the code base has been made possible by the new API, simplifying the implementation of new architectures and metrics. The usage of the ipython environment is now much easier, allowing for faster code prototyping.

References

- [1] R. Lot, F. Pellegrini, Y. Shaidu, and E. Kucukbenli, “PANNA: Properties from artificial neural network architectures” *Comput. Phys. Commun.* vol. 256, p. 107402, Nov. 2020.

P10. Two-dimensional metal structures revealed by evolutionary computations: Pb/Al₁₃Co₄(100) as a case study

Florian Brix¹, Emilie Gaudry¹

¹ Université de Lorraine, CNRS, Institut Jean Lamour, Campus Artem, Nancy, France

We have combined extensive Density Functional Theory calculations with an evolutionary algorithm to investigate possible structural models for two-dimensional (2D) Pb films supported on the Al₁₃Co₄(100) quasicrystal approximant. The minimization of the total energy with the constraint of maximizing the atomic density in the layer leads to 2D atomic arrangement with pentagonal motifs, reflecting the symmetry of the substrate. Our findings show that the 2D Pb structure can be interpreted as a stable structure, with 16 Pb atoms in the surface cell, in good agreement with the measured coverage and scanning tunnelling microscopy images. Alternatively, a metastable 15-atom 2D film also fits with the experimental observations. This study opens a route towards the prediction of supported complex 2D films [1].

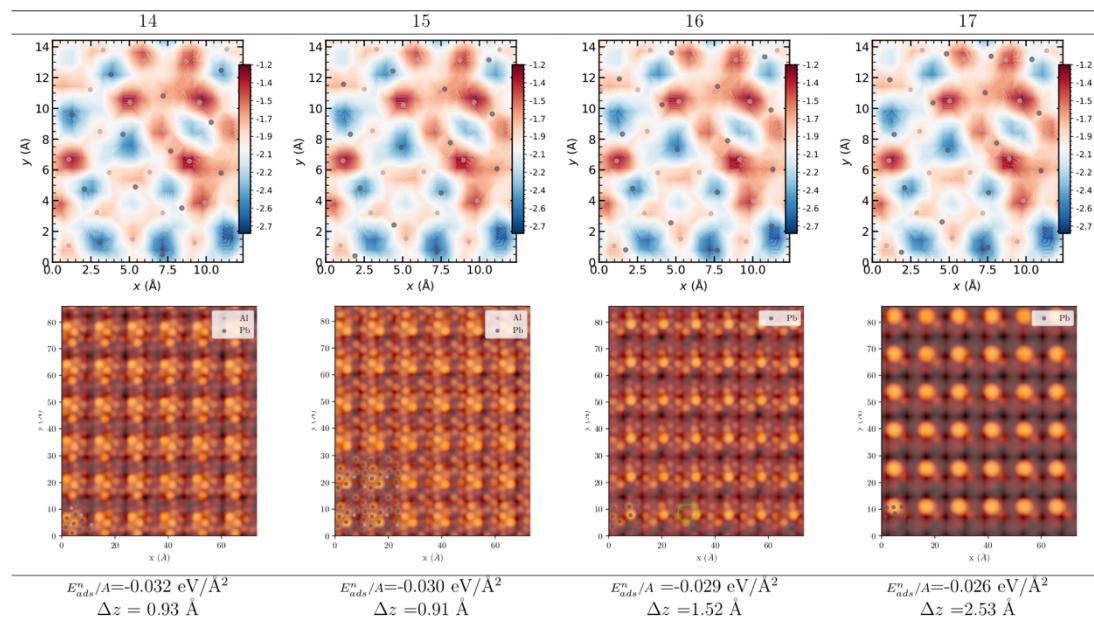


Figure 1 : Adsorption energies, structures and simulated STM (scanning tunneling microscopy) images for 2D Pb-films with different atomic densities on Al₁₃Co₄(100).

References

- [1] F. Brix and E. Gaudry, Two-dimensional metal structures revealed by evolutionary computations: Pb/Al₁₃Co₄(100) as a case study, submitted

INFO PRATIQUES

VENIR AU LAAS-CNRS

Adresse postale :

LAAS-CNRS
7, avenue du Colonel Roche
BP 54200
31031 Toulouse cedex 4

En arrivant à Toulouse en Train à la gare Matabiau

Bus : Le LAAS-CNRS est accessible en bus – direct depuis la Gare Matabiau par la Ligne 27 direction Ramonville.

Métro + bus ou Métro + à pied : La Ligne B du Métro dessert le Campus Paul Sabatier jusqu'à la station Faculté de Pharmacie notamment :

- puis une connexion en Bus par la ligne 78 est direct jusqu'au LAAS-CNRS.
- Le trajet à pied de la station de Métro Faculté de Pharmacie jusqu'au LAAS-CNRS vous prendra 16 min et vous permettra d'enjamber le Canal du Midi !

En arrivant à Toulouse en avion, à l'aéroport de Toulouse Blagnac

Tram + Métro : Tram ligne T2 jusqu'au terminus Palais de justice, puis prendre le métro ligne B direction Ramonville, descendre station Faculté de Pharmacie, puis prendre le bus 78 direct jusqu'au LAAS-CNRS.

Bus + Métro : prendre la navette aéroport jusqu'au centre-ville (30 min) arrêt : station Jeanne d'Arc, prendre le métro ligne B direction Ramonville, descendre station Faculté de Pharmacie, puis prendre le bus 78 direct jusqu'au LAAS-CNRS.

Métro Arrêt « Faculté de pharmacie »
Ligne B – connexion bus 78

Bus Stop « LAAS »
Lignes 27, 37, 78

Bus Stop « ENAC »
Lignes 27, 37, 78

