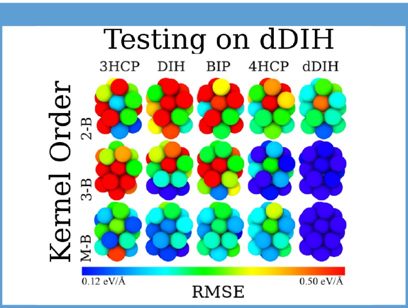
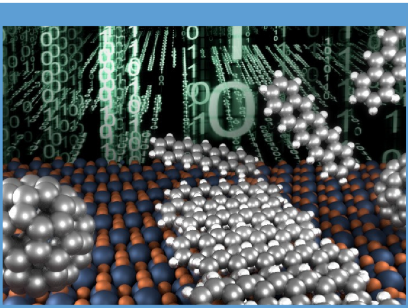
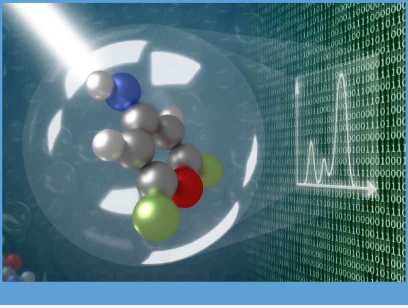
Young Researchers' Workshop on Machine Learning for Materials Science

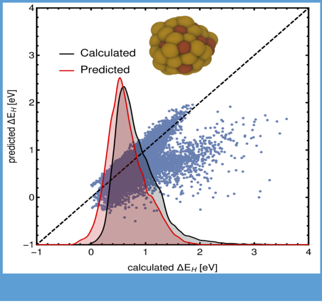
ML for MS

Abstract Book









6–10 May 2019
Aalto University, Helsinki, Finland

Contents

General 3
Organizers . . . 3
Venues . . . 3
Sponsors . . . 3
Hands-on computational tutorial . . . 4
Poster session and Workshop dinner . . . 5

Programme 6

Monday 7

Tuesday 10

Wednesday 11

Thursday 21

Friday 34

Posters 41

Index 88

General

Organizers

Milica Todorović, Aalto University, Finland
Claudio Zeni, King's College London, UK
Kevin Rossi, École polytechnique fédérale de Lausanne, Switzerland
Aldo Glielmo, King's College London, UK
Adam S. Foster, Aalto University, Finland
Patrick Rinke, Aalto University, Finland

Venues

Main venue: The Stage, Aalto Design Factory, Betonimiehenkuja 5C, 02150 Espoo
CSC tutorial: Keilaranta 14, 02150 Espoo
Poster session: Aalto School of Science, Otakaari 1, 02150 Espoo

Sponsors

Aalto University
School of Science

Aalto University
Digi Platform

Aalto-yliopisto
Design Factory

cecam
Centre Européen de Calcul Atomique et Moléculaire

Aalto University
Materials Platform

CSC

www.q-chem.org
International Journal of
QUANTUM CHEMISTRY

Journal of
COMPUTATIONAL CHEMISTRY
www.c-chem.org

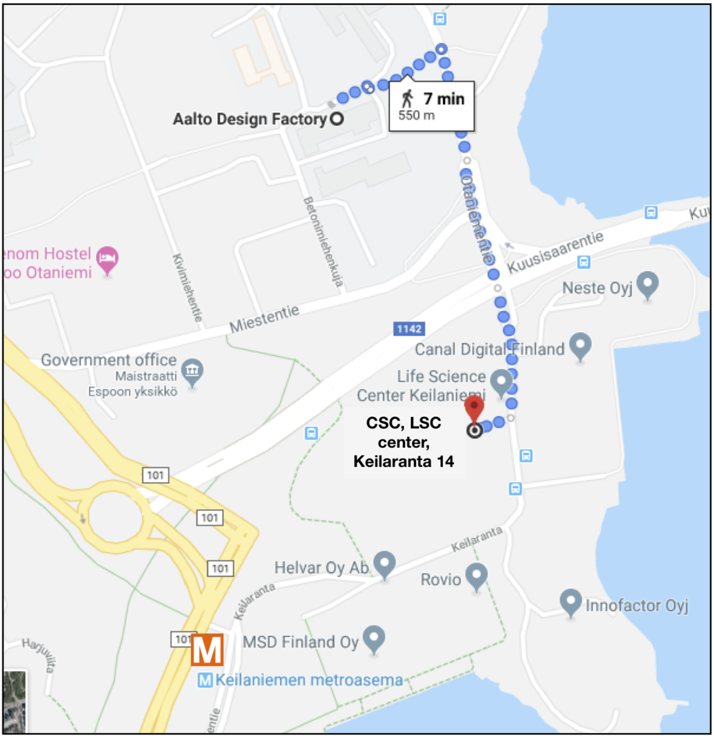
Hands-on computational tutorial

The hands-on computational tutorial will be held at two locations: the CSC building (Keilaranta 14, Espoo) and the ADF (Betonimiehenkuja 5C, 02150 Espoo) at one of the three venues:

- CSC Dogmi room: live instruction (recorded), 24 participants, workstations available.
- CSC Debatti room: webcast instruction, 20 participants, own laptop required.
- ADF Stage room: webcast instruction, all participants, own laptop required.

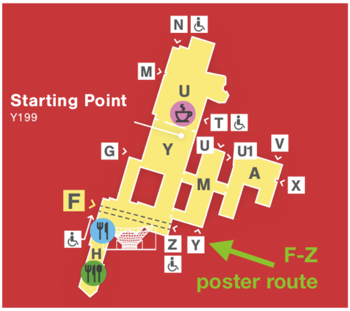
The CSC building can be accessed within a 10min walk from the ADF venue, or the Keilaniemi metro station.

Route to the CSC tutorial venue



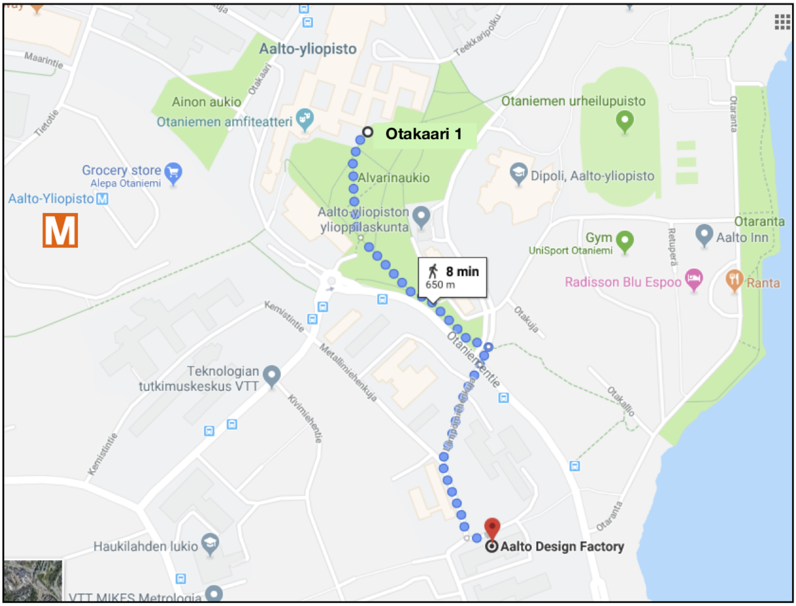
Poster session and Workshop dinner

The poster session will be held on Wednesday, 8 May, 17.30-20h at the Otakaari 1 building. Designed in 1965 by Alvar Aalto, it still serves as the Undergraduate Centre of Aalto University. It is home to Aalto School of Science and the Department of Applied Physics.

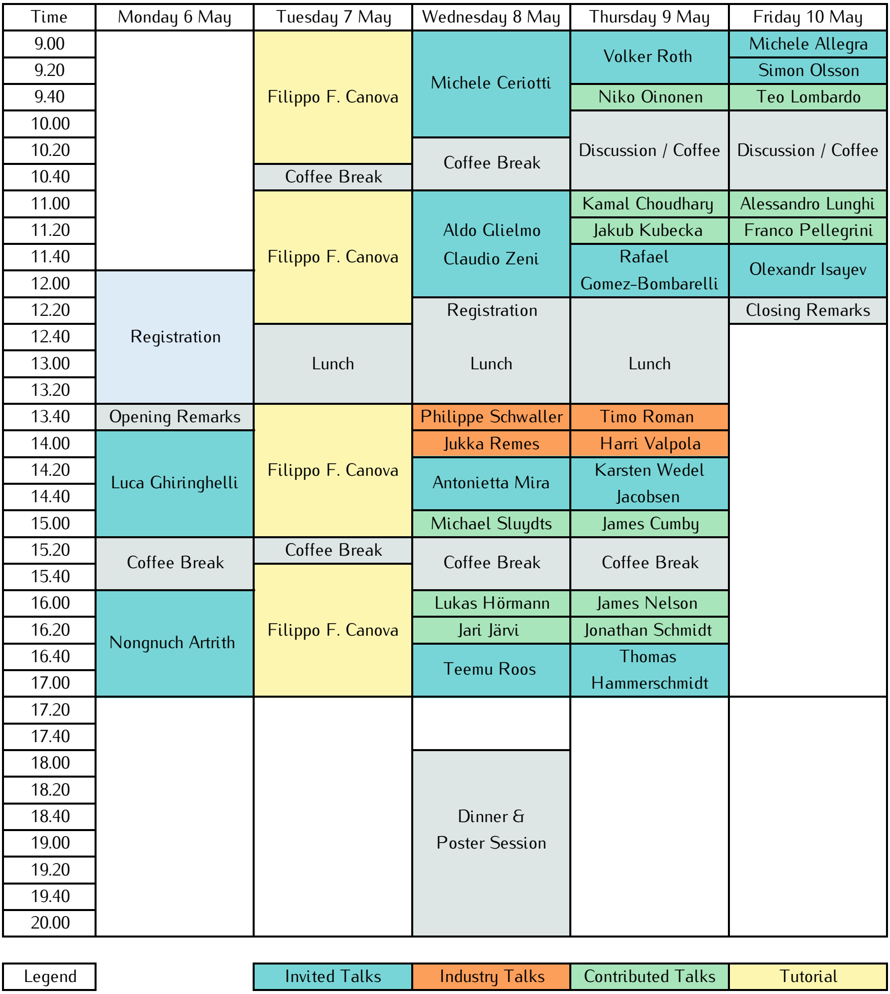


During the poster session, a warm dinner and drinks will be served at building entrance Z. The posters will be displayed along the F-Z route in the image above.

Route to the poster session venue



Programme



Time	Monday 6 May	Tuesday 7 May	Wednesday 8 May	Thursday 9 May	Friday 10 May
9.00		Filippo F. Canova	Michele Ceriotti	Volker Roth	Michele Allegra
9.20					Simon Olsson
9.40				Niko Oinonen	Teo Lombardo
10.00				Discussion / Coffee	Discussion / Coffee
10.20			Coffee Break		
10.40		Coffee Break			
11.00		Filippo F. Canova	Aldo Glielmo Claudio Zeni	Kamal Choudhary	Alessandro Lunghi
11.20				Jakub Kubecka	Franco Pellegrini
11.40				Rafael Gomez-Bombarelli	Olexandr Isayev
12.00	Registration				
12.20			Registration	Lunch	Closing Remarks
12.40		Lunch	Lunch		
13.00					
13.20					
13.40	Opening Remarks	Filippo F. Canova	Philippe Schwaller	Timo Roman	
14.00	Luca Ghiringhelli		Jukka Remes	Harri Valpola	
14.20			Antonietta Mira	Karsten Wedel Jacobsen	
14.40					
15.00			Michael Sluydts	James Cumby	
15.20	Coffee Break	Coffee Break	Coffee Break	Coffee Break	
15.40		Filippo F. Canova			
16.00	Nongnuch Artrith		Lukas Hörmann	James Nelson	
16.20			Jari Järvi	Jonathan Schmidt	
16.40			Teemu Roos	Thomas Hammerschmidt	
17.00					
17.20					
17.40					
18.00			Dinner & Poster Session		
18.20					
18.40					
19.00					
19.20					
19.40					
20.00					

Legend		Invited Talks	Industry Talks	Contributed Talks	Tutorial

Monday

12:00 - 13:40	Registration
13:40 - 14:00	Opening Remarks
14:00 - 15:20	**Luca Ghiringhelli** Intro Lecture: Exploring the materials space via regularized and symbolic regression
15:20 - 16:00	Coffe Break
16:00 - 17:20	**Nongnuch Artrith** Intro Lecture: Development of efficient and accurate artificial neural network potentials for the simulation of complex materials

Exploring the materials space via regularized and symbolic regression

Luca Ghiringhelli[1]

[1] *Fritz Haber Institut, Berlin, Germany*

The number of possible materials is practically infinite, while only few hundred thousands of (inorganic) materials are known to exist and for few of them even basic properties are systematically known. In order to speed up the identification and design of new and novel optimal materials for a desired property or process, strategies for quick and well-guided exploration of the materials space are highly needed. A desirable strategy would be to start from a large body of experimental or theoretical data, and by means of artificial-intelligence methods, to identify yet unseen patterns or structures in the data. This leads to the identification of maps (or charts) of materials where different regions correspond to materials with different properties. The main challenge on building such maps is to find the appropriate descriptive parameters (called descriptors) that define these regions of interest. Here, I will present methods for the machine-aided identification of descriptors and materials maps, which exploit in a novel fashion old workhorses of statistical analysis such as symbolic and regularized regressions, and dimensionality reduction. I will also describe relevant applications and the challenge of having open-access infrastructures for storing and accessing curated data.

Development of efficient and accurate artificial neural network potentials for the simulation of complex materials

Nongnuch Artrith[1,2]

[1] *Department of Chemical Engineering, Columbia University, New York, NY 10027, USA*
[2] *Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY 11973, USA*

Many complex materials for energy applications such as heterogeneous catalysts and battery cathode materials have compositions with multiple chemical species and properties that are determined by complex structural features. This complexity makes them challenging to model directly with first principles methods. As an alternative, machine-learning techniques can be used to interpolate first principles calculations. Such machine-learning potentials enable linear-scaling atomistic simulations with an accuracy that is close to the reference method at a fraction of the computational cost. Here, I will give an overview of the fundamentals of artificial neural network (ANN) potentials [1] and recent applications to the modeling of challenging materials classes, e.g., nanoalloys in solution [2], oxide nanoparticles [3], and amorphous materials [4, 5, 6]. The original multi-species ANN potential formalism [7] scales quadratically with the number of chemical species. This has previously prevented the modeling of compositions with more than a few elements. To overcome this limitation, we have recently developed an alternative mathematically simple and computationally efficient descriptor with a complexity that is independent of the number of chemical species [8,9]. The new methodology has been implemented in our free and open source atomic energy network (aenet) package (http://ann.atomistic.net) [9]. This development creates new opportunities for the modeling of complex materials for example for catalysis and for energy applications.

Tuesday

09:00 - 10:20	Hands on tutorial
10:20 - 11:00	Coffee Break
11:00 - 12:40	Hands on tutorial
12:40 - 13:40	Lunch
13:40 - 15:20	Hands on tutorial
15:20 - 15:40	Coffee Break
15:40 - 17:20	Hands on tutorial

Wednesday

09:00 - 10:20	**Michele Ceriotti** Intro Lecture: Machine-learning - more than potentials
10:20 - 11:00	Coffee Break
11:00 - 12:20	**Aldo Glielmo & Claudio Zeni** Intro Lecture: Gaussian processes for forcefields
12:20 - 13:20	Lunch
13:20 - 13:40	Opening Remarks
13:40 - 14:00	**Philippe Schwaller** IBM RXN for Chemistry: Predicting Chemical Reactions using the Molecular Transformer
14:00 - 14:20	**Jukka Remes** IBM PowerAI/Watson Machine Learning (WML) Accelerator - infrastructure for open source -based scalable and productive deep learning
14:20 - 15:00	**Antonietta Mira** Approximate Bayesian Computation to calibrate force-fields
15:00 - 15:20	**Michael Sluydts** Locating favorable point defects in silicon using active learning
15:20 - 15:40	Coffee Break
15:40 - 16:00	**Lukas Hörmann** Getting the Most out of Data: Smart-Data Machine-Learning for Surface Structure Search
16:00 - 16:20	**Jari Järvi** Detecting Stable Surface Adsorbates with Bayesian Optimization
16:20 - 16:40	**Teemu Roos** An overview of the goals and achieved progress in the MachQu project
17:40 - 20:20	Dinner + poster session

Wednesday

Machine-learning - more than potentials

Michele Ceriotti[1]

[1] *Institute of Materials, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.*

I will discuss machine learning approaches for atomistic modeling, taking a somewhat broader perspective than what is typically done when using ML to construct interatomic potentials. I will start by briefly reviewing the construction of feature vectors that fulfill the fundamental physical symmetries, and how these can be extended to allow for a description of properties (response tensors, density fields ...) that do not transform simply as scalars. I will also briefly discuss how some unsupervised learning techniques can be beneficial to reduce the computational cost associated with machine learning, by identifying the most important features, and the structures that are affected by the largest uncertainties in the predictions.

Gaussian processes for forcefields

Aldo Glielmo[1],Claudio Zeni[1]

[1] *Department of Physics, King's College London, UK*

We present the formalism behind Gaussian process regression and explain how this can be used in the context of force field generation from ab-initio databases. We display recent results on both bulk materials and nanoparticle systems, and showcase a method used to speed up simulations based on GP force fields by a significant factor.

IBM RXN for Chemistry: Predicting Chemical Reactions using the Molecular Transformer

Philippe Schwaller[1, 2], Teodoro Laino[1], Theophile Gaudin[1], Peter Bolgar[3], Costas Bekas[1], Alpha A. Lee[2]

[1] *IBM Research GmbH, Rueschlikon, Switzerland.*
[2] *Department of Physics, University of Cambridge, UK*
[3] *Department of Chemistry, University of Cambridge, UK*

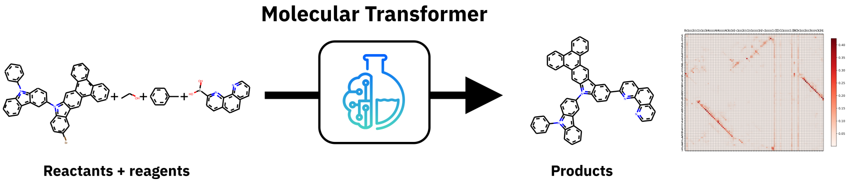


Figure 1: Predicting Chemical Reactions using the Molecular Transformer

Organic synthesis – making complex molecules out of simpler building blocks – is a crucial part in the production and therefore, also in the discovery of novel molecules and materials. One necessary yet unsolved step in synthesis planning is solving the forward problem: given reactants and reagents, predict the products. We treat chemical reaction prediction as a sequence-2-sequence problem, similar to neural machine translation, and learn to transform the SMILES strings of the reactants-reagents into the ones of the products.

We show that a multi-head attention Molecular Transformer model outperforms all algorithms in the literature, achieving a top-1 accuracy above 90% on a common benchmark dataset. Our algorithm requires no handcrafted rules, and accurately predicts subtle chemical transformations. Crucially, our model can accurately estimate its own uncertainty. Furthermore, we show that model is able to handle inputs without reactant-reagent split and including stereochemistry, which makes our method universally applicable across existing datasets.

A fully-trained model is freely available on the IBM RXN for Chemistry platform (https://rxn.res.ibm.com). Using this platform, chemists can simply draw reactants and reagents and predict the most likely outcomes of the reaction.

[1] Schwaller, P., Gaudin, T., Lanyi, D., Bekas, C. & Laino, T. "Found in Translation": Predicting outcomes of organic chemistry reactions using neural sequence-to-sequence models. Chemical Science (2018)

[2] Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Bekas, C., Lee, A. A. Molecular Transformer for Chemical Reaction Prediction and Uncertainty Estimation, http://dx.doi.org/10.26434/chemrxiv.7297379

IBM PowerAI/Watson Machine Learning (WML) Accelerator - infrastructure for open source -based scalable and productive deep learning

Jukka Remes[1],

[1] *IBM*

Deep learning (DL) can be often burdensome business (albeit interesting and at the heart of current AI development). In addition to data(+annotation) -related issues, one needs to cope with a) acquiring adequate HW capabilities for training, b) setting up versatile software environments and c) handling and running different DL experiments. A main question boils down to how much time needs to be used for setting up deep learning pre-requisites and how much time is available for actual model development. In this industry talk I will go through deep learning -environments offered on top of IBM Power hardware technology that utilize latest NVIDIA GPUs but also provide additional powerful framing around them. IBM solutions feature uniquely high data throughput rates that give significant training speed improvements. When combined with supported open source DL-frameworks (e.g. Tensorflow, Caffe, Pytorch), they enable training also models that do not fit into GPU-memory. IBM PowerAI/WML Accelerator -offering is also designed to scale across cluster effectively and includes software stack that provides different adoption alternatives wrt different usage preferences: these systems can be used similarly as any regular Linux -based Python -development environments for deep learning but provide also higher-level environments for assisted data science and e.g. machine vision & AutoML (PowerAI Vision). The idea is to improve productivity in multiple ways when utilizing deep learning.

Approximate Bayesian Computation to calibrate force-fields

Antonietta Mira[1, 2], Ritabrata Dutta[3], Zacharias Faidon Brotzakis[4]

[1] *Università della Svizzera italiana*
[2] *University of Insubria*
[3] *University of Warwick*
[4] *ETH Zurich*

I will introduce the Bayesian formalism and Approximate Bayesian Computation (ABC) a very powerful likelihood free algorithm that allows to conduct sound statistical inference (point estimation, credible intervals, hypothesis testing and model selection), even in settings where the likelihood is not available analytically but can be used to generate pseudo-data by forward simulation. ABC will be used for the calibration of force-field parameters and for their uncertainty quantification, without assuming parameter uncertainty to be Gaussian. I will illustrate the performance of the developed methodology to learn the Lennard-Jones force-field parameters of helium and TIP4P system of water implemented both for simulated and experimental datasets collected using Neutron and X-ray diffraction. For simulated data, the Bayesian estimate is in close agreement with the true parameter value used to generate the dataset. For experimental as well as for simulated data, the Bayesian posterior distribution shows a strong correlation pattern between the force-field parameters. Providing an estimate of the entire posterior distribution, our methodology also allows us to perform uncertainty quantification of model prediction. This research opens up the possibility to rigorously calibrate force-fields from available experimental datasets of any structural and dynamic property.

Locating favorable point defects in silicon using active learning

Michael Sluydts[1, 2, 3], Michiel Larmuseau[1, 2], Stefaan Cottenier[1, 2]

[1] *Center for Molecular Modeling, Ghent University, Belgium*
[2] *Department of Electrical Energy, Metals, Mechanical Constructions & Systems, Ghent University, Belgium*
[3] *ePotentia, Belgium*

Thanks to modern automation tools, it has become relatively easy to create large-scale ab initio materials datasets. These datasets are capable of providing us with important insights in large search spaces, which could never be investigated experimentally. One example is the case of point defects in silicon. Both native defects and impurities can greatly alter the electronic behavior of the silicon wafer. Defect control thus remains an important parameter in the production process of any electronic material. Creating a database of individual point defects is no problem, even when considering the entire periodic table.[1] Point defects are, however, not fixed within the crystal and, especially at higher temperatures, can interact with eachother. When taking into account the possible configurations of two defects the number of possible configurations quickly skyrockets. Given the non-negligible computation time of a single calculation this quickly becomes problematic, especially considering possible long-range interactions. Within this talk we represent a simple example, illustrating how one can use active learning to direct the screening procedure to the most important cases, as well as quantifying the uncertainty in other regions.

[1] M Sluydts, M Pieters, J Vanhellemont, V Van Speybroeck, S Cottenier Chemistry of Materials 29 (3), 975-984

Getting the Most out of Data: Smart-Data Machine-Learning for Surface Structure Search

Lukas Hörmann[1], Andreas Jeindl[1], Alexander T. Egger[1], Oliver T. Hofmann[1]

[1] *Institute of Solid State Physics, Graz University of Technology, Austria*

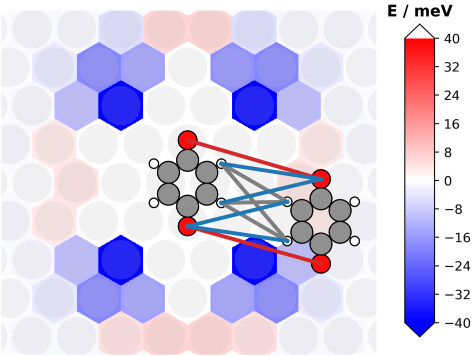


Figure 1: Interaction energies of two benzoquinone molecules at different respective positions. The lines illustrate the feature assigned to a specific molecule-molecule interaction.

The key challenge for surface structure prediction is the enormous number of possible polymorphs. Hence determining the lowest energy surface polymorph with conventional stochastic algorithms would require an insurmountable computational effort. We show that this can be overcome using the quasi-deterministic SAMPLE approach, which is based on smart-data machine-learning and experimental design theory. Only a few hundred DFT calculations of polymorphs suffice to exhaustively predict the local minima of the potential energy surface and reliably identify the global minimum.

To gain a maximum of physical insight from a limited amount of data, SAMPLE employs an energy model based on molecule-substrate and molecule-molecule interactions. Using Bayesian linear regression, it extracts these interactions directly from formation energies of polymorphs, which we calculate using dispersion corrected DFT. For this purpose, SAMPLE assigns each molecule-molecule interaction a feature based on atom distances. The features differentiate between different atom species, enabling us to separate the contributions of various molecule-fragments to the total energy of the polymorph. This allows peering deep into the interactions between individual components of the system, yielding knowledge about the physics that determine why a particular polymorph forms. This knowledge opens up the exciting possibility of designing polymorphs with highly desirable properties.

Detecting Stable Surface Adsorbates with Bayesian Optimization

Jari Järvi[1], Milica Todorovic[1], Patrick Rinke[1]

[1] *Department of Applied Physics, Aalto University, Finland*

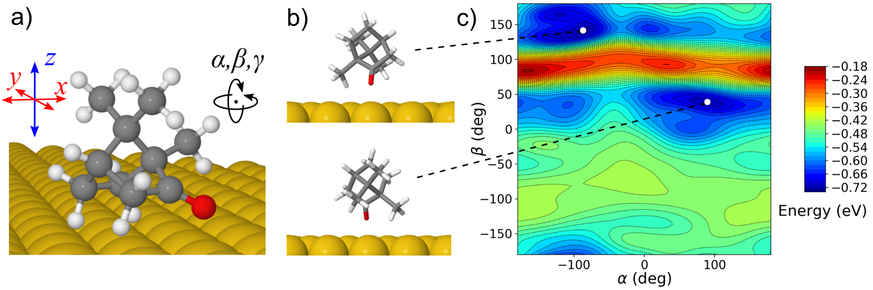


Figure 1: a) Orientational and translational search for a stable adsorption structure of camphor on Cu(11). b) Orientation of camphor in the global energy minimum (top) and in the lowest local minimum (bottom), identified in c) the 2-dimensional cross section through the 3-dimensional energy landscape with 125 sampled configurations (5 samples per dimension).

Reliable identification of stable surface adsorbates requires thorough exploration of the potential energy surface (PES) of adsorption. Simulating organic molecules on inorganic surfaces, however, is prohibitively expensive with conventional quantum mechanical methods, such as density-functional theory (DFT). Instead, traditional structure search methods have relied on chemical intuition, focusing only on the likely minimum-energy structures. At the organic-inorganic interface, this intuition is difficult to apply and can lead to biased and incorrect results.

We combine DFT simulations and artificial intelligence (AI) to resolve stable surface adsorbate structures. We apply the Bayesian Optimization Structure Search (BOSS) method [1] to study the adsorption of camphor ($C_{10}H_{16}O$) on a Cu(111) surface. BOSS is a new AI tool, which accelerates the structure search via an intelligent and unbiased sampling of the PES. BOSS minimizes the number of required energy computations and determines the complete PES for a clear identification of the most stable structure.

In this study, we first analyze camphor conformers with a 3-dimensional search of methyl group rotations. We then employ the identified minimum-energy structure of the molecule to investigate its adsorption on a Cu(111) surface (Fig. 1a). By combining QM simulations and AI, we identify the stable adsorbate structures (Figs. 1b and 1c) and the associated energy barriers for molecular rotation and translation.

[1] M. Todorovic, M. Gutmann, J. Corander, and P. Rinke, Efficient Bayesian Inference of Atomistic Structure in Complex Functional Materials, accepted in npj Comput. Mater. (2017), arXiv:1708.09274

An overview of the goals and achieved progress in the MachQu project

Teemu Roos[1],

[1] *Department of Computer Science, University of Helsinki, Finland*

The MachQu project will develop new machine learning methods to develop new efficient quantum mechanics-based simulation tools. These tools are a key to understanding atomic migration processes in a complex configuration space determined by potential energy barriers. The results of the materials simulations can be used for knowledge-based design of complex alloyed materials that can withstand extreme conditions of fusion reactors and particle accelerators.

Thursday

09:00 - 09:40	**Volker Roth** Deep latent variable models for exploring the chemical space
09:40 - 10:00	**Niko Oinonen** Automated Structure Discovery in Atomic Force Microscopy
10:00 - 11:00	Coffee + discussion
11:00 - 11:20	**Kamal Choudhary** JARVIS-ML: Physics inspired AI for fast and accurate screening of materials
11:20 - 11:40	**Jakub Kubečka** Molecular clusters: configurational sampling and global minimum search
11:40 - 12:20	**Rafael Gomez-Bombarelli** Coarse-graining atomistic simulations with deep autoencoders.
12:20 - 13:40	Lunch
13:40 - 14:00	**Timo Roman** Towards autonomous driving with deep learning
14:00 - 14:20	**Harri Valpola** Combining machine learning and reasoning tasks
14:20 - 15:00	**Karsten W. Jacobsen** Predicting material properties without knowing where the atoms are.
15:00 - 15:20	**James Cumby** Ellipsoids as a coordination descriptor
15:20 - 15:40	Coffee break
15:40 - 16:00	**James Nelson** Machine Learning Many-Body Models
16:00 - 16:20	**Jonathan Schmidt** Machine learning the exchange-correlation functional of DFT: a Jacob's elevator through modern neural networks architecture
16:20 - 17:00	**Thomas Hammerschmidt** Machine-learning material properties with electronic-structure based descriptors

Deep latent variable models for exploring the chemical space

Volker Roth[1],

[1] *Department of Mathematics and Computer Science, University of Basel, Switzerland*

Latent variable models are powerful statistical models that relate observations to unobserved latent variables. Most models of this kind admit a probabilistic interpretation in which the latent structure reflects our assumption about the underlying data generation process. The combination of ”classical” latent variable models with highly nonlinear deep neural network architectures is commonly referred to as ”deep latent variable models”. We show that in the context of Material Science, such nonlinear models are highly useful for addressing the problem of predicting molecular properties, as well as the inverse problem of suggesting novel molecules that share some properties. We discuss the general structure of such models, certain modeling and algorithmic challenges, and possible applications. In particular, we use these model in an explorative way to gain a better understand of the space of molecules.

Automated Structure Discovery in Atomic Force Microscopy

Benjamin Alldritt[1], Prokop Hapala[1], Niko Oinonen[1], Fedor Urtev[1, 2], Ondrej Krejci[1], Filippo Federici Canova[1, 3], Fabian Schultz[4], Juho Kannala[2], Peter Liljeroth[1], Adam Foster[1, 5, 6]

[1] *Department of Applied Physics, Aalto University, Finland*
[2] *Department of Computer Science, Aalto University, Finland*
[3] *Nanolayers Research Computing Ltd, UK*
[4] *IBM Research—-Zurich, Switzerland*
[5] *Graduate School Materials Science in Mainz, Germany*
[6] *WPI Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Japan*

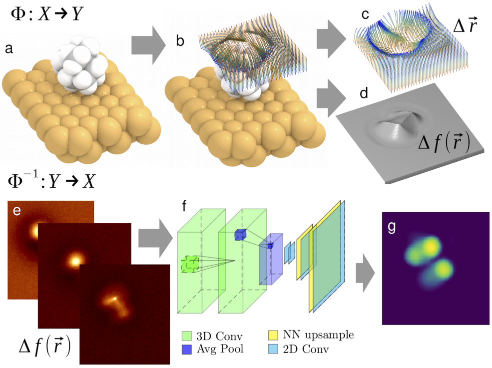


Figure 1: Schematic illustration of molecular recognition as an inverse imaging process.

Non-contact atomic force microscopy (AFM) with molecule-functionalized tips has emerged as the primary experimental technique for probing the atomic structure of organic molecules on surfaces. Most experiments have been limited to nearly planar aromatic molecules, due to difficulties with interpretation of highly distorted AFM images originating from non-planar molecules. Here we develop a deep learning infrastructure that matches a set of AFM images with a unique descriptor characterizing the molecular configuration, allowing us to predict the molecular structure directly. We apply this methodology to resolve experimentally several distinct adsorption configurations of 1S-camphor on Cu(111) based on low-temperature AFM measurements. This approach will open the door to apply high resolution AFM to a huge variety of systems where routine atomic and chemical structural resolution on the level of individual objects/molecules can be a major breakthrough.

JARVIS-ML: Physics inspired AI for fast and accurate screening of materials

Kamal Choudhary[1],

[1] *Materials measurement laboratory, National Institute of Standards and Technology, USA*

In this work, we developed a complete set of chemo-structural descriptors to significantly extend the applicability of machine-learning (ML) in material screening for multicomponent systems. These new descriptors allow differentiating between structural prototypes, which is not possible using the commonly used chemical-only descriptors. We developed ML models for formation energies, bandgaps, static refractive indices, magnetic properties, modulus of elasticity, solar-cell efficiency, k-point integration grid and plane-wave cutoffs for 3D materials as well as exfoliation energies of two-dimensional (2D) layered materials. We used a gradient boosting decision tree (GBDT) algorithm and the training data consisted of 24549 bulk and 616 monolayer materials taken from JARVIS-DFT database. Additionally, we integrate our formation-energy ML model with a genetic algorithm for structure search. Our learned model is publicly available on the website (https://www.ctcms.nist.gov/jarvisml).

[1] Kamal Choudhary, Brian DeCost, and Francesca Tavazza, Phys. Rev. Materials 2, 083801 (2018)

Molecular clusters: configurational sampling and global minimum search

Jakub Kubečka[1], Vitus Besel[1], Hanna Vehkamäki[1], Theo Kurtén[1]

[1] *INAR - Institute for Atmospheric and Earth System Research, University of Helsinki, Finland*

Theoretical studies of atmospherically relevant molecular clusters require knowledge of cluster properties such as energies, structures, etc. In the past, the easiest way to construct the structure of a molecular cluster was assembling molecules manually based on chemical intuition. We studied one of the possible paths of a systematical search for the global minimum using a “build approach”; exploring the potential energy surface by a genetic algorithm and further reoptimize, filter or sample in several steps. The configurational sampling is discussed also in the context of machine learning and the main obstacles of sampling are mentioned.

Finally, we present several atmospherically relevant molecular clusters containing molecules such as sulphuric acid, ammonia, guanidine, highly oxygenated molecules etc. The clusters are discussed in the context of New Particle Formation (NPF) in the atmosphere. In our work, we also use Atmospheric Cluster Dynamics Code (ACDC) and present several results to emphasize the importance of a proper configurational sampling process. Moreover, we have developed the Jammy Key for Configurational Sampling (JKCS) which is a program that operates with files (structures, outputs etc.) and 3rd-party computational programs needed to explore Potential Energy Surface of molecular clusters and search for a global minimum and also energetically low-lying local minima.

Coarse-graining atomistic simulations with deep autoencoders.

Rafael Gomez-Bombarelli[1],

[1] *Department of Materials Science and Engineering, Massachusetts Institute of Technology*

Unsupervised learning approaches have proven very successful at learning complex statistical distributions: human faces, poetry, music, computer code, molecular structures, etc. Furthermore, the combination of deep learning algorithms with physical simulations are producing exciting results in quantum mechanics, statistical mechanics, fluid dynamics, etc, both by building fast proxies for expensive simulations and by explicitly including physics-based priors in deep neural networks.

Molecular dynamics simulations in soft matter can predict many relevant properties for understanding and designing materials. However, they are typically hindered by the statistical complexity of sampling high-dimensional potential energy surfaces and by the short time steps needed to track fast motions. Coarse graining methods address those by building reduced-dimension systems that can be simulated at larger length- and time-scales. Here, we will describe how semi-supervised deep autoencoders can be used to parametrize and execute coarse-grained simulations. Trained on all-atom trajectories, the learning framework automatically identifies the optimal coarse graining function and the potential of mean force that governs its dynamics.

Towards autonomous driving with deep learning

Timo Roman[1],

[1] *NVIDIA, Helsinki, Finland*

Cars are becoming computers on wheels and getting those to drive autonomously is currently a very active and challenging area of research and development. Artificial deep neural networks have revolutionized computer vision and will soon allow taking a leap forward towards autonomous driving. Deep learning is at the core of autonomous driving solutions developed by NVIDIA, which GPUs play a central role in by accelerating neural network training as well as real-time inference on embedded automotive platforms. This talk will provide an overview of how NVIDIA applies deep learning and AI at scale to revolutionize perception for self-driving cars.

Combining machine learning and reasoning tasks

Harri Valpola[1],

[1] *Curious AI*

Curious AI has developed reasoning technology for control and decision making problems. This talk introduces the technology through example applications and explores how it could be applied to design tasks in material sciences.

Predicting material properties without knowing where the atoms are.

Karsten W. Jacobsen[1],

[1] *Department of Physics, Technical University of Denmark, Denmark*

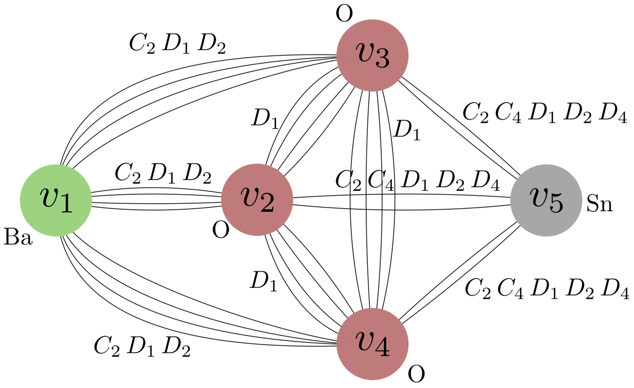


Figure 1: Symmetry labeled crystal graph for a cubic perovskite.

Computational materials screening studies require calculations of the properties of many materials typically performed using Density Functional Theory. The size of the materials space that can be investigated is therefore limited by the computational speed of the Density Functional Theory calculations. Machine learning can potentially speed-up screening studies by several orders of magnitude by either replacing the DFT calculations or by guiding the search in the materials space.

In the talk I shall discuss different possibilities for machine-learning assisted computational screening studies, and in particular I shall address the question of how to search for new materials where the atomic positions are not known in advance. This is an important aspect of finding new materials: if the atomic positions of a material are already known they are usually obtained either from experiment – in which case the material is not really new – or from a Density Functional Theory calculation – in which case the calculation has already been performed, so there is no need for machine learning. We therefore need to describe materials without detailed atomic positions and I shall discuss one approach to this using symmetries and Voronoi crystal graphs.

Ellipsoids as a coordination descriptor

James Cumby[1], J. Paul Attfield[1]

[1] *Centre for Science at Extreme Conditions and School of Chemistry, University of Edinburgh, UK*

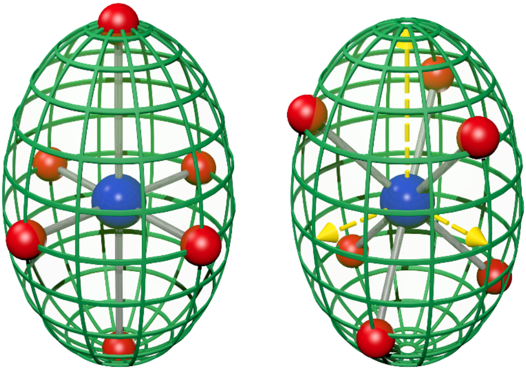


Figure 1: Ellipsoidal fits to two distorted octahedra; tetragonally elongated (left) and trigonally distorted (right).

Coordination polyhedra are prevalent across materials chemistry; in fact, many structure-property relationships can be described in terms of interactions between rigid coordination units, such as perovskite phase transitions and negative thermal expansion. Generally polyhedra are not rigid, instead displaying distortion; analysing this can be difficult, particularly if the polyhedron deviates both in bond lengths and angles. Local symmetry mode analysis can give a comprehensive structural picture in many cases, but is often difficult to interpret if multiple distortions act cooperatively. Alternatively, continuous shape measures [1] can be used to quantify deviations from a reference polyhedron, but this is not always trivial to define. Here, we present a much simpler measure of polyhedron distortion, based on the fitting of ellipsoids to coordination environments.

We have applied this new general method to a range of metal oxide materials, revealing new insights into their properties; from a switching bistability in symmetry-disallowed perovskite phase transitions to a weak off-centre 'd^5-effect' for Fe^{3+} ions, that could potentially be exploited in multiferroic materials.[2] The method is available as the python package PIEFACE.

While the method has currently been applied to metal oxides, it could easily be generalised to other materials that lend themselves to a polyhedral description, including coordination complexes, framework materials and organic molecules.

[1] S. Alvarez et al, Coord. Chem. Rev., 249, 1693, 2005.

[2] J. Cumby and J. P. Attfield, Nat. Commun. 8, 14235, 2017.

Machine Learning Many-Body Models

James Nelson[1], Stefano Sanvito[1]

[1] *School of Physics, AMBER and CRANN Institute, Trinity College, Dublin 2, Ireland*

The exact solution of many-body models for a large number of atoms is often computationally expensive or at worst completely unfeasible, due to the enormous size of the relevant basis. In the last few years, there has been much interest in applying Machine Learning (ML) to many-body systems and on the face of it, this seems like an ideal candidate problem for ML. Recently, we have demonstrated that ML can be used to find Density Functionals for such many-body models [1]. The next step is to show that ML can offer a speed up or compute previously intractable quantities. Here we now offer a potential methodology for utilising information from small cheaply calculated systems to predict quantities for larger systems using ML. Specifically, we are interested in predicting the ground state energy and learning thermodynamic functions of the disordered one-dimensional Hubbard Model.

[1] Nelson, J., Tiwari, R. and Sanvito, S., 2019. Machine learning density functional theory for the Hubbard model. Physical Review B, 99(7), p.075132.

Machine learning the exchange-correlation functional of DFT: a Jacob's elevator through modern neural networks architecture

Jonathan Schmidt[1], Carlos Benavides Riveros[1], Miguel Marques[1]

[1] *Institut für Physik, Martin-Luther-Universität Halle-Wittenberg, Germany*

Using the automatic differentiation of modern machine-learning techniques, we develop a new neural network functional for the exchange-correlation energy and the exchange-correlation potential of density functional theory (DFT). Training the neural network with exact results for a large set of 2-electrons systems in 1D, we demonstrate that the functional achieves great accuracy in self-consistent calculations for systems in the test set as well as other molecular systems. We investigate the locality and various other important properties of the exchange-correlation functional of DFT and compare the properties of the neural network with traditional DFT functionals (e.g. LDA). Using a local functional allows for a simple extensive scaling to larger systems and drastically reduces the size and complexity of the neural network. Furthermore, by changing the degree of locality one can make the functional flexible to the required accuracy for an application.

Machine-learning material properties with electronic-structure based descriptors

Thomas Hammerschmidt[1],

[1] ICAMS, Ruhr University Bochum, Germany

Machine learning the properties of materials relies on descriptors of local atomic environments. In a recent work [1] we introduced a descriptor that is based on the moments of the local electronic density of states (DOS). The moments are explicitly linked to the crystal structure and the chemistry through the moments theorem [2]. This makes it possible to construct electronic-structure based descriptors of the local atomic environment that have an immediate relation to the binding energy. In order to rule the descriptors suitable for practical machine-learning applications, we determine the moments in a numerically efficient way by coarse-grained electronic-structure methods. In particular, we use tight-binding Hamiltonians and approximate solutions for the corresponding local DOS from analytic bond-order potentials. With this methodology we demonstrate that descriptors using only four moments are already sufficient to establish a structure-energy relation because the lowest moments, calculated from the closest atomic neighborhood, carry the largest contributions to the local bond energy. Taking into account more moments leads to highly competitive descriptors for machine-learning applications [3].

[1] Jenke et al., Phys. Rev. B, 98, 144102, 2018

[2] Cyrot-Lackmann, Adv. Phys. 16, 393, 1967

[3] Sutton et al., arXiv:1812.00085

Friday

09:00 - 09:20	**Michele Allegra** Clustering by the local intrinsic dimension
09:20 - 09:40	**Simon Olsson** Machine learning molecular kinetics and coarse-grained molecular dynamics forcefields
09:40 - 10:00	**Teo Lombardo** Deep Neural Network to guide the Coarse-Grained Molecular Dynamics parameterization
10:00 - 11:00	Coffee + discussion
11:00 - 11:20	**Alessandro Lunghi** General description of covalent bonds with machine-learning potentials
11:20 - 11:40	**Franco Pellegrini** PANNA - Properties from Artificial Neural Network Architectures
11:40 - 12:20	**Olexandr Isayev** Neural networks learning quantum chemistry
12:20 - 12:40	Closing remarks
12:40 - 13:40	Lunch

Clustering by the local intrinsic dimension

Michele Allegra[1],

[1] *Institut de Neurosciences de la Timone, Marseille, France*

Data defined in high-dimensional spaces often lie on a manifold of low intrinsic dimension (ID). This fact is well known, as it provides the main justification for dimensional reduction. What is less commonly known is that the ID can vary within the same dataset. This fact has been highlighted in technical discussions, but seldom exploited to gain insight into the data structure. We developed a simple and robust approach to cluster regions with the same local ID in a given data landscape. We found that several real-world data sets contain regions with widely heterogeneous dimensions. These regions host points differing in core properties: folded vs unfolded configurations in a protein molecular dynamics trajectory, active vs non-active regions in brain imaging data, and firms with different financial risk in company balance sheets. Our results show that a simple topological feature, the local ID, is sufficient to uncover a rich structure in high-dimensional data landscapes.

Machine learning molecular kinetics and coarse-grained molecular dynamics forcefields

Simon Olsson[1],

[1] *Freie Universität Berlin*

In this talk, I will outline recent results from our group related to the modeling of molecular kinetics of proteins as well as using neural network strategies to parameterize molecular dynamics forcefields. The first part of the talk will briefly outline two problems in molecular kinetics: accuracy of classical MD forcefields used to parameterize kinetic models and the growth of model complexity with the size of the molecular system. I will then present outline Augmented Markov models (Olsson et al PNAS 2017) as a remedy to improve the accuracy of models of molecular kinetics and Dynamic Graphical Models (Olsson & Noé bioRxiv) as a possible path towards treating larger molecular systems. In the second part of the talk, I will outline recent results leveraging the model capacity of artificial neural networks to learn coarse-grained forcefields from all-atom MD data (Wang, Olsson et al ACS Central Science, in press).

Deep Neural Network to guide the Coarse-Grained Molecular Dynamics parameterization

Teo Lombardo[1, 2, 3], Jean-Baptiste Hoock[3], Alain Ngandjong[1, 2], Alejandro A. Franco[1, 2, 4, 5]

[1] *Laboratoire de Réactivité et Chimie des Solides (LRCS)*
[2] *Réseau sur le Stockage Electrochimique de l'Energie (RS2E)*
[3] *Université de Picardie Jules Verne (UPJV)*
[4] *ALISTORE-European Research Institute*
[5] *Institut Universitaire de France (IUF)*

Today's Li-ion batteries (LIBs) have reached very high performance but they need further optimization in order to reach higher energy densities in particular regarding electric vehicle applications. The ARTISTIC project, funded by the European Research Council [1] aims at developing a multiscale modelling platform that can simulate the fabrication processes of LIB electrodes and that can predict their corresponding electrochemical performance. Within this platform, Coarse-Grained Molecular Dynamics (CGMD) is used to simulate slurries and the coating along the manufacturing process. In this presentation we discuss opportunities for the use of Deep Neural Networks to guide the CGMD parameterization and we present the overall adopted approach.

[1] ERC Project ARTISTIC: https://www.u-picardie.fr/erc-artistic/?L=0

General description of covalent bonds with machine-learning potentials

Alessandro Lunghi[1], Stefano Sanvito[1]

[1] *School of Physics, CRANN and AMBER, Trinity College Dublin*

The description of potential energy surfaces in terms of machine-learning algorithms represents the most promising approach to the development of universal and accurate force fields. This computational framework is revolutionizing the fields of chemistry and material science and in the last few years great results have been achieved in both the development of structural descriptors and regression methodologies.

In this contribution we show that a combination of ridge regression and bi-spectrum components as structural fingerprints, known as SNAP potentials [1], are able to describe the potential energy surface of molecular compounds in a very general fashion and with chemical accuracy (1 kcal/mol) [2]. We show applications for both organic and coordination compounds. In particular, we show that an accurate representation of the most common bonds encountered in transition metals chemistry is possible. This includes subtle features like Jahn-Teller distortion. We show that SNAP potentials are stable enough to be used for MD simulations and that they can be generated in an automatic way with affordable computational costs.

[1] A. P. Thompson et al., Journal of Computational Physics, 285, 316-330, 2015

[2] A. Lunghi and S. Sanvito, Submitted.

PANNA - Properties from Artificial Neural Network Architectures

Ruggero Lot[1], Franco Pellegrini[1], Yusuf Shaidu[1], Emine Kucukbenli[1]

[1] *SISSA - Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy*

We present PANNA, a package to train neural networks for the prediction of properties in material systems.

PANNA is a Python code based on TensorFlow, allowing fast model training with a simple interface. At the moment, PANNA supports local environment descriptors and training of feedforward neural network with species-resolved architectures, various activation functions and regularizations. Models can be trained and evaluated with the addition of forces, and they can be exported as empirical potentials to be used in LAMMPS simulations. The TensorFlow infrastructure allows PANNA to run efficiently on different architectures and distributed systems.

I will present the code and some early results of different systems and architectures, with particular emphasis on the use of minimal model and dataset size to obtain the best results with the least effort.

Neural networks learning quantum chemistry

Olexandr Isayev[1],

[1] *University of North Carolina at Chapel Hill, Chapel Hill, NC, USA*

In this talk, we will present a fully transferable deep learning potential that is applicable to complex and diverse molecular systems well beyond the training dataset. Recently we introduced ANAKIN-ME (Accurate NeurAl networK engINe for Molecular Energies) or ANI in short. ANI is a new method and sampling procedure for training neural network potentials that utilizes a special kind of symmetry functions to build single-atom atomic environment vectors (AEV) as a molecular representation.

The AI methods that focus on the use of large and diverse data sets in training new potentials, has consistently proven to be universally applicable to systems containing the atomic species in the training set. Focusing on parametrization for organic molecules (with CHNOSFCl atoms so far), we have developed a universal neural network potential which is highly accurate compared to reference QM calculations at speeds 10^7 faster. The potential is shown to accurately represent the underlying physical chemistry of molecules through various test cases including chemical reactions (both thermodynamics and kinetics), thermochemistry, structural optimization, and molecular dynamics simulations. The results presented in this talk will provide evidence of the universal applicability of deep learning to various chemistry problems involving atomistic simulations.

Posters

Steven Baksa
Photocatalytic Activity of Semiconducting Oxides and Machine-Learning Applications in Materials Discovery
Luca Bellucci
Amyloid peptides in solution and at the gold/water interface.
Vitus Besel
New Particle Formation Involving Charged Sulfuric Acid – Ammonia Clusters
Jesper Byggmästar
Gaussian Approximation Potentials for radiation damage in bcc metals
Juan Santiago Cingolani
Understanding the dynamics of graphene flakes on liquid Cu using molecular dynamics simulations
Anais Colibaba
Clustering of transition metal dichalcogenides
Karen D. Dedecker
Systematically identifying the collective variables describing phase transformations in flexible materials
Martin Deimel
Development of Descriptors for the Prediction of Adsorption Energies on Transition Metal Catalysts and their Alloys
Francesco Delfino
Proteins structural transition states explored with minimalist coarse grained models
Laia Delgado Callicó
Moving and Vibrating: Study of the Structural and Magnetic, Electronic and Vibrational Properties of Ni_{19}
Nirmal Ganguli
Two-dimensional antiferromagnetic spintronics at perovskite oxide interfaces
Danilo González
BCN Models: A User Friendly Tool to Generate of Stoichiometric Wulff Shaped Nanoparticles
Hamidreza Hajiani
Surface termination and composition control of activity of the Co_x Ni_{1-x} Fe_2 O_4 (001) surface for water oxidation: insights from DFT+U calculations
Soumyajyoti Haldar
Noncollinear spin density of an adatom on a magnetic surface
Ali Hamedani
Extending the applications of machine learning interatomic potentials: radiation damage analysis in silicon
Lauri Himanen
DScribe: A python library of descriptors for machine learning in materials science
Marc Jäger
Machine learning hydrogen adsorption on nanoclusters through structural descriptors
Venkat Kapil
Ab inito multiple time stepping using machine learning
Konstantin Karavaev
Estimating mechanical properties of Al-based alloys via High-throughput approach
Ilia Kichev
Assessment of similarity patterns by chemometric tools in clinical data of patients diagnosed with diabetes mellitus type 2
Michiel Larmuseau
Towards active learning using reliable uncertainty estimates
Shikun Li
Theoretical study on the structures and thermodynamics of A-type rare earth sesquioxides by HSE and PBE+U methods
Yunpei Liu

A Practice in Automation and Machine Learning about Chemistry

Teng Long

Importance of crystal representation in machine learning of thermodynamic and magnetic properties

Sabine Matysik

Adsorption, Decomposition and Surface Dynamics of Formic Acid on Cu531

Madhulika Mazumder

Role of Cation Doping in NMV (Na(3+x)M(x)V(2-x)(PO4)3 Type NASICON Systems to Regulate Phase Stability and Electrochemical Behaviour as Sodium Ion Battery Cathodes - A First Principles Study

Eamon McDermott

ML assisted phase contrast of STEM imaged HfO_2 memory filaments

Joaquin Miranda

Atomic ordering and transport properties in $Nb_x CoSb$ half-Heuslers

Olga Miroshnichenko

Oxidation states of binary oxides from data analytics of the electronic structure

Kim Andrea Nicoli

Deep Learning For Predicting and Analyzing Atomistic Systems

Rodrigo P. de Carvalho

A machine learning approach to predict redox potentials of organic materials for energy applications

Heesoo Park

Prediction of Octahedral Deformation in Perovskite by Applying to Machine Learning

Mihir Ranjan Sahoo

First-principles study of a vertical spin switch in atomic scale two-dimensional platform

Henri Salmenjoki

Machine learning plastic deformation of crystals

Andrea Silva

Multi-scale modelling of Ti-MoS_2 compounds

Murat Cihan Sorkun

Limits of using experimental labeling of materials on the ML development: a case study on aqueous solubility of compounds

Štěpán Sršeň

Towards Efficient Electronic Spectra Modelling with Machine Learning

Carsten Staacke

Combining polarizable electrostatic with a machine learned potential

Sina Stocker

Predicting Reaction Energetics with Machine Learning

Behnam Parsaeifard

Performance of various fingerprints in analyzing atomic environments

Matteo Peluso

Force Field Parameterization of Halide Anions by means of a Machine Learning Procedure

Paola Torche

Neural Network for fast material characterisation in Friction Force Microscopy

Simon Wengert

Looking for Needles in the Haystack: Improving Approximate Electronic Structure Methods via a Machine-Learning Correction for Reliable Crystal Structure Screening

Jingfang Xiong

Workflow Platform Framework and its Application

Yihuang Xiong

High-throughput investigation of band alignments of cubic perovskites in solvated environments

Photocatalytic Activity of Semiconducting Oxides and Machine-Learning Applications in Materials Discovery

Steven Baksa[1], Ismaila Dabo[1]

[1] *Department of Materials Science and Engineering & Materials Research Institute, The Pennsylvania State University, United States*

Artificial photosynthesis is a sustainable technological option to store solar energy through the photocatalytic conversion of carbon dioxide and water into chemical fuels. The main challenges with this photocatalytic process include tuning the band gap of the material to match the visible region of the solar spectrum (between 1.8 and 3.1 eV) and ensuring its stability in aqueous environments. An appealing approach for developing such photocatalysts consists of intercalating functional layers into metal oxides, as is the case for, e.g., perovskite-derived compounds from the Aurivillius ($Bi_2A_{n-1}B_nO_{3n+3}$) family, which have been shown to be photoactive under visible light. Our current work focuses on calculating the surface structures and surface energies of these compounds, determining their potential-pH stability, and applying Monte Carlo sampling to elucidate the microstates of the semiconductor-water interface under an applied voltage. Time permitting, the use of machine-learning techniques to predict the interfacial properties of photocatalytic oxides will also be discussed.

Amyloid peptides in solution and at the gold/water interface.

Luca Bellucci[1],

[1] *CNR-NANO*

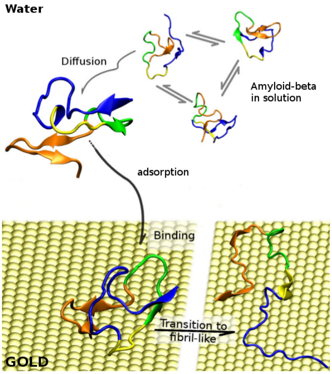


Figure 1: The main phases of the adsorption of Aβ42 onto the gold surface in water. After binding, Aβ42 can form amorphous protein-corona structures or relax towards extended conformations forming a fibril-template.

Recent experimental works tackled the effect, in vitro, of nanoparticles (NPs) on amyloid peptide fibrillations, which are involved human diseases as for example the Alzheimer's desease. The adsorption of such peptides on NPs is a complex process that involves many dynamical steps from the recognition of the peptide by the surface to the conformational rearrangement of the adsorbed peptide [1,2,3]. In our recent study [2], we revealed that a gold surface may induce fibrillation-prone conformations of the amyloid β peptide-42 (Aβ42). This result stemming from a statistically significant enhanced sampling atomistic molecular dynamics simulations over a total of 40.000 ns. Nevertheless, to investigate in details the peptide conformations during the adsorption process, more advanced techniques, as the machine learning (ML), are required. In this perspective, ML can be relevant to discover how inhibit/enhance the fibrillation process by using NPs.

[1] G Brancolini, L Bellucci, MC Maschio, R Di Felice, S Corni,Current Opinion in Colloid & Interface Science, 14 (2019)

[2] L Bellucci, G Bussi, R Di Felice, S Corni Nanoscale 9 (6), 2279-2290

[3] L Bellucci, A Ardèvol, M Parrinello, H Lutz, H Lu, T Weidner, S Corni Nanoscale 8 (16), 8737-8748

New Particle Formation Involving Charged Sulfuric Acid – Ammonia Clusters

Vitus Besel[1], Jakub Kubecka[1], Theo Kurtén[1], Hanna Vehkamäki[1]

[1] *Institute for Atmospheric and Earth System Research*

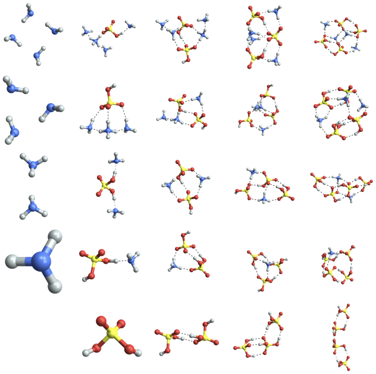


Figure 1: Sampled sulfuric acid - ammonia clusters with the number of sulfuric acid on the x-axis and the number of ammonia on the y-axis.

Atmospheric aerosol particles have impact onto the global climate in various ways, for example in form of cloud condensation nuclei. A bulk of these aerosols is formed by gas-to-particle nucleation (1), which is a still largely unexplained process.

Recently, sulfuric acid has been identified to play a key role in this new particle formation enhanced by bases, such as DMA(2) or ammonia(3). We conducted a computational study for identifying characteristics of cluster formation and nucleation involving sulfuric acid and ammonia in neutral, positive and negative mode.

We used a layered approach of configurational sampling for molecular clusters starting from utilizing a genetic algorithm in order to explore the whole potential energy surface (PES) of the clusters for all geometrical minima, however, with very unreliable energies. The structures are first optimized by a semi-empirical method, further on DFT level with the ωB97X-D functional, and filtered after each step to obtain the global minimum configuration (Cf. attached image).

Further, high level of theory (DLPNO-CCSD(T)) is used to obtain electronic energies and together with DFT frequency analysis the Gibbs free energies of formation are calculated. These are passed to the Atmospheric Cluster Dynamics Code (ACDC)(4) for studying the evolution of cluster populations.

[1] J. Merikanto, D. V. Spracklen, G. W. Mann, S. J. Pickering, and K. S. Carslaw(2009). Atmos.Chem. Phys., 9, 8601-8616

[2] R. Zhang (2010). Science, 328, 1366-1367

[3] T. Anttila, H. Vehkamäki, I. Napari, M. Kulmala (2005). Boreal Env. Res., 10, 523

[4] M.J. McGrath, T. Olenius, I.K. Ortega, V. Loukonen, P. Paasonen, T. Kurten, M. Kulmala (2012). Atmos. Chem. Phys., 12, 2355.

Gaussian Approximation Potentials for radiation damage in bcc metals

Jesper Byggmästar[1], Ali Hamedani[1], Kai Nordlund[1], Flyura Djurabekova[1]

[1] *Department of Physics, University of Helsinki, FIN-00014 Finland*

The wall materials in future fusion reactors must be able to withstand extreme neutron and ion irradiation. High-energy neutrons initiate collision cascades in the material, producing permanent damage in the crystal structure. Collision cascades take place on time scales best captured by classical molecular dynamics simulations, which rely on interatomic potentials to describe the interactions between atoms. Machine learning has during the last decade lead to development of new types of interatomic potentials with superior accuracy compared to traditional analytical potentials. Here, we present early progress on developing Gaussian Approximation potentials (GAP) [1][2] primarily for tungsten, with the future extension to other bcc metals and multi-component alloys. The GAP formalism uses Gaussian process regression to interpolate the potential energy surface based on a training database of reference structures, and has already been applied to a wide range of materials. In order for the GAP to be applicable to radiation damage simulations, particular focus has to be put on reproducing the strong repulsion at short interatomic distances, while simultaneously describing the near-equilibrium properties well. We discuss our fitting strategy to achieve this, which will allow molecular dynamics simulations of collision cascades with unprecedented accuracy.

[1] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons, Phys. Rev. Lett., 104, 136403, 2010

[2] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, Phys. Rev. B, 87, 184115, 2013

Understanding the dynamics of graphene flakes on liquid Cu using molecular dynamics simulations

Juan Santiago Cingolani[1], Mie Andersen[1], Karsten Reuter[1]

[1] *Chair of Theoretical Chemistry, Technical University of Munich, Germany*

It has been shown that high quality single layer graphene can be obtained through chemical vapor deposition on liquid Cu [1]. The role of the liquid surface during growth as well as in defect healing is not yet well understood. For instance, one remarkable aspect of this synthesis that remains to be explained is the self-alignment of growing flakes on the liquid catalyst. Aiming to develop a clearer understanding of the mechanisms behind these processes we carried out molecular dynamics (MD) simulations of graphene flakes on liquid Cu using the third-generation charge optimized many body potential (COMB3).

During our simulations we observed an unexpected “sunken-in” configuration where a graphene flake lies at or below the level of the liquid surface. We then investigated the influence of the flake on the surrounding Cu atoms and calculated a series of properties that are accessible to experimental measurements and can be further used in coarser models. We also carried out ab-initio MD simulations that, while more limited in scope, can be used to validate our results. All in all, with this work we contribute to a more detailed atomic-scale picture of the interactions of graphene flakes on liquid Cu.

[1] L. Tan, M. Zeng, T. Zhang, L. Fu, Design of catalytic substrates for uniform graphene films: from solid-metal to liquid-metal, Nanoscale, 7, 9105-9121, 2015

Clustering of transition metal dichalcogenides

Anais Colibaba[1], Stefano Sanvito[1]

[1] *School of Physics and CRANN Institute, Trinity College, Dublin 2, Ireland*

"Band-nesting" occurs in all two-dimensional transition metal dichalcogenides (2D-TMDs) and it corresponds to singularities in the joint density of states (JDOS) [1]. The JDOS is linearly dependent on the Brillouin zone (BZ) integrated inverse of $|\nabla(E_c - E_v)|$ where E_c and E_v are all conduction and valence band energies. Therefore the regions in reciprocal space where $|\nabla(E_c - E_v)| = 0$ can potentially give rise to high absorption peaks. In order to systematically screen 2D TMDs for band nesting, we created a database of 289 2D-TMD stacks that have similar lattices. Hence, the irreducible BZ of the stack is approximated to be equivalent to the ones of the TMD building blocks. We investigated $|\nabla(E_{c1} - E_{v1})|$ for these 2D-TMD stacks where E_{c1} and E_{v1} were chosen as the first conduction band of one of the TMDs and as the last valence band of the other TMD. We used dimensionality reduction techniques such as t-distributed stochastic neighbor embedding (t-SNE) and clustering algorithms such as k-means to classify all the TMD stacks based on their respective $|\nabla(E_{c1} - E_{v1})|$ gradient maps on the BZ. Six distinct clusters were identified and all of them show different regions in the irreducible BZ with very low values of $|\nabla(E_{c1} - E_{v1})|$. This analysis opens a coarse yet efficient path towards exploring large sets of 2D TMD stacks for high optical peaks.

[1] A. Carvalho, R. M. Ribeiro, A. H. Castro Neto, "Band nesting and the optical response of two-dimensional semiconducting transition metal dichalcogenides", Phys. Rev. B 88, 115205 (2013)

Systematically identifying the collective variables describing phase transformations in flexible materials

Karen D. Dedecker[1], Ruben Demuynck[1], Sven M.J. Rogge[1], Louis Vanduyfhuys[1], Michel Waroquier[1], Veronique Van Speybroeck[1]

[1] *Center for Molecular Modeling, Ghent University, Belgium*

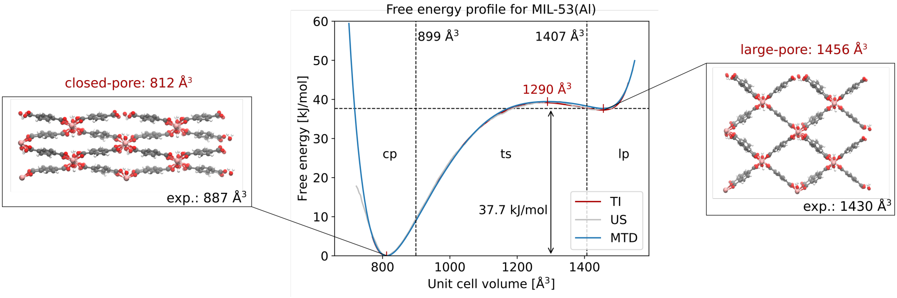


Figure 1: Free energy profile for MIL-53(Al) displaying phase transitions between a closed-pore and a large-pore phase.

Flexible metal-organic frameworks (MOFs) are a promising class of materials that possess a highly ordered network but yet show the ability to structurally transform between different crystalline phases [1]. In order to understand the molecular origin of this behavior, knowledge of the free energy surface (FES) of the material as a function of the relevant molecular degrees of freedom or collective variables (CVs) is essential. However, as yet it remains elusive how CVs can be selected for complicated transformations, and solving this problem requires a systematic protocol for the selection of CVs. We use the well-studied breathing MOF MIL-53(Al) as a benchmark system [2], for which the unit cell volume is a good CV [3]. After selecting a set of CVs, enhanced sampling methods are performed using in-house developed force fields, and the free energy profile as a function of the proposed CVs is computed. The results are then compared for various sets of CVs. One way to select these variables is by employing a time-lagged independent component analysis [4]. This linear dimensionality reduction method identifies the “slow” subspace of a large set of variables, and greatly ameliorates the obtained results compared to the results obtained with variables that were selected based on physical intuition [5]. Also non-linear dimensionality reduction methods are promising techniques to allow for the selection of CVs for materials of increasing complexity.

[1] S. Kitagawa et al., Angew. Chem. Int. Ed., 43: 2334, 2004.

[2] Y. Liu et al., J. Am. Chem. Soc., 130:11813, 2008.

[3] R. Demuynck et al., J. Chem. Theory Comput., 13:5861, 2017.

[4] F. Nüske et al., J. Chem. Theory Comput., 10:1739, 2014.

[5] R. Demuynck et al., J Chem. Theory Comput., 14:5511, 2018.

Development of Descriptors for the Prediction of Adsorption Energies on Transition Metal Catalysts and their Alloys

Martin Deimel[1], Karsten Reuter[1], Mie Andersen[1]

[1] *Theoretical Chemistry, Technical University of Munich, Germany*

The construction of microkinetic models for catalytic processes requires the knowledge of both adsorption energies of the different species and reaction barriers. In the pursuit of discovering new catalysts with improved activity and selectivity, computational screening renders the explicit calculation from first principles intractable. In order to reduce the computational effort, it is desirable to identify descriptors that allow for a cheap and accurate prediction of these quantities. In the present work we apply the recently developed compressed-sensing method Sure Independence Screening and Sparsifying Operator (SISSO) [1] to identify cheaper and more accurate descriptors for adsorption energies on transition metals (TMs) and their binary alloys [2]. These descriptors are constructed as non-linear functions of properties of the clean catalyst surface (primary features). The different metal atom composition of sites on surface alloys introduces an additional tunable parameter. In combination with distinct site preferences and coordination patterns of different adsorbates, this provides the opportunity to break the interdependence of adsorption energies imposed by scaling relations. Such interdependence is often viewed as one of the major limiting factors for identifying even better catalysts than we know today.

[1] Runhai Ouyang, Stefano Curtarolo, Emre Ahmetcik, Matthias Scheffler, Luca M. Ghiringhelli, SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates, Physical Review Materials, 2, 083802, 2018

[2] Mie Andersen, Sergey V. Levchenko, Matthias Scheffler, Karsten Reuter, Beyond Scaling Relations for the Description of Catalytic Materials, ACS Catalysis, 9, 2752-2759, 2019

Proteins structural transition states explored with minimalist coarse grained models

Francesco Delfino[1], Valentina Tozzini[2]

[1] *I.M. Sechenov First Moscow State Medical University, Russia*
[2] *National Research Center (CNR) - Institute of nanoscience, Italy*

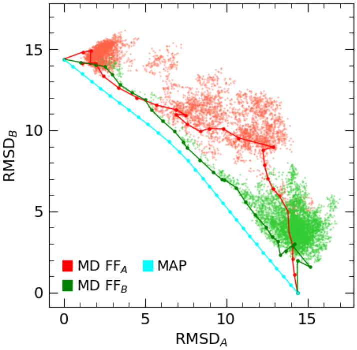


Figure 1: RMSD_A vs RMSD_B plot of Calmodulin transition trajectories between the apo-state (A) and the Ca-bound state (B), evaluated by molecular dynamics (MD) simulations at room temperature with a coarse grained force field biased towards structure A (FF_A, red), or towards structure B (FF_B, green); compared with MinActionPath [3] trajectory (MAP, cyan). The full conformational space explored during MD simulations is represented by small red (FF_A) and green (FF_B) dots.

Transitions between different conformational states are ubiquitous in proteins. However, modeling such processes is extremely difficult, due to the need to efficiently sample a vast conformational space to look for the actual transition path for highly complex systems [1]. One strategy that simplifies this task is to first apply a minimalist coarse-grained model [2] to the protein, to facilitate the exploration of the transition paths performed through molecular dynamics simulations. We then select representative structures along the trajectory based on a structural clustering algorithm and build a cleaned-up trajectory with them. Our analysis indicates that this method returns trajectories capable of exploring intermediate states with physical meaning, but the selection of representative structures remains a sensitive issue. Our method could greatly benefit from the use of more advanced machine learning techniques, which might help in strengthening the selection process.

[1] T. Sikosek, H. Krobath, H. S. Chan. Theoretical Insights into the Biophysics of Protein Bi-stability and Evolutionary Switches. PLoS Comput Biol 12(6): e1004960. (2016).

[2] V. Tozzini. Coarse-grained models for proteins. Current opinion in structural biology 15 (2), 144-150 (2005).

[3] J. Franklin, P. Koehl, S. Doniach, M. Delarue. MinActionPath: maximum likelihood trajectory for large-scale structural transitions in a coarse-grained locally harmonic energy landscape. Nucleic Acids Res. 35:W477-W482 (2007).

Moving and Vibrating: Study of the Structural and Magnetic, Electronic and Vibrational Properties of Ni_{19}

Laia Delgado Callicó[1], Kevin Rossi[2], Francesca Baletto[1]

[1] Department of Physics, King's College London, Strand, London WC2R 2LS, United Kingdom
[2] Department of Engineering, Ecole Polytechnique Federale de Lausanne, Switzerland

Based on first-principles all-electron calculations, the structural stability and magnetic variety of Ni_{19} nanoparticles has been studied. Exploring the interplay between atomic vibrations and magnetic properties represents an initial and fundamental step in the rational design of nanostructures for application in catalysis and optical devices. Accurate all-electron density functional calculations and adiabatic ab initio molecular dynamics simulations at several temperatures have been performed on Ni_{19} considering perfect and distorted structures. A non-trivial relationship between the relative stability of the nanostructures and the total magnetization for each local minimum was found, as well as a structure-dependent influence of the total magnetization on the calculated vibrational spectra. No structural or phase transitions were observed during the dynamics simulation at a range of temperatures from 0K to 900K.

Two-dimensional antiferromagnetic spintronics at perovskite oxide interfaces

Jayita Chakraborty[1], Nirmal Ganguli[1],

[1] *Department of Physics, Indian Institute of Science Education and Research Bhopal, Bhauri, Bhopal 462066, India*

Antiferromagnetic spintronics is a promising area of research for future technology because of its advantages over conventional ferromagnetic spintronics. Besides being robust against perturbation due to magnetic fields, antiferromagnetic materials produce no stray field and display ultrafast dynamics. In the absence of inversion symmetry in the structure, a strong spin-orbit interaction may lead to Dresselhaus or Rashba effect, facilitating the transfer of angular momentum between the orbital angular momentum of carriers and the spin angular momentum of the localized electrons through spin-orbit torques. Our calculations for one unit cell thick SrIrO3 on SrTiO3 substrate reveal an antiferromagnetic insulating ground state with an isolated Ir-5d band lying right above the Fermi level. Combining this nonpolar heterostructure with LaAlO3, a polar perovskite oxide along the 001 direction, we achieve charge transfer at the interface that partially fills the isolated Ir-5d band. This partially filled band exhibits Rashba effect in a two-dimensional antiferromagnetic electron gas confined to the plane of the interface. Our systematic calculations suggest that spin-orbit interaction helps to stabilize antiferromagnetic interaction among the Ir atoms. In view of the previous discussion, the prediction of substantial Rashba effect in a two-dimensional antiferromagnetic electron gas is promising for further developments of antiferromagnetic spintronics technologies.

BCN Models: A User Friendly Tool to Generate of Stoichiometric Wulff Shaped Nanoparticles

Danilo González[1], Bruno Camino[1], Albert Rimola[1], Xavier Solans-Montfort[1], Luis Rodríguez-Santiago[1], Mariona Sodupe[1]

[1] *Departament de Química, Universitat Autònoma de Barcelona, Spain*

The most recent advances on several scientific fields like electronics, medicine, and catalysis are closely related to the development of novel materials, especially at nanomaterials like nanoparticles. The physical-chemical properties of nanoparticles are controlled and tailored by changing their size and shape extending their applicability. Combination of experimental and computational techniques have demonstrated to be a successful strategy for the understanding of the relation between structural and chemical properties of nanoparticles. Nevertheless, computational modelling of nanoparticles is not trivial because of their variety of sizes and shapes and the difficult to construct rational models. With the aim of generating atomistic nanoparticles models, we present a user-friendly web server Bulk Cut Nanoparticles Models (BCN Models) that is capable to generate and characterize stoichiometric Wulff-like nanoparticles in a systematic way for a large variety of ionic binary compounds.

Surface termination and composition control of activity of the $\mathrm{Co}_x\ \mathrm{Ni}_{1-x}\ \mathrm{Fe}_2\ \mathrm{O}_4$ (001) surface for water oxidation: insights from DFT+U calculations

Hamidreza Hajiyani[1], Rossitza Pentcheva[1]

[1] *IDepartment of Physics and Center for Nanointegration (CENIDE), University of Duisburg-Essen, Duisburg, Germany*

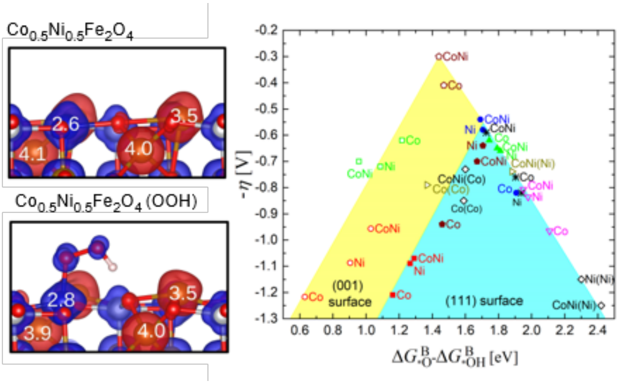


Figure 1: Overpotential versus the binding energy difference of intermediates for different reaction sites and terminations.

We used density functional theory calculations with a Hubbard U term (DFT+U) calculations to explore the origin of improvement on the catalytic activity of $\mathrm{CoFe}_2\ \mathrm{O}_4$ by partial substitution of Co by Ni. Different reaction sites (Fe, Co, and Ni) were investigated at three terminations: the B-layer and with an additional half and full monolayer of Fe. Our results indicate that for the latter termination the OER activity is significantly higher than the previously investigated (111) surface and the B-layer. Surface Co cations are identified as the active sites and the ones at the A-layer termination exhibit the lowest theoretically reported overpotential of 0.30 V. We find that the scaling relationship of the binding energies of $\Delta G^{B}_{(*OOH)}$ versus $\Delta G^{B}_{(*OH)}$ is overall fulfilled, however a deviation of ±0.40 eV from the linear fit occurs for the most active terminations that explains the reduction of overpotential which is placed above the activity volcano [1].

.

[1] [1] H. Hajiyani, R. Pentcheva , ACS Catal., 8, 11773-11782 (2018)

Noncollinear spin density of an adatom on a magnetic surface

Soumyajyoti Haldar[1], Stefan Heinze[1]

[1] *Institute of Theoretical Physics and Astrophysics, University of Kiel, Leibnizstr. 15, 24098 Kiel, Germany*

Today, noncollinear spin structures at surfaces and interfaces receive great attention due to potential applications in spintronic devices. In such magnetic structures, the spin direction changes from atom to atom. Besides this interatomic noncollinear magnetism, there is also intraatomic noncollinear magnetism in which the spin direction varies for different orbitals of an atom [1]. It can occur due to spinorbit coupling or due to a noncollinear spin structure.

Here, we demonstrate that intraatomic noncollinear magnetism can occur for adatoms on a magnetic surface with a noncollinear spin structure [2]. We study Co and Ir adatoms on Mn/W(110) using DFT. We find that the canted spin structure of the Mn surface layer is encoded into different orbitals of the adatoms. Our conclusions apply in general to adatoms on surfaces with a noncollinear magnetic structure e.g. spin spirals, skyrmions or domain walls and explain recent experimental results of spin-polarized STM experiments [3].

[1] L. Nordström et al., Phys. Rev. Lett.76, 4420 (1996)

[2] S. Haldar and S. Heinze, Phys. Rev. B98, 220401 (R) (2018).

[3] D. Serrate et al., Phys. Rev. B93, 125424 (2016)

Extending the applications of machine learning interatomic potentials: radiation damage analysis in silicon

Ali Hamedani[1], Jesper Byggmästar[1], Kai Nordlund[1], Ghasem Alahyarizadeh[2], Reza Ghader[2], Flyura Djurabekova[1], Gabriele C. Sosso[3]

[1] *Department of Physics, University of Helsinki, P.O. Box 43, FI-00014, Helsinki, Finland*
[1] *Engineering Department, Shahid Beheshti University, G.C, P.O. Box 1983969411, Tehran, Iran*
[1] *Department of Chemistry and Centre for Scientific Computing, University of Warwick, Coventry, UK*

Machine learning interatomic potentials build a bridge between the accuracy of first principles calculations and fastness of the empirical interatomic potentials; making it possible to simulate thousands of atoms with DFT accuracy. The effects of radiation on the properties of silicon are of significant interest in technological and theoretical contexts. Gaussian Approximation Potential (GAP) was used to simulate radiation damage in Si for the very first time. This potential accurately reproduces density functional theory reference results for a wide range of observable properties, including crystal, liquid, and amorphous bulk phases, as well as point, line, and plane defects. Displacement threshold energy (Ed) as the fundamental parameter in radiation damage analysis was calculated in [111], [100] and [110] directions by the original potential. Moreover, MD-DFT calculation of Ed was performed to validate acquired values. Excellent agreement was obtained in [111] direction and the consistency in [100] is superior to Tersoff and Stillinger-Weber forcefields which have been used for decades in cascade simulations. The repulsive response of the potential in the high energy collision regime was explored comprehensively. Dmol, an all-electron ab initio repulsive pair potential, was added to the original potential and the behavior of the modified potential was tuned via controlling reproduced total energy of various structures containing shorter interatomic distances against DFT.

DScribe: A python library of descriptors for machine learning in materials science

Lauri Himanen[1], Eiaki Morooka[1], Marc Jäger[1], Filippo Federici Canova[1, 4], David Gao[4], Patrick Rinke[1], Adam S. Foster[1, 2, 3]

[1] *Department of Applied Physics, Aalto University, Finland*
[2] *Graduate School Materials Science in Mainz, Germany*
[3] *WPI Nano Life Science Institute, Kanazawa University, Japan*
[4] *Nanolayers Ltd., UK*

Figure 1: DScribe: A python library of descriptors for machine learning in materials science. See homepage at https://singroup.github.io/dscribe/

Applying machine learning models for the prediction of physical properties of atomistic systems is a highly active area of research. A problem related to this activity is the conversion of atomistic structural and chemical information into well-performing input features that can be used in such machine learning models. A common approach is to introduce a feature transformation that transforms the original atomic geometry into a set of input features, here called a descriptor, that can more easily be interpreted by a machine learning algorithm. Various such descriptors have been proposed, and often multiple approaches have to be tested to find a suitable one for a specific task [1].

We introduce the python package DScribe, that provides implementations for multiple feature transformations, including Smooth Overlap Of Atomic Positions (SOAP) [2], Atom-centered Symmetry Functions (ACSF) [3], Many-body Tensor Representation (MBTR)[4], Coulomb Matrix [5], Sine Matrix [6] and Ewald Matrix [6]. Our purpose is to accelerate the application of machine learning in this field by providing a user-friendly, off-the-shelf implementation that can be easily validated and extended. This library has a coherent python-interface with C/C++ extensions for the computationally intensive tasks. We use a set of regression tests and continuous integration to ensure the validity of the implementation, and provide the source code together with tutorials and documentation.

[1] Faber, Felix A. et al., Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error, Journal of Chemical Theory and Computation, 13, 5255–5264, 2017

[2] Albert P. Bartók, Risi Kondor, and Gábor Csányi, On representing chemical environments, Phys. Rev. B, 18, 184115, 2013

[3] Jörg Behler, Atom-centered symmetry functions for constructing high-dimensional neural network potentials, The Journal of Chemical Physics, 134, 074106, 2011

[4] Haoyan Huo, Matthias Rupp, Unified Representation of Molecules and Crystals for Machine Learning, arXiv:1704.06439, 2017

[5] Rupp, Matthias and Tkatchenko, Alexandre and Müller, Klaus-Robert and von Lilienfeld, O. Anatole, Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, Phys. Rev. Lett., 108, 058301, 2012

[6] Faber, Felix and Lindmaa, Alexander and Lilienfeld, O. Anatole von and Armiento, Rickard, Crystal structure representations for machine learning models of formation energies, International Journal of Quantum Chemistry, 115, 1094–1101, 2015

Machine learning hydrogen adsorption on nanoclusters through structural descriptors

Marc Jäger[1], Eiaki Morooka[1], Filippo Federici Canova[1, 2], David Gao[2], Lauri Himanen[1], Adam Foster[1, 3, 4]

[1] *Department of Applied Physics, Aalto University, Finland*
[2] *Nanolayers Ltd., UK*
[3] *Graduate School Materials Science in Mainz, Germany*
[4] *WPI Nano Life Science Institute, Kanazawa University, Japan*

Catalytic activity of the hydrogen evolution reaction on nanoclusters depends on diverse adsorption site structures. Machine learning reduces the cost for modelling those sites with the aid of descriptors. We analysed the performance of state-of-the-art structural descriptors Smooth Overlap of Atomic Positions, Many-Body Tensor Representation and Atom-Centered Symmetry Functions while predicting the hydrogen adsorption (free) energy on the surface of nanoclusters [1]. The 2D-material molybdenum disulphide and the alloy copper–gold functioned as test systems. Potential energy scans of hydrogen on the cluster surfaces were conducted to compare the accuracy of the descriptors in kernel ridge regression. By having recourse to data sets of 91 molybdenum disulphide clusters and 24 copper–gold clusters, we found that the mean absolute error could be reduced by machine learning on different clusters simultaneously rather than separately. We concluded that fitting of potential energy surfaces could be reduced significantly by merging data from different nanoclusters. We also develop an high-throughput workflow to systematically produce large nano cluster datasets and allow for rapid ML analysis.

[1] M. O. J. Jäger, E. V. Morooka, F. F. Canova, L. Himanen, and A. S. Foster, Npj Comput Mater 4, 37 (2018).

Ab inito multiple time stepping using machine learning

Venkat Kapil[1], Joost VandeVondele[2], Michele Ceriotti[1]

[1] Institute of Materials, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[2] Swiss National Supercomputing Centre, Zurich

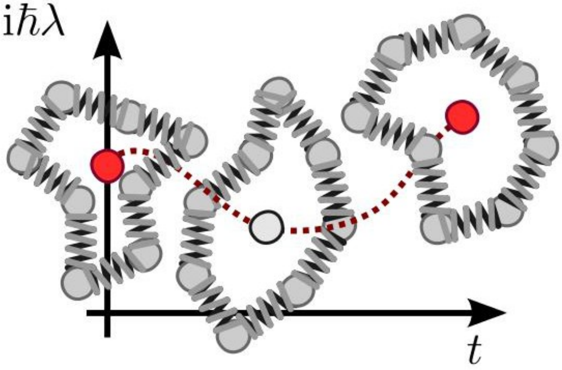


Figure 1: Multiple time stepping in real and imaginary time

The development and implementation of increasingly accurate methods for electronic structure calculations mean that, the statistical sampling procedure used to compute thermodynamic properties has become the bottleneck to reach the ultimate level of accuracy. The computational cost of performing molecular dynamics remains large, due to the need of evaluating the energetics of several tens of thousands of configurations. Furthermore, to account for the quantum nature of light nuclei, the overhead is typically a factor of 10 to 100 times larger, making it prohibitive when combined with advanced electronic structure methods. Here, we present how machine learning potentials can be used to extend multiple time step integrators [1] to the of case of ab initio molecular dynamics. We show that it is possible to dramatically reduce the cost of statistical sampling while describing interparticle interactions at high levels of electronic structure theory [2,3] . Furthermore, we demonstrate that it is possible to drive the overhead of modelling nuclear quantum effects virtually to zero [2], by performing multiple time stepping in imaginary time [4]. The basic idea is demonstrated for a combination of MP2 and semi-local density functional theory using our implementation in i-PI, the universal force engine.

[1] Tuckerman et. al Journal of Chemical Physics, 97, 1990 (1992)

[2] Kapil et. al Journal of Chemical Physics, 144, 054111 (2016)

[3] Kapil et. al Journal of Physical Chemistry C, 122 (22), pp 6048–6054 (2018)

[4] Markland, Manolopolous Journal of Chemical Physics, 129, 024105 (2008)

[5] Kapil et. al Computer Physics Communications 236, 214-223

Estimating mechanical properties of Al-based alloys via High-throughput approach

Konstantin Karavaev[1], Ekaterina Smirnova[1], Igor Abrikosov[2]

[1] *Materials Modeling and Development Laboratory, NUST "MISIS", 119049 Moscow, Russia*
[2] *Department of Physics, Chemistry and Biology (IFM) Linköping University, SE-581 83, Linköping, Sweden*

Considerable amounts of data on materials properties could be effectively generated using modern High-throughput methods. Available "Big Data" in combination with novel machine learning methods allows one to obtain trends and dependencies between external parameters and properties, which opens fundamentally new possibilities for research. Design of new aluminum alloys is one of the well-studied and technologically important problems. In this work, binary aluminum alloys with 1 and 2 % impurity concentration were selected for our high-throughput analysis. Comprehensive automatic work-flow was developed for calculation of physical properties. Ab initio simulations of aluminum alloys was performed using Exact Muffin-Tin Orbitals method (EMTO) [1] with Coherent potential approximation used to treat disorder of alloys. We had studied binary aluminum alloys with 44 alloying components. For each alloy, we determined its equation of state, equilibrium lattice parameters and bulk moduli. Besides, we calculated the single-crystal elastic constants, the Young's moduli and the shear moduli. We compared these results with the available experimental data. Analysis of results shows that variation of studied parameters depends on electronic concentration of the alloy. Obtained data from our study of Al alloys shows that this approach gives valid information about mechanical properties of Al alloys. Our next step is to perform similar calculation for multicomponent aluminum-based system.

[1] L. Vitos, Computational Quantum Mechanics for Materials Engineers: The EMTO Method and Applications, Springer-Verlag, London, 2007.

Assessment of similarity patterns by chemometric tools in clinical data of patients diagnosed with diabetes mellitus type 2

Ilia Kichev[1], Miroslava Nedyalkova[1], Dimitar Dimitrov[2], Vasil Simeonov[2]

[1] Chair of General and Inorganic Chemistry, Faculty of Chemistry and Pharmacy, University of Sofia "St. Kl.Okhridski", J. Bourcher Blvd. 1, 1164 Sofia, Bulgaria
[2] Chair of Analytical Chemistry, Faculty of Chemistry and Pharmacy, University of Sofia "St. Kl. Okhridski", J. Bourcher Blvd. 1, 1164 Sofia, Bulgaria

The present study deals with the assessment and interpretation of clinical data for patients diagnosed with diabetes mellitus type 2 by the use of intelligent data analysis. One hundred patients were involved, each of them characterized by 33 clinical indicators related to testing blood sugar, renal and liver functions, and anthropometric data. The assessment method was non-hierarchial cluster analysis, being very convenient for supervised pattern recognition tasks. The major goal of the study was to try to find patterns of similarity (clusters) in the data set in order to improve the overall information, needed by the medics for specific treatment and checking the health status. Five clusters were interpreted with respect to the similarities between the clinical indicators. The same number of clusters was a priori chosen for interpretation of similarity between the patients. Each cluster is specifically explained and discussed. In further research other methods, such ass fuzzy clustering, are going to be applied.

Towards active learning using reliable uncertainty estimates

M. Larmuseau[1, 2, 3], M. Sluydts[1, 2], T. Dhaene[3], S. Cottenier[1, 2]

[1] *Center for Molecular Modeling, Ghent University, Belgium*
[2] *Department of Electrical Energy, Metals, Mechanical Constructions Systems, Ghent University, Belgium*
[3] *SUMOLab, Ghent University, Belgium*

Good machine learning models require a large set of labeled training data. Deciding which labeled data is most essential can be done on-the-fly using a technique called active learning. Using active learning both the predictions as the uncertainty on the predictions are combined to direct the data collection algorithms towards a desired objective, implicitly defining a good balance between exploration and exploitation. In this way, it is possible to optimize the properties one is interested in, while improving the model performance. In previous work, we focused on using Gaussian Processes in an active learning framework to predict ab initio formation energies of geometrically similar structures, which lead to great speedups in discovering new quaternary materials. Deep graph learning offers new potential, by allowing us to incorporate the geometric structure of the atomic systems into the model. Unlike Gaussian processes, however, deep learning methods do not have implicit access to uncertainty quantification. The following question thus arises: ”How can reliable uncertainty estimates be included in a deep graph learning model with out significantly affecting either performance or training time?”. We explicitly demand that the training time does not increase too much as, in active learning, the model has to be completely retrained on a regular basis. This implies that a purely Bayesian approach, which would be the preferred approach to deal with this issue, is not a real option as it quickly becomes too slow for large datasets. Uncertainty estimation in deep neural networks is a hot topic in current machine learning literature, but so far no clear answer has emerged. We discuss different state-of-the-art approaches, which rely either on explicitly modeling the uncertainty on the predictions or ensemble methods and compare how well they perform in discovering new materials.

Theoretical study on the structures and thermodynamics of A-type rare earth sesquioxides by HSE and PBE+U methods

Shikun Li[1], Yong Li[1], Marcus Bäumer[1], Lyudmila Moskaleva[1]

[1] *Institute of Applied and Physical Chemistry, Universität Bremen, Germany*

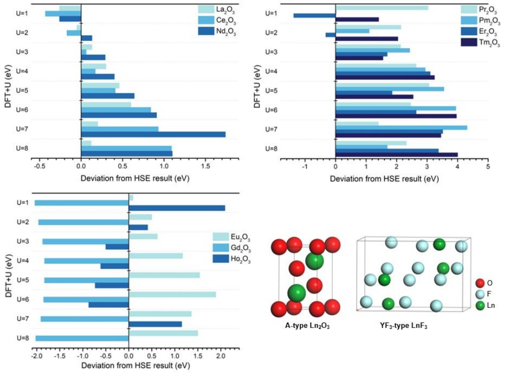


Figure 1: The deviation of PBE+U from HSE06 results for the reference reaction energies and the corresponding geometries

Rare earth oxides (REOs) have shown promising performance in the oxidative coupling of methane (OCM), which has a tremendous economic potential because of the possibility of methane conversion to C_2 hydrocarbons. In view of the difficult handling of f-electrons in rare earth elements by conventional DFT, PBE+U approach has been chosen and systematically compared to the results achieved with a hybrid functional HSE[1,2]. In doing that, a range of U values of PBE+U have been employed to find the best fit to HSE results. The imaginary reaction $2LnF_3 + 3H_2O \rightarrow Ln_2O_3 + 6HF$, has been taken as reference. The best agreement with the HSE results was found for the smaller U values (U=1-3). On the basis of the obtained U values, we will be able to use a significantly less expensive than the hybrid functional but relatively accurate method PBE+U for the complicated systems and problems, such as the adsorption and reaction mechanisms on REOs surfaces and nanoparticles etc.

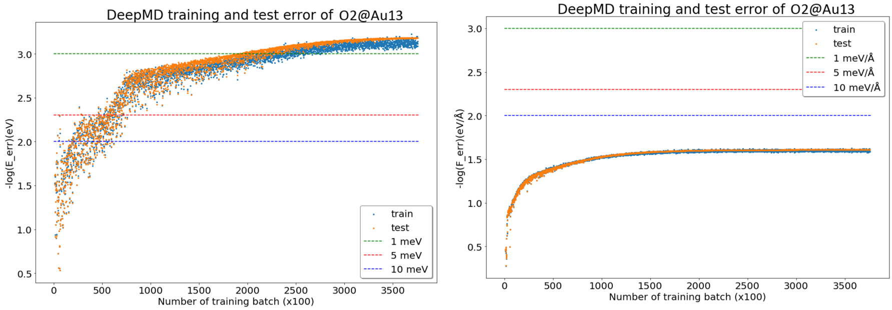
[1] Da Silva, J. L., Ganduglia-Pirovano, M. V., Sauer, J., Bayer, V., & Kresse, G. Hybrid functionals applied to rare-earth oxides: The example of ceria. Physical Review B, 75(4), 045121, 2007

[2] Brugnoli, L., Ferrari, A. M., Civalleri, B., Pedone, A., & Menziani, M. C. Assessment of Density Functional Approximations for Highly Correlated Oxides: The Case of CeO2 and Ce2O3. Journal of Chemical Theory and Computation, 14(9), 4914-4927, 2018

A Practice in Automation and Machine Learning about Chemistry

Yunpei Liu[1], Jianxing Huang[1], Jun Cheng[1]

[1] *Department Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Fujian, China*



(a) **(b)**

Figure 1: The error of the NNP model trained from O2@Au3 clusters.

The material genome project hopes to combine theoretical calculation and experimental synthesis with data analasis, which links high-throughput experiment, calculation and database together. With the participation of automate, high-throughput calculation could be easy. On the other hand, machine learning technology could greatly benefit the prediction of material properties, based on the large amount of data collected. So we tried the construction of workflow for calculation with the help of AiiDA, as well as the simulation of PES with the help of DeePMD-kit.

[1] Pizzi, G.; Cepellotti, A.; Sabatini, R.; Marzari, N.; Kozinsky, B., AiiDA: automated interactive infrastructure and database for computational science, Computational Materials Science, 2016, 111, 218-230.

[2] Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W., Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics, Physical Review Letters, 2018, 120 (14), 143001.

[3] Wang, H.; Zhang, L.; Han, J.; E, W., DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics, Computer Physics Communications, 2018, 228, 178-184.

Importance of crystal representation in machine learning of thermodynamic and magnetic properties

Teng Long[1], Qiang Gao[1], Ingo Opahle[1], Hongbin Zhang[1]

[1] *Institute of Materials Science, TU Darstadt, 64287 Darmstadt, Germany*

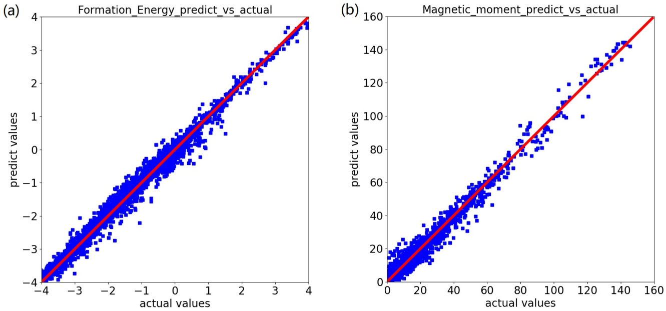


Figure 1: (a) Comparison of predicted and actual formation energy from Materials Project; (b) comparison of predicted and actual magnetic moment from high throughput calculation.

As an emergent field, materials informatics based on statistical machine learning (ML) techniques has attracted intensive attention.[1] It is realized that not only the chemical information but also more importantly a proper representation of crystal structures are required to perform accurate ML modeling.

In this work, we aspire to apply Smooth Overlap of Atomic Positions (SOAP) as a crystal representation on ML modeling of the thermodynamic and magnetic properties.[2] To consider the binary and ternary compounds at equal foot, we modified the SOAP descriptor to generate atomic specific descriptors, which are fed into the recurrent neural network.[3] We have implemented and tested the algorithm on the formation energies of 40, 000 compounds from Materials Project[4] and the magnetization of 15, 000 compounds from our high throughput DFT calculations. The accuracy (R2) of 96.8% and 91% has been achieved respectively as shown in Figure 1.

[1] Rickman, J. M., T. Lookman, and S. V. Kalinin. ”Materials Informatics: From the Atomic-Level to the Continuum.” Acta Materialia (2019).

[2] De, Sandip, et al. ”Comparing molecules and solids across structural and alchemical space.” Physical Chemistry Chemical Physics 18.20 (2016): 13754-13769.

[3] Géron, Aurélien. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. ” OŔeilly Media, Inc.”, 2017.

[4] Jain, Anubhav, et al. ”The Materials Project: Accelerating Materials Design Through Theory-Driven Data and Tools.” Handbook of Materials Modeling: Methods: Theory and Modeling (2018): 1-34.

Adsorption, Decomposition and Surface Dynamics of Formic Acid on Cu531

Sabine Matysik[1], Stephen Jenkins[1]

[1] *Department of Chemistry, University of Cambridge, United Kingdom*

The decomposition of formic acid has been studied on numerous low-index metal and metal oxide surfaces, both for its industrial relevance and as a model system for the catalytic properties of various surfaces. In this dispersion-corrected density functional theory study, we focus on the adsorption of HCOOH and formate as its first decomposition intermediate on the chiral Cu531 surface to study the interplay between chirality and surface dynamics. Using the transition state of this dissociation reaction as a starting point for molecular dynamics simulations enables us to analyse only reactive trajectories that include the adsorption process.

Role of Cation Doping in NMV (Na(3+x)M(x)V(2-x)(PO4)3 Type NASICON Systems to Regulate Phase Stability and Electrochemical Behaviour as Sodium Ion Battery Cathodes - A First Principles Study

Madhulika Mazumder[1], Swapan K Pati[2]

[1] *New Chemistry Unit, SAMAt (School of Advanced Materials), JNCASR, Bangalore 560064, India.*
[2] *Theoretical Sciences Unit, SAMAt (School of Advanced Materials), JNCASR, Bangalore 560064, India*

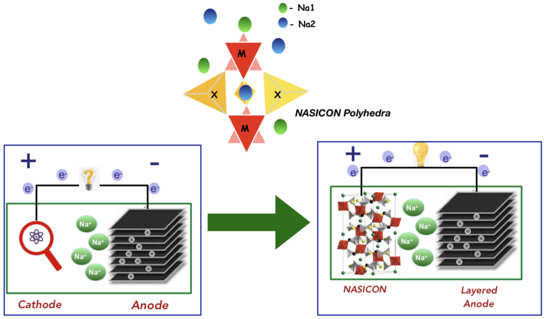


Figure 1: Will the Quest for the Best Cathode ever Stop?!

With the emergent competition between growing energy demand and development of proficient EES (Electrical Energy Storage) Systems, Sodium-Ion Batteries(SIB) have drawn great attention on grounds of abundance and low expense. However, the architecture for large-scale application and commercialization is yet to overcome pitfalls of poor specific capacity, output voltage etc. In recent years, polyanionic NASICON(Na-Super-Ionic-Conductors) frameworks are being extensively explored for their long-term cyclability and high Na+ mobility. This work focuses on $Na_{(3+x)}M_xV_{(2-x)}(PO_4)_3$ (M=Mn,Fe,Cr,Ti) NASICON systems as candidates for SIB cathode and correlates their electronic microstructure with observed electrochemical performance. First-Principles Density-Functional-Theory has been employed to predict the ground state stability and activity of aliovalently doped NMV species. The difference in the coordination environment of V/M is reflected in the preferential geometry as well as electrochemistry of the configurations. An idea about Na^+ ionic migration and rates of diffusion in each species is also acquired through Ab-Initio Molecular Dynamics Studies.

[1] Weidong Zhou, Leigang Xue, Xujie Lü, Hongcai Gao, Yutao Li, Gengtao Fu, Zhiming Cui, Ye Zhu and John B. Goodenough, "Na_xMV(PO4)3 (M = Mn, Fe, Ni) Structure and Properties for Sodium Extraction" Nano Lett. 2016, 16, 7836-7841.

[2] Kieu My Bus, Van An Dinh, Susumu Okada, Takahisa Ohno, " Hybrid functional study of the NASICON-type Na3V2(PO4)3: crystal and electronic structures, and polaron–Na vacancy complex diffusion" Phys. Chem. Chem. Phys., 2015, 17, 30433.

[3] Fan Chen, Vadim M. Kovrugin, Rénald David, Olivier Mentré, François Fauth, Jean-Noël Chotard and Christian Masquelier, "A NASICONType Positive Electrode for Na Batteries with High Energy Density: Na4MnV(PO4)3 "Small Methods 2018, 1800218

[4] Shuangqiang Chen, Chao Wu, Changbao Zhu, Yuanye Huang, Kai Xi, Joachim Maier, Yan Yu, "Challenges and Perspectives for NASICONType Electrode Materials for Advanced SodiumIon Batteries "Adv. Mater, 2017, 29, 1700431.

ML assisted phase contrast of STEM imaged HfO_2 memory filaments

Eamon McDermott[1],

[1] *Independent Researcher*

HfO_2 remains a promising active material for the large-scale manufacture of resistive memory devices. However, the amorphous structure of HfO_2 in prototype devices makes their characterization difficult, with downstream implications in device design and quality control. It would be desirable to use STEM to characterize the atomic-scale structure of a resistive memory device as it is cycled in situ, but a lack of chemical contrast in the 2D core-level EELS map provided by STEM currently prevents this.

I am implementing a ML-assisted method (as opposed to statistical methods such as PCA) to provide this chemical contrast using ab-initio simulated EELS produced by the WIEN2k code. The focus is on an extensible method that can be applied to EELS spectra of other materials at scale through the use of simulation databases (i.e. materialsproject.org). The long-term goal is an automated material classification model for users of EELS and XPS spectroscopies.

Atomic ordering and transport properties in Nb$_x$CoSb half-Heuslers

Joaquin Miranda[1],

[1] *Biomaterials, University of Bayreuth, Germany*

Defective half-Heuslers (HHs) has been recently proposed as a new class of materials for thermoelectric applications. In Nb$_x$CoSb the amount of Nb plays a pivotal role, not only for tuning the electronic properties, but also to create an effective way to scatter phonons. In this work, we study by a combination of first principles, Monte Carlo simulations, and Molecular Dynamics two fundamental aspects of the Nb$_x$CoSb compounds. First. In order to understand the relation between structure and electronic properties, we set a cluster expansion based on DFT energies. The structures obtained by MC simulations of the energy expression in thousands of atoms allow us to analyze in real- and k- space the short and long-ordering. Second. Understand the relationship between atomic-vacancy arrangements and phonon transport. For that, we collect a large data set of forces and stresses obtained from several configurations, where each configuration includes a couple of hundred atoms interactions. Importantly, the data have their origin from ab-initio molecular dynamics which initial structures were suggested by the Monte Carlos simulations at different temperatures and Nb-concentrations. Using this large data set we create effective potentials for the atomic interactions that allow us to calculate lattice thermal conductivity after performing classical molecular dynamics.

Oxidation states of binary oxides from data analytics of the electronic structure

Olga Miroshnichenko[1], Sergei Posysaev[1], Matti Alatalo[1], Duy Le[2], Talat S. Rahman[2, 3]

[1] *Nano and Molecular Systems Research Unit, University of Oulu, Finland*
[2] *Department of Physics, University of Central Florida, USA*
[3] *Department of Applied Physics, Aalto University, Finland*

Oxidation state (OS) is a concept used in many fields of chemistry and physics. Computational techniques for determining OS are demanding calculations, available in a limited number of codes. Bader charge analysis [1], based on electron density partitioning, might be an easier tool for OS determination. The problem is that connection between the oxidation state (OS) and Bader charge has been missing so far. Several studies have tried to establish the connection, and it has proven to be a nontrivial task.

Nowadays the popularity of open and big data gives rise to continuously growing open material databases with DFT results for millions of compounds. Examples of such databases are: AFLOW [2], NOMAD [3] and others. Our aim is to find a dependency between OS and Bader charge, using a large number of compounds from an open database AFLOW, and to provide a possibility to determine the oxidation state from ordinary DFT calculations. Our focus is on binary oxides, i.e. compounds of two elements, one of which is oxygen. They were chosen for simplicity as they contain only two species. Additionally, binary oxides are abundant on Earth and they are among the most calculated compound types in material databases. To our knowledge, all previous work tried to connect OS with Bader charges only in few compounds. We show that a correlation indeed exists between OSs and Bader charges and discuss the applicability of determining OS by Bader charges in mixed-valence compounds [4].

[1] Bader charge analysis code, http://theory.cm.utexas.edu/henkelman/code/bader/

[2] S. Curtarolo et al., Computational Materials Science 58, 227 (2012).

[3] M. Scheffler, C. Draxl, The NoMaD Repository, (2014).

[4] S. Posysaev, O. Miroshnichenko, M. Alatalo, D. Le, and T.S. Rahman, Computational Materials Science, 161, 403, (2019).

Deep Learning For Predicting and Analyzing Atomistic Systems

Kim Andrea Nicoli[1], Kristof Schütt[1], Pan Kessel[1], Michael Gastegger[1], Alexander Tkatchenko[2],
Klaus Robert Müller[1, 3]

[1] *Machine Learning Group, Technische Universität, Berlin (Germany)*
[2] *Physics and Materials Science Research Unit, University of Luxembourg, Luxembourg*
[3] *Department of Brain and Cognitive Engineering, Korea University, Seoul (Republic of Korea)*

The computational cost for exploring the chemical world at quantum scales constitutes a considerable challenge. In recent years, machine learning, and in particular deep learning, has found wide-spread and successful applications in quantum chemistry. Recent architectures, e.g. wACSF[1], DTNN[2], MPNN[3] and SchNet[4] , led to accurate prediction of chemical properties from reference calculations. Here, we present SchNetPack[5]: a toolbox based on the PyTorch deep learning framework that allows to efficiently train different types of atomistic networks.It provides access to common benchmark datasets and fundamental building blocks to quickly prototype and evaluate customized models. Using this framework, we implemented scSchNet[6], an extension of SchNet with skip layer connections between interaction blocks. The presence of such connections enables the analysis of the relative importance of each level of interaction for both property prediction and molecular dynamics. Our analysis shows that, despite a similar performance in terms of prediction accuracy, the relative importance of each interaction layer strongly depends on the chemical composition and configurational degrees of freedom of the molecule.

[1] M. Gastegger et al.WACSF — Weighted Atom-Centered SymmetryFunctions as Descriptors in Machine Learning Potentials, The Journalof Chemical Physics 148, 28 Jun 2018
[2] K.T. Schütt et al.,Quantum-chemical insights from deep tensor neuralnetworks, Nat. Comm. 8, 1389, Jan 2017.
[3] Gilmer J. et al.,Neural Message Passing for Quantum Chemistry, ICML2017, 12 Jun 2017.
[4] K.T. Schütt, H.E. Sauceda, P.-J. Kindermans, A. Tkatchenko, K.-R.Müller,SchNet – A deep learning architecture for molecules and mate-rials, The Journal of Chemical Physics 148, 29 Mar 2018.

[5] K.T. Schütt, P. Kessel, M. Gastegger, K.A. Nicoli, A. Tkatchenko, K.R.Müller,SchNetPack: A Deep Learning Toolbox For Atomistic SystemsJournal of chemical theory and computation 15, 1, 448–455 2018

[6] K.A. Nicoli, P. Kessel, M. Gastegger, K.T. Schütt,Analysis of AtomisticRepresentations Using Weighted Skip-ConnectionsMachine Learning forMolecule and Materials (NeurIPS 2018)

A machine learning approach to predict redox potentials of organic materials for energy applications

Rodrigo P. de Carvalho[1], Cleber F. N. Machiori[1], Daniel Brandell[2], C. Moyses Araujo[1]

[1] *Materials Theory Division, Department of Physics and Astronomy, Uppsala University, Sweden*
[2] *Department of Chemistry - Ångström Laboratory, Uppsala University, Sweden*

Energy consumption in our planet is expected to face drastic increase in the coming years demanding a novel supply of energy that is secure, clean and sustainable [1,2]. From energy harvesting devices [3] to energy storage systems [4], organic materials have been rising as a promising alternative for the next generation of energy related technologies. However, the proper design of new organic based materials demands the full understanding of their physical and chemical properties. In this regard, we have been developing a database of organic materials for energy applications that contains information about molecular geometries and high-level features extracted from DFT calculations. Based on this database, we present in this work the performance of a machine learning approach to predict the redox potentials by giving only chemical species and molecular structures, completely by-passing a set of DFT calculations needed to compute the oxidation and reduction potentials. These potentials are of fundamental importance for energy applications since they can describe energy transfer in chemical reactions[5]. A further investigation of learning algorithms and molecular representations like Coulomb matrices and the many-body tensor representation [6] has been done to benchmark our machinery. This study reveals the possibility to have a first assessment on the redox activity of organic materials, creating a framework that can aid the designing of new materials for energy applications.

[1] Pain, S. (2017). Nature, 551(7682), S134-S137

[2] Lewis, N. S., & Nocera, D. G. (2006). Proceedings of the National Academy of Sciences, 103(43), 15729-15735

[3] Schultz, D. M., & Yoon, T. P. (2014). Science, 343(6174), 1239176

[4] S. Renault, V. A. Oltean, C. M. Araujo, A. Grigoriev, K. Edström, D. Brandell, Chem. Mater. 2016, 28, 1920

[5] Steinfeld, J. I., Francisco, J. S., & Hase, W. L. (1989). Chemical kinetics and dynamics (Vol. 3). Englewood Cliffs (New Jersey): Prentice Hall

[6] Huo, H., & Rupp, M. (2017). arXiv:1704.06439, 13754-13769

Prediction of Octahedral Deformation in Perovskite by Applying to Machine Learning

Heesoo Park[1], Raghvendra Mall[2], Fahhad H Alharbi[1, 3], Stefano Sanvito[4], Nouar Tabet[1, 3],
Halima Bensmail[2], Fedwa El-Mellouhi[1]

[1] *Qatar Environment and Energy Research Institute, Hamad Bin Khalifa University, Qatar*
[2] *Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar*
[3] *College of Science and Engineering, Hamad Bin Khalifa University, Qatar*
[4] *School of Physics, AMBER and CRANN Institute, Trinity College, Ireland*

Since many research reports had proposed low-temperature solution processed methylammonium lead triiodide ($MAPbI_3$) as the absorber for solar cells, we have witnessed that the power-conversion-efficiency (PCE) of the solar cell has improved competitively. However, the poor intrinsic instability and toxicity of $MAPbI_3$ are still the obstacles against its usage in daily life. The efforts to stabilize the perovskites in a solar-cell often happened at the expense of PCE reduction suggesting that we need to put more efforts in this direction. In this work, we perform high-throughput density functional theory (DFT) calculations exploring ABC_3 chalcogenide perovskites (I-V-VI_3) and halide perovskites (I-II-VII_3) by including alkali metal elements as well as organic molecules at the A sites. We paid particular attention to structures featuring 3-dimensional BC_6 octahedral networks because the A-site cations play a significant role in determining the symmetry of the crystal. We trained and validated the performance of five different machine learning algorithms for the prediction of the octahedral deformation enabling us to correlate the accuracy of the descriptors of the formability of a perovskite structure.

Acknowledgements: This work is supported by the Qatar National Research Fund (QNRF) through the National Priorities Research Program (NPRP8-090-2-047). Computational resources have been provided by the research computing team at Texas A&M University at Qatar.

First-principles study of a vertical spin switch in atomic scale two-dimensional platform

Mihir Ranjan Sahoo[1], Ranjit Pati[2], Saroj Kumar Nayak[1]

[1] School of Basic Sciences, Indian Institute of Technology Bhubaneswar, India
[2] Department of Physics, Michigan Technological University, Michigan, USA

High in-plane charge carrier mobility and long spin diffusion length makes graphene a unique material for spin-based devices. Here, a vertical spin switch has been designed by embedding a single layer of graphene as a tunnel layer between the Ni (111) substrate. Periodic density functional approach in conjunction with Julliere's model is used to calculate the tunnel magnetoresistance (TMR). Further, single-layered hexagonal Boron Nitride (h-BN) is sandwiched between the graphene and Ni (111) substrate to understand the role of hybridization at the interface on TMR. Our calculation shows that in contrast to the graphene junction, a much higher TMR value is obtainable in the case of the graphene/h-BN multi-tunnel junction (MTJ). The TMR in graphene junction is found to decrease with the increase of an externally applied electric field, and drops to zero for a field greater than equal to 0.14 eV/Å. Similar phenomenon was observed in the case of h-BN/graphene MTJ, where TMR value remains unchanged for electric field up to 0.1 eV/Å beyond which it drops to zero. The change in hybridization and charge-carrier-population at the interface modifies the magnetic exchange interaction and magnetic anisotropy resulting in a spin flip at interface, leads to rapid drop in TMR after a threshold electric field. The high and tunable TMR value suggests h-BN assisted high performance graphene based vertical spin switch.

Machine learning plastic deformation of crystals

Henri Salmenjoki[1], Mikko J. Alava[1], Lasse Laurson[1, 2]

[1] *Department of Applied Physics, Aalto University, Finland*
[2] *Computational Physics Laboratory, Tampere University, Finland*

The irreversible changes in crystalline materials during plastic deformation are governed by the motion of dislocations – the line defects of the crystal structure. Experiments have shown that plastic deformation progresses through bursty events and in micron-scale samples these avalanches cause the stress response to fluctuate drastically from sample to sample. Here our aim was to study if the stress response of such sample could be predicted from the initial dislocation configuration. We simulated a crystalline solid with a two-dimensional discrete dislocation dynamics (2D DDD) model. The initial systems were then characterized with a set of inputs that we fed to a neural network predicting the stress for certain strain. Although the predictability vs strain evolved in a non-monotonic fashion as it was affected by the dislocation avalanches, the performance of the network with large strains was promising. Additionally, the predictability improved when the system size was increased.

Multi-scale modelling of Ti-MoS_2 compounds

Andrea Silva[1], D. Kramer[1], T. Polcar[1,2]

[1] *University of Southampton*
[2] *Czech Technical University in Prague*

Recent studies have highlighted the remarkable graphene-like properties of transition metal dichalcogenides (TMDs) and how their chemical composition make them suitable for functionalization and wide range of applications [1,2]. In particular, MoS_2-Ti coatings are known to yield excellent tribological results in industrial testing [3] and its tribological properties has been recently characterised by means of atomistic simulations [4]. Despite the great interest and numerous applications of this material, an understating of its thermodynamically stable structure and phase diagram is lacking. In this work we makes a step in the direction of filling this knowledge gap. In order to address the challenging task of determining the phase-stability of a new compound, we map energy landscapes obtained with DFT onto a cluster-expansion Hamiltonian and iteratively search for low energy orderings of the atoms inside the given host through Monte Carlo simulations. This methodology allows us to explore the Ti-Mo-S phase space, identify stable configurations of the form $Ti_xMo_{1-x}S_2$, quantify miscibility gaps and competition with ternary oxides as function of temperature. Moreover we study the effect of dimensionality on the stability of the system by simulating both 3D bulk structures and isolated 2D layers.

[1] Nat. Chem. 5, 263 (2013)

[2] Nat. Nanotechnol. 7, 699 (2012)

[3] Surf. Coatings Technol. 228, 275 (2013)

[4] Inorg. Chem. 54, 5739 (2015)

Limits of using experimental labeling of materials on the ML development: a case study on aqueous solubility of compounds

Murat Cihan Sorkun[1, 2], Abhishek Khetan[1, 2], Süleyman Er[1, 2]

[1] *DIFFER-Dutch Institute for Fundamental Energy Research, De Zaale 20, 5612 AJ Eindhoven, The Netherlands*
[2] *Center for Computational Energy Research, DIFFER - Dutch Institute for Fundamental Energy Research, De Zaale 20, 5612 AJ Eindhoven, the Netherlands*

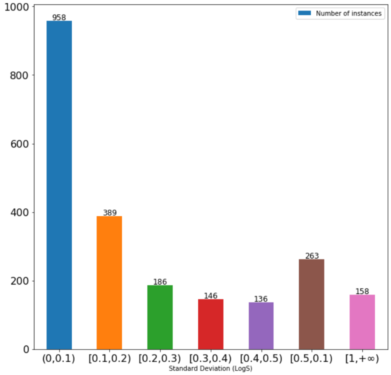


Figure 1: Distribution of number compounds according to the standard deviation of data from different experiments.

For many years, experimental methods have been used to measure the properties of materials. Today, many ML algorithms are trained with data that is labeled by experimental methods. Even though experimental methods are considered as ground truth, they also can have significant error margins. Experimental errors are barriers to the accuracy of ML models. In this study, we analyzed experimental errors that are found in the literature on aqueous solubility of compounds. For this purpose, we studied nine different publicly available aqueous solubility datasets. In these datasets, we compared different experimental measurements for the same compounds. We found that 2,236, a surprisingly large number, of compounds had at least two different experimental values reported. We analyzed the distribution and deviation of experimental values in detail and found that more than half of the total instances contained remarkable differences for the same compounds (Figure 1). We outline the effects of the experimental errors on the development of accurate aqueous solubility prediction models. We propose new methods to overcome the experimental error effect on the aqueous solubility of compounds.

Towards Efficient Electronic Spectra Modelling with Machine Learning

Štěpán Sršeň[1], Petr Slavíček[1]

[1] *Department of Physical Chemistry, University of Chemistry and Technology, Prague, Czech Republic*

We show application of various machine learning approaches to electronic spectra modelling based on the path integral molecular dynamics and the so-called reflection principle.[1] In electronic spectroscopy we are interested in excitation energies and transition probabilities. We use kernel ridge regression to obtain these excitation properties efficiently. We compare different molecular representations and select the which are most suitable for prediction of these global intensive properties. We also propose some modifications to existing representations to make them more efficient for spectroscopic purposes. In order to reconstruct continuous spectrum from single point calculations, we need to use some kind of probability density estimation in the reflection principle.[2] A non-parametric method should be used if the form of the underlying probability density function is not known. We present here a novel approach in which we iteratively combine kernel density estimation with density-based soft clustering in order to calculate the broadening hyperparameter (so-called bandwidth) for each peak separately. We also use the approach in combination with simulated annealing in order to reduce the number of data points which are needed for the modelling. Financial support from specific university research (MSMT No 21-SVV/2019) and support of Czech Science Foundation project No. 17-04068S are gratefully acknowledged.

[1] M. Ončák, L. Šištík, P. Slavíček, Can Theory Quantitatively Model Stratospheric Photolysis? Ab Initio Estimate of Absolute Absorption Cross Sections of ClOOCl. J. Chem. Phys., 133, 174303, 2010

[2] Š. Sršeň, D. Hollas, P. Slavíček, UV Absorption of Criegee Intermediates: Quantitative Cross Sections from High-Level Ab Initio Theory. Phys. Chem. Chem. Phys., 20, 6421–6430, 2018

Combining polarizable electrostatic with a machine learned potential

Carsten Staacke[1], Johannes Margraff[1], Christoph Scheurer[1], Alan Luntz[2], Karsten Reuter[1]

[1] *Theoretical Chemistry, TU München, Germany*
[2] *SunCat Center, Stanford University, USA*

All-solid-state Li-ion batteries promise gains in safety and durability by replacing the liquid by a solid-state electrolyte (SSE). In this context, SSEs from the xLi_2S $(1-x)P_2S_5$ (LPS) class, especially $Li_7P_3S_{11}$, have gained much attention. While $Li_7P_3S_{11}$ displays low conductivity in the amorphous glass phase, a meta stable glass-ceramic state shows very promising performance [1]. To understand this large difference, a detailed atomistic description of both phases is necessary. This is not directly accessible from experiment. Unfortunately, the material complexity and timescales involved in realistic glass-ceramic SSE simulations precludes the use of density-functional theory (DFT) based molecular dynamics. We have therefore engaged in the development and first-principles validation of a polarizable interatomic potential for the LPS SSE. In particular, the combined description of long-range electrostatic interactions via a classical core-shell model and short-range covalent interactions via a machine-learning based Gaussian Approximation Potential (GAP) [2] are discussed and evaluated.

[1] A. Hayashi, A. Sakuda, M. Tatsumisago, Front. Energy Res.,2016, 4, 25

[2] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi Phys. Rev. Lett.,2010, 104, 136403

[3] CoA. P. Bartók, M. C. Payne, R. Kondor, G. Csányi Phys. Rev. Lett., 2010, 104, 136403

Predicting Reaction Energetics with Machine Learning

Sina Stocker[1], Johannes T. Margraf[1], Karsten Reuter[1]

[1] *Technische Universität München, Germany*

Predictive-quality first-principles based microkinetic models are increasingly used to analyze (and subsequently optimize) reaction mechanisms in heterogeneous catalysis. In full rigor such models require the knowledge of all possible elementary reaction steps and their corresponding reaction barriers. Unfortunately, for complex catalytic processes, such as the generation of synthetic fuels out of syngas, the number of such steps becomes so large that an exhaustive first-principles calculation of all barriers becomes prohibitively expensive.

As a remedy, we explore the possibility of machine learning (ML) approaches to the prediction of the reaction energetics. An essential component in such data-driven approaches are efficient molecular representations (descriptors). We test a range of such representations that have been suggested to describe properties of closed-shell molecules and specifically assess their capabilities in describing open-shell systems and consequently reaction energetics. The obtained overall promising performance confirms the potential of ML approaches for a high-throughput screening of elementary steps in large reaction networks.

Performance of various fingerprints in analyzing atomic environments

Behnam Parsaeifard[1]

[1] *Physics dept, Universitaet Basel, Switzerland*

We introduce the overlap matrix descriptor (OM) which is based on the overlap between atomic orbitals in the system. We compare the OM descriptor with other well-known descriptors namely Behler symmetry functions and SOAP descriptors by presenting correlation plots between similarity/dissimilarity quantities based on these descriptors. We show that the OM descriptor can well distinguish different atomic environments and demonstrate automatic grain boundary detection as an application. Finally we show its good performance in machine learning schemes.

Force Field Parameterization of Halide Anions by means of a Machine Learning Procedure

Matteo Peluso[1], Francesco Fracchia[1], Giordano Mancini[1], Vincenzo Barone[1]

[1] *Scuola Normale Superiore, Pisa Italy*

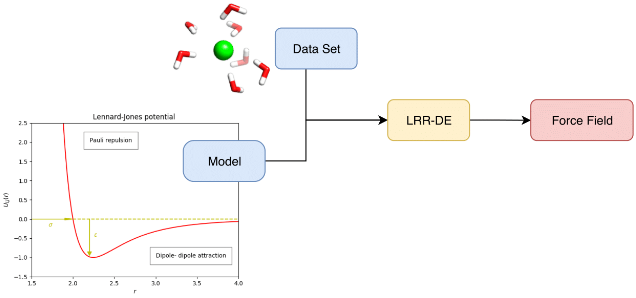


Figure 1: Force Field Parameterization

In Molecular Mechanics, the definition of an accurate, yet computationally non onerous force field and the relative parameters, is crucial for a correct description of the interaction within a system as it directly affects all the properties extracted from the simulation analysis. In my research group has been developed a tool [1] to extract, from quantum mechanical calculations, information to describe the inter-molecular components of a force field for metal ions in condensed phase. In this work, the parameter optimization of the inter-molecular components of a force field of the halide ions, Cl-, Br-, F-, I-, has been addressed with the aim of generalizing the procedure to a broader range of atomic environment. We exploited the well validated linear ridge regression differential evolution fitting recipes. The resulting parameters, thus obtained, despite the limits of the physic model describing such a chaotic environment, well reproduce the experimental observable.

[1] F. Fracchia, G. Del Frate, G. Mancini, W. Rocchia, V. Barone, Force Field Parametrization of Metal Ions from Statistical LearningTechniques, Journal of Chemical Theory and Computation.

Neural Network for fast material characterisation in Friction Force Microscopy

Paola Torche[1], Tomas Polcar[1], Ondrej Hovorka[1]

[1] *Faculty of Engineering and the Physical Sciences, University of Southampton, United Kingdom*

Friction plays a fundamental role in energy loses, functioning, and lifespan of many technological applications. Nanoscale friction is measured in Friction Force Microscopy by recording the force of a substrate on a sharp tip as it is being dragged tangent to the substrate surface. The measured force is ideally fully determined by the physical parameters of the surface and measurement system (lattice spacing, potential of interaction of the tip-substrate, elasticity constant of the dragging device, etc), however, stochasticity is also present in the measurement due to the small scales.

We study the use of Machine Learning to learn the patterns in stochastic measurements of nanoscale friction, and relate the measured force with the physical parameters of the surface. We use synthetic measurements produced by the thermal Prandlt-Tomlinson model for training a Neural Network. The aim is the designing of a tool that allows for a fast characterisation of certain physical properties (e.g. lattice spacing, potential of interaction) from Friction Force Microscopy measurements, as an alternative to individual computationally expensive fittings to molecular-level models.

[1] Gnecco, E., Roth, R., & Baratoff, A. (2012). Analytical expressions for the kinetic friction in the Prandtl-Tomlinson model. Physical Review B - Condensed Matter and Materials Physics, 86(3), 1–6. https://doi.org/10.1103/PhysRevB.86.035443

[2] Jinesh, K. B., Krylov, S. Y., Valk, H., Dienwiebel, M., & Frenken, J. W. M. (2008). Thermolubricity in atomic-scale friction. Physical Review B - Condensed Matter and Materials Physics, 78(15), 1–12. https://doi.org/10.1103/PhysRevB.78.155440

Looking for Needles in the Haystack: Improving Approximate Electronic Structure Methods via a Machine-Learning Correction for Reliable Crystal Structure Screening

Simon Wengert[1], Chiara Panosetti[1], Johannes T. Margraf[1], Karsten Reuter[1]

[1] *Technische Universität München, Germany*

Still largely a Holy Grail of modeling, the capability to reliably predict the structure of (molecular) crystals would allow for the in silico design of substances with desired characteristics. The main challenges within the process of crystal structure prediction (CSP) are the large number of possible polymorphs and the small energy differences between them. In practice, there is thus a trade-off between the ability to screen a wide range of candidates (which requires some fast evaluation of stability) and applying levels of theory that adequately describe the subtle interplay between intermolecular interactions such as H-bonding and dispersion effects. Dispersion-corrected density-functional tight binding (DFTB+TS) represents a numerically efficient method for high-throughput screening. The underlying approximations nevertheless lead to an inadequate description of many-body effects in both electronic structure and dispersion that can jeopardize a reliable stability ranking of crystals polymorphs. We therefore strive to account for these many-body effects via a machine-learning based correction term to DFTB+TS. To this end, we generate a training set that covers the different space-groups and crystal environments that are sampled during a CSP process. The correction is then trained based on relative energies, forces and stresses obtained from a higher-level theory that accounts for the many-body effects, here density-functional theory plus a many-body dispersion correction.

Workflow Platform Framework and its Application

Jingfang Xiong[1], Jianxing Huang[1], Jun Cheng[1]

[1] *College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, China*

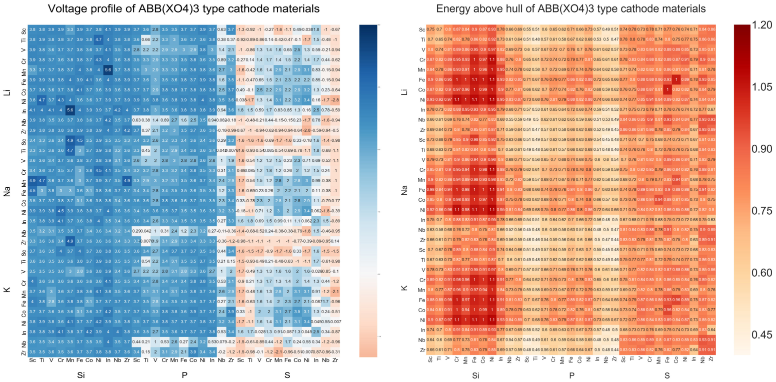


Figure 1: Main data about ABB(XO4)3 type cathode materials

In June 2011, American President Obama announced the launch of Materials Genome Initiative(MGI). MGI want to build a database to explore the relationship between structures and features. In order to get more data to support MGI, we need find a method to produce data reasonably and efficiently.

Calculation workflows are good way to get properties of materials and supply data. We can calculate and collect data automatically and high-throughput by workflows. Thus, we decided to choose atomate[1] as the base of our workflow platform.

We also used our platform to scan 1188 cathode materials that are based on $ABB(XO_4)_3$ type (in which, A = Li, Na, K; B = Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Nb, In, Zr; X = S, P, Si). And we find something interesting. When X changes from Si to P to S, the voltage is reducing in general, but when X equals P, it has the best thermodynamic stability on the whole. However, when A changes from Li to Na to K, there is no significant change in the two indicators.

[1] Mathew, K., Montoya, J. H., Faghaninia, A., Dwarakanath, S., Aykol, M., Tang, H., Chu, I., Smidt, T., Bocklund, B., Horton, M., Dagdelen, J., Wood, B., Liu, Z.-K., Neaton, J., Ong, S. P., Persson, K., Jain, A., Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. Comput. Mater. Sci. 139, 140-152 (2017).

High-throughput investigation of band alignments of cubic perovskites in solvated environments

Yihuang Xiong[1], Weinan Chen[1], Wenbo Guo[2], Hua Wei[2], Ismaila Dabo[1]

[1] *Department of Materials Science and Engineering, Materials Research Institute, The Pennsylvania State University*

[2] *College of Information Sciences and Technology, The Pennsylvania State University*

Discovering novel semiconductor photocatalyst that is capable of driving the overall water-splitting reaction at the electrochemical interfaces is of significant research interest. Among several catalytic performance descriptors, band alignment across the semiconductor-solution interfaces exerts primary influence over material's water-splitting capabilities. With the development of the density-functional theory under the framework of self-consistent continuum method (SCCS), we can now efficiently calculate the band alignment taking into account the electrolytic environmental effect. Thus, we conducted a high-throughput investigation of more than 1600 interface structures of cubic Perovskite structures in the vacuum and solvated environment. Our results highlight that under the implicit solvation schemes, both the solvation effect and the interface structures strongly affect the flat-band potentials. Furthermore, we utilize a descriptor-based machine learning method to predict band alignments. The model not only delivers a robust prediction but also revealed the correlation between atomic descriptors and interfacial properties. By combining those outputs with the DFT-derived results, we aim to understand the underlying electrochemical correlations that could emerge from the statistical learning model.

Index

Abrikosov, 61
Alahyarizadeh, 57
Alatalo, 71
Alava, 76
Alldritt, 23
Allegra, 35
Andersen, 47, 50
Araujo, 73
Artrith, 9
Attfield, 30

Baksa, 43
Baletto, 52
Barone, 83
Bekas, 14
Bellucci, 44
Benavides Riveros, 32
Bensmail, 74
Besel, 25, 45
Bolgar, 14
Brandell, 73
Brotzakis, 16
Byggmästar, 46, 57
Bäumer, 64

Camino, 54
Ceriotti, 12, 60
Chakraborty, 53
Chen, 87
Cheng, 65, 86
Choudhary, 24
Cingolani, 47
Colibaba, 48
Cottenier, 17, 63
Cumby, 30

Dabo, 43, 87
Dedecker, 49
Deimel, 50
Delfino, 51
Delgado Callicó, 52
Demuynck, 49
Dhaene, 63
Dimitrov, 62
Djurabekova, 46, 57
Dutta, 16

Egger, 18
El-Mellouhi, 74
Er, 78

F. N. Machiori, 73
Federici Canova, 23, 58, 59
Foster, 23, 58, 59
Fracchia, 83
Franco, 37

Ganguli, 53
Gao, 58, 59, 66
Gastegger, 72
Gaudin, 14
Ghader, 57
Glielmo, 13
Gomez-Bombarelli, 26
González, 54
Guo, 87

H Alharbi, 74
Hajiyani, 55
Haldar, 56
Hamedani, 46, 57
Hammerschmidt, 33
Hapala, 23
Heinze, 56
Himanen, 58, 59
Hofmann, 18
Hoock, 37
Hovorka, 84
Huang, 65, 86
Hörmann, 18

Isayev, 40

Jacobsen, 29
Jeindl, 18
Jenkins, 67
Jäger, 58, 59
Järvi, 19

Kannala, 23
Kapil, 60
Karavaev, 61
Kessel, 72
Khetan, 78
Kichev, 62
Kramer, 77
Krejci, 23
Kubecka, 45
Kubečka, 25
Kucukbenli, 39
Kurtén, 25, 45

Laino, 14
Larmuseau, 17, 63
Laurson, 76
Le, 71
Lee, 14
Li, 64
Liljeroth, 23
Liu, 65
Lombardo, 37
Long, 66
Lot, 39
Luca Ghiringhelli, 8
Lunghi, 38
Luntz, 80

Mall, 74
Mancini, 83
Margraf, 81, 85
Margraff, 80
Marques, 32
Matysik, 67
Mazumder, 68
McDermott, 69
Mira, 16
Miranda, 70
Miroshnichenko, 71
Morooka, 58, 59
Moskaleva, 64
Müller, 72

Nayak, 75
Nedyalkova, 62
Nelson, 31
Ngandjong, 37
Nicoli, 72
Nordlund, 46, 57

Oinonen, 23
Olsson, 36
Opahle, 66

P. de Carvalho, 73
Panosetti, 85
Park, 74
Parsaeifard, 82
Pati, 68, 75
Pellegrini, 39
Peluso, 83
Pentcheva, 55
Polcar, 77, 84
Posysaev, 71

Rahman, 71
Remes, 15
Reuter, 47, 50, 80, 81, 85
Rimola, 54
Rinke, 19, 58
Rodríguez-Santiago, 54
Rogge, 49
Roman, 27
Roos, 20
Rossi, 52
Roth, 22

Sahoo, 75
Salmenjoki, 76
Sanvito, 31, 38, 48, 74
Scheurer, 80
Schmidt, 32
Schultz, 23
Schwaller, 14
Schütt, 72
Shaidu, 39
Silva, 77
Simeonov, 62
Slavíček, 79
Sluydts, 17, 63
Smirnova, 61
Sodupe, 54
Solans-Montfort, 54
Sorkun, 78
Sosso, 57
Sršeň, 79
Staacke, 80
Stocker, 81

Tabet, 74
Tkatchenko, 72
Todorovic, 19
Torche, 84
Tozzini, 51

Urtev, 23

Valpola, 28
Van Speybroeck, 49
VandeVondele, 60
Vanduyfhuys, 49
Vehkamäki, 25, 45

Waroquier, 49
Wei, 87
Wengert, 85

Xiong, 86, 87

Zeni, 13
Zhang, 66